Browsing Similar or Related Data Entities by Breadth-First Search in the Semantic Web

Karol RÁSTOČNÝ *

Slovak University of Technology
Faculty of Informatics and Information Technologies
Ilkovičova 3, 842 16 Bratislava, Slovakia
xrastocny@stuba.sk

The difficulty of finding relevant data on the Web is increasing as web repositories grow. Therefore, we propose an approach for browsing the Semantic Web, which can help users find relevant results, i.e. how to find data in the Semantic Web, and how to browse similar and/or related data entities.

We extend the faceted browser Factic[1], which displays results in thumbnail matrix/list view with an additional view where results are categorized in hierarchical clusters. This view helps users to browse large numbers of results in a more organized way. We propose an approach to hierarchical cluster creation and labelling using semantic similarity computed from metadata.

Our clustering approach is based on Bordogna's and Pasi's hierarchical-hyperspherical divisive fuzzy c-means clustering method [1]. This approach has good results, but there are performance issues. We see a problem in the process of identification of optimal count of clusters in the first level of clustering. We propose an approach in which we make an interval approximation of the optimal number of clusters. As optimal number of clusters we select the best based on the quality function defined by Mecca et al. [3]. Because this quality function is based on the removal of edges with maximal length from a minimum spanning tree, where vertices represent clustered items and length of edges represents similarity, it may not return the optimal number of clusters for fuzzy clustering. To address cluster label creation, we propose a novel method based on common facets in the similarity vector, where we proceed from the lowest level of clusters to the topmost level.

While faceted browsers help users to find information they offer less support for the exploration of resources similar to an already found result. We address this via view-based search within the Semantic Web using navigation in a 2D graph. As with other tools for navigation (e.g. Paged Graph Visualization [2]), the process starts with one central node, which represents the initial result and some nodes around it

* Supervisor: Michal Tvarožek, Institute of Informatics and Software Engineering
[1] http://mirai.fiit.stuba.sk/meia

Spring 2010 PeWe Workshop, April 18, 2010, pp. 15–16.

representing its facets. After that, users can browse the graph by expanding nodes representing facets. To prevent the graph from becoming unclear due to node expansion, which can result in many (irrelevant) nodes, we propose tools based on result clustering, facet marking and the hiding of nodes and graph components.

Clusters are created from results that have the same facets displayed in the graph and behave in the same way as results (see Figure 1). This means that clusters are connected to facets and users can display their facets. We give users the ability to filter new results after expanding nodes by marking facets as wanted, unwanted and visible. This marking sets if new results new results should, may or are not allowed to have direct connections to marked facet. The hiding of nodes gives users the opportunity to choose which nodes they do not want to see anymore. The hiding of some nodes in the graph can segment the graph into several components, thus enabling the hiding of whole graph components instead of sequentially hiding individual nodes.

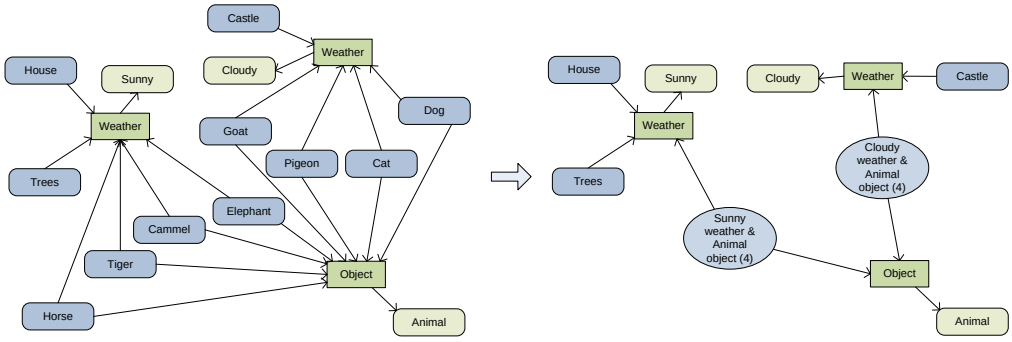

Figure 1. Example of cluster creation in a graph.

Acknowledgement. This work was partially supported by the Cultural and Educational Grant Agency of the Slovak Republic, grant No. KEGA 345-032STU-4/2010.

References

[1] Bordogna, G., Pasi G.: Hierarchical-Hyperspherical Divisive Fuzzy C-Means (H2D-FCM) Clustering for Information Retrieval. In: *Proc. of the 2009 IEEE/WIC/ACM international Joint Conference on Web intelligence and intelligent Agent Technology*, IEEE Computer Society, Washington (2009), 614-621.

[2] Deligiannidis, L., Kochut, K.J., Sheth, A.P.: RDF data exploration and visualization. In: *Proc. of the ACM first workshop on CyberInfrastructure: information management in eScience*, ACM, New York (2007), 39-46.

[3] Mecca, G., Raunich, S., Pappalardo, A.: A new algorithm for clustering search results. *Data & Knowledge Engineering*, Vol. 62, No. 3 (2007), 504-522.