**Technology:** Semantic Search

1. **Brief Overview**

Despite the success of available search engines especially Google, the traditional IR model still suffers from many limitations that arise primarily from a lack of understanding of the query and its context. Semantic search is a paradigm that is expected to increase the accuracy of search results by adding layers of context, and meaning to search algorithms. It is expected that the use of semantic search would move the search model from the document level to that of entities and knowledge. Semantic search considers the meaning of words, phrases, or even larger abstractions of text that represent a query and which occur in a document. Meaning is captured and represented in machine readable format through an ontology which is formalized using Semantic Web languages such as RDF or OWL. By “understanding” the meaning of a query and its possible dimensions, it is likely that results returned to the user will be more relevant, and that resources that would have otherwise been missed, will be retrieved. Because semantic search promises to revolutionize IR (by complementing it rather than by replacing it), even search engines that currently dominate the web, the more notable of which are Google, Yahoo, and recently Bing, are making a move towards that technology (Perez, 2009), (Krill, 2010), (BBC news, 2008).

1. **State of the Art (For Latin Languages)**
	1. Technology and Future Trends

The notion of semantic search existed even before the vision of the semantic web. However, the proposal of the semantic web provided a consistent and standardized framework within which semantic search can be enabled. Achieving semantic search within that framework typically involves the following steps:

* Semantic annotation: Semantic annotation refers to the processing of tagging or annotating, documents, words, textual units or resources using an ontology and a semantic web language such as XML, RDF, RDFS, or OWL. Such an annotation can be carried out manually, but since this is very expensive in terms of user time, a lot of work has been carried out in order to devise means for automatically or semi-automatically carrying out the task of semantic annotation. Information extraction and named entity recognition are among the two key fields being utilized in this task.
* Semantic data acquisition: Semantic data acquisition refers to the task of collecting, indexing and storing semantically annotated resources. Typically this can be done through the use of web crawlers that are capable of understanding annotated text, or of converting a structured or semi-structured data store to a standard representation through the use of a semantic web language.
* Semantic search services: After collecting and storing semantic information, semantic search services can then be carried on top of those. These services can vary from basic document oriented enhance search to entity and knowledge search. They can also provide the basis for applications such as question answering.

Since Ontologies aren’t always readily available, a lot of work is also being carried out in the area of ontology learning from textual resources. Resolution of conflicting information collected from the web is another area were work is being carried out and so is the integration of information from multiple resources.

* 1. Applications

Since Semantic search has promised to change the way search is being carried out for the better, a lot of work has been carried out in this area producing a considerable number of applications and systems. A good overview of some of these systems can be found in (Wei et al, 2008).

 One of the earliest Semantic Search systems was SHOE (Heflin and Hendler, 2000). SHOE allowed users to construct constrained logical queries by specifying attribute values of ontology classes to retrieve precise results. The system required to resources to be manually annotated and has no inference support. OWLIR (Mayfield and Finin, 2003) and KIM (Kiryakov et al, 2004) follow an integrated approach that combines logical inference and traditional information retrieval techniques. Semantic annotations in the form of semantic markup as is exploited by logical reasoners to provide enhanced search results. If no results are retrieved using the semantic search system, the system defaults to a traditional search model on the original text to return results. Both works do not take into consideration that sources may publish different Ontologies in similar domains. A number of semantic search engines were built for searching scientific publications. For example, IRIS was specifically developed for intelligently searching computer science literature. FacetedDBLP was also developed for allowing users to browse scientific publications based a number of facets.

AquaLog (Lopez et al, 2005) is a semantic based question-answering system. It employs services for deriving similarity between relations and classes and for the identification of synonyms of verbs and nouns using Word-Net. Another based question-answering system is SemSearch which translates natural language queries into formal queries for reasoning (Lei et al, 2006).

Other semantic search systems include: AquaBrowser (AquaBrowser 2008), TAP (Guha et al 2003), Flink (Mika 2005), BrowseRDF (Oren et al, 2006), DBin (Tummarello et al, 2006), Freebase (FreeBase 2010), Ginseng (Bernstein et al, 2009), H-Dose(Bonino et al, 2004), mspace (ref), MuseumFinland (Ruotsalo et al, 2009), OntoKhoj (Patel et al, 2003), OntoWiki (Auer et al, 2006), Squiggle (Celino et al, 2006), Swoogle, (Ding et al, 2004), Hakia, SenseBot, Powerset, DeepDyve, and Cognition (Pandia 2008).

 A brief feature outline for each of these and more can be found in (Hildebrand et al, 2007).

In addition, general purpose Web based search engines such as Google, Yahoo , Bing, and Ask.com, are all moving towards semantic search (Perez, 2009), (Krill, 2010), (BBC news, 2008), (Zafra, 2009).

1. **State of the Art (For Arabic Language)**
	1. Technology and Future Trends

Very little work has been carried out on Arabic semantic search. The little work that has been carried follows more or less the same technology as for latin languages.

* 1. Current and Envisioned Applications and Market Priorities

As stated before, very few Arabic semantic search systems have been developed.

The work of (El-Beltagy et al, 2003), exploited the use of adding meta-data annotations to Arabic agricultural snippets, and using that to retrieve only snippets that are relevant to a user’s query. The work however had a number of short comings: it did not employ a formal ontology, was tailored to a specific application and did not offer a uniform method for annotating documents outside the document set being addressed by the work.

In (Zaidi and Laskri, 2005) an ontology enabled Web-based multilingual tool for information retrieval in the ‘legal’ domain, is presented. To build an ontology, and to retrieve information from annotated documents, the authors have used Protégé and its query engine. It is not clear however how legal documents were annotated with semantic metadata.

(Qawaqneh et al, 2007) present a method ranking Arabic search results using ontology concepts. The adopted ontology is that of e-commerce. The presented method ranks documents according to the frequency of ontology concepts appearing in the documents. The work suffers from a number of limitations. First, the ontology used seems to be a flat one with very few or no relationships between its concepts. In fact, what is being used seems to be a terminology list in the domain rather than an ontology. There is no methodology for making use of relations had these existed, and the documents being considered are not semantically annotated at any point in time.

The work carried out by the Organizational Web mining group, which is part of the center of excellence for computer modeling and data mining, and which has produced a framework for intelligently annotating and retrieving content in an organization, is perhaps the most comprehensive work addressing the area of Arabic semantic search so far. The search component of the work is yet to be documented.

Recently, the Semantic MediaWiki (SMW), which “helps to search, organize, tag, browse, evaluate, and share” the contents of wikis built using MediaWiki (such as Wikipedia) (Krötzsch, 2010), has been extended to work with Arabic (Wiki news, 2008). The “SMW adds semantic annotations that let you easily publish Semantic Web content, and allow the wiki to function as a collaborative database” (Krötzsch, 2010).

In addition, a number of startup companies are stating that they will provide semantic enabled Arabic search engines. Kngine (kngine, 2010) and AlKhawarizmy (The Next Web, 2010), Yajeel (Yajeel, 2010) are examples of such search engines. Kngine and AlKhawarizmy have been launched by Egyptian based startup companies while Yajeel was launched by a Dubai based company called Taya IT. Technical details of how these systems work, are not available.

1. **Language Resources**
	1. Available Resources (English, Arabic)

Work on semantic web tools that ultimately effect semantic search has really taken off ever during the last 10 year or so. The result being that there is an incredible amount of tools for the semantic Web. (Bergman, 2006) lists 250 Semantic Web Tools. Further additions to this list, bringing it up to 800+ tools is also available in (Bergman, 2006). (Al-Khalifa and Al-Wabil, 2007) estimate that of the 250 tools listed in (Bergman, 2006), approximately 12% support Arabic.

* 1. Needed Resources (English, Arabic)

But based on available data, it can be deduced that the needed resources entail powerful information extraction tools, as well as efficient architectures for supporting an Arabic semantic web. Having ontology authoring and learning tools is also of paramount importance, as Arabic Ontologies are very scarce and are an essential part of any semantic enable search system.

1. **Strengths, weaknesses, opportunities and threats**
	1. Strengths

There are capable researchers in Egypt that can participate in the development of Semantic enabled search engines and tools.

* 1. Weaknesses
* There are very few Arabic enabled Semantic web tools and these are essential for building a semantic search system.
* Building an open domain semantic search system can be expensive and may require a good number of years before being ready for publication.
	1. Opportunities
* There is a big market for a powerful Arabic search Engine and semantic search is pointing the way to the future of search engines in general.
* By selecting appropriate key domains and building semantic search tools just for those, problems associated with semantic search may be simplified. The key is idenfiying particular domains that might be of great interest to web users.
* There are very few Arabic enable Semantic web tools and availing these in itself can be a contribution.
	1. Threats

 Arabic semantic enable search engines that have recently emerged, may dominate the market.

1. **References**
	1. Al-Khalifa, H. A. and Al-Wabil. (2007). “The Arabic language and the semantic web: Challenges and opportunities”, The 1st International Symposium on Computers and Arabic Language & Exhibition 2007,
	2. AquaBrowser (2008). <http://www.medialab.nl/>
	3. Auer, S.; Dietzold, S.; Riechert, T. (2006). “OntoWiki – A Tool for Social, Semantic Collaboration”, 5th International Semantic Web Conference, , GA, USA. In I. Cruz et al. (Eds.): ISWC 2006, LNCS 4273, pp. 736–749, 2006. Springer-Verlag Berlin Heidelberg.
	4. BBC news (2008). “Yahoo makes semantic search shift”, <http://news.bbc.co.uk/2/hi/technology/7296056.stm>
	5. Bergman, M. (2006). “Comprehensive Listing of 250 Semantic Web Tools”. http://mkbergman.com/?p=291
	6. Bergman, M. (2010). Sweet Tools (Sem Web).<http://www.mkbergman.com/new-version-sweet-tools-sem-web/>
	7. Bernstein, A; Kaufmann, E; Kiefer, C. (2009). “Querying the semantic web with ginseng - A guided input natural language search Engine”, In: Clematide, S; Klenner, M; Volk, M. Searching answers: Festschrift in honour of Michael Hess on the occasion of his 60th birthday. Münster, 1-10.
	8. Bonino, D. Bosca, A., Corno, F., Farinetti, L. , and Pescarmona, F. (2004). "H-DOSE: an Holistic Distributed Open Semantic Elaboration Platform," In proceedings of SWAP2004: 1st Italian Semantic Web Workshop, Ancona, Italy.
	9. Celino, I., Valle, E. D., Cerzza, D., and Turati, A. (2006) “Squiggle: a semantic search engine for indexing and retrieval of multimedia content”, In proceedings of SAMT 2006, pp. 20–34.
	10. Diederich, J., Balke, W. T., and Thaden, U. (2007). “Demonstrating the semantic growbag: automatically creating topic facets for faceteddblp. JCDL 2007, ACM, p. 505.
	11. Ding, L., Finin, T., Joshi, A., Pan, R., Cost, R. S., Sachs, J., Doshi, V., Reddivari, P. and Peng, Y. (2004). “Swoogle: A Search and Metadata Engine for the Semantic Web”, in proceedings of the 13th ACM Conf. on Information and Knowledge Management, Washington DC.
	12. El-Beltagy, S., R., Rafea, A., and Abdelhamid, Y. (2003). “Chapter 13: Using Dynamically Acquired Background Knowledge for Information Extraction and Intelligent Search”. In Editor (Masoud Mohammadian), Intelligent Agents for Data Mining and Information Retrieval, Hershey, PA, USA: Idea Group Publishing.
	13. FreeBase (2010). <http://www.freebase.com/>
	14. Guha, R. V., McCool, R., and Miller,E. (2003). “Semantic search”, In proceedings of WWW 2003, pp. 700–709.
	15. Harris, C., Owens, A., Russell, A. and Smith, D.A.(2004). “mSpace: Exploring The Semantic Web”, a Technical Report in Support of the mSpace software framework, ECS, University of Southampton, Southampton.
	16. Heflin, J. and Hendler, J. (2000). “ Searching the web with SHOE”. Artificial Intelligence for Web Search Menlo Park, CA, pp. 35–40.
	17. Hildebrand, M. et al. (2007). “Semantic Search Survey”, Wikipedia, <http://swuiwiki.webscience.org/index.php/Semantic_Search_Survey>
	18. Kiryakov, A., Popov, B., Terziev, I., Manov, D. and Ognyanoff, D. (2004) “Semantic annotation, indexing, and retrieval”, Journal of Web Semantics, Vol. 2, No. 1, 2004, pp. 49–79.
	19. Kngine. (2010). <http://kngine.com/>
	20. Krill, P. (2010). “Microsoft to update Bing with semantic search”, InfoWorld. <http://news.techworld.com/applications/3211273/microsoft-to-update-bing-with-semantic-search/?olo=rss>
	21. Krötzsch, M. (2010). “Semantic MediaWiki”, <http://semantic-mediawiki.org/wiki/Semantic_MediaWiki>
	22. Lei, Y., Uren, V. S. and E Motta, E. (2006) “Semsearch: A search engine for the semantic web”. EKAW 2006, pp. 238–245.
	23. Lopez, V., Pasin, M. and Motta,E. (2005). “Aqualog: An ontology-portable question answering system for the semantic web”. ESWC 2005, pp. 546–562.
	24. Mayfield, J. and Finin, T. (2003). “Information retrieval on the semantic web: Integrating inference and retrieval. In proceedings of the Workshop on Semantic Web at SIGIR 2003.
	25. Mika,P. (2005). “ Flink: Semantic web technology for the extraction and analysis of social networks”, Journal of Web Semantics, Vol. 3, No. 2-3, pp. 211–223.
	26. Oren, E., Delbru, R., Decker,S. (2006) “Extending Faceted Navigation for RDF Data”, In proceedings of the International Semantic Web Conference pp. 559-572, 2006.
	27. Pandia (2008). <http://www.pandia.com/sew/1262-top-5-semantic-search-engines.html>
	28. Patel, C., Supekar, K., Lee, Y., Park E. K. (2003). “OntoKhoj: a semantic web portal for ontology searching, ranking and classification”, In Roger H. L. Chiang, Alberto H. F. Laender, Ee-Peng Lim, editors, Fifth ACM CIKM International Workshop on Web Information and Data Management (WIDM 2003), New Orleans, Louisiana, USA, November 7-8, 2003. pages 58-61, ACM.
	29. Perez, J,. C. (2009) “Google Rolls out Semantic Search Capabilities”, <http://www.pcworld.com/businesscenter/article/161869/google_rolls_out_semantic_search_capabilities.html>
	30. Qawaqneh, Z., El-Qawasmeh, E., and Kayed, A. (2007). “New Method for Ranking Arabic Web Sites Using Ontology Concepts”, in proceedings of IEEE ICDIM, p. 649-656.
	31. Ruotsalo, T., Aroyo, L., and Schreiber, G. (2009). “Knowledge-Based Linguistic Annotation of Digital Cultural Heritage Collections”, IEEE Intelligent Systems, vol. 24, no. 2, pp. 64-75, IEEE Computer Society.
	32. The next Web. (2010). “Is Arabic search about to change?”, <http://thenextweb.com/me/2010/01/29/ar-search/>
	33. Tummarello, G. , Morbidoni, C., and Nucci, M. (2006). “Enabling Semantic Web communities with DBin: an overview”, in proceedings of the Fifth International Semantic Web Conference ISWC 2006, Athens, GA, USA.
	34. Wei, W., Barnaghi, P. M. and Bargiela, A. (2007). “The Anatomy and Design of A Semantic Search Engine”, Tech. rep. UNMC-CS-200712-1,, School of Computer Science, University of Nottingham Malaysia Campus, 2007.
	35. Wei, W., Barnaghi, P. M. and Bargiela, A. (2008). “Search with Meanings: An Overview of Semantic Search Systems”, International journal of Communications of SIWN, Vol. 3, pp. 76-82.
	36. Wiki news. (2008). “SMW now available in Arabic”, <http://semantic-mediawiki.org/wiki/SMW_now_available_in_Arabic>
	37. Yajeel. (2010). <http://yajeel.com/>
	38. Zafra, A. (2009). “Ask.com Focuses on Semantic Search”, <http://www.searchenginejournal.com/askcom-focuses-on-semantic-search/8252/>
	39. Zaidi, S., Laskri, M. (2005). A cross-language information retrieval based on an Arabic ontology in the legal domain. The International Conference On Signal-Image Technology & Internet–Based Systems (SITIS’05), Morocco.