

APLIKASI *SEARCH ENGINE* PAPER/KARYA ILMIAH BERBASIS *WEB* DENGAN METODE *FUZZY RELATION*

Bernard Adytia Darmadi

Alumnus Fakultas Teknologi Industri, Jurusan Teknik Informatika, Universitas Kristen Petra
e-mail: bernard.ad@gmail.com

Rolly Intan

Fakultas Teknologi Industri, Jurusan Teknik Informatika, Universitas Kristen Petra
e-mail: rintan@petra.ac.id

Resmana Lim

Fakultas Teknologi Industri, Jurusan Teknik Elektro, Universitas Kristen Petra
e-mail: resmana@petra.ac.id

ABSTRAK: Banyaknya jumlah *paper* yang dikoleksi sebuah lembaga pendidikan setiap tahun akan bertambah. Seiring dengan pertambahan jumlah *paper* tersebut maka diperlukan sebuah metode untuk mencari *paper* agar bila membutuhkan referensi maka *paper*/dokumen yang diperlukan dapat dengan mudah dapat ditemukan. Sejauh yang ada saat ini, kebanyakan mesin pencari masih mengandalkan pencarian dengan menggunakan *keyword matching/string matching* sehingga mengakibatkan hasil pencarian hanya akan menampilkan *paper-paper* yang mempunyai *keyword*/kata kunci yang dicari. Penelitian ini membahas sebuah sistem pencarian dengan menggunakan metode *fuzzy relation*, dimana dengan *fuzzy relation* didapatkan hubungan antara *keyword* dan *paper*. Dengan metode *fuzzy relation* maka sebuah pencarian mempunyai kemungkinan menampilkan hasil berupa *paper* yang tidak mengandung *keyword* yang dicari. Karena kata yang mengakibatkan *paper* (yang tidak mengandung *keyword* muncul) mempunyai hubungan dengan *keyword* yang dimasukkan.

Kata kunci: mesin pencari, relasi *fuzzy*, sistem cerdas.

ABSTRACT: *The number of paper collected by an educational institution is increasing each year. The increasing number of paper collected demand a method in order to find the right paper everytime there is someone who needs a reference. By far, most search engine still depend on keyword matching / string matching to find the appropriate result. This method will only find the appropriate paper based on the occurrence of the inserted keyword on the paper. This research will discuss a searching system using fuzzy relation, by using fuzzy relation the relation between keyword and paper is found and determined. Searching system using fuzzy relation allows the search result include paper that do not have the keyword to be shown as a result. This result is made possible because the word which occur in the paper is related to keyword inserted.*

Keywords: *search engine, fuzzy relation, intelligent system.*

PENDAHULUAN

Dokumen apa pun bentuknya pastilah disimpan dengan sebuah metode tertentu, dengan harapan bila dikemudian hari dokumen / data yang terkandung didalam dokumen diperlukan maka cukup dengan melakukan pencarian maka data yang diinginkan akan didapatkan dengan cepat dan juga hasilnya cukup relevan dengan apa yang dicari. Namun semakin banyak dokumen yang disimpan maka waktu pencarian juga akan meningkat dengan signifikan ditambah lagi juga dengan hasil pencarian yang tidak relevan karena ada banyak dokumen yang harus dipilah dan diteliti relevansinya dengan subjek yang dicari.

Oleh karena itu diperlukan sebuah metode untuk melakukan pencarian terhadap karya ilmiah yang telah dikoleksi. Mencari sebuah *pattern/pola* hubungan data pada kumpulan data yang tidak terstruktur adalah kunci dari *knowledge discovery* [3] dimana dengan adanya *pattern/pola* hubungan yang tepat untuk *knowledge discovery* maka informasi yang dihasilkan lebih relevan dan tidak terbatas pada bentuk masukan kata. Bentuk pencarian yang umum digunakan adalah dengan mencari seberapa sering sebuah kata kunci yang dimasukkan (dimasukkan sebagai pencarian) terdapat pada karya ilmiah, semakin banyak kata kunci yang ditemukan pada sebuah karya ilmiah maka dianggap bahwa karya

ilmiah tersebut semakin dekat hubungannya dengan kata kunci yang dimasukkan (*keyword/string matching*).

Penelitian ini akan menjabarkan sebuah metode pencarian dimana pencarian tidak hanya berdasarkan pada jumlah ditemukannya kata kunci pada karya ilmiah (*number of word occurrence*) namun juga melihat pada hubungannya dengan karya ilmiah dan kata kunci yang lain. Hubungan antara kata kunci dengan karya ilmiah dinyatakan dengan metode *fuzzy relation*.

FUZZY RELATION [2]

Fuzzy relation dijelaskan sebagai sebuah metode untuk menggambarkan hubungan antara dua objek (kata) yang berbeda. Sebagai contoh kata "genetic algorithm" dan "soft computing" merupakan dua kata yang berbeda. Secara sintaks maka tidak ada hubungan antara kata "genetic algorithm" dan "soft computing". Namun bila ada satu atau lebih karya ilmiah yang mengandung kata "genetic algorithm" dan "soft computing" maka dianggap bahwa terdapat sebuah hubungan antara keduanya.

Kemudian lebih jauh hubungan antara keduanya belum tentu kedekatannya. Semisal hubungan kata "soft computing" dengan "genetic algorithm" memiliki nilai 0.6 maka belum tentu sebaliknya hubungan kata "genetic algorithm" kepada "soft computing" memiliki nilai yang sama. Dari sini bisa terlihat juga kata mana yang mempunyai arti lebih luas dari yang lain.

Fuzzy relation pada penelitian ini meliputi 4 tipe yaitu :

1. *keyword to paper*
2. *paper to paper*
3. *paper to keyword*
4. *keyword to keyword*

Ke-empat tipe diatas diurutkan berdasarkan pada langkah – langkah perhitungannya.

Untuk langkah pertama yaitu *keyword to paper*. Nilai didapatkan dari hasil pencarian pada karya ilmiah dengan menggunakan *keyword/string matching* dengan melihat pada jumlah ditemukannya kata pada karya ilmiah. Kemudian dilakukan penentuan nilai fuzzy dengan rumus:

$$\text{nilai_keyword_terhadap_paper} = \min\left(1, \frac{n}{20}\right) \quad (1)$$

n = jumlah ditemukannya kata pada karya ilmiah (*number of word occurrence*). Nilai 20 merupakan nilai subjektif dimana sebuah karya ilmiah dikatakan *related* (bernilai fuzzy 1) bila kata ditemukan sebanyak lebih dari sama dengan 20 kali.

Berikut ini adalah ilustrasi perhitungan bila dilakukan pencarian dengan kata kunci jurusan teknik informatika.

Kombinasi kata:

"jurusan teknik"	-> 20 Occurance a)
"jurusan teknik informatika"	-> 10 Occurance b)
"jurusan informatika"	-> 5 Occurance c)
"teknik informatika"	-> 7 Occurance d)
"jurusan"	-> 30 Occurance e)
"teknik"	-> 20 Occurance f)
"informatika"	-> 25 Occurance g)

Perhitungan nilai *occurencinya*
occurance =

$$\begin{aligned} \dots a) & 20 * 2/3 = 13.33333 \\ \dots b) & 10 * 3/3 = 10 \\ \dots c) & 5 * 2/3 = 3.333333 \\ \dots d) & 7 * 2/3 = 4.666667 \\ \dots e) & 30 * 1/3 = 10 \\ \dots f) & 20 * 1/3 = 6.666667 \\ \dots g) & 25 * 1/3 = 8.333333 + \\ & 56.333333297 \text{ occurrence} \end{aligned}$$

Dengan menggunakan rumus (1) maka nilai fuzzy (*keyword to paper*) dari karya ilmiah yang bersangkutan dengan kata kunci jurusan teknik informatika adalah 1. kemudian dengan cara yang sama dimisalkan terdapat hubungan *keyword to paper* sebagai berikut:

1. Diasumsikan terdapat hubungan *keyword to paper* sebagai berikut:

$$\begin{aligned} P &= \{P_1, P_2, \dots, P_n\} \text{ adalah set of papers} \\ D &= \{D_1, D_2, \dots, D_m\} \text{ adalah set of keywords} \\ P_1 &= \{0.3/D_2, 0.7/D_5, 1/D_7, 1/D_8\}, \\ P_2 &= \{1/D_2, 0.8/D_5, 0.8/D_7, 1/D_8\}, \\ P_3 &= \{0.9/D_1, 0.9/D_3, 1/D_4, 0.8/D_6\}, \\ P_4 &= \{1/D_1, 0.5/D_3, 0.8/D_4, 0.8/D_6\}, \\ P_5 &= \{0.1/D_2, 0.7/D_5, 1/D_4, 1/D_8\}, \\ P_6 &= \{0.9/D_2, 1/D_5, 0.8/D_4, 1/D_8\}. \end{aligned}$$

2. Kemudian akan dilakukan pencarian hubungan *paper to paper* dengan menggunakan rumus [2]:

$$R(P_i, P_j) = \frac{\sum_D \min(\mu_{P_i}(D), \mu_{P_j}(D))}{\sum_D \mu_{P_j}(D)} \quad (2)$$

Keterangan:

R : Relasi

P_i : *Paper* / dokumen ke – i

P_j : *Paper* / dokumen ke – j

D : *Keyword*

μ : *Membership function*

sebagai suatu *mapping*

$\mu_{P_i} : D \rightarrow [0,1]$.

Contoh perhitungan *paper to paper* :

$$R(P_1, P_2) = \frac{0,3 + 0,7 + 0,8 + 1}{1 + 0,8 + 0,8 + 1} = 0,78$$

Hasil selengkapnya seperti pada tabel 1.

Tabel 1. Hubungan paper terhadap paper

X/Y	P ₁	P ₂	P ₃	P ₄	P ₅	P ₆
P ₁	1.00	0.78	0	0	0.64	0.54
P ₂	0.93	1.00	0	0	0.64	0.73
P ₃	0	0	1.0	0.97	0.36	0.22
P ₄	0	0	0.83	1.00	0.29	0.22
P ₅	0.60	0.50	0.28	0.26	1.00	0.70
P ₆	0.67	0.75	0.22	0.26	0.93	1.00

3. Perhitungan berikutnya adalah perhitungan *paper to keyword*. Perhitungan ini merupakan kebalikan dari langkah 1. dan dirumuskan sebagai berikut [2]:

$$\eta_{D_j}(P_i) = \frac{\mu_{P_i}(D_j)}{\mu_{P_i}(D_1) + \mu_{P_i}(D_2) + \dots + \mu_{P_i}(D_m)} \quad (3)$$

Keterangan :

R : Relasi

P_i : Paper / dokumen ke – i

P_j : Paper / dokumen ke – j

D : Keyword

μ : Membership function

sebagai suatu mapping

$\mu_{P_i} : P \rightarrow [0,1]$.

Contoh perhitungan:

mencari bobot *keyword* (D₂) pada *paper* 1

$$\eta_{D_2}(P_1) = \frac{0,3}{0,3 + 0,7 + 1 + 1} = \frac{0,3}{3} = 0,1$$

Hasil selengkapnya :

D₁ = {0.25/P₃, 0.32/P₄},

D₂ = {0.1/P₁, 0.28/P₂, 0.06/P₅, 0.24/P₆},

D₃ = {0.25/P₃, 0.16/P₄},

D₄ = {0.28/P₃, 0.26/P₄, 0.36/P₅, 0.27/P₆},

D₅ = {0.23/P₁, 0.22/P₂, 0.25/P₅, 0.27/P₆},

D₆ = {0.22/P₃, 0.26/P₄},

D₇ = {0.33/P₁, 0.22/P₂},

D₈ = {0.33/P₁, 0.28/P₂, 0.36/P₅, 0.27/P₆}

4. *Similarity* antara 2 *keyword* dinyatakan dengan suatu fungsi R dimana $R : D \times D \rightarrow [0,1]$ sebagaimana tertulis pada rumus [2] :

$$R(D_i, D_j) = \frac{\sum_P \min(\eta_{D_i}(P), \eta_{D_j}(P))}{\sum_P (\eta_{D_j}(P))} \quad (4)$$

Keterangan :

R : Relasi

P_i : Paper / dokumen ke – i

P_j : Paper / dokumen ke – j

D : Keyword

μ : Membership function

sebagai suatu mapping

$\mu_{P_i} : P \rightarrow [0,1]$.

Contoh perhitungan :

$$R(D_1, D_3) = \frac{0,25 + 0,16}{0,25 + 0,16} = 1$$

$$R(D_3, D_1) = \frac{0,25 + 0,16}{0,25 + 0,32} = 0,72$$

Hasil selengkapnya:

Tabel 2. Hubungan Keyword Terhadap Keyword

X/Y	D ₁	D ₂	D ₃	D ₄	D ₅	D ₆	D ₇	D ₈
D ₁	1.00	0	1.00	0.44	0	1.00	0	0
D ₂	0	1.00	0	0.26	0.64	0	0.58	0.53
D ₃	0.72	0	1.00	0.35	0	0.79	0	0
D ₄	0.89	0.44	1.00	1.00	0.54	1.00	0	0.51
D ₅	0	0.91	0	0.44	1.00	0	0.82	0.86
D ₆	0.84	0	0.93	0.41	0	1.00	0	0
D ₇	0	0.47	0	0	0.46	0	1.00	0.49
D ₈	0	1.00	0	0.54	1	0	1.00	1.00

Dari serangkaian perhitungan diatas maka akan didapatkan hubungan antara keyword dan paper serta kombinasi dari keduanya.

Perhitungan diatas masih belum cukup untuk menghasilkan pencarian dimana juga melibatkan hubungan dengan kata dan atau karya ilmiah lain pada halaman yang sama. Untuk mewujudkan hal tersebut maka dilakukan lagi satu tahapan perhitungan yaitu untuk menentukan hubungan dengan kata atau karya ilmiah lain serta menentukan juga bentuk urutan keluaran yang diinginkan (rangking). Perhitungan tersebut disebut sebagai *extended fuzzy*.

Perhitungan *extended fuzzy* dibahas dengan menggunakan contoh kasus sebagai berikut: Kata pencarian adalah K₁ dan K₁ mempunyai hubungan dengan kata yang lain. Hubungannya dijabarkan sebagai berikut: K₁ = {1/K₁, 0.8/K₇, 0.5/K₈}

Maka pencarian juga melibatkan kedua *keyword* diatas sehingga pencarian akan mencari K₁, K₇, K₈

Kemudian dari hasil pencarian masing – masing *keyword* didapatkan 3 hasil *fuzzy* set yang masing – masing mempunyai nilai sendiri – sendiri.

Mis :

P5 = {1/K₁, 0.5/K₇}

P6 = {1/K₇}

P7 = {0.4/K₁, 0.8/K₇}

P8 = {0.8/K₈}

$P_9 = \{0.7/K_8\}$
 $P_{10} = \{0.9/K_8\}$

Hubungan antara ketiga *fuzzy sets* diatas dibentuk dengan mengalikan nilai (*Dot Product*) relasi *keyword* dengan nilai element pada hasil, berikut adalah bentuk perhitungannya:

- K_1 dengan $P_5 \Rightarrow 1 * 1 = 1$
- K_1 dengan $P_8 \Rightarrow 1 * 0.8 = 0.8$
- K_1 dengan $P_7 \Rightarrow 1 * 0.4 = 0.4$
- K_7 dengan $P_6 \Rightarrow 0.8 * 1 = 0.8$
- K_7 dengan $P_7 \Rightarrow 0.8 * 0.8 = 0.64$
- K_7 dengan $P_5 \Rightarrow 0.8 * 0.5 = 0.40$
- K_8 dengan $P_9 \Rightarrow 0.5 * 0.7 = 0.35$
- K_8 dengan $P_{10} \Rightarrow 0.5 * 0.9 = 0.45$

Secara umum maka hasil pencarian dapat juga dituliskan menjadi sebagai berikut:

$\{1/P_5, 0.8/P_8, 0.4/P_7, 0.8/P_6, 0.64/P_7, 0.40/P_5, 0.35/P_9, 0.45/P_{10}\}$

bila terdapat kesamaan hasil (mis P_5 mempunyai nilai 1 dan 0.4) maka nilai tertinggi yang diambil, bila terjadi kesamaan nilai maka diambil salah satu.

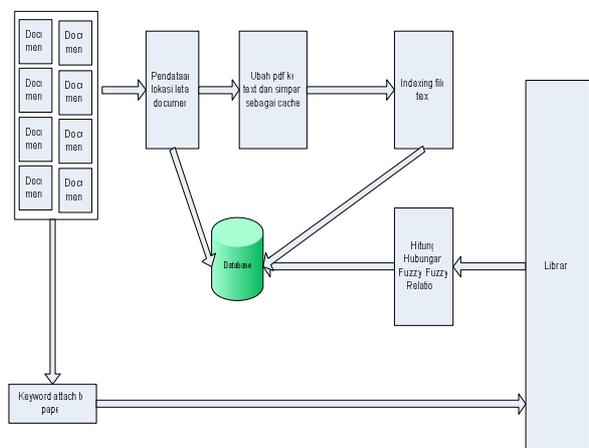
Sehingga hasil pencarian adalah:

$\{1/P_5, 0.8/P_8, 0.4/P_7, 0.8/P_6, 0.35/P_9, 0.45/P_{10}\}$

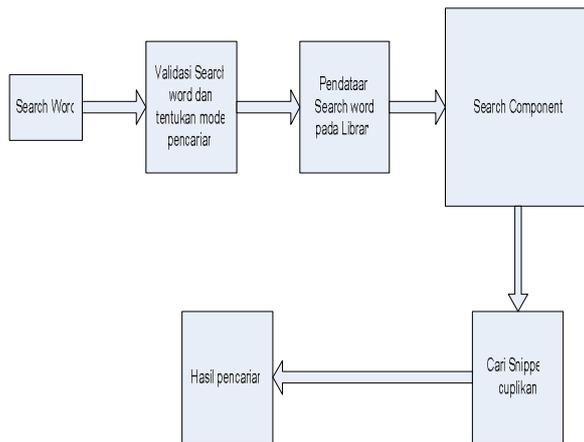
Terlihat bahwa bila dicari dengan cara biasa (*ordinary*) maka hasil pencarian adalah $\{1/P_5, 0.8/P_8, 0.4/P_7\}$

ANALISIS DAN DESAIN SISTEM

Secara umum maka desain sistem yang digunakan adalah sebagai berikut:



Gambar 1. Sistem Indexing



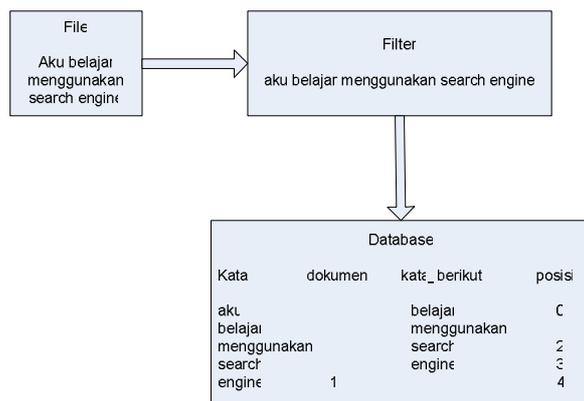
Gambar 2. Sistem Searching

Seperti pada umumnya sebuah *search engine* maka diperlukan sebuah mekanisme untuk mendapatkan data, dalam penelitian ini proses indexing adalah proses untuk mendapatkan data dan membentuk data tersebut untuk kemudian disimpan agar dapat dengan mudah diambil dan memenuhi kriteria pencarian yang diinginkan.

Pada penelitian ini, *search engine* yang dibuat dapat melakukan pencarian dengan bentuk inputan berupa kata (*word*) ataupun frase (*phrase*). Beda antara keduanya dilihat dengan ada tidaknya tanda petik (“) yang mengapit kata inputan.

Contoh: “fuzzy logic” merupakan frase
 fuzzy logic merupakan kata

Untuk mendukung pencarian dengan frase dan kata maka ada sebuah tabel yaitu *p_indexing*, yang didesain sedemikian rupa agar mempercepat proses pencarian.

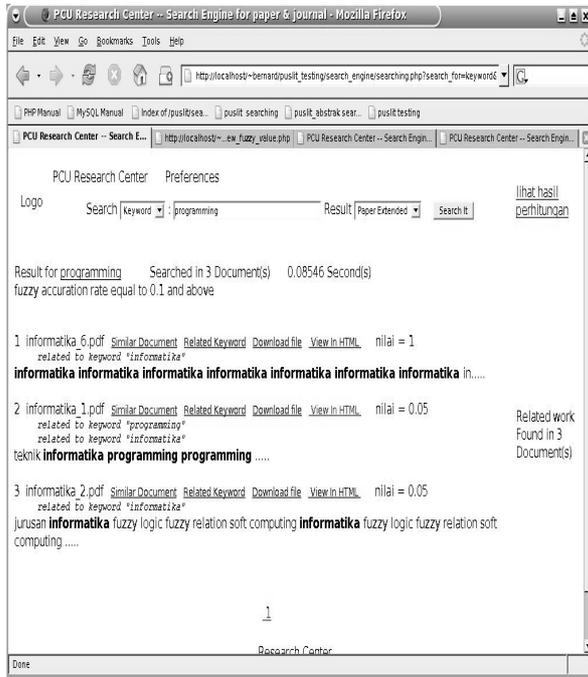


Gambar 3. Bentuk Tabel p_indexing

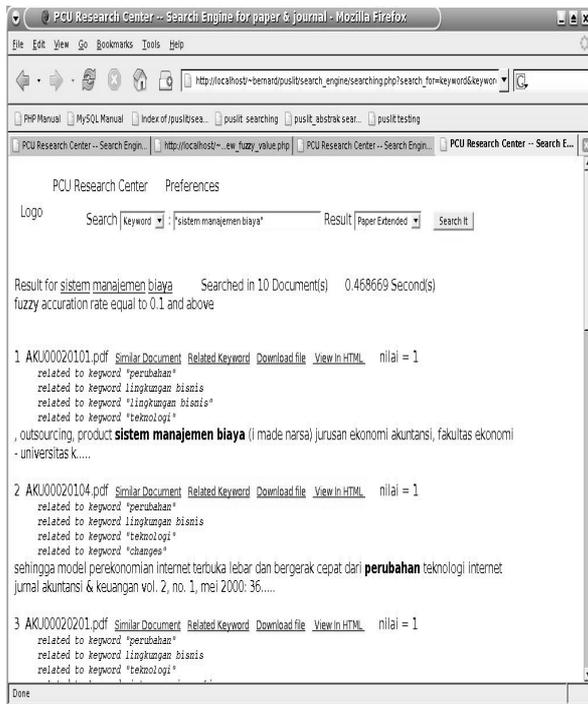
Dengan menggunakan desain seperti yang ada pada Gambar 3 maka jumlah pembacaan menjadi sebanyak $n-1$ kali, dimana bila pencarian melibatkan 3 kata maka pencarian mencari ke dalam database sebanyak 2 kali saja.

HASIL DAN PENGUJIAN

Pengujian dilakukan dengan dua cara yaitu dengan dokumen percobaan dan dokumen karya ilmiah asli yang dimiliki oleh pusat penelitian UK Petra.



Gambar 4. Hasil Pencarian dengan Menggunakan Dokumen Percobaan



Gambar 5. Hasil Pencarian dengan Menggunakan Dokumen Karya Ilmiah

Dari Gambar 4 terlihat bahwa dengan inputan kata programming, dokumen karya ilmiah yang tidak mengandung kata programming juga ikut keluar sebagai hasil pencarian.

Hal ini disebabkan karena kata informatika mempunyai hubungan dengan kata programming sehingga karya ilmiah yang mengandung kata informatika juga akan di tampilkan manakala kata programming dimasukkan sebagai kata pencarian.

Gambar 5 merupakan implementasi dari *search engine* pada dokumen karya ilmiah yang sebenarnya. Satu hal yang menjadi kendala dalam perhitungan *fuzzy relation* adalah masalah borosnya waktu perhitungan. Waktu perhitungan akan bertambah seiring dengan bertambahnya dokumen dan kata pencarian.

KESIMPULAN

Berdasarkan pada hasil pengujian maka dapat disimpulkan beberapa kesimpulan sebagai berikut:

1. *Search engine* dengan *fuzzy relation* mempunyai kelemahan yaitu waktu perhitungan yang cukup lama (bila dibandingkan dengan penggunaan *web*). Namun pada saat yang sama mempunyai keunikan yaitu dengan dapat ditemukannya relasi antara kata dan dokumen
2. *Search engine* biasa/*based on string matching* mempunyai kelebihan dalam kecepatan namun tidak dapat memberikan/menemukan relasi antara kata dan dokumen
3. Pembuatan *search engine* sebaiknya memisahkan dua aktifitas yaitu *indexing* dan *searching*. Karena proses *indexing* akan memakan *load server*.
4. Perhitungan *fuzzy relation* akan bertambah lama bila melibatkan lebih banyak dokumen dan lebih banyak *keyword* yang ada pada *library*

DAFTAR PUSTAKA

1. Castagnetto, Jesus, et al.(eds). *Professional PHP Programming*. PDF.
2. Intan, Rolly, and Mukaidono, Masao. "Toward a Fuzzy Thesaurus Based on Similarity in Fuzzy Covering". *Australian Journal of Intelligent Information Processing Systems*, vol 8 no 3, 2000.
3. Latiri, CH. C and Yahia, S. Ben. *Textmining: Generating association rules from textual data*. Computer Science department, Faculty of sciences of tunis, Tunisia