

MESIN PENCARI MEMANFAATKAN DBPEDIA KNOWLEDGE BASE

Esha Al Ma'arif

*Magister Teknik Informatika, Universitas AMIKOM Yogyakarta
Jl. Ring Road Utara, Condong Catur, Depok, Sleman, Yogyakarta 55281
Email : esha.a@students.amikom.ac.id*

Abstrak. Komunitas DBpedia mengekstrak pengetahuan terstruktur dari Wikipedia dan membuatnya bebas tersedia di Web menggunakan teknologi Semantic Web dan Linked Data. DBpedia adalah inti dari Linked Open Data Cloud dan banyak digunakan dalam penelitian dan aplikasi. Penulisan ini membahas tentang pembuatan mesin pencari berbasis web dengan memanfaatkan DBpedia sebagai knowledge base. Bahasa query yang digunakan adalah SPARQL untuk mengambil RDF data dari DBpedia. Aplikasi mengirimkan query menggunakan cURL ke DBpedia Query Builder, kemudian hasil data yang diterima akan ditampilkan kepada pengguna sesuai dengan keyword dan kategori yang diinputkan oleh pengguna.

Kata kunci : DBpedia, mesin pencari, SPARQL, cURL.

1. Pendahuluan

Kebutuhan masyarakat akan informasi saat ini sudah menjadi kebutuhan yang mendasar untuk menentukan rencana yang akan diambil ke depannya. Berbagai sarana yang dimanfaatkan sebagai sumber informasi pun sudah berkembang dengan pesat. Selain dari media, sumber informasi yang sangat populer di kalangan masyarakat saat ini adalah internet. Dengan akses internet saat ini, seakan dapat menjawab semua keingintahuan masyarakat akan segala informasi yang dibutuhkannya.

Untuk bisa mendapatkan informasi apa saja yang sesungguhnya benar-benar dibutuhkan, maka penggunaannya harus dapat memilah-milah website mana saja yang akan diakses. Beberapa website pun telah muncul untuk memberikan fasilitas layanan informasi data yang dimaksud, dan salah satunya adalah Wikipedia. Wikipedia adalah sebuah situs yang berisi tentang artikel ensiklopedia dan referensi online tingkat dunia yang bisa di akses oleh siapa saja, dimana saja selama bisa terhubung dengan internet. Wikipedia seringkali dijadikan referensi atas apapun yang dicari para pengunjung di internet.

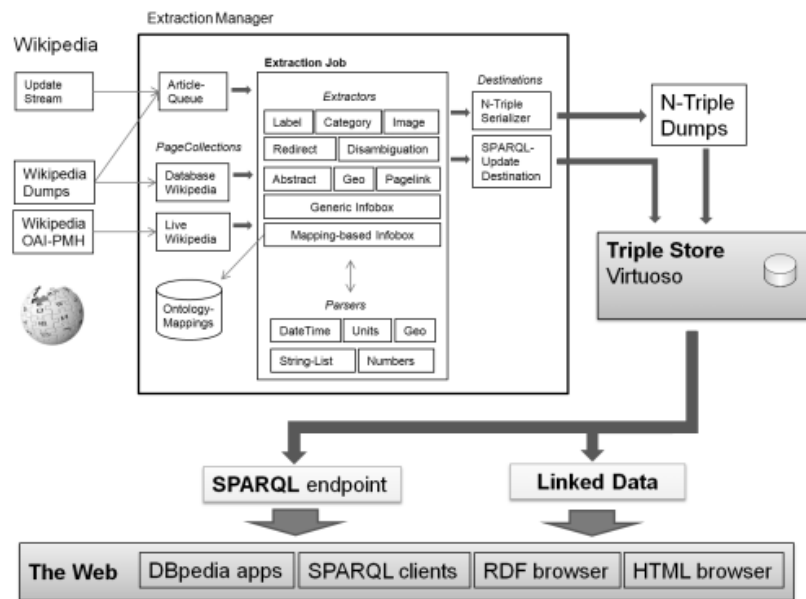
DBpedia memungkinkan pengguna mencari hubungan dan properti yang berkaitan dengan sumber daya Wikipedia. DBpedia adalah sebuah proyek yang bertujuan mengambil konten terstruktur dari informasi yang tercipta di Wikipedia. Informasi terstruktur ini tersedia di *World Wide Web*[1]. DBpedia memiliki cakupan pengetahuan manusia yang luas. Keunggulan ini menjadikannya penghubung alami untuk menghubungkan dataset, tempat dataset luar dapat bertaut pada konsep-konsepnya[2]. Dataset DBpedia saling tertaut di tingkat RDF dengan sejumlah dataset Data Terbuka lainnya di web.

Pada penelitian ini penulis menggunakan SPARQL untuk melakukan *query* dari DBpedia. SPARQL (*Simple Protocol and RDF Query Language*) merupakan suatu bahasa *query* yang dapat digunakan untuk mengakses data pada *semantic web*. Dengan menggunakan SPARQL memungkinkan untuk melakukan beberapa hal antara lain. Mengambil nilai dari data yang terstruktur maupun data yang semi terstruktur. Mengembangkan data dengan melakukan *query* terhadap suatu relasi yang tidak diketahui. Dapat melakukan *query* operasi join yang kompleks pada *database* yang berlainan secara lebih sederhana. Mengubah suatu data RDF (*Resource Description Framework*) menjadi *vocabulary* yang lain. Hasil dari *query* SPARQL dapat mengembalikan nilai dalam beberapa format data yang antara lain : XML, RDF, dan HTML[3].

SPARQL *Endpoint* disediakan untuk melakukan *query* pada *knowledge base* DBpedia. Untuk melakukan *query* sebaiknya mempelajari *ontology* yang digunakan DBpedia. Dalam membuat *query*, perlu mengetahui properti-properti yang digunakan pada DBpedia melalui halaman suatu entitas[4].

Ontologi DBpedia adalah ontologi lintas-domain dangkal, yang telah dibuat secara manual berdasarkan infobox yang paling umum digunakan di Wikipedia. Ontologi saat ini mencakup 685 kelas yang membentuk hierarki subsumption dan digambarkan oleh 2.795 properti yang berbeda[5].

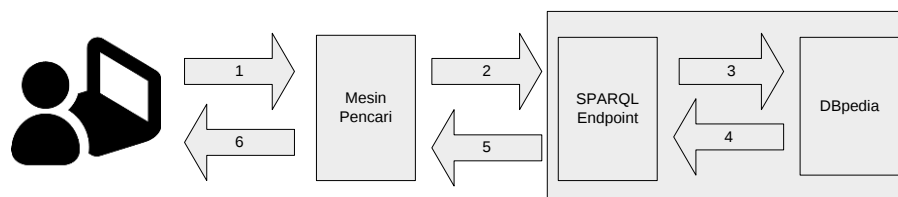
Untuk mengakses sebuah entitas dapat menggunakan URI *resource* dari entitas tersebut. Apabila URI *resource* diakses menggunakan *Semantic Web agents* maka mengembalikan RDF *descriptions*, sedangkan apabila diakses menggunakan *Web Browser* akan menampilkan informasi dari entitas yang diakses dalam tampilan HTML sederhana[4]. Gambar 1 menunjukan alur dan rangkaian ekstraksi data DBpedia dari Wikipedia.



Gambar 1. Ikhtisar kerangka ekstraksi Dbpedia^[6]

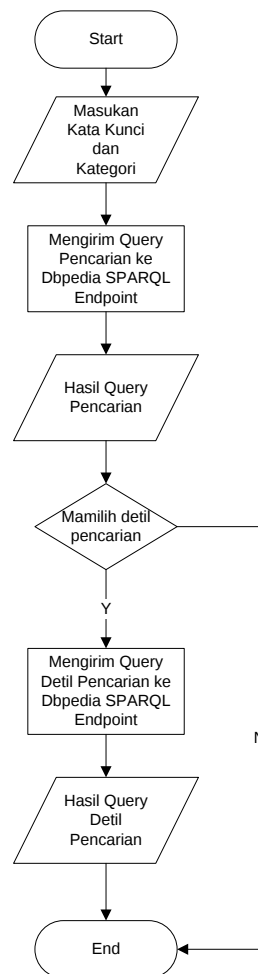
2. Pembahasan

Mesin pencari yang diimplementasikan pada penelitian ini berupa aplikasi berbasis web menggunakan bahasa pemrograman PHP. Aplikasi mengirimkan query SPARQL ke DBpedia SPARQL Endpoint menggunakan cURL. Kemudian query tersebut menghasilkan data dari DBpedia yang ditampilkan kepada user sesuai kata kunci dan kategori yang dimasukkan. Gambar 2 menunjukkan gambaran umum perancangan proses pengiriman data pada mesin pencari.



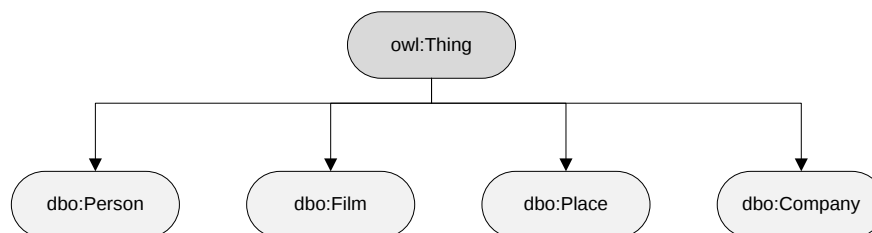
Gambar 2. Gambaran umum perancangan proses pengiriman data

Pada Gambar 2 mendeskripsikan proses pertama pengguna memasukkan kata kunci dan kategori. Kedua mesin pencari mengirimkan *syntax query* ke SPARQL *Endpoint*. Ketiga SPARQL *Endpoint* melakukan *query* berdasarkan *syntax* yang dikirimkan. Ke-empat hasil *query* dari DBpedia didapatkan oleh SPARQL *Endpoint*. Kelima SPARQL *Endpoint* mengirimkan hasil *query* ke mesin pencari. Ke-enam mesin pencari menampilkan hasil pencarian kepada pengguna. *Flowchart* proses pencarian yang dilakukan oleh pengguna melalui mesin pencari berbasis web ini dapat dilihat pada Gambar 3.



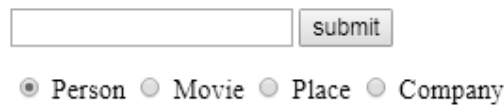
Gambar 3. *Flowchart* proses pencarian

Pada pembuatan aplikasi ini penulis menganalisa *ontology* yang ada pada DBpedia dan mengambil empat kelas yang menjadi batasan kategori pencarian. Pada aplikasi mesin pencari ini penulis mengambil batasan empat macam kategori diantaranya: *Person* (orang), *Movie* (film), *Place* (tempat), *Company* (perusahaan). Gambar 4 menunjukkan pembagian kelas ontologi DBpedia yang digunakan dalam mesin pencarian.



Gambar 4. Ontologi kelas DBpedia yang digunakan pada mesin pencarian

Melalui form aplikasi tersebut pengguna menulis kata kunci ke dalam *textbox* yang disediakan, kemudian memilih kategori dan menekan tombol submit. Selanjutnya aplikasi akan mengirimkan query ke SPARQL *Endpoint* yang disediakan oleh Dbpedia. Pada aplikasi ini SPARQL *Endpoint* yang digunakan adalah: <http://dbpedia.org/sparql>. Tampilan antarmuka mesin pencari dapat dilihat pada Gambar 5.



Gambar 5. Tampilan antarmuka mesin pencari

Sebagai contoh pengguna memasukkan kata kunci “morgan freeman” dan memilih kategori *Person*, syntax query yang digunakan ditunjukkan pada Gambar 6. Pada query tersebut setidaknya menggunakan empat *prefix* yang ditunjukkan pada Gambar 7.

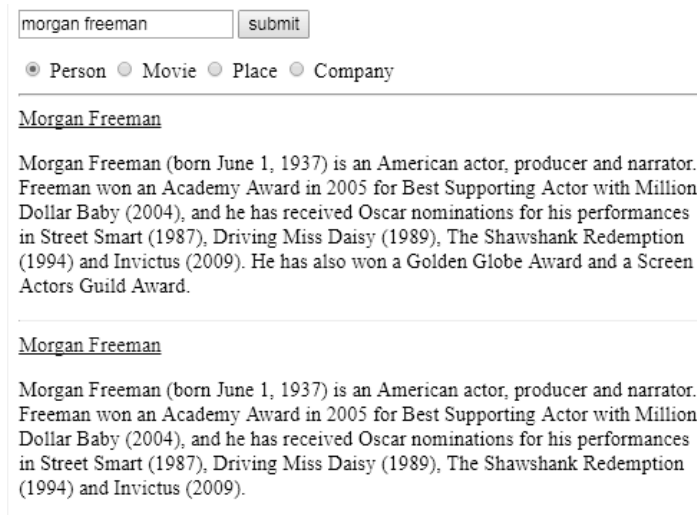
```
SELECT ?name ?person ?description WHERE {  
  ?person rdf:type dbo:Person.  
  ?person foaf:name ?name .  
  ?person rdfs:comment ?description .  
  FILTER (regex(?name, 'morgan freeman', 'i')) .  
  FILTER (lang(?description) = 'en')  
  } ORDER BY ?name
```

Gambar 6. Query pencarian “morgan freeman” dengan kategori *Person*

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>  
PREFIX dbo: <http://dbpedia.org/ontology/>  
PREFIX foaf: <http://xmlns.com/foaf/0.1/>  
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
```

Gambar 7. *Prefix* yang digunakan dalam pencarian

Dari hasil *query* SPARQL tersebut aplikasi menampilkan hasil pencarian seperti yang ditampilkan pada Gambar 8. Pengguna dapat melihat detil informasi dari link hasil pencarian tersebut seperti yang terlihat pada Gambar 9.



Gambar 8. Tampilan hasil pencarian



Gambar 9. Tampilan detail pencarian

Hasil pengujian mesin pencari dapat dilihat pada Tabel 1, berdasarkan beberapa kata kunci dan kategori yang dimasukkan. Dari Tabel 1 kita dapat melihat jumlah hasil pencarian dan kecepatan waktu pencarian.

Tabel 1. Hasil Pengujian Mesin Pencari

No	Kata Kunci	Kategori	Jumlah Hasil Pencarian	Waktu Pencarian (detik)
1	taylor swift	Person	1	1.091
2	jackie chan	Person	4	1.167
3	joko widodo	Person	1	0.990
4	kim jong-un	Person	1	1.248
5	albert einstein	Person	2	1.010
6	home alone	Movie	8	1.823
7	twilight	Movie	45	2.074
8	insidious	Movie	5	1.734
9	iron man	Movie	11	1.860
10	terminator	Movie	10	1.385
11	jakarta	Place	43	5.802
12	tokyo	Place	60	9.890
13	amazon	Place	32	6.017
14	barcelona	Place	53	6.604
15	borobudur	Place	1	1.145
16	apple	Company	27	3.303
17	microsoft	Company	18	1.995
18	coca-cola	Company	14	1.508
19	adidas	Company	3	2.424
20	samsung	Company	23	1.839

3. Simpulan

Pada penelitian ini penulis mengembangkan mesin pencarian berbasis web dengan memanfaatkan DBpedia *knowledge base*. Mesin pencari mengirimkan *query* ke SPARQL *Endpoint* kemudian berdasarkan *query* tersebut mengambil *knowledge* yang ada pada DBpedia yang merupakan hasil ekstraksi dari Wikipedia. Hasil *query* tersebut kemudian menjadi informasi bagi pengguna. Hasil pencarian yang didapatkan dapat dilihat kecepatan pencarian dan jumlah hasil pencarian. Saran untuk penelitian selanjutnya dengan memanfaatkan *knowledge base* yang ada pada DBpedia dapat mengembangkan berbagai aplikasi yang bermanfaat, sehingga dapat menghasilkan semakin banyak sarana informasi bagi pengguna.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada kedua orang tua atas segala dukungan dan doa yang selama ini diberikan. Terima kasih juga kepada para dosen dan teman-teman Magister Teknik Informatika

Universitas Amikom, beserta semua pihak yang telah membantu memberikan ide dan pengarahan dalam penyusunan makalah ini.

Daftar Pustaka

- [1]. Bizer, Christian; Lehmann, Jens; Kobilarov, Georgi; Auer, Soren; Becker, Christian; Cyganiak, Richard; Hellmann, Sebastian (September 2009). "DBpedia - A crystallization point for the Web of Data". *Web Semantics: Science, Services and Agents on the World Wide Web* 7 (3): 154–165. ISSN 1570-8268
- [2]. E. Curry, A. Freitas, and S. O’Riáin, "The Role of Community-Driven Data Curation for Enterprises," in *Linking Enterprise Data*, D. Wood, Ed. Boston, MA: Springer US, 2010, pp. 25-47.
- [3]. Awaludin , Muhammad. 2009. *Sistem Navigasi dan Pencarian Berbasis Konteks Pada Konten E-Learning Menggunakan Teknologi Web Semantik*. Institut Teknologi Sepuluh November. Surabaya.
- [4]. Id.DBpedia.org, "DBpedia Indonesia", <http://id.dbpedia.org/wiki/>, diakses tgl 30 Desember 30, 2017.
- [5]. DBpedia.org, "Ontology", <http://wiki.dbpedia.org/services-resources/ontology>, diakses tgl 6 Januari 2018.
- [6]. Wiki.DBpedia.org, "Former PHP-based Information Extraction Framework", <http://wiki.dbpedia.org/services-resources/documentation/phpframework>, diakses tgl 30 Desember 2017.