Identifying people who are talking about the same topic in social networks even having a different cultural background

Gilberto Astolfi 1

Vanessa Maia Aguiar de Magalhães 1

Marcos Alexandre Rose Silva 1

Junia Coutinho Anacleto 1

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 paper 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.

1 Federal University of São Carlos, UFSCar, Computing Department
`{gilberto_astolfi,vanessa_magalhaes,marcos_silva,junia}@dc.ufscar.b`
r

1   Introduction

Nowadays, 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 paper 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. [2]. 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 "Brazilian beach" 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 "wonderful beach" 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., "Brazilian beach" and "wonderful beach" 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 Brazil, 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 [11], 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 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 Brazil, which is the source of culture used in this work. Section 4 presents the approach proposed in this work. Finally, Section 5 shows an experiment performed with final users and Section 6 there are conclusion and future work.

2   Related works

Literature provides several works that intend to 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 whole participants in congress have common interest. 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. [17] developed a system called POLYPHONET that from a user name builds a social network. Tang et. al. [7] 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] consider the possibility of two people share the same bag of words when they type profile information, photos comments, etc. Finally, based on theories related to psychology, the work presented by Nunes [8] proposes to model, formalize and store a users' psychological profile. This process compares some profile information among the users in order to indentify similarities.

The process presented in this paper differs from the others related here, because we have not omitted the culture of the people in the search process. Considering the cultural knowledge database it is possible to find people who express differently, but still about the same issue, increasing the likelihood individuals, identifying around similar issue.

3   Cultural knowledge database

The cultural knowledge is obtained from a project called Open Mind Common Sense Project Brazil (OMCS-Br) [9]. 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(a). For instance, template: Rio de Janeiro can be called as Marvelous city.

Figure 1. (a) Example of template. (b) Illustration of the Cultural Knowledge Base.

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 base. 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 [11]. 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 [10].

The process separates the template in two concepts and joins them through Minsky's relations [12]. For instance, the template relation "*Rio de Janeiro*" can be called as "*Marvelous city*" (see Figure 1 (a)) is a Minsky's relation called *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 1 (b)). 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 Minsky's relation in this case is *IsA*. (see Figure 1 (b)). Others Minsky's relations are also possible, such as: *PropertyOf, MotivationOf, UsedFor, CapableOf*, etc [12].

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 about a specific issue, in these examples about "*Rio de Janeiro*". Finally, the whole cultural information stored as semantic network we called as Cultural Knowledge Base (see Figure 1 (b)).

4   Considering cultural knowledge to identify similar issue

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 issue, 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 "*Beautiful city keep beautiful*." When we read these two sentences, we can identify that both represent a consensus of two people over a certain issue, that is, they said the same things differently.

The use of common search engines, e.g. Google API or the Social Matching Systems 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 Matching 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. We developed an algorithm in order to realize this "cultural search" in SNSs. This algorithm works taking into consideration some steps described below.

4.1 Extracting semantic of a sentence

First, we need a sentence typed by user, for instance a commentary that a user typed on an image in the SNS Flickr (www.flickr.com), or a post on Twitter (www.twitter.com), etc. One example of a sentence is: "*Rio de Janeiro continues beautiful*".

The algorithm "read" this sentence and through of a syntactic parser for Portuguese [13] called PALAVRAS, 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*). PALAVRAS also normalizes each word in this structure, i.e., nouns and adjectives of the sentence need are in singular and verbs in infinitive form, etc. This normalization is useful 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 et. al. [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 meta-relation *mr = 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). It is important to observe that this meta-relation structure is similar to structure used in Minsky'relation, which Minsk's relation in this case is the verb.

4.2 Cultural translation

It is possible to build many *mr* through one *mr* taking account the same issue. For instance, *mr = continue (Rio de Janeiro, beautiful)*. The cultural translation can generate new others *mr*, for example, $mr_1 = continue$ *(Rio de Janeiro, perfect)*, $mr_2 = keep$ *(Rio de Janeiro, beautiful)* and $mr_3 = continue$ *(Marvelous city, beautiful)*. Observe that these three *mr* are identical in issue, ranging only linguistically, showing the main purpose of cultural translation. 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. This cultural translation taking into consideration two Minsk's relation from cultural knowledge database, such as: *IsA* e *DefinedAs* .

IsA is considered a weak relation [12] 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)*, i.e., the whole features of "*city*" are part of features of "*Rio de Janeiro*". The *IsA* role used in cultural translation is: If *X* is a specialization of *Y*, then feature $Y \subset$ feature *X*. Therefore, the concept represented by *Y* may be used to represent *X*, but not vice-versa. For instance, if there is the knowledge *IsA (Rio de Janeiro, city)*, it is possible to refer to "*Rio de Janeiro*" using the word "*city*", but you cannot refer to any "*city*" using the word "*Rio de Janeiro*".

DefinedAs is a type of relation that uses synonyms to represent the meaning of something [12]. 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)*, the whole characteristics of "*Beautiful*" are like characteristics of "*Perfect*". The *DefinedAs* role used in cultural translation is: If *X* is synonym of *Y*, then feature $X \Leftrightarrow$ feature *Y* e vice-versa. Therefore, the concept represented by *X* may be used to represent *Y* and vice versa. For instance, if there is the knowledge *DefinedAs(Beautiful, Perfect)* may refer to "*Beautiful*" using the word "*Perfect*" and vice versa.

We are going to show how to expand culturally a *mr* for a set $\beta = \{mr_1...mr_n\}$ where each mr_n owns the same meaning that the *mr* base, i.e., without losing the issue.

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 2 (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 2 (b)).
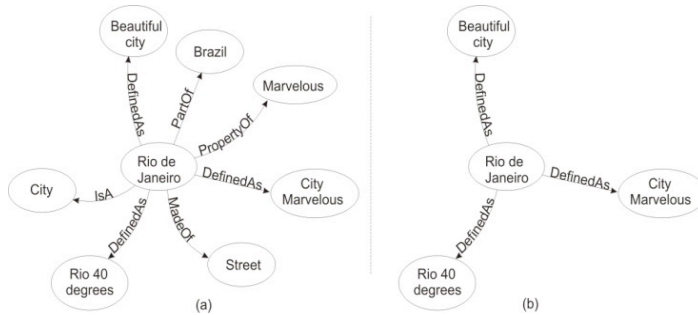

Figure 2. Example of search about cultural knowledge base.

To translate a *mr* culturally is used the second form of search, using specific relations such as *IsA* and *DefinedAs*. Parameters used in the search are the components of a m_r (*verb*, *subject* and *object*). For instance, *DefinedAs (X, (verb, subject* or *object))*. The relation *IsA* is more limited between the two, when it is used *X* is always fixed, for example, *IsA ((verb, subject* or *object), Y)*, because as previously defined, we can use only the word in *Y* to represent the word in *X*, but not vice-versa. 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 OMCS-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*
IsA (beautiful, Y)	*There was no return*
DefinedAs(X, Rio de Janeiro)	*Wonderful city*
DefinedAs(Rio de Janeiro, Y)	*Marvelous city*
DefinedAs(X, beautiful)	*beautiful*
DefinedAs(beautiful, Y)	*gorgeous*
DefinedAs(X, continue)	*Keep*
DefinedAs(continue, Y)	*There was no return*

The results obtained, in the search shown in Table 1, are used to translate 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 *mr₁ = continue (Marvelous city, beautiful)*, because in the search *DefinedAs ((subject, object or verb), Y) Y*

can be used to represent *X*. It is shown in Figure 3 an example of the cultural expansion, which black words are from sentence typed by user and, green words are from cultural knowledge base.
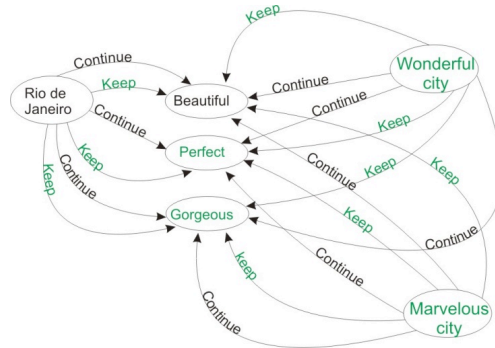

Figure 3. Example of the cultural expansion.

4.3   Search for people in SNSs

Finally, the result of this cultural expansion is used to search for users in SNSs who talk about the same issue, but express themselves the different ways or no. In the search, subject and object from each *mr* are used in order to find co occurrence among them in a post typed by a user. If there is this co occurrence, the whole text between subject and object is obtained. For instance, the text obtained from this post *"…Yes!* **Wonderful city continues beautiful***! Even many want on the contrary…"* is "Wonderful city continues beautiful!", because *"Wonderful city"* represent the subject and *"beautiful"* the object.

The whole steps, described since Section 4.1, are applied in this text obtained in order to generate a cultural network, as shown in Figure 3. After, our algorithm compares this cultural network in order to identify similarities among it and people´s posts in Social Network Services (SNSs). When words from one are similar from other, we can suppose that these users are talking about the same issue.

5   Feasibility study

We conducted a feasibility study with some people, in July 2010, in order to observe the results obtained from our algorithm. Nineteen full literate people participated of this study. Only full literate people were considered in the study once it was expected from them to be able to evaluate whether a post from Orkut Communities, selected by the algorithm, is related to the same issue used for the search, even if it is written differently. According to INSTITUTO PAULO MONTENEGRO [14], literate people have the ability to read and interpret a text.

This study took into consideration **three sentences**:
(1) "*Does Rio de Janeiro continue beautiful?*".

(2) "*Michael Jackson dies at age 50*".

(3) "*Lula defends Jose Sarney and says that complaints are endless.*"

The questionnaire shown in Figure 4 was given to each person who participated of study. It was based on the QUIS (Questionnaire for User Interaction Satisfaction) purpose for laboratory researchers of the Human-Computer Interaction the University de Maryland (USA) [15]. To answer it, people took into consideration "*the text to left*" - the posts typed by user in SNSs (Figure 4 - area A); the sentence utilized to find the posts (Figure 4 - area B) and; four questions created to allow the participants give their opinions (Figure 4 - area C).

Sentence used as issue: *Rio de Janeiro continues beautiful* B				
...But the **Rio** D **continues wonderful** and is the only place that always causes me to want to come back and spend my holidays... A	You agree that the person who wrote the text to the left is talking about the same issue of the sentence? C	Considering only the highlighted words in the text left, do you think it has the same meaning as the sentence used as an issue?	Do you think the person who wrote the text left has an interest on the issue?	If the issue was of your interest, and if the person that wrote the text was user of the Social Networking Sites, would you have interest to make contact with her?
	( ) I definitely agree ( ) I agree ( ) I lightly agree ( ) I disagree ( ) I definitely disagree ( ) I can no comments	( ) very possess ( ) possess ( ) little possess ( ) no possess ( ) it has nothing to do ( ) I cannot opine	( ) he has a lot ( ) he has ( ) he has little ( ) he has not ( ) never ( ) I cannot opine	( ) I have a lot ( ) I have ( ) I have little ( ) I do not have ( ) never ( ) I cannot opine

Figure 4. Adapted QUIS.

The main objective of this study was to collect people´s opinion about the result of the algorithm taking into consideration four topics:

- First – identify people who talk about the same issue in Social Networking Services;

- Second – observe if cultural expansion is a good way to obtain different words related to the same issue;

- Third – find people, in Social Networking Services, who are talking about the same issue in different ways;

- Fourth – introduce people to each other when they are talking about the same issue.

It was created one question according each topic. First question was: "*You agree that the person who wrote the **text to the left*** (Figure 4 - area A) *is talking about the same **issue*** (Figure 4 - area B) *of the sentence?*". It is important to describe that it were obtained many posts according the tree sentences (1) (2) (3).

In the graph, presents in Figure 5 (a), shows the results of this first question taking into consideration three sentences.

Considering only the two alternatives (*"I definitely agree"* and *"I agree"*, Figure 5 (a)) as the most relevant to the study, it is possible to observe that 80.89% of people, who

answered the questionnaire, thing that algorithm works well, because it identified users who wrote about the same issue.
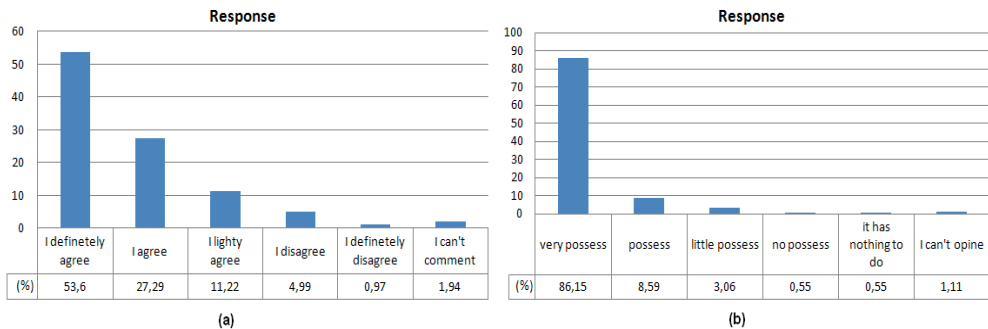

(%)	I definetely agree	I agree	I lighty agree	I disagree	I definetely disagree	I can't comment
	53,6	27,29	11,22	4,99	0,97	1,94

(a)

(%)	very possess	possess	little possess	no possess	it has nothing to do	I can't opine
	86,15	8,59	3,06	0,55	0,55	1,11

(b)

Figure 5. (a) Results obtained from the first question. (b) Results obtained from the second question.

Using the search from Orkut, the possibilities are very limited. Because, in it just identical texts are considered, i.e., sentences, such as "*Rio de Janeiro will continue beautiful*" that is semantically similar to that used in the search "*Rio de Janeiro continues beautiful*", is not considered. In this case, the change of the verb is ignored, something that the method proposed by this work is not lost. Another example, "Rio keep wonderful," which is semantically similar to the sentences used in the search, the method available in Orkut also ignores. Therefore, this search does not consider the user´s culture.

Considering the last two alternatives ("*I disagree*" and "*I definitely disagree*" Figure 5 (a)) as the event of failure of the method, there were 5.96%. The sentence "*Beach continue beautiful*" is an example of a sentence, identified in a post, which there was a failure. In this case the problem was identified on *mr continue* (*beach, beautiful*) who identified the post. It was generated from Minsky's relation the structure *DefinedAs (beach, beautiful)*, i.e., "*Rio de Janeiro*" was replaced by "*beach*".

In this case, it is very difficult to identify that this beach is localized in "Rio de Janeiro". Because of that, a post was obtained by our algorithm "*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, wonderful nostalgia! Good to know that the* **beach continue beautiful** *and loved by many! Kisses to all*."

Second question was "*Considering only the* **highlighted words** (Figure 4 - area D) *in the* **text to left**. *Do you think it has the same meaning as the* **sentence used as an issue** (Figure 4 - area B)?*".

This question intends to allow people to compare these two sentences in order to collect their opinion about the meaningful similarity among them. For instance, people needed to say if there is the same meaningful between the post "*…Yes!* **Wonderful city**

*continues beautiful! Even many want on the contrary…"*and the sentence "*Rio de Janeiro continues beautiful"*. This question is interesting in order to observe the quality of the results from the algorithm, i.e., if there is the same semantic textual, such as same context, issue and etc., in both post and sentence.

In the graph, presents in Figure 5 (b), is showed the overall results of this second question for the three sentences used in the search. Observing Figure 5 (b), it is possible to observe that the algorithm achieved satisfactory results in semantic textual comparisons. Considering two alternatives ("very posses" and "posses") of question 2 as a benchmark for success, the method got 94.74% accuracy.

Third question was: "*Do you think the person, who wrote the **text to left** (Figure 4 - area A), has an interest on the **issue*** from the sentence (Figure 4 - area B)?".
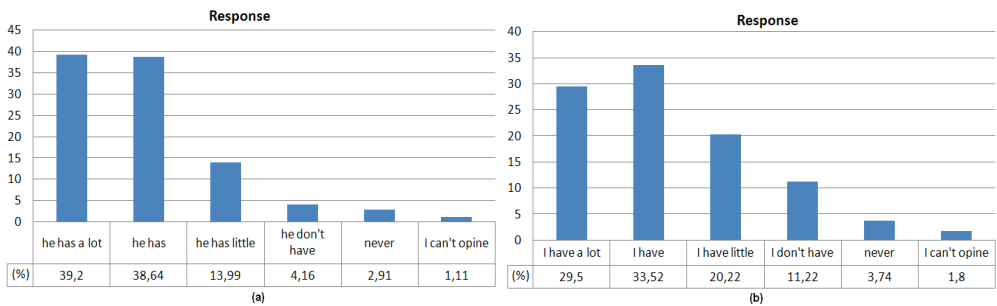

Figure 6. (a) Measures related to user's interest demonstrated in the post. (b) Simulation of Social matching.

This question intends to collect some information in order to observe the potential of the algorithm. In this case, its capacity to identify users who comment about the same issue and, to identify users who have interest about it. This question is important, because we believe that when a person comments about a certain issue; he cannot have interest about it. One possibility to notice this interest, it is reading the post. Nevertheless, it is important to say that people, who participated of study, notice on the content of each post, and this observation can vary in the interpretation of each.

The results of this evaluation is presented in the Figure 6 (a). According people´s opinion 77.84% ("*he has a lot*" and "*he has*") of the posts are related to the sentence utilized to find them. Finally, the last and fourth question was: "If you were a user in Social Network Service and had some interest about the issue of the sentence. Would you like to talk to user who comments the post (Figure 4 - area A)? Why?

Through this question we intend to collect some information about the possibility to connect users when the algorithm identifies similarities between them [16]. There are some results about this question in Figure 8 taking into consideration three sentences used in the study. Considering two alternatives ("*I have a lot*" and "*I have*") (Figure 6 (b)), it was

possible to observe that 63.02% of people had interest to connect with users who wrote the posts obtained by algorithm.

In this question, people could to explain about "Why?" they want to connect with the user or no. Some negative answers with the people´s profile are presented in Table 2.

Table 2. Comments about "Why" would you like to talk to user who comments the post?

People´s Profile	Answer	Comment
30-45 years old; doesn´t use SNSs	"Never"	*"He is very annoying and ignorant"*
18-29 years old; uses SNSs (Orkut).	"I don't have"	*"He is very dull. Person speaks a lot and comments something many times."*
18-29 years old; uses SNSs (Orkut).	"I don't have"	*"very rude"*

These results, presented in Figure 7 (a), shown that users who comment about the some issue (*"I definitely agree"* and *"I agree"*, 80,89%), or users who have interest about the some issue (*"he has a lot"* and *"he has"*, 77.84%), they can not want to connect themselves ("*I have a lot*" and "*I have*", 63.02%).

Performance of the Method for the three Issue

	Alternative 1	Alternative 2	Alternative 3	Alternative 4	Alternative 5	Alternative 6
Same issue	53,6	27,29	11,22	4,99	0,97	1,94
Interest	39,9	38,64	13,99	4,16	2,91	1,1
Matching social	29,5	33,52	20,22	11,22	3,74	1,8

(a)

Semantic search *versus* Cultural translation

Semantic search 47%

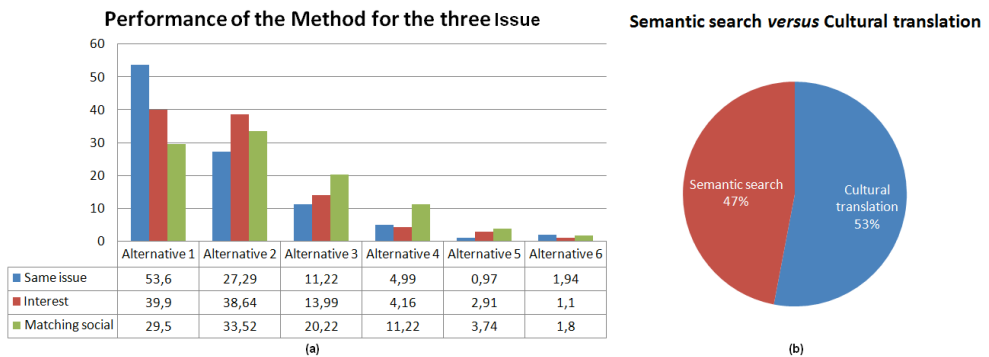Cultural translation 53%

(b)

Figure 7. (a) Comparison among identification of users that talk about the same issue, user's interest and social matching. (b) Total of the posts obtained by algorithm taking into consideration the three sentences and the use of cultural translation or just semantic search i.e., similarities between original words in the sentences and posts.

Therefore, it was possible to observe that there are others factors which are important to take into consideration in order to connect users (see Table 2), i.e., others factors than just text similarities as Chen et. al. [3]'s project or culturally similarities as the algorithm presented in this paper. Nevertheless, 60.02% of people had interest to connect with users. We believe that it is a good result when we work with a subjective issue.

An interesting result is related to support of cultural knowledge database in order to search posts. The result is shown in Figure 7 (b).

Considering Figure 7 (b), it is possible to observe that 53% of the posts (related to three sentences) obtained by algorithm took into consideration information from cultural knowledge database, i.e., it was necessary the cultural translation to find them. For example, "**Rio continues wonderful** and it is a particular city that I want to go back and stay...". In this case, there were cultural translation "*Rio de Janeiro*" to "*Rio*" and "beautiful" to "wonderful". This information *definedAs (Rio de Janeiro, Rio)* and *definedAs (beautiful, wonderful)* are presented in cultural knowledge database.

This result shows the potential of the algorithm, because it identified users who wrote about the same issue in different ways. Therefore, this cultural translation allows finding users who have different background, culture, etc. In this case, the cultural difference can be a factor to stimulate people to connect themselves, because exchange experiences, culture, knowledge can be interesting in a Social Network Service (SNS).

6   Conclusion and further works

This paper 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. We intend to consider other Minsky's relations to improve the cultural expansion of semantic relationships and develop the algorithm for better performance.

7   References

1. HOWARD, B., 2008, "Analyzing online social networks". In: *Communications of ACM,* v.51, n. 11, 14-16.
2. BOYD, D. M.; ELLISON, N. B., 2007, "Social network sites: Definition, history, and scholarship". In: *Journal of Computer-Mediated Communication*, v. 13, n. 1, article 11.

3. CHEN, J.; GEYER, W.; DUGAN, C.; MULLER, M.; 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 (CHI '09)*, Boston, MA, USA, pp. 201-210.

4. HAMASAKI, M.; MATSUO, Y.; ISHIDA, K.; NAKAMURA, Y.; NISHIMURA, T.; TAKEDA, H., 2006, "Community focused social network extraction". In: *Proceedings of the Semantic Web (ASWC 2006),* Beijing, China, pp. 155-161.

5. KAUTZ H.; SELMAN, B.; SHAH, M. A., 1997. "The Hidden Web". In: *AI Magazine,* v. 18, n. 2, pp. 27-36.

6. KAUTZ, H.; SELMAN, B.; SHAH, M. A., 1997, "Referral Web: combining social networks and collaborative filtering". *Communications of the ACM,* v. 40, n. 3, pp. 63-65.

7. TANG, J.; ZHANG, J.; YAO, L.; LI, J., 2008, "Extraction and mining of an academic social network". In: *Proceeding of the 17th International Conference on World Wide Web (*WWW '08*),* Beijing, China, pp. 1193-1194.

8. NUNES, M. A.; CERRI, S. A.; BLANC, N., 2008, "Towards user psychological profile". In: *Proceedings of the 8th Brazilian Symposium on Human Factors in Computing Systems,* Porto Alegre, RS, Brazil, pp. 196-203.

9. SILVA M. A. R.; SILVA, J. C. A., 2009, "Promoting Collaboration through a Culturally Contextualized Narrative Game". In: *Proceedings of the 11th International Conference Enterprise Information Systems (ICEIS 2009),* Milan, Italy, pp. 870-881.

10. ANACLETO, J. C.; DE CARVALHO, A. F. P.; PEREIRA, E. N.; FERREIRA, A. M.; CARLOS, A. F., 2008, "Machines with good sense: How can computers become capable of sensible reasonig?". In: *Bramer, M., Artificial Intelligence in Theory and Practice II (WCC 2008)*, Boston, 2008, v. 1, pp. 195-204.

11. MINSKY, M., 1986. *The Society of Mind. 1th ed.*, Simon and Schuster.

12. LIU, H.; SINGH, P., 2004, "ConceptNet — A Practical Commonsense Reasoning Tool-Kit". In: *BT Technology Journal,* v. 22, n. 4, pp. 211-226.

13. BICK, E., "The Parsing System "PALAVRAS": Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*".* 2000. PhD thesis - Department of Linguistcs, University of Arhus, DK., 2000.

14. INSTITUTO PAULO MONTENEGRO, 2009, "*Indicador de Alfabetismo Funcional – INAF BRASIL 2009*". Disponível em <http://www.acaoeducativa.org/images/stories/pdfs/inaf2009.pdf>. Acesso em: 09/04/2010.

15. CHIN, J.P.; DIEHL, V. A.; NORMAN, K. L., 1998, "Development of an instrument measuring user satisfaction of the human-computer interface". In: *16th International Conference on Human Factors in Computing Systems (CHI '98).*

16. TERVEEN, L.; MCDONALD, D. W., 2005, "Social matching: A framework and research agenda". *ACM Trans. Comput.-Hum. Interact*, v. 12, n. 3, pp. 401-434.

17. MATSUO, Y.; HAMASAKI, M.; NAKAMURA, Y.; NISHIMURA, T.; HASIDA, K.; TAKEDA, H.; MORI, J.; BOLLEGALA, D.; ISHIZUKA, M., 2006. "Spinning multiple social networks for semantic web". In: *Proceedings of the 21th National Conference on Artificial intelligence (AAAI),* Boston, Massachusetts, USA, 2006. pp. 1381-1386.