

Modelling Image Semantic Descriptions from Web 2.0 Documents using a Hybrid Approach.

Lailatul Qadri Zakaria¹, Wendy Hall² and Paul Lewis³

School of Electronics and Computer Science,
University of Southampton,
United Kingdom.

lqz06r@ecs.soton.ac.uk¹, wh@ecs.soton.ac.uk², phl@ecs.soton.ac.uk³

ABSTRACT

With the increasing amount of multimedia content on the web added as user generated content in Web 2.0 websites, conventional multimedia information retrieval is presented with new challenges. It is no longer possible to rely only on meta-data based retrieval but to consider also content based techniques combined with the collective knowledge generated by users' contributions and geo-referenced meta-data. Tagging is a modest way to annotate such documents and fails to capture a full semantic description of the document content. This report concerns ongoing research to investigate a means to identify, model and utilise semantic descriptions of the user-generated content in Web 2.0 documents using a hybrid approach. The approach consists of three main components, natural language processing, image analysis and a shared knowledge base. In this paper we describe the complete model but, as the image analysis component is in its early stages, the results focus on the natural language processing and the knowledge base. We show that the additional use of these components can improve retrieval and analysis performance over that based only on Web 2.0 tags.

Keywords

Image semantic description, natural language analysis, knowledge base, ontology, semantic web.

1. INTRODUCTION

The Web is a multimedia environment, which makes for complex semantics [1]. 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 [2].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

iiWAS2009, December 14–16, 2009, Kuala Lumpur, Malaysia.
Copyright 2009 ACM 978-1-60558-660-1/09/0012...\$10.00.

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. Flickr had more than 3 billion images by the end of 2008 and is adding thousands of images per minute. 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. Photos are often poorly annotated and the query is typically done by two to three words only.

Furthermore, 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 the semantics to represent the main concepts of the document [3]. For instance, searching for images related to two keywords "Tourism" and "Malaysia", using the Flickr searching mechanisms returns 9336 images using a full text method and 8105 images using the tag based search¹. To use the results, a user needs to filter the answers themselves by reading the returned documents.

Managing information within Web 2.0 documents presents new challenges to conventional multimedia information retrieval (MIR). There is a need to rely not only 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 [4]. We believe that content generated by users contains fruitful information which could be very useful to improve the information retrieval. However, its informal nature makes it much harder to structure them using traditional annotation. Therefore, we propose a hybrid approach to Web 2.0 information management, an integration of natural language analysis, image analysis and the use of knowledge bases, to tackle the richness of user generated content in Web 2.0 documents and to exploit some of the benefits of semantic web technologies.

This paper is organized into the following sections. The next section provides the background and related work. Section 3 describes the approach employed in modelling the semantic description of the documents. Section 4 provides preliminary results and discussion and finally, section 5 concludes the work done and briefly describes the future work.

¹ The search was done on April 15, 2009

2. RESEARCH BACKGROUND AND RELATED WORK

Community involvement in annotating (tagging) multimedia content form a Web scale collaboration activity and seem 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 [5]. These folksonomies will become more stable, gradually maturing over time [6] and may eventually provide a good solution to overcome knowledge acquisition problems - previously stated as the major issue for many knowledge based systems. However, folksonomies may be too semantically loose to be able to guarantee sufficient accuracy in information content representation [7].

Furthermore, based on our own observations, users are not providing enough information to describe their own material effectively, which may leads to issues in information retrieval and trustworthiness of the provided information. Therefore, even though tags do provide additional information to documents, relying on users alone to tag/annotate such documents may not necessarily solve the problem. To overcome this, a formal knowledge representation or ontology is used to bridge the semantic gap between the emergence of folksonomies and the Semantic Web. Ontologies can be used for standardization and, through their use, the content and knowledge generated from Web 2.0 can be structured.

Considering the Semantic Web, there is a gap between textual and non-textual (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. Nevertheless, for multimedia documents, there is a significant growth in Web 2.0 but exploiting Semantic Web technologies for these 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 [8] while for the Semantic Web, the main intention is for machine readable and processable content [9]. The realization of the Semantic Web requires information to be explicit and accessible directly by machine but some of its methodologies can assist with retrieval for humans as well.

2.1 Related Work

We classified the related work into three groups.

2.1.1 Annotation of multimedia content using tags.

There is an increasing interest in annotating multimedia content (namely images) in Web 2.0 documents by creating tag recommendation systems [14]-[17]. In [14] and [17], a tag co-occurrence method which is an asymmetric metric has been used as the key to their approach to identifying related tags, while [15] used a symmetric metric to find similar tags. Nevertheless, the analysis done in such research does not convey any semantics in representing the content of the image. Unlike [14], [15], and [17], the work done in [16] attempted to tag images semantically based on categories such as landmarks, places and visual descriptors. The analysis was done using only geo-tagged images, as the focus is to find places and landmarks of the images.

2.1.2 Information identification using Natural Language Processing

As text is the most used medium in delivering information on Web document content (including Web 2.0), it is crucial to be able to identify and represent the content in an efficient and effective manner. Natural language processing is a common approach used to tackle the issue. Most of the research is concerned with identifying and extracting information from unstructured text in HTML format and translating such information into a semantic web language such as XML or RDF [10, 11]. Analyses have been done on conventional Web documents, which are well documented and created by experts. In these investigations, the analysis is semi-automatic and is supported by user feedback and ontology.

2.1.3 Image Semantic Analysis

Describing image semantics often revolves around four key questions which are *who*, *what*, *where* and *when*, which would help in understanding what the image is and what it is about [19]. These questions were reformulated to describe the semantic facets of an image which combine *object*, *spatial*, *temporal* and *activity/event* facets from [20] with *abstract* and *related concepts* facet, and also *context* and *topic* facet, which capture the highest level in global semantic content of the image [21]. Nevertheless, current content based image retrieval systems typically index the images based on their low level features such as colour, shape and texture and very few attempt to really capture the higher semantic levels for retrieval. Many papers have documented the value of content based image analysis including for example previous work here on museum collections [22].

3. MODELING IMAGE SEMANTIC DESCRIPTIONS – THE HYBRID APPROACH

Figure 1 illustrates the overall process in our hybrid approach. In general, the approach consists of three main components, text analyzer, image analyzer and the knowledge bases. The text analyzer aims to capture text-based information that can be useful to represent the content and context of an image. The aim of the image analyzer is to identify objects and scenes from the image itself, and to analyse its Exif metadata.

3.1 Experiment Setup

To initiate the study, we are using Web documents from the Flickr website as our input resources. 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)). The Flickr repository can be accessed directly by using the Flickr API². So far, we have extracted 2000 documents related to Tourism in Malaysia. For each image extracted, we store its EXIF metadata and its description in XML format. At this stage, the analysis is only done on the textual description of the images and the Exif metadata will be analysed in the Image Analysis Stage.

² A set of keys is required and these numbers can be obtained from Flickr.

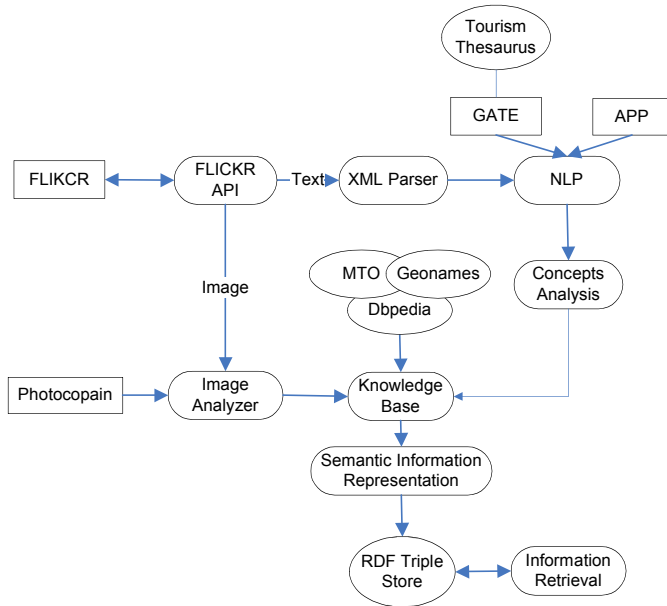


Figure 1. The Processes Workflow Overview of the Hybrid Approach

3.2 Text Analyser Component

For each image submitted for analysis, its textual description is processed by the text analyzer component. The text analyzer component consists of two NLA tools which are GATE [13] and the Apple Pie Parser (APP) [12].

GATE: 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 name of person, time, location, address and organization. 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 most common concepts that can be used to describe/define information related to the tourism domain such as attractions, environments and transportations. The thesaurus is provided by the World Tourism Organization specification. Table 1 shows some common concepts for the tourism domain.

Table 1. Common Concepts in Tourism Domain

Attraction	Environment	Transportation	Activity
Tower	Beach	Airplane	Dancing
Gallery	Mountain	Bicycle	Hiking
Bridge	City	Bus	Snorkelling
Garden	Island	Car	Paragliding
Museum	Lake	Coach	Climbing

APP: 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 are sometimes unpredictable (e.g. incomplete or including abbreviations). Figure 2 shows an example of APP output – a syntactic parse tree

that represents the syntactic structure of words based on formal grammar. The parse tree can be used to extract noun phrases which are a good indicator for identifying concepts. For example, in Figure 2, four concepts can be extracted which are *sunset*, *kuala beach*, *langkawi* and *kedah*.



Figure 2. Visualization of Parse Tree generated by APP

Each of the concepts identified by GATE and APP will pass 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 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 higher semantic information description. Furthermore, a sound semantic image description does require feeling and thought. Table 2 shows a result example generated from the text analyser.

Table 2. Text Analyser Component Output

Concept	Class	Frequency	Root word
Islands	Attraction	3	Island
Island	Attraction	3	Island
Kapas island	-	2	Invalid ³
Malaysia	-	2	Malaysia
Terengganu	-	1	Terengganu

3.3 Knowledge Bases Component

The concepts extracted from text analyzer component will then pass to the knowledge base component. The knowledge bases component consists of ontologies and an open knowledge base, which substitute for experts in the domain of interest. This component is vital to provide us with the information needed to identify concepts that are related to the domain.

³ Root word for concept that is more than one word is invalid.

3.3.1 Malaysia Tourism Ontology (MTO):

The Malaysia Tourism Ontology is a specific domain ontology, which is created to store information related to tourism in Malaysia. 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.

3.3.2 Geonames Ontology

Geonames is a geographical database that can be used in finding specific 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.

3.3.3 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. Results given by Dbpedia may vary depending on the availability of structured information. 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 of multiple knowledge disciplines.

Each of the concepts extracted from the text analyzer will be compared with the entries in the knowledge bases. The result of the analysis is stored in RDF format. Table 3 shows some concepts that can be identified by using the knowledge base component.

Table 3. Knowledge Base Component Output

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

The output produced by the approach can be visualized as Figure 3. Concepts that are identified by the tool 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 have been captured by the tool are *tourist*, *tourism*, *stone*, *sky*, *sharing*, *power*, *portugis*, *protugal*, *nice*, *n21*, *lovely historical*, *history*, *colonialism*, *exploration*, *clouds*, *cannon*, *beautiful*, *architectural gems*, *ancient and alfonsoalbuquerque*.

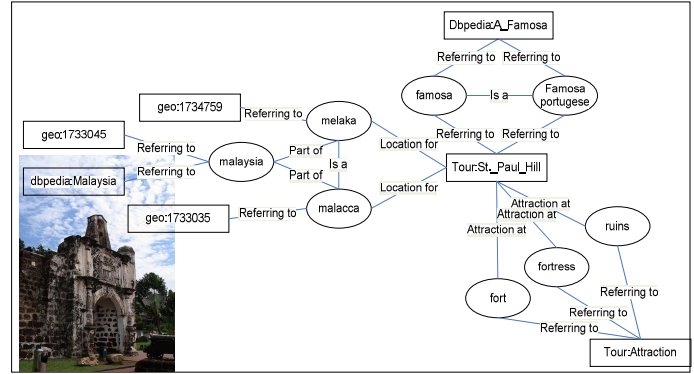


Figure 3. Visualization for information extracted for Image ID 1460920756.

In Figure 4, the image has triggered many interests which have significantly increased its document length. The length of a document corresponds directly with the significance of using text analysis to analyse the document. In this example, there is some ambiguity where the word beaches is also found in the document. Nevertheless the frequency of the word is lower than other words matched with attraction type.

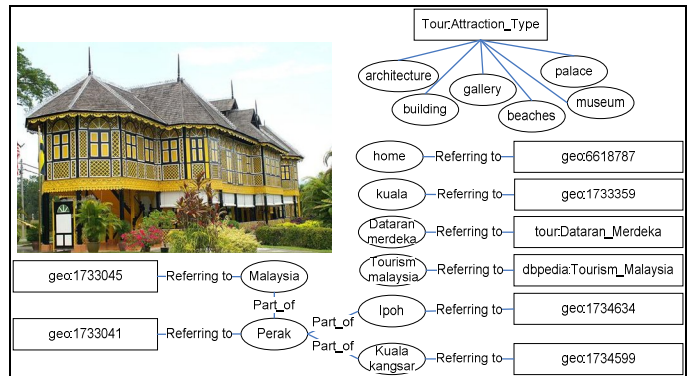


Figure 4. Visualization for information extracted for Image ID 1203148615.

The lengths of the documents in our corpus vary from one to another due to two main factors which are (1) lack of description provided by the author of the document, and (2) the interest that the image has triggered which usually will be reflected in the numbers of comments/ feedback left by viewers. In the cases where the length of the document is short, the analysis of term frequencies 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 for enhancing the description of the image.

3.4 Image Analysis Component

The development of the Image analysis component is still in its early stages, thus we will briefly describe the aims and tasks in this component. The aim of the image analysis component is to analyse the image itself to provide support for the information gathered in the previous components thus giving higher confidence when describing the images. The component will be handling two main tasks, which are analysing the camera metadata parameters (the Exif data) and analysing the image content (pixel data).

3.4.1 Camera metadata parameter analysis.

Two most common camera metadata types embedded in images are Exif and XMP metadata. Exif metadata is generated automatically and captures the information of the camera settings during the creation of the image. Exif metadata cannot be altered and the data is embedded directly in the image. Moreover the information of the Exif metadata could be lost if the user makes modifications to the image. In contrast, the XMP metadata is created manually by the user using the XMP tool. The XMP tool allows the user to participate in adding more metadata to the document such as a caption abstract via a caption writer editor, object name, title and keywords via a general editor.

3.4.2 Image analysis.

The latter task will be executed by employing 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 [18]. Photocopain will be used to help us to identify scenes, such as views, objects (such as monument, building, mountain) and festivals, which would provide a clue about the main interest of the images. For example, information representation for an image with a scenery view would possibly be closely related with information such as location and temporal information of the image, while an image with an object oriented scene, such as a museum artefact, would trigger interest in different information such as the name of the object and its location. Moreover, if we could identify an image of a festival, then it would be useful to integrate information about the festival with the image. Such information will provide additional input and will be integrated with information that has already been identified during the text analysis stage to improve the reliability and thus enrich the information representation of the images.

3.5 Information Retrieval Component

To ease the information retrieval process, we have created a prototype for retrieving from an RDF triple store. The searching can be done in two modes, which are, general search mode and specific search mode. During the general search mode, the searching is done by mapping terms added by users with the concepts within each documents and the result will be presented to users in the form of a list of the images. The semantic representation of the information within each document will be ignored during this mode as the query is done by using a flat text index. To find information based on General Query Mode, the SPARQL formulation query is as follow:

```
Query A: Find images X.
SELECT ?x
WHERE{
?term foaf:name ?x.
}
```

In the specific search mode, the searching mechanism will undergo specific information mapping. To ease the query, we will provide a simple form which will allow users to add information that they knew and mark information that they want. Such information will be used to create a query model in SPARQL language. This query model than will be matched with the semantic information representation of the documents (RDF

format). For example, finding information such as specific location and attraction for images will be formulated as bellow:

```
Query B: Find information of location Y and attraction Z.
SELECT ?location ?attraction
WHERE{
?termLocation foaf:name ?y
?termLocation dbase:indicate "location"
?termLocation foaf:name ?z
?termLocation dbase:indicate "attraction"
}
```

The general search mode will generate a list of images that matched with the query while the later search mode will generate a list of images with additional information tailored to the query. Searching by using the specific mode is done to identify specific images or specific information related to the images.

4. Preliminary Results and Discussion

We conducted an information retrieval experiment to observe the results produced by combining the natural language processing and knowledge based approaches (KB + NLP). The experiment tests the ability of the approach to find correct images compared to the conventional full text (F) and tags based (T) searches. The analysis is done by observing and comparing the semantic information generated with a bag of words and tags. Tagging does not have a standard method for different concepts, for example, images related to "Mabul island" could be represented in different patterns such as in the phrase (*mabul island*), one single word (*mabulisland*) or with two words (*mabul, island*). In cases where querying information involves phrases, for tagging, we assumed queries matching to any of these patterns are correct. For the KB + NLP approach, to ensure the quality of the information used to describe the images, we only consider it is correct if the query phrase found an exact match with the phrase itself. The result of the analysis is shown in Table 4.

In this experiment, three approaches to search have been used to identify images related to the queries, which are i) full text search, ii) tags only and iii) knowledge base and natural language analysis. In the first query, finding images of island, the knowledge base and natural language processing (KB+NLP) approach has found 41 images out of 43 relevant images (R'I), while the full text (F) and tags (T) based search has identified 17 and 7 images respectively. In the second query, finding images of *mabul island* as two words, the KB+NLP approach has identified 36 from 39 relevant images, while F search has identified 15 correct images and the T search has identified 5 correct images. In the third search, finding images of "*mabul island*" as one phrase, the KB+NLP approach has returned the similar result as in Query 2, while the F and the T based approach has found 10 and null images respectively. Similar pattern can be seen in Query 4, 5 and 6. The KB+NLP approach has found all relevant images for Query 4, Query 5 and Query 6 while F approach has identified 37, 7 and 5 images respectively and T approach has identified 30, 23 and null images respectively.

The advantages of using the KB+NLP approach is clearly on recall, identifying more images than the F and T approaches. The use of information in the knowledge bases has increased the capabilities of the approach in expanding existing information describing the images. For example, in this case, alternative names for the *mabul island* entry such as *mabul* and *pulau mabul*

thus allowing the approach to associate/link these words with *mabul island*. Moreover, the concept *mabul island* in MTO is also associated with other information such as environment *island* and attraction *coral reef* and activity *scuba diving* which provide useful rich information to improve the capabilities of information retrieval.

The natural language processing is useful to identify concepts in the form of phrases. In this experiment, *mabul island* was identified from the textual description of the images despite it being tagged as one word *mabulisland* or two different words “*mabul*” and “*island*”. For our approach, we only consider the correct image was found when the query matched with the phrase *mabul island* thus other patterns will be considered as incorrect. For Q2 and Q3, the phrase *mabul island* is treated as two separate words and one phrase respectively. The effect of this can be seen clearly in the full text and tags approaches.

Even though the KB+NLP approach found more images, it has a low precision (0.015) compared to the full text search. In Q2 and Q3, the approach has identified other islands such as *Denawan island* and *Sipadan island* which maybe tagged or commented with the keyword *mabul* or *Mabul island*. These three islands are geographically located near to each other thus it is common to see images of any of these islands described/commented together. For Q5 and Q6, all of the approaches missed one image due to misspelling of the word *thaipusam*.

Table 4. Preliminary results For Precision And Recall

Query	Approach						R' I
	F		T		KB+NL P		
	Retrieved (R)	Correct (C)	R	C	R	C	
Q1: Island	17	17	7	7	41	41	43
Q2:mabul island	16	15	7	5	39	36	39
Q3:“mabul island”	10	10	0	0	39	36	39
Q4: festival	37	37	30	30	41	41	41
Q5:thaipusam festival	27	27	23	23	31	31	32
Q6:“thaipusam festival”	5	5	0	0	31	31	32

Query	Precision			Recall		
	F	T	KB+NLP	F	T	KB+ NLP
Q1	1.000	1.000	1.000	0.395	0.162	0.953
Q2	0.937	0.714	0.923	0.384	0.128	0.923
Q3	1.000	0.000	0.923	0.256	0.000	0.923
Q4	1.000	1.000	1.000	0.902	0.731	1.000
Q5	1.000	1.000	1.000	0.843	0.741	0.989
Q6	1.000	0.000	1.000	0.156	0.000	0.989
Average	0.989	0.619	0.974	0.489	0.293	0.962

To test the usefulness of semantic annotation proposed by our approach, we tried to find specific information about location, environment and attraction of images. The semantic annotation allows the information retrieval to narrow down the search and identify images specifically tailored to users’ needs. Figure 5 depicts some of the images corresponding to the query.



Location: Kuala Lumpur
Environment: City
Attraction: building



Location: Langkawi
Environment: Island
Attraction: bridge



Location: Kuala Lumpur
Environment: city
Attraction: Mosque, buildings



Location: Mabul Island
Environment: Island
Attraction: coral reef

Figure 5. Images with semantic annotations.

5. CONCLUSION AND FUTURE WORK

In this study, we have been focusing on analysing text descriptions within Web 2.0 documents. The use of knowledge bases is vital to provide information needed to identify concepts to describe the image. The integration of natural language tools and knowledge bases is useful not only to identify the important concepts but also to link the concepts to create a semantic description of the image. We conclude that capturing a semantic description of an image would help to tackle the looseness of the semantics within Web 2.0 documents and thus could be useful to improve information retrieval.

For future work, we are going to focus on analysing the images themselves. The focus of this analysis is to identify information conveyed by the images by focusing on Exif metadata parameters. In our study, we would like to bridge the gap between images and their text descriptions by integrating the information conveyed by the images using an image analysis approach and the corresponding information captured in text descriptions using a natural language approach. Such sets of information hopefully will complement each other to increase the reliability and enrich the information used to represent the images and hence to improve the information retrieval.

6. ACKNOWLEDGMENTS

The first author is funded by the Malaysian Government, IPTA Academic Training Scheme. The second and third authors acknowledge support from the Living Knowledge FP7 project number 231126.

7. REFERENCES

- [1] Berners-Lee, T., Hall, W., Hendler, J. A., O'Hara, K., Shadbolt, N. and Weitzner, D. J.A . 2006. Framework for Web Science. *Foundations and Trends in Web Science*, 1 (1). pp. 1-130.
- [2] T. O'Reilly. 2005. "What Is Web 2.0? Design Patterns and Business Models for the Next Generation of Software," 30 Sept.
- [3] Kang, B.-Y., & Lee S.-J.. 2005. "Document indexing: a concept-based approach to term weight estimation." *Information Processing and Management: an International Journal*. 41(5): 1065 – 1080.
- [4] R. van Zwol, S. Rger, M. Sanderson, and Y. Mass. 2007. Multimedia information retrieval: "new challenges in audio visual search". SIGIR Forum, 41(2):77–82.
- [5] N. Shadbolt, T. Berners-Lee, and W. Hall. 2006. The semantic web revisited. *IEEE Intelligent Systems*, 21(3):96–101, ISSN 1541-1672.
- [6] A. Hotho, R. Jschke, C. Schmitz, and G. Stumme. 2006. 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 LNCS, pages 56–70, Heidelberg, Springer. ISBN 3-540-49335-2.
- [7] L. J. B. Nixon. 2006. Multimedia, web 2.0 and the semantic web: A strategy for synergy. First International Workshop on Semantic Web Annotations for Multimedia (SWAMM).
- [8] J. van Ossenbruggen, J. Geurts, F. Cornelissen, L. Hardman, and L. Rutledge. 2001. 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, USA. ACM. ISBN 1-58113-348-0.
- [9] T. Berners-Lee, J. Hendler, and O. Lassila. 2001. The semantic web: Scientific american. *Scientific American*.
- [10] Shahrul Azman Mohd. Noah, Lailatulqadri Zakaria, Arifah Che Alhadi: 2005. An Ontological Approach to Semantic Information Extraction and Integration of Web Documents. *iiWAS 2005*: 223-231
- [11] Alani, H., Kim, S., Millard, D., Weal, M. Hall, W., Lewis, P. and Shadbolt, N. 2003. Automatic Ontology- Based Knowledge Extraction from Web Documents. *IEEE Intelligent Systems* 18(1).pp.14-21.
- [12] Sekine S. 2002. Proteus Project - Apple Pie Parser (Corpus based Parser). <http://nlp.cs.nyu.edu/app>
- [13] H. Cunningham. 2002. GATE, a General Architecture for Text Engineering. *Computers and the Humanities*, 36:223–254.
- [14] A. Anderson, K. Ranghunathan and A. Vogel. 2008. TagEz: Flickr Tag Recommendation. Association for the Advancement of Artificial Intelligence.
- [15] Lindstaedt, S., Pammer, V., Mörzinger, R., Kern, R., Mülner, H., and Wagner, C. 2008. Recommending Tags for Pictures Based on Text, Visual Content and User Context. In *Proceedings of the 2008 Third international Conference on internet and Web Applications and Services* (June 08 - 13, 2008). ICIW. IEEE Computer Society, Washington.
- [16] E. Moxley, J. Kleban, J. Xu, and B. S. Manjunath, "Not all tags are created equal: Learning Flickr tag semantics for global annotation," in Proceedings of IEEE International Conference on Multimedia & Expo, Cancun, Mexico, June 2009.
- [17] Sigurbjörnsson, B. and van Zwol, R. 2008. Flickr tag recommendation based on collective knowledge. In *Proceeding of the 17th international Conference on World Wide Web* (Beijing, China, April 21 - 25, 2008). WWW '08. ACM.
- [18] Tuffield, M., Harris, S., Dupplaw, D. P., Chakravarthy, A., Brewster, C., Gibbins, N., O'Hara, K., Ciravegna, F., Sleeman, D., Wilks, Y. and Shadbolt, N. R. 2006. Image annotation with Photocopain. In: First International Workshop on Semantic Web Annotations for Multimedia (SWAMM 2006) at WWW2006, May 2006, Edinburgh, United Kingdom.
- [19] Shatford, S. 1986. Analyzing the subject of a picture: A theoretical approach. *Cataloging & Classification Quarterly*, 6(3), 39-62.
- [20] Layne, S. S. 1994. Some issues in the indexing of images. *J. Am. Soc. Inf. Sci.* 45(8) pp.583-588.
- [21] Hare, J. S., Lewis, P. H., Enser, P. G., and Sandom, C. J. 2007. Semantic facets: an in-depth analysis of a semantic image retrieval system. In *Proceedings of the 6th ACM international Conference on Image and Video Retrieval* (Amsterdam, The Netherlands, July 09 - 11, 2007). CIVR '07. ACM, New York, NY.
- [22] Lewis, P. H., Martinez, K., Abas, F. S., Ahmad Fauzi, M. F., Addis, M., Lahanier, C., Stevenson, J., Chan, S. C. Y., Mike J., B. and Paul, G. 2004. An Integrated Content and Metadata based Retrieval System for Art. *IEEE Transactions on Image Processing*, 13 (3). pp. 302-313.