

WEB SEMANTIK SEBAGAI MEDIA DISTRIBUSI DATA BIODIVERSITAS DENGAN MEKANISME GRDDL

Ovan S. Pulu

²*Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Gunadarma
Jl. Margonda Raya no. 100, Depok 16424, Jawa Barat
ovan@staff.gunadarma.ac.id*

Abstrak

Pemanfaatan internet khususnya website sebagai media sosialisasi dan distribusi data semakin mengalami peningkatan hal ini mendorong teknologi yang digunakan pun menjadi semakin berkembang ditandai dengan hadirnya web semantik. Web semantik memungkinkan proses ekstraksi data bisa dilakukan oleh perangkat lunak secara mandiri karena menggunakan format dokumen XHTML dengan memanfaatkan XSLT sebagai media transformasi data untuk menghasilkan format dokumen standar XML. Pada format inilah data Biodiversitas bisa diekstrak dengan lebih mudah oleh perangkat lunak lainnya. Adapun data hasil ekstraksi dokumen tersebut mengacu pada Ontologi Darwin Core.

Kata Kunci : *Web semantic Biodiversitas, GRDDL*

WEB SEMANTIK AS MEDIA DISTRIBUTION DATA BIODIVERSITY WITH MECHANISM GRDDL

Abstract

Utilization of internet especially website as media of socialization and distribution of data progressively increase this matter pushing used technology also become progressively marked by presence of semantic web. Semantic Web allows data extraction process can be done by software independently because it uses XHTML document format by utilizing XSLT as data transformation medium to generate XML standard document format. In this format Biodiversity data can be extracted more easily by other software. The data extracted the document refers to Darwin Core Ontology.

Keywords : *Semantic Web Biodiversity, GRDDL*

PENDAHULUAN

Data biodiversitas di Indonesia tidak bisa dipungkiri masih cukup sulit dalam hal pengaksesannya, keterbukaan informasi tidak serta merta mendorong data ini bisa diakses secara bebas dan mudah oleh masyarakat luas. Hampir semua situs yang berhubungan dengan data ini cukup sulit untuk memberikan data yang

lengkap, distribusi data masih sangat minim ditambah lagi format data yang tidak mendukung interoperabilitas dan hanya bisa dibaca oleh perangkat lunak tertentu.

Pemanfaatan Web semantik sebagai media distribusi data menjadi penting dikarenakan telah mendukung sistem pertukaran data yang bebas dan terformat

tanpa perlu memikirkan perbedaan platform. Tim Banners-Lee pencetus istilah ini mengatakan web semantik memberikan makna dari informasi yang terformat dengan baik, serta memungkinkan data bisa diproses secara langsung maupun tidak langsung oleh komputer.

METODE PENELITIAN

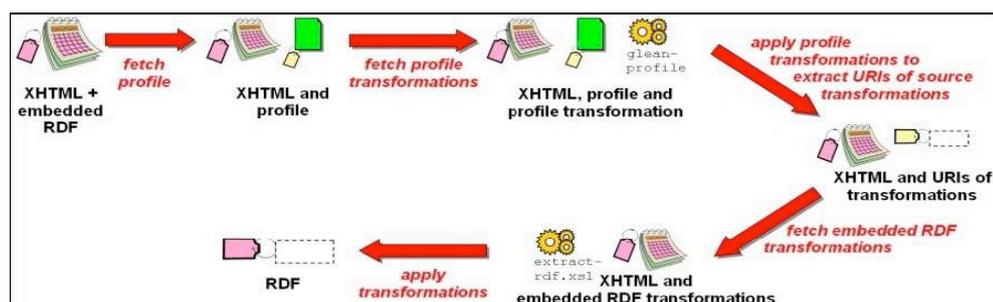
Gleaning Resource Descriptions from Dialects of Languages (GRDDL) merupakan teknik untuk memperoleh data Resource Description Framework (RDF) dari dokumen Extensible Markup Language (XML) dan eXtensible HyperText Markup Language (XHTML) tertentu yang diasosiasikan dengan algoritma transformasi untuk menterjemahkan, biasanya direpresentasikan dalam bentuk XSLT. Pengguna membaca dokumen dengan mengikuti link yang ada di web dengan menggunakan teknik yang dijelaskan dalam spesifikasi GRDDL untuk menemukan transformasi yang sesuai. Pada gambar 2.1 digambarkan bagaimana suatu dokumen XHTML yang sudah disisipkan RDF sampai dengan proses transformasinya. Pada gambar diatas sumber data berasal dari dokumen XHTML yang sudah dilengkapi dengan metadata RDF, kemudian dilakukan pemisahan antara XHTML dan profile yang digunakan. Profile ini berisi kumpulan metadata yang digunakan pada dokumen tersebut. Selanjutnya metadata tersebut diidentifikasi transformasinya sesuai dengan aturan pada Darwin Core. Pada tahapan selanjutnya dilakukan

pemisahan antar profile sesuai dengan referensi masing-masing pada alamat transformasi untuk memudahkan proses ekstraksi data. Setelah dokumen XHTML dan alamat transformasi lengkap maka bisa dilakukan ekstrak data dengan memanfaatkan teknologi *Extensible Stylesheet Language Transformations (XSLT)* menggunakan XSL untuk merepresentasikan skema data beserta bahasa pemrograman PHP. Hasil ekstraksi XSLT ini akan menghasilkan dokumen dengan RDF yang sudah terformat.

Penggunaan mekanisme GRDDL untuk web semantik bertujuan untuk menghasilkan sebuah dokumen RDF yang bisa digunakan sebagai format umum untuk pertukaran data sehingga menjadi lebih mudah dalam pemrosesan serta mempunyai kesamaan format dokumen.

Darwin Core

Darwin Core adalah sebuah badan standardisasi keanekaragaman hayati, yang mendefinisikan daftar istilah untuk memudahkan dalam berbagi informasi keanekaragaman hayati dengan disertai referensi definisi, contoh maupun komentar. Aturan yang digunakan pada aplikasi ini sesuai dengan persyaratan yang terdapat pada portal biodiversity information standards Darwin Core terms (<http://rs.tdwg.org/dwc/terms>). Tabel 1, 2 dan 3 merupakan beberapa aturan standar yang didefinisikan oleh Darwin Core yang diimplementasi ke dalam web site ini beserta penjelasannya



Gambar 1. Mekanisme kerja GRDDL

Tabel 1. Family

Term Name	Family
Identifier	http://rs.tdwg.org/dwc/terms/family
Label	Family
Definition	The full scientific name of the family in which the taxon is classified.
Namespace	http://digir.net/schema/conceptual/darwin/2003/1.0/

Tabel 2. Genus

Term name	genus
Identifier	http://rs.tdwg.org/dwc/terms/genus
Label	Genus
Definition	The full scientific name of the genus in which the taxon is classified
Namespace	http://digir.net/schema/conceptual/darwin/2003/1.0/

Tabel 3. Species

Term name	Species
Identifier	http://digir.net/schema/conceptual/darwin/2003/1.0/Species
Label	Species
Definition	The specific epithet of the organism
Namespace	http://digir.net/schema/conceptual/darwin/2003/1.0/

```
<header profile="http://www.w3.org/2003/g/data-view">
  <link rel="transformation" href="http://microdata.dev/app/services/xsl/specimen.xsl"/>
</header>
```

Gambar 2. Attribute profile

HASIL DAN PEMBAHASAN

Teknik GRDDL untuk mengekstrak data pada web semantik adalah dengan memanfaatkan format dokumen XML khususnya XHTML atau microformat. Untuk membuat dokumen XHTML terdapat beberapa penambahan microdata pada dokumen HTML dimana dokumen HTML ini adalah dasar dari pembuatan web semantik ini. Pada dokumen HTML yang dibuat terdapat atribut *profile* dan *transformation* yang menunjukkan alamat referensi transformasi dari dokumen tersebut dan diletakkan pada tag header HTML. Adapun atribut *profile* dan *transformation* yang dimaksud adalah seperti gambar 1.

Penggalan script diatas menunjukkan pada tag header dokumen ini terdapat atribut *profile* yang mereferensi ke alamat <http://www.w3.org/2003/g/data-view>. Alamat referensi ini merupakan profil metadata untuk *data-view* pada

dokumen ini yang disesuaikan dengan aturan penulisan HTML sedangkan untuk alamat transformasinya merujuk pada file *specimen.xsl*. Gambar 2 adalah contoh sintaks dokumen HTML dengan penambahan atribut.

Pada skrip HTML diatas terdapat sebuah tag `` yang diikuti oleh metadatanya untuk menjelaskan arti dari data tersebut, misalnya angka 10227210 adalah data *taxonID*. inilah yang membuat data dalam dokumen XHTML tersebut bisa ditransformasikan kedalam bentuk format RDF. Penamaan atribut pada tag `` diatas bisa bermacam-macam tidak ada batasan nama yang digunakan untuk selektor tersebut namun dari sisi sintaks, HTML5 sudah menyediakan beberapa atribut baku yang biasa digunakan secara umum diantaranya *class*, *id*, *title*.

Sampai pada tahap ini halaman yang sebelumnya hanya bersifat HTML sudah berubah menjadi dokumen XHT-

ML, dokumen inilah yang digunakan pada web semantik biodiversitas ini. Dengan format inilah teknik GRDDL bisa dilakukan untuk mengekstrak data.

Desain Halaman XHTML

Web semantik ini terdapat 2 contoh halaman XHTML yang valid dan bisa diakses untuk dilakukan ekstrak data. Secara rinci akan dijelaskan secara terpisah. **Daftar Specimen.** Tampilan antarmuka dari halaman daftar specimen seperti yang ada pada gambar 3 dengan bantuan plugin datatables untuk menampilkan datanya dalam bentuk tabel. Pada halaman ini terdapat 4 kolom informasi mengenai data specimen. Kolom ID Specimen digunakan untuk menampilkan informasi ID unik dari species, selanjutnya kolom Species merupakan nama dari

species diikuti oleh kolom Family yang digunakan untuk menampilkan informasi nama Family dari species tersebut serta kolom Genus yang menginformasikan nama Genus dari species tersebut. Dari sisi tampilan halaman ini tidak berbeda dengan halaman HTML biasa yang bisa diakses dari browser namun sebenarnya ini adalah halaman XHTML. Dibalik tampilan sederhana tersebut terdapat metadata pada setiap elemennya. Metadata tersebut hanya bisa dilihat ketika kode sumbernya dibuka. Setiap data specimen yang terdapat pada halaman ini sudah diberi atribut untuk memudahkan proses transformasi data ke dalam format RDF. Adapun cuplikan kode sumber pembentuk halaman tersebut adalah seperti terlihat code pada gambar 4.

```
<span property="dc:taxonID">10227210</span>
```

Gambar 3. Contoh Sintaks Specimen

ID Specimen	Species	Family	Genus
10000047	beckingii	ARAUCARIACEAE	Agathis
10000047	celebica	ARAUCARIACEAE	Agathis
10000048	beckingii	ARAUCARIACEAE	Agathis
10000048	rumphii	ARAUCARIACEAE	Agathis
10000048	philippinensis	ARAUCARIACEAE	Agathis
10000049	philippinensis	ARAUCARIACEAE	Agathis
10226948	cordifolia	THEACEAE	Adinandra
10226950	cordifolia	THEACEAE	Adinandra
10227023	cordifolia	THEACEAE	Adinandra
10227024	cordifolia	THEACEAE	Adinandra

Gambar 3. Halaman Daftar Specimen

```
<tr class="taxonlist" href="http://microdata.dev/specimen/detail/id/10000072">
  <td><span class="taxonID">10227210</span></td>
  <td>
    <a href="http://microdata.dev/specimen/detail/id/10000072">
      <span class="species">virilis</span></a>
    </td>
  <td><span class="family">ACANTHACEAE</span></td>
  <td><span class="genus">Acanthea</span></td>
</tr>
```

Gambar 4. Proses transformasi data ke format RDF

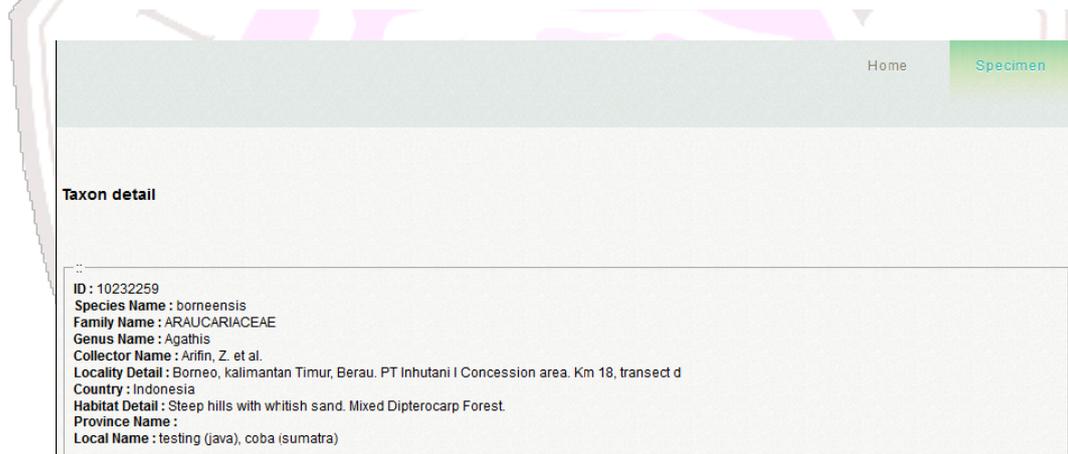
Skrip pada gambar 4 merupakan contoh dari 1 record specimen yang ada pada halaman daftar specimen, record tersebut dibuat dalam bentuk tabel yang ditandai dengan adanya tag <tr> dan <td> yang menunjukkan table row dan table column. Pada tag <tr> terdapat metadata taxonlist yang menjelaskan bahwa baris/section ini merupakan data taxon dan merferensi ke alamat yang tercantum pada atribut href-nya. Kemudian pada bagian kolom atau tag <td> mendeskripsikan data berupa taxonID, species, family dan genus dari specimen tersebut.

Detail Specimen

Melanjutkan dari halaman daftar specimen sebelumnya, terdapat tautan dari

setiap record data spesies yang ada di daftar menuju pada halaman detail dari species tersebut untuk melihat informasi yang lebih rinci seperti yang terlihat pada gambar 5.

Pada gambar diatas terdapat informasi lengkap dari salah satu species dimulai dari ID Spesiesnya (ID) hingga dengan Nama Lokal (Local Name) dari species tersebut. Tidak berbeda jauh dengan halaman daftar species sebelumnya, halaman ini juga merupakan halaman web dengan format XHTML yang sudah dilengkapi dengan metadata pada elemennya. Adapun cuplikan kode sumber halaman ini adalah seperti pada gambar 6.



Gambar 5. Halaman Detail Specimen

```
<div class="taxon">
  <span class="taxonID">10232259</span>
  <span class="specificEpithet">borneensis</span>
  <span class="family">ARAUCARIACEAE</span>
</div>
```

Gambar 6. Data Taxon

Penggalan kode HTML diatas menggambarkan data taxon yang terdiri dari taxonID, species, dan family. Tag <div> digunakan sebagai parent elemen dengan isi atribut adalah taxon, kemudian didalam tag ini terdapat tag yang menjelaskan data taxonID, species dan

family dari specimen tersebut. Atribut yang digunakan pada kode sumber diatas berupa class.

XSLT Template

Template XSL sangat berperan penting dalam hal transformasi data dari

dokumen XHTML yang sudah diberi atribut pada kontennya ke dalam format dokumen RDF. Dokumen ini berbentuk XML, umumnya dokumen ini digunakan untuk mentransformasikan dokumen XML ke dokumen XHTML atau XML ke dokumen lain namun pada penelitian ini digunakan logika terbalik dalam mendesain dokumen ini yaitu membalik fungsi transformasi data dari XHTML ke dalam bentuk dokumen XML. Struktur file ini akan dijelaskan sesuai dengan bagiannya

Header

Pada bagian header terdapat beberapa definisi format yang akan digunakan pada dokumen ini, seperti encoding, namespace hingga dengan output. Berikut adalah skrip header yang ada pada dokumen ini.

Dokumen ini menggunakan format XML dengan versi 1.0 dan encoding UTF-8, kemudian terdapat deklarasi dari xsl namespace referensi yaitu xmlns:xsl, xmlns:rdf, xmlns:dc, dan xmlns:h. Masing-masing namespace ini mengacu pada alamat yang berbeda beda sesuai dengan definisinya, salah satu contoh misalnya pada namespace xmlns:xsl mengacu pada <http://www.w3.org/1999>

[/XSL/Transform](#) sebagai aturan transformasinya

Parameter

Terdapat beberapa parameter yang digunakan dalam dokumen ini antara lain TaxonNS, about, source dan list seperti terlihat pada penggalan script yang tertera pada gambar 8.

Salah satu parameter yang digunakan pada dokumen ini adalah TaxonNS yaitu namespace untuk definisi dari taxon yang mereferensi pada situs Darwin Core yang merupakan lembaga standardisasi untuk keanekaragaman hayati. Penggunaan atribut yang merujuk pada data specimen di dalam dokumen ini akan mengikuti aturan yang sudah ditentukan oleh Darwin Core sesuai dengan alamat url yang tercantum pada bagian namespace tersebut.

Template

Bagian ini digunakan untuk mendesain struktur RDF data specimen. Data specimen tersebut bersumber dari halaman XHTML sebelumnya yang kemudian ditransformasi dengan dokumen ini untuk diekstrak datanya, adapun template yang dimaksud adalah seperti pada contoh code program pada gambar 9.

```
<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc="http://rs.tdwg.org/dwc/terms/"
xmlns:h="http://www.w3.org/1999/xhtml" version="1.0"
>
<xsl:output indent="yes" method="xml" encoding="UTF-8"/>
```

Gambar 7. Format penulisan header

```
<xsl:param name="TaxonNS">http://rs.tdwg.org/dwc/terms/</xsl:param>
<xsl:param name="XdtNS">http://www.w3.org/2001/XMLSchema#</xsl:param>
<xsl:param name="about">
  <xsl:value-of select='h:html/h:header/h:base/@href' />
</xsl:param>
```

Gambar 8. Parameter Taxon NS

```

<xsl:call-template name="textProp">
  <xsl:with-param name="property">taxonID</xsl:with-param>
</xsl:call-template>

```

Gambar 9 Dokumen XSL untuk Desain Struktur Data Spesimen

```

public function SpecimenDetail($specimenid)
{
    $basedomain = "http://microdata.dev/";
    $urlData = $basedomain. 'specimen/detail/id/'. $specimenid;
    $urlxsl = $basedomain. 'xls/single-specimen.xsl';
    $XML = new DOMDocument();
    $XML->load($urlData);

    # START XSLT
    $xslt = new XSLTProcessor();
    $XSL = new DOMDocument();
    $XSL->load($urlxsl, LIBXML_NOCDATA);
    $xslt->importStylesheet( $XSL );

    return $xslt->transformToXML( $XML );
}

```

Gambar 10 Scriptk untuk generate dokumen RDF

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```

- <rdf:RDF>
- <rdf:Description>
  <title value="Microdata Specimen">Microdata</title>
</rdf:Description>
- <rdf:Description rdf:about="http://localhost/microdata/specimen">
- <rdf:Description rdf:about="http://localhost/microdata/specimen/detail/id/10000072">
  <dc:taxonID xml:lang="en">10227210</dc:taxonID>
  <dc:scientificName xml:lang="en">virilis</dc:scientificName>
  <dc:genus xml:lang="en">Acanthea</dc:genus>
  <dc:family xml:lang="en">ACANTHACEAE</dc:family>
</rdf:Description>
- <rdf:Description rdf:about="http://localhost/microdata/specimen/detail/id/10000634">
  <dc:taxonID xml:lang="en">10226948</dc:taxonID>
  <dc:scientificName xml:lang="en">cordifolia</dc:scientificName>
  <dc:genus xml:lang="en">Adinandra</dc:genus>
  <dc:family xml:lang="en">THEACEAE</dc:family>
</rdf:Description>
</rdf:Description>
</rdf:RDF>

```

Gambar 11 Dokumen RDF

Penggalan kode program diatas menunjukkan bagian dari dokumen XSL untuk melakukan pemanggilan section pada halaman XHTML dengan nama atribut

textProp. Di dalam section tersebut ada pencarian atribut lagi dengan nama property yang isinya adalah taxonID. Maka secara otomatis data dari taxon ID yang

terdapat pada halaman XHTML sebelumnya bisa didapatkan dengan menggunakan perintah diatas. Begitulah dokumen ini melakukan transformasi data dari halaman XHTML.

Generate RDF

Untuk menguji apakah dokumen XHTML tersebut bisa ditransformasikan oleh dokumen XSL sesuai dengan yang diharapkan maka dibutuhkan sebuah bahasa pemrograman yang bisa menjalankan proses ekstraksi kedua dokumen tersebut secara paralel. Adapun bahasa pemrograman yang digunakan pada penelitian ini menggunakan PHP. Untuk menggenerate dokumen RDF dibutuhkan sebuah perintah spesifik pada bahasa pemrograman PHP yaitu XSL Processor. Dengan menggunakan perintah XSL Processor web server akan menjalankan perintah untuk melakukan transformasi data dari halaman XHTML kedalam format dokumen RDF dalam bentuk XML.

Adapun perintah yang digunakan untuk menggenerate dokumen ini hampir sama antara halaman daftar specimen dan detail specimen, perbedaannya hanya terdapat pada link halamannya. Berikut ini adalah contoh script yang digunakan untuk menggenerate dokumen RDF pada halaman XHTML detail specimen Blok program diatas merupakan sebuah fungsi dengan 1 parameter yaitu specimenid.

Parameter ini digunakan oleh perintah selanjutnya untuk menunjuk halaman XHTML dari specimen tersebut seperti yang ditunjukkan oleh variabel urlData. Halaman specimen tersebut kemudian di load dengan menggunakan fungsi PHP DOMDocument. Perintah selanjutnya adalah mengimport file XSL yang sudah dibuat sebelumnya untuk bisa diproses bersamaan dengan dokumen XHTML. Setelah kedua dokumen tersebut berhasil di load langkah selanjutnya adalah melakukan proses transformasi dengan mengeksekusi perintah `$xslt->transformToXML($XML)` maka proses ekstrak da-

ta dari halaman XHTML yang disesuaikan dengan template XSL akan dieksekusi. Hasil dari perintah tersebut akan menghasilkan dokumen RDF seperti yang terlihat pada gambar 11. Terlihat bahwa sebuah file RDF sudah terbentuk setelah melewati proses ekstraksi data dengan menggunakan pemrograman PHP, format dokumen inilah yang menjadi tujuan dari penelitian ini. Pada dokumen tersebut terdapat judul dokumen hingga dengan data dari specimen yang berhasil diekstrak yang dibedakan berdasarkan tag description.

SIMPULAN DAN SARAN

Distribusi data tumbuhan yang ada di Indonesia secara lebih luas sangat mungkin dilakukan dengan adanya web semantik, dikarenakan faktor fleksibilitas dari web ini membuat proses konsumsi data bisa lebih maksimal. Dibandingkan dengan website biasa yang menampilkan informasi yang hanya dimengerti oleh manusia, web semantik melakukan lebih dari itu yaitu memungkinkan mesin bisa memahami isi dari website tersebut. Kuncinya adalah pemanfaatan dokumen XHTML untuk menyimpan informasi dari specimen. Namun yang menjadi tantangan adalah bagaimana membuat dokumen XHTML yang sesuai agar proses ekstraksi data bisa lebih mudah, hal yang harus diperhatikan kedepannya adalah mendefinisikan standar penggunaan atribut microdata pada dokumen XHTML agar nantinya hanya terdapat 1 template umum XSL untuk melakukan semua transformasi dokumen XHTML.

DAFTAR PUSTAKA

- [1] Gleaning Resource Descriptions from Dialects of Languages (GRDDL). Available <http://ids.snu.ac.kr/w/images/b/bd/SC05.pdf>
- [2] Gruber, T. Toward Principles for the Design of Ontologies Used for

- KnowledgeSharing.<http://tomgruber.org/writing/onto-design.pdf>
- [3] Cahyono, E. B. Modul XSLT untuk OpenCMS. Lembaga Penelitian Universitas Muhammadiyah Malang
- [4] Halpin, H. Procedural versus declarative semantics on the web: Microformats, grddl, and rdfa.
- [5] Semantic Extensions to Defuddle: Inserting GRDDL into XML <https://pdfs.semanticscholar.org/c1fe/bd6666f2fd4d3b0742db462316035ec70cb2.pdf>
- [6] Mapping between RDF and XML with XSPARQL . <https://link.springer.com/content/pdf/10.1007%2Fs13740-012-0008-7.pdf>
- [7] Yu, L. (2011). A Developer's Guide to the Semantic Web. Springer.
- [8] Herman, I. (2006). Swtutorial. [Online]. Available: <http://www.w3.org/People/Ivan>

