

Understanding Cultural Differences to Identify People with Common Interest in Social Network

Gilberto Astolfi, Vanessa M. A. Magalhães, Marcos Alexandre R. Silva, Junia Coutinho Anacleto

Advanced Interaction Laboratory – LIA
Federal University of São Carlos – UFSCar
Rod. Washigton Luis KM 235 – São Carlos – SP – Brazil
{giberto_astolfi, vanessa_magalhaes, marcos_silva, junia}@dc.ufscar.br

ABSTRACT

Social media are assumed here as a blending of technology and social interaction for the co-creation of value, as in social networks. We discuss that in such environments, it is possible to find people with common interests that potentially can promote discussions, teaching and learning from each other. However, the approaches to recommend people to interact with in the social network still do not consider the users' cultural background. This article describes an approach to identify people who are talking about the same topic in social networks, even having a different cultural background, in order to introduce to each other and leverage interaction to exchange experience, knowledge, etc. Some tests' results applying the proposal are presented. The approach was adopted in three different topics of discussion and the results suggest that the generated cultural understanding improves the chances of identifying people with similar interests.

Keywords

Cultural translation, cultural differences, similar interest, Social Network, Common Sense.

INTRODUCTION

Currently, due to the variety and ease of the use multi-media features and the expansion of the Web, people's interest in being connected through the internet is increasing, leading us to the concept of Social Media. Observing one type of social media, social networks, we can report that people also have great interest in knowing and having contact to each other and it also motivates them to be in the virtual world. Many people are focused on having more contact with other people to entertain themselves, improve their education, find a new job, among other things.

This article focus on online social networks as a particular way that allows people to keep connected to each other, meet new friends and talk about many topics, share experiences, etc. [14]. In this sense, Social Networking Services (SNSs), may be a useful tool to help people to meet others with similar interests, thus allowing a wider range of contacts and interaction among them. SNSs as

Orkut (www.orkut.com), LinkedIn (www.linkedin.com), Facebook (www.facebook.com), Hi5 (www.hi5.com) LiveMocha (www.livemocha.com), among others, have hundreds of millions of users [1]. It is important to note that each SNS may have different focus. LinkedIn, for instance, aims to connect professionals, while Orkut, Hi5 and Facebook are geared towards entertainment and LiveMocha intends to support the language teaching.

Through social networks, it is possible to approach people with common interests or topics allowing discussions, teaching and learning from each other. Some people tend to use various services such as forums or chat rooms for this purpose, but although these services are frequently used, they still do not consider the users' cultural background.

Depending on the culture, country, state, region, among others, people can express themselves differently, but with the same goal. In a search process, a person may search for "experiment" and find people who talk about it using exactly the same word, but would also be interesting to find people who talk about "experiment", but using the word "test". In this context, this paper proposes an approach that allows people to seek and find others with similar interests even though they are from different cultures, or have different ways of expressing themselves. Consider the difference may be a good way to allow people to exchange experiences, that is, cultural differences can be used to bring people together and not alienate them.

For example, a person who wants to know more about "beach Brazilian" can type those words into a search field and find many people, forums or chat rooms that discuss about this, but the probability to find something that addresses this same topic that using the term "beach wonderful" is low. In this case it is necessary to note that people with similar interests can meet or write the same things differently, i.e., "beach Brazilian" and "beach wonderful" are common topics, but the SNSs do not consider this fact in a search, in other words, they do not consider people's culture.

The cultural issues are addressed in this work considering a knowledge database from the Open Mind Common Sense, a database which stores knowledge of the people's daily lives, such as vocabularies, beliefs and myths. This

knowledge is mapped by semantic relations which, according to Minsky [13], can map this cultural knowledge to the machines. It is important to mention that the problem discussed in this paper is not about the way these services work, but in how people find others that are talking about similar topics, aiming to improve the chances of finding people whom would like to interact through a SNSs.

This paper is organized as follows: Section 2 shows some related works. Section 3 presents the knowledge database from the Open Mind Common Sense, which is the source of culture used in this work. Section 4 presents the approach proposed in this work. Finally, Section 5 shows and experiment performed with final users and Section 6 concludes.

RELATED WORKS

Literature provides several works that help people in searching for other people and/or services. Hamasaki et al. [4] propose the formation of social networks to participants in a scientific congress, considering that the participants have common interest. The authors use data mining techniques to search for information on the web about each participant and connect them in the form of social networking. After, each participant in the newly formed network may interfere in the process, eliminating people they do not know or do not want to know. Kautz et. al. [5] use data mining techniques that seek co occurrence of names of people on web sites, paper publications, university charts, etc., Forming social networks with the most co occurring people. Using the same technique as [5], but only considering the web pages, Matsuo et. al. [15] developed a system called POLYPHONET that from a user name builds a social network. Tang et. al. [6] also use the web, but are concentrated in seeking people's names and their specialties co occurrences.

The work presented by Chen et. al [3] suggests the development of techniques which provide increasing effective social network of a user. Chen et. al. considers the possibility of two people post content related to the same topic, and if that happens, they have great chances to be introduced to each another, that is, measured by comparing the sets of words typed by users, where the profile information, comments and photos are extracted. The same principle, people talking about the same topic may be introduced to each other is shared in this work.

Based on theories related to psychology, the work presented by Nunes et. al. [10] proposes to model, formalize and store a user's psychological profile, which is called User Psychological Profile (UPP). The UPP is generated by taking into account the responses to a questionnaire. It can be used by the user to identify their "inner identity" or by their "friends" with the goal of achieving the user's "Social Identity". The experimental results showed that the user's storage and processing by a recommendation system can provide a better quality recommendation. However, the big challenge is to

encourage users to answer the extensive questionnaire, consisting of 900 questions used to generate the UPP. As it can be seen, the studies mentioned here do not take into account people's cultural aspects.

This paper differs from the others related here once we have not omitted the culture of the people in the search process. Considering the cultural knowledge it is possible to find people who speak differently, but still about the same topic, increasing the likelihood individuals cluster around similar topics.

CULTURAL KNOWLEDGE DATABASE

The cultural knowledge base is obtained from a project called Open Mind Common Sense Project Brazil (OMCS-Br) [11]. OMCS-Br project has been collected culture of a general public through a web site which can be accessed by anyone through <http://www.sensocomum.ufscar.br>. After entering, the person can register and have access to various activities and themes available in this site. Most of the activities and themes are templates as shown in Figure 1. For instance, template: **Rio de Janeiro** can be called as **Marvelous city**.

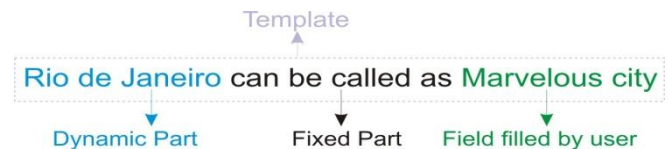


Figure 1. Example of Template.

Templates are simple grammatical structures. They have fix and dynamic parts. Dynamic parts (blue part) change when they are presented to users. They are filling out with data from other users' contribution already registered on the site. Therefore this base uses the stored knowledge to collect new one. Templates also have a field (green part) to be filled by users considering their everyday experiences, knowledge and culture. Words typed by users are stored.

These words are in natural language. Because of this, it is necessary to process them to computer to be able to use them. There is a complex process in order to create a semantic network with them. This way to store information was created by Marvin Minsky [13], researcher of Intelligence Artificial Area, whom started studying this cultural information. He believed that computers could store all the data through binary relations. In short storing data modeled as a semantic network. His theory have showing useful for the culture sensitive software development [12].

The process separates the template in two concepts and joins them through Minsky relations. For instance, the template relation "Rio de Janeiro" can be called as "Marvelous city" (see Figure 1) is DefinedAs. Because the user typed that "Rio de Janeiro" can be defined as "Marvelous city". This template is stored *DefinedAs* (*Rio de Janeiro, Marvelous city*), (see Figure 2).

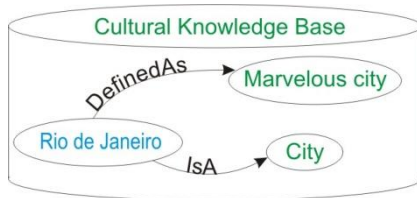


Figure 2. Illustration of the Cultural Knowledge Base.

A step in this process is Normalization, because nouns and adjectives of the sentence need are in singular and verbs in infinitive form. Avoiding that same concept can be stored in many forms, such as: Marvelous city, Marvelous cities, etc.

There are many templates to collect cultural information; another example is “Rio de Janeiro” IsA “city”. The template relation in this case is IsA. (see Figure 2). Others relations are also possible, such as: PropertyOf, MotivationOf, UsedFor, CapableOf, etc [14]. These relations are used to connect the whole information in the cultural knowledge base.

It is important to observe that there is the cultural knowledge in this base because people from different cultures, regions, etc., type what they know how about a specific topic, in these examples about “Rio de Janeiro”. Finally, the whole cultural information stored as semantic network we called as Cultural Knowledge Base (see Figure 2).

CONSIDERING CULTURAL KNOWLEDGE TO IDENTIFY SIMILAR TALKING TOPICS

Anywhere in the world, within the same language, there are varieties of vocabulary that people use to communicate. Thus, following this reasoning, it is assumed that SNSs users interact with their natural vocabularies when they need to write a text to communicate. However, taking into account this reality, it is believed that there are many users in SNSs, with the same consensus and interests related to a certain topic and context, but expressing themselves in different ways. As an example, we can consider two sentences written by two different people: "Rio de Janeiro continues beautiful" and "City beautiful remains beautiful." When we read these two sentences, we can identify that both represent a consensus of two people over a certain topic and context, that is, they said the same things differently.

The use of common search engines, e.g. Google API or the system of recommendation from people of Chen et.al. [3] does not support the identification of similarities among these people because these algorithms seek equality between words rather than consensus. In this context, this paper presents an approach that may allow Social Recommender Systems to identify and process cultural information to identify sentences written in different ways by people in SNSs, in which are influenced by their cultures, beliefs, knowledge, etc.

The approach presented in this paper is divided into three parts. First, you must have as input a sentence typed by a user in any SNS, such as a comment on a photo on Flickr (www.flickr.com) or a post on Twitter (www.twitter.com). This sentence is processed in order to construct a semantic relation among words that compose it and to represent the values of a user in relation to an topic and context. Second, the knowledge base is used to expand the culturally semantic relations previously achieved. Such expansion is represented by a directed graph, thus making it possible that people who have the same interest, but who express themselves in different vocabularies, know each other. Third, a search is conducted to identify people who are talking about the same topic, which are represented by the directed graph previously constructed.

In the following sections each of the three parts of the algorithm are explained. Subsection 5.1 illustrates the first part, the subsection 5.2 the second and finally the third part is explained in subsection 5.3.

Extracting Semantic of a Sentence

First, we need a sentence typed by user. For instance, we use a commentary that a user typed on an image in the SNS Flickr (see Figure 3).



Figure 3. Example of a sentence typed by user.

The algorithm “read” this sentence and through of a parser, called PALAVRAS a syntactic parser for Portuguese [15], identifies the grammar structure, i.e., if it has a subject, verb and object. It is possible to observe that in this example, has a subject (Rio de Janeiro), verb (continues) and object (beautiful), (see Figure 4).



Figure 4. Example of syntactic parser and Normalization.

The next step is Normalization (see Figure 4) of the concepts (words), because nouns and adjectives of the sentence need are in singular and verbs in infinitive form. Normalization is necessary to increase the potential search in cultural knowledge base, because the whole concepts in it are normalized. In addition, it is necessary to disregard the time of the verbs before searching in the SNS. Other projects as Chen [3], described in related works, do not do this process, because of this, the sentences “Rio de Janeiro continues beautiful” and “Rio de Janeiro has continued beautiful” is considered as different sentences.

Finally, next step is to create a semantic relation with components normalized of sentence. There are in this semantic relation the subject (*s*) and object (*o*) connected by verb (*v*), i.e., represented by $mr^1 = v(s, o)$. For instance, through the components of sentence: *s* = Rio de Janeiro, *v* = continue and *o* = beautiful was created: *continue (Rio de Janeiro, beautiful)* (see Figure 5).

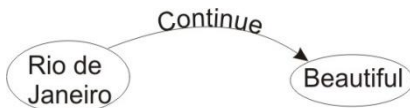


Figure 5. Semantic Relation with sentence typed by user.

The *mr* is built through algorithm that submit sentence to PALAVRAS. It returns a document “file.XML” with the parsing it. This document is analyzed in order to get *s*, *v* and *o*, i.e., the components of *mr*.

Making a summary of the whole process, basically we use a sentence typed by user to create a computational representation (semantic relation) that represents the values of the user taking into consideration the topic in a certain context.

Cultural Expansion

It is possible to build many *mr* through one *mr* taking account the same topic and context. For instance, $mr = continue (Rio de Janeiro, beautiful)$. The expansion can generate new others *mr*, for example, $mr_1 = continue (Rio de Janeiro, perfect) (Rio de Janeiro continues perfect)$, $mr_2 = keep (Rio de Janeiro, beautiful) (Rio de Janeiro keeps beautiful)$ and $mr_3 = continue (Marvelous city, beautiful) (Marvelous city continues beautiful)$. Observe that these three *mr* are identical in topic and context, ranging only linguistically, showing the main purpose of semantic expansion.

It is important to describe that this approach does not use dictionary synonymous. It is supported by the base of cultural knowledge from the project OMSC-Br, which provides knowledge as beliefs, customs, rules seen as popular laws, specific vocabulary from any group, etc. For

instance, dictionary synonymous does not have the information that “Benjamin”, “Benji” and “Franklin” are nicknames to say a hundred dollars, although this kind of knowledge is part of everyday life for many people, i.e., culture.

In order to expand culturally the semantic relation, this approach considers just two Minsky’s relation, because many semantic relations, that compose the OMSC-Br, are not suitable for this purpose. Based on our experiments we elected two relations, such as: *IsA* and *DefinedAs*. They are explained in the following sections.

IsA

IsA is considered a weak relation [14] and its purpose is to specialize something hierarchically. It is represented as follows: $IsA (X, Y)$, where *X* is a specialized concept about on generic concept *Y*. In other words, there are in *X* the whole characteristics from *Y*, because the features are derived from *Y*, but there are others features in *X*. For instance, $IsA (Rio de Janeiro, city) (Rio de Janeiro is a city)$, i.e., the whole features of “city” are part of features of “Rio de Janeiro”. On the other hand, there are others features in “Rio de Janeiro” like “own a beach”. Figure 6 shows the graphical representation of the relation *IsA*.

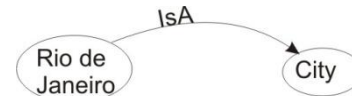


Figure 6. Graphic representation of the relation *IsA*.

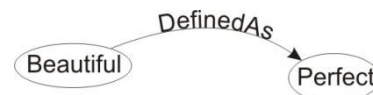
Considering before explanation, this paper considers the following definition:

- (1) If *X* is a specialization of *Y*, then *feature (Y) ⊂ feature (X)*.

This definition claims that $(X \rightarrow Y)$, i.e., **the concept represented by *Y* may be used to represent *X*, but not vice versa**. For instance, if you have the knowledge $IsA (Rio de Janeiro, city)$, you may refer to “Rio de Janeiro” using the word “city” ($Rio de Janeiro \rightarrow city$), but you cannot refer to any “city” using the word “Rio de Janeiro”.

DefinedAs

DefinedAs is a type of relation that uses synonyms to represent the meaning of something [14]. It is represented as follows: $DefinedAs (X, Y)$, where *X* is a concept with the same essence the concept *Y*. What we mean is that, there are the same characteristics in *X* and *Y*. For instance, $DefinedAs (Beautiful, Perfect) (Beautiful is defined as perfect)$, the whole characteristics of “Beautiful” are part of “Perfect”. Figure 7 shows a graphical representation of the relation *DefinedAs*.



¹ meta relation

Figure 7. Graphic representation of the relation DefinedAs.

Thus it is possible to make the following definition:

- (1) If X is synonym of Y , then $feature(X) = feature(Y)$.

This definition ensures that feature of X equal feature Y , this lead to $(X \leftrightarrow Y)$, i.e., **the concept represented by X may be used to represent Y and vice versa**. For instance, if you have the knowledge $DefinedAs(Beautiful, Perfect)$ you may refer to "Beautiful" using the word "Perfect" and vice versa.

Cultural expansion of a mr

We will show how to expand culturally a mr for a set $\beta = \{mr_1, \dots, mr_n\}$ of other new mr , where each mr owns the same meaning and semantic value that the mr base, i.e., without losing the topic and context.

When is conducted a search of cultural knowledge in the base OMSC-Br is used a concept as a reference. Then all knowledge related to the concept in question is retrieved from the cultural knowledge base. For Instance, through the submission of concept "Rio de Janeiro" there is a return of cultural knowledge similar to Figure 8 (a). In addition, it is possible to do a search involving a concept and a relation, for example, $DefinedAs(Rio de Janeiro, Y)$. Thus, everything that is related to "Rio de Janeiro" through relation $DefinedAs$ are retrieved from the base (see Figure 8 (b)).

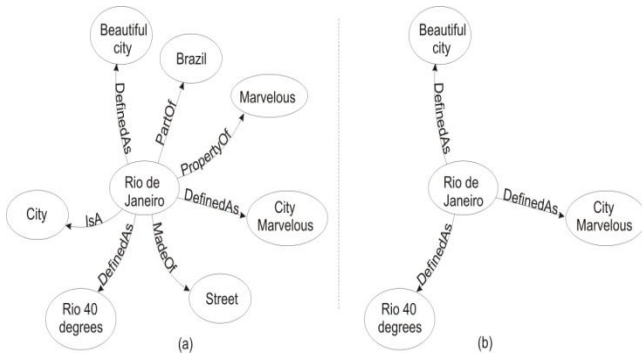


Figure 7. Example of search about cultural knowledge base.

To expand a mr culturally is used the second form of search, i.e., a concept associated with a relation. In addition, we use only relations IsA and $DefinedAs$, because they are single relations of the OMSC-Br that have the potential to represent variations in language, as shown in the above definitions (1) and (2). Parameters used in the search are the components of a mr (verb, subject and object). For instance, $DefinedAs((verb, subject or object), Y)$.

For a coherent search to the goal, which is to identify cultural knowledge that can be used to expand a mr

culturally, there are certain criteria that we must consider in the use of components of mr as parameters:

- When using the subject: $IsA(subject, Y)$, $DefinedAs(X, subject)$ and $DefinedAs(subject, Y)$;
- When using the object: $IsA(object, Y)$, $DefinedAs(X, object)$ e $DefinedAs(object, Y)$;
- When using the verb: $DefinedAs(X, verbo)$ and $DefinedAs(verb, Y)$;

The relation IsA is more limited between the two, when it is used X is always fixed, because as previously defined, if $(X \rightarrow Y)$ we can use only the word in Y to represent the word in X , but not vice versa. For instance, with the search $IsA(Rio de Janeiro, Y) Y = "city"$, the word "city" can be used to refer to the word "Rio de Janeiro", but the word "Rio de Janeiro" can not be used to refer to the "city", because when we point to "Rio de Janeiro" saying it is a "city" no impact on the meaning, but when we point to any one "city" saying it is "Rio de Janeiro" there is a great inconsistency.

We can also observe that when using the verb the search for cultural knowledge we use only relation $DefinedAs$, because there is not hierarchical representation among verbs..

To illustrate the search we going to consider $mr = continue(Rio de Janeiro, beautiful)$. Table 1 shows the search parameters and results obtained of the base of the OMSC-Br using IsA and $DefinedAs$.

Table 1. Example of search using mr component.

Parameters of search	Return in X or Y
$IsA(Rio de Janeiro, Y)$	City, state
$IsA(lindo, Y)$	There was no return
$DefinedAs(X, Rio de Janeiro)$	Rio 40 degrees
$DefinedAs(Rio de Janeiro, Y)$	Marvelous city, Beautiful city
$DefinedAs(X, lindo)$	Perfect, beautiful
$DefinedAs(lindo, Y)$	beautiful, gorgeous
$DefinedAs(X, continuar)$	keep
$DefinedAs(continuar, Y)$	There was no return

The results obtained in the search, according to Table 1, are used to expand culturally $mr = continue(Rio de Janeiro, beautiful)$. Each concept obtained in the result can be used to replace the respective component (verb, subject or object) of mr used as parameter in the search, resulting in a new mr . For instance, the concept "Marvelous city" obtained in the search with $DefinedAs(Rio de Janeiro, Y)$ can replace the component "Rio de Janeiro" (subject) in $mr = continue(Rio de Janeiro, beautiful)$, deriving it to $mr_1 = continue(Marvelous city, beautiful)$, because in the search

DefinedAs ((subject, object or verb), Y) is adopted ($X \leftrightarrow Y$), i.e., in this case Y can be used to represent X. In addition, it is generated new *mr* by permutation, i.e., we can derive another *mr* from *mr*₁, for example, *mr*₁ = continue (Marvelous city, beautiful) would generate *mr*₂ = keep (Marvelous city, beautiful), this case “keep” replace “continue”, because in the search DefinedAs(X, continue) we got the concept “keep”. It is made clear that the permutation is only possible between the same components of *mr*, i.e., subject to subject, verb to verb, and object to object.

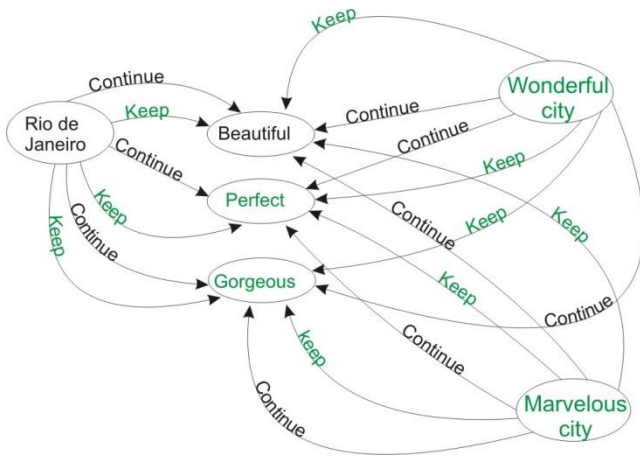


Figure 8. Example de cultural expansion.

Figure 8² shows the whole process of cultural expansion, which results in $Z = mr \cup \beta$, i.e., *mr* built from sentence typed by user and the cultural expansion its. This set of new *mr* represents a single topic concerning certain context, but so diverse that explores the possible variations of the language. It is noticed that all *mr* own the same semantic meaning and can replace one another in any circumstances.

Search for consensus SNSs

Finally, Z is used to search for users in SNSs who share the same interests, taking into account each one’s culture. In the search process only the pairs *subject* and *object* of each *mr* are sought at some block of text typed by the user in SNSs. These blocks can be found in the Orkut forums, posts on Facebook or any other SNS that enables the user to enter sentences. Once retrieved, the text blocks are examined to find the *subject* and *object*, used in the search, in the same sentence. If there is such an occurrence, the sentence is subjected to a process similar to that one described in Section “Extracting Semantic of a Sentence”, which generates a *mr*. Finally, the result is compared with *mr* used in the search to identify similarities between them. Should the successful user of the SNS, responsible for the text block, is selected to be talking on the topic and context

² In this figure is showed part of the cultural expansion, some concepts found in the search were omitted.

represented by Z. At the end of the search process, when all *mr* \cup Z are used, there is a group of users that is talking on a certain topic and context, regardless of how they express themselves.

Next section presents an experiment that investigates the potentiality of the proposed approach.

EXPERIMENT

Aiming to validate the proposed approach, an experiment was proposed to verify if the cultural expansion proposed in this work matched with real meanings as interpreted by people, allowing the recognition of similar talking topics. The experiment was performed in July, 2010 and 3 different topics were evaluated. Nineteen people with different profiles answered a questionnaire as described in the following sub-section.

Methodology

Nineteen full literate people participated in the experiment. Only literate people were considered in the experiment once it was expected from them to be able to evaluate whether a post from Orkut, selected by the method, is related to the same subject used for the search, even if it is written differently. According to [16], literate people have the ability to read and interpret a text.

Each person, who from this point will be called participant received a questionnaire with a number of posts retrieved by method, where for each post it should answer both questions. The posts were related to three separate topics coming from Orkut posts: "Does Rio de Janeiro continue beautiful?", "Michael Jackson dies at age 50" and "Lula defends Jose Sarney and says that complaints are endless."

Figure 9 illustrates part of the questionnaire applied.

Sentence used as subject: "Lula defends Jose Sarney and says complaints is endless"			
Texts	You agree that the person who wrote the text to the left is talking about the same issues of the sentence (in red above)?	Considering only the words highlighted in the text left, do you think he has the same meaning as the sentence (in red above) used as a subject?	Do you think the person who wrote the text on the left has an interest on the subject of prayer (prayer in red above)?
President Sarney advocates and calls of complaints One day after the speech of José Sarney on the crisis of ethics that reaches the Senate (chaired by him), President Luiz Inácio Lula da Silva, traveling in Asia, criticized the basis of complaints of irregularities. "They have no end, and then nothing happens," said Lula. Sarney, President of the Senate, used the rostrum of the House yesterday (16) to defend himself from accusations of using secret acts to appoint their relatives.	I definitely agree ()	Very possess ()	I have a lot ()
	I agree ()	Possess ()	I have ()
	I lightly agree ()	Little possess ()	I Have little ()
	I disagree ()	No possess ()	I do not have ()
	I definitely disagree ()	It has nothing to do ()	Never ()
	I can not comment ()	I can not opine ()	I can not opine ()

Figure 9. Part of the questionnaire applied.

This questionnaire was based on the QUIS (Questionnaire for User Interaction Satisfaction) purpose for laboratory researchers of the Human-Computer Interaction the University de Maryland (USA) [17].

The next section presents the results of the study.

Results

Figure 10 presents a graph showing the distribution of all posts of the three topics used in the search. Fifty-three percent are posts that were recovered with the support of cultural knowledge of the Project OMCS-Br.

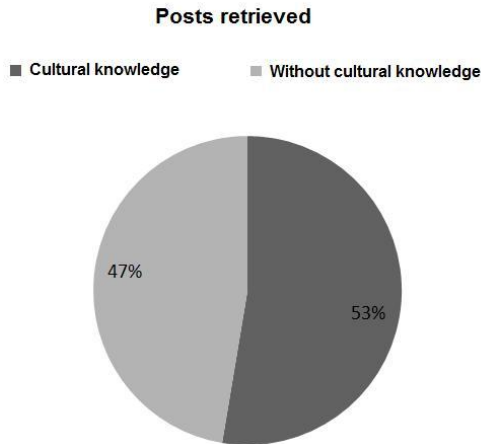


Figure 10. Graph showing the distribution of all posts of the three topics

For example, the posting: "... But Rio continues wonderful and is the only place that awakens me always want to come back and spend my holidays despite...". It was recovered due the support base of cultural knowledge, because prayer original used as subject "Rio de Janeiro continues beautiful."

In this case, "Rio de Janeiro" was replaced by cultural synonym "Rio", found in the search base as defined (Rio de Janeiro, Rio). Already "beautiful" has been replaced by "wonderful" found at the base as definedAs (beautiful, wonderful). The end of the process so as sentence found in the post, in which case it is "River still wonderful."

The Question 1 (second column of the table in Figure 9) aims verifying whether the method was able grouping people who are talking about the same topics. In the graph in Figure 11 shows the overall results to this question for the three subjects used in the search.

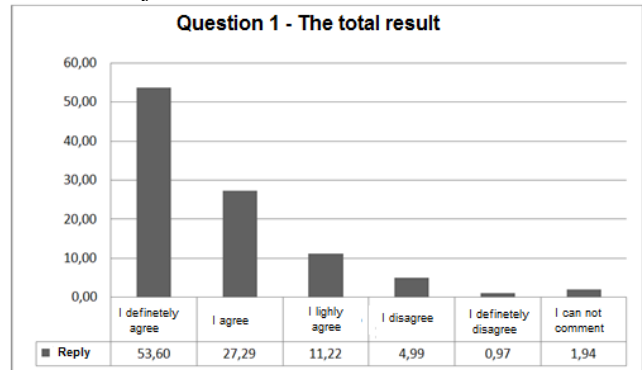


Figure 11 - Results of the search the peoples the talk about the same topics

Considering only the two alternatives (strongly agree and agree) as the most relevant to group people and say that they are talking about the subjects used in the search, the method achieved 80.89% success.

Now, if we consider the three alternatives, the method got 92.10% success. In the search, using the method available in Orkut, the possibilities are very limited. For example, using the sentence "Rio de Janeiro continues beautiful" in quotes, the best way to consider the subject, the search has virtually the same efficiency that the method proposed by this work.

The difference is that in search of Orkut are considered only identical texts, ie, sentences, such as "*Rio de Janeiro continues beautiful*," that is semantically similar to that used in the search, is not considered. In this case, the temporal variation of sentence is ignored, something that the method proposed by this work is not lost.

Also, posts that have sentences written differently, as for example, "*Rio still beautiful*," which is considered semantically similar to the sentences used in the search, the method available in Orkut ignores. This type of comparison, consider that people's culture, is the main point of the method proposed by this work.

In the graph of Figure 12 shows the comparison between the result of "use" and "no use" in search of cultural knowledge, or are considered only posts with prayers, such as "Rio de Janeiro continues beautiful" , "Rio de Janeiro continues beautiful," "Rio de Janeiro continues beautiful," etc.. and the result with the "use" of cultural knowledge in the search, namely the posts are considered as sentences, eg, "City beautiful remains beautiful," "Rio remains beautiful," etc.

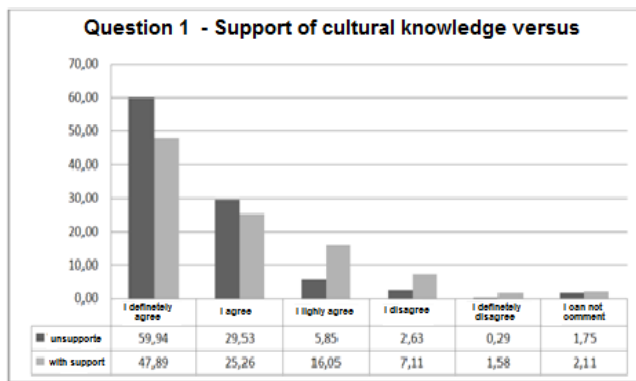


Figure 12 - Comparison between the result of "use" and "no use" in search of cultural knowledge

We realize that "use" of cultural knowledge takes a slight disadvantage in terms of efficiency in relation to the search. For if we consider only the first two alternatives (agree and strongly agree) as a success, "no use" is a hit, in the opinion of the participants of the case study of 89.47% against 73.15% of "use" of knowledge cultural.

One advantage that "use" of cultural knowledge has led in the case of this case is in relation to the number of posts retrieved, ie 380 against 342 with the "no use". That would not be possible using only textual comparisons as a search of Orkut does.

The result obtained with the "use" of cultural knowledge shows that the use of the base of OMCS-Br as a bank of synonyms cultural satisfactory result was obtained for this work. Therefore been grouped people who talk about the subject matter, but that are expressed differently, for example, people who have cultural differences.

Thus, culture is used here not to alienate people, but as an alternative to approach them.

Now consider the last two alternatives to Question 1 (disagree and strongly disagree) as the event of failure of the method, the "use" of cultural knowledge had 8.69% error, ie, identified posts where people were not talking about the subjects in question. But the "no use" of cultural knowledge had 2.93% error.

The sentence "*The beach is still beautiful*" is an example of a prayer, identified in a post, which brought the method to commit such errors. In this case the problem was identified on *mr* continue (*beach, beautiful*) who identified the post. It was generated from the ratio of Minsky DefinedAs (*Rio de Janeiro, beach*), that is, "Rio" was replaced by "*the beach*".

It is noticed that the *mr* is not clear that the matter is that search is over "*Rio de Janeiro continues beautiful*." Despite the "*Beach*" be defined as synonymous with "*Rio*" for a project collaborator OMCS-Br, when using this knowledge is distorted cultural semantics of prayer, causing it to express something completely undefined in the subject: "*Beach is still beautiful*." What is "*Beach*"?

Below is an example of a post that caused the error:

"Hi!

Ditto, ditto!

Pleeease someone here is the time to ride the scooter at nine was everything!?

When night had no lamp on the beach, the day had crab and everybody could get the yacht to swim?

Ahhhhh, nostalgia delicious!

Good to know that the beach is still beautiful and loved by many!

Kisses to all."

The error perceived in relation to "use" of cultural knowledge is very difficult controlling because the captured knowledge is very difficult controlling because the culturally the concept sought, as is the case of "Rio" is considered synonymous with "Beach", when applied in a larger context, as a post from a user, if this work could end up being meaningless.

The Question 2 (third column of Figure 9) aims to establish whether the *mr* used as a parameter in searches are faithfully represent the topics in question. Moreover, there is also the *mr* derived from the use of basic cultural knowledge representation and not lose the context of the subject. They are expected to observe whether the semantic textual comparisons as proposed by this work are being carried out successfully.

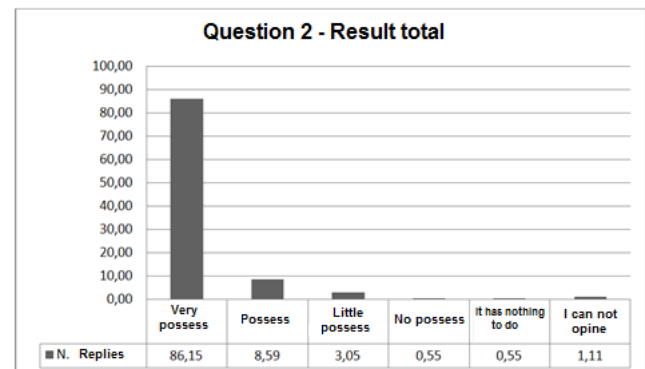


Figure 13 - Results in semantic textual comparisons

In general the graph shown in Figure 13 shows that the method achieved satisfactory results in semantic textual comparisons. For if we consider the three alternatives (It has very, owns and has little) of question 2 as a benchmark for success, the method got 97.78% accuracy.

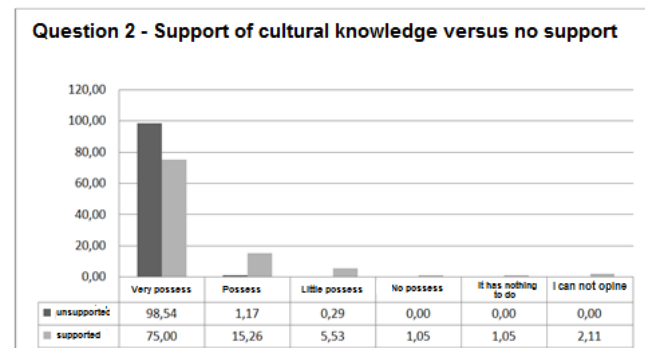


Figure 14 - Comparison between the use of cultural knowledge in search and non-use

In Figure 14 shows a comparison between the use of cultural knowledge in search and non-use. It is observed that whereas the first three alternatives (very own, owns and has little) of Question 2 was 95.79% success with the use of cultural knowledge. It is no use, was more effective, with 100% accuracy.

This difference between the two figures is explained by the fact itself perceived in the analysis of results for Question 1. That is, a mistake that is noticeable only when cultural knowledge extracted from the base is used in a larger context. For the analysis of the response to Question 1, there was an error of 8.69% in this case the error was 2.10%.

This difference between the two types of errors is a question because in the case study participants considered the post as a whole, including its context. In the case of Question 2, participants saw only the highlighted text snippet. Why we observed more errors in the analysis of Question 1.

The Question 3 (fourth column of the table in Figure 9) aims to verify if people grouped by the method could have interest in the subjects used in the search.

Here it is assumed that when a person talks about a subject, does not mean she has the same interest. For this reason the questionnaire was administered to participants.

Also, that the evaluations made by participants are based on the content of each post, and can vary in the interpretation of each.

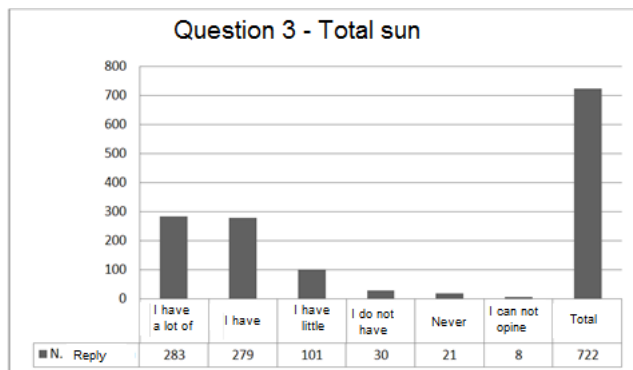


Figure 15 – Measures of the user’s interest demonstrated in the post

In the Figure 15 are presents the data of this evaluation. We observed that the study participants felt that in 77.83% of user’s posts have demonstrated interest in the topics concerned. This is satisfactory considering how the first two alternatives in question 3 (Lots and has).

CONCLUSION AND FUTURE WORK

This article focused on the challenge of giving some support to identify in virtual communities a certain group of people by identifying common interests by understanding their different cultural expressions and

attitudes and ‘translating’ their differences to see if they really share those interests. Through the approach described in this paper, SNSs users can identify the people who are talking about the same topics in different ways, so they can establish a relationship to talk about those topics, exchanging experiences, knowledge, solutions, among other things. This approach can be used in systems as Social Matching Systems, Management Systems or any tool that needs to improve its search engine considering cultural background. In this case the methodology was tested to search for people, but the same process can be used to identify educational materials or reports with the same content but with different names, influenced by the culture of the people.

It is intended to consider other Minsky’s relations to improve the cultural expansion of semantic relationships and develop the algorithm for better performance.

ACKNOWLEDGMENTS

We thank CNPq, FAPESP, CAPES and EMBRAPA for financially supporting this research. We also thank all the collaborators of the Open Mind Common Sense in Brazil Project who have been building the common sense knowledge base considered in this research.

REFERENCES

- Howard, B. 2008. Analyzing online social networks. *Commun. ACM* 51, 11 (Nov. 2008), 14-16. DOI= <http://doi.acm.org/10.1145/1400214.1400220>.
- Boyd, D. M. and Ellison, N. B. 2007. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), article 11. <http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html>.
- Chen, J., Geyer, W., Dugan, C., Muller, M., and Guy, I. 2009. Make new friends, but keep the old: recommending people on social networking sites. In *Proceedings of the 27th international Conference on Human Factors in Computing Systems* (Boston, MA, USA, April 04 - 09, 2009). CHI '09. ACM, New York, NY, 201-210. DOI= <http://doi.acm.org/10.1145/1518701.1518735>
- Hamasaki, M., Matsuo, Y., Ishida, K., Nakamura, Y., Nishimura, T. and Takeda, H. 2006. Community focused social network extraction. In: *Proceedings of the Semantic Web* (Beijing, China, September 03 - 07, 2006). ASWC 2006, Springer, Heidelberg, 155-161. <http://10.1007/11836025>.
- H. Kautz, B. Selman, and M. A. Shah. 1997. The Hidden Web. *AI Magazine* 18, 2, 27-36.
- Kautz, H., Selman, B., and Shah, M. 1997. Referral Web: combining social networks and collaborative filtering. *Commun. ACM* 40, 3 (Mar. 1997), 63-65. DOI= <http://doi.acm.org/10.1145/245108.245123>.

7. Tang, J., Zhang, J., Yao, L., and Li, J. 2008. Extraction and mining of an academic social network. In *Proceeding of the 17th international Conference on World Wide Web* (Beijing, China, April 21 - 25, 2008). WWW '08. ACM, New York, NY, 1193-1194. DOI=<http://doi.acm.org/10.1145/1367497.1367722>.
8. Spertus, E., Sahami, M., and Buyukkokten, O. 2005. Evaluating similarity measures: a large-scale study in the orkut social network. In *Proceedings of the Eleventh ACM SIGKDD international Conference on Knowledge Discovery in Data Mining* (Chicago, Illinois, USA, August 21 - 24, 2005). KDD '05. ACM, New York, NY, 678-684. DOI=<http://doi.acm.org/10.1145/1081870.1081956>.
9. Chen, W., Chu, J., Luan, J., Bai, H., Wang, Y., and Chang, E. Y. 2009. Collaborative filtering for orkut communities: discovery of user latent behavior. In *Proceedings of the 18th international Conference on World Wide Web* (Madrid, Spain, April 20 - 24, 2009). WWW '09. ACM, New York, NY, 681-690. DOI=<http://doi.acm.org/10.1145/1526709.1526801>.
10. Nunes, M. A., Cerri, S. A., and Blanc, N. 2008. Towards user psychological profile. In *Proceedings of the VIII Brazilian Symposium on Human Factors in Computing Systems* (Porto Alegre, RS, Brazil, October 21 - 24, 2008). ACM International Conference Proceeding Series, vol. 378. Sociedade Brasileira de Computação, Porto Alegre, Brazil, 196-203.
11. Silva M. A. R., and Silva J. C. A. 2009. Promoting Collaboration through a Culturally Contextualized Narrative Game. In *Proceedings of the 11th International Conference Enterprise Information Systems* (Milan, Italy, May 6-10, 2009). ICEIS 2009. Springer, Heidelberg, 870-881. DOI=http://10.1007/978-3-642-01347-8_72.
12. Anacleto, J. C., De Carvalho, A. F. P., Pereira, E. N., Ferreira, A. M., and Carlos, A. F. 2008. Machines with good sense: How can computers become capable of sensible reasoning?. In: Bramer, M. (Org.). *Artificial Intelligence in Theory and Practice II*. WCC 2008, Springer, Boston, 195-204.
13. Minsky, M. 1986. *The Society of Mind*. Simon and Schuster, New York.
14. Liu, H. and Singh, P. 2004. ConceptNet — A Practical Commonsense Reasoning Tool-Kit. *BT Technology Journal* 22, 4 (Oct. 2004), 211-226. DOI=<http://dx.doi.org/10.1023/B:BTTJ.0000047600.45421.6d>.
15. Bick, E. *The Parsing System "PALAVRAS": Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*. PhD thesis, Department of Linguistics. University of Aarhus, DK., 2000.
16. Indicador de Alfabetismo Funcional – *INAF BRASIL 2009* (2009). São Paulo: Instituto Paulo Montenegro. Disponível em: <<http://www.acaoeducativa.org/images/stories/pdfs/inaf2009.pdf>>.
17. CHIN, J.P.; DIEHL, V. A.; NORMAN, K. L. "Development of an instrument measuring user satisfaction of the human-computer interface". In *Proceedings CHI'98*, 1998