Ontology Based Modeling and Visualization of Social Networks for the Web

Tamara Oellinger, Pinar Oezden Wennerberg

SES Support to External Security – Web Intelligence and Monitoring Group
JRC Joint Research Center of the European Commission
IPSC SES TP 267 I-21020 Ispra (VA) Italy
tamara.oellinger@jrc.it
pinar.oezden@jrc.it

Abstract: Exploring and processing the relationships between social entities by mapping them to mathematical graph models has been a rapidly growing research area known as Social Network Analysis (SNA). The majority of models adapt statistical approaches such as data mining to discover social relationships in given resources. This paper introduces an initial work that takes a knowledge representation approach utilizing ontologies in order to map the relationships between social entities to logical models. This approach enables the application of inference mechanisms by exploiting the logical structure of the model. As a result, we gain automatically additional knowledge from the resources at hand. This is realized in our work distinguishing it from other related research, which is based on statistical or knowledge representation approaches that we have been aware of. Additionally, we use this structure and the automatically gained information to visualize the social network. This enables information seekers to have an explicit view of the entities and the relationships between them. The interactive visualization acts as an exploration tool which allows users to find pertinent information. As a demonstration, we use the ontological model to discover and to visualize the relationships between the terrorist organization Al Qaeda, the people and the events associated with it.

1 Introduction

Identifying relationships between social entities, events and locations may not be obvious at first glance. Taking advantage of Social Network Analysis (SNA), which has been a rapidly growing area of research in the recent decades [Wa99, Sc04, DF99], such hidden relationships can be discovered by explicitly stating them in form of a network model. In the literature, Social Network Analysis is defined as the analysis of "relationships among social entities and on the patterns, structures and implications of relationships" [Wa99].

In computer science, ontologies are understood as devices that bring a machine-readable conceptual structure to a domain of interest. More formally, ontologies are defined as "a formal explicit specification of a shared conceptualization" [GCF04]. In recent years the application of ontologies to the Web has received a high level of attention as a research area [Us03, LLH01]. As such, ontologies provide the necessary conceptual and logical framework to capture the information present in a social network and analyze it using hierarchical structures, axioms and rules. In addition to statistical approaches [Ne06, KSS97], modeling of social networks can be done with ontologies. [De06]. We see several advantages in taking the latter approach. Firstly, we can explicitly specify the concepts and the relationships relevant to the domain of interest. This ensures that the model best fits our requirements and is error free. Secondly, using reasoning and inference, ontologies can guarantee that the modeling of contradictory or inconsistent information is not allowed; hence the validity of the information encoded is verified. Thirdly, ontologies combined with the inference mechanism result in automatic information gain through the rules that can then infer new information. Inference mechanism can be facilitated over ontology-based social networks to automatically come up with new relations (typically more generic) and concepts beyond the ones existing about or between the entities. Despite certain drawbacks, such as human intervention, time and resource demands and restrictions to particular domains, we believe that the advantages of the ontology-based approach outweigh the disadvantages. Our work differs from that of other ontology-based work mentioned above in that we deploy reasoning and inference on our model.

We used the ontology to explicitly model and visualize the relationships between the social entities connected to the Al Qaeda terrorist organization. Related social entities can not only be persons and organizations but also events and locations. The domain of terrorism is rather suitable to demonstrate the application of ontology-based Social Network Analysis because it is a restricted domain with well defined entities (such as S*uicide Bomber* or *Terrorist Organization*) and well defined relationships (such as *organizer of*, *follower of*). The following scenario describes our view of how this could be achieved. For example, from distributed resources we may collect the following information: Abdullah Al Reshood *roommateOf* Aafia Siddique, Aafia Siddique *isMemberOf* Al Queda, Aafia Siddique *traveledTo* Pakistan, Mounir-al Motassad *traveledTo* Pakistan. Once such a social network model is captured in a logical structure through an ontology, it becomes available for an inference engine to run its rules over the model to find out if any further information can be inferred. In case of success, the resulting implicit information can be explicitly stated and can be further tracked: Abdullah Al Reshood *isRelatedTo* Al Queda? Al Queda *isRelatedTo* Pakistan? Hence, knowledge discovered in this fashion, could be used as input to discover and visualize on the Web further information related but not limited to security issues.

The proper presentation of information, and especially that of large data sets, remains crucial as poorly presented information can be unclear, confusing and possibly misleading. In this respect, the web-based interactive visualization interface is an essential complement to the project for several reasons. Firstly, unlike textual information representation, visualisation offers an intuitive way of displaying the information including that which is newly inferred by the model, and understanding the connections between entities. Secondly, the interactivity permits the user to selectively focus on his area of interest. Thirdly, the Internet is well suited for the dissemination of information as it is increasingly accessible to a greater number of users.

2 Design and Implementation of the Social Network

The design and the implementation process involved two tasks. The first task was to set up the conceptual structure, i.e. the ontology, whilst the second task was to build a Web application using the ontology in such way that it is accessible to end-users on the Web. In the following subsections we first describe the details about the design of the ontology and then the implementation of the application that builds on it. Henceforth we refer to the application as the Knowledge Base. In Section 3 we explain the details of the visualization of the Knowledge Base.

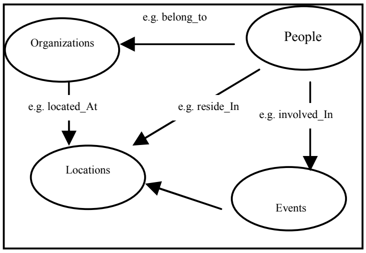
2.1 The Ontology Component

The Knowledge Base deploys four different ontologies encoded in OWL[1] (Web Ontology Language); People, Organizations, Events and Locations. In our opinion, this model reflects the relationships between the associated entities of the given organization. This separate ontology approach also facilitates modularization. This way modifications and extensions can be made to any of the four ontologies without affecting the structure of the other ones. The classes, the attributes and the relations to be included in the ontologies are determined on the basis of their suitability to capture the information in the online news articles.

The table below presents some relationships, concepts and attributes from the People, the Organizations, the Locations and the Events ontologies.

Ontology	Concepts	Attributes	Relations
People	*Politician, Spokesperson…*	*hasDateofBirth, hasGender…*	*isFounderOf, isOrganizerOf….*
Organizations	*GovernmentOrganization, ReligiousOrganization, TerroristOrganization*	*hasFoundationDate*	*isOrganizerOf, hasMember, isInConflictWith*
Locations	*Country,City, CaptiaCity*		*isLocationOf, countryHasProvince, countryHasState*
Events	*InstantEvents, RecurringEvents, Election, Meeting*	*hasStartDate, hasEndDate*	*hasLocation, hasPeopleKilled*

[1] http://www.w3.org/TR/owl-features/



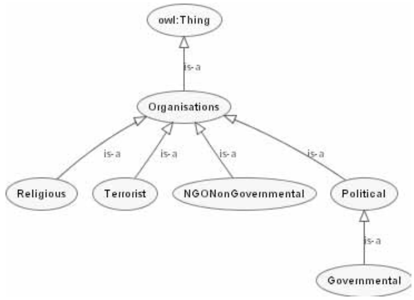


Figure 1: The figure on the left hand side
demonstrates the conceptual structure of the Knowledge Base that reflects our view of how the social entities in the network relate to each other. The figure on the right hand side demonstrates the hierarchical structure of the Organizations ontology.

Typically, when an instance is added to the ontology it inherits all the relations and the attributes defined on the class in which it is inserted into. More concretely, if Al Qaeda is declared to be the instance of the class *Terrorist Organization*, it is expected that it is equipped with the attributes *hasLeader*, *hasLocation*, *isInConflictWith* that would take on values such as *UsamaBinLaden, Afganistan, USGovernment* etc. respectively.

2.2 The Data Component

Having constructed the conceptual structure, the second step was to populate the ontologies with instances in a semi-automatic way. For that, data about events, people, organizations and locations in the domain of terrorism was used. This was done by extracting the information automatically from a relational database that has been constructed manually to store the information.

The extraction was accomplished using the RAP - RDF API for PHP V0.9.3[2] and Powl[3] that operate on the RDF[4] model. We consider the RDF model as most suitable for our purposes for at least three reasons. First of all, the RDF model is itself a network model. It is based on the so-called subject-predicate-object triplets, where the subject and the object can be considered as vertices and the predicates as edges in a graph. Secondly, it is based on the XML[5] language so it enables using information from distributed, heterogeneous resources regardless the format and structure. Hence, it can be used to gather information from various news resources in different formats on the Web The Web ontology language OWL operates on the RDF model extending it with "meaning" and enabling the incorporation of logic. Thirdly, the above mentioned APIs incorporate an inference mechanism that operates on the RDF model. Finally, OWL and RDF are considered as the relevant Web standards, for which many tools and services exist[6].

---

[2] http://www.wiwiss.fu-berlin.de/suhl/bizer/rdfapi/
[3] http://powl.sourceforge.net/
[4] http://www.w3.org/RDF/
[5] http://www.w3.org/XML/[5]
[6] http://planetrdf.com/guide/

2.3 The Application with the Inference Mechanism Involved

The Knowledge Base has been constructed by reading the conceptual structure (the ontologies) and the real world data (the instances) at the same time and by converting them into the corresponding RDF model. In a second step, queries are performed on the RDF model using the query language RDQL[7] to extract the relevant parts of the entire data resulting in so-called views. Such views can be *"All terrorist organizations at a given location"* or *"The suicide bomber in a given event"*. The APIs mentioned above incorporate the RDQL query mechanism as well as the inference engine, such that while querying the entire model to output the views, the inference mechanism is deployed and additional information is inferred. These output views containing all the information about the instances are produced in RDF format and serve as input for the social network visualization. In a final step, the whole model is serialized to a human-readable format and can be accessed and queried through a web-based user interface.

3 Visualization of the Social Network

This part of the paper describes the requirements for a web based interactive social network visualization, followed by the description of the design process and the implementation of the prototype.

3.1 Requirements Review

Bosley and Straub [BS02] suggest 10 heuristic design rules for developing interactive statistical data mining interfaces. These can be partly applied to design social networks exploration interfaces. The interface should give an overview of the available data. Metadata should be clearly defined and displayed. Short and cryptic labels should be avoided, unfamiliar terms explained. The interface is supposed to offer clear and sufficient instructions in order to facilitate its use. Users should not find themselves searching for non-existent or non-available data. If a query returns a null result the user should be informed.

As social networks are large data sets with a numerous data entry points, it is considered inefficient to offer pre-generated visualizations. The user should have the possibility to choose his point of interest and from there to interactively explore the network. As there are no well-established graph formats for social network visualizations, the design of an easy-to-interpret output format is considered a challenge. Additionally, the user should have the possibility to explore the data in textual format as "Graphics aid learning when they present the same information as text in a different format and also when they present information complementary to textual information" [BTB02]. Finally, the loading time of the application should be very short.

---

[7] http://www.w3.org/Submission/RDQL/

3.2 The Design

The social network visualization was designed for the web. The application consists of a menu which contains entities, a graph surface on which the visualization is drawn, and textual information displayed on request. The menu of the application is separated in three different parts reflecting three different ontologies: *people, events, and organisations*. Indentations detail the hierarchy of the ontology in the menu (from general to more specific). Ontology concepts and their instances are further distinguished by using different fonts. It is not only possible to hide single instances but also whole concepts. This enables the user to study the social network in detail, by focusing on people, or people that are also terrorists whilst masking other instances. Edward Tufte considers rrevealing the data at several levels of detail, from a broad overview to the fine structure as an important issue in visualizations [Tu98].

Figure 2: Screenshot of the menu showing the ontology hierarchy; items highlighted in dark gray are currently displayed in the graph surface (not showed) whilst those in light gray are masked.

Instances can be dragged on the graph surface, where their relations with other instances will be shown. In doing so, the visualized instances are clearly highlighted in the menu offering the user a better overview in comparison to menus without this functionality, e.g. drop down menu[8]. Related instances are connected with lines with labels indicating the type of relation. As the relations are not weighted, the distance between two

[8] For example, visit http://www.trackingthethreat.com

instances does not imply the strength of the relationship, therefore the symbols can be freely rearranged by the user. The ability to move elements about enhances comprehension of the relationships unlike a centred view on one entity[9]. Different symbols are used in the graph surface to facilitate the distinction between people, events, organisations, and locations. Upon mouse over, a small popup menu appears with options to get textual information, display further relations or mask the instance. The textual information is contained in separate pages and offer additional information which is not displayed in the visualization (e.g. long descriptions, date of birth, etc…). Being in another format, it facilitates understanding.

3.3 The Implementation

For the visualization, Macromedia Flash was preferred to SVG[10] as its plug-in (Flash Player) is extensively distributed [11].

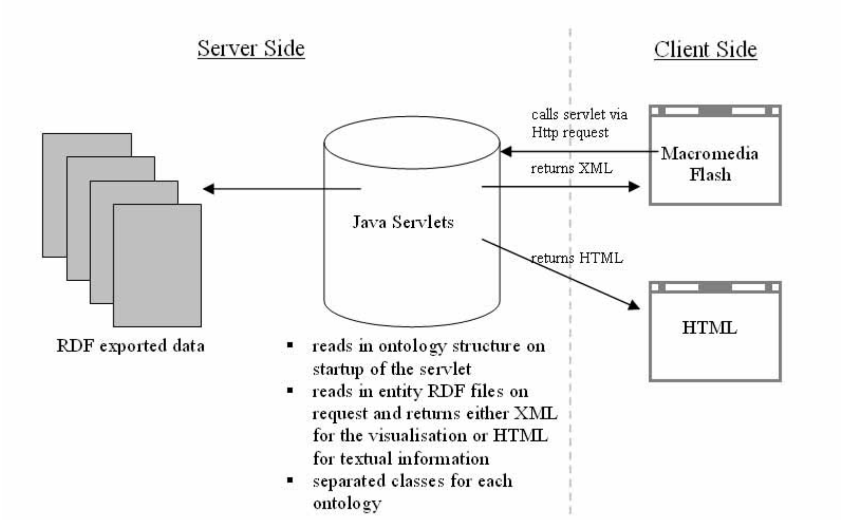


Figure 3: System Architecture

An important requirement was the separation of data modelling and administration from data visualization. New instances should be automatically added to the applications without any changes to the software, and modifications to ontologies should be easy implement able with little programming effort. Each ontology is represented by a unique Java Servlet. Depending on the Servlet parameters, it returns the menu data (in XML format), the relations for an instance (in XML format), or the detail page for an instance

---

[9] For example, visit http://www.der-mo.net/relationBrowser/index.html
[10] For more information see: http://www.w3.org/Graphics/SVG/
[11] Source: http://www.adobe.com/products/player_census/flashplayer/

(in HTML format). By changing one ontology's structure, only parts of the web application need to be adapted. As the menu structure is built by parsing the RDF file of each ontology on startup of the web application, new OWL subconcepts can be added by simply restarting it. On the other hand, new instances are added automatically without restarting. To minimize data traffic and loading time, data loading occurs only when requested. When the application starts, only the data needed to build the menu is loaded. Further loading depends solely on the user's actions, such as visualizing an instance or its related textual info.

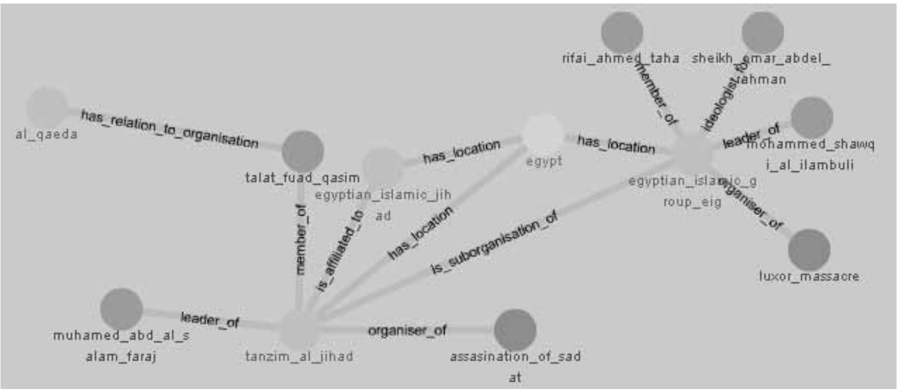


Figure 4: Screenshot of the social network visualization

4 Conclusions

We have shown that the combination of RDF models and ontologies provide a suitable means to model and visualize social networks. The deployment of inference mechanisms and the possibility to extend and refine the network with further (sub)concepts are major advantages. A second step will be to merge the RDF query possibilities and the visualization thereby visualizing the query outputs directly.

Information that is modelled and visualized this way can easily be adapted to other social networks models. An interesting example to explore with such social network models would be that of business enterprises with global reach. The entities and relationships could be discovered in such enterprises, whose employees, partners etc. collaborate heavily through the Web.

As mentioned previously, we believe our results demonstrate that working with small and well defined social network models prove to be more efficient and more easily realizable than attempting to capture too general large knowledge models.

References

[BTB02] Bauer-Morrison, J.; Tversky B.; Betrancourt, M.: Animations: Does It Facilitate Learning. International Journal of Human-Computer Studies, 57:247-262, October 2002.

[BS02] Bosley, J. J.; Straub, K. A.: Data Exploration Interfaces: Meaningful Web Database Mining by Non-Statisticians. In: Proceedings of the IBM Make IT Easy 2002, 2002.

[DF99] Degenne, A.; Forse, M.: –Introducing Social Networks Statistical Methods Series, Sage Publications Ltd., 1999.

[De06] Deligiannidis, L.; Sheth, A. P.; Smyth, P.; Aleman-Meza, B.: Semantic Analytics Visualization Lecture Notes in Computer Science, Vol. 3975 (January 2006) ; pp.48-59.

[GCF04] Gómez-Pérez, A.; Corcho, O.; Fernandez-Lopez, M.: Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web. First Edition (Advanced Information and Knowledge Processing), Springer Verlag, London, 2004.

[KSS97] Kautz, H.; Selman, B.; Shah, M.: Referral Web: Combining Social Networks and Collaborative Filtering. Communications of the ACM, 40(3) pp.:63–65.

[LLH01] Lassila, O.; Lee, T.B.; Hendler, J.: The Semantic Web. Scientific American, 184(5), 2001; pp. 34—43.

[Ne06] Newman, D.; Chemudugunta, C.; Smyth, P.; Steyvers, M.: Analyzing Entities and Topics in News Articles Using Statistical Topic Models? Lecture Notes in Computer Science, Vol. 3975 (January 2006), pp. 93-104.

[Sc04] Scott, J.P: Social Network Analysis – A Handbook, Sage Publications Ltd., 2004.

[Tu98] Tufte, E. R.: The Visual Display of Quantitative Information. Graphics Press, Cheshire, Connecticut, 16. Edition, 1998.

[Wa99] Wassermann, S.; Faust, K.; Iacobucci, D.; Granovetter, M.: Social Network Analysis – Methods and Applications (Structural Analysis in the Social Sciences), Cambridge University Press, 1999.

[Us03] Uschold, M.: Where Are The Semantics in the Semantic Web? AI Magazine, 24(3), 2003; pp. 25--36.