Emerald Insight

Library Hi Tech

Constructing linkage between libraries and up-to-date news
Wei Yu Junpeng Chen

Article information:

To cite this document:
Wei Yu Junpeng Chen , (2016),"Constructing linkage between libraries and up-to-date news", Library Hi Tech, Vol. 34 Iss 2 pp. 301 - 313
Permanent link to this document:
http://dx.doi.org/10.1108/LHT-11-2015-0109

Downloaded on: 10 November 2016, At: 20:39 (PT)
References: this document contains references to 22 other documents.
To copy this document: permissions@emeraldinsight.com
The fulltext of this document has been downloaded 97 times since 2016*

Users who downloaded this article also downloaded:

(2016),"Internet of Things – potential for libraries", Library Hi Tech, Vol. 34 Iss 2 pp. 404-420 http://dx.doi.org/10.1108/LHT-10-2015-0100

(2016),"Low-barrier-to-entry data tools: creating and sharing humanities data", Library Hi Tech, Vol. 34 Iss 2 pp. 268-285 http://dx.doi.org/10.1108/LHT-07-2015-0073

Access to this document was granted through an Emerald subscription provided by emerald-srm:563821 []

For Authors

If you would like to write for this, or any other Emerald publication, then please use our Emerald for Authors service information about how to choose which publication to write for and submission guidelines are available for all. Please visit www.emeraldinsight.com/authors for more information.

About Emerald www.emeraldinsight.com

Emerald is a global publisher linking research and practice to the benefit of society. The company manages a portfolio of more than 290 journals and over 2,350 books and book series volumes, as well as providing an extensive range of online products and additional customer resources and services.

Emerald is both COUNTER 4 and TRANSFER compliant. The organization is a partner of the Committee on Publication Ethics (COPE) and also works with Portico and the LOCKSS initiative for digital archive preservation.

*Related content and download information correct at time of download.

Downloaded by TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES At 20:39 10 November 2016 (PT)

The current issue and full text archive of this journal is available on Emerald Insight at:
www.emeraldinsight.com/0737-8831.htm

Constructing linkage between libraries and up-to-date news

Wei Yu
Jiangsu Key Laboratory of Data Engineering and Knowledge Service, Information Management School, Nanjing University, Nanjing, China, and
Junpeng Chen
School of Information Engineering, Nanjing University of Finance and Economics, Nanjing, China

Received 17 November 2015
Revised 5 February 2016
15 March 2016
Accepted 6 April 2016

Abstract

Purpose – The purpose of this paper is to construct the linkage between libraries and up-to-date news. This study developed a system to recommend libraries' resources to those daily news readers who are interested in the topics of the target news. The analysis of experiments results served as the reference for the development and improvement of linking libraries' resources with other web resources.

Design/methodology/approach – Up-to-date news were gathered through the news feeds to make the integration with the libraries' records. In task 1, the libraries' records were linked and recommended to the target libraries' records which are of the same topics. In task 2, the system aimed to find the relevant libraries' records for target news. Three recommendation methods were compared in both tasks to find the most effective approach to the system.

Findings – Experiment results showed that: at first, in task 1, the system can assign the libraries' records of the related topics effectively; second, in task 2, the recommending system can obtain a satisfied recall hit rate through human evaluation. Therefore, regarding the popularity of the daily news online, the linkage and recommendation with the libraries' resources can increase the visibility of the libraries' resources and eventually promote the information consuming in libraries.

Practical implications – The authors have confirmed, using three matrix factorization methods, that weighted matrix factorization used in the libraries' records recommendation system, could achieve better performance than the other two. Based on the research, the libraries could incorporate the online news and libraries' resources in practice.

Originality/value – To increase the visibility and promote information consuming of libraries, this study proposed a novel method to construct the linkage between library and up-to-date news. The results of data analysis indicate that recommendation of libraries resources through the daily news can achieve effective performance. Thus, it can be inferred that the research results of this study are representative and have practical values in real world practice.

Keywords Library services, Web resources, Information technologies, Libraries resources recommendation, Up-to-date news, WMF

Paper type Research paper

Introduction

Over the last decade, libraries have experienced a steady decline in the number of users of the online services, which in turn translates into a substantial decrease in the number of information consumers who use library resources for their information needs. Compared to the web search engine and the news website, libraries may be too far away from our everyday life: people pay more attention to online news that will bring visible influence to our lives. Reportedly, the great majority of the information seeking activities start from search engines such as Google instead of library websites (De, 2005).

Though the daily news can bring fresh information and hot spots, it can also cause confusion for people who have no background information or domain knowledge about the news. To help people grasp the gist of the news article, Wei and Gao (2014)

Library Hi Tech
Vol. 34 No. 2, 2016
pp. 301-313
© Emerald Group Publishing Limited
0737-8831
DOI 10.1108/LHT-11-2015-0109

Downloaded by TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES At 20:39 10 November 2016 (PT)

Downloaded by TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES At 20:39 10 November 2016 (PT)

extracted the indicative tweets to automatic extract news highlights. HEADY (Alfonseca *et al.*, 2013) presented an abstractive headline generation system based on the generalization of syntactic patterns by means of a noisy-or Bayesian network for headline generation from news collections. These works adopted natural language processing methods to help users get more clear vision about the facts and opinions reported in the news, but they still leave the problem of how to find the deep-seated problems and profound analysis to users' themselves. Stajner *et al.* (2013) discussed how to automatic select the social media responses to news. They only emphasized the sentiment of these social media responses, not providing semantic analysis based on the content of the news.

As the problem described above, there is a lack of method to provide domain knowledge and depth analysis about the news, through which people can have a thorough grasp and have extending reading about the news. On the other hand, in the publications and articles stored in the libraries, most depth thoughts and valuable views that may be useful are often ignored by the web users. Hence, to construct the linkage between libraries' resources and daily news may give users extending reading lists of library resources which can provide much knowledge on users' subject of interest. Through this linkage each news article will be linked to the records of relevant publications in libraries, which in turn provide bibliographic information on the publications of interest and directs users to their local or digital libraries to access those publications. Constructing this linkage will consequently improve the visibility of the library resources which are currently overlooked to a large extend by information consumers.

This newswire-library linkage offers a number of unique advantages to both library and information seekers. First, the up-to-dateness: due to the nature of daily news, the library can be introduced to the daily life of the information seekers, which can increase the information consuming of the library resources. Second, the extensive coverage and comprehensiveness: the library resources can contains more background knowledge and profound argument than the short news articles can offer, so that these library resources can help people learned the news more accurately and comprehensively, and finally lead to more appropriate decision. Third, the rich description: library articles provide rich descriptive content for the represented concepts in newswires. According to the report published by Online Computer Library Centre (OCLC) in 2009 (Calhoun *et al.*, 2009), the majority of end users of online library catalogues surveyed considered adding more subject information to improve the discovery-related data quality of the catalogues. Using daily news as a complementary information resource in library catalogues addresses this need by offering a lot of daily life subject headings to complement the traditional library vocabularies.

For all above, constructing the linkage between library records and daily news may be considered the first step towards full newswires-library integration. However, the high cost of manual linking for newswires-library links poses a major obstacle to achieving this goal. Therefore, considering libraries limited resources, our aim is to reduce the cost of such integration as much as possible. To this end, in this paper we describe the design and development of a novel method for constructing linkage between library and up-to-date news.

Literature review

Everyday information is proved to be critical to build information community and has an important impact on people's information behaviour (Khoir *et al.*, 2015). As social and cultural institutions, libraries have a role to play in enabling individuals to acquire

Downloaded by TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES At 20:39 10 November 2016 (PT)

the type of skills from the information society and engaging with information community so as to extending "civilizational competence" beyond the realm of the library walls to encompass practice that translate into various spheres of individuals' lives (Caidi, 2006; Goulding, 2009). Since Web 2.0 is becoming a part of everyday life online, libraries can make their materials accessible to people all over the world through internet and web (Holmberg *et al.*, 2009). Hence, libraries have been among the first to realize the great potential and importance of the web in fulfilment of information needs and expectations (Pisanski and Žumer, 2005).

To provide a wide variety of web services to a broad range of users, many libraries chose to build libraries' website to help users easily access to library information services. National Library of New Zealand design a website to meet the information needs of diverse international users (Booth and Napier, 2003). The key principle for the website development is to organize the resources in an efficient way. Shanghai Jiao Tong University Library developed an information portal to integrate its information resources and services (Jin and Peng, 2009). The information portal could provide cross-searching, resources linking, and service integration supported by Metalib SFX and Web 2.0 technologies. A survey conducted with 23 accessible and English language websites of national libraries in Asia in 2011 (Zarei and Abazari, 2011) showed that the national libraries of eight countries provided web-based services in a range of 25-52.29 per cent. In the study, the authors suggested the libraries make high use of traditional library tools and resources to offer efficient online services for every target users.

With the development of web services technologies, libraries also implement web service applications to increase usage rate of libraries on the web. To find out the actual usage of the mobile library, a study developed a Mobile Library APP system and invited students of National University of Tainan in Taiwan to be experimental subject (Pu *et al.*, 2015). The analysis of experiment results showed the students were highly positive to accept the Mobile Library App service for it can help them work more efficiently. A study from University of Barcelona (Rodrigues-Gairin and Somoza-Fernandez, 2014) described the use of web services to interconnect the GTBib Interlibrary Loan (ILL) programme which is widely adopted by Spanish university libraries with the OCLC WorldShare platform. The study aims to link the two platforms to make the request/response through the resources sharing platforms more efficiently. A case study on a library's information retrieval system effectiveness shows that the IR component used in the current publishers ILS could be adequately replaced with an open source component (Marijan and Leskovar, 2015). Based on the research, the publisher could incorporate suggested open source IR components in practice. All the studies above mainly focused on constructing web services and homepages for libraries' resources to attract users and information consuming. But this information consuming needs the users to visit websites and information portals at first. There is a lack of effective methods to interlink libraries to other web resources and "push" libraries' resources to users who may not visit the libraries websites.

For this problem, there has been substantial research carried out in relation to library records with other web resources such as Wikipedia (Joorabchi and Mahdi, 2014) or web pages (Golub, 2006). Rodriguez *et al.* (2015) recommend non-traditional educational resources to users with ontology. These linkages among libraries and other web resources can increase the visibility of the libraries resources and attract more information consumers. But these studies are mainly on linking library to controlled vocabularies and classification systems, which may be far away from our daily lives and need high-information literacy skills to use compared to online news.

Downloaded by TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES At 20:39 10 November 2016 (PT)

In this paper, we construct the linkage between libraries and up-to-date news. Once a user is on the newswire page of a topic, she/he will be provided with a "further reading" list. This reading list consists of a set of library resources most relevant to the target news topic. For example, for such a news article "Climate meeting to discuss future of fossil fuels-After concluding that global warming almost certainly is man-made and poses a grave threat to humanity, the UN sponsored expert panel on climate change is moving on to the next phase: what to do about it". The particular topic about this news article may be "global warming", "climatic changes", and the most relevant subject topics from library including "greenhouse effect", "Nature-effect of human beings", etc. From these subject topics, we can find the corresponding catalogue records and recommend the books to users who are interested in this news article, such as "Dire predictions: understanding global warming", "The greenhouse effect", and "Climate: present, past and future". These books can help users get further reading and provoke deep thinking.

Methodology

Figure 1 illustrates the proposed framework to recommend libraries catalogue records for up-to-date news. In this framework, first up-to-date news data and libraries catalogue records are gathered and preprocessed. Second, content-analysis and metadata-analysis are carried out upon these datasets. In metadata-analysis, we only select the title and contents/descriptions of the news articles and libraries catalogue records and make the title and contents/descriptions a concatenation for content-analysis. Natural language processing such as stop words removing or lemmatization is adopted in the content-analysis at first. Then, we build a co-occurrence matrix composed of the constituent words from the contents of the libraries catalogue records.

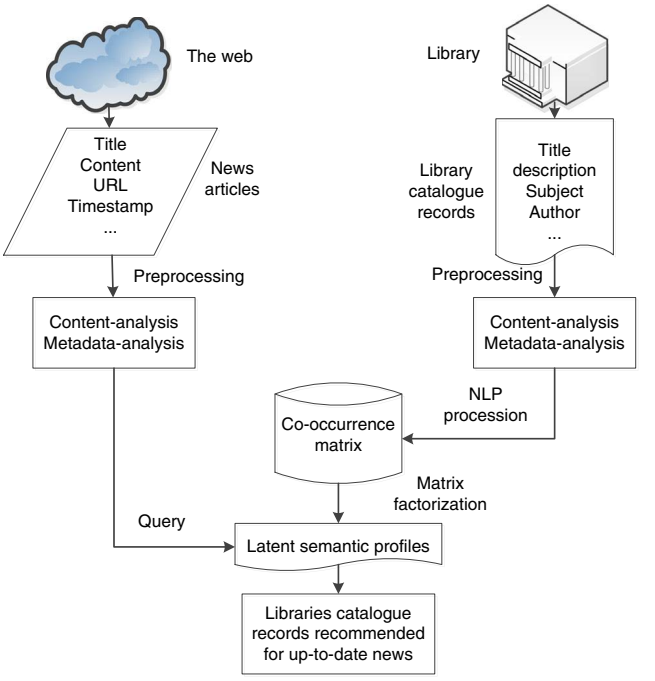

Figure 1.
The framework for libraries records recommendation

Downloaded by TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES At 20:39 10 November 2016 (PT)

In this recommendation framework, the latent semantics are extracted from the contents of the libraries catalogue records to calculate the semantic similarity. However, the challenge is that the contents of the libraries catalogue records are very short and the co-occurrence matrix from the constituent words is too sparse to learn the accurate semantics. Hence, we used the matrix factorization methods to find latent semantics profiles for the libraries catalogue records.

Suppose there are news articles set $N=(n_1, n_2, \ldots, n_b)$, and libraries catalogue records set $D=(d_1, d_2, \ldots, d_n)$. The constituent words of the libraries catalogue records can be represented as $T=(t_1, t_2, \ldots, t_m)$.Then the co-occurrence matrix $X_{m\times n}=TD$. This matrix X comprises the tf-idf values in each x_{ij} cell, namely, that tf-idf value of word t_i in libraries catalogue record d_j.

Our approach to automatically link library records through daily news articles adopted the weighted matrix factorization (WMF). This method allows for direct control on each matrix cell on document-word matrix and hence can assign a flexible weighting measure (Steck, 2010). The model factorizes the original matrix X_{ij} into two matrices such that $X\approx P^T Q$, where P is a $D\times M$ matrix, and Q is a $D\times N$ matrix.

The model parameter (vectors in P and Q) are optimized by minimizing the objective function:

$$\sum_i \sum_j W_{ij}\left(P_{\cdot,i}\cdot Q_{\cdot,j}-X_{ij}\right)^2+\lambda\|P\|_2^2+\lambda\|Q\|_2^2 \tag{1}$$

where λ is a free regularization factor and the weight matrix W defines a weight for each cell in X. In application, we can assign a small weigh:

$$W_{i,j}=\begin{cases}1, & \text{if } X_{ij}\neq 0\\ w_m, & \text{if } X_{ij}=0\end{cases} \tag{2}$$

Experiments

Experiment setting

Fulfilling the proposed vision of news-library integration requires two major developments: annotation of library records with news topics as subject metadata, and annotation of news articles with citations to the most relevant library resources.

Corpora. We use the catalogue records corpora: WorldCat-million dataset released by the OCLC in 2012 (Murphy, 2012). The dataset contains metadata records of nearly 1.2 million library materials most widely held in libraries. The metadata contains approximately 80 million linked data triples, which can help user find the linked resources easily on the web. Every item in this catalogue records contains about 15 properties, including title of the book, author, publisher, genre/form, database, description, rating, subject, physical format, additional, material type, document type, contributors, ISBN, OCLC number, series, etc. The "description" property includes the summary, abstract, and description of the book. These descriptive words contain abundant information about what the publication talked about and hence may be the most important property for getting the point of the book besides the "title" and "subject". In order to accurately link catalogue records and news articles, we need to use the "description" and "title" properties of the catalogue records in experiments. Hence, we filtered "trivial" catalogue records which lack such properties. The final dataset results in 217,147 catalogue records.

Downloaded by TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES At 20:39 10 November 2016 (PT)

For the corpus of news articles, we collect the news articles of Yahoo! news from RSS feeds, dated from 5 April to 7 July 2014, totally 95 days. Each news articles has the published time, title, news summary, URL. We remove the duplicate news to get 4,702 news articles, most of which are mainly about the current events, including music, sports, technology, international affairs, domestic news, such as "Climate meeting to discuss future of fossil fuels", or "NASA training humanoid robot to perform surgery", etc.

A catalogue record is represented as the concatenation of the "title" property and "description" property, while a news article is the concatenation of its title and its summary. This concatenation can yield better performance. All the datasets are tokenized, POS-tagged by Stanford POS tagger and lemmatized by snowball, generating 216,147 short text and 131,295 words from catalogue records set for training and developing, and the other datasets are left for test. We also filter the "noisy" words whose frequency are more than 4,000 ($> 4{,}000$) or less than 3 (< 3) and retained 67,932 words from the training set. To get rid of these stop words can improve the performance greatly. The value of each word in matrix X is its tf-idf value in the short text.

Baselines. We present two baselines: first, latent semantic analysis (LSA) that uses svds() function in Matlab; and second, non-negative matrix factorization (NMF) that uses nnmf() function in Matlab. To curtail variation in results due to randomness, the number of NMF and WMF is the average of 10 runs. The similarity between different documents is measured by cosine similarity.

Evaluation

In order to evaluate the proposed linking system, a collection of library metadata records manually linked with news articles was required to train and test the system. However, to the best of our knowledge no such dataset has been built before and therefore, we had to build our own. For this purpose, we decide to perform two tasks to evaluate the quality of our system.

Task 1. In this task, when a catalogue record is submitted, the system is required to recommend the relevant catalogue records. It is worth noting that the catalogue records have the similar format as the news articles: they all have title and short description. Hence, we can use catalogue records to evaluate the performance of our linking and recommending system as the news articles can do to some extent. The main difference between the catalogue records and news articles is that the catalogue records have been classified by authority and been tagged with controlled vocabularies, such as Dewey Classification, Library of Congress Subject Headings (LCSH), Faceted Application of Subject Terminology (FAST), etc. We used these classification tags to evaluate the performance, since these classification tags have been adopted in book taxology for a long time and can represent the expert knowledge from the authoritative book classifiers.

At first, we randomly selected 1,000 catalogue records from 216,147 short text as development set, based on which all the parameters are tuned. Then the 1,000 catalogue records which have not been used for training and developing are used to evaluate the performance of systems. For every target catalogue records, we collected the classification tags of the top 50 most relevant catalogue records and ranked the terms with their frequencies. In order to examine if the system is able to rank the target catalogue records in the first few returned results, we used top ten recall hit rate to evaluate whether the classification tags of the target catalogue record are in the top ten highest frequently appeared classification tags of the top 50 most relevant catalogue records.

Downloaded by TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES At 20:39 10 November 2016 (PT)

Task 2. This task is given the text (title and news summary) in news article, a system aims to find the relevant catalogue records. For our task evaluation, ideally, we would like the system to be able to identify the catalogue records specifically referred to by the subject of each news article. However, this is very difficult given the large number of potential candidates, especially those with slight variations. Therefore, for every target news article, we collected the classification tags, i.e., the subject topics of the top 50 most relevant catalogue records the system recommended, and ranked these terms with their frequencies.

In order to examine if the system is able to recommend the relevant catalogue records in the first few returned results, we used top ten recall hit rate to evaluate whether the top ten highest frequently appeared classification tags of the top 50 most relevant catalogue records are relevant with the target news article. Three human raters were required to select the relevant subject topics from the ten terms and the average number of rated numbers by these three human raters will be the final value. This evaluation metric can greatly reduce the heavy and fallible labourious work for reading the large scale of text from catalogue records. Additionally, rating only the subject topics can avoid severe subjective conflicts among the human raters for these subject topics are authoritative controlled vocabularies which are widely accepted.

Experiment results

Task 1. Table I summarizes the performance of the baselines and WMF at latent dimension $D = 100$. All the parameters are chosen based on the development set. For WMF, we try different values of w_m. Since λ is a free regularization factor, we set $\lambda = 1.0$. We evaluate the performance of the different methods with different classification tags: Dewey, FAST, and LCSH.

WMF outperforms LSA and NMF by a large margin. This is because WMF takes advantage of much more missing words to learn more robust latent semantics vectors. LSA achieves better scores than NMF, although the idea of NMF is commonly used in word sense disambiguation (Chen *et al.*, 2014) as well as image processing. The reason is words are enough to capture semantics for LSA, while NMF are used for a denser matrix. For different authority classifications, Dewey, FAST, and LCSH, WMF gets a better performance on FAST than on Dewey and LCSH. There are few differences for the performance of the two baselines, LSA and NMF on these three authority classifications.

Different values for w_m result in different values. When $w_m = 0.01$, we get the highest values on development and test set for FAST and test set for LCSH, they are 63.0, 61.6, and 73.2 per cent, respectively. When $w_m = 0.005$, the performances on

		Dewey		FAST		LCSH	
Models	Parameters	Dev.	Test	Dev.	Test	Dev.	Test
LSA	–	54.3	56.1	43.1	46.4	56.3	54.8
NMF	–	51.7	51.2	37.1	38.8	48.3	49.0
WMF	$w_m = 0.001$	67.0	65.4	59.8	60.6	70.3	70.7
WMF	$w_m = 0.005$	*68.2*	*68.5*	62.4	61.0	*73.0*	71.9
WMF	$w_m = 0.01$	67.4	68.4	*63.0*	*61.6*	72.6	*73.2*
WMF	$w_m = 0.02$	66.8	67.2	59.6	59.7	71.1	70.1
WMF	$w_m = 0.05$	66.0	64.9	59.1	58.4	69.1	68.7

Table I. Top ten recall hit rate performance for task 1(in per cent)

Downloaded by TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES At 20:39 10 November 2016 (PT)

development and test set is best for Dewey, they are 68.2 and 68.5 per cent. The highest performance for LCSH on development set is obtained when $w_m = 0.005$. These experiments show that when $w_m = 0.01$, WMF performances best on test set. Although a larger D yields a better result, we still use a 100 due to computational complexity.

Task 2. In order to get an objective observation of the performance, we randomly selected 500 news articles (about 10 per cent of the news articles set) for evaluation. The results are showed in Table II, from which we can see WMF has better performance than LSA and NMF. From analysis, we noted that some news may be pieced together into a whole story. For example, a series news articles is about the Ukraine conflicts, such as the Obama's statement about the conflict, Putin's attitude towards Ukraine, and the east rebels and conflicts of Ukraine. For different aspects of the whole story, our method can find various topics, such as "Nationalism-Ukraine-history", "Democratization-Ukraine", which the information seekers may be interested in and can have extensive reading on.

Also, we can find many analogy cases that users can have further reading. For example, for the news "Idaho just hosted the best political debate of the year", the recommended subject topics include "debates and debating-Missouri-Kansas City". People who are interested in debate in Idaho may be also interested in debate in Missouri. Additionally, our method can help word sense disambiguation and name entity recognition in the news articles. For instance, when "Spelling Bee" appeared in the news, we find the similar subject topic is about "spelling ability". This may be because the subject topics used by the catalogue records are control vocabularies, which are less ambiguous than the free form of the newswires.

Case study and analysis

In order to show the effect of our recommendation methods, we randomly selected five examples from our libraries-newswire linkages experiments for analysis. The cases are shown in Table III.

From the Table III, we can note that every news are related to some topics, that is, keywords or subjects. The corresponding library's records are ranked with the similarity to the target news in ascending order. Due to limited space, we only selected the top five keywords and top five library's records for the target news in each case.

In case one, the news is about the end of the operating system "Windows XP", which had been supported by Microsoft. In the top five keywords, we can find "software", "operating systems", "computer", etc. The corresponding library's records include "Windows 7 quicksteps", "guide to Microsoft Windows 7" and "Windows XP". It is worth noting that "Windows 7" is a kind of operating system supported by Microsoft and seemed as the successor of the "Windows XP". Hence, if users are interested in the "end of Windows XP support" and are disturbed by this situation, these books about "guide to Windows 7" may be very useful to them and can help them make their way easily.

Table II.
Top ten recall hit rate performance for task 2 (in per cent)

Models	Parameters	Performance
LSA	–	45.4
NMF	–	31.4
WMF	$w_m = 0.01$	*70.4*

Downloaded by TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES At 20:39 10 November 2016 (PT)

Daily news	Keywords/subjects	Library catalogue records
1. Title: end of Windows XP support spells trouble for some Contents: Microsoft will end support for the persistently popular Windows XP on Tuesday, and the move could put everything from the operations of heavy industry to the identities of everyday people in danger	1. "Microsoft Windows (Computer file)" 2. "Operating systems (Computers)" 3. "Microsoft PowerPoint (Computer file)" 4. "Microcomputers" 5. "Presentation graphics software"	1. "The complete idiot's guide to Microsoft Windows 7" 2. "Windows 7 quicksteps" 3. "Windows XP all-in-one desk reference for dummies" 4. "Windows XP visual encyclopedia" 5. "Windows 7 inside out"
2. Title: Ebola fears spread as outbreak worsens Contents: emergency measures taken across West Africa to contain one of worst-ever outbreaks, threatening every country in region	1. "Africa" 2. "Epidemics – Europe – History" 3. "South Africa" 4. "AIDS (disease) – Africa" 5. "Africa, Sub-Saharan"	1. "Africa from 1945" 2. "Africa, a continent self-destructs" 3. "Africans: the history of a continent" 4. "Popular politics in the history of South Africa, 1400-1948" 5. "Black death: AIDS in Africa"
3. Title: AP PHOTOS: tornadoes rip across Midwest, South Contents: with parts of the USA recovering from deadly tornadoes, more heavy storms are making their way across the south	1. "Tornadoes" 2. "Hurricanes" 3. "Storms" 4. "USA" 5. "Natural disasters"	1. "Hurricanes, storms, tornadoes" 2. "Tornadoes" 3. "Storm warning; the story of hurricanes and tornadoes" 4. "Historical catastrophes: hurricanes & tornadoes" 5. "Disastrous hurricanes and tornadoes"
4. Title: Jihadist Islamic State seizes key Syria oil field Contents: the insurgent group took control of a major Syrian oil field on the Iraqi border	1. "Petroleum industry and trade" 2. "Petroleum reserves – Forecasting" 3. "PETROLEUM" 4. "Iraq" 5. "Energy policy – USA"	1. "Petrotyranny" 2. "Oil, power, and principle: Iran's oil nationalization and its aftermath" 3. "Crude power: politics and the oil market" 4. "Black gold: the story of oil" 5. "Crude awakenings: global oil security and American foreign policy"
5. Title: so what? USA regains jobs lost in the recession Contents: analysts downplay milestone as economy is still millions of jobs short of goal	1. "Job hunting" 2. "Unemployment – USA" 3. "Employment interviewing" 4. "Vocational guidance" 5. "Labor market – USA"	1. "The job description handbook" 2. "The job search solution the ultimate system for finding a great job now!" 3. "The job-hunter's survival guide: how to find hope and rewarding work even when\"there are no jobs\" 4. "Everything you need to know about getting a job" 5. "The coming jobs war: what every leader must know about the future of job creation"

Table III.
Cases for linkages between daily news and library

Downloaded by TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES At 20:39 10 November 2016 (PT)

For the second case, the news announced that a terrible disease Ebola was threatening the Africa. The keywords include "Africa", "history", and "AIDS". In the recommended library's records, users can compare the Africa to Europe, and can learned much about the geography and history of the Africa. Also, users can learn the "black death-AIDS" in Africa and can compare it to Ebola, both of which are serious disease and big dangers to the Africa.

The news in the third case talked about tornadoes in the USA. The keywords are mainly about "tornadoes", "hurricanes", "storms", etc. The corresponding library records include "storm warning", "historical catastrophes", or "disastrous hurricanes and tornadoes". This list contains much information about the various storms including tornadoes.

In the fourth example, the news described a group's control of oil fields at Iraqi border. The keywords include "petroleum", "Iraq", "energy policy", etc. The recommended records are related to "oil" or "black gold", and are mainly about politics and market.

The last example is about job lost in the recession of USA. The keywords include "job hunting", "unemployment", "vocational guidance", etc. From the corresponding records, users can find how to get a job, such as "the job search solution the ultimate system for finding a great job now", "the job-hunter's survival guide", etc. Besides, users can get in-depth analysis in the situation, such as "the job description handbook" or "the coming job war".

From all these cases, we can see that users can get not only the related topics but the related library's records from our library-newswire linkages. When users are interested in some topics in the daily news, she/he can easily find the similar topics and corresponding library's records from our recommending methods. This solution can bring users and applications to libraries as well as improving books utilization and providing in-depth thoughts to users. After learning more about the interested topics, users can have more background knowledge and can make more appropriate decision.

Since the library's records in our experiments are in linked data formats, they can make contribution to knowledge discovery and information sharing besides the benefits discussed above. For example, in the fifth case we can see the related records online, which include "the job description handbook". The simplified record in our experiment is as following:

> 190968
>
> www.worldcat.org/oclc/173248969
>
> 2008.
>
> < http://dewey.info/class/658.306/e22/ > .
>
> The job description handbook.
>
> "Introduction – An overview of job descriptions – Avoiding legal pitfalls – Analyze and define the job – Writing the job description – Using job descriptions – Troubleshooting." "Provides a guide for managment and human resources personael to create effective job descriptions, and describes how to establish a list of duties, requirements, qualifications, and other data; and includes a CD-ROM with a PowerPoint presentation".

The record includes the record's template id in our experiments, URLs of the resource, date, Dewey class, topic, and descriptions. The corresponding resource provided by OCLC is shown in figure.

From the Figure 2 we can find many useful metadata besides the metadata already used in our recommending methods, such as writer, publisher, database, format, etc.

Figure 2.
A record published by OCLC

More applications and recommendation can be applied based on these metadata. For example, after read a book recommended by our recommendation system through the daily news, users may be also interested in the books written by the same author. Also, they may be interested in the books published by the same publisher. There could be more utilization of the library's resources based on the results of our recommendation, and these applications can also promote information sharing and knowledge discovery.

Conclusion

In this paper, we have proposed a novel approach to link the library to everyday life through daily news using the WMF model. In the proposed method, we construct a newswire-library linkage by offering the relevant catalogue records from library to users who are interested in the subject topics of the target news articles. A number of experiments have been carried out and the results showed that our method can find out the relevant subject topics for both catalogue records and news articles. The performance of our method also overtakes the baselines in our experiments.

In future work, we will adopt more metadata from libraries in knowledge discovery and information sharing. A new model utilizing more semantic information should be found to improve accuracy and provide better results for the libraries' resources recommendation.

Acknowledgements

The authors thank for the precious work of the anonymous reviewers. This work is supported by the Ministry of Education of Humanities and Social Science Youth funded project (Grant No. 15YJC870002), the project of Social Science Foundation of Jiangsu Province, China (Grant No. 14TQC003), General University Science Research Project of Jiangsu Province (Grant No. 15KJB520013), the Major Projects of National Social Science Fund of China (Grant No. 13&ZD174), Fundamental Research Funds for the Central Universities (20620140647).

Downloaded by TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES At 20:39 10 November 2016 (PT)

Downloaded by TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES At 20:39 10 November 2016 (PT)

References

Alfonseca, E., Pighin, D. and Garrido, G. (2013), "HEADY: news headline abstraction through event pattern clustering", *Proceedings of ACL 2013, The 51st of Annual Meeting of Association of Computational Linguistics*, *Sofia*, available at: www.aclweb.org/anthology/P13-1122 (accessed 6 June 2014).

Booth, K. and Napier, J. (2003), "Linking people and information: web site access to National Library of New Zealand information and services", *The Electronic Library*, Vol. 21 No. 3, pp. 227-233.

Caidi, N. (2006), "Building 'civilizational competence': a new role for libraries?", *Journal of Documentation*, Vol. 62 No. 2, pp. 194-212.

Calhoun, K., Cantrell, J., Gallagher, P., Hawk, J. and Gauder, B. (2009), "Online catalogs: what users and librarians want: an OCLC report", OCLC, Dublin, OH, available at: www.oclc.org/content/dam/oclc/reports/onlinecatalogs/fullreport.pdf (accessed 6 June 2014).

Chen, J.P., Feng, S. and Liu, J. (2014), "Topic sense induction from social tags based on non-negative matrix factorization", *Information Sciences*, Vol. 280, pp. 16-25.

De Rosa, C. (2005), "Perceptions of libraries and information resources: a report to the OCLC membership", OCLC Online Computer Library Center, Dublin, OH, available at: www.oclc.org/reports/2010perceptions.en.html (accessed 6 June 2014).

Golub, K. (2006), "Automated subject classification of textual web pages, based on a controlled vocabulary: challenges and recommendations", *New Review of Hypermedia and Multimedia*, Vol. 12 No. 1, pp. 11-27.

Goulding, A. (2009), "Engaging with community engagement: public libraries and citizen involvement", *New Library World*, Vol. 110 Nos 1/2, pp. 37-51.

Holmberg, K., Huvila, I., Kronqvist-Berg, M. and Widen-Wulff, G. (2009), "What is Library 2.0?", *Journal of Documentation*, Vol. 65 No. 4, pp. 668-681.

Jin, Y. and Peng, J. (2009), "Information portal development and practice at Shanghai Jiao Tong University Library", *Online Information Review*, Vol. 33 No. 3, pp. 537-547.

Joorabchi, A. and Mahdi, A.E. (2014), "Towards linking libraries and Wikipedia: automatic subject indexing of library records with Wikipedia concepts", *Journal of Information Science*, Vol. 40 No. 2, pp. 211-221.

Khoir, S., Du, J.T. and Koronios, A. (2015), "Everyday information behaviour of Asian immigrants in South Australia: a mixed-methods exploration", *Information Research*, Vol. 20 No. 3, available at: www.informationr.net/ir/20-3/paper687.html#.VrQSnfSl_ak (accessed 20 September 2015).

Marijan, R. and Leskovar, R. (2015), "A library's information retrieval system (In)effectiveness: case study", *Library Hi Tech*, Vol. 33 No. 3, pp. 369-386.

Murphy, B. (2012), "OCLC provides downloadable linked data file for the 1 million most widely held works in WorldCat", OCLC, Dublin, OH, available at: www.oclc.org/news/releases/2012/201252.en.html (accessed 6 June 2014).

Pisanski, J. and Žumer, M. (2005), "National library web sites in Europe: an analysis", *Program: Electronic Library and Information Systems*, Vol. 39 No. 3, pp. 213-226.

Pu, Y.-H., Chiu, P.-S., Chen, T.-S. and Huang, Y.-M. (2015), "The design and implementation of a Mobile Library APP system", *Library Hi Tech*, Vol. 33 No. 1, pp. 15-31.

Rodrigues-Gairin, J.-M. and Somoza-Fernandez, M. (2014), "Web services to link interlibrary software with OCLC WorldShare", *Library Hi Tech*, Vol. 32 No. 3, pp. 483-494.

Rodriguez, A.C., Gago, J.M.S., Rifon, L.E.A. and Rodriguez, R.P. (2015), "A recommender system for non-traditional educational resources: a semantic approach", *Journal of Universal Computer Science*, Vol. 21 No. 2, pp. 306-325, available at: www.jucs.org/jucs_21_2/a_recommender_system_for/jucs_21_02_0306_0325_rodriguez.pdf (accessed 15 July 2015).

Stajner, T., Thomee, B., Popescu, A.-M., Pennacchiotti, M. and Jaimes, A. (2013), "Automatic selection of social media responses to news", *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Chicago, IL*.

Steck, H. (2010), "Training and testing of recommender systems on data missing not at random", *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC*.

Wei, Z. and Gao, W. (2014), "Utilizing microblogs for automatic news highlights extraction", *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics, Dublin*, available at: http://aclweb.org/anthology/C14-1083 (accessed 15 July 2015).

Zarei, H. and Abazari, Z. (2011), "A study of web-based services offered by Asian national libraries", *The Electronic Library*, Vol. 29 No. 6, pp. 841-850.

About the authors

Wei Yu is currently an Associate Professor in the Department of Information Management at the Nanjing University, China. She received a PhD Degree in the Computer School from the Wuhan University, China in 2007, and MSc Degree in the Computer School from the Wuhan University, China in 2004, and a BSc Degree in the Computer School from the Wuhan University, China in 2002. Her research interests include mainly digital library, semantic web, natural language processing, etc.

Junpeng Chen is currently a Lecturer in the School of Information Engineering at the Nanjing University of Finance & Economics, China. He received a PhD Degree in the Computer School from the Wuhan University, China in 2012, and MSc Degree in the Computer School from the Wuhan University, China in 2004, and a BSc Degree from the Wuhan University, China in 2001. His research interests include mainly natural language processing, text mining, information organization, etc. Junpeng Chen is the corresponding author and can be contacted at: luckjp@163.com

Downloaded by TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES At 20:39 10 November 2016 (PT)

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: **permissions@emeraldinsight.com**