

USING CULTURAL DIFFERENCES TO JOIN PEOPLE WITH COMMON INTERESTS OR PROBLEMS IN ENTERPRISE ENVIRONMENT

Gilberto Astolfi, Vanessa M. A. de Magalhães, Marcos A. R. Silva and Junia C. Anacleto
Computing Department, Federal University of São Carlos, Washington Luis Road, São Carlos, Brazil

Keywords: Culture, Enterprise environment, Social networking service, Collaborative work.

Abstract: This paper describes a methodology to identify people with common subjects but different cultures. Each person has a culture and it influences in his way to express himself or to write, etc. Because of that, nowadays, on the Web, it is very difficult a person who writes 'validating enterprise system' finds another who writes 'attesting B2B solution' because there are different words. On the other hand, there is a common subject that search engines can not identify, because it is necessary taking into consideration the cultural knowledge and experience from each person. In this context, the methodology presented here intends to consider this cultural information in order to identify when two or more people are talking about the same subject.

1 INTRODUCTION

Nowadays, the most of the people are concerned about having more experience e knowledge in order to improve their abilities in work, study and life. There are many ways to achieve these objectives, such as: studying in a university, attending a specific course, working in some companies, among others.

On the other hand, there is a particular way, called social networking service (SNSs) (Boyd and Ellison, 2007), which many people are using to know others people, to talk about many subjects, to talk with their friend, coworkers, etc. Because of this, this service can be a useful tool to help people to know others, in the some or in different areas, to talk about any situation, technology, problem, etc., i.e., people can learn and improve their experience with others through this service.

According to Hovard (Howard, 2008), the use of this type of service is very common between people, for instance, Orkut (www.orkut.com), LinkedIn (www.linkedin.com), Facebook (www.facebook.com), Hi5 (www.hi5.com), among others, have been billions of accesses. There are two of them with a bigger number of accesses than others in the world: Facebook and Orkut, mainly in Brazil.

It is a well-known fact that most of people whom surface on the web have a good chance of participating in an on-line social network. Therefore,

these social networks are improving their services to connect more and more people. For example, Orkut recommends people to connect with based on a "friend-of-a-friend" approach. Facebook and Sonico (www.sonico.com) added import wizards, which allow importing contacts from email and instant messaging clients.

Through this new world that is connecting people, it is also possible to use it in order to allow people to learn with each other, i.e., allow people with common subjects or problems, etc., meet themselves on the web and talk about these themes.

Some users usually use some services, such as: forums or chats, to have any contact with others, but there is a problem. It important to describe that there is not problem in these services; the problem is in how people find these forums or chats. For example, a person who wants to know more about Enterprise System can type these words in a search field to find many people, forums and chats that like to talk about it, but it will more difficult to find any people that like also it, but they write as B2B Solution.

In this case, it is necessary to observe that people with common interests can know or write the some thing in different ways, i.e., Enterprise System and B2B Solution are common themes, but the social networks do not consider this different to do the search, in other words, they do not consider people culture.

Depends on culture, country, state, region, among

others, the people can express themselves in different ways, but with the same objective. A person can search about evaluation and just find people that talk about it, but it is also interesting to find people that talk about test.

In this context, we propose, in this paper, a methodology that allows people to search and to find people with common interests in different culture, because this difference can be a good way to allow people exchange experience, knowledge, etc. i.e., the cultural difference can be used to join people not to distance them.

There are some projects which intend to connect people but they do not consider the cultural difference, i.e., if a person writes in a different way than other, they do not know to each other easily. Some projects are described below.

2 RELATED WORKS

There are many research works in literature whose goals are to identify group of people and recommend them to social networks formation. These studies have diverse way, for example, to propose social networks to people, to suggest communities to people in social networks, etc.

Some studies are focused on applications of algorithms designed to identify and join people related to a specific professional community, such as researchers. In (Tang et al., 2008) is proposed the formation of social networks to participate in a scientific congress, considering that the participants have some common research interests. The authors use data mining techniques to search information about people on the web and relate to each other, after which participants interfere with this network to refine it excluding people they do not know.

Kautz et. al. (Kautz et al., 1997b) use data mining techniques that consider the names of individuals co-occurring in web pages, in the publication of articles, on network news, and charts (university departments) are probably potential training social networks. In addition, Kautz et. al. (Kautz et al., 1997a) taking into consideration the same data source to propose a system to enhance searching for people by combining collaborative filtering and social networks. Through this same strategy, related to data mining, Matsuo et. al. (Matsuo et al., 2006) developed a system called POLYPHONET, which to measure the co-occurrence of names on web sites and, the social networking is built with the names more co-occurrences. In Tang et. al. (Tang et al., 2008) also use the web to search data that can identify the expertise of people to relate

them according to this information.

The authors in (Spertus et al., 2005) also experimented algorithms in order to propose communities to users on the Orkut. Communities are suggested from a communities base that certain user is a member. The similarity between communities is calculated overlapping them, e.g., based on the number of users in each community and the number of common users between them, so if two communities have a large number of common users they are considered similar. Using the same reasoning (Chen et al., 2009b) evolved two types of algorithms to recommend customized communities for users in Orkut.

The work presented in (Chen et al., 2009a) adopts techniques that have proved effective to recommend people for social networking. The authors improved four types of algorithms that were applied on a social networking site. The most important from our perspective is the one that stress the possibility of two people to post content related to the same subject and, if it happens, those people have great chance to be similar; this is measured by comparing a set of words posted by each user, extracted from their profile's information, comments on photos, shared lists, etc.

As can be observed, the studies mentioned here do not take into account the cultural aspects of people in searching for similarity to recommend social networking. The most is done is to consider some words or specialty that people share among them. Our work differs in that point, because we believe that taking into consideration the culture we can find more people with same interests in different cultures, i.e., country region, among others.

3 POTENTIAL OF CULTURAL

Our methodology, described here, considers the use the semantic context to verify similarities among people taking into consideration a subject and context. We assume that if two people have the same consensus related to a certain subject and context, they have good chances to share the same interests and consequently be similar to each other in that context.

In any country in the world there is a diversity of vocabularies. Thus, following the reasoning set, we hypothesize that the user uses in social interaction (mainly in SNSs) their natural vocabularies when they need to write a text. For example, in order to exchange messages. Therefore, taking into consideration this reality, we believe that there are many users, in SNSs, with same consensus or interests related to a subject and a context, but they can express themselves in a different way. For example, see this both sentences

from two different users: "Usability Test validates Enterprise Systems" and "Evaluation Usability attests B2B Solutions". When we read these both sentence, we can identify that both represent a consensus related to a subject and context, i.e., they say something in different way, but the computer, just trough these words, can not process these sentences and identity that they are similar using comparison words (Chen et al., 2009a). In this context, this paper presents a methodology that allows computer identify and process cultural information in order to identify familiar sentences wrote in different way by people because their culture, knowledge, etc. This methodology taking into consideration a cultural knowledge base.

3.1 Cultural Knowledge Base

The cultural knowledge base is obtained from a project called Open Mind Common Sense Project Brazil (OMCS-Br) (Silva and Anacleto, 2009). OMCS-Br project has been collected culture of a general public through a web site (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: Enterprise System can be called as B2B Solutions.

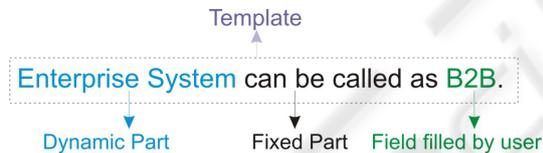


Figure 1: Template example.

Templates are simple grammatical structures. They have fix and dynamic parts. Dynamic parts 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 to be filled by users considering their everyday experiences, knowledge and culture. Finally, the fixed part is strategically defined by the project coordinators to consider the theory defined by Marvin Minsky (Minsky, 1988) and it keeps rich semantic relation (Liu and Singh, 2004) that represent the common sense of different people. For instance, the template relation "Enterprise System can be called as B2B Solutions" (Figure 1) is *DefinedAs*. Because the user typed that "Enterprise System can be defined as B2B Solutions". This template is stored (*DefinedAs 'Enterprise System' 'B2B Solutions'*), see Figure 2.

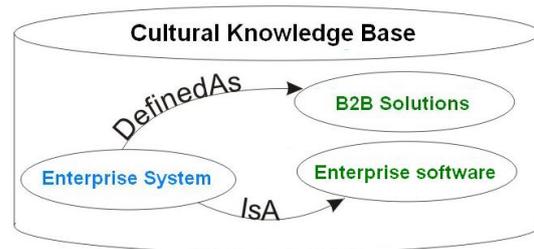


Figure 2: Simple example of semantic network.

There are many templates to collect cultural information; another example is "Enterprise System IsA Enterprise software". The template relation in this case is *IsA*. (see Figure 2). Others relations are also possible, such as: *PropertyOf*, *MotivationOf*, *Used-For*, *CapableOf*, etc (Liu and Singh, 2004). 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 subject, in these examples about Enterprise System. Finally, the whole cultural information stored as semantic network we called as Cultural Knowledge Base (see Figure 2).

Our methodology taking into consideration relations: *DefinedAs* and *IsA*, which represents synonymous and specialization respectively:

- $IsA(concept1, concept2)$, where *concept1* is a specialization of *concept2*;
- $DefinedAs(concept1, concept2)$, where *concept1* and *concept2* are synonymous.

According to these data we can defined that:

- If *concept1* is a specialization of *concept2*, then $feature(concept1)$ contains $feature(concept2)$;
- If *concept1* is synonymous of *concept2*, then $feature(concept1)$ is equal $feature(concept2)$.

In other words, the concepts connect with one of these relations are similar, i.e., strongly connect.

4 HOW THE METHODOLOGY WORKS?

First the methodology needs a sentence typed by user. The user can be typed this sentence in forums, chats, search field, etc. For example: "Usability Test validates Enterprise Systems" The computer "read" this sentence and through of a parser, called PALAVRAS a syntactic parser for Portuguese (Bick, 2000), identifies the grammar structure, i.e., if it has a subject,

verb and object. It is possible to observe that in this example, Figure 3, has a subject (Usability Test), verb (Validates) and object (Enterprise Systems).

The next step is Normalization (Figure 3) 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.

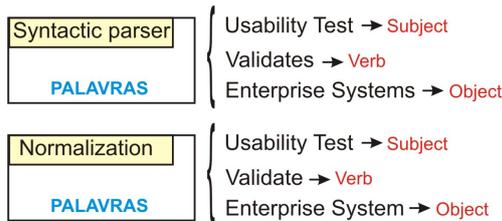


Figure 3: Normalization by PALAVRAS.

The next step is to create a semantic network with this sentence. There are in this semantic network the subject and object connected by verb. For instance, trough the sentence: subject (Usability Test), verb (Validate) and object (Enterprise System) was created: *validate(Usability Test, Enterprise System)* (see Figure 4).



Figure 4: Semantic Network with user's sentence.

It is important to write that the Normalization allows creating the same semantic network with different sentences typed by many users, as Table 1. This table shows four different sentences typed by four different users and just one semantic network as result.

Other projects as Chen (Chen et al., 2009a), described in related works, do not do this process, because of this, the sentences (*Usability Test, validates, Enterprise System*) and (*Usability Test, validated, Enterprise System*) will be considered as different sentences.

Our methodology through the Semantic Network does some searches in the Cultural Knowledge Base in order to expand it, as Figure 5.

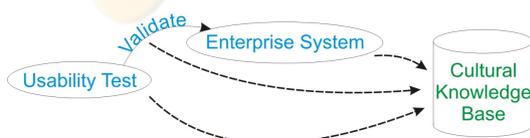


Figure 5: Searching in the Cultural Knowledge Base.

Table 1: Sentences from different users and their Semantic Network.

Users	Sentences	Semantic Network
user 1	Usability Test validates Enterprise System	validate(Usability Test, Enterprise System)
user 2	Usability Test validated Enterprise System	validate(Usability Test, Enterprise System)
user 3	Usability Test validating Enterprise System	validate(Usability Test, Enterprise System)
user 4	Usability Test has validated Enterprise System	validate(Usability Test, Enterprise System)

The main objective of doing this search, in this base, is to identify others cultural concepts related with concepts in Semantic Network. For instance, it is possible to identify what concepts are connected with Enterprise System, taking into consideration two Minskys relation: *DefinedAs* and *IsA*. In this case, the results were: B2B Solution, Enterprise software, etc. Therefore, there are other people that know Enterprise System, but they call it with different names, such as: B2B Solution, Enterprise software. This difference can be influenced by culture, experience, knowledge, etc., but, there is this information stored in Cultural Knowledge Base. There is the some process to identify cultural concepts related to each concept in Semantic Network, because of that, the next steps are to search with the concepts: Usability test and Validate, as show in Figure 6.

The Figure 6 shows the user's Semantic Network. It is possible to search in any forums, chats, etc., who are people talk about the same issue typed by user in the sentence "Usability Test validates Enterprise Systems". Because of that, different users that talk something with different ways can be identified. For example, trough this methodology, it is possible to observe that a user that write a sentence with "B2B Solution", "attest" and "evaluation usability" can be interested in knowing other user that write a sentence with "usability test", "validate" and "enterprise system", because there are different words but there is a common issue, as show in Figure 7.

5 EXPERIMENT

We conducted a survey by email in February 22th, 2010 to a group user Twitter (www.twitter.com) in order to observe the use of methodology described in

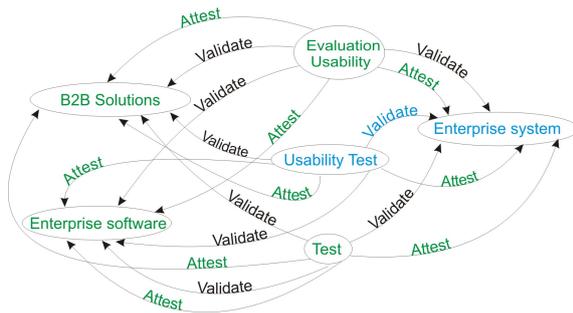


Figure 6: Semantic Network with Cultural concepts.

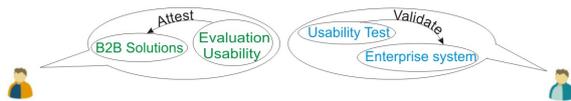


Figure 7: People with common interests but different cultures.

this paper. There were three steps in this experiment: First, it was sent an email requesting a sentence to each user, which showed their interest about a certain issue and context. We guide each one, because they needed to provide and to type a sentence according to the grammatical structure required by our methodology, i.e., subject, verb and object. Second, using the sentence of the each user and API from Twitter it was applied our methodology in order to find users who have some interests taking into consideration a certain issue and context. Third, we sent to each user a new email with the recommendation and we also asked their opinion about the results through this email. The questions were:

- Do you already "follow" this person? [yes/no];
- Is this a good recommendation? [yes/no].

5.1 Results

There were more thousand "tweets" (user's post in the Twitter) recovered of different users of the Twitter taking into consideration the sentence typed by participants of the experiment. Through this amount only 3.36% of users were chosen to be recommended to participants of the experiment. This is possible because we firstly seek "tweets" with the subject and object of each one (see Figure 6) and after we analyze the semantic relation between them in order to verify whether the "tweet" is within the issue and context desired by the user, i.e., if they have the same interests.

The Figure 8 shows the relation among the "tweets" recovered in the search. The part in yellow (96.64%) shows the "tweets" discarded, because words were not related semantically. The red part

(2.87%) shows the "tweets" were found with the semantic relation generated from the sentences provided by the participants of the experiment. The green part (0.49%) shows the "tweets" were found with the aid of the cultural expansion.

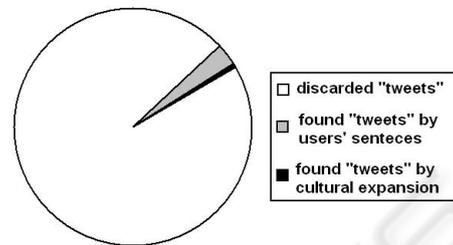


Figure 8: Resulted web mining in the Twitter.

These results show that there was low number of users selected for recommendation. On the other hand, there were good chances that they are inserted into the same issue and context. If we recommender 100% of users selected without considering the semantics involved between the words used in the search, we would do certainly mistaken recommendation and we would not satisfy the user's intention, which is to find people with similar interests. In addition, 0.49% of the "tweets" found from the cultural expansion justify the use of our methodology to find people with same interests but different cultural knowledge, i.e., to search and find others people who express themselves differently. As an example, a user, of the experiment, typed the sentence "Amazon filters the oxygen of the world" and from this sentence was found another user typed the sentence "Amazon forest strainers O2 of the world, because of that it is very important to consider human life and the cultural knowledge... ", i.e., different in expression, but inserted on the same issue and context.

Overall, our users rated 88% of the good recommendations; followed by 12% of bad. Considering these results the methodology proved to be useful in order to find people who express themselves differently in the same issue and context. Although he has been experiment found few people with common interest, our methodology was efficient to find people considering the same words, semantically connected, that people typed, as well as, others concepts obtained through cultural knowledge, i.e., just the same interests.

6 CONCLUSIONS

This paper described a methodology in order to identify users with some interests but different cultures.

Through this methodology, the users can identify who are people talking about the some subject with different ways, then, they can talk about it and exchange experience, knowledge, solutions, among others. Because of that, employees in an enterprise environment have another opportunity to use social networking service to find people, with or without the same culture, to talk, to ask help, etc.

This methodology can be used in many systems, such as: social networking service, enterprise system or any tool that needs to improve the search mechanism because in this case, the methodology was used to identify people but the same process can used to identify educational materials, reports which each user defined a different name taking into consideration his culture, among others. These results suggest that if the Social Match Systems considers semantics, issue, context and culture to search similar people, they could do recommendation more robust.

ACKNOWLEDGEMENTS

We thank CNPq, FAPESP, EMBRAPA and CAPES for partial financial support to 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

- Bick, E. (2000). *The Parsing System "PALAVRAS": Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*. PhD thesis, Department of Linguistics - University of Aarhus.
- Boyd, D. M. and Ellison, N. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1).
- Chen, J., Geyer, W., Dugan, C., Muller, M., and Guy, I. (2009a). Make new friends, but keep the old: recommending people on social networking sites. In *CHI '09: Proceedings of the 27th international conference on Human factors in computing systems*, pages 201–210, New York, NY, USA. ACM.
- Chen, W.-Y., Chu, J.-C., Luan, J., Bai, H., Wang, Y., and Chang, E. Y. (2009b). Collaborative filtering for orkut communities: discovery of user latent behavior. In *WWW '09: Proceedings of the 18th international conference on World wide web*, pages 681–690, New York, NY, USA. ACM.
- Howard, B. (2008). Analyzing online social networks. *Commun. ACM*, 51(11):14–16.
- Kautz, H., Selman, B., and Shah, M. (1997a). Referral web: combining social networks and collaborative filtering. *Commun. ACM*, 40(3):63–65.
- Kautz, H., Selman, B., and Shah, M. A. (1997b). The hidden web. *AI Magazine*, 18:27–36.
- Liu, H. and Singh, P. (2004). Conceptnet — a practical commonsense reasoning tool-kit. *BT Technology Journal*, 22(4):211–226.
- Matsuo, Y., Hamasaki, M., Nakamura, Y., Nishimura, T., Hasida, K., Takeda, H., Mori, J., Bollegala, D., and Ishizuka, M. (2006). Spinning multiple social networks for semantic web. In *AAAI'06: proceedings of the 21st national conference on Artificial intelligence*, pages 1381–1386. AAAI Press.
- Minsky, M. (1988). *The Society of Mind*. Simon and Schuster, New York, NY, USA, 2nd edition.
- Silva, M. A. R. and Anacleto, J. C. (2009). *Enterprise Information Systems*, volume 24 of *Lecture Notes in Business Information Processing*, chapter Promoting Collaboration through a Culturally Contextualized Narrative Game, pages 870–881. Springer Berlin/Heidelberg, Berlin, Germany.
- Spertus, E., Sahami, M., and Buyukkokten, O. (2005). Evaluating similarity measures: a large-scale study in the orkut social network. In *KDD '05: Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pages 678–684, New York, NY, USA. ACM.
- Tang, J., Zhang, J., Yao, L., and Li, J. (2008). Extraction and mining of an academic social network. In *WWW '08: Proceeding of the 17th international conference on World Wide Web*, pages 1193–1194, New York, NY, USA. ACM.