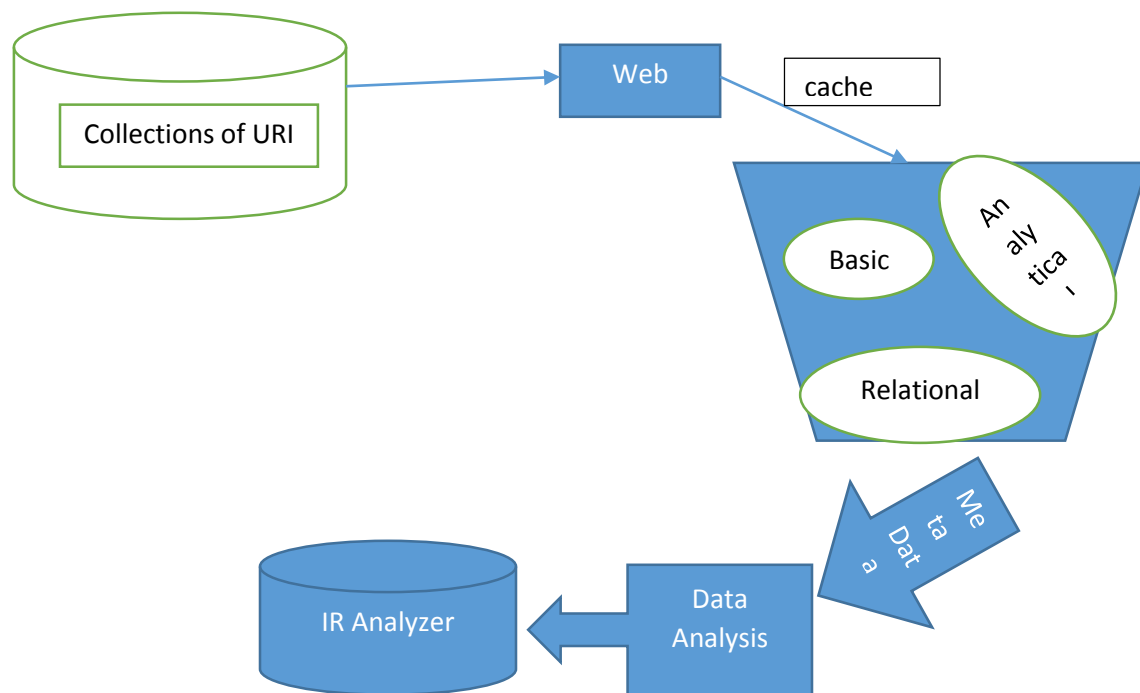# Assignment
Semantic Web Search Engines

swoogle

Swoogle is a crawler-based indexing and retrieval system for the Semantic Web documents [SWD] – i.e., RDF or OWL documents. A crawler is a software that systematically browse and extract information from them for Indexing. So there is no need to develop some prototype systems (like OILIR, SWANGLER) to perform Information Retrieval [IR]. It analyzes the documents it discovered to compute useful metadata properties and relationships between them.

Now, the Semantic Web has in RDF and OWL document formats, is basically a web universe equal to the web of HTML documents. And web search engines like Google, didn't work well with documents encoded in the semantic web languages RDF and OWL. These retrieval systems are designed to work with natural languages and expect documents to contain unstructured text composed of words. It performing poor job of tokenizing semantic web documents and didn't understand XML namespace.

Swoogle generates metadata for given SWD's and lists Ontologies related to given keywords. It is better than other prototype systems like OWLIR, SWANGLER. Because it requires building of Custom Indexing Module. They use their own ontology standards which are not suitable for SWD's. SWOOGLE stores metadata about RDF documents in its database so that it can retrieve SWD's based on Classes(C), Properties (P) and Individuals (I). SWOOGLE is designed to work with all SWDB's and is better than current web search engines like Google because Google work with natural languages only.

Collections of URI

Web cache

An aly tica l

Basic

Relational

Me ta Dat a

Data Analysis

IR Analyzer

Swoogle Architecture

And Swoogle has some issues like crawler is running as a single thread on one machine and Limited number of SWDs discovered and revisited. And it possible to make solution to use of Hadoop Architecture and Grub

The current version of swoogle has discovered and analyzed over 11,000 semantic web documents. A second version has been designed and partially implemented that will also store Meta data on classes and properties and is designed to support millions of documents.

WebKB

WebKB simply known as Knowledge Based servers, but not a web search engine. It allows to store, organize and retrieve knowledge or Document Elements (DEs) in Web-accessible files. It is Web accessible Knowledge Base System (KBS). It's kind of DMBS, unlike RDBMS, ODBMS, deductive DBMS to add, modify large amount of schemas and definitions.

It support logic-based semantic networks and logical inferences to check knowledge entering and knowledge retrieval. It also an on-line private/shared knowledge-based annotation tool. It means WebKB is a knowledge representation tool extended to permit the indexation of information with knowledge statements. And the retrieval of this information via knowledge queries.

It used for manually indexing and then retrieving documents (i.e) retrieving precise information, not simply documents. Hence, it is **unrelated to** classic document retrieval systems like Altavista or Yahoo. And it aims to automatically extract some knowledge from Web documents for classifying them and then easing their retrieval and browsing.

WebKB-1: a private annotation tool

WebKB-1 permits Web users with some aptitude in knowledge representation to store, organize and retrieve knowledge or document elements (DEs) within Web-accessible files.

WebKB-2: a shared annotation tool

WebKB-2 permits Web users to store, organize and retrieve knowledge in a large single knowledge base on WebKB server machine.

Intellidimension:

Intellidimension was founded in June of 2000 and is one of first companies to bring a Semantic Web product to the market. They focus on creating Semantic Web technology that allows applications to handle information intelligently. Intellidimension is self-funded and privately held.

Other Semantic Search Engines are:

Hakia: Semantic Rank algorithm independently ranks content on the basis of more sentence analysis.

Kosmix: Its content aggregating technology will become more important as content on the web grows.

Exalead: It is image search engine based on size and color. The Company has been focusing on the enterprise search market, essentially attempting to solve the problem of search for content where link analysis is of little help.

SenseBot: There is Link Sensor, a tool that can be used on major blogging platforms (WordPress, Blogger, etc.) For automatically picking up key concepts from the post and linking them to related articles from the same blog or publisher.

Cognition Search: It can use to extract relevant results from their contents.

It consists of linguistic algorithms for generating semantic representations that fuel computational parsing of documents.
Algorithms are based on:

- Ontology – To find the senses of a word

     e.g. officer means executives and board members.

- Morphology – To finalize the different forms of a word

     e.g. available and availability.

- Synonymy – To link words to concepts

     e.g. "gulp down" means to drink.

Lexxe: It is a good site to try out ego surfing (how popular you are on the web). The engine also provides keywords that represent categories for the results, clicking which takes you to more relevant topics for the query.

Powerset: After acquired by Microsoft, the changes to live search were noticeable in the related searches and content returned from Wikipedia. All Changes are transparent but definitely in the longer run we can expect more additions to live search. These include faster return of results, more accurate results for less keywords (three to four words maximum) and more awareness on the side of users.

Factbites: It is a very good question and answer search engine ONLY if your queries are relatively simple. It is also a good tool to use if you need a simple answer quickly. It searches for matches on the basis of your whole topic area, not just your keyword. The Factbites engine focuses on finding genuine, meaningful content. This makes it very good at filtering out spam sites.

Semantic Blogs:

Semi-Blog: It's also Known as semantic blogging is the use of rich metadata to transform blogs from simple online diaries to full users in an information sharing ecosystem that blogs enriched with semantic, machine-understandable metadata.

Adding of semantic metadata would make content published in a user's blog easier to find and browse that. Building a semantic blog, require software that integrates tightly with the user's desktop environment.

Examples of Semi-Blogs:

https://www.flickr.com/

http://blogs.msdn.com/
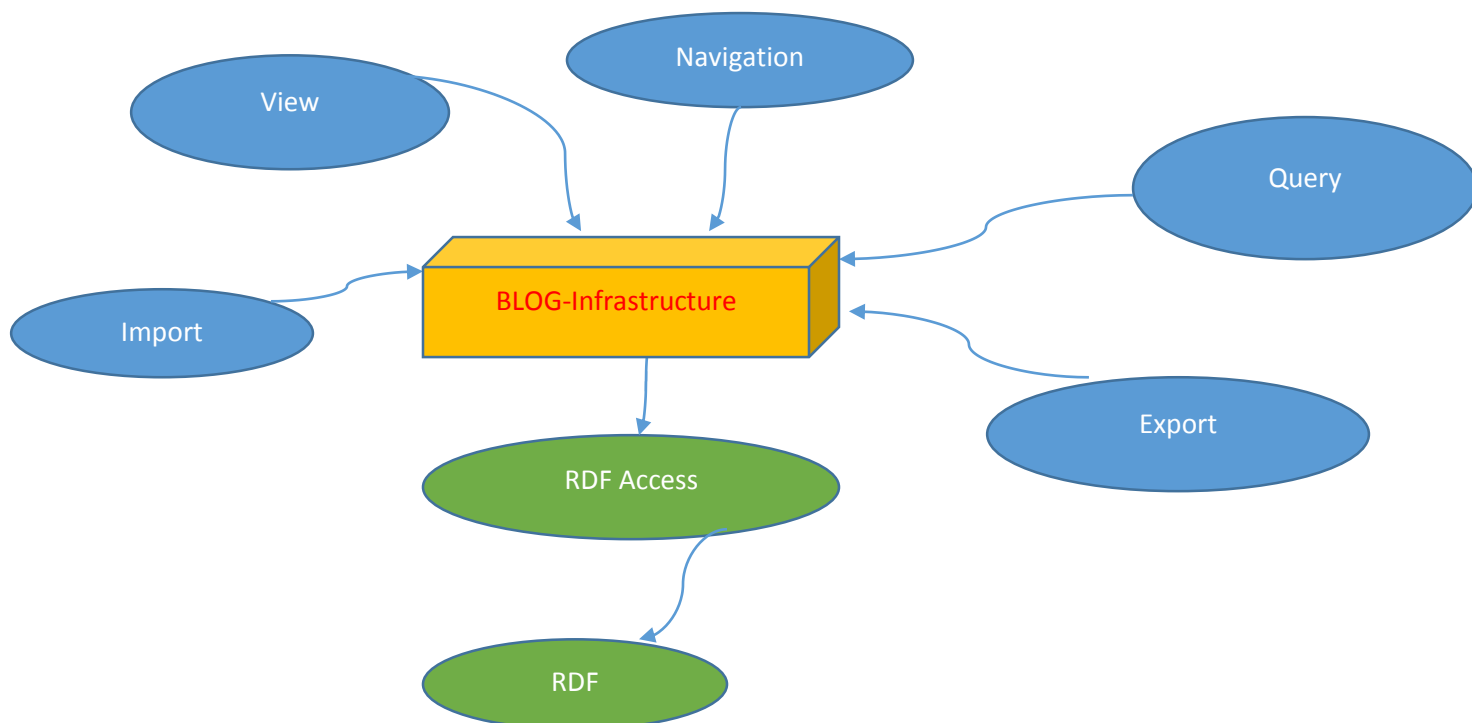
https://devzone.nordicsemi.com/blogs/

HashTags:

A Hashtag is a string of characters preceded by the hash ($\#$) character and it is used to categories text data. It is a label for content and helps to others who are interested in a certain topic, quickly find content on that same topic. The meaning of hashtag is usually difficult to analyze often abbreviated or concatenated concepts.

Hashtags are mostly used on social media sites. But it fame on Twitter. And now you can use hashtags on other social platforms, such as Instagram, Facebook, Pinterest and Google+.

Twitter user have developed tagging culture by adding Hash symbol ($\#$) in front of short Keyword. Collective usage of hashtags indeed information about their requirement. We need to research on semantics of social media content because twitter generates more textual information.

Eg.

#Hashtag, #SemanticAnalysis, #bigdata

Semantic Digital Library:

JeromeDL:

It is a Semantic Digital Library engine. It is used Semantic Web and Social Networking technologies to improve browsing and searching for resources.

Every library user can bookmark interesting books, articles, or other materials in semantically annotated directories in social and semantic services. Users can allow others to see their bookmarks and annotations and share their knowledge within a social network. It can also treat a single library resource as a blog post. Users can comment the content of the resource and reply to others' comments and this way create new knowledge. All data are stored in the library. It is also available in RDF format for querying and processing by other applications and result of every search or browse action is immediately available as a link to RDF. The knowledge created by the users along with the library resource could also be find and used outside the library world. Innovative MBB (MultiBeeBrowse) offers multifaceted navigation and provides SOA for integration with other applications. TagsTreeMaps (TTM) allows to easily filter out resources using clustered tags and presented with tree maps layout. Users can also navigate through the presented search results using SIMILE Exhibit component.

BRICKS:

It establishing the organizational and technological foundations of a Digital Library at the level of a European Digital Memory. A "digital library" in this context refers to a networked system of services over globally available collections of multimedia digital documents, providing a variety of knowledge layers for a variety of users and access modalities.

The BRICKS vision is an integrated system that offers functionality for new generation of Digital Libraries, Like Digital Museums, Digital Archives and other kinds of digital memory systems. The mission of the BRICKS Factory is the definition, development and maintenance of a user and service-oriented space to share knowledge and resources in the Cultural Heritage domain.

It is Peer-to-Peer network: a decentralized network that allows for easy integration of new nodes

- Multi-media digital library: not only for text documents, but also pictures, videos, audio, etc.

- Content Management: member organizations of BRICKS can store digital content in BRICKS or make content stored in other systems available to BRICKS
- Cross-lingual Semantic Search: search for concepts rather than for keywords – will find also synonyms, narrower terms and documents in languages other than the language of the query
- Ability to structure content according to different taxonomies
- Users can comment on content items as well as other people's comments.

References

1. http://www.w3.org/2001/sw/Europe/reports/demo-lessons-report/

2. http://ceur-ws.org/Vol-295/paper07.pdf

3. http://www2013.wwwconference.org/companion/p621.pdf

4. http://smallbiztrends.com/2013/08/what-is-a-hashtag.html

5. http://ceur-ws.org/Vol-295/paper07.pdf

6. https://www.flickr.com/

7. http://blogs.msdn.com/

8. https://devzone.nordicsemi.com/blogs/