Semantic Bookmarking for Reference Management

Alex Logvynovskiy, Mohammad Dastbaz, Chris Imafidon

School of Computing, Information Technology and Engineering,
University of East London
Docklands Campus
University Way
London E16 2RD
{a.logvynovskiy,m.dastbaz,c.imafidon}@uel.ac.uk

1 Abstract

Finding publications relevant to research is crucial to its success. With the number of scientific publications ever-increasing, a well-annotated and structured bibliography can be seen as a highly-detailed map of the *publication space*. Social bookmarking applications have shown considerable advantage over traditional reference applications in creating and managing such 'publication maps' by means of personalisation, recommendation and global availability. On the other hand, advances in semantic web and linked data are increasing the need for means of *guided semantic navigation* that the existing bookmarking systems lack.

As a solution, we propose a collaborative reference management framework based on concept of semantic bookmarking. The primary aim of the framework is to separate bookmark management functions of the underlying citation network into a set of structure and semantics modules in order to improve navigation and provide richer explanation output to the user.

The rationale behind the *separation of the structural and semantic layers* can be outlined as following: the structure of the citation network changes differently to its semantics; structure-related operations can be optimised by means of a purposefully constructed index; semantic relationships are more dynamic and normally require in-context evaluation on a per-request basis; both structural and semantic layers can be comprised of several modules.

Semantic pertinence module of the framework deals with the matching of a query against the associations of the network. The output is a subset of the network such that its elements are relevant to the query. In its simplest version, the semantic matching function merely verifies the existence of a publication-tag association. A more complex version of the semantic function would exploit logical reasoning over the tag ontologies in order to retrieve additional semantic relationships, eg. hierarchical structure among the tags. It might also interact with the structural connectivity module to obtain information about ancestor-descendant relationships without the costly semantic traversal.

Whenever the outcome of the semantic matching passed into the structural connectivity module, the results are further clustered according to their mutual relationship

within the citation network. A cluster is an equivalent of a graph component within the network, i.e. a group of publications connected with each other directly or through intermediary references. The number of components among the results, therefore, corresponds to the number of the clusters.

The structural connectivity module rates clusters as well as publication nodes within the clusters in order to achieve a relevant recall. Rating by size is chosen on the grounds of simplicity, but can be replaced with a more sophisticated ranking algorithm without affecting the rest of the framework. The ratings of the publications within the cluster are determined by the structural relatedness. Similarly to the semantic relatedness function, structural relatedness functions can be more complex and may interact with the semantic module in order to improve the recall.

An example of the relatedness module results is given in Fig. 1 below.

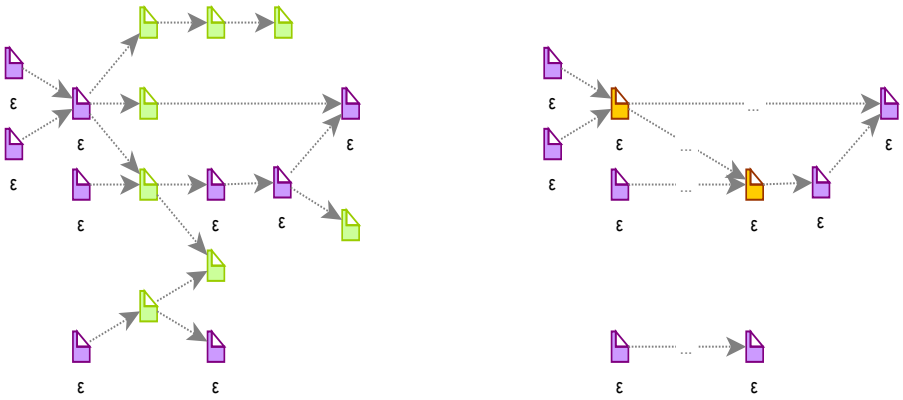

Fig. 1. Relatedness module results after semantic (left) and structural (right) matching.

The graph on the left represents a sample network of publications with its nodes selected by the semantic pertinence module highlighted in purple. These are then passed to the structural connectivity module (shown on the right) where it is split into clusters according to the structural index. In the current example, there are two clusters, the larger being rated as more relevant due to its size. The publications within the cluster are rated according to the number of components they produce if removed from the cluster (highlighted in the example).

In order to evaluate the proposed framework, we developed a prototype system that implements its modular architecture. The main features of the system are visualisation according to underlying structure, navigable results, uniform search/browsing interface and explained relevance**.**

As a future work, we plan to expand the ontology sets for tags, bookmarking as well as their alignment in order to provide better support for ‘weaker semantics’. We also consider experiments with several structural and semantic modules combined by a weighting function for clustered output. More elaborate functions based on the relevance of individual publications within the cluster are to be explored. Similarly, additional ranking measures within the cluster will be investigated.