Evaluating Tagclouds for Health-Related Information Research

Paul Clough and Barbara Sen

Department of Information Studies, University of Sheffield, Sheffield, UK
p.d.clough/b.a.sen@sheffield.ac.uk

This paper investigates the potential use of "tagclouds" for health-related research. Twelve researchers (experts in their field) participated in a user experiment using PubCloud. Findings indicate that results from PubCloud provide effective summaries, users are moderately satisfied, and tagclouds often throw-up unexpected (but useful) terms. Comments from participants fell into two main categories: technical functionality and visualisation of the output. It is expected that results from this study could help inform use of tagclouds for health information researchers, inform the uses of tagclouds in health informatics education, and feed into the design and development of applications involving tagcloud visualisations.

Keywords
Tagclouds, Web 2.0, Visualisation, Evaluation

1. Introduction

This paper investigates the potential use of "tagclouds" for health researchers. The tagcloud (also known as a word or term cloud) is an approach used to present a visual summary from collections of texts [1, 2], commonly associated with web2.0 or the "social web" [3]. Sets of terms (words or phrases) are selected to form the tagcloud and textual attributes (e.g. colour and font size) are used to represent features of the associated terms (e.g. the frequency, relevance, popularity or recency of a term). Tags are linked to the underlying content and when visualised (usually alphabetically), frequently occurring topics may be depicted in a larger font size. Other aspects such as the recency of terms can be visualised using different colours.

Viégas and Wattenburg [2:52] discuss tagcloud visualisations as social signallers or analytical tools. In the health sector, health information researchers may find them useful for the visualisation of research concepts, as an aid to building search strategies, for information retrieval, as an analytical tool, as a way of presenting their research findings, and as a tool for learning and teaching. Despite their appeal, Rivadeneira et al. [1] comment on the lack of experimental studies evaluating the effectiveness of tagclouds in various contexts. This study contributes to existing work on tagcloud visualisations by focusing on the utility of tag clouds in health-related information research. It complements the work of Kuo et al [4] who developed PubCloud, an application for use with biomedical text which summarises the results returned by PubMed searches.

To test the utility of tagclouds in visualising health-related topics, a group of 12 researchers (experts in their field) participated in an interactive user experiment using PubCloud. Our findings indicate that results from PubCloud provide effective summaries, users are moderately satisfied, and the tagclouds often throw up unexpected (but useful) terms. It is expected that the results of this study will help inform the use of tagclouds for health information researchers and in health informatics education, and feed into the development of information access applications involving tagcloud visualisations. Section 2 describes related work; Section 3 describes our experimental setup to test the utility of tagclouds in health-related information research; Section 4 presents the results of our study and discussion of the findings; and Section 5 concludes the paper and comments on future work.

2. Related Work

Hearst and Rosner [5] identify the lack of tagcloud usability studies. Rivadeneira et al [1] considered tag layout and gisting and found font size, but not layout, significant. Halvey and Keane [6] found alphabetical listings more effective than spatial organisation. Hearst and Rosner [5] surveyed both users and developers to ascertain the advantages and disadvantages of tagclouds. They concluded that the main functionality of tagclouds is as a "suggestive device" rather than a "precise depiction" of the phenomenon. They suggest that the strength of tagclouds is not as a navigational tool.

Sinclair and Cardew-Hall [7] evaluated how useful tagclouds were for finding information; they concluded that tagclouds are best suited for general or simple information search tasks. They found tagclouds to also have positive value in terms of summarising and scanning content. Identified limitations of tagclouds included: difficulty in searching for specific information, technical problems and inaccessibility of the underlying tagged content.

Research is ongoing with regard to tagcloud development looking at ways to improve functionality and visualisation of the data. Developments have moved from "first generation" tagclouds, with limited functionality, to a "second generation" of tagclouds [8], with improved design to enable increased user interactivity and visualisation [9, 10]. This study builds on the work of Kuo et al [4] by testing tagcloud usability for health-specific information search tasks working with health information experts.

3. Experimental Setup

3.1 PubCloud

PubCloud[1] is an experimental application for use with biomedical text which summarises the results returned by PubMed[2] searches [4]. Visualised tags are actually words (or terms) which are extracted from various fields of the top *n* results (e.g. abstract, author field or assigned MeSH terms). Font size of the tags indicates frequency in the results, and font colour indicates recency of the results (average publication date for the documents containing the word).

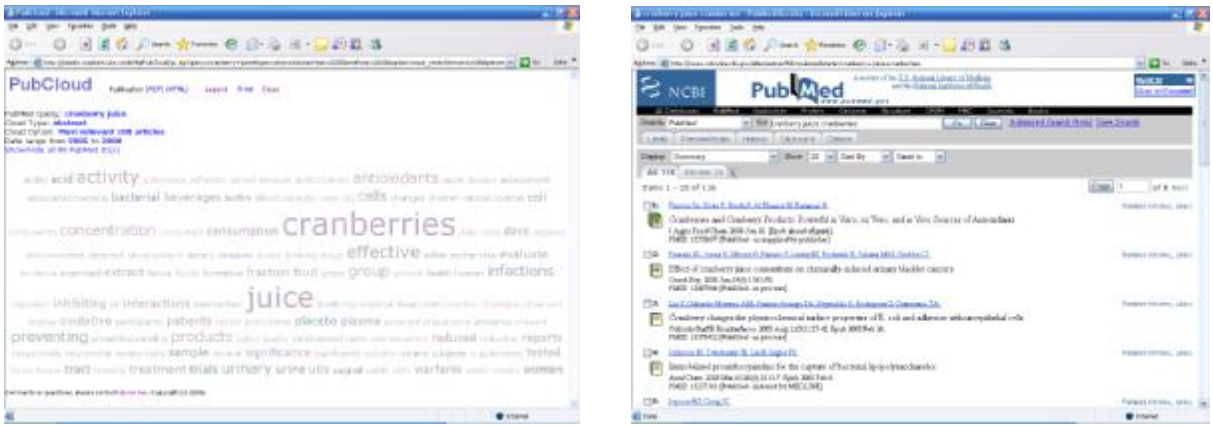

Figure 1 Example tagcloud for query "cranberry juice" (left), selecting largest tag "cranberries" produces PubMed results (right) for the search "cranberry juice cranberries"

The user can select how many tags to display in the tagcloud and tags are hyperlinked to a page displaying records in PubMed containing the term (and the original search terms). In this case the tagcloud is acting as an intermediate layer between specifying the search and results (Figure 1). The user can set a number of preferences when generating the tagcloud such as a date range, the number

[1] Pubcloud [Online] Available at URL: http://bioinfo.icapture.ubc.ca:8090/PubCloud/ Last access 14th July 2008.
[2] Medline [Online] Available at URL: http://www.ncbi.nlm.nih.gov/pubmed/ Last accessed 15th August 2008.

of tags to display, and which field to generate the tagcloud from (abstract, author field or assigned MeSH terms).

3.2 Methodology

A group of approximately 20 researchers (experts in their field) were contacted to participate in an experiment using PubCloud. Participants first carried out a questionnaire to establish a user profile, then familiarised themselves with the PubCloud tool, before carrying out the main experiment: searches on two topics within their area of expertise comprising of a "simple" and more "complex" search (i.e. 4 searches in total). Participants were free to search on any subject area, but to select topics within their area of expertise and/or research (see Table 1 for example searches). After completing the searches, participants were asked to complete a final questionnaire to establish their overall satisfaction with the tagclouds and gather feedback regarding utility within health-related information research. Questions regarding the utility of tagclouds were based on those used by Rivadeneira et al [1].

Table 1 Example "simple" and "complex" searches conducted by participants.

Simple search	**Complex search**
Consumer health information	Effect of HAART on information needs of people living with HIV/AIDS,
Malaria treatment	Mapping reviews or scoping reviews or systematic reviews and qualitative data
Ebola epidemic	Use of pattern detection in epidemiology since 1998
Bulimia	Effectiveness of St John's wort in treating depression 2001-2008
HIV and AIDS	

3.3 Participants

In total, 12 health-related professionals participated in the study from a variety of roles including librarian, lecturer, PhD student, communications manager, information and documentation officer and researcher. The median age of participants was 30-34 (25% 18-24; 10% 50-59), all were educated to a minimum of Bachelors-level (9 at a postgraduate level), all use the Internet at least once a day, and all search for health information more than once a week. All participants use PubMed more than once a month (50% more than once a month; 42% more than once a week; 8% more than once a day), and most were unfamiliar with the notion of a tagcloud (33% never heard of them; 33% familiar).

4. Results and Discussion

Based on the average (mean and median) of scores across all four search tasks undertaken by participants, Tables 2 and 3 show the results from the post-task questionnaire. The figures for *percent sum* are computed by totalling scores given by all participants as a proportion of the maximum score possible (i.e. if all participants gave maximum score, percent sum would be 100%). Overall, results show that the generated tagclouds provide effective summaries (61.5%), users are reasonable satisfied (58%) with content and appearance, the tagclouds often (74%) contain unexpected terms, but also contain irrelevant terms (67%). Firstly, considering the search task (simple or complex), Table 2 shows that overall the type of search task does not greatly affect the satisfaction of the tagcloud

generated by PubCloud (although definitely showing room for improvement). There were also definite types of queries which caused "poor" tagclouds (e.g. ambiguous terms).

Table 2 Mean, median and percent sum across all four tasks (1=definitely not; 5=very much so).

	Mean		Median		Percent Sum	
	Simple	**Complex**	**Simple**	**Complex**	**Simple**	**Complex**
Provides effective summary	3.03	3.08	3	3	61%	62%
Satisfied with tagcloud	2.86	3.00	3	3	57%	60%
Contains unexpected terms	3.81	3.58	4	4	76%	72%
Contains irrelevant terms	3.33	3.36	3	3	67%	67%

Table 3 shows results across all tasks, again, but this time considering the source of the tagcloud: abstract terms, authors or assigned MeSH terms. Overall, results are similar for the abstract and MeSH terms; the author tagclouds providing the least satisfying results. This is because author terms contain far less repetition and end up producing tagclouds with less pronounced terms, i.e. less aesthetically appealing. Results also indicate that MeSH results provide the most irrelevant terms (expected because the vocabulary is controlled), but that the author tagcloud also provides the most unexpected terms (i.e. highlighting new related authors).

Table 3 Mean, median and percent sum across all four tasks (1=definitely not; 5=very much so).

	Mean			Median			Percent Sum		
	Abstract	**Author**	**MeSH**	**Abstract**	**Author**	**MeSH**	**Abstract**	**Author**	**MeSH**
Provides effective summary	3.13	2.83	3.21	3	3	3	63%	57%	64%
Satisfied with tagcloud	3.08	2.75	2.96	3	3	3	62%	55%	59%
Contains unexpected terms	3.71	3.96	3.42	4	4	3.5	74%	79%	68%
Contains irrelevant terms	3.42	3.46	3.17	3	3	3	68%	69%	63%

Table 4 shows the results of the post-experiment questionnaire (ranked in descending order of percent sum score) with findings similar to those from previous research: that users find the tagclouds easy to use and intuitive and that the main utility of the tagcloud is for navigating to underlying content and to gain an impression of a subject area, rather than for performing search or browse. Participants also found the tagclouds useful as a means of suggesting alternative search terms (i.e. prompting further query terms).

Table 4 Mean, median and percent sum of final questions (1=definitely not; 5=very much so**)**.

	Mean	**Median**	**Percent Sum**
Overall easy to use	3.90	4	83.3%
Overall intuitive to use	3.48	4	76.7%
Overall useful for suggesting alternative terms	3.27	4	68.3%
Useful to gain impression of underlying subject area	3.22	3.5	68.3%
Useful for navigating to PubMed content	3.31	4	66.7%
Overall helpful to me	3.10	3.5	66.7%
Useful for mapping out findings	2.97	3	66.7%
Useful for browsing results	2.80	3	61.7%
Useful for performing search	2.47	2	58.3%

4.2 Participant's Comments

We asked participants for their comments on the generated tagclouds. Six respondents liked the emphasis given to the terms retrieved by the use of larger fonts and one person thought this preferable to a list. Three respondents liked the visualisation of the concept produced e.g. *"easy to gain an impression of the existing literature."* They found the tool easy to use, describing it as *"a simple idea"*, *"easy to scan"* and *"clear and concise."* One respondent found it useful for navigation: *"occasionally helped me to make links between keywords that otherwise hadn't occurred to me."*

With regards to search complexity, the reaction was mixed: two respondents commented explicitly that they found the tool more useful with simple searches; whilst one found it more useful with complex searches. One person felt its usefulness was most evident when scoping a topic: *"It felt as if the tag cloud would be at its most useful if I was to conduct my very first/early searches on a topic. The more familiar I am with the topics searched, the more restrictive the tag cloud felt."*

The author function received criticism for being messy, confusing and not producing a useful author map. This idea of being messy, busy and confusing/distracting was also identified by other respondents (5): *"... too many other words that distracted me.", "… there is a feeling that the cloud may be acting as a barrier to getting deeper results"* and *"Alphabetical organisation seemed clunky – a distraction from the more free-flowing map between terms."* A further concern regarded the accuracy and relevancy of displayed results: *"I would be concerned over how accurate it is. But I wouldn't rely on the tool alone so this doesn't really pose a problem"* and *"Sometimes so many terms that are not related to the query are displayed."*

4.3 Discussion

There is continuing debate on the utility and usability of tagclouds [8, 5, 2]. The developers of PubCloud carried out their own usability evaluation and found that although a need for improvement was identified in terms of summarising results, response to PubCloud was generally positive [4]. The findings from this study confirm these findings and indicate that overall the tool was found to be easy to use and useful. In fact, so easy that one respondent commented on the "addictive" quality of PubCloud, a characteristic attributed to other Web 2.0 applications [3], suggesting that this could be exploited in healthcare settings.

Table 5 Respondents suggestions for PubCloud regarding *visualisation of output*.

Comments from this study	**Related literature**
"Little visible information for each tag."	[4] discuss colour and layout.
"…adding a note in the result page showing more frequently used tag might be larger or brighter', …may be useful to guide users, particularly novices at PubCloud".	Tag weighting discussed by [9].
"Perhaps random order words would be useful..."	Alphabetisation addressed by [6].
"It would be useful to draw clearer size distinctions in the 'author' search. The authors often came up with the same prominence – making it difficult to establish who the key authors were."	Font size addressed by [6].
"They do help zero-in on important topics. However, they are a bit off-the-mark at times and the enlarged font size of a red herring will be obviously misleading…"	Presentation of tags addressed by [6]; [5] discuss visualisation design flaws.
"Too many terms. Most of them not particularly relevant to the search. I guess this problem is faced by most search tools".	[9] discuss specificity.

General comments given on usability were mainly positive, including: *"I think it is great, very useful."; "It has potential if developed further"*, *"Certainly provides an interesting alternative to regular search engines, and I'd like to see it more widely available on browsers"*, *"The tag cloud certainly "softens" the health information search process, making it more accessible to non-academics"* and *"It is highly recommended in an information health setting."* Overall tagclouds and PubCloud in particular, were considered useful and to have potential, although a need for development and refinement was identified other than those discussed which can be categorised as visualisation of output (Table 5) and technical functionality (Table 6).

Table 6 Respondents suggestions for PubCloud regarding *technical functionality*.

Comments from this study	**Related literature**
"Need more options to narrow articles."	[9] discuss specificity.
"Maybe this tool could be modified to be more context-aware in conducting searches".	[7] conclude tag-clouds best for broad categorisation.
"…but there is a problem with the number of search terms it can take in one go…when I typed in a number of terms in one sequence, it couldn't generate any results".	[7] identified problems with inaccessibility.
"…providing a function to sort results by frequency or recency which can help user find their targets quicker…"	[4, 6, 7] address speed in relation to information retrieval and tagclouds.
"Of course it's only as good as the algorithm that creates the tag cloud." *"I like the overall concept but not the execution."*	[9, 10] consider ways to improve tagcloud functionality, e.g. clustering tags.
"Develop the advanced search options so that they mirror Pubmed functionality more closely" *"Let it be linked to other health information databases."* *"MeSH terms need to have slight improvement."*	Health specific development not currently addressed in the literature

The respondents were asked if they could see potential uses for the tag clouds, both for the health professional or the academic: *"…in journals to accompany an abstract alongside keywords particularly e-journals", "As a teaching tool to support information literacy."*, *"Perhaps as a visualisation for popular searches…"* And for patients: *"For use with newly diagnosed patients, to be able to create a tag cloud for their condition…"*, *"…Perhaps a tag cloud of 5-10 words only would best help the less IT literate to operate PubMed more effectively than if they used the PubMed interface direct."* and *"Perhaps it could be delivered through attractive, colourful touch-screen in GP surgeries, pharmacies and public libraries."*

5. Conclusions

This paper reports an experiment to assess the utility of tagclouds in health-related information research. A total of 12 expert users participated in the study by carrying out a range of simple and complex searches on preferred topics which were familiar to them using PubCloud, a tagcloud visualisation based on PubMed. Participants completed a series of questions to assess their satisfaction with PubCloud and provide feedback on areas for improvement and potential applications of tagclouds in health-related information research. Findings indicate that Pubcloud is easy to use and intuitive. It is a useful tool for gaining an impression of the underlying subject area and for suggesting alternative terms. It is less useful for performing searches. Users suggested a number of improvements, but also saw potential for both health professionals and healthcare consumers. Comments from participants fell into two main categories: technical functionality and visualisation of

the output. Future work might include exploring instances when the tagclouds do not work in more detail (e.g. identifying potential query failure), carrying out a larger lab-based contrastive user study, and improving the PubCloud visualisation based on current research from the second-generation tagcloud research.

Acknowledgements

Our thanks go to the participants for their very helpful comments. Particular thanks go to Byron Kuo for permitting the use of PubCloud. This work is partly supported by EC under the Information Society Technologies (IST) programme of the 6th FP for RTD - project MultiMATCH contract IST-033104.

References

[1] Rivadeneira, A. W., Gruen, D. M., Muller, M. J., and Millen, D. R. *Getting our head in the clouds: toward evaluation studies of tagclouds.* In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (San Jose, California, USA, April 28 - May 03, 2007). CHI '07. ACM, New York, NY, 995-998.

[2] Viégas, F. B. and Wattenberg, M. *TIMELINES Tag clouds and the case for vernacular visualization*. Interactions 15(4) (Jul. 2008), 49-52.

[3] Boulos, M. N. K. and Wheeler, S. *The emerging Web 2.0 social software: an enabling suite of sociable technologies in health and health care education*. Health Information Libraries Journal, 2007, 24 (1) 2-23.

[4] Kuo, B. Y., Hentrich, T., Good, B. M., and Wilkinson, M. D. *Tag clouds for summarizing web search results*. In Proceedings of the 16th international Conference on World Wide Web (Banff, Alberta, Canada, May 08 - 12, 2007). WWW '07. ACM, ACM 978-1-59593-654-7/07/0005, New York, NY, 1203-1204.

[5] Hearst, M. A. and Rosner, D. *Tag clouds: Data analysis tool or social signaller?* In. Proceedings of 41st Hawaii International Conference on System Sciences – 2008. 1530-1605.

[6] Halvey, M. J. and Keane, M. T. *An assessment of tag presentation techniques.* In Proceedings of the 16th international conference on World Wide Web, (Banff, Alberta, Canada, May 08 – 12, 2007). WWW '07, ACM, New York, NY, 1313-1314.

[7] Sinclair, J. and Cardew-Hall, M. *The folksonomy tag cloud: when is it useful?* Journal of Information Science. 2008, 34 (1) 15-29.

[8] Nielson, M. *Functionality in a second generation tag cloud.* Masters Thesis. Department of Computer Science and Media Technology. Gjøvik University College, Norway. 2007.

[9] Hassan-Montero, Y. and Herrero-Solana, *V. Improving tag-clouds as visual information retrieval* interfaces. In Proceedings of International Conference on Multidisciplinary Information Sciences and Technologies, InSciT2006. Mérida, Spain. October 25-26, 2006.

[10] Begelman, G., Keller, P., and Smadja, F. Automated Tag Clustering: Improving search and exploration in the tag space. In Proceedings of the Fifteenth International World Wide Web Conference (WWW2006) (Edinburgh, Scotland, May 22-26, 2006). ACM Press, New York, NY, 2006.