Scientific Research Publishing

Journal of Computer and Communications, 2020, 8, 72-85
https://www.scirp.org/journal/jcc
ISSN Online: 2327-5227
ISSN Print: 2327-5219

A New Approach of Time Series Variation Based on Power Links and Field Association Words

Zohair S. Malki[1], El-Sayed Atlam[1,2], Talal H. Noor[1], Ahmad Reda Alzighaibi[1], Ghada Elmarhomy[1], Abdallah A. Mohamed[3,4]

[1]College of Computer Science and Engineering, Taibah University, Yanbu, KSA
[2]Faculty of Science, Tanta University, Tanta, Egypt
[3]Mathematics and Computer Science Department, Faculty of Science, Menoufia University, Menoufia, Egypt
[4]Computer Engineering & Computer Science, Unversity of Louisville, Louisville, USA
Email: satlam@yahoo.com

How to cite this paper: Malki, Z.S., Atlam, E.-S., Noor, T.H., Alzighaibi, A.R., Elmarhomy, G. and Mohamed, A.A. (2020) A New Approach of Time Series Variation Based on Power Links and Field Association Words. *Journal of Computer and Communications*, **8**, 72-85.
https://doi.org/10.4236/jcc.2020.83008

Received: January 29, 2020
Accepted: March 13, 2020
Published: March 16, 2020

Abstract

This paper has proposed a new methodology extracting stability classes of field association words depending on automatically power link analysis to enhance the precision of decision tree. In this paper, we have studied the effects of the time variation based on the frequencies of specific words called field association words that connected to documents using power link in a specific period. The stability classes have referred to the popularity of field association words based on the change of time in a given period. The new approach has evaluated by conducting experiments simulating results of 1575 files (about 5.16 MB). Based on these experiments, it has turned out that, the F-measure for ascending, stable and descending classes have achieved 93.6%, 99.8% and 75.7%, respectively. These results mean that F-measure was increasing by 12%, 4% and 34% than traditional methods because of the power link analysis.

Keywords

Information Retrieval, Time Changing, Power Link Analysis, Field Association Words, Decision Tree

1. Introduction

Recently, there are a huge amount of texts which are processed automatically in computers. Documents can be managed by computers to retrieve important information that can be used for searching, clustering and classifying, etc. Many words occur frequently in documents and generally strongly related to the field of that document. Extraction of important words from documents is highly use-

ful and conclusive for Information Retrieval (IR) branch. IR refers to the process of finding relevant information in topics that concern the user and then retrieving that information. Generally, the frequencies of words in texts change by the time and often linked with particular period of time. For example, words as "flu" is more spread in winter and "stormy" is more spread when there are strong winds and usually rains. Moreover, word groups that linked with words of search on a search engine as Yahoo and Google are changing with the time Ohkubo *et al.* [1]. Fields like sport and political figures always change with the time variant.

Traditional approaches [2]-[8] for searching, classifying, clustering and analyzing texts are incapable to determine specific words in a particular period of time. Atlam *et al.* [9] [10] approach points out that the popularity of words change according to the time using the ordinary keywords that represent documents. However, ordinary keywords are not representing documents correctly, because Atlam approach ignored the importance of connection between words and their fields.

By finding specific words in the document the reader can easily decide the field of the document without a need to read the whole text. These specific words or units are called Field Association (FA) words. FA words are the smallest number of words by which the reader of the text can decide the field of the text. Other method by Atlam *et al.* [11] represents FA words across the time based on the traditional FA algorithm of Atlam *et al.* [12]. However, this algorithm had some drawbacks, because some of irrelevant FA words that are not restricted to special field are selected as FA words candidates, while that are restricted to that particular field are not selected to be FA words [13] [14].

Power Links Analysis (PLA) was developed by Rokaya and Atlam [15] to analyze the co-occurrences of terms in the publications of a given topic which solve the previous drawback of Atlam approach. This approach is based on the advanced form of frequency as well as the distance between different instances of the given words. This additional information activates the concept of distribution of terms among different parts of a given document. PLA is based on the assumption that document words can give an enough representation of the document content and depends on the quantified frequencies and the distances between different instances of the two terms. Moreover, a time-series of such maps can trace the dynamic changes in this conceptual space. Further, PLA can be used to extract more relevant FA words from text documents [16].

The novelty of this paper is using Power Links Analysis (PLA), however the traditional methods have discussed the effects of the time change on the frequencies of FA words only. This means that there are some changes in the levels of some words after using their PLAs, which will affect on the DT results and improve the performance of the precision and recall as well as F-measure as shown in the paper results. Our paper suggests a methodology for automatic evaluation of the Stability (St) classes using the decision tree C4.5 algorithm of Quinlan [17] on the FA terms based on PLA. This methodology is assumed to

indicate the popularity of FA terms relying on the change of time and to improve the DT precision.

This paper is organized as follows. Section 2 introduces related work and FA words with their levels.

Section 3 presents the concept of PLA. Section 4 introduces our new methodology for evaluating St classes that indicate the popularity of FA words using PLA in a certain time period. Section 5 presents the experimental observations. Finally, conclusion and possible future work are introduced in section 6.

2. Related Work and FA Words

2.1. Related Work

Usually, IR field collection of documents changes based on time passes. The collection at time t contains, for an instance, M_t documents, K_t tokens, and term q which has q_t frequency. Whilst at time $t+1$, these computations will be changed. Therefore, the documents are added to or deleted from a collection according to the time and also the collection frequencies of words change. Atlam *et al.* [11] proposed a method to introduce the popularity of words with time based on their frequency in the past years texts data. This approach defines number of attributes and three classes of stability as the index of spread of words to obtain the frequency change of words quantitatively. Furthermore, decision tree is used to estimate these classes. However, this approach used the common keywords to represent the documents which are not the best representative. This method neglected entirely the importance of the relationship between words and their fields.

Atlam *et al.* [9] introduced the effects of changing the time on the frequency of specified words called FA words using the decision tree. They presented number of features to study the changing of FA words frequency according to the time and three stabilization classes that refer to the popularity of FA term across time. However, this method was depended on the traditional FA algorithm of Atlam *et al.* [10] which caused to produce some irrelevant FA terms that are not restricted to the specific field.

Co-word analysis the dynamics of science as a result of actor strategies. This technique should allow the reader in principle to identify the actors and explain the global dynamic [18]-[26]. Rokaya and Atlam [15] proposed a method of building dynamic FA words dictionary using PLA. Furthermore, this algorithm presented new rules to enhance the quality of FA terms dictionary in English. Moreover, the PLA algorithm used a technique to extract and refine the confusion sets to provide context-sensitive spell checking based on FA words [27]. This technique joined between the advantages of statistical and machine learning method of Rokaya *et al.* [27] which used to build a real word spell checker in English and Arabic. Also, the PLA approach was presented to classify the important and advertising messages and spam the reduction one [28] [29].

2.2. Field Association (FA)

1) ***FA WORDS***

In this paper, a document field is a fundamental and popular knowledge which can be utilized in human communication, for example, <MIDICINE/Diseases/Pollen Allergy> explains the path on tree with super-field <MIDICINE> with subfield <Diseases> and terminal field <Pollen Allergy> [5]. The tree structure was organized to illustrate the associations among document fields through the field tree Dozawa [30].

2) ***FA WORDS LEVELS***

Some FA words can be decided only by a specific field, whilst others may be decided by two or more fields. Relying on FA word success in referring to specific fields, there are five different levels as follows:

a) Ideal FA words: words associated with one sub-field (e.g., influenza, chemotherapy and insulin).

b) Semi-ideal FA words: words associated with some of sub-fields in one super-field (e.g., sneeze and cough).

c) Medial FA words: words associated with single super-field (e.g., blood and hospital).

d) Various FA words: words associated with some of sub-fields of different super-fields (e.g., program and win).

e) Non-FA words: not related to any field or decide it (e.g., rule and size).

The traditional algorithm of Atlam *et al.* [10] was used to judge these levels and to determine automatically the FA words based on term frequency and concentration ratio. The resulted FA terms are used as input to the algorithm of computing the ranks of FA words depending on PLA [31]. The traditional algorithm [10] takes as input the list of words selected from a corpus which comprises of groups of documents in different fields to judge the level of FA words.

3. Power Link Analysis (PLA)

The concept of PLA reflects the value of the word in terms of its relation to the words in the document. Moreover, each document will be presented by the average of the PLA between the current term and the terms in the same document.

PLA Steps

In the following sub-sections, we will introduce three main concepts for power link analysis as follows:

1) ***TERM to TERM PLA***

Supposed we have two terms t_1 and t_2 belongs to a document D, sometimes there is a link between t_1 and t_2 such link will be measured by the function $LT(t_1,t_2)$:

$$LT(t_1,t_2)=\frac{|D|*cr(t_1,t_2)}{average\ L(t_{1i},t_{2j})}$$

where $|D|$ is the number of different terms in document D, $cr(t_1,t_2)$ is the co-occurrence frequency of the two terms t_1 and t_2 in the document D, $L(t_{1i},t_{2j})$ is the distance between any two successive instants t_{1i} and t_{2j} of the terms t_1 and t_2. The value $average\ L(t_{1i},t_{2j})$ represents the average distance between any instants t_{1i} and t_{2j} of the terms in the document D. The function $LT(t_1,t_2)$ is symmetric, which means $LT(t_1,t_2)=LT(t_2,t_1)$.

2) *TERM TO DOCUMENT PLA*

By using the PLA between two terms as mentioned above, the PLA of a term to a document related to a given field can be represented by Figure 1.

The PLAs for n of FA words t and document D related to a given field $\langle S\rangle$ can be represented by $LTD(t,D,\langle S\rangle)$ as follows:

$$LTD(t,D,\langle S\rangle)=\sum_{fi\in\langle S\rangle}LT(t,fi)/n$$

where n is the number of FA words related to a field $\langle S\rangle$ and exist in document D.

3) *TERM TO FIELD PLA*

Term t to field $\langle S\rangle$ power link is represented by LTS as follows:

$$LTD(t,\langle S\rangle)=\frac{\sum_i LTD(t,\langle S\rangle)*crs(t,\langle S\rangle)}{nd}$$

where at least one FA words can exist in document D_i which is related to field $\langle S\rangle$, $crs(t,\langle S\rangle)$ represents the co-occurrence of the words t and the field $\langle S\rangle$, where $crs(t,\langle S\rangle)=\{f_i: f_i\in S,\min cr(f_i,t)>0\}+1$ In other words, $crs(t,\langle S\rangle)$ represents the number of FA terms that identify the field $\langle S\rangle$ and occur in a document whenever the term t occurs, and nd represents the number of documents that contain FA words that identify the field S and the term t.

Our new approach will use the algorithm of evaluating the levels of FA terms based on PLs [25] and then, check the effectiveness of the time change using new methodology. The selection of FA words is depended on using the PLAs.

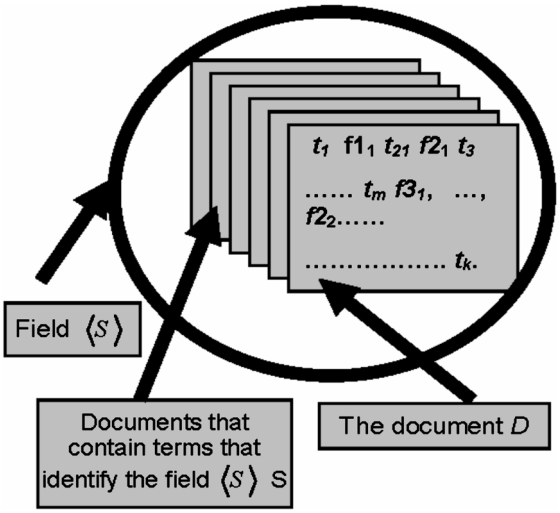

Figure 1. Terms to documents and field PL candidate.

4. Suggested Methodology

4.1. System Outcome

Figure 2 shows the outlines of the suggested approach. In this approach $Frequency\left(FA_k^{p_i}\right)$ represents the frequency of FA words depending on the PL k in a particular period p_i. However, $Total\_Frequency\left(FAs, p_i\right)$ represents the total frequencies of all FA words based on PLAs that lie in p_i. In order to accommodate the influence by the difference of FA words in each period with the changing of time rightly, the normalization frequency of FA words k in that period of time p_i, $Normalize\_Frequency_{ki}\left(FA_k^{p_i}\right)$, is represented by the following formula:

$$Normalize\_Frequency_{ki}\left(FA_k^{p_i}\right)=\frac{Freq\left(FA_k^{p_i}\right)}{Total\_Frequency\left(FAs, p_i\right)}$$

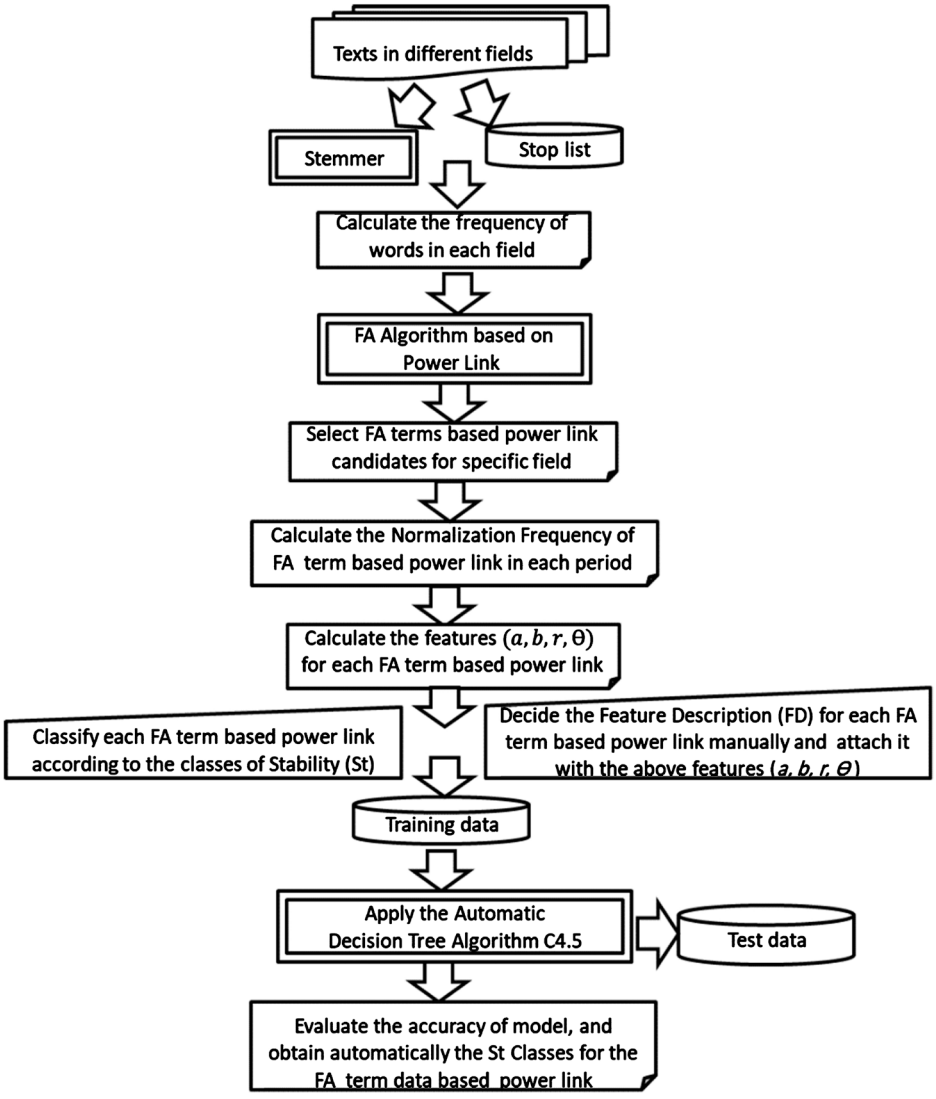

Figure 2. Suggested approach outline for judging the publicity of FA words with PLA according to time changing.

Definition 1: The FA words based on PLAs with increasing frequency according to the time change are called ascending FA words based on PLAs and its class is called ascending class. Whilst the FA words based on PLAs with descending frequency according to the time change is called descending FA words and its class is called descending class. Moreover, the FA words based on PLAs with stable frequency according to the time change are called stable FA words and its class is called stable class. These three classes are called FA words based on PLAs Stability (St) classes and are defined to determine how much the publicity of a specific FA word based on PLAs with the time change based on the changing of frequency of FA words in a given data.

The updated algorithm based on PLA:

Input: Corpus (collection of documents) $S=\{S_1,S_2,\cdots,S_m\}$, where each field S_j comprises of a set of documents $\{S_{j1},\cdots,S_{jn}\}$ that grouped according to a set of particular time periods $P=\{p_1,\cdots,p_n\}$, and words $w_1,w_2,\cdots,w_n$ that have PLA with that fields.

Output: Evaluation of St classes for the list of FA terms based on PLA.

Steps:

1. apply the FA algorithm of Rokaya *et al.* [25] that based on PLAs in each $p_i \in P$ on the collection of extracted Words (W) in field S with their attributes.

2. $W=\left[\left(w_1,\left[f_{11},f_{12},\cdots,f_{1m}\right]\right),\cdots,\left(w_t,\left[f_{t1},f_{t2},\cdots,f_{tm}\right]\right)\right]$.

3. where f_{qj} is the frequency of word w_q in field S_j, to get FA words collection based power link with the above five levels in each p_i, we will select the FA words levels for the topic of medical field. The output will be collection consists of FA terms lists based power link for the medical field in each given time, $FAT=\left[FA_{p_1},\cdots,FA_{p_n}\right]$, where $FA_{p_i}=\left[FA_1^{p_i},\cdots,FA_q^{p_i}\right]$ in period p_i, $i=1,2,\cdots,n$.

4. for each FA_{p_i} in FAT, do.

5. for each $FA_k^{p_i}$ in FA_{p_i}, do.

6. calculate $Norm\_Freq_i\left(FAT_i,P\right)=\dfrac{Freq\left(FAT_i,P\right)}{Total\_Freq\left(FATs,P\right)}$.

is the normalization frequency of FA word based on power link k in each p_i, $k=1,2,\cdots,q$ and $i=1,2,\cdots,n$.

7. end.

8. end.

9. get $FA\_NormFreq=\left\{\left(FAT_1;Norm\_Freq_1,\cdots,Norm\_Freq_n\right),\cdots,\left(FAT_n;Norm\_Freq_1,\cdots,Norm\_Freq_n\right)\right\}$.

10. for each K in FA_NormFrequency, do.

11. calculate $a=\dfrac{n\sum_{i=1}^{n}x_iy_i-\sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i}{n\sum_{i=1}^{n}x_i^2-\left(\sum_{i=1}^{n}x_i\right)^2}$ where a is the slope of the trend line, $t_i=p_i$ (the values of periods), $y_i=Normalize\_Frequency_{ki}\left(FA_k^{p_i}\right)$ (the values of the time series of normalization frequency of FA term based power link

$k$in each period of time p_i).

12. calculate $b = \frac{\sum_{i=1}^{n} y_i - a\sum_{i=1}^{n} x_i}{n}$, where b is the slice of the trend line.

13. calculate $r = (\text{sign of } a)\sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$, $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$, $\hat{y}$ represents the estimated values of y_i, $\hat{y}_i = ax_i + b$.

14. calculate $\theta = \tan^{-1}\frac{a_1 - a_2}{1 + a_1 a_2}$, a_1 and a_2 represent the gradients of first line obtained from ancient data (first two years) and second line obtained from all data ($a_2 = a$) respectively.

15. end.

16. let $FA\_Features = \left[(FAT_1; a, b, r, \theta, FD, St), \cdots, (FAT_q; a, b, r, \theta, FD, St)\right]$, where $FA\_Features$ is the FA words list based on power link with their features (a, b, r, θ, FD) and St classes, which is obtained after appending FD and St to the former calculations, FD represents the Feature Description of the FA terms based power link that are decided manually as described in sub-section 4.2.1.

17. DT C4.5 algorithm applied on the five features of FA terms based power link with their stability classes $(a, b, r, \theta, FD, St)$ as a training data.

18. To evaluate the accuracy of this model test data is used to obtain automatically the St Classes for the new test data.

In summary, the suggested algorithm is using for evaluating the stability (St) classes of FA words based on PLA automatically. The St classes refer to the publicity of FA words based power link across the time by depending on their frequencies.

4.2. FA Term Based on PLA Features

In the following sub-sections, five features of FA term based on PLAs are introduced as follows: Feature Description (FD), Slice of trend line (b), Slope of line (a), Angle between two lines (θ) Correlation confection (r) and the resulting amounts from these features are used as a training data for DT C4.5. These features are utilized to measure the frequency change of FA words based on PLAs with respect to time change to judge the stabi1lity (St) classes for FA words. More details about these attributes were discussed in [5]. Moreover, in this paper, we use the FD to describe terms of the medical field, as an example, as follows: a) "Disease-name: e.g., Cancer", b) "Treatment-name: e.g., Avastin", c) "Doctor-name: e.g., Magdi Yacoub", d) "Organization name: WHO", e) "Medical-name: e.g., Patient". These features are useful to study the influence of the time change on the frequencies and determine the St classes better and easily.

5. Experimental Observation

5.1. Experimental Data

We trained our new approach using a collection of data (corpus) obtained from

Internet web sites such as *Medical News Today* and *Independent News* (2014-2017). The total number of files is 1575 file with size about 5.16 MB. From the collected data corpus, the FA words based on PLAs with their levels are extracted. After that, select the first three specific levels of FA words based on PLAs that are related to the sub-field <diseases> under the medical field considering the time change of the frequencies for these selected FA words. The sub-field <diseases> of the medical field is privileged by frequent and constant articles every year, therefore, its terms are unique and tend to change with respect to time change. The data is divided into two groups; one is considered as the training data which is introduced to DT C4.5 as an input and the other group represents the test data which is totally different from the training data. The features of both groups are determined by using that frequency of FA words that result from using PLAs changes by the time. Table 1 shows a sample of the DT data for produced FA words based on PLA, which are Diabetes, FDA, Insulin and Plague.

5.2. Experimental Results

Firstly, the resulted FA words are improved by using the PLA. Table 2 shows comparison between samples of the FA words before and after using PLA. For

Table 1. A sample of dt data based on PLA.

a	*b*	*r*	*θ*	FD	class
−0.00620612473293	0.029561492688	−0.834218970995	0.0275088295695018	Di_names	i
−0.00117000712178	0.00502336680825	−0.969192210922	0.02950772989560018	O_names	c
−0.00206022993183	0.0106119100027	−0.708290516692	0.20377137307855486	T_names	i
−2.54349374302e−05	8.47831247667e−05	−0.866025403785	−0.0014573145647215727	Di_names	d

Where *a* mean Slope of line, *b* means Slice of trend line, *r* means Correlation confection, *θ* Angle between two lines, *FD*. Feature Description, class means stability *class* increasing, decreasing and constant.

Table 2. Comparison between new and traditional approach based on PLS.

Samples of words	Traditional system	New Approach based on power link
Cancer	Perfect term on E.1.0 (Cancer)	Perfect term on E.1.0
Sugar	Perfect term on E.1.3 (Diabetes)	Not Perfect term on E.1.3
Chemotherapy	Perfect term on E.1.0 (Cancer)	Perfect term on E.1.0
Viral	Perfect term on E.1.1 (Hiv/Aids)	Not Perfect term on E.1.1
Flu	Perfect term on E.1.7 (Influenza)	Perfect term on E.1.7
Aspirin	Perfect term on E.1.0 (Cancer)	Not Perfect term on E.1.0
Insulin	Perfect term on E.1.3 (Diabetes)	Perfect term on E.1.3
microbe	Perfect term on E.1.1 (Hiv/Aids)	Not Perfect term on E.1,1
fever	Perfect term on E.1.6 (Pollen Allergy)	Not Perfect term on E.1.6
pandemic	Perfect term on E.1.7 (Influenza)	Not Perfect term on E.1.7

example, the term "viral" took level 1 for <Hiv/Aids> field but after using the PLA the level changed to be level 4. Therefore, the PLA improved the results for our new approach as shown in Table 2.

From Table 2, it is clear that there are some changes in the levels of some words after using their PLAs, which will be affected on the DT results and improve the performance of the precision and recall.

Secondly, after training the DT by the training data, a comparison was done between automatic results by DT and manual results by human with respect to the classification of St classes of the tested data as in Table 3.

Table 3 represents the final result of the DT. Shaded rectangles mean the FA words based on PLA number that is determined correctly in both manual and automatic system of DT. The columns have FA words based on PLA number that evaluated by the DT and the rows have FA words based on PLA number that evaluated manually in each St Class.

To evaluate the DT system, three main terms in IR are called Recall (R), Precision (P) and F-measure are applied to each St class and defined as follows:

$$\text{Precision} = \frac{\text{Correct classified FA terms determined by decision tree}}{\text{Total classified FA terms determined by decision tree}}$$

$$\text{Recall} = \frac{\text{Correct classified FA terms determined by decision tree}}{\text{Total corrected classified FA terms determined by Human}}$$

$$\text{F\_measure} = \frac{2 * \text{Precision} * \text{Recal}}{\text{Precision} + \text{Recall}}$$

Table 4 introduces the accuracy of the new approach using the three evaluation terms R,P and F-measure rates to determine correctly the classified FA words based on PLA that are evaluated by the DT C4.5 depend on the frequency change with the time change.

Table 3. Resulted DT with manually evaluation.

DT Evaluation				
Inc. Class	Const. Class	Dec. Class		
650	1	76	Inc. Class	Human Evaluation
0	2837	8	Const. Class	
12	1	151	Dec. Class	

Table 4. Evaluate St classes based on PLA using R, P and F-measure.

New Methodology DT Evaluation depend on Power Link			
Evaluation Rate/St.	**Ascending**	**Stable**	**Descending**
P	98.2	99.9	64.3
Re	89.4	99.7	92.1
F-measure	93.6	99.8	75.7

5.3. Traditional and New Method Results Comparison

In this paper, the F-measure estimates the accuracy of the new methodology and the traditional method by Atlam *et al.* [12]. The traditional method uses the traditional algorithm [2] of FA terms which neglects the links among terms, documents and fields. Table 5 shows the P, R and F-measure rates for both the traditional and the new method.

Table 5 and Figure 3 show the F-measure rates for traditional and new based on PLA methods. Based on the evaluation results, it turns out that the performance is better when using the new method that depends on FA words with PLA. Moreover, the result of F-measure using the new based on PLA method in sub-section 5.2 is more correct than the traditional method. Generally, stable class has slightly improved. This is logic valid because it has stable frequencies of FA words depend on PLA according to time change.

Finally, it is clear that the effectiveness of the new methodology based on PLA is confirmed by using F-measure with improvement of the accuracy for ascending class by 12%, stable class by 5% and descending class by 34%, respectively.

6. Conclusions

This paper presented a new methodology to produce St classes of classified FA words based on PLA automatically. We have provided a detailed overview of the suggested method and its algorithm and presented our evaluation. The results from our evaluation indicate that the performance of our new method is better than traditional method performance. In conclusion, the effectiveness of the new methodology based on PLA is confirmed by using F-measure for ascending class

Table 5. Comparison of new and traditional methods using F-measure.

	Traditional Method DT Evaluation			New Method DT Evaluation		
Evaluation Rate/St.	Ascending	Stable	Descending	Ascending	Stable	Descending
P	82.8	95.6	43.8	98.2	99.9	64.3
R	80.2	96.4	39.8	89.4	99.7	92.1
F-measure	81.5	95.998	41.7	93.6	99.8	75.7

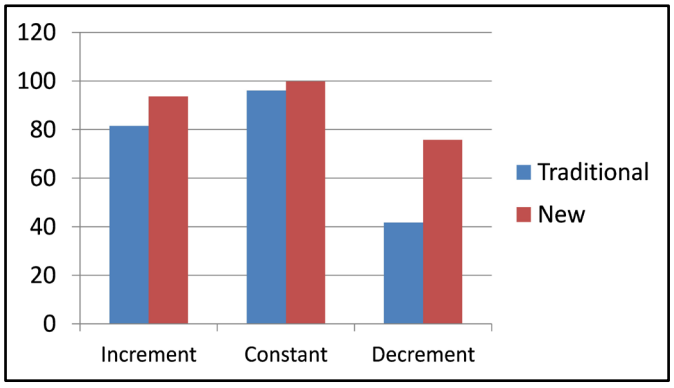

Figure 3. F-measure evaluation for traditional and new based power link methods.

as 93.6%, stable class as 99.8% and descending class as 75.7%, respectively.

Future work could focus on using compound FA words based on PLA and build Arabic dictionary based on FA words that can be produced from using PLA. Moreover, multi-language approach can be applied for this system to make cross-language information retrieval.

Conflicts of Interest

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

References

[1] Ohkubo, M., Sugizaki, M., Inoue, T. and Tanaka, K. (1998) Extracting Information Demand by Analyzing a WWW Search Login. *Transactions of Information Processing Society of Japan*, **39**, 2250-2258.

[2] Azzopardi, J. and Staff, C. (2012) Incremental Clustering of News Reports. *Algorithms*, **5**, 364-378. https://doi.org/10.3390/a5030364

[3] Chen, Y., Wang, W. and Liu, Z. (2011) Searching, Analyzing and Exploring Databases. In: Yu, J.X., Kim, M.H. and Unland, R., Eds., *Database Systems for Advanced Applications*. *DASFAA* 2011. *Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg, 65-88.

[4] Fukumoto, F., Yuzuki, S. and Fukumoto, J.I. (1996) An Automatic Clustering of Articles Using Dictionary Definitions. *International Journal of Transaction of Information Processing Society of Japan*, **37**, 1789-1799. https://doi.org/10.3115/992628.992699

[5] Hara, M., Nakajima, H. and Kitani, T. (1997) Keyword Extraction Using Text Format And Word Importance in Specific Field. *International Journal of Transaction of Information Processing Society of Japan*, **38**, 299-309.

[6] Liman, J. (1996) Cue Phrase Classification Using Machine Learning. *Journal of Artificial Intelligence Research*, **5**, 53-94. https://doi.org/10.1613/jair.327

[7] Prabaharan, S. and Wahidabanu, R.S.D. (2012) Support Vector Machine Based Classification of Clicked Document Using Topic Ontology for Profile Generation. *Information Technology Journal*, **11**, 1007-1015. https://doi.org/10.3923/itj.2012.1007.1015

[8] Sakurai, T. and Utsumi, A. (2004) Query-Based Multi-Document Summarization for Information Retrieval. *Proceedings of NTCIR*, Tokyo, Japan, 452-458.

[9] Atlam, E.-S., Makoto, O., Masami, S. 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*, **8**, 157-171. https://doi.org/10.1016/S0306-4573(01)00028-0

[10] Atlam, E.-S., Abdelrahim, E.M.D. 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*, **7**, 21-30.

[11] Atlam, E.-S., 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

[12] Atlam, E.-S., Morita, K., Fuketa, M. and Aoe, J. (2002) A New Method for Selecting English Field Association Terms of Compound Terms and Its Knowledge Representation. *International Journal of Information Processing and Management*, **38**, 807-821. https://doi.org/10.1016/S0306-4573(01)00062-0

[13] Atlam, E.-S., Elmarhomy, G., Fuketa, M., Morita, K. and Aoe, J. (2006) Automatic Building of New Field Association Word Candidates Using Search Engine. *International Journal of Information Processing and Management*, **42**, 951-962. https://doi.org/10.1016/j.ipm.2005.08.006

[14] Sharif, U.M., Elmarhomy, G., Atlam, E.-S., 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] Rokaya, M. and Atlam, E.-S. (2010) Search Engines Results Based on Power Links, *Journal of Computer Applications in Technology Special Issue on*: *Intelligent Text Processing with Its Applications and Computational Linguistics*, **38**, 298-305. https://doi.org/10.1504/IJCAT.2010.034530

[16] Atlam, E.-S., Fayed, G., Dawlat, A. El Mohamed, A. 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**, No. 2, 236-241. https://doi.org/10.14569/IJACSA.2018.090233

[17] Quinlan, J.R. (1993) C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, San Francisco, CA.

[18] Callon, M., Courtial, J.P. and Laville, F. (1991) Co-Word Analysis as a Tool for Describing the Network of Interactions between Basic and Technological Research: The Case of Polymer Chemistry. *Scientometrics*, **22**, 155-205. https://doi.org/10.1007/BF02019280

[19] Cao, L.J. and Tay, F.E.H. (2003) Support Vector Machine with Adaptive Parameters in Financial Time Series Forecasting. *IEEE Transaction on Neural Networks*, **14**, 1506-1518. https://doi.org/10.1109/TNN.2003.820556

[20] Hipel, K.W. and McLeod, A.I. (1994) Time Series Modeling of Water Resources and Environmental Systems. Elsevier, Amsterdam.

[21] Hyndman, R.J. and Athanasopoulos, G. (2013) Forecasting: Principles and Practice. OTexts, Online Open-Access Textbooks.

[22] Landgrebe, D. and Safavian, S.R. (1991) A Survey of Decision Tree Classifier Methodology. *IEEE Transactions on Systems*, *Man and Cybernetics*, **21**, 660-674. https://doi.org/10.1109/21.97458

[23] Chen, X., Chen, J., Wu, D.S., Xie, Y. and Li, J. (2016) Mapping the Research Trends by Co-Word Analysis Based on Keywords from Funded Project. *Procedia Computer Science*, **91**, 547-555. https://doi.org/10.1016/j.procs.2016.07.140

[24] Ying, D., Gobinda, G. and Schubert, F. (2001) Bibliometric Cartography Of Information Retrieval Research by Using Co-Word Analysis. *Information Processing & Management*, **37**, 817-842. https://doi.org/10.1016/S0306-4573(00)00051-0

[25] Ravikumar, S., Agrahari, A. and Singh, S.N. (2015) Mapping the Intellectual Structure of Scientometrics: A Co-Word Analysis of the Journal Scientometrics (2005-2010). *Scientometrics*, **102**, 929-955. https://doi.org/10.1007/s11192-014-1402-8

[26] Zhang, Y. (2015) Research Patterns and Trends of Recommendation System in China Using Co-Word Analysis. *Information Processing & Management*, **51**, 329-339. https://doi.org/10.1016/j.ipm.2015.02.002

[27] Rokaya, M. and Aljahdali, S. (2013) Building a Real Word Spell Checker Based on Power Links. *International Journal of Computer Applications*, **65**, 14-19.

[28] Rokaya, M. (2015) Arabic Semantic Spell Checking Based on Power Links. *ISSN International Information Institute*, **18**, No. 11.

[29] Rokaya, M. and Hemdan Dalia, I. (2016) Bibliometric Cartography of Nutrition Science Researches Based on Power Links Analysis. *ISSN International Information Institute*, **19**, No. 9(B).

[30] Dozawa, T. (1999) Innovative Multi-Information Dictionary, Imidas' 99. Annual Series, Zueisha Publication Co., Japan.

[31] Knees, P., Pampalk, E. and Widmer, G. (2004) Artist Classification with Web-Based Data. *Proceedings of the 5th International Symposium on Music Information Retrieval*, Barcelona, Spain, October 2004, 517-524.