Advances in Engineering Software 70 (2014) 25–35

Contents lists available at ScienceDirect

Advances in Engineering Software

journal homepage: www.elsevier.com/locate/advengsoft

Semantic requirements sharing approach to develop software systems using concept maps and information entropy: A Personal Health Information System example

Varadraj P. Gurupur [a,*], Unal Sakoglu [b], G. Pankaj Jain [b], U. John Tanik [c]

[a] Health Informatics and Information Management, Louisiana Tech University, United States
[b] Department of Computer Science and Information Systems, Texas A&M University – Commerce, United States
[c] Department of Computer Science, Indiana University – Purdue University Fort Wayne, United States

ARTICLE INFO

Article history:
Received 3 May 2013
Received in revised form 20 November 2013
Accepted 4 January 2014
Available online 2 February 2014

Keywords:
Semantic Web Technology
Structural properties
Concept maps
Semantic gap
Information entropy
Node source connectivity strength

ABSTRACT

In this paper we describe a process of developing software systems by capturing the conceptual domain knowledge of the problem domain using concept maps. To illustrate the implementation of this process we have used the example of developing Personal Health Information Systems. In addition to the aforementioned development process, the paper also describes an evaluation metric developed using Design Structure Matrix and information entropy to measure the structural properties of the concept map. The determination of entropy is based on the information derived from the hierarchical structure of a concept map. The probability distributions and the information entropy were calculated defining a new metric: node source connectivity strength based on the number of unique paths from a node to another. The results were compared by using a more standard metric, the graph node connectivity.

1. Introduction

In this paper we attempt to provide a solution that mitigates the semantic gap problem associated with the implementation of software systems using the example of a Health Information System [1,2]. Our attempt to mitigate this problem is based on capturing the research process using a process model and then capturing the semantic of the process using concept maps. A broad overview of possible solutions is provided in Section 2. In Section 3, a detailed overview of the solution to this problem has been provided. In Section 4, the solution is analyzed by applying Design Structure Matrix (DSM) [3,4] and information entropy [5,6] to mitigate the problem of semantic gap. We compare the semantic gap results of the new entropy measure based on the number of unique paths from a source node (concept map) with that of standard connectivity.

1.1. Semantic gap in building Personal Health Information Systems

A gap exists between user needs and software implementation efforts in the area of software development [7]. This problem persists mainly because software systems often do not fully support user needs [8]. This gap, termed a semantic gap (Fig. 1), exists mainly for two reasons: (a) lack of communication between the users and developers of the system and (b) lack of integrated support tools across the different software development phases, which further leads to a breakdown in communication. Chomsky [9] states that "grammar of a language can be viewed as a theory of the structure of this language." Since it is difficult to limit the grammar of a language to a particular group of sentences, the ambiguity involved in natural languages sometimes intensifies the breakdown in communication between user needs and implementation. This problem is critical because it leads to major inefficiencies in the operation of enterprises [10] and therefore results in increased costs and low productivity. Many enterprises have failed to succeed due to software failures caused by implementations not meeting the user needs [11]. The existence of a semantic gap may have the following effects:

(i) Difficulty in translating abstract information into an executable entity.
(ii) Difficulty in communication between the user and developer as a result of lack of domain expertise.
(iii) Increase in time required to build the implementation as per user needs.

There is a need for software requirements to be developed in a manner that allows them to be converted appropriately into

* Corresponding author. Tel.: +1 318 257 2667.
E-mail address: gurupur@latech.edu (V.P. Gurupur).

0965-9978/$ - see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.advengsoft.2014.01.001

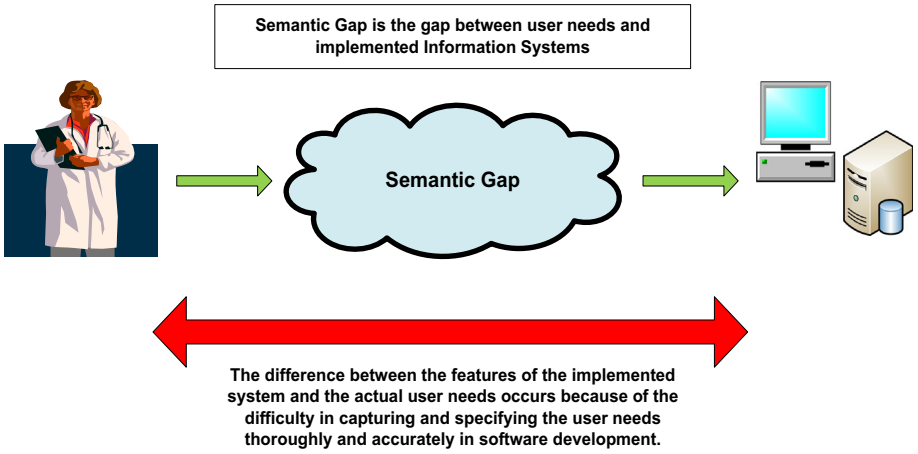

Fig. 1. Defining the semantic gap problem.

software design [12]. Usually, the problem domain is well understood by the domain expert, who educates the Requirements Engineer about the problem domain. The Requirements Engineer uses his/her understanding of the domain to build the requirements document. It must be noted that the Requirements Engineer may not fully comprehend the problem domain, especially if he/she must deal with unfamiliar terminologies. For example, if the Requirements Engineer needs to build software for biomedical research, then it is imperative that he/she be well versed in some of the complex terminologies and concepts used in health information systems.

To provide a streamlined approach to the aforementioned problem, we have devised a two phased approach that involves using semantic relationships between the attributes of a dataset. First, we facilitate the required identification of the process where a system is divided into subsystems involved in the process and their relationships. Second, we build a repository of these relationships and use it to extract data along with its semantic relationships from the database.

Usually any manual process requires the user to know the intricacies of the semantics of the data of database or an information system. If the process has to be automated by using a web-based application, the developer needs to comprehend the intricacies of the terminologies involved. This may involve additional learning time which may delay the progress of building the application. If the information system does not provide the semantics of the relationships of the various data elements, then the user is expected to know them. Therefore, we have built an information system using concept maps, and semantic web technologies to facilitate the storage and retrieval of these relationships.

2. Addressing the semantic gap problem

2.1. Need for Semantic Web Technology in health information systems

The process of health informatics could be further enhanced by using the tools and technologies associated with the semantic web available to the scientific community [13]. A shift from research based on painstakingly acquired evaluation, experience, and knowledge of the domain experts to that of a decision support system can accelerate the research process. Therefore, we propose the development of an integrated tool using concept maps with Web Ontology Language (OWL) [14] as an effective resource for enhancing biomedical informatics research. OWL is an essential part of the semantic web which is now being used to build infrastructure for storing metadata on a particular problem domain. OWL provides a high level of web automation by enabling information to be represented in the form of machine-processible ontologies for Personal Health Information Systems.

2.2. Need to develop a process model for a health research process

The process of building an information system for health research involves identifying the processes involved in the research [2]. In this paper we perceive a process as a collection of resources, tasks performed on these resources, and the flow of control of these tasks. A process model helps an individual clearly understand the research process. While the processes for enterprises have used process modeling paradigms used by methodologies such as CommonKADS [15], in this paper we describe processes by using Task System Models (TSM). A TSM is defined by Mills and Tanik [16] as "a process engineering formalism that supports the development and maintenance of distributed process systems." TSM are commonly used in designing operating systems. However, Sadasivam [2] in his dissertation explained that they can also be used to describe Health Information research processes.

2.3. Need for concept map-based approach

A concept map is a tool for representing and organizing knowledge [17] which is usually formed by assembling a constellation of concepts. This is done by encapsulating objects that can form complex relationships between concepts indicated by a connecting line linking them all together [18]. A particular concept can be depicted by using concept map tools such as IHMC C-Map Tools COE developed by the Institute for Human Machine Cognition (IHMC). These C-Map Tools software can be effectively used by the clinicians to represent an idea or a concept in a graphical format [19,20].

A concept map allows an ordinary computer literate to illustrate an idea, concept or domain knowledge to be described in a graphical form [21–23]. Novak and Cañas [22] explains that concept maps have many such features that can be effectively used in representing domain information. Table 1 provides the key features of the concept maps required for representing domain information.

Concept maps allow a two level representation of domain knowledge information:

(i) Concept maps allow the domain expert to build a schematic representation of the domain knowledge. This information can be easily comprehended by every individual involved in the development of a software system [8,23].