

GlobeMash: a Mashup for Accessing GLOBE

Joris Klerkx

(Katholieke Universiteit Leuven, Belgium
Joris.Klerkx@cs.kuleuven.be)

Erik Duval

(Katholieke Universiteit Leuven, Belgium
Erik.Duval@cs.kuleuven.be)

Abstract: In this paper, we present GlobeMash, a mashup web application that uses standardized data formats like XML, JSON, LOM, SVG, CAM and protocols like SOAP, HTTP, to enable users to access the repositories of the Global Learning Objects Brokered Exchange (GLOBE) consortium. GlobeMash uses the Google Maps API to display the repositories and results on a geographical map, the Timeline API of the SIMILE project and the federated search layer of the GLOBE infrastructure. It enables users to query all the repositories in GLOBE and to get an insight in their search history by visualizing the latter as a combination of an extended tag-cloud and a synchronized timeline.

Keywords: Hypermedia systems, web-based services, AJAX, information visualization.

Category: H3.3

1 Introduction

This paper presents GlobeMash, a mashup application that focuses on opening up a collection of learning object repositories. More specifically, we focus on the Global Learning Objects Brokered Exchange (GLOBE) consortium [GLOBE, 04], which aims to unlock the ‘deep’ web of the learning repository networks that its members (Ariadne, Edna, Merlot, Nime, LORnet and KERIS) maintain. GlobeMash enables users to

- find learning objects within the GLOBE network of repositories,
- obtain a visual insight in the search history of users.

These functionalities are necessary for users to be able to access learning objects of multiple repositories in an efficient and flexible way. Saving the history of a user is a functionality that resulted from usability research of user interfaces to access libraries [Dorner, et al., 03]. Furthermore, this history feature can be used to enable support systems for administration decisions. It can provide insight into the user’s behaviour thus enabling management to recognize needs and potential savings (e.g. cheaper learning material) earlier [Najjar, et al., 06].

A mashup is a web application that uses services from more than one source to create a new service. Services used in mashups are typically sourced from a third party via a public interface or API. GlobeMash uses

- the SQI API [van Assche, et al., 06] for searching within GLOBE,
- the Google Maps API [GoogleMaps, 07] to show geographic information about the GLOBE Repositories and their results, and

- the Timeline API of the SIMILE Project [SIMILE, 03] to visualize a history of keyword queries.

In section 2, we discuss the features of GlobeMash that distinguishes it from the existing interfaces. Section 3 goes deeper into the technical aspects of GlobeMash and we conclude this paper in section 4.

2 A Mashup for GLOBE

In this section we will present the mashup that was created to access GLOBE. First of all, we situate GlobeMash in the context of the existing web interfaces that each of the members of GLOBE has implemented. In the following sections, we will describe the features of GlobeMash one-by-one. We will not include keyword searching within GLOBE, as this is the most basic functionality that every interface should offer. We will only include those features that are specific for GlobeMash.

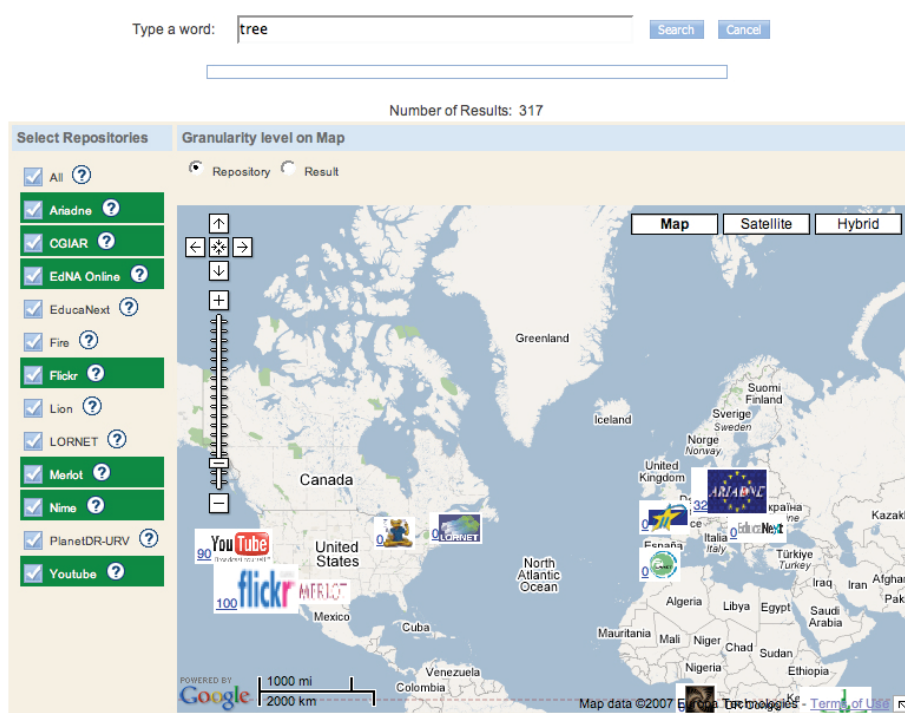


Figure 1: *GlobeMash: A Mashup Application to Search within GLOBE.*

2.1 Situation

Each member in GLOBE has implemented a web interface for querying the GLOBE repositories. They all have the following features in common:

- Users can search in all repositories by filling out a simple query in an electronic form.
- Search results are always presented as a human readable abstract of the metadata that describes the learning object.
- Users can consult a result from a particular repository through the web interface of the repository where the result was found.

Before researchers can investigate the behaviour of the users in GLOBE, they need to find a way to collect and merge the logs for all these separate applications. A possible solution would be to use the Contextualized Attention Metadata (CAM) scheme [Wolpers, et al., 06] for logging events so that the log events are standardized and easily distributable for researchers.

Title	Description	Repository
Phylogeny and Reconstructing Phylogenetic Trees	A java application that lets you investigate the problem of creating a phylogenetic tree, there is some explanation of the problem of drawing phylogenetic trees and a couple of versions of the applet. The applets let you randomly generate trees and also let you infer a phylogenetic tree with a couple of different methods. Number of species, mutation rate and the method for calculating the distance matrix can all be varied.	AMERLOT
Seed Threshing	A short video clip demonstrating the process of seed threshing. This is a part of Session 2 of the Tree Seeds for Farmers Toolkit.	CESAR
Seed Threshing	A short video clip demonstrating the process of seed threshing. This is a part of Session 2 of the Tree Seeds for Farmers Toolkit.	AMERLOT
Seed Sorting	A short video clip depicting the process of sorting seeds on a tarp. This is a part of Session 2 of the Tree Seeds for Farmers Toolkit.	CESAR
Seed Sorting	A short video clip depicting the process of sorting seeds on a tarp. This is a part of Session 2 of the Tree Seeds for Farmers Toolkit.	AMERLOT
Seed Propeny	An Image depicting species selection. Session 2 of the Tree Seeds for Farmers Toolkit	AMERLOT
From little acorns... An evaluation of The Knowledge Tree - an e-journal of flexible learning in VET	An evaluation of the Australian Flexible Learning Framework e-journal, The Knowledge Tree.	
Tree of Life	As described at the site, "The Tree of Life is a project containing information about the diversity of organisms on Earth, their history, and characteristics. The information is linked together in the form of the evolutionary tree that connects all organisms to each other." Biosciences section of the Directory of the Virtual Library, an expert-run catalog of sections of the	AMERLOT

Figure 2: Result list presentation of *GlobeMash*; results are sorted on the description field.

2.2 Result presentation

All existing interfaces in GLOBE present results of a keyword query as a list of objects, with a small number of metadata fields. These fields typically include the title, a description, the author(s) and the repository where the result was found. In figures 1 and 2, the reader can see that we integrated both such a list presentation and on top of that a googlemap with geographic information. In the following subsections, we discuss all methods of representing results within *GlobeMash*. Besides those presentation methods, a user can always look at the full metadata instance (LOM) that was returned by the *Globe* network, which is not always the case in existing interfaces.

2.2.1 Geographic information of the repository

When a user selects this level of granularity, all repositories that are searched on, are displayed on the map, as a logo at the geographic location of the headquarters of the organisation that maintains it. This can be seen in Figure 1. If the user issues a keyword query, a number is shown to the left of the icon that presents the number of search results that are returned. The size of the icon grows proportionally to the

number of results as a clear visual cue. This way of presenting the result information provides a clear visualization of the relation between the repositories and the results. We can for instance imagine that a user wants to see which repositories contain large numbers of learning resources on a specific topic like economy. The user can click on the repository logo of his choice to only see the results of that repository in the list (Figure 2) below the map so the map can also be used as a filtering mechanism.

2.2.2 Geographic information of the individual results

When a user selects this level, all results are indicated on the map. To be able to do so, we need a mapping from the metadata of the result to a location. We try to automatically deduce this location from the metadata of the result. First of all, if present, we map the address of the institution of the author(s) through an open geolocation database. If the metadata lacks the address of the institution, we look at the email address of the author(s) and map the domain in this address to an IP address. We can then again map this IP address to a location. Clearly, we cannot always do this. For instance, if the email address of the author(s) is not within the domain of his institution but rather from a domain like gmail.com or hotmail.com, the resulting location of the result will not be correct. Always available in the returned metadata is the location of the description of the result on the web in the form of a handle. This location is mapped to a geolocation in case that all previous ways fail.

With all results displayed on the map, one can get an impression about the distribution of learning objects on specific topics over the world.

2.2.3 List presentation of the results

A table of results is always shown when the user performed a query. The columns indicate the title, a description, the author(s) and the repository where the result was found. All the results can be sorted on every column as research proofs that this is a usability requirement for searching libraries and databases [Dorner, et al., 03]. This functionality is not available in the existing interfaces to find material in GLOBE. When the user selects a repository in the map, the table filters out the results of the other repositories so the list is synchronized with the map.

Another usability requirement is that a user should be able to merge results from different repositories [Dorner, et al., 03] in one list. Therefore we added the functionality to enable users to create their own list with results of their interests that came in from the network of repositories. The user can click on a result and look at the object itself without having to leave the current webpage. Results from multiple queries to the GLOBE network can be added to this list.

2.3 A history visualization

In order to create a feedback loop that enables learning from the way people actually use the search interface within GLOBE, it is essential to track the behaviour of users with the web application. It is needed to answer questions like

- Which keywords are used how many times?
- Which keywords are used that have many or no results?
- Which keywords have results in multiple repositories?

- Which keywords or topics have the attention of a particular user over time?
- What is the behaviour of users that search within the GLOBE community?
- What is the time spent with the web application?
- How many results does a user look at in detail after performing a query?
- When did a user perform a query and which keywords did she use?
- Etc.

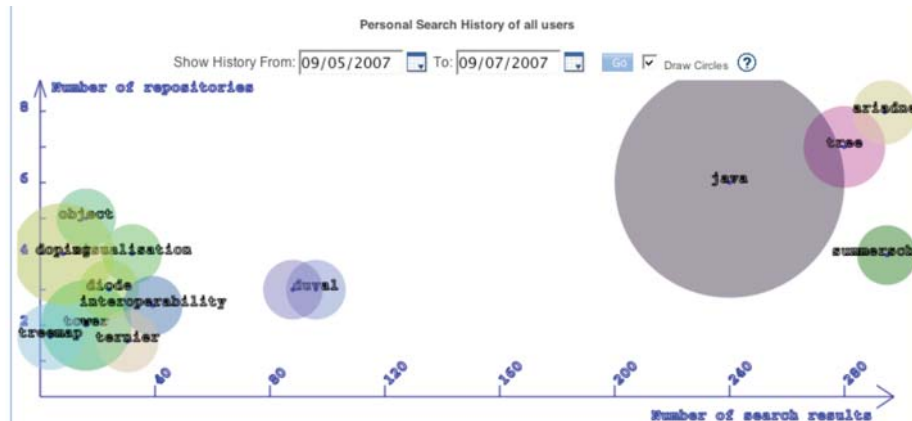


Figure 3: A Tag Cloud Visualization of Users' Search history

Therefore we created a visualization that enables users to get an insight in the questions above. Every event in the web application is logged in the Contextualized Attention Metadata (CAM) specification [Wolpers, et al., 06].

With the logged history, a visualization of the history of keyword queries from users is built and shown in GlobeMash. The visualization consists of a tag-cloud visualization, which we extended with extra information levels, and a timeline that displays all search events on it. This can be seen in figures 3 and 4. If a user does not login, the history of every user of the last time period is visualized without giving details on which query belongs to which user. If the user does login, he can see his own history. With the current software, we are able to visualize the events of multiple users. A user should be able to publish his extended tag-cloud to other users like people can do in the social bookmarking tool del.icio.us where they can publish their tag-cloud that is based on their bookmarks. This is however not yet functional GlobeMash but it will be added in the near future.

2.3.1 Extended Tag-Cloud

A tag-cloud is a list of the most popular tags, usually displayed in alphabetical order, and visually weighted by font size. In a tag-cloud, when a user clicks on a tag, one obtains an ordered list of tag-described resources, as well as a list of other related tags [Hassan-Montero, et al., 06]. In our version, the tags represent a keyword query of a user. If we would use an alphabetical order to place the keywords, users would not be able to get an insight in questions like

- Which keywords are used that have many or no results?

- Which keywords have results in multiple repositories?

Therefore, the keywords are not displayed in alphabetical order but are placed in a coordinate system where the X-axis represents the average number of search results and the Y-axis the number of repositories where results were found, both in relation to the performed keyword query. The scale of the X-axis and the Y-axis are automatically computed from the data that has to be drawn in the visualization.

On top of that, the visualization offers the functionality to draw circles around a tag or keyword. The bigger the radius of a circle, the higher the frequency of the keyword query through time. The color-coding of these circles can be random or it can be used to encode the number of different users that actually used the keyword query, depending on the selection of the user. Another idea for the color-coding would be the number of results that were downloaded and used by e.g. a teacher in his course.

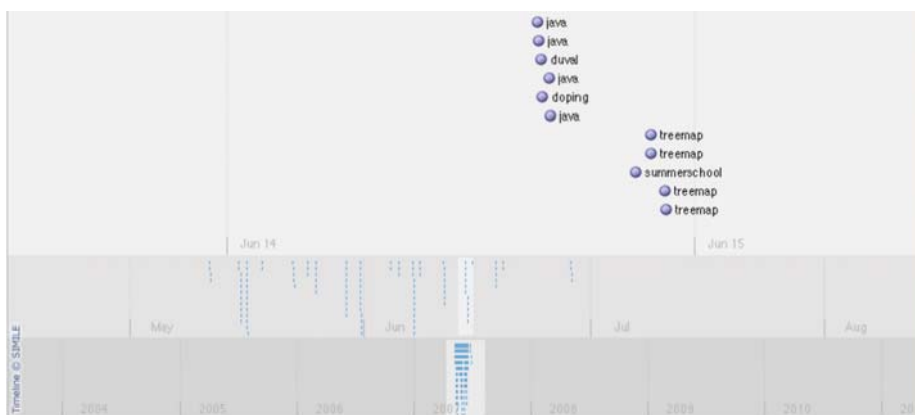


Figure 4: A Timeline visualization of users' search history

2.3.2 Timeline

The tag-cloud visualization gives an overview of the keyword used in queries in the federated network of repositories. It however does not show when a keyword is used. Therefore we integrated a timeline that we synchronized with the tag-cloud visualization. On the timeline, all keyword queries are displayed at the exact time and date that they were issued. In this way, a user can get an idea of the topics he was interested in over time. If a user scrolls the timeline to a specific date, he can choose to only show the keywords from a start date to the chosen date on the visualization. This can be used to filter out keywords that were used in a time period that does not have the interest of the user. If a user clicks on a keyword in the tag-cloud visualization, the timeline will smoothly scroll to the last timestamp that the keyword was used.

3 Underlying Technology

Our objective was to design and implement a mashup web application which is modular and uses standardized data formats (LOM [LOM, 04], XML [XML, 06], JSON [JSON, 06], SVG [SVG, 03], CAM [Wolpers, et al., 06]) and protocols (HTTP, SOAP) to ensure that different parts of GlobeMash can also be used in other contexts than GLOBE. The visualization of the history can for instance be used in other tools that collect attention metadata. GlobeMash is created with the help of the Dojo toolkit which is an Open Source DHTML toolkit written in JavaScript. It allows developers to easily build dynamic capabilities into web pages [DOJO, 06].

The federated search layer of GLOBE is accessed when a user issues a keyword query. This works by sending an AJAX request from the client to a java servlet at our server that uses the CEN/ISSS Simple Query Interface (SQI) [van Assche, et al., 06] for querying the repositories within GLOBE. The results that are returned to the servlet are in the IEEE LTSC Learning Object Metadata (LOM) format [LOM, 04]. These are returned in JSON format [JSON, 06] to the client where we parse the JSON data and add the results to the map and the list presentation. The map is rendered using the Google Maps API [GoogleMaps, 07]. The geographical locations of the results are found by performing three look-up requests in GeoIP [GeoIP, 07], NetGeo [Moore, et al., 00] and Hostip [HostIp, 07]. By performing multiple requests, it is more likely to get a valid location as the three geolocation databases together have a better coverage of the world.

The history visualization consists of two parts. The upper part is an SVG (Scalable Vector Graphics) drawing that is rendered by the Dojo framework in the client browser. The timeline visualization part is created using the SIMILE project and is a DHTML-based AJAX widget for visualizing time-based events [SIMILE, 03]. We synchronized the SVG visualization and the timeline widget by using the javascript event system.

All events in the web application are logged in the Contextualized Attention Metadata (CAM) specification and stored in a PostgreSQL database. The events are transferred to the client browser by mapping the attention metadata to JSON after which the history visualization gets created by executing a number of our javascript functions in the client browser.

4 Conclusion

This paper is part of our continuing research on accessing learning object repositories in an efficient and flexible way. In [Klerkx, et al., 05], we created an open and extensible information visualization framework, which we used to access the Ariadne Knowledge Pool System. In this paper we broadened our scope by moving from one LOR towards opening up a collection of repositories, namely GLOBE. Therefore, we created GlobeMash: a mashup application that distinguishes itself from existing interfaces by displaying the repositories on a geographical map, by adding a visualization of the history of the keyword queries of the users and by adding functionalities that improve the usability of search by e.g. enabling users to merge results from multiple repositories.

Logfile analysis has been the starting point in order to begin to know what happens inside GlobeMash. We know the user attention from looking at the attention metadata that is collected and from using the history visualization. But this approach is insufficient as an evaluation of the usability of the mashup from an end user point-of-view. We want to know what users like and what they dislike, what they find useful, what they find hard to understand, etc. Therefore, we are planning to perform a thorough usability test. We will evaluate GlobeMash against a number of usability criteria that were defined in [Dorner, et al., 03]. The results will be used to add new features or leave out the features that are not useful in the web application.

Acknowledgements

We gratefully acknowledge the financial support of the K.U.Leuven research council through the BALO project, the Interdisciplinary Institute for Broadband Technology (IBBT) through the Acknowledge project, and the European Commission through the ProLearn Network of Excellence on Professional Learning.

References

- [van Assche, et al., 06] F. van Assche, E. Duval, D. Massart, D. Olmedilla, B. Simon, S. Sobernig, S. Ternier, and F. Wild, "Spinning Interoperable Applications for Teaching & Learning using the Simple Query Interface", *Educational Technology & Society* 9 (2) (2006), 51-67.
- [DOJO, 06] The Dojo Toolkit: <http://dojotoolkit.org>
- [Dorner, et al., 03] Dorner, D.G., Curtis, A. (2003), "A comparative review of common user interface software products for libraries", National Library of New Zealand
- [GeoIP, 07] GeoIp Technology, <http://www.maxmind.com/app/ip-locate>
- [GLOBE, 04] Global Learning Objects Brokered Exchange: www.globe-info.net/
- [GoogleMaps, 07] Google Maps API: www.google.com/apis/maps/
- [Hassan-Montero, et al., 06] Y. Hassan-Montero and V. Herrero-Solana. "Improving Tag-Clouds as Visual Information Retrieval Interfaces, Proc. InSciT 2006.
- [HostIp, 07] hostIP geolocation database: www.hostip.info
- [JSON, 06] RFC4627: JavaScript Object Notation (JSON)
- [Klerkx, et al., 05] J. Klerkx, M. Meire, S. Ternier, K. Verbert, and E. Duval, "Information Visualization: Towards an Extensible Framework for Accessing Learning Object Repositories"; Proc. ED-Media'05, AACE, Montreal, Canada (2005), 4281-4287
- [LOM, 04] 1484.12.1: IEEE Standard for Learning Object Metadata
- [Moore, et al., 00] D. Moore, R. Periakaruppan, J. Donohoe, K. Claffy, "Where in the World is netgeo.caida.org?" CAIDA Publications, (2000)
- [Najjar, et al., 06] J. Najjar, M. Wolpers and E. Duval, "Attention Metadata: Collection and Management", WWW2006 workshop on Logging Traces of Web Activity: The Mechanics of Data Collection, Edinburgh, Scotland, May 23, 2006
- [SIMILE, 03] The SIMILE project: <http://simile.mit.edu/wiki/SIMILE>About>
- [SVG, 03] W3C Scalable Vector Graphics (SVG) format
- [Wolpers, et al., 06] M. Wolpers, G. Martin, J. Najjar and E. Duval, Attention Metadata in Knowledge and Learning Management, Proc. IKNOW'06, Graz, Austria
- [XML, 06] W3C Extensible Markup Language (XML)