Building Customized Search Engines: An Interoperability Architecture

Cecil Eng Huang Chua, Nanyang Technological University, Singapore
Roger H. L. Chiang, University of Cincinnati, USA
Veda C. Storey, Georgia State University, USA

ABSTRACT

Search engines are essential, ubiquitous tools for seeking information from the Internet. Prior research has also demonstrated that combining features of separate search engines often improves retrieval performance. However, such feature combination is often difficult, because developers don’t consider other developers when building their software. To facilitate the development of search engines, we propose a customized search engine approach to integrating appropriate components from multiple search engines. This article presents an interoperability architecture for building customized search engines. To achieve this, authors analyze existing search engines and decompose them into self-contained components that are classified into six categories. The Automated Software Development Environment for Information Retrieval (ASDEIR) was developed as the prototype of the proposed interoperability architecture to test its feasibility, robustness, and usefulness. ASDEIR incorporates intelligent features that detect and attempt to resolve conflicts between components. [Article copies are available for purchase from InfoSci-on-Demand.com]

Keywords: Componentization; Design Architecture; Interoperability; Search Engine

INTRODUCTION

With the continual increase in information available on the World Wide Web, and the growth of global commerce, search engines continue to play an integral part in the world economy. Substantial research on information retrieval focuses on improving existing search algorithms, or developing new
and better search algorithms, query expansion methods, and meta-search amalgamation algorithms. Search engines, many incorporating intelligent algorithms, have been developed to support a variety of information search needs. However, much research effort is wasted, because search technologies developed by one researcher cannot be easily combined with technologies developed by another.

This research develops an interoperability architecture that helps developers build customized search engines by combining technologies (Papazoglou & Heuvel, 2007). There are two objectives of this research. The first objective is to componentize search engines by analyzing and decomposing existing search engines into a set of self-contained components, which we classify into six categories. This approach is consistent with many engineering disciplines that recognize how various parts of distinct tools perform the same task (Bucchiarone, et al., 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, et al., 1998) can all be componentized as self-contained search engine components.

The second objective is to propose an interoperability architecture as the foundation for building customized search engines. We advocate a customized search engine approach where search engine developers can 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. By necessity, the component integration must be achieved with the support of intelligent interfaces to bridge these components.

A software architecture is commonly validated via a case study (Dashofy, et al., 2005; Hayes-Roth, et al., 1995; Xu, et al., 2004). Our research tries to go one step further. It not only recounts how the architecture can be designed and implemented, but also how the deliverable can simulate and behave like existing software artifacts (i.e., customized search engines). In other words, evaluations performed on search engines developed using our architecture share performance characteristics of more traditional search engines. The contributions of the research are to both create and evaluate the proposed architecture. We evaluate and demonstrate that the proposed architecture is:

- Feasible: it can be applied and used to build customized search engines;
- Robust: it 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
- Useful: the customized search engines built could improve retrieval accuracy through a simulation experiment.
adapt@Agent.Hospital: Agent-Based Organization & Management of Clinical Processes
www.igi-global.com/article/adapt-agent-hospital/2378?camid=4v1a