

Displaying Hidden Information in Glossaries

S

Marcela Ridao

Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina

Jorge Horacio Doorn

Universidad Nacional del Oeste, Argentina & Universidad Nacional de La Matanza, Argentina

INTRODUCTION

All models created during the software development process and specifically during the Requirements Engineering Process, are created with structures and purposes clearly defined. These structures have been conceived to maximize the expressiveness of the model for its objective.

Despite this, and very possibly because of this, it may occur that this models hold information no perceptible during its routine use. From an epistemological point of view, this situation is very similar to what happens in data mining (Artz, 2009). For particular case of Requirements Engineering models, which are written in natural language, it has been observed that at least part of this hidden information can be discovered.

In this chapter, a strategy to visualize the grouping of terms in a glossary called Language Extended Lexicon (LEL) is proposed. Experiments performed on real world cases have shown that the clusters obtained using syntactic resources of the LEL model, coincide with semantic nuclei of the application domain. The proposed strategy is based on graphs constructed using hypertext links embedded in the LEL model.

Disciplines dedicated to the comprehension of phenomena whose dominant aspect is the structural complexity instead of the essential complexity of their components have been acquiring more importance (Barabasi, 2002; Dorogovtsev

& Mendes, 2003). There are many examples, in several disciplines, where the detection of groupings provides a notorious contribution to the comprehension of the phenomenon being studied (Sarmah, Kalita, & Bhattacharyya, 2011; Mo, Cao, & Wang, 2012; Zimmermann, Ntoutsis, Siddiqui, Spiliopoulou, & Kriegel, 2012). Most evident examples of this sort of problems are the organizational networks, social networks, bibliographic references networks and interest groups among many others. The classic visual representation of these networks is done by means of graphs. Actually, it happens that when the number of nodes surpasses a moderated limit, those graphs become not useful to observe the relevant characteristics of the network. Configurations of nodes and connections also occur in a great diversity of other applications. They may represent physical networks, such as electrical circuits, roadways, or organic molecules. They are also used when representing less tangible interactions as might occur in ecosystems, sociological relationships, databases, or in the flow of control in a computer program (Gross & Yellen, 2003, 2006; Kamada & Kawai, 1989).

The requirements prioritization method proposed by Duan, Laurent, Cleland-Huang, & Kwiatkowski (2009), analyzes documents of the Requirements Engineering process by means of detecting clusters of requirements. In Duan proposal, the requirements are included in clusters by

means of an iterative algorithm. Those clusters are object of negotiations among the stakeholders to assign priorities. Requirements inherit the priority assigned to the cluster to which they belong.

Some of the models of the Requirements Engineering may be studied from the structural point of view. Particularly, one of the most promising is the LEL (Leite & Franco, 1990). This model records the vocabulary used in the application domain. It describes the words and phrases used by clients and users with a meaning specific to the application domain.

Observing the LEL from a structural point of view, it may be immediately seen that it can be depicted using a graph where the vertices symbolize the LEL entries and the edges represent the hypertext links among entries.

From this point of view, the LEL may be visualized as a sort of linguistic network with a complex structure; in such a way that besides the explicit information stored in each symbol, there is implicit information stored in the relationships among symbols. To build and to analyze the graph of the symbols of the LEL is a sort of knowledge mining.

The most basic knowledge acquirable from the LEL graph is the existence of clusters of entries, which become roots for any taxonomy of the business process. Intuitively it may be said that if the application domain is divided into areas of interest or organization fragments, then it may be expected some coupling among the terms used in such area.

This chapter describes a mechanism to build the graph of any LEL, offering a good visualization of the existing symbol clusters.

The following section analyses the Language Extended Lexicon (LEL), a glossary built during Requirements Engineering process. Then, a section describing force directed algorithms applied to graph visualization is included. After that, the use of these methods in LEL graph construction is studied. And finally, some results, future work and conclusions are presented.

BACKGROUND

LEL: Language Extended Lexicon

The creation of a glossary defining the jargon used in the application domain has been highlighted since many years ago (Arango, 1994; Leite & Franco, 1990). Moreover, several experiences along decades have shown that these glossaries are, by themselves, an important source of information to acquire knowledge about the application domain. (Ben Achour, Rolland, Maiden, & Souveyet, 1999; Oberg, Probasco, & Ericsson, 1998; Prakash, Aurum, & Kox, 2004; Regnell, 1999; Rolland & Ben Achour, 1998). This fact is very well known by linguistics, since they have established that language reflects environment and technology. Just a few examples of that are i) Arabic which has 80 words for camel, ii) Japanese having more than 20 words for rice and iii) Inuit with more than 20 words for snow and ice. It is not needed to explain the importance of camel, rice, snow and ice for those peoples. (Nettle & Romaine, 2000).

The successful use of glossaries by actual practitioners has been reported for more than two decades (Weidenhaupt, Pohl, Jarke, & Haumer, 1998). The authors of this article have a library of more than 200 real cases glossaries developed by undergraduate students, graduate students, practitioners and themselves. The oldest cases were developed as early as 1990 (Leite & Franco, 1990; Rivero, Doorn, del Fresno, Mauco, Ridao, M., & Leonardi, 1998). Some of the most relevant are:

- A reflexive application of the requirement engineering process to the requirement engineering process itself (García & Gentile, 1999).
- Acquisition, administration and maintenance of ATM equipments, 2003.
- Quality Control of Medicaments Supplies, 2007.

9 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/displaying-hidden-information-in-glossaries/184439

Related Content

Rough Set Based Similarity Measures for Data Analytics in Spatial Epidemiology

Sharmila Banu K. and B.K. Tripathy (2016). *International Journal of Rough Sets and Data Analysis* (pp. 114-123).

www.irma-international.org/article/rough-set-based-similarity-measures-for-data-analytics-in-spatial-epidemiology/144709

Local Government, Decentralisation, Devolution, and Service Delivery in Zimbabwe

Tawanda Zinyama (2021). *Encyclopedia of Information Science and Technology, Fifth Edition* (pp. 1597-1611).

www.irma-international.org/chapter/local-government-decentralisation-devolution-and-service-delivery-in-zimbabwe/260290

Hybrid TRS-PSO Clustering Approach for Web2.0 Social Tagging System

Hannah Inbarani H, Selva Kumar S, Ahmad Taher Azar and Aboul Ella Hassanien (2015). *International Journal of Rough Sets and Data Analysis* (pp. 22-37).

www.irma-international.org/article/hybrid-trs-pso-clustering-approach-for-web20-social-tagging-system/122777

Complexity Analysis of Vedic Mathematics Algorithms for Multicore Environment

Urmila Shrawankar and Krutika Jayant Sapkal (2017). *International Journal of Rough Sets and Data Analysis* (pp. 31-47).

www.irma-international.org/article/complexity-analysis-of-vedic-mathematics-algorithms-for-multicore-environment/186857

Discovering Patterns using Process Mining

Ishak Meddah and Belkadi Khaled (2016). *International Journal of Rough Sets and Data Analysis* (pp. 21-31).

www.irma-international.org/article/discovering-patterns-using-process-mining/163101