

Visualization of Uncertain Information in Tag Clouds

Manolis Wallace and Thanos Triantos



Knowledge and Uncertainty Research Laboratory
Department of Informatics and Telecommunications
University of Peloponnese
Tripolis, Greece 22 100
gav@uop.gr
<http://gav.uop.gr>

Abstract. Tag clouds provide an excellent means of visualization of weighted semantic information. With their generation depending on given or calculated weights, their use is not possible when these weights are not known with certainty. In this paper we propose an extension to support the notion of uncertainty and provide the tools to implement it.

Keywords: Tag cloud, uncertainty, visualization, transparency

1 Introduction

Tag clouds are a very useful and intuitive way to present verbal information. Going beyond their initial use as site navigation tools, lately they have received more attention as a way to visualize textual corpora, when frequency of occurrence or some other measure of importance is of interest. Tag clouds can actually be (and often are) used in any context where weighted textual information needs to be visualized. Still, existing approaches do not allow for the depiction of uncertainty in an intuitive manner.

As a solution we propose the utilization of opaqueness/transparency to depict degrees of certainty.

1.1 An example

As an example, consider table 1.1. In it we summarize the upcoming year's expected budget for a research group. There are three already running projects, for which next year's budget is already secured and an agreement with industry to implement a project next year with the contract still pending but almost certain. But there are also submitted proposals to calls of different difficulties. And on top of that there is the knowledge that some amount is typically secured during the course of every year.

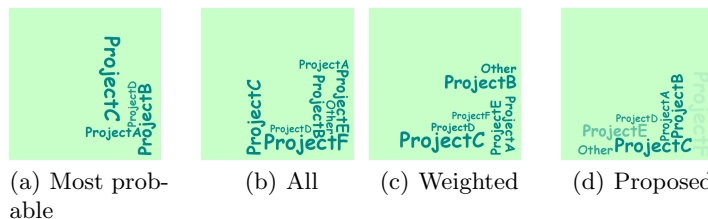
Should we wish to depict this information in a tag cloud, we would be faced with the decision of how to visualize the different degrees of certainty related to

Table 1. A research group’s expected budget for the next year

Funding source	Amount in Keuros	Comments
ProjectA	50	Running project
ProjectB	100	Running project
ProjectC	150	Running project
ProjectD	30	Agreed project, to be contracted
ProjectE	100	Submitted proposal, easy call
ProjectF	200	Submitted proposal, very competitive call
Other	50	Additional funding typically attracted each year

each one of the table’s entries. One way would be to only depict most probable options, as shown in figure 1(a). Alternatively, in figure 1(b) we depict all options, still ignoring some of the information in the table. In figure 1(c) we incorporate uncertainty by weighing amounts proportionally to their probability. Although this is valid from a risk analysis perspective, the visualization is counter intuitive and misleading: project F is depicted as small in scale which is absolutely false; either project F will bring a large budget or none at all.

The error in figure 1(c) stems from joining volume and uncertainty, which are unrelated in our example, in one visualization parameter. In order to overcome this we propose presenting volume using text size and degree of uncertainty using transparency, as shown in figure 1(d).



2 The tool in use

There are various libraries available for the creation of tag clouds. In this work we have used the Kumo - Java Word Cloud and extended it to accept degrees of certainty as input and visualize it as degrees of opaqueness.

Similarly to the Kumo library, our finalized tool and accompanying libraries will be made freely available under a GPLv3 licence.

References

1. Kenny Cason, Kumo - Java Word Cloud, <http://kennycason.com/posts/2014-07-03-kumo-wordcloud.html>