Swoogle

search and metadata for the semantic web

Presented by ebiquity lab
University of Maryland Baltimore County

Partial research support was provided by DARPA contract F30602-00-0591 and by NSF by awards NSF-ITR-IIS-0326460 and NSF-ITR-IDM-0219649.
Motivation

- (Google + Web) has made us all smarter
- something similar is needed by people and software agents for finding information on the semantic web
Swoogle Demo

1. Find “Time” Ontology (Swoogle Search)
2. Digest “Time” Ontology
   - Document view
   - Term view
3. Find Term “Person” (Ontology Dictionary)
4. Digest Term “Person”
   - Term definition
   - Class-Property bond
5. Swoogle Statistics
   - Swoogle Today
6. Ontology Rank
   - Swoogle’s top 10
7. Swoogle Services
   - Swoogle Triple Shop
   - Swoogle Web Service
   - Swoogle RDF Directory
   - Swoogle Firefox Plug-in
Find “Time” Ontology

We can use a set of keywords to search ontology. For example, “time, before, after” are basic concepts for a “Time” ontology.

http://www.isi.edu/~cpan/damtime/time-entry.owl

http://cidoc.jisc.forth.gr/rdfs/cidoc_crm_v3.4.3.rdf

http://pervasive.semanticweb.org/ont/dev/time
Digest “Time” Ontology (document view)

http://www.isi.edu/~pan/damltime/time-entry.owl

Annotition
- Label: none
- Version: none
- Authors: unknown
- Topics: unknown
- Comment: An entry sub-ontology of time.
- Endorsers: none

Other properties
- Size in bytes: unknown
- cache: Cached: Original File N-Triples
- Swoogle URI ID: $6399

Relations with other semantic web documents
Imports: none
Imported by:
- http://ontology.ihmc.us/Acknowledgement.owl
- http://ontology.ihmc.us/ActionStatus.owl
- http://ontology.ihmc.us/ActivityStatus.owl
- http://ontology.ihmc.us/ActionStatus.owl

References:
- http://www.isi.edu/~pan/damltime/time-entry-case1.owl
- http://ontology.ihmc.us/ActionStatus.owl
- http://ontology.ihmc.us/ActivityStatus.owl

Referenced by:
- http://www.isi.edu/~pan/damltime/time-entry-case1.owl
- http://ontology.ihmc.us/ActionStatus.owl
- http://ontology.ihmc.us/ActivityStatus.owl

Basic properties
- Encoding: RDF
- Triples: 308
- Classes Defined: 17
- Instances Defined: 7
- weight: 308
- Swougle Rank Score: 0.653507
- Type: Ontology/Instance mixture (67%)

Referenced by:
- http://www.isi.edu/~pan/damltime/time-entry-case1.owl
- http://ontology.ihmc.us/ActionStatus.owl
- http://ontology.ihmc.us/ActivityStatus.owl

Extended by:
- http://ontology.ihmc.us/ActionStatus.owl
- http://ontology.ihmc.us/ActivityStatus.owl

See all
- Extends: none
- Mapped to: none
- Prior versions: none
- Newer versions: none
- Similar docs: EQ: http://www.isi.edu/~pan/damltime/time-entry.owl

See all
- Swougle Search ©

Copyright © 2004, eRigality Research Group, UMBC v2.0
Summary: Document Metadata

- Web document metadata
  - When/how discovered/fetched
  - Suffix of URL
  - Last modified time
  - Document size

- SWD metadata
  - Language features
    - OWL species
    - RDF encoding
  - Statistical features
    - # of Defined/used terms
    - # of Declared/used namespaces
    - Ontology Ratio
  - Ontology Rank

- Ontology annotation
  - Label
  - Version
  - Comment

- Relations
  - Links to other SWDs
    - Imported SWDs
    - Referenced SWDs
    - Extended SWDs
    - Prior version
  - Links to terms
    - Classes/properties defined
    - Classes/properties used
Document-Term Relation

http://www.cs.umbc.edu/~finin/foaf.rdf

rdf:type
foaf:Person

foaf:mbox
finin@umbc.edu

http://foo.com/foaf.rdf

rdf:type
foaf:Person

http://foo.com/foaf.rdf#finin

foaf:mbox
finin@umbc.edu

http://xmlns.com/foaf/1.0/

rdfs:subClassOf
foaf:Person

wordNet:Agent

rdfs:Class

foaf:mbox

rdf:Property

rdfs:domain

rdfs:Class

rdfs:subClassOf

defined Class

defined Property

defined Individual

populated Class

populated Property

Swoogle

ebiquity lab, CSEE, UMBC
Find Term “Person”

Ontology Dictionary: Term Lookup

- Person
- Prefix
- Suffix
- Fuzzy

Ontology Dictionary: Term Lookup

- 1 of total 376 results for person in 0.0087 seconds
- URI: http://xmlns.com/foaf/0.1/person
- Type: Family, rdfs:owl
- Defined by 16 ontologies
- Populated by 51626 documents and 954433 class instances

Not capitalized! URIref is case sensitive!
Hence, C-P bonds will emerge from instance data instead of ontology definition.
Summary: Term Metadata

**C-P bond (from SWO)**
- foaf:mbox
- foaf:name

**C-P bond (from SWI)**
- foaf:name
- dc:title

**Onto 1**
- foaf:mbox
- foaf:name
- rdfs:domain

**Onto 2**
- rdf:type
- owl:Class
- rdfs:label
- “Person”

**SWD3**
- rdf:type
- foaf:name
- “Tim Finin”

**Term Definition**
- rdfs:subClassOf -- foaf:Agent
- rdfs:label – “Person”
Swoogle Statistics

- Summarize the dataset collected by Swoogle
  - Swoogle Watch
    - Swoogle Today
    - Distribution of visited URLs
    - Document discovery log
    - Term discovery log
  - Semantic Web Watch
    - SWD distribution by last-modified month
    - SWD distribution by website
    - SWD distribution by suffix
  - Ontology Watch
    - Term (class/property) usage
    - Namespace usage
### SWD per Website (Top 500 website)

(Generated at: Thu, 6 Jan 2005 12:49:49 -0500)

There are **130882** websites that hosting **336456** SWDs

<table>
<thead>
<tr>
<th>rank</th>
<th>host name</th>
<th>swd count</th>
<th>swd percentage</th>
<th>google estimation</th>
<th>nswd count</th>
<th>nswd/swd</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>blog.livedoor.jp</td>
<td>15599</td>
<td>4.64 %</td>
<td>73000</td>
<td>2589</td>
<td>0.1660</td>
</tr>
<tr>
<td>2</td>
<td><a href="http://www.ecademy.com">www.ecademy.com</a></td>
<td>10052</td>
<td>2.99 %</td>
<td>0</td>
<td>82</td>
<td>0.0082</td>
</tr>
<tr>
<td>3</td>
<td><a href="http://www.uklug.co.uk">www.uklug.co.uk</a></td>
<td>8717</td>
<td>2.59 %</td>
<td>0</td>
<td>1</td>
<td>0.0001</td>
</tr>
<tr>
<td>4</td>
<td><a href="http://www.livejournal.com">www.livejournal.com</a></td>
<td>8338</td>
<td>2.48 %</td>
<td>0</td>
<td>2295</td>
<td>0.2752</td>
</tr>
<tr>
<td>5</td>
<td>xml.tabulets.com</td>
<td>7826</td>
<td>2.33 %</td>
<td>0</td>
<td>2167</td>
<td>0.2769</td>
</tr>
<tr>
<td>6</td>
<td><a href="http://www.deadjournal.com">www.deadjournal.com</a></td>
<td>7224</td>
<td>2.15 %</td>
<td>0</td>
<td>6761</td>
<td>0.9359</td>
</tr>
<tr>
<td>7</td>
<td>testers.cpan.org</td>
<td>5678</td>
<td>1.69 %</td>
<td>0</td>
<td>21</td>
<td>0.0037</td>
</tr>
<tr>
<td>8</td>
<td>blog.goo.ne.jp</td>
<td>5374</td>
<td>1.60 %</td>
<td>25600</td>
<td>350</td>
<td>0.0651</td>
</tr>
<tr>
<td>9</td>
<td><a href="http://www.amagle.com">www.amagle.com</a></td>
<td>5090</td>
<td>1.51 %</td>
<td>10500</td>
<td>5799</td>
<td>1.1393</td>
</tr>
<tr>
<td>10</td>
<td>yaplog.jp</td>
<td>4420</td>
<td>1.31 %</td>
<td>10900</td>
<td>4015</td>
<td>0.9084</td>
</tr>
<tr>
<td>11</td>
<td>blogs.dion.ne.jp</td>
<td>3723</td>
<td>1.11 %</td>
<td>14300</td>
<td>2735</td>
<td>0.7346</td>
</tr>
<tr>
<td>12</td>
<td><a href="http://www.w3.org">www.w3.org</a></td>
<td>3071</td>
<td>0.91 %</td>
<td>11700</td>
<td>3735</td>
<td>1.2162</td>
</tr>
<tr>
<td>13</td>
<td>ameblo.jp</td>
<td>2983</td>
<td>0.89 %</td>
<td>0</td>
<td>90</td>
<td>0.0302</td>
</tr>
<tr>
<td>14</td>
<td>mirrors.webthing.com</td>
<td>2895</td>
<td>0.86 %</td>
<td>2650</td>
<td>2023</td>
<td>0.6988</td>
</tr>
<tr>
<td>15</td>
<td><a href="http://www.mypress.jp">www.mypress.jp</a></td>
<td>2891</td>
<td>0.86 %</td>
<td>12400</td>
<td>1018</td>
<td>0.3521</td>
</tr>
<tr>
<td>16</td>
<td>blog.drecom.jp</td>
<td>2552</td>
<td>0.76 %</td>
<td>6770</td>
<td>850</td>
<td>0.3331</td>
</tr>
<tr>
<td>17</td>
<td><a href="http://www.kwark.org">www.kwark.org</a></td>
<td>2543</td>
<td>0.76 %</td>
<td>43</td>
<td>468</td>
<td>0.1840</td>
</tr>
<tr>
<td>18</td>
<td>iw.stanford.edu</td>
<td>2501</td>
<td>0.74 %</td>
<td>1730</td>
<td>18</td>
<td>0.0072</td>
</tr>
<tr>
<td>19</td>
<td>tinysofa.iszerviz.hu</td>
<td>2471</td>
<td>0.73 %</td>
<td>12200</td>
<td>1089</td>
<td>0.4407</td>
</tr>
<tr>
<td>20</td>
<td>testers.astray.com</td>
<td>2291</td>
<td>0.68 %</td>
<td>0</td>
<td>1</td>
<td>0.0004</td>
</tr>
</tbody>
</table>
SWD distribution by Last Modified Time

generated by swoogle.umbc.edu at 6-APR-2005

log(Document Count) vs Year

- Red: swd
- Blue: R_onto>0
- Green: R_onto>0.8
Swoogle ranks SWDs with “ontology rank”

This report is dynamically generated based on the latest data, and it will take 5 to 10 seconds.
Swoogle Services

- Swoogle Triple Shop
  - Save user selected RDF documents in cart
  - Query cart using SPARQL or RDQL
- Swoogle Web Service
  - [http://swoogle.umbc.edu/webservice/](http://swoogle.umbc.edu/webservice/)
  - Document Search ; Term Search ; Term Digest
- Swoogle RDF Index (Beta): publishing available RDF documents under a web directory.
- Swoogle cache service (Beta): snapshot and achieving the change of RDF documents
- Swoogle provenance service (Alpha): tracking provenance for any given RDF graph.
Query result is encoded in OWL

```
<?xml version="1.0" encoding="WINDOWS-1252"?>
<rdf:RDF xmlns:swoogle="http://swoogle.umbc.edu/swoogle#"
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xmlns:dc="http://purl.org/dc/elements/1.1/">
    <swoogle:Response>
        <rdfs:comment>This result is generated by SWOOGLE2 at
        http://swoogle.umbc.edu.</rdfs:comment>
        <swoogle:queryLimit>10</swoogle:queryLimit>
        <swoogle:queryTime>1.024</swoogle:queryTime>
        <swoogle:totalResults>689</swoogle:totalResults>
        <swoogle:querySuccess>true</swoogle:querySuccess>
        <swoogle:hasResult>
            <rdf:Seq>
                <rdf:li>
                    <swoogle:Document
                        rdf:about="http://www.cs.umbc.edu/~finin/finin.rdf"/>
                </rdf:li>
                <rdf:li>
                    <swoogle:Document
                </rdf:li>
                <rdf:li>
                    <swoogle:Document
                        rdf:about="http://www.csee.umbc.edu/~finin/finin.rdf"/>
                </rdf:li>
                <rdf:li>
                    <swoogle:Document
                        rdf:about="http://www.csee.umbc.edu/~yzou1/ccpp9.rdf"/>
                </rdf:li>
                <rdf:li>
                    <swoogle:Document
                        rdf:about="http://www.cs.umbc.edu/~finin/foaf.rdf"/>
                </rdf:li>
            </rdf:Seq>
        </swoogle:hasResult>
    </swoogle:Response>
</rdf:RDF>
```
Motivation
Shop RDF documents and query them.
No hassle in installing other RDF tools.
Motivation
• Publish URLs of RDF documents for agents.
• Reduce crawling traffic.
• Similar to Google’s “site” query
• Output RDF/XML for agents and HTML for human

Mode 1: Global Swoogle RDF Directory Service

Mode 2: Local RDF Directory Service