Lejla Kodrić, Sanjin Kodrić,
University of Sarajevo,
Faculty of Philosophy
Teo Eterović
University of Sarajevo,
Faculty of Electrical Engineering

ONLINE PRESENCE OF MEŠA SELIMOVIĆ'S LITERARY HERITAGE: DIGITAL HERITAGE REPOSITORY IN THE FUNCTION OF LITERARY AND CULTURAL STUDIES

Abstract. Potentials of scientific research in general are obviously subject to changes in accordance with current digital information and communication environment, which is capable to give new, advanced performances, or those essentially different from the traditional ones, even in the domain of humanities, including literary studies as well. In this connection, developing a digital heritage repository, as a way of organizing and composing objects of heritage in selected thematic entities, is demanding, but also inspiring possibility in the case of literary heritage as well. The aim of the proposed paper is to examine potentials and advantages of developing a digital repository of Meša Selimović's literary work. The focused problem is going to be approached in a holistic manner, which means in terms of its practical and cultural values, but also in terms of its feasibility. The authors of the proposed paper are taking interdisciplinary methodological sides, which is caused by the authors' different academic education and positions within the academic community (library and information sciences, literary history and literary and cultural theory, as well as electrical engineering).

Keywords: literary and cultural heritage, digital heritage repository / digital library, Meša Selimović

Online Presence of Literary Heritage

Potentials of scholarly research in general are obviously subject to changes in accordance with current digital information and communication environment, which is capable to give new, advanced performances, or those essentially different from the traditional ones, even in the domain of humanities, including literary and cultural studies as well. In this connection, developing a digital heritage repository / digital library, as a way of organizing and composing objects of heritage into selected thematic entities, is a demanding, but also inspiring possibility in the case of literary heritage too. This is exactly why it is important to think about developing a strategy for building a particular, although until now not existing, but as a scholarly, educational and teaching aid increasingly more interesting digital repository of the literary work of Meša Selimović (Tuzla, 1910 – Belgrade, 1982), which is one of the most important canonical values of the entire Bosnian-Herzegovinian culture, as well as the wider South-Slavic cultural context in general, especially after his great novels *Derviš i smrt* (1966) and *Tvrđava* (1970), the author's genuine masterpieces.

In view of this, the aim of the proposed paper is to examine potentials and advantages of developing a digital repository of M. Selimović's literary work. The focused problem is going to be approached in a holistic manner, which means in terms of its practical and cultural values, but also in terms of its feasibility.

The authors of the proposed paper are taking interdisciplinary methodological sides, which is caused by different academic education and positions within the academic community of the authors (library and information sciences, literary history and literary and cultural theory, as well as electrical engineering). This is exactly what gives an opportunity to observe potentials of developing a literary heritage digital repository from different perspectives: the perspective of increasingly more necessary information sciences and information technology, but also the perspective of literary and cultural studies. Finally, efforts of this kind are even more important in the year 2010, since it is the year when the entire Bosnian-Herzegovinian culture celebrates the 100th anniversary of M. Selimović's birth.

Repository of this kind would initiate a unique gathering of media information sources completely different thematically, chronologically and/or in terms of genres, authors etc., addressing in a way the work and personality of an individual author – in this case M. Selimović. Having such a repository, literary and cultural scholar, researcher of another sort or even a common reader would be able to use at least:

- every M. Selimović's work on a single computer interface, even those works that are less accessible in broader context,
- manuscripts, critical editions and/or early versions of canonical works, unpublished and/or unfinished works, interviews in different media, writing notes, personal letters, private photos,
- works of authors who have dealt with M. Selimović's work and personality in any genre or medium, and
- critical works and commentaries in full text or partially, bibliographies, works of art, such as metaliterary works (poems, stories, theater plays, TV programs, films etc.) referring in a way to M. Selimović, pictures, drawings, graphics, sculptures, music compositions inspired by the author and/or his work.

Digital Heritage Repository / Digital Library

The digital heritage repository / digital library has been implemented on top of the Unified Meta Data Format (UMDF) as described in [1]. The use of the UMDF enables us to bind metadata and media (audio, video, books, 3D models) into one file, so that we can design a simple digital library application model with an ability of full-text search through the digital repository, no matter which media type is used. Also, the encapsulated metadata is practical for clustering and classification algorithms. The library supports the following formats: any format of video, any format of sound, any format of books, online created articles (reviews, references, etc.) and documents.

Indexing / Search Engine

We implemented a customized indexing engine based on Apache Lucene that indexes the UMDF format iterating trough the repository folder and extracting the data from the zip file format and indexing it as follow (boosting score factor):
The metadata from the meta/plain.xml:

- title boost score 5.0,
- summary boost score 4.0,
- document filename boost score 3.0.

Also, the resulting documents are ranked based on the Vector Space Model and cosine similarity query matching where we represent the document and the query as vectors $d = (\omega_1, \omega_2, \ldots, \omega_{|T|})$ and the query as $d = (\omega_{q1}, \omega_{q2}, \ldots, \omega_{q|T|})$ respective in a multidimensional space. We use the query terms as dimensions and the documents and the query as vectors in that space as shown in Figure 1.
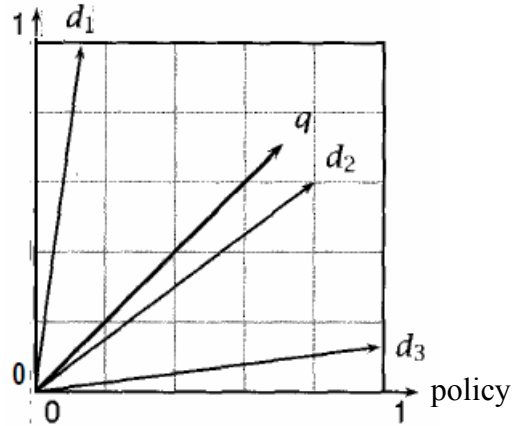
Figure 1: Graphic presentation of the Vector Space Model for query matching, where d_1, d_2, d_3 are the documents, q is the query and *car* and *policy* are the query keywords represented as dimensions

We use the cosine similarity model to calculate the distance between the document and the query and use it as our relevance score. We can calculate the angle which represents the similarity as shown in 1.2, where A and B are vectors (for more information see [2]).

$$Sim(A, B) = cos(\theta) = \frac{A \; B}{|A||B|} = \frac{x_1 \cdot x_2 + y_1 \cdot y_2}{(x_1^2 + y_1^2)^{\frac{1}{2}} \cdot (x_2^2 + y_2^2)^{\frac{1}{2}}} \quad (1.2)$$

We use tf-idf with length normalization because the sound, video and picture resources have less textual content than electronic documents and we need to make the content proportional. As shown in 1.3 [3], the $tf_i \log(N/N_i)$ represents the term-frequency × inverse document frequency, where i represents the i-th term in the document d and the number of occurs of i inside d is represented as N that is reverse proportional to N_j – the number of documents that contain the term i inside the corpa. The $\sum_{j=1}^{|T|} [tf_j \cdot \log(N/N_j)]$ **Error! Bookmark not defined.**part represents the normalization, where |T| is the number of unique terms inside the document.

$$\omega_i = \frac{tf_i \cdot \log(N/N_i)}{\sum_{j=1}^{|T|} [tf_j \cdot \log(N/N_j)]} \quad (1.3)$$

We also made some modifications to the search engine in order to support multilingual searching. For that purpose, we used the Google translate API that automatically detects the language in which the query was made and translates it into other languages and finally extends the query with the translated keywords.

The results are formatted depending on the UMDF metadata specifications. The search engine supports customized sound, video and image presentation. A quick sound and video

preview have been implemented in Flash and the image preview using DHTML and JavaScript as shown in 1.X.
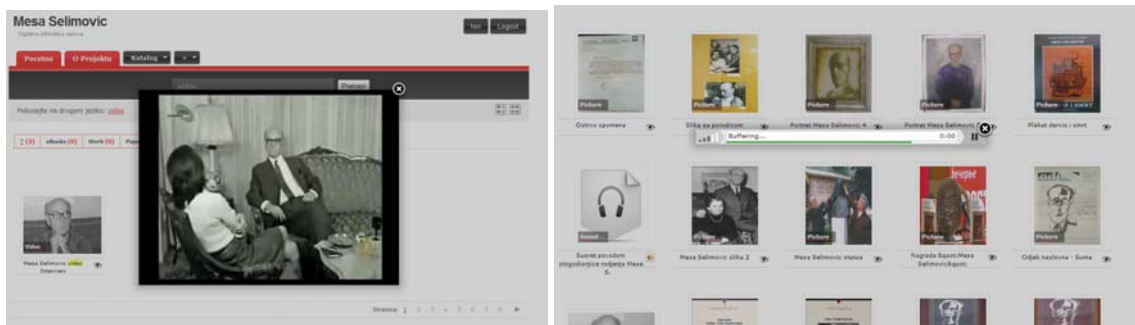

Figure 1.X: Video and audio preview

Navigation

The digital library has three types of content navigation: a catalog made by a domain expert, an 'intelligent algorithm' clustered view and the search engine. The search engine has been described before; it is a fast way to navigate through the data – with only two clicks plus keyword typing. The domain expert catalog has been implemented through a list of categories that the domain expert can select when adding the resource. The category view has been implemented through hyperbolic graph as shown in 1.4.
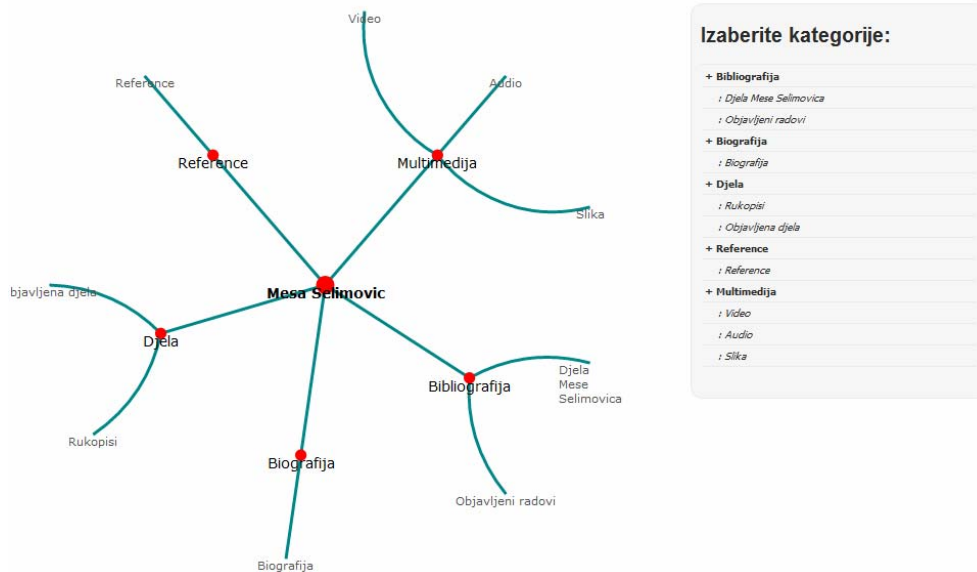

Figure 1.4: Hyperbolic graph used for category presentation

The clustering algorithm has been implemented trough a 'label first' algorithm. The idea is to find, first, meaningful cluster labels through Singular Value Decomposition and term co-relevance using suffix arrays and, then, to assign the documents to the clusters using the Vector Space Model in order to make the cluster-document relations harder – more relevant. We used a cluster graph to visualize it as shown in 1.5.
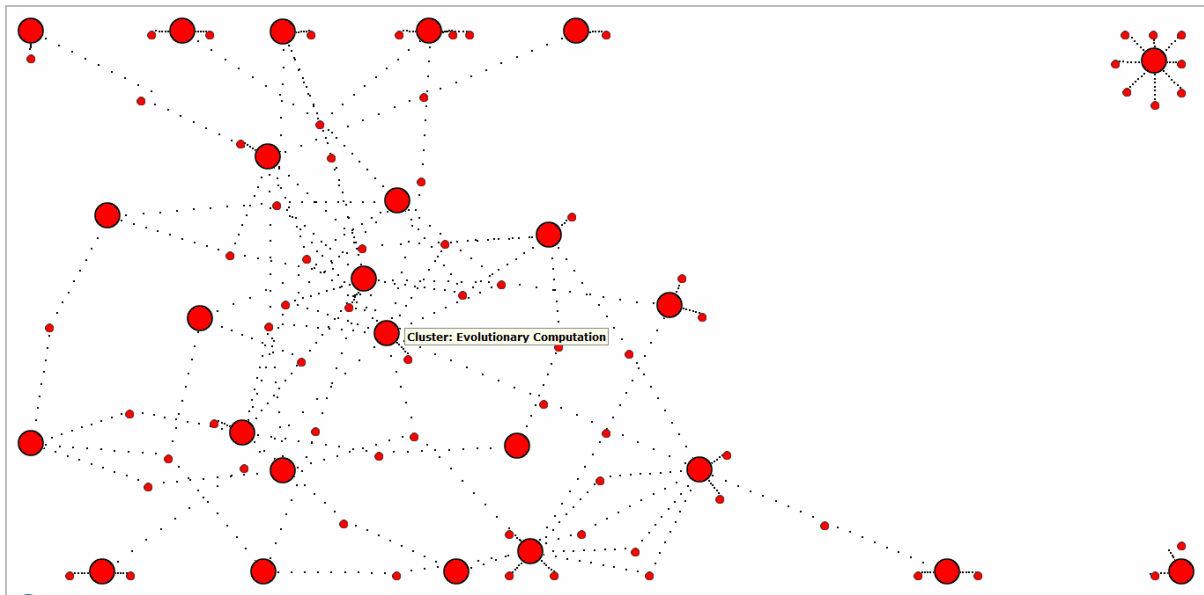
Figure 1.5: Cluster map

View

The document view has been designed in spirit of Web 2.0 concepts – in a meaningful way. We implemented the following Web 2.0 concepts:

- wiki – in order to edit the resource description. We believe it can be useful to give website users such as literary and cultural scholars, students and researches as well as other users of any orientations and interests an ability to describe the resources such as pictures, audio-video interviews and books authored by M. Selimović in a collaborative way,
- comments – users are able to have conversation about a specific book, interview, picture etc.,
- score – users are able to rate quality of a resource.

References

The use of the UMDF enables us to view the various media type in a unified way, so it is easy to refer one resource to another. For example, we referred a critical review of M. Selimović's novel to this author's work. We can also refer another critical work to a critical work that refers an initial literary work, as shown in picture 1.6.

Figure 1.6: References

Final Remarks

As shown in M. Selimović's case, using digital information resources offers a general single narrative sequence presentation of the author's personality and literary work for scholarly and non-scholarly purposes. It also gives an opportunity of simultaneous using digital information resources stored within a number of heritage institutions and private collections throughout Bosnia-Herzegovina and South-East Europe. It is even more important if we have on mind that – due to different reasons – M. Selimović's literary legacy is traditionally disunited and stored in numerous places (mostly in Sarajevo, Tuzla and Belgrade). Through its digital versions, the author's literary heritage would be, for the first time, given as 'physically' united and disclosed to a researcher or any other user as a single, coherent system of literary heritage, much more simple and accessible for any literary / cultural research or other, even non-professional usage. As it usually happens in projects of this kind, the boundaries of this literary digital repository would terminate the necessity of depositing from only one or few institutions / collections, making the 'space' which would, with its digitized information objects referring to the author, attract attention of many individual and institutions of different interests. In this way, institutional and geographical distance of traditional sources of information in connection with the author would be overcome.

Project of this kind opens possibilities of approaching the author and other, wider literary / cultural phenomena in the way of exploiting much bigger number of different information sources than it is possible in the analog world. An earlier principle of selectivity and restriction to only one medium or only one institutional framework referring to the author's work, now exceeds simultaneous and parallel using of different sources, bibliographical, critical and other texts, such as audio-visual records etc. that can be used to help research process and general reception of the opus or another literary / cultural issue discussed. These 'thematic exhibitions', typical for digital heritage repositories of today, are way of providing valuable information for literary / cultural researchers, students of literature and other cultural fields, but also for common or so-called 'free' readers. This is especially important in the cases of heritage institutions dealing with the literary heritage of a given author which are for some reason inaccessible for a user (e.g. long distance and other technical inconveniences, special membership and other special requirements, physically disabled user etc.).

Heritage digital repositories / digital libraries of this kind are also a guarantee of recognition of specific identities of 'small' languages and 'small' cultures in relation to cultures and literatures of 'great' nations under conditions in which most researchers in their scholarly endeavors prefer to use the so-called 'space flows'. In addition to the improved access to literary and cultural heritage, a digital heritage repository / digital library of this kind

in both the author's and general case would ensure protection of traditional heritage resources, especially manuscript and other original and rare heritage materials. In this way, manuscripts and similar heritage materials would still be available for interested public, while being protected from possibility of destruction during immediate handling.

In addition to a number of other ways of usage, the idea of online presence of M. Selimović's literary heritage could:

- serve as a potential model for other similar projects, including those much wider (e.g. digital repository / digital library of Bosnian-Herzegovinian literature),
- connect synergicly humanities, social sciences and modern technology, and
- benefit the author, different sorts of researchers of his work and personality, students, common readers etc., as well as Bosnian-Herzegovinian culture and society as a whole, especially in the way of promoting authentic cultural values.

References

[1] T. Eterović. N. Šrndić, Introducing The Unified E-Book Format And A Hybrid Library 2.0 Application Model Based On It, Review of the National Center for Digitization, pp 43–52, 2008.
[2] E.Gracia,"Cosine Similarity and Term Weight Tutorial", MiIslita SEWF, 2006.
[3] S. Robertso, Understanding Inverse Document Frequency: On theoretical arguments for IDF, Journal of Documentation 60 no. 5, pp 503–520.
[4] L. Kodrić. Model digitalnih informacijskih usluga u baštinskim ustanovama: doktorska disertacija, Filozofski fakultet Sveučilišta u Zagrebu, 2010.
[5] S. Kodrić, Kulturalno pamćenje i reprezentacija prošlosti u bošnjačkoj književnosti 20. stoljeća (Od utilitarizma preporodnog doba do savremene poetike svjedočenja): doktorska disertacija, Filozofski fakultet Univerziteta u Sarajevu, 2009.

lejla.kodric@ff.unsa.ba
sanjin.kodric@ff.unsa.ba
teo.eterovic@etf.unsa.ba