Using Font Attributes in Knowledge Maps and Information Retrieval

Richard Brath[1] and Ebad Banissi[1]

London South Bank University, London, U.K.

Abstract. Font specific attributes, such as bold, italic and case can be used in knowledge mapping and information retrieval to encode additional data in texts, lists and labels to increase data density of visualizations; encode data quantitative data into search lists; and facilitate text skimming and refinement by visually promoting of words of interest.

1 Why Font Attributes?

Information visualization (infovis) transforms data into visual representations. In knowledge mapping, visualizations are used to gain insight into the structure of large scale information spaces. In knowledge maps, similar to geographic maps, text should have an inherent role to help viewer comprehend information, however, the use of font-specific attributes, such as bold, italic, caps, etc., in infovis is uncommon for encoding additional information. In information retrieval, search results may use a few font-attributes, e.g. bold, underline, serif/sans serif, to differentiate classes of metadata.

The goal of this paper is to illustrate that font-specific attributes can be used to: 1) facilitate skimming texts such as abstracts or lead paragraphs; 2) encode quantitative data using a novel technique of proportional encoding in search results and facets; and 3) encode multiple data attributes in labels.

2 Background

Knowledge maps frequently use text labels: Places & Spaces (scimaps.org) is a repository of information visualizations and maps typically organizing large information spaces (i.e. knowledge maps). Of 144 maps, 80% use some form of text in the central visualization. When text is used, 2/3 use traditional infovis attributes of size and color (e.g. text size corresponding to size of a region or size of a node). Text-specific attributes are used in 28% of the examples, however, these are typically used only to differentiate between compositional elements (e.g. labels, axes, tick labels, hyperlinks, city, region, body of water in a map). In only a few instances (mostly maps, infographics and a few infovis) are a broader mix of font attributes used, e.g. case, italics and spacing [7, 18].

Information retrieval infovis has a similar usage. Of 45 examples in Hearst's infovis chapters [8], half use traditional visual attributes of size and/or color. There are 13 examples using one type-specific attribute, either bold, caps, or font family, and in most cases these are used to either highlight a search term or differentiate between types of data, e.g. category title vs. category instance; axis title vs. tick label.

Other infovis also use font attributes, e.g. italics [16], uppercase [5] or bold [6]. Innovators include Baecker & Marcus [2] who utilize bold, italics, font size, underlines, serif/sans-serif to enhance readability of computer code - a practice now commonplace in most code editors. Fat fonts [15], is a specialized font that varies font weight per character so that the ink varies in proportion to the numeric value represented. Muriel Cooper's Visible Language Workshop explored 3D typographic spaces with variations in size, case, color and font family, e.g. [20]. Typographic maps [1] uses only type to create geographic maps.

Typography and **cartography** have centuries of history with innovative font encoding of information in documents hundreds of years old (e.g. fig. 1). There are many techniques for creating emphasis and differentiation with font, with various guidelines and conventions (e.g. typographic [21, 13], cartographic [11, 17], user interface design [12, 10, 22]).

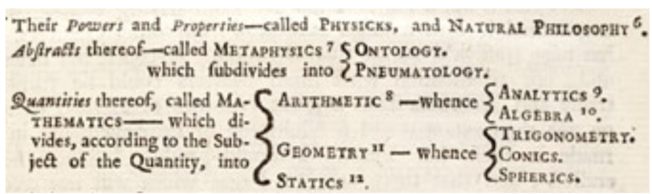
Their *Powers* and *Properties*—called PHYSICKS, and NATURAL PHILOSOPHY 6.
Abstracts thereof—called METAPHYSICS 7 which subdivides into { ONTOLOGY. PNEUMATOLOGY.
Quantities thereof, called MATHEMATICS—which divides, according to the Subject of the Quantity, into { ARITHMETIC 8 —whence { ANALYTICS 9. ALGEBRA 10. } GEOMETRY 11 — whence { TRIGONOMETRY. CONICS. SPHERICS. } STATICS 12.

Fig. 1. Structure of knowledge from *Cyclopaedia* (1728) using italics for broad topics, small caps for specific fields and roman for explanatory text.

Historically, user interfaces recommended against type attributes due to low resolution displays. New higher resolution devices, improved font rendering technology, a wide range of typefaces designed for the screen and rich markup formats, now result in recommendations to more broadly use type-specific attributes for user interfaces and web.

In contrast to infovis for information retrieval, text-based search results and navigation interfaces typically differentiate metadata associated with a document using type size, color, underlines (e.g. links) or font family (e.g. titles). Bold or color is frequently used to highlight search terms. See Hearst [8] or popular search interfaces, e.g. Google, Yahoo, Bing, Ebay, Amazon, NYTimes, LinkedIn, etc.

Notation systems such as chemical formulas (e.g. $[As@Ni_{12}As_{20}]^{3-}$), mathematical formulas (e.g. $\mu_e(A) = \inf\{\lambda_*(O) \mid O \in \mathcal{O}, A \subset O\}$) and markup notation (e.g. <div class="body">Text </div>) use different type elements to emphasize, delineate or otherwise add information to text.

Based on an review of these above domains, a list of font-specific properties (not including generic color and size visual attributes) include:

- **Weight** (bold) can have up to six weights for screen and up to 9 for print.
- *Italic* or *Oblique* are both sloped fonts but italics have different letterforms and there are instances of reverse italics and vertical italics.

- CASE includes UPPER, lower, Mixed and SMALL CAPS. Uppercase is designed to standout from lowercase while small caps blend in.
- S p a c i n g . Tracking (space between letters) and leading (between lines).
- Typeface indicates font family: sans, blackletter, *script*, `source`, MATHBOLD,...
- Underline can be distracting. Typographers recommend subtle variants.
- Condensed/Expanded. Similar to, but better than, horizontally scaled fonts.
- Superscript and subscript encode via size and position to adjacent text.
- "Paired delimiters" evoke enclosure by pairing the same (or mirrored) shapes.
- Alphanumeric glyphs (A,B,C,1,2,3) can literally encode data and are uniquely orderable. Glyphs not native to the viewer (e.g. α, β, γ) are also orderable, but symbols (e.g. $\infty, \forall, \flat$) are not orderable.

3 Type for Knowledge Maps and Information Retrieval

Type attributes can be used to encode additional information at different levels of use in knowledge mapping and information retrieval, ranging from low-level document views, to search lists, to the macro-level overviews. A few examples:

3.1 Type Visualization on Texts

In some search results, full sentences, abstracts or lead paragraphs are presented (e.g. BioText Search Engine, Wikipedia's Today's featured article archive). At the micro-level of documents and texts, type attributes can be used to aid comprehension without changing layout. Text skimming is a reading technique of rapid eye movement across a large body to text to get the main ideas and content overview. At a low level, the strategy requires the reader to dip into the text looking for words such as proper nouns, unusual words, enumerations, etc. Word frequency analysis can be used to weight the least common words (fig. 2).

> *The* **flights** *of the* 1902 **glider** had **demonstrated** *the* **efficiency** *of our* system *for* **maintaining equilibrium,** *and* also *the* **accuracy** *of the* **laboratory** work *upon* which *the* **design** *of the* **glider** was **based.** *We* then felt *that we* were prepared *to* **calculate** *in* **advance** *the* **performance** *of* **machines** *with a* degree *of* **accuracy** *that* had never been possible *with the* **data** *and* **tables possessed** *by our* **predecessors.** *Before* leaving camp *in* 1902 *we* were already *at* work *on the* general **design** *of a* new **machine** which *we* **proposed** *to* **propel** *with a* **motor.**

Fig. 2. Lead paragraph of *How We Made the First Flight* by Orville Wright formatted for skimming using font weight and with search term underlined.

Font weight draws visual attention to the highest contrast which are the least frequent words as per text skimming strategy. In the above introductory paragraph on flight, the terms *glider* and *motor* have the heaviest weight and are possible terms for query refinement. Similarly, visual weighting of proper nouns, enumerations and unusual words facilitates fact-finding tasks. Attributes could instead be based on specific domain vocabulary [14] with interactions (e.g. drag and drop) to facilitate guided search. Other type attributes can also be applied, e.g. in the above figure dotted underlines indicate search terms. Also, less important parts of speech (e.g. articles, pronouns, etc) are italicized to enhance figure-ground separation between the heavy-weight words and background.

3.2 Type Visualization on Lists

Query results can be visualized and may include quantitative data e.g. readership, relevance or number of citations. Newsmap.jp displays news headlines in a treemap, with headline size indicating readership (fig. 3).

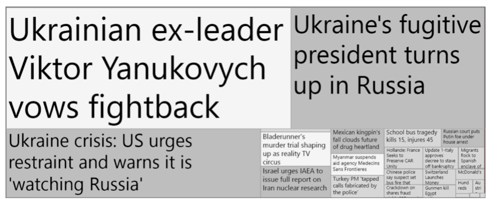

Fig. 3. Top: Portion of Newmap [23] from 02/28/2014. Size represents readership.

Instead, using fixed size text enables more legible headlines to be displayed and then font-weight can be used to encode readership, either by setting the weight per headline (fig. 4-left), or proportionally encoding readership by setting bold to proportionally correspond to the magnitude (fig. 4-right).

Fig. 4. Top: Left: Same headlines as fig. 3 with font weight indicating readership. Right: Same with proportion of bold indicating readership.

Any visualization technique has some degree of lossiness as data is transformed into a visual encoding. Lossiness can be evaluated in the above representations by measuring:

1. *Number of readable headlines:* if a headline is too truncated, too small to read or does not appear, it is considered unreadable. Six point was used as the threshold for too small for a 96dpi screen.

2. *Number of uniquely perceivable sizes:* can be estimated using prior experimental data [9], e.g. area ± 5% or length ±2.5%. For font weight, a show of hands in seminar settings yielded the most responses for 4 levels of weights.
3. *Area of the overall plot.*

Information density is a function of the measures (#readable x #sizes / area). Density was measured across all three variants and repeated for different aspect ratios with different numbers of items (e.g. sparse scenario, dense scenario). Information density for the proportional encoding consistently outperformed the treemap by a factor of 2 while the font weight encoding could underperform:

Variant	Normal	Dense	Sparse
Font Weight	1.42x	0.67x	0.86x
Proportion	2.68x	2.22x	2.07x

These three different encodings were presented to three different groups of infovis researchers, each with more than 10 attendees. Proportional encoding received a positive response. None of the participants were confused by the encoding and consistently scored 3 or 4 on a 1-4 scale on three questions indicating desirability, ease of use and likelihood of ease for others to understand.

One concern expressed with proportional encoding is that the maximum proportion length is limited to the shortest headline. This can be addressed by adding the beginning of the lead sentence to the end of the headline, similar to email lists (fig. 5). Also, the approach can be extended to convey multiple data attributes, e.g. (font weight and underline).

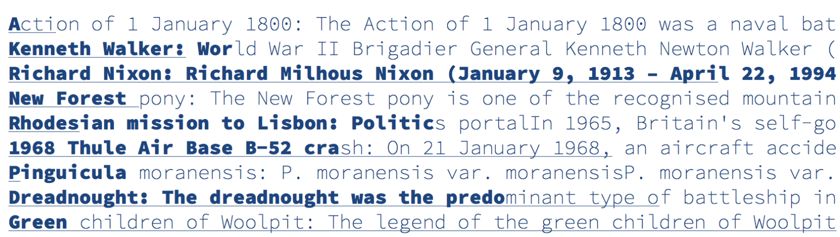

Fig. 5. Wikipedia articles with bold length indicating number of authors and underline length indicating number of readers.

This approach could lead to new techniques for rapidly scanning through search results, particularly in applications where each result is presented as an individual row. For example, news search on financial terminals (e.g. Reuters, Bloomberg) typically result in dense lists of headlines where space is at a premium, so visual techniques such as icons or color are used to indicate additional information; and font-based approaches could also be used to provide information without requiring any additional space. Similarly, facets for query refinement may have additional metadata and typically exist in narrow side panels. For example, fig. 6 represents facets for query refinement provided by Amazon on a search for *information retrieval* where the length of underline has been added to indicate the quantity of matching items.

Network Administration | Computers & Technology | Databases |
Programming | Web Development & Design | Reference

Fig. 6. Facets with length of underline indicating the quantity of matching items.

3.3 Type Visualization on Macro-views

Macro-scale visualizations of large domains of information usually have some text but sometimes text can be minimal. For example, choropleth maps color regions to indicate data values and are extremely popular. However, choropleth maps have issues, including: 1) *Small areas* can be invisible (e.g. Dubai, Singapore are not visible on a world map); 2) *Identification* of a selected area or finding a named target can be difficult e.g. 63% of young adults in USA could not locate Iraq on a map in a National Geographic survey in 2006; and 3) *Data encoding* is typically limited to a single value, such as hue or brightness.

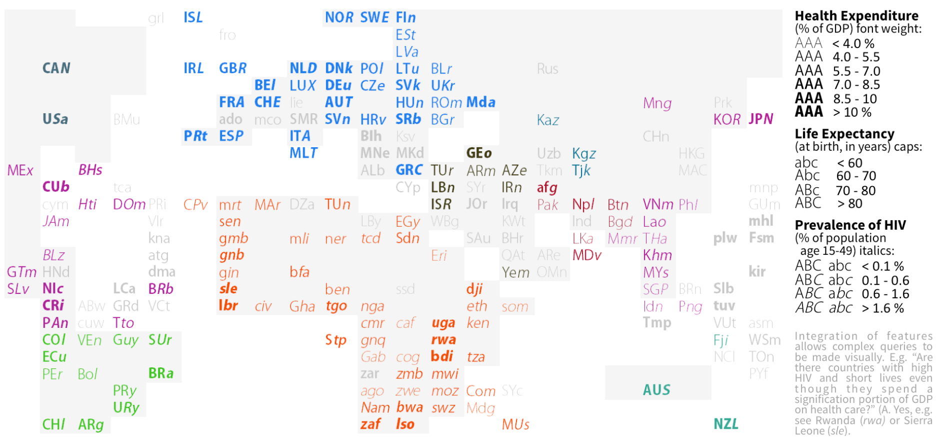

Fig. 7. Map indicating data via label with color, bold, italic and case.

Instead of using shapes to identify regions, mnemonic codes (e.g. ISO country codes) are used. With some spatial adjustment to remove overlap, all labels are visible while maintaining local proximity. Color can still be used, as well as type-specific attributes to encode additional data. Health expenditures, life expectancies and HIV data are represented using bold, case and italic into a single label in fig. 7 allowing complex queries to be made visually, e.g. Are there countries with high HIV (italics) and short lives (lowercase) even though a signification portion of GDP is spent on health care (ultrabold)? (A. Yes, e.g. Rwanda (***rwa***) or Sierra Leone (***sle***)).

In addition to offering greater data density and no loss of small countries, a similar map (using only labels and color) resulted in 62% correct responses on identification tasks compared to 17% correct for a shape-based map with no labels (i.e. choropleth map); in surveys in a seminar setting. This indicates a strong potential for richly labeled maps to offer greater information density and lower data lossiness.

In knowledge maps (e.g. scimaps.org) representations range from highly labeled maps such as Skupin's self-organizing maps (e.g. [19]) to graphs with many or no labels (e.g. [16, 3]). This approach can be applied to any map or graph with minimally overlapping labels (e.g. fig. 8).

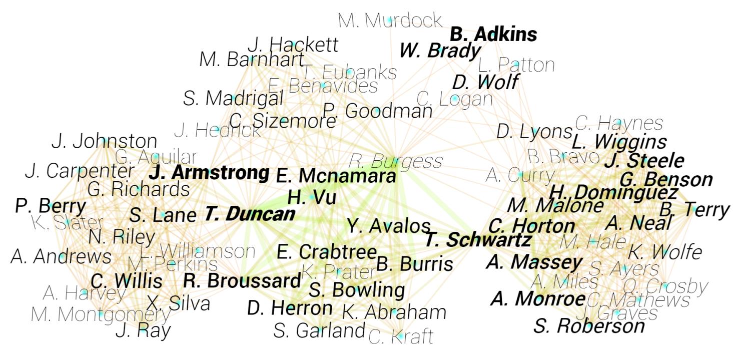

Fig. 8. Graph of people indicating data by font weight and oblique angle.

4 Conclusion

Feedback, metrics and informal surveys are promising and indicate that font-specific attributes can be used to increase data density in texts (e.g. abstracts, lead paragraphs) and lists (e.g. results lists, facets) to add additional information to aid tasks such as fact finding, information gathering and query refinement (e.g. fig. 2,5,6). Font-specific attributes could also be potentially used on labeled knowledge maps to add additional information (e.g. fig. 7).

These typographic attributes may not have the same degree of effectiveness as other visual channels (e.g. size, hue, angle, shape). Visual channels have been researched in more detail (e.g. Bertin, Cleveland, MacKinlay, Wilkinson, Mazza, Munzner), although a definitive list ranking visual channels for different types of tasks and number of uniquely perceivable levels does not exist. Font-specific attributes can be mapped back to these well-known visual channels [4] and then visual attribute heuristics can be used for guidance. For example, font weight, which utilizes visual channels of size (i.e. line width) and intensity, will rank higher for effectiveness of ordered or quantitative data encodings than font family, which utilizes the visual channel of shape. These existing heuristics can also be used to identify combinations of font-attributes that may be integral (e.g. capitalization + italic) vs. separable (e.g. font weight + underline). While the mapping of font-specific attributes to more general visual channels can provide an initial indication of promising attributes and combinations, usability testing should be done in the future to validate these approaches.

Future work should also consider different representations e.g. there may be other novel visualizations; as well as systems implemented to evaluate applicability and efficacy to specific information retrieval tasks.

References

1. Afzal, S., Maciejewski, R., Jang, Y., Elmqvist, N., Ebert, D.S.: Spatial text visualization using automatic typographic maps. Visualization and Computer Graphics, IEEE Transactions on 18(12), 2556–2564 (2012)
2. Baecker, R., Marcus, A.: Human Factors and Typography for More Readable Programs. Addison-Wesley (1990)
3. Boyack, K.W., Klavans, R.: Creation of a highly detailed, dynamic, global model and map of science. Journal of the Association for Information Science and Technology 65(4), 670–685 (2014)
4. Brath, R., Banissi, E.: The design space of typeface. Visualization and Computer Graphics, IEEE Transactions on (2013)
5. Chan, M., Podlaseck, M.: A personalized navigation tool for online listening and free browsing: The glass engine. In: Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2002. pp. 263–264 (2002)
6. Dobson, T., Ruecker, S., Gabriele, S., Sinclair, S.: The mandala browser (2005), http://mandala.humviz.org/help/
7. Ellingham, H.: A chart illustrating some of the relations between the branches of natural science and technology. http://bit.ly/1hlJ6VK (1948), accessed 11/02/2013
8. Hearst, M.: Search User Interfaces. Cambridge University Press (2009)
9. Heer, J., Bostock, M.: Crowdsourcing graphical perception: Using mechanical turk to assess visualization design. In: ACM Human Factors in Computing Systems (CHI). pp. 203–212 (2010)
10. Heim, S.: The Resonant Interface: HCI Foundations for Interaction Design. Pearson Education Inc. (2008)
11. Hodges, E.R.S.: The Guild Handbook of Scientific Illustration. John Wiley (2003)
12. Kahn, P., Lenk, K.: Principles of typography for user interface design. Interactions-New York pp. 15–29 (1998)
13. Lupton, E.: Designing Type. Yale (2004)
14. Mayr, P., Mutschke, P., Petras, V.: Reducing semantic complexity in distributed digital libraries: Treatment of term vagueness and document re-ranking. Library Review 57(3), 213–224 (2008)
15. Nacenta, M., Hinrichs, U., Carpendale, S.: Fatfonts: combining the symbolic and visual aspects of numbers. In: Proceedings of the International Working Conference on Advanced Visual Interfaces. pp. 407–414. ACM (2012)
16. Paley, B.: Map of science. http://wbpaley.com/brad/mapOfScience/index.html (accessed: 11/12/2013)
17. Robinson, A., Morrison, J., Muehrcke, P., Kimerling, A., Guptill, S.: Elements of Cartography. John Wiley & Sons, New York, NY (1995)
18. Skupin, A.: The world of geography: Visualizing a knowledge domain with cartographic means. Proceedings of the National Academy of Sciences of the United States of America 101(Suppl 1), 5274–5278 (2004)
19. Skupin, A., Biberstine, J.R., Börner, K.: Visualizing the topical structure of the medical sciences: a self-organizing map approach. PloS one 8(3), e58779 (2013)
20. Small, D.: Navigating large bodies of text. IBM Systems Journal 35, 515–525 (1996), http://diglib.eg.org/EG/DL/conf/EG2013/stars/039-063.pdf
21. Squire, V., Willberg, H.P., Forssman, F.: Getting it Right with Type. Laurence King Publishing (2006)
22. Teague, J.: Fluid Web Typography. New Riders, Berkeley, CA (2009)
23. Weskamp, M.: Projects: Newsmap. http://marumushi.com/projects/newsmap (2004), accessed: 03/03/2014