

# Universiteti i Evropës Juglindore

Fakulteti i shkencave dhe teknologjive Bashkëkohore

Tema:

TETOVË 2015

# Informatat e shteteve duke përdorur DBpedia

Prof. Xhemal Zenuni

Studenti: Ylber Xhambazi

# Përmbajtja

Abstrakti	2
Hyrje	3
Kërkesat, implementimi dhe instalimi	
SPARQL	
EasyRDF	6
Linked Data	
DBpedia	
Ndërtimi i aplikacionit	10
Përfundimi	12
Lista e figurave	15
Bibliografia	16

### Abstrakti

Ky projekt paraqet ndërtimin e një ueb aplikacioni për gjenerimin e profileve të shteteve nga të dhëna të ndryshme ku si burim kryesorë konsiderohet Linked Open Data me RDF serializime te ndryshme. Kryesisht janë përdorur teknikat dhe teknologjitë e ueb – it semantik për përdorimin dhe vizualizimin e këtyre të dhënave. Me përdorimin e SPARQL kemi arritur që nga datasete të ndryshme të gërshetojmë informacione në baze të disa kritereve. DBpedia është një përpjekje për nxjerrjen e informacioneve të strukturuara nga Wikipedia dhe për të bërë këtë informacion në dispozicion të ueb – it. Në DBpedia përshkruhet nxjerrja e dataseteve të ndryshme dhe si informacion rezulton se është publikuar në ueb për konsumin me qenie njerëzore dhe makinë. Së fundmi ne paraqesim statusin aktual të interlinking DBpedia me datasete te hapura ne ueb dhe që përshkruajnë se si DBpedia mund të shërbejë si një bërthamë për zhvillimin e të dhënave të hapura ne ueb.

## Hyrje

Në kuadër të përpjekjeve të W3C Linked Open Data (LOD) një numër vazhdimisht ne rritje i të dhënave kanë filluar të publikohen në ueb dhe interlinking data sipas parimeve të krijuesit të internetit Tim Berners – Lee. Ueb – i i të dhënave aktualisht përbëhet nga miliarda treshe RDF dhe mbulon fusha të tilla si informacionet gjeografike, njerëzit, kompanitë, komunitetet online, filma, muzikë, libra dhe publikime shkencore. Përveç publikimit dhe interlinking datasete DBpedia punon edhe me Linked Data shfletues (ang. browsers), Linked Data crawlers, ueb i të dhënave në motorët e kërkimit (ang. search engines), dhe shumë aplikacione që përdorin Linked Data nga ueb.

DBpedia RDF i dataseteve ku mbahen dhe publikohen duke përdorur OpenLink Virtuso ku Virtuso ka infrastrukturë që siguron qasje në të dhënat e DBpedia RDF nëpërmjet një SPARQL endpoint që së bashku me mbështetjen e HTTP për standardin e çdo ueb klienti ku merr si html ose rdf përfaqësime që kanë burimet e DBpedias<sup>1</sup>.

Ky projekt përbehet nga baza e DBpedia që funksionon me llojin e pyetësorit (ang. query) që

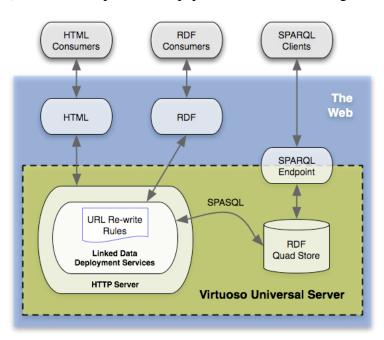


Figure 1 Arkitektura e DBpedia

quhet SPARQL dhe nga këtu nxirren te dhënat me anë te pyetjeve që bëhen ne SPARQL, edhe aty ka një lloj të gjuhës si që është SQL te baza e të dhënave por duhet që të shihen datasetet se nga ku e kërkon atë të dhënë që të duhet.

Në të ardhmen pritet që të ketë një funksionalitet shumë më të lart se sa që ka sot sepse si është duke shkuar ecuria e internetit shihet që e gjitha baza e të dhënave do të zhvendoset ne internet dhe nëpërmjet pyetësorëve që na i ofrojnë llojet e bazave të ndryshme edhe pyetësorët që i kanë të ndryshëm ua lehtësojnë punën përdoruesve që me disa pyetësorë të gjej at që kërkon.

<sup>&</sup>lt;sup>1</sup> http://www.informatik.uni-leipzig.de/~auer/publication/dbpedia.pdf

# Kërkesat, implementimi dhe instalimi

EasyRdf përmban të gjitha karakteristikat e tij te shkruar ne bibliotekën e PHP dhe kërkon që të jetë i instaluar PHP 5.2, 5.3 dhe 5.4 ose versionet më të reja që do të dalin ne vazhdim. Që të funksionojë EasyRdf po ashtu njëlloj si PHP si funksionon nëpërmjet xamp ose wamp serverit dhe apache dhe mund të gjendet në appache friends<sup>2</sup>.

Për të punuar në EasyRdf duhet që të kemi të instaluar XAMPP dhe disa file prej te që duhen shkarkuar prej te EasyRdf<sup>3</sup> versionin më të ri dhe duhet që ta lejmë te skedari (ang. folder) i XAMPP që gjendet në diskun C:\xampp për të funksionuar duhet që në të njëjtin skedar (ang. folder) që të bëhet shkarkimi (ang. download) edhe SPARQL realisht është ARC2<sup>4</sup>.

Këto të dyja i bëjmë bashkë në një skedar (ang. folder) por prapë është problem se duhet që të rregullohet path-i që bënë shumë problem sepse fajllat (ang. files) që ekzekutohen në htdocs duhet që të u tregohet rruga se ku i kanë libraritë, path – i është ky më poshtë.

```
set_include_path("C:/xampp/arc2/easyrdf/lib");
    require_once "EasyRdf.php";
    include ('C:\xampp\arc2\easyrdf\examples\html_tag_helpers.php');
Me këtë që shpjegova më lartë tregoj se si funksionon rdf dhe se si mund të shfaqet rezultati që kërkojmë.
```

Ndryshe nga kjo që shpjegova më lartë ka edhe një mundësi tjetër për të funksionuar EasyRDF ku duhet pjesën që e marrim nga EasyRDF d.m.th skedarin që e marrim duhet që aty ta punojmë edhe projektin ku edhe path apo vendosja se prej ku merren file të EasyRDF duket kështu

```
set_include_path(get_include_path() . PATH_SEPARATOR . '../lib/');
require_once "EasyRdf.php";
require_once "html_tag_helpers.php";
```

<sup>&</sup>lt;sup>2</sup> http://www.apachefriends.org/en/xampp-windows.html

<sup>&</sup>lt;sup>3</sup> http://www.easyrdf.org/downloads

<sup>&</sup>lt;sup>4</sup> https://github.com/semsol/arc2/wiki

### **SPARQL**

Shumë njerëz pyesin se çfarë mund të bëhet me SPARQL që nuk mund të bëhet me SQL, kur në të vërtet ata kujdesen për atë çfarë mund të bëhet ne RDF që nuk mund të bëhet me databazën relacionare. Të dyja këto gjuhë japin akses te përdoruesit për të krijuar, kombinuar dhe të konsumojnë të dhëna të strukturuara. SQL e bën këtë duke hyrë në tabelat në bazat e të dhënave relacionale kurse SPARQL e bën këtë duke hyrë në një rrjetë të Linked Data por SPARQL është projektuar që të bashkojë burimet e ndryshme në të dhënat.

Të dyja gjuhët janë përpjekur për të pasur më shumë aftësi të ngjashme.

- SQL dhe SPARQL kanë një grup të ngjashëm të funksioneve agregate të përdorura nga operatori GROUP BY
- Për zgjedhjen e të dhënave ku disa vlera nuk ekzistojnë, SQL përdor vlera null si DIAGNOSIS.name = NULL. Për zgjedhjen e të dhënave ku vlerat nuk përputhen SQL përdor NOT EQUAL. SPARQL përdor NOT EXISTS filtër për qëllime të ngjashme.
- SQL përdor LEFT JOIN OUTER kurse SPARQL përdor fjalën OPTIONAL për të treguar se zgjidhjet nuk duhet të eliminohen edhe pse disa prej tyre nuk janë të pranishme në të dhënat ku kërkohen<sup>5</sup>.

Gjithnjë e më shumë është duke u përdorur SPARQL (ang. SPARQL Protocol and RDF Query Language) që është një pyetësor (ang. Query) që shërben për të tërhequr të dhëna nga një koleksion në rritje të të dhënave publike dhe private. Drejtori i W3C apo shpikësi i internetit Tim Berners – Lee duke u përpjekur për të përdorur ueb semantik pa SPARQL është njëjtë si të përdorësh një bazë të dhënash pa sql.

SPARQL nuk është projektuar për të qenë pyetësor te të dhënat relacionale por është projektuar për të qenë pyetësor te modeli i të dhënave të RDF (ang. Resource Description Framework).

Në SPARQL një procesor mund të jetë një program I pavarur i ndërtuar në një rezervë të trefishtë ose pjesë e një shtrese të middleware që ju lejon të dërgoni pyetjet ne SPARQL për një grup të dhënash që nuk mund të pranojnë ndryshe. Procesori ARQ është i dobishëm kur kemi

5

<sup>&</sup>lt;sup>5</sup> http://www.topquadrant.com/2014/05/05/comparing-sparql-with-sql/

mësuar gjuhën SPARQL, sepse më shpejt që japim një pyetje një fotografi të të dhënave ose diçka tjetër do të shohim rezultat shumë shpejt<sup>6</sup>.

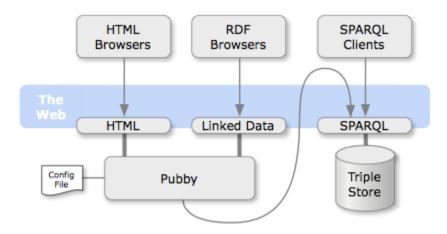


Figure 2 Arkitektura e SPARQL

## **EasyRDF**

Mund të quhet si Ueb 3.0, të dhënat e uebit ose Ueb Semantik, i përket një gjenerate tjetër që është Ueb aplikacion dhe do të duhet të mbështetet në burimet e shpërndara të informacionit. Është një gjuhë skriptuese si PHP që ka qenë gjithmonë një zgjedhje e mirë për mjedise dinamike dhe detyrave integruese. PHP ndihmon të zvogëlojë kohën e zbatimit dhe shpesh luan një rol qëndror duke bërë përparime në sisteme të sofistikuara që e bën atë të përshtatshme për projekte në ueb semantik<sup>7</sup>.

EasyRdf është një bibliotekë e PHP e projektuar për të bërë të lehtë për të konsumuar dhe për të prodhuar RDF. RDF ishte projektuar për përdorimin në ekipe të përziera të zhvilluesit e me përvojë dhe pa përvojë. Ishte e shkruar në objektet e orientuara në php dhe është testuar gjerësisht duke përdorur PHPUnit.

Pas analizimit të EasyRdf ndërton një grafik të objekteve php të cilat pastaj mund të ecin rreth për të marrë të dhëna për t'u vendosur në faqe. Metodat lëshojnë dhe tregojnë që janë në dispozicion për të inspektuar sepse të dhënat janë gjatë zhvillimit. Të dhënave u është ngarkuar në

<sup>&</sup>lt;sup>6</sup> Bob DuCharme, Learnin SPARQL (Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopolm CA 95472, 2011), Page 19-29.

<sup>&</sup>lt;sup>7</sup> http://www.eventbrite.com/e/workshop-easy-rdf-and-sparql-for-lamp-systems-building-semantic-web-apps-on-php-mysql-tickets-333851559

mënyrë tipike një objekt i EasyRdf\_Graph nga dokumentacioni burimor i RDF i ngarkuar nga ueb – i nëpërmjet HTTP. Klasa EasyRdf\_GraphStore thjeshton ngarkimin dhe ruajtjen e të dhënave në një SPARQL. Pyetje SPARQL mund të bëhen mbi HTTP për një Triplestore duke përdprur klasën EasyRdf\_Sparql\_Client. SELECT dhe ASK pyetje do të kthehen në një objekt EasyRdf\_Sparql\_Result dhe CONSTRUCT dhe DESCRIBE pyetje do të kthehen në një objekt EasyRdf\_Graph<sup>8</sup>.

#### Linked Data

Ideja e Linked Data qëndron në atë se është me e re se ajo e ueb – it semantik, por ndonjëherë është më e lehtë për të menduar se si u ndërtua ueb semantik prapa ideve të Linked Data. Linked Data nuk është një specifikim por një grup i praktikave më të mira për të siguruar një infrastrukturë të dhënave që e bën më të lehtë për të ndarë të dhënat në të gjithë ueb – in. Pastaj mund të përdoren teknologjitë në ueb – in semantik të tilla si RDFS, OWL dhe SPARQL për të ndërtuar aplikime rreth atyre të dhënave.

Një rast tipik i një Linked dataset të madhe është DBpedia që në thelb përmban Wikipedia në dispozicion me RDF. Rëndësia e DBpedia nuk është vetëm se ajo përfshin të dhëna nga Wikipedia por edhe se ajo përfshin lidhje të dhënash të tjera ne ueb si p.sh, në Geonames etj etj. Duke ofruar këto lidhje shtesë (në aspektin e RDF tresheve) aplikacionet mund të shfrytëzojnë shtesë njohuri nga datasets të tjera kur zhvillimin e një aplikimi duke u mbështetur në fakte të integruara nga disa datasets<sup>9</sup>.

<sup>&</sup>lt;sup>8</sup> https://github.com/njh/easyrdf

<sup>&</sup>lt;sup>9</sup> http://www.w3.org/standards/semanticweb/data

# **DBpedia**

DBpedia është një projekt që synon të nxjerrë përmbajtjen e strukturuar nga informacioni i krijuar si pjesë e projektit Wikipedia. DBpedia është si një lloj pasqyre e wikipedias, wikidata në ueb semantik. Ky informacion i strukturuar është bërë atëherë në dispozicion të WWW. DBpedia lejon që përdoruesit me pyetjet e drejtuara nga marrëdhëniet dhe vetit që lidhen me burimet e Wikipedia, duke përfshirë edhe lidhjet me datasets të tjera të ngjajshme. DBpedia është përshkruar nga Tim Berners – Lee si një nga pjesët më të famshme të përpjekjeve të decentralizuara Linked Data. Projekti ka filluar nga njerëzit në Universitetin e Berlinit dhe Universitetin e Lajpcigut, në bashkëpunim me OpenLink Software, dhe dataset në dispozicion të parë të publikut u botua ne vitin 2007 dhe prapavija e DBpedia ishte si ne figurën më poshtë.

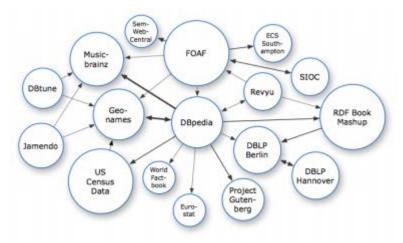


Figure 3 DBpedia në vitin 2007

Artikujt në Wikipedia përbëhen kryesisht prej tekstit të lirë por përfshijnë edhe informacione të strukturuar ngulitur në artikuj. Disa kategori të informacionit janë: kategorizimi, imazhet, gjeo – koordinimet dhe lidhjet me ueb faqet e tjera.

DBpedia është pjesë e projektit të komunitetit W3C Linked Open Data, ku një përpjekje për të publikuar dhe Interlink burime të ndryshme të dhënave të hapura. Që nga shtatori i 2011 kjo përpjekje ka ndërtuar një rrjet të burimeve të të dhënave të ndërlidhura (ang. interlinking) që arrin në më shumë se 31 miliardë RDF treshe. Si wikipedia që përmban artikuj për shumë koncepte me qëllim të përgjithshëm edhe DBpedia gjithashtu mund të shihet si një ontologji e madhe që ngarkon URI në shumë koncepte dhe mbështet këto URI me përshkrime dereferenceable RDF. Nëse ju keni

një profil të FOAF dhe keni nevojë për kushte për përshkrimin e interesave të juaja ose vendndodhjen tuaj ju mund të konsideroni përdorimin e DBpedia URI – ve. Kjo do të lejojë shfletuesit RDF si një lloj i përzier që hedh në tabelë ose ueb shfletues OpenLink i të dhënave për të shfletuar nga profili juaj foaf në DBpedia<sup>10</sup>.

Kurse në ditët e sotme është zhvilluar tepër DBpedia dhe ka miliona informata në të gjitha drejtimet, pritet të përmirësohet edhe më shumë ne vitet ne vazhdim kurse në vitin 2014 është zhvilluar tepër si ne fotografinë më poshtë<sup>11</sup>.

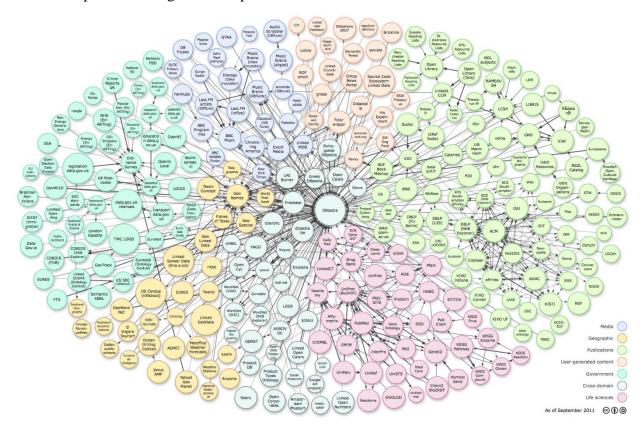


Figure 4 DBpedia në vitin 2014

9

 $<sup>^{10}\,\</sup>underline{\text{http://wiki.dbpedia.org/services-resources/interlinking}}$ 

<sup>11</sup> http://en.wikipedia.org/wiki/DBpedia

# Ndërtimi i aplikacionit

Karakteristikat dhe definicionet që i kam përmendur më lartë tash do të sqaroj pjesë të kodit me poshtë se si kam ardhur deri te ky përfundim.

Për të përdorur SPARQL dhe RDF duhet qe te përdoreshin nëpërmjet PHP, në fillim shkova prej EasyRDF dhe morra shembullin se si të përdor RDF edhe për të shfaqur të dhëna që i dua unë. RDF punon nëpërmjet prefikseve, këto të fundit shërbejnë se prej ku i merr të dhënat edhe i ekzekuton në rast mungese të prefiksit atëherë nuk gjen se ku do te lidhet edhe te nxjerr gabim sintaksorë. Prefikset janë pjesë e rëndësishme tjetër është edhe prej ku ekzekutohen prej kujt baze vijnë të dhënat dhe si ekzekutohen.

```
set_include_path(get_include_path() . PATH_SEPARATOR . '../lib/');
    require_once "EasyRdf.php";
    require_once "html_tag_helpers.php";

EasyRdf_Namespace::set('category', 'http://dbpedia.org/resource/Category:');
EasyRdf_Namespace::set('dbpedia', 'http://dbpedia.org/resource/');
EasyRdf_Namespace::set('dbpedia', 'http://dbpedia.org/ontology/');
EasyRdf_Namespace::set('dbp', 'http://dbpedia.org/property/');
EasyRdf_Namespace::set('dcterms', 'http://purl.org/dc/terms');
EasyRdf_Namespace::set('owl', 'http://www.w3.org/2002/07/wslth);
EasyRdf_Namespace::set('val', 'http://www.w3.org/2002/07/wslth);
EasyRdf_Namespace::set('rdfs', 'http://www.w3.org/2000/1/rdf-schema#');
EasyRdf_Namespace::set('rdf', 'http://www.w3.org/1999/02/22-rdf-syntax-ns#');
EasyRdf_Namespace::set('dcterms', 'http://purl.org/dc/terms/');
EasyRdf_Namespace::set('dcterms', 'http://purl.org/dc/terms/');

$sparql1 = new EasyRdf_Sparql_Client('http://dbpedia.org/sparql');
$sparq12 = new EasyRdf_Sparql_Client('http://dbpedia.org/sparql');
```

Figure 5 Tregon prefiksat se prej ku merren datasetet

```
<html>
<head>
 <title>dbpedia project</title>
 <meta http-equiv="content-type" content="text/html; charset=utf-8" />
</head>
<body>
<h1>List of the Countries</h1>
<form action="" method="get">
<select name="abc">
<?php
    $result = $sparql1->query(
        'SELECT * WHERE { '.
        ' ?country rdf:type dbo:Country .'.
        ' ?country rdfs:label ?label .'.
        ' ?country dc:subject category:Member_states_of_the_United_Nations .'.
        ' FILTER ( lang(?label) = "en" ) '.
        '} ORDER BY ?label'
   );
    foreach ($result as $row) {
       echo "<option value=".($row->label).">".($row->label)."</option>";
?>
</select>
<input button class="" name="button1" type="submit">
</form>
Total number of countries: <?= $result->numRows() ?>
```

Figure 6 Pyetësori nga ku nxirren të dhënat nga DBpedia

# List of the Countries



Total number of countries: 188

Figure 7 Lista e shteteve

```
<?php
error_reporting(0);
if(isset($_GET['abc'])) {
    $abc = $_GET['abc'];
    $result1 = $sparql1->query(
        'SELECT distinct ?ekrye ?long ?flamuri ?eparaja ?web ?lang ?desc ?popullsia WHERE { '.
        '?country rdf:type dbo:Country;'.
        'rdfs:label ?label;'.
        'dbo:capital ?krye.'.
            '?country dbo:thumbnail ?flamuri.'.
            '?country dbo:abstract ?desc.'.
            'OPTIONAL{?country dbo:currency ?paraja}'.
            'OPTIONAL{?country dbp:populationCensus ?popullsia}'.
            'OPTIONAL{?paraja rdfs:label ?eparaja}'.
            'OPTIONAL{?country foaf:homepage ?web.}'.
            'OPTIONAL { ? country dbp:officialLanguages ? lang } '.
            '?country dbp:conventionalLongName ?long.'.
            '?country dcterms:subject category:'.$abc.'.'.
            '?krye rdfs:label ?ekrye.'.
        'FILTER(langMatches(lang(?ekrye), "EN"))'.
        'FILTER(langMatches(lang(?label), "EN"))'.
        'FILTER(langMatches(lang(?eparaja), "EN"))'.
        'FILTER(langMatches(lang(?desc), "EN"))'.
        '} ORDER BY ?label'
```

Figure 8 Kodi i inormatave te se cilit shtet

```
foreach ($result1 as $row) {

    echo "Name of the state:".($row->long)."";
    echo ""echo """echo """echo """"echo """"echo """"echo """""echo """""echo """""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""""<th
```

Figure 9 Informatat që nxirren në faqe

#### List of the Countries

Albania ▼ Submit

Total number of countries: 188

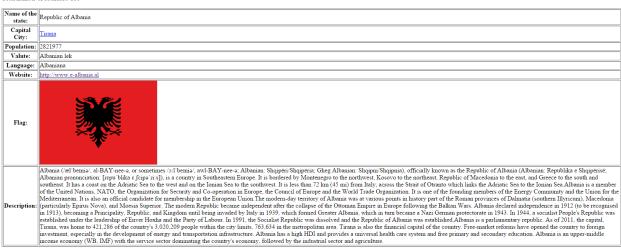


Figure 10 Komplet faqja

# Përfundimi

Me këtë ueb aplikacioni apo me këtë faqe u përpoqa për të përdorur çdo gjë që unë di në lidhje me semantikën. Krijimi i këtij ueb aplikacioni është një sfidë, por unë nuk mendoj qe do të ndalem këtu mendoj që do të përpiqem për të zhvilluar atë edhe më shumë se kaq sepse ky është qëllimi i semantikës që lidhet diçka me çdo gjë.

# Lista e figurave

Figure 1 Arkitektura e DBpedia	3
Figure 2 Arkitektura e SPARQL	
Figure 3 DBpedia në vitin 2007	
Figure 4 DBpedia në vitin 2014	
Figure 5 Tregon prefiksat se prej ku merren datasetet	
Figure 6 Pyetësori nga ku nxirren të dhënat nga DBpedia	
Figure 7 Lista e shteteve	11
Figure 8 Kodi i inormatave te se cilit shtet	12
Figure 9 Informatat që nxirren në faqe	12
Figure 10 Komplet faqja	13

# Bibliografia

- 1. <a href="http://www.informatik.uni-leipzig.de/~auer/publication/dbpedia.pdf">http://www.informatik.uni-leipzig.de/~auer/publication/dbpedia.pdf</a>
- 2. <a href="http://www.topquadrant.com/2014/05/05/comparing-sparql-with-sql/">http://www.topquadrant.com/2014/05/05/comparing-sparql-with-sql/</a>
- 3. http://www.apachefriends.org/en/xampp-windows.html
- 4. http://www.easyrdf.org/downloads
- 5. https://github.com/semsol/arc2/wiki
- 6. Bob DuCharme, Learnin SPARQL (Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopolm CA 95472, 2011).
- 7. <a href="http://www.eventbrite.com/e/workshop-easy-rdf-and-sparql-for-lamp-systems-building-semantic-web-apps-on-php-mysql-tickets-333851559">http://www.eventbrite.com/e/workshop-easy-rdf-and-sparql-for-lamp-systems-building-semantic-web-apps-on-php-mysql-tickets-333851559</a>
- 8. <a href="https://github.com/njh/easyrdf">https://github.com/njh/easyrdf</a>
- 9. <a href="http://www.w3.org/standards/semanticweb/data">http://www.w3.org/standards/semanticweb/data</a>
- 10. http://en.wikipedia.org/wiki/DBpedia
- 11. http://wiki.dbpedia.org/services-resources/interlinking