

## Telling Visual Stories about Document Collections from Machine Readable Languages

Many human activities revolve around text documents. People make sense of documents, analyze, synthesize, summarize, and do many other activities. Because of text, these activities are time-consuming, labour-intensive, and cognitively-taxing. Current information retrieval systems (such as search engines and online catalogues), however, do not make these activities easier to perform. These information systems allow users only to search, browse, preview, open, and read documents. All other interactions such as grouping, filtering, extracting, annotating, and many other take place outside of the information systems.

One way to reduce the complexity of activities is to visualize machine-readable languages. Machine-readable languages facilitate document retrieval (Library of Congress, 2009; Taylor, 2000). They describe basic information about documents such as titles, subjects, languages, years of publication, places of publication, and some other information. In visualizations machine readable descriptions of multiple documents can be grouped, filtered, sorted, and combined in many other ways on graphical representations. Examples of graphical representations include Tag Clouds, geographic maps, pie charts, timelines, Kohonen maps, scatter plots, and other. Representations are not perfect; they have pros and cons (Peterson, 1996; Markman, 1998). Each representation can support only a limited number of tasks. For example, maps can tell where documents were created; Tag Clouds can help users identify frequencies of tags or subjects; pie charts show proportions of languages or some other descriptors. Despite these limitations, taken together, representations can tell stories about collections of documents and serve as useful summaries of document collections.

In information visualization, however, visualizations consist not only of representations, but also interactions (Spence, 2007). Interactions are low-level actions that users perform on representations followed by the response that they get (e.g., filtering, annotating, extracting, grouping, etc.) (Sedig, 2009). In this sense, visualizations should not only represent machine readable languages but they should also allow users to change representations and engage in interactive discourse. Specifically, users should be able to *ask* “What if?” questions and see how representations respond. For example, users should be able to communicate with representations in the following way: What happens if I remove documents published before 1990? or What if I extract documents published in XX location?

The purpose of this presentation is to explain not only the theory behind visual storytelling, but also to demonstrate how visual storytelling from machine readable languages can help users understand and make sense of text collections. The presentation will be supplemented with a demonstration of a map-based visualization prototype VIsual COLlection Explorer (VICOLEX), which was designed to help users make sense of a collection of documents about local history of Ukraine from the Library of Congress Catalogue (Buchel & Sedig, 2011).

## References:

Buchel, O., & Sedig, K. (2011). "Extending map-based visualizations to support visual tasks: The role of ontological properties." *Knowledge Organization*, 38(3), pp. 204-229.

Library of Congress. (2009). "Understanding MARC bibliographic: Machine-readable cataloging." Available at: <http://www.loc.gov/marc/umb/>.

Markman, A.B. (1998). *Knowledge representation*. Mahwah, NJ: Lawrence Erlbaum Associates.

Peterson, D. (Ed.). (1996). *Forms of representation: An interdisciplinary theme for cognitive science*. Exeter, UK: Intellect, Ltd.

Sedig, K. (2009). "Interactive mathematical visualizations: Frameworks, tools, and studies." In E. Zudilova-Seinstra, & e. al. (Eds.), *Trends in Interactive Visualisation: A State-of-the-Art Survey* (pp. 112-132). Springer Verlag.

Spence, R. (2007). *Information visualization: Design for interaction*. 2<sup>nd</sup> ed. Saddle River, NJ: Prentice Hall.

Taylor, A. G. (2000). *Wynar's introduction to cataloging and classification*. 9th ed. Englewood, Colorado: Libraries Unlimited, Inc.