

Collaborative Building of Controlled Vocabulary Crosswalks

ABSTRACT: The purpose of this paper is to present JOnto, a framework for annotating resources, in a wider context of Semantic Social Collaborative Filtering idea. I will also show what kind of problems does this combination cause and how they shall be solved.

One of the main features of classic libraries which was inherited by the Semantic Digital Libraries is describing resources with various metadata. Many different types of Knowledge Organization Systems are used to perform this task – ranging from controlled vocabularies and taxonomies to semantic nets and ontologies. JOnto project was carried out to make this process easier, as far as taxonomies are concerned. It allows users to select some desired annotations from classification schemata like WordNet, DMoz, DDC and PKT or write them manually, and use them to describe a specific resource. All work is done using simple and clear AJAX-based web interface.

However, the important drawback of this solution is that there is no support for swift taxonomy to taxonomy transition or conversion. This means that a resource has to be described independently in each taxonomy, even if the concepts taken from them are very similar. It makes a work with JOnto pretty uncomfortable.

Another disadvantage is not exploiting the whole potential of user community. Great success of so called Web 2.0 realized everyone that the future belongs to the software which is community-aware above all. To respond to this trend, digital libraries introduced an idea of Semantic Social Collaborative Filtering (SSCF). It's foundation is to treat each user not like a consumer of the content delivered, but rather as one of its creators. SSCF enable users to annotate and evaluate resources and assign them to previously created bookmarks, as well as exchange these bookmarks with their friends. Then it utilizes this user created content, for example, to enhance searching effectiveness.

Both, allowing conversions between different taxonomies and enhancing user engagement in work with JOnto, are the basis of the master thesis written at Gdansk University of Technology entitled Collaborative Building of Controlled Vocabulary Crosswalks. First task of it would be extending JOnto with various controlled vocabularies, apart from those already implemented. Then a testing environment should be set to collect large amount of data from the community, i.e. users annotations. Apparently the best way to resolve the problem with conversions between taxonomies is to create a proper algorithm based on user delivered data. The reason is that similar contents will be described by equivalent concepts taken from different taxonomies. The more annotations collected, the more precise the result of this association will be.

After implementing the conversion algorithm in JOnto it will be possible to include interesting and useful additional functionalities. Most of all – automated suggestions of concepts similar to those just typed (from the same or different taxonomies). Another improvement would be autocompletion of typed words based either on WordNet or on annotations written by the user (possibly also his friends) so far.

As you could realize from this paper, the process of annotating resources in digital libraries, although very important, still can be improved in many fields. Referenced master thesis is assumed to solve as least some of them.