NEW PERSPECTIVES FOR SEARCH IN SOCIAL NETWORKS
A Challenge for Inclusion

Júlio Cesar dos Reis
Institute of Computing (UNICAMP) and CTI Renato Archer
Rodovia Dom Pedro I, km 143,6, 13069-901, Campinas, São Paulo, Brazil

Rodrigo Bonacin
CTI Renato Archer and Faculty of Campo Limpo Paulista (FACCAMP)
Rua Guatemala, nº 167, 13231-230, Campo Limpo Paulista, São Paulo, Brazil

M. Cecilia C. Baranauskas
Institute of Computing at University of Campinas (UNICAMP)
Caixa Postal 6176, 13083 970, Campinas, São Paulo, Brazil

Keywords: Semantic Web, Semantic Search, Inclusive Social Network Services, Organisational Semiotics.

Abstract: The world is populated with many scenarios characterized by a diversity of cultures and social problems. Thus it is necessary to investigate computational solutions that respect this diversity. The use of search engines is one of the main mechanisms to provide the access to information generated in the Social Network Services (SNS). These mechanisms are currently built through lexical-syntactical processing resulting in barriers for many users to access correct and valuable information in the Web. Novel search mechanisms could effectively help people to recover and use information through Inclusive Social Networks Services (ISN), promoting the universal access to information. This paper shows results of search activities in an ISN that point out how to improve search engines considering aspects related to social and digital inclusion. Inspired in these results, we outline an approach based on Organisational Semiotics to build Web ontology, which is used by an inclusive search engine drawn up in this paper. Actually, this proposal combines different strategies to provide better search results for all.

1 INTRODUCTION

Web systems and portals are available to a vast number of people with socio-cultural differences. Within a Social Network, people communicate and behave according to commitments, linguistic constraints, culture and other social aspects. Social Network Services (SNS) of nowadays could be more appropriate to the context of people´s life, considering their differences, in order to promote social and digital inclusion.

This scenario becomes even more challenging and important in contexts such as Brazil, where there are several social problems and a huge cultural diversity. In this perspective, one of the fundamental points is to provide barriers free access to information to every citizen. This could help to bring great social benefits and contribute to a profound social transformation. Thus, it is extremely important that all people have the opportunity to retrieve, access and use information provided in the digital media in a smooth way.

The SNS represents an opportunity to interaction, access to information and knowledge through the Web. These systems primarily allow individuals to share their interests and activities, constituting communities. The e-Cidadania project (Baranauskas, 2007) aims at transforming a SNS into an engine for digital inclusion and citizenship. The networks with such characteristics can be defined as "Inclusive Social Network Services" (ISN) (see Neris *et al.*, 2009). The use of search

Reis J., Bonacin R. and C. Baranauskas M. (2010).
NEW PERSPECTIVES FOR SEARCH IN SOCIAL NETWORKS - A Challenge for Inclusion.
In *Proceedings of the 12th International Conference on Enterprise Information Systems - Human-Computer Interaction*, pages 53-62
DOI: 10.5220/0002897000530062

engines is one of the primary ways to find and to make access to information generated in these systems. However, search mechanisms are currently built based on comparisons of keywords and lexical-syntactical information processing (syntax search). These mechanisms are not sufficient and adequate to effectively make sense to individuals in an inclusive scenario within social networks.

Based on empirical results, which will be discussed in this paper, we have observed that people organized into virtual communities bring to this space their own vocabularies and meanings, and also develop their own local vocabularies through interaction and communication using technology. The results pointed out the need for novel search mechanisms considering the diversity of users' competencies and inclusion aspects.

A more appropriate inclusive search solution to an ISN should reflect the semantics used by participants of the social network. In few words, a search engine should take into account the local meanings created, shared and used by people organized into a community. In this paper we argue that the quality and response accuracy of a search mechanism are intrinsically associated to the proximity of the semantics shared by people. Thus, it is necessary to identify the meanings used in the network and to represent its semantic aspects. This could actually contribute to make the information accessible to everyone, including people with low educational levels that have difficulty to access online information due to their simple vocabulary or their deficiency in writing. Usually, these people use an informal (colloquial) or local vocabulary in the search. With the proposed solution they could find the correct information in an easier and more precise way, and learn from it.

In this paper we show results of search activities within an ISN, conducted in the context of the e-Cidadania project. The goal of these activities was to observe a set of search scenarios with potential users of an ISN, and to understand how these users make sense of a search mechanism. Based on the results, we present a proposal of developing a more appropriate search mechanism for an ISN with foundations in Organisacional Semiotics (OS) (Stamper 2001; Liu 2000). In our approach the goal is to expand and to improve the search technologies and techniques of the Semantic Web field. Besides, the representational structure (semantic model) used by the search mechanism is based on data from the interaction and communication of users in the social network system. Thus, the search engine will take into account the meanings shared and created by people (including the informal terms) in their interaction with the system aiming to provide better results.

The paper is organized as follows: Section 2 presents the concept of ISN and the importance of search mechanisms for the universal access to information; Section 3 presents the analysis of the empirical experiment with ISN users; Section 4 details the proposed approach; Section 5 makes a discussion about the approach and related works; and Section 6 concludes presenting further works.

2 UNIVERSAL ACCESS AND INCLUSIVE SEARCH

According to Boyd & Ellison (2008) since the beginning of the Social Networking, sites such as *MySpace, Facebook, Orkut* and others have attracted millions of users and many of them have integrated these sites into their daily practices.

Online Social Networks or "communities of members" have great relevance in the Web as users spend much time of navigation on them. According to Nielsen (2009), social networks are more popular than e-mail, with 66.8% of global reach. Around the world it represents the fourth most used resource in the Internet and 85.2% of penetration are in the portals and communities of general interest. Additionally, 85.9% of Internet users use search engines, which is one of the most common activities.

Despite these great numbers and the success of Social Network sites among Internet users, in social contexts like Brazil and other developing countries, there are yet a lot of people without access to the Internet and consequently without opportunities to access information and knowledge. Social indicators shown by the PNAD (National Survey by Household Sample) produced by IBGE (Brazilian Institute of Geography and Statistics) (IBGE, 2008) points out that in 2008, 65% of the population did not have access to the Internet.

In addition, important data from the Ministry of Education in Brazil (MEC, 2007) reveal that about 30 million of Brazilians are functionally illiterate, defined as the population over 15 years old and less than 4 years of schooling (21.6% of the population). Using a broader concept of functional illiteracy, according to a survey from Paulo Montenegro held in 2007 (IPM, 2007), the majority (64%) of Brazilians between 15 and 64 years old and more than 4 years of schooling reach no more than the degree of rudimentary literacy, *i.e.*, they have only

the ability to locate explicit information in short texts or make simple math, they are not able to understand longer texts. This data illustrates only part of the challenge that we face in terms of designing systems, which should include all these users. In this context it has become a major concern to allow access to online content available from SNS to all people in a more "natural" and efficient way.

Thus, it is extremely important to recreate methods to permit the effective access and use of information conveyed in digital media, for all. This could be materialized with the ISN concept. We understand ISN as a "virtual communication space" based on the concept of social networks, which is inclusive and allows the community to share knowledge about the community know-how. This space has to facilitate "exchange" (of knowledge, goods and services) in accordance to the collaborative (project team, partners, community) system conception.

It is also worth to mention that in an ISN there are not target users, but all users are relevant and should be included without discrimination. Thus, there may be people without skills to handle certain technological features of the system and consequently without knowledge to find information that they need in the system. Moreover, those users most often use colloquial terms to express themselves through the system. For example, they may use the term "*postinho*" (in Portuguese) instead of "Basic Health Unit" (formal). They use terms that make sense to them, but in fact, these expressions semantically mean the same. So when someone is trying to retrieve information from the ISN, these factors must be taken into account by the search engine. On the other hand, when a user searches for something in a non-formal or not refined way and, the same concept but in its formal way (cult) is returned, this represents an opportunity for learning.

Therefore, we should seek for a computational search solution that takes into consideration the meaning that is adopted or emerges in the context of use of that network; *i.e.* the meaning that people bring to the network, and those that are constructed by using the system over time (through interaction). This may facilitate and provide better access to the content generated by users of the network.

3 ANALYSING SEARCH SCENARIOS OF AN ISN

From a practical point of view, the e-Cidadania project resulted in the *'Vilanarede'* (www.vilanarede.org.br) ISN system. This system has represented an opportunity to investigate the interaction behavior of representative users in a developing country. As a direct activity of the project, we have conducted the 8th Participatory Workshop, in a telecenter located at 'Vila União', neighborhood of the Campinas, Brazil. In this workshop we developed an activity related to search in the ISN. The objective was to observe some major points including: (1) How would the users build understanding of the search engine? (2) Which key-words would they use? (3) Would they have any difficulty in completing the proposed scenarios? and (3) What would be their satisfaction with the search results?

A task sheet with 4 search scenarios was presented to each pair of participants, and a form was prepared to the observers (researchers) of the activity. We had 7 pairs of users in total. An initial instruction about the activity was given to the participants. The pairs were formed by the users themselves, and for each scenario the pair should write the words used in the search and the title of the announcements found. Resulting from this activity, we had both the sheet tasks filled by the pairs of users and the observation forms filled out by the observers. Besides, the activity was filmed and there was audio recording of each pair during the task execution. The 4 search scenarios were:

Scenario 1: Find out announcements on how to popularize the 'VilanaRede'.

Scenario 2: Find out announcements of mango (fruit) in 'VilanaRede'.

Scenario 3: Find out announcements related to food in'VilanaRede'.

Scenario 4: Find out announcement related to religion item combined with handicraft in 'VilanaRede'.

Each scenario intended to verify whether semantic capacity was needed for the search mechanism. The time for the completion of the scenarios was approximately 45 minutes. After the execution of the search scenarios, a general discussion was conducted to get the general impression of users about the activity. During this discussion, several interesting stories were collected.

In Scenario 1, we wanted to observe whether users would use synonymous of "popularize" to find the announcements. Some pairs had difficulty to understand the scenario, as well as difficulty in choosing the terms for the search. However some pairs associated the word "popularize" to "divulge"

and quickly found related announcements. In this scenario one pair used some unusual keywords such as: "*boca-a-boca*" (a popular expression used in Brazil that means "orally passing information from person to person"), "email", "phone" and "posters". By using the term "*boca-a-boca*" in order to find announcements about how to divulge the site, unusual results also appeared as an advertisement for "*Bife de casca de banana*" (steak of banana peel). It happened because in one of the comments of this announcement we find "I'm with water in my mouth (*boca*)" in reference to the announcement of "steak of banana peel". Phrases for search like "divulgation of the *'Vilanarede'*" or verbs such as "to popularize announcement" or "advertising Vila" were also used in this scenario.

In Scenario 2, we wanted to verify if users would find any announcement related to mango (fruit) in the application. There was no announcement about the mango fruit in the system. However there was an announcement about *mangá* (cartoon) and it was written without the acute accent in the word ('*manga*' in Portuguese, which is mango fruit in English). In this scenario, users mainly used the keywords: "mango (fruit)", "fruit", "mango", "mango fruit", "mango / fruit". Some pairs were uncertain if they would have to put the keyword "fruit" or not. Note that in a semantic search, by putting the key-word "fruit", the application should return all the announcements with mango (fruit), in the case of announcements semantically related to fruit.

In Scenario 3, we wanted to see whether users would use the keyword "food" in the search or they would make a search for specific foods through the search engine. As a result, when users tried the keyword "food", the system returned nothing. However there are several announcements on food in the system: the sale of "*salgadinhos*" (homemade snacks), "*pão de queijo*" (cheese bread) and others. Among the relevant considerations from the observers, during the execution of this scenario users said that the system should relate "*salgadinhos*" (homemade snacks), "*pão de queijo*" (cheese bread) and "*Bife de casca de banana*" (steak of banana peel) with the concept of food. And this makes sense because semantically all of these are food. During the discussion phase one of the users commented: "Using food is easier because it already covers everything," *i.e.,* all types of food in the system. Another said: "To be more 'lean' and practical for those who are starting (in terms of computer literacy), like us, when we enter "food", it should return a variety of foods due to our difficulty." Yet another user says: "Maybe to use food does not help in the search of something more specific, but if it is something that we have no knowledge of the domain, or we do not know what to look for, the tool would be useful and helpful." The main keywords used in this scenario were "food", "*comida caseira*" (homemade cooking), "food sale", "*salgado*" (homemade snacks), "*salgadinhos*"(small homemade snacks), "*salgadinho frito*" (fried homemade snacks), pies, "*doces*" (sweets), "*pão-de-queijo*" (cheese bread), "*docinhos*" (small sweets), cake, pastel and "*brigadeiro*" (chocolate sweet). Note that users utilize several variations in words such as "homemade snack", "small homemade snack" and "fried homemade snacks".

With Scenario 4, we aimed to determine which keywords users would use when looking for a specific announcement. One of the observers indicated that the pair found the "Saint Anthony" because they already knew that this announcement was in the system. The same was reported by several other observers. The vast majority of the pairs used the keywords: "homemade craft", "Crafts saint", "holiness", "holy" and "saints". Users found the desired information successfully. But one of the pairs put keywords like: "Orisha", "Orisha of cloth", "religious" and "sculpture" and didn't find out any announcement. Several observers noticed that the subjects utilize terms from their own colloquial language in the search; examples can be seen as "*manga rosa*" (pink mango), "*manga coquinho*" (coconuts mango), "*tutu de feijão*" (tutu bean), "*boca-a-boca*" for the term "divulge", "small sweets", "little homemade snack" and "Orisha". Also in several occasions the pairs discussed before reaching an agreement on which word to use in the search.

Another interesting result was obtained by the analyze of the interaction of a deaf-mute user with search mechanisms. We could observe that this user has some difficult related to word meanings. We observed that he does not know some words, and use the same hand signal to several different words; he has difficult in writing and make different meaning associations, because what make sense for them (many deaf-mute people) is not words, but signals. The user had difficult in understanding the scenario 1, since the words popularize, advertise, advertisement and disclose have the same hand signal in his language. Moreover we could see that the user understands different some words that has the same meaning; his behavior during the search was not confident neither independent on performing the tasks, and the user asked a lot of

questions to the observer.

Additionally, general results indicate that users from this context under study (prospective users of an ISN) had difficulty with the search button; in other words, they do not have a clear concept about the act of "searching" in an application on the Internet. Some users had no idea about the scope of searching. They do not know if that search is only for the announcements in the *'Vilanarede'* system which they are using or whatever the term they type the system will return successful results. This stay explicit in a description from a user who said: "Search fondée because it is something chic". However another report from other user says: "Fondée is very chic, we do not have it here in our network... we will not put it in the search because the network is ours, it is "poor"... and it will not have fondée...". This statement shows that the second user has the notion of the search scope, which will be just within the announcements from that social network system, so as there were any announcements about fondée, nothing would be returned.

Even with this lack of sense about the search scope, one of the observers explains that users stay surprised with the power of the search, and they explored and tested it easily. This is confirmed due to the users' behavior during the execution of the activity. Positive results were obtained from it. From the data filled by the observers' formularies, it is possible to indicate that approximately 80% of the pairs proved be safe during the performance of the task. Approximately 80% of the pairs proved be independently by performing the scenarios. Around 60% of the pairs did not make a lot of questions to the facilitator during the task. In the discussion after the activity, users said that they had to reason more to perform the search, making exchange of words and testing various alternatives. More details from this experiment can be seen in Reis *et al.* (2010a).

These practical results show that users' colloquial language should be considered during the development of more appropriate search engines. Moreover, people in a social network can create their vocabulary, sharing meanings in the community. The results showed us that it is necessary to construct computationally tractable models from the semantic point of view that come out from the network itself. Semantics here is understood as the interpretation of signs (Peirce 1931-1958, cf 2.228) by individuals and their association with real world elements. This interpretation is socially contextualized; *i.e.* individuals and communities may have different interpretations for the same sign and a sign may connote different meanings depending on the context applied.

As a theoretical reference for our proposed approach, we have used the Semantic Analysis Method (SAM) from the Organisational Semiotics (Liu, 2000). This method assists the users or problem owners in eliciting and representing meanings in a formal and precise model. An Ontology Chart (OC) describes a view of responsible agents in the focal domain and their behavior or action patterns named affordances (Liu, 2000). Some basic concepts of SAM adopted in this paper are based in Liu (2000). In the next section we outline a new approach to build Web ontologies based on this method as well as a novel search mechanism that utilizes it.

4 TOWARD A NEW SEARCH MECHANISM

General difficulties faced by users to get information in the Web can be mainly explained by: (1) overload of information presented in the system; and (2) problems related to the contextualization of meaning for the terms used. As an attempt to solve this problem, we have investigated approaches that can result in better and more appropriate search engines for ISN.

In a social network the "emergence" of meaning is an ongoing process in which meanings and interpretations are constructed, used and shared through the system based on the interactions and expressions of users. These interpretations expressed by users in the system could be computationally represented. Several improvements could be achieved such as semantic models that can make the social network context more faithful resulting in more adequate search engines.

To accomplish that, we propose a search engine informed by a Semiotic approach. We have developed a semi-automatic process to model the semantics in the ISN using SAM, and we use the outcome of this process in the search engine.

4.1 Modelling the Ontologies for ISN

In the *'VilanaRede'* system, users express themselves through their profiles, announcements of products, services and ideas created by them; and they communicate mainly through commentaries about the announcements and chats between

members of the network. These data are stored in the ISN system database and from these dada we represent the semantics used in the social network in a structure called 'Semiotic Web ontology'.

This structure is a semantic model (computationally tractable ontology) constructed from a semi-automatic process based on the SAM along with the vocabularies shared in the social network. The theoretical and methodological concepts described in the SAM are used in conjunction with other technologies from the Semantic Web field to describe computationally tractable ontologies using the Web Ontology Language (OWL) (W3C, 2004). The idea is to incorporate the concepts of particular Agents (roles) and Affordances (pattern of behavior) arising from the SAM into an expanded and more representative Semantic Web ontology. It is worth to mention that it is not the goal to create a "perfect ontology" from a theoretical point of view, but to produce practical and immediate results for search in ISN. Therefore some properties from the OC may not be fully transcribed to OWL, while other aspects such as agent-affordance relationship are emphasized.

This approach is justified from a Semiotic perspective, since the signs are socially constructed. Thus, a computational model that represents the semantics of an SNS should contain the agents that interpret the socially shared concepts. With this approach we incorporate and take into account to Semantic Web ontologies concerns and possible representations arising from the Ontology in a Semiotic perspective. In addition to agents and affordances, we have observed that Semantic Web ontologies also do not incorporate (at least explicitly) the idea of ontological dependency relations, the existential relation in the model. The approach is also justified by the representational limitations shown in literature (*e.g.* Tanasescu & Streibel (2007)) regarding the use of ontologies in computing and their expressivity.

As a conceptual model, in this computational approach based on SAM, the agents have behavior(s) (affordances) related to a concept. For example, a seamstress, which is an agent, can sew a "*manga"* (it means sleeve in English). Sewing is a pattern of behavior of a seamstress (in other words an affordance). "*Manga*" is a concept that can have several different meanings in Portuguese (It can mean sleeve, fruit, color, etc.), but in this context due to the affordance and the agent, the meaning of "*manga"* is more closely linked to shirt and not, for example, to "*manga"* fruit (mango in English) that can also be represented in the model, as illustrated by Figure 1.

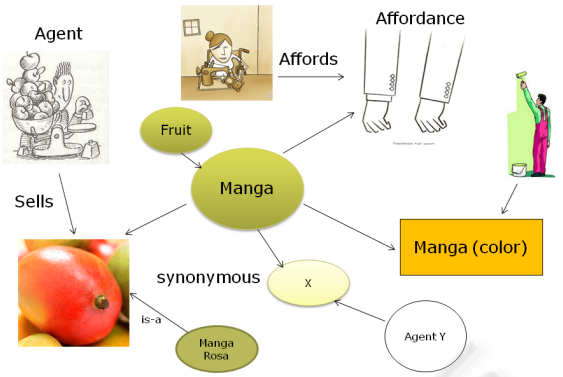

Figure 1: Modelling meanings: an example of polysemy using agents and affordances.

Figure 1 illustrates an example of modeling in which the grocer and the seamstress are agents that have affordances connected to specific concepts. And this model can also have specific 'is-a' relationships; for example '*manga rosa*' is a specific kind of mango. This also shows that concepts can be related to several agents and affordances and with other concepts, constituting relations and representations that make more complete ontologies compared to conventional ontologies described purely for a domain. For example, '*manga*' can also mean a color for a painter who is searching something in the network, as well as '*manga*' can have any synonym that makes sense for an agent 'Y' modeled from the data of the social network. We can see other examples like: 'crane' mean a bird or a type of construction equipment and we can model it using the agents and their affordances in a 'Semiotic Web Ontology'.

To develop this representation, we propose an assisted method (semi-automatic) with several distinct steps, which is a novel approach to create ontologies; the method is illustrated in Figure 2. It includes: (1) the extraction of terms from the database of the ISN system; (2) the creation of an OC (from SAM); and (3) the creation of the final OWL ontology.

In this assisted process, the first step is to process the data from the system database. This step takes into account the social relations in the network, and provides the necessary well defined data (a list of concepts, agents and affordances, etc) to build the semantic model. Related to this phase, we apply algorithms for text extraction and analysis and data from the database of the ISN system. For this we propose the use of keyphases extractors like KEA (Medelyan, 2006) and tools for term extraction and creation of concepts such OntoLP (Ribeiro & Vieira, 2008).

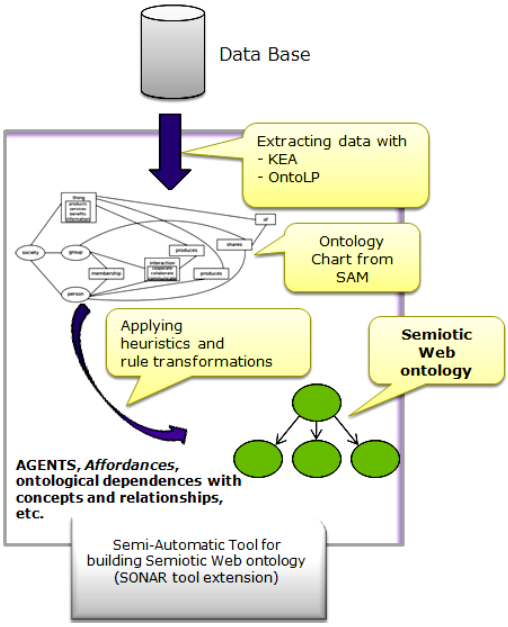

Figure 2: Illustration of the Semi-Automatic Tool.

The next step is the building of an OC (from SAM) by an ontology engineer. This intermediate ontology diagram is important to identify the possible agents in the ISN and their patterns of behavior.

In the third step, from the OC, a set of specific heuristics and rule transformations is applied to derive an initial OWL ontology (computationally tractable), extending the computational development of the SONAR CASE tool (see Santos *et al.*, 2008). Bonacin *et al.* (2004) proposed a heuristic to transform OC into system design diagrams; however those heuristic must be adapted to our purpose. In this step the ontology engineer can also be assisted by existing tools for Ontology Learning and other data extracted from the system. As suggested by Maedche & Staab (2001), Ontology Learning can be helpful for ontology engineers to build these artifacts. In the next sub-section we draw up an Inclusive Search mechanism.

4.2 Outlining an Inclusive Search Engine

After the creation of the semantic model, it is used by the ISN search engine. When the user is logged in the social network system and he/she enters with some search term(s) in the search engine, the system starts a process of making relationships of this/these keyword(s) with the available 'Semiotic Web Ontology'. The purpose is to perform a search in the database system and to return semantically related results. For example, suppose the user types the term "small snack". If there is nothing in the system with this expression, from the analogies and semantic relations made, the system may return some other types of food semantically close. Likewise if the user enters the word 'food', all advertisements related to food should be returned.

There are several architectural proposals for semantic search solutions, as described in Mangold (2007) and also in Fang *et al.* (2005). The decisions and architectural strategies for resolving the semantic search in this implementation is carried out in accordance with the requirements of an ISN. The main difference in the search solution of this proposal is to take into account information regarding the user that is making the search (from his/her profile) and the user that produces the content. See Figure 3 that illustrates this idea.

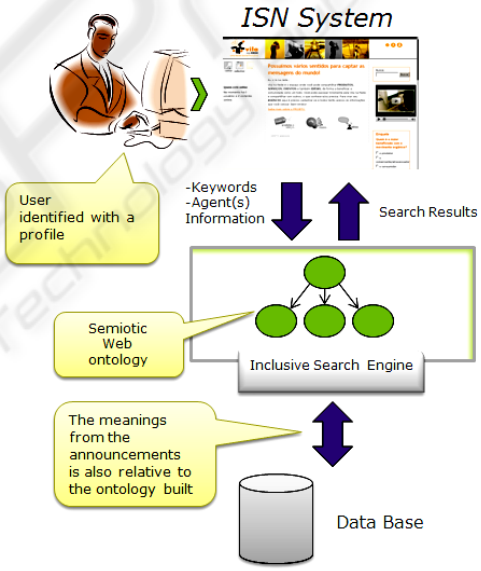

Figure 3: Illustration of the Search Mechanism.

In this strategy, the user profile is important due to the possibility of discovering a context for the search terms. From the user profile, we aim to identify the agent represented in the ontology. Thus we can prioritize (or even limit) the search space, making a relation between the user and the semantic model generated. For instance, imagine that a biologist is logged into the system (we could find that a user is a biologist based on his/her profile) and requests a search with the keyword 'crane'. If there exist a relation between the 'biologist' agent and the term 'crane' in the ontology, most likely the results (announcements) that could be returned first (ranked

first) should be related to the concept of crane as a 'bird', not the other meaning(s) of this word. However, to a civil engineer that searches the same word, maybe the results that most interest him / her refers to the construction equipment and not the bird. We do not mean that other results are not required or may not be returned in response to the search, (the engineer may want to know about this kind of bird), but the announcements from the social network that relates 'crane' with a construction equipment must have greater relevance in the ranking of results.

The agent-affordance relation is also used to indicate the probable meaning of the words in the announcement. For example, we could verify whether the word 'crane' is about 'bird' or 'construction equipment' in a particular announcement that mentions the word 'crane' based on the user that entered such information, or based on the commentaries made on this announcement. Then, if the user who submitted the announcement is a 'biologist' agent, 'crane' would be most likely a 'bird' in this announcement. In a similar way, if the advertiser is a civil engineer, also 'crane' would mean most likely a 'construction equipment'. We also could have relationships between agents and could verify how much an agent is semantically close to another and indicate the probable meaning based on this aspect.

5 DISCUSSION

The support for better results from the search engine in this approach demands a careful modeling procedure. Different signs with the same meaning (synonyms) coming from different virtual communities of the social network can be discovered having the opportunity to be represented in the ontology; such signs and meanings can be purely regional. They could not be present in formal dictionaries or thesauri generally used by conventional search mechanisms. Furthermore because they are cultural expressions emerging from the social network, the ontology would potentially provide smarter and richer search results when compared to ontologies based on domains or formal definitions.

The approach provides means to discover and distinguish the meanings used in the SNS, representing them through the agents in the 'Semiotic Web ontology'. Differently from conventional computing ontologies, and other approaches to semantic representation, our proposal involves adding the agents and affordances concept. This addition can cooperate for richer search results treating the polysemy problem in not restricted or controlled language contexts. Moreover, the inclusion of the agents and other concepts from SAM in the Web ontology can help improving the search mechanism, generating results more adequate to a SNS context.

In the case of users with limited literacy and with difficulty in dealing with technological artifacts (digitally illiterate), it is important to give to them the opportunity to perform the search using their daily language since usually is what make sense to them, and to provide search results more natural and adequate to their live. Thus, the search engine should reflect the semantic reality of the social network users, and therefore the required information will be easier to find. A search engine with such characteristics could create opportunities for inclusion, since the method for building the semantic model as well as the strategies in the search engine to use the ontology suggests that the search results returned will tend to make more sense for the user that searches.

Some recent studies in the literature address search solutions for SNS (*e.g.* Vieira *et al.* (2007), Perisic & Haynes (2009) and Gürsel & Sen (2009)). These works are particularly focused on searching just users' profile in the network and to develop solutions in this direction; the work of Choudhari *et al.* (2008) makes progress in the development of semantic search in social networks, however their work have the same limitation and does not use ontologies to perform the search.

Regarding semantic search but not strictly related to SNS context, there exist various proposals and solutions as illustrated by the survey of Mangold (2007) and Wei *et al.* (2007). Ontology based semantic search solutions (*e.g.* Bonino *et al.* (2004) and Fang *et al.* (2005)) as well as ontology based query expansion (*e.g.* Hoang & Tjoa (2006) and Bhogal *et al.* (2007)) have enhanced techniques for semantic search applications. In order to implementing a solution and make improvements to a search engine of an ISN, future research of our work includes a detailed observation of more ontology based query expansion approaches in the literature to using the 'Semiotic Web ontology' method. Other approaches (*e.g.* Tvarozek *et al.* (2008)) have trying to take advantage of the 'faceted browsing paradigm' and employ an integration solution between semantic search and visual navigation in graphs using the idea of social networks.

Previous work conducted by Reis *et al.* (2010b)

have develop a study on the challenges related to search in ISN; the authors propose recommendations for a search engine better suited to this kind of system. To the best of our knowledge were not found in the literature so far investigations that have specifically focused on semantic search in SNS considering aspects of accessibility and inclusion. We agree that the development of a search engine on an ISN should include these new challenges.

We also argue that the approach developed in this paper can improve and expand methodologically and technologically Semantic Web techniques, such as Web ontologies, illustrating immediate and practical results for better SNS search engines. This approach differs from others, since the search solution outlined tries to derive the term meanings of the search and also the term meanings of the content produced by users in the system, as information from the users profile together with their behaviour and roles. The solutions presented in literature so far do not deal directly with this approach. Future experiments with real users can show whether our approach can bring promising benefits revealing search results more suitable to the context of social and digital inclusion, and also to SNS in general.

6 CONCLUSIONS

Social network systems may provide inclusive access to digital information for people, creating situations where the users' diversity is respected and the access difficulties minimized. This is the purpose of the Inclusive Social Networks (ISN). It is important to provide information retrieving in a more natural way from the user's viewpoint, with results that make sense to people. Therefore more suitable mechanisms for search should take into account the meanings created, shared and used by people in the social network.

This paper presented new perspectives for search in ISN which consider the inclusive social context. It showed the outcomes of an analysis about how to improve a search mechanism considering aspects related to the digital and social inclusion. We could verify with real users, that semantic aspects can make a difference for the users to reach information, and that the current syntactic search engines are not enough to an ISN context. Inspired on the practical context of ISN users, this paper outlined a novel method based on Semantic Analysis Method from Organisational Semiotics to build more representative Web ontologies, as well as requirements of a software tool that materializes these ideas as a process. Besides, it showed strategies for using the proposed ontology with the search engine to provide more precise search results.

As further work, the goal of this research is to improve (in the implementation sense) the ideas drawn up for the search mechanism described in this paper. For that, we aim to develop the semi-automatic tool for building 'Semiotic Web ontology', as an extension of the SONAR tool, including the heuristics and rule transformations to build the OWL ontology aided from OC. Furthermore we intend to develop a pilot implementation of this search engine based on the 'Semiotic Web ontology' in the '*VilanaRede*' system, using and improving the strategies mentioned in this paper; also the work involves new practical experiments with real users, utilizing this novel search mechanism in order to evaluate and validate the solution with empirical results.

ACKNOWLEDGEMENTS

This work is funded by Microsoft Research - FAPESP Institute for IT Research (proc. n. 2007/54564-1) and by CNPq/CTI (680.041/2006-0). The authors also thank colleagues from CTI, IC – Unicamp, NIED, InterHAD and Casa Brasil for insightful discussion.

REFERENCES

Baranauskas, M. C. C., 2007. *e-Cidadania: Systems and Methods for the Constitution of a Culture mediated by Information and Communication Technology*. Proposal for the Microsoft Research-FAPESP Institute.

Bhogal, J., Macfarlane, A., Smith, P., 2007. *A review of ontology based query expansion.* Information Processing & Management, Vol. 43, No. 4, pp. 866-886.

Bonacin, R., Baranauskas, M. C. C., Liu, K., 2004. *From Ontology Charts to Class Diagrams: semantic analysis aiding systems design*. In: Proc. of the 6th International Conference on Enterprise Information Systems. Porto, Portugal.

Bonino, D., Corno, F., Farinetti, L., Bosca, A. 2004. *Ontology Driven Semantic Search*. WSEAS Transaction on Information Science and Application, Issue 6, Vol. 1, pp. 1597-1605.

Boyd, D. M. & Ellison, N. B., 2008. *Social Network Sites: Definition, History, and Scholarship*. Journal of Computer-Mediated Communication. Number 13. pp. 210-230.

Choudhari, A., Jain, M., Sinharoy, A., Zhang, M., 2008. *Semantic Search in Social Networks*. <http:// www.cc.gatech.edu/projects/disl/courses/8803/2008/project/project_deliverables/group22/proposal.pdf>. Accessed on January, 2010.

Fang, W., Zhang, L., Wang, Y., Dong, S., 2005. *Toward a Semantic Search Engine based on Ontologies*. In Proc. of the 4th International Conference on Machine Learning and Cybernetics. Guangzhou, China. Volume: 3, pp. 1913-1918.

Gürsel, A & Sen, S., 2009. *Improving search in social networks by agent based mining*. In Proc. of the 21st International joint conference on Artificial intelligence table of contents. Pasadena, California, USA. pp. 2034-2039.

Haynes, J. & Perisic, I., 2009. *Mapping Search Relevance to Social Networks*. In Proc. of 3rd Workshop on Social Network Mining and Analysis. International Conference on Knowledge Discovery and Data Mining. Paris, France.

Hoang, H. H. & Tjoa, M., 2006. *The State of the Art of Ontology-based Query Systems: A Comparison of Existing Approaches.* In Proc. of IEEE International Conference on Computing and Informatics (ICOCI). Kuala Lumpur, Malaysia.

IBGE, 2008. *Acesso à Internet e posse de telefone móvel celular para uso pessoal 2008*. Retrieved January, 2010 from <http://www.ibge.gov.br/home/estatistica/populacao/acessoainternet2008>

IPM, 2007. *Indicador Nacional de Analfabetismo Funcional*. Retrieved January, 2010 from <http://www.smec.salvador.ba.gov.br/site/documentos/espaco-virtual/espaco-dados-estatisticos/indicador%20de%20analfabetismo%20funcional%202007.pdf>.

Liu, K., 2000. *Semiotics in information systems engineering*, Cambridge University Press.

Maedche, A. & Staab, S., 2001. *An Ontology Learning for the Semantic Web*. IEEE Intelligent Systems. Volume 16, Issue 2. pp. 72 – 79.

Mangold, C., 2007. *A survey and classification of semantic search approaches*, Int. J. Metadata, Semantics and Ontology, Vol. 2, No. 1, pp.23–34

MEC, 2007. Retrieved November, 2009 from <http://portal.mec.gov.br>.

Medelyan, O., 2006. *Semantically Enhanced Automatic Keyphrase Indexing*. In Proc. of Women in Machine Learning (WiML) Workshop co-located with the Grace Hopper Celebration of Women in Computing. San Diego, USA.

Neris, V. P. A., Almeida, L. D., Miranda, L. C., Hayashi, E., Baranauskas, M. C. C., 2009. *Towards a Socially-constructed Meaning for Inclusive Social Network Systems.* In Proc. of International Conference on Informatics and Semiotics in Organisations - ICISO. Beijing, China. pp. 247-254.

Nielsen, 2009. *Global Faces and Networked Places. The Nielsen Company USA*. Retrieved July, 2009 from http://blog.nielsen.com/nielsenwire/wp-content/uploads/2009/03/nielsen_globalfaces_mar09.pdf.

Peirce, C. S., 1931-1958. *Collected Papers.* Cambridge, Mass: Harvard University Press.

Reis, J. C., Bonacin, R., Baranauskas, M. C. C., 2010a. *Analysing Search in Inclusive Social Network Services*. IC-10-02, www.ic.unicamp.br/~reltech.

Reis, J. C., Baranauskas, M. C. C., Bonacin, R., 2010b. *Busca em Sistemas Sócio-Culturais Online: Desafios e Recomendações* (Paper submitted to XXXVII Seminário Integrado de Software e Hardware - SEMISH 2010) (In Portuguese).

Ribeiro, L. C. & Vieira, R., 2008. *OntoLP: Engenharia de Ontologias em Língua Portuguesa*. In Seminário Integrado de Software e Hardware - SEMISH. Anais do XXVIII Congresso da Sociedade Brasileira de Computação. Porto Alegre: SBC, v. 1. p. 1-15. (In Portuguese).

Santos, T. M., Bonacin, R., Baranauskas, M. C. C., Rodrigues, M. A., 2008. *A Model Driven Architecture Tool Based on Semantic Analysis Method*. In Proc. of 10th International Conference on Enterprise Information Systems. Barcelona, Spain. V. 2 pp. 305-310.

Stamper, R. K., 2001. *Organisational Semiotics: Informatics without the Computer?* In Information, Organisation and Technology: Studies in Organisational Semiotics, eds. K. Liu, R. Clarke, P. B. Andersen and R. K. Stamper. Kluwer Academic Publishers.

Tanasescu, V. & Streibel, O., 2007. *Extreme Tagging: Emergent Semantics through the Tagging of Tags*. In International Workshop on Emergent Semantics and Ontology Evolution (ESOE2007). Busan, Korea.

Tvarožek, M., Barla, M., Frivolt, G., Tomša, M., Mária, B., 2008. *Improving Semantic Search via Integrated Personalized Faceted and Visual Graph Navigation.* In SOFSEM 2008: Theory and Practice of Computer Science. Lecture Notes in Computer Science - Springer Berlin / Heidelberg. Vol. 4910, pp. 778-789.

Vieira, M. V., Fonseca, B. M., Damazio, R., Golgher, P. B., Reis, D. C., Neto, B. R., 2007. *Efficient Search Ranking in Social Networks*. In Proc. of 16th ACM Conference on information and knowledge management. Lisbon, Portugal. pp. 563-572.

Wei, W., Barnaghi, P. M., Bargiela, A., 2007. *The Anatomy and Design of a Semantic Search Engine*. Tech. rep., School of Computer Science, University of Nottingham Malaysia Campus.

W3C, 2004. "*OWL-Web Ontology Language*", Recommendation 10 February 2004. Retrieved March, 2009 from http://www.w3.org/TR/owl-features.