

FAIRness at University of Novi Sad

- Discoverability of PhD research results for Non-Serbian scientific community -

Introduction

The University of Novi Sad has more than 50,000 students and 5,000 employees and as such it is one of the largest educational and research centres in Central Europe. The University of Novi Sad offers 300 accredited study programmes on all study levels, undergraduate, masters, specialist and doctoral studies, which are carried out at the Faculties and University Centres for Interdisciplinary and Multidisciplinary Studies.

PHD UNS is digital library of PhD dissertations defended at University of Novi Sad. It is a result of DOSIRD UNS project (<http://dosird.uns.ac.rs/>). Moreover, PHD UNS is integrated with CRIS UNS system [1] and it is based on CERIF compatible data model [2, 3].

About 5,000 PhD dissertations have been defended at University of Novi Sad since its foundation, and more than 1,000 in the last four years. More than 95% of those dissertations have been written in the national language (Serbian). Furthermore, each PhD dissertation starts with Key words documentation usually in two languages: Serbian and English. Key words documentation is prescribed by University of Novi Sad board and contains the following metadata: title, author, commission members, institution, abstract, advisor, publication year, publisher, etc. The PHD UNS digital library preserves:

1. Metadata prescribed by the University for all dissertations defended at the University;
2. PDF files representing PhD dissertation for dissertations defended in the last ten years;
3. Open-access license signed by a PhD student for dissertations defended in the last four years.

A web page for searching dissertations has been implemented (<http://www.cris.uns.ac.rs/searchDissertations.jsf>). Dissertations are searchable by metadata in both languages (Serbian and English) and by full-text (extracted from PDF files). Search results can be represented in visual (in the form of words cloud) and textual mode. Textual mode includes representation of results in four formats: Harvard citation style, Dublin Core, MARC 21, ETD-MS. If PDF file is attached to dissertation record and dissertation is published under open-access licence, a link for downloading dissertation is available. Before a download is performed, a web page with information about signed open-access license is shown to the user. A PhD author can select and sign one of six Creative Commons licences. Each download is stored in download logs and is used as an input to automatic recommendation system (reference). Metadata about all dissertations published under open access licence are available for exporting via OAI-PMH protocol to service providers (DART-Europe, OpenAIRE+, NDLTD, etc.).

All previously stated features implemented in PHD UNS improves FAIRness of research information available in PhD dissertations, but there is still a language barrier to take into account in that most

dissertations are written in the Serbian language. Metadata and data should be findable and understandable to humans and machines according to the **FAIR** principles [4]. Authors of scientific publications, including PhD dissertations, usually want to attract interest to their work to enhance their reputation both nationally and internationally. Although metadata about dissertations expressed in two languages (Serbian and English) improves discoverability of Serbian PhD dissertations, non-Serbian researchers still can't read and search by full-text PhD dissertations written in the Serbian language. Although FAIR principles have been implemented at the University of Novi Sad digital library, the topic of this paper is a specific aspect of FAIR principles - making findable research information stored in PHD UNS dissertations to the non-Serbian scientific community.

English scientific community

English has become global language in the economic, scientific and cultural world largely dominated by Anglo-American countries. However, even if English is the dominant language in science, it is certainly not the native language for the majority of scientists and there is much to be discovered in academic material not presented in English. More than 95% of PhD dissertations of the republic of Serbia have been written in the national language (Serbian). However, the Serbian Ministry of Science prescribed a rule that each dissertation before it is defended has to be verified by publishing its main results in at least one paper published in a journal belonging to the Web of Science JCR list. The objective of this is to ensure results published in the Serbian language PhD dissertation are also visible to the global science community through associated journal papers which are usually published in English. The same approach can be adopted for all non-English digital libraries of research publications all over the world.

The full version of this paper would describe an approach as to how the connection between those journal papers and PhD dissertations is established at the University of Novi Sad using a digital library and CRIS system. The approach has been implemented at the University of Novi Sad PHD UNS repository and CRIS UNS system by developing the following features:

1. The connection between a PhD dissertation and journal paper which presents the main results of PhD research can be established by the author of the dissertation or by librarian of the institution where the dissertation has been defended.
2. If a PhD dissertation is a result of local searching of the PHD UNS digital library, a reference to the associated journal paper written in English is provided.
3. If a PhD dissertation has been exported to ETD repositories networks, such as DARTEurope (www.dart-europe.eu/full.php?id=1058428), a link (identifier) to the PhD dissertation stored in PHD UNS opens web page will provide a reference to the associated journal papers written in English (www.cris.uns.ac.rs/record.jsf?recordId=93783&source=DARTEurope&language=en). Connected journal papers can be searched and downloaded using other scientific output repositories: Scopus, Google Scholar, etc.

A similar approach can be adopted for establishing the connection between data sets and PhD dissertations taking into account that the CRIS UNS system can also store data set using the `cfResultProduct` entity of CERIF data model.

Ex-Yugoslav scientific community

Since the beginning of national conflicts and Yugoslav wars (in the 90s), all ex-Yugoslav republics became independent countries. The official language of Yugoslavia was the Serbo-Croatian language. Today, we have similar languages by morphology and vocabulary derived from Serbo-Croatian language: Serbian, Croatian, Montenegrin and Bosnian. As more than 20 million people from the Ex-Yu region have reading skills in all of the languages of the region, multi-lingual search capabilities over texts and metadata written in those languages should be implemented in order to enable scientists to query the system in their native language and find publications in other regional languages that would otherwise be overlooked due to the language differences. There are two approaches for developing ex-Yugoslav language tools which would support those multi-lingual search capabilities. The first one includes development of common preprocessing techniques including stemming and parallel vocabularies. This approach is possible because of very similar morphology and vocabulary of ex-Yugoslav languages. The second approach includes creation of semantic space using Latent Semantic Analysis (LSA). An LSA approach would entail a creation of a virtual semantic space that relies on the assumption that, provided the languages are using the same script (latin or cyrillic) the underlying similarities between the languages can be exploited to perform a semantically driven search taking into account synonymy and co-occurrence information in all the languages that are to be supported (Serbian, Croatian, Montenegrin and Bosnian). The objective of both those approaches is to ensure results published in the Serbian language PhD dissertation are also visible to the scientists from ex-Yugoslav region, and will be presented in the full version of paper.

References

1. Ivanovic, L., Ivanovic, D., Surla, D. (2012), "Integration of a Research Management System and an OAI-PMH Compatible ETDs Repository at the University of Novi Sad, Republic of Serbia", *Library Resources and Technical Services*, Vol. 56, No. 2, pp. 104-112, DOI 10.5860/Irts.56n2
2. Ivanovic, L., Ivanovic, D., Surla, D. (2012), "A data model of theses and dissertations compatible with CERIF, Dublin Core and EDT-MS", *Online Information Review*, Vol. 36, No. 4, pp. 568-586, DOI 10.1108/14684521211254068
3. Ivanović, D., Surla, D. & Konjović, Z. (2011), "CERIF compatible data model based on MARC 21 format", *The Electronic Library*, Vol. 29, No. 1, pp. 52-70, DOI: [10.1108/02640471111111433](https://doi.org/10.1108/02640471111111433)
4. LiberEUROPE, (2017), "Implementing FAIR Data Principles: The Role of Libraries"
5. Layfield, C., Ivanovic, D., & Azzopardi, J. (2017). Multi-Lingual LSA with Serbian and Croatian : An Investigative Case Study. In Proceedings of the 3rd International KEYSTONE Conference IKC 2017. Gdansk.