Bottom-up Ontology for IT Skills

Pedro Miguel Nunes Cunha
Instituto Superior Técnico
Lisboa, Portugal
pcunha288@gmail.com

ABSTRACT

The recruitment process carried out by companies specialized for the purpose has been automated over the years to reduce time and costs while improving its accuracy. Approaches, such as ontologies, have been used to represent skills, position categories and other information necessary for the matching process between candidates and job positions. However, existing ontological developments do not take advantage of the data used by these companies daily, but rather, by the data available from other sources, which may not correctly represent the domain in which the company operates.

Using the Design Science Research Methodology to guide the work and the "Ontology Development 101" Methodology to guide the development, a bottom-up ontology was developed, with terms gathered from the database (DB) of a recruitment company. The focus was on skills and job categories in the IT field, on the skills hierarchy, on the relationships between skills and job categories, and the relations among skills themself. For an exhaustive evaluation process, four approaches were conducted: Reasoners, Competency Questions, Data-Driven and by Comparison with other IT Skills Ontology.

The results obtained through the analysis of the evaluation process demonstrate that the developed IT Skills Ontology is indeed a tool that proves to be useful in the recruitment process, more specifically, in the match between candidates with job categories.

Author Keywords

Ontology; Bottom-up Ontology; IT Skills Ontology; Ontology Development; Ontology Evaluation; IT Recruitment Match

INTRODUCTION

A vast amount of people, are often looking for jobs. There may be several reasons for that, from finishing college to looking for new challenges. At the same time, companies are looking for persons to fill their needs. For this end, and to help both companies and people in the this selection and recruitment process, specialized companies where founded.

The process of selection and recruitment of candidates for job positions, conducted by recruitment companies, is done less automatically than what they would possibly want, or could do [25], therefore taking more time and with higher costs. This work usually falls on human resources workers [34].

It is also important to highlight that since the industry is in constant change [23, 26], it becomes more difficult to be an expert and most of the time recruiters are not specialists in the field for which they are searching or evaluating [27]. For that reason, they usually follow traditional or ad-hoc techniques to conduct the "match" between candidates and positions, in which, if the candidate fulfills the requirements, then he or she should be contacted to move forward in the process of recruitment.

The recruitment company that drove this research specializes in the IT field, which means most of the job advertising (for job positions) and candidates they look up for are of the IT field.

The selection process that is carried out by this company can follow several paths. For this work we focused on the standard case, in which the candidate, through the platform, applies for a published job and there is a recruitment specialist that evaluates the possible match.

These recruitment specialists are someone that sits between the external and internal personal of the company, who possesses a more broad knowledge regarding IT skills (or at least this is expected). They have the function of evaluating the information from both the candidate and the requirements for the job, among other important aspects.

In general, the way the recruitment process, and more specifically the skills assessment between job advertisement requirements and the candidate's skills needs a unified mechanism. A mechanism that more easily identify what is associated with a skill or category in the IT field.

Therefore, there is a need for a solution that can help automate and conduct this process more precisely. By giving recruitment companies the means to do it, even when resorting to people not qualified for it and in addition to this, allowing the unambiguous identification and definition of each concept with which it is necessary to interact during the recruitment process, will help to simplify it. One solution to implement these changes, and that fits the requirements in need, are ontologies.

There is a vast amount of ontologies, in the most diverse fields, including in the IT field [5]. From those, and important for our research, are the ontologies that identify, qualify and categorize skills of the IT field. Although ontologies have been developed specifically for this purpose or even adapted from other main objectives, from the state-of-art conducted for this research, they lack being developed based on existing data (possibly in recruitment companies DB) but rather from other sources [12, 20]. This process of using empirical data for the development of ontologies is formally named bottom-up ontology development.

The Design Science Research Methodology (DSRM) methodology, was chosen to help guide this research. DSRM provides a process model to conduct research in the field of Information Systems [17]. The goal with DSRM, is to produce artifacts to serve as solutions for identified organizational problems. The method is an iterative cycle focused on the development of said artifacts, by getting feedback on the work developed, it allows the researcher to improve the artifact quality.

In this research, we aim to develop a bottom-up IT Skills ontology following the Ontology Development 101 methodology [22], composed of seven iterative steps. The skills will be categorized into IT categories, with the ontology terms coming from the content stored in the company's DB, specifically data referring to categories and skills used to describe the candidates and job advertisement requirements.

With the developed solution, we aim to serve both recruitment companies and candidates. Companies that will now have a method, possibly more effective than the ones they use, to conduct the skills assessment process (in the recruitment process). For candidates that with this solution can know which category fits best, based on their skills.

To verify the ontology quality five types of evaluation will be performed: Competency Questions, Reasoners, Data-Driven with Job Advertisement and Candidates data from the company DB, by Comparing it to another IT Skills Ontology and an overall assessment.

THEORETICAL BACKGROUND

We started by conducting a study on the theoretical background of skills, ontologies fundamentals, engineering, bottom-up development and ontologies in the domain of IT skills.

Skills

A skill in its core definition [1], is the ability to do something well, implying understanding or knowledge. The word may be ambiguous, since it is not clear either if it indicates mere adequacy or superior, extraordinary ability.

Some literature identifies a small semantic core, composed by the terms "ability", "aptitude", "capability", "competence", "effectiveness" and "skill" [36], defining competence in several fields, as a specialized system of an individual, with the necessary or sufficient requirements to achieve a specific objective.

Other, identifies many levels, at which competence can be defined [18]. For example, competence at a functional level is described as the things that a person who works in a given occupational field should be able to do and be able to demonstrate.

The HR-XML Consortium workgroup propose the following definition of competence: *"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 posses and which is necessary for, or material to, the performance of an activity within a specific business context."*.

Francis Green [13] also reported his understanding of the term skill and its similarity to other terms used by various fields to refer to the concept. He ends up defining three key points that a skill must have, thus creating the *PES concept of skill*: Productive - using a skill at work brings a productive value; Expandable - a skill is subject to training and development resulting in an improvement; Social- skills are socially determined.

Ontology Fundamentals

Our ontological study, started by learning about what an ontology is and why to use this type of technology as well as the most direct benefits of their usage.

Communication is fundamental for people, organizations and software systems [32]. Differences amongst themselves, in how they communicate and how they demonstrate their viewpoints results in a lack of a shared understanding, that leads to poor communication.

Therefore, there was a need to establish a unified framework to help reduce or eliminate conceptual and terminological confusion, leading to a shared understanding. To formally represent this shared understanding, this knowledge regarding some domain, a conceptualization can serve as the basis [14]. It is an abstract view of the domain we want to portray composed of objects, concepts and the relationships amongst them. A solution to represent this knowledge, in the field of computer science, was to use ontologies.

What are ontologies?

Several definitions for the term "ontology" have been given over the years. Each author provides their view over the subject, but the definitions provided are quite alike amongst themselves.

When describing ontologies, to be used in the computer science field, Gruber described them as *"an explicit specification of a conceptualization. For knowledge-based systems, what "exists" is exactly that which can be represented."*[14]. This is, until the current days, one of the most quoted definitions, and based on it, many others were proposed. Uschold defined ontology as *"Ontology is the term used to refer to the shared understanding of some domain of interest which may be used as a unifying framework to solve the above problems in the above-described manner."*[32], which was the adopted definition for this research.

Corcho et al. identified the above mentioned and other definitions in the literature: *"Ontologies are defined as a formal specification of a shared conceptualization"*; *"a logical theory which gives an explicit, partial account of a conceptualization"* [7, 28]. For van der Vet et al. *"an ontology serves as a partial specification of the knowledge representation to be built in a later stage. The specification is partial because it supplies concepts in which states of affairs can be expressed but does not actually specify states of affairs."*[33].

Ontological benefits

Ontologies can be beneficial for various purposes. Following the findings in the literature, Uschold et al. identifies three levels of benefits in using ontologies [32]. At the *communication* level, they enable shared understanding and facilitate communication between people; *Inter-operability* since users need consistent tools that enable the integration of various software tools and to exchange data; *Systems engineering* as to applications that support the design and development of software systems.

More recently, two new levels of benefits have been highlighted [37]. *Understanding*, meaning that an ontology can serve as documentation for the conceptualisation of a domain; *Reuse*, the existence of an ontology similar to the one intended to be developed.

Additionally, doubt arises whether the usage of ontology can bring such significant benefits as some say. Therefore, the costs of implementation must be taken into account and verify if they are justifiable.

Ontology Engineering

Initial research has proposed general criteria to guide the development of ontologies [14]. Other researches state-of-art, portray several approaches that have been devised since then, for the process of design and development of ontologies, namely:

- Methontology [10];
- Ontology Development 101 [22];
- On-To-Knowledge [29];
- Toronto Virtual Enterprise Method (TOVE) [15];
- Uschold and King's [32];
- KACTUS [2];
- NeOn [30].

These methodologies are the most used or analyzed by those who study or develop ontologies, and based on them, three phases that are common to the majority have been identified.

For the first phase, the main goal is to identify the domain that the ontology will cover, what will it be used for, the questions it is supposed to answer and who will use and maintain the ontology.

The second phase is the identification and definition of the main components of ontologies, which are the classes, relations and instances [8].

Finally, the third phase corresponds to the evaluation of the ontology. From the literature, ontology evaluation can be divided into three topics [3, 11, 24, 31, 35], in which it is decided the criteria to evaluate it (i.e. consistency, completeness or Computational efficiency), the level at which it is going to be evaluated (i.e. Hierarchy, syntactic level or structure) and the approach taken to perform it (i.e. Data-driven evaluation).

Bottom-up Ontology Development

Following a bottom-up approach means that the terms of the ontology, and possibly the relationships between them, are obtained based on empirical data. There are three approaches to acquire the domain terms and relations: Manual -The ontology is built by hand; Semi-automatic - Computer system that recommends the addition of new ontological components depending on analysis on the domain data; Automatic - The ontology is developed by the system, with no human intervention required.

Skills Ontology

There are some ontologies developed for the domain of skills, specially regarding IT skills. An ontology-based algorithm, in the form of a graph, was proposed in [19]. The principals of the ontology are that each sub-ontology represent a domain, the nodes represent skills and directional edges representing the relationships and their relative weight to the existing relationship between the nodes. The process of matchmaking itself is done calculating similarities between the job requirements and the candidates' skills.

To guide students to understand, based on their training (acquired skills) and their interests related to the IT context, Nguyen, et al. proposed an ontology of IT skills [21]. It consists of 214 terms, selected from the curriculum and job requirements, which are grouped into specific IT categories. Nguyen et al. then proceed to the application of the created ontology for the matching between students and job positions.

Mochol et al. proposed a human resources ontology [20]. The ontology was intended to be used as a job portal to allow job and candidate posting and also providing support as matchmaking, in job seeking.

One of the thirteen sub-ontologies that compose the Reference Ontology [12], corresponds to the Skill Ontology. This ontology has two concepts, the concept of Skill and a concept of ICT Skill, the last is subclass of the first. For example, the Hardware skill has ICT Skill Hardware programming. It is based on the European Dynamics Skill classification, which has 291 skills.

When our work was already in the ontology development phase, more specifically in the first phase of the evaluation, we became aware of a new proposal for an IT Skills Ontology [6] developed in the same domain as our own. It is an ontology developed for the IT recruitment area, which eventually was adopted by the company that motivated this work. It is more of a literature-focused approach, in which the terms that compose it were acquired through an extensive research of skills for the various IT categories. We took advantage of this ontology to evaluate ours and thus improve the final quality in terms of structure and knowledge generated.

RESEARCH PROBLEM

This section corresponds to the first step of the DSRM "Problem identification and Motivation", where the research problem is identified.

Many companies choose to delegate the recruitment process to specialized companies for this purpose [16]. These recruiting companies, usually have a large pool of candidates, according to their field(s) of expertise, to facilitate the process. This means that, according to job requirements, provided by the client company, they do the filtering and preliminary interviews from the candidates within their candidate pool. So the company itself, only enters the process in a more advanced stage, with fewer and better candidates for screening.

The more precise the candidate filtering process is, for the first interviews and for later sending the information to the company, the better the outcome. Also, the less time it takes to complete the candidate selection process, ensuring accuracy in the selection, the faster they can move on to the next advertisement job. Therefore, creating a higher probability of generating profit to the company.

To achieve this goal, these recruitment companies need means, which has been the interest of several researchers and organizations [9]. More automatized and accurate methods or tools, otherwise, they may miss ideal candidates and keep taking too long, leading to money loss.

Ontology-based techniques, are one of the matchmaking approaches [9], between candidates and job positions, identified in the literature and having already been considered, and in some cases applied. In which the match between candidate and job position is based on the skills set they possess or require.

For the specific case of the recruitment company that motivated this work there is a need for an ontology that is directed to the IT field. Focused on the domain in which the company performs its work.

The categories and skills used to define a job advertisement or a candidate are stored in specific tables in the recruitment company's DB. With this information (containing around 700 terms), it is possible to develop an ontology in a bottom-up approach, differentiating it from other ontologies. Furthermore, with this approach, we will be able to portray a more precise domain in which IT recruitment companies operate. As far as it was possible to ascertain, in previous research, there were no developments of bottom-up ontologies for IT skills.

Therefore, the problem identified in this research is the **lack of ontologies for IT skills, developed taking advantage of empirical information existing in the recruitment companies themselves to better portray the domain in which they work.**

DEVELOPMENT

Covering the DSRM's Design and Development step, we chose to follow as a guide the Ontology Development 101 methodology [4, 22], which consists of seven steps for the development of an ontology. The choice of this methodology was due in the first instance to the description of its application is made using the same software chosen for the development of our ontology. Also, the simplicity and clarity in applying the various steps compared to other methodologies was an important fact since it was the first contact with the development of ontologies. The methodology process can be better visualized in the Figure 1.

Being the development of an ontology such an iterative process, we knew that there could be a need to go back some steps to improve the quality of the final ontology. With this in

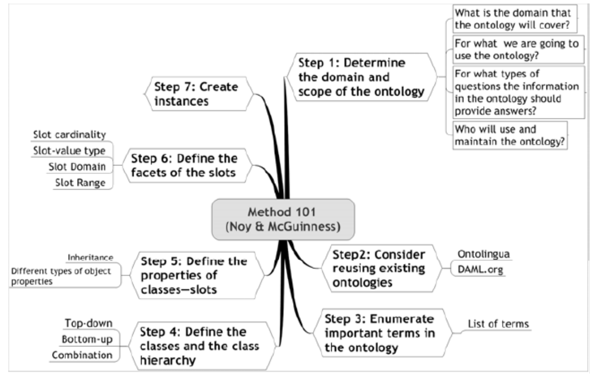

Figure 1. Ontology Development 101 Method [4]

mind, we ended up doing two iterations on the course of the development the first one from steps one to six and followed by an intermediate evaluation, and the second one where we went back to perform steps five, six, and seven. To perform the development, we choose to use the Protégé tool, for being the most used one, having the highest amount of accessible documentation and due to the Cellfie plug-in, that automates the mapping of excel files content to the ontology.

1) Determine the domain and scope The first step consists of defining the domain and scope, answering some predefined questions of the methodology: a) What is the domain that the ontology will cover?, b) For what we are going to use the ontology?, c) For what types of questions the information in the ontology should provide answers? and d) Who will use and maintain the ontology?. The answers to these questions may change during the ontology-design process, but at any given time they help limit the scope of the model.

In our work, the domain of IT skills and IT job categories is portrayed. Recruitment companies can use this ontology to facilitate the selection and recruitment process of candidates for job positions basing on the candidate's skills and the requirements for a given job. The maintenance process is to be the responsibility of whoever uses the ontology. Competence questions were defined to help determine the ontology scope:

1. What are the best suited category for a candidate with the skills PHP, HTML and Javascript?
2. Which skills one should have to be a good Front-end developer?
3. Java is a good to have skill for which categories?
4. Is Bootstrap a good complementary skill for HTML?
5. What are the skills shared by a Front-end developer and a Data Scientist?
6. What are the programming frameworks associated with Python?
7. Can the ontology provide match's between candidates and jobs ads?
8. Can the ontology provide match's between jobs ads and categories?

9. Can the ontology provide match's between candidates and categories?

10. Can the defined properties provide inferred knowledge ?

2) Consider reusing existing ontologies

In this step, we should consider the use of ontologies developed by other persons, to the same end of ours or by means of refinement and extension for our particular domain.

Since our problem focuses on the lack of developed ontologies, using the existing data in recruitment companies as an advantage, we did not reuse any existing ontology identified in the ontology, but the fact that a database in its nature is an ontology, we could say that there was a re-utilization of an ontology. Nevertheless, we took advantage of an existing ontology as a mechanism to evaluate by comparing it to our ontology.

3) Enumerate important terms

The terms used to be part of the ontology are acquired at this stage. A manual approach to gather the terms was followed, in which we searched the terms in the company's DB and extracted them to excel files so that we could import them to Protégé with Cellfie. We ended with three files:

- One with all the skills - 761 terms, which were analysed one by one, to assess if they were useful or not. At this point we eliminated those that were duplicated (i.e. "JIRA" and "jira"), non-valid IT terms(i.e. "veterinary"), terms that were considered as a category (i.e. "Backend Development"), some were changed to create a more uniform collection (i.e. "Sharepoint" became "Microsoft Sharepoint"), and we even found cases were the terms corresponded to organizations/companies (i.e. SAP Solutions);
- One with all the categories - 16 terms, among them we found three that are very borderline. They are not entirely related to IT, but since they portrayed the vision of the company at a given point, we decided to keep them;
- One with all the existing relations between skills and categories - 126 relations that would prove useful in the next steps;

4)Define classes and the class hierarchy

From the lists of terms obtained in the previous step, a main excel sheet was created, with the final set of skills, categories, relations and the position in the hierarchy that a given skill had.

The skill hierarchy was design prior to the process of identification of relations, providing for a better organization of skills and more easy identification of relationships between the classes in the next step of the methodology. It took into consideration the skills from the database that fit with our domain and with the PES skill concept. The resulting hierarchy can be seen in Figure 2.

5)Define class properties

Classes alone are not enough to define the ontology clearly, nor to answer the competency questions presented in the first step. Therefore, after defining the classes, we described their relations, that are represented with properties.

Figure 2. Class hierarchy for skills

To define these properties to relate skills and categories, we used the description previously made for each skill and conducted an online search to identify which category or categories would that skill be associated with. The needed properties that were defined are shown in Table 1.

Table 1. Properties

Relation Name	Domain	Ranges
uses	Categories	Skills
hasKnowledgeOf	Categories	Knowledge_type
isProgrammerOf	Categories	Programming_ Language_type
isFrameworkOf	Frameworks_type	Programming_ Language_type

The "uses" property relates categories with skills that represent development tools and frameworks, "hasKnowledgeOf" relates categories with skills that represent knowledge in the IT area, "isProgrammerOf" relates categories with skills that represent programming languages and finally "isFrameworkOf" represents the existing relations between skills, and more specific Programming Languages and their Frameworks. This last property, was only added in a second iteration of the development process, after an initial round of evaluations perform.

6)Define the facets of the properties

Allowed classes for slots of type Instance are often called a range of a property. For example, a possible class named Programming Languages is the range of the"isProgrammerOf"property. The classes to which a property is attached is called the domain of the property. So the *Programming Languages* class has as domain the entire set of job categories such as the *Frontend developer*.

Additionally, they can have facets describing the types and number of values (cardinality). We used the cardinality facet, to define the minimum number of association with a specific property, for example, to be a back-end developer one must possess at least two skills of the"ProgrammingType".

7)Create instances

This step was only performed in the second iteration of the methodology. Only by having classes we were not able to correctly infer the ontological properties defined, therefore we needed to add instances to achieve better results. The instances added represent not only the entire set of skills but also candidates and job advertisements.

During this development step, we also conduct an analysis on the skills that we had previously identified as being Programming Frameworks. This served to understand and identify with which Programming Languages they were associated. This information, was all portrayed in a new excel sheet, in which we identified the information regarding a given skill, as Table 2 shows. The class Python, which represents the programming language, has as an instance Python_I and is not a framework of any other programming language(or skill). On the other hand, the class Espresso, which represents the framework, has as an instance Espresso_I and is a framework of Java_I and Kotlin_I (that represent the classes Java and Kotlin).

Table 2. Instances sample

Class	Instance	Framework of
Python	Python_I	——
Django	Django_I	Python_I
Espresso	Espresso_I	Java_I, Kotlin_I

All the terms to be represented as skills or categories in the ontology and the instances were automatically defined through the Cellfie plug-in. Therefore, we wrote transformation rules that allowed us to convert the contents of the excel sheets into the right place in the ontology. Figure 3 portrays one of the used rules.

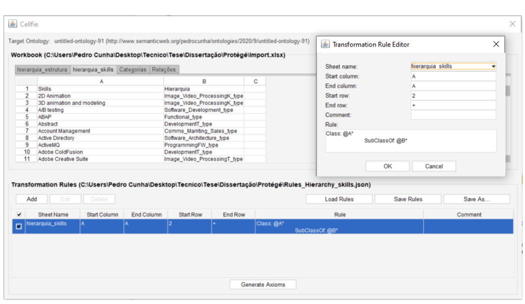

Figure 3. Example of a Cellfie transformation rule

DEMONSTRATION

This section covers the Demonstration step of the DSRM, where for the exact purpose of demonstrating the feasibility of the developed solution, data from candidates and job advertisements sourced from the company's DB will be used.

Data from the candidates and job advertisements was obtained by performing a new search to the company DB. The information that we retrieved for the candidates was ids and skills, and for the job advertisements, we retrieved ids and the skills posted as requirements.

Again, similar to what was the analysis made to find out which tables were useful for the acquisition of terms for skills and categories, we study the DB and concluded which were the tables useful to our needs: One with all the candidates; One that related the candidates and skills that they possess; One with all the required job advertisement information; One table of descriptions for the codes used to reference skills in all the previous ones.

We extracted this information to two excel sheets to work offline, not needing to search every time to add one candidate or job to the ontology. Similar to what happened in the case of Skills and Categories, in this case, two top classes were added to the ontology, *Candidates* and *Job_Ads*, under which we manually created classes to represent each candidate and job advertisement.

The obtained results, after the ontology had been classified and inferred by the reasoner, can be seen in Figure 4, which portrays the match between a candidate and a job advertisement, based on the skills that the candidate possesses and the ones required for that specific job.

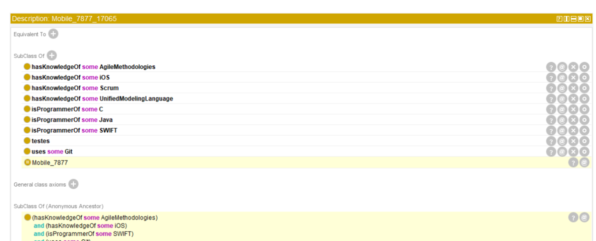

Figure 4. Example of the match between a candidate and a job advertisement

As for Figure 5, it portrays the match between a job advertisement with the *MobileAppsDeveloper* category, based on the skills required for it.

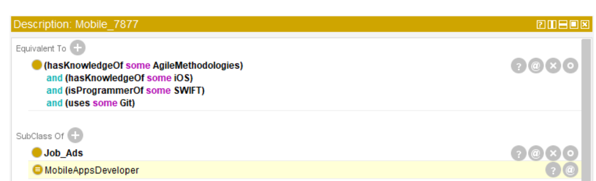

Figure 5. Example of a match between a job advertisement and a category

There were also cases when no match occurs, either for candidates and for job advertisements, as can be seen in Figures 6 and 7.

Figure 6. Example of a job advertisement matching no category

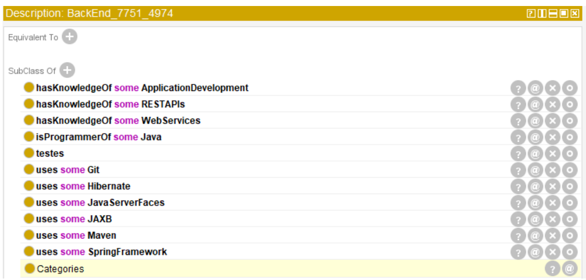

Figure 7. Example of a candidate matching no category

The instance "Candidate_9164" represents one candidate of the company's DB, with the skills he possesses, defined as object property assertions. With his set skill, after running the reasoner, the profile match's the "Back-end developer" and the "ElectronicsAndTelecommunications" categories, as Figure 8 shows.

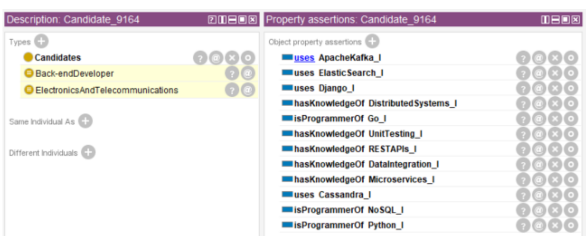

Figure 8. Instance of the candidate 9164

Then we removed the object property assertion of *"isProgrammerOf Python_I"* thus, eliminating the direct relation between the candidate and the *Python* skill. Due to the inference of the rule previously defined, when running the reasoner, the ontology classifies the candidate as being able to program in *Python* since he uses the *Django* framework. As shown in Figure 9.

The last example describes the case when the candidate matches with a given job advertisement and category, but on the other hand, there is no match between job advertisement and category. It is the case of the "Candidate_9127", in which the skills required for the job are not enough to match a category, but the candidate does match both the job advertisement and the "Full-stack Developer" category. Figure 10 portrays this case.

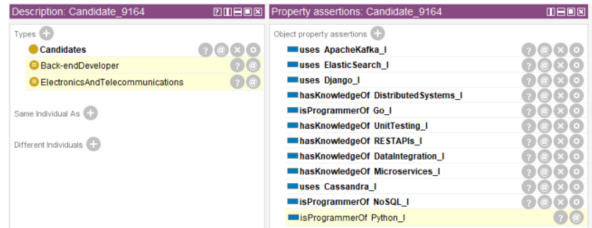

Figure 9. Instance of the candidate 9164 without the Python skill

Figure 10. Job example that does not match a category but a candidate match's the job and a category

EVALUATION

This sections regards the evaluation phase of the DSRM methodology where our artifact, our IT Skills Ontology was evaluated, following four approaches to conduct the assessment on the quality and knowledge that the ontology is able to generate.

Reasoners

We used reasoners available in the Protégé, to ensure that nothing was being wrongly made, using mostly the Pellet reasoner. The usage of reasoners helped us identify problems regarding the definition of classes and properties in the Protégé. Running the reasoner classify and infer the ontology, given what was defined, and highlights problems structural problems of the development.

Comparing With Another Ontology

By comparing our ontology with other IT Skills Ontology [6], developed in the same context as ours, but following an approach focusing more on the literature, enabled us to identify some properties that we did not defined in our own. Those properties, and specially the one that we ended up defining in our ontology, that relates skills that represent Programming Frameworks and Programming Languages, helped to improve the knowledge that the ontology can generate. With it, whomever uses the ontology can easily know with which Programming Language skills does one Framework skill relates. Of course, together with the increase knowledge generated with this added property, the quality of the ontology was also increased.

Competency Questions

With regard to the defined competency questions, the results were as follows:

1. Front-end Developer and Full-stack Developer;
2. One should program in at least two programmig languages and use one programmig framework. Due to a very high level of description there are numerous skills associated with Front-end Development;

3. Back-end Developer, DevOps and SysAdmin, Fullstack Developer and MobileApps Developer;
4. Yes, every category that has HTML as a skill also has CSS. Bootstrap is a framework of CSS, and therefore it is a good skill combination;
5. There are no skills shared by this two categories in the ontology;
6. Django, Falcon, Flask, Keras, NumPy, Pandas, Pyramid, Qlik Sense, SQLAlchemy, Starlette, TensorFlow and ZeroMQ;
7. Yes, but only if the candidate has all the job ad skills.
8. Yes, but many job ads have few skills, which having such high-level categories makes hard to have match's;
9. Yes, although Q7 problem is also verified adding to the fact that to many candidates have more than 10 skills, which facilitates the match;
10. Yes. The isFrameworkOf provides inferred knowledge, since just after inferred, will it compile and be represented in the ontology due to the defined rules.

Data-Driven with Candidates and Jobs Advertisements

The Demonstration step performed in both iteration provided a Data-driven evaluation of our solution, where we took advantage of the information regarding candidates and job advertisements present in the company's DB. With it, we were able to better adjust the needed minimum skills that one must possess to match to a given category and also prove that the ontology can match candidates with job advertisements.

Performing this evaluation, we notice that some candidates and job advertisements had few skills and would not match any category. To a certain point, it was reasonable to reduce the category number of minimum skills required to match but reducing more would lead to one candidate or job advertisement match several categories based on a handful of skills.

While in the first iteration, these small modifications were applied to the defined relations, increasing the generated knowledge, in this second iteration, we did not change them since they were at a point that no increase of knowledge was to be gained with modifications.

In the first iteration, the evaluation of candidates (and job advertisements) focused more on the classes, while in the second one, it focused more on the individuals rather than as classes. This because it allowed to test the rules, and therefore, the added new property, to understand if the improvement helped in the ontological classification process.

Evaluation Assessment

This section presents the result of the evaluation assessment on the ontology, conducted through the first and second iteration, and follows the criteria and the chosen approach levels identified in the theoretical background research.

Regarding the criteria parameters, the results were the following:

- *Accuracy* - Although no user evaluation was conducted, the ontology captures and represents the view of a company in the recruitment for the IT field. To that extent, it is accurate;
- *Completeness* - If we consider the whole IT area as a domain, then the ontology does not cover it. However, considering as domain the universe in which the company was developing its work, at the time in which the data was acquired, and the capacity to answer the formulated competency questions, then the ontology covers the domain to a certain point;
- *Computational efficiency* - As we added new candidates and job advertisements to the ontology, it would take more time for reasoners to process it. We can expect that the ontology, at least if used directly in Protégé, although being able to grow, it will take more time for reasoners to process it;
- *Conciseness* - In fact, it does include some irrelevant axioms. There were three categories in the data that we opted to include in the ontology;
- *Consistency* - The usage of reasoners helped to ensure in fact that the developed ontology is consistent;
- *Expandability* - The ontology is quite tolerant to the addition of new definitions and concepts. A practical example is the addition of the most recent property and the relationships that were defined from it, in which there was no need to make changes to the existing structure;
- *Organizational fitness* - We were not able to evaluate this parameter since our ontology was not deployed in an organizational environment;
- *Sensitiveness* - The process of identifying the minimum skills needed to match a candidate or job advertisement with a category (or more) has shown that there is considerable sensitivity on the part of ontology. Sometimes reducing the minimum number of a given relation by one was enough to infer three or four new matches (in a universe of 50 candidates and job advertisements).

As previously mentioned, we conducted a data-driven evaluation of the ontology, which acts on three of the six levels identified in theoretical background research. Our analyses of them is as follows:

- *Lexical, vocabulary, or data layer* - All the vocabulary used for the construction of the ontology, from the terms of skills, categories to properties and the hierarchy class, are terms whose source is the company's DB or other such as existing IT ontologies.
- *Hierarchy or taxonomy* - This ontology's main objective is to serve as a tool to facilitate the recruitment process, and in particular, to help in the match process between candidates and categories or job advertisements. Therefore, this was our main goal during the development, so the created concept hierarchy, and here we included not only the relationship between skills and categories but also the class hierarchy, was undoubtedly our focus.

- *Other semantic relations* - The calculation of the accuracy performed in the first iteration set, with 20 candidates and 32 job advertisements, resulted in a 57% precision, meaning that in 57% of the cases, there was a match between candidates or job advertisements and at least one category. This value increased slightly to 60% after the improvements of the second iteration.

  It was also possible to verify that adding the new property and rule returned eight inferences. Meaning that in eight cases, it generated knowledge of programming in a particular language from a possessed or required framework.

CONCLUSION

This research followed the DSRM, which comprises six steps, where after a literature review in the context of this work, we defined the problem in the first step as the lack of ontologies for IT skills, developed taking advantage of empirical information existing in the recruitment companies themselves to better portray the domain in which they work. From the identified problem, a proposal for the development of an IT Skills Ontology was outlined. It followed the Ontology Development 101 methodology, which consists of seven steps for an ontology development, having been the terms of the ontology gathered from an IT recruitment company's DB.

Both methodologies that guided this work contemplate iterations. Therefore, the development was performed in two iterations of the DSRM, targeting the Demonstration and Evaluation steps, being the Ontology Development 101 also a target of these iterations.

With the second iteration, which was based on the results obtained from the evaluation performed in the first iteration, it was possible to make improvements to the ontology. These improvements made the ontology more complete and provide a higher quality, making it possible to infer more knowledge.

Five approaches were used to evaluate the ontology: Comparative Evaluation, Reasoners, Data-Driven, Competency Questions and an Overall Assessment.

Limitations and challenges

During the development of this research, we faced some limitations and challenges:

- In terms of methodologies for ontology development, most of the ones we studied did not contain practical examples of how they were applied;
- Somewhat related to the previous point, the ontologies we found, from IT and other domains, made little reference to the applied development methodology;
- Lack of information on IT Skills Ontologies, especially regarding the lack of detail on how their development;
- The process of identifying each skill to know its functionality and consequent identification of the category to which it belongs;
- The limitation of the Cellfie plug-in at the mapping level, which led to more manual work in defining the relationships identified for the ontology;
- The defined categories were of such a high level that it was complex to define the necessary minimums for the match.
- The biggest one was the change of context for which the ontology was developed. The initial idea was to be an ontology to serve as a comparison with a second ontology, developed following a non-bottom-up approach, in which the goal was to arrive at a final ontology that was composed of the best aspects of the two.

Future Work

The results of the evaluation of the ontology allowed us to verify that there are certain aspects of the ontology which can improve as a way to increase the quality and inferred knowledge.

By the evaluation conducted, it is clear that the ontology lacks detail as to the categories that it represents. In terms of high-level representation, the ones defined are adequate for the IT field. However, being so high level, the specification of each one is reduced.

Another feature regarding the ontology properties that would increase the knowledge generated by the resulting ontology is a property that relates similar skills, for example, skills as PostgreSQL and MySQL.

New instance properties, such as the version of a given skill, the experience a candidate has with a given skill, or the experience needed with a given skill for a given job.

To reduce the required time for the classification and inference process, thus boosting the amount of knowledge generated at once, it may be considered the development of an application whose core is the ontology and around it a system capable of taking a higher advantage of it. A system that also facilitates the introduction of new elements such as candidates and job advertisements, with a more user-friendly design.

REFERENCES

[1] Paul Attewell. 1990. What is skill? *Work and occupations* 17, 4 (1990), 422–448.

[2] Amaia Bernaras, Iñaki Laresgoiti, and Jose Manuel Corera. 1996. Building and Reusing Ontologies for Electrical Network Applications. In *ECAI*.

[3] Janez Brank, Marko Grobelnik, and Dunja Mladenic. 2005. A survey of ontology evaluation techniques. In *Proceedings of the conference on data mining and data warehouses (SiKDD 2005)*. Citeseer Ljubljana, Slovenia, 166–170.

[4] Karin Koogan Breitman, Marco Antonio Casanova, and Walter Truszkowski. 2007. Methods for ontology development. *Semantic Web: Concepts, Technologies and Applications* (2007), 155–173.

[5] Coral Calero, Francisco Ruiz, and Mario Piattini. 2006. *Ontologies for software engineering and software technology*. Springer Science & Business Media.

[6] Beatriz Carrilho Toscano, Miguel Mira da Silva, and Rubén Sousa Pereira. 2020. IT Skills Ontology.

https://fenix.tecnico.ulisboa.pt/cursos/meic-a/dissertacao/846778572213024. (2020). (Accessed on 02/07/2021).

[7] Oscar Corcho, Mariano Fernández-López, and Asunción Gómez-Pérez. 2003. Methodologies, tools and languages for building ontologies. Where is their meeting point? *Data & knowledge engineering* 46, 1 (2003), 41–64.

[8] Oscar Corcho, Mariano Fernandez-Lopez, and Asuncion Gomez-Perez. 2007. Ontological engineering: what are ontologies and how can we build them? In *Semantic web services: Theory, tools and applications*. IGI Global, 44–70.

[9] Maryam Fazel-Zarandi and Mark S Fox. 2009. Semantic matchmaking for job recruitment: an ontology-based hybrid approach. In *Proceedings of the 8th International Semantic Web Conference*, Vol. 525.

[10] Mariano Fernández-López, Asunción Gómez-Pérez, and Natalia Juristo. 1997. Methontology: from ontological art towards ontological engineering. (1997).

[11] Asunción Gómez-Pérez. 2004. Ontology evaluation. In *Handbook on ontologies*. Springer, 251–273.

[12] Asunción Gómez-Pérez, Jaime Ramírez, and Boris Villazón-Terrazas. 2007. Reusing Human Resources Management Standards for Employment Services.. In *FIRST*. 28–41.

[13] Francis Green. 2013. *Skills and skilled work: An economic and social analysis*. Oxford University Press.

[14] Thomas R Gruber. 1995. Toward principles for the design of ontologies used for knowledge sharing? *International journal of human-computer studies* 43, 5-6 (1995), 907–928.

[15] Michael Grüninger and Mark S Fox. 1995. Methodology for the design and evaluation of ontologies. (1995).

[16] Myrna L Gusdorf. 2008. Recruitment and selection: Hiring the right person. *USA: Society for Human Resource Management* (2008).

[17] Alan R Hevner, Salvatore T March, Jinsoo Park, and Sudha Ram. 2004. Design science in information systems research. *MIS quarterly* (2004), 75–105.

[18] Françoise Delamare Le Deist and Jonathan Winterton. 2005. What is competence? *Human resource development international* 8, 1 (2005), 27–46.

[19] Hexin Lv and Bin Zhu. 2006. Skill ontology-based semantic model and its matching algorithm. In *2006 7th International Conference on Computer-Aided Industrial Design and Conceptual Design*. IEEE, 1–4.

[20] Malgorzata Mochol, Holger Wache, and Lyndon Nixon. 2007. Improving the accuracy of job search with semantic techniques. In *International Conference on Business Information Systems*. Springer, 301–313.

[21] Cuong Duc Nguyen, Khoi Duy Vo, Dang Bach Bui, and Dung Tien Nguyen. 2011. An ontology-based IT student model in an educational social network. In *Proceedings of the 13th International Conference on Information Integration and Web-based Applications and Services*. 379–382.

[22] Natalya F Noy, Deborah L McGuinness, and others. 2001. Ontology development 101: A guide to creating your first ontology. (2001).

[23] Jim Pinkelman. 2013. Computing changes: an industry perspective. *ACM Inroads* 4, 4 (2013), 39–42.

[24] Helena Sofia Pinto and João P Martins. 2004. Ontologies: How can they be built? *Knowledge and information systems* 6, 4 (2004), 441–464.

[25] Chris Piotrowski and Terry Armstrong. 2006. Current recruitment and selection practices: A national survey of Fortune 1000 firms. *North American Journal of Psychology* 8, 3 (2006), 489–496.

[26] Michael H Randall and Christopher J Zirkle. 2005. Information technology student-based certification in formal education settings: Who benefits and what is needed. *Journal of Information Technology Education: Research* 4, 1 (2005), 287–306.

[27] Usman Shahbaz, Amin Beheshti, Sadegh Nobari, Qiang Qu, Hye-Young Paik, and Mehregan Mahdavi. 2018. irecruit: Towards automating the recruitment process. In *Service Research and Innovation*. Springer, 139–152.

[28] Steffen Staab and Rudi Studer. 2010. *Handbook on ontologies*. Springer Science & Business Media.

[29] Steffen Staab, Rudi Studer, H-P Schnurr, and York Sure. 2001. Knowledge processes and ontologies. *IEEE Intelligent systems* 16, 1 (2001), 26–34.

[30] Mari Carmen Suárez-Figueroa, Asunción Gómez-Pérez, and Mariano Fernández-López. 2012a. The NeOn methodology for ontology engineering. In *Ontology engineering in a networked world*. Springer, 9–34.

[31] Mari Carmen Suárez-Figueroa, Asunción Gómez-Pérez, Enrico Motta, and Aldo Gangemi. 2012b. Introduction: Ontology engineering in a networked world. In *Ontology engineering in a networked world*. Springer, 1–6.

[32] Mike Uschold and Michael Gruninger. 1996. Ontologies: Principles, methods and applications. *The knowledge engineering review* 11, 2 (1996), 93–136.

[33] Paul E Van Der Vet and Nicolaas JI Mars. 1998. Bottom-up construction of ontologies. *IEEE Transactions on Knowledge and data Engineering* 10, 4 (1998), 513–526.

[34] Patrick Van Esch, J Stewart Black, and Joseph Ferolie. 2019. Marketing AI recruitment: The next phase in job application and selection. *Computers in Human Behavior* 90 (2019), 215–222.

[35] Denny Vrandečić. 2009. Ontology evaluation. In *Handbook on ontologies*. Springer, 293–313.

[36] Franz E Weinert. 2001. Concept of competence: A conceptual clarification. (2001).

[37] Burcu Yildiz. 2007. *Ontology-driven information extraction*. na.