



Web 2.0 Broker: a tool for massive collection of user information

Manuela Núñez, Laura Díaz, Carlos Granell, and Joaquín Huerta
Universitat Jaume I – Institute of New Imaging Technologies, Castellón, Spain

Web 2.0 services have been raised swiftly during the latest years increasing the amount of publicly available information provided by users. One of the consequences of this participative scenario is the massive existence of user information that differs, not only semantically, but also in its nature and format. As a result to be able to consume all this amount of information we must consider the discovery and access of different and specific web-based sharing services such as YouTube which supports video data, Twitter that offers a social networking and microblogging text services, Flickr that carries out online photo management and sharing, OpenStreetMap that provides extensive free spatial data or Geonames which is a geographical mapping data base. The bigger this amount of services (and data) becomes the more time takes for a user to find detailed information related to a specific subject.

The tool we propose is called the Web 2.0 Broker, a mashup of several services acting as adapters to bring the opportunity to the user of searching in several Web 2.0 services at the same time by using one unique entry and query. The Web 2.0 Broker adapts and propagates the user query to a set of Web 2.0 Services, it filters the results obtained and finally offers to the end user the results in a standardized way.

To increase interoperability and scalability of this component we have chosen the OpenSearch discovery protocol. OpenSearch sets out a collection of technologies which allows publishing of search results in a standard and accessible format.

The Web 2.0 Broker implements the Opensearch specification as a query interface, the use of this standard assures its reusability in different scenarios, its internal adapters transform this OpenSearch query to the concrete API of this different Web 2.0 services to access.

As a proof of concept we have developed a web client using GWT (Google Web Toolkit) technology, which is an open source set of tools to create faster AJAX and complex JavaScript front-end applications in Java, based on RPC (Remote Procedure Call) Protocol to communicate both client and server-side. While client-side presents the user interface and gathers the user query, the server-side unify the request thanks to the descriptors of OpenSearch and broadcast the request through the different search engines, having the central process of dealing with each API and reading data from several third party server in real time. Once the information is joined, this could be sent to the user in a lightweight data-interchange format like JSON (JavaScript Object Notation) or Atom, the web client also includes a map client based on Openlayers to show graphically the geo-referenced results.

The massive collection, mixing and combining of specific content from the web together into an integrated environment could help either users to retrieve daily information or researchers to increase the amount of retrieved information to increase accuracy of analysis and statistics.