International Journal of Intelligent Computing and Information Science

SEARCHSENSE: A SEMANTIC META SEARCH ENGINE

Maged F. El-Sayed

Department of Information Systems and Computers, Faculty of Commerce, Alexandria University, Alexandria, Egypt
maged@alexu.edu.eg

Abstract: *Despite the recent advancements in information retrieval research and in online search engines, satisfying the needs of search engines users remains largely challenging. This is due to many reasons including the size and richness of information available over the internet and the semantic gap between the intention of search engine user and how search engines understand that intention.*

This article highlights the main reasons for ineffective web searches and sheds the lights on the status of ongoing research stream, the SearchSense project, which focuses on improving the effectiveness of search engines in satisfying users' needs. The research has been carried out through multiple projects that were implemented over several years with main overall focus on designing an effective semantic search engine. SearchSense employs the semantic technology to bridge the gap between search engines and their users and to provide a better presentation of web search results. SearchSense could be used as a Meta Semantic Search Engine on top of regular search engine or could easily be incorporated in any information retrieval system.

The article describes the overall framework of the solution and outlines its main components. Details on the technicalities of the solution components are presented in relevant articles. The article also provided a summary of the results of the experiments that have been conducted to test the effectiveness of the solution.

Keywords: *Information Retrieval, Semantic Web, Semantic Search Engines, Ontologies.*

1. Introduction

With the rapid increase in the size of information over the internet, the importance of search engines as the main tool for searching and retrieving information is continuously increasing. Results returned by traditional search engine are typically large in size and contain many irrelevant documents. Search engine users typically need to go through numerous returned documents to reach to answers for their queries. One of the reasons that prevent a typical user from making the best search results is that traditional search engines and information retrieval systems are largely lexical; meaning that they rely on matching keywords in the user's search queries and do not necessarily understands the user's intention.

Many researchers have recently focused on semantic technology as a mean to improve web search effectiveness. Semantic search is a search that is based on the meaning rather than the syntax of the query. Semantic Search has been largely studied during the last few years, [1] provide a recent survey of this research. In addition to research solution many Semantic Search Engine have recently emerged, for example: Hakia, Kosmix, Swoogle, Cognition, Lexxe, Kngine, Powerset, DuckDuckGo, Sensebot [2][3][4]. Google search engine has also been improving its semantic search abilities as it answers an increasing fraction of natural language queries [1].

According to [1], all Semantic Search techniques introduced so far merely simulate an understanding of the meaning of the search rather than providing real understanding of the search. [5] have presented four common issues that face intelligent semantic search engines, namely: (1) low precision and high recall, (2) identifying intention of the user, (3) extrapolation of users' search patterns, and (4) handling inaccurate queries. Despite the many attempts for creating a semantic search engines, none of these attempts has managed to address all those common issues and none of these search engines has gained wide recognition among internet users. *SearchSense* attempts to address and resolve all of the four common issues discussed by [5]. In particular, it improves the precision and recall of retrieved result, identifies intention of the user, uses patterns learned from previous user's searches, and handles the problem of users' inaccurate queries.

[4] have identified four main approaches that could be used for developing semantic information retrieval systems: (1) contextual analysis, (2) ontology, (3) natural language understanding and (4) reasoning. Contextual analysis focuses on disambiguating queries. Reasoning attempts to infer additional information from existing facts. Natural language understanding involves identifying the intent of the information. The forth approach uses ontology to represent knowledge about the domain and to enrich queries. While the Ontology approach is the most used by semantic retrieval system some systems use mix and match between the four approaches [4]. One very important open research question is which combination of the four approaches is better [6]. We do believe that building a full-fledged semantic information retrieval system requires the integration of all the four approaches. To the best of the author's knowledge, *SearchSense* is the only solution that provides a clear and effective integration of all the four approaches.

Unlike many other solutions, *SearchSense* is not limited to keyword or structured searches as it also supports natural language web searches. *SearchSense* does not require the user to formulate queries in a formal language or to provide more than keyword such as query-by-example solutions; this makes *SearchSense* more user-oriented. While *SearchSense* can be considered a Meta search engine in which it could be placed on top of any subordinate search engine, it can also be incorporated in any information retrieval system or search engine.

Based on the classification scheme for semantic search engines presented by [7], the main features of *SearchSense* are summarized as follows:

- **Architectures:** although *SearchSense* is designed as a meta search engine that can work on top of any subordinate search engine. It could also be developed as a stand-alone semantic search engine. And due to the modularity of *SearchSense* design the semantic components of *SearchSense* could be easily used to upgrade any traditional search engine.
- **Coupling level of documents and ontologies**: *SearchSense* does not require the documents to have semantic annotations which is flexible and realistic given that most internet documents do not have semantic annotations. Although such loose coupling between documents and ontologies traditionally leads to difficulty of selecting appropriate ontology for a given domain, *SearchSense* avoids such drawback through the powerful semantic classification feature of *SearchSense* that enables the system to select ontologies relevant to retrieved documents.

- **Transparency:** *SearchSense* is a hybrid system as it can act as a transparent system (meaning that the semantic capabilities of the system are invisible to the user and the system does not request additional information from the user) and at the same time it can utilize users' interactions.
- **User's Context utilization:** although *SearchSense* is capable of categorizing some queries into question-categories through the use of search patterns, the system should not be categorized as "Hard-coded" instead it should be classified as "Learning" since the system is capable of guessing the intention of the user and his/her information needs through extracting the user context. Such context is used by *SearchSense* for query modification and reformulation.
- **Query modification:** *SearchSense* performs two query modification techniques: query rewriting and Graph-driven conversion. Query rewriting optimizes the query through augmentation, trimming, and substantiation. Graph-driven conversion utilizes graph structure and spreading-activation algorithm for performing query modification and rewriting.
- **Ontology structure:** *SearchSense* utilizes combination of Anonymous properties, Standard properties and Domain specific properties. The Standard properties that the system use are: synonym_of, hypernym_of, meronym_of, instance_of and negation_of. This gives the system better semantic search capabilities and more flexibility in handling different domains.
- **Ontology technology:** *SearchSense* uses OWL as the ontology description language which enables reusability and interoperability of ontologies.

This research aggregates the reasons for ineffective web searches into three issues:

1. **Difficulty of determining user's intentions:** Queries submitted to search engine are typically short, ambiguous and not well articulated. This causes a semantic gap between the user and the search engine. Traditional search engines and retrieval solutions rely largely on retrieving documents that contain terms similar to the terms in the user's query. An effective web search solution needs to understand the intention of the user and the semantics of his/her query and to provide answers that better satisfy his/her needs. *SearchSense* tackles this issue through the following novel techniques:
    - Classifying the web search query into a well-defined web search classification [8]. This allows the determination of user's intention.
    - Reformulating the web search query to a semantically richer query that approximates the user's intention to the search engine [9] [10]. This enables the search engine to provide a better result.
2. **Diversity of terms semantics in different domain:** web terms are largely diverse and rich in semantics. A term may have different semantics depending on the context and the domain. This issue contributes largely to those non-relevant web pages that appear in search engine result. An effective retrieval solution must be able to take these differences into consideration and process the right document semantics depending on the domain of the web search. *SearchSense* tackles this issue through the use of two types of ontologies: a generic ontology that models widely known concepts and their relationships and a collection of domain-specific ontologies that model the semantics of web document terms in different domains [9]. These two types of ontologies are used in the classification of queries and in the reformulation of new semantically richer queries.
3. **Result representation:** majority of search engine represent the search result in the form of a list of URL of web documents that are retrieved based on the search query. Despite the efforts given by search engines to improving the ordering of retrieved lists to provide the most relevant

result first, the big picture of the retrieved result remains largely hidden from the user due to the difficulty of comprehending large parts of the result. We believe that such big picture is very important for educating the user of what is available over the internet in relationship to his/her web search. To provide such big picture of the result, an effective retrieval solution should extract knowledge from the entire result (or a considerable part of it) and provide it to the user in a form that is easy to comprehend. *SearchSense* tackles this issue through two alternatives:

- Semantically clustering the web search result, where the result is represented as a number of clusters of retrieved web documents [11]. These clusters are built through a novel semantic similarity solution rather than just performing traditional term-based similarity on retrieved documents. This provides a better clustering quality were the clustering is done in a way similar to what a human would do but with added benefit of being able to process large results that is not feasible for a human to process.
- Representing the search result visually as a Semantic Tag Cloud. Semantic Tag Cloud is an enhancement over the traditional tag cloud in which the cloud represent better knowledge about the retrieved documents rather than a disperse collection of terms.

2. SearchSense: The Research Project

This paper reveals the overall *SearchSense* project which is the result of a stream of research over several years led by the author of this article. This research stream is composed of several research projects [12] [10] [9] [8] [11] each of which is contributing a component to the larger solution shown in Figure 1. The following subsections discuss the main components of *SearchSense*.

2.1 Classification of Web Searches

As part of solving the problem of determining the user's intention, one subproject [8] of *SearchSense* focuses on classifying web search queries. By determining the class/category of the web search, the intention of the user and the type of information needed could be identified. Table 1 shows the 11 classes of web searches that are used in this project. These classes of web searches are based on the work of [13] and [14]. Each of these classes/categories has well-defined characteristics that enable determining the intention of the search. For example: the query "Wold War II war" is considered Informational –Undirected, where the purpose of queries of this category is to know anything and everything about a topic, most queries in this type are related to science, medicine, history and news and celebrities.

To extract patterns of web searches, 80,000 randomly selected queries from AOL 2006 datasets were analyzed [8]. Based on this analysis 1182 unique *Search Type Patterns* were extracted and classified into the 11 classes of web searches shown in Table 1. A *Search Type Pattern* is composed of a sequence of Web Terms. Table 1 shows the classification of web search patterns used by *SearchSense* and the number of patterns in each search classification. In addition to the *Search Type Patterns*, a taxonomy that contains 10,440 terms of most frequently used web search terms was constructed.

Using both the *Search Type Patterns* and web search terms taxonomy it became possible to classify new web searches into a well defined category and hence determine the user's intention. For example, the web search query: "list of movies by steven spielberg" matches the Search Pattern <CN_IFT + PP + CN_Ent + PP + PN>, where CN_IFT represents Informational Term (e.g. the term "List"), PP represents Preposition, CN_Ent represents Entertainment term, PP represents Preposition, PN represents Proper Noun. This Search Pattern is classified by the solution as "Informational-List Search Type" which represents the class of web searches where the user is trying to find a list of suggested websites or documents or to find a list of suggestions for further research.

Experiments have shown a very promising result where the solution was able to correctly classify 85.5% of randomly selected test search queries given the 1182 unique Search Type Patterns and the web search terms taxonomy. More details of the experiments are presented in the experiments section.

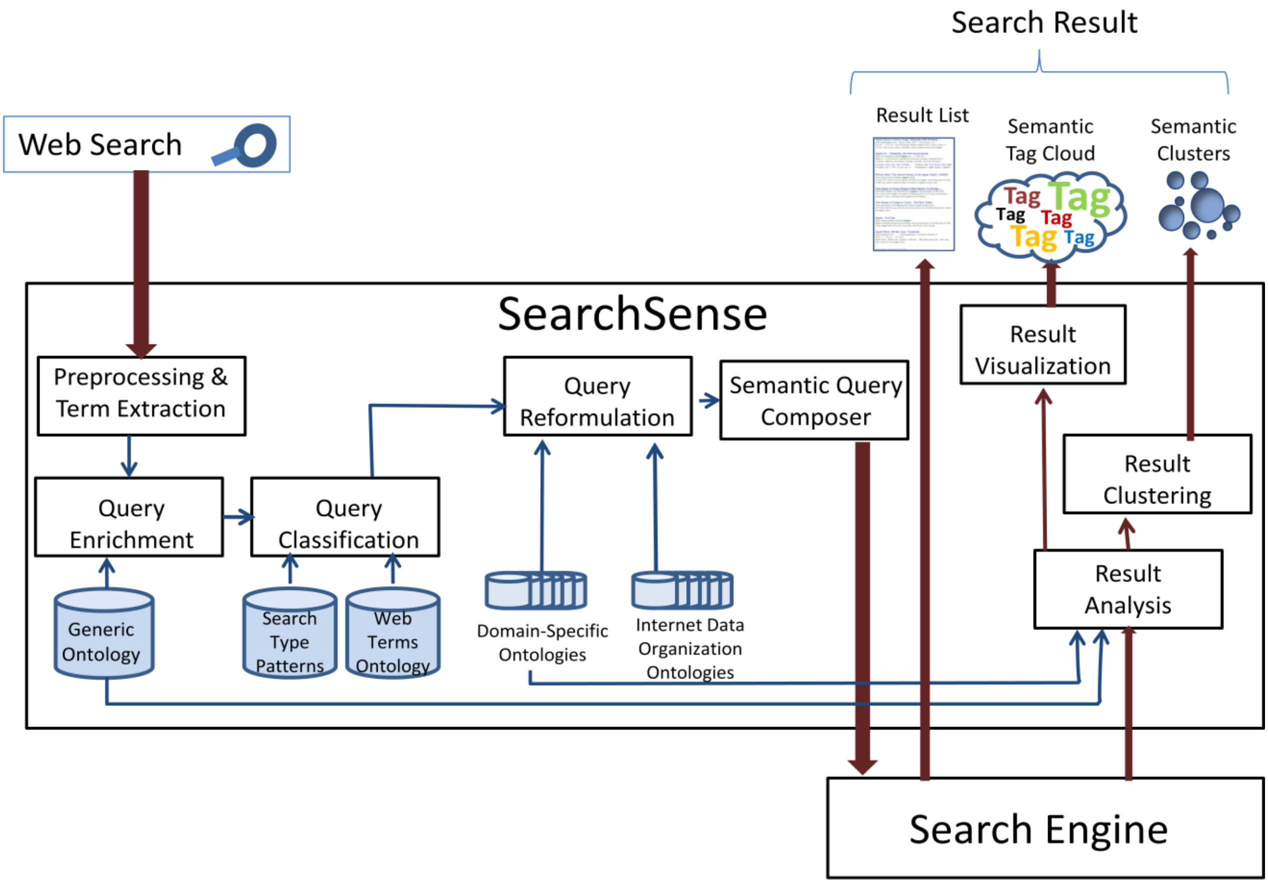

Figure 1: Outline of SearchSense

In addition to the technique described above and as part of *SearchSense* research, a fuzzy approach was developed for classifying web searches [10]. The approach relies on fuzzy membership calculation and KNN classification and has two phases:

- The training phase: in this phase membership of query terms to given classes are computed. The fuzzy degree value of membership of each term ranges between 0 and 1. The larger the value, the better the match of keyword to the class. In particular, a membership of 1 reflects a perfect match of the term to the class and that this term does not belong to other classes. A membership of 0 reflects that the term does not belong to the class. The membership value of a term k_i to a class c_j is computed as follows:

$$\mu\left(k_i, c_j\right) = \frac{\sum W_{c_j}(k_i)}{\sum W(k_i)}$$

Where:

$\mu\left(k_i, c_j\right)$: is the membership of a term k_i to a class c_j
$W_{c_j}(k_i)$: is the occurrence frequency of term k_i in class c_j
$W(k_i)$: is the occurrence frequency of term k_i in all classes

- The classification phase: in this phase terms are extracted from the new query, these terms are semantically enriched through ontology annotation then the query is classified using a fuzzy similarity measure. The fuzzy similarity measure relies on computing the membership value of each semantically enriched query term to given classes. Then query classification is performed using the k-nearest neighbor (KNN) classifier based on the similarity computation discussed above. The fuzzy similarity between a query q and a class c_j is calculated as follows:

$$Sim\left(q, c_j\right) = \frac{\sum_{k_i \in q} \mu(k_i, c_j) \otimes \mu_q(k_i)}{\sum_{k_i \in q} \mu\left(k_i, c_j\right) \oplus \mu_q(k_i)}$$

Where:
$\mu_q(k_i)$: is membership value of term k_i to query q which is a Boolean value representing the occurrence of term k in query q (it takes the value 1 if the term is present in the query and 0 if the term is not present in the query)
$\otimes$: is the fuzzy conjunction (t-norm) operator
$\oplus$: is the fuzzy disjunction (t-conorm) operator.

Our experiments show that this fuzzy classification technique provides an average classification accuracy of 89.2%.

Table 1: Classification Web Searches and number of patterns in each class

Classification	Number of patterns
Informational -List	155
Informational -Find	164
Informational -Advice	121
Informational -Undirected	51
Informational -Directed-Open	113
Informational -Directed -Closed	234
Transactional -Obtain - Online	59
Transactional -Obtain -Offline	76
Transactional -Interact	28
Transactional -Download –Free	104
Transactional -Download -not Free	69

2.2 Formulating Semantically-rich Queries

A semantic gap between the user and the search engine exists when the search engine is unable to understand the user's intention. Such semantic gap causes low result precision and high recall [5].

The approach taken by *SearchSense* in [9] to solve this problem is reformulating the user's web search into a more semantically rich web search that approximates the user's intention to the search engine. This is done through three main phases:

1. Query enrichment: where the user's query is annotated with additional terms that are obtained through processing a generic ontology to make it more semantically rich. Determining which terms to be obtained from the generic ontology is done through a spreading activation algorithm [15] and a shortest path algorithm. This step facilitates the query classification step discussed next.
2. Query classification: the enriched query is classified into one of web search domains. For obtaining better result for classification, we apply a semantic classification technique [12] that compute the similarity between a semantic representation of the query and a semantic representation of the class using a cosine similarity function. These semantic representations are obtained through the use of spreading activation. We may also use the fuzzy classification presented above [10].
3. Query reformulation: after enriching the query and determining its domain, a domain specific ontology called IDOO (Internet-Data-Organization Ontology), and a special algorithm are used to reformulate the query. This reformulation process involves adding terms and deleting terms to the query. IDOO models how terms are organized and expressed in a specific domain on the internet. We believe that the consideration of such valuable knowledge is very important to query reformulation.

For example, given the query "Where can I find a Cheap Kindle?", the query enrichment phase adds the following terms to it: Question, Investigation, Action, E-reader, Electronics, Product, Price, Criteria and Judgment Measure, and Location. Next in the query classification phase the system semantically classifies the query into the shopping domain. In the query reformulation phase, a number of reformulated queries are generated such as: <Kindle "super deal" "Buy now">, <Kindle "bargain" "add to cart"> , <Kindle "floor model" "Buy web only">, .. etc.

Experiments confirm that this solution improves average precision of query results, more details on experiments are presented in the experiments section.

2.3 Clustering Search Engine Result

The result of web search is typically large and diversified. Most search engine users view only the first few pages of the result rather than viewing the entire result or major part of it. We believe that processing a large part of the search engine result is very important for understanding the big picture of the result. One possible useful way of presenting the large result is in the form of clusters, where each cluster groups similar web documents. For example the web search: "Apple" might return clusters representing: electronics, company, fruit,.. etc.

Unlike traditional clustering solutions that rely on syntactic term similarity, *SearchSense* takes a more human-like approach where the similarity of documents is determined by the semantics of the contents of documents. For example, given four documents: $D1$ = {Apple:1, Headphone:1} , $D2$ = {Apple:1, diet:1} , $D3$ = {Orange:1, diet:1} , and $D4$ = {Samsung galaxy:1, Bluetooth:1}, a traditional clustering solution would assign the following similarities between each two documents: 50% for D1-D2, 0% for D1-D3, 0% for D1-D4, 50% for D2-D3, 0% for D2-D4, and 0% for D3-D4. Such similarity does not take into consideration the semantics of the terms and will result in bad clusters. *SearchSense* takes a different approach [11] were the clustering of the documents is determined by the semantic similarity of the document contents. For the same four documents, *SearchSense* generates the following similarities: 18% for D1-D2, 0% for D1-D3, 64% for D1-D4, 83% for D2-D3, 0% for D2-D4, and 0% for D3-D4. This result is closer to human perception and would create better clusters than those created by traditional clustering solutions.

SearchSense performs semantic clustering through the following main steps:

- Semantically modeling each retrieved document: This involves feature extraction and semantic annotation (through the use of a generic ontology) to construct a semantic representation of the document.
- Enriching the semantics of the document model through the use of activation spreading algorithm after determining the most relevant features to be enriched though shortest path algorithm.
- Computing similarity among documents in the web search result based on the most relevant features in the document model.
- Applying clustering algorithm to obtain the clusters.

Experiments have confirmed that this semantic clustering technique provides high precision and are close to human clustering.

2.4 Web Search Result Visualization Using Semantic Tag Clouds

A Tag Cloud is a visual representation for textual contents that is widely used to summarize contents of websites. The size and/or the color of a term in the Tag Cloud relates to its frequency in the original document. Figure 2(a) shows an example of a traditional Tag Cloud that contains terms extracted from an electronics website. The number associated with each term reflects its frequency in the document. Traditional Tag Clouds suffer from many issues including semantic density [16] and Poor understandability of structure and relation [17]. Varity of solutions has been introduced for these issues, among them solutions that rely on semantic clustering and aggregation [18] and solution that use WordNet ontologies and semantics [19]. Yet these solutions mainly consider the lexical similarity of terms and do not capture richer semantics.

SearchSense introduces an improved version of Tag Cloud, called the Semantic Tag Could. A Semantic Tag Could provides a thematic presentation of the web document where not only lexical term relationships are considers but also other semantics are taken into consideration. *SearchSense* approach mainly relies on domain-specific ontologies and a special algorithm in generating a Tag Cloud that provides better representation of the document. Figure 2(b) shows a Semantic Tag Cloud which represents *SearchSense*'s alternative to the traditional Tag Cloud in Figure 2(a). Semantic Tag Cloud is not only capable of semantically grouping and aggregating concepts from the document but also capable of reflecting additional concepts that contribute to the thematic presentation of the documents.

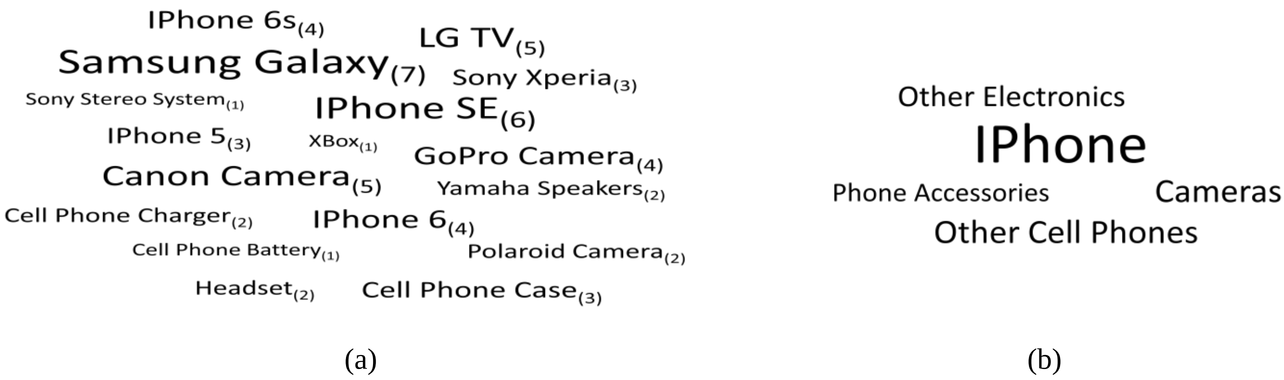

Figure 2: (a) A typical Tag Cloud with frequencies, and (b) Semantic Tag Cloud for the terms shown in (a)

3. Experimental Evaluation

Throughout the *SearchSense* project experience a variety of experiments has been conducted to evaluate the validity and performance of SearchSense. Evaluation is done using prototype software developed using Java programming language.

For the purpose of experiments a number of Ontologies were built including a generic ontology and domain-specific ontologies. Ontologies were constructed using Protégé ontology builder. Access to ontologies is done using Jena API programmatic environment and using the SPARQL query language where RDF and OWL data models were queried. The prototype also uses a database that contains 1182 unique Search Type Patterns and a taxonomoy of web search terms that contains 10,440 terms. Search Type Patterns and the web search terms taxonomoy are constructed through analyzing 80,000 randomly selected queries from AOL 2006 datasets.

The focus of the experiments was on testing different features of the system including: semantic clustering, web search classification, query reformulation and improvement in search engine results.

For testing the query reformulation feature of the system and the improvement in the returned results a set experiments were used. The experiments mainly measure the percentage of relevant URLs in the top 20 URLs returned by the web search. This should be sufficient to judge the quality of the returned result since a typical internet user mostly examines the first few documents in the result [20]. We ran several experiments using keyword search engines, such as Google and Yahoo. We also experimented against semantic search engines, such as Kngine and Hakia [2].

The results of the experiments show that our solution enhances both the results retrieved by keyword search engines as well as by semantic search engines. The best result achieved was when our solution was used on top of Google search engine; the average precision of the result in this case was 80% while the average precision of Google alone during time of experiments was 33%. Our solution has also improved the average precision of Yahoo from 16% to 75%, average precision of Kngine from 19% to 46%, and average precision of Hakia from 27% to 59%. The experiments also measured the average decrease in result size, and it was found that due to the reformulated queries the result size of Google has decreased by an average of is 85%.

The semantic clustering feature of *SearchSense* was tested using ten different queries of diverse semantics, namely: Apple", "Paris", "Jaguar", "Hollywood", "RedHotChiliPeppers", "Mac", "SnowLeopard", "Lion", "Tiger", and "Mouse". For the sake of computing precision, human clustering was taken as the reference for correctness. *SearchSense* scored average precision of 90%. Five of the queries that have been used in experiments were also used by [21] to test a clustering solution that was used. The average precision reported by [21] was 78.2%.

The web search classification feature of *SearchSense* was tested using 10,000 queries that are randomly selected from AOL 2006. Results of the experiment show that *SearchSense* is able to identify and classify 7754 of the queries. The unclassified 2246 queries contain 927 quires that are considered vague or contain mistakes. This makes the accuracy of the classification 85.5% of the valid queries. We have also tested our fuzzy classification technique [10] and experiment result showed average accuracy of 89.2%.

4. Conclusion

Due the size and diversity of web contents, the retrieval of relevant documents remains to be an important and challenging task. This article reviews the status of *SearchSense* research project that focuses on improving the effectiveness of web searches. *SearchSense* encompasses several projects, each of which contributes an important component of the solution. *SearchSense* employs the semantic technology for bridging the semantic gap between the search engines and their users and to provide a better presentation of the result.

In this article the main changes facing web search has been identified and discussed. These main challenges are: difficulty of determining User's intentions, variations in terms semantics among web documents, and result size. The techniques presented in this article for solving these challenges include: semantic clustering of web search results, classification of web searches, reformulations of web searches into more semantically-rich web searches, web search result clustering and visualization using Semantic Tag Clouds. A variety of experiments has been conducted to evaluate the proposed solutions. These experiments have shown very promising results.

As future work, ontologies used by *SearchSense* could be expanded and integrated with online ontologies to further enhance the effectiveness of the system. More integration and optimization of different components of the system will be investigated. Finding more web search patterns and Self-learning capability of the system will also be considered. One additional area of future research is the optimization of the system performance.

References

[1] H. Bast, B. Buchhold, and E. Haussmann, "Semantic Search on Text and Knowledge Bases," *Found. Trends® Inf. Retr.*, vol. 10, no. 2–3, pp. 119–271, 2016.

[2] M. Manuja and D. Garg, "Semantic web mining of un-structured data: Challenges and opportunities," *Int. J. Eng.*, vol. 5, no. 3, pp. 268–276, 2011.

[3] S. K. Sahu *et al.*, "Comparative Analysis of Semantic and Syntactic Based Search Engines By," *J. Adv. Comput. Commun. Technol.*, vol. 4, no. 2, pp. 22–27, 2016.

[4] G. Sudeepthi, G. Anuradha, and M. Babu, "A survey on semantic web search engine," *Int. J. Comput. Sci.*, vol. 9, no. 2, pp. 241–245, 2012.

[5] G. Madhu, A. Govardhan, and T. K. V. Rajinikanth, "Intelligent Semantic Web Search Engines: A Brief Survey," *Int. J. Web Semant. Technol.*, vol. 2, no. 1, pp. 34–42, 2011.
[6] A. Azizan, Z. A. Bakar, N. Khairuddin, and N. L. Saad, "Ontology-Based Information Retrieval : A Review," in *Proceedings of International Symposium on Mathematical Sciences and Computing Research*, 2013, no. December 2013.
[7] C. Mangold, "A survey and classification of semantic search approaches," *Int. J. Metadata, Semant. Ontol.*, vol. 2, no. 1, p. 23, 2007.
[8] A. Mohasseb, M. El-sayed, and K. Mahar, "Automated Identification of Web Queries using Search Type Patterns," in *Proceedings of 10th International Conference on Web Information Systems and Technology*, 2014, pp. 295–304.
[9] A. Awad, M. El-sayed, and Y. El-Sonbaty, "Approximating User'S Intention for Search Engine Queries," in *Proceedings of the 8th International Conference on Web Information Systems and Technologies*, 2012, pp. 422–425.
[10] S. M. Fathalla, Y. F. Hassan, and M. El-Sayed, "A hybrid method for user query reformation and classification," in *Proceedings of 22nd International Conference on Computer Theory and Applications, (ICCTA)*, 2012, pp. 132–138.
[11] S. S. Soliman, M. F. El-Sayed, and Y. F. Hassan, "Semantic Clustering of Search Engine Results," *Sci. World J.*, vol. 2015, 2015.
[12] L. Mostafa, M. Farouk, and M. Fakhry, "An Automated Approach for Webpage Classification," in *Proceedings of the 19th International Conference on Computer Theory and Application*, 2009.
[13] A. Broder, "A Taxonomy of Web Search," *ACM SIGIR Forum*, vol. 36, no. 2, pp. 3–10, 2002.
[14] B. J. Jansen, D. L. Booth, and A. Spink, "Determining the informational, navigational, and transactional intent of Web queries," *Inf. Process. Manag.*, vol. 44, no. 3, pp. 1251–1266, 2008.
[15] G. Salton and C. Buckley, "On the Use of Spreading Activation Methods in Automatic Information Retrieval," in *Proceedings of the 11th annual international ACM SIGIR conference on Research and development in information retrieval*, 1988, pp. 147–160.
[16] Y. Hassan-Montero and V. Herrero-Solana, "Improving Tag-Clouds as Visual Information Retrieval Interfaces," in *Proceedings of International Conference on Multidisciplinary Information Sciences and Technologies, InSciT2006*, 2006.
[17] M. A. Hearst and D. Rosner, "Tag clouds: Data analysis tool or social signaller?," in *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2008.
[18] Y. Chen, R. Santamaría, A. Butz, and R. Therón, "TagClusters : Semantic Aggregation of Collaborative Tags beyond TagClouds," *Lect. Notes Comput. Sci.*, vol. 5531, pp. 56–67, 2009.
[19] A. M. Rinaldi, "Improving tag clouds with ontologies and semantics," in *Proceedings of International Workshop on Database and Expert Systems Applications (DEXA)*, 2012, pp. 139–143.
[20] E. Agichtein, S. Lawrence, and L. Gravano, "Learning search engine specific query transformations for question answering," in *Proceedings of the tenth international conference on World Wide Web - WWW '01*, 2001, pp. 169–178.
[21] H. O. Borch, "On-Line Clustering of Web Search Results," 2006.