

University of Southampton Research Repository ePrints Soton

Copyright © and Moral Rights for this thesis are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given e.g.

AUTHOR (year of submission) "Full thesis title", University of Southampton, name of the University School or Department, PhD Thesis, pagination

UNIVERSITY OF SOUTHAMPTON

Analysing The Content of Web 2.0 Documents by Using A Hybrid Approach.

by

Lailatul Qadri binti Zakaria

A thesis submitted in partial fulfillment for the
degree of Doctor of Philosophy

in the

Faculty of Engineering and Applied Science
Department of Electronics and Computer Science

June 2011

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF ENGINEERING AND APPLIED SCIENCE
DEPARTMENT OF ELECTRONICS AND COMPUTER SCIENCE

Doctor of Philosophy

by Lailatul Qadri binti Zakaria

User involvement in Web 2.0 has made a significant contribution to the increase in the amount of multimedia content on the Web. Images are one of the most used media, shared across the network to mark user experience in daily life. Interactive applications have allowed users to participate in describing these images, usually in the form of free text, thus gradually enriching the images' descriptions. Nevertheless, often these images are left with crude or no description. Web search engines such as Google and Yahoo provide text based searching to find images by mapping query concepts with the text description of the image, thus limiting the information discovery to material with good text descriptions. A similar issue is faced by text based search provided by Web 2.0 applications. Images with less description might not contain adequate information while images with no description will be useless as they will become unsearchable by a text based search. Therefore, there is an urgent need to investigate ways to produce high quality information to provide insight into the document content. The aim of this research is to investigate a means to improve the capability of information retrieval by utilizing Web 2.0 content, the Semantic Web and other emerging technologies. A hybrid approach is proposed which analyses two main aspects of Web 2.0 content, namely text and images. The text analysis consists of using Natural Language Processing and ontologies. The aim of the text analysis is to translate free text descriptions into a semantic information model tailored to Semantic Web standards. Image analysis is developed using machine learning tools and is assessed using ROC analysis. The aim of the image analysis is to develop an image classifier exemplar to identify information in images based on their visual features. The hybrid approach is evaluated based on standard information retrieval performance metrics, precision and recall. The example semantic information model has structured and enriched the textual content thus providing better retrieval results compared to conventional tag based search. The image classifier is shown to be useful for providing additional information about image content. Each of the approaches has its own strengths and they complement each other in different scenarios. The thesis demonstrates that the hybrid approach has improved information retrieval performance compared to either of the contributing techniques used separately.

Contents

Acknowledgements	xi
1 Introduction	1
1.1 Research Motivation	1
1.2 Research Background and Problem Statements	2
1.3 Research Aims and Objectives	7
1.4 Research Scope	8
1.5 Research Contributions	9
1.6 Chapter Organization	10
2 Literature Review	13
2.1 Multimedia on the Web	13
2.1.1 Web Generations	13
2.1.2 Multimedia Representation On The Web	17
2.2 Layers of Semantic Image Descriptions	20
2.3 Annotating Images with Semantic Descriptions	23
2.3.1 Folksonomy or Tag Based Approach	23
2.3.1.1 Reviews	23
2.3.1.2 Discussion	25
2.3.2 Natural Language Processing Approach	26
2.3.2.1 Reviews	27
2.3.2.2 Discussion	28
2.3.3 Automatic Image Annotation	29
2.3.3.1 Reviews	29
2.3.3.2 Disucssion	31
3 Methodology For The Hybrid Approach	35
3.1 The Preliminary Analysis	35
3.1.1 A Brief Description about Flickr	36
3.1.2 Experiment Setup and Word Analysis	37
3.1.3 Tags and Text Content Observation	38
3.1.4 Text Content Comparison	39
3.1.5 Discussions and Conclusion	41
3.2 Introduction to the Hybrid Approach	43
3.3 Multimedia Corpus	44
3.4 Conclusion	46
4 Modelling Semantic Image Descriptions using Text Analysis	47

4.1	Work flow in Modelling Semantic Image Descriptions	47
4.2	General and Specific Information identification and Extraction	48
4.3	Specific Information Identification Using Ontologies And Knowledge Bases	51
4.3.1	Malaysia Tourism Ontology (MTO)	51
4.3.2	Geonames Ontology	54
4.3.3	DBpedia	54
4.3.4	Mapping Extracted Concepts To The Knowledge Base	55
4.3.5	Semantic Information Model Description	57
4.4	Comparing Extracted Concepts with Tags: Discussion	59
4.4.1	Conclusion	63
5	Image Classification By Using Image Content Analysis	65
5.1	Image Classifier Design	66
5.2	The Line Analysis	68
5.3	The Colour Analysis	71
5.4	Finding the Optimal Threshold Value To Identify Building Images	72
5.4.1	Experiment 1: Optimal Threshold Value Selection Using Line His- tograms	72
5.4.2	Experiment 2: Optimal Threshold Selection by Observing Colour Histograms	81
5.4.3	Experiment 3: Optimal Threshold Selection by Observing Line and Colour Histograms	82
5.5	Classifying building Images and Non building Images: Results and Dis- cussions	84
5.5.1	Classification Results for Line Analysis	84
5.5.2	Classification Results for Colour Analysis	86
5.5.3	Classification Results for Integrated Analysis	87
5.6	Conclusions	93
6	Evaluation: Results and Discussions	95
6.1	Introduction	95
6.2	Task 1: Text based searching.	96
6.2.1	Extracting information related to location.	99
6.2.2	Locating information related to attractions and events tourism in Malaysia	100
6.3	Task 2: Image Analysis based search compared with text based search . .	103
6.4	Task 3: Integrating Text and Image based search: the Hybrid Approach. .	105
6.5	Task 4: Using image classifiers to classify other images than building/city landscape	107
6.6	Conclusion	112
7	Concluding Remarks and Future Work	115
7.1	Thesis Summary	115
7.2	Summary of the Text Analysis	116
7.3	Summary of the Image Analysis	118
7.4	Concluding Remarks for The Hybrid Approach	119
7.5	Future Work	120

7.5.1	Adding multimedia ontology for information representation inter-operability	120
7.5.2	Adding emotions towards higher semantic image annotations . . .	122
7.5.3	EXIF metadata	122
7.5.4	Classification for other Landscapes and other objects	123
7.5.5	Using a Hybrid interface to Support the hybrid approach	123
A	Preliminary Analysis Documents Examples	125
B	Stop Words	129
C	Tourism Thesaurus	133
D	Text Analysis Examples	139
E	Malaysia Tourism Ontology	143
F	Text Analyser Interface	149
	Bibliography	153

List of Figures

2.1	Information translations from Web 2.0 to the Semantic Web formats. . . .	17
2.2	The Pyramid of Jørgensen (2003).	21
3.1	Process flow of user generated content in Flickr.	36
3.2	An Example of Flickr Entries.	37
3.3	The Hybrid Approach Work Flow Framework	43
4.1	The construction of the semantic information model	48
4.2	Visualization of Parse Tree generated by APP	50
4.3	Main roots in the Malaysia Tourism Ontology. Descriptions for each root is presented in Appendix E	52
4.4	Sample of information collected to develop the tourism ontology for Malaysia. .	53
4.5	Visualization for information extracted for Image ID 1460920756	58
4.6	Visualization for information extracted for Image ID 1203148615.	58
4.7	This figure shows some examples of images, tags and concepts that were extracted by using text analysis techniques.	60
4.8	Example for storing Geonames entries in RDF format	60
4.9	Sample for incomplete Geonames entries	61
4.10	This figure shows some examples of images, tags and concepts that were extracted by using text analysis techniques.	62
5.1	The construction of City Landscape and Non City Landscape Image Clas- sifier	67
5.2	This Figure shows (a) example of directions and angles used to observe edges, (b) illustrates orientation angles and (c) shows list pixels count for each directions and its normalized value.	69
5.3	Illustration of RGB colour cube	71
5.4	building Image and non building image distributions using line histograms. Each of the boxes represents an image in the training set. The blue boxes refer to building images and the red boxes refer to non building image. The x-axis shows a cut-off point / threshold value candidate, and the Y-axis show number of building and non-building images in the training set.	76
5.5	The ROC Curves generated based on Sensitivity and 1- Specificity for Line Histogram Analyses from Table 5.7.	80
5.6	ROC Curves for Colour Histogram Analysis. The 3D Colour Histogram [216 bins] and 3 Colour Histograms [168 bins] have produced a perfect classification while 3D Colour Histogram with 64 bins has produced a negative curves.	81

5.7	Curves for line and colour histograms integrations. The training sets have produced a perfect classification.	83
5.8	ROC Curves illustrating image classification performance results using Line Analysis.	86
5.9	ROC Curves illustrating image classification performance results using Colour Analysis.	88
5.10	Overall Image Classification Performance Comparisons and Conclusion. . .	91
5.11	ROC Curves comparison between all tests done in image classification . .	94
7.1	Annotating a region of an image using MPEG-7 standard	121
7.2	Segment annotation using COMM standard	122
A.1	Document T1: <i>Through the mist</i>	125
A.2	Document T3: <i>Roman Masterpiece</i>	126
A.3	Document T17: <i>Kids having a splash behind Taj</i>	127
D.1	Image ID: 399296817.	139
D.2	Image ID: 399296817.	141
E.1	An Overview of Relationships between classes and Properties in the Malaysia Tourism Ontology	143
E.2	Protege screenshot shows an example of event added in the Malaysian Tourism Otology.	147
E.3	Figure 2: Protege screenshot shows an exmple of attraction added in the Malaysian Tourism Otology.	148
F.1	Text Analyser Interface screenshot.	150
F.2	The Natural Language Processing Component Output.	151
F.3	The Knowledge Base Component Output.	152

List of Tables

2.1	Different Layers of Facet Descriptions for Object, Spatial, Temporal, Activity and Abstract of Images.	22
2.2	Comparisons for tag based related analysis	26
2.3	Comparisons for NLP related work in text analysis	28
2.4	A comparison for related work in image classification	32
3.1	Percentage tags in the whole text description	40
3.2	Sample of information generated and stored in the multimedia corpus . .	46
4.1	Common Concepts in the Tourism Domain	49
4.2	Text Analyser Component Output. Invalid* refers to root word for concept that is more than one word. Text descriptions for this example is presented in Appendix D	51
4.3	Entry example for Attraction: Kapas Island	53
4.4	Entry example for Event:Citrawarna Festival	54
4.5	Geonames entries matched to <i>Kapas Island</i>	54
4.6	Kapas Island entry example in Dbpedia	55
4.7	Petronas Twin Tower entry example in Dbpedia	55
4.8	Knowledge Base Component Output	57
5.1	Sample of building and non building images used in the training set . . .	68
5.2	Examples of city and non city images with their normalized line histograms. The red bar in the line histograms represents long line data while the blue bar represents short line data.	70
5.3	Examples of building and non building images and their colour histograms generated by extracting colour features from the images. The second column shows the 3-D colour histograms and the third column shows the separate R,G and B histograms superimposed in different colours. For the 3D colour histogram, the x-axis represents the bin number while the fraction of pixels for each bin is indicated in the y-axis. For the 3 separate colour histograms, the x-axis bin numbers indicate the colour intensity while the y-axis indicates the fraction of pixels with that intensity for that each colour.	73
5.4	Unmerged and merged line histograms	74
5.5	Prediction Values for building and Non building Images. (Prediction value is Probability of images being a building images). Value generated by the Inference Engine for Line 1: Long Lines [72 dirs].	75
5.6	Data Matrix	77

5.7	Line Histogram Analyses Result for Sensitivity and 1-Specificity calculated based on Confusion Matrix. The Line Histogram Analyses consist of three sub experiments which are Line 1, Line 2 and Line 3.	80
5.8	Result for Finding Optimal Threshold value by observing colour histograms.	82
5.9	Thresholds values identified for building and non building selected based on trainings results.	83
5.10	Training sets results for colour and line histogram integration. Note: <i>Sens</i> represnt <i>Sensetivity</i> while <i>speci</i> denotes <i>Specificity</i>	85
5.11	Tests Results for Line Analysis	86
5.12	Quantitative Analysis Result for Line 1: Long Lines [72 Dirs]	87
5.13	Quantitative Analysis Result for Line 3: Short and Long [48 Dirs]	87
5.14	Test Result for Colour Analysis	88
5.15	Quantitative Analysis Result for Colour 2: 3D Histogram [216 bins]	88
5.16	Quantitative Analysis Result for Colour 3: 3H Histogram [196 bins]	89
5.17	Tests Result for Integrating Line and Colour Analysis	90
5.18	Quantitative Analysis Result for TEST 1	92
5.19	Quantitative Analysis Result for TEST 2	92
5.20	Quantitative Analysis Result for TEST 3	92
5.21	Quantitative Analysis Result for TEST 4	92
5.22	Quantitative Analysis Result for TEST 5	93
5.23	Quantitative Analysis Result for TEST 6	93
6.1	Queries and results for text based search analysis	98
6.2	Results for finding admin location for known object/location.	100
6.3	SPARQL Query and Results of attractions related to Sabah	101
6.4	SPARQL Query and Results of attractions related to Citrawarna Festival.	102
6.5	Example of information related to Petronas Towers extracted from Dbpedia resource.	103
6.6	Image Classification Results for CASE 1	104
6.7	CASE 1-Results for identifying building images by using text based search and image analysis.	104
6.8	CASE 3: Results for identifying building images by using text based search and image analysis.	105
6.9	Case 1: Results for integrating text based searching (tags and semantic model) with image analysis.	106
6.10	CASE 3: Results for integrating text based searching (tags and semantic model) with image analysis.	106
6.11	Results for recall and precision values for 100 images with the highest sunset prediction value.	108
6.12	Lists of 20 images with the highest probability of being sunset images.	109
6.13	Lists of 20 images with the highest probability of being beach images.	110
6.14	Results for classifying beach images by using Image Analysis (Line and 3D colour histograms). The result shows correct and incorrect images, precision and recall for 10 to 100 observed images.	111
6.15	Results for classifying beach images by using Image Analysis (3D colour histogram). The result shows correct and incorrect images, precision and recall for 10 to 100 images observed.	112

E.1	List of Event properties and its descriptions	144
E.2	Table 2: List of Attraction properties and its description	145
E.3	List of Datetime properties and its descriptions	145
E.4	list of Location properties and its description	145

Acknowledgements

This research project would not have been possible without the support of many people. I am heartily thankful to my supervisors, Prof. Dame Wendy Hall and Prof. Paul Lewis who were abundantly helpful and offered invaluable assistance, support and guidance. Special thanks also to all graduate friends, especially AIM group for sharing the thought and invaluable assistance. I would also like to convey thanks to the Ministry of Education Malaysia for providing the financial means. Finally, I wish to express my love and gratitude to my beloved families and friends; for their understanding and endless love, through the duration of the study.

Lailatul Qadri binti Zakaria

Chapter 1

Introduction

This chapter commences with an overview of the research in multimedia information retrieval analysis focusing on exploring rich user experience information within Web 2.0 documents. It will briefly provide research background and state problems that triggered the study. Next it presents objectives, research scope and research contributions. Finally chapter outlines for the rest of the thesis are presented.

1.1 Research Motivation

Two observations form the starting point for this research. The first is that Web 2.0 and the growth of social networking are providing a substantial body of valuable, but not easily accessible content on the web. For example, Wikipedia contains more than 91,000 active contributors working on more than 17,000,000 articles in more than 270 languages. Hundreds of thousands of visitors from around the world, produce tens of thousands of edits and add thousands of new articles daily ([Wikipedia \(2011\)](#)). The use of Hypertext Markup Language (HTML) is well known for its ability to provide an infrastructure to present and structure information in a way desired by its users, but at the same time, it only offers little or non-explicit information about the content of the information itself to allow direct access by machines. As a result, this collective knowledge is only well presented for human and is not easily accessible by machine.

The second is that the Semantic Web technologies provide a valuable set of tools and representations for content on the web, facilitating enhanced retrieval both by human and machine. Adapting Semantic Web technologies seems to be promising to overcome the earlier issue. The Semantic Web has gained many research interests such as web document content extraction ([Ciravegna and Wilks \(2003\)](#), [Wu et al. \(2008\)](#) and [Ruiz-Casado et al. \(2008\)](#)), knowledge representation in ontology development (Geonames ¹

¹<http://www.geonames.org/>

and DBpedia²) and multimedia semantic annotation (Pastorello et al. (2008), Grosky et al. (2008) and Garca et al. (2008)). For example, knowledge from Wikipedia has been extracted in the form of the Semantic Web standards and presented in the DBpedia knowledge base (Hellmann et al. (2009)).

Our research interest is on multimedia assets in Web 2.0 content focusing on image and text descriptions. Hence, the aim of this research is to investigate whether it is possible to improve the retrieval of Web 2.0 content on the web automatically by extracting, enriching and structuring Web 2.0 content using Semantic Web and other emerging technologies.

1.2 Research Background and Problem Statements

The Web is a multimedia environment, which makes for complex semantics (Berners-Lee et al. (2006)). Web documents might contain multimedia asset such as text, image audio and video. In the structure level, web document contents are interlinked from one to another thus creating a complex hypermedia network. In the content level, each media item can be represented in a semantic information model, and might contain relationships to the knowledge sources it represents.

The use of the Web to publish items such as pictorial collections, music and videos in great numbers is supported by the improvement in storage and network technologies. Websites such as Youtube, Flickr, Facebook, and Fotopages are among the most popular examples of the Web 2.0 trend. They allow people and communities to share, tag and describe their multimedia content in an interactive environment. The Web 2.0 term was coined not primarily to introduce a vision, but to describe the current state in Web engineering (Oreilly (2005)).

The popularity of Web 2.0 has made a significant contribution to the increase in the number of web pages and multimedia content on the Web. It presents new challenges to conventional multimedia information retrieval (MIR) not only to rely on meta-data but also on content based information retrieval combined with the collective knowledge generated by users' contributions and geo-referenced meta-data that is captured during the creation process (van Zwol et al. (2007)). Flickr had more than 3 billion images by the end of 2008 and is adding thousands of images per minute. There are already 80 million user generated content contributors in the US alone and the number is predicted to rise up to 115 million in 2013³.

²<http://wiki.dbpedia.org/About>

³The data was accessed on 10th June 2009 from
<http://www.emarketer.com/Report.aspx?code=emarketer2000549>

The differences between user generated content and expert generated content may be presented as follows. User generated content is generally unpredictable content. The purpose of creating the content is based on the user's own requirements, either for public or private access. The information added is based on user knowledge (shallow or in depth) and the length of the content is usually increased gradually by other users' efforts in providing feedback to the material added by the user. The information could be unclear or missing, and the trustworthiness of the information is sometimes questionable. The structure of user generated content in Web 2.0 is usually controlled by the applications. Examples of user generated content are documents produced in Web 2.0 applications such as Flickr, Wikipedia and Facebook.

In contrast, expert generated content is usually presented in a well structured way with a well defined format, hyperlinks with additional supportive materials such as images, video or audio. The information added both in general and specific information is usually based on specific end user requirements, either for the general public or a specific target audience. Moreover, the quality of the information added by experts is clear and mostly reliable. Therefore, analysing user generated content would be much more difficult than expert based content. Examples of expert generated content are documents from authorized sources such as medical, university and news web pages.

Multimedia is mainly treated and annotated using text, allowing the media to be accessed by text based searching. Most commercial search engines are relying on text based searching instead of content based search to deal with multimedia due to the many limitations in current content based search and retrieval when applied to Web-scale data. When the average query length is only between two and three words long ([Jansen and Spink \(2006\)](#) & [Spink et al. \(2001\)](#)), it is difficult to be able to provide accurate information for the required multimedia search. Therefore, although most of the Web 2.0 websites do provide text based searching to find images by mapping query concepts with words in image titles, descriptions or tags, access has become more difficult as the number of photos has increased. Moreover, Web 2.0 documents are based on HTML syntax which focuses exclusively on the information presentation while the information content itself remains essentially invisible to machines. The use of the HTML syntax has become the major challenge in information retrieval. Due to its lack of semantic information available to describe the document content, most of the interesting information which the machine could process directly is hidden. The documents/images are usually represented in the form of a flat text index. The text index consists of terms, frequencies based on the occurrences of the terms and term weights which statistically indicate their importance. The statistical methods lack precision and they fail to extract semantic indexes to represent the main concepts of the document ([Kang and Lee \(2005\)](#)).

With the emergence of massive multimedia repositories on the Web, traditional text based search suffers from limitations (Abdel-Mottaleb et al. (1996), Djeraba (2002), van Ossenbruggen et al. (2004), Shah et al. (2004), Geurts et al. (2005) and Troncy (2003)). Some problems and issues associated with text based annotation of multimedia information are listed below:

1. It would be much easier to annotate earlier rather than later. Usually, most of the technical information that is needed for annotation is captured during production. For example, EXIF metadata of the picture taken is able to capture information such as date, lens setting and geo location, while, information scripts, story boards and edit discussion lists are sometimes available in the creative industries. Furthermore, adding metadata during the production process is much cheaper and yields higher quality annotation than adding metadata in a later stage.
2. In general, manual annotation can provide image descriptions at the right level of abstraction. Nevertheless, manual annotation takes too much time and is thus expensive. Moreover, manual annotations are not scalable. Even though annotation based on automatic feature analysis is fast, cheap and more consistent, it is only able to extract low level signatures, while most high level description still requires human intervention.
3. Human interpretation leads to a high discrepancy in subjective perception resulting in inconsistencies within multimedia collections. It is difficult to describe multimedia information accurately given that different users have different concepts and context (thoughts and feelings) even when dealing with the same multimedia material.
4. Standardisation is required to ensure interoperability between systems, devices, applications or components. Since there are different file formats and tools available today, reusing metadata created by different tools is often hindered by the lack of interoperability. Annotator's tools are different from end user's tools. Thus, there is a need to provide the necessary interoperability by providing a high degree of standardisation both in the syntax and at the semantic level. This will ensure that one tool can parse the other tools' formats. To support semantic interoperability and automatic inference, the tool must be able to "understand" different terms and relations between terms used in different tools.
5. There are various approaches in classifying metadata, either describing properties of the image itself or describing the subject matter of the image. In the first category, typical annotations provide information such as title, creator and brief description of image properties. The second category describes what is depicted in the image. It is also useful to distinguish between objective observations (*the mouse is in front of the cat*) versus subjective interpretations (*the mouse seems to*

be chased by the cat). As a result, one sees a large variation in vocabularies used for this purpose.

There is an urgent need to investigate ways to produce high-quality information by providing additional knowledge and deeper understanding to provide insight into the document content. The Semantic Web technologies provide promising infrastructures to overcome these issues. The Semantic Web is introduced as an extension of the current Web in which content is given a well defined meaning, better enabling computers and people to work in cooperation (Berners-Lee et al. (2001) & Shadbolt et al. (2006)). Firstly, Semantic Web languages offer a formal representation such as Extensible Markup Language(XML), Resource Description Framework (RDF) and Web Ontology Language (OWL) that can be used to semantically describe Web content. Secondly, ontologies play an important role to facilitate information sharing by enabling a common understanding of domain knowledge for communication between people and across applications (Decker et al. (2000)). Ontologies are used to expose implicit knowledge in the semantic layer and provide a conceptualisation of the domain and associations with the media objects to facilitate semantic navigation, browsing and retrieval. These technologies will bring benefits in enhancing information retrieval and improved interoperability (Uren et al. (2006)).

Adding multimedia to the Semantic Web requires information to be explicitly annotated. Multimedia annotation is an initial step to facilitate effective multimedia services such as enhancing multimedia information retrieval, intelligent processing and presentation generation. Embedding multimedia formal knowledge representation into Semantic Web standards is crucial to ease information retrieval. There has been an increasing interest in multimedia content and context analysis which comes from a broad range of research communities; machine learning, computer vision, model based approaches and knowledge discovery. Nevertheless, all of them are facing one common problem: the semantic gap. The semantic gap is defined as *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)). To address the semantic gap issue, high semantic descriptions need to be generated to allow reliable content location, access and navigation services. The semantic description is important for media elements understanding, synthesizing meaning, manipulating perception and generating a message with a systematic study of media productions (Dorai and Venkatesh (2001)).

Since multimedia resources are highly complex and content rich, they require appropriate technologies to support semantic analysis and annotation processes. An image, for example, can be described by different facets of semantics such as object, location, spatio-temporal and events/activity. Furthermore, an image can be viewed differently, either through perceptual content which derives from our perception, or through inferential reasoning by our knowledge, personal experience and cultural condition. In order to

allow multimedia content to be reached directly by machines, it needs to be properly annotated and tagged. Standard technologies such as language representation proposed by the Semantic Web can be used to annotate multimedia content in order to enhance interoperability and standardization. Moreover, appropriate annotation is essential to represent the different levels of information and their association to support semantic retrieval.

Multimedia annotation requires tools to discover useful knowledge from multimedia objects, and enable innovative and intelligent application, filtering and multimedia information retrieval ([Benitez and Chang \(2003\)](#)). Currently, most of the material added in Web 2.0 documents are described by using free text information in title, caption and comments; and annotated in the form of tags. Even though tagging is a useful method to characterize this information, it misses a lot of semantics of the data ([W3C \(2007\)](#)). The interest of this research is therefore to analyse the content of Web 2.0 documents and improve the information representations of the content in order to provide a better understanding thus enhancing information retrieval.

The in depth analysis on this research focuses on handling image and text components of Web 2.0 content. With the increase of storage capacities for digital cameras, they can generate hundreds of photos on an ongoing basis, and hence contribute to the huge number of photos uploaded in Web 2.0 websites. The ranges of the photos are massive, from vacation, sports, weddings, parties, travelling, friends, hobbies, pets, every day life and many more. Typically, based on the owner's interest and time to spend, these photos are stored with some description such as title, captions and tags. Based on the study done by [Ames and Naaman \(2007\)](#), the motivation for tagging can be classified into two main reasons: personal and social. The owner might tag the photos for personal organisation and communication which will add future recall and to facilitate remembering details about the photos. By adding tags to the photos, the owner will ensure that the photos can be easily found by specific people with whom the user might want to share or they may be discovered by anyone who may be interested in the photos.

Nevertheless, these images are usually too often left without any or with very crude descriptions. The only way for the images to be found is by browsing the directories, their name providing usually the date and the description with one or two words of the original event captured by the specific photos. Although most Web 2.0 websites do provide text based searching to find images by mapping query concepts with words in image's title, description or tags, the access has become more difficult as the number of photos increased, the photos are ill annotated and the query is done by two to three words only ([Jansen and Spink \(2006\)](#) & [Spink et al. \(2001\)](#)).

For instance, searching for images related to two keywords *Tourism* and *Malaysia*, using the Flickr searching mechanisms returns 16,528 images using a full text method and 11,134 images using the tag based search². Images with less tags or description would be more difficult to find, while images with no tags and description will be useless, unsearchable by the text based search. To use the results, a user needs to filter the answers themselves by reading the returned documents. The practical effect of the situation is that at a certain stage, the user will be confronted with more information than they can effectively process. Information overload is a common term used by the Web information retrieval community to describe the fact that the information received becomes a hindrance rather than a help, even though the information is potentially useful (Bawden and Robinson (2009)). It is associated to three main factors, which are quantity, format and quality (Ho and Tang (2001)). Web 2.0 documents seem to fill in all the given factors, and as a consequence, Web 2.0 could lead to information overload (Heylighen (2004)).

In an attempt to overcome these problems, this thesis presents a hybrid approach. The hybrid approach is an integration of a method from text and a method from image analysis. The image analysis is used to analyse image content directly while the text analysis is used to tackle the text descriptions associated with the images. Each approach has its own advantages in the analysis. For example, in cases where images are well described, there would be sufficient text descriptions to use text analysis alone. On the other hand, if the images are lacking tags or descriptions or both, image analysis would have an advantage over the text analysis in handling such content. Therefore, we hope the two parts of a hybrid approach would compliment each other in providing better information retrieval performance.

1.3 Research Aims and Objectives

The aim of the research is to improve the capability of information retrieval by utilizing the Web 2.0 document content, the Semantic Web and other emerging technologies. We explore how to optimise the use of textual description generated by user rich experience for knowledge discovery, followed by image feature analysis to classify the images with similar features. This research will demonstrate the use of a text and image analysis hybrid in order to enhance information retrieval quality.

²The search was done on 16 June 2010

1. The main objective for the text analysis is to generate a semantic information model to describe the image. The sub-objectives for text analysis are presented as follows:
 - (a) to analyse and extract information in textual content by using Natural Language Processing and Ontologies.
 - (b) to expand information by using an online knowledge base such as Dbpedia and Geonames.
 - (c) to model information by using RDF description representations that are tailored with Semantic Web standards.
2. The main objective for the image analysis is to develop an image classifier exemplar to add classification information to images based on content. In the first instance, a classifier to tag images containing buildings is proposed. The sub-objectives for image analysis are presented as follows:
 - (a) To classify images based on visual features
 - (b) To evaluate the image classifier
3. Objectives for the hybrid approach evaluation based on information retrieval.
 - (a) to evaluate the semantic information model
 - (b) to evaluate the image classifier
 - (c) to evaluate the semantic information model and image classifier in the form of a hybrid approach.

1.4 Research Scope

We focus our interest into two main media, which are images and text. Web 2.0 documents from the Flickr website are used to initiate the experiments. Although there is MIR Flickr ([Huiskes and Lew \(2008\)](#)), a benchmark for image information retrieval developed from Web 2.0 resource, it does not provide sufficient text descriptions for text analysis because the images are only described with tags and EXIF Metadata. The Flickr website provides a good media resource for two reasons. Firstly, it is an online photo management website, thus the main media is image. Secondly, their users are encouraged to describe their own material (image) by using tags and free text description in title, caption and comment sections. This study will be focusing on the tourism domain of interest. In the tourism domain, we are interested to obtain information related to events and attractions. In order to initiate text analysis, 200 of the most relevant documents (based on Flickr ranking) that responded to the *Tourism* and *Malaysia* tag search are downloaded into our repository. As for image analysis, 1250 images ranging

from various landscape views are used. For evaluation purpose, 1034 Flickr documents (image and text description) are used. Since the study will be using English based linguistic analysis tools; Apple Pie Parser and GATE, only English based documents will be used for advanced analysis. A domain specific ontology, the Malaysia Tourism Ontology, is developed in order to support information discovery in the selected domain of interest.

1.5 Research Contributions

From this research, four major contributions can be identified:

- Contributions from text analysis are:
 1. A semantic information model to translate the textual content of Web 2.0 documents in a semantic manner. The semantic model would provide a better understanding compared to tags as the related concepts will be linked together. Furthermore, the semantic model is opened up by extending the model with the open knowledge bases, Dbpedia and Geonames ontology.
 2. Malaysia Tourism Ontology (MTO) which is developed to support information discovery tailored to the specific domain of interest. Since there is no specific ontology describing tourism in Malaysia available, MTO is developed to store information related to tourism in Malaysia and replace experts in the domain of interest. It consists of two main classes which are event and attraction.
 3. A prototype tool to assist the generation of the semantic information model. For the semantic context extraction and description of Image material is presented. The development of the annotation tool prototype should aid users to participate in semantic annotation and description of multimedia documents and the bridging of the semantic gap by producing a consistent description and a standard representation for image annotation. The representation follows guidelines of Semantic Web standards to provide appropriate annotations to populate multimedia content in the Semantic Web.
- Contribution from image analysis:
 1. A building landscape classifier is developed to identify city/building images. By analysing low level features of the image, a prediction is made to indicate whether the image is a building image or not. A prototype for the classifier can be used as a template to identify other sets of images such as sunsets, beach scene or mountain scene.

In this research, we have demonstrated a hybrid approach to integrate results from text and image analysis. Text retrieval and image based retrieval both have advantages in their own domain, thus it is valuable to investigate if these approaches complement each other to provide better information retrieval performance. We have evaluated our approach in two different situations, which are (a) images with adequate text descriptions, and (b) images with lack of text description. In the situation where text description is rich, text based searching has performed better in finding documents/images compared to image based search. In the situation where insufficient text description is available, text based search will fail completely and image based search will be useful to find the required information.

We have shown how text based search could be used to support image based search by producing a training set for the image classifiers. This feature could help the image classifier to identify images beyond the classification domain itself. Integrating both approaches in different situations has further alleviated the information searching. Therefore, the integrated approach of text and image analysis has generated a significant impact in improving the performance of information retrieval for user generated content in Web 2.0 documents.

1.6 Chapter Organization

The remainder of this thesis is structured as follows:

- Second Chapter : Literature Review

This chapter provides a literature review of the research areas. Firstly, it will cover topics such as multimedia on the web, Web 2.0 and the Semantic Web. Secondly, it will review some technical approaches related to text and image analysis towards adding semantic value in the information description followed by related work done closely to the research interest.

- Third Chapter : Methodology for the Hybrid Approach

This chapter will introduce our approach to integrating text and image analysis. Firstly, a preliminary justification analysis is done to identify issues and requirements for adding semantics in image representations. This is followed by a framework model for the hybrid approach.

- Fourth Chapter : Modelling Semantic Image Descriptions Using Text Analysis

This chapter will describe the text based approach to generate the semantic image information model. Text descriptions are analysed by using natural language processing and ontologies. The main objective is to identify and extract important information and link this information to produce the semantic information model and represent it in the form of Semantic Web standards.

- Fifth Chapter : Image Classification By using Image Content Analysis

This chapter presents an image based approach to annotate images. The image annotation is done by using an image content classifier. It demonstrates image classification for identifying images that contain particular object (in this instance buildings) by utilizing low level image features which are colour and edge. These features are analysed by using Bayesian Inference tools, and the classification performances rate is evaluated.

- Sixth Chapter : Evaluations: Results and Discussions

This chapter will provide experiments used to evaluate the hybrid approach against the individual approach. Three main experiments are used which are text, image and hybrid based search. The main objectives of the experiments are to evaluate each approach based on standard information retrieval performance metrics, precision and recall.

- Seventh Chapter: Conclusions and Future Work

This section provides a summary of the results obtained in the thesis. It briefly summaries how each of the objectives presented in Section 1.3 are achieved. Finally, a plan for future work is presented which would be interesting to tackle but out of the research scope for this thesis.

Chapter 2

Literature Review

This chapter is divided into two main sections. It commences with an overview of multimedia on the Web. Initially, an overview of web generations is provided. Next, it explores the Web 2.0 influences on multimedia growth, followed by multimedia representations on the web and then, Semantic Web standards. In the next section, the focus is on methods for analysing multimedia content, primarily in the text and image domains. It reviews state of the art approaches in image and text analysis directed towards analysis of Web 2.0 content. This chapter concludes by reviewing some related work close to our hybrid approach.

2.1 Multimedia on the Web

2.1.1 Web Generations

The early generation of the Web was developed to accommodate document sharing by research scientists on the internet. It provides the fundamental and friendly infrastructure to store and publish information using HypertText Markup Language (HTML) language and Hyper-Text Transfer Protocol (HTTP). With the enthusiastic acceptance by global users, Web content has been growing rapidly ever since. HTML is well known for its ability to provide an infrastructure to present and structure information as desired by its users. Nevertheless, HTML only offers little or non-explicit information about the content of the information itself to allow direct access by machines. As a result, with the increasing amount of unstructured information and crude information representation on the Web, information retrieval is facing information overload, making efficient and effective discovery of resources a hard task ([Montebello \(1998\)](#)).

The key element to alleviate information discovery in large numbers of websites requires improving the ability of information agents to communicate directly with one another and providing explicit information representations in a way that is far more accessible by information agents (Alesso and Smith (2008)). Towards these issues, the Semantic Web is introduced as an extension of the current Web in which content is given a well defined meaning, better enabling computers and people to work in cooperation (Berners-Lee et al. (2001)). The aim is to make the semantics of Web content explicit and machine-readable by concentrating on the meaning/semantics instead of Web content structure and presentation. The semantic web offers two key enabling technologies which are i) rich markup languages to annotate and describe information and ii) ontologies, a specification of conceptualizations to provide a formal shared description for the purpose of enabling information sharing and reuse (Gruber (2005)).

Extensible markup language (XML) was created to overcome HTML inadequacy. Like HTML, XML is based on tags, but with a different purpose. HTML tags provide information about the formatting of the documents. It only focuses on how to display the information and has fixed tags such as list, bold and colour. On the other hand, XML separates content from formatting. It focuses on describing the content itself by using unfixed tags to allow users to structure and define their own tags to best reflect the content of the information. Therefore, the XML description of the content would be explicit, structured and also machine-readable.

XML does provide description about data/metadata too but it does not facilitate any direct access to the semantics. Unlike XML, Resource Description Framework (RDF) is about adding semantics to the data. It is based on three fundamentals concepts which are resource, properties and statement. Resource is an object that we want to describe such as image, event or person. Properties refer to relationships between resources. Statement denotes object-property-value triplets which describe resources which have properties which have values. The values could be other resources or literals (instances). On top of RDF is an ontology language, a formal semantics to allow people/agents to reason about the information. RDF Schema (RDFS) and Web Ontology Language (OWL) are two examples for ontology languages. The main concern in RDFS is organising vocabularies in a type of hierarchy such as subclass and subtype relationships, domain and range restrictions, and instances of classes. Nevertheless, it does lack some features such as the ability to handle disjoint classes (e.g male and female).

OWL is an extension of RDFS with a full logic. There are three sublanguages of OWL, in decreasing complexity levels, which are OWL Full, OWL Description Logic (OWL DL) and OWL Lite. Each sublanguage has its own advantages and disadvantages. OWL Full is fully compatible with RDF, both syntactically and semantically, thus any valid RDFS is also valid in OWL Full. The drawback of OWL Full is that the language has become large and complicated and logically, dashing any hope of complete (efficient) reasoning support. The advantage of OWL DL is it permits efficient reasoning support by

restricting the use of RDF constructors. The disadvantage is it loses full compatibility with RDF. In a way, RDF documents need to be extended to be compatible with OWL DL, whereas all OWL DL is compatible with RDF documents. The OWL Lite is easier to implement but comes with restricted expressivity. The choice of which ontology languages to use depends on the level of expressiveness required to describe information in the ontology.

Semantic Web technologies can be used to support activities in annotation, presentation, sharing and retrieval of multimedia documents. Annotating multimedia content using a Semantic Web language such as RDF would improve the information representation by attaching metadata to the content or media it describes. For example, an image may have metadata properties added manually by the owner such as title and descriptions, and also metadata properties generated automatically during the production of the image which store technical details such as the device used to capture the image, focal length, flash (on or off) and aperture value.

Multimedia documents could also be annotated based on domain of interest by using domain ontologies such as geo spatial (Geonames), cultural heritage (CRM-CIDOC¹) and medical (MeSH Ontology²). These ontologies were developed by using the Semantic Web languages (RDFS and also OWL) and can be used to improve information presentation and facilitate information sharing among users. Furthermore, there are also multimedia ontologies such as MPEG-7 and COMM which could be used to facilitate information retrieval of media assets and documents containing them by providing a means to associate semantics with particular sections of the media assets. Thus it can be seen that Semantic Web technologies would be useful to bring benefits in improving multimedia content representation and retrieval on the Web.

Even though the Semantic Web does provide fundamental facilities for information sharing, it is too complicated for wide acceptance by the public thus limiting interest to those actively involved in its development. Web 2.0 is taking a different approach in information sharing. The term Web 2.0 was actually not introduced to refer to a vision, but to describe the current state in web engineering (Oreilly (2005)). Earlier, Web 2.0 was described as changing from hand written HTML to machine generated and often called active HTML pages. Later, Web 2.0 emerges into the Social Web - a user-centric publishing platform. Web 2.0 is characterised by tagging, social networks and dynamic pages generated by user contributions (Treese (2006)).

Folksonomy is a new term derived from active communities (folks) generating taxonomy, used to classify and index Web content using tags (Wal (2007) & Voß (2007)). The user supplies the content of Web 2.0 documents through dynamic and interactive web pages. Two key enabling technologies for creating dynamic and interactive web applications

¹<http://www.cidoc-crm.org/>

²<http://mklab.iti.gr/content/mesh-ontologies>

are a really Simple Syndication (RSS) and Asynchronous JavaScript and XML (Ajax). With RSS, Web 2.0 applications can integrate certain services from different websites to allow users to subscribe to content that is tailored to their needs or interest and can be sent the data in a format they request. Ajax allows Web applications to retrieve data from a server asynchronously in the background, inferring about the display and behaviour of the existing page (Moore (2008)). These technologies allowed web sites to be turned from a set of static documents connected by hypertext links into a customized page layout, quickly and dynamically changing its view and personalizing its content to meet the users' needs.

The Semantic Web and Web 2.0 are two distinguishable cultures yet sharing in the single Web. The Semantic Web marks its status by focusing on providing rich technical infrastructures to improve information accessibility and support information exchange. Web 2.0 treasures its users by providing an interactive and dynamic environment with silent technologies which reflect the future of the Web that is enriched with media content, interactive platforms and mostly user oriented (Nixon (2006)).

Sir Tim Berners Lee has anticipated the benefits of integrating Web 2.0 and the Semantic Web in the following quotation: *I think maybe when you've got an overlay of scalable vector graphics - everything rippling and folding and looking misty - on Web 2.0 and access to a semantic Web integrated across a huge space of data, you'll have access to an unbelievable data resource* (Shannon (2006)).


Echoing this statement, merging the Semantic Web and Web 2.0 to facilitate interlinked data and customisable portable applications seems to be one of the ways towards the next web generation (Kroeker (2010) & Silva et al. (2008)). Web 2.0 has become a collaborative knowledge source, locked in, poorly structured, with high subjectivity and often buried in low-quality content (Cena et al. (2009)). Thus the next generation of the web will help to filter the *crowd wisdom* to provide much more valuable information by sharing personal experiences that may benefit others. Semantics would be essential for having a more intelligent Web. It will allow information to be structured in a way that machines could read and understand it. Thus sophisticated queries and intelligent search agents could identify information needed, in the same way as we would do ourselves. One of the challenges now is to enable the general users who are unaware of the complexity of these technologies to create machine-understandable semantic content. Furthermore, it would also effect future search engines to be more media centric, taking media as input and be able to search not only based on text but also media content features. Thus, user interfaces need to evolve into intelligent and multi modal interactions.

Figure 2.1 shows some examples to clarify the differences between Semantic Web and Web 2.0. Web 2.0 is concerned with information sharing among humans, therefore most of Web 2.0 applications such as Wikipedia and Flickr provide friendly and interactive interfaces. The Semantic Web however is meant for information sharing among machines

(information agents), therefore information need to be presented in a way that can be understood by machines. The development of DBpedia is a clear effort to translate Web 2.0 content into Semantic Web formats. In this research, our interest is to represent Flickr document contents semantically tailored to the Semantic Web.

Web 2.0

Example: Wikipedia:



Mabul Island
From Wikipedia, the free encyclopedia

Mabul is a small island in the south-eastern coast of Sabah in Malaysia. The island has been a fishing village since 1970s. Then in 1990s, it has become popular to divers due to its beautiful scenery to Sipadan island.

Located 16km from Sipadan, this 20-hectare piece of land surface 5-10 meters above sea level, consists mostly flat grounds and several area of tall-shaded Sumbawa oak. A sandy beach, perched on the northwest corner of a large two square kilometer reef.

The Semantic Web

Example: Dbpedia :

```
<?xml version="1.0" encoding="utf-8" ?>
<rdf:app xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dbpedia="http://www.w3.org/2003/01/rdf-schema#"
xmlns:label="http://dbpedia.org/resource/Mabul"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:owl="http://www.w3.org/2002/07/owl#"
xmlns:geo="http://www.w3.org/2003/01/geo"
xmlns:skos="http://www.w3.org/2004/02/skos#"
xmlns:dbpedia-owl="http://dbpedia.org/resource/Mabul"
xmlns:dbprop="http://dbpedia.org/property"
xmlns:dbpedia-owl:abstract="http://www.w3.org/2002/07/owl#"
xmlns:dbpedia-owl:photo="http://dbpedia.org/ontology"
xmlns:lang="en"
>
<rdf:description of="http://dbpedia.org/resource/Mabul"
xmlns:dbpedia="http://dbpedia.org/resource/Mabul"
xmlns:dbprop="http://dbpedia.org/property"
xmlns:dbpedia-owl="http://dbpedia.org/ontology"
xmlns:lang="en"
>
Mabul is a small island eastern coast of Sabah in Malaysia. The island has been a fishing village in 1990s. It first became popular to divers due to its proximity to Sipadan Island. From Sipadan, this 20-hectare piece of land surface 5-10 meters above sea level, consists mostly flat grounds and aerial view is oval-shaped, surrounded by
```

FIGURE 2.1: Information translations from Web 2.0 to the Semantic Web formats.

The web has gone through different generations and optimistically there will be more to come. A new field called Web Science is an interdisciplinary approach to study the Web as an important entity (Berners-Lee et al. (2006) & Hendler et al. (2008)). Each web evaluation has its own power factor and issues. This research is only focusing on a small fraction of today's Web issues, analysing the content of Web documents for better retrieval, and there are other issues such as safety and legal rights which are pointed to by Shadbolt and Berners-Lee (2008). With massive multimedia repositories, it is hard to prevent people from taking advantage of illegal transmission of digital media properties such as picture, song and motion picture (Davis (2001)). Laws related to intellectual property and copyright for digital materials are already being debated. The web provides the medium for information sharing, and Web users need a better way to filter the reliability of the information, whether it can be trusted or not. Moreover, too much data exposure could lead to a safety risk (Felt et al. (2008)). Today, the social web has already been urged to provide safety features for secure environments (News (2010)). Therefore it is crucial to understand these issues and to be able to engineer the future web tailored to its need which would provide a better web (Hall et al. (2009)).

2.1.2 Multimedia Representation On The Web

In the early web, adding multimedia content to the internet was an unappealing task. The content needed to be coded by using HTML manually, or assisted with a crude interface. Most websites have a limited content allowance size thus most people were involved as end users, receiving the information and not as content providers. Crawler-s/spiders (software used in search engines to scan through document content) will filter this information and represent the content in the form of a bag of words (called indexes)

and used in information retrieval. As a result, only text content can be retrieved by this type of text based search.

Over time, web technologies have seen a profound improvement from bandwidth support to dynamic and interactive interfaces, thus shifting users from passive to active information providers. Today's web technologies have a significant contribution in adding multimedia content on the web. Its interactive platform brings a tremendous amount of informal knowledge added by user contributions. Specific skills are not required to join the communities and within a short period, a large number of people have joined in.

Web 2.0 shares these common features; users who become members will have the authorisation to upload and label their resources. Resources will be connected with tagged hyperlinks. Users may follow the links to other resources that share the same tag, browse other resources uploaded by other users and see how others label their resources. Users play two important roles as a content provider in Web 2.0; as authors and active viewers. Authors are responsible for uploading multimedia objects and describing them with terms or concepts that best reflect the content while active viewers interact with the website by giving feedback or comments. As a result, Web 2.0 applications are collecting a large amount of user experience data.

Community involvement in annotating multimedia content forms a Web scale collaboration activity and seems to be very supportive to cope with the increasing amount of multimedia content on the Web. Folksonomies arise when a large number of people are interested in a particular domain and are encouraged to describe it, creating a loose taxonomy ([Shadbolt et al. \(2006\)](#)). These folksonomies will become more stable, gradually maturing over time ([Hotho et al. \(2006\)](#)) and hope to provide a good solution to overcome knowledge acquisition problems - previously stated as the major issue for many knowledge based systems. Nevertheless, such folksonomies are too semantically loose to be able to guarantee sufficient accuracy in information content representation ([Nixon \(2006\)](#)). It does not provide semantic information about either terms in tags or links between these terms.

Although multimedia content in Web 2.0 is still based on HTML code, the content representation is slightly better by introducing tags, a list of keywords to describe the content explicitly to the user (human reader), yet still a bunch of words to the crawlers³. Therefore, web information retrieval remains the same, based on text centric and not media centric. There is a lot of research interested in multimedia information retrieval, but none has found a solution to tackle multimedia content at web scale. Most of this research is domain specific and only works for a small corpus. We will describe some of the approaches which are related to our interest later in the next section.

³The crawlers or also known as web spiders are programmed by search engines that browse around the Web and create an index of all collected information.

Considering the Semantic Web, there is a gap between textual and multimedia materials on the Web. Researches in textual material have already made a remarkable impact on the Semantic Web and there are many text content analysis approaches which have been applied successfully for extracting semantic descriptions to cope with the requirement of the Semantic Web. However, for multimedia documents, there is a significant growth in Web 2.0 but shifting to the Semantic Web is still in its infancy. The Web 2.0 intention is to provide flexibility in terms of presentation and user interaction primarily to human readers ([van Ossenbruggen et al. \(2001\)](#)) while for the Semantic Web, the main intention is on machine readable and processable content ([Berners-Lee et al. \(2001\)](#)).

The realization of the Semantic Web requires information to be explicit and accessible directly by machine. The Semantic Web technologies could be used to assist multimedia content. Firstly, it provides languages (such as XML, RDF and OWL) which can be used to represent multimedia content. Secondly, it has ontologies. Ontologies could play two roles in multimedia representation. Firstly, it works as an information provider to support information discovery in multimedia content analysis. Secondly, ontologies are used to provide standards for describing multimedia content. There are several ontologies introduced to standardize multimedia content description such as Dublin Core ([DCMI \(2008\)](#)), MPEG-7 ([Hunter \(2001\)](#) & [Arndt et al. \(2007\)](#)), COMM([Arndt et al. \(2008\)](#)). Nevertheless, they are not widely used for several reasons. Firstly, it is difficult and time consuming to annotate multimedia content manually. Secondly, the complexity of many standards makes the multimedia representation process difficult. These ontologies only provide a standard for describing the content, while the extraction of the descriptions and content annotation with the corresponding metadata is out of these standards ([Bloehdorn et al. \(2005\)](#) & [Nitta et al. \(2005\)](#)).

In this research, we are focusing on analysing multimedia content consisting of text and images which are gathered from Web 2.0 documents. The information related to images will be analysed and represented in the form of Semantic Web languages. These descriptions will reflect the image content tailored to our domain of interest. During the analysis process, domain ontologies and an open knowledge base will be used to support information discovery. Due to multimedia ontologies complexity, the representation of the multimedia description will be using Dublin Core standards, which will provide a simple standard in information representation. We do acknowledge the importance for proper multimedia ontologies to support integration across multimedia applications. Therefore, the prototypes will carry this limitation but we opened up the content representation by linking the representation with well known knowledge bases such as Geonames and Dbpedia. The standardization issue will be discussed further in the future work. In the next section, we will review some approaches to analysing multimedia components in Web 2.0 documents.

2.2 Layers of Semantic Image Descriptions

This section reviews some methods for describing still images. A survey performed by [Petrie et al. \(2005\)](#) has identified a list of important information required when describing an image on the web which are object, building, people in the image, what is happening in the image, purpose of the image, colour in the image, emotion or atmosphere expressed by the image and locations. Such information provides a general description about semantic information conveyed by the image. A number of eminent works are presented to express the richness of semantic content and content interpretable within an image. Earlier in 1962, Panofsky, an art historian, published three modes in describing an image of Renaissance art work; pre-iconography, iconography and iconology ([Panofsky \(1962\)](#)). Pre-iconography represents generic image subject matter, factual and expressional facets. Iconography indicates specific subject matter and iconology represents the symbolic meaning of the image.

[Shatford \(1986\)](#) extended the first two modes to characterize images into *what is it of?* and *what is it about?*. These correspond with the factual (objective) and expressional (subjective) components of pre-iconography; abstract or symbolic about-ness of the picture. Shatford's generalization of Panofsky's analytic modes beyond the realm of fine art has been formalised in a mode/facet matrix, in which each of Panofsky's modes represented more simply as Generic (pre-iconographic), Specific (iconographic) and Abstract (iconological) and extended the model further by breaking each of these three levels into four facets which are *who*, *what*, *where* and *when*.

[Jørgensen \(2003\)](#) has described a more elaborate conceptual framework, based on the same principles, called the Pyramid (Figure 2.2). This structure contains ten levels. Levels 1 to 4 refer to syntactic visual content. Of these, the first level describes physical typology, which describe what sort of image it is such as photograph, print or painting. Levels 2 to 4 refer to visual primitives of colour texture and shape. These syntactic or perceptual attributes are interpretation-free responses to a visual stimulus. The second layer which is Global distribution describes the overall colour or texture of the image. The third layer is Local Structure which describes the colour or texture of particular features within the image, followed by the fourth layer Global composition which refers to the arrangement or layout of these features. The rest of the levels (level 5 to 10) describe semantic or interpretative attributes which require both interpretation of perceptual cues and application of a general level of knowledge or inference from that knowledge to name the attribute ([Jørgensen \(1996\)](#)). The six semantic layers are defined in terms of generic, specific and abstract for object and scenes which can be seen as incorporating the Panofsky/Shatford model (generic, specific, abstract).

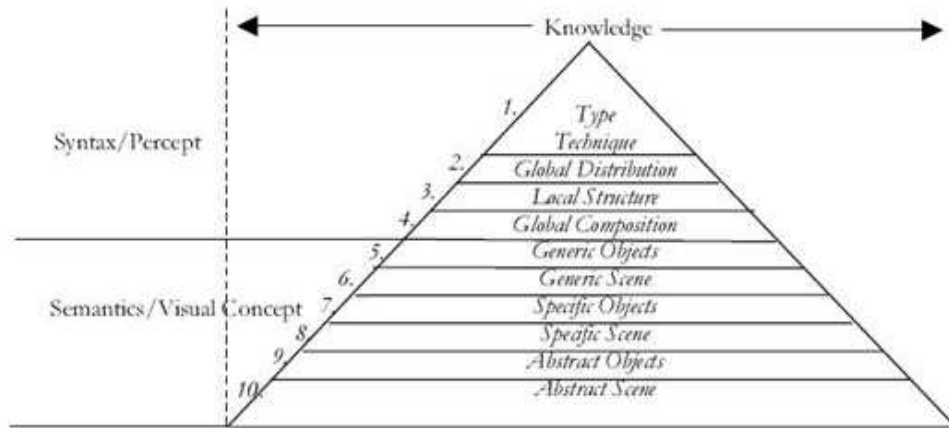


FIGURE 2.2: The Pyramid of Jörgensen (2003).

Eakins and Graham (1999) used a similar approach for the categorisation of queries, but using only three levels corresponding with visual primitives, logical or derived features and abstract features. A three-level hierarchical model of perception consistent with Eakins and Graham has been described by Greisdorf and OConnor (2002).

Hare et al. (2007) combined object, spatial, temporal and activity facets advocated by Layne (1994) as key elements of the subject component in document cataloguing (Table 2.1(i-iv)). Abstract and related concept facets and context and topic facets are added to capture the global semantic content of the image. The facets can be represented in terms of generic-specific classes and instance. The facets provide a conceptual framework for representing semantic content in terms of either natural language or a controlled, keyword based vocabulary, or both (Table 2.1(v)).

Acquiring such semantic descriptions is founded on collective knowledge, cultural conditioning, personal knowledge and experience (Enser (2008)). Furthermore, describing images is a subjective task where different people might have different views about an image (Shatford (1986)), and also different meaning to the same person at different times or under different conditions (Enser et al. (2005)).


Facets	Descriptions
 <p>Image ID: 153554716</p>	
i. Object Facet	
Generic object instance	Ocean, water, boats, sky
Generic name object class	Ocean view
Specific name object Class	Ocean village
Specific name object instance	Penyabong beach
ii. Spatial Facet	
Generic location	Outside, bay
Specific location	Penyabong, Johore, Malaysia
iii. Temporal Facet	
Generic time	Evening, sunset
Specific time	10 Jan 2007
 <p>Image ID: 473775905</p>	
iv. Activity/Event Facet	
Generic activity	Dancing, performance
Generic event	Traditional dancing
Specific named event	Colours of Malaysia
Specific event instance	Colours of Malaysia an annual event highlighting Malaysian arts, traditions and culture via performances dance and music
v. Abstract/Semantic Facet	
Topic	Traditional dancing
Related concept/object class	Dancers, traditional clothes
Abstract concept context	Dancers are performing traditional dancing during Colours of Malaysia festival.

TABLE 2.1: Different Layers of Facet Descriptions for Object, Spatial, Temporal, Activity and Abstract of Images.

2.3 Annotating Images with Semantic Descriptions

This section reviews some related work towards automatic image annotation. The reviews are divided into three main groups which are tag or folksonomy, text and image based approaches. Tag based approaches consist of some related work in tag propagation using tag recommendation systems and adding semantic value into tags. Secondly, the text analysis is focused on natural language processing and ontology based approach towards analysing free text description. It also covers a cross domain approach between tag and image analysis and text and image analysis. Finally, the image analysis is emphasised on scene and object identification techniques to support automatic image annotation which provides a general view and clue for the semantic content of the image.

2.3.1 Folksonomy or Tag Based Approach

Tagging is one of the most useful approaches in populating text description to describe images on the Web. Based on inspection, images are tagged with information such as location, artefacts/object, people/group, action/events and time ([Sigurbjörnsson and van Zwol \(2008\)](#)). Although words presented in tags are lacking semantics, a study done by [Al-Khalifa and Davis \(2007\)](#) shows that tags agree more closely to human generated keywords compared to those automatically generated from word analysis. Nevertheless, a tagging system has its own limitations. Tags depend entirely on the user to provide description and in some cases resources are left with no tag at all. In terms of navigation and retrieval, the tag mechanism is limited to keywords/tag based and tag cloud navigations. Tag based search also suffers from poor precision and recall due to basic variation problems such as people from different background and expertise may use different words to describe similar images, and also syntactic similarity resulting from singular and plural words, and also concatenated/misspelled tags. There is an increasing research interest in utilizing tags with folksonomies to alleviate the above limitations. We classified the tag based approach into two main efforts which are developing tag based recommendation systems and adding semantics to tags.

2.3.1.1 Reviews

exFlore ([Angeletou et al. \(2008\)](#)) was developed to enrich tags with formal semantics by utilizing the loose semantics from the folksonomies as well as a formal knowledge resource. The system analyses the context surrounding both tags and resources to determine their correct sense, and then assigns them to formal semantics. The result is a semantic layer that imposes a clearly defined structure of the tags and resources. Adding semantics to tags enables formal queries and intelligent search on the original resource space.

[Sigurbjörnsson and van Zwol \(2008\)](#) is focused on utilizing tags to recommend a list of tags that can be added to a photo. Their image dataset comes from Flickr database. They used Wordnet to classify tags into five main classes which are location, object/artefact, people/group, action/event and time. The distance between two words in a tag is calculated by using symmetric and asymmetric normalization metrics. In their research, they are using an asymmetric metric to measure how often the two tags occur together. The evaluation is done by comparing tags defined by the user, tags recommended by the system and tags accepted from the recommendation tag list. Their result shows that the recommendation system has generated useful additional tags to the user-defined tags especially for two classes which are location and object or artefact.

Research done by [Moxley et al. \(2009\)](#) and [Rattenbury et al. \(2007\)](#) are focused on learning tag semantics by using geo-referenced images in the Flickr collection to identify tags related to events and places. Event tags are defined as tags whose distribution is expected to follow temporal patterns while place tags are expected to have significant spatial patterns. [Rattenbury et al. \(2007\)](#) has introduced a Scale-structure identification (SSI) approach to extract information about events and places. SSI is used to measure how similar the data is to a single cluster at multiple scales [Rattenbury et al. \(2007\)](#). The idea behind SSI was if tag x is an event then the points in T_x , the time usage distribution, should appear as a single cluster at many scales. The approach was tested using the Flickr dataset focusing on images related to San Francisco Bay Area. The SSI approach is claimed to perform better than Naive Scan and Spatial Scan methods in identifying tags to be an event or place.

[Moxley et al. \(2009\)](#) has extended work done by [Rattenbury et al. \(2007\)](#) by adding visual term extraction and landmark detection approaches. The visual term extraction approach is used to find tags which are visually identifiable from image content (such as sky, beach and flower). A landmark identification approach is used to detect tags which describe landmarks by querying Dbpedia database which lists georeferenced Wikipedia entries to extract the proper name of the landmark. The additional information provided by the approaches has provided a better result in identifying tag places compared to SSI approach alone. TagEz ([Anderson et al. \(2008\)](#)) and Tagr ([Lindstaedt et al. \(2008\)](#)) are developed to predict tags for Flickr images by using multimodal features which are tags and images. The tag analysis is done by using asymmetric metric in order to calculate the relatedness between one tag to another which is similar to [Sigurbjörnsson and van Zwol \(2008\)](#). The image analysis is integrated with ALIPR tool⁴. The ALIPR tool is used to predict a list of keywords from a new image by inferring basic texture features

⁴ALIPR is a semi automatic tool to assist users in image annotation and image discovery. New images will be analyzed by using image analysis where low level features are extracted and inferred to predict tags. A list of tags will be suggested to the user and users may choose words listed in the tags or describe the image with their own words. Next, the system tool will predict the feeling conveyed by the image by listing a list of affective words which the user can choose from or providing the information based on their own feeling. The interface of ALIPR can be found in <http://alipr.com/>.

and segmentation of the image. Borda voting⁵ is used to aggregate both tags and image components. The evaluation is done by using a test set of 924 images with some original tags for each image. Standard information retrieval matrix is used to assess the system by comparing tags recommendation by analysing tags and the image. Their result shows that their tags analysis has widely outperformed image analysis. The low score for image analysis was due to the use of Coral images as the training set.

Tagr's domain of interest is images of fruit and vegetables gathered from Flickr group. Tagr consists of five modules which are image classification, image similarity, user similarity, tag association and Wordnet term association. The image classification automatically produces one or more tags for an untagged image. The images are analysed based on colour and texture features, and for the classifier, they used the SVM approach. The images are classified into a number of fruit and vegetable classes such as banana, blueberry and orange, and the classification accuracy rate is up to 71% (Lindstaedt et al. (2009)). Image similarity is done by analysing colour features to explore user defined annotations of visually similar images. Wordnet is used to find synonym and hypernym sense between tags. Tags distribution is based on the statistical distribution of tags to calculate tags similarity. User similarity is calculated based on similar tags used by a similar group of people. Tagr's evaluation is based on a user based study to assess the usefulness of tag recommendation to users based on each module. The result shows that their test users felt most supported by tags distribution and Wordnet module which give direct support for the task of finding the right tags. The image classification and similarity give conditional support, depending on how similar the received image is with the image to be tagged. The users felt least supported by the user similarity.

2.3.1.2 Discussion

A comparison for each research for tag based analysis is presented in Table 2.2. As stated earlier, we have divided the tag based approach into two main groups which are tags recommendation systems and adding semantics to tags. exFlore (Angeletou et al. (2008)), Moxley et al. (2009) and Rattenbury et al. (2007) are some research examples which are interested in enriching semantic value into tags. ExFlore has used ontologies to discover sense of relation between tags. Rattenbury et al. (2007) has used tag's usage patterns based on timeline and locations to extract semantics for tags. The approach was extended by Moxley et al. (2009) with additional visual term extraction and landmark detection approach. ExFlore and Moxley et al. (2009) have integrated ontologies in their work with different uses. ExFlore used ontologies to transform a flat folksonomy tag space into a rich semantic structure while Moxley et al. (2009) has integrated Dbpedia knowledge base to identify tags associated to landmarks. Sigurbjörnsson and van Zwol (2008), Tagz and Tagr are some examples for research which is focusing on the second

⁵Borda voting is originated from election theory.

group, tags recommendation approach. In Sigurbjörnsson and van Zwol (2008), tag recommendations are done by using statistical methods to identify clusters of related tags without defining the exact relations among them. Tagz and Tagr have integrated tags and image based approaches in automatic tag propagation. All of these researches have used the Flickr dataset.

Research	Approach	Domain	Test Set
exFlore	adding semantic to tag	General	Flickr
Sigurbjörnsson and van Zwol (2008)	tag recommendation	General	Flickr
Rattenbury et al. (2007)	adding semantics to tag	Place and Landmarks	Flickr
Moxley et al. (2009)	adding semantics to tag	Place and Landmarks	Flickr
TagEz	tag recommendation and image analysis	General	Flickr
Tagr	tag recommendation and image analysis	Fruit and veg-etable	Flickr

TABLE 2.2: Comparisons for tag based related analysis

In comparison to such experiments, our research also shares a similar interest which is trying to translate rich user experience into semantic structure but with a broader scope, where we are not only focusing on tags, but also image caption, title and user comments are taken into account. Such information is usually in the form of free text, therefore natural language analysis is required. In the next section, we review work related with natural language analysis. As far as we are concerned, most research in analysing text description in the Web 2.0 domain is focusing on tags only, thus this review covers natural language analysis in web documents in general.

2.3.2 Natural Language Processing Approach

Text is the most used medium in delivering information including describing other media such as image, video and audio, thus it is crucial to be able to identify and represent such content in an efficient and effective manner. Natural language processing (NLP) plays an important role in indexing of multimedia documents (Declerck et al. (2004)). NLP techniques could be used to extract semantic triplets (a set of two concepts and the relation between them) from the caption text, which constitute higher level indexing than isolated keywords (Rowe and Guglielmo (1993)). The semantic triplets could be used to generate semantic indexes which can then be searched in a semantic manner instead of keywords based search. This section reviews some NLP related research in analysing images and also Web document content.

2.3.2.1 Reviews

Névéol et al. (2009) has integrated text and image analysis to tackle multimodal documents in the medical domain. The image and text documents consists of 180 radiographs with text descriptions (captions or paragraphs describing the image). In the text analysis, NLP integrated with IRMA (Lehmann et al. (2003)) and MeSH knowledge bases are used for the semantic indexing task. The image analysis is done by integrating texture features, the cross-correlation function (CCF) and the image distortion model. Such information is analysed by using a k-nearest neighbour classifier. Their experiments show that image analysis provides a better result in indexing and retrieval compared to text analysis. Integrating text and image analysis has outperformed both independent image and text analysis in the indexing task but not in retrieval where the image analysis gave similar results to the integrated.

Pastra et al. (2003) has used an NLP approach to analyse images of crime scenes with captions. The aim of the approach is to extract relationships between objects in the image description by using a named entity analyser to identify specific entities. The caption is pre-processed by using the GATE tool (Cunningham et al. (2002)) which consists of a simple tokeniser that identifies words and spaces, a sentence segmenter and a named entity recogniser. In order to identify specific terms in the named entity analysis, Ontocrime, a specific hierarchy ontology which contains structures of relevant concepts to crime scene investigation is used. Extracted information is stored in the form of a triple, object - relationship - object. The same semantic representation is then used to support image retrieval by using free text queries. Similarity scores between query triples and indexing triples in domain specific ontology are computed by the retrieval system. Furthermore, the specific ontology also can be used in query expansion.

The use of NLP and ontologies to support information extraction in analysing Web document content is demonstrated in Artequakt (Alani et al. (2003)) and Noah et al. (2009). Artequakt searches the web and knowledge about artists is extracted based on specific ontology and stored in a knowledge base in the form of triplets (subject-relation-object). The ontology was constructed from CIDOC Conceptual Reference Model (CRM) ontology to present the domain of artists and artefacts. Two NLP tools, GATE and Apple Pie Parser (APP) are used to assist text analysis. APP is used to classify grammatically related phrases such as noun phrases and verbs while GATE and WordNet are used to identify named entities.

Noah et al. (2009) focuses on web documents related to the medical domain. A specific ontology for heart disease is developed based on Medical Subject Heading (MeSH)⁶ to support information extraction and semantic information modelling. The information

⁶The Medical Subject Headings comprise the National Library of Medicine's controlled vocabulary used for indexing articles, for cataloguing books and other holdings, and for searching MeSH-indexed databases, including MEDLINE (MeSH (2010))

extraction is divided into two subsequent stages; syntactic and semantic analysis. Syntactic analysis is done by integrating the APP tool. The semantic analysis is performed by extracting the semantic relationships between the selected concepts, which are done by either the use of domain specific knowledge automatically or by analysing sentences with the help of the user. The extracted semantic information is then presented in XML. In terms of the concept extracted, the approach has a better result compared to the keyphrase extractor presented by KEA (Witten et al. (1999)).

2.3.2.2 Discussion

A comparison of the use of NLP approaches to analyse text descriptions from each research project is presented in Table 2.3. In this review, all of the research that has used NLP methods is in specific domains such as the medical domain (Név      et al. (2009) and Noah et al. (2009)), cultural and heritage (Alani et al. (2003)) and security (Pastra et al. (2003)). All of the approaches were integrated with Semantic Web technologies such as (a) specific domain ontologies to assist the text analysis in information identification, and (b) semantic languages such as XML and RDF to improve information representation and retrieval. The textual description comes from a different medium such as descriptions or captions of images and web content. In general, the text description used by these researches are provided by experts in each domain of interest. In N      et al. (2009), the NLP and image analysis was integrated and a better result was produced compared to using each approach on its own. Based on these reviews, there are two most used NLP tools which are GATE and APP.

Research	Approach	Test Set
N����� et al. (2009)	NLP and Image Analysis	Radiographs with descriptions
Pastra et al. (2003)	NLP and domain specific ontology	Crime Scene Images with captions
Alani et al. (2003)	NLP and domain specific ontology	Web documents related to artist and cultural heritage
Noah et al. (2009)	NLP and domain specific ontology	Web documents related to heart disease

TABLE 2.3: Comparisons for NLP related work in text analysis

Our research shares a similar approach with N      et al. (2009), where the NLP approach is integrated with image analysis to tackle two different media, text and images. Nevertheless, they have used images with a text description which is provided by domain experts while our documents are generated by users from different knowledge levels. We have provided some comparison between expert generated content with user generated content in the first chapter.

2.3.3 Automatic Image Annotation

Capturing the semantic description of images is usually motivated by the need for improved Content Based Image Retrieval (CBIR). CBIR is interested in organizing images based on their visual features such as colour, texture and shape. Query by Image Content (QBIC) (Flickner et al. (1995)) and VisualSEEK (Smith and Chang (1996)) are some CBIR examples which support content based retrieval by the visual features such as *'find images with a red round object'*. Nevertheless, these visual features do not allow users to query images by semantic meaning and most users are familiar with high level concepts which are normally presented in the form of text. In order to overcome this issue, image annotation is required to label image with visual terms which provide semantic understanding about the image content. The most common approach in performing automatic image annotation is by using object or scene classifications (such as Szummer and Picard (1998) and Serrano et al. (2002)). There are also other approaches such as semantic spaces techniques (Hare et al. (2008)) and co-occurrence model to keywords and low level features (Mori et al. (1999)). In this section, it focuses on the earlier technique of automatic image annotation which is by using object and scene classification. Image classification is used to identify object or scene classes by using machine learning techniques such as probabilistic classifiers k-nearest neighbour and support vector machines.

2.3.3.1 Reviews

Szummer and Picard (1998) has used colour and texture features with the k-nearest neighbour (KNN) technique to classify indoor and outdoor images. The colour features are based on Ohta colour space⁷ and the image is divided into 32 bins per-channel (32 x 3 channel) while the texture features are computed base on multiresolution, simultaneous autoregressive model (MSAR). The research has used an image dataset from Kodak. The test was divided into two strategies which are by calculating features in the whole image and secondly by dividing the image into 4 x 4 sub blocks and computing the features separately over each block. Szummer and Picard (1998) reports classification performance of 75.6% for colour features and 83.0% for texture feature in the first strategy. In the second strategy, each sub-block is classified independently, followed by another classification on a result generated independently of block. Their result shows that, individual sub-block classifiers have produced less accuracy compared to a whole image classifier. In their report, integrating both features, colour and texture, has produced a stronger result which is 90.3%.

⁷(Ohta color histogram was used here as colour features. The Ohta colour space is a linear transformation of the RGB space, its colour channels is defined by Yu-Ichi Ohta and Sakai (1980): $I1 = (R+G+B)/3$, $I2 = (R-B)/2$ and $I3 = (2G-R-B)/4$. $I1$ is the intensity component, whereas $I2$ and $I3$ are roughly orthogonal colour components, these two channels somewhat resemble the chrominance signals produced by the opponent color mechanisms of human visual system M. et al. (2002)).

[Serrano et al. \(2002\)](#) has improved the approach to indoor/outdoor classification done by [Szummer and Picard \(1998\)](#). The test was done by using similar low level features (colour and texture) and image dataset (Kodak). Compared to [Szummer and Picard \(1998\)](#), they are using LST colour space instead of Ohta colour space and, wavelet decomposition to compute texture features compared to MSAR. The wavelet decomposition is used to reduce feature dimensionality, therefore decreasing the classification complexity. The Support Vector Machine (SVM) is chosen over KNN due to the fact that it theoretically produces more efficient classification than KNN classifier ([Serrano et al. \(2002\)](#)). Using a similar strategy in [Szummer and Picard \(1998\)](#), the classification rate for an entire image is 74.5 and 83.0, for colour and texture respectively. The performance rates for the second strategy are 70.7 and 74.4 for colour and texture respectively. Similar to [Szummer and Picard \(1998\)](#), the second strategy has produced a lower result because there are fewer and weaker signatures in image subsections. Although with less feature dimensions, integrating both features has achieved a success rate of 90.2%.

[Vailaya et al. \(2001\)](#) & [Vailaya et al. \(1999\)](#) have used colour and edge features with a Bayesian framework. The classification is done based on hierarchical classification of vacation images. Firstly, the image is classified into indoor or outdoor. A subset of outdoor is furthered classified into city or landscape, followed by a subset of landscape is then classified into sunset, forest and mountain classes. Indoor/outdoor classification is done by inferring 10 x 10 sub-block colour using LUV colour space, while city/non city is classified by observing edge directions histogram and finally sunset/forest/mountain classification is identified by using colour features in HSV space. The performance rate for indoor/outdoor classification is 90.5% which is comparable to [Szummer and Picard \(1998\)](#). Edge features has provided the best individual performance rate of 95.3% for city/landscape images. For landscape classifications, colour feature has provided the best accuracy of 96.6 for sunset/forest classification and 96% for forest/mountain classification.

[Luo and Savakis \(2001\)](#) has used low level features (colour and texture) and mid level/semantic features (grass and sky) with a Bayesian approach for outdoor and indoor classifier. The colour feature was calculated based on Ohta colour space with 64 x 3 bins while the texture features was based on MSAR. The research has used an image dataset from Kodak. The classification performance rate for colour and texture are 74% and 82% respectively, while integrating both features has increased the performance to 82.3%. The grass and sky features are identified based on colour/texture feature classification with of 95% correct with 10% false positive. Integrating semantic features with low level features has improved the performance rate to 84.7%, which is better than using low level features alone.

[Payne and Singh \(2005\)](#) has performed indoor/outdoor classification by only concentrating on edge features. The tests were done on less than 900 in a mixture of vacation photographs and comparing two methods which are Rule based approach and KNN.

The Rule based approach does not require any training set. The classifier result for the Rule based approach is 87.7% which is better than KNN (84.5%).

[Hervé and Boujemaa \(2007\)](#) has used colour, shape, texture and local edge orientation histogram with Support Vector Machine to classify scenes based on the nature of the image (such as artistic representation, colour photograph, black and white photograph) and the context of the image (such as indoor or outdoor, day or night, and natural or urban). They are using HSV colour histogram (120 bins), integrating colour and shape (218 bins), and texture and colour with (218 bins). The edge histogram is used to assist in urban image classification. The tests were made on ImageEVAL benchmark. The performance was calculated based on 13 classification queries and the Mean Average Precision (MAP) shows that integrating colour and texture has improved the classification performance compared to using only colour and edge histograms providing a huge contribution for queries related to urban natural images.

[van de Sande et al. \(2010\)](#) has used different colour descriptors (such as RGB and HSV) to classify scene and object. Such colour descriptors with different invariance such as invariance to light intensity, light intensity shifts and light colour change are tested. Furthermore, the SIFT descriptor which describes the local shape of a region using edge orientation histograms is also used. The analysis is done by employing SVMs, a similar method as presented by [Zhang et al. \(2007\)](#). The test images are provided by PASCAL Visual Class Challenge which contains nearly 100000 images of different objects such as aeroplane, bicycle, horse and person. The result shows that integrating colour and SIFT variants have the best performance compared to others in classifying objects in the test set.

2.3.3.2 Disucssion

A comparison for each research for automatic image annotaion is presented in Table 2.4.

In general, most of the reviewed works have focused on natural and artificial scene classification. Features such as colour, texture, shape and edges have been used in the classification. Research done by [Oliva et al. \(1999\)](#) provides general clues to classify natural and artificial scenes based on edges/shape features. Artificial scenes such as can be characterized by vertical and horizontal structure, for example a city building composed of tall building exhibiting vertical structure, while more balance vertical and horizontal directions (cross shape) for indoor scenes such as kitchen or living room. Natural panoramic scenes such as beach and field can be characterized by a vertical line. Oblique line (mainly orientations at 45 degree plus or minus 15 degree) can be found in natural images such as mountain, canyons and valleys. Circular orientation is commonly identified in highly texture environments such as forests and fields. In [Hervé and Boujemaa \(2007\)](#) & [Vailaya et al. \(2001\)](#), they both agree edges are a good identifier

Research	Scene/Object Classified	Features	Classification Method	Test Set
Szummer and Picard (1998)	indoor/outdoor	colour, texture	KNN	Kodak
Serrano et al. (2002)	indoor/outdoor	colour, texture	SVMs	Kodak
Vailaya et al. (2001) & Vailaya et al. (1999)	indoor/outdoor, city/landscape, sunset/forest/mountain	colour, edges	Bayesian	General vacation photograph
Payne and Singh (2005)	indoor/outdoor	edges	Rule based approach, KNN	General Photograph
Luo and Savakis (2001)	indoor/outdoor	colour, texture	Bayesian	Kodak
Hervé and Boujemaa (2007)	artistic representation, colour photograph, black and white photograph, indoor/outdoor, day/night, natural/urban	colour, texture, shape, edges	SVM	ImageEVAL
van de Sande et al. (2010)	objects and scenes classified on test set	colour, edge	SVMs	PASCAL VOC Challenge 2007

TABLE 2.4: A comparison for related work in image classification

for classifying building/city images and integrating edge with other features such as colour does improve the performance. In most research, natural scene classification is by using colour and texture features. In comparison, colour features have performed better than texture and integrating both of these features has yielded a better performance rate ([Hervé and Boujemaa \(2007\)](#) and [Serrano et al. \(2002\)](#), [Luo and Savakis \(2001\)](#)).

The most common test set used in these analyses is Kodak and Coral dataset. Kodak test set only consists of 1342 images which consists of typical family and vacation scenes while Kodak was criticised for its simplicity and being too easy([Hervé and Boujemaa \(2007\)](#)). Other researchers used images of vacation photographs from unknown sources ([Vailaya et al. \(2001\)](#) & [Payne and Singh \(2005\)](#)). For the purpose of analysing information related to Web 2.0 documents, none of the image sets reviewed in the related work provide images with user generated text description. Even MIR Flickr fails to meet the requirement needed for this research. Therefore, we have gathered our own images from Flickr API. The source allows us to download not just images, but also text description associated to the images, which is required for image and text analysis respectively.

In term of classification strategies, Bayesian, SVM and KNN are imperfect. KNN is criticised for having a poor run-time performance thus, generally slow and difficult to determine the correct value of k from the validation set ([Serrano et al. \(2004\)](#)). The SVM model run time is also time consuming and it is too large to be used in a practical system with limited space ([Zhang et al. \(2002\)](#)). The Bayesian approach disadvantage is the normality assumption in general pattern recognition literature and neural networks are usually hard to optimise for generalisation ([Payne and Singh \(2005\)](#)). It is also hard to compare the classifier performances between these researches due to two main factors. Different researches have used different learning approaches and different benchmark/image sets to test on. It is recorded that the Bayes classifier has the ability to classify images more effectively compared to SVM and KNN classification methods in image segmentation ([Rahimizadeh et al. \(2009\)](#)). Hence we use a Bayes based machine learning tool from Microsoft called Infer.NET⁸ for our classifier in Chapter 5.

⁸Infer.NET is a framework for running Bayesian inference in graphical models. It can be used to solve image classification problems by using light weight programming based on C#.

Chapter 3

Methodology For The Hybrid Approach

This chapter describes our work in modelling semantic image information in Web 2.0 documents. Firstly, it reports preliminary work to justify Web 2.0 content issues. Requirements identified for adding semantics to image representations are presented. Next, it introduces the hybrid framework model to integrate text and image analysis to tackle the given issues.

3.1 The Preliminary Analysis

The aim of this preliminary analysis is to observe user generated content in Web 2.0 documents. In detail, we observed the text description which is used to describe the image in order to assess its usefulness for retrieval and potential for enriching and structuring the document content. The objectives of this analysis are:-

- to analyse the text description generated by user contribution by using simple text analysis.
- to observe the content of text description and its compatibility to be used in image description

Most Web 2.0 research tends to focus on utilizing tags and in this research we broaden the scope to all the text descriptions that describes the images which include title, caption, tags, and comments. We believe such items contain fruitful information for image description. In order to set up the analysis, we used Web 2.0 document content provided by Flickr.

3.1.1 A Brief Description about Flickr

To initiate the study, we have created a multimedia document corpus originated from the Flickr website. Flickr is an exemplar of a Web 2.0 website - an online photography management website that provides a means for photo publication, storing, sharing and searching (Flickr (2007)). Figure 3.1 summarizes the processes and flow of user generated content in Flickr. Flickr provides an interactive environment for users to create the entry by adding the image, assigning information to the image (tags, title, caption) and classifying an image based on user interest. After the entry is published to the public, other members may provide additional information to the image based on their point of view or interest by adding more tags, leaving feedback/comments or inviting the image to a group that share the same interest. Usually, the length of the web document is gradually increased by these activities. Figure 3.2 shows an example of Flickr entries. Flickr uses text based searching mechanisms and it will return results by comparing the search query with words in the title, caption or tags.

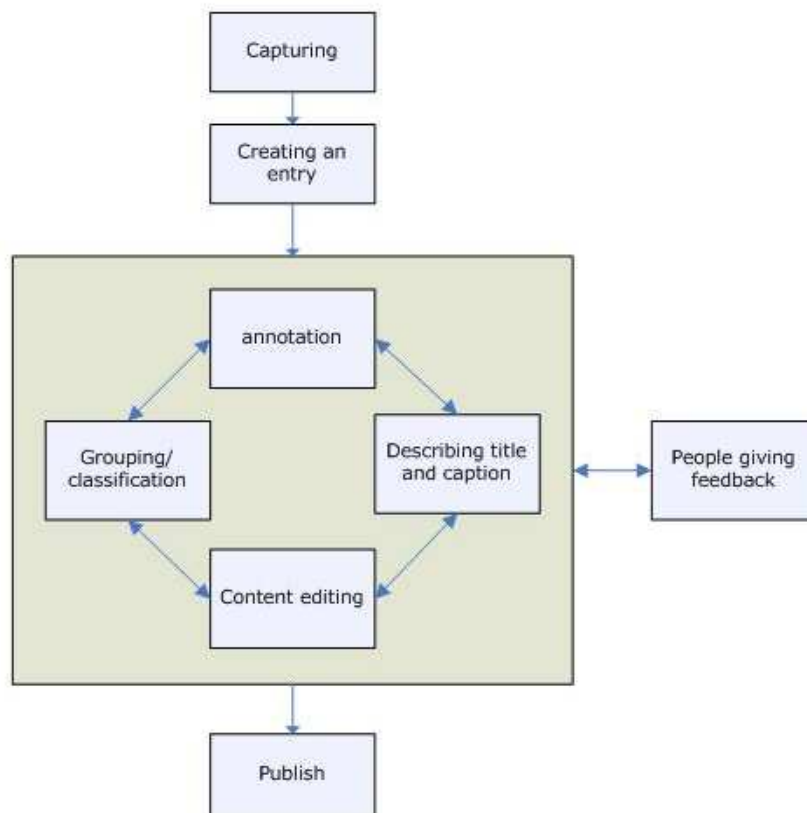


FIGURE 3.1: Process flow of user generated content in Flickr.

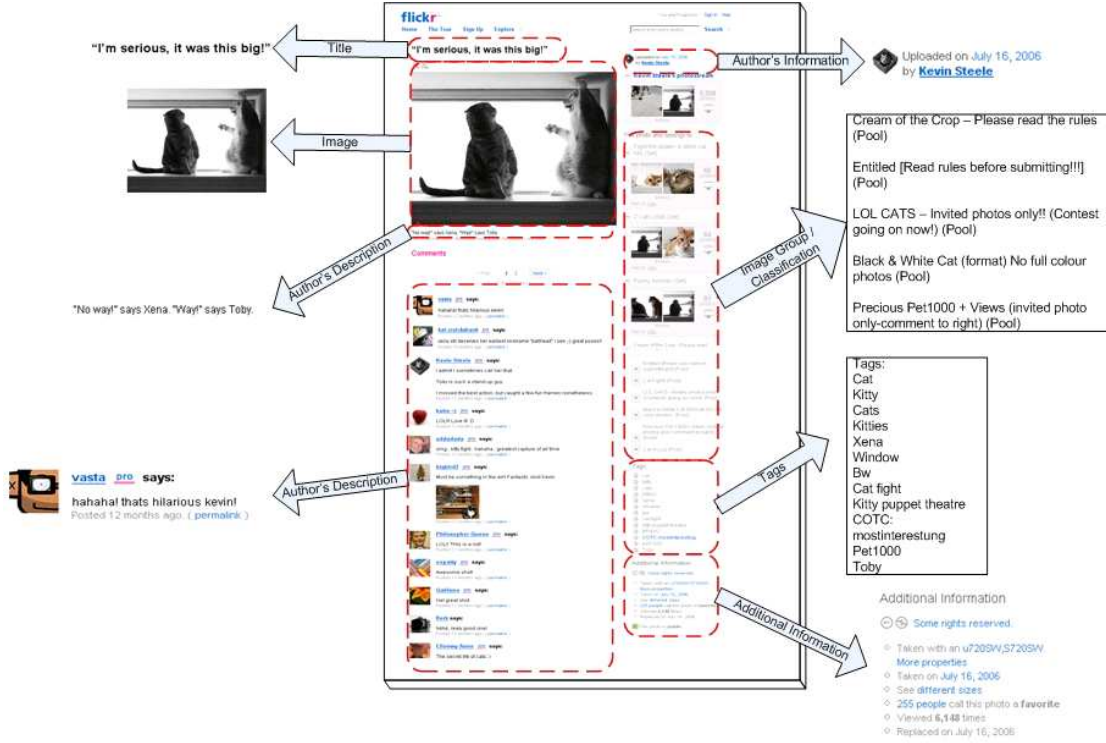


FIGURE 3.2: An Example of Flickr Entries.

3.1.2 Experiment Setup and Word Analysis

Flickr offers information search based on tags and full text. The users queries will be matched to the tags or the full text. However, these searches will only covers the text description in tags (tags based search) and in both the tags and descriptions (full text based search). Text descriptions generated in the comments section are inaccessible by the retrieval. Therefore, we could not achieve our aims in the preliminary analysis by looking at frequencies over whole of Flickr. Nevertheless, tags and full text based search could be useful to generate tag clouds which would provide a general view of word or tag frequencies used in any domain of interest. Therefore, we have gathered some documents from Flickr which are tailored to our research needs.

In this experiment, text descriptions will refer to title, captions, tags and comments. 44 Flickr documents related to two domains, travel and cat, were used to initiate the preliminary research. We have only chosen 44 documents in order to allow us to understand the basic structure and document contents in detail. The main concern of this preliminary analysis is to observe textual content describing an image which is generated by user contributions. The result of the analysis will be used to establish whether the whole textual content have more information to describe the image compared to just using text from tags alone. The analysis was done by calculating the proportions of words matched in two sets which are the tags and the full text set. The sample was relatively small, enabling us to observe these documents in great details.

Earlier in this research, the Flickr documents were downloaded manually from the websites. These documents are in the form of HTML tags therefore it needs to be cleaned in order to be able to analyse with text analysis. A simple tool was created by using JTidy¹ to generate plain text from the Web documents.

Next, the plain text is sent for word analysis. The plain text is presented in the form of a list of words and each word is passed through three sub analyses. Firstly is stop words analysis. A word that is matched to a word in the stop word list is eliminated. The stop word list contains words such as *are, they, you, a, an, but* and *of*. A full list of stop words used in this research is presented in Appendix B. Such words need to be removed because they do not carry any semantic meaning about the document content and tend to occur frequently in any text documents. Next is Porter Stemmer analysis. Each word is stemmed into its root word by using the Porter Stemmer algorithm (Giustina (2009)). Finally, the words are calculated based on their root words and a list of word frequencies is generated.

3.1.3 Tags and Text Content Observation

Tags are labels or words used by user to describe an image. The free text descriptions are a set of information added by users which are title, tags, captions and also feedback by other users. Here, we have observed the information diversity in tags and free text descriptions. Based on our observation, the range of the number of tags used with the images in the Cat domain is between 6 and 30 words. Most words listed in the tags consist of objects occurring in the image such as *cat, kitten, window* and *stair*. Some words might be unknown, for example words such as *gato, chat* and *neko* which are other words for cat in Portuguese and Spanish, French, and Japanese respectively. To understand such labels prior linguistic knowledge may be required. In general, tags also consist of other information such as:

- personal information - specific names of objects.
- the activity or event in the image - such as *playing* and *jumping*.
- abstract concepts - such as *funny* and *adorable*.
- the instrument used to capture the images such as *Canon EOS 5D*

In the travel domain, the range of words labelled in the images is between 5 and 85 words. Words listed in the tags vary and they characterize objects, time, location and activity information of the images. For example, in document titled *Through the Mist* (Figure A.1 in Appendix A), the tags contain words such as *abstract, people, yellowstone, mist, fog, geysers, life, travel, camping, wyoming, national parks, yellowstone national*

¹JTidy (Jtidy (2009)) is an HTML parser that can be modified into a tool for cleaning HTML tags

park, *moments*, *summer* and *interestingness*. Such words can be classified into layers of semantic description as discussed in Chapter. The word *interestingness* is an example of Flickr related terms which are used by Flickr.

- object : *people, fog, geysers*
- time: *summer*
- location: *Wyoming, national park, yellowstone national park*
- activity/event: *camping, traveling*
- semantic or abstract concepts : *sweet memories*

Some tags also contain information of monuments or buildings represented in the images. For example, a document titled *Kids Having Splash Behind Taj* (Figure A.3 in Appendix A) is labelled with 85 words. Words such as *mausoleum*, *dome*, *tomb* and *taj mahal* reflect the generic or specific object class and instance. Moreover, words that are related to it also contain words that reflect related concepts of the object such as *shahjahan* and *mumtaz*. The image has also been labelled with *17thcentury* that might indicate the time of the monument creation.

Unlike tags, word frequencies provide information about the most used words in text descriptions. The word frequencies have captured the other side of the image characteristic. After inspection, words that reflect the affective level of the images dominated and scored among the highest in the word frequency lists. Some common examples for such words are *adorable*, *beautiful*, *funny*, *fun*, *amazing*, *lovely* and *silly*. It also contains words which are useful to describe the images but not listed in tags. For example, in image ID T3 (Figure A.2 in Appendix A), the text analysis has identified words such as *sky*, *blue*, *engineers* and *building* which are missing from tags and useful to describe the image.

3.1.4 Text Content Comparison

The list is then compared with words in tags. Tags are words which are manually picked by users to describe their material. Research done by Al-Khalifa and Davis (2007) shows that tags can be used to represent document content more than automatically generated keywords. By assuming tags are baseline information to describe the image, we calculate the average number of words listed in the word frequency table which matched to one of the tags. The result will provide brief information about the proportion of words/information used to describe the content of the whole document if only tags were used. The formulation to calculate the word portion is presented as follow:

$$\text{Percentage} = 100 \times T \cup F/T$$

Where, F denotes number of words in the word frequency, T refers to number of words occurring in the tag list and $T \cup F$ represents the number of matched words between the frequency list and the tag.

Table 3.1 shows results for the percentage of words matched between tags and word frequencies for each analysed document. The proportion average for cat and travel domains are 0.47% and 0.32% respectively. It is clear that that focusing on tags only reflect less than 50% off the overall textual content. Therefore, analysing tags alone would miss the other words which might be useful to describe the image. We choose to analyse all the text descriptions that are surrounding the image in order to fully utilize information provided by users' contributions. The word analysis might be too crude to generate adequate words/concepts because sentences are chunked into words without considering any phrases. A more advanced text analysis is presented which includes natural language processing to provide proper language analysis and ontologies to provide prior knowledge for information identification.

Doc. ID (Cat)	Percentage	Doc ID (Travel)	Percentage
C1	0.73	T1	0.25
C2	0.67	T2	0.30
C3	0.57	T3	0.40
C4	0.71	T4	0.67
C5	0.50	T5	0.33
C6	0.50	T6	0.47
C7	0.40	T7	0.25
C8	0.24	T8	0.27
C9	0.58	T9	0.19
C10	0.60	T10	0.45
C11	0.82	T11	0.46
C12	0.43	T12	0.29
C13	0.38	T13	0.34
C14	0.65	T14	0.41
C15	0.20	T15	0.17
C16	0.36	T16	0.30
C17	0.36	T17	0.10
C18	0.28	T18	0.25
C19	0.23	T19	0.20
C20	0.45	T20	0.50
C21	0.38	T21	0.25
C22	0.23	T22	0.19
Average (Cat)	0.47	Average(Travel)	0.32

TABLE 3.1: Percentage tags in the whole text description

3.1.5 Discussions and Conclusion

The result shows the *beauty*, the *neutral* and the *ugly* side of tags and text content generated actively by viewers.

The beauty : As stated earlier, tags may provide additional information that might not be represented in the title or the description. By inspection, we conclude that users tend to describe the affective aspect of the image. Therefore, words that are expressing feeling and thought are captured and listed as among the highest in the word frequency list. It shows the beauty side of user's involvement in web document generation. These user comments are similar to a conversation, thus the information given is at a higher semantic level. They share and express feelings and thoughts, and words such as *funny*, *adorable*, *beautiful* are captured with high frequency. Such words represent the abstract concepts of the image, which is more difficult to capture by any advanced image processing. However, further analysis is still required to establish whether the semantics of such words are in the right context or not.

The neutral : These labels provide potentially useful information to represent the image, but they are not very meaningful on their own where they are discrete and lack semantic information. It would be very useful if we could clarify all of this information, classify it into layers of semantics such as object, spatial, temporal and event or activity. For example, information for location can be represented in hierarchical order such as *Agra is part of India*, and *India is part of Asia* instead of just lists of words *Agra*, *India* and *Asia*. In order to clarify, classify and semantically characterize the image, firstly, we need to have prior knowledge related to the images.

The ugly : Viewer comments may generate too much information that is stereotyped, redundant and lacking meaningful information for other viewers. It would be much more informative if the comments added by viewers contain information that will enrich the knowledge for the author and other viewers. Moreover, sometimes, not all words used to label the image reflect the content or context of the image.

The study has shown significant problems with the descriptions of images in Web 2.0 documents which may be summarised as follows;

- Tags do provide additional information to the images, but they are too discrete (independent) and too loose (unconstrained). Tags are embedded with hyperlinks only for the purpose of navigation for images that share the same label.
- Information is coded using HTML tags - human friendly (useful for information display) but providing no meaning for information retrieval.
- Image grouping or classification is done manually by creating a group and inviting users who have images of common interest.

- Dynamic content generated by user comments contains valuable information describing the affective/semantic level of the images.
- The Flickr searching mechanism only compares the search query with words in the title, caption and images. Content generated by user comments are completely ignored during search processing.
- Web 2.0 documents depend on users to include information and most of the Web 2.0 users are not skilled to provide the appropriate information for more general retrieval.
- The use of abbreviation, short forms of words and phrases, in describing information or providing feedback in comment sections such as *“Luv this piece! @ wat time did u take this pic?”* (taken from comment in document ID: 153550434).

Based on the analysis and observation, we conclude that text descriptions do contain valuable information to describe the image. In order to tackle the looseness in Web 2.0 documents, we would like to investigate a means to add semantics to generate more information that offers meaningful description related to the images. For the purpose of information reuse and information retrieval, information needs to be represented using Semantic Web standards. In order to overcome the problems, several requirements have been identified:

1. Adding more meaningful information and connecting related information in a hierarchy or as a semantic link. It will provide more clear descriptions for the information.
2. Embedding a layer of information representation conforming to Semantic Web standards to allow easy access for information retrieval mechanisms and to ease information reuse.
3. Provide prior information related to the domain of interest. For example, for the tourism domain, information such as tourism event, place of interest, main attractions and list of activities that can be done is useful to be included to enrich images related to tourism.
4. Classify information based on the image’s semantic layers and its visual features. Information could be classified based on semantic layers such as objects, locations, time and activities or events which need to be identified. Image visual features could be analysed to provide additional information in order to identify items in the image scene.

3.2 Introduction to the Hybrid Approach

This section will introduce our hybrid approach in order to fully utilize the richness of user generated information in Web 2.0 documents and fulfil the requirements listed in the previous section. The framework of the hybrid approach consists of two sections which are text and image analysis. Figure 3.3 presents the framework for the Hybrid Approach.

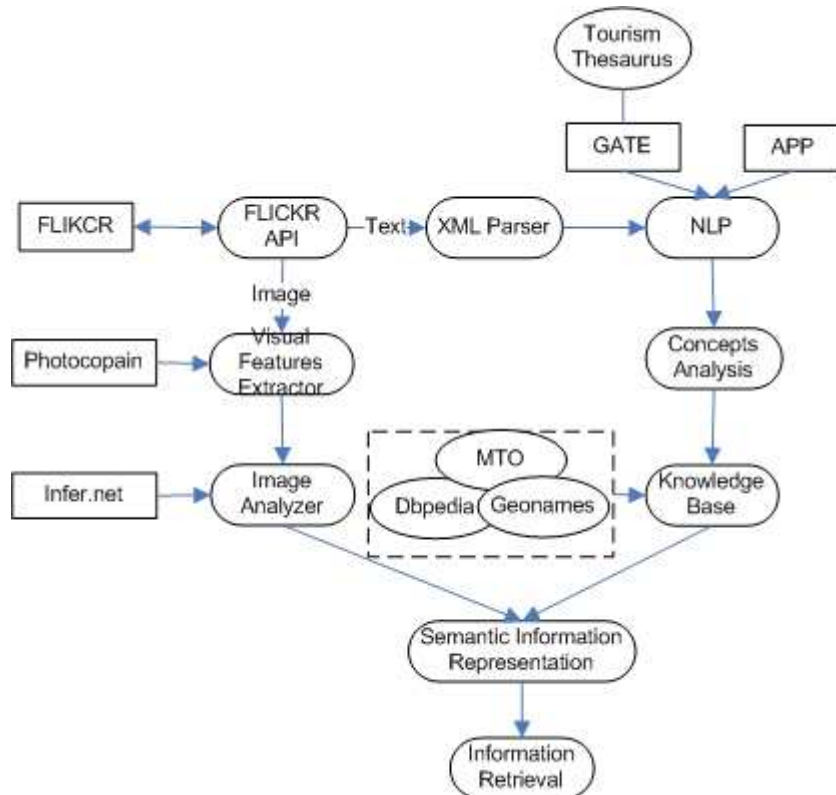


FIGURE 3.3: The Hybrid Approach Work Flow Framework

The aim of text analysis is to translate text description into a more structured semantic information representation. Firstly, textual description is parsed to generate plain text for natural language processing (denoted as NLP in the Table 3.3). The NLP is integrated using two linguistic analysers which are GATE and Apple Pie Parser. The linguistic analysers will parse sentences into a parse tree, a syntactic structure of words according to some formal grammar, which contains valuable information to identify concepts. Identified concepts are then mapped to concepts in the knowledge bases and the accumulative information from the analysis is presented in a semantic information representation. Chapter 4 describes the text analysis in detail.

The aim of the image analysis is to be able to annotate images by using an image classification approach in order to identify object or scene according to their visual content. The image analysis starts with extracting visual features from the images. Two visual features are considered which are colours and edges. The Photocopain tool

(Tuffield et al. (2006)) is used to extract edges and convert the information into lines. Features are presented in the form of a histogram and submitted to the Image Analyzer. The image analyzer is developed based on the Infer.net tool to provide an engine to analyse such features and classify the images. The image classification information could also be represented using a semantic representation. Chapter 5 represents the development of the image classifier in detail.

The hybrid approach evaluations are presented in Chapter 6. The assessments are made using standard information retrieval evaluation techniques. The evaluations are divided into three stages which are text, image and the hybrid approach. In order to initiate the research, a multimedia corpus is developed and presented in the next section.

3.3 Multimedia Corpus

The aim of this stage is to create an accessible multimedia corpus in order to replace the tedious work of manually selecting documents from Flickr in the preliminary work. Earlier, the documents were manually downloaded from the Flickr websites and stored in HTML form. The content of these documents was extracted by creating a parser to remove the hard coded HTML tags. We then improved our multimedia document corpus by developing a tool to automatically search, extract and download images and descriptions from an authorized Flickr database.

The tool is created using the Flickr API and Java application. In order to connect and pull information directly from the Flickr database, a set of numbers (keys) is required and these numbers can be obtained from Flickr. The tool allows information downloading for 200 Flickr documents per-search (which was the maximum allowance of images that can be downloaded using the Flickr API at that time (December 2008)), thus creating a bigger corpus for a larger scale study. Nowadays, Flickr has increased the maximum allowance of the images to 4500 images per-search. The corpus consists of images and their corresponding textual descriptions, which are restructured into an XML representation to improve content accessibility and ease information reuse for further analysis. The collection of multimedia documents is gathered by searching either by keyword or tag matching, which so far, will return the most relevant public images added by the Flickr users.

The following information is extracted from the database: Image title, URL, caption, list of tags and feedback or comment from other users. Such information is restructured and stored into an XML representation as follows:

```

<photo id="">
  <owner nsid="" />
  <title> Image Name </title>
  <url> Image URL </url>
  <description> Image Description</description>
  <geo>
    <latitude></latitude>
    <longitude> </longitude>
  </geo>
  <tags>
    <tag id="N" > </tag >
  </tags>
  <comments>
    <comment id="N" > </comment >
  </comments>
</photo>

```

note: *N* = numbers of tags /comments

The collection of multimedia documents in our corpus is created based on tag searching. As stated in the research scope, the study will be focusing mainly on documents related to the tourism domain of interest in Malaysia. To initiate text analysis, 200 documents that responded to *Tourism* (AND - unification) *Malaysia* tag query were selected. Since the query is set to find the most relevant documents based on the Flickr ranking approach, these documents have a variety of document lengths generated by the users. In the early stage of the study, we were using the most recent documents added to the Flickr database. However, searching for new documents means most of the returned documents lack additional information which comes from other users. Based on observation, the images gathered in the corpus represent locations, festivals and people in Malaysia. Furthermore, each of the documents is given an identification similar to Flickr's image identification to allow us to track the document in the Flickr database if it is required. Each document will be stored in two different folders holding the image description and the image its self, respectively. Table 3.2 shows an example of a Flickr document stored in the corpus. To initiate image analysis, more than 1000 images with different scenes (such as beach, island, mountain, forest and city) and different places from around the world were downloaded. Finally, for the hybrid approach evaluation, more than 1000 documents related to Tourism and Malaysia were added to the multimedia corpus. These documents were carefully selected and irrelevant documents were filtered. The irrelevant documents denote documents which contain images from outside Malaysia and sensitive images.

TABLE 3.2: Sample of information generated and stored in the multimedia corpus

Image Example:



XML description examples:

```

<title> MALAYSIA </title>
<url> http://www.flickr.com/photos/jingle526/1203148615/ </url>
<descriptions>This former palace of the Sultan of Perak was built without the use of a single nail.
It is now a museum. It's on EXPLORE Aug. 29. </descriptions>
<tags> <tag1>Malaysia</tag1> <tag2>kualakangsar</tag2> <tag3>perak</tag3>
<tag4>travel</tag4> <tag5>tourism</tag5> <tag6>perspective</tag6> <tag8> asia</tag8>
<tag9>impressedbeauty</tag9> </tags>
<comments> <comment1>Really stunning, just so awesome. XXX wow, so beautiful,
Nice change of scenery jingle526 :) </comment1>
<comment2>This is an amazing looking building and couldn't believe it was built without a
single nail! </comment2>
<comment3>Beautiful capture! Your fantastic color picture is my winner! Please add this
photo to www.flickr.com/groups/colorphotoaward/ <comment3>
<comment4> its in kuala kangsar...</comment4>
<comment5>I love the color and the window shapes Beautiful architecture,
nicely framed.<comment/5>
<comment6>I get to travel around the world with flickr! A very beautiful building and
a great shot! Such a fantastic building. That is a sight to behold Neat capture! I guess
they used pegs instead of nails?</comment6>

```

3.4 Conclusion

In this chapter, we have presented the preliminary data collection and analysis prior to processing free text descriptions in Web 2.0 documents. In general, the user generated content can be classified into three main groups we name beauty, neutral and ugly classes. The preliminary results indicate significant issues with describing images in Web 2.0 documents and a list of recommendations is presented to overcome such issues. In this research, we have proposed a hybrid approach using both text and image content to fulfil such recommendations and these are discussed further in the next chapter. Finally, the multimedia corpus was presented to initiate the research.

Chapter 4

Modelling Semantic Image Descriptions using Text Analysis

This chapter concerns the first stage of the hybrid approach which involves text analysis supported by knowledge bases. In this stage, we analyse user generated information within the textual description of the image by using natural language processing tools and ontologies. The main objective of the analysis is to identify and extract important concepts and represent these concepts using the language of the Semantic Web.

4.1 Work flow in Modelling Semantic Image Descriptions

Based on observations in our preliminary work (see Chapter 3), user generated content in Flickr textual descriptions do provide valuable information to describe the images. Nevertheless, the rich user experience information needs to be extracted and represented semantically in order to improve its accessibility for effective reuse and retrieval. Flickr textual descriptions such as title, caption, tags and comment are stored with the images they represent. The interest of the research is to identify information related to tourism such as locations, attractions and events in these textual descriptions. This information will be identified, extracted and modelled as a semantic description. The objectives of the text analysis approach are:

1. to obtain a list of concepts that are useful to represent the image.
2. to link and expand these concepts via the knowledge bases
3. to represent the information semantically in the standard form of a Semantic Web language such as RDF

The workflow for generating the semantic information model is illustrated in Figure 4.1. In general, text analysis can be divided into two phases; general information identification and specific information identification. A detailed description of each phase is covered in Section 4.2 and 4.3 respectively. The general and specific information identification is achieved by using Natural Language Processing tools. The tools are used to help us identify concept candidates that can be used for image description. The ontologies provide the required information related to the domain of interest and also guidelines for ensuring consistency in information description. Results from each analysis will be translated into the semantic information model in RDF format.

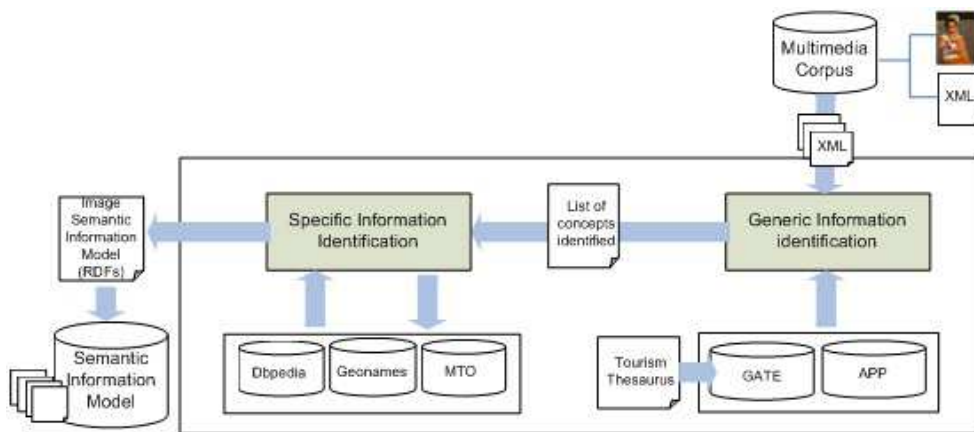


FIGURE 4.1: The construction of the semantic information model

4.2 General and Specific Information identification and Extraction

Generally, the Natural Language Analysis (NLP) process can be divided into two sequential stages; the morphology and syntactic analysis; and the semantic analysis. The morphology and syntactic analysis is the generic information identification process while the semantic analysis is the specific information identification process.

The generic information identification component consists of two NLP tools which are GATE (Cunningham et al. (2002)) and the Apple Pie Parser (APP)(Skine (2008)). GATE is used to recognize specific elements which are already predefined in its knowledge base (gazetteer). By default, GATE is useful for identifying information such as the names of people, times, locations, addresses and organizations. In order to maximise GATE capabilities for our application, we have enriched the GATE knowledge base by adding a tourism thesaurus. The tourism thesaurus contains lists of the most common concepts that can be used to describe information related to the tourism domain. The thesaurus is provided by the World Tourism Organization (WTO (2008)).

Table 4.1 shows some samples of information added into the GATE knowledge base in order to help GATE to identify concepts related to the tourism domain. Four main concepts in the tourism domain have been added to the knowledge base; environment, attraction, transportation and activity. The full list of these concepts are listed in Appendix C. The location information consists of states and capital cities of Malaysia are also included into GATE. Such information will help us to capture information relating to tourism in Malaysia in general.

Attraction	Activity	Transportation	Environment
theater	acrobatics	air	beach
national park	folk dance	airplane	mountain
craft	adventure	bicycle	island
garden	firework	bus	city
architecture	festival	camel	forest
building	aerobics	car	coast
amphitheatre	paragliding	coach	city
ancestral home	hiking	ferry	city centre
city hall	dance	funicular railway	riverside
palace	scuba diving	helicopter	lake

TABLE 4.1: Common Concepts in the Tourism Domain

APP is a light weight domain free analyser that can handle incomplete sentences, thus making it very suitable to handle text from user generated documents, which is sometimes poorly structured (e.g. incomplete or including abbreviations). Figure 4.2 shows an example of APP output – a syntactic parse tree that represents the syntactic structure of words based on formal grammar. Syntactic labels of each of the words in the sentence were obtained, such as *noun phrase*, *prepositional phrase* and *verb*. The parse tree can be used to extract noun phrases which are a good indicator for identifying concepts. For example, in Figure 4.2, four concepts can be extracted which are *sunset*, *kuala beach*, *langkawi* and *kedah*.



FIGURE 4.2: Visualization of Parse Tree generated by APP

Compared to GATE, APP is a domain free application. These extracted concepts could be used to find clues about the content of the document. In order to help us finding more information about the concepts, it will require more knowledge from other resources such as ontologies and knowledge bases.

Each of the concepts identified by GATE and APP is submitted to the next process, concept analysis. The concept analysis stage will refine the concepts, whereby each concept will undergo word stemming and concept frequency analysis. In this analysis, the Porter Stemmer ([Giustina \(2009\)](#)) is used to parse words to their root words. Nevertheless, over parsing is a common error in word stemming. For example, word *festival* could be parsed into *festive* while *computer* could become *compute*. In cases where there are articles in the front of nouns such as *the beach* or *a beautiful beach*, the articles (such as ‘*the*’, *a* and *an*) will also be removed. Sometime, the extractor will encounter noun phrases that contain adjectives and affective words such as the highest mountain or beautiful sunset. Even though the study is not focusing on the affective aspect of describing information, we do support the use of affective words such as beautiful sunset for describing an image instead of just sunset to facilitate a higher level of semantic information description. Furthermore, a sound semantic image description does require feeling and thought. Table 4.2 shows an example result generated from the generic information identification and extraction process.

The GATE and APP analysis will produce a list of concepts which will be submitted to the next analysis; specific information analysis. In this analysis, we translates these isolated concepts into a semantic model. An ontology is used to assist the translation process because it provides shared knowledge which could be used to classify these concepts with similar conceptual references.

Text descriptions example:

Cotton Island. The spectacular sunset over the jetty was a sight to behold. Pulau Kapas or Cotton Island inherited its name from the native because of its incomparable white beaches. Surrounded by crystal clear ocean, Pulau Kapas promises a spectacular getaway from the hustle and bustle of city life to quiet natural retreats with abundant sunshine and crisp clean air. An island renowned for its clear waters, sandy white beaches and swaying palms, it is relatively isolated. Home to an infinite variety of hard and soft corals, the waters around the island abound with sea-shells, fish and turtles... ..

Concept	Class	Frequency	Root word
Islands	Environment	4	Island
Island	Environment	4	Island
Kapas island	-	1	Invalid*
Pulau Kapas	-	2	Invalid*
beaches	Environment	2	beach
Terengganu	-	1	Terengganu
Snorkeling	Activity	1	snorkel

TABLE 4.2: Text Analyser Component Output. Invalid* refers to root word for concept that is more than one word. Text descriptions for this example is presented in Appendix D

4.3 Specific Information Identification Using Ontologies And Knowledge Bases

The semantic analysis refers to linking isolated concepts (extracted from the analysed text description) that share the same conceptual reference (such as beach, island, mountain are referred to as attractions). In the general information identification, the tourism thesaurus was used to provide a list of concepts that are commonly used in the tourism domain. However, specific ontologies and knowledge bases are required to capture the semantic information about tourism in Malaysia, which substitute for experts in the domain of interest. It is vital to provide the information needed to identify concepts that are related to the domain. In this direction, three knowledge bases were used which are Malaysia Tourism Ontology, Geonames Ontology and Dbpedia Open Knowledge base.

4.3.1 Malaysia Tourism Ontology (MTO)

The Malaysia Tourism Ontology is a specific domain ontology, which is created to store information related to tourism in Malaysia. MTO development was based on the Harmonise Ontology. The Harmonise Ontology enables information-based business to exchange travel and tourism information between organizations. MTO consists of two main roots which are *Attraction* and *Event*. The ontology was developed by using Protege presented in the form of OWL. In general both the *Event* and *Attraction* root will have information such as names, descriptions and locations. An *Event* is a temporal

activity, therefore it has a *timeline* property that indicates when and how long the event is going to be. An *Attraction* has a *type* property which will provide type of attraction such as *national park*, *historic building* or *island*. If the entry for both *Attraction* and *Event* has more than one name, the alternative name will also be stored to increase the chance of finding related concepts. Figure 4.3 illustrates properties for Event and Attraction in the tourism ontology. Figure E.2 and Figure E.3 in Appendix E show Protege screenshot examples for adding event and attraction instances.

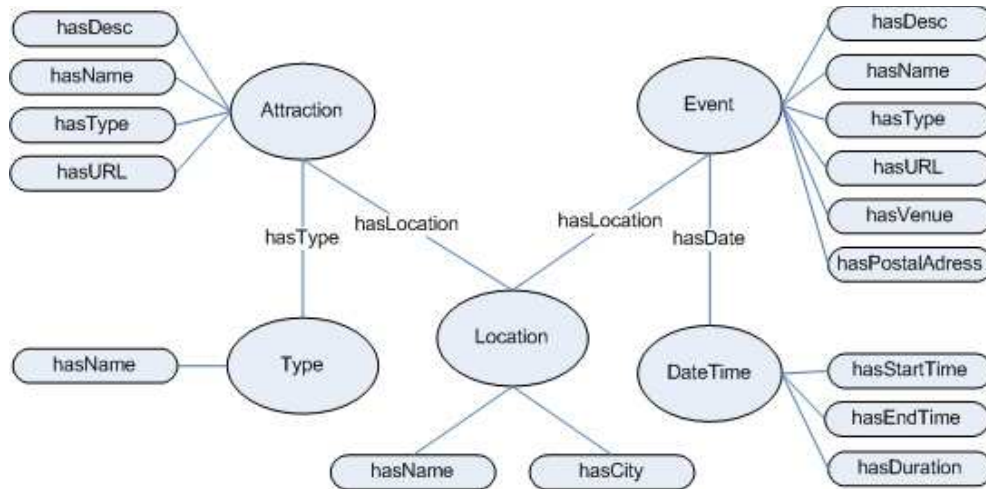


FIGURE 4.3: Main roots in the Malaysia Tourism Ontology. Descriptions for each root is presented in Appendix E

In order to capture the semantic information about tourism in Malaysia, the ontology is enhanced with Tourism Malaysia instances. These instances are gathered from the Ministry of Malaysia Tourism Portal ([Malaysia \(2009\)](#)) and Virtual Malaysia Portal ([Tourism \(2009\)](#)). Instances are added into MTO ontology and were gathered manually from the portals using Protege. Figure 4.4 illustrates the instances added to the MTO.

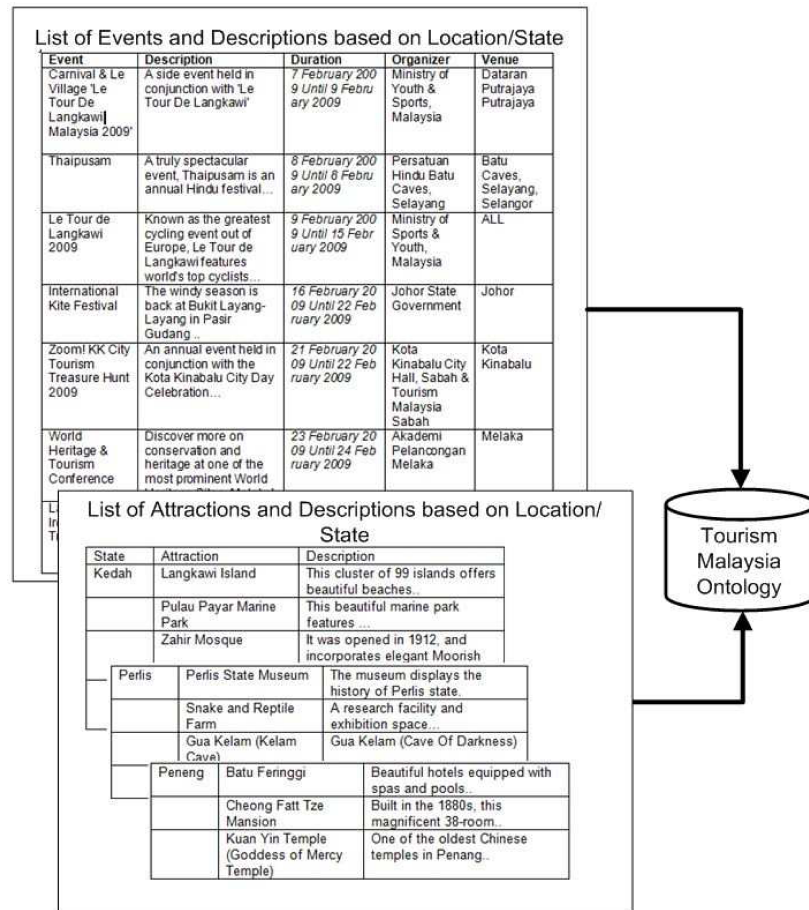


FIGURE 4.4: Sample of information collected to develop the tourism ontology for Malaysia.

Table 4.3 shows *kapas island* one of the entries added to MTO. In this example, the *kapas island* entry has two names which are *kapas island* (*hasName kapas island*) and *pulau kapas* (*hasName pulau kapas*) and has *island* attraction type (*hasType island*). The concept is also linked to its source in the Malaysia Tourism Portal (*hasURL http://www.tourism.../item.asp?item=pulaukapas*) and linked to where *kapas island* is located. In this example, *kapas island* is located in *Terengganu*. Instead of listing the location in the form of a concept, information for the attraction's location is extended to information about Terengganu in the Geonames Ontology (*hasPostalAddress http://www.geonames.org/1733036/*).

Property	value
hasName	Pulau kapas, kapas island
hasDescription	Popular among Malaysian and international travellers alike,....
hasType	island
hasURL	http://www.tourism.../item.asp?item=pulaukapas
hasPostalAddress	http://www.geonames.org/1733036

TABLE 4.3: Entry example for Attraction: Kapas Island

The following example (Table 4.4) shows information about *Citrawarna festival*, one of the entries added for *Event*. Similar to the previous example, it has names such as *citrawarna*, *colours of Malaysia* and *citrawarna Malaysia*. This event is celebrated all around Malaysia (*hasVenue all around Malaysia*) and the celebration lasts for 1 month (*hasDuration 1 month*).

Property	value
hasName	Citrawarna, colours of Malaysia, citrawarna malaysia
hasVenue	All Around Malaysia
hasDescription	A nationwide Colours of Malaysia extravaganza, lovingly dubbed Citrawarna..
hasDuration	1 month

TABLE 4.4: Entry example for Event:Citrawarna Festival

4.3.2 Geonames Ontology

Geonames is a geographical database which provides spatial geographical information about locations. We use the Geonames ontology to help us identify concepts that can be associated with a location. Other related parameters about the concept can be extracted to expand the knowledge about the concept such as longitude and latitude of the location, the geographical features and other names used to describe the place. Table 4.5 shows two entries in Geonames matched to concepts *kapas island*.

Parameter	Entry 1	Entry 2
ID Location:	1734708	1734707
Location:	Pulau Kapas	Pulau Gumia
Country:	Malaysia	Malaysia
Feature:	Mountain, hill, rock,...	Mountain, hill, rock,...
Lat:	5.1266667	3.233333
Long:	103.266667	103.233333
Alt Name Long:	Pulau Kerengga, Pulo Kapas	Pulau Kapas Kechil
Admin Name:	Terengganu	Terengganu

TABLE 4.5: Geonames entries matched to *Kapas Island*

4.3.3 DBpedia

Dbpedia is an open knowledge base created by the community to extract structured information from Wikipedia and to make this information available on the Web. Dbpedia uses the RDF language for representing the extracted information. Dbpedia is used to expand our information by matching the concepts found in the document against the Dbpedia dataset. Unlike Geonames, DBpedia is created based on the collaborative work of extracting information from the Wikipedia website which has been producing a wide range of information from multiple knowledge disciplines. Results given by Dbpedia may

vary depending on the availability of structured information. For example, the entry for an *island*, *Kapas Island* would consist of location (such as latitude and longitude), while the entry for a building such as *Petronas Twin Tower* consists of information such as surpassed by building, length of construction and how many floors it has (Table 4.6 and Table 4.7).

Concept	Kapas Island
Resources	http://dbpedia.org/resource/Category: Island_of_Malaysia ,
Abstract	Kapas Island or Pulau Kapas is an island 6 kilometer off the coast of Terengganu, Malaysia. “Kapas” is the Malay word for cotton....
References	http://serimanjung.blogspot.com , http://www.qimichaletkapas.com
Lat	5.215555556
Long	103.271385918711

TABLE 4.6: Kapas Island entry example in Dbpedia

Concept	Petronas Twin Tower
Abstract	The Petronas Twin Tower (also known as the Petronas Tower or Twin Tower), in Kuala Lumpur, Malaysia were the world’s tallest building before being surpassed by Taipei 101.
Surpassed by	Taipey 101
Construction	1992 – 1998
Floor	88
Location	Malaysia, Kuala Lumpur

TABLE 4.7: Petronas Twin Tower entry example in Dbpedia

4.3.4 Mapping Extracted Concepts To The Knowledge Base

Each query concept extracted from the text analyzer will be mapped to the entries in the knowledge bases. The knowledge base is in the form of the Semantic Web language, RDF and therefore SPARQL can be used to access MTO and Dbpedia while the Geonames Web Service is used to access Geonames ontology. The mapping process between identified concepts in the generic information identification and concepts in the knowledge base is presented as follows:

1. to identify if the concept is related to events, attraction or location by matching the extracted concepts with the MTO ontology. If the input concept matched to a concept in MTO, its corresponding information will be saved. For example, concept *palace* has found a match in the MTO ontology. Based on MTO ontology, *palace* is classified under *castle* and it is an *Attraction*. An example for *palace* information is presented as follows:


```

<ie:Result>
<ie:KnowledgeExtracted>
<ie:name >palace</ie:name>
<ie:indicate>Attraction</ie:indicate>
<ie:frequency>1</ie:frequency>
<ie:rootWord>palace</ie:rootWord>
<ie:referTo rdf:resource="http://www...../tourMalaysia.owl/castle">
</ie:KnowledgeExtracted>
</ie:Result>

```

2. to identify if the query concept is location related information. If a match is found in the Geonames entries, the corresponding information about location will be saved. If the query concept has a match with previous task (task 1), it would try to use the alternative names to find a match in Geonames. An example for *Kuala Kangsar* is presented as follows:

```

<geo:Feature rdf:about="http://www.geonames.org/1734599/">
<geo:name>Kuala Kangsar</geo:name>
<geo:featureClass>city, village,...</geo:featureClass>
<wgs84_pos:lat>4.7666667</wgs84_pos:lat>
<wgs84_pos:long>100.9333333</wgs84_pos:long>
<geo:adminCode>07</geo:adminCode>
<geo:adminName>Perak</geo:adminName>
</geo:Feature>

```

3. to identify if the identified concept has a similar entry in the Dbpedia knowledge base. Similar to task 2, if the query found a match with task 1, the alternative names will also be used to find a match in the Dbpedia entries. Since Dbpedia consists of entries from many domains, a constraint is set to allow our tool to match the identified concept with entries related only to our domain of interest. The matching is done by identifying entries that are related to the Skos category which is associated to Malaysia. For example, *Kapas Island* is identified based on searching under *Category:Islands_of_Malaysia* while *Thaipusam Festival* is found based on searching *Category:Religious_of_Malaysia*.

Table 4.8 shows the list of concepts that were identified by the Knowledge based component. In this example, it shows that some concepts might be identified in more than one knowledge base. For example, concept *Melaka* was found in MTO and Geonames ontology while Famosa was identified in MTO and Dbpedia. Finally, the results of the general and specific information identification are extracted and stored in the RDF format.

Concept	MTO	Geonames	Dbpedia
Melaka	Tour:Melaka	Geo:1734756	-
Malacca	Tour:Melaka	Geo:1733035	-
Famosa	Tour:St_Paul_Hill	-	Dbpedia:Famosa
Fortress	Tour:Attraction	-	-
Fort	Tour:Attraction	-	-
Malaysia	-	Geo:1733045	Dbpedia:Malaysia

TABLE 4.8: Knowledge Base Component Output

4.3.5 Semantic Information Model Description

This section provides a description of some examples of the semantic information model generated by the text analysis. The output produced by analysing document ID 1460920756 is visualized in Figure 4.5. Concepts that are identified are linked to others by information identified in the knowledge bases. For example, the word *famosa* is identified in two sources which are Dbpedia and MTO. Other concepts that were captured during generic information identification are *tourist*, *tourism*, *stone*, *sky*, *sharing*, *power*, *portugis*, *protugal*, *nice*, *n21*, *lovely historical*, *history*, *colonialism*, *exploration*, *clouds*, *cannon*, *beautiful*, *architectural gems*, *ancient* and *alfonsodalbequerque*. This result shows a very good example of how knowledge bases are useful to map a semantic description of the image which constitutes a higher level of indexing than isolated terms.

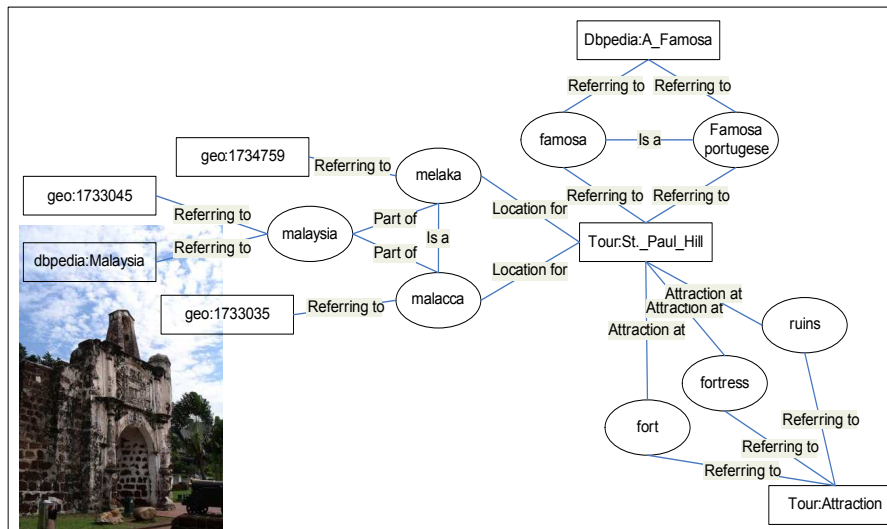


FIGURE 4.5: Visualization for information extracted for Image ID 1460920756

The image in Figure 4.6 only consists of simple descriptions as follows:

Title : Malaysia

Tags : malaysia, kualakangsar, perak, travel, tourism, perspective, asia, impressed-beauty,

Caption : This former palace of the Sultan of Perak was built without the use of a single nail. It is now a museum. It's on EXPLORE Aug. 29.

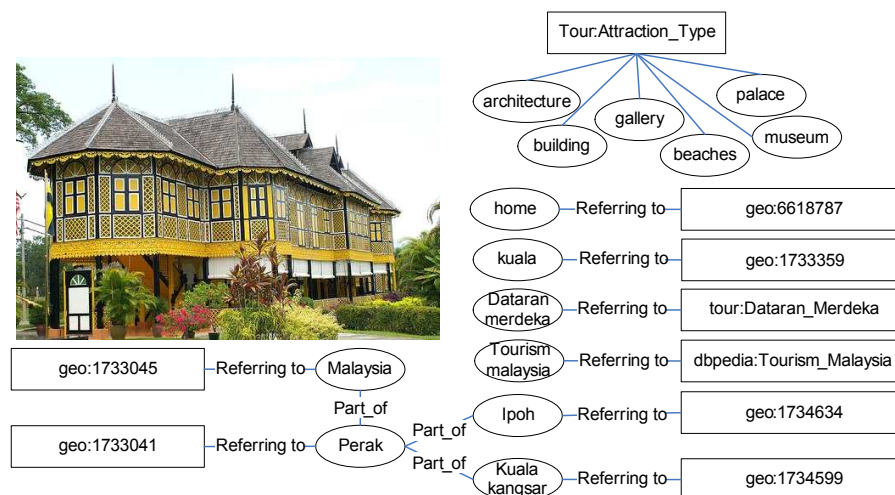


FIGURE 4.6: Visualization for information extracted for Image ID 1203148615.

Nevertheless, the image has triggered many additional items of information which have significantly increased its document length (as presented for Figure D.2 in Appendix D). Even though there is a lot of feedback about the images added by other users, the accurate specific term to describe the building in the image is *istana kenagan* or *kenagan palace* which is still not found in the document. There is some ambiguity such as the word *beaches* which is also found in the documents. In this case, the length of the document

is quite long and therefore the word *beaches* can easily be ignored by comparing word frequency with other attraction types that have been captured. Nevertheless, the length of the documents in our corpus is inconsistent from one to another due to two main factors which are lack of description provided by the author of the document, and the interest that the image has triggered which usually will be reflected by the numbers of comments or feedback left by viewers. In the cases where the length of the documents is short, the term frequency does not provide any significant information to provide a hint of the terms that can be used to represent the image. In these cases, image analysis would be very useful to develop an enriched description of the images. In the next chapter, the second stage of the analysis, namely the image analysis, is presented.

4.4 Comparing Extracted Concepts with Tags: Discussion

In this section, we provide observations on our approach in extracting and identifying words used to describe information in Web 2.0 documents. Figure 4.7 and Figure 4.10 show some examples of data table of tags and concepts that have been discovered during the analysis. The result shows a list of concepts extracted from the document. Some of the concepts are with identifiers and frequencies of the concepts are also provided.

Based on observation, most of the images are described by using discrete words and not in the form of phrases. For example, document ID 473775905 shows a woman dancing in traditional dress. The document has been tagged with 23 words such as (cultural, dance, event, women and *kingbirthdaycelebration*) to describe the image but all of these words are disjointed (discrete) from one to another. As can be seen, a phrase *king birthday celebration* has been tagged by one word *kingbirthdaycelebration*. The analysis has enabled the concepts to be discovered in the form of phrases such as *traditional malay dance*, *malay dance*, *freedom square* and *coronation celebration*, which are indicated with yellow highlights, are significant to describe the image. By using a combination of natural language processing tools APP and GATE, the phrase *freedom square* and *Kuala Lumpur* are characterized as information to indicate location. Furthermore, it also found *kuala lumpur* concept in the Geonames Ontology entry and the information is captured in an RDF representation as presented in Figure 4.8.



Image	Tags	Concepts Discovered
ID: 153554713 	shamshahrin shamsudin imagemaker wiredlens malaysia kedah langkawi island sunset beaches destination travel tourism places sky sun sea scenery outdoor nature water clouds	Location : Malaysia [2] Transportation : water [1] Environment : island [1] State : kedah [1] Location : Langkawi [1] Location : Kuah Beach [1] Temporal : Sunset [2] Scenery & Natural [1] wiredlens [1] beaches [1] destination [1] travel [1] tourism [1] sky [1] sun [1] sea [1] scenery [2] nature [1] clouds [1] beautiful sunset [1]
ID : 473775905 	kuala Lumpur malaysia cultural dance event kingsbirthdaycelebration daulattuanku dancers malay women tourism yellow dancer performance art visitmalaysia jalalspagespeoplesalbum ~vivid~ wwwow anawesomeshot shot blueribbon	Date : 26 April 2007 [1] Location : Kuala Lumpur [1] Location : Freedom Square [1] Date : Coronation [1] malay dance [1] traditional malay dance [1] king [1] coronation celebration [1] dataran merdeka [1] malaysia [1] dance [1] event [1] kingsbirthdaycelebration [1] dancers [1] women [1] tourism [1] dancer [1] performance [1] art [1] culture [1] beautiful colors [1] perfect timing [1] colors [1] country [1] vivid [1] color [1] magic [1] fantastic color [1]

FIGURE 4.7: This figure shows some examples of images, tags and concepts that were extracted by using text analysis techniques.

```

<geo:Feature rdf:about = "http://http://sws.geonames.org/1735161/">
  <geo:name>Kuala Lumpur</geo:name>
  <geo:featureClass>city, village,...</geo:featureClass>
  <wgs84_pos:lat>3.1666667</wgs84_pos:lat>
  <wgs84_pos:long>101.7</wgs84_pos:long>
  <geo:adminCode>14</geo:adminCode>
  <geo:adminName>Kuala Lumpur</geo:adminName>
</geo:Feature>
</rdf:RDF>

```

FIGURE 4.8: Example for storing Geonames entries in RDF format

The other example shows the analysis done for document ID 153554713. The given title for the document is LGK Kuah0846. The description for the image is *Sunset at Kuah Beach in Langkawi, Kedah*. The image was originally tagged with 22 words, words to identify location (*langkawi, kedah, beaches* and *island*), words related to object (*sky, sun, sea* and *water*) and temporal (*sunset*). Based on our domain of interest, we have identified some of these words and other words within the document as location (*Malaysia, langkawi* and *kuah beach*), environment (*island*) and temporal (*sunset*). We also found affective words such as *beautiful sunset* phrase in the document.

In this example, two concepts were found that match to Geonames Ontology entries: *Kedah* or *Negeri Kedah* and *Langkawi* is shown in the example as highlighted in grey colour (Figure 4.9). Since the Geonames Ontology is also a work in progress, it is not possible to find missing information in the Geonames Ontology. For this example, *Langkawi* is actually part of *Kedah* which mean that adminName and adminCode for *Langkawi* is supposedly inherited from *Kedah* entry. The missing of information in the Geonames Ontology would influence the semantic link that we are trying to capture especially in describing the location of the images such as *Kuah Beach* is part of *Langkawi* and *Langkawi* is part of *Kedah*.

```
<geo:Feature rdf:about = "http://http://sws.geonames.org/1733048/">
  <geo:name>Negeri Kedah</geo:name>
  <geo:featureClass>country, state, region,...</geo:featureClass>
  <wgs84_pos:lat>6.0</wgs84_pos:lat>
  <wgs84_pos:long>100.6666667</wgs84_pos:long>
  <geo:adminCode>02</geo:adminCode>
  <geo:adminName>Kedah</geo:adminName>
</geo:Feature>

<geo:Feature rdf:about = "http://http://sws.geonames.org/6301256/">
  <geo:name>Langkawi</geo:name>
  <geo:featureClass>spot, building, farm</geo:featureClass>
  <wgs84_pos:lat>6.329728</wgs84_pos:lat>
  <wgs84_pos:long>99.728667</wgs84_pos:long>
  <geo:adminCode></geo:adminCode>
  <geo:adminName></geo:adminName>
</geo:Feature>
```

FIGURE 4.9: Sample for incomplete Geonames entries

Document ID 1203148615 (Figure 4.10) shows a former palace that has been converted into a museum. The document has been tagged with eight tags; *kualakangsar, perak, travel, tourism, perspective, asia, impressedbeauty* and *Malaysia*. In this example, the owner of the image tends to describe more about location than the object of the image (*kualakangsar, perak, asia* and *Malaysia*). Moreover, the example also shows the same pattern in tagging terms that consist of two words, *impressedbeauty* and *kualakangsar*. The *impressedbeauty* tag is from *impressed beauty*, while the *kualakangsar* tag is from the phrase *kuala kangsar* which indicates a specific location in State of Perak. Moreover, our hybrid approach can identify more interesting concepts and phrases such as *former palace, museum, kuala kangsar* and *beautiful architecture*.

FIGURE 4.10: This figure shows some examples of images, tags and concepts that were extracted by using text analysis techniques.

Image	Tags	Concepts Discovered
ID: 1203148615 	kualakangsar perak travel tourism perspective asia impressedbeauty malaysia	Location : Malaysia [23]
		Location : Thailand [6] ? - confuision Attraction : palace [3] Location : Ipoh [1] Location : Kangsar [1] Location : Kuala [1] State : perak [4] Attraction : museum [2] Person : Sultan [2] Kuala kangsar former palace [1] travel [1] tourism [1] asia [1] scenery [1] amazing looking building [1] window shapes [1] beautiful architecture [1] nicely [1] wonderful place [2] beautiful colors and place [1] congratulations [1] weather [1] beaches [1] ambient [1] nails [5] wonderful colors and nice [1] beautiful building [1] pangkor laut [1] architecture [1] house [1] stream [2] wonderful architecture [1] beautiful lovely yellow [1] delicate pattern [1] nail [1] wonderful registry [1] color and architecture [1] beautiful home [1]

Based on tourism vocabularies provided by the World Tourism Organization, words such as *palace* and *museum* are characterized as attraction. A few ambiguous entries are found in this example. Firstly, the word *Thailand* is captured in the concept extracted alongside the word *Malaysia* and both are identified as location. In this case, the location of the image is chosen by comparing their frequencies. Secondly, the ambiguity continues in the result provided by the Geonames Ontology which shows three different adminNames which indicate different states which are *State of Perak* (*Ipoh* and *Kuala Kangsar*), *Selangor* (*Home*) and *Sabah* (*Kuala*). This ambiguity can be cleared by also

comparing the entry with its frequencies.

4.4.1 Conclusion

In this section, we have presented a natural language approach and the use of Semantic Web technologies to tackle text based information produced by user contributions in Web 2.0 documents. Based on our observation, the identified concepts are useful to represent the content of the corresponding images. Ontologies and knowledge bases have enriched the concepts with additional information related to the domain of interest. Finally, the information is stored in the form of RDF to represent such information semantically.

Chapter 5

Image Classification By Using Image Content Analysis

This chapter discusses the second stage of the hybrid approach, that is, the image analysis stage. Image analysis and in particular automatic image classification or annotation, often begins with feature extraction for content representation. The content representation can be based on low level and high level feature extraction. In this research, low level features such as colours and edges constitute the extracted information that can be derived directly from the image itself without requirements to refer to any additional semantic information.

The image analysis is required to characterise or identify objects or scenes depicted by the image. In our approach, the image analysis stage involves the extraction of two features which are edges (or lines) and colour information. The aim of the image analysis is to classify images based on their low level features by using Bayesian Inference. Recent work on automatic image classification and annotation was reviewed in Chapter 2.

In this chapter, we develop and demonstrate a simple classifier which attempts to classify images which contain buildings and those which do not. Equivalently it can be regarded as an annotator which aims to annotate images with the tag “buildings” when appropriate. Buildings are example of an image class that can be identified using this method. The method could be extended into other classes such as mountains, beaches and forest. This image classification is important because it can enable users to retrieve images that may not be well tagged and also to annotate images with information that they may want to use for retrieval purposes but not necessarily for explicit annotation. The performances are assessed using confusion matrices and ROC curves.

5.1 Image Classifier Design

An artificial versus natural feature extractor developed in Photocopain (reference) uses classification of the edge direction coherence vector to classify image content, based on the assumption that artificial structures tend to be created with straight-edges, whereas natural structures do not [Tuffield et al. \(2006\)](#). In our development of a classifier we combine the edge/line features used in Photocopain with colour features to classify building and non building images. The aims of the analysis are:

1. To develop a classifier which distinguishes between images which predominantly contain buildings and those which do not.
2. To find optimum values for parameters in the classifier using a training data set.
3. To classify images from a test set using the values identified.
4. To evaluate the performance of the image classifier.

The image classifier construction is illustrated in Figure 5.1. Images are divided into two sets which are a training set and a test set. The training set consists of 210 images (105 building and 105 non building images). These images were obtained from Flickr and carefully allocated to the appropriate set. Table 5.1 shows some examples for building and non building images used in the training set. In general, the building image set (Table 5.1(a)) has a specific focus mainly on structure such as buildings and houses, while the non building image set (Table 5.1 (b)) has a wider focus covering such topics as people, flower, boat and etc. The test set consists of 1040 images (534 building images and 506 non building images).

To initiate the analysis, known building and non building images (in the training set) are submitted to low level feature extractors for colour and line features. The City Landscape Identifier (CLI) from Photocopain ([Tuffield et al. \(2006\)](#)) is used in the line analysis to generate line histograms. Detailed descriptions for each analysis are presented in Section 5.2 and Section 5.3 respectively. Both of these analyses produce results in the form of histograms. These histograms are normalized and used as the input to the inference Inference Engine. The Inference Engine is implemented using Infer.net to provide a Bayesian Inference tool. The Inference Engine uses the training data in order to generate prediction values in the range 0 to 1.0 for each image in the set. A high value indicates a building image and a low value a non building image. A threshold value on the prediction must be chosen, above which images are tagged as building images. The query image / unknown image (presented as image X in Figure 5.1) is submitted and follows a similar root to images in the training set. Line and colour histograms are extracted and passed to the Inference Engine and a probability value for image X is generated and the threshold applied.

Section 5.4 describes experiments carried out in obtaining the threshold values, followed by an analysis of the building and non building image classification in Section 5.5. Even though the image classifier is only focusing on identifying building and non building images, it could be expanded to include other classifiers to detect the presence or absence of beaches/ocean, mountains, sunset and forest etc. To create classifiers for generating a wide range of tags it may be necessary to use more powerful low level features such as the visual terms used by some researchers[Hare and Lewis (2010) and Sivic and Zisserman (2003)].

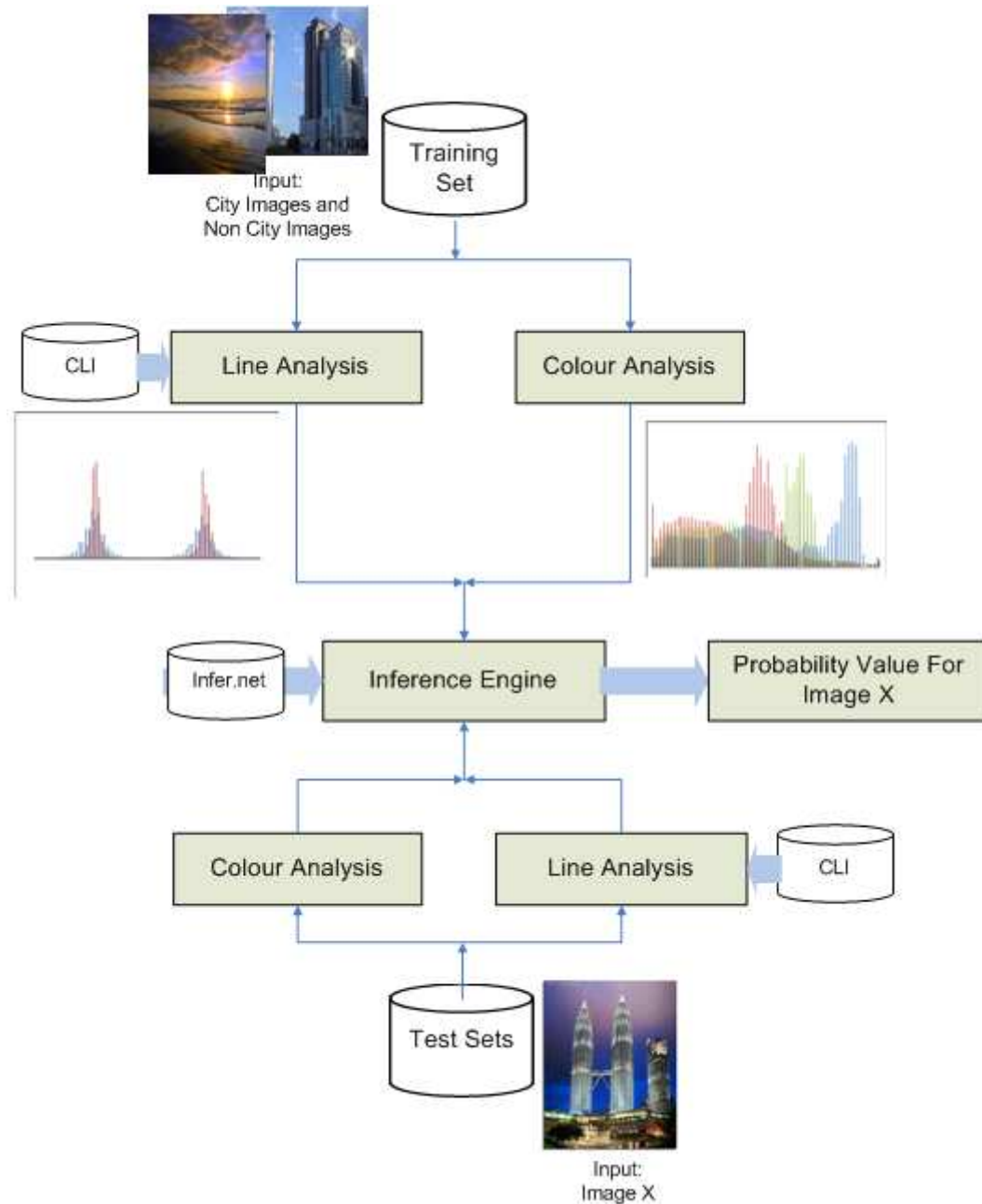


FIGURE 5.1: The construction of City Landscape and Non City Landscape Image Classifier



TABLE 5.1: Sample of building and non building images used in the training set

5.2 The Line Analysis

The extraction of line information is done by using an algorithm from the Photocopain system. Photocopain is a content based annotation tool which is integrated with the AKTiveMedia image annotation system to allow users to annotate images semi-automatically (Tuffield et al. (2006)). Photocopain is used to identify scenes, such as city views, objects (such as monument, building), providing clues about the main interest of the images. City Landscape Identifier (CLI) is one of the Photocopain components. The CLI is used to represent image content by calculating the edge direction coherence vector for artificial and natural features. The assumptions are that straight edges tend to appear more in artificial structures such as in buildings and bridges, rather than natural structures such as mountains and clouds. The CLI used the Canny edge-detector to extract edge information from the image. Short edges are identified and represented as an incoherent edge vector histogram. Next, an edge tracing algorithm is used to find those edges that are long (coherent) and these values are represented in a coherent edge vector histogram. The two histograms make up the edge direction coherence vector. In Photocopain artificial and natural scenes are assessed by observing these two histograms.

The Line Analysis works as follows. An image is submitted to the CLI analyzer and the analyzer identifies edges at 72 directions. The edges are converted into lines and the number of pixels in a straight line section represents its length. Lines shorter than a particular threshold are regarded as incoherent and longer lines are regarded as coherent. Histograms representing the incoherent lines and the coherent lines are created. The histograms have 72 bins indicating the 72 edge directions and the numbers of pixels in lines at each of those directions are stored in the bins for the incoherent and coherent lines separately. The data in the histograms are normalized by summing up the total number of pixels in each direction and dividing each value by the total value. The normalized data gives the proportion of pixels as a percentage for each direction. Figure 5.2 c shows a sample output generated by the CLI analyzer for *Image ID 18396625*. In Figure 5.2(c), the analysed image has generated high calculation of pixel counts in 0, 90, 185 and 360 degree. In general, images of buildings will produce a high pixel counts in such a range of angles. Based on the observation, we have decided to analyse images based on certain directions and compared them with the results for all directions. Table 5.2 shows an example of normalized line histograms for building and non building images.

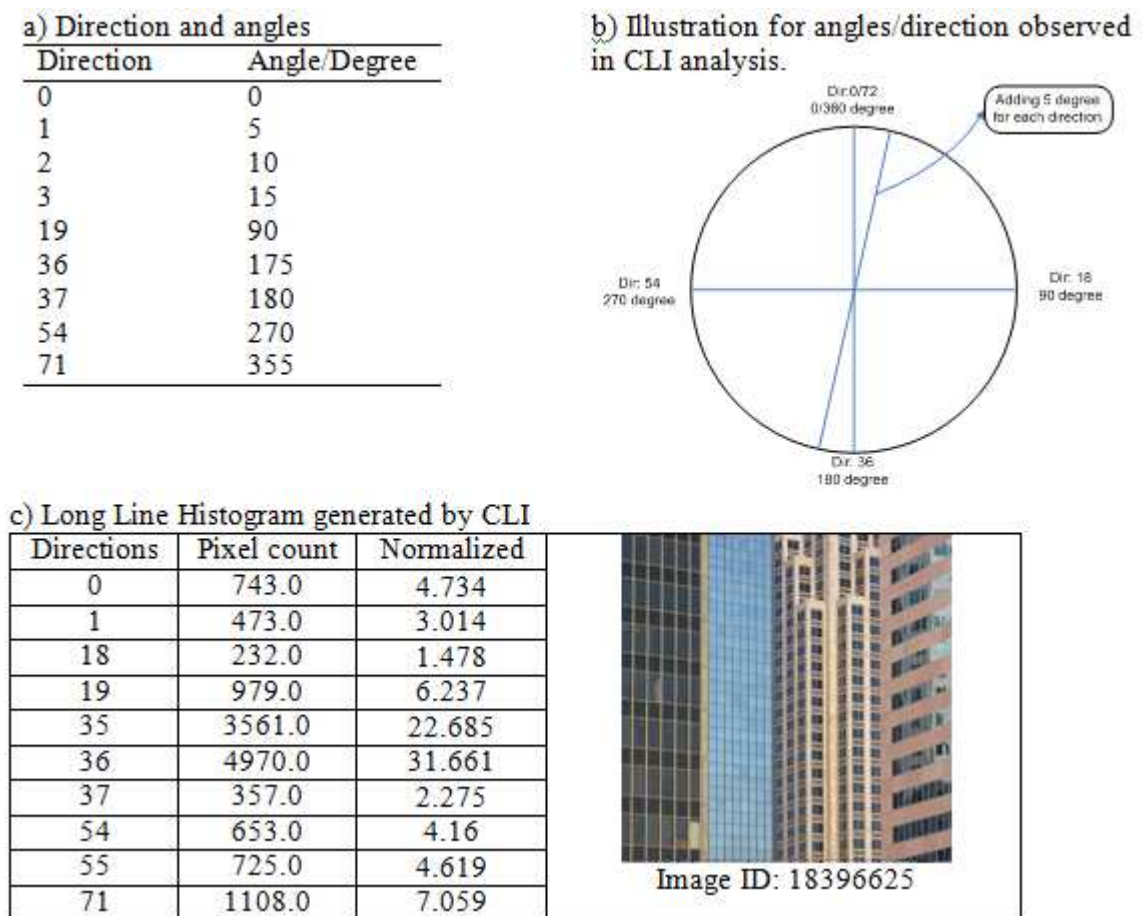


FIGURE 5.2: This Figure shows (a) example of directions and angles used to observe edges, (b) illustrates orientation angles and (c) shows list pixels count for each directions and its normalized value.


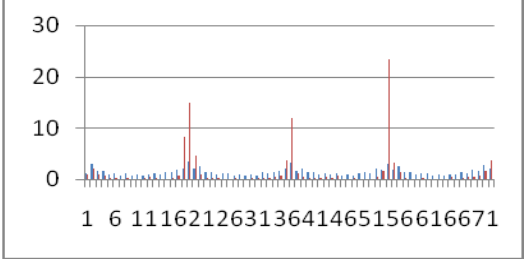

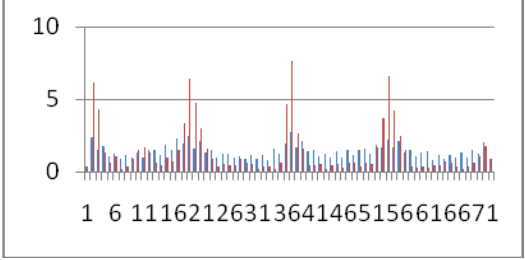

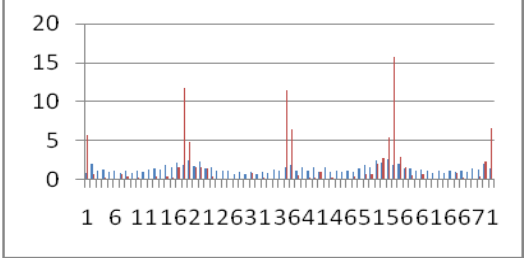

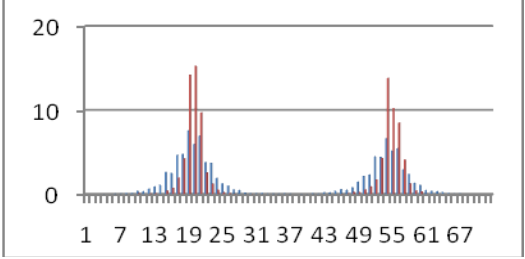

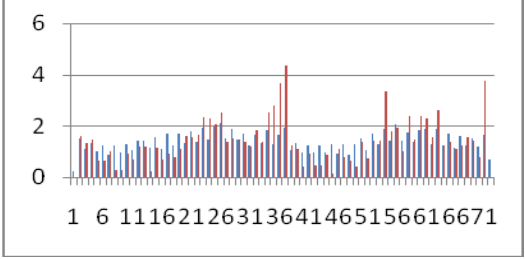
Input Images	Line Analysis Output
 <p>(a) ID: 173113612</p>	
 <p>(b) ID: 175003809</p>	
 <p>(c) ID: 213076494</p>	
 <p>(d) ID: 27976290</p>	
 <p>(e) ID: im164</p>	

TABLE 5.2: Examples of city and non city images with their normalized line histograms. The red bar in the line histograms represents long line data while the blue bar represents short line data.

5.3 The Colour Analysis

Colour analysis here involves extracting a colour histogram for each image to be used as the colour feature in the machine learning process. In the first instance we use the RGB colour model although it would be interesting to experiment with alternative colour models. For each image, the colour histogram is generated to provide information about the distribution of colours in the image. There are two different methods for generating the colour histogram. One is to divide the 3-D RGB colour space into cells and count the number of pixels falling in each cell. The other is to calculate a histogram for each of the three (R,G and B) colour channels separately.

- **3D Colour Histogram** In the first method, each pixel in the images is projected into a 3D RGB colour space as illustrated in Figure 5.3 . In this work the 3D colour space is divided into 4x4x4 and 6x6x6 cells which generates colour histogram with 64 bins and 216 bins respectively. The number of pixels in each cell are counted and are stored in the colour histogram. The second column in Table 5.3 shows some examples for colour histograms generated by extracting colour feature based on 3D colour space. The histogram is normalized by summing up the total number of pixels in each bin and dividing each value by the total value. The normalized data gives the proportion of pixels as a percentage for each bin.
- **Separate Colour Histogram** In the second method, the intensity histogram of each colour, red, green and blue, is observed separately. In an RGB image each pixel is represented by three colours, which are red green and blue, and the intensity level for each colour is observed and extracted to generate three colour histograms. Each colour histogram is divided into 64 bins producing 3 colour histograms for each image. The third column in Table 5.3 shows some examples of colour histograms generated by extracting colour features based on the separate colours. The colour histogram is also normalized by using the same normalisation as described earlier.

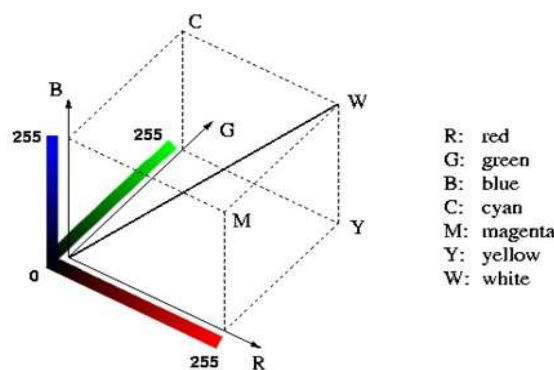


FIGURE 5.3: Illustration of RGB colour cube

5.4 Finding the Optimal Threshold Value To Identify Building Images

In this section, we present our approach to finding the optimal threshold value to identify building and non building images. Three experiments are involved;

1. Experiment 1: Finding threshold values using line histograms
2. Experiment 2: Finding threshold values using colour histograms
3. Experiment 3: Finding threshold values using line and colour histograms

These experiments are done using just the training set of images as this is all part of the classifier training process. All of the images in the training set are already labelled with either building or non building tags, thus making it possible to evaluate classification performance with different thresholds. The classification performances are evaluated by using confusion matrices and ROC curves to identify optimum threshold values to classify building and non building images. Detailed descriptions for the processes involved in the experiments are presented for Experiment 1, and the processes are repeated for Experiment 2 and Experiment 3.

5.4.1 Experiment 1: Optimal Threshold Value Selection Using Line Histograms

The line histograms generated in the line analysis consist of two types of data which are the short line and the long line histograms. In this experiment, we divided the analysis into three sub experiments:

1. Line 1 : observing long lines only with all directions. In this analysis, all 72 data values in the long line histogram are observed. This analysis is referred as **Line 1: Long Line [72 Dirs]**.
2. Line 2 : observing long lines with merged directions. In this analysis, only 24 directions are observed. The 72 bins in the long line histogram are merged into 24 bins by summing 3 bins for each direction. This analysis is referred to as **Line 2: Long Line [24 Dirs]**.
3. Line 3 : observing short and long lines with merged directions. In this analysis, 24 bins are used in the short line histogram and 24 bins in the long line histogram. This analysis is referred to as **Line 3: Short and Long [48 Dirs]**.

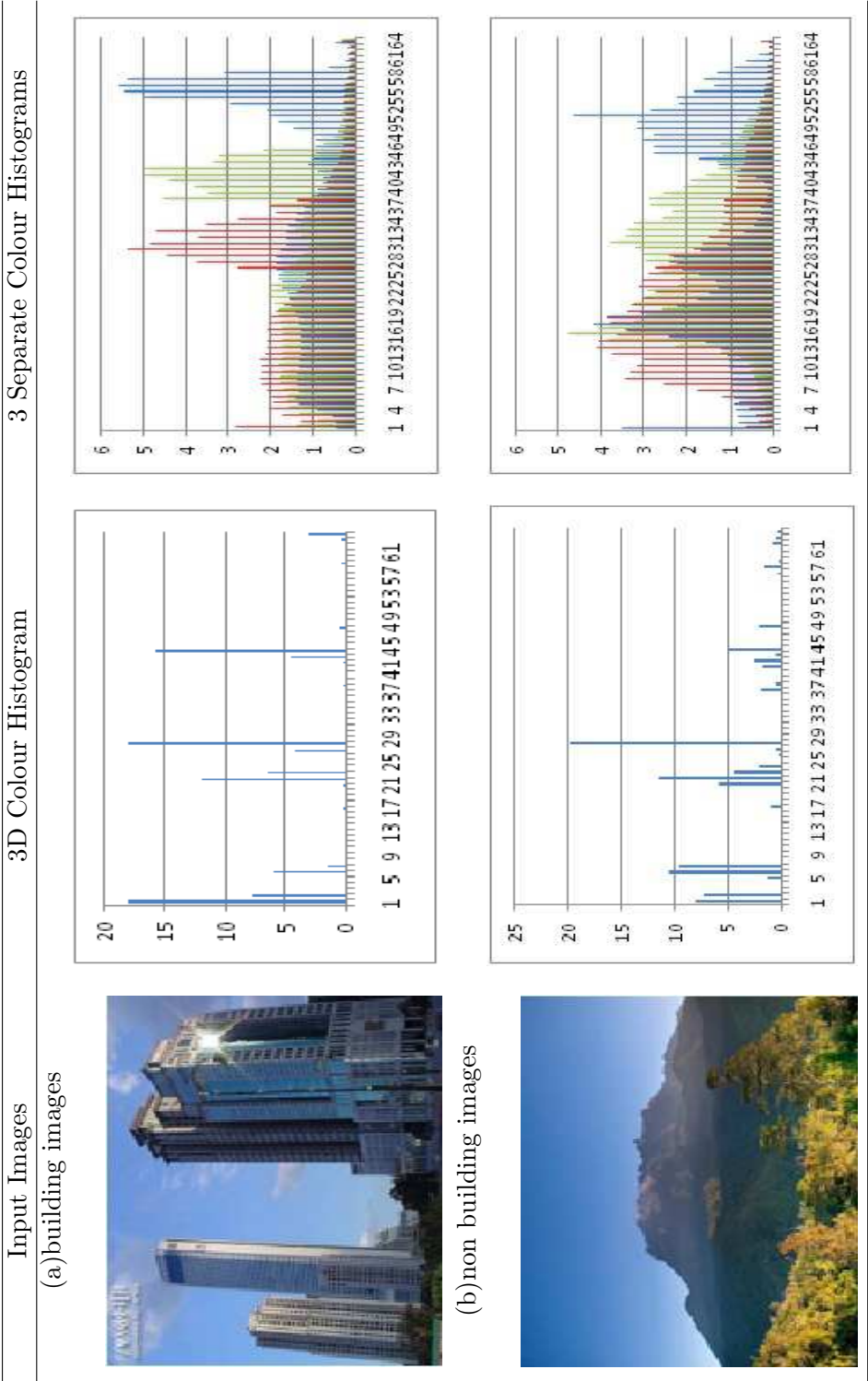
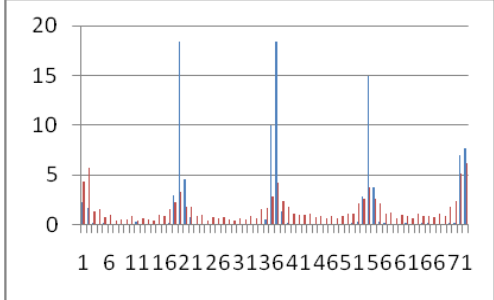


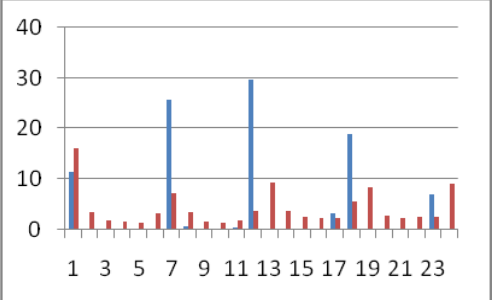
TABLE 5.3: Examples of building and non building images and their colour histograms generated by extracting colour features from the images. The second column shows the 3-D colour histograms and the third column shows the separate R,G and B histograms superimposed in different colours. For the 3D colour histogram, the x-axis represents the bin number while the fraction of pixels for each bin is indicated in the y-axis. For the 3 separate colour histograms, the x-axis bin numbers indicate the colour intensity while the y-axis indicates the fraction of pixels with that intensity for that each colour.

Using data from all directions means all 72 data/value are used. Each data item in the line histogram indicates the percentage of pixels with edge directions in each 5 degree range. *Merged direction* refers to merging the data to generate a wider orientation direction. Table 5.4 shows a sample of data in the 72 bin form a (Table 5.4(a)) and the merged form (Table 5.4(b)).

Full/Unmerged Data Observation			Merged Data Observation		
Directions	Short Line	Long Lines	Directions	Short Lines	Long Lines
1	4.30	2.21	1 [72,1,2]	16.13	11.50
2	5.64	1.69	2 [3,4,5]	3.41	0.33
3	1.26	0.13	3 [6,7,8]	1.83	0.05
4	1.46	0.00	4 [9,10,11]	1.54	0.32
...
71	5.04	6.92	23 [66,67,68]	2.57	0.05
72	6.18	7.60	24 [69,70,71]	9.04	7.17



(a) Unmerged Data



(b) Merged Data

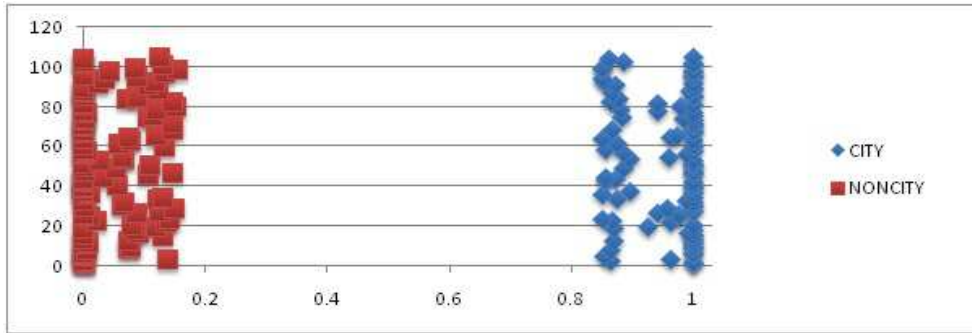
TABLE 5.4: Unmerged and merged line histograms

Table 5.5 shows examples of the building and non building image prediction generated by the Inference Engine. The prediction values range from 0 to 1. If the image prediction value is close to 1, it shows high probability that the image is a building image. If the image prediction value is close to 0, it shows high probability that the image is a non building image.

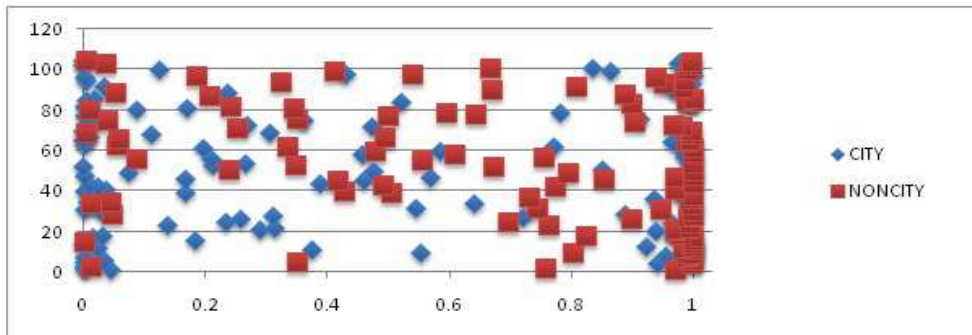
Building Image Id	Prediction	Non Building Image Id	Prediction
1	1.0000	1	0.0006247
2	1.0000	2	1.04E-13
3	0.8629	3	1.96E-07
4	0.9607	4	0.1375
5	0.8539	5	1.30E-16
6	1.0000	6	1.43E-06
7	1.0000	7	0.0003746
8	0.8625	8	0.005515
9	1.0000	9	0.07277
10	1.0000	10	0.07526
11	0.9999	11	1.90E-10
12	1.0000	12	0.005949
...
105	1.0000	105	4.81E-13

TABLE 5.5: Prediction Values for building and Non building Images. (Prediction value is Probability of images being a building images). Value generated by the Inference Engine for Line 1: Long Lines [72 dirs].

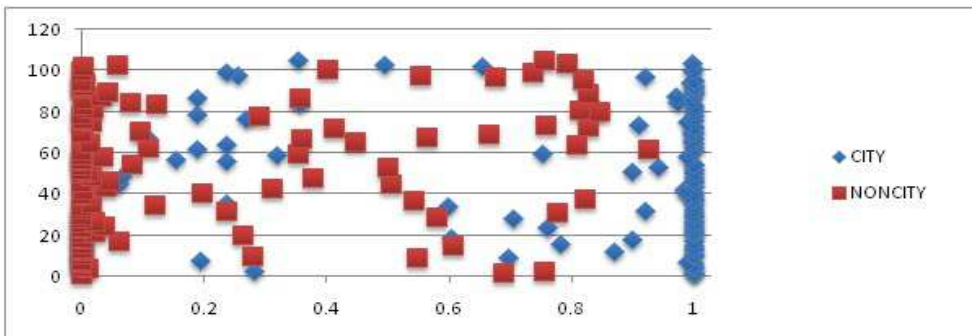
Full results for line histogram analyses are illustrated in the form of image distributions in Figure 5.4. Building images are represented in blue boxes while non building images are represented in red boxes. Figure 5.4(a) shows an excellent classification. There is a clear separation between building and non building images. building images are well classified with cut-off point 0.8 and above, while non building images are classified using cut-off point 0.2 and under. Figure 5.4(c) shows a good classification. Most of the building images fall close to 1, while there are a few images distributed in the range of 0 to 0.8, and vice versa. Most of non building images are distributed close to 0, while a few images are distributed in range of 0.2 to 1. Figure 5.4(b) shows a bad classification. Most of building images are distributed evenly at each end. For non building images, most of them are distributed close to 1, indicating these images have low probability to be identified as non building images, but high probability to be identified as building images.



(a) Experiment Line 1 by observing long Lines with 72 Directions



(b) Experiment Line 2 by observing long lines with 24 directions each.



(c) Experiment Line 3 by observing short and long lines only with 48 directions.

FIGURE 5.4: building Image and non building image distributions using line histograms. Each of the boxes represents an image in the training set. The blue boxes refer to building images and the red boxes refer to non building image. The x-axis shows a cut-off point / threshold value candidate, and the Y-axis show number of building and non-building images in the training set.

The classification performances for each experiment are analysed to identify the optimum threshold value to identify building and non building images by using Confusion Matrix and ROC Curve data analysis. The Confusion Matrix¹ is used to display actual and predicted classification by the image classifier. The entries in the Confusion Matrix have the following meaning in the context of our study:

Actual\Predicted	Negative (Non building)	Positive (building)
Negative (Non building)	a	c
Positive (building)	b	d

TABLE 5.6: Data Matrix

The entries in the confusion matrix have the following meaning:

- a is the number of correct predictions that an instance is negative
- b is the number of incorrect predictions that an instance is negative
- c is the number of incorrect predictions that an instance is positive
- d is the number of correct predictions that an instance is positive

or in our case:

- a is the number of correct non building images identified as non building
- b is the number of incorrect building identified as non building
- c is the number of incorrect non building images identified as building
- d is the number of correct building images identified as building

¹ Kohavi and Provost, 1998

Six standards indicators that can be generated from the Confusion Matrix are as follows:

- The **True Positive Fraction** or **sensitivity** (TP) is the proportion of positive cases that are correctly identified, as calculated using the equation:

$$\text{True positive fraction (Sensitivity)} = d/(b+d)$$

- The **False Positive Fraction** (FP) is the proportion of negative cases that are incorrectly classified as positive.

$$\text{False positive fraction} = c/(a+c)$$

- The **True Negative Fraction** or **specificity** (TN) is defined as the proportion of negatives cases that are classified correctly, as calculated using the equation:

$$\text{True negative fraction (specificity)} = a/(a+c)$$

- The **False Negative Fraction** (FN) is the proportion of positive cases that are incorrectly classified as negative, as calculated using the equation:

$$\text{False negative fraction} = b/(b+d)$$

- The **Accuracy** (AC) is the proportion of the total number of predictions that are correct.

$$\text{Accuracy} = (a+b)/(a+b+c+d)$$

- The **Precision** (P) is the proportion of the predicted positive cases that are correct, as calculated using the equation:

$$\text{Precision} = d/(c+d)$$

Two indicators are used to evaluate the image classification performance, the True Positive Fraction or *sensitivity* and the True Negative fraction (TNF) or *specificity*. TPF is a fraction of building images that are correctly identified as building image and TNF is a fraction of non building images that are correctly identified as non building image. Table 5.7 shows results of *sensitivity* and *1-specificity* for **Experiment 1**. Figure 5.5 shows ROC curves produced from Table 5.7. The ROC curves are generated by plotting *1-specificity* on X-axis and *Sensitivity* on Y-axis. The ROC curve allows visual representation analysis of the trade off between sensitivity and specificity.

Figure 5.5 shows the classifier has achieve perfect classification for experiment **Line 1: Long Lines [72 Dirs]** and a good classification in experiment **Line 3: Long and Short Lines [48 Dirs]**. Nevertheless, for **Line 2: Long Lines [24 Dirs]**, the negative curve shows most of the building images and non building images are distributed at the wrong end of the classification groups. Based on the calculations and visualization given by the Confusion Matrix and the ROC Curves, the optimum threshold values for classifying building images and non building images are identified. The Optimum threshold value to identify building images is at cut-off point 0.8, while 0.2 for non building images. Therefore, images with prediction value of 0.8 or higher would have a greater possibility to be identified as building images and images with prediction value 0.2 or lower would have a greater possibility to be identified as non building images.

Selecting the optimum threshold value is a trade off task between Sensitivity and specificity. The threshold values are selected based on results from Experiment Line 1 (Table 5.7). The table shows that all the building images are identified at cut-off point 0.8, and all non building images are correctly identified at cut-off point 0.2. Similar threshold values can be agreed by observing results in **Experiment Line 3**. 74% of building images are correctly identified and 8% of non building images were wrongly classified as building images at cut-off point 0.8. For non building images, 65% of non building images were correctly identified and only 8% of building images were incorrectly identified as building images.

The threshold values can be adjusted to increase sensitivity rate at the cost of a decrease in the specificity. Increasing the building threshold to a higher cut-off point would increase the credibility of the classification to identify building images by lowering the percentage of non building images to be identified as building images (increasing the specificity), but at the same time, it could miss some of the building images (decreasing the sensitivity).

Threshold/ cut off point	Experiment Line 1 Long Lines [72 Dirs]	Experiment Line 2 Long Lines [24 Dirs]	Experiment Line 3 Long and Short Lines [48 Dirs]	
	Sens.	1-spec.	Sens.	1- Spec.
0.95	0.66	0.00	0.19	0.44
0.90	0.70	0.00	0.24	0.47
0.85	0.95	0.00	0.26	0.51
0.80	1.00	0.00	0.28	0.53
0.75	1.00	0.00	0.30	0.58
0.70	1.00	0.00	0.31	0.60
0.65	1.00	0.00	0.31	0.64
0.60	1.00	0.00	0.31	0.66
0.55	1.00	0.00	0.34	0.68
0.50	1.00	0.00	0.36	0.70
0.45	1.00	0.00	0.40	0.73
0.40	1.00	0.00	0.45	0.76
0.35	1.00	0.00	0.44	0.78
0.30	1.00	0.00	0.47	0.82
0.25	1.00	0.00	0.51	0.83
0.20	1.00	0.00	0.54	0.86
0.15	1.00	0.02	0.59	0.87
0.10	1.00	0.27	0.62	0.87
0.05	1.00	0.46	0.65	0.91

TABLE 5.7: Line Histogram Analyses Result for Sensitivity and 1-Specificity calculated based on Confusion Matrix. The Line Histogram Analyses consist of three sub experiments which are Line 1, Line 2 and Line 3.

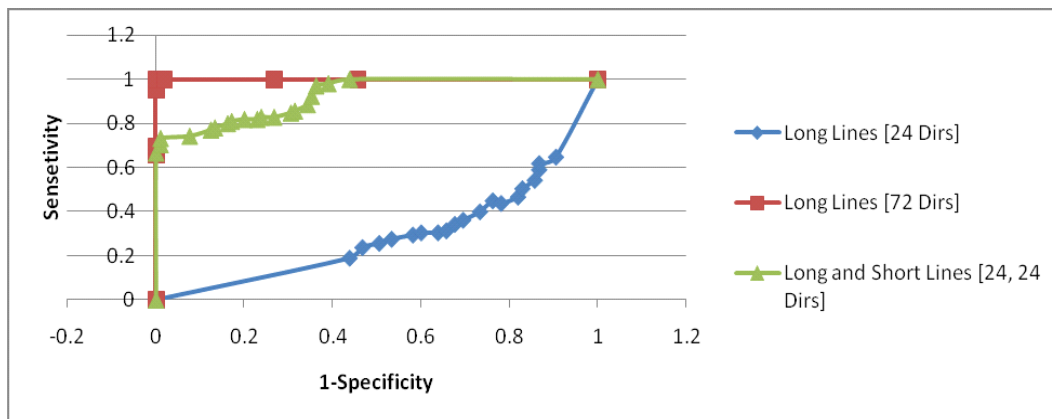


FIGURE 5.5: The ROC Curves generated based on Sensitivity and 1- Specificity for Line Histogram Analyses from Table 5.7.

5.4.2 Experiment 2: Optimal Threshold Selection by Observing Colour Histograms

In Experiment 2, colour histograms are observed. The experiment is divided into three sub experiments:

1. **Colour 1** : observing 3D colour histograms by using 64 bins (RGB Colour Space is divided into 4x4x4 cells)
2. **Colour 2** : observing 3D colour histograms by using 216 bins (RGB Colour Space is divided into 6x6x6 cells)
3. **Colour 3** : observing 3 colour histograms (Red, Green and Blue) by using 64 bins for each colour separately ($3 \times 64 = 192$ bins)

The experimental processes are similar to Experiment 1. The classification performance for different thresholds for each of the sub experiments is presented in Table 5.8 and the ROC curves for each test performance are plotted in Figure 5.6. Perfect classification was achieved by **Colour 2** and **Colour 3**. Compared to **Colour 3**, the best classification is achieved by **Colour 2** as all building images were correctly identified using cut-off point = 0.85 and non building images using cut-off point = 0.15. **Colour 1** has produced a negative curve as most of the building images have prediction values lower than 0.5 and non building images are higher than 0.5. The optimum threshold values selected for this experiment were at 0.8 for building and 0.2 for non building images. The values are selected based on the **Colour 3** results. Using **Colour 2** results would increase the threshold value for building images and decrease the threshold value for non building images. Nevertheless, it might also lower the chance of images being identified as building images and non building images.

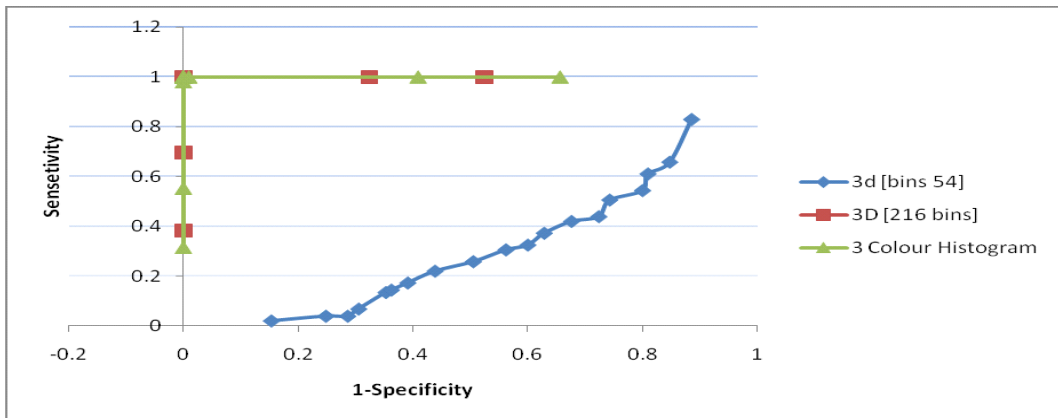


FIGURE 5.6: ROC Curves for Colour Histogram Analysis. The 3D Colour Histogram [216 bins] and 3 Colour Histograms [168 bins] have produced a perfect classification while 3D Colour Histogram with 64 bins has produced a negative curves.

Threshold	Colour 1 3D His- tograms [64 bins]	Colour 2 3D His- togram [216 bins]	Colour 3 3 Colour Histograms
	Sensitivity 1- Specificity	Sensitivity 1- Specificity	Sensitivity 1- Specificity
0.95	0.02	0.15	0.38
0.90	0.04	0.25	0.70
0.85	0.04	0.29	1.00
0.80	0.07	0.30	1.00
0.75	0.13	0.35	1.00
0.70	0.14	0.36	1.00
0.65	0.17	0.39	1.00
0.60	0.22	0.44	1.00
0.55	0.26	0.50	1.00
0.50	0.30	0.56	1.00
0.45	0.32	0.60	1.00
0.40	0.37	0.63	1.00
0.35	0.42	0.68	1.00
0.30	0.44	0.72	1.00
0.25	0.50	0.74	1.00
0.20	0.54	0.80	1.00
0.15	0.61	0.81	1.00
0.10	0.66	0.85	1.00
0.05	0.83	0.89	1.00

TABLE 5.8: Result for Finding Optimal Threshold value by observing colour histograms.

5.4.3 Experiment 3: Optimal Threshold Selection by Observing Line and Colour Histograms

In Experiment 3, both line histograms and colour histograms are used in the classification. The experiment is divided into six sub experiments:

- Training 1: 3D Colour Histogram [64 Bins] with Short and Long Line Histogram
- Training 2: 3D Colour Histogram [64 Bins] with Long Line Histogram
- Training 3: 3D Colour Histogram [216 Bins] with Short and Long Line Histogram
- Training 4: 3D Colour Histogram [216 Bins] with Long Line Histogram
- Training 5: 3 Colour Histograms with Short and Long Line Histogram
- Training 6: 3 Colour Histograms with Long Line Histogram

The experimental processes are similar to Experiment 1. The classification performance for each test is presented in Table 5.10 and plotted in the form of ROC Curves in Figure 5.7. All of the trainings runs have produced perfect classifications. Earlier in the colour analysis experiment, the colour histogram with a small number of bins has produced a negative result. Training 1 and 2 shows integrating the colour histogram (small number of bins) with the line histograms could improve the results. The optimum threshold values selected for each test are presented in Table 5.9.

Trainings	Threshold for building	Threshold for Non building
Training 1	= 0.85	= 0.15
Training 2	= 0.85	= 0.20
Training 3	= 0.85	= 0.15
Training 4	= 0.85	= 0.15
Training 5	= 0.80	= 0.20
Training 6	= 0.85	= 0.20

TABLE 5.9: Thresholds values identified for building and non building selected based on trainings results.

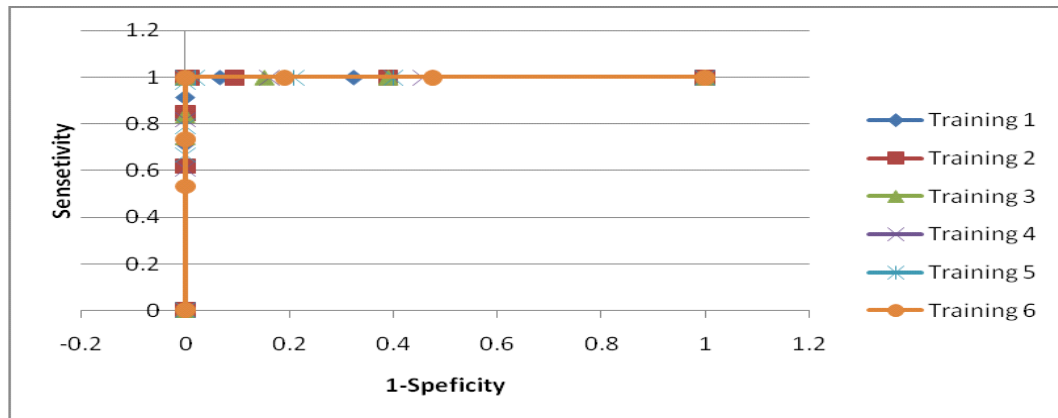


FIGURE 5.7: Curves for line and colour histograms integrations. The training sets have produced a perfect classification.

5.5 Classifying building Images and Non building Images: Results and Discussions

This section presents image classification results for line, colour and integrating line and colour analysis. The test set consists of Test Set 1 and Test Set 2. Each of the test sets consists of 534 building images and 506 non building images respectively. The classifications were done based on building and non building image thresholds identified in Section 5.4.

5.5.1 Classification Results for Line Analysis

The image classifier results for line analysis are presented in Table 5.11. The blue mark on the table shows the threshold point to identify building images while the red mark shows the threshold point to identify non building images. The ROC curves in Figure 5.8 are generated from the Table 5.11. **Line 1: Long Line [72 Dirs]** has produced a slightly better curve than observing **Line 3; short and Long Lines [48 Dirs]**. Table 5.12 and Table 5.13 show quantitative analysis for **Line 1: Long Lines [72 Dirs]** and **Line 3: Short and Long Lines [48 Dirs]** respectively. The analyses are done based on threshold values identified in the Section 5.4. Labels in the tables are described as follows:

- Total = total number of building images and non building images identified in Test Set 1 and Test Set 2.
- Total (%) = Percentage of building and non building images identified in test set.
- Correct (%) = Percentage of building or non building images correctly identified.
- Incorrect (%) = Percentage of building and non building images incorrectly identified.

Line 2 is ignored because it has produced a negative ROC Curve result during the training phase. The training results were presented in Section 5.4.1. In **Line 1**(Table 5.12), 485 images are identified as building images, 437 images as non building image, leaving 11.35% or 118 images falling in the unknown category. In **Line 3**(Table 5.13), the classifier was able to identify more building images which are 495 images, adding 58 more images from **Line 1**. Nevertheless, only 332 images are classified as non building images. Although, **Line 3** has found fewer images as non building images, its accuracy rate is higher than **Line 1**. For line analyses, we conclude that image classification has performed better by observing long lines for identifying building images and observing short and long lines for identifying non building images.

Threshold	TRAINING 1		TRAINING 2		TRAINING 3		TRAINING 4		TRAINING 5		TRAINING 6	
	Sens	1-Speci	Sens	1-Speci	Sens	1-Speci	Sens	1-Speci	Sens	1-Speci	Sens	1-Speci
1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.95	0.71	0.00	0.62	0.00	0.74	0.00	0.61	0.00	0.67	0.00	0.53	0.00
0.90	0.91	0.00	0.85	0.00	0.82	0.00	0.82	0.00	0.79	0.00	0.73	0.00
0.85	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	0.98	0.00	1.00	0.00
0.80	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
0.75	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
0.70	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
0.65	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
0.60	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
0.55	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
0.50	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
0.45	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
0.40	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
0.35	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
0.30	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
0.25	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
0.20	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
0.15	1.00	0.00	1.00	0.01	1.00	0.00	1.00	0.00	1.00	0.02	1.00	0.00
0.10	1.00	0.07	1.00	0.01	1.00	0.15	1.00	0.16	1.00	0.21	1.00	0.19
0.05	1.00	0.32	1.00	0.39	1.00	0.39	1.00	0.46	1.00	0.40	1.00	0.48
0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

TABLE 5.10: Training sets results for colour and line histogram integration. Note: Sens represent Sensitivity while speci denotes Specificity.

Threshold	Long Lines [72 Dirs]		Long and Short Lines Histograms [24 + 24 Dirs]	
	Sensitivity	1-Specificity	Sensitivity	1-Specificity
1.00	0.00	0.00	0.00	0.00
0.95	0.60	0.24	0.57	0.16
0.90	0.61	0.26	0.62	0.19
0.85	0.63	0.27	0.64	0.22
0.80	0.64	0.28	0.68	0.26
0.75	0.66	0.29	0.71	0.28
0.70	0.67	0.30	0.71	0.31
0.65	0.68	0.31	0.73	0.33
0.60	0.70	0.32	0.75	0.35
0.55	0.70	0.33	0.78	0.36
0.50	0.71	0.33	0.78	0.37
0.45	0.71	0.34	0.79	0.39
0.40	0.72	0.34	0.80	0.41
0.35	0.72	0.36	0.81	0.43
0.30	0.73	0.37	0.82	0.44
0.25	0.74	0.39	0.84	0.48
0.20	0.75	0.40	0.85	0.50
0.15	0.76	0.41	0.87	0.55
0.10	0.78	0.44	0.89	0.59
0.05	0.79	0.48	0.91	0.63
0.00	1.00	1.00	1.00	1.00

TABLE 5.11: Tests Results for Line Analysis

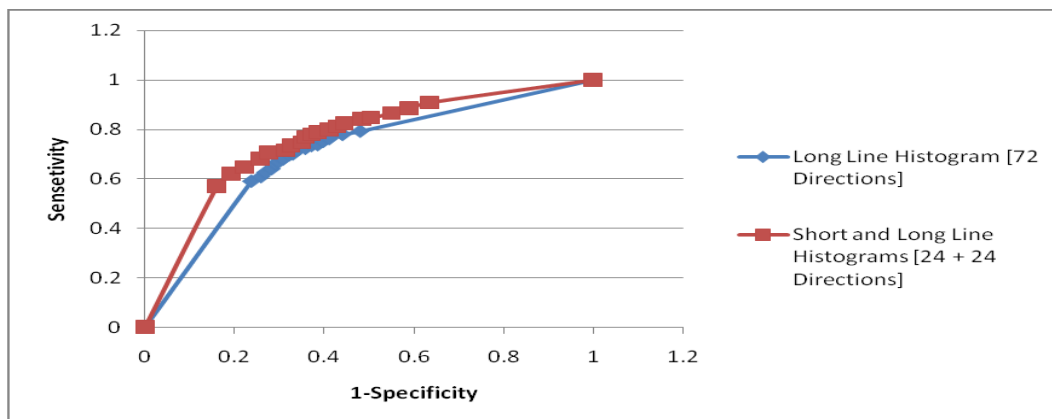


FIGURE 5.8: ROC Curves illustrating image classification performance results using Line Analysis.

5.5.2 Classification Results for Colour Analysis

The image classification results based on Colour Analysis are presented in Table 5.14. The results were plotted in ROC curves as illustrated in Figure 5.9. Based on the graph generated as Figure 5.9, **Colour 2** clearly has produced a better result than **Colour 3**. Table 5.15 and Table 5.16 shows quantitative analysis based on **Colour 2** and **Colour 3** threshold values. Results for **Colour 2** (Table 5.15) shows that 326

Threshold	building	Non building	Unknown	Total
	X = 0.8	X = 0.2	0.8 < X < 0.2	
Test Set 1 (building)	342	132	60	534
Test Set 2 (Non building)	143	305	58	506
Total Identified	485	437	118	1040
Identified (%)	46.63%	42.02%	11.35%	100%
Correct (%)	70.52%	67.79%	-	-
Incorrect (%)	29.28%	30.21%	-	-

TABLE 5.12: Quantitative Analysis Result for **Line 1: Long Lines [72 Dirs]**

Threshold	building	Non building	Unknown	Total
	X = 0.8	X = 0.2	0.8 < X < 0.2	
Test Set 1 (building)	364	81	89	534
Test Set 2 (Non building)	131	251	124	506
Total Identified	495	332	213	1040
Identified (%)	47.60%	31.92%	20.48%	100%
Correct (%)	73.54%	75.60%	-	-
Incorrect (%)	26.46%	24.40%	-	-

TABLE 5.13: Quantitative Analysis Result for **Line 3: Short and Long [48 Dirs]**

images are identified as building images, 149 as non building images, leaving almost half of other images unidentified. Results for **Colour 3** (Table 5.16) show that 323 images were identified as building images, 318 as non building images, leaving 401 images in the unknown category.

The classifier was able to identify more building images using **Colour 2** than using **Colour 3** analysis. The accuracy rates for building images are higher in **Colour 2** than **Colour 3**, which are 75.47% and 57.89% respectively. For non building images, the classifier has identified more than double the non building images by using **Colour 3**, compared to **Colour 2**. Nonetheless, the accuracy rate for identifying non building images is higher in **Colour 2** which is 61.74% compared to **Colour 3** which is 58.86%. For colour analyses we can conclude that, image classification has performed better by observing 3D colour histograms with 216 bins for building images and 3 separate colours histograms with 196 bins for non building images.

5.5.3 Classification Results for Integrated Analysis

The image classifications results for integrating line and colour analyses are presented in Table 5.17. Six tests are conducted and the performance results for these tests are illustrated in the ROC curves (Figure 5.10). The ROC curves show all tests have produced positive results. Test 1, Test 2, Test 3 and Test 4 have produced similar curves, indicating performances for these tests are similar. Tests 5 and Test 6 have produced lower curves indicating poorer performances. The quantitative analysis shows that Test 2 has identified more building and non building images than the rest of the tests which is

Threshold	3D [216 bins]		3 Colour Histogram	
	Sensitivity	1- Specificity	Sensitivity	1- Specificity
0.95	0.26	0.07	0.20	0.11
0.90	0.32	0.09	0.25	0.16
0.85	0.39	0.13	0.32	0.23
0.80	0.46	0.16	0.35	0.27
0.75	0.54	0.22	0.40	0.31
0.70	0.61	0.28	0.43	0.34
0.65	0.66	0.34	0.45	0.37
0.60	0.71	0.43	0.49	0.40
0.55	0.74	0.53	0.52	0.42
0.50	0.76	0.58	0.55	0.45
0.45	0.79	0.63	0.58	0.49
0.40	0.81	0.67	0.60	0.52
0.35	0.84	0.72	0.63	0.53
0.30	0.86	0.75	0.68	0.56
0.25	0.88	0.77	0.71	0.60
0.20	0.89	0.82	0.76	0.63
0.15	0.91	0.87	0.77	0.68
0.10	0.92	0.90	0.82	0.72
0.05	0.94	0.93	0.89	0.78

TABLE 5.14: Test Result for Colour Analysis

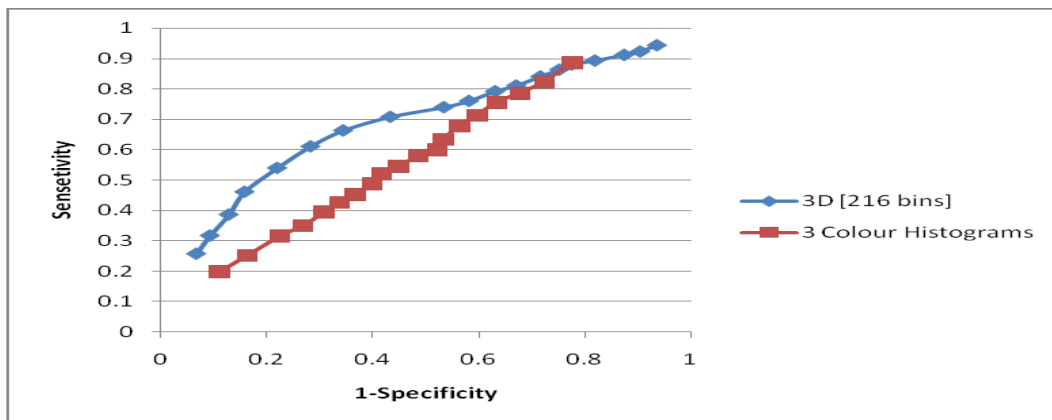


FIGURE 5.9: ROC Curves illustrating image classification performance results using Colour Analysis.

Threshold	building	Non building	Unknown	Total
	X = 0.8	X = 0.2	0.8 < X < 0.2	
building	246	57	231	534
Non building	80	92	334	506
Total	326	149	565	1040
Total (%)	31.35	14.33	54.33	100%
Correct (%)	75.47	61.74	-	-
Incorrect (%)	24.53	38.26	-	-

TABLE 5.15: Quantitative Analysis Result for **Colour 2: 3D Histogram [216 bins]**

Threshold	building	Non building	Unknown	Total
	X = 0.8	X = 0.2	$0.8 < X < 0.2$	
building	187	130	217	534
Non building	136	186	184	506
Total	323	316	401	1040
Total (%)	31.06	30.38	38.56	100%
Correct (%)	57.89	58.86	-	-
Incorrect (%)	42.11	41.14	-	-

TABLE 5.16: Quantitative Analysis Result for **Colour 3: 3H Histogram [196 bins]**

729 images, with 431 building images and 298 non building images. Based on accuracy rate, Test 4 has the highest accuracy rate for finding building images which is 96.74%, while Test 6 has the highest accuracy rate for identifying non building images which is 91.67%. Nevertheless, Test 6 only found less than 30 images as building and non building images, showing use of strict threshold values has increased quality in image classification, but also decreased the quantity of images that could be identified.

The performances for all tests in building and non building image classification are presented in Figure 5.11. A reminder of the tests is given as follows:

1. Line Analyses:

- (a) Line 1: Long Lines Histogram with 72 directions
- (b) Line 3: Short and Long Line Histograms with 48 directions

2. Colour Analyses:

- (a) Colour 1: 3D Colour Histograms with 64 bins
- (b) Colour 2: 3D Colour Histograms with 216 bins
- (c) Colour 3: 3 Colour Histograms with 196 bins

3. Integrated Analyses

- (a) Test 1: Colour 1 with Line 1
- (b) Test 2: Colour 1 with Line 3
- (c) Test 3: Colour 2 with Line 1
- (d) Test 4: Colour 2 with Line 3
- (e) Test 5: Colour 3 with Line 1
- (f) Test 6: Colour 3 with Line 3

Threshold	TEST 1		TEST 2		TEST 3		TEST 4		TEST 5		TEST 6	
	Sens	1-Speci	Sens	1-Speci	Sens	1-Speci	Sens	1-Speci	Sens	1-Speci	Sens	1-Speci
1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.95	0.55	0.11	0.54	0.12	0.02	0.00	0.10	0.01	0.00	0.00	0.00	0.00
0.90	0.60	0.14	0.61	0.15	0.08	0.01	0.18	0.02	0.00	0.01	0.00	0.00
0.85	0.64	0.16	0.64	0.17	0.14	0.02	0.29	0.05	0.00	0.01	0.00	0.01
0.80	0.69	0.21	0.68	0.22	0.25	0.04	0.37	0.07	0.02	0.02	0.01	0.03
0.75	0.72	0.23	0.71	0.24	0.36	0.06	0.47	0.10	0.03	0.04	0.03	0.05
0.70	0.76	0.26	0.73	0.26	0.47	0.11	0.58	0.15	0.09	0.07	0.07	0.09
0.65	0.78	0.28	0.76	0.28	0.58	0.16	0.65	0.19	0.20	0.11	0.18	0.12
0.60	0.80	0.30	0.78	0.30	0.66	0.22	0.76	0.24	0.34	0.16	0.36	0.19
0.55	0.81	0.32	0.80	0.32	0.76	0.29	0.80	0.29	0.57	0.23	0.57	0.29
0.50	0.82	0.35	0.82	0.35	0.85	0.38	0.86	0.40	0.78	0.35	0.79	0.42
0.45	0.84	0.36	0.83	0.37	0.90	0.48	0.89	0.49	0.90	0.49	0.91	0.56
0.40	0.85	0.38	0.84	0.40	0.93	0.59	0.92	0.62	0.96	0.65	0.96	0.72
0.35	0.87	0.41	0.86	0.44	0.95	0.70	0.94	0.70	0.97	0.81	0.98	0.83
0.30	0.88	0.45	0.87	0.48	0.96	0.80	0.96	0.79	0.99	0.89	0.99	0.91
0.25	0.89	0.48	0.88	0.51	0.98	0.87	0.97	0.84	0.99	0.94	0.99	0.95
0.20	0.90	0.52	0.89	0.54	0.99	0.91	0.98	0.88	1.00	0.96	1.00	0.98
0.15	0.92	0.56	0.91	0.57	0.99	0.94	0.99	0.91	1.00	0.96	1.00	0.99
0.10	0.94	0.62	0.93	0.64	1.00	0.96	0.99	0.95	1.00	0.99	1.00	1.00
0.05	0.96	0.71	0.96	0.70	1.00	0.98	1.00	0.97	1.00	1.00	1.00	1.00
0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

TABLE 5.17: Tests Result for Integrating Line and Colour Analysis

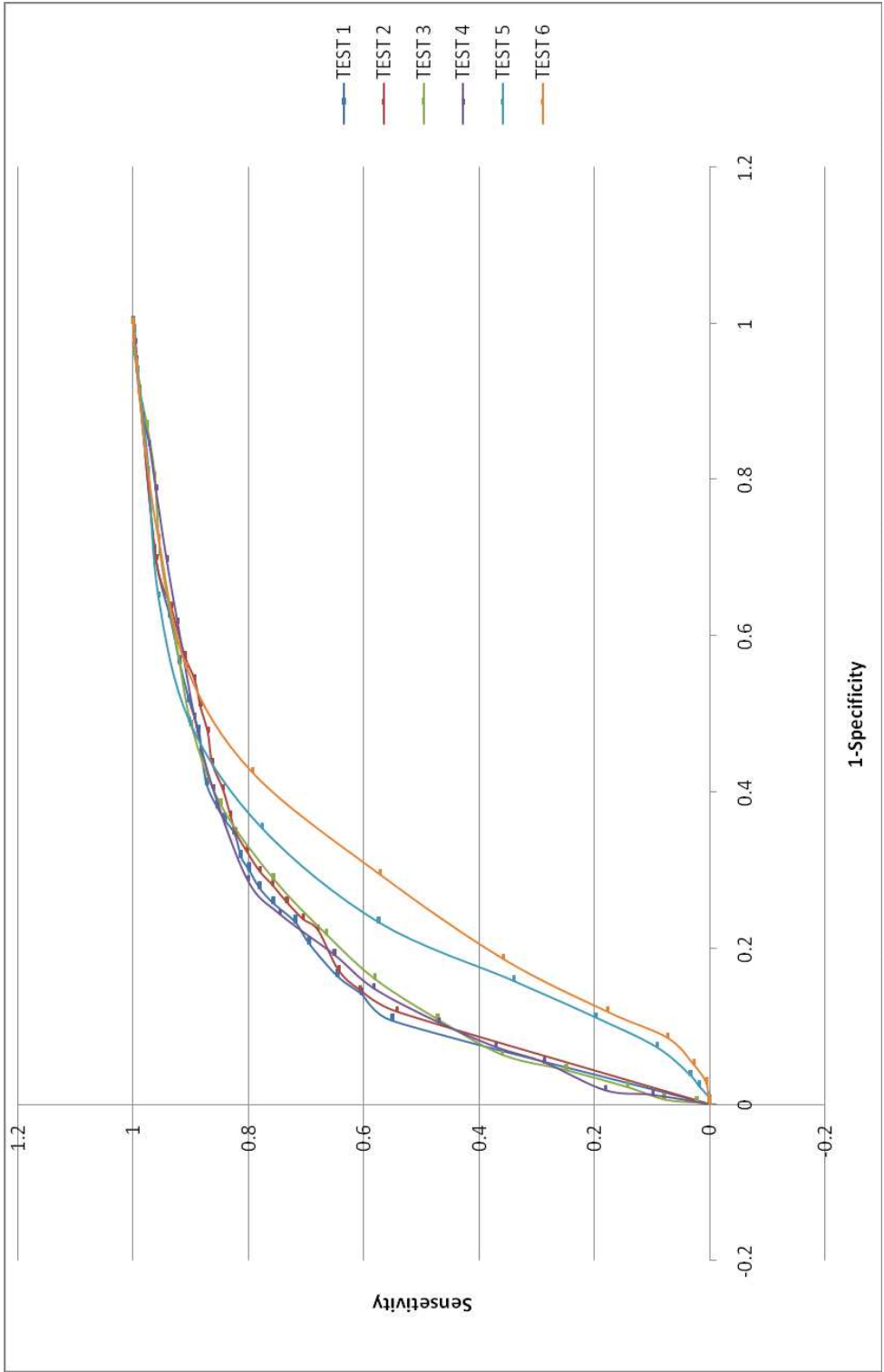


FIGURE 5.10: Overall Image Classification Performance Comparisons and Conclusion.

Threshold	building	Non building	Unknown	Total
	X = 0.85	X = 0.15	$0.85 < X < 0.15$	
building	245	43	246	534
Non building	83	220	203	506
Total	328	263	449	1040
Total (%)	31.54%	25.29%	43.17%	100%
Correct (%)	74.70%	83.65%	-	-
Incorrect (%)	25.30%	16.35%	-	-

TABLE 5.18: Quantitative Analysis Result for **TEST 1**

Threshold	building	Non building	Unknown	Total
	X = 0.85	X = 0.2	$0.8 < X < 0.2$	
building	344	57	133	534
Non building	87	241	178	506
Total	431	298	311	1040
Total (%)	41.44%	28.65%	29.90%	100%
Correct (%)	79.81%	80.87%	-	-
Incorrect (%)	20.19%	19.13%	-	-

TABLE 5.19: Quantitative Analysis Result for **TEST 2**

Threshold	building	Non building	Unknown	Total
	X = 0.85	X = 0.15	$0.85 < X < 0.15$	
building	76	4	454	534
Non building	11	32	463	506
Total	87	36	917	1040
Total (%)	8.37%	3.46%	88.17%	100%
Correct (%)	87.36%	88.89%	-	-
Incorrect (%)	12.64%	11.11%	-	-

TABLE 5.20: Quantitative Analysis Result for **TEST 3**

Threshold	building	Non building	Unknown	Total
	X = 0.85	X = 0.15	$0.8 < X < 0.2$	
building	153	27	354	534
Non building	5	44	457	506
Total	158	71	811	1040
Total (%)	15.19%	6.83%	77.98%	100%
Correct (%)	96.84%	61.97%	-	-
Incorrect (%)	3.16%	38.03%	-	-

TABLE 5.21: Quantitative Analysis Result for **TEST 4**

The line analyses are indicated with diamond marks. The colour analyses are indicated with triangle marks while the integrated analyses are indicated with small bar marks. The integrated analyses, which are Test 1, Test 2, Test 3 and Test 4 have performed better than the rest of the analyses. Line analyses have better results compared to colour analysis. The lowest classification performance is given by Colour 3 as the ROC curves appeared to be the nearest to the neutral line. In conclusion, image classification

Threshold	building	Non building	Unknown	Total
	$X = 0.8$	$X = 0.2$	$0.8 < X < 0.2$	
building	10	1	523	534
Non building	12	18	476	506
Total	22	19	999	1040
Total (%)	2.12%	1.85%	96.06%	100%
Correct (%)	45.45%	94.74%	-	-
Incorrect (%)	54.55%	5.26%	-	-

TABLE 5.22: Quantitative Analysis Result for **TEST 5**

Threshold	building	Non building	Unknown	Total
	$X = 0.85$	$X = 0.15$	$0.85 < X < 0.15$	
building	1	1	532	534
Non building	14	11	481	506
Total	15	12	1013	1040
Total (%)	1.44%	1.15%	97.40%	100%
Correct (%)	6.67%	91.67%	-	-
Incorrect (%)	93.33%	8.33%	-	-

TABLE 5.23: Quantitative Analysis Result for **TEST 6**

can be improved by integrating colour and line analyses rather than using line or colour analysis alone.

In the 2008 and 2009 TRECVid challenges one of the tasks involved the development of a cityscape detector, similar but not identical in functionality to our building detector. The best top 10 inferred average precision (infAP) achieved was in the range 0.28 to 0.37 (Awad et al. (2008) and Over et al. (2010)). Although these infAP values are lower than our classification performances, their object identification tasks are based on finding object in a shot, which contains multiple frames while our research is only focusing on identifying a specific object in a single image. However, it would be interesting to use the object detectors developed in TRECVid with our own images to compare the results more thoroughly.

5.6 Conclusions

In this chapter we have developed a classifier which can be used for automatically annotating images with a “buildings” or “non-buildings” flag. Although just a proof of concept it can be seen that other image classifiers could be developed to add additional annotations to images based on their content and hence provide enhanced retrieval facilities for those images.

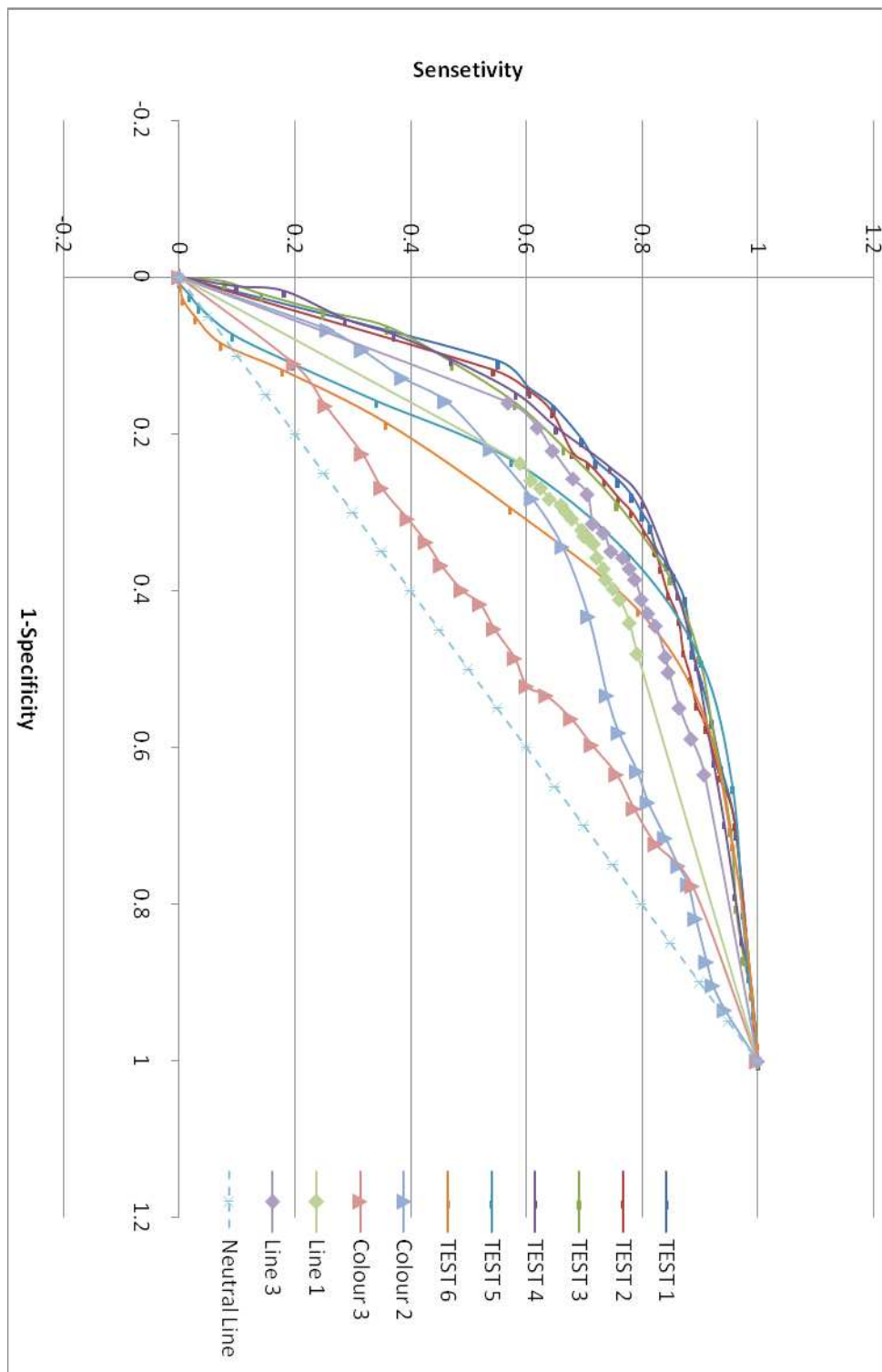


FIGURE 5.11: ROC Curves comparison between all tests done in image classification

Chapter 6

Evaluation: Results and Discussions

This chapter discusses a range of experiments to evaluate the techniques developed in earlier chapters. Firstly, the semantic model produced by text analysis, is compared with conventional tag based searching. Then a comparison is made between searching using image classifier results and text based searching. This is followed by an evaluation of an integrated approach using a hybrid of the text and image based methods. The main objective of the experiments is to evaluate each approach based on standard information retrieval performance metrics, precision and recall.

6.1 Introduction

Use case scenario: A user may be searching for information related to the Malaysian tourism domain and using different methods, for example, based on text only, image only or integrating text and images. Different document sets are presented to reflect real case scenarios of searching for images with rich descriptions and images with poor descriptions. The evaluation of the approaches used in the study is based on their ability to find relevant information/images tailored to the user queries. Two evaluation metrics are used:

- *precision* refers to the proportion of a set of retrieved images that are relevant to the query
- *recall* refers to the proportion of all the relevant images in the search dataset which are retrieved.

The evaluation processes are divided into four main tasks which are:

1. Task 1: Text based searching versus tag based searching. The objective of this task is to evaluate search using the concepts in the semantic information model generated by text analysis and to compare the result with conventional tag based searching. Two cases are considered. The first uses a set of documents which are well tagged and described. The second uses a set of documents which are less well tagged or described. Results and discussion for Task 1 are given in Section 6.2.
2. Task 2: Image based searching versus text based search. The objective of this task is to evaluate the two approaches (text and image search separately) based on identifying building/city images. Two cases were used in this experiment and both of the cases consist of well tagged/described images, where the first case has a wider range of images such as landscapes, people and celebration, and the second case focuses on images tagged with city. Results and discussion for Task 2 are given in Section 6.3.
3. Task 3: Integrating text and Image approaches. The objective of this experiment is to evaluate how much the text and image based approaches complement each other by evaluating a hybrid approach and using similar cases to those in Task 2. Results and discussion for Task 3 are given in Section 6.4.
4. Task 4: Using image classifiers to classify a wider range of images than building/city landscape. The objective of Task 4 is to demonstrate on integrating text based search with the image analysis approach using a wider range of classifiers than those used in Task 2 and Task 3. Results and discussion for Task 4 are given in Section 6.5.

6.2 Task 1: Text based searching.

Does the semantic model generated by text analysis improve retrieval compared to using tags alone?

To answer this question, we conducted an information retrieval experiment to observe the results produced by combining natural language processing and knowledge based approaches in the Semantic Model introduced in Chapter 4. The objective for Task 1 is to evaluate the ability of the Semantic Model to find correct images compared to a conventional tag based search. The analysis is done by observing and comparing the semantic model and tags in the form of a bag of words. Tagging does not have a standard method for describing concepts that have two or more words. For example, images related to “*Mabul island*” could be represented with different patterns such as in the phrase (mabul island), one single word (*mabulisland*) or with two words (*mabul, island*). In cases where the query information involves phrases, for tagging, we assumed queries matching to any of these patterns are correct. For the semantic model approach, to

ensure the quality of the information used to describe the images, we only consider it is correct if an exact match is found with the query phrase itself.

Two types of querying are considered. In the first, the query concept is matched to the tag lists in each document. We refer to this as tag based search. In the second, the query concept is matched to the list of concepts from the semantic model which includes information from tags, extracted concepts and enriched information from the knowledge base. This is referred to as Semantic Model based search. The ground truth for these experiments was obtained manually by looking at each image. Here, annotations are based both on what the image contains and also what it is related to. For example, an image may be labelled with *KL Tower*, either because the *KL tower* is in the image or because the image was taken from the **KL tower**. 36 queries were used in this assessment ranging from finding general images (such as **beach**, **island** and **mountain**) to specific images (such as *tioman island*, *citrawarna festival* and *mount kinabalu*). Two standards information evaluation metrics were used which are precision and recall.

Two cases of Flickr documents were used in this experiment.

- CASE 1 contains popular images and consists of 370 images with full descriptions, most of which are properly tagged and have substantial feedback from Flickr's users. Statistically, the highest number of tags in CASE 1 is 73 tags while the lowest is 5 tags. The average numbers of tags per image in CASE 1 is 20.93.
- CASE 2 contains unpopular images/documents which consists of 270 images and most of these are lacking tags or descriptions or both. Statistically, the highest number of tags in CASE 2 is 60 while the lowest is none. The average number of tags per image in CASE 2 is 4.73.

Queries, samples and results for the Task 1 evaluation are presented in Table 6.1. It can be seen that the semantic model has produced better results compared to tag based search in both precision and recall for both cases. The semantic model's average precision value across the two cases is 0.62 which is 0.14 higher than the average tag based search precision value. The average recall value for the Semantic model is 0.26 higher than the average recall value for tag based search. In CASE 1, the precision value for tag based search is 0.63 while for the semantic model it is 0.81, which has improved by 0.18 from the tag based search. The recall value for tag based search is 0.40 and the value has increased by 0.30 to 0.70 for the semantic model. Lack of textual descriptions in CASE 2 has resulted in lower results in precision and recall values for both tag and semantic model search. Here, the precision value for tag based search is 0.32, and it has increased to 0.44 by using the semantic model. The recall value for tag based search is 0.20, and the value has increased by 0.23 to 0.42 for the semantic model.

i) Sample queries and results observed based on tags based and semantic based search for CASE 1.

Queries	Tags		Semantic Model		Relevant Images
	Identified	Correct	Identified	Correct	
Q1: Beach	76	70	85	78	134
Q2: Island	137	113	149	122	135
Q3: Mountain	13	10	14	11	21
Q4: Building	48	42	54	47	76
Q5: tioman+island/ tiomanisland/ tioman island	22	16	19	16	17
Q6: “tioman island”	1	0	19	16	17

ii) Precision and Recall generated from observing results in CASE 1 and CASE 2.

CASE	Precision		Recall	
	Tags	Semantic Model	Tags	Semantic Model
1	0.63	0.81	0.40	0.70
2	0.32	0.44	0.20	0.42
Average	0.48	0.63	0.30	0.56

TABLE 6.1: Queries and results for text based search analysis

The use of knowledge bases has increased the capabilities of the semantic model approach to extend existing information describing the images with information in the knowledge bases. For example, alternative names for the tioman island entry such as *tioman* and *pulau tioman* (a Malay word for *tioman island*) are found in the Malaysian Tourism Ontology thus allowing these words to be associated with *tioman island*. Moreover, the concept *tioman island* is also identified in the Geonames ontology and Dbpedia which provides a richer and more useful range of information to improve the capabilities of information retrieval and which can be used beyond the capability of the tag based search itself.

The natural language processing has played an important part to identify concepts in the form of phrases. In this experiment, *tioman island* was identified from the textual description of the images despite it being tagged as one word *tiomanisland* or two different words *tioman* and *island*. For our approach, we only consider the correct image was found when the query matched with the phrase *tioman island* thus other patterns will be considered as incorrect. For Q5 the phrase *tioman island* is treated as two separate words, one word and a phrase while for Q6 it is treated as one phrase exclusively. The effect of this can be seen clearly in Q5 (the tags approach as presented in Table 6.1(i)) where the tag based search has found more images than using the semantic model. For Q5, although the tag based search found more images, it has a lower precision (0.72) compared to the semantic model search (0.84).

Beyond the capability of the tags based search, the semantic model offers a more specific searching capability which allows the information retrieval to reuse information in the knowledge bases. The semantic model allows the information retrieval to narrow or broaden the search in order to identify specific images or information tailored to the requirements of the queries. Two examples are given in the following subsections.

6.2.1 Extracting information related to location.

The most accurate approach to identify information about location is based on the Geonames ontology. For example *Where is Mount Kinabalu located?*. If *Mount Kinabalu* has entries in the Geonames ontology, the question could be easily answered by locating *admin name* in the entry. For this question, the answer would be *Sabah* or *The State of Sabah*. A SPARQL query example for finding broader term/location for the given query is presented as follows:

```
{ ?termGeo geo:name ?query.
  ?termGeo geo:adminName ?location .
  FILTER regex(?query, "mount kinabalu" , "i") }
```

Furthermore, the Geonames ontology also contains information such as location coordinates and geographical features for each location. Table 6.2 shows admin name results for each concept/object queried which are *Sabah* for *mount kinabalu*, *Serawak* for *Bako national park*, *Selangor* for the *Twin Towers* and *Sabah* for *perhentian*. Table 6.2 also shows other information related to the queries which are coordinates and features. In cases where images have coordinate information which is usually stored in the image's EXIF metadata, the coordinates can be used to identify the location for the images by querying the Geonames ontology. Most of the time, the coordinates will identify the very specific location where the images were taken.













Location (?query)	Mount Kinabalu	Bako National Park	Twin Towers	Perhentian
Admin Name (? location)	Sabah	Serawak	Selangor	Terengganu
Coordinates	Lat : 6.0833 Long : 116.55	Lat : 1.7167 Long : 110.4667	Lat : 3.158 Long : 101.7116	Lat : 5.915 Long : 102.7397
Features	mountain,hill,rock	parks,area,	spot, building, farm,	mountain,hill,rock
Geonames Source/ID	Geo: 1736632	Geo: 1781760	Geo: 6619683	Geo: 1750625
Images for RDF Files which contain the queried concept.	 1813186786.jpg  2100904083.jpg  3074479670  3899564198	 1346636304.jpg	 959060474.rdfs  1105570354.rdf  2402758780	 3344089984.jpg  3345994729.jpg  3346597188.jpg  3346597190.jpg

TABLE 6.2: Results for finding admin location for known object/location.

6.2.2 Locating information related to attractions and events tourism in Malaysia

Information related to attraction and events could be identified by binding queried concepts with concepts in the Tourism Malaysia Ontology (MTO). For example, “*What are the attractions and types of attractions in Sabah?*” and “*Give me information about the Citrawarna Festival?*” As discussed in Chapter 4, MTO contains two main roots which are Attraction and Event.

Attraction: Each Attraction entry consists of information such as *name* (*hasName*), *location* (*hasPostalAddress*) and *attraction type* (*hasAttType*). The *hasPostalAddress* information is linked with the Geonames entry. For example, the Geoname entry for *Sabah* is <http://www.geonames.org/1733039>. There are 14 entries in MTO for attractions related to *Sabah*. Table 6.3 shows a SPARQL query example, some samples for the entries along with other related information such as *hasName* and *hasAttType* and images that correspond to the examples.



i) SPARQL Query:			
<pre>?entryMTO tour:hasPostalAddress <http://www.geonames.org/1733039> ?entryMTO tour:hasName ?name ?entryMTO tour:hasAttType ?type</pre>			
ii) Results example for attraction entry			
Query	MTO entry	hasName (?name)	hasAttType (?type)
Sabah	Mabul Island	Mabul island, mabul ,pulau mabul	Coral reef, fishing village, island, scuba diving
Sabah	Sipadan Island	Sipadan Island, sipadan, pulau sipadan	National park, coral reef, island, fishing village
Sabah	Maliau Basin	Maliau Basin	National park,
Sabah	Kinabalu Park	Mount Kinabalu, kinabalu, kinabalu park, gunung kinabalu	Mountain, national park
iii) Images related to attraction in Sabah			
			
Mabul Island 479883807.jpg	Kinabalu Park 4567532077.jpg	Sipadan Island 4537572847.jpg	Maliau Basin 2408903375.jpg

TABLE 6.3: SPARQL Query and Results of attractions related to Sabah

Event: Event entries have similar attributes to Attraction, but with additional information which is venue (hasVenue) and date (hasTimePeriod). The date information is linked to DatePeriod Class which consists of start date (hasStartDate), end date (hasEndDate) and duration of the event (hasDuration). Results for finding information about the Citrawarna festival are presented in Table 6.4.

i) SPARQL Query:	
<pre> ?eventMTO tour:hasName citrawarna festival ?eventMTO tour:Description ?description ?eventMTO tour:hasVenue ?venue ?eventMTO tour:hasTimePeriod ?time ?time tour:hasDuration ?duration </pre>	
ii) Results example for event entry	
Query	Citrawarna Festival
hasDescription	A nationwide Colours of Malaysia extravaganza....
hasName	Citrawarna Malaysia, Colour of Malaysia, Citrawarna Festival, Citrawarna
hasVenue	All around Malaysia
hasTimePeriod	<Citrawarna_Festival_Date>
*hasTimePeriod is linked to other class as follows:	
Query	Citrawarna_Festival_Date
hasDuration	One Month
iii) Images related to Citrawarna Festival	
	
521030838.jpg	187784965.jpg
	
2522518439.jpg	2522513957.jpg
	
187784963.jpg	

TABLE 6.4: SPARQL Query and Results of attractions related to Citrawarna Festival.

In some cases, Dbpedia entries provide richer sets of information beyond the capabilities of the MTO and Geonames ontologies which have fixed parameters. Table 6.5 shows an example of information related to Petronas Towers which is extracted from Dbpedia. The information allows information retrieval to find answers to such queries as *What is the floor count in Petronas towers?* However, since Dbpedia information depends on user generated content, the information is sometimes inconsistent.

Dbpedia	Petronas _{Towers}
URI	http://dbpedia.org/resource/Petronas_{Towers}
Property:label	Petronas Towers
Property:buildingName	Petronas Twin Towers
Property:cost	\$ 1,6 billion
Property:year	1998 “2003”
Property: architect	C?@sar Pelli, Mahathir bin Mohamad
Property:floorCount	88
surpassedByBuilding	http://dbpedia.org/resource/Taipei101

TABLE 6.5: Example of information related to Petronas Towers extracted from Dbpedia resource.

6.3 Task 2: Image Analysis based search compared with text based search

i. Does the city/building classifier improve search capability for identifying building images compared to using text based searching alone?

To answer this question, the text based searching approaches, tag based and semantic model based, were compared to image based search. Here, image based search means search which uses the results of an image classifier which identifies an image as building or not building as described in Chapter 5. Again, two sets of documents were used which are CASE 1 and CASE 3.

The image analysis is done by analysing building and non building images based on threshold values observed in Chapter 5. The threshold values for identifying building images are at 0.85 and above while for non building images are at 0.15 and below. Result for the image analysis is presented in Table 6.6. The building classifier has identified 79 images out of 370 images as building images. There are only 76 relevant building images in CASE 1, therefore the proportion of relevant building images in CASE 1 is only 20%. Results for classifying building and non building images for CASE 1 is presented in Table 6.6. With the precision value of 0.47, it shows that almost half of the identified building images are correctly classified. The precision and recall value for finding non building images are better compared to identifying building images.

Threshold	Building	Non building	Relavant Images	Precision	Recall
	X \geq 0.85	X \leq 0.15			
building	37	14	76	0.47	0.49
Non building	42	164	294	0.92	0.56
Total	79	178	370		

TABLE 6.6: Image Classification Results for **CASE 1**

In this task, we are only interested in finding images annotated with 'building'. As stated in Task 1, tag based search has identified 42 of these correctly as building images and the semantic model has correctly identified 48 images. Results for identifying building images by using text (tags and semantic model) and image based searching are presented in Table 6.7. For CASE 1 text based searching (tags and semantic model) has a better recall and precision performances compared to the image analysis approach. The recall value for tag and semantic model are 0.55 and 0.62 respectively while the image analysis is 0.49. The precision value for tag and semantic model are 0.88 and 0.87, while the precision value for image analysis is 0.47.

Approach	Building		Relevant Images	Precision	Recall
	Identified	Correct			
Image Analysis	79	37	76	0.47	0.49
Tag	48	42	76	0.88	0.55
Semantic Model	54	47	76	0.87	0.62

TABLE 6.7: CASE 1-Results for identifying building images by using text based search and image analysis.

The image set CASE 3, consists of 400 most relevant documents (according to Flickr API) gathered by querying "*City*" and "*Malaysia*" using a Flickr tag based search. As for CASE 1, most of these images are well tagged and described. Statistically, the highest number of tags in CASE 3 is 75 tags while the lowest is 5 tags. The average number of tags per image in CASE 3 is 24.72, which is slightly higher than CASE 1. Based on inspection, 254 images in CASE 3 are building and 146 images are not. The precision value for querying *City* and *Malaysia* in Flickr is 0.64. Results for CASE 3 analysis are presented in Table 6.8 together with precision and recall values. The image analysis has classified 174 images as building of which 142 are correct classifications. The tag based search found 117 images as building images of which 90 were correctly tagged. Based on these results, image analysis based retrieval has achieved the highest precision which is 0.82, followed by the semantic model and tag based approaches. Despite having the highest precision, the image analysis retrieval has lower recall which is 0.56, short by 0.11 compared to the semantic model.

Approach	Building		Relevant Images	Precision	Recall
	Identified	Correct			
Image Analysis	174	142	254	0.82	0.56
Tag	117	90	254	0.77	0.35
Semantic Model	210	170	254	0.81	0.67

TABLE 6.8: CASE 3: Results for identifying building images by using text based search and image analysis.

The results shows that, in CASE 1 where images are well tagged and described, text based search has the advantages over the image based search. Manual annotation (tagging) has provided the highest score in precision followed closely by the semantic model. The Semantic model based approach has the highest precision compared to the tag and image based search. It also shows that manual annotating (tagging) done by Web 2.0 users can sometimes be influenced by their knowledge background. Based on observation of CASE 2, Malaysian city landscape images are usually dominated by the two most prominent building landscapes, which are *Kuala Lumpur twin towers building* and *Kuala Lumpur tower*. In some cases, these images are tagged with specific information such as *klcc*, *petronas twin towers* and in some cases the images were tagged with generic information such as *buibuilding*. Although the images were tagged with specific information, the semantic model approach has the advantage to identify these images as building if the text analysis could identify specific concepts such as (*klcc*, *petronas twin towers*, *kl twin towers* and *twin towers*). The information retrieval has matched these concepts with concepts in the MTO, and infers these concepts as *building*.

6.4 Task 3: Integrating Text and Image based search: the Hybrid Approach.

Does integrating text based search with image based search improve the performance level in the information retrieval?

To answer this question, the two text approaches, which are tag and semantic model, were first each integrated separately with image based searching.

Table 6.9 shows the results of this integration for CASE 1. It can be seen that both integrated approaches show an increase in recall over the recall for the unintegrated approaches (Table 6.7). Integrating tags and image analysis has increased the recall value to 0.68, increases of 0.20 and 0.13 compared to using image analysis or tags separately. Integrating the semantic model with image analysis has provided a better recall which is 0.74, improving on each of the separate approaches (as presented in Table 6.7) which are tag based search by 0.21, image based search by 0.27 and semantic model by 0.14. The precision values for the integrated approaches have fluctuated over the unintegrated approaches. Integrating tags and image analysis has increased the precision value to 0.51,

a slight increase of 0.04 to using image analysis, but it has decreased to 0.38 to using tag based search. Integrating the semantic model with image analysis has generated a better precision which is 0.52 compared to the tag and image integration. However, the precision value is only improved for the image based search, but the value is lower than each of the text base searches.

Approach	Building		Relevant Images	Precision	Recall
	Identified	Correct			
Tag and Image Analysis	102	52	76	0.51	0.68
Semantic Model and Image Analysis	107	58	76	0.52	0.76

TABLE 6.9: Case 1: Results for integrating text based searching (tags and semantic model) with image analysis.

Table 6.10 shows the results for integrating each of the two text approaches with the image based approach for the image set CASE 3. It can be seen that integrating the semantic model and image analysis has improved the recall value by around 0.14 and produced a slight improvement in precision compared to integrating the tag and image analysis approaches.

From Table 6.8 and Table 6.10 together it can be seen that the integrated semantic model and image analysis approach has found 110 and 74 additional images compared to using image analysis and the semantic model separately. The recall value of integrated tag and image analysis retrieval has increased by almost 0.40 compared to using tags based search alone. However, the precision values for the integrated approaches are slightly lower than using the approaches separately. The precision value of integrated tag and image analysis is 0.77, which is similar to tag based approach and 0.04 lower than using image analysis separately. The precision value for the integrated semantic model and image analysis is 0.78, which shows a slight decrease of around 0.04 and 0.03 than separately using image analysis and the semantic model respectively.

Approach	Building		Relevant Images	Precision	Recall
	Identified	Correct			
Tag and Image Analysis	240	185	254	0.77	0.73
Semantic Model and Image Analysis	284	220	254	0.78	0.87

TABLE 6.10: CASE 3: Results for integrating text based searching (tags and semantic model) with image analysis.

The results for both cases show that integrating text and image analysis is better than using the approaches separately. In cases where building images are not tagged with building, these images would be unsearchable by text based search. If the image based search is able to classify these images, it complements the text based search by finding new images for text based search. However, in cases where images were not classified

as building images by image based search, text images could complement image based search by identifying these images if they have been tagged manually with building.

6.5 Task 4: Using image classifiers to classify other images than building/city landscape

Can image based searching be improved by going beyond building classifier specification?

This experiment demonstrates the use of the capability of our image classifiers to work beyond the building classification domain. Two examples are used to explore the issue which are searching for sunset and searching for beach images which represent two different difficulty levels. Searching for sunset images should be less difficult compared to searching for beaches images. Sunset is a general phenomenon and can occur in any instances of outdoor images regardless of the landscape. Only colour features are used to classify sunset images. Unlike beach images, the classifier needs to distinguish between beach images and other landscapes. Here, two features are used which are colour and line histogram. The results presented in this experiment are based on observing images with the highest predicted probability of being the requested images. It is different from the evaluation approach for finding building/city images where the threshold value for building images were obtained from intensive image analysis experiments (see Chapter 5).

During text based search, we have found 54 sunset images in CASE 1 and only 2 out of 72 sunset images in CASE 2. By using images identified in CASE 1 as training images, the experiment tried to find other sunset images in CASE 2 by using the image classifier approach. Based on observing these 54 sunset images, there are 49 correct sunset images and 5 incorrect sunset images. Images in CASE 2 are used as a test set. With a slight modification, the image classifier is easily transformed into a sunset classifier. Based on our experience in previous experiments, the 3D colour histogram with higher bin count will produce a better result than with lower bin count (see chapter 5). Therefore, the experiment is based on a 3D colour histogram with 216 bins. The results for classifying sunset images in CASE 2 are presented in the Appendix. Table 6.11 shows precision and recall values based on observing 100 images with the highest sunset prediction value.

Number of Images	Correct	Incorrect	Precision	Recall
10	8	2	0.80	0.13
20	16	4	0.80	0.26
30	23	7	0.77	0.38
40	27	13	0.68	0.44
50	34	16	0.68	0.56
60	36	24	0.60	0.59
70	45	25	0.64	0.74
80	49	31	0.61	0.80
90	53	37	0.59	0.87
100	54	46	0.54	0.89

TABLE 6.11: Results for recall and precision values for 100 images with the highest sunset prediction value.

Table 6.12 shows lists of 20 images with the highest probability of being sunset images. The image with the highest sunset probability value is on the top left. The images are presented in rank order from left to right. For the first 10 and 20 images, the precision value is 0.8, showing 80% of the images identified were correct leaving only 4 incorrect images. As can be seen in the Table 6.12, the incorrect images are number 6, 10, 11 and 13. These images have a similar background colour as sunset which might be produced by sources such as building painting, colour reflection and artificial light. The precision value is decreased as the number of observed images increased. By contrast, the recall value is increased as the number of observed images increased. Based on observation of the 100 images with the highest sunset prediction results, the precision value went down to 0.54 while the recall value went up to 0.89. Although the precision value is almost 50:50, the recall value for the image classifier shows that it was able to identify almost 90% of the sunset images in CASE 2, which is better than the 2 images out of 72 found using the text based search.





















				
0.9996 (1) 3150060907.jpg	0.9993 (2) 1858827239.jpg	0.999 (3) 3140386489.jpg	0.9988 (4) 3232976437.jpg	0.9984 (5) 4202407600.jpg
				
0.9983 (6) 1266117254.jpg	0.9976 (7) 3405922879.jpg	0.9975 (8) 4332083816.jpg	0.997 (9) 3843866254.jpg	0.9967 (10) 3131137552.jpg
				
0.9965 (11) 524885771.jpg	0.9963 (12) 4513894127.jpg	0.9963 (13) 3289669699.jpg	0.996 (14) 3388906818.jpg	0.9959 (15) 3538776170.jpg
				
0.9957 (16) 1853056119.jpg	0.9957 (17) 2419024638.jpg	0.9957 (18) 2191786925.jpg	0.995 (19) 4381426087.jpg	0.9949 (20) 3375265668.jpg

TABLE 6.12: Lists of 20 images with the highest probability of being sunset images.

Table 6.13 lists the 20 images with the highest probability of being beach images. Each image is presented with rank number and prediction value. The image with the highest sunset probability value is on the top left. The images are presented in rank order from left to right. In this example, the text based search has found 85 images in CASE 1 with 78 correct images and 7 incorrect images. However, the image retrieval only found 4 out of 60 beach images in CASE 2. The beach results in CASE 1 are used as training images

to classify beach images. Results for the beach classifications are given in Table 6.14. Starting with a promising result, the image classifier has found 7 out of 10 images with highest beach prediction value are correct. Nevertheless, the precision value drops to 50:50 for 20 to 40 observed images and continues to decline to 0.43 at 100 observed images. Table 6.13 shows 20 images that have the highest beach prediction value for CASE 2. For the first 20 images, the image classifier has confused beach images with building (6, 9, 10 and 20), lake (16 and 17), and cloud (15) images.





















 0.9236 3538776170.jpg	 0.9195 3355636424.jpg	 0.8909 4151693456.jpg	 0.8574 3369576324.jpg	 0.8508 3394770387.jpg
 0.8459 437147208.jpg	 0.8452 2545344331.jpg	 0.8378 4442877784.jpg	 0.8236 3131137552.jpg	 0.8158 3598090976.jpg
 0.8069 4259986034.jpg	 0.7812 3665210486.jpg	 0.7735 2472912285.jpg	 0.7478 361732529.jpg	 0.7471 3150060907.jpg
 0.7449 3236995216.jpg	 0.7424 3391748385.jpg	 0.7405 3004947580.jpg	 0.7405 3234047982.jpg	 0.7372 493600065.jpg

TABLE 6.13: Lists of 20 images with the highest probability of being beach images.

Number of Images	Correct	Incorrect	Precision	Recall
10	7	3	0.70	0.11
20	10	10	0.5	0.16
30	15	15	0.5	0.25
40	20	20	0.5	0.33
50	23	27	0.46	0.38
60	29	31	0.48	0.47
70	34	36	0.49	0.56
80	36	38	0.45	0.59
90	42	48	0.47	0.69
100	43	57	0.43	0.70

TABLE 6.14: Results for classifying beach images by using Image Analysis (Line and 3D colour histograms). The result shows correct and incorrect images, precision and recall for 10 to 100 observed images.

For the purpose of comparison, the experiment for beach images is repeated by using 3D colour histogram with 216 bins. The result for the experiment is given in Table 6.15. Based on observing the first 10 images with highest beach prediction values, the image classifier has classified 6 out of 10 correct images. The first 10 images shows that integrating two features (colour and line histograms) is better than one feature (colour histogram) for finding beach images. Nevertheless, after observing 100 images, the value for both precision and recall for using two features are lower than using colour feature alone.

Both examples have shown that the image classifier is extensible to other domains of interest and significantly able to tackle images with lack of textual description. Using the predefined images retrieved from CASE 1 as a training set, the image searches have identified more images than using text based search alone. These results are also good evidence to support an answer to the given questions asked in Section 6.4, which is whether integrating text based search with image based search would improve the information retrieval compared to using them separately. In this case, the image based search completely depends on results given in text based search. If the text based search is unable to provide any information, then the image based search would not be useful.

By using predefined images which are retrieved from CASE 1 as a training set, these images were observed to distinguish between correct and incorrect images. Based on the observation, the text based search was able to identify more correct images than incorrect images, thus providing many true images and only a few false images as a training set into the image based search. Lack of false images might have caused inadequate information to identify false images which is crucial to discriminate between false and true during the classification process. This issue could be overcome by adding a user feedback mechanism, where the user can participate to improve the searching capability by selecting good or bad images which would be updated to the image based search to improve the classification of the requested images. The idea will be extended further in the Future Work.

Number of Images	Correct	Incorrect	Precision	Recall
10	6	4	0.60	0.10
20	11	9	0.55	0.18
30	17	13	0.57	0.28
40	23	17	0.58	0.38
50	28	22	0.56	0.46
60	32	28	0.53	0.52
70	36	34	0.51	0.59
80	41	39	0.51	0.67
90	44	46	0.49	0.72
100	46	54	0.46	0.75

TABLE 6.15: Results for classifying beach images by using Image Analysis (3D colour histogram). The result shows correct and incorrect images, precision and recall for 10 to 100 images observed.

6.6 Conclusion

In this chapter, four experiments were presented to evaluate the text based, image based and the hybrid based approaches. Three sets of Flickr documents were used in these experiments. CASE 1 represents images with much user generated content and CASE 2 represents images with lack of description. CASE 3 represents images that are tagged with city.

Results for Task 1 have shown that the semantic model generated by text analysis has improved recall and precision value in information retrieval compared to using tag based searching for both CASE 1 and CASE 2. The use of natural language tools has allowed the semantic model to be represented and queried not only with words but also in the form of phrases. The knowledge base and ontologies have enriched the semantic model by providing additional information and thus improved the information retrieval.

In Task 2, the image classifier has performed better than text based search in precision while the text based search has performed better in recall. The use of a Bayesian inference approach in the image classifier has allow the image classifier to identify building images more precisely compared to text based search even though all the images in CASE 3 are tagged with city.

Task 3 shows that integrating text and image based search as a hybrid approach has increased the capability of information retrieval to identify images relevant to the query for both image and text based search. Results for the hybrid approach shows that the two components, which are text based and image based, have complemented each other, thus increasing the chance to find the required images.

In Task 4, the image classifier is modified to classify images beyond city and building landscape. Using results in task 1 as input, the image classifier has found additional requested images in CASE 2 which are not searchable by using the text based method due

to lack of textual description. The results show that the image classifier is extensible and also supports the answer given to the question in Task 3. It is also clear from published work on image classification that more powerful classifier could be developed using more sophisticated image features.

Chapter 7

Concluding Remarks and Future Work

The research in this thesis shows how ideas from two different areas, language processing and computer vision, can be brought together and extended to improve information retrieval from the web. It mainly focuses on the analysis of the content of multimedia documents on Web 2.0, the social web. A hybrid approach is developed which is an integration of natural language processing and image annotation ideas together with semantic web technologies in order to support and enhance information discovery.

7.1 Thesis Summary

Embedding additional information in Web documents is crucial for easy information retrieval. In the early stages of Web development, additional information was embedded in the Web documents by adding META tags. These tags consist of information such as author, keywords and descriptions which are invisible to the users and are usually only used by the information retrieval systems. Tagging in Web 2.0 has transformed this situation by providing visible additional information. Instead of hiding them, tags are shown to users and hyperlink mechanisms are applied on tags to aid navigation around the Web. Currently, images in Web 2.0 documents are described using free text information in title and caption, while tags provide additional information to the images. Nevertheless tagging is limited in its ability to make explicit a lot of the semantics of data ([W3C \(2007\)](#)). Content generated by user comments or views tend to describe the affective side of images, thus only limited information can be extracted to describe the images objectively. The initial focuses of this study were on 1) analysing the structure of the Web 2.0 documents and, 2) analysing the basics of the document content (text and image). This initial study has shown significant problems with describing images in Web 2.0 documents as stated in Section 3.2. In general, these problems are closely

related to the representation of the information (hard coded and lack of semantics) which causes the limitations in searching capabilities. Furthermore, the used of free text descriptions in describing and communicating the content of the documents can only be understood by humans not machines. In order to tackle the looseness in Web 2.0 documents, there is an urgent need to add semantics to generate more information that offers meaningful descriptions related to the images. For the purpose of enhanced information reuse and information retrieval, the information needs to be represented in a more highly structured way by using for example the Semantic Web standards. In order to overcome the problems, several requirements have been identified such as:

- extracting the Web 2.0 resources based on domain of interest to capture the important information that can be used in describing the images.
- mapping the extracted information with domain knowledge to create a conceptual information representation which will provide a more clear description of the images
- extending the extracted information with other more reliable resources such as open generic knowledge bases (Dbpedia), specific domain knowledge bases (Geonames ontologies) and specific authorized webpages
- using the Semantic Web Standards to represent such information to ensure information visibility for future access and reuse.

To initiate the experiments, we created a multimedia corpus using a set of 200 documents that were collected from the Flickr website based on a Tourism and Malaysia query. The multimedia corpus consists of images and text descriptions in the XML format and unstructured text, which are extracted from the Flickr documents. The hybrid approach is implemented in the Image Semantic Information Extraction (ISIE) framework.

7.2 Summary of the Text Analysis

Firstly, to deal with textual description in Web 2.0 documents, a combination of Natural Language Processing and Ontologies were used in information identification and extraction processes. Natural Language Processing tools were used to analyse text descriptions. Ontologies were used to provide a guideline in information representations and to replace the task of the domain expert to provide knowledge to the application. As presented in Chapter 1, the main objective of text analysis is to generate semantic information models to describe the image, and three sub objectives identified:

1. to analyse and extract information in textual content by using Natural Language Processing and Ontologies.

2. to classify and expand information by using ontologies/knowledge base.
3. to model information by using RDF description representation that is tailored with Semantic Web standards.

The text analysis was divided into three experimental tasks to achieve the objectives listed above. The descriptions for each task were given as follows:

- Task 1: Analysing and Extracting Resource

The first objective is achieved in Task 1. In this task, a list of concepts were obtained from the resources that might be meaningful to describe the images. Two NLP tools (APP and GATE) were integrated to assist in identifying concepts. The concepts identifications are done by extracting noun phrases from the APP parse tree output. The main advantage of using APP is, concepts can be extracted in the form of phrases such as *kenagan palace* or *traditional malay dance*. GATE is applying Named Entity Analysis (NEA) to recognize basic concepts such as person, location, address (URL) and organization. In addition, the GATE application is integrated with tourism related information gathered from the Tourism Thesaurus provided by World Tourism Organization (WTO). The thesaurus provides lists of common concepts for attraction, environment, event/activity and temporal/time frame, location (state in Malaysia).

- Task 2: Classifying Concepts based on the domain of interest

The first half of the second objectives is achieved in Task 2. In this task, images that share the same characteristics were classified with concepts ranging from general to specific. The aim of this task is to classify the concepts that have been identified based on tourism interest: attraction, environment, event/activity and temporal/time frame and location. The tourism thesaurus added in GATE might give a clue to answer general or specific information such as object, attraction, temporal, location or event of the image. The identification of specific information related to location is done by matching extracted concepts with the Geonames Ontology. This concept classification provides meaningful information to a concept. For example, an image may be enhanced with additional information for extracted concepts such as *pulau tioman* which is classified as location, island as it's environment, sunrise as it's time frame and finally based on Geonams Ontology, *pulau tioman* is located in the State of Pahang. As a result, having such additional information tagged to these concepts increases the semantic value to the image description thus enabling information retrieval to expand or narrow down the searching to meet the user's query.

- Task 3: Extending / Linking with Other Resource.

Objectives two and three were completed in this task. It is valuable to expand the information extracted with other related Semantic Web resources especially

well known standards such as Dbpedia and the Geonames Ontology to make it shareable. Since well published tourism information for tourism in Malaysia in the form of Semantic Web standards are still not available, a specific domain ontology was developed specifically for this purpose. The Malaysia Tourism Ontology was produced by gathering information from authorized web pages such as Tourism Malaysia and Virtual Malaysia. The extracted information was linked to the other resources by using a standard representation which allows other tools that are used for other resources to access and reuse our information in the future. Finally, all the extracted information is translated into RDF. The RDF representation of the extracted information will provide an extended layer (to the Web 2.0 document) of information representation conforming to Semantic Web standards, which will increase the accessibility for information retrieval mechanisms and to ease information reuse.

7.3 Summary of the Image Analysis

The Image analysis was presented to tackle image based information by developing an image classifier example. In this analysis, image low level features, colour and lines were extracted and analysed by using a Bayesian approach which is powered by the Infer.net tool. The Bayesian approach was used to develop the image classifier which is used to identify images that might represent certain objects or landscapes. The classifier was evaluated by using analyses based on the confusion matrix and ROC curves. For the purpose of experiment, the first instance of the classification was to identify images containing buildings. Two image sets were used, a training set which contains 210 images and a test set which contains 1040 images. Three objectives were identified to develop the image classifier:

1. To find optimum cut off points to define threshold values to classify building images and non building images.
2. To classify images based on the threshold values identified.
3. To evaluate the image classifier in the building and non building domain.

The development of the image classifier was divided into two main tasks as follows:

1. Task 1: Finding the optimal threshold value The first objective was achieved in the first task.

Images in the training set were used and its low level features, which are colour and line, were extracted and represented in the form of histograms. Three experiments were conducted to identify the optimal threshold value for identifying building images by observing each feature separately and also in combination approach.

2. Task 2: Image Classification.

The second and third objectives are achieved in Task 2. Images in the test set were used is similar to task 1, its low level features were extracted and represented in the form of histograms. Several experiments were conducted in two different modes, observing each feature separately and in combination. Based on the experiments, the image classifier is shown to be capable of identifying building images and the combination approach, which is colour and line histogram has produced a better result compared to analysing the features separately.

7.4 Concluding Remarks for The Hybrid Approach

The hybrid approach is evaluated in three search modes which are text based search, image based search and integrated search. In text based search, concepts identified in the semantic model have achieved a better result in information retrieval compared to tags. The results show that the use of natural language processing was able to identify more concepts which might be useful to represent the content of the image while ontologies have allowed the concepts to be expanded thus enriching the information representation. Beyond the capability of tag based search, the semantic model generated by text analysis in the hybrid approach is presented in the form of RDF which allows the search to be done in a generic and specific manner.

The image based search was evaluated by comparing the information retrieval result with text and tag based search. Image and text based search have produced higher results compared to tag based search. The evaluations were done using two sets of documents, the first set contains images with full description and the second set contains images with minimal descriptions. The text based approach has generated better results compared to tags and image based search in the first set while the image based search has surpassed the text and tag based approach in the second set. The result shows the text based approach is best used in well described images, while the image based approach is a promising approach to tackle images with minimal descriptions. We have presented an image classifier to identify images which contain specific object, buildings. Some experiments were also presented to show the potential to expand the image classifier capability in identifying landscapes such as beach and sunset.

The hybrid based approach is evaluated by integrating tag and image based search as one search and image and semantic model based search as another search. The result for both searches show that the later integration has provided better results compared to the first integration. Moreover, the integration approach is better than using the approaches separately. The text and image based es complement each other to handle information shortage in text description, therefore integrating

both approaches is a promising start to handle user generated content which has increased rapidly.

7.5 Future Work

Following the research described in this thesis, a number of improvements are suggested in the next subsections.

7.5.1 Adding multimedia ontology for information representation interoperability

As stated earlier in this thesis, extracted information is presented in the form of a semantic model. The semantic model contains factual information related to the tourism domain which reflects conceptual information indicated by the analysed image. Nevertheless, the semantic model approach is focusing on the conceptual description of the image and not the image itself. Really, a multimedia ontology is required to improve the representation of the semantic model to support image content description in the media-data such as colour, texture and media spatiotemporal structure which can be extracted from within the image. Most importantly, the multimedia ontology is required to support interoperability of the semantic model with different models which are generated by other tools. There are a number of multimedia ontologies available, such as Moving Picture Experts Group (MPEG-7) ([Hunter \(2001\)](#)) and Core Ontology for Multimedia (COMM)([Arndt et al. \(2008\)](#)).

MPEG-7 metadata standard provides a set of elements for describing the semantic content of audiovisual (AV) material. The aim is to standardise quantities measured as AV features and the structure of descriptions and their relationship which can be used to enable fast efficient retrieval from digital archives (pull applications) as well as filtering of streamed audiovisual broadcasts on the Internet. MPEG-7 provides a framework for interoperable multimedia content delivery services and plays an important role in manipulation and management of multimedia content and its metadata. MPEG-7 consists of two main components; (1) a core set of Descriptors (Ds) that can be used to describe the various features of multimedia content, and (2) pre-defined structures of Descriptors and their relationships, called Description Schemes (DSs). Usually, images are tagged by setting:

- the boundary of a region (decomposition of media asset). It provides descriptors for spatial, temporal, spatiotemporal and media source decompositions of multimedia content.
- annotate the content of the region with metadata (semantic annotation of its parts).

The annotation descriptors can be :

- administrative metadata: creation and production
- content-based metadata: audio/visual descriptors
- semantic metadata: interface with domain specific ontologies

Figure 7.1 shows an example to annotate an image by using the MPEG-7 approach. Nevertheless, the extraction and annotation of the content process with the corresponding metadata is out of scope of the standard.

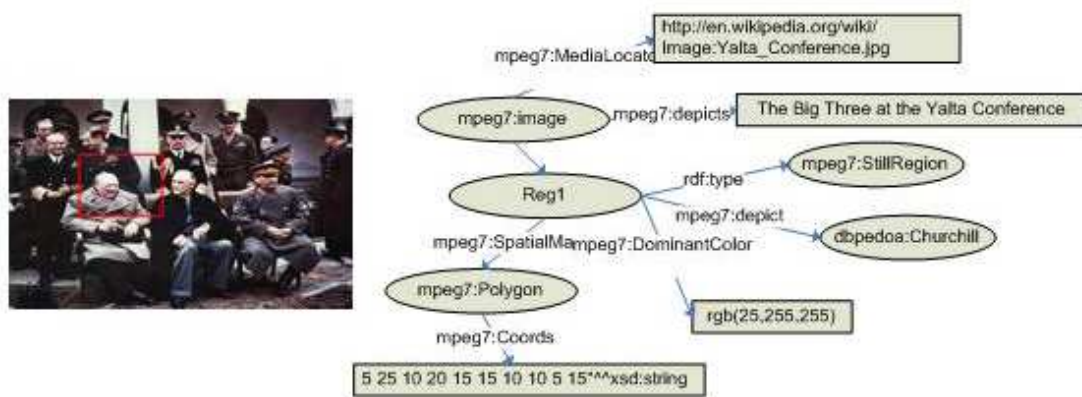


FIGURE 7.1: Annotating a region of an image using MPEG-7 standard

A Core Ontology of Multimedia (COMM) is a domain independent vocabulary that explicitly includes formal foundation categories, such as process or physical objects and eases linkage of domain specific ontologies (Arndt et al. (2008)). COMM is developed based on both the MPEG-7 standard and the Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) foundational ontology. DOLCE provides two design patterns (1) Descriptions and Situations (D&S) and Ontology of Information Objects (OIO). OIO can be used to formalize contextual knowledge, while D&S implements a semiotics model of communication theory. The DOLCE patterns need to be extended for representing MPEG-7 concepts since they are not sufficiently specialized to the domain of multimedia annotation. MPEG-7 provides two important roles in multimedia annotation, (1) the decomposition of a media asset and, (2) the annotation of its parts. The example of segment annotation using COMM is shown in Figure 7.2.

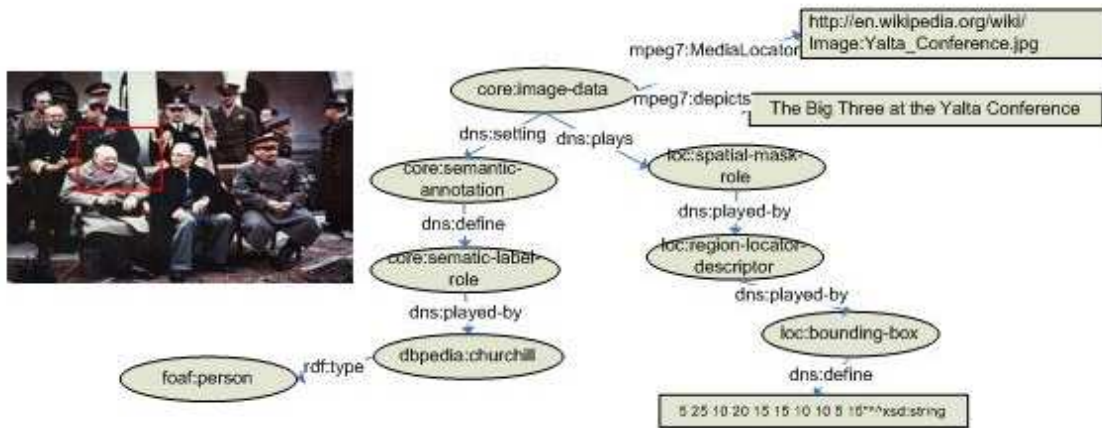


FIGURE 7.2: Segment annotation using COMM standard

7.5.2 Adding emotions towards higher semantic image annotations

In the preliminary experiments, we found that content generated by user comments or views tend to describe the affective words such as beautiful, wonderful and lovely. It would be interesting to study the relationship between images, the emotions they convey and affective words used to describe them. There are some linguistic based researches towards classifying words based on class of emotions such as Nastase et al. (2007) and Lee et al. (2004). Based on a word sounds, Nastase et al. (2007) shows that words expressing the same emotion have more in common with each other than with words expressing other emotions. Lee et al. (2004). have introduced five classes (vowel, stop, glide, nasal and fricative) to classify utterances into four classes which are angry, happy, neutral and other. Current image analysis techniques allows for automatic extraction of low level features such as colour, edge, texture, it is still in its infancy to extract high level features such as object, behaviour and emotions. Emotions are usually described in adjective forms thus user generated content would be a good source to initiate the research.

7.5.3 EXIF metadata

For each image added in Flickr, Exif metadata for the images will be stored automatically into *additional information section*. The EXIF metadata can contain a wide range of information such as date, geo location, technical detail of the camera setting (such as Full Length, Exposure time, Aperature FNumber, FlashUsed, Subject distance, etc), description and copyright information. Even though such information is extractable and was made publically available in Flickr, Flickr only uses Exif metadata to set the date the image was taken and organizing images based on the instrument used to capture the images. The full potential of using

Exif metadata within Flickr to improve information retrieval is yet to be explored. The Exif metadata parameters could be used to identify scene mode such as outdoor, indoor, sunset, portrait, landscape and portrait. In order to classify image scene, camera metadata can be classified into three groups: brightness, subject distance and flash setting (fired or not). Scene brightness includes parameters such as exposure time, aperture number and shutter speed. In general outdoor sun lighting is brighter than indoor artificial light; the exposure time corresponds to shutter speed. At a bright scene, both exposure-time and shutter speed would become shorter but the value is longer for a darker scenes. The subject distance in auto-focus mode could be useful to provide hints for an object oriented scene or not. For example, a longer subject distance value could suggest an image of a landscape while a shorter subject distance could indicate an object oriented scene. The absence or the presence of the flash setting is useful to indicate indoor or outdoor and night and daylight scenes. The flash is usually used during low light such as indoors and at night.

7.5.4 Classification for other Landscapes and other objects

Earlier in Chapter 6, we have provided some examples to expand the image classifier capability in identifying other landscapes such as beaches and sunsets. Increasing the capability of image classifiers would allow the classifier to annotate more unknown images which are unsearchable by the text based search. Moreover, there are also images which capture the same object but with different angle and situations which could be grouped into certain domains of interest. For example, if the image classifier could identify specific images such as *Kuala Lumpur City Center*, unknown images that share similar features with the images will be annotated with specific information thus increasing searching capability to find such images.

7.5.5 Using a Hybrid interface to Support the hybrid approach

Conventionally, an information retrieval interface would only support text based searching. With the hybrid approach, we could introduce a hybrid user interface, which allows queries in the form of text and image. The text based query will support text query input while the image based query will support image query input. If the user does not have keywords to start a search, alternatively, an image could be used as an input for search. Integrating text and image based search in the information retrieval interface would require relevance feedback and technical supports. The relevance feedback is required in searching refinement which would allow the system to re-analyse results based on user feedback, while the technical support is required to increase the analysis capability which is normally time consuming.

As a conclusion, we have shown that this hybrid approach provides improved information retrieval performances compared to either of the contributing techniques used separately. During this research, we have encountered other interesting topics to be considered in the future work. We believe, the future work extensions could improve the capability of the hybrid approach towards a better retrieval.

Appendix A

Preliminary Analysis Documents Examples

Travel: Document T1



FIGURE A.1: Document T1: *Through the mist.*

Title: Through the Mist

Description: Sun sets along the Middle Geyser Basin in Yellowstone National Park while a fellow tourist takes a moment to contemplate the area.

Tags: people, yellowstone, mist, fog, geysers, life, travel, camping, wyoming, national parks, yellowstone national park, moments, summer, interestingness, interesting, t10, BRAVOA, BigFave, been@1of100, usa, fivestarsgallery

Comments:

-great!!!

-Fantastic shot! (1-2-3 Nature)

-sweet. great capture!! – Seen in 1-2-3 Nature (1 SHOT/DAY!!!READ THE RULES) (?)

-Another truly awesome pic. A Big Fave Please add this to www.flickr.com/groups/bigfave

-Awesome sunset. The bridge is perfect and i really love the mist! Great compo-

sition!

-wow... what a shot. Love the smoky effect. *seen in contacts pool*

-This is a remarkable photo! Well done!

-Gorgeous color!

-very nice!

.....

Simple text analysis result: word/rootword (frequency) fave/fave (20), faved/fave (20), shot/shot (19), great/great (17), wow/wow (16), love/love (14), loved/love (14), karma/karma (10), big/big (9), wonderful/wonder (9), explore/explore (9), awesome/awesome (8), mist/mist (7), fantastic/fantast (7), saw/saw (7), post/post (7), amazing/amaz (7), posting/post (7), favorite/favorite (6), day/day (6), composition/composit (6), color/color (6), nice/nice (6), colors/color (6), interestingness/interesting (6), days/day (6), travel/travel (6), favorites/favorite (6), world/world (5), capture/capture (5), please/please (5), art/art (5), thanks/thank (5), beautiful/beauti (5), picture/picture (5), thank/thank (5), photography/photography (4), best/best (4), yellowstone/yellowstone (4), nature/nature (4), add/add (4), gorgeous/gorgeou (4), incredible/incred (4), dylela: feel/feel (4), stunning/stun (4), lighting/light (4), interesting/interest (4), brilliant/brilliant (4), bravo/bravo (4), feeling/feel (4), light/light (4),...

Travel: Document T3



FIGURE A.2: Document T3: *Roman Masterpiece*.

Title: Roman Masterpiece

Description: The Pont du Gard (Roman Aqueduct), in France, was built shortly before the Christian era to allow the aqueduct of Nmes (which is almost 50 km long) to cross the Gard river. The Roman architects and hydraulic engineers who designed this bridge, which stands almost 50 m high and is on three levels the longest measuring 275 m created a technical as well as an artistic masterpiece.

Tags: Gaetan Bourque, Pont, Gard, Bridge, France, Travel, wow, favme, wonder, Imapix, Favpix, BRAVO, Voyage, Europe, Trip, blue, TopFavPix, Colors, Gatan

Bourque, Copyright 2006 Gatan Bourque. All rights reserved, specobject, france Tourism, Most interesting, FrHwoFavs, France Landscapes

Comments:

- Now I have to put France on my travel list. Your pictures are too inspiring.
- Very good shot, beautiful angle and a perfect weather, I was not so lucky when I visited Provence ... :)
- Beautiful view of the blue sky through the arches!
- Gorgeous!
- Just beautiful!!!
- the tree, the light, the clouds - it's just right.
- I've just found your photos and I hardly know where to begin - everything here is so very beautiful and gives me the feeling you get when you reach the top of the hill and see the world unrolled around you. It's all just wonderful, including this pic.
- Great, taken in full beauty. I love it and your photo.

Simple text analysis result: word/rootword (frequency) shot/shot (25), shots/shot (25), great/great (21), beautiful/beauti (21), faves/fave (14), roman/roman (13), ponts/pont (13), gards/gard (15), du/du (11), wonderful/wonder (11), wonders/wonder (11), wonder/wonder (11), aqueduct/aqueduct (10), aqueduct-s/aqueduct (9), well/well (9), nice/nice (9), france/france (7), amazing/amaz (7), captured/captur (7), masterpiece/masterpiece (6), favorite/favorite (6), bridge/bridge (6), wow/wow (6), view/view (6), blue/blue (6), top/top (6), love/love (6), color/color (6), bridges/bridge (6), bravo/bravo (6), colors/color (6), part/part (6), views/view (6), favorites/favorite (6), viewed/view (6), engineers/engine (5), sky/sky (5), ve/ve (5), pic/pic (5), build/build (5), building/build (5), engineering/engine (5), work/work (5), terrific/terrific (5), pl

Travel: Document T17



FIGURE A.3: Document T17: *Kids having a splash behind Taj.*

Title: Kids Having Splash Behind Taj

Description: They were beating the heat at Yamuna. This is one of my favourite

shots. I like it very much, in spite of poor contrast, it was a dull day and almost against light.

Tags: India, boy, funsafety, shore, coast, banks, young, small, kid, boat, wooden, oldtimes, asia, taj, monument, mahal, mughal, reflection, yamuna, river, mausuleum, famous, famed, building, architecture, moghul, dome, spiral, tower, towers, pollution, people, indian, agra, tour, tourism, holiday, tourist, travel, muslim, religion, religious, hindu, hinduism, tomb, sevenwonders, UNESCO, wonder, beauty, shah-jahan, mumtaz, mumtaj, marble, art, carving, memory, landscape, culture, heritage, water, symbol, attraction, spectacular, outside, exterior, carvings, sculpture, sculpted, pillars, burial, destination, grave, historical, social, 17thcentury, islam, minarette, inspiration, icon, landmark, mausoleum, white, site, boys

Comments:

- very cool. Free india
- very nice shot from the director of SPLASH!
- Fantastic shot, my friend, I love it!
- Very good capture.... great timing...
- Indeed, an outstanding shot, my captain. :) The freedom is on the air!
- Its wonderful. Fun shot, Doug
- Great view!
- This is a beautiful shot, I love it.
- Great capture! The kids running happy, their reflection, the Taj in the background. A powerful photo!
- wonderful scene, my friend!!!! excellent captured!!! have a great day and thanks for your comments!!!
-

Simple text analysis result: word/rootword (frequency) shots/shot (30), shot/shot (30), great/great (24), f :18 love/love (18), taj/taj (16), travel/travel (14), traveller/travel (14), india/india (13), capture/capture (13), beautiful/beauti (11), world/world (10), nice/nice (9), day/day (8), good/good (8), view/view (8), mahal/mahal (8), viewed/view (8), wonderful/wonder (7), thanks/thank (7), action/action (7), amazing/amaz (7), thank/thank (7), please/please (7), called/call (7), wonders/wonder (7), wonder/wonder (7), call/call (7), photography/photography (6), reflection/reflect (6), excellent/excel (6), reflections/reflect (6), see/see (6), adding/ad (6), post/post (6), seeing/see (6), admin/admin (6), added/ad (6), people/people (6), kids/kid (5), friends/friend (5), cool/cool (5), friend/friend (5), look/look (5), eyes/eye (5), wow/wow (5), life/life (5), eye/eye (5), looking/look (5), hi/hi (5), kid/kid (5), sharing/share (4), digital/digit (4), rights/right (4), fantastic/fantast (4), fun/fun (4), better/better (4), suresh/suresh (4), limitation-s/limit (4), not/not (4), shared/share (4), right/right (4), moment/moment (4), limits/limit (4),...

Appendix B

Stop Words

a	around	description	give	inc	mm
about	as	did	go	incl	mo
above	assume	discover	gone	indeed	more
abs	at	dl	got	into	moreover
accordingly	be	do	gov	investigate	most
across	became	does	had	is	mostly
after	because	done	has	it	mr
afterwards	become	due	have	its	much
again	becomes	during	having	itself	mug
against	becoming	each	he	just	must
all	been	ec	hence	keywords	my
almost	before	ed	her	keep	myself
alone	beforehand	effected	here	kept	namely
along	being	eg	hereafter	kg	nearly
already	below	either	hereby	km	necessarily
also	beside	else	herein	last	neither
although	besides	elsewhere	hereupon	latter	never
always	between	enough	hers	latterly	nevertheless
am	beyond	especially	herself	lb	next
among	both	et	him	ld	no
amongst	but	etc	himself	letter	nobody
an	by	ever	his	like	noone
analyze	came	every	how	ltd	nor
and	can	everyone	however	made	normally
another	cannot	everything	hr	mainly	nos
any	cc	everywhere	i	make	noted
anyhow	cm	except	ie	many	nothing
anyone	come	find	if	may	now
anything	compare	for	ii	me	nowhere
anywhere	could	found	iii	meanwhile	obtained
applicable	de	from	immediately	meta	of
apply	dealing	further	importance	mg	off
are	department	gave	important	might	often
arise	depend	get	in	ml	on

Continued: Stop Words

one	recently	soon	throughout	where
only	refs	specifically	thru	whereafter
onto	regarding	still	thus	whereas
other	relate	strongly	to	whereby
others	said	studied	together	wherein
otherwise	same	sub	too	whereupon
ought	seem	substantially	toward	wherever
our	seemed	such	towards	whether
ours	seeming	sufficiently	try	which
ourselves	seems	take	type	while
out	seen	tell	ug	whither
over	set	th	under	who
overall	seriously	than	unless	whoever
owing	several	that	until	whom
own	shall	the	up	whose
or	she	their	upon	why
oz	should	theirs	us	will
particularly	show	them	use	with
per	showed	themselves	used	within
perhaps	shown	then	usefully	without
pm	shown	thence	usefulness	wk
precede	shows	there	using	would
predominantly	significantly	thereafter	usually	wt
present	since	thereby	various	yet
presently	slightly	therefore	very	you
previously	so	therein	via	your
primarily	some	thereupon	was	yours
promptly	somehow	these	we	yourself
pt	someone	they	were	yourselves
quickly	something	this	what	yr
quite	sometime	thorough	whatever	
rather	sometimes	those	when	
readily	somewhat	though	whence	
really	somewhere	through	whenever	

Flickr related words. Flickr related words are added to the stop words list to filter out common words used in describing images especially in tags. These words often appear in tags list and do not carry semantic meaning.

comment	hey	way
comments	gd	pic
commented	tag	wins
photo	great	win
group	fave	shot
pool	friends	picture
flickr	friend	share
www	nice	maybe
com	nearby	subject
title	ahh	doubt
titles	inside	enter
says	outside	bet
image	add	care
document	remove	official
photo	nie	best
uploader	thanks	super candid photographer
information	thank	candid human expression award
classification	1	tho
uploaded	2	ha
additional	3	typ
tags	4	thanks
list	5	looky
taken	6	gota
view	7	abigfave
wow	8	dex
ing	9	wowiekazowie
oh	invite	lot
admin	invites	len
tagline	heart	max
please	hearts	dont
superb	top	oe
start	term	sis
end	terms	sic
post	well	

Appendix C

Tourism Thesaurus

Environment

beach	road
beaches	rocky beach
beachside	rural
beauty spot	rural town
big city	seafront
center	seafront area
city	ski
city centre	skiing area
coast	spa town
coastal town	tranquil at night
countryside	tranquil, day and night
cross-country skiing	upland area
edge of forest	valley
forest	village
gorge	volcano
high mountains	woods
hinterland	
historic district	
hunting reserve	
island	
islands	
lakeside	
lowland area	
mountain	
motorway campsite	
mountain town	
mountains	
nature reserve	
outskirts	
resort town	
riverside	

Event

acrobatics	brass bands	daily market	flower market
acting	bullfight	dance	flowers festival
adventure	bullfighting	dance festival	flying
adventure tour	business	dance music	folk
aerial tour	cabaret	dancing	folk dance
aerobics	caf concert	decorative	folk music
agriculture	caf theatre	designing	football
agriculture fair	canoeing	do-it-yourself	fundraise
agriculture open farm	batik canting	drawing	gambling
air	card games	eco tour	gardening
air festival	carnival	education	gastronomic festival
annual festival	cattledriving	electronic music	glider
archaeology	caving	equine	gliding
archery	celebration	ethnic dance	go-kart racing
astronomy	ceremony	exhibition	gold panning
athletics	chart music	expo	golf
athletics	children's games	fair	gondola tour
auction	choral	fair market	gourmet tour
badminton	choral music	fair trade	graphic
ballet	circus	fashion	guided tour
ballooning	classical music	fencing	gymnastics
baseball	clay-pigeon shooting	festival	gymnastics
basketball	collecting	festival arts	handball
bathing	competition	festival drama	handicraft market
bicycle touring	computers	fete festival	harpooning
bingo	concert	film	health spa
bird watching	country music	film cartoons	heritage railway
bivouacing	courses	film documentaries	hiking
boating	cricket	film festival	hockey
bobsleigh	croquet	fireworks	horse-drawn holidays
body building	crosswords	fish market	hunting
books fair	cultural	fishing	ice hockey
botany	cultural reading	fjord tour	ice skating

Continued: Event

bowling	wrestling	flea market
boxing	cycling	floor show
jazz music	open day	religious music
jogging	opera	riding
judo	operetta	rock climbing
karate	organised tour	roller-skating
kayaking	paintball	rowing
kiting	painting	rugby
land sailing	painting	running
landsailing	parachuting	safari tour
lecture	paragliding	sailing
listening to music	parascending	sailing
literature	performance	scale modelling
long-distance riding	photo festival	scuba diving
market	photo safaris	sculpture
mime	photographic	sculpture
mineralogy	photography	sea fishing
motocross	picnicking	seasonal festival
motor tour	poetry	shooting
motorcycle touring	polo	shopping
motorsport	pop/rock	sing along
motorsport	popular festival	singalong
motorsport	printing	singing
mountain	public holiday	skateboarding
mountain bike	puppet theatre	soccer
mountainbike	radio	soccer
mountaineering	rafting	speaking
music	rambling	speaking foreign languages
music	rap/hiphop	special events
music festival	reading	sport
music flamenco	reggae	squash
national festival	religion	street market
nature	religion open air service	stretching
newage music	religion parade	surfing
nightlife	religion pilgrimage	swimming
nostalgia tour	religion service	snorkeling
open air concert	religion spectacle	snorkelling

Continued: Event

indie music
international festival
skydiving
table tennis
technology fair
television
tennis
tennis
theatre
thematic fair
tour
tour group
tourism fair
traditional celebration
traditional dancing
traditional festival
traditional singing
trekking
trolley driving
ultralight
video games
volcanology
volleyball
walking
walking tour
water polo
water skiing
water festival
watersports
weekly market
wilderness
window shopping
wine festival
wine tour
world aids day

Attraction

art	city	minaret
arts	town excavations	monastery
craft	cloister	museum
handcraft	collegiate church	monastery convent
architecture	convent	monolith
acropolis	crafts Show	monument
amphitheatre	cultural Interest Property	monumental historic site
ancestral home	diocesan Museum	monumental site
apartment building	dolmen	mosque
aqueduct	event center	municipal museum
arch	excavation Site	museum
archaeological site	farmhouse	museum collection
basilica	fort	national heritage
basilica	fortified farmhouse	national historic
boat trips	fortress	artistic monument
bridge	fortress	national monument
building	fountain	national museum
buildings	funfair	national park
burial mound	galician country house	nature preserve
castle	gallery	nature reserve
castles	garden	nature trail
castle ruins	gate	necropolis
catacombs	gorge/ravine	nostalgic railroad
cathedral	great thematic park	outings
cathedral museum	guided tours	day trips
cave	hermitage	palace
grotto	Hill-fort	pantheon
chapel	historic building	pedestrian precinct
chapels	island museum	place of pilgrimage
church	local museum	Planetarium

Continued: Attraction

prehistoric cave	wine road
prehistoric shelter	world heritage
private puseum	zoological garden
procathedral	game reserve
provincial museum	zoo
public building	zoos
pyramid	theme parks
regional museum	former palace
reservoir	jetty
sanctuary	churhes
scenic drive	circus
series of sculptures	citadel
show kitchen	menhir
show mine	military building
small palace	mill
sports facilities	partico
synagogue	observatory
taules,	
talaiots	
navetas	
temple	
temple-fortress	
theatre	
theme Road	
thermal baths	
tomb	
tower	
university	
vantage point	
vintners lane	
walled enclosure	
walls	
waterfall	

Appendix D

Text Analysis Examples

Document ID: 399296817



FIGURE D.1: Image ID: 399296817.

Text Description: Title: cotton sunset

Url: <http://www.flickr.com/photos/dcsuave/399296817>

Description: The spectacular sunset over the jetty was a sight to behold. Pulau Kapas or Cotton Island inherited its name from the native because of its incomparable white beaches. Surrounded by crystal clear ocean, Pulau Kapas promises a spectacular getaway from the hustle and bustle of city life to quiet natural retreats with abundant sunshine and crisp clean air. An island renowned for its clear waters, sandy white beaches and swaying palms, it is relatively isolated. Home to an infinite variety of hard and soft corals, the waters around the island abound with sea-shells, fish and turtles. The island's laid back atmosphere is ideal for relaxation but the more adventurous will find it is also a haven for swimming, snorkeling, windsurfing, kayaking, boating and fishing Location: Kapas Island, Terranganu, Malaysia. Tokina 12-24

Tags: island, kapas, malaysia, sunset, jetty, AnAwesomeShot

Comments: - Rated 9/5 by the Rate My: Nature group - Spectacular view. You are invited to add this image to An Awesome Shot! Please tag the photo with

.AnAwesomeShot. - Rated 5/5 by the Rate My: Nature group. I want to go there!
- Rated 6/5 by the Rate My: Nature group. Seen in Rate My: Nature - Rated 5/5
by the Rate My: Nature group this is too perfect!! i feel like deleting my photo...
:) - Rated 5/5 by the Rate My: Nature group - Rated 4/5 by the Rate My: Nature
group - ... nice atmosphere ... :-) - beautiful place. :D - absolutely stunning!!!!
Seen in Shooter of the day pool. - Superb Love this shot! Please Share your best
pictures of the world with us. What a Wonderful World Group. - Your work has
inspired me!! You are INVITED to post it in.....

DOCUMENT ID: 1203148615



FIGURE D.2: Image ID: 399296817.

Text Description

Title: MALAYSIA

Captions: This former palace of the Sultan of Perak was built without the use of a single nail. It is now a museum. It's on EXPLORE Aug. 29.

Tags : malaysia, kualakangsar, perak, travel, tourism, perspective, asia, impressed-beauty,

Comments: -This Great Photographic Art was made by a Diamond Class Photographer! -Please add your photo to Flickr Diamond The Diamond Class Photographer -Read the group rules please and tag your photo DiamondClassPhotographer -You may tag your photo again as flickrdiamond -Really stunning, just so awesome. XXX -wow, so beautiful, -Nice change of scenery jingle526 :) This is an amazing looking building and couldn't believe it was built without a single nail! Beautiful capture! -Your fantastic color picture is my winner! -Please add this photo to -www.flickr.com/groups/colorphotoaward/ -its in kuala kangsar... -Beautiful -beautiful -Wonderful -Seen in my contacts' photos. (?) -This is an invitation to add your Ultimate shot at - POPs Gallery for Ultimate Shots by Invitation only - Please consider adding your photo to our Gallery. www.flickr.com/groups/popsgallery/ Please remember to tag this photo POPsGallery - Thank you Ersama for the comments and award. - Thank you @schnetler75. - Thank you @kimtojin for the compliments and award. There was no plan also. - Terima Kasih Raheem, it is on the Tag. - Thank you Diego. - Thank you Dear Anne. It is beautiful and being kept like when it was built. - wow, great architecture. I love the color and the window shapes - Beautiful architecture, nicely framed. - Thank you my friends for the faves and comments. - @Jeanie - @Popotito and for the invite. - @Dear LenSOP. - I SNIFFED out this wonderful image! - You deserve this nose worthy award! - Please ADD your - Impressively beautiful photo to - Impressed by your Beauty! (Invite only images) - Please tag your photos ImpressedBeauty - nice and interesting capture...like to see Malaysia, too .) - Hi, I'm an admin for a group

called The Coolest Damn Cool Photographers In The World!, and we'd love to have your photo added to the group. - This is beautiful!! - What a wonderful place and what a wonderful shot.. - wow beautiful colors and place - Congratulations! You have been awarded the - Daark Goblet Award - You are invited to post this photo at - I AM - an Amazing Amateur Photographer - Please Tag Image AmazingAmateur - what a beautiful please! i wish i could visit someday! - Thank you my dear friends for the fave, comments and awards and invitations. - @Karen - @Gudi, you will enjoy the weather in Malaysia and their beaches. - @HakanGil, I see a lot of Turkish visitors here. - @islanddreams132 - @Kim, nice to see you back. You should try to visit Malaysia then you can stop over in Thailand which is just next door. -Wow Jingle....so beautiful building and ambient...If you have more photos post on flickr - Wow, that's a really beautiful building. And no nails? That's amazing. I wonder how it all fits together. I agree with venkane. If you have more photos of Malaysia, please share them. They are so interesting! I like seeing how things look in different places of the world.....

Appendix E

Malaysia Tourism Ontology

Malaysia Tourism Ontology (MTO)

Description for Malaysian Tourism Ontology: The Malaysia Tourism Ontology is a specific domain ontology, which is created to store information related to the tourism in Malaysia. The ontology was created based on the Harmonise Project. It consists of two main roots which are Attraction and Event. The Attraction and Event instances are added from information gathered based on the Ministry of Malaysia Tourism Portal. Figure E.1 shows an overview for the Malaysia Tourism Ontology.

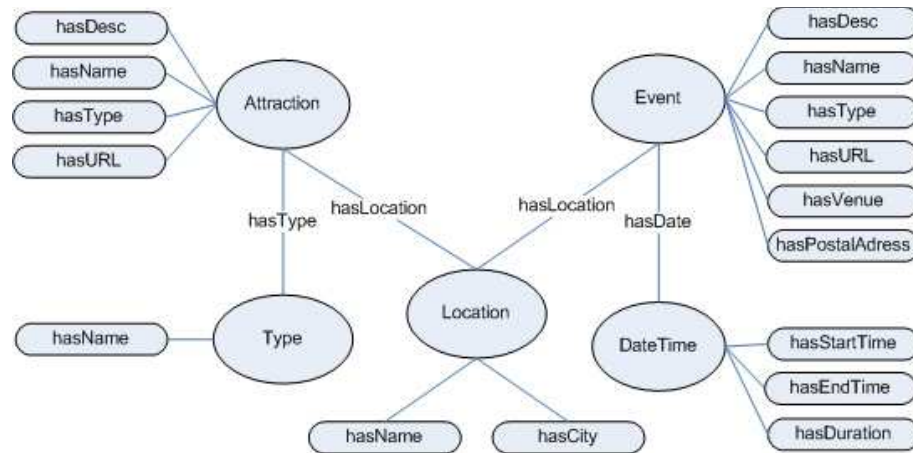


FIGURE E.1: An Overview of Relationships between classes and Properties in the Malaysia Tourism Ontology .

EVENT CLASS

Description for the Event Class:

Event class is used to store information related to specific events or activities held in Malaysia. Event class may contain information such as name, venue and date. Such information is stored in the event properties as presented in Table E.1. Figure E.2 shows protege screenshot for adding event value in the ontology.

Property	Description
hasDesc	Provides general information related to the event instance. The description is captured from authorized Malaysian Tourism web pages or Wikipedia resources.
hasName	Provides alternative names for the instances. In many cases, instances may have more than one names or labels due to dialect or language differences and also abbreviations used to describe the instances.
hasType	Shows list of attraction type for the instances.
hasURL	Shows a list of URL resources from authorized portal, Wikipedia and also Geonames webpage.
hasDate	Linked to DateTime Class
hasVenue	Shows a list of location where the event is celebrated for example through out Malaysia (if the event is celebrated country based event), list of states (if the event is celebrated in certain state or region), specific location such as park or hotel.
hasPostalAddress	Shows state location in the form of geonames url for state based event.

TABLE E.1: List of Event properties and its descriptions

ATTRACTION CLASS

Description for Attraction Class:

Attraction class is used to store information related to specific attractions in Malaysia. Attraction class may contain information such as name, location and list of attraction. Such information is stored in event properties as presented in Table 2. Figure 2 shows protege screenshot for adding attraction value in the ontology.

Property	Description
hasDesc	Provides general information related to the attraction instance. The description is captured from authorized Malaysian Tourism web pages or Wikipedia resources.
hasName	Provides alternative names for the instances. In many cases, instances may have more than one names or labels due to dialect or language differences and also abbreviations used to describe the instances.
hasType	Shows list of attraction type for the instances.
hasURL	Shows a list of URL resources from authorized portal, Wikipedia and also Geonames webpage.
hasLocation	Provides location information.

TABLE E.2: Table 2: List of Attraction properties and its description

DATE CLASS

Description for DateTime Class :

DateTime class is used to store information relate to temporal information. DateTime class may contain information such as time duration and specific date. Specifically, this class is related to Event. Table 3 shows information stored in the DateTime properties.

Property	Description
hasStartDate	Provides start date for an event instance
hasEndDate	Provides end date for an event instance
hasDuration	Provides length for an event instance

TABLE E.3: List of Datetime properties and its descriptions

LOCATION CLASS

Description for Location Class::

Location class is used to store information related to state of Malaysia. Each state is stored in location instance which contain information such as state name and capital city for each state. Location properties are presented in Table 4.

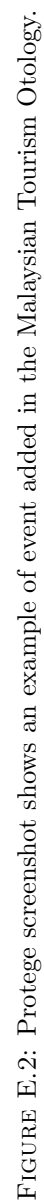
Property	Description
hasName	Provides a list of state names. Each state may contain more than one name due to language differences and abbreviation.
hasCity	Provides a capital city for each state

TABLE E.4: list of Location properties and its description

TYPE CLASS

Description for Type Class :

Type class is used to store information related to attraction type. Specifically, this class is linked to attraction class. The type class only contain one property which is `hasName`. The `hasName` allows can be used to store alternative names for attraction type. For example, *beach* is consists of other names such as *beaches*, *white sandy beach* and *rocky beach*.



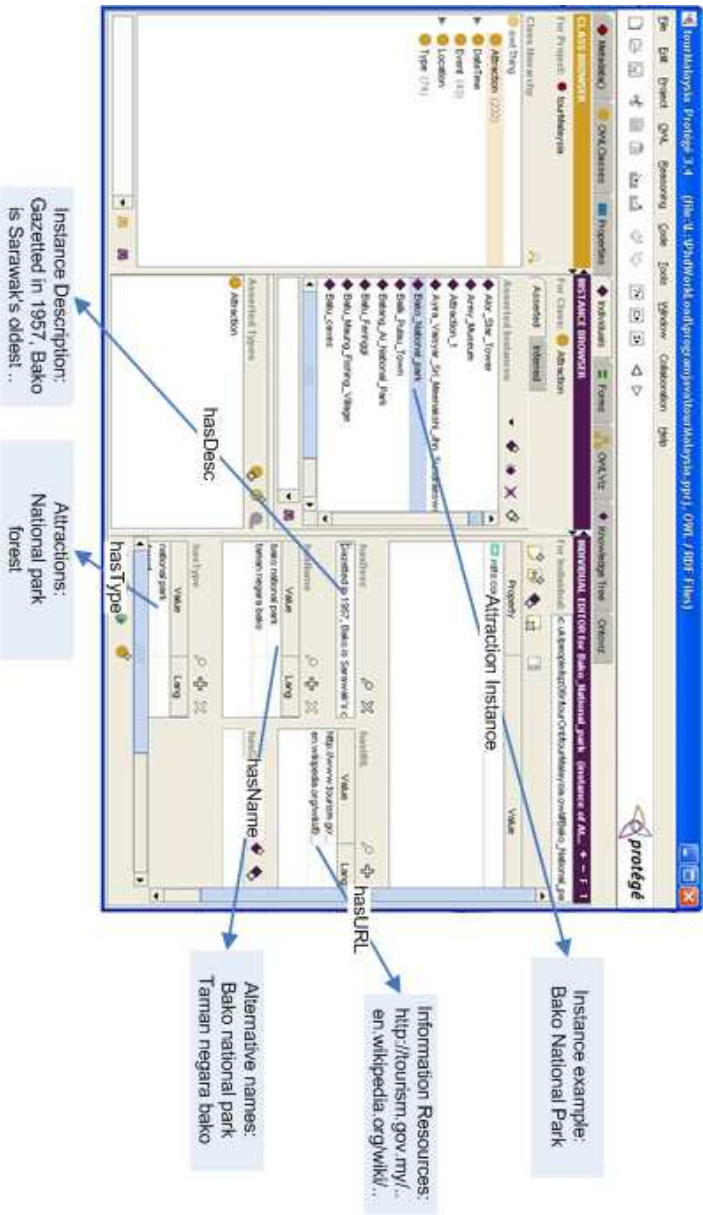


FIGURE E.3: Figure 2: Protege screenshot shows an example of attraction added in the Malaysian Tourism Otology.

Appendix F

Text Analyser Interface

The text analyser will assist user in analysing text descriptions. The text analyser consists of two main components which are a natural language analysis and a knowledge base. Figure [F.1](#) shows the text analyser interface. To start the analysis, a list of images will be presented to users to be selected. The selected image will be displayed along with its text descriptions. The text analysis consists of two components which are a natural language processing and a knowledge base. Each component is performed by clicking the "NLP Analyser" and "Knowledge Base" buttons, and results for each component is presented in Figure [F.2](#) and Figure [F.3](#) respectively

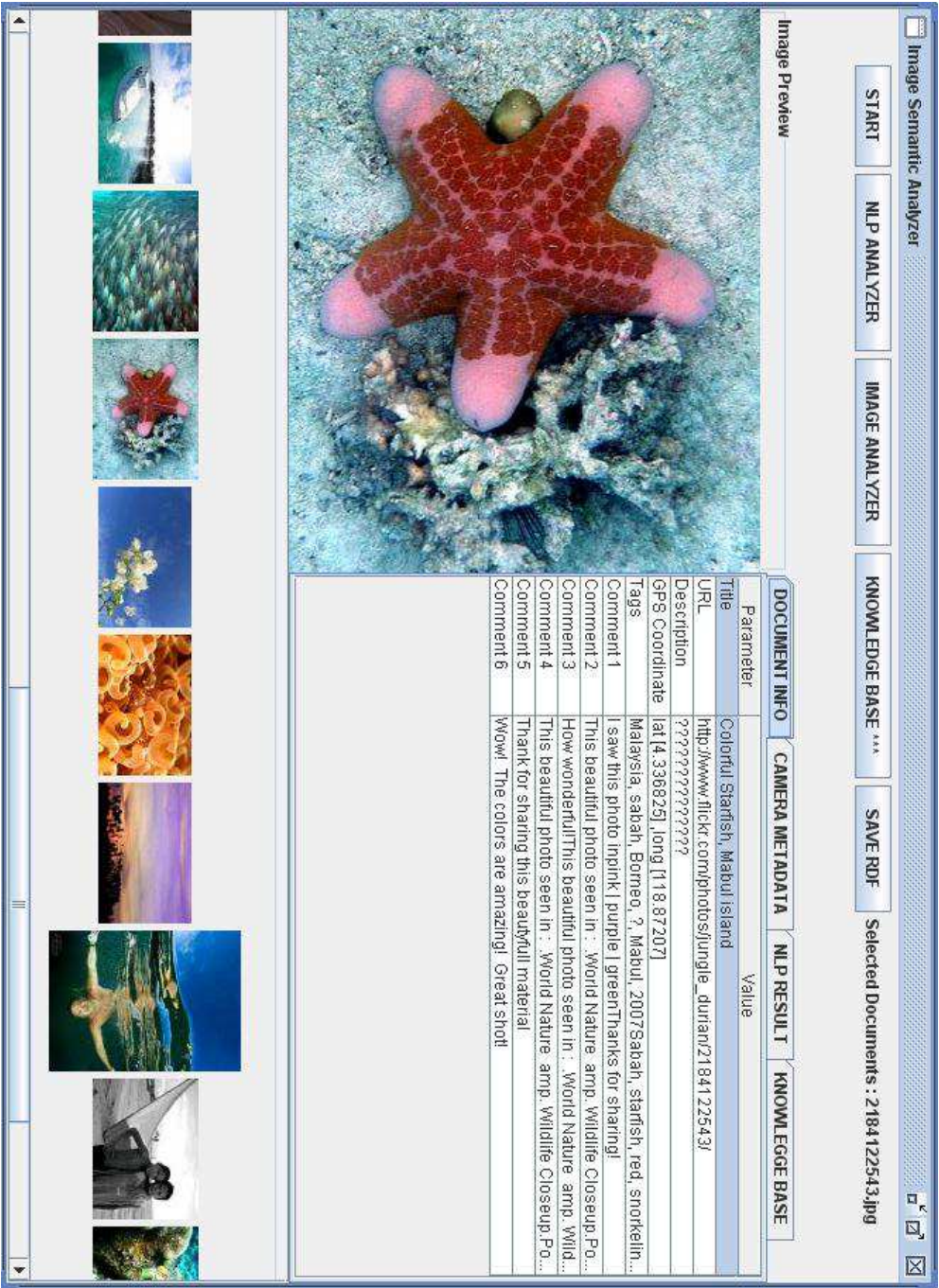


FIGURE F.1: Text Analyser Interface screenshot.

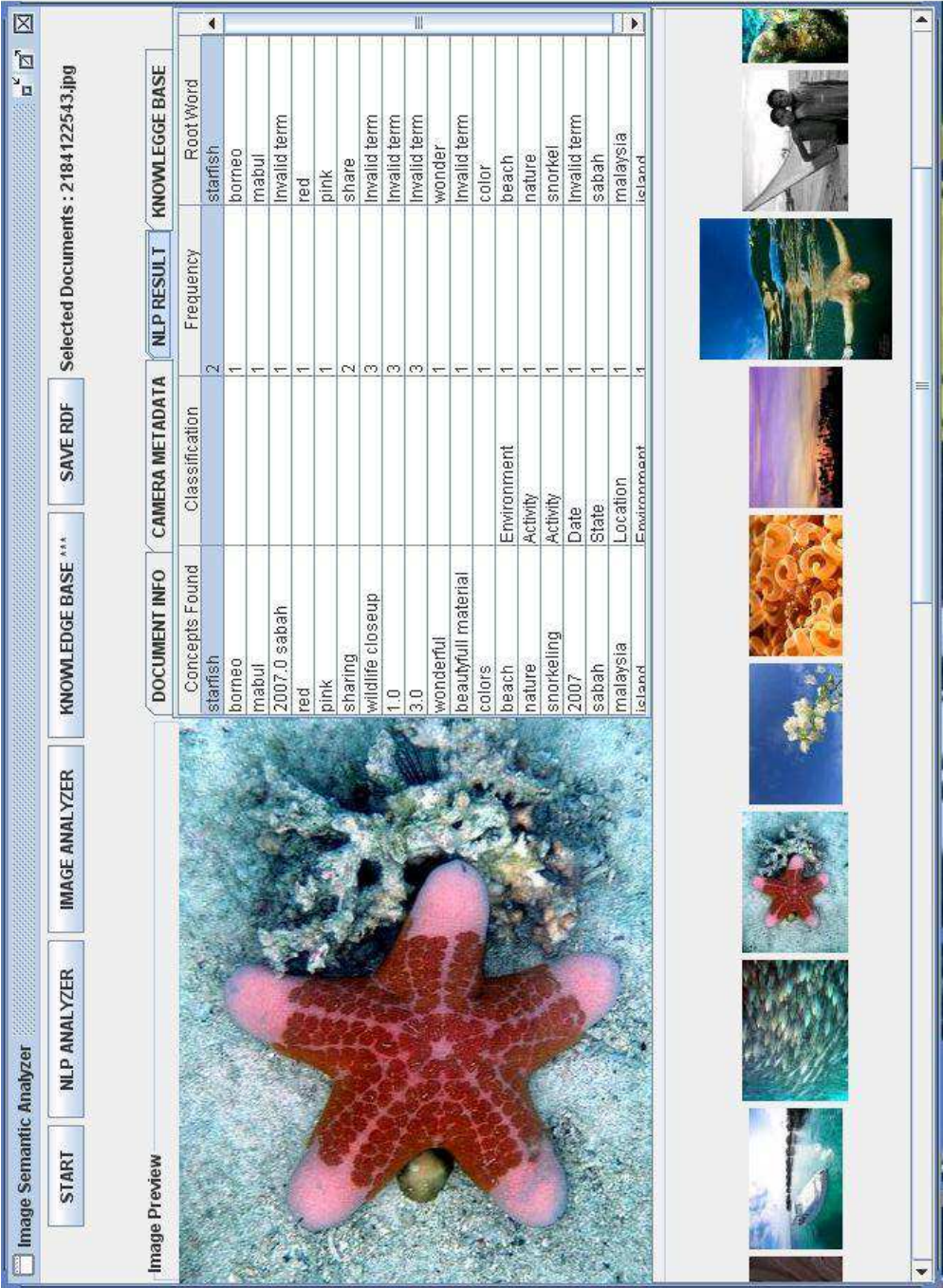


FIGURE F.2: The Natural Language Processing Component Output.

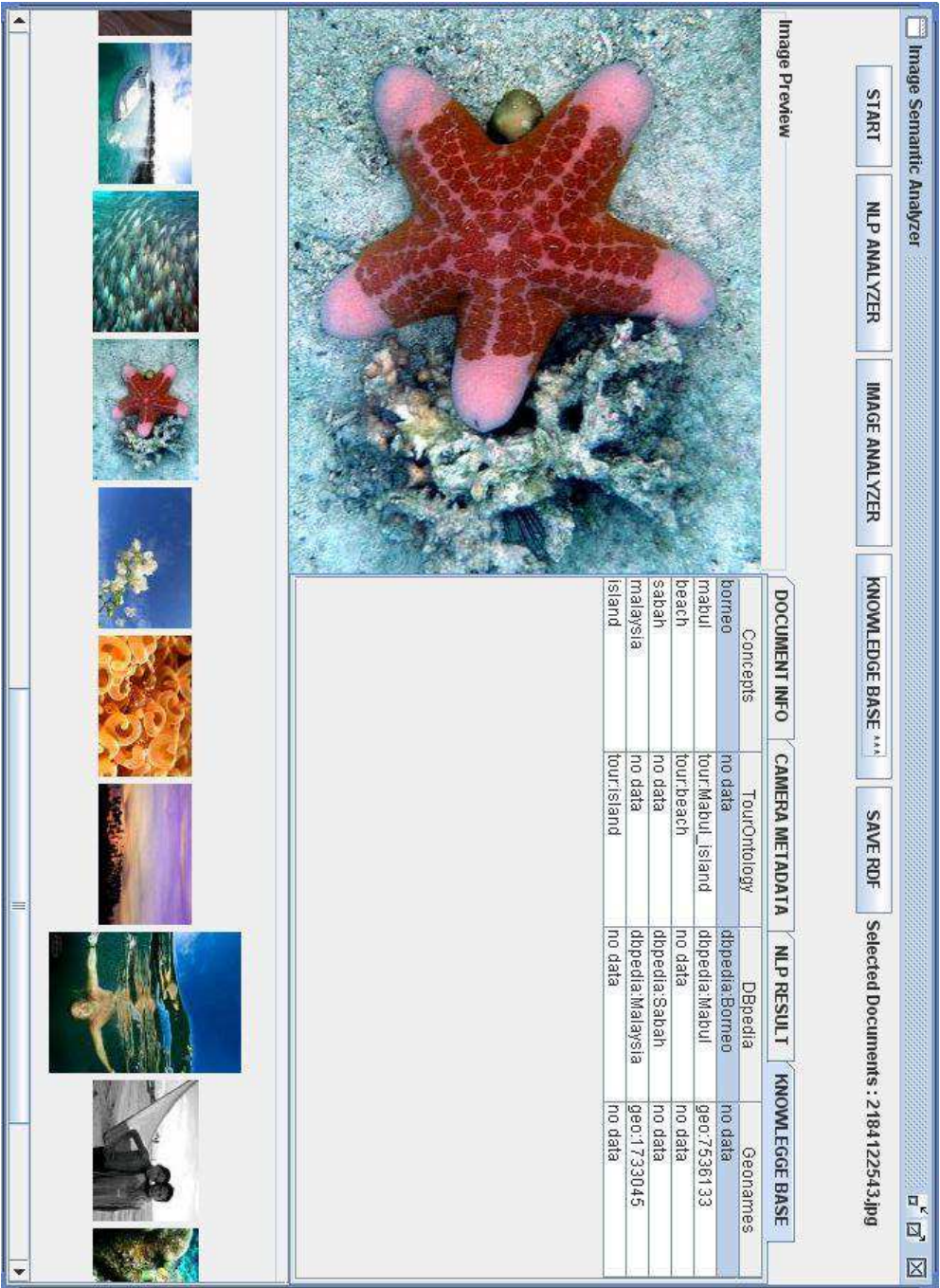


FIGURE F.3: The Knowledge Base Component Output.

Bibliography

- M. Abdel-Mottaleb, Wu H.-L., and N. Dimitrova. *Aspects of multimedia retrieval*. *Philips Journal of Research*, 50(1-2):227 – 251, 1996. ISSN 0165-5817.
- Hend S. Al-Khalifa and Hugh C. Davis. *Fasta: A folksonomy-based automatic metadata generator*. In *EC-TEL 2007 - Second European Conference on Technology Enhanced Learning*. Springer Verlag, 2007.
- Harith Alani, Sanghee Kim, David E. Millard, Mark J. Weal, Wendy Hall, Paul H. Lewis, and Nigel R. Shadbolt. Automatic ontology-based knowledge extraction from web documents. *IEEE Intelligent Systems*, 18(1):14–21, 2003. ISSN 1541-1672.
- H. P. Alesso and C. F. Smith. *Thinking on the Web: Berners-Lee, Godel and Turing*. Wiley-Interscience, New York, NY, USA, 2008. ISBN 0471768669.
- M. Ames and M. Naaman. Why we tag: motivations for annotation in mobile and online media. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 971–980, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-593-9.
- Ashton Anderson, Karthik Raghunathan, and Adam Vogel. Tagez:flickr tag recommendation. 2008.
- Sofia Angeletou, Marta Sabou, and Enrico Motta. *Semantically enriching folksonomies with flor*. In *In Proc of the 5th ESWC. workshop: Collective Intelligence amp; the Semantic Web*, 2008.
- R. Arndt, Troncy R., S. Staab, and L. Hardman. *Adding formal semantics to mpeg-7: Designing a well-founded multimedia ontology for the web*. Technical Report 4, Department of Computer Science, University of Koblenz, Universitätsstraße 1, 56070 Koblenz, 2 2007.
- R. Arndt, R. Troncy, S. Staab, L. Hardman, and M. Vacura. *Comm: Designing a well-founded multimedia ontology for the web*. In *ISWC 2007 + ASWC 2007*, pages 30–43, 2008. ISBN 978-3-540-76297-3.

- Paul Over George Awad, R. Travis Rose, Jonathan G. Fiscus, Wessel Kraaij, and Alan F. Smeaton. Trecvid 2008 - goals, tasks, data, evaluation mechanisms and metrics. In *TRECVID*, 2008.
- David Bawden and Lyn Robinson. The dark side of information: overload, anxiety and other paradoxes and pathologies. *J. Inf. Sci.*, 35(2):180–191, 2009. ISSN 0165-5515.
- A. B. Benitez and S.-F. Chang. Automatic multimedia knowledge discovery, summarization and evaluation. *IEEE Transactions on Multimedia*, 2003 (submitted), 2003.
- T. Berners-Lee, W. Hall, J. A. Hendler, K. O’Hara, N. Shadbolt, and D. J. Weitzner. **A framework for web science**. *Foundations and Trends in Web Science*, 1(1), 2006.
- T. Berners-Lee, J. Hendler, and O. Lassila. **The semantic web: Scientific american**. *Scientific American*, May 2001.
- Stephan Bloehdorn, Kosmas Petridis, Carsten Saathoff, Nikos Simou, Yannis Avrithis, Siegfried H, Yiannis Kompatsiaris, and Michael G. Strintzis. Semantic annotation of images and videos for multimedia analysis. In *In Proceedings of the 2nd European Semantic Web Conference, ESWC 2005*, pages 592–607, 2005.
- Federica Cena, Rosta Farzan, and Pasquale Lops. Web 3.0: Merging semantic web with social web. In *HT ’09: Proceedings of the Twentieth ACM Conference on Hypertext and Hypermedia*, New York, NY, USA, July 2009. ACM.
- Fabio Ciravegna and Yorick Wilks. Designing adaptive information extraction for the semantic web in amilcare. In *Annotation for the Semantic Web, Frontiers in Artificial Intelligence and Applications*. IOS. Press, 2003.
- Hamish Cunningham, Diana Maynard, Kalina Bontcheva, and Valentin Tablan. **GATE: A framework and graphical development environment for robust NLP tools and applications**. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, PA, USA*, 2002.
- Randall Davis. The digital dilemma. *Commun. ACM*, 44(2):77–83, 2001. ISSN 0001-0782.
- DCMI. **Dublin core metadata element set, version 1.1**, 2008.
- S. Decker, D. Fensel, F. Van Harmelen, I. Horrocks, S. Melnik, M. Klein, and J. Broekstra. Knowledge representation on the web. In *In Proc. of the 2000 Description Logic Workshop (DL 2000*, pages 89–97, 2000.

- Thierry Declerck, Jan Kuper, Horacio Saggion, Anna Samiotou, Peter Wittenburg, and Jesus Contreras. [Contribution of nlp to the content indexing of multimedia documents](#). In Peter Enser, Yiannis Kompatsiaris, Noel E. OConnor, Alan F. Smeaton, and Arnold W. M. Smeulders, editors, *Image and Video Retrieval*, volume 3115 of *Lecture Notes in Computer Science*, pages 647–647. Springer Berlin / Heidelberg, 2004. 10.1007/978-3-540-27814-6_71.
- C. Djeraba. [Content-based multimedia indexing and retrieval](#). *Multimedia, IEEE*, 9(2): 18–22, 2002.
- C. Dorai and S. Venkatesh. Computational media aesthetics: Finding meaning beautiful. *IEEE MultiMedia*, 8(4):10–12, 2001. ISSN 1070-986X.
- J. P. Eakins and M. E. Graham. [Content-based image retrieval: A report to the jisc technology applications programme](#). Technical report, Institute for Image Data Research, University of Northumbria at Newcastle, 1999.
- Peter G. B. Enser. Visual image retrieval. *Annual Review of Information Science and Technology.*, 42(1):1–42, 2008.
- Peter G.B. Enser, Christine J. Sandom, and Paul Lewis. [Automatic annotation of images from the practitioner perspective](#). LNCS 3:497–506, 2005.
- Adrienne Felt, Pieter Hooimeijer, David Evans, and Westley Weimer. Talking to strangers without taking their candy: isolating proxied content. In *SocialNets '08: Proceedings of the 1st Workshop on Social Network Systems*, pages 25–30, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-124-8.
- M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Qian Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker. Query by image and video content: the qbic system. *Computer*, 28(9):23 –32, September 1995. ISSN 0018-9162.
- Flickr. [Welcome to flickr - photo sharing](#), 2007.
- Roberto Garca, Chrisa Tsinaraki, scar Celma, and Stavros Christodoulakis. [Multimedia content description using semantic web languages](#). In Yiannis Kompatsiaris and Paola Hobson, editors, *Semantic Multimedia and Ontologies*, pages 17–54. Springer London, 2008. ISBN 978-1-84800-076-6. 10.1007/978-1-84800-076-6_2.
- J. Geurts, J. van Ossenbrugen, and L. Hardman. Requirements for practical multimedia annotation. In *Multimedia and the Semantic Web, 2nd European Semantic Web Conference*, 2005.
- Fabrizio Giustina. [The porter stemming algorithm](#), 2009.
- H Greisdorf and B. OConnor. Modelling what users see when they look at images: a cognitive viewpoint. *Journal of Documentation*, 58(1):6–29, 2002.

- William I. Grosky, Rajeev Agrawal, and Farshad Fotouchi. **Mind the gaps-finding the appropriate dimensional representation for semantic retrieval of multimedia assets.** In Yiannis Kompatsiaris and Paola Hobson, editors, *Semantic Multimedia and Ontologies*, pages 229–252. Springer London, 2008. ISBN 978-1-84800-076-6. 10.1007/978-1-84800-076-6₉.
- T. Gruber. **Ontology of folksonomy: A mash-up of apples and oranges**, 2005.
- Wendy Hall, David De Roure, and Nigel Shadbolt. **The evolution of the web and implications for eresearch.** *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 367(1890):991–1001, March 2009.
- J. S. Hare, P. H. Lewis, P. G. B. Enser, and C. J. Sandom. Semantic facets: an in-depth analysis of a semantic image retrieval system. In *CIVR '07: Proceedings of the 6th ACM international conference on Image and video retrieval*, pages 250–257, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-733-9.
- Jonathan Hare and Paul Lewis. **Automatically annotating the mir flickr dataset: Experimental protocols, openly available data and semantic spaces.** In *MIR '10: Proceedings of the international conference on Multimedia information retrieval*, pages 547–556. ACM, March 2010.
- Jonathon S. Hare, Sina Samangooei, Paul H. Lewis, and Mark S. Nixon. Semantic spaces revisited: investigating the performance of auto-annotation and semantic retrieval using semantic spaces. In *CIVR*, pages 359–368, 2008.
- Sebastian Hellmann, Claus Stadler, Jens Lehmann, and Sören Auer. **Dbpedia live extraction.** In *Proceedings of the Confederated International Conferences, CoopIS, DOA, IS, and ODBASE 2009 on On the Move to Meaningful Internet Systems: Part II*, OTM '09, pages 1209–1223, Berlin, Heidelberg, 2009. Springer-Verlag. ISBN 978-3-642-05150-0.
- James Hendler, Nigel Shadbolt, Wendy Hall, Tim Berners-Lee, and Daniel Weitzner. Web science: an interdisciplinary approach to understanding the web. *Commun. ACM*, 51(7):60–69, 2008. ISSN 0001-0782.
- Nicolas Hervé and Nozha Boujemaa. Image annotation: which approach for realistic databases? In *CIVR '07: Proceedings of the 6th ACM international conference on Image and video retrieval*, pages 170–177, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-733-9.
- Francis Heylighen. **Complexity and information overload in society: why increasing efficiency leads to decreasing control. technological forecasting and social change.** *Bulletin of the Medical Library Association*, 87:2, 2004.
- Jinwon Ho and Rong Tang. Towards an optimal resolution to information overload: an infomediary approach. In *GROUP '01: Proceedings of the 2001 International ACM*

- SIGGROUP Conference on Supporting Group Work*, pages 91–96, New York, NY, USA, 2001. ACM. ISBN 1-58113-294-8.
- A. Hotho, R. Jschke, C. Schmitz, and G. Stumme. **Trend detection in folksonomies**. In Yannis S. Avrithis, Yiannis Kompatsiaris, Steffen Staab, and Noel E. O’Connor, editors, *Proc. First International Conference on Semantics And Digital Media Technology (SAMT)*, volume 4306 of *LNCIS*, pages 56–70, Heidelberg, dec 2006. Springer. ISBN 3-540-49335-2.
- Mark J. Huiskes and Michael S. Lew. The mir flickr retrieval evaluation. In *MIR ’08: Proceeding of the 1st ACM international conference on Multimedia information retrieval*, pages 39–43, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-312-9.
- J. Hunter. Adding multimedia to the semantic web - building an mpeg-7 ontology. In *In International Semantic Web Working Symposium (SWWS)*, pages 261–281, 2001.
- B. J. Jansen and A. Spink. How are we searching the world wide web?: a comparison of nine search engine transaction logs. *Inf. Process. Manage.*, 42(1):248–263, 2006. ISSN 0306-4573.
- C. Jørgensen. **Image retrieval: theory and research**. *ASIS 1996 Annual Conference*, 1996.
- Corinne Jørgensen. *Image retrieval: theory and research*. Scarecrow Press, 2003.
- Jtidy. **Jtidy, html parser and pretty-printer in java**, 2009.
- Bo-Yeong Kang and Sang-Jo Lee. Document indexing: a concept-based approach to term weight estimation. *Inf. Process. Manage.*, 41(5):1065–1080, 2005. ISSN 0306-4573.
- Kirk L. Kroeker. Engineering the web’s third decade. *Commun. ACM*, 53(3):16–18, 2010. ISSN 0001-0782.
- Chul Min Lee, Serdar Yildirim, Murtaza Bulut, Abe Kazemzadeh, Carlos Busso, Zhigang Deng, Sungbok Lee, and Shrikanth Narayanan. Emotion recognition based on phoneme classes. In *Proc. ICSLP04*, pages 889–892, 2004.
- Thomas M. Lehmann, Henning Schubert A, Daniel Keysers B, Michael Kohnen A, and Berthold B. Wein A. B.b.: The irma code for unique classification of medical images. In *In: Medical Imaging. Volume 5033 of SPIE Proceedings*, pages 109–117, 2003.
- Stefanie Lindstaedt, Roland Mörzinger, Robert Sorschag, Viktoria Pammer, and Georg Thallinger. **Automatic image annotation using visual content and folksonomies**. *Multimedia Tools Appl.*, 42(1):97–113, 2009. ISSN 1380-7501.
- Stefanie Lindstaedt, Viktoria Pammer, Roland Mörzinger, Roman Kern, Helmut Mülner, and Claudia Wagner. Recommending tags for pictures based on text, visual content and user context. In *ICIW ’08: Proceedings of the 2008 Third International Conference on Internet and Web Applications and Services*, pages 506–511, Washington, DC, USA, 2008. IEEE Computer Society. ISBN 978-0-7695-3163-2.

- Jiebo Luo and A. Savakis. Indoor vs outdoor classification of consumer photographs using low-level and semantic features. *Image Processing, 2001. Proceedings. 2001 International Conference on*, 2:745–748 vol.2, October 2001.
- Pietikäinen M., Mäenpää T., and Viertola J. **Color texture classification with color histograms and local binary patterns**. In *Proc. 2nd International Workshop on Texture Analysis and Synthesis*, 2002.
- Tourism Malaysia. **Tourism malaysia official website**, 2009.
- MeSH. **Medical subject headings - home page**, June 2010.
- M. Montebello. Information overload-an ir problem? *String Processing and Information Retrieval: A South American Symposium, 1998. Proceedings*, pages 65–74, Sep 1998.
- J. Moore. **What is ajax?**, 2008.
- Y. Mori, H. Takahashi, and R. Oka. **Image-to-word transformation based on dividing and vector quantizing images with words**, 1999.
- Emily Moxley, Jim Kleban, Jiejun Xu, and B. S. Manjunath. Not all tags are created equal: learning flickr tag semantics for global annotation. In *ICME'09: Proceedings of the 2009 IEEE international conference on Multimedia and Expo*, pages 1452–1455, Piscataway, NJ, USA, 2009. IEEE Press. ISBN 978-1-4244-4290-4.
- Vivi Nastase, Marina Sokolova, and Jelber Sayyad Shirabad. Do happy words sound happy? a study of relations between form and meaning for english words expressing emotions. In *Proceedings of RANLP 2007*, Borovets, Bulgaria, 2007.
- Aurélié Névéal, Thomas M. Deserno, Stéfan J. Darmoni, Mark Oliver Güld, and Alan R. Aronson. Natural language processing versus content-based image analysis for medical document retrieval. *J. Am. Soc. Inf. Sci. Technol.*, 60(1):123–134, 2009. ISSN 1532-2882.
- BBC News. Facebook urged to add panic button at meeting with ceop, 2010.
- Naoko Nitta, Noboru Babaguchi, and Tadahiro Kitahashi. Generating semantic descriptions of broadcasted sports videos based on structures of sports games and tv programs. *Multimedia Tools Appl.*, 25(1):59–83, 2005. ISSN 1380-7501.
- L. J. B. Nixon. **Multimedia, web 2.0 and the semantic web: A strategy for synergy**. *First International Workshop on Semantic Web Annotations for Multimedia (SWAMM)*, 2006.
- Shahrul Azman Noah, Lailatulqadri Zakaria, and Arifah Che Alhadi. Extracting and modeling the semantic information content of web documents to support semantic document retrieval. In *APCCM '09: Proceedings of the Sixth Asia-Pacific Conference on Conceptual Modeling*, pages 79–86, Darlinghurst, Australia, Australia, 2009. Australian Computer Society, Inc. ISBN 978-1-920682-77-4.

- Aude Oliva, Antonio Torralba, Anne Guerin Dugue, and Jeanny Herault. Global semantic classification of scenes using power spectrum templates. 1999.
- T. Oreilly. *What is web 2.0: Design patterns and business models for the next generation of software*. *Social Science Research Network Working Paper Series*, 2005.
- Paul Over, George Awad, Jon Fiscus, Martial Michel, Alan F. Smeaton, and Wessel Kraaij. *Trecvid 2009 - goals, tasks, data, evaluation mechanisms and metrics*. National Institute for Standards and Technology (NIST), 2010.
- E. Panofsky. *Studies in Iconology: Humanistic Themes in the Art of the Renaissance*. Harper & Row, 1962. ISBN 0064300250.
- G.Z. Pastorello, J. Daltio, and C.B. Medeiros. Multimedia semantic annotation propagation. In *Multimedia, 2008. ISM 2008. Tenth IEEE International Symposium on*, pages 509–514, dec. 2008.
- Katerina Pastra, Horacio Saggion, and Yorick Wilks. Nlp for indexing and retrieval of captioned photographs. In *EACL '03: Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics*, pages 143–146, Morristown, NJ, USA, 2003. Association for Computational Linguistics. ISBN 1-111-56789-0.
- Andrew Payne and Sameer Singh. Indoor vs. outdoor scene classification in digital photographs. *Pattern Recogn.*, 38(10):1533–1545, 2005. ISSN 0031-3203.
- H. Petrie, C. Harrison, , and S. Dev. Describing images on the web: a survey of current practice and prospects for the future. In *Describing images on the web: a survey of current practice and prospects for the future*, 2005.
- Hamid Rahimizadeh, M.H Marhaban, R.M Kamil, and N.B Ismail. Color image segmentation based on bayesian theorem and kernel density estimation. *European Journal of Scientific Research*, 26:430–436, 2009.
- Tye Rattenbury, Nathaniel Good, and Mor Naaman. Towards automatic extraction of event and place semantics from flickr tags. In *SIGIR '07: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 103–110, New York, NY, USA, 2007. ACM. ISBN 978-1-59593-597-7.
- Neil C. Rowe and Eugene J. Guglielmo. Exploiting captions in retrieval of multimedia data. volume 29, pages 453–461, Tarrytown, NY, USA, 1993. Pergamon Press, Inc.
- Maria Ruiz-Casado, Enrique Alfonseca, Manabu Okumura, and Pablo Castells. *Information extraction and semantic annotation of wikipedia*. In *Proceeding of the 2008 conference on Ontology Learning and Population: Bridging the Gap between Text and Knowledge*, pages 145–169, Amsterdam, The Netherlands, The Netherlands, 2008. IOS Press. ISBN 978-1-58603-818-2.

- Navid Serrano, Andreas Savakis, and Jiebo Luo. A computationally efficient approach to indoor/outdoor scene classification. *International Conference on Pattern Recognition (ICPR'02)*, 4:40146, 2002.
- Navid Serrano, Andreas E. Savakis, and Jiebo Luo. Improved scene classification using efficient low-level features and semantic cues. *Pattern Recognition*, 37(9):1773–1784, 2004.
- N. Shadbolt, T. Berners-Lee, and W. Hall. The semantic web revisited. *IEEE Intelligent Systems*, 21(3):96–101, 2006. ISSN 1541-1672.
- Nigel Shadbolt and Tim Berners-Lee. **Web science emerges**. *Scientific American*, October 2008.
- B. Shah, V. Raghavan, and P. Dhatric. Efficient and effective content-based image retrieval using space transformation. In *MMM '04: Proceedings of the 10th International Multimedia Modelling Conference*, page 279, Washington, DC, USA, 2004. IEEE Computer Society. ISBN 0-7695-2084-7.
- Victoria Shannon. **A 'more revolutionary' web**, 2006.
- S. Shatford. Analyzing the subject of a picture: A theoretical approach. *Cataloging & Classification Quarterly*, 6(3):39–62, 1986.
- Börkur Sigurbjörnsson and Roelof van Zwol. Flickr tag recommendation based on collective knowledge. In *WWW '08: Proceeding of the 17th international conference on World Wide Web*, pages 327–336, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-085-2.
- Juan M. Silva, Abu Saleh Md. Mahfujur Rahman, and Abdulmotaleb El Saddik. Web 3.0: a vision for bridging the gap between real and virtual. In *CommunicabilityMS '08: Proceeding of the 1st ACM international workshop on Communicability design and evaluation in cultural and ecological multimedia system*, pages 9–14, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-319-8.
- Josef Sivic and Andrew Zisserman. **Video Google: A Text Retrieval Approach to Object Matching in Videos**. *Computer Vision, IEEE International Conference on*, 2:1470–1477 vol.2, April 2003.
- S. Skine. **Proteus project - apple pie parser (corpus based parser)**, 2008.
- A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(12):1349–1380, 2000. ISSN 0162-8828.
- John R. Smith and Shih-Fu Chang. **Visualeek: a fully automated content-based image query system**. In *Proceedings of the fourth ACM international conference on Multimedia, MULTIMEDIA '96*, pages 87–98, New York, NY, USA, 1996. ACM. ISBN 0-89791-871-1.

- Amanda Spink, Dietmar Wolfram, Major B. J. Jansen, and Tefko Saracevic. Searching the web: the public and their queries. *J. Am. Soc. Inf. Sci. Technol.*, 52(3):226–234, 2001. ISSN 1532-2882.
- Martin Szummer and Rosalind W. Picard. Indoor-outdoor image classification. In *CAIVD '98: Proceedings of the 1998 International Workshop on Content-Based Access of Image and Video Databases (CAIVD '98)*, page 42, Washington, DC, USA, 1998. IEEE Computer Society. ISBN 0-8186-8329-5.
- Malaysia Tourism. [Malaysia tourism](#), 2009.
- W. Treese. Web 2.0: is it really different? *netWorker*, 10(2):15–17, 2006. ISSN 1091-3556.
- R. Troncy. [Integrating structure and semantics into audio-visual documents](#). In *2nd International Semantic Web Conference (ISWC'03)*, volume LNCS 2870, pages 566–581, Sanibel Island, Florida, USA, October 2003.
- Mischa Tuffield, Stephen Harris, David P. Dupplaw, Ajay Chakravarthy, Christopher Brewster, Nicholas Gibbins, Kieron O'Hara, Fabio Ciravegna, Derek Sleeman, Yorick Wilks, and Nigel R. Shadbolt. [Image annotation with photocopain](#). In *First International Workshop on Semantic Web Annotations for Multimedia (SWAMM 2006) at WWW2006*, 2006.
- V. Uren, P. Cimiano, J. Iria, S. Handschuh, M. Vargas-Vera, E. Motta, and F. Ciravegna. [Semantic annotation for knowledge management: Requirements and a survey of the state of the art](#). *Web Semantics: Science, Services and Agents on the World Wide Web*, 4(1):14–28, January 2006.
- Aditya Vailaya, Anil Jain, Mario Figueiredo, and HongJiang Zhang. Content-based hierarchical classification of vacation images. In *ICMCS '99: Proceedings of the IEEE International Conference on Multimedia Computing and Systems*, page 9518, Washington, DC, USA, 1999. IEEE Computer Society. ISBN 0-7695-0253-9.
- Aditya Vailaya, Associate Member, Mario A. T. Figueiredo, Anil K. Jain, Hong-Jiang Zhang, and Senior Member. Image classification for content-based indexing. *IEEE Transactions on Image Processing*, 10:117–130, 2001.
- Koen E.A. van de Sande, Theo Gevers, and Cees G.M. Snoek. Evaluating color descriptors for object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32:1582–1596, 2010. ISSN 0162-8828.
- J. van Ossenbruggen, J. Geurts, F. Cornelissen, L. Hardman, and L. Rutledge. Towards second and third generation web-based multimedia. In *WWW '01: Proceedings of the 10th international conference on World Wide Web*, pages 479–488, New York, NY, USA, 2001. ACM. ISBN 1-58113-348-0.

- J. van Ossenbruggen, F. Nack, and L. Hardman. That obscure object of desire: Multimedia metadata on the web, part 1. *IEEE MultiMedia*, 11(4):38–48, 2004. ISSN 1070-986X.
- R. van Zwol, S. Rger, M. Sanderson, and Y. Mass. **Multimedia information retrieval: "new challenges in audio visual search"**. *SIGIR Forum*, 41(2):77–82, 2007.
- J. Voß. **Tagging, folksonomy & co - renaissance of manual indexing?**, 2007.
- W3C. **W3c semantic web frequently asked questions**. online, 2007.
- T. V. Wal. **Folksonomy definition and wikipedia :: Off the top :: vanderwal.net**, 2007.
- Wikipedia. **Wikipedia:about**, 2011.
- Ian H. Witten, Gordon W. Paynter, Eibe Frank, Carl Gutwin, and Craig G. Nevill-Manning. Kea: practical automatic keyphrase extraction. In *DL '99: Proceedings of the fourth ACM conference on Digital libraries*, pages 254–255, New York, NY, USA, 1999. ACM. ISBN 1-58113-145-3.
- WTO. **World tourism organization, committed to tourism and the millennium development goals**, 2008.
- Fei Wu, Raphael Hoffmann, and Daniel S. Weld. **Information extraction from wikipedia: moving down the long tail**. In *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '08, pages 731–739, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-193-4.
- Takeo Kanade Yu-Ichi Ohta and Toshiyuki Sakai. Color information for region segmentation. pages 222–241, 1980.
- J. Zhang, M. Marsza, S. Lazebnik, and C. Schmid. Local features and kernels for classification of texture and object categories: A comprehensive study. *Int. J. Comput. Vision*, 73(2):213–238, 2007. ISSN 0920-5691.
- Lei Zhang, Mingjing Li, and Hong-Jiang Zhang. Boosting image orientation detection with indoor vs. outdoor classification. In *WACV '02: Proceedings of the Sixth IEEE Workshop on Applications of Computer Vision*, page 95, Washington, DC, USA, 2002. IEEE Computer Society. ISBN 0-7695-1858-3.