Mediapedia: Mining Web Knowledge to Construct Multimedia Encyclopedia

Richang Hong, Jinhui Tang, Zheng-Jun Zha
Zhiping Luo, and Tat-Seng Chua

Computing 1, 13 Computing Drive, 117417, Singapore.
{hongrc,tangjh,zhazj,luozhipi,chuats}@comp.nus.edu.sg

Abstract. In recent years, we have witnessed the blooming of Web 2.0 content such as Wikipedia, Flickr and YouTube, etc. How might we benefit from such rich media resources available on the internet? This paper presents a novel concept called *Mediapedia*, a dynamic multimedia encyclopedia that takes advantage of, and in fact is built from the text and image resources on the Web. The *Mediapedia* distinguishes itself from the traditional encyclopedia in four main ways. (1) It tries to present users with multimedia contents (e.g., text, image, video) which we believed are more intuitive and informative to users. (2) It is fully automated because it downloads the media contents as well as the corresponding textual descriptions from the Web and assembles them for presentation. (3) It is dynamic as it will use the latest multimedia content to compose the answer. This is not true for the traditional encyclopedia. (4) The design of *Mediapedia* is flexible and extensible such that we can easily incorporate new kinds of mediums such as video and languages into the framework. The effectiveness of *Mediapedia* is demonstrated and two potential applications are described in this paper.

Keywords: Web Knowledge, Multimedia Encyclopedia.

1   Introduction

The word "encyclopedia" comes from the classical Greek and was first used in the title of a book in 1954 by *Joachimus Fortius Ringelbergius* [7]. The encyclopedia as we recognize it was developed from the dictionary in the 18th century. However, it differs from dictionary in that each article in encyclopedia covers not a word, but a subject. Moreover, it treats the published article in more depth and conveys the most relevant accumulated knowledge on the subject. Thus encyclopedia is a wealth of human knowledge and has been widely acknowledged.

Most early encyclopedias are laid out using plain text with some drawings or sketches [7]. Their presentation is also somewhat plain and not so vivid. This aroused because of the initiation of print and photograph technologies at that time. In the 20th century, the blooming of multimedia technology has fostered the progress of encyclopedia. One landmark development in encyclopedia is the production of Microsoft's Encarta, which was published on CD-ROMs and supplemented with videos and audio files as well as high quality images. Recently
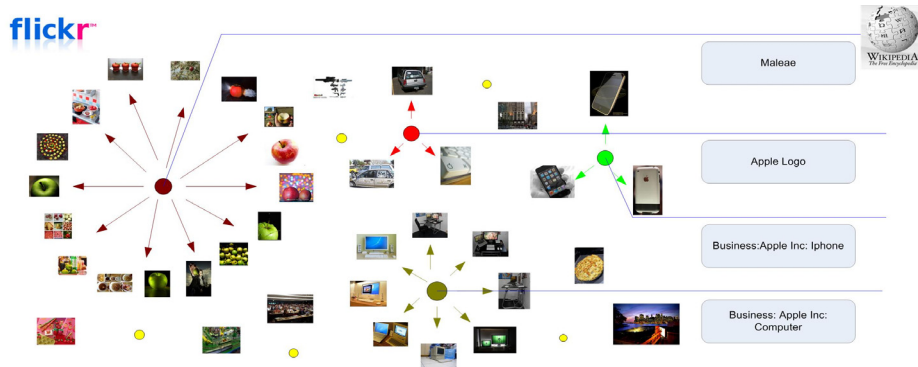
Fig. 1. The manually grouped Flickr's top 60 images accompanied with the disambiguation entries for "apple" on Wikipedia (Yellow dots denote some other images in top 60). We can see the diversity and somewhat noisy nature of Flickr images and the inherent ambiguity of the concept.

web based encyclopedias such as Wikipedia emerge by leveraging on the hypertext structure and the attributes of user contributed contents. Although there are some criticisms on Wikipedia's bias and inconsistencies, it is still the most popular encyclopedia due to the timeliness of its contents, its online accessibility and it is free of charge. To date, Wikipedia contains more than 13 million articles, of which about 2.9 million are in English[1]. Considering the success of Wikipedia, is it the ultimate form of encyclopedia, or are there other ways to construct a more interesting, useful and attractive encyclopedia?

As we know, Web 2.0 content such as Flickr, Zoomr, YouTube, etc. allow users to distribute, evaluate and interact with each other in the social network. Take Flickr as an example, it contains more than 3.6 million images as of June 2009 and many of the images are in high resolution. Thus the characteristics of Web 2.0 enrich the resources available online. Is it then possible to utilize these rich multimedia repositories to offer the dynamic meaning of concepts as well as new concepts by ways of automatically assembling them for multimedia encyclopedia? Actually projects such as Everything, Encarta and Wikipedia have included some images, audios, and even videos. However, they only appear in a limited number of but not all entries, and they may not be the latest and most representative. Moreover, the presentation is somewhat tedious and unattractive as it focuses mostly on textual description with multimedia contents used mostly as illustrations.

In this paper, we propose a multimedia encyclopedia called *Mediapedia* which is automatically produced and updated by leveraging on the online Web 2.0 resources. The novel form of encyclopedia interprets the subject in a more intuitive and vivid way. The key characteristics of *Mediapedia* that distinguishes it from other encyclopedias are that: (1) the presentation is in form of video; (2) it is

---

[1] http://en.wikipedia.org/wiki/Wikipedia
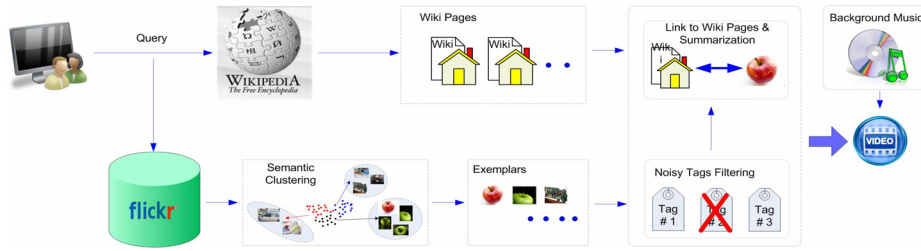
Fig. 2. The system framework of the proposed *Mediapedia*. It mainly consists of: (1) image clustering for producing exemplars; (2) association of exemplars to Wikipedia and (3) multimedia encyclopedia presentation.

fully automatically produced, (3) it is updated dynamically; and (4) the whole framework is flexible and extensible that facilitates more potential applications. Through *Mediapedia*, the users can choose to view the concept based on its most "common" meaning or in "diverse" form, which affects the duration of presentation for the concept. When users input a query, the system first crawls the diverse images from Flickr and generates the exemplar images; it then associates the exemplars with Wikipedia summary after noise tag filtering; and finally it automatically produces the multimedia encyclopedia for the concept with synchronized multimedia presentation.

Although *Mediapedia* is promising and desirable, we have to face many challenges. As an example, Fig. 1 illustrates the manually grouped Flickr's top 60 images accompanying with disambiguation entries for the concept "apple" on Wikipedia. We can see that the retrieved images are diverse and somewhat noisy. The disambiguation page on Wikipedia identifies different senses or sub-topics of the concept. Thus we have to find the exemplars from the piles of images and associate them with the corresponding concise Wiki description. We conclude the challenges as follows. (1) How to make the tradeoff between "typicality" and "diversity". In other words, face with the list of retrieved images, which ones are more typical for characterizing the concept and to what extent they are sufficient for showing the diversity. (2) Where and how to discover the corresponding textual contents and prune them to describe the image exemplars. (3) How to present the multimedia content (e.g., text, image and audio) to ensure coherence and elegance of the multimedia encyclopedia. (4) The final and most important challenge is why do we do this work? Are there any potential applications based on this work? In next section, we will answer the "how to construct" question to tackle the challenges (1), (2) and (3). Section 3 evaluates the performance of *Mediapedia*. We describe several potential applications in Section 4, and conclude the paper in Section 5.

2   How to Construct

This section describes the system framework and the algorithms involved. Figure 2 illustrate the proposed framework for *Mediapedia*. We first elaborate on the

image clustering for producing the exemplars. We then discuss the association of exemplars with Wikipedia that aims to associate the exemplars with user contributed contents on Wiki [9][17][18]. We finally assemble the exemplars and the concise descriptions to produce the multimedia encyclopedia, where images, transcripts and the background music are presented in an attractive and vivid way. We describe the detailed algorithms in the following subsections.

2.1   Image Clustering for Producing Exemplars for Concepts

Considering the attributes of images from Flickr, a question naturally arise is how to efficiently present the representative images for the concept to users? Some works have been proposed to organize the retrieved images into groups for improving user experiences[1][3]. However, these works are based on traditional clustering algorithm, and although they can produce a more organized result, how to present the clusters to users is still challenging. The studies on finding exemplars from piles of images can be seen as a further step to solving the problem, where the most popular way may be the k-centers algorithm such as [3]. Frey *et al.* proposed affinity propagation to discover exemplars from a set of data points and it has been found to be more effective than the classical methods [4]. It can also be considered as an effective attempt to tackle the problem of finding image exemplars [12][5], which is consistent with the first of the four challenges. Here, we take advantage of the Affinity Propagation (AP) algorithm [4] to acquire exemplars for presentation.

We denote a set of n data points as $\mathcal{X} = \{x_1, x_2, \cdots, x_n\}$ and the similarity measure between two data points as $s(x_i, x_j)$. Clustering aims at assembling the data points into $m(m < n)$ clusters, where each cluster is represented by an "exemplar" from $\mathcal{X}$. Two kinds of messages are propagated in the AP algorithm. The first is the "responsibility" $r(i, k)$ sent from data point i to data point k, which indicates how well k serves as the exemplar for point i taking into account other potential exemplars for i. The second is the "availability" $a(i, j)$ sent from data point k to data point i, which indicates how appropriate for point i to choose point k as exemplar taking into account the potential points that may choose k as their exemplar. The messages are iterated as:

$$r(i, k) \leftarrow s(i, k) - \max_{k' \neq k}\{a(i, k') + s(x_i, x_{k'})\}, \tag{1}$$

$$a(i, k) \leftarrow \min\{0, r(k, k) + \sum_{i' \notin \{i,k\}} \max\{0, r(i', k)\}\}. \tag{2}$$

where the self-availability is updated in a slightly different way as:

$$a(k, k) \leftarrow \sum_{i' \neq k} \max\{0, r(i', k)\}. \tag{3}$$

Upon convergence, the exemplar for each data point x_i is chosen as $e(x_i) = x_k$ where k maximizes the following criterion:

$$\arg\max_{k} a(i, k) + r(i, k) \tag{4}$$

2.2   Association of Exemplars to Wikipedia Pages

The AP algorithm facilitates the finding of exemplars for each cluster. Given the exemplars, we have to face the second challenge of where and how to obtain the corresponding textual description and prune it for those exemplars. As we know, most uploaded images contain a large number of user contributed tags in social media network. However, the tags tend to contain a lot of noise, and are inadequate to help users understand the inherent meaning of the images. An intuitive way is to associate the exemplars with Wikipedia by leveraging the exemplar's tags. We thus need to remove the noisy tags first and then analyze the correlation of remaining tags of each exemplar with its corresponding Wiki pages. We also need to prune the Wiki pages by summarization techniques to produce a brief description in *Mediapedia*.

Noisy Tag Filtering Since we aim to describe the exemplars with their corresponding text descriptions in Wikipedia by leveraging on the tags of the exemplars, the quality of the tags should therefore be of high quality. In other words, we need to remove insignificant tags such as typo, number, model ID and stop-words, etc., from the tag list. WordNet[2] is a popular lexical database and has been widely used in eliminating noisy tags [8]. Here we list the tags in their respective word group and removing those tags that do not appear in WordNet. We denote the tags after noise filtering as $\mathcal{T} = \{t_{ij}, 1 \leq i \leq m, 1 \leq j \leq N(t_i)\}$, where j indicates the tag in the group t_i and $N(t_i)$ indicates the total number of tags in t_i.

We then utilize the Normalized Google Distance (NGD) [10] between the concept and its associated tags in each cluster as a metric for the semantic relationship between them. Since NGD is a measure of semantic inter-relatedness derived from the number of hits returned by Google search engine, it can be used to explore the semantic distance between different concept-tag pairs. Given the concept q and tag t_{ij}, the NGDbetween them is defined as:

$$ngd(q, t_{ij}) = \frac{\max\{\log f(q), \log f(t_{ij})\} - \log f(q, t_{ij})}{\log M - \min\{\log f(q), \log f(t_{ij})\}} \tag{5}$$

Here, M is the total number of retrieved web pages; $f(q)$ and $f(t_{ij})$ are the number of hits for the concept q and tag t_{ij} respectively; and $f(q, t_{ij})$ is the number of web pages containing both q and t_{ij}. We then use the NGD metric to rank the tags with respect to the concept q. Note that some tags may appear in more than one cluster since a concept usually possesses a variety of presentations and the AP algorithm only groups the images with feature similarity.

Linking to Wikipedia As aforementioned, for each concept, we obtain several exemplar images via clustering the retrieved Flickr images. We then collect the corresponding documents from Wikipedia by Wiki dump[3]. By considering these exemplar images as characterizing the various aspects of the query in a visual

---

[2] http://wordnet.princeton.edu/
[3] http://en.wikipedia.org/wiki/Dump

manner while the Wiki documents provide textual descriptions of different senses of a concept, we thus argue that the seamless combination of the visual and textual description can help users better comprehend the concept[15]. However, it is nontrivial to automatically associate each image with the corresponding document.

Here we resort to Latent Semantic Analysis (LSA) method to calculate the similarity between each image and the documents, while the similarity is in turn used to associate the images with the documents [14]. LSA is an approach for automatic indexing and information retrieval that maps document as well as terms to a representation in the so-called latent semantic space. The rationale is that documents which share frequently co-occurring terms will have a similar representation in the latent space. We give the necessary mathematical equations and methodology for LSA here. More detailed theoretical analysis can be found in [14].

LSA represents the association of terms to documents as a term-document matrix:

$$\mathcal{R} = \begin{pmatrix} r_{11} & \cdots & r_{1n} \\ \vdots & \ddots & \vdots \\ r_{m1} & \cdots & r_{mn} \end{pmatrix} = [d_1, \cdots, d_n] = [t_1, \cdots, t_m]^T \tag{6}$$

where element r_{ij} denotes the frequency of term i occurs in document j and properly weighted by other factors [16]. d_i is the i-th column of $\mathcal{R}$ corresponding to the i-th document, and t_j is the j-th row of $\mathcal{R}$ corresponding to the j-th term. The documents and terms are then re-represented in a k-dimensional latent vector space. This is achieved using the truncated singular value decomposition as follows:

$$\mathcal{R} = \mathbf{U}_k \mathbf{\Sigma}_k \mathbf{V}_k{}^T \tag{7}$$

where $\mathbf{U}_k$ and $\mathbf{V}_k$ are orthonormal matrices consisting of the left and right singular vectors respectively, and $\mathbf{\Sigma}_k$ is a diagonal matrix containing the singular values.

Given a concept q consisting of the tags of a image, its corresponding representation in k-dimensional latent semantic space is $q^T \mathbf{U}_k$, and the documents are represented as columns of $\mathbf{\Sigma}_k \mathbf{V}_k{}^T$. The similarities between the concept and documents are calculated as:

$$sim = (q^T \mathbf{U}_k)(\mathbf{\Sigma}_k \mathbf{V}_k{}^T) \tag{8}$$

where the i-th column of sim is the similarity between the concept and the i-th document.

2.3   Encyclopedia Presentation and User Interface

Given the exemplars and the associated Wiki pages, the next problem is how to present the multimedia content to ensure coherence and elegance of the multimedia encyclopedia. Since the articles from Wikipedia are highly structured and cover as many sub-topics about the concept as possible, it is infeasible to
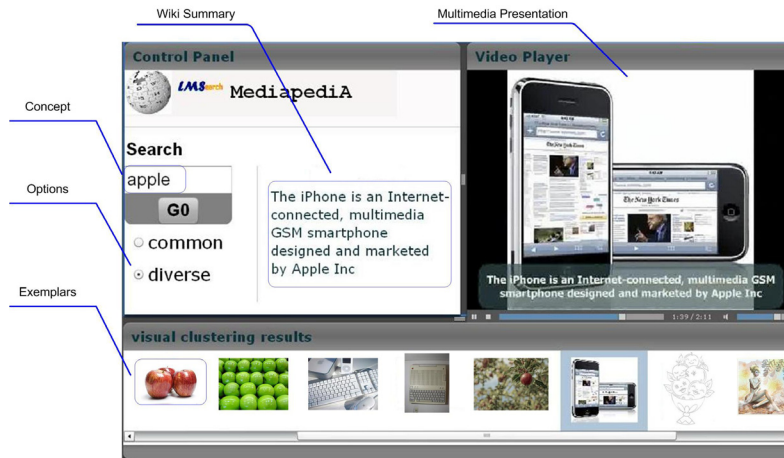
Fig. 3. The user interface of *Mediapedia*. The five annotations indicate the functionalities of the corresponding parts.

directly incorporate the Wiki content into *Mediapedia*. A concise summary of the concept would be more desirable.

There are many methods focusing on the summarization of individual document or multiple documents. In the scenario of web documents, the performance of those methods can be improved by leveraging the hypertext structure including anchor text, web framework etc. Ye *et al.* proposed to summarize the Wikipedia pages by utilizing their defined Wiki concepts and inforbox structure through the extended document concept lattice model [6]. Moreover, it produces summary with various lengths to suit different users' needs. We therefore employ the summarization method in [6] to provide a concise description for the concept where the length of summary can be controlled by the users.

In the presentation of encyclopedia, we claim that video or well structured audio visual presentation is more attractive and vivid as compared to plain text even with hypertext links to sound, images and motion. In this study, we employ the APIs from *imageloop*[4] for presenting the multimedia encyclopedia by way of slideshow which consecutively displays the images ordered by the size of clusters. The display effect is constrained to the options of *imageloop*. In order to demonstrate the performance of exemplars generation and Wiki page association, we design the user interface to show the composed slideshow accompanying with the exemplars and text summary from Wiki on the same panel. Figure 3 illustrates the user interface. We embed background music to *Mediapedia* for enhancing its presentation. In our rudimentary system, the background music is randomly selected from a pool of *Bandari Rhythms*[5].

---

[4] http://www.imageloop.com/en/api/index.htm
[5] http://en.wikipedia.org/wiki/Bandari_music

3   Evaluation of Mediapedia

In this section, we briefly evaluate the performance of the generation of exemplars and their links to Wikipedia pages respectively. More detailed analysis of the performance on individual components can be referred to the literatures [4][11][6][14]. We choose 20 concepts, most of which are from the "NUS-WIDE" concept list [8] and crawl the top ranked 1000 returned images for each concept. For simplicity, we adopt the 225-D block-wise color moment as the visual features. The similarity between two images is considered as the negative distance between their feature vectors [4]. The damping factor in AP algorithm is set to 0.5 and all the preferences, i.e., the diagonal elements of the similarity matrix are set to the median of $s(i,k), i \neq k$. After clustering, we rank the clusters according to their size. Figure 4(a) illustrates the top 12 exemplars for the concept "apple" generated by the AP algorithm.


(a)   (b)

Fig. 4. Evaluation on the modules. (a) the top 12 exemplars for the concept "apple" generated by affinity propagation. (b) two exemplar images for the concept "apple" and their associated Wiki summaries with the option of two sentences.

For the association of exemplars to Wiki pages, we download the Wikipedia entries for each concept as well as the disambiguation entries for those concepts that contain ambiguity. Here, the concepts are directly used as queries for downloading. We utilize NGD to filter some irrelevant tags with respect to the concept. We then apply LSA to the relationship matrix constructed by the tags and the documents. For each image, the value of its relevance to Wiki pages is accumulated by the normalized similarity between each tag-document pair, which is similar to the metric of Normalized Discounted Cumulative Gain (NDCG). Note that if the concept is without ambiguity (by referring to Wikipedia), the exemplar images and Wiki pages are directly associated. For concise presentation, Wiki pages are summarized by the method in [6]. Figure 4(b) illustrates the two exemplar images associated with the summary from Wiki pages. Note that the number of the sentences is optional and here we show the summaries with two sentences only. The tags in yellow font are removed by the step of noisy tag filtering.

We briefly introduce the process of testing and present some experimental results for the concept "apple". Since the individual component has been evaluated in other works[4][11][6][14], we only briefly test the overall performance of this system. In this paper, we define three metrics for user based evaluation: 1) experience, do you think the system is interesting and helpful? 2) informative,
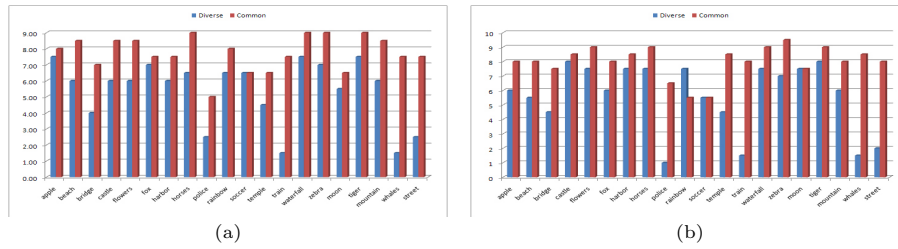
Fig. 5. User based evaluation on the overall performance: (a) experience and (b) informative.

do you feel the system can properly describe the concept (taking account of both "typicality" and "diversity")? Four students are involved in the user study. For the metrics, each student is asked to score the system performance on 10 levels ranging from 1 (unacceptable) to 10 (enjoyable). Figure 5 demonstrates the user based evaluation. We can see that the overall performance is not as good as expected and should be improved further. On the one hand, the images from Flickr are too diverse and noisy to be ideally grouped. On the other hand, the rational of our study is based on the assumption that the distribution of images from Flickr can automatically make a tradeoff between "typicality" and "diversity". However, in some cases, this assumption may not agree with the truth. Thus we can see for the option of "common", the performance is acceptable while should be improved for the option of "diverse".

4 The Potential Applications

This section answers the final challenge about why we do this work. Similar to encyclopedia which has been widely used all around our lives, *Mediapedia* can also be applied to many fields: as the coinage *Mediapedia* indicates, we can lookup *Mediapedia* for interpretation with multimedia form. Imaging the scenario of child education, such a vivid and colorful encyclopedia would be very attractive. Here, we simply present another two potential applications to show how we can benefit from *Mediapedia*.

4.1 Definitional Question-Answering

Definitional QA in multimedia is emerging following the blooming of internet technology and the community contributed social media. Chua *et al.* presented a survey on Multimedia QA and the bridge between definitional QA in text and multimedia [2]. Analogous to textual definitional QA, it is a way of Multimedia QA by incorporating some defined key-shots, which equate the sentences in textual QA and indicate the key sub-topics of the query, into the answers. The idea has been approved successfully in event driven web video summarization, i.e., multimedia QA for events [11]. *Mediapedia* can be viewed as another way for multimedia QA.

The attributes of intuitiveness and vividness of *Mediapedia* facilitate the application in multimedia QA. Compared to the process of determining key-shots first and then combining them into the answer, the way of *Mediapedia*

would be more direct and flexible. Moreover, we can apply some image processing techniques such as the film digital effects to produce more coherent and elegant video answers.

4.2   Tourist Site Snapshot

There are a huge number of images related to some popular tourist sites on Flickr. Moreover, more photos are taken on some landmark buildings or scenes. Take "Paris" as an example, the "Eiffel Tower" can be deemed as one of the famous landmarks. Thus in our system, the images from those sites will be grouped into comparatively larger clusters by the AP algorithm. After association to the entry on Wikipedia, the produced slideshow will comprise of views of the popular landmark and their corresponding descriptions. This will provide informative snapshots of the tourist sites to the users.

It is more exciting that many images from the social sharing websites offer geo-information which specifies the specific world location of the photo taken [13]. Considering the fact that Flickr alone carries over one hundred million geo-tagged photos (out of a total of 3 billion photos), the use of these images would improve the performance of selecting the exemplar images and provide better tourist site snapshots, even tourism recommendation.

5   Discussions and Conclusions

This paper presented a novel concept named *Mediapedia* which aims to construct multimedia encyclopedia by mining web knowledge. The *Mediapedia* distinguishes itself from traditional encyclopedia in its multimedia presentation, full automated production, dynamic update and the flexible framework where each module is extensible to potential applications. In the proposed system, we employed the AP algorithm in producing the exemplars from image pool, while using LSA to associate exemplars to Wiki pages and utilizing document lattice model to perform Wiki pages summarization. We finally assembled them for multimedia encyclopedia. Two potential applications are described in detail.

This study can be deemed as an attempt at constructing *Mediapedia* by leveraging on web knowledge. The experimental results, however, were not as good as expected and should be improved further. This may be aroused by the assumption that the distribution of images from Flickr can automatically make a tradeoff between "typicality" and "diversity". However, this is not the truth for all the concepts. Improvement can be made by taking into account the tags in producing the exemplar, and leveraging the images embedded in Wikipedia to facilitate better association and so on. An alternative approach to tackle the problem is to start our *Mediapedia* from Wikipedia, by identifying different senses of the concept by Wikipedia first and then associating them with images/audios. Thus, the framework proposed in this study is still evolving. The main contribution of this work can be seen as an attempt to construct multimedia encyclopedia by mining the rich Web 2.0 content. Moreover, we provide an interesting system which performance is acceptable. In future works, we will improve the performance of *Mediapedia* by incorporating interactivity and exploring another design alternative.

References

1. Cai Deng, He Xiaofei, Li Zhiwei, Ma Wwei-Ying, and Wen Ji-Rong: Hierarchical clustering of WWW image search results using visual, textual and link information. Proc. of ACM MM'04, (2004) 952-959
2. Chua Tat-Seng, Hong Richang, Li Guangda, Tang Jinhui: From Text Question-Answering to Multimedia QA. To appear in ACM Multimedia Workshop on Large-Scale Multimedia Retrieval and Mining (LS-MMRM) (2009)
3. Gonzalez T.: Clustering to minimize the maximum intercluster distance. Theoretical Computer Science. (1985) 38(2):293-306
4. Frey B. and Dueck D.: Clustering by Passing Messages Between Data Points. Science, (2007) 315(5814):972
5. Jia Yangqing, Wang Jingdong, Zhang Changshui, Hua Xian-Sheng: Finding Image Exemplars Using Fast Sparse Affinity Propagation. ACM International Conference on Multimedia, ACM'MM 08, Vancouver, BC, Canada. (2008)
6. Ye Shiren, Chua Tat-Seng, Lu Jie: Summarization Definition from Wikipedia. ACL-IJCNLP. Singapore, 2-7, August. (2009)
7. Sorcha Carey: Two Strategies of Encyclopaedism. Pliny's Catalogue of Culture: Art and Empire in the Natural History. Oxford University Press. (2003)
8. Chua Tat-Seng, Tang Jinhui, Hong Richang, Li Haojie, Luo Zhiping, Zheng Yantao: NUS-WIDE: A Real-World Web Image Databased From National University of Singapore. ACM International Conference on Image and Video Retrieval. Greece. Jul. 8-10, (2009)
9. Wang Meng, Hua Xian-Sheng, Song Yan, Yuan Xun, Li Shipeng, Zhang Hong-Jiang: Video Annotation by Semi-Supervised Learning with Kernel Density Estimation, ACM'MM, (2006)
10. Cilibrasi Rudi and Vitanyi Paul: The Google Similarity Distance, ArXiv.org or The Google Similarity Distance, IEEE Trans. Knowledge and Data Engineering, 19:3 370-383. (2007)
11. Hong Richang, Tang Jinhui, Tan Hung-Khoon, Yan Shuicheng, Ngo Chong-Wah, Chua Tat-Seng: Event Driven Summarization for Web Videos. To appear in ACM Multimedia Workshop on Social Media. (WSM) (2009)
12. Zha Zheng-Jun, Yang Linjun, Mei Tao, Wang Meng, Wang Zengfu.: Visual Query Suggestion. To appear in ACM'MM, Beijing, China, (2009)
13. Zheng Yan-Tao, Zhao Ming, Song Yang, Adam Hartwig, Buddemeier Ulrich, Bissacco Alessandro, Brucher Fernando, Chua Tat-Seng, Neven Hartmut: Tour the World: building a web-scale landmark recogntion engine. In Proceedings of CVPR'09, Miami, Florida, U.S., June 20-25, (2009)
14. Ding C. H. Q.: A Similarity-Based Probability Model for Latent Semantic Indexing, Proc. of SIGIR, (1999)
15. Li Haojie, Tang Jinhui, Li Guangda and Chua Tat-Seng: Word2Image: Towards Visual Interpretation of Words. In Proceedings of ACM'MM 08. (2008)
16. Dumais S.: Improving the retrieval of information from external sources. In Behavior Research Methods, Instruments and computers, 232, 229-236, (1991)
17. Wang Meng, Yang Kuiyuan, Hua Xian-Sheng, Zhang Hong-Jiang: Visual Tag Dictionary: Interpreting Tags with Visual Words. To appear in ACM Multimedia Workshop on Web-Scale Multimedia Corpus. (2009)
18. Tang Jinhui, Yan Shuicheng, Hong Richang, Qi Guojun, Chua Tat-Seng: Inferring Semantic Concepts from Community-Contributed Images and Noisy Tags. To appear in ACM'MM09. (2009)