Contemporary Engineering Sciences, Vol. 8, 2015, no. 33, 1551 - 1563
HIKARI Ltd, www.m-hikari.com
http://dx.doi.org/10.12988/ces.2015.510282

Semantic Architecture for the Analysis of the Academic and Occupational Profiles Based on Competencies

Alexandra González Eras

Department of Computer Sciences
Universidad Técnica Particular de Loja, Ecuador

Jose Aguilar

CEMISID, Dpto. de Computación, Universidad de Los Andes, Mérida Venezuela
Prometeo researcher, Universidad Técnica Particular de Loja, Ecuador

Abstract

This research shows a semantic architecture for the extraction, comparison and feedback of professional and educational competencies. The main product is a scheme that facilitates the detection of competencies based on the skills and knowledge. The scheme carries out tasks of natural language processing and of similarity calculation, among others, in order to determine the differences among the professional and educational competencies.

Keywords: Semantic measures, Natural Language Processing, Competency

1 Introduction

For universities and employers, the management of the competencies is a two way process, including knowledge identification to create new graduate profiles and determination of the qualified professional skills to fill a working place. However, in reality it is almost impossible to compare professional and educational competencies, mainly due to incompatible profiles [1]. In addition, the information published on university websites and work platforms is unstructured, ambiguous, and sometimes incomplete [2]. Looking for a solution, models and platforms have been proposed in order to standardize the profiles [3], and compare through competency frameworks. Nevertheless, the actors rarely use these tools, or they have only been proposed for one language or context, without a real application in others. As a result, universities cannot identify the requirements of employers, and employers cannot identify new graduate profiles aligned with their job offers.

The goal of this work is to establish a management system of academic and labor profiles, allowing the extraction, the comparison and the update of the competencies. First, we give an overview of the system architecture, and then focus on the different modules of the architecture. The resulting scheme is the management system cornerstone, which is applied on competency profiles in Spanish.

Notably, we focused on the analysis of the skills and knowledge, because they are the most common concepts associated with the definition of competencies in both contexts [4, 5]: labor and academic; in addition, skill and knowledge patterns are traceable in the profiles [6, 7]. This allows us to perform a less subjective characterization of competencies, compared to other constituent elements, such as attitude and value.

2 Academia versus occupational environment

Training for work is a mix between education, work experience and specific training acquired throughout life; hence competencies are defined and constructed in social practice, as a joint effort between companies, workers and educators [6]. Typically, professional skills and knowledge are acquired at university, where students develop a competence profile. The term competency in the work context can be used sometimes to refer to actions and their consequences, sometimes as cognitive abilities and personal characteristics [1]; whereas in academia competencies are expressed in terms of qualifications and certifications (such as diplomas) [6], or in terms of learning outcomes within educational processes [9]. Normally, the students use the profile to create a summary and seek employment [10]. Similarly, companies seek candidates based on the profile of universities but also consider other aspects such as additional training and experience [11].

We need to establish a common environment to compare professional and educational competencies. Through similarity measures and a description of the notion of competencies based on the natural language processing, we can define a semantic architecture that covers aspects such as extraction, comparison and prediction of competencies. This paper pursues this objective.

3 Semantic Architecture

Fig. 1 shows our semantic architecture, in which the academic and job profiles get into an iterative pipeline process comprising 4 phases: characterization, extraction, comparison and updating. The *extraction phase* includes the development of a scheme based on linguistic patterns, described like logical descriptions, which allow the recognition of the competency descriptor elements (in our case, skills and knowledge). The *characterization phase* analyzes documents using these patterns defined in an ontological model. The *phase of comparison* identifies similarity measures between skill, knowledge and competency patterns (identified in the previous phase), for the determination of the levels of closeness between them. To do this, our scheme combines different similarity measures [12, 17]. This phase has two outputs: the groups of similar competencies, and the remaining competencies. Finally, the *phase of update* analyzes groups of similar and dissimilar competencies, to determine levels of relationship and difference. The outputs of this phase are recommendations such as groups of new competencies, more common skills, new knowledge domains, etc., that provide feedback both for the labor and academic contexts (their competency ontologies). In this work we focus on the phases of characterization, extraction and comparison.

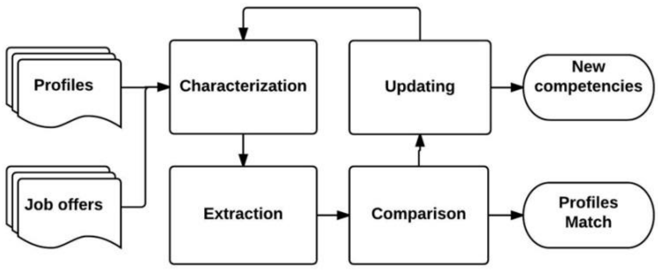

Fig. 1 System's architecture

4 Characterization Phase

The first step is to select the textual documents relating to academic and professional profiles; then it uses linguistic patterns to analyze the documents to recognize the competency elements (skills and knowledge) in them (in each domain, academic and professional). The linguistic patterns have a logical description defined in an ontological model. So, this phrase defines the semantic structure of the ontological models used by our system. Now we explain the steps of this phase.

4.1. Pattern recognition

According our linguistic patterns, the system takes as starting point the basic elements of the competencies: skill and knowledge. Then, it reviews the linguistic structure of each candidate sentence, to detect lemmas and linguistic groups, and finally, establish patterns for each basic concept [6, 13, 14]. In general, we assume that competency is the union of a verb phrase and noun phrase, where the noun phrase (NP) represents knowledge and verbal phrase (VP) represents skills. Table 1 presents these patterns like axioms, as well as examples found in the profiles.

Table 1 Competency patterns

Element	Pattern	Example
Knowledge	Noun Phrase (NP): (NN/S) [(NN/S)(NN/S)] [(NN/S)(IN)(NN/S)]	Proyecto Sistema Operativo Programa de software
Skill	Verb Phrase (VP): (VP)[VB]	Diseñar Gestionar Gestión
Competency	NP + VP	Diseñar programas de software Gestión de sistemas operativos

4.3. Model definition

The ontological model consists of three levels (see Fig. 2), one conceptual where there are definitions based on the concepts of competency, skill and knowledge; the next one has the production rules, where are described the general linguistic patterns of competency, skill and knowledge; and the third one defines the grammatical categorization, which is used to define the linguistic characteristics (textual constituents) and patterns for each component. The levels are implemented in Protégé [16] tool, using SWRL [18] and GATE [19].

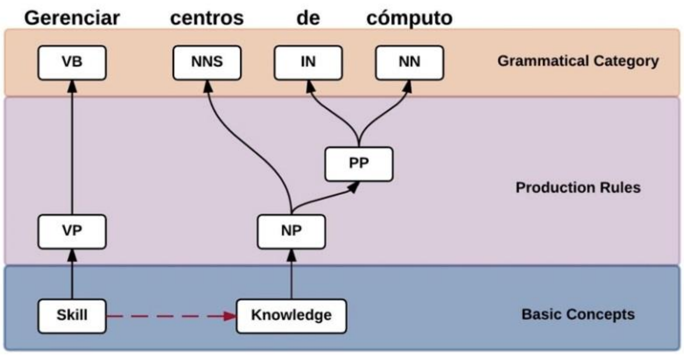

Fig. 2 Model definition

Particularly, the basic level defines both Skill and Knowledge, related to the concept Competency as subclasses (hasClass) through the isfollowedBy relationship, indicating the existence of a predetermined order, in which Skill is followed by Knowledge. The level of production rules presents the definition of the concepts of Skill and Knowledge, using their linguistic equivalences (see table 1). Fig. 3 shows an example of a production rule: it indicates that the superclass Phrase includes subclasses Noun_phrase (Nominal Phrase) and Verb_phrase (Verbal Phrase), that a Verb_phrase is considered a skill and a Noun_phrase is considered a Knowledge. The hasNext property [15] is included to indicate the sequence of the constituents of the competencies.

```
Rules
Verb_phrase(?x) -> Skill(?x)
Noun(?x), hasNext(?x, nil) -> Noun_phrase(?x)
Noun(?x), Noun(?y), hasNext(?x, ?y), hasNext(?y, nil) -> Noun_phrase(?x)
Noun_phrase(?x) -> Knowledge(?x)
Verb(?x), hasNext(?x, nil) -> Verb_phrase(?x)
```

Fig 3. Production rules based on Table

Level 3 describes the linguistic characterizations of noun phrases and verb phrases [20, 21], based on Table 1. For this purpose, it is required that each constituent is associated with a grammatical category (Gramatical_Category). In this case, a verb belongs to the grammatical category VB, a noun to the grammatical category NN or NNS, and a preposition to the grammatical category IN.

5 Extraction Phase

The extraction phase contemplates the population of the ontological models with individuals that are extracted from the corpus. To do this, first the system applies extraction techniques of the natural language processing (NLP) area on the corpuses, in order to identify verbs and nouns in the candidate sentences [21, 22]. Then, it combines the discovered elements according to the production rules. This allows building two ontologies with the same structure but with different instances (individuals): the first with the instances found in academic profiles (academic Ontology), and the second with the instances found in the job offers (work Ontology). In general, this phase defines the individuals (population) of the ontological models used by our system, using the corpuses defined in the first phase.

5.1. Linguistic annotation

The candidate sentences enter the tagging process defined by the NLP schemes [22], in order to label the words according to their types: verbs, prepositions or nouns. The words labeled are entered in the ontology models (are their individuals), according to the associated grammatical category (Grammatical_Category) and the linguistic rules defined for each component of the ontological model.

5.2. Inference by production rules

Instances of verbs (VB), prepositions (IN) and nouns (NN) are associated to nominal and verbal phrases, and are combined to identify skills, knowledge and competencies. Some of the production rules can be seen in Table 2, which cover the most frequent cases (patterns) in the profiles.

Table 2 Production rules

Definitions	Production rules
Competency is the join of skill and knowledge	C → SK
Knowledge is a Noun Phrase	K → NP
Skill is a Verb Phrase	S → VP
Noun Phrase is a noun or a prepositional phrase	NP → NN NN → PP
Prepositional Phrase is the join of a preposition and Noun Phrase	PP → IN NN
Noun Phrase is the join of two nouns	NP → NN NN
Noun Phrase is the join of noun preposition and noun	NP → NN IN NN
Verb Phrase is a verb	VP → V VP → VB
Verb Phrase is the join of one or more verbs	VP → V V V

6 Comparison Phase

The comparison process is based on the measure of taxonomic similarity proposed on [12], which defines that: The "Similarity Measure" between C and C' concepts is based on three aspects (see equation 1): the Similarity of Ancestor of C and C' (SA), the Similarity of Siblings of C and C' (SS) and the Similarity of Descendants of C and C' (SD).

$$MS(C,C') = \frac{SA(C,C') + SD(C,C') + SS(C,C')}{3} \quad (1)$$

The equations 2, 3 and 4 define the above-mentioned measures

$$SA(C,C') = \frac{1}{n}\sum_{i=1}^{n} \max(Sim(Anc_i(C), Anc_1(C')), \dots, Sim(Anc_i(C), Anc_m(C'))) \quad (2)$$

$$SS(C,C') = \frac{1}{n}\sum_{i=1}^{n} \max(Sim(S_i, S_1'), \dots, Sim(S_i, S_m')) \quad (3)$$

$$SD(C,C') = \frac{1}{n}\sum_{i=1}^{n} \max(Sim(H_i, H_1'), \dots, Sim(H_i, H_m') \quad (4)$$

Similarity of Ancestor of C and C` (SA)

The similarity of C and C' concepts (equation 2) will be proportional to the similarity of ancestral concepts. In this case, it is considered the average of the maximal similarities of each ancestor of concept C´ with each ancestor of C, where:

- **Anc_i(C):** ancestor i between concept C and the root.
- **Anc_j(C'):** ancestor j between concept C' and the root.
- **Sim(Anc_i(C),Anc_j(C')):** similarity measure between ancestors of concept C and concept C', based on lexical similarity measures.
- **n:** It is the number of ancestors of concept C.
- **m:** It is the number of ancestors of concept C'.

Similarity of Siblings of C and C' (SS)

The similarity of C and C' concepts (equation 3) will be proportional to the similarity of the siblings. In this case, it is considered the average of the maximal similarities of the siblings of the concept C and concept C', where:

- **S_i:** It corresponds to the i sibling of C concept.
- **S'_j:** It corresponds to the j sibling of C' concept.
- **Sim(S_i,S'_j):** similarity measure between the siblings of C and C' concepts.

Similarity of Descendants of C and C' (SD)

The similarity between two concepts C and C' (equation 4) will also be proportional to the similarity of their direct descendants. In this case, it is considered the average of the maximal similarities of the children of the concept C with the children of the concept C', where:

- **H_i:** It corresponds to the i child of the concept C.
- **H'_j:** It corresponds to the j child of the concept C'.
- **Sim(H_i,H'_j):** similarity measure between the children of C and C' concepts.

The combined measures are used to compare the instances in the ontologies

7 Experiments

In this section we present the utilization of our architecture.

Corpus Description

The sources of our two corpuses are: 1) professional career profiles of Computer Science, who are taken from the websites of Latin American Universities; and 2) job profiles extracted from employment platforms found on Internet. To set up the corpus are selected candidate phrases in each profile, which are chosen from paragraphs corresponding to sections as: description, occupational field, skills, and knowledge areas.

Extraction schema

By applying the schemes of characterization and extraction of our proposition over the paragraphs, in a set of 20 professional profiles and 100 job offers, we can measure the accuracy of the recognition of competencies on them (see Table 3). We see that our system can recognize the competencies, and their components.

Table 3 Accuracy levels of extraction schema

Profile	Skill	Knowledge	Competency	Av. Acc.
Professional	0,7	0,8	0,7	0,73
Job Offer	0,7	0,8	0,7	0,73

Comparison Process

Our proposal of comparison includes the following: first, the selection of a thesaurus against which to compare the individuals; second, the selection of lexical measures to obtain the value similarity between the individuals and the thesaurus; and third, the definition of an algorithm to calculate these measures of section 6.

Regarding the selection of thesaurus, we chose the thesaurus DISCO, which "improves and extends terminological support for the description and translation of abilities, skills, and competences in the contexts of the job market and education"[1] [8]. DISCO offers a taxonomic structure of related terms and phrases to the Computer Sciences area. With regard to measures of lexical similarity, we chose two, the Levenshtein distance measure, which determines the number of changes required for a word to become the other; and the Strikematch[2] measure, which provides greater robustness with respect to change of word order. The first experiments with the two measures confirm the need to implement a combined measure (lexical-taxonomic), also show that Strikematch gives better results.

Fig. 4 presents the algorithm for comparing two individuals C and C´ (in the work and academic ontologies, respectively) using the corpus of DISCO.

```
Start
Variables
  string C, C'
  double SM, SA, SS, SD, Sim
  integer n, m, med
  struct tesauri
   get (C, C')
   calculate n level of (C)
   calculate m level of (C')
   For i=1 to n
     SA= SA + calculate max ancestor (Ci, C'j)/n ∀j=1 to m
   For i=1 to n
      SS = SS + calculate max siblings (Ci, C'j)/n ∀j=1 to m
   For i=1 to n
       SD = SD + calculate max descendants (Ci, C'j)/n
   SM = calculate taxonomical similarity(C, C')
   get SM
```

Fig. 4 Calculate Similarity Algorithm

Table 4 shows the results for various words extracted in the previous phases, we see that individuals have the biggest similarity when they have the same parents, children and siblings (in this case is equal to 1), while when they are very dissimilar this value tends to 0, which corresponds to our desire to determine similarity (synonymy) between words collected in academic and work profiles.

[1] DISCO II, available online at http://disco-tools.eu/disco2_portal/terms.php

[2] Strikematch, available online at http://www.catalysoft.com/articles/StrikeAMatch.html

Table 4. Different comparisons of individuals from the Professionals and Academic ontologies

N	Word from Acad. Ontol.	Word from Prof. Ontol	SA	SS	SD	SM
1	Computacion en paralelo	Computacion distribuida	1	1	1	1
2	Computacion en paralelo	Computacion en red	0.82	0.12	0.05	0.33
3	Computacion en paralelo	Software	0.02	0.03	0.02	0.02
4	Sistemas informaticos	Software	0.2	0.05	0.1	0.09

Table 4 presents the calculation of the similarity for different words extracted from the work and academic ontologies. The first case represents two words with the same parents and siblings (they have not kids, see Fig 6). This is the case with more similarity (we suppose when they have not kids, that SD by defect is 1, and it is similar for the case when they have not parents and siblings). Fig. 5 is the case of lowest similarity, when they do not share elements of the tree (sub-trees) of DISCO taxonomy (it is the experiment 4 of the Table 4). According to our results, when SM has a value superior to 0.3 we consider similar the words, but when this value is close to 1 they are synonymous.

It is important to note that these first results of the comparison is for the case of knowledge. Upcoming work will define the scheme of comparison of skills, which has different characteristics, because it requires the comparison of verbs.

8 Related Works

There are several works that compare competencies through different strategies (semantic models, similarity measures, etc.). For example, [7] proposes a hybrid scheme to extract and compare, based on an ontological model. In [6] and [10] propose an extraction scheme based on comparing vocabularies and skills through ratings. In [9] creates a semantic model where competencies are compared through a difference between levels. The levels correspond where are located the concepts in the ontology. In [1] a measure of similarity that takes into account the distance between concepts of an ontological model is used to compare profiles. In general, none previous work presents a proposal to perform a pattern comparison based on taxonomic structures like thesauri, mixed with similarity measures.

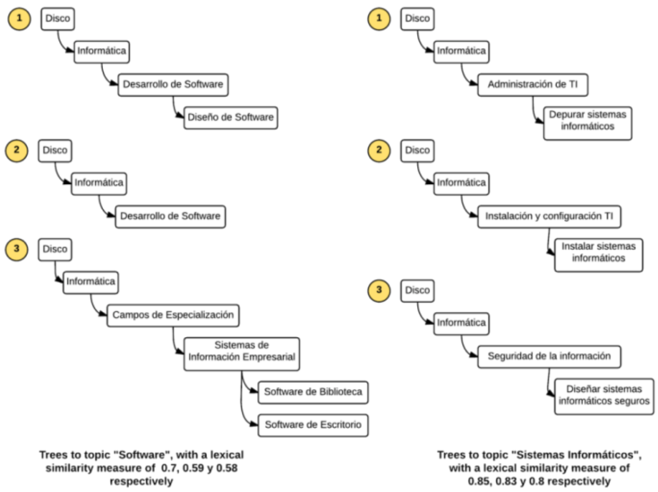

Fig. 5 Case of low similarity between "Software" and "Sistemas Informáticos"

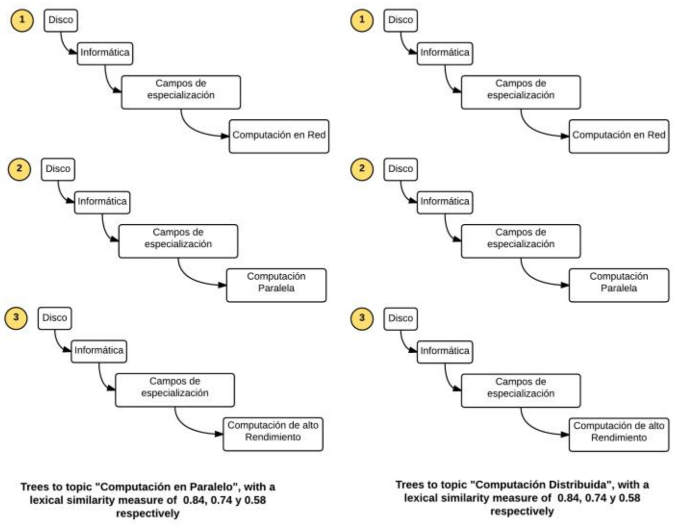

Fig. 6 Case of high similarity between "computación en paralelo" and "computación distribuida"

9 Conclusions

This research presents a model for comparing the competency profiles, using the linguistic patterns of knowledge and skills, through a measure of similitude composed of a lexical and taxonomic part. The linguistic patterns allow considering the variations of competencies found in the texts, which constitutes an advantage of our model with respect to previous works based on ontologies. The preliminary results obtained in the comparison process are encouraging, because it detects the similitude (synonymies) and dissimilitude of the individuals (they represent knowledge) of the work and academic ontologies, using specialized thesaurus. Future works must define the scheme of comparison of skills, which has different characteristics because it must be based on the comparison of verbs.

Based on these results, we will extend our work including alignment techniques of ontologies at the level of the individuals (normally, it is at level of concepts). Also, future jobs must consider the inclusion of other concepts during the comparison process (for example, skills), and exploit the inference process on the academic and job ontologies obtained in this work.

Acknowledgements. Dr Aguilar has been partially supported by the Prometeo Project of the Ministry of Higher Education, Science, Technology and Innovation of the Republic of Ecuador.

References

[1] M. Fazel-Zarandi and M.S Fox, Semantic Matchmaking for Job Recruitment: An Ontology-Based Hybrid Approach, *Proceedings of the 3rd International Workshop on Service Matchmaking and Resource Retrieval*, Semantic Web at the 8th International Semantic Web Conference ISWC, 2010.

[2] M. Fazel-Zarandi, *Representing and Reasoning about Skills and Competencies over Time*, Doctoral dissertation, University of Toronto, 2013.

[3] F. Draganidis and G. Mentzas, Competency based management: A review of systems and approaches, *Information Management & Computer Security*, **14** (2006), no. 1, 51 - 64. http://dx.doi.org/10.1108/09685220610648373

[4] Reference Points for the Design and Delivery of Degree Programmes in Business. Universidad de Deusto, *IEEE Trans. Software Eng.*, **32** (2006).

[5] The European Qualifications Framework for Lifelong Learning. Luxembourg: Office for Official Publications of the European Communities, (2008), ISBN 978-92-79-08474-4. http://www.ond.vlaanderen.be/hogeronderwijs/bologna/news/EQF_EN.pdf

[6] N. Malzahn, S. Ziebarth and H. Hoppe, Semi-automatic creation and exploitation of competence ontologies for trend aware profiling, matching and planning, *Knowledge Management & E-Learning: An International Journal*, **5** (2013), no. 1 84 - 103.

[7] V. Janev and S. Vranes, Ontology-based Competency Management: the Case

Study of the Mihajlo Pupin Institute, *Journal of Universal Computer Science*, **17** (2011), no. 7, 1089 -1108.

[8] H. Muller-Riedlhuber, The European dictionary of skills and competencies (disco): an example of usage scenarios for ontologies, *I- SEMANTICS*, (2009), 467 – 479.

[9] G. Paquette, An ontology and a software framework for competency modeling and management, *Educational Technology & Society*, **10** (2007), no. 3, 1 – 21.

[10] J. Dorn and M. Pichlmair, A competence management system for universities, *European Conference on Information Systems*, (2007), 759 – 770.

[11] J. Dorn, Tabbasum N. and M. Pichlmair, Ontology development for human resource management, *Proceedings of the 4th International Conference on Knowledge Managements*, (2007), 109 – 120. http://dx.doi.org/10.1142/9789812770592_0010

[12] M. Mendonza, N. Perozo and J. Aguilar, An approach for Multiple Combination of Ontologies based on the Ants Colony Optimization Algorithm, *Asia-Pacific Conference on Computer Aided System Engineering*, (2015), 140 - 145. http://dx.doi.org/10.1109/apcase.2015.32

[13] P. Buitelaar and T. Eigner, Topic extraction from scientific literature for competency management, *The 7th International Semantic Web Conference*, 2008.

[14] X. Xiang, D. Zhang, J. Qin and Y. Fu, Ant colony with genetic algorithm based on planar graph for multiple sequence alignment, *Information Technology Journal*, **9** (2010), no. 2, 274 - 281. http://dx.doi.org/10.3923/itj.2010.274.281

[15] T. Rodriguez, J. Aguilar, Aprendizaje ontológico para el marco ontológico dinámico semantico, *DYNA*, **81** (2014), no. 187, 56 - 63.

[16] M. Horridge, H. Knublauch, A. Rector, R. Stevens and C. Wroe, *A Practical Guide to Building OWL Ontologies Using the Protégé-OWL Plugin and CO-ODE Tools*, University of Manchester, 1 2004.

[17] D. Metzler, S. Dumais and C. Meek, *Similarity Measures for Short Segments of Text*, Springer Berlin Heidelberg, 16 – 27, 2007. http://dx.doi.org/10.1007/978-3-540-71496-5_5

[18] I. Horrocks, P.F. Patel-Schneider, H. Boley, S. Tabet, B. Grosof and M. Dean, *SWRL: A Semantic Web Rule Language Combining OWL and RuleML*, W3C Member Submission, 21, 2004.

[19] H. Cunningham, et al, *Text Processing with GATE (Version 6)*, University of Sheffield Department of Computer Science, 2011.

[20] C. D. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing*, MIT Press, Cambridge, 1999.

[21] J. Pustejovsky and J. A. Stubbs, *Natural Language Annotation for Machine Learning*, O’Reilly Media, 2013.

[22] H. Cunningham, D. Maynard, K. Bontcheva and V. Tablan, GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications, *Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics ACL'02*, 2002.

Received: October 18, 2015; Published: December 2, 2015