Navigating and Annotating Semantically-Enabled Networks of People and Associated Objects

Sheila Kinsella[1], Andreas Harth[1], Alexander Troussov[2], Mikhail Sogrin[2], John Judge[2], Conor Hayes[1], John G. Breslin[1]

[1] Digital Enterprise Research Institute, National University of Ireland, Galway, Ireland

sheila.kinsella@deri.org, andreas.harth@deri.org, conor.hayes@deri.org, john.breslin@deri.org

[2] IBM LanguageWare, Dublin Software Lab, Ireland

atrousso@ie.ibm.com, sogrimik@ie.ibm.com, johnjudge@ie.ibm.com

Abstract: Social spaces such as blogs, wikis and online social networking sites are enabling the formation of online communities where people are linked to each other through direct profile connections and also through the content items that they are creating, sharing and tagging. As these spaces become bigger and more distributed, more intuitive ways of navigating the associated information become necessary. The Semantic Web aims to link identifiable objects to each other and to textual strings via relationships and attributes respectively, and provides a platform for gathering diverse information from heterogeneous sources and performing operations on such linked data. In this paper, we will demonstrate how this linked semantic data can provide an enhanced view of the activity in a social network, and how the Galaxy tool described in this work can augment objects from social spaces, by highlighting related people and objects, and suggesting relevant sources of knowledge.

This work has been partially funded by Science Foundation Ireland under Grant No. SFI/02/CE1/I131 and a Supplemental Equipment Grant.

1. Introduction

The ability to link to other pages and objects is a key facility of the World Wide Web architecture. It has enabled every web site to become part of a global network of information. More recently, new client server applications such as wikis and blogs have made writing and linking on the Web extremely easy for the average user. The result has been the creation of vast amounts of user-generated content, often organised within on-line communities. In order to take advantage of the huge store of knowledge which is amassing online, we require new methods of navigating this data. The problem is not simply one of countering information overload, although this is certainly pertinent, but of providing links to relevant sources of information, possibly scattered across several domains. The goal is to enable the user to move through the information space quickly and intuitively by suggesting relevant related people, concepts and objects at every step.

One problem is that the current link mechanism on the Web does not differentiate between different types of links and does not allow different types of relationships to be expressed. Data is presented as a set of documents and other files, interconnected by hypertext links. The concepts represented in the documents and the types of the relationships between them are not explicitly stated, and can be hard for a computer to infer. This has meant that applications that do provide navigation aides, such as recommender systems, tend to concentrate on single types of relationships (e.g. likes and dislikes) and tend to be located within single domains (e.g. Amazon.com). Preference and relationship data accumulated by one user in a particular domain cannot be easily transferred to another domain. For instance, a blogging community may be dispersed over numerous different sites and platforms, and an interest group may share photos on

Flickr, bookmarks on del.icio.us, and hold conversations on a discussion forum. A single person may have several separate online accounts, and may have a different network of friends on each. Therefore, the information existing in online social spaces forms massive, intricate and generally disjoint networks of people and objects.

In short, the lack of standards for expressing semantic data in links in Web 1.0 has meant that semantic relationships have had to be inferred and stored, generally by single organisations in large centralised silos – with negative implications for interoperability, ownership and privacy.

Semantic Web research (Berners-Lee et al. 2001) offers the possibility of overcoming many of these problems by enabling the description of arbitrary objects or concepts, and the relationships between them, using shared machine-readable formats. Each object has a unique Uniform Resource Identifier (URI), which allows it to be referenced across sources. Semantic data can be viewed as a directed graph where the vertices represent objects or concepts, and the edges represent semantic relationships. A fundamental part of the Semantic Web is the ontology, a data structure specifying the concepts that are needed to understand a domain, and the vocabulary and relationships required to enter into a discourse about it.

Representing Web data in this way allows the expression of different types of relationships between people, between people and concepts or objects, and so forth. Furthermore, these types of relationships are expressed in non-proprietary formats and can be transferred and understood in the different domains or communities. For example, the Friend-of-a-Friend (FOAF)[1] vocabulary allows for the expression of the links between people and the things they create and do. The relationships between communities of friends represented in FOAF can be processed in any program that understands the FOAF vocabulary.

Creating a graph on the Web of different types of objects linked by different types of relationships is a major step towards realising navigation aides or recommender systems that can process various kinds of relationships and objects. However, to fully realise the power of these new representation models, users require techniques to extract knowledge from the Semantic graph and to infer associations between objects that may not be explicitly linked. In Semantic Web research, the standard way to do this is to use an inference engine based on a logic framework such as the Web Ontology Language (OWL) to allow logic reasoning on the Web (Dean and Schreiber 2004). An inference engine operating on the Semantic Web graph generates new triples, i.e. the output model is the input model plus additional inferred triples.

However, inferring general relationships from graphs can be achieved using techniques other than logic. In this paper we demonstrate how relevant related information can be extracted from Semantic Web data using the Galaxy tool where the output of inferred triples is generated by a spreading activation technique over weighted links. A related method has been applied (Amitay et al. 2004) to derive a geographical focus from a text, based on locations which are mentioned in the text, but that algorithm can operate only on a hierarchical network. Spreading activation has been applied to semantic networks for social network analysis in applications including recommender systems (Liu et al. 2006), community detection (Alani et al. 2003), and modeling trust propagation (Ziegler and Lausen 2005).

To demonstrate our technique, we gather information represented in common formats and represent the data as a semantic graph, consisting of interrelated people, objects and their associated semantic terms. This data is used as input to the Galaxy tool which provides a generic way of ontology-based network mining. We attempt to locate a set of related items within our dataset, given some text referring to a particular person or object, or to a set of people and objects. Our approach makes use of the network of links existing between people, including not only social connections, but also semantic connections via shared interests or other areas of common ground. The analysis extends further than people and objects that are closely related, to three degrees of separation and beyond.

The main contributions of this paper are as follows:

- We describe how a semantic data model of social spaces can give improved insight into the activity of a social network
- We explain the capabilities of the Galaxy tool in mining social semantic networks to provide an enhanced view of networked data
- We present initial results of experiments carried out on a data set extracted from the Semantic Web

2. Object-centered networks

Jyri Engeström, co-founder of the micro-blogging site Jaiku, has theorized that the longevity of social networking sites is proportional to the "object-centered sociality"[2] occurring in these networks, i.e. where people are connecting via items of interest related to their jobs, workplaces, favourite hobbies, etc. On the Web, social connections are formed through the actions of people - via the content they create together, comment on, link to, or for which they use similar annotations.

Adding annotations to items in social networks (e.g., using topic tags, geographical pinpointing, etc.) is an especially useful aid for browsing and locating both interesting items and related people with similar interests. Some popular types of content items include blog entries, videos, and bookmarks. These objects serve as the lodestone for social networks, drawing people back to check for new items and for any updates from those in their network who share their interests. On Flickr, people can look for photos categorized using an interesting "tag", or connect to photographers in a specific community of interest. On Upcoming, events are also tagged by interest, and people can connect to friends or like-minded others who are attending social or professional events in their own locality.

The figure below is illustrative of an object-centered social network for three people, showing their various user accounts and the things that they create and do using these accounts. Bob and Carol are connected through bookmarked websites that they both have annotated and also through events that they are both attending, and Alice and Bob are using matching tags on media items and are subscribed to the same blogs.

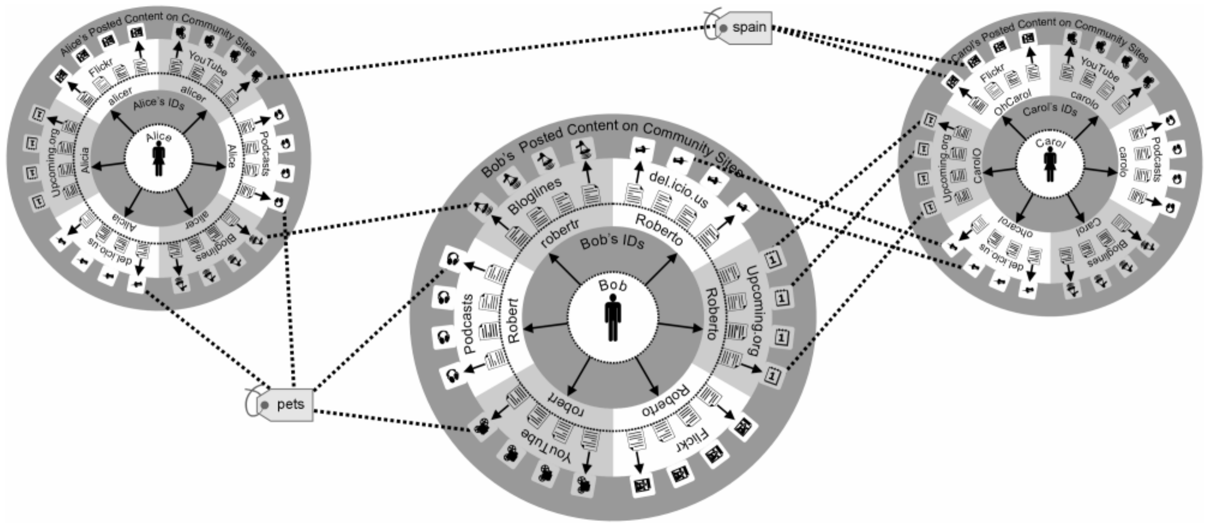

Figure 1: Object-centered social networks are formed by people (using their online accounts) and the content items they act upon

As the connections between people become intertwined with their real-world interests, it is probable that people's social networking methods will move closer towards simulating their real-life social interaction, so that people will meet others through something they have in common, and not by randomly approaching each other.

Since more interesting social networks are being formed around the connections between people and their objects of interest, and as these object-centered social networks grow bigger and more diverse, more intuitive methods of navigating the associated information contained in these networks have become necessary – both within and across social networking sites (e.g., a community of interest for mountaineering may consist of people and content distributed across photo-, bookmark- and event-centred social networks).

Person- and object-related data can also be gathered from various social networks and linked together using a common representation format. This linked data can provide an enhanced view of individual or community activity in a localized or distributed object-centered social network(s) ("show me all the content that Alice has acted on in the past three months").

The Semantic Web, which aims to link identifiable objects to each other and to textual strings, can be used for linking the diverse information from heterogeneous social networking sites and for performing operations on such linked data. The involvement of objects in social networks on the Semantic Web has been investigated (Kinsella et al. 2007). The Semantic Web is already being used by various efforts to augment the ways in which content can be created, reused and linked by people on social networking and media sites. These efforts include the Friend of a Friend (FOAF) project, ontology-enhanced wikis such as the Semantic Media Wiki, the NEPOMUK social semantic desktop[3], and the Semantically-Interlinked Online Communities (SIOC) initiative. In the other direction, object-centered networks can serve as rich data sources for Semantic Web applications. Tim Berners-Lee said in a 2005 podcast, "I think we could have both Semantic Web technology supporting online communities, but at the same time also online communities can support Semantic Web data by being the sources of people voluntarily connecting things together." Users of social networking sites are already creating extensive vocabularies and annotations through "folksonomies" (collections of free-text keywords that are used to tag content items). Since the meaning of these terms is being produced through a consensus of community users, these terms are serving as the objects around which more tightly-connected social networks are centred and formed.

3. Semantic Web

The purpose of the Semantic Web is to enable the online description of arbitrary objects in such a way that software can be used to automatically combine, mine, process, and manipulate data from the Web. Machine-readable descriptions of objects and the relationships between them on the Web enable universal knowledge representation mechanisms on a global scale. For the simplest form of object identification, the same Uniform Resource Identifier is used across multiple sources to reference an object. In many people using the same URI for a particular object, the available data pieces mesh up and form a well-connected and richly-interlinked information space with structured representation features. Layered on top of the foundational URI naming mechanism are a number of other technologies to enable knowledge representation features of increasing complexity and sophistication:

- Resource Description Framework (RDF): a universal way of identifying and talking about entities, basic type system (Manola and Miller 2004)
- RDF Schema (RDFS): a vocabulary with terms for describing classes and properties, subclass and subproperty relationships (Brickley and Guha 2003)
- Web Ontology Language (OWL): terms for describing classes, inverse properties, cardinality constraints; subset of first order logics

Information on the Semantic Web is commonly expressed using the RDF language. An RDF document is composed of a sequence of statements of the form <*subject*, *predicate*, *object*>, indicating a directed link from the subject node to the object node, where the predicate link describes the relationship between them.

RDF uses the concept of URIs to name all sorts of objects; for example: *http://www.w3.org/People/Berners-Lee/card#i* to denote Tim Berners-Lee, *http://sws.geonames.org/2964180/* to denote the city Galway, *http://deri.ie/* to denote the research institute, and *http://purl.uniprot.orgs/uniprot/Q91474* to denote the protein

SHNF1. Objects identified via URIs typically have one or many associated types e.g. *http://xmlns.com/foaf/0.1/Person* or *http://swrc.ontoware.org/ontology#FullProfessor*. The relationships between objects are denoted using URIs, such as the instance-to-type relation *rdf:type*. Namespace prefixes (such as *rdf:*) can be used to abbreviate URIs.

On the level of RDFS, the nodes represent instances of classes, and the links represent instances of properties. Classes and the possible properties which can exist between them are defined in RDFS or OWL. The description of classes and properties form a vocabulary that can be created or extended as required. For example, vocabularies exist to describe conferences, projects, communities, geographical information, trust networks, and many other domains. Sometimes namespace prefixes indicate the schema to which classes and properties belong. Dereferencing the namespace prefix URI via the Hypertext Transfer Protocol HTTP can provide a machine-interpretable description of the classes and properties.

OWL is an expressive language inspired by Description Logics, a subset of first order logics. In this study, we use OWL's logical constructs mainly for cleaning up the dataset, and rather focus on the benefits of RDF in data integration scenarios, leveraging its directed labeled graph data model to describe real-world objects.

4. Dataset

We use the Galaxy tool to analyse a dataset consisting of social network information focused around the Semantic Web community. Our model includes people and related entities, specifically interests, documents, workplaces, projects and schools. The data under analysis is part of a web crawl of RDF data that was carried out during June/July 2007 using MultiCrawler (Harth et al. 2006). The initial dataset originates from approximately 85,000 sources and consists of over 35 million statements. Object consolidation (Hogan et al. 2007) was performed in order to merge identifiers of equivalent instances occurring across different sources. From the original crawl, we extracted a smaller sub-graph for analysis. The sub-graph is based around the URIs of four people in the Semantic Web community: Tim Berners-Lee, Dan Brickley, Andreas Harth and Tim Finin. We used YARS2 (Harth et al. 2007) to extract all people connected within three links of the root nodes, via *foaf:knows* relations. We also included any other nodes connected to these people. The resulting dataset consists of a vast amount of information in many different vocabularies; the total number of statements is shown in Table 1. In order to process the data using the Galaxy tool, it was necessary to convert the required information into a format which it can read.

Distance from root nodes	No. of Statements
1 hop	3889
2 hops	94540
3 hops	1237021

Table 1: Number of statements found at 1, 2, and 3 hops from the root nodes, before conversion and filtering

The current version of Galaxy is an early prototype which takes input data expressed in an XML format. However it is planned that RDF support will be available in the near future. We developed a program which can extract specific information from RDF, and map it to the required format. For this initial work, we only include a small set of link types, but it would be possible to extract a much broader range of data. The information we extract is a subset of three vocabularies. Most of the extracted data is described using the Friend of a Friend vocabulary (shorthand:*foaf*), which enables the description of people and their relationships with other resources. It also enables the expression of other information relating to a person, such as contact details, workplace and school details, as well as publications and other items they have created. Anyone can create their own FOAF file describing themselves and their social network, and the information from multiple FOAF files can easily be combined to obtain a higher-level view of the network. We also include some data expressed using the RDF Schema (shorthand:*rdfs*) and Dublin Core[4] (shorthand:*dc*), both of which include properties commonly used to specify the names of resources. There are two main steps to the conversion process - extraction of nodes and links, and extraction of text labels.

We derive information from RDF statements based on predicates. All extracted nodes and links are assigned a type. For instance, all object nodes which occur with the predicate *foaf:interest* are mapped to type 'interest'. The predicates which we extracted are shown in Table 2, along with the link type each predicate was mapped to.

Predicate	**Link Type**
foaf:knows	knows
foaf:interest	hasInterest
foaf:currentProject	hasProject
foaf:pastProject	
foaf:workInfoHomepage	hasWorkplace
foaf:workplaceHomepage	
foaf:schoolHomepage	hasSchool
foaf:made	isMakerOf
foaf:maker	madeBy

Table 2: RDF predicates which were extracted and the link type to which they were mapped

Figure 2 shows the node types which exist in our data model, and the link types which connect them together. The link type "madeBy" is considered to be the inverse of the link type "isMakerOf"; in other words, they represent the same relationship, but in opposite directions.

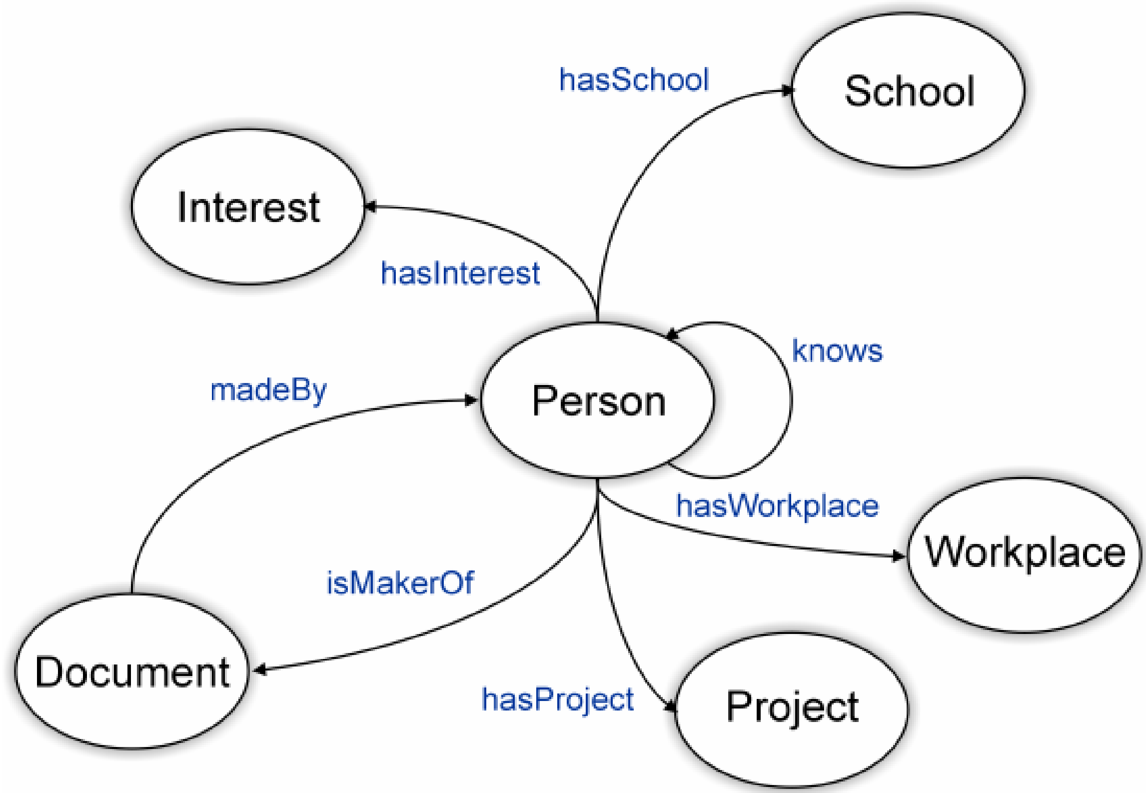

Figure 2: Node and link types in the data model

We also extract names for nodes, so that textual references to a particular node will be recognised by the Galaxy tool. For each node type, we compiled a list of the predicates which indicate that the object node is a name for the subject node. For example, where the subject node is of type Person, this list includes predicates such as *foaf:name* and *foaf:nick*. This information is applied to the RDF document and the relevant object nodes are then attached as labels to their respective subject nodes. Table 3 shows for each node type the predicates which we assume to indicate names. Some nodes may have several different labels. Other nodes do not have any name specified, in which case we use the URI of the node as a label.

Type	Names
Person	foaf:nick, foaf:name, foaf:firstName, foaf:givenname, foaf:family_name, foaf:surname
Interest	dc:title, dc:subject, rdfs:label
Project	dc:title, dc:subject, rdfs:label
Workplace	dc:title, dc:subject, rdfs:label
School	dc:title, dc:subject, rdfs:label
Document	dc:title, dc:subject, rdfs:label

Table 3: Node types and the predicates which indicate names

Our dataset contains 16468 entities and 25028 relationships. Most of the entities are people. The composition of the dataset is shown in detail in Table 4.

Node Type	Instances	
Person	11314	(68.7%)
Interest	2228	(13.5%)
Document	1956	(11.9%)
Workplace	443	(2.7%)
Project	339	(2.1%)
School	188	(1.1%)

Table 4: Frequencies of node types in the network

5. Galaxy

Galaxy is an ontological network miner designed for the Nepomuk project by the IBM LanguageWare Team[5] for application to tasks in social semantic computing. The Galaxy tool uses a spreading activation algorithm to perform clustering on semantic networks. Instead of the traditional method of hard clustering, which partitions a graph into different groups, Galaxy performs soft clustering, which involves taking a sub-graph based around a set of input nodes, and finding the focus of this sub-graph. The method can be applied to social networks, company organisation charts, or any other set of graph-structured data. Initially, an ontological network of concepts and related terms must be generated based on data provided by the user. Galaxy can then process documents, and identify their main concepts, based on the ontological information. The two main steps to this process are the mapping of terms to concepts, and the location of the main concepts.

The Galaxy tool takes a piece of text as input, and then maps terms in the text to concepts in the ontological network. If necessary, the topology of the graph is used in disambiguating terms in the document. The concepts which are identified as corresponding to terms in the text act as input nodes for the spreading activation algorithm. The result of the algorithm is a set of focus nodes, which can be interpreted as those nodes which are most central in the sub-graph based around the input set.

The spreading activation algorithm uses the principle of light propagating in a graph via its links to identify the focus nodes for a query. The input nodes of a query act as sources of light which spreads outwards. As the light travels farther from a source node, it gets dimmer. If the light emitted from multiple nearby source nodes combines, the point at which they overlap will be illuminated to a greater degree. The nodes which accumulate the most light are deemed to be the focus nodes, and the concepts which they represent are displayed to the user.

Figure 3 shows an undirected graph where each of the four corner nodes have been activated and act as sources of light. Figure 4 shows the same graph after light has been propagated around the graph.

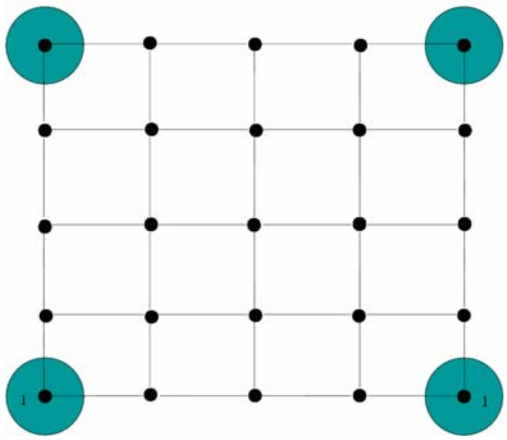

Figure 3: Graph before propagation of light

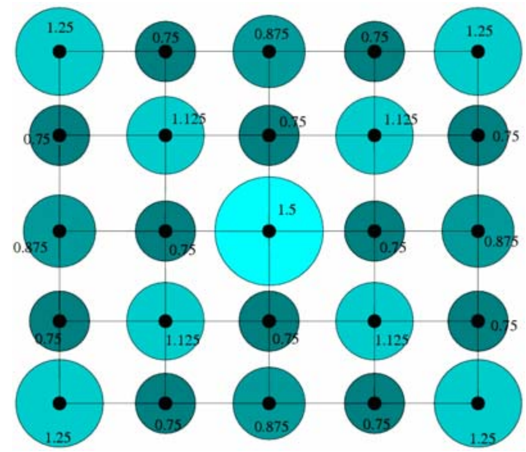

Figure 4: Graph after propagation of light

Galaxy can be used with any kind of graph or tree, and allows for both directed and undirected links. Various parameters can be tuned to alter the behaviour of the algorithm. This allows a domain-expert to stipulate the properties of a semantic graph that are most important for a particular task. For example, in a graph with different types of relations, some may be considered more relevant than others, depending on the application. The Galaxy tool allows for link types to be weighted in order to reflect the relative significance of different relationships.

Galaxy can be customised to a range of different tasks. Possible application areas include expert-finding, metadata creation, and community detection.

6. Results

In the following we present the results of some sample queries for the semantic graph described in Section 4. In each case, we provide Galaxy with a short piece of text, and it uses the topology of the semantic network to extract the main concepts of the document based on terms mentioned in the text. Each instance is represented by a URI corresponding to that resource, but here we display text names so that the results are human-readable. Our queries are based around three people: John Breslin, Tim Berners-Lee and Andreas Harth. Firstly, we perform queries for each of these individuals in order to obtain an ego-centric view of their network. Secondly, we perform queries involving pairs of individuals as a means of detecting the community to which they belong.

The objective of the ego-centric queries is to derive an overview of the most relevant available content relating to a particular person. The results for Query 1, “John Breslin”, are shown in Table 5. For this query, Galaxy identifies the people John Breslin and Hannes Gassert, as well as several entities directly related to John Breslin and two entities related to Hannes Gassert (Semantic Web at del.icio.us and mediagonal).

<table>
<tr><th>Type</th><th>Instances</th></tr>
<tr><td rowspan="2">Person</td><td>John Breslin</td></tr>
<tr><td>Hannes Gassert</td></tr>
<tr><td rowspan="3">Interest</td><td>Semantic Web at del.icio.us</td></tr>
<tr><td>Semantic Web</td></tr>
<tr><td>RDF</td></tr>
<tr><td>Document</td><td>John Breslin's blog</td></tr>
<tr><td rowspan="5">Workplace</td><td>Semantic Web Cluster, DERI</td></tr>
<tr><td>DERI</td></tr>
<tr><td>DERI Galway</td></tr>
<tr><td>Líon Project, DERI</td></tr>
<tr><td>mediagonal</td></tr>
<tr><td>School</td><td>National University of Ireland, Galway</td></tr>
</table>

Table 5: Results for Query 1: “John Breslin”

The results for Query 2, “Tim Berners-Lee”, are given in Table 6. Galaxy locates the appropriate person and additionally one interest and several documents.

Type	Instances
Person	Tim Berners-Lee
Interest	Semantic Web
Document	FOAF Document for Tim-Berners Lee
	Tim Berners-Lee's blog
	N3Logic : A Logic For the Web
	Creating a Policy-Aware Web: Discretionary, Rule-Based Access for the World Wide Web
	Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web
	Semantic Web Boot Camp 2007 data

Table 6: Results for Query 2: "Tim Berners-Lee"

Query 3 for "Andreas Harth" locates the person Andreas Harth, one interest and two projects, as shown in Table 7.

Type	Instances
Person	Andreas Harth
Interest	Knowledge Representation
Project	YARS
	SWSE

Table 7: Results for Query 3: "Andreas Harth"

The data on which these results are based originates not just from the FOAF files of the individuals involved, but also from other documents which contain references to these people. Results like these could be useful to someone who has come across a reference to these people on the Web and is interested in finding out more related information.

We also experimented with using Galaxy identify a community, starting with multiple individuals within that community. We chose two queries, each mentioning two people: "John Breslin, Tim Berners-Lee" and "John Breslin, Andreas Harth". The results of the query "John Breslin, Tim Berners-Lee", filtered to return only nodes of type Person, are shown in Table 8.

Type	Instances
Person	John Breslin
	Tim Berners-Lee
	Dan Brickley
	Eric Miller
	James Hendler
	Henry Story
	Charles McCathieNevile

Table 8: Results for Query 4: "John Breslin, Tim Berners-Lee", Persons only

The subjects of our first query, John Breslin and Tim Berners-Lee, are both involved in Semantic Web research. However they are not directly connected to each other. The results show that Galaxy has identified a set of individuals who are located around the two subjects in our query, resulting in a broad view of the Semantic Web community. These people were not identified as relevant to either of our initial separate queries for John Breslin and Tim Berners-Lee, however when we take the two people together they are found to be important. The results in Table 8 are based on data aggregated from Tim Berners-Lee's FOAF file, John Breslin's FOAF file, and other documents. This overview of the network is not possible without considering information from multiple sources in our dataset.

Type	Instances
Person	John Breslin
	Andreas Harth
	Hannes Gassert
	Aidan Hogan
	Matteo Magni
	Fergal Monaghan
	Sheila Kinsella
	Siegfried Handschuh
	Axel Polleres
	Knud Möller

Table 9: Results for Query 5: "John Breslin, Andreas Harth", Persons only

Query 5 involves John Breslin and Andreas Harth. The results, filtered to return only nodes of type Person, are shown in Table 9. In this query the two people are again Semantic Web researchers, however in this case they work closely together within the same research institute. The second query therefore has a much narrower focus than the first. All of the people identified by Galaxy for the query "John Breslin, Andreas Harth" are either members or former members of the Digital Enterprise Research Institute, and are closely connected to one or both subjects of the query. Most of them were not identified in Queries 1 or 3. As for the previous query, the results are enabled by the aggregation of social networks expressed in multiple interconnected FOAF files.

Although all of the queries mentioned above are very simple, longer text documents can be analysed with Galaxy, for example e-mails and blog posts.

7. Discussion

The examples we have shown in this paper indicate that mining the graph of Semantic Web data using a spreading activation approach allows for the discovery of new relationships between nodes. Evaluating this type of system in a more objective way will be a difficult task. This is due to difficulty in establishing reasonable baselines to compare the system with, and in judging which criteria to use when making this evaluation. There are many aspects we could analyse - for instance the ease with which it can be used is certainly an important factor for success. This has much to do with HCI factors such as presentation and the interaction model employed.

However, the most common evaluation approaches for recommender-type systems are performed off-line using techniques from machine learning and information retrieval such as cross validation and measures of recall/precision (Hayes et al. 2002). Off-line evaluation requires labeled training and test data in order to measure how many relevant results are retrieved for each test query. Unfortunately, such data is hard to come by, particularly in the domain of the Semantic Web, where the relevance of query results can not be easily defined in advance.

There is still a lot of research to be done in order to find means of expressing end-user queries over semantic web data. Depending on how precise the queries are, only one valid URI would be relevant. Large scale corpora are only beginning to appear, and it is hard to assign relevancy scores when both the queries and the data are unclear. In principle large datasets could be labeled, but there is significant human effort involved in such a task.

Some approaches approximate the relevance of search results by comparing them with the results achieved by querying another 'authoritative' Web source such as Google (Hayes and Avesani 2007). This is not a satisfactory approach where different types of data objects are returned as in the Semantic Web approach described here.

Alternatively, on-line or live approaches evaluate performance in real time with real users. On-line evaluation is problematic because of the need to field a fully engineered system and to find a representative community of users. However, the results from such an approach can be extremely useful in detecting the incorrect assumptions and biases underlying any one particular navigation or recommender strategy. Typically, the strategy being evaluated would be deployed concurrently with a rival 'baseline' strategy. Neither the evaluator nor the test subject should be aware of which strategy is being deployed – the classic double blind test.

Future evaluations of the technique we have described in this paper will incorporate both off-line and on-line strategies. We will test our system off-line using small labeled data sets to help us develop our hypotheses and then we will test our hypotheses on real users.

8. Conclusions

This paper demonstrates how relevant related information can be extracted using the Galaxy tool from a set of Semantic Web data obtained from multiple online sources, where the output of related items is generated by a spreading activation technique over weighted links. We began with an outline of object-centered networks, and described how a semantic data model of social spaces can give an improved insight into the activity of a social network. We then explained the capabilities of the Galaxy tool in ontology-based mining of social semantic networks, and showed how it can provide an enhanced view of networked data. Finally, we presented initial results of experiments carried out on a data set extracted from the Semantic Web, which makes use of the network of links existing between people, including not only social connections, but also semantic connections via shared interests or other areas of common ground. The analysis extends further than people and objects that are closely related, to three degrees of separation and beyond.

References

Alani, H., Dasmahapatra, S., O'Hara, K. and Shadbolt, N. (2003). "Identifying Communities of Practice through Ontology Network Analysis." IEEE Intelligent Systems 18(2): 18-25.

Amitay, E., Har'El, N., Sivan, R. and Soffer, A. (2004). "Web-a-where: geotagging web content". Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval, NY, USA.

Berners-Lee, T., Hendler, J. and Lassila, O. (2001). "The semantic Web." Scientific American 284(5): 28-37.

Brickley, D. and Guha, R. V. (2003). "RDF Vocabulary Description Language 1.0: RDF Schema." W3C Working Draft.

Dean, M. and Schreiber, G. (2004). "OWL Web Ontology Language Reference."

Harth, A., Umbrich, J. and Decker, S. (2006). "MultiCrawler: A Pipelined Architecture for Crawling and Indexing Semantic Web Data". Proceedings of the 5th International Semantic Web Conference, Athens, GA, USA.

Harth, A., Umbrich, J., Hogan, A. and Decker, S. (2007). "YARS2: A Federated Repository for Searching and Querying Graph-Structured Data". Proceedings of the 6th International Semantic Web Conference, Busan, Korea.

Hayes, C. and Avesani, P. (2007). "Using Tags and Clustering to Identify Topic-Relevant Blogs". Proceedings of the 1st International Conference on Weblogs and Social Media, ICWSM 07, Boulder, CO, USA.

Hayes, C., Massa, P., Avesani, P. and Cunningham, P. (2002). "An on-line evaluation framework for recommender systems". Proceedings of the Workshop on Recommendation and Personalization in eCommerce, at AH2002, Malaga, Spain.

Hogan, A., Harth, A. and Decker, S. (2007). "Performing Object Consolidation on the Semantic Web Data Graph". Proceedings of I3: Identity, Identifiers, Identification, Workshop at 16th International World Wide Web Conference (WWW2007), Banff, Alberta, Canada.

Kinsella, S., Harth, A. and Breslin, J. G. (2007). "Network Analysis of Semantic Connections in Heterogeneous Social Spaces". Proceedings of the UK Social Network Conference, London, United Kingdom.

Liu, H., Maes, P. and Davenport, G. (2006). "Unraveling the Taste Fabric of Social Networks." International Journal on Semantic Web and Information Systems 2(1): 42-71.

Manola, F. and Miller, E. (2004). "RDF Primer."

Ziegler, C. N. and Lausen, G. (2005). "Propagation Models for Trust and Distrust in Social Networks." Information Systems Frontiers 7(4): 337-358.

---

[1] http://www.foaf-project.org/

[2] http://www.zengestrom.com/blog/2005/04/why_some_social.html

[3] http://nepomuk.semanticdesktop.org/

[4] http://dublincore.org/

[5] http://www.alphaworks.ibm.com/tech/lrw