INSTITUTO DE COMPUTAÇÃO
UNIVERSIDADE ESTADUAL DE CAMPINAS

Modeling Meanings from Inclusive Social Network Services

Júlio Cesar dos Reis
Maria Cecília Calani Baranauskas
Rodrigo Bonacin

Technical Report - IC-11-03 - Relatório Técnico

January - 2011 - Janeiro

Modeling Meanings from Inclusive Social Network Services

Júlio Cesar dos Reis[1 2], M. Cecília C. Baranauskas[1] , Rodrigo Bonacin[2]

[1] Departament of Information System at Institute of Computing
University of Campinas – UNICAMP
13083-970, Campinas, SP, Brazil

[2] Center for Information Technology Renato Archer - Rodovia Dom Pedro I, km 143,6,
13069-901, Campinas, SP, Brazil

{*julio.reis and rodrigo.bonacin*}*@cti.gov.br, cecilia@ic.unicamp.br*

Abstract

Inclusive Social Network Service (ISN) can be defined as a Social Network Service (SNS) with resources that promote access for all, including those in the margin of the digital culture. An ISN must include adequate means to recover information that make sense for all. A search mechanism capable of understanding the shared meanings used by the ISN users is still needed. In this sense methods and approaches should support capturing the social and cultural aspects from the ISN including its colloquial language and the shared meanings. In order to achieve a better understanding and representation of the semantics utilized by ISN members, this technical report presents the application and the analysis of a semantic modeling method proposed to represent meanings of terms adopted by ISN users. The outcome of the method is intended to be used by an inclusive search mechanism. This approach can enable novel ontology-based search strategies that potentially provide more adequate semantic search results.

Keywords: Inclusive Social Networks, Semantic Analysis Method, Organizational Semiotics, OWL, Semantic Search, Semiotic Web Ontology.

1 Introduction

Brazil and many other developing countries are replete of scenarios characterized by cultural diversity, as well as social and digital inclusion problems. SNSs may have a dominant role in influencing the adoption and use of information and communication technologies [1]. Moreover, they can mean an opportunity for socio-digital inclusion [2], promoting a broad access to knowledge. Nevertheless, the universal access depends also on the capabilities of the search mechanism.

An ISN can be understood as a "SNS for all"; *i.e.* a system which allows the widest diversity of people to integrate a group and interact to produce information (tangible and not tangible things) that can be shared with other persons and groups [2]. An ISN differs from other SNSs in its special attention to provide universal access to its functionalities (*e.g.* a search mechanism). The ISN functionalities are designed considering the diversity of their target audience, and its potential to help people feel comfortable while interacting with computer systems.

Based on empirical investigations, which analyzed search scenarios of an ISN, our previous research [3] has pointed out the importance of considering semantic aspects in search mechanisms. Reis *et al.* [3] argue that it is necessary to consider the influence of users' daily and ordinary language employed in the ISN. Besides, the semantic aspects can make the difference for the non-expert users, especially by the representation of the local/colloquial meanings.

Thus, in order to develop a search mechanism more suitable for ISNs, such solution should reflect the semantics used by the participants of the system [3]. The meanings must be modeled according to the patterns of behaviour shared in the ISN. For that, methods that represent the complex aspects related to the use of signs in the social network are necessary. These methods should result in computational models that represent the meanings relative to a context and to a person or a social group, which share signs and behave according to commitments, linguistic constraints, culture and other social aspects.

Within this context, and motivated by the necessity to improve Semantic Web (SW) [7] ontologies representation power and also to go toward a more human-representative Web ontology, Reis *et al.* [8] have proposed a semiotic-based approach to design Web ontologies. In this approach the meanings are first modeled using the Semantic Analysis Method (SAM) [10] from Organizational Semiotics (OS) [11, 12]; then, heuristics and transformations rules [14] are applied to get a 'Semiotic Web ontology' (SWO). An SWO is a semantic model (computationally tractable ontology) described in *Web Ontology Language*[1] (OWL) which includes concepts from SAM.

In this technical report we present the construction of SWO using real content from the *VilanaRede*[2] ISN. We show how we instantiated the method proposed by Reis *et al.* [3] in order to model the meanings from the *VilanaRede* system. All the steps are illustrated, from the analysis of the ISN content to the construction of the resulting SWO (OWL ontology). Along the method, results from text mining tools were used by (prospective) system analysts aiming to verify the support provided by these tools during the modeling process.

---

[1] www.w3.org/TR/owl-features
[2] www.vilanarede.org.br

This technical report is organized as follows: Section 2 presents the method used to represent the meanings of the terms in an ISN; Section 3 illustrates the steps performed to achieve the SWO; Section 4 presents a discussion and Section 5 draws conclusions.

2 Modeling the Semantics of an ISN

The ISN is concerned with the development of systems that make sense to a community of people aiming at promoting a digital culture. In this direction, investigations regarding more adequate search mechanism for an ISN have taken into account the user's language aspects and their applied meanings [3]. According to results described in Reis *et al.* [17], within an ISN, the search mechanism should be more precise delivering search results that make sense to users. This can generate new opportunities and facilities to access information, since users (mainly those in the process of literacy) make use of terms that are highly connected to their daily life. However, the terms employed in an ISN are not always part of the cult or formal language. Then, the meanings must be represented as they are defined in the network, thus computational artifacts that have the capability to represent the semantic of the language that come out from the ISN interactions are necessary.

Nevertheless, the discovery and representation of the employed semantics in a social network is set up as a complex modeling of a socio-cultural system [17]. There is a huge cultural diversity in the produced information besides the informality and variety of the subjects. The challenge is to deal with unstructured and informal information in natural language, and from that to achieve a well modeled and formalized knowledge, structured, interoperable, computer-manipulable and tractable allowing inference (*i.e.* a computational ontology).

Most of the information available in the Web environment and also in the SNS are in a format that can not the interpreted by machines. The search tools have not been designed to interpret or represent the meanings of search terms. They have being fundamentally constructed to respond based on keyphrases comparison and on lexical-syntactic processing, *i.e.* without considering semantic knowledge. Consequently, they do not have the capacity to differentiate the relevant and irrelevant search results from the semantics point of view. Such differentiation is mainly difficulted by polissemy factors (*e.g.* a word that has various meanings) and synonymies (*e.g.* various different words with the same meaning).

Tazi [4] argues that knowledge can be represented with the Sowa's Conceptual Graphs. The knowledge can also be represented by ontologies. It is worth differentiating Ontology that is a "branch" of Philosophy that studies the nature of the existence and the structure of reality [5] from ontologies in Computer Science context. According to Studer *et al.* [6] ontology in this context is a shared and common understanding of some domain that can be communicated between people and computers; it is a formal specification that should be readable and understandable by machines.

Semantic Web ontologies can contribute to the task of semantic search engines, since they are computer interpretable artifacts that allow inference and are part of the *World Wide*

Web Consortium[3] (W3C) standards. These ontologies have been widely accepted and used by SW communities and initiatives, and frequently they are described by computational languages for ontology description such as OWL. Wei *et al.* [20] argue that ontology and knowledge base are fundamental cornerstones for designing useful semantic search services. However, effectiveness of the semantic search depends largely on the quality and coverage of the underlying knowledge base. Thus knowledge acquisition is a bottleneck for semantic-enhanced applications besides its trust and quality.

According to Reis *et al.* [3], for an inclusive search mechanism, the data to be used by the semantic modeling phase (ontology construction) must come from the ISN. The analysis of this data may provide a way to develop a search mechanism more adequate to the language context of ISN users. Besides, an inclusive search mechanism informed by a semiotic approach can take into account the human aspects in the semantic modeling [3]. Figure 1 presents the overview of the process to construct SWOs.

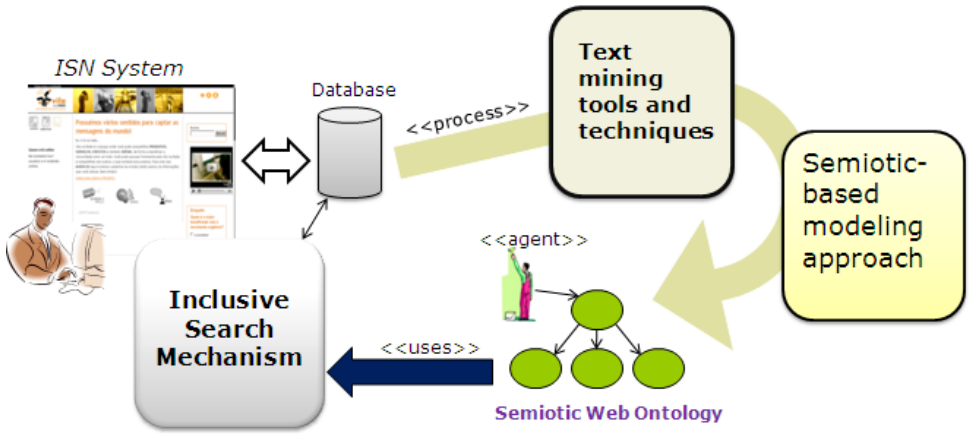


Figure 1: Method for building SWO from ISN content (adapted from [3])

From a Semiotics perspective the signs are socially constructed. Thereby, a computational model that represents the semantics from a Social Web application should contain the agents that interpret the socially shared concepts. This approach incorporates and takes to SW ontologies concerns and possible representations arising from a Semiotic perspective [8]. We have pursuing a method that is based on a subjectivist paradigm [10], which understands reality as a social construction based on the behaviour of agents participating on it, instead of a objectivist paradigm which presupposes that there exists a

[3] http://www.w3.org/

world independent of the observer. Therefore, we have used the SAM to model the semantics of an ISN. Moreover, we believe that using a traditional Web ontology approach grounded in an objectivist paradigm seems not adequate to model the meanings of an ISN, since in this particular context methods to deal with human aspects are necessary.

2.1 The Semantic Analysis Method

Using a subjectivist 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 behaviour [16]. SAM supports the analysis, specification and representation of a social system and is divided into four phases: problem definition, candidate affordance generation, candidate grouping and ontology charting [10].

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 graphical representation of a conceptual model that describes a view of responsible agents in the focal domain including their pattern of behavior named affordances and the ontological dependences between them [10]. The OC is read from left to right; anything connected to the right is dependent for its existence on the left affordance(s).

The most important concepts in SAM are Affordance and Agent. Affordance is a concept introduced by Gibson [9] that can be used to express the invariant repertories of behaviour of an organism made available by some combined structure of the organism and its environment. In SAM the concept introduced by Gibson was extended by Stamper [13] to include invariants of behavior in the social world; affordances are social constructs in a certain social context [10]. The Agent is a special kind of affordance, which can be defined as something that performs responsible behaviour. Agents are affordances that can take responsibility both for their own actions and 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.). The SAM determines the affordances, agents and the relationships between them. In the OC the agents are represented by circle, affordance by rectangles and the ontological dependencies by lines connecting agents and affordances.

SAM provides a different and independent view of the social context under study addressing issues that are not represented in any Web ontology. It provides a different way of thinking about the representation of the meanings if compared with traditional methods to model Web ontologies. One important reason for using SAM in our approach for modeling the meaning of an ISN is that, besides its semantic richness modeling based on existential relationships, one of the SAM principles according to Liu [10] is that is not allowed for an analyst to invent artificial terms or introduce new concepts when modeling the agents' actions. 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 putting them into a context of actions which 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 by the problem owners and the analyst. The reason for this is that the world to be modeled 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.

Briefly, according to Salter [16] SAM can be conducted through few steps. First the problem or the context under study is defined, usually from a written description. Following a linguistic analysis looking for verb and noun phrases applied to the domain description results in a list of potential candidate for affordances and agents. In this report, this linguistic analysis was aided by text mining tools that provide terms to support the modeling. In the final stage a complete OC can be produced which gives a full indication of that social context including the ontological dependency between the affordances. The SAM has been explained in much greater detail elsewhere [11], [10].

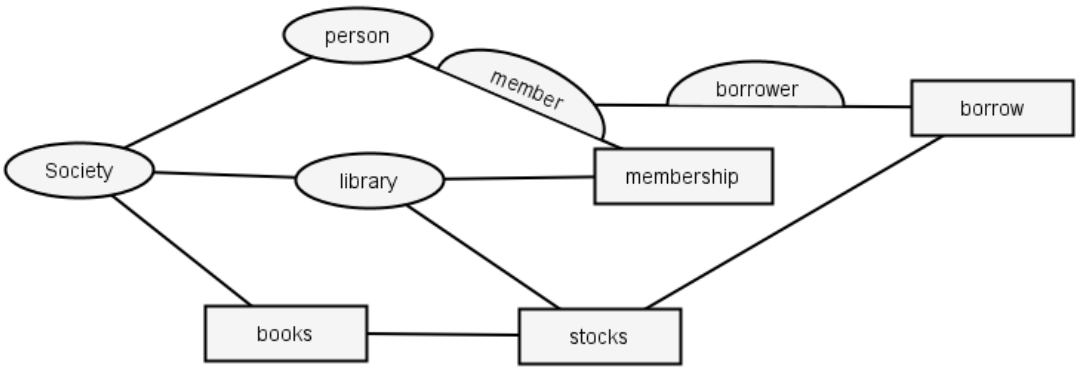


Figure 2: An example of Ontology Chart of a library domain adapted from Salter [16]

Figure 2 illustrates an example of OC adapted from Salter [16]. In this OC, if there is a library and there is a person then that person can have membership of the library. Note that both the library and the person are defined in respect to Society and are both agents. Membership of the library (an affordance) is dependent on the existence of two antecedents the library and the person. A person who has membership of the library is defined to have the role of member. Observe that a book is defined in respect to Society, although written by a person they continue to exist after the person has finished existing so are dependent on Society. Existence of the library and the book leads to the affordance 'stocks' and a combination of library membership and the library stocking books leads to the affordance 'borrow', members can borrow stock from the library. At this point the member has the role borrower. These ontological dependencies are illustrated graphically in the form of a chart (Figure 2). Following the OC reading, to borrow stocks, the stocks and membership must exist, which depend on the existence of the library.

SAM allows us to discover the agents and their respective pattern of behaviour (affordances) in an ISN. In the proposed process the result of the SAM is required to build the SWO [14]. In this process some theoretical and methodological concepts of SAM are used in conjunction with technologies from the SW to describe computationally tractable

ontologies using OWL [8]. The next section shows the steps involved in the modeling process that starts from the ISN content towards a SWO.

3 Building SWOs from the *VilanaRede* ISN

Semantic models adherent to the ISN context can better inform the search mechanism generating as a consequence more precise and adequate semantic search results to users. In order to achieve that, we have proposed to model the shared meanings from the ISN using SWO. Text mining tools are suggested to aid this modeling process in an attempt to deal with the heterogeneity problem. The case study was performed using the *VilanaRede* system.

3.1 The *VilanaRede* Content

VilanaRede is an ISN system 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 or with low literacy levels. *E-Cidadania*[4] is a Brazilian research project that has taken the challenge of developing systems that allow the access and that make sense to the community of users, contributing to the promotion of a digital culture and respecting the diversity of the population. As a result, the project has launched *VilanaRede* that has been designed with, by and for Brazilian people. Users in *VilanaRede* collaborate announcing and sharing goods and services, events and ideas.

The content domain of the whole set of announcements available in *VilanaRede* is wide-ranging; *i.e.* there are announcements about various contexts. Actually, one of the main challenges in modeling 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*", which is constituted by content from various areas and subjects independent from a specific domain. Thus the content presented in this system tends to be heterogeneous with respect to: content's subject, ways of expression, users' language, age, physical and sensorial limitations, identity, among other factors. Such announcements are diversified regarding the content as: sale of various handmade products, meals, electronic products, advocacy services, events including debate about education and "June parties"[5]; as well as ideas that cover various subject as: recipes, environmental awareness, health tips, and so on.

This investigation considered 232 announcements in *VilanaRede* distributed among products, services and ideas. The SAM was applied using these real announcements created

---

[4] www.nied.unicamp.br/ecidadania

[5] It is typical parties that occurs in some regions of Brazil in the winter season

by users of the *VilanaRede* in an attempt to model the shared meanings. These announcements also include commentaries of other users that constitute the main data considered during the modeling. Altogether 10 groups of announcements were selected according to common subjects. Thus, based on them the SAM was applied to each group resulting in 1 or 2 OC(s) for each group. The subjects of the groups are: cooking and meal ordering, sale of products and services, cultural events, divulging *VilanaRede*, physical exercises and health promotion, social projects including inclusion and citizenship, offer of courses and seminars, health-oriented food, handicraft and environment.

3.2 The Mining Text Tools as a Support for Modeling

The ontology construction is usually a hard and slow task that requires combining the knowledge of a domain specialist, with the ability and experience of an ontology engineer in a unique effort. This difficulty is still aggravated by the growing huge amount of data available together to the informality (*e.g.* in an ISN) and complexity of the subjects. Thus (semi) automatic computational solutions are necessary to aid the ontology construction. For the ISN context, Reis *et al.* [15] have conducted a study analyzing the application of possible text mining tools and techniques for the identification of relevant concepts and semantic relationships that come out from the ISN. Then the *VilanaRede* content (*i.e.* its announcements) was used to conduct such study. The authors aimed to design a strategy based on those tools to possibly assist the construction of ontology that model the meanings of the ISN system.

The tools chosen to be analyzed were: the keyphrase extractor KEA [18], ExATOlp [19] (a tool for term extraction in the Portuguese Language), and CLUTO[6], which use algorithms for clustering. Results showing positive and negative cases of the outcomes were elicited. Moreover, algorithmic procedures were created to verify the intersection of results from the outcome of each tool. Tables illustrating the results are also described to show the terms extracted with their relative and absolute frequency. The results indicate that the more adequate approach to analyze the ISN information is to verify both: the data captured by individualized announcements with data independent of any announcement. The keyphrases extracted by KEA on each announcement inform about the subjects discussed in the social network, while the approach utilized by ExATOlp provides a general view of all considered announcements. The post-processing carried out with terms organized by semantic categories shows to be useful to analyze the terms repeated by all tools, since this can indicate the concepts that are mandatory during the ontology modeling.

---

[6] http://glaros.dtc.umn.edu/gkhome/views/cluto/

3.3 Modeling OCs from Content of the *VilanaRede* ISN

The conducted experiment used the outcomes of the text mining tools to assist the OC modeling. This activity was done by sixteen (prospective) analysts from the graduate course in Computer Science at UNICAMP-Brazil who were studying Organizational Semiotics. They were divided into groups of two or three people. Each group received one subject (one of those already mentioned) collected from the *VilanaRede's* announcements. Each set of data includes among twenty to thirty announcements. Besides the text of the announcement, they also received the keyphrases extracted from each announcement. In addition, each group received two lists of terms: one was organized according to their absolute and relative frequency, and other organized by their semantic categories. Supplementary explanations of how the data was collected can be found in [15].

The first step was to teach the students on how to apply SAM and to construct the OC. After that, each group should model an OC that might represent the semantics of those *VilanaRede* announcements. The students could also use a Case Software Tool called SONAR [21] to model the OCs. They were free to define their own strategies to build up the OCs. The groups had about one week to explore the announcements and to construct their OC. In the end of the activity, each group presented the results that they have got during the modeling process; six OCs in an initial version were created by the students. Four of them were chosen to be presented in detail here. This version of the OCs is not the same of those initially delivered by the groups, since those diagrams were discussed in class and modified. The Appendix A (In Portuguese) presents the final version of the OCs.

As illustrated in Figure 3, the first OC models courses and seminars offered in *VilanaRede*. The main affordance in this context is *course*. The group expressed that a *person* can have two roles: *teacher* and/or *student* that affords *teaches* and *attends*, respectively. Both affordances depend on the affordance *course*. *Course* depends on the affordance *information*. Another agent is *institution*, and *hire* depends on it and on the role of *teacher*. Other role-name is *researcher:* a person that *realizes* research in a *topic* and *participates* in a *project*. The affordance *uses* depends on *course* and *tools* to exist. In the announcements there is a huge quantity of particular topics for the *course*, and *area of operation* for the *institution*. Such particular information was not modeled as affordances in this diagram due to the modeling granularity. They can be possible values for the determiners. Section 4 presents a discussion about granularity aspects of OCs.

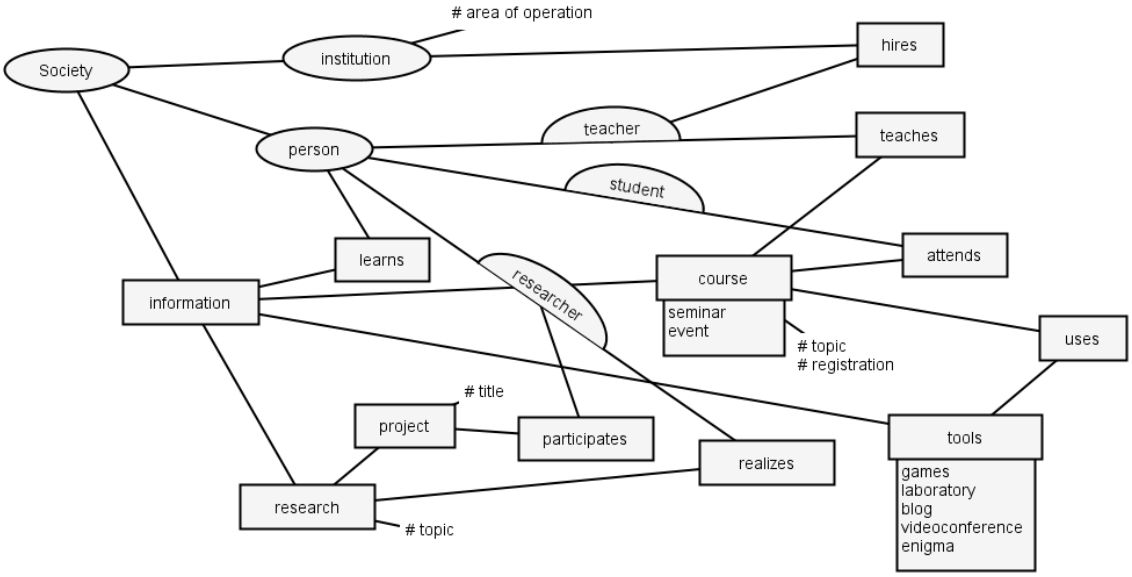


Figure 3: OC regarding courses and seminars offers modeled from *VilanaRede* announcements (original terms in Portuguese)

Figure 4 shows another OC that models announcements about health-oriented alimentation. The announcements mainly argue about the importance of the food and the consumption of some products to improve the health conditions. In this OC a main affordance is *consumption*, which has three specific kinds: *sustainable*, *healthy* and *therapeutic*. Those depend on *food* and on *consumes*. A person in a role of an *informant informs* about this kind of *information*. The affordance *consumes* depends on a *person* (in the role of *consumer*) and on *food*. *Improvement of health* is ontologically dependent of the *consumption*, and it also affords to *treat* various kinds of *diseases*.

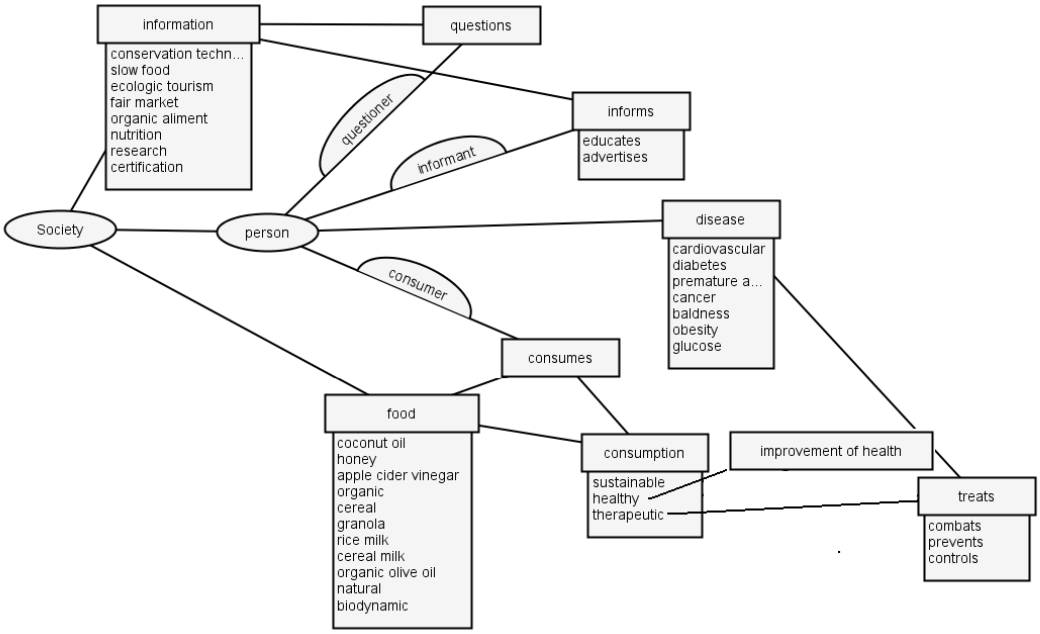


Figure 4: OC regarding health-oriented alimentation modeled from *VilanaRede* announcements (original terms in Portuguese)

The third set of announcements in *VilanaRede* is related to the advertisement of cultural events. Figure 5 shows the OC for these announcements. The main affordance in this context is the *cultural event* that depends on *funds* and *organizes*. An *institution* in the role of *sponsor* affords *funds* and an *institution* in the role of *promoter* affords *organizes*. *Participates* depends on the *cultural event* and also on the *person* in the role of *participant*. The affordance *presents* also depends on the existence of the *cultural event* and *artistic group* in the role of *presenter*. Another important affordance is *divulges* that depends on *person* (in the role of *spreader*) and *cultural event*. Furthermore, there are some specific types of cultural events and this affordance has some determiners.

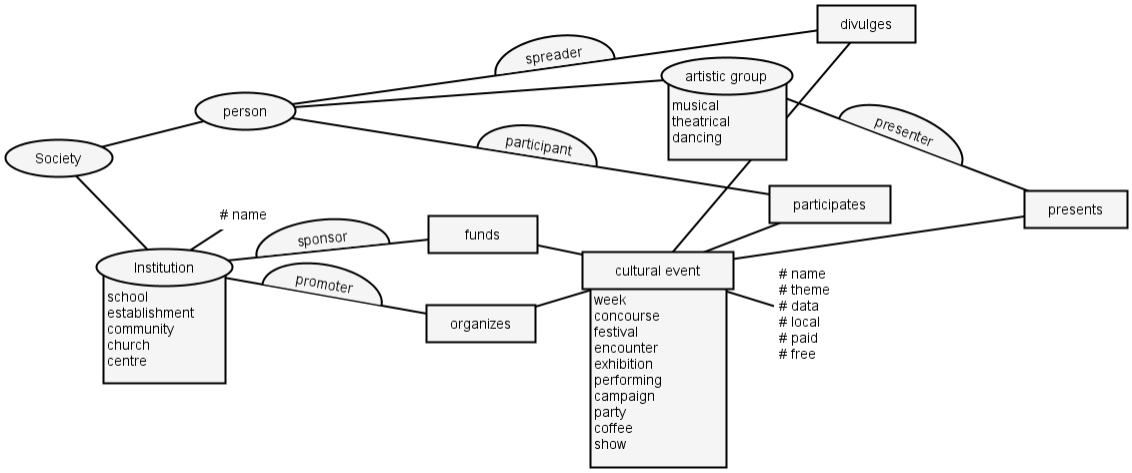


Figure 5: OC regarding cultural events modeled from *VilanaRede* announcements (original terms in Portuguese)

Another example of an OC modeled by a group of students using *VilanaRede* announcements is illustrated by Figure 6. This OC is about handicraft.

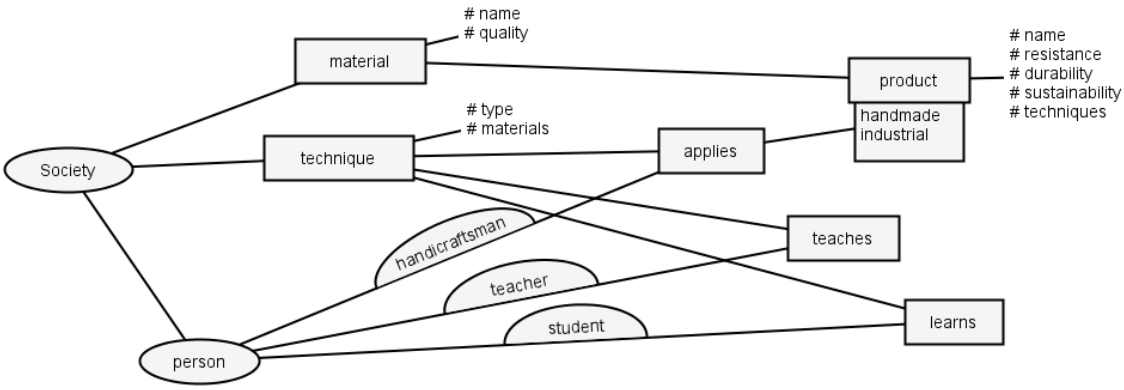


Figure 6: Other example of an OC modeled from *VilanaRede* announcements

In the end of the experiment all the students answered a questionnaire about the proposed activity. The results of this questionnaire are shown in the following section.

3.3.1 Evaluating the Support of the Mining Tools

The main aim of this part of the study was to evaluate how (prospective) analysts deal (in the modeling) with situations of informality and variety of the content. Moreover, it is important to know the difficulties of the modeling process in this context, and also how the results of the mining tools were applied. We tried to answer these questions based on the students'

statements after the activity. These data can help us to know more about the process, and consequently improve it.

At the end of the activity, the students were asked to answer five questions and to give their opinion about factors that could contribute or facilitate the modeling process. The five questions were:

1. Did the keyphrases extracted from each announcement help the diagram construction? Why? How?
2. Did the list of terms with frequencies of occurrence help the diagram construction? Why? How?
3. Did the terms divided by semantic categories help the diagram construction? Why? How?
4. Did the informality and variety of the content make the process difficult?
5. Did you face any difficulty during the utilization of the SONAR tool? Do you have any suggestion?

We used a *Likert* scale to collect the answers of the questionnaire. The participants also had a free space to comment each question. Table 1 summarizes the total of answers for all possible answers in each question for a total of sixteen participants. These results are discussed as follows.

Table 1: Summarization of the participants' answers

	Not	**Little**	**Indifferent**	**Yes**	**A lot**
Question 1	3	7	4	2	0
Question 2	9	4	1	2	0
Question 3	9	2	3	1	1
Question 4	2	2	1	8	3
Question 5	7	3	3	3	0

(1) Did the keyphrases extracted from each announcement help the diagram construction?

There are different statements about this question. The main consideration is that the keyphrases in some cases can help identifying possible agents and affordances. They might help to define possible words that could compose the diagram, but most of the words did not cover all necessary information for the construction of the diagram. Nevertheless, such keyphrases were used for most of the groups as a complement to reaffirm the findings obtained with the previous analysis of the announcements. The text of the announcements in its full version was considered important for the analysis. This happens because such keyphrases are too vague when considered alone; as they do not carry enough semantics, the

full announcement is necessary to contextualize it. The participants also stated that analyzing the full text was important to get familiarity with the subject.

Participants also said that some keyphrases were not important, since such terms seemed to be without importance in the announcement, and also sometimes the group has not used it. A few participants preferred to create a synthesis using the texts and from it to extract their own keyphrases to build the diagram. Other approach used by some groups was to create other sub-sets, grouping the announcements by pragmatic affinity inside a set of announcements.

Six participants stated that most of the announcements did not present relevant keyphrases, and that they seemed randomly extracted. Some groups concluded that the keyphrases would demand more work to filter what does not matter and add what was missing than manually extract the keyphrases from the text. Figure 7 presents graphically the results for this question. Most of the participants answered that the keyphrases present little help to the process.

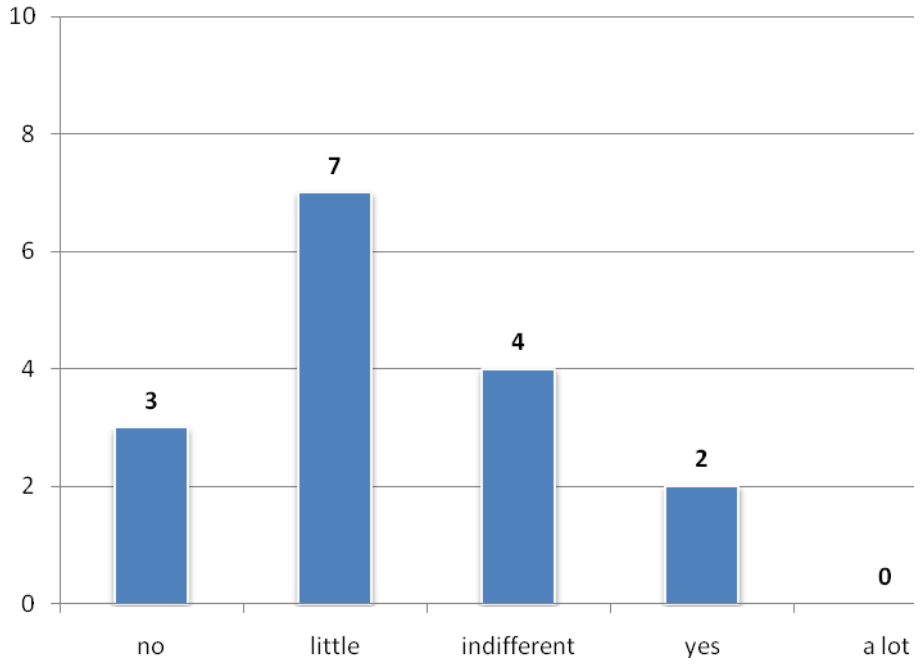


Figure 7: Results for the first question

(2) Did the list of terms with occurrence frequencies help the diagram construction?
For this question there are both positive and negative statements. As illustrated by Figure 8 most of the people answered that this list did not aid in the modeling. Some (three) positive statements say that such list is important since it shows the most frequent keyphrases. These participants also said that since the goal was to model 'a kind of high level' (abstracting the information of such context), then to know the keyphrases occurrence was important to find

the affordances that should appear in the model. The list also helps when the reading of all announcements cannot be done, thus the list is relevant to not forget some terms or ideas that could be important. Some people (three) said that the list was consulted, but it was not considered for the construction of the diagram.

Most of the participants said that the big problem was that the list covered all announcements of *VilanaRede*, and not the specific domain that they were working on (see Figure 8). Moreover, they argued that when consulting the list, some keyphrases of their context under study were in positions of little relevance in the list (*i.e.* low frequency). According to them, the list could be more useful if applied just to the set of specific announcements.

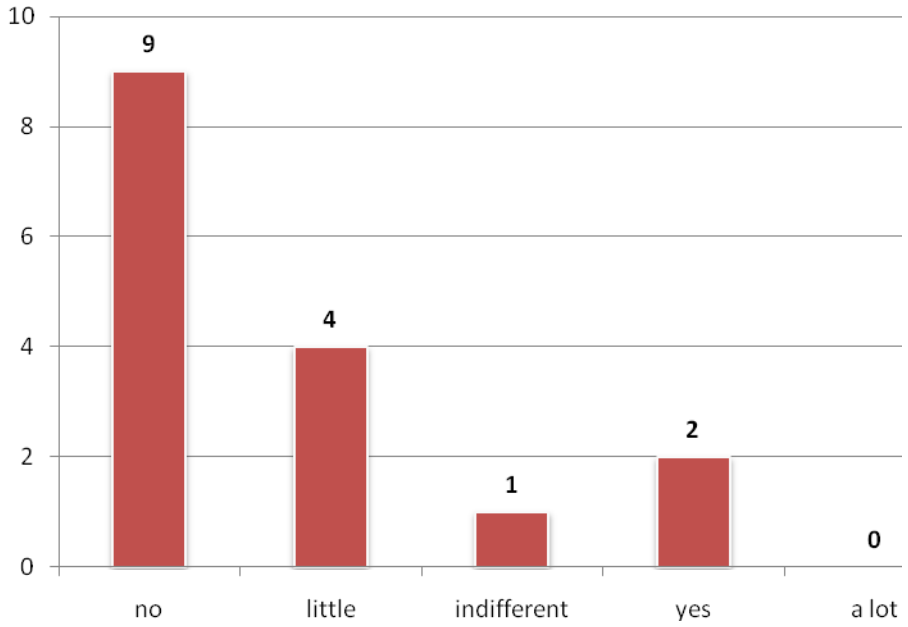


Figure 8: Results for the second question

(3) Did the terms divided by semantic categories help the diagram construction?
Concerning this list, some participants (five) stated that it was consulted but was not considered during the construction of the diagram. One participant said that the way the terms were presented were very confuse. However, a few (two) argued that the division of the terms by semantic categories allows a better visualization of the topic, and should be very important during modeling if such information were necessary, or also if the group had difficulties to categorize some word.

Some people (three) reported a problem similar to that described in the last question, since this list (of semantic categories) also mixed keyphrases from all the announcements. Nevertheless, this list could be excellent if such information referred to each announcement or to a specific set of announcements. Figure 9 shows the answers for this question.

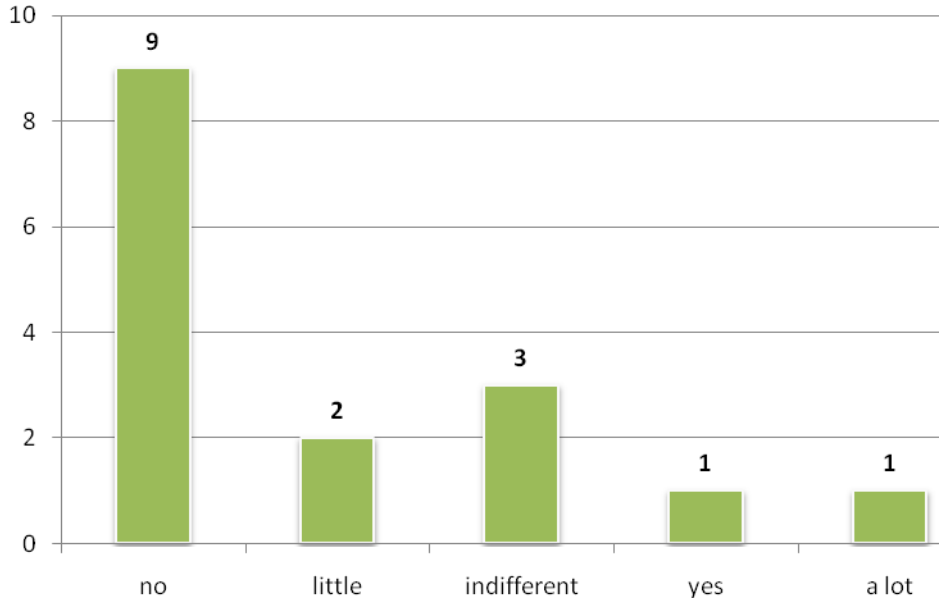


Figure 9: Results for the third question

(4) Did the informality and variety of the content disturb the process?

Some participants (four) said that they had no difficulties to deal with the informality or variety, however one of them noticed that if the announcements were not divided by sets of specific themes, those factors could introduce difficulties. A participant argued that this kind of range of informality and variety is found in a bigger proportion in the Web environment, thus this must be natural to any analysis. Most of the participants (twelve) stated that such situation makes the process difficult as presented in Figure 10.

One participant said that the variety is a hard problem, since it increases the complexity and range of the ideas to be modeled, making the modeling confuse. The variety increases the difficulties to define the core of the problem definition, and also to choose the agents and affordances. This happens mainly in the begging of the process when each announcement should be read and synthesized without the idea of the whole.

According to the majority of participants, the informality and the variety of the content complicated the extraction of representative information and the relation between them. Frequently the subject of an announcement was "distant" from other of the same set, making difficult to associate this concept with the others. The informal language used in the commentaries to the announcements was also pointed out as a problem during modeling.

Moreover, the content informality and variety implies directly in the granularity of the diagram. Sometimes, the inclusion of a new term may cause a huge modification in the diagram. One person argued that in his opinion the informality is not the problem, but the variety generates confusion. Thus, a good capacity to abstract the problem is necessary. Most of the participants agreed that it is much easier to model a context with well-defined and formalized rules in closer domains.

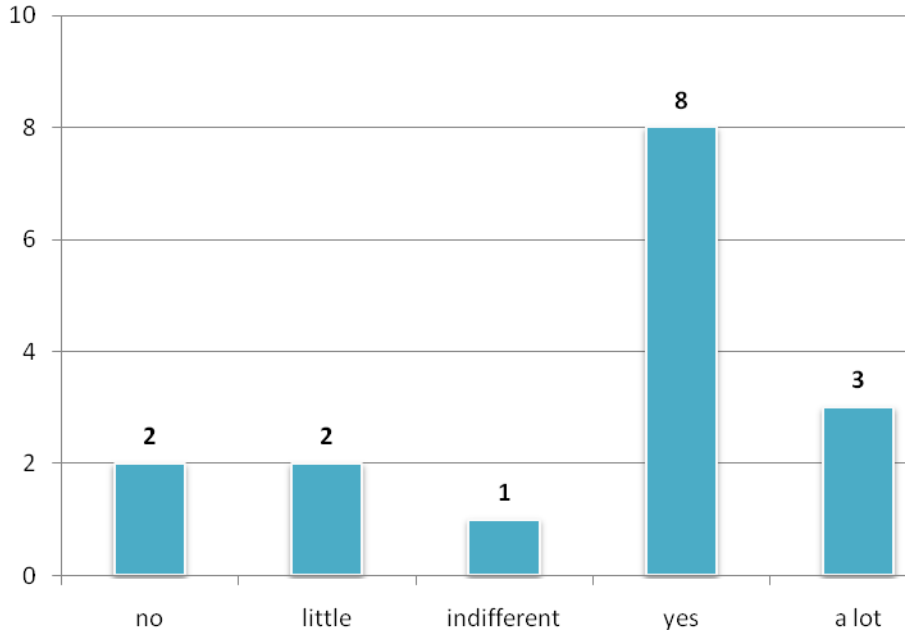


Figure 10: Results for the fourth question

(5) Did you face any difficulty during the utilization of the SONAR tool?
The use of SONAR tool was well accepted in the activity. Figure 11 shows that most of the participants had 'no' or had 'little' difficulty when using the SONAR tool. People stated that the tool is simple, efficient and functional, facilitating the modeling phase. Moreover, they said that the tool is easy to learn and has good usability. One participant highlighted that the visual metaphors of the tool offers the opportunity for a faster modeling. However, some problems and improvements were also pointed out such as: better treating the scroll bars when the diagram is too big, problems when exporting the diagram as image, and difficulties in refreshing the diagram.

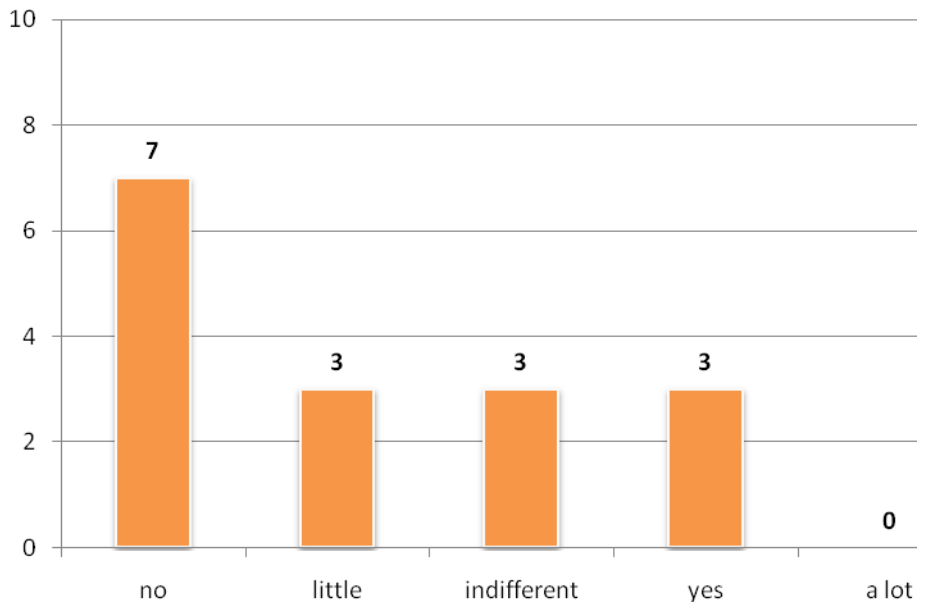


Figure 11: Results for the fifth question

The participants also made some suggestions to facilitate the modeling process from their point of view. The most common suggestions were about improving the SONAR's usability, and about new functionalities that could be aggregated to it. One participant suggested that SONAR could be integrated to a new tool that provides the frequencies and semantic categories. Other participant stated that SONAR could include new features to support the full OC notation set. They also suggested to include features for online and collaborative modeling, and to allow the use of colored lines in order to facilitate the reading of more complex diagrams. Another suggestion is the inclusion of a space where people could discuss about the ontological dependences and make it externally available.

3.4 From OCs to Web Ontologies: Modeling SWOs

The SAM and its successful application are crucial to the proposed method, since it provides important outcomes (*e.g.* OC(s)) to the next steps. In this study, once the SAM was applied based on ISN content, and the OCs were modeled, the next step proposed is to model an SWO from it. For that, the semi-automatic process proposed in [14] is applied. The process includes heuristics and transformation rules for deriving an initial Web ontology described in OWL from OC. This is an assisted process since it demands human intervention during the transformation process.

After the OC construction, the analyst must specify the affordances that should be mapped to OWL classes and the affordances that should be mapped as object properties. From that, the implemented transformation rules are able to construct an OWL file from

the OC. This OWL file includes classes, object properties and data properties derived from the heuristics.

Transformation rules, which follow the heuristics, were implemented in SONAR tool. This tool constructs one OWL for each OC. The details of how the transformation occurs and also example of OWL codes generated using the tool can be found in [14]. Figure 12 shows the OWL classes hierarchy resulted from the OC of Figure 3. The OWL class named *Agent* is the super-class of all the other agents modeled in the OC. Figure 13 shows the Object properties and Datatype properties generated by the transformation.

Figure 12: OWL class hierarchy after applying the transformation

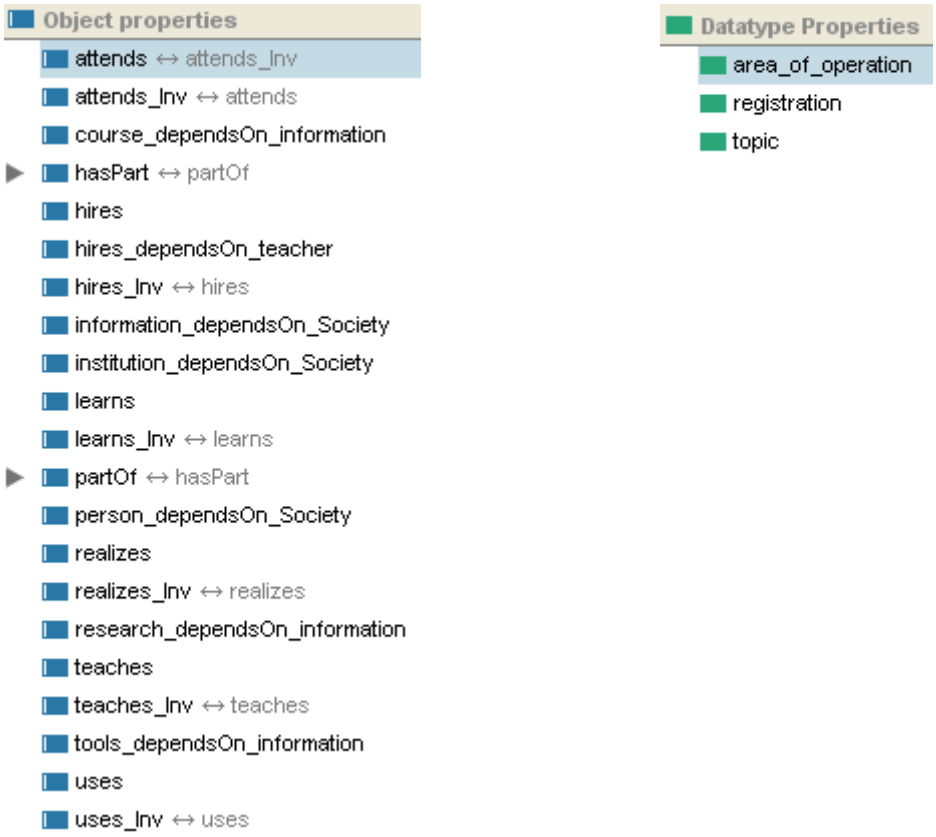


Figure 13: OWL Object and Datatype properties generated

In order to make a practical use of such model (*e.g.* by an inclusive search engine), it is necessary to merge all OWL ontologies created in this step, and also to present more details.

3.5 Merging and Detailing the SWOs

Up to this point the results provided by the method furnish diagrams that represent pieces of knowledge in a fragmented way, since it is a hard work (may be impossible) to model all the variety of contents at the same time. However, those diagrams need to be centralized and detailed with a wider coverage, since this knowledge representation must be used in practice in a semantic search engine. For that, the ontologies must be merged. After merging such ontologies, the result must be detailed, once the process so far delivers an initial version of OWL ontologies, and it is necessary that particular information of such modeled classes appear in the final SWO. Therefore, new classes can be modeled to build new relationship as well as instances of classes must be created, range of values can be added, rules defined, among other information to complement the ontology.

The merge in this work was done manually. We recognize that there are a few techniques in literature to deal with it in semi-automatic way, as ontology matching and alignment techniques. Such investigation can be addressed in further works. Classes with the same label are one of the main recurrent problems when merging ontologies. For example, most of the OCs has modeled the affordance *person*. And also, most of the time such concept is transformed to an OWL class (according to the transformation rules). Therefore, such situations must be analyzed to avoid mistakes and wrong representations. Figure 14 shows an example of merging two ontologies (the OCs of Figure 3 and Figure 6). Observing the OCs, the role-names *teacher* and *student* appear in both, as well as the affordance *teaches* and *learns*. The merged ontology will have just one class to represent such role-name (Figure 14).

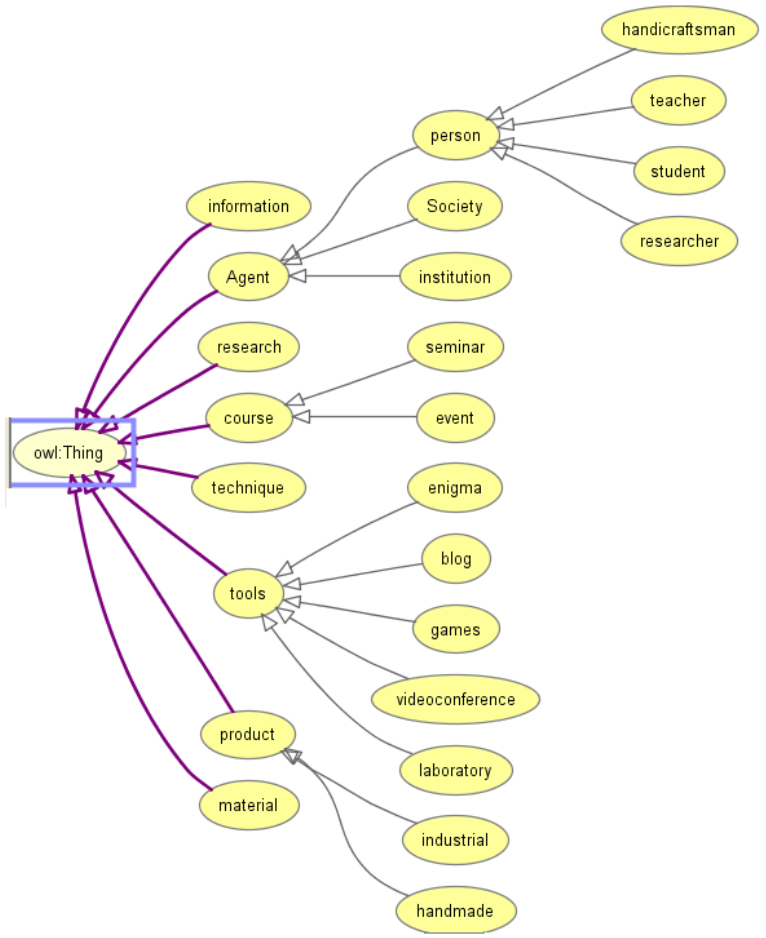


Figure 14: An example when merging the SWOs

Besides the classes, it is also necessary to pay attention to the object and data properties. According to Figure 15 both object properties *teaches* and *attends* appears once in the merged

ontology. The range in this context is manually modified since in one OC *teaches* is related to course and in the other with technique. Other properties that appear in both were also merged (*e.g. person_dependsOn_Society*). The Datatype properties contain all the determiners of both OCs.

Figure 15: Showing the merged properties

From this step on, the analyst can keep detailing the ontology using an Ontology Editor Software as Protégé[7], adding as information as necessary and also including Semantic Web Rules[8].

4 Discussion

Web ontologies have a great potential to be very useful for the creation of Web applications adequate to the diversity presented in the contemporary society. However, it is necessary to wonder whether it has the necessary modeling power to reflect complex social concepts and meanings. In general, the application of Web ontologies to domains that are not clearly delineated is still a big challenge, since such artifacts are very hard to be created and sustained (evolved) over time.

Ontology can have a fundamental role to represent semantics. To create useful ontologies with a suitable representation of reality is still a north to be followed. In this work we have pursuing the construction of more representative Web ontologies for ISN. For that, we have based on SAM, which tries to model the agents and respective affordances. The approach is different since the meanings of the words are synthesized in existential relationships. The fact of thinking in agents, affordances and ontological dependences can result in a differentiated representation of the knowledge of the context under study.

Regarding the informal contexts we have worked on, the experience in modeling the OCs have faced two main issues: the relation between the Universal versus the particular, and the diagram granularity. Due to the variety of the content, in most of the announcements used, too much things were identified as particular things. However, in the OC what should be modeled are the Universal concepts. In order to achieve it, oftentimes abstractions should be done and the particular stayed as possible values of a determiner, for example. Moreover, some affordances modeled are applied to specific interpretation that comes from the ISN content (*e.g.* a refrigerator was set as a kitchen utensil). Such affordance at the same time could be interpreted as an appliance or as furniture, or a material or a commodity, etc. Thus the meaning of the context is made by those who use it (*i.e.* the agent). In an announcement, when a refrigerator is being sold, it will be treated as a product. The modeling considers that context of interpretation (*i.e.* what make sense for the agent or the role-name that has been modeled).

---

[7] protege.stanford.edu
[8] www.w3.org/Submission/SWRL

Diagram granularity was another recurrent problem in our experience. Many words can be considered as determiner values or as affordances in the modeling. In this case, we could analyze which details would enrich the search. When such words are set as possible values of a determiner are not possible to increase the details about it (*e.g.* include a new determiner or new ontological dependence from or to it). Thus, sometimes the better option might be to model a word as an affordance in order to be able to add new modeling details, but it is necessary to know a limit of detail to stop the modeling. How much the ontology is detailed richer the context under study can be described, and as consequence the semantic information used during the search to be improved. However, to model a deep and wide context is a hard time-consuming task. Moreover, since we have made the modeling from unstructured texts described in natural language, too many details are lost during the modeling; this can also impact the semantic search. This is a recurring problem.

During the proposed activity for modeling OC from the *VilanaRede* content, various OCs were modeled, and some intersection between them could be identified. Based on the results of the activity, it is possible to observe that the variety and informality of the domain is really an aggravating modeling problem. These features decrease the efficiency of the text mining tools and also disturb the modeling process.

The experiment with the students brought important feedback about the process and about the tools support. The text mining tools outcomes were used in practice as well as the SONAR was evaluated in a real modeling context. The results point out that SONAR can be used in the process to model SWO, and also improvements could tune and facilitate the modeling process.

Additionally, we have noticed that the outcomes of the text mining tools could be more helpful if the lists were created using a specific announcement set. We considered positive and valid the use of such tools as a support, and their importance can be stressed in a situation when modeling a huge amount of information, that should be impossible to read without the help of a software tool. Nevertheless, we recognize that more investigations are necessary to improve the quality of the keyphrases. It also includes other factors that can help the analyst, and show the problem of applying it to such informal context. Another factor to be considered during the experiment was the few experience of the participants in applying SAM. They presented difficulties mainly related to ontological dependences.

We could observe that a lot of work must be done in the area of natural language processing to improve the results of the tools, and also to indicate and filter the information in the context under study. There is still some important information to be identified, *e.g.* word polissemy, *i.e.* the possible meaning of a word related to a determined agent, and also synonymous. This is an aspect that could be semi-automatic modeled as rules in the ontology. Other possibility is to identify directly a list of possible agents or role-names.

The proposed method applied in this study can be reached due to the transformations ready to use in the SONAR tool, although some information (*e.g.* possible SWRL rules) still needs to be manually modeled. Future challenges involve automatically merge and evolve the ontologies.

5 Conclusion

Search mechanisms can be decisive for recovering adequate information. This mechanism is even more crucial in ISN systems, which presupposes to promote the universal access to the knowledge for all. The semantic model that underlies and informs the mechanism is a key point to enable a suitable search mechanism for ISN. In this technical report we presented a practical experience in modeling the shared meanings from an ISN. We applied the SAM to an informal and diverse content, evaluating possible tools to help in the modeling step. It was an attempt to observe the use of such tools in practice, and also the usefulness of the created ontologies. Thus we started from the ISN content executing all proposed steps to achieve a SWO. The use of the tools was positive for the modeling steps, but they still need improvements to give better support to the process.

Further work includes the use of the created ontologies to reach a stable and detailed SWO, and the application of the SWO in an inclusive search mechanism. This mechanism is under development in *VilanaRede*. Other possibilities involve including OS Norm Analysis Method in the process and discuss its possible improvements to the process. Methods and tools to capture and integrate new affordances in the ontology under use can also be proposed. These affordances come from the ISN activities. Knowing how to deal with them, maintaining consistency between the ISN content and the ontologies is also future work.

Acknowledgments

This work was funded by Microsoft Research - FAPESP Institute for IT Research (proc. n. 2007/54564-1) and by CNPq/CTI (680.041/2006-0). The authors specially thank the participants involved in the modeling process (in the role of analysts) and also the colleagues from CTI, IC (UNICAMP), NIED, InterHAD and Casa Brasil for insightful discussions.

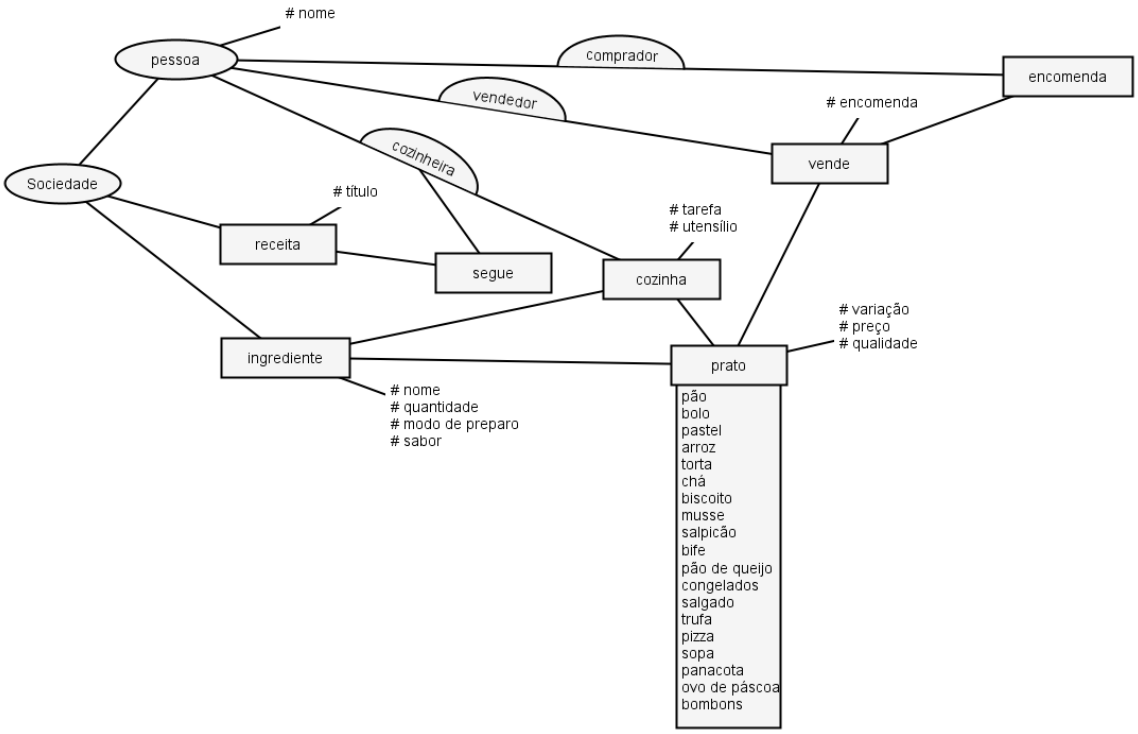
References

1. 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
2. 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.
3. REIS, J. C.; BONACIN, R.; BARANAUSKAS, M. C. C. 2010. **Prospecting an Inclusive Search Mechanism for Social Network Services.** In Lecture Notes in Business Information Processing (LNBIP), Springer-Verlag. (to be published)

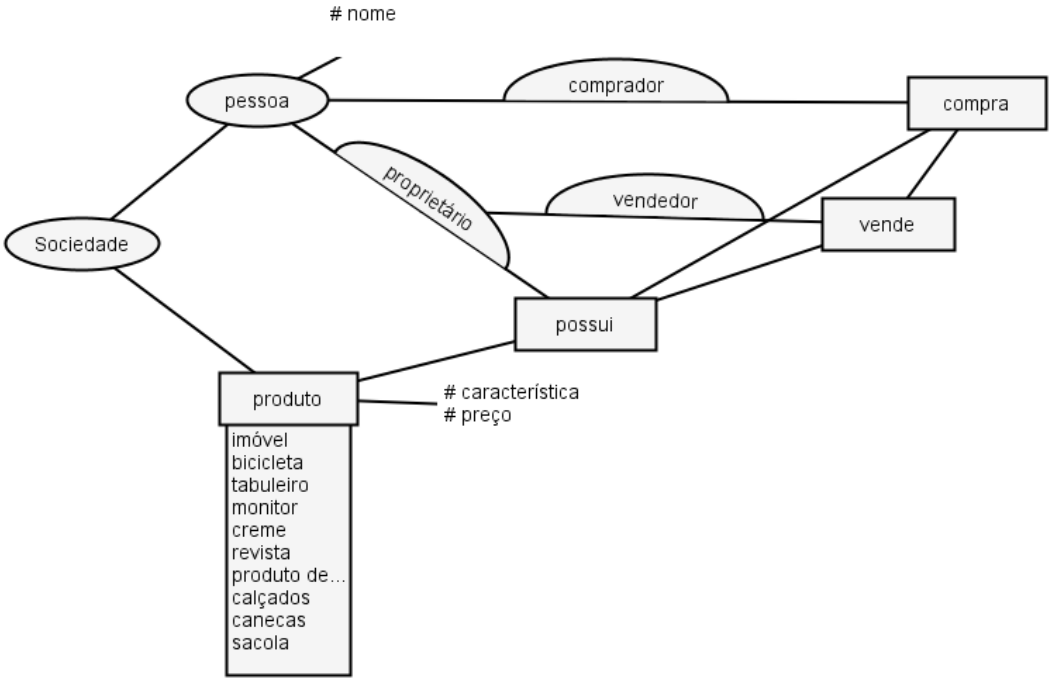
4. TAZI, S., 1994. **Using Sowa's conceptual graphs for enhancing hypertext readers performances**. Intelligent Hypertext Workshop, Washington. citeseer.ist.psu.edu/12096.html
5. JACOB, ELIN K., 2005. **Ontologies and the Semantic Web**. Bulletin of the American Society for Information Science and Technology. Volume 29 - Issue 4, pp. 19 – 22.
6. STUDER, R. *et al.* 1998. **Knowledge engineering: principles and methods**. Data & Knowledge Engineering, v.25, n.1/2.
7. BERNERS-LEE, T., HENDLER, J., LASSILA, O., 2001. **The Semantic Web**, Scientific American.
8. REIS, J. C., BONACIN, R. AND 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.
9. GIBSON, J.J., 1977. **The Theory of Affordances**. In Perceiving, Acting, and Knowing. Eds. Robert Shaw and John Bransford.
10. LIU, K., 2000. **Semiotics in information systems engineering**. Cambridge University Press.
11. STAMPER, R. K.; ALTHANS, K.; BACKHOUSE, J., 1988. **Measur: Method For Eliciting, Analysing and Specifying User Requirements**. In: Computerized Assistance During the Information Systems Life Cycle. North-Holland, pp. 67-115.
12. STAMPER, R., LIU, K; M. HAFKAMP, Y., ADES, 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.
13. STAMPER, R.K., 1993. **Social Norms in requirements analysis - an outline of MEASUR**. In Jirotka M, Goguen J, Bickerton M. (eds) Requirements Engineering, Technical and Social Aspects. New York.
14. REIS, J.C., BONACIN, R., BARANAUSKAS, M. C. C., 2011. **From Ontology Charts to Web Ontologies: Heuristics and Transformation Rules**. IC-11-02, www.ic.unicamp.br/~reltech
15. REIS, J. C., BONACIN, R., BARANAUSKAS, M. C. C., 2010. **Identificando Semântica em Redes Sociais Inclusivas *Online*: Um estudo sobre ferramentas e técnicas**. IC-10-28, www.ic.unicamp.br/~reltech. (In Portuguese)
16. 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.
17. 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)
18. 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.

19. 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.
20. 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.
21. 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.

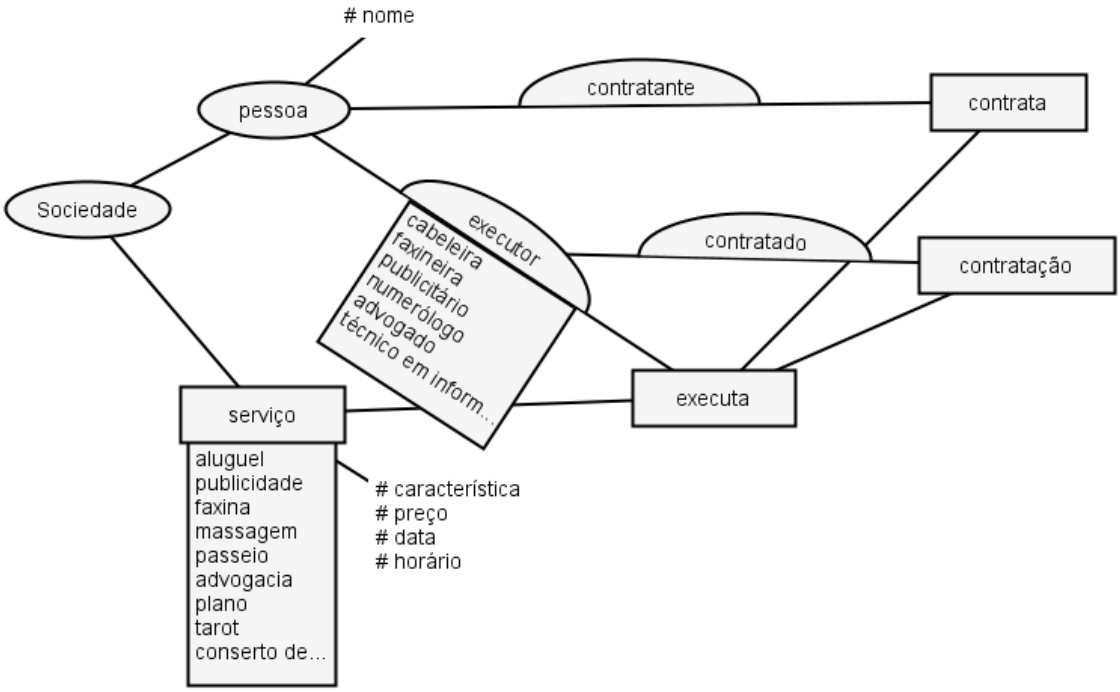
Appendix A: OCs modeled from the *VilanaRede* ISN (In Portuguese)



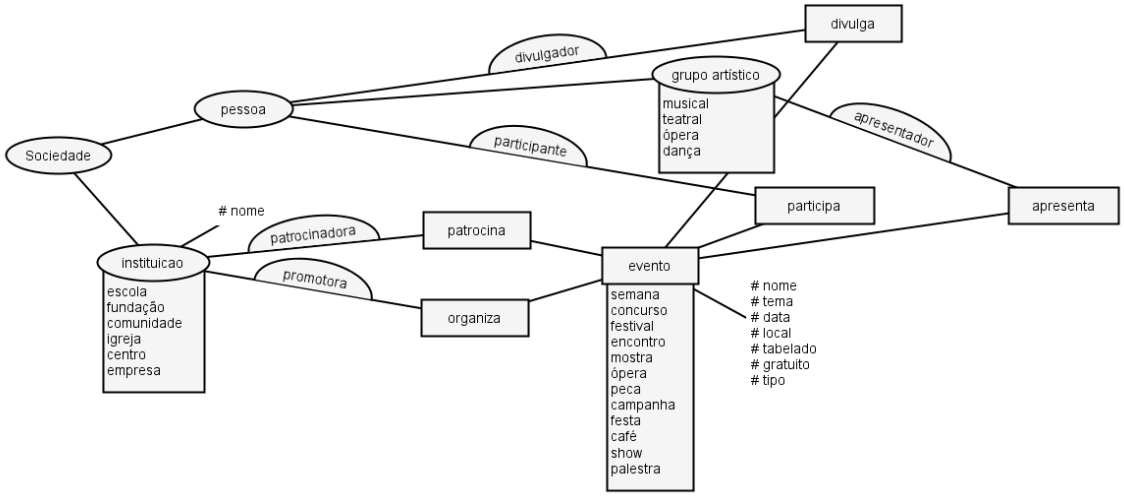
Cooking and meal ordering



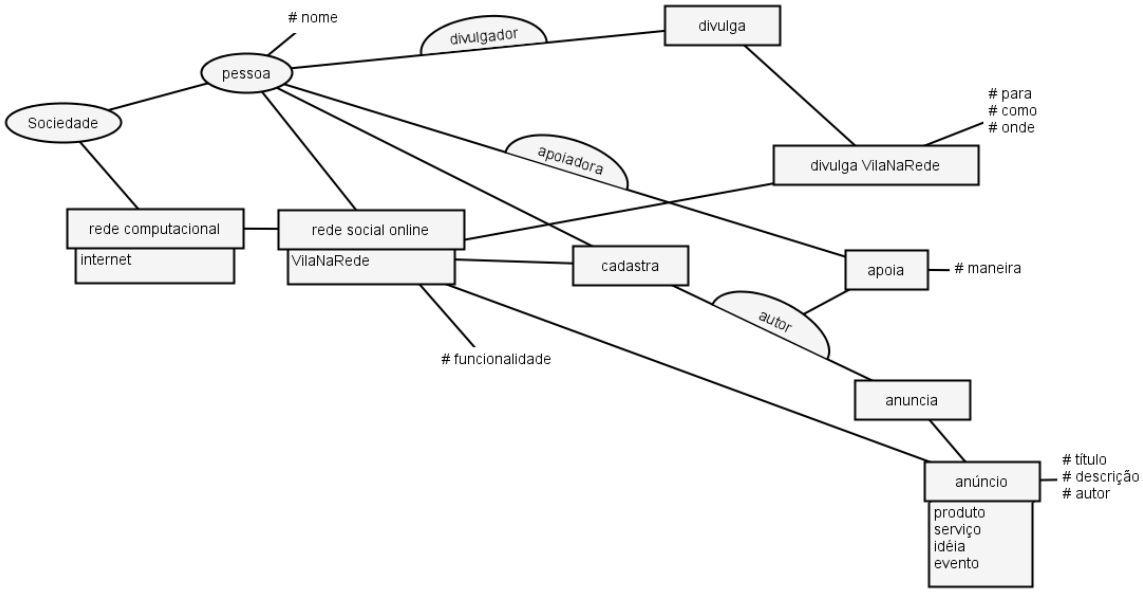
Sale of products



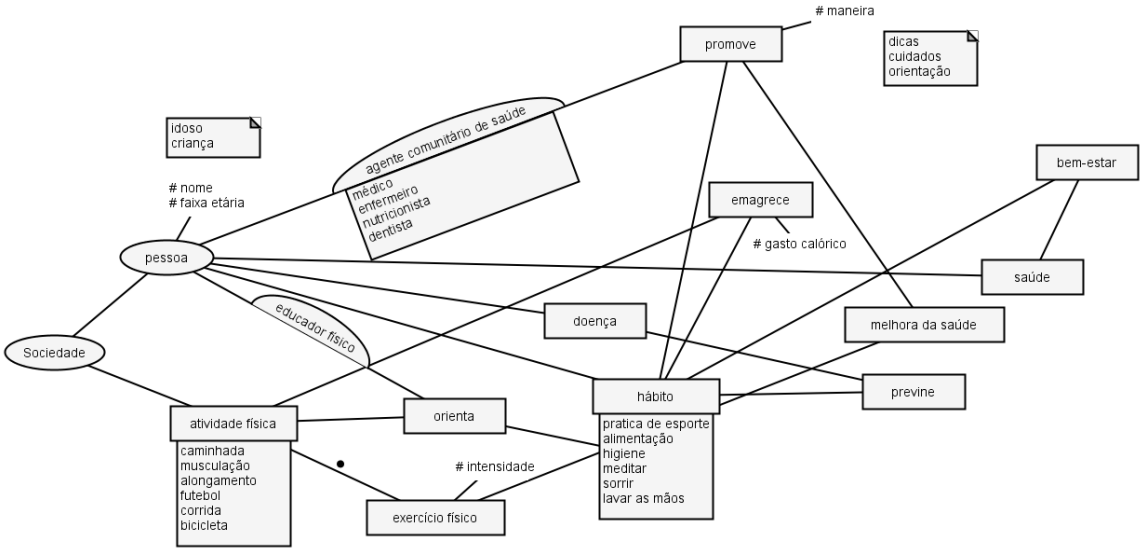
Sale of services



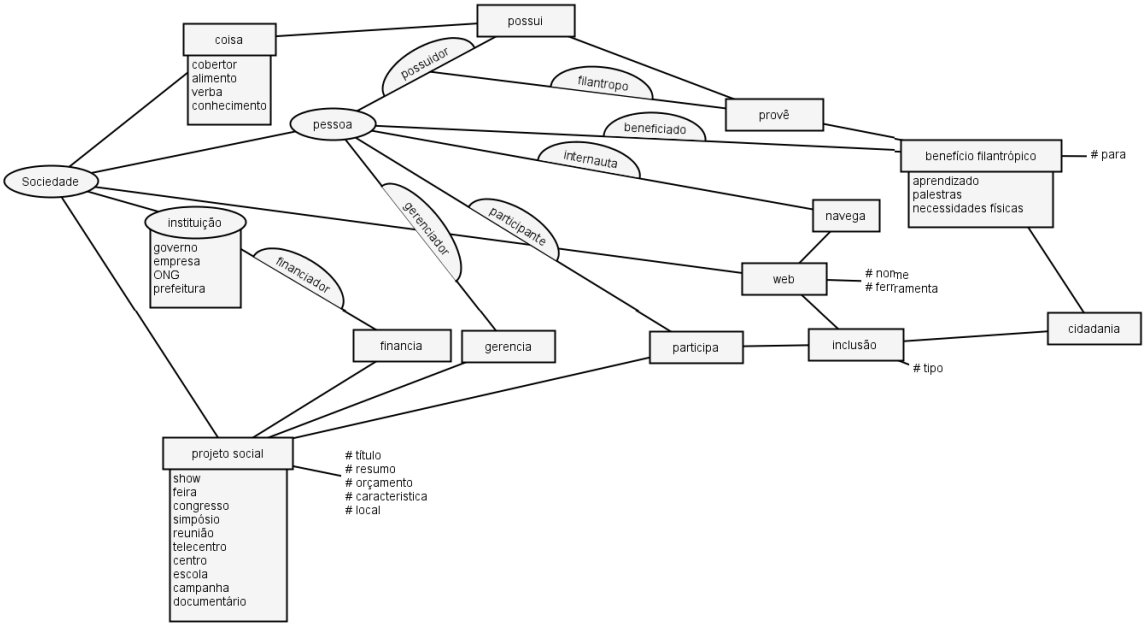
Cultural events



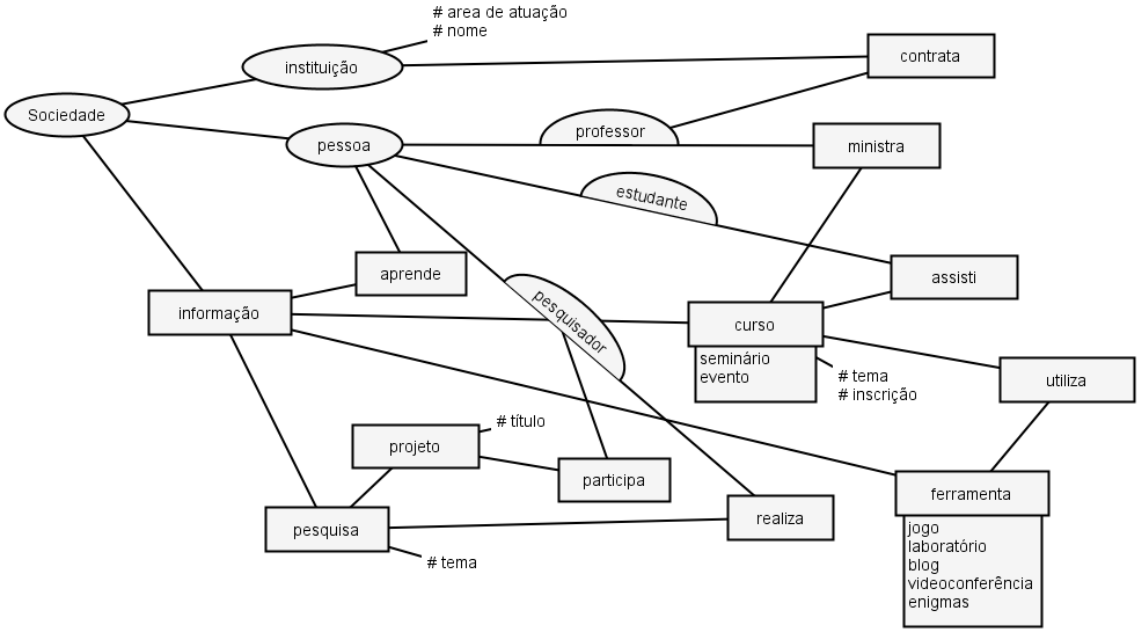
Divulging *VilanaRede*



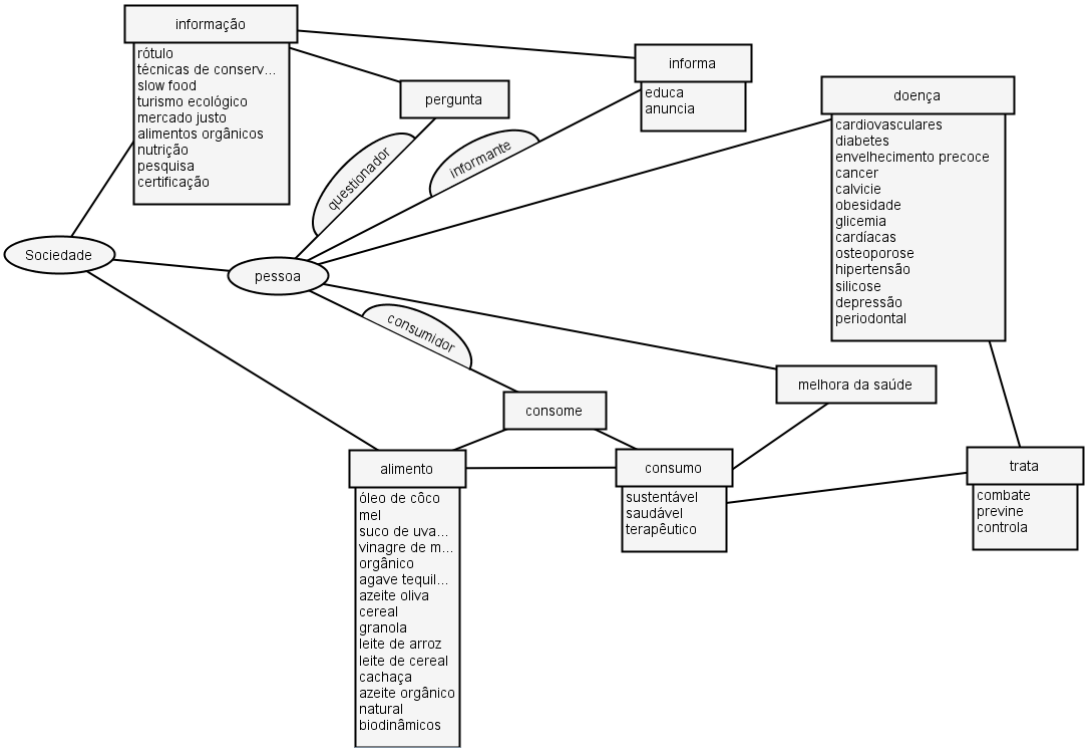
Physical exercises and health promotion



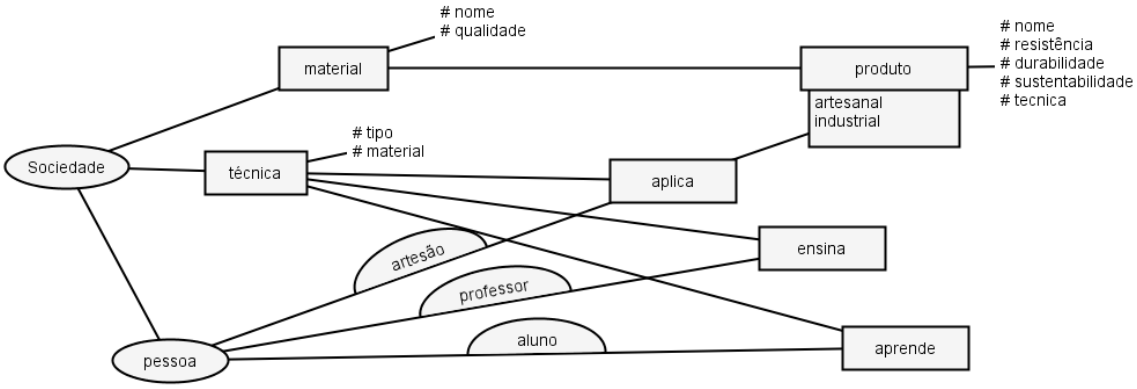
Social projects including inclusion and citizenship



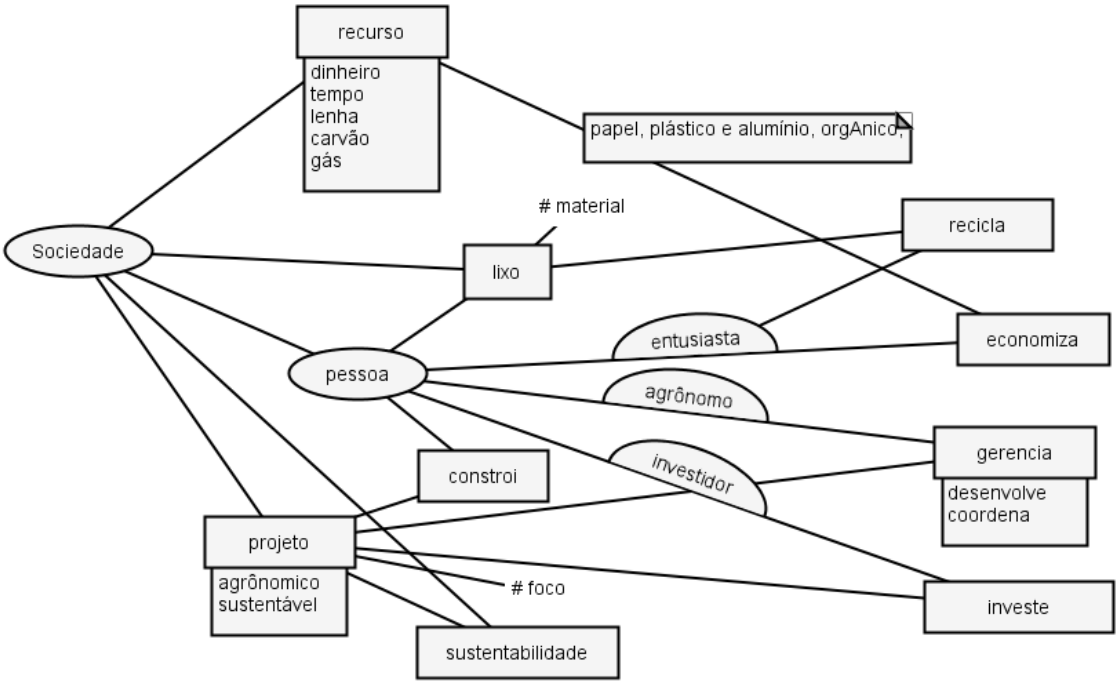
Offer of courses and seminars



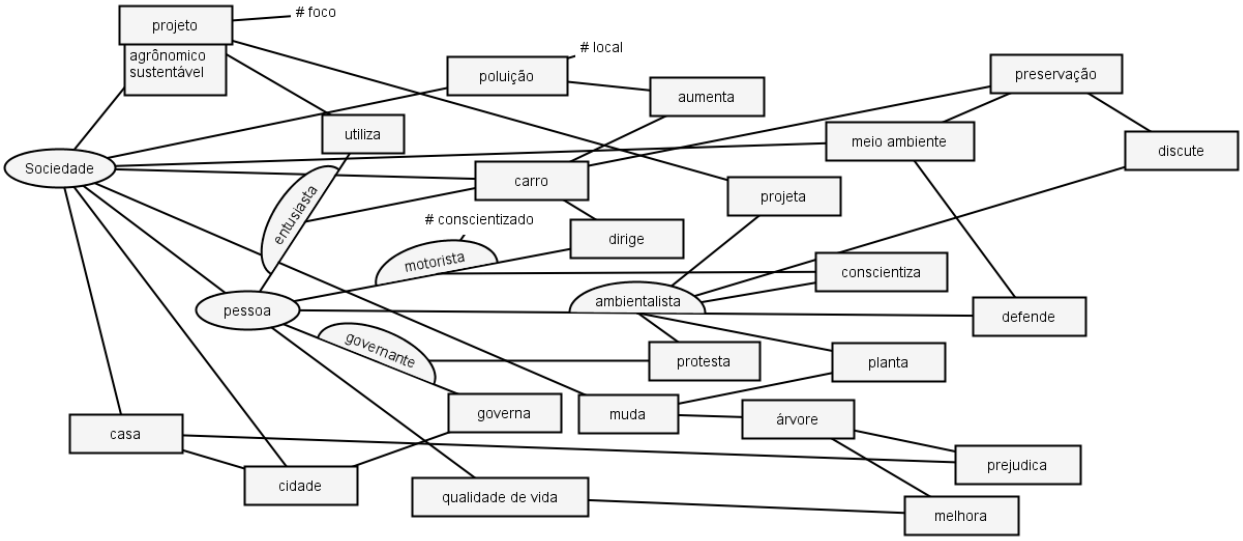
Health-oriented food



Handicraft



Environment



Environment