

## Chapter 16

# A User–Aware and Semantic Approach for Enterprise Search

**Giacomo Cabri**

*University of Modena and Reggio Emilia, Italy*

**Riccardo Martoglia**

*University of Modena and Reggio Emilia, Italy*

### ABSTRACT

*This article describes how in addition to general purposes search engines, specialized search engines have appeared and have gained their part of the market. An enterprise search engine enables the search inside the enterprise information, mainly web pages but also other kinds of documents; the search is performed by people inside the enterprise or by customers. This article proposes an enterprise search engine called AMBIT<sup>1</sup>-SE that relies on two enhancements: first, it is user-aware in the sense that it takes into consideration the profile of the users that perform the query; second, it exploits semantic techniques to consider not only exact matches but also synonyms and related terms. It performs two main activities: (1) information processing to analyse the documents and build the user profile and (2) search and retrieval to search for information that matches user's query and profile. An experimental evaluation of the proposed approach is performed on different real websites, showing its benefits over other well-established approaches.*

### INTRODUCTION

Enterprises produce and rely on a large amount of information. A small part of information is available by public web sites, while the most part is exploited by employees of the enterprise itself by means of intranet, and by customers of the enterprise who have access to some information for business purposes.

In this scenario, the capability of searching for needed information plays a fundamental role. On the one hand, enabling internal employees to find the needed information in a short time is not only useful to speed up their work, but also to avoid or decrease the frustration of long and unsuccessful searches. On the other hand, precise and relevant answers to customers that exploit the company web sites for both searching and interacting can grant a high degree of customer satisfaction.

DOI: 10.4018/978-1-7998-0951-7.ch016

In this scenario, the authors point out two aspects that can improve the use of search engines in an enterprise context by providing more relevant search results: user-awareness and semantics. Existing studies and surveys in general information management contexts have highlighted the benefits that can be brought to search results by the former (Xiang et al., 2010) and latter (Mangold, 2007). User-awareness means to exploit the knowledge of the user in terms of profile and context to effectively tailor the search on the base of the available information. Semantics can be useful to overcome the limitations of a syntactic approach, which is often exploited but does not consider similar pieces of information expressed in different ways. As far as the authors know, there are no enterprise search engines that exploit both aspects in a single approach.

Starting from the above considerations, the general semantic foundations introduced in (Martoglia, 2015) are exploited to proceed towards the goal of achieving a user-aware semantic enterprise search engine. The search engine is called AMBIT-SE (AMBIT Search Engine). It is not a generic search engine, but a search engine dedicated to searches in an enterprise scenario. The AMBIT-SE approach improves search results by:

- Taking advantage of textual information, including user information. Indeed, text is the primary component of the documents that should be presented / suggested to users, and also one of the main information characterizing user profiles. Consider, for instance, the contents of user browsing history, the description of users' interests, and so on;
- Exploiting semantic techniques: instead of a pure syntactic matching between the query keywords and the words in the available documents, it relies on their meaning and takes into account synonyms and related terms.

The approach presented in this paper brings the following novel contributions with regard to the initial idea of an enterprise search engine sketched in (Cabri, 2016) and to the state of the art:

- Differently from the state of the art on available enterprise search engines, it is able to combine semantics and user-awareness without requiring any manual work (e.g. for annotating documents, describing user profiles, etc.). Novel semantic and user-aware techniques allow the engine to go beyond standard syntactic search in a completely automatic way;
- The semantic text analysis techniques are adapted and refined from previous authors' studies on the effectiveness of semantic text management in specific subject areas such as software engineering (Bergamaschi et al., 2015; Martoglia, 2011), agricultural (Beneventano et al., 2016) and user-centric cultural enhancement data (Martoglia, 2015). In the presented approach, the techniques are generalized to work in a non-specialized enterprise search setting with general purpose ontologies and new weighting schemes, allowing them to be directly compared to the new class similarity contribution;
- The strength of the semantic text analysis contribution is added to the new contribution given by semantic categorization. Categorization classifies documents on the basis of a well-known taxonomy (defined by IPTC<sup>2</sup>) in order to provide improvements in the retrieval effectiveness. To this end, a novel class similarity metric is introduced, with a weighting scheme exploiting the novel concept of inverse (document) class frequency;

18 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/a-user-aware-and-semantic-approach-for-enterprise-search/239942](http://www.igi-global.com/chapter/a-user-aware-and-semantic-approach-for-enterprise-search/239942)

## Related Content

---

### The French Digital Kitchen: Implementing Task-Based Language Teaching Beyond the Classroom

Paul Seedhouse, Anne Preston, Patrick Olivier, Dan Jackson, Philip Heslop, Thomas Plötz, Madeline Balaamand Saandia Ali (2014). *Computational Linguistics: Concepts, Methodologies, Tools, and Applications* (pp. 968-986).

[www.irma-international.org/chapter/the-french-digital-kitchen/108760](http://www.irma-international.org/chapter/the-french-digital-kitchen/108760)

### Newness and Givenness of Information : Automated Identification in Written Discourse

Philip M. McCarthy, David Dufty, Christian F. Hempelmann, Zhiqiang Cai, Danielle S. McNamaraand Arthur C. Graesser (2012). *Applied Natural Language Processing: Identification, Investigation and Resolution* (pp. 457-478).

[www.irma-international.org/chapter/newness-givenness-information/61065](http://www.irma-international.org/chapter/newness-givenness-information/61065)

### LinkedVis an Information Visualisation Toolkit for RDF Data

Antonio Garroteand María N. Moreno García (2014). *Computational Linguistics: Concepts, Methodologies, Tools, and Applications* (pp. 799-815).

[www.irma-international.org/chapter/linkedvis-an-information-visualisation-toolkit-for-rdf-data/108752](http://www.irma-international.org/chapter/linkedvis-an-information-visualisation-toolkit-for-rdf-data/108752)

### Model of the Empirical Distribution Law for Syntactic and Link Words in "Perfect" Texts

Pavel Makagonov (2015). *Modern Computational Models of Semantic Discovery in Natural Language* (pp. 27-47).

[www.irma-international.org/chapter/model-of-the-empirical-distribution-law-for-syntactic-and-link-words-in-perfect-texts/133874](http://www.irma-international.org/chapter/model-of-the-empirical-distribution-law-for-syntactic-and-link-words-in-perfect-texts/133874)

### Semantics of Techno-Social Spaces

Sergey Maruev, Dmitry Stefanovskyand Alexander Trousov (2015). *Modern Computational Models of Semantic Discovery in Natural Language* (pp. 204-234).

[www.irma-international.org/chapter/semantics-of-techno-social-spaces/133880](http://www.irma-international.org/chapter/semantics-of-techno-social-spaces/133880)