DOI: 10.20472/IAC.2015.015.185

ELENA ALEXANDRA TOADER

The Bucharest Academy of Economic Studies, Romania , Romania

USING STANFORD PARSER METHOD FOR ASSESSING THE COMPETENCIES OF IT PROFESSIONALS

Abstract:

The paper deals with the presentation of the Stanford Parser method and its implementation into a competency assessment tool. The competency assessment tool is designed for appraising the performance that IT professionals have. Using the Stanford Parser package, the opinions of technical employees can be measured and their level of specific competencies can be revealed. The present research tries to provide evidence on the competencies that IT Romanian professionals have. The research is conducted on implementing the Stanford Parser into a web based assessment tool. This software application represents an efficient and transparent method to evaluate the competencies of technical professionals.

Acknowledgment
This papers was co-financed from the European Social Fund, through the Sectorial Operational Programme Human Resources Development 2007-2013, project number POSDRU/159/1.5/S/138907 "Excellence in scientific interdisciplinary research, doctoral and postdoctoral, in the economic, social and medical fields -EXCELIS", coordinator The Bucharest University of Economic Studies.

Keywords:

Stanford Parser, competency model, competency assessment, IT professional, assessment tool

JEL Classification: C52, M51

1 Introduction

One of the most important types of online resources are the opinions expressed by the employees on web platforms like blogs, forums. A method of performing opinion mining analysis is processing the language used to express it (Natural Language Processing).

In the recent year, the syntactic parsers for English language shave gained an increasing interest due to their features: a large applicability domain, high accuracy: over 90% reported on Wall Street Journal, section of Penn Tree Bank (McDonald, Crammer & Pereira, 2005); efficiency and robustness.

The syntactic parsers are mainly based on dependencies representations which are simple description of the grammatical relationships in a sentence used by people to extract textual relations (Marneffe & Manning, 2011). There are many dependency parser methods that can be made through dynamic programming or graph algorithms.

The Stanford Parser is a package containing implementation of probabilistic natural language parsers in Java programming language, which can be incorporated into different types of applications with a wide range of applicability: academia, industry, government.

The latest version of Stanford Parser is based on a neuronal network which offers a high speed of text parsing and also the performance and accuracy of parsing.

The current research is focusing on implementation of Stanford Parser in a web based competency assessment tool. The IT professionals must evaluate their competencies using a transparent and efficient method that can provide a valid score of a competency assessment. The Stanford Parser represents an innovative method that can be applied in assessment of competency domain.

2 Opinion extraction using Stanford Parser

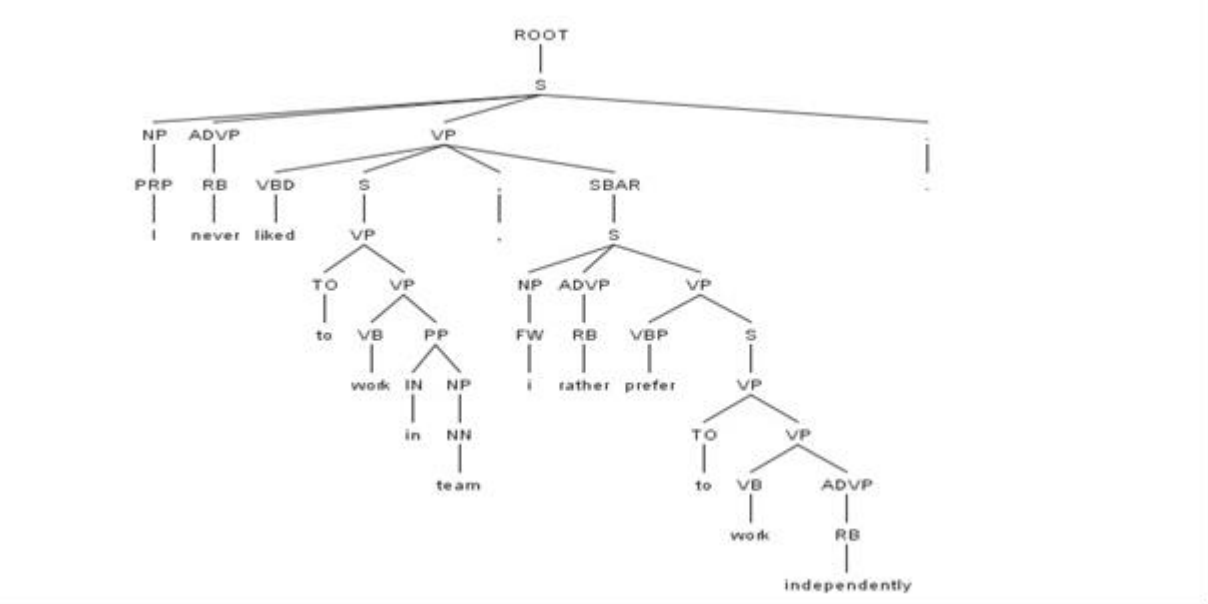
The analysis of an opinion or a speech expressed by someone can be made through a program that examines the grammar structure of the sentences, grouping the words in parts of speech. Those applications are using knowledge's of the language in which the opinions were expressed. The accuracy of the syntactic parser in English language is over 90% (Hudson, 1984).

Stanford dependencies are expanding the Graph Algorithms presented by Klein & Manning (2003) and creates a representation of the grammatical relations between words and their part of speech in the sentence. The output of the parsed sentence is represented by a tree containing those grammar relations. The design of the tree is very simple to understand and efficient in order to extract the opinion expressed by a person. The Stanford Dependencies are represented by a triplet: the name of the relation, the governor and the dependent.

Using the relation between words, it can be represented the sentences through a graph, where nodes are the words and the connection links are the abbreviated name of dependence. In Figure 1 is described an example of the Stanford Parser dependency for the sentence: „I never liked to work in team, i rather prefer to work independently" using the Dependency Stanford Parser (Klein

& Manning, 2003). The sentence is referring to the competency „team work" defined within the competency model defined by Bodea & Toader (2013).

Figure 1: An example of Stanford Parser dependency tree



Source: Own adjustment based on Stanford Parser

3 Stanford Parser and neural networks

The new version of the Dependency Parser is based on neural networks and has multiple advantages than previous versions. The parsing speed and the performance that the neural networks has given, the transition-based dependencies are few advantages provided by the new version. The parser is capable to analyze more than a thousand sentences per second, with a degree of accuracy of 92% (Chen & Manning, 2014).

Although the parser has multiple advantages, there are some disadvantages. Due to a statistical perspective, the parser has poorly estimated feature weights. That means that there is insufficient data for correctly weight the features. There is some exception, when the Word class feature has been introduced, that meant an improvement for parsing performance (Koo et al., 2008). Another disadvantage is that the parser is based on manually templates that needs time and expertise to be completed (He et al., 2013).

The last version of Stanford Dependency Parser defined by Chen & Manning (2014) was created to parse the opinion expressed by a person using a transition-based dependency parser.

The neural network is composed by a vector containing the representations of words, POS tags (part of speech) and dependency labels. The parser is using two different dependency representations (CoNLL and Stanford Dependencies).

Transition-based dependency parsing has the role to predict a transition sequence from the initial to terminal configuration, resulting a dependency parse tree as it can be observed in Figure 1.

The POS tagging for the sentence: „I never liked to work in team, I rather prefer to work independently" is described in Figure 2.

Figure 2: The POS tagging of the sentence

I	never	liked	to	work	in	team,	i	rather	prefer	to	work	independently
NN	RB	VBD	TO	VB	IN	NN ,	FW	RB	VBP	TO	VB	RB

Source: Own adjustment based on Stanford Parser

The parse of the sentence is described in Figure 3 and is representing the tree in a transition-based dependency parser (Klein & Manning, 2008)

Figure 3: The parse of the sentence

```
(ROOT
 (S
  (S
   (NP (PRP I))
   (ADVP (RB never))
   (VP (VBD liked)
    (S
     (VP (TO to)
      (VP (VB work)
       (PP (IN in)
        (NP (NN team))))))))
  (, ,)
  (NP (FW I))
  (ADVP (RB rather))
  (VP (VBP prefer)
   (S
    (VP (TO to)
     (VP (VB work)
```

(ADVP (RB independently))))))))

Source: Own adjustment based on Stanford Parser

The dependencies resulted after the parsing using the Stanford Parser (Klein & Manning, 2008), are described in Figure 4.

Figure 4: The dependencies resulted

root (ROOT-0, prefer-11)

nsubj (liked-3, I-1)

neg (liked-3, never-2)

ccomp (prefer-11, liked-3)

aux (work-5, to-4)

xcomp (liked-3, work-5)

prep (work-5, in-6)

pobj (in-6, team-7)

nsubj (prefer-11, i-9)

advmod (prefer-11, rather-10)

aux (work-13, to-12)

xcomp (prefer-11, work-13)

advmod (work-13, independently-14)

Source: Own adjustment based on Stanford Parser

The neural networks parser have been studied previously by defining a simple synchronized network to be applied in the Penn Tree-Bank parser (Henderson, 2004). The neural network predicted the parse decision in a constituency parser. There is evidence that a word representation is used, instead of multiple distributed representation (Marshall et al. 2005).

Sigmoid Belief Networks have been applied in order to parse constituency transformation (Titov & Henderson, 2007). Extending the previous studies, it was introduced a transition-based dependency parser in order to use a Temporal Restricted Boltzmann Machine (Garg & Henderson, 2011). There were developed different neural network architectures, but the advantage of the latest Stanford parser version is the unrestricted vocabulary used. The one of recently study based on neural network for transition-based dependency parsing was made by Stenetorp (2013), and the conclusion was that the empirical performance for the model defined was unsatisfactory.

After multiple experimental evaluations, the Stanford Dependency Parser defined by Klein & Manning (2008) exceeds other dependency parsers in accuracy and speed.

The innovation of the Dependency Parser is given by the grammar representation of words, POS tags and links between. The use of the Dependency Parser in assessment of competency represents an innovative method that can be applied in establishing an evaluation score of the opinion expressed by the IT professional.

4 Implementing Stanford Dependency Parser in competency assessment

Measuring the success of an organization involve measuring the PM competency of the employees, the firms' performance and projects' success. Extensive research has been done and the conclusion was that the human factor is the most important factor that influences the success of an IT organization (Bodea & Dascalu, 2010).

The development of the web based application that assesses the competencies of the IT professionals is based on the PM competency model defined by Bodea & Toader (2013).

The PM competency model consists of three factors of influences: methodical, personal-social and strategic-organizational grouping eighteen PM competencies: knowledge of applied PM methods; technical analysis of information; identifying problems and finding solutions; automating and optimizing work processes; evaluation, review and quality assurance of work; implementation of the maintenance technique; knowledge of software requirements; teamwork; creativity; vigilance; efficiency; motivation; ethics; stress resistance; permanent organization; health, security, safety and work environment; ideas to improve work; respect of work methods and procedures.

The principal aim of the study was to develop a web-based system for assessing the performance of IT specialists and implementing Stanford Dependency Parser for evaluating the opinion expressed by the professionals about acquiring certain competencies. Dependency Parser can be useful for the assessment of the PM competency in determining a measure of an opinion, for assuring an accuracy result and a transparent process.

The application components are: the database developed in MySQL tool that contains: the competency ontologies defined by Bodea & Toader (2012), the employees' information, all the questions, answers that assessment contains, the answers provided by the employees during the evaluation, the score obtained for each response and the feedback for each answer.

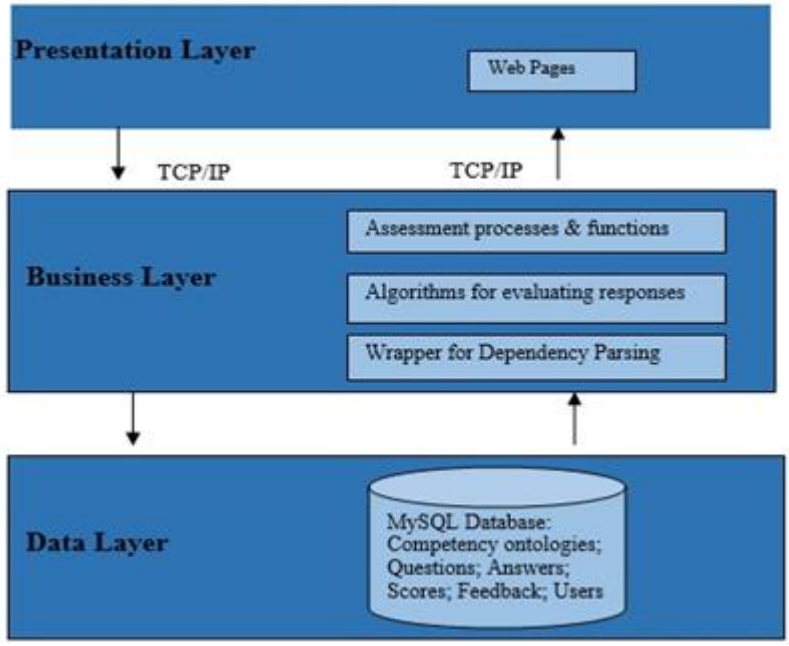
The competency ontology combines concepts of knowledge, skills, target population and performance indicators for IT domain. Also provides ways for modeling IT competencies, for defining competencies and an evaluation criteria for skill performance (Bodea & Toader, 2012).

The business layer contains assessment processes and functions, algorithms for responses evaluation and a Wrapper for Dependency Parsing which contains the access to Stanford Dependency Parser.

The presentation layer contains the web pages through is made the assessment. The application architecture is outlined in Figure 5.

The application is based on modules for each participant on the evaluation flow: employee, employer and system administrator. The application was developed in PHP programming language and assesses all the competencies from the PM competency model defined by Bodea & Toader (2013).

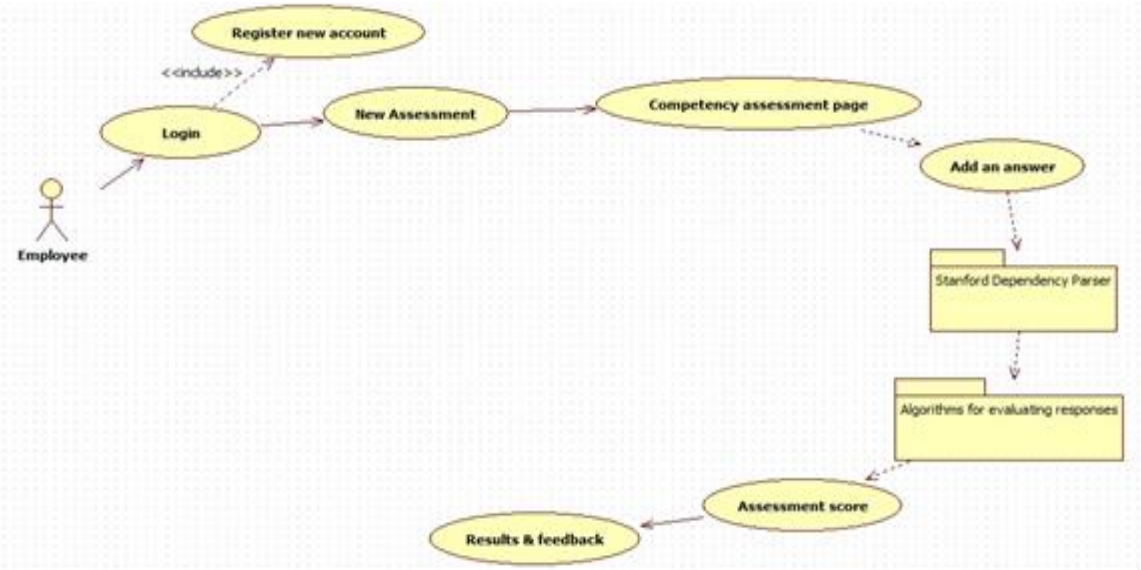
Figure 5: The Application architecture



Source: Own development

The employee module offers the following features: the login into the application (users have the possibility to register a new account) is required. Then, the employee can take a new assessment and has to respond to an open question related to a competency level it holds when performing daily tasks. After he has expressed the opinion related to a certain competency, the answer is parsed with the Stanford Dependency Parser. The parsing score obtained will be integrated in different algorithms for evaluating responses, resulting a final assessment score. At the end of the evaluation, there is a results page containing a table with the given response for the question and a feedback as an online resource link that the employee must access it in order to improve the competency level. In Figure 6 is described a use case for the assessment of an open question and response evaluated by Dependency Parser.

Figure 6: A use case example for the assessment of a competency



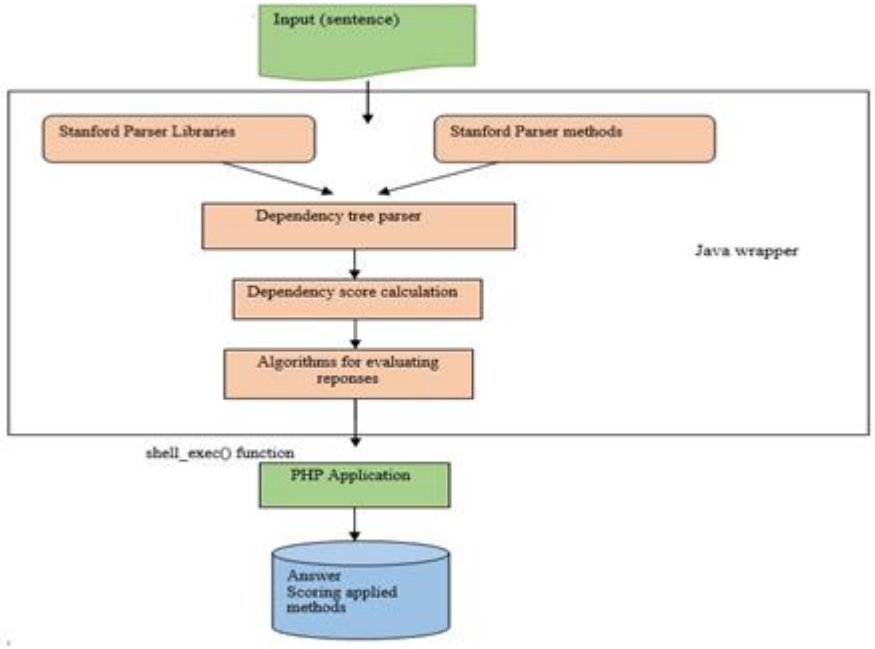
Source: Own development

5 The Dependency Parser wrapper

The wrapper for parsing the opinion expressed was developed in Java, using the Stanford Parser libraries. The version of the Stanford Parser used is 3.5.0 and it was released in October 2014 (available http://nlp.stanford.edu/software/lexparser.shtml#History).

The connection between the java wrapper and the PHP application where the assessment is made, is carried out through the invocation of „shell_exec" php function. This function invokes the parser. In Figure 6 is described the architecture of the Wrapper for Stanford Dependency Parser.

Figure 7: Dependency Parser wrapper architecture



Source: Own development

There are a variety of tools and also indicators that can be used to evaluate performance of IT professionals, but the major problem is that they are not effective in measuring competency and they don't assure the objectivity of the process. The assessment tool is embedding different innovative methods to assess the competency and assure the validity and transparency of the process.

6 Conclusions

This study investigated the implementation of the Stanford Dependency Parser for evaluating the opinion expressed by the professionals about acquiring certain competencies.

The Stanford Dependency Parser has many advantages: can be applied on a large variety of domains and also presents high accuracy and efficiency. The latest Stanford Parser is based on neural network algorithms which offers a high speed of text parsing and a good performance. A web-based application tool was developed in order to assess the competencies of the IT

professionals. The architecture of the application was outlined as well as the use case of one module: employee in assessing a competency by expressing an opinion related to it.

The Wrapper for Dependency Parsing includes the Stanford Parser libraries and can measure an opinion about the competency level acquired by an IT professional. The limitation of the current implementation is that the application does not contain algorithms for evaluating responses. The benefit of the proposed assessment method is the reducing time of the entire process and the accuracy of the opinion score. Future studies will include the analysis and implementation of different methods for analyzing the responses.

Acknowledgment

This papers was co-financed from the European Social Fund, through the Sectorial Operational Programme Human Resources Development 2007-2013, project number POSDRU/159/1.5/S/138907 "Excellence in scientific interdisciplinary research, doctoral and postdoctoral, in the economic, social and medical fields -EXCELIS", coordinator The Bucharest University of Economic Studies.

Reference

McDonald, R., K. Crammer, and F. Pereira. 2005. Online large-margin training of dependency parsers. In Proceedings of ACL, pp. 91–98.

Marneffe Marie-Catherine and Christopher D. Manning, Stanford typed dependencies manual, 2011.

R. Hudson. 1984. Word Grammar. Blackwell.

Dan Klein and Christopher D. Manning. 2003. Accurate Unlexicalized Parsing. Proceedings of the 41st Meeting of the Association for Computational Linguistics, pp. 423-430, Available: http://nlp.stanford.edu:8080/corenlp/

C. Bodea, E-A. Toader, “Development of the PM competency model for IT professionals, base for HR management in software organizations”, 12th International Conference on Informatics in Economy (IE 2013), Education, Research and business Technologies, Bucharest, April 2013

Danqi Chen and Christopher Manning. 2014. A Fast and Accurate Dependency Parser Using Neural Networks. In Proceedings of EMNLP, 2014.

Terry Koo, Xavier Carreras, and Michael Collins. 2008. Simple semi-supervised dependency parsing. In ACL

He He, Hal Daum´e III, and Jason Eisner. 2013. Dynamic feature selection for dependency parsing. In EMNLP.

James Henderson. 2004. Discriminative training of a neural network statistical parser. In ACL.

Marshall R. Mayberry III and Risto Miikkulainen. 2005. Broad-coverage parsing with neural networks. Neural Processing Letters.

Ivan Titov and James Henderson. 2007. Fast and robust multilingual dependency parsing with a generative latent variable model. In EMNLP-CoNLL.

Nikhil Garg and James Henderson. 2011. Temporal restricted Boltzmann machines for dependency parsing. In ACL-HLT.

Pontus Stenetorp. 2013. Transition-based dependency parsing using recursive neural networks. In NIPS Workshop on Deep Learning.

Xavier L., Marquez L., 2008 A Joint Model for Parsing Syntactic and Semantic Dependencies, Proceedings of CoNLL-2008

Klein D., Manning C. D., 2003. Fast Exact Inference with a Factored Model for Natural Language Parsing. In Suzanna Becker, Sebastian Thrun, and Klaus Obermayer (Eds), Advances in Neural Information Processing Systems 15 (NIPS 2002). Cambridge, MA: MIT Press, pp. 3-10.

Bodea C. N., Dascalu M. - Competency-based E-assessment in Project Management and Firm Performance: a Case Study, Proceedings of the 4th EUROPEAN COMPUTING CONFERENCE (WSEAS), Bucuresti, Romania, 20-22 aprilie 2010, pp. 76-81

C. Bodea, E-A. Toader, “Ontology-based modeling of the professional competencies - a comparative analysis”, 11th International Conference on Informatics in Economy (IE 2012), Education, Research and business Technologies, Bucharest, May 2012, pp. 452-458;