**Social Networks and Application Tools International Workshop SoNet-08,
Skalica, Slovakia, September 19-21, 2008, pp. 5-12**

# Navigating Networked Data using Polycentric Fuzzy Queries and the Pile UI Metaphor

Alexander Troussov[1], John Judge[1], Mikhail Sogrin[1], Cristian Bogdan[2], Pär Lannerö[2], Henric Edlund[2], and Yngve Sundblad[2]

[1] IBM Dublin Lab, Building 6, IBM Technology Campus, Damastown Ind. Est., Mulhuddart, Dublin 15, Ireland
{atrousso, johnjudge, sogrimik}@ie.ibm.com
[2] CSC-School of Computer Science and Communication, KTH – Royal Institute of Technology, SE-10044 Stockholm, Sweden,
{cristi, pbb, hed, yngve}@csc.kth.se

**Abstract.** The proliferation of Web 2.0 and Enterprise 2.0 technologies has lead to the emergence of massive networks connecting people and various digital artefacts. The efficiency of human navigation in such networks depends on the availability of suitable user interfaces powered by an "intelligent" backend which provides guidance and recommendations. In this paper we describe how the "pile" based GUI (Graphical User Interface) called Nepomuk Simple and the IBM graphmining library Galaxy can be used for such guided navigation through the Personal Information Model (PIMO) ontology in the scenario of social semantic desktop as pertaining to the EU 6th framework project Nepomuk.

Firstly, we describe a method for graph based related item recommendation. The initial data for recommendations which allow browsing from one single thing to another, one-to-one correspondences might be properly treated as an egocentric query. Following this logic, initial data for recommendations which allow "many-to-many", browsing, must be treated as a polycentric query. Secondly, we present Nepomuk Simple a GUI for managing "piles" of desktop resources. The contents of a pile can be processed by engines like Galaxy, and recommendations can be presented to the user so that she can modify the "pile" used to drive the process. We conclude that the process of incremental modification of a "pile", assisted by recommendations resulting from fuzzy polycentric queries, might be formally viewed as the process of browsing and exploring data provided by networks not by browsing from a single object to another, but by dealing with several objects at the same time.

**Keywords:** Pile UI, spreading activation, social web, semantic web

# 1 Introduction

The proliferation of Web 2.0 and Enterprise 2.0 technologies has lead to the emergence of massive networks interconnecting people and various digital artefacts. The composition of these networks frequently includes nodes and links which are more related to the technologies underpinning the functioning of these networks, than to the potential interpretation of these networks by humans. The content of these networks is influenced by premises which encourage utilising data before providing structure, the result being that often the content of these networks is usually of mixed quality.

The efficiency of human navigation in such networks depends on the availability of suitable user interfaces powered by an "intelligent" back end which provides guidance and recommendations based on soft computing methods. In this paper we describe how the "pile" based GUI (Graphical User Interface) called Nepomuk Simple and the IBM library Galaxy [1] can be used for such guided navigation through the Personal Information Management Ontology (PIMO) in the scenario of social semantic desktop as pertaining to the EU 6th framework project Nepomuk.

Firstly, we describe a method for graph based related item recommendation. It is a problem which is well addressed by methods of soft graph clustering and fuzzy inference. The initial data for recommendations which allow browsing from one single thing to another, one-to-one correspondences, might be properly treated as an egocentric query on a large interconnected multidimensional network. Following this logic, initial data for recommendations which allow many-to-many, browsing, must be treated as a polycentric query on the network. We present the IBM Galaxy library as the reference implementation of unified API for processing such queries.

Secondly, we present Nepomuk Simple, a GUI for managing "piles" of desktop resources. The contents of a pile can be processed by engines like Galaxy, and recommendations can be presented to the user so that she can modify the "pile" used to drive the process. In this way, various kinds of recommendations can be presented to the user so that she can improve the "pile", making it more suitable to the task at hand, and/or easier to view and understand.

We conclude that the process of incremental modification of a "pile", assisted by recommendations resulting from fuzzy polycentric queries, might be formally viewed as the process of browsing and exploring data provided by networks not by browsing from a single object to another, but by dealing with several objects at the same time maintaining a "pile" of current objects in focus. Such an "ambient navigation" tool might be used for exploring various massive multidimensional networks, which extends beyond the scope of scenarios explored in the EU 6$^{th}$ framework project Nepomuk where these ideas have been developed.

# 2 Pile UI Metaphor

The most pervasive computer interactive environments are now based on the desktop metaphor, which imitates traditional office environments (see, e.g. [2]). "Real" desktops usually have piles of things on them where we have (consciously or

unconsciously) grouped together items which are related to each other or to a task. The so called “Pile” UI imitates this type of organisation (for example, IBM Activity Explorer). [2], [3] and other papers indicate that such casual organisation of information helps to avoid premature categorisation and reduces the retention of useless documents.

Nepomuk Simple is a Pile-based UI, which is under development by KTH for the EU project Nepomuk. The goal of Nepomuk Simple is a knowledge workbench with an easy-to-use interaction method. The intended users are knowledge workers that should be able to work with and understand semantic information. These users can browse and edit Nepomuk knowledge structures. Workflow is facilitated by Piles and (later) Templates. Some information is displayed in specialized Views, including a timeline and a map.

Nepomuk Simple is based on the knowledge created during field studies performed early in the Nepomuk project. The strategy for this implementation is to make “the simplest thing that could possibly work” towards its priorities. This is a work in progress, both conceptually and implementation-wise. For this we have chosen to mainly use a task oriented scenario where a fictitious user of the system, Claudia, is organizing a business trip. The scenario is referred to in the Nepomuk community as “Claudia is organising her trip to Belfast“. Putting the Pile metaphor of Nepomuk Simple in such a scenario allows us to make a semantic workbench which is useful in a realistic situation. Much effort has been put into also making it useful as a generic interface to semantic components of different sorts.

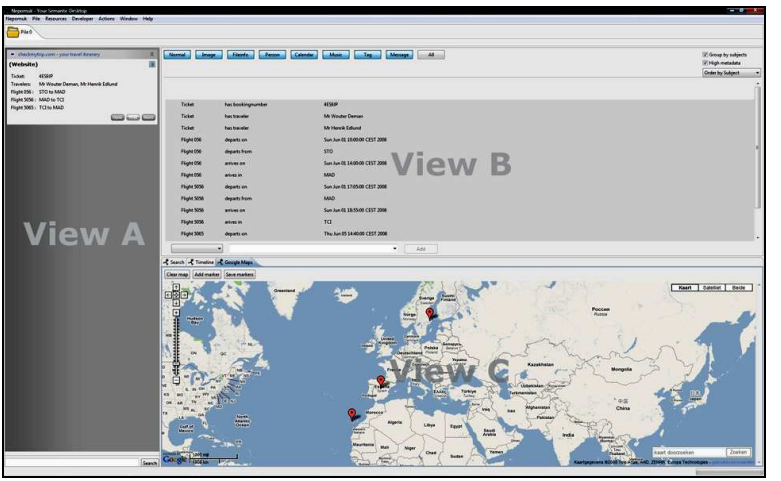


**Fig. 1.** The overall concept of the Nepomuk Simple Pile UI

The application is divided into three parts:

*View A – The pile*. Here the user can take any resource from the desktop by drag-and-drop to put into the pile. This is derived from the scenario part of when Claudia has booked everything and needs some sort of bundle to keep all the relevant resources (email, tickets, bookings, etc.) in one context.

*View B – The semantics of the content in A is extracted and presented here*. This view gives Claudia the opportunity to see vital data from all parts of the Belfast trip pile in one place. She can also make annotations to the pile, and edit the semantics. By drag-and-drop she can create new piles based on one or more of the resources shown. This is somewhat similar to browsing the web using tabs.

*View C- Different views of the resources in A* combined with the annotations in B enables Claudia to look at her trip by a timeline or map view. The intention with this field is to keep it open for custom made plug-ins.

Currently, the metadata shown in the view B is only the metadata that has been added to the triple store. In this paper we describe ongoing work for providing recommendation of related items using IBM Galaxy library.

## 3 Related Item Recommendation as a Problem of Processing Fuzzy Polycentric Queries on Multidimensional Networks

The proliferation of Web 2.0 technologies has lead to the emergence of massive networks connecting people and various digital artefacts. Enterprise 2.0 usually adds new dimensions and new connections (for example, since identity management on the intranet is simple, it is easy to add additional links from a corporate remake of Facebook to a corporate remake of Delicious).

Communication networks are of particular interest to business and security applications. Lexico-semantic resources (such as WordNet or medical ontologies) are an important part of knowledge-based methods in language engineering. Semantic web relies on the use of ontologies, and the objects of piles in the Nepomuk simple are objects from Personal Information Management Ontologies (PIMO).

Also all of the above mentioned domains might probably benefit from the use of activation spread methods as a scalable efficient solution the search and navigation, tolerant to errors and inconsistencies in the data.

We believe that recommendation of a related items for a set of nodes on a network is a problem which can be addressed by a blend of fundamental graph-mining techniques like soft clustering and fuzzy inferencing; moreover, the user herself frequently would not be able to specify what exactly she would like to have as a recommendation (done in the mode of a predictive search).

On one hand, one can think that related items are those items which are connected to all of the items in a query by task dependent links of various types which represent the relationship(s) between those items. For example, if a query is a set of nodes corresponding to the concepts of “researcher”, “Karlsruhe”, “cat”, etc. Related items could be, for example, all the nodes representing people with the properties: “is a” researcher, lives or works in “Karlsruhe”, “owns” pets some of which are “cats”, etc. However, there is a danger that the query will produce too many results, so that ranking is needed. Also, the more nodes and dimensions that are added to the query, the likelihood that the query returns too few results will increase. Therefore the problem of related item recommendation should be treated by fuzzy logic, which allows easy parameterisation so that the logical AND can be easily and smoothly converted to the operation OR for query relaxation. Instead of retrieving ALL the

nodes representing people with properties: “is a” researcher, lives or works in “Karlsruhe”, “owns” pets some of which are “cats” etc; the query processor retrieves a sufficiently long list of people which have some of the required properties ranked according to how many properties are satisfied.

On the other hand, we can suggest that related items are those nodes on the network which minimise the weighted distance to all nodes in the query, and therefore related items are those nodes which might be considered as centroids of clusters induced by queries. Since partitioning of all the nodes into clusters is not needed, this is soft clustering.

## 4 Using Related Item Recommendations within the Pile UI Metaphor

In this section we describe the generic architecture of related item recommendation on networked data and ongoing work on integrating this architecture with the pile UI metaphor.

**Processor of fuzzy polycentric queries.** In our work we use IBM’s Galaxy library as a reference implementation for processing fuzzy polycentric queries. In this scenario a query is a real-valued function on nodes of the network. The processor uses this function as initial “activation” on the network, and then propagates this activation to other nodes using the links that connect the nodes to form the network. The output of the processor is a list of the most activated nodes ranked according the level of activation.

**Setting parameters of the query processor**. In our current work we focus on how to create a user friendly interface for setting parameters for the spread of activation module in the Galaxy library.

**Query generator**. In the pile UI metaphor the query might be created by taking all the nodes of the network corresponding to the content of the pile and by assigning the activation 1.0 to all these nodes. In our future work we are looking at how the query might be modified to indicate the context of the query and user focus.

**Post processing**. Applications which utilise the architecture described here can set additional parameters for query processing. For instance, in applications for community-based metadata recommendations in collaborative bookmarking systems, the query processor will return nodes corresponding to resources, tags, users, and instances of tagging. Post processing is needed to filter out all nodes which are not tags.

**Explanatory module**. The explanatory module applies the specifics of the applications to the results generated by the query processor. For instance, when used in collaborative bookmarking systems, this module is able to generate explanations for the various instances of tagging, and present these instances to the user.

### 4.1 Strategies of applying Spread of Activation to the Pile

We have found a number of strategies to use spread of activation in conjunction with Nepomuk Simple and its Pile metaphor. We will introduce them with a symbolic equation

*Dictionary + input data + configuration → Polycentric search results*

For example the current Galaxy library demo configurations work like this:

*XML nodeset +text+ link resistances → ranked list of related nodes*

In other words Galaxy takes a number of XML nodes as its dictionary, and a text as input, and its configuration consists of activation resistances of various graph edges that link the nodes. As a result the Galaxy demo displays a ranked list of nodes related to the polycentric query members.

We have found a number of ways to combine data that we have available in the Pile that allows us to put Galaxy to full use in the Nepomuk Simple context. We use data on any of the left-side members of the above "equation". We will exemplify some of these strategies below.

*RDF storage+ text + link resistances → ranked list of related URIs*

In this case, we take the entire Nepomuk RDF storage as dictionary, and given a text we are able to point the most relevant objects (by way of their URIs). A problem occurring with this approach is that sometimes the labels from the RDF store are not very meaningful, hence their use in the dictionary may not always give good results.

*PIMO nodes +text+ link resistances → ranked list of related PIMO concepts*

In this variation, we only analyze the PIMO parts of the RDF repository. In this case we should have better labels, hence a higher quality of results. One variant of this is to still derive the dictionary from the entire RDF storage but sort out only the PIMO concepts for the output.

*RDF storage+ list of URIs + predicate resistances → ranked list of related URIs*

Unlike in all the variants above, in this case we do not have text in the input, but a list of URIs, i.e. all, or some of the Pile members. Also for configuration we assign activation resistances to various kinds of predicates of items in the RDF storage.

*RDF storage, several ontologies + list of URIs + predicate resistances → ranked list of related URIs*

In this complex variant we are trying to make use of most of the information we have at hand. Thus we consider not only the URIs of the resources but also their text (where available, e.g. for text file URIs, whether they are plain text or some other content type) and their metadata that Nepomuk-simple currently shows in View B.

### 4.2 Mutual Collaborative Spread of Activation

The current spread of activation usage scenarios assume a lone user looking for interesting issues based on some focus items. In collaborative settings (as most knowledge workers find themselves working in) there may be items that other users

would like (to some extent) our user (or some of their co-workers) to look at. In such a case, these items should get even more powerful activation for all searches made by the user, and maybe also if searches are not made explicitly, i.e. the user would get a notification on items that their co-workers changed and were searched for in the past, or are simply related to some members of their Piles.

The theoretical framework for such a spread of activation, which we could call Mutual collaborative SoA, is described in [5]. The set of items that are in the user interest (hence set the starting point for a polycentric query) are in the user *focus.* The focus is thus covered by the current spread of activation scenarios. The set of items that others (co-workers) would like the user to look at are said to have a *nimbus* towards the user. The nimbus is not considered by the current SoA scenarios, however, it is easy to notice that the nimbus can spread through the network exactly as the focus does, using the same Spread of Activation methods.

Focus is a bit different from a simple polycentric search in that it has *persistence* dimensions. The focus models the “interests” of a user, which can be extracted from the explicit searches the user makes, but also from other sources, such as the Piles created, the objects looked at, etc. Nimbus also has similar long-term properties. Also important is the *time evolution* of focus and nimbus: for example the nimbus of a “meeting” calendar item will depreciate a lot after the meeting time. In spread of activation terms, this will lead to much lower (or zero) initial activation.

### 4.3 The Experimentation Playground

While our current pile implementation (Nepomuk Simple) seems promising as a tool for organizing desktop resources of all kinds, and the Galaxy library can provide relevant suggestions on related resources, Nepomuk Simple is not an adequate tool for *experimenting* with the recommendations. Whenever a user interacts with a pile, the modified pile is saved in the system. So if a user wants to keep her existing pile, but still see what recommendations would come up if she did some or other modification, the user must first make a copy of the existing pile. The experience from the web is that people like to see what happens when they pursue all kinds of associative threads. Tabbed browsing is a popular GUI feature supporting this tentative navigation. If you follow a link in a new tab it is easy to go back to where you were before following the link.

In Nepomuk Simple, tabs are used to separate piles, and creating a new tab means creating a new pile. Generally, such a tab is either empty or it contains only the resource dropped there. To provide a "playground" for experimenting with possible changes to an existing pile, we plan to provide the option of creating a (temporary) copy of a pile, in a new tab.

Having a copy of a pile is a good starting point for playing with different pile configurations, and see how they affect recommendations. However, playing with resources can be even further supported if we introduce the concept of a "shelf". On a shelf, the user can put (iconic representations of) resources which at some point appear on the screen, and which the user considers potentially interesting candidates for inclusion in the pile. Resources put on the shelf can be easily re-located, included in the pile, put back on the shelf, or removed from the shelf.

We believe that this GUI element would support the user in his or her mental processing of the pile topics. The usage scenario would be somewhat similar to everyday web searching, where queries are often successively refined by adding and subtracting terms to the latest query string. Yet another option for more fully exploiting the capabilities of the Galaxy algorithms would be to provide one "positive" and one "negative" area in the pile. Resources placed in the negative area would provide input to Galaxy to lessen the ranking of resources having a strong relation with these resources, while resources put in the positive area would be treated just as ordinary initial activation nodes in the Galaxy spread of activation algorithms.

## 5 Conclusions and Future Work

In this paper we have outlined the problem of related item recommendations on networked data and suggested that it might be viewed as the problem of processing fuzzy polycentric queries, such as those provided by the pile UI. For the processing of such queries we suggest the use of spreading activation methods, since they provide a blend of soft clustering and fuzzy inferencing, which is needed for adequate recommendations in the scenarios we consider.

The initial data for recommendations which allow browsing from one single thing to another, one-to-one correspondences can be treated as an egocentric query on the network. Similarly, many-to-many correspondences are treated as polycentric network queries.

We conclude that the process of incremental modification of a “pile”, assisted by recommendations resulting from fuzzy polycentric queries, might be formally viewed as the process of browsing and exploring data provided by networks. This type of navigation is not simply browsing from a single object to another, but by dealing with several objects at the same time in a process similar to how one browses in a library or shop. Such an “ambient navigation” tool might be used for exploring various massive multidimensional networks, which occur in the socio-semantic information space which we encounter in the modern information age. This usage extends beyond the scope of scenarios explored in the EU 6$^{th}$ framework project Nepomuk where these ideas have been developed. For instance, Nepomuk Simple powered by Galaxy might be used to navigate not Nepomuk PIMO, but social networks.

The ambient navigation described here is still a work in progress. Integration efforts are ongoing and the (tentative) results thus far have been encouraging. Our future work in this area will focus on finishing of integration Nepomuk Simple with the Galaxy library.

## References

1. Troussov, A., Sogrin, M., Judge, J., Botvich, D.: Mining Socio-Semantic Networks Using Spreading Activation Technique. International Workshop on Knowledge Acquisition from the Social Web (KASW'08), (2008)

2. Kaptelinin, V, Czerwinski, M. (ed.): Beyond the Desktop Metaphor: Designing Integrated Digital Work Environments. MIT Press (2007)
3. Malone, T.W. How do people organize their desks? Implications for designing office information systems, ACM Transactions on Office Information Systems, 1983, 1, 99-112
4. Agarawala, A., Balakrishnan, R.: Keepin' It Real: Pushing the Desktop Metaphor with Physics, Piles and the Pen. Proceedings of the SIGCHI conference on Human Factors in computing systems, 2006. p. 1283-1292
5. Sandor, O., Bogdan, C., Bowers, J.: Aether: An Awareness Engine for CSCW. Proceedings of the fifth conference on European Conference on Computer-Supported Cooperative Work (ECSCW'97), 1997