Search Engine Application using Fuzzy Relation Method for E-Journal of Informatics Department Petra Christian University

Leo Willyanto Santoso
Petra Christian University
Siwalankerto 121-131
Surabaya 60236 INDONESIA
+6231-2983455
leow@petra.ac.id

Rolly Intan
Petra Christian University
Siwalankerto 121-131
Surabaya 60236 INDONESIA
+6231-2983455
rintan@petra.ac.id

Prayogo Probo Susanto
Petra Christian University
Siwalankerto 121-131
Surabaya 60236 INDONESIA
+6231-2983455

ABSTRACT
Nowadays, scientific articles are easily obtained, because many researchers who conduct research discover new things. However, the increasing number scientific articles is not accompanied by the availability of applications to assist in the search for relevant articles. Today, available search engine applications perform only a search process based on string matching of search terms. In this research, a search engine application based on keyword relevance by using fuzzy relationship was developed.

This search engine application is built using PHP programming language and mysql as its database. Windows XP is used as the operating system. The used methods in fuzzy relationship are keyword used to paper, paper to paper, and paper to keywords and keyword to keyword. In addition, the components used to convert pdf files into plain text format.

Based on the results of experiments conducted, the process of searching for the 25 articles takes less than 5 seconds. For the indexing process, it is influenced by the number of pages per article.

Keywords
Fuzzy Relation, Search Engine, Paper.

1. INTRODUCTION
At this time, journal is one of the many forms of documents selected by the researchers and scientists to put the results of experiments or research that have been conducted. Through the journal, the researchers poured all aspects of the research that been conducted by attaching a detailed information about the research he had conducted.

A journal as a medium of information from the experts/researchers to the public media has an important role and very strategic. For example in the field of education, the journal serves as a good material for teaching materials (for teachers) or as a reference for students to learn a new science.

Currently, website has been highly developed. This resulted in a website used as a medium of publication of a journal from experts/researchers to the public or known by the name of the e-journal. But more and more of the e-journal are not followed by the use of search engines technology. The search engine on each of the e-journal that is useful to facilitate a user who wants to do a search on a journal and other journal/articles that may still relate to one another is needed.

The problem is how to design search engines on e-journal that can produce a related mutually journal to one another based on the keywords that are input by the user.

This paper presents a new search engine applications that do not only search on the similarity keyword provided by journal or scientific paper, but also provide a reference paper which relate to each other as well as journal is desired by the user.

The remaining part of this paper is organized as follows. Section 2 presents an overview of current proposal for dealing with fuzzy relation. Section 3 depicts the approach that we have delineated to solve the proposed problems. Section 4 discusses the performance of proposed methods. Finally, section 5 concludes the paper.

2. FUZZY RELATION
Fuzzy relation is a method for explaining the relationship of two different things (completely different). As illustration, the word "apple" (apple) and "tiger" (tiger) then in general the two words are not related. In general, the word "apple" refers to the name of the fruit and the "tiger" refers to the name of wild animal.

In the computer world there is manufacturer software, Macintosh (Mac). Mac has the brand "apple" so often referred to as the Apple Macintosh. Recently, Mac issued a new operating system called "Tiger" OS. From the relationship with Mac as the word "Apple" and "Tiger" is actually not related in general and in writing, have a relationship in the world of computers. Given the fuzzy relation then this kind of relationship will be examined with an assumption and goal that by knowing the relationship closeness/kinship between the two word/object. In relation to the world of search (searching), then by inserting the word "apple", there is the possibility of the word tiger will also be a result of output. Not because the results wrong, but because between the word "apple" and "tiger" there is kinship [4].

Explanation of fuzzy relation can also be described as follows: two words that completely unrelated (eg: "apple" and "tiger"), will have a relationship when both the word is addressed in one document. More and more documents that discuss both the relationship between the two words ("apple" and "tiger") will be getting closer. Fuzzy Relation will search 4 links from a combination of words (keywords) and documents (paper) these relationships are:

- *Keyword* to *paper*
- *Paper* to *paper*
- *Paper* to *keyword*
- *Keyword* to *keyword*

Explanation of each of this relationship along with the calculation process is described as follows:

1. At this step is assumed relationship between the keyword to the weight of the paper has value to the paper of the following keywords:

   $P = \{P_1, P_2, \ldots Pn\}$ is a *set of papers*

   $D = \{D_1, D_2, \ldots Dn\}$ is a *set of keywords*

   For example from the data obtained by paper and keyword relationship expressed as a fuzzy set of papers on the following keywords:

   $P_1 = \{0.3/D_2, 0.7/D_5, 1/D_7, 1/D_8\}$,

   $P_2 = \{1/D_2, 0.8/D_5, 0.8/D_7, 1/D_8\}$,

   $P_3 = \{0.9/D_1, 0.9/D_3, 1/D_4, 0.8/D_6\}$,

   $P_4 = \{1/D_1, 0.5/D_3, 0.8/D_4, 0.8/D_6\}$,

   $P_5 = \{0.1/D_2, 0.7/D_5, 1/D_4, 1/D_8\}$,

   $P_6 = \{0.9/D_2, 1/D_5, 0.8/D_4, 1/D_8\}$

   For $P = \{P_1, P_2, \ldots P_6\}$ and $D = \{D_1, D_2, \ldots D_8\}$, where each *paper*/document regarded as a fuzzy set of keywords so we get $\mu P_5 (D_1) = 0.1$.

2. *Similarity* between 2 *papers* expressed as a function of R where $R: P \times P \rightarrow [0.1]$ [5]

$$R(P_i,P_j) = \frac{\sum_D (\mu_{Pi}(D), \mu_{Pj}(D))}{\sum_D \mu_{Pj}(D)} \quad (2.1)$$

   Where:

   R: Relation

   P_i: *Paper*/document i

   P_j:*Paper*/document j

   D: *Keyword*

   μ: *Membership function* as a *mapping* $\mu P_i: D \rightarrow [0,1]$.

   Can find a relationship between a paper with one another, eg:

$$\text{relation } R(P_1,P_2) = \frac{0.3+0.7+0.8+1}{1+0.8+0.8+1} = \frac{2.9}{3.6} = 0.78$$

   The calculation of paper to paper as a whole can be seen in Table 1.

Table 1. Relation *Paper* to *Paper*

X / Y	P1	P2	P3	P4	P5	P6
P1	1,00	0,78	0	0	0,64	0,54
P2	0,93	1,00	0	0	0,64	0,73
P3	0	0	1,00	0,97	0,36	0,22
P4	0	0	0,83	1,00	0,29	0,22
P5	0,60	0,50	0,28	0,26	1,00	0,70
P6	0,67	0,75	0,22	0,26	0,93	1,00

3. From the existing data of keywords related to the paper, then we will get the paper on the relationship between keywords. Relationship of paper to keyword can be calculated using formula 2.2 [5]:

$$H_{Dj}(P_j) = \frac{\mu_{pi}(D_j)}{\mu_{pi}(D_1)+\mu_{pi}(D_2)+\ldots+\mu_{pi}(D_m)} \quad (2.2)$$

   Where:

   R: Relation

   P_i: *Paper*/document i

   P_j: *Paper*/document j

   D: *Keyword*

   μ: *Membership function* as a *mapping* $\mu P_i: D \rightarrow [0,1]$.

   Example: Calculate the weight of *keyword* (d2) for *paper* 1

$$\eta_{Dj}(P_j) = \frac{0.3}{0.3+0.7+1+1} = \frac{0.3}{3} = 0.1$$

   so the final result is:

   $D_1 = (0.25/P_3, 0.32/P_4)$,

   $D_2 = (0.1/P_1, 0.28/P_2, 0.06/P_5, 0.24/P_6)$,

   $D_3 = (0.25/P_3, 0.16/P_4)$,

   $D_4 = (0.28/P_3, 0.26/P_4, 0.36/P_5, 0.27/P_6)$,

   $D_5 = (0.23/P_1, 0.22/P_2, 0.25/P_5, 0.27/P_6)$,

   $D_6 = (0.22/P_3, 0.26/P_4)$,

   $D_7 = (0.33/P_1, 0.22/P_2)$,

   $D_8 = (0.33/P_1, 0.28/P_2, 0.36/P_5, 0.27/P_6)$

4. *Similarity* between 2 *keywords* expressed as a function of R where $R: D \times D \rightarrow [0,1]$ as written in the formula 2.3 [5]:

$$R(D_i,D_j) = \frac{\min(\quad_i(p), \quad_j(p))}{(\quad_{D_i}(p))}$$