

Chapter 1

Improving Domain Searches through Customized Search Engines

Cecil Eng Huang Chua
University of Auckland, New Zealand

Roger Hsiang-Li Chiang
University of Cincinnati, USA

Veda C. Storey
Georgia State University, USA

ABSTRACT

Search engines are ubiquitous tools for seeking information from the Internet and, as such, have become an integral part of our information society. New search engines that combine ideas from separate search engines generally outperform the search engines from which they took ideas. Designers, however, may not be aware of the work of other search engine developers or such work may not be available in modules that can be incorporated into another search engine. This research presents an interoperability architecture for building customized search engines. Existing search engines are analyzed and decomposed into self-contained components that are classified into six categories. A prototype, called the Automated Software Development Environment for Information Retrieval, was developed to implement the interoperability architecture, and an assessment of its feasibility was carried out. The prototype resolves conflicts between components of separate search engines and demonstrates how design features across search engines can be integrated.

1. INTRODUCTION

Arguably, the most important driver in the growth of the Internet and e-commerce is the existence

of easy to use and effective search engines. This makes search engines an integral part of the world economy. Unfortunately, there is no single best search engine for all contexts. Algorithms suited for a domain such as medical research (Mao & Tian, 2009) are not effective for searching the Se-

DOI: 10.4018/978-1-60960-595-7.ch001

mantic Web (Li, Wang, & Huang, 2007). Similarly, algorithms optimized for the Semantic Web are not as efficient for searching blogs as specialized blog search algorithms (Thelwall & Hasler, 2007). It is also costly to develop new search engines for emergent domains from scratch. However, certain aspects of search engines are shareable across domains. For example, search interfaces for blogs and semantic search can be similar. Unfortunately, search technologies developed by one researcher cannot be easily combined with technologies developed by another, resulting in wasted efforts when developing advanced search technologies.

The objective of this research is to propose an interoperability architecture to help developers build customized search engines by combining existing and developing technologies (Papazoglou & van den Heuvel, 2007). There are two tasks involved. The first is to analyze and decompose existing search engines into a set of self-contained components, and to create a meaningful set of categories in which to classify them. This is consistent with many engineering disciplines that exploit how various parts of distinct tools perform the same task (Bucchiarone, Pelliccione, Polini, & Tivoli, 2006). For example, in Google, the vector query interface, the content repository of billions of web pages, and the PageRank search algorithm (Brin & Page, 1998; Page, Brin, Motwani, & Winograd, 1998) can all be self-contained search engine components. The second task is to craft an interoperability architecture as the basis for building customized search engines.

The contribution of this research is to enable search engine developers to identify and integrate self-contained search engine components based on the search needs of a particular domain, instead of building a domain-specific search engine from scratch. Component integration must be achieved with the support of intelligent interfaces to bridge components.

Software architecture is usually validated via a case study (Dashofy, Hoek, & Taylor, 2005;

Hayes-Roth, Pflieger, Lalanda, Morignot, & Balabanovic, 1995; Xu, Yang, & Huang, 2004). This research, however, tries to demonstrate how the architecture can be designed and implemented, and how the deliverable can simulate and behave like existing software artifacts (customized search engines). In this way, the evaluations performed on search engines developed using the proposed architecture share performance characteristics of more traditional search engines. The contributions of the research are both the creation and evaluation of the proposed architecture. The evaluation of the proposed architecture is based on the following.

- *Feasibility*: the architecture can be applied and used to build customized search engines;
- *Robustness*: the architecture encompasses a wide range of search engines and demonstrates that components of existing search engines can be easily assembled to build a customized search engine; and
- *Usefulness*: the customized search engines built improve retrieval accuracy.

The research is carried out in three stages.

1. **Componentization of Search Engines.** Existing search engines are analyzed and decomposed into a set of self-contained components. Section 2 presents a classification taxonomy of these self-contained search engine components.
2. **Interoperability Architecture.** The interoperability architecture for building customized search engines is developed. This involves identifying the relationships between search engine components, and proposing mechanisms to resolve incompatibilities. Section 3 presents the proposed architecture.
3. **Prototype and Evaluation Experiments.** A prototype is developed to evaluate the interoperability architecture. The prototype

20 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/improving-domain-searches-through-customized/54422

Related Content

Application of Artificial Intelligence Technology in Optimizing Control Parameters of Traffic Signal Group Systems

Xiuli Liand Huachang Miao (2024). *International Journal of Intelligent Information Technologies* (pp. 1-19). www.irma-international.org/article/application-of-artificial-intelligence-technology-in-optimizing-control-parameters-of-traffic-signal-group-systems/355013

Ensemble of SVM Classifiers for Spam Filtering

Ángela Blancoand Manuel Martín-Merino (2009). *Encyclopedia of Artificial Intelligence* (pp. 561-566). www.irma-international.org/chapter/ensemble-svm-classifiers-spam-filtering/10303

Intelligent Agents for E-Learning

Ralf Brunsand Jürgen Dunkel (2008). *Intelligent Information Technologies and Applications* (pp. 25-42). www.irma-international.org/chapter/intelligent-agents-learning/24258

Artificial Immune Systems for Anomaly Detection in Ambient Assisted Living Applications

Sebastian Bersch, Djamel Azzi, Rinat Khusainovand Ifeyinwa E. Achumba (2013). *International Journal of Ambient Computing and Intelligence* (pp. 1-15). www.irma-international.org/article/artificial-immune-systems-for-anomaly-detection-in-ambient-assisted-living-applications/101949

Distributed Data Real-Time Transaction Calculation Based on Collaborative Optimization and Multi-Objective Genetic Algorithm

Li Liao (2024). *International Journal of Intelligent Information Technologies* (pp. 1-16). www.irma-international.org/article/distributed-data-real-time-transaction-calculation-based-on-collaborative-optimization-and-multi-objective-genetic-algorithm/333632