GEIST research on Context-Aware systems

The concepts of context and context awareness have been studied for more than 20 years in the field of artificial intelligence, computer and cognitive science. However, it has been still identified by Gartner, alongside cloud computing, business impact of social computing, and pattern based strategy, as being one of the broad trends that will change IT and the economy in the next 10 years [1].

Moreover these IT and economical changes are reflecting themselves also onto business applications. Applications are simplifying, are becoming mobile, are moving to the cloud, are getting more social and user focused [2]. Hence we are faced with a series of new challenges in the context of developing future business apps.

They need to:

- be highly adaptive;
- provide UIs that are user specific;
- provide means for users to modify them by themselves according to their needs, goals, context, while still keeping the underlying infrastructure in place;
- be interactive;
- be distributed, at the same time cloud computing ready;
- support both desktop and mobile environments, while providing a similar experience, finally
- developed with business users, and vendors and for customers, while hiding as much as possible all technical requirements.

Considering how wide the research area is, providing a holistic, yet analytic perspective on these concepts remains a challenge. We employ a new research methodology that aims to address and visualize the topic of context and context awareness from a holistic point of view, by means of text mining and text clustering.

Research methodology

There is a huge amount of work that tackles the problem of context and context awareness in different fields and from different aspects. However, there is no unified view on the matter, nor – to the best of our knowledge – there is any approach that provides a holistic view on the subject. Therefore we propose a research methodology which takes advantage of the existing techniques for text clustering and text mining to get a broader view on the research that has been done on the side of context and context awareness.

The motivation to use text mining and clustering techniques is very simple. Too many papers that need to be organized, make the task almost impossible to fulfill. Moreover such an approach will provide an automatic way to extract related terms, topics and directions of research.

We present our methodology in a form of a simple workflow, modeled as a business process model, designed using the BPMN [3] notation and depicted in Figure 1. The model presents the steps that we took in our research approach and the ordering of those steps.

Figure 1: Research Methodology - Business Process Model (BPMN notation)

We have compiled a bibliography file which so far contains 94 carefully selected bibliographic entries that spans over a period of more than 20 years, starting 1991-2013. The quality of the papers is also an important factor. There are two ways to weight and asses the quality of the papers. One way is objective as it is given by the number of citation a paper has. We have extracted the number of citation, where this number existed, for a paper from digital libraries websites: CiteSeer, Google Scholar, ACM Digital Library, IEEE Xplore Digital library. In the cases where there is no available citation number, we can not know for sure if a paper has been cited or not, therefore it is up to the researcher to read and asses the quality of a paper. This approach one could say that is rather subjective.

The steps for compiling this bibliographic collections are depicted in Figure 1. We start by searching via Google for context related keywords i.e. *context*, *context-awareness*, *context-aware surveys*. A survey is a better entry point as it provides a wider view on a subject. These are just entry search terms. The more you search and read, the more terms can be further used. Besides the "random" search, we followed (searched) also concrete references that were indicated in the initial papers that we retrieved and read.

The next step in the process (See Figure 1) is to add bibliographic entries. We used for the clustering algorithms the abstract of each paper, if there was one. In consequence a bibliographic entry, if there is one, needs to have an abstract. Some of the papers also contained keywords. We have also used when available the keywords associated. These were combined with the abstract.

We used JabRef to compile our bibliography. JabRef offers the functionality of an export layout, which we are using to export the bibliographic information into Carrot2 input format. Carrot2 as stated on the project website is an "Open Source Search Results Clustering Engine. It can automatically organize small collections of documents (search results but not only) into thematic categories". The reason for using Carrot2 over other tools (such as Lemur, Terrier) is its simplicity. It was very simple to write an export layout from JabRef to Carrot2 XML input format. The export layout is also available online at the previously given address. And also the results are by default given also in several visual formats.

Although Carrot2 provides several search algorithms we used Lingo and K-Means algorithms as they provided the best results. Unfortunately the free version of Carrot2 does not provide options to addresses issues such as synonyms in order to improve the results. Arthur and Vassilvitskii state in [4] that the K-Means method is a well known geometric clustering algorithm based on work by Lloyd [5]. Though the K-Means term has been first used by MacQueen [6]. According to Arthur and Vassilvitskii, given a set of n data points, the algorithm uses a local search approach to partition the points into k clusters. Lingo [7] as described by the authors is able to capture thematic threads in a search result, that is discover groups of related documents and describe the subject of these groups in a way meaningful to a human.

Figure 2: K-Means - Foam Visual Representation

Figure 3: Lingo - Foam Visual Representation

Figures 2 and 3 depict the results of running the K-Means and respectively Lingo algorithms over our bibliographic collection. The results are visualized in a Foam representation. Results are similar but not the same. We can easily visualize directions of research and words related with the context concept. Having similar results it helps to verify the output of the clustering algorithms. Having differences helps to identify what each algorithm has missed with respect to the other.

The authors of "Contextualization as an independent abstraction mechanism for conceptual modeling" [8] already identified that context is of fundamental importance for cognitive psychology, and computer science. Furthermore it states that in computer science the notion of context has been addressed in several areas such as: artificial intelligence, software development, databases, data integration, machine learning and knowledge representation. Since all these directions have been also identified by our research approach we argue that results are satisfactory in terms of how adequately the mining and clustering algorithms have performed.

In addition based on the information depicted in Figures 2 and 3, context has been used to address many of the future business apps challenges we have enumerated in Section I: adaptation, mobile computing, flexibility, user, modeling, task management, distributed systems, business process models.

Files

- Context bibliography file (bibtex inside zip) - contextbiblio.bib.zip
- JabRef2Carrot2 Export Filter - carrot2jabrefexportfilter.zip

How to

JabRef

1. Install custom export format
   1. Extract carrot2jabrefexportfilter.zip archive
   2. In JabRef menu: Options→Manage custom exports→Add new
   3. Use for name: Carrot2
   4. Browse and use for "Main layout file": carrot2xml.layout
   5. For "File extension" use: .xml
2. Open contextBiblio.bib file
3. File→Export and choose for "File format": Carrot2 (*.xml)

Carrot2 Workbench

- In the "Search" panel use for "Source": XML
- Choose the clustering algorithm: Lingo, K-means
- Browse and use the file you exported to the bibliography from JabRef

Findings

There has been done a huge amount of work that addresses the problem of context. And although this work has tackled different aspects and research directions, i.e. modeling, reasoning, data-bases etc., we argue that all this work, from the focus point of view, follows two major directions: context-aware applications that are system-centric (most part of the work) context-aware applications that are user-centric. These two directions act as an analysis framework for us and our further assertions revolve around these directions.

❌

Figure 4: Mind Map of Context Related Concepts for the User-centric perspective

Figure 4 depicts a mind map with the context related concepts for the user-centric perspective. We argue that the combination of these concepts together with proper techniques for modeling, reasoning and system specific execution facts can address the challenges we

Tools

- JabRef - http://jabref.sourceforge.net/
- Carrot2 - http://project.carrot2.org/

Papers

Team

- Emilian Pascalau
- Grzegorz J. Nalepa
- Szymon Bobek
- Krzysztof Kluza

[1)]
M. Wang, "Context-aware analytics: from applications to a system framework", http://eresearch.csm.vu.edu.au/files/apweb2012/download/APWeb-Keynote-Min.pdf, 2012.

[2)]
C. McLellan, T. Hammond, L. Dignan, J. Hiner, J. Gilbert, S. Ranger, P. Gray, K. Kwang, and S. Lui, The Evolution of Enterprise Software. ZDNet and TechRepublic, 2013. [Online]. Available: http://www.zdnet.com/topic-the-evolution-of-enterprise-software/

[3)]
OMG, "Business process model and notation (bpmn)," OMG, Tech. Rep. formal/2011-01-03, November 2011.

[4)]
D. Arthur and S. Vassilvitskii, "How slow is the k-means method?" in Proceedings of the 2006 Symposium on Computational Geometry (SoCG), 2006.

[5)]
S. P. Lloyd, "Least squares quantization in pcm," IEEE Transactions on Information Theory, vol. 28, no. 2, pp. 129–136, 1982.

[6)]
J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," in Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability, 1967, pp. 281–297.

[7)]
S. Osinski, J. Stefanowski, and D. Weiss, "Lingo: Search results clustering algorithm based on singular value decomposition," in Intelligent Information Systems, 2004, pp. 359–368.

[8)]
A. Analyti, M. Theodorakis, N. Spyratos, and P. Constantopoulos, "Contextualization as an independent abstraction mechanism for conceptual modeling," INFORMATION SYSTEMS, vol. 32, pp. 24–60, 2007.

From:
https://geist.re/ - **GEIST Research Group**

Permanent link:
https://geist.re/pub:research:context_aware?rev=1369412950

Last update: **2013/05/24 16:29**