# SOMLib-based Interfaces to Digital Libraries

The SOMLib Digital Library Project aims at the development of a digital library system supporting intuitive, user friendly browsing of document collections by combining the benefits of conventional library organization with the possibilities offered by digital collections. It is based on the self-organizing map (SOM), a popular unsupervised neural network, used to organize documents by content. Different visualization methods, including map-based metaphors of islands, or bookshelf metaphors, address different user needs. The result from this is a map of a document archive, where documents are grouped according to their semantic similarity, i.e. documents on similar topics are located close to each other on the map.  The map may be depicted as a cartographic map, where islands (topic clusters) float in the sea, or using a bookshelf metaphor, where different parts of metadata information are depicted in a graphical way.

Different methods for labeling the resulting maps, ranging from keyword selection methods to simple NLP tools, automatically  provide descriptive labels of the topic clusters, characterizing what to find where on the map. This allows users to orient themselves on the map (Figure 1, 2, 3). The system serves as an additional access method to large document collections, complementing traditional metadata or full-text search. The main goal of the system is to provide users with a possibility to quickly obtain an overview of which information is present in a document collection (something that is impossible with a traditional query interface). It furthermore makes search results more intuitively comprehensible by, for example, showing in which sub-section of the collection (i.e. in which topic areas) hits have been found. Last, but not least, it allows to locate documents that do not contain the specific keyword searched for, but which – due to their topical similarity – are located in the same topical cluster.

It can be incorporated into basically any traditional DL system to complement traditional search interfaces. It offers a novel, exiting interface, allowing users to orient themselves in the collection, getting an overview of their holdings. Current prototype studies are in progress for e.g. the Austrian Research Documentation (Figure 1), or an integration into the Greenstone DL system (Figure 4), and may also be used for audio digital libraries, organizing music by sound similarity based n frequency spectra analysis, both on desktop as well as portable/mobile devices (Figure 5).
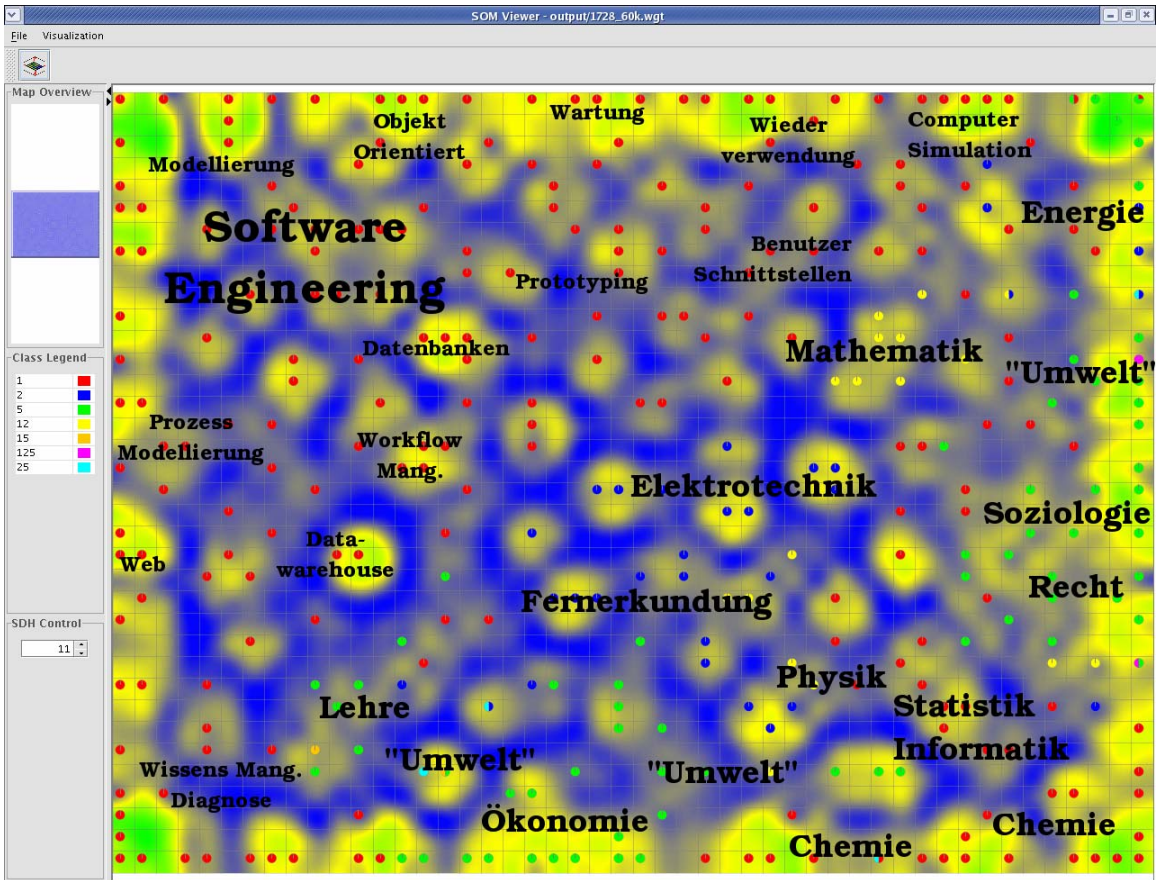
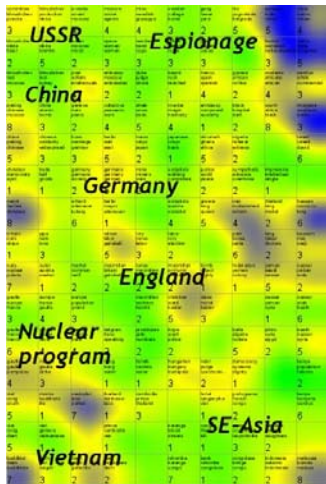Figure 1: SOMLib view of the Austrian Research Documentation (subset)

Figure 2: SOMLib view of TIME Magazine Data, LabelSOM and abstracted labels
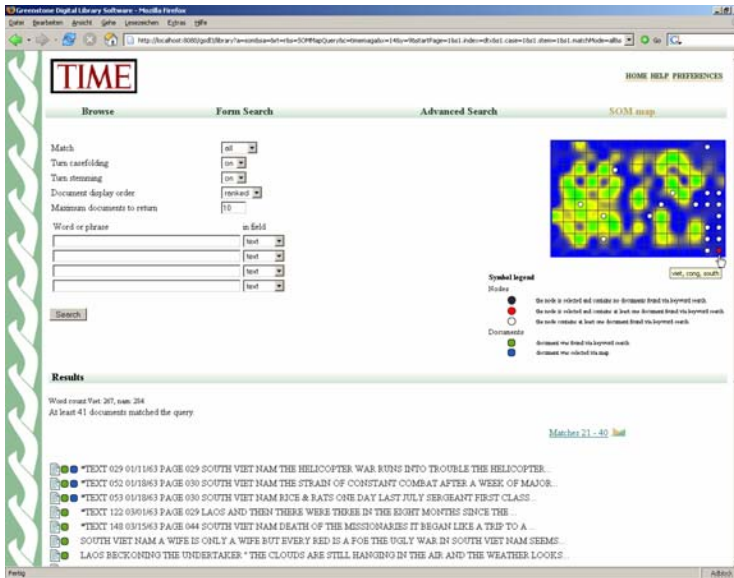
Figure 3: LabelSOM labels (detail)

**Figure 4: SOMLib integrated in Greenstone DLS**



**Figure 5: MobileSOMPlayer on a PDA**

Contact Info:
*Andreas Rauber*
*Vienna University of Technology*
*e-mail: rauber@ifs.tuwien.ac.at*
*URL: http://www.ifs.tuwien.ac.at/~andi*