

Swoogle

search and metadata for the semantic web



Swoogle is a crawler based search & retrieval system for semantic web documents (SWDs) in RDF, Owl and DAML. It discovers SWDs and computes their metadata and relations, and stores them in an IR system.

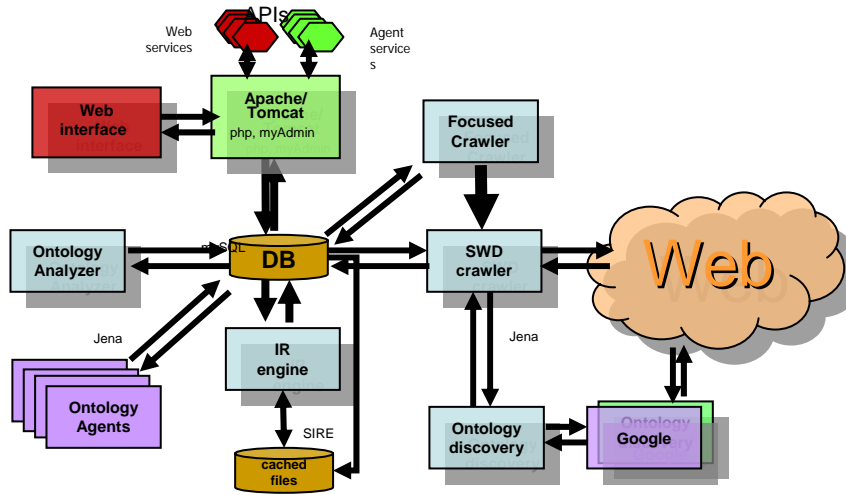
SWD Properties

Language and level; encoding, number of triples, defined classes, defined properties, & defined individuals; type (SWO, SWI); form (RSS, FOAF, P3P, ...); rank; weight; annotations; ...

SWD Relations

Binary: $R(D1, D2)$

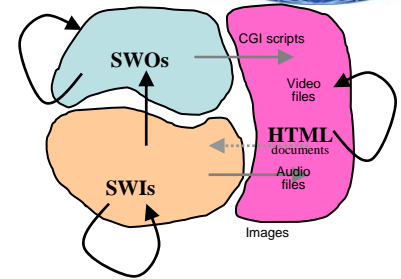
- **IM:** D1 owl:imports D2
 - **IMstar:** transitive closure of IM
 - **EX:** D1 extends D2 by defining classes or properties subsumed by D2's
 - **PV:** owl:priorVersion & subproperties
 - **TM:** D1 uses terms from D2
 - **IN:** D1 uses individual defined in D2
 - **MAP:** D1 maps some of its terms to D2's
 - **SIM:** D1 & D2 are similar
 - **EQ:** D1 & D2 are identical
 - **EQV:** D1 & D2 have the same triples
- Ternary:** $R(D1, D2, D3)$
- **MP3:** D1 maps a term from D2 to D3 using owl:sameClass, etc.



Swoogle uses two kinds of crawlers to discover semantic web documents and several analysis agents to compute metadata and relations among documents and ontologies. Metadata is stored in a relational DBMS.

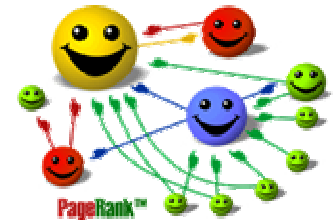
<http://swoogle.umbc.edu/>

Swoogle v1 has ~12K SWDs & 100K relations.
v2 will also catalog classes and properties and their metadata and have >1.6M SWDs.



SWD = SWO + SWI

The web, like Gaul, is divided into three parts: the regular web (e.g. HTML), Semantic Web Ontologies (SWOs), and Semantic Web Instance files (SWIs)



SWD Rank

A SWD's rank is a function of its type (SWO/SWI) and the rank and types of the documents to which it's related.



SWD IR Engine

Swoogle puts documents into a character n-gram based IR engine to compute document similarity and do retrieval from queries