9414bd54d3eb219f301608695ffcb629
ebrary

CHAPTER 6

TEXTUAL STRUCTURES

Ben Fry, U.S.: "On the *Origin of Species*: The Preservation of Favoured Traces," 2009.

The interactive online visualization depicts changes in the six editions of Darwin's *On the Origin of Species*. Each edition is color coded, allowing, for example, at a glance to see how entire volumes were changed over the course of fourteen years. Given the scope of Darwin's work and the limited space we have on the screen, Fry enables one to read text by clicking on the colored blocks. The bottom image shows how the words are also color coded, highlighting changes and refinements in the text over the years. Fry explains, "We often think of scientific ideas, such as Darwin's theory of evolution, as fixed notions that are accepted as finished. In fact, Darwin's *On the Origin of Species* evolved over the course of several editions he wrote, edited, and updated during his lifetime. The first English edition was approximately 150,000 words and the sixth is a much larger 190,000 words. In the changes are refinements and shifts in ideas—whether increasing the weight of a statement, adding details, or even a change in the idea itself."[13] The application was built with Processing, an open source Java-based programming language he developed with collaborator Casey Reas.

http://benfry.com/traces

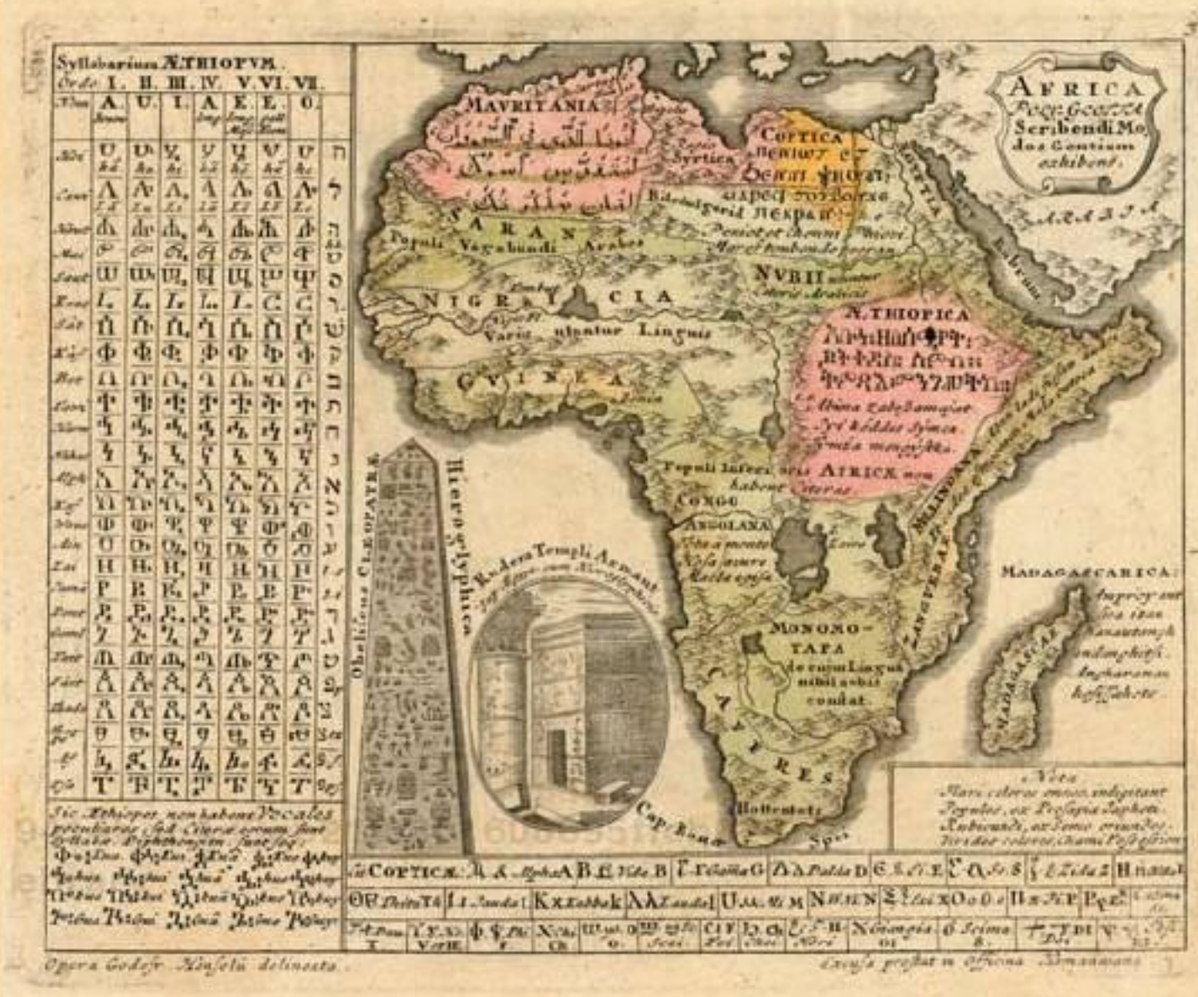
Recent advances in information storage and computational power have affected and largely facilitated the analysis of natural-language data. Large amounts of historical as well as contemporary documents are available in digital format, opening up new and powerful ways of examining literary data. Furthermore, online social interactions and conversations, mostly textual, are providing new data sources that, coupled with new research questions, are prompting understanding of social phenomena never before possible.

Methods and tools for the visualization of textual data are scarce. Examination of early books on visualization of information, including those by Willard Brinton, Jacques Bertin, and even Edward Tufte, reveal the lacuna. To my knowledge, the first book to dedicate a chapter on document visualization is *Using Vision to Think* by Card

9414bd54d3eb219f301608695ffcb629
ebrary

Meirelles, Isabel. Design for Information : An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations. Osceola, WI, USA: Rockport Publishers, 2013. ProQuest ebrary. Web. 10 May 2016.
Copyright © 2013. Rockport Publishers. All rights reserved.

9414bd54d3eb219f301608695ffcb629
ebrary



Gottfried Hensel published a series of maps in 1741 in Nürnberg, depicting the use of languages in geographic space. The language usages are demarcated in the map by means of written samples separated by dotted lines. The samples are mostly translations of the first words of the Lord's Prayer into local languages. Robinson speculates that these maps are the first ones to use colors to represent categorical data. He writes, "Hensel's map of Africa uses color to show locations of the descendants of Shem, Ham, and Japheth. His maps may be the first to use color to distinguish areas on a thematic map."[14] The use of colors is explained in the African map on the bottom-right corner as a note in Latin: the colors mark areas settled by descendants of the three sons of Noah: Japhet ("rubicundi," pink), Shem ("oriundos," yellow-orange), and Ham ("virides," olive green).[15]

9414bd54d3eb219f301608695ffcb629
ebrary

Meirelles, Isabel. Design for Information : An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations. Osceola, WI, USA: Rockport Publishers, 2013. ProQuest ebrary. Web. 10 May 2016.
Copyright © 2013. Rockport Publishers. All rights reserved.

9414bd54d3eb219f301608695ffcb629
ebrary

and colleagues in 1999. The introduction to the chapter "Data Mining: Document Visualization" elucidates the focus: "Emerging technology trends imply that document visualization will be an important visualization application for the future.... These trends [the World Wide Web, digital libraries, communication advances] portend a vast information ecology in which information visualization could have a major role."[1]

Indeed, we see more research directed at parsing large text datasets that includes the emerging field of digital humanities, characterized by interdisciplinary collaborations, and the use of other analytical tools, often in combination with the more traditional interpretative methods of inquiry. In his seminal book *Graphs, Maps, Trees*, Moretti argues for a "distanced reading" of literature that calls for models rather than text. The method proposes processes of reduction and an abstraction of literary corpus instead of the reading of individual works—i.e., a quantitative approach. Moretti contends, "Quantitative research provides a type of data which is ideally independent of interpretations ... and that is of course also its limit: it provides data, not interpretation."[2]

Outside the academic domain, the largest contribution to the visualization field has come from the collaborative team of Fernanda Viégas and Martin Wattenberg, who together have devised and made public several tools available through the IBM website ManyEyes (www-958.ibm.com). For example, *Phrase Net* and *Word Tree* are tools widely used by both the general public and academics (see pages 196–203). When asked about new frontiers in visualization in a 2010 interview, Viégas and Wattenberg argued, "One of the things I think is really promising is visualizing text. That has been mostly ignored so far in terms of information visualization tools, and yet a lot of the richest information we have is in text format."[3]

9414bd54d3eb219f301608695ffcb629
ebrary

Nominal Data

Objects, names, and concepts are examples of nominal data. We distinguish nominal datum on the basis of quality: A is different from B. The questions we ask about nominal data are what and where. Nominal data have no implicit quantitative relationship or inherent ordering, and questions such as how much don't apply. Consider the following nominal data: trouser, shirt, banana, fish. We cannot say that trousers are ranked higher than bananas without adding other kinds of information. We can organize the data, but we need to make use of external methods, such as organizing alphabetically, for example.

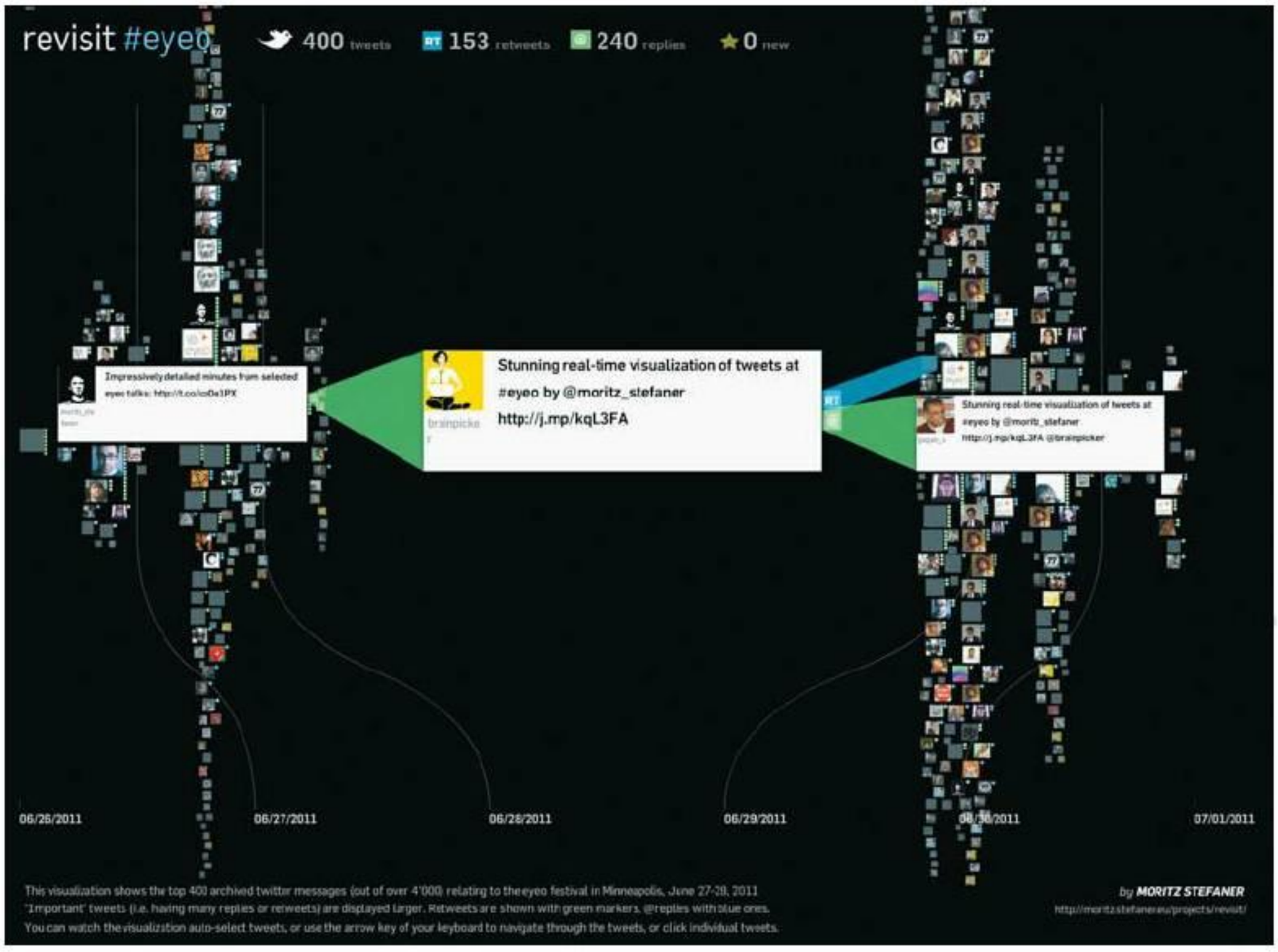
When we organize nominal data, changes in the data type might happen. For example, if we decide to count how many times each word appears in this book, we would be able to order the words according to their frequency in the text, but what started as nominal data now becomes ordinal data. In other words, ordering or sequencing doesn't apply to nominal data, unless we impose some external order that might change their nature.

Nominal datum can share characteristics that might distinguish it from others, and more important, allow grouping. Bananas and trousers are different kinds of stuff: the first we normally eat, and the latter we normally wear. On the other hand, we can eat bananas and fish as well as group them under a food category, even though one would be a member of a fruit subcategory and the other would not. Because categorization plays a major role in manipulating nominal data, it is often called categorical data.

Nominal data are considered qualitative and are rarely visualized without correlating to other kinds of data. For example, we could rank (ordinal) countries (nominal) according to the amount of exports (quantitative) of bananas (nominal).

Meirelles, Isabel. Design for Information : An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations. Osceola, WI, USA: Rockport Publishers, 2013. ProQuest ebrary. Web. 10 May 2016.
Copyright © 2013. Rockport Publishers. All rights reserved.

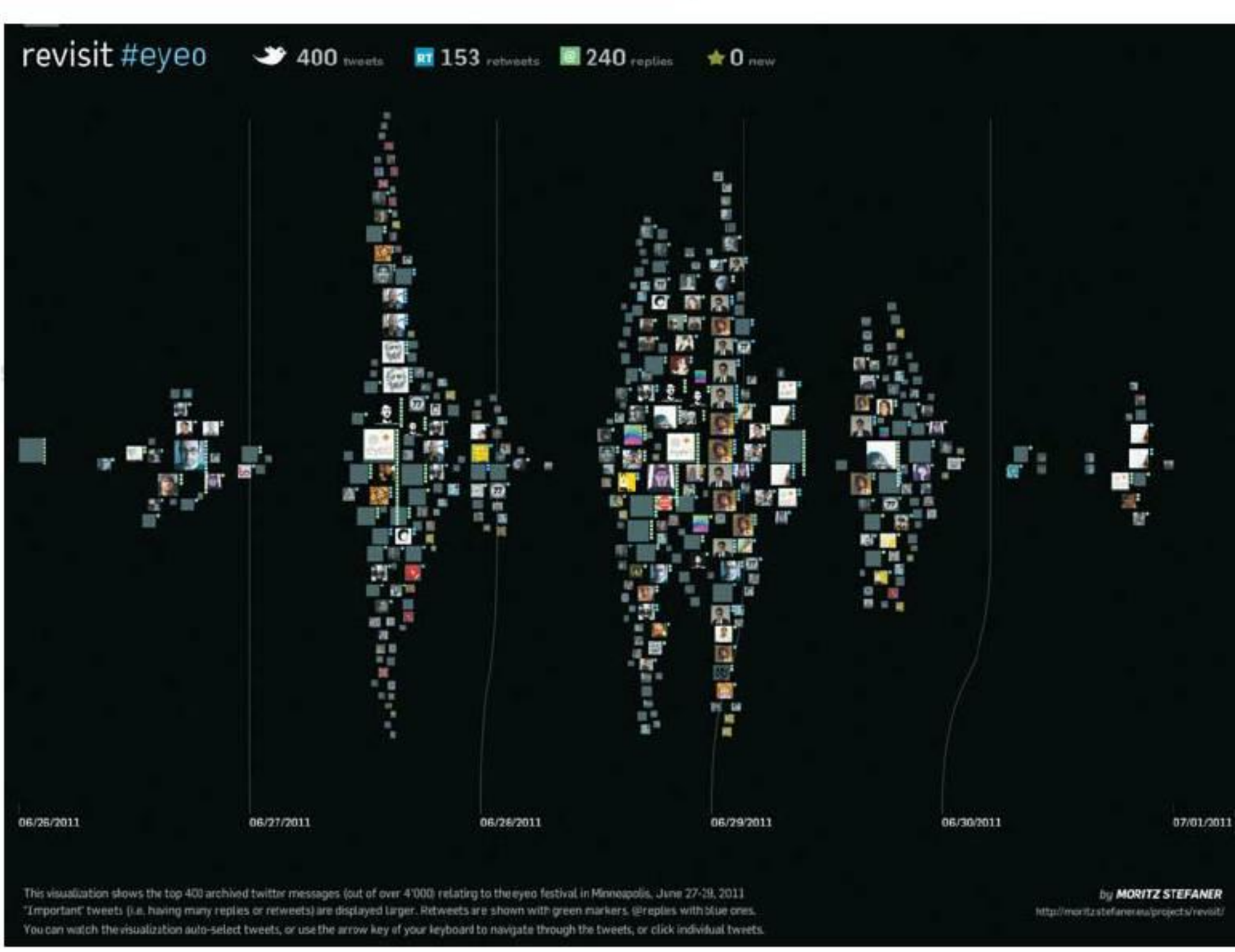
9414bd54d3eb219f301608695ffcb629
ebrary



Moritz Stefaner, Germany: "Revisit," 2010.

"Revisit" by Moritz Stefaner (2010) is a real-time visualization of Twitter messages around a specific topic. The system has been used at numerous conferences as a visual backchannel, including SEE Conference, Alphaville, VisWeek, and Eyeo Festival. The interactive application depicts flows of tweets while showing their connections. The network of tweets is organized horizontally by time, with earlier time to the left-hand side. Tweets are connected if they share content, either by the action of retweet (depicted by the blue color) or by @-reply (green). Individual tweets are represented by the squared icon of its author, with its size proportional to its importance, given by frequency of retweets or replies connected to each tweet. As Stefaner explains, "In contrast to other Twitter walls used at public events, it provides a sense of the most important voices and temporal dynamics in the Twitter stream, and reveals the conversational threads established by retweets and @-replies."[16]

http://moritz.stefaner.eu/projects/revisit-twitter-visualization



9414bd54d3eb219f301608695ffcb629
ebrary

Meirelles, Isabel. Design for Information : An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations. Osceola, WI, USA: Rockport Publishers, 2013. ProQuest ebrary. Web. 10 May 2016.
Copyright © 2013. Rockport Publishers. All rights reserved.

9414bd54d3eb219f301608695ffcb629
ebrary

Image from fourteenth-century illuminated manuscript Codex St. Peter perg 92, leaf 11v, depicting Raimundus Lullus and Thomas le Myésier: *Electorium parvum seu breviculum* (after 1321).

9414bd54d3eb219f301608695ffcb629
ebrary

TYPES OF VISUALIZATIONS

Most text documents such as books, news articles, tweets, and poems are unstructured data, in that they do not have predefined data models. Searching for words, sentences, and topics in documents might yield the distribution of themes or frequency of words, for example. Data mining and text analytic techniques offer methods to extract patterns and structure that provide meaning to these documents. Ward and colleagues define three levels of text representation that can be used to convert unstructured text into some form of structured data for subsequent generation of visualizations:[4]

- Lexical: Transforms a string of characters into a sequence of atomic entities for further analysis.
- Syntactic: Examines and defines the function of each token. Decisions on which language model and grammars to use will further define the analytical approach.
- Semantic: Extracts the meaning of the structure derived from the syntactic level toward an analytic interpretation of the full text within a specific context.

9414bd54d3eb219f301608695ffcb629
ebrary

The goal of most natural-language data analysis is to look for patterns, structures, or relationships within a collection of documents (corpus). Depending on the task of interest (i.e., co-occurrences, relationships, evolution of topics), different types of visualizations are required. Marti Hearst identifies three types of visualizations of textual data:[5]

- Visualizations of connections among entities within and across documents: Applications are in the field of text mining, and as Hearst explains, they aim at "the discovery by computer of new, previously unknown information, by automatically extracting information from different written resources."[6]
- Visualizations of document concordances and word frequencies: Applications are in the field of literature analysis, linguistics, and other fields for which the goal is to understand the properties of language, such as language patterns and structure.
- Visualizations of relationships between words in their usage in language and in lexical ontologies: Applications are mostly in the fields of literary analysis and citation analysis.

9414bd54d3eb219f301608695ffcb629
ebrary

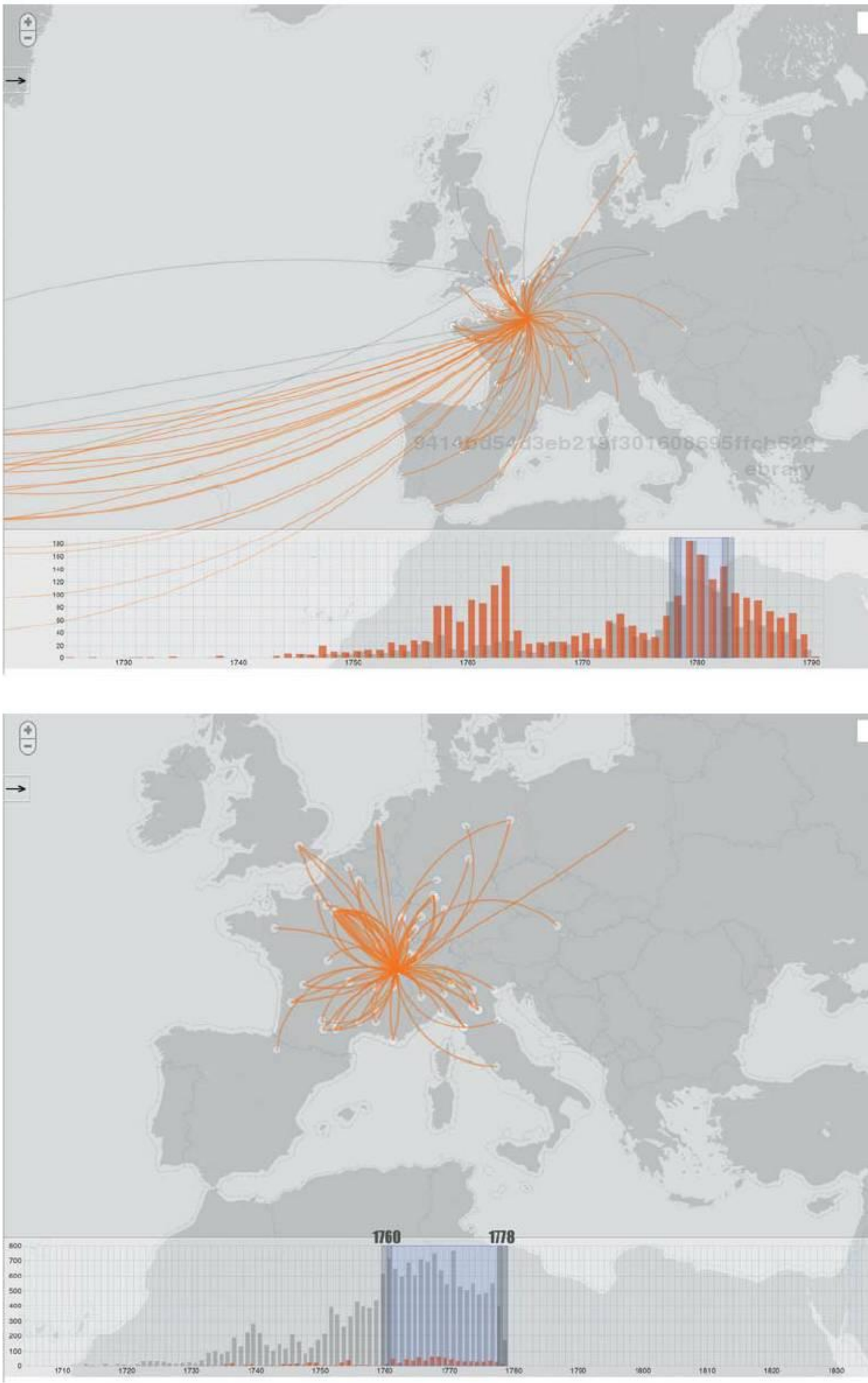
VISUAL LANGUAGE AND VERBAL LANGUAGE

Visualization of texts can be divided roughly into two large groups in what concerns the types of structures and visual elements used in the display. One group uses language, per se, as the atomic visual element in displaying linguistic data. The other uses external forms of data structures to visualize textual data, such as when we employ geographical or statistical methods to depict patterns in texts.

9414bd54d3eb219f301608695ffcb629
ebrary

Meirelles, Isabel. Design for Information : An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations. Osceola, WI, USA: Rockport Publishers, 2013. ProQuest ebrary. Web. 10 May 2016.
Copyright © 2013. Rockport Publishers. All rights reserved.

9414bd54d3eb219f301608695ffcb629
ebrary



"Mapping the Republic of Letters"team at Stanford University, U.S.: *"Corrispondenza,"* 2010.

Corrispondenza is a geographic correspondence viewer combined with a focusable timeline created at Stanford University for the "Mapping the Republic of Letters" collaborative project in the digital humanities. The goal of the visualization is to depict spatially and temporally the correspondences among early-modern scholars.

The tool uses a timeline depicting two data measures by year: the letters plotted on the map and those not plotted. They explain, "We added to this a feature that shows on the map connections that do not have dates, so, letters that do not appear on the timeline. If there is no date for a letter, there is no place to put it on the timeline. As long as we have a source and a destination, we indicate that line as a gray line that is persistent, i.e. does not change with the change in time period."[17] This feature can be seen in the top image depicting the Franklin letters. The visualization includes both letters that are missing location information, which are represented by gray bars in the timeline, and letters that are missing dates, which appear as gray lines on the map.

The bottom image shows the Voltaire letters. It shows letters without location information. The result is quite dramatic, as it shows that there are many more letters not plotted than those plotted.

https://republicofletters.stanford.edu/tools

9414bd54d3eb219f301608695ffcb629
ebrary

Meirelles, Isabel. Design for Information : An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations. Osceola, WI, USA: Rockport Publishers, 2013. ProQuest ebrary. Web. 10 May 2016.
Copyright © 2013. Rockport Publishers. All rights reserved.

9414bd54d3eb219f301608695ffcb629
ebrary

The *New York Times*, U.S.: "Inaugural Words: 1789 to the Present," 2011.

The visualization "Inaugural Words: 1789 to the Present" was published in 2011 at the *New York Times* online. It looks at the language of presidential inaugural addresses. The most-used words in each address are sized according to the frequency of use, and ordered accordingly. Words that were used significantly more in an address than average appear highlighted in yellow. Selecting a word opens a window with the parts of the transcript where the words were enunciated. In addition, there is an interesting histogram comparing the use of the word with that of other presidents.

www.nytimes.com/interactive/2009/01/17/washington/20090117_ADDRESSES.html

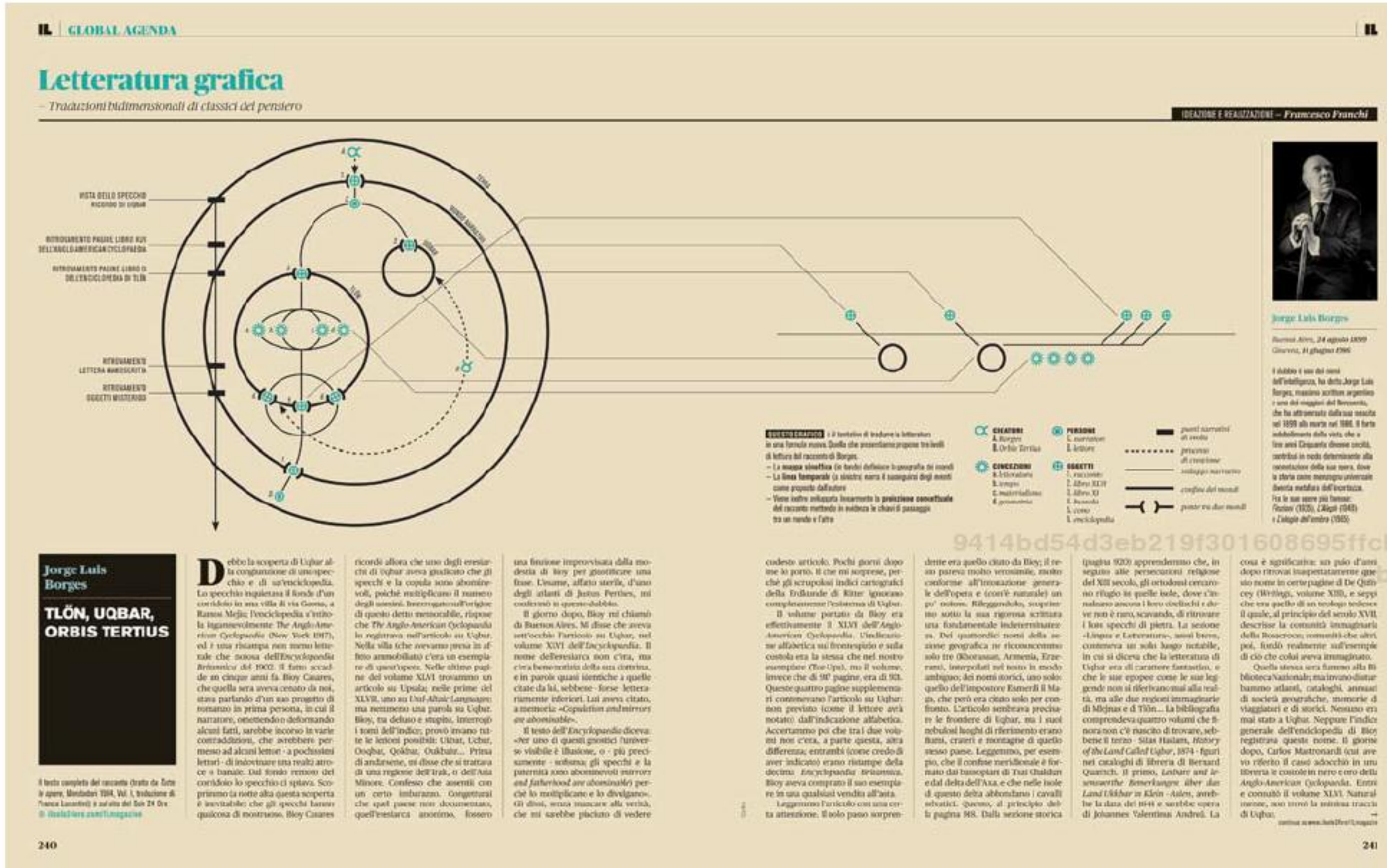
The examination of literary content by means of other data structures, such as maps, is further combined with other literary analytical methods, because they help explain all that texts can offer. Moretti explains, "What do literary maps do … First, they are a good way to prepare a text for analysis. You choose a unit—walks, lawsuits, luxury goods, whatever—find its occurrences, place them in space … or in other words: you reduce the text to a few elements, and abstract them from the narrative flow, and construct a new, artificial object like the maps that I have been discussing. And, with a little luck, these maps will be more than the sum of their parts: they will possess 'emerging' qualities, which were not visible at the lower level."[7] An example is the interdisciplinary and international project in the digital humanities centered at Stanford University, "Mapping the Republic of Letters." Since 2008, the initiative has developed several visual analytical tools that include the use of maps and quantitative approaches to examining the correspondence, travel, and social networks of early-modern scholars in the world. Another example is the quantitative analysis of the frequency and evolution of regular and irregular verbs in English language led by linguist Steve Pinker.[8]

The focus of this chapter is on visualizations that examine linguistic data within a document or corpus by using written language to represent itself—in other words, when a typographic system is the main visual system in conveying information. Though in high demand, due to the growing need to analyze large amounts of

9414bd54d3eb219f301608695ffcb629
ebrary

Meirelles, Isabel. Design for Information : An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations. Osceola, WI, USA: Rockport Publishers, 2013. ProQuest ebrary. Web. 10 May 2016.
Copyright © 2013. Rockport Publishers. All rights reserved.

9414bd54d3eb219f301608695ffcb629
ebrary



unstructured data available digitally, analytical methods that use a language–typography correspondence are small in number. Hearst contends, "Nominal or categorical variables are difficult to display graphically because they have no inherent ordering. The categorical nature of text, and its very high dimensionality, make it very challenging to display graphically."[9]

Francesco Franchi, Italy: "Jorge Luis Borges," 2008.

The infographic was published in 2008 in "Letteratura Grafica," a column of the Italian monthly newsmagazine *IL–Intelligence in Lifestyle.* It depicts three levels of the Argentinean writer Jorge Luis Borges's oeuvre: geographical (circular part), temporal (left-most text), and conceptual (linearly). Francesco Franchi, the art director, explains, "The column is an attempt to translate some pieces of literature classics in a nonlinear way through two dimensions, graphics and maps. The goal is to produce synoptic maps that allow the relationships between the elements of literary narrative to be seen, and specifically, to show complex relationships in a more easily understood way using linear forms."[18]

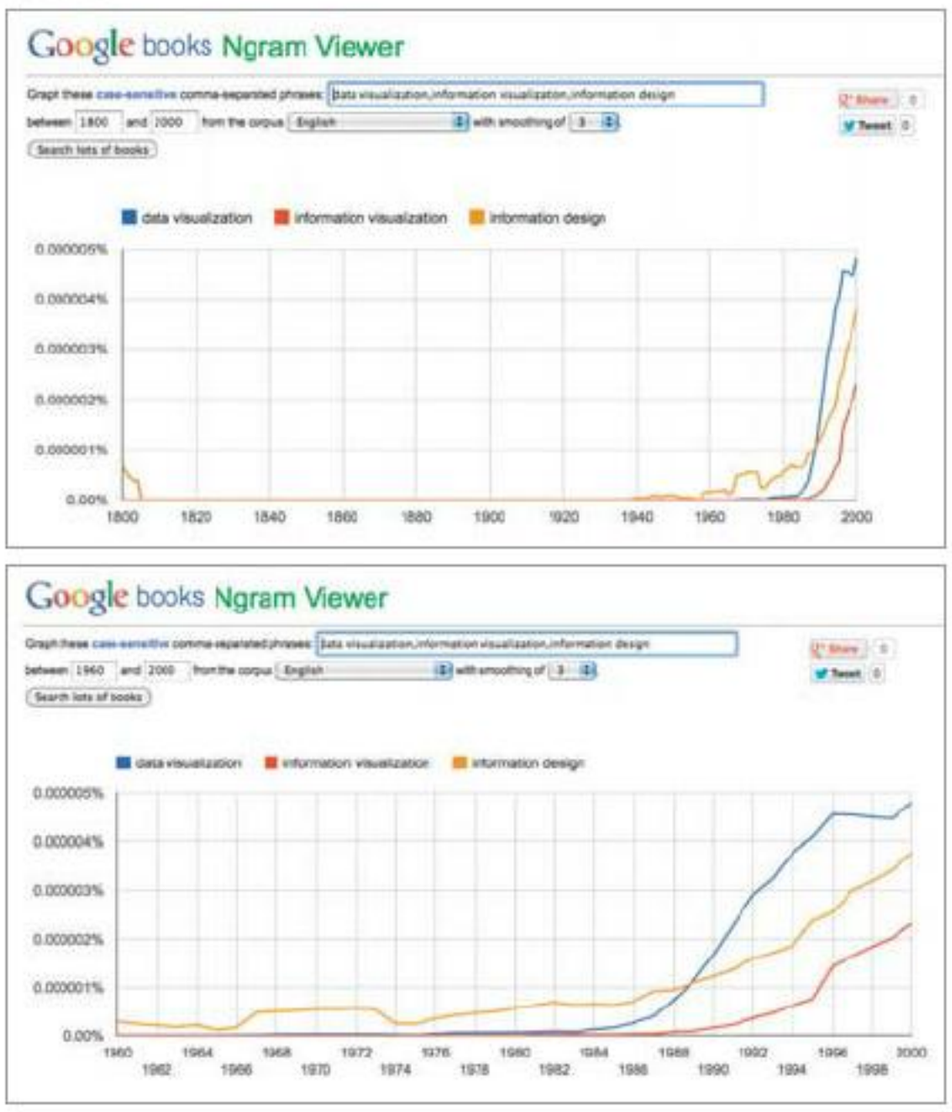
HOW WE PROCESS TEXTUAL INFORMATION

Ware explains that, under the dual coding theory, there are two fundamentally different types of information stored in distinct working memory and long-term memory systems: *imagens*, characterized by mental representations of visual information, and *logogens*, denoted by mental representations of language information, except for the sound of words.[10] He further elucidates, "Visual text is processed visually at first, but the information is rapidly transformed into nonvisual association structures of *logogens*. Acoustic verbal stimuli are processed primarily through the auditory system and then fed into the logogen system. Logogens and imagens, although based on separate subsystems, can be strongly interlinked; for example, the word cat and language-based concepts related to cats will be linked to visual information related to the appearance of cats and their environment."[11]

9414bd54d3eb219f301608695ffcb629
ebrary

Meirelles, Isabel. Design for Information : An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations. Osceola, WI, USA: Rockport Publishers, 2013. ProQuest ebrary. Web. 10 May 2016.
Copyright © 2013. Rockport Publishers. All rights reserved.

9414bd54d3eb219f301608695ffcb629
ebrary



Google Books initiative, U.S.: "Ngram Viewer," 2010.

Devised in 2010 by Google Books initiative, the Ngram Viewer allows anyone to search a word (1-gram) or several words or phrases (n-grams) in a corpus of books and examine usage over time. The top two line graphs show my searches for the usage of the terms "data visualization," "information visualization," and "information design" in the English books corpus. The topmost graph shows usage for the terms between 1800 and 2000, and below it I narrowed the search to start in 1960, because this date shows the beginning of a trend, with a growing usage for the three terms starting in the '90s. In the bottom graph, I compare trends in usage for the two possible spellings of "visualization" and "visualisation" in the same corpus of English books.

In the article "Natural Language Corpus Data," Peter Norvig argues that counting the number of appearances of words is relevant: "Why would I say this data is beautiful, and not merely mundane? Each individual count is mundane. But the aggregation of the counts—billions of counts—is beautiful, because it says so much, not just about the English language, but about the world that speakers inhabit. The data is beautiful because it represents much of what is worth saying."[19]

http://books.google.com/ngrams

Different from images and diagrams, which are understood in parallel, natural languages—whether spoken, written, or signed—are taken serially. There is an inherent temporal nature to language that transforms language into a sequence of mentally recreated dynamic utterances.[12]

PROBLEMS OF USING TYPOGRAPHY AS VISUAL ELEMENTS

There are several problems with using typography as the main visual element in visualizations, especially when using most Western writing systems. Long words occupy more space than small ones do, thus resulting in a misconceived impression of weight, given that we tend to associate size with importance. The issue is even more prominent when other visual variables, such as color and weight, are added to the typographic system, because they influence the perception of hierarchy in the graphic. A similar problem was discussed in relation to choropleth maps and how the sizes of geographic space coupled with the color encoding system mislead the interpretation of information by providing an erroneous impression of importance (see page 142).

On the other hand, when we substitute words by graphical elements other than typography we hide the information that we intend to reveal. The absence of written language in a display depicting linguistic data restricts the possibilities of interpretation of the intended information, especially when reading content is of importance. As explained in the box Nominal Data (see page 187), we understand nominal data through differentiation—in other words, by distinguishing whether two concepts are the same or different. This is one of the reasons behind labels in most graphic displays. For example, in a map with dots representing cities, we are able to differentiate cities by reading their names.

In previous chapters, we examined data structures using typographic elements to depict information in visualizations, and those are affected by the same constraints described here. What follows are three case studies that use typographic systems to depict textual data in informational displays: *Wordle*, *Phrase Net*, and *Word Tree*.

9414bd54d3eb219f301608695ffcb629
ebrary

Meirelles, Isabel. Design for Information : An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations. Osceola, WI, USA: Rockport Publishers, 2013. ProQuest ebrary. Web. 10 May 2016.
Copyright © 2013. Rockport Publishers. All rights reserved.

9414bd54d3eb219f301608695ffcb629
ebrary

CASE STUDY

Wordle

www.wordle.net

9414bd54d3eb219f301608695ffcb629
ebrary

The image and video-sharing online community Flickr devised in 2002 *Tag Cloud*, a tool that serves as both navigation and a graphic depiction of the most popular tags by their users. The method has since gained wide use, not only among tag-based websites, who use it mostly as a tag aggregation tool while affording access to content, but mostly as a means to analyze and graphically present the frequency of words in a corpus. The latter is commonly called a "word cloud."

Both representations encode the variable of word frequency to the visual property of type size. In addition, word clouds tend to include other visual parameters, mostly for aesthetic purposes, such as direction and color. For example, the website *Wordle* invites the user to define visual parameters of the graphic by offering several color schemes, fonts, and two options for word placement: alphabetical (as in all tag clouds) or center line.

9414bd54d3eb219f301608695ffcb629
ebrary

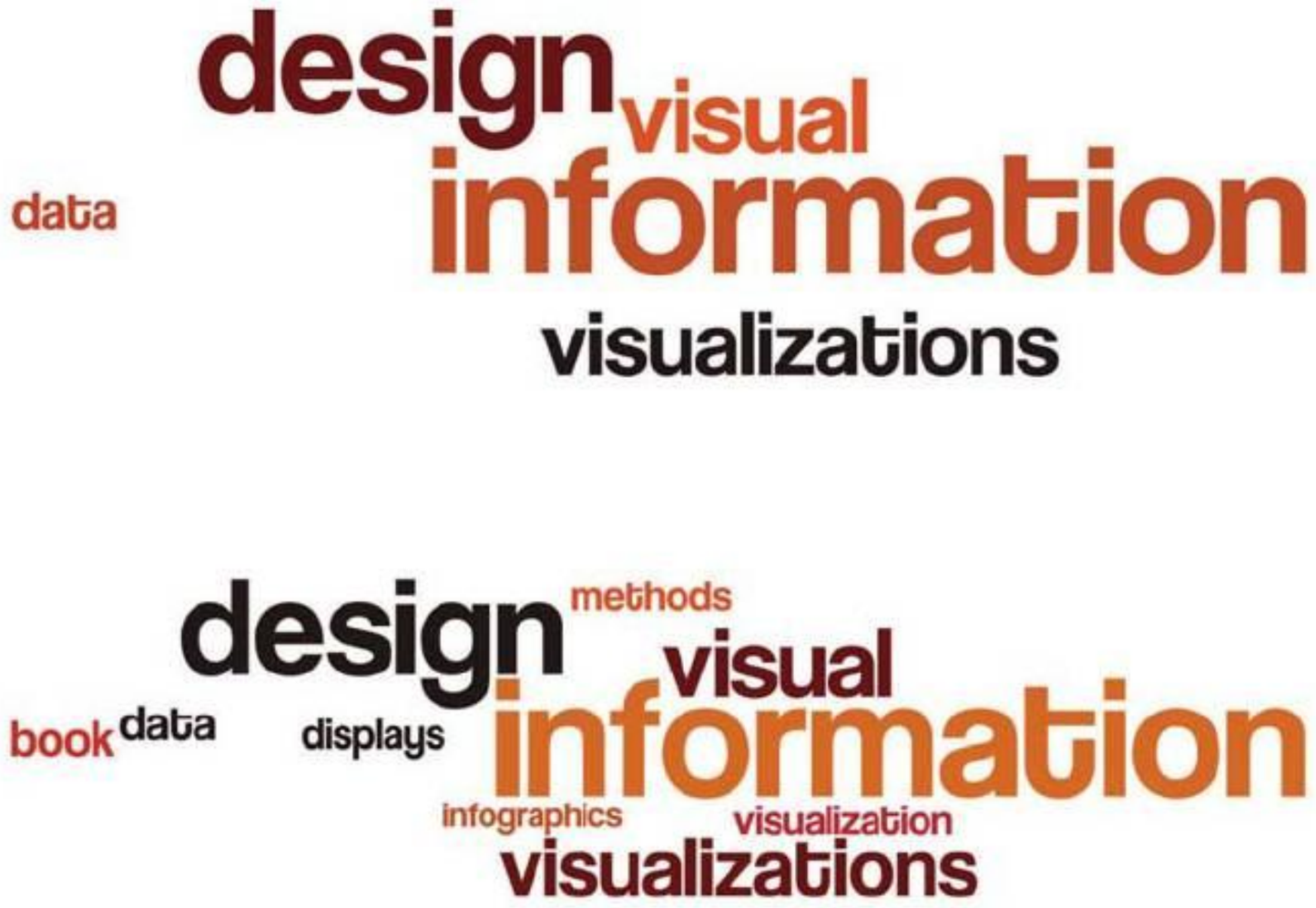
Wordle is an online tool for making "word clouds" created by Jonathan Feinberg in 2008. The online Java applet allows anyone to paste a text, choose some visual parameters, and output a word cloud for later use or sharing purposes. Similar to other textual analysis tools, Wordle removes "stop words," or high-frequency words, such as the, it, to, because otherwise the graphic would mostly contain only those. Feinberg warns that word clouds are constrained as a visualization method and points to four major caveats: word sizing is deceptive given that two words with the same frequency will be perceived differently depending on their length (number of letters); color is meaningless, in that it doesn't encode any variable and is used for aesthetic appeal; fonts are fanciful, and favors expressiveness; word count is not specific enough, because "merely counting words does not permit meaningful comparisons of like texts."[20] Yet, it is extremely popular and has been widely used.

Despite the low efficacy, word clouds have become quite popular, especially in education settings. In an investigation about usability of word clouds, Viégas and colleagues contend that learning and memory are two cognitive processes supported by word clouds, despite the fact that most people surveyed did not understand the encoding system (type size) in the graphic. They argue, "The feeling of creativity is central to the experience of using *Wordle*. Even the examples where Wordle aids learning and memory include elements of creation. For people making mementos, creativity is key to the experience; many people relate *Wordle* to scrapbooking. In the classroom, *Wordle* is not just a broadcast medium, but something that students can use themselves. One typically does not think of visualization as a creative outlet, any more than one would think of a microscope as an authoring tool. Rather than a scientific instrument, however, the type of visualization represented by *Wordle* may be more like a camera: a tool that can be used to document and create."[21]

9414bd54d3eb219f301608695ffcb629
ebrary

Meirelles, Isabel. Design for Information : An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations. Osceola, WI, USA: Rockport Publishers, 2013. ProQuest ebrary. Web. 10 May 2016.
Copyright © 2013. Rockport Publishers. All rights reserved.

9414bd54d3eb219f301608695ffcb629
ebrary



9414bd54d3eb219f301608695ffcb629
ebrary

I used the introduction of this book to generate in *Wordle* the word clouds reproduced here. All outputs used options offered in the site: Coolvetica font, Horizontal layout, alphabetical order, and the "kindled" color palette. They differ in relation to the maximum number of words in each layout, that are from top to bottom: three, five, ten, twenty-five, and finally fifty words. The larger the number of words, the harder it is to discern relevant information. Also note the changes in font size and font color among the versions due to the random way the application renders the *word clouds*.

9414bd54d3eb219f301608695ffcb629
ebrary

Meirelles, Isabel. Design for Information : An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations. Osceola, WI, USA: Rockport Publishers, 2013. ProQuest ebrary. Web. 10 May 2016.
Copyright © 2013. Rockport Publishers. All rights reserved.

9414bd54d3eb219f301608695ffcb629
ebrary

Phrase Net

www-958.ibm.com/software/data/cognos/manyeyes/page/Phrase_Net.html

9414bd54d3eb219f301608695ffcb629
ebrary

Designers are familiar with the potentials and constraints of using typography, and to what extent rendering type on a surface, be it a book or a screen, affects or affords legibility. As explained previously, there are several issues with using typography in information displays. On the other hand, natural language imposes constraints that need to be respected when the purpose of the visualization is the interpretation of meaning. For example, the ordering of words is relevant, because it indicates certain groupings that affect the semantics of the text. Viégas and Wattenberg further explain, "The conflict between positioning and legibility can lead to displays that are hard to read or where spatial position is essentially random."[22]

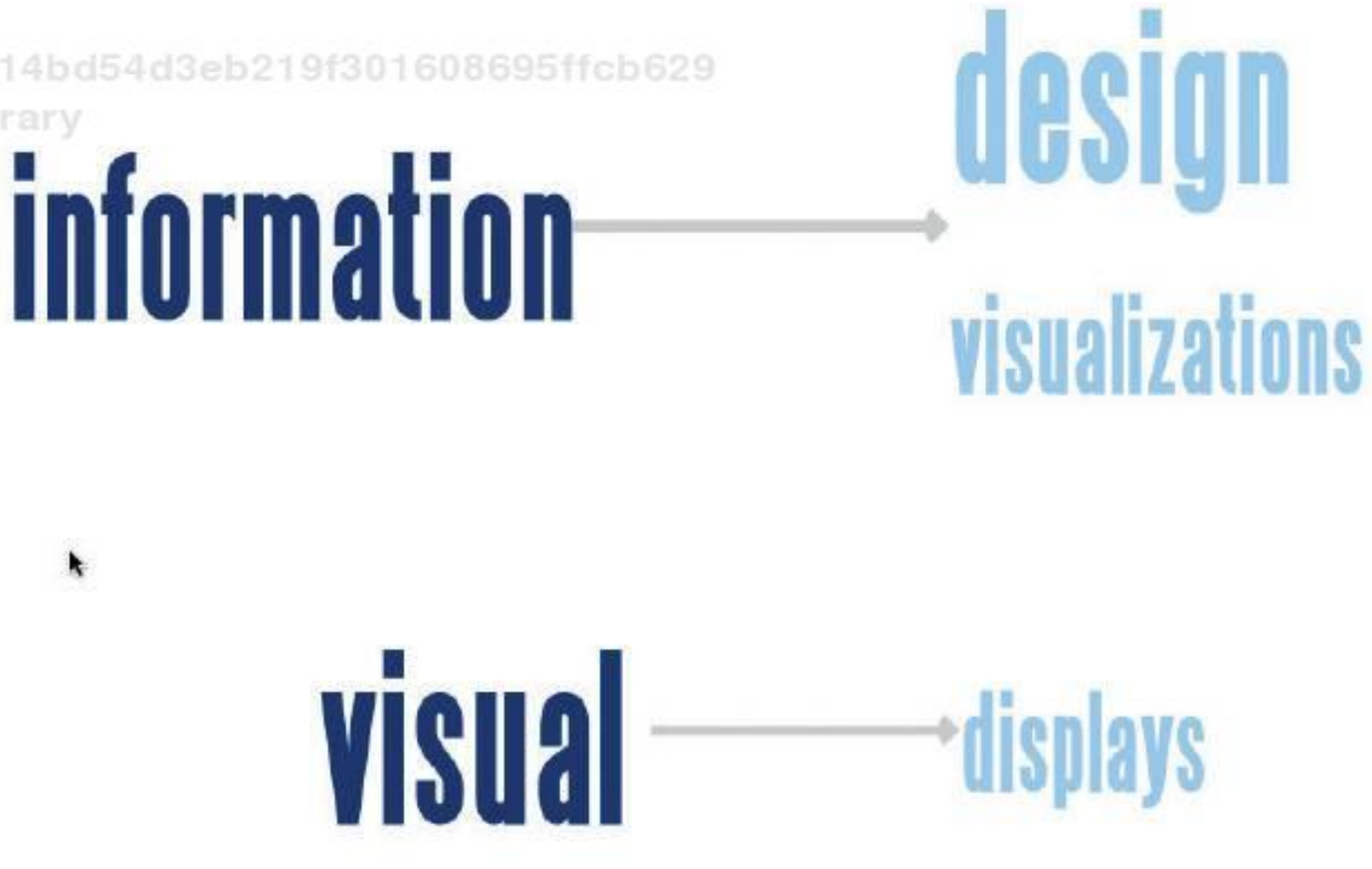
Phrase Net is an online visualization that diagrams the relationships between words in a text. The technique was devised by Fernanda Viégas and Martin Wattenberg in 2009 for IBM's site, Many Eyes. The unit of analysis is the phrase, and relationships among words in a phrase are depicted as networks while respecting syntactic ordering. The application examines how pairs of words are combined according to the parameters defined by users. For example, among the connectors in the list we find *and*, *at*, *'s*, and so on. One can also define a connector that might be appropriate to the text at hand. After the extraction of the pairs, the program then renders the result as a network, where the nodes are the words represented by means of typography, and the links are lines depicting the connections based on the selected pattern "A <connector> B." The links are weighted according to in and out connections, with the line weight representing the amount and the arrows pointing to the direction of word ordering. The type size of words represents the total number of occurrences of the term. Type is rendered in a sequential blue color palette, with the shades standing for the ratio of out-degree to in-degree, where dark blue signifies high ratio—in other words, more out-links than in-links.

9414bd54d3eb219f301608695ffcb629
ebrary

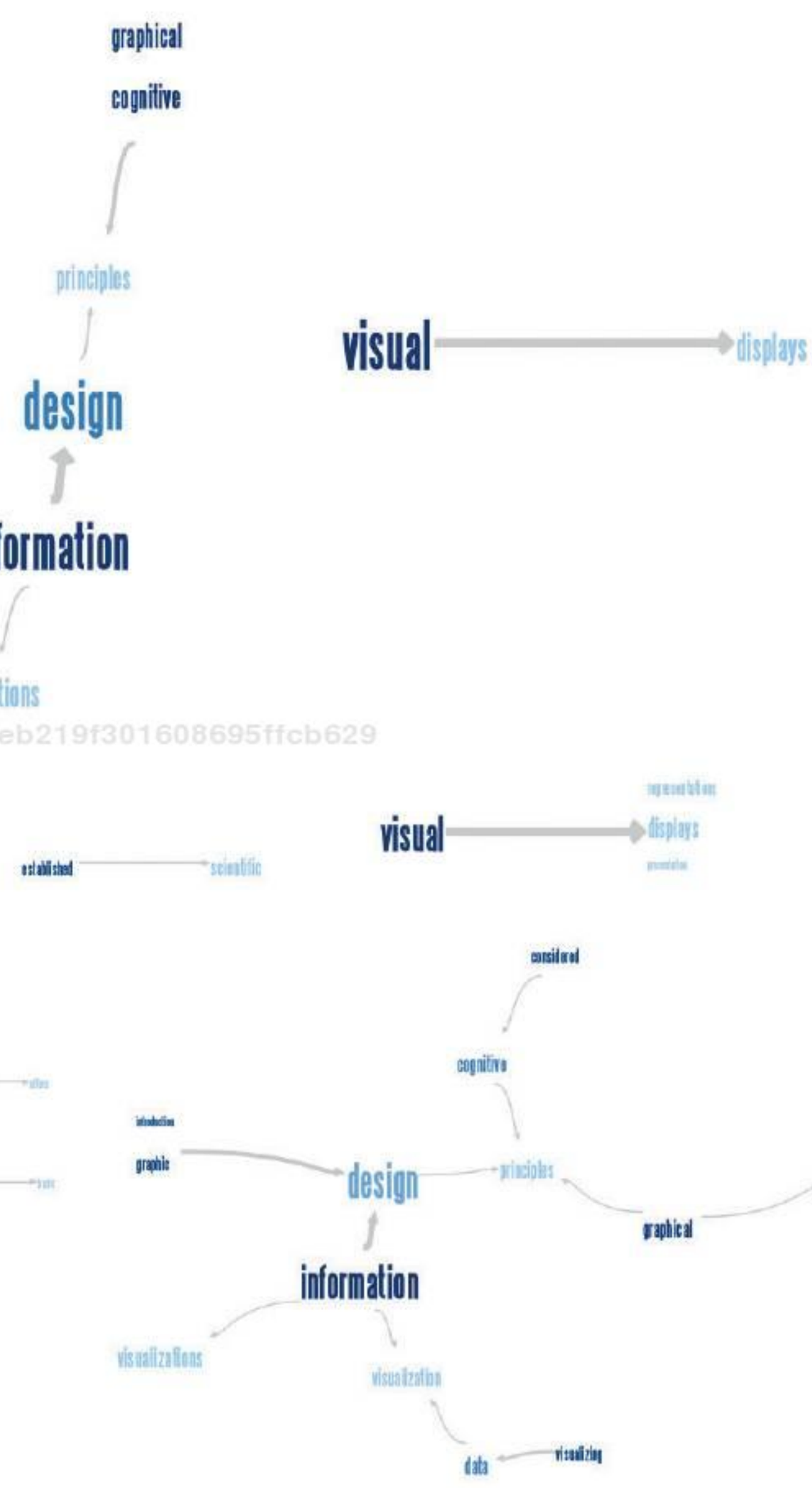
9414bd54d3eb219f301608695ffcb629
ebrary

Meirelles, Isabel. Design for Information : An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations. Osceola, WI, USA: Rockport Publishers, 2013. ProQuest ebrary. Web. 10 May 2016.
Copyright © 2013. Rockport Publishers. All rights reserved.

9414bd54d3eb219f301608695ffcb629
ebrary



9414bd54d3eb219f301608695ffcb629
ebrary



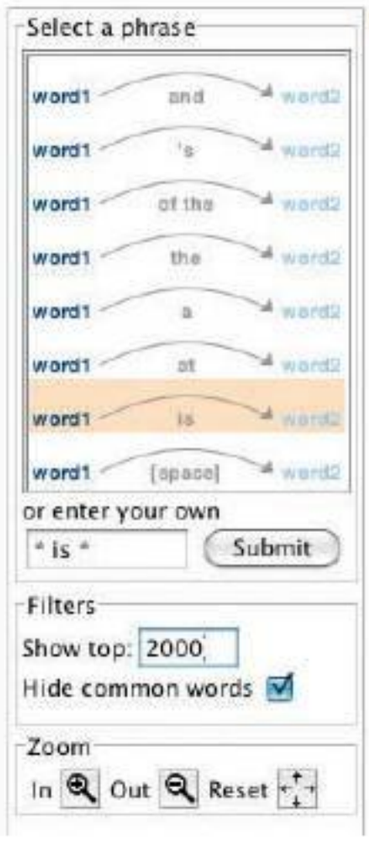
9414bd54d3eb219f301608695ffcb629
ebrary

I used the introduction of this book to generate the diagrams reproduced here. They examine pairs of words connected by a space between them. From top to bottom the diagrams show the top five, ten, and twenty-five words. In contrast to word clouds, Phrase Net renders relationships between words, including the direction of the connection, that is the word order.

9414bd54d3eb219f301608695ffcb629
ebrary

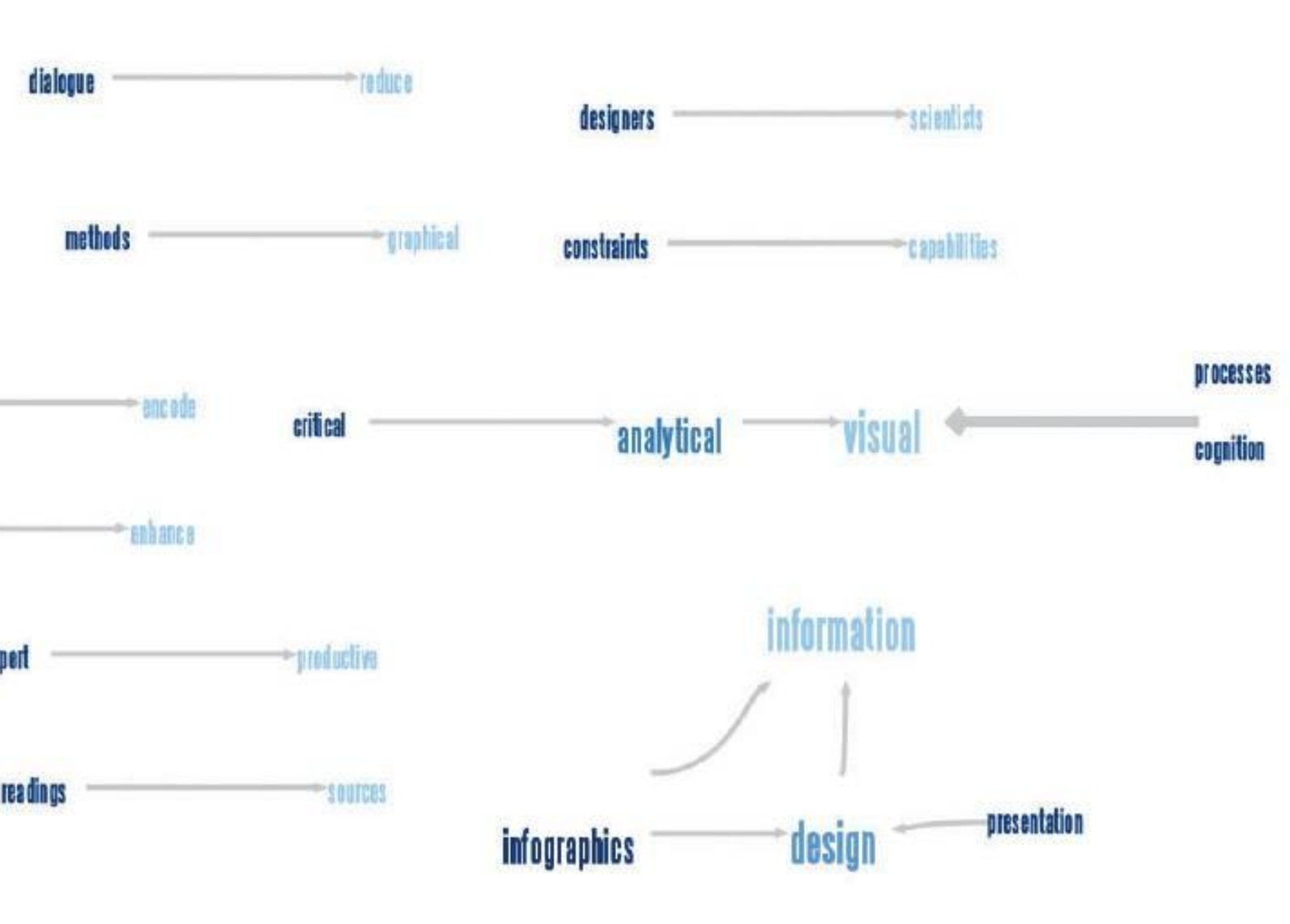
Meirelles, Isabel. Design for Information : An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations. Osceola, WI, USA: Rockport Publishers, 2013. ProQuest ebrary. Web. 10 May 2016.
Copyright © 2013. Rockport Publishers. All rights reserved.

9414bd54d3eb219f301608695ffcb629
ebrary



This Phrase Net diagram visualizes 2000 words connected by the verb *is* in the introduction of this book. The result is quite interesting, and something worth remembering: data perception is essential to visualization.

9414bd54d3eb219f301608695ffcb629
ebrary

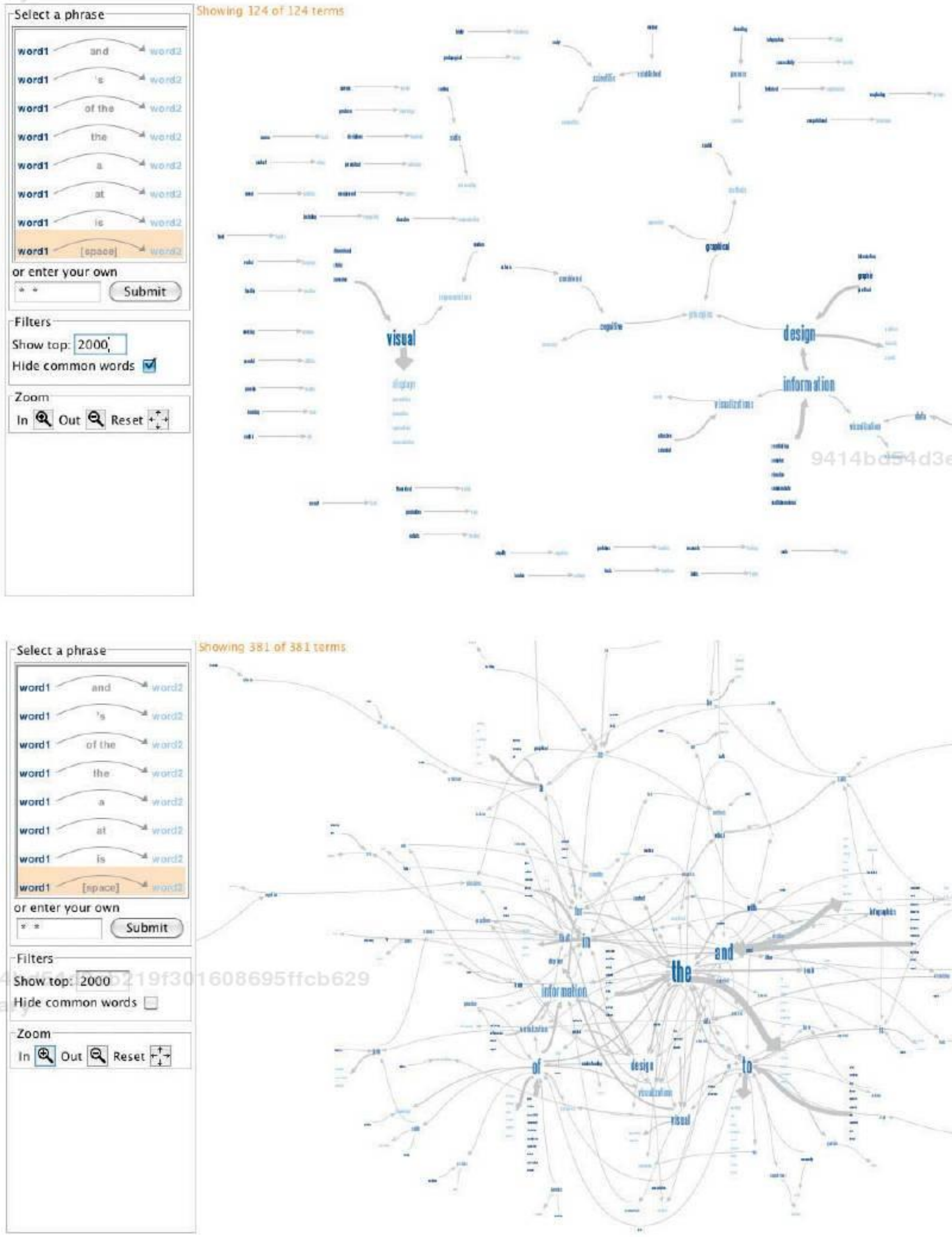


This Phrase Net diagram visualizes twenty-five words connected by the conjunction *and* in the introduction of this book.

9414bd54d3eb219f301608695ffcb629
ebrary

Meirelles, Isabel. Design for Information : An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations. Osceola, WI, USA: Rockport Publishers, 2013. ProQuest ebrary. Web. 10 May 2016.
Copyright © 2013. Rockport Publishers. All rights reserved.

9414bd54d3eb219f301608695ffcb629
ebrary



These Phrase Net diagrams visualize the same number of words (2000) connected by a space in the introduction of this book. The one at the bottom was rendered with common words in the representation. As previously discussed in chapter 2, when there are too many connections, occlusions occur and the graph becomes too complicated to be easily understood.

9414bd54d3eb219f301608695ffcb629
ebrary

Meirelles, Isabel. Design for Information : An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations. Osceola, WI, USA: Rockport Publishers, 2013. ProQuest ebrary. Web. 10 May 2016.
Copyright © 2013. Rockport Publishers. All rights reserved.

9414bd54d3eb219f301608695ffcb629
ebrary

CASE STUDY

Word Tree

www-958.ibm.com/software/data/cognos/manyeyes/page/Word_Tree.html

9414bd54d3eb219f301608695ffcb629
ebrary

Word Tree is a visual search tool for unstructured text. The technique was created by Fernanda Viégas and Martin Wattenberg in 2007 for IBM's site, Many Eyes. The visualization starts when we select a word or a phrase as the search term. Then the program looks for all occurrences of the term within the given text. It finally builds a tree structure of the content, with branches rendered until it finds a unique phrase used exactly once. There are three options for arranging the branches: alphabetically, by frequency (largest branches first), and by order of first occurrence, which reflects the original text.

The authors explain that the tool Word Tree is a visual version of a traditional concordance, also known in computer science as the visual version of a suffix tree. Besides preserving the context in which the term occurs, the method also preserves the linear arrangement of the text.

Similar to word clouds, font size represents term occurrence, with the font size proportional to the square root of the frequency of the term. Different from most text visualization methods, Word Tree does not discard stop words or punctuation, because those are considered critical for purposes of context.

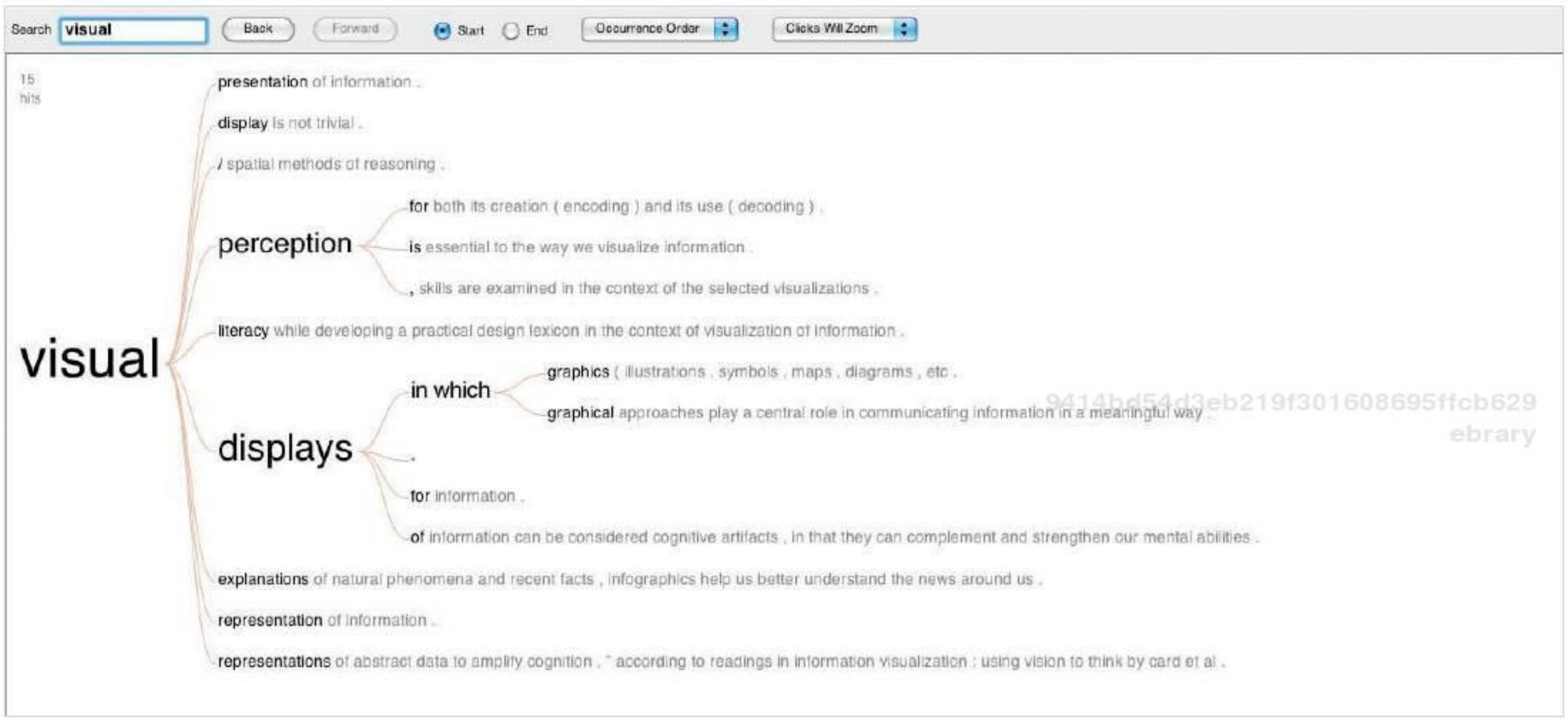
9414bd54d3eb219f301608695ffcb629
ebrary

I used the introduction of this book to generate these Word Tree diagrams. They show the content structure starting (at the top) and ending (at the bottom) with the adjective *visual*.

9414bd54d3eb219f301608695ffcb629
ebrary

Meirelles, Isabel. Design for Information : An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations. Osceola, WI, USA: Rockport Publishers, 2013. ProQuest ebrary. Web. 10 May 2016.
Copyright © 2013. Rockport Publishers. All rights reserved.

9414bd54d3eb219f301608695ffcb629
ebrary



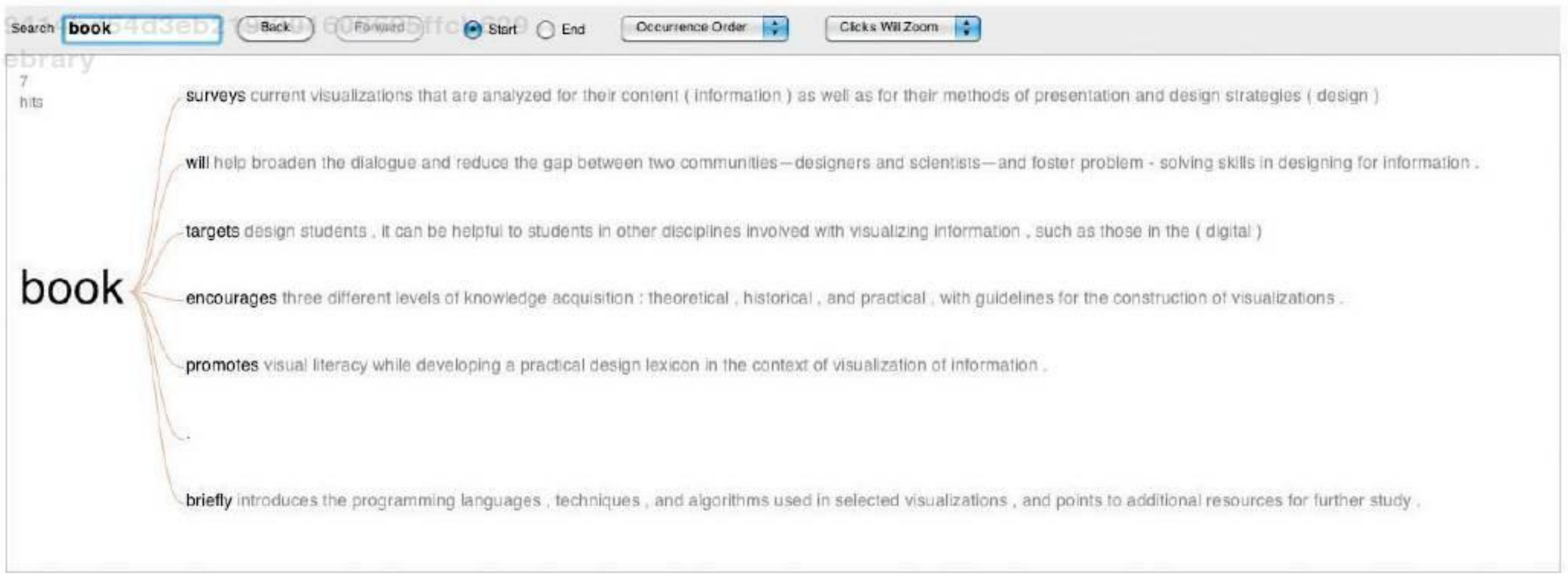
9414bd54d3eb219f301608695ffcb629
ebrary

9414bd54d3eb219f301608695ffcb629
ebrary

Meirelles, Isabel. Design for Information : An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations. Osceola, WI, USA: Rockport Publishers, 2013. ProQuest ebrary. Web. 10 May 2016.
Copyright © 2013. Rockport Publishers. All rights reserved.

9414bd54d3eb219f301608695ffcb629
ebrary

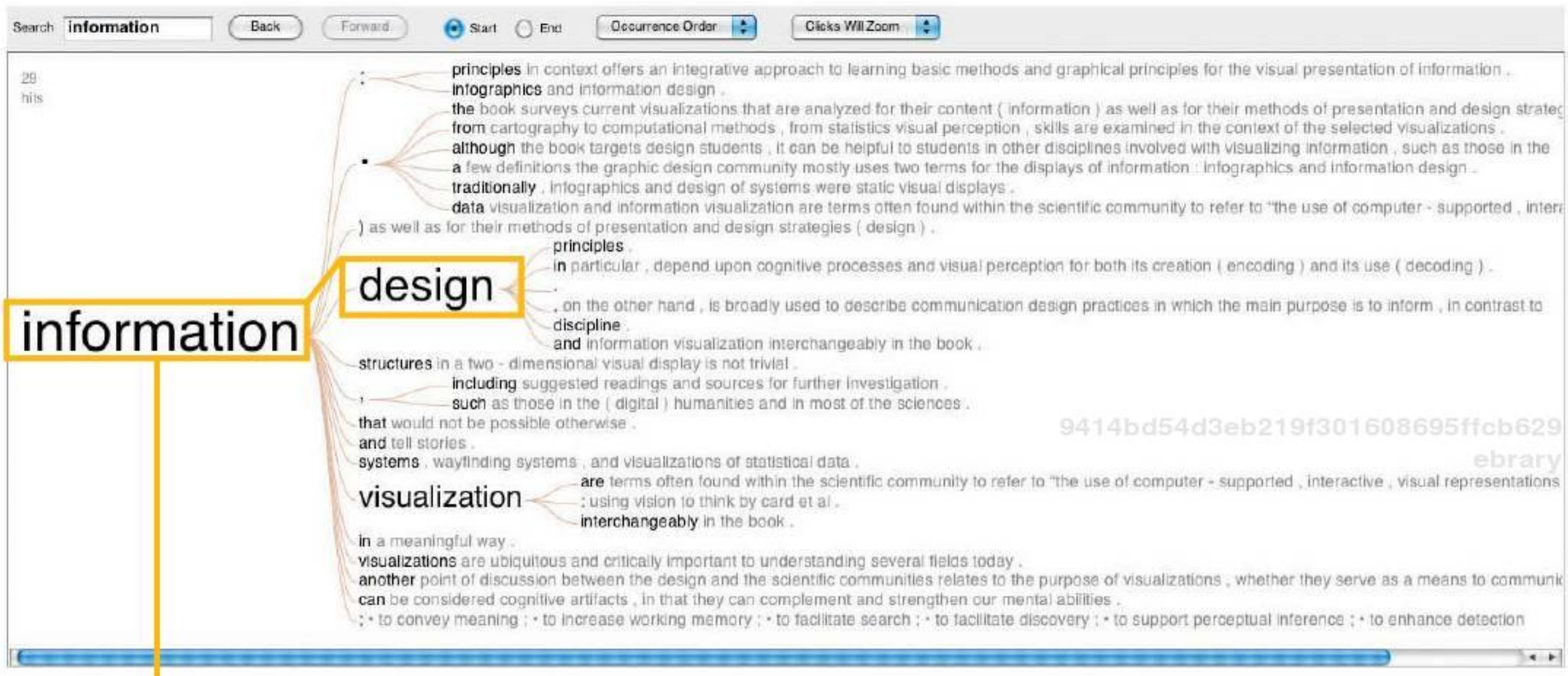
The Word Tree diagrams show content from the introduction of this book. I first searched for the term *information*, and then visualized it at the end (left) and at the beginning of sentences (right). Next, I combined the word *design* to the initial search, and the result is the diagram at the bottom right. The diagram below reveals content structure for occurrences starting with the term *book*.

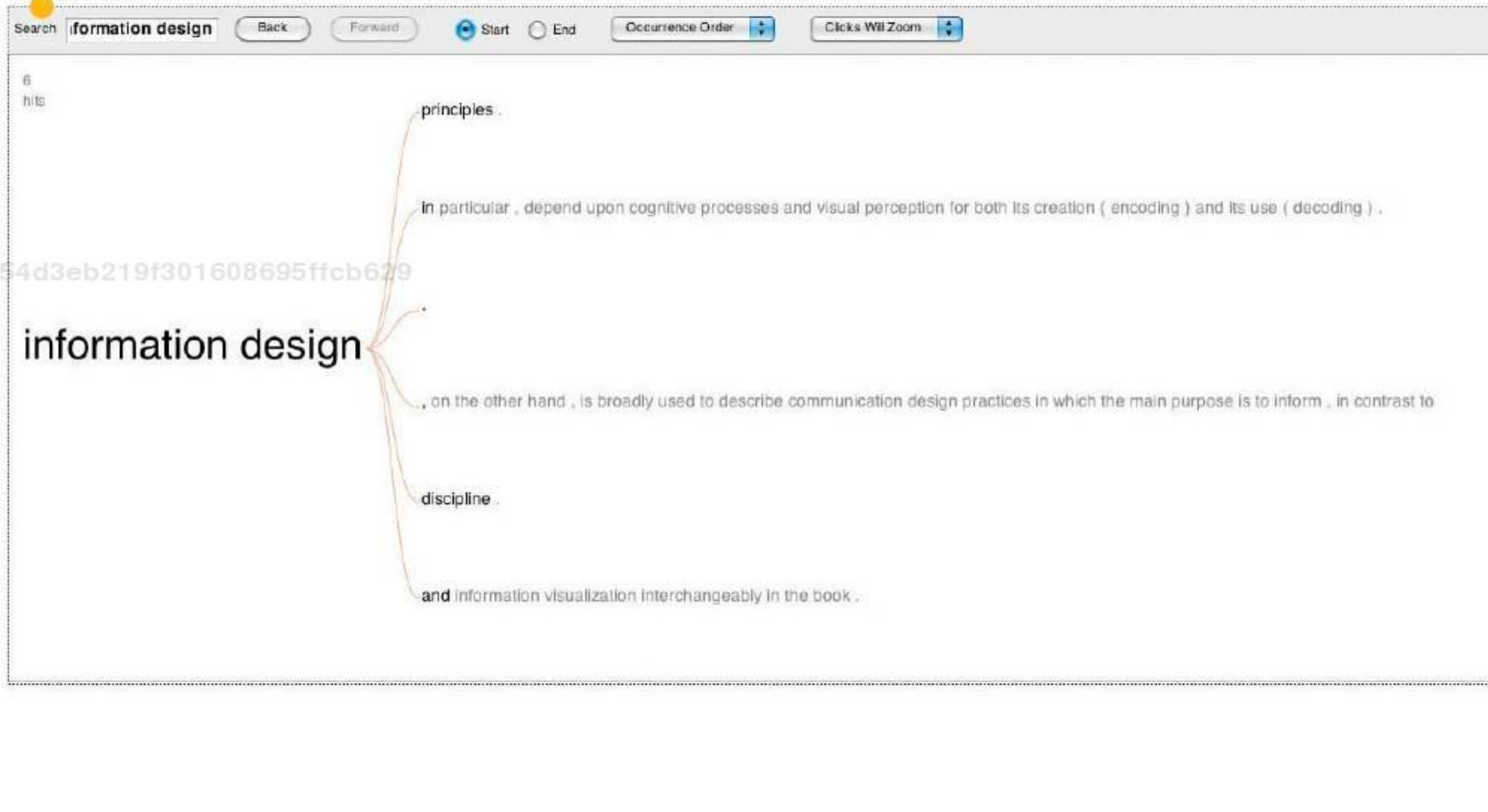


9414bd54d3eb219f301608695ffcb629
ebrary

Meirelles, Isabel. Design for Information : An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations. Osceola, WI, USA: Rockport Publishers, 2013. ProQuest ebrary. Web. 10 May 2016.
Copyright © 2013. Rockport Publishers. All rights reserved.

9414bd54d3eb219f301608695ffcb629
ebrary





9414bd54d3eb219f301608695ffcb629
ebrary

Meirelles, Isabel. Design for Information : An Introduction to the Histories, Theories, and Best Practices Behind Effective Information Visualizations. Osceola, WI, USA: Rockport Publishers, 2013. ProQuest ebrary. Web. 10 May 2016.
Copyright © 2013. Rockport Publishers. All rights reserved.