

HASIL PENCARIAN DARI MESIN PENCARI YANG RELEVAN DENGAN *QUERY EXPANSION* MENGGUNAKAN ONTOLOGI

Maskur

Prodi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Malang

Jl Raya Tlogomas, No.246, Malang

E-mail: maskur.informatika@gmail.com

ABSTRAK

Mesin pencari seperti Google, Yahoo dan bing digunakan untuk mencari dan memperoleh data atau informasi yang tersebar di Internet. Hasil pencarian oleh mesin pencari kadang-kadang tidak sesuai yang diharapkan atau kurang relevan. Hal ini karena proses pencarian tidak mempertimbangkan konteks informasi atau konten yang dicari. Teknologi web semantik mampu memahami makna dari sebuah kata dan konsep serta hubungan keduanya. Teknologi ini menggunakan sumber data yang didefinisikan dalam ontologi berdasarkan *Resource Description Framework* (RDF). Penelitian ini akan diterapkan ke mesin pencari yang menggunakan web semantik untuk mencari istilah dalam proses Teknologi Informasi. Hasil pengolahan semantik akan dilakukan perluasan permintaan untuk meningkatkan kinerja pencarian. Berdasarkan hasil pengujian perbandingan nilai *precision* hasil pencarian dari mesin pencari, dapat disimpulkan 26 dari 30 kata kunci berhasil dilakukan *query expansion* dan dilakukan perhitungan nilai *precision*. Perbandingan masing-masing nilai *precision@5*, *precision@10*, dan *precision@20* antara hasil pencarian biasa dengan hasil pencarian *query expansion* memperlihatkan dari ke 26 kata kunci tidak ada yang mengalami penurunan nilai *precision* dan hampir sebagian besar mengalami peningkatannya nilai *precision*.

Kata kunci : *Mesin Pencari, Semantik Web, Ontologi, Query Expansion, RDF.*

ABSTRACT

Search engines like Google, Yahoo and bing used to search for and obtain data or information that is spread across the Internet. Search results by search engines sometimes not as expected or less relevant. This is because the search process does not consider the context or content of information being sought. Semantic web technologies are able to understand the meaning of a word and concept as well as their relationship. This technology uses a data source defined in an ontology based on the Resource Description Framework (RDF). This research will be applied to the search engine using semantic web to search for terms in the process of Information Technology. The results of semantic processing will be done in expanding the demand to improve search performance. Based on the results of comparative testing precision value of search engine results, it can be concluded 26 of the 30 successful keyword query expansion and calculation precision value. Comparison of each precision value @ 5 @ 10 precision, and precision @ 20 between the regular search results with search query results showing expansion of the 26 keywords no impaired precision and almost most experienced peningkatannya precision value

Keywords: *Search Engines, Semantic Web, Ontology, Query Expansion, RDF.*

PENDAHULUAN

Perkembangan internet yang meningkat dengan cepat berpengaruh pada semakin bertambahnya jumlah pengguna internet di dunia. Ini tentunya tak lepas dari penyebaran data dan informasi melalui internet tanpa ada batasan waktu dan tempat. Semua orang di dunia satu sama lain dapat saling berbagi informasi dan data serta mempublikasikan sumber daya yang mereka miliki melalui internet. Untuk mencari dan mendapatkan informasi yang beredar di internet, bisa menggunakan beberapa mesin pencari yang juga sudah banyak tersedia, seperti *Google*, *Yahoo* dan *Bing*.

Sebagian besar mesin – mesin pencari tersebut melakukan pencarian berdasarkan kata kunci yang dimasukkan oleh pengguna, dimana kata kunci tersebut terkait dengan informasi atau data yang mereka butuhkan. Kemudian, mesin pencari berusaha mencocokkan kata kunci tersebut dengan isi dokumen yang tersebar di internet. Terkadang hasil pencarian dengan menggunakan metode ini banyak memberikan hasil yang tidak sesuai atau tidak relevan dengan konteks informasi yang diinginkan. Ini terjadi karena pada proses pencariannya tidak mempertimbangkan konteks dari informasi atau konten yang dicari, melainkan hanya mencocokkan kata – kata yang ada pada suatu dokumen di internet dengan kata kunci yang dimasukkan. Dengan kata lain hasil pencarian tersebut memiliki ketepatan yang sangat rendah. Namun ada pula dokumen yang isinya sebenarnya diperlukan, namun ternyata tidak ditampilkan oleh mesin pencari atau dengan kata lain memiliki tingkat *recall* yang rendah [1].

Salah satu pendekatan yang dapat digunakan untuk menjawab beberapa permasalahan di atas adalah dengan konsep pencarian cerdas. Yaitu melakukan pencarian data atau informasi melalui mesin pencari yang menggunakan semantik web dari sumber data yang didefinisikan dalam

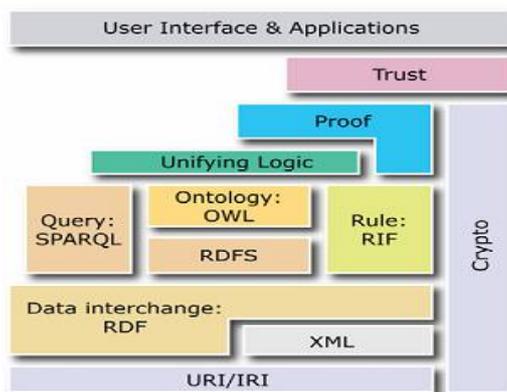
suatu ontologi. Semantik web mampu memahami tidak hanya makna dari suatu kata dan konsep, namun juga hubungan logis diantara keduanya, sehingga mesin pencari nantinya dapat menghasilkan suatu informasi yang sesuai dengan keinginan pengguna [2].

Pada penelitian ini penulis akan menerapkan konsep semantik web ke dalam suatu mesin pencari yang ditujukan untuk pencarian informasi dan istilah – istilah dalam mata kuliah dibidang teknik informatika guna menunjang mahasiswa dalam proses pembelajaran. Dengan harapan hasil pencariannya akan lebih relevan dibandingkan dengan melakukan pencarian menggunakan mesin pencari biasa.

METODE

Web Semantik

Sebuah web semantik terbentuk terdiri dari berbagai macam komponen-komponen yang saling berhubungan satu sama lain. Adapun komponen-komponen tersebut dapat digambarkan pada lapisan-lapisan yang ada pada layer web semantik Gambar 1 [3].



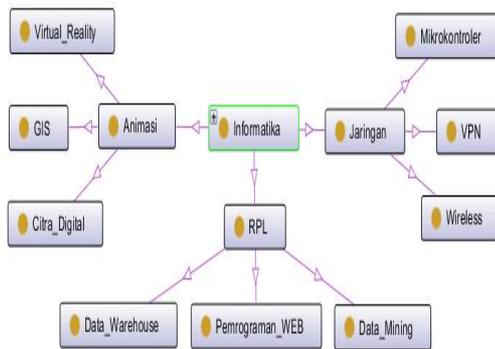
Gambar 1 Semantic Web Layer

Ontologi

Ontologi pada penelitian ini digunakan untuk mengetahui hubungan antar kata dalam istilah yang ada pada mata kuliah di jurusan teknik informatika. Struktur ontologi nantinya diharapkan dapat dimengerti dengan baik oleh komputer yaitu dengan cara

mendefinisikan sebuah informasi yang mempunyai hubungan antara subjek, predikat, dan objek.

Selain itu tujuan dari perancangan ontologi ini adalah dapat digunakan untuk proses pencarian, sehingga kata kunci yang dicari tidak hanya kata kunci yang dimasukkan oleh pengguna melainkan juga kata kunci tersebut memiliki sinonim dan keterkaitan atau hubungan dengan kata-kata yang tersusun secara semantik di dalam ontologi [3].



Gambar 2 Ontologi Mata Kuliah Teknik Informatika

Setiap ontologi pada semantik web berisi tentang kumpulan pengetahuan pada domain tertentu yang terdiri dari konsep (*class*) dan relasi (*properties*). Pada ontologi Gambar 1 terdapat 3 *class* utama, yaitu: Desain, RPL, dan Jaringan. Pada saat mendefinisikan sebuah atau beberapa *class*, maka harus mendefinisikan atau menjelaskan struktur internal dari konsep yang akan dibangun dari beberapa *class* tersebut. Setelah *class* diciptakan langkah selanjutnya adalah mendefinisikan *properties*, dari sini sebuah *class* akan mempunyai nilai tambah dalam hal ini adalah informasi.

Resource Description Framework (RDF)

Dokumen RDF yang mendefinisikan daftar istilah-istilah yang ada pada masing-masing subclass. Sebagian isi dari dokumen RDF yang telah diterjemahkan berdasarkan ontologi tersebut dapat dilihat pada Gambar 2 [4].

```
<?xml version="1.0"?>
<rdf:RDF xmlns="http://www.umm.ac.id/"
  xml:base="http://www.umm.ac.id/"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-
  schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:www="http://www.umm.ac.id/"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema
  #"
  xmlns:rdf="http://www.w3.org/1999/02/22-
  rdf-syntax-ns#">
  <owl:Ontology
    rdf:about="http://www.umm.ac.id/">
  <owl:Class rdf:about="&www;Informatika"/>
  <!-- http://www.umm.ac.id/Desain -->
  <owl:Class rdf:about="&www;Desain">
    <rdfs:subClassOf
      rdf:resource="&www;Informatika"/>
    <rdfs:adalah>Desain</rdfs:adalah>
  </owl:Class>
  <!-- http://www.umm.ac.id/GIS -->
  <owl:Class rdf:about="&www;GIS">
    <rdfs:subClassOf
      rdf:resource="&www;Desain"/>
    <rdfs:adalah>GIS</rdfs:adalah>
  </owl:Class>
```

Gambar 3 Isi Dokumen RDF

Perancangan Eksekusi Query

Tujuan dari perancangan eksekusi *query* ini adalah agar memudahkan sistem dalam memaknai inti dari sebuah kata kunci. Terkadang kata kunci yang diberikan oleh pengguna nantinya ada yang dapat langsung mengarah pada inti dari makna yang ada di dokumen RDF ada pula yang tidak.

Sebelum kata kunci dari pengguna dieksekusi dalam proses semantik, kata kunci tersebut akan diolah terlebih dahulu yaitu dengan memotong atau memisahkannya menjadi beberapa bagian. Hal ini dilakukan untuk memudahkan proses eksekusi *query* terhadap dokumen RDF agar hasil yang di dapat nantinya lebih maksimal [5].

SPARQL

SPARQL (*Simple Protocol And RDF Query Language*) merupakan bahasa *query* untuk RDF/OWL. SPARQL menyediakan fasilitas untuk mengekstrak informasi dalam bentuk URI, blank node dan literal, mengekstrak subgraf RDF, dan membangun graf RDF baru berdasar pada informasi dari graf yang dilakukan *query*. *Query* SPARQL didasarkan pada pencocokan pola graf. Pola graf yang paling sederhana adalah

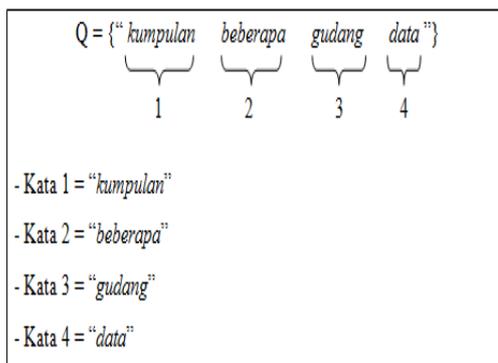
triple pattern yang mirip dengan RDF *triple*, hanya saja pola pada *query* dimungkinkan pemberian nama diluar terminology RDF pada posisi subyek, predikat dan obyek [6].

JENA

JENA – Semantic Web Toolkit merupakan API (*Application Programming Interface*) dan *software tool* untuk membangun aplikasi *Semantic Web* yang berbasis Java. JENA menyediakan kumpulan *tools* dan *libraries* untuk membantu pengembang dalam membangun dan memanipulasi RDF, RDFS, RDFa, OWL dan SPARQL

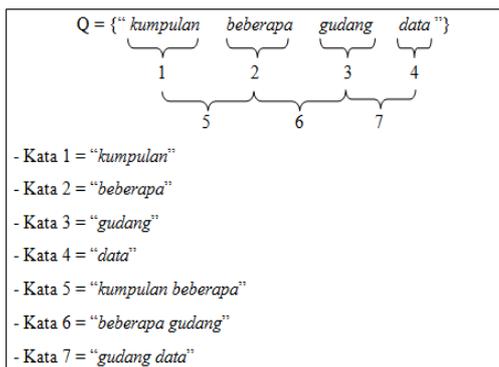
Proses *Tokenizing*

Pengguna memasukkan kata kunci $Q = \{\text{"kumpulan beberapa gudang data"}\}$. Dari kata kunci yang merupakan sebuah kalimat tersebut akan dipotong atau dipisah menjadi perkata [8].



Gambar 4 Pemotongan Kalimat (Satu Kata)

Pemotongan tersebut mengubah pola kalimat menjadi bentuk perkata yang masing-masing hanya terdiri dari satu kata. Selain itu sistem juga akan melakukan pemotongan menjadi perkata yang masing-masing terdiri dari dua kata. Hal ini dilakukan untuk memaksimalkan proses pencarian di dokumen RDF, karena suatu inti makna belum tentu terdiri dari satu kata saja tetapi dua kata. Untuk proses pemotongannya tetap memanfaatkan proses pemotongan yang pertama, untuk lebih jelasnya dapat dilihat pada Gambar 4.



Gambar 5 Pemotongan Kalimat (Dua Kata)

Setelah dua proses pemotongan dilakukan maka akan dihasilkan beberapa kata dimana selanjutnya sistem akan melakukan proses semantik yaitu proses dimana sistem akan melakukan pencarian istilah ke dalam dokumen RDF menggunakan *query* SPARQL berdasarkan kata kunci tersebut [7].

Nilai *Precision*

Precision dapat diartikan sebagai kepersisan atau kecocokan (antara permintaan informasi dengan jawaban terhadap permintaan itu). Jika seseorang mencari informasi di sebuah sistem, dan sistem menawarkan beberapa dokumen, maka kepersisan ini sebenarnya juga adalah relevansi, artinya seberapa persis atau cocok dokumen tersebut untuk keperluan pencari informasi, bergantung pada seberapa relevan dokumen tersebut bagi si pencari [8]. Persamaan 1 digunakan untuk menghitung nilai *precision*.

$$precision = \frac{\text{dokumen relevan}}{\text{dokumendit emukan}} \quad (1)$$

HASIL DAN PEMBAHASAN

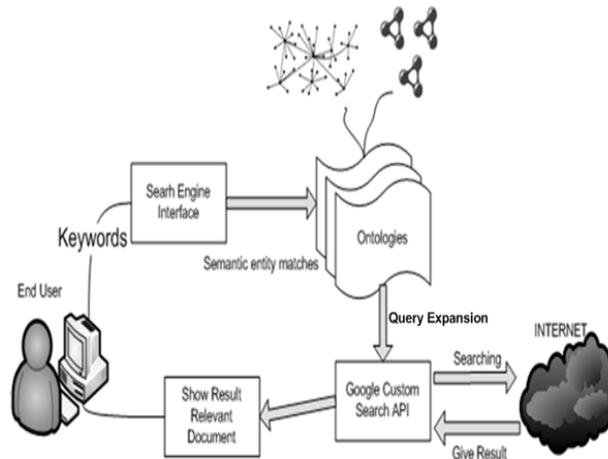
Arsitektur sistem ini memberikan gambaran ketika pengguna mengkases mesin pencari, didalamnya terdapat alur proses dari komponen atau perangkat yang saling terhubung satu sama lain.

Dari mesin pencari menggunakan semantik web yang telah dibangun, kemudian dilakukan pengujian dengan memasukkan 30 kata kunci yang berbeda, dengan rincian yaitu 10 kata kunci dibidang Rekayasa Perangkat

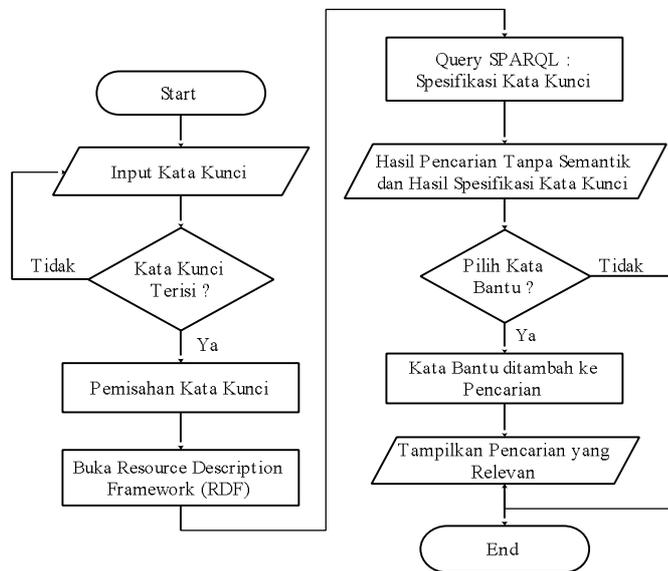
Lunak (RPL), 10 kata kunci dibidang Jaringan Komputer, dan 10 kata kunci dibidang Animasi Grafis.

Dimana alur proses tersebut tidak terlihat oleh pengguna, seperti proses semantik yaitu pencocokan kata kunci dengan istilah yang terdaftar di ontologi. Yang selanjutnya hasil pencocokan tersebut akan dilakukan *query expansion*. *Query expansion* adalah proses mereformulasikan kembali *query* awal

dengan melakukan penambahan beberapa kata pada *query* untuk meningkatkan performa dalam proses *information retrieval*. Dalam konteks *web search engine*, ini termasuk evaluasi input *user* dan memperluas *query* pencarian untuk mendapatkan dokumen yang cocok atau relevan [9]. Arsitektur sistem dari mesin pencari yang menggunakan semantik web ini dapat dilihat pada Gambar 6.



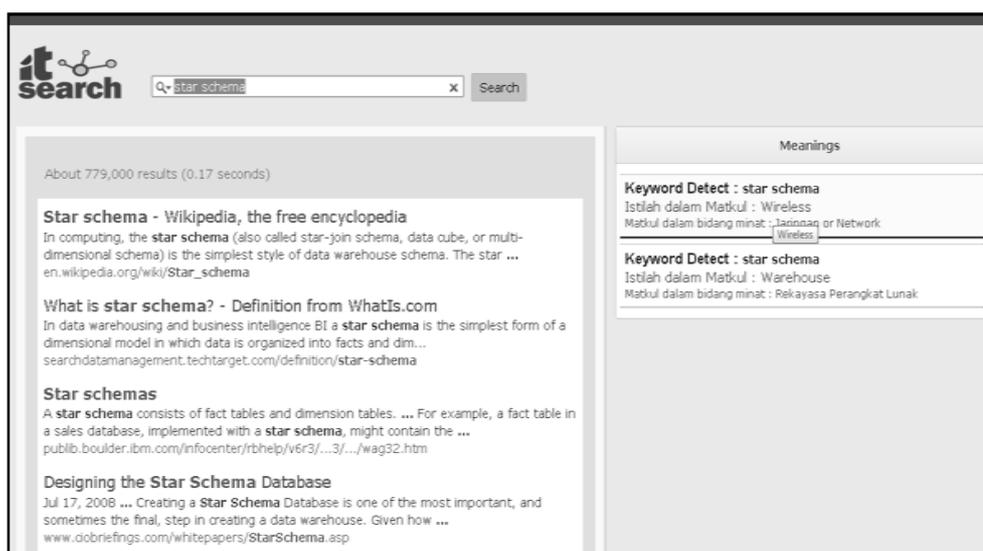
Gambar 5 Arsitektur Sistem



Gambar 6 Diagram Alir Mesin Pencari

1. Pertama, pengguna memasukkan kata kunci pencarian. Kemudian akan dicek apakah kata kunci sudah terisi dengan benar atau belum.
2. Kedua, Jika kata kunci telah terisi dengan benar, langkah selanjutnya adalah memotong kata kunci menjadi bentuk kata jika kata kunci yang

3. Ketiga, sistem akan menghubungkan dengan ontologi yaitu dengan membuka dokumen RDF. Dimana didalam file RDF ini berisi domain pengetahuan yang saling berelasi.
4. Keempat, Untuk memanggil kata kunci yang memiliki keterkaitan dengan kata yang ada didalam file RDF, diperlukan *query* SPARQL untuk menjalankannya.
5. Kelima, Sistem akan menampilkan hasil pencarian tanpa menggunakan semantik, disertai dengan spesifikasi kata kunci hasil dari *query* SPARQL terhadap file RDF.
6. Keenam, Jika pengguna memilih kata kunci yang lebih spesifik tersebut, maka sistem akan melakukan proses pencarian lagi dan menampilkan hasil pencarian yang lebih relevan.



Gambar 7 Hasil pencarian biasa.



Gambar 8 Hasil pencarian semantik dan *query expansion*

Nilai Precision

Adapun nilai *precision* yang dihitung dibagi menjadi 3 bagian berdasarkan peringkat hasil pencarian, yaitu:

1. *Precision@5*, nilai *precision* dihitung berdasarkan 5 teratas hasil pencarian. Sehingga rumusnya menjadi.

$$precision @ 5 = \frac{dokumen\ relevan}{5} \times 100\%$$

2. *Precision@10*, nilai *precision* dihitung berdasarkan 10 teratas hasil pencarian. Sehingga rumusnya menjadi:

$$precision @ 10 = \frac{dokumen\ relevan}{10} \times 100\%$$

3. *Precision@20*, nilai *precision* dihitung berdasarkan 20 teratas hasil pencarian. Sehingga rumusnya menjadi:

$$precision @ 20 = \frac{dokumen\ relevan}{20} \times 100\%$$

Tabel 1 Hasil Pengujian Nilai Precision

Daftar Pencarian	<i>Precision@5</i>		<i>Precision@10</i>		<i>Precision@20</i>	
	<i>P</i>	<i>QE</i>	<i>P</i>	<i>QE</i>	<i>P</i>	<i>QE</i>
1 jquery	100%	100%	95%	100%	95%	100%
2 klasterisasi	80%	80%	50%	90%	60%	95%
3 histogram	20%	100%	20%	100%	25%	95%
4 routing	40%	40%	40%	50%	25%	35%
5 stereoscopic	80%	100%	90%	100%	80%	90%
6 Html5	100%	100%	100%	100%	100%	100%
7 model data raster	100%	-	100%	-	95%	-
8 remote sensing	100%	100%	90%	100%	90%	95%
9 pengertian olap	60%	60%	60%	60%	80%	85%
10 contoh AVR	40%	60%	40%	80%	60%	85%
11 tutorial json	80%	80%	50%	90%	60%	95%
12 proses OLTP	40%	-	40%	-	25%	-
13 bahasa assembly	60%	60%	60%	80%	80%	85%
14 definisi bluetooth	100%	100%	100%	100%	100%	100%
15 Sistem informasi peta	80%	80%	50%	80%	60%	80%
16 What tunneling means	60%	60%	60%	60%	80%	85%
17 Kepanjangan dari vrml	40%	-	40%	-	60%	-
18 Penjelasan tentang IPsec	80%	-	80%	-	80%	-
19 istilah penggalian informasi	40%	80%	60%	85%	60%	90%
20 adhoc tanpa kabel	100%	100%	100%	100%	100%	100%
21 algoritma naive bayes	100%	100%	80%	100%	80%	100%
22 steganografi	80%	100%	80%	100%	80%	100%
23 codeigniter	100%	100%	100%	100%	100%	100%
24 Komunikasi serial sinkron	60%	90%	50%	90%	60%	90%
25 Data spasial	60%	80%	60%	80%	60%	90%
26 Kepanjangan ETL	40%	80%	60%	85%	60%	90%
27 wimax	100%	100%	100	100%	100	100%
28 image enhancement	60%	60%	60%	60%	80%	85%
29 GPRS	100%	100%	100%	100%	100%	100%
30 Penjelasan georeferensi	80%	100%	90%	100%	80%	90%

Dari Tabel 1 dapat dilihat bahwa terdapat 4 kata kunci yang tidak dapat dilakukan *query expansion* dan dihitung nilai *precision*-nya. Ini terjadi karena istilah yang ada pada 1 kata tersebut tidak terdaftar di ontologi.

Namun 26 kata kunci lainnya berhasil dilakukan *query expansion* dan dilakukan perhitungan nilai *precision*-nya. Serta kata kunci tidak ada yang mengalami penurunan nilai *precision*, namun justru hampir sebagian besar mengalami peningkatannya nilai *precision*. Sehingga hasil pencarian yang dilakukan

mesin pencari yang menggunakan semantik web ini mendekati relevan.

SIMPULAN

Berikut beberapa kesimpulan yang dapat diambil dari hasil penelitian ini antara lain :

1) Semantik web dapat diimplementasikan pada sistem mesin

pencari untuk melakukan proses navigasi dan pencarian suatu informasi di internet pada domain pengetahuan tertentu guna mendapatkan hasil yang relevan.

- 2) Teknologi semantik web dengan pendekatan ontologi merupakan salah satu dasar struktur informasi yang digunakan untuk melakukan proses pencarian informasi yang relevan.
- 3) Berdasarkan pengujian perbandingan nilai *precision* hasil pencarian dari mesin pencari, disimpulkan 26 dari 30 kata kunci berhasil dilakukan *query expansion* dan dilakukan perhitungan nilai *precision*. Perbandingan masing-masing nilai *precision@5*, *precision@10*, dan *precision@20* antara hasil pencarian biasa (*plain*) dengan hasil pencarian *query expansion* (QE) memperlihatkan dari ke 26 kata kunci tidak ada yang mengalami penurunan nilai *precision*.
- 4) Berdasarkan pengujian jumlah hasil pencarian dari mesin pencari, bahwa tingkat keberhasilan berkurangnya jumlah hasil pencarian dari mesin pencari yang menggunakan semantik web ini mencapai nilai 80%. Sehingga hasil pencarian dari mesin pencari yang menggunakan semantik web ini mendekati relevan.
- 5) Kemampuan mesin pencari untuk dapat meningkatkan hasil pencarian yang relevan tergantung dari banyak tidaknya istilah mata kuliah yang didaftarkan dalam ontologi.

SARAN

Beberapa saran yang dapat diberikan dari penulis sebagai hasil penelitian ini antara lain:

- 1) Perlu adanya penambahan istilah-istilah dari masing-masing mata kuliah dengan tujuan lebih memperkaya kosa kata dan pengetahuan dari ontologi.
- 2) Ditambahkannya fungsi untuk dapat menambahkan istilah pada ontologi secara otomatis melalui sistem mesin pencari.
- 3) Dibutuhkannya sistem basis data untuk menampung hasil pencarian terlebih dahulu dari *Google Custom Search Engine*, agar nantinya dapat dilakukan proses pencarian berdasarkan konteks (*context based*

search) guna meningkatkan nilai *precision*.

- 4) Untuk mendapatkan hasil pencarian yang lebih relevan dapat dikolaborasikan dengan metode *indexing* terhadap hasil pencarian, salah satunya dengan teknik pembobotan.

DAFTAR PUSTAKA

- [1] Toba, Hapnes, “*Pencarian Cerdas dengan Penggunaan Semantic Web*” Universitas Kristen Maranatha Bandung, Indonesia, 2005
- [2] R.D.Agushinta., H.P Permatasari, S. Soeparlan, “*Aplikasi Pencarian Karya Ilmiah Berbasis Web Semantik*” Universitas Gunadarma, Indonesia, 2005
- [3] D.O. Siahaan, “*Graphical Notations For Semantic Web Language*” Institut Teknik Sepuluh November (ITS), Indonesia, 2006
- [4] T. Gruber. “*Towards Principles for the Design of Ontologies Used for Knowledge Sharing*” Int. Journal of Human-Computer Studies, 43:907-928, 1995.
- [5] I.W.S. Wicaksana. “*Ontology: Bahasa dan Tools Protégé*”, Universitas Gunadarma, Indonesia, 2006.
- [6] Basil El, et al. “*SPARQL Query Verbalization for Explaining Semantic Search Engine Queries*”, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, 2014.
- [7] J.J. Carroll, et al. *Jena: Implementing the Semantic Web Recommendations*. HP Laboratories Bristol, Inggris, 2003.
- [8] A. Wibowo. “*Pengujian Kerelevanan Sistem Temu Kembali Informasi*”, Politeknik Negeri Batam, 2011.
- [9] Qiu, et al, HP. “*Concept Based Query Expansion*”, Swiss Federal Institute of Technology Zurich. Switz, 1997