Zuzana Nevěřilová

Implementing Dynamic Visualization as an Alternative Interface to a Digital Mathematics Library

In: Petr Sojka (ed.): Towards a Digital Mathematics Library. Paris, France, July 7-8th, 2010. Masaryk University Press, Brno, Czech Republic, 2010. pp. 63--68.

Persistent URL: http://dml.cz/dmlcz/702574

Implementing Dynamic Visualization as an Alternative Interface to a Digital Mathematics Library

Zuzana Nevěřilová

Masaryk University, Faculty of Informatics
Botanická 68a, 602 00 Brno, Czech Republic
xpopelk@fi.muni.cz

Abstract. This paper presents an alternative interface for browsing in the Czech Digital Mathematics Library (DML-CZ) using our Visual Browser web browsing tool. Using dynamic visualization, we have created a tool for browsing the library graphically. Visualization can help users orient themselves in complex data and at the same time reveal sometimes unexpected relationships among units; it at least speeds up browsing.
This work follows the metadata processing undertaken on DML-CZ and visualizes all reasonable and useful relationships among journals, issues, articles, authors, classification, keywords, references and similar articles. We converted metadata to RDF and use a Visual Browser Java Applet that runs in a web browser. We describe briefly the metadata nature, then server and client side of the visualization including data formats and conversions. There follows a description of the interaction between visual and textual interfaces.

Key words: visualization, RDF, visual interface, Visual Browser, DML-CZ, EuDML

1 Introduction

This paper presents dynamic visual interface for browsing the Czech Digital Mathematics Library (DML-CZ) as an alternative to a textual listing. We are offering the interface to the ongoing EuDML project[1] [13]. The DML-CZ [1] currently contains more than 28,000 articles in 11 journals, 5 proceedings series and 28 monographs [6]. Users usually do not browse within such a vast amount of data, rather they search for titles or authors.

On the search results page users can see the number of search results and the list of articles. When clicking on an article, the information listed below is shown:

- bibliographic information about the article (author, title, serial, year, Mathematics Subject Classification (MSC) [7], ...);
- preview of the article and link to the PDF;
- link to similar articles;
- references with links to articles where possible.

[1] The European Digital Mathematics Library – http://www.eudml.eu/

Petr Sojka (editor): DML 2010, Towards a Digital Mathematics Library, pp. 63–68.
 ISBN 978-80-210-5242-0

The particular advantage of the DML-CZ interface is that it finds *similar articles* in search results. Three methods for calculating similarities are used [10] and the percentages are expressed graphically. This is so far the only information that is visualized. Nevertheless according to [14] a good visualization helps accelerating the cognitive process, since the eyes can pick up details of the visualization and keep a holistic overview at the same time. Visualization is most suitable for complex and relatively sparse data and this is precisely the case of library data.

Google has started to offer a graphical interface for search results in addition to the standard view: their so-called *Google Wonder Wheel*[2] has both plain text and timelines. Information seekers who would tend to use it, are likely to appreciate it for not only Google searches.

The structure of paper is as follows. In Section 2 we describe the server side including data formats provided by the server. Section 3 briefly describes the Visual Browser and shows the interaction between the Visual Browser and the textual listing on the web page. Section 4 contains both the conclusion and the future development that the dynamic visual interface may undergo.

2 Server Side

Since the amount of data in DML-CZ is very large, a client-server architecture is the most suitable. The server has to store the data, provide a method for its retrieval and quickly return a small amount of the data requested.

2.1 Data Formats

Because the client side uses RDF [2], the server has also to provide this format. We had to convert the existing XML format of metadata to RDF. This conversion required the following steps:

- selecting only the appropriate data for visualization (some information is omitted);
- assigning IDs for articles, issues, journals and authors;
- adding short titles for the visualization;
- conversion of the `lang` attribute according to RFC 3066 sec. 2.3 [12];
- adding information about similar articles;
- adding MSC labels.

2.2 RDF Server

Joseki RDF Server[3] was used. It offers SPARQL [9] as a query language. Joseki was selected because of the Jena Framework[4] used in the client. Nevertheless, the server side can be substituted by any other RDF server if needed. The data is stored in a relational database.

[2] `http://www.googlewonderwheel.com/`
[3] `http://joseki.sourceforge.net/`
[4] `http://jena.sourceforge.net/`

3 Client Side

On the client side two interfaces are used: a traditional textual interface (a list of authors and articles) and the Visual Browser [8]. The latter is a tool for the dynamic (animated) visualization of RDF graphs. It provides flexible visualization thanks to the two-layer architecture:

- first layer—the data stored in RDF (whether in RDF/XML, N3 [3] or Turtle [5]);
- second layer—*perspective of view*, an XML description of graphic representation of nodes and edges of the graph.

The visualization of different types of data is described below.

The Visual Browser exists either as a standalone Java application or as a Java applet. The applet can communicate with textual parts of the search results page. The interaction Java Applet—web page was made through AJAX[5] plus JavaScript to communicate with the applet.

Submitting the search field or browsing data in one of the interfaces results in a SPARQL query. The server evaluates the query and returns an RDF graph. The XSLT [4] conversion is made and the result is returned as a list of authors and titles. The communication between the applet and the web page is bi-directional: clicking on a name or title in the list renders a set of nodes and edges in the visual interface, a set of nodes and edges can be displayed as a list of authors and articles.

We expect users to type (part of) a name or title in the search box. Then users can browse either the more familiar textual interface (as they are used to), or the visual one. Conversely, when viewing a particular subgraph in the Visual Browser, users can click to have it appear it in the textual interface as seen in Figure 1.

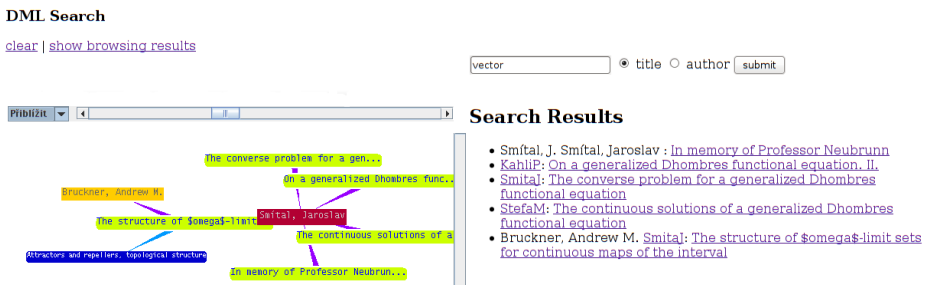

Fig. 1. Visual and textual interfaces to search results within DML-CZ. It allows users to choose how they browse the results. For this purpose, it is necessary that the interfaces are able to communicate.

[5] Asynchonous JavaScript And XML

3.1 Visualizing Metadata

In this visualization, nodes represent units such as authors, articles, issues, journals, and keywords and MSC as well. Different classes of units are represented by different colours and shapes. Mapping from logical entities to their visual attributes is fully configurable in Visual Browser.

Edges represent authors and their articles, articles in issues, issues in journals, as well as links between similar articles. Some of these relations are *structural* (e.g. articles in issues), some are *semantic* (e.g. classification of articles), some have both aspects (authors of articles). We have to evaluate users' behaviour to decide what types of relations are useful for browsing. Even though we expect semantic relations to be more important than the structural ones, we nevertheless display both types of relations. Similarly to the visualization of nodes, the appearance of the edge (its colour, shape and length) distinguishes different classes of edges.

Pointing cursor on a node opens a small po-up window with a short text. This can be helpful when displaying titles or even abstracts as seen in Figure 2.

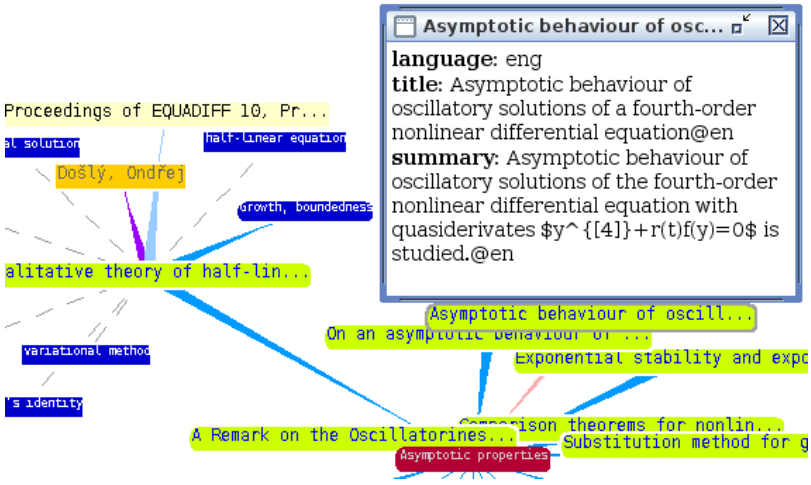

Fig. 2. Visualization of texts: pointing the cursor on a node, more information pop-up in a small window

3.2 Visualizing Similarities

The current interface for DML-CZ provides information about semantically similar articles. Similarities have been pre-calculated [11] using three different methods [10]. Similar articles are connected with edges of different lengths; the shorter the edge, the more similar two articles are (see Figure 3 on the next page).

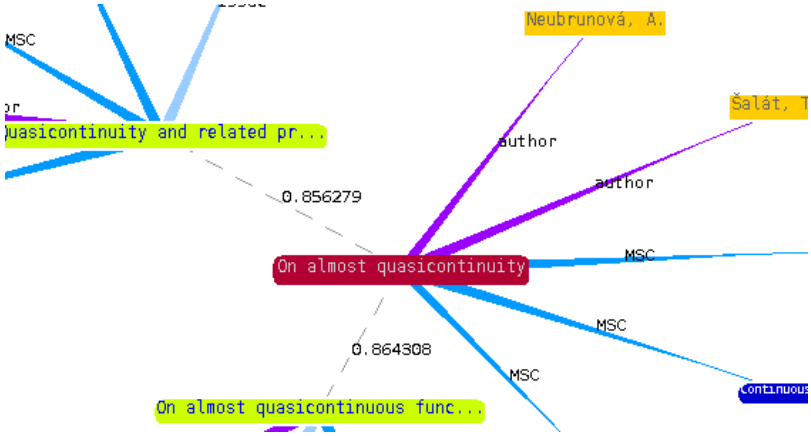

Fig. 3. Visualization of similarities: the length of the edge is also a bearer of meaning; with edge labels displayed, one can also see similarities expressed by numbers

3.3 Visualizing References

Scientific articles usually cite other sources. These citations (references) are related to a topic mentioned in that article and therefore can help users who have already read the article and are looking for further reading. The required state will be that users can browse references to articles regardless of the repository these articles come from. Achieving a high coverage of at least articles' metadata is one of the major goals of the EuDML project.

4 Conclusion and Future Work

We have presented an alternative to the current DML-CZ interface. Visual interfaces are more attractive and can help orientation in complex data such as library records. So far we do experiment with this, targeting at possibility to include it in the official DML-CZ site and offering it to the EuDML project.

Future work comprises monitoring users' preferences on interfaces and their possible feedback. It will probably take some time before users get accustomed to utilize the visual interface, since it is far from the traditional way of browsing. But we hope that users will appreciate the holistic overview of complex information.

Our immediate plans include working on the design of the search result interfaces. For this, users' feedback will be necessary. We also have to test the RDF Server on the significant loads that are expected within DML-CZ and EuDML. These conditions seem necessary for usability within any real-world DML. Working prototype can be seen on

`http://dmlsearch.dml.cz/`.

Acknowledgments

This research has been partially supported by the grant reg. no. 1ET200190513 of the Academy of Sciences of the Czech Republic (DML-CZ), and by EU project # 250,503 in CIP-ICT-PSP.2009.2.4 (EuDML).

References

1. Bartošek, M., Lhoták, M., Rákosník, J., Sojka, P., Šárfy, M.: DML-CZ: The Objectives and the First Steps. In: Borwein, J., Rocha, E.M., Rodrigues, J.F. (eds.) CMDE 2006: Communicating Mathematics in the Digital Era, pp. 69–79. A. K. Peters, MA, USA (2008)
2. Beckett, D., McBride, B.: RDF/XML syntax specification (February 2004), `http://www.w3.org/TR/rdf-syntax-grammar/`
3. Berners-Lee, T.: Notation 3 (2008), `http://www.w3.org/DesignIssues/Notation3`
4. Clark, J.: XSL transformations (XSLT) (1999), `http://www.w3.org/TR/xslt`
5. David Beckett, T.B.L.: Turtle – terse RDF triple language (2008), `http://www.w3.org/TeamSubmission/turtle/`
6. DML: The Czech Digital Mathematics Library – news (2010), `http://dml.cz/news`, retrieved April 20, 2010
7. Ion, P., Eilbeck, C.: Mathematics Subject Classification 2010 (2010), `http://msc2010.org/`
8. Nevěřilová, Z.: Visual Browser: A tool for visualising ontologies. In: Proceedings of I-KNOW'05. pp. 453–461. Know-Center in coop. with Graz Uni, Joanneum Research and Springer Pub. Co., Graz, Austria (2005)
9. Prud'hommeaux, E., Seaborne, A.: SPARQL query language for RDF (2008), `http://www.w3.org/TR/rdf-sparql-query/`
10. Řehůřek, R., Sojka, P.: Automated Classification and Categorization of Mathematical Knowledge. In: Autexier, S., Campbell, J., Rubio, J., Sorge, V., Suzuki, M., Wiedijk, F. (eds.) Intelligent Computer Mathematics—Proceedings of 7th International Conference on Mathematical Knowledge Management MKM 2008. Lecture Notes in Computer Science LNCS/LNAI, vol. 5144, pp. 543–557. Springer-Verlag, Berlin, Heidelberg (Jul 2008)
11. Řehůřek, R., Sojka, P.: Software Framework for Topic Modelling with Large Corpora. In: Proceedings of LREC 2010 workshop New Challenges for NLP Frameworks. pp. 45–50. Valletta, Malta (2010), software available at `http://nlp.fi.muni.cz/projekty/gensim`
12. RFC3066: Tags for the identification of languages (January 2001), `http://potaroo.net/ietf/idref/rfc3066/`
13. Sylwestrzak, W., Borbinha, J., Bouche, T., Nowiński, A., Sojka, P.: EuDML—Towards the European Digital Mathematics Library. In: Sojka, P. (ed.) Proceedings of DML 2010, pp. 11–26. Masaryk University Press, Paris, France (Jul 2010)
14. Tufte, E.: Envisioning Information. Graphics Press (1990)