iULib: Where UDL and Wikipedia Could Meet

Yonghong Tian, Tiejun Huang, Wen Gao

School of EE & CS, Peking University, Beijing 100871, China

ABSTRACT

Empowering the group collaboration and knowledge-sharing capabilities for the Universal Digital Library (UDL) is definitely an important work after more than 1.5 million digitalized books were open to access online. One motivation of developing such a platform is the emergence of Web 2.0 in recent years, especially with the rapidly increased popularity of Wikipedia. This paper presents our vision, which we call *iULib*, about where and how UDL and Wikipedia could meet. In the first phase, we directly apply the Wiki architecture and software in UDL to upgrade the digital library as an interactive platform that facilitates community and collaboration. Preliminary implementation shows the feasibility and reliability of our design. Furthermore, as a free encyclopedia that assembles contributions from different users, Wikipedia may also be used as a knowledge base for UDL. As a result, UDL can be upgraded as an intelligent platform for information retrieval and knowledge sharing. Our practice at the WikipediaMM task in the ImgeCLEF 2008 shows that the knowledge network constructed from Wikipedia can be used to effectively expand the query semantics of image retrieval. It is expected that Wikipedia and digital library can integrate each other's valuable results and best practices to benefit each other.

Keywords: Digital library, Wikipedia, Group collaboration, Knowledge sharing, Knowledge network, Query expansion, iULib

1. INTRODUCTION

There are more than 1.5 million digitalized books (e-books for short) in the Universal Digital Library (UDL). When such a huge collection was open to access online since Nov. 2007, the first goal of Million Book Project (MBP) had been realized successfully. Clearly, one million digitized books are not the ultimate goal of the project, but a first step towards universal access to human knowledge.[1] Towards this end, UDL should constantly improve its collaboration and knowledge-sharing capabilities, and be gradually upgraded from a "digitalized form of traditional library" to a "knowledge library."

However, the traditional UDL platform was merely to provide services like metadata-based or text-based book retrieval and book browsing, which to some extent restricted the exploitation of the full potential of UDL. With such a simple service pattern, UDL faces at least two challenges: 1) E-books are simply presented in UDL as isolated units. In fact, an ideal model of digital library is network model, in which books are associated for better services. For instance, when a user input a query "digital library technology," a good service platform should not only return the books and chapters that directly match the query, but also recommend books of related subjects such as "book digitization" and "book retrieval." 2) Often, the typical usage pattern in UDL is book retrieval. That is, users input the retrieval request and get books that they need. No user participation and group collaboration are involved in UDL. It is clear that UDL should not only be a large collection of e-books, but also a user community for knowledge sharing.

To address these challenges, UDL should further improve its service platform. As a matter of fact, we can also see that more and more libraries are taking up these challenges by creating social platforms for knowledge sharing both in physical and digital spaces, e.g., the library catalog where users can create, add, tag, remix, recommend and share resources.

In recent several years, we have made some efforts to improve the UDL service platform. We designed three experimental systems that enhance book presentation patterns and integrate user interaction: (a) IQuery, a novel e-book search system that supports multi-granularity e-book retrieval with more retrievable units and multi-facet navigation. It extracted a knowledge network from e-books and implemented a powerful information visualization module for book navigation; (b) Illustrator, a semantic illustration retrieval system from a very large data set that contains about 8.7 million illustrations; (c) KnowMap, a visual, hierarchical e-book browsing

Imaging and Printing in a Web 2.0 World II, edited by Qian Lin, Jan P. Allebach, Zhigang Fan, Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 7879, 78790G · © 2011 SPIE-IS&T · CCC code: 0277-786X/11/$18 · doi: 10.1117/12.876534

and retrieval system based on topic map. However, we haven't solved the problem of users' strong participation and collaboration. Despite that the knowledge network extracted from books are visualized for users' query expansion, it is still far from users' prospect.

This paper reports our recent work and future vision on how to combine UDL with Wikipedia. To a large extent, our research is enlightened by the concept of Web 2.0. Compared with Web 1.0, Web 2.0 is marked by more user participation in various ways. Exactly motivated by Web 2.0, "library 2.0" has been proposed as a new generation of library. Some researchers (e.g., Ref. 2) also have tried to take it into practice. However, it is not clear how Web 2.0 and digital library integrate each other's valuable results and best practices to benefit each other. Therefore, in this paper we will try to explore where and how UDL and Web 2.0 could meet by using an experimental system, *iULib*, as the illustrative example. By integrating UDL and Wikipedia, an incentive application of Web 2.0 for users' strong involvement and knowledge sharing, iULib can be used as both an interactive platform that facilitates community and collaboration, and an intelligent platform for information retrieval and knowledge sharing in UDL.

The rest of this paper is organized as follows: In Section 2, we discuss the possible interaction of Web 2.0 and Wikipedia with UDL. In Section 3 and 4, we introduce iULib 1.0, a UDL Wiki system, and iULib 2.0 where Wikipedia serves as a knowledge base for UDL, respectively. Future work is discussed in Section 5. Finally, Section 6 concludes this paper.

2. WHAT WIKIPEDIA COULD GIVE TO UDL

Web 2.0 refers to an updated and improved version of WWW that allows the users to communicate, collaborate, and share information online in completely new ways. The central point of Web 2.0 is the user's stronger involvement and participation in the Web, forming social networks and virtual communities in a global network. Key Web 2.0 applications include blogging, Wiki, tagging and social bookmarking, multimedia sharing, content syndication.[3]

With over 2.3 million entries, Wikipedia is a typical Wiki system that allows users to view, edit and discuss term pages in the online encyclopedia. It's marked by the following features:

(1) An online encyclopedia that assembles human knowledge of all fields. It differs from online dictionaries which explain terms and forums by exploiting user participation in simple means.

(2) Well Organized. Articles are structurally parallel with links between each other. The structure of hypertext is continuously changing and evolving.

(3) Openness. Following the GNU license, the content of Wikipedia is free to be replicated, edited and distributed partially or fully. Pages can be easily maintained with simple format and tagged links. Moreover, the evolvement of Wikipedia pages is observable. Users are free to access history of content changes.

2.1 Related work on applying Wiki into libraries

Interestingly, we now see trends to add Wiki and other Web 2.0 applications to traditional libraries.[4] OCLC employed open software WikiD to build a Wiki system that supports capture of user contributions such as reviews and table of contents to associate with Open WorldCat entries. Frumkin[5] discussed how collaborative tools such as Wikis, could be utilized in digital libraries, then proposed three potential applications of digital library Wiki: the Wiki as a knowledge base tool, as a content management tool and as a tool to empower interactive finding aids. Lim et al.[2] studied the task of extracting metadata from Wikipedia articles and integrating Wikipedia in Web-based geography digital library. However, these researches are either exploiting Wiki applications in simple tasks or have not been implemented.

2.2 Our solution of combining UDL with Wikipedia

This paper presents our vision about where and how UDL and Wikipedia could meet. In the first phase, we directly apply the Wiki architecture and software in UDL to upgrade the digital library as an interactive platform that facilitates community and collaboration. This system, named *iULib* 1.0, holds thousands of nodes each of which represents one e-book. Such a system provides a platform where users can interact and share knowledge

with each other. It is imaginable that UDL would be enhanced by using the Wiki architecture. Preliminary implementation shows the feasibility and reliability of our design.

Furthermore, as a free encyclopedia that assembles contributions from different users, Wikipedia may also be used as a knowledge base for UDL. The characteristics of Wikipedia are not only the ability of covering various fields but also a dense link structure. Therefore it's possible to construct a knowledge base, such as thesaurus, from Wikipedia via NLP technologies and link mining strategies. By using the knowledge base constructed from Wikipedia, UDL can be upgraded as an intelligent platform for information retrieval and knowledge sharing. For UDL, the knowledge base extracted from Wikipedia can be utilized in many aspects. Especially, query expansion techniques can be used with the knowledge base to improve book retrieval performance. Our practice at the WikipediaMM task in the ImgeCLEF 2008 shows that the knowledge network extracted from Wikipedia can be used to effectively expand the query semantics of image retrieval. The Wikipedia knowledge network can also serve as a supplement for metadata-based TermNet[6] and as a navigation tool.

3. IULIB 1.0: APPLYING WIKI ARCHITECTURE IN UDL

To enhance the group collaboration and knowledge-sharing capabilities of UDL, we propose a solution that transfers digital collections to Wiki communities, where book information pages are available for anyone who accesses the book.

3.1 Implementation of iULib 1.0 using MediaWiki

We referred iULib 1.0 to be a system that could hold millions of records and permit random users to modify its entries and keep a full history of their changes. By transferring our digital collection to Wiki software system, all authorized users are capable to edit the content. And more importantly, historical changes are stored to cope with data restoring and reversion mining tasks. Subsystems in different languages can also be created with contribution from users. Like multiple language editions of Wikipedia, users can translate a page of book information into their preferred languages or even create subsystems in new languages.

So far, we implemented iULib 1.0 (with more than 150 thousands English e-books at North Technical Center of MBP in China) using MediaWiki software (http://www.mediawiki.org) which is widely employed in building Wiki sites. MediaWiki has a specific page source code style that defines the page links, page structures and text formats. Therefore, we converted the book metadata from UDL database to text that satisfies MediaWiki page source code style, and then used a batch processing tool to import these texts and book cover images.

For each e-book in UDL, we created one Wiki page which consists of the following parts:

(1) *Title*: Title of the e-book, including both title and subtitle.

(2) *Metadata*: Metadata of the e-book, including title, author, subjects, language, publish year, publisher, country, page number, etc. In MBP, they were manually produced during the digitization process according to the USMARC21 entries of Library of Congress (LC) or OCLC. Some semantic-relevant fields, such as field 650 and its subfields, are visualized in the e-book Wiki page. Meanwhile, DC metadata stored in TOC.xml file is also available for each e-book.

(3) *Summarization*: In iULib 1.0, there is no summarization for any e-book initially. Nevertheless, users or automatic summarization technologies can generate the summarization for each e-book.

(4) *Navigation*: Table of content or even full text of the e-book is shown here for users to access the e-book resources. Here table of content is a hierarchical structure of the e-book extracted from TOC.xml. Meanwhile hyperlinks are provided to refer to the full text of the e-book. Currently, UDL 1.0 merely links to the corresponding pages of e-books at North Technical Center of MBP in China. However, it is easy to develop a unified local e-book reader with standardized API of security and access control to automatically locate and access e-books with various data formats and stored in all UDL databases.

(5) *Link to similar books*: Hyperlinks to similar e-books. Similar books here can be obtained by using two methods: a) Index all books' metadata and find similar e-books with metadata (such as title); b) Select correlated e-books from BookNet constructed in our previous work.[6]

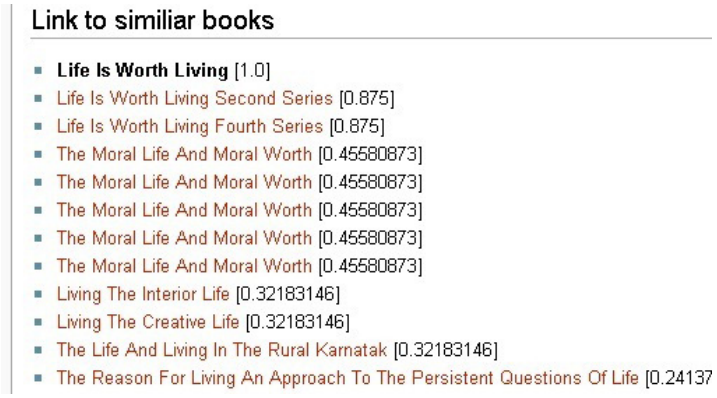

Figure 1. Similar e-books found for "Life is Worth Living" by Searching titles.

Given an e-book set, the key issue of finding similar e-books is to define an appropriate similarity measure between e-books. Vector Space Model (VSM) represents each e-book b_i as a vector $\mathbf{v}_i$ of feature values. Thus the similarity between two e-books can be calculated by the cosine value of two vectors. In iULib 1.0, we used TFIDF to calculate the feature values of vectors.

For example, when only the titles of e-books are used, it is a simple and cost-effective to find similar e-books by calculating cosine similarity between titles of e-books. Fig. 1 shows such an example of found similar e-books for the given book "Life is Worth Living."

In our previous work, BookNet was used to quantitatively represent the association between e-books.[6] Naturally, the BookNet can be directly used to find similar e-books in iULib 1.0. For more details about the construction of BookNet, please refer to Ref. 6.

Note that the two methods can be combined to archive overall similarity between books. That is,

$$Sim_F(b_i, b_j) = \omega_{VSM} \times Score_{VSM}(b_i, b_j) + \omega_{Net} \times Score_{Net}(b_i, b_j), \quad (1)$$

where ω_{VSM} and ω_{Net} are the combination weights.

(6) *Link to author(s)*: Author(s) profiles and associated links. As initial information here may be incomplete, Wiki community users can contribute to this part. The Wiki source converter can produce links to those authors of the e-books though the authors' information pages may be there waiting for users to create and edit. These links may appear as a form like "Author: James Martin" which shows that these entities are defined in a special namespace and may follow a template.

(7) *Copyright Status*: It's clarified in this part whether the current book is free to access or reuse. Copyright status information here contains the copyright owner, copyright reason and copyright expiry date.

(8) *Review*: Users can post their comments and recommendations for the e-book in this part.

3.2 Characteristics of iULib 1.0

We emphasis, herein, that iULib 1.0 has the following major characteristics:

(1) *Integration of user participation.* iULib 1.0 is different with most digital libraries as it makes users involved in various aspects including data creation, content contribution, as well as their comments.

(2) *Dynamic data collection.* Since users can edit the page source and contribute to the system, the data collection would not be just the original e-books, but a dynamic data collection which contains information that could never been produced in the digitization process, such as author profile pages and the summarization of e-books.

(3) *Strong semantic correlations.* A dense link structure is a common feature of Wiki community. Thereby it is easy to understand that pages in iULib 1.0 are connected with links which stand for semantic correlations. For

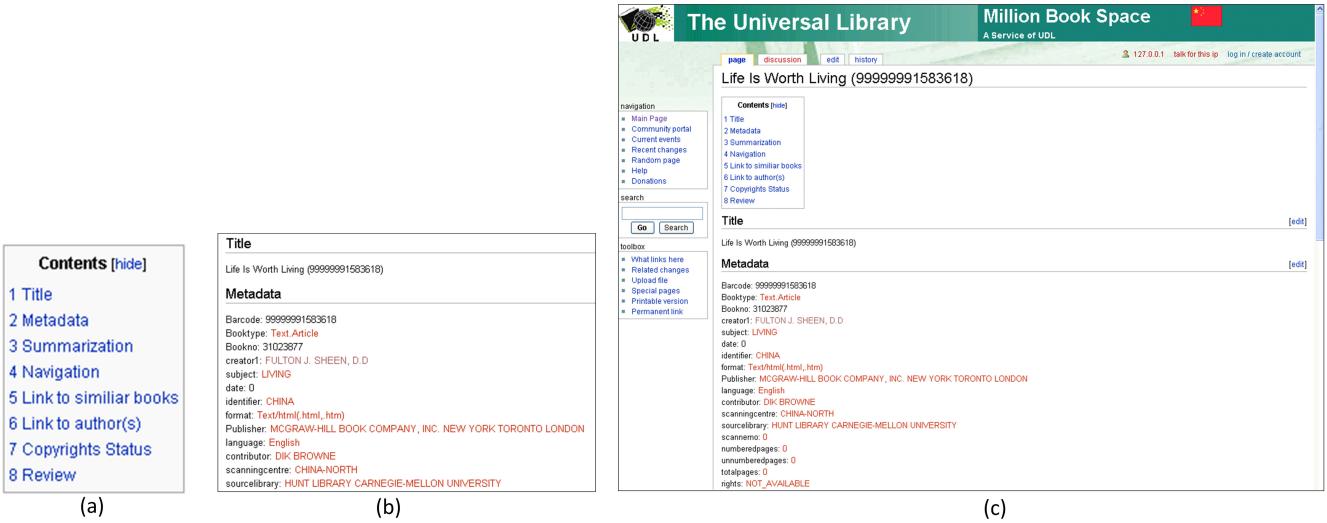

Figure 2. The snapshot of iULib 1.0: (a) table of contents; (b) metadata description; (c) the user interface.

instance, in an author's page, list of his/her works may be posted. Consequently, these e-books are correlated since they are written by the same author. Moreover, methods of extracting similar e-books mentioned above make it possible to build a book network which enhances the correlations among pages in iULib 1.0.

Fig. 2 shows the snapshot of iULib 1.0. This system can also be accessed online at http://www.ulib.org.cn.

4. IULIB 2.0: WIKIPEDIA AS A KNOWLEDGE BASE FOR UDL

Wikipedia may also be used as a knowledge base for UDL. Towards this end, we try to design iULib 2.0. The purpose of iULib 2.0 is to upgrade UDL to an intelligent platform of knowledge organization and information retrieval.

4.1 Wikipedia as a knowledge base

Wikipedia is not only a Web collection but also an online knowledge center which assembles all users' intelligence. Therefore, it is naturally attractive and promising that this open and constantly evolving encyclopedia can yield inexpensive knowledge structure.

In our previous work,[6] we presented a novel e-book retrieval system, called IQuery, which provides multi-granularity and multi-facet e-book retrieval. IQuery extended its capability of retrieval and navigation with a pre-constructed TermNet. This TermNet was built with the "subject" domain in metadata as well as the occurrences of terms in e-books. That is, if two terms co-occur in the same e-book, it can be deduced that there may be some semantic association between them. Table 1 shows the top 10 related terms returned by IQuery for term "Microsoft." At the WikipediaMM task in ImageCLEF 2008, we also constructed a knowledge base from Wikipedia by combining keyphrase extraction and link analysis (Keyword-IBF method), which helped improve the retrieval performance.[7]

These practices show that it's helpful to build a knowledge base for enhancing information retrieval. It's also possible to extract a knowledge network from online encyclopedias such as Wikipedia.

4.2 Constructing Knowledge Network from Wikipedia

Typical knowledge base extraction methods from Wikipedia are link structure mining[8] and content mining. We experimented three typical strategies for building a knowledge base from Wikipedia:

(1) TF-IDF

Table 1. Top 10 related terms returned by IQuery for term "microsoft."

Related Term	**Score**
Microsoft Corporation	0.872
Gates Bill	0.848
Computer software industry	0.848
Businessmen	0.848
Computer industry	0.294
Computer software	0.257
Automobiles	0.245
Computers	0.228

Similar to cosine similarity described in Section 3, we treated each article as a VSM vector. Thus a direct way to build a knowledge network from Wikipedia is to find the most similar article titles.

(2) Link Co-occurrence Analysis

Co-occurrence analysis has been proved to be effective widely in thesaurus construction. Thus we can employ this method to build a knowledge network from Wikipedia. Intuitively, if two concepts have many links to the same other concepts, they can be considered semantically similar. We thus define link co-occurrence lc between c_i and c_j:

$$lc(c_i, c_j) = \frac{|L_{c_i} \cap L_{c_j}|}{|L_{c_i} \cup L_{c_j}|}, \tag{2}$$

where L_{c_i} is a set of links that concept c_i refers to. To minimize the number of candidate related concepts, we only choose concepts linked to c_i with a path length of l.

(3) PF-IBF

PF-IBF (Path Frequency - Inverse Backward link Frequency) which was proposed in Ref. 8, is a link structure mining method for building concept networks from Wikipedia. This method considers two statistical features: *pf* and *ibf*. *pf* is used with the assumption that if there are many paths between two concepts, then they are possible to have strong correlation:

$$pf(c_i, c_j) = \sum_{n=1}^{N} \frac{l_n}{f(n)}, \tag{3}$$

where l_n denotes the count of paths with length of n, and $f(n)$ is a monotonically increasing function which indicates that a longer path length would lead to a smaller weight.

On the other hand, some articles may have links to others. In this case, these articles would have a big *pf* value to many articles. Thus a statistical feature *Inverse Backward link Frequency* (ibf) is calculated as:

$$ibf(t) = \log \frac{N}{bf(t) + \beta}, \tag{4}$$

where $bf(t)$ is the number of backward links of article t, N denotes the total number of articles, and β is a parameter in case $bf(t)$ is zero. Therefore, the *relatedness* between two articles can be calculated as

$$pfibf(c_i, c_j) = pf(c_i, c_j) \times ibf(c_j). \tag{5}$$

Figuer 3 shows some examples of correlated terms extracted by the above three methods and *TermNet*. TF-IDF method can extract similar terms and performs well only when the input query is specific. In case the input query contains ambiguous terms such as "apple" which can refer to both "apple (fruit)" and "Apple Company", the method may present unsatisfactory results. Moreover, via the TF-IDF method, we can not find out latent correlated terms which contain no common words with the input query.

Query	Extracted correlated terms			
	TF-IDF	Co-occurrence	PF-IBF	TermNet
Data Mining	Relational data mining, Data stream mining, Java Data Mining, Web data mining, Oracle Data Mining, Weather Data Mining	Data, Knowledge discovery, Data warehouse, Enterprise resource planning, Java Data-Mining	Human rights, Reliable sources, Open Directory Project, Information, Database, Education, Data warehouse	MINERA, Database management, Database searching, Temporal databases, Computer algorithms
Thesaurus	CrispThesaurus, RCA Thesaurus, OpenThesaurus, Thesaurus.com	Historical Thesaurus of English, Roget's Thesaurus, Terminology, Synonyms	Wiktionary, Dictionary, Oxford English Dictionary, Controlled vocabulary	N/A
Land Use	Nonconforming use (land use), Special use permit (land use), Variance (land use), Land use planning, Land use forecasting, Locally unwanted land use, Land Use and Development Act	Variance (land use), Zoning, Deforestation, Wilderness, Salinization, Land use planning, Natural resources	Biodiversity, Zoning, Natural environment, Deforestation, Anthropogenic, Soil erosion	City planning, Town planning, Environmental policy, Regional planning, Transportation, Urban renewal, Agriculture
City Planning	Planned city, Garden city (city planning), List of planned cities, Master of City Planning, City planning effort	Regional planning, Satellite campus, Tract housing, Urban design, Cluster development	Developments, Zoning, Europe, Industrial park, Science park, Cemetery, Architecture, Land use planning, Business cluster	Urbanisme, Town planning, Cities and towns, Urban renewal, Urban policy, Regional planning
Peking University	University of Peking, Peking Medical University, High School of Peking University, Peking University alumni, Peking University faculty	Tsinghua University, Fudan University, Zhejiang University, Hong Kong University of Science and Technology	Date of establishment, Universityof Tokyo, Postgraduate education, List of universities in the People's Republic of China, Tsinghua University	N/A
Iraq War	Iraq War/Iraq Insurgency, Iraq War/Iraq Timeline, 2003_war_in_Iraq, 2003_Iraq_war, The Iraq War, 1991_Iraq_war	Iraq, Sectarian, Saddam_Hussein, Iraqi_resistance_movement	Saddam Hussein, Iraq, Sunni	United States, War on Terrorism, World politics, Imperialism

Figure 3. Samples of queries and the extracted correlated terms.

Link-based methods, including link co-occurrence and PF-IBF, have an obvious advantage that they can handle latent correlations. As shown in Figuer 3 , PF-IBF can extract meaningful semantic related terms. However, this method is unstable as sometimes frequently linked concepts that may be not general but authoritative could lead to a low *ibf*. In contrast, link co-occurrence analysis method performs quite stably. This method has the capacity of filtering system concepts such as dates. Actually, it trends to extract related terms that have similar link distribution to that of the input queries.

The above methods provide several optional strategies to extract correlated concepts for given terms. In other words, these methods can be used to build a knowledge network from Wikipedia.

4.3 Book retrieval with knowledge network

In our previous work,[6] we have demonstrated that *TermNet* can be used in IQuery to extend the query terms by selecting related terms. However, there are some drawbacks of this query expansion approach:

a) MARC or DC subject terms are defined with subject headings or thesauruses. Thus there may be no correlated terms returned for an input query by using word matching method proposed in[6] (See "Thesaurus" and "Peking University" in Figure 3).

b) Subject terms are not evolving. Consequently, this subject term network may not match users' changing needs as new knowledge and techniques appear. For instance, we could not find any extended terms for "Google" in *TermNet* as "Google" is too new to be included in subject headings like LCSH.

Therefore, *TermNet* is not enough for query expansion in our e-book retrieval system. As mentioned above, the knowledge base built from Wikipedia may be able to enhance the retrieval performance. Since Wikipedia covers a wide range of knowledge, we can always find related terms in Wikipedia knowledge network. New terms and recent events may also be discussed in Wikipedia before any book about them is included in UDL. Therefore, theoretically these drawbacks of *TermNet*-based query expansion method could be solved via integrating Wikipedia into e-book retrieval.

By using query extension with the knowledge base constructed from Wikipedia, we achieved the best retrieval results at the WikipediaMM task in ImageCLEF 2008 which aims to investigate effective retrieval approaches in the context of a large-scale and heterogeneous collection of Wikipedia images.[7] Our results demonstrate that the knowledge network extracted from Wikipedia can be used to effectively expand the query semantics of image retrieval. To further illustration, some experimental results at this task are given in the following.

Table 2. The testing results of image retrieval via query extension at the WikipediaMM task

Run ID	QE	Modality	MAP	P@5	P@10
No-QE	without	TXT	0.2565	0.4427	0.3747
CBIR	without	IMG	0.1928	0.5307	0.4507
QE-SWIBF	with	TXT	0.2609	0.4400	0.3693
QE-Text-Semi	with	TXT	0.3444	0.5733	0.476

At the WikipediaMM task, we evaluated the performance of knowledge network based query expansion method (denoted by QE-SWIBF). As shown in Table 2, results indicate that this query expansion method could slightly improve the MAP performance of text-based image retrieval systems. It's understandable that the automatically extracted knowledge base could not be directly employed in some applications. The experimental results indicate that a good knowledge network construction algorithm with reasonable manual confirmation (denoted by QE-Text-Semi) can remarkably improve the retrieval performance. The results also show us a vision of using Wikipedia as a knowledge base for query expansion in UDL. In the ongoing work, we will implement this method in the IQuery system for UDL.

5. DISCUSSION AND FUTURE WORK

The implementation of iULib shows clearly that for the e-books in UDL, it is possible to build a Wiki like dynamic book space with strong inter-book correlations and user participation, and to build knowledge networks from Wikipedia that help improving e-book retrieval. However, there are still several challenges for iULib.

5.1 Challenges for iULib

Surely there are some drawbacks to applying Wiki to UDL. First of all, iULib 1.0 needs a group of users that share knowledge and collaborate for a common object. It's easy to apply Wiki architecture and software to UDL but very difficult to keep a large number of active users. Just like what Web 2.0's central point is, iULib 1.0 would not achieve its goal if user involvement is insufficient.

Secondly, as e-book information pages can be edited by any user, no one can ensure the authority of e-book information unless corresponding control strategies are conducted.

5.2 Applying other Web 2.0 applications to UDL

Wikipedia is just one form of Web 2.0 applications. There are many other Web 2.0 applications that could be applied in and benefit UDL.

(1) *Social Tagging.* In social tagging applications such as del.icio.us, users can add tags to their interested content. These tagging systems, also referred to as folksonomies, enable people to manage their information with flat and free keywords. When tagging is enabled in UDL, e-books would become easier for users to catalog and manage. In addition, these tags may enrich the knowledge network of UDL and help generate personalized knowledge network for navigation and retrieval.

(2) *Social Bookmarking*. With the help of metadata, social bookmarking is a method for users to organize and search web pages. Social bookmarking, together with social tagging, can be employed in UDL to enhance user participation and sharing. One possible application in UDL may be management of users' book list, and reading notes. In an open library community, this social bookmarking system may create an environment for collaborative learning and interpersonal knowledge sharing.

5.3 UDL Everywhere: Open API based applications

We now have more than one thousand Open APIs on the Web such as Google Maps API which are widely used to build new applications. Mashups of services and these APIs make it easier for users to access data resources and experience creative ideas. Usually, digital libraries provide portals for users to search and view e-books and other digital resources. However, digital libraries could have done much better with open API by means of developing applications that access their resources with third party open APIs, or providing open platforms which allow developers to create new applications that exploit digital e-books and resources in various ways.

To our knowledge, openlibrary.org is the first open platform that provides interfaces to access book metadata. For UDL, the basic functions of the open platform may contain e-book retrieval, illustration retrieval, copyright control, etc.

6. CONCLUSION

With the rapid increasing popularity of Wikipedia, this paper presents our vision of bridging Wikipedia to UDL, which we call iULib. As a first step, we have implemented iULib 1.0, which directly applied the Wiki architecture and software in UDL to upgrade the digital library as an interaction platform. Secondly, we propose iULib 2.0 where Wikipedia can serve as a knowledge base for UDL. We have also discussed the possibility and approaches of building a knowledge network from Wikipedia, and analyzed how this knowledge network could be used in e-book retrieval. It is clear that Wikipedia and digital library can integrate each other's valuable results and best practices to benefit each other.

ACKNOWLEDGMENTS

This work is supported by a grant from the CADAL project.

REFERENCES

[1] Huang, T., Tian, Y., Wang, C., Shi, X., and Gao, W., "Towards a multilingual, multimedia and multimodal digital library platform," *J. Zhejiang Univ. SCI. A* **6**(11), 1180–1184 (2005).

[2] Lim, E., Wang, Z., Sadeli, D., Li, Y., Chang, C., Goh, D., Theng, Y., Zhang, J., and Sun, A., "Integration of wikipedia and a geography digital library," in [*Digital Libraries: Achievements, Challenges and Opportunities*], *Proc. 9th Int'l Conf. Asian Digital Libraries, LNCS 4312*, 449–458 (2006).

[3] Anderson, P., "What is web 2.0? ideas, technologies and implications for education," *JISC Technology and Standards Watch* , 2–64 (2007). http://www.jisc.ac.uk/media/documents/techwatch/tsw0701b.pdf.

[4] Maness, J. M., "Library 2.0 theory: Web 2.0 and its implications for libraries," *Webology* **3**(2), Article 15 (2006).

[5] Frumkin, J., "The wiki and the digital library," *OCLC Systems & Services: International Digital Library Perspectives* **21**(1), 18–22 (2005).

[6] Huang, C., Tian, Y., Zhou, Z., and Huang, T., "Towards multi-granularity multi-facet e-book retrieval," *Proc. 16th Int'l World Wide Web Conf.*, 1331–1332 (2007).

[7] Zhou, Z., Tian, Y., Li, Y., Huang, T., and Gao, W., "Large-scale cross-media retrieval of wikipediamm images with textual and visual query expansion," in [*Evaluating Systems for Multilingual and Multimodal Information Access*], *LNCS 5706*, 763–770 (2009).

[8] Nakayama, K., Hara, T., and Nishio, S., "A thesaurus construc-tion method from large scale web dictionaries," *Proc. IEEE Int'l Conf. Advanced Information Networking and Applications*, 932–939 (2007).