FAIRness at University of Novi Sad - Discoverability of PhD research results for Non-Serbian scientific community –

Ivanović Dragan¹ and Ivanović Lidija¹  Layfield Colin²

¹University of Novi Sad
²University of Malta

FAIRness of research information, CRIS 2018
Outline

1. Introduction
   - University
   - DOSIRD UNS
   - Motivation

2. English scientific community
   - Community
   - Solution

3. Ex-Yugoslav scientific community
   - Ex-Yugoslav region
   - Common preprocessing tool
   - Ex-Yugoslav common semantic space

4. Conclusion
Introduction

University
- DOSIRD UNS
- Motivation

English scientific community
- Community
- Solution

Ex-Yugoslav scientific community
- Ex-Yugoslav region
- Common preprocessing tool
- Ex-Yugoslav common semantic space

Conclusion
University of Novi Sad

The first faculty in Novi Sad was founded in 1954
The University of Novi Sad was founded on 28th of June 1960
Today, UNS represents an autonomous institution for education, science and arts
The first faculty in Novi Sad was founded in 1954
The University of Novi Sad was founded on 28th of June 1960
Today, UNS represents an autonomous institution for education, science and arts
The first faculty in Novi Sad was founded in 1954
The University of Novi Sad was founded on 28th of June 1960
Today, UNS represents an autonomous institution for education, science and arts
Rectorate building
University of Novi Sad Cities

dragan.ivanovic@uns.ac.rs
Outline

1. Introduction
   - University
   - DOSIRD UNS
   - Motivation

2. English scientific community
   - Community
   - Solution

3. Ex-Yugoslav scientific community
   - Ex-Yugoslav region
   - Common preprocessing tool
   - Ex-Yugoslav common semantic space

4. Conclusion
About project

- Software infrastructure for educational domain - well developed, but lack of software infrastructure for science
- The DOSIRRD UNS project has been started in 2009 - http://dosird.uns.ac.rs/
- The project team - professors, scientists, PhD students and librarians of the University of Novi Sad
- The main goal was to develop software infrastructure for research domain
Current Research Information System of University of Novi Sad

Under development since 2009

Fulfills all local requirements prescribed by
- the University
- government of the Autonomous province of Vojvodina and
- government of the Republic of Serbia

In accordance with international well-known research domain standards and protocols

Data about scientific outputs are stored in MARC 21 format

Architecture of the system enables an easy integration with LIS and CRIS
Usage statistic

- 700 registered users
- 30,000 scientific publications
Digital library of dissertations defended at the University of Novi Sad

Under development since 2010

Integrated as part of CRIS UNS

The first version was put into operation in 2013 - link
Usage statistic

- Metadata about all dissertation defended at the University since its foundation - 5,000
- Attached pdf files - 1,400
- Open-access pdf files - 1,250
- 250,000 downloads
Outline

1. Introduction
   - University
   - DOSIRD UNS
   - Motivation

2. English scientific community
   - Community
   - Solution

3. Ex-Yugoslav scientific community
   - Ex-Yugoslav region
   - Common preprocessing tool
   - Ex-Yugoslav common semantic space

4. Conclusion
Discoverability of PHD dissertations

- Metadata and data should be findable and understandable to humans and machines according to the FAIR principles.
- Authors want to attract interest to their work to enhance their reputation both nationally and internationally.
- Metadata about dissertations in two languages (Serbian and English) improves discoverability of Serbian PhD dissertations, non-Serbian researchers still can’t read and search by full-text.
Goal

- Make findable and accessible research information stored in PHD UNS dissertations to the non-Serbian scientific community
  - English scientific community
  - Ex-Yugoslav scientific community
Goal

- Make findable and accessible research information stored in PHD UNS dissertations to the non-Serbian scientific community
  - English scientific community
  - Ex-Yugoslav scientific community
Goal

- Make findable and accessible research information stored in PHD UNS dissertations to the non-Serbian scientific community
  - English scientific community
  - Ex-Yugoslav scientific community
Outline

1. Introduction
   - University
   - DOSIRD UNS
   - Motivation

2. English scientific community
   - Community
   - Solution

3. Ex-Yugoslav scientific community
   - Ex-Yugoslav region
   - Common preprocessing tool
   - Ex-Yugoslav common semantic space

4. Conclusion
Non-English science

- English is global language
- Any scientist should master English
- But it is certainly not the native language for the majority of scientists
- There is a science beyond science published in English
- 25 winners of the Nobel Prize in Literature up to 2001, wrote only 9 of their masterpieces in English and remaining 16 laureates in other native languages
Some countries are making extra efforts to gain scientific visibility in their national language

- SciELO (Scientific electronic library online - http://www.scielo.br/)
Non-English theses and dissertations

- A lot of ETD repositories containing mainly non-English theses and dissertations
- Serbian case - more than 90% of PhD dissertations in Serbian language
- Authors of scientific publications want to attract interest to their work to enhance their reputation
  - nationally
  - and internationally
Outline

1. Introduction
   - University
   - DOSIRD UNS
   - Motivation

2. English scientific community
   - Community
   - Solution

3. Ex-Yugoslav scientific community
   - Ex-Yugoslav region
   - Common preprocessing tool
   - Ex-Yugoslav common semantic space

4. Conclusion
Three general approaches

- Writing in English - Is that good for national heritage and science?
- Translation into English - Time-consuming process
- Writing in non-English language, publishing research parts in English
Three general approaches

- Writing in English - Is that good for national heritage and science?
- Translation into English - Time-consuming process
- Writing in non-English language, publishing research parts in English
Three general approaches

- Writing in English - Is that good for national heritage and science?
- Translation into English - Time-consuming process
- Writing in non-English language, publishing research parts in English
Writing in non-English language, publishing research parts in English

- Writing dissertation in non-English language
- Publishing parts of PhD research in journals, conference proceedings, etc.
- Publishing parts of PhD research is natural and very common process
- There are regional or national rules which prescribe this as mandatory
The Serbian Ministry of Science prescribed a rule - publishing an article in a journal belonging to the Web of Science JCR list is mandatory.

Results published in Serbian language in a PhD dissertation are also visible to world science community through journal articles published in English.
Our approach

- Associated journal articles can be stored in CRIS UNS system
- The connection between a PhD dissertation and journal article is established by author or librarian
- If a PhD dissertation is a result of local searching of PHD UNS, a reference to the associated journal article written in English is provided
- If a PhD dissertation has been exported to ETD repositories networks, link (identifier) to the PhD dissertation stored in an ETD node should open web page which provides a reference to the associated journal articles written in English as well
Our approach

- Associated journal articles can be stored in CRIS UNS system
- The connection between a PhD dissertation and journal article is established by author or librarian
  - If a PhD dissertation is a result of local searching of PHD UNS, a reference to the associated journal article written in English is provided
  - If a PhD dissertation has been exported to ETD repositories networks, link (identifier) to the PhD dissertation stored in an ETD node should open web page which provides a reference to the associated journal articles written in English as well
Our approach

- Associated journal articles can be stored in CRIS UNS system
- The connection between a PhD dissertation and journal article is established by author or librarian
- If a PhD dissertation is a result of local searching of PHD UNS, a reference to the associated journal article written in English is provided
- If a PhD dissertation has been exported to ETD repositories networks, link (identifier) to the PhD dissertation stored in an ETD node should open web page which provides a reference to the associated journal articles written in English as well
Our approach

- Associated journal articles can be stored in CRIS UNS system.
- The connection between a PhD dissertation and journal article is established by author or librarian.
- If a PhD dissertation is a result of local searching of PHD UNS, a reference to the associated journal article written in English is provided.
- If a PhD dissertation has been exported to ETD repositories networks, link (identifier) to the PhD dissertation stored in an ETD node should open web page which provides a reference to the associated journal articles written in English as well.
Establishing connection

FAIRness at University of Novi Sad, CRIS 2018, Umea, Sweden
Searching PHD UNS

Search results

Number of results is 1
Sort by: relevance

MILJIĆ, U. (2015) Production and quality assessment of fruit wines from native plum (Prunus domestica L.) varieties. (PhD dissertation), Faculty of Technology at Novi Sad


**Title**: Production and quality assessment of fruit wines from native plum (Prunus domestica L.) varieties  
**Author**: Milić Uroš  
**Subject(s)**: Plum, fruit wine, fermentation, quality assessment, characterization, Sljiva, voćno vino, fermentacija, ocena kvaliteta, karakterizacija  

**Abstract**:

The aim of this PhD thesis was to assess the possibility of using three native plum varieties, with different ripening periods (Čačanska rana, Čačanska lepota and Požeška), as raw material for the production of fruit wines. Determined mechanical composition and chemical characteristics of fruit pulp and juice indicate that the Čačanska lepota and Požeška are considered as better raw materials for the production of fruit wine compared to Čačanska rana. Optimization of fermentation conditions (temperature, pH, duration of fermentation and the dose of pectolytic enzyme) was conducted. This step also included investigation of the different pectolytic enzymes use for the treatment of pomace and evaluated the impact of using different yeast strains, as well as the effect of different production microorganisms on the plum wine quality. It was found that among the tested production microorganisms, Saccharomyces (S. cerevisiae) yeasts give the plum wines of best quality. Numerical optimization procedure resulted with the following values of the process parameters of plum wine fermentation: temperature 25 °C, pH value 3.5 and pectolytic enzyme dose 0.5 g/100 kg. Under these conditions the obtained fitted models predict the ethanol yield of 7.9%, glycerol yield 5.5%, the wine yield of 45% (49 ml from 100 g of pomace) and the formation of 710 mg/l of methanol. Characterization of the produced plum wines included the determination of the most important ingredients: alcohol, acids, minerals, phenolic and aromatic compounds, as well as evaluation of their functional characteristics (antiradical, antimicrobial and antiproliferative activities). Finally, the possibility of reducing the must-sediment production in plum wine was estimated by applying different physico-chemical treatments of the pomace. Significantly higher efficiency of procedures that involve some form of heat treatment of pomace, compared to treatments which involve the use of certain enzymatic mix, was observed.

**Publisher**: University of Novi Sad, Faculty of Technology  
**Date**: 2015-04-06  
**Type**: PhD thesis  
**Identifier**: http://www.cris.uns.ac.rs/record.jsf?recordId=93703&source=DAExecutiveLanguage-en
PHD UNS record page

Record preview

Thesis or dissertation
Author: Milija Uroš
Publication language: Serbian
Title: Production and quality assessment of fruit wines from native plum (Prunus domestica L.) varieties
Theses/dissertation type: PhD dissertation
Institution: Faculty of Technology at Novi Sad, University of Novi Sad
Year: 2015
Download:

Related records


Outline

1. Introduction
   - University
   - DOSIRD UNS
   - Motivation

2. English scientific community
   - Community
   - Solution

3. Ex-Yugoslav scientific community
   - Ex-Yugoslav region
   - Common preprocessing tool
   - Ex-Yugoslav common semantic space

4. Conclusion
After 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.

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.
Two approaches

- There are two approaches for developing ex-Yugoslav language tools which would support those multi-lingual search capabilities
  - Development of common preprocessing techniques including stemming and parallel vocabularies
  - Creation of semantic space using Latent Semantic Analysis (LSA)
Two approaches

There are two approaches for developing ex-Yugoslav language tools which would support those multi-lingual search capabilities:

- Development of common preprocessing techniques including stemming and parallel vocabularies
- Creation of semantic space using Latent Semantic Analysis (LSA)
There are two approaches for developing ex-Yugoslav language tools which would support those multi-lingual search capabilities.

- Development of common preprocessing techniques including stemming and parallel vocabularies.
- Creation of semantic space using Latent Semantic Analysis (LSA).
Outline

1. Introduction
   - University
   - DOSIRD UNS
   - Motivation

2. English scientific community
   - Community
   - Solution

3. Ex-Yugoslav scientific community
   - Ex-Yugoslav region
   - Common preprocessing tool
   - Ex-Yugoslav common semantic space

4. Conclusion
The tool

- Ex-Yugoslav language tool performs preprocessing of PDF texts and metadata before its indexing.
- This language tool is developed as a Lucene Analyzer.
- The developed ex-Yugoslav language tool has been put into operation in 2013.
- About 5% of total number of downloads is performed from Croatia, Montenegro and Bosnia and Herzegovina (three ex-Yugoslav countries).
The preprocessing steps

- Case folding
- Alphabet conversion
- Stop word removal
- Stemming
- Vocabularies translation
- Characters conversion
The preprocessing steps

- Case folding
- Alphabet conversion
  - Stop word removal
  - Stemming
- Vocabularies translation
- Characters conversion
The preprocessing steps

- Case folding
- Alphabet conversion
- Stop word removal
  - Stemming
  - Vocabularies translation
  - Characters conversion
The preprocessing steps

- Case folding
- Alphabet conversion
- Stop word removal
- Stemming
- Vocabularies translation
- Characters conversion
The preprocessing steps

- Case folding
- Alphabet conversion
- Stop word removal
- Stemming
- Vocabularies translation
- Characters conversion
The preprocessing steps

- Case folding
- Alphabet conversion
- Stop word removal
- Stemming
- Vocabularies translation
- Characters conversion
Drawbacks

- Time-prone process of creation parallel vocabularies - it is almost impossible to create complete parallel vocabulary for Serbo-Croatian languages
- Even though we have complete parallel vocabulary, it should be maintained taking into account that languages evolve
Outline

1. Introduction
   - University
   - DOSIRD UNS
   - Motivation

2. English scientific community
   - Community
   - Solution

3. Ex-Yugoslav scientific community
   - Ex-Yugoslav region
   - Common preprocessing tool
   - Ex-Yugoslav common semantic space

4. Conclusion
An exploratory case study was performed in 2017 that investigated the creation of multi-lingual semantic spaces for Information Retrieval.

This experiment targeted the Serbian and Croatian languages.

These languages are similar - it is possible to create a joint, or mixed, semantic space utilising Latent Semantic Analysis, for semantic similarity comparison purposes.

Latent Semantic Analysis, or LSA, is an Information Retrieval technique whose origins are found in factor analysis.

Accuracy is about 90%.
Future Research Directions Utilising LSA

- Check results with four languages: Serbian, Croatian, Montenegrin and Bosnian
- The automatic generation of a synonymy vocabulary list between the different languages contained in a semantic space
Future Research Directions Utilising LSA

- Check results with four languages: Serbian, Croatian, Montenegrin and Bosnian
- The automatic generation of a synonymy vocabulary list between the different languages contained in a semantic space
Summary

- More than 90% of PhD dissertations of the republic of Serbia have been written in national language (Serbian).
- Authors of scientific publications usually want to attract interest to their work to enhance their reputation both nationally and internationally.
- An approach which makes accessible research presented in Serbian PhD dissertations for the English scientific community through connected journal papers have been described.
- Two approaches for making findable research presented in Serbian PhD dissertations for the ex-Yugoslav scientific community have been described.
More than **90%** of PhD dissertations of the republic of Serbia have been written in **national language (Serbian)**

Authors of scientific publications usually want to **attract interest** to their work to enhance their reputation both **nationally and internationally**

An approach which makes **accessible** research presented in Serbian PhD dissertations for the **English scientific community** through connected journal papers have been described

Two approaches for making **findable** research presented in Serbian PhD dissertations for the **ex-Yugoslav scientific community** have been described
Summary

- More than 90% of PhD dissertations of the republic of Serbia have been written in national language (Serbian).
- Authors of scientific publications usually want to attract interest to their work to enhance their reputation both nationally and internationally.
- An approach which makes accessible research presented in Serbian PhD dissertations for the English scientific community through connected journal papers have been described.
- Two approaches for making findable research presented in Serbian PhD dissertations for the ex-Yugoslav scientific community have been described.
Summary

- More than **90%** of PhD dissertations of the republic of Serbia have been written in **national language (Serbian)**
- Authors of scientific publications usually want to **attract interest** to their work to enhance their reputation both nationally and internationally
- An approach which makes **accessible** research presented in Serbian PhD dissertations for the **English scientific community** through connected journal papers have been described
- Two approaches for making **findable** research presented in Serbian PhD dissertations for the **ex-Yugoslav scientific community** have been described
Thank you for your attention!!!
If you have any questions, please do not hesitate to
  • ask me now
  • ask me during the conference breaks
  • contact me via email address dragan.ivanovic@uns.ac.rs