Int J Digit Libr
DOI 10.1007/s00799-016-0168-4

Font attributes enrich knowledge maps and information retrieval

Skim formatting, proportional encoding, text stem and leaf plots, and multi-attribute labels

Richard Brath[1] · **Ebad Banissi**[1]

Received: 29 April 2015 / Revised: 27 December 2015 / Accepted: 11 January 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract Typography is overlooked in knowledge maps (KM) and information retrieval (IR), and some deficiencies in these systems can potentially be improved by encoding information into font attributes. A review of font use across domains is used to itemize font attributes and information visualization theory is used to characterize each attribute. Tasks associated with KM and IR, such as skimming, opinion analysis, character analysis, topic modelling and sentiment analysis can be aided through the use of novel representations using font attributes such as skim formatting, proportional encoding, textual stem and leaf plots and multi-attribute labels.

Keywords Font attributes · Text visualization · Information density · Alphanumeric glyphs · Quantitative typography

1 Introduction

The potential of typography to convey information is often overlooked in many *knowledge maps* (KM) and *information retrieval* (IR) displays. But fonts have properties, such as bold and italic, which can make some text visually pre-attentive and otherwise add more information into textual displays.

In a review of KM and IR visualizations [10], only a quarter use font attributes, suggesting a missed opportunity. To effectively leverage font attributes in KM and IR interfaces it is necessary to:

1. *Assess existing KM and IR interfaces* including deficiencies which could be improved through the use of font attributes.
2. *Itemize font attributes* based on a review of current usage to encode data in various domains.
3. *Characterize font attribute capabilities* based on information visualization (infovis) theory.
4. *Identify KM and IR applications*, including:
    - *skimming texts* such as lead paragraphs;
    - encoding quantitative data in search lists using the novel technique of *proportional encoding*;
    - profiling metadata associated with topics, entities and facets via extensions to *stem and leaf plots*; and
    - *multivariate labels* for data-dense knowledge maps.

The contribution of this article includes an overview of font attributes and their application to tasks associated with KM and IR, including skimming, opinion analysis, character analysis, topic modelling and sentiment analysis.

2 Existing KM and IR interfaces

Information retrieval is a complex activity comprising of search, browsing, filtering, formulating queries, identifying entities, skimming content, comparing results, and so on [48]. Therefore, systems for information retrieval may contain a variety of components for interacting with results to support these different tasks (Fig. 1), such as:

(a) *List* of resulting documents (e.g. web search interfaces).
(b) *Detailed text* for one or a few matching items, such as lead sentence, lead paragraph, abstract, keyword in context (KWIC), or an integrated document viewer (e.g. TextArc [46], TRIST[49]).

✉ Richard Brath
brathr@lsbu.ac.uk

[1] London South Bank University, London, UK

Published online: 08 February 2016

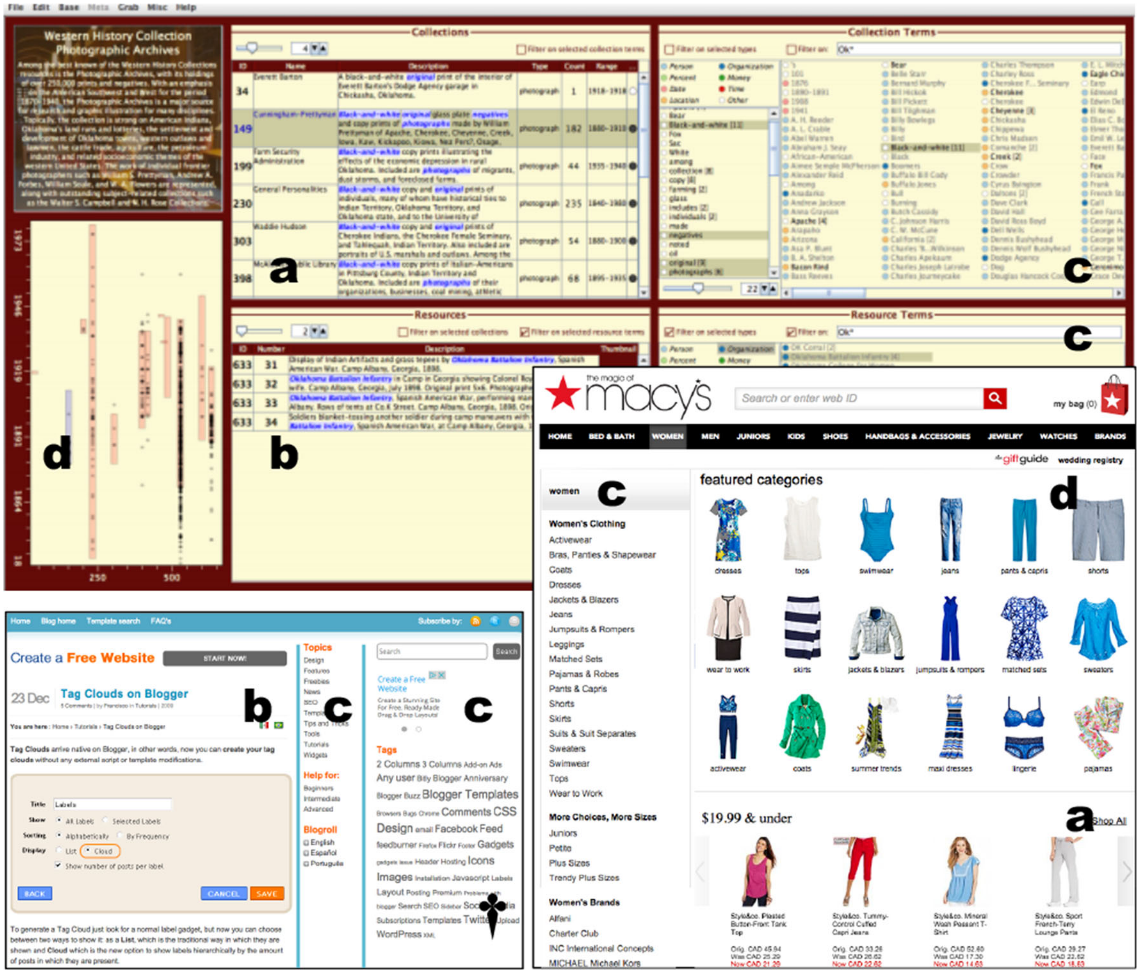


Fig. 1 Multi-component IR interfaces: improvise [63], a blog post with topics and tags (http://blog.btemplates.com/tag-clouds-on-blogger), http://macys.com department browsing

(c) *Facet lists*, such as topics, publication dates, named entities, tags, etc. used for profiling results and query refinement (e.g. many shopping web sites, TRIST, UO Western History Collections Improvise interface [63]).

(d) *Corpus overview* such as a top level categories (e.g. http://macys.com) or knowledge map(s).

Knowledge maps are multi-variate visualizations that provide an overview of large-scale knowledge spaces. Hundreds to billions of items are spatially located and visual attributes such as colour and size depict data. They may be large-scale static images or interactive visualizations enabling exploration and drill down (Fig. 2).

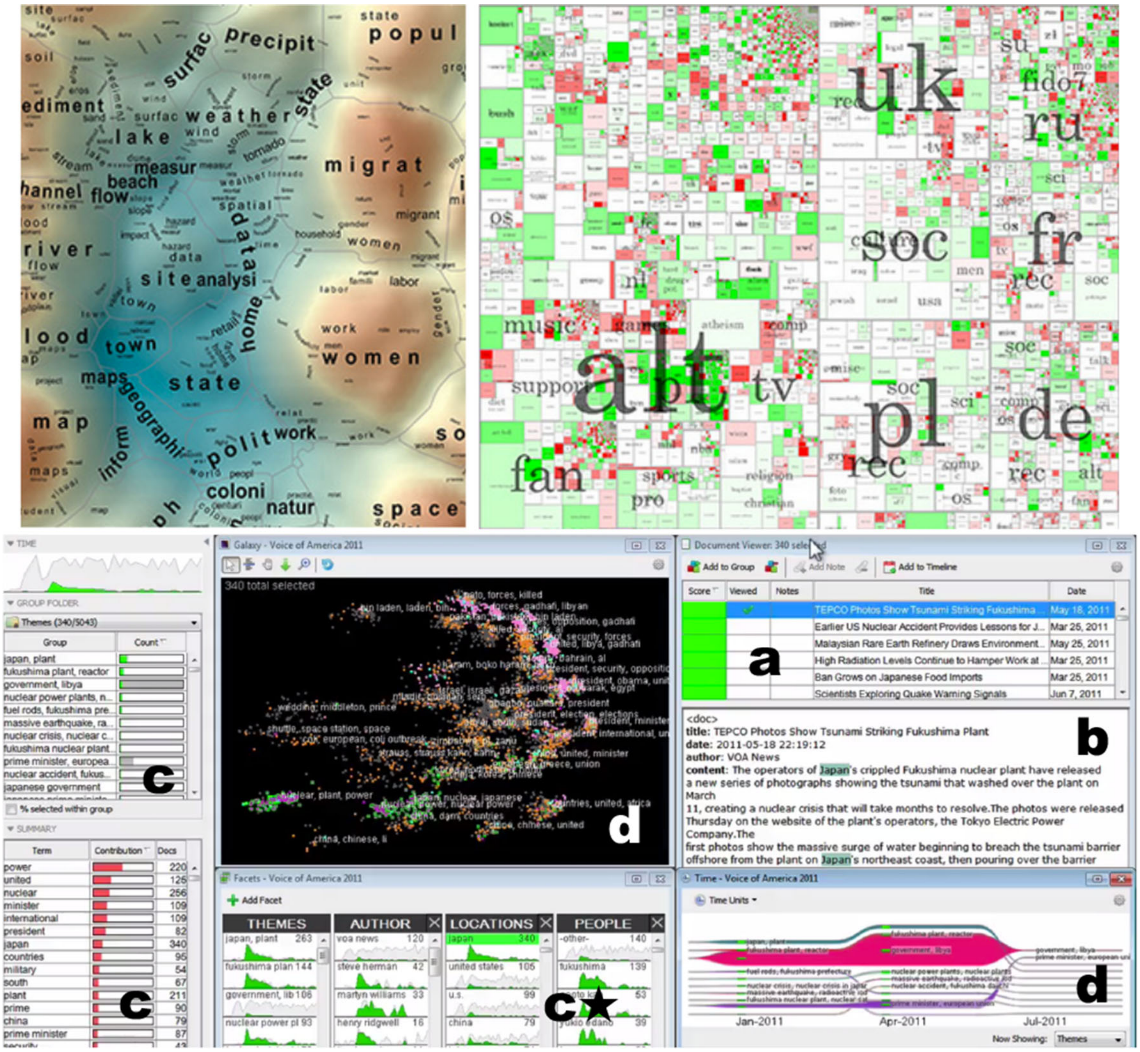
Information visualizations may be used in these IR components with techniques such as highlighting query terms, visualizing elements associated with results such as facets, or relationships between words and documents [27,61]. Knowledge maps are one type of visualizations that may be a component in an IR system, e.g. InSpire (e.g. Fig. 2d [66]).

Other IR components may also be presented as visualizations, e.g. facet lists as tag clouds (Fig. 1†) or as time series charts (Fig. 2∗); the document overview may be a visualization [46] (Fig. 3, left); or the list results presented as icons [49] (Fig. 3, right).

Various benefits can be achieved with KM and IR visualizations. For example, TRIST enabled junior analysts to review twice as many documents in half the time compared to traditional internet search (e.g. Google). KMs can reveal adjacencies, patterns of collaboration, social networks, gaps, frontiers and dynamics [7]. The Bohemian Bookshelf [58] facilitates serendipitous discoveries in digital libraries.

2.1 Problems with KM and IR visualizations

In visualization, data are transformed into visual attributes such as size, colour, bold or italic. In some KM and IR visualizations, these encodings are problematic. For example:

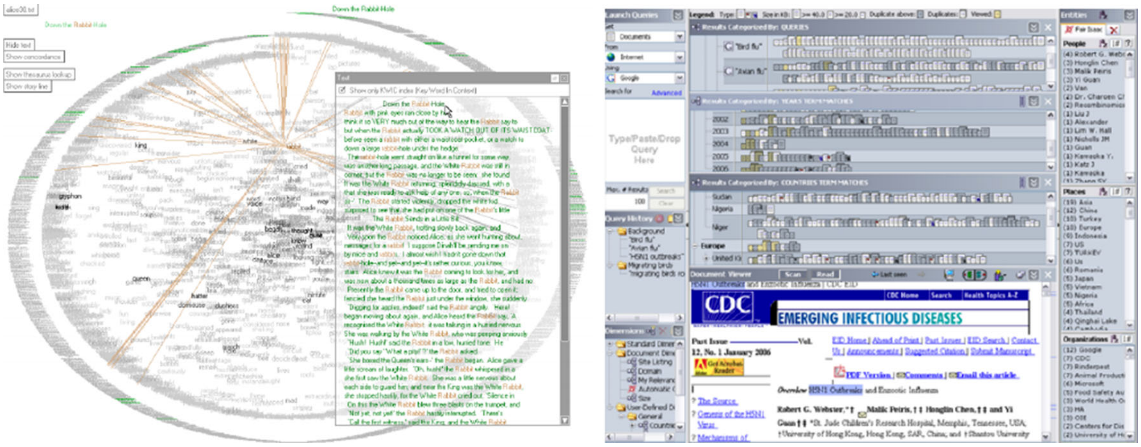


Title	Date
TEPCO Photos Show Tsunami Striking Fukushima...	May 18, 2011
Earlier US Nuclear Accident Provides Lessons for J...	Mar 25, 2011
Malaysian Rare Earth Refinery Draws Environment...	May 25, 2011
High Radiation Levels Continue to Hamper Work at...	Mar 25, 2011
Ban Grows on Japanese Food Imports	Mar 25, 2011
Scientists Exploring Quake Warning Signals	Jun 7, 2011

<doc>
title: TEPCO Photos Show Tsunami Striking Fukushima Plant
date: 2011-05-18 22:19:12
author: VOA News
content: The operators of Japan's crippled Fukushima nuclear plant have released a new series of photographs showing the tsunami that washed over the plant on March 11, creating a nuclear crisis that will take months to resolve.The photos were released Thursday on the website of the plant's operators, the Tokyo Electric Power Company.The first photos show the massive surge of water beginning to breach the tsunami barrier offshore from the plant on Japan's northeast coast, then pouring over the barrier

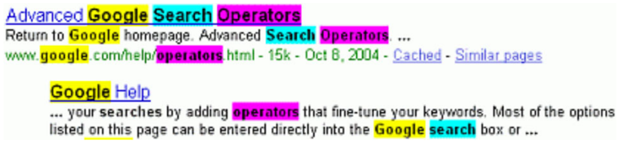
Term	Docs
power	220
united	125
nuclear	256
minister	109
international	109
president	82
japan	340
countries	95
military	54
south	67
plant	211
prime	90
china	79
prime minister	87

Fig. 2 *Example knowledge maps*: Skupin's self-organizing map [52]; Usenet treemap [25] and InSpire [66] with galaxy map (*top centre d*) and flow map of themes (*bottom right d*)



Down the Rabbit Hole
Hide text
Show concordance
Show thesaurus lookup
Show stop list
CDC
EMERGING INFECTIOUS DISEASES
Type/Paste/Drop Query Here
Results Categorized By: QUERIES
Results Categorized By: YEARS
Results Categorized By: COUNTRIES
Document Viewer
Entities
People
Places
Organizations

Fig. 3 Visualization-focused IR: TextArc [46] with the document view as a visualization; TRIST [49] with the search result list as icons



Advanced Google Search Operators
Return to Google homepage. Advanced Search Operators. ...
www.google.com/help/operators.html - 15k - Oct 8, 2004 - Cached - Similar pages

Google Help
... your searches by adding operators that fine-tune your keywords. Most of the options listed on this page can be entered directly into the Google search box or ...

Fig. 4 Search term *highlights* can reduce text legibility, e.g. *blue on purple* (colour figure online)

Query term highlighting Text legibility depends on the contrast between the characters and background [50]. When multiple data attributes are encoded with colour there is the potential for bad combinations. For example, Fig. 4 is an early version of Google toolbar (2009) using foreground text colour to differentiate links, text fragments and metadata; and background color for term highlights. This results in a difficulty to read blue on purple combination. Poor design choices can result in illegible text, but the text is a critical element in an IR solution.

Tag clouds (Fig. 1†) and *Treemaps* (Fig. 2, top right [25]) are popular visualization techniques that vary word size in proportion to data (e.g. tag frequency, size of topic area). However, tag clouds and treemaps are spatially inefficient: some text is larger—taking up more space, and other text become smaller, some of which may become too small to read. Jakob Nielsen is critical:

> [Compared to tag clouds] a one paragraph summary would probably be more enlightening, faster to scan, take up less screen space allowing for more items to be summarized on any given page [45].

2.2 Why font attributes

Font attributes such as bold, italics and case have existed for hundreds of years, but their use in KM, IR and infovis is currently low:

- http://Scimaps.org is a repository of knowledge maps—only 28 % of these visualizations use font-specific attributes, and in most cases they only differentiate between compositional elements (e.g. tick labels, item labels, legend).
- Similarly, in Hearst's examples of IR infovis, only 29 % use font-specific attributes [27].
- The lists of visual attributes compiled by infovis researchers relegate text to a single entry [3,17,19,37,40,42,62,65] with no indication of font attributes, although Börner recently included some [6].

Several current factors suggest font attributes may provide unique opportunities to encode information:

Improved technology Infovis, similar to the Internet, evolved in an era when screen displays were 72–96 pixels per inch (PPI). Subtle variations in fonts are not possible at small sizes and low resolutions, and user interface guidelines in the 1990s recommended against italics, weights, small caps, and so on [32]. These constraints are not an issue on current high-resolution devices (e.g. 150–300 PPI) which can better represent the subtle details of typography (Fig. 5). New font rendering technology (e.g. Windows ClearType, Mac Quartz) improves the display of screen fonts. A wide range of typefaces have been designed for the screen (e.g. 500+ free web fonts on http://google.com/fonts) and better control over markup formats are available (e.g. CSS, HTML5, SVG). Web interface guidelines have evolved into recommendations to more broadly use font attributes in user interfaces and web including mixing type families, weights, italics and capitalization (e.g. [57]).

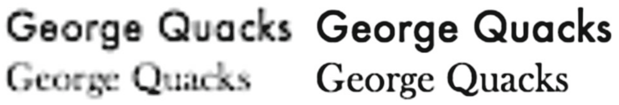


Fig. 5 Enlarged nine-point Futura and Baskerville fonts on an 86-PPI screen (*left*) vs the same on an iPad3 with 264 PPI (*right*). With more than nine times resolution, the fine details of text are legible

Unique capabilities of text vs. other glyphs The use of glyphs is becoming popular in infovis [5,38]. While glyphs may work well for concrete nouns that can be used across languages (e.g. http://thenounproject.com), text glyphs are a unique type of glyph that offer capabilities beyond other glyphs (such as pictographs) including:

- *Font attributes* such as bold, italic and case are not available to other glyphs.
- *Literal encoding* with text is unambiguous compared to pictographic glyphs [3].
- *Letter order* with text can be used to created ordered representations (i.e. alphabetic order, such as an index).
- *Glyph design* can be difficult, particularly when there are many categories to encode. Corresponding text does not require a design task.

Extensions to text-based interfaces One problem with tag clouds is "many users do not know how to use them" [45], even though they show similar information to most facet lists (i.e. a set of terms for refinement plus to size indicate term frequency). In general, new visualization techniques need to be learned as they may look and interact very differently than familiar techniques. Applying font attributes to existing textual interfaces has the potential to reduce learning as the representation is simply an extension to an already familiar interface.

Given the problems with some types of visualizations, plus with the improved capabilities of visual displays, the renewed interest in glyphs and extensibility of existing textual interfaces, it is appropriate to further investigate the opportunities that font attributes may provide to all the components of an information retrieval system.

3 Font attributes across domains

While font attributes may not have been formally investigated by infovis researchers, font attributes are sometimes used in infovis and have a long history in other domains.

Knowledge maps frequently use text labels: *places and spaces* (http://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 colour (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 [24,52].

Information retrieval infovis has similar usage. Of 45 examples in Hearst's infovis chapters [27], half use traditional visual attributes of size and/or colour. There are 13 examples using one type-specific attribute, either bold, caps, or font family. In most cases these are used simply: to either highlight a search term or differentiate between types of data, e.g. category title vs. instance; axis title vs. tick label. Similarly, van Hoek's IR infovis survey [61] has only a passing mention of boldface.

Text visualization systems, such as Termite or Tiara [18,35], often utilize well-known infovis techniques (e.g. adjacency matrices, stream graphs) and may utilize the visual attributes of hue and size but ignore the opportunity of font attributes. Early innovators include Baecker and 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 (e.g. WebStorm). Weaver's interactive markdown [63] exposes many HTML tags to format textual metadata including bullets, lines, bold, italics, super/subscript and so on.

Other infovis also use font attributes, e.g. italics [47], uppercase [16] or bold [23]. Fat fonts [43] is a specialized font that varies as 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, colour and font family, e.g. [54]. Typographic maps [1] use only type to create geographic maps.

Typography and *cartography* have centuries of history with innovative font encoding of information. For example, Cyclopaedia from 1728 (Fig. 6, top) provides an early knowledge map as a hierarchy with italics denoting broad topics, small caps for specific fields, roman for explanatory text and superscript for chapter numbers. The genealogy chart from 1820 (Fig. 6, bottom) uses bold uppercase for major branches, plain uppercase for small branches, mixed case for direct descendants, sovereign rulers in small caps and spouses in italics.

There are many techniques for creating emphasis and differentiation with font, with various guidelines and conventions (e.g. typographic [36,55], cartographic [30,51], and user interface design [29,32,57]). Some cartographic texts explicitly indicate font attributes as a means for encoding data (e.g. [34]) and enumerate conventions for use, such as italics for water features (e.g. [22]). Figure 7 uses font family, case, spacing, multiple underlines to encode data.

Text-based search results in information retrieval and navigation interfaces differentiate metadata associated with a document using type size, colour, underlines (e.g. links) or font family (e.g. titles). Bold or colour is frequently used to highlight search terms; see Hearst [27] or popular search interfaces, e.g. Google, Yahoo, Bing, Ebay, Amazon, NYTimes, LinkedIn, etc.

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

After reviewing these above domains, a list of font-specific properties (not including generic colour and size visual attributes) can be compiled.

4 Font attribute capabilities

All font attributes can be used simply to encode differentiation (i.e. a binary encoding) and in many use cases, (e.g. magazines, web sites), a small mix of 2–3 typefaces, size, bold, case, etc. may be used to create an information hierarchy (e.g. title, headings, body, caption, labels, ticks, etc.) differentiating among elements on the screen and aiding the viewer to orient themselves in relation to the content [36]. However, many of the font attributes can go beyond differentiation and be used to encode more levels of either categoric or quantitative data. Furthermore, font attributes have different levels of effectiveness; for example, **bold** or *italic* may visually be detected almost immediately but the difference between sans and serif may not be noticeable if the two typefaces are very similar.

- *Weight* (bold) can range up to nine levels of weight. Weight can be used to encode quantitative data, with heavier text indicative of greater magnitude [21,34].

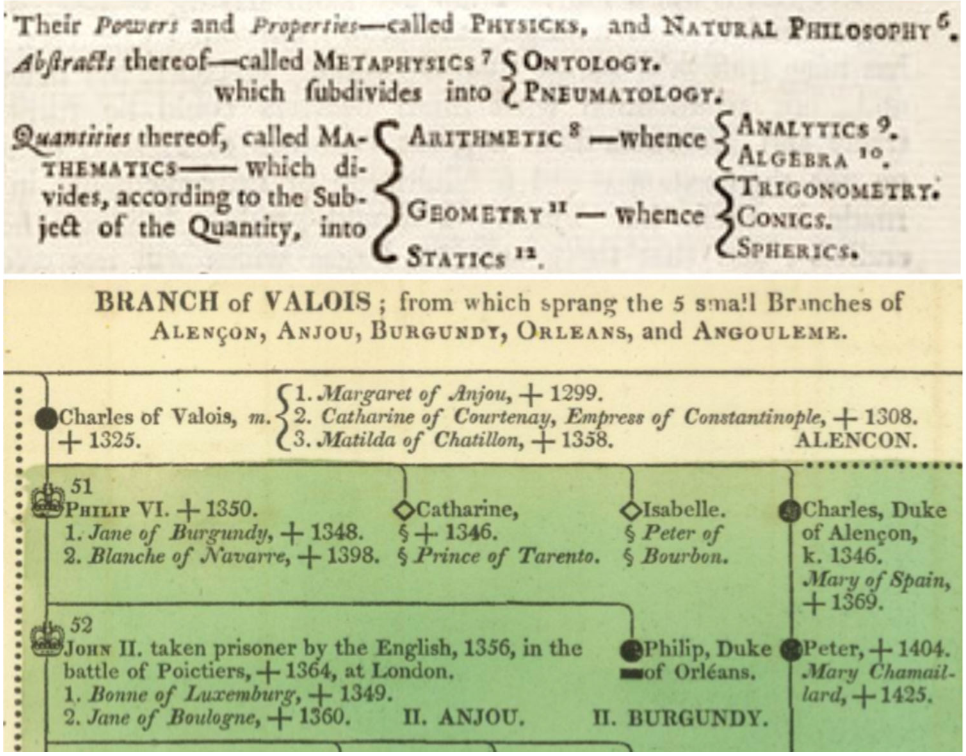

Their *Powers* and *Properties*—called PHYSICKS, and NATURAL PHILOSOPHY 6.
Abſtracts thereof—called METAPHYSICS 7 ONTOLOGY.
which ſubdivides into PNEUMATOLOGY.
Quantities thereof, called MA-THEMATICS—— which di-vides, according to the Sub-ject of the Quantity, into
ARITHMETIC 8 —whence
ANALYTICS 9.
ALGEBRA 10.
GEOMETRY 11 — whence
TRIGONOMETRY.
CONICS.
SPHERICS.
STATICS 12.
BRANCH of VALOIS ; from which sprang the 5 small Branches of
ALENÇON, ANJOU, BURGUNDY, ORLEANS, and ANGOULEME.
Charles of Valois, *m.* + 1325.
1. *Margaret of Anjou*, + 1299.
2. *Catharine of Courtenay, Empress of Constantinople*, + 1308.
3. *Matilda of Chatillon*, + 1358.
ALENCON.
51
PHILIP VI. + 1350.
1. *Jane of Burgundy*, + 1348.
2. *Blanche of Navarre*, + 1398.
Catharine,
§ + 1346.
§ *Prince of Tarento.*
Isabelle.
§ *Peter of*
§ *Bourbon.*
Charles, Duke of Alençon, k. 1346.
Mary of Spain, + 1369.
52
JOHN II. taken prisoner by the English, 1356, in the battle of Poictiers, + 1364, at London.
1. *Bonne of Luxemburg*, + 1349.
2. *Jane of Boulogne*, + 1360.
II. ANJOU.
Philip, Duke of Orléans.
II. BURGUNDY.
Peter, + 1404.
Mary Chamaillard, + 1425.

Fig. 6 *Upper* text tree from Cyclopaedia [15] uses *italics, small caps, roman text* and *superscripts* to differentiate elements. *Lower* genealogy chart [14] uses *bold, italics, case* and *small caps* in this text graph

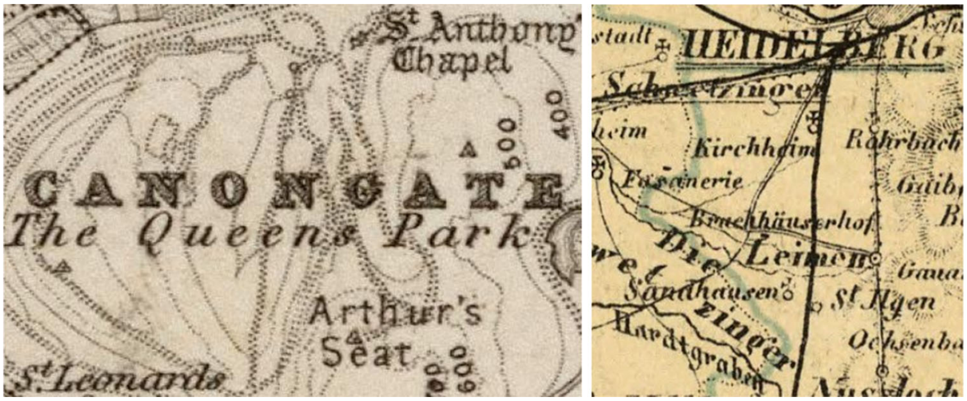


Fig. 7 The *left map* [31] uses four different font families (*blackletter, outlined slab-serif, italic serif* and *sans-serif*). The *right map* [33] uses *single* and *multiple underlines* (*top*) and both *forward* and *reverse* sloping *italics* (*bottom left*)

- *Italic* or *oblique* are both sloped fonts but italics have different letterforms. Sloped fonts vary in slope angle including instances of reverse italics and even vertical italics [36]. Slope can be used to encode a diverging scale, ranging from reverse, to vertical, to forward slope.
- CASE includes UPPER, lower, Mixed and SMALL CAPS. Uppercase is designed to standout from lowercase while small caps blend in. Case is sometimes used in cartography for ordered data, e.g. states in uppercase and counties in lowercase [34].
- Typeface indicates font family, e.g. sans, blackletter, *script*, `source`, MATHBOLD, etc. Typeface is best used to symbolize categoric information. Many typographic and cartographic references suggest using no more than two typefaces although there are examples with more (e.g. Fig. 7). Historically there was bias to using simpler sans serif fonts on screen due to limited resolutions but more detailed fonts are now used in web design, e.g. [8,41].
- Underline can be distracting. Many typographers and cartographers recommend against underline or subtle variants. Underlines on SMALLCAPS or UPPER case do not interfere with descenders. Underlines can express ordered data (e.g. dot, dash, single, double as in Fig. 7).
- Width of a string is used in cartography to indicate the range of an area feature, and thus has possibilities for encoding quantities. Different ways to set width include:
  - Ultra-condensed Condensed Normal Extended are specifically designed widths of a given typeface for applications such as tight spaces. Few fonts are available in a range of widths.

Table 1 Font attributes relation to visual channels and encodings

	Visual Channel							
Font Attribute	Position	Length/Size	Intensity	Orientation	Shape	Containment	Preattentive Potential†	**Best for encoding:** Q: quantitative O: ordered C: categoric G: grouping/relationship L: literal
Font weight		⬩	◆				**HP**	**Q** (2-9 levels)
Oblique/Italic				◆			**HP**	**C**, **Q** using slope angle
Case		⬩			◆		**P**	**C**, possible O (2-3 levels)
Typeface					◆		**P**	**C** (2-6 levels)
Underline	◆	⬩					**HP**	**C**, **O**, **Q** (using length)
Condensed		◆	⬩				**HP**	**Q**, **O** (2-4 levels)
Squished		◆	⬩				**HP**	**Q**
Spacing		◆	⬩				**HP**	**Q**
Super/subscript	◆	⬩					**HP**	**C** (2 levels)
Delimiters						◆	D	**G**
Text Glyph					◆		D	**L**, **O**
Symbols					◆		D	**C**

◆ / ⬩ indicates primary / secondary visual channel for font attribute
† HP: Highly probable, P: probable, D: doubtful

- Squished/Stretched are horizontally scaled fonts. Typographers recommend against these distortions.
- Spacing includes adjusting the space between letters (tracking) and between lines (leading).

- Superscript and subscript encode via size and position relative to adjacent text. They can be used to encode a high number of categories (e.g. Fig. 10 in [56]).
- 'Paired delimiters' evoke enclosure by pairing the same (or mirrored) shapes, e.g. ([ ],{ },“ ”,* *, etc.)
- 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. ∞, $\forall$, $\flat$) are not orderable. While other font attributes may visually be perceived without active attention for quick perception, words and phrases must be actively read which is slower.

These typographic attributes have different levels of effectiveness, just as other visual channels (e.g. size, hue, orientation, shape). Visual channels have been researched in detail for their ability to be detected rapidly (i.e. pre-attentiveness e.g. [26,67]); and accuracy of perceiving magnitude (e.g. [20,28]). Font-specific attributes can be mapped to these well-known visual channels [10] as summarized in Table 1. The diamonds indicate the primary visual channel that the font attribute uses, with a small diamond indicating a secondary visual channel that may be used. The pre-attentive potential column indicates whether the font attribute is likely to be perceived pre-attentively (i.e. perceived almost immediately, ahead of active focused attention) based on the properties of the underlying visual channel [67].

Visualization researchers have created rankings of visual channels (e.g. [3,19,37,62,65]), although a definitive ranking for different types of encodings and number of uniquely perceivable levels does not exist. These rankings can be used as a heuristic to determine which kind of data is best encoded with a particular font attribute (last column on Table 1). For example, font weight, which utilizes visual channels of size (i.e. line width) and intensity (i.e. amount of ink per glyph) will rank higher for effectiveness of ordered or quantitative data encodings than font family, which utilizes the visual channel of shape. Font family, however, will be effective for encoding categoric data.

Encoding data using typography utilize the inherent properties of a well-designed font. Consistency in stroke angle, weight, heights, etc. establishes a consistency across all glyphs within the same font family: disrupting this expectation causes a visual anomaly that visually *stands out* such as a shift to italic. Furthermore, a well-designed font has tweaked letterform shapes, spacing and weight such that if one squints at a page of text, the paragraphs appear as gray blocks. Typographers refer to this as *colour*, not to be confused with *hue* as discussed in infovis. As a result, a viewer does not perceive any particular letters standing out in a field of type all using the same typeface. The even pattern and density of type can then have information added using font attributes to make desired subsets of text standout. Additionally, these font attributes have been designed to be combined together while retaining legibility, such as underline + ***bold*** + ***italic*** + ***ALL CAPS***.

5 Fonts for knowledge maps and information retrieval

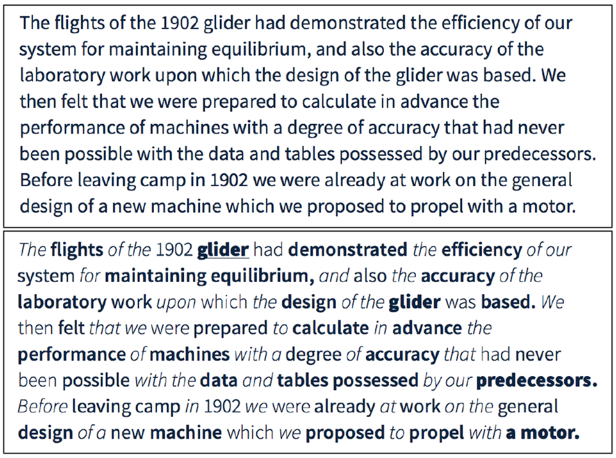
Font attributes can be used to encode additional information at different levels of use in knowledge mapping and information retrieval, ranging from low-level document views, to search lists, to the macro-level overviews. A few examples:

5.1 Font visualization on texts

In some search results, previews of full sentences, abstracts or lead paragraphs are presented (e.g. BioText Search Engine, Wikipedia's *Today's Featured Article* archive). Font attributes can be used to adjust words to facilitate rapid comprehension without changing layout.

5.1.1 Skim formatting of previews

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,



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.

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. 8 Lead paragraph of *How We Made the First Flight* by Orville Wright *before* and *after* formatting for skimming using font weight and *italic with search term underlined*

unusual words, enumerations, etc. Word frequency analysis can be used such that the least common words in the corpus have the heaviest weight while the most frequent words have the lightest weight (Fig. 8).

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 visually pop-out. They are unambiguous words and could be terms for query refinement. Visual weighting of proper nouns, enumerations and unusual words facilitates fact-finding tasks.

Evenly weighted text (for reading) and skim-formatted text (for skimming) can be toggled between. Note that the layout of the words remain in similar places before and after formatting. Text formatted for reading has a consistent medium weight (i.e. referred to as *book weight* in typography). The skim formatting adjusts some words to a heavy weight (which is slightly wider than book text) and some words to a lighter weight (which is slightly narrower than book text)—thereby retaining words in similar positions between the two views. This consistency of position between modes allows the viewer to maintain reading position and use spatial memory to move around the text even when toggling between modes.

Instead of weighting words based on English language word frequency, specific domain vocabularies could be used [39]; or, more advanced text analytics could be used to identify salient entities, phrases and sentences, which in turn are weighted more heavily in relation to other phrases and sentences with lower scores. Other font attributes can also be applied to the text. For example, in Fig. 8 underline indicates search terms and less important parts of speech (e.g. articles, pronouns, etc.) are italicized to create greater differentiation to enhance figure-ground separation between the heavy-weight words and background.

Expert feedback has been collected via interviews with information visualization researchers and prospective users in news organizations. There appears to be a strong appeal with responses ranging from visceral to observational:

> "Can you install this on my iPad now?"
> "I can see using this immediately in my own visualization research."
> "This is similar to how we used multiple underlines in our paper textbooks in college."
> "The ability to toggle is key: people who consume news all day will need to move back and forth between reading and skimming."
> "The technique can work well by aiding recognition of keywords instead of relying on searching (recall)."
> "Perhaps the same technique could be used to make the words pop-out that make the text more memorable, the way that Kennedy or Martin Luther King used spoken emphasis on words."

5.2 Font visualization on lists

Search results generate lists: lists of documents, lists of facets, lists of topics, lists of words, and so on. Visualization techniques that enhance text lists could improve tasks associated with scanning query results, comparing alternative queries, filtering facets, assessing topics and comprehending analytics such as opinion analysis.

5.2.1 Query result lists and proportional encoding

Query results may include quantitative data, e.g. relevance, score, readership, number of citations, etc. Newsmap.jp displays news headlines in a treemap, with headline size indicating readership and background brightness indicating recency (Fig. 9).

Instead, using a fixed size text enables a greater number of legible headlines to be displayed and then font weight can be used to encode readership, either by setting the weight per headline (Fig. 10, top), or *proportionally encoding* readership by setting bold to proportionally correspond to the magnitude (Fig. 10, bottom).

Any visualization technique has some degree of lossiness as data are 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 96-dpi screen.

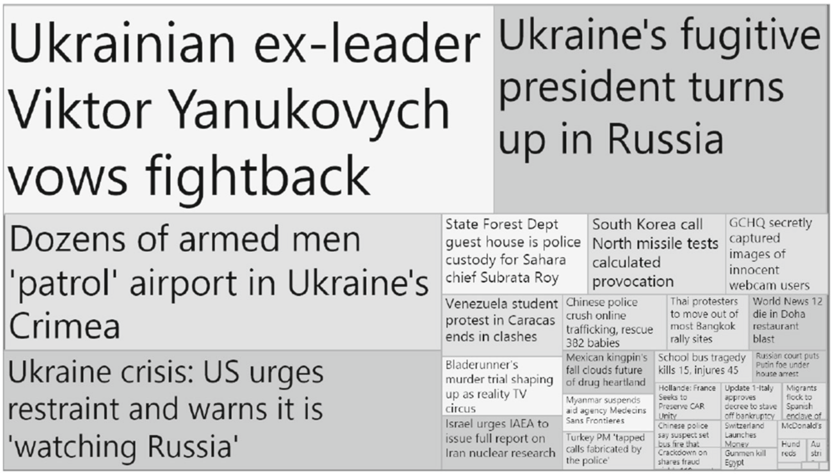
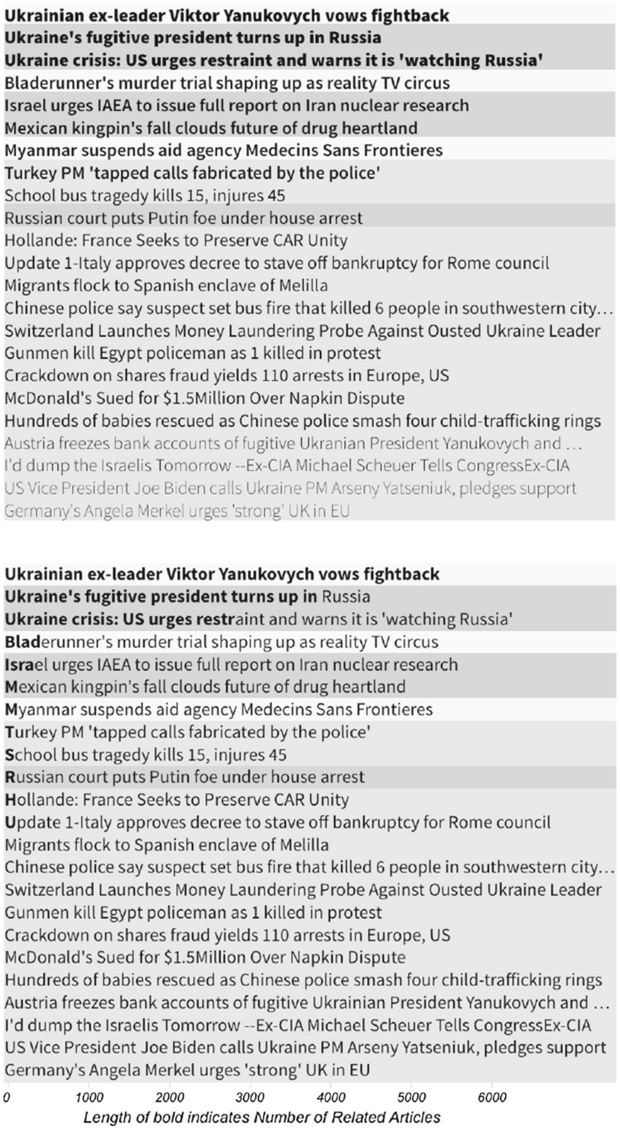


Fig. 9 *Top* portion of Newmap [64] from 02/28/2014. *Size* represents readership


Ukrainian ex-leader Viktor Yanukovych vows fightback
Ukraine's fugitive president turns up in Russia
Ukraine crisis: US urges restraint and warns it is 'watching Russia'
Bladerunner's murder trial shaping up as reality TV circus
Israel urges IAEA to issue full report on Iran nuclear research
Mexican kingpin's fall clouds future of drug heartland
Myanmar suspends aid agency Medecins Sans Frontieres
Turkey PM 'tapped calls fabricated by the police'
School bus tragedy kills 15, injures 45
Russian court puts Putin foe under house arrest
Hollande: France Seeks to Preserve CAR Unity
Update 1-Italy approves decree to stave off bankruptcy for Rome council
Migrants flock to Spanish enclave of Melilla
Chinese police say suspect set bus fire that killed 6 people in southwestern city...
Switzerland Launches Money Laundering Probe Against Ousted Ukraine Leader
Gunmen kill Egypt policeman as 1 killed in protest
Crackdown on shares fraud yields 110 arrests in Europe, US
McDonald's Sued for $1.5Million Over Napkin Dispute
Hundreds of babies rescued as Chinese police smash four child-trafficking rings
Austria freezes bank accounts of fugitive Ukranian President Yanukovych and ...
I'd dump the Israelis Tomorrow --Ex-CIA Michael Scheuer Tells CongressEx-CIA
US Vice President Joe Biden calls Ukraine PM Arseny Yatseniuk, pledges support
Germany's Angela Merkel urges 'strong' UK in EU

Ukrainian ex-leader Viktor Yanukovych vows fightback
Ukraine's fugitive president turns up in Russia
Ukraine crisis: US urges restraint and warns it is 'watching Russia'
Bladerunner's murder trial shaping up as reality TV circus
Israel urges IAEA to issue full report on Iran nuclear research
Mexican kingpin's fall clouds future of drug heartland
Myanmar suspends aid agency Medecins Sans Frontieres
Turkey PM 'tapped calls fabricated by the police'
School bus tragedy kills 15, injures 45
Russian court puts Putin foe under house arrest
Hollande: France Seeks to Preserve CAR Unity
Update 1-Italy approves decree to stave off bankruptcy for Rome council
Migrants flock to Spanish enclave of Melilla
Chinese police say suspect set bus fire that killed 6 people in southwestern city...
Switzerland Launches Money Laundering Probe Against Ousted Ukraine Leader
Gunmen kill Egypt policeman as 1 killed in protest
Crackdown on shares fraud yields 110 arrests in Europe, US
McDonald's Sued for $1.5Million Over Napkin Dispute
Hundreds of babies rescued as Chinese police smash four child-trafficking rings
Austria freezes bank accounts of fugitive Ukrainian President Yanukovych and ...
I'd dump the Israelis Tomorrow --Ex-CIA Michael Scheuer Tells CongressEx-CIA
US Vice President Joe Biden calls Ukraine PM Arseny Yatseniuk, pledges support
Germany's Angela Merkel urges 'strong' UK in EU

Fig. 10 *Top* same headlines as Fig. 9 with font weight indicating readership. *Bottom* same with proportion of *bold* indicating readership

Table 2 Relative information density of headlines vs. treemap

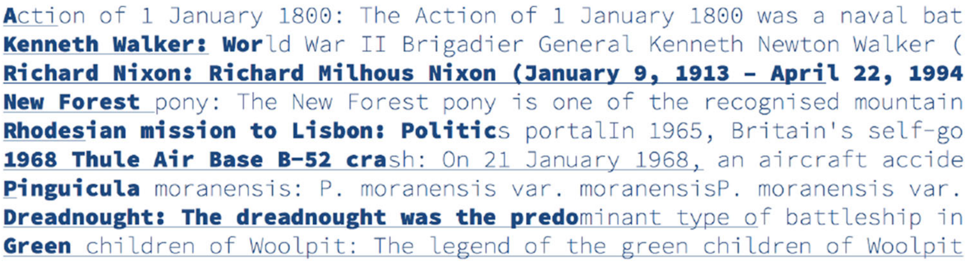
Variant	Normal	Dense	Sparse
Font weight	1.42×	0.67×	0.86×
Proportion	2.68×	2.22×	2.07×

2. *Number of uniquely perceivable sizes* can be estimated using prior experimental data [28], 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 perceptible weights.
3. *Area of the overall plot.*

Information density is a function of the measures (#readable x #sizes / area). Density was measured across three variants and repeated for different aspect ratios with different numbers of items (i.e. in addition to the example shown, a sparse example started with a treemap where the area was large and almost all the initial headlines were readable; and a dense example started with a treemap where the area was small and most of the initial headlines were too small to read). Information density for the proportional encoding consistently outperformed the treemap by a factor of 2 while the font weight encoding could underperform as shown in Table 2 (see [11] for details).

These three different encodings were presented to three different groups of infovis researchers, each with more than ten 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, understanding and perceived ease of use for general population.

While the technique scores well, interviews indicate some hesitation for some participants: they fully understand how the technique addresses the shortcomings of a treemap, but feel that a treemap offers some other benefits including



Action of 1 January 1800: The Action of 1 January 1800 was a naval bat
Kenneth Walker: World War II Brigadier General Kenneth Newton Walker (
Richard Nixon: Richard Milhous Nixon (January 9, 1913 – April 22, 1994
New Forest pony: The New Forest pony is one of the recognised mountain
Rhodesian mission to Lisbon: Politics portalIn 1965, Britain's self-go
1968 Thule Air Base B-52 crash: On 21 January 1968, an aircraft accide
Pinguicula moranensis: P. moranensis var. moranensisP. moranensis var.
Dreadnought: The dreadnought was the predominant type of battleship in
Green children of Woolpit: The legend of the green children of Woolpit

Fig. 11 Wikipedia articles with *bold length* indicating number of authors and *underline length* indicating number of readers

(a) ease of understanding—no axis, explanation or note is required; and (b) aesthetic appeal of variously sized boxes offers visual interest that proportional encoding does not offer.

The approach can be extended to convey multiple quantitative data attributes using separable font attributes, such as font weight and underline. Figure 11 shows a list of Wikipedia articles, based on a *Featured Articles* search. Bold length indicates the number of authors and underline length indicates the number of readers. This combination of font attributes visually shows that authorship and readership are not correlated: the obscure article *Green Children of Woolpit* has the most readers and very few authors while *Richard Nixon* has the most authors. Layering this additional information into the text can aid analysis, e.g. editors (looking for highly edited articles), researchers (looking for popular articles), or analysts (looking for relationships between both).

In addition to comparison *between rows* proportionally encoded, the approach can be used to compare *between result sets*. For example, the movie review web site http://rottentomatoes.com provides a key quote for each movie reviewer set out on a web page. However, with 100+ movie reviews and no macro-ordering the viewer is simply reading random quotes. Figure 12 shows a list of quotes evenly sample from an ordered set of reviews. The proportion of bold along the quote indicates the reviewer's score. Various visual comparisons can be made between the two sets of opinions:

- The amount of bold, per movie, is indicative of the overall rating. In this sample it can be seen that *Despicable Me 2* has less bold than *How to Train your Dragon* indicating a lower rating.
- Within a single movie, the slope formed by the boundary between bold and non-bold text provides an indication of dispersion. Between movies, the slope can be compared, e.g. *Despicable Me 2* has a shallower slope than *How to Train your Dragon* indicating higher dispersion for the former.
- The initial portion of the reviewer's opinion can be read and compared, for example, the worst review for each movie at the top; or the best reviews for each movie at the bottom.

One early concern expressed with proportional encoding is that the maximum proportion length is limited to the shortest headline (Fig. 10). This can be addressed by adding the beginning of the lead sentence to the end of the title (Fig. 11); or appending ellipses to end, similar to entries in a table of contents (Fig. 12).

The approach of proportional encoding could lead to new techniques for rapidly scanning through search results and comparing multiple search results. For example, news search on financial information systems (e.g. Reuters, Dow Jones, Bloomberg) typically results in dense lists of headlines on screens where space is at a premium. Font-based approaches can be used to provide additional information without requiring any additional space.

5.2.2 *Word lists and stem and leaf text plots: for facets, categories, topics, entities, keywords, descriptors, etc.*

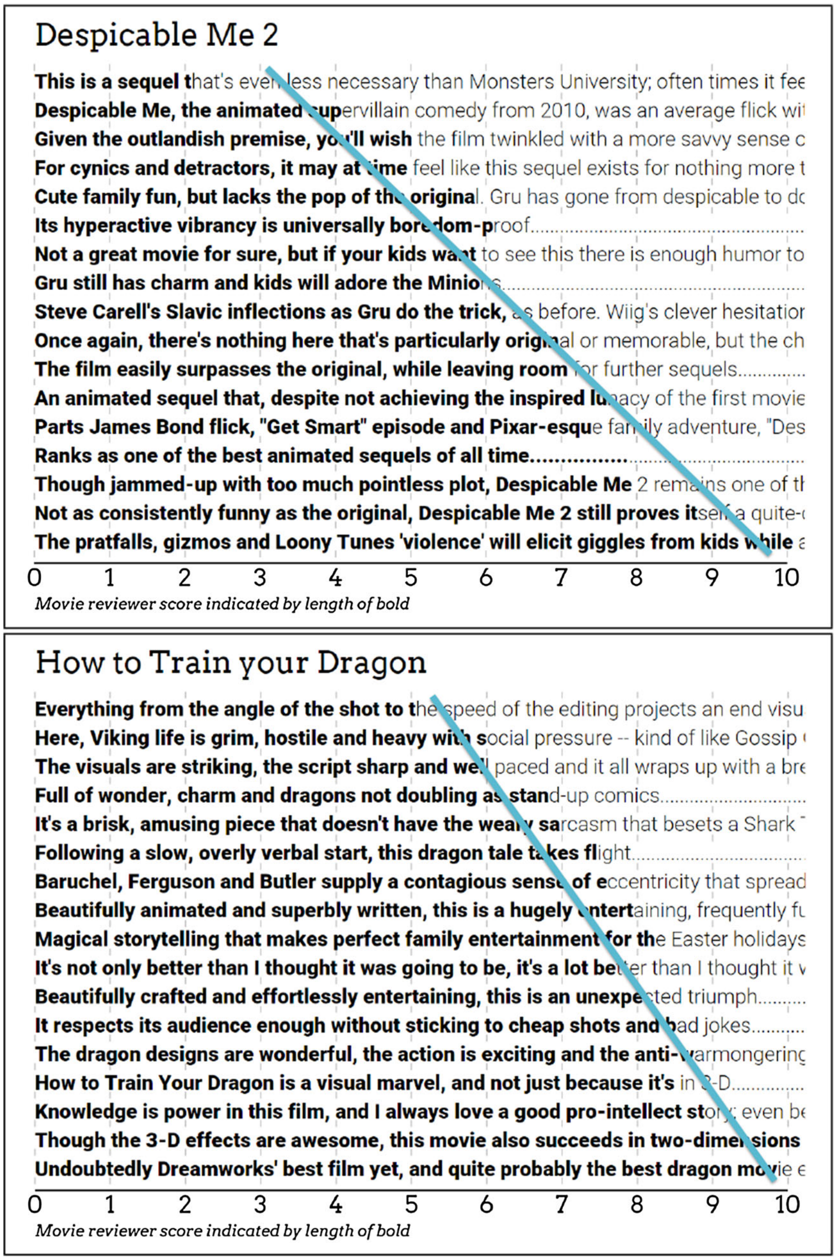
With query results there may also be a variety of additional metadata. For example, lists of facets (keywords) are often in narrow side panels and used to filter results.

In some cases, proportional encoding may be effective. Figure 13 represents facets for query refinement where the length of underline has been added to indicate the quantity of matching items (this example based on an Amazon search for *information retrieval*).

In other uses proportional encoding may not be effective. Sometimes there are many categories, making a list cumbersome to get an overview as the entire list may not fit on screen when set out linearly, or there may be lists of lists, such as a list of topics with each topic containing a list of keywords, or entity extraction with each entity containing a list of descriptors. In these cases, *stem and leaf plots* (as popularized by [59,60]) can be extend with font attributes:

Entity extraction is a subtask associated with search and information retrieval that seeks to extract and classify elements in text such as the names of persons, organizations and locations. The entities can be enhanced with additional information from the text, such as sentiment, keywords or other descriptors associated with those entities.

Figure 14 is an extension of a *stem and leaf plot* to text. The stem, left side, lists characters from *Grimms' Fairy Tales*. The leaves, right side, list adjectives occurring within ±3 words



Despicable Me 2

This is a sequel that's even less necessary than Monsters University; often times it fee
Despicable Me, the animated supervillain comedy from 2010, was an average flick wi
Given the outlandish premise, you'll wish the film twinkled with a more savvy sense c
For cynics and detractors, it may at time feel like this sequel exists for nothing more t
Cute family fun, but lacks the pop of the original. Gru has gone from despicable to dc
Its hyperactive vibrancy is universally boredom-proof......................................
Not a great movie for sure, but if your kids want to see this there is enough humor to
Gru still has charm and kids will adore the Minions..
Steve Carell's Slavic inflections as Gru do the trick, as before. Wiig's clever hesitatior
Once again, there's nothing here that's particularly original or memorable, but the ch
The film easily surpasses the original, while leaving room for further sequels.............
An animated sequel that, despite not achieving the inspired lunacy of the first movie
Parts James Bond flick, "Get Smart" episode and Pixar-esque family adventure, "Des
Ranks as one of the best animated sequels of all time...........................................
Though jammed-up with too much pointless plot, Despicable Me 2 remains one of tl
Not as consistently funny as the original, Despicable Me 2 still proves itself a quite-
The pratfalls, gizmos and Loony Tunes 'violence' will elicit giggles from kids while a

Movie reviewer score indicated by length of bold

How to Train your Dragon

Everything from the angle of the shot to the speed of the editing projects an end visu
Here, Viking life is grim, hostile and heavy with social pressure -- kind of like Gossip
The visuals are striking, the script sharp and well paced and it all wraps up with a bre
Full of wonder, charm and dragons not doubling as stand-up comics...........................
It's a brisk, amusing piece that doesn't have the weary sarcasm that besets a Shark
Following a slow, overly verbal start, this dragon tale takes flight................................
Baruchel, Ferguson and Butler supply a contagious sense of eccentricity that spread
Beautifully animated and superbly written, this is a hugely entertaining, frequently fu
Magical storytelling that makes perfect family entertainment for the Easter holidays
It's not only better than I thought it was going to be, it's a lot better than I thought it v
Beautifully crafted and effortlessly entertaining, this is an unexpected triumph.........
It respects its audience enough without sticking to cheap shots and bad jokes..........
The dragon designs are wonderful, the action is exciting and the anti-warmongering
How to Train Your Dragon is a visual marvel, and not just because it's in 3-D..............
Knowledge is power in this film, and I always love a good pro-intellect story, even be
Though the 3-D effects are awesome, this movie also succeeds in two-dimensions
Undoubtedly Dreamworks' best film yet, and quite probably the best dragon movie e

Movie reviewer score indicated by length of bold

Fig. 12 Movie review quotes from *Rotten Tomatoes* ordered by reviewer rating, evenly sampled, and proportion of *bold* indicating rating

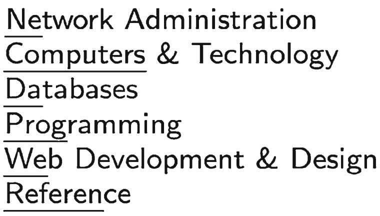


Fig. 13 Facets with *underline length* indicating number of matching items

from each character, sorted and weighted by frequency. At a macro-level, the longest adjective lists in these fairy tales are associated with *birds, kings, princesses* and *wives*. Font weight makes more frequently used adjectives visually standout. Focusing on characters with frequent adjectives, the most frequent adjectives are as expected: *princesses* tend to be ***beautiful***, whereas *kings* tend to be ***old*** and ***great***, *girls* are ***little*** and ***poor***, and *witches* tend to be ***old*** more frequently than they are wicked. The visualization can also act as an interface to both target entities (by clicking on the entity) and specific sentences associated with descriptors (by clicking on the adjectives).

Topic models may be used to characterize themes in documents, such as latent Dirichlet allocation (LDA) [4]. These models generate topics based on word co-occurrence across a corpus returning a list of words per topic. Topic model out-

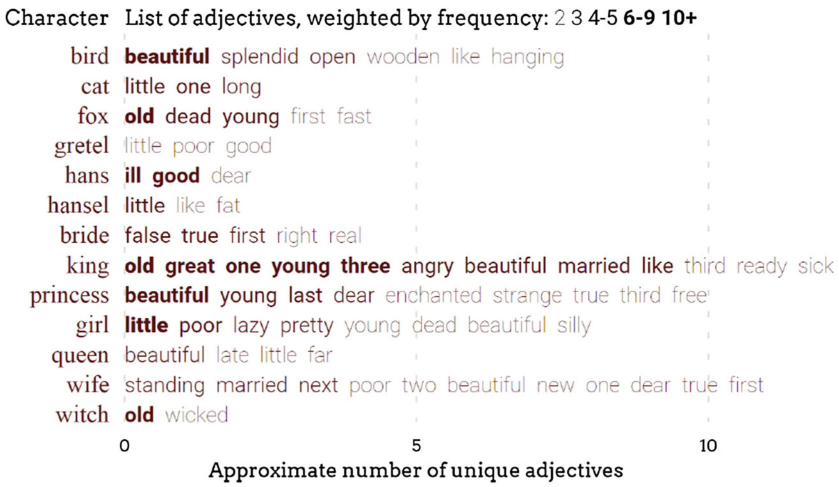


Fig. 14 Characters from *Grimms' Fairy Tales* and associated adjectives weighted by frequency

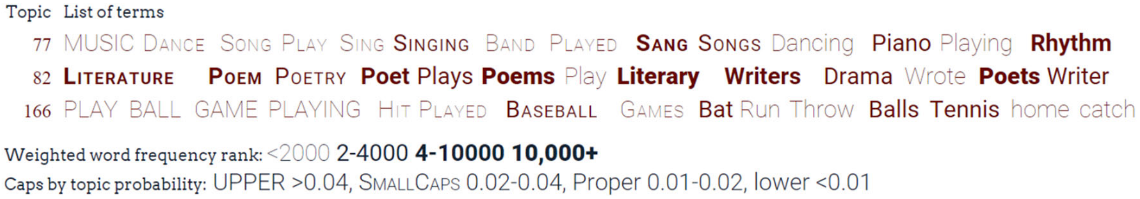

Topic List of terms
77 MUSIC Dance Song Play Sing Singing Band Played Sang Songs Dancing Piano Playing Rhythm
82 Literature Poem Poetry Poet Plays Poems Play Literary Writers Drama Wrote Poets Writer
166 PLAY BALL GAME PLAYING Hit Played Baseball Games Bat Run Throw Balls Tennis home catch
Weighted word frequency rank: <2000 2-4000 4-10000 10,000+
Caps by topic probability: UPPER >0.04, SmallCaps 0.02-0.04, Proper 0.01-0.02, lower <0.01

Fig. 15 Three topics generated by latent Dirichlet allocation. *Case* indicates probability of word in topic, *weight* indicates overall word frequency with low-frequency, less ambiguous words in higher weights

puts may need to be verified to ensure the words correspond to meaningful topics and evaluations of topic quality rely on examination of the words associated with topics (e.g. [44]). Each topic is simply a list of words and the probability associated with each word for that topic. Interfaces for displaying topics, in many cases, may simply be a list of top *n* words per topic. Some interfaces highlight topic words in the context of the original texts with markup such as bold, colour, boxes and superscripts.

Figure 15 is a view of three topics, one topic per row, based on topic analysis of the TASA corpus by Steyvers and Griffiths [56]. Words in each row are listed in order of probability occurrence which is also reinforced with capitalization from highest probability to lowest: ALL CAPS, SmallCaps, Proper Case, lower case.

There can be many challenges with topic models, for example, the meaning of words may be ambiguous. In this example, all three topics contain the word *play*, which has a different meaning in the context of each topic. The top five words in topic 166 are all ambiguous words with multiple meanings. Less frequent words (based on English language usage) may have less ambiguity and be more salient to the analysis: by weighting the infrequent words with the heaviest weight, these words visually stand out. In topic 166, **Baseball** and ***Tennis*** are more weighted and are less ambiguous terms than the higher probability terms PLAY or BALL, which are both ambiguous and both can be associated either with sports or with the arts.

In addition to probability and salience, there can be many attributes of interest to display along with keywords, such as word count, sentiment, pointwise mutual information (PMI) score, human-selected best words, other ratios. Combining these many different data sets into a visual representation is a problem of representing *set membership*.

Set membership of items that may belong to many different sets can be challenging to encode in a small space. Given a large space, a table can list each item (rows) and set properties (columns), but the table uses more space than may be available to a facet list. Instead, combinations of visual attributes can be used. However, traditional visual attributes may interfere with one another, for example, size and shape combined can result in small dots of indistinguishable shape, or multiple uses of a visual attribute such as colour may result in potentially illegible combinations (Fig. 4). Set membership of text lists can be facilitated through the use of multiple font attributes. Bold, italics, underline, font family and case can be combined together retaining legibility while maintaining the ability to comprehend the separate attributes.

Entities in facet lists may have many attributes, e.g. people have gender, age, family connections and so forth, which may be relevant to the task. For example, searching for information on the *RMS Titanic*, there are many stories regarding passengers and crew (e.g. on Google 90m Titantic results,

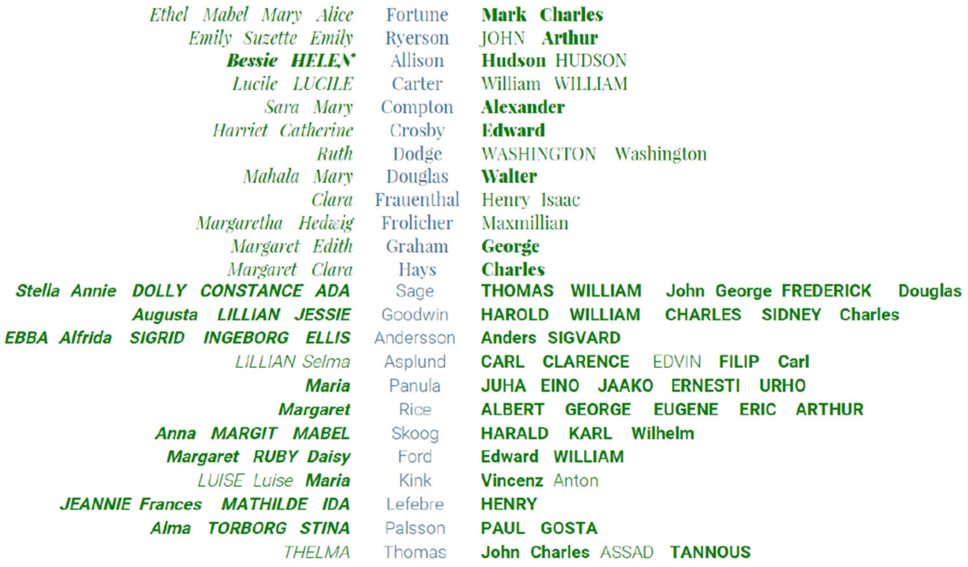


Fig. 16 24 large families from the *Titanic* with first-class families *top* and third-class families *bottom*. *One row* per family, women *left* and men *right*. *Bold* indicates death, *uppercase* indicates children, *serif font* indicates first class while *sans serif* indicates third class

2m passenger and crew results). Figure 16 is a double-sided stem and leaf plot showing a small set of *Titanic* passengers from http://encyclopedia-titanica.org with sets for gender, survival, age group and class. Each row is a family: the centre column is the family surname, with given names for female on the left and male on the right. The upper half of the image is a subset of first class passengers and the lower half a subset of third class passengers. In this example:

1. *Font weight* indicates survival: bold indicates death.
2. *Italics* represent gender: italic indicates female.
3. CASE indicates age: all caps makes children stand out.
4. Font family indicates class: a plain font for third class vs. a serif font for first class.

One can see patterns among this small set *Titanic* passengers, without a need to click through to detailed data. For example, among the first class (serif) there is more bold (deaths) lowercase (adults) non-italic on right-hand side (male). This is seemingly consistent with phrase *women and children first*. However, this pattern is not apparent in the third class: unfortunately most passengers die (bold). Similarly visually enhanced facet lists could provide insight into broad patterns and outliers at a high level otherwise not visible.

5.3 Font visualization on macro-views

Knowledge maps are macro-scale visualizations of large domains of information which provide an overview of the information structure. Labels (or interactions such as tooltips) are required to identify information such as categories, topics and entities. Font attributes can be utilized to increase the information content without requiring interaction thereby (1) providing faster access to details (a glance is faster than movement), (2) increasing information density and (3) providing the opportunity to visibly identify meaningful patterns that are otherwise not visible, i.e. serendipity [12,58].

5.3.1 Multi-attribute labels

In some knowledge maps, large amounts of information can be organized and each discrete entity plotted directly, such as the US Senate (100 senators), countries (200), stock index (30–1000) or the Internet (hundreds of thousands of web sites). However, in some cases, these maps are visualized where traditional visual attributes such as size, shape and colour may be used without any labels, for example, the senate as a floor plan (e.g. Sunlight Foundation bit.ly/1KeteBH) or just a few labels (e.g. http://internet-map.net) as in Fig. 17. To access additional information, interaction such as tooltips, zoom or reconfiguration is required, but active interaction is slower than shifting attention.

Instead discrete items can be represented as a text label gaining the benefit of identifying each unique entity, plus additional attributes conveyed in font attributes. For example, Fig. 18, shows similar information to Fig. 17, plus individual senator names, gender, tenure, age category, education, ethnicity and party affiliation. All this information can be represented in the same area while retaining legibility.

Another example is *choropleth maps* (Fig. 19) which use coloured regions to indicate data values. These maps are extremely popular, however, they have issues, including:

1. *Small areas* can be invisible (e.g. Dubai, Singapore are not visible on a world map).

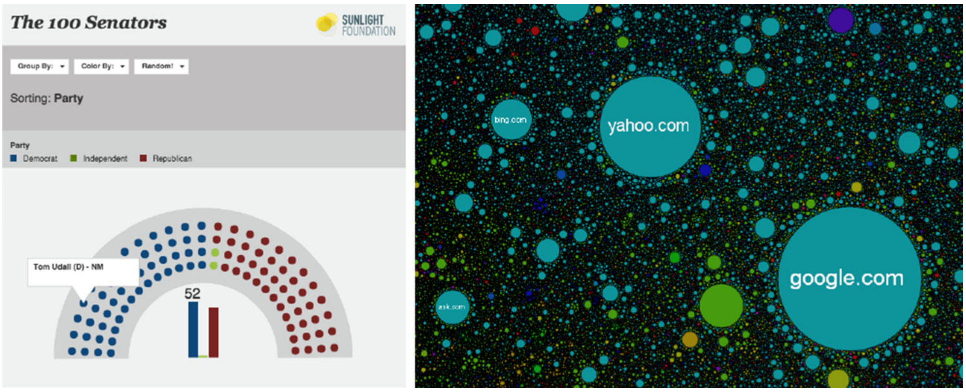
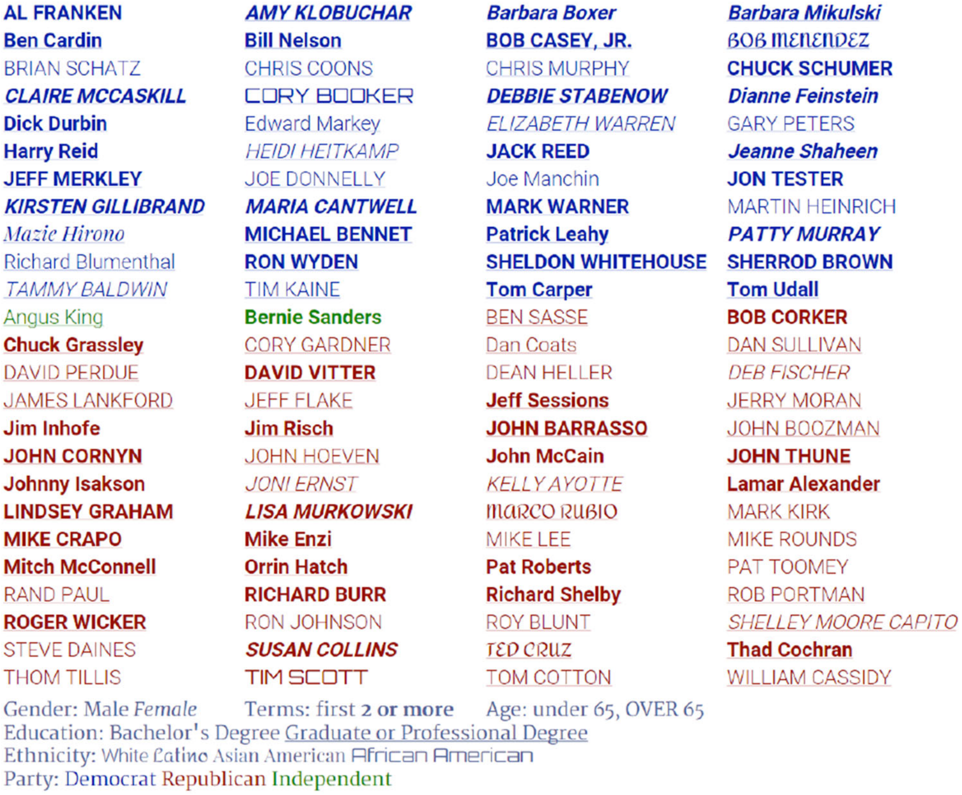


Fig. 17 Knowledge maps with low label usage. *Left* US Senate map (http://bit.ly/1KeteBH), *Right* Internet map (http://internet-map.net)



AL FRANKEN	AMY KLOBUCHAR	Barbara Boxer	Barbara Mikulski
Ben Cardin	Bill Nelson	BOB CASEY, JR.	BOB MENENDEZ
BRIAN SCHATZ	CHRIS COONS	CHRIS MURPHY	CHUCK SCHUMER
CLAIRE MCCASKILL	CORY BOOKER	DEBBIE STABENOW	Dianne Feinstein
Dick Durbin	Edward Markey	ELIZABETH WARREN	GARY PETERS
Harry Reid	HEIDI HEITKAMP	JACK REED	Jeanne Shaheen
JEFF MERKLEY	JOE DONNELLY	Joe Manchin	JON TESTER
KIRSTEN GILLIBRAND	MARIA CANTWELL	MARK WARNER	MARTIN HEINRICH
Mazie Hirono	MICHAEL BENNET	Patrick Leahy	PATTY MURRAY
Richard Blumenthal	RON WYDEN	SHELDON WHITEHOUSE	SHERROD BROWN
TAMMY BALDWIN	TIM KAINE	Tom Carper	Tom Udall
Angus King	Bernie Sanders	BEN SASSE	BOB CORKER
Chuck Grassley	CORY GARDNER	Dan Coats	DAN SULLIVAN
DAVID PERDUE	DAVID VITTER	DEAN HELLER	DEB FISCHER
JAMES LANKFORD	JEFF FLAKE	Jeff Sessions	JERRY MORAN
Jim Inhofe	Jim Risch	JOHN BARRASSO	JOHN BOOZMAN
JOHN CORNYN	JOHN HOEVEN	John McCain	JOHN THUNE
Johnny Isakson	JONI ERNST	KELLY AYOTTE	Lamar Alexander
LINDSEY GRAHAM	LISA MURKOWSKI	MARCO RUBIO	MARK KIRK
MIKE CRAPO	Mike Enzi	MIKE LEE	MIKE ROUNDS
Mitch McConnell	Orrin Hatch	Pat Roberts	PAT TOOMEY
RAND PAUL	RICHARD BURR	Richard Shelby	ROB PORTMAN
ROGER WICKER	RON JOHNSON	ROY BLUNT	SHELLEY MOORE CAPITO
STEVE DAINES	SUSAN COLLINS	TED CRUZ	Thad Cochran
THOM TILLIS	TIM SCOTT	TOM COTTON	WILLIAM CASSIDY

Gender: Male *Female* Terms: first **2 or more** Age: under 65, OVER 65
Education: Bachelor's Degree Graduate or Professional Degree
Ethnicity: White Latino Asian American African American
Party: Democrat Republican Independent

Fig. 18 US Senators. Set membership indicated by *italic, bold, caps, underline, typeface* and *colour* (colour figure online)

2. *Perception* can be biased by large regions that are more prominent than small regions but the task may require that regions be perceived equally (e.g. comparison of policies).
3. *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 (http://on.natgeo.com/1Vzfiay).
4. *Data encoding* is typically limited to a single value, such as hue or brightness.

Instead of using shapes to identify regions, mnemonic codes (e.g. ISO country codes) can be used. With some spatial adjustment to remove overlap, all labels are visible while maintaining local proximity. Visual attributes such as colour or font weight encode data, e.g. Fig. 20.

The labelled cartogram map (Fig. 20) and a corresponding geographic choropleth map with no labels (Fig. 19) have been evaluated on tasks for location and identification with two small groups of people (ten undergraduate university students, seven information visualization professionals). The identification task marked a particular country on the map and required the viewer to name the country. The location task required the viewer to find a named country on the map and report the colour of that country. Each viewer had a set of eight questions evenly distributed between the two task types and the two map types. The labelled cartogram outperforms the unlabelled choropleth map in both tasks, with an overall 2.2× performance improvement (Table 3).

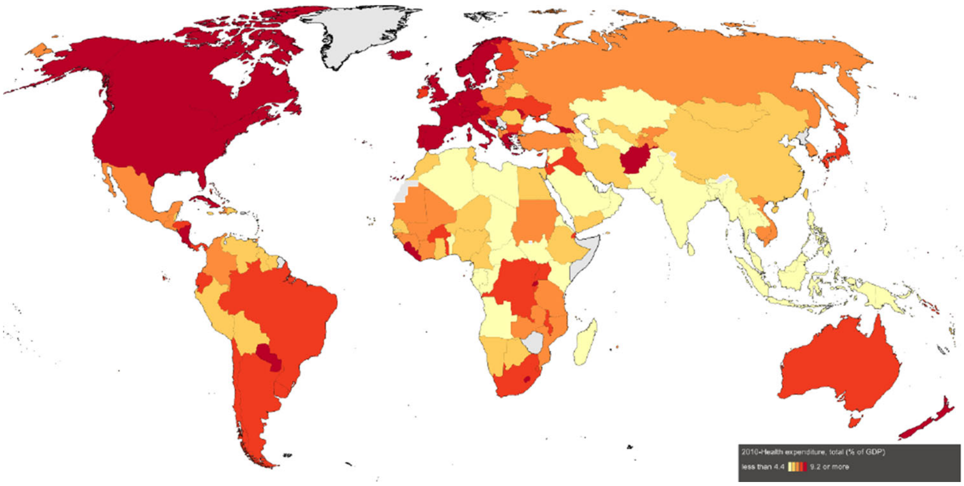


Fig. 19 Choropelth map of country health expenditure as % of GDP, from http://data.worldbank.org/indicators 04/09/2013

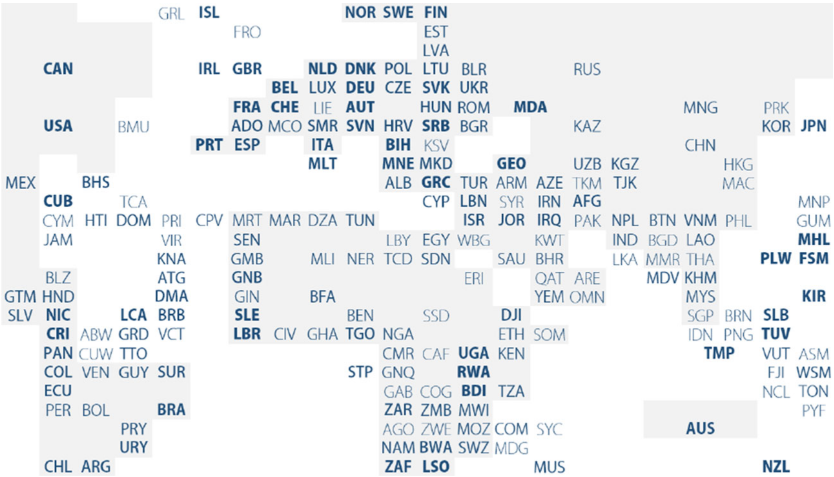

GRL ISL NOR SWE FIN
FRO EST
LVA
CAN IRL GBR NLD DNK POL LTU BLR RUS
BEL LUX DEU CZE SVK UKR
FRA CHE LIE AUT HUN ROM MDA MNG PRK
USA BMU ADO MCO SMR SVN HRV SRB BGR KAZ KOR JPN
PRT ESP ITA BIH KSV CHN
MLT MNE MKD GEO UZB KGZ HKG
MEX BHS ALB GRC TUR ARM AZE TKM TJK MAC
CUB TCA CYP LBN SYR IRN AFG MNP
CYM HTI DOM PRI CPV MRT MAR DZA TUN ISR JOR IRQ PAK NPL BTN VNM PHL GUM
JAM VIR SEN LBY EGY WBG KWT IND BGD LAO MHL
KNA GMB MLI NER TCD SDN SAU BHR LKA MMR THA PLW FSM
BLZ ATG GNB ERI QAT ARE MDV KHM
GTM HND DMA GIN BFA YEM OMN MYS KIR
SLV NIC LCA BRB SLE BEN SSD DJI SGP BRN SLB
CRI ABW GRD VCT LBR CIV GHA TGO NGA ETH SOM IDN PNG TUV
PAN CUW TTO CMR CAF UGA KEN TMP VUT ASM
COL VEN GUY SUR STP GNQ RWA FJI WSM
ECU GAB COG BDI TZA NCL TON
PER BOL BRA ZAR ZMB MWI PYF
PRY AGO ZWE MOZ COM SYC AUS
URY NAM BWA SWZ MDG
CHL ARG ZAF LSO MUS NZL

Fig. 20 Cartogram of equal-sized labels with font weight encoding data

Table 3 Performance of identification and location tasks

Task	Choropleth map (%)	ISO code map (%)	ISO code performance
Identify	15	65	4.4×
Locate	53	85	1.6×
Total	34	75	2.2×

Once labels are used instead of shape, visual attributes can be used to encode multiple data attributes. Health expenditures, life expectancies and HIV data are represented using bold, case and italic into a single label in Fig. 21; as well as colour to indicate region. Note that italics and case are encoded using a proportional encoding along the length of the string enabling these attributes to express four levels of quantities rather than only two (e.g. the mnemonic code USA using proportional case can represent four different values as USA, USa, Usa and usa). This single label allows complex queries to be made visually, e.g. Are there countries with high HIV (italics) and short lives (lowercase) even though a significant portion of GDP is spent on health care (ultrabold)? [Answer: yes, e.g. Rwanda (***rwa***) or Sierra Leone (***sle***)].

This mnemonic map can be compared for information density and relative lossiness to the choropleth map in a similar approach as discussed earlier (Sect. 5.2.2). The choropleth map loses information as small countries become too small to identify (e.g. Dubai, Singapore) as shown in Table 4. The mnemonic map does not have the loss of small countries and thus results in an incremental information density. Furthermore, the mnemonic map enables the encoding of multiple data attributes simultaneously, significantly increasing data density.

Given greater information density, lower data lossiness and improved task performance, there should be a strong potential for richly labeled maps.

5.3.2 Multi-level labels and font attributes

In many knowledge maps (e.g. scimaps.org) the designer requires a strategy for revealing labels at different scales from high-level categories and topics down to lower level entities and individual documents. Some approaches include:

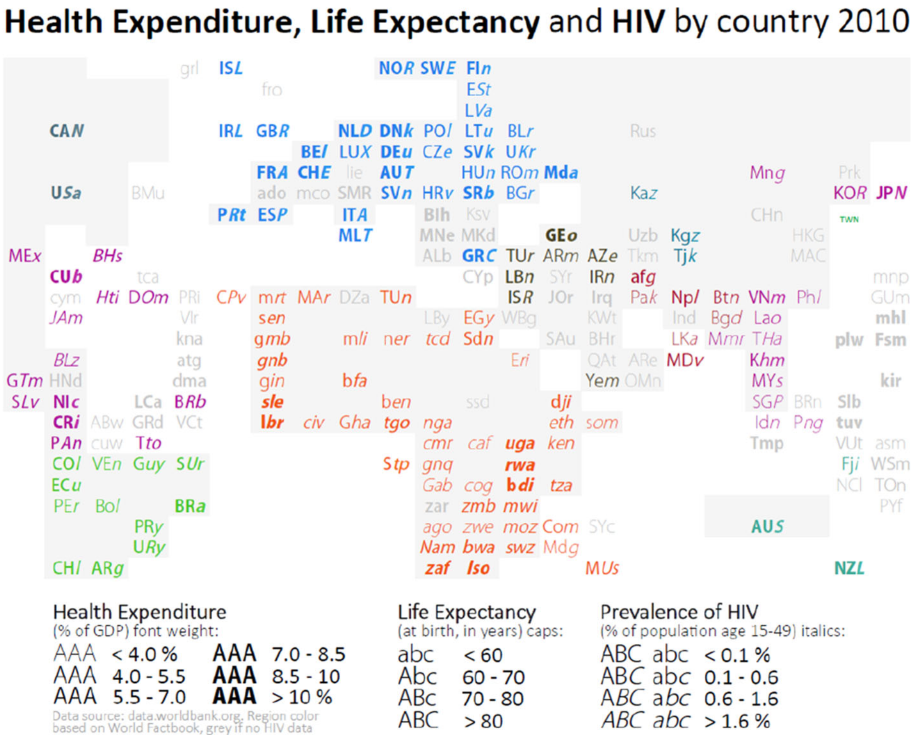

Health Expenditure, Life Expectancy and HIV by country 2010
grl ISL NOR SWE FIn
fro ESt
LVa
CAN IRL GBR NLD DNk POl LTu BLr Rus
BEl LUX DEu CZe SVk UKr
FRA CHE lie AUT HUn ROm Mda Mng Prk
USa BMu ado mco SMR SVn HRv SRb BGr Kaz KOR JPN
PRt ESP ITA Blh Ksv CHn TWN
MLT MNe MKd GEo Uzb Kgz HKG
MEx BHs ALb GRC TUr ARm AZe Tkm Tjk MAC
CUb tca CYp LBn SYr IRn afg mnp
cym Hti DOm PRi CPv mrt MAr DZa TUn ISR JOr Irq Pak Npl Btn VNm Phl GUm
JAm VIr sen LBy EGy WBg KWt Ind Bgd Lao mhl
kna gmb mli ner tcd Sdn SAu BHr LKa Mmr THa plw Fsm
BLz atg gnb Eri QAt ARe MDv Khm
GTm HNd dma gin bfa Yem OMn MYs kir
SLv NIc LCa BRb sle ben ssd dji SGP BRn Slb
CRi ABw GRd VCt lbr civ Gha tgo nga eth som Idn Png tuv
PAn cuw Tto cmr caf uga ken Tmp VUt asm
COl VEn Guy SUr Stp gnq rwa Fji WSm
ECu Gab cog bdi tza NCl TOn
PEr Bol BRa zar zmb mwi PYf
PRy ago zwe moz Com SYc AUS
URy Nam bwa swz Mdg
CHl ARg zaf lso MUs NZL
Health Expenditure
(% of GDP) font weight:
AAA < 4.0 %
AAA 4.0 - 5.5
AAA 5.5 - 7.0
AAA 7.0 - 8.5
AAA 8.5 - 10
AAA > 10 %
Data source: data.worldbank.org, Region color based on World Factbook, grey if no HIV data
Life Expectancy
(at birth, in years) caps:
abc < 60
Abc 60 - 70
ABc 70 - 80
ABC > 80
Prevalence of HIV
(% of population age 15-49) italics:
ABC abc < 0.1 %
ABC abc 0.1 - 0.6
ABC abc 0.6 - 1.6
ABC abc > 1.6 %

Fig. 21 Map indicating data via label with *colour, bold, italic* and *case* (colour figure online)

Table 4 Information density of various maps

Map variant	# Countries discernable	# Attributes	Relative info density
Choropleth 1 var	163	1	1.0
Mnemonic 1 var	187	1	1.15×
Mnemonic 4 vars	187	4	4.6×

- Skupin's self-organizing maps [53] labels with size indicating topic scale and breadth. The labels, regardless of size, do not overlap.
- Paley's graph of science [47] uses tiny discrete labels for specific topics overlaid on top of large labels for broad topics.
- Boyack and Klavans [9] use colour to indicate broad topics and overlay labels to call out specific items of interest (e.g. emerging topics).

Font attributes can be used on labels in knowledge maps and more generally any label in a visualization, such as graphs and scatterplots [12]. Figure 22 shows a knowledge map depicting the top 500 largest public companies in the US. The base map is organized as a spring-based graph layout with links based on company co-citations among news sources, research analysts, and portfolio modellers. With hundreds of companies, the company names are not depicted as there is insufficient space. Instead, the base map includes large-scale sector labels (e.g. Financials, Info Tech) and smaller industry labels (e.g. Banks, Chemicals, Software). Zoom adds successively smaller industries (e.g. Regional banks, Cruises, Internet software, etc.) and successively smaller companies.

Overlaid on the base map are news data:

- *Node colour* represents news sentiment over the day per company (a red to green gradient).
- *Label weight* indicates news volume: heavier than normal news is set in heavier font weights. Industry and sector label weights indicate net news volume for constituents.
- *Label italic angle and colour* indicate sentiment. This is a double encoding: the same variable is encoded to two different visual attributes reinforcing each other.
- *Individual company labels* are represented only if their news volume is over a user-defined threshold, are set on a dark background overlaid on the base map.
- *Uppercase company labels* indicate companies with fresh news stories in the last 10 min.

This example uses the traditional attribute of font size on the base map to distinguish between levels of hierarchy. It then adds additional attributes using colour, font angle, font weight and uppercase. In this particular example, there is not a lot of news in materials (light-weight). There is negative news in gas utilities (red and reverse italics). There is fresh news on a few companies in uppercase, such as Express Scripts, General Electric and Altera.

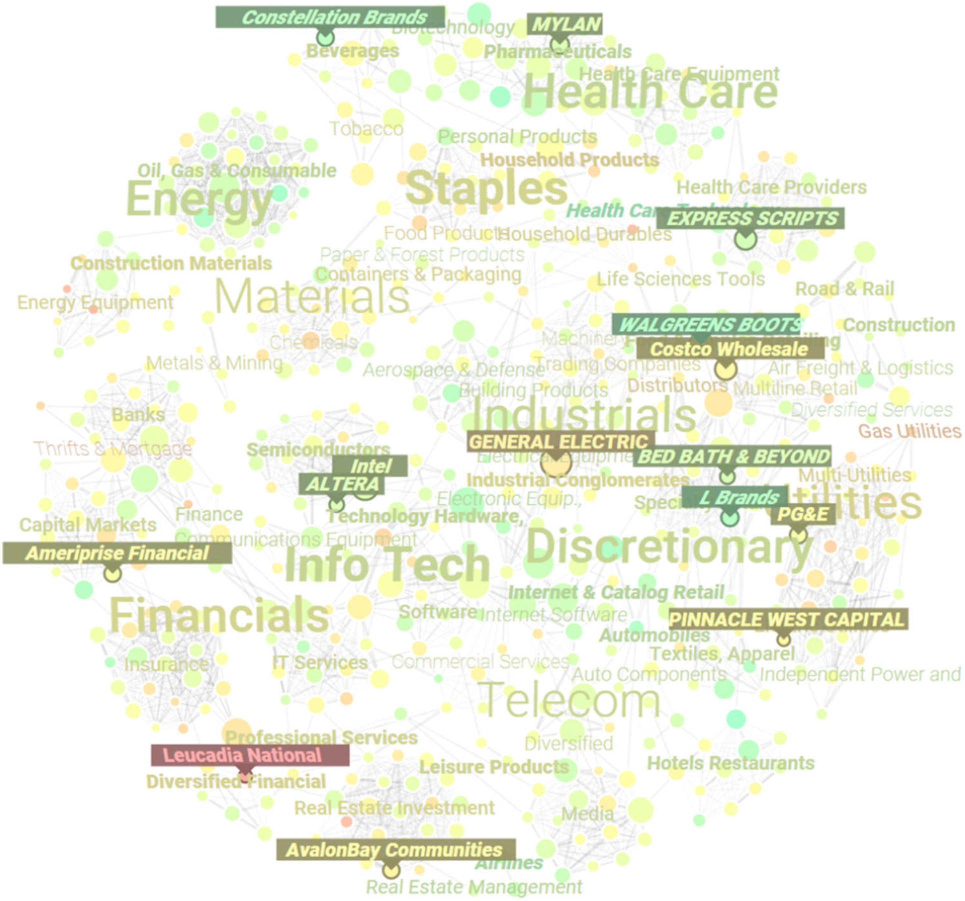


Fig. 22 Knowledge map of hundreds of companies using font angle, weight and caps to convey news sentiment, volume and recency

Compared to a traditional treemap or tag cloud, this labeled graph representation indicates three data attributes (sentiment, volume, recency via font angle, font weight and caps), plus explicit depictions of linkages and summaries at multiple levels; whereas a treemap or tag cloud typically indicates two data attributes (via size and hue) and may not represent linkages or summaries.

6 Discussion

As shown, there are eight different font attributes (weight, italic, case, typeface, underline, width, super/sub script, paired delimiters) in addition to alphanumeric glyphs. These have been characterized in relation to known visual channels and characterized for different types of data to encode. These attributes can be used to encode additional data in many application areas in KM and IR (e.g. skimming, readership, facets, entity lists, topic models, set membership, knowledge maps). The large number of font attributes, different ways of encoding data (e.g. proportional encoding, words) and the many application areas imply that there is a large design space with many potential visualizations of which only a few have been explored and presented herein. When confronted with a large design space, one should consider a broad range of alternatives across the design space rather than focusing on a narrow subset of design space, e.g. [13,42], meaning that other novel visualization techniques in KM and IR using font attributes should be explored.

The examples shown have been small-scale prototypes with limited datasets. They have been sufficient to engage with experts, conduct surveys, perform measurements and otherwise generate insight. For example, a journalist first noticed that the skim formatting preserved word location and that he could toggle interactively back and forth between unformatted and skim-formatted texts while reading.

With regards to the specific applications shown, evaluation is ongoing. Some early evaluation includes:

Expert evaluation in infovis has been conducted via interviews with a half dozen infovis practitioners, each with more than 20 years experience. Response has been been positive, with a particularly strong positive reaction to text skimming. The expert feedback has been generally positive to textual stem and leaf plots and multi-attribute labels, in part because these techniques are addressing open issues such as opinion analysis and topic analysis. Proportional encoding is understood as addressing shortcomings with other visualization techniques although there is some hesitation regarding the aesthetic appeal and time to understand, both of which should be investigated further.

Some *surveys* have been conducted including more than 20 computer science graduate level students; more than 20 actively working programmers, designers and journalists. In some cases, time was limited and the surveys brief. One survey asked participants to score different visualizations and in particular focused on the novel visualization technique of proportional encoding (Fig. 11) in terms of (a) perceived desirability, (b) understanding and (c) perceived ease of use for general population. In all three questions, proportional encoding scored either 3 or 4 on a scale from 1–4 indicated agreement or strong agreement with the statement and indicates that all participants understood the encoding technique.

Desirability, by ranking different visualization techniques, resulted in a ranking with skim formatting and proportional encoding top, followed by stem and leaf, with multi-label attributes last (e.g. Fig. 21). The latter item may have ranked lowest, because, perhaps, it has the highest number of different data attributes encoded: in the skim formatting, proportional encoding and stem and leaf encodings, typically only one or two data attributes were added into the text whereas the multi-attribute label had three to five attributes encoded. When more data attributes were encoded in a single visualization this tended to generate more questions from the participants, e.g. "what does the colour mean in that label" or "what does case mean in that stem and leaf display". This suggests where multiple data attributes are encoded into a single item there is greater potential for misinterpreting the encoding, perhaps because the cognitive load is increased when the viewer needs to remember each encoding. Visualization design-time heuristics can aid in reducing cognitive load, for example, using intuitive mappings (e.g. a data attribute representing a magnitude maps well to font weight); or by avoiding mappings where multiple visual attributes interfere with the perception of the individual attributes (e.g. integral vs. separable attributes [62]).

Measured tasks have been used to compare choropleth maps vs. labelled cartogram, specifically comparing identification and location tasks, which showed higher performance for the labelled map (Table 3). However, this result is specific to mapping of geographic entities and has not been tested more generally against a broader range of knowledge maps, such as graphs or self-organizing maps.

Measurement of lossiness and fidelity is a general approach to measuring and comparing information content across alternative visualization techniques [11] as shown in Tables 2, 3 and 4. While these calculations have not been shown for the other visualizations herein, in general, the use of font attributes increases information density compared to plain text. For example, Fig. 15 shows topic words ordered by probability, with caps indicating probability and weight indicating word frequency: the original plain text version does not show the information encoded in these attributes and has a lower information density assuming other attributes remain the same. In general, font attributes may have lower lossiness compared to some other visual attributes, such as size (e.g. Fig. 10) or shape (e.g. Fig. 19).

Design critiques from typographers were sought as typographers have deep knowledge on the subtleties of fonts, their use and applications. At a high level, critiques were positive across the wide range of new visualization techniques, with some specific criticisms related to individual techniques, e.g.:

> "This represents a whole new way of thinking about type."
> "These are very creative new uses of typography."
> "This is what typographers have always done: changing type attributes for different applications. What is new is the application of type to show quantities."
> "Type attributes visually work as validated over centuries of empirical refinement. These are new applications of proven techniques."
> "Multivariate labels are engaging and can stimulates analysis."
> "Mnemonic labels are interesting: if you do not recognize a code immediately, you have still got the code and adjacent codes that you can use as a cue to search your memory."
> "You have to be careful with the encoding. If the encoding is intuitive as in these examples, it makes the application easy to understand."
> "When you change the font, you change the semantics. So using the right attributes for the target application is required to express what you are trying to encode."
> "The skim format approach could instead encode semantics. For example, comics use conventions for shout, whisper, and so on."
> "Multivariate encoding works well for labels, but for prose readability may be impacted."
> "You have to be careful when mixing many attributes together in prose. Readability can be impacted."

While broadly positive, the typographers raise issues not identified by other evaluations, particularly with regards to issues such as readability—topics generally not considered in infovis and suitable for further investigation.

7 Conclusion

The design space for the use of text and font attributes within information retrieval and knowledge maps is large and this article illustrates a number of emerging visualization techniques. These techniques are applicable to the different components that comprise an information retrieval system, including:

- Result lists and contextual sentences via proportional encoding (Figs. 10, 11, 12).
- Paragraph and document views via skim formatting (Fig. 8).
- Facets (e.g. entities, topics) via textual stem and leaf plots (Figs. 14, 15, 16).
- Knowledge maps via multi-attribute labels (Figs. 18, 21, 22).

One challenge is current software limitations. Off-the-shelf visualization tools rarely expose font attributes. Basic bold and italic may be accessible in some tools but multiple weights, slopes and widths require programming and access to large font families. Some of these are addressable with flexible visualization libraries (e.g. D3.js) combined with commercial fonts (e.g. Adobe, Linotype) or high-quality open source fonts (e.g. http://google.com/fonts). Adobe's open source Source Sans Pro font was used in many of the examples.

Given that the use of font attributes to encode data is a newly evolving area, KM and IR researchers in the near future should acquire an understanding of techniques from infovis, typography and cartography such as texts (e.g. [6,36]) design courses, and conferences. KM and IR projects, whether short term hackathons or large-scale systems, should also consider cross-disciplinary collaboration between domain experts, text analytics experts, and researchers from infovis, typography and/or cartography.

Future work should consider different representations, i.e. there may be other novel visualizations. For example, the semantics of the text, rather than word frequency, readership or other statistics, could be reflected in the typographic attributes—such as the formatting of words to indicate emotions, or the examples shown here have focused at the level of words, sentences and paragraphs: how might variation of font attributes be utilized within a single word, for example, to indicate word stems in a keyword search or topic analysis.

Feedback, metrics and surveys collected so far are promising and indicate that font-specific attributes could be used to increase data density in texts (e.g. abstracts, lead paragraphs), lists (e.g. result lists and word lists) to aid tasks such as fact finding, information gathering and query refinement, and also add information to knowledge maps, (e.g. use of labels and data added to labels). There are many additional different levels of evaluation that should be considered including, evaluation of the efficacy of font attributes to encode data; issues of interference when multiple attributes are used together on the same label; the usability and effectiveness of specific novel font-based techniques; and the creation and evaluation of a system that uses a combination of these novel techniques in a larger KMIR system.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Afzal, S., Maciejewski, R., Jang, Y., Elmqvist, N., Ebert, D.S.: Spatial text visualization using automatic typographic maps. IEEE Trans. Vis. Comput. Graph. **18**(12), 2556–2564 (2012)
2. Baecker, R., Marcus, A.: Human Factors and Typography for More Readable Programs. Addison-Wesley, Menlo Park (1990)
3. Bertin, J.: Sémiologie Graphique. Gauthier-Villars, Paris (1967)
4. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003)
5. Borgo, R., Kehrer, J., Chung, D.H., Maguire, E., Laramee, R.S., Hauser, H., Ward, M., Chen, M.: Glyph-based visualization: foundations, design guidelines, techniques and applications. In: Eurographics State of the Art Reports, Eurographics Association, EG STARs, pp. 39–63 (2013)
6. Börner, K.: Atlas of Knowledge: Anyone Can Map. MIT Press, Cambridge (2015)
7. Börner, K., Klavans, R., Patek, M., Zoss, A.M., Biberstine, J.R., Light, R.P., Larivière, V., Boyack, K.W.: Design and update of a classification system: the UCSD map of science. PloS One **7**(7), e39464 (2012)
8. Bosler, D.: Mastering Type: The Essential Guide to Typography For Print and Web Design. How Books, Cincinnati (2012)
9. Boyack, K.W., Klavans, R.: Creation of a highly detailed, dynamic, global model and map of science. J. Assoc. Inf. Sci. Technol. **65**(4), 670–685 (2014)
10. Brath, R., Banissi, E.: The design space of typeface. In: IEEE Conference on Information Visualization [Poster paper] (2014)
11. Brath, R., Banissi, E.: Evaluating lossiness and fidelity in information visualization. Proc. SPIE 9397, Visualization and Data Analysis 2015, 93970H. doi:10.1117/12.2083444 (2015)
12. Brath, R., Banissi, E.: Using text in visualizations for micro/macro readings. In: IUI Workshop on Visual Text Analytics (2015b)
13. Buxton, B.: Sketching User Experiences: Getting the Design Right and the Right Design. Morgan Kaufmann, San Francisco (2010)
14. Carey, M., Lavoisne, J.: A Complete Genealogical, Historical, Chronological, and Geographical Atlas. J. Barfield Publisher Philadelphia, USA (1820)
15. Chambers, E.: Cyclopaedia. London, UK (1728)
16. Chan, M., Podlaseck, M.: A personalized navigation tool for online listening and free browsing: the glass engine. Proc. World Conf. Educ. Multimed. Hypermed. Telecommun. **2002**, 263–264 (2002)
17. Chen, M., Floridi, L.: An analysis of information visualization. Synthese **190**(16), 3421–3438 (2013). doi:10.1007/s11229-012-0183-y
18. Chuang, J., Manning, C.D., Heer, J.: Termite: visualization techniques for assessing textual topic models. In: Proceedings of the International Working Conference on Advanced Visual Interfaces, pp. 74–77. ACM (2012)
19. Cleveland, W.: Elements of Graphing Data. Hobart Press, Summit (1985)

20. Cleveland, W.S., McGill, R.: Graphical perception: theory, experimentation, and application to the development of graphical methods. J. Am. Stat. Assoc. **79**(387), 531–554 (1984)
21. Craig, J., Scala, I.K., Bevington, W.: Designing with Type: The Essential Guide to Typography, 5th edn. Watson-Guptill, New York (2006)
22. Cuff, D.J., Mattson, M.T.: Thematic Maps: Their Design and Production. Metheun, New York (1982)
23. Dobson, T., Ruecker, S., Gabriele, S., Sinclair, S.: The mandala browser (2005). http://mandala.humviz.org/help/. Accessed 17 June 2014
24. Ellingham, H.: A chart illustrating some of the relations between the branches of natural science and technology (1948). http://bit.ly/1hlJ6VK. Accessed 11 Feb 2013
25. Fiore, A., Smith, M.A.: Treemap visualizations of newsgroups. Technical Report, Microsoft Research, Microsoft Corporation, Redmond (2001)
26. Healey, C.G., Enns, J.T.: Attention and visual memory in visualization and computer graphics. IEEE Trans. Vis. Comput. Graph. **18**, 1170–1188 (2011)
27. Hearst, M.: Search User Interfaces. Cambridge University Press, Cambridge (2009)
28. 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)
29. Heim, S.: The Resonant Interface: HCI Foundations for Interaction Design. Pearson Education Inc (2008)
30. Hodges, E.R.S.: The Guild Handbook of Scientific Illustration. Wiley, New York (2003)
31. James, H.: Ordnance Survey. Treasury minute, dated 18 May 1855, and previous papers, relating to the Ordnance Survey. Ordnance Survey, Southampton (1857)
32. Kahn, P., Lenk, K.: Design: principles of typography for user interface design. Interactions **5**(6), 15–29 (1998)
33. Kiepert, H.: Allgemeiner Hand-Atlas der Erde und des Himmels nach den besten astronomischen Bestimmungen, neuesten Entdeckungen und kritischen Untersuchungen entworfen. Weimar Geographisches Institut, Weimar (1856)
34. Krygier, J.: Making Maps: A Visual Guide to Map Design for GIS. Guildford Press, New York (2005)
35. Liu, S., Zhou, M.X., Pan, S., Song, Y., Qian, W., Cai, W., Lian, X.: Tiara: interactive, topic-based visual text summarization and analysis. ACM Trans. Intell. Syst. Technol. **3**(2), 25:1–25:28 (2012)
36. Lupton, E.: Designing Type. Yale (2004)
37. MacKinlay, J.: Automating the design of graphical presentations of relational information. ACM Trans. Graph. **5**(2):110–141 (1986)
38. Maguire, E.: Systematising glyph design for visualization. PhD thesis, Oxford University, Oxford (2015)
39. Mayr, P., Mutschke, P., Petras, V.: Reducing semantic complexity in distributed digital libraries: treatment of term vagueness and document re-ranking. Libr. Rev. **57**(3), 213–224 (2008)
40. Mazza, R.: Introduction to Information Visualization. Springer, New York (2009)
41. Muehlenhaus, I.: Web Cartography: Map Design for Interactive and Mobile Devices. CRC Press, Boca Raton (2014)
42. Munzner, T.: Visualization Analysis and Design. CRC Press, Boca Raton (2015)
43. 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)
44. Newman, D., Noh, Y., Talley, E., Karimi, S., Baldwin, T.: Evaluating topic models for digital libraries. In: Proceedings of the 10th Annual Joint Conference on Digital Libraries, pp. 215–224. ACM (2010)
45. Nielsen, J.: Tag cloud examples (2009). https://www.nngroup.com/articles/tag-cloud-examples. Accessed 29 Jan 2015
46. Paley, W.B.: Textarc: showing word frequency and distribution in text. In: Poster Presented at IEEE Symposium on Information Visualization, vol. 2002 (2002)
47. Paley, W.B.: Map of science. http://wbpaley.com/brad/mapOfScience/index.html. Accessed 11 Dec 2013
48. Pirolli, P., Card, S.: The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. Proc. Int. Conf. Intell. Anal. **5**, 2–4 (2005)
49. Proulx, P., Tandon, S., Bodnar, A., Schroh, D., Harper, R., Wright, W.: Avian flu case study with nspace and geotime. In: 2006 IEEE Symposium On Visual Analytics Science and Technology, pp. 27–34. IEEE (2006)
50. Robinson, A.: The Look of Maps. The University of Wisconsin Press, New York (1952)
51. Robinson, A., Morrison, J., Muehrcke, P., Kimerling, A., Guptill, S.: Elements of Cartography. Wiley, New York (1995)
52. Skupin, A.: The world of geography: visualizing a knowledge domain with cartographic means. Proc. Natl. Acad. Sci. USA **101**(Suppl 1), 5274–5278 (2004)
53. 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)
54. Small, D.: Navigating large bodies of text. IBM Syst. J. **35**, 515–525 (1996). http://diglib.eg.org/EG/DL/conf/EG2013/stars/039-063.pdf. Accessed 28 June 2014
55. Squire, V., Willberg, H.P., Forssman, F.: Getting it Right with Type. Laurence King Publishing London, UK (2006)
56. Steyvers, M., Griffiths, T.: Probabilistic topic models. Handb. Latent Semant. Anal. **427**(7), 424–440 (2007)
57. Teague, J.: Fluid Web Typography. New Riders, Berkeley (2009)
58. Thudt, A., Hinrichs, U., Carpendale, S.: The bohemian bookshelf: supporting serendipitous book discoveries through information visualization. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1461–1470. ACM (2012)
59. Tufte, E.: The Visual Display of Quantitative Information. Graphics Press Cheshire, Connecticut, USA (1983)
60. Tukey, J.: Some graphic and semigraphic displays. In: Statistical Papers in Honor of George W Snedecor 5 (1966)
61. Van Hoek, W., Mayr, P.: Assessing visualization techniques for the search process in digital libraries (2013). arXiv:1304.4119
62. Ware, C.: Information Visualization: Perception for Design. Springer, New York (2000)
63. Weaver, C.: Embedding interactive markdown into multifaceted visualization tools. In: IUI Workshop on Visual Text Analytics (2015)
64. Weskamp, M.: Projects: Newsmap (2004). http://marumushi.com/projects/newsmap. Accessed 03 March 2014
65. Wilkinson, L.: The Grammar of Graphics. Springer, New York (1999)
66. Wise, J., Thomas, J.J., Pennock, K., Lantrip, D., Pottier, M., Schur, A., Crow, V., et al.: Visualizing the non-visual: spatial analysis and interaction with information from text documents. In: Information Visualization, 1995. Proceedings., pp. 51–58. IEEE (1995)
67. Wolfe, J., Horowitz, T.S.: Visual search. Scholarpedia **3**(7), 3325 (2008). (Revision 145401)