CeADAR – Centre for Applied Data Analytics Research

Enterprise Ireland Data Analytics Technology Centre

Search & Retrieval - State of the Art Review

Document Type:	Literature Review
Project Title:	CeADAR
DDN Theme:	1 – Intelligent Analytic Interfaces
Theme Leader:	Sarah Jane Delany (DIT)
DDN Sub-Theme:	Ease of Interaction
Authors:	Eoghan O'Shea, Robert Ross
Document Version:	1.0
Date of Delivery to ISG:	2nd December, 2014
Number of pages:	13

CeADAR is a research partnerhip comprising University College Dublin, University College Cork, and Dublin Institute of Technology.

http://www.ceadar.ie

Contents

1 Executive Summary

In this review, we discuss the state-of-the-art in the use of smart search techniques that can be applied in structured and semi-structured information sources in the Enterprise. Such enterprise information sources are typically forms of domain knowledge that range from customer records to company knowledge that drives business processes. Such knowledge is often of extremely high value but may be only stored in an ad-hoc fashion.

This review briefly summarises the state-of-the-art in practical Information Retrieval and Knowledge Engineering techniques that can be used for managing and querying domain knowledge in an organisation. The first such group of systems are free-text oriented Information Retrieval (IR) systems which are used by large organisations to mine internal documentation and produce query engines that replicate web search experiences in the Enterprise. The second such group of systems are based on the application of formal Knowledge Engineering techniques to the curation and querying of domain specific information stored directly within, or with the assistance of, ontological reasoning systems. We also show that recent trends point towards a hybridisation of these techniques.

This state of the art review is oriented specifically towards search and retrieval in the domain of foods, nutritional information, and recipes. This choice of a specific domain is useful for both structuring and the explanation of the state-of-the-art. However, it should be emphasised that the technologies highlighted have close analogues across other business domains.

2 Introduction

In companies that deal with high volumes of domain specific information, search and retrieval of company information can be a burden. Employees on a regular basis have to query internal knowledge bases or consult with external sources of information to gather sufficient knowledge to solve domain specific tasks.

Take for example the domain specific problem of companies working with many different services or products. Searching through product or service queries can be a key part of the data analysis and testing process within the organisation. Typically this is done based on manually selected search terms, which can be time consuming and ineffective. For example, an analyst in a company with text descriptions of many food items might wish to retrieve all items containing the ingredient "cod". While it is easy to retrieve food items such as "cod fillets" or "cod liver oil", ideally an automated retrieval system should also return "fish fingers", "baccala" and other food items containing cod. Similarly, a user might wish to search for items that belong to the same category, e.g. breakfast cereals in a database of food items, and to combine these items together under a single header. At present, a typical keyword-based search will not retrieve the wealth or depth of information required in such an application. While reference to an internet search engine might assist with such a task, such assistance is ad-hoc, sometimes unreliable, and certainly time-consuming.

Smarter solutions to this search and retrieval task can broadly be split into two categories. The first of these would be to help someone locate relevant information by using document repositories and Information Retrieval mechanism such as Vector Space and Probabilistic Models [25], or graph based analysis [4]. Public search engines such as Google or Bing are the classic examples of systems which pull together classic information retrieval techniques with graph based analysis to allow querying of unstructured (and sometimes semi-structured) information. While early generations of search engines were often very literal in analysis, public search engines now answer queries containing everything from poor spelling to metanyms. While public search engines are extremely powerful, it is non-trivial to customise them to specific tasks or apply them to proprietary or otherwise sensitive information sources.

The second approach to a smarter solution to the search and retrieval problem is through the use of formal knowledge engineering or semantic search techniques. At the core of such techniques are two premises: (a) that the meaning of words rather than literal representations are searched for; and (b) that a structured form of representation is used directly to capture domain knowledge or to assist in the processing of queries with respect to a knowledge base. Such techniques are based on earlier work on Knowledge Representation & Reasoning (see [3] for more information), but more recently have been developed in the R&D context of the semantic web, e.g. see [6], [11] or [2]. In a recent review of semantic web search engines [24] identifies four different approaches to semantic search: (i) contextual analysis; (ii) reasoning; (iii) natural language understanding; and (iv) the use of ontologies as an underlying technology. These technologies when taken together enable powerful application specific search and query while being more constrained than full-blown web search, and can be used directly within custom applications.

In this state of the art we will review a range of applied technologies that can enable advanced search and retrieval. We begin with Information Retrieval derived methods in Section 3 before considering the Knowledge Engineering derived methods in Section 4, and hybrid methods in Section 5. In Section 6 we consider the domain of food and nutrition and review publicly available resources that can be used for search and retrieval type queries.

3 Information Retrieval in the Enterprise

In this section we consider Smart Search & Retrieval systems based directly on Information Retrieval methods.

3.1 Content & Document Management Systems

Information Retrieval based on either Text Mining or Graph based search techniques are the core technologies behind web search engines and large scale systems for searching unstructured text documents. The technical theory and details of Information Retrieval systems have been covered extensively in scholarly publications elsewhere and will not be repeated here. Readers who are interested in an approachable introduction to IR theory are directed towards [22].

While IR systems form the basis of commercial search engines that operate on the World Wide Web or more generally the Internet, the same search techniques have also been used in a normal of commercial and open-source tools that may be used for searching intranets and company specific document repositories. Generally these systems are collectively termed Document Management Systems or Frameworks. This is different from Content Management Systems which is a term that is typically used to describe the back-end of a commercial publication website such as that operated by a news organisation.

In terms of open solutions to Information Retrieval, the widest used and built upon Document Management Systems are based on Apache Lucene[1]. Lucene is an Information Retrieval Software Library supported by the Apache Foundation and available in a number of different programming languages. Lucene provides document index and search functionality that can be used as the kernel of either an internet search engine or a local search engine. Many different document types can be indexed - Lucene does not assume that the docs to be indexed are HTML files. However, Lucene does not contain crawling or parsing functionality. Other projects add this layer of functionality on top of Lucene. The most notable of these is Apache Solr[2] which provides an enterprise search server that can be freely deployed within an organisation.

A number of other projects are available that build on top of or are similar in nature to Lucene and its derived projects. OpenKM[3] for example is an open Knowledge Base application that is built on top of the open-source Apache Jackrabbit document management service[4]. Project

[1] http://lucene.apache.org/
[2] http://lucene.apache.org/solr/
[3] http://www.openkm.com/en/
[4] http://jackrabbit.apache.org/

Lemur[5] meanwhile is a distinct research and development project that can be used "for developing search engines, text analysis tools, browser toolbars, and data resources in the area of IR". Lemur, and its search engine Indri, has full text indexing capabilities including the use of stemmers and stop word filters. a structured query language, clustering, synonyms, and symmetrisation. In terms of commercial search, ElasticSearch[6] is a commercial distributed search engine backed by a company with substantial seed funding which is moving into the Enterprise search space.

3.2 Corporate Search in the Enterprise

The advantage of IR projects such as Lemur and Solr is that they allow sophisticated commercial search solutions to be built within an enterprise without risk of sensitive commercial knowledge being exposed to public search engines. Nevertheless Commercial Search Companies such as Google or Microsoft's Search division have provided commercial solutions that can be deployed safely within an enterprise to enable Smart Search.

The Google Search Appliance (GSA)[7] is a commercial solution that allows a medium to large enterprise to deploy a Google search style solution to internal company documents. The GSA combines Google Software and Dell hardware to provide a concrete solution. The Google software directly supports advanced search techniques including for example sensitivity to synonyms use. Quoting GSA's documentation:

> Synonyms that give alternate terms for your search. E.g. when user types "cell phone" Search will add suggestions e.g. "mobile phone" to the result set

Microsoft's Enterprise Search solutions are built around the MS SharePoint[8] application framework rather than MS Bing. While SharePoint incorporates a wide family of technologies, at its heart it provides an Enterprise Document Management and Search Service that can be deployed to enable smart search and retrieval. The level of technological capability in search is however very much dependent on the specific variant of SharePoint which is deployed. Search capabilities range from minimal document indexing in the Microsoft SharePoint Foundations to a much more complete solutions in the SharePoint Enterprise deployment. Underlying search technologies include the FAST enterprise search technologies which were acquired in 2008, and the Microsoft Search Server which is also available as a standalone component.

4 Knowledge Engineering Based Approaches

IR based technologies generally assume no underlying structure in the data. Knowledge Engineering based solutions on the other hand assume a high degree of underlying structure in the data and take advantage of this structure to provide a higher level of complexity in answers than can be provided in IR based solutions. This improvement comes however at a higher computational cost and fragility in design. This section first briefly introduces the

[5] http://www.lemurproject.org/

[6] http://www.elasticsearch.org/

[7] https://www.google.com/work/search/products/gsa.html

[8] http://office.microsoft.com/en-ie/sharepoint/

fundamentals of ontological reasoning and representation, before looking specifically at semantic search and applications and APIs that are available for Knowledge Engineering based search.

4.1 Ontology Reasoning & Representation

An ontology is an explicit specification of a conceptualisation [9]. Ontologies were envisioned as a means to represent knowledge that could be understood, used and shared among distributed applications and agents [8]. An ontology provides a vocabulary for representing and communicating knowledge about some topic and the set of relationships that hold among the terms in that vocabulary [20]. To deal with the task of accessing linked ontologies, ontology languages such as the Resource Description Framework (RDF) and the Web Ontology Language (OWL) were developed[9]. Using these underlying technologies systems can be developed which allow search over one or a number of (linked) ontologies in particular domains.

The topic of ontology design and use is closely related to research and development topics such as *Linked Data* and the *Semantic Web*. The Semantic Web was once popularly seen as the next generation of the world wide web which was assumed to be based on highly structured data as distinct from the current unstructured data that constitutes 99% of the current web. More correctly and recently the semantic web is seen as a development of standard interconnected services based on structured information that can be used to facilitate a range of web based applications. Linked Data is the most concrete manifestation of the real 'semantic web'. Linked Data is about using the Web to connect related data that was not previously linked, or using the Web to lower the barriers to linking data currently linked using other methods. Wikipedia defines Linked Data as "a term used to describe a recommended best practice for exposing, sharing, and connecting pieces of data, information, and knowledge on the Semantic Web using URIs and RDF".

Ontologies and related search techniques allow us reason on and query complex knowledge sources. In short inference techniques based on rule-based systems allow us to infer knowledge that is not represented explicitly in the underlying data set. The classic example of such explicit information is inferring familial relationships from a base set of family relationships, i.e., inferring a grandmother relationship holding between X and Z from from base statements relating X and Y, and Y and Z with for example a Mother and Father relationship directly. Such inference based search capabilities can therefore allow us to answer a wider range of queries than can be handled through only literal querying of a knowledge base.

4.2 Semantic Search

The use of ontologies to provide or facilitate search results is an example of *semantic search*. Generally semantic search can be categorised into a number of different categories such as document oriented search, entity and knowledge oriented search, multimedia oriented

[9] We refer the reader to an excellent review of linked data and the semantic web at (LinkedDataTools.com) (`http://www.linkeddatatools.com/semantic-web-basics`) for a thorough introduction to the underlying technologies.

search, relation oriented search, semantic analytics and mining based search ([26]).

Rather than covering the entire range of semantic search types, we limit ourselves to *relation oriented search*. We refer the reader to reviews such as [17], [10], [7], [21], [18], [26], [24] for a more in depth exploration of the subject area.

Much early research on the semantic search dealt with augmenting traditional text search with semantic techniques, such as using ontological techniques to increase recall or precision [18]. Some work such as that by [19] and [5] made use of the WordNet (reference) ontology to expand keyword search terms to their synonym and meronym sets. In a similar vein [14] developed a wrapper for WordNet (Clever Search) that allows a user to select a particular meaning for a word in a clarification dialogue before being added to the search keywords.

CIRI [1] has been developed as an ontological front-end, allowing users to search through available ontologies and select concepts from them (by opening and closing branches in a tree hierarchy of concepts), which can then be used to constrain the search. The search is carried out using standard text search engines based on the chosen concepts (selected concepts and expansion level). CIRI uses the Protege ontology editor for ontology creation.

Aqualog [16] meanwhile is a question answering system which tries to make sense of natural language queries by looking at the structure of the ontology and the information available on the semantic web, as well as using string similarity matching, generic lexical resources such as WordNet and a domain-dependent lexicon ontained through the use of a learning mechanism. Similarly [15] developed the SemSearch search engine for the semantic web. This system provides a Google-like query interface where the user can specify the queried subject and a combination of keywords. The query is then created by matching the search keywords against the subject, predicate or object(instance) in the knowledge base using a number of different combinations and the result returned.

More recently [20] have developed a domain specific (agricultural information) ontology based semantic search engine. The heart of this system is a mapping technique between instance and class which finds the right result while preventing the inaccurate, non-meaningful results. Also recently [12] discusses the development of a domain specific semantic search engine based on the travel domain that uses a web crawler to perform focused crawling of the web to create a knowledge base set up as an OWL ontology. The web crawling was carried out using Nutch[10], while the ubiquitous Protege Ontology Editor[11] was used for the building of the ontology.

4.3 Applications & APIs

Ontological representation and querying engines can be built independently of domains, but are most commonly investigated with respect to particular business or research domains. In this section we briefly review a number of the more prominent examples of domain specific ontology application and API development. The topic of Food Based Knowledge Resources will be handled in detail in the next section.

[10] `http://nutch.apache.org/`

[11] `http://protege.stanford.edu/`

- Financial Industry Business Ontology[12]: The Financial Industry Business Ontology "is a collaborative effort among industry practitioners, semantic technology experts and information scientists to standardise the language used to precisely define the terms, conditions, and characteristics of financial instruments; the legal and relationship structure of business entities; the content and time dimensions of market data; and the legal obligations and process aspects of corporate actions."
- Ontology Systems[13]: "Ontology Systems applies the power, simplicity and speed of search to business applications across the enterprise. This means the ability to search and link applications, databases, files and spreadsheets."
- Ontotext [14]: provides a complete set of semantic technologies to identify meaning across diverse databases and massive amounts of unstructured data. Ontotext blends text mining, powerful structured queries, semantic annotation and semantic search with an RDF graph database. It allows companies the ability to integrate their ontologies with ontotext.
- Edaman [15]: Edaman in conjunction with OntoText uses semantic technology to organise information about food.
- GoodRelations [16]: is a lightweight ontology for annotating offerings and other aspects of e-commerce on the Web. GoodRelations is the only OWL DL ontology officially supported by both Google and Yahoo.
- eClassOWL [17]: is the Web Ontology for Products and Services. eClassOWL is an OWL ontology for describing the types and properties of products and services on the Semantic Web (also known as the "Web of Linked Data"). eClassOWL is meant to be used in combination with the GoodRelations ontology for e-commerce.
- Yummly [18] Yummly is a recipe platform that utilises semantic search web technology to create robust search, filter, and recommendation features [19]. Yummly has created a food ontology and uses natural language processing to extract information about a food's ingredients.

5 Semantic Search In IR

Rather than considering Smart Search to be a choice between either Information Retrieval or Knowledge Management based approaches, the reality is that the most versatile smart search solutions will likely make use of a combination of these underlying technologies. While ad-hoc combinations are certainly possible, interesting hybrid solutions have already been developed and deployed. In this section we will focus on SmartLogic's Semaphore as a case study in hybrid approaches to smart search.

Another more prominent example of hybridisation in Smart Search is IBM's Watson and derived software platform. The IBM Watson solutions were built upon a vast collection of

[12] http://www.edmcouncil.org/financialbusiness
[13] http://www.ontology.com/
[14] http://www.ontotext.com/
[15] https://www.edamam.com/
[16] http://www.heppnetz.de/projects/goodrelations/
[17] http://www.heppnetz.de/projects/eclassowl/
[18] http://www.yummly.com/
[19] http://tinyurl.com/lk9sms6

data. This data ranged from vast amounts of unstructured content – including the English Wikipedia – to highly structured content. Watson made use of numbers of individual answering experts to determine the most probable solutions to problems. These solutions are thus potentially based on both unstructured and structured search results. This collaborating experts view of hybridisation is a different view of smart search which we believe will have significant consequences into the future.

5.1 Semaphore Enterprise Semantic Platform

The Semaphore Enterprise Semantic Platform[20] allows developers to improve the performance of common search and document management services with sophisticated meta-data and concept processing capabilities. Semaphore augments traditional information management systems, e.g. content management systems, by extracting the most important topics in a document, converting these into a model (a list, taxonomy or ontology), and then using this model to classify content and enrich it with metadata. Semaphore consists of four core modules: (i) Ontology Server & Manager (for the development and management of ontologies), (ii) the Advanced Linguistics Pack (which provides Natural Language Processing tools such as text mining and entity extraction, (iii) the Classification Server (providing accurate meta-data tagging of content), and (iv) the Semantic Enhancement Server (enhances search engines such as Microsoft Sharepoint or Apache Lucene/Solr with semantics to improve search capabilities). See the white paper on Semaphore from Smartlogic at [21]. Semaphore, uses semantics, text analytics and visualisation software technologies to perform five primary functions [22]: (i) Ontology and taxonomy management; (ii) Auto-classification of unstructured data; (iii) Text analysis (including entity, fact and sentiment extraction); (iv) Metadata management; and (v) Content visualisation. The metadata that Semaphore is able to extract can be used in a large number of ways, e.g., text analytics, content visualisation, decision support using information locked-up in content, etc. Moreover, Semaphore has out of the box integration with a range of platforms including SharePoint, Marklogic, Solr, LucidWorks, the Google Search Appliance, and OpenText.

6 Domain Resource Analysis: Food & Nutrition

Whether based on IR or KB based solutions, smart search requires underlying data sources to operate over. In this section we will briefly summarise information sources that will be of particular use to any project working in the food domain.

A number of ontologies have been developed to specify types of food and the construction of recipes. By and large these are high-level ontologies which do not specify recipes explicitly, but rather provide the class and relation definitions that can be used to specify specific recipes. In the listing below we will specifically reference where a project includes a substantial number of recipe instances.

- Kolchin & Zamula[13] describe a food product ontology built as an extension to the widely used Good Relations ontology discussed in a previous section. Their ontology

[20] http://www.smartlogic.com/home/products/products-overview
[21] http://www.barbador.com/assets/kb-smartlogic/Semaphore-White-Paper.pdf
[22] see footnote 20

defines a number of new classes and properties related to the food domain including "Food", "Ingredient", "Food Additive" and "Food Category". They evaluate their designed system with SPARQL queries built from it. The ontology itself is publically available [23].

- Snae & Bruckner[23] present the design of a system for food or menu planning called the Food-Oriented Ontology-Driven System (FOODS). FOODS combines a food ontology with a user interface and has been designed to find the appropriate dish for a user. Unfortunately the related ontology does not appear to be accessible online.
- The Linked Recipe schema is intended to support the expression of recipes using RDF technology (`http://linkedrecipes.org/schema`). It contains a number of classes such as "Ingredients", "Food", "Dish", "Recipe", etc.
- Linked Open Vocabulary Food Ontology (`http://lov.okfn.org/dataset/lov/details/vocabulary_food.html`). This ontology is used by the Open Food Facts (`http://datahub.io/dataset/open-food-facts`) dataset.
- BBC Food Ontology (`http://www.bbc.co.uk/ontologies/fo`)
- LIRMM ontologies publishing platform: Food Ontology (`http://data.lirmm.fr/ontologies/food`). This ontology models the Food domain. It allows one to describe ingredients and food products.
- BEVON: Beverage Ontology (`http://rdfs.co/bevon/latest/html`)
- Semantic Diet (`http://semanticdiet.com/data.event`) Semantic Diet uses semantic web technologies to bring together information about food and nutrition from the U.S. Dept. of Agriculture (USDA), recipes contributed by users and crawled off the web, and personal dietary needs.
- dbpedia - ingredient ontology (`http://dbpedia.org/ontology/ingredient`). This is part of the large dbpedia knowledge base, `http://dbpedia.org/About`.

While not ontologies themselves, the USDA National Nutrient Database for Standard Reference (`http://ndb.nal.usda.gov/`) and AGROVOC (`http://aims.fao.org/agrovoc`), a multilingual thesaurus made by the Food and Agriculture Organisation of the United Nations (FAO), may provide useful information for the construction of standalone ontologies based in the food/drink domain.

Other sources of useful information on recipes, useful for the construction of standalone ontologies, are the APIs provided by the recipe websites, Yummly (Yummly Recipes API[24]) and Big Oven (Big Oven Recipe API[25]).

7 Conclusions

In this review of the state-of-the-art in the field of search & retrieval, we outlined the different research and/or systems presently available that augment traditional keyword search with semantic search techniques. In particular we focused on Information Retrieval and ontology based semantic web search engines that can be applied to the domain of food and nutrition. The choice of IR or Knowledge Representation engine should depend on the nature of internal data held by a company. Where information can be structured, a Knowledge Engineering

[23] `https://raw.githubusercontent.com/ailabitmo/food-ontology/master/food.owl`
[24] https://developer.yummly.com/
[25] http://api.bigoven.com/

approach remains most attractive. For higher volumes of data that are less suited to curation into a structured or semi-structured organisation the Information Retrieval option remains most appropriate. In the long run we believe that hybridisation through for example collaborating expert systems such as IBM's Watson architecture will provide a natural means for integrating Information Retrieval frameworks with Expert Systems.

References

[1] Eija Airio, Kalervo Järvelin, Pirkko Saatsi, Jaana Kekäläinen, and Sari Suomela. Ciri-an ontology-based query interface for text retrieval. In *Web Intelligence: Proceedings of the 11th Finnish Artificial Intelligence Conference*, 2004.

[2] Sören Auer and Jens Lehmann. Creating knowledge out of interlinked data. *Semantic Web*, 1(1):97–104, 2010.

[3] Ronald Brachman and Hector Levesque. *Knowledge Representation and Reasoning*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2004.

[4] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. In *Seventh International World-Wide Web Conference (WWW 1998)*, 1998.

[5] David Buscaldi, Paolo Rosso, and Arnal Emilio Sanchez. A wordnet-based query expansion method for geographical information retrieval. In *Working notes for the CLEF workshop*, 2005.

[6] Mathieu d'Aquin, Enrico Motta, Marta Sabou, Sofia Angeletou, Laurian Gridinoc, Vanessa Lopez, and Davide Guidi. Toward a new generation of semantic web applications. *Intelligent Systems, IEEE*, 23(3):20–28, 2008.

[7] Hai Dong, Farookh Khadeer Hussain, and Elizabeth Chang. A survey in semantic search technologies. *Second IEEE International Conference on Digitial Ecosystems and Technologies (IEEE DEST 2008)*, 2008.

[8] Miriam Fernandez, Iván Cantador, Vanesa López, David Vallet, Pablo Castells, and Enrico Motta. Semantically enhanced information retrieval: an ontology-based approach. *Web Semantics: Science, Services and Agents on the World Wide Web*, 9(4):434–452, 2011.

[9] Thomas R Gruber. A translation approach to portable ontology specifications. *Knowledge acquisition*, 5(2):199–220, 1993.

[10] Ramanathan Guha, Rob McCool, and Eric Miller. Semantic search. In *Proceedings of the 12th international conference on World Wide Web*, pages 700–709. ACM, 2003.

[11] Pascal Hitzler and Frank van Harmelen. A reasonable semantic web. *Semantic Web*, 1(1):39–44, 2010.

[12] S Sowmya Kamath, Dhivya Piraviperumal, Garima Meena, Srijana Karkidholi, and Keerthana Kumar. A semantic search engine for answering domain specific user queries. In *Communications and Signal Processing (ICCSP), 2013 International Conference on*, pages 1097–1101. IEEE, 2013.

[13] Maxim Kolchin and Dmitry Zamula. Food product ontology: Initial implementation of a vocabulary for describing food products. In *Proceedings of the 14th Conference of Fruct Association*, pages 191–196, 2013.

[14] Peter M Kruse, André Naujoks, Dietmar Rösner, and Manuela Kunze. Clever search: A wordnet based wrapper for internet search engines. *Working Notes for the CLEF Workshop*, 2005.

[15] Yuangui Lei, Victoria Uren, and Enrico Motta. Semsearch: A search engine for the semantic web. In *Managing Knowledge in a World of Networks*, pages 238–245. Springer, 2006.

[16] Vanessa Lopez, Michele Pasin, and Enrico Motta. Aqualog: An ontology-portable question answering system for the semantic web. In *The Semantic Web: Research and Applications*, pages 546–562. Springer, 2005.

[17] G Madhu, Dr A Govardhan, and Dr TV Rajinikanth. Intelligent semantic web search engines: A brief survey. *arXiv preprint arXiv:1102.0831*, 2011.

[18] E. Makela. Survey of semantic search research. In *Proceedings of the seminae on knowledge management on the semantic web*, 2005.

[19] Dan I Moldovan and Rada Mihalcea. Using wordnet and lexical operators to improve internet searches. *IEEE Internet Computing*, 4(1):34–43, 2000.

[20] Debajyoti Mukhopadhyay, Aritra Banik, Sreemoyee Mukherjee, Jhilik Bhattacharya, and Young-Chon Kim. A domain specific ontology based semantic web search engine. *arXiv preprint arXiv:1102.0695*, 2011.

[21] Walter Renteria-Agualimpia, Francisco J López-Pellicer, Pedro R Muro-Medrano, Javier Nogueras-Iso, and F Javier Zarazaga-Soria. Exploring the advances in semantic search engines. In *Distributed Computing and Artificial Intelligence*, pages 613–620. Springer, 2010.

[22] Mark Sanderson and W Bruce Croft. The history of information retrieval research. *Proceedings of the IEEE*, 100(Special Centennial Issue):1444–1451, 2012.

[23] Chakkrit Snae and M Bruckner. Foods: a food-oriented ontology-driven system. In *Digital Ecosystems and Technologies, 2008. DEST 2008. 2nd IEEE International Conference on*, pages 168–176. IEEE, 2008.

[24] G Sudeepthi, G Anuradha, and M Surendra Prasad Babu. A survey on semantic web search engine. *International Journal of Computer Science Issues*, 9, 2012.

[25] Howard R Turtle and W Bruce Croft. A comparison of text retrieval models. *The computer journal*, 35(3):279–290, 1992.

[26] Wang Wei, Payam M Barnaghi, and Andrzej Bargiela. Search with meanings: an overview of semantic search systems. *Int. J. Communications of SIWN*, 3:76–82, 2008.