REPRESENTATION WITH WORD CLOUDS AT THE PHD UNS DIGITAL LIBRARY

Georgia Kapitsaki[1] and Dragan Ivanović[2]

[1]University of Cyprus, Cyprus
[2]University of Novi Sad, Serbia

ABSTRACT

Many systems provide search and recommendation capabilities to scholars that search for scientific documents including research papers and dissertations. The appearance of search results may largely affect the system use. Traditional approaches provide textual formats for showing the results to users, whereas more recent approaches concentrate on other forms, e.g., on two dimensions. Moreover, this presentation may be adapted to user needs providing a personalised user experience combined with other contextual factors, such as enriching user search with keywords from recently used documents. In this paper, we present our work on results representation in the framework of a dissertation search engine in the Serbian language with the ultimate aim to provide a more personalised experience to users. We have integrated our approach in the PhD UNS digital library system of the University of Novi Sad, a research information, library and educational information system, and are discussing as early evaluation how users are perceiving this approach outlining also our vision for a context-aware digital library system. The initial results demonstrate the usefulness of providing more choices to the users adapting application to their needs.

KEYWORDS

PHD UNS, word cloud, search results, log analysis

1. INTRODUCTION

Discovering information on the web is not always a trivial task for researchers that aim to examine previous research works that can be used as basis or reference for future research. Many systems provide the opportunity to scholars to search for papers and dissertations providing also relevant recommendations to users based on their areas of interest. Although the Google search engine is considered a superior solution to more elaborated library systems, such elaborated systems may provide more benefits in specific contexts, e.g., when searching for dissertations in specific languages, countries or institutes as addressed also in the framework of the current work [1].

Different information retrieval approaches are used in these systems for search purposes, whereas many systems integrate also recommendation mechanisms recommending relevant scientific papers or dissertations to users[2]. Although the Google Scholar recommendation is a very popular solution for scholars, additional systems may utilize different information about users in order to retrieve information. For instance, the Scienstein research paper recommender system enhances the keyword-based search by combining it with citation analysis, author analysis, source analysis, implicit ratings and explicit ratings [3].

Regardless of whether a more general or more specific system is used, the presentation of search results to users is important, as it widely affects how they perceive the system and may reduce or increase the chances of using the system and the frequency of use. In recommender systems, it is argued that the system's user interface in general (e.g., the display of predictions at the time users rate items) may even affect user's opinion [4].The visualization may then be adapted to user needs, presenting results either in a textual or an alternative format depending on how users respond to the alternative presentations of the system. This can form part of a personalised context-aware system that considers user's environment, history and interaction with the system in order to act proactively and adapt the input and result to each user. Context-awareness is an inherent part of many systems in different domains, e.g., web services, mobile computing, where the application or system functionality adapts to the context of use [5, 6]. Context refers in most cases in any information that is relevant to the user, the system and any interaction between the user and the system[7].

Taking into consideration the above, in this paper we are presenting our work toward results representation in the framework of context-aware information provision for dissertations for scholars. In the framework of the PhD UNS digital library (DL) we are aiming at providing personalised services to the users[8]. The vision of this process is briefly described in a previous work of the authors [9].The PhD UNS digital library includes doctoral dissertations in the Serbian language with the motivation of providing access to research data as a step toward the development of a knowledge-based society. In this paper, we focus on the presentation of the results visualization component for the search results for the users of the PhD UNS system. A new way of presenting the search results to the users was conceived, designed and implemented. Specifically, the presentation of the content of a PhD dissertation as a word cloud was addressed. Word clouds are currently widely used in different systems. A word or tag cloud is a visual representation of word content commonly used to represent content in different environments [10].

Subsequently, users have the opportunity to provide their feedback on this visualization indicating in essence whether they prefer the textual or the new graphical presentation of the results (changing from one representation to the other). The feedback was then used to adapt the results based on user preference. We are using this component as the initial step toward a fully personalised system, where different context parameters will be considered for providing a personalised context-aware user experience. At the current state, the user feedback is provided for personalisation purposes for the results appearance and is used in subsequent uses of the system. Personalisation has also been integrated into the recommendations provided by the system as presented in a previous work of the authors, whereas additional considerations are outlined as future work [11].

The contributions of this work is twofold:

-   We perform a study of word clouds as a visualization approach for digital libraries search results and we evaluate this approach in the framework of the PhD UNS digital library. Although the results have been used in a digital library in the Serbian language, they can be easily replicated in libraries implemented in different languages.

-   A side contribution can be found in the vision of results personalisation introduced toward a context-aware user experience.

The rest of the paper is structured as follows. Section 2 outlines previous related work in the area of recommendations for scholars and the results visualization considering also word clouds. Section 3 presents the PhD UNS digital library and its use. Section 4 is dedicated to the presentation of the new word cloud generation component including also implementation details.

The integration and use of the component in the framework of PhD UNS digital library is presented in section 5. Section 6 is dedicated to evaluation results describing how users have perceived the system and finally, section 7 concludes the paper outlining also future research directions.

2. RELATED WORK

2.1. Results visualization

Previous works have focused on personalising search or recommendation results. The process of presenting to users results in formats other than textual has been studied by many researchers in the past, in order to improve user experience in search engines, information retrieval approaches and recommender systems.

A controlled comparison of text, 2D, and 3D approaches to a set of typical information seeking tasks on a collection of 100 top ranked documents retrieved from a much larger document set was presented in [12]. The experiments conducted included the participation of 15 individuals. The study revealed that although a visualization can assist the reduction of the mental workload for interpreting the results, these reductions and their acceptance depend on an appropriate mapping among the interface, the task and the user. In relevance to the above, our approach lies in the area of 2D display of information, but instead of focusing on basic text information we have adopted newer approaches found in word clouds. Visualization has also been addressed in even earlier works in the framework of database search [13].

Most and more recent workshave examined visualization in web search, such as in [14] that presents an approach for the clustering of search engine results that relies on the semantics of the retrieved documents. The approach takes into consideration both lexical and semantics similarities among documents and applies activation spreading technique, in order to generate clusters based on semantic properties. In [15], a model for web search visualization is proposed, where physical location, spatial distance, color and movement of graphical objects are used to represent the degree of relevance between a query and relevant web pages considering this way the context of users' subjects of interest. Previous works are thus trying to use different document properties in order to improve results visualization. However, we rely mostly on the document content, as it can better summarize the dissertation but focus on providing a different presentation.

2.2. Word clouds

According to Wikipedia, a word cloud is a "*visual representation of text data, typically used to depict keyword metadata (tags) on websites, or to visualize free form text. Tags are usually single words, and the importance of each tag is shown with font size or color.*" As aforementioned, work clouds are used in different environments, whereas they are a popular way of representing information on the web summarizing the content of documents and other sources of information. Previous works have introduced various algorithms for the tag selection or new ways for the word cloud creation [16, 17, 18].

Tag clouds have been used in PubCloud for the summarization of results from queries over the PubMed database of biomedical literature[19].PubCloud responds to queries of this database with tag clouds generated from words extracted from the abstracts returned by the query. The authors found that the descriptive information is this way provided in a better way to users. However, the discovery of relations between concepts is rendered less effective. This approach has similarities

with our work, since it addresses the visualization in the framework of scientific literature, but focuses on database queries and considers only document abstracts for the tag cloud generation.

2.3. Context-awareness

Context-aware services are relevant in diverse domains. Mobile computing and pervasive computing offer the necessary information from sensors on the mobile device and in user's environments for context-aware application provision [20].Personalisation may adapt various features (e.g., presentation, structure) in order to address specific needs of each individual [21].In the framework of the web, and in web search and information retrieval systems, many systems utilize user's search history in order to offer personalised search. In the work of [22] it was found that personalisation based on short-term history or "within-session" behavior is less valuable than long-term or "across-session" personalisation.

In the area of digital libraries, an approach to construct personalised digital libraries that satisfy a user's necessity for information is introduced in [23]. Adaptive digital libraries are libraries that automatically learn user preferences and goals and personalise their interaction using this information. Based on this work he authors go further to develop a personalised digital library to suit the needs of different cognitive styles[24]. Adaptability versus adaptivity were investigated in a digital library and it was found that users performed better and perceived the adaptive version more positively. Both studies indicated that cognitive styles influence users' preferences for the use of digital libraries and are for this reason considered also in the framework of our work.

A number of tools have focused on providing personalised recommendations [25, 26]. However, we are not considering such works further, as the presentation of the recommendation process in PhD UNS digital library is outside the scope of the current paper.

2.4. Contribution of our work

We share similarities with previous works in terms of techniques used, as for instance word clouds have been used in other systems as well in order to improve the user experience. However, in contrast to previous works we apply a new visualization technique in a specific context, a Serbian digital library, allowing automatic adaptation for the appearance of search results based on user's reaction. This visualization concept tested in a real setting forms part of a wider context-aware experience for users.

3. PHD UNS

The current Research Information System of the University of Novi Sad (CRIS UNS) has been developed since 2008[27]. Digital Library of Dissertations of the University of Novi Sad (PhD UNS) has been developed and integrated with the CRIS UNS system since 2011 [28, 29]. The implemented digital library enables support for the processes relevant to the educational aspect of PhD studies (release thesis for public review, promotion of PhDs, etc.). Someone can see this as an integrated system of research information system, library and educational information system. This integrated system contributes to: 1) avoiding duplicated inputs on the three platforms and thus decrease number of the university stuff necessary for this work; 2) increasing metadata quality, reliability and reusability; 3) increasing quality level of services based on these metadata. PhD UNS is being used since December 2013. The University of Novi Sad Senat passed the decision that the upload of PhD thesis in PhD UNS system is the obligatory step before defending the thesis. However, all dissertations defended before December 2013 can also be scanned using Quidenus Mastered Book Scan 3.0 and be stored in the PhD UNS digital library.

Besides e-thesis, signed licenses for copyrights transfer are also stored in PhD UNS. There were two options by December of 2014: publishing dissertation under one of the six levels of Creative Commons License and publishing dissertation under a non-open-access license. Dissertations published under Creative Commons License can improve dissemination of knowledge stored in dissertations. This fact was recognized by the academic community at the University of Novi Sad. More than 90% (232 of 255) dissertations defended in the first year after putting the digital library into operation (December of 2013 – December of 2014) have been published under open-access licenses. Since December of 2014, PhD. candidates must sign open-access statement which is in accordance with the new Serbian Regulations on Higher Education.

Cataloguing within integrated system of CRIS UNS and PhD UNS is done using the MARC 21 format [30, 31]. This enables the easy integration with library information systems based on MARC formats, such as BISIS system used by Central Library of University of Novi Sad. There is a web application for searching digital library which enables search by metadata and full text of PhD dissertations [32]. Metadata for all 5,000 dissertations ever defended at the University of Novi Sad (since 1955) are stored in Digital Library of Dissertations and can be searched via web application. The search engine of the digital library has been implemented using the Apache Lucene information retrieval library [33]. The set of catalogued metadata and full text of dissertations have been pre-processed and indexed using the custom Lucene Analyzer. The Analyzer includes the following processing steps: transformation from the Cyrillic alphabet to the Latin alphabet; stop word removal based on a stop word list for the Serbian language; and stemming. More than 250,000 downloads of dissertations using the web application for searching the PhD UNS digital library have been recorded in the system logs till this point.

Also, the model of integrated system enables export of dissertation's metadata through OAI-PMH protocols in Dublin Core, MARC 21, ETD-MS and CERIF format. It enables interoperability of the PhD UNS digital library and OAI-PMH compatible institutional repositories, CERIF based information systems, MARC based library information systems, network of digital libraries based on OAI-PMH protocol, such as NDLTD, DART Europe and OATD. More than 800 open-access dissertations are exported via the OAI-PMH protocol to those networks.

4. WORD CLOUD GENERATOR COMPONENT

The work cloud generator component forms now part of the PhD UNS digital library. As aforementioned, its aim is to present user search results in a word cloud representation. Specifically, users can search for dissertations using a number of keywords as performed in similar scholar systems. The new component targets the way of presenting results to users with the aim of adapting the results visualization to the user's preferred mode, i.e., textual versus graphical. In order to achieve this, the word cloud generator component performs a number of actions on the dissertation texts as displayed in Figure 1.

The word cloud component was implemented in Java and creates as output an image (currently in PNG format) with a work cloud for the text of a PhD dissertation.The tool uses as input the PDF file of the dissertation, it then parses the textual content of the file, performs a transformation from the Cyrillic alphabet to the Latin alphabet, in order to allow the easier subsequent processing, and performs actions traditionally used in information retrieval, in order to process the document text and choose the most important words from the text : stop word removal based on a stop word list for the Serbian language and stemming. The result of pre-processing is list of pairs containing original version of word from the text and its stem. The details of the tool utilized for this pre-processing step can also be found in a previous publication [11].
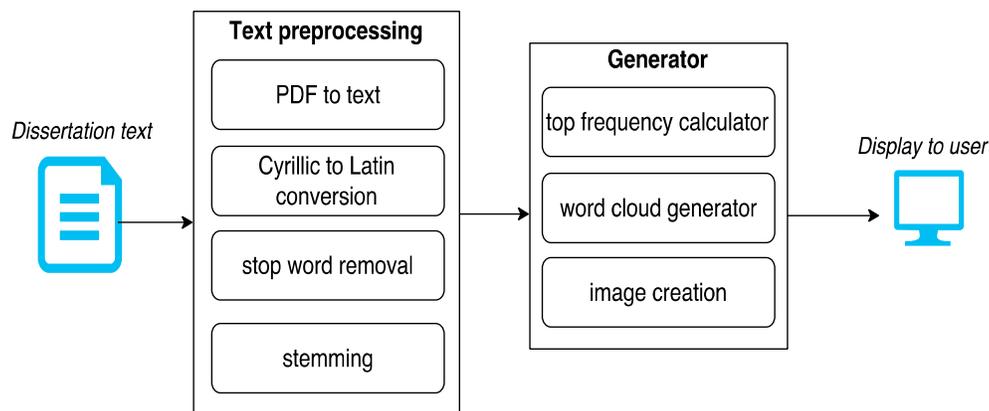
Figure 1. Word Cloud generator steps

The tool then proceeds to calculate the top frequencies of words in the text, generates the word cloud and creates an image file. All words are considered equal indetermining the importance of a keyword., while they are tokenized based on the existence of white spaces between the words in each sentence examined. The whole document text is considered instead of the abstract for instance, as we did not want to omit important words that appear in the text and may repeat many times. It is also usual for some abstracts to appear in English, but we wanted to capture the whole text. When calculating the word frequencies, the stemmed version of the words is used, in order to consider the different appearances of the same word (e.g., as noun, verb, etc.). Note however, that the words in the cloud are displayed in their original – and not the stemmed – version, in order to be better understandable to the users. The original word with the highest frequency for the respective stemmed word is presented to the user. Numbers and short words (this number of letters for small words is currently set to 4) are ignored.

For implementation purposes, the Kumo library in Java was used [34]. Kumo carries the MIT license and its code has been extended to accommodate the needs of the PhD UNS digital library. The tool can be adapted to consider a different number of keywords or use different colors for the word cloud creation.

5. INTEGRATION TO PHD UNS

The word cloud generator component described in the previous section has been integrated in the PhD UNS digital library application and has been put into operation in April, 2017. The component accepts a PDF file as input and generates an image (PNG file) as output. The information retrieval process includes two phases: indexing and searching. The purpose of creation and storing an index (indexing) is to optimize speed and performance in finding relevant documents for a search query. Without an index, searching would require considerable time and computing power. Taking into account that the word cloud generator is time-prone and a high-computing process, it is invoked in the phase of indexing and generated image is stored as supplement material to a PhD dissertation in the server file system. Figure 2 presents a Unified Modeling Language (UML) activity diagram which describes the process of adding new dissertation to the PhD UNS digital library. The activity *Generate word cloud image* is highlighted with red background and represents invoking the execution of the word cloud component described in the previous section. Moreover, the activity *Create Lucene index* includes the same steps for text pre-processing as the steps described in the word cloud generator component.
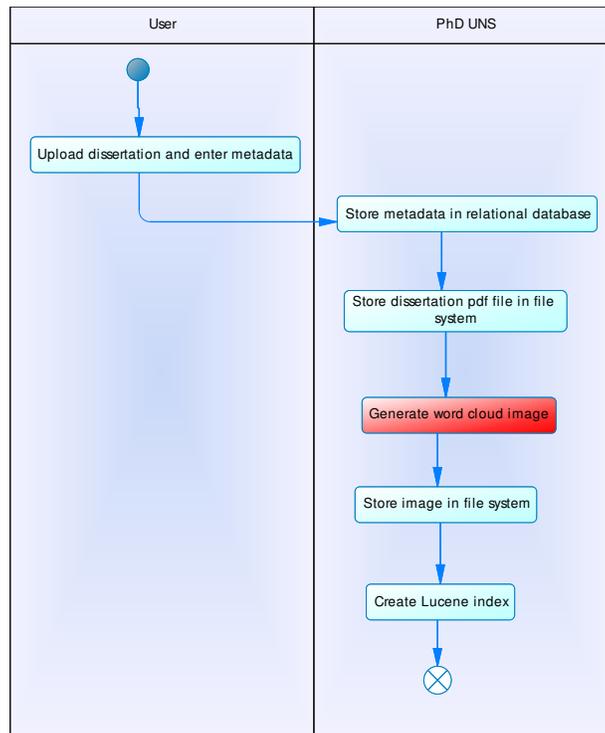
Figure 2. Adding a dissertation to the system

The default way of representing the search results to the users which access the PhD UNS search web page for the first time is randomly selected among these two choices:1) display the results as references in Harvard representation style or 2) display the results as a word cloud image. Portions of screenshots depicting how the results are displayed in textual and word cloud representation are found in Figure 3. However, the user can request the change of the representation style as depicted in the process of Figure 4 (style option in the screenshots of Figure 3), providing this way her feedback and indicating her preference for the results visualization
.

6. EARLY EVALUATION

If the representation style is changed by a user of the digital library, the relevant message is stored in the server log files along with other information for the use of the system. Listing 1 shows an example of an entry in the log file with the various data stored for each user interaction with the PhD UNS system. Information, such as date and time of access, location of access, IP address, user device and representation style, are available. The representation style is also stored in cookies on the browser on the client side as the user preferred representation style and will be considered as the default visualization style for the certain user for any future access to the PhD UNS search web page. We have used the log analysis, in order to perform an evaluation of the early results of our approach and examine how users reacted to the new representation. Note that we are currently utilizing only a subset of the information available in the log files. Additional data will be used in the future in order to add more context-aware features to the digital library.

Number of results is 493

Sort by: relevance   Style: structured

- БЈЕЛИЦА, М. (2013)*Methods of implementation of context-aware platforms and context-aware user interfaces for applications in consumer electronics.* (PhD dissertation), Faculty of Technical Sci
- MILANKOV, V. (2016)*Phonological awareness deficit in children with dyslexia and dysorthography.* (PhD dissertation), Faculty of Medicine at Novi Sad
- BIKICKI, N. (2010)*UTICAJ ZNAČENJA PARTIKULE PROZIRNIH FRAZALNIH GLAGOLA NA NJIHOVO USVAJANJE.* (PhD dissertation), Faculty of Philosophy at Novi Sad
- SLADIĆ, G. (2011)*Context Sensitive Access Control Model TI for Business Processes.* (PhD dissertation), Faculty of Technical Sciences at Novi Sad
- МАТАНОВИЋ, Ј. (2016)*The significance of demographic and psychologicalcharacteristics for realistic and intended consumer behavior.* (PhD dissertation), Faculty of Philosophy at Novi Sad
- ERDEJI, I. (2017)*Individual and organisational predictors of prosocial service behaviour amongemployees in cruise line industry.* (PhD dissertation), Faculty of Sciences at Novi Sad
- ЛАЈШИЋ, Х. (2016)*IT-Supported Development of Human Resources Management Model.* (PhD dissertation), Doktorske disertacije iz interdisciplinarne odnosno multidisciplinarne oblasti na Unive
- ANIĆ, I. (2011)*Cognitive processes in solving mathematical problems in real context.* (PhD dissertation), Faculty of Sciences at Novi Sad
- ГАК, Д. (2016)*Designing Business English Course for the Purposeof Developing Pragmatic Competence Important forthe Advancement of Business Communication.* (PhD dissertation), Faculty of
- RADOVANOVIĆ, M. (2011)*High-Dimensional Data Representations and Metrics for Machine Learning and Data Mining.* (PhD dissertation), Faculty of Sciences at Novi Sad
- NEDELJKOVIĆ, U. (2016)*Universal Type: Modernist Utopia or Current Communication Requirement.* (PhD dissertation), Faculty of Technical Sciences at Novi Sad
- MIŠKOVIĆ, D. (2017)*Context-Dependent Speech Recognition in Human-Machine Interaction.* (PhD dissertation), Faculty of Technical Sciences at Novi Sad

(a)

(b)

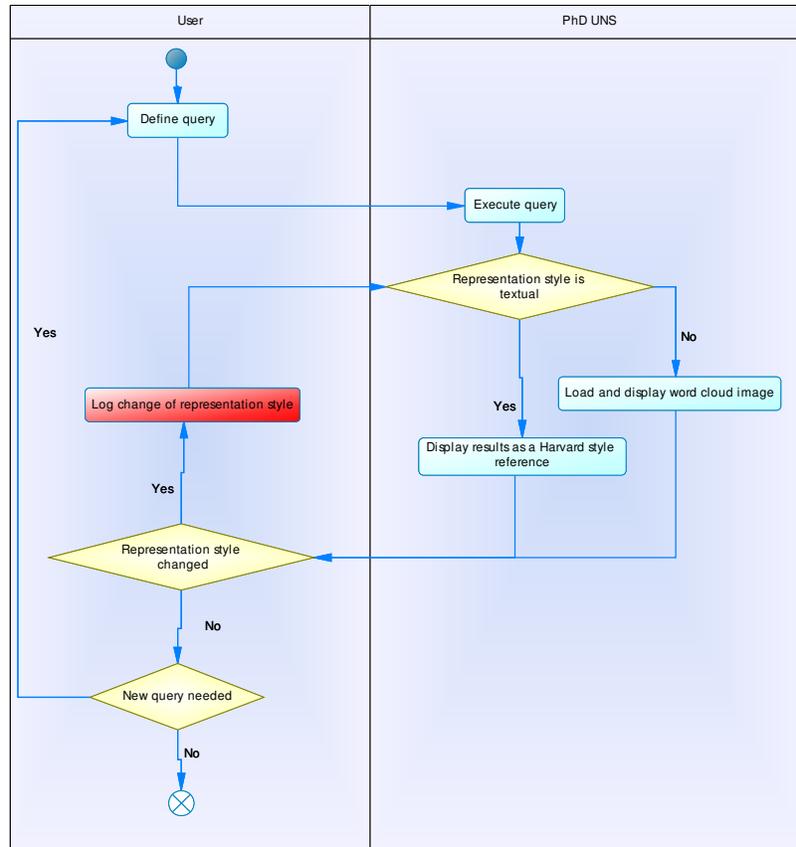Figure 3. PhD UNS results for search keyword "context-awareness" in (a) textual and (b) word cloud representation

Figure 4. Query execution for representation feedback

[INFO] 26.04.2017. 12:52:16 (SearchDissertationsManagedBean:setRepresentationStyle)
Date and time: Wed Apr 26 12:52:16 CEST 2017|
miliseconds: 1493203936644| +
session id: FFBAA80BECE0B982EBA313A2DE38CCC9|
userId: 149320389739914|
ip address: 147.91.177.241(proxy = 147.91.173.31)|
location: city: Belgrade, **postal code**: null, **regionName**: null (region: 00), **countryName**: Serbia (country code: RS), **latitude**: 44.818604, **longitude**: 20.468094|
user agent (device): Mozilla/5.0 (Windows NT 5.1) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/49.0.2623.112 Safari/537.36|
new representation style: wordCloud

Listing 1. A log message example

Information from the users of the PhD system for the first months of use of the new representation component were collected. Log messages for the first four months of the new feature usage (April – August 2017) were for this purpose imported into a MySQL database for the purpose of log analysis. A total of 7,052 queries were defined and executed by 3,023 PhD UNS users during this period. The randomly selected representation style was changed 528 times by the PhD UNS users: it was changed 450 times to the textual representation style and 78 times to the visual (word cloud image) representation style. We can conclude that generally a lot of users accept both types of representation style taking into account that 2,495 of 3,023 users did not choose to change the representation style selected for them by the system. Furthermore, there are much more users which changed the representation style to textual than to visual representation style. This may be attributed to the fact that users are more familiar with textual

representation in their interactions with scholar systems (e.g., reference style of research papers). The provision of additional adaptability in the visualization (e.g., provide to the user the possibility to change the colors or the size of the word cloud) may also help in understanding whether the representation parameters affect the way users react to it.

However, we also observed that there are users which changed representation styles a few times; we suppose that this change was performed since they needed a different representation style for different types of information needs. We have not noted however, any decrease in the number of visitors during the period the word cloud representation as introduced, indicating that the user's frequency of use of the system and user's opinion was not largely affected. Further log analysis will be performed in the next period of use of the system, as the personalised features will expand allowing to draw more useful conclusions, where a larger number of users will be considered.

7. CONCLUSIONS

In this paper, we have presented our work on the visualization of recommendations for scholars in the framework of the PhD UNS digital library. We have used a word cloud in order to study how users react to this new representation and then adapt the default results presentation for each user based on the feedback received by users. To the best of our knowledge, this is the first work that addresses the study of results representation for the Cyrillic alphabet. The initial results obtained from analyzing the log files of the system demonstrate that most users accept both representation styles, as only a small percentage of users chose to change the randomly chosen representation style during the use of the system. Further evaluation is required, in order to study whether the adaptability of this word cloud representation can improve the way users interact with PhD UNS.

As future work, we intend to extend the personalised features of the system providing context features to the user searchers via the enrichment of user queries and results for scientific documents retrieval with context information. We intend to use different information sources for this purpose, such as keywords from dissertations texts, user device motion information(when a mobile device is used to access the system), user device battery information, keywords from previous user searches, and keywords from user's publications and the publications of user's collaborators (available only for registered users). These adaptations will realize our vision for a personalised digital library system for the Serbian language. Note that we want to preserve also user's privacy at the time of providing a personalised solution. Although all user data is currently anonymous and most visitors are guest users of the system, we will target in the future additional privacy protection mechanisms for registered users.

ACKNOWLEDGEMENTS

REFERENCES

[1]   Brophy, Jan, & David Bawden (2005) "Is Google enough? Comparison of an internet search engine with academic library resources",Aslib Proceedings. Vol. 57. No. 6. Emerald Group Publishing Limited.

[2]   Beel, Joeran, Langer, Stefan, Genzmehr, Marcel & Nuernberger. Andreas (2013) "Introducing Docear's research paper recommender system", In Proceedings of the 13th ACM/IEEE-CS joint conference on Digital libraries (JCDL '13), ACM, pp459-460.

[3]   Gipp, Bela, Beel, Joeran& Hentschel, Christian (2009) "Scienstein: A research paper recommender system." Proceedings of the international conference on emerging trends in computing (icetic'09).

[4]   Cosley, Dan, Shyong K. Lam, Istvan Albert, Joseph A. Konstan& John Riedl (2003) "Is seeing believing?: how recommender system interfaces affect users' opinions", In Proceedings of the SIGCHI conference on Human factors in computing systems, ACM, pp585-592.

[5]   Hong, Jong-yi, Suh, Eui-ho & Kim, Sung-Jin (2009) "Context-aware systems: A literature review and classification." Expert Systems with Applications, Vol. 36, No. 4, pp8509-8522.

[6]   Kapitsaki, Georgia M., Prezerakos, George N., Tselikas, Nikolaos D.& Venieris, Iakovos. S. (2009)"Context-aware service engineering: A survey", Journal of Systems and Software, Vol. 82, No. 8, pp1285-1297.

[7]   Abowd, Gregory, Dey, Anind, Brown, Peter, Davies, Nigel, Smith, Mark & Steggles, Pete (1999) "Towards a better understanding of context and context-awareness", In Handheld and ubiquitous computing, Springer Berlin/Heidelberg, pp304-307.

[8]   Borgman, Christine L. (1999) "What are digital libraries? Competing visions", Inf. Process. Manage., Vol. 35, No. 3, pp227-243.

[9]   Ivanovic, Dragan, & Kapitsaki, Georgia M., (2015) "Personalisation of Keyword-based Search on Structured Data Sources", 1st International KEYSTONE Conference (IKC 2015).

[10]   Scanfeld, Daniel, Vanessa Scanfeld, & Elaine L. Larson, (2010) "Dissemination of health information through social networks: Twitter and antibiotics", American journal of infection control, Vol. 38, No. 3, pp182-188.

[11]   Azzopardi, Joel, Ivanovic, Dragan, Kapitsaki, Georgia M. (2016) "Comparison of Collaborative and Content-Based Automatic Recommendation Approaches in a Digital Library of Serbian PhD Dissertations", Proceedings of the International KEYSTONE Conference 2016, pp100-111.

[12]   Marc M. Sebrechts, John V. Cugini, Sharon J. Laskowski, Joanna Vasilakis, & Michael S. Miller. (1999) "Visualization of search results: a comparative evaluation of text, 2D, and 3D interfaces", In Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '99), ACM, pp3-10.

[13]   Bowers, Frank H., and Stuart K. Card, (1996) "Method and apparatus for visualization of database search results", U.S. Patent No. 5, pp546,529.

[14]   Soliman, Sara Saad, El-Sayed, Maged F. & Hassan, Yasser F. (2015) "Semantic Clustering of Search Engine Results", The Scientific World Journal.

[15]   Nguyen, Tien, & Zhang, Jin (2006) "A novel visualization model for web search results",IEEE transactions on visualization and computer graphics, Vol. 12, No. 5.

[16]   McNaught, Carmel& Lam, Paul (2010) "Using Wordle as a supplementary research tool", The qualitative report, Vol. 15, No. 3, pp630.

[17]   Cui, Weiwei, Yingcai Wu, Shixia Liu, Furu Wei, Michelle X. Zhou, & Huamin Qu, (2010) "Context preserving dynamic word cloud visualization", In IEEE Pacific Visualization Symposium (PacificVis), pp121-128.

[18]   Hassan-Montero, Yusef & Herrero-Solana, Victor (2006) "Improving tag-clouds as visual information retrieval interfaces", In Proceedings of the International conference on multidisciplinary information sciences and technologies, pp25-28.

[19] Byron Y-L Kuo, Thomas Hentrich, Benjamin M. Good& Mark D. Wilkinson. (2007) "Tag clouds for summarizing web search results", In Proceedings of the 16th international conference on World Wide Web (WWW '07). ACM, pp1203-1204.

[20] Raento, M., Oulasvirta, A., Petit, R., & Toivonen, H. (2005) "Context Phone: A prototyping platform for context-aware mobile applications", IEEE pervasive computing, Vol. 4, No. 2, pp51-59.

[21] Fink, Josef &Kobsa, Alfred (2000) "A review and analysis of commercial user modeling servers for personalisation on the world wide web",User Modeling and User-Adapted Interaction, Vol. 10, No. 2, pp209-249.

[22] Yoganarasimhan, Hema, (2016) "Search personalisation using machine learning".

[23] Frias-Martinez, E., Magoulas, G., Chen, S. & Macredie, R., (2006)"Automated user modeling for personalised digital libraries" ,International Journal of Information Management, Vol. 26, No. 3, pp234-248.

[24] Frias-Martinez, Enrique, Sherry Y. Chen, & Xiaohui Liu, (2009), "Evaluation of a personalised digital library based on cognitive styles: Adaptivity vs. adaptability", International Journal of Information Management, Vol., 29, No. 1, pp48-56.

[25] Tejeda-Lorente, Álvaro, Carlos Porcel, Eduardo Peis, Rosa Sanz, & Enrique Herrera-Viedma, (2014) "A quality based recommender system to disseminate information in a university digital library", Information Sciences, Vol. 261, pp52-69.

[26] Beel, Joeran, Langer, Stefan, Kapitsaki, Georgia, Breitinger, Corinna & Gipp, Bela (2015) "Exploring the potential of user modeling based on mind maps", In Proceedings of the International Conference on User Modeling, Adaptation, and Personalisation, Springer, pp3-17.

[27] Surla, Dušan, Ivanović, Dragan & Konjović, Zora (2013) "Development of the software system CRIS UNS." In Proceedings of the Intelligent Systems and Informatics (SISY), 2013 IEEE 11th International Symposium, pp111-116.

[28] CRIS PhD UNS, http://dosird.uns.ac.rs/phd-uns-digital-library-phd-dissertations.

[29] Ivanović, Lidija, Ivanović, Dragan & Surla, Dušan (2012) "Notes on Operations: Integration of a Research Management System and an OAI-PMH Compatible ETDs Repository at the University of Novi Sad, Republic of Serbia", Library Resources & Technical Services, Vol.56, No. 2, pp 104-112.

[30] Ivanović, Dragan, Surla, Dušan & Konjović, Zora (2011)"CERIF compatible data model based on MARC 21 format", The Electronic Library, Vol. 29, No. 1, pp52-70, DOI: 10.1108/02640471111111433.

[31] Ivanović, Lidija, Ivanović, Dragan& Surla, Dušan(2012)"A data model of theses and dissertations compatible with CERIF, Dublin Core and EDT-MS", Online Information Review, Vol. 36, No. 4, pp568-586, DOI 10.1108/14684521211254068.

[32] Ivanović, Lidija, Ivanović, Dragan, Surla, Dušan & Konjović, Zora (2013) "User interface of web application for searching PhD dissertations of the University of Novi Sad", In Proceedings of the Intelligent Systems and Informatics (SISY), 2013 IEEE 11th International Symposium, pp117-122.

[33] Apache Lucene, https://lucene.apache.org/.

[34] Kumo, https://github.com/kennycason/kumo.

AUTHORS

Georgia Kapitsaki is an Assistant Professor at the Department of Computer Science of the University of Cyprus and faculty member of the Software Engineering and Internet Technologies (SEIT) laboratory in UCY. She received her PhD from the National Technical University of Athens, Greece (2009). Her research interests include: software engineering, privacy enhancing technologies (PETs), context-awareness and service-oriented computing. She has published over 40 papers in international conferences and journals. She has participated in conference organisation (e.g. ICSR 2016) and has served as a TPC member in repudiated journals and conferences. She has been involved in EU projects and has worked as a software engineer in the industry.

Dragan Ivanović is an associate professor within Faculty of Technical Sciences, University of Novi Sad. His research interests are focused on digital documents management, information retrieval, research management information systems, software architectures and library standards. Dragan holds courses Digital documents management on master studies and Selected Topics of Digital Archives on PhD studies at the Computer Science Department – these courses include topics related to: library standards and formats, digital repositories (including OA policies and OA repositories), information retrieval, search personalisation, etc. Dragan has been involved in more than 10 research projects and published over 40 scientific papers.