THESIS SUMMARY

to the Ph.D. dissertation by

Ildikó Borbásné Szabó

Design of Higher Education Portfolio

Supervisor:

András Gábor Ph.D.
associate professor

Budapest, 2012
Department of Information Systems

THESIS SUMMARY
to the Ph.D. dissertation by

Ildikó Borbásné Szabó
Design of Higher Education Portfolio

Supervisor:

Gábor András Ph.D.
associate professor

© Ildikó Borbásné Szabó
Table of contents
I. The relevancy of this topic ........................................................................................................4
   I.1 Related projects ..................................................................................................................4
II. Research methodology ..........................................................................................................6
III. The results of the thesis .....................................................................................................7
IV. Bibliography .......................................................................................................................16
V. Publication related to the topic ..............................................................................................20
I. The relevancy of this topic

One assumption of sustaining competitive edge of an economy is to transfer and usage of relevant knowledge to the right place and in the right time. It means to overcome the barriers to mobility and to create a competitive education area. The financial crisis emphasized the importance of higher education in qualifying students in the light of the needs of labor market and „creating and maintaining a broad, advanced knowledge base and stimulating research and innovation” (BolognaProc, 2009).

The Hungarian government recognized that there is a mismatch between the learning outcomes of higher education qualifications and the competences required by companies. Therefore, one goal of the Hungarian higher education reform taking place nowadays is to rationalize qualification obtained in higher education in the light of requirements of the world of labor. (SZKTerv, 2012 pp.47)

So my research focus was to develop an information system that is capable of evaluating the HE qualifications in regard of missing and surplus competences needed by companies.

I.1 Related projects

Two EU-funded projects – OntoHR\(^1\) and SAKE\(^2\)– have already dealt with this problem. The SAKE project aimed at „specifying, developing and deploying a holistic framework and supporting tools for an agile knowledge-based e-government that will be sufficiently flexible to adapt to changing and diverse environments and needs“(SAKE, 2006). In 2008 MEC pilot as a Hungarian case has already showed that competencies had constituted an appropriate base to compare the requirements of labor market (demand side) with the learning outcomes of a certain qualification (supply side of labor market). Having transformed the degree programs based on the requirements of demand side, a viable higher education portfolio can be created (Kovács, 2010).

The European Commission’s OntoHR project aimed “to create more specific qualification to job matching, with the overriding purpose of tackling the conversion of vocational education qualifications into job related competencies“ (Kismihók-Mol, 2011) (Mol et al., 2012).

---

1 SAKE - Semantic-enabled Agile Knowledge-based eGovernment (IST 027128) is a research co-financed from the 6th Framework Programme for Research and Technological Development. It run from 2006 to 2009. (Source: www.sake-project.org).

2 Az OntoHR (Ontology Based Competency Matching between the Vocational Education and the WoRkplace, 504151-LLP-1-2009-1-HU-LEONARDO-LMP) is a Leonardo Da Vinci project. It run from 2009 to 2011 (Source: www.onto.hr).
These projects aimed to identify the shortcomings of higher and vocational education learning outcome through matching a Job Role Ontology (JRO) based on competences retrieved from job role descriptions and a Learning Outcome Ontology (LOO) based on competences claimed and/or extracted from descriptions of a given training program.

SAKE project concerned several ICT job profiles and Business Informatics degree program, whilst the goal of OntoHR was to build an ontology-based selection and training system based on Information System Analyst (ISA) job role. One module of this system dealt with the evaluation of the ICT degree programs. They didn’t execute the matching process dynamically. In these projects, the JRO ontology reflected only a static moment of requirements of the labor market. In SAKE project, the job advertising documents were downloaded and tagged manually. In OntoHR, the ontology elements were extracted from the detailed descriptions of ISA job profile given by public organizations (e.g., O*Net) or by projects concerned job analysis (e.g., EUQuaSIT). The current research aimed at formalizing the job requirements derived from IT/Telecommunication category of a popular job recruitment portal (Profession.hu) into the Job Role Ontology and matching this ontology to the Learning Outcome Ontology, which is created by the learning outcomes, and materials of Business Informatics Bachelor’s degree program (Szabó, 2011).

In these projects ontology-based approach provided a framework for the matching. The network of DEHEMS project aimed „to study determinants of graduate’s early career success in selected professional domains and fields of study“. A knowledge platform was created by the consortium to optimalize quality assurance and performance of HE institutions. Having analyzed related disciplines and data, the needs of the labor market related to the learning outcome of qualifications were examined by a series of interviews executed in six countries. (e.g. the requirement for balancing the scientific and practical skills in the degree programs) (Melink-Pavlin, 2012).

The European Centre for the Development of Vocational Training (CEDEFOP) elaborated a methodology to discover the skill discrepancies between both sides of the labor market using forecasting models (Cedefop, 2012).

Primarily my research focus was to develop an information system to examine the alignment of the qualifications to the actual requirements provided by the labor market. But in a later phase, I would like to extend this system with forecasting models (for example to calculate the student’s number of a given HE qualification). Both sides of the labor market have a special viewpoint in this domain which can be handled by the ontology-based approach well. So the research had been built on the results of the above mentioned projects (OntoHR and SAKE), but it handled dynamically the needs of labor market as against these projects. The implementation of all degree programs needs too
much resource, so the output of this research is a prototype system, which dynamically examines in what measure the learning outcomes of Business Informatics Bachelor’s degree program at Corvinus University of Budapest are matched to the requirements given by the Software developer job role (job requirements). The system can be customized for optional qualifications and job roles within its boundaries.

The dynamism is provided by collecting these requirements per period, using them for executing the matching procedure and evaluating the results.

My research aimed to answer the next questions:

- What kinds of documents and related processing methods facilitate the dynamic management of needs?
- Is the concept of competence really suitable for comparing the two sides?
- Is the ontology approach an appropriate methodology to present the job requirements and the structure of qualifications and to execute the matching procedure in dynamic manner?
- How the results of the matching process can be used in decision preparation processes related to the higher education?

My hypotheses corresponded to the system development phases outlined by these questions:

- Ontology development: collecting the documents and sources following the changes of the labor market, formalizing them into ontologies
- Ontology matching: examining the alignment of the two ontologies.
- System integration: evaluating the results, examining the adaptability of this system in decision making processes.

**II. Research methodology**

The goal of this research was to develop a system which is capable of mapping the similarity and discrepancies between the two sides of labor market in regard of qualification. It didn’t want to discover rules and connexions that explain interferences among people, so it doesn’t fit into the world of social science research (Babbie, 2001 pp.4).

Instead, it aims to develop a system to prove the correctness of the approach. This way, the research follows the two-phased incremental software development method (Sommerville, 2007 pp. 73), where the hypotheses represented the functional requirements of the increment in every certain
stage. Correctness of the hypotheses is investigated based on the output of the connected system development phase.

The adaptability of deductive or inductive thinking method (Babbie, 1996), and quantitative or qualitative methodology (Balaton-Dobák, 1991) in the research were investigated in regard of the hypotheses.

The applied models in the research had to mirror reality and its connexions, so qualitative techniques (Denzin-Lincoln, 2000) seemed to be applicable to build these models. The research used ontology-based approach to unify the viewpoints appeared in the two sides of labor market.

The deductive thinking method facilitates to verify the connexions derived from the adaptability of the prototype. But the system integration into the decision making process can be achieved with the help of inductive thinking method

III. The results of the thesis

On the demand side of labor market, the importance of this concept was showed by the advantages of switching from job-based to competency-based organizational approach (Lawler, 1994), by its strategic importance presented by Schoonover and Andersen (Schoonover et al., 2000), and by the role of updating competency models and job descriptions in talent specific succession planning (Egodigwe, 2006).


So, the competence seems to be an appropriate base to achieve the comparison between the two sides of the labor market. (In English the competence and competency concept are distinguished. The first interpretation focuses on the background of this concept, rather than the last one (Rowe, 1995). So following the guideline of Hungarian public education too, the thesis uses that interpretation (Tót, 2009). But sometimes the competency word describes its elements too.)

The competence includes „knowledge, skills and abilities” in the Act CXXXIX of 2005 on Higher Education. OECD considers this concept wider than just knowledge and skills. „It involves the ability to meet complex demands, by drawing on and mobilizing psychosocial resources (including skills and attitudes) in a particular context” (DeSeCo, 2005:4).

In the world of labor the competences are always a close relationship with an activity, which requires them. The concept means „a specific, identifiable, definable, and measurable knowledge, skill, ability and/or other deployment-related characteristic (e.g. attitude, behavior, physical
ability) which a human resource may possess and which is necessary for, or material to, the performance of an activity within a specific business context.” (Allen, 2001:3).

In summary, the competence includes knowledge, abilities and/or skills, behavioral and motivational attitudes. So the thesis uses knowledge, skill and attitude triple to describe this concept.

The competence elements must be placed in its context. So they must be connected to activities, tasks, functions or roles that require them.

The organizations use competence models to connect job requirements to the personal competence lists (Henczi-Zöllei, 2007 pp.81). General models – like HLD model (Henczi-Zöllei, 2007 pp.91-92.), ETA model (Ennis, 2008) –, and management specific models too – like Spencer developed competence dictionary (Bakacsi et al., 2005 pp. 144-145.), the framework elaborated by Management Charter Initiative (MCI)3 (Barker, 1993) were examined. In the process of analysis the next questions were investigated to evaluate the usability of these models in the research:

- How does the model describe the competence elements?
- How does the model present their context of use?
- How does the model draw the relationships between the competence concepts (e.g. one competence is a premise of other competences)?
- Is there any importance of the model on the education side?

Among the four models, MCI and ETA has performed the best, however MCI can only be applied in the case of IT managers, while in the case of the latter the problem is the context-dependent buildup of competences instead of a structure based on competences being preconditions of other competences.

It can be seen therefore that such a model is needed that can draft the content and application environment of competences in a general way and gets updated according to a certain job afterwards. Ontology-based approach contributed largely to the elaboration of this model.

In the ontology development process, we apply a discipline to describe “a theory of what entities can exist in the mind of a knowledgeable” (Van Heijst et al., 1997 pp. 191). So it provides a unified view where “the objects, concepts, and other entities that are assumed to exist in some area of interest” are mapped and specified (Guarino et al., 2009 pp. 3). (We have to distinguish the taxonomy from ontology, because the ontology concept includes the description of concepts and objects too (Sántáné-Tóth, 2006)).

---

3 An organization who is responsible for the elaboration of NVQ qualifications.
However, even when the higher education training programs are constructed with the agreement of the employment and education side (as an initiative aiming this was seen in the case of the National Vocational Qualification), originally formulated learning outcomes can change during the course of the training program – including the time needed for announcing the program and matriculation. When this situation occurs, the reconciliation between the two parties is not common at all. It can occur this way that the same competence is interpreted on the two sides in different aspects, in different context and by emphasizing different essential elements. This way, in order to realize the ontology-based comparison formulated as the goal of this research, mapping or formal specification of competence content available on both sides into two appropriate ontological structures was needed, then performing the investigation of fitting these ontologies to each other was also required. Afterwards, the analysis of the fitting of the elaborated system into the decision preparation process affecting current state of higher education was performed.

The Business Informatics degree program and the Software developer job role were chosen to underlie the prototype. This degree program runs at our department, so we can change its learning outcomes, if it is necessary. The developer role was needed in 25% of the job advertisements appeared in October 2009 and October 2012. It was a huge need, so only 13 roles appeared in 90% of the job advertisements. Although this degree program qualifies not developers primarily, but this role is enough needed to underlie the prototype.

The goal of the system is to process documents that follow the changes of competences on each side of the labor market. On the education side, the descriptions about the Business Informatics Bachelor’s Degree Program and the experiences of lecturers constituted the source of the ontology. On the occupational side, the job offers downloaded from Internet served as a basis of the ontology.

The functional requirements related to the system were the next ones:

- collecting job requirements from the Internet in an automatic manner, extracting knowledge elements of them and forming these elements into the Job Role Ontology in a self-automatic manner; formalizing the actual status into the Learning Outcome Ontology;
- achieving the matching process between Competence classes or its subclasses of both ontology and evaluating the results.

These requirements delineated into two development phases, so the incremental system development methodology was usable (Sommerville, 2007 pp. 73).

My research hypotheses equaled to the requirements of the development phases, so after evaluating the output of the phase, they were accepted or rejected.

The H.1 hypothesis stated that the system must reflect the structure of qualification and the occupation geographically and periodically. So the meta-models, as the outputs of this phase, had
to include the competence elements as the base of the comparison and to mirror the contexts of the competences on each side. So the meta-model of Job Role Ontology had to represent the world of organization, the meta-model of Learning Outcome Ontology had to reflect the knowledge flow on the education side.

The Competency Ontology (Draganidis-Mentzas, 2006 pp.4), the HR ontology (Mochol et al., 2004 pp.4), the Higher Education Ontology (Kő-Futó, 2008) served as a basis of creating the first meta-model of JRO ontology. The LOO ontology model was created as an extended version of Education Ontology Model (Vas, 2007) (Kismihók et al., 2012). The meta-model of the ontologies is presented by Figure 1.

In the meta-model of Job Role Ontology, the Industry, Organization, Department and Position classes put the competences into an organizational context. Within an organization, the business processes consist of tasks that roles and responsibilities belong to. In the backward direction, the Role as the parts of the Position class determines the entities of the Task and the Responsibility class. Competence(s) are required to execute a Task. The attitudes of the Competence class facilitate to execute the comparison at the appropriate Level, in the Period of Validity and in a given Region.
The meta-model of the Learning Outcome Ontology is an extended version of the OntoHR project’s Educational Ontology by the Description of the Degree Program, as sources of the competences, and by the attitudes of the Competence class.

The elements of competences mentioned in Section 2 (like Skill, Attitude and Knowledge) represent the basis of the comparison, but in the prototype we used only the Knowledge class as we will see in the next section.

Four ontology development tools – PowerLOOM\(^4\), OntoStudio\(^5\), Protégé 3.x/4.x\(^6\) and KAON2\(^7\) – were studied to implement the above mentioned meta-models. These tools and its predecessors were presented in thesis of Vas (2007). They were examined by the next criteria:

- The general features included the name of development organization, version number and functions of the tool, and the ontology languages used by the tool.

Automatic data import were investigated by

- open source code: this ensures the usability of the tool in long term period.
- documentation: this facilitates to transform and improve the tool.
- supported import file formats: what document format can be handled by the tool. The source documents may be transformed by using document processing tools (like gawk etc.) into the required format (e.g. XML).

The process of ontology matching was studied by next features:

- modularity, extendability: this feature is necessary for improving the original source code.
- built-in reasoner: how can this reasoner run a new algorithm?
- existence of a built-in ontology matching procedure.

Although Ontostudio has built-in ontology matching procedures and it is extendable by plugins, the non available source code and the dependency on license make difficult to customize this tool – e.g. for handling the Hungarian language or importing the downloaded advertisements in automatic manner – in change management.

The open source programs are extendable or improved by a Java program or a plugin-in, to execute matching procedures or import data automatically. Protégé 3.x and 4.x have much functionality, but 4.X can handle OWL 2, so I chose it and its knowledge representation method to implement the

\(^4\) http://www.isi.edu/isd/LOOM/PowerLoom/
\(^5\) http://www.softpedia.com/get/Science-CAD/OntoStudio.shtml
\(^6\) http://protege.stanford.edu
\(^7\) http://kaon2.semanticweb.org/
meta-models. For the creation of domain ontologies, new classes were added to the meta-models by extracting new knowledge from documents and other sources represented each side of the labor market.

The learning outcomes of the above-mentioned degree program have not been changed since 2005 and they are written too general in related documents, so the **structure of qualification** was created by the experiences of lecturers. The improvement of skills is not so stressful in this degree program, so we rather focused on Knowledge area class in the process of extending the meta-model of Learning Outcome Ontology. This competence element is measurable better than the other ones.

The organizations need to prepare themselves to adapt successfully to a rapidly changing environment. The job roles may be shaped due these changes. The online job advertisements are easy accessible and contain the descriptions of the job roles, so they were used to extend the Job Role Ontology meta-model with new classes in the process of creating the **structure of occupation**. They didn’t constitute a representative sample, but two researches (HRToborzás, 2011) (AllasTrend, 2009) showed that the organizations preferred Internet for advertisement. So a Java crawler was used to download job offers appeared on Profession.hu as a popular job recruitment portal in IT/Telecommunication category (IT Development/Programming category from September 2012).

The files downloaded from 27th December 2010 to 10th April 2011 served as basis to build the first version of JRO ontology. The XHTML format of files and the customized advertisements caused some trouble (Szabó, 2011). Having resolved these problems, just 10,76% of the downloaded files remained useful. Document processing tools (DOS Batch files, gawk and gsar scripts, JAVA programs) were used to transform their relevant contents into an appropriate format.

The Task class played an important role in the construction of the Job Role Ontology. But too many positions and related tasks appeared in the job advertisement collection, so we had to choose a position (like Software Developer position), its roles (Developer role) and related tasks (Designing the software development process, Preparing specification, Program coding, Program testing, Bug fixing and Communicating) to create the first version of the ontology.

The pattern-based ontology learning technique (Zhou, 2007) (Haase-Völker, 2008) seemed to be useful to extend this version with new classes, because classes from simple text files were extracted. A dictionary based on these patterns was created, where typical verbs or open queries served as expressions (for example the question was (design of, what)) and the expression was the **design of embedded software**. The ontology creation process using this dictionary has already written in (Szabó, 2012). Having executed this process, 18,6% of the tasks appeared in the job offer collection.
were processed. Having added new elements to the dictionary (e.g. new verbs, or new queries), its 33% were processed. The expressions described 91% of new classes well, but the algorithm put 27.6% of new classes into wrong classes. In the future, new queries, not new verbs, related to the appropriate class can improve the hit rate.

But the review of an expert was needed to verify and refine the ontology built on this way.

The advertisers describe the required competences slightly in many cases. Having cleaned the ontology by an expert, the competence elements (mainly the knowledge elements) of The Open Group Architecture Framework (OpenGroup, 2009:694-699) were assigned to the appropriate Task subclasses in manual or semi-automatic manner. The TOGAF is an industry standard, so it reflects the requirements of the organizations for Enterprise Architecture job role including the Software Developer role too.

The program is extendable by an algorithm, which can examine in what measure the TOGAF knowledge elements cover the actual requirements appeared in the job offers (for example Generic knowledge in Unix / Linux, AIX or Windows).

Having created the first version of Job Role Ontology and Learning Outcome Ontology, the instances of their classes were created. As it was mentioned earlier, the matching process was executed through the Knowledge class. The instances of this class were represented by the level, region and period of validity attributes. The instances of LOO ontology had Budapest as Region attribute, BSc as Level attribute and 2012 as Period of validity attribute. In October 2009 and October 2012, 74% of the advertised jobs were required in Budapest. Considering the willingness of people to move, these attributes characterized the underlying degree program of the prototype well.

A collection of job advertisements downloaded in October 2012 served as a basis to create instances of Competence class of JRO ontology. The algorithm extracted tasks from this collection, and searched the related Task class in the ontology using a version of the above mentioned dictionary. This algorithm processed 49.75% of the tasks advertised for bachelor students. The expressions of dictionary identified the 75% of these tasks well. This led to create the instances of 7 knowledge area and 4 skills. The information appeared in the job offers about region, level, and period of validity were set as attributes of these instances.

The H.2 hypothesis aimed to find an appropriate information technology tool to compare structure of qualification with occupational structure. So the outputs of the previous phase constituted the input of this stage. It was necessary to integrate the output of Protégé 4.1 ontology development environment with the product of this phase.
The ontology mapping deals with the matching of ontologies (Su, 2002)(Ehrig-Sure, 2004) (Kalfoglou-Schorlemmer, 2003). Alasoud et al. (2008) define ontology matching problem as follows: “given ontologies O1 and O2, each describing a collection of discrete entities such as classes, properties, individuals, etc., we want to identify semantic correspondences between the components of these entities.”

In the first version of the prototype, the comparison between both sides was executed through the Knowledge class, because the knowledge elements can be measured and can be assigned to the tasks more unambiguously than the other elements. This research concerned on finding the semantic and/or structural correspondences between the instances of the Knowledge class of both ontologies. So ontology matching systems proposed by Choi (2006) (Glue, Mafra, Lom, Qom, Onion, Omen) and offered by Noy (2000) (Prompt, IF-Map) were investigated. As Protégé 4.x uses the OWL Diff plugin and Compare Ontologies function to achieve matching, so I examined them instead of Prompt. The following features served as a basis to analyze:

- ontology matching is achieved in dynamic manner:
  - automatic, semi-automatic or non automatic working: the level of human intervention
  - the handling of changes occurred in the ontology
- reusability:
  - usage of different ontology format in matching process
  - type of matching method
  - support for modularity, integration with other systems
  - adaptability in Hungarian language environment.

Based on these characteristics, MAFRA (Maedche, 2002) and OWL Diff (OWLDiff, 2008) (or Compare Ontologies as its built-in version into Protégé 4.2) ontology matching tools seemed to be most suitable to achieve matching process. They are free downloadable, to execute from command prompt or a Java program automatically, to support RDF(S) or OWL languages and to handle changes occurred in the ontology through the usage of a semantic bridge or Protégé ontology editor. These are the most advantages of these programs compared to the others. However, they need human intervention as against IF-MAP. Unfortunately, MAFRA Toolkit hasn’t been supported since 2009, so I opted for OWL Diff and Compare Ontologies application.

Before executing the ontology matching procedure, I had to create two new versions of both ontologies. So I had to unify their name space and leave only Knowledge class into them. Both OWL Diff plugin and Compare Ontologies application assigned the instances with same name perfectly, but they weren’t capable of handling synonyms and set operations between attributes
with the same name. But these are open source tools, so their algorithm can be improved by any other algorithm of the above-mentioned tools (e.g. LOM and QOM). OWL Diff presented more transparent report than the other tool.

**In summary,** the developed system satisfied the requirements related to the system development phases, so the hypotheses were accepted. But the system has boundaries that may be eliminated in the future. The strengths and weaknesses of this system are summarized in the next table.

<table>
<thead>
<tr>
<th>The strengths of the system</th>
<th>Development ways</th>
</tr>
</thead>
<tbody>
<tr>
<td>The meta-models reflect reality and contexts.</td>
<td></td>
</tr>
<tr>
<td>The expressions in the dictionary describe the ontology elements in high precision manner.</td>
<td></td>
</tr>
<tr>
<td>The matching tool assigns the ontology elements with same name and attributes with each other in 100%.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>The weaknesses of the system</th>
<th>Development ways</th>
</tr>
</thead>
<tbody>
<tr>
<td>Just one part of job offers is usable due to their semi-structured format (e.g. XHTML format).</td>
<td>To stand for filtering more relevant text blocks (not only Task: and Requirement: blocks)</td>
</tr>
<tr>
<td>Less than 50% of the tasks can be processed in the ontology development.</td>
<td>To extend the dictionary with new expressions To eliminate the tasks related to not the given role.</td>
</tr>
<tr>
<td>The algorithm places tasks into wrong classes.</td>
<td>To extend the dictionary with open queries instead of verbs.</td>
</tr>
<tr>
<td>The matching tool doesn’t handle synonyms generally and set operations between the attributes with same name.</td>
<td>To improve its algorithm.</td>
</tr>
<tr>
<td>The whole process stops, if one application (e.g. OWL Diff) requires different input format as provided by the other application (e.g. Protégé API).</td>
<td>An algorithm is needed to convert these formats into each other.</td>
</tr>
</tbody>
</table>

Table 1. The strengths and weaknesses of the system

Among these boundaries the system is usable and customized by other Profession.hu job offers and higher education training programs. So this prototype proved the thesis given by András Gábor, i.e. the qualifications can be verified in the function of job requirements using competence based approach and information technology tools.

To put forecasting models into the system may provide more information to underlie the decision making process in higher education. Finally, a competitive higher education portfolio may be created.
The usage of this system in the actual decision making process (H.3 hypothesis) is not verifiable in nowadays, due to the use of different approach in the education policy. But it will be tested in SMART project ((Skill MAthching for Regional development) which follows competence approach in comparing structure of qualification with occupational structure (Anon, 2012 pp.24).

IV. Bibliography


Borbásné Szabó, I. A felsőokatási képzések munkaerői piaci szempontból való vizsgálata. Vezetéstudomány.(in press)


V. Publication related to the topic

**HUNGARIAN JOURNAL ARTICLES**


**INTERNATIONAL JOURNAL ARTICLES**


**INTERNATIONAL CONFERENCE PAPERS WITH PROCEEDINGS**


