Crew: Cross-modal Resource Searching by Exploiting Wikipedia

Chen Liu, Beng Chin Ooi, Anthony K. H. Tung, Dongxiang Zhang
School of Computing, National University of Singapore
13 Computing Drive, 117417, Singapore
liuchen,ooibc,atung,zhangdo@comp.nus.edu.sg

ABSTRACT

In Web 2.0, users have generated and shared massive amounts of resources in various media formats, such as news, blogs, audios, photos and videos. The abundance and diversity of the resources call for better integration to improve the accessibility. A straightforward approach is to link the resources via tags so that resources from different modals sharing the same tag can be connected as a graph structure. This naturally motivates a new kind of information retrieval system, named cross-modal resource search, in which given a query object from any modal, all the related resources from other modals can be retrieved in a convenient manner. However, due to the tag homonym and synonym, such an approach returns results of low quality because resources with the same tag but not semantically related will be directly connected as well. In this paper, we propose to build the resource graph and perform query processing by exploiting Wikipedia. We construct a concept middle-ware between the layer of tags and resources to fully capture the semantic meaning of the resources. Such a cross-modal search system based on Wikipedia, named **Crew**, is built and demonstrates promising search results.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Query formulation, Search process

General Terms

Algorithms, Experimentation

Keywords

Web 2.0, Wikipedia, Cross-Modal Search

1. INTRODUCTION

In the Web 2.0 age, we have witnessed the success of many folksonomy systems such as Delicious[1], Flickr[2] and Youtube[3], in which users are free to create and share diverse resources, including news, photos and videos. Moreover, users are encouraged to add extra textual terms as the description or summarization for the resources to bridge the semantic gap. Resources of different types are said to be related as long as they share identical tags, leading to a new research topic named cross-modal search. Intuitively, the cross-modal search system aims to treat the resources from different modal in a uniform manner. Both the query and the related results can be from any supported modal. In this way, the resources from different websites or of different types can be better integrated and organized to support new applications. For example, in travel applications[2], when a user is browsing a beautiful scene photo, he can submit the photo to the system and obtain the introduction of the scene, the travel tips and similar photos from other scenes. Similarly, when he is reading a travel blog, he can submit the blog to the system to get the photos of places mentioned in the article.

However, simply connecting the resources via sharing tag has its drawbacks. On one hand, tagging is a subjective process and easy to result in noise. Different users are likely to assign different tags to the same subject. On the other hand, the tag homonym(one tag with multiple meaning) and tag synonym(multiple tags with one meaning) also cause ambiguity and low-quality search results. Figure 1 shows an example of connecting resources via tags. We can see that all the resources in the figure are associated with the tag "Orange" and become related to each other. However, "Orange" could mean fruit, color, location, album, etc. Sharing the same tag does not mean real semantic relationship. For example, there is no direct connection between the fruit "orange" and the person "Orange Ferriss".

In this paper, we propose to utilize Wikipedia, the largest online encyclopedia, to build a concept layer to connect the resources. The web resources are no longer related by sharing identical tags. Instead, we map the resources to the nodes in the Wikipedia concept layer to eliminate the ambiguity of tags. The relationship between the resource and the concept is many-to-many. Each resource can be associated with multiple concept and different resources can be mapped

[1] http://delicious.com/
[2] http://www.flickr.com/
[3] http://www.youtube.com/

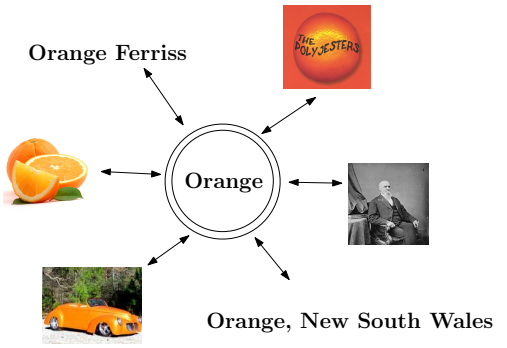

Figure 1: A resource graph connected by "Orange"

to the same concept. Only if the concept are related can we say the associated resources are related as well. Figure 2 illustrates the data model of resources with tag "Orange" based on the Wikipedia concept nodes. The resources are no longer directly connected to the tag "Orange". Instead, we add a concept middle-ware between the tag layer and the resource layer. Since "Orange" can be associated with multiple different concept, we extract the concept out and assign the resources to related concept nodes. Therefore, with the concept layer added, more semantic meaning of the resources can be captured.

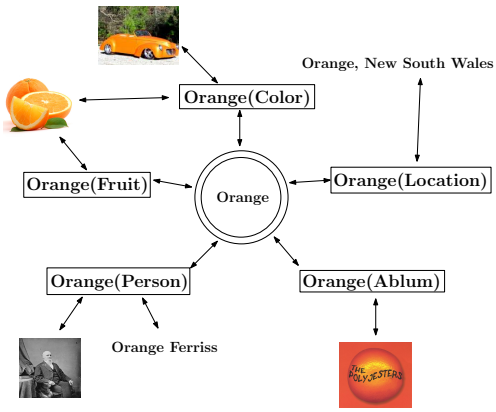

Figure 2: A resource graph connected by "Orange" based on Wikipedia

1.1 System Overview and Implementation

The architecture of our system is illustrated in Figure 3. The system consists of three layers from the bottom level to the top:

1. The crawler to collect various media resources from the Web 2.0 systems publicly accessible on the web.
2. The resource mapping and indexing component to integrate various modal resources to the concept space derived from Wikipedia.
3. The search engine to accept queries of different media types and execute the cross-modal search.

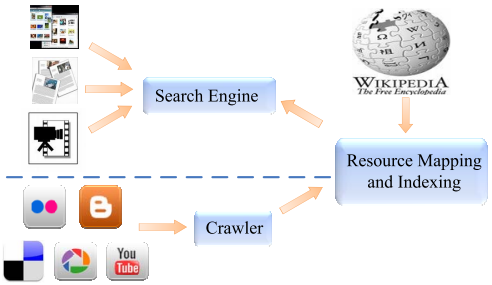

Figure 3: System Framework

We will explicitly explain each component in the following subsections.

1.2 The Crawler

Due to the proliferation of Web 2.0 social and community systems, a large number of multimedia objects with various media types are publicly available. For example, users can share and tag their photos on Flickr and Picasa Web Albums[4]. Youtube and Hulu[5] are websites to provide services for video sharing. Delicious provides web page resources well tagged. Most of these sites have published the API to facilitate the data retrieval. Otherwise, we would simply crawl the web pages using the robot and design our parser for information extraction. The objects as well as the associated tags, descriptions and other resources that can be leveraged to bridge the semantic gap are all retrieved for index.

1.3 Resource Mapping and Indexing

In our system, we utilize the Wikipedia as the schema for data storage and indexing. The Wikipedia is organized in the graph structure, with the nodes representing the individual concepts and the edges capturing the relationship between these concepts. A simple way to model the relationship is to use the hyperlink between the wiki pages. Concept A is related to concept B when A refers to B in its page content.

Each object in the database is mapped to the Wikipedia concept nodes based on their semantic similarity. The mapping process can take advantage of the textual descriptions associated with the object. If there are tags or summaries available, we adopt the uniform data model in [5] except that we ignore the spatial component in our approach. Each object is now represented as the object reference as well as the associated tags. We will first filter out the spam tags via a stoplist generated manually. For example, the words "canon" and "nikon" frequently occur in the tag list of Flickr data set. They are related to the photographic equipment of that photo instead of the semantic meaning and thus can be removed. The leftover tags are assumed to present certain semantic meaning. To make a further cleaning on the noise tags, we perform clustering on the associated tags for each object. The most relevant tags are selected and the left are discarded. Clarity score [3] is used as the distance

[4] `http://picasaweb.google.com/`
[5] `http://www.hulu.com/`

function to measure the similarity between two tags. If the keyword distribution in the retrieved results by using tags A and B to query the Wikipedia database is distinct from the overall distribution, we consider A to be similar with B. In this case, we can map the objects with high quality tags to the wiki pages. We adopt the method in [4] to measure the semantic relationship between the tags and the Wikipedia concepts.

For the media objects without any textual description, we use the content-based media retrieval methods to find the similar objects which are associated with some amounts of refined tags. These tags are retrieved and ranked to fill the semantic gap. The ranking takes into account of the similarity score, concept importance based on PageRank [1]. With the derived tags, we can map them to the Wikipedia concepts. To facilitate the query processing, we build inverted list for each concept to record all the associated media objects. The elements in the list are ordered by their relevance score.

1.4 Search Engine

Our system is designed to support various types of input, including keywords and media objects. Given a query, we first map it to the Wikipedia concepts using the same procedure described above. In this way, we can first get an initial set of concept nodes. We select the most important top-k concepts. Based on them, we also extend the concept scope to include those closely related ones which do not explicitly contain the query tags. Then, we retrieve all the resources mapped to these concepts using the inverted index. All the related media objects are ranked and returned to the user.

2. DEMONSTRATION

In our demonstration of the system, we crawled around $110K$ document from Delicious and another $140K$ images from Flickr using the published API. For current stage, our system supports three types of query input. The input can be keywords specified by users, or a textual document, or an image. The user interface is illustrated in Figure 4 and is simple and easy to use just as the general search engines. We have a textbox to accept the keywords and an upload component to support search by document or search by image. We also provide filtering options to retrieve the results users are interested. Users can choose to return only the document resources or images. Otherwise, all the related resources will be returned by default. An example of search by keywords "Formula 1" is illustrated in Figure 5. We can see that the document results from Delicious and photos from Flickr demonstrate promising accuracy.

When users are browsing the results returned by our system, they are allowed to directly submit a query using the item they are interested. This provides great convenience for users to jump across different modals and obtain more useful and interesting information.

Finally, we provide an uploading engine for users to complete our database by contributing new resources. Each time an object is uploaded, we return to users with a collection of Wikipedia concepts obtained using our method. Users can manually select the concepts they consider relevant. Given this feedback, we can insert the object into our database by mapping it to the user defined concepts. The process is human intelligence motivated and thus can improve the quality of our database. The upload interface is shown in Figure 6. The user submits an image displayed in the bottom. The returned Wikipedia concepts are mainly about "CHANGYI AIRPORT" in Singapore.

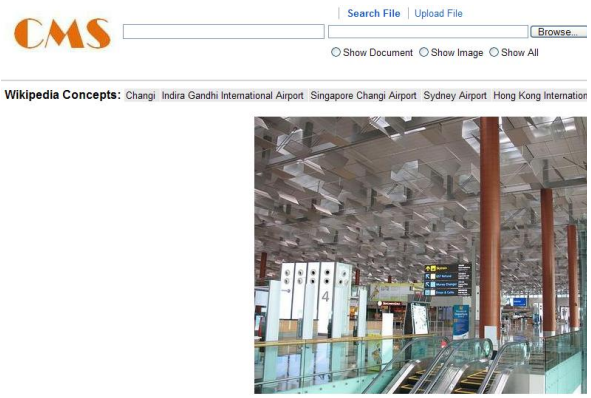

Figure 6: Upload interface

3. REFERENCES

[1] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. *Comput. Netw. ISDN Syst.*, 30(1-7):107–117, 1998.

[2] Y. Chen, S. Chen, Y. Gu, M. Hui, F. Li, C. Liu, L. Liu, B. C. Ooi, X. Yang, D. Zhang, and Y. Zhou. Marcopolo: a community system for sharing and integrating travel information on maps. In *EDBT '09: Proceedings of the 12th International Conference on Extending Database Technology*, pages 1148–1151, New York, NY, USA, 2009. ACM.

[3] S. Cronen-Townsend and W. B. Croft. Quantifying query ambiguity. In *Proceedings of the second international conference on Human Language Technology Research*, pages 104–109, San Francisco, CA, USA, 2002. Morgan Kaufmann Publishers Inc.

[4] D. Milne and I. H. Witten. An effective, low-cost measure of semantic relatedness obtained from wikipedia links. In *In Proceedings of AAAI 2008*, 2008.

[5] D. Zhang, B. C. Ooi, and A. K. H. Tung. Locating mapped resources in web 2.0. In *ICDE*, pages 521–532, 2010.

Figure 4: User Interface

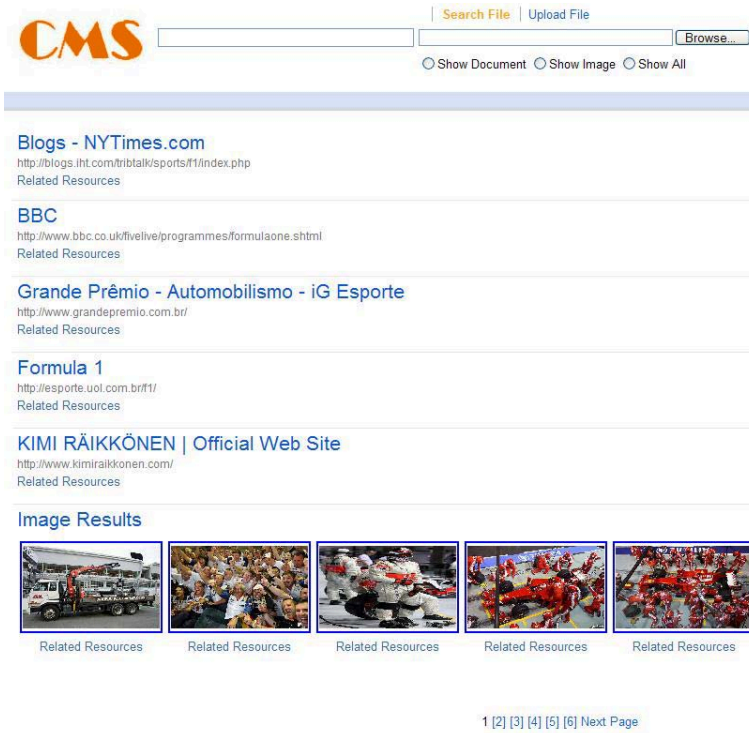

Figure 5: Example result