

PENCARIAN CERDAS DENGAN PENGGUNAAN SEMANTIC WEB

Hapnes Toba

Program Studi Diploma III Teknologi Informasi Fakultas Teknologi Informasi
Universitas Kristen Maranatha Bandung
Jl. Suria Sumantri No. 65 Bandung 40164

Abstrak

Pada artikel ini akan dibahas studi literatur mengenai teknologi pencarian data dengan menggunakan konsep pencarian cerdas. Perkembangan data di Internet ataupun sumber data elektronik tertentu sangat sulit diikuti, setiap saat informasi pada suatu web site bisa berubah isinya, baik secara dinamik maupun statik. Salah satu pendekatan yang dapat digunakan untuk mengatasi masalah di atas adalah dengan melakukan pencarian data melalui mesin pencarian yang berbasis pada semantik dari sumber data yang didefinisikan dalam ontology. Teknologi Internet yang digunakan dalam ontology XML (*eXtensible Markup Language*) dan RDF (*Resource Description Framework*). Keunggulan teknologi XML dan RDF adalah adanya jaminan untuk interoperabilitas antara sumber data yang tersebar di berbagai tempat dengan platform yang sama ataupun berbeda. Hasil dari penelitian literatur ini ditindaklanjuti dengan perancangan dan pembuatan suatu mesin pencarian cerdas yang ditujukan untuk pencarian data yang menunjang mahasiswa dalam pembuatan kerja praktek atau tugas akhir. Pencarian tersebut akan dapat dilakukan di Internet ataupun suatu sumber data internal tertentu, seperti misalnya database perpustakaan atau pengelolaan dokumen elektronik lainnya.

Kata kunci: pencarian cerdas, ontology, XML, RDF, web semantik, dokumen elektronik.

1. Pendahuluan

Perkembangan data dan informasi pada Internet menuntut setiap pengguna untuk dapat menyaring informasi yang benar-benar diperlukannya. Latar belakang studi literatur ini adalah adanya kesulitan untuk mencari informasi yang diperlukan dari sumber data elektronik, baik pada Internet ataupun sistem pengarsipan tertentu secara tepat. Setiap saat informasi pada suatu *web site* atau sistem pengarsipan bisa berubah isinya, baik secara dinamik maupun statik. Permasalahan timbul ketika komputer yang digunakan untuk mencari informasi tidak dapat memfilter hasil pencarian. Pertanyaan yang dapat diajukan disini adalah: apakah ada cara yang efisien untuk mencari data di Internet, sehingga para pencari data, bahkan pada tingkat pemula pun, mampu mendapatkan informasi yang diinginkannya, tanpa harus melakukan seleksi yang rumit. Salah satu pendekatan yang dapat digunakan untuk menjawab pertanyaan di atas adalah dengan melakukan pencarian data melalui mesin pencarian yang berbasis pada semantik dari sumber data yang didefinisikan dalam *ontology*.

Pada akhir artikel akan dituliskan suatu skenario aplikasi untuk melakukan pencarian cerdas yang ditujukan untuk pencarian data guna menunjang mahasiswa dalam pembuatan kerja praktek atau tugas akhir. Pencarian tersebut akan dapat dilakukan di Internet ataupun suatu sumber data internal tertentu, seperti misalnya database perpustakaan atau pengelolaan dokumen elektronik lainnya.

2. Mesin Pencarian Data

Dalam paragraf ini akan dibahas berbagai permasalahan yang ada pada mesin pencarian data di Internet, dan kemungkinan untuk mengatasinya dengan mesin pencarian cerdas.

2.1 Precision dan Recall

Selama ini jika seseorang mencari data di Internet, akan disugahi hasil berbagai jenis informasi yang mungkin saja tidak diperlukan atau dengan kata lain pencarian tersebut memiliki ketepatan (*precision*) yang sangat rendah. Bisa juga terjadi bahwa ada dokumen yang isinya sebenarnya diperlukan, namun ternyata tidak ditampilkan oleh mesin pencari atau dengan kata lain tingkat *recall* rendah. Sebagai contoh jika seseorang ingin mencari data tentang "web" melalui mesin pencarian Lycos, Infoseek, Yahoo, Excite, AOL, MSN Search, HotBot, LookSmart, Snap, WebCrawler, AltaVista atau Google, maka akan muncul segala informasi tentang "web", dan disinilah harus pandai-pandai menyeleksi apa yang sebenarnya diperlukan.

2.2 Pencarian Cerdas

Semua mesin pencarian yang disebutkan di atas, menggunakan konsep *URL indexing* dan rangkuman terhadap suatu web-site tertentu. Ada berbagai masalah yang mungkin muncul (1), antara lain: index URL yang ditawarkan mungkin sudah tidak lagi aktif, mesin URL sedang *down*, dokumen yang direferensikan mungkin saja sudah dihapus atau dipindahkan, ataupun mungkin saja waktu tunggu yang terlalu lama untuk mengakses

dokumen (*request timeout*). Bahkan mereka yang sudah mahir menggunakan mesin pencarian pun, terkadang masih kesulitan untuk menyeleksi informasi mana yang ditawarkan oleh mesin pencarian yang layak untuk diakses.

Sebagaimana telah disebutkan pada paragraf 1, diperlukan suatu mekanisme yang tepat dalam pencarian data di Internet. Konsep yang dibutuhkan adalah pencarian cerdas, yaitu komputer dapat menyeleksi secara otomatis informasi mana yang sebenarnya diperlukan oleh pencari data.

3. Ontologi dan Software Agent

Pencarian cerdas memerlukan suatu mekanisme untuk melakukan seleksi informasi. Salah satu pendekatannya adalah dengan menggunakan keseragaman makna antara informasi yang dicari dan apa yang akan ditawarkan oleh mesin pencarian (2).

Keseragaman makna (*semantic*) tentang sesuatu term pencarian tersebut dapat diimplementasikan dalam suatu sumber data yang disebut ontologi. Ontologi adalah sumber data yang berisi representasi pengetahuan tentang suatu domain pada dunia riil (3). Lebih jauh lagi ontologi ini harus dilengkapi dengan penggunaan terminologi, anotasi dan rekomendasi (*inference rules*) untuk suatu domain tertentu. Dengan penggunaan keempat komponen semantik tersebut maka diharapkan setiap data yang ada dapat “mendeskripsikan” dirinya sendiri, sehingga mesin pencari data “mengerti” isi data tersebut.

Selain menggunakan sumber data yang dapat mendeskripsikan dirinya sendiri, untuk membuat suatu mesin pencarian yang cerdas, diperlukan konsep juga “agent”. Software agent ini akan bertindak sebagai mesin pencari data yang sesungguhnya. Agent akan “menentukan” informasi mana yang diperlukan oleh pencari data melalui penggunaan ontologi yang telah didefinisikan sebelumnya.

4. Hasil Penelitian

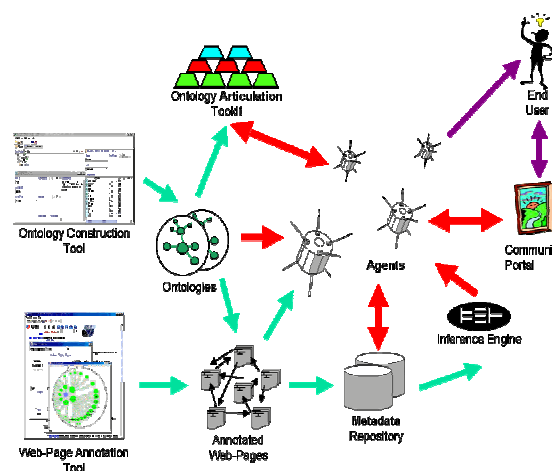
Ditinjau dari sudut teknologi, semantic web seperti yang telah dituliskan pada paragraf 3, dapat dilakukan dengan menggunakan XML dan RDF. XML adalah penerus HTML yang selama digunakan untuk pembuatan dokumen web di Internet. XML secara lebih luas mengijinkan pembuat dokumen untuk menentukan tag-tag yang akan digunakan dalam dokumennya. Pembuat dokumen dengan XML dapat mendeskripsikan dokumennya melalui tag-tag yang digunakannya. XML memungkinkan pencarian secara langsung dalam teks-teks yang ditampilkan dalam web-site tertentu secara terstruktur. Di sisi lain, RDF (4) memungkinkan pendefinisian tag-tag secara lebih terstruktur sebagaimana bahasa manusia dengan adanya subyek, predikat dan obyek. RDF membuat

suatu dokumen XML secara lebih terstruktur, melalui pendekatan dari bahasa pemrograman, yaitu: obyek (*objects / things*), properti (*properties*) dan nilai (*values*). XML dan RDF akan dihubungkan melalui URI (*Uniform Resource Identifier*). Penggunaan XML, RDF dan URI akan disatukan dalam *software agent* yang melakukan pencarian. Contoh penggunaan XML dan RDF adalah di bawah ini:

```
XML
<dokumen>
  <server
    href="http://itmaranatha.org">
  research </server> menyimpan
  <tugas_akhir> SemanticWeb
  </tugas_akhir>
</dokumen>

RDF
<http://itmaranatha.org>
<http://itmaranatha.org/research>
<http://itmaranatha.org/research/semanticweb.pdf>
```

Penggunaan XML dan RDF menguntungkan penggunaan konsep *agent*, bahwa informasi akan disaring oleh si *agent*, sehingga *agent* tersebut dapat menentukan apakah suatu dokumen merupakan sesuatu yang diperlukan. Sebagai suatu ringkasan, semua penjelasan di atas dapat digambarkan pada Figur 1:

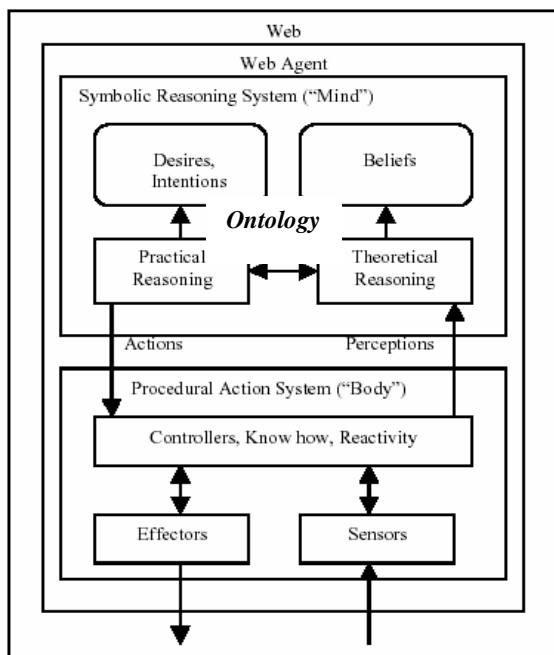


Figur 1. Semantic Web, Ontology dan Agent

4.1 Arsitektur Web Agent

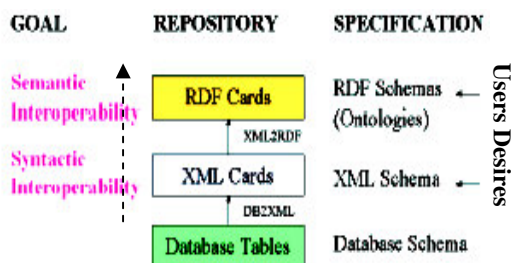
Studi literatur mengenai Web Semantic telah menghasilkan suatu rancangan *generic search agent*, yang dapat digunakan dalam berbagai domain/ontologi. Agent yang dirancang tersebut akan dapat menelusuri suatu ontologi tertentu (yang diimplementasikan sebagai suatu basis data), kemudian melakukan pencarian ke Internet, setelah mempertimbangkan strategi pencarian dari input yang dibaca melalui ontologi. Sebagaimana suatu agent, diperlukan adanya “*mind*” (pikiran) sebagai dasar untuk bertindak, dan “*body*” (tubuh: sensor

dan effector) untuk bergerak. Arsitektur dari Search Agent tersebut adalah sebagaimana tergambar dalam Figur 2. Arsitektur ini merupakan ekstensi dari arsitektur Bell (6), dengan penggunaan ontologi pada sistem penalarannya.



Figur 2. Web Search Agent

Ontologi akan menentukan dan menyaring apakah informasi yang diperoleh telah sesuai dengan keinginan ("desires") dari si pencari informasi. Ontologinya sendiri akan memiliki struktur sebagaimana digambarkan pada Figur 3.



Figur 3. Struktur Ontology

Terlihat dari Figur 3, bahwa data-data mengenai suatu dokumen yang tersimpan (misalnya metadata Dublin Core) akan diubah menjadi XML, kemudian akan dikelola dengan RDF. Dengan transformasi seperti ini akan menjamin adanya interoperabilitas secara semantik.

Untuk sensor, operator yang diperlukan antara lain:

- o getUrl(URL, Text)
- o getValue(URL, Property, Value)
- o isValid(URL)
- o isPermissible(URL, Permission)

Untuk effector, operator yang diperlukan antara lain:

- o setText(URL, String, NewString, NewURL)
- o setLink(URL, String, Link, NewURL)

Guna mendukung interoperabilitas, agent yang dirancang ini dapat berkomunikasi dengan web agent lainnya di web. Suatu domain XML yang sudah dibuat untuk keperluan komunikasi antar web agent, adalah KQML (*Knowledge Query Manipulation Language*) (7). Keuntungan dengan penggunaan konsep *Web Agents* adalah: *agent* bisa mencari data dengan informasi yang tidak lengkap, *agent* bisa bergerak pada lingkungan yang berubah-ubah (*dynamic*), *agent* dapat digunakan untuk bergerak dalam berbagai lingkungan (salah satunya dengan *Semantic Web*), dan *agent* dapat di-update ilmunya.

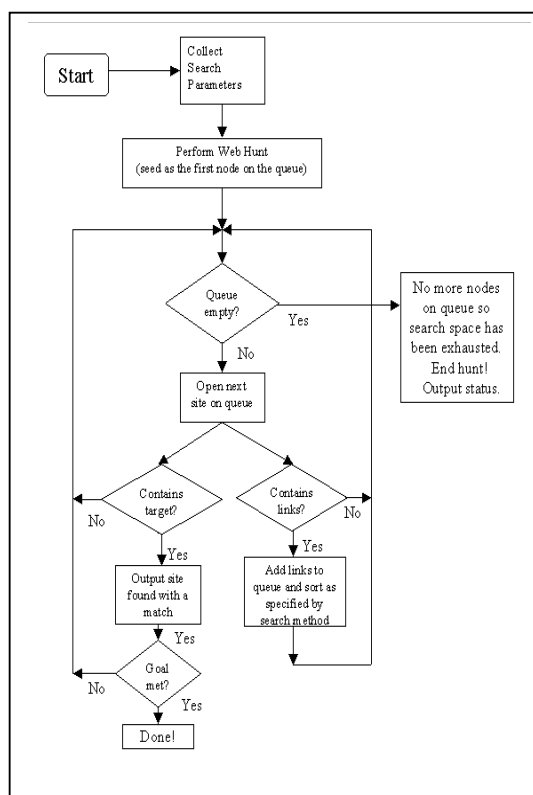
4.2 Skenario Aplikasi

Rancangan Agent yang telah dituliskan pada paragraf 4.1, akan diaplikasikan untuk membuat suatu mesin pencarian data bagi mahasiswa yang memerlukan data-data (artikel, dokumen, buku, ataupun web site) pendukung tugas akhir ataupun kerja praktek. Data-data mengenai dokumen yang diperlukan akan disimpan dalam suatu ontologi. Ontologi tersusun sebagai suatu sumber pengetahuan dalam suatu domain, misalnya: kedokteran, finansial, psikologi, komputer, dsb. Dengan demikian pencarian akan semakin terarah karena semantik dari suatu term pencarian akan terlebih dahulu dicocokkan dengan domain dimana pencarian tersebut dilakukan.

Skenario: Pengguna adalah seorang mahasiswa yang memerlukan informasi ketersediaan dokumen-dokumen di perpustakaan atau di Internet untuk referensi tugas akhirnya. Dokumen tersebut dapat berupa buku, laporan tugas akhir / kerja praktek, dokumen HTML. Pengguna memasukkan subyek dokumen yang diperlukannya tersebut. Agent tidak akan menggunakan operator boolean sebagaimana lazimnya *advanced search*. Agent akan langsung mengubah subyek yang diinput menjadi term pencarian, dan membandingkannya dengan isi ontologi (terminologi), guna meyakinkan bahwa term yang diinput tersebut masih memungkinkan adanya sinonim lainnya atau tidak. Jika ada maka term pencarian akan diperluas (meningkatkan *recall*). Hasil pencarian kemudian akan dicocokkan kembali dengan ontologi untuk meyakinkan bahwa semua hasil pencarian sesuai dengan domain yang diinginkan (meningkatkan *precision*). Hasil pencarian yang akan ditawarkan kepada pengguna akan disampaikan dalam bentuk dokumen HTML, dengan isi data adalah *links* ke setiap dokumen.

Pencariannya sendiri akan mengambil strategi BSF (*Breadth Search First*), yaitu mencoba

satu-persatu sumber data yang terdaftar dalam ontologi. Atau jika ada tawaran dari *Web Agent* lainnya dengan parameter pencarian yang sama, maka *Web Agent* disarankan untuk menggunakan juga sumber data yang ditawarkan oleh agent lainnya. Strategi ini dipilih untuk mempercepat waktu pencarian, dan meyakinkan bahwa agent tidak “terjebak” melakukan pencarian terlalu mendalam pada suatu sumber data. Algoritma *reasoning* untuk pencarian dapat dilihat pada Figur 4. Parameter pencarian akan diambil dari term pencarian setelah disesuaikan melalui ontologi.



Figur 4. Algoritma Pencarian BSF (*reasoning*)

Formalisasi dari skenario aplikasi di atas dapat dituliskan sebagai berikut:

Beliefs (Theoretical Reasoning): untuk mengubah term pencarian dari suatu perintah pencarian Web, menjadi term yang sesuai dalam domain tertentu.

```
GetURL(URL, terms, domain) →
    Ontology(terms, domain)
    PerformBSFSearch(domainTerms,
        dbLists)
SetText(URL, terms, domainTerms, dbLists)
```

Desires / Intentions (Practical Reasoning): untuk mengubah beliefs menjadi aksi untuk mengakses sumber data dalam domain tertentu sesuai dengan data-data dari ontologi.

```
SetLink(URL, domainTerms, Link, dbLists)
```

5. Kesimpulan

Hasil studi literatur mengenai *Semantic Web* ini telah memberikan suatu strategi baru untuk meningkatkan efektifitas pencarian dokumen di Internet ataupun sumber data elektronik tertentu. Pencarian tersebut menggunakan sumber data yang dikenal sebagai ontologi, dan penggunaan teknologi internet dengan XML dan RDF. Penggunaan ontologi ini diharapkan akan dapat meningkatkan *recall* dan *precision*, pada waktu pencarian dokumen. Konsep lain yang juga harus dipahami bila ingin menggunakan *Semantic Web* adalah konsep mengenai *Software Agent*.

Lebih jauh lagi, hasil studi literatur ini telah memberikan inspirasi untuk membuat suatu mesin pencarian cerdas melalui penggunaan ontologi. Mesin pencarian cerdas tersebut diharapkan dapat membantu mahasiswa untuk melakukan pencarian dokumen di Internet atau perpustakaan sesuai dengan pengetahuan dalam domain tertentu.

6. Referensi

- [1] Hongxue Wang dan John Slaney; The design of an intelligent agent for retrieving information over the Web; Department of Computer Science, Australian National University; Automated Reasoning Project, Australian National University Australia 2000
- [2] Sue Bowness; Spinning a Smarter Web; Information Highways, March/April 2004.
- [3] Berners-Lee, T., J. Hendler, dan O. Lassila, The semantic web. *Scientific American*, 284(5):34-43, 2001.
- [4] <http://www.w3.org/RDF/>; akses April 2005
- [5] Eero Hyvonen, Mirva Salminen, Miikka Junnila, dan Suvi Kettula; A Content Creation Process for the Semantic Web; Helsinki Institute for Information Technology (HIIT), University of Helsinki, 2005
- [6] Jennings, N.R.; Sycara, S.; and Wooldridge, M. 1998. "A Roadmap of Agent Research and Development", *Autonomous Agents and Multi-Agent Systems I*. Kluwer, 275-306.
- [7] Labrou, Y. and Finin, T. 1994. "A Semantics Approach for KQML – a general purpose communication language for software agents", Proceedings of the 3rd International Conference on Information and Knowledge Management, ACM Press.