Integrating Semi-formal Knowledge Organization Structures

Gilles Falquet, Luka Nerima, Claire-Lise Mottaz Jiang, Jean-Claude Ziswiler

Centre universitaire d'informatique CUI, University of Geneva

24, rue du Général-Dufour, 1211 Geneva 4, Switzerland

{falquet, nerima, mottaz, ziswiler}@cui.unige.ch

ABSTRACT

The last years, we had been working on hyperbook structures to build digital libraries. A hyperbook is made of a domain ontology containing the most important concepts of the field or subject in question and of information fragments linked to the ontology's concepts. Fragments are text junks and serve primarily to define a concept, but they also can describe different aspects of the concept or can contain examples, references, etc. Optionally, links between fragments and concepts can be typed. The digital library is build by alignment of the different hyperbook ontologies that identifies equivalent and similar concepts. The aim is to create an extended view of each hyperbook in the form of a virtual document that provides readers with supplementary information found in the other hyperbooks, like additional examples, term definitions, more detailed or more general information, etc.

Much in the spirit of Marshall and Shipman outlining that «The difficulty of knowledge acquisition, representation and reasoning has a long history of being underestimated», the aim of inventing hyperbooks is to build a knowledge organization structure that is as easily to construct as low structured KOS (for instance glossaries or metadata annotated models like learning objects), but has a stronger semantic structure that can be used for the integration process.

Many research communities proposed to write full-fledged ontologies that result in a KOS with a strong semantic structure. With such kind of ontologies, it might be possible to process logic reasoning, which might become more difficult with a hyperbook structure that just contains a small domain ontology and textual fragments. On the other side, ontology built according to specifications like the ones proposed in the RDF/OWL family are time-consuming to construct and suitable only for homogenous domains. For instance, it might be possible to create an OWL ontology describing all elements of a house, but it seems nearly impossible to write an ontology about the United Nations under OWL specification.

Anyway, we found evidence through different example that the hyperbook structure is suitable to integrate hyperbooks into a digital library of hyperbooks. But concepts must be linked to representative fragments that either define, or describe, or show examples, or simply refer to the concept.

Last year, we tried to integrate two hyperbooks about agriculture politics made by domain specialists. A complete automatic integration approach allowed sorting out relations indicating equivalent and similar concepts.

Last winter, we let graduated students of a computer science course model hyperbooks about the topics of the course. We found a clear difference when comparing the students' hyperbook with the one build by domain specialists. Students found appropriate concepts, but finally didn't take a lot of care to select the fragments. This probably because we provided them with slides out of the course presentation and with selected publications around the course topics, so fragments we not easily to find and to write. Domain specialists can take advantage of documents of their daily work, so it might be easier for them to create well-done hyperbooks. We conclude that hyperbook creation is fastest when there exists already adequate material in a knowledge base that easily can be fragmented. Particularly, glossaries or similar KOS might be the best starting point for the construction of hyperbooks.

We propose the following integration process to assemble the digital library: First, we compute semantic similarities between concepts of the hyperbook. The mapping approach relies on both conceptual structure comparison (based on word matching, semantic neighbourhood matching and the positions in the "is-a" and "part-of" hierarchies) and fragment comparison. The existence of semantic similarity between fragments increases the concepts' similarity.

Secondly, the weighted similarity links are used to generate a reading interface of an extended hyperbook by presenting the book content within its semantic context.

We built a prototype to generate virtual documents of formal hyperbooks and to apply filtering, organization and assembling mechanisms. To avoid information overflow by attaching any kind of links to the initial hyperbook, we designed a graphical user interface generator that produces expand-in-place links for larger textual fragments that are showed to users after activating the corresponding link.

In the example with graduated students, it was more difficult to find appropriate similarities in a fully automatic integration process as with the hyperbooks built by domain experts. In this case, we need an alternative way to validate the determined relations.

In social navigation and social bookmarking when a user follows a link or bookmarks a page, this action increases the score (or weight) of this link respectively page. So each one can benefit from everyone's experiences, discoveries, etc. We use a similar principle to establish more reliable semantic relations between ontologies if concept description is not well done. It process as follows:

When a user reads a hyperbook, the system automatically proposes links to (or expansion in place of) fragments from other hyperbooks according to the alignment described above.

Then, the user can do three things:

1)   Follow (respectively expand) the link

2)   Definitely accept the link

3)   Definitely reject the link

1 and 2 reinforce the value of the similarity of the concepts involved in the link inference (2 is stronger), 3 weakens the similarity value. The underlying hypothesis is that the user will immediately see what is completely irrelevant.

Future work consists of adding functions to "create your own book". Such a tool would contain link proposals that the user can include or not, or introduce a very simple "language" to express inter-book queries to create dynamically derived fragments, e.g. "Find examples of this concept in other books" or "find more descriptions of this concept". Here again, the user should be able to accept or reject answers