Chapter 2
Overview of Text Visualization Techniques

Abstract The increasing availability of electronic document archives, such as webpages, online news sites, blogs, and various publications and articles, provides an unprecedented amount of information. This situation introduces a new challenge, which is the discovery of useful knowledge from large document collections effectively without completely going through the details of each document in the collection. Information visualization techniques provide a convenient means to summarize documents in visual forms that allow users to fully understand and memorize data insights. In turn, this process facilitates data comparison and pattern recognition. Many text visualization techniques have been extensively studied and developed for different purposes since the 1990s. In this chapter, we briefly review these techniques to provide an overview of text visualization. Our survey is based on studies summarized in the online text visualization browser (http://textvis.lnu.se/). We classify different text visualization techniques regarding their design goals, which largely group existing techniques into five categories. These categories include techniques developed (1) for visualizing document similarity, (2) for revealing content, (3) for visualizing sentiments and emotions of the text, (4) for exploring document corpus, and (5) for analyzing various domain-specific rich-text corpus, such as social media data, online news, emails, poetry, and prose. Based on this taxonomy, we introduce the details of the primary text visualization research topics in the following chapters of this book.

2.1 Review Scope and Taxonomy

In this book, we have reviewed over 200 papers summarized in the Text Visualization Browser [61] (Fig. 2.1), which was developed by the ISOVIS Research Group from Linnaeus University in Sweden. This online tool provides the most comprehensive and up-to-date summary of text visualization that has been published. This browser enables a user to filter these techniques interactively according to tasks, data, application domain, and visual design styles, which greatly support the writing of this book, especially this chapter.

C. Nan and W. Cui, *Introduction to Text Visualization*, Atlantis Briefs
in Artificial Intelligence 1, DOI 10.2991/978-94-6239-186-4_2

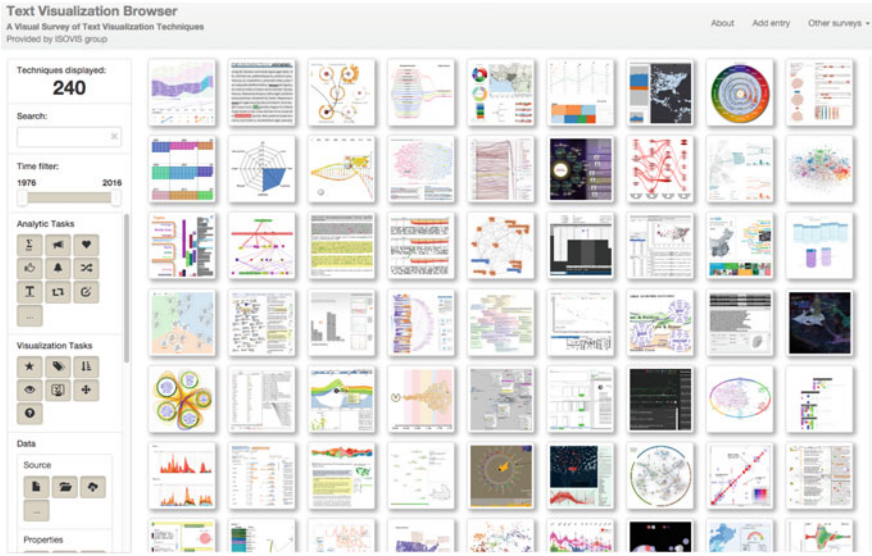

Fig. 2.1 TextVis Browser, an interactive online browser of existing text visualization techniques

Among all the studies listed in the Text Visualization Browser, a set of 120 core text visualization papers were reviewed; these were published in key conferences and journals in three different but highly related areas, namely, visualization, data mining, and human-computer interaction. In particular, our review focuses on the related papers published in (1) visualization-related conferences and journals, such as IEEE International Conference on Information Visualization, IEEE International Conference on Visual Analytics Science and Technology, IEEE Transactions on Visualization and Computer Graphics, IEEE Computer Graphics and Applications, Computer Graphics Forum, and Information Visualization; (2) data mining-related conferences and journals, such as ACM sigKDD Conference on Knowledge Discovery and Data Mining, IEEE Transactions on Knowledge and Data Engineering, SIAM International Conference on Data Mining, IEEE International Conference on Data Engineering, and ACM International Conference on Information and Knowledge Management; (3) human-computer-interaction related conferences and journals including ACM sigCHI Conference on Human Factors in Computing Systems and ACM International Conference on Intelligent User Interfaces.

Text Visualization Browser provides four different ways to categorize existing techniques, i.e., categorizing by task (either analysis or visualization), by data to be visualized, by application domain, and by style of visualization design. However, providing a clear taxonomy with minimum overlap among different technique categories is difficult. To address this issue, we provide a simple yet clear taxonomy based on development goals and purposes of the existing works. In particular, we first separate the techniques into two parts, namely, visualization or interaction technique and various systems developed for different domains by employing these techniques. In particular, in terms of visualization technique, we classify related research into three categories based on goals. These categories include techniques developed for (1) visualizing document similarity, (2) revealing the content of the text data, and (3) showing sentiments and emotions. In this way, we discuss different types of techniques in this book clearly, as well as illustrate their applications and show examples of using these techniques together for solving application problems in different domains.

The rest of this chapter and the book is presented by following this taxonomy. We first briefly review existing techniques and applications in this chapter and describe the detail of major and important techniques in the subsequent chapters.

2.2 Visualizing Document Similarity

Representing content similarities at the document level is one of the most traditional visualization techniques produced for summarizing document collections. These visualizations share a similar representation in which documents are visualized as points on a low-dimensional (2D or 3D) visualization plane. The distance between each pair of points represents the similarities between the corresponding two documents, and follows the rule of the closer, the more similar. Many similar techniques have been extensively studied and have been categorized as either (1) projection-oriented or (2) semantic-oriented.

2.2.1 Projection Oriented Techniques

Projection oriented techniques visualize documents through a dimension reduction procedure. In these techniques, a document is represented as a bag of words and formally described by an N-dimensional feature vector. To compute this vector, a set of most informative words W ($|W| = N$) that best differentiates each document (i.e., best captures the features of different documents) is ranked out from the entire document collection based on "Term Frequency Inverse Document Frequency (TF-IDF)". As a well-known numerical statistic method designed to help extractive word features for document classification [78], the process computes a TF-IDF score for each word in the document collection. The word with a higher score is considered to be more informative, i.e., more useful than other words for classifying different documents. Before computing TF-IDF, stop words are removed and word stems are extracted to ensure the informativeness of each word and the correctness of the word frequency calculation in TF-IDF. Based on these words, an N-dimensional feature vector is produced for a document with each field indicating a word with top-ranking TF-IDF score and the field value indicating its frequency in the given document. Based on this feature vector, projection-oriented techniques visualize a document from the N-dimensional feature space into a 2D or 3D visualization space via a dimension reduction algorithm.

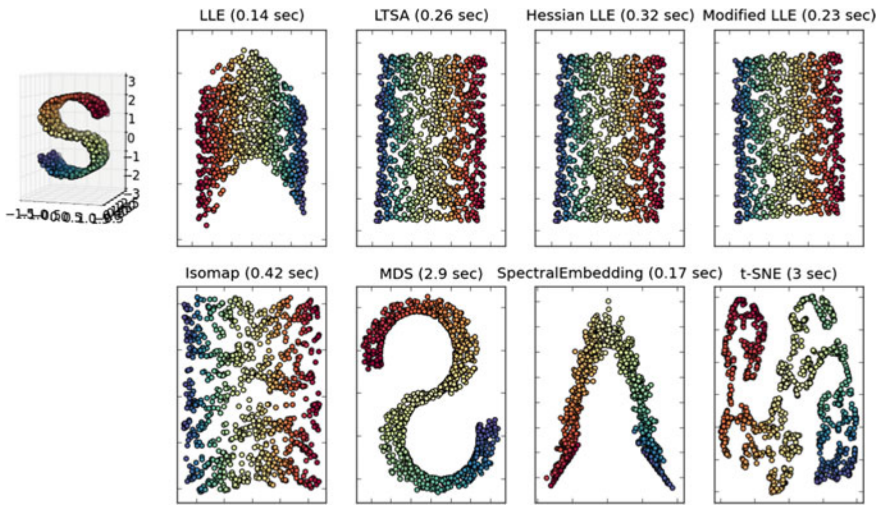

Fig. 2.2 Comparison of different Non-Linear Projection Techniques (also known as manifold learning). The results shown in this figure are is produced based on scikit-lear, which illustrates the results of projecting a "S-shaped" three dimensional dataset (*the left most figure*) onto a 2D plane. The visualization results illustrate that how the original distance in 3D space is preserved in a 2D plane via different algorithms

Generally, the projection oriented techniques can be further separated into (1) linear projections (i.e., linear dimension reduction) and (2) non-linear projections (also known as manifold learning or non-linear dimension reduction). Representative techniques in linear projection include *Principle Component Analysis* (*PCA*) [51] and *Linear Discriminant Analysis* [5]. Both techniques could be formulated in consistent form in which pair-wise distances between data items are maximized and guided by weights that indicate the importance of separating pairs of points in the results [58]. These techniques, although effective in terms of computation, usually fail to capture data similarities when they are non-linear. Therefore, many non-linear projection techniques have been developed and extensively studied. Existing methods could be further classified into (1) distance-oriented techniques and (2) probabilistic formulation-based techniques. A comparison of different techniques is shown in Fig. 2.2.

Specifically, the distance oriented non-linear projection techniques such as *Multidimensional Scaling* (*MDS*) [60] and *Locally Linear embedding* (*LLE*) [92] introduce different methods to preserve the distances in the high-dimensional feature space in a low (2D or 3D) dimensional visualization space. The probabilistic formulation based techniques such as *Stochastic Neighbor Embedding* (*SNE*) [45] and *t-Distributed Stochastic Neighbor Embedding* (*t-SNE*) [73] formulate document similarity via statistical models, in which the similarity between two documents i and j is captured by the conditional probability of $P_{(j|i)}$, i.e., given that document i the probability of document j are in the neighborhood of i in the feature space. Compared with distance oriented techniques, these probabilistic based approaches can be more effectively computed and can produce results with improved quality [73].

2.2.2 Semantic Oriented Techniques

These approaches represent document similarity via latent topics extracted from text data. Studies in this direction are mainly inspired by topic modeling techniques such as *Probabilistic Latent Semantic Analysis* (*PLSA*) [46], *Latent Dirichlet Allocation* (*LDA*) [6], *Spherical Topic Model* (*SAM*) [87], and *Non-Negative Matrix Factorization* (*NMF*) [64]. Although widely used for analysis, these topic modeling techniques are not designed for visualization purpose, thereby directly showing the analysis results is usually difficult for users to interpret. For example, PLSA and LDA analyze topics in a simplex space which is shown as a triangle on the 2D Euclidean visualization plane. Therefore, these methods cannot display more than three topics a time (Fig. 2.3d). Semantic oriented techniques have proposed to produce improved visual representations. These techniques were pioneered by *Probabilistic Latent Semantic Visualization* (*PLSV*) [47] which embeds the latent topics and docu-

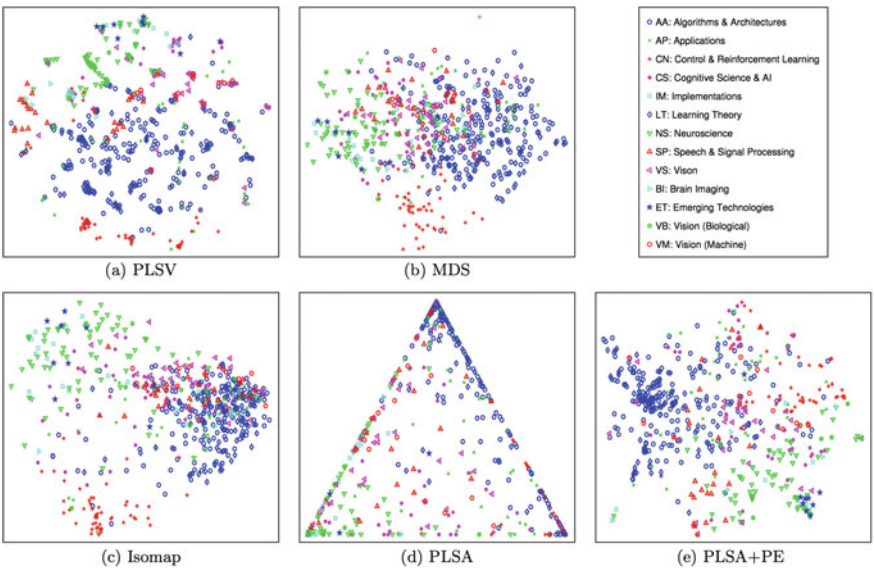

Fig. 2.3 Visualizing document similarity based on semantic oriented techniques (**a**, **d**, **e**) and non-linear projection (**b**, **c**). Each *point* in the diagram is a document colored by their primary topics that are extracted based on topic modeling. The distances between documents encode their similarities, following the rule of "the more similar, the closer". This figure represent visualizations of the same data, i.e., papers published in NIPS, which was first published in [47]

ments in the generic Euclidean space at the same time as the distances directly encode similarities among documents, regarding their shared topics (Fig. 2.3a). Following this work, other techniques such as Spherical Semantic Embedding (SSE) [63] were also developed, which provides improved approximations of similarities among documents. When compared to the projection oriented techniques (Fig. 2.3b, c), semantic based approaches usually produce more meaningful results that are easier for users to understand.

2.3 Revealing Text Content

Visually representing the content of a text document is one of the most important tasks in the filed of text visualization. Specifically, visualization techniques have been developed to show the content of documents from different aspects and at different levels of details, including summarizing a single document, showing the words and topics, detecting events, and creating storylines.

2.3.1 Summarizing a Single Document

Existing visualization techniques summarize a document through two main aspects: (1) content such as words and figures and (2) features such as average sentence length and number of verbs.

In terms of showing the content of a document, Collins et al. [19] introduce DocBurst which decompose a document into a tree via its innate structures such as sections, paragraphs, and sentences that are illustrated in a SunBurst visualization [97] (Fig. 2.4). Rusu et al. [93] visualize the content of a document via a node-link diagram based on a semantic graph extracted from the document. Strobelt et al. [99] introduce a system that transforms a document into cards, in which the content of the document is summarized via keywords and critical figures that are extracted from the document (Fig. 2.5). Stoffel et al. [98] propose a technique for producing the thumbnail of a document based on keyword distortion. This technique produces a focus+context representation of document at the page level. In particular, on each page of the document, important words are shown in a larger font whereas the rest ones are suppressed as the context, thereby compressing the entire page into a small thumbnail without losing the key information of each document page.

Despite the aforementioned studies, document fingerprint [50, 55, 80] is another typical visualization technique that is developed to summarize a single document. Instead of showing words and figures, this technique captures the key features of a document from multiple aspects through a heatmap visualization in which each cell represents a text block (e.g., a paragraph or a sentence) with color showing its feature value (Fig. 2.6). A set of linguistic features were used to measure the document from different aspects, as summarized in [55], including (1) statistical

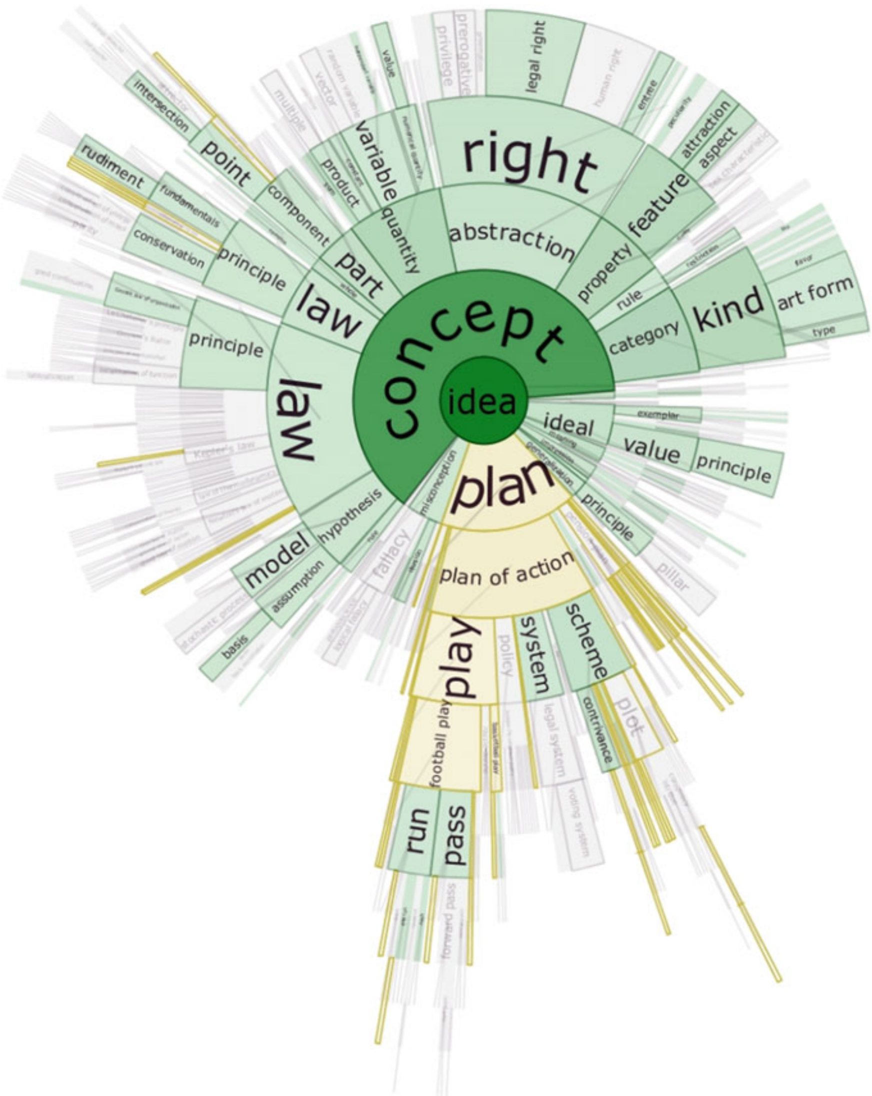

Fig. 2.4 DocuBurst visualization of a science textbook rooted at idea. A search query for words starting with pl has been performed. Nodes matching the query are highlighted in *gold*

features such as average word length, average number of syllables per word and average sentence length, (2) vocabulary features such as the frequencies of specific words and vocabulary richness measured by Simpson Index, and (3) syntax features computed based on syntax trees [54].

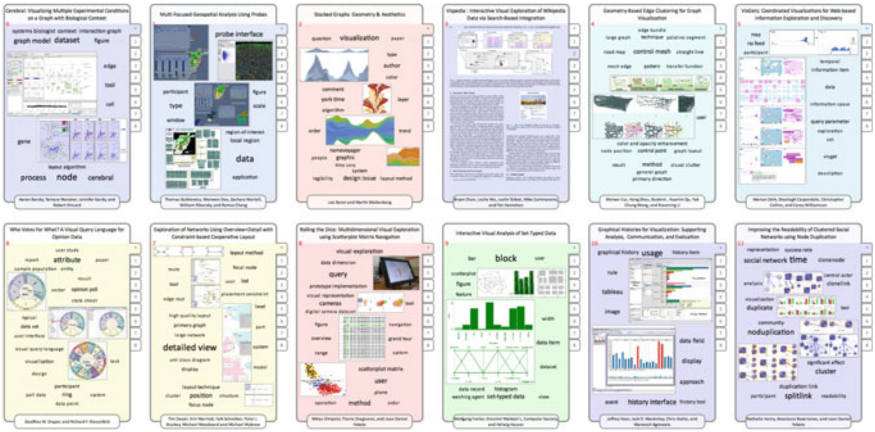

Fig. 2.5 Summarization of the IEEE InfoVis 2008 proceedings corpus in Document Cards (a portion). Referring to [98] for the complete visual summarization of the whole proceeding

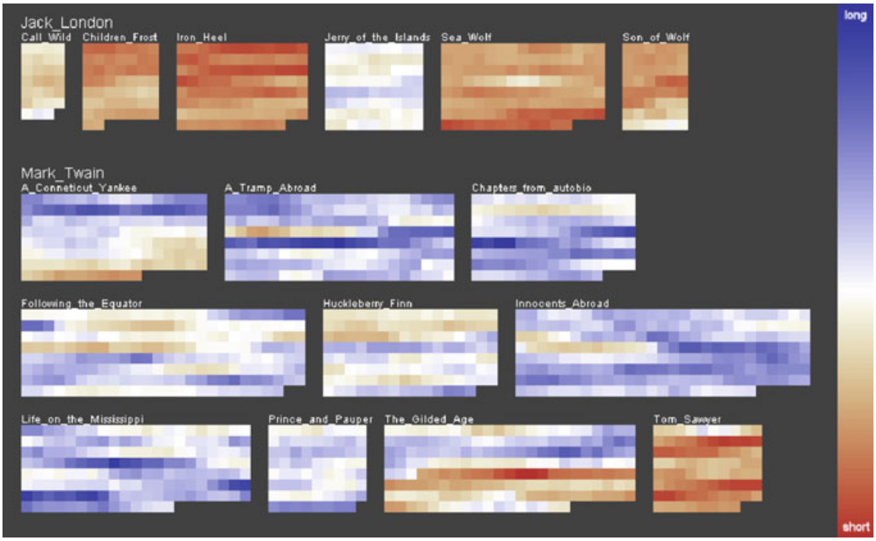

Fig. 2.6 The literature fingerprint visualization showing the feature "average sentence length" of books written by Jack London and Mark Twain

2.3.2 *Showing Content at the Word Level*

Directly illustrating the keywords of a document is the most intuitive approach to present document content. Existing visualization techniques in this category are largely developed to address three general problems: (1) how to represent the words esthetically in a visual form to clearly depict the content of the text; (2) how to

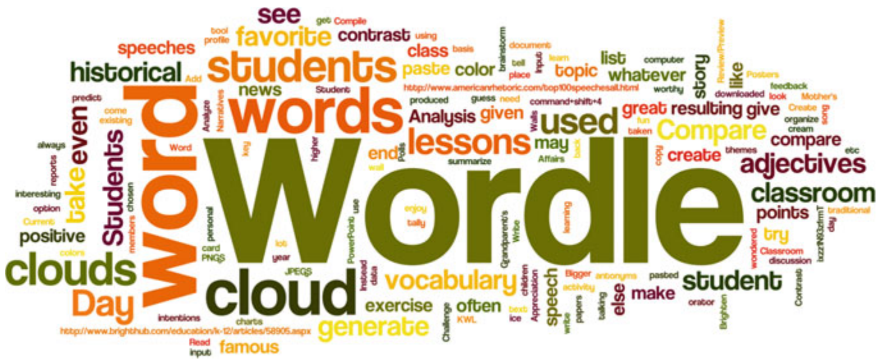

Fig. 2.7 Wordle visualization of a bag of words extracted from text data

summarize and represent the semantic relationships such as "A is B" and "A of B" between words in the text, and (3) how to reveal word-level patterns such as repetitions and co-occurrences.

TagCloud [53] is one of the most intuitive and commonly used techniques for visualizing words. It illustrates a bag of words that summarize the content of the input text data in a cloud form, in which words, with font size indicating their importance, are packed together without any overlap. Traditional TagCloud aligns words line by line, which is most commonly used in webpages to show the content of, for example, the current web. Different packing strategies will result in various types of TagClouds [13, 14, 20, 35, 104, 114], among which Wordle is the state-of-the-art technique that produces aesthetic word packing results by precisely calculating the word boundary and randomly inserting the word into empty spaces guiding by a spiral line (Fig. 2.7). Despite these static techniques, dynamic word clouds [24] are also developed for showing the changes of the text content of a streaming corpus such as Twitter and publication dataset over time.

Although widely used, TagClouds fail to uncover the word relationship as usually the words are randomly placed. Therefore, many tree or graph based visualization techniques are introduced. For example, WordTree [109] (Fig. 2.8) summarizes text data via a syntax tree in which sentences are aggregated by their sharing words and split into branches at a place where the corresponding words in the sentences are divergent. Another example is the PhraseNet [103] (Fig. 2.9). This visualization employs a node-link diagram, in which graph nodes are keywords and links represent relationships among keywords which are determined by a regular expression indicated by users. For example, as shown in Fig. 2.9, a user can select a predefined regular expression from a list to extract a relationship such as "X and / is / of Y" or by inputting the regular expression by their own.

Despite relationships, visualizations are also developed to illustrate highly detailed patterns such as word co-occurrences and repetitions [4, 49, 108]. For example, Wattenberg introduced the Arc diagram (Fig. 2.10), which is one of the earliest visu-

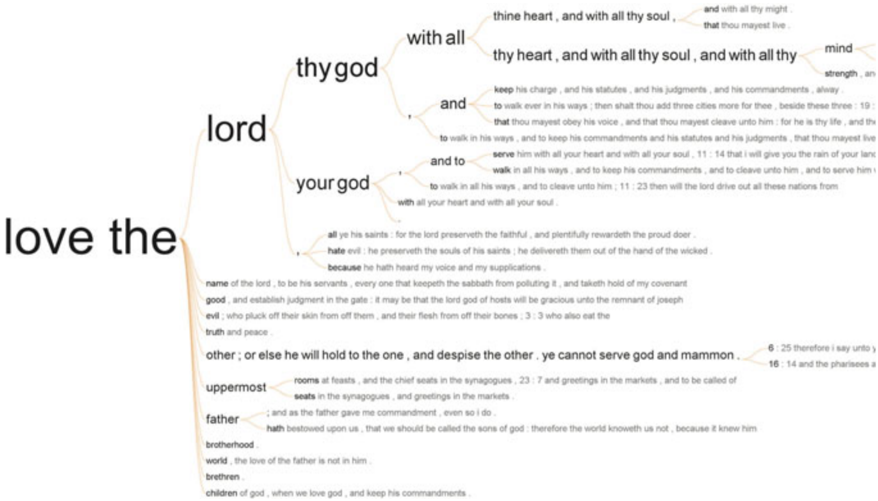

Fig. 2.8 Word tree of the King James Bible showing all occurrences of *love the*

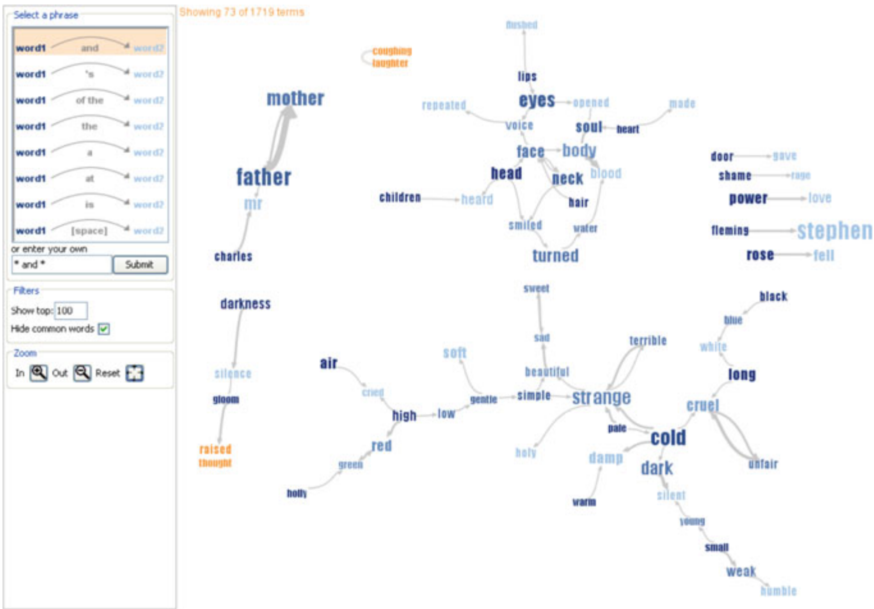

Fig. 2.9 The Phrase Net user interface applied to James Joyce Portrait of the Artist as a Young Man. The user can select a predefined pattern from the list of patterns on the left or define a custom pattern in the box below. This list of patterns simultaneously serves as a legend, a list of presets and an interactive training mechanism for regular expressions. Here the user has selected X and Y, revealing two main clusters, one almost exclusively consisting of adjectives, the other of verbs and nouns. The highlighted clusters of terms have been aggregated by our edge compression algorithm [103]

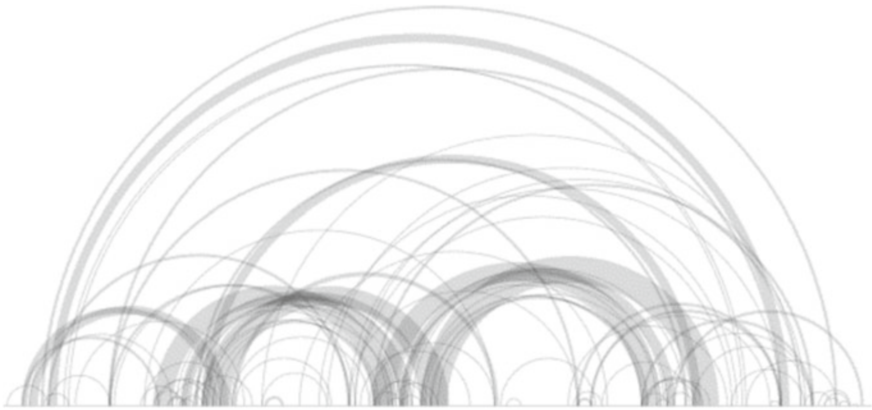

Fig. 2.10 The arc diagram visualization of a HTML webpage in which repeated strings (e.g., words or sentences) are connected by arcs

alizations designed to illustrate repetition patterns. It represents an input string in a raw and connects the repeated symbols or words via arcs, thereby illustrating a clear visual pattern when repetition occurs.

2.3.3 Visualizing Topics

Accompanying the rapid development of topic analysis techniques, visualizing topics has become a highly interesting research direction in the field of text visualization in recent years. Compared with word-level visualizations, showing topics helps to capture more semantics of the data, thereby producing text visualizations that are easier to interpret. A growing number of visualization techniques are developed to (1) summarize and explore static topic information, (2) illustrate the topic dynamics over time, (3) help with topic comparison, and (4) illustrate events and storylines.

Even before the invention of modern topic modeling techniques such as PLSA [46] and LDA [6], visualization systems such as Topic Island™ [76] and IN-SPIRE™[1] had been introduced to illustrate and explore static topic themes extracted from text data. Research in this direction is significantly accelerated by the development of topic analysis techniques. Many attempts have been made to represent the topic analysis results. In particular, Cao et al. introduced ContexTour [67] (Fig. 2.11), FacetAtals [17], and SolarMap [15] in a row based on a so-called "multifaceted entity-relational data model". In particular, they decompose the text corpus into this data model based on a series of text analysis approached including (1) topic analysis, (2) name entity identification, and (3) word co-occurrence detection. The resulting visualization illustrates static topics and their corresponding relationships from

[1] http://in-spire.pnnl.gov/.

Fig. 2.11 Visualizing research topics in a publication dataset of papers published in the computer science conferences and journals in 2005 in ContexTour, in which the background contour produces a density map based on kernel density estimation, showing the underlying distribution of the words. Topics are shown as TagCoulds on *top* of the contour map

multiple information facets. Following these works, many similar techniques such as VisClustering [65] and TopicPanoram [69] were also developed. When compared to the aforementioned techniques, these works employed similar visual designs and provided similar functions in terms of topic representation and exploration, but were developed to focus on different analysis tasks (Fig. 2.12).

Capturing the topic dynamics is another research direction that attracts great attention in the field of text visualization. In particular, ThemeRiever [42] is one of the earliest techniques developed to show how the frequency of the keywords are changed over time. It visualizes a set of keywords as stripes (i.e., themes) whose thicknesses change over time, indicating the change of frequencies of the corresponding keywords, which are shown on top of the stripe. This design was later extended by Liu

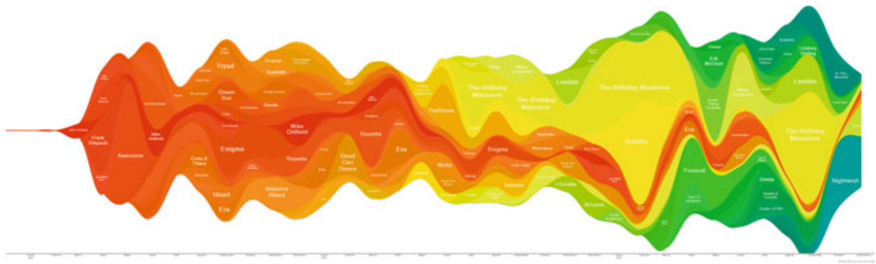

Fig. 2.12 ThemeRiver [42]: keywords in a document collection are shown as *colored* "*stripes*", i.e. themes, with width indicating the occurrence frequency of keywords at different times

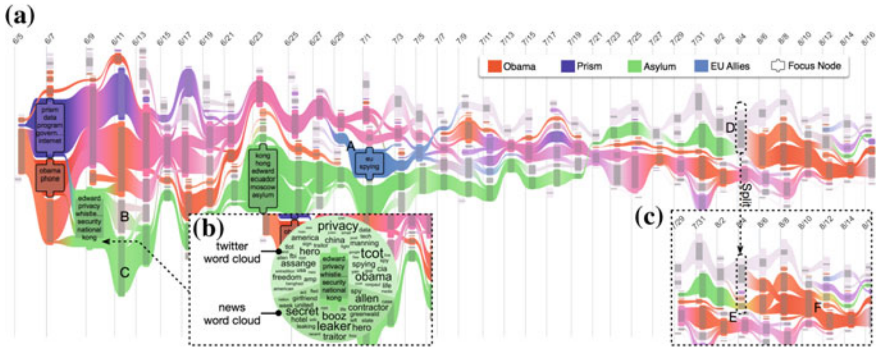

Fig. 2.13 RoseRiver, a visual analytics system for exploring evolutionary hierarchical topics. **a** Overview of the Prism scandal (June 5–Aug 16, 2013). Four *colors* represent the four major topics. Topics are displayed as *vertical bars*. The *color stripes* represent the evolving relationships between topics. **b** Comparison of the prominent keywords in tweets and news articles of the topic. The arc lengths encode the news article and tweet numbers (in log scale). **c** The new layout generated by splitting the gray topic

et al. [71] by using strips to represent changing topics, on top of which topic keywords are visualized as TagClouds. Following this research direction, Cui et al. [22] introduce TextFlow, the state-of-the-art design that captures the dynamics of topic evolution patterns such as splitting and merging over time. Subsequently, the authors refined their original design and algorithm to produce a hierarchical topic flow [23] that supports different levels of detail. This technique precisely captures the topic evolution patterns as shown in Fig. 2.13. Despite showing the overview of topic evolution trend, visualization techniques were also introduced to capture other temporal patterns in text data. For example, Gad et al. [34] introduced ThemeDelta, which integrates with topic modeling algorithms to identify change points when significant shifts in topics occurred. Liu et al. [68] introduced a technique for exploring topic lead-lag, i.e. a pattern illustrating how a former topic result in the occurrence of a latter topic

When multiple document collections are visualized together at the same time, a spontaneous analysis task is to compare to find their common and distinct topics. To this end, many visualization techniques are introduced. Diakopoulos et al. [25] develop Compare Clouds, which is a TagCloud visualization designed to compare the topic keywords of two sets of input documents. Oelke et al. [81] also introduce a visual analysis system for comparing and distinguishing different document collections. In particular, it detects discriminative and common topics and visualizes each topic in a circular glyph, in which topic keywords are shown as a TagCloud. Glyphs are laid out based on topic similarities, i.e., the glyphs of similar topics are placed close to each other whereas the dissimilar ones are separated apart. In this way, common topics are placed at the center of the view and the discriminative ones are clearly separated into different topic regions, thereby forming a visualization that facilitates topic differentiations and comparisons.

2.3.4 Showing Events and Storyline

Finding topics in a collection of documents is generally a clustering approach. Documents that have similar contents (essentially using similar words) are clustered together to constitute a topic or a theme. By contrast, event analysis focuses on a different type of information that has time and space as the primary attributes [48]. An event is generally considered as an occurrence at a given space-time that is perceived by an observer to have a beginning and an end. Human beings are gifted with the ability to perceive and make sense of real-world activities as consisting of discrete events with orderly relations [116]. Thus, when visualizing events, researchers more focus on understanding the four Ws that characterize these events: who, what, when, and where.

A large amount of structured or semi-structured textual data explicitly contain event information, such as crime incidents or accidents recorded by police departments [11, 66], patient records [21, 32, 39, 86, 111–113], and customer purchase logs [12]. For these datasets, the primary goal of visualization is to provide visual summaries and to support efficient queries. For example, as an early work, LifeLines [86] provides a general visualization that allows users to explore details of a patient's clinical records (Fig. 2.14). A subsequent version, LifeLines2 [106] enhances the visualization and introduces three general operators, namely, align, rank, and filter, for interactive exploration of categorical, health-related datasets. PatternFinder [32] is designed to help users visually formulate queries and find temporal event patterns in medical record datasets. LifeFlow [112] and Outflow [111] aggregate multiple event sequences into tree or graph visualizations and provide users with highly scalable overviews (Fig. 2.15).

In real-world situations, people also often segment activities into events at multiple timescales [62]. Small events are grouped to constitute a large and complex event, such as acts in a play. In addition, events may also share elements, such as participants

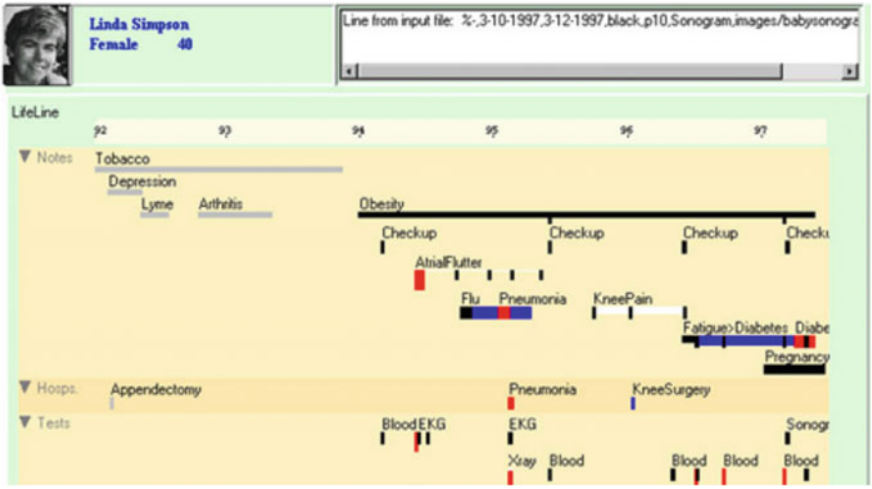

Fig. 2.14 Screenshot of LifeLines [86]: *colored horizontal bars* show the time of occurrence and duration of clinical events for a patient, such as medical incidents, treatments, and rehabilitation. Additional information is encoded by the height and color of individual bars. Multiple facets of the records, such as notes and tests, are stacked vertically, and can be expanded and collapsed as needed

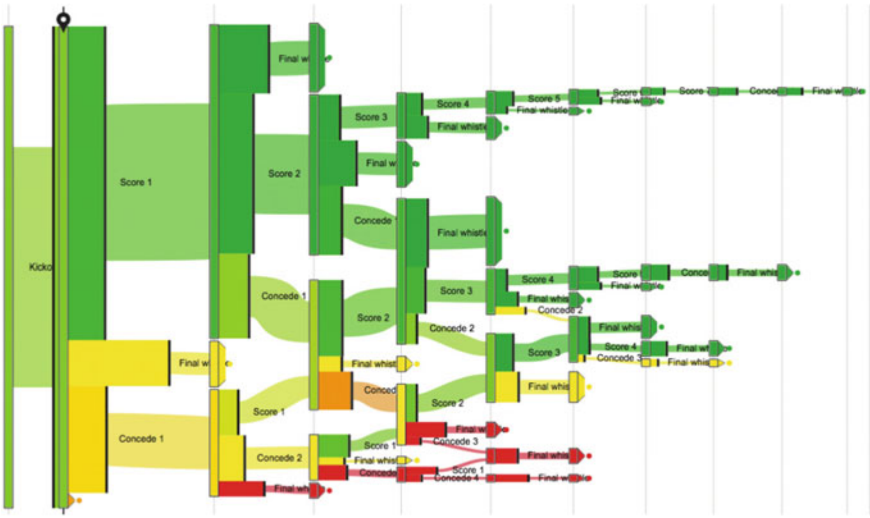

Fig. 2.15 Screenshot of Outflow [111] visualization that shows the scores of Manchester United in the 2010–2011 season. *Green* indicates pathways of winning, while *red* shows pathways of losing

and locations, with one another. To represent relationships between events, Burch et al. [12] use a horizontally oriented tree layout to represent the hierarchical relationships of transaction sequences along the timeline. André et al. [2] present Continuum to visually organize large amounts of hierarchical events and their relationships.

Recently, significant research has been conducted to extract and explore events in unstructured textual data, such as news articles [28, 95] and microblogs [26, 75]. Since event information in these data is generally implicit, it requires text mining techniques, such as topic detection and tracking, are necessary to extract events for further visualization. For example, EventRiver [72] presents events in a river-metaphor based on event-based text analysis. In EventRiver visualization, an event is represented by a bubble that floats on a horizontal river of time. The shape of the bubble indicates the intensity and duration of the corresponding event. The color and vertical position of the bubble are used to indicate relationships among different events. Krstajic et al. [59] propose a visualization technique that incrementally detects clusters and events from multiple time series. To explore events in microblogs, Marcus et al. [75] describe a visual analytics system that allows users to specify a keyword of interest, and then visually summarizes events related to the query. DÖrk et al. [26] also propose an approach called visual backchannel that integrates text, images, and authors extracted from Twitter data. Their system does not require keywords of interest. Instead, it monitors evolving online discussions on major events, such as political speeches, natural disasters, and sport games. Twitter posts are summarized as a temporally adjusted stacked graph. Related authors and images are also visualized as a People Spiral and an Image Cloud, respectively, to help users track the event evolutions. LeadLine (Fig. 2.16) combines topic analysis and event detection techniques to extract events from social media data streams. Various information, such as topic, person, location, and time, is identified to help users reason about the topic evolutions.

Recently, storyline visualization has emerged and attracted significant attention. Unlike events and topics that focus on activities and themes, storyline visualizations switch the focus to entities and relationships between entities. In a typical storyline visualization, x-axis represents time, and an entity is represented as a line that extends

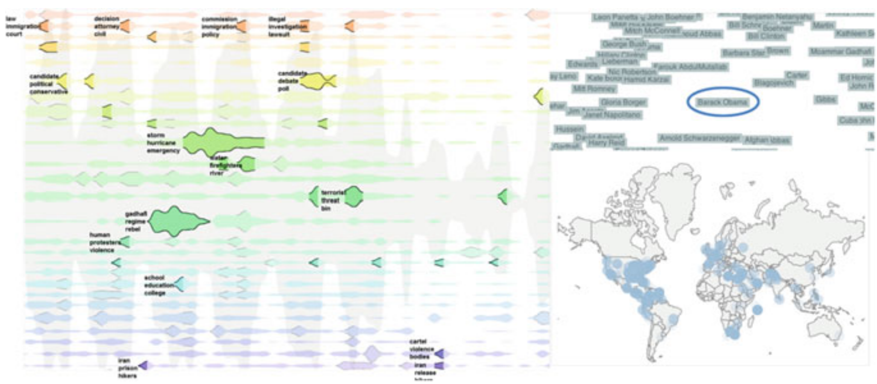

Fig. 2.16 Screenshot of LeadLine visualization that summarizes CNN news stories from Aug 15, 2011 to Nov 5, 2011. *Bottom right* locations that are related to President Obama are marked on the map. *Left* events that are related to the president are highlighted by *color-coded bursts*

horizontally from left to right. The relationships between entities, which may change over time, are encoded by the vertical distances between the lines. Figure 1.1 shows an example of such visualizations. The figure shows the main story in the book of *Lord of the Rings*. Although this chart is manually made by the author, it inspires a set of approaches that aim to automatically generate similar visualizations. For example, Ogievetsky [83] builds an online tool to allow users to interactively generate and adjust a storyline visualization. However, to find a visually satisfying arrangement of lines is the key issue for this visualization. To solve this problem, Ogawa and Ma [82] raise several criteria and propose a rule-based algorithm to automatically generated a storyline-like visualization to help experts track software evolution (Fig. 2.18). Later, Tanahashi and Ma [101] further formulate the storyline layout process as an optimization problem and solve the problem with a genetic method. Although time-consuming, their algorithm can generate results that are comparable to those handmade by XKCD. Based on their work, another optimization process is proposed by Liu et al. [70]; This process is time efficient enough to support real-time interaction and still be able to maintain the same level of aesthetics (Fig. 2.19). Recently, Tanahashi et al. [100] have extended storyline visualizations to streaming data, and further provide users with the ability to follow and reason dynamic data.

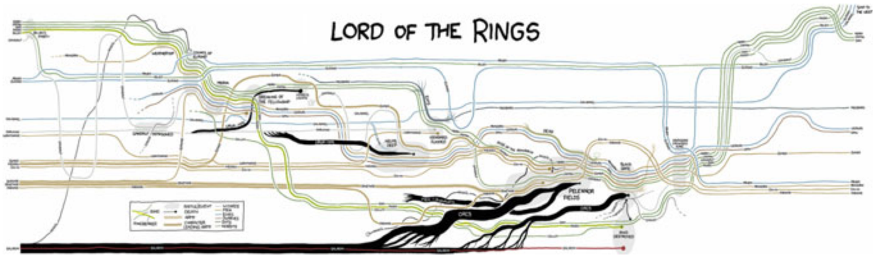

Fig. 2.17 XKCD's movie narrative chart of *Lord of the Rings* [77]

Fig. 2.18 Storyline that shows the development of Python source codes [82]

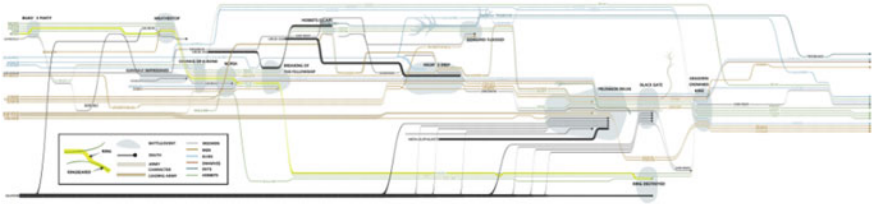

Fig. 2.19 Reproduction of the same chat in Fig. 2.17 using StoryFlow algorithm [82]

2.4 Visualizing Sentiments and Emotions

Many visualization techniques have been developed to illustrate the change of sentiments over time regarding to a given streaming text corpus such as news corpus, review comments, and Twitter streams. This goal can be achieved, as shown in Fig. 2.20, by showing the sentiment dynamics in a time-series diagram, in which the time-series curve illustrates the change of sentiment scores computed across the entire dataset at different time points. However this simple visualization is too abstract to display information details such as the causes behind the sentiment shifts. Therefore, many other more advanced techniques have been introduced to illustrate and interpret the sentiment dynamics from different perspectives.

Most techniques are developed to compute and visualize the sentiments of a focal group of people based on the text data produced by them. The resulting visualization forms a "happiness indicator" that captures the sentiment change of the focal group over time. For example, Brew et al. [9] introduce SentireCrowds, which represents the sentiment changes of a group of twitter users from the same city in a timeline view

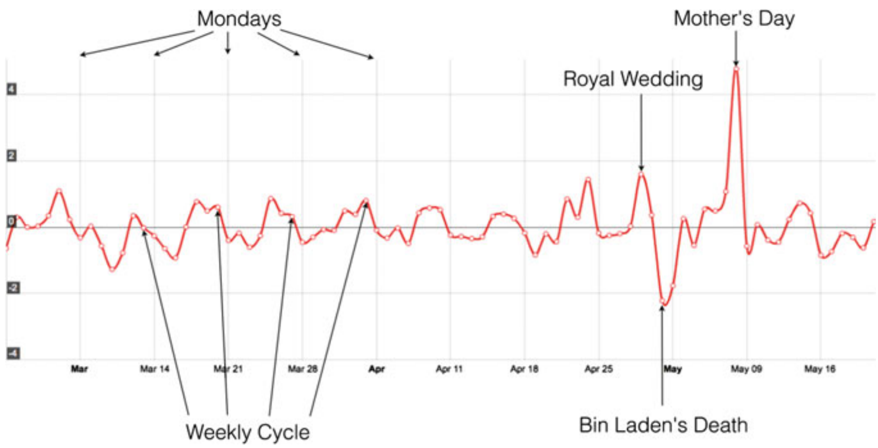

Fig. 2.20 Sentiment Indexing of Twitter data in a time-series diagram. This figure shows that the public sentiment may change dramatically regarding to different events in our real-life

and summarizes the potential underlying event that causes the changes in a multi-level TagCloud designed based on Treemap. Zhao et al. [118] introduce PEARL, which visualizes the change of a person's emotion or mood profile derived from his tweets in a compound belt visualization. The belt groups a set of emotion bands, each indicating a type of emotion differentiated by colors. The thickness of the band changes over time, representing the portion of the corresponding emotion at different time. Guzman et al. [40] visualize the change of emotions of groups of different developers in various software development projects. Hao et al. [41] analyze sentiments via geo-temporal term associations based on a streaming dataset of customer's feedback. Kempter et al. [57] introduce a fine-grained, multi-category emotion model to classify users' emotional reactions to public events overtime and to visualize the results in a radar diagram, called EmotionWatch, as shown in Fig. 2.22.

Despite the preceding visualizations, some visual analysis systems have also been developed to assist with dynamic sentiment analysis. For example, Wanner et al. [107] develop a small multiple visualization view to conduct a semi-automatic sentiment analysis of large news feeds. In this work, a case study on news regarding the US presidential election in 2008 shows how visualization techniques will help analysts draw meaningful conclusions without exerting effort to read the content of the news. Brooks et al. [10] introduce Agave, a collaborative visual analysis system for exploring events and sentiment over time in large Twitter datasets. The system employs multiple co-ordinated views in which a streamgraph (Fig. 2.21) is used to summarize the changes of the sentiments of a subset of tweets queried based on users' preferences. Zhang et al. [117] introduce a spacial-temporal view for visualizing the sentiment scores of micro-blog data based on an electron cloud model intruded in physics. The resulting visualization maps a single sentiment score to a position inside a circular visualization display.

More sophisticated systems are also developed to analyze the change of sentiments based on streaming text data. For example, Rohrdantz et al. [91] introduce a visual analysis system to help users to detect interesting portions of text streams,

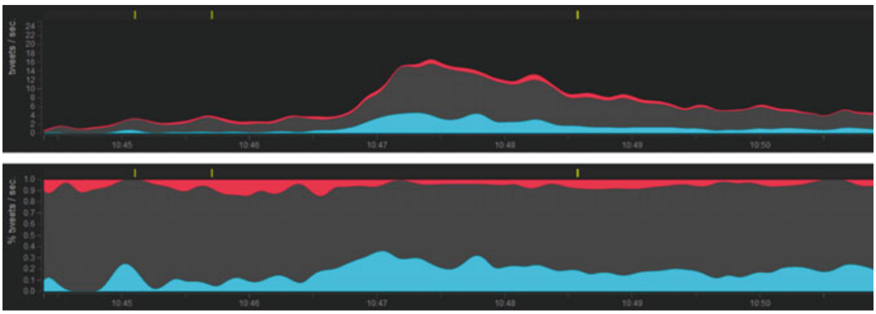

Fig. 2.21 Sentiment streamgraphs for the keyword search Flacco, the Super Bowl MVP in a Twitter dataset using Agave [10]. Negative is *red*, neutral is *gray*, and positive is *blue*. *Top* overall frequency of tweets, divided by sentiment type. *Bottom* sentiment as percent of overall volume

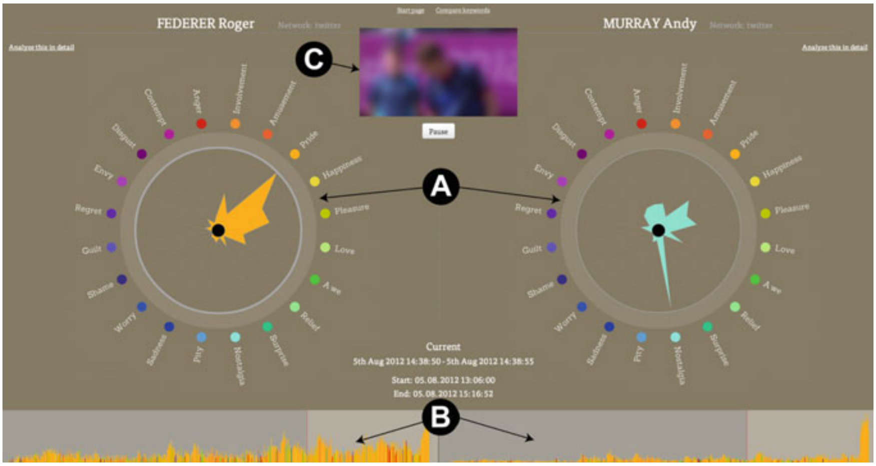

Fig. 2.22 Comparison of two emotion profiles of Roger Federer and Andy Murray (two tennis athletes) after a tennis game in EmotionWatch [57]; (*A*) the EmotionWatches, (*B*) timelines showing the two emotion flows, and (*C*) video

regarding to the change of sentiments, data density, and context coherence based on a set of features extracted from the text. Wang et al. [105] introduce SentiView, which employs advanced sentiment analysis techniques as well as visualization designs to analyze the change of public sentiments regarding popular topics on the Internet. Other systems are designed for analyzing sentiment divergences (i.e., conflicting of opinions) that occur between two groups of people. For example, Chen et al. [18] introduce the first work in this topic based on a simple time-series design that summarizes the overall conflicting opinions based on the Amazon review data. Following this topic, Cao et al. [16] introduce a more advanced technique called SocialHelix, which extracts two groups of people having the most significant sentiment divergence over time from Twitter data and illustrates their divergence in a Helix visualization to show how the divergence occurred, evolved, and terminated.

In terms of application, a large set of such techniques are developed to represent the customer's sentiments based on the review data. Alper et al. [1] introduce OpinionBlocks, an interactive visualization tool to improve people's understanding of customer reviews. The visualization progressively discloses text information at different granularities from the keywords to the phrases in which the keywords are used, and to the reviews containing the phrases. The information is displayed within two horizontal regions, representing two types (positive and negative) of different sentiments. Gamon et al. [36] introduce Pulse for mining topics and sentiment orientation jointly from free text customer feedback. This system enables the exploration of large quantities of customer review data and was used for visually analyzing a database for car reviews. Through this system, users can examine customer opinion at a glance or explore the data at a finer level of detail. Oelke et al. [79] analyze to determine customers' positive and negative opinions through the comments or rat-

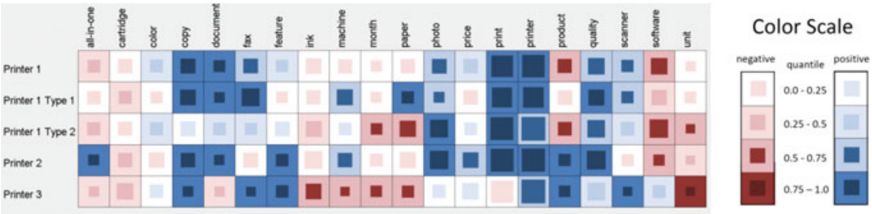

Fig. 2.23 Summary Report of printers: each *row* shows the attribute performances of a specific printer. *Blue color* represents comparatively positive user opinions and *red color* comparatively negative ones (see *color scale*). The size of an *inner rectangle* indicates the amount of customers that commented on an attribute. The larger the rectangle the more comments have been provided by the customers

ings posted by the customers. This system visualize the analysis results in a heatmap view showing both volume of comments and the summarized sentiments (Fig. 2.23). Wu et al. [115] introduce OpinionSeer, which employs subjective logic [52] to analyze customer opinions on hotel rooms based on their review data inside a simplex space, which is visualized in a triangle surrounded by context about the customers such as their ages and their countries of origin. More generic systems are also developed. For example, Wensel [110] introduce VIBES, which extracts the important topics from a blog, and measures the emotions associated with those topics that are illustrated through a range of different visualization views. Makki et al. [74] introduce an interactive visualization to engage the user in the process of polarity assignment to improve the quality of the generated lexicon used for sentiment or emotion analysis via minimal user effort.

2.5 Document Exploration Techniques

With a large document collection, how to effectively explore the data to find useful information or insightful data patterns is always a challenge that attracts many research attentions. Many visualization systems are designed to support an effective exploration of big text corpus. Many studies are focused on inventing or improving the text data exploration techniques. A large category of them are query-based systems in which full text indices are built so that users can query to retrieve data based on their interests. Based on these techniques, many systems are developed to explore text collected from various application domains. In this section, we review these exploration techniques as well as their applications.

Fig. 2.24 Data Mountain visualization shows 100 webpages

2.5.1 Distortion Based Approaches

As early as the 1990s, some distortion based techniques have been developed to assist in text data exploration. For example, Robertson and Mackinlay introduced Document Lens [90] which introduced a focus+context design inspired by the magnifier lens. In this visualization, the focused content of a document is shown in details in the view center, surrounded by other parts of the content that provides an overall impression of the text data. Despite showing the content, a similar idea is also used to visualize documents. For example, Data Mountain [89] employs a focus+context view based on a prospective projection, in which the focused documents are shown in a larger size with additional details in front of other documents, whereas the unfocused ones are shown at the back side in a smaller size (Fig. 2.24). Users can interactively switch between the focus and context by clicking on the documents.

2.5.2 Exploration Based on Document Similarity

Exploring documents in a similarity view is another early but popular technique that is extensively used in many text visualizations. These systems, such as InfoSky [3] and $IN\text{-}SPIRE^{TM}$ [44] and ForceSPIRE [31], use an overview that summarizes the entire document collection based on document similarities (see, Sect. 2.1) and employs a multiple coordinated view to reveal the document details from various aspects such as keywords and topics. Users can navigate through the similarity view based on interactions such as zooming and panning, thereby showing different levels of details [30].

2.5.3 Hierarchical Document Exploration

Exploring big document collection based on hierarchical clustering is another commonly used approach. For example, Paulovich and Minghim introduce HiPP [85], which lays out documents via a hierarchical circle packing algorithm, in which a document is shown as a circle. Dou et al. [29], Brehmer et al. [8], as well as Pascual-Cid and Kaltenbrunner [84] develop different types of document exploration systems that are all based on hierarchical clustering. In these systems, documents are hierarchically clustered based on their topic similarity and the cluster results are shown in a tree view to guide the data navigation.

2.5.4 Search and Query Based Approaches

Full text search and document query are also widely used to support document exploration since the very beginning of the text visualization [43, 94, 96]. Instead of showing a ranked list of related documents regarding the query keywords, most of the existing visualization techniques transform the search results into a visual representation to illustrate the insight of content relationships among documents. Graph layout and projection-based approach are commonly used to represent the search and query results showing the relationships among documents [7], or text snippets such as words [88] or collections of topic keywords [33, 37]. More advanced techniques are also developed. For example, Isaacs introduced Footprints to support an effective and interactive procedure to retrieve information in a desired subject from a large document collection. In this system, the topics queried by a user are visually summarized in an iconic representation as shown in Fig. 2.25. The user can click the topic to load a collection of related document in the document list and the content of a selected document can be further shown in the document viewer. A set of filters also helps users extract the most interesting data.

Exploration in Coordinated Views. The aforementioned document exploration techniques are usually combined with multiple coordinated visualization views that illustrate different aspects of the input document collection. Thew views are usually connected together through interactions such as "Linking and Brushing".[2] In these

[2] "The idea of linking and brushing is to combine different visualization methods to overcome the shortcomings of single techniques. Interactive changes made in one visualization are automatically

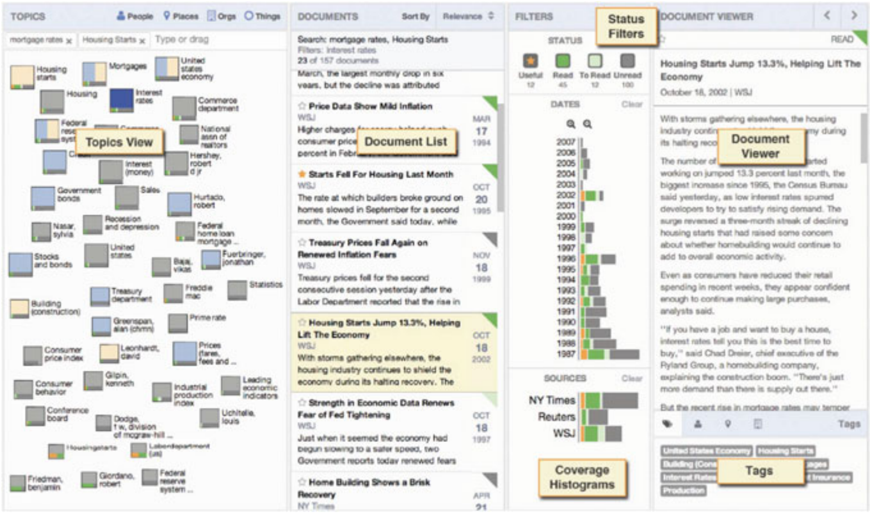

Fig. 2.25 Footprints, a topics-based document search tool, which supports exploratory search to help analysts **a** retrieve unknown information the goal of query is missing, **b** track the query results from multiple aspects to avoid missing important information and help to determine when should the query procedure be stopped

techniques, topic analysis is one of the most important aspects. For example, Dou et al. [27] introduce ParallelTopics, which guides the document exploration via multiple views. Specifically, it employs a parallel coordinates to illustrate how a document distributed in different topics, employs a TagCloud view to illustrate the keywords inside each topic, and employs the theme river to show the change of topics over time. Many other similar systems are available, such as Jigsaw [38] and IVEA [102], which are discussed in details in Chap. 4.

2.6 Summary of the Chapter

In this chapter, we reviewed more than 200 papers in the field of text visualization to provide an overview. This chapter provides readers a brief idea of what is text visualization and what is the research focus of this field. In particular, we summarize the existing works into three major categories based on the type of information to be shown in a visualization. Specifically, these techniques include those for (1) visualizing document similarities, (2) revealing text content, and (3) visualizing

(Footnote 2 continued)
reflected in the other visualizations. Note that connecting multiple visualizations through interactive linking and brushing provides more information than considering the component visualizations independently.” [56].

semantics and emotions. We also reviewed the most commonly used text exploration techniques, including (1) distortion-based approaches, (2) exploration based on document similarity, (3) hierarchical document exploration, (4) search and query-based approaches, and (5) exploration in coordinated views. In the following chapters, we focus on detailed techniques.

References

1. Alper, B., Yang, H., Haber, E., Kandogan, E.: Opinionblocks: visualizing consumer reviews. In: IEEE VisWeek 2011 Workshop on Interactive Visual Text Analytics for Decision Making (2011)
2. André, P., Wilson, M.L., Russell, A., Smith, D.A., Owens, A., et al.: Continuum: designing timelines for hierarchies, relationships and scale. In: Proceedings of the 20th Annual ACM Symposium on User interface Software and Technology, pp. 101–110. ACM (2007)
3. Andrews, K., Kienreich, W., Sabol, V., Becker, J., Droschl, G., Kappe, F., Granitzer, M., Auer, P., Tochtermann, K.: The infosky visual explorer: exploiting hierarchical structure and document similarities. Inf. Vis. **1**(3–4), 166–181 (2002)
4. Angus, D., Smith, A., Wiles, J.: Conceptual recurrence plots: revealing patterns in human discourse. IEEE Trans. Vis. Comput. Graph. **18**(6), 988–997 (2012)
5. Balakrishnama, S., Ganapathiraju, A.: Linear Discriminant Analysis—A Brief Tutorial, vol. 18. Institute for Signal and information Processing, Starkville (1998)
6. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003)
7. Bradel, L., North, C., House, L.: Multi-model semantic interaction for text analytics. In: 2014 IEEE Conference on Visual Analytics Science and Technology (VAST), pp. 163–172. IEEE (2014)
8. Brehmer, M., Ingram, S., Stray, J., Munzner, T.: Overview: the design, adoption, and analysis of a visual document mining tool for investigative journalists. IEEE Trans. Vis. Comput. Graph. **20**(12), 2271–2280 (2014)
9. Brew, A., Greene, D., Archambault, D., Cunningham, P.: Deriving insights from national happiness indices. In: 2011 IEEE 11th International Conference on Data Mining Workshops (ICDMW), pp. 53–60. IEEE (2011)
10. Brooks, M., Robinson, J.J., Torkildson, M.K., Aragon, C.R., et al.: Collaborative visual analysis of sentiment in Twitter events. In: Cooperative Design, Visualization, and Engineering, pp. 1–8. Springer, Berlin (2014)
11. Buetow, T., Chaboya, L., OToole, C., Cushna, T., Daspit, D., Petersen, T., Atabakhsh, H., Chen, H.: A spatio temporal visualizer for law enforcement. In: Intelligence and Security Informatics, pp. 181–194. Springer, Berlin (2003)
12. Burch, M., Beck, F., Diehl, S.: Timeline trees: visualizing sequences of transactions in information hierarchies. In: Proceedings of the Working Conference on Advanced Visual Interfaces, pp. 75–82. ACM (2008)
13. Burch, M., Lohmann, S., Beck, F., Rodriguez, N., Di Silvestro, L., Weiskopf, D.: Radcloud: visualizing multiple texts with merged word clouds. In: 2014 18th International Conference on Information Visualisation (IV), pp. 108–113. IEEE (2014)
14. Burch, M., Lohmann, S., Pompe, D., Weiskopf, D.: Prefix tag clouds. In: 2013 17th International Conference on Information Visualisation (IV), pp. 45–50. IEEE (2013)
15. Cao, N., Gotz, D., Sun, J., Lin, Y.R., Qu, H.: Solarmap: multifaceted visual analytics for topic exploration. In: IEEE International Conference on Data Mining, pp. 101–110. IEEE (2011)
16. Cao, N., Lu, L., Lin, Y.R., Wang, F., Wen, Z.: Socialhelix: visual analysis of sentiment divergence in social media. J. Vis. **18**(2), 221–235 (2015)

17. Cao, N., Sun, J., Lin, Y.R., Gotz, D., Liu, S., Qu, H.: Facetatlas: multifaceted visualization for rich text corpora. IEEE Trans. Vis. Comput Graph. **16**(6), 1172–1181 (2010)
18. Chen, C., Ibekwe-SanJuan, F., SanJuan, E., Weaver, C.: Visual analysis of conflicting opinions. In: 2006 IEEE Symposium on Visual Analytics Science and Technology, pp. 59–66. IEEE (2006)
19. Collins, C., Carpendale, S., Penn, G.: Docuburst: visualizing document content using language structure. Comput. Graph. Forum **28**(3), 1039–1046 (2009)
20. Collins, C., Viegas, F.B., Wattenberg, M.: Parallel tag clouds to explore and analyze faceted text corpora. In: IEEE Symposium on Visual Analytics Science and Technology, 2009. VAST 2009, pp. 91–98. IEEE (2009)
21. Cousins, S.B., Kahn, M.G.: The visual display of temporal information. Artifi. Intell. Med. **3**(6), 341–357 (1991)
22. Cui, W., Liu, S., Tan, L., Shi, C., Song, Y., Gao, Z.J., Qu, H., Tong, X.: Textflow: towards better understanding of evolving topics in text. IEEE Trans. Vis. Comput. Graph. **17**(12), 2412–2421 (2011)
23. Cui, W., Liu, S., Wu, Z., Wei, H.: How hierarchical topics evolve in large text corpora. IEEE Trans. Vis. Comput. Graph. **20**(12), 2281–2290 (2014)
24. Cui, W., Wu, Y., Liu, S., Wei, F., Zhou, M.X., Qu, H.: Context preserving dynamic word cloud visualization. In: IEEE Symposium on Pacific Visualization, pp. 121–128 (2010)
25. Diakopoulos, N., Elgesem, D., Salway, A., Zhang, A., Hofland, K.: Compare clouds: visualizing text corpora to compare media frames. In: Proceedings of IUI Workshop on Visual Text Analytics (2015)
26. Dörk, M., Gruen, D., Williamson, C., Carpendale, S.: A visual backchannel for large-scale events. IEEE Trans. Vis. Comput. Graph. **16**(6), 1129–1138 (2010)
27. Dou, W., Wang, X., Chang, R., Ribarsky, W.: Paralleltopics: a probabilistic approach to exploring document collections. In: 2011 IEEE Conference on Visual Analytics Science and Technology (VAST), pp. 231–240. IEEE (2011)
28. Dou, W., Wang, X., Skau, D., Ribarsky, W., Zhou, M.X.: Leadline: interactive visual analysis of text data through event identification and exploration. In: 2012 IEEE Conference on Visual Analytics Science and Technology (VAST), pp. 93–102. IEEE (2012)
29. Dou, W., Yu, L., Wang, X., Ma, Z., Ribarsky, W.: Hierarchicaltopics: visually exploring large text collections using topic hierarchies. IEEE Trans. Vis. Comput. Graph. **19**(12), 2002–2011 (2013)
30. Endert, A., Burtner, R., Cramer, N., Perko, R., Hampton, S., Cook, K.: Typograph: multiscale spatial exploration of text documents. In: 2013 IEEE International Conference on Big Data, pp. 17–24. IEEE (2013)
31. Endert, A., Fiaux, P., North, C.: Semantic interaction for visual text analytics. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 473–482. ACM (2012)
32. Fails, J.A., Karlson, A., Shahamat, L., Shneiderman, B.: A visual interface for multivariate temporal data: finding patterns of events across multiple histories. In: 2006 IEEE Symposium on Visual Analytics Science and Technology, pp. 167–174. IEEE (2006)
33. Forbes, A.G., Savage, S., Höllerer, T.: Visualizing and verifying directed social queries. In: IEEE Workshop on Interactive Visual Text Analytics, Seattle, WA (2012)
34. Gad, S., Javed, W., Ghani, S., Elmqvist, N., Ewing, T., Hampton, K.N., Ramakrishnan, N.: Themedelta: dynamic segmentations over temporal topic models. IEEE Trans. Vis. Comput. Graph. **21**(5), 672–685 (2015)
35. Gambette, P., Véronis, J.: Visualising a text with a tree cloud. In: Classification as a Tool for Research, pp. 561–569. Springer, Berlin (2010)
36. Gamon, M., Aue, A., Corston-Oliver, S., Ringger, E.: Pulse: mining customer opinions from free text. In: Advances in Intelligent Data Analysis VI, pp. 121–132. Springer, Berlin (2005)
37. Gomez-Nieto, E., San Roman, F., Pagliosa, P., Casaca, W., Helou, E.S., de Oliveira, M.C.F., Nonato, L.G.: Similarity preserving snippet-based visualization of web search results. IEEE Trans. Vis. Comput. Graph. **20**(3), 457–470 (2014)

38. Gorg, C., Liu, Z., Kihm, J., Choo, J., Park, H., Stasko, J.: Combining computational analyses and interactive visualization for document exploration and sensemaking in jigsaw. IEEE Trans. Vis. Comput. Graph. **19**(10), 1646–1663 (2013)
39. Gschwandtner, T., Aigner, W., Kaiser, K., Miksch, S., Seyfang, A.: Carecruiser: exploring and visualizing plans, events, and effects interactively. In: Visualization Symposium (PacificVis), 2011 IEEE Pacific, pp. 43–50. IEEE (2011)
40. Guzman, E.: Visualizing emotions in software development projects. In: IEEE Working Conference on Software Visualization, pp. 1–4. IEEE (2013)
41. Hao, M.C., Rohrdantz, C., Janetzko, H., Keim, D.A., et al.: Visual sentiment analysis of customer feedback streams using geo-temporal term associations. Inf. Vis. **12**(3–4), 273 (2013)
42. Havre, S., Hetzler, B., Nowell, L.: Themeriver: visualizing theme changes over time. In: IEEE Symposium on Information Visualization, 2000. InfoVis 2000, pp. 115–123. IEEE (2000)
43. Hearst, M.A., Karadi, C.: Cat-a-cone: an interactive interface for specifying searches and viewing retrieval results using a large category hierarchy. ACM SIGIR Forum **31**(SI), 246–255 (1997)
44. Hetzler, E., Turner, A.: Analysis experiences using information visualization. IEEE Comput. Graph. Appl. **24**(5), 22–26 (2004)
45. Hinton, G.E., Roweis, S.T.: Stochastic neighbor embedding. In: Advances in neural information processing systems, pp. 833–840 (2002)
46. Hofmann, T.: Probabilistic latent semantic indexing. In: Proceedings of International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 50–57. ACM (1999)
47. Iwata, T., Yamada, T., Ueda, N.: Probabilistic latent semantic visualization: topic model for visualizing documents. In: Proceedings of SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 363–371. ACM (2008)
48. Jain, R.: Out-of-the-box data engineering events in heterogeneous data environments. In: Proceedings. 19th International Conference on Data Engineering, 2003, pp. 8–21. IEEE (2003)
49. Jänicke, S., Geßner, A., Büchler, M., Scheuermann, G.: Visualizations for text re-use. GRAPP/IVAPP, pp. 59–70 (2014)
50. Jankowska, M., Keselj, V., Milios, E.: Relative n-gram signatures: document visualization at the level of character n-grams. In: 2012 IEEE Conference on Visual Analytics Science and Technology (VAST), pp. 103–112. IEEE (2012)
51. Jolliffe, I.: Principal Component Analysis. Wiley Online Library (2002)
52. Jøsang, A.: The consensus operator for combining beliefs. Artif. Intell. **141**(1), 157–170 (2002)
53. Kaser, O., Lemire, D.: Tag-cloud drawing: algorithms for cloud visualization. arXiv preprint cs/0703109 (2007)
54. Kaster, A., Siersdorfer, S., Weikum, G.: Combining text and linguistic document representations for authorship attribution. In: SIGIR Workshop: Stylistic Analysis of Text for Information Access (2005)
55. Keim, D., Oelke, D., et al.: Literature fingerprinting: a new method for visual literary analysis. In: IEEE Symposium on Visual Analytics Science and Technology, 2007. VAST 2007, pp. 115–122. IEEE (2007)
56. Keim, D., et al.: Information visualization and visual data mining. IEEE Trans. Vis. Comput. Graph. **8**(1), 1–8 (2002)
57. Kempter, R., Sintsova, V., Musat, C., Pu, P.: Emotionwatch: visualizing fine-grained emotions in event-related tweets. In: International AAAI Conference on Weblogs and Social Media (2014)
58. Koren, Y., Carmel, L.: Visualization of labeled data using linear transformations. In: Proceedings of IEEE Symposium on Information Visualization, pp. 121–128 (2003)
59. Krstajić, M., Bertini, E., Keim, D.A.: Cloudlines: compact display of event episodes in multiple time-series. IEEE Trans. Vis. Comput. Graph. **17**(12), 2432–2439 (2011)
60. Kruskal, J.B.: Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika **29**(1), 1–27 (1964)

61. Kucher, K., Kerren, A.: Text visualization browser: a visual survey of text visualization techniques. Poster Abstracts of IEEE VIS (2014)
62. Kurby, C.A., Zacks, J.M.: Segmentation in the perception and memory of events. Trends Cogn Sci **12**(2), 72–79 (2008)
63. Le, T., Lauw, H.W.: Semantic visualization for spherical representation. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1007–1016. ACM (2014)
64. Lee, D.D., Seung, H.S.: Algorithms for non-negative matrix factorization. In: Advances in Neural Information Processing Systems, pp. 556–562 (2001)
65. Lee, H., Kihm, J., Choo, J., Stasko, J., Park, H.: ivisclustering: an interactive visual document clustering via topic modeling. Comput. Graph. Forum **31**(3pt3), 1155–1164 (2012)
66. Levine, N., et al.: Crimestat iii: a spatial statistics program for the analysis of crime incident locations (version 3.0). Ned Levine & Associates, Houston/National Institute of Justice, Washington (2004)
67. Lin, Y.R., Sun, J., Cao, N., Liu, S.: Contextour: contextual contour visual analysis on dynamic multi-relational clustering. In: SIAM Data Mining Conference. SIAM (2010)
68. Liu, S., Chen, Y., Wei, H., Yang, J., Zhou, K., Drucker, S.M.: Exploring topical lead-lag across corpora. IEEE Trans. Knowl. Data Eng. **27**(1), 115–129 (2015)
69. Liu, S., Wang, X., Chen, J., Zhu, J., Guo, B.: Topicpanorama: a full picture of relevant topics. In: 2014 IEEE Conference on Visual Analytics Science and Technology (VAST), pp. 183–192. IEEE (2014)
70. Liu, S., Wu, Y., Wei, E., Liu, M., Liu, Y.: Storyflow: tracking the evolution of stories. IEEE Trans. Vis. Comput. Graph. **19**(12), 2436–2445 (2013)
71. 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. (TIST) **3**(2), 25 (2012)
72. Luo, D., Yang, J., Krstajic, M., Ribarsky, W., Keim, D.: Eventriver: visually exploring text collections with temporal references. IEEE Trans. Vis. Comput. Graph. **18**(1), 93–105 (2012)
73. Van der Maaten, L., Hinton, G.: Visualizing data using t-sne. J. Mach. Learn. Res. **9**(2579–2605), 85 (2008)
74. Makki, R., Brooks, S., Milios, E.E.: Context-specific sentiment lexicon expansion via minimal user interaction. In: Proceedings of the International Conference on Information Visualization Theory and Applications (IVAPP), pp. 178–186 (2014)
75. Marcus, A., Bernstein, M.S., Badar, O., Karger, D.R., Madden, S., Miller, R.C.: Twitinfo: aggregating and visualizing microblogs for event exploration. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 227–236. ACM (2011)
76. Miller, N.E., Wong, P.C., Brewster, M., Foote, H.: Topic islands TM—a wavelet-based text visualization system. In: IEEE Symposium on Information Visualization, pp. 189–196. IEEE (1998)
77. Munroe, R.: Movie narrative charts. http://xkcd.com/657/. Accessed Jan 2016
78. Neto, J.L., Santos, A.D., Kaestner, C.A., Freitas, A.A.: Document clustering and text summarization. In: Proceedings of the International Conference Practical Applications of Knowledge Discovery and Data Mining, pp. 41–55. The Practical Application Company (2000)
79. Oelke, D., Hao, M., Rohrdantz, C., Keim, D., Dayal, U., Haug, L.E., Janetzko, H., et al.: Visual opinion analysis of customer feedback data. In: IEEE Symposium on Visual Analytics Science and Technology, 2009. VAST 2009, pp. 187–194. IEEE (2009)
80. Oelke, D., Kokkinakis, D., Keim, D.A.: Fingerprint matrices: uncovering the dynamics of social networks in prose literature. Comput. Graph. Forum **32**(3pt4), 371–380 (2013)
81. Oelke, D., Strobelt, H., Rohrdantz, C., Gurevych, I., Deussen, O.: Comparative exploration of document collections: a visual analytics approach. Comput. Graph. Forum **33**(3), 201–210 (2014)
82. Ogawa, M., Ma, K.L.: Software evolution storylines. In: Proceedings of the 5th International Symposium on Software Visualization, pp. 35–42. ACM (2010)
83. Ogievetsky, V.: PlotWeaver. http://ogievetsky.com/PlotWeaver/. Accessed Jan 2016

84. Pascual-Cid, V., Kaltenbrunner, A.: Exploring asynchronous online discussions through hierarchical visualisation. In: 2009 13th International Conference on Information Visualisation, pp. 191–196. IEEE (2009)
85. Paulovich, F.V., Minghim, R.: Hipp: a novel hierarchical point placement strategy and its application to the exploration of document collections. IEEE Trans. Vis. Comput. Graph. **14**(6), 1229–1236 (2008)
86. Plaisant, C., Mushlin, R., Snyder, A., Li, J., Heller, D., Shneiderman, B.: Lifelines: using visualization to enhance navigation and analysis of patient records. In: Proceedings of the AMIA Symposium, p. 76. American Medical Informatics Association (1998)
87. Reisinger, J., Waters, A., Silverthorn, B., Mooney, R.J.: Spherical topic models. In: Proceedings of International Conference on Machine Learning, pp. 903–910 (2010)
88. Riehmann, P., Gruendl, H., Potthast, M., Trenkmann, M., Stein, B., Froehlich, B.: Wordgraph: keyword-in-context visualization for netspeak's wildcard search. IEEE Trans. Vis. Comput. Graph. **18**(9), 1411–1423 (2012)
89. Robertson, G., Czerwinski, M., Larson, K., Robbins, D.C., Thiel, D., Van Dantzich, M.: Data mountain: using spatial memory for document management. In: Proceedings of the 11th Annual ACM Symposium on User Interface Software and Technology, pp. 153–162. ACM (1998)
90. Robertson, G.G., Mackinlay, J.D.: The document lens. In: Proceedings of the 6th Annual ACM Symposium on User Interface Software and Technology, pp. 101–108. ACM (1993)
91. Rohrdantz, C., Hao, M.C., Dayal, U., Haug, L.E., Keim, D.A.: Feature-based visual sentiment analysis of text document streams. ACM Trans. Intell. Syst. Technol. (TIST) **3**(2), 26 (2012)
92. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. Science **290**(5500), 2323–2326 (2000)
93. Rusu, D., Fortuna, B., Mladenić, D., Grobelnik, M., Sipos, R.: Document visualization based on semantic graphs. In: 13th International Conference Information Visualisation, pp. 292–297. IEEE (2009)
94. Sebrechts, M.M., Cugini, J.V., Laskowski, S.J., Vasilakis, J., Miller, M.S.: Visualization of search results: a comparative evaluation of text, 2d, and 3d interfaces. In: Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 3–10. ACM (1999)
95. Smith, D.A.: Detecting events with date and place information in unstructured text. In: Proceedings of the 2nd ACM/IEEE-CS Joint Conference on Digital Libraries, pp. 191–196. ACM (2002)
96. Spoerri, A.: Infocrystal: a visual tool for information retrieval & management. In: Proceedings of the Second International Conference on Information and Knowledge Management, pp. 11–20. ACM (1993)
97. Stasko, J., Zhang, E.: Focus+ context display and navigation techniques for enhancing radial, space-filling hierarchy visualizations. In: IEEE Symposium on Information Visualization, 2000. InfoVis 2000, pp. 57–65. IEEE (2000)
98. Stoffel, A., Strobelt, H., Deussen, O., Keim, D.A.: Document thumbnails with variable text scaling. Comput. Graph. Forum **31**(3pt3), 1165–1173 (2012)
99. Strobelt, H., Oelke, D., Rohrdantz, C., Stoffel, A., Keim, D., Deussen, O., et al.: Document cards: a top trumps visualization for documents. IEEE Trans. Vis. Comput. Graph. **15**(6), 1145–1152 (2009)
100. Tanahashi, Y., Hsueh, C.H., Ma, K.L.: An efficient framework for generating storyline visualizations from streaming data. IEEE Trans. Vis. Comput. Graph. **21**(6), 730–742 (2015)
101. Tanahashi, Y., Ma, K.L.: Design considerations for optimizing storyline visualizations. IEEE Trans. Vis. Comput. Graph. **18**(12), 2679–2688 (2012)
102. Thai, V., Handschuh, S., Decker, S.: Tight coupling of personal interests with multi-dimensional visualization for exploration and analysis of text collections. In: International Conference on Information Visualisation, pp. 221–226. IEEE (2008)
103. Van Ham, F., Wattenberg, M., Viégas, F.B.: Mapping text with phrase nets. IEEE Trans. Vis. Comput. Graph. **15**(6), 1169–1176 (2009)

104. Viegas, F.B., Wattenberg, M., Feinberg, J.: Participatory visualization with wordle. IEEE Trans. Vis. Comput. Graph. **15**(6), 1137–1144 (2009)
105. Wang, C., Xiao, Z., Liu, Y., Xu, Y., Zhou, A., Zhang, K.: Sentiview: sentiment analysis and visualization for internet popular topics. IEEE Trans. Hum. Mach. Syst **43**(6), 620–630 (2013)
106. Wang, T.D., Plaisant, C., Shneiderman, B., Spring, N., Roseman, D., Marchand, G., Mukherjee, V., Smith, M.: Temporal summaries: supporting temporal categorical searching, aggregation and comparison. IEEE Trans. Vis. Comput. Graph. **15**(6), 1049–1056 (2009)
107. Wanner, F., Rohrdantz, C., Mansmann, F., Oelke, D., Keim, D.A.: Visual sentiment analysis of rss news feeds featuring the us presidential election in 2008. In: Workshop on Visual Interfaces to the Social and the Semantic Web (VISSW) (2009)
108. Wattenberg, M.: Arc diagrams: visualizing structure in strings. In: IEEE Symposium on Information Visualization, pp. 110–116. IEEE (2002)
109. Wattenberg, M., Viégas, F.B.: The word tree, an interactive visual concordance. IEEE Trans. Vis. Comput. Graph. **14**(6), 1221–1228 (2008)
110. Wensel, A.M., Sood, S.O.: Vibes: visualizing changing emotional states in personal stories. In: Proceedings of the 2nd ACM International Workshop on Story Representation, Mechanism and Context, pp. 49–56. ACM (2008)
111. Wongsuphasawat, K., Gotz, D.: Exploring flow, factors, and outcomes of temporal event sequences with the outflow visualization. IEEE Trans. Vis. Comput Graph. **18**(12), 2659–2668 (2012)
112. Wongsuphasawat, K., Guerra Gómez, J.A., Plaisant, C., Wang, T.D., Taieb-Maimon, M., Shneiderman, B.: Lifeflow: visualizing an overview of event sequences. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1747–1756. ACM (2011)
113. Wongsuphasawat, K., Shneiderman, B.: Finding comparable temporal categorical records: a similarity measure with an interactive visualization. In: IEEE Symposium on Visual Analytics Science and Technology, 2009. VAST 2009, pp. 27–34. IEEE (2009)
114. Wu, Y., Provan, T., Wei, F., Liu, S., Ma, K.L.: Semantic-preserving word clouds by seam carving. Comput. Graph. Forum **30**(3), 741–750 (2011)
115. Wu, Y., Wei, F., Liu, S., Au, N., Cui, W., Zhou, H., Qu, H.: Opinionseer: interactive visualization of hotel customer feedback. IEEE Trans. Vis. Comput. Graph. **16**(6), 1109–1118 (2010)
116. Zacks, J.M., Tversky, B.: Event structure in perception and conception. Psychol. Bull. **127**(1), 3 (2001)
117. Zhang, C., Liu, Y., Wang, C.: Time-space varying visual analysis of micro-blog sentiment. In: Proceedings of the 6th International Symposium on Visual Information Communication and Interaction, pp. 64–71. ACM (2013)
118. Zhao, J., Gou, L., Wang, F., Zhou, M.: Pearl: an interactive visual analytic tool for understanding personal emotion style derived from social media. In: 2014 IEEE Conference on Visual Analytics Science and Technology (VAST), pp. 203–212. IEEE (2014)

http://www.springer.com/978-94-6239-185-7

Introduction to Text Visualization
Cao, N.; Cui, W.
2016, VIII, 114 p. 81 illus., 71 illus. in color., Hardcover
ISBN: 978-94-6239-185-7
A product of Atlantis Press