



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk menggubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

BAB V

IMPLEMENTASI DAN PENGUJIAN

5.1 Implementasi

Implementasi pada *real environment* dilakukan di *server* Universitas Multimedia Nusantara dengan alamat 192.168.2.111/08110110048. Adapun spesifikasi *server* yang digunakan adalah,

- a. *Processor*: Intel Xeon CPU E5540 @2.53GHz,
- b. RAM: 4096MB,
- c. *Operating System*: Windows Server Enterprise (6.0, Build 6001), dan
- d. Web Server Apache 2.0.

Sedangkan *browser* yang digunakan dalam pengujian aplikasi adalah Mozilla Firefox 11.0 yang dijalankan pada resolusi layar 1280 × 800.

5.2 Proses Uji Coba dan Dataset

Uji coba dilakukan dengan proses *training* yang melibatkan seluruh dokumen web di dalam *resources*. Kemudian, hasil dari *training* tersebut dianalisa untuk dibandingkan dengan penelitian Darmawan Fatriananda[3] dan diukur rata-rata waktu eksekusi proses propagasinya. Sedangkan proses *training* ini dilakukan dengan dua tipe yaitu menggunakan proses *harvesting* dan tanpa proses *harvesting*. Adapun langkah-langkah yang dilakukan untuk melakukan proses *training*:

- a. Menentukan nilai *pk* yaitu jumlah maksimum kata pada *keywords*, di mana *pk* mempengaruhi penggolongan dokumen web yang dianggap miskin metadata dengan dokumen web yang dianggap kaya metadata. Sedangkan uji coba pada penelitian ini, nilai *pk* = 20.

- b. Menentukan nilai rk yaitu jumlah maksimum kata yang direkomendasikan kepada setiap dokumen web untuk melengkapi *keywords*. Sedangkan uji coba pada penelitian ini, diawali dengan $rk = 4$.
- c. Menentukan nilai minimum bobot *edge* yang menjadi tolak ukur kualitas hubungan *cooccurrence* antar *vertex*, di mana nilai tersebut mempengaruhi status keterhubungan antar *vertex* dan secara tidak langsung mempengaruhi jumlah kata yang direkomendasikan. Sedangkan uji coba pada penelitian ini, nilai minimum bobot *edge* adalah nilai rata-rata dari keseluruhan bobot *edge* yang ada pada FCON.
- d. Jika suatu proses *training* tidak merekomendasikan kata baru atau nilai *precision* mendekati 1, maka nilai rk diperbarui menjadi $rk = 2 \times rk$.
- e. Proses *training* dilakukan berulang kali hingga nilai $rk = pk$.

Dataset yang digunakan sebagai input aplikasi berupa dokumen XML yang diperoleh dari *tools* proximity. Proximity sendiri merupakan sistem *open-source* untuk penemuan pengetahuan relasional yang dirancang dan diimplementasikan oleh *Knowledge Discovery Laboratory* pada *Department of Computer Science* di University of Massachusetts Amherst[25]. Dokumen XML tersebut bernama proxwebkb.xml (Gambar 5.1) dengan ukuran 25,673,147 *byte*, yang berisi informasi tentang sekumpulan dokumen web dan dokumen multimedia beserta keterhubungannya.

Properti *keywords* diperoleh berdasarkan informasi pada tag `<ATTRIBUTE NAME="page_words_top100" ITEM-TYPE="O" DATA-TYPE="str">`. Sedangkan pada penelitian ini, *object* yang dapat diterima oleh aplikasi adalah *object* dengan *keywords* yang terdiri dari minimal satu *keyword*. Berdasarkan proses eksplorasi tersebut, diperoleh 3026 dokumen web, 8 dokumen multimedia, dan 12 relasi antara dokumen web dengan dokumen multimedia.

6	<!DOCTYPE PROX3DB SYSTEM "prox3db.dtd">
7	<PROX3DB>
8	<OBJECTS>
4145	<LINKS>
5081	<ATTRIBUTES>
5082	<ATTRIBUTE NAME="school" ITEM-TYPE="O" DATA-TYPE="str">
7489	<ATTRIBUTE NAME="url_server_info" ITEM-TYPE="O" DATA-TYPE="str">
0193	<ATTRIBUTE NAME="url_protocol" ITEM-TYPE="O" DATA-TYPE="str">
2600	<ATTRIBUTE NAME="url_hierarchy1" ITEM-TYPE="O" DATA-TYPE="str">
2226	<ATTRIBUTE NAME="url_hierarchy2" ITEM-TYPE="O" DATA-TYPE="str">
7538	<ATTRIBUTE NAME="url" ITEM-TYPE="O" DATA-TYPE="str">
9945	<ATTRIBUTE NAME="page_num_outlinks" ITEM-TYPE="O" DATA-TYPE="int">
2352	<ATTRIBUTE NAME="page_num_inlinks" ITEM-TYPE="O" DATA-TYPE="int">
4759	<ATTRIBUTE NAME="url_hierarchy1b" ITEM-TYPE="O" DATA-TYPE="str">
4385	<ATTRIBUTE NAME="page_words_top100" ITEM-TYPE="O" DATA-TYPE="str">
8249	<ATTRIBUTE NAME="pagetype" ITEM-TYPE="O" DATA-TYPE="str">
5656	<ATTRIBUTE NAME="link_id" ITEM-TYPE="L" DATA-TYPE="str">
8460	<ATTRIBUTE NAME="link_tag" ITEM-TYPE="L" DATA-TYPE="str">
1264	</ATTRIBUTES>
1265	</PROX3DB>

Gambar 5.1 Struktur Proxwebkb.xml

5.3 Proses Pembentukan FCON Pada Aplikasi

Pembangunan FCON dimulai dengan melakukan iterasi kepada setiap dokumen web yang terdapat pada *database*. Kemudian pada setiap dokumen web dilakukan iterasi pula terhadap *keywords* yang terdapat pada dokumen web terkait. Sebagai contoh, terdapat tujuh dokumen web yang dilengkapi dengan *keywords*, seperti yang ditunjukkan pada Tabel 5.1.

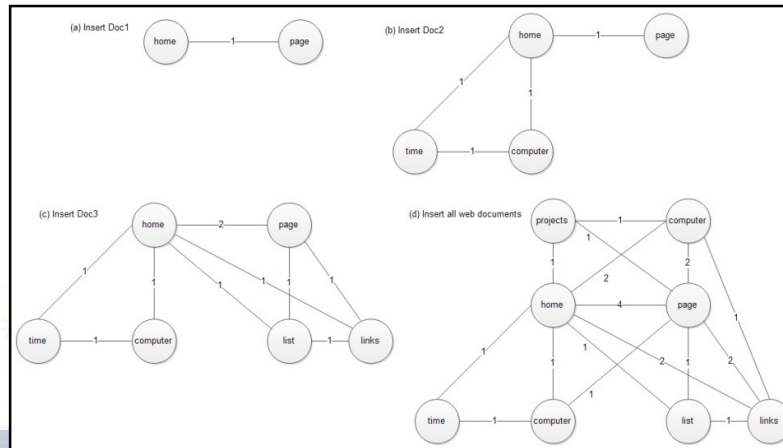
Tabel 5.1 Contoh Dokumen Web dan Keywords

ID Document	URL	Keywords
Doc1	http://www.cs.wisc.edu/~jglarson/jglarson.html	Home, Page
Doc2	http://www.cs.washington.edu/homes/chou/personal.html	Home, Computer, Time
Doc3	http://www.cs.cornell.edu/info/people/rvr/iajb/index.html	Home, List, Page, Links
Doc4	http://www.cs.cornell.edu/info/people/scytacki/links.html	Home, Page, Computer, Links
Doc5	http://www.cs.washington.edu/education/courses/557/95w/images.html	Home, Page, Computer, Project
Doc6	http://www.cs.washington.edu/homes/chou/dict/dicttop.html	Page, Computer
Doc7	http://www.cs.wisc.edu/~ammons/ammons.html	Page

Pada Doc1, terdapat *keywords* yaitu Home dan Page. Kedua *keywords* tersebut ditransformasikan ke dalam FCON, di mana Home dan Page merupakan dua *vertex* yang dihubungkan oleh *edge* berbobot 1. Gambar 5.2-(a) mengilustrasikan FCON setelah menambahkan *keywords* pada Doc1.

Pada Doc2, terdapat *keywords* yaitu Home, Computer, dan Time, di mana Home telah terdapat di dalam FCON sehingga tidak dibuat *vertex* baru untuk Home. Sedangkan Computer belum terdapat di dalam FCON sehingga dibuat *vertex* baru untuk Computer dan dihubungkan dengan *vertex* Home dengan bobot *edge* 1. Demikian pula dengan Time, dibuat *vertex* baru untuk Time yang dihubungkan dengan *vertex* Computer dan Home dengan bobot kedua *edges* yang sama yaitu 1, seperti yang diilustrasikan pada Gambar 5.2-(b).

Pada Doc3, terdapat *keywords* yaitu Home, List, Page, dan Links, di mana Home telah terdapat di dalam FCON sehingga tidak dibuat *vertex* baru untuk Home. Sedangkan List belum terdapat di dalam FCON sehingga dibuat *vertex* baru untuk List dan dihubungkan dengan *vertex* Home dengan bobot *edge* 1. Selain Home, Page juga telah terdapat di dalam FCON sehingga tidak dibuat *vertex* baru untuk Page. Namun, bobot dari *edge* $e(\text{Home}, \text{Page})$ ditambah sebesar 1 sehingga bobot pada *edge* tersebut menjadi 2. Kemudian dibuat pula *edge* baru yang menghubungkan antara *vertex* Page dengan List dan bobot dari *edge* baru tersebut diinisialisasi dengan 1. Selanjutnya, *vertex* baru kembali dibuat untuk Links dan *vertex* ini dihubungkan dengan *vertex* Home, List, Page yang dihubungkan dengan *edge* berbobot masing-masing 1. Gambar 5.2-(c) mengilustrasikan FCON setelah menambahkan *keywords* pada Doc3. Setelah semua dokumen web dimasukkan ke dalam FCON, hasil akhir dari pembentukan FCON dapat dilihat pada Gambar 5.2-(d).



Gambar 5.2 Ilustrasi Pembentukan FCON

5.4 Perhitungan Precision dan Recall Pada Aplikasi

Perhitungan *precision* dan *recall* dilakukan pada setiap proses *training* untuk masing-masing dokumen yang memperoleh rekomendasi metadata. Kedua *variable* tersebut merupakan tolak ukur guna menilai kualitas dari hasil rekomendasi metadata. Berdasarkan dari contoh hasil rekomendasi metadata pada Tabel 5.2 dan Persamaan 1 dan 2, diperoleh hasil perhitungan *precision* $Pr(\mu)$ dan *recall* $Re(\mu)$ sebagai berikut,

- a. <http://www.cs.wisc.edu/~dailey/saramail.html>

$$Pr(\mu) = \frac{|meta(n_1, \mu) \cap rec(n_1, \mu)|}{|rec(n_1, \mu)|} = \frac{3}{4} = 0.75$$

$$Re(\mu) = \frac{|meta(n_1, \mu) \cap rec(n_1, \mu)|}{|meta(n_1, \mu)|} = \frac{3}{6} = 0.5$$

- b. <http://www.cs.wisc.edu/~horn/hornfamily.html>

$$Pr(\mu) = \frac{|meta(n_2, \mu) \cap rec(n_2, \mu)|}{|rec(n_2, \mu)|} = \frac{0}{4} = 0$$

$$Re(\mu) = \frac{|meta(n_2, \mu) \cap rec(n_2, \mu)|}{|meta(n_2, \mu)|} = \frac{0}{7} = 0$$

- c. <http://www.cs.wisc.edu/~hcl/cs302/grading.html>

$$\Pr(\mu) = \frac{|meta(n_3, \mu) \cap rec(n_3, \mu)|}{|rec(n_3, \mu)|} = \frac{4}{4} = 1$$

$$\text{Re}(\mu) = \frac{|meta(n_3, \mu) \cap rec(n_3, \mu)|}{|meta(n_3, \mu)|} = \frac{4}{8} = 0.5$$

d. <http://www.cs.wisc.edu/~sastry/album.html>

$$\Pr(\mu) = \frac{|meta(n_4, \mu) \cap rec(n_4, \mu)|}{|rec(n_4, \mu)|} = \frac{4}{4} = 1$$

$$\text{Re}(\mu) = \frac{|meta(n_4, \mu) \cap rec(n_4, \mu)|}{|meta(n_4, \mu)|} = \frac{4}{4} = 1$$

Tabel 5.2 Contoh Hasil Rekomendasi Metadata

No.	URL	Keywords		Total Keywords		
		Old	Rec.	meta(n _i , μ)	rec(n _i , μ)	meta(n _i , μ) ∩ rec(n _i , μ)
1.	http://www.cs.wisc.edu/~dailey/saramail.html	edu; mail; address; cs; page; washington	edu; cs; page; home	6	4	3
2.	http://www.cs.wisc.edu/~horn/hornfamily.html	two; work; time; research; new; postscript; computer	edu; cs; page; home	7	4	0
3.	http://www.cs.wisc.edu/~hcl/cs302/grading.html	information; class; page; pages; home; cs; edu; computer	edu; cs; page; home	8	4	4
4.	http://www.cs.wisc.edu/~sastry/album.html	home; page; edu; cs	edu; cs; page; home	4	4	4

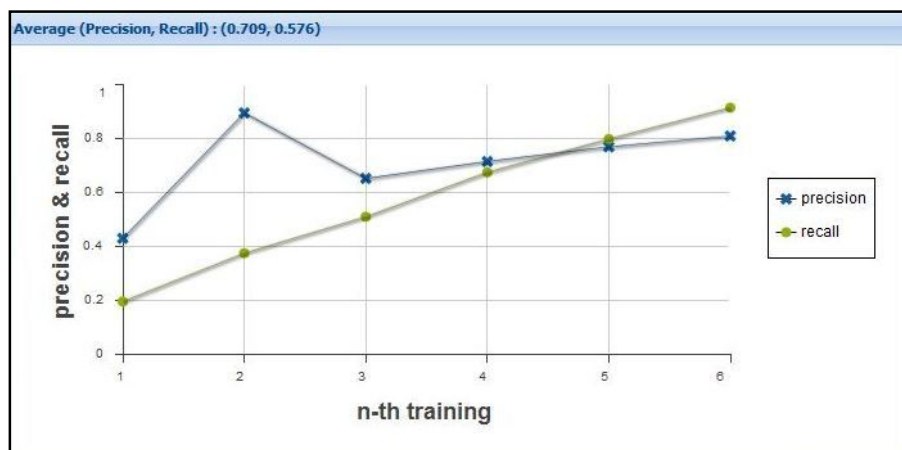
5.5 Hasil Uji Coba

Proses propagasi metadata yang dilakukan tanpa menggunakan metode *harvesting* menghasilkan rata-rata *precision* sebesar 0.709 dengan nilai maksimum yang dicapai adalah 1 dan rata-rata nilai maksimum sebesar 0.959. Sedangkan untuk rata-rata *recall* sebesar 0.576 dengan nilai minimum yang dicapai adalah 0.05263 dan rata-rata nilai minimum sebesar

0.386. Selain itu, keberhasilan merekomendasikan metadata mencapai angka 87.667%. Sedangkan hasil uji coba secara rinci dapat dilihat pada Tabel 5.1 dan Gambar 5.2.

Tabel 5.3 Hasil Pengujian 3026 Dokumen Web dengan FCON Tanpa Harvesting

No.	Parameter				Generate Metadata	Precision		Recall	
	Harvesting	Max. Poor Keyword	Max. Rec. Keyword	Min. Weight Link		Average	Max	Average	Min
1.	No	20	4	116	86%	0.42767	1	0.19358	0.05263
2.	No	20	4	125	42%	0.89237	1	0.3727	0.15789
3.	No	20	8	127	98%	0.64943	1	0.50765	0.21053
4.	No	20	12	141	100%	0.71248	0.91667	0.67246	0.42105
5.	No	20	16	159	100%	0.76693	0.9375	0.79595	0.63158
6.	No	20	20	178	100%	0.80701	0.9	0.9121	0.84211
Average					87.667%	0.709	0.959	0.576	0.386



Gambar 5.3 Hasil Pengujian 3026 Dokumen Web dengan FCON Tanpa Harvesting

Pada percobaan selanjutnya, proses propagasi metadata dilakukan menggunakan metode *harvesting*. Namun, pada *training* pertama nilai *precision* dan *recall* mencapai angka

yang sama dengan *training* pertama pada percobaan yang dilakukan tanpa menggunakan metode *harvesting*, yaitu dengan nilai *precision* 0.42767 dan *recall* 0.19358. Hal ini disebabkan karena perbandingan jumlah dokumen web dengan jumlah dokumen multimedia terlampaui jauh, yaitu antara 3026 dengan 8. Selain itu, jumlah relasi antar keduanya hanya terdapat 12 relasi. Oleh karena itu, guna menguji pengaruh *harvesting* pada dataset yang digunakan, dilakukan eliminasi terhadap dokumen web. Eliminasi tersebut menghasilkan 67 dokumen web tanpa mempengaruhi jumlah dokumen multimedia dan relasi keduanya.

Pada dataset yang baru, yaitu dengan 67 dokumen web, 8 dokumen multimedia, dan 12 relasi antara dokumen web dan dokumen multimedia, proses propagasi metadata yang dilakukan tanpa menggunakan metode *harvesting* menghasilkan rata-rata *precision* sebesar 0.779 dengan nilai maksimum yang dicapai adalah 1 dan rata-rata nilai maksimum sebesar 0.943. Sedangkan untuk rata-rata *recall* sebesar 0.705 dengan nilai minimum yang dicapai adalah 0.11111 dan rata-rata nilai minimum sebesar 0.489. Selain itu, keberhasilan merekomendasikan metadata mencapai angka 71.889%. Sedangkan hasil uji coba secara rinci dapat dilihat pada Tabel 5.2 dan Gambar 5.3.

Tabel 5.4 Hasil Pengujian 67 Dokumen Web dengan FCON Tanpa Harvesting

No.	Parameter				Generate Metadata	Precision		Recall	
	Harvesting	Max. Poor Keyword	Max. Rec. Keyword	Min. Weight Link		Average	Max	Average	Min
1.	No	20	4	2	97%	0.34583	1	0.3254	0.11111
2.	No	20	4	3	39%	0.82627	1	0.51898	0.23077
3.	No	20	4	3	3%	0.99153	1	0.56958	0.25
4.	No	20	8	3	100%	0.58898	0.875	0.65904	0.38462
5.	No	20	12	3	100%	0.69351	0.83333	0.78346	0.53333

Tabel 5.5 Hasil Pengujian 67 Dokumen Web dengan FCON Tanpa Harvesting (Lanjutan)

No.	Parameter				Generate Metadata	Precision		Recall	
	Harvesting	Max. Poor Keyword	Max. Rec. Keyword	Min. Weight Link		Average	Max	Average	Min
6.	No	20	12	4	51%	0.95763	1	0.799	0.57895
7.	No	20	16	4	100%	0.80603	0.875	0.86821	0.68421
8.	No	20	16	5	57%	0.96467	1	0.88717	0.78947
9.	No	20	20	5	100%	0.83452	0.9	0.93829	0.84211
Average					71.889%	0.779	0.943	0.705	0.489

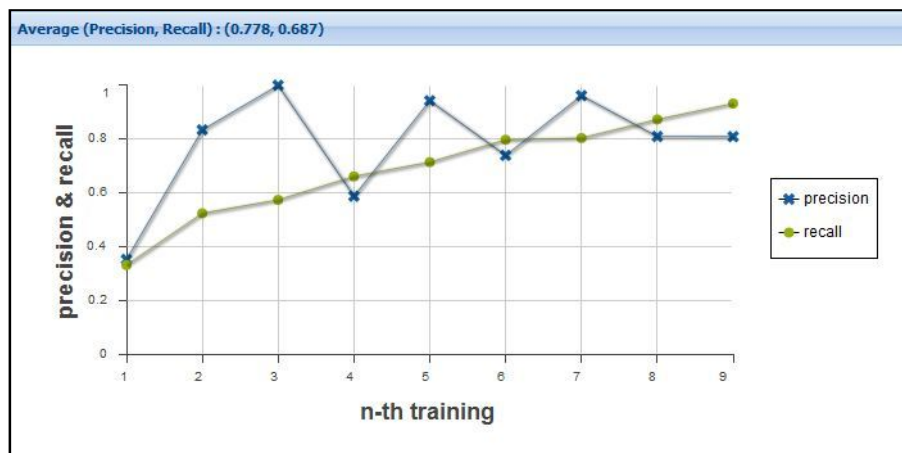


Gambar 5.4 Hasil Pengujian 67 Dokumen Web dengan FCON Tanpa Harvesting

Selanjutnya, proses uji coba dilakukan dengan menerapkan metode *harvesting*. Uji coba tersebut menghasilkan rata-rata *precision* sebesar 0.778 dengan nilai maksimum yang dicapai adalah 1 dan rata-rata nilai maksimum sebesar 0.93. Sedangkan untuk rata-rata *recall* sebesar 0.687 dengan nilai minimum yang dicapai adalah 0.11111 dan rata-rata nilai minimum sebesar 0.461. Selain itu, keberhasilan merekomendasikan metadata mencapai angka 71.222%. Sedangkan hasil uji coba secara rinci dapat dilihat pada Tabel 5.3 dan Gambar 5.4.

Tabel 5.6 Hasil Pengujian 67 Dokumen Web dengan FCON dan Harvesting

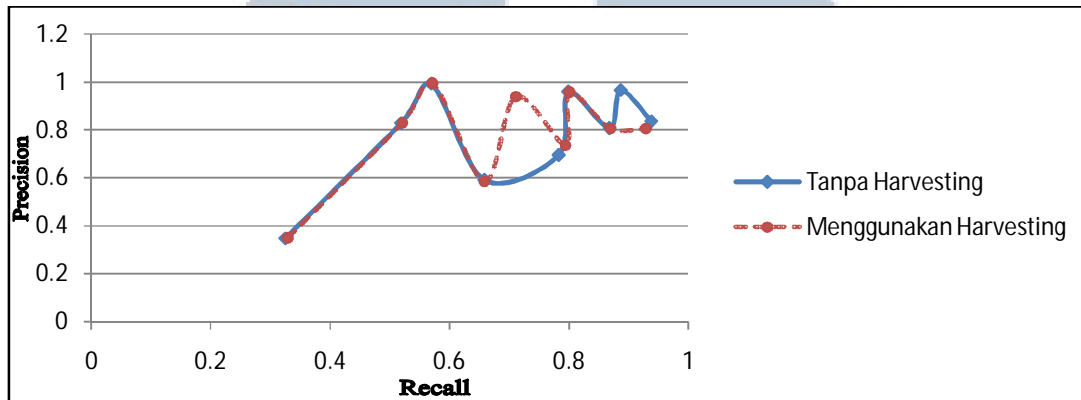
No.	Parameter				Generate Metadata	Precision		Recall	
	Harvesting	Max. Poor Keyword	Max. Rec. Keyword	Min. Weight Link		Average	Max	Average	Min
1.	Yes	20	4	2	97%	0.35	1	0.32956	0.11111
2.	Yes	20	4	3	42%	0.83051	1	0.52136	0.25
3.	Yes	20	4	3	2%	0.99576	1	0.57123	0.27273
4.	Yes	20	8	3	100%	0.58475	0.75	0.65831	0.35714
5.	Yes	20	8	3	49%	0.93856	1	0.71122	0.47059
6.	Yes	20	12	3	100%	0.73588	0.83333	0.79409	0.5625
7.	Yes	20	12	4	51%	0.95763	1	0.80053	0.57895
8.	Yes	20	16	4	100%	0.80603	0.9375	0.86932	0.70588
9.	Yes	20	20	5	100%	0.80556	0.85	0.92848	0.84211
Average					71.222%	0.778	0.93	0.687	0.461



Gambar 5.5 Hasil Pengujian 67 Dokumen Web dengan FCON dan Harvesting

Berdasarkan Tabel 5.4 dan Tabel 5.5, dibuatlah grafik guna menjelaskan secara *visual* tentang perbandingan hasil uji coba antara pembentukan FCON tanpa *harvesting* dengan

pembentukan FCON yang menggunakan *harvesting*. Sedangkan grafik tersebut dapat dilihat pada Gambar 5.6.



Gambar 5.6 Perbandingan Pembentukan FCON Tanpa dan Menggunakan Harvesting

Bukan hanya *precision* dan *recall* yang menjadi perhatian dalam proses uji coba, waktu yang diperlukan untuk proses propagasi metadata pun turut diukur. Pengukuran atau pencatatan waktu dilakukan terhadap proses uji coba yang pertama yaitu 3026 dokumen web tanpa *harvesting*. Hasilnya diperoleh rata-rata waktu yang diperlukan untuk proses propagasi metadata adalah 162.93 detik. Sedangkan hasil uji coba secara rinci dapat dilihat pada Gambar 5.5.



Gambar 5.7 Waktu Propagasi Metadata

5.6 Pembahasan

Berdasarkan Tabel 5.1 dan hasil uji coba yang dilakukan oleh Darmawan Fatriananda[3], diperoleh hasil analisis sebagai berikut,

- a. Nilai *precision* meningkat sebesar 16.07%, yaitu dari 0.6108 menjadi 0.709.
- b. Nilai *recall* menurun sebesar 27.11%, yaitu dari 0.79025 menjadi 0.576.
- c. Keberhasilan merekomendasikan metadata meningkat sebesar 75.33%, yaitu dari 50% menjadi 87.667%.
- d. Pada penelitian ini, nilai *precision* telah mencapai angka lebih besar dari nilai *recall*, yaitu $0.709 > 0.576$.

Berdasarkan Tabel 5.2 dan Tabel 5.3, diperoleh hasil analisis sebagai berikut,

- a. Pembentukan FCON tanpa metode *harvesting* menghasilkan *precision* sebesar 0.779. Sedangkan pembentukan FCON yang diawali dengan proses *harvesting* membuat nilai *precision* menurun sebesar 0.13%, yaitu menjadi 0.778.
- b. Pembentukan FCON tanpa metode *harvesting* menghasilkan *recall* sebesar 0.705. Sedangkan pembentukan FCON yang diawali dengan proses *harvesting* membuat nilai *recall* menurun sebesar 2.55%, yaitu menjadi 0.687.
- c. Pembentukan FCON tanpa metode *harvesting* mempunyai tingkat keberhasilan rekomendasi metadata sebesar 71.889%. Sedangkan pembentukan FCON yang diawali dengan proses *harvesting* membuat tingkat keberhasilan rekomendasi metadata menurun sebesar 0.93%, yaitu menjadi 71.222%.
- d. Pada pembentukan FCON, baik tanpa metode *harvesting* maupun menggunakan metode *harvesting*, keduanya menghasilkan nilai *precision* yang telah mencapai angka lebih besar dari nilai *recall*, yaitu masing-masing $0.779 > 0.705$ dan $0.778 > 0.687$.

Sedangkan waktu yang dibutuhkan untuk propagasi metadata pada penelitian ini yaitu 162.93 detik. Namun, angka tersebut tidak dapat dibandingkan dengan penelitian Darmawan Fatriananda[3] karena pada penelitiannya tidak melakukan analisa terhadap waktu eksekusi tersebut.



BAB VI

SIMPULAN DAN SARAN

6.1 Kesimpulan

Kesimpulan yang diperoleh dari penelitian ini:

- a. Metode *harvesting* mampu memberikan pengaruh positif terhadap proses pembentukan FCON karena dengan proses ini maka jumlah metadata yang menjadi *node* dalam *network* semakin bertambah. Selain itu, pengaruh positif dapat pula dilihat dari penurunan *recall* sebesar 2.55%, meskipun terjadi penurunan juga terhadap *precision* tetapi penurunan tersebut hanya mencapai 0.13%. Berdasarkan presentase ini, terlihat bahwa penurunan *recall* yang terjadi masih lebih besar dari pada penurunan *precision*.
- b. Metode FCON dapat digunakan sebagai *network* dalam melakukan pembentukan metadata secara otomatis dan menghasilkan nilai *precision* yang lebih besar dari nilai *recall*, yaitu $0.709 > 0.576$ dengan keberhasilan rekomendasi metadata sebesar 87.667%.
- c. Algoritma *quicksort* dapat digunakan sebagai pendukung proses propagasi metadata, di mana waktu yang dibutuhkan untuk melakukan propagasi terhadap 3026 dokumen web adalah 162.93 detik.
- d. Jika dibandingkan dengan penelitian Darmawan Fatriananda[3], maka pada penelitian ini nilai rata-rata *precision* meningkat sebesar 16.07% dan rata-rata *recall* menurun sebesar 27.11%. Di mana kedua *variable* tersebut telah memenuhi standar kualitas pembentukan metadata yaitu *precision* > *recall*. Selain itu, terjadi peningkatan pula terhadap angka keberhasilan rekomendasi metadata sebesar 75.33%.

6.2 Saran

Saran pengembangan penelitian ini:

- a. Guna memperoleh hasil analisa yang lebih akurat terhadap pengaruh metode *harvesting* dalam pembentukan FCON, diperlukan dataset dengan dokumen web yang berfokus pada pengelolaan dokumen multimedia seperti *.pdf dan *.doc.
- b. Menerapkan metode tertentu untuk menentukan bobot *edge* yang menjadi syarat kelayakan sebuah *vertex* untuk direkomendasikan, sehingga dapat diperoleh hasil optimal.

