

INFORMASI EKSTRAKSI BERBASIS WEB SEMANTIK PADA PEMBELAJARAN ONLINE

Rusdy Agustaf

Staf Pengajar Jurusan Teknik Informatika, Universitas Janabadra
Jalan TR Mataram No. 55-57 Yogyakarta 55231 Telp./Fax. (0274) 543676

ABSTRACT

World Wide Web (WWW) is currently very rapid progress, especially providing information and presented in the form of documents. Wwww progress has increased the demand relating to the content, and can be used to facilitate data in semantic search, so it will be a purposeful and intelligent. Access and knowledge extraction from documents online is an extremely important part in the realization of semantic web services make available better knowledge. This study aims to provide an online learning dynamic and creative through the availability of learning materials available and is sourced from the web, automatically based on ontology, the reconstruction in accordance with user preferences, the authors raised the title: "Information Extraction Web Based Semantic On Learning On line" Ontology plays an important role in the semantic web related semantic enrichment, interpretation, ontology is the domain of the courses, learning materials to produce structures and extraction of information. The main part of the ontology is generated through the Conceptual Reference Model (CRM) ontology.

Altova is an ontology editor based on graphics, providing options for defining ontology in the format RDF Schema (RDFS) have reasoning abilities, to increase the RDFS ontology reasoning is complex in semantic web.

In this paper we will present the methods we currently use to model, consolidate and store knowledge extracted from the web so that it can be re-purposed as adaptive e-Learning content. We look at how Semantic Web technology could be used within this process and also how such techniques might be used to provide content to be published via the Semantic Web.

Keywords : Information Extraction, Semantic Web, Ontology.

PENDAHULUAN

Kemajuan *World Wide Web* (WWW), dan juga sistem informasi yang sangat pesat pada belakangan ini. Kemajuan www. akan tersaji dalam bentuk dokumen, sehingga kemajuan www. dan informasi serta perkembangannya telah meningkatkan permintaan yang terkait pada konten, dan dapat digunakan untuk memfasilitasi pencarian data secara semantik, sehingga penemuan informasi yang terbaru dan pengolahan informasi ini akan terarah dan menjadi lebih cerdas. Akses dan ekstraksi pengetahuan dari dokumen secara online merupakan bagian yang sangat penting

dalam merealisasikan web semantik dan akan membuat tersedianya layanan pengetahuan yang lebih baik.

Penelitian yang dilakukan ini, bertujuan untuk menyediakan pembelajaran secara online yang dinamis dan kreatif melalui ketersediaan materi pembelajaran yang ada dan bersumber dari web, serta dapat digunakan secara otomatis yang berbasis pada ontology, kemudian melakukan rekonstruksi yang sesuai dengan preferensi penggunaannya, berdasarkan pembicaraan diatas penulis mengangkat judul :

"Informasi Ekstraksi Berbasis Web Semantik Pada Pembelajaran Online"

Dari uraian diatas, rumusan masalah yang diambil adalah sebagai berikut:

Melalui penelitian yang dilakukan ini akan dikembangkan bagaimana cara untuk dapat menghasilkan pengetahuan yang sesuai dengan struktur pembelajaran secara online sehingga bisa digunakan dalam proses anotasi dokumen web secara semantik.

Permasalahan yang dibicarakan khusus dibatasi pembelajaran secara online, yang dapat digunakan dalam proses anotasi dokumen web secara semantik.

METODE PENELITIAN

Penelitian dilakukan penulis dengan menggunakan metode sebagai berikut :

1. Penelitian Literatur dilakukan dengan mencari sumber teori, informasi dari literatur yang ada dipustaka, atau dari situs-situs pada internet / sumber-sumber lain yang benar.
2. Penelitian dilakukan lewat web untuk memperoleh layanan pengetahuan yang lebih baik, juga diperlukan cara pencarian dan juga ekstraksi pengetahuan dari dokumen web.

LANDASAN TEORI

Proses anotasi dokumen web yang tersedia merupakan kendala dalam mewujudkan web semantik [6],[5]. Anotasi yang dilakukan secara manual sangat tidak praktis dan sangat sulit sedangkan proses otomatisasi dalam proses yang ada masih belum banyak berkembang saat ini. Sehingga untuk memperoleh layanan yang ada dengan pengetahuan yang lebih baik diperlukan cara untuk pencarian dan ekstraksi pengetahuan dari dokumen web sesuai dengan domain konsep ontologi yang sudah didefinisikan.

RANCANGAN SISTEM

Pada gambar 1 berikut menunjukkan tiga klasifikasi komponen kunci yang terlibat dalam arsitektur sistem, yaitu :

Ekstraksi pengetahuan :

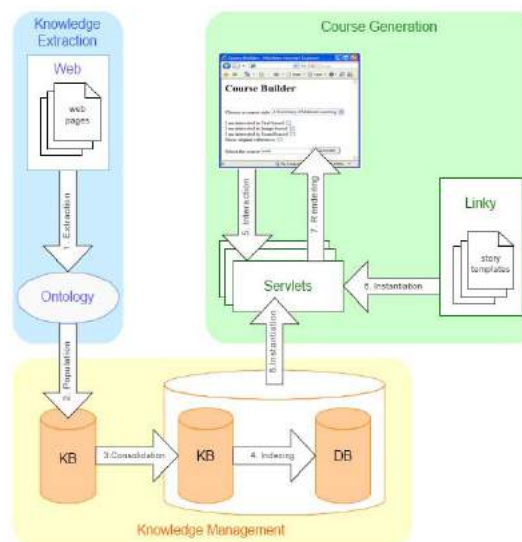
Digunakan untuk bisa melakukan ekstraksi informasi faktual untuk secara bersama dengan kalimat dan paragraf yang bersumber dari dokumen web. Informasi diperoleh dalam format RDF.

Pengelolaan Pengetahuan yang ada dan Penyimpanan

Informasi RDF disimpan pada server yang mengimplementasikan ontologi dan disimpan sesuai definisi struktur pengetahuan. Indeks basis data digunakan untuk meningkatkan kecepatan akses ke konten berbasis dokumen web.

Menciptakan materi pembelajaran :

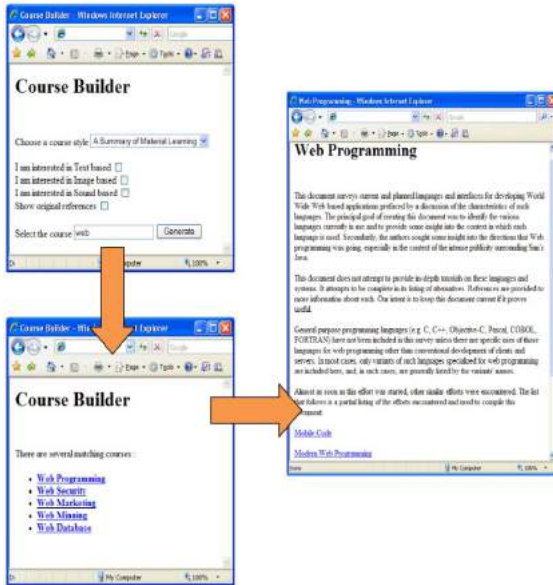
Model pembelajaran online digunakan untuk query struktur pengetahuan dan basis data. Model pembelajaran merupakan halaman web yang adaptif.



Gambar 1 Arsitektur Sistem

Pada gambar 2 ditunjukkan interaksi pengguna dengan sistem. Halaman awal pengguna menentukan mata kuliah dan type materi pembelajaran yang akan di hasilkan secara otomatis. Jika terdapat lebih dari satu mata kuliah yang memiliki struktur pengetahuan bersesuaian maka mata kuliah yang ada dapat dipilih.

Selanjutnya pengguna memperoleh materi pembelajaran yang sesuai.



Gambar 2. Otomatisasi Halaman Mata Kuliah, Materi Pembelajaran

MENGELOLA DAN EKSTRAKSI PENGETAHUAN

Ontologi memainkan peran yang penting dalam web semantik berkaitan dengan pengayaan semantik dan interpretasi [7]. Ontologi merupakan domain dari mata kuliah dan materi pembelajaran yang digunakan untuk menghasilkan struktur dan ekstraksi informasi. Bagian utama dari ontologi dihasilkan melalui Conceptual Reference Model (CRM) ontologi [3].

Proses pengembangan menggunakan bantuan Altova. Altova merupakan editor ontologi yang berbasis pada grafis [2]. Altova menyediakan pilihan untuk mendefinisikan ontologi dalam format RDF Schema (RDFS) serta memiliki kemampuan penalaran dasar sebuah ontologi. RDFS merupakan cara untuk meningkatkan penalaran ontologi secara kompleks dalam web semantik.

OWL (Ontology Web Language)

OWL dibangun dari RDF dan RDF Schema dan menambahkan vocabulary untuk

lebih mendeskripsikan property dan classes. OWL merupakan bahasa ontologi untuk Web semantik yang direkomendasikan oleh World Wide Web Consortium (W3C). OWL dibagi dalam 3 kelompok sublanguage berdasarkan ekspresi bahasanya yang digunakan sesuai kebutuhan [10] :

1. OWL Lite

Mendukung pengguna yang memerlukan klasifikasi berdasarkan hirarki dan constraint sederhana. Misalkan untuk constraint kardinalitas, OWL Lite hanya memungkinkan nilai kardinalitas 0 atau 1. OWL Lite merupakan sublanguage OWL yang memiliki ekspresi bahasa yang paling sederhana, memiliki formalitas bahasa yang lebih rendah namun lebih tinggi dari RDF Schema. OWL Lite dapat disebut sebagai ekstensi dari RDFS.

2. OWL DL (Description Logic)

Mendukung pengguna yang menginginkan ekspresi maksimal sambil tetap mempertahankan kelengkapan komputasional (segala konklusi dijamin dapat dikomputasi) dan decidability (semua komputasi akan selesai pada waktu yang terbatas). OWL DL menambahkan beberapa fitur dari yang terdapat pada OWL Lite, antara lain membuat operas himpunan seperti unionOf, intersectionOf, complementOf. OWL DL dinamakan dari korespondensinya dengan description logics, bidang penelitian yang mempelajari logics yang membentuk pondasi formal dari OWL.

3. OWL Full

Ditujukan untuk pengguna yang menginginkan ekspresi maksimum dan kebebasan syntax RDF tanpa jaminan komputasional. Sebagai contoh, pada OWL Full sebuah class dapat dianggap sebagai koleksi individual dan juga sebagai individual itu sendiri. OWL Full merupakan sublanguage yang paling kompleks dengan

batasan syntax yang begitu kecil dibanding sublanguage lainnya. Namun hal ini menyebabkan tak adanya jaminan komputasi complete. Misalnya pada OWL Full, suatu reasoning tidak dapat dijamin dapat selesai pada waktu yang finite (terbatas).

Setiap sub language ini merupakan ekstensi dari pendahulunya yang lebih sederhana [9]. Relasi berikut ini berlaku (inversenya tidak) :

- a. Setiap ontologi OWL Lite legal adalah ontologi OWL DL legal juga.
- b. Setiap ontologi OWL DL legal adalah ontologi OWL Full legal juga.
- c. Setiap konklusi OWL Lite yang valid adalah konklusi OWL DL yang valid juga.
- d. Setiap konklusi OWL DL yang valid adalah konklusi OWL Full yang valid juga.

Pengembang ontologi yang menggunakan OWL perlu mempertimbangkan sublanguage yang paling cocok dengan kebutuhannya. Pilihan antara OWL Lite dan OWL DL bergantung pada cukup tidaknya ekspresi yang terdapat pada OWL Lite sehingga memerlukan ekspresi dari OWL DL. Pilihan antara OWL DL dan OWL Full bergantung pada lebih penting melakukan automated reasoning atau memberikan ekspresi yang lebih tinggi pada model seperti memberikan meta classes (classes of classes) [4].

Ekstraksi Informasi

Ekstraksi Informasi (EI) adalah salah satu cara dari sejumlah metode yang dapat digunakan untuk memperkaya dokumen web dengan melakukan interpretasi secara semantik. Dikarenakan waktu dan upaya yang diperlukan untuk proses anotasi manual terhadap banyaknya halaman dokumen serta prasyarat dan jenis informasi maka diperlukan teknik ekstraksi yang baik [5].

Seperti diketahui bahwa informasi pada dokumen web menggunakan kosa kata

terbatas, struktur dan gaya komposisi untuk mendefinisikan konten yang sama. Hal ini membuat teknik EI sulit mengatasi variasi yang ada tersebut. Lebih penting lagi bahwa EI tidak memiliki domain pengetahuan yang diperlukan dalam menentukan relasi entitas yang akan di ekstraksi.

Sehingga dalam penelitian ini digunakan ontologi yang digabungkan dengan tujuan umum dari basis data lexical, WordNet [1], entity-recogniser, GATE (General Architecture for Text Engineering [8]) sebagai panduan untuk identifikasi pengetahuan tidak hanya entitas tetapi juga relasinya. Otomatis perluasan istilah berbasis WordNet digunakan untuk meningkatkan ruang lingkup dari analisa teks yang sesuai pola sintaksis dan definisi kata. Ketika pengguna mencari mata kuliah, dimana mata kuliah tersebut belum ada dalam basis data pengetahuan maka informasi akan di ekstraksi. Suatu proses query dilakukan pada mesin pencari seperti Google dan Yahoo.

Selanjutnya dipilih dokumen Web yang bersesuaian (sesuai dengan kriteria dan berelasi dengan topik lainnya), digunakan kata kunci untuk proses ekstraksi dari situs web yang ada dan dilakukan proses similaritas antara hasil query dan pencarian tersebut.

Untuk membangun semantik yang kaya informasi, maka perlu dilakukan ekstraksi antara relasi sehingga diperoleh identifikasi entitas pada relasi tersebut [7].

Oleh karena itu pengetahuan terkait semantik domain yang spesifik sangat diperlukan, sehingga pada akhirnya dapat digunakan untuk membuat kesimpulan dari ontologi yang ada. Dalam penelitian ini proses query ke server ontologi dilakukan untuk memperoleh pengetahuan. Pada WordNet (sinonim, hipernim dan hiponim) digunakan untuk mereduksi permasalahan variasi linguistik antara entitas sintaksis yang berbeda. Output dalam bentuk RDF (gambar 3) disimpan pada server ontologi dan menjadi sebuah pengetahuan. Sehingga RDF dapat digunakan untuk proses anotasi halaman dokumen web.

Original Sentences

Web Programming has 2 categories that is client side and server side. Client side need web browser like Internet Explorer, Opera or Mozilla Firefox. Example languages programming for client side are HTML, CSS, and Javascript. Server side need web server like Internet Information Services (IIS) or Apache. Example languages programming for server side are ASP, JSP, PHP and CGI-Perl.

```
<kb:Course rdf:intro="&kb;Course_1" kb:name="Web Programming" rdfs:label="Course_1">
  <kb:has_categories rdf:resource="&kb;Category"/>
</kb:Course>
<kb:Course rdf:intro="&kb;Category" rdf:about="&kb;Category" rdfs:label="Category">
  <kb:category_1="&kb;Client_Side"/>
  <kb:category_2="&kb;Server_Side"/>
</kb:Course>
<kb:category_1 rdf:intro="&kb;Client_Side" kb:name_tools="Web Browser" kb:example="
Internet Explorer, Opera, Mozilla Firefox" rdfs:label="Client_Side">
  <kb:has_client_programming_language rdf:resource="&kb;Client_Programming_Language"/>
</kb:category_1>
<kb:category_2 rdf:intro="&kb;Server_Side" kb:name_tools="Web Server" kb:example="
Internet Information Services (IIS), Apache" rdfs:label="Server_Side">
  <kb:has_server_programming_language rdf:resource="&kb;Server_Programming_Language"/>
</kb:category_2>
<kb:Client_Programming_Language rdf:intro="&kb;Client_Programming_Language"
kb:example="HTML, CSS, Javascript" rdfs:label="Client_Programming_Language">
</kb:Client_Programming_Language>
<kb:Server_Programming_Language rdf:intro="&kb;Server_Programming_Language"
kb:example="ASP, JSP, PHP, CGI-Perl" rdfs:label="Server_Programming_Language">
</kb:Server_Programming_Language>
```

Gambar 3. RDF Representasi Paragraf

KESIMPULAN DAN SARAN

Kesimpulan

Hasil penelitian mengenai ”Informasi Ekstraksi Berbasis Web Semantik Pada Pembelajaran Online” seperti yang telah diuraikan diatas, maka dapat disimpulkan hal-hal sebagai berikut:

Sistem yang dikembangkan melalui penelitian ini mengintegrasikan berbagai alat untuk otomatisasi proses perolehan pengetahuan berbasis ontologi dan mengelola pengetahuan tersebut untuk menghasilkan mata kuliah dengan materi pembelajaran yang sesuai, dengan melalui proses otomatisasi ini maka pembelajaran yang dihasilkan dapat lebih sesuai dengan konteks domain yang sedang dipelajari dan menjadi lebih adaptif.

Saran-Saran

Sesuai hasil penelitian diatas perlu disampaikan saran-saran sebagai berikut :

1. Solusi Pembelajaran yang akan datang diharapkan dapat mengembang Informasi Ekstraksi Berbasis Web Semantik Pada Pembelajaran Online ke bentuk-bentuk lain seperti workshop atau seminar dengan waktu yang diatur atau ditentukan.

2. Pembelajaran yang akan ada, selanjutnya diharapkan dapat mengembang Informasi Ekstraksi Berbasis Web Semantik Pembelajaran Online dengan ber-macam macam sistem operasi

DAFTAR PUSTAKA

- [1] G. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. Miller. Introduction to wordnet: An on-line lexical database. Technical report, University of Princeton, U.S.A., 1993.
- [2] H. Eriksson, R. Fergeson, Y. Shahr, and M. Musen. Automatic generation of ontology editors. In 12th Workshop on Knowledge Acquisition, Modelling, and Management (KAW99), Canada, 1999.
- [3] N. Crofts, D. Dionissiadou, and M. Stiff. Definition of the cidoc object-oriented conceptual reference model. Technical report, International Organization for Standardization, 2000.
- [4] Stanford, 2000, Protégé tutorial, How do I create ontology, http://protege.stanford.edu/doc/tutorial/get_started/how.html, diakses 17 September 2011.
- [5] R. Yangarber and R. Grishman. Machine learning of extraction patterns from unannotated corpora: Position statement. In In Proceedings of Workshop on Machine Learning for Information Extraction, pages 76–83, ECAI, Berlin, 2001.
- [6] J. Kahan and M.-R. Koivunen. Annotea an open RDF infrastructure for shared web annotations. In World Wide Web, pages 623–632, 2001. <http://citeseer.nj.nec.com/kahan01annotea.html>.
- [7] T. Berners-Lee, J. Hendler, and O. Lassila. Semantic web. Scientific American, 2001.
- [8] H. Cunningham, K. Bontcheva, V. Tablan, C. Ursu, and M. Dimitrov. Developing language processing components with gate (user’s guide). Technical report, University of Sheffield, U.K., 2002. available in <http://www.gate.ac.uk/>.

- [9] Preece, A., 2003, OWL: the Web Ontology Language, <https://users.cs.cf.ac.uk/A.D.Preece/talks/OWLTheWebOntologyLanguage/OWLTheWebOntologyLanguage.pdf>, diakses 15 September 2011.
- [10] Obitko, M., 2007, Ontologies and Semantic Web, <http://www.obitko.com/tutorials/ontologies-semantic-web>, diakses 10 Agustus 2011.