Empowering the Search of One's Social Directory

João Guerreiro, Daniel Gonçalves, Helena Sofia Pinto
Technical University of Lisbon / INESC-ID
Rua Alves Redol, 9. 1000-029, Lisboa, Portugal.
joao.p.guerreiro@ist.utl.pt, daniel.goncalves@inesc-id.pt, sofia@inesc-id.pt

Abstract

The frequent search for information about people comes across several difficulties in current social applications. Users still have to navigate among sources to find what they look for. Besides, it gets even more difficult when the search is about fitting broader criteria than to searching for a specific someone. In this paper, we present PersOntology, an ontology to describe people, which specifies the most prominent attributes and relations in a people search context. This specification allowed gathering information from multiple sources in a single knowledge base (KB), providing an overview of a person's Social Directory. We built an interface, based on PersOntology, able to search and browse person-related information. A preliminary evaluation enabled the improvement of the ontology towards suitability with future social-based interfaces. This preliminary assessment showed that such enriched KB enables novel and more efficient interfaces, represented here as attribute search.

Keywords

Social Directory, Person Ontology, Social Networking Sites, Profile Attributes, Person Search, Attribute Search.

1. INTRODUCTION

It is very frequent to search for information about people, mainly in Social Networking Sites (SNS) [Joinson2008] as users want to know more information about others or just keep up-to-date. In addition, not only does our personal space comprise an extensive *Social Directory* with information about friends and acquaintances, but also about people that we do not fully remember or barely know [Zhou2012]. All this person-related information can be leveraged to help users when searching for information about someone or trying to find people that fit some specific criteria. For instance, we may need to find someone interested in photography or with a background in computer science. Still, we have to navigate among several sources to satisfy our needs.

Previous research already tried to take advantage of multiple sources to support people search (e.g. [Zhou2012, Guerreiro2012b]), but do not consider the relevance of attributes, which vary in this context [Stecher2008, Guerreiro2012a]. Moreover, a specification of the attributes to consider would help merging the information from several sources and deal with their heterogeneity.

Literature reviews have documented the advantages to integrate information from multiple sources and highlighted the effectiveness of ontologies in their ability to cope with the resultant heterogeneity [Wache2001]. Ontologies enable a clear definition of the terms and relations that are important to a specific domain. Yet, current user interfaces are taking little advantage of the semantics that those ontologies convey [Fluit2006]. *Google Knowledge Graph* and *Facebook Graph Search* are exceptions in the person-search domain, but neglect the user personal space as results are restricted to *Google+* (except famous people) and *Facebook* content, respectively.

In this paper, we will present an ontology (*PersOntology*) and an interface that support the search for information about people. The clear link between them brings to bear the relevance of such scaffold when designing new interfaces. A preliminary evaluation with users, interviews and a brainstorming session enabled several suggestions to improve the ontology towards suitability with a person's *Social Directory*. This preliminary assessment showed that such enriched knowledge base enables novel and more efficient interfaces, represented in this preliminary assessment by the ability to search by attribute.

2. THE SOCIAL DIRECTORY SCAFFOLD

The *Friend-of-a-Friend*[1] (FOAF) ontology is the most popular ontology used to describe people and their relationships. However, it still lacks the ability to support a *Social Directory* that resorts to multiple sources. First, it does not consider where the information came from (provenance) and the trust on the on the correctness of each item. Second, it lacks a few attributes that were found important for this context in previous studies.

This fact led to the development of *PersOntology*, which was based on our previous study on profile attribute's relevance [Guerreiro2012a]. We do not claim that *PersOntology* is the *FOAF* substitute (can be an extension). Instead, we believe that current ontologies still need further research and that the outcomes of our preliminary evaluation will help the discussion on this subject.

[1] http://www.foaf-project.org/ (Last visit in 09/2013)

2.1 PersOntology – Purpose and Scope

The starting point to build an ontology is defining its purpose and scope (e.g. [Uschold1995]). This ontology aimed at providing a *Social Directory* resorting to the sources that contain information about people. It has to be able to specify, store and query the relevant attributes used to define people and the relations between them.

Determining the scope was divided in two phases: obtaining the key concepts/relationships and the competency questions (CQ), which are a list of representative questions that an ontology should be able to answer. As an ontology to define people, the basic concepts are the person and the personal attributes used to define her. Thus, the most important relation is *person-hasPersonAttribute-PersonAttribute* (Figure 1). As a support of the main features are relationship type, source and weight. RelationshipType arises from a direct mapping from our previous study [Guerreiro2012a]. Source designates the information source from where each specific attribute was extracted and weight the trust/relevance of that item.

The CQ are enumerated below, with a comparison to what FOAF is able to respond to:

1) *What is John's professional history?* FOAF has several properties that stem from an iterative development process. Yet, we identified additional attributes that are important for our context.
2) *Which are John's Interests? And which are the most relevant ones?* FOAF is able to answer to the first question, but not to the second. It is difficult to assure information correctness and to determine the relevance of an item. Yet, crossing information between sources (some more reliable than others) may help determining these values.
3) *How do I know that John is interested in Sports?* Without provenance information, it is not possible to know where an attribute came from. Access to this information may enable filtering the data. To cite one example, one may erase all attributes extracted from an application that contains false information.
4) *Who lives in Los Angeles? And who is a Professor?* It is possible to obtain information by attribute with FOAF, except for attributes that are not specified therein. For example, *workInfoHomepage* refers to the website that contains work information, but the information itself is not in the ontology.
5) *What friends of mine live in Los Angeles?* FOAF relies on the property *knows* to establish relations between two persons, but this question specifies friends. Although the *xfn ontology* extends this property with other options, we rely on the relationship types used on our previous study: friend, acquaintance or famous.

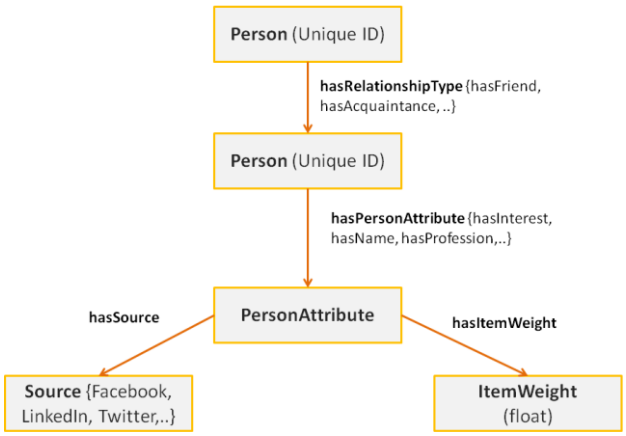

Figure 1. *PersOntology* conceptual model

2.2 PersOntology – Conceptual Model

To define our ontology we used Protegé[2] for the Web Ontology Language (OWL)–DL. Figure 1 shows the ontology concepts and their relations. The user has a *RelationshipType* (responds to CQ5) with other person that has several *PersonAttributes* from [Guerreiro2012a] (CQ1, CQ4). The *PersonAttributes* contain the properties *hasWeight* and *hasSource* (CQ2, CQ3).

3. PROTOTYPE

We developed an interface to search and visualize a user's *Social Directory*, which depicts the information extracted from 3 SNS (*Facebook*, *twitter* and *LinkedIn*). This interface was used in a preliminary experiment where users could search for either people or attributes. Such artifact was a means to discuss the ontology and its capabilities to support a person's *Social Directory*.

3.1 Populate the Ontology

We used the Python Language and the *Fuxi framework*[3], built on *rdflib*, to both data extraction and management. They allowed us to populate the ontology. When new data is extracted from one of the sources, it is mapped as a graph of instances, which is used to reason. This step allows inferring logical consequences from a set of facts.

3.2 Search for People and Attributes

One may search both for people and for a specific attribute. When searching for a person, the results page presents all the attributes about her (Figure 2). Therein, the results are grouped by attribute and are by default collapsed. Figure 2 depicts some of *John Smith*'s attributes.

Searching by attribute arose from the need to obtain information about people that are related with a specific attribute. In current solutions, we have to navigate among and within applications. In fact, this feature enables reaching to information that could otherwise be missed, enhancing the quality of the information retrieved. Figure 3 shows a search for *Profession*, which presents all professions on the *Social Directory*.

The interface does not show the items weight and sources. We wanted to perceive if the participants of the upcoming experiment would notice their absence, thus helping to determine their relevance.

4. PRELIMINARY EVALUATION

A first attempt to evaluate the ontology assured that it could answer to all CQ. We populated the ontology with data from 2 different persons and performed queries to the ontology using *Sparql*, a query language for *RDF*.

[2] http://protege.stanford.edu/ (last visit in: 09/2013)

[3] http://code.google.com/p/fuxi/ (last visit in: 09/2013)

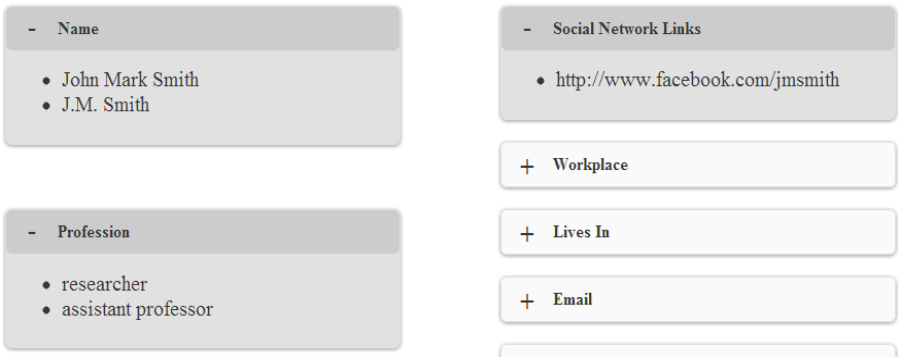

Figure 2. Part of the result's page of a search by person (*John Smith*)

Our main objective with the preliminary evaluation was to understand the strengths that can be leveraged by *PersOntology* and other ontologies with similar purposes. Moreover, we wanted to identify possible improvements to such ontology able to represent a person's *Social Directory*. We gathered 4 *Computer Science* volunteers, which are familiar with web search and SNS, to take part in a study that included the exploration of our interface, an interview and a brainstorming session.

4.1 Explore the Interface

Participants explored the interface without information regarding the ontology. The tasks were performed using the *think-aloud* method, where users were encouraged to mention difficulties and suggestions about features they miss or dismiss. The tasks they had to perform were:

1) *Extract Data* - Use the interface to extract information from *Facebook*, *Twitter* and *LinkedIn* accounts.
2) *Search for* a *Person* – A person of their choice.
3) *Search for an Attribute* – An attribute of their choice.
4) *Search for a Specific Person* – A friend that all users shared and find his *Profession*. It aimed at providing a common ground for brainstorming.
5) *Search for a Specific Attribute* – Similarly, it aimed at providing a common ground; they had to search for the attribute *Profession* and find all *Researchers*.
6) *Search in SNS* – Search for the same person and attribute of 2) and 3). It aimed at a comparison with the direct use of SNS.

4.2 Semi-Structured Interviews

Participants were encouraged to mention important features that our interface did not provide in comparison to SNS, its strengths and possible improvements. We also brought to discussion the value of adding the sources and weight. While they were exploring the interface and thinking-aloud, we gathered a set of questions regarding their comments in order to get detailed explanations.

4.3 Brainstorming Session

The brainstorming session was conducted in the same day of the remaining phases and included the same 4 participants and the interviewer (one of the authors). The topics for discussion were based on the most mentioned features in the interviews and those with more disagreement. We wanted them to reach a consensus about improvements to be made. At first, this discussion was centered on the interface since users had not seen what was on the background. Afterwards, their awareness of the ontology provided more topics for discussion. The use of a white board helped participants exposing their opinions, mark their differences and try to reach a consensus about the features that should be removed, added or changed.

5. RESULTS

Both interviews and brainstorming sessions provided good insights on the advantages and possible future improvements. Herein, we discuss the most relevant ones.

5.1 Searching by Attribute Rocks!

Most SNS contain information about people's attributes. They are prepared to answer to a search about a person and it works fairly well when we know whom to search for. Although, users found our interface "*good to search for a specific person since it is easier to find the information we are looking for*", they were very excited when performing the attribute search tasks. The ability to search by attribute was found the greatest advantage supported by the ontology. Participants mentioned they "*could not perform the same task resorting to any application they use*" and that they "*remember occasions where it would be useful in the past*" and "*imagine several scenarios where it may be very useful*".

5.2 What about Interactions?

All participants claimed that information about interactions and social activity would be a valuable addition.

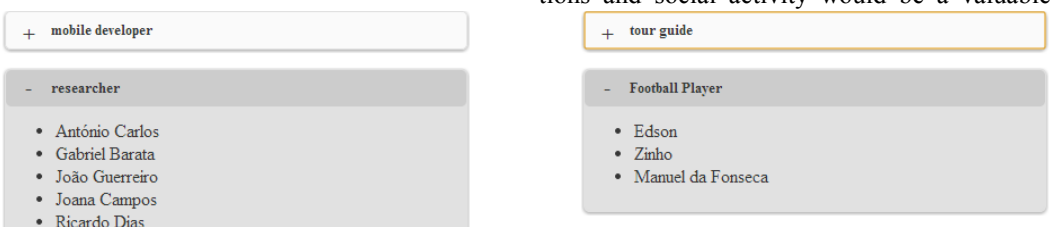

Figure 3. Part of the result's page of a search by attribute (*profession*)

Since it was so popular, they discussed about which information should be considered if it was to be stored. Participants reached a consensus (Figure 4) about the most important features to store for each interaction: *who*, *when*, *what*, *where* and the *content*. *Who* states the people involved in that interaction and can be classified as *public* (e.g. a tweet), *restricted* (e.g. e-mail for multiple receivers) and *private* (between two persons). *When* states the date and time of the interaction and *Where* regards to the information source (e.g. *Gmail*). *What* indicates the type of interaction, such as "received e-mail" or "sent Wall Post" and the *Content* contains the interaction itself. They found the most recent updates (even if *public*) more important, as they felt the need to know what others have been doing, a need mentioned in [Joinson2008].

5.3 Sources for Contradictory Information

The usefulness of knowing where the information came from was an assumption we made when building the ontology. In the brainstorming session, participants agreed that when the information is believed to be true, knowing the source is not that important, but when there are contradictory results it is important to know where the information comes from. In light of this, we believe that it is useful to maintain this information in the ontology, but only show it in particular situations or on demand.

5.4 Weight for Trustfulness and Relevance

When building the ontology we considered the weight of information items, thinking on how relevant and trustworthy it is. When asked about this feature users considered it as two separate things (trustfulness and relevance). Some information can be more consistent and trustworthy if it appears in many sources, but it does not mean it will be more relevant for the user. That information may have to take into consideration the user's interests and PI.

5.5 Interface Potential

The interface is very simple, but the users mentioned several opportunities enabled by the ontology. First, the search could benefit from the use of contextual cues, such as current location. To cite one example, when travelling to *LA* and searching for some attribute, it may present first the people living or working in *LA*. Second, they agreed that the results' page should consider other semantic relations, such as presenting all professions related to computer sciences together. Third, the participants reported difficulties to obtain an overview about the search results. Most participants referred to the impact of the top section of the *Facebook Timeline* as a good example of graphical information used to aid in a rapid overview. We believe that the best solution is to combine both and provide means to get an overview at first, but also enable this more detailed search. Fourth, they suggested us to allow filtering multiple attributes to find someone.

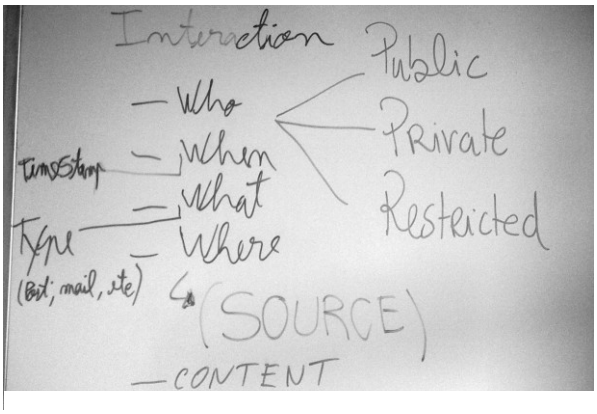

Figure 4. *Consensus* on interactions' properties

6. CONCLUSION

We have presented *PersOntology*, an ontology that specifies the information to be extracted and stored in order to build a *Social Directory* to back-up people-search. Upon that, we built an interface that collects information from 3 SNS and supports searches by person and attribute. A preliminary evaluation where participants had to search for people and attributes, and then brainstorm about it, led to suggestions to improve *PersOntology* and other ontologies with similar purposes. Searching by attribute was considered our interface's main feature, as it enables finding people that fit a certain criteria.

7. ACKNOWLEDGMENTS

Project supported by the Portuguese FCT: individual grant SFRH/BD/66550/2009; project Pest-OE/EEI/LA0021/2013 and project PAELife AAL/0014/2009.

8. REFERENCES

[Fluit2006] Fluit, C. et al (2006). Ontology-Based Information Visualization: Toward Semantic Web Applications. In *Visualizing the Semantic Web*, 45–58.

[Guerreiro2012a] Guerreiro, J. and Gonçalves, D. (2012). "Tell me What I Want to Know!": The Effect of Relationship Closeness on the Relevance of Profile Attributes. *Proceedings of the ACM Conference on Hypertext and Social Media*, 230-235.

[Guerreiro2012b] Guerreiro, J. and Gonçalves, D. (2012). Visualiz'em: "Show me more about him!" *Proceedings of the International Working Conference on Advanced Visual Interfaces,* 445-448.

[Joinson2008] Joinson, A. N. (2008). Looking at, Looking up or Keeping up with People ? Motives and Uses of Facebook. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1027-1036.

[Stecher2008] Stecher, K. and Counts, S. (2008). Thin Slices of Online Profile Attributes. *Proceedings* ICWSM.

[Uschold1995] Uschold, M. and King, M. (1995). Towards a Methodology for Building Ontologies. In Skuce, D., *Methodology*, volume 80, 275–280, Citeseer.

[Wache2001] Wache, H. et al (2001). Ontology-Based Integration of Information:" A Survey of Existing Approaches. *Proceedings of IJCAI Workshop: Ontologies and Information Sharing*, 2001(4):108–117.

[Zhou2012] Zhou, M. X. et al (2012). Finding Someone in My Social Directory Whom I Do not Fully Remember or Barely Know. *Proceedings of the ACM international conference on Intelligent User Interfaces*, 203-206.