

PENGGUNAAN REPOSITORY SEMANTIC WEB CONTENT MODEL (SWCM) UNTUK METADATA PUBLIKASI ILMIAH

Devi Munandar¹⁾, Taufiq Wirahman²⁾

Pusat Penelitian Informatika Lembaga Ilmu Pengetahuan Indonesia^{1,2)}

Kompleks LIPI Gedung 20 Lantai 3

Jl Sangkuriang No 21/154D Cisitu Bandung 40135

Telepon: 022-2504711 Faks: 022-2504712

E-mail : devi@informatika.lipi.go.id¹⁾, taufiq@informatika.lipi.go.id²⁾

Abstrak

Teknologi Semantic Web telah berkembang sejalan dengan senantiasa diupdatenya teknologi web dari Web1.0 yang memenuhi kebutuhan home page, portal hingga teknologi Web 2.0 yang dirancang untuk menampilkan micro content. Saat ini belum adanya sebuah model yang mampu dan dapat mengakomodasi serta merepresentasikan secara aktual sebuah web of content resources dengan metadata dari semantic web. Walaupun semantic web merupakan perluasan dari sebuah web namun belum ada suatu model formal untuk medeskripsikan hasil kombinasi hal tersebut. Pada makalah ini akan dicoba membuat content management model yang dikombinasikan dengan metadata dalam penggunaan sebuah web secara ekspresif dan fleksibel pada sebuah semantic web. RDF selain menguraikan tentang meta-data di web juga merupakan representasi format dasar untuk sebuah knowledge semantic web yang mengandung konsep data biner. Keterbatasan penggunaan RDF triple stored dalam hal penyimpanan data biner yang besar membutuhkan sebuah model yang dapat menguraikan, mengakses serta melakukan perubahan pada resources dalam sebuah content web. Metodologi untuk persyaratan SWCM dapat menggunakan metode Granularitas, Ekspresivitas, kompatibilitas, Penamaan, Pencarian, Cara pandang penyajian, Hubungan antar relasi, Formalisasi, Hak Akses serta penalaran. Dengan menggunakan metodologi tersebut SWCM diimplementasikan dalam 2 layer yaitu core layer yang mengandung Binary Store (BinStore) dan repository layer untuk pemeliharaan item yang merupakan sentral konsep yang menjembatani RDF dengan web dengan pengalamanan via URI dan merujuk ke web content di saat run-time. Metadata publikasi ilmiah merupakan sarana yang digunakan pada makalah ini yang ditempatkan pada Repository untuk SWCM.

Kata Kunci : meta-data, semantic web, content management, RDF.

PENDAHULUAN

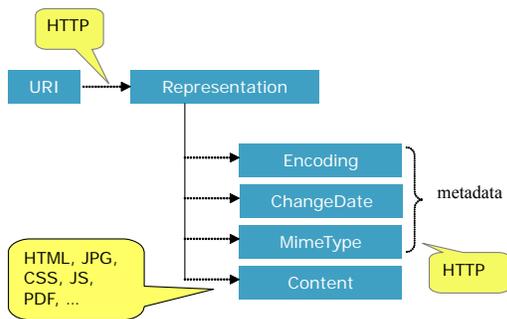
Dunia maya atau Internet telah banyak mempengaruhi kehidupan di setiap lapisan masyarakat di dunia saat ini. Dengan banyaknya orang yang menggantungkan perkembangan kebutuhan informasinya kepada Internet, sehingga teknologi yang dipergunakan dalam pembangunan sebuah situs web pun terus berkembang. Dari kemunculan pertama kalinya perkembangan web diawali dengan teknologi Web 1.0, dalam hal ini pengguna di hadapkan pada fitur yang hanya bisa mencari (searching) dan melihat-lihat (browsing) data informasi yang ada di web, kemudian bergeser pada era pengembangan web kedua (Web 2.0) di mana pengunjung mulai dapat melakukan interaksi dengan diatur oleh sistem yang ada pada web. Adapun Jenis interaksi yang dapat dilakukan pada generasi web kedua ini antara lain untuk saling bertukar informasi (sharing), eksploitasi informasi, dan juga pembuatan

komunitas-komunitas online seperti yang marak saat ini, seperti Friendster, Multiply, YouTube, dan lain-lain. Masing masing komunitas ini mempunyai kepentingannya sendiri dalam saling bertukar data maupun informasi yang mereka himpun. Dalam era inilah sebenarnya interaksi sosial dalam dunia maya mulai dikembangkan. Dan mulai dari era ini pulalah ide untuk mengembangkan aspek sosial sebuah web mulai dipikirkan sehingga lahirlah yang semantic web. Secara umum semantic web hanyalah perluasan dari teknologi web yang ada, bukan suatu teknologi yang berbeda sama sekali. Hanya saja, semantic web tidak hanya menampilkan informasi dalam bahasa alami yang hanya dipahami oleh manusia tetapi juga dalam bentuk yang dapat dimengerti, diinterpretasi dan diolah oleh mesin. Dalam perkembangannya semantic web menghadapi kendala yang beragam. Format publikasi isi Web juga masih menjadi perdebatan karena untuk menampilkan sebuah informasi akan diperlukan dua

format yang berbeda, untuk dilihat manusia dan dapat diproses oleh mesin. Pada saat ini belum ada sebuah formal model yang menjabarkan bagaimana perpaduan dari content web dengan meta data. Pada generasi pendahulunya, semantic web mempunyai kekuatan pada pendefinisian Resource Description Framework (RDF), namun RDF dianggap sesuatu bentuk yang lemah sebagai content binary atau XHTML karena tidak dapat di integrasikan dengan web content. Dengan menggunakan teknologi arsitektur Semantic Web Content Repository maka permasalahan ini coba diminimalisasi dengan menggabungkan kedua bentuk tersebut dengan membuat semantic web content model sehingga RDF dan web arsitektur merupakan superset dari model tersebut.

Web

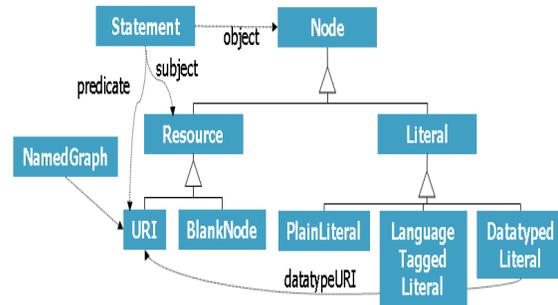
Untuk membangun sebuah web biasanya digunakan style arsitektur untuk mendefinisikan setiap komponen yang nantinya merupakan bagian dari word wide web (WWW). Pendeskripsian sebuah addressable resources yang ditunjukkan dengan merespon pengiriraman self-describing yang dikirimkan ke www tersebut. Salah satu penggunaan arsitektur untuk membangun sebuah web menggunakan metoda REST (*Representational State Transfer*) yang digunakan bagaimana caranya mendistribusikan, menguraikan komponen-komponen pembentuk web serta mempunyai batasan-batasan bagi setiap elemen pembentuknya. Salah satu pernyataan batasan yang digunakan dengan metoda REST adalah “Hypertext merupakan mesin dari sebuah aplikasi”, maksudnya adalah masing-masing respon yang diterima berisikan URIs dari *resources* terkait tanpa dapat di integrasikan dengan *resources* lainnya. Dengan asumsi tersebut maka tidak ada suatu model pun yang dapat membuat keterkaitan antar resources selain menggunakan RDF (*Resources Description FrameWork*).



Gambar 1. Model Web [1]

Didalam www resources dialamatkan menggunakan URIs yang disajikan mengandung dengan karakter stream dan berupa encoding metadata, isi web, jenis metadata lainnya seperti perubahan tanggal dan jenis lainnya. Pada kenyataannya sangat banyak jenis

metadata lainnya yang diikuti sertakan pada sebuah web seperti isi web yang di kompres serta pengontrolan cache.[1]



Gambar 2. Model Semantic Web [1]

Semantic Web

Resouces Description FrameWork (RDF) merupakan presentasi format dasar untuk knowledge pada semantic web. Pada awalnya RDF digambarkan sebagai suatu format untuk menguraikan metadata mengenai *resources* di web. Pada Gambar 2 dapat di lihat sebuah RDF model berikut dengan notasi *NamedGraphs*-nya yang digunakan pada SPARQL. Walaupun RDF secara konseptual mengandung *binary data* yang dapat menyimpan *xsd:base-64Binary datatype literal*, namun tidak ada cara yang digambarkan untuk mengubungkan URIs dengan isi dari web. Saat ini RDF triple store tidak dirancang untuk menyimpan data biner dalam kapasitas besar. Pembuatan Library untuk RDF merupakan cara mengurangi beban untuk mendefinisikan, mengakases, mengubah isi dari suatu web. Masalah kedua adalah kurangnya tools yang tersedia untuk penulisan RDF. Oleh karena itu pembuatan RDF dapat dibagi menjadi 2 classes:(1) *generic*: dimana pengguna dapat merubah skema di saat run-time, dan (2) *fixed-schema*: skema merupakan *pre-defined*, sebagai contohnya *tools* nya adalah sebuah *address book editor* tanpa output data dalam sebuah *fixed RDF format*.

Pembuatan RDF secara generic tanpa pendefinisian skema terlebih dahulu dianggap sangat fleksibel. Contoh penggunaannya misalkan setiap RDF resource tanpa isi atau kosong, berisi satu atau multipel label. Dengan menggunakan aplikasi maka contoh tersebut dapat ditangani. RDF dapat dipanggil oleh sebuah bahasa assembly untuk data yang dapat menampilkan hampir segalanya dari RDF, namun kekurangannya adalah pada high-order untuk membuat lebih efisien untuk berinteksi langsung dengan manusia.

METODOLOGI

Kriteria semantic web content model

Kriteria pembuatan sebuah semantic web content model dapat di definisikan mempunyai beberapa acuan.

Granularitas adalah semantic web dapat mendeskripsikan content web yang bisa berkolaborasi dengan web lainnya dan dapat berisikan micro-content yang berisikan alamat berupa tag komentar, blog, imej ataupun video. **Ekspresivitas** adalah sebuah model yang mempunyai RDF yang fleksibel dan terdefinisi untuk menghubungkan web-content, dengan menggunakan skema dan mind-mapping RDF diharapkan mampu mengintegrasikan data dari resource yang berbeda. **Kompatibilitas** adalah bagaimana menggunakan semantic web content model bersama-sama dengan framework yang diinginkan. Terutama penggunaan background-knowledge yang terdapat dalam RDF secara bersamaan dengan semantic web content model. **Penamaan** adalah URL yang mengandung isi web atau file name atau sebuah path, penamaan disarankan memudahkan pengguna untuk mengingatnya, karena kita mengetahui semantic web menggunakan URIs dengan penamaan yang unik untuk mendefinisikan bermacam-macam resources. **Pencarian** merupakan kriteria yang dibutuhkan oleh semantic web content model untuk kemampuan melakukan query pada pencarian meta data di semantic web. **Representasi Penyajian** adalah setiap model dapat dipakai oleh end-users, oleh karena itu ada beberapa persyaratan yang harus dipenuhi pembuatan struktur meta-data. Penyajian metada merupakan representasi human-readable yang lebih dimengerti oleh pengguna. Selain itu ada kriteria lainnya yang dipersyaratkan seperti Relasi inverse, Kebebasan menentukan formalisasi, hak akses, serta penalaran.

Arsitektur SWCM

Pada pembuatan SWCM perancangannya menggunakan 2 layer: Model layer utama (*core*) dan *repository layer* untuk penyimpanan item-item. Pada **Core Layer** mengandung sebuah *simple binary store* (*BinStore*), *RDF Named Graph*, dan indeks teks. *BinStore* dapat mengakses isi data secara acak, dengan menggunakan *interface BinStore* dapat melakukan pembacaan berulang-ulang dengan hanya melakukan sekali penulisan kode. RDF *repository* dimodelkan sebagai sebuah Set Model. Set Model yang dirancang merupakan abstraksi layer dari RDF-triple dan quad yang memudahkan *programmer* dalam melakukan pengembangan. Saat ini *OpenRDF* digunakan dalam berbagai implementasi. Untuk meningkatkan performa dan kecepatan dalam pencarian dengan menggunakan *Query*, maka *Apache Lucene* digunakan sebagai sebuah indeks yang full teks. Indeks disimpan sebagai *inverse mappings* untuk item URI, konten maupun URI, *Property URI*, *RDF literal content*. Untuk pengindekan *Binary content*, semua *Binary Content* di indeks setelah semuanya di tulis di *BinStore*. Cara ini mirip dengan indeks yang dilakukan terhadap RDF *literal* setelah semuanya di tulis terlebih dahulu. Pola *Proxy* digunakan untuk memisahkan API dari indeks. Dengan menghapus *resources* maka indeks juga turut di

hilangkan. SPARQL digunakan untuk mengakses RDF dan *full text index*, secara umum idenya adalah memisahkan *query* mengeksekusinya secara terpisah: satu bagian *query* digunakan untuk RDF store dan satu lagi digunakan untuk *full text index*. Setelah hal ini dilaksanakan maka dilakukan penggabungan oleh item URIs, tergantung dari ukuran yang dihasilkannya. Pembentukan pertama adalah full text query dan kemudian disatukan dengan URI didalam RDF query yang menghasilkan URIs. Core layer pada intinya hanya melaksanakan perintah *starting*, *running*, *stopping* pada 3 komponen: *RDF Store*, *text index* dan *BinStore*. Pada **Repository Layer** digunakan untuk *update* dan melakukan query terhadap *Core Layer*. Ini lebih pada menyederhanakan proses *debug* untuk dapat melakukan *sharing* secara mudah terhadap item data. Pada *repository*, disaat *runtime* masing-masing item bisa dikunci. Sesudah dikunci, item tidak dapat di edit oleh pemakai lain. Proses dapat juga mengunci item secara otomatis apabila pengguna melakukan permintaan untuk mengunci proses tersebut supaya di saat pengeditan pengguna lain hanya dapat mengakses secara *read only*.

IMPLEMENTASI

Pada implementasi penulisan makalah ini, digunakan data publikasi ilmiah sebagai metadata untuk perancangan SWCM. Metadata disimpan dalam bentuk RDF. Dua RDF model digunakan untuk setiap SWCM, pertama untuk data aktual sebagai model eksplisit oleh pengguna (*user model*) dan yang kedua sebagai hasil dari proses RDF yang digunakan untuk *query* (*index model*). Sebuah Item x disimpan sebagai

```
<x> a swcm:Item
```

konten dari Item disimpan dalam *BinStore*. NameItem dengan URI x dan nama konten "TechnicalReport" direpresentasikan dalam RDF sebagai:

```
<x> a swcm:NameItem; swcm:hasContent  
"TechnicalReport"
```

Maksudnya konten dari NameItems saat ini disimpan dalam RDF. Selanjutnya dapat disimpan dalam *BinStore*, supaya memudahkan untuk melakukan debug dan implementasi dari mesin query. Sebuah Relasi p ("work for") dengan inverse q ("employs"). Di representasikan sebagai:

```
<p> a swcm:Relation; swcm:hasContent  
"works for"; swcm:hasInverse <q>.  
<q> a swcm:Relation; swcm:hasContent  
"employs"; swcm:hasInverse <p>.
```

sebuah *statement* s dari a ke b dengan relasi p di representasikan sebagai:

```
<s> a swcm:Statement;  
swcm:hasSource <a>; swcm:hasTarget  
<b>; swcm:hasRelation <p>
```

untuk statement ini, maka data juga ditulis dalam Model Index RDF:

```
<a> <p> <b>. and <b> <p-inverse> <a>.
```

maksud pernyataan diatas, melakukan inverse triple kedalam model index

```
:SWIBIB a swcm:NameItem , swcm:Item
;swcm:hasChangeDate "2008-10-13T08:45:02Z"^^xsd:dateTime
;swcm:hasContent "SWIBIB Semantic Web For Bibliography Text".
:employs a swcm:NameItem ,
swcm:Item , swcm:Relation
;swcm:hasAuthor swcm:anonymous-author
;swcm:hasChangeDate "2008-10-13T08:45:03Z"^^xsd:dateTime
;swcm:hasContent "employs"
;swcm:hasInverse :employedBy .
:worksFor a swcm:NameItem ,
swcm:Item , swcm:Relation
;swcm:hasAuthor swcm:anonymous-author
;swcm:hasChangeDate "2008-10-13T08:45:04Z"^^xsd:dateTime
;swcm:hasContent "works for"
;swcm:hasInverse :employs .
```

KESIMPULAN

Pada makalah ini dibahas mengenai Semantic Web Content Model yang di gunakan unuk content management. Model yang di hasilkan adalah kombinasi fitur ekspresif model semantic web yang lebih di kenal dengan nama RDF. Hasil dari content model di tempatkan dari berbagai cara sebagai persyaratan sebuah content management. Walaupun hasil dari model yang didapatkan sama seperti RDF, SWCM ditargetkan untuk browsing dan authoring oleh pengguna. Dengan penambahan batasan, semua content yang mempunyai URI, maka setiap relasi mempunyai inverse, dapat juga tersedia dua relasi yang dipunyai oleh label yang sama. SWCM beserta implementasinya digunakan pada Semantic Web dengan metadata bibliography text dengan pendekatan iMapping dan konseptual struktur data. Data disimpan dalam BinStore dalam repository supaya memudahkan untuk melakukan debug dan implementasi dari mesin query. Dengan mendefinisikan statement maka data yang ditulis dalam index.rdf akan mempunyai inverse triple kedalam model index.

DAFTAR PUSTAKA

- [1] Max Volkel. (2007). *A Semantic Web Content Model and Repository*. Proceeding of the 3rd International Conference on Semantic Technologies.
- [2] H.Peter Alesso, Craig F.Smith (2005). *Developing Semantic Web Services*. A.K Peter Ltd

- [3] The Semantic Web Real World Applications From Industry. Springer Science+Business Media, LLC.(2008)
- [4] Nuescheler, D. (2005). Content repository api for java technology specification. Technical Report Java Specification Request 170, Day Management AG, Switzerland
- [5] Haller, H., Volkel, M., Kugel, F. (2006). Mapping wikis - towards a graphical environment for semantic knowledge management. In Schaert, S., Volkel, M., Decker, S., eds.: Proceedings of the First Workshop on Semantic Wikis – From Wiki To Semantics.
- [6] <http://semanticweb.org/wiki/CDS>
- [7] <http://semanticweb.org/wiki/RDF2Go>
- [8] <http://code.google.com/p/swecr/wiki/SWCM>