An Improvement of FA Terms Dictionary using Power Link and Co-Word Analysis

El-Sayed Atlam
Information Sceinec and Intelligent System, Faculty of Enginerring, Tokushiam, Japan
Computer Science Division, Mathematical department, Tanta, Egypt

Fayed Ghaleb
Department of Mathematics, Faculty of Science, Ain Shames University, Cairo, Egypt

Dawlat A. El A.Mohamed
Department of Mathematics, Faculty of Science, Ain Shames University Cairo, Egypt

Doaa Abo-Shady*
Computer Science Division, Mathematical department, Tanta, Egypt

***Abstract*—Information retrieval involves obtaining some wanted information in a database. In this paper, we used the power link to improve the extracted field association terms from corpus by the proposed algorithm to support the machine to take the right decision and attach the candidate words in their convenient position in dictionary of the field association terms. Power Link is used as a quantitative tool to compute the co-citation relation among two words depending on the co-frequency and distances among instances of the words. In this paper, concept of the Power Link as well as modifications of the rules is used to classify the scientific papers into its proper field. In this paper, instead of whole document, a given document will be divided into three parts, namely, title, abstract and body. A given term will be given a weight that depends on the location of the term inside a specific document. The greatest weight will be given to the title then the abstract then the body, respectively. Results show an improvement in precision, recall and F measure.**

***Keywords*—Information retrieval; FA terms; co-word analysis; power link; precision; recall**

I. INTRODUCTION

Information retrieval (IR) defined as the activity of finding information resources related to an information need from a group of information resources. Searches can depend on whole document text or other content-based indexing. To provide automatic information retrieval systems, we can use several different retrieval techniques based on Field Association (FA) Terms and this paper concentrate on the concept of FA terms with co-word analysis [3].

Humans can understand the field of the scientific papers through detecting the particular terms, these terms called FA terms. Field of a document can be classified as: a super field, a sub field and terminal field, and the representation scheme of the document field called field tree [12]. For example, the path <Science& Technology/ COMPUTER/ Programming> expresses super field < Science& Technology > having sup field < COMPUTER > and terminal field < Programming > and the field code of this path can be defined by K.12.5.

FA terms are collected according to how well they refer to particular field. For example, "Communication network" and "compiler" are FA Terms of sup-field < COMPUTER >. As an FA Term may relate to more than one field, there are five levels used to rank FA terms as in [12]:

Level 1: The terms that specified to only one subfield and called Perfect FA terms.

Level 2: The terms that specified to more than one subfield and in only one super-field and called Imperfect FA terms.

Level 3: The terms that specified to one super-field and called Super FA Terms.

Level 4: The terms that specified to more than one subfield of more than one super field and called Cross FA Terms.

Level 5: The terms that do not assign any subfield or super-field and called Non FA Terms.

To choose the helpful FA terms need to consider the relations among simple and compound FA terms and field ranking. So, we need to use the co-word analysis and the Power Link concepts [18].

The co-word analysis is a quantitative study of relations between elements (i.e., terms or noun phrases or topics or fields). The inclusion and proximity indexes are used to compute the strength of relations among elements, these indexes depended on the co-occurrence frequency of elements. Co-word analysis focus on the dynamics of science as an outcome of actor methods. Changes in the content of a topic area are the common impact of a great number of individual strategies. This method must let us in principle to identity the actors and describe the global dynamic as in [11].

In [6], author presented an approach using the passage retrieval to improving constructing FA terms dictionary. They suggested a new method for locating FA terms using passage (parts of a document text) method instead of locating them from the full documents.

In [10], author provided the algorithm based on Power Link concept which explained and computed the relation among two words depended on the co-frequency and the relative locations of various successive instances. If words have nearer relative locations then the Power Link become bigger for those words.

In [13], author presented a method based on the Power Link concept to improve the classification of search engines results. This method depends on ranking the terms in a given field.

Depending on the absolute frequencies reflects the documents length rather than the weight of words, so recent works depend on normalized frequencies instead of absolute frequencies [10], [13], [19] and [20]. Also, recent works used the co-occurrence frequencies to reflect the relation between terms [4]. Power Link method uses the normalized frequencies, co-occurrence frequencies and considered the relative distances between terms.

While Power Link algorithm considers the whole documents, and gives the same weight for all parts of scientific paper, we will give different weight for different parts of a given scientific paper. In this work, the Power Link algorithm will be implemented, in addition to the another algorithm detect the pre-defined errors in Pre-text processing step presented by [7] to improve the quality of results and purge files from the resulting errors.

After collecting the corpus, in the pre-processing phase, every scientific paper will be divided into three parts, title, abstract and body. Each part will be given a different weight based on its importance. The title contains the most related terms to the topic and reflects the field of the document more than other parts. The abstract contains related terms to each other and reflects the field of the document more than the remainder body. So, we propose to give the terms that occur in the title the highest weight, then the abstract and give the body the least weight in the processing phase, the Power Link will be used to improve the FA terms dictionary. As a result, the proposed idea improved the Perfect FA terms (Level 1) and not improved in results of Imperfect and super FA terms (Level 2 and 3) so, level 1 is enough in our data. This idea can be used in many applications in information retrieval field.

The precision, recall and F measure values referred that the presented algorithm produced in average 0.90%, 0.85% and 0.87% respectively which means that the algorithm effective performance. The F value refers the strength of the algorithm.

The rest of article proceeds as the following: In Section 2, we presents a summary discuss of some definitions and modified algorithm. Sections 3 provide the modified algorithm for determining the Perfect FA terms (Level 1). Section 4 includes the results and discussion then in Section 5.

II. DEFINITIONS

A. Power Link Analysis

Power Link is a quantitative tool to determine the co-citation relationship among two terms depending on the frequency and the distances among instances of the terms [21]. In this paper, we used the Power Link as a tool to improve the extracted field association terms from corpus by the proposed algorithm.

The Power Link algorithm presented in [10] was provided calculations for how tow terms tend to occur altogether in a specific corpus. The Power Link value among two terms was high, if these terms are related together strongly.

The link between any two terms t1 and t2 in document D can calculated by the function of power link $LT(t_1, t_2)$ defined in Section 3.

B. Continuity and Transition Theme

Continuity and transition theme is a method to detect or determine the field of each part of a given document. The features of a subject are given based on continuity and transition. The theme field is defined as the field that a sentence presents, which is denoted by F_{theme} [14]. F_{theme} is preserved by *continuity* or changed by *transition* through sentences [9].

Let F_{theme} is field of sentence S that includes FA terms, then the power link among S and F_{theme} is computed by the field that gives $\max_j P(S, F_j)$ where, $P(S, F_j)$ is the Power Link among S and whole fields which expressed by the formula $P(S, F_j) = \sum_{i=1}^{n} P(FA_i, F_j)$ for each FA Term in F. So, the existing sentence is attached to the same passage If it has the equal F_{theme} as the previous sentence, or has no F_{theme}, or has no field. And S is delimited and a new passage starts if the existing sentence S has a different F_{theme} from the previous sentence, for more details see [5], [8] and [10].

Here, we can detect the three parts (title, abstract and body) by determining the head word of every part (i.e., abstract and introduction). If the head words are not present or repeated then we need to apply continuity and transition theme in this case. Always the first sentence on any document is the title that contains the most related words together and indicates to the field of the paper, the second paragraph usually is the abstract that contains a summary of all important information about the paper. So it contains the most important FA terms that indicate to the field of the paper and the power link between these terms should be high. So according to the previous rules we can detect and extract the abstract part from the document.

C. Real Word Spell Checker

Many words with multiple meanings exist in the English language. Technically, almost every word has a multiple meaning. How often do you go into the dictionary to look up a word, and find that only one meaning is listed next to it? Practically never! Many words have slightly varying meanings, or they can be used as different parts of speech.

For example (right: You were right./Make a right turn at the light, type: He can type over 100 words per minute./That dress is really not her type), (ate/eight, blew/blue, fair/fare, no/know).

To solve these problems, some algorithms were proposed to automatically detect such errors in syntax or meaning. In this work, to avoid these problems, we use the Real Word Spell Checker algorithm in Pre-text processing step. This

method depends on automatic building of errors that called confusion sets for a specific terms dictionary and corresponding corps. For more details see [7].

1) Algorithm for Calculating the Perfect FA Terms (PFAT)

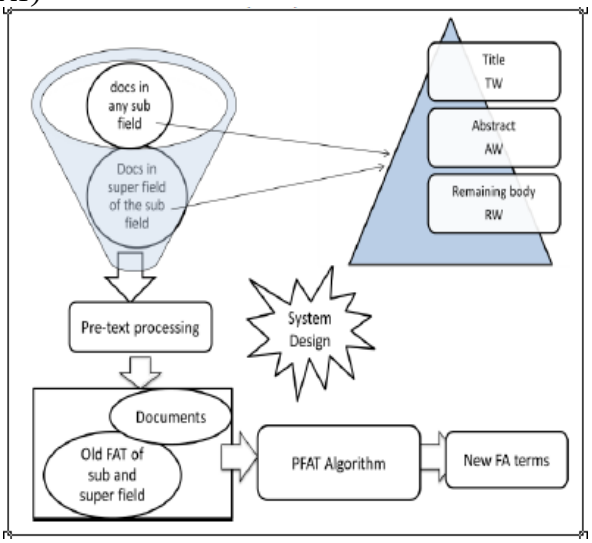

Fig. 1. System design.

Inputs:

a) Documents in any specified sub field and its super field after indexing to new terms candidate (ranked depending on their occurrence in document after stemming, removing stop words) to extract the new field association terms from them, as in Fig. 2 and Table I.

b) FA terms dictionary (by traditional algorithm by [12]) of sub and super field to be used in the Power Link calculations among them and the candidate terms in each document. Also, data of super field will be used to calculate the concentration ratio.

Programming Wireless Sensor Networks:
Fundamental Concepts and State of the Art
Abstract
Wireless sensor networks (WSNs) are attracting great interest in a number of application domains concerned with monitoring and control of physical phenomena, as they enable dense and untethered deployments at low cost and with unprecedented exibility.
1. INTRODUCTION
Wireless sensor networks (WSNs) are distributed systems typically composed of embedded devices, each equipped with a processing unit, a wireless communication interface, as well as sensors and/or actuators. Many applications have been proposed to date that show the versatility of this technology, and some are already ending their way into the mainstream.
2 L. Mottola and G.P. Picco
available to application developers [OnWorld ; CONET]. However, of the several experiences reported in the literature where WSN applications have been deployed in the real-world, only a few exceptions rely on some high-level programming support [Ceriotti et al. 2009; Buonadonna et al. 2005; Whitehouse et al. 2004].

Fig. 2. Sample of Docs Input.

Output:

A new set of improved FA terms:

We can demonstrate the system design and proposed algorithm in Fig.1 and 3 by the four main steps: 1) Power Link calculation, 2) Compute the Candidate Terms Frequency, 3) compute the concentration ratio, and 4) Compute the Precision and Recall Values, more details about those steps will be discussed in the following sub sections.

TABLE I. CANDIDATE TERMS AFTER INDEXING

program	**cost**	**technolog**
wireless	**unpreced**	**alreadi**
sensor	**exibl**	**nding**
network	**introduct**	**way**
fundament	**wireless**	**Mainstream**
concept	**sensor**	**mottola**
state	**network**	**picco**
art	**wsn**	**avail**
abstract	**distribut**	**applic**
wireless	**system**	**develop**
sensor	**typic**	**onworld**
network	**compos**	**conet**
wsn	**embed**	**howev**
attract	**devic**	**sever**
great	**equip**	**experi**
interest	**process**	**report**
number	**unit**	**literatur**
applic	**wireless**	**wsn**
main	**commun**	**applic**
concern	**interfac**	**deploy**
monitor	**well**	**real**
control	**sensor**	**world**
physic	**actuat**	**except**
phenomena	**mani**	**reli**
enabl	**applic**	**high**
dens	**propos**	**level**
Unteth	**date**	**program**
deploy	**show**	**Support**
low	**versatil**	

2) Power Link Calculations

For each candidate term t in each document D compute the following Power Link calculations:

a) Compute the Power Link between the term t and the sub field <S>:

$$\mathrm{LTS}(t, <S>) = [\sum_i \mathrm{LTD}(t, D_i, <S>) * \mathrm{crs}(t, <S>)] \; / \; nd \quad (1)$$

where: D_i is includes at least one FA term belong to $< S >$.

$\mathrm{LTD}(t, D_i, <S>)$ is the Power Link between t and D that will be compute it in b.

$\mathrm{crs}(t, <S>)$ is co-occurrence of term t and $< S >$ S.T: $\mathrm{crs}(t, <S>) = |\{f_i : f_i \in <S>, min\ cr\ (f_i, t) > 0\}| + 1$ is the number of FA terms identify $< S >$ and appear in D that the t appears.

nd is the number of documents that includes FA terms that identify $< S >$ and t.

b) Compute the Power Link between t and D respect to a given terminal field $< S >$:

$$\text{LTD}(t, D, < S >) = \sum_{f_i \in <S>} \frac{LT(t, f_i)}{n} \quad (2)$$

S.T.: n is number of $f_i = f_1, f_2, f_3, \ldots$ that are FA terms that identify $< S >$ and in D.

$\text{LT}(t, f_i)$ is link between t and f_i that will be compute it in (c).

c) Compute the Power Link between two terms t_j and f_i based on dividing the document:

Firstly: we have two constant terms (stems) in every doc are "abstract" and "introduct" according to the corpus are scientific papers.

let $\text{word1}_k =$ "abstract" and $\text{word2}_m =$ "introduct".

S.T: k is the index of word1 in D.

m is the index of word2 in D.

j is the index of t in D.

There are three cases to compute LT according to term t position:

Suppose ($\text{TW} > \text{AW} > \text{RW}$ are the title, abstract and reminder body weights respectively) so:

case 1: if $j > k$ (S.T: t position is in the title) then

$$\text{LT}(t_j, f_i) = \frac{|D| \times \text{cr}(t_j, f_i)}{\text{average } L(t_{j_r}, f_{i_s})} * \text{TW} \quad (3)$$

else,

case 2: if $k < j < m$ (S.T: t position is in the abstract) then

$$\text{LT}(t_j, f_i) = \frac{|D| \times \text{cr}(t_j, f_i)}{\text{average } L(t_{j_r}, f_{i_s})} * \text{AW} \quad (4)$$

else,

case 3: if $j > m$ (S.T: t position is in the body of paper) then

$$\text{LT}(t_j, f_i) = \frac{|D| \times \text{cr}(t_j, f_i)}{\text{average } L(t_{j_r}, f_{i_s})} * \text{RW} \quad (5)$$

where: $|D|$ is the number of different terms in document D, co-occurrence frequency $\text{cr}(t_j, f_i)$ of t_j and f_i in D and $L(t_{j_r}, f_{i_s})$ is the distance between any two successive instants t_{j_r} and f_{i_s} of t_j and f_i, such that there are no other instants of the term t_j or f_i between the instants t_{j_r} and f_{i_s} in D, note that, the extremes values are neglected. TW, AW and RW are reflects how much the relation between terms in each part of a document (i.e. Title, abstract and body, respectively). such that TW bigger than AW and AW bigger than RW because as usual the terms are more related together in the title more than the abstract also more than the body of the scientific researches and its values are determined by experiments.

Also, we used the continuity and transition to determine the abstract in case if the doc has problems to detect this part.

3) Compute the Candidate Terms Frequency

The frequency of a term t in a sub field $< S >$ is denoted by $F(t, < S >)$ then

$$F(t, < S >) = \sum_{D_i} f(t, D_i) \quad (6)$$

S.T: D_i is a document that includes FA terms that identify $< S >$ and $f(t, D_i)$ is defined as this formula:

$$f(t, D_i) = \log(\text{dtf}(t) + 1) / \left(\sum_{y \in D_i} \log(\text{dtf}(y) + 1) * \frac{U}{(1 + 0.0115 * U)}\right)$$

S.T.: $\text{dtf}(t)$ is number of times that term t occur in D_i .

$\left(\sum_{y \in D_i} \log(\text{dtf}(y) + 1)\right) =$ sum of $\log(\text{dtf} + 1)$ for whole terms in the D_i .

The local information and the normalization factor are given as these parts $\log \frac{\text{dtf}(t)+1}{\sum_{x \in d} \log(\text{dtf}(x)+1)}$ and $\frac{U}{(1+0.0115*U)}$ respectively [2].

U is the number of unique terms in D_i.

This formula is derived from the classic known formula 'TF $*$ IDF' (Term Frequency-Inverse Document Frequency) of Salton and used it in this algorithm instead of the traditional methods [12], [15], [16] and [6] that used the absolute frequency that only depend on the number of a term repetition in the document and not effective enough [1].

4) Compute the Concentration Ratio

The concentration ratio $\text{PL}(t, < S >)$ that based on the frequency and Power Link calculations can be used to judge whether or not the term t is a *Perfect FA term* and defined as:

$$\text{PL}(t, < S >) = \frac{F(t, <S>) * LTS(t, <S>)}{F(t, <S`>) * LTS(t, <S`>)} \quad (7)$$

Where $F(t, < S >)$ and $LTS(t, < S >)$ are frequency and Power Link calculations that will be computed in previous steps , since $< S >$ is the sub field, $< S` >$ is the super field of this sub field and by using threshold α to judge the levels of FA terms. Such that, If PL is less than value of α then t is not perfect term else t is perfect term.

5) Compute the Precision and Recall Values

To test the efficiency of the system we used the measurement of precision and recall to reach the best result of FA terms and its measure are

$$P_i = \frac{Number\ of\ Relevant\ FATs\ extracted\ by\ system}{total\ no.of\ FATs\ extracted\ by\ system} \quad (8)$$

$$R_i = \frac{\text{Number of Relevant FA terms extracted by system}}{\text{Total Number of FA words extracted Manually}} \quad (9)$$

where the termination condition of algorithm is $P_i - P_{i-1} < \mathcal{E}1$, $R_i - R_{i-1} < \mathcal{E}2$ where $\mathcal{E}1$ and $\mathcal{E}2$ are most low value as we wish, such that:

if $P_i - P_{i-1} < \mathcal{E}1$ and $R_i - R_{i-1} < \mathcal{E}2$ then algorithm will be the terminated, else repeat the processes.

(***End of algorithm)***

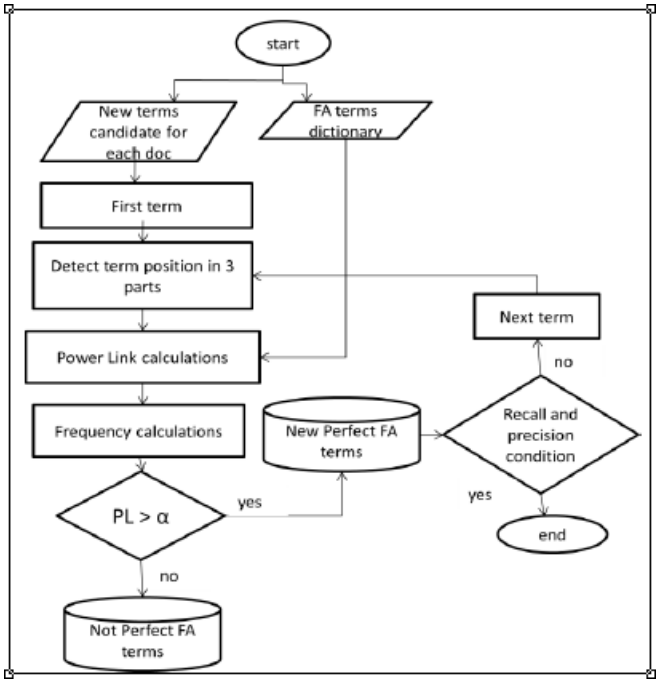

Fig. 3. PFAT Algorithm.

III. Experiments and Results

The experiments used to validate the advantage of the newly approach and that was the main purpose of it. Furthermore, we choose a most efficient weights along group of trials to provide good algorithm performance. Also, we write the code of our system by Python language that can be easily satisfied for any process on the text but there was a lot of challenges in pre-processing the text files to be formulated, like to convert from PDF file to txt file where some data can lost and there are not function in python can read from PDF file.

In this paper, we focus on a super field science and technology(K) and its sub-field Computer (K. 12) with corps size 12.2 MB about 4741 candidate terms were extracted.

Used the Real Word Spell Checker algorithm in pre-processing step led to discovery and correction 5% errors of the terms. Also we detect the three parts in 100 documents by use the continuity and transition theme. After the comparative analysis of the power link algorithm presented by [10], the proposed algorithm and some research information systems on scientific researches, it was recognized that giving different weights for each part could be improved selection of Perfect FA Terms (Level 1) but not improved of level 2 and 3 in our data. Table II show samples of perfect and not perfect FA terms that resulting from proposed algorithm (PFAT) and traditional algorithm [10], note that terms "Data, keyword and system" are detected as perfect by old method but they are not perfect FA terms in <Science& Technology\ Computer> field. We use $\mathcal{E} = 0.001$ and the threshold value = 0.9 that showed the best one for the concentration values in [17]. So, this threshold used as a fixed threshold for the concentration values in all loops and the average values of precision and recall are 0.90% and 0.85% respectively, as in Fig. 4. The results showed that the power links by weights do better than the random that produced the values of precision and recall in average 0.80% and 0.70%, respectively, as in Fig. 5. This means that, in this random data, the algorithm has efficiency 100% and to ensure the strong of the results, F is also calculated using the formula.

$$F = \frac{2*Precision*Recall}{Precision+Recall} \quad (10)$$

The average of new value of F is 0.87% while it was 0.74% using traditional method which refers a high performance of the system.

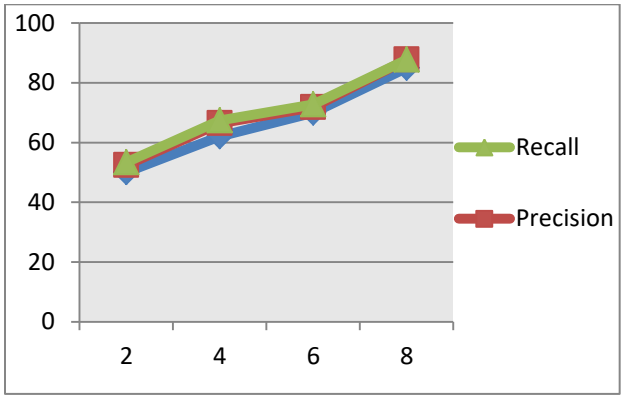

Fig. 4. Precision, recall and F measure by new approach.

TABLE II. Comparison of New and Traditional Approaches.

Term Samples	Old system	New system
Network	Perfect term in K.12	Perfect term in K.12
Data	Perfect term in K.12	Not Perfect term in K.12
Softwar	Perfect term in K.12	Perfect term in K.12
Keyword	Perfect term in K.12	Not Perfect term on K.12
Hardwar	Perfect term in K.12	Perfect term in K.12
System	Perfect term in K.12	Not Perfect term in K.12
Algorithm	Perfect term in K.12	Perfect term in K.12
Memori	Perfect term in K.12	Not Perfect term in K.12
Control	Perfect term in K.12	Not Perfect term in K.12
Structur	Perfect term in K.12	Not Perfect term in K.12
Code	Not Perfect term in K.12	Perfect term in K.12
Goal	Not Perfect term in K.12	Not Perfect term in K.12
Select	Not Perfect term in K.12	Not Perfect term in K.12
Imag	Not Perfect term in K.12	Not Perfect term in K.12

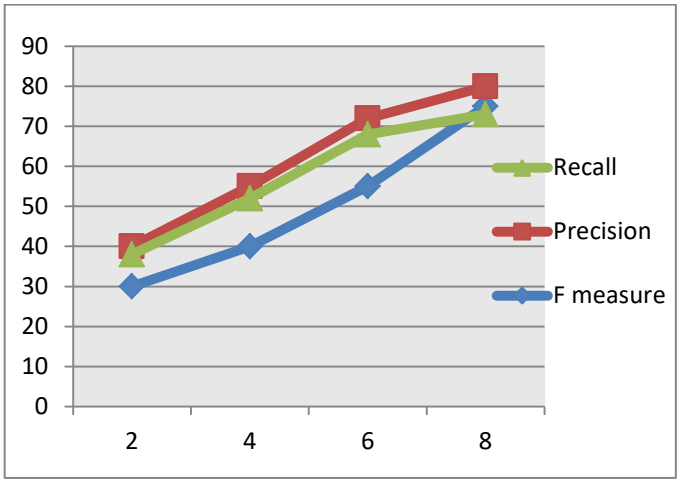

Fig. 5. Precision, recall and F measure by traditional approach.

IV. Conclusion

In this work we proposed an approach to produce an improvement FA terms dictionary by used Power Link concept and give different weights to terms according to their position in the document. The precision achieved using the new method 0.90. Hence, the algorithm succeeded to improve the values of precision by 10% than traditional approach.

Future work could focus on the importance to consider the difference between languages and cultures between English and Arab countries in the Middle East. Different languages can be implemented by doing some natural language processing & speech recognition researches using English, Japanese and Arabic languages. Also, this method can used in Building a comprehensive FA terms dictionary and can apply it in many of the applications especially in text summarization, text classifications, Extraction, filtering and machine translation.

Furthermore, we can apply the Power Link analysis using different weights not only on the scientific research but also any type of unstructured documents.

References

[1] Song S-K, Myaeng SH,"A novel term weighting scheme based on discrimination power obtained from past retrieval results, " Inf Process Manag, 2012 . http://dx.doi.org/10.1016/j.ipm.2012.03.004

[2] Cummins R, O'Riordan C, "Evolving local and global weighting schemes in information retrieval, " J Inf Retr vol.9, pp. 311–330, 2006.

[3] Chun H-W, Jeong C-H, Song S-K, Choi Y-S, Jeong D-H, Choi S-P, Sung W-K "Smart searching system for virtual science brain," LNCS 6890, pp. 324–332, 2011.

[4] Mahmoud Rokaya *, Elsayed Atlam, Masao Fuketa,Tshering C. Dorji, Jun-ichi Aoe, "Ranking of field association terms using Co-word analysis," Information Processing and Management, 2007.

[5] Mahmoud B. Rokaya, "Automatic text extraction based on field association terms and power links," *International Journal of Computer and Information Technology (ISSN: 2279 – 0764) vol. 02– no 06, November 2013.*

[6] Uddin, S., Elmarhomy, G., Atlam, E., Fuketa, M., Morita K. and Aoe, J., "Improvement of automatic building field association term dictionary using passage retrieval," *Information Processing & Management Journal*, vol. 43, 2007.

[7] Mahmoud Rokaya, AbdAllah Nahla, Sultan Aljahdali, "Context-sensitive spell checking based on field association terms," IJCSNS International Journal Of Computer Science And Network Security, vol. 12 no. 3 pp. 64-68, 2012.

[8] Oi Mean Foong1, Alan Oxley1 And Suziah Sulaiman1, "challenges and trends of automatic text summarization," International Journal Of Information And Telecommunication Technology, vol. 1, no 1, pp 34-39, 2010.

[9] Mahmoud Rokaya, "Automatic summarization based on field coherent passages," International Journal of Computer Applications, Published by Foundation of Computer Science, New York,USA, October 2013.

[10] Mahmoud Rokaya and El-Sayed Atlam, "Building of field association terms based on links," *Int. J. Computer Applications in Technology, vol. 38, no. 4, 2010.*

[11] Callon, M., Courtid, J. and Ladle, F, "Co-word analysis as a tool for describing the network of interactions between basicand technological research: the case of polymer chemistry," *Science Metrics*, vol. 22, no. 1, pp.155–205, 1991.

[12] Atlam, E., Morita, K., Fuketa, M. and Aoe, J, "Documents similarity measurement using field association terms," *Information Processing and Management*, Vol. 39, pp.809–824, 2003.

[13] Mahmoud Rokaya, "Improving Ranking of Search Engines Results Based on Power Links," *IPASJ International Journal of Information Technology (IIJIT)*, vol 2, no 9, September 2014.

[14] Lee, S., Shishibori, M., Sumitomo, S., Aoe, J., " Extraction of field-coherent passages," Information Processing And Management, vol. 38, pp. 173-207, 2002.

[15] Fuketa, M., Lee, S., Tsuji, T., Okada, M. and Aoe, J, "A document classification method by using field associationwords," *Information Science*, vol. 126, pp.57–70, 2000.

[16] Elmarhomy, G., Atlam, E., Fuketa, M., Morita K., Sumitomo, T. and Aoe, J, "Automatic deletion of unnecessary field association word using morphological analysis," *Journal of Computer and Mathematics*, vol. 83, no. 3, pp.247–262, 2006.

[17] Atlam, E., Elmarhomy, G., Morita, K., Fuketa, M. and Aoe, J, "Automatic building of new field association word candidates using search engine," *Information Processing &Management Journal*, vol. 42, no. 4, pp.951–962, 2006.

[18] Atlam E-S., Ghaleb F., Taha A., Ismail A., "A new retrieval method based on time seriesvariation using field association terms," Mathematical Methods in the Applied Sciences, in press.

[19] Mahmoud Rokaya, Dalia I. Hemdan, "Bibliometric cartography of nutrition science researches based on Power Links analysis," International Information Institute, vol.19, no.9(B), 2016.

[20] Mahmoud Rokaya, "Arabic Semantic Spell Checking Based on Power Links," International Information Institute, vol.18, no.11,2015.

[21] Mahmoud Rokaya, "Spam Reduction Based on Power Link Analysis," International Information Institute, vol.19, no.6(A),2016.