

Refining an Ontology by Learning Stakeholder Votes from their Texts

Olga Tatarintseva^{1,2} and Vadim Ermolayev¹

¹ Department of IT, Zaporozhye National University,
66 Zhukovskogo st., 69063, Zaporozhye, Ukraine

tatarintseva@znu.edu.ua, vadim@ermolayev.com

² Satelliz, 158 Lenina st., P.O. Box 317, 69057, Zaporozhye, Ukraine

Abstract. This paper reports on our experiments evaluating the improvement of OntoElect approach to ontology refinement in the case study with the ICTERI Scope Ontology. OntoElect is based on collecting and assessing the commitment of domain knowledge stakeholders for ontological refinement offerings. We report the improvement with respect to the previous results. Our first experiment evaluates the change in the quality of ontology due to the involvement of domain knowledge stakeholders in semantic annotation of their papers, compared to the previous study in which the annotations were done by knowledge engineers. Our second experiment checks if the result became better after the introduction of the automated term extraction from the full texts of ICTERI papers. Extracted terms are compared to the manual annotations. The results of the experiments verify the proposed ontology changes and are further used for the ICTERI Scope ontology refinement.

Keywords. ICTERI Scope ontology, ontology engineering, domain knowledge stakeholder, term mining, evaluation, refinement

Key terms. KnowledgeEngineeringMethodology, SubjectExpert, SubjectDomain, Metric, Ontology

1 Introduction

Maintaining an ontology in its lifecycle that fits all the requirements of the subject domain stakeholders is a complicated task in ontology engineering which does not have a complete solution so far. The problem is to a large extent in devising a methodology for ontology refinement that enables a complete and timely account for those requirements and maps them to the updated revision of the ontology. One complication is that the stakeholders who own the requirements need to be committed to provide their inputs for ontology refinement. Furthermore, those inputs need to be meas-

ured and applied correspondingly to the utility of their contribution and in a harmonized way to ensure the consistency and validity of result.

This paper reports on the improvement of our OntoElect approach for iterative ontology refinement [1, 2]. The approach has been proposed in [1] using the allusion of elections in which different “ontology offerings” compete for the commitment of the pool of the relevant domain knowledge stakeholders being the “electorate”. OntoElect has been basically validated in an experiment reported in [2] where the approach was detailed by offering voting metrics for the ICTERI ontology built and refined iteratively based on semantic annotation of the pool of papers of ICTERI 2011 conference.

The results of our previous experiment suggested several important technical aspects [2] for improving OntoElect methodology as a whole and the accuracy of our measurements in particular. Some of those aspects have been addressed in the work reported in this paper.

Firstly, our previous experiment was based substantially on the manual annotation of papers. A knowledge engineer assigned key terms or suggested missing terms based on her personal interpretation of the abstract of a paper. By that we mimicked voting by paper authors while keeping them free of extra annotation effort. The lowlights of this approach to annotation were that:

- Domain knowledge stakeholders (paper authors) were in fact not involved in the workflow and therefore not motivated to be committed to the resulting ontology refinement
- The quality of semantic annotations we obtained has been perceived as fairly low because (a) done by a knowledge engineer who is not a subject expert with respect to the annotated paper; (b) the source for this work was just an abstract, not a paper, and its meaning has been interpreted by a knowledge engineer.

To overcome those shortcomings we first decided to involve the authors more actively by requesting that they themselves semantically annotate their submissions to ICTERI 2012¹. It has also been considered as promising to refine the approach by automated extraction of terms from the papers authored by our subject domain knowledge stakeholders. Here we present the results of our experiments which checked how these two refinements helped improving the quality and adequacy of the ICTERI Scope ontology.

Further, the document corpus used in the previous experiment was fairly small in size for assuring reliable judgements about the opinion of the stakeholder community. For improving on that we continued the collection of ICTERI papers which has been extended by all papers of ICTERI 2012.

We first repeat the previous experiment [2] based however on the document corpus of ICTERI 2012 papers semantically annotated by their authors. We then focus on answering the question about annotation quality by: (i) performing automated term extraction from the full texts of our complete document corpus (ICTERI 2011 and 2012); and (ii) comparing the results of automated term extraction to the outputs of manual semantic annotation.

¹ See <http://isrg.kit.znu.edu.ua/icteriwiki/index.php/ICTERI-Terms>

The remainder of the paper is structured as follows. Section 2 briefly reviews the related work in relevant fields. Section 3 outlines the OntoElect approach to ontology refinement and presents the case study dealing with ICTERI Scope Ontology as well as the document corpus at our disposal. Section 4 sets up our experiments by describing the workflow, evaluation metrics, and used tools. Section 5 presents and discusses the results of our experiments. The paper is further concluded and our plans for the future work are outlined.

2 Related Work

One of the possible ways to check if a conceptualization of a domain is correct and complete is to evaluate the model against the interpretation of the meaning of the representative set of relevant documents. The document corpus will be relevant and representative if it covers the majority of the views by the domain knowledge stakeholders. Their interpretations may be collected and further analysed for refining the ontology using different techniques which may be sought in several areas of research and development. In this section we briefly outline the related work in the relevant fields of research and refer to our previous publication [1] for a more in-depth and detailed coverage.

One of the popular relevant research areas studying how interpretations are collected is collaborative or social tagging and annotation. A good survey of the field is [3] where the use of tags for different purposes and associated shortcomings are analysed. Semantic annotation and tagging approaches further refine social tagging techniques by offering the collections of terms that are taken from taxonomies, folksonomies, or thesauri [4]. Hybrid approaches for collaborative tagging and annotation aiming at the enrichment of seed knowledge representations by a user community are reported for example in [5].

One of the promising approaches focused, besides collecting interpretations or subjective conceptualizations, on motivating more people to take part in developing or refining ontologies is offering a game with a purpose to intended users. Following this approach, ontology development or refinement can be implicitly embedded in a game software. There ontology elements are created, updated, and validated implicitly in the background [6]. Gaming approach has also been tried for evaluating how well ontological specifications fit to the interpretations of random users (FACTory Game by Cycorp, <http://game.cyc.com/>). Several game scenarios have been developed [5] for ontology building and refinement, ontology matching, annotating content using lightweight ontologies. Those are similar to our OntoElect approach. Both approaches offer possibilities to identify whenever users start to agree on and share commitment to certain ontological items.

Social and gaming approaches that involve the direct participation of human stakeholders are complemented by the plethora of research results in automated knowledge extraction or ontology learning. This strand of research involves the stakeholders indirectly – through making use of their professional outputs, like authored texts. A comprehensive survey of the techniques used to learn ontologies from texts is [7]. In

the second experiment we present in this paper only term extraction using the TerMine tool [8] has been performed.

Yet one more important aspect in developing or refining an ontology is the re-use of the other ontologies or their most relevant parts to the developed ontology. In this context the ontology meaning summarization approach [#] makes good sense for helping an ontology engineer choose the most relevant and valuable parts for re-use. The approach is based on detecting the “key concepts” of an ontology under analysis which best characterize its meaning. The key concepts are determined using a combination of criteria from lexical statistics, taxonomy graph analysis, and popularity based on a number of hits. Especially in using the popularity and coverage metrics, this approach coincides well with our approach (OntoElect). OntoElect is however used not for summarizing but refining an ontology based, among other things, on assessing the coverage and popularity of the Key Terms. Besides that the mechanisms of obtaining the measures are different. From the other hand, OntoElect does not yet consider ontology re-use as one of important mechanisms for refinement. Hence, combining some features of [9] in OntoElect may be enriching.

3 ICTERI Case Study

The idea of OntoElect approach [1] was inspired by public election campaigns. Just as the leader in a public election campaign gets the major part of the electorate’s commitment to win, the extent of the domain knowledge stakeholders’ commitment hints about the quality and completeness of the ontology. Following this allusion, the votes of the domain knowledge stakeholders for alternative ontology offerings are collected and used as the measure of their commitment. The ontology offering that collects the biggest share of votes could therefore be considered as the best and most complete.

In our case study the OntoElect approach is applied for refining the ICTERI Scope Ontology in the iterative ontology engineering experiment. Our domain knowledge stakeholders are the authors of ICTERI papers. Ontology offerings in the reported work are the structural contexts² in the five thematic areas of the ICTERI scope offered to the authors for choosing the appropriate key terms to annotate their papers.

As this data had to be selected we decided to simulate the opinions of the electorate by annotating the papers of ICTERI 2011 manually. For this we extracted the terms which were specified as the list of ICTERI Key Terms if it was possible. In some cases we had to add Missing Concepts (also called Missing Key Terms) for the papers, if such terms did not exist in the list.

So, we received three semantic annotation types:

- KeyWord – for the key words, which were selected by the authors
- KeyTerm – for the terms which were selected manually and were found in the list of the ICTERI terms

² A structural context, as suggested e.g. in [10], is composed of a central concept with all his domain and object properties and the concepts connected to the central concept by these object properties.

- MissingConcept - for the terms, which did not exist in the list of the ICTERI terms, but were covered during annotation

One use of the particular term was considered to be one vote for the selected term. The votes were normalized as frequencies of use. Such information allowed us to measure the popularity of each semantic context, circumscribe the most frequently demanded part of the ontology and make suggestion about the completeness of the ontological offerings.

For the papers of ICTERI 2012 we requested that the authors annotate their papers not only using the freely chosen key words, but also using the terms found in ICTERI scope ontology. As the result the corpus for the further analysis was increased. The data provided by the authors can be accepted as more authentic than that which was obtained ourselves, as they are the real domain experts for the field they study. The analysis of the received data is presented in Section 5.

But even when we use the information presented by the authors, and the results of our manual annotation we can't guarantee that this information is accurate enough for applying it to the ontology refinement process. To obtain the experiment we needed to have results received in different ways because we wanted to achieve the impartial assessment of OntoElect approach. Before applying the results in ontology refining process we decided to check them with freely available tool for text mining.

For our experiment we chose one of the services provided by the National Centre for Text Mining (NaCTeM). As reported in the official website of NaCTeM³ it is the first publicly-funded text mining centre in the world. It provides text mining services in response to the requirements of the UK academic community. NaCTeM is operated by the University of Manchester.

4 Experimental Set-up and Tools

To control the results of the experiment we have to understand which main questions we are going to answer after its realization and how to measure these results.

Our measurable objectives for the experiment have been formulated as follows [2]:

- Does the ontology fit to the requirements of the subject experts in the domain?

The fitness of the ontological offering will be measured as a ratio of the average frequency of use of the available Key Terms (positive votes) to the similar for the missing Key Terms (negative votes). Special attention will be paid to the freely chosen key words that are identical to the available Key Terms. Those will be considered as extra positive votes for the semantic context of the Key Term.

- Is there a particular part in the ontology that is the most important for the stakeholders?

The importance of an ontology fragment comprising particular concepts will be measured as frequency of use of these concepts (positive votes). Fragments of different importance will also be presented as percentiles.

³ Official web site of National Centre for Text Mining <http://www.nactem.ac.uk/>

- Is there a part in the ontology that could be dropped as the stakeholders do not really require it?

Similarly to importance, these ontology fragments will be outlined using low frequency of use percentiles.

- What would be a most valuable addition to the ontology that will substantially improve stakeholders’ commitment to it?

The papers have been annotated using missing Key Terms and freely chosen keywords. Those missing Key Terms that are frequently used will form the core of this effective extension. If some of the keywords are also used frequently by the authors they may become good candidates for the inclusion in the effective extension as well. Special attention will be paid to the freely chosen key words that are identical to the missing Key Terms. Those will reinforce the votes on the addition to the ontology.

The flow of activities has been organized in three consecutive phases as presented in Fig. 1. The description of each phase in details is presented in [2].

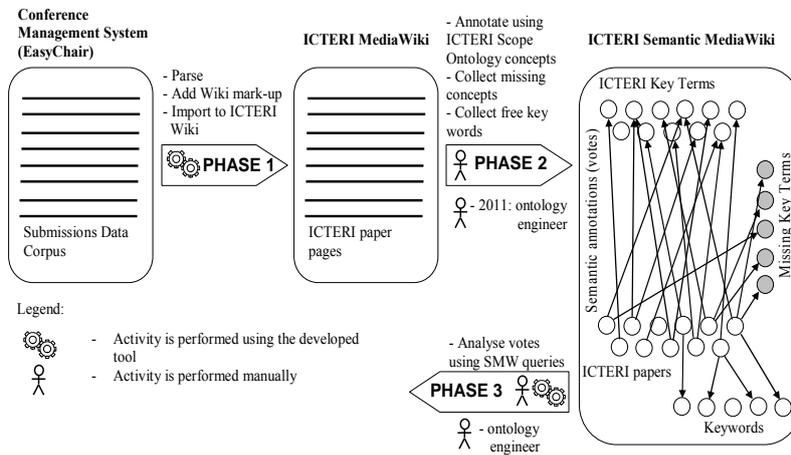


Fig. 1. The workflow for processing ICTERI papers and collecting stakeholders’ votes repeated according to [2]

At phase 1 we have extracted the semi-structured information about the papers accepted for ICTERI 2012 and transformed these into the collection of paper articles in the ICTERI Wiki. At Phase 2 we extracted the freely chosen KeyWords and the KeyTerms from the ICTERI Scope ontology assigned by the authors and added these to the semantic annotations of the papers. In several cases we detected considerable meaning gaps between the extracted key words and Key Terms when annotated the papers manually. Therefore, we opted to add the missing Key Terms to the corresponding semantic annotations. As a result of this Phase the semantic relationships between the pages of `Category:Paper` and the pages of `Category:Concept` have been specified as semantic properties. These semantic properties allowed us to receive all the measurements planned for the evaluation experiment. These measurements have been done using different Semantic MediaWiki queries at Phase 3.

Compared to the previous year experiment [2], we automated the extraction of the frequency of use statistics which made the process less error prone and faster. For that the SMWAskAPI⁴ extension of the Semantic MediaWiki has been used. This extension supports semantic queries of #ask and enables the use of the corresponding API for executing Semantic MediaWiki ask queries.

Each page of the ICTERI Wiki uses semantic tagging. An example of the Semantic properties specified for pages in `Category:Paper` is given in Fig. 2.

For our analysis we used the pages from `Category:Paper` and `Category:Workshop` with the property `hasPublicationYear` equal to 2012, namely the values of the semantic properties `hasKeyWord`, `hasKeyTem`, and `MissingConcept`.

Is Your Ontology a Burden or a Gem? – Towards Xtreme Ontology Engineering	
HasAuthor	Olga Tatarintseva + ⓘ , Vadim Ermolayev + ⓘ , Anna Fensel + ⓘ
HasKeyTerm	KnowledgeEngineeringMethodology + ⓘ , SubjectExpert + ⓘ , Collaboration + ⓘ , Approach + ⓘ
HasKeyWord	Ontology + ⓘ , Stakeholder commitment + ⓘ , Collaboration + ⓘ , Ontology engineering + ⓘ , Ontology election + ⓘ
HasLanguage	English + ⓘ
HasPresentation	Presentation: Is Your Ontology a Burden or a Gem? – Towards Xtreme Ontology Engineering + ⓘ
HasPublicationYear	2011 + ⓘ
HasTitle	Is Your Ontology a Burden or a Gem? – Towards Xtreme Ontology Engineering + ⓘ
MissingConcept	ConceptualModeling + ⓘ
Modification date	31 March 2012 18:59:28 + ⓘ
PublicationURL	Http://oeur-ws.org/Vol-716/ICTERI-2011-CEUR-WS-paper-4-p-65-81.pdf + ⓘ
Categories	Paper
<small>hide properties that link here</small>	
Is Your Ontology a Burden or a Gem? – Towards Xtreme Ontology Engineering + ⓘ , Presentation: Is Your Ontology a Burden or a Gem? – Towards Xtreme Ontology Engineering + ⓘ	HasTitle
Burden-or-Gem + ⓘ	redirect page

Fig. 2. Semantic properties for the ICTERI Wiki page in the `Category:Paper`

The scripts for analyzing these values were coded in Python. Some steps were also implemented using shell scripting. As outputs we have received:

- The list of KeyWords, KeyTerms, and MissingConcepts for each article, if they were defined
- The overall number of the papers according to the values of the properties `hasPublicationYear`, and the selected `Category`
- The number occurrences of each KeyWord, KeyTerm and MissingConcept.

The analysis and discussion of the results is given in Section 5.

To perform our second experiment we applied the Term Management System named TerMine, which identifies key phrases in text. It uses C-value [8], a domain-independent method for automatic term recognition (ATR) which combines linguistic and statistical analyses with the emphasis on the statistical part. The linguistic analysis enumerates all candidate terms in a given text by applying part-of-speech tagging, extracting word sequences based on adjectives/nouns, and stop-list. The statistical

⁴ See the description of the SMWAskAPI on <http://www.mediawiki.org/wiki/Extension:SMWAskAPI>

analysis assigns a candidate term to a termhood by using the following four characteristics:

- The occurrence frequency of the candidate term
- The frequency of the candidate term as part of other longer candidate terms
- The number of these longer candidate terms
- The length of the candidate term

The data corpus for this term extraction and analysis was the merge of the pools of ICTERI 2011⁵ and ICTERI 2012⁶ papers published in the respective proceedings, and consisted of 63 papers. The papers from both proceedings volumes have been merged in a single file and uploaded for processing by TerMine. The workflow for the second experiment is pictured in Fig. 3.

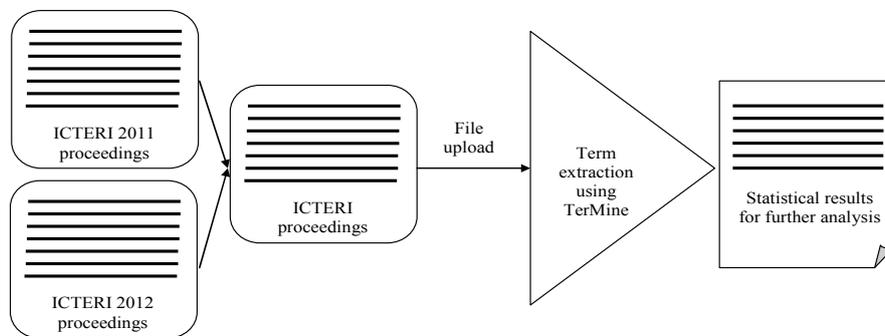


Fig. 3. The workflow for conducting the second experiment for term mining and analysis

The data processed in the pipeline is illustrated by the example of a single paper [1] in Fig. 4.

The results of term mining were provided in several forms. All the terms defined in the text were highlighted by colour markings (upper part of Fig. 4a). The information about the overall number of the terms mined from the text was also given (433 terms listed – in the bottom of Fig. 4a). The terms were also presented in the table view, each preceded with the assigned rank number and followed by the statistical score measure (lower part of Fig. 4a). The rank of a term means the position of each term in the table sorted by the score; the rank values of the terms with the same score are equal. The scores were computed automatically using the Term Recognition technique [8] which uses the information about the frequencies of term occurrence. This approach is essentially a shallow bag of terms extraction technique – therefore the output needs to be post-processed as described using our single paper data example. For this example the number of extracted terms was 433 which is obviously too many. To compare, the authors were advised to assign 3-5 KeyTerms to their papers which best describe its meaning. Among those extracted terms that we needed to sort out were also names, affiliations, cities, etc, which had no semantic relationship to the

⁵ <http://ceur-ws.org/Vol-716/>

⁶ <http://ceur-ws.org/Vol-848/>

meaning of the paper. Also, it has been assumed that the terms with a low number of occurrences in text have a negligent semantic contribution and may also be filtered out – so only the higher ranked part of the term list may be considered.

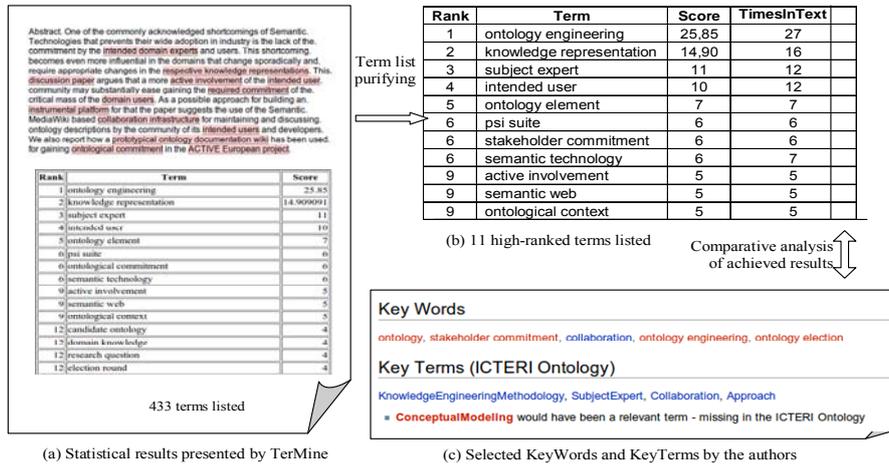


Fig. 4. An example of the data processed in the term mining experiment

While post-processing the list of mined terms we decided to leave only the terms, which were used more than 5 times, and had score more than 5 points. Applying this threshold returned 11 of 433 terms for the example outlined in Fig. 4., which constitutes only 2.54 per cent of the overall number of the mined terms. The manual check of the example however indicates that these 11 high ranked terms indeed contribute most significantly to describing the semantics of the corresponding paper (Fig. 4b).

The right part of Fig. 4 allows to compare the result of term extraction (Fig. 4b) with the output of manual semantic annotation (Fig. 4c) for the selected example paper. A mechanical comparison reveals substantial difference, which however is not that big after manual mapping of the extracted terms to the concepts of the ICTERI Scope ontology. In fact there is a subset of extracted terms that could be directly mapped into the Key Terms of the ontology: subject expert; ontology engineering (as a methodology). Another group is relevant to the assigned KeyWords: ontology, stakeholder commitment, ontology engineering. Some are synonymic in the context of this paper: stakeholder and intended user. Some represent the meaning which is too fine-grained for a semantic annotation: ontology element. And, which is most important, some are the new valid candidates for the inclusion into the ICTERI Scope ontology: knowledge representation, semantic technology, semantic web.

5 Results and Discussion

In this section we present the results of the experiment. The set up of all its stages is described in Section 4. The discussion of the experiment results is structured along the measurable items.

The frequency of use diagrams (Fig. 5 and 6) are built in regard to the total amount of the papers and the number of occurrences of a particular term.

Similar work was reported in our previous publication [2]. In it we described the mechanism of ranging. The actual experiment is based on the previous results. But they changed as the document corpus we are working with has increased and the data to work with has changed. As we consider OntoElect approach in the case study of iterative refinement of the ICTERI Scope Ontology such changes are greatly important. The diagrams which show these changes are shown below:

- For the KeyWords which were selected by the authors manually (only those, which were chosen by at least two authors, Fig. 5)
- For the KeyTerms, which were selected from the ICTERI ontology terms (Fig. 6)

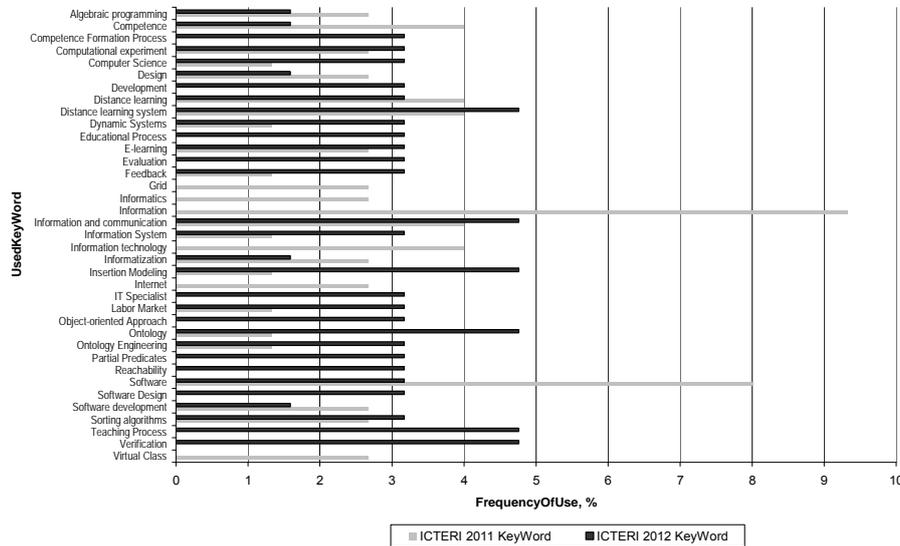


Fig. 5. The frequency of use of the freely chosen KeyWords

We did not provide a frequency of use diagram comparing the Missing Key Terms because the difference in the results of 2011 and 2012 is tiny and could be neglected. We provided the comparison analysis for KeyWords and Missing Key Terms lists (Fig. 7). To compute the range of use of each term we divided each frequency of use index by the frequency value of the most popular term, which range of use was taken as 100 per cent. These terms are not the part of the ontology, but are the most possible candidates.

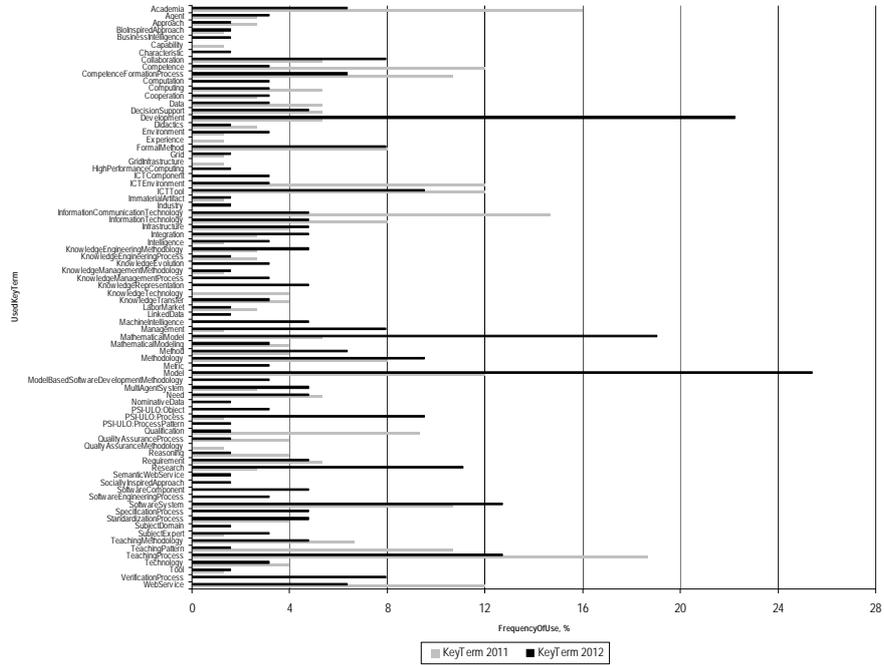


Fig. 6. The frequency of use of available KeyTerms

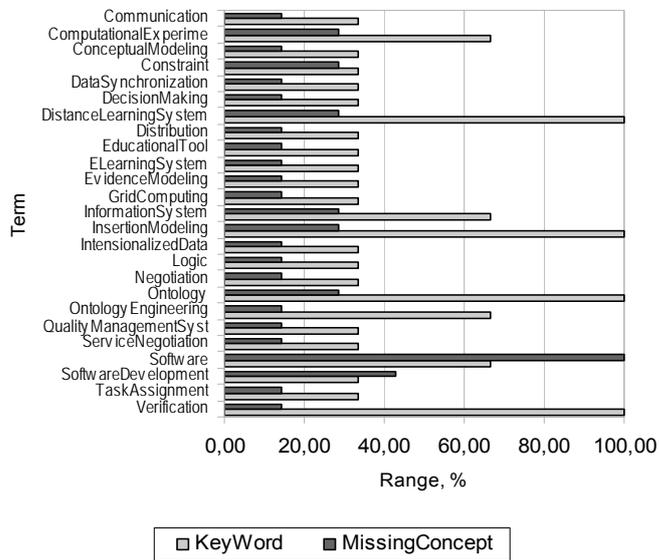


Fig. 7. The range of use for Missing Key Terms and KeyWords

The results which we received at the previous step can be merged and we can receive the new version of the potential ontology offering. But before doing this we will look into the results of the text mining experiment.

Using TerMine tool for automation of the knowledge mining process we received some interesting results. The overall number of the found terms was 8487. All of the selected terms were graduated and received the position in the rank table. Studying the results it is obvious that the number of the terms selected by the data mining tool is too big.

It was decided to leave in the rank table only those terms which have the score of 10 and more. The popularity of the terms which were picked up is evident. The total number of such terms is 157. It is easy to count that it makes up less than 2% of all the terms proposed by the tool. After refining the list and deletion of the superfluous information only 140 terms left.

To compare the frequencies of use provided by the TerMine and our own calculations we decided to use percentage method (similar to that used for building the diagram in Fig. 7). We took the maximal value for each group of concepts as 100 per cent and divided it by the frequency of use value of a particular term. As a result each term got the value, called the range, which could be compared with the ranges of the other terms. We analyzed the three groups of mined term matches to the: (i) KeyWords; (ii) MissingConcepts; and (iii) KeyTerms. As the number of the Missing Concepts is not too big we decided to combine them with the Key Words in the diagram (Fig. 8).

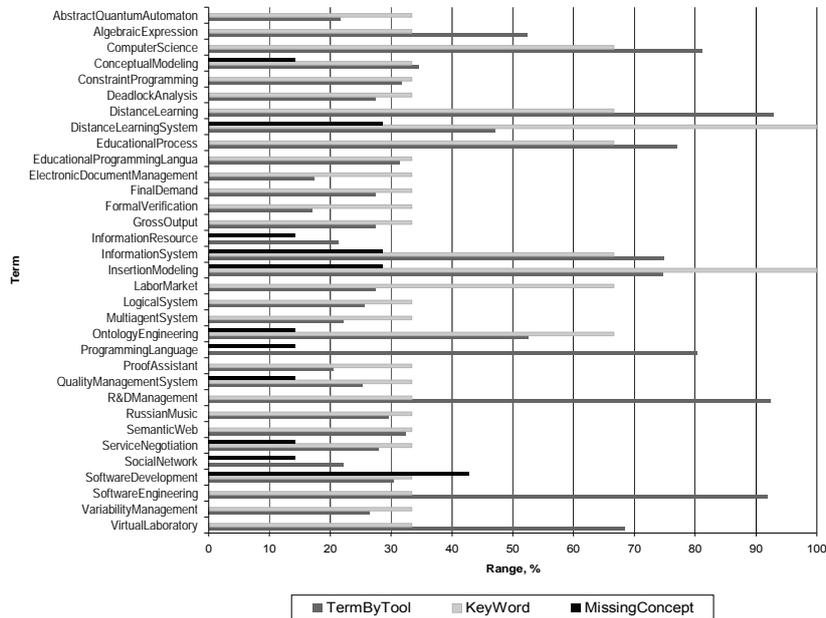


Fig. 8. The range of use for terms detected by tool, Key Words and Missing Concepts. Range values were normalized by the frequency of use of the highest scored extracted term (100)

After the automated search of the identical terms in the lists of KeyTerms and terms mined by tool we discovered that some of them were missed as the search did not use the rules of common sense and the relations described in the ontology. For example, according to the ICTERI Scope ontology the term Integration subsumes to PSI-ULO:Process. Knowing this fact we understand that the term IntegrationProcess is just the same as the term Integration. But this match is not obvious for the simple search and will not be detected.

Therefore, to find the matches in the lists of KeyTerms and terms mined by the tool we decided to undertake a more careful analysis. We scanned the list of the KeyTerms for matching the ToolTerms manually. Besides for this process we used the whole pool of 8487 terms mined by the tool. The result is pictured in the diagram (Fig. 9).

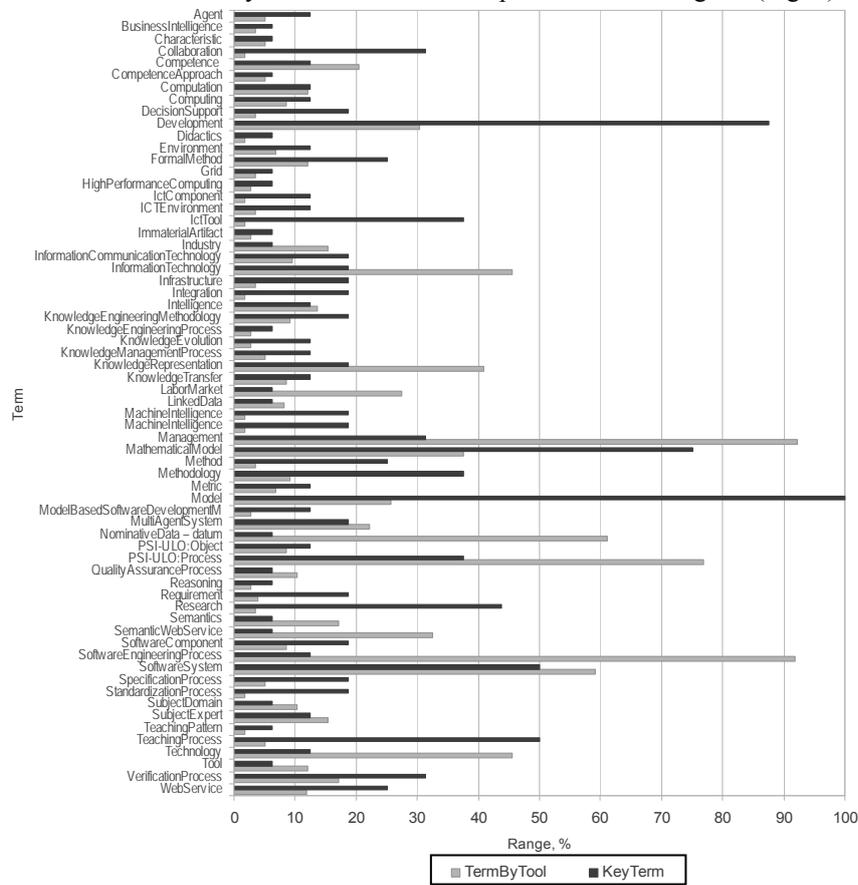


Fig. 9. The range of use for the KeyTerms and the terms extracted by the TerMine tool

6 Concluding Remarks and Future Work

The paper reported on the experiment evaluating the improvement of OntoElect approach to ontology engineering in the case study with the ICTERI Scope Ontology. In particular, the approach has been used to evaluate the validity of the papers' annotation by drawing knowledge stakeholders to their own papers' annotation and by studying their quality using voting for full texts.

The experiment consisted of two stages. The first one was based on the comparison analysis of the papers presented during international conferences ICTERI 2011 and ICTERI 2012. The second one was dedicated to performing automated term extraction from the full texts and comparing the results of automated term extraction to the outputs of manual semantic annotation.

Achieved results stress the important parts of the ontology and those which are less popular among the authors. The comparison analysis of the first experiment shows how the situation changed during two years. The KeyWords and MissingConcepts which have high frequency of use values, especially if they are named in both lists, are the first candidates to become the new part of the ontology.

The second experiment shows which ontological offerings agree with the terms mined by tool and which numeric characteristics these matches have. The terms extracted by the tool and their matches with the KeyWords and MissingConcepts, which have range more than 50 per cent, are also good candidates to be added to the ontology. Besides, the degree to which the extracted terms match the KeyWords and KeyTerms indicate about the adequacy of paper annotation. Overall the overlap between the meanings of the extracted terms and the KeyTerms measures the range of so to say the similarity in the meanings of the papers within the corpus and the ontological offerings aimed at covering these meanings. The quantitative results of our experiments still need to be processed and analyzed more thoroughly before deciding about the implementation of the changes to the ontology. Besides that several other aspects still need to be researched in our future work.

Firstly, the document corpus used in the case study, though growing, is still not very big to allow robustly applying the majority of traditional knowledge extraction techniques. At the moment it could only be stated that the information we have now is enough to prove the concept – i.e. the validity of the approach based on the assessment of and account for domain knowledge stakeholder opinions, implicitly reflecting their needs. After applying the refinements suggested by the stakeholders, the ontology still needs to be evaluated and validated using other methods.

Secondly, in this paper we reported about only a partial and shallow way of extracting knowledge from paper texts. A possible refinement to this preliminary solution could be sought in using a hybrid iterative knowledge extraction workflow that incrementally adds ontology elements to the “ontology learning layer cake (c.f. [11])”.

References

1. Tatarintseva, O., Ermolayev, V., Fensel, A.: Is Your Ontology a Burden or a Gem? – Towards Xtreme Ontology Engineering. In: Ermolayev, V. et al. (eds.) Proc. 7-th Int. Conf. ICTERI 2011, Kherson, Ukraine, May 4-7, 2011, CEUR-WS.org, vol-716, ISSN 1613-0073, 65–81, online (2011)
2. Tatarintseva, O., Borue, Yu., and Ermolayev, V.: OntoElect Approach for Iterative Ontology Refinement: a Case Study with ICTERI Scope Ontology. In: Ermolayev, V. et al. (eds.) Proc. 8-th Int. Conf. ICTERI 2012, Kherson, Ukraine, June 6-10, 2012, CEUR-WS.org, vol-848, ISSN 1613-0073, 244, online (2011)
3. Gupta, M., Li, R., Yin, Z., Han, J.: Survey on Social Tagging Techniques. SIGKDD Explorations 12(1), 58–72 (2010)
4. Uren, V., Cimiano, P., Iria, J., Handschuh, S., Vargas-Vera, M., Motta, E., Ciravegna, F.: Semantic annotation for knowledge management: Requirements and a survey of the state of the art. *Science. Services and Agents on the World Wide Web* 4(1), 14–28 (2006)
5. Hunter, J., Khan, I., Gewrber, A.: HarvANA – Harvesting Community Tags to Enrich Collection Metadata. In: Paepcke A, Borbiha J, Naaman M (eds.) 8th ACM/IEEE-CS Joint Conference on Digital Libraries, 147–156. ACM New York, New York (2008)
6. Siorpaes, K., Hepp, M.: Games with a Purpose for the Semantic Web. *IEEE Intelligent Systems* 23(3), 50–60 (2008)
7. Wong, W., Liu, W., and Bennamoun, M.: Ontology learning from text: A look back and into the future. *ACM Comput. Surv.*, 44(4), Article 20, 36 pages. DOI=10.1145/2333112.2333115 (September 2012)
8. Frantzi, K., Ananiadou, S. and Mima, H.: Automatic recognition of multi-word terms. *Int. J. of Digital Libraries* 3(2), pp.117-132 (2000)
9. Peroni, S., Motta, E., d'Aquin, M. Identifying Key Concepts in an Ontology, through the Integration of Cognitive Principles with Statistical and Topological Measures. In: Proc. 3rd Asian Semantic Web Conference (ASWC 2008), Dec 08-11, 2008, Bangkok, Thailand (2008)
10. Ermolayev, V., Copylov, A., Keberle, N., Jentzsch, E., Matzke, W.-E.. Using Contexts in Ontology Structural Change Analysis.. In: Ermolayev, V., Gomez-Perez, J.-M., Haase, P., Warren, P, (eds.) CIAO 2010, CEUR-WS, vol. 626 (2010)
11. Wong, W., Liu, W., Bennamoun, M.: Ontology Learning from Text: a Look Back and into the Future. *ACM Comput. Surv.*, 44(4), Article 20, 36 p., <http://doi.acm.org/10.1145/2333112.2333115> (2012)