

Title :	Integration of automatic indexing system within the document flow in grey literature repository
Author(s) :	Jindřich Mynarz and Ctibor Škuta NTK, National Technical Library (Czech Republic)
Session Three Paper 2	

The Web empowered the authors of grey literature to publish their work on their own. In case of self-published works their author is also their indexer. And because not many of the grey literature authors are professional indexers, this may result in poor or no indexing.

Even though the Web made publishing easier, indexing is still hard. Nevertheless, we believe that the Web technologies and machine learning algorithms may help to reduce the cognitive overhead involved in indexing, and make it eventually as easy as publishing on the Web is.

To help overcome the issue of quality and consistency of subject indexing automatic indexing systems can be used. Given enough full-texts already equipped with the terms from the controlled vocabulary that is to be used, machine learning algorithms can be employed.

Our aim is to provide human-competitive automatic indexing to authors and producers of grey literature. We demonstrate how an automatic indexing system based on machine learning can be integrated into the document flow in an open source digital repository of grey literature. We build upon open source tools and a controlled subject heading vocabulary available in an open standard format.

We will be using Maui Indexer as automatic indexing system, CDS Invenio as digital repository software, and Polythematic Structured Subject Heading System (PSH) as knowledge organisation system. Both Maui Indexer and CDS Invenio are open source, and CDS Invenio's modular architecture makes it possible to extend it with new functionality. Maui Indexer works with controlled vocabularies expressed in Simple Knowledge Organisation System format in which the PSH is available.

From these components combined we will try to put together a solution for automatic indexing aimed at grey literature in the Czech language environment. Maui Indexer is domain and language independent so it is possible to adapt it for the field of Czech grey literature. The document samples we will test on will come from the National Repository of Grey Literature which is maintained by National Technical Library of Czech Republic.

In the end, we will discuss integration of the automatic indexing component from the user perspective and sketch out how the user can interact with it through the user interface. Also we will provide details around the actual realization of the proposed system. The conclusion will deal with the evaluation of benefits of the implemented system for grey literature authors.