Tag cloud

From Wikipedia, the free encyclopedia

A **tag cloud** (**word cloud**, or **weighted list** in visual design) is a visual representation of text data, typically used to depict keyword metadata (tags) on websites, or to visualize free form text. Tags are usually single words, and the importance of each tag is shown with font size or color.[2] This format is useful for quickly perceiving the most prominent terms and for locating a term alphabetically to determine its relative prominence. When used as website navigation aids, the terms are hyperlinked to items associated with the tag.

foundation-l word cloud, created with the complete gzip'ed list archives (without duplicate emails from archives and all headers and quoted text in body), using IBM Word Cloud Generator build 32.[1]

Contents

- 1 History
- 2 Types
  - 2.1 Frequency
  - 2.2 Categorization
- 3 Visual appearance
- 4 Data clouds
- 5 Text clouds
- 6 Collocate clouds
- 7 Perception of tag clouds
- 8 Creation of a tag cloud
- 9 See also
- 10 References
- 11 External links

A tag cloud with terms related to Web 2.0

History

In the language of visual design, a tag cloud (or word cloud) is one kind of "weighted list", as commonly used on geographic maps to represent the relative size of cities in terms of relative typeface size. An early printed example of a weighted list of English keywords was the "subconscious files" in Douglas Coupland's *Microserfs* (1995). A German appearance occurred in 1992.[3]

The specific visual form and common use of the term "tag cloud" rose to prominence in the first decade of the 21st century as a widespread feature of early Web 2.0 websites and blogs, used primarily to visualize the frequency distribution of keyword metadata that describe website content, and as a navigation aid.

The first tag clouds on a high-profile website were on the photo sharing site Flickr, created by Flickr co-founder and interaction designer Stewart Butterfield in 2004. That implementation was based on Jim Flanagan's Search Referral Zeitgeist,[4] a visualization of Web site referrers. Tag clouds were also popularized around the same time by Del.icio.us and Technorati, among others.

Oversaturation of the tag cloud method and ambivalence about its utility as a web-navigation tool led to

a noted decline of usage among these early adopters.[5][6] (Flickr would later "apologize" to the web-development community in their five-word acceptance speech for the 2006 "Best Practices" Webby Award, where they simply stated "sorry about the tag clouds.")[7]

A second generation of software development discovered a wider diversity of uses for tag clouds as a basic visualization method for text data. Several extensions of tag clouds have been proposed in this context. Examples include Parallel Tag Clouds,[8] SparkClouds,[9] and Prefix Tag Clouds.[10] The Word Cloud Explorer, written in Adobe Flex, combines tag clouds with a number of interactive features for text analysis.[11]

Types

There are three main types of tag cloud applications in social software, distinguished by their meaning rather than appearance. In the first type, there is a tag for the frequency of each item, whereas in the second type, there are global tag clouds where the frequencies are aggregated over all items and users. In the third type, the cloud contains categories, with size indicating number of subcategories.

Frequency

In the first type, size represents the number of times that tag has been applied to a single item.[12] This is useful as a means of displaying metadata about an item that has been democratically "voted" on and where precise results are not desired. Examples of such use include Last.fm (to indicate genres attributed to bands) and LibraryThing (to indicate tags attributed to a book).

In the second, more commonly used type, size represents the number of items to which a tag has been applied, as a presentation of each tag's popularity. Examples of this type of tag cloud are used on the image-hosting service Flickr, blog aggregator Technorati and on Google search results with DeeperWeb.

A data cloud showing the population of each of the world's countries. Created in R with wordcloud package. Data from Country population. Note that the proportional sizes of China and India were divided in half.

Categorization

In the third type, tags are used as a categorization method for content items. Tags are represented in a cloud where larger tags represent the quantity of content items in that category.

There are some approaches to construct tag clusters instead of tag clouds, e.g., by applying tag co-occurrences in documents.[13]

More generally, the same visual technique can be used to display non-tag data,[14] as in a word cloud or a data cloud.

The term **keyword cloud** is sometimes used as a search engine marketing (SEM) term that refers to a group of keywords that are relevant to a specific website. In recent years tag clouds have gained popularity because of their role in search engine optimization of Web pages as well as supporting the user in navigating the content in an information system efficiently.[15] Tag clouds as a navigational tool make the resources of a website more connected,[16] when crawled by a search engine spider, which

may improve the site's search engine rank. From a user interface perspective they are often used to summarize search results to support the user in finding content in a particular information system more quickly.[17]

Visual appearance

Tag clouds are typically represented using inline HTML elements. The tags can appear in alphabetical order, in a random order, they can be sorted by weight, and so on. Sometimes, further visual properties are manipulated in addition to font size, such as the font color, intensity, or weight.[18] Most popular is a rectangular tag arrangement with alphabetical sorting in a sequential line-by-line layout. The decision for an optimal layout should be driven by the expected user goals.[18] Some prefer to cluster the tags semantically so that similar tags will appear near each other.[19][20][21] Heuristics can be used to reduce the size of the tag cloud whether or not the purpose is to cluster the tags.[20]

A data cloud showing stock price movement. Color indicates positive or negative change, font size indicates percentage change.

Data clouds

A **data cloud** or **cloud data** is a data display which uses font size and/or color to indicate numerical values.[22] It is similar to a tag cloud[23] but instead of word count, displays data such as population or stock market prices.

Text clouds

A **text cloud** or **word cloud** is a visualization of word frequency in a given text as a weighted list.[25] The technique has recently been popularly used to visualize the topical content of political speeches.[24][26]

Collocate clouds

Extending the principles of a text cloud, a **collocate cloud** provides a more focused view of a document or corpus. Instead of summarising an entire document, the collocate cloud examines the usage of a particular word. The resulting cloud contains the words which are often used in conjunction with the search word. These collocates are formatted to show frequency (as size) as well as collocational strength (as brightness). This provides interactive ways to browse and explore language.[27]

Text cloud comparing 2002 State of the Union Address by U.S. President Bush and 2011 State of the Union Address by President Obama.[24]

Perception of tag clouds

Tag clouds have been subject of investigation in several usability studies. The following summary is based on an overview of research results given by Lohmann et al.:[18]

- Tag size: Large tags attract more user attention than small tags (effect influenced by further

   properties, e.g., number of characters, position, neighboring tags).
- Scanning: Users scan rather than read tag clouds.
- Centering: Tags in the middle of the cloud attract more user attention than tags near the borders (effect influenced by layout).
- Position: The upper left quadrant receives more user attention than the others (Western reading habits).
- Exploration: Tag clouds provide suboptimal support when searching for specific tags (if these do not have a very large font size).

Creation of a tag cloud

In principle, the font size of a tag in a tag cloud is determined by its incidence. For a word cloud of categories like weblogs, frequency, for example, corresponds to the number of weblog entries that are assigned to a category. For smaller frequencies one can specify font sizes directly, from one to whatever the maximum font size. For larger values, a scaling should be made. In a linear normalization, the weight t_i of a descriptor is mapped to a size scale of 1 through f, where t_{min} and t_{max} are specifying the range of available weights.
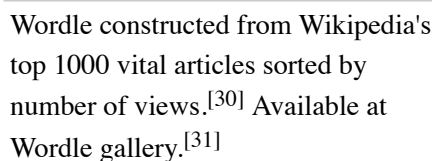
$$s_i = \left\lceil \frac{f_{\max} \cdot (t_i - t_{\min})}{t_{\max} - t_{\min}} \right\rceil \text{ for } t_i > t_{\min}; \text{ else } s_i = 1$$

- s_i: display fontsize
- $f_{\max}$: max. fontsize
- t_i: count
- $t_{\min}$: min. count
- $t_{\max}$: max. count

Since the number of indexed items per descriptor is usually distributed according to a power law,[28] for larger ranges of values, a logarithmic representation makes sense.[29]

Implementations of tag clouds also include text parsing and filtering out unhelpful tags such as common words, numbers, and punctuation.

There are also websites creating artificially or randomly weighted tag clouds, for advertising, or for humorous results.

See also

- Concordance
- Folksonomy
- Keywords
- tf-idf

References

1. Word-Cloud Generator   (archive) (http://web.archive.org /web/20110719002236 /http: //www.alphaworks.ibm.co m/tech/wordcloud)

Wordle constructed from Wikipedia's top 1000 vital articles sorted by number of views.[30] Available at Wordle gallery.[31]

2. Martin Halvey and Mark T. Keane, An Assessment of Tag Presentation Techniques (http://www2007.org/htmlposters/poster988/), poster presentation at WWW 2007, 2007

3. Gilles Deleuze, Felix Guattari (1992). *Tausend Plateaus. Kapitalismus und Schizophrenie*. ISBN 3-88396-094-2.

4. A copy of Jim Flanagan's Search *Referral Zeitgeist* was available at archive.org (http://web.archive.org/web/20041204231120 /http://twiki.tensegrity.net/bin/view /Main/SearchReferralZeitgeist) but has since been blocked. In the comments of a blog entry (http://www.37signals.com/svn/archives /000937.php), a user identified as Steve Minutillo attribute the idea to Jim Flanagan, stating that Flanagan's site had such displays in 2002.

5. "Tag Clouds R.I.P.?". Readwriteweb.com. 2011-03-30.

6. "The Death of Tag Clouds". Woorkup.com. Retrieved 2011-05-23.

7. "Welcome to the Webby Awards". Webbyawards.com. 2011-10-28. Retrieved 2013-07-27.

8. Collins, C., Viegas, F. and Wattenberg, M., Parallel Tag Clouds to Explore and Analyze Faceted Text Corpora. Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST 2009), pp. 91–98, IEEE, 2009.

9. Lee, B., Riche, N., Karlson, A., and Carpendale, S., SparkClouds: Visualizing Trends in Tag Clouds (http://research.microsoft.com/en-us /um/people/nath/docs /sparkclouds_infovis2010.pdf). IEEE Transactions on Visualization and Computer Graphics, vol. 16, no. 6, pp. 1182–1189, 2010.

10. Burch, M., Lohmann, S., Pompe, D., Weiskopf, D., Prefix Tag Clouds (http://www.vis.uni-stuttgart.de/uploads/tx_vispublications /PrefixTagClouds-IV2013.pdf). Proceedings of the 17th International Conference on Information Visualisation (IV 2013), pp. 45-50, IEEE, 2013.

11. http://wordclouds.visualdataweb.org/

12. Bielenberg, K. and Zacher, M., Groups in Social Software: Utilizing Tagging to Integrate Individual Contexts for Social Navigation (http://bielenberg.info/thesis.pdf), Masters Thesis submitted to the Program of Digital Media, Universität Bremen (2006)

13. Knautz, K., Soubusta, S., & Stock, W.G. (2010). Tag clusters as information retrieval interfaces (http://www.phil-fak.uni-duesseldorf.de /fileadmin/Redaktion/Institute /Informationswissenschaft/stock /Knautz_Soubusta_Stock.pdf). Proceedings of the 43rd Annual Hawaii International Conference on System Sciences (HICSS-43), January 5–8, 2010. IEEE Computer Society Press (10 pages).

14. Kamel Aouiche, Daniel Lemire, Robert Godin, Collaborative OLAP with Tag Clouds: Web 2.0 OLAP Formalism and Experimental Evaluation (http://arxiv.org/abs/0710.2156), *Webist* 2008, 2008.

15. Helic, D.; Trattner, C.; Strohmaier, M.; Andrews, K. (2011). "Are Tag Clouds Useful for Navigation? A Network-Theoretic Analysis" (PDF). *Journal of Social Computing and Cyber-Physical Systems*. **1** (1): 33–55. doi:10.1504/IJSCCPS.2011.043603.

16. Trattner, C.:Linking Related Content in Web Encyclopedias with search query tag clouds (http://www.austria-lexikon.at/attach /User/Trattner%20Christoph /ctrattner_IADIS_WWW_journal.pdf). IADIS International Journal on WWW/Internet, Volume 9, Issue 2, 2011

17. Tratter, C., Lin, Y., Parra, D., Yue, Z., Brusilovsky, P.: Evaluating Tag-Based Information Access in Image Collections (http://www.austria-lexikon.at/attach /User/Trattner%20Christoph/ht076-trattner.pdf). In Proceedings of the 23rd ACM Conference on Hypertext and Social Media (HT 2012). ACM, New York, NY, USA, 2012

18. Lohmann, S., Ziegler, J., Tetzlaff, L. Comparison of Tag Cloud Layouts: Task-Related Performance and Visual Exploration (http://www.uni-due.de/~s400268/Lohmann09-interact.pdf), T. Gross et al. (Eds.): INTERACT 2009, Part I, LNCS 5726, pp. 392–404, 2009.

19. Hassan-Montero, Y., Herrero-Solana, V. Improving Tag-Clouds as Visual Information Retrieval Interfaces (http://www.nosolousabilidad.com/hassan /improving_tagclouds.pdf). InSciT 2006: Mérida, Spain. October 25–28, 2006.

20. Owen Kaser and Daniel Lemire, Tag-Cloud Drawing: Algorithms for Cloud Visualization (http://arxiv.org/abs/cs/0703109), Tagging and Metadata for Social Information Organization (WWW 2007), 2007

21. Salonen, J. 2007. Self-organising map based tag clouds - Creating spatially meaningful representations of tagging data (http://matriisi.ee.tut.fi/hypermedia/julkaisut /2007-salonen-som-clouds.pdf). Proceedings of the 1st OPAALS conference, 26–27 November 2007, Rome, Italy.

22. Apel, Warren. "ManyEyes Visualization and Commentary: *World Population Data Cloud*.". Retrieved 2007-08-26.

23. Wattenberg, Martin. "ManyEyes Visualization: *Ad cloud*". Retrieved 2007-03-12.

24. Steinbock, Daniel. "TagCrowd visualization: State of the Union". Retrieved 2011-03-05.

25. Lamantia, Joe. "Text Clouds: A New Form of Tag Cloud?". Retrieved 2008-09-11.

26. Mehta, Chirag. "US Presidential Speeches Tag Cloud". Retrieved 2008-09-11.

27. "Collocate cloud". Retrieved 2008-12-05.

28. Jakob Voss: *Collaborative thesaurus tagging the Wikipedia way*. April 2006 [1] (http://arxiv.org/abs/cs/0604036)
29. "Kentbyte: "Tag Cloud Font Distribution Algorithm". June 2005". Echochamberproject.com. Retrieved 2013-07-27.
30. "Monthly wiki page Hits for en.wikipedia". Wikistics.falsikon.de. 2009-08-31. Retrieved 2013-07-27.
31. "WikipediaTop1000VitalArticleHits". Wordle. Retrieved 2013-07-27.

External links

- Understanding Tag Clouds (http://www.joelamantia.com/ideas/tag-clouds-evolve-understanding-tag-clouds) - an information design analysis of tag clouds
- Tag Clouds Gallery: Examples and Good Practices (http://www.smashingmagazine.com/2007/11/07/tag-clouds-gallery-examples-and-good-practices/) - comparison of tag cloud visual designs
- Design tips for building tag clouds (http://www.onlamp.com/pub/a/onlamp/2006/06/08/designing-tag-clouds.html) - software development guide from O'Reilly's ONLamp
- 9 Word Cloud generators in one place (http://www.smashingapps.com/2011/12/15/nine-excellent-yet-free-online-word-cloud-generators.html) - these are all free to experiment with.
- AnswerGarden (http://answergarden.ch) - A realtime social tagcloud creation tool.

Wikimedia Commons has media related to *Word clouds*.

Retrieved from "https://en.wikipedia.org/w/index.php?title=Tag_cloud&oldid=740416063"

Categories: Web 2.0 neologisms | Visualization (graphic)