JICTR International Journal of Information & Communication Technology Research

Volume 6- Number 1- winter 2014 (69-78)

A Semantic Domain-Specific Framework to Assist Researchers in Screening Contents

Maryam Tayefeh Mahmoudi
School of ECE, College of Engineering
University of Tehran, Tehran, Iran
Multi-Media Group
IT Research Faculty, ICT Research Institute (ITRC)
Tehran, Iran
mahmodi@csri.ac.ir

Fattaneh Taghiyareh
School of ECE
College of Engineering
University of Tehran
Tehran, Iran
ftaghiyar@ut.ac.ir

Received: Janruary 9, 2013- Accepted: October 6, 2013

***Abstract*— Screening and selecting appropriate research documents out of enormous existing contents, has drawn a wide range of research activities these days. Some researchers focus on developing automatic content assessment systems, while the others propose and expand some semantic rules and structures to facilitate the assessment process. There exist various content assessment methods which usually consider at least one of syntactic, semantic and structural perspectives through information retrieval or machine learning algorithms. In this paper, a semantic domain-specific framework is presented to assist researchers in their screening, selecting and recommending activities. The proposed framework is equipped with the ontology of key segments to assess various parts of research content, as well as WordNet and domain ontology, to reinforce semantic rules. The proposed framework is examined on a dataset of contents, and is also compared to the experts' assessment of the same research materials. The comparison results reveal that the proposed semantic researcher-assisting framework has been successful in almost 70% of cases.**

Keywords- Content screening; content assessment; content review; researcher assisting system; semantic rules; ontology.

I. INTRODUCTION

Screening, selecting and recommending appropriate research and educational materials from huge amount of contents has turned into a serious topic for researchers. Taking advantage of some automatic content assessment algorithms or web-based open review platforms plays a significant role in facilitating this process [1, 2]. Although, statistical approaches have been successful in performing effective content quality assessment, they have some inherent deficiencies which call for syntactic, semantic and structural assessment techniques to be overcome. Taking these considerations for the assessment process gives us the chance to improve the capability of automatic assessment systems in some way. There exist various algorithms and tools in this regard which are used to automatically screen and analyze contents.

Within this scope, semantic screening of research contents such as papers and reports has been a key factor for content assessment purposes. Another area of text understanding that is of high interest to researchers, concerns the processes that occur during the summary phases of reading [3]. In all these cases, there have been several methods for assessing the contents of a text, such as spotting keywords [4], utilizing dictionary of affective concepts, lexicons and commonsense knowledge-base [5], exploiting multiple ontologies [6], and domain specific classification [7].

The focus of this paper is semantic screening of research contents which facilitates the content selection process for researchers. In other words, the proposed approach can be regarded as a pre-reviewing process which is useful for selecting, filtering and recommending purposes being popularly used by instructors, researchers as well as conference and journal reviewers. To increase the semantic potential of the process, we have applied the ontology of content's key segments together with WordNet. Besides, pre-requisite of implementing the whole process at the first glance, is to extract a set of key words and phrases as indicators of different categories in the content, corresponding to the defined ontology. Subsequently, there is a need for some semantic rules, which in our framework call for usage of keywords, ontology and WordNet to determine different categories of content. There is no doubt that, proposing some complementary rules based on well-reputed journal's policies of reviewing articles, can help finalize the screening process and select the appropriate papers, as well. In this way, researchers may confront a limited set of automatic pre-screened contents that save their time and efforts in investigating plenty number of contents. In the meantime, it can leverage researcher's knowledge in meeting his/her targeted information needs.

The rest of the paper is organized as follows: Section II reviews some of the existing approaches in the area of automated assessment of the content. Section III describes the proposed semantic domain-specific framework. In section IV, experimental results are analyzed and followed by the section on conclusion and future works.

II. EXISTING APPROACHES TO CONTENT ASSESSMENT

A major difficulty with respect to assessing and selecting research and educational contents is huge amount of existing documents, as well as the few number of experts, researchers and advisors. Moreover, having various levels of expertise on domain specific contents makes them unusable for a wide spectrum of researchers [8]. In this respect, making use of automatic recommender systems that are capable to handle pre-screening or filtering the contents, seems to be helpful. Taking this point into account, each research content can be assessed from semantic, syntactic and structural perspectives. Having a survey on existing approaches for assessing contents reveals that the majority of them are either based on information retrieval or on machine learning algorithms.

In information retrieval approach, the similarities between the query and groups of documents are calculated and documents are ranked based on their similarities. In these systems queries are considered as small documents of words. It should also be mentioned that in automatic assessment of contents like summaries, similarities between the answers given by experts and users are considered. Also, the complementary role of semantic component models to the full text and keyword indexing, makes these methods appropriate for content assessment purposes [9].

Some of the important information retrieval-based assessment approaches can be enumerated as N-gram co-occurrence, Latent semantic analysis (LSA), BLUE and ROUGE [14, 15, 16]. N-gram is a substring with a length of N words. This procedure determines the similarity between two contexts by using the average number of matched N-grams [17]. Latent semantic analysis is based on the close relationship between concepts and words used in the text. LSA algorithm stores the word and its frequency in a matrix which is changed with the singular value decomposition function. Here cosine correlation function is responsible for measuring similarity between two matrices [10]. Blue is capable of assessing a content like an essay or summary of relevant information by matching it against the model content stored in the system. The main idea of the comparison is to measure the closeness of the candidate to the reference content [1]. Rouge is the mostly used procedure to evaluate the automatically generated summaries. It gets the automatically generated summary and summaries of some references as its inputs, and calculates their similarities through comparing the number of the matched N-grams. Since this procedure is completely automatic and its algorithms are well-documented, it is the best choice to evaluate users' answers to the designed questions [18]. Additionally, the combined procedure of N-gram co-occurrence and LSA can also represent appropriate ensemble results [19]. Among the different approaches to calculate semantic similarity, information content (IC) of concepts from the knowledge provided by ontologies from the one side, and semantic network of resources and properties, from the other side, show promising results in text searching and screening [6, 20].

The other approaches benefited by analytical and combinational potentials, are those which are based on machine learning algorithms widely used for content assessment purposes [1, 10, 11]. These approaches mostly make use of semantic structures like frames, rules, ontology, information concept, etc., that enable a better understanding of the content's semantics [12, 13].

Analytical learning algorithms are privileged by explanation-based methods, as well as case-based and instance-based learning algorithms which have shown promising results for content assessment purposes [21, 22, 23]. There also exist some combinational learning methods which take advantage of neural networks, reinforcement learning, intuitive and bio-inspired learning algorithms (ex. genetic algorithms [24], ant colony [25], artificial immune system [26, 27], etc.). These algorithms can be widely used for e-learning and e-research purposes in activities like text classification [28, 29], document clustering [30], content assessment [6], text reconstruction [31] and so on.

Exploiting the semantic capabilities of afore-mentioned approaches, leads us to apply semantic and syntactic assessment methods in order to screen and review the research contents like scientific papers and reports.

III. Proposed Semantic Domain-Specific Framework

A. Overall Structure

For many years, research papers and reports have been the major tools for scientific progress [32]. However, by the increasing amount of these contents, there is a need to have fast and accurate actions of screening, assessing, reviewing and recommending as well. To overcome such a problem, we propose a semantic framework to assist researchers in screening contents in the domain of agent science and technology. The infrastructure of the proposed framework is based on semantic rules, WordNet and ontologies developed for content's key segments in general, and "agent science and technology" domain in particular. In this way, both semantic and syntactic analyses of contents assist researchers to select contents appropriately. Figure1 illustrates such a framework.

B. Domain- Specific Ontology

Domain ontology is defined in terms of formal and explicit specifications of a shared conceptualization in a specific domain [33]. The construction of domain ontologies relies on domain modelers and knowledge engineers, which are typically overwhelmed by the potential size, complexity and dynamicity of a specific domain.

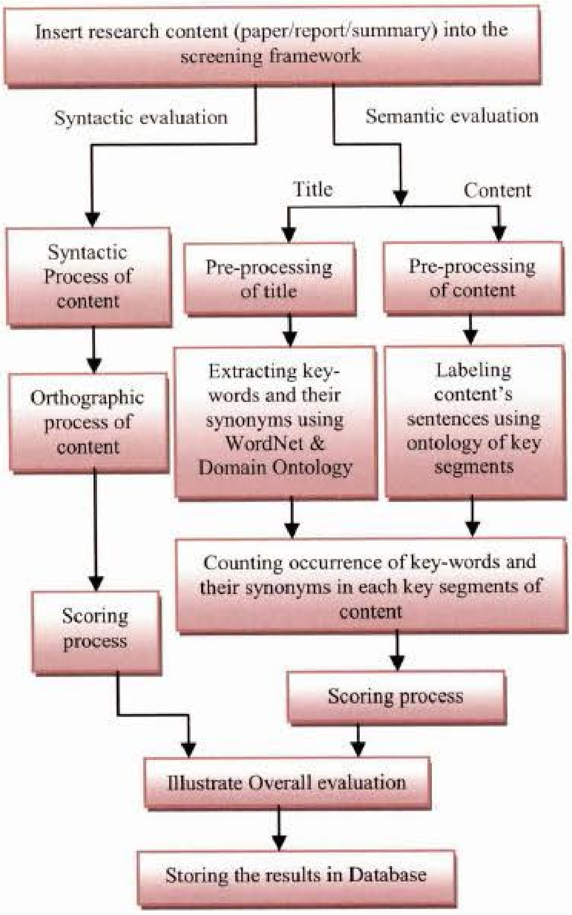
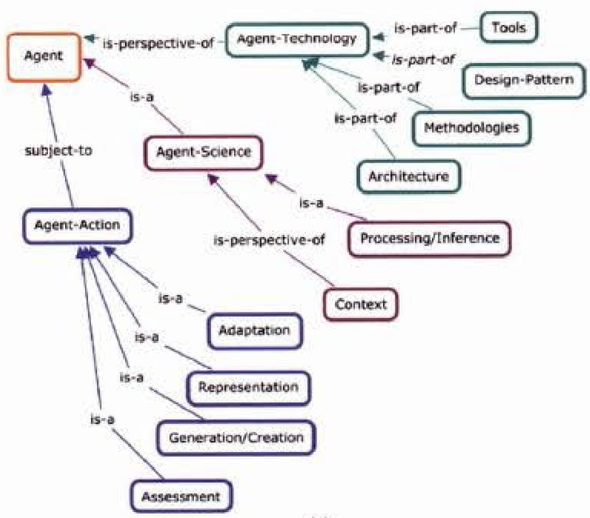


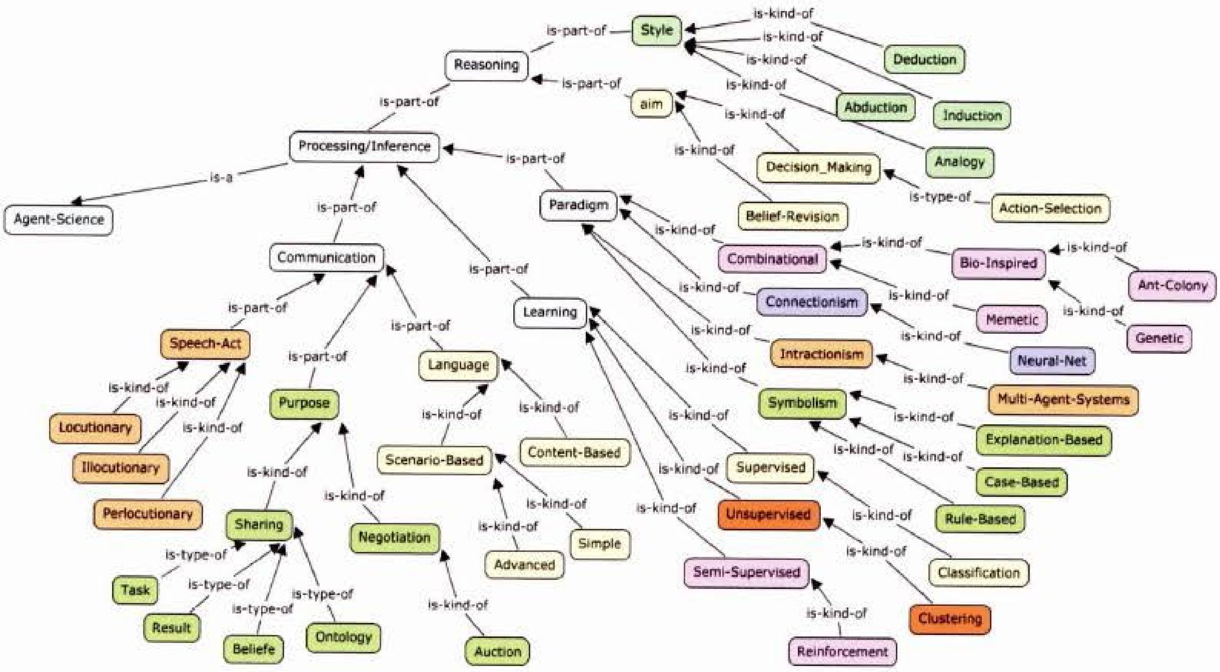
Fig. 1 Structure of the proposed framework

Information systems increasingly depend on ontology to structure data in a machine readable format and ensure satisfactory performance. Within this scope some generic ontologies like WordNet, are available, whereas most applications need specific domain ontology to describe concepts and relations in the related domain [34].

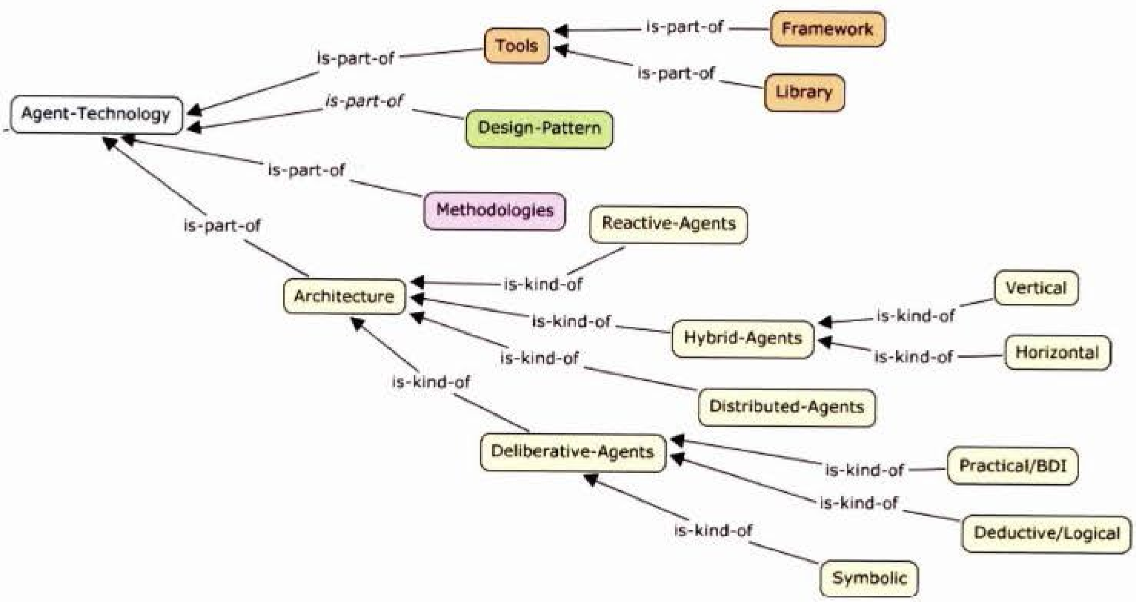
For the current research work, as we had the permission of accessing to reviewers' comments and the papers in the domain of agent science & technology (as a part of the conference on Robotics and Artificial Intelligence), there was a need to have a domain ontology in the same field. It should be noted that the domain ontology used in our framework is a result of continuous investigation of the existing research materials (books and articles) with regard to the corresponding domain, let say, "agent science and technology" in the present research. This ontology is more comprehensive compared to the existing ontologies, already suggested for the corresponding domain [35, 36, 37, 38]. Additionally, the domain ontology used in this research has already been applied to a wide range of issues such as: "content indexing/ annotation", "search/ retrieval" and "content personalization", and the results have been sufficiently satisfactory [39, 40, 41]. It is therefore expected that such an ontology would be mature enough to respond appropriately to the present application as well. Figures 2 (a), (b) and (c) illustrate some major parts of this ontology. With regard to the proposed framework, domain ontology is applied whenever synonyms of key-words do not exist in the WordNet. In such a situation domain ontology is used to extract the corresponding words of the major existing key-words in titles and body of content, in order to determine the semantic evaluation score for a content. As the reviewing rules are completely independent of the domain ontology, there is no doubt that based upon constituting ontologies for other domains, the proposed approach can be equally used to find synonyms of the required technical keywords which do not exist in the WordNet.



(a)



(b)



(c)

Fig.2 Some parts of the ontology of agent science & technology

C. Content's Key Segments Ontology

Automatic screening of contents cannot be realized unless the key segments of content have been determined properly. For this purpose, we have to look for the concepts that can be generally meaningful for a wide range of possible contents with no particular emphasis on the application domain in which the content is created for [42, 43, 44]. We therefore should emphasize on aspects such as: (i) whether the content has been benefited by a glance of the part to show the genealogy of the problem, (ii) the focal points of the existing approaches with regard to an existing problem (including strengths & weak-points), (iii) the key point of the ongoing approach, (iv) the way the ongoing approach can show its strengths with regard to the existing problem, (v) the format according to which the ongoing approach can be assessed, and finally (vi) the new horizons for the ongoing approach with regard to possible application domains that may be somewhat important in the future trend of the existing problem [42].

Taking the above points into account, the ontology of content's key segments is illustrated in Figure 3. As it is seen from the figure, labels such as "general background", "existing viewpoints", "key issues", "realization/ implementation", "comparative analysis & capability interpretation", and "conclusion & prospect anticipation" seem to cover the characteristics discussed above, and are in reality being used by a wide range of researchers to disseminate and screen research works in terms of appropriate research contents. Let say, in this way the fitness of the content is assured in a systematic manner. Moreover, by checking the existence of items propounded in content, the screening process will be facilitated [42].

IV. CONTENT PROCESSING AND EXPERIMENTAL RESULTS

As it was discussed earlier, semantic evaluation of content is performed through two major processes; processing the title of the content and processing the whole content. In technical contents especially papers of the conferences or journals which usually discuss applied technical issues of the domain and not the basic and philosophical metaphors, main key-words in the title and respectively subtitles have the potential to show the pattern of the major discussion in the paper. It has been found from screening a wide range of papers that, those keywords can occur in different parts of the content. In order to save computational cost, looking for the title's keywords together with their synonyms from WordNet and domain ontology is suggested first, and next it is suggested to process the whole contents so as to determine the degree of content's relevance to different content's key segments (illustrated in Figure 3).

A. Processing the Title

To process the title of the content, it has to be tokenized first. Afterwards, as the stop words do not reflect significant meaning to the objective of the title and have no use in semantic assessment, they have to be omitted.

The process is followed by determining major key-words in the title and subtitlesm Our emphasis in this paper is the types of materials which are mostly concerned with applied aspects, wherein the significant keywords are expected to show up in the titles (or/ and subtitles) in some way. By in "someway", we mean that a term or a phrase which is semantically similar to a keyword, would appears in the title in a way a WordNet or domain ontology can later be used to find the corresponding synonyms to be subsequently searched for in the content. So they can be taken as appropriate signs for fitness evaluation of different categories within the content. One should notice that, stemming key-words of the title by the use of WordNet is an unavoidable phase. For instance, for the title "A model of normative power", stop words such as: "a" and "of" are omitted, and then through connecting to the WordNet, stems of significant key-words are retrieved as: "model", "norm" and "power". The rest of the process belongs to finding stem words or their synonyms (which have been retrieved from WordNet or domain ontology) in different segments of content.

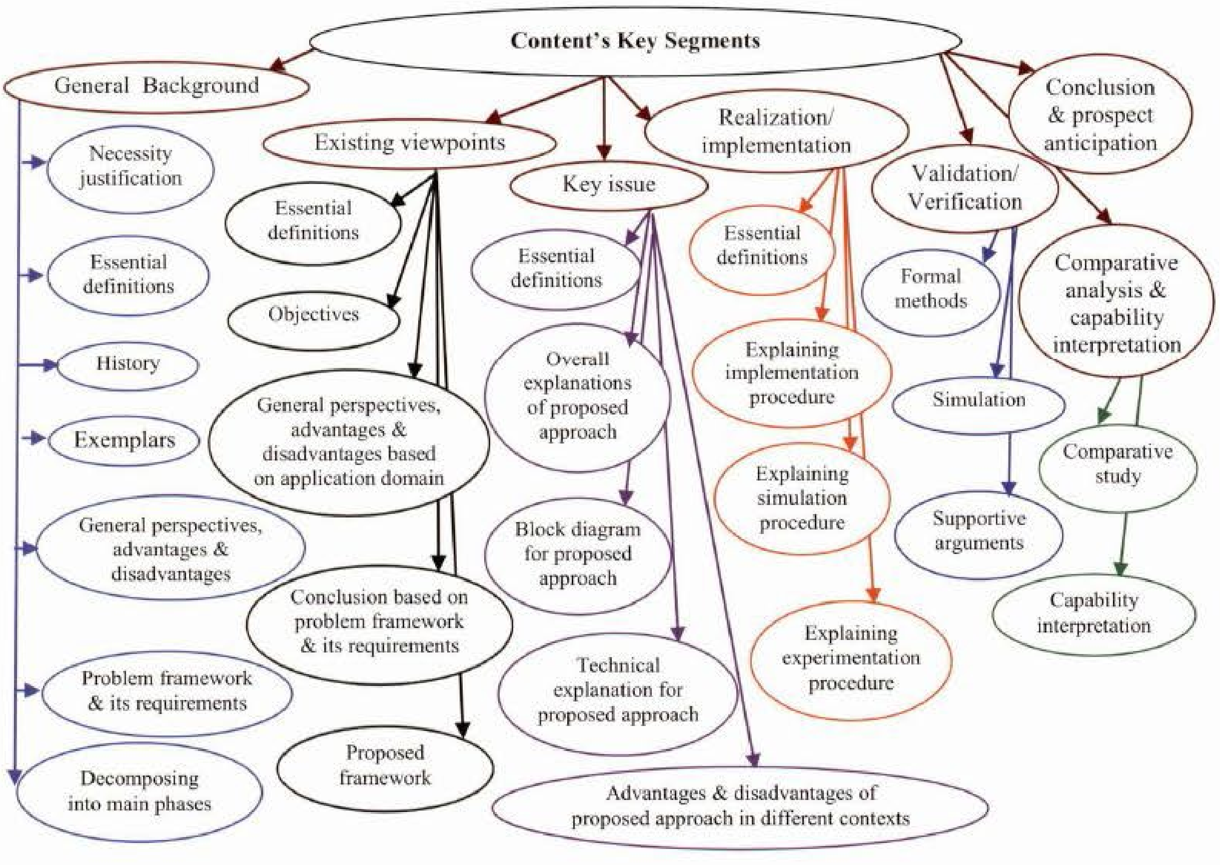


Fig.3 Ontology of Content's Key Segments

B. Processing the Content

In order to process the whole content of a research document and distinguish its key segments, the related significant words in each category should be recognized. To realize this, it is necessary to parse the sentences within the content and break each sentence into its significant key-words using WordNet and domain ontology. Having a review of several research contents like papers or reports, also the major flashers which are discussed in technical writing books [44,45], enables us to detect relevant words and phrases for each category. Facing the mentioned words or phrases while processing the content, indicates the existence of each significant category in the content. Table I, illustrates some samples of the related words and phrases to each category. The automatic process of retrieving these words and phrases from highly cited papers through information retrieval methods is a part of our future work.

For experimentation, we have focused on summaries of reports and abstracts of papers to show the capability of the suggested semantic framework. In this regard, five major categories (BPMRC0 have been defined for an abstract or a summary as follows: "Background information", "Purpose of study", "Methodology", "The most important result", and finally, "Conclusion and future prospects". Figure 4 shows the ontology of summary or abstract with respect to these categories, which have been used in our experimentations [44].

TABLE I. SOME PHRASES AND WORDS USED FOR SUMMARY CATEGORIZATION

Categories within the Summary	Some Related Words & Phrases to Each Category				
Background information	Literature review	Survey of existing methods	Related Existing methods	Historical background of	Surveys
Purpose of study	We propose	Our approach provides	This paper focuses	This paper Present	We are concerned
Methodology	We compute	Our account	We introduce	We bring	We approach
Important result	We examine	Results Demonstrate	As a result	Experimental results show	Results reveal
Conclusion	Finally discuss	The results suggest	Experiments suggest	We thus argue	We prove

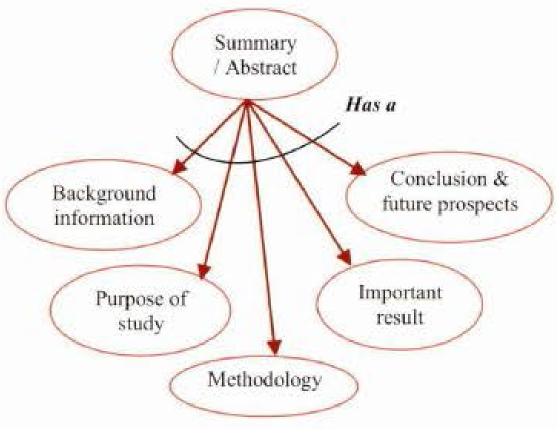


Fig. 4 Ontology of a summary

In order to process the summary and distinguish segments of BPMRC, the related significant words belonging to each category of BPMRC should be recognized. Thus, it is necessary to parse the summary and find the significant key-words corresponding to each segment. Having a review on several summaries yields some relevant words and phrases for each category of BPMRC, which are shown in Table1 and are extractable through the parsing process.

C. Semantic & Overall Screening

After processing summary and title, for each of the five mentioned categories, number of words (Wn) and occurance rates of these words resulted from the earlier title processing stage (in each segment of content (fn)), together with the ratio of the fn / Wn (that can be taken as a representative ratio to distinguish the potential of each category) will be calculated. Afterwards, based upon the threshold values obtained through experience, each category will be assigned label of "good", "moderate" or "weak". It is to be mentioned that, in case that one cannot detect a category in a summary, this summary will be tagged by "weak" label. Figure5 illustrates the evaluation process.

In order to evaluate the summaries, we have applied some rules that were determined based on threshold values and the value of fn / Wn ratio. Here is an example of the rules used for this purpose:

- Was not found → Weak
- If (f_n/ W_n < 0.07) → Moderate
- If (f_n/ W_n > 0.07) → Good

Once "evaluation" and "prescreening" stages were accomplished, the proposed framework automatically decides to tag each content as "accepted in this level", "rejected in this level", "needs major revision" or "needs minor revision". Thus, two phases of screening for each category are needed to make a final decision. Here are some of the rules we have applied to finalize the decision on "Purpose of study" part of the content:

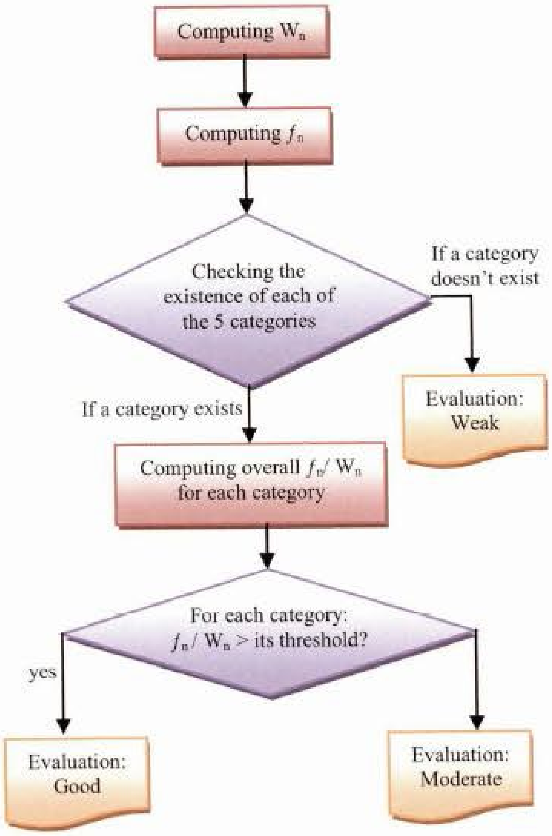


Fig.5 Flowchart of the evaluation & screening process

- Purpose of Study: Weak → rejected at this level
- Purpose of study: Moderate → needs major revision (MaR)
- Purpose of Study: Good → accepted in this level

Finally, the evaluation results for each segment of BPMRC are calculated, and based upon the pre-determined rules, final decisions will be made. As the policy of decision making on reviewers' comments for each journal and conference differs and are hidden from others, it is not possible to find these rules automatically through semi-supervised manner out of a corpus of published manuscript. That is why a paper rejected by a conference or journal, may be accepted by another conference or journal. In this respect, these rules have been determined based on the real screening and reviewing procedures with regard to an existing editorial board of some journals such as International Journal of ICT.

- One or more rejects → not recommended
- One "Major Revision" → moderately recommended
- Two or more "Major Revision" → not recommended
- If all categories are accepted → strongly recommended

D. Experimental Results

In order to assess the proposed framework, it was necessary to have a comparison between the real screening/reviewing results and automatically calculated results on the same dataset. In this respect, 54 research papers in the domain of "agent science & technology" were selected from the proceedings of Iranian Computer Society Conference (2010). Since we needed to have the comparison between the existing review results by real experts and the automatic results achieved by the proposed framework, we were obliged to contend to a limited data set with limited number of real reviewing results. Figure 6 illustrates the pseudo code of our suggested framework.

As it is shown in the pseudo code, syntactic evaluation of summary is followed by tagging several parts of summary into five categories of Background information, Purpose, Methodology, Results and Conclusion based on extracting related keywords via WordNet. The ratio of significant words to total number of words in the related part (fn /Wn), determines the potential of each part of the summary. Finally the overall screening of summary is realized based on the pre-determined rules for evaluation. Figure7 illustrates the tagged parts of a summary based on suggested categories. It is to be mentioned that, each summary of selected dataset has been evaluated by experts based on the above mentioned categories, and in the meantime by the proposed framework.

Table 3 illustrates a comparison between the results obtained from manual and automatic screening process. The results presented in Table 3 reveal that, among 54 summaries, 7 summaries were definitely recommended, 10 summaries were moderately recommended and 6 summaries were not recommended both by experts and by our proposed framework. Our suggested framework has wrongly recommended 2 summaries, while the experts have rejected them. In the meantime, our framework has not recommended 3 summaries, while the experts have recommended them.

```
Insert into screening framework (research content);
Syntactic evaluation;
Extract key-words of title;
Finding synonyms using WordNet & domain ontology
(key-words);
BPMRC Categorization (summary);
Computing (W_n);
Computing (f_n);
If (a category wasn't detected)
   Evaluation of the category = weak;
Else
   Compute f_n /W_n ;
    For each category:
        If (f_n /W_n > corresponding threshold)
            Evaluation of the category = good;
        Else
            Evaluation of the category = moderate;
Overall recommendation based on defined rules;
Storing results in database;
```

Fig.6 Pseudo code for proposed screening framework

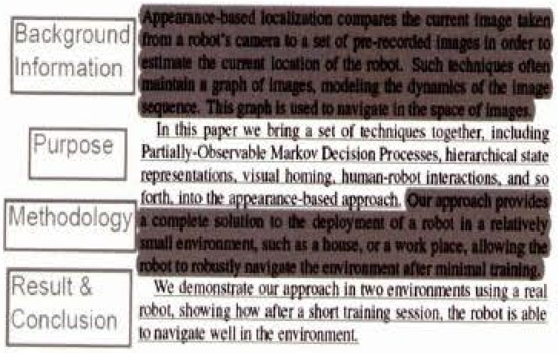


Fig.7 A sample summary and its categorization

Based on the results shown in Table 3, we can infer that in 23 cases, the results of our framework exactly match the expert's analysis of the summaries. However, in 5 cases, the results obtained by our framework are in contrast with experts' views. From the 26 results left, 13 research contents seem to have been distinguished correctly.

There could be several reasons for these mismatches. The main reason is the deficiencies in a set of relevant phrases and keywords in a summary. This is especially evident in the background information category which was not detectable in some summaries, and therefore made the proposed framework incorrectly evaluate the background information category as weak, and resultantly those summaries as not recommendable. Another reason for these mismatches is the imprecise definition of thresholds in the proposed assessment rules. Hence, refining these works can be of great importance as a future research work.

TABLE II. RESULTS OF AUTOMATIC ASSESSMENT OF SUMMARY IN FIGURE7

Categories of summary	Results
Background information	Moderate
Purpose	Moderate
Methodology	Good
Important result	Good
Conclusion	Good
Grammar & Spelling	Good
Overall	Recommended

TABLE III. COMPARISON OF AUTOMATIC RESULTS WITH EXPERT'S RESULTS

Expert's Results / Framework's Results	Recommended	Moderately Recommended	Not Recommended
Recommended	7	5	2
Moderately Recommended	8	10	8
Not Recommended	3	5	6

Closer looks into the details of the results reveal that the evaluation of our framework mostly differs from those of the experts with regard to the background information and conclusion categories. Therefore, as a future work we should concentrate more on these categories. Also, finding better phrases and more relevant patterns as well as applying better rules for the evaluation of these categories can definitely enhance our assessment performance. We also would like to extend the proposed framework to the whole parts of the research content. In this regard, we are negotiating with some conferences in agent science domain to access to their reviewing results so that we may assess our screening framework in a more realistic way.

V. CONCLUDING REMARKS

In this paper, we proposed a semantic domain-specific framework which is capable of assisting researchers by pre-screening the contents automatically. In this manner, researchers can have the privilege of pre-filtering contents automatically to save their cost and time. To perform this task, the ontology of content's key segments together with WordNet and a domain ontology (here ontology of agent science & technology) were applied. As the present suggested framework has just been applied to a limited number of contents, no serious problem would exist with regard to time consumption.

Experimental results on 54 research contents reveal that the proposed framework is successful in almost 70% of cases. There are some reasons for such a deficiency like: considering inappropriate phrases and keywords, as well as determining imprecise thresholds. As a future work we should concentrate on finding better phrases and more relevant patterns, and also applying better rules for evaluation of the entire content as well as the related summaries. For the moment, extracting reviewing rules has been performed through a process of acquisition from experts. However, such an extraction based on a process of statistical data mining is suggested as future research work. The automatic process of retrieving related words and phrases from highly cited papers through information retrieval methods is also a part of our future work.

ACKNOWLEDGMENT

Research Institute for Research Institute for ICT (ITRC) is appreciated for its financial support of the current research.

REFERENCES

[1] F. Noorbehbahani and A. A Kardan, "The Automatic Assessment of Free Text Answers Using a modified BLEU Algorithm," Computers & Education, July 2010.

[2] Y. A. Le Borgne and A. Campo, "Open review in computer science Elsevier grand challenge on executable papers", Procedia Computer Science, Vol. 4, 2011, pp. 778-780.

[3] R. Olmos, J. A. León, G. Jorge-Botana, and I. Escudero, "New algorithms assessing short summaries in expository texts using latent semantic analysis," Behavior Research Methods, Vol. 41, No. 3, 2009, pp. 944-950.

[4] A. Valitutti, C. Strapparava, and O. Stock, "Developing Affective Lexical Resources," PsychNology Journal, Vol. 2, No. 1, 2004, pp. 61–83.

[5] H. Liu, H. Lieberman, and T. Selker, "A Model of Textual Affect Sensing using Real-World Knowledge," Proc. IUI, Miami, FL, ACM Press, New York, 2003, pp. 125–132.

[6] D.Sanchez and M. Batet, "A Semantic Similarity Method Based on Information Content Exploiting Multiple Ontologies", Expert Systems with Applications, Vol.40, Issue 4, March 2013, pp. 1393-1399.

[7] T. Nasukawa and J. Yi, "Sentiment Analysis: Capturing Favorability Using Natural Language Processing," Proc. K-CAP, ACM Press, New York, 2003, pp. 70–77.

[8] M. S. Paukkeri, M. Ollikainen, T. Honkela, "Assessing User-Specific Difficulty of Documents", Information Processing & Management, Vol. 49, Issue 1, Jan. 2013, pp. 198-212.

[9] S.L.Price, M.L.Nielsen, L.M. Delcamber, J. Steinhauer, "Using Semantic Components to Search for Domain-Specific Documents: An Evaluation from the System Perspective and the User Perspective", Information Systems, Vol. 34, Issue 8, Dec. 2009, pp.724-752.

[10] Y. He, S. Hui, and T. Quan, "Automatic summary assessment for intelligent tutoring systems," Computers & Education, Vol. 53, No. 3, Nov 2009, pp. 890-899.

[11] F. Figueiredo, L. Rocha, T. Couto, T. Salles, M. Gonçalves, and Jr. W. Meira, "Word co-occurrence features for text classification," Information Systems, Vol.36, Jul 2011, pp.843-858.

[12] D.Sanchez, M.Batet, D.Isern, "Ontology-Based Information Content Computation", Knowledge-Based Systems, Vol. 24, Issue 2, March 2011, pp. 297-303.

[13] W. IJntema, J. Sangers, F. Hogenboom, F. Frasincar, "A lexico-semantic pattern language for learning ontology instances from text, Web Semantics: Science, Services and Agents on the World Wide Web", Vol. 15, Sept. 2012, pp. 37–50.

[14] D. Santos, M. B. Ribeiro, and R. Bastos, "Developing a Conference Management System with the Multi-Agent Systems Unified Process: A Case Study," AOSE 2007, LNCS 4951, 2008, pp. 212–224.

[15] V. Soo, S. Lin, S.Yang, S.Lin, and S.Cheng, "A cooperative multi-agent platform for invention based on patent document analysis and ontology," Expert Systems with Applications, Vol. 31, No. 5, 2006, pp.766–77.

[16] M. M. Hasan," Can Information Retrieval Techniques Meet Automatic Assessment Challenges?," 12th Intl. Conf. on Computer and Information Technology, December 2009.

[17] F. Figueiredo, L. Rocha, T. Couto, T. Salles, M. Gonçalves, and Jr. W. Meira, "Word co-occurrence features for text classification," Information Systems, Vol.36, Jul 2011, pp.843-858.

[18] C. Lin, "ROUGE: a Package for Automatic Evaluation of Summaries," Workshop on Text Summarization (WAS 2004), Barcelona, Spain, 2004.

[19] Y. He, S. Hui, and T. Quan, "An Ensemble Approach for Semantic Assessment of Summary Writings," ICCET '09: Proceedings of the 2009 International Conference on Computer Engineering and Technology, Vol. 01, Jan 2009.

[20] M. Lee, W. Kim, S. Park, "Searching and Ranking Method of Relevant Resources by User Intention on the Semantic Web", Expert Systems with Applications, Vol.39, Issue 4, March 2012, pp.4111-4121.

[21] M. Healy, S. J. Delany, A. Zamolotskikh, "An Assessment of Case Base Reasoning for Short Text Message Classification", 15th Irish Conf. on Artificial Intelligence and Cognitive Sciences (AICS'04), 2004, pp.9-18.

[22] R. Weber, D. W. Aha, N. Sandhu, H. Munoz-Avila, "A Textual Case-Based Reasoning Framework for Knowledge Management Applications", 9th German Workshop on Case-Based Reasoning Shaker Verlag, 2001, pp. 244-253.

[23] D. Bridge,P. Healy, "The GhostWriter-2.0 Case-Based Reasoning System for Making Content Suggestions to the Authors of Product Reviews", Knowledge-Based Systems, Vol. 29, May 2012,pp. 93–103.

[24] W. Song and S. Park, "Genetic algorithm for text clustering based on latent semantic indexing", Computers & Mathematics with Applications, Vol. 57, Jun 2009, pp.1901-1907.

[25] M. Hosseinzadeh-Aghdam, N. Ghasem-Aghaee, and M. Basiri, "Text feature selection using ant colony optimization", Expert Systems with Applications, Vol. 36 (3 Part 2), April 2009, pp. 6843-6853.

[26] D. Dasgupta, S. Yu, and F. Nino, "Recent Advances in Artificial Immune Systems: Models and Applications", Applied Soft Computing, Vol. 11, No. 2, Mar 2011, pp. 1574-1587.

[27] N. Forouzideh, M. T. Mahmoudi, K. Badie, "Organizational Texts Classification Using Artificial Immune Recognition Systems", IEEE Symp. Series on Computational Intelligence (SSCI 2011), Paris, France, April 11-15, 2011.

[28] M. Ghiassi, M. Olschimke, B. Moon, P. Arnaudo, "Automated text classification using a dynamic artificial neural network model Original Research Article", Expert Systems with Applications, Vol. 39, Issue 12, Sep. 2012, pp.10967-10976.

[29] M. T. Mahmoudi, F. Taghiyareh, B. N. Araabi, "A Neuro-Fuzzy Immune Inspired Classifier for Task-Oriented Texts", Journal of Intelligent & Fuzzy Systems. (Accepted for Publication)

[30] Q. Guo and M. Zhang, "Multi-documents Automatic Abstracting based on text clustering and semantic analysis", Knowledge-Based Systems, Vol. 22, No. 6, 2009, pp. 482-485.

[31] L. Gao, Y. Wang, Z. Tang, X. Lin, "Newspaper article reconstruction using ant colony optimization and bipartite graph",Applied Soft Computing, In Press, Corrected Proof, Available online 21 July 2012.

[32] G. R. Brammer, R. W. Crosby, S. J. Matthews, and T. L. Williams, "Paper M^ach´e: Creating Dynamic Reproducible Science", Procedia Computer Science, Vol.4, 2011, pp. 658–667.

[33] D. Sanchez, A. Moreno, "Learning Non-Taxonomic Relationships from Web Documents for Domain Ontology Construction", Data & Knowledge Engineering, Vol. 64, 2008, pp. 600–623.

[34] C. S. Lee, Y. F. Kao, Y. H. Kuo, M. H. Wang, "Automated Ontology Construction for Unstructured Text Documents", Data & Knowledge Engineering, Vol. 60, 2007, pp.547–566.

[35] F. Zini, L.Sterling, "Designing Ontologies for Agents", Joint Conference on Declarative Programming (AGP'99), , 1999, pp.29-42.

[36] A. Adamo, L. Cafaro, V. Loiay, C. Romanoz and M. Venieroy, "Multi-layered Agent Ontology System for Resource Inventory", . IEEE International Symposium on Industrial Electronics (ISIE 2008), pp. 2317 – 2322.

[37] M. Obitko, V.Marik, "Ontologies for multi-agent systems in manufacturing domain", 13th International Workshop on Database and Expert Systems Applications, 2002, pp. 597 – 602.

[38] http://agentowl.sourceforge.net/api/agent/core/onto/Ontology.html

[39] M.T.Mahmoudi, F.Taghiyareh, K.Rajavi, M.S.Pirouzi, "A Context-Aware Framework for Semantic Indexing of Research Papers", The Fourth Inl. Conf. on Information, Process, and Knowledge Management (eKnow 2012), Valencia, Spain, 2012.

[40] M. T. Mahmoudi, F. Taghiyareh, K. Badie, "A Semantic Rule-based Framework for Efficient Retrieval of Educational Materials", The Electronic Journal of E-Learning, Volume 11, Issue 3, 2013, pp182-192.

[41] L. Zeynalvand, "Applicational Development of Academic Article Search System Based on Domain Ontology", B.Sc. Thesis, University of Tehran, 2013.

[42] K. Badie, M. Kharrat, M. T. Mahmoudi, M. S. Mirian, T. M. Ghazi, S. Babazadeh, "Ontology-Driven Creation of Contents: Making Efficient Interaction between Organizational Users and Their Surrounding Tasks", Chapter 10 in Book of User Interface Design for Virtual Environments: Challenges and Advances (Editor: Badrul H.

Khan), Information Science Reference, IGI Global USA, 2011.

[43] M. T. Mahmoudi, F. Taghiyareh, K. Rajavi, F. Shokri, L. Khamnian, "Semantic Advisor-Assisting Framework to Select Learning Materials", 3rd Intl. Conf. on e-learning & e-teaching, ICELET 2012, Tehran.

[44] T. N. Huckin, L. A. Olsen, "English for Science and Technology: A Handbook for Nonnative Speakers", Mcgraw-Hill College, 1983.

[45] T. N. Huckin, L.A. Olsen, "Technical Writing and Professional Communication for Non-Native Speakers", McGraw-Hill Humanities/Social Sciences/Langua; 2 Sub edition, 1991.

Maryam Tayefeh Mahmoudi is a Ph.D. candidate at department of machine intelligence in the University of Tehran majoring in artificial intelligence with emphasis on intelligent organization of educational contents. Within the past years, she has been involved in a variety of research works at IT Research Faculty in ICT Research Institute (ITRC), working on issues like automatic generation of ideas and contents, decision support systems for research & education purposes, as well as conceptualization of IT research projects. She is a co-author of many research papers in different journals and proceedings of conferences.

Fattaneh Taghiyareh is a faculty member at the Department of Software engineering & Information Technology, School of Electrical and Computer Engineering, University of Tehran since 2001. She received her Ph.D. from the Tokyo Institute of Technology in 2000. Her research interests are in Technology Enhanced Learning, Multi-Agent Systems, and Human-Centered Computing Systems. Her current research involves the intersection of Web-Services, Personalization, and adaptive LMSs as well as applying ontology to semantic web. She is the ex Head of the Department of Information Technology at the School of Electrical & Computer Engineering, and the former Director of Technology Incubator, University of Tehran. Recently she has established e-learning Laboratory with the mission of providing standard views for e-learning materials and platforms. Currently she is supervising IT Foundations Lab, e-learning Lab, and Multi Agent Systems Lab as well.