

Contextual presentation of medical forum's discussions

Ghazar Chahbandarian¹, Mustapha Mojahid², Nathalie Bricon-Souf³

Université de Toulouse; UPS, IRIT; Elipse, 118 route de Narbonne, F-31 062 Toulouse Cedex 9,
France

¹Ghazar.Chahbandarian@irit.fr

²Mustapha.Mojahid@irit.fr

³Nathalie.Souf@irit.fr

Abstract: Medical forum sites contain a large amount of data. In this paper we designed a model that reads the contents of a thread in a medical forum and analyzes the contexts used in it to visualize the discussion in a better way. We implemented the model in an interactive tool that analyzes any thread in Doctissimo's forum by entering the thread's URL to visualize it in three different ways. The implemented tool helps the users to see the active authors who use the medical words most and to understand the progress of a thread in the forum. The project has promising results although it is still at its early stages.

Key words: Context, medical forums, Health informatics.

1 Introduction

Forums are the place where people go to discuss and find solutions to their problems. One of the special kinds of forums is Online Health Forums which are specialized in the discussion of medical issues. However the large amount of data in the medical forums and the overlapping relations between the data make them harder to analyze and to extract useful information. Furthermore, it's difficult to deal with posts containing inappropriate language and to construct a global view of the interactions between authors. The objective of this paper is to design a tool that analyzes medical forums thread and to provide visual presentations other than the standard web presentation with the ability to change between the views interactively according to the users' preference. Therefore, different users have better global understanding and they can figure out and focus on the important information in the medical forums.

2 Background and Related works

We made the hypothesis that, with a tool which allows the user to propose different contextual displays of forum information, we could discover how to select context to highlight some features of the discussions such as the pertinence of the exchange according to a medical question or the dynamicity of the post. This section describes some background knowledge about forums and context elicitation.

2.1 Forums

Forums are the place on the internet where people interact in an open discussion and exchange ideas through posted messages which in the most cases are archived to be accessed any time. Moreover, they provide easy usage thanks to their beautiful interfaces and diverse

functionalities. These forums are used increasingly in medical domains especially by the patients to find answers to their health problems. These forums are led most of the time by an individual who is usually non-professional, this person tries to pursue the problem that the group is interested in (Lederman 2013).

Most of the research focuses on extracting data from forums to store it in the search engine's indexes, (Limanto 2005) which deduces the html template of the forum and extracts the information about the posts eg., author, title, content etc.

Other research discusses the use of data mining techniques to develop a tool that helps the teachers to evaluate and observe the discussion in school forums by indicating the amount and the frequency of the participations (Dringus 2005).

In this paper we discuss and build a tool that analyzes a medical forum and visualizes it with different contextual presentations using the available contexts.

2.2 What is context

A lot of work has been done regarding context and context-aware applications. However, it is not always clear, what context really means. Different context definitions exist, one of the most famous is "*Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves.*" (Dey 2001). This definition helps the application developers to choose the right piece of information that characterizes the situation and to consider the information as a context for a specific application scenario, whereas in Brezillon's opinion context impacts more on the relationship between the pieces of information than on the pieces of information themselves and his definition is "*Context is that which constrains something without intervening in it explicitly*" (Brezillon 2007).

Context can be classified in two levels, low level context and high level context; the former is any information collected from direct sensors, it is acquired without any further interpretation, whereas the latter is calculated by interpreting the low level context. For instance, the GPS location for someone is a low level context, while deducing that he is in an office is a high level context.

Context in health care area is still modest and research is ongoing. Context elicitation is used to deal with the selection of relevant contexts in the application and it is often considered as major problem, care recommendation is different according to the available health care devices, for instance, Magnetic Resonance Imaging is not available in all hospitals. "*Contextual situations are frequent and for each situation it is difficult to determine what the contextual elements are*" (Bricon-Souf 2013).

3 Model designing

In order to handle the challenges we have first to process medical forums and that is the preliminary stage on which we are now working. In this section we describe a model to process medical forums. The idea is to be able to capture some threads, extract different information contained in them and choose what could be relevant context in order to propose an interpretable view of the data.

The process of generating a context-based view of a forum can be summarized in four steps. 1) Reading a forum thread. 2) Identification of low level context. 3) Calculation of high level context. 4) Visualization of information. A brief explanation of each step is presented in the following:

3.1 Reading a forum thread

The first step of our model prepares the data by automatically reading the content of any desired thread to process it in upcoming steps. The content might be stored locally or it could be accessed remotely by simply using the thread's URL.

3.2 Identification of low level context

The objective of this step is to identify relevant information stated explicitly in a forum's thread, such as author's name, the date and the time of the participation, citations, etc. The identified information is considered as low level contexts due to the extraction without any calculations. We identify the contexts by analyzing the structure of the page. Then we figure out the information patterns. Finally we extract and store the information to be used in the next step.

3.3 Calculation of high level

This section does calculations on the low level context to get information of a higher context level. For instance we have performed calculations to get: *Total duration of the forum* by subtracting the last post's time from the first one; *Sub discussion count* by counting how many hibernates are there in a discussion; *Participant count* by counting all the authors in a discussion; *Period of the day* by analyzing the time; *Length of the post* by counting the words in a post; *Medical or not* by counting how many medical words exist in a post using medical words dictionary. We also show the author's expressiveness by focusing on the punctuations and the smileys he is using in a post.

3.4 Visualization

The last step of the model is responsible of the visualization of the information in a forum. To provide a better understanding of a thread's content we used the emerging context from the previous steps in order to propose a contextual visualization of the threads according to the user's choices.

We define three main types of contextual visualization:

- Time line: it displays all the occurrences of the posts with the same order than in the original forum content. The distances between occurrences are proportional to the duration of the posts.
- Citation: the authors often quote each other; this choice displays the posts according to the citing relationships between them.
- Author's Time Line: the posts are classified according to their authors and are ordered in chronological order.

We see in figure 2 and figure 3 examples of an implementation of context visualization.

4 Implementation

We implemented a functional prototype of the context-based model in a tool written in java designed with an interactive interface. The input of the tool is an html file or a web URL. It identifies all the contexts, and presents them depending on the user's choices and configurations.

“Doctissimo medical forum¹” is one of the most famous French medical forums and lots of people rely on it to search for their medical problems. As a first step our tool analyzes the threads of this forum which contains a lot of threads. Each thread discusses one issue,

¹ <http://forum.doctissimo.fr/>

generally first post of the thread represents the question or the issue to be discussed, after that users express their ideas to discuss and solve the issue. The tool will be generalized later to analyze any medical forum.

In the following sections we describe the interface of our tool and follow it up with a case discussion.

4.1 Toolbox area

This area allows the user to choose what to display and how to display it. It's located on the left side of the interface with the ability to be hidden when a wider preview area is needed, in the toolbox the user chooses the configuration of the context to use in the visualization. The toolbox consists of the following options and configurations: *File chooser* the area where the user chooses the html file or the URL of the forum thread to be analyzed; *Low level context area* and *High level context area*: it consists of sliders and checkboxes to give the user the ability to choose which context to display; *Relationship area*: where the user chooses how to represent the relationships between the contexts.

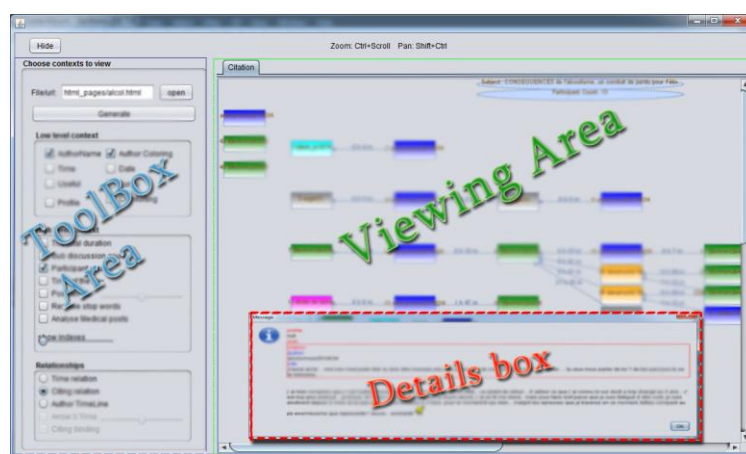


FIGURE 1 – The areas of the interface.

4.2 Viewing area

This area shows the contextual visualization of the thread's content. It is located on the right side of the interface; it is responsible to display all the contextual information and relationships between each other depending on the user's configurations in the toolbox area. Some of available functionalities in this area are: Zooming and Panning. Moreover, the user can find more information by clicking on any post which brings him/her to the details box.

4.3 Details box

This box contains additional context information, when the user clicks on any representation of a post, a window will pop up with the complete information of the original post such as the author's profile, the original text and citations if there are any. The surface indexes² or the medical words will be highlighted if the associated checkboxes are selected.

In order to explain the viewing area and the details box we will show screen captures of our tool.

² surface indexes: punctuation marks, numerations, smileys ...

Figure 2 represents a screen capture of the viewing area, the illustrated view is zoomed out and the toolbox area is hidden to provide a wider view area, the blue lines in the figure represent the author's time line, the boxes represent the posts and the red arrows represent the citation relations i.e. when an author cites other posts. The small red box inside each post represents how many medical words are used in the post. We are now testing with a very simple medical word list but we intend to use more sophisticated terminology such as SNOMED for the next step.

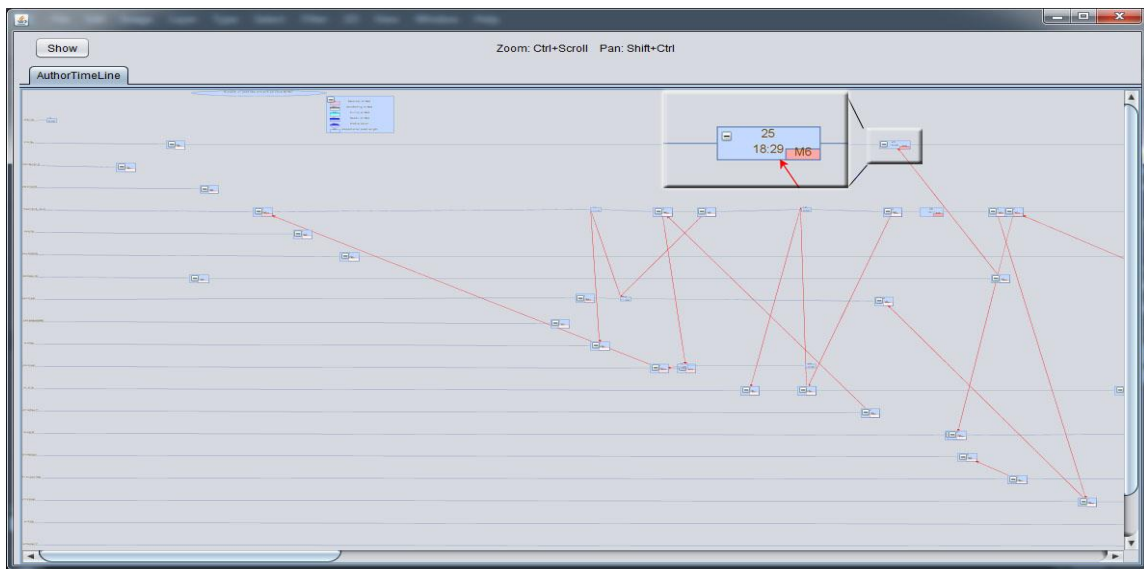


FIGURE 2 – Capture screen of the tool showing the author's time line.

Figure 3 represents a screen capture of the details box where the configuration of viewing medical words and surface indexes are chosen, we see that all the medical words are bold, all the surface indexes are black, in contrast all other characters are gray, in this way the user can focus on the chosen contexts more easily.

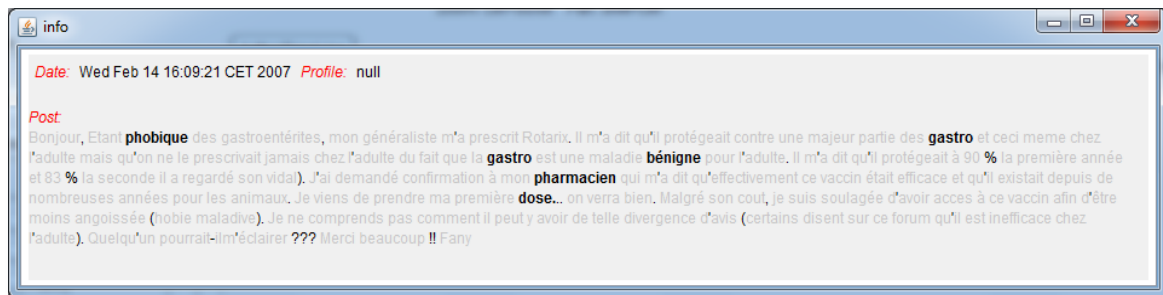


FIGURE 3 – The details box of a post.

5 Discussion

The tool allows getting different views of the posts in a thread. We hope that it helps to observe specific behavior patterns and provide better understanding to the forum.

Figure 2 is an example where the user chooses the author time line view with the medical words and citation binding options. One of the important observations is detecting the active authors, which is clear in the right part of the view area; we observe as well the fifth author from the top uses a lot of medical words and citations. Furthermore, the author who started the thread never participated anymore after asking the question.

Moreover, by examining each post we observe the authors' language level, the information quality, expressiveness and the use of medical words.

We notice in figure 3 that the author uses some medical words and a lot of punctuations such as brackets, exclamation marks and full stops. We can conclude that the author tries to organize his ideas and make them clear to the viewers and that increases the quality of his participation.

6 Conclusion and future work

In this paper, we presented our work to process and analyze medical forums to provide better visualization. We provided an interactive tool to observe the authors' behavior and their interactions between each other, in addition to the overall progress of the thread.

As a result we made a step forward in realizing the challenges by detecting the qualified users in the medical forums and detecting the participations with highest quality in general.

The next step will be evaluating our work by testing the tool with different user categories such as sociologists, patients and doctors. We will prepare different testing scenarios and ask the users for their remarks and their satisfaction to improve our tool.

This work is an initial phase of a PhD research that seeks to study what makes a specific context more important than the others in general and we would like to apply the results in synthesizing electronic health records to a wide variety of contexts.

Acknowledgement

We thank the Midi-Pyrénées region and the metropole Castres-Mazamet as well as the Languedoc-Roussillon region through the MSH-M “parlons de nous” project and inter-MSH “A quoi pensent les patients” project.

References

- BELLAVISTA P., & AL. (2012). A Survey of Context Data Distribution for Mobile Ubiquitous Systems. *ACM Computing Survey* 44 (4) (August 1): p. 1-45
- BRÉZILLON P. (2007). Context modeling: Task model and model of practices. In *Sixth international and interdisciplinary conference on modeling and using context* p. 122–135. Roskilde, Denmark.
- BRICON-SOUF N., & NEWMAN C. R. (2007). Context awareness in health care: a review. *International Journal of Medical Informatics*, 76(1), p. 2–12.
- BRICON-SOUF N., & CONCHON E. (2013). Context Awareness for Medical Applications. *Medical Applications of Artificial Intelligence*, 355.
- DEY A. K. (2001). Understanding and Using Context. *Personal and Ubiquitous Computing* 5(1) (February 28): p. 4-7.
- DRINGUS L. P., & ELLIS T. (2005). Using data mining as a strategy for assessing asynchronous discussion forums. *Computers & Education*, 45(1), p. 141–160.
- KIM J., & CHUNG K.-Y. (2011). Ontology-based healthcare context information model to implement ubiquitous environment. *Multimedia Tools and Applications*. *Multimedia Tools Appl.*
- LEDERMAN R., & AL. (2013). Who can you trust? Credibility assessment in online health forums. *Health Policy and Technology*, p. 1–13.
- LIMANTO H. Y., & AL. (2005). An information extraction engine for web discussion forums. *Special Interest Tracks and Posters of the 14th International Conference on World Wide Web*.