FOLKSONOMY AND HEALTHCARE SOCIAL NETWORKS – PERSPECTIVES IN MACEDONIA

Zoran Mandev
FEIT
Skopje, Macedonia

Suzana Loskovska
FEIT
Skopje, Macedonia

Ivan Chorbev
FEIT
Skopje, Macedonia

ABSTRACT

Collaborative tagging nowadays provides enormous amount of user generated taxonomic information. This paper presents the basic concepts of folksonomy - collaborative tagging. The special case and circumstances in Macedonia are taken in consideration. An analysis of the usage of social networks in Macedonia is presented as well as the current situation in the domain of health care web sites. We present an overview of current problems that seekers for health care information in Macedonia are facing, especially citizens in Macedonia not proficient in English. A hierarchic model for organizing user-generated tags is proposed.

I. INTRODUCTION

The expansion of the Internet provided a lot of information practically about everything. Of course, the validity of the available information is a problem by itself. However, the amount of information causes another problem – it is very difficult to find the relevant information, and also the classification process becomes more expensive. Even now when there are effective and powerful search engines like Google and Bing, which use special algorithms for extracting the most relevant information, it is very difficult to find what we are looking for. The problem is more intensified when it comes to finding information about healthcare. For instance: How a person diagnosed with an illness or condition (acute or chronic) will be able to find the information that he or she is looking for, without being swamped in the massive amounts of data and useless commercials that exist on the web. For this kind of people, finding the relevant resources is very important because it can help them in decision making about treatments, lifestyle and activities connected to their disease which finally will lead to them feeling better.

This problem is bigger in the countries where English is not the native language (the most of the information on the web is in English). English is not the native language in Macedonia. There are a lot of people who do speak English, but most only have basic understanding. Very few people know words and phrases that are closely connected to specific medical domain. Only the ones that are language experts and the ones that work in the field (in this case medical and health care) are familiar with the English terms that are exclusive for the specific domain.

We propose a model for tagging the content which will try to overcome this problem and enable the users to generate their own dependencies between the tags and content (close related to some specific decease).

There are 830000 (February, 2011) Facebook users in Macedonia [1]. Considering this fact and the unofficial numbers about the amount of visits to the existing web sites in Macedonia, we conclude that there is room for implementing system for collaborative tagging in social networks that focuses on healthcare. People are used to socialising on Facebook and other international social networks. The tags that they generate in these places are frequently used for targeting ads, but they are also used for generating feeds that are closely related to customer needs.

The main purpose of using tags is gathering user generated knowledge which will help creating a system which will be semi-automated in the beginning and fully automated at the end. With this approach the final version of the system will be able to work independently. In other words, there will be no need of domain experts to closely monitor the system (in the medicine and healthcare area that would be doctors, medical students or pharmacists). This will enable lower cost of the building and assigning resources compared to the classical taxonomies. Having this knowledge the system will be able to provide more relevant information to the end user. It will also provide statistic data about the existence of the specific decease or condition.

The collection of tags that will be gathered as the result of the process is called folksonomy. The word is a combination of the two words: folks and taxonomy which means that the taxonomic information is gathered by common people (folks).

In the proposed model we are using horizontal and vertical relations between tags, which will enable to link tags with the same meaning in different languages

(Macedonian and English). This will provide linking to the Macedonian names of the deceases and medical conditions with their English and Latin names, which will further enable the users to read the contents connected to tags and search for similar conditions and deceases. The system will enable adding new tags and linking them to the existing ones, and will provide solution for determining the relevance of the relationships between the tags and the content.

II. RECENT WORK

In the last several years there were a number of projects where computer engineers worked together with doctors, students and researchers in the area of medical and health care in order to provide better services for the patients by using the advantages of Web 2.0.

In the Great Britain in 2006 the National Health Service came to a conclusion that in order to extend the life expectancy of patients and to reduce the cost of medical care, the government should promote a campaign which will lead to patient-led revolution in the health care. The proposed system would enable direct participation of the citizens in planning and developing the healthcare they use. The same conclusion was reached by the leaders in the computing industries in the United States. They pointed out that using the advantages of Web 2.0 promotes the emergence of the so called Health 2.0 [2], providing various benefits for the patients. The patients, by their direct inclusion, will be able to lobby for better healthcare laws, they will also be included in decision making about medical care. Better conditions for acceleration of biomedical scientific discovery will be provided. Opposed to the positive expectations, there are some concerns bothering public health officials. Some of the concerns are focused on the validity of available data. There are fears that the large amount of user generated content produced by non experts in the area of medicine and healthcare can mislead some of the patients. Patients can be lead to decisions which in the future can jeopardize their health or even have fatal consequences.

In 1999 Karen Parles launched a web site that focuses on patients that suffer from various kinds of lung cancer. This site and many others have contributed in the expansion of the Health 2.0. These sites provide communication among patients with same or similar illnesses. The sites also provide anonymity leading to more honest and effective communication. By sharing patients can cope with their condition more easily, and can be advised by people who experienced the same or similar illness [3].

The increasing amount of medical information that is user generated can be useful not only for patients and their families but also for medical professionals and caregivers. They are able to use the statistic information in order to provide better service. Physicians are able to focus on the current problems that most of the patients have. This however raises other concerns, primarily about privacy issues. The question how data from users should be handled and whether statistic data can be used even when we don't know where it came from.

When social networks are mentioned today, most of us will think of Facebook or Twitter. The majority of social networks these days are focused on the teenagers and young people in general. With the popularization of smart phones, the easy access to broadband internet, social networks become increasingly popular. Some applications have strictly defined use - specialized apps for searching the nearest healthcare center [4]. mCare is a wireless mobile system based on a social networking platform in health care. It is a prototype of a mobile application that provides its users several functionalities essential for building social network communities. The users can search for physicians and specialists; they can also put comments and ratings about physicians. There is a module that provides asking questions, searching the database for questions and answers. Every answer can be commented and rated by the users, and the answers can be sent to the user's friends.

Web sites about health care should provide quality and relevance of the information they provide. They should be user sensitive, smart and they need to be able to identify user needs. BCKOnline is a portal whose target group are breast cancer patients and their families, carers and friends. The resources that are included in the knowledge repository are described by using metadata from a user perspective. Besides description of the resource the metadata embraces description of the audience (age, stage of the cancer, etc). By using of these attributes the users can profile their information needs.[5].

Besides the user generated content and user generated taxonomies, there is a need for experts in the specific domain, who will be included in identifying the resources. In other words, when knowledge repository about some illness or condition is created or maintained in order to provide better quality of the content medical or health care expert is needed.

Almost all social networks now use collaborative tagging. Users are allowed to add their own annotations on the resources. This provides a large amount of user generated content, but also creates a problem when it comes to classifications of the tags [6]. Although health care is a specific domain, further classification of the tag's different sub domains is needed. There are several approaches for classification, like forest of trees, directed acyclic graphs, etc.

For better search output for tags related to a tag called core-tag, a core-tag oriented spectral clustering algorithm is proposed [8]. It introduces core-tag oriented space and a method called core-tag oriented spectral clustering. With this approach, when the users search for a tag, called core tag, they will get a result which will contain the related tags, but it will exclude the tags that are very close to each other in the cluster. Exclusion of the tags that are close to the core tag lead to simplifying the search output, which ultimately gives more accurate results.

III. FOLKSONOMY

The word folksonomy comes from the word folks meaning people and taxonomy. It represents a type of a distributed classification system which is created by a group of the resource users [7]. Users meet this kind of features at almost every site for social networking (ex: Facebook). Digital resources, images, texts, videos, posts on the social network site are one of the many things that can be tagged. The tags can be observed as means for organizing information. Because of the fact that the tags are not generated by the experts in the domain, they can't be used for formal classification.

Generation of the content by users comes with its faults. The tags the user adds may have meaning that everybody will understand, but in the most cases the tags are personalized (it make sense only to the creator of the tag, and possibly some small group of friends); every different word has unique meaning. All of this leads to chaotic set of tags which can make searching almost impossible. But, this is not always recognized as a problem, because of the fact that the tags are primarily meant to ease the use of the one submitting the tag.

Basically the tag is a set of freely chosen textual words (keywords, metadata, etc.). They do not follow any formal rules, because they are not generated by experts in the specific domain. Also, the tags can be in every possible language, which in some cases may not be the same as the language of the text being tagged. The user can tag the digital resource with any word that makes sense to him/her. When making a system that can provide features that can ease the searching of the tags, understanding the behaviour of the users is needed. There is little known about how the users make decisions in the process of creating tags. The best approach is to monitor the most popular resources (the ones with the most tags) and to monitor the behaviour on a single user. Although the tags may have meaning only to the creator, usually there are some tags that are used by others too, having socially shared meaning. This can happen when the users come from similar surroundings and share common interests. This kind of approach enables extracting the most popular (in some cases most relevant) tags from the system.

A test taken on tags on Flickr [7] shows that the popularity of a tag decreases rapidly as the number of tags is increased. As the number of users is increasing the number of tags used only once is increasing. This happens because of the fact that people come from different places, with different background and they mostly speak different languages.

Users don't pay much attention when tagging, and even when the tagging is in only one language there are single-used tags just because of few common mistakes like: errors in spelling, badly encoded tags, not following the convention for using singular and plural, etc. One of the solutions for reducing the number of single-used tags is proposing a standard to the users which can include some rules that they should follow when tagging. Some of these propositions include: using plural, grouping words using underscore, using lower case and following conventions started by others. There is no guarantee that the users will accept the suggestions, and this becomes even impossible when the number of users becomes very big (ex: Facebook, Twitter, Flickr).

Many sites offer the users the possibility of editing their tags, which can ultimately lead to some improvement. Also, in the editing process the system can have a list of similar popular tags which can lead the user into correcting the possible mistakes. Ultimately, a perfect balance cannot be found, because as the number of the users/tags rises (the site gathers more user generated content - knowledge) the number of "irrelevant" tags will rise too.

IV. SOCIAL NETWORKS TAGGING

Today almost all social networks use tagging. These are user generated tags which can be added freely or can be used from list of suggestions.

On Facebook all user generated content, with the exception of posts and messages, can be tagged. The user chooses between the tagging of some of his friends or adding a custom (set of keywords) tag. Although there is content available to everyone (and content that specific user has access -content generated by friends), Facebook still does not allow tags based search. It uses the tags to achieve better user experience, ex: displays the activity of the users that have common images (both users are tagged on the same image), if not manually set differently, suggests common friends, uses boxes for displaying mutual images when the user sends message to his/hers friend, etc.

V. TWITTER ON THE OTHER HAND USES TAGS IN THE POSTS, ALLOWS SEARCHING FOR SOME TAG AND ENABLES DISPLAYING ALL THE POSTS THAT CONTAIN THAT TAG. THE TAGS HERE ARE IN THE POST (THE POST CONTAINS THE TAG), AND ADDITIONAL TAGS CAN'T BE ADDED AFTER THE POST IS SUBMITTED.OUR ARCHITECTURE

We propose a model which offers an improvement in the organizing and structuring of user generated tags. This concept helps the users getting the most relevant search results. The goal is to enable users to reach the desired content without knowing the exact Latin or English medical terms. Figure 1 presents the architecture of the proposed model.

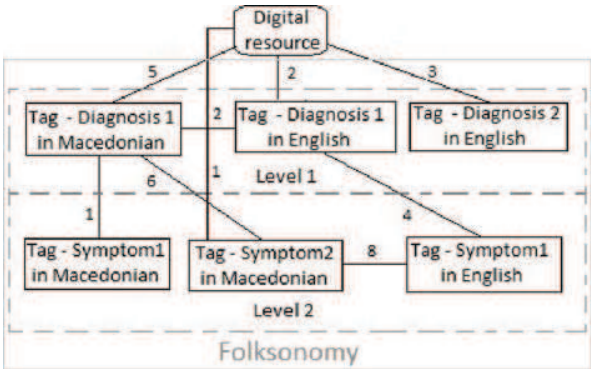

Figure 1: Architecture of the model

The users can add tags to some digital resource about health care, by choosing the level of the tag and assigning a name for the tag. We propose 2 types of tags, a 2 level architecture. In the first level, the tags contain the names of the diagnosis that are mentioned in the resource. In the second level, the tags describe the conditions connected with the diagnoses in question.

The tags in the first level can be connected (assigned) to the resource, but they can also be connected to another tag on the same level, for the same resource. This feature enables connecting the names of the diagnoses in different languages and creating connections between similar diagnoses. Having this kind of connections eases the search in the tags when it comes to searching for the words (names of conditions and diagnoses) that are in a language that is not the same as the language the resource was written in. The output of the search provides the resource found, but also the tags (if there are any) that are connected to the keyword the user is looking for.

Second level tags can be assigned to the resource, can be connected to a tag from the same level – same as first level tags, but they can also be assigned to a tag from the first level. Tags about symptoms can be assigned to only one first level tag (diagnosis). These limitations are introduced to minimize the appearances of the single use tags.

Besides the ability of assigning tags, every user is able to vote (the maximum number of votes per connection is limited to 1) for the connections between the tags and the resource. Having weight of the connection gives better output for searching and the users get more accurate information based on social knowledge.

For further preventing of single use tags that have no votes, we set a timeout period for every tag (connection). After the connection is assigned, a countdown timer starts and within that period the connection should get an expected number of votes in order to stay visible to everyone after the timeout. If the timeout expires without achieving the expected votes, the tag or connection will be visible only to the user that created it. The timeout period and the number of votes are set by the domain expert. The domain expert can also change the weight of the connections which improves the quality of the gathered knowledge.

VI. FUTURE WORK

The next phase of our work is the experimentation with the model and tuning its parameters. One of our future goals is finding an efficient solution for easing the communication of the users in such way that the system will automatically suggest friends to the user based on their tagging. With this feature, the end users will be able to get in touch with people that have common interest in some digital resources about health care and medicine which will lead to better user experience in the social network.

Algorithms for extracting statistical data from the user generated tags should be applied in order to ease

the analysis of the user's behaviour which can further be used in improving the model. The same data can be used by the physicians and government officials in order to retrieve information about what the receivers of health care are interested in.

VII. CONCLUSION

The expansion of the Internet and Web 2.0 enables improvements in the social network sites including the ones that are specialized in health care. Besides the content that is generated by the domain experts (doctors, medical students, etc.) it is allowed to the users to generate their own content – to be able to assign tags to the digital resources available online. This enables gathering large amount of user generated social knowledge. Although the tags in the vast amount of cases are used only once, some classification can be done. The extracted knowledge can be used to improve the user experience by providing better search results, providing the most relevant resource and enabling the users to get in touch with users that experience similar problems. All of this will enable better experience for the users, especially the ones who Internet for finding information about their illness or medical condition.

As mentioned before, there is big difference in naming the diagnosis and the condition in every language. Because of the fact that the most of the content on the Web is in English and often people do not know the correct terms in the domain of medicine and healthcare, finding the right resource on the internet is very difficult. Furthermore there are a great number of resources that are completely irrelevant, and the users are overwhelmed with information. The proposed model enables marginalizing the boundaries between the English language and other languages, reduces the appearance of single used tags and gives solution for better organization of the tags which provides better accuracy for searching.

REFERENCES

[1] http://www.facebook.com, February 2011

[2] Bradford W. Hesse, Derek Hansen, Thomas Finholt and Sean Munson and John C. Thomas "Social Participation in Health 2.0", *IEEE Computer*, pp 45 – 52, November 2010.

[3] Mari Carmen Domingo, "Managing Healthcare Through Social Networks", *IEEE Computer*, pp 20 – 25, July 2010.

[4] Weider D. Yu and Asim Siddiqui, "Towards A Wireless Mobile Social Network System Design in Healthcare", *2009 Third International Conference on Multimedia and Ubiquitous Engineering*, June 2009 Qingdao, China.

[5] Joanne Evans, Rosetta Manaszewicz and Jue Xie, "The Role of Domain Expertise in Smart, User-sensitive, Health Information Portals", in *Proceedings of HICSS'2009*. pp.1-10, 2009.

[6] Jonathan Klinginsmith, Malika Mahoui, Yuqing Wu and Josette Jones, "Discovering Domain Specific Concepts within User-Generated Taxonomies", in *Proceedings ICDM Workshops*, 2009, pp.19-24.

[7] Marieke Guy and Emma Tonkin, "Folksonomies. Tidying up Tags?" *in D-Lib Magazine*, Vol. 12, No. 1, 2006

[8] Yexi Jiang, Changjie Tang, Kaikuo Xu, Yu Chen, Liang Tang, "CTSC: Core-Tag oriented Spectral Clustering Algorithm on Web2.0 Tags", *2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery*, August 2009, Tianjin, China