ARTICLE

Personalization of Search Results Representation of a Digital Library

Ljubomir Paskali, Lidija Ivanovic, Georgia Kapitsaki, Dragan Ivanovic, Bojana Dimic Surla, and Dusan Surla

ABSTRACT

The process of discovering appropriate resources in digital libraries within universities is important, as it can have a big effect on whether retrieved works are useful to the requester. The improvement of the user experience with the digital library of the University of Novi Sad dissertations (PHD UNS) through the personalization of search results representation is the aim of the research presented in this paper. There are three groups of PHD UNS digital library users: users from the academic community, users outside the academic community, and librarians who are in charge of entering dissertation data. Different types of textual and visual representations were analyzed, and representations which needed to be implemented for the groups of users of PHD UNS digital library were selected. After implementing these representations and putting them into operation in April 2017, the user interface was extended with functionality that allows users to select their desired style for representing search results using an additional module for storing message logs. The stored messages represent an explicit change in the results representation by individual users. Using these message logs and ELK technology stack, we analyzed user behavior patterns depending on the type of query, type of device, and search mode. The analysis has shown that the majority of users of the PHD UNS system prefer using the textual style of representation rather than the visual. Some users have changed the style of results representation several times and it is assumed that different types of information require a different representation style. Also, it has been established that the most frequent change to the visual results representation occurs after users perform a query which shows all the dissertations from a certain time period and which is taken from the advanced search mode; however, there is no correlation between this change and the client's device used.

INTRODUCTION

In order to place their current work within a framework of previous methods or identify research gaps, researchers often need to identify and study previous research. Discovering information on the web is not always a trivial task. Many systems allow scholars to search for research papers, dissertations, and other technical reports, providing at the same time relevant recommendations to users based on their areas of interest or previous searches. Although web search engines are considered a superior solution to more specialized digital library systems, these specialized systems may provide more benefits in specific conditions, e.g., when searching for dissertations in specific languages, or by affiliated countries or institutions.[1] Nowadays, digital libraries are widely used by diverse communities of users for diverse purposes.[2] Xie and colleagues conducted an analysis in 2018 to compare similarities and differences in perceptions of the importance of

Ljubomir Paskali (ljubomir.paskali@gmail.com) PhD Student, University of Novi Sad, Serbia. **Lidija Ivanovic** (lidija.ivanovic@uns.ac.rs) Assistant Professor, University of Novi Sad, Serbia, she is a corresponding author. **Georgia Kapitsaki** (gkapi@cs.ucy.ac.cy) Associate Professor, University of Cyprus, Cyprus. **Dragan Ivanovic** (dragan.ivanovic@uns.ac.rs) Full Professor, University of Novi Sad, Serbia. **Bojana Dimic Surla** (bdimicsurla@raf.edu.rs) Full Professor, Union University, Serbia. **Dusan Surla** (surla@uns.ac.rs) Professor Emeritus, University of Novi Sad, Serbia. © 2021.

different digital libraries evaluation criteria by heterogeneous stakeholders in academic settings.[3] Specifically, they surveyed three groups of stakeholders (scholars, librarians, and digital library users), and through their analysis of the survey's responses, they identified differences in opinions not only between user expectations and the digital library practice but also between what is desirable and what is possible in the academic environment.

Regardless of whether more general (i.e., web search engines) or more specific (i.e., digital libraries) systems are used, the presentation of search results to users is important. It significantly affects how they perceive the system and may reduce or increase the chances of using the system and the frequency of use, as usability is an important aspect in any system. The visualization of search results may be adapted to user needs. Presenting results either in a textual or an alternative format depends on how users respond to the alternative presentations of the system and this process is followed in different domains, such as recommender systems.[4] This can form part of a personalized context-aware system that considers users' environment, history, and interaction with the system in order to act proactively and adapt the presentation of search results to each user.

Based on the gaps identified above, the focus for this study is the improvement of the user experience with the PHD UNS digital library through the personalization of the offered service for system users. It is necessary to provide the user with a choice between textual and visual search results representation according to user preferences and needs. This study reveals a new way of presenting search results that has been analyzed, designed, and implemented by the authors. More concretely, presentations of the bibliographic metadata in a standardized citation style (Harvard style) and bibliographic formats (MARC 21, Dublin Core, ETD-MS) have been analyzed and implemented. Word cloud format is widely used in different systems and this representation has been implemented in the PHD UNS system. Finally, it is necessary to determine if the initial search results representation should be stored in the history of users' queries, device types, and search mode. Users have the ability to provide their feedback on the visualization of the search results, therefore indicating if they prefer a textual or new visual results representation (by changing search results representation style). The feedback received is used to adapt the results representation based on the user preferences. This component represents the first step towards a completely personalized system, in which different contextual parameters will be used for providing a personalized context-based user experience. At this point, user feedback is used for personalization, search results representation, and subsequent system use. A preliminary version with preliminary results regarding the word cloud component is described by Kapitsaki and Ivanovic.[5] In respect to this previous work, we are presenting the evolvement of personalization in the PHD UNS system and a more thorough evaluation that allows us to perform statistical analysis and draw more generic conclusions.

Accordingly, the motivation for this research is the personalization of the search results representation of a digital library, and the research questions to which this research should provide answers have been identified. We are discussing our results based on these questions:

1. **RQ1:** What are the users' profiles of the PHD UNS digital library?
2. **RQ2:** How could search results best be presented to different users within PHD UNS's digital library collections?
3. **RQ3:** Can the search results representation with PHD UNS's digital library depend on the history of users' queries, device types, and search mode?

RELATED WORK

DOSIRD UNS

The DOSIRD UNS project (http://dosird.uns.ac.rs/) was launched in 2009 with the aim to develop software infrastructure for the research domain of the University of Novi Sad (UNS). The CRIS UNS system (www.cris.uns.ac.rs) is the first result of this project. This system represents the information system of the research domain of the UNS. The development of the system started with the beginning of the project in 2009 and is still active. The digital library of theses and dissertations (PHD UNS), which is the topic of this paper, is integrated within CRIS UNS. The complete CRIS UNS system was developed in accordance with the recommendations of the euroCRIS (www.eurocris.org) non-profit organization.

Systems which contain the published scientific results were analyzed and, on the basis of these analyses, a set of metadata describing the scientific and research result in CRIS UNS was created.[6] A paper by Ivanović et al. described the CERIF compatible data model based on the MARC 21 format which maps part of the CERIF data model to the MARC 21 format data model.[7] The MARC 21 format is a standardized format for storing bibliographic data. CRIS UNS has been built on this model. The system architecture and implementation are described in previous publications.[8]

The development of the Digital Ph.D. Dissertations Library (PHD UNS) began in 2010. In December 2012, the Senate of the University of Novi Sad approved the commissioning of a public service for the search of a digital library of dissertations defended at the University (https://cris.uns.ac.rs/searchDissertations.jsf). PHD UNS has been implemented with the following characteristics:

- The digital library of e-theses is integrated into the information system of the scientific research activity of the University of Novi Sad (CRIS UNS).
- The digital library is CERIF compatible, that is, it can exchange metadata with CERIF-compatible systems of scientific research activity.
- E-theses are described by a set of metadata which includes all the metadata prescribed by the Dublin Core and the ETD-MS metadata format, that is, the system can exchange the data in Dublin Core or ETD-MS format via the OAI-PMH protocol.
- The digital e-thesis library has a data model and architecture that can be easily integrated with a bibliographic system based on the MARC 21 bibliographic format.
- The user interface allows a user to enter the thesis and dissertation data without knowing the standardized metadata formats on which the digital library is built.

The integration of PHD UNS within CRIS UNS involved the following four steps:

1. The CRIS UNS data model has been extended with entities and properties for describing PhD theses in accordance with CERIF, Dublin Core, and ETD-MS data models.[9]
2. The CRIS UNS software architecture and user interface has been extended in order to support basic functionality of cataloguing theses.[10]
3. Theses' metadata have been imported from the previous source.[11]
4. The web page for searching among the collection of theses has been implemented.[12]

Searching Personalization

The findings and analysis of scientific results described in papers, theses, and dissertations is an important part of research activities in the scientific community. Therefore, the use and development of the tools and the bibliographic systems which enable advanced search is becoming increasingly more common.

The personalization of search results can include automatic recommendations to users.[13] Moreover, part of search personalization refers to the personalization of results representation. Similar to popular web search engines like Google, the way the search results are represented is very important to users. The way the results are represented can affect user perception of the system and the frequency of its use. The results can be presented to users in formats other than textual in order to improve the user experience in search tools, as well as to improve access to finding information and in recommendation systems.[14]

Ferran and colleagues described browsing and searching personalization systems for digital libraries.[15] Their approach is based on the use of ontologies for describing the relationships between elements that determine the functionalities of the desired personalization system. Those elements include the user's profile, including navigational history and the user preferences, as well as the information collected from the navigational behavior of the digital library users. Such a personalization system can improve digital library users' experience.

Sebrechts and colleagues presented 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.[16] The conducted experiments included 15 participants. The study revealed that although 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 (see the Visual Results Representation section later in this article), but instead of focusing on basic text information we have adopted newer approaches found in word clouds. Bowers and Card analyzed visualization in the framework of database search.[17]

Soliman et al. presented an approach for the clustering of search engine results that relies on the semantics of the retrieved documents.[18] The approach takes into consideration both lexical and semantic similarities among documents and applies activation spreading technique, in order to generate clusters based on semantic properties. Nguyen and Zhang proposed a model for web search visualization, 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.[19]

A word or tag cloud is a visual representation of word content commonly used to represent content in different environments.[20] Several past works have introduced various algorithms for the tag selection or new ways for the word cloud creation.[21]

Tag clouds have been used in PubCloud for the summarization of results from queries over the PubMed database of biomedical literature.[22] 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.

Context Awareness

Context awareness is a part of many systems in various domains, where the application or system functionality adapts to the context of use, such as in mobile computing or pervasive computing applications.[23] The first definition of context was given by Abowd and colleagues, where they defined context "as any information that is relevant to the user, to the system and to any interaction between the user and the system."[24] Applications utilize context data in order to provide context-aware services to users. Context information, such as user location or user preferences, are used to adapt the application functionality or presentation to a specific user.

Mobile computing and pervasive computing offer the necessary information from mobile device sensors and in users' environments for context-aware application provision.[25] Fink and Kobsa claimed that personalization may adapt various features in order to address the specific needs of each individual.[26] Many systems utilize users' search history in order to offer personalized search in the framework of the web information retrieval systems. Yoganarasimhan found that personalization based on short-term history or within-session behavior is less valuable than long-term or across-session personalization.[27] Behnert and Lewandowski analyzed the application of web search engines ranking approaches on digital libraries, and they argued for a user-centric view on ranking, taking into account that ranking should be for the benefit of the user, and user preferences may vary across different contexts.[28]

Frias-Martinez and colleagues defined an approach to constructing personalized digital libraries. Adaptive digital libraries automatically learn user preferences and goals and personalize their interaction using this information.[29] Based on previous work, Frias-Martinez and colleagues developed a personalized digital library to suit the needs of different cognitive styles.[30]

Contribution of Our Work

We share similarities with previous work in terms of techniques used, as for instance the Harvard reference style, standardized bibliographic formats (MARC 21, Dublin Core, ETD-MS) and word clouds, which have been used in other systems for representation of search results in order to improve the user experience. However, in contrast to previous work, we apply techniques in a specific context for a Serbian digital library, allowing automatic adaptation for the representation of search results based on a user's type, history, and reaction. Those search results representations are implemented and integrated within a real system (PHD UNS) and are tested with real users' feedback. That users' feedback is analyzed using ELK stack technologies, making our main conclusions useful for similar systems and future research on personalization in digital libraries.

METHODOLOGY

The main requirements for implementation of the PHD UNS digital library were for the system to be compatible for integration with other systems of scientific research activity, support for data exchange in different standardized formats, and provision of representation of the results to users of different categories and profiles (researchers, scientists, librarians, users from outside the academic community, etc.). For these reasons, existing formats for representation of references, bibliographic metadata formats, as well as techniques for visual representation of textual publications, were analyzed. The format adopted for the representation of references was Harvard-style implemented with the FreeMarker template. FreeMarker
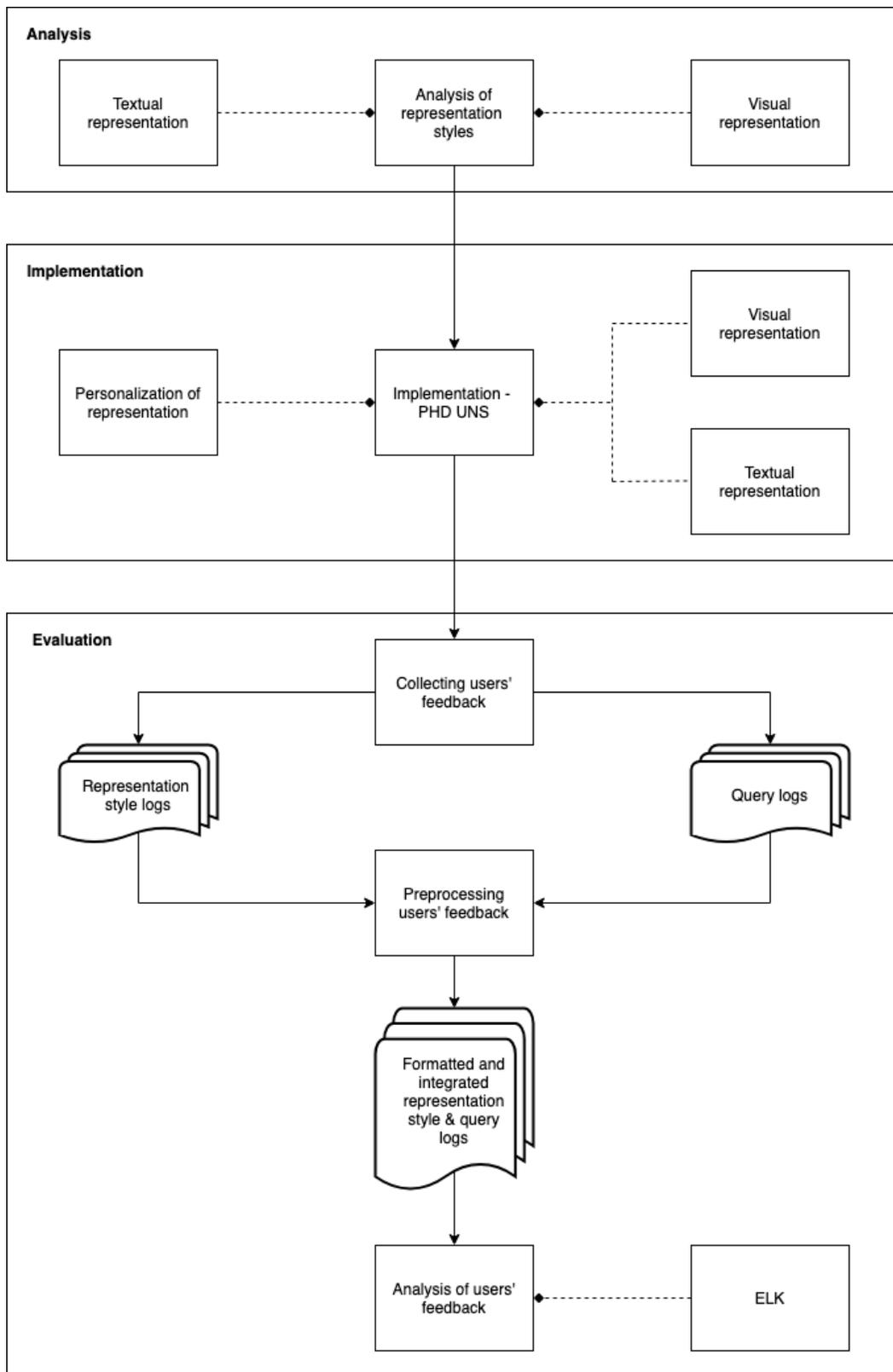
(https://freemarker.apache.org/) is an open-source template engine for Java that assists in separating the web user interface from the main system functionality, following the MVC (model view controller) pattern. The analysis established that it was necessary to implement the search results representation in three bibliographic formats: MARC 21, Dublin Core, and ETD-MS. For each of these formats, appropriate mappers/serializers were made which transform the data from a database into the XML (eXtensible Markup Language) representation of the previously mentioned bibliographic formats. For visual representation, word cloud style was adopted. A component for generating word cloud images was implemented and integrated into the PHD UNS digital library.

When presented with search results, users are able to choose which desired representation format is stored and used as preferable for the given user in future search results representations. A logging system was implemented to assist with this; this module is invoked when the user changes the default data view mode. Received messages are preprocessed for the purpose of analyzing messages and obtaining a more accurate evaluation. The aim of the analysis of the messages on the work of the PHD UNS system is to obtain the desired statistics:

- Distribution of used representation styles
- Top queries executed before changing into the textual style, as well as into the visual style
- Distribution of devices used before changing into the textual style, as well as into the visual style
- Distribution of search modes before changing into the textual style, as well as into the visual style

These statistics are analyzed to determine the user behavior patterns depending on the type of search (basic or advanced), the search device used, the executed query, etc. Based on the established patterns it is possible to determine the representation style for future searches of the new users. The results of the analysis are graphically represented using ELK stack technologies and are presented below in the Evaluation section. The methodological approach is shown in figure 1.

The rest of this paper is organized in accordance with the methodology steps shown in figure 1.

Analysis

Textual representation → Analysis of representation styles ← Visual representation

Implementation

Personalization of representation → Implementation - PHD UNS → Visual representation / Textual representation

Evaluation

Collecting users' feedback → Representation style logs / Query logs → Preprocessing users' feedback → Formatted and integrated representation style & query logs → Analysis of users' feedback ← ELK

Figure 1. The methodology of the study presented in this paper.

ANALYSIS OF SEARCH RESULTS' REPRESENTATION STYLES

Textual Representation

For the needs of the PHD UNS library, three types of search results' textual representation for three types of PHD UNS users were analyzed:

Reference representation. A citation style is defined as a set of rules for citing sources in academic writing and those rules prescribe style for in-text citations, as well as style for reference representation in the references' list. This textual approach is intended for users from the academic community—researchers, teaching staff of universities, and PhD students. Since the majority of PHD UNS users belong to this group, this is the default representation in a textual representation of the search results. This group of users is familiar with this type of representation. From this type of representation, the users can easily recognize basic data of interest and can use this representation for citing and referencing the dissertations which have been retrieved as a result of executing a query.

Taking into account that there are currently many different styles for representing references and citations (e.g., APA, MLA, Harvard, Vancouver, and Chicago, among others), and that they might change in the future with the emergence of new trends in science (for example, the emergence of open science and the need to cite data sets, not just publications), it is necessary to create a scalable component for representing the results in a form of a reference. Based on this analysis, we decided that the architecture of this component will be based on FreeMarker, which makes the introduction of new templates for the output format easier, and that the first FreeMarker template should be created for Harvard style.

Structured representation. This textual approach is intended for users outside the academic community who want to search the digital library. This type of representation represents only the data from the digital library database in a legible format that is represented in the web browser.

Bibliographic formats representation. This textual approach is intended for librarians who are in charge of entering and maintaining data in the digital library. In addition to one central library of the University of Novi Sad, there are libraries in every department within the University. Most of these libraries use the BISIS library system, which is based on MARC format. Therefore, it can be concluded that the majority of the librarians who enter data into the PHD UNS library are familiar with the MARC format. Librarians can use the representation of metadata about dissertations in the MARC 21 format to check if all of the information about a dissertation is entered correctly.

- *MARC 21* bibliographic format supports not only descriptions of theses and dissertations but also other published scientific results, such as a paper published in a journal, a monograph, a paper published in conference proceedings, etc. There are several examples where theses and dissertations are described using the MARC 21 format in the bibliographic information systems of some universities.[31]
- *Dublin Core* (http://dublincore.org) is the most commonly used format for data exchange between different information systems, and data are exported in this format via the OAI-PMH protocol from the PHD UNS system into a network of digital libraries, such as DARTEurope, OATD, and NaRDuS. The Dublin Core XML schema is available online at www.openarchives.org/OAI/2.0/oai_dc.xsd. The representation in Dublin Core format can

be used by librarians to check if the metadata will be correctly exported to the previously mentioned aggregation systems.

- ***Electronic Theses and Dissertations Metadata Standard (ETD-MS)*** (www.ndltd.org/standards/ metadata) is an extension of the Dublin Core format with new features/properties. The standard defines a set of metadata that is used to describe a master's thesis or a doctoral dissertation. The metadata of this standard describe the author, his/her paper, and the context in which this paper has been created in a way that will be useful not only to the researcher, but also to the librarian and/or the technical staff in charge of maintaining the paper in electronic form. This format is used within the NDLTD worldwide network of digital theses and dissertations and in this format the data is exported via the OAI-PMH protocol from the PHD UNS system to this network. The XML schema of the ETD-MS format is available online at www.ndltd.org/standards/metadata/etdms/1.0/etdms.xsd. The representation in the ETD-MS format can be used by librarians to check if the metadata will be correctly exported to the NDLTD network (http://union.ndltd.org/).

Visual Representation

A word cloud is a visual representation of textual content, with the importance of each word indicated with a different font size and/or color. Word clouds are often used in many digital libraries to represent textual content.[32] As previously written, the word cloud is used in different environments and is a popular way to represent web results by summarizing the content of documents and other sources of information. We adopted a word cloud approach for visual representation of the user search results in the PHD UNS library.

Various tools for generating word clouds are available, such as the tool offered by Jin.[33] Based on the characteristics of available tools, we decided to implement the Kumo library available in the Java programming language (https://github.com/kennycason/kumo) that allows easier integration within the PHD UNS digital library.

IMPLEMENTATION DETAILS

This section presents the implementation of the textual and visual search results representation, as well as the implementation of the search results personalization.

Textual Results Representation

Based on the analysis presented in the previous section, we decided to implement the following functionality in order to enhance the PHD UNS digital library: a structured representation for users outside the academic community, a representation in the form of references for scholars, and a representation of bibliographic and library formats for librarians in charge of data in the PHD UNS digital library.

Reference representation. Figure 2 depicts the architecture of the module for generating PhD dissertations' representations in the form of references for scholars.
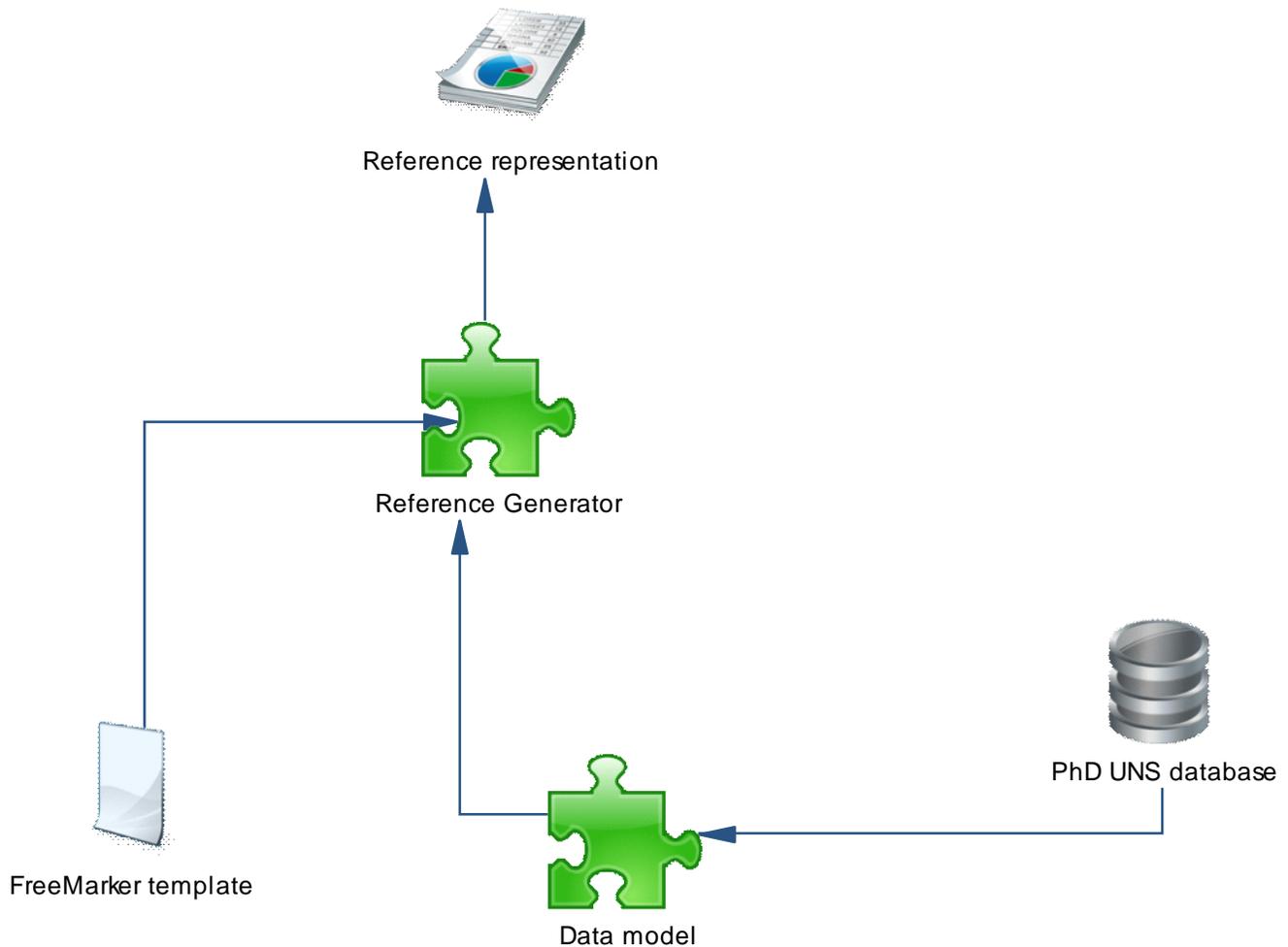
Figure 2. Architecture of module for generating reference representation.

The model of the reference generator component is shown as the class diagram in figure 3. This component can be used to generate textual representation to all publications from the data model component (figure 4) in the chosen reference style (FreeMarker template—see listing 1). The central class is *TemplateRunner* which includes the necessary operations to generate reports. The *TemplateHolder* represents the template container and has operations for adding new templates and selecting template for generating report. The *Template* class is the model of the template for one reference style and one publication type.

The component architecture described in the class diagram of figure 3 is independent of the number of templates, whereas adding a new template to the component requires creating a new instance of the *Template* class. As similarly performed in the CRIS UNS system, the implementation of these instances of the *Template* class is done in FreeMarker that does not require the recompilation of the source code.
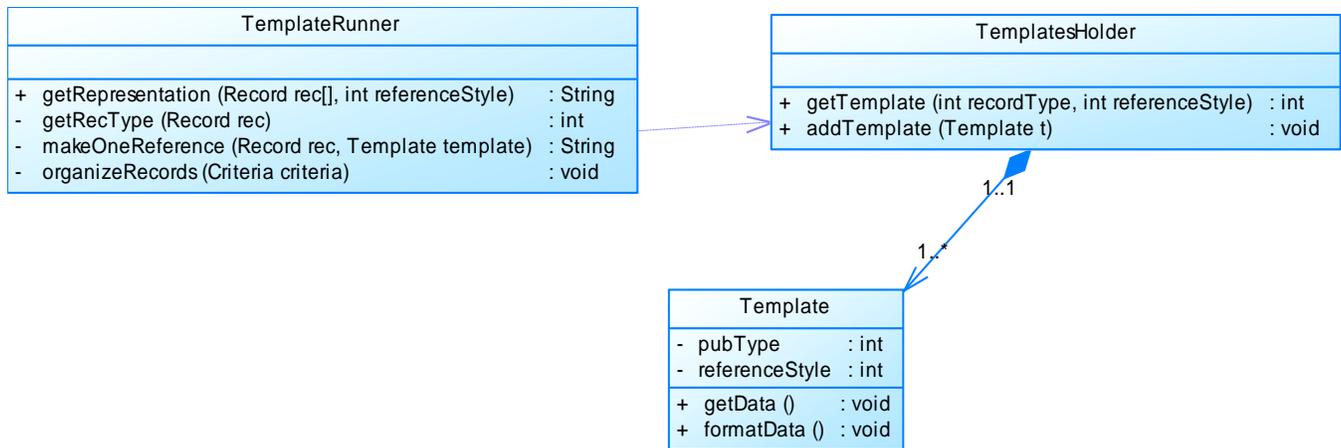
Figure 3. Architecture of the component for generating template.

```
<#macro nameInitial name>
<@compress>
<#t><#if name?length &gt;1>
<#t><#if (name?upper_case?starts_with("LJ") ||
    name?upper_case?starts_with("NJ"))>, ${name?substring(0,2)?upper_case}.
<#t><#else>, ${name?substring(0,1)?upper_case}.</#if>
<#t><#elseif name?length=1>, ${name?upper_case}.</#if>
</@compress>
</#macro>

<#t>${author.name.lastname?upper_case}
<#t><@nameInitial author.name.firstname/>
<#t><#if publicationYear??> (${publicationYear})</#if>
<i>${someTitle!""}</i>. (${localizedStudyType}), ${institution.someName}
```

Listing 1. Harvard-style FreeMarker template

Structured representation. The simplified version of the bibliographic records data model that is used in the CRIS UNS system is shown in figure 4. The CRIS UNS system contains also other publication entities, such as monograph, journal paper, etc. The PHD UNS digital library is integrated into the CRIS UNS system and uses the entities shown in figure 4.
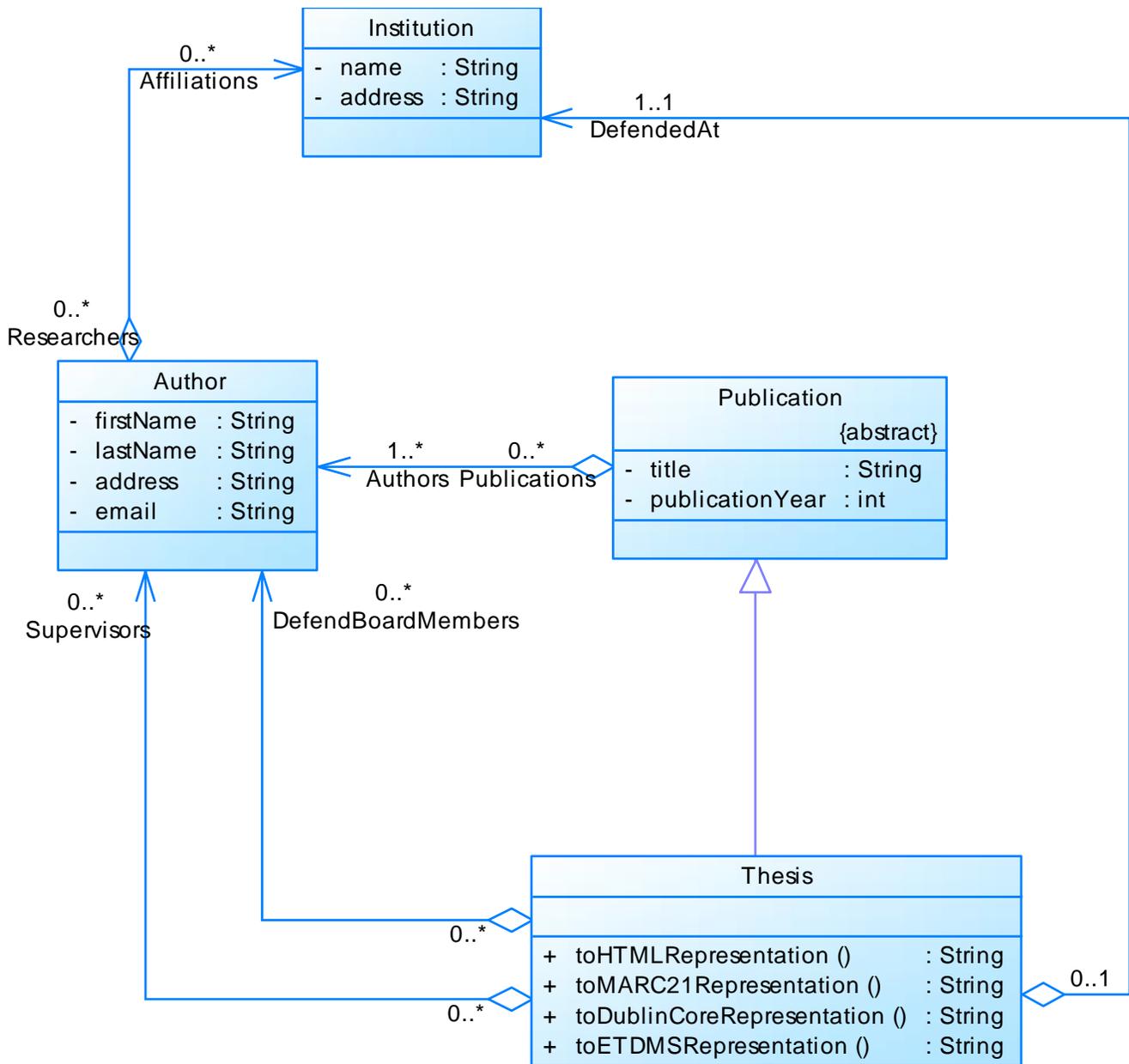
Figure 4. Data model.

A structured representation for users outside the academic community is implemented with the help of the *toHTMLRepresentation* method contained in the *Thesis* class. This method forms a structured representation of the dissertation with the help of HTML markup. When storing a dissertation in a database, HTML representation is generated and stored in Lucene indexes for faster representation of search results.

Bibliographic formats representation. After analyzing the bibliographic and library formats (see the Methodology section above) we concluded that we should implement the search results representation in MARC21, Dublin Core, and ETD-MS bibliographic formats for the needs of librarians who are in charge of entering data in the PHD UNS library. The representation of these formats is implemented in a similar way as the representation of structured data outside the academic community, with the help of the following methods: the *toMARC21Representation*,

toDublinCoreRepresentation, and *toETDMSRepresentation* in the *Thesis* class (figure 4). These methods generate an XML representation of these formats. When storing a dissertation in a database, these XML representations are stored in Lucene indexes for faster retrieval of search results.

Screenshots of user interface. Figure 5 presents the textual search results representation. The basic representation contains the metadata of the dissertation presented as a Harvard-style reference. This is the basic representation because the researchers from the academic environment are the most common users of the PHD UNS library. Additional metadata are represented by pressing the button 🛈 which is located next to the reference and represents the data structured for the needs of users outside the academic community (see fig. 6).

Figure 5. Results representation in a textual format.

In addition, the representation of the dissertation metadata is also available to library users in MARC 21, Dublin Core, and ETD MS formats (see fig. 6).

Figure 6. Structured and bibliographic formats representation.

Visual Results Representation

This section describes the implementation of the graphical (visual) search results representation. The graphical representation is realized using a word cloud to represent the content of a dissertation.

Word cloud generator component. The word cloud generator component forms a new part of the PHD UNS digital library. The aim of this new component is to present the user search results in a word cloud representation (see fig. 7).
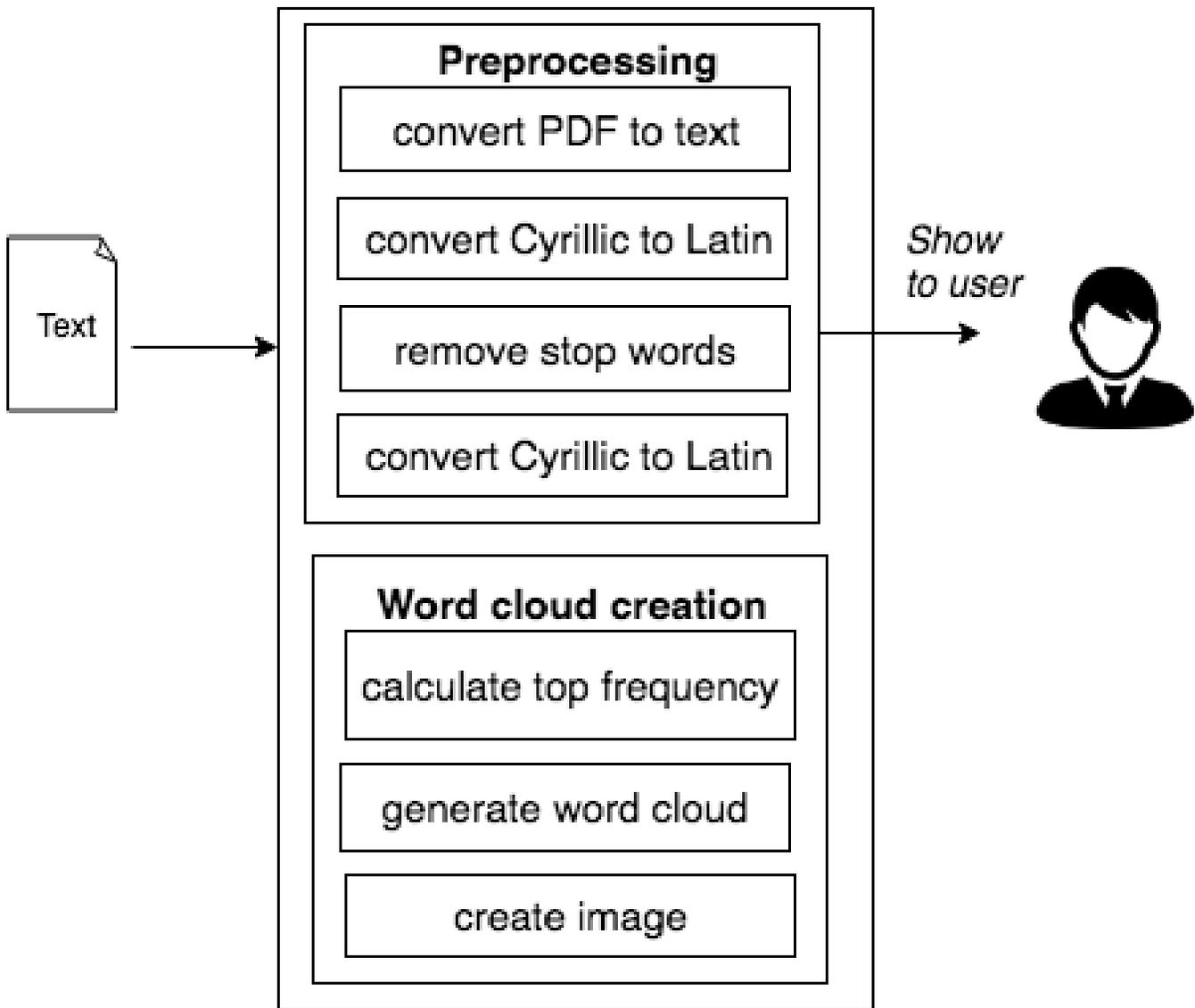
Figure 7. Word cloud generator steps.

The word cloud component was implemented in Java. The component accepts a PDF file as input and generates an image (PNG file) as output. The tool uses as input the PDF file of the dissertation; it then parses the textual content of the file and performs a preprocessing of the text. The result of the preprocessing is a list of pairs containing the original version of each word from the text and its stem. The details of the tool utilized for this preprocessing step can also be found in the existing publication.[34] The tool then calculates the top frequencies of words in the text, generates the word cloud, and creates an image file.

As aforementioned, for implementation purposes, the Kumo library was used (https://github.com/kennycason/kumo). Kumo is an open source software that carries the MIT license. The source code has been extended to accommodate the needs of the PHD UNS digital library.

Integration to the PHD UNS system. The word cloud generator component described in the previous section has been integrated into the PHD UNS digital library application and was put into operation in April 2017, although some necessary adaptations have been performed since then and have all been integrated as well. Taking into account that the word cloud generator is lengthy and creates a computationally intensive process, it is invoked in the indexing phase and the generated image is stored as supplementary material to a PhD dissertation in the server file system. Figure 8 presents a Unified Modeling Language (UML) activity diagram which describes the process of adding a new dissertation to the PHD UNS digital library. The activity "Generate word cloud image" is highlighted in red and represents invoking the execution of the word cloud component. Moreover, the activity "Create Lucene index" includes the same steps for text preprocessing as the steps described in the word cloud generator component (see fig. 7).
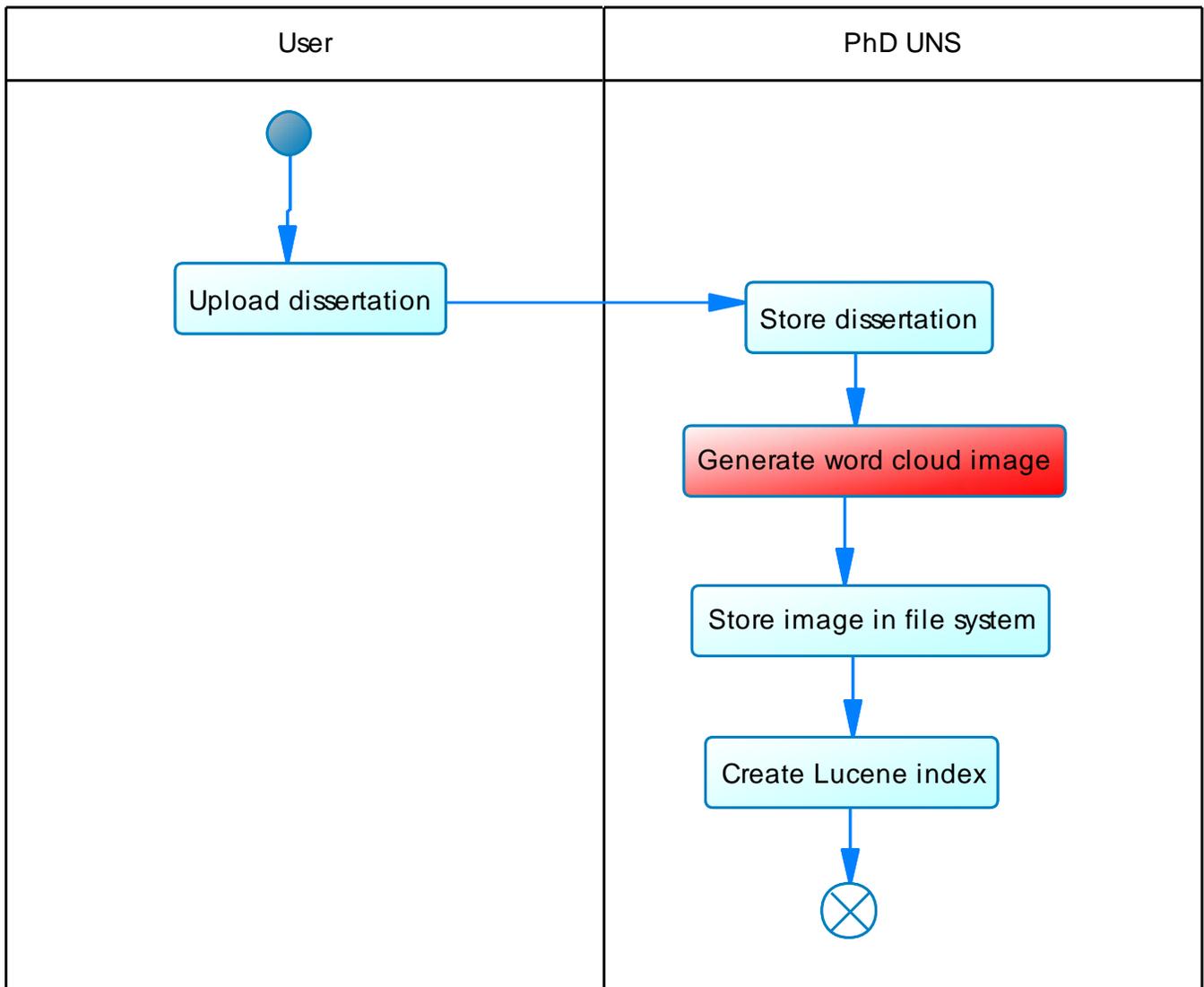
Figure 8. Adding a new dissertation into the PHD UNS system.

The search results representation in the form of a word cloud is enabled via the user interface page for representing the search results of the PHD UNS digital library (see fig. 9).

Figure 9. Results of the search of the PHD UNS system in a word cloud format.

Personalization of Representation

This section describes the implementation related to personalization of the search results representation. The user can select the desired style of representation, and the representation history of the results is recorded in order to personalize the results representation and customize the user's profile and information needs.

The initial search results representation style for users who search for dissertations in the PHD UNS system for the very first time is the random selection of one of the two options:

- result representation in a textual format
- result representation as word cloud images

After analyzing the logs from changing the results representation (see next section), this random selection could be replaced with a choice that depends on the context: queries, devices, types of searches, etc.

The parts of the page which represent how the results are presented in the textual and word cloud representations are shown in figures 5 and 9, respectively. Users can change the representation style from the page. In this way, users give feedback and indicate their preference for visualization of the results which is used in the future results representations for that user.

EVALUATION

Collecting User Feedback

If a digital library user changes the style of results representation, the relevant message about the change of the representation style together with the user metadata is recorded using the log4j. This process is shown in red in the activity diagram in figure 10.
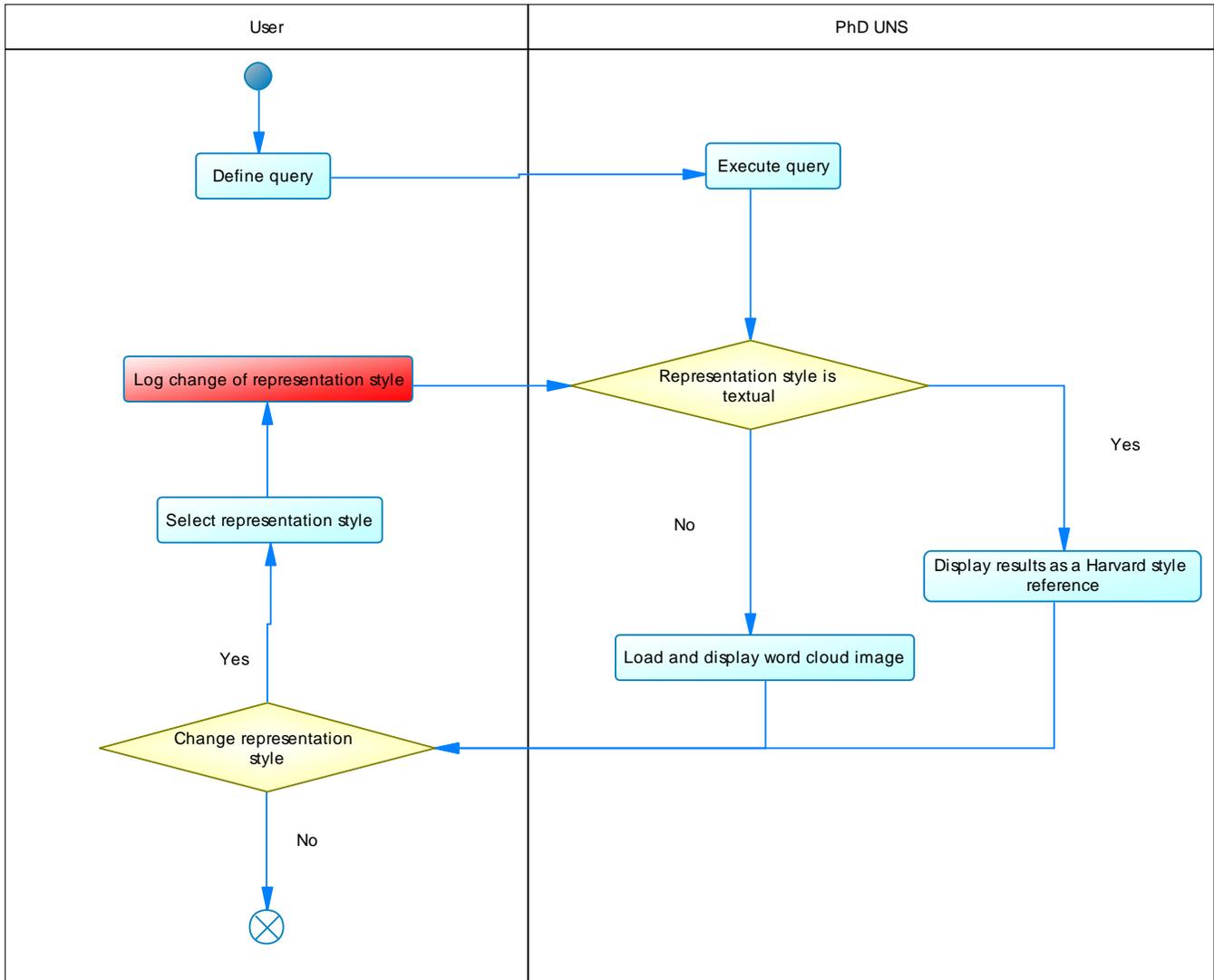
Figure 10. The process of executing queries and giving feedback on the representation style.

Listing 2 is an example of a recorded message about the change of the representation style containing user metadata from the PHD UNS system. Information, such as the time and territorial determinant of a web client, the agent used, and the representation style, are also recorded. The representation style is stored on the user browser in the form of cookies and represents the basic style for representing results in future searches of dissertations in the PHD UNS system. By analyzing the messages about the change of the representation style, we evaluate the results of our approach and examine how the users respond to the new style of representation.

```
[INFO] 22.08.2017. 16:07:33 (SearchDissertationsManagedBean:setRepresentationStyle) Date and
    time: Tue Aug 22 16:07:33 CEST 2017| miliseconds: 1503410853455| + session id:
    2A4CE66932D0C3C8DB97098DFF956074| userId: 150341083728649| ip address: 188.2.29.239|
    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 10.0; WOW64) AppleWebKit/537.36 (KHTML, like Gecko)
    Chrome/60.0.3112.78 Safari/537.36 OPR/47.0.2631.55| new representation style: wordCloud
```

Listing 2. Example of a message about the change of the representation style

Preprocessing Users' Feedback

As already indicated, each change in the representation style of the search results causes the creation of an appropriate message (see listing 2). In order to better understand the context of use and the reason for changing the representation style, these messages are preprocessed and supplemented with information on the type of search and the given query which preceded the change in the representation style (highlighted in yellow in listing 3). By analyzing additional information, we can understand which context of usage and user actions preceded the change of the representation style. Additional information is obtained from the received queries for the PHD UNS system and is mapped by using a unique user session identifier. An example of a message after preprocessing is shown in listing 3.

```
[INFO] 22.08.2017. 16:07:33 (SearchDissertationsManagedBean:setRepresentationStyle) Date and
    time: Tue Aug 22 16:07:33 CEST 2017| miliseconds: 1503410853455| + session id:
    2A4CE66932D0C3C8DB97098DFF956074| userId: 150341083728649| ip address: 188.2.29.239|
    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 10.0; WOW64) AppleWebKit/537.36 (KHTML, like Gecko)
    Chrome/60.0.3112.78 Safari/537.36 OPR/47.0.2631.55| new representation style: wordCloud|
    query: internet| searching mode: basic
```

Listing 3. Example of a message about the change of the representation style after preprocessing

Analysis of User Feedback

Messages such as the one in listing 3 with additional information about contextual use are suitable for further analysis using ELK stack technologies. Messages in a given format are collected from logs of the PHD UNS system using the Logstash Grok filter. This filter is used for parsing, statistical analysis based on field values, data filtering, and advanced search using multiple filters. The parsed messages have been forwarded to Elasticsearch components of ELK stack technology. The Grok pattern definition, which represents the rules and instructions for parsing messages, is located in the configuration files that are forwarded as a parameter when running the tool. An example of a configuration file is shown in listing 4.

```
input {
  file {
    path => "/config-dir/logs-style-formatted/*.log"
    start_position => beginning
  }
}
filter {
  grok {
    break_on_match => false
    match => { "message" => "%{LOGLEVEL:logLevel}" }
    match => { "message" => "Date and
    time: %{DAY:logDay} %{MONTH:logMonth}   %{MONTHDAY:logMonthday} %{NUMBER:logHour}:%{NUMBE
    R:logMinute}:%{NUMBER:logSecond} %{WORD:logTimezone} %{YEAR:logYear}" }
```

```
   match => { "message" => "userId: %{NUMBER:userId}\|" }
   match => { "message" => "city: %{DATA:city}," }
   match => { "message" => "countryName: %{DATA:country} \(" }
   match => { "message" => "user agent \(device\): %{DATA:userAgent}\|" }
   match => { "message" => "new representation style: %{DATA:newStyle}\|" }
   match => { "message" => "query: %{DATA:query}\|" }
   match => { "message" => "searching mode: %{DATA:searchingMode}\|" }
 }
}
output {
  elasticsearch { hosts => ["elasticsearch:9200"] }
}
```

Listing 4. An example of a Grok pattern used to analyze the message about the change of the representation

The analysis of the messages about the work of the PHD UNS system is presented in this section. The results are represented using the Kibana graph component of the ELK stack. This component is used for visualization and data exploration, analysis of logs at specified time intervals, and real-time monitoring of applications.

The word cloud generating component was put into operation in April 2017. Log messages were analyzed from then until the end of 2019. In total, there were 17,474 analyzed messages about changing the style of search results representation. In these messages, the style was changed into a textual representation 16,032 times, while it was changed into a visual representation style in the form of a word cloud image 1,442 times. Thus, most of the users of the PHD UNS system changed the representation style to textual rather than visual format. This tells us that the majority of users are more familiar with the textual style of representing search results in interaction with scientific systems. Based on this analysis, it can be concluded that the random selection of the representation style of the results is not a good choice. We also analyzed the client devices used when changing representation style (textual and visual). Computers were used considerably more frequently than mobile devices. Devices with larger resolution screens are more suitable for presenting search results in different formats. Distribution of the change in the representation style of the search results is similar for computers and mobile devices, and based on the device, we cannot conclude which representation style is more suitable for the user.

The queries that were submitted by users before changing the style of representation were also analyzed; in other words, which of the queries and results representations initiated the change into the other style of representation. Figure 11 and figure 12 show the most commonly executed queries before changing the style of representation into the textual and visual format, respectively.

Figure 11 shows the most commonly executed queries before changing into the textual style of representation. Some of the queries shown on this figure represent the names of faculties of the University, such as

- fakultet tehnick nauka (Department of Technical Sciences)
- filosofsk fakultet (Department of Philosophy)

Figure 11. Top queries before changing into the textual style

The most commonly executed queries before changing into the visual representation style are shown in figure 12. Some queries shown on this figure represent scientific fields, such as

- doktor medicinskih nauka (doctor of medical science)
- doktor geografskih nauka (doctor of geographical science)

Figure 12. Top queries before changing into the visual representation style.

Based on figures 11 and 12, we can conclude that the queries users submitted before the change in the style of representing the results are of a general type, that is, they represent the queries in faculties or by scientific fields. These types of queries give long lists of results. For queries over longer periods of time where the representation of all dissertations defended in a certain period is required, users changed the representation style into visual.

The search mode used before the change into the textual style is shown in figure 13, while the search mode used before the change into the visual representation style is shown in figure 14.
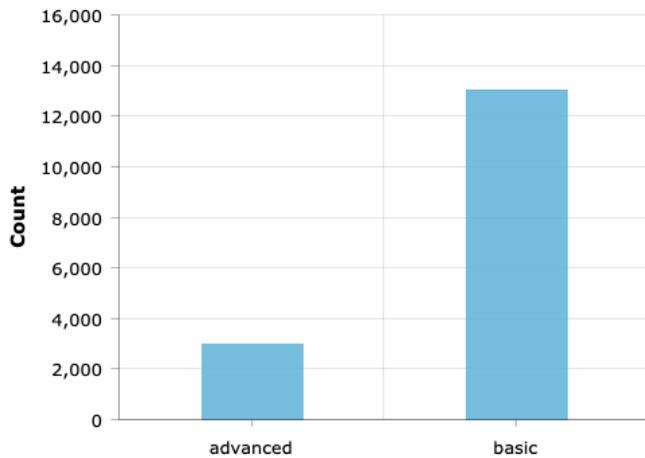
Figure 13. Search mode before changing into the textual style.
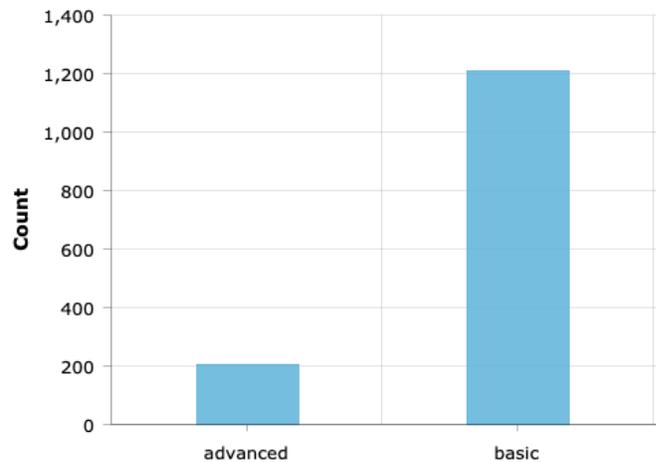
Figure 14. Search mode before changing into the visual style.

By analyzing figures 13 and 14, we conclude that most queries preceding the change of the representation style are set from the basic search mode (labeled *basic* on the figures), which is the default search mode. Also, we notice that there is an increase in the percentage when changing into the visual style of the advanced search mode as compared to the basic search mode. This is in compliance with the analysis following figures 11 and 12, because in the advanced search mode we make queries for a time range that gives long lists of results.

Also, we notice that some users have changed the style of results representation several times, so it is assumed that different types of information require a different representation style. There has been no reduction or enlargement in the number of users since the introduction of the word cloud generating component, which indicates that the introduction of the new component has not affected the frequency of the system use significantly.

CONCLUSION

This paper describes one improvement on user experience performed for the users of the PHD UNS digital library. This improvement was implemented through the personalization of the search results representation which was put into operation in April 2017. Users of the PHD UNS digital library are using desktop and laptop computers considerably more than mobile devices (**RQ1**). Moreover, besides specific exploratory queries, the users are raising general queries by scientific fields, faculties, or in the time range. The PHD UNS digital library has three user groups: those from the academic community, those from outside the academic community, and librarians in charge of entering the dissertation data. For these three groups of users, the following textual search results representations (**RQ2**) have been selected and implemented: Harvard-style representation of the dissertation in the form of references for users from the academic community; HTML structured results representation for users outside the academic community; and MARC 21, Dublin Core, ETD-MS bibliographic records for the library users. For the visual representation, word cloud presentation based on the complete text from the PDF file of the

dissertation has been selected and implemented. It is possible to select the desired search results representation which initiates storing of the messages about the representation style of the results, client device used, time, etc. This message is joined with a preset query message to analyze the patterns of system usage and establish a correlation between the change of the representation style and the type of query, device, and search mode (**RQ3**).

Based on the conducted analysis, we reached the following conclusions:

- A significantly larger number of users of the PHD UNS system use the textual representation style rather than the visual representation. This tells us that a larger number of users is more familiar with the textual style of representing search results in interaction with scientific systems and that a random selection of the representation style of the results used since April 2017 was not a good choice for the first-time user. Because of this observation, the initial selection of the representation style for the first-time user was changed to the textual search results representation (**RQ3**).
- Some users changed the representation style of the results several times and it is assumed that different types of information require a different representation style. Based on this, we can conclude that the possibility of personalizing the search results representation is a useful functionality that contributes to the improvement of the PHD UNS system and the user experience.
- It has been established that the most frequent change of the visual results representation is after a query which shows all the dissertations from a certain time period taken from the advanced search mode, but there is no correlation between this change and the device being used. Based on this, it can be concluded that in certain cases for queries which show long lists of results, it is more transparent to see the results in the visual mode (**RQ3**). It is necessary to collect more data and carry out additional analysis, in order to be able to precisely establish the correlation or to precisely determine for which queries and for which types of users this applied to, so that the system could automatically change the style of representation in certain cases.

Directions for future research and application development include the following. It is planned to collect and analyze additional messages about the work of the digital library in order to further enhance the user experience. Also, it is necessary to follow the trends of the results representation due to the change of standardized reference styles, bibliographic formats, technologies and hardware devices, and it is further necessary to coordinate the results representation with these trends. Differences between the behavior of the different user groups will also be examined further.

ENDNOTES

[1] J. Brophy and D. Bawden, "Is Google Enough? Comparison of an Internet Search Engine with Academic Library Resources," *Aslib Proceedings* 57, no. 6 (2005): 498–512, https://doi.org/10.1108/00012530510634235.

[2] A. F. Smeaton and J. Callan, "Personalisation and Recommender Systems in Digital Libraries," *International Journal on Digital Libraries* 5, no. 4 (2005): 299–308, https://doi.org/10.1007/s00799-004-0100-1.

[3] Iris Xie, Soohyung Joo, and Krystyna K. Matusiak, "Multifaceted Evaluation Criteria of Digital Libraries in Academic Settings: Similarities and Differences from Different Stakeholders," *The Journal of Academic Librarianship* 44, no. 6 (2018): 854–63, https://doi.org/10.1016/j.acalib.2018.09.002.

[4] Theodora Nanou, George Lekakos, and Konstantinos Fouskas, "The Effects of Recommendations' Presentation on Persuasion and Satisfaction in a Movie Recommender System," *Multimedia Systems* 16, No. 4–5 (August 2010): 219–30, https://doi.org/10.1007/s00530-010-0190-0.

[5] Georgia Kapitsaki and Dragan Ivanović, "Representation with Word Clouds at the PHD UNS Digital Library," *Computer Science & Information Technology* 21 (2017), https://doi.org/10.5121/csit.2017.71102.

[6] Dragan Ivanović, "Software Systems for Increasing Availability of Scientific-research Outputs," *Novi Sad Journal of Mathematics – NS JOM* 42, no. 1 (2012): 37–48.

[7] Dragan Ivanović, Dušan Surla, and Zora Konjović, "CERIF Compatible Data Model Based on MARC 21 Format," *The Electronic Library* 29, no. 1 (2011): 52–70, https://doi.org/10.1108/02640471111111433.

[8] Dragan Ivanović, "A Scientific-research Activities Information System," (PhD thesis, University of Novi Sad, 2010); D. Ivanović, G. Milosavljević, B. Milosavljević, and D. Surla, "A CERIF-compatible Research Management System based on the MARC 21 Format," *Program: Electronic Library and Information Systems* 44, no. 1 (2010): 229–51; Dragan Ivanović and Branko Milosavljević, "Software Architecture of System of Bibliographic Data," in *Proceedings of the XXI Conference on Applied Mathematics PRIM 2009*, 85–94.

[9] Lidija Ivanović, Dragan Ivanović, and Dušan Surla, "A Data Model of Theses and Dissertations Compatible with CERIF, Dublin Core and EDT-MS," *Online Information Review* 36, no. 4: 548–67, https://doi.org/10.1108/14684521211254068.

[10] Lidija Ivanović, Dragan Ivanović and Dušan Surla, "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* 56, no. 2: 104–12, https://doi.org/10.5860/lrts.56n2.104.

[11] Lidija Ivanović and Dušan Surla, "A Software Module for Import of Theses and Dissertations to CRISs," in *Proceedings of the CRIS 2012 Conference,* (Prague, June 6–9, 2012): 313–22.

[12] Lidija Ivanovic, "Search of catalogues of theses and dissertations," *Novi Sad Journal of Mathematics – NS JOM*, 43, no. 1 (2013): 155–65; Lidija Ivanović, Dragan Ivanović, Dušan Surla and Zora Konjović, "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*: 117–22.

[13] Joel Azzopardi, Dragan Ivanović and Georgia Kapitsaki, "Comparison of Collaborative and Content-Based Automatic Recommendation Approaches in a Digital Library of Serbian PhD

Dissertations," in *Proceedings of the International KEYSTONE Conference 2016*: 100–11, https://doi.org/10.1007/978-3-319-53640-8_9.

[14] Dragan Ivanović and Georgia Kapitsaki, "Personalisation of Keyword-based Search on Structured Data Sources," in *Proceedings of the 1st International KEYSTONE Conference (IKC 2015)*.

[15] Núria Ferran, Enric Mor, and Julià Minguillón, "Towards Personalization in Digital Libraries through Ontologies," *Library Management* 26, no. 4/5 (2005): 206–17, https://doi.org/10.1108/01435120510596062.

[16] Mark M. Sebrechts et al., "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)*: 3–10, https://doi.org/10.1145/312624.312634.

[17] Frank H. Bowers and Stuart K. Card, "Method and Apparatus for Visualization of Database Search Results," *U.S. Patent* no. 5,546,529.

[18] Sara Saad Soliman, Maged F. El-Sayed, and Yasser F. Hassan, "Semantic Clustering of Search Engine Results," *The Scientific World Journal* (2015), https://doi.org/10.1155/2015/931258.

[19] Tien Nguyen and Jin Zhang, "A Novel Visualization Model for Web Search Results," *IEEE Transactions on Visualization and Computer Graphics*, 12, no. 5 (2006), https://doi.org/10.1109/TVCG.2006.111.

[20] Daniel Scanfeld, Vanessa Scanfeld, and Elaine L. Larson, "Dissemination of Health Information through Social Networks: Twitter and Antibiotics," *American Journal of Infection Control* 38, no. 3 (2010): 182–88, https://doi.org/10.1016/j.ajic.2009.11.004.

[21] Carmel McNaught and Paul Lam, "Using Wordle as a Supplementary Research Tool," *The Qualitative Report* 15, no. 3 (2010): 630; Weiwei Cui et al., "Context Preserving Dynamic Word Cloud Visualization," in *IEEE Pacific Visualization Symposium (PacificVis)* (2010): 121–28, https://doi.org/10.1109/PACIFICVIS.2010.5429600; Yusef Hassan-Montero and Herrero-Victor Solana, "Improving Tag-clouds as Visual Information Retrieval Iinterfaces," in *Proceedings of the International Conference on Multidisciplinary Information Sciences and Technologies* (2006): 25–28.

[22] Byron Kuo, Thomas Hentrich, Benjamin Good, and Mark Wilkinson, "Tag Clouds for Summarizing Web Search Results," in *Proceedings of the 16th International Conference on World Wide Web (WWW '07)* (2007): 1203–04, https://doi.org/10.1145/1242572.1242766.

[23] Jong-yi Hong, Eui-ho Suh, and Sung-Jin Kim, "Context-aware Systems: A Literature Review and Classification," *Expert Systems with Applications* 36, no. 4 (2009): 8509–22, https://doi.org/10.1016/j.eswa.2008.10.071; Georgia M. Kapitsaki, George N. Prezerakos, Nikolaos D. Tselikas, and Iakovos S. Venieris, "Context-aware Service Engineering: A Survey," *Journal of Systems and Software* 82 No. 8 (2009): 1285–97, https://doi.org/10.1016/j.jss.2009.02.026.

[24] Gregory D. Abowd et al., "Towards a Better Understanding of Context and Context-awareness," in *Handheld and Ubiquitous Computing* (Springer: Berlin/Heidelberg, 1999): 304–07, https://doi.org/10.1007/3-540-48157-5_29.

[25] Mika Raento, Antti Oulasvirta, Renaud Petit and Hannu Toivonen, "ContextPhone: A Prototyping Platform for Context-aware Mobile Applications," *IEEE Pervasive Computing* 4 no. 2 (2005): 51–59, https://doi.org/10.1109/MPRV.2005.29.

[26] Josef Fink and Alfred Kobsa, "A Review and Analysis of Commercial User Modeling Servers for Personalisation on the World Wide Web," *User Modeling and User-Adapted Interaction* 10, no. 2 (2000): 209–49, https://doi.org/10.1023/A:1026597308943.

[27] Hema Yoganarasimhan, "Search Personalization using Machine Learning," *Management Science* 66 no. 3 (2020): 1045–70, https://doi.org/10.1287/mnsc.2018.3255.

[28] Cristiane Behnert and Dirk Lewandowski, "Ranking Search Results in Library Information Systems—Considering Ranking Approaches Adapted from Web Search Engines," *The Journal of Academic Librarianship* 41 no. 6 (2015): 725–35, https://doi.org/10.1016/j.acalib.2015.07.010.

[29] Enrique Frias-Martinez, George Magoulas, Chen Sherry, and Robert Macredie, "Automated User Modeling for Personalised Digital Libraries," *International Journal of Information Management* 26 no. 3 (2006): 234–48, https://doi.org/10.1016/j.ijinfomgt.2006.02.006.

[30] Enrique Frias-Martinez, Chen Sherry, and Liu Xiaohui, "Evaluation of a Personalised Digital Library based on Cognitive Styles: Adaptivity vs. Aadaptability," *International Journal of Information Management* 29 no. 1 (2009): 48–56, https://doi.org/10.1016/j.ijinfomgt.2008.01.012.

[31] Magda El-Sherbini and George Klim, "Metadata and Cataloging Practices," *The Electronic Library* 22 no. 3 (2004): 238–48, https://doi.org/10.1108/02640470410541633; Shawn Averkamp and Joanna Lee, "Repurposing ProQuest Metadata for Batch Ingesting ETDs into an Institutional Repository" (University Libraries Staff Publications, 2009): 38; Sai Deng and Terry Reese, "Customized Mapping and Metadata Transfer from DSpace to OCLC to Improve ETD Workflow," *New Library World*, 2009, 110, no. 5/6 (2009): 249–64, https://doi.org/10.1108/03074800910954271.

[32] Steffen Lohmann, Jürgen Ziegler, and Lena Tetzlaff, "Comparison of Tag Cloud Layouts: Task-related Performance and Visual Exploration," in *Proceedings of the IFIP Conference on Human-Computer Interaction* (2009): 392–404, https://doi.org/10.1007/978-3-642-03655-2_43.

[33] Yuping Jin, "Development of Word Cloud Generator Software based on Python," *Procedia Engineering* 174 (2017): 788–92, https://doi.org/10.1016/j.proeng.2017.01.223.

[34] Joel Azzopardi, Dragan Ivanović and Georgia Kapitsaki, "Comparison of Collaborative and Content-Based Automatic Recommendation Approaches in a Digital Library of Serbian PhD Dissertations," in *Proceedings of the International KEYSTONE Conference 2016*: 100–11, https://doi.org/10.1007/978-3-319-53640-8_9