Exploiting Tag Clouds
for Database Browsing and Querying

Stefania Leone, Matthias Geel, and Moira C. Norrie

Institute for Information Systems, ETH Zurich
CH-8092 Zurich, Switzerland
{leone|geel|norrie}@inf.ethz.ch

Abstract. We show how tag clouds can be used alongside more traditional query languages and data visualisation techniques as a means for browsing and querying databases. Our approach is based on a general, extensible framework that supports different modes of visualisation as well as different database systems. A number of demonstrator databases and interfaces will be used to show how tag clouds can be used to visualise and browse data or metadata and even a mix of both in object databases and relational databases. Further, we will demonstrate synchronised browsing based on tag clouds as well as ways in which tag clouds can be combined with other forms of querying and data visualisation.

Keywords: tag cloud, data visualisation, database interface

1   Introduction

Tag clouds are widely used in Web 2.0 applications for visualising user-generated tags and folksonomies of specific web sites such as Flickr[1]. The presentation and layout of tags can be controlled so that features such as the size, font and colour can be used to give some measure of the importance of a given tag, while the positioning of tags may be based on pure aesthetics or some form of relationship between tags.

Given the flexibility of tag clouds in terms of information representation together with the simplicity of the associated style of navigation, it is natural that database researchers should consider exploiting the concept of tag clouds to address the longstanding problems of database usability [1]. The use of a query language requires the user to master not only the query language but also the database schema. To allow users to view the data in a natural way, a higher-level presentation of the database content such as a visual schema browser and query interface is needed. Another approach is to focus on the data rather than the schema as supported in keyword search interfaces to databases. Tag clouds have been proposed as a means of summarising and refining the results of keyword searches as presented in [2, 3]. In this case, the term *data cloud* is used to refer to their particular adaptation of tag clouds for this purpose. An interesting

---

[1] http://www.flickr.com

feature of their approach is that since it was developed for relational databases, the developer of a data cloud application specifies how application entities can be composed from the relations in the database in order that keyword search can be applied to entities rather than simple attributes or tuples. The keyword search is based on a traditional information retrieval approach where entities are considered as documents and attribute values as weighted terms. Another project that uses tag clouds for summarising query results is PubCloud [4] for searching the PubMed biomedical literature database. In this case, the tag clouds are generated from words extracted from the abstracts returned by the query.

Our goal was to investigate the extent to which tag clouds could be exploited to support more traditional forms of database browsing and querying, either replacing existing query languages and other modes of data visualisation or being used alongside them. Our tag clouds therefore mainly represent data and metadata values rather than terms occurring within them. To support our investigations, we have developed a general, extensible framework that supports different modes of data visualisation, including customisable tag clouds. We have also designed it so that different types of databases can be accessed and currently have implementations for both object databases and relational databases.

A key advantage of the tag cloud approach is that it is data-driven rather than schema-driven which is particularly beneficial to users with no experience of databases and query languages. Our initial user studies have shown that even users with low computer literacy and no previous experience of tag clouds were able to find the results of non-trivial queries using our system. At the same time, expert users also gave favourable feedback about the system and particularly liked the fact that it could be combined with query expressions.

Our contributions include:

- A data browser that allows any data source to be browsed and queried using tag clouds.
- Experimentation with text and position features of tags in a tag cloud to make clouds more informative.
- A tool that serves different purposes: Novice users are able to access structured data sources without knowing the query language and schema, while expert users can browse a data source in order to get to know the schema and thus be enable to express complex queries over the data source.
- An extensible and flexible platform for experimentation where new data sources and new visualisation techniques can be added.

In the following sections, we provide an overview of the data browser, the architecture and also the demonstration.

2   Data Browser

As highlighted in [5], tag clouds serve multiple purposes. They can be used for searching for specific information, browsing a data collection without a specific target, as a tool for impression formation and gisting, and to recognise what a

data collection is about. In the Web, tags of a tag cloud are usually hyperlinks that lead to a collection of items that are associated with a tag. Tag clouds are graphically appealing due to different visualisation features. Tag cloud features include text features, such as the tag content, the size, font style and colour as well as the positioning and order of tags in a cloud. A lot of studies, such as [5–7] have experimented with tag cloud features and positioning and their impact on users. According to [5, 6], font size, font weight and intensity are the most important features. While topic-based layouts of tags can improve search performance for specific search tasks compared to random arrangements, they still perform worse than alphabetic layouts according to [7].

We adapted these concepts to browse structured data where tags represent attribute values. Clicking on a tag initiates a selection for data items with the corresponding attribute value. In the case of object databases, the result would be a collection of objects, while in the case of a relational database it would be a collection of tuples, i.e. a relation. We note that concepts similar to those proposed for data clouds in [2, 3] could be adopted to return entities rather than tuples for specific applications. Similarly, it is possible to mix different attribute values in a single tag cloud or to form tag clouds from combined attribute values. In addition, we use these concepts to also browse metadata and have even experimented with a mix of metadata and data within tag clouds.

We now explain these concepts further by means of an example based on a database with information about contacts and their locations. Generally, we define a data source to be a set of data collections, where each collection contains data items of a specific type. These collections are either class extents or sets of objects of a specific type in object databases, while they are relations in relational databases.
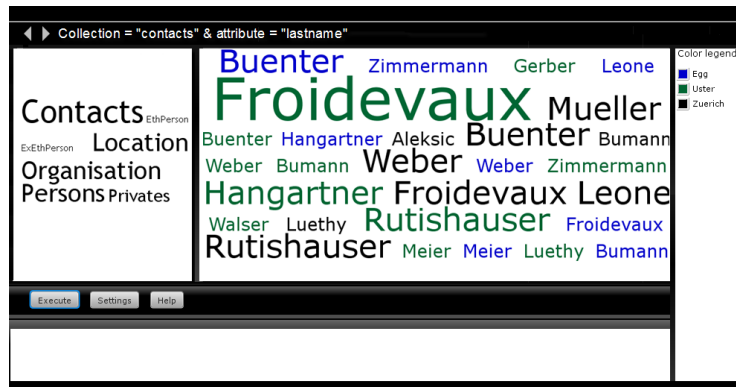

Fig. 1. Schema and Data Browsing

The metadata that defines the schema of a database can itself be represented by a tag cloud as shown in Figure 1. On the lefthand side of this figure, the tag

cloud gives the names of the various collections of data items within the database. The default is to have the size of the tags represent the relative cardinality of the collection.

A user can start browsing a database either by entering a query expression in the window below the tag clouds or by selecting one or more of the tags in the schema tag cloud. Each collection can have a default attribute or set of attributes specified for its visualisation as a tag cloud. However, the user can also specify this by means of a simple selection of attributes through checkboxes. Alternatively, one can display the attributes themselves as a tag cloud in the lefthand window and allow the users to select one or more attributes as tags. In this way, we support synchronised browsing across the metadata and data through the adjacent tag clouds.

In the example of Figure 1, the attribute `lastname` is displayed in the tag cloud on the right as indicated by the navigation path shown on top of the cloud window. The size of the tags in this tag cloud represents how many data items have that attribute value. In this way, the tag cloud can be considered as a visualistaion of the attribute value frequency. The user can now click on a tag and further refine their selection. When hovering over a `lastname` tag, a user gets detailed information about the number of objects that have this attribute value, or in the case of only a single object, we get the set of attribute values.
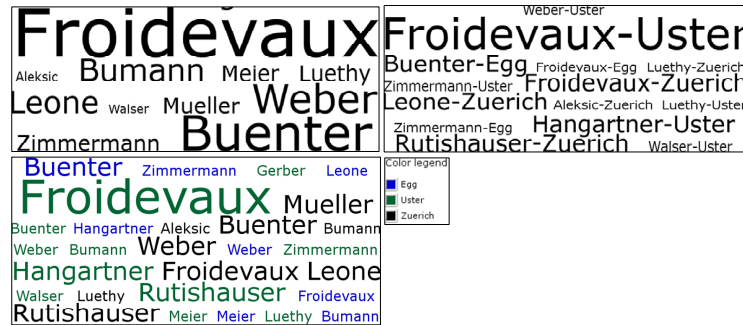

Fig. 2. Exploiting tag cloud features

We offer different modes for visualisation, as depicted in Figure 2. In the tag cloud in the upper-left corner, the contacts are displayed by lastnames. In the upper-right corner, two attributes are bound to the tag content feature, namely the attribute `lastname` of `contacts` as well as the attribute `city` of the associated location objects. The tags thus represent the number of contacts with a given name that live in the same city. In this example data set, the tag Froidvaux-Zurich represents the set of contacts with lastname 'Froidvaux' who live in 'Zurich'. As one can see in this example, more people with the name 'Froidvaux' live in 'Uster', than in 'Zurich'. In the lower-left corner of Figure 2, we added colour as an additional visualisation dimension: The attribute `lastname` is

bound to the tag content, while the attribute `city` from the associated location is bound to the colour feature. As one can see from the index on the righthand side of the figure, each distinct attribute value of the city attribute is assigned a specific colour. We have experimented with these different tag features in a user study. Care has to be taken in choosing the right attributes to bind to the colour feature. It only makes sense, if the distinct set of values is not too large, since otherwise the index becomes very large and the tag colours are not very informative.
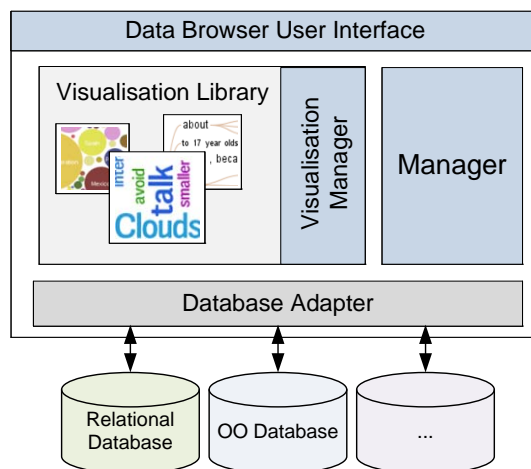
3   Architecture

Fig. 3. System Architecture

Figure 3 gives an overview of the system architecture. The manager component is the heart of the system and responsible for handling requests from the user interface, forwarding these to the database through the database adapter and invoking the visualisation manager to transform the results into the appropriate visual elements to be returned and displayed in the GUI. Our framework is extensible in multiple ways. Firstly, we provide a data adapter interface which can be implemented for any data source. At the moment, we have an implementation for the object databases db4objects[2], OMS Avon[3] and OMSPro[4] as well as a MySQL implementation. Secondly, the visualisation manager can manage different kinds of visualisation techniques. Therefore, we provide a visualisation

---

[2] http://www.db4o.com/

[3] http://maven.globis.ethz.ch/projects/avon/

[4] http://www.globis.ethz.ch/research/oms/platforms/omspro

interface which has to be implemented to add a new technique to the visualisation library. We currently provide a tag cloud visualisation, and are working on a bubble chart visualisation. Our data browser application is flexible and configurable and is currently used as a platform for experimentation in our research group.

4   Demonstration

In our demonstration, we will show how users can browse both relational and object databases using our data browser. The demonstration will include showing tag clouds over data, metadata and a mix of data and metadata. We will provide a set of demonstrator databases including a contacts database and a publications database implemented using both relational and object databases. Visitors will be able to freely browse these databases, pose queries and exploratively get an impression of the schema and the data. We will also provide a list of query tasks from our user study so that users can experience how query results can be obtained using only the bowser, using only query expressions and using the browser in conjunction with query expressions.

References

1. Jagadish, H.V., Chapman, A., Elkiss, A., Jayapandian, M., Li, Y., Nandi, A., Yu, C.: Making Database Systems Usable. In: Proc. ACM SIGMOD'07. (2007)
2. Koutrika, G., Zadeh, Z.M., Garcia-Molina, H.: Data Clouds: Summarizing Keyword Search Results over Structured Data. In: Proc. EDBT'09. (2009)
3. Koutrika, G., Zadeh, Z.M., Garcia-Molina, H.: CourseCloud: Summarizing and Refining Keyword Searches over Structured Data. In: Demo Proc. EDBT'09. (2009)
4. Kuo, B.Y.L., Hentrich, T., Good, B.M.., Wilkinson, M.D.: Tag Clouds for Summarizing Web Search Results. In: Proc. WWW'07. (2007)
5. Rivadeneira, A.W., Gruen, D.M., Muller, M.J., Millen, D.R.: Getting our Head in the Clouds: Toward Evaluation Studies of Tag Clouds. In: Proc. CHI '07. (2007)
6. Bateman, S., Gutwin, C., Nacenta, M.: Seeing Things in the Clouds: The Effect of Visual Features on Tag Cloud Selections. In: Proc. 19th ACM Conf. on Hypertext and Hypermedia. (2008)
7. Schrammel, J., Leitner, M., Tscheligi, M.: Semantically Structured Tag Clouds: An Empirical Evaluation of Clustered Presentation Approaches. In: Proc. CHI'09. (2009)