

RIDAR – Relevant Internet Data Resource Identification

Zoltán Balogh

Institute of Informatics, Slovak Academy of Sciences, Dubravská cesta 9,
845 07 Bratislava, Slovak Republic
balogh@savba.sk

Abstract. Information acquiring systems often require identifying primary internet resources. RIDAR allows exploiting existing search engines to retrieve links to relevant Internet resources based on users-supplied search terms or more complicated search expressions. Details about identified resources (URL, title, etc.) are stored into databases.

Keywords: Information Resource Identification, Internet Search Services Integration

Introduction

This tool exploits the potential of existing search engines to identify relevant information resources on the Internet based on users-supplied search terms or more complicated search expressions. Tool can integrate any search engine which exposes a web service API. Currently, RIDAR supports and had integrated the following search engines:

- Google
- Yahoo!

RIDAR provides generic interfaces which allow integrating search engines as well as targets for storing search results (databases).

Implementations

In order to access API of any search engine, one must register to obtain an application ID (or license key). License key must be used each time the API is accessed. License key for Google and application ID for Yahoo was obtain just for the purpose of the

2 Zoltán Balogh

NAZOU project. Limitation of such registration is that the number of queries is limited for each license key.

RIDAR also allows storing retrieved results into any target such as database or generic file. Currently MySQL target is implemented in RIDAR.

References

1. Google Web APIs (beta). <http://www.google.com/apis/>
2. Yahoo! Search Web Services. <http://developer.yahoo.net/search/index.html>Alberto