

Visualization of Uncertainty in Tag Clouds

Nikos Platis, Manolis Wallace^(✉), and Thanos Triantos

Knowledge and Uncertainty Research Laboratory, Department of Informatics
and Telecommunications, University of the Peloponnese, 22 100 Tripolis, Greece
{nplatis,wallace}@uop.gr, thanos.tri@hotmail.com
<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 of tag clouds to support the notion of uncertainty and explore some properties of this new representation. Furthermore, we present a tool implementing it.

Keywords: Tag cloud · Uncertainty · Visualization · Transparency

1 Introduction

Tag clouds are a very useful and intuitive way to present textual information. They are consistently used as a way to visualize textual corpora, when frequency of occurrence or some other measure of importance is of interest. More generally, tag clouds can actually be (and often are) used in any context where weighted textual information needs to be visualized. See for example Fig. 1 where free text responses from 4000 individuals [3] are summarized in one brief picture.

In all existing applications of tag clouds, the information depicted is certain and complete. However, real life semantic information rarely is, and various forms of uncertainty are inherent in it. In this paper we propose the utilization of the opaqueness of the tags to depict degrees of certainty on a tag cloud.

The rest of the paper is organized as follows: In Sect. 2 we outline related background such as information visualization, uncertainty, and tag clouds. In Sect. 3 we present our proposed approach using a simple example and discuss its most important aspects. In Sect. 4 we focus on the software tool that implements it. Finally, in Sect. 5 we list our concluding remarks.

2 Related Work

2.1 Information Visualization

Tag clouds are a characteristic tool of Information Visualization. Information Visualization uses graphical representations of data in order to enhance human



Fig. 1. Word cloud of open ended responses from the Wikipedia Readers Survey [6].

cognition [7], to ease understanding of the data, to allow the viewer to form a mental model of it.

The graphical representations devised are most often able to convey, in an intuitive way, multiple properties (or dimensions) of the data at once, which would be impossible with a simple textual listing of the same data; for example, points on a x - y diagram could depict many more properties apart from those assigned to their x and y coordinates by varying their size, color, and shape.

2.2 Uncertainty

Uncertainty is an inherent feature of human life and we are accustomed to dealing with it in almost any information that we are faced with, often even sub-consciously. In fact, the term “uncertainty” may refer to any situation in which something is not known certainly and/or accurately, which encompasses many heterogeneous types of information.

In this paper we deal only with uncertainty that refers to the very existence of the respective data. This includes, for instance, *probabilistic information* (events that may happen with known probability) and *possibilistic information* (events that may happen with unknown probability). Therefore, we deal with cases where uncertainty is orthogonal to the magnitude of the respective data and not correlated, i.e. it is an additional dimension of the data.

Other types of uncertainty, with which we do not deal here, refer to the magnitude of the data itself, for example *imprecision* stems from the finite precision of a measuring tool.

2.3 Tag Clouds

A tag cloud is a visualization of a set of tags, each of which is associated with a weight (frequency, importance, etc.) The tags are drawn using different *font sizes*, relative to their weight, and arranged so as not to overlap.

The other drawing parameters are chosen according to the application at hand. Tags may be arranged alphabetically (when discoverability is important), or such that the more important ones are placed near the center of the tag cloud (which emphasizes those with higher weights, cf. Fig. 1), or completely at random

aiming at a more aesthetically pleasing result. They may be placed all horizontally or both horizontally and vertically (which leads to more interesting cloud shapes). Tags may be colored according to their weight (which, again, emphasizes those with higher weights, cf. Fig. 1), or with different colors according to some other characteristic (eg. category), or completely at random. Finally, the tag cloud may assume a specific shape or not.

3 Visualization of Uncertainty

As has been explained in the previous section, in the conventional tag cloud the weight of each tag is visualized by controlling the size of the font. In this work we aim to also visualize the uncertainty of each tag when it is not correlated to its weight; therefore we need to utilize a different graphical characteristic of the visualization.

Our proposal is to visualize the uncertainty of tags by controlling their *opaqueness*. Tags that are absolutely certain are printed “normally”, absolutely improbable cases are not depicted at all, and intermediate cases are drawn with varying levels of transparency. Figure 2 demonstrates the concept.

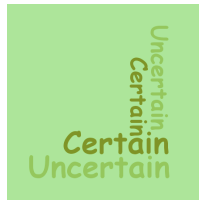


Fig. 2. The uncertain tag cloud concept.

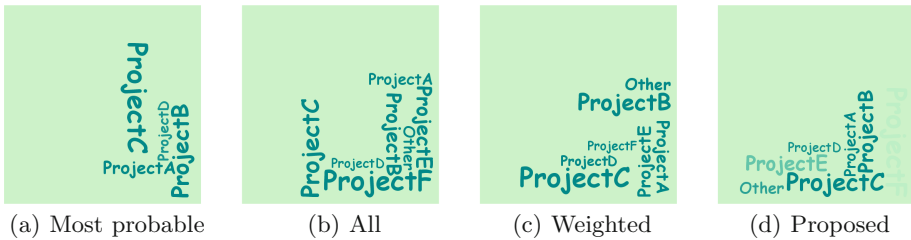
3.1 An Example

As an example, consider Table 1. In it we summarize the upcoming year’s expected budget for a research group. There are three projects already running, for which next year’s budget is secured, and an agreement with industry to implement a project next year with the contract still pending but almost certain. But there are also two proposals submitted to calls of different difficulties. And on top of that, there is the knowledge that some amount is typically allotted during the course of every year by the department.

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 each one of the table’s entries. One way would be to only depict most probable options, as shown in Fig. 3(a); this clearly omits some information and we should certainly do better. Alternatively, in Fig. 3(b) we depict all entries in a simple tag cloud, choosing to hide the fact that we already know that some of these tags

Table 1. Expected budget of a research group for the next year.

Funding source	Amount (K euros)	P	Comments
Project A	50	1.0	Running project
Project B	100	1.0	Running project
Project C	150	1.0	Running project
Project D	30	0.9	Agreed project, to be contracted
Project E	100	0.5	Submitted proposal, easy call
Project F	200	0.1	Submitted proposal, very competitive call
Other	50	0.8	Additional funds typically attracted per year

**Fig. 3.** Different ways to visualize the data of Table 1.

correspond to improbable situations; again, not all the information available in the table is represented.

In Fig. 3(c) we incorporate uncertainty by weighing amounts proportionally to their probability. From an economic or risk analysis perspective this is the optimal approach, as what is depicted is the real economic value of each project at the present time. Still, from an information visualization point of view this is counterintuitive and misleading: for example, consider project F which is depicted as small in scale; this is inaccurate in all cases as project F will either bring in a large amount or none at all. The problem in this representation stems from joining volume and uncertainty, which in our case are unrelated, into one visualization parameter.

In order to overcome this, we apply our proposed solution of representing magnitude using font size, as usual, and degree of uncertainty using transparency; see Fig. 3(d). Compared to the previous approaches, we observe that this visualization contains all of the information in the table and communicates it accurately in a straightforward manner.

3.2 Discussion

Perhaps the greatest attractions of the conventional tag cloud are its simplicity and intuitiveness: larger tags stand out immediately as more important in the context of the image. Therefore, in extending the tool one has to be careful in order to retain these desirable properties.

We feel that our proposed extension of the tag cloud, which employs the opaqueness of the tags to convey their certainty, is very natural and intuitive as well: the less probable a tag, the more it faints into the background. The viewer can understand the relative certainty of the tags based on their opaqueness, while still being able to measure their relative importance based on their size.

The main parameter that influences the effectiveness of the uncertain tag cloud seems to be the choice of colors for the tags and for the background of the picture. In fact, our preliminary testing shows that if the tag color has enough contrast with the background color then the tag cloud delivers the information successfully. Figure 3(d) shows that even though the tag and the background color are both shades of green, the transparency effect is clear.

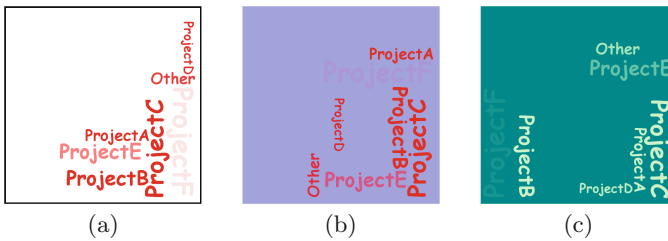


Fig. 4. Different color combinations.

Figure 4 demonstrates some other color combinations. Figure 4(a) uses a white background, which is possibly the safest one to use in conjunction with a strong tag color. Figure 4(b) utilizes two colors whose blending is rather well known, so the effect still gets through. Finally, Fig. 4(c) uses the colors of Fig. 3(d) inverted: a light green for the tags and a dark one for the background; we believe that the opacity effect is more intuitive with light backgrounds.

Another interesting aspect of the uncertain tag cloud would be to assess its effectiveness when many terms are depicted. Our simple test case uses few terms to convey the basic principle of this new representation; we believe that real-world uncertain tag clouds will be equally effective as conventional ones, since the effect of opacity is similar to the one of using different shades of color as in Fig. 1.

Of course, formal user evaluation of the proposed visualization would provide better insight into its properties, strengths and weaknesses.

4 Implementation

Several libraries are available for the creation of tag clouds [1, 2, 4, 5], offering varied choices regarding the placement of words, text colors, overall size and shape, etc.

Our proposed approach to the representation of uncertainty is not related to these choices, and therefore it may be combined with any suitable software.

In order to experimentally demonstrate the effectiveness of the proposed approach we have chosen to extend the Kumo - Java Word Cloud library [5] to accept degrees of certainty as input and visualize it as degrees of opaqueness.

Kumo was chosen because it is open source software provided as a library, thus easily extensible. Furthermore, it provides enough flexibility for our purposes, allowing the user to customize all the graphical parameters of the tag cloud (dimensions, font size, colour palettes, tag direction, tag cloud shape).

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

5 Conclusions

Tag clouds are an established tool for the visualization of semantic textual information. In this work we proposed to extend them for the visualization of uncertain information by utilizing the opaqueness of the tags to indicate their degree of certainty. We discussed the properties of this new representation and developed a software library to showcase it.

This work will be followed by a formal user evaluation of the proposed visualization in order to establish its characteristics on sound evidence and formulate guidelines for its application. Moreover, we plan to investigate other metaphors for the transparency in tag clouds and investigate situations where it might be intuitive.

References

1. d3-cloud. <http://github.com/jasondavies/d3-cloud/>
2. OpenCloud. <http://opencloud.mcavallo.org/>
3. Research: Wikipedia Readership Survey 2011/Results (2011). https://meta.wikimedia.org/wiki/Research:Wikipedia_Readership_Survey_2011/Results
4. WordCram. <http://github.com/danbernier/WordCram/>
5. Cason, K.: Kumo - Java Word Cloud. <http://kennycason.com/posts/2014-07-03-kumo-wordcloud.html>
6. Pande, M.: Word Cloud of open ended responses from Wikipedia's readers survey. via Wikimedia Commons
7. Ware, C.: Information Visualization, Perception for Design, 3rd edn. Morgan Kaufmann, Amsterdam (2012)