Vision for Search String Translator

Preliminary Draft

September 26 - 27, 2004

Introduction

We envision a Search String Translator (SST) that will translate search strings entered in a common format into equivalent search strings for various target search engines. Target engines could include search engines available on the World Wide Web as well as the searchable subscription databases commonly available in libraries. The intention is to eliminate the “accidental” difficulties of searching – remembering the different search string formats used by different search engines and the different features they support – so that researchers can concentrate on the “essential” problem of formulating an effective search strategy (see Fred P. Brooks, “No Silver Bullet: Essence and Accident in Software Engineering” for a discussion of these terms). We also envision this as a teaching tool for courses that include a component on conducting research on the Internet.

Positioning

Business Opportunity

This part is a summary of the business case for the software. This is important during the Inception phase, in which we are trying to determine whether it is worthwhile to continue.

This section still needs to be written!

Problem Statement

Although it’s relatively easy to use Internet search engines, it is often difficult to use them effectively. Part of this difficulty is inherent in the task of formulating a strategy to search for information in a large and loosely organized collection of information. Choosing the combination of search terms that will properly focus a search takes thought and effort. More recent search engines provide help in the form of “clustering” of sites that deal with different parts of a topic (Vivisimo), concept maps (Kartoo) or “resource
sites” with numerous links to presumably relevant sites (Teoma), but the essential difficulty remains.

However, another difficulty is not inherent in searching – the different search string formation rules required, and different features supported, by different search engines. This is the difficulty that the proposed translator proposes to address.

It also proposes to provide help for people who are trying to learn to search more effectively and the instructors (including both classroom instructors and librarians) trying to help them.

**Product Position Statement**

The professional literature for librarians includes discussions of federated searching – searching in several places at once. This system is clearly related to such federated search tools. I need to review this literature with a librarian. Possible unique features of this translator include the following:

- The ability to create search environments – including source string format specifications, supported operations and features, and available target search engines. Instructors could use these to create restricted environment for educational purposes. Librarians could use this feature to configure a default environment tailored to the resources available in the library.
- The ability to create and run tutorials to help students and other researchers become more proficient in Web-based research using search engines.

**Stakeholder Descriptions**

**Researchers** using the Internet or library databases want fast accurate translations of the common-format search strings that they enter into the format required by each target search engine. If a translation is not possible for a target engine, they want a clear explanation of why not, and possibly an indication of how to modify the source string to make a translation possible. They want a system that is easy to use for people with limited experience with information technology.

**Students** want a system that promotes learning how to develop effective search strategies. They may want tutorials to help them. If so, they want an evaluation of their work on tutorials after each step of the tutorial with a summary evaluation at the end.

**Instructors** teaching courses that include using search engines and similar software want to be able to create and maintain search environments that promote student learning. They also want to be able to create tutorials – possibly with associated search environments – for this purpose. They may want information about student performance on tutorials, both for grading and to be able to improve the tutorial.
Course coordinators may want to maintain a repository of instructional search environments and tutorials that have proven to be effective.

Librarians want to be able to create default search environments for their patrons and staff that are adapted to the search resources available in their libraries. (Each library will have its own set of resources.) Since they may be required to run training sessions or create tutorials for staff or patrons, they also have many of the same interests as instructors.

SST Administrators want to insure that only authorized users can create, retrieve, update and delete environments or tutorials. They also want to be able to manage the set of target search engines (add new engines, modify the description of existing engines, delete engines that are no longer supported, etc.). They also want error logs of problems encountered by other users of the system.

System Administrators want to be sure that the Search String Translator does not consume too many computer resources (memory, CPU time, etc.) They also want a system that is easy to install and maintain.

Other stakeholders...?
## Functional Requirements (System Features)

<table>
<thead>
<tr>
<th>Use Case</th>
<th>Actor</th>
<th>Goals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translate Search String</td>
<td>Researcher</td>
<td>Fast accurate translation of search strings from a common format to the format required by various search engines. Clear explanations of why a translation fails.</td>
</tr>
<tr>
<td>Report Failure</td>
<td>Any User</td>
<td>Report failure of system to perform properly.</td>
</tr>
<tr>
<td>Manage Search Environments</td>
<td>Search Environment Developer (includes librarians, instructors, and course coordinators)</td>
<td>Instructors, course coordinators, and librarians: create, modify, and delete instructional search environments. Librarians also want to create and modify the default environment for library patrons and staff.</td>
</tr>
<tr>
<td>Manage Target Search Engines</td>
<td>SST Administrator</td>
<td>Wants to maintain current information about target search engines. Add, delete, or update search engine descriptions.</td>
</tr>
<tr>
<td>Create Tutorial</td>
<td>Instructor (includes Librarian)</td>
<td>Wants to create a tutorials to aid learning.</td>
</tr>
<tr>
<td>Review Tutorial Results</td>
<td>Instructor</td>
<td>Wants to get a summary of the performance of students who took the tutorial, for evaluation of students work and/or evaluation of tutorial’s effectiveness.</td>
</tr>
<tr>
<td>Take Tutorial</td>
<td>Student</td>
<td>Wants to learn about search engines and how to use them effectively. Wants quick feedback to help her evaluate her progress.</td>
</tr>
<tr>
<td>Review failed transactions</td>
<td>SST Administrator</td>
<td>Wants to know about system failures so that she can address them.</td>
</tr>
<tr>
<td>Manage Authorized Users</td>
<td>SST Administrator</td>
<td>Wants to ensure that only authorized users can create, modify, delete, or activate search environments or tutorials.</td>
</tr>
</tbody>
</table>

*I’m sure I’ve forgotten something...*
Other Requirements

Usability: The intended users will have varying degrees of fluency in using information technology. A relatively naïve user should be able to use the basic functionality of the SST with minimal instruction. The SST should provide clear explanations when it is not possible to translate a source string for a particular target search engine. On-line documentation should be available.

Accuracy: The SST should provide correct translations of the source string in 99% of instances in which translation is possible. It should correctly identify the sources of the problem for 99% of untranslatable source strings.

Response Time: The SST should return translations of 99% of source strings in under one second.

Resource Usage: For translation, the translation process should run locally on the user’s computer after the search environment is established from a server. Storage of search environments should take up minimal space in the server.

Maintainability: Since search engine formats and features change frequently, it should be possible to adapt to most changes with no modifications to existing code. At worst, minimal modifications – mostly in the form of additional code – should be required. This issue will need to be revisited in the anticipated second version of the SST.

Extensibility: An anticipated future version of the SST will send the translated search strings to the target engines and display the results of the searches. The system should be designed to easily accommodate this extension.

Implementation: The SST will be open-source code written in the Java programming language.