



HAL
open science

Addressing universal access in social networks: an inclusive search mechanism

Julio Cesar dos Reis, Rodrigo Bonacin, Maria Cecilia Calani Baranauskas

► **To cite this version:**

Julio Cesar dos Reis, Rodrigo Bonacin, Maria Cecilia Calani Baranauskas. Addressing universal access in social networks: an inclusive search mechanism. *Universal Access in the Information Society*, 2013, <10.1007/s10209-013-0290-7>. <hal-00790033>

HAL Id: hal-00790033

<https://hal.science/hal-00790033v1>

Submitted on 19 Feb 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Addressing Universal Access in Social Networks: An Inclusive Search Mechanism

Julio Cesar Dos Reis¹, Rodrigo Bonacin², M. Cecília C. Baranauskas³

¹*CR SANTEC, Public Research Centre Henri Tudor and LRI, University of Paris XI – 2A, Kalchesbrück, L–1852, Luxembourg +352 42 59 91 3358*

julio.dosreis@tudor.lu

²*FACCAMP and CTI Renato Archer – Rod.Dom Pedro I, km 143,6 Campinas, SP – Brazil +55 19 37466192*

rodrigo.bonacin@cti.gov.br

³*Department of Information Systems, Institute of Computing, University of Campinas (UNICAMP) – Campinas, SP – Brazil +55 19 35215870*

cecilia@ic.unicamp.br

LONG PAPER

Abstract: Social Network Services (SNSs) have brought new possibilities and challenges to the design of software environments that respect people's cultural differences. These systems may represent an opportunity for social and digital inclusion. However, search mechanisms in these systems impose serious barriers for people in the process of acquiring digital literacy. One of the barriers is the difficulty of using the adequate terms/keywords to perform content searches. This paper presents an approach to allow ordinary, non technology proficient people, to access the content of a network through the use of search parameters that make sense to them. The proposal is grounded on Semantic Web technologies (Web ontology) combined with Organizational Semiotics concepts and methods to identify the users' profile and language. A case study was conducted with the search mechanism integrated into a SNS, and a preliminary evaluation reveals the advantages and drawbacks of the approach.

Keywords: Inclusive Social Network, Inclusive Search Mechanism, Information Retrieval, Ontology, Semantic Search and Semiotics.

1. INTRODUCTION

The advent of the Social Web [2] has brought new opportunities for knowledge access and sharing. However, there are still many barriers that prevent people from interacting with Information and Communication Technologies (ICTs) [1]. Social Network Services (SNSs), Wikis, and Blogs are examples of Social Web applications that generate a huge amount of information, and consequently require

specific mechanisms to recover relevant information for the users. While SNSs impose challenges regarding information access, they may also have a dominant role in influencing the adoption and use of the ICTs [4]. In developing countries, a small part of the population is considered as ICTs experienced users; an even smaller part of the population has effective access to valuable and desired information through the Web. Ideally, a beneficial cycle should be promoted in their learning process, *i.e.*, as the value of recovered information increases, so should the ease to recover more information.

In this paper, the concept of Inclusive Social Network (ISN) refers to a type of SNS for all [5], *i.e.*, a system that potentially allows the vastest diversity of people to integrate a group, and interact to produce information (tangible and non-tangible objects) that can be shared with other people and groups. In other words, a SNS with resources to promote access for all, including those at the margin of the digital culture, may be defined as an ISN. Within an ideal ISN all people should have the means to recover information in a way that makes sense to them. Previous studies [6] point out that people who are developing digital literacy (and who frequently have limited schooling) have difficulty in providing the adequate terms/keywords for the search mechanism. A major problem is the use of informal/colloquial terms in the search box; as a consequence, the search mechanism usually produces unexpected results. In this context, the objective of this paper is to propose an approach to search mechanisms that consider the users' profile and their use of colloquial language.

The main contribution of this paper concerns ideas and methods as to how to produce search mechanisms that enable users in the process of acquiring digital literacy to reach written information using well defined/formal words and standard grammar rules. It is expected that users will be able to learn from the recovered information (once the informal terms are associated with the formal ones), which will potentially facilitate the next search iterations. In this sense, the search mechanism may represent an instrument of digital inclusion, where experts and novices effectively communicate and share information. The focus of this paper is not on the design of search mechanisms for digitally illiterate people. Rather, the focus is on the design of search mechanisms that also consider/include non-technology proficient people. With the universal design principles in mind,

methods are explored to model users and the user's relation to the use of colloquial language and shared meanings.

The main shortcomings of traditional search solutions is that, in most cases, they consider that people have already mastered the use of this technology, and the main approaches are syntactic, *i.e.*, they are based on a lexical-syntactic process of information, and not on meaning representations. In recent years the semantic aspects of information have been better explored by the Semantic Web (SW) [19] technology. However, search mechanisms are based on semantic representations disconnected from the users' daily language, since they usually make use of standard terminologies and vocabularies. Moreover, most of the semantic search solutions often require users to point out the context of the meanings (*i.e.*, disambiguation) in order to improve their performance.

With the objective of optimizing search results, researchers in the field of semantics have incorporated techniques from a variety of other research fields, and implemented a number of practical systems [16]. Search mechanisms with semantic characteristics demand that the mechanism be based on the knowledge model of the domain, *i.e.*, the knowledge must be computationally represented so that the machine can "interpret" it. The knowledge can be represented by means of ontologies. Ontologies, mainly in the context of the SW, usually are described by computational ontology languages such as the *Web Ontology Language* (OWL) [21].

This paper proposes an Inclusive Search Mechanism (ISM) that aims to contribute to participatory and universal access to knowledge. The final objective of the proposed solution is to return semantic search results well-suited for all, and, at the same time, respecting the individuality of each user. An ontology-based approach to such mechanisms is proposed. This solution is based on a new approach to the design of Web ontologies [7], which articulates aspects of semantic modeling with strategies to create the ontology based on the content that comes from the ISN system. The overall solution was applied in a case study with real users in the context of the *e-Cidadania* project [3], which investigated methods and tools for the construction of ISNs. A preliminary evaluation of a prototype based on the proposed approach was carried out to show how the overall solution (method and mechanism) affects the user's access to information,

based on the search results provided. Limitations of the proposed approach are also discussed in the paper.

The remainder of the paper is organized as follows: Section 2 presents the motivation for this investigation. Section 3 presents the background of the adopted methodology. Section 4 describes the proposed design method for ISMs. Section 5 proposes a solution (algorithms) for an ISM grounded in the presented method. Section 6 presents the case study showing how the proposed approach (method and mechanism) was instantiated and implemented in an ISN system. This section also presents the results of a preliminary evaluation of the mechanism with a target audience. Section 7 discusses the solution and relates it to other relevant work. Finally, Section 8 concludes the paper and points to further research needed.

2. SEARCH MECHANISMS: A BARRIER TO UNIVERSAL ACCESS IN SNS

Commonly referred to as the ‘digital divide’, there are considerable inequalities in access to information through the Internet, especially in developing countries. Even more important than the physical access to the Internet, is the universal access to the knowledge generated by digital media, which is at the core of a more equitable information society. A search mechanism designed to facilitate access to information for all is motivated and justified by the context of the differences that need to be addressed.

2.1 Social Context and Diversity

In developing countries such as Brazil, India, and China, access to information and knowledge is still restricted to a small portion of the population. There are technological, educational, cultural, social, and economic barriers that have prevented access to, and interaction with, technology [1]. In this context there are still many people without access to the Internet and, consequently, without opportunities to access information and knowledge.

The Brazilian Internet Steering Committee [8] states that in 2010 46% of the population had never used a computer and 52% had never accessed the Internet. Moreover, data from the Internet World Stats (IWS) [9] in 2010 revealed

that in India, with a population of over 1 billion people, despite having the fastest growing rate of new Internet users, only 8.5% of the population uses the Internet.

Regarding education, the Indicator of Functional Literacy in Brazil (INAF in its Portuguese acronym) [10] points out that in 2009 (last survey available), 27% of the Brazilian population between 15 and 64 years old were considered functionally illiterate, defined as the population with less than 4 years of schooling, and unable to perform simple tasks that involve understanding written words and phrases. Using a broader definition of functional illiteracy, according to the same Indicator, the majority (52%) of Brazilians reach the degree of rudimentary literacy, *i.e.*, they have only the ability to locate explicit information in short texts or do simple math, and they are not able to understand longer texts; 9% of these individuals are considered absolutely illiterate. Only 27% of the adult population is considered fully literate.

This scenario illustrates only part of the challenges to be faced in terms of designing systems which should improve these people's social condition. According to Baranauskas & Souza [1], this problem is difficult because it is unique. Its treatment requires multidisciplinary expertise, with methods and techniques for systems that enable the establishment of a digital culture by providing barrier free access to information to every citizen, respecting their differences.

One of the barriers in providing illiterate people with access to digital information is related to their difficulty in using adequate terms/keywords to perform content searches. This difficulty is not the major problem for the absolutely illiterate population, who have other educational issues to deal with, nor is it a problem for fully literate people with good informatics skills. However, when considering individuals with literacy problems (*i.e.*, individuals that are not fully literate), or individuals in process of developing digital literacy, the use of search mechanisms is a barrier to information access. This hypothesis was analyzed in preliminary studies that pointed out these user's main difficulties when using syntactic search mechanisms.

2.2 Preliminary Usage Scenarios

The investigation reported here started with an analysis of the behavior of the prospective users of an ISN when they use search mechanisms. This study was

conducted with real users using a syntactic search mechanism of an SNS system. The activity was carried out in a telecenter in Campinas city, Brazil. Fourteen members of the target community were invited to participate in the experiment. All the participants are considered “immigrants”, coming from various regions of Brazil, usually the poorest states. Their ages ranged from 22 to 60 years old. Considering the schooling level, 40% of the participants have university degree, 30% have high school degree while 10% have incomplete high school; 10% have the elementary school education and 10% have not completed the elementary school. The experiment included people of various occupations, such as housewife, cook, handicraftsman, hairdresser, seamstress, retired people, teacher, student and others. Concerning their experience with ICT or other electronic devices, all of them have at least one TV set at home and two have never used automated teller machines. Only one person does not have a mobile phone, all the others use the mobiles on a daily basis. Six of them use Short Message Service (SMS), but they do not use Internet on their mobiles, despite its drop in price. Two out of the fourteen participants do not have a computer at home, and two of those who do have a computer at home do not use it. Those who do have a computer at home also have Internet access, three of them with dial up access. However, only two participants were familiar with online social networks.

The participants were asked to search within proposed scenarios; the objective was to observe their behavior while using the search mechanism to assess whether semantic aspects could make a difference in their search situations. In this study, a task sheet with 4 search scenarios was presented to each of the 7 pairs of participants. The users formed the pairs themselves, and for each scenario the pair would write the words used in the search and the title of the information (announcements) retrieved from the system [6]. The 4 search scenarios were:

Scenario 1: Find announcements on how to popularize the ‘*VilanaRede*’.

Scenario 2: Find announcements of mango (fruit) in ‘*VilanaRede*’.

Scenario 3: Find announcements related to food in ‘*VilanaRede*’.

Scenario 4: Find announcement related to religion items combined with handcraft in ‘*VilanaRede*’.

Taking Scenario 3 as an example, the objective was to see whether users would use the keyword “food” in the search, or if they would make a search for specific foods through the search engine. When users tried the keyword “food”,

the system found nothing. However, there are several announcements for food in the system: the sale of “homemade snacks”, “cheese bread”, and others. Among the relevant considerations noted by the observers in the experiment was that, during the execution of this scenario, users said that the system should have associated “homemade snacks”, “cheese bread”, and “steak of banana peel” with the concept of “food”. This makes sense, since semantically all of these items are types of 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 'clean' and practical for those who are starting (in terms of acquiring computer literacy), like us, when we enter “food”, it should return a variety of foods due to our difficulty.” Yet another user said: “Maybe using food does not help in the search for 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”, “homemade cooking”, “food sale”, “homemade snacks”, “small homemade snacks”, “fried homemade snacks”, “pies”, “sweets”, “cheese bread”, “small sweets”, “cake”, “pastel” and “chocolate sweet”. Note that the participants used several variations in the words such as “homemade snack”, “small homemade snack” and “fried homemade snacks”.

Several observers noticed that the subjects used terms from their own colloquial language in the search. For example: “pink mango”, “coconut mango”, “tutu beans”, “by word of mouth”, “small sweets”, “small homemade snack” and “*Orisha*”. In addition, in several occasions the participants discussed amongst themselves before reaching an agreement as to which word to use in the search. The description of the other scenarios, and a deeper analysis and discussion of results are presented in [6].

This analysis revealed the importance of considering semantic aspects in the design of a search mechanism that is more adequate for the ISN concept [6]. To improve search results for this context it is necessary to take into account the way in which the users’ ordinary language influences meaning in the social network (*i.e.*, users’ local/colloquial meaning should be considered during the development of more appropriate ISMs). Thus, such mechanisms should take into account the meanings created, shared, and used by people through the use of the system [6]. These meanings are expressed both in the content posted and in the

search queries created by users. Moreover, the scenarios also points out that it is necessary to construct computationally tractable models from the semantic point of view derived from the SNS itself.

It was noticed that users formulate queries that are related to a given social context. The more adequate search results (*i.e.*, content from the SNS) must be meaningfully related to the user that performs the search [6]. The meaning of words in an inclusive mechanism must be considered according to who produces and who consumes the information. The main challenge in this context is to provide a harmonious experience between the mechanism and the users in their colloquial language [11].

Thus, in order to develop a search mechanism more suitable for the ISN concept, which considers the users' informal language, the solution should reflect the semantics used by the participants of the system [6]. Besides taking into account the content from the ISN system to model the semantics, the meanings must be modeled according to this social context. It is necessary to clarify the pattern of behaviors shared in the network. In this way, methods are needed to computationally model the meanings related to a context and to a person or a social group that share meanings and behave according to commitments, linguistic constraints, culture, and other social aspects. Once based on Semiotic theory [30], semantics is understood as the interpretation of signs by individuals in association with real world elements. This interpretation is socially contextualized, *i.e.*, individuals and communities may have different interpretations. Organizational Semiotics (OS) concepts and methods were considered in the definition of the approach, which takes into account these interpretations through the identification of patterns of behaviors.

3. ORGANIZATIONAL SEMIOTICS THEORY AND METHODS

This section presents a brief overview of OS as the theoretical background for this work. The Semantic Analysis Method (SAM) is also presented in order to clarify the proposed solution.

3.1 Organizational Semiotics

Semiotics, the doctrine of signs, leads to an understanding of information as properties of signs. Anything standing for something or used to signify something else [30] is an example of a sign: words, sentences, traffic lights, diagrams, the wave of a hand, and facial expressions. OS understands an organization as a system of signs, and studies organizations using concepts and techniques rooted in Semiotics [32, 33, 34]. OS can be understood as one of the branches of the Semiotics developed by Peirce [30] and others. The rationale behind OS is based on the assumption that any organized behavior is affected by the communication and interpretation of signs by people, both individually and in groups [34].

OS studies the nature, characteristics, functions, and effects of information and communication in organizational contexts. Organization is considered a social system in which people behave in an organized manner by conforming to a certain system of norms [34]. In this work, a social network is also understood as an organization, since there are communication rules, and some behaviors are probably expected in the communities. Social rules, often implicit, state what is allowed and what is not.

From a philosophical point of view, OS understands reality as a social construction based on the behavior of the agents who participate in it. People share patterns of behavior governed by a system of signs. This philosophical position states that, for all practical purposes, nothing exists without a perceiving agent or without the agent engaging in actions [34]. Methods from the OS are useful to delineate the influence of the social aspects in the organizations.

Among the methods employed by the OS community is a set of methods known as MEASUR (Methods for Eliciting, Analyzing and Specifying Users' Requirements) [36], which deal with the use of signs, their function in communicating meanings and intentions, and their social consequences. MEASUR involves the analysis of stakeholders in a focal problem, their needs and intentions, and the constraints and limitations related to the prospective software system. In this work the Semantic Analysis Method (SAM), from MEASUR, is adopted as part of the proposed approach to model the knowledge coming from the ISN. The following section presents a summary of SAM.

3.2 The Semantic Analysis Method

Using a subjective philosophical stance and an agent-in-action ontology the SAM determines the underlying semantics of a social context and the relationship between the human agents and their patterns of behavior [37]. SAM supports the analysis, specification, and representation of a social system.

The SAM assists users or problem-owners in eliciting and representing their meanings in a formal and precise semantic model, the Ontology Chart (OC). OC is a graphic representation of a conceptual model that describes a view of responsible agents in the focal domain including their pattern of behavior, referred to as affordance(s), and the ontological dependences between them [34]. The meanings are specified in the OC, which represents an agent-in-action ontology. The OC is read from left to right; any element is dependent for its existence on the affordance(s) to its left to which it is connected. The meaning of the words used in the semantic model is treated as a relationship between the signs and the appropriate actions by the agents.

In the SAM, “the world” is socially constructed by the agents’ actions, on the basis of what is offered by the physical world itself [34]. It is worth mentioning that this concept of ontology and of agent (from SAM) is not the same used by the SW community. An OC represents a domain under study, which can be described by the concepts, the ontological dependencies between the concepts, and the norms detailing the constraints at both the universal and the instance levels [38]. Some basic concepts of SAM adopted in this paper are based on Liu [34], and are briefly presented as follows:

“**Affordance**”, is a concept introduced by Gibson [39] that can be used to express the invariant repertoires of behavior of an organism made available by some combined structures of the organism and its environment. In SAM [36] the concept introduced by Gibson was extended by Stamper to include invariants of behavior in the social world, and affordances are social constructs in a certain social context [34]. The social world acts as the environment that is constantly affecting the agents’ behavior, and at the same time it is affected by the agents’ actions.

“**Agent**” is a special kind of affordance, which can be defined as something that has responsible behavior. Agents are affordances that can take responsibility both for their own actions and as for the actions of others. An agent

can be an individual person, a cultural group, a language community, a society, *etc.* (an employee, a department, an organization, *etc.*);

“Ontological dependency” is formed when an affordance is only possible if certain other affordances are available. Saying that the affordance “A” is ontologically dependent on the affordance “B” means that “A” exists only when “B” does. For example, for a person to be able to stumble, he/she must first walk; for two people to divorce, they need to be married; thus there exists an ontological dependency between stumbling and walking, and also between divorce and marriage.

The SAM has other important characteristics to be considered. One of the SAM principles, according to Liu [34], is that an analyst is not allowed to invent artificial terms or introduce new concepts when modeling the agents’ actions in the OC. The purpose of this is to force the analyst to speak the same language as the problem-owners. Any ambiguity in the terms or concepts used in describing the problem should be resolved by putting them into a context of actions that are already described and understood. When doing so, if the problem-owners are inspired with some new terms, they may be used only after a careful justification is provided by the problem owners and the analyst. The reason for this is that the modeled world is constructed by the community of agents, *i.e.*, the problem owners. The agents know the meaning of words in their own world, their interpretations are the only ones justified.

4. THE SEMIOTIC INFORMED WEB ONTOLOGY DESIGN

Since the meanings are socially determined, a new approach for the design of Web ontologies grounded on semantics was proposed to inform ISMs. This proposal includes the role of the individual who interprets signs (represented as concepts in ontology) within the Web ontology. Therefore, the concepts of ‘Agents’ and ‘Ontological Dependences’ that come from the SAM outcome are included in the Web ontologies described in OWL [7]. With that, both the ‘Agents’ and ‘Affordances’ are transformed into OWL classes, and are related to each other following the ontological dependences modeled into the OC.

In this approach, instead of modeling the meanings from the social network using OWL directly, they are first modeled in an OC using SAM. In

order to achieve an OWL, ontology, heuristics, and transformation rules presented in [40] are applied to produce a ‘Semiotic Web Ontology’ (SWO) [7]. SWO is a semantic model (computationally-tractable ontology) constructed by including SAM concepts as Agents, Affordances, and Ontological dependences, thereby combining SAM concepts with SW technologies. The heuristic’s aim is to transform the OC artifact into OWL code. Some relations between the models are mapped, and one model supports the construction of the other, providing benefits presented in the different viewpoints. The heuristics represent a procedure that makes explicit the relationships between the models, and makes possible the construction of one diagram from the other. They were implemented in the SONAR Case Software Tool that assists designers in a semi-automatic transformation process [41].

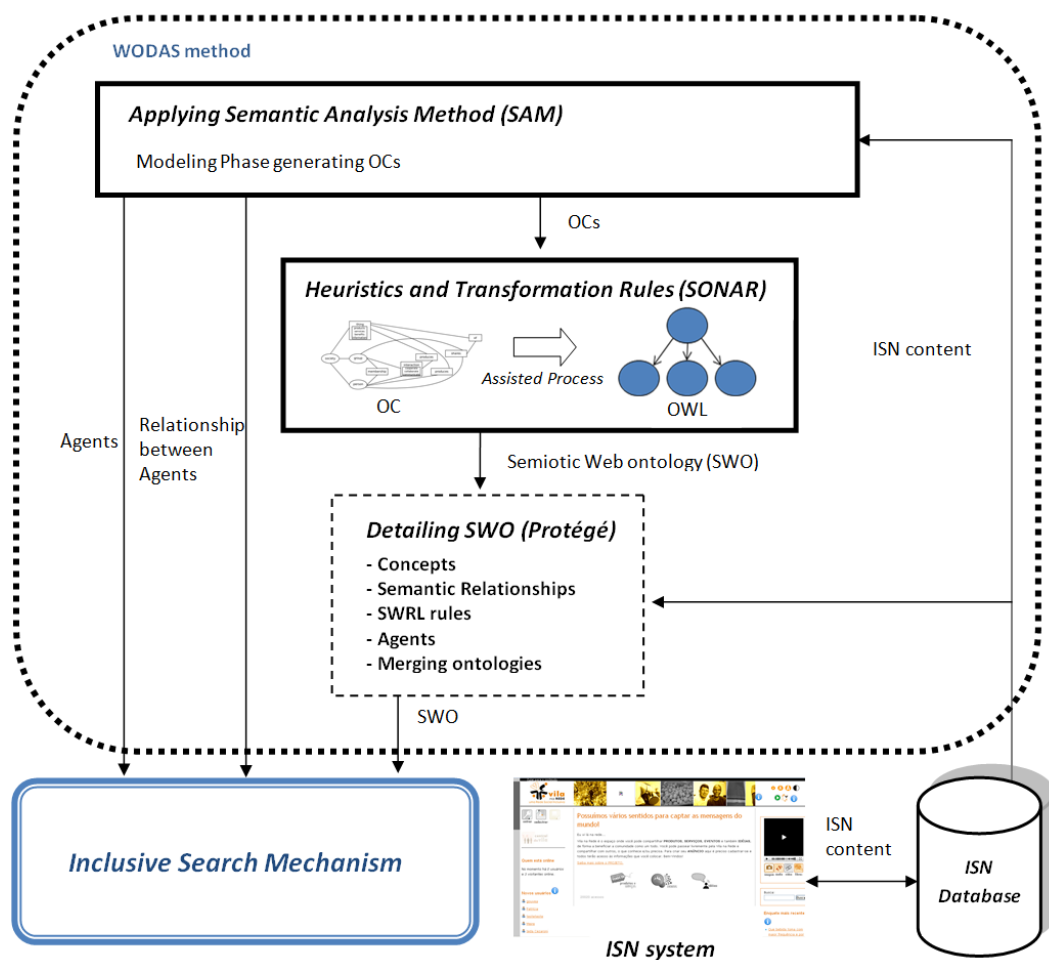


Figure 1. The WODAS method to create SWO

Using the SWO approach to inform a search mechanism may lead to search results that are more adequate to the ISN concept, since the shared meanings were modeled using an agent-in-action ontology grounded in the concept of affordances. In this sense, the concepts modeled in a SWO are extracted from affordances modeled in SAM. The meaning of these concepts is not determined by the object's properties itself, but rather they are a combination of aspects and depend on who (*i.e.*, agent) interprets them.

Figure 1 presents the WODAS (**Web Ontology Design Aided by Semiotics**) method to create SWO in an ISN context. Based on ISN content, SAM is applied generating one or more OCs. The generated OCs are used to produce the SWOs through the heuristics and transformation rules [40]. Such OWL ontologies can be merged as well as detailed (*e.g.*, including more classes, instances and also SWRL – Semantic Web Rule Language – rules) using an OWL editor. Besides the detailed SWO, the ISM uses the agent's models, and also possible relationship between the agents, during the search time.

The search solution is based on SWO semantic information. It is used to classify words (both search keywords and words from published contents) and also to make query expansion during search time. Besides the SWO, the ISM requires information about the relationship between each SNS user and the agents modeled, and also possible relationship between different agents. Information from the users' profile and their activity in the SNS are necessary to connect each user to a possible agent. Figure 2 shows a general vision of the relation between the ISM and the WODAS method.

The proposed ISM solution is built on the probable stated meanings based on the identified agents and their relationship with SNS users. This approach tries to treat polysemy, synonymous, and other aspects of human language during the search, through the identification of "local" meanings. For that, it is proposed to connect users to agents, as explained below.

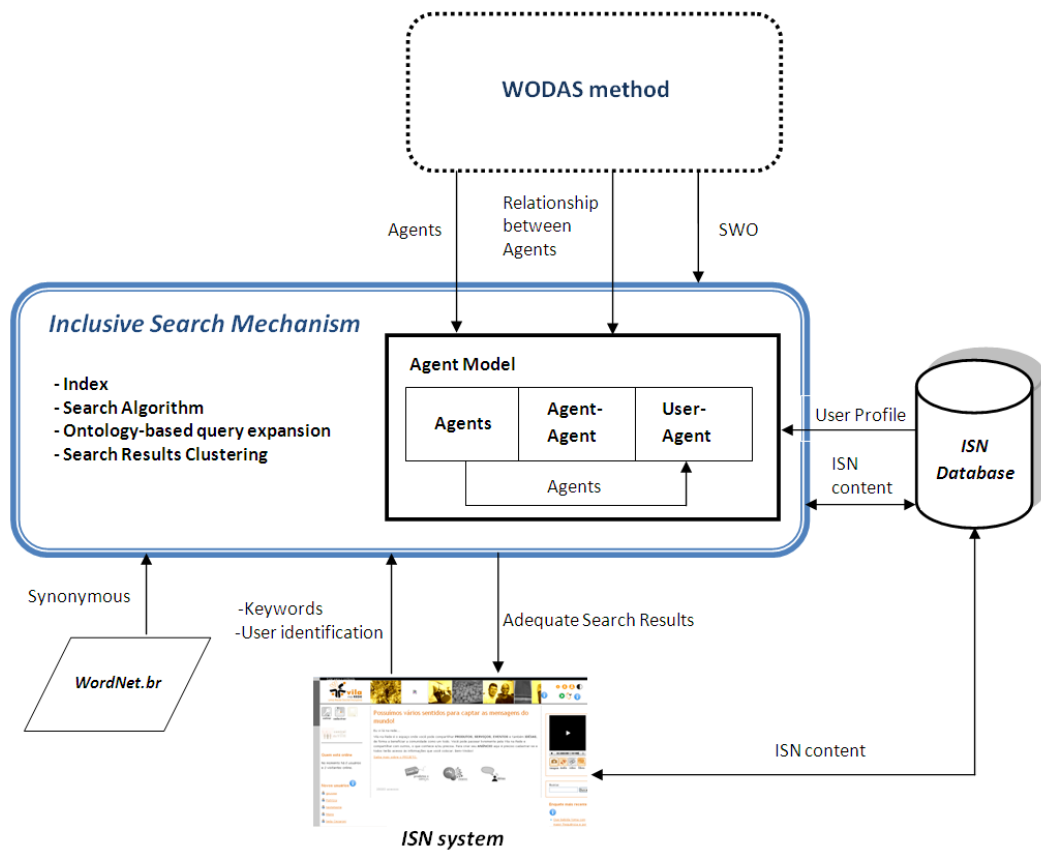


Figure 2.The Inclusive Search Mechanism (ISM)

Modeling user-agent relationship: During search time, the ISM uses the connections made between the agents represented in the ontology and the SNS users. Therefore, it is necessary to create a user-agent relationship beforehand. Users themselves specify their profiles using ISN system features. The present proposal only demands that users make a choice of a pre-defined combo-box list. Thus, users indicate their profession and the related activities in which they are involved among the concepts that can socially characterize them. The ontology engineer makes a relationship (connection) between a particular user group in the SNS system and appropriate multiple agents represented in the OC based on this information, as well as on the user's activities in the SNS (*e.g.*, the content that he/she has shared). This can be modeled by using a priority rank (specified by the ontology engineer) for the different agents that suit the user. Additionally, it is necessary to define the relationship between the agents, *i.e.*, how much one agent is semantically close to another, *e.g.*, a 'doctor' agent is more related to 'medical laboratory scientist' agent than to a 'hairdresser' agent. The ontology engineer also has the role of modeling the relationship between agents. Ideally, the combo-boxes can potentially be substituted by the detection of

the profiles from the content using Social Network Analysis (SNA) techniques. SNA may also help to discover the relationship between the agents. However, the application of these techniques requires long-term investigation.

Modeling SWRL rules to get meanings: SWRL rules are described to computationally handle the agent-term relationship. These rules are modeled into the SWO by the ontology engineer. The rules are used both during the search index processing and during the search in an attempt to semantically classify the words. There is not a minimum or a maximum number of rules to be modeled. This is determined by the number of agents, and the possible polysemy and synonymous terms in the specific context that needs to be modeled. These terms are found based on the social network activities. It is the content posted by users. A study [51] using text mining tools has been conducted to find possible terms to aid ontology modeling, as well as to identify the possible polysemy and synonymous terms which may support the ontology engineer creating the SWRL rules.

Search Index: During the modeling phase, an inverted search index has to be created in order to associate words and their possible meanings. The possible pair (word, meaning) has a relationship with the identification of the contents that contain that word. The inverted index is automatically created from the processing of all possible contents that can be retrieved from the SNS during the search. Thus, for each content that has an author who was connected to some agent during the user-agent modeling, words from this content are analyzed in order to remove stop words, and are confronted with the modeled SWRL rules in the ontology, trying to find a possible meaning for each word based on the author's agents. An ontology class (*i.e.*, a meaning) is set to a term, depending on the SWRL rules available. The same word can be set up with different meanings (OWL classes) according to the way different modeled agents conform to the SWRL rules, pointing to different SNS contents. Therefore, it is possible to recognize the appropriate meaning of a word according to the author's agents. The search index must be automatically updated when new content is published.

For instance, a user is related to an agent named 'hairdresser' in the user-agent relationship. SWRL rules specify that the meaning of the term 'cut' for a 'hairdresser' is related to the concept of 'hair'. Therefore, when a post is published in the SNS by a user with the profile of 'hairdresser', it is processed in

the search index. If in this post the term ‘cut’ is found, as a consequence the meaning attributed to this term is related to ‘hair’, since the SWRL rule associated the term ‘cut’ according to the agent ‘hairdresser’ with the meaning ‘hair’ (hair cut). Additionally, other disambiguations are also processed by the mechanism, including distinguishing ‘haircut’ associated with the ‘hairdresser’ from ‘haircut’ in the finance domain.

5. AN INCLUSIVE SEARCH MECHANISM

The ISM finds out the possible modeled meanings for the words in the ISN content (that is stored in the search index). Moreover, it matches and retrieves the content(s) that contain a word with the meaning that is likely to be more suitable to the user who does the search (*i.e.*, matching the meaning of the input search keywords with the word/class pair in the index).

The key idea is to find out if the most likely meaning of the search keyword(s) matches the likely meaning of the word in the ISN content. Thus, the proximity or similarity scale between the search keywords and the content words is determined by the most probable agent’s interpretation of the word (producer and consumer), instead of being determined by a fixed vocabulary.

A scenario that exemplifies the potential and the practical aspects of the ISM proposal is as follows: the term ‘shingle’ can have different meanings depending on the context (polysemy). ‘Shingles’ means a viral disease in a medical context, as well as a roof in an architectural context. As an example, suppose that a user (a doctor) logged into the ISN system and searched for ‘shingles’. If there is a relationship between such user and an agent in the OWL code, it is possible to find the probable meaning for this keyword according to the agent, and consequently to present more suitable search results to him/her. For instance, if a user containing a doctor agent is searching, then the mechanism should first return the results that contain ‘shingles’ as a disease. Thus, it is necessary to describe the Web ontology (including SWRL rules) stating that if the agent is a doctor, the meaning of ‘shingles’ is closer to ‘disease’, as well as if the agent is a ‘civil engineer’ the meaning of ‘shingles’ is relative to ‘roof’, and so on. As described by the rule (1), for example, if one instance of the Agent class has a role equal to ‘doctor’ and the instance of the Input class is equal to ‘shingles’, then this instance must be of the class ‘Viral_Disease’. When modeling the SWRL

rules, it is not necessary to have all the possible combinations of an agent's role with all terms, but it is necessary to know, in the considered ISN, the possible ambiguous terms for each agent.

$$(1) \text{ Agent}(?x) \wedge \text{role}(?x, ?y) \wedge \text{swrlb:equal}(?y, \text{"doctor"}) \wedge \text{Input}(?z) \wedge \text{data}(?z, \text{"shingle"}) \rightarrow \text{Viral_Disease}(?z)$$

The rules can also be generalized to deal with domain specific vocabularies, terminologies, and ontologies associated with the agents. The intersection of the domain specific terms of these external sources and the ontology can be used for disambiguation. For example, the rule (1) can be generalized in the following way: if one instance of the Agent class has a role equal to 'doctor' and the instance of the Input class is equal to a term of ICD10¹, then this instance must be of the class 'Disease'. Unfortunately, in many cases these external vocabularies, terminologies, and ontologies do not exist or are not available; in such cases, the ontology engineer may use natural language processing tools to assist the construction of the rules as described in [55].

A strategy based on semantic proximity is proposed in order to rank the search results. The search algorithm selects the results, and clusters them according to their meanings. If some ISN content contains a word classified with the same meaning as the input search keyword for the user, then this content is set as an appropriate result for this user, and must be presented to the user first. In the same way, if a word has the same meaning as the input search keyword, then the content that contains this word is also an appropriate result.

To organize the search results, the ISM solution distinguishes between two principal groups of results: (1) the group that is most likely to be relevant (personalized) to the user, and (2) the less relevant group. Group 1 contains search results in which the meaning of the words is the same as the input keyword. Group 2 contains all other possible meanings for the input keyword or other related words. Inside each group (1 and 2) the semantic results are organized in: "synonymous results", "more specific results", "more generic", and "related results". Results that contain an exact match with some input search keyword are shown first in the group; the results retrieved by the ontology-based query expansion are organized by internal groups. For example, results containing synonyms in group 1 are those in which the word has the same meaning as an

¹ <http://www.who.int/classifications/icd/en/>

input search keyword. Results that contain synonyms for the input search term, but have no classification or a different classification (meaning), are presented in the synonym results of the second group. Similar schemes are used for all other internal groups (semantically organized) for both group 1 and 2. The search results are organized as hide boxes and internal boxes. The understanding of this group of results by low literacy users is analyzed in the evaluation of this investigation. The ISN content is clustered in each group and internal group according to the search algorithm presented in the next section.

5.1 The search algorithm

The search algorithm receives the input search keywords and the user's identification. The output of the algorithm is the search results organized by lists of the respective groups and internal groups. Given the user's identification, the algorithm gets the users' agents. For each input search keyword, the first step is to find its likely meaning. This is performed through the interpretation of the representation of the agents in the Web ontology, the user-agent relationship, and SWRL rules. It is possible to have situations in which no meaning is found given the combination of the users' agents and the search keyword, due to the lack of rules to treat it. Figure 3 shows the proposed search algorithm.

<p>Input: <i>input search keywords and the user's identification</i></p> <p>Output: <i>The semantic search results organized by lists</i> <i>user's agents=Get user's agent (user's identification)</i></p> <p>Foreach <i>search keyword do</i></p> <p style="padding-left: 40px;"><i>meaning=Find meaning in the SWO (keyword, user's agents)</i></p> <p style="padding-left: 40px;"><i>index classes=Query in the index (keyword)</i></p> <p style="padding-left: 40px;"><i>expanded_words=Perform the query expansion in the SWO (keyword)</i></p> <p style="padding-left: 40px;"><i>SearchByClasses (keyword, expanded_words, index classes, meaning)</i></p> <p style="padding-left: 40px;"><i>SearchByAgents (keyword, expanded_words, user's agents)</i></p> <p>End</p>

Figure 3.The proposed search algorithm

Next, the algorithm queries in the index ask for the search keyword. If the keyword exists in the index, the algorithm returns the possible classes (*i.e.*, the meanings) for that keyword. However, such query made in the index may not find the input keyword, so no defined class may be returned. This situation happens

when the word is not in the content processed by the index, or there is not a SWRL rule that treats the searched combination pair (keyword, agent).. It is possible to have situations in which a word has a classification for one content (depending on the user's agents that have published it) and no classification for other contents, depending on the agents and SWRL rules. Since neither input search nor content words may be classified using the SWRL rules, they are recorded in the index without a classification, *i.e.*, with an unknown meaning.

The algorithm performs the (**'Query expansion'**) on the SWO available based on the input search keyword. The query expansion occurs for each input search keyword. The SWO is used to perform the query expansion (*i.e.*, to find new possible words to be searched in the index that are semantically related to those of the input). The SWO is analyzed to find synonyms, more generic, more specific, and also related words. The modeled ontological dependencies that come from the OC to the Web ontology are also used to find related words to be searched. For example, if an input keyword A depends ontologically on B, then B will be used as a related word to perform the ontology-based query expansion search. All words found in the ontology for each type of operation mentioned (synonymous, specific, generic, and related) are used to make the query expansion, and to return other possible semantic search results. Alternatively, to increase the synonym list, the solution uses standard vocabularies, such as '*WordNet*'.

After the query in the index that looks for the input search keyword, the available meaning classes for this keyword, and the query expansion execution, a set of expanded words and classes may become available. Afterwards, the search by classes (**'SearchByClasses'**) method is executed using the following parameters: (1) the search keyword entered by the user, (2) the words retrieved from the ontology expansion (query expansion), (3) the index of meanings found from the search keyword according to the user's profile, and (4) the classes retrieved in the inverted index.

The algorithm queries each word in the index (including the expanded words) for results (contents) related to each class. For example, if two classes were found in the index for a certain search keyword, this keyword, and all the others found in the ontology query expansion, will be queried twice in the index combined with a class (one for each different class). If some of these retrieved

classes in the index were equal to the class retrieved for the search keyword (the meaning in the algorithm), then the possible search results are set in group 1; if not, the results are set in group 2. Figure 4 illustrates this method step by step. The results for the expanded words are listed in the appropriated internal groups.

```

Foreach word and expanded_words do
  Foreach class do
    results=Query in the index (word, class)
    If (class== meaning) then
      group_1=group_1+results
    Else
      group_2=group_2+results
    End
  End
End

```

Figure 4.The method search by classes

It is possible to observe that if an undefined class is found in the search and the word exists in the index, all the search results, including those retrieved from the query expansion, are set to group 2. Consequently, the number of search results in group 1 tends to increase in function of the number of SWRL rules. There is always a query expansion for each keyword, and the search results for the possible words found are queried in the index, combined with the possible meaning found for the input keyword search.

After the search by class execution, the search is also performed by agents (**‘SearchByAgents’**). If the search keyword is not found in the index, the probability of the **‘SearchByClasses’** finding personalized results (*i.e.*, in group 1) is zero. Besides, when a meaning for the search keyword relative to the user that performs the search is not found, all the possible results found by the **‘SearchByClasses’** through the index are also set as less relevant results (*i.e.*, group 2), regardless of the classes found in the index. This is because the classes found in the index do not match a class for the user that is performing the search, since a class was not found based on the user’s agents who perform the search.

In an attempt to find results for group 1, the search results from queries in the index with the input keyword and with the expanded words are classified as group 1 or 2 according to the user’s agents that have published it. In this situation the lexical-syntactical search results (regardless of the class meaning to which they belong) are selected as group 1 or group 2 according to the agent that

produced the content. Therefore, if an agent that belongs to the user that has published the content found from the index matches an agent associated with the user that performed the search, then this search result is presented in group 1; otherwise the search result is set to group 2. The ontology engineer determines the similarities between two agents at modeling time. The ontology engineer is able to do this based on the OC diagrams, which have the agent's relationship. For instance, if an agent A depends ontologically on agent B, then this indicates that a relationship between them maybe specified. Moreover, a weighed value between agents may also be considered when determining their similarities. The search mechanism will tend to return the content as an adequate result to the user that performs the search. Sometimes this agent's matching is not so clear, *i.e.*, the agents (from the search and from the content) are not the same. Thus, similarities, *i.e.*, relationships modeled between the agents, are used to 'infer' the meaning represented by the list of relationships between the agents. Figure 5 illustrates this method step by step.

```

Foreach word and expanded_words do
  Results = Query in the index (word)
  Foreach results do
    agents = Get user's agent (agent result)
    If (agents == user's agents) or is_similar() Then
      group_1 = group_1 + result
    Else
      group_2 = group_2 + result
    End
  End
End

```

Figure 5. The method search by agents

The final result of the search algorithm is a union of results from 'SearchByClasses' and 'SearchByAgents' algorithms. This union is the set of pointers for the content in groups 1 and 2 (including their internal groups). The proposed algorithms are not able to handle superposition and intersection of result clusters. Moreover, the disambiguation is not performed by the words itself; their meanings are determined by the agents and modeled through the rules. After determining the possible meaning of each input search keyword, based on the agents associated with the user that performs the search, such meanings are

compared with all possible meanings for the terms retrieved from the index. Thus, the search results are clustered into groups 1 or 2 according to this comparison.

6. THE *VilanaRede* CASE STUDY

The WODAS method to model SWOs and the proposed ISM were implemented and evaluated in the context of the *VilanaRede*² system, an ISN. *VilanaRede* was developed as a product of *e-Cidadania*'s Project, with the objective of being accessible for the widest variety of users, including those less familiar with technology and those with low literacy levels. *E-Cidadania*³ is a Brazilian research project that has taken on the challenge of developing systems that give access to users in a way that makes sense within the context of the community. It may contribute to the promotion of a digital culture that respects the diversity of the population. *VilanaRede* users participate in the ISN by making announcements, collaborating in announcements, sharing goods and services, and sharing events and ideas. The content initially available in *VilanaRede* as well as some of its users are considered in the Case Study. The WODAS method was applied using the real content available in the *VilanaRede* system to construct the SWOs. The following sections show how the SWOs were modeled using the *VilanaRede* data, the prototype developed, and its preliminary evaluation.

6.1 Building SWOs from the *VilanaRede* content

The content domain of the announcements available in *VilanaRede* is wide-ranging, *i.e.*, there are announcements about various domains. In fact, one of the main challenges in modeling this content is to deal with the open and informal domain of the announcements. Such contents created by users of the *VilanaRede* system form a “reference corpus”. Thus, the contents presented in this system tend to be heterogeneous with respect to contents’ subject, ways of expression, users, and identity, among other factors. The announcements are diversified regarding the contents and include sale of various handmade products, meals, electronic products, advocacy services, events that include debates about education, and

² www.vilanarede.org.br

³ www.nied.unicamp.br/ecidadania

“June parties”⁴. Ideas also cover a variety of subjects, such as: recipes, environmental awareness, health tips, and so on.

This investigation considered 230 announcements in *VilanaRede* distributed among products, services, and ideas. Each announcement has an associated forum with questions and answers produced by various users (comments exchanged among users). A total of 1633 documents were taken into account for the initial construction of the models and the preliminary evaluation of the proposed mechanism. The number of announcements considered was limited by the relative size of the network in the beginning of the *VilanaRede* usage. The idea was to produce an initial model and develop this model as the size of the network increased. The study described here points out aspects of the construction of this initial model. The document’s texts were analyzed using three mining tools: Kea [53], ExATOlP [54] and CLUTO⁵. A study has been conducted [51] investigating, with real contents of *VilanaRede*, text mining tools for the identification of concepts and semantic relationships that come out of the ISN data.

The SAM was applied using the announcements created by users of *VilanaRede* in an attempt to model the meanings intended to be shared in the network. User’s commentaries about the announcements were also considered during the modeling. Altogether, 10 groups of announcements were created according to subjects’ sets. The group’s subjects were: cooking and meal ordering, sale of products and services, cultural events, announcing *VilanaRede*, physical exercises and health promotion, social projects that involve inclusion and citizenship, offerings of courses and seminars, health-oriented food, handicrafts, and the environment. SAM was applied to each group resulting in 1 or 2 OC(s) for each group.

With the results of the three text mining tools mentioned above, sixteen graduate students in Computer Science who were studying OS, were involved in the task in the role of analysts. They were divided into groups of two or three people. Each group got one subject (of those mentioned before) collected from the *VilanaRede*’s announcements. Each set of data included approximately twenty to thirty announcements with its respective commentaries, and the results of the text mining tools.

⁴ typical parties that occur in some regions of Brazil in the winter season

⁵ <http://glaros.dtc.umn.edu/gkhome/views/cluto/>

After applying SAM, each group modeled an OC that might represent the semantics of the *VilanaRede* announcements. The analysts were supported by the SONAR tool [41] in order to model the OCs. One week later, each group presented the results they had achieved during the modeling process. The analysts created six initial versions of OCs. Figure 6 shows one of the OCs modeled from the *VilanaRede* content related to courses and seminars.

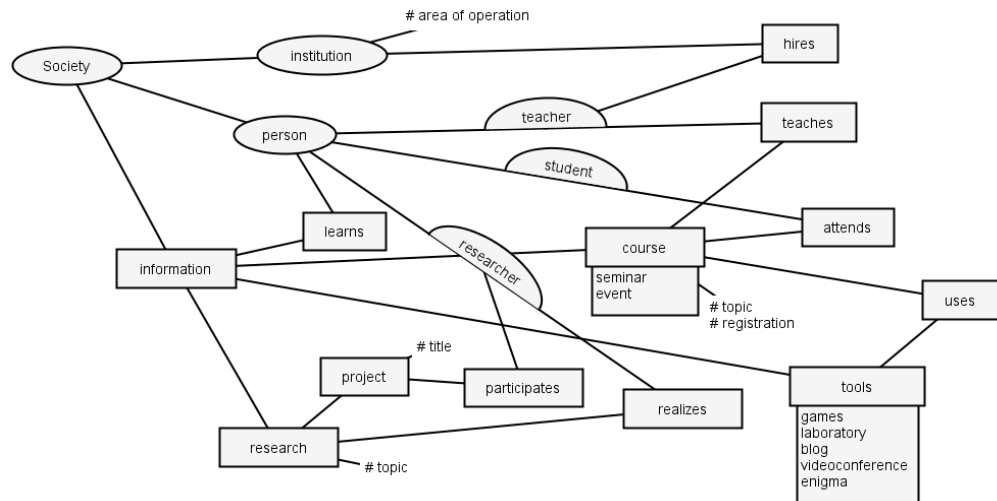


Figure 6. An example of an OC modeled from *VilanaRede* contents [42]

The next step involved modeling an SWO from the OC. For that, the semi-automatic process that includes heuristics and transformation rules for deriving an initial Web ontology described in OWL from OC was applied. In this process the analysts specified the affordances that should be mapped to OWL classes, as well as the affordances that should be mapped to object properties using the SONAR [41] tool. From that, the implemented transformation rules were able to construct an OWL file for each OC. This OWL file includes classes, object properties, and data properties derived from the heuristics.

At this point the diagrams represent pieces of knowledge in a fragmented way, since it is difficult to model the variety of contents at the same time. However, these diagrams need to be centralized and detailed for a wider coverage, since this knowledge representation must be used by the ISM. For this purpose, the ontologies were merged into a single ontology. At this point in the investigation, an ontology engineer manually merged the ontologies. Though knowing that ontology merging is still an open research problem, over the last years the scientific SW community has evolved regarding this problem. Many

approaches and solutions for matching algorithms, as well as for ontology alignment and mappings, may provide and improve the merging solutions in future work. Moreover, some tools for ontology merging may support the ontology engineer in this task (*e.g.*, Prompt⁶).

After the merging, the process delivered an initial version of OWL ontologies up to that point; additional information was necessary regarding the modeled classes in the final SWO. Therefore, new classes were elaborated to build new relationships. In addition, to complement the ontology, instances of classes were created, ranges of values added, and SWRL rules were defined. The details of how the transformation occurs, and also an example of OWL codes generated using the SONAR tool (that implements the transformation rules), can be found in [40]. More details regarding the application of the method to the *VilanaRede* context can be found in [42].

6.2 The Prototype Construction

The *VilanaRede* system uses the Content Management System⁷ Drupal 5 to manage users' accounts, content types, and other general functions; the ISM was developed as Web Services connected to the Drupal modules and to the ISN modules. The ISM Web Services manage all the necessary information and provide all the services to handle the search requests, the index processing, the adding of agents and user-agent relationships, as well as the ontology management for getting meanings and query expansion. The search services include the search algorithm that provides the search results organized by lists, which are interpreted on the Drupal side after the search service request. Moreover, in the Profile User Interface (UI) at *VilanaRede*, users can choose a possible agent as their social profile that is stored using these services.

The UI of search results in *VilanaRede* was developed following the organization of search results as proposed in section 5. Figure 7 shows a screen snapshot after the search request processing. The 'Personalized' box contains the results for group 1 while the 'Other' box contains the results for group 2. The internal boxes (*e.g.*, synonym results) organize the semantic search results into both 'Personalized' and 'Other' boxes. It is important to observe that the results from the extended semantic of search keywords are also provided in the light

⁶ protege.stanford.edu/plugins/prompt/prompt.html

⁷ Drupal.drupal.org

green internal boxes. The user can expand these boxes to see the results, for instance, if he is interested in synonym results. This is a way to present related semantic search results.

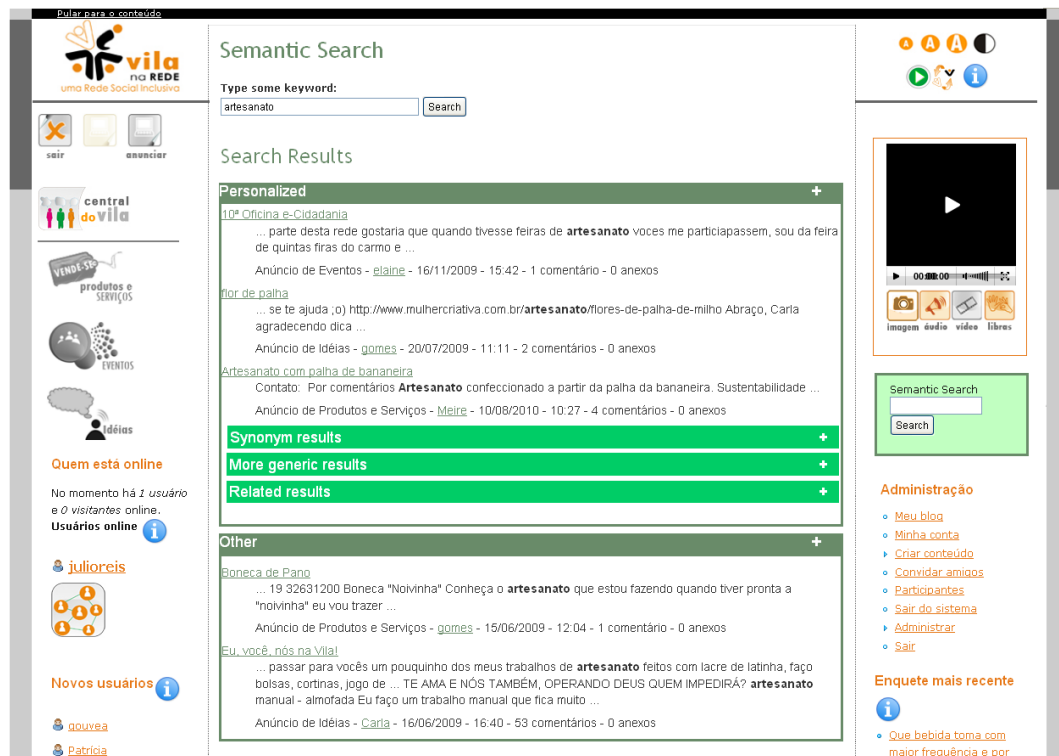


Figure 7. UI of semantic search results in VilanaRede

As an example of search results based on the existential relationships modeling (ontological dependencies in OC), Figure 8 presents a search result regarding “health”, when the keyword used in the search is “wellbeing” (in the OC modeling, “wellbeing” was modeled as ontologically dependent on the concept “health”).

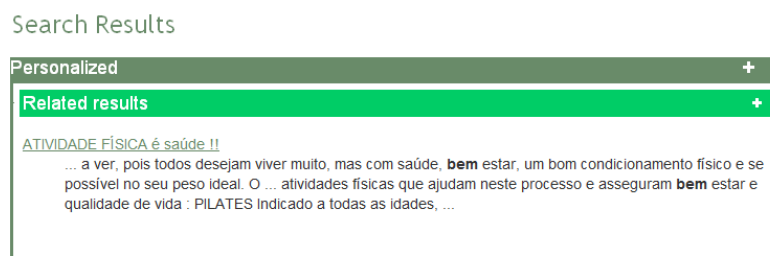


Figure 8. A specific search result exemplifying the existential relationship.

6.3 Preliminary Evaluation of the Search Mechanism

Besides the semantic aspects regarding the search results, it is also important to analyze how users make sense of the UI solution. For example, would people recognize the search boxes and the organization proposed for the semantic search results? Would they easily interact with the boxes that contain the results?

Subjects: An evaluation was carried out with 25 real users. The evaluation included both people that had already interacted with *VilanaRede* before, (16 people) as well as people that had never used *VilanaRede* (9 people). The group represented a diversity of profiles and included both technology informed people, and people with little experience with computers. The participants lived in two cities of São Paulo State (Brazil), Pedreira (4 people), and Campinas (21 people). All people performed the proposed activity individually. People from Pedreira carried out the activity in a telecenter, while people from Campinas did it at home or at their place of work.

As Figure 9 shows, with regard to age, 48,0% of the participants were between 51 and 60 years old, and 40,0% between 21 and 30 years old; 68,0% of the participants were women. Considering their level of education, 36,0% of the participants held a university degrees; 24,0% had high school degrees, while 8% had not completed high school; 20,0% had just elementary school education, and 12,0% had not completed elementary school. The rationale was to reproduce a heterogeneous population in terms of educational attainment, represented as dashed lines in Figure 9 (around 1/3 of the participants with university degree (36,0%), 1/3 of the people from elementary school to high school (32,0%), and 1/3 of the people with elementary school education or only a few years of formal education (32,0%)). As the ISM is intended to be “for all”, heterogeneity is a key factor to be considered.

The experiment also included people with different social profiles such as housewife, cook, handicraftsman, hairdresser, seamstress, retiree, teacher, student, researcher, and others. Heterogeneity also entails language disparities, resulting in difficulties for the search mechanism. In fact, if the modeling and case study were focuses on a narrow user profile, for example housewife with elementary school

education, the results probably would be better (in terms of search precision). This heterogeneity is also present in the target users of *VilanaRede*.

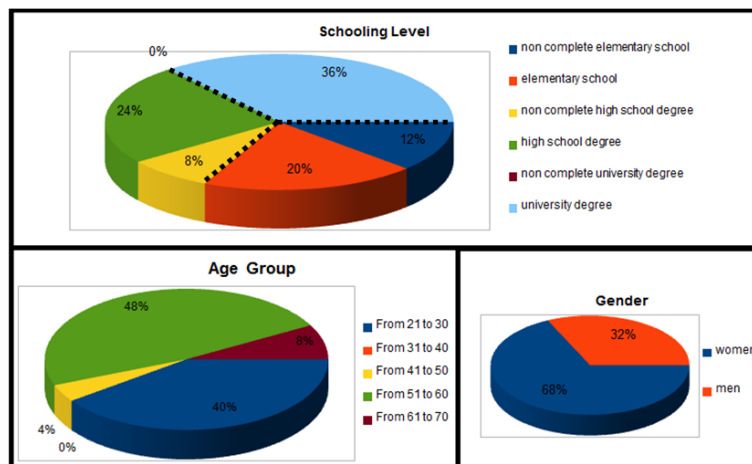


Figure 9. Graphics showing the subjects' characteristics

Furthermore, the heterogeneity considered experience with computers. Although the majority of the participants (about 72,0%) had a computer at home, and 60,0% had connection to the Internet (including dial-up connections), this does not mean that they had a high frequency of computer and Internet use, and much less that they had experience with the Internet. Around 60,0% of the participants declared that they did not use the computer frequently. Most of the participants who used the computer more frequently (40,0%) declared that younger family members assisted them (usually their sons or daughters). In this sense, the study comprises users who do not use computers, who have low contact with computers, who own computers but are not frequent users, those who own a computer and use it frequently while assisted by others, and also users who use computers frequently without assistance. Moreover, a previous experiment had been targeted to understand how these people make sense of search mechanisms, and how their educational and computer skills limitations influence their search behavior [6].

Material: Five distinct groups of people were selected to conduct the experiment. For each group, three specific search scenarios were elaborated. People of the same group performed the same three search scenarios. The search scenarios were designed specifically for each group aiming to observe and evaluate different aspects of the search (*e.g.*, polysemy in words). The search scenarios were also designed to encompass the different subjects of the *VilanaRede* content. Each

group of people was constituted according to their social identity (*i.e.*, regarding the agent concept).

Some scenarios asked the participant to search for things within a wide range of subjects (*e.g.*, to find announcements related to ‘project’), since there are different types of projects announced in *VilanaRede*, such as social projects and sustainable projects. Other scenarios asked the participant to find very specific information (*e.g.*, to find announcements that contain diseases that physical exercises may prevent).

Since the results tend to be personalized, the profile set during the user-agent modeling for each person that was in a group aimed to observe whether the results would make sense for each person. For example, a scenario asked participants to find announcements related to ‘material’ (*i.e.*, any kind of material available in the announcements). Since there are various announcements related to different kinds of material (*e.g.*, didactic material for courses, handicraft materials, recyclable material, *etc.*), and there were various profiles for the people in the group (*e.g.*, student, environmentalist, craftsman), the results regarding material should suit their profile. For instance, for a student, the personalized results should be related to didactic material.

Moreover, the aim of the scenarios was to observe whether people would recognize the synonymous search results provided, as well as the more generic, specific, and related results that could be returned; *e.g.*, if someone searched for ‘handicraft’, then the ‘specific results’ should return the different examples of handicraft available at *VilanaRede*.

Procedure: During the activity each participant of the same group performed the three search scenarios twice. Each participant did not necessarily perform the activity at the same time. Each participant used the ISM developed in the *VilanaRede* system, named here “mechanism *I*”, and also used the default syntactic search mechanism available by Drupal adopted in the *VilanaRede* system, named here “mechanism *S*”. Half of the people were chosen to start the activity using mechanism *I*, while the other half started the activity using mechanism *S*. This division was carried out within each group, *i.e.*, while half of the group started with mechanism *I*, the other half started with mechanism *S*. In groups that were composed of an odd number of people, the difference was compensated between the groups, *i.e.*, in a group of five people, two people

started with mechanism *I* and three with mechanism *S*, while in a different group of five people, three of them would start with mechanism *I* and two with mechanism *S*. The choice of which mechanism people would start with was made randomly.

Each participant had three attempts to find all possible announcements related to the question made by each proposed scenario (*i.e.*, a maximum of three attempts for mechanism *I* and three attempts for mechanism *S*, or vice-versa). Participants received support from the researcher throughout the activity. For each scenario executed, they would fill out a form answering a few questions about the search results retrieved for each mechanism in each search attempt. They would answer whether the results provided by the search mechanism were sufficient and relevant according to the scenario's goal, and would point out the more relevant announcements of the search. The participants were allowed to use any keywords for each attempt.

The participants would also answer a few questions after the finalization of each scenario (*i.e.*, after completing the attempts with both search mechanisms *I* and *S*). The questions were, for example: In what search mechanism did he/she get better results (*i.e.*, provided more important and relevant search results)?, During which attempt?; In what mechanism did the participant find the required announcements more rapidly according to the scenario's goal (*i.e.*, sufficient announcements without many attempts)?; Did the search results' ranking correspond to his/her expectations in both kinds of mechanisms?; Did the participant have any difficulty in interacting with the search boxes in the UI during the mechanism search? (*i.e.*, Did he/she recognize the search boxes or did he/she have any difficulty in opening or closing it?). In this investigation relevant search results and sufficient search results have different meanings. Relevant search results means the search results regarding the proposed scenario are pertinent in the perception of the user, while sufficient results means, the quantity of search results is enough for the user. Those terms substitute the terms *Precision* and *Recall* usually adopted in the information retrieval field, since the focus here is to evaluate the adequacy of the results from users' point of view. *Precision* and *Recall* presuppose a well defined set of relevant documents to be compared; in this work we deal with informality, including the language of people with low literacy skills, constituting a fuzzy comparison set.

In addition, at the end of the activity after performing all three search scenarios, each participant was asked which of the two search mechanisms, *I* or *S*, he/she preferred and if he/she would use the search mechanism while interacting in *VilanaRede*. Participants were also asked which kind of organization for the search results they would prefer to use during interaction with *VilanaRede* search, the *I* or the *S* mechanism.

As the experiment was with 25 people and each of them executed 3 search scenarios answering the questionnaires, 75 scenario executions were obtained for each mechanism *I* and *S*. It is worth mentioning that not all 3 attempts to perform a search were used by the participants. Results are presented in the following section.

6.3.1 Results

According to the participants' answers, 87,0% of the scenario executions indicate that better search results were found using mechanism *I*. Regarding these scenarios, 60,0% found the more important and relevant results in the first search attempt, while 29,0% found the more important results in the second attempt.

As

Figure 10 and Figure 11 show, during the execution of the search attempts with mechanism *S*, 25,3% of the scenario executions returned sufficient results (as defined in the last section) in the first search attempt, while 52,0% of the scenario executions returned relevant search results (as defined in the last section) in the first attempt. Regarding the first attempts using mechanism *I*, 54,6% of the scenario executions were considered to return sufficient search results, and 80,0% were considered to return relevant search results. Considering the first search attempt, mechanism *I* was superior to mechanism *S*.

Sufficient Results

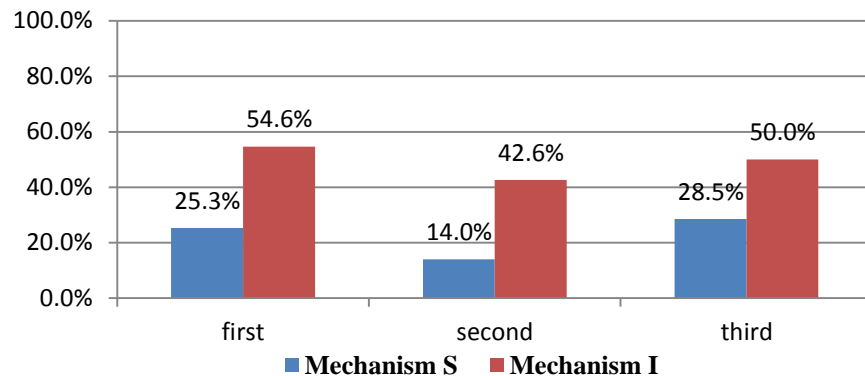
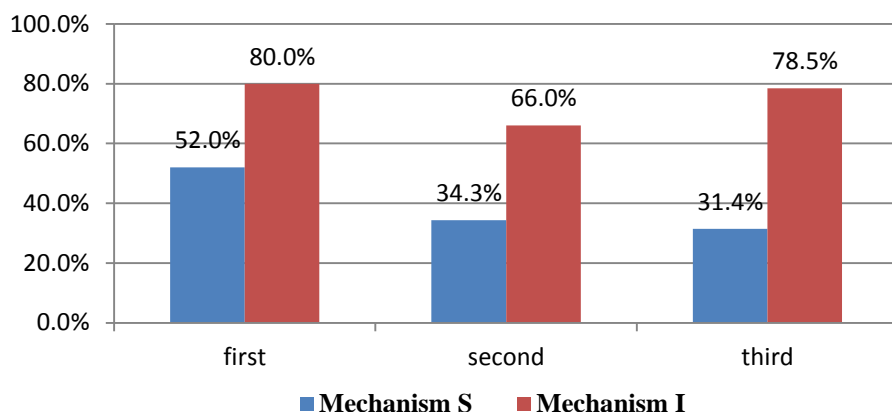


Figure 10. Comparing the mechanisms regarding sufficient results in the first, second and third search attempts

In the execution of scenarios during the second attempt using mechanism S, 14,0% were reported to return sufficient results, and 34,3% returned relevant results. In the second attempt using mechanism I, 42,6% of the executed scenarios returned sufficient search results, and 66,0% were considered to return relevant results.

Figure 10 compares the mechanisms regarding sufficient results returned, while Figure 11 shows a comparison of the mechanisms regarding the relevant

Relevant Results



results.

Figure 11. Comparing the mechanisms regarding relevant results

In the third attempt, 28,5% of the scenario executions using mechanism S returned sufficient results, while when using mechanism I 50,0% of the executions presented sufficient results. In the third attempt, 31,4% of the scenarios

executed with mechanism *S* returned relevant search results, while with mechanism *I* about 78,5% of the results were considered relevant. It is important to note that with mechanism *S*, during the three attempts, the number of executions that returned relevant results decreased for each attempt, while with mechanism *I*, despite a small drop in the second attempt, the third attempt maintained about the same level as the first (see Figure 11).

Figure 12 compares the proportion of executed scenarios in each search attempt. In all the scenarios the first attempt was executed using both mechanisms. As shown in Figure 12, mechanism *S* demanded a second attempt (according to the participants) in 85,4% of the scenarios, while using mechanism *I*, the second attempt was necessary in 72% of the scenarios. This shows that a second search attempt was necessary for more scenarios with mechanism *S*, and that the first attempt using mechanism *I* was more effective. While in 46,7% of the scenarios it was necessary to perform the third attempt using mechanism *S*, new attempt executions were necessary in 37,4% of the scenarios using mechanism *I*.

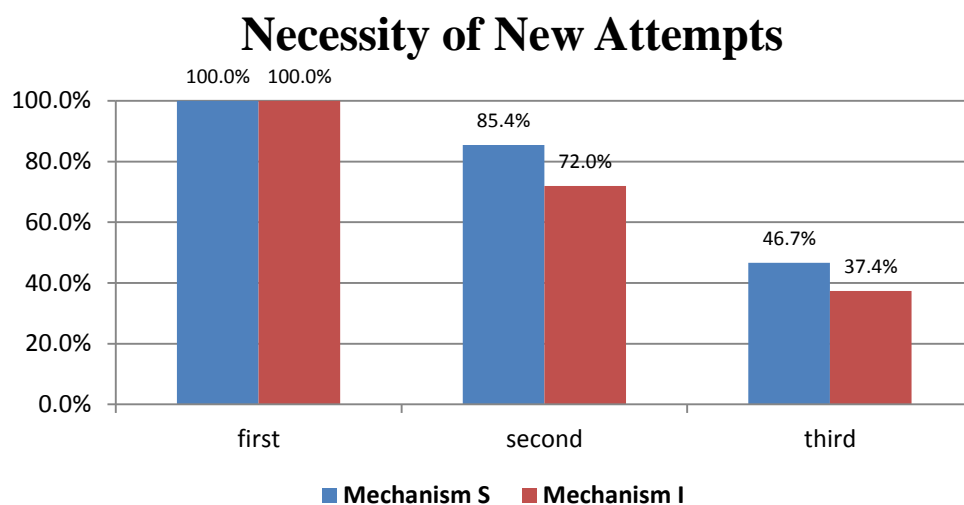


Figure 12. Comparing the mechanisms regarding necessity of new search attempts

Moreover, based on the participants' answers, the results indicate that in 88,0% of the scenario executions the results were found quickly with mechanism *I* (*i.e.*, the participant considered that he did not have to make many attempts to obtain the desired information). For mechanism *S*, 38,6% of the executions were considered to reach the desired announcements quickly.

Regarding the search results' ranking, in 86,6% of the scenario executions the search results' ranking was considered to match the users' expectations with

mechanism *I*, while with mechanism *S* this number was 49,3%. Moreover, the users wondered whether there were other relevant announcements, beyond those presented by the search results in 49,3% of the executions with mechanism *I*, while with mechanism *S* the users wondered about other relevant announcements in 82,6% of the executions. This means that when using mechanism *I*, users believed they had retrieved more relevant search results than when using mechanism *S*.

Regarding general answers after the experiment, 96% of the participants answered that they would prefer mechanism *I* for daily use in *VilanaRede*. About 76% of the users answered that they would prefer search results divided by boxes over a traditional search list; but in 24% of the scenario executions they had some difficulty interacting with the search boxes. The difficulties included understanding the purpose of the search boxes, their labels, and how to open or close them. However, it was noticed that such difficulties were concentrated in first search executions.

In addition to the quantitative results, some examples may illustrate how the semantics and social aspects make a difference in the ISM. For instance, one of the participants, a 60-year old hairdresser with incomplete elementary school education, who does not have a computer at home, illustrates some interesting points when performing a search scenario regarding “cut”. The word “cut” in Portuguese language has different meanings such as: “haircut” or “wood cutting” or a “kind of bruise”. In *VilanaRede* there are announcements regarding all these meanings for the word “cut”. Therefore, when she was performing a scenario with the mechanism *S*, the announcement regarding “haircut” was not so visible to her, but when performing the same search scenario with the mechanism *I*, the announcement regarding “haircut” was the first one in the personalized box. The hairdresser noticed such a difference, and liked to see this announcement as the first suggestion in the personalized box. This illustrates how the ISM may motivate people during the search, and improve search results according to social profiles.

Furthermore, from a qualitative point of view, participants who preferred to have the search results divided by boxes indicated that this facilitated the visualization of the results. The boxes allow a better view of the results, improving their understanding, as well as the semantic organization of more

specific results. One participant said: *“It is easier to locate what you are looking for or something similar (a related subject)”*. Those participants who preferred the traditional search list claimed that this way was easier for them, since they were more familiar with it, as it was the way they learned. One of the participants remarked that it would be interesting to join both approaches. Thus, it would be possible to join the more restricted and controlled results provided by mechanism *S* according to the input search keywords with the broader results provided by mechanism *I*, which could favor the discovery of useful and important related search results.

The main arguments given by the participants concerning their preference for mechanism *I* includes that it returns more complete and wider search results with more alternatives. One participant said: *“The information is returned in a more detailed way, more complete, but in an organized form presenting all possible search results related to the input search keywords”*. They also stated that, using mechanism *I*, they could find more satisfactory results, and that the ‘Personalized box,’ and the ‘Other box’ lead to a better division of the results and was more practical for finding information. Additionally, one participant explained that mechanism *I* was mainly useful to suggest semantic search results related to the input search keywords.

7. DISCUSSION AND RELATED WORKS

Ontologies have a fundamental role in representing semantics. To create useful ontologies with a suitable representation of reality is still an open goal; their use in services such as semantic searches involves many challenges still to be addressed. This paper has proposed a method for the construction of Web ontologies focusing on universal access to information. The method is based on SAM, which tries to model the agents and their respective affordances within a social network. This approach is different from others, as the meanings of the words are synthesized in existential relationships. In addition, social aspects are captured in the approach. First, the social agents that compose the OC diagram are extracted from the social network activity. These agents are connected to each participant in the SNS. This information is used to determine the meanings of words when creating the search index, as well as to determine the most appropriate search results. The interplay between the social agents and the semantics of the terms

directly affects the search results provided. Moreover, the proposed solution explores the combination of OS concepts with SW technologies, aiming at creating Web ontologies which can be used by an ISM following the ISN concept. The instantiation of the proposed approach, using real ISN contents, and its use for an ISM developed in the *VilanaRede*, has shown potential. The results display the impact of the solution within the developed search mechanism prototype.

Related to this work, solutions for search mechanisms in SNS have been presented in the literature which rank the search results based on various strategies (*e.g.*, recommendation and trust using the social distance between people in the SNS) [12, 13, 14, 15]. However, the currently available search solutions assume that the majority of people master the use of this technology. Therefore, these solutions are typically generic, impersonal, and constructed primarily with keyword comparisons using lexical-syntactical information processing (syntactic search). Search engines may provide a huge number of answers in response to a user query, whereas some of the most relevant are not found.

The traditional search engines are no longer able to provide precise results due to the huge volume and complexity of the information [22]. The main deficiency of these mechanisms is that they are not capable of dealing with polysemy and synonyms aspects. Thus, many times they return results that do not meet the users' needs. Semantic search has become an alternative to overcome the deficiencies of such traditional mechanisms. Works such as Heflin & Hendler [23], Guha *et al.* [24], besides Bonino *et al.* [25] and Fang *et al.* [26], have introduced the idea of using ontologies in semantic search mechanisms.

There are several architectural proposals for semantic search solutions. Mangold [27], Wang *et al.* [28], Hoang & Tjoa [45], and Hildebrand *et al.* [44] have made an extensive revision of the main proposals for semantic search solutions in the literature. However, few approaches target semantic search for SNS [11]. For instance, Choudhari *et al.* [29] describe the architecture for developing a semantic search for SNSs by using an approach based on grouping semantically related terms. This approach differs from the one presented here, since it focuses on social searches (searches for users) and does not deal with the contents in terms of their semantic aspects, which is generated and shared by the individuals. While the social semantic search is certainly important for connecting people and for many other activities in a SNS, it does not solve the problems

related to universal access to SNS content, which is approached in this paper's proposal.

Moreover, as semantic searches usually presuppose the use of ontologies, one of the major issues to consider is how to assist or automatically construct ontologies. The identification or extraction of concepts and the discovery of semantic relationships is a problem that is tackled in the Ontology Learning area. Literature reports advances in the field of developing ontologies from texts [46], as well as from semantic tags [47]. Ontology Learning in the SNS context has been explored in some studies [47, 48, 49, 50]. However, most of the experiments and solutions are still highly dependent on restricted domains and idioms.

Various techniques and tools for text mining, automatic text classification and semantic extraction can be used along with the approach proposed in this paper (section 4). In fact, they are indispensable for the scalability of the method, when considering the huge quantity of text to be analyzed and an ontology to be constructed from it. As the focus of this paper is on the design of ISM (including the ontology construction), it proposes the use of techniques and tools for text mining, automatic text classification, and semantic extraction. However, it is beyond the scope of this paper to go deeper into the analysis of these techniques and tools. During the execution of the case study (section 6) three tools were considered. In this specific case study, one crucial requirement was the application of the system to an open domain, and the use of the Portuguese language. The analysis of the tools and how they supported the construction of the ontologies is detailed in [55].

Other types of information usually provided by SNS have not yet been explored. For instance, the association of announcements with "tags" was not successful in the workshops with users in the context of the *e-Cidadania* Project. The application of WODAS in systems with a high use of "tags", keywords, or folksonomies (*e.g., Youtube* and *Twitter*) might facilitate the construction of the OCs, since they might complement or, in some cases, even substitute the analysis of the texts. One of the main research issues to be considered in this case is how to obtain reliable tags from beginners. This issue has to be addressed in future research.

These aspects are important, since many particular cases that represent social network culture might not be included in dictionaries or formal thesauruses,

which are generally used by conventional search mechanisms. An ontology model based on data from the ISN system may potentially lead to more appropriate semantic relations that will provide possibilities for semantically useful and meaningful search results. The proposed approach gives a general method for how to discover the meanings within the ISN, representing them by agents in the ontology.

The WODAS, grounded on SAM, has shown to have a positive influence for a more inclusive search solution. Such an approach enables the identification of the possible agents and their relationships, bringing a new way of representing the semantic information that supports the search mechanism. The identified and modeled agents are part of the search solution, being a special cornerstone and differential of the proposed solution. By associating them to the search, the OC may enable new semantic search possibilities, besides the approaches based on a strict hierarchy of classes. The OC enables finding existential relationships, and discovering more representative search keywords to be used by the ontology query expansion. Besides the existential relationships, the agents and their possible patterns of behavior (affordances) enabled by the method are used by the search algorithm to find more adequate and personalized search results.

During the evaluation, it was noticed that users in an initial phase of digital literacy do not have a clear concept of a search mechanism. Sometimes it was difficult for them to answer whether the search results were relevant and sufficient, since the participants would have liked to get results that did not exist within the *VilanaRede* content. In fact, they did not know that if the system does not contain a specific announcement, there can be no results for it. Due to these limitations, sometimes they had difficulty expressing their opinion and suggesting a search keyword during the activity. In this same context concerning keywords some participants expressed what they understood by certain keywords. For instance, some users expected that a keyword in singular form would return just one result (*e.g.*, ‘plate’ would return just one result for plate and ‘plates’ would return various results).

Regarding the UI proposal to organize the semantic search results, users had problems understanding the label’s meaning in the search boxes. This was observed mainly with those not used to digital concepts. Most of the novice users did not identify the search results in the boxes during their first search attempt. It

was necessary to call their attention to the boxes, and then they could understand and recognize them. They usually knew that there were more search results, and they also recognized the icon to open the boxes, but they did not comprehend what kind of results to expect inside the boxes by just reading their labels. When looking at the results inside the boxes, some people recognized the concepts described in the labels as “more generic” and “more specific results”.

Some users also did not understand the purpose of the ‘Personalized box.’ However, most of the times they liked the results available in that box, and considered the results from mechanism *I* richer than those generated by mechanism *S*. For example, a handicraftsman recognized the results in the ‘Personalized box’ as those of a different handicraftsman that was also a *VilanaRede* user. But the Personalized box should not be considered as the only aspect of the personalization; it should be combined with relations and other strategies. In addition, the possibility of finding specific concepts derived from generic ones, and vice-versa, was relevant and interesting in mechanism *I*, since users could find other related results.

The case study conducted with participants using the developed ISM in the *VilanaRede* produced satisfactory results showing the potential of the proposed approach for ISMs. It was important to preliminarily evaluate the solution by showing the possible improvements and limitations of the approach. One limitation of this evaluation is due to the lack of a comparison with a syntactic search mechanism, as no similar approach (concerning universal access and social network) was found in the Semantic Web field. In fact, the syntactic search still is the usual approach for SNS content search, and the social and digital inclusion aspects are not considered in the design of such search mechanisms. In this sense, this work goes one step further by providing ISM based on methods that model the users’ colloquial language. However, it is also necessary to recognize that there is still much work to be done to create a solution truly adequate for the context under study, which has many challenges and requires further research.

Scalability is a crucial factor to be addressed. The approach presented here is expected to be applied in SNSs with hundreds or thousands of documents, or in specific communities inside a bigger social network, where users intend to share information. As mentioned before, some tools were used to support the SAM. However, when considering huge SNSs with millions of users, a higher degree of

automation in the process is needed. The ideas and methods presented in this paper must be explored with other methods to provide scalable solutions. Some techniques, which can be used with WODAS to make it more scalable, are:

- Text mining tools, automatic text classification, semantic extraction, and SNA techniques. These may be used to provide a (semi)automatic construction of the OCs, and to automatically detect the users' profiles (user-agent relationship modeling). Studies in SNA [52] are important for the scalability of the solution. Especially in huge SNSs, SNA techniques may be helpful in the semi (automatic) identification of the users' profiles and relations. Such results can be combined with the models and techniques proposed in this paper. The study of such alternatives is beyond the scope of this paper, and should be the object of future research.
- Techniques based on Tags, Keywords, and Folksonomy may also be used in the creation of the OCs, minimizing the ontology engineer's efforts;
- Techniques for ontology alignment, merging, and evolution may be applied to assist (or automate) the ontology engineering when creating bigger ontologies from fragments identified in the analysis of the network, thus making it feasible to deal with possible intersections and conflicts;
- Ontology evolution methodologies and techniques may be explored, since the ontologies should evolve with the network content over time;
- Ontology reuse and high-level ontologies could be used to facilitate the construction of the network ontology. The proposal could be enriched and completed with an automatic text analysis component that reads the user's context and builds up the user's context, semantically based on some existing high-level ontologies. This could improve the semantic information concerning the users. Studies of the SNS interactions can also be combined with the models proposed in this paper.
- Weighed ontologies, although they might introduce additional complexity to the modeling and representation, could be used to produce more precise results in complex networks.

In general, in agreement with Hendker & Berners-Lee [55], one of the greatest challenges to SW is to turn messy human knowledge into a shared information space that is useful to everyone. In this sense, it is expected that the work reported in this paper will contribute with ideas and methods to this open challenge, especially since it includes ordinary and not technically experienced people.

8. CONCLUSION

Search mechanisms are fundamental for retrieving information. These mechanisms are even more crucial in ISN systems, which have the purpose of promoting universal access to knowledge. The semantic model that underlies and informs the search mechanism is a key point for a suitable search mechanism for ISN. In this context, Web ontologies have the potential of being very useful for the creation of Web applications, which are more adequate for diverse populations regarding their inclusion in the digital world. However, methods that are able to reflect complex social concepts and meanings are needed. In general, the application of Web ontologies to domains that are not clearly delineated is still a challenge, since such artifacts are very hard to create and to sustain (*i.e.*, they evolve over time).

This paper has proposed an approach to construct representative Web ontologies which can be used by semantic search mechanisms following the ISN concept. The Semiotic Web ontology generated and constructed from a real ISN content, was used to inform an ISM. A prototype of this ISM was developed and integrated in the *VilanaRede* system. A case study using this ISM prototype was conducted with real users in order to evaluate the approach. This evaluation included users of different social profiles, as well as users with different levels of computer experience, seeking a solution that could reach and make sense for each user, *i.e.*, a personalized solution. Positive feedback was obtained from the users involved in the case study.

Future work involves trying to produce more empirical results to evaluate the proposed approach more deeply, and facing new challenges for ontology evolution through maintaining consistency over time between the ISN content and the ontologies. Future work also includes exploring other techniques to develop

ontology-based query expansion, and to explore visualization techniques for the semantic search results.

ACKNOWLEDGEMENT

This work was funded by Microsoft Research – FAPESP Institute for IT Research (#2007/54564-1), CNPq/CTI (#680.041/2006-0), and by EcoWeb Project (#560044/2010-0). The authors thank the commitment and collaboration provided by the people of ‘Vila União’ and by the Telecenter ‘Vila Monte Alegre’. The authors also thank colleagues from CTI Renato Archer, IC/UNICAMP, InterHAD, and NIED/UNICAMP.

REFERENCES

1. Baranauskas, M.C.C. & Souza, C.S. (2006): Participative and universal access to knowledge for the Brazilian citizen. Grand Challenges in Computer Science Research in Brazil – 2006 – 2016. Available at: <www.sbc.org.br/index.php?option=com_jdownloads&Itemid=195&task=finish&cid=12&catid=50>. Accessed January 2011.
2. Boyd, D. M. & Ellison, N. B. (2008): Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*. Num. 13, pp. 210-230.
3. Baranauskas, M.C.C. (2007): e-Cidadania: Systems and Methods for the Constitution of a Culture mediated by Information and Communication Technology. Research Proposal for the Microsoft Research-FAPESP Institute.
4. Ramachandran, Divya; Kam, Matthew; Chiu, Jane; Canny, John; Franke, James L., (2007): Social Dynamics of Early Stage Co-Design in Developing Regions. In Proc. of the SIGCHI Conference on Human Factors in Computing Systems. California, USA. pp. 1087-1096
5. 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: International Conference on Informatics and Semiotics in Organisations. Beijing. pp. 247-254.
6. Reis, J.C., Bonacin, R., Baranauskas, M. C. C.(2011): Prospecting an Inclusive Search Mechanism for Social Network Services. ICEIS 2010 - Revised Selected Papers. In J. Filipe and J. Cordeiro (Eds.), *Lecture Notes in Business Information Processing (LNBIP)* Vol. 73, pp. 555-570.
7. Reis, J.C., Bonacin, R., Baranauskas, M. C. C., (2010): A Semiotic-based Approach to the design of Web Ontologies. In Proc. of 12th International Conference on Informatics and Semiotics in Organisations – ICISO 2010. Reading, UK. pp. 60-67.

8. Brazilian Internet Steering Committee. Survey on the use of Information And Communication Technologies in Brazil. 2010<www.cgi.br/english/index.htm> Accessed February 2012.
9. Internet World Stats. 2010. <<http://www.internetworldstats.com/stats3.htm>> Accessed February 2012.
10. Institute Paulo Montenegro. Indicator of Functional Literacy (2009). <www.ipm.org.br> Accessed February 2012.
11. Reis, J.C., Baranauskas, M. C. C., Bonacin, R. (2010): Busca em Sistemas Sócio-Culturais Online: Desafios e Recomendações. In Seminário Integrado de Software e Hardware (SEMISH). In Proc of the XXX Congress of the Computer Brazilian Society. Belo Horizonte: SBC, v. 1. pp. 380-394. (In Portuguese)
12. Yu, Bin & Singh, Munindar P. (2003): Searching social networks. In Proc. of the 2th International joint Conference on Autonomous agents and multiagent systems. New York, NY, USA. ACM Press. pp. 65–72
13. 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 the 16th ACM Conference on information and knowledge management. Lisbon, Portugal. pp. 563-572.
14. Haynes, J. & Perisic, I. (2009): Mapping Search Relevance to Social Networks. International Conference on Knowledge Discovery and Data Mining. In Proc. of the 3rd Workshop on Social Network Mining and Analysis. Paris, France. Article N° 2.
15. Pereira, R & Silva, S.R.P. (2008): Folksonomias: uma análise crítica focada na interação e na natureza da técnica, in Proceedings of the VIII Brazilian Symposium on Human Factors in Computing Systems. ACM International Conference Proceeding Series; Vol. 378. Porto Alegre, RS, Brazil, pp. 126-135
16. Wei Wang, Payam M. Barnaghi, Andrzej Bargiela (2008): Search with Meanings: An Overview of Semantic Search Systems. International Journal of Communications of SIWN, Vol. 3, June 2008, pp. 76-82.
17. Jacob, Elin K., 2005. Ontologies and the Semantic Web. Bulletin of the American Society for Information Science and Technology. Volume 29 - Issue 4, pp. 19 – 22.
18. Studer, R.; Benjamins, V. R; Fensel, D. (1998): Knowledge engineering: principles and methods. Data & Knowledge Engineering, v.25, n.1/2, pp. 161-197.
19. Berners-Lee, T.; Hendler, J.; Lassila, O. (2001): The Semantic Web, Scientific American, May 2001 issue.
20. Gruber, T. R. (1993): A translation approach to portable ontologies. Knowledge Acquisition, Vol. 5, Num. 2, pp. 199-220.
21. OWL Web Ontology Language. W3C Recommendation 10 February 2004. <www.w3.org/TR/owl-features> Accessed October 2010.
22. Kassim, J. M. & Rahmany, M. (2009): Introduction to Semantic Search Engine. In International Conference on Electrical Engineering and Informatics. Selangor, Malaysia. Vol. 02, pp. 380-386.

23. Heflin, J. &Hendler, J. (2000): Searching the web with SHOE. Artificial Intelligence for Web Search. In AAAI Workshop, WS-00-01, AAAI Press, Menlo Park, CA, pp. 35–40.
24. Guha, R.; McCool, R.; Miller, E. (2003): Semantic Search. Proceedings of the 12th international conference on World Wide Web. pp. 700-709. Budapest, Hungary
25. Bonino, D.; Corno, F.; Farinetti, L.; Bosca, A. (2004): Ontology Driven Semantic Search. WSEAS Transaction on Information Science and Application, Vol. 1, pp. 1597-1605.
26. Fang, Wei-Dong; Zhang, Ling; Wang, Yan-Xuan; Dong, Shou-Bin. (2005): Toward a Semantic Search Engine based on Ontologies. In Proc. of the 4th International Conference on Machine Learning and Cybernetics. pp. 1913-1918.
27. Mangold, C. (2007): A survey and classification of semantic search approaches. Int. J. Metadata, Semantics and Ontology, Vol. 2, No. 1, pp.23–34
28. 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.
29. Choudhari, Ajay; Jain, Mohit; Sinharoy, Avik; Zhang, Min. (2008): SmartSeek: A Semantic Search in Social Networks < www.cc.gatech.edu/projects/disl/courses/8803/2008/project/project_deliverables/group22/final_report.pdf >. Accessed January 2011.
30. Peirce CS (1931-1958): Collected Papers. Harvard University Press, Cambridge, Massachusetts.
31. Calegari, Silvia &Pasi, Gabriella (2008):Personalized Ontology-Based Query Expansion. In Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. Web Intelligence & Intelligent Agent. IEEE Computer Society Washington, DC, USA. Vol. 03, pp. 256-259.
32. OSW (1995): The circulation document in the Organizational Semiotic Workshop, The 1st International Workshop on Organizational Semiotics, Enschede, The Netherlands.
33. Stamper, R. K. (2000):Organisational Semiotics: Informatics http://www.cc.gatech.edu/projects/disl/courses/8803/2008/project/project_deliverables/group22/final_report.pdf without the Computer? In In-formation, Organisation and Technology: Studies in Organisational Semiotics, eds. K. Liu, R. Clarke, P. B. Andersen and R. K. Stamper. Kluwer Academic Publishers, 2001.
34. Liu, K. (2000): Semiotics in information systems engineering. Cambridge Uni. Press.
35. Stamper , R., Liu, K., Hafkamp, M. and Ades, Y. (2000): Understanding the Role of Signs and Norms in Organisations: A semiotic approach to information systems design. Journal of Behaviour and Information Technology, 19(1): 15-27.
36. Stamper, RK (1993): Social Norms in requirements analysis – an outline of MEASUR. In: Jirotk M, Goguen J, Bickerton M. (eds) Requirements Engineering, Technical and Social Aspects. Academic Press, New York.
37. Salter, A; Liu, K. (2002): Using Semantic Analysis and Norm Analysis to Model Organisations. In 4th International Conference on Enterprise Information Systems, pp. 847-850.

38. Liu, K., Sun, L. and Fu, Y. (2008): Ontological modeling of content management and provision. *Information and Software Technology*, v.50 n.11, p.1155-1164.
39. Gibson, J.J.,(1977). *The Theory of Affordances*. In *Perceiving, Acting, and Knowing*. Eds. Robert Shaw and John Bransford.
40. Reis, J.C.,Bonacin, R., Baranauskas, M. C. C., (2011): From Ontology Charts to Web Ontologies: Heuristics and Transformation Rules. Technical Report (IC-11-02), Institute of Computing, University of Campinas, Brazil. Available at www.ic.unicamp.br/~reltech
41. 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 the 10th International Conference on Enterprise Information Systems – ICEIS 2008*. Barcelona, Spain, v. 2. pp. 305-310.
42. Reis, J.C.,Bonacin, R., Baranauskas, M. C. C., (2011): Modeling Meanings from Inclusive Social Network Services. Technical Report (IC-11-03), Institute of Computing, University of Campinas, Brazil. Available at www.ic.unicamp.br/~reltech
43. Salter, A; Liu, K. (2002): Using Semantic Analysis and Norm Analysis to Model Organisations. In *4th International Conference on Enterprise Information Systems (ICEIS)*, pp. 847-850.
44. Hildebrand, M.; Ossenbruggen J., and van Hardman, L. (2007): An analysis of search-based user interaction on the semantic web. Report, CWI, Amsterdam, Holland.
45. Hoang,H.&Tjoa,A. (2006): The State of the Art of Ontology-based Query Systems: A Comparison of Existing Approaches.In:*Proceedings of the IEEE International Conference on Computing&Informatics*, Paper-Nr. 99, 06.
46. Buitelaar, P., Cimiano, P., Magnini, B. (2005): Ontology learning from text: An overview. In *Ontology Learning from Text: Methods, Evaluation and Applications*, v. 123 of *Frontiers in Artificial Intelligence and Applications*. IOS Press.
47. Mika, P. (2005): Ontologies are us: A unified model of social networks and semantics. *Proc. of the 4th International Semantic Web Conference*. LNCS 3729, Springer-Verlag, pp. 522-536.
48. Mori, J., Ishizuka, M., Matsuo, Y. (2007): Extracting Keyphrases to Represent Relations in Social Networks from Web. *Proceedings of the 20th Inter. joint Conf. on Artificial Intelligence table of contents*. Hyderabad, India, pp. 2820-2825.
49. Hamasaki, M., *et al.*, (2008): Ontology Extraction by Collaborative Tagging with Social Networking. *Proceedings of the 17th International WWW Conference*. Beijing, China.
50. Egger, M., Fischbach, K., Gloor, P., Lang, A, Sprenger, M. (2009): Deriving taxonomies from Automatic Analysis of Group Membership Structure in Large Social Networks. In: *Lecture Notes in Informatics*, vol 154, In *Proceedings of Informatik*, Lübeck.
51. Reis, J. C.; Bonacin, R.; Baranauskas, M. C. C. (2010): Ferramentas e Técnicas de Mineração de Texto em Redes Sociais Inclusivas: Promovendo Meios para o Acesso Universal ao Conhecimento. In *Proceedings of the IADIS Ibero-Americana WWW/Internet 2010 (CIAWI 2010)*. Carvoeiro, Algarve – Portugal, pp. 155-162 (In Portuguese).

52. Pfeil, U. & Zaphiris, P. (2010): Applying qualitative content analysis to study online support communities. In *Journal of UAIS*. V.16(1), pp. 1-16.
53. Medelyan, O. and I. Witten, H. (2008): Domain-independent automatic keyphrase indexing with small training sets. *Journal of the American Society for Information Science and Technology*. V. 59, I. 7, pp. 1026- 1040.
54. Lopes, L., Fernandes, P., Vieira, R., Fedrizzi, G. (2009): ExATOlP: An Automatic Tool for Term Extraction from Portuguese Language Corpora. In *Proc. of the 4th Language and Technology Conference*. pp. 427-431.
55. Hendker, J. & Berners-Lee, T. (2010): *From the Semantic Web to Social Machines: A research challenge for AI on the World Wide Web*. Artificial Intelligence. Elsevier. 174, pp.156-16.