

# Text Rank for Domain Specific Using Field Association Words

Omnia G. El Barbary<sup>1</sup>, El Sayed Atlam<sup>1,2</sup>

<sup>1</sup>Mathematics Department, Faculty of Science, Tanta University, Tanta, Egypt

<sup>2</sup>College of Computer Science and Engineering, Taibah University, Yanbu, KSA

Email: satlam@taibahu.edu.sa

**How to cite this paper:** El Barbary, O.G. and Atlam, El-Sayed (2020) Text Rank for Domain Specific Using Field Association Words. *Journal of Computer and Communications*, 8, 69-79.

<https://doi.org/10.4236/jcc.2020.811005>

**Received:** October 3, 2020

**Accepted:** November 17, 2020

**Published:** November 20, 2020

Copyright © 2020 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

Text Rank is a popular tool for obtaining words or phrases that are important for many Natural Language Processing (NLP) tasks. This paper presents a practical approach for Text Rank domain specific using Field Association (FA) words. We present the keyphrase separation technique not for a single document, although for a particular domain. The former builds a specific domain field. The second collects a list of ideal FA terms and compounds FA terms from the specific domain that are considered to be contender keyword phrases. Therefore, we combine two-word node weights and field tree relationships into a new approach to generate keyphrases from a particular domain. Studies using the changed approach to extract key phrases demonstrate that the latest techniques including FA terms are stronger than the others that use normal words and its precise words reach 90%.

## Keywords

Text Rank, Keyphrase Extraction, Field Association Words, Information Retrieval

---

## 1. Introduction

The knowledge available through a web is infinite most days. It frequently includes data of great quality in the form of online pages. But identifying relevant information automatically and choosing the highest set of data for a specific information need isn't an easy task. Text Rank is a natural language rating algorithm based on the general concept of a graph [1] [2] [3]. Essentially, Page Rank runs on a diagram precisely designed for an exactly NLP task. For keyphrase extraction [4] [5], it builds a graph using some set of text units as vertices. Keyphrase collection is essential for too many problems relating to NLP, along with

text summaries, text categorization, retrieving data, etc. The maximum is the collection of significant and topical sentences from either a text or subject corpus. However, the major issue is far from being solved: state-of-the-art accomplishment in the extraction of keyphrases was much significantly smaller than many core NLP tasks [5]. The similitude of extracting keyphrases from either a document or a field is that they're using similar algorithms. Given the fact that Term Frequency and Inverse Document Frequency (TFIDF) is used to measure the domain weighted in a strengthened Text Rank system for Keyphrase extracting in prior gallate [6] [7], it is badly performed when retrieving domain-specific keyphrases. The difficulty comes throughout the particular field. Keyphrase excavation requires so many regions related to key knowledge whereas the extraction of data file keyphrases hardly concerns the topic of a single document. The primary objective of the whole article is to study how to use field association words to strengthen domain-specific keyphrase extraction based on Text Rank using field association words. The extracting of Domain-specific Keyphrase requires 3 stages. First, we defined a framework of the domain corpus. Secondly, extricate a list of different domain words or phrases using field association words algorithm, and obtain great and semi-perfect FA words. Keyphrases can annotate a domain's main meaning and are normally nouns, adjectives and verbs, but shouldn't be worthless words such as stop words. Thereby also, in order to obtain a decent major party key phrase, we should use the stop-word list to erase stop words [8] [9]. Third, then choose key phrases from the selection key sentence menu utilising supervised or unsupervised methods.

The remainder of the paper is organized as follows: Section 2 illustrates FA words and their methodology for extraction. Text Rank for extracting of FA words is defined in Section 3. Section 4 shows corpus construct and experimental results.

## 2. Field Association Words

All traditional methods of text classification and document similarity are based on word information in the whole documents. The key idea in our new study is to extract a new term called (FA) words that can recognize fields by using specific words without reading the whole document. For example, word "election" can indicate the document filed "Politics".

Document fields can be decided efficiently if there are many FA words and if the frequency rate is high. Therefore, five levels of FA words can be described. Traditional method was building FA words by adding new FA word candidates to FA word dictionary manually, but there are many FA words not appended to the dictionary, and much time needed to revise that dictionary. A new method for selecting English field association terms of compound words and a method to append these FA words to that dictionary automatically in [10]-[17]. Using these specific words and new FA words dictionary, our target is to make a new research in all old information retrieval areas (Ex. Document classification and

summarization, etc.) by using these specific words, which will be more effective than using whole documents as traditional methods. Research has shown that (FA) Words are valuable in document classification [12] [18], similar file retrieval [19] [20] and passage retrieval [21], and holds a lot of potential for applications in natural language processing and information retrieval. Therefore, this chapter presents a method to extract candidates for FA words from large domain specific corpora.

## 2.1. FA Words

**Definition 1:** A standard FA word indicates a minimal unit (word) with intended meaning defining a given area.

**Definition 2:** A composite term FA is composed of multiple single words of FA. For example, in machine learning, the word compound FA is the information system.

**Definition 3:** A field tree is a graphical description of associations between domains of the documents. Leaf nodes suit terminals in field tree as baseline of information. The knowledge base will constitute a dictionary of FA Terms.

A field tree which contains 11 superfields, 70 median fields and 321 terminal fields (subfields) is used in our analysis. In **Figure 1**, for example, the path defines superfield “electronics”.

Every FA Word is connected within a dynamic field tree to a particular specialty the one shown in **Figure 1**. Since an FA Term may relate to even more only one field, it is likely at much more than one node that the very same FA Term may be related to the field tree. In the FA Words database, its degree reflects that an FA Word belongs to more than one area or not.

*01. electronics*

*01.0 MICROELECTRONICS*

*01.1 INFORMATION PROCESSING*

*01.2 IT AND TELEMATICS*

*01.3 TELECOMMUNICATIONS*

*01.3.0 Radar*

*01.3.1 Research Networking, GRID*

*01.3.2 Mobile Communications*

*01.3.3 Audiovisual Equipment*

.....

*01.4 MULTIMEDIA*

*01.4.0 Cultural Heritage*

*01.4.1 E-Learning*

*01.4.2 Visualisation,*

*01.4.3 Human Language Technologies*

-----

*01.5 NETWORKING*

**Figure 1.** A part of field tree.

## 2.2. Levels of FA Words

Many FA Words may know a given field individually, although some FA Words might refer to several or even more domains. Therefore each FA concept has a broader focus for associating with a sector. So take that into consideration, FA Terms are graded into five distinct levels [11] depending on how well they represent specific fields as illustrated in **Table 1**. There are five levels of FA terms with strength degree. The first level with the most strong terms called Perfect Field Association words (PFA) Associate the terminal area to one. The second level less than the first in strength of words called Semiperfect Field Association words (SPFA) Associate several terminal fields into one medium field. The third level less than the second in strength called Medium field association words Just combine with one medium field. The fourth level called Multiple-Field Association words (MuFA) Associate several terminal fields and a medium field. The last field is called Non-Specific Field Association words (NSFA) which do not identify Medium or Terminal fields. NSFA includes the words stop (e.g. articles, prepositions, pronouns).

**Table 1** shows some examples of FA words and their ranks. The word “Microelectronics” in the field association path “<technology\electronics>” considered PFA words. The word “Biocontrol” in the field association path <technology\biological science> and <technology\agriculture> considered SPFA.

## 2.3. Comparison with Traditional Words

We mean either index term or terminology by traditional Words. An index phrase is a term which takes the meaning of a document’s subject matter and is usually used in document retrieval [22]. Index word constitutes a standardized vocabulary for vocabulary use and is used as keywords for retrieving documents and text in an information system such as a catalog or search engine.

Via contextual expression compares the efficiency, for the purpose of recording and promoting correct usage [23]. It is commonly used in translation and in the representation of knowledge in a given domain. However, the words “terminology” and “index terminology” are often used interchangeably as in [22] [24].

**Table 1.** An example of the levels of FA words.

FA word	Field association path	Levels
Microelectronics	<technology\electronics>	1
Biocontrol	<technology\biological science> <technology\agriculture>	2
Information	<technology>	3
Seed coating	<technology\agriculture> <industry\agrofood industry>	4
The, he she, it, in, on,...		5

FA words provide a more structured analysis of words which can distinguish between different areas. In a category strategy called field tree, FA mainly related to the fields and the object of FA words are to develop a comprehensive depth of knowledge. At the level of individual terms, both indexes word and terminology are similar to FA Terms, but in several respects they are also distinct. FA words are equivalent to index phrases in that they both consist of selecting a string of letters that are statistically and contextually significant, together with the names of persons and places, etc. used in a particular article. But index words are less clearly defined than FA words, and the choosing of index phrases can very much depend on the user and the reason for which they are chosen. Contrary to FA words, sometimes indebted words may also compose of word stems.

Nominal terms and concept study may rule out the naming of persons and places that qualify as FA words. Likewise, certain topic terminologies that do not appear in the records will not count as FA terms. Also, for each domain, traditional words are typically handled in isolation [19] unlike the FA words which are handled as a knowledge base. FA words obtain from a text by taking into consideration their occurrence in both the document and the particular corpus of the whole domain, whereas index Words can be extracted immediately from different files. Thus FA terms have a higher field specificity while index terms cannot be said the same. Furthermore, as stated in Section 2.1, by definition, the terms FA are limited to the minimum word or phrase that can identify a document field.

### 3. Improving Text Rank Using FA Word Extraction

Therefore, the words to rank are sequences of one or more lexical term extracted from text, and these describe the vertices assigned to the text graph [4] [21]. Any relationship that can be established between two linguistic items is a useful link (edge) that can connect two vertices of this nature. We use a co-occurrence relationship, streaked by the distance between occurrences of words:  $joi$  is two vertices joint if they're identical lexical units co-occur within a field of maximum words. Co-occurrence links obvious relations between syntactic elements.

#### 3.1. FA Word Weights

If FA words are often used in papers to index each document collection, database  $D_i$  may be represented as a vector of terms in which the document-term weight is represented [12]. A weight of 0 is assumed for terms not assigned to a given document.

Two key words-weight elements of FA exist:

- 1) Rate of presence of the FA term  $FAW_k$  or  $FAtf_{ik}$  or text in question  $D_i$ .
- 2) Inversion level of FA word, text  $FAT_k$  or  $FAidf_k$ . And its computed as

$$FAW_{ik} = \frac{FAtf_{ik}}{\sqrt{\sum_k (FAtf_{ik})^2}} = \frac{FAtf_{ik} \cdot FAidf_k}{\sqrt{\sum_k (FAtf_{ik} \cdot FAidf_k)^2}}$$

The formula of Text Rank is proposed in [12] shown in the formula (1).

$$WS(V_i) = (1-d) + d * \sum_{V_j \in In(V_i)} \frac{FAw_{jk}}{\sum_{v_k \in Out(V_j)} FAw_{jk}} WS(V_j) \tag{1}$$

The  $d$  value is normally set at 0.85.  $FAw_{jk}$  is the weight of the edge from the prior node  $V_j$  to the current node  $V_k$ .  $In(V_j)$  is the set of nodes that point to it (predecessors).  $Out(V_k)$  is the set of nodes that node  $V_j$  points to (successors).

$\sum_{v_k \in Out(V_j)} FAw_{jk}$  is the summation of all edge weights in the previous node  $V_k$ .

### 3.2. Text Rank Domain Specific Algorithm

**Algorithm 1** describes our method for ranking text for specific fields. the inputs for this algorithm is the selected set of FA words and threshold  $\infty$ . in this algorithm we calculate the concentration ratio as follows, For the parent  $\langle S \rangle$ , the child field  $\langle C \rangle$ , the concentration ratio(Concentration ( $w, \langle C \rangle$ )) of the *FA word*  $w$  in the field  $\langle C \rangle$  is defined as in line 4. For the root =  $\langle S \rangle$ , the child field =  $\langle S/C \rangle$  of the field tree, the Formula in line 5 is used to judge whether or not the word  $w$  is a *Perfect-FA word*. After that, calculate the weight of these words in step 6. And repeat these steps for semi-perfect FA words at steps 7, 8 and 9.

**Algorithm 1:**

**Input:** 1-  $w$ , is a set of candidates FA

2- Norm( $w, \langle S \rangle$ ) for  $w$  and  $\langle S \rangle$

3-  $\infty$  a threshold to judge FA words ranks

**Output:** weighted PFA, SPFA words

**Method:**

1) Set PFA={}, SPFA={}

2) set root= $\langle S \rangle$ , set child $\langle s/c \rangle$ .

3) for the root $\langle S \rangle$  and any child $\langle S/C \rangle$

4) calculate  $conc(w, \langle S \rangle) = ((Normalization(w, \langle C \rangle)) / (Normalization(w, \langle S \rangle)))$ ,

$Normalization(w, \langle T \rangle) = ((Frequency(w, \langle T \rangle)) / (Total-Frequency(\langle T \rangle)))$

5) if ( $conc(w, \langle S \rangle) \wedge conc(w, \langle S/c \rangle) \geq \infty$ )

Then set  $w$  in class PFA

$$6) PFAw_{ik} = \frac{PFAtf_{ik}}{\sqrt{\sum_k (PFAtf_{ik})^2}} = \frac{PFAtf_{ik} \cdot PFAidf_k}{\sqrt{\sum_k (PFAtf_{ik} \cdot PFAidf_k)^2}}$$

$$7) WS(V_i) = (1-d) + d * \sum_{V_j \in In(V_i)} \frac{PFAw_{jk}}{\sum_{v_k \in Out(V_j)} PFAw_{jk}} WS(V_j)$$

Else

8) if ( $conc(w, \langle S \rangle) \geq \infty \wedge conc(w, \langle S/c \rangle) < \infty$ )

Then set  $w$  in class SPFA

$$9) SPFAw_{ik} = \frac{SPFAtf_{ik}}{\sqrt{\sum_k (SPFAtf_{ik})^2}} = \frac{SPFAtf_{ik} \cdot SPFAidf_k}{\sqrt{\sum_k (SPFAtf_{ik} \cdot SPFAidf_k)^2}}$$

$$10) WS(V_i) = (1-d) + d * \sum_{V_j \in In(V_i)} \frac{SPFAw_{jk}}{\sum_{v_k \in Out(V_j)} SPFAw_{jk}} WS(V_j)$$

End

Apply Algorithm 1 for some documents first extracts PFA and SPFS as in **Figure 2**. **Table 2** shows the weights representation of the selected PFA and SPFA.

## 4. Experimental Results

### 4.1. Corpus

Field association terms serve as a highly simplified description for a domain, which can be used as domain labels and used in text classification, retrieval of information, and so on. The approach presented is tested by applying it to 38 MB of text databases (newspaper, technical report, etc.) comprising 11 fields, 70 median fields and 321 terminal fields.

### 4.2. Experiment Design

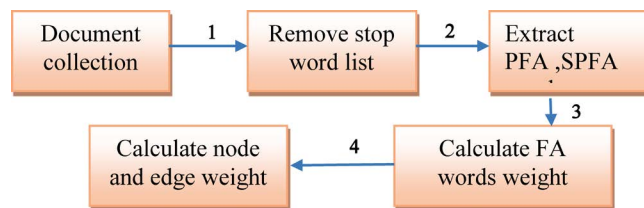
**Figure 3** shows the flow chart of extracting domain specific Text Rank.

```
{'TELECOMMS': 1,
'MICROELECTRONICS': 1,
'INFORMATION PROCESSING': 1,
'Graphics': 1,
'Middleware': 1,
'MULTIMEDIA': 1,
'NETWORKING': 1,
'MODELLING': 1}
```

**Figure 2.** The result of extracts PFA and SPFS words.

**Table 2.** Weights representation of the selected PFA and SPFA.

	TELECOMMS	MICROELECTRONICS	INFORMATION PROCESSING	Graphics	Middleware	MULTIMEDIA	NETWORKING	MODELLING
$d_1$	0.24	0.25	0	0.59	1.17	0.2	0	0
$d_2$	0	0	0	0	0	0.44	0.54	0.6
$d_3$	0.25	0.24	0.21	0.44	0	0	0	0
$D_4$	0.22	0.24	0.22	0	0	0	0.57	0



**Figure 3.** Flow chart of domain specific text rank.

First ever step is data preparation of field documents, involving the dataset of the domain file, word division, eliminating stop words and choosing candidate terms. The second step is to extract FA terms perfect and semi-perfect. Third, calculate the weight of the words FA. Next, use the ranking formulas on the prepared dataset to measure node weight or edge weight and FA word ratings. Finally, take the PFA top PFA and SPFA words as domain keyphrases.

### 4.3. Experimental Evaluation

Our experiment considers a comparison for Text Rank using normal keywords and FA words.

Precision, Recall and F-measure are used to evaluate relevancies of the given methods and defined as follows:

$$\text{Recall}(R) = \frac{\text{Correct Classified Documnts}}{\text{Total Corrected Classified}}$$

$$\text{Precision}(P) = \frac{\text{Correct Classified Documnts}}{\text{Total Retrieved Classified}}$$

$$\text{F-measure} = \frac{2 \times P \times R}{P + R}$$

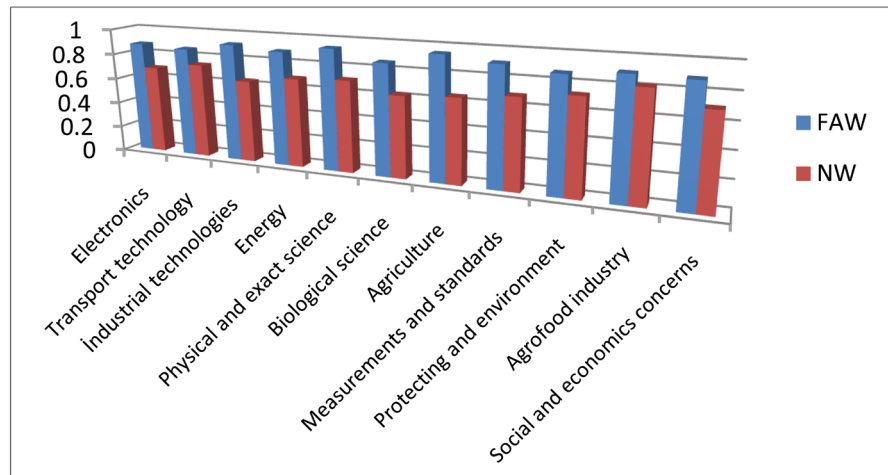
Precision, Recall and F-measure for 11 super-fields are measured using normal keywords and FA words and as shown in **Table 3** & **Figure 4**.

From the evaluation results it turns out that the best performance is recorded in Text Rank with FA-words as obvious in **Table 2**. Moreover, the calculation of F-measure for each class separately using FA-words are more accurate than normal keywords.

**Table 3.** The results using Normal (N) keywords and FA words.

Name of field	Precision		Recall		F-measure	
	NW	FAW	NW	FAW	NW	FAW
Electronics	0.59	0.78	0.79	0.98	0.69	0.88
Transport technology	0.22	0.84	0.62	0.90	0.74	0.86
Industrial technologies	0.48	0.89	0.74	0.95	0.64	0.92
Energy	0.71	0.89	0.79	0.98	0.69	0.89
Physical and exact science	0.48	0.90	0.63	0.99	0.71	0.94
Biological science	0.55	0.80	0.64	0.93	0.63	0.86
Agriculture	0.77	0.85	0.81	0.9	0.65	0.95
Measurements and standards	0.64	0.88	0.75	0.95	0.69	0.91
Protecting and environment	0.59	0.81	0.64	0.95	0.73	0.87
Agrofood industry	0.61	0.84	0.72	0.98	0.82	0.90
Social and economic concerns	0.43	0.86	0.69	0.97	0.70	0.89





**Figure 4.** F-measure comparison between text rank using FA words and normal words.

## 5. Conclusions

Extraction of domain keyphrase is important for many tasks in the processing of language by nature. This method extracted keyphrases from all domains. We analyzed the development of Text Rank using field association words and a node weight. Experiment indicates that the Text Rank accuracy will grow markedly with Algorithm 1. Experiments show that the PFA and SPFA weights are of the highest precision when we extract Top words from a domain corpus.

Experiments also show that FA Text Rank's extraction precision will hit 90 percent when our corpus extracts keyphrases.

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

- [1] Li, W.G. and Zhao, J.B. (2016) Text Rank Algorithm by Exploiting Wikipedia for Short Text Keywords Extraction. *3rd International Conference on Information Science and Control Engineering*, Beijing, 8-10 July 2016, 683-686. <https://doi.org/10.1109/ICISCE.2016.151>
- [2] Yu, S.S., Su, J.D., Li, P.F. and Wang, H. (2016) Towards High Performance Text Mining: A Text Rank-Based Method for Automatic Text Summarization. *International Journal of Grid and High Performance Computing*, **8**, 58-75. <https://doi.org/10.4018/IJGHPC.2016040104>
- [3] Mihalcea, R. and Tarau, P. (2004) Text Rank: Bringing order into Texts. *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, Barcelona, July 2004, 404-411.
- [4] Bellaachia, A. and Al-Dhelaan, M. (2012) NE-Rank: A Novel Graph-Based Keyphrase Extraction in Twitter. *2012 IEEE/WIC/ACM International Conferences on Web Intelligence & Intelligent Agent Technology*, Macau, 4-7 December 2012, 372-379.
- [5] Frank, E., Paynter, G.W., Witten, I.H., Gutwin, C. and Nevill-Manning, C.G. (1999)

- Domain-Specific Keyphrase Extraction. *Proceedings of 16th International Joint Conference on Artificial Intelligence*, Stockholm, 31 July-6 August, 668-673.
- [6] Fuketa, M., Lee, S., Tsuji, T., Okada, M. and Aoe, J. (2000) A Document Classification Method by Using Field Association Words. *Information Sciences*, **126**, 57-70. [https://doi.org/10.1016/S0020-0255\(00\)00042-6](https://doi.org/10.1016/S0020-0255(00)00042-6)
- [7] Fuketa, M., Mizobuchi, S., Hayashi, Y. and Aoe, J. (1998) A Fast Method of Determining Weighted Compound Keywords from Text Databases. *Information Processing and Management*, **34**, 431-442. [https://doi.org/10.1016/S0306-4573\(98\)00012-0](https://doi.org/10.1016/S0306-4573(98)00012-0)
- [8] Witten, I.H., Paynter, G.W., Frank, E., Gutwin, C. and Nevill-Manning, C.G. (1999) KEA: Practical Automatic Keyphrase Extraction. In *Proceedings of the 4th ACM Conference on Digital Libraries*, Berkeley, August 1999, 254-255. <https://doi.org/10.1145/313238.313437>
- [9] Abu El-Khair, I. (2006) Effects of Stop Words Elimination for Arabic Information Retrieval: A Comparative Study. *International Journal of Computing & Information Sciences*, **4**, 119-133.
- [10] Atlam, E., Okada, M., Shishibori, M. and Aoe, J. (2001) An Evaluation Method of Words Tendency Depending on Time-Series Variation and Its Improvements. *International Journal of Information Processing and Management*, **38**, 157-171. [https://doi.org/10.1016/S0306-4573\(01\)00028-0](https://doi.org/10.1016/S0306-4573(01)00028-0)
- [11] Atlam, E., Morita, K., Fuketa, M. and Aoe, J. (2002) A New Method for Selecting English Field Association Terms of Compound Words and Its Knowledge Representation. *Information Processing and Management*, **38**, 807-821. [https://doi.org/10.1016/S0306-4573\(01\)00062-0](https://doi.org/10.1016/S0306-4573(01)00062-0)
- [12] Atlam, E., Fuketa, M., Morita, K. and Aoe, J. (2003) Document Similarity Measurement Using Field Association Terms. *Information Processing & Management Journal*, **39**, 809-824. [https://doi.org/10.1016/S0306-4573\(03\)00019-0](https://doi.org/10.1016/S0306-4573(03)00019-0)
- [13] Atlam, E., Ghada, E., Fuketa, M., Morita, K. and Aoe, J. (2006) An Automatic Deletion of Unnecessary Field Association Word Using Morphological Analysis. *International Journal of Computer and Mathematics*, **83**, 247-261. <https://doi.org/10.1080/00207160600875234>
- [14] Atlam, E., Sharif, U.M., Ghada, E., Fuketa, M., Morita, K. and Aoe, J. (2007) Improvement of Building Field Association Term Dictionary Using Passage Retrieval. *Information Processing and Management*, **43**, 1793-1807. <https://doi.org/10.1016/j.ipm.2006.12.006>
- [15] Atlam E., Abdelrahim, E. and Mansour, R.F. (2016) Retrieving and Building Structure Approach of NLP Knowledge to Improve the Disambiguation of Word Semantics. *The Institute of Electronics, Information and Communication Engineers (IEICE)*, **7**, 21-30.
- [16] El-Sayed,A., Ghaleb, F. Taha, A. and Ismail, A. (2017) A New Retrieval Method Based on Time Series Variation Using Field Association Terms. *International Journal of Mathematical Method Application and Science*, **41**, 5780-5791. <https://doi.org/10.1002/mma.4713>
- [17] Atlam, E., Dawlat, M., Ghaleb, F. and Abo-Shady, D. (2018) An Improvement of FA Terms Dictionary Using Power Link and Co-Word Analysis. *International Journal of Advanced Computer Science and Applications*, **9**, 236-24. <https://doi.org/10.14569/IJACSA.2018.090233>
- [18] Hasan, K.S. and Vincent, N. (2014) Automatic Keyphrase Extraction: A Survey of the State of the Art. *Proceedings of the 52nd Annual Meeting of the Association for*

*Computational Linguistics*, Baltimore, June 2014, 1262-1273.

- [19] Hasan, K.S. and Vincent, N. (2010) Conundrums in Unsupervised Keyphrase Extraction: Making Sense of the State-of-the-Art. *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, Beijing, 23-27 August 2010, 365-373.
- [20] Jones, K.S. (2004) A Statistical Interpretation of Term Specificity and Its Application in Retrieval. *Journal of Documentation*, **60**, 493-502.  
<https://doi.org/10.1108/00220410410560573>
- [21] Lee, S., Shishibori, M., Sumitomo, T. and Aoe, J. (2002) Extraction of Field-Coherent Passages. *Information Processing & Management*, **38**, 173-207.  
[https://doi.org/10.1016/S0306-4573\(01\)00032-2](https://doi.org/10.1016/S0306-4573(01)00032-2)
- [22] Saneifar, H., Bonniol, S., Laurent, A., Poncelet, P. and Roche, M. (2009) Terminology Extraction from Log Files. In: Bhowmick, S.S., Küng, J. and Wagner, R., Eds., *Database and Expert Systems Applications. DEXA 2009. Lecture Notes in Computer Science*, Vol. 5690, Springer, Berlin, Heidelberg, 769-776.  
[https://doi.org/10.1007/978-3-642-03573-9\\_65](https://doi.org/10.1007/978-3-642-03573-9_65)
- [23] Jiang, G., Sato, H., Endoh, A., Ogasawara, K. and Sakurai, T. (2005) Extraction of Specific Nursing Terms Using Corpora Comparison. *Proceedings of the AMIA Annual Symposium*, Washington DC, 22-26 October 2005, 997.
- [24] Hulth, A., Karlgren, J., Jonsson, A., Boström, H. and Asker, L. (2001) Automatic Keyword Extraction Using Domain Knowledge. *Proceedings of the 2nd International Conference on Computational Linguistics and Intelligent Text Processing*, Mexico, 18-24 February 2001, 472-482. [https://doi.org/10.1007/3-540-44686-9\\_47](https://doi.org/10.1007/3-540-44686-9_47)