

**Semantic Multimedia
Modelling & Interpretation for
Search & Retrieval**



**Middlesex
University**

Nida Aslam

Submitted in partial fulfilment of the
Requirements for the degree
of Doctor of Philosophy

School of Engineering and Information Sciences

MIDDLESEX UNIVERSITY

May- 2011

© 2011

Nida Aslam

All rights reserved

Declaration

I certify that the thesis entitled Semantic Multimedia Modelling and Interpretation for Search and Retrieval submitted for the degree of doctor of philosophy is the result of my own work and I have fully cited and referenced all material and results that are not original to this work.

Name NIDA ASLAM

Date MAY 20th, 2011

Abstract

“A photograph is a secret about a secret. The more it tells you the less you know.”
Diane Arbus, photographer, 1923 – 1971

With the axiomatic revolutionary in the multimedia equip devices, culminated in the proverbial proliferation of the image and video data. Owing to this omnipresence and progression, these data become the part of our daily life. This devastating data production rate accompanies with a predicament of surpassing our potentials for acquiring this data. Perhaps one of the utmost prevailing problems of this digital era is an information plethora.

Until now, progressions in image and video retrieval research reached restrained success owed to its interpretation of an image and video in terms of primitive features. Humans generally access multimedia assets in terms of semantic concepts. The retrieval of digital images and videos is impeded by the semantic gap. The semantic gap is the discrepancy between a user’s high-level interpretation of an image and the information that can be extracted from an image’s physical properties. Content- based image and video retrieval systems are explicitly assailable to the semantic gap due to their dependence on low-level visual features for describing image and content. The semantic gap can be narrowed by including high-level features. High-level descriptions of images and videos are more proficient of apprehending the semantic meaning of image and video content.

It is generally understood that the problem of image and video retrieval is still far from being solved. This thesis proposes an approach for intelligent multimedia semantic extraction for search and retrieval. This thesis intends to bridge the gap between the visual features and semantics. This thesis proposes a Semantic query Interpreter for the images and the videos. The proposed Semantic Query Interpreter will select the pertinent terms from the user query and analyse it lexically and semantically. The proposed SQI reduces the semantic as well as the vocabulary gap between the users and the machine. This thesis also explored a novel ranking strategy for image search and retrieval. SemRank is the novel system that will incorporate the Semantic Intensity (SI) in exploring the semantic relevancy between the user query and the available data. The novel Semantic Intensity captures the concept dominancy factor of an image. As we are aware of the fact that the image is the combination of various

concepts and among the list of concepts some of them are more dominant than the other. The SemRank will rank the retrieved images on the basis of Semantic Intensity.

The investigations are made on the LabelMe image and LabelMe video dataset. Experiments show that the proposed approach is successful in bridging the semantic gap. The experiments reveal that our proposed system outperforms the traditional image retrieval systems.

Acknowledgment

*"I can no other answer make but thanks, and thanks, and ever thanks,"
William Shakespeare, Twelfth Night, Act III,
scene III*

I was once told that PhD is a long journey of transformation from a 'novice' to 'professional' researcher. To succeed, an individual can never do it alone; there must be someone who is there for them, providing all the helping hands and supports. I can't agree more and as for my case, I have a lot of people to thank.

I owe my gratitude to all those people who have made this thesis possible. It is a pleasure to thank them all in my humble acknowledgment. In the first place, I am heartily thankful to my supervisor Dr. Jonathan Loo (Middlesex University) for helpful discussions, guidance and support throughout my PhD studies. His scientific intuition and rigorous research attitude have enriched me with the in-depth understanding of the subject and the passion of dedicating myself to research, which I will benefit from in the long run. His gracious support in many aspects has enabled me to overcome difficulties and finish my study. I also extend my thanks to my second supervisor Dr. Martin Loomes.

Further, I would like to express my special gratitude to my colleague and best friend Irfan Ullah whose dedication, love, and persistent care has diminished all the difficulties of studying and living in the unfamiliar land far away from home.

I am grateful to my financial supporters, including the Kohat University of Science & Technology, Pakistan. Many thanks also go to my friends shafi Ullah khan, Tahir Naeem who accompanied me during the last three years, making my time in London most enjoyable. Finally, I pay tribute to the constant support of my family, but especially of my parents, Brothers, sisters and my uncle for their untiring efforts and praying. Without them, my whole studies would have been impossible and whose sacrifice, I can never repay. This one is for you!

Dedication

I dedicate this thesis to my late Grandmother, (may her soul rest in peace) for their unconditional support and prayers, tireless love and motivation throughout my studies.

Everything I have accomplished, I owe to her.

List of associated Publications

1. Nida Aslam, Irfan Ullah, Jonathan Loo, RoohUllah " A Semantic Query Interpreter Framework by using Knowledge Bases for Image Search and Retrieval" IEEE International Symposium on Signal Processing and Information Technology December 15-18, 2010 - Luxor – Egypt.
2. Nida Aslam, Irfan Ullah, Jonathan Loo, RoohUllah "SemRank: Ranking Refinement Strategy by using the Semantic Intensity" World Conference on Information Technology, 2010.
3. N. Aslam, I. Khan, K.K. Loo, "Limitation and Challenges: Image/Video Search & Retrieval", International Journal of Digital Content Technology and its Applications (JDCTA), ISSN: 1975-9339, Vol. 3, No. 1, March 2009.
4. N. Aslam, I. Khan, K.K. Loo, "Growing Trend from Uni to Multimodal Video Indexing", International Journal of Digital Content Technology and its Applications (JDCTA), ISSN : 1975-9339, Vol. 3, No. 2, June 2009.
5. Nida Aslam, Irfanullah, Jonathan Loo, Roohullah. "Automatic Semantic Query Interpreter (SQI) for Image Search and Retrieval". Springerlink International Journal on Digital Libraries (Submitted Feb 2011).
6. Nida Aslam, Irfanullah, Jonathan Loo, Roohullah. "Exploiting the knowledgebases for Semantic Query Expansion for the Video Retrieval". Elsevier Journal of King Saud University – Computer and Information. (Submitted Feb 2011).
7. Nida Aslam, Irfanullah, Jonathan Loo, Roohullah. "Exploiting the Semantic Intensity for Semantic Ranking of Images". Elsevier Journal of Visual Communication and Image Representation (Submitted Feb 2011).
8. Nida Aslam, Irfanullah, Jonathan Loo, Roohullah. A survey on the Content based Image and Video Retrieval
9. Nida Aslam, Irfanullah, Jonathan Loo, Roohullah. A survey on Semantic based Image and Video Retrieval.

Table of Contents

List of Figures.....	XIII
List of Tables.....	XVI
List of Acronyms.....	XVIII
Chapter 01 - Introduction.....	1
<i>1.1 Introduction</i>	<i>2</i>
<i>1.2 Motivation and Application</i>	<i>6</i>
<i>1.3 Main Aim and Objectives.....</i>	<i>9</i>
<i>1.4 Problem and Challenges</i>	<i>11</i>
<i>1.5 Research Direction</i>	<i>13</i>
1.5.1 High Level Semantic Concepts and Low-level Visual features.....	13
1.5.2 Human Perception of Image and Video Content.....	14
1.5.3 Image and Video Retrieval.....	15
1.5.4 Diverse Nature of the Bench Mark Datasets.....	15
<i>1.6 Proposed Research Contributions</i>	<i>15</i>
1.6.1 Semantic Query Interpreter for Image Search and Retrieval	16
1.6.2 SemRank: Ranking Refinement by using Semantic Intensity	17
1.6.3 Semantic Query Interpreter for Video Search and Retrieval.....	17
<i>1.7 Organization of the Thesis.....</i>	<i>17</i>
Chapter 02 - Basic Concepts & Literature Review	20
<i>2.1 Information Retrieval</i>	<i>22</i>
<i>2.2 General Multimedia Retrieval.....</i>	<i>23</i>
<i>2.3 Image Retrieval.....</i>	<i>25</i>
<i>2.4 Video Retrieval.....</i>	<i>26</i>
<i>2.5 Approaches for Image and Video Retrieval</i>	<i>27</i>
2.5.1 Text based Image and Video Retrieval.....	28
2.5.2 Content based Image and Video Retrieval.....	32

2.5.2.1	Feature Extraction	35
2.5.2.2	Similarity Measurement	50
2.5.2.3	Query Paradigm.....	52
2.5.2.4	Existing Content Based Retrieval Systems.....	55
2.5.3	Semantic Based Retrieval	63
2.5.3.1	Semantic Concept Detection.....	65
2.5.3.2	Automatic Image and Video Annotation	67
2.5.3.3	Relevance Feedback	68
2.5.3.4	Ontologies for Image and Video Retrieval	70
2.5.3.5	Multi-modality Information Fusion	71
2.5.3.6	Semantic Based Queries Paradigm	73
2.6	<i>Evaluation Measure</i>	74
2.6.1	Precision	76
2.6.2	Recall	77
2.6.3	F-Measure.....	79
2.7	<i>Chapter Summary</i>	80
Chapter 03 - Semantic Query Interpreter for Image Search & Retrieval.....		81
3.1	<i>Introduction</i>	83
3.2	<i>State-of-the-Art.....</i>	86
3.2.1	Probabilistic Query Expansion	88
3.2.2	Ontological Query Expansion.....	89
3.3	<i>Proposed Framework</i>	92
3.3.1	Core Lexical Analysis.....	96
3.3.1.1	Pre-processing.....	96
3.3.1.2	Candidate Term Selection	99
3.3.1.3	Lexical Expansion Module	99
3.4.1	Common Sense Reasoning.....	102
	ConceptNet	103
3.4.2	Candidate Concept Selection	105
3.4.2.1	Semantic Similarity Calculation	106
3.4.3	Retrieval and Ranking of Result.....	110
3.4	<i>Experiments</i>	112
3.5	<i>Chapter Summary</i>	126

Chapter 04 - SemRank: Ranking Refinement Strategy by using the Semantic Intensity...	128
4.1 Introduction	129
4.2 State-of-the-Art Ranking Strategies.....	132
4.2.1 Retrieval Model	133
4.2.1.1 Boolean Model	133
4.2.1.2 Statistical Model.....	135
4.3 Proposed Semantic Ranking Framework.....	147
4.3.1 Semantic Intensity.....	148
4.4 Experimental Study	156
4.5 Chapter Summary	162
Chapter 05 - Semantic Query Interpreter for Video Search & Retrieval.....	164
5.1 Introduction	166
5.2 Video Structure and Representation	167
5.2.1 Shot Boundary Detection.....	169
5.2.2 Key Frame Selection	171
5.2.3 Feature Extraction	172
5.2.3.1 Visual Modality.....	173
5.2.3.2 Audio Modality	174
5.2.3.3 Textual Modality	174
5.3 State of the Art.....	175
5.3.1 Content Based Video Retrieval (CBVR)	176
5.3.2 Video Semantics	178
5.3.3 Query expansion.....	178
5.4 Proposed Contribution.....	180
5.5 Evaluation.....	181
5.5.1 LabelMe Videos	181
5.6 Experimental Setup	182
5.7 Chapter Summary	187
Chapter 06 - Conclusion & Perspectives.....	188
6.1 Introduction	189
6.2 Research Summary.....	189

6.2.1	Semantic Query Interpreter for Image Search and Retrieval	189
6.2.2	SemRank.....	190
6.2.3	Semantic Query Interpreter for Video Search and Retrieval.....	191
6.3	<i>Future Perspective</i>	191
6.3.1	Semantic Query Interpreter extension	192
6.3.2	Semantic Encyclopedia: An Automated Approach for Semantic Exploration and Organization of Images and Videos.....	192
6.3.3	SemRank for Videos.....	194
Appendix		195
Rerences		230

List of Figures

Figure 1.1	Multimedia data growth rate	3
Figure 1.2	Data Production Exponential rate [John et al. 2008]Amount of Digital Information Created and Replicated each year.	4
Figure 1.3	Semantic Gap	6
Figure 1.4	Image is Worth than a thousand of Words. The image can be perceived as an image of lion, sign of Bravery, sign of Fear, image of a zoo, lion in the jungle, lion in the forest, Franklin Park Zoo etc.	8
Figure 1.5	Long tail Problem	11
Figure 1.6	Proposed Research Contributions	15
Figure 2.1	Typical Information Retrieval process	22
Figure 2.2	Progression in the Multimedia Retrieval	23
Figure 2.3	Semantically rich images and ambiguous images	25
Figure 2.4	Multiple interpretation of same images Park like Tree, Sky, Horse, People, Ridding, Sunny Day, Outdoor	28
Figure 2.5 a	Same name different Semantics i.e. jaguar car and the jaguar animal.	30
Figure 2.5 b	Same name different Semantics i.e. Apple a company brand and the name of fruit	30
Figure 2.6	Typical Architecture of Content Based Retrieval	32
Figure 2.7	Colour based image interpretation	35
Figure 2.8	The additive colour model HSV	36
Figure 2.9 (a)	RGB: Additive Colour for light-emitting computer monitors. Each coloured light "add" to the previous coloured lights.	37
Figure 2.9 (b)	CMYK: Subtractive colours for Printer. Each colour added to the first colour blocks the reflection of colour, thus 'subtracts' colour.	38
Figure 2.10	Same Car with different colour composition	39

Figure 2.11	Colour Histogram	40
Figure 2.12	Images with Similar Colour Composition But different Semantics	43
Figure 2.13	Different Colour Composition but Similar Semantic idea	44
Figure 2.14	Various types of Textures	46
Figure 2.15	Different Query Paradigm	52
Figure 2.16	Category of the data that will either retrieved by using the particular retrieval system and the data in the corpus that will not retrieved	74
Figure2.17	Venn diagram for Precision and Recall	76
Figure2.18	Interpretation of precision-recall curves	77
Figure 3.1	Overall Semantic Query Interpreter	91
Figure 3.2	Query Expansion along with semantic similarity by WordNet. The WordNet attaches the synset of the cars like motor car, railway car, railcar, machine, cable car, automobile, railroad car, auto, gondola, elevator car etc. The figure contains the lexical expansion along with the semantic similarity value. As we know motor car relates more with the car that's why its Semantic similarity value is 1. The greater the semantic similarity value greater will be the relevancy degree.	92
Figure 3.3	Query Expansion along with semantic similarity by ConceptNet. The ConceptNet attaches the following concepts with the keyword car like brake, day, drive, front part, good appearance, hood, its head, lane, light, long distance, motor cycle, mountain, other person, person's mother, plane, pollutant, right behing, road trip, bed etc. The figure also contains the conceptual expansion of the car along with the Semantic similarity value. Greater the Semantic similarity value greater will be the relevancy degree. Among the expanded terms some of them are noises that will significantly decrease the precision of the system.	93

Figure 3.4	Query Expansion along with semantic similarity by Semantic Query Interpreter. The Semantic Query Interpreter expansion contains the selected lexical and conceptual expansion of the keyword car. The figure contains the selected expansion terms according to the threshold, and the semantic similarity value between the original query term and the expanded terms.	94
Figure 3.5	Example of synsets and semantic relations in WordNet	98
Figure 3.6	An illustration of a small section of ConceptNet	101
Figure 3.7	Representation of the Vector Space Model	108
Figure 3.8	Precision of five queries of three different types using proposed SQI	113
Figure 3.9	Recall of five queries of three different types using proposed SQI	115
Figure 3.10	F-Measure of five queries of three different types using proposed SQI	116
Figure 3.11	Precision comparison of the LabelMe system, WordNet based expansion, ConceptNet based expansion and the Proposed SQI at the different precision level	125
Figure 4.1	Typical IR Paradigm	131
Figure 4.2	Vector Measure Cosine of theta	136
Figure 4.3	Bayesian Inference IR Model	140
Figure 4.4	Typical Learning to Rank Paradigm	141
Figure 4.5	Both A and B figure represents images of the car. The frequency of the car in the image B is greater than the image A. But the image A depicts the car concept more clearly. Hence image A has a greater relevancy degree than image B even though image B has greater frequency	147
Figure 4.6	The image is taken from the LabelMe dataset. Image depicts a list of concepts like road, vehicles, signs, buildings, sky, trees, umbrella, buildings, street, cross walk, highlight, flags	148
Figure 4.7: (a)	Regular Polygon	149

Figure 4. 7: (b)	IRRegular Polygon	149
Figure 4.8	'CAR' Query Output using VSM and SQI	152
Figure 4.9	Semantic Intensity of the images	153
Figure 4. 10	Comparison of the LabelMe system, Vector Space Model and SemRank	157
Figure 4.11	Precision Recall curve of Single Word Single Concept Queries	158
Figure 4.12	Precision Recall curve of Single Word Multi Concept Queries	159
Figure 4.13	Precision Recall curve of Multi Word Multi Concept Queries	160
Figure 5.1	A hierarchical decomposition and representation of the video contents.	167
Figure 5.2	Analysis of the video contents	169
Figure 5.3	Key frame identification	171
Figure 5.4	Multimodal Video Content	172
Figure 5.5	Typical Content Based Video Retrieval	176
Figure 5.6	Different precision values for the five randomly selected user queries of three different categories on the LabelMe video corpus.	182
Figure 5.7	Different recall values for the five randomly selected user queries of three different categories on the LabelMe video corpus.	184
Figure 5.8	Different F-measure values for the five randomly selected user queries of three different categories on the LabelMe video corpus.	185

List of Tables

Table 3.1	Five randomly selected single word single concept query, the expanded query terms by using the Semantic query Interpreter and the top ten retrieved results.	117
Table 3.2	Five randomly selected single word multi- concept query, the expanded query terms by using the Semantic query Interpreter and the top ten retrieved results	119
Table 3.3	Five randomly selected multi word multi concept query, the expanded query terms by using the Semantic query Interpreter and the top ten retrieved results	121
Table 3.4	Precision comparison of the LabelMe system, WordNet based expansion, ConceptNet based expansion and the Proposed SQI at the different precision level	123
Table 4.1	Comparison of the LabelMe system, Vector Space Model and SemRank at different precision values.	156

List of Acronyms

ASR	Automatic Speech Recognizer
BIR	Binary Independence Model
CBAR	Content-Based-Audio-Retrieval
CBIR	Content-Based-Information-Retrieval
CBIRD	Content-Based Image Retrieval from Digital libraries
CBMR	Content-Based-Multimedia-Retrieval
CBR	Content-Based-Retrieval
CBVIR	Content-Based-Visual-Information-Retrieval
CBVR	Content-Based-Video-Retrieval
CCM	Colour Coherence Matrix
CCV	Colour Coherence Vector
CMY	Cyan, Magenta, Yellow
CMYK	Cyan-Magenta-Yellow-blackK
CORR	Colour Auto Correlogram
CSAIL	Computer Science and Artificial Intelligence Laboratory
FCH	Fuzzy Colour Histogram
FFT	Fast Fourier Transform
HSB	Hue, Saturation, Brightness
HSV	Hue, Saturation, Value
IC	Information Content
IR	Information Retrieval
IRank	Interactive Ranking
JACOB	Just A Content Based
LCS	Least Common Subsumer

LM	LabelMe
LPC	Linear Predictive Coding
LSCOM	Large Scale Concept Ontology Model
LSI	Latent Semantic Indexing
LTR	Learning to Rank
MARS	Multimedia Analysis and Retrieval System
MFCC	Mel Frequency Cepstral Coefficient
MIR	Multimedia-Information-Retrieval
NIST	National Institute of Standards and Technology
NLP	Natural Language Processing
OCR	Optical Character Recognizer
OQUEL	Ontology Query Language
OWL	Web Ontology language
PRP	Probability Ranking Principal
QBIC	Query by Image Content
QPRP	Quantum Probability Ranking Principal
RDF	Resource Description Framework
RGB	Red Blue Green
SemRank	Semantic Ranking
SI	Semantic Intensity
SIMBA	Search IMages By Appearance
SQI	Semantic Query Interpreter
SUMO	Suggested Upper Merged Ontology
SVM	Support Vector Machine
TREC	Text Retrieval Conference
VIPER	Visual Information Processing for Enhanced Retrieval
VSM	Vector Space Model
WSD	Word Sense Disambiguation

ZCR

Zero Crossing Rate

Chapter 01

Introduction

*"We are drowning in information, but starving for knowledge."
JOHN NAISBITT American writer (born 1929)*

We are in an era where the technologies that enable people to easily capture and share digitized video data are rapidly developing and becoming universally available. Personal computers are continually getting faster, smaller, and cheaper, while high speed and reliable networking has shifted towards mobile and wireless access. Gone are the days where a television set and an inane amount of cables was necessary to watch and edit video. To date, handheld devices and the Internet have become a common method to create and transport video documents. As a result, there has been a huge increase in the utilization of videos and images as one of the most preferred types of media due to its content richness for many significant applications. In order to support and maintain video and image growth, further enhancement on the current solutions for Images and Video Retrieval is required.

1.1 Introduction

The prevalence of the digital world in our global village has entrenched the digital media in our lives. We live in the digital revolutionary era, overwhelmed by data overdose. With the rapid progression in the digital technology like digital cameras, handheld devices, mobile phones, digital video recorders and scanners, airborne radars, digital synthesizers and PDA's (Personal Digital Assistant) etc. hand-in-hand with the low cost efficient storage devices have given birth to an unprecedented rate of the increase in the production of images and videos.

Owing to the inexpensive and easy way to acquire images and videos, the average users have massive volume of digital data on their personal computer. The extremely widespread utilization of digital technologies mounts millions of digital data daily. This has had the twofold impact of both decreasing the cost and enabling easy dissemination of digital media which becomes the catalyst for increasing acquisition of digital information in the form of images and videos. Now producing a digital data is at our finger tips.

This ubiquitous ness of multimedia data alters the user's behaviour and perception. These images and videos can be readily available both online with applications such as the Flickr, Facebook, twitter, you tube, Google, etc. as well as offline in personal computers, mobile phones. Flickr uploads 3,000 images per minute. According to the statistics of flicker on September 2010 five billion photos has been hosted on flicker. According to the statistics

on January 2010 more than three billion photos has been uploaded on Facebook¹ [Data Statistic]. These sites contain the immense repositories of the user generated multimedia contents. This growth increases the size of the proverbial haystack of images and videos, through which a user has to search. As a result, it becomes harder to find the desired data. All these advancements equip us with the wealth of information but come up with a disaster as well. Today we are flooding in data while starving in knowledge.

“Data is of no use unless you can access it”

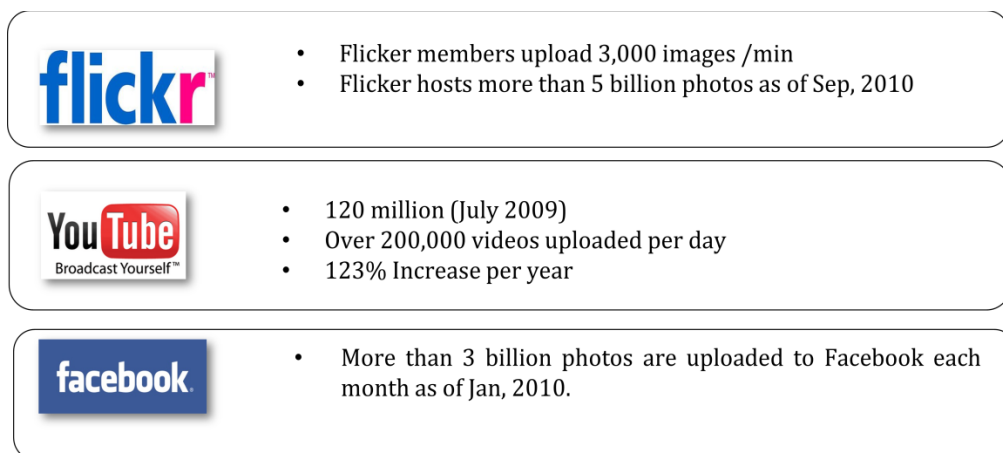


Figure 1.1: Multimedia data growth rate

Such a gigantic wealth of information is just worthless if you can't access or retrieve the appropriate information. The exponential growth of the multimedia data production from the last five years is shown in the Figure 1.2. However, the tremendous aggregate of multimedia data has exacerbated the problem of finding a desired data. How to locate required information has become the prevalent issue while facing the extensive ocean of digital data. In order to make use of such ever increasing endless digital data there needs an intuitive way of exploring the multimedia data.

¹ <http://blog.flickr.net/en/2010/09/19/5000000000/>

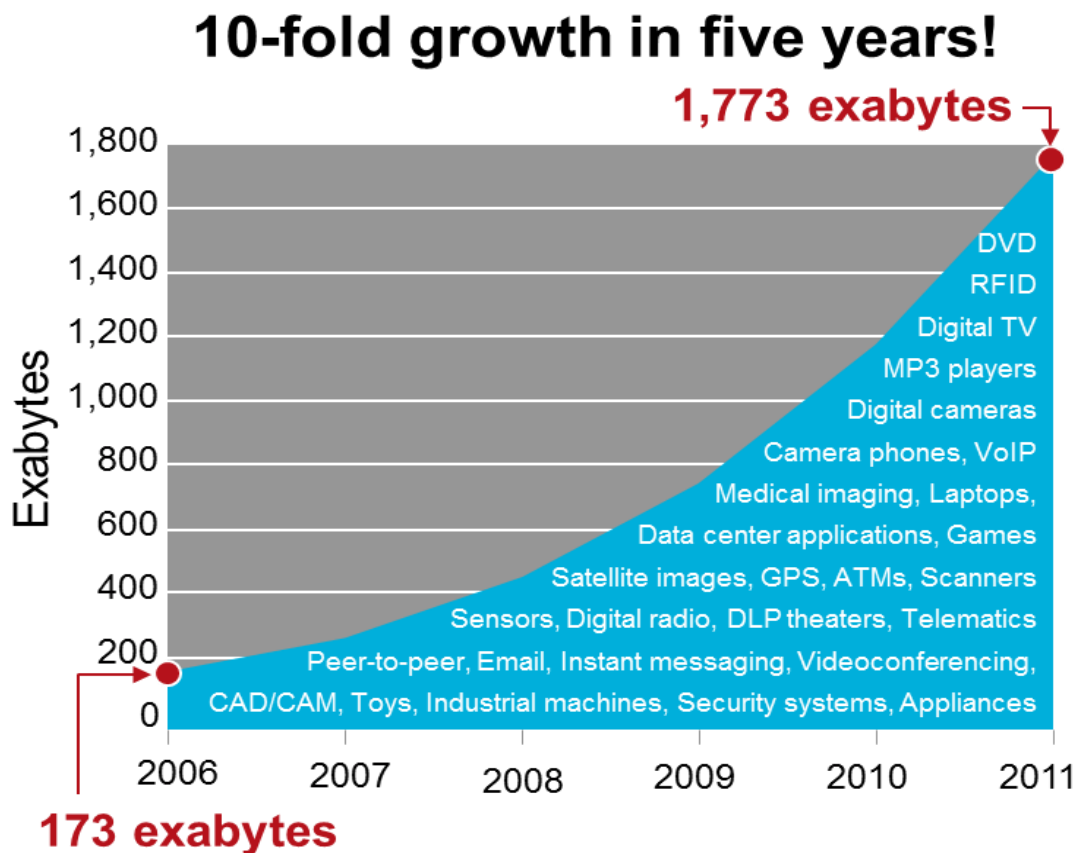


Figure 1.2: Data Production Exponential rate [John et al. 2008] Amount of Digital Information Created and Replicated each year.

Contemporary proliferation of the digital devices successively with the inexpensive storage media to cumulate immense volumes of digital data, have accompanied with the demand of an intelligent way to organize, manage and access the relevant data. Numerous information retrieval systems developed in recent years. Text retrieval systems satisfy user's demands by using the keyword matching technique. Google, Yahoo, Msn etc. are among the top retrieval systems, which have billions of hits a day. While the images, audio and video data the retrieval systems have not achieved success as that of text retrieval systems. Even though the area has already been investigated by many researchers but is not yet matured. In order to cope with the user's demand, researchers are continuously probing an intelligent way to retrieve the image and video data successfully like textual information retrieval.

The first attempts to organize images were based on textual description and started in 1970s [Tamura et al. 1984]. Images or videos are automatically saved in alphanumeric keywords, which are meaningless. Mostly the users annotate their digital data by manually assigning the names. The data is mostly archived according to the event, venue or by person name. Which is very time consuming and tedious. For finding the desired images one has to go through all the data in the collection manually. Searching the data of the interest is manageable for the small corpus or archive. However for the huge amount of data it is impractical. Finding an image that meets the user's requirement in a large and varied collection is a troublesome, prohibitive and time consuming task. These difficulties of finding the relevant information became acute with the increase of the size of the image collections. Hence, the research community is continuously exploring new and compelling ways to access stored images and videos.

Traditionally the text based retrieval systems mostly rely on the textual keywords i.e. the manual annotation that are attached with the images and videos for describing their contents. They use the manual annotation as the most intuitive way of retrieving the desired data. But unfortunately these annotations cannot describe the image semantically. Because the manually user annotated data contains the user specific data i.e. birthday pictures, wedding videos etc. and sometimes the user annotates with the words like X123, abc, xyz etc. that doesn't have any proper meaning at all. This way of extracting multimedia semantics is flawed and costly. Images and videos can depict more than one concept or different people interpret the same image differently and same is the case with the videos. Sometime the same user annotates the same image differently at different time. All these differences are due to the difference in the user perception, background or difference in the environment. So it will be difficult to find the words that will define the image and videos intuitively. Such a bleak situation has led researchers to explore the idea of the content based retrieval for image and video data.

In light of the above issues the content based image retrieval (CBIR) is introduced by the researcher in 1990's. The CBIR has been extensively studied, and numerous methods and systems have been developed for extracting the contents of the image by using these primitive features (low level feature). Increasingly, this technique has drawn more and

more attention from the research community. CBIR was first introduced for the image but later on the CBR was also extended for the videos as well. In CBIR the digital data is retrieved on the basis of the content that is extracted automatically by using the intrinsic, low level feature (colour, shape, texture etc.) extraction techniques. The image is defined in terms of the pixel representation.

Despite of the advancement in the CBIR, the users still have the problem in finding relevant information from the large heterogeneous corpus. The content in the CBR refers to the properties of the image or the video. It defines what is inside the image or video rather than what is happening inside the image and video as shown in the Figure 1.3. But the contents can't depict the entire semantics. There is a lack of synergy between the various approaches to attack the image retrieval problems. CBIR techniques suffer from the semantic gap, which is the discrepancy between the information that can be derived from the low-level image data and the interpretation that users have of an image.

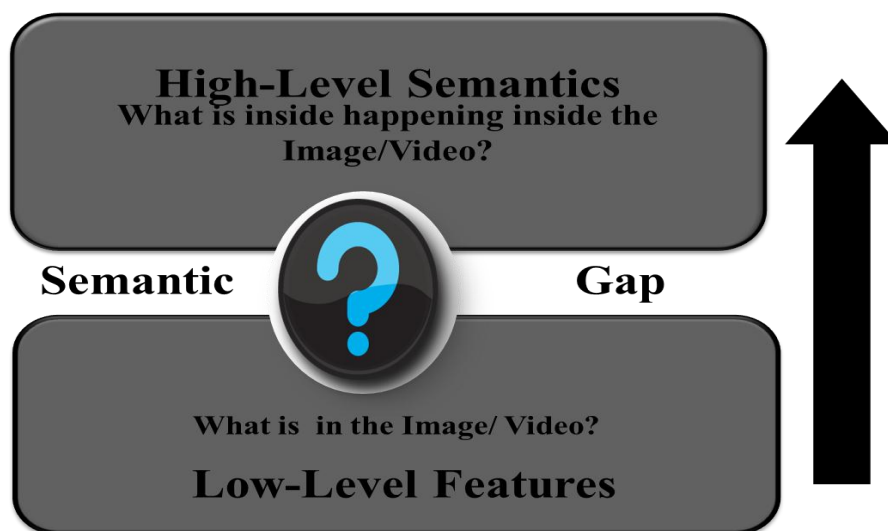


Figure 1.3: Semantic Gap

1.2 Motivation and Application

The advancement in the digitalization have resulted in the availability of ever increasing volume of digital data also will bring a need for the better way to organize and search. Despite of all the advancement in digital technology, the task of managing such a

large and ever increasing repositories is non-trivial task. The search engines like Yahoo, Google, YouTube, Facebook, Msn, Flickr etc. are quite capable of retrieving the document but the result is not optimal. Until now, the advancement in the digital data analysis have mostly stressed on devising an automated ways of extracting information from them using the low-level features. These efforts have shown an improvement over the traditional ones in terms of automation by using the CBR techniques. Ideally CBR techniques will be able to automatically retrieve the image or video on the basis of primitive features. The CBR systems mostly rely on the visual feature of an image or video. But the visual similarity is not the semantic similarity. The semantics refers to intended meaning of media such as images and video. Sometimes the thing that looks similar doesn't contain similar concepts. CBIR captured the interest of both industrial as well as the academic research communities. Unfortunately, the CBR is still far from optimal.

A proverb ascribed to the Chinese philosopher Confucius “a picture is worth a thousand words²” highlights one of the main reasons for the failure of the CBR systems. The proverb suggests that image can be replaced by the words but these words assigned to an image can also differ from person to person. The single image can be explained by many words as shown in the Figure 1.4. There are no specific words to define the image and words for describing an image are subjective to the person perception and background. Due to the rich contents of an image different people perceive the same image differently same is the case with the video data as well. This perception subjectivity leads to the poor retrieval performance. Computers are still yet not being able to cover all the human perception and to cope with the flexible nature of a human. As humans perceive the things differently from a computer. Unfortunately, a CBIR system is unable to understand the data as how human can interpret it. It is difficult for the CBIR system to interpret coloured pixels in the image into higher level representations which are used by humans to understand the image. Therefore, such CBIR retrieval results cannot be as accurate compared to judgment made by human beings. There is a gap between the human perception and the computer interpretation for an image this gap is known as the semantic gap. According to Smeulders [Smeulders et al. 2000] Semantic gap is

² A Confucius Chinese proverb "The Home Book of Proverbs, Maxims, and Familiar Phrases".

“The lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation.”



Figure 1.4: Image is Worth than a thousand of Words. The image can be perceived as an image of lion, sign of Bravery, sign of Fear, image of a zoo, lion in the jungle, lion in the forest, Franklin Park Zoo etc.

Semantic gap is a gap between high-level semantic concept by which human understands an image and low-level features used by a computer to classify images. The current CBR system are mature enough to extract what is in the image or the video but flunks to extract the semantic in the images and videos i.e. what is actually happening inside the video. How human beings perceive and identify images, are typically based on high-level abstractions such as activities, objects identifications, events, or emotions. The effective and efficient image and video retrieval system cannot be attained by just considering only simple independent visual features. The semantic gap exists because the features or information that we can currently extract from the pixel content of digital images are not rich or expressive enough to capture higher-level information which on the other hand is readily performed by human beings. Extraction of visual features from an image has long history, but still it is very difficult to use such features to represent high-level semantics. It is a challenging problem to bridge the semantic gap. Research in this problem area is gaining attention to make

computers capable of extracting high-level semantic concepts from images and videos as human do.

The key motivation for our research is the lack of effective solutions to the current problem of retrieving semantically similar images and videos from a colossal, overwhelming and ever increasing datasets. Much of the efforts have already been made in retrieving the data from these huge corpuses but yet no breakthrough results have been made. We have tried to bridge the semantic gap by analysing the user request semantically and ranked the output based on the semantic relevancy order.

The intended audience for this specific research comprises of Content Owners and production companies like BBC, CNN, Geo, TV Service Providers, Satellite & Cable companies, Electronics Manufacturers like Mobile, DVRs, Digital media players, Content-Service Providers, Content monitoring companies which provide push and pull services, Web-Content aggregators, Companies that aggregate digital media like Google, Yahoo, YouTube, Content-repackaging companies, Companies that acquire content like sports videos and TV programs and repackage it according to user needs etc. This research helps in managing multimedia data effectively and efficiently, helps in searching and retrieving the particular information from the large dump of information and attempts to make media search and retrieval easy.

1.3 Main Aim and Objectives

The production of digital data is increasing expeditiously due to the advancement digital technology and proliferation of cheap storage media. It is undoubtedly seen that the number of images and other media are increasing dramatically. Such a sheer amount of digital data disqualifies the manual browsing for finding the relevant data. Millions of images and videos are uploaded by the users on Google, flicker, Facebook and You Tube etc. Images and videos are now become a vital part of our life and a widely used medium for education, communication, education as well as for entertainment. As video is the best way of communicating or expressing an idea because it uses all the modalities simultaneously. Despite of large number of digital data repositories available both online as well as offline, accurate image and video retrieval system are still rare. In order to overcome the bottleneck

and to cater the needs arising from such increasing image and video corpus intelligent systems are required to search, filter and retrieve the relevant data. There is a strong urge to efficiently and automatically extract high level information from images and videos. The production and use of digital data in such a ubiquitous manner has brought yet another challenge i.e.

- How to organise and manage large collections of digital data?
- How to facilitate efficient and easy retrieval of and access to a desired data in a collection of several millions of available images and videos?
- How to bridge the gap between the user perception and computational analysis of an image or video?

The intention of this research is to explore an intelligent ways for retrieving the relevant data from the highly dynamic and huge digital data repository on the basis of Semantic features not on primitive features. Although reasonable attempts have been made, but still the problem is not yet completely resolved. We focus on an efficient and effective Semantic multimedia modelling and interpretation framework for search and retrieval particularly in terms of retrieval accuracy.

More specifically, the objectives of the research are as follows:

- To explore a way to reduce the semantic gap.
- To investigate and develop techniques to extract semantics or the intended meaning behind the group of words in the form of user query.
- To investigate a way of ranking the output on the basis of the semantics rather than on the basis of the frequency.
- Explore a way to display the data according to the semantic relevancy.
- To implement the proposed approach to improve semantic multimedia retrieval.

This work is concerned with Semantic based retrieval not on the basis of contents to improve the retrieval performance. The overall aim of this thesis is to explore new ways of retrieving the digital data based on the semantic features with a main focus on bridging the semantic gap.

1.4 Problem and Challenges

Taking into account the development in the digital technology and tremendous surge of digital data production has brought challenges and new opportunities for the research community. The production of digital data is increasing at an incredible velocity. This colossal of digital data being generated resulted in the difficulty of finding the relevant data. . Due to the unavailability of the efficient systems to index, analyse the digital data semantically, significant amount of useful data becomes useless. Most of the relevant data cannot be retrieved due to poor annotation. Only few data will be mostly retrieved and some of the relevant data will be rarely retrieved. This problem is known as long tail problem as shown in the Figure 1.5. There is a need for the system for organizing and retrieving the relevant information to meets the user's need.

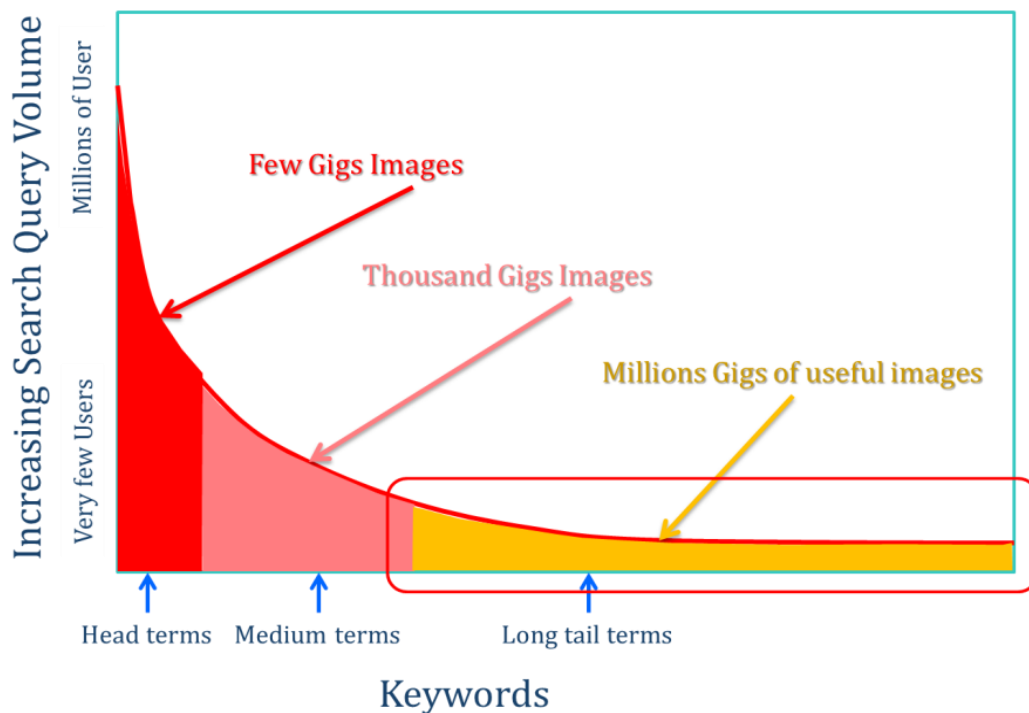


Figure 1.5: Long tail Problem

The main challenge is that for a computer the image contents are just the combination of pixels that are characterized by the low-level colour, shape, texture etc. And the video is ultimately a group of frames with a temporal feature and the frames are basically an image. For them it refers to not the contents that are appearing but rather on the semantic concepts that what is happening inside the video or image. The same video contents or images have more than one meaning or we may say as for the same image and video different people extract different meanings. A content-based description of an image and video can differ considerably from a human user's description, just as two people can provide different descriptions of the same scene. Due to the flexible nature of the human and the hard-coded computer nature there appears a problem known as the semantic gap.

The foremost challenge faced by the research community while retrieving the relevant information is the "Semantic Gap". The Semantic Gap is defined as "the disparity between a user's high-level interpretation of an image and the information that can be extracted from an image's low-level physical properties". The CBR systems rely on the machine-interpretable description of an image i.e. the content. Sometimes the contents cannot describe the overall

semantics of the image and the video. For instance if in an image there is grass, people, trees, sky, ball, benches etc. These are the contents and these contents by itself can only depicts that what is inside the image but flunks to define what is actually happening inside the image.

Although the state-of-the-art retrieval techniques or the CBR systems are day by the day getting more and more powerful and can now achieve the accuracy up to some extent. The performance of the CBR system is inherently constrained by the use of the low-level features, and cannot give satisfactory retrieval results. Indeed, the power of these tools doesn't reduce the semantic gap. Nevertheless, we are still far away from a complete solution of reducing the semantic gap because of the following challenges.

- i. How can we interpret the user requirements that are given to the system in the form of query?
- ii. How to reduce the semantic gap?
- iii. How the background knowledge can be extracted from the digital data?
- iv. How to bring the efficiency to the system along with the semantic accuracy?

1.5 Research Direction

Many advances have been made in various aspects of image and video retrieval, including primitive feature extraction, multi-dimension indexing, object detection and recognition. However, there are still many research issues that need to be solved. These include

1.5.1 High Level Semantic Concepts and Low-level Visual features

Human beings sense their surroundings extensively by audio-visual mode. The human brain and visual system together with the human auricular skills provide outstanding capabilities to process audio-visual information, and instantly interpret its meaning based on experience and prior knowledge. Audio-visual sensation is the most convenient and most effective form for humans to consume information, we believe what we can see and hear, and we prefer to share our experiences by aural and visual description. In particular for complex

circumstances, visualization is known to convey the facts of the matter best. The growth of visual information available in images and videos has spurred the development of efficient techniques to represent, organize and store the video data in a coherent way.

Current computer vision and the data mining techniques had made it possible to automatically extract the low level features from the images and the videos. These techniques interpret the videos as a group of frames within which each key frame is treated like an image. Each image is interpreted in terms of group of pixels that represent the entire image or video. These primitive features are then used to find the high level semantic concepts. But unfortunately it is not possible to extract the semantics from these low level features (discussed in chapter 2 figure 2.12 and 2.13). There is a need for a system to establish a link between low and high level semantic features. In order to bridge the semantic gap several attempts have been made like Concept detectors [Naphade et al. 2004], Semantic annotation [Carneiro et al. 2007], ontologies [Wei et al. 2010]etc. But yet no universally accepted solution has been made. For an efficient utilization of the images and video data there is a need for the system that can perform the extraction of the high level semantic concepts accurately.

1.5.2 Human Perception of Image and Video Content

The ultimate end user of an image and video retrieval system is human. The user is the most significant part of the Information retrieval mechanism. The study of the human perception for the images and video is very crucial. The ultimate aim of the information retrieval process is the user satisfaction. Unfortunately this aim is hard to achieve. It is due to the fact that human perception difficult to interpret. The same thing will be interpreted differently by different users. While sometimes a same user can interpret the same thing differently. The subjective human perception makes it difficult to interpret. The hard coded computer had a difficulty in coping with the flexible human nature.

This topic is gaining increasing attention in recent years, aiming at exploring how human perceive image and video content and how can we integrate such a “human model” into the image and video retrieval systems. This is because, after realizing the difficulty in

interpreting human perception subjectivity of image and video content, they naturally resorted to different semantic query analysis techniques to “decode” human perception.

1.5.3 Image and Video Retrieval

The Proliferation of digital camera and the low cost storage devices have drastically increased the multimedia content. Substantially searching the multimedia data is more strenuous than the text document. This is partly because multimedia content can be difficult to articulate, and partly because articulation can be subjective. For example, it is difficult to describe a desired images and videos entirely by using low-level features such as color, texture. Moreover, different users may perceive the same image differently; and even if an image is perceived as similar, users may use different concepts to depict it. To better organize and retrieve the almost unlimited information, semantic based retrieval systems are highly desired. Such solution exists for text-based information. For images and videos, even though some good work has been done, technical breakthroughs are needed to make image and video search comparable to their text-based counterparts. Semantic analysis of image and videos is an important step to this goal. One major technical barrier lies in linking the low-level visual features used in most systems today to the more desire semantic-level meanings is using the primitive features for interpreting the semantics.

1.5.4 Diverse Nature of the Bench Mark Datasets

The availability of datasets and their annotation standard is another core challenge. The datasets like Coral, TRECVID, Image CLEF, PASCAL and LabelMe are developed by keeping different aspects of the annotations in mind. This increase the complexity in firming a flexible system for all types of datasets.

1.6 Proposed Research Contributions

Keeping these in mind, we propose a Semantic Multimedia modelling and interpretation framework that can offers a semantic accuracy in terms of search and retrieval. The main aim of this thesis is to propose a novel framework for searching and retrieving the multimedia data semantically. It is in this scope that we try to solve one of the most

challenging issue of the semantic multimedia retrieval i.e. the Semantic gap and address three main elements: Semantic Query Interpreter for images, Semantic Ranking and Semantic query interpreter for videos. Figure 1.6 depicts the contributions of the proposed work along with the addressed problems. Despite of the previous work that emphasises on what is in the image or video. With this approach we try to investigate a way to explore what is actually happening in an image or video. We have substantially reduced the semantic gap by achieving a noticeable precision. An overview of the proposed contributions is discussed below.

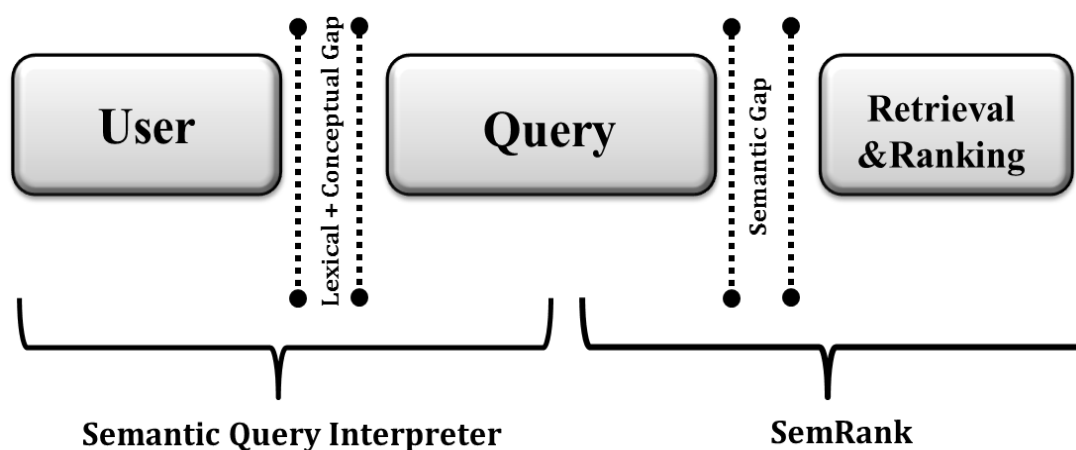


Figure 1.6: Proposed Research Contributions

1.6.1 Semantic Query Interpreter for Image Search and Retrieval

Query plays a vital role in the Information Retrieval process. The proposed semantic query interpreter will expand the user queries for all possible dimensions. The semantic query interpreter expands the user query by using the knowledgebases. The expansion can be done on two main dimensions i.e. lexical and conceptual. The lexical expansion can be done by using largest open source lexical knowledgebase i.e. WordNet. While the conceptual expansion can be done by using the open source conceptual reasoning knowledgebase i.e. ConceptNet. The semantic query interpreter framework along with the integration of these knowledgebases attempts to reduce the semantic as well as vocabulary gap. The efficiency of

the proposed system can be validated by using the well-known retrieval model known as the vector space model.

1.6.2 SemRank: Ranking Refinement by using Semantic Intensity

Extracting the relevant information from the corpus and then rank the information according to the relevancy order is one of the main functions in the IR systems. Users are highly keen in the top ranked retrieved results. We have proposed a semantic ranking technique known as SemRank, to sort the retrieved results on the basis of the semantic relevancy order. The SemRank rank the output by using semantic intensity of an image. The Semantic Intensity is the concept dominancy with in an image. An image is an integration of different semantic concepts. Some of the concepts in the image are more dominant than the other. The SemRank ranks the retrieved documents in the decreasing order of their concept dominancy factor. The effectiveness of the proposed model is compared against the most widely used traditional retrieval model known as vector space model.

1.6.3 Semantic Query Interpreter for Video Search and Retrieval

The proposed semantic query interpreter for the images had been extended for the videos as well. It will expand the user query lexically and conceptually for retrieving the relevant videos. For evaluating the effectiveness of the proposed framework we have used the LabelMe Video dataset.

A detailed discussion on all these contribution has found in the fort coming chapter 3, 4 and 5.

1.7 Organization of the Thesis

The thesis is organized in the following manner.

In Chapter 2, an extensive discussion on the current achievements concerning the components of image and video search and retrieval is provided. The main aim of this chapter is to survey the state of the art in the respective field. This includes general information retrieval overview along with the overview of the multimedia data and retrieval discussion. The next discussion focuses on the structure of text, Image and video and their retrieval

techniques. The next discussion related to various retrieval techniques like the text based, content based and semantic based retrieval and compared all the approaches for image and video retrieval. The chapter also discusses several approaches that aim to bridge the gaps between high-level and low-level feature. This chapter intent to explore the current work that has been done so far for reducing the semantic gap in image and video retrieval. It includes various image and video analysis techniques, different types of queries, existing systems etc. This discussion provides a motivation for semantic based retrieval as one of the most promising approaches.

In Chapter 3, a proposed algorithm semantic query Interpreter for Images is discussed. The chapter also explored the recent work in the area of the query analysis and expansion along with their pros and cons. The efficiency of the proposed system is tested in terms precision, recall and F-measure. The experiments are performed on open source image LabelMe dataset to prove the semantic accuracy of the proposed system.

In Chapter 4, a semantic ranking approach for the image retrieval is introduced to rank the retrieved results on the basis of semantic relevancy. A brief discussion on the current work on ranking the retrieved images has also been made. The next discussion focuses on the novel concept Semantic Intensity for sorting the output. A performance analysis, which is carried out with a same dataset that contains object annotated images, will be reported to demonstrate that the proposed scheme is effective and reliable for the semantic ranking of output.

In Chapter 5, proposed Semantic query interpreter for video search is discussed in details. This chapter researches the entire structure of the video along with the various video analysis techniques. The performance analysis has been made on the video datasets like LabelMe dataset. Results are reported to verify the effectiveness of the proposed model.

Finally, in chapter 6 we conclude with a summary of achievements and the future work are discussed. Chapter 6 is followed by appendices and references.

The appendices contain the implementation of the proposed contributions.

It is to be noted that all the main chapters are presented with a self-contained set of introduction, main concepts, experimental results, and conclusion.

Chapter 02

Basic Concepts & Literature Review

“I find that a great part of the information I have was acquired by looking up something and finding something else on the way”.

Franklin P. Adams, 1881–1960

Today the technological advancement has achieved a point that was thought of nothing more than science fantasy a few decades ago. With the emergence of multimedia enable devices users start accumulating large repositories of digital data. Due to the low cost digital storage these archives are progressively being digitalized. These progressions persist at an incredible velocity. These advancements in the information technology have culminated in a collection of immense amounts of information that are consumed by user on a routinely basis, including military, government agencies, new agencies like BBC, Security agencies like CCTV, education and several other purposes as well. In addition, individuals also consume this information. Nowadays majority of the people have digital cameras even in the hand held devices like mobile phones and they use these cameras to generate their own personal image or video collections. Moreover, current low-cost data storage technology permits for huge sizes of digital data collections. All these advances yet bought up with a new demand of effective multimedia data retrieval systems.

The paramount concern is the organization of these gigantic volumes of digital data collections for potent use and accessibility. Utilization of these worthy digital data collections requires an effective image searching, browsing and retrieval. Owing to this fact, in the 1970s a new research area, image search and retrieval, was innate [Cristina et al. 2001]. Image retrieval is concerned as a system that can allow the user to find the relevant information from such a colossal data corpus.

Multimedia data comprises of text, images, audio and video data. Need of the efficient system for discovering the required data in this overload of data. Today's retrieval systems have coped the users need for the textual data but for the multimedia data like images and videos it's still at the infancy stage. The reason is that *"a word is easily identifiable in a text and is usually associated with a concept at the level of human communication. An automatically identifiable feature of an image, such as a colour distribution, does not provide a retrieval system with a concept that is equally useful for user interrogation"* [Cristina et al. 2001] *and is therefore, not practical for indexing as is required by search engines.* The existing Content Bases retrieval Systems (CBRS) relies on the wealth of integrated primitive features of the image or video, i.e. either colour, shape, texture, motion or the combination of them. These systems are seldom adequate for attaining accurate results.

The major stumbling block in the CBR systems is that the low-level features like colour, shape, texture can be automatically interpreted by the machine while the high level features or semantic features cannot be extracted by a machine in a reliable way. These high level features are mostly extracted and annotated by humans. In spite of the fact that sufficient research works have been done so far in the field, but still no universally accepted model has yet been developed.

The overall aim of the chapter is to systematically explore the key elements and the state of the art done for retrieving the images and videos according to the users. The research will be built on the background of semantic based image and video retrieval, keeping in view the challenge of the semantic gap and the limitations of the current Content Based Retrieval Systems. In this Chapter, we will focus on explaining the background of our research by explaining the Information Retrieval (2.1), general multimedia retrieval (2.2), Image retrieval (2.3), video retrieval (2.4), approaches for image and video retrieval and the query paradigms (2.5), the Evaluation (2.6) and finally we conclude the chapter in the chapter summary (2.7).

2.1 Information Retrieval

Information Retrieval (IR) is an art of retrieving the required information from the collection of data or a term used for the task of making information accessible and available to the user. Information can loosely be defined as some sort of organized data, which is considered useful or meaningful to someone. Information Retrieval deals with the representation, storage, organization of, and access to information items [Baeza-Yates et al. 1999]. Information can be anything either textual data, audio, visual data (images) or a video clips. Traditional IR systems deal mostly with text. Information retrieval tasks include organize, store, browse, index, search and retrieve the required data. The general IR diagram is represented in Figure 2.1.

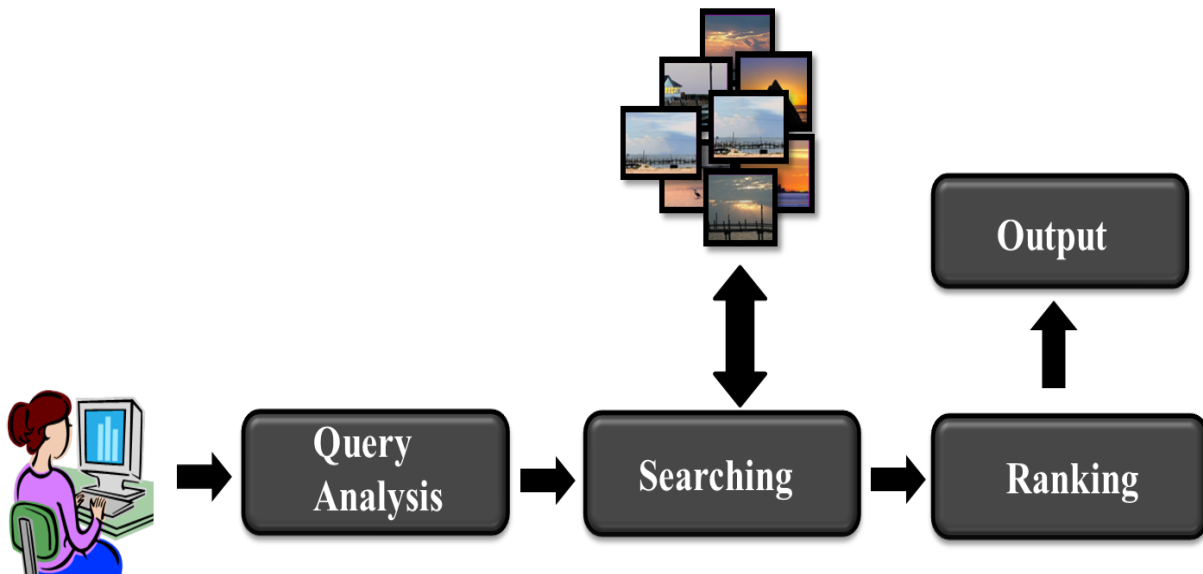


Figure 2.1: Typical Information Retrieval process

2.2 General Multimedia Retrieval

The technological, cultural and economic drastic variations that appeared about with the emergence of the information age are rich in prospective benefits, but yet exhibit a number of challenges. The information is copious. It is in many cases accessible and available, but its sheer volume and diversity make it difficult to pinpoint the actual useful information. Knowledge stored in books or transmitted by oral tradition, has long ago outgrown the human synthetic capability Paper based catalogues. Indexes and digests can still be beneficial but with today's digital technology we are getting closer to building a deal information "system".

The key ability of such a system is multimedia retrieval. In simple terms, a user explains to the system what they needed information is and the system's role is to find it, structure it and present it to the user. The outcomes of such an interaction consist of a set of digital documents that contain the desired information. The medium that would convey this information could be textual, still images, video, audio or a free mixture thereof. This polymorphic nature requires flexible need formulation and effective information-storage, access and transmission strategies.

These individual sub-problems have already been studied. The domain of multimedia retrieval is at meeting point of different research areas databases, signal processing, communications, artificial intelligence, human-computer interaction and many more.

Multimedia information retrieval has been a very active research topic in recent years.

Progression in the multimedia enabled devices together with the low cost storage devices has raised the production of multimedia data like images, audio and video. These multimedia objects now become a part of our everyday lives. Presently producing and storing the multimedia data is at our finger tips. This omnipresence of the digital data has led to the demand for the information retrieval systems that can organize, store, manage and access the relevant data. Traditional multimedia information retrieval systems are designed for managing the textual data Retrieval of the textual data relies on the comparison of the text in the query with that within the document. These methods are sufficient for the textual data but don't provide precise results for the multimedia data. As the semantics within the images, video and audio can't be as easily interpreted as text [H. Tamura at al. 1984, Chang et al. 1992]. Text based information has been comprehensively researched and implemented successfully. While for the images and videos it is still worth investigating. The progression of the multimedia search and retrieval is shown in the Figure 2.2.

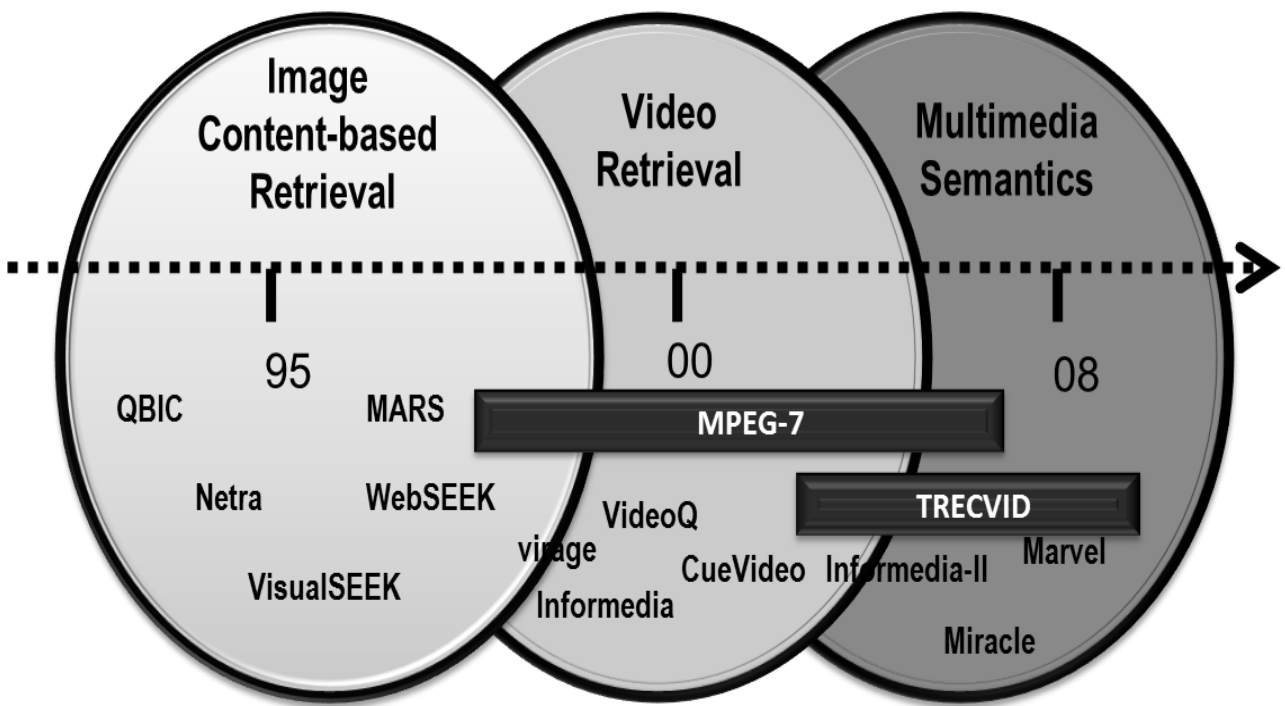


Figure 2.2: Progression in the Multimedia Retrieval

2.3 Image Retrieval

Image retrieval sprouted from the fertile soil of database research intermixed with the hint of information theory and document processing. Image retrieval can be defined as a way of retrieving the relevant images according to the user query from the set of images in the large image corpus. The image corpus may consist of the photos of the party, home photos, holiday photos, photos uploaded by the user on the social media system like Flickr, Facebook etc. Overwhelming amount of the digital data production is available online, e.g. Google, Yahoo, Facebook, You Tube, etc. as well as offline. A successively growing aggregate of information is available in digital formats. The information is not restrained to textual documents, but can additionally be delineated as images, audio and videos. Owing to the number of entities held in such aggregations, users must be equipped with various methods to retrieve the information they are searching for.

"A picture is worth a thousand words"¹, even a simple image can convey more than one semantic idea. This can be abstracted with in the contents of the image and can't be easily extracted like a textual data. Some of the images are semantically rich and present multiple concepts. Extracting and interpreting the semantics from semantically rich and ambiguous images like shown in the Figure 2.3 is very difficult. This reveals that searching an image is difficult than searching for text. The first research was aimed at text retrieval and passed through numerous phases, starting from abstract retrieval (medical, legal, etc.) to heterogeneous retrieval .With the proliferation of media such as images, videos and audio, and the steadily increasing availability of computing power and storage space, the angle of research started widening to accommodate these new types of content. New ingredients were poured into the brew of information retrieval, starting from signal processing, digital image processing, cognitive science through artificial intelligence and bio-inspired techniques, up to more empirical and subjective issues of the HCI and computer vision. All these disciplines find in information retrieval, a challenging area for endeavour.

¹ A Confucius Chinese proverb "The Home Book of Proverbs, Maxims, and Familiar Phrases".

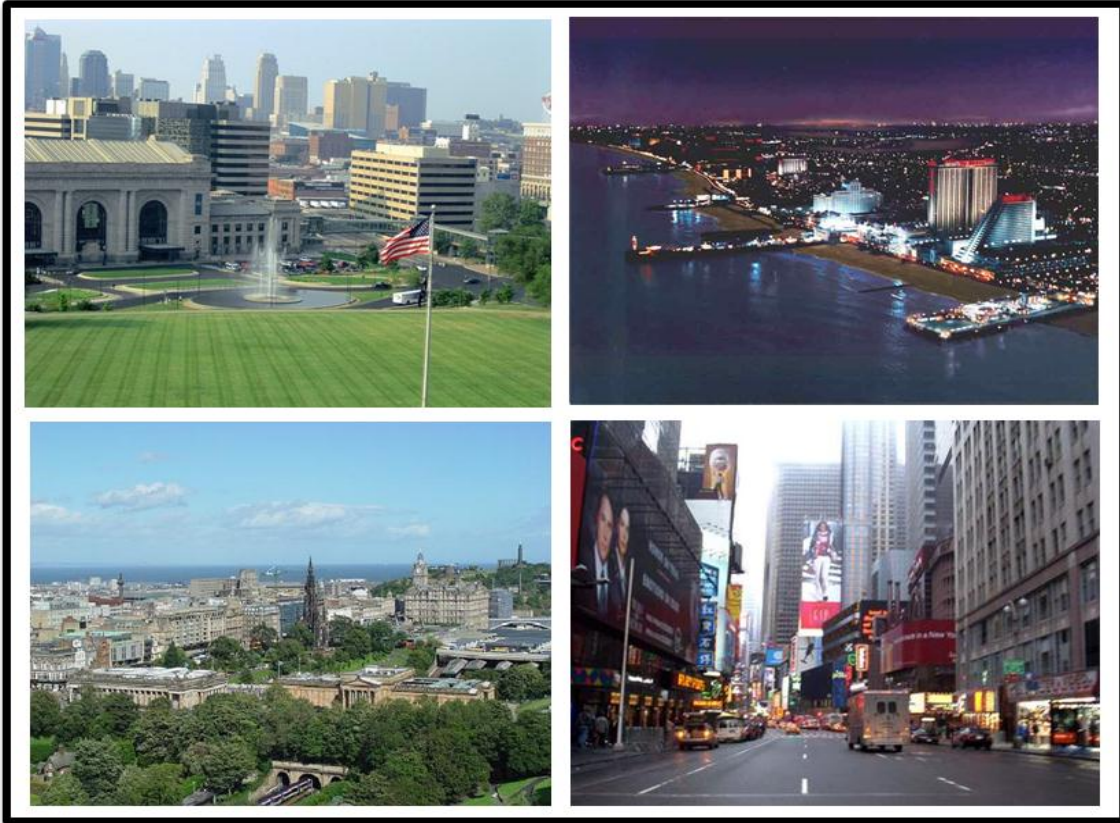


Figure 2.3: Semantically rich images and ambiguous images

Image retrieval is concerned with searching and retrieving the relevant digital images from an image dataset. This field has been explored since the 1970s [Cristina et al. 2001]. Preliminary image retrieval techniques were mostly based on the textual descriptions and captions of images. These systems were meta-data (textual annotation, user specified text) for the retrieval [Datta et al.2008]. Later these become the standards for the retrieval systems e.g. Google Image search , Flickr, Yahoo Image Search , Imagery Search Interface, Pic Search, Alta Vista, Pixsy, FeelImage image search, Photo Bucket etc. Today, image retrieval is one of the demanding applications that develop along with the advancement of digital imaging technologies. The exponential number of digital images produced by consumers need to be searched and allocated for future use. Consequently, image retrieval has been diversely researched by the research community.

2.4 Video Retrieval

In 1960 the first computer-based multimedia data was developed which combine the text and the images [Randall et al. 2001]. . After then all the other modalities (textual, audio,

visual) are continue on adding to the system e.g. animations, audio and eventually when all these modalities are added then it constitute to something that we called as video. Owing with the invention of the digital technology and the multimedia enable devices, the world seems to be smaller enough to accommodate such exponentially increasing production of digital data. The cost of the storage decreases and the amount of multimedia data increases. Finding the data of the interest become harder and harder.

Ubiquitous presence of the multimedia data including broadcast news, documentary videos, meeting, movies etc. confronts the users with the problems of organizing, storing and finding the relevant information. The most prevalent issue is the information overload. This overload surpasses our capabilities of accessing and processing the information. Furthermore, video data interpretation and representation is very different from those encountered in the textual. Most of the users are now accustomed of simple and an intuitive way of finding the relevant information. Among all the multimedia data the video is the most efficient way to transmit the information. As the video uses all the modality such as audio, visual and textual simultaneously so it is most complex to deal with. The techniques of the textual and visual information retrieval systems are not applicable here for the accurate and efficient retrieval. The main stumbling block is that the video can simultaneously depicts more than one semantics. Due to this explosive growth there is a strong urge for the technologies that can access and retrieved the desired video from the large video corpus with the accuracy.

Apparently the manual analysis of the video data seems to be sufficient enough for the small collection. Unfortunately it will be not practical to use it for the colossal and ever increasing video corpus. In order to address these issues, research communities have been exploring the ways to retrieve videos efficiently and effectively from the large video corpus. The large amount of video data is available both offline and online including Youtube, Google Video, Blinkx, Pixsy, fooooo, VideoSurf, Truveo, MSN videos, Yahoo videos etc.

2.5 Approaches for Image and Video Retrieval

There are various image and video retrieval techniques. These techniques can be categorized according to whether they are based on text, content, or semantic concepts. We distinguish these approaches by the type of characteristics that are employed to denote the images and videos and the approaches that are used to retrieve relevant data.

The text-based image and video retrieval techniques use keywords, the CBIR techniques use low-level image and video features and the semantic-based techniques use concepts.

2.5.1 Text based Image and Video Retrieval

The most common way to manage large image and video collections is to store text in the form of keywords together with the image or video. Rather recently, conventional image and video retrieval employed text as the paramount mode by which to represent and retrieve images and video from corpus [Bimbo. 1999]. Since then 1970's, the image can be exhibited by the textual description and text based retrieval techniques have been used for retrieving the relevant data [Thomas et al.1997].

We categorize the text based retrieval technique into two categories i.e.

- Retrieval technique that is based on the text surrounds the image or video.
- Retrieval technique that based on the annotation (meta-data) attach with the each image or video in the corpus.

The approach that deals with adjacent text searches the keywords that are physically similar to the image. Search engines that exploit this approach are Google, Yahoo, and AltaVista. This approach is based on the assumption that the surrounding text describes the image and videos. The technique relies on text surrounding the image and videos such as filenames and the paragraphs close to the image and videos with potential relevant text. The main impediment with these techniques is that the search engine deliberates an image or video relevant because it is annotated with a specific keyword.

Sometimes a relevant image might be left out owed to the absence of specific keywords. While often there might be no relevant text surrounding the images or videos, but they are relevant. In fact, there might exist images or videos where the surrounding text has nothing to do with them. In these cases, these returned results might be irrelevant and have nothing in common with the required images and videos.

The other approach uses the annotation of the images and videos and is often a manual task. The text-based technique first annotates with text, and then uses text-based retrieval techniques to perform image and video retrieval. Annotation of images and videos lets the user

to annotate the image with the text (metadata) that is considered relevant. The text can be time, event, location, participants or whatever the user finds relevant.

Limitations of Text based Image Retrieval

Nevertheless, there exist two major difficulties, especially when the volume of images and videos collection is large with hundreds of thousands samples. One is the huge amount of human labour required in manual image and video annotation and is very time-consuming. Textual based retrieval cannot append the perceptual significant visual features like colour, shape, texture [Bimbo. 1999]. The other difficulty comes from the rich content in the images and videos and the subjectivity of human perception which is more essential. The annotation of the image and videos completely depends on the annotation interpretation [Enser et al 1993] i.e. different people may perceive the same image differently as shown in the Figure 2.4. The Figure depicts the multiple interpretation of the single image. The perception subjectivity and annotation impreciseness may cause unrecoverable mismatches in later retrieval processes. And to retrieve the required data the user constructs a query consisting of the keywords that describes the desired image and video. Although the text based retrieval system has gained benefits of traditionally successful information retrieval algorithms and techniques.



Figure 2.4: Multiple interpretation of same images Park like Tree, Sky, Horse, People, Ridding, Sunny Day, Outdoor

Critics of text-based approach dispute that for accurate image annotation it must be automated. The automatic annotation is limited due to its deficiency of extracting semantic information from the images and videos. Only automatic annotation of images and videos in integration with pure text-based image and video retrieval will be inadequate. The available metadata is mostly restricted to the technical information surrounding the image or video, such as time, resolution of the image or video and their name.

The users may find it difficult to use text to perform a query for some portion of the content of an image and video. Text-based retrieval techniques are absolutely limited to search the metadata that is tagged to the image and video. If the text queried is not annotated with the same tag as attached with the image and video, the data will not be returned. This means that if a particular piece of the image or video is interesting this must be explicit included in the metadata. If the desired object is not a main part of the image or video, sometimes it may happen that is not described in the metadata and hence cannot be a retrieve as a result from a query describing such portions of the image or video.

One of the disadvantages of text-based image retrieval is that a word can have different meanings. This problem is best illustrated with an example, searching for the images or videos of jaguar or Apple. The system can't differentiate either the user is looking for the jaguar car or jaguar animal as shown in the Figure 2.5(a, b). The two concepts have the same name but contain an entirely different semantic idea. The retrieval systems don't have reliable ways to separate the concepts. These problems are present even in systems with automatic synonym lists or thesaurus capabilities [Schank et al. 2004]. There exist several text-based image and video retrieval services today, Google is a large player. Google is the largest player but still faces the same problem.

Attempts have been made to make the tags attached to the images and videos more flexible by attaching vast number of descriptive words. The thesaurus based annotation or knowledge based annotation has gained much of the researchers attention [Tring et al. 2000].

Consideration to the demands, researchers concluded that visual features play a crucial role in the effective retrieval of digital data. This initiates to the development of the content based image and video retrieval [Venters et al. 2000].



Figure 2.5 a: Same name different Semantics i.e. jaguar car and the jaguar animal.

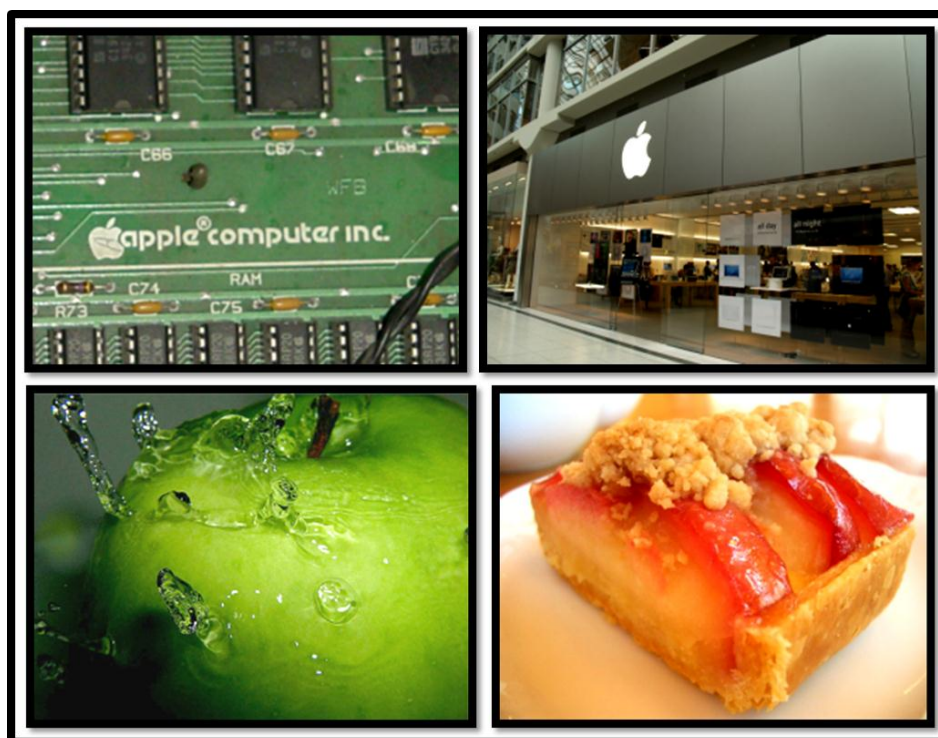


Figure 2.5b: Same name different Semantics i.e. Apple a company and the name of fruit

2.5.2 Content based Image and Video Retrieval

The need to manage these images and videos, to locate target images in response to user queries has become a significant problem. One way to solve this problem would be describing the image and videos by keywords. The keyword based approach has a bottleneck of manually annotating and classifying the images and videos, which is impractical for the overwhelm corpuses. The human perception subjectivity problem may affect the performance of the retrieval system.

Current commercial image and video search engines retrieve the data mainly based on their keyword annotations or by other data attach with it, such as the file-name and surrounding text. This relinquishes the actual image and video more or less ignored and has been following limitations. First, the manual annotation of images requires significant effort and thus may not be practical for large image collections. Second, as the complexity of the images increases, capturing image and video content by text alone becomes increasingly more difficult.

In seeking to overcome these limitations, content-based retrieval (CBR) was proposed in the early 1990's [Baeza-Yates et al. 1999]. "Content-based" means that the technology makes direct use of content of the image and video rather than relying on human annotation of metadata with keywords. Content-based retrieval (CBR) research endeavours to devise a retrieval system that exploits digital content in the retrieval process in a manner that is eventually independent of manual work. CBR is an umbrella term for content-based multimedia retrieval (CBMR), content based visual information retrieval (CBVIR), content-based image retrieval (CBIR), content-based video retrieval (CBVR) and content-based audio retrieval (CBAR). CBR may also be termed as multimedia information retrieval (MIR).

Content based retrieval extract the feature of the image or video themselves and use it for retrieval rather than the user generated meta data. CBR uses the primitive features of the image and video like the colour, shape, texture, motion etc. [Sharmin et al. 2002]. Content based system index the images and videos automatically by using different techniques for their visual contents.

For the computer, a video is merely a group of frames with a temporal feature, where each frame is basically an image. The computer take each image as a combination of pixels

characterize by the low-level colour, shape and texture. CBR represents these features in the form of vectors called the descriptors of the image or video. CBR extract these primitive features by using automated techniques and then further use it for searching and retrieval. Thus, these low-level visual features extraction from images and videos has initiated to the many research in the CBR [Veltkamp et al 2000].

A typical CBIR system should be able to interpret the content of the images in a query and a collection, compare the similarity between them, and rank the images in the collection according to their degree of relevance to the user's query [Tamura et al. 1984]. The Figure 2.6 shows the typical content based retrieval system, where the user query was initially analyse to extract the pertinent information from it either in the form of text, visual data or the combination of different modalities i.e. video. After the extraction the query will be represented in the form of finite set of feature vectors. The data in the corpus are also delineated with the set of vectors. The relevancy between the query feature vector and the corpus feature vector is calculated by performing various similarity computation methods and the data are then index and retrieved according to the user demands. The accuracy of the retrieved results can be judged by performing the relevance feedback mechanism Retrieval deals with the problem of finding the relevant data from the collection of images or videos according to the user request. The user request may be in the form of the textual data or in the form of query by example. It's relatively easy to extract the low level features from the images and videos in the query as well as in the collection and then compare it.

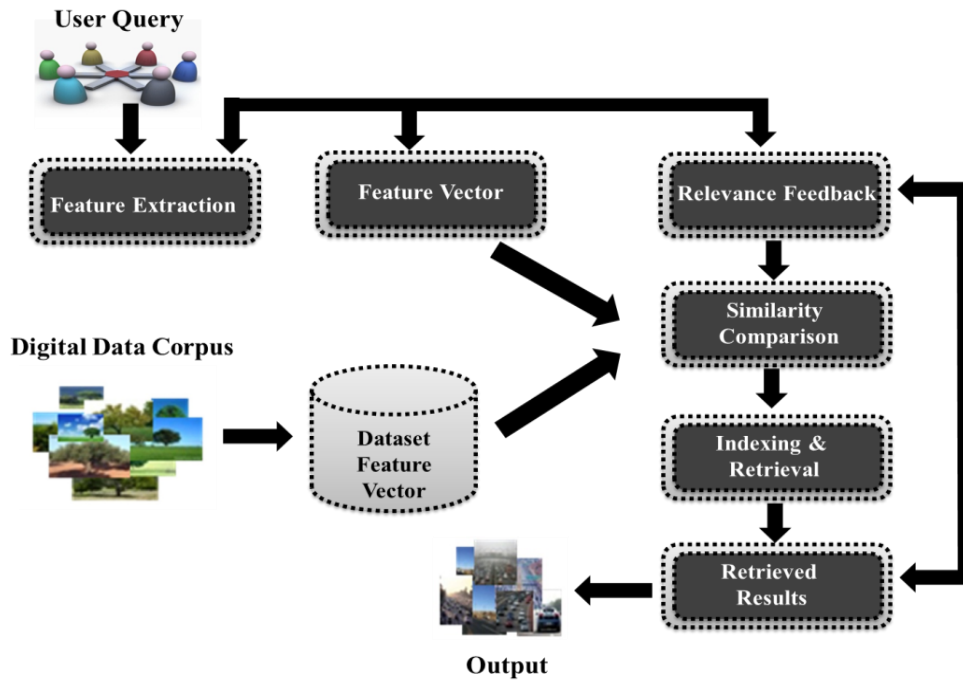


Figure 2.6: Typical Architecture of Content Based Retrieval

The paramount objective of CBR is efficiency during image and video search and retrieval, thereby reducing the need for human intervention. Computer can retrieve the images and videos by using CBR techniques from the large corpus without the human assumption. These low level extracted features then represent the image or video and these features are used later on for performing the similarity comparison between the other images or videos in the corpus. These extracted features serve like a signature for images and videos. Images and videos are compared by using different similarity comparison techniques. They are compared by calculating the dissimilarity of its characteristic components to other image and video descriptors.

CBR approach shows substantial results with the queries like “show me the images or videos of the red colour”, “Show me the image with blue colour is above the green colour” etc. The available automated CBR techniques deal such a type of queries elegantly but flunk to cope with the high level semantic queries like “Show me the images or videos of the people in the park”, people on the beach, car on the road etc. Such type of queries cannot be tackled successfully by the CBR systems. These queries require more sophisticated techniques to extract the actual semantics abstracted inside it. Related work in CBR from the perspective of images can be found from the overview studies of [Rui et al. 1999], [Smeulders et al. 2000],

[Vasconcelos et al. 2001], [Eakins 2002], [Kherfi et al. 2004], [Datta et al. 2005], [Chen et al. 2004], [Dunckley 2003], [Santini. 2001] [Santini et al.2001], [Lew et al. 2001], and [Bimbo et al. 1999].

CBIR has received considerable research interest in the last decade [Vasconcelos et al. 2001] and has evolved and matured into a distinct research field. The CBIR mainly comprises of two main steps feature extraction and the similarity measurement. These key technical components of the CBIR system will be introduced in the following sections.

2.5.2.1 Feature Extraction

"Images are described by visual words just like text is defined by textual words"

In fact, an image or a video frame is merely a rectangular grid of coloured pixels for a computer. And to a computer an image doesn't mean anything, unless it is told how to interpret it. Image and video descriptors are intended for the motive of image or video retrieval. Descriptors seek to apprehend the image or video characteristics in such a way that it is facile for the retrieval system to identify how similar two images or videos are according to the user's interest. CBR system index images or videos by using the low-level features of the image and videos itself, such as colour [Pass et al. 1998, Smith et al. 1996a, Swain et al. 1991], texture [Manjunath et al.1996, Sheikholeslami et al. 1994, Smith et al. 1996b], shape [Safar. M et al. 2000, Shahabi et al. 1999, Tao et al. 1999], and structure features [Pickering et al. 2003, Howarth et al. 2005]. The colour, shape and texture are the principal features of the images. The visual contents of images and videos are then symbolized as a feature vector of floating numbers. For example, the colour, texture and shape features extracted from an image form an N-dimensional feature vector, and can be written as

$$N = \{n_1, n_2, n_3\} \quad (2.1)$$

Where n_1, n_2, n_3 is a vector of its own, and n_1 is the colour, n_2 is texture and n_3 is the shape. While for the video there is an additional vector n_4 , where n_4 is the motion. In the following section, we introduce the visual features to give an impression of how images and video frames can be converted into a representation that the retrieval system can work with.

A Colour

A very common way to see at images is by analysing the colours they contain. Colour is the most prominent visual feature in CBIR since it is well correlated with human visual perceptions of objects in an image. A digital colour image is represented as an array of pixels, where each pixel contains three or four tuples of colour components represented in a numerical form. The abstract mathematical representation of colours that computers are able to use is known as the colour model.

The similarity between the images and the videos is calculated by using the colour histogram value. The histogram depicts the specific values of the pixels inside the image or video frame. The current colour based retrieval techniques divides the image into regions by using colour proportion. The colour based technique doesn't depend on the size and orientation of an image.

Since 1980s various colour based retrieval algorithms have been proposed [Smith et al. 1996 c]. A most basic form of colour retrieval involves specifying colour values that can be further used for retrieval. Indeed, Google's image and Picasa 3.0 can also provide the facility to the user to search the images that contain homogenous colour composition. The most common representation of colour information is in the form of colour histogram and colour moment. Colour anglogram [Zhou X.S. et al. 2002], correlogram [Huang J. et al 1997], color co-occurrence matrix (CCM) [Shim S. et al. 2003] are some of the other feature representations for colour. The general colour based interpretation of an image is shown in Figure 2.7.

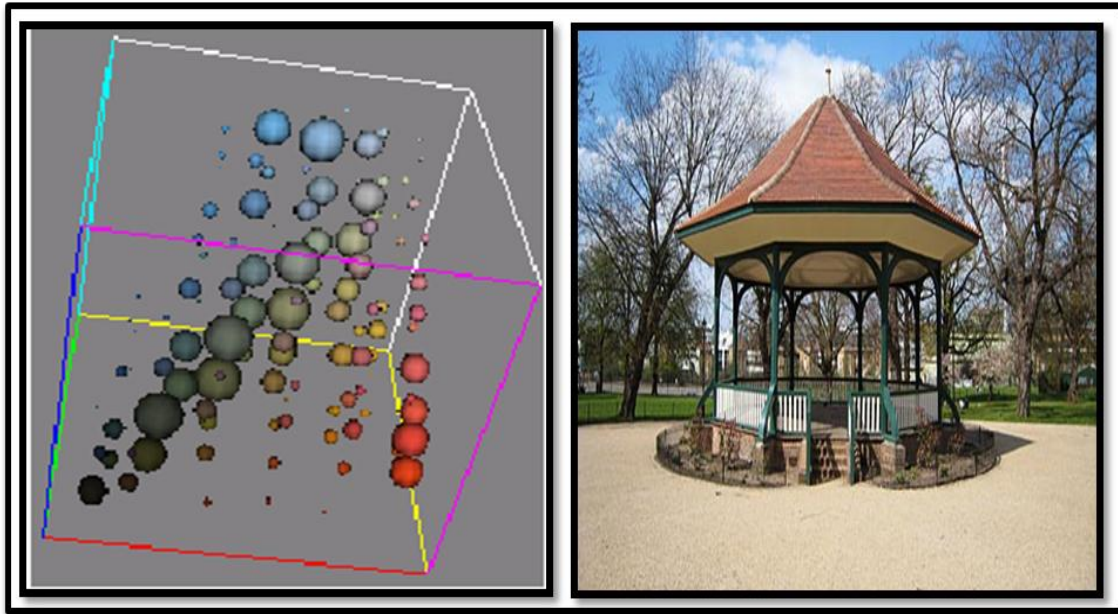


Figure 2.7: Color based image interpretation

Colour Spaces

There are many colour spaces designed for different systems and standards, but most of them can be converted by a simple transformation.

- i. **RGB (Red-Green-Blue):** Digital images are normally represented in RGB colour space; it is the most commonly use colour space in computers. It is a device dependent colour space, which used in CRT monitors.
- ii. **CMY (Cyan-Magenta-Yellow), CMYK (CMY-Black):** It is a subtractive colour space for printing, it models the effect of colour ink on white paper. Black component is use for enhancing the effect of black colour.
- iii. **HSB (Hue, Saturation, Brightness) or HSV (Hue, Saturation, Value):** It was used to model the properties of human perception. It is an additive colour model. However it is inconvenient to calculate colour distance due to its discontinuity of hue at 360° . The HSV model is shown in the Figure 2.8.
- iv. **YIQ, YCbCr, YUV:** Used in television broadcast standards. Y is the luminance component for backward compatibility to monochrome signal and other components are for chrominance. It is also used in some image compression standards (e.g. JPEG) that process luminance and chrominance separately.

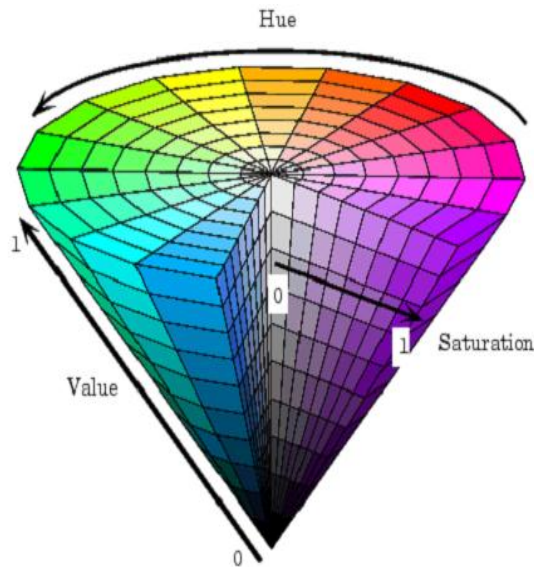


Figure 2.8: The additive colour model HSV

Colour Models

A colour model is an abstract mathematical model describing the way colours can be represented as tuples of numbers, typically as three or four values or colour components. When this model is associated with a precise description of how the components are to be interpreted (viewing conditions, etc.), the resulting set of colours is called colour space [Wiki colour Space]. A colour model is a formularized system for composing different of colours from a set of primary colours. There are two types of colour models, subtractive and additive.

An **additive colour model** uses light emitted directly from a source. The additive colour model typically uses primary colour i.e. red, green and blue light to produce the other colours. Combination of any two of these additive primary colours in equal amounts produces the additive secondary colours or primary subtractive model colours i.e. cyan, magenta, and yellow. Integration of all these three colours RGB in equal intensities constitute white as shown in the Figure 2.9 a.

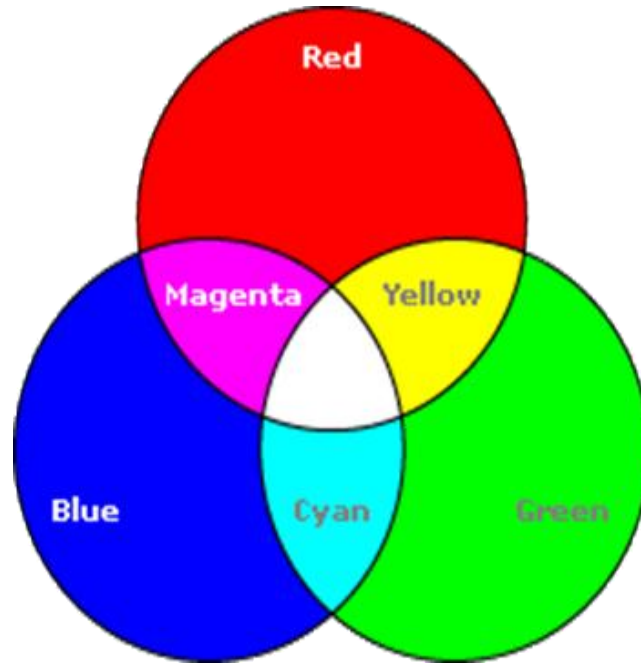


Figure 2.9 (a): RGB: Additive Colour for light-emitting computer monitors. Each coloured light "add" to the previous coloured lights.

A subtractive colour model illustrates the blending of paints, dyes, and natural colorants to produce a full series of colours, each generated by subtracting (absorbing) some wavelengths of light and reflecting the others. Colours observed in subtractive models are the due to reflected light. Different wavelength lights constitute different colours. The CMYK model (Cyan-Magenta-Yellow-blackK) model is the subtractive model. The combination of any two of these primary subtractive model colour i.e.(Cyan, Magenta, Yellow) results in the primary additive model or secondary subtractive model colour i.e. red, blue, green and the convergence of it constitute black colour as shown in Figure 2.9 b.

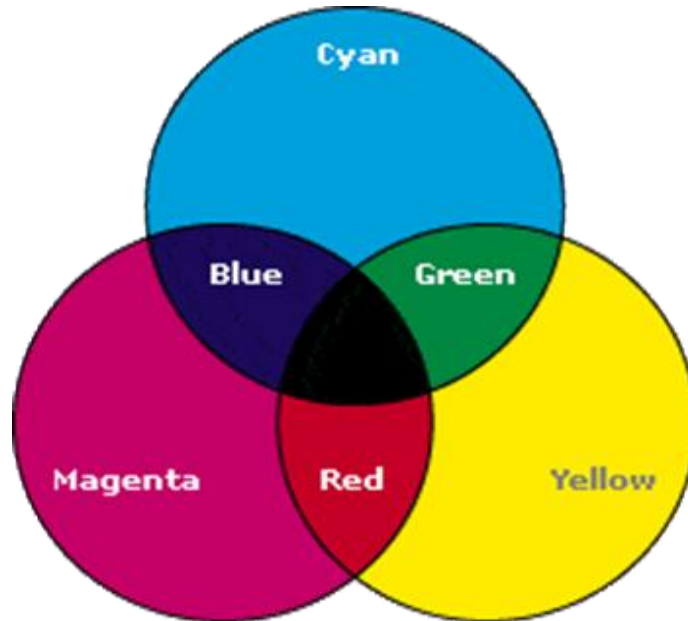


Figure 2.9 (b): CMYK: Subtractive colours for Printer. Each color added to the first colour blocks the reflection of colour, thus 'subtracts' colour.

For some of the concepts the colour scheme helps in achieving suitable results like forest, sky, tree, grass, sea etc. The colour descriptor will help in retrieving the accurate results. But for the categories like the car, house, street etc. Colour descriptors can't play a vital role. The colour descriptor will fail in a situation of the same car with different colours as shown in Figure 2.10. The colour based retrieval system fails to find the relevancy between the two images containing the semantically identical but visually different objects. For the retrieval based on the colour two most frequently used representatives are colour histogram and colour moment. These representatives are represented in the section below.



Figure 2.10: Same Car with different color composition

a Colour Histogram

A histogram provides a summary of the distribution of a set of data. A colour histogram provides a comprehensive overview of the image or video frame in terms of color. A colour histogram for a coloured image describes the different intensity value distributions for colours found in the image. The histogram intent to define the number of times each colour appears in an image and video frame. Statistically, it utilizes a property that images having similar contents should have a similar colour distribution. One simple approach is to count the number of pixels of each colour and plot into a histogram. The histogram h of an image I is represented as

$$H(I) = \sum_{i=1}^N P_i \quad (2.2)$$

Where p_i is the percentage of i -th colour in the colour space, N is the number of colours in the colour space. To enable scaling invariant property, the histogram sum is normalized to 1. The percentage is proportional to the number of pixels in the image.

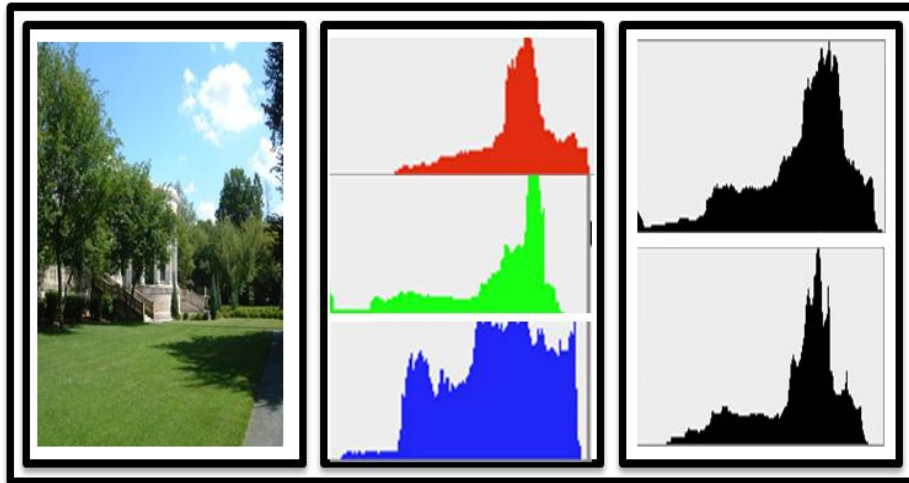


Figure 2.11: Colour Histogram

Mostly commercial CBR systems like Query-By-Image-Content uses colour histogram as one of the feature for the retrieval. Colours are normally grouped in bins, so that every occurrence of a colour contributes to the overall score of the bin it belongs to. The bin explains the intensities of different primary colour i.e. quantity of red, blue or green for a particular pixel. It doesn't define individual colour of the pixels. Histograms are usually normalized, so that images of different sizes can be fairly compared. Figure 2.11 depicts the histogram representation of an example image where x axis in the color histogram represent the color of each pixel while y axis define the intensity of each pixel color in RGB space. The colour histogram is the most commonly and effectively used colour feature in CBIR [Swain et al. 1991, Faloutsos et al. 1994, Stricker et al. 1995, Deselaers et al. 2008, Chakravarti et al. 2009 and Smeulders et al. 2000]. Retrieving images based on the colours technique is widely used because it does not depend on image size or orientation.

The most common method to create a colour histogram is by splitting the range of the RGB intensity values into equal-sized bins. For example, a 24-bit RGB colour space contains 224 possible (RGB) values. Since this gives us approximately 16.8 million bins, it will be too large to be dealt with efficiently. Therefore, we need to quantize the feature space to a smaller number in order to reduce memory size and processing time, as examples [Stricker et al. 1995, Swain et al. 1991] have proposed techniques for colour space quantization. After having defined the bins, the numbers of pixels from the image that fall into each bin are counted. A colour histogram can be used to define the different distributions of RGB intensity values for a whole image, known as a global colour histogram, and for specific regions of an image,

known as a local colour histogram. For a local colour histogram, the image is divided into several regions and a colour histogram is created for each region.

A histogram refinement strategy has been proposed by Pass for comparing the images [Pass et al.1996]. Histogram refinement splits the pixels in a given bucket into several classes, based upon some local property. Within a given bucket, only pixels in the same class are compared. They describe a split histogram called a colour coherence vector (CCV), which partitions each histogram bucket based on spatial coherence. [Han et al. 2002] proposed a new colour histogram representation, called fuzzy colour histogram (FCH), by considering the colour similarity of each pixel's colour associated to all the histogram bins through fuzzy-set membership function. This approach is proves very fast and is further exploited in the application of image indexing and retrieval.

The paradigm of the colour histogram works on the assumption that all the images or videos frames with the similar colour composition are similar [Jain et al. 1995]. It will retrieve all the data whose colour composition is similar to the given query. This will be true in some cases. Colour composition can't be the identity of the image or object inside the image.

b Colour Moment

Colour moment approach was proposed by [Stricker et al. 1995]. It is a very compact representation of colour feature. The mathematical meaning of this approach is that any colour distribution can be characterized by its moments. Moreover, most of the information is concentrated on the low-order moments, only the first moment, second and third central moments (mean, variance and skewness) were extracted as the colour feature representation. Colour similarity can be measured by Weighted Euclidean distance. Due to the ease and sound performance of colour histogram technique it is widely used in colour based retrieval systems.

Colour is the human visual perceptual property. Human discriminate an images or objects initially on the basis of colours. Colour can be extracted from the digital data easily and automated and effective functions are available for calculating the similarity between the query and the data corpus. Colour feature are effectively used for indexing and searching of colour images in corpus.

The existing CBIR techniques can typically be categorized on the basis of the feature it used for the retrieval i.e. colour, shape, texture or combination of them. Colour is an

extensively utilized visual attribute that plays a vital role in retrieving the similar images [Low et al. 1998]. It has been observed that even though colour plays a crucial role in image retrieval, when combined with other visual attributes it would yield much better results [Hsu et al. 1995]. This is because, two images with entirely similar colour compositions, may have different Semantic idea and sometimes two images have same colour composition but they are not similar as shown in the Figures below. Hence something that looks similar is not semantically similar. The colour composition of both the images in Figure 2.12 is same but they depict the entirely different semantic idea. By analysing both the images using the colour based retrieval techniques both the images are similar but unfortunately there is no semantic similarity between the two images. While in the Figure 2.13 contain two groups of images with the similar semantic idea but have completely different colour composition. The colour based retrieval system flunks to recognize the relevancy between the two images.





Figure 2.12: Images with Similar Colour Composition But different Semantics





Figure 2.13: Different Colour Composition but Similar Semantic idea

B Shape

Among the primitive features shape is one of them. Shape in an image refers to the shape of the regions in an image. Shape is well defined concept used for the computing the similarity between the images rather than the texture and colour. Shape deals with the spatial information of an image. [Biederman. 1987] proved that natural objects can be recognized by their shape. Shapes of all the objects in an image are computed to identify the objects in the image. Shapes with in the image can be represented in terms of curves, lines, Eigen shapes, points, medial axis etc. Shapes can be expressed in terms of various descriptors like moments, Fourier descriptors, geometric and algebraic invariants, polygons, polynomials, splines, strings, deformable templates, and skeletons [Kimia. 2001].

Study [Posner et al.1989], [Biederman. 1985] showed that shape based retrieval has been less investigated by researcher than the colour and texture. The reason is due to its complexity of identifying the shapes of the objects in an image. There are no standards available for the shape extraction and matching.

Shapes are described in terms of the numeric values known as the shape descriptors. The shape descriptors contain the feature vector of a given shape. These feature vectors uniquely identifies the particular shape, which are then further use for the shape matching. The shape analysis can be divided into two main categories i.e. boundary (external) and global (internal) [Loncaric et al. 1998], [Veltkamp et al. 2000].

Shape based retrieval mainly include the image segmentation i.e. dividing an image into various segments. Image segmentation includes techniques to locate objects and boundaries (lines, curves, etc.) in images. Image segmentation is a combination of segments that collectively constitute the entire image. The need to describe shapes mathematically leads to the two general methods for shape representation and description: region-based methods and contour-based methods [Zhang et al. 2004].

In **region-based methods**, the features are extracted from the whole region. Such region based features include area, length and angle of major and minor axes, and moments. Area is the total number of pixels inside a region. Based on the area, we can obtain the mean colour of the region that is the average colour value within the region. This is the sum of the colour values of all the pixels in the shape divided by the number of pixels.

Contour-based methods represent a shape by a coarse discrete sampling of its perimeter. Contour-based shape descriptors include perimeter, Hausdor distance, shape signature, Fourier descriptor, wavelet descriptor, scale space, auto regression, elastic matching and shape context [Zhang et al. 2004].

Region-based shape descriptors are often used to discriminate between regions with large differences [Zhang et al. 2004], and are usually combined with contour-based features. Shape matching is performed by comparing the region-based features using vector space distance measures, and by point-to-point comparison of contour-based features. Measuring shape complexity is necessary to recognize the shapes.

Among the simple complexity shape descriptors are circularity and compactness (also known as thinness ratio and circularity ratio). These two shape descriptors belong to both region-based and contour based methods. Circularity is calculated as

$$\text{Circularity} = \frac{\text{Perimeter}^2}{\text{Area}} \quad (2.3)$$

Compactness reflects how circular the shape is. It is calculated using the formula [Costa et al. 2000]

$$\text{Compactness} = 4\pi \left(\frac{\text{Area}}{\text{Perimeter}^2} \right) \quad (2.4)$$

C Texture

Texture refers to visual patterns in images and their spatial definition. Textures are denoted by **texels** which are then located into a number of sets, relying on how many textures an image comprises. These sets define the texture along with their location.

Textures comprise of a specific type of pattern, which generally have a very homogeneous structure, however this is not always the case. It is an intrinsic property of virtually all surfaces, such as clouds, trees, bricks, grass, etc. some of the textures are shown in the Figure 2.14. It comprises of significant information about the structural interpretation of surfaces and their relationship to the adjacent environment. Since last few decades much of the research has been done in the pattern recognition and computer vision. Now, many of these techniques are applied on CBIR. Textures are an important part of life, since they often are an intrinsic quality of a particular object.



Figure 2.14: Various types of Textures

Texture is a troublesome concept to illustrate. The recognition of particular textures in an image is accomplished mainly by modelling texture as a two-dimensional gray level variation. The relative brightness of pairs of pixels is computed such that degree of contrast,

regularity, coarseness and directionality may be estimated [Tamura et al. 1978]. However, the problem is in recognizing patterns of co-pixel variation and relating them with specific classes of textures such as a silky, or rough.

Co-occurrence Matrix

Haralick [Haralick et al. 1973] proposed the co-occurrence matrix representation of texture features. This approach explored the gray level spatial dependence of texture. Initially the co-occurrence matrix is constructed by using the distance between image pixels and orientation. These matrixes are then used to represent the various textures.

Tamura Texture

Tamura et al. [Tamura et al. 1978] explored the texture representation from a different angle. They developed computational approximations to the visual texture properties and applied psychology studies. The six visual texture properties were

- i. **Coarseness:** Related to the size of the image elements in a texture.
- ii. **Contrast:** Related to the sharpness of the edges and period of repeating of texture elements.
- iii. **Directionality:** The shape and placement of the texture elements.
- iv. **Line Likeness:** Decide the texture element is like a line or not.
- v. **Regularity:** Variation between placements of the texture elements in the texture.
- vi. **Roughness:** Measure the texture is rough or smooth.

Wavelets texture representations

After the introduction of the wavelet transformation technique in early 1990s, researchers have explored the ways to apply the wavelet to represent the textures in images [Gross et al. 1994]. In a more recent paper written by Ma and Manjunath [Ma et al. 1995], they evaluated the texture image annotation using various wavelet transform representations, including orthogonal and bi-orthogonal wavelet transforms, the tree-structured wavelet transform, and the Gabor wavelet transform. Smith and Chang [Chang et al. 1996] found that the Gabor transform was the best among the tested candidates which matched human vision study results.

D Motion features

Motion is the prime attribute expressing temporal information of videos. And it is more reliable compared to other features such as colour and texture, etc. [Wu et al. 2002]. Efficient motion feature extraction is a significant progression for content-based video retrieval. Motion feature is significant for the description of video content. Video retrieval based on motion features is one important part of retrieval applications in video database. For instance, when browsing the video obtained by surveillance system or watching sports programs, the user always has the need to find out the object moving in some special direction.

In MPEG-7 [He. 2000], parametric motion descriptor is defined which represents the global motion information in video sequences with 2D motion models. Global motion is the movement of background in a frame sequence and it is mainly caused by camera motion. Global motion information represents the temporal relations in video sequences. Compared with other video features, it can represent the high-level semantic information better. And it is important for motion based object segmentation, object tracking, mosaicing, etc. Motion-based video retrieval can be implemented by parametric motion descriptor on the basis of appropriately defined similarity measure between motion models.

Some related work has been done in the aspect of extraction of motion descriptors. Jeannin et al. [Jeannin et al. 2000] proposed their algorithms for extraction of camera motion descriptor and motion trajectory descriptor. In their algorithm, the extraction of motion trajectory descriptor was based on the assumption that the object was already segmented correctly and they didn't deal with the problem of object segmentation. Kang et al. [Kang et al. 1999] proposed their algorithm on compressed domain data, and only did their work on camera motion analysis. Divakaran et al. [Divakaran et al. 2000] focused on motion activity descriptor, which described the activity in a video sequence in a whole.

2.5.2.2 Similarity Measurement

After the extraction of the visual features of images and videos, these features are then compared against the query to find the degree of relevancy among them. For CBIR we need to be able to take a query and the feature space and produce a ranked order of images and videos that reflect the user need. A large numbers of different similarity measures are used by the research community. The choice of similarity measure depends on the chosen image descriptor,

and may require designing a unique similarity measure if no existing ones are suitable. In this section we will discuss a selection of widely used measures for metric-based and histogram-based image descriptors.

Metrics

When the image descriptor consists of a coordinate vector that indicates a point in a multi-dimensional metric space, the similarity between descriptors is commonly determined by calculating the distance between their points in space. Various metrics can be used for this calculation.

Manhattan metric

Also known as the L1 distance, this similarity metric measures the distance d between two points $x = \{x_1, x_2, \dots, x_n\}$ and $y = \{y_1, y_2, \dots, y_n\}$ as the sum of their absolute coordinate differences

$$d_{L_1} = \sum_{i=1}^n |x_i - y_i| \quad (2.5)$$

Other names for this metric are the city block distance and taxicab distance, since they refer to the shortest distance between two points in a city where the streets are laid out in a rectangular grid, such as is the case on Manhattan Island, New York.

Euclidean Metric

This metric is commonly referred to as the L2 distance, and measures the shortest path between the two points

$$d_{L_2}(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2.6)$$

When a researcher simply states "the distance between points A and B is X", without specifying which distance measure is used, the Euclidean distance is generally implied, since it is the most commonly used similarity measure.

Minkowski metric

The Minkowski metric is a generalization of the L_1 and L_2 metrics, where the order parameter p controls how the distance is calculated

$$d_{L_p}(x, y) = (\sum_{i=1}^n |x_i - y_i|^p)^{1/p} \quad (2.7)$$

Choosing $p=1$ results in the Manhattan distance, choosing $p=2$ in the Euclidean distance and choosing $p=\infty$ in the Chebyshev distance. Fractional distances can be obtained by choosing $0 < p < 1$. Note that such distances are not metric because they violate the triangle inequality.

There is a clear gap between the cognitive nature of a human similarity assessment and the deterministic similarity function.

2.5.2.3 Query Paradigm

Query refers to as "a formal specification of the user's information need". Queries play a vital role in the success of the retrieval system. Query processing in the content based retrieval systems finds the similar images and video according to the query and then ranked in the descending similarity order. Queries for similar images and videos are normally posed via a user interface. This user interface can be in the form of text or more commonly a graphical user interface. Queries can be either the simple textual description or an example image or video. Content based systems can have the text based as well as content based queries. While the text

based retrieval system have only textual queries. Text-based queries can be formulated in free-text or according to a query structure. Free-text queries are normally formulated for retrieving images/videos using the full-text information retrieval approach. Structured queries are used in image retrieval systems that are based on a particular structured representation such as XML, RDF, and OWL etc. Content-based queries use the extracted image and video features as the query examples. These features are compared to the image and video features in the dataset, and similar images and videos are retrieved. Some of the basic CBR query paradigms are discussed below.

a) **Sketch Retrieval based Query**

One of the preliminary studied approaches for retrieving the multimedia data is query by sketch. By means of this paradigm, the user sketch their needs, the sketch may be either by using colour, by drawing different geometrical shapes or by providing different textures. The system then extracts the features from these drawing and then uses these features for searching the visually similar images. Relevance feedback is commonly used. The shortcoming of this approach is that it is very time consuming. It requires the user to have the complete grasp over drawing the sketches because these sketches are used to extract their information needs. Due to these drawbacks this method is not widely used.

b) **Query-by-Example**

Sketch based method is limited because of its time consuming nature and inability of the system to extract the exact features from the visual sketches. This leads to the surge for the system that will be easy to the ordinary user and involves less human intervention. In order to cope with these needs the researchers comes up the new query paradigm i.e. query by example [Faloutsos et al. 1994].

Query by example refers to the technique which includes the input query in the form of image and video example. Which will be further used for searching and retrieval? This technique work on the principal that first the features from the query example are extracted and were used to index the data. Query by example systems also contain the relevance feedback approach in order to check the accuracy of the result. Relevance feedback comprehends the technique of taking the initially retrieved results from a given query and exploits these results in order to check whether those results are relevant or not to perform a new query. The

aspiration behind the technique is that “the images and videos are difficult to define in terms of words”. The features that are used to perform the search may be either colour, shape, texture, motion or the combination of them. The wide range of different interpretations of an example makes this approach more useful when the user provides more than one example to disambiguate the information need [Heesch. 2005], [Rui et al. 1997]. The phenomena of the simple query by example technique are shown in the Figure 2.15.



Figure 2.15: Different Query Paradigm

c) Query by Keyword

Query-by-keyword is by far the most popular method of search query. The user interprets their needs in the form of the single keywords or the combination of keywords. These systems mostly rely on the annotation (meta-data) tag with the images and videos. The system then uses these textual data to search for a particular data [Magalhaes et al. 2007], [Yavlinsky. 2007]. One of drawback is that data different people define and interpret it differently. If the same data is not defined with the same vocabulary then the system will not retrieve the data even though they are relevant.

d) Spatial Queries

Spatial queries are related with object spatial positions. Objects positions can be queried in three different ways i.e. spatial relations between two objects can be queried; locations of object can be queried, and object trajectories can be queried. Koprulu [Koprulu et al. 2004] support regional queries, fuzzy spatio-temporal queries, and fuzzy trajectory queries.

2.5.2.4 Existing Content Based Retrieval Systems

With the exponential proliferation in the volume of digit data, searching and retrieving in an immense collection of un-annotated images and videos are attaining the researcher's attention. Content-Based Retrieval (CBR) systems were proposed to remove the tedious task of annotation to computer. Many efforts have been made to perform CBR on the efficient premise based on feature, colour and texture and shape. Comparatively, a few models have been developed, which doesn't retrieve the images or videos on the basis of CBR. Since early 1990's, many systems have been proposed and developed, some of the models are QBIC, Virage, Pichunter, VisualSEEK, Chabot, Excalibur, Photobook, Jacob, UC Berkeley Digital Library Project. [Smeulders et al. 2000], [Rui et al. 1999], [Fend et al. 2003], [Singhai et al. 2010] give general surveys on the technical achievements in the domain of content-based image and video retrieval. These surveys review the different techniques that are used for the extraction of the visual features from the images, the methods that are used for the similarity measurement.

a) QBIC (Query by Image Content)

QBIC is abbreviated as Query by image content is a classical commercial image retrieval model developed by IBM [Niblack et al. 1994]. It was developed to retrieve the images used in art galleries and art museums [Seaborn. 1997]. It employs the primitive feature, i.e. colour, shape and texture for retrieval. It executes CBIR by exploiting various perceptual features, according to the user requirements. In QBIC, it supports query by example, sketch based query, query on the basis of particular colour and texture etc. For the colour representation QBIC employs a partition-based approach and colour histogram [Faloutsos et al. 1994], for pre-filtering the candidate images the average Munsell transformation is used, for texture pattern it uses the refined version of tamura (coarseness, contrast and directionality) texture representation technique, a moment based shape feature to describe shapes includes area, circularity and eventricity. The QBIC system also supports multi-dimensional indexing by using orthogonal transformation, such as the Karhunen-Loeve Transform (KLT) to perform dimension reduction with an R*-tree used as an underlying indexing structure [Faloutsos et al. 1994].

b) **Virage**

Virage was developed by Virage Inc. is a CBIR system for retrieving the ophthalmologic images [Seaborn. 1997]. It exploits colour, texture, and shape for the retrieval. It supports visual queries like query by image and sketch based query. The queries were performed by using global colour, local colour, texture classification and structure. Virage interpret the image in to following layer domain objects and relations, domain events and relations, image objects and relations, and image representations and relations in order to provide the flexibility of simultaneously viewing the data from various abstraction levels. The system also grants the user with an ability to calculate the weight assignment among the visual features [Xu et al. 2000], [Gupta et al. 1991]. Its main features are the ability to perform image analysis (either with predefined methods or with methods provided by the developer) and to compare the feature vectors of two images.

c) **Pichunter**

Pichunter was developed by NEC Research Institute, Princeton, NJ, USA. It utilizes colour histogram and colour spatial distribution together with the textual annotation. Besides a 64-bin HSV histogram, two other vectors - a 256-length HSV colour auto correlogram (CORR) and a 128-length RGB colour-coherence vector (CCV) - are describing the colour content of an image [Cox et al. 2000]. It implements a probabilistic relevance feedback technique by using Bayesian probability theory [Seaborn. 1997]. This system was initially tested on Corel stock photographs. It supports the query by example approach.

d) **VisualSEEK**

VisualSEEk was developed by Image and Advanced Television Lab, Columbia University, NY [Smith et al. 1997 a]. It is a heterogeneous system that deploys the colour percentage method for the retrieval. It combines image feature extraction based upon the representation of colour, texture and spatial layout. The prime uniqueness of the system is that the user diagrammatically forms the queries based on the spatial arrangement. The system is capable of executing a vast variation of complex queries due to an efficient indexing, and also because spatial issues such as adjacency, overlap and encapsulation can be addressed by the

system. The retrieval process is accentuated by using binary-tree based indexing algorithms. The results of a query are demonstrated in decreasing order of similarity. The results are displayed along with the value of the distance to the query image.

e) **Image Rover**

ImageRover [Sclaroff et al. 1997] is an image retrieval tool developed at the Boston University. The system combines both visual and textual statistics for computing the image decompositions, textual associations, and indices. The extracted visual features are stored in a vector form using colour and texture-orientation histograms, while the textual features are captured using Latent Semantic Indexing based on associating the related words in the containing HTML document [Cascia et al. 1998]. The relevance feedback technique is used to refine the initial query results. The system performs relevance feedback using Minkowski distance metric, and the retrieval process is accentuated by using an approximate K-nearest neighbours indexing scheme [Sclaroff et al. 1997], [Taycher et al. 1997]. At the beginning of a search session, the user types a set of keywords connected to the images he/she is looking for. In the further stages, the user adds/removes images from a relevant images set

f) **Chabot**

It was developed by Department of Computer Science, University of California, Berkeley, CA, USA to retrieve the images. Chabot intended to integrate text based descriptions with image analysis in retrieving images from a collection of photographs of the California Department of Water Resources [Virginia. et al. 1995]. Queries are composed of textual information, date information, numerical information and colour information reflecting the target image's data.

g) **Excalibur**

Excalibur was developed by Excalibur Technologies. It is the hybrid system that incorporates some of the properties of QBIC and Virage like standard metrics, colour, texture and shape and benefits the image ration property of Pichunter. It admits queries by example based on HSV colour histograms, relative orientation, curvature and contrast of lines in the image, and texture attributes, that measure the flow and roughness in the image. The user initially specifies the desired visual similarity by specifying the relative importance of the

above image attributes, and then selects one of the displayed images as query [Seaborn. 1997]. The images are shown without an explicit ordering.

h) Photobook

Photobook was developed by Vision and modelling Group, MIT Media Laboratory, Cambridge, MA [Pentland et al. 1996]. It benefits statistical analysis, color percentage and texture for retrieval. Photobook executes three different approaches to constructing image representations for querying purposes, each for a specific type of image content, faces, 2D shapes and texture images. Picard and Minka [Picard et al. 1995] suggested incorporating human in the image annotation and retrieval loops in the latest version of Photobook. The face recognition technology of Photobook has been used by Visage Technology in a FaceID package, which is used in several US police departments. Experimental results reveal the effectiveness of their approach in interactive image annotation.

Queries are composed using example images either single image or multiple images. The system interacts with the user and performs retrieval based on text annotations. Other visual attributes are also integrated in order to improve the quality of the retrieval. The comparison is carried out on the extracted feature vectors with consideration on invariance to scaling and rotation. The Photobook 5 caters a library of matching algorithms for calculating the linear distance among a set of images. While the Photobook 6 allows for matching user-defined algorithms via dynamic code loading. The system includes a distinct interactive agent (referred to as FourEyes) which is capable of learning from the user selection [Hsu et al. 1995].

i) Jacob

JACOB is abbreviated as Just A Content Based query system for video databases was developed by Computer Science & Artificial Intelligence Lab, University of Palermo, Italy. It utilizes the texture, motion and colour feature to retrieve the video from the corpus. The system performs queries based on colour and texture features. Colour is illustrated by a histogram in the RGB space. Texture features used are to measures extracted from the grey-level co-occurrence matrix, the maximum probability, and the uniformity [Cascia et al. 1996]. The queries may be direct or by example. A direct query is made by inserting a few values describing the colour histogram or the texture features or the integration of both. Two colour histograms are compared using the distance measure. The results are arranged in the

descending order of similarity. The number of returned frames is chosen by the user. By choosing the returned frame the user can view the connected shot.

j) WebSEEk

WebSEEk was developed by Image and Advanced Television Lab, Columbia University, NY and is a text and image search engine [Smith et al. 1997 b]. WebSEEk is a catalog-based search engine for the World Wide Web and makes text-based and content based queries for the images and videos. The results are displayed according to the decreasing colour similarity to the selected item. Colour is represented by means of a normalized 166-bin histogram in the HSV colour space. A manipulation (union, intersection, subtraction) of the search result lists, in order to reiterate a query, is possible [J. R. Smith 1997]. It comprises of various modules image and video collection module, indexing module, search module, classification module, and browse and retrieval module.

k) Blob world

Blob World system was developed by Computer Science Division, University of California, Berkeley. This system is presented in [Carson et al. 2002], [Carson et al. 1998]. The system uses image features which are extracted using segmentation of images. This segmentation is done using an EM-style algorithm clustering the image pixels using colour, texture, and position information. To query the database the user selects a region from an image and the system returns images containing similar regions.

Queries used the following primitive features i.e. colour, texture, location, and shape of regions (blobs) and of the background. Queries are composed of regions, selected from an example image contained in the database. Regions (called blobs by the authors) must be chosen from a segmented version of the image, and a maximum of two regions can be selected, one being the main subject of the query, while the second represents the background. For each region, an overall importance level can be defined (very or somewhat important). The importance of colour, texture, position and shape of the selected blob can also be defined (not important, somewhat important, and very important). The retrieved images displayed in linear order, along with the segmented version.

l) MARS

MARS is abbreviated as Multimedia Analysis and Retrieval System was developed by Department of Computer Science, University of Illinois at Urbana-Champaign, further developed at Department of Information and Computer Science, University of California at Irvine, CA [Rui et al. 1997]. The system supports queries on integration of content based (colour, texture, shape) and textual based. Colour is represented using a 2D histogram, texture is represented by two histograms, for coarseness and directionality and one scalar defining the contrast. Histogram intersection is used to compute the similarity distance between two colour histograms. While the similarity between two textures of the whole image is determined by a weighted sum of the Euclidean distance between contrasts and the histogram intersection distances of the other two components, after a normalization of the three similarities. MARS formally proposes relevance feedback architecture in image retrieval and integrates such technique at various levels during retrieval. Images are listed in order of decreasing similarity.

m) Netra

Netra a region-based system was developed by Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA [Ma et al. 1999]. Images are retrieved using the primitive features like shape, texture, colour and spatial location. Images are divided into various regions according to the colour homogeneity. From these regions the primitive features are extracted.

The user selects any one of image as the query image. The user can select on one of the regions and select one of the any image attribute colour, spatial location, texture, and shape. The images are match according to the linear ordered. The query is composed of a region coming from a pre-segmented database image. The user selects the region and can query the database according to similarity in the colour, texture, shape and position domains. It is also possible to directly select colours from a colour codebook and to draw a rectangle to specify the position of the wanted region.

n) SIMBA(Search IMages By Appearance) system

SIMBA were developed by Institute for Pattern Recognition and Image Processing, Freiburg University, Germany [Siggelkow et al. 2001a], [Siggelkow et al 2002], [Siggelkow et

al. 2001b], [Siggelkow et al. 1997]. It uses the feature using colour and texture and uses features invariant against rotation and translation. By a weighted combination the user can adapt the similarity measures according to user needs.

o) VIPER (Visual Information Processing for Enhanced Retrieval)

Viper is proposed by D. Squire, W. Muller, H. Muller [Squire et al. 1999]. Queries are composed of a set of relevant images and another set composed of non-relevant images. The user can refine a query by selecting images in the query output as relevant or not relevant. Retrieval by global similarity in a heterogeneous image database. The image representation is borrowed from text retrieval: each image is represented by a set of more than 80'000 binary features (terms), divided into colour and texture features. These features simulate stimuli on the retina and early visual cortex. Colour features are obtained by quantizing the HSV space. A histogram is computed for the whole image, as well as for recursively divided blocks (each block contains 4 subblocks). Texture features are computed using Gabor filters at 3 scales and 4 orientations.

p) SIMPLIcity

SIMPLIcity is the region based system. It integrates the region-based approach to the semantic classification technique, and segments an image into regions. It partitions the image into blocks of 4 x 4 pixels and extracts a feature vector of six features from each block. Three features represent the average colour components, the remaining three representing texture information. After the partition, the images are then cluster according to their feature vector by using the k-means algorithm, each cluster corresponds to one region in the segmented image. SIMPLIcity performs only global search (i.e. uses inclusive properties of all regions of images) and does not allow the retrieval based on a particular region of the image. Simplicity [Wang et al.2001] incorporates the properties of all the segmented regions so that information about an image can be fully used. To segment an image, the systems partition the image into blocks and extract a feature vector for each block. Feature clustering is performed by using the k-means clustering algorithm. The feature vectors are cluster in to various classes and every class represents the one region. There are six features that are used for segmentation. Three of them are colour components (LUV colour space), and the other three represent energy in high frequency bands of the wavelet transform.

q) **CBIRD (Content-Based Image Retrieval from Digital libraries)**

CBIRD is an image retrieval system that combines automatically generated keywords and several visual features and feature localization to index both images and videos in the web [Sebastian et al. 2002]. This system uses colour channel normalization for finding similar images present in different illumination conditions. They also present a technique to search by object model.

r) **Informedia**

Informedia is a very interesting system developed by Wactlar et al [Wactlar et al. 1996] to perform video retrieval by using speech and image processing [Posner. 1989]. The features from the videos are by using the following techniques colour histograms, speech, motion vectors and audio tracks. Videos are then indexed according to these features. For examples include objects in images, important words from audio tracks, captions etc. Since the system uses so many features from videos, it is ideally suited for applications such as retrieval from news and movie archives.

Summary of Existing CBR Systems

Most of those aforementioned systems and much of the past research have procured the CBR from its infancy to the matured stage. Even though in some cases these systems exhibit substantial outcomes but still have limited efficiency. They have concentrated on the extraction of the low-level features. They emphasized on the explicit features of the images and videos. These features are automatic but omit to study the implicit meaning behind the image and video. The hidden meaning or the semantic idea of the image and video can be interpreted solely by analyzing its contents. This leads to the semantic gap. Until and unless, we study the various implicit meanings of images and videos, which cannot be discernible by the content the semantic gap will not reduce.

Limitations of CBIR system

Existing CBR systems have made many significant results for a specific domain but yet haven't made any breakthrough output. Despite of the apparent success of the CBR, existing system are still far from retrieving the relevant result according to the user demand. Number of

research has already been made in CBIR. The commercial image providers, for the most part, are not using these techniques. The main reason is that most CBIR systems require an example image and then retrieve similar images from their databases. Real users do not have example images they start with an idea, not an image. Some CBIR systems allow users to draw the sketch of the images wanted. Such systems require the users to have their objectives in mind first and therefore can only be applied in some specific domains, like trademark matching, and painting purchasing.

The CBR techniques rely on the visual features, ranks results based on similarity to query examples. The CBR system handle the images as the sequence of pixels but these pixel sequence cannot depicts the implicit idea behind them. This is explicitly encouraging for image retrieval, where visually similar objects can depict good query results. CBR systems can accurately retrieve the visually similar results. However, this poses the problem of how to find images that are not visually similar but are semantically similar. Mostly the visually similar results are not semantically similar as shown in the Figure 2.12. Both the images are visually similar but depict completely complementary idea. Hence these visual features are not amenable for searching the semantically relevant results. To solve this problem, Semantic based retrieval has emerged and gained more attention.

2.5.3 Semantic Based Retrieval

Unfortunately, for a computer an image is simply a matrix of n-tuples or a group of pixels and the video is a group of frames, and each frame is perceived like an image. Nowadays, we are in the technological revolution era, still computers are not efficient analysing an image like a human can and extracting semantic content from it. Computers be capable of simply trying to predict what is inside the image by extracting primitive information, like colour histograms, textures, regions, shapes, edges, spatial positioning of objects, but they aren't able to detect the high level semantics like the presence of burning of wood in the street, people playing in the park on the sunny day. For semantic extraction and analysis the manual annotation is done even though the manual process has a lot of weaknesses as already explained above. Presently, the majority image and video retrieval systems only exception to some of them leans on primitive features to extract the syntactic idea i.e. content from images and videos. These systems can effectively extract what is inside the image or video.

Humans can readily perceive the events, scene, people and objects inside the image and video. They can easily comprehend what is indeed happening inside the image or video, i.e. actual semantic. When the user hunts for a specific image and video from the corpus he/she had an idea of the particular data which depends upon his/her perception capability and experience. Human perception is not just an interpretation of a retinal patch, but it is an integration of the retinal patch and our understanding about objects [Datta et al. 2008], which highly depends upon users experiences and background. The two persons can perceive the same image/video differently. This is due to the flexible nature of human. Hence a retrieval system should be capable enough to cope with the flexible nature of human.

Extracting the semantics from the video and image is still an open challenge. There always exists a gap between the low level syntactic feature and high level abstract semantic feature. This gap is due to the difference in the nature of the flexible human perception and hard coded computer. Attempts have already been made to reduce this gap, yet no breakthrough results have been achieved. It is due to the reason that high level semantic idea can't be extracted by using the low level primitive feature. These low level can depicts the contents of the image/video. But unfortunately contents cannot cover the entire meaning inside them. Sometimes the user's needs cannot be expressible in terms of low level features. All these drawbacks inherently constrained the performance of the content based retrieval systems.

Approaches for Semantic Based Retrieval

It is true that combining context with semi-automated high-level concept detection or scene classification techniques, in order to achieve better semantic results during the multimedia content analysis phase, is a challenging and broad research area for any researcher. Although the well-known "semantic gap" [Tamura et al. 1978] has been acknowledged for a long time, multimedia analysis approaches are still divided into two rather discrete categories; low-level multimedia analysis methods and tools, on the one hand (e.g. [Taycher et al. 1997]) and high-level semantic annotation methods and tools, on the other (e.g. [Swain et al. 1996], [Howarth et al. 2005]). It was only recently, that state-of-the-art multimedia analysis systems have started using semantic knowledge technologies, as the latter are defined by notions like ontologies, folksonomies [Rui et al. 1997] and the Semantic Web standards. Their advantages, when using them for creation, manipulation and post-processing of multimedia metadata, are

depicted in numerous research activities. The core idea is to combine such formalized knowledge and a set of features to describe the visual content of an image or its regions, like, for instance, in [Biederman. 1985], where a region-based approach using MPEG-7 visual features and ontological knowledge is presented.

The principal obstacle to comprehend actual semantic-based image retrieval is that semantic description of image is troublesome. Image retrieval based on the semantic meaning of the images is currently being explored by many researchers. This is one of the efforts to close the semantic gap problem. In this context, there are following main approaches:

- Semantic Concept Detection
- Automatic Image and video Annotation
- Relevance Feedback
- Ontologies for Image and Video Retrieval
- Multimodal Fusion

2.5.3.1 Semantic Concept Detection

The prevalent problem of bridging the semantic gap has investigated by many researchers by automatically detecting the high level concepts detectors. The inspiration behind it is that extracting an objects, scene, or events from images and videos will increase the retrieval performance. Detecting high level concepts in image and video domain is an important step in achieving semantic search and retrieval. The research community has long struggled to bridge the semantic gap from successfully implemented automatic low-level feature analysis and extraction (colour, texture, shape) to the semantic content description of images and videos. An emergent research area is the specification of the semantic filters for the detection of a semantic concepts and help in accurate search, and retrieval.

To reduce the semantic gap, one approach is to employ a set of intermediate semantic concepts [Naphade et al. 2004] that can be used to describe visual content in video/image collections (e.g. outdoors, faces, animals). It is an intermediate step in enabling semantic image/video search and retrieval. These semantic concepts comprises to various concepts [Chang et al. 2005] such as those related to people, acoustic, objects, location, genre and production etc. the techniques that are mostly used for these intermediate concept detection are

object detection, object recognition, face detection and recognition, voice and music detection, outdoor, indoor location detection etc.

One of the significant achievements in current years includes automatic semantic classification images and videos in to a large number of predefined concepts that are pertinent and amenable to searching. Automatic semantic classification produces semantic descriptors for the images and videos, analogous to the text documents are represented by some of the textual terms. It can be beneficial and worthwhile for semantically accurate search and retrieval. The cardinal approach in interpreting semantic concepts is extracting low-level features using texture, colour, motion and shape on an annotated data set, and then ranking and retrieving data using the models trained for each concept.

The semantic classification can be applied by using the well-known codebook model [Agarwal et al. 2008]. The codebook model represents the image in terms of the visual vocabulary. The vocabulary contains the semantic modeling of the images at various levels i.e. word level [Boutell et al. 2006], [Gemert et al. 2006], [Mojsilovic et al. 2004] , [Vogel at al. 2007], topic level [Boutell et al. 2006], [Agarwal et al. 2008], [Fei at al. 2005], [Bosch at al. 2008], [Larlus et al. 2006], phrases level (image spatial layout) [Boutell et al. 2006], [Agarwal et al. 2008], [Lazebnik et al. 2006] [Moosmann et al. 2008], [Sudderth et al. 2008].

The code book model visual word vocabulary may be constructed by using different techniques like k-means clustering on image features [Bosch at al. 2008], [Lazebnik et al. 2006], [Nowak et al. 2006], [Winn et al. 2005], [Sudderth et al. 2008]. K-means reduces the variation amongst the clusters and the data, placing clusters near the most frequently occurring features. In comparison to clustering, a vocabulary may be procured by manually labeling image patches with a semantic label [Mojsilovic et al. 2004], [Gemert et al. 2006], [Boutell et al. 2006], [Vogel et al. 2007]. For example, Vogel et al. construct a vocabulary by labeling image patches of sky, water or grass. Semantic vocabulary represents the meaning of an image.

Recent studies reveals the significance of Codebook approach in detecting the semantic concept [Sande et al, 2010 a], [Snoek et al, 2008], [Sande et al, 2010 b], [Jurie et al. 2005]. Several other classification approaches are also available for the semantic concept detection like decision tree classifier (DT), support vector machine classifier (SVM), Association Rule Mining [Witten et al. 2005], Association Rule Classification (ARC) or Associative Classification (AC) [Liu et al, 1998], [Lin et al 2009] . Systems with the best performance in

image retrieval [Iek et al. 2007], [Sande et al, 2010 a] and video retrieval [Snoek et al, 2008], [Wang et al, 2007] use combinations of multiple features for concept detection.

The Large-Scale Visual Concept Detection Task [Nowak et al. 2009] evaluates 53 visual concept detectors. The concepts used are from the personal photo album domain: beach holidays, snow, plants, indoor, mountains, still-life, small group of people, portrait Set of semantic concepts can be defined based on prior human knowledge for developing the semantic concept detectors. The ground truth annotation of each of the concepts is collected. The widely used annotation forum is the TRECVID. The TRECVID'3 has successfully annotated 831 semantic concepts on a 65-hour development video collection [Lin et al. 2003]. The automatic semantic concept detection has been surveyed by many researchers in recent years [Barnard et al. 2003], [Naphade et al. 1998] , [Lin et al. 2003], [Yan et al. 2005], [Yang et al. 2004], [Wu et al. 2004], [Jeon et al. 2003]. Their successes have demonstrated that a large number of high-level semantic concepts are able to be interpreted from the low-level multi-modal features of video collections. In the literature, most concept detection methods are evaluated against a specific TRECVID (TREC Video Retrieval Evaluation) benchmark dataset which contains broadcast news video or documentary video and The Large-Scale Concept Ontology for Multimedia (LSCOM) project was a series of workshops held from April 2004 to September 2006 [Naphade, et al. 2006] for the purpose of defining a standard formal vocabulary for the annotation and retrieval of video.

2.5.3.2 Automatic Image and Video Annotation

The automatic annotation has gained a lot of attraction of the research community in the recent year as an attempt to reduce the semantic gap. The aim of auto-annotation techniques is to attach textual labels (meta data) to un-annotated images/video, as the descriptions of the content or objects in the images. Association of textual descriptions with visual feature is a stepping stone towards bridging the semantic gap problem. This has led to a new research problem known as automatic image and video annotation [Datta et al. 2008], also known as automatic image/video tagging, auto-annotation, linguistic indexing or automatic captioning, automatic Annotation.

Automatic image and video annotation is the attempt to discover concepts and keywords that represent the image and videos. This can be done by predicting concepts to which an object belongs. When a successful mapping between the visual perception and

keyword is achieved, the image annotation can be indexed to reduce image search time. Hence, text-based image retrieval can be semantically more meaningful than search in the absence of any text.

Automated image annotation, intends to and the correlation between low-level visual features and high-level semantics. It emerged as a remedy to the time-consuming and laborious task of annotating large datasets. Most of the approaches use machine learning techniques to learn statistical models from training set of pre-annotated images and apply them to generate annotations for unseen images using visual feature extracting technology.

Automated image annotation can be categorized with respect to the deployed machine learning method into co-occurrence models, machine translation models, classification approaches, graphic models, latent space approaches, maximum entropy models, hierarchical models and relevance language models. Another classification scheme makes reference to the way the feature extraction techniques treat the image either as a whole in which case it is called scene-orientated approach or as a set of regions, blobs or tiles which is called region-based or segmentation approach.

Currently various approaches to automatically annotating images have been proposed [Yang et al. 2006], [Carneiro et al. 2007]. Many statistical models, the translation model [Duygulu et al. 2002], cross-media relevance model (CMRM) [Jeon et al. 2003], Continuous Relevance Model (CRM) [Lavrenko et al. 2004], multiple Bernoulli relevance model (MBRM) [Feng et al. 2004], maximum entropy (ME) [Deselaers et al. 2007], and Markov random field (MRF) [Carlos et al. 2007], Word co-occurrence [Jair et al. 2008] are proposed. Although the keyword distribution carries some semantic information about the image content, its estimation from the co-occurrence of image and keywords often faces severe data sparsity.

In early work on automatic image annotation, Saber and Tekalp [Saber et al. 1996] used colour, shape and texture features; they reported on several algorithms for automatic image annotation and retrieval using region-based colour, region-based shape, and region-based texture features. Another approach is to use the salient objects identified in the images.

2.5.3.3 Relevance Feedback

Rocchio (1971) introduced relevance feedback, an IR technique for improving retrieval performance by optimizing queries automatically through user interaction [Rocchio. 1971].

Relevance feedback is among one of the approaches that are intended to bridge the semantic gap. The relevance feedback intends to obtain the results that are initially returned from a given query and utilize information about whether or not those results are relevant to perform a new query. In relevance feedback, human and computer interact to transform high-level queries to models that are based on low-level features. Relevance feedback is an effective technique employed in traditional text-based information retrieval systems. In some CBIR systems, users are asked to provide the system, as a part of the query, with some extra information such as the level of importance for each feature, or suggesting a set of features to be used in image retrieval. It seems to be an efficient way to help the user modelling his query and to establish a link between the low level and high level features; however, different users may have a different perception of the notion of similarity between image properties. Furthermore it may not even be applicable to explicit the information need of a user exactly as a weighted combination of features of a single query image.

Providing single user and multi-user relevance feedback during the image retrieval process could also be used to alleviate the problems in understanding the semantics in an image as well as to automatically annotate semantic concepts with the low-level image features. [Yang et al. 2006] proposed the S-IRAS system which uses a semantic feedback mechanism in order to improve the automatically derived annotations based on low-level features. It is different from the ordinary CBIR relevance feedback, where the knowledge gained from the relevance feedback is incorporated directly at the semantic level. During the semantic feedback process, the image annotations were learned using two strategies, namely short-term and long-term learning.

In the short-term learning, the query semantics were correlated with the semantic expressions (concepts) based on the example images in the training set

The long-term learning involves refining the semantic expression based on the positive examples learned through the semantic feedback mechanism.

Using multi-user relevance feedback, Chen et al. [2007] constructed a user-centered semantic hierarchy based on the low-level image features. A collective community vote approach was used to classify the images into a specific semantic concept. These concepts are then used to support semantic image browsing and retrieval.

2.5.3.4 Ontologies for Image and Video Retrieval

When focusing solely on the problem of the semantic gap, it is true that ontology bases retrieval remains still an inevitable. Sufficient works has been done in solution of this problem. The term ontology has been used by philosophers to describe objects that exist in the world and their relationships. Ontology consists of a set of definitions of concepts, properties, relations, constraints, axioms, processes and events that describe a certain domain or universe of discourse. Ontology can be defined as “an explicit specification of a conceptualization” [Gruber. 1995].

In the recent years several standard description languages for the expression of concepts and relationships in domain ontologies have been defined, among these the most important are, Resource Description Framework Schema (RDFS), Web Ontology Language (OWL) and, for multimedia, the XML Schema in MPEG-7. Using these languages metadata can be fitted to specific domains and purposes, yet still remaining interoperable and capable of being processed by standard tools and search systems. Nowadays, ontologies are used to appropriately represent a structured knowledge for a domain [Hare et al., 2006b]. Image and video retrieval using ontology is a form of structured-text information retrieval. The ontology can be represented by various ontology representation languages, and XML is the base language used for constructing ontology.

The integration of an ontology in image retrieval can either be used as a guide (for example WordNet) during the retrieval process or as a repository that can be queried from [Hollink et al., 2003; Jiang et al., 2004; Wang et al., 2006, Harit et al., 2005]. [Town et al. 2004a] shows that the use of ontologies to relate semantic descriptors to their parametric representations for visual image processing leads to an effective computational and representational mechanism. Their ontology implemented the hierarchical representation of the domain knowledge for a surveillance system. Town also proposed an ontological query language (OQUEL). The query is expressed using a prescriptive ontology of image content descriptors. The query approach using OQUEL is similar to the approach presented by Makela et al. [Makela et al. 2006] who implement a web system known as “Ontogator “ to retrieve images using an ontology. [Mezaris et al. 2004], [Mezaris et al. 2003] propose an approach for region-based image retrieval using an object ontology and relevance feedback. The approach utilizes an unsupervised segmentation method for dividing the images into

regions that are later indexed. The object ontology is used to represent the low-level features and act as an object relation identifier for example the shape features are represented as slightly oblong, moderately oblong, very oblong. Hollink et al. [Hollink et al. 2004] add the spatial information of the objects as part of the semantic annotations of images. They adopt the spatial concepts from the Suggested Upper Merged Ontology (SUMO) [Niles et al. 2001].

The ontology learning can be categorized into six sub categories learning terms, synonyms, concepts, concept hierarchies, relations, and rules [Cimiano et al. 2006]. Ontological learning can be categorized into four groups [Wei et al. 2010], Lexicosyntactic-based approach [Cimiano et al. 2006], [Ponzetto et al. 2007], [Suchanek et al. 2007], [Navigli et al. 2004], Information Extraction [Cimiano et al. 2005], [Kiryakov et al. 2004], Machine Learning [Fleischman et al. 2002], [Suchanek et al. 2006], [Pasca. 2005] and Data Co-Occurrence Analysis [Sanderson et al. 1999], [Diederich et al. 2007]. A detailed survey of ontology learning methods is also provided by Biemann [Biemann. 2005].

2.5.3.5 Multi-modality Information Fusion

In recent times, multimodal fusion has gained much attention of many researchers due to the benefit it provides for various multimedia analysis tasks. The integration of multiple media, their associated features, or the intermediate decisions in order to perform an analysis task is referred to as multimodal fusion. A multimedia analysis task involves processing of multimodal data in order to obtain valuable insights about the data, a situation, or a higher level activity. Examples of multimedia analysis tasks include semantic concept detection, audio-visual speaker detection, human tracking, event detection, etc.

Research in the CBMR is motivated by a growing amount of digital multimedia content in which video data has a big part. Video data comprises plentiful semantics such as people, scene, object, event and story etc. many research efforts has been made to negotiate the “semantic gap” between low level features and high level concepts. In general, three modalities exist in video namely the image, audio and textual modality. How to utilize multi-modality features of video data effectively to better understand the multimedia content remains a great challenge.

Multimedia data like audio. Image and video are delineated by features from multiple sources. Traditionally, images are represented by keywords and perceptual features such as

colour, texture, and shape. Videos are represented by features embedded in the visual, audio and caption tracks. For example, when detection concept from video, non-visual features were extracted, such as audio features [Pradeep et al. 2010], [Adams et al. 2003], automatic speech recognizer (ASR) Transcript Based Features and Video Optical Character Recognition and Metadata. After the extraction these characters are fused together for extracting the semantic concept.

A multimodal analysis approach for semantic perception of video incorporates a fusion step to integrate the outcomes for various single media analysis. The two main strategies of fusion are early fusion and late fusion. And most of the existing methods for video concept detection are based in these two strategies.

The most widely used strategy is to fuse the information at the feature level, which is also known as early fusion. The other approach is decision level fusion or late fusion [Hall et al. 1997], [Snoek et al. 2005] which fuse multiple modalities in the semantic space. A combination of these approaches is also practiced as the hybrid fusion approach [Wu et al. 2006]. A hybrid system has been proposed that utilizes the benefits of both the strategies of feature and decision level.

- i. **Visual features.** It may include features based on color (e.g. color histogram), texture (e.g. measures of coarseness, directionality, contrast), shape (e.g. blobs), and so on. These features are extracted from the entire image, fixed-sized patches or blocks, segmented image blobs or automatically detected feature points.
- ii. **Text features.** The textual features can be extracted from the automatic speech recognizer (ASR) transcript, video optical character recognition (OCR), video closed caption text, and production metadata.
- iii. **Audio features.** The audio features may be generated based on the short time Fourier transform including the fast Fourier transform (FFT), mel-frequency cepstral coefficient (MFCC) together with other features such as zero crossing rate (ZCR), linear predictive coding (LPC), volume standard deviation, non-silence ratio, spectral centroid and pitch.
- iv. **Motion features.** This can be represented in the form of kinetic energy which measures the pixel variation within a shot, motion direction and magnitude histogram, optical flows and motion patterns in specific directions.

Existing surveys [Pradeep et al. 2010] in this direction are mostly focused on a particular aspect of the analysis task, such as multimodal video indexing [Chang et al. 2005], [Snoek et al. 2005], automatic audio-visual speech recognition [Potamianos et al. 2003], biometric audiovisual speech synchrony [Bredin et al. 2007], multi-sensor management for information fusion [Xiong et al. 2002], face recognition [Zhao et al. 2003], multimodal human computer interaction [Jaimes et al. 2005], [Oviatt et al. 2003], audio-visual biometric [Aleksic et al. 2006], multi-sensor fusion [Luo et al. 2002] and many others. By observing the related work, the successful techniques for multimodal combination in video retrieval have so far been late fusion, linear combinations, lexical and visual features, and query class-dependent weighting.

2.5.3.6 Semantic Based Queries Paradigm

Semantic based multimedia systems have already proliferated in the multimedia information retrieval community providing various search paradigms. There are three different semantic search paradigms that users can exploit to satisfy their information need. These search paradigms work on a high-level feature space that is obtained through different methods.

a) Keyword Based Queries

The direct persistence of keyword annotations, i.e. high-level features, permits the user to enumerate a set of keywords to search for multimedia content comprising these concepts. This is already a large step towards more semantic search engines. However, comparatively beneficial in some situations this still might be too limiting, semantic multimedia content apprehends knowledge, which goes apart from the ordinary listing of keywords. These semantic structures are the characteristics that humans lean on to manifest some information need. Natural language based queries and semantic example based queries investigate these aspects.

The developed high-level analysis algorithm equips a set of keyword that empowers multimedia information to be searched with a vocabulary of predefined keywords. The implemented search-by-keyword paradigm permits the user to submit a query in the form of keywords and produce one or more query vectors that are then used to search for the documents that are most similar to that query vector.

b) **Natural Language based Queries**

In text based information retrieval the user submits the query in the form of text by using different techniques like in the form of vectors or Simple Boolean expressions that are further used in inference networks [Croft et al. 1991]. These sorts of query expressions are now practicable in multimedia information retrieval exploiting the algorithms that can discover multimedia concepts. Recently, [Town et al. 2004b] proposed an ontology based search paradigm for visual information that allows the user to express his query as a sentence, e.g., “red flower with sky background

c) **Semantic Example based Queries**

The implemented search-by-semantic-example paradigm applies the high-level analysis on the query example to obtain the corresponding keyword probabilities. To find the documents that are most similar to the query vector we use the same strategy as for the previous case. Several examples can be provided and they are combined according to the logical expression submitted by the user. Moreover, both search-by-keyword and search-by-semantic-example can be employed concurrently to ameliorate the expressiveness of the user information requirements.

Several semantic example based query techniques has been proposed to bridge the semantic gap [Rasiwasia et al. 2007], [Rasiwasia et al. 2006]. These sorts of approaches can demonstrate good results, but it inflicts an extra overload on users who now have to describe their idea in terms of all possible instances and variations, or express it textually. Thus, in these cases users should be able to formulate a query with a semantic example of what they want to retrieve. Of course, the example is not semantic per se but the system will look at its semantic content and not only at its low-level characteristics, e.g., colour or texture. This means that the system will infer the semantics of the query example and use it to search the image dataset

2.6 **Evaluation Measure**

The standard process of scientific research is to evaluate hypotheses and research questions based on clear and justified standards. Image and video retrieval is a subclass of information retrieval and inherits therefore many of the aspects that encompasses information retrieval. Image and video retrieval is concerned with retrieving images and videos that are

relevant to the user's request from collections of images and videos. The essential aims of information retrieval are to be efficient and effective. Efficiency means delivering information quickly and without excessive demands on resources, even when there is a massive amount of information to be retrieved. Clearly efficiency is extremely relevant to information retrieval where late response is often useless information. Effectiveness is concerned with retrieving relevant documents. This implies that the user finds the information useful.

A significant landmark in the evaluation of information retrieval systems was the Cranfield experiments, in which the measurement of recall and precision was first established [Cleverdon, 1967]. Many alternatives have been proposed later, containing fallout (the proportion of returned documents out of those irrelevant), F-measure, etc. The retrieved results can be evaluated by means of the various evaluation techniques. Information retrieval systems have been evaluated for many years. Evaluation is the major part of the retrieval systems. Information science has developed many different criteria and standards for the evaluation e.g. effectiveness, efficiency, usability, satisfaction, cost benefit, coverage, time lag, presentation and user effort, etc. Among all these evaluation technique precision which is related to the specificity and recall which are related to the exhaustively are the well-accepted methods. In our approach, we use average precision as well as recall for evaluating the performance. For calculating, the precision and recall the retrieved, relevant and irrelevant as well the non-retrieved relevant as well as the relevant information must be available.

In IR system returns the two sets of documents i.e. the relevant and irrelevant. The relevant documents are the documents that belong to the category that is defined by the user while the irrelevant ones don't belong to that specific category. Figure 2.16 illustrate the categorizes of the data in the corpus and the data that will retrieved by using the retrieval systems.

Irrelevant	Retrieved & Irrelevant	Not retrieved & Irrelevant
Relevant	Retrieved & Relevant	Not Retrieved but Relevant
	Retrieved	Not Retrieved

Figure 2.16: Category of the data that will either retrieved by using the particular retrieval system and the data in the corpus that will not retrieved

2.6.1 Precision

Precision is the ratio of the number of relevant records retrieved to the total number of irrelevant and relevant records retrieved.

$$\text{Precision} = \frac{|A \cap B|}{|B|} \quad (2.8)$$

Where

A is set of relevant Images.

|A|= No of relevant Images in the dataset

B is the set of retrieved images.

|B|= No of retrieved Images.

Precision is a measure of the proportion of retrieved relevant documents. It is important in information search. Considering that users often interact with few results only, the top results in a retrieved lists are the most important ones. An alternative to evaluate these results is to measure the precision of the top-N results, P@N. P@N is the ratio between the number of

relevant documents in the first N retrieved documents and N. The P@N value focuses on the quality of the top results, with a lower consideration on the quality of the recall of the system.

2.6.2 Recall

While recall is the ratio of the number of relevant records retrieved to the total number of relevant records in the database.

$$\text{Recall} = \frac{|A \cap B|}{|A|} \quad (2.9)$$

The recall measures the proportion of relevant documents that are retrieved in response to a given query. A high recall is important especially in copyright detection tasks. Both precision and recall values are single-value metrics that consider the full list or retrieved documents. Since most retrieval systems, however, return a ranked list of documents, evaluation parameters should allow to measure the effectiveness of this ranking. One approach to combine these metrics is to plot precision versus recall in a curve. The Venn diagram for the precision recall is shown in the Figure 2.17.

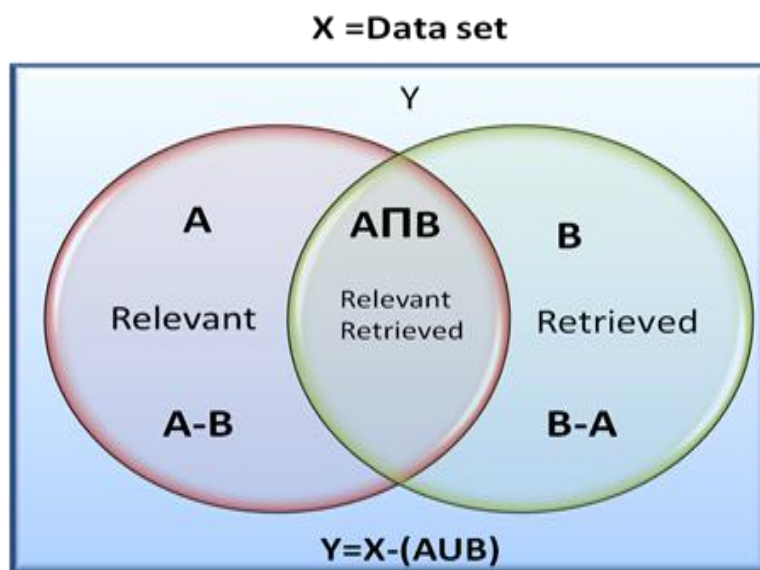


Figure 2.17: Venn diagram for Precision and Recall

Precision-recall curves are another useful way of visualizing a system's retrieval effectiveness in detail. Figure 2.18 presents the examples of three systems. These curves are obtained by plotting the evolution of the precision and recall measures along the retrieved rank [Joao. 2008]. An ideal system would achieve both 100% precision and 100% recall. Practically there is a trade-off between precision and recall.

In information retrieval, an ideal precision is 1.0, depicts that all the retrieved documents are relevant, even though if some of the relevant documents in the corpus are not retrieved. While the ideal recall is 1.0 delineates that all the relevant documents in the corpus was retrieved, even though if they contain many of the retrieved irrelevant documents as well.

Often, precision and recall are the inverse of each other, one is increasing at the cost of other. There is a trade-off between the precision and recall e.g. when an IR system increase the precision by retrieving only relevant documents and decreasing the irrelevant ones at the cost of missing some of the relevant documents in the corpus and vice versa. In IR context, precision and recall are described in terms of a set of retrieved documents.

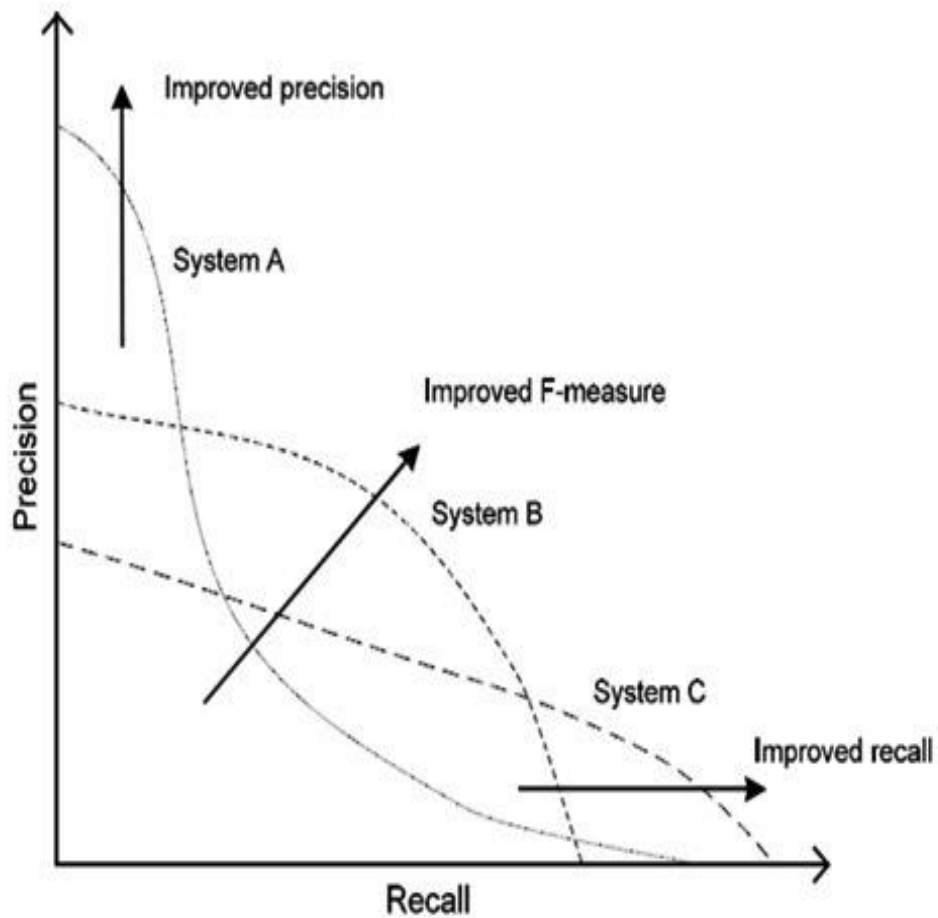


Figure 2.18: Interpretation of precision-recall curves

2.6.3 F-Measure

A measure that combines precision and recall is the harmonic mean of precision and recall, the traditional F-measure or balanced F-score [Rijsbergen, 1979]. The F-measure is the harmonic mean of precision and recall and is computed as

$$\text{F-Measure} = 2 \cdot \frac{\text{Pre} \cdot \text{Rec}}{\text{Pre} + \text{Rec}} \quad (2.10)$$

In statistics, the F1 Score (also F-score or F-measure) is a measure of a test's accuracy. The F1 score can be interpreted as a weighted average of the precision and recall, where an F1 score reaches its best value at 1 and worst score at 0. Since all these measures are used to investigate the efficiency of the IR system.

2.7 Chapter Summary

In this chapter, we surveyed the several different principles that are used in the image and video retrieval. We first discussed the general structure of information retrieval, multimedia retrieval, followed by the general overview of the text retrieval. The detailed discussion about the image and video retrieval and the various techniques that are used for the image and video retrieval and their pros and cons. We have survey the retrieval techniques in terms of the low level i.e. Content based Retrieval techniques and high level analysis i.e. Semantic based Retrieval. We have also explored various evaluation techniques used for investigation of the information retrieval systems. To achieve a more comprehensive understanding of the field, we concluded a thorough research of previous work. We have done a detailed survey of all these techniques and concluded that semantic based retrieval outperforms the content based retrieval techniques. Keeping this in mind we have further contributed in the Semantic based retrieval by proposing three main contributions which are discussed in the fort coming chapters. The detailed discussion of the first contribution the Semantic query interpreter will be found in the next chapter 3.

Chapter 03

Semantic Query Interpreter for Image Search & Retrieval

“There is no problem so complicated that you can’t find a very simple answer to it if you look at it right. Or put it another way: The future of computer power is pure simplicity.”

The Salmon of Doubt by Douglas Noel Adams

Due to the ubiquitousness of the digital media including broadcast news, documentary videos, meeting, movies, etc. and the progression in the technology and the decreasing outlay of the storage media leads to an increase in the data production. This explosive proliferation of the digital media without appropriate management mimics its exploitation. Presently, the multimedia search and retrieval are an active research dilemma among the academia and the industry. The online data repositories like Google, YouTube, Flickr, etc. provides a gigantic bulk of information but finding and accessing the data of interest becomes difficult. Due to this explosive proliferation, there is a strong urge for the system that can efficiently and effectively interpret the user demand for searching and retrieving the relevant information. The effective search and retrieval of multimedia contents subsists one of the major issues among researchers. There exists a gap between the user query and the results obtain from the system, so called semantic gap. The keyword based system mostly rely on the syntactic pattern matching, and the similarity can be judge by using the string matching not concept matching. The focus of this chapter is to bridge the semantic gap. A major part of the retrieval systems is the query interpreter. In order to cope with these problems, we are proposing a novel technique for automatic query interpretation known as the Semantic Query Interpreter (SQI). In SQI, the query is expanded dimensionally by means of lexically by using the open source knowledgebase i.e. WordNet and semantically by using the common sense knowledgebase i.e. ConceptNet in order to retrieve more relevant results. We evaluate the effectiveness of SQI using the open benchmark image dataset, the LabelMe. We use two of the eminent performance evaluation method in Information Retrieval (IR) i.e. the Precision and Recall. Experimental results manifest that SQI shows substantial rectification over the traditional ones.

The remainder of the chapter is as follows. Section 3.1 discusses the introduction about the chapter. Section 3.2 reviews the existing techniques or state of the art in the field of query expansion and analysis along their pros and cons. Section 3.3 includes the new proposed framework. Section 3.4 contains how to evaluate the performance of the proposed algorithm and the experimentation setup i.e. to evaluate and compare the proposed technique with existing ones. Finally, in section 3.5 we summarize the chapter.

3.1 Introduction

With the rapid evolution in the digital technologies, has steered to the stunning amount of the image data. The unprecedentedly high production of multimedia data, boosts the expectation that it can be as easily manage as text. Keeping this, it is impracticable for the user to manually search the relevant information. Several search engines have been developed to overwhelm this. Researcher community is continuously exploiting the techniques for effectively and efficiently managing these data. However, the dilemma is yet not figured out completely. The majority of the retrieval systems work efficient for the simple queries. Sometimes the relevant data is available in the corpus, but it cannot be annotated with the particular word in the query. This problem is known as the vocabulary gap.

The diversity of the human perception and vocabulary difference is the main stumbling block in the performance of the information retrieval system. According to Bates [Bate. 1986], "*the probability of two persons using the same term in describing the same thing is less than 20%*", and Furns et al. found that "*the probability of two subjects picking the same term for a given entity ranged from 7% to 18%*" [Furn et al. 1987]. Single concept can be verbalized into different words. All these inconveniences steered to the substantial magnitude of data retrieval but among the entire retrieved outcome, merely some of them are relevant. However, until now, cardinal challenge is taking the user demand, interpreting it precisely for finding the data of the user's interest. Several attempts have been made for retrieving the relevant images, but still it's frustrating. Thus, there is a great need for the system that can find the relevant information from this over whelming archive.

Searching and Retrieving the textual information is not a laborious task. While in case of audio, images and video data it is not a facile task because as it is said that *A picture is worth a thousand words*¹. It is usually recognized that the current state of the art image analysis techniques is not proficient at apprehending all implications that an image may have. To elevate the retrieval process, it is beneficial to integrate the image features with other sources of knowledge. Low-level visual features exclusively do not comprehend the concepts of an image, and text based retrieval (annotation) by themselves are extensively directly associated to the high-level semantics of an image. Integration of visual features and annotations can supplement each other to cater results that are more precise.

¹ A Confucius Chinese proverb "The Home Book of Proverbs, Maxims, and Familiar Phrases

The possibility for searching the images by utilizing the keywords is a remarkably wholesome and restrains numerous unnecessary images from the retrieved results. It is not backbreaking to believe that manually annotating monumental aggregations of images is an extremely monotonous and subjective task, and there is a risk of inconsistent word assignment or vocabulary difference, unless a fixed set of terms is used. Therefore, more and more research is directed at reducing this vocabulary. Although, the performance of the current techniques is not adequate yet, the results are promising and the quality of the text-based results is likely to improve over the next years.

To date the semantic based image retrieval problem is not yet solved completely. However, the textual information retrieval is full-blown. The image retrieval mostly rely on the textual information i.e. metadata and the contents of the image i.e. Content Based Image Retrieval (CBIR). The textual descriptor also called the annotation though cannot capture the overall semantic content of the image. Sometimes the textual description allied with the image could be ambiguous or inaccurate in describing the image semantic and some irrelevant images come out as a result of the user query. The conventional Content Based Image Retrieval techniques are still exploiting the image based on low- level features like colour, shape, texture, etc. These techniques do not exemplify the noteworthy efficiency. CBIR techniques are interpreting an image analogous to a computer. They are rendering an image just as the composite of pixels that are characterized by colour, shape and texture. However, for the user, the image is the combination of objects instead of pixels, delineating some concepts. For them, it does not only refer to the content of the image that is appearing, but rather the semantic idea it is exemplifying.

It is worth saying that for the same image can be interpreted differently by different people. The content based system will be beneficial for simple queries but it will comes to grief for the complex queries like query on the basis of scene, event or some complex concepts etc. These let the user to hunt for the images by the queries like “show me the images of Bush”, “show me the images of the car” etc. but these types of systems will flunk in extracting the high-level semantics. These are the reasons that lead to the poor retrieval performance. This bleak situation leads to the need for the system that can extract the semantic from the image or video that can elaborate the semantic concept in to the object level i.e. what objects will constitute to create the following scene, event or situation. Owing to the flexible nature of the human and the hard coded computer, nature there appears a problem known as the semantic gap. It is due to the

difference between the user interpretation and the machine understanding. Semantic gap can be defined as the “The lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation” [Smeulders et al. 2000]. Bridging the semantic gap has been declared key problem Information Retrieval (IR) systems since a decade. The efficiency of the retrieval system relies on the ability of the system to comprehend the high-level features or semantics.

The success of the retrieval system depends on the number of relevant documents it retrieves. Higher the number of relevant document it retrieves higher will be its precision and efficiency. One of the major challenges in the search and retrieval system is the difficulty of interpreting the user requirement correctly or describing the demands precisely in the form of query so that the system can process it accurately.

Sometimes the word in the query does not match the words in the corpus even though they contain the same concept or information. However, the semantic of both the concepts is same but the vocabulary may be different. This vocabulary difference problem is known as the word mismatch or the Vocabulary gap, which causes the efficiency of the retrieval performance low. It is the “*lack of coincidence between the word used in the query to retrieve the particular document and the word used to annotate the particular concept*”. For example, someone might be interested in the images of the "rock" and some of the images of the "rock" are annotated with the word "stone" in the corpus. One concept may be expressed by different words i.e. vehicle may be expressed in auto or automobile, etc. Word mismatch is among one of the causes of retrieval system failure. The word mismatch problem, if not appropriately addressed, would degrade retrieval performance critically of an information retrieval system.

Secondly, due to the difference in the user's background, experiences and perception regarding the same image. It is impossible for the machine to completely cope with the flexible nature of human. When the user enters the query it is more likely that the system is not capable of wholly comprehend what the user wants or sometimes the user cannot express its requirements properly. The success of the retrieval system much more relying on the ability of the system to understand the query and then find the appropriate data in relation to the query specification.

As an attempt to rectifying these stated problems, Query expansion has been gaining more and more importance from the recent years [Jin et al. 2003]. Query expansion is a

technique of expanding the query by adding some additional terms to the query that are closely related to the query terms. From the last few decades, different of the query expansion techniques have been proposed by different researchers from the manually constructed thesauri to the open source knowledgebase. Query expansion with domain specific knowledge sources has substantially improved the retrieval performance [Stokes et al. 2009]. All these methods manifest the efficacious performance.

Different query expansion techniques have been continuously investigating by the researcher since decades, but still some of the issues are remain at their infancy. In this thesis, we are proposing automatic Semantic Query Expansion techniques, where the query after the preprocessing will be expanded lexically by using the open source lexical knowledgebase i.e. WordNet [Fellbaum et al. 1998] and then expanded conceptually by using the largest conceptual open source knowledgebase i.e. ConceptNet [Liu et al. 2004a]. These knowledgebase attaches the list of related words with the query that will make the query more flexible or increase the recall but will simultaneously decrease the precision. For achieving the precision among the list, some of the concepts will be prune by using the candidate concepts selection module. That will use the semantic similarity technique of WordNet. The final list of expanded concepts will be applied on the open source benchmark LabelMe dataset. The results are retrieved and ranked by using the vector space model. The experimental results demonstrate that our method achieved significant improvement in terms of precision and recall over the existing. This scheme seems effective for interpreting the user requirement semantically in terms of both theoretical as well as experimental analysis. It can provide the semantic level query expansion.

3.2 State- of-the-Art

The performance of the information retrieval system is highly affected by the query engine. Most of the data available are unstructured that is only understandable by the human. One of most important factor is to get what the user needs. Word mismatch is one of the most commonly occurring problems in IR. Word mismatch occur when the user particular concept is annotated with the different vocabulary and the user uses different. This leads to serendipitous results [Nekrestyanov et al. 2002]. Various techniques and approaches has been proposed in order to cope with this problem query expansion is among one of them. Query expansion is used to remove the vocabulary mismatch problem or to reduce the vocabulary gap [Poikonen et al. 2009].

Clearly, such vocabulary gaps make the retrieval performance non-optimal. Query expansion [Voorhees, 1994] [Mandala et al. 1999] [Fang et al. 2006] [Qiu et al. 1993] [Bai et al. 2005] [Cao et al. 2005] is a commonly used strategy to bridge the vocabulary gaps by expanding original queries with related terms. Expanded terms are often selected from either co-occurrence-based thesauri [Qiu et al. 1993] [Bai et al. 2005] [Jing et al. 1994] [Peat et al. 1991] [Fang et al. 2006] or handcrafted thesauri [Voorhees. 1994] [Liu et al. 2004a] or both [Cao et al. 2005] [Mandala et al. 1999].

Query expansion is a promising approach to ameliorate the retrieval performance by adding some additional terms to the query that are closely related. A search for e.g. “automobile” should also return results for its synonym “vehicle”. Therefore, the aim is to expand queries with their synonyms and other related words in order to receive results that are more relevant. For instance, a related word to the query “crocodile” might be “alligator”. To find such query expansion terms several techniques have been developed in the recent years, of which some of the literature will be discussed here.

The idea of query expansion has been exploited for decades but still it is worth investigating. The goal of the query expansion is to improve either the precision or the recall and to increase the quality of the search engines. Query expansion is an effective technique in information retrieval to improve the retrieval performance, because it often can bridge the vocabulary gaps between queries and documents. Another way to improve retrieval performance using WordNet is to disambiguate word senses. Voorhees [Voorhees, 1993] showed that using WordNet for word sense disambiguation degrades the retrieval performance. Liu et al. [Liu et al. 2004a] used WordNet for both sense disambiguation, query expansion, and achieved reasonable performance improvement. However, the computational cost is high and the benefit of query expansion using only WordNet is unclear.

Query expansion can be classified as manual query expansion and automatic query expansion. The manual query expansion involves much user intervention. The user is implicated in the process of selecting the supplementary terms [Ekmekcioglu et al 1992], [Beaulieu et al. 1992] [Wade et al. 1988]. However, manual query expansion reliance profoundly on the user and experiments using this technique do not result in considerable enhancement in the retrieval effectiveness [Ekmekcioglu et al 1992]. While the automatic query expansion can be done

automatic without much user intervention. The automatic query expansion outperforms the manual query expansion and makes the information retrieval process facile and efficient.

The automatic query expansion is more efficient than the interactive query expansion. One of the approaches used to increase the uptake of the interactive query expansion is by displaying the summaries of the overviews [Gooda et al. 2010]. Therefore, in this thesis we focus on automatic query expansion particularly on expansion using knowledgebase. Query expansion can be categorized as probabilistic query expansion and expansion by using ontologies.

3.2.1 Probabilistic Query Expansion

Probabilistic query expansion generally based on calculating co-occurrences of terms in documents and selecting the most related to the query. Several probabilistic query expansion methods have been proposed relevant feedback [Ponte et al. 1998] [Miller et al. 1990], Local co-occurrence method [Jin et al. 2003], [Rocchio 1971] and Latent Semantic Indexing (LSI_based) [Hong-Zhao et al. 2002], [Deerwester et al. 1990] have been proposed. Most probabilistic methods can be categorized as global or local.

Global techniques extract their co-occurrence statistics from the whole document collection and might be resource intensive as the calculations can be performed off line.

Local techniques extract their statistics from the top-n documents returned by an initial query and might use some corpus wide statistics such as the inverse document frequency. The calculation for the local probabilistic query expansion is done on-line One of the first successful global analysis techniques was term clustering [Jones 1971]. Other global techniques include Latent Semantic Indexing [Deerwester et al. 1990], and Phrase finder [Jing et al. 1994].

These techniques utilize different approaches to build a similarity matrix of terms and select terms that are most related to the query terms. Local techniques assumed that the top-n documents are relevant to the query. This assumption is called pseudo-relevance feedback and has verified to be a modest. In pseudo-relevance feedback, the decision is made without the user intervention. However, it can cause a considerable discrepancy in performance relying on whether the documents retrieved by the initial query were indeed relevant. The method of relevant feedback altered the query terms according to the distribution of the terms in the relevant and irrelevant documents that are retrieved in response to the query. This method is a prevailing technique and can ameliorate the retrieval result in most cases [Ponte et al. 1998],

[Miller et al. 1990]. Conversely, this method is relying on the first-retrieved top-relevant documents. If the first result is not worthy, relevant feedback will culminate in even worse results. Most local analysis methods use the notion of Rocchio's [Rocchio, 1971] ideal query as a start point.

A number of approaches have been proposed, which vary on how they select the terms from the top-n documents and their endeavours to reduce the influence of irrelevant documents returned by the initial query [Mitra et al. 1998] [Lu et al., 1997] [Buckley et al., 1998]. Local Contest Analysis is nevertheless, the most flourishing local analysis method [Xu et al, 2000]. Local co-occurrence is a probabilistic method based on the co-occurrence frequency of the words in the training corpora. Local co-occurrence method has shown the substantial results for the IR system [Ponte et al. 1998], [Milne et al. 2007], but it collapses with meaning clustering.

Latent Semantic Indexing (LSI) is a powerful method, which can be implemented by two kinds of algorithms, i.e. singular value decomposition [Zhao et al. 2002] and probabilistic LSI [Deerwester et al. 1990]. The method builds a semantic space, map each term into this space and cluster automatically according to the meaning of terms. However, it is difficult to control the query expansion degree and the modified queries may contain many irrelevant terms, which can be seen as noise.

3.2.2 Ontological Query Expansion

Ontological methods suggest an alternative approach, which uses semantic relations drawn from the ontology to select terms. Ontology based query expansion have been studied for a long time [Jin et al. 2003] [Mandala et al, 1999]. By using this approach, query expansion is done semantically and users are able to have a faster access to their required information. For this purpose, Fu, Navigli and Andreou have been presented various methods and algorithms [Fu et al. 2005], [Andreou et al. 2008]. The leading precedence of the probabilistic methods is that the association between the expanded terms and the original query terms are readily generated from the corpus. However, there are a significant number of manually edited large repositories of relations between concepts stored in ontologies and using those data for query expansion is covered in the literature. Most approaches use large lexical ontologies usually WordNet, ConceptNet or Cyc because they are not domain specific and because their relations are not sparse. In our thesis we use the ontological query expansion by the integration of the lexical

knowledgebase i.e. WordNet and conceptual knowledgebase i.e. ConceptNet. A brief overview of both the knowledgebases will be discussed in the section below 3.3.

One of the previous works is Ontology-based query expansion. Ontology is a resource, which provides the relation information between two concepts. The relation types include coordination, synonyms, hyponym and other semantic relation. Some of the former works [Jin et al. 2003] [Zhang et al. 2002] show that WordNet can be used as ontology in query expansion, but it is strongly depending on the characteristics of queries. Even For some queries, the more expansion will be resulted in a worse performance.

Mihalcea and Moldovan [Mihalcea et al. 1999] and Lytinen et al. [Lytinen et al. 2000] used Word-Net [Miller et al. 1990] to obtain the sense of a word. In contrast, Schutze and Pedersen [Schutze et al. 1995] and Lin [Lin et al. 1998] used a corpus-based approach where they automatically constructed a thesaurus based on contextual information. The results obtained by Schutze and Pedersen and by Lytinen et al. are encouraging. However, experimental results reported in [Gonzalo et al. 1998] indicate that the improvement in IR performance due to WSD is restricted to short queries, and that IR performance is very sensitive to disambiguation errors. Harabagiu et al. [Harabagiu et al. 2001] offered a different form of query expansion, where they used WordNet to propose synonyms for the words in a query, and applied heuristics to select which words to paraphrase. Ingrid et al. uses WordNet for the query expansion for obtaining the semantic information. They expand the user query by using the WordNet and then query reduction for removing those terms that can distract the query result. They performed experiments on the TREC8, 9 and 10 queries. Moreover, concluded that their approach enhanced the average number of correct documents retrieved by 21.7% and average successfully processed query enhanced by 15% [Ingrid et al. 2003]. Hsu expands the user query by integrating the ConceptNet as well as the WordNet and selecting the candidate terms by using the spreading activation technique [Hsu et al. 2008].

A thesaurus is defined as a dictionary of synonyms or a data structure that defines the semantic relatedness between words [Schutze et al. 1997]. Thesauri are used to expand the seed query to improve the retrieval performance [Fang et al. 2001]. Thesauri are also known as semantic networks. There are two types of thesauri hand crafted thesauri and automatically generated thesauri. The hand crafted thesauri is developed manually by peoples is the hierarchal form of the related concepts. While the automatically constructed thesauri is the dictionary of the

related terms that are derived from the lexical or semantic relationship among them. The thesauri may be the general purpose or the domain specific. Query expansion by using thesauri and automatic relevance feedback shows an effective improvement for web retrieval [Jian et al. 2005]. Query Expansion by using the knowledgebases has gain a considerable researcher attention from the last few decades.

The semantic networks like WordNet are able to attach the synsets to each word in the query. It contain the words their definitions along with relationships. WordNet [Fellbaum et al. 1998], Cyc [Lenat. 1995] and ConceptNet [Liu et al. 2004a] are considered the widest commonsense knowledgebase currently in use. WordNet has been used as tool for query expansion and various experiments have been performed on the TREC collection. The query terms are expanded by using the synonyms, antonyms, hypernyms and hyponyms. The Wikipedia and WordNet is used for query reformulation and to extract a ranked list of related concepts [Adrian et al. 2009] The performance of the Wikipedia is improved by bringing together WordNet hierarchical knowledgebase with the Wikipedia classification [Simone et al. 2009]. The ambiguity of the Wikipedia is removed by using the WordNet synsets. The following are the list of projects that uses the WordNet knowledgebase that are SumoOntology, DB pedia, Open CYC, Euro WordNet, eXtended WordNet, Multi WordNet, Image Net, Bio WordNet, Wiki Tax 2 WordNet [Wiki WN]. The experiments show that the retrieval performance has improved a lot for the short queries but for the complex and Long queries the results have not been very successful [Voorhees. 1994].

Semantic query expansion is still an exigent issue. Early work mostly relies on the text matching techniques. However, subsequently the trend was moved to the semantic expansion of the user queries. Those systems heavily rely on lexical analysis by using lexical knowledgebase such as one of the one of the largest open source lexical knowledgebase i.e. WordNet. However, the WordNet is suitable for the single keyword based query but it flunks in the case of the complex queries like the multi concept queries. It does not find the semantic relatedness or have no potential for the common sense reasoning.

Despite of the fact, that lexical analysis plays an imperative role in the extracting the meaning from the user request, the common sense reasoning also plays a focal role. Common sense knowledge includes knowledge about the spatial, physical, social, temporal and psychological aspects of everyday life. WordNet has been used mostly for the query expansion.

WordNet has been used usually for the query expansion. It has made some rectification but was limited. The query expanded by using the WordNet shows better performance than the query without using it. It will increase the recall but the precision of such type of queries is not so optimal.

The common sense knowledge represents the deeper analysis of the word or a concept. For achieving, the IR accuracy there is a need for the system to understand and interpret the user request fully by using the commonsense, which is not present in the computer only human possess. Computer is superior to a human in the computational task but weak in the common sense reasoning. Several studies reveals the importance of common sense reasoning in information retrieval, data mining, data filtering etc. [Lieberman et al. 2004].

Query expansion has been applied by various researchers in different domain like in health information i.e. Electronic health records [Keselman et al. 2008], [Hersh. 2009], genomic information retrieval, geographic information retrieval and for various languages [Mojgan et al. 2009]. The PubMedTM search engine uses the automatic query expansion technique known as the automatic term mapping [Lu et al. 2009], [Yeganova et al. 2009]. They applied their expansion technique on TREC Genomics data in both 2006 and 2007 and showed the effectiveness of the query expansion. The flexible query expansion technique raises the retrieval performance of the genomic information retrieval systems [Xiangming et al. 2010]. Queries are expanded by means of the authoritative tags to refine the user query and all these tags are stored in the users profile for future use [Pasquale et al. 2010]. The proposed SQI integrate both the lexical as well as the commonsense knowledge for achieving the accuracy.

3.3 Proposed Framework

As already discussed, the query expansion is one of the ways to increase the efficiency of the information retrieval systems. In our research we explore a semantic approach for expanding the query both lexically and semantically by using knowledgebases and selecting the candidate terms for expansion, then retrieve, and rank the data by using one of the well-known retrieval model the vector space model. We use the semantic similarity function to make comparison between the expanded terms. For extracting the semantics from the user query, we anchor the intended senses or concepts with the pertinent query terms. The overall Semantic Query Interpreter is shown in the Figure 3.1.

If the user enters a keyword based query or an object based query e.g. car he can only get the images which are indexed by a keyword car. We use the WordNet as well as the ConceptNet to expand the query. With the WordNet we expand the query by taking the synonyms i.e. synsets along with their semantic similarity. The original query may be expanded to include auto, automobile, machine, motorcar, railcar, railway car, railroad car, cable car, gondola, elevator car etc. After the Lexical analysis then the semantic knowledge can be applied by means of the ConceptNet knowledgebase. The ConceptNet expand the query by adding following concepts along with their semantic similarity e.g. bed, brake, car, day, drive, front part, good appearance, hood, its head, lane, light, long distance, motorcycle, mountain, other person, person's mother, plane, pollutant, right behing, road trip etc. All these expanded terms raised the system recall but simultaneously decrease the system's precision. Because some expanded concepts are relevant while some of them are irrelevant i.e. the noise. In order to maintain the precision we have to remove these noises. These noises can be removed by means of the candidate concept selection algorithm. The synset and the concepts are retrieved along with their semantic similarity as shown in the figures. The semantic query interpreter contains the following components.

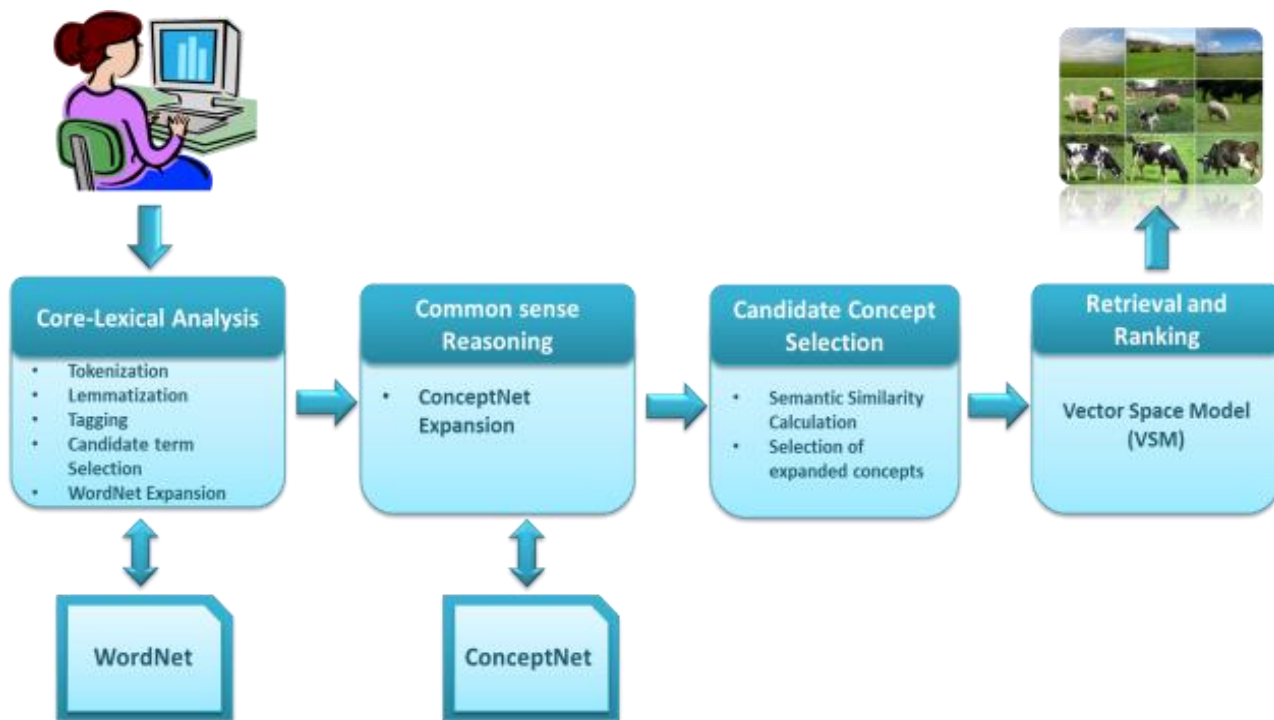


Figure 3.1: Overall Semantic Query Interpreter

- i. Core Lexical Analysis
- ii. Common Sense Reasoning
- iii. Candidate Concept Selection
- iv. Retrieval and Ranking of Results

The detailed description of each of the component is described below.

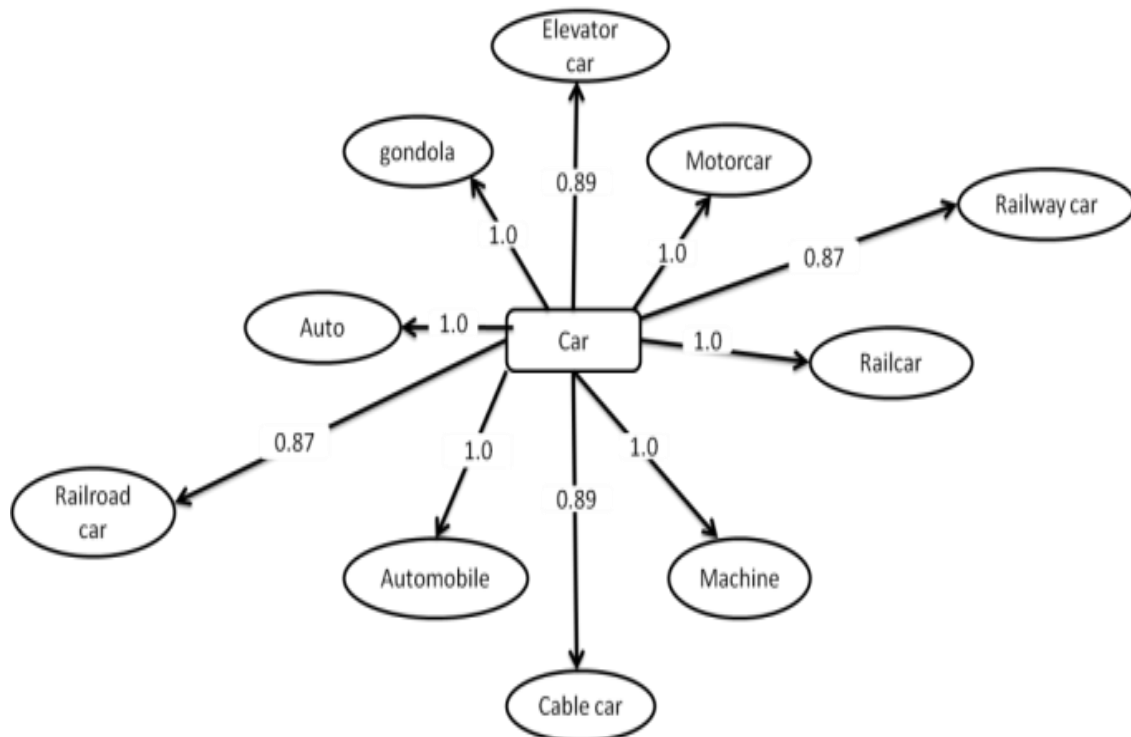


Figure 3.2: Query Expansion along with semantic similarity by WordNet. The WordNet attaches the synset of the cars like motor car, railway car, railcar, machine, cable car, automobile, railroad car, auto, gondola, elevator car etc. The figure contains the lexical expansion along with the semantic similarity value. As we know motor car relates more with the car that's why its Semantic similarity value is 1. The greater the semantic similarity value greater will be the relevancy degree.

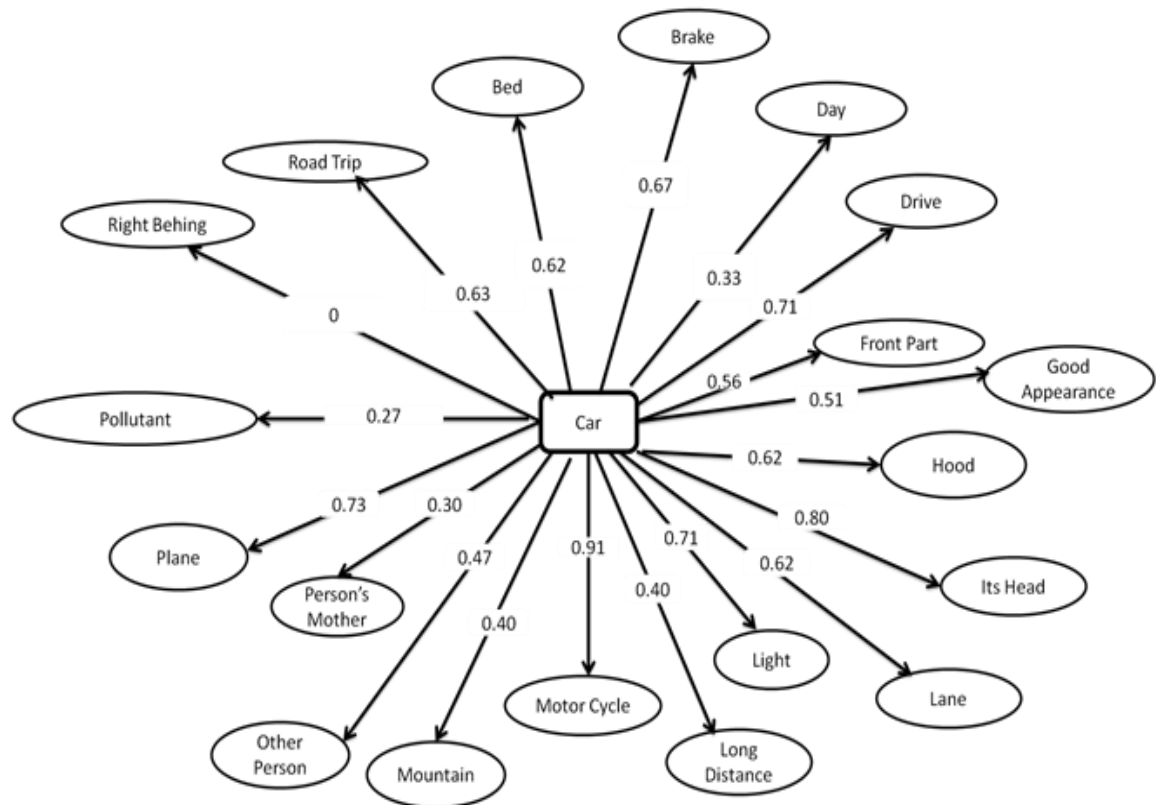


Figure 3.3: Query Expansion along with semantic similarity by ConceptNet. The ConceptNet attaches the following concepts with the keyword car like brake, day, drive, front part, good appearance, hood, its head, lane, light, long distance, motor cycle, mountain, other person, person's mother, plane, pollutant, right behing, road trip, bed etc. The figure also contains the conceptual expansion of the car along with the Semantic similarity value. Greater the Semantic similarity value greater will be the relevancy degree. Among the expanded terms some of them are noises that will significantly decrease the precision of the system.

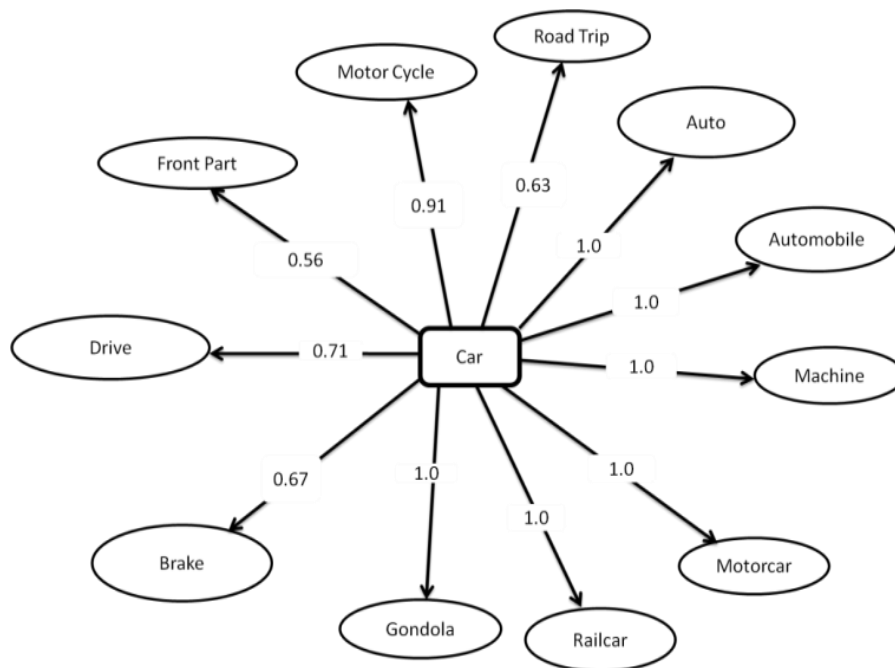


Figure 3.4: Query Expansion along with semantic similarity by Semantic Query Interpreter. The Semantic Query Interpreter expansion contains the selected lexical and conceptual expansion of the keyword car. The figure contains the selected expansion terms according to the threshold, and the semantic similarity value between the original query term and the expanded terms.

3.3.1 Core Lexical Analysis

The user's query is the significant part in the information retrieval system. However, it will not always contain the sufficient words to accurately explain the user requirement or sometimes they cannot express query request in the proper form. Core Lexical analysis converts a stream of the words of the query into a stream of concepts that can be used for expansion. Not every word counts the same indeed in the query. Thus, one of the prime intentions of the lexical analysis phase is the recognition of pertinent words in the query. The query can be expressed as the combination of events, concepts and objects.

The core Lexical analysis is the key element of the Semantic Query Interpreter. The core lexical analysis contains the pre-processing module and the lexical expansion module.

3.3.1.1 Pre-processing

The pre-processing includes the basic natural language processing (NLP) functions. The pre-processing module consist of the following main steps

- i. Tokenization
- ii. Lemmatization
- iii. Part-of-Speech tagging

Tokenization

Tokenization is the process of separating and perhaps categorizing sections of a string of input characters. Tokenization is the process of truncating a stream of text up into words, phrases, symbols, or other meaningful elements called tokens. The lists of tokens are then passed for further processing and lexical analysis. In languages such as English (and most programming languages) where words are delimited by white space (space, enter, and tab characters). White space characters, such as a space or line break, or by punctuation characters, separate tokens.

Lemmatizer

Lemmatizer is one of the module of Montylingua [Covington, et al. 2007], is an automatic NLP tool that first tags input data with a tagger that the creator [Hugo Liu, 2004] claims exceeds the accuracy of the Transformation-based Part of Speech Tagger. The lemmatizer strips the suffixes from plurals and verbs and returns the root form of the verb or noun. Lemmatization is the procedure of deciding the lemma for a given word. So various inflected forms of a word can be investigated as a single item. It does a similar task with stemming but answer the dictionary form of a word and save the part of speech information for us and convert the diverse morphological form to the base form. We run the Lemmatization instead of Stemming on the datasets.

Some examples of the lemmatization output,

- Walks, walk, walking, walked → walk.
- striking → striking
- loves, loved → love
- are, am, is → be
- best, better → good

Part-of-Speech-Tagging

Part-of-speech tagging or grammatical tagging or word-category disambiguation is the process of characterizing up the words in a text (corpus) as corresponding to a specific part of

speech, according to its definition and its context, i.e. relationship with adjacent and related words in a phrase, sentence, or paragraph. Part of speech tagging is depending on the meaning of the word and the relationship with adjacent words. There are seven parts of speech for English, i.e. noun, verb, adjectives, pronoun, adverb, preposition, conjunction, interjection. For computational intentions however, each of these major word classes is ordinarily subdivided to manifest further granular syntactical and morphological structure.

A POS categorizes the words in the sentences based on its lexical category. POS tagging is conventionally performed by rule-based, probabilistic, neural network or hybrid systems. For languages like English or French, hybrid taggers have been able to achieve success percentages above 98% [Schulze et al.1994].

Montylingua [Montylingua] is a natural language processing engine primarily developed by Hugo Liu in MIT Media Labs using the Python programming language, which is entitled as “an end-to-end natural language processor with common sense ” [Liu et al. 2004a]. It is a complete suite of several tools applicable to all English text processing, from raw text to the extraction of semantic meanings and summary generation. Commonsense is incorporated into MontyLingua's part-of-speech (POS) tagger, Monty Tagger, as contextual rules.

MontyTagger was initially released as a tagger like the Brill tagger [Brill. 1995]. Later on, the MontiLingua complete end-to-end system was proposed by Hugo Liu [Liu et al. 2004a]. A Java version of MontyLingua, built using Jython, had also been released. MontyLingua is also an integral part of ConceptNet [Liu et al. 2004a], presently the largest commonsense knowledgebase [Hsu et al. 2006], as a text processor and understander, as well as forming an application programming interface (API) to ConceptNet. MontyLingua consists of six components: MontyTokenizer, MontyTagger, MontyLemmatiser, MontyREChunker, MontyExtractor, and MontyNLGenerator. MontyTokenizer, which is sensitive to common abbreviations, separates the input English text into constituent words and punctuations. MontyTagger is a Penn Treebank Tag Set [Marcus et al., 1993] part-of-speech (POS) tagger based on Brill tagger [Brill. 1995] and enriched with commonsense in the form of contextual rules.

3.3.1.2 Candidate Term Selection

Candidate term selection module refers to the process of eliminating the terms that occur in the user query but do not contribute a lot in interpreting the semantics from the query. Some of the words in the query have the grammatical significance but do not supplement in discriminating the relevant and irrelevant results. For example, some of the frequently occurring terms like the, is, a etc. is the part of the user query. If these terms will be passed to the expansion module, they would not have any significant result on the precision of the output. Rather it will create the noises and make the increase the number of irrelevant terms for the expansion. Articles, prepositions, and conjunctions will be purged from the user query prior to the query expansion. From the list of the tagged terms, only some of the terms will be selected. Mostly, nouns can be used to extract the concepts from the image, e.g. car, sky, people, etc. nouns are the entities but always entities alone cannot define the overall query. From the list of the tagged words, the nouns (entities), verbs (events) and adjectives (properties) are selected. The selected candidate terms are then passed to the lexical expansion module for appropriate lexical and conceptual expansion.

3.3.1.3 Lexical Expansion Module

The lexical expansion module comprises of the technique for expanding the selected terms lexically by using the one of the largest English language thesaurus i.e. WordNet.

WordNet

WordNet [Carneiro et al. 2005] is an electronic thesaurus that models the lexical knowledge of English language. The facial feature of WordNet is that it arranges the lexical information in relations of word meanings instead of word forms. Particularly, in WordNet words with the same meaning are grouped into a “synset” (synonymous set), which is a matchless representation of that meaning. Consequently, there exists a many-to-many relation between words and synsets: some words have several different meanings (a phenomena known as polysemy in Natural Language Processing), and some meanings can be expressed by several different words (known as synonymy). In WordNet, a variety of semantic relations is defined between word meanings, represented as pointers between synsets.

WordNet is separated into sections of five syntactical categories: nouns, verbs, adjectives, adverbs, and function words. In our work, only the noun category is explored due to the following two reasons (1) nouns are much more heavily used to describe images than other classes of words, and (2) the mapping between nouns and their meanings, as well as the semantic relations between nominal meanings are so complicated that the assistance from thesaurus becomes indispensable. WordNet [Miller GA. 1990] contains approximately 57,000 nouns organized into some 48,800 synsets. It is a lexical inheritance system in the sense that specific concepts (synsets) are defined based on generic ones by inheriting properties from them. In this way, synsets establish hierarchical structures, which drive from generic synsets at higher layers to specific ones at lower layers. The relation between a generic synset and a specific one is called Hypernym/Hyponym (or IS-A relation) in WordNet. For example, conifer is a hyponym of tree, while tree is a hypernym of conifer. Instead of having a single hierarchy, WordNet selects a set of generic synsets, such as {food}, {animal}, {substance}, and treats each of them as the root of a separate hierarchy. All the rest synsets are assigned into one of the hierarchies starting with these generic synsets. Besides the Hypernym/Hyponym relation, there are some other semantic relations such as Meronym/Holonym (MEMBER-OF), and Antonym. Some synsets and the relations between them are exemplified in Figure 3.5.

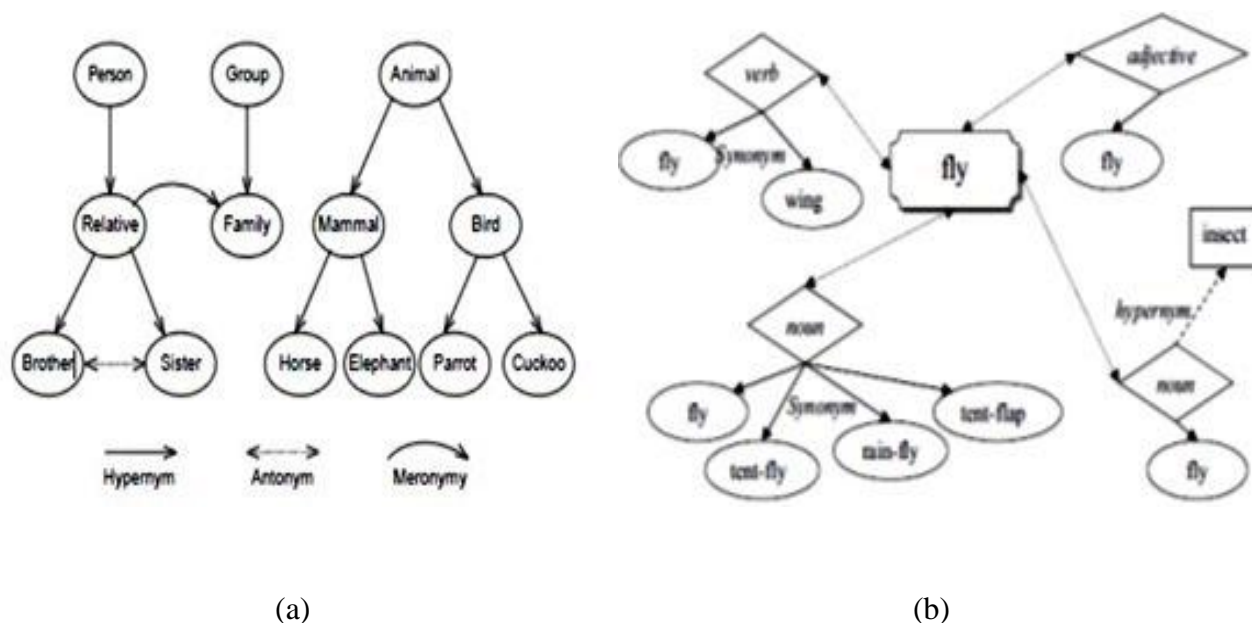


Figure 3.5: Example of synsets and semantic relations in WordNet

Words are arranged semantically and not alphabetically unlike most dictionaries. The potential benefit that WordNet has over other dictionaries is the assembling, which has been applied to each word. Words are harmonized together to form synsets (synonym sets), which represent a single sense. In this thesis, we used WordNet to remove the problem of Word Sense Disambiguation and Vocabulary gap.

The core lexical analysis takes the use query as a string or a keyword. The user query may be a single word single concept, single word multi-concept or sometimes multi-word multi-concept.

$$Q = \{K_1, K_2, \dots, K_t\} = \bigcup_{i=1}^t K_i \quad (3.1)$$

Where K_i is the set of token of the given user query.

The output of the core lexical analysis is list of expanded terms of the user query. These expanded terms are the list of synonyms.

$$Q' = \bigcup_{i=1}^{i=t'} K'_i \quad (3.2)$$

Where Q' the list of is refine concepts and their synonyms and K' is the sub concepts of the keywords i.e the expanded list of concepts.

The overall core lexical analysis algorithm is shown below.

Proposed Algorithm 3.1: Core Lexical Analysis

Input: $Q \rightarrow \cup_{i=1}^t K_i$ // List of original query terms

Output: $Q' \rightarrow \cup_{i=1}^{t'} K'_i$ // List of expanded query terms

Method:

Procedure #1: Drop some of the common words

Procedure#2: set of rules for selecting some of the terms from the list of tagged words for finding the synonyms

$LE \leftarrow$ Lemmatization (Q)

$LBT \leftarrow$ Montylingua POS(LE)

$S \leftarrow \{ 'ADV', 'NNP', 'VPZ' \}$

$Q' \leftarrow$ candidate Terms (LBT, S)

$i \leftarrow$ Length (Q')

$Q'(i). \text{Synset} \leftarrow$ WordNet. getSynSet ($Q'(i). \text{Keyword}$)

3.4.1 Common Sense Reasoning

After the core lexical analysis that attach the appropriate synsets with the original query word. The selected pre-processed query terms are then transferred into the common sense reasoning phase that attach the context or the concepts rather than the words by using the common sense knowledgebase i.e. ConceptNet. ConceptNet covers a wide range of common sense concepts along with its more diverse relational ontology as well as its large number of inter conceptual relations.

In our model, we extract the common sense reasoning by using the Knowledge-Lines also called K-Lines from ConceptNet. K-Lines are the Conceptual correlation. ConceptNet contain the eight different kinds of K-Line categories that combine the K-Line into the ConceptNet twenty relationships. That helps in the conceptual reasoning. The overview of the ConceptNet is given below

ConceptNet

ConceptNet [Liu, et al. 2004] is a commonsense knowledgebase. ConceptNet 2.1 also encompasses MontyLingua, a natural-language-processing package. ConceptNet is written in Python but its commonsense knowledgebase is stored in text files. Unlike other knowledgebases like CYC, FrameNet and Wikipedia, ConceptNet is based more on context and allow a computer to understand new concepts or even unknown concepts by using conceptual correlations called Knowledge-Lines. ConceptNet is at present deliberated to be the biggest commonsense knowledgebase [Liu, et al. 2004], [Hsu, et al. 2008]. It is composed from more than 700,000 free text contributors' assertions. Its nodes core structure is concepts, which each of which is a part of a sentence that expresses a meaning. ConceptNet is a very wealthy knowledgebase for several aspects: First, it includes an immense number of assertions and nodes. Second, it has a broad range of information. Finally, it has different kinds of relationships, including description parameters. Figure 3.6 presents a snapshot that includes useful relationships between concepts. In the last version of ConceptNet "ConceptNet4", each relationship has several fields expressing its score, polarity and generality. This information is automatically inferred by examining the frequency of the sentences that provoked this relationship.

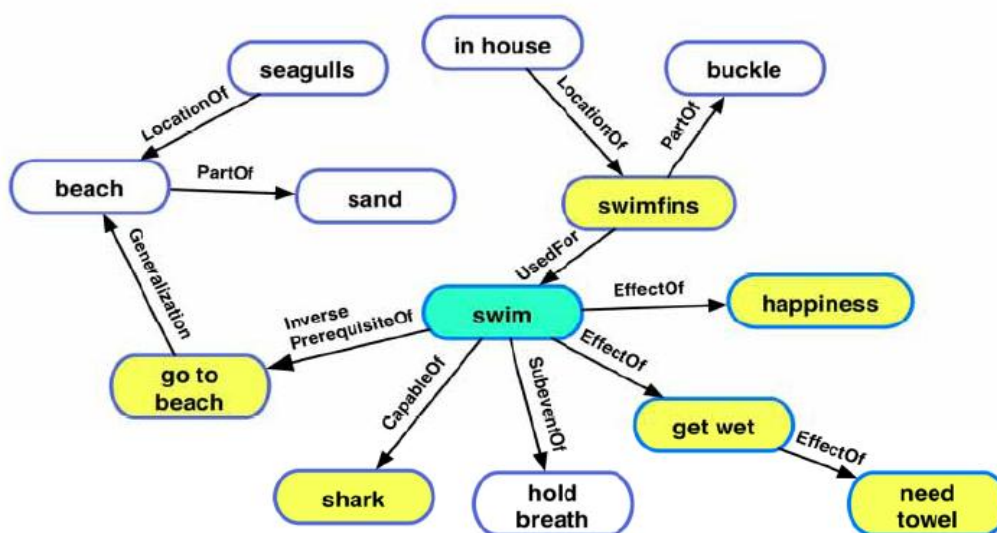


Figure 3.6: An illustration of a small section of ConceptNet

ConceptNet is a contextual common sense reasoning system for common sense knowledge representation and processing. ConceptNet is developed by MIT Media Laboratory and is presently the largest common sense Knowledgebase [Liu et al. 2004b]. ConceptNet enable

the computer to think like a human. ConceptNet is the semantic network representation of the OMCS (Open Mind Common Sense) knowledgebase. It contains 300,000 nodes, 1.6 million edges and 20 relations that are IsA, HasA, PartOf, UsedFor, AtLocation, CapableOf, CreatedBy, MadeOf, HasSubevent, HasFirstSubevent, HasLastSubevent, HasPrerequisite, MotivatedByGoal, Causes, Desires, CausesDesire, HasProperty, ReceivesAction, DefinedAs, SymbolOf, LocatedNear, ObstructedBy, conceptuallyRelatedTo, InheritsFrom etc.

ConceptNet has not been so much well known in the IR like the WordNet. Only few people have used it for the expanding the query with the related concepts [Liu et al. 2002], [Li et al. 2008], [Hsu et al. 2008]. Common sense reasoning also used in image retrieval by expanding the Meta data attached to the image with the spatially related concepts. The experiments are conducted on the Image CLEF 2005 data set and proved that the common sense reasoning improves the retrieval performance. ARIA (Annotation and Retrieval Integration Agent) contain both the annotation as well as the retrieval agent. The annotation agent apply the common sense reasoning to annotate the images while the retrieval phase perform the common sense reasoning to bridge the semantic gap and to retrieve the relevant images [Lieberman et al., 2001]. Several surveys have conducted to show the importance of the common sense reasoning for several applications [Lieberman et al. 2004]. Nevertheless, the improvement in the precision accounts for the interest of introducing Common sense reasoning in the information retrieval systems. The comparison of the WordNet and the ConceptNet is conducted on the TREC-6, TREC-7 and TREC-8 data sets and concluded that WordNet have higher discriminative ability while ConceptNet have higher concept diversity [Hsu et al. 2006].

As we have discussed above the common sense reasoning module extend the user query conceptually. The output of the candidate term selection module of the core lexical analysis serves as an input into the common sense reasoning. The input is in the form of list of the selected query terms. The output of the common sense reasoning is the list of the selected concepts attached with the query.

Proposed Algorithm 3.2: Common Sense Reasoning

Input: $Q' \rightarrow \cup_{i=1}^{t'} K_i$ // List of selected candidate query terms

Output: $Q'' \rightarrow \cup_{i=1}^{t''} K_i''$

Method:

$i \leftarrow \text{Length}(Q')$

$Q'(i). \text{ConceptSet} \leftarrow \text{ConceptNet.getConceptSet}(Q'(i). \text{keyword})$

3.4.2 Candidate Concept Selection

Our candidate concept selection employs the semantic similarity function for estimating the similarity between terms that are relatively semantically similar in order to reduce the noise. The semantics of keywords are identified through the relationships between keywords by performing semantic similarity on them [Fang et al. 2005] [Andrea et al. 2003] [Varelas et al. 2005] [Bonino et al. 2004] [Khan et al. 2006] [Sayed et al. 2007]. Experiment results show that all the similarity functions improve the retrieval performance, although the performance improvement varies for different functions. We find that the most effective way to utilize the information from WordNet is to compute the term similarity based on the overlap of synset definitions. Using this similarity function in query expansion can significantly improve the retrieval performance.

The WordNet semantic similarity function is used to calculate the semantic similarity between the original selected query terms with the expanded term. The lexical expansion module and the common sense reasoning module come up with the set of the expanded terms. These expanded terms will probably increase the recall but sometimes significantly decrease the precision of the system. Some of them are noises and it will derogate the retrieval performance. However, if the two terms have not that much in common then it will increase the recall at the expense of precision. This is one of the drawbacks of the query expansion. We have tried to control the bottleneck of selecting some of the most similar concepts by utilizing the candidate concept selection module. The candidate concept selection uses the semantic similarity function to prune the less semantically similar expanded terms. The semantic similarity between the selected candidate terms and the set of lexical and conceptual expanded term is computed. The

threshold has been defined for selecting the list of candidate concepts. The threshold can be set by taking the average mean of all the semantic similarity value of the expanded terms. The expanded terms with the semantic similarity value below the threshold are prune and the rest of them are selected and passed to the retrieval model for retrieving the result according to the expanded query terms.

Semantic similarity can be measured in order to filter the concepts. This will significantly increase the precision of the system. The various semantic similarity measures are discussed below.

3.4.2.1 Semantic Similarity Calculation

Due to the subjectivity in the definition of the semantic word similarity, there is no unique way to compute the performance of the proposed measures. These measures are folded into two groups in [Mihalcea et al. 2006], corpus-based and knowledge-based similarity measures. The corpus-based measure attempts to recognize the similarity between two concepts exploiting the information exclusively derived from large corpora. The knowledge-based measures try to quantify the similarity using the information drawn from the semantic networks.

Knowledge-based Word Similarity Measures

The knowledge-based technique measures the similarity between two concepts employing the information drawn from the semantic networks. Most of these measures use WordNet [Miller et al. 1990] as the semantic network. The similarity between two concepts and two words is not same. Some words have different senses or different concepts. In order to compute the semantic similarity all the sense of the words are considered. The score are assigned to all the sense of words and then select the highest similarity score. Some of these similarity measures use information content (IC) which represents the amount of information belonging to a concept. It is described as

$$IC(c) = -\log (P(c)) \quad (3.3)$$

Where $IC(c)$ is the information content of the concept c , and $P(c)$ is the probability of encountering an instance of the concept c in a large corpus. Another used definition is the least common subsumer (LCS) of two concepts in taxonomy. LCS is the common ancestor of both concepts, which has the maximum information content.

Leacock & Chodorow Similarity

This similarity measure is introduced in [Leacock. et al. 1998]. The similarity between two concepts is defined as

$$\text{Sim}_{\text{lech}}(c_i, c_j) = \log \left(\frac{\text{length}(c_i, c_j)}{2 \times D} \right) \quad (3.4)$$

where c_i, c_j are the concepts, $\text{length}(c_i, c_j)$ is the length of the shortest path between concepts c_i and c_j using node counting, and D is the maximum depth of the taxonomy.

Lesk Similarity

In Lesk measure, [Lesk. et al. 1986] similarity of two concepts is defined as a function of overlap between the definitions of the concepts provided by a dictionary. It is described as

$$\text{Sim}_{\text{lesk}}(c_i, c_j) = \frac{\text{def}(c_i) \cap \text{def}(c_j)}{\text{def}(c_i) \cup \text{def}(c_j)} \quad (3.5)$$

Where $\text{def}(c)$ represents the words in definition of concept c . This measure is not limited to semantic networks, it can be computed using any electronic dictionary that provides definitions of the concepts.

Wu & Palmer Similarity

This similarity metric [Wu. et al. 1994] measure the depth of two given concepts in the taxonomy, and the depth of the LCS of given concepts, and combines these figures into a similarity score

$$Sim_{wnp}(c_i, c_j) = \frac{2 \times depth(LCS(c_i, c_j))}{depth(c_i) + depth(c_j)} \quad (3.6)$$

Where $depth(c)$ is the depth of the concept c in the taxonomy, and $LCS_{(c_i, c_j)}$ is the LCS of the concepts c_i and c_j .

Resnik Similarity

Resnik similarity measure [Resnik et al. 1995] is defined as the information content of the LCS of two concepts

$$Sim_{res}(c_i, c_j) = IC(LCS(c_i, c_j)) \quad (3.7)$$

Lin's Similarity

The key idea in this measure is to find the maximum information shared by both concepts and normalize it. Lin's similarity [Lin et al. 1998] is measured as the information content of LCS, which can be seen as a lower bound of the shared information between two concepts, and then normalized, with the sum of information contents of both concepts. The formulation is as below

$$Sim_{lin}(c_i, c_j) = \frac{2 \times IC(LCS(c_i, c_j))}{IC(c_i) + IC(c_j)} \quad (3.8)$$

Jiang & Conrath Similarity

This measure is introduced in [Jiang et al. 1997]. This measure also uses IC and LCS. It is defined as below

$$Sim_{jnc}(c_i, c_j) = \frac{1}{IC(c_i) + IC(c_j) - 2 \times IC(LCS(c_i, c_j))} \quad (3.9)$$

Hirst & St-Onge Similarity

This measure is a path based measure, and classifies relations in WordNet as having direction. For example, is-a relations are upwards, while has-part relations are horizontal. It establishes the similarity between two concepts by trying to find a path between them that is

neither too long nor that changes direction too often. This similarity measure is represented with Sim_{hso} . Detailed description of this method can be found in [Hirst et al. 1998].

The proposed candidate concept selection selects the candidate concepts from the list expanded terms in order to prune the noises. The candidate terms are selected on the basis of semantic similarity between the expanded terms and the original query terms are computed and then selected according to the threshold. The threshold can be computed by taking an average mean of the expanded terms and the query terms. The output of the algorithm is the list of selected candidate terms and these selected expanded terms are then used for further retrieval and ranking. The overall algorithm of proposed module is given below.

Proposed Algorithm 3.3: Candidate Concept Selection

Input: $Q \rightarrow \cup_{i=1}^t K_i$ // List of original query terms, Synset and ConceptNet

Output: $Q' \rightarrow \cup_{i=1}^{t'} K'_i$ // List of selected query terms

Method:

$I \leftarrow \text{Length}(Q)$

// Adding Semantic Similarity
 $Q(i).keyword \leftarrow Q(j).keyword$

$J \leftarrow \text{Length}(Q(i).synset)$
 $Q(i).synSet(j).SS \leftarrow \text{WordNet.SemSim}(Q(i).SynSet(j).Word)$
 $Q(i).ConceptSet(j).SS \leftarrow \text{WordNet.SemSim}(Q(i).keyword, Q(i).ConceptSet(j).Word)$

//Select candidate terms from the SynSet
 $TH \leftarrow Q(i).SynSet.AvgMean()$

$K \leftarrow \text{Length}(Q(i).Synset)$
 $Q' \leftarrow Q(i).Keyword$
If $(Q(i).SynSet(k).SS \geq th)$
 $Q' \leftarrow Q' + Q(i).SynSet(k).keyword$

//Select candidate term from ConceptSet
 $Th \leftarrow Q(i).ConceptSet.AvgMean()$

$H \leftarrow \text{Length}(Q(i).ConceptSet)$
IF $(Q(i).ConceptSet(h).Keyword$

3.4.3 Retrieval and Ranking of Result

For retrieving and ranking the results, we use the one of the standard model the Vector Space Model (VSM) that is for information filtering, information retrieval, and indexing and relevancy rankings. This model has been used for the last few decades in information retrieval. This model is based on linear algebra. The vector space model (VSM) [Salton et al. 1975] is one of the renowned models in IR. The most popular retrieval model that of the vector model, allows each document to be weighted on a sliding scale. This allows documents to be ranked according to degree of similarity and was chosen as the most suitable method. Other models were not pursued owing to poor performance and over complexity for the task in hand. The VSM operates by delineating each document as an n dimensional vector space. The similarity between the query and the document is compared by using the cosine measure. The smaller the angle, the similar is the document.

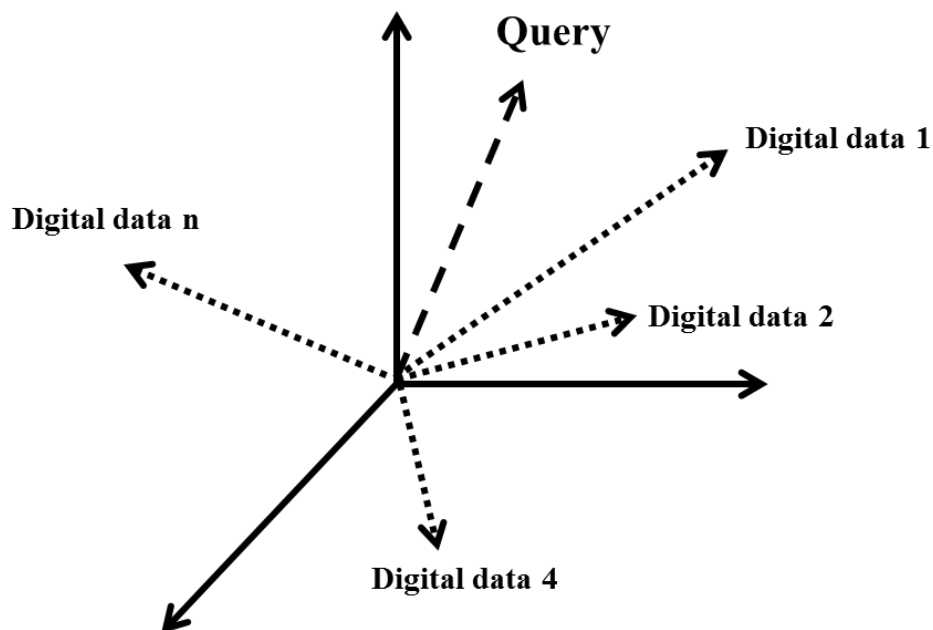


Figure 3.7 Representation of the Vector Space Model

Retrieved documents using the vector space model are ranked according to the weights according to term frequency (tf) and the inverse document frequency ($tf-idf$). The tf value measures the salience of a term within a document, and the idf value measures the overall importance of the term in the entire document collection. A high term frequency and inverse documents represents a high frequency of the term with in the document. The higher the $tf \times idf$

weight, the more relevant a given document is to a given term. The following calculations compute the *tf* and *idf* measures respectively.

$$Tf_{ij} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (3.10)$$

Where $N_{i,j}$ is the number of occurrences of the term I in document j and the denominator is the total number of occurrences of all terms in the document d_j

$$idf_i = \log \frac{|D|}{|\{d:t_i \in d\}|} \quad (3.11)$$

Where $|D|$ is the total number of documents and the denominator is the number of documents that contain the term t_i

We used it as baseline method for checking our algorithm. It also computes the similarity between the expanded query terms and the images. In VSM, the images as well as the query terms are represented in the form of vectors. The similarity between a document and a query is calculated by the cosine of the angle between the image vector and the query vector. The expanded query is compared against the meta data (annotation) attach with the images in the corpus and then the results are ranked accordingly. The term frequency and the inverse document frequency are the widely used factors for calculating the weight of the image. Therefore, the images with the largest number of concepts are ranked better. The overall algorithm of the retrieval and ranking the images is discussed below. The input is the list of selected candidate concepts and the output is the list of the ranked images according to the query terms.

3.4 Experiments

However, the semantic accuracy is the main focus of our research. Bridging the semantic gap is the overall theme of our research. All experiments and evaluation of proposed framework have been performed on the LabelMe datasets, available freely for research. LabelMe is a project created by the MIT Computer Science and Artificial Intelligence Laboratory (CSAIL) which provides a dataset of digital images with annotations. The dataset is dynamic, free to use, and open to public contribution. As of October 31, 2010, LabelMe has 187,240 images, 62,197 annotated images, and 658,992 labelled objects.

The aspiration behind creating LabelMe comes from the history of publicly available data for computer vision researchers. Most available data was tailored to a specific research group's problems and caused new researchers to have to collect additional data to solve their own problems. LabelMe was created to solve several common shortcomings of available data. The following is a list of qualities that distinguish LabelMe from previous work.

- Designed for recognition of a class of objects instead of single instances of an object. For example, a traditional dataset may have contained images of dogs, each of the same size and orientation. In contrast, LabelMe contains images of dogs in multiple angles, sizes, and orientations.
- Designed for recognizing objects embedded in arbitrary scenes instead of images that are cropped, normalized, and/or resized to display a single object.
- Complex annotation: Instead of labelling an entire image (which also limits each image to containing a single object), LabelMe allows annotation of multiple objects within an image by specifying a polygon bounding box that contains the object.
- Contains a large number of object classes and allows the creation of new classes easily.
- Diverse images: LabelMe contains images from many different scenes.
- Provides non-copyrighted images and allows public additions to the annotations. This creates a free environment.

The number of images in the datasets is continuously increasing day by day. As the researchers are adding new images along with the annotation data. The experiment has been conducted on some of the categories from the LabelMe 31.8 GB dataset. We have selected 181, 932 images with 56946 annotated images, 352475 annotated objects and 12126 classes for performing the experiments.

The experiments are firstly performed to make a comparison between the LabelMe Query systems, WordNet based expansion, and ConceptNet based expansion and the proposed Semantic Query Interpreter. The LabelMe query (LM query) system works on the text matching technique. The LM query module compares the text in the query with the tags attached with the image. The LM is the open annotation tool any one can annotate the LabelMe images. The WordNet has been used in the LabelMe web based annotation tool in order to remove the problem of sense disambiguation and to enhancing object labels with WordNet. The LM query system works well for the single word single concept query but flunks in the case of multi-concept queries or the complex queries.

As we, all are well aware, that query plays a primal role in the IR systems. Moreover, the query is the translation of the user's requirements and needs. The retrieved results can be evaluated by means of the relevancy with the information need. Information retrieval systems have been evaluated for many years. Evaluation is the major part of the retrieval systems. Information science has developed many different criteria and standards for the evaluation e.g. effectiveness, efficiency, usability, satisfaction, cost benefit, coverage, time lag, presentation and user effort, etc. Among all these evaluation technique precision which is related to the specificity and recall which are related to the exhaustively are the well accepted methods. In our approach, we use precision, recall and F-Measure for evaluating the performance. For calculating, the precision and recall the retrieved, relevant and irrelevant as well the non-retrieved relevant as well as the relevant information must be available. While for the F-measure, we need the value of precision and recall.

As we, all are well aware, that query plays a primal role in the IR systems. Moreover, the query is the translation of the user's requirements and needs. The retrieved results can be evaluated by means of the relevancy with the information need. Information retrieval systems have been evaluated for many years. Evaluation is the major part of the retrieval systems. Information science has developed many different criteria and standards for the evaluation e.g.

effectiveness, efficiency, usability, satisfaction, cost benefit, coverage, time lag, presentation and user effort, etc. Among all these evaluation technique precision which is related to the specificity and recall which are related to the exhaustively are the well-accepted methods. In our approach, we use precision, recall and F-Measure for evaluating the performance. For calculating, the precision and recall the retrieved, relevant and irrelevant as well the non-retrieved relevant as well as the relevant information must be available. While for the F-measure, we need the value of precision and recall.

Precision is the fraction of the documents retrieved that are relevant to the user’s information need. The precision can be calculated by the formula given below

$$\text{Precision} = \frac{|A \cap B|}{|B|} \quad (3.12)$$

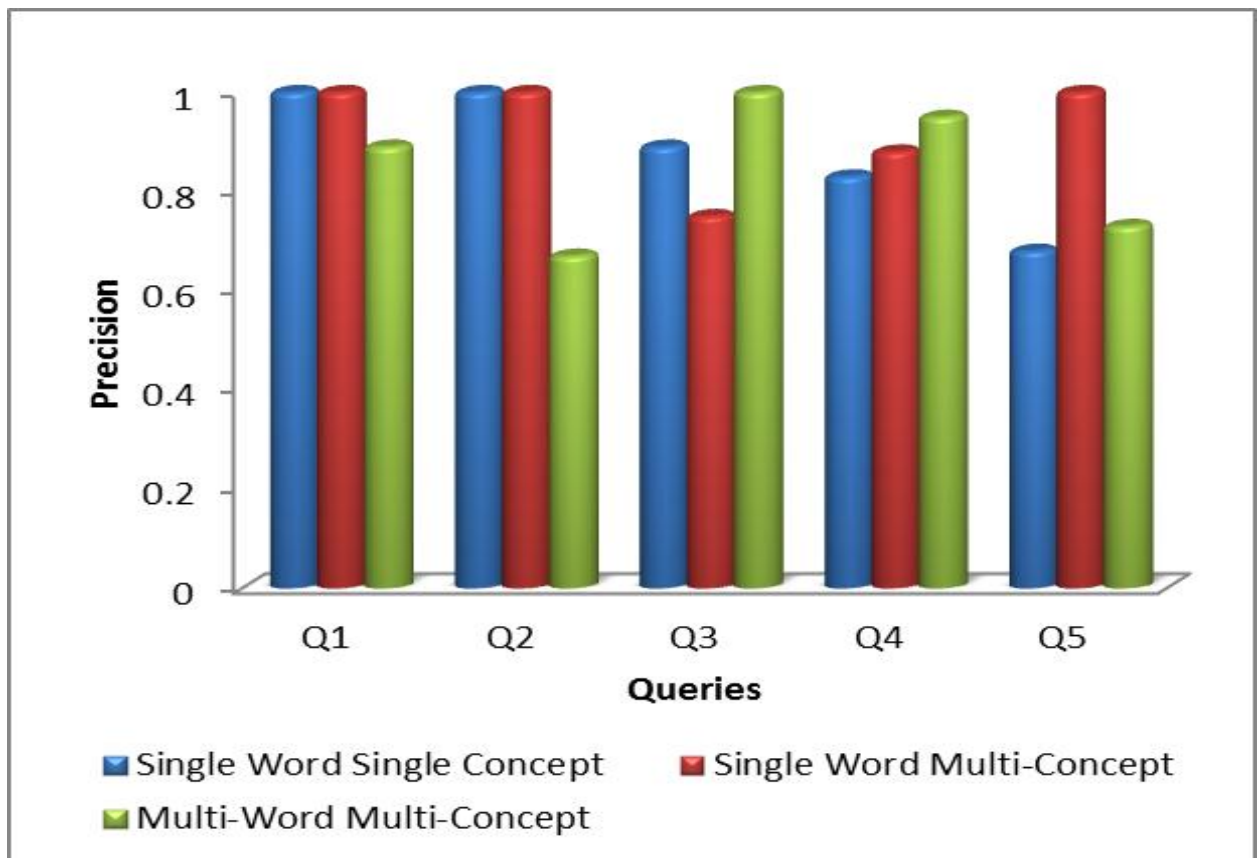


Figure3.8: Precision of five queries of three different types using proposed SQI

Figure 3.8 shows the precision of the five randomly selected different queries for each of the three categories i.e. Single Word Single Concept, Single Word Multi-Concept and Multi-Word Multi-Concept queries. The five randomly selected single word single concept queries are car, building, tree, sky, and house. The five randomly selected single word multiconcept queries are street, park, transport, game and office. While the five randomly selected multiword multiconcept queries are car on the road, people in the park, allow me to view building in the street, people on the seaside and people sitting on the benches. The results show the substantial improvement of the retrieval precision in case of the single word single concept and single word multiconcept. The mean average precision of the single word single concept queries is 0.88, the mean average precision of the single word multi-concept queries is 0.96 and the mean average precision of the multi word multi-concept queries is 0.85. The result showed that the system works very well for many cases i.e. queries but for some cases, there is little bit variation. The difference in the precision level of the different types of queries is due to the query complexity and due to the poor annotation. As with the increase in the complexity, there is a decrease in the performance efficiency. The system can expand the queries but fails to contribute in the annotation. Our proposed Semantic query interpreter has shown the significant precision level over the LabelMe dataset. As we know that sometimes, the query expansion increases the recall of the system and decreases the precision. We have maintain the precision of the by selecting the candidate concepts selection module (see section 3.4.2). This module pruned the most semantically relevant concepts among the expanded concepts to the original selected query terms. The query terms are selected by using candidate term selection module (see section 3.3.1.2) of the core lexical analysis. The candidate concepts selection module intents to maintain the precision of the system by selecting the expanded concepts based on semantic similarity between them.

While Recall is the fraction of the documents that are relevant to the query that are successfully retrieved, and can be calculated

$$\text{Recall} = \frac{|A \cap B|}{|A|} \quad (3.13)$$

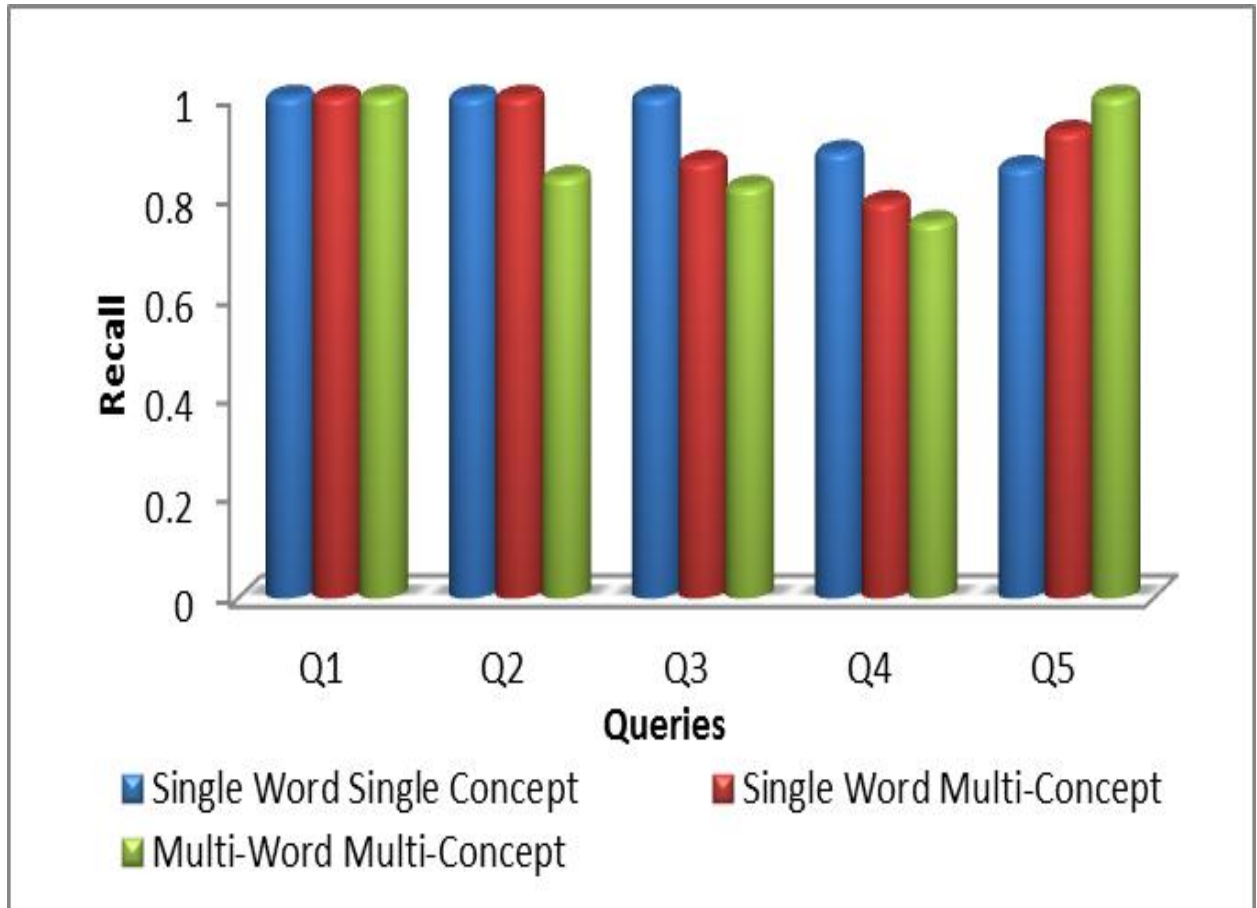


Figure 3.9: Recall of five queries of three different types using proposed SQI

Figure 3.9 shows the recall of the five randomly selected different queries for each of the three categories i.e. single Word Single concept, Single Word Multi-Concept and Multi-Word Multi-Concept queries. The same five randomly selected three different categories of queries that are used for computing the precision is used for recall computation as well. The result shows the substantial improvement of the recall of the proposed model. The recall of the system can increase more if we remove the candidate concept selection module (see section 3.4.2) of the proposed framework. The mean average recall of the single word single concept queries is 0.95, the mean average recall of the single word multi-concept queries is 0.92 and the mean average recall of the multi-word multi-concept queries is 0.89.

Since Precision and Recall specify the performance of a system from two very different points of view, we also used a combined measure of them, namely F-Measure (Baeza-Yates and Ribeiro-Neto 1999). F-Score, weighted harmonic mean or F-Measure can be defined as

$$\text{F-Measure} = 2 \cdot \frac{\text{Pre} \cdot \text{Rec}}{\text{Pre} + \text{Rec}} \quad (3.14)$$

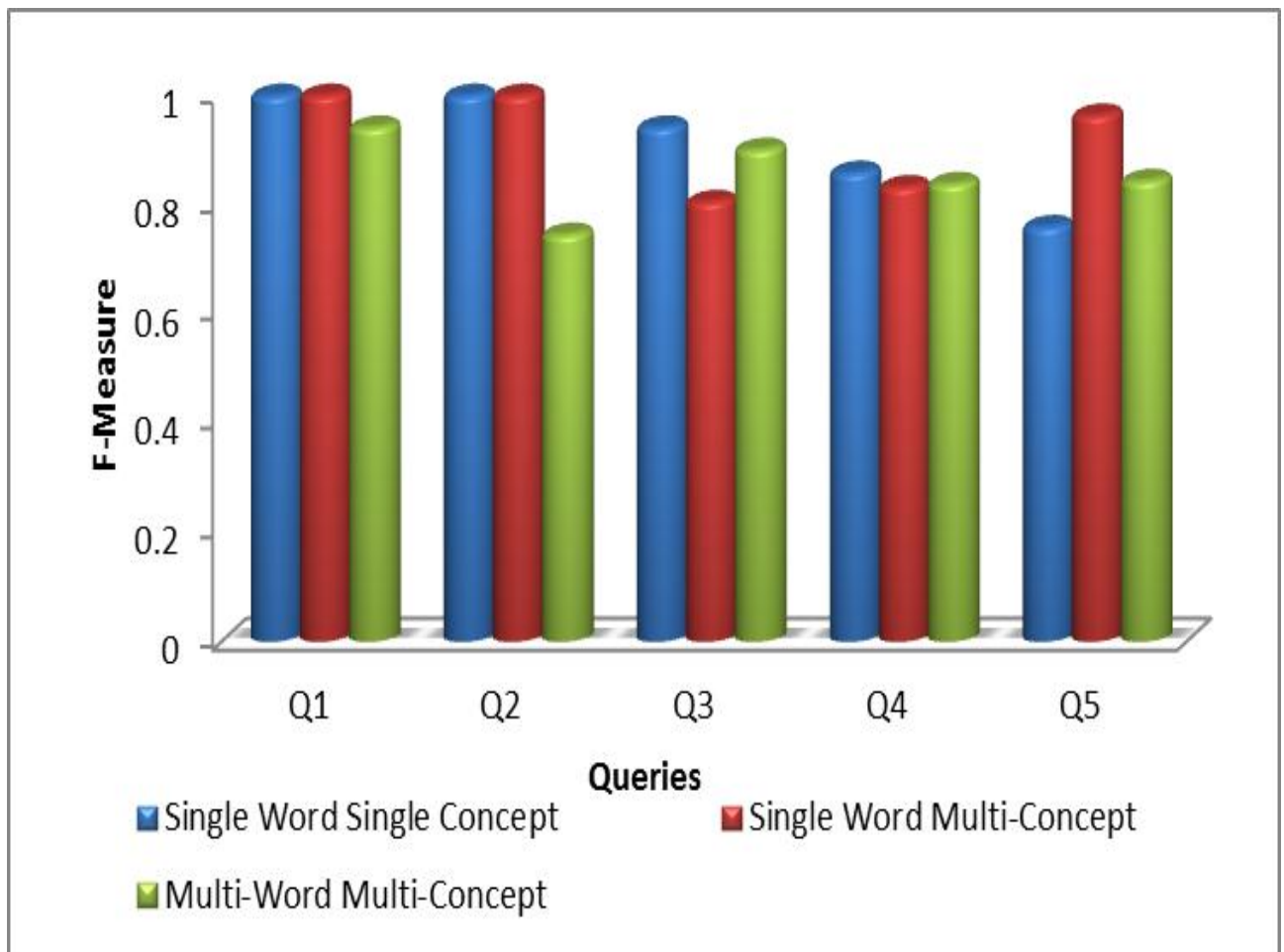





Figure 3.10: F-Measure of five queries of three different types using proposed SQI



F-Measure ranges in the real interval [0, 1], the higher it is the better a system works. Ideally, the recall and precision should both be equal to one, meaning that the system returns all

relevant documents without introducing any irrelevant documents in the result set. Unfortunately, this is impossible to achieve in practice. If we try to improve recall (by adding more disjunctive terms to the query, for example), precision suffers; likewise, we can only improve precision at the expense of recall. Furthermore, there is often a tradeoff between retrieval effectiveness and computing cost. As the technology moves from keyword matching to statistical ranking to natural language processing, computing cost increases exponentially.

Figure 3.10 shows the F-measure of the five randomly selected different queries for each of the three categories i.e. single Word Single concept, Single Word Multi-Concept and Multi-Word Multi-Concept queries. The mean average F-measure of the single word single concept queries is 0.91, the mean average F-measure of the single word multi- concept queries is 0.92 and the mean average F-measure of the multi-word multi-concept queries is 0.86. The mean average F-measure of the multi-word multi-concept query is lesser than the single word single concept and single word multi-concept. It is because with the increase in the complexity the efficiency decreases and is difficult to deal with.

Table 3.1: Five randomly selected single word single concept query, the expanded query terms by using the Semantic query Interpreter and the top ten retrieved results.





Single Word Single Concept		
Queries	Expanded terms	Outputs
Car	auto(1),automobile(1),machine(1), motorcar(1), railcar(1), railway car(0.87), railroad car(0.87), cable car(0.89), gondola(1), elevator car(0.89)	
Building	building(1), edifice (1), construction (1), aisle (0.8), apartment (0.8), apartment building (0.93), city (0.8), difference (0.63), one level (0.86), persons own restaurant business (0.67), second floor (0.86), shape (0.71), tank (0.71), toilet (0.75), window (0.71)	
Tree	tree (1), corner (1), shoetree(1), animal (0.67), child (0.62), difference (0.57), ground (0.57), large plant (0.88), person (0.67), pink ball (0.78), shape (0.77)	

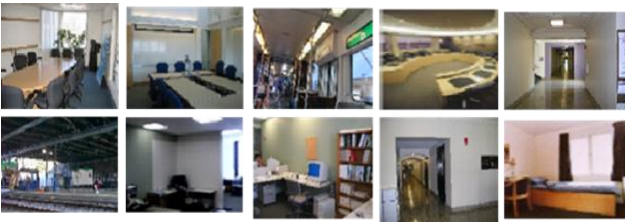
<p>Sky</p>	<p>sky (1), air balloon (0.61), earth (0.5), forest fire (0.44), rain (0.53), three hot air balloon (0.38), water particle(0.58)</p>	
<p>House</p>	<p>house (1), family (1), household (1), menage (1), theatre (1), firm (1), business firm (0.98), sign of the zodiac (0.94), sign (1), mansion (1), bed (0.71), dinner (0.71), family (1), gold-tipped fountain pen (0.55), group (0.73), jack (0.59), one part (0.91), open bottle (0.59), role (0.78), shore (0.55), table (0.8)</p>	

The Table 3.1 shows the randomly selected five Single Word Single Concept queries and their expanded terms along with the semantic similarity values, by using the Semantic Query Interpreter and the top 10 ranked retrieved results. The first query in the Table 3.1 is car was expanded with the following terms auto, automobile, machine, motorcar, railcar, railway car, railroad car, cable car, gondola and elevator car. The SQI successfully tagged most of the relevant terms with the query and make it more flexible and easy to deal with. The SQI retrieve most of the relevant results according to the query from the dataset. But unfortunately, among these results, some of the most relevant ones are displayed after the less relevant ones. SQI have successfully retrieved the relevant results but the ranking of the results was not appropriate. The second query building is also the single word single concept query and was expanded using proposed technique by the following concepts building, edifice, construction, aisle, apartment, apartment building, city, difference, one level, persons own restaurant business, second floor, shape, tank, toilet and window. The SQI successfully expanded the query and same problem held with the building query i.e. the ranking problem. The third query tree and their expanded terms are tree, corner, shoe tree, animal, child, difference, ground, large plant, person, pink ball and shape and the fourth query is sky and their expanded terms are sky, air balloon, earth, forest fire, rain, three hot air balloon and water particle and the last query is house and their expanded terms are house, family, household, menage, theatre, firm, business firm, sign of the zodiac, sign, mansion, bed , dinner, family, gold-tipped fountain pen, group, jack, one part, open bottle, role, shore and table. We have controlled the noisy terms from the expanded term by using the candidate concepts selection module. But still a little amount of noisy terms are there. It is due to the reason that these noisy terms are included in the concept tree of the knowledgebases. We

cannot completely remove the noisy terms but we can controlled and in the propose framework we attempts to control these noisy terms by using the candidate concept selection module. Many of the systems easily process such type single word single concept queries but the problem lies for the single word multiconcept and multiword multi concept queries. SQI have successfully retrieved the relevant results but the ranking of the results was not appropriate. We attempts to remove the bottleneck of the traditional ranking technique by proposing the Semantic ranking strategy known as SemRank, the detail discussion will be made in chapter 4. The results for the single word multi-concept and multi-word multi-concept queries are shown below.

Table 3.2: Five randomly selected single word multi- concept query, the expanded query terms by using the Semantic query Interpreter and the top ten retrieved results.





Single Word Multi-Concept		
Queries	Expanded terms	Outputs
Street	car(1),Road(1),auto(1),automobile(1),machine(1),motorcar(1),railcar(1),gondola(1),route(1),bed(0.62),brake(0.67),drive(0.71),front part(0.56),hood(0.62),itshead(0.8),lane(0.62),light(0.71),motorcycle(0.91) ,plane(0.73) ,road trip(0.63),bluecircle(0.64) ,car(0.62),line(0.93) ,mirrorimage(0.65),othercar(0.62), tree (0.88)picture(0.67),rearfender(0.67),road(1),slushy road(0.67),street(1),tire(0.62)	
Park	people (1), park (1), citizenry (1), multitude (1), masses (1), mass (1), parkland (1), commons (1), common (1), green (1), ballpark (1), parking lot (0.98), activity person (0.67), child (0.73), city (0.71), family (0.67), Frisbee (0.67), front (0.62), garage (0.71), parking area (0.93), path (0.67), place (0.82), slide (0.67)	
Transport	building (1), street(1), edifice (1), construction (1), aisle (0.80), apartment (0.80), apartment building (0.93), city (0.80), difference (0.63), one level (0.86), person's own restaurant business (0.67), second floor (0.86), shape (0.71), tank (0.71), toilet (0.75), window (0.71), avenue (0.95), ball (0.62), driveway (0.82), flow (0.67), gas station (0.65), park (0.59), same street (1), subway (0.63), surface (0.67)	
Game	people (1), seaside (1), citizenry (1), multitude (1), masses (1), mass (1), seaboard(1), many thing (0.40)	

<p>Office</p>	<p>image (1), people (1), bench (1), picture (1), place (1), pose (1), position (1), lay (1), setup (1), localize (1), place (1), lay out (1), terrace (1), judiciary (1), workbench (1), photograph (0.94), statue (0.67), water (0.67), flower (0.17), person's house (0.24), person (0.71), be two woman (0.62), long chair (0.90), park (0.56), person (0.57), person's own pair (0.54), tree(0.43)</p>	
----------------------	---	--

The Table 3.2 shows the randomly selected five Single Word Multi- Concept queries and their expanded terms along with the semantic similarity values, by using the Semantic Query Interpreter and the top 10 ranked retrieved results. The first query street that is the single word abstract concept and contain several other concepts like tree, road, building, car etc. That combinely illustrate a concept street. The expanded concepts for the first query street are car, road, auto, automobile, machine, motorcar, railcar, gondola, route, bed, brake, drive, front part, hood, itshead, lane, light, motorcycle, plane, road trip, blue circle, car, line, mirror image, other car, tree, picture, rearfender, road, slushy road, street and tire. The system successfully processed the query street. The second query i.e. park is also multi concept query that contains other concepts like people, garden, tree, Frisbee, place, garage, parking lot, motor cycle etc. Park is an ambiguous query or heteronyms (words that have same spelling with different meanings). Park shares two concepts i.e. car park and recreation park. These types of queries are difficult for the system to be dealt with, even when the query is too short. Our system displays most of the recreational park images for park query. The second query park and their expanded concepts are people, park, citizenry, multitude, masses, mass, parkland, commons, common, green, ballpark, parking lot, activity person, child, city, family, frisbee, front, garage, parking area, car, path, place and slide. Among the expanded concepts, we can see it contain the concepts of both car park and recreational park. The retrieved results, some of the results are irrelevant. It is because these images are tag with these concepts. The third query transport is the multi-concept query that contains any type of transport. The transport query is the generalization of the query car. The expanded concepts of the query transport are building, street, edifice, construction, aisle, apartment, apartment building, city, difference, one level, person's own restaurant business, second floor, shape, tank, toilet, window, avenue, ball, driveway, flow, gas station, park, same street, subway and surface. The query game is also a generalization of all different types of games. The expanded concepts of the query game are people, seaside, citizenry, multitude, masses, mass, seaboard and many things. While the last query office and their expanded terms

are image, people, bench, picture, place, pose, position, lay, setup, localize, place, lay out, terrace, judiciary, workbench, photograph, statue, water, flower, person's house, person, be two woman, long chair, park, person, person's own pair and tree. The SQI successfully retrieve most of the relevant queries like the single word single concepts but the ranking is not appropriate. The results for the multi-word multi-concept queries are shown below.

Table 3.3: Five randomly selected multi-word multi-concept queries, the expanded query terms by using the Semantic query Interpreter and the top ten retrieved results.

Multi-Word Multi-Concept		
Queries	Expanded terms	Outputs
Car on the road	street(1), avenue (0.95), ball (0.62), driveway (0.82), flow (0.67), gas station (0.65), park (0.59), same street (1), subway (0.63), surface (0.67)	
People in the park	park (1), parkland(1), commons (1), commons(1), green (1), ballpark(1), parking lot(0.98), activity person (0.67), child (0.73), city(0.71), family (0.67), Frisbee (0.67), front (0.62), garage (0.71), parking area (0.93), path (0.67), place (0.82), slide (0.67)	
Allow me to view building in the street	transport(1), conveyance (1), transportation (1), shipping (1), ecstasy (1), rapture (1), exaculation (1), ratus (1), aircraft (0.82), car (0.82), carry (1), form (0.75), ground (0.80), ground transportation (0.93), item (0.67), long distance (0.67), stand (0.78), transportation (1), transportation device (0.95), truck (0.80)	
People on the Seaside	game (1), biz(1), plot (1), brain (0.80), break (0.71), checkers (0.84), eight ball (0.83), gambling (0.89), game (1), injury (0.82), jack (0.89), more fun (0.88), other person (0.77), sixteen ball (0.83), two different fact (0.70)	

<p>People Sitting on the Benches</p>	<p>office(1),business office(0.98) ,agency(1), bureau(1), authority(1) ,function(1) ,part(1),role(1),power(1),office staff(0.98) ,position(1), post(1), berth(1), spot(1), billet(1), place(1), situation(1),boss(0.71) ,dentist work(0.64),eraser holder(0.58),fouryear(0.62), location(0.82),many piece(0.62), paper(0.59), paperwork(0.84), report(0.67), similarity(0.62), tape dispenser(0.61), tool(0.59), window(0.62)</p>	
---	--	--

The Table 3.3 shows the randomly selected five Multi-Word Multi-Concept queries and their expanded terms along with the semantic similarity values, by using the Semantic Query Interpreter along with the top 10 ranked results. Most of the multi-word multi-concept queries are successfully expanded the query term by using the SQI and most of the relevant results are retrieved but among these results. The first multiword query in the table is the “car on the road” which is the combination of two main concepts i.e. car and the road. While road contain several other concepts like street, driveway etc. The expanded concepts for the first query are street, avenue, ball, driveway, flow, gas station, park, same street, subway and surface. The second query “people in the park” also contain two main concepts like people and park. The word park is heteronyms but the word people in the query, points towards the recreational park and makes the query clear. Even though the people are also available in the car park but mostly, found in the recreational park. If in the query park comes with the word car then it points towards the parking lot or area. The expanded concepts for the query “people in the park” are park, parkland, commons, commons, green, ballpark, parking lot, activity person, child, city, family, Frisbee, front, garage, parking area, path, place and slide. The next query “building in the street” is also the combination of two main concepts building that is the single word single concept and street that is the single word multi-concept. The query building in the street is the integration of single concept and multi-concept words. The expanded terms of the query are “allow me to view building in the street” are transport, conveyance, transportation, shipping, ecstasy, rapture, exaculation, ratus, aircraft, car, carry, form, ground, ground transportation, item, long distance, stand, transportation, transportation device and truck. The other query “people on the seaside” and their expanded concepts are game, biz, plot, brain, break, checkers, eight ball, gambling, game, injury, jack, more fun, other person , sixteen ball and two different fact. While the last query “people sitting on the benches” are office, business office, agency, bureau, authority, function, part, role, power ,office staff, position, post, berth, spot, billet, place, situation, boss, dentist work, eraser holder, four year, location, many piece, paper, paperwork, report, similarity,

tape dispenser, tool and window. The system retrieved most of the relevant results but the ranking of the result is not appropriate it is because of the bottleneck of the retrieval model that we have used. In the next chapter, we have tried to remove the drawback ranking by proposing the semantic ranking strategy.

In the following, we study the performance improvement of proposed Semantic query interpreter compared to that of WordNet based expansion, ConceptNet based expansion and the LabelMe query system. We study the improvements for randomly selected queries and the retrieval performance is measured using the following p@5 - precision in top-5 retrieved documents, p@10 - precision in top-10 retrieved documents, p@15 - precision in top-15 retrieved documents, p@20 - precision in top-20 retrieved documents, p@30 - precision in top-30 retrieved documents, p@100 - precision in top-100 retrieved documents.

Table 3.4: Precision comparison of the LabelMe system, WordNet based expansion, ConceptNet based expansion and the Proposed SQI at the different precision level

Technique	P@5	P@10	P@15	P@20	P@30	P@100
LabelMe Query System	0.513	0.427	0.331	0.3128	0.245	0.1301
WordNet based Expansion	0.785	0.728	0.5012	0.4264	0.3263	0.1921
ConceptNet Based Expansion	0.6873	0.6147	0.459	0.391	0.315	0.1721
Proposed Semantic Query Interpreter	0.884	0.82	0.5333	0.465	0.3818	0.2157

We have made the comparison between the WordNet and ConceptNet expansion in order to judge either the lexical expansion alone works well or the conceptual reasoning is also required. Moreover, investigate the efficiency of our proposed approach. The Table 3.4 and the Figure 3.11 shows the comparison of all these approach. The experimental results show that, WordNet-based query expansion methods brings little improvement, the average precision for

the top 5 retrieved images will increase from 51.3 % to 78.5%. While the ConceptNet based expansion increases the precision from 51.3% to 68.73 %. While the proposed semantic query interpreter will improve the precision of the system to 88.4%. The improvement is quite significant.

Sometimes most of the relevant information is available in the corpus but unfortunately cannot retrieve. It is due to the fact that either the annotation is not accurate or sometimes it is not tagged with the appropriate word even though it contain shares same semantic idea. This bottleneck has been removed by using the lexical expansion. It is the well-known fact that the image contains different object and these objects combine to constitute different semantic concepts. If the image is not tagged with the particular concepts then the images will not retrieved even though they contain the all the object that constitute that particular concept. This will decrease the performance of the system. The proposed framework attempts to rectify this problem by using the conceptual expansion. In practice, users may concern more about precision. If we remove the candidate concept module, from the proposed framework, it will significantly increase the recall but the precision will decrease. The reason could be that too much worthless information added to the query. The focus of our research is to bridge the semantic gap by retrieving the most of semantically relevant results. The Figure 3.11 depicts the comparison between the LabelMe query systems, WordNet based expansion, ConceptNet based expansion and the proposed semantic query interpreter. The Figure 3.11 shows the substantial performance improvement of the proposed system over the other ones. It is clear from the result that the lexical as well as the conceptual expansion is necessary to increase the performance of the IR system.

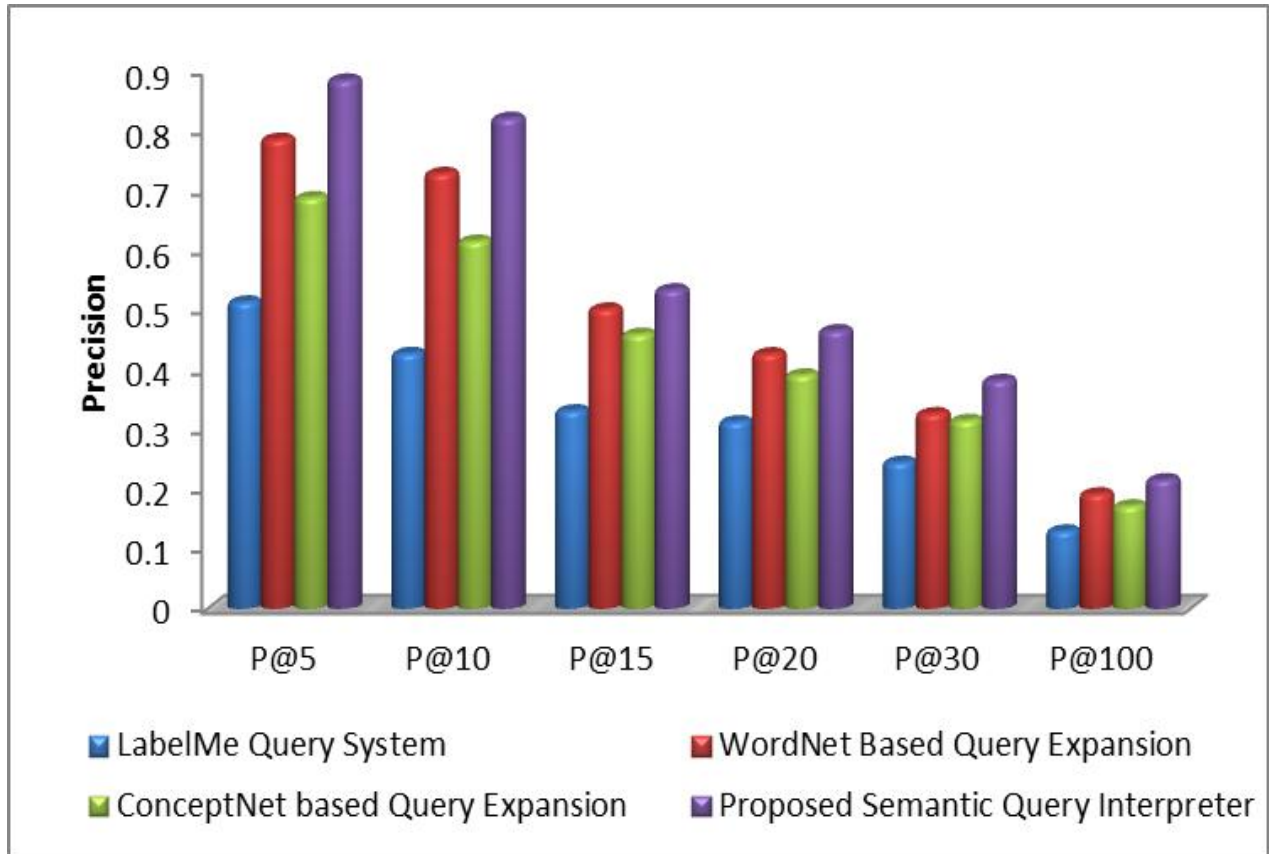


Figure 3.11: Precision comparison of the LabelMe system, WordNet based expansion, ConceptNet based expansion and the Proposed SQI at the different precision level

3.5 Chapter Summary

A large number of researchers actively investigated the image and video retrieval problem in recent years. Moreover, the researchers are long struggling to reduce the semantic gap. To overcome this gap, we proposed an automatic semantic query expansion technique called the semantic query interpreter (SQI), which automatically interpret the user query semantically as well as lexically. We used WordNet the online lexical knowledgebase to add the lexical paraphrase i.e. synonyms with the user query. In addition, we used the common sense knowledgebase i.e. ConceptNet to the common sense reasoning to the user query. From the list of the expanded concepts, some of the semantically relevant terms are selected.

For the retrieval and ranking, we used the Vector Space Model (VSM). We used the precision and recall as the evaluation factor to evaluate the performance of the proposed technique. We assumed the queries as a single word queries that may be single concept or multi

concept or it may be multiword multi concept queries. Compared to the state of the art keyword based query interpreter, which uses the lexical sources our method is much more accurate and precise. Compared to the content-based systems our method relies on the lexical as well as the common sense knowledgebases. SQI leads to improvement in search and retrieval. The experimental results have confirmed that the proposed approach showed the better performance than the previous ones.

We have already explored the significance of the user query in the information retrieval process and the detailed overview of all the query analysis techniques. In particular, several different query analysis techniques have been used we have particularly explored the query expansion by using the knowledgebases. We have explored our proposed semantic query interpreter that will extract the relevant information from the data corpus. But, unfortunately fails to rank the output based on the semantic relevancy between the query and the retrieved results. In the next chapter, we proposed the novel ranking strategy based on the semantic relevancy between the query and the document.

Chapter 04 .

SemRank: Ranking Refinement Strategy by using the Semantic Intensity

"My job is to make images and leave the decision making and conclusion drawing to other people."

Laurie Anderson

The ubiquity of the multimedia has raised a need for the system that can store, manage, structured the multimedia data in such a way that it can be retrieved intelligently. One of the current issues in media management or data mining research is ranking of retrieved documents. Ranking is one of the provocative problems for information retrieval systems. Given a user query comes up with the millions of relevant results but if the ranking function cannot rank it according to the relevancy than all results are just obsolete. However, the current ranking techniques are in the level of keyword matching. The ranking among the results is usually done by using the term frequency. This chapter is concerned with ranking the document relying merely on the rich semantic inside the document instead of the contents. Our proposed ranking refinement strategy known as SemRank, rank the document based on the semantic intensity. Our approach has been applied on the open benchmark LabelMe dataset and compared against one of the well-known ranking model i.e. Vector Space Model (VSM). The experimental results depicts that our approach has achieved significant improvement in retrieval performance over the state of the art ranking methods.

The remainder of this chapter is organized as follows. Section 4.1 included the introduction of the chapter. In Section 4.2 we surveyed the existing state of the art techniques used to rank the output. Section 4.3 introduces the proposed semantic ranking strategy for ranking the output based on the semantic intensity of the concepts with in the image. The description of the novel Semantic Intensity concept is also discussed in section 4.3. The Section 4.4 compares the SemRank approach with the existing selective retrieval approaches. Section 4.5 contains the summary of this chapter.

4.1 Introduction

An increasing immensity of procurable digital data online as well as offline has simulated recent research into digital data mining, data management, data filtering and information retrieval. Due to omnipresence of these data, acquisition becomes a bottleneck. So there is an urge for the efficient and effective retrieval techniques. The scarcity of rigid structure and the immense mass of information pose stupendous challenges to research community, and create several intriguing works for the organization and management of this colossal data for the academic community. The number of digital is continuously increasing.

Systems for retrieving specific objects in large scale image datasets have seen tremendous progress over the last five years [Jegou et al. 2008] [Nister et al. 2006] [Philbin et al. 2007]. It is now possible to retrieve objects from datasets of millions of images [Jegou et al. 2009a], [Perdoch et al. 2009] and performance on standard retrieval benchmarks has improved significantly [Chum et al. 2007], [Jegou et al. 2009b] [Perdoch et al. 2009] [Philbin et al. 2008]. Merely finding the relevant information is not the only task of IR systems. Instead the IR systems are supposed to retrieve the relevant information as well as rank or organize according to its degree of relevancy with the given query.

Information retrieval system intends to retrieve the relevant document according to the user request and then rank the output according to the relevancy order. The efficiency of the IR system relies heavily upon the ability of the system to prune the irrelevant information and return only relevant documents. The aspiration of a retrieval system is to find the relevancy between the data e.g. text, images, audio, video according to the user's requirement. Such information need is delineated in the form of a query, which usually corresponds to a bag of words. A 100% efficiency is impractical to accomplish because the user doesn't always provide the enough information need in the form of query. The traditional IR process bases on the string matching technique. The document's relevance was a function of the number of times each query word appeared in the document. Furthermore, the retrieved information items should be ranked from the most relevant to the least relevant. Unfortunately, these systems are not precise enough to retrieve the relevant information.

One key question in document retrieval is how to arrange documents based on their degrees of relevance to a query. This problem is effectively tackled by a ranking function which sorts the retrieved documents according to the relevancy degree. However, many of returned images are noisy, disorganized, or irrelevant. Ranking the IR result has gained much of the researcher attention. Traditionally, IR system uses the term frequency, inverse document frequency for defining the relevancy degree. Thus, it is possible to empirically tune the parameters of ranking functions [Salton et al. 1971]. Documents are judged within two categories: relevant and irrelevant.

Ranking is one of the intriguing issues in the IR systems. Ranking deals with sorting the retrieved results according to the relevancy with the given query. However, the result is the combination of the relevant as well as irrelevant data. The relevant document may have

different degree of relevancy. The relevancy degree is defined as a “function that determines the degree of semantic relatedness between the query and the retrieved results”. To achieve high precision the relevant document must be top ranked. Retrieving the relevant information without appropriate ranking is obsolete. The goal of ranking is to define an ordered list of documents such that documents similar to the query occur at the very first positions.

The main stumbling block in ranking is to classify which documents are relevant and which are irrelevant. The first key problem in the retrieval system is to find which of the documents are relevant. Research community has spent a lot of effort in sorting the results according to the user’s interest. Existing ranking techniques mostly rely on keywords to judge the relevancy of the data with the given query. The relevancy was defined in terms of number of times the words that is in the query appear in the document i.e. term frequency. The document with the greater term frequency will be top ranked. The current techniques mostly rely on the keyword matching technique for finding the document relevancy with the query. But unfortunately the keywords alone cannot capture the entire semantics behind the query. The word relevant means that retrieved documents should be semantically related to the user information need. The typical flow of IR model is shown in the Figure 4.1. The systems works well for simple object based queries. However, for the complex queries it’s trivial and leads to the poor retrieval performance. This is the one of the main handicap of the traditional IR systems.

In order to achieve effective retrieval performance, instead of using the keyword or text matching technique for ranking, it must be done by exploring the intended meaning behind the group of words or keyword. There is a demand for the system that can rank the output by considering multiple features instead of single feature for exploring the semantics.

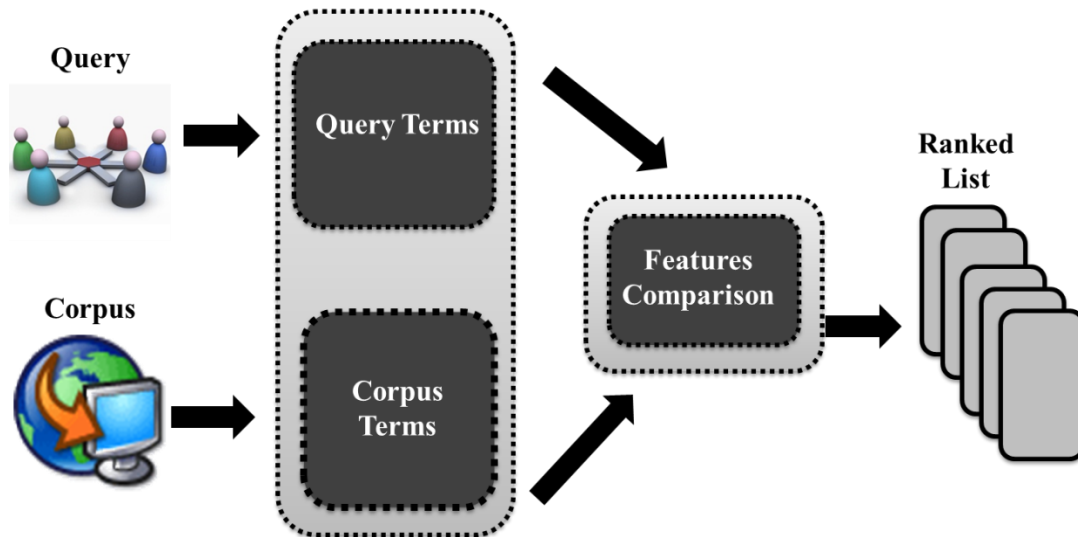


Figure 4.1: Typical IR Paradigm

To tackle this problem, we propose a novel ranking strategy known as SemRank, which rank the retrieved results on the basis of the Semantic Intensity (SI), which is the “concept dominance factor, the greater SI value of the image will have greater relevancy with the query”. The aspiration of this research is to enhance the quality of results obtained from a traditional information retrieval (IR) system by considering the semantic intensity of the relevant objects in the image instead of the frequency. The inspiration for the SemRank is that retrieving the relevant information is not a difficult task for the state of the art IR systems but ranking of the required document is still an open challenge. We focus on improving the precision of the IR system by ranking the documents on the semantic similarity between the retrieved document and the user query. Our method, rank the result on the basis of the semantic dominance of the concept in the retrieved images. Based on the semantic intensity the retrieved documents are then ranked.

4.2 State-of-the-Art Ranking Strategies

Extracting the relevant information from the corpus and then rank the information according to the relevancy order is one of the main functions in the IR systems. The ranking area in the data mining and IR has already been investigated by many researchers by assigning the calculating the frequency of the term in the query and the frequency of the query terms in the document, assigning the weights to the objects, etc. It is worth saying that a true ranking

strategy is the one in which the relevant documents come before the irrelevant and less relevant ones.

Over the past years, IR models, such as: Boolean, Vectorial, Probabilistic and Language models have represented a document as a set of representative keywords and defined a ranking function to associate the degree of relevancy for each document with its respective query [Baeza et al. 1999]. The language model in the IR works on the principal that relevancy between the document and the query can be judge by measuring how often the word in the query appears in the document. In general, these models are designed in an unsupervised manner and thus the parameters of the underlying ranking functions, if exist, are usually tuned empirically. However, the ability of these models to effectively rank relevant documents is still limited by the ability of the user to compose an appropriate query. Despite this fact, IR models flunk when the user's requirement is not explicitly defined in the user query. Some of the traditional retrieval models are discussed in the section below.

4.2.1 Retrieval Model

With a given query, an ideal IR system should only return relevant documents and ranks these documents in decreasing order of relevancy. The relevance degree between the document and the user query can be estimated by various IR models. A ranking strategy is used to compare the similarity between the modelled query and each respective modelled document. The documents that are more similar to the user are retrieved. In a nut shell, A retrieval model is an application of a mathematical framework to measure the distance between a document d and query q and the relevance of document d w.r.t query q . The traditional text based retrieval system categorize into two categories i.e. Boolean model and the statistical model [Korfhage. 1997], [Baeza et al. 1999].

4.2.1.1 Boolean Model

Boolean model is the classical and one of the oldest IR models. It is based on classical set theory and Boolean algebra [Sparck et al. 1997]. In the Boolean model, the query defines a Boolean logic for terms that are required if a document is to be retrieved. Queries are delineated by the set of keywords. The relevance of the result or document with the query can be judge by using the Boolean operators like AND, OR, NOT. Queries in the Boolean model

are formulated as Boolean expressions. These Boolean expressions have been made by using the three Boolean operators. The Boolean model uses the Boolean algebra.

The Boolean model works well for the simple queries while it fails for the complex queries. The Boolean model assigns equal weights to all the relevant documents. This results in the difficulty of ranking the most relevant one than the less relevant. Actually The Boolean Model is a simple retrieval model based on set theory and Boolean algebra that Documents are represented by the index terms assigned to the document. There is no degree of relevancy between the query and the documents. The weights of the retrieved document are either 0 or 1.

The bottleneck of the traditional Boolean model are string matching and may led to either too many or few retrieved documents. While the ranking of the retrieved document is a problem because all the retrieved documents have equal weights. But unfortunately all the retrieved documents don't have equal degree of relevancy. The user's queries are very difficult to transform into Boolean expression.

Extended Boolean Model

The Extended Boolean Model was proposed by Salton [Salton et al. 1983]. The extended model intends to rectify the bottlenecks of the Boolean information retrieval model. The Boolean model lacks the relevancy degree weights between the query and the document. Extended Boolean model is a type of hybrid model that contain all the features of Boolean model with a partial matching feature of vector space model. The vector space model weighting feature is used as a remedy to the relevancy degree problem of Boolean model.

The extended Boolean model can be considered as an integration of both the Boolean and vector space model. In the Extended Boolean model, a document is delineated as a vector like the vector space model.

The weight of term K_x associated with document D_j is measured by its normalized Term frequency and can be defined as:

$$W_{x,j} = f_{x,j} \times \frac{Idf_x}{\max_i Idf_x} \quad (4.1)$$

Where Idf_x is inverse document frequency.

The weight vector associated with document d_j can be represented as:

$$V_{d_j} = [W_{1,j}, W_{2,j}, \dots \dots \dots, W_{i,j}] \quad (4.2)$$

Considering the space composed of two terms K_x and K_y only, the corresponding term weights are W_1 and W_2 . Thus for query $q_{or} = (K_x \vee K_y)$, we calculate the similarity with the following formula.

$$Sim(q_{or}, d) = \sqrt{\frac{W_1^2 + W_2^2}{2}} \quad (4.3)$$

For query $q_{and} = (K_x \wedge K_y)$, we can use:

$$Sim(q_{and}, d) = 1 - \sqrt{\frac{(1-W_1)^2 + (1-W_2)^2}{2}} \quad (4.4)$$

4.2.1.2 Statistical Model

A statistical model represents the document and the user formulation in the form of mathematical equations. There are two main types of statistical model i.e. Vector space model and the probabilistic model. The statistical model computes relevancy between the user query and the document by using the term frequencies. The statistical retrieval models rectify some of the drawbacks of Boolean retrieval methods, but they have their own drawbacks.

The statistical model represents the query and the document as a bag of words. The similarity between the query and the document is calculated on the occurrence frequencies.

Vector Space Model

In contrast to the Boolean model, several other IR models have been proposed to estimate the relevance of a document. One such model is the vector space model [Salton et al. 1986], where both queries and documents are represented as vectors in the same space.

The Vector Space Model (VSM) is one of the well-known traditional retrieval models. That uses the bag of words approach for the text retrieval. The vector space model denotes the documents and the user queries as an n dimensional space of vectors. These dimensions are the terms employed to build an index to represent the documents [Salton et al. 1983] as shown in the Figure 3.7. The index terms are formed by using lexical scanning to identify the significant terms, by using some of the natural language processing techniques like morphological analysis for converting a word into its inflected form and then the occurrence of those inflected form is computed.

A Vector space model associate similarity measures between a query and a document by using direction and distance. The retrieval problem is then reduced to how close the document and the query in in the vector space as in Figure 4.2. This removes the binary limitation of the Boolean model by allowing partial matching between query and document.

The most widely used Vector model uses the cosine measure between the query and the document to determine the similarity [Baeza et al. 1999]. This is given by the following equation.

$$\text{Cosine } \theta_{d_j} = \text{Sim}(q, d_j) \quad (4.5)$$

$$\text{Sim}(d_j, q) = \frac{\vec{d}_j \cdot \vec{q}}{|\vec{d}_j| \times |\vec{q}|} = \frac{\sum_{i=1}^t w_{i,j} \times w_{i,q}}{\sqrt{\sum_{i=1}^t w_{i,j}^2} \times \sqrt{\sum_{i=1}^t w_{i,q}^2}} \quad (4.6)$$

Where the query vector is defined as $q = (w_{1,q}, w_{2,q}, \dots, w_{t,q})$ where t is the total number of index terms in the database. Similarly, the document vector is defined as $\vec{d}_j = (w_{1,j}, w_{2,j}, \dots, w_{t,j})$. Let $w_{i,q}$ and $w_{i,j}$ be the weight of the i th term in the query q or

document d_j respectively. The most famous way of calculating these terms weights is the term frequency inverse document frequency (*tf-idf*) method [Salton et al. 1998].

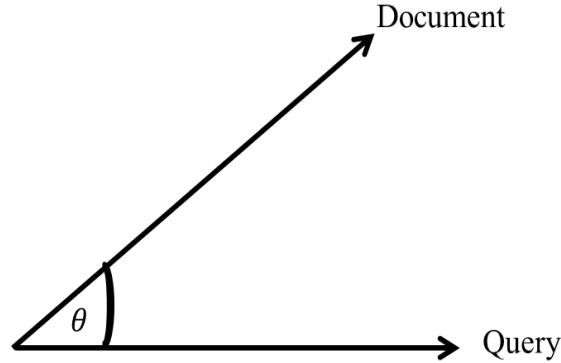


Figure 4.2: Vector Measure Cosine of theta

The *tf-idf* is characterized by two components. The first component is the term frequency factor *tf*. It is a measure of the frequency of the term in the document and is calculated by

$$f_{i,j} = \frac{\text{freq}_{i,j}}{\max_l \text{freq}_{l,j}} \quad (4.7)$$

Where $\text{freq}_{i,j}$ is the frequency of the term K_i in the document d_j . $F_{i,j}$ is the normalized frequency of the term K_i in document D_j (i.e. the number of times the term K_i is mentioned in the text document D_j). The $\max_l \text{freq}_{l,i}$ is calculated over all terms which are mentioned in the document. Because terms which appear in many documents are not as useful in distinguishing relevance, a second component for the method was introduced. It is the inverse document frequency or *idf* factor and is given by

$$\text{Idf}_i = \log \frac{N}{n_i} \quad (4.8)$$

Where N is the total number of documents and n_i is the number of documents which contain the K_i term. Combining the tf and idf components we get the $tf-idf$ term weight scheme which is given by:

$$w_{i,j} = f_{i,j} \times \log \frac{N}{n_i} \quad (4.9)$$

A high value in w is reached by a high term frequency (in the given document) and a low document frequency of the term in the whole collection of documents, as a result, the strategy tends to filter out common terms. Generally the vector method performs well and provides a similarity measure which can be used for ranking documents.

Limitation of the vector space model is that it focuses on the frequencies of the terms that are tagged with the image during annotation while doesn't consider the data inside the image. The vector space model relies on the text matching technique and is unable to consider the structural information. However, sometimes the cosine similarity between the query and the image is high but the semantic similarity between the image and query is low.

Probabilistic Model

Both the boolean and vector methods do not handle uncertainty or missing data. The probabilistic models also known as inference network calculate the relevancy of the document by using the probabilistic techniques. Rank the document by calculating the ratio between the relevant as well as the irrelevant one.

The Probability Ranking Principle (PRP) is the most widely used and accepted ranking criteria for the retrieval of documents [Robertson, 1977]. The classical probabilistic retrieval model [Robertson, 1977], [Robertson et al, 1976] of information retrieval has received recognition for being theoretically well founded. For the probabilistic retrieval models, we estimate two probabilistic models for each query: relevant class and non-relevant class. The probability ranking principle [Robertson, 1977] suggests ranking documents by the log-odds

ratio of being observed in the relevant class against the non-relevant class. Robertson [Robertson, 1977] has proved that ranking documents by the odds of being generated by the relevant class against non-relevant class optimizes the retrieval performance under the word independence condition.

Since we are assuming that each document is described by the presence/absence of index terms any document can be represented by a binary vector,

$$x = (x_1, x_2, \dots, x_n) \quad (4.10)$$

Where $x_i = 0$ or 1 indicates absence or presence of the i^{th} index term. We also assume that there are two mutually exclusive events,

$w_1 =$ document is relevant

$w_2 =$ document is non-relevant.

The probabilistic method can handle these problems through probability theory. The frequency of the terms is used to determine if a document is relevant. Bayes theory can then be used to determine if the document is relevant to the query. The two main probabilistic models are BIR model and Bayesian inference.

Binary Independence Retrieval (BIR)

The BIR model was developed by Robertson and Sparck Jones [Robertson et al, 1976]. BIR uses an interactive approach to calculate the similarity ranking value between the query and the document. The first part of BIR is to guess R , the set of document to be relevant.

Let $P(R | \vec{d}_j)$ be the probability that the document d_j is relevant to the query q .

Let $P(\bar{R} | \vec{d}_j)$ be the probability that the document d_j is not relevant to the query q .

The similarity $\text{Sim}(d_j, q)$ of the document d_j to the query q is defined by:

$$\text{Sim}(d_j, q) = \frac{P(R | \vec{d}_j)}{P(\bar{R} | \vec{d}_j)} \quad (4.11)$$

As explained by [Baeza et al 1999], after applying Bayes rule and assuming independence of index terms and taking logarithms the expression becomes.

$$\text{Sim}(d_j, q) \approx \sum_{i=1}^t w_{i,q} \times w_{i,j} \times \left(\log \frac{P(k_i|R)}{1 + P(k_i|R)} + \log \frac{1 - P(k_i|\bar{R})}{P(k_i|\bar{R})} \right) \quad (4.12)$$

Bayesian Inference

Bayesian inference models the retrieval process as an evidential reasoning process [Turtle et al. 1991], [Turtle et al. 1990]. It associates random variables with the index terms, documents and user queries. The documents are investigated individually as evidence and the degree of relevancy in the query is calculated and ranked for each document (i.e. calculate $P(q | \vec{d}_j)$). The documents that return the highest degree of belief in the query are the documents that are retrieved by the system as the relevant documents for that query.

The basic inference network to model information retrieval is given in the Figure 4.3. The document is the root node of the network. Each document is made of index terms and has a causal relationship with them. An arc from the document to the index term implies that there is a causal relationship between that document and index terms and that the observation of one causes a change in belief of other.

Probabilistic model assumed to rank the document according to the probability of relevance to the given query. The probabilistic model uses the probability ranking principle [Belkin et al. 1992]. The relevancy between the relevant and irrelevant document is measured using statistical distribution. There is an uncertainty between the users need.

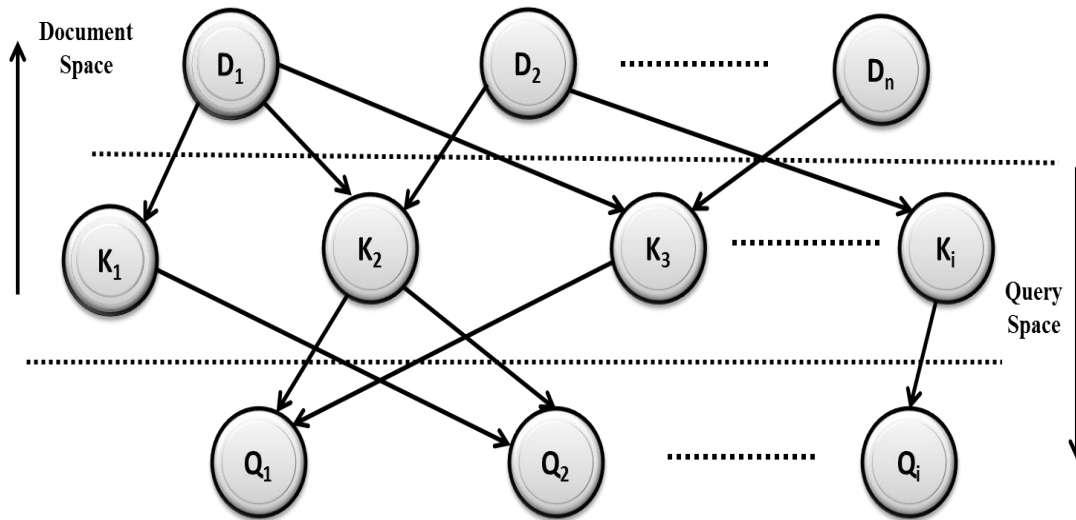


Figure 4.3: Bayesian Inference IR Model

The problem of effectively estimating the relevant and non-relevant models remains a major obstacle in the practical applications of the probabilistic retrieval models.

Latent Semantic Indexing

The vector space model has been extended by incorporating several statistical and AI techniques to remove the bottleneck of the VSM. The different techniques are incorporated to exploit in finding the association between the term and the document, one of such type of technique is the Latent Semantic Indexing (LSI). This technique doesn't work on the simple text matching technique it will even retrieve such type of an output in which the query and the document doesn't have any common word. This technique calculates the association between the term and the document and then used it for further retrieval process. The LSI apprehend the in depth analysis for finding the correlation between the terms and the process is also automatic. The LSI represent the term in the reduced dimension space which is the key difference between the vector space model and the LSI technique. In VSM the term weighting technique and the relevance feedback substantially boost the VSM performance over the LSI.

Over the past years, IR models, such as Boolean models, vector models, probabilistic models, and language models, have represented a document as a set of representative keywords and defined a ranking function to associate a relevance degree with a document and a query

[Baeza et al. 1999]. In general, these models are designed in an unsupervised manner and thus the parameters of the underlying ranking functions, if exist, are usually tuned empirically.

Recently, much of the work has been done in the retrieval models with relevance judgement like learning to rank method. The learning to rank is a type of supervised learning based method for learning a ranking function from the training corpus automatically [Burges et al. 2005], [Cao et al. 2007], [Crammer et al. 2002], [Freund et al. 2003] [Herbrich et al. 2000].

Learning to Rank

Learning to Rank is an effective ranking technique for the information retrieval and data mining. Learning to rank exploits the machine learning technique for the generation of the ranking function. The typical learning to rank paradigm is shown in the Figure 4.4 below.

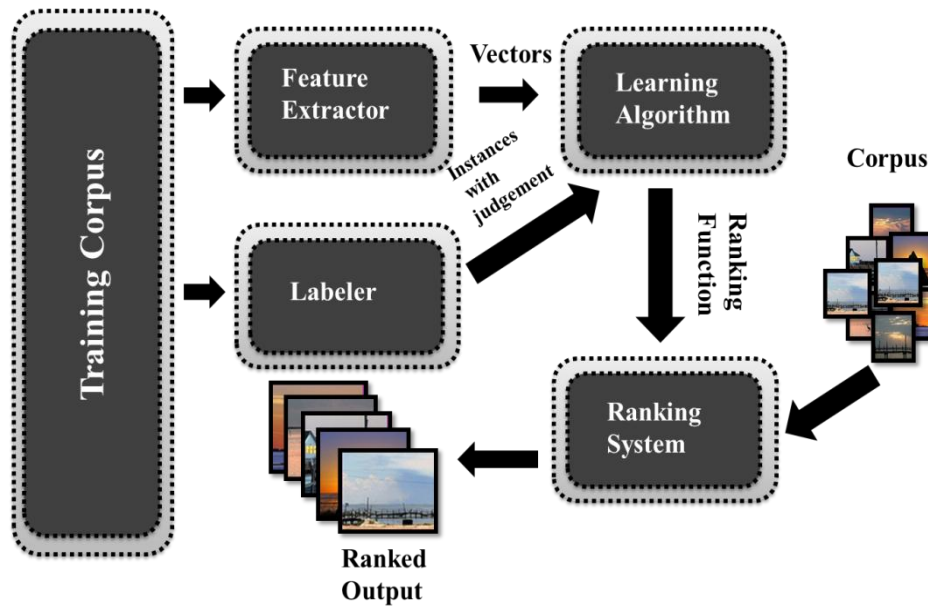


Figure 4.4: Typical Learning to Rank Paradigm

General Learning to Rank Process

There are two main steps of the Learning to Rank function i.e. training and the test. Given a user query is the collection of the terms

$$Q = \{ q_1, q_2, \dots \dots q_n \} = \cup_{i=1}^n q_i \tag{4.13}$$

And the document collection $D = \{d_1, d_2, \dots, d_m\}$,

$$D = \bigcup_{j=1}^m D_j \quad (4.14)$$

The training corpus is created for the each query and the document $(q_i, d_j) \in Q \times D$, which will be later used for deriving the relevance judgement between the d_j and q_i . The relevance judgement can be of various types i.e. a label i.e. relevant or irrelevant, a rating e.g. definitely relevant, possibly relevant, irrelevant, an order e.g. for a specific query like q_i the document d_j is at the K^{th} position, or a score e.g. $Sim(q_i, d_j)$ that delineate the degree of relevance between the query and the document.

The inputs to the LTR consist of training instances, feature vectors, and the corresponding relevance judgments. The output is a ranking function, f , where $f(q_i, d_j)$ is supposed to give the “true” relevance judgment for q_i and d_j . During the training process, the learning algorithm attempts to learn a ranking function such that a performance measure like classification accuracy, error rate, Mean Average Precision etc. with respect to the output relevance judgments can be optimized. In the test phase, the learned ranking function is applied to determine the relevance between each document d_j in D and a new query q .

Categorizes of Learning to Rank

As distinguished by [Cao et al. 2007] and [Cao et al. 2006], and fall into three categories:

- Point-wise approach
- Pairwise approach
- List-wise approach

In the point-wise approaches, each training instance is associated with a rating. The learning is to find a model that can map instances into ratings that are close to their true ones. A

typical example is PRank [Crammer et al. 2002], which trains a Perceptron model to directly maintain a totally-ordered set via projections.

The pair-wise approaches take pairs of objects and their relative preferences as training instances and attempt learning to classify each object pair into correctly-ranked or incorrectly-ranked. Indeed, most existing methods are the pair-wise approaches, including Ranking SVM [Herbrich et al. 2000], RankBoost [Freund et al. 2003], and RankNet [Burges et al. 2005].

Finally, the list-wise approaches use a list of ranked objects as training instances and learns to predict the list of objects e.g. ListNet.

Several different learning to rank techniques have been proposed in the literature [Burges et al. 2005], [Cao et al. 2007], [Herbrich et al. 2000], [Nallapati. 2004] [Xu et al. 2007]. They mainly differ in terms of the loss function used to guide the learning process [Liu, 2009]. Much effort has been placed on the development of ranking strategies including RankBoost [Freund et al. 2003] [Rudin et al. 2009], RankNet [Burges et al. 2005], ListNet

[Cao et al. 2007] , Page Rank [Jing et al. 2008], Vector Space Model (VSM)[Salton et al. 1975], iRANK (Interactive Ranking) [Furu et al. 2009], fRank [Ming-Feng et al. 2007] ,PPRank (Predict Popularity rank) [Crammer et al. 2002] , Ada Boost (Adaptive Boosting) [Jun et al. 2007] , HostRank [Xue et al. 2005], topical PageRank [Nie et al. 2006], Quantum Probability Ranking Principle (QPRP) [Zuccon et al. 2010], LexRank [Erkan et al. 2004] etc.

Several learning to rank techniques, such as RankNet [Burges et al. 2005], RankBoost [Freund et al. 2003], and Ranking SVM [Joachims. 2002], learn a ranking function for a specific task by optimizing a selected loss function. However, for these, the loss function may only be loosely related to standard IR evaluation measures. This could result in the obtained ranking function deviating from the target evaluation measure and producing poor retrieval performance.

To avoid this issue, Xu & Li proposed the AdaRank algorithm, which is a boosting-based method and employs an exponential loss function based on IR evaluation metrics [Jun et al. 2007]. Similar to the AdaBoost algorithm [Freund. 1995], AdaRank can focus more on the difficult queries during the construction of a ranking function.

RankGP, based on genetic programming (GP) is developed to learn a ranking function by integrating following three features, including content features, structure features, and

query-independent features. RankGP represents a potential solution (i.e., a ranking function) as an individual in a population of GP [Yeh et al. 2007]. The LexRank is a ranking strategy using the graph based approach for computing the relevancy degree for the textual information retrieval [Erkan et al. 2004].

RankBoost uses the boosting approach for combining the preferences. RankBoost is another boosting technique based on AdaBoost [Rudin et al. 2009]. RankBoost is suitable for ranking the results with a certain criteria and is used to document ranking in Information Retrieval (IR). Boosting refers to a general method of building a single strong learner by repeatedly constructing a weak learner with respect to a specific distribution and adding it to the strong learner [Freund et al. 1999]. A weak learner is a type of a classifier which is only slightly correlated with the true classification. While a strong learner is a classifier that is arbitrarily well correlated with the true classification. Based on the boosting technique, [Freund et al. 2003] proposed an efficient learning technique, called RankBoost. Similar to other boosting algorithms, the RankBoost algorithm builds a document ranking function by combining several “weak” rankers of a set of document pairs.

Burges et al. (2005) proposed the RankNet algorithm, which learns a retrieval function by employing a probabilistic cost function on a set of pairs of training examples. RankNet uses the gradient descent algorithm to train the neural network model for ranking [Burges et al. 2005]. ListNet uses the probabilistic approach for ranking. It uses a list wise approach and used objects as an instance. ListNet [Cao et al. 2007] introduces a probabilistic-based list-wise loss function for learning. Neural network and gradient descent are employed to train a list prediction model. QPRP has been proposed to remove the document dependency problem of Probability Ranking Principle (PRP). The main drawback of the Probability Ranking Principle (PRP) is that it does not cater for dependency between documents. Using Quantum Theory within IR was originally proposed by van Rijsbergen [Rijsbergen et al. 2004], and has been subsequently developed in a number of ways [Melucci. 2008], [Piwowarski et al. 2009], [Hou et al. 2009], [Flender et al. 2009], [Rosero et al. 2009]. Recently, the extension of the Probability Ranking principal has been proposed to remove the document relevancy bottleneck, the model is known as the Quantum Probability Ranking Principal (QPRP). QPRP apprehend the relevance dependency between the documents by using “Quantum Interference” [Zuccon et al. 2009], [Zuccon et al. 2010], [Khrennikov. 2009].

Support Vector Machine (SVM) has been widely and effectively used for binary classification in many fields. For instance, in information retrieval (IR), SVM is used to classify documents [Nallapati. 2004] or to estimate whether the most frequent terms in the pseudo-feedback documents are useful or not for query expansion [Cao et al. 2007]. However, SVM cannot indicate the ranking sequence among multiple objects (e.g. documents) because it is a binary classifier.

A learning algorithm on the basis of the support vector machine (SVM) has been developed for ranking known as Ranking SVM [Cao et al. 2006]. Support vector machine (SVM) is a machine learning technique for ranking. Chapelle et al propose new methods for optimizing the Rank SVM training by using primal Newton method [Chapelle et al. 2009]. Different researchers are trying to optimize the state of the art techniques like [Agarwal et al. 2010] proposes an algorithm to remove the hinge loss on SVM. A new learning strategy has been proposed known as learning to rank (LTR) it uses several document features [Peng et al. 2010]. LTR selects appropriate ranking function for each query. The inspiration of LTR is, it is not necessary that a ranking function which works well for the single query will work well for all the other set of queries. Different ranking function suits different queries. The ranking fusion technique has been also used to make the significant improvement in retrieval of hand writing recognition systems [Pena et al. 2010]. All these approaches aim at producing the efficient ranking algorithm in order to optimize the retrieval performance.

IRank is an interactive ranking framework which uses the “rank-learn-combine” [Furu et al. 2009]. There are two types of IR approaches based on the sequence of rank-learn-combine e.g. combine the features first and then use it for ranking while the second approach uses the ranking aggregation approach to fuse all the ranking results. The second approaches are also known as ensemble ranking, the most popular implementation of which is to linearly combine the ranking features to obtain an overall score which is then used as the ranking criterion. However, both of the above-mentioned “combine-then- rank” [Dwork et al.2001] and “rank-then-combine” [Pickens et al. 2008] approaches have a common drawback of not effectively utilizing the information supplied by different ranking function and ignore the interaction between the functions prior to the combination.

Okapi BM25 is another probabilistic ranking function widely used by search engines to rank documents according to their relevance to a given query. It uses the bag of words model

for ranking the documents. The BM25 approach fails to find the interrelationship between the terms within the documents [Stephen et al. 2009]. Yuchi proposes the Probabilistic hypergraph ranking for the images [Yuchi et al. 2010]. The images are delineated by vertices of the graphs. The probabilistic hypergraphs are used to exploit the relevance between the images. They propose a transductive learning problem for content based image retrieval.

Another concept of focused retrieval has been proposed by the research community for the textual passage retrieval, element retrieval and question answering systems [Shafiq et al. 2007]. Focused Retrieval (FR) is relatively a new area of research which deals with retrieving specific information to the query rather than state of the art information retrieval systems (search engines), which retrieve documents. Pehcevski et al. [Pehcevski et al. 2010] proposes a ranking function for Wikipedia known as the entity ranking system. They utilize the known categories, the link structure of Wikipedia (Wikipedia category score), as well as the link co-occurrences (link score), document score to retrieve relevant entities in response to the query. The concept of focused retrieval has been investigated by many researchers in the textual domain [Paris et al. 2010], [Kaptein et al. 2010], [Andrew et al. 2010], [Arvola et al. 2010].

In the above many approaches to ranking are discussed which have been used for the text as well as the image retrieval. Although the area related to the text retrieval are matured but image retrieval is worth investigating. Most of these techniques retrieved and ranked the images on the basis of visual similarity. But still the precision of the system is low because the visual similarity is not the semantic similarity.

4.3 Proposed Semantic Ranking Framework

With the development of IR, people find that the conventional IR models could not satisfy with practical requirements such as high precision and low human consumption. And machine learning methods can be helpful to improve models' performance. Evaluation results from various experiments indicate that current information retrieval methods are effective to retrieve relevant documents, but they have severe difficulties to generate a pertinent ranking of them.

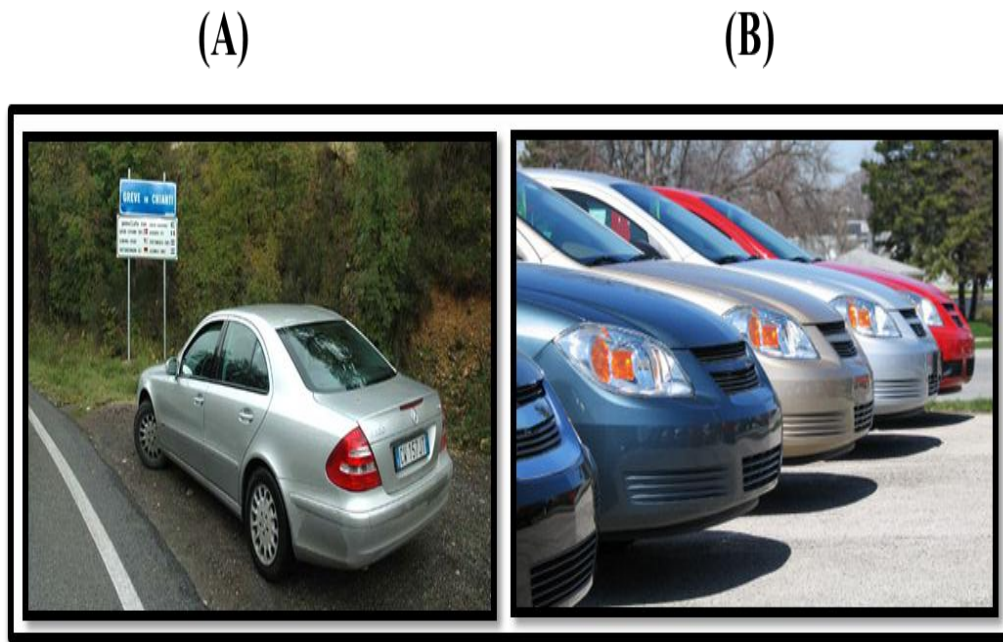


Figure 4.5: Both A and B figure represents images of the car. The frequency of the car in the image B is greater than the image A. But the image A depicts the car concept more clearly. Hence image A has a greater relevancy degree than image B even though image B has greater frequency.

In this study, we employ new methods for semantic ranking for the image search and retrieval systems. Our line of research focuses on the ranking on the basis of the semantics not on the basis of frequency comparison between the query and the documents. The frequency doesn't depict the semantics inside the data. Let's consider the simple example shown in the Figure 4.5. We envision that in order to achieve the effective retrieval performance semantic similarity should be consider instead of the visual or the textual similarity between the query and the information obtain from the tags attach with the image i.e. annotation. Based on this institution, we exploit the Semantic Intensity (SI) for ranking the images.

4.3.1 Semantic Intensity

The Semantic Intensity can be defined as the “concept dominancy factor with in the image”. While image is the combination of different objects, these objects constitute to form different semantic idea. Different combination of objects depicts different concepts. The images can depict different semantic idea simultaneously. However, these semantic ideas have

different dominance degree. Some of the ideas in the image are more dominant than the other as shown in the Figure 4.6.

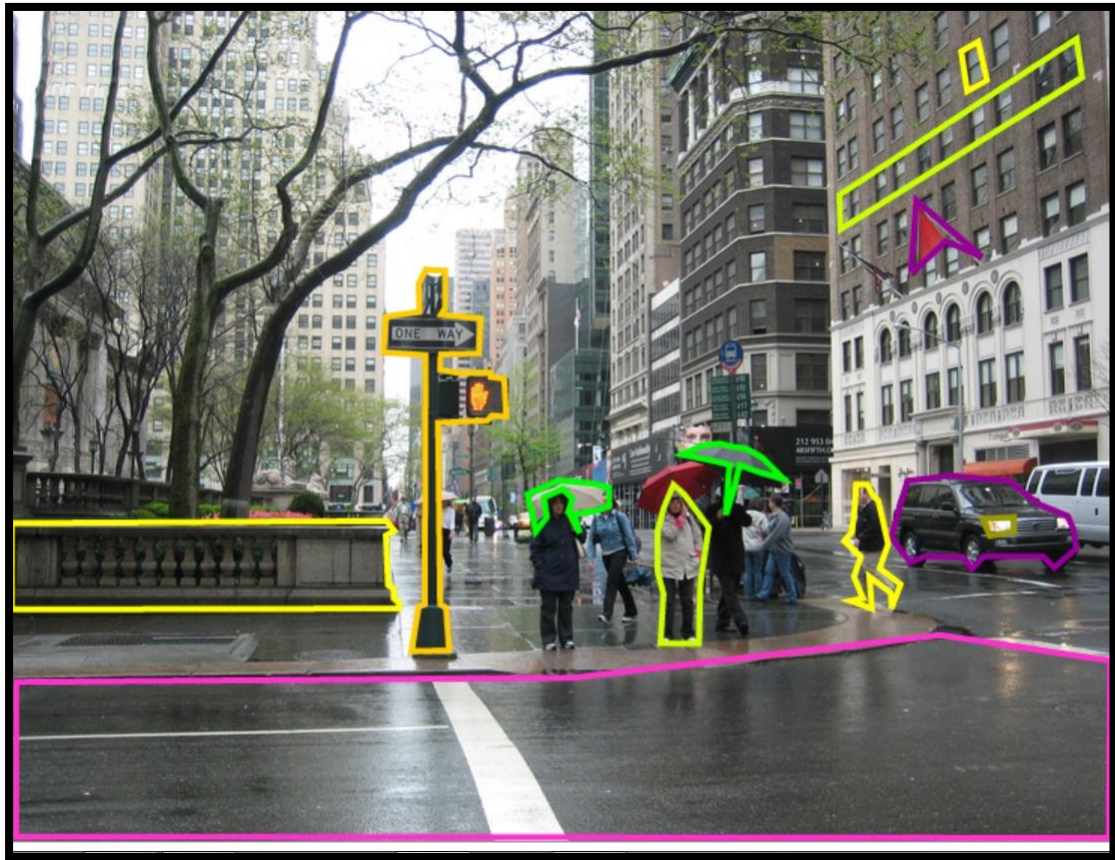


Figure 4.6: The image is taken from the LabelMe dataset. Image depicts a list of concepts like road, vehicles, signs, buildings, sky, trees, umbrella, buildings, street, cross walk, highlight, flags etc. and some hidden concept like rain. Among all the concepts some are more dominant like street, building etc.

We have implemented a semantic Intensity concept on the LabelMe dataset which is open source dataset available for academic and research, the object in the LabelMe dataset images is represented by a set of points known as polygon. The polygon may be either a regular or irregular polygon.

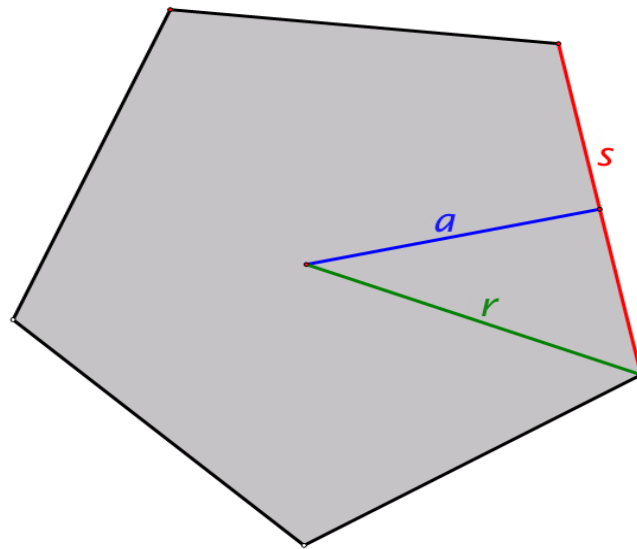


Figure 4. 7: (a) Regular Polygon

The area A of a regular n -sided polygon having side s , apothem a , and circum-radius r is given by

$$A = \frac{1}{2} nsa = \frac{1}{4} ns^2 \cot \frac{\pi}{n} = na^2 \tan \frac{\pi}{n} = \frac{1}{2} nr^2 \sin \frac{2\pi}{n} \quad (4.15)$$

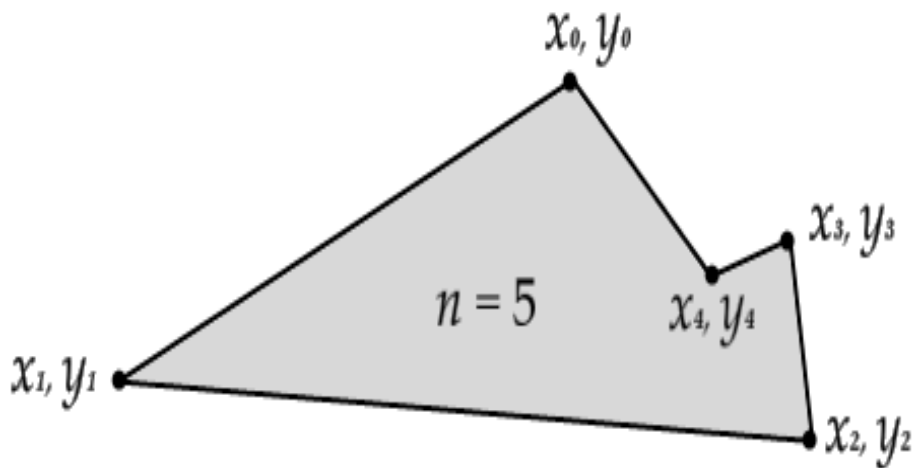


Figure 4.7:(b) IRRegular Polygon

While area of the irregular polygon is

$$A_{poly} = \frac{1}{2} \sum_{i=0}^{n-1} (x_i y_{i+1} - x_{i+1} y_i) \quad (4.16)$$

The A_{poly} is the object dominance degree with in the image. The greater the A_{poly} value greater will be the object dominance degree.

$$A_{poly} = OD \quad (4.17)$$

The Concept Dominancy for the given object can be calculated as

$$CD = \frac{A_{poly}}{I_s} = \frac{OD}{I_s} \quad (4.18)$$

Where $I_s = h * w$, represents size of the image.

The Semantic Intensity (SI) for a particular concept relevant to the given query is calculated by the following expression.

$$SI = R.Q'.SS * CD \quad (4.19)$$

Where $R.Q'.SS$ are the expanded query terms with their appropriate semantic similarity value of a particular query. Analogy to the image the query is also the combination of different

concepts. Some of the concepts in the query are more dominant than the other. We have calculated the dominance level of the different concepts in the query by using the Semantic Query Interpreter. The Semantic Similarity value retrieved with the selected expanded terms depicts the dominance degree of the concepts with the query.

Let's consider the scenario of the simple query car. The car query processed by the Semantic Query Interpreter module and the results are retrieved by using the VSM, we have taken some of the top ranked results as shown in the Figure 4.8.

In the Figure 4.8 it is clear that semantic query interpreter retrieve most of the relevant results, however the ranking is not appropriate some of the less relevant results come before the more relevant ones. Let's consider the scenario of the third and sixth one. The sixth one is more relevant to the query than the third one. The semantic intensity of the sixth image is greater than the third image as shown in the Figure. Hence the greater the SI value of the image relevant to the query, the higher is the rank.



Figure 4.8: 'CAR' Query Output using VSM and SQI



Figure 4.9: Semantic Intensity of the images

The proposed model works on the principle of the semantic Intensity for the ranking of a result. Let the initial user query, which is the combination of different keywords be applied on the Semantic Query Interpreter for expanding the query lexically and conceptually.

$$Q = \{t_1, t_2, \dots, t_n\} = \bigcup_{i=1}^n t_i \quad (4.20)$$

Where Q is the query with set of terms 't'.

$$Q' = \{(t, SS)_1, (t, SS)_2, \dots, (t, SS)_n\} = \bigcup_{i=1}^n (t, SS)_i \quad (4.21)$$

While Q' is the expanded or enhanced query with their semantic similarity values.

After the expansion of the user query from the SQI, the query is applied on the corpus C. The system must return a subset of images C' from the corpus C, where C is set of images with their annotation represented by the following equation. The images are represented by M.

$$C = \{M_1, M_2, \dots, M_m\} = \bigcup_{j=1}^m M_j \quad (4.22)$$

Where M_1, M_2 is the number of images in the corpus.

Where $M = \{O_1, O_2, \dots, O_z\} = \bigcup_{k=1}^z O_k$, then equation (4.22) become

Where O_1, O_2 is the number of objects with in the images

$$C = \{M_1, M_2, \dots, M_m\} = \bigcup_{j=1}^m (\bigcup_{k=1}^z O_k)_j \quad (4.23)$$

The returned results are then passed to the SemRank module to rank the output on the basis of the Semantic Intensity rather than the frequency or visual similarity. The over-all algorithm of the SemRank is given below.

Proposed Algorithm 4.1: SemRank

Input: $Q' \rightarrow \bigcup_{i=1}^{t'} (K', SS)_i$
 $C \rightarrow \bigcup_{j=1}^n (\bigcup_{x=1}^m O_x)_j$

Output: $R' = \text{SemRank Result}$

Method:

// applying enhance query on the corpus C
 $R \leftarrow Q' \cap C$, where $R \in Q' \cap C'$ and $C' \leq C$

For each C' . Image in R . C'

For each C' . Image . Object in R . C' . Image

//Calculate object dominancy OD for each object

$$OD \leftarrow \frac{1}{2} \sum_{i=1}^{n-1} (x_i y_{i+1} - x_{i+1} y_i)$$

//Calculate the Concept Dominancy of each concept tag with the object in the image

$$CD \leftarrow \frac{OD}{I_s}, \text{ Where } I_s = H \times W \text{ of the image}$$

// calculate the Semantic Intensity (SI) for concepts relevant to the query

$$SI \leftarrow R \cdot Q' \cdot SS \times CD$$

// Where $R \cdot Q' \cdot SS$ is the semantic similarity value of each term

// Calculate netSI at $R.C'$. Imagelevel

$$R.\text{SetSI} \leftarrow \sum_{i=1}^n (SI)_i$$

// Where n is the number of concept tag with object per imageSort the result in descending order

$$R' \leftarrow \text{Sort}(R.\text{netSI}, \text{Descending})$$

4.4 Experimental Study

Measuring relative performance of information retrieval (IR) systems is essential for research and development and for effectiveness. In this chapter, we have presented many different retrieval techniques for building IR systems. A natural question arises on how to evaluate the performance of an IR system. The evaluation of an IR system is the process of investigating the effectiveness of the system in terms of the user satisfaction [Voorhees. 2001].

By submitting a query to an IR system, a set of documents in the collection is returned. With the relevance assessments set for this query, the performance of the IR system can be evaluated by examining whether each returned document is relevant to the query. The conventional way of measuring the quality of the results returned by a system in response to a query is to use precision and recall.

A comprehensive empirical performance study, using both Vector Space Model and SemRank has been made. The experiments were conducted on some of the categories from the LabelMe 31.8 GB dataset. Which contain total of 181,983 images, 56,943 annotated images and 125,040 images are still not annotated. The study is made with the objective to test the result of the proposed method against the traditional IR model i.e. VSM. Several experiments were conducted using different set of queries like keyword based queries which may either single concept or multi-concept and multi-word queries i.e. multi-word multi concept etc.

In order to provide an objective comparison of the retrieval performance of the algorithms, we used the quantitative evaluation criterion, the precision and recall graph. Retrieval precision is defined as the proportion of the images among all those retrieved that are truly relevant to a given query, recall is defined as the proportion of the images that are actually retrieved among all the relevant images to a query.

Specifically, a comparative analysis has been made among the SemRank, Vector Space Model and the simple LabelMe query engine. The Table 4.1 shows the comparison of all the three techniques and reveals that the SemRank significantly outperforms the commonly used Vector Space Model.

P @ n (Precision at Position n)

For a given query, its precision of the top n results of the ranking list is defined as Eq.

$$P @ n = \frac{\text{\# of relevant documents in the top } n \text{ results}}{n}$$

Table 4.1: Comparison of the LabelMe system, Vector Space Model and SemRank at different precision values.

Technique	P@5	P@10	P@15	P@20	P@30	P@100
LabelMe Query System	0.513	0.427	0.331	0.3128	0.245	0.1301
Vector Space Model	0.884	0.82	0.5333	0.465	0.3818	0.2157
SemRank	0.972	0.895	0.725	0.615	0.473	0.358

We have produced ranked results for three different query categories for the test. We judge our rankings qualitatively by showing some highly ranked results for the three methods i.e. Proposed SemRank, traditional Vector Space Model and the LabelMe system. We also judge our results quantitatively according to the results of a user study which compares the goodness of our top ranked images to top ranked images ranked using the two alternative methods. The Table 4.1 depicts the comparison of the LabelMe system, VSM and the proposed SemRank at different precision values. At P@5 there is a drift in the precision of 7.9% from VSM to the SemRank. It is due to the fact that most of the top retrieved results are relevant. As the SemRank judge the relevancy in terms of Semantic Intensities of the Concept with in the image and the Semantic Intensities of the concepts with in the query instead of the occurrence frequency. The Semantic Intensities of the concepts with in the query is triggered by calculating the Semantic Similarity between the query term and the expanded terms. The Semantic Similarity computation in the Semantic Query Interpreter aims to evaluate that among the expanded terms which ones are more relevant than the other. The results of the SemRank for other precision level also depict the substantial improvement over the LabelMe and VSM.

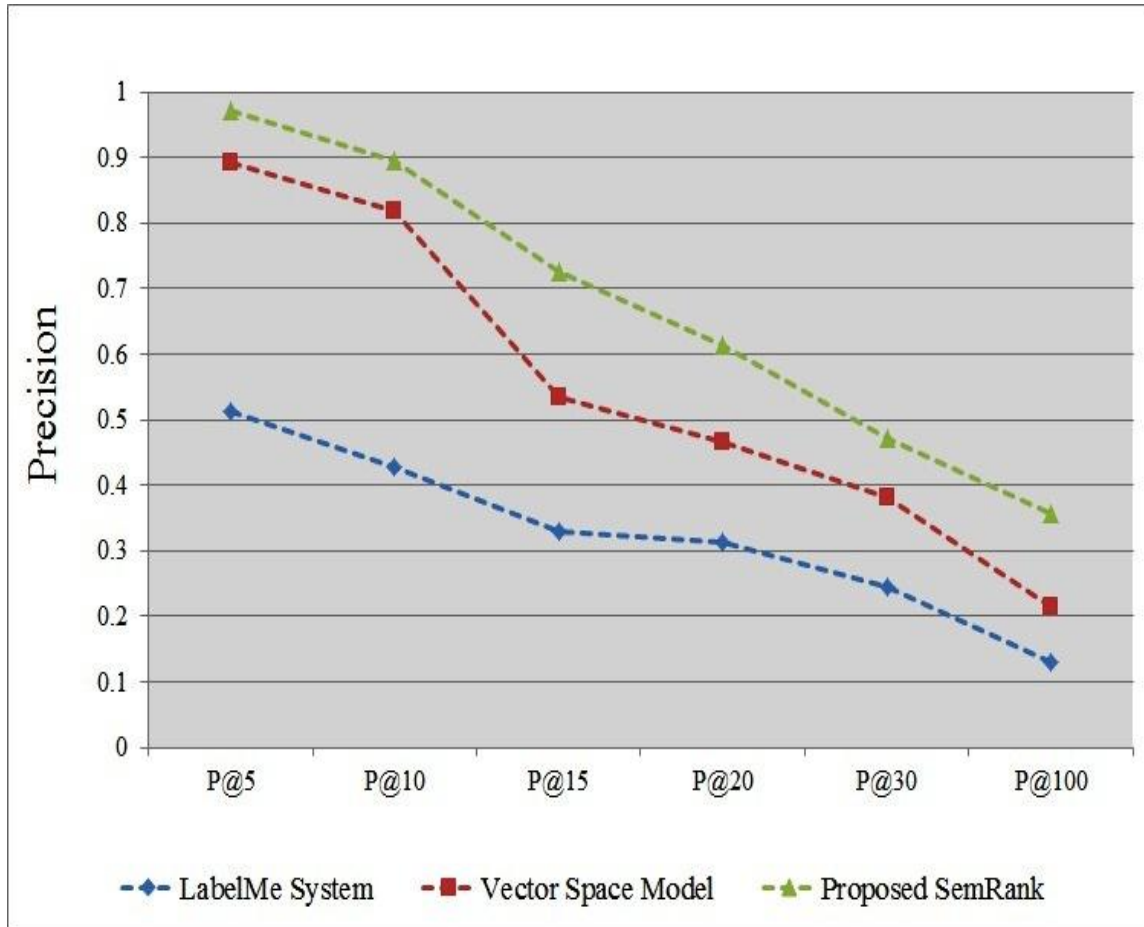


Figure 4. 10: Comparison of the LabelMe system, Vector Space Model and SemRank

The Figure 4.10 depicts the comparison results of our experimental study. The randomly selected queries of all the three predefined category is applied on all the three models and the top retrieved images are selected in order to test how precisely and accurately each model rank the images. The above numbers correspond to the precision at different levels i.e. P@5 (Precision for the top 5 retrieved images), P@10 (Precision for the top 10 retrieved images), P@15 (Precision for the top 15 retrieved image), P@20 (Precision for the top 20 retrieved images), P@30 (Precision for the top 30 retrieved images) and P@100 (Precision for the top 100 retrieved images). As can be seen from the random results, that the SemRank out performs at all the precision level. Our proposed ranking strategy that incorporates semantic intensity information performs improved results than the LabelMe system and the VSM. The significance in the outcome is due the fact that our proposed technique judges the degree of relevancy between the user query and the retrieved results on the basis of the semantic intensity. It is the well-known fact that an image is the combination of different objects and these combines to constitute different semantic idea. Within a single image some of the

concepts are more dominant than the other. Even though sometimes an image contain the same objects but constitute different semantic concepts like the images of street and the car park mostly contain same objects but the it is the dominancy factor that differentiate the images of the car park with the street like both the images may contain sky, road, tree, vehicles, people and buildings etc. The vector space model outperforms the LabelMe system because the LabelMe system simply takes the query terms and match with the terms tagged with the images and retrieved the results. While in the Figure 4.10 the VSM approach takes the output of our proposed Semantic Query Interpreter (see chapter 3) as an input for retrieving and ranking the retrieved results. Therefore the output of the VSM is better than the LabelMe system even though the VSM relies on the number of occurrences of terms of user queries occurs in the corpus. The degree of relevancy can be judged on the basis of frequencies. While proposed SemRank technique enhanced the performance Semantic Query Interpreter significantly. From the Figure 4.10 we can judge that incorporating the Semantic Intensity or the concept dominancy makes a clear, obviously useful difference for our system.

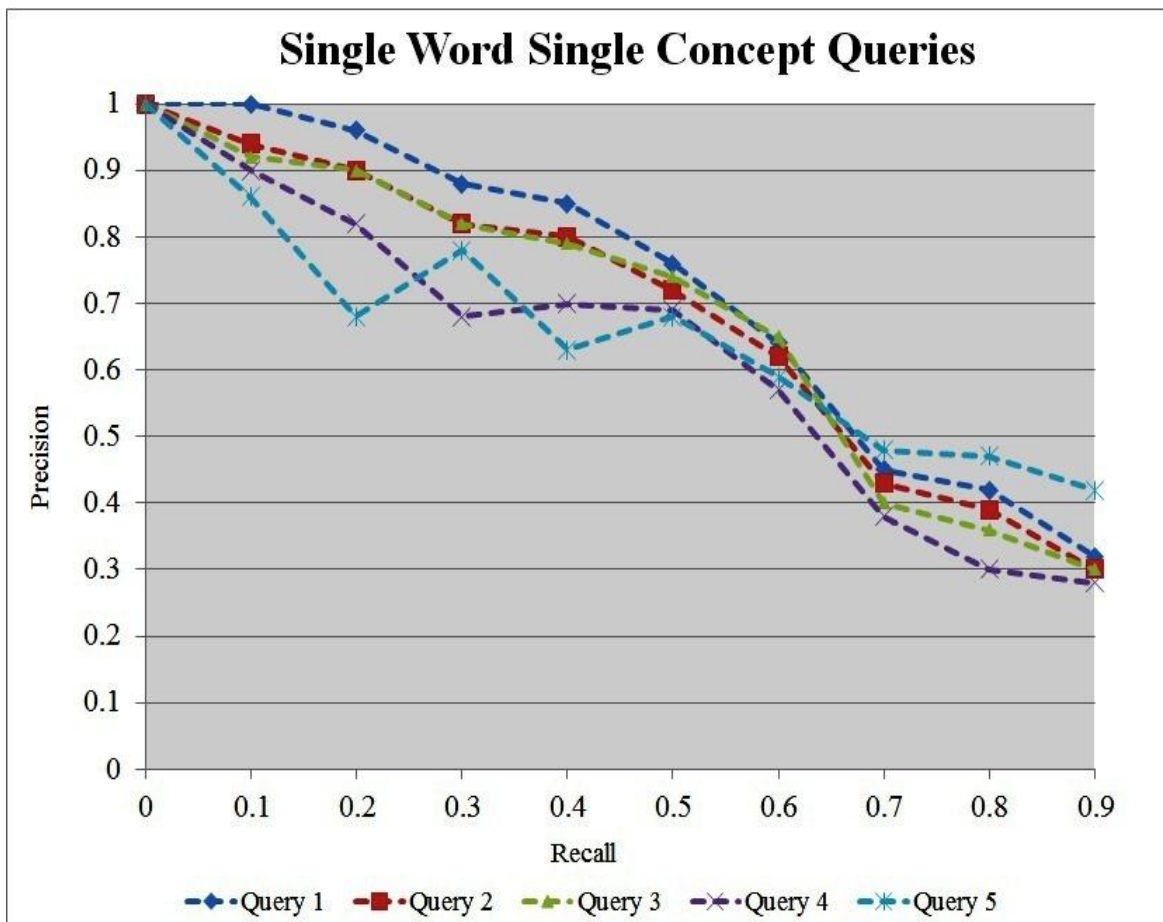


Figure 4.11: Precision Recall curve of Single Word Single Concept Queries

The Figure 4.11 depicts the precision recall curve for the five randomly selected single word single concept queries. We have applied the queries that are already expanded by the semantic query interpreter on the SemRank model. The list of five single word single concept queries is car, building, tree, sky and house. The outcome of the Semantic Query Interpreter serves as an input into the SemRank. The output of the Semantic Query Interpreter is the original query terms and expanded terms along with their semantic similarity value. The semantic similarity values helps to compute the semantic intensity of the concepts with in the query. The variation in the outcomes of various types of single word single concept is due to the annotation. The SemRank enhance the efficiency of the Semantic Query Interpreter by ranking the images on the basis of semantic relevancy.

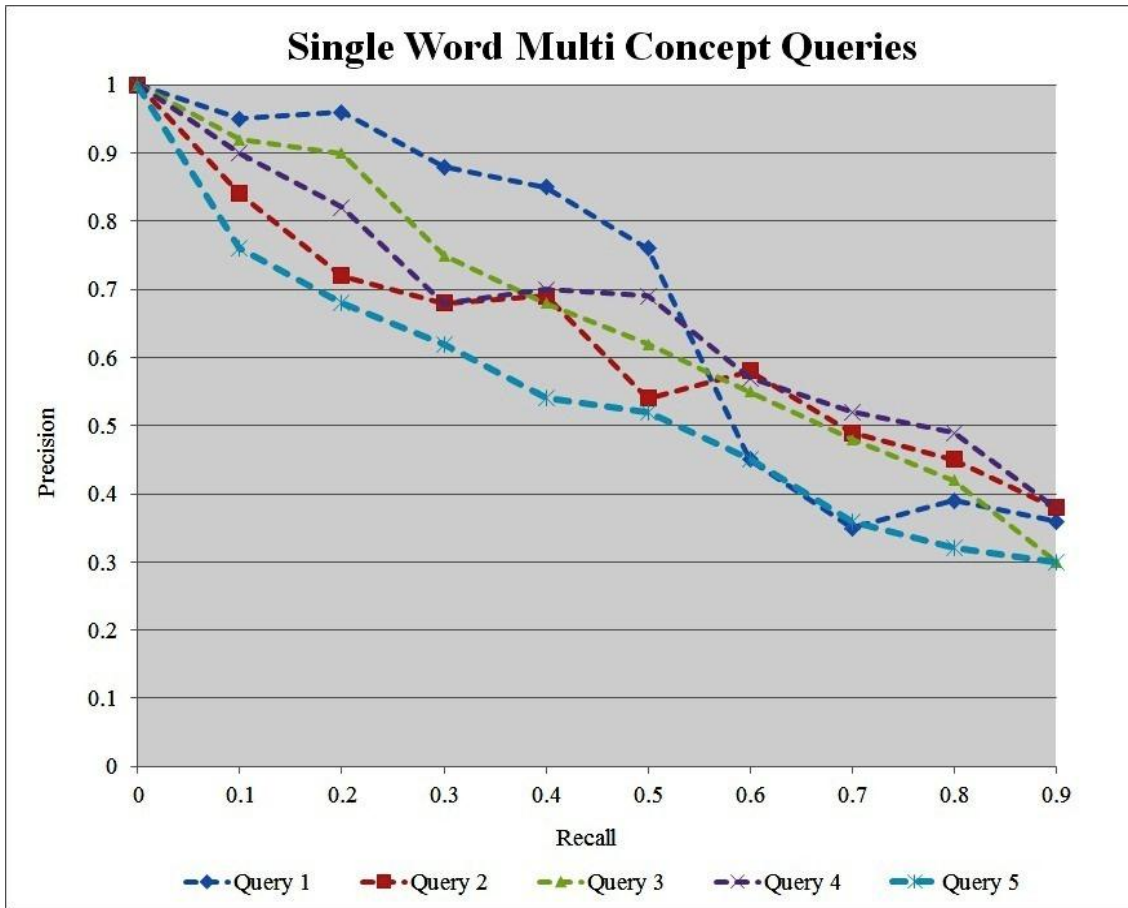


Figure 4.12: Precision Recall curve of Single Word Multi Concept Queries

The Figure 4.12 depicts the precision recall curve for the five randomly selected single word multi-concept queries. We have applied the queries that are already expanded by the semantic query interpreter on the SemRank model. The list of five single word multi-concept queries is street, office, transport, park and game. As the multi-concepts queries is the combination of several other concepts. Among them some are more related than the other. We have input the original query terms along with the expanded terms and the Semantic Similarity value between them. This will help in computing the semantic intensity of the concept with in the query and the retrieves most of the relevant results.

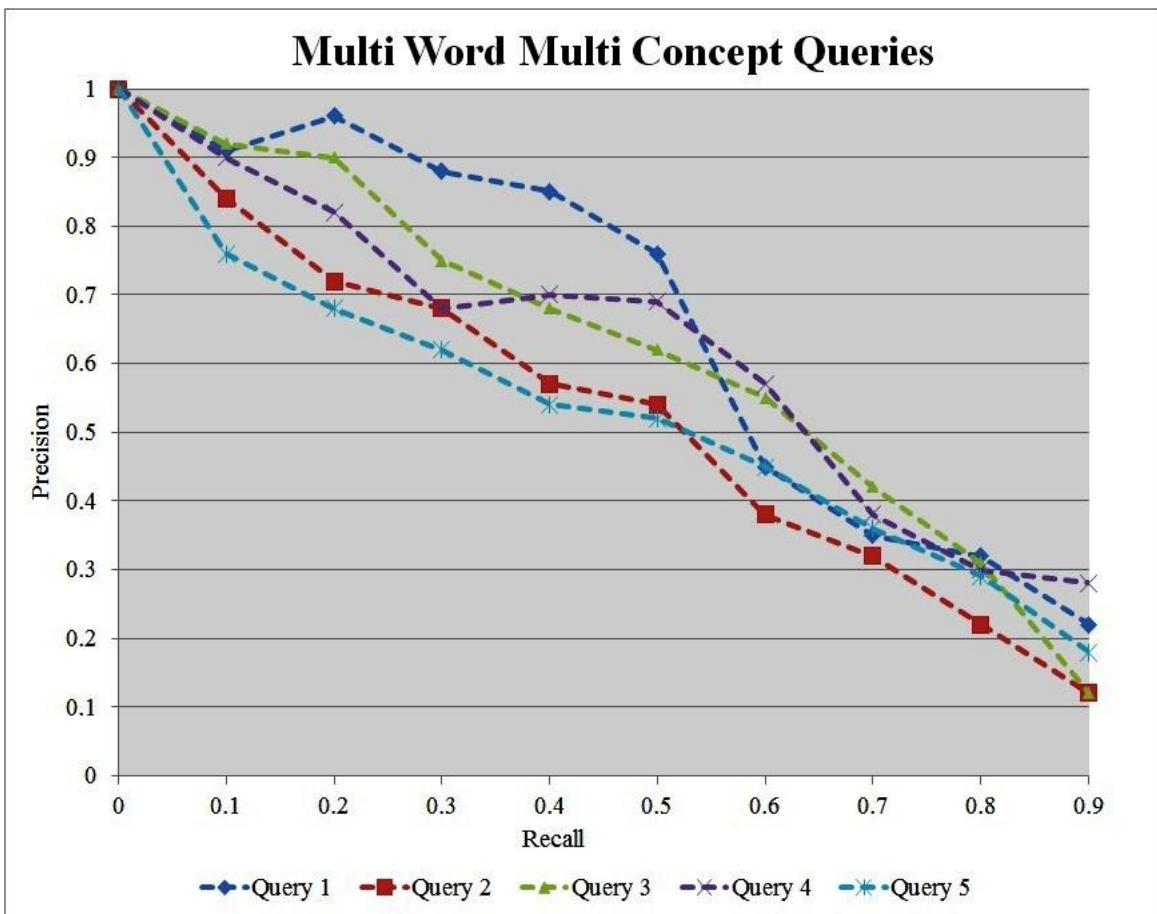


Figure 4.13: Precision Recall curve of Multi Word Multi Concept Queries

The Figure 4.13 depicts the precision recall curve for the five randomly selected multi word multi-concept queries. We have applied the queries that are already expanded by the semantic query interpreter on the SemRank model. The list of five multi-word multi-concept queries is car on the road, people in the park, Allow me to view building in the street, people on the seaside, I want to see images of people sitting on the benches. As the multi-concepts

queries is the combination of several other concepts and several categories of word like it may either a single concept of multi concept like in the “people in the park” the word people is the single concept while the word park is multiconcept word. While some of the queries are the combination of single concept words like “car on the road”. In this query car and road both the single concept words.

Among the expanded terms some are more related than the other. For finding the degree of relevancy between the original query terms and expanded term we have used the semantic similarity function. We have input the original query terms along with the expanded terms and the Semantic Similarity value between them. This will help in computing the semantic intensity of the concept with in the query and the images.

The interpolated precision-recall curve for the different categorizes of the queries. The randomly selected five queries for each type (single word single concept, single word multi concept, multi word multi concept) queries are selected to further evaluate the performance of the proposed system. As shown in Figures 4.10, 4.11 and 4.12, the proposed SemRank may improve the performance of the retrieved results. It shows the significant result for the simple keyword based queries to the multi word multiconcept queries. It is due to the fact that the SemRank find the intensity of different concept dominancy hierarchy. This concept dominancy or semantic Intensity helps the ranking function to sort the results based on the degree of relevancy between the concepts instead of the frequency of the occurrences.

4.5 Chapter Summary

In conclusion we would like to accentuate that 100% retrieval performance is an exceedingly hard dilemma to achieve. The main problem lies in the deed that we don't totally formularize the term relevant and irrelevant. In this chapter, we have explored the proposed new ranking strategy known as SemRank which uses the SI measure to calculate the image relevancy weights against the query. It has an advantage that it can employ the semantics inside the image and the query in determining the ranking order. We have compared our model with the Vector Space Model (VSM). Experimental results showed that SemRank approach has better retrieval performance than the VSM. We believe that considering the Semantic Intensities of the images enhance the precision of the IR systems. In future, we plan to exercise our approach on other image as well as on video datasets.

We have presented an overview of various IR models in this chapter, from the boolean model to various statistical models and the merits and demerits of these models for the ranking. In particular, several different features are used for ranking data have been described, such as visual similarity, term frequency etc. And in this chapter we have also shown that how to build an effective ranking strategy that will sort the output according to the semantic relevancy degree because the users are mostly interested in the top ranked results. After investigating the Semantic query interpreter and the SemRank efficiency on the images. In the next chapter, we extend our proposed Semantic Query Interpreter for the videos as well.

Chapter 05

Semantic Query Interpreter for Video Search & Retrieval

“Whatever the device you use for getting your information out, it should be the same information.”

Tim Berners-Lee

Spurred by technology modernizations, there has been a gigantic upsurge in the utilization of video, as is the most widely exploited media type owed to its content richness, for many significant applications. To sustain an on-going rapid growth of video information, there is an emerging demand for a sophisticated video retrieval system. However, current video retrieval solutions are still immature and lack of any standard. As a remedy to the problems different approaches have been developed feature based indexing is among one of them. The automated process of the video feature indexing replaces the manual process. But unfortunately the low level feature extraction is unable to interpret the semantics of the video intuitively. This predicament is due to the semantic gap. The semantic gap is due to the high level user requirement for the video and the computer interpretation of video as an arbitrary sequence of audio-visual tracks

With the ubiquitous use of digital video capturing devices and low cost storage devices, most users are now accustomed to the easy and intuitive way when searching with large video sources. Witnessing the overwhelming of digital video media, the research community has elevated the issue of its worthwhile use and management. Stored in immense multimedia databases, digital videos need to be retrieved and structured in an intelligent way, relying on the content and the rich semantics involved. Focusing on video media as the most complex form of multimedia, in particular the semantic aspect of it is the most challenging task of semantic based video retrieval. In order to achieve semantic and conceptual retrieval of videos from large repositories, there is an urge for the system that can interpret the user's demand semantically and retrieve the result accurately. In order to cope with these problems, we are proposing a technique for automatic query interpretation known as the Semantic Query Interpreter (SQI) that can expand the user query at the lexical and conceptual level rather than the visual. SQI interprets the user query both lexically and semantically by using open source knowledgebases i.e. WordNet and ConceptNet. Effectiveness of the proposed method is explored on the open-benchmark video data set the LabelMe video data set. Experimental results manifest that SQI shows substantial rectification over the traditional ones.

The remainder of the chapter is as follows. Section 5.1 discusses the introduction about the chapter. Section 5.2 reviews the video structure and representation. Section 5.3 explores the state of the art. Section 5.3 includes the new proposed technique. Section 5.5 discusses the datasets that will be used for investigation. Section 5.6 contains the experimentation setup i.e. to evaluate and compare the proposed technique with existing ones. Finally in we conclude the chapter in Section 5.7.

5.1 Introduction

Recently the availability of the multimedia data has increased exponentially together with the development of the techniques to facilitate the storage and access of these contents. A new landscape for business and innovation opportunities in multimedia content and technologies has naturally emerged from this evolution, at the same time that new problems and challenges arise. The web also contains a wide variety of media, roughly 60,000 new videos are uploaded to YouTube.com per day, and in the image domain, several thousand images per minute are added to Flickr.com. Websites like YouTube, Daily motion has led to a rather uncoordinated publishing of video data by users worldwide [Cunningham et al. 2008]. Due to the sheer amount of large data collections, there is a growing need to develop new methods that support the users in searching and finding videos they are interested in.

There is an immense diversity of video collections exist, ranges from small personal video collections to monumental archives of TV, CCTV, news documentary, etc. This leads to the challenge of how to extract the required information from such a colossal data and how to make this wealth of information worthwhile and easily accessible to the user. In order to attain this purpose it is inevitable to be able to apprehend the semantics that lies inside the group of frames.

The most unique characteristic of a video document is its ability to convey a rich semantic presentation through the synchronized audio, visual and text presentations over a period of time. Often, the videos are related to a particular topic which is described using both images and text. This makes it more difficult, as the user needs visual information like key frames or video playback to judge if a video clip is relevant or not. The contents alone are not

sufficient enough to find the desired video clip. Previous research has been concentrated on content based retrieval, so it is a well-studied process. However, semantic based video retrieval as a research field is nearly untouched.

5.2 Video Structure and Representation

With the evolutions in data capturing, storing, and transferring technologies, video usage has increased in various applications such as documentaries, security cameras, distance learning, video conferencing, and so on. Proliferated video data requires effective and efficient data management. Video data is distinctive from textual data since video has image frames, sound tracks, texts that can be extracted from image frames, spoken words that can be interpreted from the audio track, temporal, and spatial dimensions. Multiple sources of the video surge the magnitude of video data and make it obstinate to store, manage, access, reuse, and compose.

A video clip comprises of a succession of still images shown rapidly in a sequence. Typical frame-rates are around 24 or 30 images per second. There can also be a sound track associated with the video sequence. Videos can thus be stored hierarchically so that the image frames and sound tracks are sub-objects in an object tree. Alternatively, usually the important key frames of the video as separate images, and have the entire video clip as the parent object. If the video is too long then the video is first converted into various segments and then the key frames are selected for every segment.

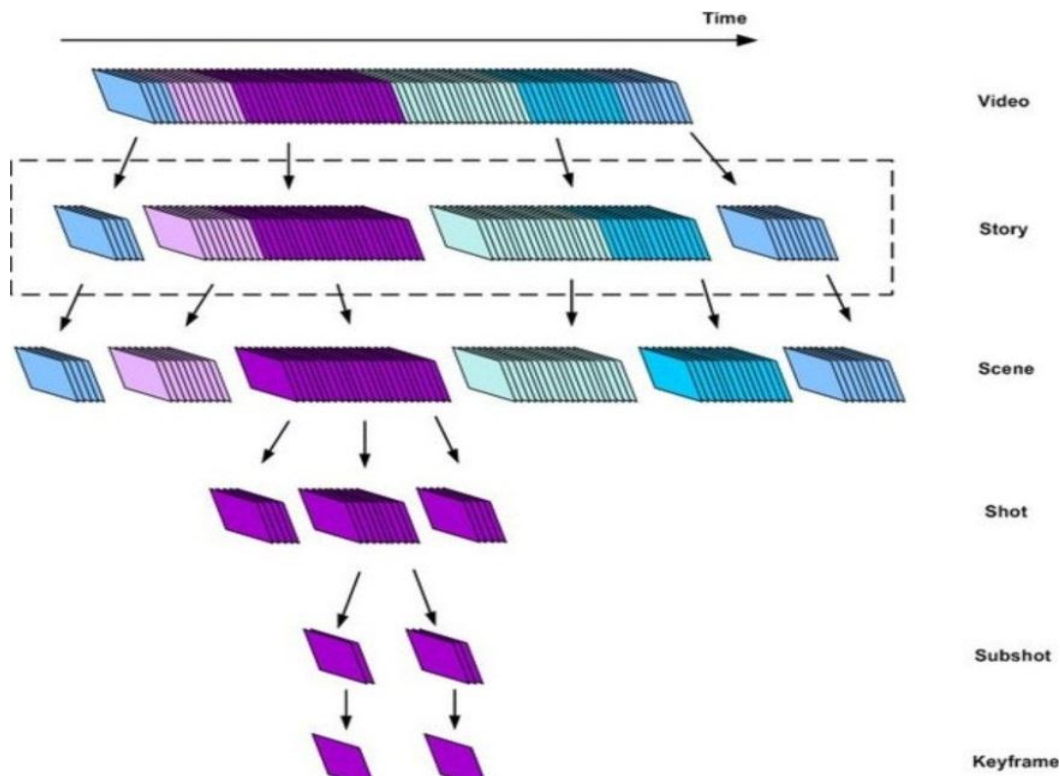


Figure 5.1: A hierarchical decomposition and representation of the video contents¹.

In hierarchical video structure, complex video units are divided into elementary units recursively. The most often proposed hierarchies have a segment-scene-shot-frame structure. Video stream consists of frames, shots, scenes and sequences.

Frames are single pictures and the elementary video units. There are 14-25 frames per second, so frame sequences give more meaning than individual frame. Physically related frame sequences generate video shots.

Shots are segmented based on low level features and shot boundary algorithms can detect shots automatically. Shots are not sufficient for extracting the semantics from the video because there are too many shots in a long video and shots do not capture the semantic structure of video. Therefore, semantically related and temporally adjoining shots are grouped into scenes.

¹ http://www.scholarpedia.org/article/Video_Content_Structuring

Scenes are segmented on the high-level features logically. The scene boundary detection is more difficult than shot boundary detection. The scenes are usable in the content-based video indexing and retrieval due to their semantic structures. Scenes may be yet not sufficient for search and retrieving very long video. It might be necessary to combine related scenes into sequences or acts. Sequence extraction is also difficult and needs human assistance. Figure 5.1 shows the hierarchical structure of video. The technique used for the converting the video into hierarchal structure is discussed below.

5.2.1 Shot Boundary Detection

The atomic unit of access to video content is often considered to be the video shot. Monaco [Monaco. 2009] defines a shot as a part of the video that results from one continuous recording by a single camera. It hence represents a continuous action in time and space in the video. Especially in the context of professional video editing, this segmentation is very useful. Consider for example a journalist who has to find shots in a video archive that visualise the context of a news event. Shot segmentation infers shot boundary detection, since each shot is delimited by two consecutive shot boundaries. Hanjalic provide a comprehensive overview on issues and problems involved in automatic shot boundary detection [Hanjalic. 2002]. A more recent survey is given by Smeaton et al. [Smeaton et al. 2010].

Shots are the smallest semantic units of a video and consist of a sequential set of frames. A scene is composed of a number of shots. The gap between two shots is called a shot boundary. Two shots are separated by a transition, like a fade-over or simply a hard cut. According to Zhang et al. [Zhang et al. 1993], there are mainly four different types of common shot boundaries within shot

- **A cut:** It is a hard boundary or clear cut which appears by a complete shot over a span of two serial frames. It is mainly used in live transmissions.
- **A fade:** A fade can be either the fade-in or the fade-out. The fade-out emerges when the image fades to a black screen or a dot. The fade-in appears when the image is displayed from a black image. Both effects last a few frames.

- **A dissolve:** It is a synchronous occurrence of a fade-in and a fade-out.
- **A wipe:** This is a virtual line going across the screen clearing the old scene and displaying a new scene. It also occurs over more frames.

In text retrieval, documents are treated as units for the purpose of retrieval. So, a search returns a number of retrieved results. It is easy to design a system that retrieves all documents containing a particular word. The user can browse through the results easily to find parts of interest. If documents are too long, techniques have been developed to concentrate on the relevant sections [Salton et al., 1993].

This practice cannot be used for videos. If videos are treated as units of retrieval, it will not lead to a satisfactory result. After relevant videos have been retrieved, it is still an issue to find the relevant clip in the video. Especially as most clips have a duration of only a few seconds. Even if these small clips are seen as associated stories of several minutes of length, it is not optimal. It is time consuming to browse through all video sections to find the relevant part [Girgensohn et al., 2005]. Visual structures such as colour, shape and texture can be used for detecting shot boundaries and for selecting key frames [Aigrain et al., 1996].

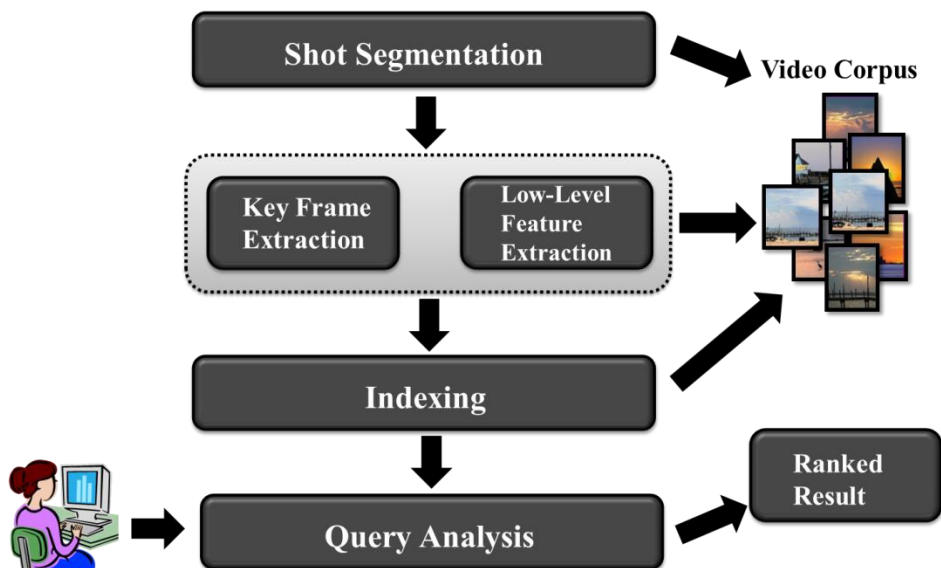


Figure 5.2: Analysis of the video contents

5.2.2 Key Frame Selection

Key frames are still images extracted from original video data that best represents the content of the shot in an abstract manner. Key frames have been frequently used supplement the text of a video log, identifying them was frame done manually in the past. The effectiveness of key-frames depend on how well they are chosen from all frames of a sequence. The image frames within a sequence are not all equally descriptive. Certain frames may provide more information about the objects and actions within the clip than other frames. In some prototype systems and commercial products, the first frame of each shot has been used as the only key frame to represent the shot content.

Key frame based representation views video abstraction as a problem of mapping an entire segment to some small number of representative images. The key-frames needs to be based on the basic principle of content based so that they retain the eminent content of the video while purging all redundant information. In theory, semantic primitive of video, such as interesting objects, actions, and events should be used. However, such general semantic analysis is not currently feasible, especially when information from sound tracks and/or closed caption is not available. One possible and simple solution to detect key frames is to take any frame e.g. the first or the middle one as a key frame.

One merit of key frame extraction is to only process key frames instead of all frames, while not losing too much discriminative information. On a shot level, it has been shown that using key frames instead of either regularly sampled frames or the first frame of a shot improves performance. Since key frames are extracted within a shot, a possible problem is that they might repeat themselves in different shots.

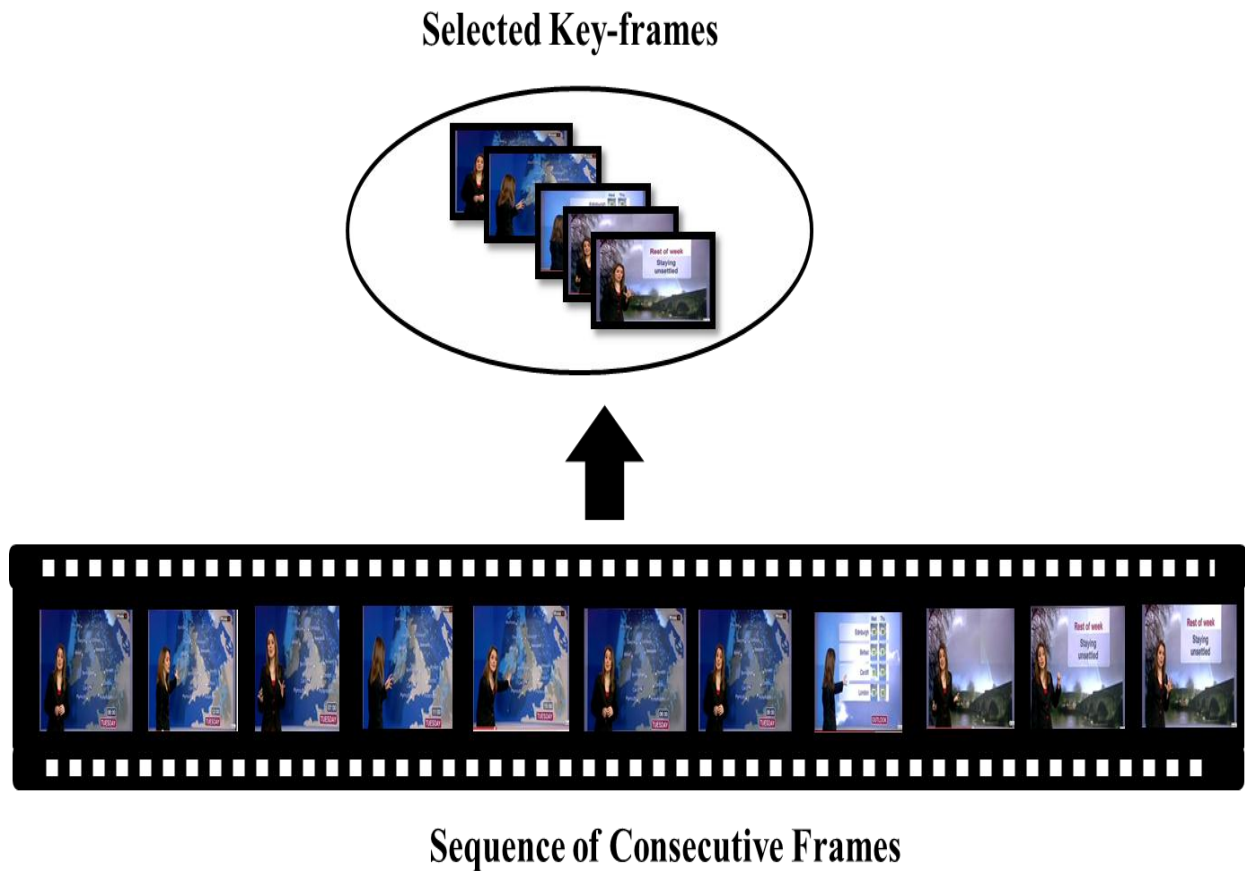


Figure 5.3: Key frame identification

5.2.3 Feature Extraction

In CBVIR system videos, clips or single frames should be represented as points in an appropriate multidimensional metric space where dissimilar videos are distant from each other, similar videos are close to each other, and where the distance function captures well the user's concept of similarity. A picture is worth a thousand words, and thus a profound challenge comes from the dynamic interpretation of videos under various circumstances. A video will first be pre-processed (e.g. shot boundary detection, key frame selection), followed by the feature extraction step, which will emit a video description.

Features have to describe videos with as few dimensions as possible, while still preserving properties of interest. Different modalities to satisfy these requirements will be explained below.

The content-based video indexing and retrieval intents at retrieving video content efficiently and effectively. Most of the studies have reinforced on the visual component of video content in modelling and retrieving the video content. Besides visual components, much worthy information is also conveyed in other media constituents such as superimposed text, closed captions, audio, and speech that shepherd the pictorial component. The multimodal nature of the video makes it more strenuous to process it. There are three modalities of the video i.e. visual, auditory, and textual modalities in video.

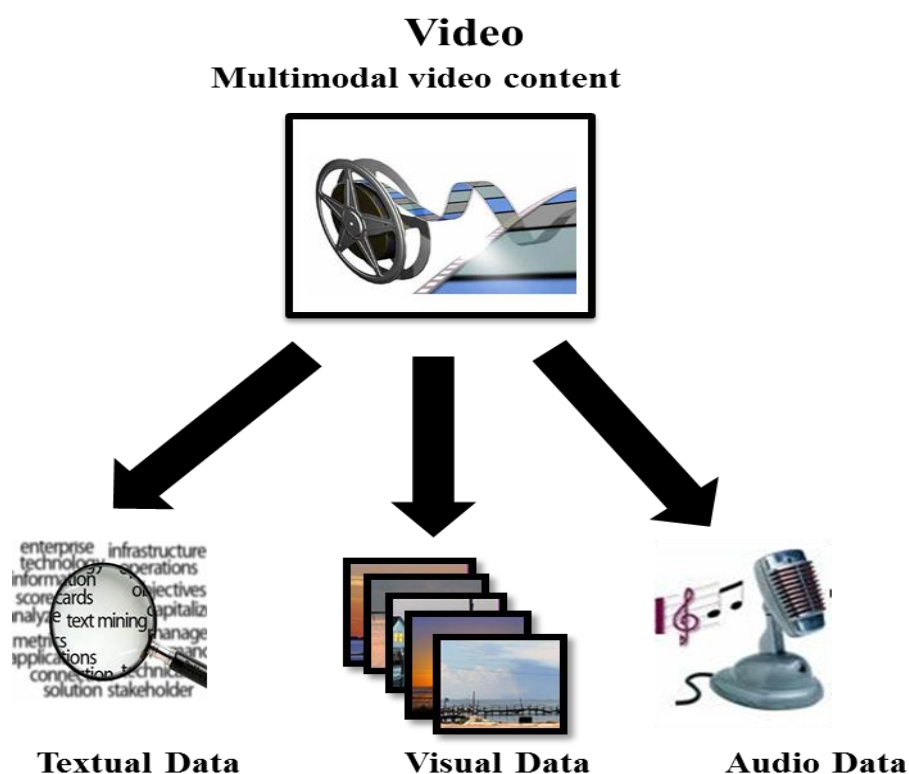


Figure 5.4: Multimodal Video Content

5.2.3.1 Visual Modality

Visual modality concerns with everything that can be viewed in the video. The visual data can be acquired as a stream of frames at some lines of resolution per frame and an associated frame rate. The elementary units are the single image frames. Consecutive video frames give a sense of motion in the scene. Visual perception is elementary information while watching video. Visual information caters for perceptual properties like colour, texture, shape

and spatial relationships; semantic properties like objects, events, scenes and the meaning with the combination of these features. Erol proposed the shape based retrieval of the video objects [Erol et al. 2005].

Visual object acts as a source of visual data. There are a lot of visual objects, but salient objects are more considerable for viewer. Visual events are consecutive frame groups, which give semantic meaning.

5.2.3.2 Audio Modality

Auditory modality concerns with everything that can be heard in the video. Audio refers to generic sound signals, which include speech, dialog, sound effects, music and so on. The audio information often reflects directly what is happening in the scenes and distinguishes the actions. Speech is related to the story and you cannot understand the content well without listening to it. Music changes atmosphere of scene and physiological viewpoint of audience, horrible films do not scare people without sudden sound increases.

The audio data play back in simultaneously with the playback of the video frames. The audio data may cover speech, music, sound effects and different sound tracks, etc. Each of these characteristic sound tracks can be characterized using their own domain specific sound events and objects as Hunter et al. [Hunter et al. 1998]. Moriyama et al [Moriyama et al. 2000] divide audio component into four tracks, namely speech by actors, background sounds, effect sounds, and BGMs.

Shot represents pictorial changes in visual modality. The BGM represents music superimposed on the video. Effect sounds superimposed on video and have no melody such as fight effect For some of the categorizes of the video audio plays a very most significant role and exclusively can provide worthy information like news.

5.2.3.3 Textual Modality

Textual modality includes texts and speech transcripts. Textual modality contains everything that can be converted into text document in the video document. Text can be

thought as a stream of characters. There are mainly two types of textual information in video i.e. visible texts and transcribed speech texts.

- Visible texts are superimposed text on the screen such as closed captions or natural parts of scenes such as logos, billboard texts, writings on human clothes, etc.
- Another text source is speech that can be transcribed into text [Mihajlovic et al. 2001].

Texts play important role in illuminating the video content. Especially in news, in documentary videos and in distance learning videos, texts are heart of the video content. For broadcast news videos, text information may come in the format of caption text strings in video frames, as close caption, or transcripts. This textual information can be used high-level semantic content, such as news categorization and story searching. In documentary videos, speech is more dominant while clarifying the subject. In distance learning videos, all stuff can be converted to text from teacher speaking to board content.

In textual information retrieval of video area, the Infromedia [Infromedia] project has a leading role. This project aims to automatically transcribe, segment, and index the linear video using speech recognition, image understanding, and natural language processing techniques. Video Optic Character Recognition (VOCR) techniques are used for extraction text from video frames and Automatic Speech Recognition (ASR) techniques are used for conversion of speech to text. Their system indexes news broadcasts and documentary programs by keywords that are extracted from speech and closed captions.

5.3 State of the Art

Unlike still images, videos are dynamic in nature and are visual illustrations of information. The continuous characteristic and immense data volume make it further challenging to process and manage videos. On the other hand, as more information, particularly temporal and motion, is contained in videos, we have a better opportunity to analyse visual content inside video. Furthermore, although videos are continuous media, the semantics contained within a video program is difficult to extract.

During recent years, methods have been developed for retrieval of images and videos based on their visual features. Commonly used visual similarity measurements are colour, shape, texture and spatio-temporal [Niblack et al. 1993]. Two typical query modalities include query by example and query by text [Chang et al. 1997]. A number of studies have been conducted on still image retrieval. Progresses have been made in areas such as feature extraction [Chang et al. 1995], similarity measurement, vector indexing [Chang et al. 1987], [Rowe et al. 1994] and semantic learning [Minka et al. 1996]. Content-based video retrieval includes scene cut detection, key frame extraction [Meng et al. 1995], [Zhang et al. 1994]. Extraction of constituent objects and discovery of underlying structures has been already investigated by the research community and many breakthrough results have been made. While a problem of extracting the semantics from the video has been addressed by many researcher but had yet not been solved.

While image retrieval techniques can be applied to video searching, unique features of video data demand solutions to many new challenging issues. The video retrieval work can be mainly divided into two main areas i.e. content based video retrieval and Semantic based Video retrieval.

5.3.1 Content Based Video Retrieval (CBVR)

In past the research community has proposed the content based video retrieval for the enhancement of the traditional video search engines [Steven et al. 2007]. The content based video retrieval intends to retrieve the required video segments on the basis of the content of the video with the user intervention [Jiunn et al. 2006]. Current based video indexing and retrieval systems face the problem of the semantic gap between the simplicity of the available visual features and the richness of user semantics. Content based video and retrieval has been the focus of the research community during last few years. The main idea behind this is to access information and interact with large collections of videos referring to and interacting with its content, rather than its form. Content-based video retrieval (CBVR) tasks such as auto-annotation or clustering are based on low-level descriptors of video content, which should be compact in order to optimize storage requirements and efficiency. Shanmugam et al uses the color, edge and motion feature as a representative of the extracted key frame. These are stored

in the feature library and are further used for content based video retrieval [Shanmugam et al. 2009]. Although there has been a lot of effort put in this research area the outcomes were relatively disappointing.

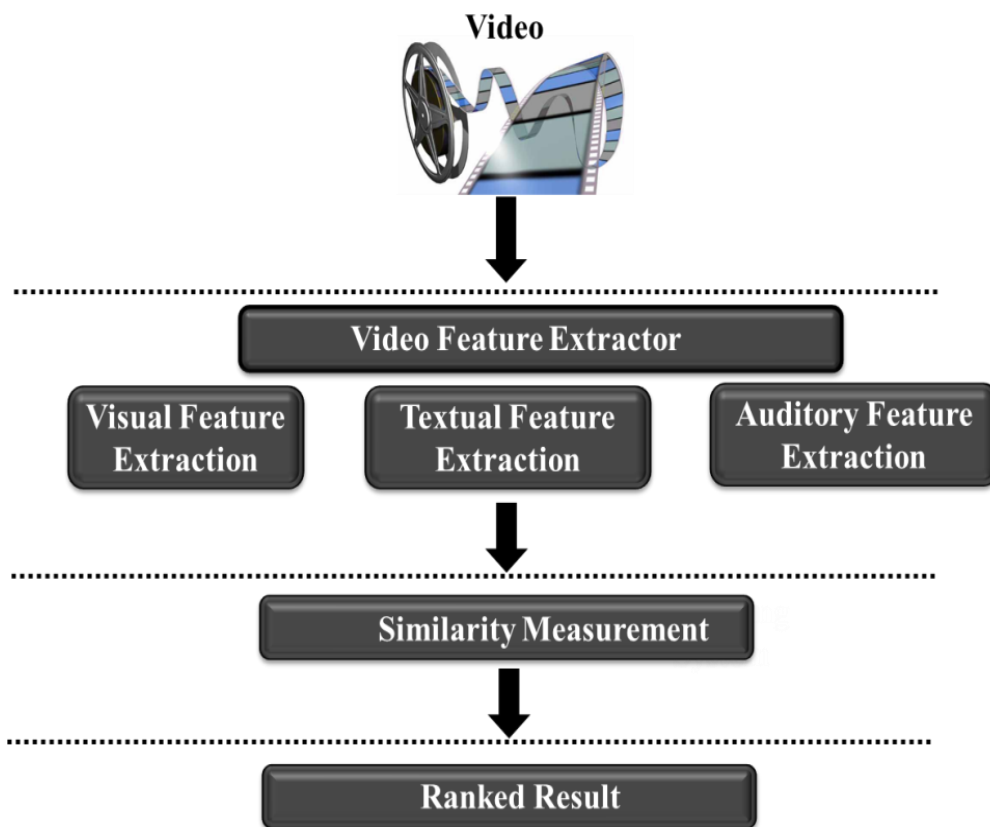


Figure 5.5: Typical Content Based Video Retrieval

So as to digest the vast amount of information involved in the construction of video semantics it is substantial to define appropriate video representation in a CBVIR system.

Video analysis is the backbone of all video retrieval engines. Analysis approaches aim to develop effective and efficient methodologies for accessing video contents. As we have discussed in Section 5.3.3 a video document consists of several modalities, e.g. a video document is made up of audio tracks, visual streams and different types of annotations. Thus, video analysis has to take numerous modality features into consideration. Moreover, these features are of various natures. Video analysis techniques can be split into two main categories

i.e. content-based analysis and semantic based analysis. In this thesis we will focus on the Semantic analysis instead of the content based.

5.3.2 Video Semantics

Video as a carrier of information presents a prominent role in sharing information today. The most significant advantage of video is its capacity to transmit information in a manner that serves the human perception best to perceive and consume information by audio visual means.

Current trends in video semantics suggest a great deal of enthusiasm on the part of researchers. Semantic based search and retrieval of video data has become a challenging and important issue. Video contains audio and visual information that represent a complex semantics which are difficult to extract, combine in video information retrieval.

Extracting the semantic content is complex due to requiring domain knowledge and human interaction. The simplest way for modelling video content is free text manual annotation, in which video is first divided into segments and then every segment is described with free text. There has been a plethora of interesting research work presented recently that focuses on problem of bridging this semantic gap [Hauptmann et al. 2007], [Hoogs et al. 2003], [Snoek et al. 2007]. In our thesis we use these annotations to extract the semantic form the videos and the user requests by using the knowledgebases. We have proposed the exploit the knowledgebases for the semantic extraction from the video at the query and ranking level.

5.3.3 Query expansion

The human brain and visual system together with the human auricular skills provide outstanding capabilities to process audio-visual information, and instantly interpret its meaning on the basis of experience and prior knowledge. Audio-visual sensation is the most convenient and most effective form for humans to consume information we believe what we can see and hear, and we prefer to share our experiences by aural and visual description. In particular for complex circumstances, visualization is known to convey the facts of the matter

best. The growth of visual information available in video has spurred the development of efficient techniques to represent, organize and store the video data in a coherent way.

We argued that when interacting with a video retrieval system, users express their information need in search queries. The underlying retrieval engine then retrieves relevant retrieval results to the given queries. A necessary requisite for this IR scenario is to correctly interpret the users' information need. As Spink et al. [Spink et al. 1998] indicate though, users very often are not sure about their information need. One problem they face is that they are often unfamiliar with the data collection, thus they do not know what information they can expect from the corpus [Salton et al. 1997]. Further, Jansen et al. [Jansen et al. 2000] have shown that video search queries are rather short, usually consisting of approximately three terms. Considering these observations, it is hence challenging to satisfy users' information needs, especially when dealing with ambiguous queries.

Triggering the short search query "Apple", for example, a user might be interested in videos about the company called Apple, fruit. Without further knowledge, it is a demanding task to understand the users' intentions. Semantic based information retrieval aims at improving the traditional content based retrieval model.

Video retrieval based query expansion approaches include [Volkmer et al. 2006], who rely on textual annotation (video transcripts) to expand search queries. Within their experiment, they significantly outperform a baseline run without any query expansion, hence indicating the potentials of query modification in video search. Similar results are reported by Porkaew [Porkaew et al. 1999] and Zhai et al. [Zhai et al. 2006], who both expand search queries using content-based visual features.

The original, manually entered query is most important as there are many different ways to describe the same object or event. However, it is nearly impossible to formulate a perfect query at first attempt due to the uncertainty about the information need and lack of understanding on the retrieval system and collection. The original query indicated what the searcher really wants, but a problem is, that a query might not be precise enough or that

retrieval misses videos that have semantic similarities but no speech similarities. Different query expansion techniques have been tested, e.g. [Beaulieu, 1997, Efthimiadis, 1996].

In [Zhai et al., 2006], the authors propose an automatic query expansion technique. It expands the original query to cover more potential relevant shots. The expansion is based on an automatic speech recognition text associated to the video shots. Another approach, the interactive query expansion is discussed e.g. in [Magennis et al. 1997]. The idea is that the automatically-derived terms are offered as suggestions to the searcher, who decides which to add. All of the above approaches prove the usefulness of the automatic query expansion techniques. Current query expansion techniques for the videos lack the semantic based query expansion. The detailed state of the art for the query expansion was already discussed in chapter 3.

5.4 Proposed Contribution

In light of the above stated problems we have proposed a semantic query interpreter for the videos as well. The semantic query interpreter will expand the user query lexically as well as semantically. The main theme of the Semantic Query Interpreter for the video is same as the images. We have evaluated the SQI for the images on the LabelMe image dataset. The SQI for the video will be evaluated on the LabelMe video dataset.

We have applied our research work on the LabelMe videos, the structure of the LabelMe video datasets structure is similar as that of the LabelMe images, as the video is the sequential combination of the images. Based on this, the LabelMe video is handled, and the other difference is that they are not only dealing the objects tracking, but also capture events in the videos. The user begins the annotation process by clicking control points along the boundary of an object to form a polygon. When the polygon is closed, the user is prompted for the name of the object and information about its motion. The user may indicate whether the object is static or moving and describe the action it is performing, if any. The user can further navigate across the video using the video controls to inspect and edit the polygons propagated across the different frames.

To correctly annotate moving objects, The LabelMe web tool allows the user to edit key frames in the sequence. Specifically, the tool allows selection, translation, resizing, and editing of polygons at any frame to adjust the annotation based on the new location and form of the object. For the event annotation, the users have an option to insert the event description in the form of sentence description. When the user finishes outlining an object, the web client software propagates the location of the polygon across the video by taking into account the camera parameters. Therefore, if the object is static, the annotation will move together with the camera and not require further correction from the user. With this setup, even with failures in the camera tracking, the user can correct the annotation of the polygon and continue annotating without generating uncorrectable artifacts in the video or in the final annotation.

The Semantic Query Interpreter module for the videos is same like an image. The same four modules it contain i.e. core lexical analysis, common sense reasoning, candidate concepts selection and ranking and retrieval module. The results of the SQI for the videos are ranked and retrieved using the Vector Space Model. The detailed discussion of all the modules was already presented in chapter 3.

5.5 Evaluation

The majority of IR experiments focus on evaluating the system effectiveness. The effectiveness of the proposed system was investigated by using the same measure that we used for the images like precision, recall and F-measure (F-Score) and the significance of these evaluation parameters was already discussed in chapter 3 The experiments were performed on LabelMe video dataset. A brief over view of the LabelMe Videos is discussed in the next section.

5.5.1 LabelMe Videos

The LabelMe Videos are aim to create an open database of videos where users can upload, annotate, and download content efficiently. Some desired features include speed, responsiveness, and intuitiveness. They designed an open, easily accessible, and scalable annotation system to allow online users to label a database of real-world videos. Using the

LabelMe labelling tool, they created a video database that is diverse in samples and accurate, with human guided annotations. They enriched their annotations by propagating depth information from a static and densely annotated image database. The basic intention of this annotation tool and database is that it can greatly benefit the computer vision community by contributing to the creation of ground truth benchmarks for a variety of video processing algorithms, as a means to explore information of moving objects.

They intend to grow the video annotation database with contributions from Internet users. As an initial contribution, they have provided and annotated a first set of videos. These videos were captured at a diverse set of geographical locations, which includes both indoor and outdoor scenes. Currently, the database contains a total of 1903 annotations, 238 object classes, and 70 action classes.

The most frequently annotated static objects in the video database are buildings (13%), windows (6%), and doors (6%). In the case of moving objects the order is persons (33%), cars (17%), and hands (7%). The most common actions are moving forward (31%), walking (8%), and swimming (3%).

5.6 Experimental Setup

The experiments presented in this thesis use the Precision (P) and recall (R), F-measure (F1) and as performance measurements. Overall, it can be concluded from our experiments that semantic based query expansion can improve the performance not only for the LabelMe corpus but also for other videos dataset. The proposed semantic query interpreter works well for the images as well for the videos. Some of the variation in the result is due to the problem of poor annotation. We have applied the three categories of the queries i.e. single word single concept, single word multi-concept and multi word multi-concept for investigating the performance of our proposed Semantic Query Interpreter on video dataset.

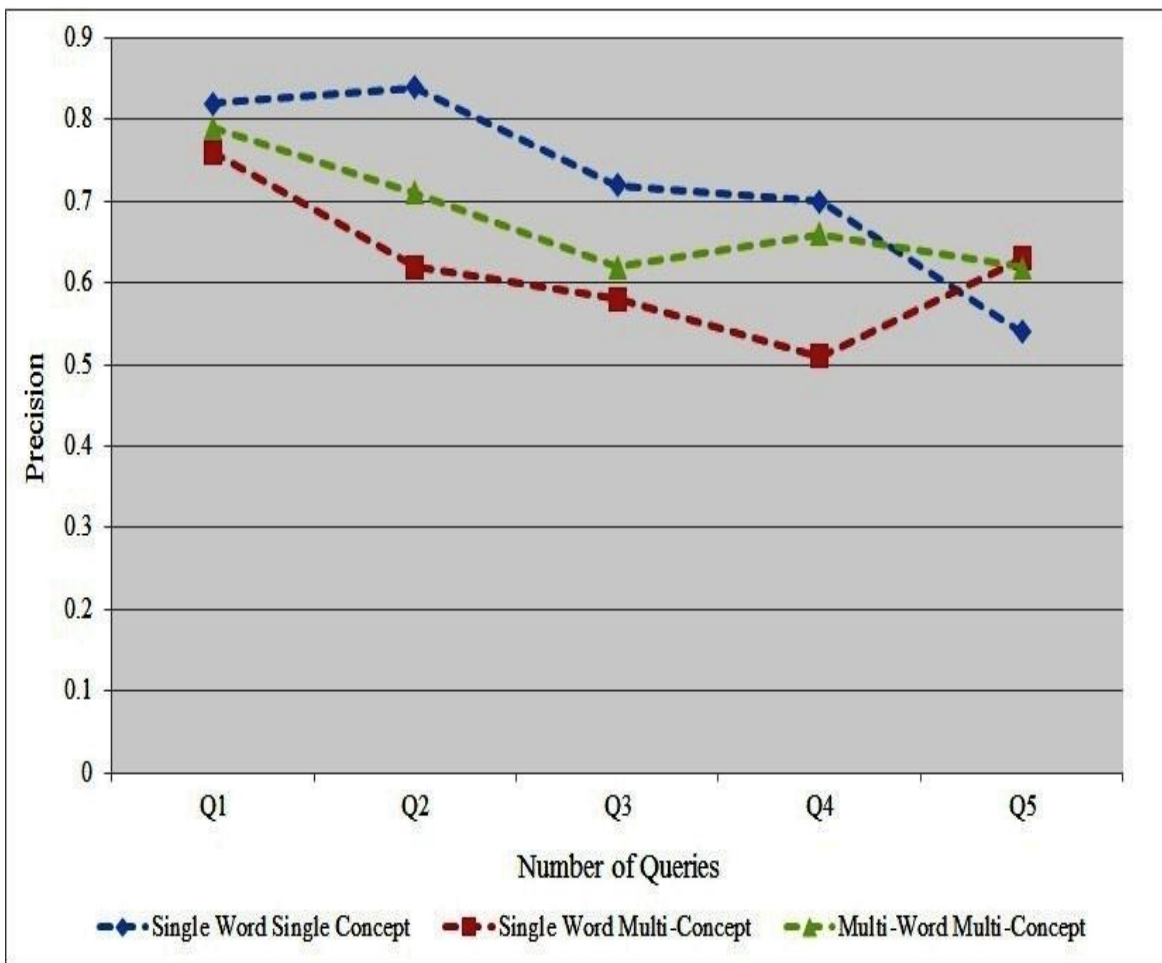


Figure 5.6: Different precision values for the five randomly selected user queries of three different categories on the LabelMe video corpus.

Figure 5.6 shows the precision of the five randomly selected different queries for each of the three categories i.e. Single Word Single Concept, Single Word Multi-Concept and Multi-Word Multi-Concept queries. The five randomly selected single word single concept queries are car, building, tree, sky, and house. The five randomly selected single word multiconcept queries are street, park, transport, game and office. While the five randomly selected multi-word multi-concept queries are car on the road, people in the park, allow me to view building in the street, people on the seaside and people sitting on the benches. The results show the substantial improvement of the retrieval precision from single word single concept to multi-word multi-concept. The mean average precision of the single word single concept

queries is 0.72, the mean average precision of the single word multi-concept queries is 0.62 and the mean average precision of the multi word multi-concept queries is 0.68. The result showed that the system works very well for many cases i.e. queries but for some cases, there is little bit variation. The efficiency of the proposed framework on video dataset is less than the images. It is due to the fact that video contain the complex nature and dealing the videos is difficult than images. The difference in the precision level of the different types of queries is due to the query complexity and due to the poor annotation. As with the increase in the complexity, there is a decrease in the performance efficiency. The system can expand the queries but fails to contribute in the annotation. Our proposed Semantic query interpreter has shown the significant precision level over the LabelMe video dataset. As we know that sometimes, the query expansion increases the recall of the system and decreases the precision. We have maintain the precision of the by selecting the candidate concepts selection module (see Chapter 3 section 3.4.2). It pruned the most semantically relevant concepts among the expanded concepts to the original selected query terms. The query terms are selected by using candidate term selection module (see Chapter 3 section 3.3.1.2) of the core lexical analysis. The candidate concepts selection module intents to maintain the precision of the system by selecting the expanded concepts based on semantic similarity between them.

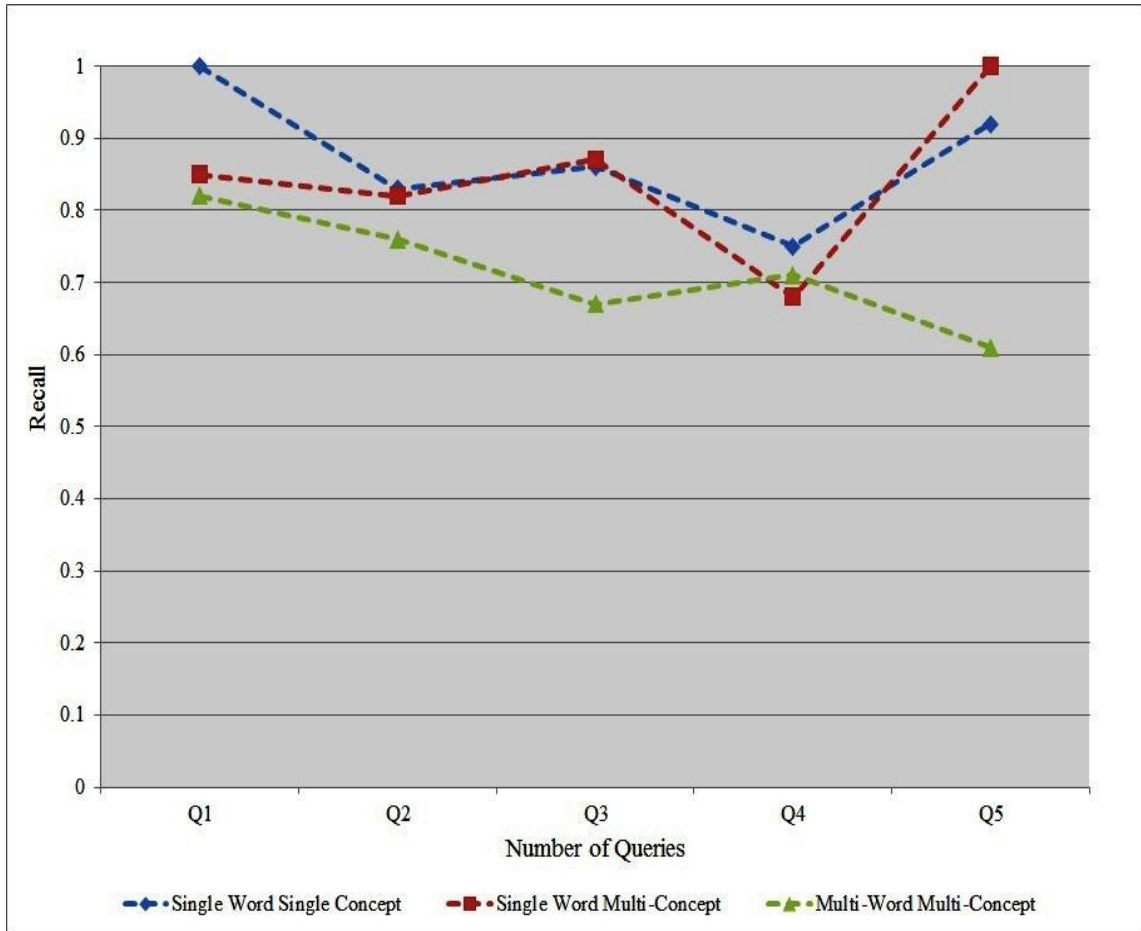


Figure 5.7: Different Recall values for the five randomly selected user queries of three different categories on the LabelMe video corpus

Figure 5.7 shows the recall of the five randomly selected different queries for each of the three categories i.e. Single Word Single Concept, Single Word Multi-Concept and Multi-Word Multi-Concept queries. The same five randomly selected three different categories of queries that are used for computing the precision is used for recall computation as well. The result shows the substantial improvement of the recall of the proposed model. The recall of the system can increase more if we remove the candidate concept selection module (see Chapter3 section 3.4.2) of the proposed framework. The mean average recall of the single word single concept queries is 0.87, the mean average recall of the single word multi concept queries is 0.84 and the mean average recall of the multi word multi concept queries is 0.71.

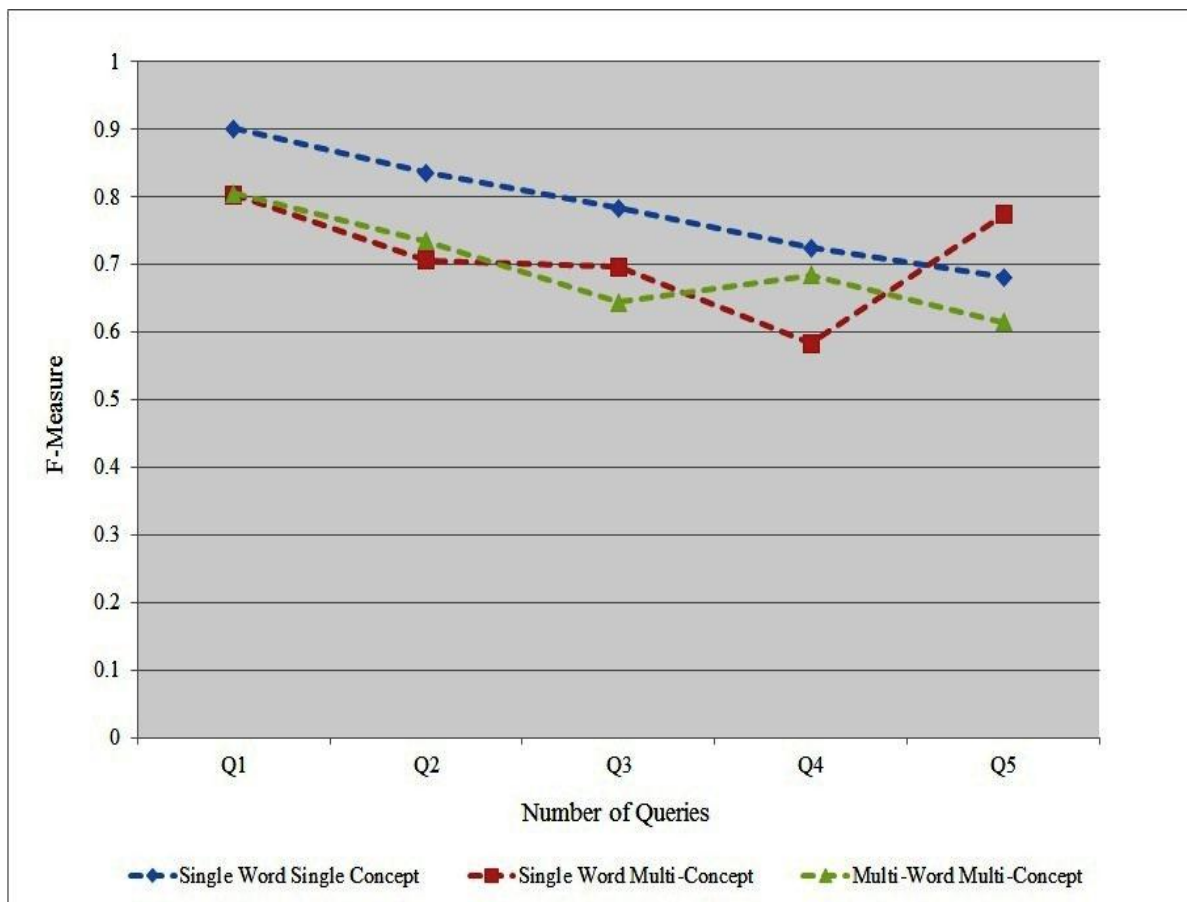


Figure 5.8: Different F-Measure values for the five randomly selected user queries of three different categories on the LabelMe video corpus.

Figure 5.8 shows the F-measure of the five randomly selected different queries for each of the three categories i.e. single Word Single concept, Single Word Multi-Concept and Multi-Word Multi-Concept queries. The mean average F-measure of the single word single concept queries is 0.78, the mean average F-measure of the single word multi- concept queries is 0.71 and the mean average F-measure of the multi-word multi-concept queries is 0.70. The mean average F-measure of the multi-word multi-concept query is lesser than the single word single concept and single word multi-concept. It is because with the increase in the complexity the efficiency decreases and is difficult to deal with

We have made the investigation of the performance evaluation of the proposed semantic query interpreter model on the LabelMe videos dataset. We have already proved in the chapter 3 that the conceptual as well as the lexical expansion boost the performance of the image retrieval system. The above results demonstrate that the proposed Semantic Query Interpreter will also enhance the performance of the video retrieval system. The overall efficiency of the semantic query interpreter for videos is less than the semantic query interpreter for the image retrieval. It is due to the fact that video has a complex structure to deal with. The Figure 5.6, 5.7 and 5.8 shows the substantial performance improvement of the proposed system ones. It is clear from the result that the lexical as well as the conceptual expansion is necessary to increase the performance of the IR system.

5.7 Chapter Summary

In this chapter, we have presented a semantic query interpreter approach for the videos. We have investigated the semantic query interpreter on the LabelMe video corpus. The proposed technique shows substantial results for the LabelMe videos. We have used the traditional similarity based retrieval model known as the Vector Space model in order to test the efficiency of the proposed semantic query interpreter on the video dataset. Experimental results for the comprehensive LabelMe video data set have demonstrated the usefulness of the proposed semantic based extraction. There are several areas for that are worth investigating. First, since it is infeasible to incorporate the proposed query interpreter to other datasets like the TRECVID, VideoCom, YouTube etc.

Chapter 06

Conclusion & Perspectives

“Solutions almost always come from the direction you least expect, which means there’s no point in trying to look in that direction because it won’t be coming from there.”

The Salmon of Doubt by Douglas Noel Adams

6.1 Introduction

The basic intention behind this chapter is giving a final reflection on the finished work and explores the directions for future work. We have addressed the main challenge of Semantic gap in Semantic Multimedia analysis, search and retrieval. We have tried to reduce this gap. This thesis has proposed solutions to the problems that help in the extraction and exploitation of the actual semantics inside the image and the video using the open source knowledgebases.

This chapter draws a conclusion in summarizing its cognitions and illustrates the course of the work. Section 6.2 summaries the findings of this thesis. In Section 6.3, the works that have not been considered in this research but that are worth being focused on in a future work.

6.2 Research Summary

Aiming to bridge the semantic gap, this thesis is presented a new paradigm of semantic based video and image search, more specifically, concept based video and image search method where the knowledgebases are used to extract the semantics in order to find the users requirements.

The following contributions have been presented in this thesis

6.2.1 Semantic Query Interpreter for Image Search and Retrieval

This thesis has proposed a Semantic Query Interpreter for image search and retrieval. The query plays a significant role in the information retrieval process. The performance of an IR system heavily depends upon the query engine. Keeping this in mind we propose a semantic query interpreter by using the query expansion technique. We expand the user query by using the open source knowledgebases. The query was expanded both lexical and conceptually. Initially the query is first pre-processed by using the basic NLP (natural language processing) function. We know that not every word in the query matters a lot. Some

of the terms in the query are more significant than the other. We have initially selected that significant terms and then expand it lexically by using the well-known lexical open source knowledgebase WordNet. While the conceptual expansion can be done by the open source conceptual reasoning knowledgebase Concept. These knowledgebases expand the user queries. Among the expanded terms some of the terms are noises that will however increase the recall but significantly reduce the precision of the system. These noises are removed by using the proposed candidate concept selection module. That can filter the noises from the expanded terms on the basis of the semantic similarity between the expanded and the original query terms. Vector Space model has been used to retrieve and ranks the result. The effectiveness of the proposed algorithm has been investigated on the LabelMe image dataset. The performance can be measured in terms of precision, recall and F-measure. Three types of queries have been applied on the proposed technique i.e. Single word Single concept, Single word Multi-concept and Multi-word Multi-concept. The proposed technique has also compared against the LabelMe query System. The result of the experiments reveals that SQI shows the substantial improvement in terms of precision and outperforms the LabelMe Query system. The proposed system has been implemented by using Matlab and C# environment. The code of the proposed contribution was available in Appendix.

6.2.2 SemRank

The proposed has solved the problem of finding the relevant data from the dynamic and ever increasing colossal data corpus. But the problem of displaying the retrieved results according to the degree of relevancy between the user query and the available data is still there. Users are mostly accustomed of the top ranked results. Despite of this fact, we proposed the ranking strategy based on the semantic relevancy between the query and the data. The proposed technique is known as SemRank, the ranking refinement strategy by using the semantic intensity. We have proposed a novel concept of Semantic Intensity. Semantic Intensity is defined as the concept dominancy factor with the image or video. The Semantic Intensity intents to explore the dominancy level of all the available semantic concepts in the image. And the SemRank rank the retrieved results according to the decreasing order of the semantic intensity values i.e. the image with the greater SI value comes before the image with low SI value. The proposed technique has been compared against the well-known retrieval

model known as the vector space model that find the relevancy between the user and the document on the basis of frequency of the terms. The SemRank has been investigated on the LabelMe image dataset. The evaluation can be made in terms of precision and recall. Five randomly selected queries of three different categories i.e. single word Single concept, Single Word Multi-concept and Multi-word Multi-concept. A comparison has also been made between the VSM, SemRank and LabelMe system. The results demonstrate the effectiveness of the SemRank over the VSM and LabelMe system. The proposed system has been implemented using the Matlab and C# environment which is available in appendix.

6.2.3 Semantic Query Interpreter for Video Search and Retrieval

The surge of digital images comes along with the video surge also. After investigating the effectiveness of the proposed Semantic Query Interpreter module on the images we have extended the SQI to the video domain as well. The semantic query interpreters have been applied on the video datasets as well in order to investigate its performance on videos as well. We have applied the proposed Semantic Query Interpreter on the LabelMe video dataset. The experimental results have been made in terms of precision, recall and F-measure. The experiments have been made by selecting randomly five different queries. The experimental results show the significant performance of SQI for the videos as well. The proposed system has been implemented using the Matlab and C# environment available in appendix.

6.3 Future Perspective

The problems addressed by this thesis are very challenging. This thesis aims at providing a solution to semantic modelling and interpretation for image and video retrieval. We have tried to propose a system that better satisfy the users' demands and needs although encouraging performance has been obtained by using proposed contributions but some of the work are worth investigating and needs further extension. In this section, we discuss some of the remaining issues in our proposed solutions.

6.3.1 Semantic Query Interpreter extension

The proposed semantic query interpreter is worth to be extended by integrating the Cyc knowledgebase. The Cyc is the largest open source knowledgebase. The Cyc is not rich in conceptual reasoning like the ConceptNet and lexically rich like WordNet. But contain more information than ConceptNet and WordNet. Some of the terms that are missing in WordNet and ConceptNet will be available in Cyc. The latest version of OpenCyc, 2.0, was released in July 2009. OpenCyc 1.0 includes the entire Cyc ontology containing hundreds of thousands of terms, along with millions of assertions relating the terms to each other, however, these are mainly taxonomic assertions, not the complex rules available in Cyc. The knowledgebase contains 47,000 concepts and 306,000 facts and can be browsed on the OpenCyc website. This makes the proposed Semantic Query Interpreter more flexible.

6.3.2 Semantic Encyclopedia: An Automated Approach for Semantic Exploration and Organization of Images and Videos

The huge increase in the number of digital photos and videos generated in recent years has put even more emphasis on the task of image and video classification of unconstrained datasets. Consumer photographs, a typical example of an unconstrained dataset, comprise a significant portion of the ever increasing digital photography and video corpus. Due to their unconstrained nature and inherent diversity, consumer photographs and videos present a greater challenge for the algorithms (as they typically do for image understanding). Fortunately, digital photographs and videos usually offer a valuable additional piece of information in the form of camera metadata that complements the information extracted from the visual image and video content.

Queries often give too many results. Some of them are relevant some are irrelevant. These documents are arranged on the basis of the semantic intensity. Semantic intensity defines the semantic similarity between the query and the output result. Users are generally looking for the best video with the particular piece of information and the efficient way of

finding the particular video. Users don't want to look through hundreds of videos to locate the information.

Basic idea of the Semantic Encyclopaedia is to give the system more semantic accuracy as well the bringing the efficiency to the retrieval process. The output results are often ranked according to the semantic similarity. Different ways of ranking the documents are used like use similarity between query and the document, other often use factor to weight ranking score like ranking on the basis of visual similarity etc. or some may use iterative search which rank documents according to similarity/dissimilarity to query .After receiving the output of the initial query, get the feedback from the user as too what videos are relevant and then add words from known videos to the query. This will bring the accuracy to the system by specifying either the output videos or results are highly accurate, satisfactory or unsatisfactory. Then rank the output of the particular query according to the feedback in the Encyclopaedia so when in future if the same query is given to the system after the query interpretation the result will be directly displayed to the user from the Encyclopaedia. The already processed query record is saved in the encyclopaedia for future reference and use.

Basically the queries that are input to the system may be either already processed query, somewhat relevant to the previous queries or it may be a completely new one. The first category of the query will be after passing from the semantic query interpreter directly passed to the semantic encyclopaedia while the second one is processed with the combine effort of the semantic model and the semantic encyclopaedia and the last one is processed by the semantic model and then passed to the semantic encyclopaedia for user relevance feedback and for future use. The idea behind is to take the results that are initially returned from a given query by the semantic model and then pass the result to semantic encyclopaedia for relevance feedback from the user in order to check either the videos that are displayed are relevant to the query or not. Relevance feedback can give very substantial gain in the query formulation as well as retrieval performance. Relevance feedback usually improves average precision at the same time increases the computational work.

6.3.3 SemRank for Videos

The proposed SemRank module is currently proposed for the images. We will extend the proposed module for the videos and investigate in various datasets like TRECVID, YouTube, Open Video Project, VideoCom etc.

Appendix A

Semantic Query Interpreter

Source Code 1.1: Main Program

```
using System;
using System.Collections.Generic;
using System.Linq;
using System.Windows.Forms;

namespace Nida
{
    static class Program
    {
        /// <summary>
        /// The main entry point for the application.
        /// </summary>
        [STAThread]
        static void Main()
        {
            Application.EnableVisualStyles();
            Application.SetCompatibleTextRenderingDefault(false);
            Application.Run(new Form1());
        }
    }
}
```

Source Code 1.2: Form1 (Handling the main GUI)

```
using System;
using System.Collections.Generic;
using System.Collections;
using System.ComponentModel;
using System.Data;
using System.Drawing;
using System.Linq;
using System.Text;
using System.Windows.Forms;

using montylingua;
using ConceptNetUtils;
using WordsMatching;
using MLApp;

namespace Nida
{
    public partial class Form1 : Form
    {
        public Form1()
        {
            InitializeComponent();
        }

        #region Variable Declaration
```

```

// Structure for the SynSet with Semantic Similarity
public struct SynSet
{
    private string SySet;
    public string setSynSet
    {
        get
        {
            return SySet;
        }
        set
        {
            SySet = value;
        }
    }
    private double SemSim;
    public double setSemSim
    {
        get
        {
            return SemSim;
        }
        set
        {
            SemSim = value;
        }
    }
}

// Matlab COM component
public static MAppClass DB; // = new MAppClass();

// Structure for the Concept and Semantic Similarity Handling
public struct ConceptStruct
{
    private string Concept;
    public string setConcept
    {
        get
        {
            return Concept;
        }
        set
        {
            Concept = value;
        }
    }
    private double SemSim;
    public double setSemSim
    {
        get
        {
            return SemSim;
        }
        set
        {
            SemSim = value;
        }
    }
}
}

```



```

// Structure for the Query Handling
public struct QueryHandlers
{
    private string Word;
    public string setgetWord
    {
        get
        {
            return Word;
        }
        set
        {
            Word = value;
        }
    }
    private string POS;
    public string setgetPOS
    {
        get
        {
            return POS;
        }
        set
        {
            POS = value;
        }
    }
    private double SSAvgM;
    public double SAvgM
    {
        get
        {
            return SSAvgM;
        }
        set
        {
            SSAvgM = value;
        }
    }
    private double CSAvgM;
    public double CAvgM
    {
        get
        {
            return CSAvgM;
        }
        set
        {
            CSAvgM = value;
        }
    }
    private SynSet[] synSet;
    public SynSet[] setgetSynSet
    {
        get
        {
            return synSet;
        }
        set
        {
            synSet = value;
        }
    }
}

```

```

    }
}
private ConceptStruct[] Concept;
public ConceptStruct[] setgetConcept
{
    get
    {
        return Concept;
    }
    set
    {
        Concept = value;
    }
}
}

// Query Handler Instance
public static QueryHandlers[] QH = new QueryHandlers[100];

// Static string variables for handling string data.
public static string Query;
public static string QcandTerms;
public static string QueryOutput;

#endregion

#region Buttons Events

private void btnLexical_Click(object sender, EventArgs e)
{
    Query = tbQuery.Text;
    SupportForms.LexicalAnlaysia LA = new SupportForms.LexicalAnlaysia();
    LA.Show();
}

private void btnPython_Click(object sender, EventArgs e)
{
    tbPython.Text = "Python.status = IN PROGRESS";
    //DB = new MAppClass();
    try
    {
        QueryHandler.MontyStart();
        tbPython.Text = "Python.status = START";
    }
    catch (Exception e1)
    {
        tbPython.Text = "Python.status = ERROR : [" + e1.Source + "]-
"+e1.Message;
    }
}

private void btnSem_Click(object sender, EventArgs e)
{
    SupportForms.ConceptExtraction CE = new SupportForms.ConceptExtraction();
    CE.Show();
}

private void btnRanking_Click(object sender, EventArgs e)
{
    DB = new MAppClass();
}

```

```

        SupportForms.Matlab M = new Nida.SupportForms.Matlab();
        M.Show();
    }
    #endregion
}
}

```

Source Code 1.3: Query Handler

```

using montylingua;
using ConceptNetUtils;
using WordsMatching;
using System.Collections;
using System;
using System.Data;

namespace Nida
{
    class QueryHandler
    {
        #region Variable Definition

        public static JMontyLingua Monty;
        public static ConceptNetUtils.Search CNSearch = new ConceptNetUtils.Search();
        public static ConceptNetUtils.FoundList CNFoundList = new
ConceptNetUtils.FoundList();
        public static ConceptNetUtils.Misc CNMisc = new ConceptNetUtils.Misc();
        public static ArrayList ALFoundList = new ArrayList();
        public static string[] POS =
{" /JJ", "/NN", "/NNS", "/NNP", "/NPS", "/RB", "/RBR", "/RBT", "/RN", "/VBG", "/VBD"};

        #endregion

        #region Function Defintion and Declaration

        // MontyLingua Object Instance
        public static void MontyStart()
        {
            Monty = new JMontyLingua();
        }

        // Query Handling Tagging
        public static void QHtaging(string text)
        {
            int i = 0, a;
            string[] tok = text.Split(' ');
            string[,] duptoken = new string[30, 2], remtoken = new string[30,2];
            string str, dupstr="";

            // ----- Refreshing GridView -----//
            for (int j = 0; j <= Form1.QH.Length - 1; j++)
            {
                Form1.QH[j].setgetWord= null;
                Form1.QH[j].setgetPOS = null;
            }

            // ----- Processing TagsPOS -----//
            foreach (string t in tok)

```

```

{
    foreach (string P in POS)
    {
        if (t.Contains(P))
        {
            str = t;
            a = str.IndexOf("/");
            str = str.Substring(a+1);
            a =str.IndexOf("/");
            str = str.Substring(a+1);
            if (!str.Equals(dupstr))
            {
                dupstr = str;
                Form1.QH[i].setgetWord = str;
                Form1.QH[i].setgetPOS = P;
                i += 1;
            }
        }
    }
}

// Query Handling SynSet
public static void QHsynSet()
{
    string[] str;
    // WnLexicon.WordInfo wordinfo;// =
WnLexicon.Lexicon.FindWordInfo(txtWord.Text, chkMorphs.Checked);
    for (int i = 0; i <= Form1.QH.Length - 1; i++)
    {
        if (Form1.QH[i].setgetWord != null)
        {
            WnLexicon.WordInfo wordinfo =
WnLexicon.Lexicon.FindWordInfo(Form1.QH[i].setgetWord, true);
            if (wordinfo.partOfSpeech == Wnlib.PartsOfSpeech.Unknown)
                continue;
            else
                str = WnLexicon.Lexicon.FindSynonyms(Form1.QH[i].setgetWord,
wordinfo.partOfSpeech, true);
            Form1.QH[i].setgetSynSet = SynSet_SemSim(Form1.QH[i].setgetWord,
str);
        }
    }
}

// Query Handling Avg Means Calculation
public static void QHAvgM()
{
    double Cval = 0.00, Sval = 0.00, S = 0.00;

    for (int i = 0;(Form1.QH[i].setgetWord!=null)&(i <= Form1.QH.Length - 1);
i++)
    {
        // Semantic Similarity Average Means
        S = 0.00;
        for (int p = 0; (Form1.QH[i].setgetSynSet[p].setSynSet != null) & (p
<= Form1.QH[i].setgetSynSet.Length - 1); p++)
        {
            S += 1;
            Sval += Form1.QH[i].setgetSynSet[p].setSemSim;
        }
        if (Sval != 0.0)

```

```

        Form1.QH[i].SAvgM = Math.Round(Sval / S, 2);
    else
        Form1.QH[i].SAvgM = 0.00;

    // ConceptNet Average Means
    S = 0.00;
    for (int j = 0; j <= Form1.QH[i].setgetConcept.Length - 1; j++)
    {
        S += 1;
        Cval += Form1.QH[i].setgetConcept[j].setSemSim;
    }
    if (Cval != 0.00)
        Form1.QH[i].CAvgM = Math.Round(Cval / S, 2);
    else
        Form1.QH[i].CAvgM = 0.00;

    Cval = Sval = 0.00;
}

// Query Handling Candidate Terms Selection
public static string QHCandTerms()
{
    string synStr = "", wordStr = "", conStr = "";
    string Str;

    for (int i = 0; i <= Form1.QH.Length - 1; i++)
    {
        if (Form1.QH[i].setgetWord != null)
        {
            wordStr += Form1.QH[i].setgetWord + "(1),";

            // SynSet Candidate Terms Selection
            for (int p = 0; p <= Form1.QH[i].setgetSynSet.Length - 1; p++)
                if (Form1.QH[i].setgetSynSet[p].setSynSet != null)
                    if (Form1.QH[i].setgetSynSet[p].setSemSim > 0 &
                        Form1.QH[i].setgetSynSet[p].setSemSim >= Form1.QH[i].SAvgM)
                        synStr += Form1.QH[i].setgetSynSet[p].setSynSet +
                            "("+Form1.QH[i].setgetSynSet[p].setSemSim+",");

            // ConceptNet Candidate Terms Selection
            for (int j = 0; j <= Form1.QH[i].setgetConcept.Length - 1; j++)
                if (Form1.QH[i].setgetConcept[j].setConcept != null) // &
                    (!Form1.QH[i].setgetWord.Equals(Form1.QH[i].setgetConcept[j].setConcept)) &
                    Form1.QH[i].setgetConcept[j].setSemSim >= Form1.QH[i].setgetAvgM) //-- Work fine but
                    split is for easy 2 understand
                    {
                        if
                        (!Form1.QH[i].setgetWord.Equals(Form1.QH[i].setgetConcept[j].setConcept))
                            if (Form1.QH[i].setgetConcept[j].setSemSim > 0 &
                                Form1.QH[i].setgetConcept[j].setSemSim >= Form1.QH[i].CAvgM)
                                    conStr += Form1.QH[i].setgetConcept[j].setConcept
                                        + "("+Form1.QH[i].setgetConcept[j].setSemSim+",");
                    }
                else break;
            }
        }
        Str = wordStr + synStr + conStr;
        Str = Str.Substring(0, Str.Length - 1);
        return Str;
    }
}

```

```

// Query Handling Concept and Semantic Similarity Calculation
public static void QHConSem(string RT)
{
    for (int i = 0; i <= Form1.QH.Length - 1; i++)
    {
        Form1.QH[i].setgetConcept = Concept_SemSim(Form1.QH[i].setgetWord, RT);
    }
}

// Query Handling supporting function for extracting concepts
// from ConceptNet and semantic similarity from WordNet
public static Form1.ConceptStruct[] Concept_SemSim(string word, string
RelationType)
{
    Form1.ConceptStruct[] CS = new Form1.ConceptStruct[20];
    SentenceSimilarity SS = new SentenceSimilarity();
    string[] Concepts = new string[100];
    if (SupportForms.ConceptExtraction.XMLPath != null)
    {
        CNSearch.XMLLoadFilePaths(SupportForms.ConceptExtraction.XMLPath);
        try
        {
            //Reset List(s) to null.
            CNSearch.Clear();
            CNFoundList.Reset();
            ALFoundList.Clear();

            //If checked in one of the , Search them...
            //Preform Search using ConceptNetUtil Class Library

CNSearch.XMLSearchForChecked(SupportForms.ConceptExtraction.XMLPath, word.Trim(),
CNMisc.RemoveCategoryString(RelationType), 20, false, null);

        to lose scope***
            int numberoflines = CNSearch.GetTotalLineCount();
            for (int j = 0; j < numberoflines; j++)
            {
                //Copy into a global ArrayList
                ALFoundList.Add(CNSearch.GetFoundListLine(j));

                //Copy into a global CNFoundList
                // CNFoundList[j] = CNSearch.GetFoundListLine(j);
            }

            System.Collections.IEnumerator myEnumerator =
ALFoundList.GetEnumerator();
            int a, k = 0;
            string st;
            while (myEnumerator.MoveNext())
            {
                st = myEnumerator.Current.ToString();
                while (st.Length > 0)
                {
                    try
                    {
                        a = st.IndexOf('('); a++; st = st.Substring(a);
                        a = st.IndexOf(' '); a++; st = st.Substring(a);
                        a = st.IndexOf(' '); Concepts[k++] = st.Substring(0,
a); a++; st = st.Substring(a);
                        a = st.IndexOf(' '); a++; st = st.Substring(a);

```



```

    }
    return SSet;
}
#endregion

#region Gridview

// Gridview showing only Word and POS
public static DataTable GDtaging()
{
    DataTable Pir = new DataTable("ConceptList");
    DataColumn Words = new DataColumn("Word");
    DataColumn POS = new DataColumn("POS");
    Pir.Columns.Add(Words);
    Pir.Columns.Add(POS);
    DataRow newRow;
    for (int i = 0; i <= Form1.QH.Length - 1; i++)
    {
        if (Form1.QH[i].setgetWord != null)
        {
            newRow = Pir.NewRow();
            newRow["Word"] = Form1.QH[i].setgetWord;
            newRow["POS"] = Form1.QH[i].setgetPOS;
            Pir.Rows.Add(newRow);
        }
    }
    return Pir;
}

// Gridview showing Word, POS, SynSet
public static DataTable GDSynSet()
{
    DataTable Pir = new DataTable("ConceptList");
    DataColumn Words = new DataColumn("Word");
    DataColumn POS = new DataColumn("POS");
    DataColumn Syn = new DataColumn("SynSet");
    Pir.Columns.Add(Words);
    Pir.Columns.Add(POS);
    Pir.Columns.Add(Syn);
    DataRow newRow;
    string str = "";

    for (int i = 0; i <= Form1.QH.Length - 1; i++)
    {
        if (Form1.QH[i].setgetWord != null)
        {
            newRow = Pir.NewRow();
            newRow["Word"] = Form1.QH[i].setgetWord;
            newRow["POS"] = Form1.QH[i].setgetPOS;
            str = "";
            for (int p = 0; p <= Form1.QH[i].setgetSynSet.Length - 1; p++)
            {
                if (Form1.QH[i].setgetSynSet[p].setSynSet != null)
                    str += Form1.QH[i].setgetSynSet[p].setSynSet + "(" +
Form1.QH[i].setgetSynSet[p].setSemSim + ")", ";
            }
            newRow["SynSet"] = str;
            Pir.Rows.Add(newRow);
        }
    }
    return Pir;
}
}

```



```

// Gridview showing Word, POS, SynSet, Concept and Semantic Similarity
public static DataTable GDConSem()
{
    DataTable Pir = new DataTable("ConceptList");
    DataColumn Words = new DataColumn("Word");
    DataColumn POS = new DataColumn("POS");
    DataColumn Syn = new DataColumn("SynSet");
    DataColumn Conc = new DataColumn("Concept(SS)");
    Pir.Columns.Add(Words);
    Pir.Columns.Add(POS);
    Pir.Columns.Add(Syn);
    Pir.Columns.Add(Conc);
    DataRow newRow;
    string str="";

    for (int i = 0; i <= 100 - 1; i++)
    {
        if (Form1.QH[i].setgetWord != null)
        {
            newRow = Pir.NewRow();
            newRow["Word"] = Form1.QH[i].setgetWord;
            newRow["POS"] = Form1.QH[i].setgetPOS;
            str = "";
            for (int k = 0; k <= Form1.QH[i].setgetSynSet.Length - 1; k++)
            {
                if (Form1.QH[i].setgetSynSet[k].setSynSet != null)
                    str += Form1.QH[i].setgetSynSet[k].setSynSet + "(" +
Form1.QH[i].setgetSynSet[k].setSemSim + "),"";
            }
            newRow["SynSet"] = str;
            str = "";
            for (int p = 0; p <= Form1.QH[i].setgetConcept.Length - 1; p++)
            {
                if (Form1.QH[i].setgetConcept[p].setConcept != null)
                    str += Form1.QH[i].setgetConcept[p].setConcept + "(" +
Form1.QH[i].setgetConcept[p].setSemSim + "),"";
            }
            newRow["Concept(SS)"] = str;
            Pir.Rows.Add(newRow);
        }
    }
    return Pir;
}

// Gridview showing All data of the Query Handler QH
public static DataTable GDAvgM()
{
    DataTable Pir = new DataTable("ConceptList");
    DataColumn Words = new DataColumn("Word");
    DataColumn POS = new DataColumn("POS");
    DataColumn Syn = new DataColumn("SynSet");
    DataColumn Conc = new DataColumn("Concept(SS)");
    DataColumn SAvg = new DataColumn("S-AvgM");
    DataColumn CAvg = new DataColumn("C-AvgM");
    Pir.Columns.Add(Words);
    Pir.Columns.Add(POS);
    Pir.Columns.Add(SAvg);
    Pir.Columns.Add(CAvg);
    Pir.Columns.Add(Syn);
    Pir.Columns.Add(Conc);
}

```

```

DataRow newRow;
string str;

for (int i = 0; i <= 100 - 1; i++)
{
    if (Form1.QH[i].setgetWord != null)
    {
        newRow = Pir.NewRow();
        newRow["Word"] = Form1.QH[i].setgetWord;
        newRow["POS"] = Form1.QH[i].setgetPOS;
        newRow["S-AvgM"] = Form1.QH[i].SAvgM;
        newRow["C-AvgM"] = Form1.QH[i].CAvgM;

        str = "";
        for (int k = 0; k <= Form1.QH[i].setgetSynSet.Length - 1; k++)
            if (Form1.QH[i].setgetSynSet[k].setSynSet != null)
                str += Form1.QH[i].setgetSynSet[k].setSynSet + "(" +
Form1.QH[i].setgetSynSet[k].setSemSim + "),"";
        newRow["SynSet"] = str;

        str = "";
        for (int p = 0; p <= Form1.QH[i].setgetConcept.Length - 1; p++)
            if (Form1.QH[i].setgetConcept[p].setConcept != null)
                str += Form1.QH[i].setgetConcept[p].setConcept + "(" +
Form1.QH[i].setgetConcept[p].setSemSim + "),"";
        newRow["Concept(SS)"] = str;

        Pir.Rows.Add(newRow);
    }
}
return Pir;
}

#endregion
}
}

```

Lexical Expansion

For lexical expansion of the query, we use the WordNet. For this purpose, the function have been taken from the code project written by Tunah available freely under GNU license for research purpose. The function that we have used for the research purpose and query expansion in lexical dimension are

We have used the following supporting code for WordNet, ConceptNet and Montylingua for the research purpose, all these code are available openly for the research purposes. Next we will describe the supporting tools one/one

1. WordNet Supporting tools:

For WordNet support, we have selected the tools from the code project written by Tunaah, for sentence similarity, word ambiguity and semantic similarity among the words. The functions that are used during the research process are

- a. ISimilarity.cs
- b. Relatedness.cs
- c. SentenceSimilarity.cs
- d. SimilarGenerator.cs
- e. WordSenseDisambiguity.cs
- f. WordSimilarity.cs
- g. Matcher.BipartiteMatcher.cs
- h. Matcher.HeuristicMatcher.cs
- i. TextHelper.Acronym.cs
- j. TexHelpre.ExtOverlapCounter.cs
- k. TextHelper.StopWordsHandler.cs
- l. TextHelper.Tokeniser.cs

These function are jointly used to calculate the semantic similarity among the words.

The source code for the semantic similarity are

For Lexical Analysis the following group of functions are used

```
using System;
using System.Collections.Generic;
using System.ComponentModel;
using System.Data;
using System.Drawing;
using System.Linq;
using System.Text;
using System.Windows.Forms;

using montylingua;

namespace Nida.SupportForms
{
    public partial class LexicalAnlaysia : Form
    {
        public LexicalAnlaysia()
        {
            InitializeComponent();
            tbOrigQ.Text = Form1.Query;
            Wnlib.WNCommon.path = "C:\\Program Files\\WordNet\\2.1\\dict\\";
        }

        #region Variable and Function
```

```

public string Tokenize()
{
    string Tokens="";
    string[] Token = tbOrigQ.Text.Split(' ');
    foreach (string word in Token)
    {
        Tokens += " [ " + word + " ]";
    }
    return Tokens;
}

#endregion

#region Buttons Events

private void btnToken_Click(object sender, EventArgs e)
{
    tbToken.Text = Tokenize();
}

private void btnLema_Click(object sender, EventArgs e)
{
    try
    {
        tbLemmatize.Text = QueryHandler.Monty.lemmatise_text(tbOrigQ.Text);
    }
    catch
    {
        MessageBox.Show("Lemmatization Problem, it may be due to proxy server
still down", "Lemmatizer Error ");
    }
}

private void btnPOS_Click(object sender, EventArgs e)
{
    try
    {
        tbPOS.Text = QueryHandler.Monty.tag_text(tbOrigQ.Text);
    }
    catch
    {
        MessageBox.Show("POS Problem, it may be due to proxy server still down
", " POS Error");
    }
}

private void btnConceptSel_Click(object sender, EventArgs e)
{
    string LemaText;
    try
    {
        LemaText =QueryHandler.Monty.lemmatise_text(tbOrigQ.Text);
        QueryHandler.QHtaging(LemaText);
        dataGridView1.DataSource = QueryHandler.GDtaging();
    }
    catch
    {
        MessageBox.Show("Concept Selection Problem, it may be due to proxy
server still down ", " Concept Selection ");
    }
}

```

```

private void btnWNSynset_Click(object sender, EventArgs e)
{
    try
    {
        QueryHandler.QHsynSet();
        dataGridView1.DataSource = QueryHandler.GDsynSet();
    }
    catch
    {
        MessageBox.Show("Tagging POS is empty", "Sysnset");
    }
}

#endregion
}
}

```

ConceptNet: The Code for this module is taken from the code project openly available for research purposes; we have modified the coder as per our requirements. The snapshot of the source code is under. These code are written for ConceptNet 2.1 version.

Function: Handling the Commonsensical Expansion and Candidate Concept Selection

```

////////////////////////////////////
//Form1.cs - version 0.01412006.0rc4
//BY DOWNLOADING AND USING, YOU AGREE TO THE FOLLOWING TERMS:
//Copyright (c) 2006 by Joseph P. Socoloski III
//LICENSE
//If it is your intent to use this software for non-commercial purposes,
//such as in academic research, this software is free and is covered under
//the GNU GPL License, given here: <http://www.gnu.org/licenses/gpl.txt>
//
using System;
using System.Drawing;
using System.Collections;
using System.ComponentModel;
using System.Windows.Forms;
using System.Data;
using ConceptNetUtils;

namespace Nida.SupportForms
{
    /// <summary>
    /// Summary description for Form1.
    /// </summary>
    public class ConceptExtraction : System.Windows.Forms.Form
    {
        private System.Windows.Forms.Label label2;
        private System.Windows.Forms.ComboBox cbRelationshipTypes;
        private System.Windows.Forms.Button btSearch;
        private System.ComponentModel.IContainer components;

        #region Variables and Functions

        //Initialize ConceptNetUtils
        ConceptNetUtils.Search CNSearch = new ConceptNetUtils.Search();

```

```

        ConceptNetUtils.FoundList CNFoundList = new ConceptNetUtils.FoundList();
ConceptNetUtils.Misc CNMisc = new ConceptNetUtils.Misc();
private BindingSource mLAppClassBindingSource;
private Panel panel2;
private TableLayoutPanel tableLayoutPanel1;
private Button button2;
private DataGridView dataGridView2;
private Button button1;
    ArrayList ALFoundList = new ArrayList();
private Button btnCandiTermSel;
private Button btnAvgMeans;
private TextBox tbOutput;
private Label label1;

public static string XMLPath="";

    public ConceptExtraction()
    {
        //
        // Required for Windows Form Designer support
        //
        InitializeComponent();

        //
        // TODO: Add any constructor code after InitializeComponent call
        //
    }

#endregion

/// <summary>
/// Clean up any resources being used.
/// </summary>
protected override void Dispose( bool disposing )
{
    if( disposing )
    {
        if (components != null)
        {
            components.Dispose();
        }
    }
    base.Dispose( disposing );
}

#region Windows Form Designer generated code
/// <summary>
/// Required method for Designer support - do not modify
/// the contents of this method with the code editor.
/// </summary>
private void InitializeComponent()
{
    this.components = new System.ComponentModel.Container();
    System.ComponentModel.ComponentResourceManager resources = new
System.ComponentModel.ComponentResourceManager(typeof(ConceptExtraction));
    this.label2 = new System.Windows.Forms.Label();
    this.cbRelationshipTypes = new System.Windows.Forms.ComboBox();
    this.panel2 = new System.Windows.Forms.Panel();
    this.tbOutput = new System.Windows.Forms.TextBox();

```

```

        this.dataGridView2 = new System.Windows.Forms.DataGridView();
        this.tableLayoutPanel1 = new System.Windows.Forms.TableLayoutPanel();
        this.btnCandiTermSel = new System.Windows.Forms.Button();
        this.btnAvgMeans = new System.Windows.Forms.Button();
        this.button2 = new System.Windows.Forms.Button();
        this.btSearch = new System.Windows.Forms.Button();
        this.button1 = new System.Windows.Forms.Button();
        this.label1 = new System.Windows.Forms.Label();
        this.mLAppClassBindingSource = new
System.Windows.Forms.BindingSource(this.components);
        this.panel2.SuspendLayout();

((System.ComponentModel.ISupportInitialize)(this.dataGridView2)).BeginInit();
        this.tableLayoutPanel1.SuspendLayout();

((System.ComponentModel.ISupportInitialize)(this.mLAppClassBindingSource)).BeginInit()
;
        this.SuspendLayout();
        //
        // label2
        //
        this.label2.Font = new System.Drawing.Font("Arial", 12F,
System.Drawing.FontStyle.Bold, System.Drawing.GraphicsUnit.Point, ((byte)0));
        this.label2.Location = new System.Drawing.Point(3, 6);
        this.label2.Name = "label2";
        this.label2.Size = new System.Drawing.Size(469, 33);
        this.label2.TabIndex = 4;
        this.label2.Text = "Select Relationship Type";
        //
        // cbRelationshipTypes
        //
        this.cbRelationshipTypes.DropDownStyle =
System.Windows.Forms.ComboBoxStyle.DropDownList;
        this.cbRelationshipTypes.Font = new System.Drawing.Font("Arial", 12F,
System.Drawing.FontStyle.Bold, System.Drawing.GraphicsUnit.Point, ((byte)0));
        this.cbRelationshipTypes.ImeMode = System.Windows.Forms.ImeMode.NoControl;
        this.cbRelationshipTypes.Items.AddRange(new object[] {
        "K-Lines: ConceptuallyRelatedTo",
        "K-Lines: ThematicKLine",
        "K-Lines: SuperThematicKLine",
        "All K-Lines",
        "Things: IsA",
        "Things: PartOf",
        "Things: PropertyOf",
        "Things: DefinedAs",
        "Things: MadeOf",
        "All Things",
        "Spatial: LocationOf",
        "Events: SubeventOf",
        "Events: PrerequisiteEventOf",
        "Events: First-SubeventOf",
        "Events: LastSubeventOf",
        "All Events",
        "Causal: EffectOf",
        "Causal: DesirousEffectOf",
        "All Causal",
        "Affective: MotivationOf",
        "Affective: DesireOf",
        "All Affective",
        "Functional: CapableOfReceivingAction",
        "Functional: UsedFor",
        "All Functional",

```

```

        "Agents: CapableOf",
        "All (Returns all results with word)"});
        this.cbRelationshipTypes.Location = new System.Drawing.Point(10, 44);
        this.cbRelationshipTypes.Name = "cbRelationshipTypes";
        this.cbRelationshipTypes.RightToLeft =
System.Windows.Forms.RightToLeft.No;
        this.cbRelationshipTypes.Size = new System.Drawing.Size(496, 37);
        this.cbRelationshipTypes.TabIndex = 5;
        //
        // panel2
        //
        this.panel2.BackColor =
System.Drawing.Color.FromArgb(((int)(((byte)(192)))), ((int)(((byte)(192)))),
((int)(((byte)(255)))));
        this.panel2.BorderStyle = System.Windows.Forms.BorderStyle.Fixed3D;
        this.panel2.Controls.Add(this.tbOutput);
        this.panel2.Controls.Add(this.dataGridView2);
        this.panel2.Controls.Add(this.tableLayoutPanel1);
        this.panel2.Controls.Add(this.cbRelationshipTypes);
        this.panel2.Controls.Add(this.label2);
        this.panel2.Location = new System.Drawing.Point(19, 92);
        this.panel2.Name = "panel2";
        this.panel2.Size = new System.Drawing.Size(1619, 909);
        this.panel2.TabIndex = 14;
        //
        // tbOutput
        //
        this.tbOutput.BackColor = System.Drawing.Color.Tan;
        this.tbOutput.Font = new System.Drawing.Font("Arial", 9.75F,
System.Drawing.FontStyle.Bold, System.Drawing.GraphicsUnit.Point, ((byte)(0)));
        this.tbOutput.Location = new System.Drawing.Point(515, 740);
        this.tbOutput.Multiline = true;
        this.tbOutput.Name = "tbOutput";
        this.tbOutput.ScrollBars = System.Windows.Forms.ScrollBars.Vertical;
        this.tbOutput.Size = new System.Drawing.Size(1093, 153);
        this.tbOutput.TabIndex = 17;
        this.tbOutput.Text = resources.GetString("tbOutput.Text");
        //
        // dataGridView2
        //
        this.dataGridView2.AllowUserToOrderColumns = true;
        this.dataGridView2.ColumnHeadersHeightSizeMode =
System.Windows.Forms.DataGridViewColumnHeadersHeightSizeMode.AutoSize;
        this.dataGridView2.Location = new System.Drawing.Point(520, 44);
        this.dataGridView2.Name = "dataGridView2";
        this.dataGridView2.Size = new System.Drawing.Size(1088, 687);
        this.dataGridView2.TabIndex = 16;
        //
        // tableLayoutPanel1
        //
        this.tableLayoutPanel1.ColumnCount = 1;
        this.tableLayoutPanel1.ColumnStyles.Add(new
System.Windows.Forms.ColumnStyle(System.Windows.Forms.SizeType.Percent, 100F));
        this.tableLayoutPanel1.Controls.Add(this.btnCandiTermSel, 0, 4);
        this.tableLayoutPanel1.Controls.Add(this.btnAvgMeans, 0, 3);
        this.tableLayoutPanel1.Controls.Add(this.button2, 0, 2);
        this.tableLayoutPanel1.Controls.Add(this.btSearch, 0, 1);
        this.tableLayoutPanel1.Controls.Add(this.button1, 0, 0);
        this.tableLayoutPanel1.Location = new System.Drawing.Point(5, 92);
        this.tableLayoutPanel1.Name = "tableLayoutPanel1";
        this.tableLayoutPanel1.RowCount = 5;

```



```

        this.tableLayoutPanel1.RowStyles.Add(new
System.Windows.Forms.RowStyle(System.Windows.Forms.SizeType.Percent, 20F));
        this.tableLayoutPanel1.RowStyles.Add(new
System.Windows.Forms.RowStyle(System.Windows.Forms.SizeType.Percent, 20F));
        this.tableLayoutPanel1.RowStyles.Add(new
System.Windows.Forms.RowStyle(System.Windows.Forms.SizeType.Percent, 20F));
        this.tableLayoutPanel1.RowStyles.Add(new
System.Windows.Forms.RowStyle(System.Windows.Forms.SizeType.Percent, 20F));
        this.tableLayoutPanel1.RowStyles.Add(new
System.Windows.Forms.RowStyle(System.Windows.Forms.SizeType.Percent, 20F));
        this.tableLayoutPanel1.Size = new System.Drawing.Size(505, 801);
        this.tableLayoutPanel1.TabIndex = 15;
        //
        // btnCandiTermSel
        //
        this.btnCandiTermSel.Font = new System.Drawing.Font("Arial", 12F,
System.Drawing.FontStyle.Bold, System.Drawing.GraphicsUnit.Point, ((byte)0));
        this.btnCandiTermSel.Image = global::Nida.Properties.Resources.Database;
        this.btnCandiTermSel.ImageAlign =
System.Drawing.ContentAlignment.MiddleLeft;
        this.btnCandiTermSel.Location = new System.Drawing.Point(3, 643);
        this.btnCandiTermSel.Name = "btnCandiTermSel";
        this.btnCandiTermSel.Size = new System.Drawing.Size(496, 152);
        this.btnCandiTermSel.TabIndex = 16;
        this.btnCandiTermSel.Text = "4. Candidate Concept";
        this.btnCandiTermSel.TextAlign =
System.Drawing.ContentAlignment.MiddleRight;
        this.btnCandiTermSel.UseVisualStyleBackColor = true;
        this.btnCandiTermSel.Click += new
System.EventHandler(this.btnCandiTermSel_Click);
        //
        // btnAvgMeans
        //
        this.btnAvgMeans.Font = new System.Drawing.Font("Arial", 12F,
System.Drawing.FontStyle.Bold, System.Drawing.GraphicsUnit.Point, ((byte)0));
        this.btnAvgMeans.Image =
global::Nida.Properties.Resources.Semantic_Intensity1;
        this.btnAvgMeans.ImageAlign = System.Drawing.ContentAlignment.MiddleLeft;
        this.btnAvgMeans.Location = new System.Drawing.Point(3, 483);
        this.btnAvgMeans.Name = "btnAvgMeans";
        this.btnAvgMeans.Size = new System.Drawing.Size(496, 151);
        this.btnAvgMeans.TabIndex = 17;
        this.btnAvgMeans.Text = "3. Avg. Means";
        this.btnAvgMeans.TextAlign = System.Drawing.ContentAlignment.MiddleRight;
        this.btnAvgMeans.UseVisualStyleBackColor = true;
        this.btnAvgMeans.Click += new System.EventHandler(this.btnAvgMeans_Click);
        //
        // button2
        //
        this.button2.BackColor = System.Drawing.Color.Transparent;
        this.button2.Font = new System.Drawing.Font("Arial", 12F,
System.Drawing.FontStyle.Bold, System.Drawing.GraphicsUnit.Point, ((byte)0));
        this.button2.Image = global::Nida.Properties.Resources.Main_Concepts;
        this.button2.ImageAlign = System.Drawing.ContentAlignment.MiddleLeft;
        this.button2.Location = new System.Drawing.Point(3, 323);
        this.button2.Name = "button2";
        this.button2.Size = new System.Drawing.Size(496, 151);
        this.button2.TabIndex = 14;
        this.button2.Text = "2. Concept(s) Selection";
        this.button2.TextAlign = System.Drawing.ContentAlignment.MiddleRight;
        this.button2.UseVisualStyleBackColor = false;
        this.button2.Click += new System.EventHandler(this.button2_Click);

```

```

//
// btSearch
//
this.btSearch.BackColor = System.Drawing.Color.Transparent;
this.btSearch.Font = new System.Drawing.Font("Arial", 12F,
System.Drawing.FontStyle.Bold, System.Drawing.GraphicsUnit.Point, ((byte)0));
this.btSearch.Image = global::Nida.Properties.Resources.Result41;
this.btSearch.ImageAlign = System.Drawing.ContentAlignment.MiddleLeft;
this.btSearch.Location = new System.Drawing.Point(3, 163);
this.btSearch.Name = "btSearch";
this.btSearch.Size = new System.Drawing.Size(496, 150);
this.btSearch.TabIndex = 10;
this.btSearch.Text = "1. Selected Term(s)";
this.btSearch.TextAlign = System.Drawing.ContentAlignment.MiddleRight;
this.btSearch.UseVisualStyleBackColor = false;
this.btSearch.Click += new System.EventHandler(this.btSearch_Click);
//
// button1
//
this.button1.Font = new System.Drawing.Font("Arial", 12F,
System.Drawing.FontStyle.Bold, System.Drawing.GraphicsUnit.Point, ((byte)0));
this.button1.Image = global::Nida.Properties.Resources.Concepts;
this.button1.ImageAlign = System.Drawing.ContentAlignment.MiddleLeft;
this.button1.Location = new System.Drawing.Point(3, 3);
this.button1.Name = "button1";
this.button1.Size = new System.Drawing.Size(496, 151);
this.button1.TabIndex = 15;
this.button1.Text = "0. ConceptNet";
this.button1.TextAlign = System.Drawing.ContentAlignment.MiddleRight;
this.button1.UseVisualStyleBackColor = true;
this.button1.Click += new System.EventHandler(this.button1_Click);
//
// label1
//
this.label1.AutoSize = true;
this.label1.Font = new System.Drawing.Font("Arial", 24F,
System.Drawing.FontStyle.Bold, System.Drawing.GraphicsUnit.Point, ((byte)0));
this.label1.Location = new System.Drawing.Point(413, 13);
this.label1.Name = "label1";
this.label1.Size = new System.Drawing.Size(523, 56);
this.label1.TabIndex = 15;
this.label1.Text = "Concept(s) Extraction";
//
// ConceptExtraction
//
this.AutoScaleBaseSize = new System.Drawing.Size(8, 19);
this.ClientSize = new System.Drawing.Size(1036, 708);
this.Controls.Add(this.label1);
this.Controls.Add(this.panel2);
this.Name = "ConceptExtraction";
this.StartPosition = System.Windows.Forms.FormStartPosition.CenterScreen;
this.Text = "Concept Extraction";
this.Load += new System.EventHandler(this.Form1_Load);
this.panel2.ResumeLayout(false);
this.panel2.PerformLayout();

((System.ComponentModel.ISupportInitialize)(this.dataGridView2)).EndInit();
this.tableLayoutPanel1.ResumeLayout(false);

((System.ComponentModel.ISupportInitialize)(this.mLAppClassBindingSource)).EndInit();
this.ResumeLayout(false);
this.PerformLayout();

```

```

    }
    #endregion

    /// <summary>
    /// The main entry point for the application.
    /// </summary>

    private void Form1_Load(object sender, System.EventArgs e)
    {
        //Select the first item in the Relationship Types ComboBox
        cbRelationshipTypes.SelectedIndex = 0;
    }

    private void btSearch_Click(object sender, System.EventArgs e)
    {
        dataGridView2.DataSource = QueryHandler.GDsynSet();
    }

    private void button2_Click(object sender, EventArgs e)
    {
        QueryHandler.QHConSem(cbRelationshipTypes.Text);
        dataGridView2.DataSource = QueryHandler.GDConSem();
    }

    private void button1_Click(object sender, EventArgs e)
    {
        OpenFileDialog FD = new OpenFileDialog();
        FD.InitialDirectory =
        Environment.GetFolderPath(Environment.SpecialFolder.Personal);
        FD.ShowDialog();
        XMLPath = FD.FileName;
    }

    private void btnAvgMeans_Click(object sender, EventArgs e)
    {
        QueryHandler.QHAvgM();
        dataGridView2.DataSource = QueryHandler.GDAvgM();
    }

    private void btnCandiTermSel_Click(object sender, EventArgs e)
    {
        Form1.QueryOutput = tbOutput.Text = Form1.QCandTerms =
        QueryHandler.QHCandTerms();
    }
}

```

Matlab: As per requirement of the research, some of our work is perform in Matlab, while for some C# tool is used. We have used the utility MLApp for C# to call the Matlab function. Further, we have handled the Matlab function execution through threading process. The source code for different purpose performs in the Matlab is (for the Matlab function are giving under the head of Matlab code). The following are the complete set of functions that is used to handle the processing between Matlab and C# environment.

Source Code 1.6: Interfacing with Matlab

```
using System;
using System.Collections.Generic;
using System.ComponentModel;
using System.Data;
using System.Drawing;
using System.Linq;
using System.Text;
using System.Windows.Forms;
using MLApp;
using System.Threading;

namespace Nida.SupportForms
{
    public partial class Matlab : Form
    {
        public Matlab()
        {
            InitializeComponent();
            tbQuery.Text = Form1.QueryOutput;
        }

        #region Matlab Functions

        public static string ConExt;
        public static string outPut;
        // MLApp.MLAppClass DB = new MLAppClass();
        public void path()
        {
            // Define Directories Path
            Nida.Form1.DB.Execute("setImagePath(' + tbHI.Text + "')");
            Nida.Form1.DB.Execute("setAnnotationPath(' + tbHA.Text + "')");
        }

        #endregion

        #region Button Events

        private void btnDBCreation_Click(object sender, EventArgs e)
        {
            string st;
            // Setting paths for images and Annotations
            tbReport.Text = "Path Setting";
            path();
            tbReport.Text = "Database Creation in progress";
            st = Nida.Form1.DB.Execute("QI_DBCreation");
            int a = st.IndexOf("@");
            int b = st.IndexOf("#");
            try
            {
                tbReport.Text = st.Substring(a + 1, b - a - 1);
            }
            catch { }
        }

        private void btnHI_Click(object sender, EventArgs e)
        {
            FolderBrowserDialog fd = new FolderBrowserDialog();
            fd.ShowDialog();
            tbHI.Text = fd.SelectedPath.ToString();
        }
    }
}
```

```

}

private void btnHA_Click(object sender, EventArgs e)
{
    FolderBrowserDialog fd = new FolderBrowserDialog();
    fd.ShowDialog();
    tbHA.Text = fd.SelectedPath.ToString();
}

private void btnResult_Click(object sender, EventArgs e)
{
    string st;
    tbReport.Text = "";
    // Setting paths for images and Annotations
    path();

    // Calling Matlab function
    int a = Convert.ToInt32(tbstart.Text), b = Convert.ToInt32(tblast.Text);
    try
    {
        tbReport.Text = "Result Display in progress...";
        st = Nida.Form1.DB.Execute("QI_resultDisplay(" + a + "," + b + ")");
        a = st.IndexOf("@");
        b = st.IndexOf("#");
        tbReport.Text = st.Substring(a + 1, b - a - 1);
    }
    catch { }
}

private void btnQuery_Click(object sender, EventArgs e)
{
    string s = tbQuery.Text, st = "";
    path();
    int a = 0, c = 0;
    bool b = true;
    a = s.IndexOf("");
    while (b)
    {
        st += s.Substring(0, a);
        s = s.Substring(a);
        // a = -10;
        a = s.IndexOf("");
        if (a <= 0) b = false;
    }
    if (s.Length > 0) st += s.Substring(1, s.Length - 1);
    tbQuery.Text = st;
    st = "";
    tbReport.Text = "Query in progress...";
    if (cboxRank.SelectedItem.ToString() == "VSM")
        st = Nida.Form1.DB.Execute("QueryInterpreter('" + tbQuery.Text + "',' +
1 + ")");
    else if (cboxRank.SelectedItem.ToString() == "SIRRS")
        st = Nida.Form1.DB.Execute("QueryInterpreter('" + tbQuery.Text + "',' +
2 + ")");

    try
    {
        a = st.IndexOf("@");
        c = st.IndexOf("#");
        tbReport.Text = st.Substring(a + 1, c - a - 1);
    }
    catch { }
}

```

```

}
#endregion

private void button1_Click(object sender, EventArgs e)
{
    string s = tbQuery.Text, st = "";
    int a = 0;
    bool b = true;
    a = s.IndexOf("");
    while (b)
    {
        st += s.Substring(0, a);
        s = s.Substring(a);
        // a = -10;
        a = s.IndexOf("");
        if (a <= 0) b = false;
    }
    if (s.Length > 0) st += s.Substring(1,s.Length-1);
    tbQuery.Text = st;
}
}
}
}

```

Source Code from MATLAB

Source Code 2.1: For database creation

```

function Report = QI_DBCreation
global DB HA;
DB = QI_LMdatabase(HA);
Report = 'Database creation completed';

```

Source Code 2.2: QI_LMDatabase

```

function [D, XML] = QI_LMdatabase(varargin)
%function [database, XML] = LMdatabase(HOMEANNOTATIONS, folderlist)
%
% This line reads the entire database into a Matlab struct.
%
% Different ways of calling this function
% D = LMdatabase(HOMEANNOTATIONS); % reads only annotated images
% D = LMdatabase(HOMEANNOTATIONS, HOMEIMAGES); % reads all images
% D = LMdatabase(HOMEANNOTATIONS, folderlist);
% D = LMdatabase(HOMEANNOTATIONS, HOMEIMAGES, folderlist);
% D = LMdatabase(HOMEANNOTATIONS, HOMEIMAGES, folderlist, filelist);
%
% Reads all the annotations.
% It creates a struct 'almost' equivalent to what you would get if you
concatenate
% first all the xml files, then you add at the beggining the tag <D> and at
the end </D>
% and then use loadXML.m
%
% You do not need to download the database. The functions that read the
% images and the annotation files can be refered to the online tool. For
% instance, you can run the next command:
%
% HOMEANNOTATIONS = 'http://labelme.csail.mit.edu/Annotations'
% D = LMdatabase(HOMEANNOTATIONS);

```

```

%
% This will create the database struct without needing to download the
% database. It might be slower than having a local copy.
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% LabelMe, the open annotation tool
% Contribute to the database by labeling objects using the annotation tool.
% http://labelme.csail.mit.edu/
%
% CSAIL, MIT
% 2006
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% LabelMe is a WEB-based image annotation tool and a Matlab toolbox that
allows
% researchers to label images and share the annotations with the rest of
the community.
% Copyright (C) 2007 MIT, Computer Science and Artificial
% Intelligence Laboratory. Antonio Torralba, Bryan Russell, William T.
Freeman
%
% This program is free software: you can redistribute it and/or modify
% it under the terms of the GNU General Public License as published by
% the Free Software Foundation, either version 3 of the License, or
% (at your option) any later version.
%
% This program is distributed in the hope that it will be useful,
% but WITHOUT ANY WARRANTY; without even the implied warranty of
% MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the
% GNU General Public License for more details.
%
% You should have received a copy of the GNU General Public License
% along with this program. If not, see <http://www.gnu.org/licenses/>.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% This function removes all the deleted polygons. If you want to read them
% too, you have to comment line (at the end): D = LMvalidobjects(D);

Folder = [];

% Parse input arguments and read list of folders
Narg = nargin;
HOMEANNOTATIONS = varargin{1};
if Narg==3
    HOMEIMAGES = varargin{2};
else
    HOMEIMAGES = '';
end

if iscell(varargin{Narg})
    if Narg == 2
        Folder = varargin{2};
        Nfolders = length(Folder);
    end
    if Narg == 3
        Folder = varargin{3};
        Nfolders = length(Folder);
    end
    if Narg == 4

```

```

        Folder = varargin{3};
        Images = varargin{4};
        Nfolders = length(Folder);
    end
else
    if Narg==2
        HOMEIMAGES = varargin{2};
    end
    if ~strcmp(HOMEANNOTATIONS(1:5), 'http:');
        folders = genpath(HOMEANNOTATIONS);
        h = [findstr(folders, pathsep)];
        h = [0 h];
        Nfolders = length(h)-1;
        for i = 1:Nfolders
            tmp = folders(h(i)+1:h(i+1)-1);
            tmp = strrep(tmp, HOMEANNOTATIONS, ''); tmp = tmp(2:end);
            Folder{i} = tmp;
        end
    else
        files = urldir(HOMEANNOTATIONS);
        Folder = {files(2:end).name}; % the first item is the main path
name
        Nfolders = length(Folder);
        %for i = 1:Nfolders
        %     Folder{i} = Folder{i};
        %end
    end
end

% Open figure that visualizes the file and folder counter
Hfig = plotbar;

% Loop on folders
D = []; n = 0; nPolygons = 0;
if nargout == 2; XML = ['<database>']; end
for f = 1:Nfolders
    folder = Folder{f};
    disp(sprintf('%d/%d, %s', f, Nfolders, folder))

    if Narg<4
        filesImages = [];
        if ~strcmp(HOMEANNOTATIONS(1:5), 'http:');
            filesAnnotations = dir(fullfile(HOMEANNOTATIONS, folder,
'*.xml'));
            if ~isempty(HOMEIMAGES)
                filesImages = dir(fullfile(HOMEIMAGES, folder, '*.jpg'));
            end
        else
            filesAnnotations = urlxmldir(fullfile(HOMEANNOTATIONS,
folder));
            if ~isempty(HOMEIMAGES)
                filesImages = urldir(fullfile(HOMEIMAGES, folder), 'img');
            end
        end
    else
        filesAnnotations(1).name = strrep(Images{f}, '.jpg', '.xml');
        filesAnnotations(1).bytes = 1;
        filesImages(1).name = strrep(Images{f}, '.xml', '.jpg');
    end
end

```



```

%keyboard

if ~isempty(HOMEIMAGES)
    N = length(filesImages);
else
    N = length(filesAnnotations);
end

fprintf(1, '%d ', N)
emptyAnnotationFiles = 0;
labeledImages = 0;
for i = 1:N
    clear v
    if ~isempty(HOMEIMAGES)
        filename = fullfile(HOMEIMAGES, folder, filesImages(i).name);
        filenameanno = strrep(filesImages(i).name, '.jpg', '.xml');
        if ~isempty(filesAnnotations)
            J = strmatch(filenameanno, {filesAnnotations(:).name});
        else
            J = [];
        end
        if length(J)==1
            if filesAnnotations(J).bytes > 0
                [v, xml] = loadXML(fullfile(HOMEANNOTATIONS, folder,
filenameanno));
                labeledImages = labeledImages+1;
            else
                %disp(sprintf('file %s is empty', filenameanno))
                emptyAnnotationFiles = emptyAnnotationFiles+1;
                v.annotation.folder = folder;
                v.annotation.filename = filesImages(i).name;
            end
        else
            %disp(sprintf('image %s has no annotation', filename))
            v.annotation.folder = folder;
            v.annotation.filename = filesImages(i).name;
        end
    else
        filename = fullfile(HOMEANNOTATIONS, folder,
filesAnnotations(i).name);
        if filesAnnotations(i).bytes > 0
            [v, xml] = loadXML(filename);
            labeledImages = labeledImages+1;
        else
            disp(sprintf('file %s is empty', filename))
            v.annotation.folder = folder;
            v.annotation.filename = strrep(filesAnnotations(i).name,
'.xml', '.jpg');
        end
    end

    n = n+1;

    % Convert %20 to spaces from file names and folder names
    if isfield(v.annotation, 'folder')
        v.annotation.folder = strrep(v.annotation.folder, '%20', ' ');
        v.annotation.filename = strrep(v.annotation.filename, '%20', '
');
    end
end

```

```

        % Add folder and file name to the scene description
        if ~isfield(v.annotation, 'scenedescription')
            v.annotation.scenedescription = [v.annotation.folder ' '
v.annotation.filename];
        end
    end

%
%     if isfield(v.annotation.source, 'type')
%         switch v.annotation.source.type
%             case 'video'
%                 videomode = 1;
%             otherwise
%                 videomode = 0;
%         end
%     else
%         videomode = 0;
%     end

% Add object ids
if isfield(v.annotation, 'object')
    %keyboard
    Nobjects = length(v.annotation.object);
    %
    [x,y,foo,t,key] = LMobjectpolygon(v.annotation);

    % remove some fields
    if isfield(v.annotation.object, 'verified')
        v.annotation.object = rmfield(v.annotation.object,
'verified');
    end

    for m = 1:Nobjects
        % lower case object name
        if isfield(v.annotation.object(m), 'name')
            v.annotation.object(m).name =
strtrim(lower(v.annotation.object(m).name));
        end

        % add id
        %     if isfield(v.annotation.object(m).polygon, 'pt')
        %         v.annotation.object(m).id = m;

        %         % Compact polygons
        %         v.annotation.object(m).polygon =
rmfield(v.annotation.object(m).polygon, 'pt');

        %         pol.x = single(x{m});
        %         pol.y = single(y{m});
        %         pol.t = uint16(t{m});
        %         pol.key = uint8(key{m});
        %         if isfield(v.annotation.object(m).polygon, 'username')
        %             pol.username =
v.annotation.object(m).polygon.username;
        %         end
        %         v.annotation.object(m).polygon = pol;
        %     else
        %         v.annotation.object(m).deleted = '1';
        %     end
    end
end
end

```

```

% store annotation into the database
D(n).annotation = v.annotation;

if nargout == 2
    XML = [XML xml];
end

if mod(i,10)==1 && Narg<4
    plotbar(Hfig,f,Nfolders,i,N);
end
end
disp(sprintf(' Total images:%d, annotation files:%d (with %d empty xml
files)', N, labeledImages, emptyAnnotationFiles))
end

if nargout == 2; XML = [XML '</database>']; end

% Remove all the deleted objects. Comment this line if you want to see all
% the deleted files.
D = LMvalidobjects(D);

% Add view point into the object name
D = addviewpoint(D);

% Add crop label:
%words = {'crop', 'occluded', 'part'};
%D = addcroplabel(D, words); % adds field <crop>1</crop> for cropped
objects

% Add image size field
% D = addimagesize(D);

% % Summary database;
%[names, counts] = LMobjectnames(D);
%disp('-----')
%disp(sprintf('LabelMe Database summary:\n Total of %d annotated images. \n
There are %d polygons assigned to %d different object names', length(D),
sum(counts), length(names)))
disp(sprintf('LabelMe Database summary:\n Total of %d annotated images.',
length(D)))
%disp('-----')
%
close(Hfig)

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function fig = plotbar(fig,nf,Nf,ni,Ni)

if nargin > 0
    clf(fig)
    ha = subplot(2,1,1, 'parent', fig); cla(ha)
    p = patch([0 1 1 0],[0 0 1 1],'w','EraseMode','none', 'parent', ha);
    p = patch([0 1 1 0]*nf/Nf,[0 0 1
1], 'g', 'EdgeColor', 'k', 'EraseMode', 'none', 'parent', ha);
    axis(ha, 'off')
    title(sprintf('folders (%d/%d)',nf,Nf), 'parent', ha)

```

```

    ha = subplot(2,1,2, 'parent', fig); cla(ha)
    p = patch([0 1 1 0],[0 0 1 1],'w','EraseMode','none', 'parent', ha);
    p = patch([0 1 1 0]*ni/Ni,[0 0 1
1], 'r', 'EdgeColor', 'k', 'EraseMode', 'none', 'parent', ha);
    axis(ha, 'off')
    title(sprintf('files (%d/%d)',ni,Ni), 'parent', ha)
    drawnow
else
    % Create counter figure
    screenSize = get(0, 'ScreenSize');
    pointsPerPixel = 72/get(0, 'ScreenPixelsPerInch');
    width = 360 * pointsPerPixel;
    height = 2*75 * pointsPerPixel;
    pos = [screenSize(3)/2-width/2 screenSize(4)/2-height/2 width height];
    fig = figure('Units', 'points', ...
        'NumberTitle', 'off', ...
        'IntegerHandle', 'off', ...
        'MenuBar', 'none', ...
        'Visible', 'on', ...
        'position', pos, ...
        'BackingStore', 'off', ...
        'DoubleBuffer', 'on');
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function files = urlxmldir(page)

files = []; Folder = [];
page = strrep(page, '\\', '/');

%page

[folders,status] = urlread(page);
if status
    folders = folders(1:length(folders));
    j1 = findstr(lower(folders), '<a href="');
    j2 = findstr(lower(folders), '</a>');
    Nfolders = length(j1);

    fn = 0;
    for f = 1:Nfolders
        tmp = folders(j1(f)+9:j2(f)-1);
        fin = findstr(tmp, '"');
        if length(findstr(tmp(1:fin(end)-1), 'xml'))>0
            fn = fn+1;
            Folder{fn} = tmp(1:fin(end)-1);
        end
    end

    for f = 1:length(Folder)
        files(f).name = Folder{f};
        files(f).bytes = 1;
    end
end
end

```

Source Code 2.3: Result Display / Query Output

```

function Report = QI_resultDisplay(t1, t2)
global Dq HI;

```

```

for n = t1: t2
    fn = fullfile(HI,Dq(n).annotation.folder, Dq(n).annotation.filename);
    figure;
    imshow(fn);
end
Report = '...@ Results are displayed...#';

```

Source Code 2.4: Query Interpreter

```

function Report = QueryInterpreter(text, flage)
global Drq Dq DB;
inde = 0;
pos = findstr(text, ',');

for i = 1:length(pos)
    inds = inde+1;
    inde = pos(i);
    token = substr(text,inds, inde-inds);
    fp = findstr(token, '(');
    lp = findstr(token, ')');
    Drq(i).Word = substr(token,0,fp-1) ;
    Drq(i).SS = str2double(substr(token,fp+1,lp-fp-1));
end
clear lp fp token inde inds pos i;

% Extracting tokens for the query
token = '';
for i = 1:length(Drq)
    token = strcat(token, lower(Drq(i).Word), ',');
end
token = token(1:end-1);

% Querying the Corpus
Dq = LMquery(DB, 'object.name', token);
clear token i;

if (flage == 2)
    SIRRS;
elseif(flage == 1)
    VSM;
end

Report = '@...Query Interpreting Process completed...#';

```

Source Code 2.4: Set Annotation Path

```

function setAnnotationPath(Path)
global HA;
HA = Path;

```

Source code 2.5: Set Image/Video Corpus Path

```

function setImagePath(Path)
global HI;
HI = Path;

```

Source Code: Semantic Intensity Ranking Refinement Strategy

```
function SIRRS
global Dq Drq;

% Adding SS to the Corpus
for i = 1: length(Dq)
    for j = 1: length(Dq(i).annotation.object)
        for k = 1: length(Drq)
            if strcmpi(Dq(i).annotation.object(j).name, Drq(k).Word)
                SS = Drq(k).SS;
                SI = str2double(Dq(i).annotation.object(j).SI);
                Dq(i).annotation.object(j).SS = SS;
                Dq(i).annotation.object(j).RS = SS * SI;
            else
                Dq(i).annotation.object(j).SS = 0;
                Dq(i).annotation.object(j).RS = 0;
            end
        end
    end
end
clear i j k SS SI Drq;

% Calculating RS at image level
for i = 1:length(Dq)
    R = 0;
    for j = 1:length(Dq(i).annotation.object)
        R = R + Dq(i).annotation.object(j).RS;
    end
    Dq(i).annotation.RS = R;
end
clear i R j;

% Sorting the resultant data for retrieval
for i = 1:length(Dq)
    for j = i:length(Dq)
        if (Dq(i).annotation.RS < Dq(j).annotation.RS)
            temp = Dq(i).annotation;
            Dq(i).annotation = Dq(j).annotation;
            Dq(j).annotation = temp;
        end
    end
end
clear i j temp m;
```

Source Code: Vector Space Model

```
function VSM
global DB Drq Dq;
Dqt = '';
% Extracting terms
t = '';
for i = 1:length(Drq)
    t{i} =lower(Drq(i).Word);
end
clear i;
```

```

load('stopwords');
% Calculating Term frequency
t = sort(t);
clc;

for i = 1:length(DB)
    g = 1;
    for j = 1:length(t)
        tf = 0;
        tsi = 0;
        for k = 1:length(DB(i).annotation.object)
            if
                (strcmpi(t{j},NI_PorterStemmer(removestopwords(DB(i).annotation.object(k).name,stopwords))))
                    tf = tf + 1;
                    tsi = tsi + str2double(DB(i).annotation.object(k).SI);
                    obj = DB(i).annotation.object(k).name;
                end
            end
            if tf>0
                Dqt(i).annotation.imagepath =
                    strcat(DB(i).annotation.folder,'\',DB(i).annotation.filename);
                Dqt(i).annotation.object(g).name = obj;
                Dqt(i).annotation.object(g).SI = tsi;
                Dqt(i).annotation.object(g).TF = tf;
                g = g + 1;
            end
        end
    end

clear tsi tf k j i;

h = 1;
Dt='';
for i = 1: length(Dqt)
    if ~isempty(Dqt(i).annotation)
        Dt(h).annotation = Dqt(i).annotation;
        h = h + 1;
    end
end
clear h i Dqt;

% Calculating df (document frequency)
% put the data in front of the term
D = length(Dt);
Q = '';
for i = 1:length(t)
    df = 0;
    for j = 1:length(Dt)
        for k = 1:length(Dt(j).annotation.object)
            if
                (strcmpi(t{i},NI_PorterStemmer(Dt(j).annotation.object(k).name)))
                    df = df + 1;
                end
            end
        end
    end
    Q(i).term = t{i};
    Q(i).df = df;
    if df>0

```

```

        Q(i).idf = log(D/df);
    else
        Q(i).idf = 0;
    end
end
clear i df k j D;

% Calculating weights for all documents
for i = 1:length(Q)
    for j = 1:length(Dt)
        for k = 1:length(Dt(j).annotation.object)
            if
                (strcmpi(Q(i).term,NI_PorterStemmer(removestopwords(Dt(j).annotation.object
                (k).name, stopwords))))
                    Dt(j).annotation.object(k).IDF = Q(i).idf;
                    Dt(j).annotation.object(k).wht =
Dt(j).annotation.object(k).TF * Q(i).idf;
            end
        end
    end
end
clear k j i;

% wht for Query
for i = 1:length(Q)
    if Q(i).df>0
        Q(i).wht = 1 * Q(i).idf;
    else
        Q(i).wht = 0;
    end
end

% Calculating Vector Length for Query
S = 0;
for l = 1:length(Q)
    S = S + pow2(Q(l).idf);
end
QVL = abs(sqrt(S));

% Calculating Vector Length for each document
for i = 1:length(Dt)
    S = 0;
    for j = 1:length(Dt(i).annotation.object)
        if isfield(Dt(i).annotation.object, 'wht')
            S = S + pow2(Dt(i).annotation.object(j).wht);
        else
            S = 0;
        end
    end
    Dt(i).annotation.VL = abs(sqrt(S));
    Dt(i).annotation.QVL = QVL;
end
clear S j i l;

% Performing dot(.) product
for i = 1:length(Dt)
    DP = 0;

```



```

    for j = 1:length(Dt(i).annotation.object)
        for k = 1:length(Q)
            if isfield(Dt(i).annotation.object,'wht')
                DP = DP + (Q(k).wht * Dt(i).annotation.object(j).wht);
            end
        end
    end
    Dt(i).annotation.DP = DP;
end
clear DP k j i;

% Calculating Rank List
for i = 1:length(Dt)
    Dt(i).annotation.RL = Dt(i).annotation.DP / (Dt(i).annotation.QVL *
Dt(i).annotation.VL);
end
clear i;

% Sorting the result descending wise
for i = 1:length(Dt)
    for j = i:length(Dt)
        if Dt(i).annotation.RL < Dt(j).annotation.RL
            temp = Dt(i);
            Dt(i) = Dt(j);
            Dt(j) = temp;
        end
    end
end
clear temp j i;

% Extracting Datasets with RL > 0
for i = 1:length(Dt)
    if Dt(i).annotation.RL > 0
        Dq(i) = Dt(i);
    end
end
clear i;

% Adding SS to each term
for i = 1:length(Dq)
    for j = 1:length(Dq(i).annotation.object)
        for k = 1:length(Drq)
            if
                (strcmpi(NI_PorterStemmer(removestopwords(Dq(i).annotation.object(j).name,s
topwords)), Drq(k).Word))
                    disp(strcat('i=',num2str(i),' j=',num2str(j),'
k=',num2str(k)));
                    Dq(i).annotation.object(j).SS = Drq(k).SS;
                end
            end
        end
    end
end
clear i;

```

References

- Adams et al. 2003 Adams, W., Iyengar, G., Lin, C., Naphade, M., Neti, C., Nock, H., Smith, J. Semantic indexing of multimedia content using visual, audio, and text cues. *EURASIP J. Appl. Signal Process.* 2003(2), 170–185 (2003).
- Adrian et al. 2009 Adrian Popescu, Herve Le Borgne, and Pierre-Alain Moellic ,”Conceptual Image Retrieval over a Large Scale Database”, *CLEF 2008, LNCS 5706*, pp. 771–778, Springer-Verlag Berlin Heidelberg 2009.
- Agarwal et al. 2008 A. Agarwal and B. Triggs, “Multilevel image coding with hyper features,” *Int. J. Computer. Vision*, vol. 78, no. 1, 2008.
- Agarwal et al. 2010 Agarwal, S., Collins, M. Maximum margin ranking algorithms for information retrieval. In: Gurrin, C., et al. (eds.) *ECIR 2010. LNCS*, vol. 5993, pp. 332–343. Springer, Heidelberg (2010).
- Aigrain et al. 1996 P. Aigrain et al.: *Content-based Representation and Retrieval of Visual Media: A State-of-the art Review*, *Multimedia Tools and Applications*, Kluwer Academic Publishers, Vol. 3, No. 3, November 1996.
- Aleksic et al. 2006 Aleksic, P.S., Katsaggelos, A.K. Audio-visual biometrics. *Proc. IEEE* 94(11), 2025–2044 (2006).
- Andrea et al. 2003 M.Andrea Rodriguez and Max J. Egenhofer, “Determining Semantic Similarity among Entity classes from Different Ontologies”, *Knowledge and Data Engineering*, *IEEE Transactions*, vol.15, Issue 2, pp. 442-456, March-April 2003.
- Andreou et al. 2008 Andreou, A., 2005. *Ontologies and Query Expansion*. M.S. thesis, School of Informatics, Edinburgh Univ., Edinburgh, UK.
- Andrew et al. 2010 Andrew Trotman, Shlomo Geva, Jaap Kamps, Mounia Lalmas, Vanessa Murdock. *Current Research in Focused Retrieval and Result Aggregation*. 2010 s Springerlink.
- Arvola et al. 2010 Arvola, P., Kekalainen, J., and Junkkari, M. (2010). *Expected Reading Effort in Focused Retrieval Evaluation*. 2010 Springerlink.
- Baeza et al. 1991 U. Manber and R. Baeza-Yates. An algorithm for string matching with a sequence of don't cares. *Information Processing Letters*, 37:133--136, February 1991.
- Baeza et al. 1999 Baeza-Yates, R. and Ribeiro-Neto, B. (1999) *Modern information retrieval*. N.Y. ACM Press.
- Baeza. 2003 Ricardo Baeza-Yates. *Information Retrieval in the Web: beyond current search engines*, *International Journal on Approximated Reasoning* 34 (2-3), 97-104, 2003

- Bai et al. 2005 J. Bai, D. Song, P. Bruza, J. Nie, and G. Cao. 2005. Query expansion using term relationships in language models for information retrieval. In Fourteenth International Conference on Information and Knowledge Management (CIKM 2005).
- Barnard et al. 2003 K. Barnard, P. Duygulu, D. Forsyth, N. de Freitas, D. Blei, and M. Jordan. Matching words and pictures. *Journal of Machine Learning Research*, 3, 2002.
- Bate. 1986 M. Bates, "Subject Access in Online Catalogs: A Design Model", *Journal of the American Society for Information Science*, 11, 357 - 376, 1986.
- Beaulieu et al. 1992 Hancock-Beaulieu, M. (1992), Query expansion Advances in research in on-line catalogues *Journal of Information Science*, 18(2), 99-110.
- Beaulieu, 1997 Beaulieu, Experiments with interfaces to support query expansion. *Journal of Documentation*. v53 i1. 8-19. 1997.
- Belkin et al. 1992 Belkin, N., and Croft, B. 1992. Information filtering and information retrieval. *Communications of the ACM* 35(12):29–37
- Biederman. 1985 Biederman, I., 1985. Human image understanding: Recent research and a theory. In *Computer Vision, Graphics, and Image Processing*, vol. 32, pp. 29-73.
- Biederman. 1987 Biederman, I (1987) "Recognition-by-components: a theory of human image understanding" *Psychological Review* 94(2), 115-147.
- Biemann. 2005 C. Biemann, "Ontology Learning from Text: A Survey of Methods," *LDV Forum*, vol. 20, no. 2, pp. 75-93, 2005.
- Bimbo. 1999 A Del Bimbo, *Visual Information Retrieval*, Morgan Kaufmann Ed., 1999.
- Bonino et al. 2004 D.Bonino, F.Corno, L.Farinetti and A.Bosca, " Ontology driven semantic Search", *WEAST Transaction on Information Science and Application*, vol 1, pp. 1597-1605, December 2004.
- Bosch at al. 2008 A. Bosch, A. Zisserman, and X. Munoz, "Scene classification using a hybrid generative/discriminative approach," *TPAMI*, vol. 30, no. 4, pp. 712–727, 2008.
- Boutell et al. 2006 M. Boutell, J. Luo, and C. Brown, "Factor-graphs for region-based whole-scene classification," in *CVPR SLAM Workshop*, 2006.
- Bredin et al. 2007 Bredin, H., Chollet, G.: Audiovisual speech synchrony measure: application to biometrics. *EURASIP J. Adv. Signal Process.* 11 p. (2007). Article ID 70186.
- Brill. 1995 Brill, Eric (1995) *Transformation-Based Error-Driven Learning and Natural Language Processing: A Case Study in Part of Speech Tagging*. *Computational Linguistics*, December 1995.

- Buckley et al. 1998 Buckley, C., Mitra, M., Walz, J., and Cardie, C. (1998). Using clustering and super concepts within smart. In Voorhees, E., editor, In Proceedings of the 6th Text Retrieval Conference (TREC-6), page 107124.
- Burges et al. 2005 Burges, C., Shaked, T., Renshaw, E., Lazier, A., Deeds, M. Hamilton, N., and Hullender, G. Learning to rank using gradient descent. In Proceedings of the 22nd International Conference on Machine Learning (Bonn, Germany. 2005).
- Cao et al. 2005 G. Cao, J. Nie, and J. Bai. 2005. Integrating word relationships into language models. In Proceedings of the 2005 ACM SIGIR Conference on Research and Development in Information Retrieval.
- Cao et al. 2006 Y. Cao, J. Xu, T.-Y. Liu, H. Li, Y. Hunag, and H.W. Hon, Adapting ranking SVM to document retrieval, SIGIR 2006.
- Cao et al. 2007 Cao, Z., Qin, T., Liu, T. Y., Tsai, M. F., and Li, H. Learning to rank: From pairwise approach to list wise approach. In Proceedings of the 24th International Conference on Machine Learning (Corvallis, OR. 2007).
- Carlos et al. 2007 Carlos Hernandez-Gracidas and L. Enrique Sucar, Markov Random Fields and Spatial Information to Improve Automatic Image Annotation, Advances in Image and Video Technology Lecture Notes in Computer Science, 2007, Volume 4872/2007.
- Carneiro et al. 2007 Carneiro, G., Chan, A.B., Moreno, P.J., Vasconcelos, N.: Supervised learning of semantic classes for image annotation and retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (2007)
- Carson et al. 1998 C. Carson, M. Thomas, S. Belongie, J. M. Hellerstein, J. Malik. Blobworld: A System for Region-Based Image Indexing and Retrieval. Proc. International Conference on Visual Information Systems, pp. 509–516, Amsterdam, The Netherlands, June 1999. Springer Verlag.
- Carson et al. 2002 C. Carson, S. Belongie, H. Greenspan, J. Malik. Blobworld: Image Segmentation Using Expectation-Maximization and its Application to Image Querying. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 8, pp. 1026–1038, Aug. 2002.
- Cascia et al. 1996 M. La Cascia and E. Ardizzone. Jacob: Just a content-based query system for video databases. In Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP96, May 7-10, '96, Atlanta, Georgia, 1996.
- Cascia et al. 1998 M. La Cascia, S. Sethi, and S. Sclaroff. Combining textual and visual cues for content-based image retrieval on the world wide web. In Proceedings of the IEEE Workshop on Content-based Access of Image and Video Libraries, pages 24–28, Santa Barbara, CA, USA, June 1998.

- Chakravarti et al. 2009 Rishav Chakravarti, Xiannong Meng . A Study of Color Histogram Based Image Retrieval. 2009 Sixth International Conference on Information Technology: New Generations.
- Chang et al. 1987 S.-K. Chang, Q. Y. Shi, and C. Y. Yan, "Iconic indexing by 2-D strings", *IEEE Trans. Pattern Anal. Machine Intell.*, 9(3):413-428, May 1987.
- Chang et al. 1995 Shih-Fu Chang and John R. Smith, "Extracting Multi-dimensional Signal Features for Content-Based Visual Query," *SPIE Symposium on Visual Communications and Image Processing*, May 1995
- Chang et al. 1996 J. R. Smith and S.-F. Chang, "Automated binary texture feature sets for image retrieval," in *Proceedings of International Conference on Acoustics Speech and 125 Signal Processing*, vol. 4, pp. 2239-2242, May 1996.
- Chang et al. 1997 S.-F. Chang, W. Chen, H. Meng, H. Sundaram, and D. Zhong, "VideoQ: An Automated Content-Based Video Search System Using Visual Cues", *ACM 5th Multimedia Conference*, Seattle, WA, Nov. 1997.
- Chang et al. 2005 Chang, S.F., Manmatha, R., Chua, T.S.: Combining text and audio-visual features in video indexing. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5, pp. 1005–1008. IEEE Computer Society, Philadelphia (2005).
- Chang et al. 2005 S. F. Chang, R. Manmatha, and T. S. Chua. Combining text and audio-visual features in video indexing. In *IEEE ICASSP 2005*, 2005.
- Chapelle et al. 2009 O. Chapelle and S. S. Keerthi. Efficient Algorithms for Ranking with SVMs, *Information Retrieval Journal*, July 2009.
- Chen at al. 2006 Yanhua Chen, Manjeet Rege, Ming Dong, Farshad Fotouhi: Deriving semantics for image clustering from accumulated user feedbacks. *ACM Multimedia 2007*: 313-316
- Chum et al. 2007 O. Chum, J. Philbin, J. Sivic, M. Isard, and A. Zisserman. Total recall: Automatic query expansion with a generative feature model for object retrieval. In *Proc. ICCV*, 2007.
- Cimiano et al. 2005 P. Cimiano and J. Volker, "Text2onto," *Proc. Int'l Conf. Natural Language to Information Systems (NLDB)*, pp. 227-238, 2005.
- Cimiano et al. 2006 P. Cimiano, *Ontology Learning and Population from Text: Algorithms, Evaluation and Applications*. Springer-Verlag New York, Inc., 2006.
- Cleverdon. 1967 Cleverdon, C. (1967). The Cranfield tests on English language devices. *Aslib Proceedings* 19(6), 173-194.

- Costa et al. 2000 L. Costa and R. M. Cesar Jr. Shape Analysis and Classification: Theory and Practice. CRC Press, Inc., Boca Raton, FL, USA, 2000.
- Cox et al. 2000 Ingemar J. Cox, Matthew L. Miller, Thomas P. Minka, Thomas Papatomas, and Peter N. Yianilos. The bayesian image retrieval system, PicHunter: Theory, implementation and psychophysical experiments. IEEE Transactions on Image Processing ,2000.
- Crammer et al. 2002 Crammer, K., and Singer, Y. PRanking with ranking. Advances in Neural Information Processing Systems, 14: 641-647. 2002.
- Cristina et al. 2001 Cristina Ribeiro, Gabriel David. A Metadata Model for Multimedia Databases, Proceedings of the International Cultural Heritage Informatics Meeting (ichim), Milan, Italy, 2001.
- Croft et al. 1991 Croft, W. B., Turtle, H. R., and Lewis, D. D. (1991). The use of phrases and structured queries in information retrieval. In ACM SIGIR Conf. on research and development in information retrieval, Chicago, Illinois, United States.
- Cunningham et al. 2008 Sally Jo Cunningham and David M. Nichols. How people find videos. In , pages 201–210. ISBN 978-1-59593-998-2.
- Datta et al.2008 Ritendra Datta, Dhiraj Joshi, Jia Li, and James Ze Wang, “Image retrieval: Ideas, Influences, and trends of the new age,” ACM Comput. Survey, vol. 40, no. 2, 2008.
- Deerwester et al. 1990 Deerwester, Furnas, Landauer and Harshman, 1990. Indexing by latent semantic analysis, JASIS 1990.
- Deselaers et al. 2007 Thomas Deselaers, Tobias Weyand and Hermann Ney, Image Retrieval and Annotation Using Maximum Entropy, Evaluation of Multilingual and Multi-Modal Information Retrieval Lecture Notes in Computer Science, 2007, Volume 4730/2007.
- Deselaers et al. 2008 T. Deselaers, D. Keysers, and H. Ney. Features for image retrieval: An experimental comparison. Information Retrieval, 2008.
- Diederich et al. 2007 J. Diederich and W.-T. Balke, “The Semantic Growbag Algorithm: Automatically Deriving Categorization Systems,” Proc. European Conf. Digital Libraries (ECDL), pp. 1-13, 2007.
- Divakaran et al. 2000 A. Divakaran, A. Vetro, “Video Browsing System Based on Compressed Domain Feature Extraction”, IEEE Transaction on Consumer Electronics, Vol. 46, No. 3, pp. 637-644, August 2000
- Dunckley. 2003 Dunckley L (2003) Multimedia Databases: An Object-Relational Approach. Addison-Wesley.
- Duygulu et al. 2002 Duygulu, P., Barnard, K., de Freitas, J.F.G., Forsyth, D.A. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary. In: European Conference on Computer Vision, pp. 97–112 (2002)

- Dwork et al.2001 C. Dwork, R. Kumar, M. Naor and D. Sivakumar. 2001. Rank Aggregation Methods for the Web. WWW, pp. 613 – 622.
- Eakins. 2002 Eakins JP (2002). Towards intelligent image retrieval. Pattern Recognition 35(1) 3-14.
- Efthimiadis. 1996 Query expansion. Annual Review of Information Systems and Technology. v31. 121-187.
- Ekmekcioglu et al. 1992 Ekmekcioglu, F. C., Robertson, A. M. and Willett, effectiveness of query expansion in ranked output document retrieval systems, Journal of Information Science
- Enser et al 1993 P. G. B Enser, Query analysis in a visual information retrieval context, Journal of Document and Text Management, pg. 25 -52, 1993.
- Erkan et al. 2004 G. Erkan and D. R. Radev. 2004. LexRank: Graph-based Centrality as Saliency in Text Summarization, Journal of Artificial Intelligence Research, 22:457-479.
- Erol et al. 2005 B. Erol, and F. Kossentini, “Shape-based retrieval of video objects,” IEEE, Trans. on Multimedia, Vol. 7, No. 1, pp 179-182, 2005
- Faloutsos et al. 1994 C. Faloutsos, R. Barber, M. Flickner, J. Hafner, W. Niblack, D. Petkovic, and W. Equitz. Efficient and effective querying by image content. Journal of Intelligent Information Systems, 3(3-4):231-262, 1994.
- Fang et al. 2001 Fang Sheng, Xieg Fan, Gary Thomas, “A Knowledge-Based Approach to Effective Document Retrieval”, Journal of Systems Integration, Vol 10, 2001.
- Fang et al. 2005 Wei-Dong Fang, Ling Zhang, Yan Xuan Wang, and Shou-Bin Dong, “Towards a Semantic Search Engine Based on Ontologies”, IEEE Proceedings of the Fourth International Conference on Machine Learning and Cybernetics Guangzhou China, pp. 1913-1918, August 2005.
- Fang et al. 2006 H. Fang and C. Zhai. 2006. Semantic term matching in axiomatic approaches to information retrieval. In Proceedings of the 2006 ACM SIGIR Conference on Research and Development in Information Retrieval.
- Fei at al. 2005 L. Fei-Fei and P. Perona, “A Bayesian hierarchical model for learning natural scene categories,” in CVPR, 2005.
- Fellbaum et al. 1998 C. Fellbaum, WordNet: an electronic lexical database. Cambridge, Mass: MIT Press, 1998.
- Fend et al. 2003 D. Fend, W. Siu, H. Zhang, editors. Multimedia Information Retrieval and Management – Technological Fundamentals and Applications. Springer Verlag, 2003.

- Feng et al. 2007 Ming-Feng Tsai, Tie-Yan Liu, Tao Qin, Hsin-Hsi Chen, Wei-Ying Ma. FRank: A Ranking Method with Fidelity Loss, ACM , 2007.
- Feng et al. 2004 Feng, S.L., Manmatha, R., Lavrenko, V.: Multiple bernoulli relevance models for image and video annotation. In: IEEE Conf. Computer Vision and Pattern Recognition (2004)
- Fleischman et al. 2002 M. Fleischman and E.H. Hovy, "Fine Grained Classification of Named Entities," Proc. Int'l Conf. Computational Linguistics (COLING '02), 2002
- Flender et al. 2009 Flender, C., Kitto, K., Bruza, P, Beyond ontology in information systems. In: QI 2009. (2009).
- Frakes et al. 1998 Frakes, William B. and Baeza-Yates, Ricardo (Eds.), Information Retrieval: Data Structures and Algorithms, Englewood Cliffs, NJ: Prentice-Hall, 1992. ISBN: 0-13-463837-9 (504 pgs.) (Revised Version - 1998) republished on a cd-rom entitled Dr.Dobbs Essential Books on Algorithms and Data Structures
- Freund et al. 1999 Yoav Freund, Robert E. Schapire. A Short Introduction to Boosting. Journal of Japanese Society for Artificial Intelligence, 14(5):771-780, September, 1999.
- Freund et al. 2003 Freund, Y., Iyer, R., Schapire, R. E., and Singer, Y. An efficient boosting algorithm for combining preferences. Journal of Machine Learning Research, 4: 933-969. 2003.
- Freund. 1995 Yoav Freund. Boosting a weak learning algorithm by majority. Information and Computation, 121(2):256–285, 1995.
- Fu et al. 2005 Fu, G., Ch. B. Jones and A.I. Abdelmoty. Ontology-based Spatial Query Expansion in Information Retrieval, Proceeding of ODBASE: OTM Confederated International Conferences, 2005 pp: 1466-1482.
- Furn et al. 1987 G. W. Furnas, T. K. Landauer, L. M. Gomez, and S. T. Dumais, "The Vocabulary Problem in Human-System Communication", Communications of the ACM, Vol. 30, No. 11, November 1987, 964-971.
- Furu et al. 2009 Furu Wei, Wenjie Li, Wei Wang, "iRANK: An Interactive Ranking Framework and Its Application in Query-Focused Summarization", 2009 , ACM.
- Gemert et al. 2006 J. van Gemert, J. Geusebroek, C. Veenman, C. Snoek, and A. Smeulders, "Robust scene categorization by learning image statistics in context," in CVPR SLAM Workshop, 2006.
- Girgensohn et al., 2005 Andreas Girgensohn , Frank Shipman , Lynn Wilcox , Thea Turner , Matthew Cooper, MediaGLOW: organizing photos in a graph-based workspace, Proceedings of the 13th international conference on Intelligent user interfaces, February 08-11, 2009, Sanibel Island, Florida, USA.

- Gonzalo et al. 1998 Gonzalo, F. Verdejo, I. Chugur, and J. Cigarran. Indexing with WordNet synsets can improve text retrieval. In Proceedings of the COLING-ACL'98 Workshop on Usage of Word-Net in Natural Language Processing Systems, pages 38–44, Montreal, Canada, 1998.
- Gooda et al. 2010 Gooda Sahib, N., Tombros, A. Ruthven, I. Enabling interactive query expansion through eliciting the potential effect of expansion terms ECIR 2010. LNCS, Springer, Heidelberg (2010).
- Gross et al. 1994 M. H. Gross, R. Koch, L. Lippert, and A. Dreger, "Multiscale image texture analysis in wavelet spaces," in Proceedings of IEEE International Conference on Image Processing, vol. 3, pp. 412-416, November 1994.
- Gruber. 1995 Gruber, T. 1995. Towards Principles for the Design of Ontologies Used for Knowledge Sharing. International Journal of Human and Computer Studies, 43(5/6): 907- 928.
- Gruber. 1996 Gruber, T. 1996. Ontolingua: A mechanism to support portable ontologies. Technical report KSL-91-66. Stanford University, Knowledge Systems Laboratory. USA.
- Gupta et al. 1991 A. Gupta, T. E. Weymouth, and R. Jain. An extended object-oriented data model for large image bases. In Proceedings of the 2nd International Symposium on Design and Implementation of Large Spatial Databases (SSD), pages 45–61, Barcelona, Spain, September 1991.
- Hall et al. 1997 Hall, D.L., Llinas, J.: An introduction to multisensor fusion. In: Proceedings of the IEEE: Special Issues on Data Fusion, vol. 85, no. 1, pp. 6–23 (1997).
- Han et al. 2002 Ju Han and Kai-Kuang Ma, IEEE "Fuzzy Color Histogram and Its Use in Color Image Retrieval", IEEE transactions on image processing, vol. 11, no. 8, august 2002.
- Hanjalic. 2002 Alan Hanjalic. Shot-Boundary Detection: Unraveled and Resolved? IEEE Transactions on Circuits and Systems for Video Technology, 12(2):90–105, 2 2002.
- Harabagiu et al. 2001 S. Harabagiu, D. Moldovan, M. Pasca, R. Mihalcea, M. Surdeanu, R. Bunescu, R. Girju, V. Rus, and P. Morarescu. The role of lexico-semantic feedback in open domain textual question-answering. In ACL01, Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics, pages 274–281, Toulouse, France, 2001.
- Haralick et al. 1973 R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification." IEEE Transactions on Systems, Man, and Cybernetics, vol. 3, no. 6, pp. 610-621, April 1973.

- Hare et al. 2006b J. S. Hare, P. A. S. Sinclair, P. H. Lewis, K. Martinez, P. G. Enser, and C. J. Sandom. Bridging the semantic gap in multimedia information retrieval: Top-down and bottom-up approaches. In Proceedings of the 3rd European Semantic Web Conference, Budva, Montenegro, June 2006b.
- Harit et al. 2005 G. Harit, S. Chaudhury, and J. Paranjpe. Ontology guided access to document images. In Proceedings of the Eighth International Conference on Document Analysis and Recognition (ICDAR '05), pages 292-296, Seoul, Korea, August 2005.
- Hauptmann et al. 2007 A. Hauptmann, R. Yan, W.-H. Lin, M. Christel, and H. Wactlar. Can high-level concepts fill the semantic gap in video retrieval? a case study with broadcast news. *IEEE Transactions on Multimedia*, 2007.
- He. 2000 Y.W. He, "Global Motion Estimation Algorithm and Its Application in Video Coding", Master's Thesis, Tsinghua University, P. R. China, October 2000
- Hearst et al. 1992 M.A. Hearst, "Automatic Acquisition of Hyponyms from Large Text Corpora," Proc. Int'l Conf. Computational Linguistics (COLING), pp. 539-545, 1992.
- Heesch. 2005 Heesch, D. The NNk technique for image searching and browsing. PhD Thesis, Department of Computing, University of London, Imperial College of Science, Technology and Medicine, London, UK. 2005.
- Herbrich et al. 2000 Herbrich, R., Graepel, T., and Obermayer, K. Large margin rank boundaries for ordinal regression. *Advances in Large Margin Classifiers*, 115-132. 2000.
- Hersh. 2009 W. Hersh, *Information Retrieval – A Health and Biomedical Perspective*, Springer, New York, 2009.
- Hirst et al. 1998 Hirst and St-Onge, D. Lexical chains as representations of context for the detection and correction of malapropisms. In *WordNet, An Electronic Lexical Database*. The MIT Press, 1998.
- Hollink et al. 2003 L. Hollink, G. Schreiber, J. Wielemaker, and B. Wielinga. Semantic annotation of image collections. Proceedings of the of the K-CAP 2003 Workshop on Knowledge Markup and Semantic Annotation (Semannot '2003), Sanibel, Florida, USA, October 2003.
- Hollink et al. 2004 L. Hollink, G. Nguyen, G. Schreiber, J. Wielemaker, B. Wielinga, and M. Worring. Adding spatial semantics to image annotations. in 4th International Workshop on Knowledge Markup and Semantic Annotation at ISWC'04
- Hoogs et al. 2003 A. Hoogs, J. Rittscher, G. Stein, and J. Schmiederer. Video content annotation using visual analysis and a large semantic knowledgebase. In *Computer Vision and Pattern Recognition*, 2003.

- Hou et al. 2009 Hou, Y., Song, D. Characterizing pure high-order entanglements in lexical semantic spaces via information geometry. In: QI 2009.
- Howarth et al. 2005 Howarth, Peter and Ruger, Stefan (2005a, March). Fractional distance measures for content-based image retrieval. In Proceedings of ECIR 2005: European conference on IR research, Santiago de Compostela, ESPAGNE. Springer, Berlin, ALLEMAGNE (2005) (Monographie).
- Hsu et al. 1995 W. Hsu, T. S. Chua, and H. K. Pung. An integrated color-spatial approach to content-based image retrieval. In Proceedings of the 3rd ACM Multimedia Conference, pages 305–313, San Francisco, CA, USA, November 1995.
- Hsu et al. 2006 Hsu, Ming-Hung and Chen, Hsin-His. Information Retrieval with Commonsense Knowledge. In Proceedings of 29th ACM SIGIR International Conference on Research and Development in Information Retrieval (2006).
- Hsu et al. 2008 Ming-Hung Hsu, Ming-Feng Tsai and Hsin-Hsi Chen, "Combining WordNet and ConceptNet for Automatic Query Expansion: A Learning Approach" ,Lecture Notes in Computer Science, 2008, Volume 4993/2008.
- Huang J. et al 1997 Huang J., Kumar S.R., Mitra M., Zhu W.J., Zabih R., "Image Indexing using Color Correlograms", IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 1997 pages 762-768.
- Hunter et al. 1998 Hunter J., Iannella R., "The Application of Metadata Standards to Video Indexing", Second European Conference on Research and Advanced Technology for Digital Libraries ECDL'98, Crete, Greece, 1998.
- Ingrid et al. 2003 Ingrid Zukerman, Bhavani Raskutti, YingyingWe. Query Expansion and Query Reduction in Document Retrieval. Proceedings of the 15th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'03) 2003 IEEE.
- Jaimés et al. 2005 Jaimés, A., Sebe, N. Multimodal human computer interaction: a survey. In IEEE International Workshop on Human Computer Interaction. Beijing (2005).
- Jain et al. 1995 Jain, Anil K., and Aditya Vailaya. Image Retrieval Using Color And Shape. Great Britain: Elsevier Science Ltd, 1995.
- Jair et al. 2008 H. Jair Escalante, Manuel Montes and L. Enrique Sucar. Improving Automatic Image Annotation Based on Word Co-occurrence. Adaptive Multimedial Retrieval: Retrieval, User, and Semantics Lecture Notes in Computer Science, 2008, Volume 4918/2008.

- Jansen et al. 2000 Jansen, B. J., Spink, A. and Saracevic, T. (2000) "Real life, real users and real needs: A study and analysis of users queries on the Web." *Information Processing and Management*, 36(2), 207-227
- Jeannin et al. 2000 S. Jeannin, B. Mory, "Video Motion Representation for Improved Content Access", *IEEE Transaction on Consumer Electronics*, Vol. 46, No. 3, pp. 645-655, August 2000.
- Jegou et al. 2008 H. Jegou, M. Douze, and C. Schmid. Hamming embedding and weak geometric consistency for large scale image search. In *Proc. ECCV*, 2008.
- Jegou et al. 2009a H. Jegou, M. Douze, and C. Schmid. Packing bag-of-features. In *Proc. ICCV*, Sep 2009.
- Jegou et al. 2009b H. Jegou, M. Douze, and C. Schmid. On the burstiness of visual elements. In *Proc. CVPR*, Jun 2009.
- Jeon et al. 2003 J. Jeon, V. Lavrenko, and R. Manmatha. Automatic image annotation and retrieval using cross-media relevance models. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 119–126, 2003.
- Jian et al. 2005 Jian-Fu Li, Mao-Zu Guo, Shu-Hong Tian, "A New Approach to Query Expansion" *International Conference on Machine Learning and Cybernetics*, Guangzhou, August 2005.
- Jiang et al. 2004 S. Jiang, T. Huang, and W. Gao. An ontology-based approach to retrieve digitized art images. In *Proceedings of the 2004 IEEE/WIC/ACM International Conference on Web Intelligence*, pages 131-137, Beijing, China, September 2004.
- Jiang, et al. 1997 Jiang, J. and Conrath, D. Semantic similarity based on corpus statistics and lexical taxonomy. In *Proceedings of the International Conference on Research in Computational Linguistics*, 1997.
- Jin et al. 2003 Jin, Q., J. Zhao and B. Xu, 2003. Query Expansion Based on Term Similarity Tree Model. *Natural Language Processing and Knowledge Engineering*, 2003. *Proceedings. 2003 International Conference*, pp: 400-406.
- Jing et al. 2008 Yushi Jing, Shumeet Baluja (2008), "PageRank for Product Image Search", *ACM*.
- Jing et al. 1994 Y. Jing and W. Bruce Croft. 1994. An association thesaurus for information retrieval. In *Proceedings of RIAO*.
- Jiunn et al. 2006 Chi-Jiunn Wu Hui-Chi Zeng Szu-Hao Huang Shang-Hong Lai Wen-Hao Wang, "Learning-Based Interactive Video Retrieval System", *Proceedings of IEEE International Conference on Multimedia and Expo.*, pp: 1785-1788, 9-12 July 2006.
- Joachims. 2002 T. Joachims, *Optimizing Search Engines Using Click through Data*, *Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD)*, *ACM*, 2002.

- Joao. 2008 Joao Miguel Costa Magalhaes, Statistical Models for Semantic-Multimedia Information Retrieval. PhD thesis September 2008.
- John et al. 2008 John F. Gantz, Christopher Chute, Alex Manfrediz, Stephen Minton, David Reinsel, Wolfgang Schlichting, Anna Toncheve. "The Diverse and Exploding Digital Universe," An Updated Forecast of Worldwide Information Growth Through 2011, March 2008.
- Jorden et al. 2007 Jorden Boyd Graben, David Blei, Xiaojin Zhu" A Topic Model for Word Sense Disambiguation" Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning.
- Ju et al. 2007 Jun Xu, Hang Li, AdaRank: A Boosting Algorithm for Information Retrieval, Proc. of SIGIR 2007, 391-398.
- Jurie et al. 2005 F. Jurie and B. Triggs. Creating efficient codebooks for visual recognition. In IEEE International Conference on Computer Vision, pages 604-610, Beijing, China, 2005.
- Kang et al. 1999 H.B. Kang, "Spatio-Temporal Feature Extraction from Compressed Video Data", TENCON 99, Proceedings of the IEEE Region 10 Conference , Vol. 2, pp. 1339-1342, 1999
- Kaptein et al. 2010 Kaptein, R., Marx, M., (2010). Focused Retrieval and Result Aggregation with Political Data. Springerlink.
- Katerina et al. 2000 Katerina Frantzi, Sophia Ananiadou, Hideki Mima, Automatic recognition of multi-word terms: the C-value/NC-value method. International Journal on Digital Libraries, Vol. 3, No. 2. (25 August 2000), pp. 115-130.
- Keselman et al. 2008 A. Keselman, A.C. Browne and D.R. Kaufman, Consumer health information seeking as hypothesis testing, Journal of the American Medical Informatics Association (2008)
- Khan et al. 2006 S. Khan and F. Marvon, "Identifying Relevant Sources in query Reformulation", In the proceedings of the 8th International Conference on Information Integration and Web-based Applications & Services(iiWAS2006), Yogyakarta Indonesia, pp. 99-130, 2006.
- Kherfi et al. 2004 Kherfi ML, Ziou D & Bernardi A, Image retrieval from the world wide web issues, techniques and systems. ACM Computing Surveys (2004) 36- 35-67.
- Khrennikov. 2009 Khrennikov, A. Interpretations of probabilities. Walter de Gruyter (2009).
- Kimia. 2001 Kimia, B., "Shape Representation for Image Retrieval", Image Databases: Search and Retrieval of Digital Imagery,2001, John Wiley & Sons, pp. 345-358.
- Kiryakov et al. 2004 A. Kiryakov, B. Popov, I. Terziev, D. Manov, and D. Ognyanoff, "Semantic Annotation, Indexing, and Retrieval," J. Web Semantics, vol. 2, no. 1, pp. 49-79, 2004.

- Koprulu et al. 2004 Koprulu M., Cicekli N.K., Yazici A., "Spatio-temporal Querying In Video Databases", *Inf. Sci.* 160(1-4), pp. 131-152, 2004.
- Korfhage. 1997 Korfhage, R. R. (1997) *Information storage and retrieval*. N.Y. Wiley Computer Publishing.
- Larlus et al. 2006 D. Larlus and F. Jurie, "Latent mixture vocabularies for object categorization," in *BMVC*, 2006.
- Lavrenko et al. 2004 Lavrenko, V., Manmatha, R., Jeon, J.: A model for learning the semantics of pictures. In: *Advances in Neural Information Processing Systems*, vol. 16 (2004)
- Lazebnik et al. 2006 S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *CVPR*, 2006, pp. 2169–2178.
- Lazebnik et al. 2006 S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *CVPR*, 2006, pp. 2169–2178.
- Leacock et al. 1998 Leacock, C. and Chodorow, M. Combining local context and WordNet sense similarity for word sense identification. In *WordNet, An Electronic Lexical Database*. The MIT Press, 1998.
- Lek et al. 2007 M. Marsza lek, C. Schmid, H. Harzallah, and J. van de Weijer. Learning object representations for visual object class recognition, 2007. Visual Recognition Challenge workshop, in conjunction with IEEE International Conference on Computer Vision, Rio de Janeiro, Brazil
- Lenat. 1995 D. B. Lenat, "Cyc: A large-scale investment in knowledge infrastructure," *Communications of the ACM*, 1995.
- Lesk et al. 1986 Lesk, M. Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone. In *Proceedings of the SIGDOC Conference*, 1986.
- Lew et al. 2006 Lew MS, Sebe N, Djeraba C & Jain R (2006) Content-based multimedia information retrieval: state of the art and challenges. *ACM Transactions on Multimedia Computing, Communications and Applications* 2(1): 1-19.
- Li et al. 2006 J. Li and J. Wang. Real-time computerized annotation of pictures. In R. Zimmermann, editor, *MULTIMEDIA '06: Proceedings of the 14th annual ACM international conference on Multimedia*, pages 911-920, Santa Barbara, CA, USA, October 2006.
- Li et al. 2008 H. Li et al "Combining WordNet and ConceptNet for Automatic Query Expansion: A Learning Approach", *AIRS 2008*, LNCS 4993, pp. 213-224, Springer Berlin Heidelberg.
- Lieberman et al. 2004 Lieberman, H., Liu, H., Singh, P., and Barry, B. Beating Common Sense into Interactive Applications. *AI Magazine* 25(4) (2004).
- Lieberman et al., 2001 Lieberman, H., E. Rosen Zweig. P. Singh, (2001). *Aria: An Agent for Annotating And Retrieving Images*, *IEEE Computer*, July 2001

- Lin et al 2009 Lin, Mei-Ling Shyu, Guy Ravitz, Shu-Ching Chen, “video semantic concept detection via associative classification”, ICME 2009.
- Lin et al. 1998 Lin, D. An Information-theoretic Definition of Similarity. In Proc. of the ICML’98, 1998.
- Lin et al. 1998 D. Lin. Automatic retrieval and clustering of similar words. In COLING-ACL’98 – Proceedings of the International Conference on Computational Linguistics and the Annual Meeting of the Association for Computational Linguistics, pages 768–774, Montreal, Canada, 1998.
- Lin et al. 2003 C. Lin, B. Tseng, and J. Smith. VideoAnnEx: IBM MPEG-7 annotation tool for multimedia indexing and concept learning. In IEEE International Conference on Multimedia and Expo, 2003
- Liu et al, 1998 B. Liu, W. Hsu, and Y. Ma, “Integrating classification and association rule mining,” in International Conference on Knowledge Discovery and Data Mining (KDD98), August 1998, pp. 80–86.
- Liu et al. 2002 Liu, H. and Lieberman, H: Robust Photo Retrieval Using World Semantics. In: Proceedings of LREC2002 Workshop: Using Semantics for IR (2002).
- Liu et al. 2004 H. Liu and P. Singh, “Conceptnet a practical commonsense reasoning tool-kit,” BT Technology Journal, vol. 22, no. 4, pp. 211–226, 2004.
- Liu et al. 2004a Liu, Hugo (2004). MontyLingua: An end-to-end natural language processor with common sense.
- Liu et al. 2004b Liu, H. and Singh, P.: ConceptNet: A Practical Commonsense Reasoning Toolkit. BT Technology Journal (2004).
- Liu et al. 2008 Yongli Liu, Chao Li, Pin Zhang, Zhang Xiong, A Query Expansion Algorithm based on Phrases Semantic Similarity. 2008 International Symposiums on Information Processing.
- Liu, 2009 Tie-Yan Liu (2009), Learning to Rank for Information Retrieval, Foundations and Trends in Information Retrieval: Vol. 3: No 3, pp. 225–331,
- Loncaric et al. 1998 Loncaric, S., 1998. A Survey of Shape Analysis Techniques. In Pattern Recognition, vol. 31, no. 8, pp. 983-1001.
- Low et al. 1998 W. C. Low and T. S. Chua. Color-based relevance feedback for image retrieval. In Proceedings of the International Workshop on Multimedia Database Management Systems, pages 116–123, Dayton, OH, USA, August 1998.
- Lu et al. 1997 Lu, A, Ayoub, M, and Dong. J. (1997). Adhoc experiments using eureka. In In Proceedings of the 5th Text Retrieval Conference, page 229240.
- Lu et al. 2009 Z. Lu, W. Kim and W. Wilbur, Evaluation of query expansion using MeSH in PubMed, Information Retrieval (2009).

- Luo et al. 2002 Luo, R.C., Yih, C.C., Su, K.L.: Multisensor fusion and integration: Approaches, applications, and future research directions. *IEEE Sens. J.* 2(2), 107–119 (2002).
- Lytinen et al. 2000 S. Lytinen, N. Tomuro, and T. Repede. The use of WordNet sense tagging in FAQfinder. In *Proceedings of the AAAI00 Workshop on AI and Web Search*, Austin, Texas, 2000.
- Ma et al. 1995 W. Y. Ma and B. S. Manjunath, "A comparison of wavelet transform features for texture image annotation," in *Proceedings of IEEE International Conference on Image Processing*, vol. 2, pp. 256-259, October 1995.
- Ma et al. 1999 Wei-Ying Ma and B. S. Manjunath. NeTra: A toolbox for navigating large image databases. *Multimedia Systems*, 7(3):184–198, May 1999.
- Magalhaes et al. 2007 Magalhaes, J., Overell, S., and Rüger, S. (2007). A semantic vector space for query by image example. In *ACM SIGIR Conf. on research and development in information retrieval, Multimedia Information Retrieval Workshop*, July 2007, Amsterdam, The Netherlands.
- Magennis et al. 1997 Mark Magennis , Cornelis J. van Rijsbergen, The potential and actual effectiveness of interactive query expansion, *Proceedings of the 20th annual international ACM SIGIR conference on Research and development in information retrieval*, p.324-332, July 27-31, 1997, Philadelphia, Pennsylvania, United States .
- Makela et al. 2006 E. Makela, E. Hyvonen, and S. Saarela. Ontogator | a semantic view-based search engine service for web applications. In I. F. Cruz, S. Decker, D. Allemang, C. Preist, D. Schwabe, P. Mika, M. Uschold, and L. Aroyo, editors, *Proceedings of the 5th International Semantic Web Conference (ISWC 2006)*, pages 847-860, Athens, GA, USA, November 2006.
- Mandala et al, 1999 Rila Mandala, Takenobu Tokuanga and Hozumi Tanaka, 1999. Combining multiple evidence from different types of thesaurus for query expansion. *SIGIR*.
- Manjunath et al.1996 Manjunath, B. S. and Ma, W. Y. (1996). Texture Features for Browsing and Retrieval of Image Data, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 18, No. 8, August 1996.
- Marcus et al., 1993 Marcus, M., Santorini, B. and Marcinkiewicz, M. (1993). Building a large annotated corpus of English: the Penn treebank. *Computational Linguistics*, 19:313-330.
- Melucci. 2008 Melucci, M. A basis for information retrieval in context. *ACM TOIS* 26(3) (June 2008).

- Meng et al. 1995 J. Meng, Y. Juan and S.-F. Chang, "Scene Change Detection in a MPEG Compressed Video Sequence," SPIE Symposium on Electronic Imaging: Science & Technology- Digital Video Compression: Algorithms and Technologies, SPIE vol. 2419, San Jose, Feb. 1995.
- Mezaris et al. 2003 V. Mezaris, I. Kompatsiaris, and M. G. Strintzis. An ontology approach to object-based image retrieval. In S. Oy, editor, Proceedings of the IEEE International Conference on Image Processing (ICIP 2003), volume 2, pages 511-514, Barcelona, Catalonia, Spain, September 2003.
- Mezaris et al. 2004 V. Mezaris, I. Kompatsiaris, and M. G. Strintzis. Region-based image retrieval using an object ontology and relevance feedback. EURASIP Journal on Applied Signal Processing, (6):886-901, 2004.
- Mihajlovic et al. 2001 Mihajlovic V., Petkovic M., "Automatic Annotation of Formula 1 Races for Content-Based Video Retrieval", Technical report no. TR-CTIT-01-41, Centre for Telematics and Information Technology, University of Twente, 2001.
- Mihalcea et al. 1990 R. Mihalcea and D. Moldovan. A method for word sense disambiguation of unrestricted text. In ACL99 – Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics, Baltimore, Maryland, 1999.
- Mihalcea et al. 2006 R. Mihalcea, C. Corley, and C. Strapparava. 2006. Corpus-based and knowledge-based measures of text semantic similarity. AAAI 2006.
- Miller et al. 1990 Miller, G., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. 1990. WordNet: An on-line lexical database. International journal of lexicography, 3(4), 235-244.
- Miller et al. 1990 Miller, G.A. et al., 1990. Introduction to WordNet: an on-line lexical database. International Journal of Lexicography, 3 (4): 235-312.
- Miller GA. 1990 Miller GA (1990) Nouns in WordNet: A Lexical Inheritance System. Int. J of Lexicography 3(4): 245-264.
- Milne et al. 2007 D. Milne, "Computing semantic relatedness using Wikipedia link structure," in Proceedings of the New Zealand Computer Science Research Student Conference, 2007.
- Minka et al. 1996 T.P. Minka and R. W. Picard, "Interactive Learning using a Society of Models", MIT Media Lab Perceptual Computing , 1996.
- Mitra et al. 1998 Mitra, M., Singhal, A., and Buckley, C. (1998). Improving automatic query expansion. In In Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 98, Melbourne, Australia, Aug. 2428).

- Mojgan et al. 2009 Mojgan Farhoodi, Maryam Mahmoudi, Ali Mohammad Zare Bidoki, Alireza Yari and Mohammad Azadnia, "Query Expansion Using Persian Ontology Derived from Wikipedia", World Applied Sciences Journal 7 (4): 410-417, 2009.
- Mojsilovic et al. 2004 A. Mojsilovic, J. Gomes, and B. Rogowitz, "Semantic-friendly indexing and querying of images based on the extraction of the objective semantic cues," Int. J. Computer. Vision, vol. 56, no. 1-2, 2004.
- Monaco. 2009 James Monaco. How to Read a Film. Oxford Press, London, 4th edition, 2009. ISBN 978-0-19-532105-0.
- Moosmann et al. 2008 F. Moosmann, E. Nowak, and F. Jurie, "Randomized clustering forests for image classification," TPAMI, vol. 30, no. 9, 2008.
- Moriyama et al. 2000 Moriyama T., Sakauchi M., "Video Summarisation based on the Psychological Content in the track structure", ACM Multimedia Workshops 2000, pp. 191-194, 2000.
- Nallapati. 2004 Ramesh Nallapati. Discriminative Models for Information Retrieval. Ramesh Nallapati, ACM-SIGIR, 2004.
- Naphade et al. 1998 M. R. Naphade, T. Kristjansson, B. Frey, and T.S. Huang. Probabilistic multimedia objects (multijects): A novel approach to video indexing and retrieval in multimedia systems. In Proc. of ICIP, 1998.
- Naphade et al. 2004 M. R. Naphade and J. R. Smith. On the detection of semantic concepts at trecvid.In. In Proceedings of the 12th annual ACM international conference on Multimedia, pages 660–667, New York, NY, USA, 2004.
- Naphade, et al. 2006 Naphade, et al., "Large-Scale Concept Ontology for Multimedia," IEEE MultiMedia, vol. 13, no. 3, pp. 86-91, July-September 2006.
- Navigli et al. 2004 R. Navigli and P. Velardi, "Learning Domain Ontologies from Document Warehouses and Dedicated Web Sites," Computational Linguistics, vol. 30, no. 2, pp. 151-179, 2004.
- Nekrestyanov et al. 2002 Nekrestyanov I.S., Pantleeva N.V., "Text Retrieval Systems for the Web" Programming and computer Software, 2002, Vol 28, No 4.
- Niblack et al. 1993 W. Niblack, et al, "The QBIC Project: Querying Images by Content Using Color, Texture and Shape", Proc. IS&T/SPIE. on Storage and Retrieval for Image and Video Databases I, San Jose, CA, February 1993.
- Niblack et al. 1994 W. Niblack, R. Barber, and et al. The QBIC project: Querying images by content using color, texture and shape. In Proceedings of SPIE Storage and Retrieval for Image and Video Databases, pages 173,187, San Jose, CA, USA, 1994.

- Nie et al. 2006 Nie, L., Davison, B. D., and Qi, X. Topical link analysis for web search. In Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (Seattle, WA. 2006).
- Niles et al. 2001 I. Niles and A. Pease. Towards a standard upper ontology. In C. Welty and B. Smith, editors, FOIS '01: Proceedings of the International Conference on Formal Ontology in Information Systems, pages 2-9, Ogunquit, Maine, USA, October 2001.
- Nister et al. 2006 D. Nister and H. Stewenius. Scalable recognition with a vocabulary tree. In Proc. CVPR, 2006.
- Nowak et al. 2006 E. Nowak, F. Jurie, and B. Triggs, "Sampling strategies for bag-of-features image classification," in ECCV, 2006.
- Nowak et al. 2009 S. Nowak and P. Dunker. Overview of the clef 2009 large scale visual concept detection and annotation task. In CLEF working notes 2009, Corfu, Greece, 2009.
- Oviatt et al. 2003 Oviatt, S.L.: Multimodal interfaces. In: Jacko, J., Sears, A. (eds.) The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications. Lawrence Erlbaum Assoc., NJ (2003).
- Paris et al. 2010 Paris, C., Wan, S., Thomas, P. (2010). Focused and Aggregated Search: A Perspective from Natural Language Generation. Springerlink.
- Pasca. 2005 M. Pasca, "Finding Instance Names and Alternative Glosses on the Web: Wordnet Reloaded," Proc. Int'l Conf. Computational Linguistics and Intelligent Text Processing (CICLing), pp. 280-292, 2005.
- Pasquale et al. 2010 Pasquale De Meo, Giovanni Quattrone , Domenico Ursino, A query expansion and user profile enrichment approach to improve the performance of recommender systems operating on a folksonomy, Springer Science Business Media (2010).
- Pass et al, 1996 Greg Pass, Ramin Zabih, "Histogram refinement for content based image retrieval" WACV '96.
- Pass et al. 1998 Pass, G., Zabih, R., Miller, J. (1996). Comparing images using color coherence vector, ACM Multimedia, Boston, MA, 1996, pp. 65-73.
- Peat et al. 1991 H. J. Peat and P. Willett. 1991. The limitations of term co-occurrence data for query expansion in document retrieval systems. Journal of the american society for information science, 42(5):378-383.
- Pehcevski et al. 2010 Pehcevski, J., Thom, J.A., Vercoustre, A.-M., Naumovski, V. (2010) Entity ranking in Wikipedia Utilising categories, links and topic difficulty prediction. Springerlink.

- Pena et al. 2010 Pena Saldarriaga, S., Morin, E., Viard-Gaudin, C.: Ranking fusion methods applied to on-line handwriting information retrieval. In: Gurrin, C., et al. (eds.) ECIR 2010. LNCS, vol. 5993, pp. 253–264. Springer, Heidelberg (2010).
- Peng et al. 2010 Peng, J., Macdonald, C., Ounis, I.: Learning to select a ranking function. In: Gurrin, C., et al. (eds.) ECIR 2010. LNCS, vol. 5993, pp. 114–126. Springer, Heidelberg (2010).
- Pentland et al. 1996 A. Pentland, R. W. Picard, and S. Sclarof. Photobook: Content-based manipulation of image databases. *International Journal of Computer Vision*, 18(33):233 - 254, 1996.
- Perdoch et al. 2009 M. Perdoch, O. Chum, and J. Matas. Efficient representation of local geometry for large scale objects retrieval. In *Proc. CVPR*, 2009.
- Perdoch et al. 2009 M. Perdoch, O. Chum, and J. Matas. Efficient representation of local geometry for large scale objects retrieval. In *Proc. CVPR*, 2009.
- Philbin et al. 2007 J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Object retrieval with large vocabularies and fast spatial matching. In *Proc. CVPR*, 2007.
- Philbin et al. 2008 J. Philbin, O. Chum, M. Isard, J. Sivic, and A. Zisserman. Lost in quantization: Improving particular object retrieval in large scale image databases. In *Proc. CVPR*, 2008.
- Picard et al. 1995 R. W. Picard and T. P. Minka. Vision texture for annotation. *Multimedia Systems: Special Issue on Content-based Retrieval*, 3(1):3-14, 1995.
- Pickens et al. 2008 J. Pickens and G. Colovchinsky. 2008. Ranked Feature Fusion Models for Ad Hoc Retrieval. *CIKM*, pp. 893-900.
- Pickering et al. 2003 Pickering, Marcus J. and Ruger, Stefan (2003, November). Evaluation of key frame-based retrieval techniques for video. *Computer Vision and Image Understanding* 92 (2-3), 217–235.
- Piwowarski et al. 2009 Piwowarski, B., Lalmas, M. A quantum-based model for interactive information retrieval. In: *ICTIR '09*. (2009).
- Poikonen et al. 2009 T. Poikonen and P. Vakkari, Lay persons' and professionals' nutrition-related vocabularies and their matching to a general and a specific thesaurus, *Journal of Information Science* (2009).
- Ponte et al. 1998 Ponte, J. and W. Croft, 1998. A language modeling approach to information retrieval. In *Proceedings of the 21st ACM SIGIR Conference on Research and Development in Information Retrieval*, pp: 275-281.
- Ponzetto et al. 2007 S. Ponzetto and M. Strube, “Deriving a Large Scale Taxonomy from Wikipedia” *Proc. 22nd Nat'l Conf. Artificial Intelligence (AAAI '07)*, pp. 1440-1447, July 2007.

- Porkaew et al. 1999 Kriengkrai Porkaew, Sharad Mehrotra, Michael Ortega, and Kaushik Chakrabarti, 1999, Similarity Search Using Multiple Examples in MARS, the 3rd International Conference on Visual Information Systems (Visual), June 2-4, 1999, Amsterdam, Netherlands
- Posner et al. 1989 Posner, M.I., 1989. Foundations of cognitive Science. Editor MIT Press, ISBN: 0262161125.
- Potamianos et al. 2003 Potamianos, G., Neti, C., Gravier, G., Garg, A., Senior, A.: Recent advances in the automatic recognition of audiovisual speech. Proc. IEEE 91(9), 1306–1326 (2003).
- Pradeep et al. 2010 Pradeep K. Atrey, M. Anwar Hossain, Abdulmotaleb El Saddik, Mohan S. Kankanhalli Multimodal fusion for multimedia analysis: a survey, Springer-Verlag Multimedia Systems (2010) 16:345–379.
- Qiu et al. 1993 Y. Qiu and H.P. Frei. 1993. Concept based query expansion. In Proceedings of the 1993 ACM SIGIR Conference on Research and Development in Information Retrieval.
- Randall et al. 2001 Randall Packer and Ken Jordan, eds., Multimedia: From Wagner to Virtual Reality (New York: W.W. Norton, 2001)
- Rasiwasia et al. 2006 Rasiwasia, N., Vasconcelos, N., and Moreno, P. (2006). Query by semantic example. In CIVR, July 2006, Phoenix, AZ, USA.
- Rasiwasia et al. 2007 Rasiwasia, N., Moreno, P., and Vasconcelos, N. (2007). Bridging the gap: Query by semantic example. IEEE Transactions on Multimedia 9 (5):923-938.
- Resnik. et al. 1995 Resnik, P. Using information content to evaluate semantic similarity. In Proceedings of the 14th International Joint Conference on Artificial Intelligence, 1995.
- Rijsbergen et al. 2004 Van Rijsbergen, C.J. The Geometry of Information Retrieval. Cambridge University Press (2004).
- Rijsbergen. 1979 Van Rijsbergen, C. J. (1979). Information Retrieval. 2nd edition. London: Butterworths. Available at: <http://www.dcs.gla.ac.uk/Keith/Preface.html>
- Roberto et al. 2009 Roberto Navigli, Word Sense Disambiguation: A survey. ACM Computing surveys Vol 41, No.2, article 10, February 2009.
- Robertson et al. 1976 S. Robertson and K. S. Jones. Relevance weighting of search term. Journal of the American Society for Information Science, 27, 1976.
- Robertson. 1977 S.E. Robertson. The probability ranking principle in IR. Journal of Documentation, 33, 1977.

- Rocchio. 1971 Rocchio, J. J. (1971). Relevance feedback in information retrieval. In G. Salton (Ed.), *The SMART retrieval system. Experiments in automatic document processing* (pp. 313–323). Englewood Cliffs, NJ: Prentice-Hall.
- Rosero et al. 2009 Huertas-Rosero, A.F., Azzopardi, L., van Rijsbergen, C.J.: Eraser lattices and semantic contents. In: *QI 2009*. (2009).
- Rowe et al. 1994 L. A. Rowe, J. S. Boreczky, and C. A. Eads, “Indexes for User Access to Large Video Databases”, *Proc. IS&T/SPIE Conf. on Storage and Retrieval for Image and Video Databases II*, San Jose, CA, 1994, pp. 150-161.
- Rudin et al. 2009 C. Rudin, R. Schapire, Margin-based ranking and an equivalence between AdaBoost and RankBoost, *Journal of Machine Learning Research*, 10: 2193—2232, 2009.
- Rui et al. 1997 Y. Rui, T. S. Huang, M. Ortega, and S. Mehrotra. Content-based image retrieval with relevance feedback in mars. In *Proceedings of IEEE International Conference on Image Processing*, pages 815-818, Washington, DC, USA, October 1997.
- Rui et al. 1999 Y. Rui, T. Huang, S. Chang. Image Retrieval: Current Techniques, Promising Directions and Open Issues. *Journal of Visual Communication and Image Representation*, Vol. 10, No. 4, pp. 39–62, April 1999.
- Saber et al. 1996 E. Saber, A.M. Tekalp, and G. Bozdagi. Fusion of color and edge information for improved segmentation and edge linking. In *Proc of 1996 IEEE Int'l Conf. on Acoustics, Speech, and Signal Processing, ICASSP'96*, volume 4, pages 2176-2179, May 1996.
- Safar. M et al. 2000 Safar. M., Shahabi, C., Sun, X. (2000). Image retrieval by shape: A comparative study, *Proceedings of IEEE International Conference on Multimedia and Exposition (ICME)*, USA, 2000.
- Salton et al. 1975 G. Salton, A. Wong, and C. S. Yang (1975), "A Vector Space Model for Automatic Indexing," *Communications of the ACM*.
- Salton et al. 1971 G. Salton. *The SMART Retrieval System: Experiments in automatic document processing*. Prentice-Hall, Englewood Cliffs, NJ, 1971.
- Salton et al. 1975 G. Salton, A. Wong, and C. S. Yang (1975), "A Vector Space Model for Automatic Indexing," *Communications of the ACM*.
- Salton et al. 1983 Salton, Gerard; Edward A. Fox, Harry Wu (1983), *Extended Boolean information retrieval*, *Communications of the ACM*, Volume 26, Issue 11.
- Salton et al. 1993 Salton, Allan, Buckley and singhal. *Automatic Analysis theme generation and summarization of machine readable texts*.1993.
- Salton et al. 1997 Gerard Salton, Amit Singhal, Mandar Mitra, Chris Buckley: *Automatic Text Structuring and Summarization*. *Inf. Process. Manage.* 33(2): 193-207 (1997)

- Salton et al. 1998 Salton, G. and C. Buckley (1998). Term-Weighting Approaches in Automatic Text Retrieval. *Information Processing and Management* 24(5): 513-523
- Sande et al , 2010 b Koen E. A. van de Sande, Theo Gevers and Arnold W. M. Smeulders, The University of Amsterdam's Concept Detection System at Image CLEF 2009, [springer link](#) *Lecture Notes in Computer Science*, 2010, Volume 6242/2010, 261-268.
- Sande et al, 2010 a K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek. Evaluating color descriptors for object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010.
- Sanderson et al. 1999 M. Sanderson and W.B. Croft, "Deriving Concept Hierarchies from Text," *Proc. ACM SIGIR*, pp. 206-213, 1999.
- Santini et al. 2001 Santini S, Gupta A & Jain R (2001) Emergent semantics through interaction in image databases. *IEEE Transactions on Knowledge and Data Engineering* 13(3): 337-351.
- Santini. 2001 Santini S (2001) *Exploratory Image Databases*. Academic Press
- Sayed et al. 2007 EI Sayed, Ahmad Hacid, Hakim Zighed and Djamel, "A new Context-Aware Measure for Semantic Distance Using a Taxonomy and a Text Corpus", *IEEE International Conference on Information Reuse and Integration*, 2007 (IRI-07), pp. 279-284, 2007.
- Schank et al. 2004 Schank Roger C., Janet L. Kolodner, Gerald DeJong, "Conceptual information retrieval." *Proceedings of the 3rd annual ACM conference on Research and development in information retrieval*, 2004, Cambridge, England.
- Schulze et al. 1994 Schulze, B.M. et al; "Comparative State-of-the-art Survey and Assessment of General Interest Tools", Technical Report D1B – I, DECIDE Project, Institute for Natural Language Processing, Stuttgart, 1994.
- Schulze et al. 1995 H. Schutze, J. O. Pedersen. *Information Retrieval Based on Word Senses. Fourth Annual Symposium on Document Analysis and Information Retrieval*, 161-175, 1995.
- Schutze et al. 1997 Schutze, H. and J. Pederson. 1997. A Cooccurrence-based Thesaurus and Two Applications to Information Retrieval. *Information Processing and Management*.
- Sclaroff et al. 1997 S. Sclaroff, L. Taycher, and M. La Cascia. Imagerover: A content-based image browser for the world wide web. In *Proceedings of the IEEE Workshop on Content-based Access of Image and Video Libraries*, pages 2–9, San Juan, Porto Rico, June 1997.
- Seaborn. 1997 Seaborn, M. A. (1997). *Image Retrieval Systems*, November 1997.

- Sebastian et al. 2002 Sebastian, T. B., Klein, P. N., and Kimia, B. B. Shock-based indexing into large shape databases. In Proceedings of the Seventh European Conference on Computer Vision (2002), pp. 731-746.
- Shafiq et al. 2007 Shafiq Rayhan Joty , Sheikh Sadid-Al-Hasan. Advances in Focused Retrieval: A General Review. Computer and information technology, 2007. iccit 2007. 10th The International Conference on Communications and Information Technology.
- Shahabi et al. 1999 Shahabi, C., Safar, M. (1999). Efficient retrieval and spatial querying of 2D objects, IEEE International Conference on Multimedia Computing and Systems (ICMCS99), Florence, Italy, 1999, pp. 611-617.
- Sharmin et al. 2002 Siddiqu Sharmin, "A Wavelet Based Technique for Analysis and Classification of Texture Images", Carleton University, Ottawa, Canada, Proj. Rep. 70.593, April 2002.
- Shanmugam et al. 2009 T.N.Shanmugam, Priya Rajendran, "Effective Content-Based Video Retrieval System Based On Query Clip", Proceeding of the 2nd International Conference On Advanced Computer Theory and Engineering, vol.2 no.5, pp.1095-1102, September 2009.
- Sheikholeslami et al. 1994 Sheikholeslami, G., Zhang, A. (1997). An approach to clustering large visual databases using wavelet transform, Proceedings of the IEEE International Conference on Image Processing, 1994, pp. 407-411.
- Shim S. et al. 2003 Shim S., Choi T.S., "Image Indexing by Modified Color Co-occurrence Matrix", IEEE International Conference on Image Processing, Sept. 2003, pages 14-17.
- Siggelkow et al. 2001a Sven Siggelkow and Hans Burkhardt. Fast invariant feature extraction for image retrieval. In Remco C. Veltkamp, Hans Burkhardt, and Hans-Peter Kriegel, editors, State-of-the-Art in Content-Based Image and Video Retrieval. Kluwer, 2001.
- Siggelkow et al. 2001b S. Siggelkow, M. Schael, H. Burkhardt. SIMBA — Search Images By Appearance. Proc. DAGM 2001, Pattern Recognition, 23rd DAGM Symposium, Vol. 2191 of Lecture Notes in Computer Science, pp. 9–17, Munich, Germany, Sept. 2001. Springer Verlag.
- Siggelkow at al. 1997 S. Siggelkow, H. Burkhardt. Local Invariant Feature Histograms for Texture Classification. Technical Report 3, University of Freiburg, Institute for Computer Science, 1997.
- Siggelkow et al 2002 S. Siggelkow. Feature Histograms for Content-Based Image Retrieval. Ph.D. thesis, University of Freiburg, Institute for Computer Science, Freiburg, Germany, 2002.
- Simone et al. 2009 Simone Paolo Ponzetto, Roberto Navigli, "Large-Scale Taxonomy Mapping for Restructuring and Integrating Wikipedia". International Joint Conference On Artificial Intelligence , ACM (2009).

- Singhai et al. 2010 Nidhi Singhai, Prof. Shishir K. Shandilya. "A Survey On: Content Based Image Retrieval Systems". *International Journal of Computer Applications* (0975 – 8887) Volume 4 – No.2, July 2010.
- Smeaton et al. 2010 Alan F. Smeaton, Paul Over, and Aiden R. Doherty. Video shot boundary detection: seven years of TRECVID activity. *Computer Vision and Image Understanding*, 4 (114):411–418, 2010.
- Smeulders et al. 2000 A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain. Content-Based Image Retrieval: The End of the Early Years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 12, pp. 1349–1380, Dec. 2000.
- Smith et al. 1996 a Smith, J. R., Chang, S. F. (1996). VisualSeek: a fully automated content-based image query system, *Proceedings of ACM Multimedia 96*, Boston, MA, 1996, pp. 87-98.
- Smith et al. 1996b Smith, J. R., Chang, S. (1996). Transform features for texture classification and discrimination in large image databases, *Proceedings of the IEEE International Conference in Image processing*, 1994, pp. 407-411.
- Smith et al. 1996c Smith, John R, and Shi-Fu Chang. "Tools and Techniques for Color Retrieval." *Electronic Imaging: Science and Technology-Storage~Retrieval for Image and Video Database IV*. San Jose: IS&T/SPIE, 1996. 1-12.
- Smith et al. 1997 a J. R. Smith and S. F. Chang. Querying by color regions using the visualseek content-based visual query system. *Intelligent Multimedia Information Retrieval*, pages 23- 41, 1997.
- Smith et al. 1997 b J. R. Smith and S. F. Chang. Visually searching the web for content. *IEEE Multimedia Magazine*, 4(3):12 - 20, 1997.
- Snoek et al, 2008 C. G. M. Snoek, K. E. A. van de Sande, O. de Rooij, B. Huurnink, J. C. van Gemert, J. R. R. Uijlings, The MediaMill TRECVID 2008 semantic video search engine. In *Proceedings of the 6th TRECVID Workshop*, Gaithersburg, USA, November 2008.
- Snoek et al. 2005 Snoek, C.G.M., Worring, M.: Multimodal video indexing: A review of the state-of-the-art. *Multimed. Tools Appl.* 25(1), 5– 35 (2005).
- Snoek et al. 2005 Snoek, C.G.M., Worring, M., Smeulders, A.W.M.: Early versus late fusion in semantic video analysis. In: *ACM International Conference on Multimedia*, pp. 399–402. Singapore (2005).
- Snoek et al. 2007 C. G. Snoek, B. Huurnink, L. Hollink, M. de Rijke, G. Schreiber, and M. Worring. Adding semantics to detectors for video retrieval. *IEEE Transactions on Multimedia*, 2007
- Sparck et al. 1997 Sparck Jones and Willett "Search term relevance weighting given little relevance information", *Journal of Documentation*, 35, 1979, 30-48. (Reprinted in *Readings in Information Retrieval*, (Ed.), 1997).

- Spink et al. 1998 Spink, Amanda, Judy Bateman, & Bernard J. Jansen. 1998. Searching the web: A survey of Excite users. 1998. Users' searching behaviour on the Excite web search engine. In National Online Meeting Proceedings, ed. Martha Williams. New York: Information Today.
- Squire et al. 1999 D. Squire, W. Muller, H. Muller, "Relevance Feedback and Term Weighting Schemes for Content Based Image Retrieval", in Proceedings of the third International Conference on Visual Information Systems (VISUAL 99), Amsterdam, The Netherlands, 1999, pp.549-556.
- Stephen et al. 2009 Stephen Robertson and Hugo Zaragoza. The Probabilistic Relevance Framework: BM25 and Beyond. Foundations and Trends in Information Retrieval 3 no. 4, 333-389 (2009).
- Steven et al. 2007 Steven C.H. Hoi and Michael R. Lyu, "A Multimodal And Multilevel Ranking Framework For Content- Based Video Retrieval", 2007 International conference on Acoustics, speech, and Signal processing, Hawaii, USA, 15-20 April 2007.
- Stokes et al. 2009 N. Stokes, Y. Li, L. Cavedon and J. Zobel. Exploring criteria for successful query expansion in the genomic domain, Information Retrieval 12(2009).
- Stricker et al. 1995 M. Stricker and M. Orengo. Similarity of color images. In W. Niblack and R. Jain, editors, Storage and Retrieval for Image and Video Databases III (SPIE), volume 2420, pages381-392, San Diego/La Jolla, CA, USA, February 1995.
- Suchanek et al. 2006 F.M. Suchanek, G. Ifrim, and G. Weikum, "Combining Linguistic and Statistical Analysis to Extract Relations from Web Documents," Proc. ACM SIGKDD, pp. 712-717, 2006.
- Suchanek et al. 2007 F.M. Suchanek, G. Kasneci, and G. Weikum, "Yago: A Core of Semantic Knowledge," Proc. 16th Int'l Conf. World Wide Web (WWW '07), pp. 697-706, 2007.
- Sudderth et al. 2008 E. Sudderth, A. Torralba, W. Freeman, and A. Willsky, "Describing visual scenes using transformed objects and parts," IJCV, vol. 77, no. 1-3, 2008.
- Sudip et al. 2007 Sudip Kumar Naskar, Sivaji Bandyopadhyay, "Word Sense Disambiguation Using Extended WordNet," iccta, pp.446-450, International Conference on Computing: Theory and Applications (ICCTA'07), 2007.
- Swain et al. 1991 Swain, M., Ballard, D. (1991). Color indexing, International Journal of Computer Vision, 7(1), 1991, pp. 11-32.
- Swain et al. 1996 M. Swain, C. Frankel, and V. Athitsos. WebSeer: An image search engine for the world wide web. Technical report tr-96-14, University of Chicago Department of Computer Science, July 31 1996.

- Tamura et al. 1978 H. Tamura, S. Mori, and T. Yamawaki, "Texture features corresponding to visual perception," IEEE Transactions on Systems, Man, and Cybernetics, vol. 8, no. 6, pp. 1264-1274, September 1978.
- Tamura et al. 1984 H. Tamura and N. Yokoya "Image database systems: A survey" Pattern Recognition 17(1) pp.29-43 (1984).
- Tao et al. 1999 Tao, Y., Grosky, W. (1999). Delaunay triangulation for image object indexing: A novel method for shape representation, Proceedings of the Seventh SPIE Symposium on Storage and Retrieval for Image and Video Databases, San Jose, CA, 1999, pp. 631-942.
- Taycher et al. 1997 L. Taycher, M. La Cascia, and S. Sclaroff. Image digestion and relevance feedback in the ImageRover www search engine. In In Proceedings of the 2nd International Conference on Visual Information Systems, pages 85–94, San Diego, CA, USA, December 1997.
- Thomas et al.1997 Thomas S. Huang, Yong Rui, and Shinh-Fu Chang, Image retrieval: Past, Present and Future, International Symposium on Multimedia Information Processing, 1997.
- Town et al. 2004 b Town, C. P., and Sinclair, D. A. (2004). Language-based querying of image collections on the basis of an extensible ontology. International Journal of Image and Vision Computing 22 (3):251-267.
- Town et al. 2004a C. P. Town. Ontology based Visual Information Processing. PhD thesis, University of Cambridge,2004.
- Tring et al. 2000 Lai.Tring-Sheng, "A Photographic Image Retrieval System", School of Computing, Engineering and Technology University of Sunderland Sunderland, United Kingdom January 2000.
- Turney et al. 2001 Turney, Mining the web for synonyms: PMI-IR versus LSA on TOEFL. In Proc. Of ECML'01, 2001.
- Turtle et al. 1990 Turtle, H., Croft, W.B, Inference network for document retrieval. 13th annual international ACM SIGIR conference on research and development in information retrieval. Brussels, Belgium.1990.
- Turtle et al. 1991 Turtle, H., Croft, W.B, Evaluation of an inference network-based retrieval model. ACM Transaction on Information Systems (TOIS), 1991.
- Varelas et al. 2005 Giannis Varelas, Epimenidis Voutsakis and Paraskevi Raftopoulou, "Semantic Similarity Methods in WordNet and their application to Information Retrieval on the web", 7th ACM international workshop on Web information and data management, pp. 10-16 November, 2005.

- Vasconcelos et al. 2001 N. Vasconcelos and M. Kunt. Content-based retrieval from image databases: Current solutions sand future directions. In B. Mercer, editor, Proceedings of the IEEE International Conference on Image Processing (ICIP 2001).
- Veltkamp et al 2000 R. C. Veltkamp and M. Tanase. Content-based image retrieval systems: A survey. Technical Report UU-CS-2000-34, Utrecht University, 2000.
- Venters et al. 2000 C. C. Venters and M. Cooper, Content-based image retrieval, Technical Report, JISC Technology Application Program, 2000.
- Virginia. et al. 1995 Virginia E. Ogle and Michael Stonebraker. Chabot: Retrieval from a relational database of images. IEEE Computer, 28(9):40-48, September 1995.
- Vogel et al. 2007 J. Vogel and B. Schiele, "Semantic modeling of natural scenes for content-based image retrieval," IJCV, vol. 72, pp. 133–157, 2007.
- Volkmer et al. 2006 Exploring automatic query refinement for text-based video retrieval. T Volkmer, A Natsev, IEEE Intl. Conf. on Multimedia and Expo (ICME'06), 2006
- Voorhees. 1993 Voorhees, Ellen M. 1993. Using WordNet to disambiguate word senses for text retrieval. In Proceedings of the Sixteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pages 171-180. ACM Press
- Voorhees. 1994 E.M. Voorhees. Query expansion using lexical-semantic relations. In Proceedings of the 17th ACM-SIGIR Conference, 1994.
- Voorhees. 2001 Voorhees, E. (Sept. 2001) Evaluation by Highly Relevant Documents. In Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, New Orleans, LA, USA, pp. 74-82.
- Wactlar et al. 1996 Wactlar HD, Kanade T, Smith MA & Stevens SM (1996) Intelligent access to digital video: informedia project. Computer 29(5): 46-52.
- Wade et al. 1988 Wade, S. J. and Willett, P. (1988) INSTRUCT: a teaching package for experimental methods in information retrieval III Browsing clustering and query expansion Program, 22(1), 44-61.
- Wang et al, 2007 D.Wang, X. Liu, L. Luo, J. Li, and B. Zhang. Video diver: generic video indexing with diverse features. In ACM International Workshop on Multimedia Information Retrieval, pages 61-70, Augsburg, Germany, 2007.
- Wang et al.2001 Wang, J.Li, J. Wiederhold, G. "SIMPLicity: Semantics-Sensitive Integrated Matching for Picture Libraries," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23(9), 2001, 947-963.

- Wei et al. 2010 Wang Wei et al. 2010 Wang Wei, Payam Barnaghi, Member, IEEE, and Andrzej Bargiela, Member, IEEE, “Probabilistic Topic Models for Learning Terminological Ontologies”, IEEE transactions on knowledge and data engineering, vol. 22, no. 7, July 2010.
- Winn et al. 2005 J. Winn, A. Criminisi, and T. Minka, “Object categorization by learned universal visual dictionary,” in ICCV, 2005, pp. 1800–1807.
- Witten et al. 2005 I. H. Witten and E. Frank, Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann, second edition, June 2005.
- Wong et al. 2007 S. Wong, T. Kim, and R. Cipolla. Learning motion categories using both semantic and structural information. Proc. IEEE International Conference on Computer Vision and Pattern Recognition, 2007. Minneapolis, MN.
- Wu et al. 2002 Chuan Wu, Yuwen He, Li Zhao, Yuzhuo Zhong, Motion feature extraction scheme for content-based video retrieval, Proceedings of SPIE - The International Society for Optical Engineering (2002), Volume: 4676, Pages: 296-305
- Wu et al. 2004 Yi Wu, Edward Y. Chang, Kevin Chen-Chuan Chang, and John R. Smith. Optimal multimodal fusion for multimedia data analysis. In Proceedings of the 12th annual ACM international conference on Multimedia, pages 572–579, 2004.
- Wu et al. 2006 Wu, Z., Cai, L., Meng, H.: Multi-level fusion of audio and visual features for speaker identification. In: International Conference on Advances in Biometrics, pp. 493–499 (2006).
- Wu, et al. 1994 Wu,Z. and Palmer,M. Verb semantics and lexical selection. In Proceedings of the Annual Meeting of the Association for Computational Linguistics, 1994.
- Xiangming et al. 2010 Xiangming Mu and Kun Lu ,Towards effective genomic information retrieval: The impact of query complexity and expansion strategies Journal of Information Science (2010).
- Xiong et al. 2002 Xiong, N., Svensson, P.: Multi-sensor management for information fusion: issues and approaches. Inf. Fusion 3, 163–186(24) (2002).
- Xu et al. 2007 Jun Xu, Hang Li, AdaRank: A Boosting Algorithm for Information Retrieval, Proc. of SIGIR 2007, 391-398.
- Xu et al. 2000 Xu, J. and Croft, W. (2000). Improving the effectiveness of information retrieval with local context analysis. Transactions on Information Systems (ACM TOIS), 18(1):79–112.
- Xu et al. 2000 B. Xu. A visual query facility for DISIMA image database management system. Master’s thesis, Department of Computing Science, University of Alberta, April 2000.

- Xue et al. 2005 Xue, G. R., Yang, Q., Zeng, H. J., Yu, Y., and Chen, Z. Exploring the hierarchical structure for link analysis. In Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (Salvador, Brazil. 2005).
- Yan et al. 2005 R. Yan and M. R. Naphade. Semi-supervised cross feature learning for semantic concept detection in video. In IEEE Computer Vision and Pattern Recognition (CVPR), San Diego, US, 2005.
- Yanagawa et al.2007 A. Yanagawa, S.F. Chang, L. Kennedy, and W. Hsu. Columbia university's baseline detectors for 374 Iscom semantic visual concepts. Columbia university ADVENT Tech. Report 222-2006-8, March 2007.
- Yang et al. 2004 J. Yang, M. Y. Chen, and A. G. Hauptmann. Finding person x: Correlating names with visual appearances. In Intl. Conf. on Image and Video Retrieval (CIVR'04), Ireland, 2004.
- Yang et al. 2006 Yang, C., Dong, M., Hua, J. Region-based image annotation using asymmetrical support vector machine-based multiple-instance learning. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (2006).
- Yang et al. 2006 Yang, Hui, and Jamie Callan. 2006. Near-duplicate detection by instance-level constrained clustering. In Proc. SIGIR, pp. 421-428. ACM Press. DOI: doi.acm.org/10.1145/1148170.1148243.
- Yavlinsky. 2007 Yavlinsky, A. Image indexing and retrieval using automated annotation. PhD Thesis, Department of Computing, University of London, Imperial College of Science, Technology and Medicine, London. 2007.
- Yeganova et al. 2009 L. Yeganova, D.C. Comeau, W. Kim and W.J. Wilbur, How to interpret PubMed queries and why it matters, Journal of the American Society for Information Science and Technology (2009).
- Yeh et al. 2007 Jen-Yuan Yeh, Jung-Yi Lin, Hao-Ren Ke, Wei-Pang Yang, Learning to Rank for Information Retrieval Using Genetic Programming, SIGIR'07, July 23 – 27, 2007, Amsterdam, Netherlands ACM.
- Yuchi et al. 2010 Yuchi Huang, Qingshan Liu, Shaoting Zhang, Dimitris N. Metaxas. Image Retrieval via Probabilistic Hypergraph Ranking, 2010 IEEE.
- Zhang et al. 1993 Zhang, H. -J., Kankanhalli, A., and Smoliar, S. W. (1993). Automatic Partitioning of full-motion video. Multimedia Systems. vo. 1, pp. 10-28.
- Zhang et al. 1994 Hong Jiang Zhang, Yihong Gong, etc. "Automatic Parsing of News Video", Proc. IEEE Int'l Conf. Multimedia Computing and Systems, IEEE Computer Society Press, Los Alamitos, Calif.,1994.
- Zhang et al. 2002 Ming Zhang, Ruihua Song, Chuan Lin, Shaoping Ma, Zhe Jiang, Yijiang Liu and Le Zhao, 2002. Expansion-Based Technologies in Finding Relevant and New Information. TREC 2002.

Zhang et al. 2004 D. Zhang and G. Lu. Review of Shape Representation and Description Techniques. *Pattern Recognition*, 37(1):1-19, 2004.

Zhao et al. 2002 Hong-zhao H.E., H.E. Pi-lian, Jian-feng Gao and Changing Huang, 2002. Query Expansion Based on the Context in Chinese Information Retrieval. *Journal of Chinese Information Processing*, 16 (6): 32-37.

Zhao et al. 2003 Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A. Face recognition: a literature survey. *ACM Comput. Surv.* 35(4), 399–458 (2003).

Zhou X.S. et al. 2002 Zhou X.S., Thomas S. Huang “Unifying Keywords and Visual Contents in Image Retrieval”, *IEEE Multimedia*, Vol. 9, Issue 2, April 2002 pages 23-33.

Zuccon et al. 2010 Zuccon, G., Azzopardi, L.: Using the quantum probability ranking principle to rank interdependent documents. In: Gurrin, C., et al. (eds.) *ECIR 2010*. LNCS, vol. 5993, pp. 357–369. Springer, Heidelberg (2010).

Zuccon et al. 2009 Zuccon, G., Azzopardi, L., van Rijsbergen, K.: The quantum probability ranking principle for information retrieval. In: *ICTIR '09*. (2009).

Zuccon et al. 2010 Zuccon, G., Azzopardi, L.: Using the quantum probability ranking principle to rank interdependent documents. In: Gurrin, C., et al. (eds.) *ECIR 2010*. LNCS, vol. 5993, pp. 357–369. Springer, Heidelberg (2010).

Chang et al. 1992 S. K. Chang and A. Hsu. Image information systems: Where do we go from here? *IEEE Transactions on Knowledge and Data Engineering*, 4(5):431442, October 1992.

Cyc <http://www.cyc.com>

Wiki WN http://en.wikipedia.org/wiki/WordNet#cite_note-20 (Last access 3-Jan-2011)

Montylingua www.web.media.mit.edu/~hugo/montylingua/

Informedia <http://www.informedia.cs.cmu.edu/>

Flicker <http://www.flickr.com/>

Facebook <http://www.facebook.com/>

Google Images <http://www.google.co.uk/img/np?hl=en&tab=wi>

Yahoo Images <http://images.search.yahoo.com/>

Imagery <http://elzr.com/imagery>

Pic search	http://www.picsearch.com/
Alta Vista	http://www.altavista.com/image/default
Pixsy	http://www.pixsy.com/
Feelimage	http://www.feelimage.net/photo/
Photobucket	http://photobucket.com/
Youtube	http://youtube.com/
Google Video	http://video.google.co.uk/?hl=en&tab=ww
Blinkx	http://www.blinkx.com/
Enfooooo	http://en.fooooo.com/
Videosurf	http://www.videosurf.com/
Truveo	http://www.truveo.com/
Msn Videos	http://uk.msn.com/?st=1
Yahoo Video	http://video.search.yahoo.com/
Wiki Color Space	http://en.wikipedia.org/wiki/Color_model (Last access 29-Dec-2010)
QBIC	http://www.qbic.almaden.ibm.com/
Virage	http://www.virage.com
Pichunter	http://www.pichunter.com/
VisualSEEK	http://www.ctr.columbia.edu/VisualSEEk
Image Rover	http://www.cs.bu.edu/groups/ivc/ImageRover/
Chabot	http://http.cs.berkeley.edu/~ginger/chabot.html
Excalib	http://vrw.excalib.com/
Photobook	http://vismod.media.mit.edu/demos/photobook/

Jacob	http://www.csai.unipa.it:80/research/projects/jacob/
WebSEEk	http://www.ctr.columbia.edu/WebSEEk/
Blobworld	http://elib.cs.berkeley.edu/photos/blobworld/
Mars	http://www-db.ics.uci.edu/pages/research/mars.shtml
Netra	http://maya.ece.ucsb.edu/Netra/
SIMBA	http://lmb.informatik.uni-freiburg.de/SIMBA
Data Statistic	http://blog.flickr.net/en/2010/09/19/5000000000/ (Last access 23-Dec-2010)
Dailymotion	http://www.dailymotion.com/gb