Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 35 (2014) 194 – 202

Procedia Computer Science

18th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems - KES2014

A case-based reasoning system to support the global software development

Rodrigo G. C. Rocha[a]*, Ryan R. Azevedo[a], Ygor Cesar Sousa[a], Eduardo de A. Tavares[a], Silvio Meira[b]

[a]*Federal Rural University of Pernambuco, Garanhuns 55292270, Brazil*
[b]*Federal University of Pernambuco, Recife 50670901, Brazil*

Abstract

In addition to the benefits brought by the use of Distributed Software Development, new challenges linked to its use also emerged. Due to lack of information companies and organizations around the globe, independently, solve these challenges in many different ways, each with their practices, some more some less efficient, where best practices are hardly widespread among DDS community. In this context, this paper aims to present a web system based on Case Based Reasoning and Natural Language Processing to extract information in text form of problems and solutions adopted by distributed software projects and to recommend similar past experiences in order to support the decisions and resolutions of problems arising from new situations in distributed projects. The feasibility proof of the method was made from experimental tests conducted to identify the success of the recommendations of previous valid cases, with a success rate of 90% for the sample used.

Peer-review under responsibility of KES International.

Keywords: Distributed Software Development; Case-Based Reasoning, Natural Language Processing.

* Corresponding author. Tel.: 55-87-**3764-5505**.
E-mail address: rodrigo@uag.ufrpe.br

Peer-review under responsibility of KES International.
doi:10.1016/j.procs.2014.08.099

1. Introduction

As a reflex of the business globalization and the evolution of the software market, new Software Engineering approaches are conceived and so are new ways to compete and collaborate. As a result, the Distributed Software Development practice evolves. In a DSD project, the teams working on it are dispersed in different locations[1].

Besides the benefits obtained by using this approach, new challenges linked to its use also emerged. Rocha et al.[2] states that in the distributed scenery, software projects assume different perspectives and consequently new risks. If good knowledge about the entire project does not exist and neither already known factors that may influence it, chances of failure will be higher. Due to lack of knowledge, companies and organizations around the world, independently attempt to patch these challenges in a variety of ways, each with its own practices; some more efficient than others, but even the efficient ones are hardly ever shared in the DDS field.

In this context, this work presents a method expressed in a web tool that uses the concept of Case-Based Reasoning and Natural Language Processing for extracting data from texts about problems and solutions adopted by companies and organizations in distributed projects. It recommends similar past experiences with the purpose of supporting decision making and solving problems raised from new situations in distributed projects. Thus, this tool recommends cases for those problems identified aiming to find solutions through the known cases by the system.

An experiment was necessary for the validation of the tool, which used 19 cases of problems and solutions retrieved from the event Distributed Software Development Workshop, from years 2007 to 2012, with a total of 51 papers analyzed. From the tests performance and the experiment, it is possible to state that the recommendation of previous valid cases was successful, with a 90% hit rate in the context of the sample used.

This work is organized as follows: Section 2 presents the background, in which the main concepts of the work are presented; Section 3 contains the proposed tool, its characteristics and features; in Section 4 the experiment used for the validation of the tool is discussed; Section 5 analyses related works, and finally, Section 6 comes with the final considerations

2. Background

This section presents the main concepts involved in this work, conceptualizing Distributed Software Development, Natural Language Processing Theory and Case-Based Reasoning.

2.1. Distributed software development

DSD is a software development model in which people involved with a certain project are in different locations[3]. According to Prikladnicki et al.[4], economy expansion, sophistication of communication means and cost pressure have encouraged the investment on DSD. The main advantages of this environment are: decrease in software development costs, access to human resources from other locations, and attending the needs of global clients[5].

However, there exists a series of challenges inherent to this context: physical distance, time zones and cultural differences. Carmel and Agarwal[6] evaluate how physical distance contributes to the complexity of the project developed under this kind of condition and define three main obstacles that DSD must overcome: the coordination, the information control and communication.

2.2. Natural language processing

NLP is a field of Artificial Intelligence and linguistics that consists in the development of computational models for performing tasks that depend on information expressed in a natural language[7]. In computing, the goal in using this concept is the evolution in the way in which the computer is used. Most of human knowledge is stored as natural language; computers able to understand natural language will be able to use this data with optimal speed and for a wide range of purposes[8]. Liddy[9] states that NLP is a set of techniques whose objective is to analyze and represent in a natural way texts that occur in one or more levels of linguistic analysis, with the purpose of reaching a language processing level similar to that of humans in different kinds of activities and applications.