Validate Conference Paper Using Dice Coefficient

Nurhilyana Anuar & Abu Bakar Md Sultan

Faculty of Computer Science and Information Technology, University Putra Malaysia

43400 UPM Serdang, Malaysia

E-mail: shee_lyana@yahoo.com, abakar@fsktm.upm.edu.my

Abstract

Dice Coefficient is the techniques to find similarity of an object and widely used in digital library, sciences and other fields. Thus, this project is the first attempts to employed Dice Coefficient for selecting paper in conference management system. An experimental result with limited test cases indicates Dice Coefficient is potentially to be used in the broad spectrum of respective application.

Keywords: Dice coefficient, Validate, Similarity search, Conference management system

1. Introduction

The world is changing from paper based system towards online system so called web based system which is more efficient and productive. Most of the developing countries have started to implement the online system or system based to the respective area that they are doing. The aim is to increase their quality, efficiency, productivity and competitiveness in their service. Nowadays, searching can be done by system which is much faster compared to the past decades, searching of file or information has been done manually. Moreover, searching technique is the most popular technique used by the search engines such as Google or Yahoo!, to find the similar files in response to the user s queries.

The potential of web technologies has attracting academia and industries to move from conventional method of searching and matching to the new spectrum in solving their respective problem. One of the most frequently used of searching and matching activities is in organising conferences. Organising conferences is a tedious and time consuming activity. Normally the event should be properly plan and manage. Recently many organisers are moving their ways in organising the conferences through the web application. This type web application system is known as Conference Management System (CMS). CMS helps to reduce the workload of the organisers on certain activities but basically the process of reviewing and validating the papers are still done manually.

CMS is a vital medium for researcher to share new findings with others. Conference organizer is the backbone to make the process work well. Usually, in order to ensure the paper organization is going smooth the administrator or organizer will have to provide features receiving paper submissions, collecting reviewers' topic preferences, collecting conflicts of interest, assigning reviewers to papers, disseminating submissions to reviewers, collecting reviews, monitoring review coverage, sharing reviews among the program committee, ensuring independence of reviews, collecting final accepted versions, creating a conference website and program, publishing proceedings and others. Therefore, in order to improve the efficiency of the tasks which are being carried out, a solution is invented to reduce the administrator's works.

This paper presents prototype system of validate conference paper using similarity search technique. The searching algorithm is Dice Coefficient used for verifying the paper submitted by the author. This technique will find similar keywords that match with the theme of the conference that had been organize. Usually, validating the paper to match with the theme is done by administrator first. Thus, the purpose of using similarity search technique of collecting the conflict of interest is achieved.

2. Literature Review

As far as we are concerned, there is no research on using similarity search technique to validate the paper in Conference Management System. We made some analysis which shows some techniques have been applied in Conference Management System such as ConfSys(Huang, Y. Feng, B.C.Desai, 2008) which is using system parameter setup and common function to manage all the process of the conference. Moreover, in ConfSys2(Huang, Y. Feng, B.C.Desai, 2003), it improves the ConfSys by introducing user-group-function management and smart daemon conference management. (S.Ferilli, N.Di Mauro, T.M.ABasile, F.Esposito, and M.Biba, 2006) have proposed exploitation of intelligent technique for indexing and retrieving documents which are automatic task paper-review assignment by extract paper topic from the title and abstract in a scientific conference management. Furthermore, MYREVIEW(P.Rigaux, 2004) system used an iterative rating method for automatic paper assignment to the reviewer.

3. Proposed Prototype System Validation

We are concerned to improve the validation paper process in Conference Management System by using similarity search technique which are simple to implement and less complexity.

The prototype system developed is to improve the validation conference paper which previously took a lot of time for administrator to find the matches paper according to the theme. Mainly, the prototype system consists of 3 processes: Validate Process, Review Process, and Status.

The flow of this prototype system such as follows. There are 30 papers selected randomly which are consist of related and unrelated papers to the theme that have been stored in the database. In each paper, it consists of title and abstract. The validation process starts when the administrator defined few keywords significant to the theme. For this paper the chosen theme is Business Process Reengineering in E-Commerce.

3.1 Validation Process

Basically we adapted similarity search technique which is called dice coefficient algorithm to validate the paper in order to find the matches paper according predefined keywords set by the administrator. The technique we applied here was proposed by (G. Kondrak, 2005). Previously, they were using same concept as dice coefficient but it was calculated in bigrams which used to calculate the similarity measure of the string.

3.2 Review Process

During this process, all the matches paper have been validated will be submitted automatically to the reviewer screen. The decision will be made by the reviewer. The status of the paper consists of accept, reject and keep in view (KIV) by the reviewer.

3.3 Status Process

Status of the paper will be displayed after the review process completed by the reviewer.

4. Result and Discussion

According to the prototype system built, it shows that the efficiency of the paper to validate is less than a nanosecond. Besides, the accuracy of finding the keyword similarity is almost 95%.

5. Conclusion and Future Work

The result from this project indicates Dice Coefficient could be used for CMS. Dice Coefficient used simple statistical approaches and easily implemented. Since this is the first attempts to investigate Dice Coefficient for CMS, the result could be questionable because limited number of the test cases. Thus our future effort are looking to implement the system in to the real world problem

Furthermore, the model built in this paper is suggested to build into web based system or real system. This can help administrator to look on the pattern directly the process recorded into similarity search system. Thus, it will enhance the decision process in the organization.

References

G. Kondrak. (2005). N-Gram Similarity and Distance. *M.Consens and G. Navarro* (Eds.): SPIRE 2005, LNCS 3772.

Huang, Y. Feng, B.C.Desai. (2008). CONFSYS: A Web-based Academic Conference Management System. ACM, Proceedings of the 2008 C3S2E Conference.

Huang, Y. Feng, B.C.Desai. (2009) CONFSYS2: An Improved Web-based Multi-Conference Management System. ACM, Proceedings of the 2009 C3S2E Conference

P.Rigaux. (2004). An Iterative Rating Method: Application to Web-based Conference Management. *ACM Symposium on Applied Computing*.

S.Ferilli, N.Di Mauro, T.M.ABasile, F.Esposito, and M.Biba. (2006). Automatic Topics Identification for Reviewer Assignment. *Springer-Verlag Berlin Heidelberg*.

Notes

Note 1. This is an example of dice coefficient using bigram concept.

    Threshold value predefined as 0.5

$$\{ co, on, nt, te, ex, xt \} \dashrightarrow object\ x$$

$$\{co, on, nt, ta, ac, ct\} \dashrightarrow object\ y$$

$$Dice\ (context, contact) = (2 \cdot 3) / (6 + 6) = \ 0.5$$

Table 1. Frequency of matches' paper by using Dice's coefficient

Keyword (X)	1	2	3	4
Frequency	8	13	15	14

Table 2. Result on test keyword #1 (BPR)

PAPER REF NO.	ACCURACY SCORE
1002	0.8
1003	1.0
1005	1.0
1007	1.0
1008	1.0
1013	1.0
1024	0.8
1030	1.0

Table 3. Result on test keyword #2 (process reengineering)

PAPER REF NO.	ACCURACY SCORE
1001	0.933
1002	1.0
1003	1.0
1004	0.933
1005	1.0
1007	1.0
1008	1.0
1010	1.0
1011	1.0
1013	1.0
1022	0.933
1024	1.0
1030	0.844

Table 4. Result on test keyword #3 (business process reengineering)

PAPER REF NO.	ACCURACY SCORE
1001	0.872
1002	1.0
1003	0.952
1004	0.872
1005	0.952
1007	0.976
1008	0.611
1010	0.976
1011	0.952
1013	0.936
1015	0.667
1022	0.533
1024	1.0
1025	0.649
1030	0.815

Table 5. Result on test keyword #4 (e-commerce business process reengineering)

PAPER REF NO.	ACCURACY SCORE
1001	0.72
1002	0.792
1003	0.8301
1004	0.72
1005	0.8302
1007	0.8462
1008	0.5106
1010	0.8462
1011	0.8302
1013	0.8621
1022	0.72
1024	0.8077
1025	0.5106
1030	0.7384

Frequency of Similar Paper
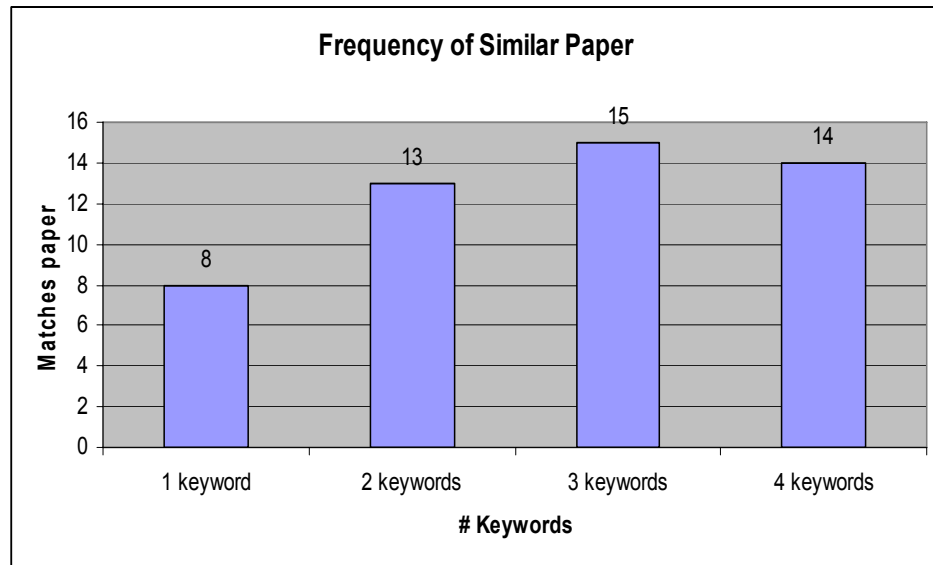
Figure 1. Frequency of similar paper

Average Efficiency
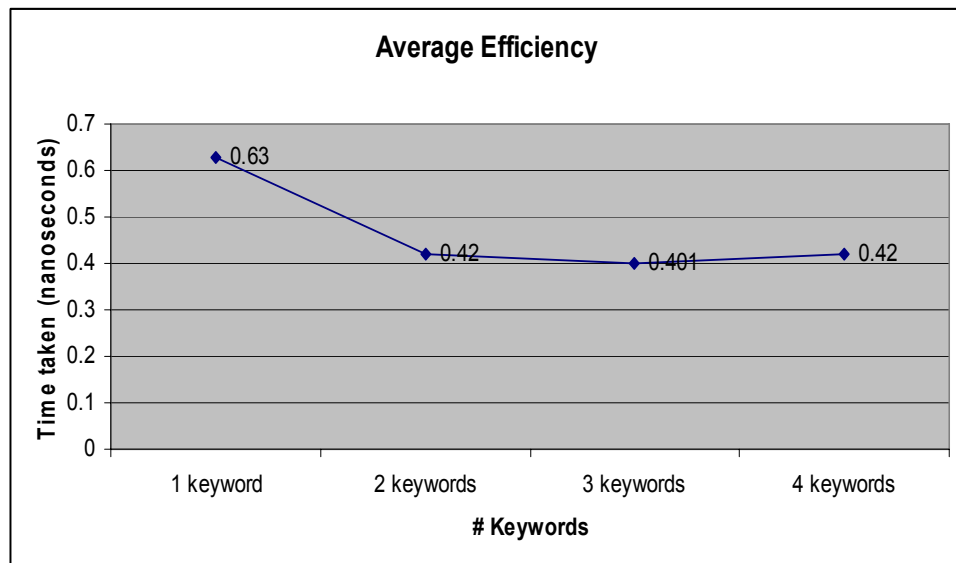
Figure 2. Average efficiency of matches' paper

Average Similarity
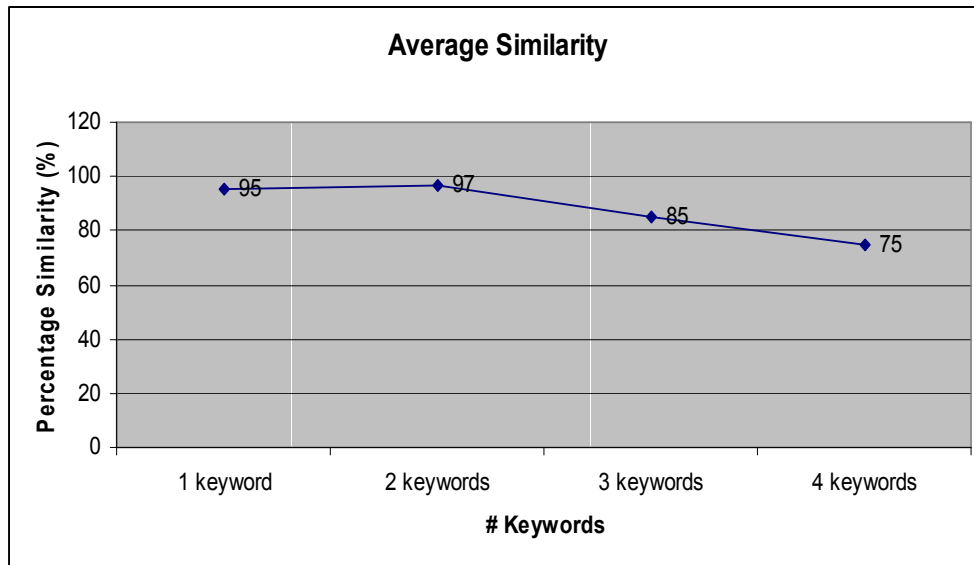
Figure 3. Average percentage of similar paper based on keywords

$$\frac{2 \times \left| n\text{-}grams(X) \cap n\text{-}grams(Y) \right|}{\left| n\text{-}grams(X) \right| + \left| n\text{-}grams(Y) \right|}$$

Figure 4. Mathematical expression for n-gram dice coefficient by G. Kondak
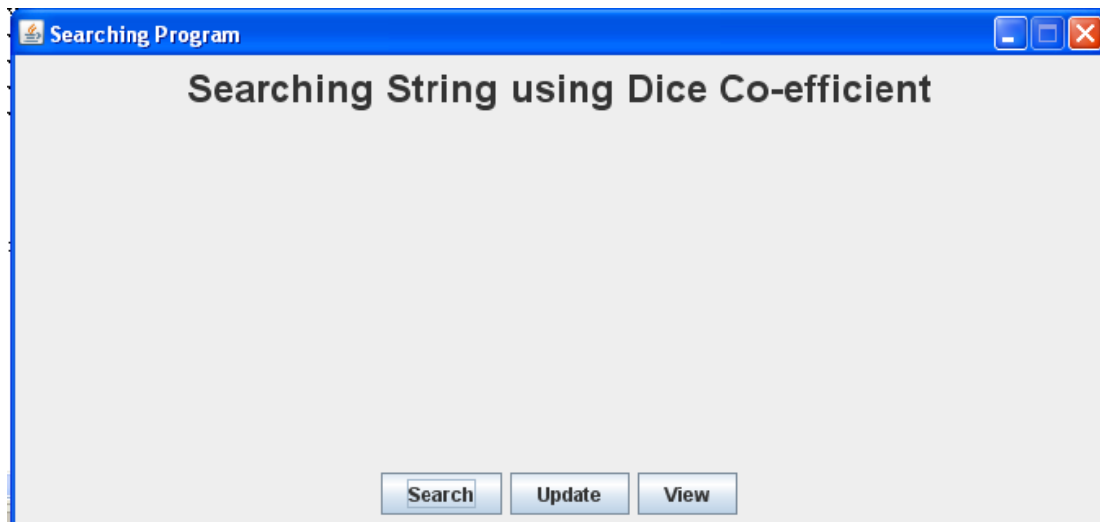
Figure 5. Screen Snapshot of Main Page
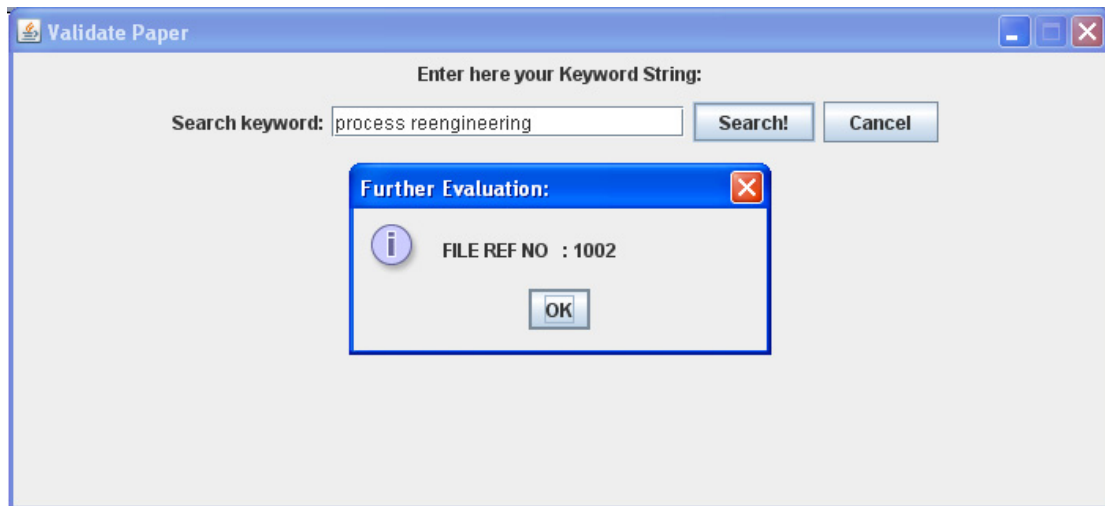
Figure 6. Screen Snapshot Validate Process
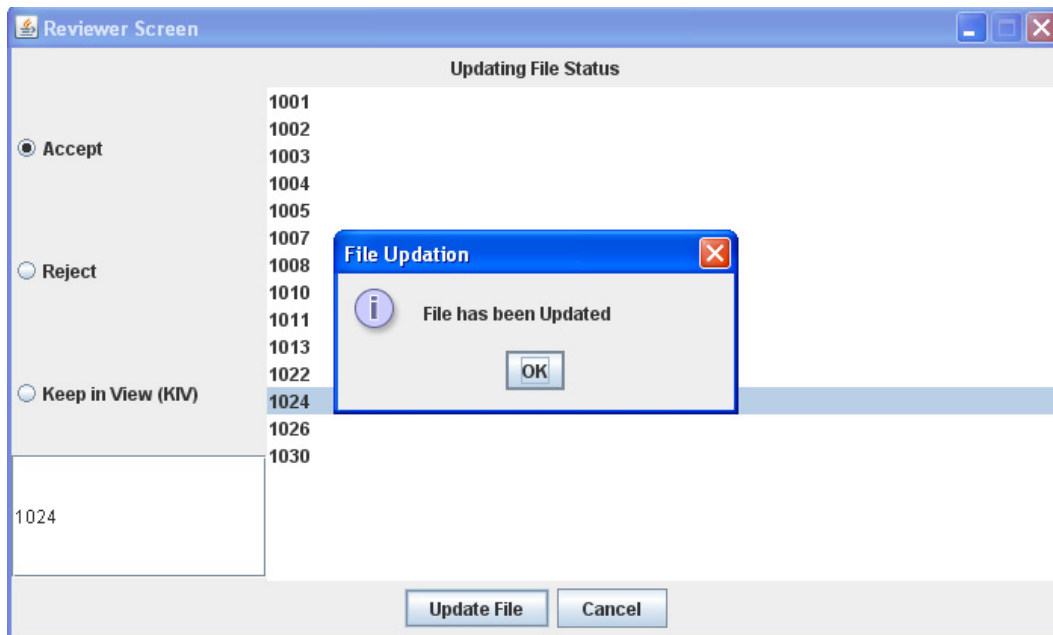
Figure 7. Screen Snapshot Review Process

Figure 8. Screen Snapshot Status