The War for Talent: Identifying competences in IT Professionals through semantics

Ricardo Colomo-Palacios, Universidad Carlos III de Madrid, Spain
Marcos Ruano-Mayoral, EgeoIT, Spain
Pedro Soto-Acosta, Universidad de Murcia, Spain
Ángel García-Crespo, Universidad Carlos III de Madrid, Spain

ABSTRACT
In current organizations, the importance of knowledge and competence is unquestionable. In Information Technology (IT) companies, which are, by definition, knowledge intensive, this importance is even more crucial. In such organizations, the models of knowledge exploitation include specific processes and elements that drive the production of knowledge aimed at satisfying organizational objectives. However, competence evidence recollection is a highly intensive time consuming task and this circumstance can be seen as the key point for our system. SeCEC-IT is a tool that based on software artifacts extracts relevant information using natural language processing techniques and enables competence evidence detection by deducing competence facts from documents in an automated way. SeCEC-IT includes in its technological picture semantic technologies, natural language processing and human resource communication standards (HR-XML).

Keywords: Semantic Technology; Knowledge Management; Competence Evidences; IT Professionals

INTRODUCTION
The use of IT solutions has become a key issue in many organizations worldwide. Organizations currently use multiple IT/IS solutions to support their activities at all management levels (Trigo, Varajao & Barroso, 2009). Software costs as a percentage of total computer system costs continue to increase; while associated hardware costs continue to decrease (Huang & Lo, 2006). Software development is a collaborative and knowledge intensive process where success depends on the ability to create, share and integrate information (Walz et al., 1993), among other factors. Software development is an intense human capital activity, more intense in intellectual capital (Sommerville & Rodden, 1996). Although the importance of human factors has been widely recognized as key for software engineering, researchers should put a larger focus on the humans involved in software engineering than what has been done to date (Feldt et al., 2008). However, poor management of human factors in technical projects, and software engineering projects can be considered as technical projects, can hinder the use and effectiveness of technology (Ives & Olsen, 1984).

Individual differences have been identified as one of the paradigms for the research of human factors in software development (Curtis, 2002). IT workers professional practice must be continually revised and improved in order to adapt workers competences’ to technical innovations
and soft skills to evolving markets (Casado-Lumbreras et al., 2009). In this scenario, competence at the individual level is required for the creation of core competence, crucial for today’s organizations at the organizational level (Bassellier, Reich & Benbasat, 2001). But in spite of this importance, the world is facing an IT professionals shortage. Thus, attracting students in order to shape tomorrow’s labor horizon has become a major issue of concern in educational institutions (Garcia-Crespo et al., 2009). According to the analysis by Morello, Kyte and Gomolsky (2007), many young people see IT as an unattractive career option: it is both hard work and “uncool”. Additionally, this negative image is confirmed by the paradox that the strategic contribution of IT is recognized within enterprises, but the status of the IT department is low (Avison, Cuthbertson, & Powell, 1999). The shortage of IT professionals has been pointed out by many works (e.g. Acharya & Mahanty, 2008; Agarwal & Ferratt, 2002; Mithas & Krishnan, 2008; Wells & Bogumil, 2001). As a consequence of this, the war for talent (Michaels, Handfield-Jones & Axelrod, 2001) in the IT sector has its battlefield outside and inside the company and the internal recruitment of professionals must be done basing selection requirements in competence evidences. But in spite of the importance of competence evidences and knowledge sharing proficiencies pointed out by Liebowitz (2009), just a few companies have access to this data and promote their recollection throughout the year.

Given the need of the corporations around the world to get competence evidences in a trusted and automatic way SeCEC-IT is presented in this paper. SeCEC-IT is a tool that based on the work performed by IT professionals in the context of software engineering development projects, extracts relevant information from software artifacts (programs, documents,….) using natural language processing and enables competence evidence detection by deducing competence facts in an automated and semantic way. These competence facts can be transferred to common human resource management tools that can exploit this information using competency management interchange standards in order to be used to internal recruiting or to support knowledge management issues.

The remainder of the paper is organized as follows. Section 2 outlines relevant literature in the collection of competence evidences. Section 3 sums up main research efforts about semantic technologies. In Section 4, the architecture for the SeCEC-IT approach is presented along with the description of the implementation of the proof of the concept architecture. Finally, conclusions and future work are discussed in Section 5.

COLLECTING EVIDENCES OF COMPETENCE

Competences and competence management has proved to be an extremely relevant area of study including fields such as pedagogy, psychology or technology. The term "competence" has been applied in reference to many different domains of behavior (Waters & Sroufe, 1983). Anderson and Messick (1974) have catalogued 29 diverse referents ranging from specific skills (fine motor dexterity) to abstract concepts such as consolidation of identity.

According to McClelland, (1973) competency is comprehended as the relation between humans and work tasks, that is, the concern is not about knowledge and skills in itself, but what knowledge and skills are required to perform a specific job or task in an efficient way (McClelland, 1973). In a subsequent analysis of the term in the scientific literature, Draganidis & Mentzas (2006) state that a competency must be defined in terms of:

- Category. A group to which homogeneous and/or similar competencies belong.
- Competency. A descriptive name for the specific competency.
- Definition. Statement(s) that explains the basic concept of this competency.
Demonstrated behavior. Behavior indicators which an individual should demonstrate if the specified competency is possessed.

The competence approach was a major innovation in the human resource development field in the 1990s (Collin & Holden, 1997). McClelland (1987) suggested that competence ought to become the basis for more effectively predicting individual performance in organizations. Moreover, competences can be defined as features related to effective working performance (Boyatzis 1982). That could be the reason why, competence is often used in the sense of performance, however, this is not entirely accurate (Bassellier, Horner Reich, & Benbasat, 2001). Nonetheless, competence is a factor that, coupled with motivation, effort and supporting conditions, may have a direct impact on performance (Schambach, 1994).

In IT field there are many attempts to adopt and study the competence paradigm in various fields (e.g. Acuña & Juristo, 2004; Colomo-Palacios et al., 2010; Ruano-Mayoral et al., 2010; Trigo et al., 2010; Turley & Bieman, 1995). However, competence evidence recollection, in general, and in software development teams, in particular, has received reduced attention in both theory and practice. In the work of Ruano-Mayoral et al. (2007) an antecedent of the system presented in this paper is presented. Referred system is a mobile tool to recollect competence evidences, however, the collection of such evidences is made in a manual way. Taking this antecedent into account, the main purpose of this paper is to present a tool aimed to detect and classify competence evidences within software development projects using software artifacts in an automated and semantic way.

SEMANTICS: A NEW PARADIGM ENABLED BY TECHNOLOGY

The information contained in Web pages was originally designed to be human-readable. As the Web grows in both size and complexity, there is an increasing need for automating some of the time consuming tasks related to Web content processing and management.

In this scenario, semantic web can be seen as a vision for the future of the Web, where the unit of information is the data, instead of the web page, as in the traditional Web. Around that vision of a web of data, the W3C consortium has promoted the development of several technologies to describe resources by means of ontologies and rules. Semantic web represents a revolution in many senses. The term “Semantic Web” was coined by Berners-Lee, Hendler & Lassila (2001), to describe the evolution from a document-based web towards a new paradigm that includes data and information for computers to manipulate. Ontologies (Fensel, 2002) are the technological cornerstones of the Semantic Web, because they provide structured vocabularies that describe a formal specification of a shared conceptualization. Ontologies were developed in the field of Artificial Intelligence to facilitate knowledge sharing and reuse (Fensel et al., 2001). Ontologies provide a common vocabulary for a domain and define, with different levels of formality, the meaning of the terms and the relations between them. Knowledge in ontologies is mainly formalized using five kinds of components: classes, relations, functions, axioms and instances (Gruber, 1993). The theory which supports the use of ontologies is a formal theory within which not only definitions but also a supporting framework of axioms is included (Smith, 2003).

Taking full advantage of ontologies, the Semantic Web provides a complementary vision as a knowledge management environment (Warren, 2006) that, in many cases has expanded and replaced previous knowledge and information management archetypes (Davies, Lytras & Sheth, 2007). Thus, Semantic Web has emerged to be a new and highly promising context for knowledge and data engineering (Vossen, Lytras & Koudas, 2007). The goals of the Semantic Web initiative include the integration of data from different sources in a machine processable
format in order to make them accessible to computer programs and facilitating the use of data in ways that have not been thought of when the data was entered or recorded (Battré, 2008). It is agreed that semantic enrichment of resources would lead to better search results (Scheir, Lindstaedt & Ghidini, 2008). In this new scenario, the challenge for the next generation of the Social and Semantic Webs is to find the right match between what is put online and methods for doing useful reasoning with the data (Gruber, 2008).

There are several works that reflect the importance of semantic technologies and their impact in competence systems and models. Semantic technology has been applied for project management teams construction (Gómez-Berbís et al., 2008), knowledge management for software projects (Colomo-Palacios et al., 2008), technical competence assessment (Colomo-Palacios et al., 2010), knowledge sharing and reuse (Lanzenberger, 2008), assist the learning process (Naeve, Sicilia & Lytras, 2008; Collazos & García, 2007), competence development efforts (Dodero et al., 2007) or assist work assignment (Macris, Papadimitriou, & Vassilacopoulos, 2008) to cite some of the most recent initiatives.

SECEC-IT: ARCHITECTURE AND CASE STUDY

One of the key elements in the SeCEC-IT picture is capturing competence evidences and enabling internal and established Human Resource Management (HRM) solutions to use them. Due to this possibility, the reliability and precision of the competence evidences and their usability will be drastically increased. On the other hand, there is a need to develop a solution that could interconnect with a set of companies. The best tool for this purpose is the HR-XML standard.

The HR-XML Consortium is an independent, nonprofit organization dedicated to the development and promotion of a standard suite of XML specifications to enable e-business and the automation of human resources-related data exchanges. SIDES, one of the recommendations published by the HR-XML Consortium can be seen as a suite of data exchange standards for staffing issues. One of the multiple parts of SIDES is a competence schema designed to fulfill the following requirements (Allen, 2003):

- The competence schema is simple and sufficiently flexible and generalized so that it is useful within a variety of business contexts.
- The schema provides structure to enable competences to be easily compared, ranked, and evaluated.
- The schema is capable of referencing competence taxonomies from which competence descriptions were taken or used.
- The competence schema is relatively simple and compact so that it does not add to the complexity of the process-specific schemas within which it is used.

For the purpose of our work, the competence schema allows the integration with other Human Resources Management Systems but, to achieve the full capacity of competence analysis that this framework seeks, it is necessary to build an extension of the competence schema to store some extra information about each competence evidence. The extension was used by authors in the past and can be found in Ruano-Mayoral et al. (2007).

In what follows an explanation of SeCEC-IT will be given showing its architecture and a use case.

ARCHITECTURE

The architecture of SeCEC-IT is based on component groups that interact among themselves, to offer an automatic a solution to the problem proposed. The conjunction of these systems
permits the correct operation of the whole set of components, and the obtention of the necessary data to achieve the desired outcome. Since interoperability is one of the most challenging problems in modern cross-organizational information systems (Mocan et al., 2009), much emphasis is put on interoperability issues, done via web services. The final architectural approach is a tailor-made value-added technological solution. Components might be related to the behavior as specified in the collaboration among those elements, turning those structural and behavioral elements into progressively larger subsystems and the architectural style that guides this organization. Figure 1 shows how these different subsystems communicate and the flow of exchanged messages, in order to process the final system response. In the following the internal working of every element will be detailed.

As mentioned above, the architecture is comprised of three operating layers or subsystems. Firstly, the interface layer is composed by a number of interfaces through which end systems can interact with SeCEC-IT. Secondly, the logic layer encompasses the reasoning, inference and business logic management functionalities. Finally, the persistence and storage layer is composed by semantic repositories, storing the competence evidence ontology. In the following, we will detail several of the core components in each layer.

- INTERFACE. This layer presents two components, namely HRM Interface and Repository crawler.
HRM Interface. This component allows the communication with external HRM solutions using a web service. Three different kinds of data is exchanged:

- Competence descriptions included in HRM Solutions. These descriptions will feed the crawler that will seek relevant information relative to these descriptions in the set of software artifacts.
- Instructions of how to locate and reach project repositories (URL).
- HR-XML formats containing information elaborated by the system that is transmitted to HRM external systems in response of a given query.

Repository Crawler. This component, given a project repository, crawls documents and sends them to the NLP engine in search of competence evidences.

LOGIC. This layer provides cutting-edge functionalities through the following components:

- NLP engine. Given a (set of) document(s) in a project repository and a set of human resources, NLP seeks for relevant competence evidences, such as participation of a programmer in a requisite extraction process. In this module several well known tools are implemented, including GATE to syntactic annotate noun phrases and JAPE to extract all phrases related to competence evidences. Once a competence evidence is found, Competence Engine will be responsible of its classification and storage.

- HR-XML Engine. This component constructs a valid HR-XML document from a query by reading data in the persistence layer and returns this document to the interface layer in order to be delivered to the external system.

- Competence Engine. Is responsible of dealing with competence ontology and stores information of competence evidences in the persistence layer. It hides competence complexity to other components of the system.

PERSISTENCE. Finally, the persistence layer stores the knowledge about the competence evidences. On the one hand, the Competence ontology defines the relevant characteristics of each competence. All this information is used to describe competences suitable for our system (technical competence). This ontology has been defined using the Ontology Web Language (OWL) (Bechhofer et al., 2004). The storage and ontology reasoning has been developed based on the Jena framework. On the other hand, the competence evidences and their location relative to their project repositories are also stored into a database. Both the competence ontology schema and its populated instances are stored in the KAON2 ontology repository. KAON2 is an infrastructure for managing OWL-DL ontologies. In the case of SeCEC-IT, Jena is the backbone technology that relies on a MySQL database. Jena is a framework for building Semantic Web applications that provides a programmatic environment for RDF, RDFS and OWL, SPARQL and includes a rule-based inference engine.

About implementation and internals, SeCEC-IT is a Web based application build under Java EE (by using Java Enterprise Edition 5 SDK). Business logic design was done using MagicDraw.
This tool enables Model Driven Architecture (MDA) architecture and automatic code generation by using AndroMDA.

The Java-based tool RACER is also implemented. RACER reasoning engine and the Jena framework are crucial for the business logic manager layer, the former for the reasoning and the latter for the RDF Management and SPARQL Querying. Lastly, JAXB is used for XML handle (in order to communicate with others) and JENA 2 for ontology information issues.

**USE CASE**

To explain the realization of SeCEC-IT in a functional environment, as referred before, a use case will be included. The software development company SEMDEV would like to implant a new knowledge management and competence management program. The final aim of this program is to assign people to projects basing these decisions in resource availability and competence. Now managers perform project staffing just using availability and using informal information, but company owners want to implement a more scientific approach that allows to assign personnel closing the gap between competence and project role.

SEMDEV sends to SeCEC-IT the information needed to start the crawling process. Firstly, the access to current software development projects repositories (giving an external granted access to a PDF and code repository), on the other hand, the set of competences and human resources aimed for SEMDEV. These competences are adapted to SeCEC-IT competence ontology and the crawling process starts. The Repository Crawler component looks for relevant information in the repository and sends relevant information to the NLP Engine. This component extracts relevant competence evidences from documents and sends this information to the Competence Engine. The Competence Engine populates the competence ontology by creating a number of instances for given competences and human resources.

For example, the analysis of a number of software artifacts may imply that the resource RCP is competent in Requirements Engineering, but his results in Software Testing are low according to several comments in their proofs. The Competence Engine will store this information in the Persistence Layer that will be stored in the KAON2 ontology repository. Once all documents are crawled, the system regularly performs an update check in order to find out if there are new versions of documents and, thus, new competence evidences.

Later on, SEMDEV implants a HRM tool (such as Meta4 PeopleNet) to support competence development and management as well as project staffing based on competences. In this new scenario, SEMDEV human resource administrators can ask SeCEC-IT for competence evidences. Let’s imagine that more information about competence evidences are needed for RCP. In this case, PeopleNet could ask to SeCEC-IT for this information using the HRM Interface. Once this order is received, the HR-XML engine extracts this information from the Ontology Repository using SPARQL and forms a correct HR-XML format in order to be sent to PeopleNet using the HRM Interface.

**CONCLUSIONS AND FUTURE WORK**

The advent of the information age represents both a challenge and an opportunity for knowledge and competence management. New forms of knowledge extraction and expert location are deeply impacting companies around the world. IT companies are facing a war for talent in which, every project must be scheduled according to the availability of resources and their competences. In this new scenario, counting on with tools to seek competency in work
environment can enable a better personnel management that could be a competitive advantage for the company.

SeCEC-IT, following the path of some previous works (Colomo-Palacios et al., 2008; Colomo-Palacios et al., 2010; García-Crespo et al., 2009; Ruano-Mayoral et al., 2007) brings new features to competence management in software development projects: the transformation of plain text to competence evidences in an automatic way. This competence evidences could be used, by means of the interfaces implemented, for staffing and teaming purposes or as a support to performance appraisal.

But SeCEC-IT is not only relevant for IT professionals and managers. Counting on a tool that can derive competency evidences directly from software artifacts could be a competitive advantage for the firm. Moreover, HRM personnel could also benefit from this. Having in mind that sometimes IT workers perform a highly technical work, it’s very difficult for HRM departments to infer competency from their work without the assistance of an IT manager a highly IT qualified individual. But, having in mind the lack of available time, it’s sometimes difficult for HRM personnel to know the competency levels of IT professionals apart from the yearly assessment. This information could be of great benefit for the corporation and its applications are multiple: improve the person-role fit, detect competency gaps, improve internal recruitment process and perk up professional development and career planning.

Taking into account the possibilities initiated by the current research effort, four separate lines of future research may be considered. In the first place, the integration of certain Web 2.0 contents as a source for competence evidences. In the second place, authors suggest to expand the possibilities of the system to deal with cultural, gender and performance differences. In the third place, integrate the tool into a wider program in which affective factors pointed out by Smith (2010) must be taken into account. Lastly, it is aimed to integrate SeCEC-IT in Computer Aided Software Engineering tools and, in particular, in effort and duration estimation tools for software development projects.

ACKNOWLEDGEMENTS

This work is supported by the Centre for the Development of Industrial Technology (CDTI) (Ministry of Science and Innovation) under Sem-IDI project IDI-20091150.

REFERENCES


Bassellier, G., Reich, B.H. & Benbasat, I., 2001. IT Competence of Business Managers: A


