J Vis
https://doi.org/10.1007/s12650-020-00689-0

REGULAR PAPER

Xinhuan Shu · Jiang Wu · Xinke Wu · Hongye Liang · Weiwei Cui · Yingcai Wu · Huamin Qu

DancingWords: exploring animated word clouds to tell stories

Received: 1 July 2020 / Accepted: 9 August 2020

Abstract By encoding semantic relations into relative positions, word clouds have shown the capability to deliver richer messages than purely visualizing word frequencies. Existing studies mainly focus on layout algorithms that cluster related words, preserve temporal coherence, and optimize spatial shapes. However, they cannot fully convey multiple relations among words and their evolvement through relative positions and static representations. In this paper, we explore animated word clouds that take advantage of storytelling strategies to present interactions between words and show the dynamic process of content changes, thus communicating the underlying stories. We initially create several exemplars of animated word clouds with designers through a structured iterative design process. These exemplars lead to a preliminary design space that distills essential narrative elements with design choices. Based on the design space, we develop a prototype tool, DancingWords, which provides story-oriented interactions and automatic layouts for users to generate animated word clouds. We evaluate the expressiveness and usefulness of our system through several example animated stories and a usability study with general users. The results show that DancingWords allows users to produce appealing storytelling videos easily and quickly for communication.

Keywords Storytelling · Animation · Text visualization · Interaction

1 Introduction

Word clouds have gained popularity for providing fast impressions of text data with prominent words (Felix et al. 2018). They usually encode the word significance by font size and spatially arrange words on the canvas. Previous research on word clouds focuses on layout generations with various objectives, such as addressing the aesthetic issues (Viégas and Wattenberg 2008; Wang et al. 2020; Wor 2020) and communicating relationships between words (Cui et al. 2010; Wu et al. 2011; Hearst et al. 2019). Recent studies

X. Shu · H. Qu
Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Hong Kong, China
E-mail: xshuaa@connect.ust.hk

X. Shu · Y. Wu (✉)
Zhejiang Lab, Hangzhou, China
E-mail: ycwu@zju.edu.cn

J. Wu · X. Wu · H. Liang · Y. Wu
State Key Lab of CAD&CG, Zhejiang University, Hangzhou, China

W. Cui
Microsoft Research Asia, Beijing, China

Published online: 04 September 2020

utilize word clouds to facilitate storytelling. For example, researchers (Cui et al. 2010; Wu et al. 2011) encoded semantic relationships by relative positions. Wang et al. (2018a) introduced context-aware editing, which enables users to order words in a semantically meaningful way.

However, current semantic word clouds are not effective in communicating stories at two key aspects. The first aspect is to illustrate complex relationships between words at the same time. A simple story like *Little Red Riding Hood* can have multiple roles and events between them. However, current encoding methods (e.g., sizes, relative positions, and colors) face challenges to express these relations clearly. For example, we can place words *Wolf* and *Grandmother* close to each other to show their co-occurrence in an event, but it is hard to reflect the status that *Grandmother* is in the stomach of *Wolf*. Besides, words in a word cloud provide different levels of information for the stories. For instance, *Old* describes *Grandmother*, while *Forest* indicates the location of the story. Second, narrative contents evolve ceaselessly with a set of events are advancing the story temporally and causally. For instance, the story initially occurs between *Little Girl* and *Wolf*; then, *Wolf* goes for *Grandmother*. Thus, a single word cloud for the whole story loses temporal information. Although consecutive changes can be visualized by word clouds side-by-side, it adds the cognitive load on viewers to track keywords, letting alone a story which might have multiple events simultaneously. Therefore, intuitively revealing word relations and event changes remain an understudied problem.

The first limitation requires an understanding of essential narrative elements and their relationships, which spurs us to propose new visual designs for enriching the expressiveness of traditional word clouds. The designs should allow audiences to quickly perceive different narrative relations and keep consistent during the whole story. The second deficiency in conveying temporal evolution inspires us to leverage the merits of animation to improve the transition (Heer and Robertson 2007) and engage viewers (Tversky et al. 2002; Amini et al. 2018). Combining animation with word clouds allows to unfold complex relationships in a story via the dimension of time.

In this paper, we explore how animated word clouds can serve as a storytelling device. To understand the potential and identify designs, we follow a structured design process to create examples of animated word clouds for specific stories. We collect a set of storytelling videos after several rounds of iterative design. Further analysis and follow-up interviews with designers result in a preliminary design space to create animated word clouds. Based on the design space, we implement a proof-of-concept prototype, *DancingWords*, to assist the creation of animated word clouds for presentations. *DancingWords* encapsulates designs for essential narrative elements, provides automatic layouts, and offers visual embellishments to augment the expressiveness of word clouds. Users can interactively specify relations between words and generate animation automatically. We demonstrate the effectiveness with several examples and further evaluate the system with a usability study and received encouraging feedback. Our contributions are:

- We present a preliminary design space of animated word clouds for telling the underlying stories through an exploratory study, where we identify key narrative elements, propose a set of design factors, and summarize design goals.
- Based on the design space, we develop *DancingWords*, an authoring tool that allows users to explore narratives and generate animation with automatic layouts interactively.

2 Related work

We reviewed previous work in two closely related topics, namely, word cloud visualization and storytelling visualization.

2.1 Word cloud-related visualization

Word clouds are widespread because of their appealing visual design and capability to provide fast impressions of text data. A word cloud automatically extracts frequent words from a text corpus and encodes the frequency by font size. The spatial arrangement of words in word clouds has received much attention. Plenty of layout algorithms have emerged with different design goals. *Wordle* (Viégas et al. 2009) arrange words in different colors and orientations for aesthetic criteria, which is widely used for text analysis (Li et al. 2018; Onoue et al. 2016). One stream of word cloud-related studies (Wang et al. 2020; Wor 2020) aims to present the whole word cloud as a specific shape to attract viewers. Furthermore, morphable word clouds (Chi et al. 2015) constrain multiple word clouds in a set of evolving shapes and design smooth shape transitions with animation. The animation layout follows the constrains of temporal coherence and compactness. In our layout algorithm, we currently do not consider the compactness of word clouds, but leave the space for word animation.

Another stream arranges words based on the relationships between each two words. For example, Wu et al. (2011) extracted semantically similar keywords from documents and positioned them close to each other. Cui et al. (2010) placed words which preserved similar context in the neighborhood, while preserving the temporal coherence among multiple word clouds. In addition, Wang et al. (2018a) provided a method that allows users to consistently edit word clouds and generate semantically meaningful layouts manually. Other works provide variants of word clouds to show context information, e.g., adding links between words to indicate co-occurrence (Wattenberg and Vigas 2008; Hu et al. 2017). Our work can be placed in this category, where we incorporate the narrative relations between words into word clouds. However, these existing methods cannot fully satisfy our requirements, since the word relations in stories are more than co-occurrence and evolve along the story advancement. Animation naturally unfolds the complex relations via the time dimension, which allows to present the story frame-by-frame. Besides, animation is regarded as one of mainstream storytelling mediums (Segel and Heer 2010), thus inspiring us to further explore how to combine animation and word clouds to tell stories.

2.2 Storytelling visualization and authoring tools

Data-driven storytelling (Segel and Heer 2010; Latif and Beck 2019) has attracted much attention in the public, which in turn spurs the need for authoring tools (Mei et al. 2018b). A palette of tools has emerged recently for different storytelling forms (e.g., Kim et al. 2017; Wang et al. 2018b; Amini et al. 2017; Mei et al. 2018a; Lu et al. 2020), of which several are designed to communicate insights for specific data or visualization types. For example, DataToon (Kim et al. 2019) focuses on creating data comics for dynamic graphs, and Timeline Storyteller (Brehmer et al. 2017) is devoted to timeline construction. Our work is specific to word clouds and contributes an authoring tool that interactively crafts word clouds and animations to generate storytelling videos.

Video authoring tools such as Adobe After Effects allow designers to create animation freely. However, the flexibility sets barriers to novices, which requires users' creativity for animated design and familiarity with tools. Data videos present different characteristics related to narrative constructs (Cao et al. 2020; Tang et al. 2020). Bulterman and Hardman (2005) argued that a structure-based paradigm is useful for authoring video-based stories. Authoring tools build up a storyboard structure based on abstract narrative components, and users directly manipulate these components to complete the story. *DancingWords* follows a similar design paradigm, where we summarize essential story elements with designs and encapsulate corresponding animated designs in story-oriented interactions. Users are able to create animated videos by interactively specifying relationships of words in a narrative order.

3 Designing animated word clouds

This section first introduces our study procedure during which we identified key narrative elements and collected example videos through several rounds of iterations with designers. Based on the study, we summarized a preliminary design space.

3.1 Study procedure

This study went through five steps to explore essential narrative elements and designs in animated word clouds for storytelling, following a similar paradigm to create graph comics (Bach et al. 2016).

3.1.1 Understand storytelling

We intended to take advantage of storytelling that helps to communicate underlying relations in word clouds. It required us to identify essential narrative elements when telling a story. We first referred to the literature in the narratology (Bal 2009; Coulter and Smith 2009) and storytelling (Blazer 2016) and analyzed examples in different forms such as fictions, comics, and animation. Since a multitude of studies has formed a large design space, we considered basic story components which contain a series of temporally and causally related events, and characters interacting in a given setting (Bal 2009). Three fundamental dimensions are included in this definition, i.e., contexts, roles, and events. *Contexts* provide the background information of a story, e.g., indicating where and when the story occurs, introducing a narrative world to the audience. *Roles* include characters and objects. Characters are actors of the story, and objects can act and interact with characters. *Events* involve various activities between roles and are arranged in a specific *order* to advance the story. Audience can naturally identify these narrative elements and understand the story. Animated word clouds aim to leverage storytelling to improve communication.

3.1.2 Design prototypes with discussion

Our second step was to explore possible designs that explicitly showed these narrative dimensions in animated word clouds. We held a three-hour brainstorming seminar with six PhD students who have experience in storytelling visualization. During the brainstorming, we asked them to independently design visual representations for each narrative dimension and then discussed in a collaborative session. The results converged to the following practices:

- *Contexts* are designed as underpaintings and introduced first.
- *Roles* are underlying topics composed by multiple adjacent words. Separated clusters of words indicate different roles, and their positions suggest their relations.
- *Events* represent the evolvement of topics and are displayed through animation.

Two of our authors then produced four prototypes based on four documents of different length, i.e., *Little Red Riding Hood*, *The Moon and Six Pence*, part of *Napoleon*, and news summary. To collect external feedback and evaluate the video quality, we presented our prototypes to two experts, namely, a senior visualization researcher and an expert in computing aesthetics, and discussed in the laboratory seminar. They all provided feedback on the perception of content and aesthetics of animation design. We further improved our videos after several iterations in this step.

3.1.3 Design principles

Based on our practices, we summarized three design principles to ensure readability. These principles referred to previous studies (Tversky et al. 2002; Heer and Robertson 2007; Amini et al. 2018) in visualization and literature on animated storytelling (Blazer 2016).

- Positions should indicate semantics. For example, closely related words are clustered to represent a role, and different roles in the same context should be identified easily.
- Each animation should convey a certain narrative element and avoid meaningless movements. For example, if a word moves from one position to another, there must be a corresponding change on the relations of words.
- Extra visual elements can be added to augment the expression, e.g., using background color or contours to show groups.

3.1.4 Create animated word clouds with designers

To validate the narrative dimensions and collect more design choices, we recruited seven video designers (five females, two males, aged between 21 and 25) from a university by disseminating advertisements through emails and among online designer groups. All participants (p1–p7) have design-related major backgrounds, including Industrial Design (3), Information Visualization (2), Product Design (1), and Digital Media (1). They have received formal training in visual communication and had professional experience in making animated storytelling videos for academic or social purposes (mean = 2.86 years, range = 2–5 years). In addition, they all have experience in creating word clouds, and five of them are familiar with information visualization. Each participant was paid ranging from \$30 to \$40 according to the video quality with diverse designs, which were judged by the first three paper authors.

The whole procedure is as follows. We first gave a brief introduction to our theme, i.e., animating word clouds that enables to communicate the underlying stories. Next, we showed participants our current videos and other animated storytelling forms, since appropriate example videos are useful to guide the design process (Kerzner et al. 2019). We also emphasized that visual representations are not limited to the given examples and encouraged novel designs for different narrative dimensions during the whole process. After the introduction, we required participants to generate animated word clouds for the story, *Little Red Riding Hood*. We chose this story to collect designs due to two major reasons. First, this story covered basic elements identified in the first step and are familiar to participants. Second, they did not need to spend time collecting extra information and reading large text documents. Before instantiating the videos, we asked participants to sketch their ideas first. We closely communicated with them during the process, ensuring that the final video should clearly convey the story and their designs conform to word cloud visualization. Sketches involved storyboarding, basic layouts at key frames, and animation designs. After that, participants produced videos using different tools. Four of them chose Adobe After Effect, and the others used Microsoft PowerPoint. After watching the first version of videos, five participants were asked to fine-tune parts of their videos which were hard to follow or interpret. It took ~ 5 hours (except the tuning time) on average for a video.

3.1.5 Post interview

After the study, we interviewed the participants about the authoring process and design choices, i.e., *what narrative elements help communication*, and *how these elements were visually presented*.

First, they agreed that the three basic dimensions (i.e., contexts, roles, and events) were critical to convey stories. In fact, these narrative dimensions guided their design process. They first roughly segmented the story according to the context information, then refined the narrative structure based on the evolvement of roles, and finally determined the storyboarding in line with events. Second, designers reported on several rationales they adopted during the process based on their previous experience. They suggested explicit and consistent designs for each narrative element can help identify the specific element and communicate the relations. In addition, they preferred to enhance the performance of animated storytelling with additional visual embellishments. For example, they added roles' pictures and figurative icons to enrich expression. Therefore, we proposed another dimension for such effects. Through the interviews, we decomposed videos with the designers and grouped design choices in narrative elements. We converged towards a design space of animated word clouds for storytelling by iterative discussions.

3.2 Design space

Based on related literature on narratology (Bal 2009; Coulter and Smith 2009) and interviews with designers, we first built the design space with four dimension (i.e., contexts, roles, events, and embellishments). To codify the videos, two authors decomposed the video with designers and corresponded the designs to each narrative dimension during the postinterviews. Thus, we collected multiple pairs of design and its narrative dimension. Moreover, considering the visual differences of designs in each dimension, we re-examined their narrative roles in the animated stories and further distinguished narrative elements after iterative discussions among authors. We present design strategies for each narrative elements (E1–E8) as follows. It is noted that we do not intend to exhaust the methods and give an inclusive design space. Figures demonstrate descriptive examples derived from our collected videos.

3.2.1 Dimension I: Display of contexts

Contexts, also referred to settings, set up the main background and atmosphere of the story. According to our interview in Sect. 3.1.5, participants declared that when they were designing the story, they first illustrated contexts, such as where and when the stories took place. It is common that a story consists of multiple contexts, which requires designs to set up a single context and show a switch process of contexts.

E1. Context setup Context information encircles a set of words in a scope using contours, background colors, and screen split (Fig. 1a). Words that specify the context information are better to be explicitly displayed around the scope. These designs conform to the law of common region in Gestalt principles (Palmer 1992). Viewers can naturally perceive that the story takes place in this certain context.

E2. Scene switch These designs (Fig. 1b) refer to similar techniques in cinematography (Mascelli 2005). Both camera motions and pushing effects trace the changing path of different contexts, while zoom shows that the story develops into the previous context.

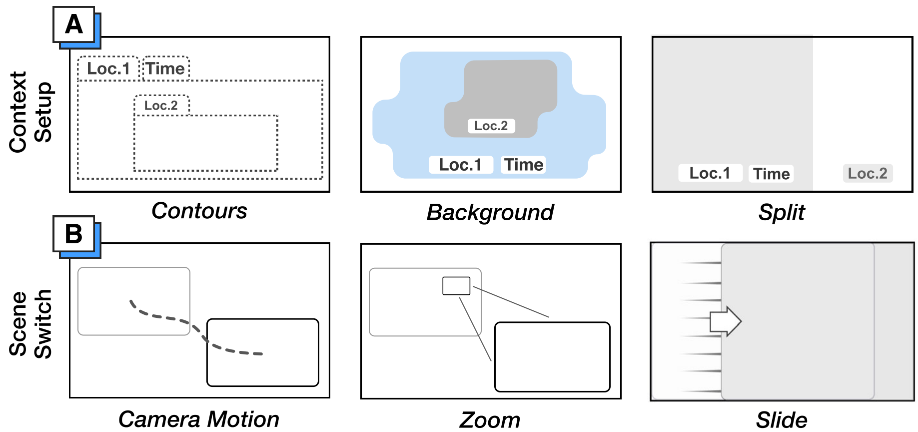

Fig. 1 Designs for contexts. **a** Context setup and **b** scene switch

3.2.2 Dimension II: Cast of roles

Roles could be a significant impetus in storytelling, which unfold the story based on their own growth or relation changes with others. The interview with designers indicated that the evolvement of roles helped determine key frames. In word clouds, we planned to reveal the changes of topics through specifying words as "roles." Specifically, we described two possible relations of roles, i.e., groups of multiple roles (E3) and attached information to a role (E4).

E3. Groups Groups are organized by roles on the same side, and different groups are identified by semantic relations. It can be displayed via contours and background colors according to the common region (Palmer 1992) or shown as network visualization (Fig. 2a).

E4. Description Description represents words that are closely related to the core topic, but cannot illustrate topics by themselves as attached information. They are placed around the related roles according to the proximity law (Max 1923). Visual metaphors like dashed links and brackets are also used for description.

3.2.3 Dimension III: Event representation

An event in the story indicates that some changes take place (Bal 2009). A series of events are arranged casually and temporally to describe the dynamic process. Words in the *event* dimension are used to explain the relations between roles. Traditional stories can contain a variety of events for expressiveness, but we do not consider to design for each event type. In this paper, we present two typical events according to different design strategies, i.e., action and dialog.

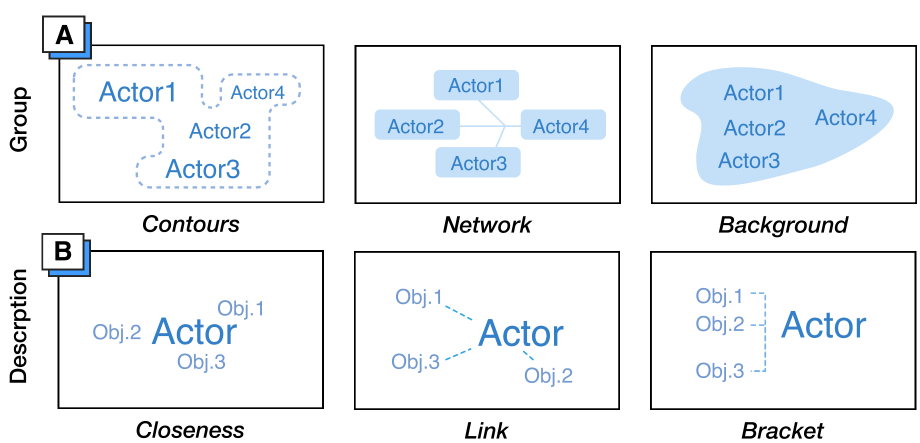

Fig. 2 Designs for roles. **a** Description and **b** group

E5. Action We found that designers leverage graphics to manifest events. For example, an arrow with words from one role to another explains a directed event (Fig. 3a-1), and dashed lines weaken the directionality between two roles (Fig. 3a-2). Figure 3a-3 shows another common action which objects are exchanged between two roles.

E6. Dialog Dialog events are commonly designed differently in comparison with the action event. Typically they use visual metaphors, e.g., speakers, dialog, and thought bubbles (Fig. 3b).

3.2.4 Dimension IV: Visual embellishments

In addition to the three essential narrative elements, designers have put great efforts in designing good storytelling effects. The postinterview in Sect. 3.1.5 reflected on their strong intentions to increase story retention and convey emotions with visual embellishments. Therefore, we propose another dimension to the animated word clouds for storytelling. There should be a variety of methods to embellish animation in fact. We describe two strategies our participants commonly used in their designs.

E7. Icons The benefit of integrating pictographs or icons to augment the expression has been widely discussed in communicative visualization (Amini et al. 2015; Bateman et al. 2010), such as lowering interpretation time and impressing viewers. Compared to plain texts, icons play vital roles in engagement.

E8. Special effects Designers include special animation designs to convey certain emotions. For example, a word "bumps" another that depicts a fierce fight. Some words circled by thought bubbles represent dialogs. These kinds of designs mimic traditional 2D animation. Most of them are designed based on well-established guidelines, such as *Disney's twelve principles of animation* (Thomas and Johnston 1995).

4 DancingWords

This section presents design goals we proposed for our authoring tool, *DancingWords*, followed by the design details of each part.

4.1 Design goals

We settled on three design goals based on the preliminary design space and observations of the authoring processes.

- *G1. Set up narrative elements interactively* The system should allow users to directly assign words with narrative elements based on their relations and provide corresponding designs. Encapsulated interactions can facilitate the story-oriented authoring process (Bulterman and Hardman 2005).

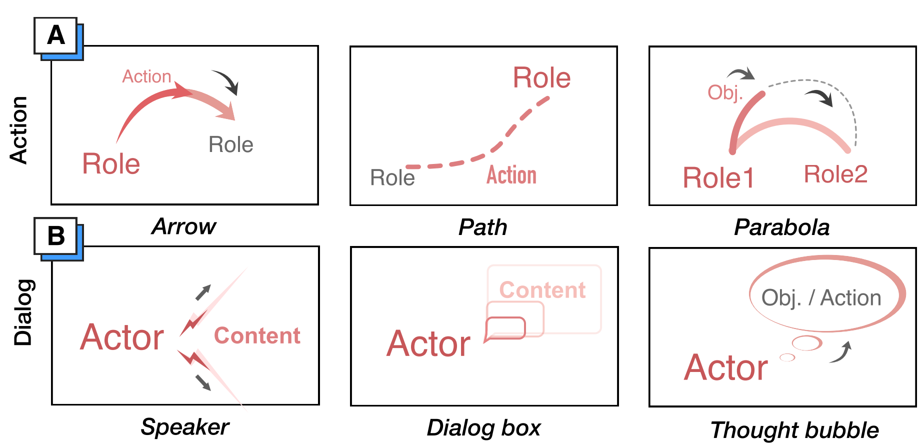

Fig. 3 Designs for events. **a** Event and **b** dialog. The black arrows indicate the animation process. The opacity distinguishes different states during the animation

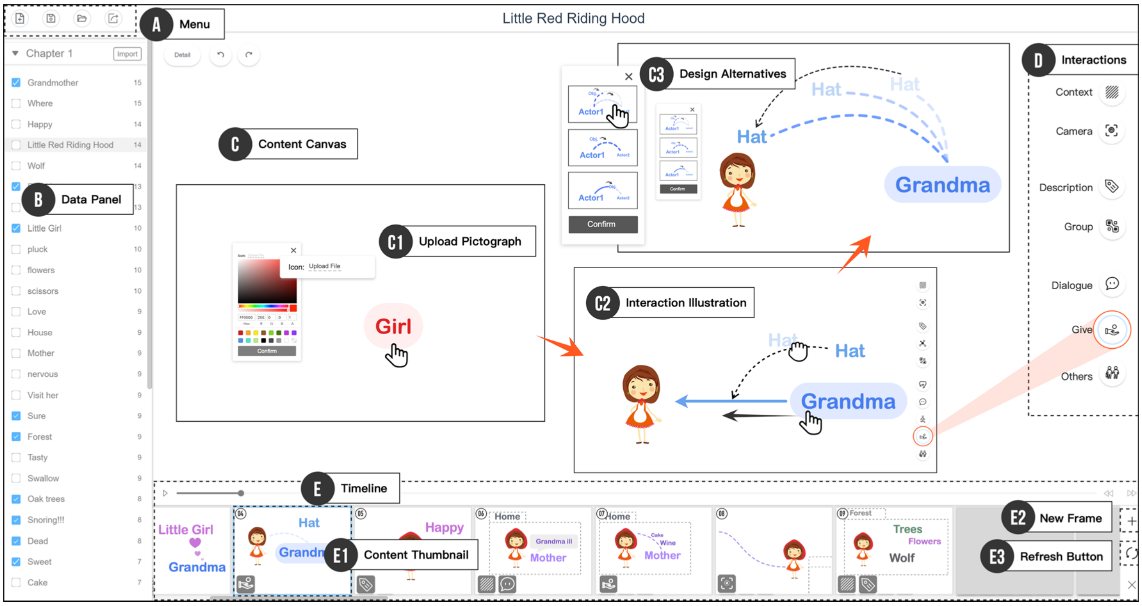

Fig. 4 *DancingWords* consists of five components. **a** A top menu bar provides basic file operations and shows the title. **b** The left data panel lists words in different chapters and allows users to add and check words. **c** The content canvas supports content authoring and reviewing with a set of story-oriented interactions (**d**). **e** The timeline panel presents a timeline with thumbnails of key frames

- *G2. Support the story-driven auto-layout* The system should organize the layout of word clouds according to different narrative elements and designs. It alleviates the efforts of manual adjustment and achieves congruence during animation.
- *G3. Incorporate visual embellishments* Visual embellishments can greatly increase the retention and engaging effects of animated storytelling as mentioned in Dimension IV. The system should allow to improve animated word clouds with E7 and E8.

4.2 System design

Following the design space and goals, we developed a prototype web-based system, *DancingWords*, to assist the creation.

4.2.1 Usage scenario

DancingWords works as a proof-of-concept tool to create animated word clouds for storytelling. Target users can be both designers who use the tool for rapid video prototyping and iteration, and novices who can easily author animated stories for presentation. We present a usage scenario to demonstrate how a hypothetical author can use *DancingWords* to quickly create an animated story, *Little Red Riding Hood*.

The user begins the creation by importing the original text document to the tool. *DancingWords* automatically extracts a set of frequent words and lists them in the decreasing order of frequency (Fig. 4b), which is similar to generate typical word clouds. The system allows users to add or modify some words and their frequencies if needed. The user then creates the animated stories by crafting key frames. Take a plot as an example where grandmother gives a hat to her little girl . The user first chooses the words *Girl*, *Grandma*, and *Hat* from the data panel (Fig. 4b). Considering the little *Girl* is the leading role of the story, the user wants to enrich the expression and uses pictographs. By clicking on the word *Girl*, the user uploads an icon to replace the plain text (Fig. 4c-1) and then chooses the *Give* interaction from the set of interactions (Fig. 4d) to complete this plot. After specifying the mode (Fig. 4c-2), the user clicks on two roles, *Grandma* and the icon, successively. The system automatically generates an arrow between the specified roles. The user then drags the word *Hat* to the arrow and finishes the event. The output animation is expected to move the word *Hat* from the word *Grandma* to the girl icon following a parabola path. The system also provides alternative designs (Fig. 4c-3) for users based on the preliminary design study. The current key frame on the timeline will be updated with the newly added interaction icon and content thumbnail (Fig. 4e-1). Moreover, users can click the refresh button (Fig. 4e-3) to re-layout this frame if unsatisfied with the current automatic layout.

After finishing the content of this frame, the user adds a new key frame (Fig. 4e-2). The words for the "context" and "role" dimensions will be checked and appear in the follow-up frame by default. The user then repeats operations, choosing words in the data panel, acting with different interactions according to the story content, and generating the animation. He could re-play the animation during the creation process to check the effects and refine the content. Finally, satisfied with the results, the user exports the video for presentations (Fig. 7).

4.2.2 *User interface*

The user interface is implemented using Vue.js. The main view uses Konva, an HTML Canvas Javascript library, to interactively manipulate 2D context. Figure 4 shows our authoring tool that allows general users to create animated word clouds in an integrated process from importing data, interacting with narrative elements, to exporting videos. The interface is composed of five UI components. The top menu bar (Fig. 4a) provides basic file operations (i.e., new, save, load, and export). The left data panel (Fig. 4b) lists all imported words in a decreasing order of frequency. Checked words are displayed in the content canvas (Fig. 4c), which supports content authoring and reviewing. A set of story-oriented interactions are arranged vertically on the right (Fig. 4d). These interactions are grouped according to the three different narrative dimensions. The bottom of the interface (Fig. 4e) presents a timeline with thumbnails of key frames, where users can review and re-edit the key frames they have created. For example, dragging the interaction icon out of the thumbnail (Fig. 4e-1) could delete this interaction.

4.2.3 *Story-oriented Interactions*

To incorporate G1, we provide the following story-oriented interactions (Fig. 5) that allow users to directly specify the narrative elements on the content canvas. These interactions help construct the content tree for the automatic layout algorithm in Sect. 4.2.8.

4.2.4 *Dimension I: Display of context*

Context mode asks users to specify words as context (E1) first: It generates a dashed contour by default. Users lasso and drag a set of words into the context scope.

Camera mode allows users to draw a camera motion path that depicts scene switching (E2). Those words appearing in the successive scenes will move along the drawn path.

4.2.5 *Dimension II: Cast of roles*

Group mode enables users to lasso multiple words to generate a group (E3). The group is highlighted with a background color based on the color of the largest word.

Description mode (E4) allows to drag description words (e.g., *weak* and *old*) to the roles (Fig. 5). The position and color of these description words will be updated according to the role.

4.2.6 *Dimension III: Event representation*

Dialog mode enables users to specify a dialog (E6) between roles. In this mode, users first specify two speakers and lasso the dialog contents to the auto-generated arrow between two speakers.

Give mode provides another typical event (E5) design (Fig. 3a-3). In this mode, users need to specify the giver and receiver successively and drag the given word to the auto-generated arrow. Words then animate from the giver to the receiver in the path.

Action mode specifies a common event (E5). The system draws an arrow or dashed links between two specified roles to demonstrate the direction. The words that explain the concrete events are then dragged to the arrow.

4.2.7 Dimension IV: Visual embellishments

Given G3, the system also provides interactions to enable users to incorporate visual embellishments in the animated word clouds, such as icons (E7). Users are allowed to change the color of the words, as well as replacing them with icons (Fig. 4c-1).

4.2.8 Layout algorithm

In accordance with G2, the system should generate layouts automatically according to different interactions. The individual layout at each key frame should be coherent with the overall visual representation that illustrates the whole story. Besides, the successive word clouds should preserve temporal coherence with smooth visual transition. However, current algorithms to generate word clouds cannot satisfy both requirements meanwhile. For example, the algorithms (e.g., Viégas et al. 2009) pursuing compactness of word clouds cannot provide enough space for the animation. Second, most dynamic word clouds layouts (e.g., Cui et al. 2010; Wu et al. 2011) maintain the global coherence, whereas the word positions in our scenario only indicate the local semantics in two successive key frames. Our proposed layout algorithm is implemented based on the force-directed graph using D3.js and adapted to keep semantic relations and preserve stability in successive key frames. In our situation, the layout is expected to satisfy several requirements.

- *R1. Keep the hierarchical structure of words* Words in a key frame are assigned with different narrative elements, presenting a story hierarchy. For example, multiple roles compose a group, and a context may contain several groups. The algorithm should ensure both words and their hierarchical structures overlapping-free.
- *R2. Maintain relative stability of word clouds* Every time we add interactions, the force-directed algorithm may generate a new layout with randomness. To avoid meaningless animation as required in Sect. 3.1.3, we need a stability algorithm.
- *R3. Reduce overlapping during animation* The whole animation should be easy to follow and without much clutter. However, the force-directed algorithm can only ensure that words are not overlapped, but the motion paths of words are overlooked. A algorithm to reduce overlaps between motion path and words is required.

Based on these requirements, we implement an algorithm at each key frame as shown in Fig. 6. First, we build a content tree for the words in a key frame based on a top-down approach (Fig. 6a). Each node of this tree represents a word. A complete tree consists of a root node, a context layer, group layer, role layer, and description layer. The content tree is consistent with the semantic hierarchy in the design space (R1). For example, the context node (E1) contains all the nodes in this context and is appended to the root node. The child nodes of the context are the groups (E3) which are composed of multiple roles. We place the events (E5) at the same layer of the role layer, since events are closely related to roles and required to arrange the positions simultaneously. The description layer is appended to the role layer, since description (E4) provides additional information of the role.

To maintain stability (R2), we record three pieces of information of each node: (1) the size of the word; (2) the position of the word in the previous frame (empty if it is a new word); and (3) should it be fixed, determined by the word-fixed algorithm. Be default, we fix all the words. A word will be unlocked (Fig. 6b), if it meets any of the following conditions: (1) it is a new word; (2) it moves to another word; (3) it joins or leaves a group; (4) it moves to another context.

Then, we implement a hierarchical force-directed algorithm (R1) in a bottom-up approach based on the constructed content tree (Fig. 6c). We first find a subtree with a height of two and then generate a local layout for this subtree based on the force-directed algorithm. We exert a repulsive force on every two words

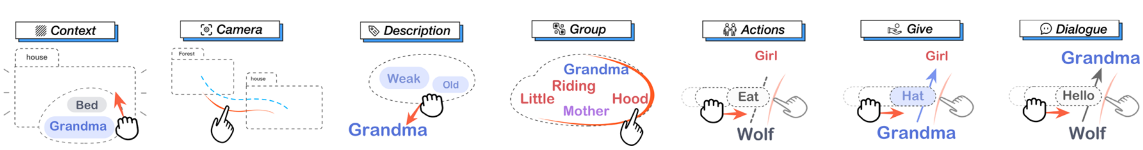

Fig. 5 An overview of interactions. The first two interactions act with the context dimension, the following two controls the role dimension, and the other two provide interactions to act with the event dimension

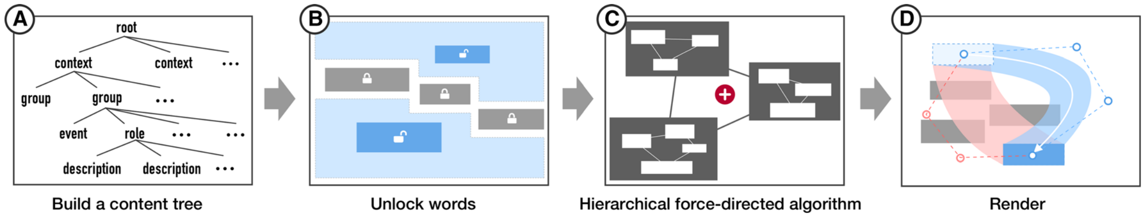

Fig. 6 The story-oriented auto-layout algorithm. **a** Building a content tree in a top-down approach; **b** unlocking words according to the word-fixed rules; **c** implementing a hierarchical force-directed algorithm in a bottom-up approach based on the content tree; **d** rendering the motion path with a heuristic method

and a central attraction force on all words. The weight of the repulsive force is adaptive depending on their narrative elements. The fixed words will be locked to their original positions. Then, we replace this subtree with a pseudonode, of which the position and size are decided by the layout of the subtree. Meanwhile, we solve the word occlusion problem in each subtree following Wang et al.'s (2018a) method, where the neighboring repulsive forces push away two collided words. The above operations are performed recursively until there is only one node in the content tree. At last, we re-calculate the position of each node by superimposing the relative displacement from the root node to the leaves.

Finally, we render the motion path on the content canvas (Fig. 6d). We propose a heuristic algorithm to derive a motion path with minimum overlapping areas (R3). There are two situations which should render a motion path: (1) draw an arrow and (2) move the word. The origin and destination of this path have been calculated based on the hierarchical force-directed layout. We use a quadratic or cubic Bezier curve as the motion path between these two end points. Specifically, we generate several different sets of control points and produce several Bezier curves correspondingly. The motion path considers the size of the moving word along the curve, as shown in the blue area in Fig. 6d. To calculate the overlapping areas, we first judge whether there are collisions between the motion path and the words around by comparing the positions of the path border and words' rectangular vertices. When there exist collisions, we calculate approximate trapezoidal areas as the overlapping areas based on the intersections of the path border and words' rectangle. Finally, we choose a motion path with minimum overlapping areas with other words on its motion path. Although we cannot guarantee no overlapping areas, it can help alleviate the issue with relatively good performance.

5 Evaluation

In this section, we demonstrate the expressiveness of *DancingWords* with several example animated stories and show the usability through a reproduction study with feedback from end users.

5.1 Example animated word clouds

In this section, we crafted four animated stories to demonstrate the expressiveness of our tool and exemplify a diversity of usage scenarios, including fairy tales, movies, and news summary. Examples are available at https://dancingwords.github.io/.

With the automatic layout algorithm, *DancingWords* can reduce users' workload in creating animated word clouds and also demonstrate good performance (G2). For example, the algorithm well preserves the hierarchical structure of the contents in Fig. 7b-3. All roles appear within the same shire context (Fig. 7b-1) and distributed according to different groups (Fig. 7b-3). Roles in the same group (i.e., Frodo, Sam, Merry, Pippin, and Aragorn) gather together within the light blue background and do not interact with other groups. Moreover, it preserves the relative spatial coherence between leading roles among the successive key frames (e.g., Fig. 7c-2, c-3). The final rendering algorithm shows a promising effect. For example, arrows in Fig. 7a-1, c-1 are both rendered with minimum overlapping. During the process, there are no obvious distracting overlaps in the videos. In addition, during the case studies, we find that the algorithm greatly saves the creation time so that we can put more efforts on sketching the expression, instead of adjusting the positions and layouts.

Besides, to show the capability of our tool for crafting complex stories, we reconstructed the story from Cui et al. (2010) which summarized 20-year news about Apple. We selected 25 most frequent words in each

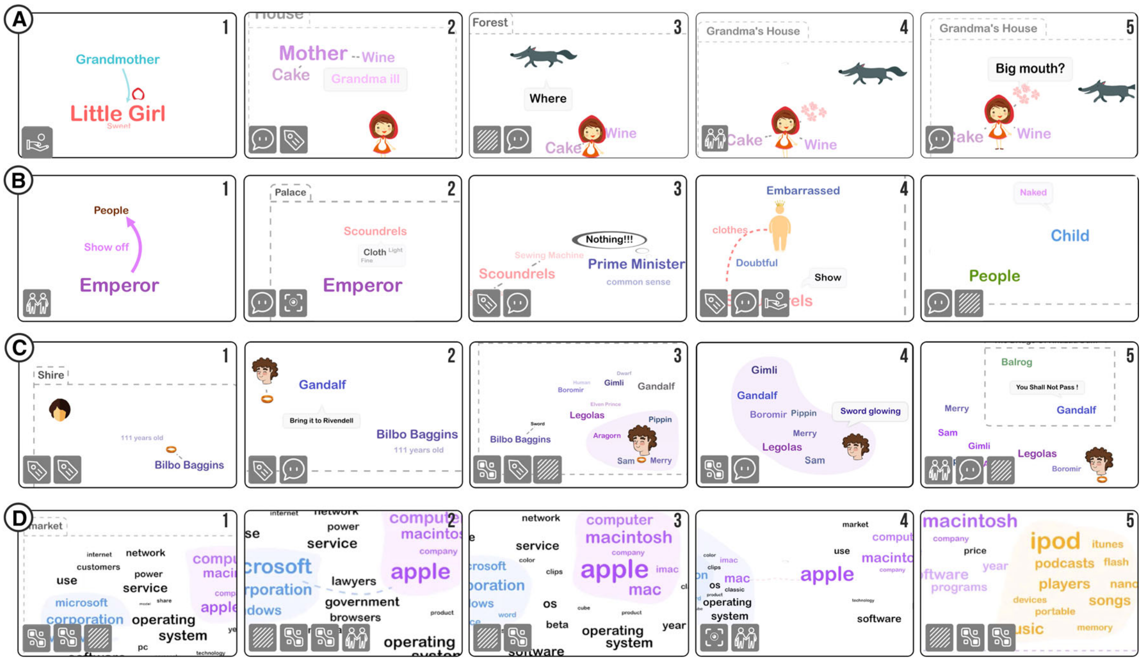

Fig. 7 Examples created with *DancingWords*. **a**, **b** Two fairy tales, *Little Red Riding Hood* (**a**) and *The Emperor's New Clothes* (**b**). **c** The beginning of Frodo Baggins's adventure from *The Lord of the Rings*. **d** 20-year news summary about Apple reconstructed from Cui et al. (2010). The small icons on the bottom left of each key frame represent the interactions involved in the frame. We provide part of screenshots from our videos, and the whole animated stories can be found on the supplementary materials

frame of their case figures. Our example aims to reveal the relations of words through setting up the narrative elements and animating the dynamic process. In the early years, Apple was competing with Microsoft in the computer market, thus being put together in the same context (Fig. 7d-1). During this period, the font size of the word *Microsoft* was changing due to the anti-monopoly lawsuit with government and other technology companies. Therefore, we add a connection between Microsoft and Apple through the action mode and put related words around the link (Fig. 7d-2). Later, when Apple enriched the diversity of their products with iPod and iPhone, we dissolved the previous context including Microsoft and Apple (Fig. 7d-3) and created a new one for Apple only. The design of camera moving effects (Fig. 7d-4) aims to show that people's focus on Apple was not binding with Microsoft and diverting to new aspects in music players and phones (Fig. 7d-5). Through the animation, we directly present the development of Apple.

By incorporating visual embellishments (G3), *DancingWords* can greatly improve the quality of the final animated word clouds and engage viewers. For example, we add character figures in Fig. 7. The dialog bubbles are also designed to be close to animation.

5.2 Usability study

To evaluate the usability of *DancingWords*, we conducted a reproduction study with users who did not take part in the previous study.

Participants We recruited 10 participants from a university. Seven of them are undergraduate (p1–p2) and postgraduate (p3–p7) students with text visualization experience (four males, three female; aged from 21 to 25), while the other three were students majored in Product Design (p8–p10) skilled in creating videos (two males, one female; aged from 20 to 24).

Procedure Participants were asked to complete two replication tasks on the Google Chrome browser on a PC with a 1920×1080 display. (1) The first task serves as a training task where participants were asked to reproduce the example video about *Little Red Riding Hood*. We first provided a 10-min tutorial about our system and provided guidance timely. (2) The second task required participants to create animated word clouds for the story, *The Emperor's New Clothes*, without our assistance and example videos. In this task, they needed to arrange the narrative structure in the video by themselves. We chose these two well-known stories in order to save participants' time to get familiar with stories and directly create the animated stories. Finally, we collected their feedback through a questionnaire using five-point Likert scale (1—strongly disagree, 5—strongly agree) and a semi-structured interview.

Results All participants completed two replication tasks in 90 min. Specifically, the first took $\sim$ 30 min (including the tutorial), while the second task lasted $\sim$ 60 min. The time used with our system to create an animated story is greatly shorter than that designers spent in the preliminary study (5 h). Based on the ratings given by participants after the study, the overall feedback is promising. They generally were satisfied with the tool, where all scores are beyond 4 in a 5-point Likert scale. Participants were mostly pleased with these intuitive story-oriented interactions (4.5/5) and agreed that they covered basic narrative elements (4.375/5). Moreover, adding visual embellishments was regarded as a useful technique to enhance the appeal (4.375/5). It indicated the strengths of *DancingWords* to create animated word clouds for storytelling based on the structured design space. The auto-layout is relatively weak (4.0/5). Opinions for the automatic layout algorithm vary among participants. Most general users (p1, p3–p6) said that they were satisfied with the results. Participants with rich video-creation experience (p8–p10) admitted that the automatic layout algorithm could greatly reduce the workload, although it did not always promise a perfect layout. P8 argued that he had to adjust the detailed layout by himself. It was reasonable that professional designers required flexibility to realize elaborate designs, which could be hard to be satisfied with auto-layouts. Overall, they agreed that the system could assist the creation process (4.125/5) and produce various stories (4.125/5). In addition, they proposed specific use cases of the generated animated stories, such as course presentation, report summarization, and quick demo. P9 said, "It allows to quickly create prototype videos for idea brainstorming and communication."

Suggestions for improvements Afterward, participants provided constructive comments on *DancingWords* for further improvements. First, they required a module to help identify words, e.g., processing text and extracting words according to the narrative elements. During the study, we observed that they spent much time on finding words in the left panel, since the current system asks users to decide the word relations themselves. It could be helpful if the system provides semantic categories for words. Another shortcoming is that the automatic layout does not always meet expectations. Users need to re-generate the layout to achieve a better effect. Other minor issues include incorporating more designs for contexts and events.

6 Discussion

This section discusses using animated word clouds for storytelling, as well as the implications and limitations of this work.

6.1 Animated word clouds for storytelling

In this work, we explore the possibility of animated word clouds to tell stories. We discuss our final representations compared to traditional word clouds, as well as the generalizability of the tool to create animated words and figures for storytelling.

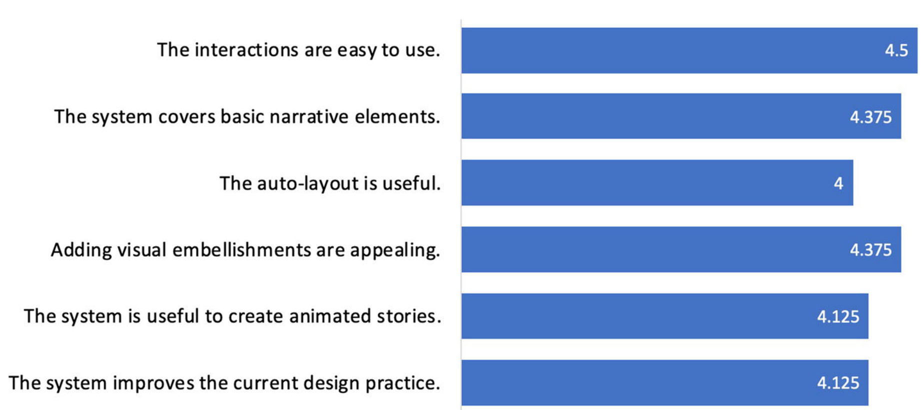

Fig. 8 Overall ratings of the usability study

During our exploratory study, we preserve the major feature of word clouds that encode word significance into font size. Although it is hard to compare the size of icons and words, users prefer to replace the major characters with icons based on our observations. Thus, the visual significance of icons is consistent with the original data. Another problem is that word clouds in each frame are not compact enough. Prior studies in word clouds (Barth et al. 2014) have evaluated that semantic word clouds cannot be as compact as Wordle. Our work aims to reveal the narrative relations among words, which preserves semantic and temporal relations simultaneously. Proper whitespace can help identify word groups and leave space for animation. The goal of our current automatic layout algorithm is to satisfy the design space and does not consider the compactness. But it is feasible to further improve the compactness after the hierarchy force-directed algorithm.

Another emerging discussion is about applying the tool to create animated words and figures for storytelling. The integration of figures and words in animated storytelling not only preserves the readability of text, but also increases the appeal and retention with figures. *DancingWords* provides an easy authoring of animated words and figures, which is appreciated during the user study. Moreover, we can improve the generalizability by enriching the designs of different narrative elements. For example, besides basic arrows to display action events (E5), the tool can provide more alternative designs or allow users to upload their own figures to convey the event. Future work can also explore how users combine words and icons for storytelling and improve the tool iteratively.

6.2 Implications and limitations

Implications The entire design process from understanding the design space to developing a proof-of-concept authoring tool has shown several implications. First, our own trials to produce examples of animated word clouds provide a preliminary characterization of the problem domain. Communicating stories with animated word clouds has a large design space given the variety of visual designs and narrative elements. We narrowed down to three essential elements which can compose a typical story based on literature review and our own practice. We further validated our designs and enriched the space by cooperating with designers. The iterative brainstorming and discussions distilled the design space and illuminated the authoring tool. Second, this work presents a storytelling authoring tool for general users. It follows a structure-based design paradigm that encapsulates common interactions in line with identified narrative elements. Users are able to directly interact with narrative elements and quickly produce animated stories.

Limitations Our work has several limitations. First, the performance of these generated animated word clouds should be further evaluated. The experts and designers in the exploratory study have agreed that current narrative elements show the structure of stories. Participants in the usability study validate that using current designs is able to generate animated stories. However, the number of these participants is limited. A further study with a broad range of viewers for specific animation effects should be conducted in the future. Second, the automatic layout algorithm is based on a heuristic method. We define and modify the different levels of forces through multiple iterations to achieve an expected result. Further extension of the current algorithm could be explored to support a more reasonable weight adjustment. Third, *DancingWords* currently requires manual efforts to select words. Incorporating advanced machine learning methods or NLP algorithms (Zhou 2019; Chawla et al. 2020) to help preprocess documents and extract relations (Yang et al. 2017; Wong and Zhang 2018; Xia et al. 2016) is worth studying.

7 Conclusion

In this paper, we explore the possible designs in animated word clouds to tell stories. Through a systematic iterative approach with designers, we collect examples of animated word clouds and summarize the design space for different narrative elements. Based on the design space, we further implement an authoring tool, *DancingWords*, which incorporates a set of intuitive interactions for users to directly manipulate the narrative elements and provides an automatic algorithm to adjust the layout. The expressiveness and usability of the tool are demonstrated through example animated stories and a reproduction user study.

There are several avenues for future work. First, we plan to deploy *DancingWords* on a public website and collect a broader range of examples from general users, as well as their comments for such animated word clouds. Second, we intend to incorporate advanced NLP algorithms or sentiment analysis methods into the tool, which can help extract the outline of stories and recommend words.

Acknowledgements The work was supported by NSFC (61761136020), NSFC-Zhejiang Joint Fund for the Integration of Industrialization and Informatization (U1609217), Zhejiang Provincial Natural Science Foundation (LR18F020001). This project was also partially funded by HKUST SSC grant F0707 and Microsoft Research Asia (FY20-Research-Sponsorship-252 and No. MRA19EG02).

References

Amini F, Riche NH, Lee B, Hurter C, Irani P (2015) Understanding data videos: looking at narrative visualization through the cinematography lens. In: Proceedings of the ACM conference on human factors in computing systems, pp 1459–1468. https://doi.org/10.1145/2702123.2702431

Amini F, Riche NH, Lee B, Monroy-Hernández A, Irani P (2017) Authoring data-driven videos with DataClips. IEEE Trans Vis Comput Graph 23(1):501–510. https://doi.org/10.1109/TVCG.2016.2598647

Amini F, Riche NH, Lee B, Leboe-McGowan J, Irani P (2018) Hooked on data videos: assessing the effect of animation and pictographs on viewer engagement. In: Proceedings of the international conference on advanced visual interfaces, pp 21:1–21:9. https://doi.org/10.1145/3206505.3206552

Bach B, Kerracher N, Hall KW, Carpendale S, Kennedy J, Riche NH (2016) Telling stories about dynamic networks with graph comics. In: Proceedings of the ACM conference on human factors in computing systems, pp 3670–3682. https://doi.org/10.1145/2858036.2858387

Bal M (2009) Narratology: introduction to the theory of narrative, 3rd edn. University of Toronto Press, Toronto

Barth L, Fabrikant SI, Kobourov SG, Lubiw A, Nöllenburg M, Okamoto Y, Pupyrev S, Squarcella C, Ueckerdt T, Wolff A (2014) Semantic word cloud representations: hardness and approximation algorithms. In: LATIN: theoretical informatics-Latin American symposium, vol 8392, pp 514–525. https://doi.org/10.1007/978-3-642-54423-1_45

Bateman S, Mandryk RL, Gutwin C, Genest A, McDine D, Brooks CA (2010) Useful junk? The effects of visual embellishment on comprehension and memorability of charts. In: Proceedings of the ACM conference on human factors in computing systems, pp 2573–2582. https://doi.org/10.1145/1753326.1753716

Blazer L (2016) Animated storytelling: simple steps for creating animation and motion graphics. Peachpit Press, Berkely

Brehmer M, Lee B, Bach B, Riche NH, Munzner T (2017) Timelines revisited: a design space and considerations for expressive storytelling. IEEE Trans Vis Comput Graph 23(9):2151–2164

Bulterman DCA, Hardman L (2005) Structured multimedia authoring. ACM Tran Multimed Comput Commun Appl 1(1):89–109. https://doi.org/10.1145/1047936.1047943

Cao R, Dey S, Cunningham A, Walsh J, Smith RT, Zucco JE, Thomas BH (2020) Examining the use of narrative constructs in data videos. Vis Inform 4(1):8–22. https://doi.org/10.1016/j.visinf.2019.12.002

Chawla P, Hazarika S, Shen H (2020) Token-wise sentiment decomposition for ConvNet: visualizing a sentiment classifier. Vis Inform 4(2):132–141. https://doi.org/10.1016/j.visinf.2020.04.006

Chi M, Lin S, Chen S, Lin C, Lee T (2015) Morphable word clouds for time-varying text data visualization. IEEE Trans Vis Comput Graph 21(12):1415–1426. https://doi.org/10.1109/TVCG.2015.2440241

Coulter CA, Smith ML (2009) The construction zone: literary elements in narrative research. Educ Res 38(8):577–590. https://doi.org/10.3102/0013189X09353787

Cui W, Wu Y, Liu S, Wei F, Zhou MX, Qu H (2010) Context-preserving dynamic word cloud visualization. In: Proceedings of IEEE Pacific visualization symposium, pp 121–128. https://doi.org/10.1109/PACIFICVIS.2010.5429600

Felix C, Franconeri S, Bertini E (2018) Taking word clouds apart: an empirical investigation of the design space for keyword summaries. IEEE Trans Vis Comput Graph 24(1):657–666. https://doi.org/10.1109/TVCG.2017.2746018

Hearst M, Pedersen E, Patil LP, Lee E, Laskowski P, Franconeri S (2019) An evaluation of semantically grouped word cloud designs. In: IEEE transactions on visualization and computer graphics, p 1

Heer J, Robertson GG (2007) Animated transitions in statistical data graphics. IEEE Trans Vis Comput Graph 13(6):1240–1247. https://doi.org/10.1109/TVCG.2007.70539

Hu M, Wongsuphasawat K, Stasko J (2017) Visualizing social media content with SentenTree. IEEE Trans Vis Comput Graph 23(1):621–630

Kerzner E, Goodwin S, Dykes J, Jones S, Meyer MD (2019) A framework for creative visualization-opportunities workshops. IEEE Trans Vis Comput Graph 25(1):748–758. https://doi.org/10.1109/TVCG.2018.2865241

Kim NW, Schweickart E, Liu Z, Dontcheva M, Li W, Popovic J, Pfister H (2017) Data-driven guides: supporting expressive design for information graphics. IEEE Trans Vis Comput Graph 23(1):491–500. https://doi.org/10.1109/TVCG.2016.2598620

Kim NW, Riche NH, Bach B, Xu G, Brehmer M, Hinckley K, Pahud M, Xia H, McGuffin MJ, Pfister H (2019) DataToon: drawing dynamic network comics with pen + touch interaction. pp 1–12. https://doi.org/10.1145/3290605.3300335

Latif S, Beck F (2019) Interactive map reports summarizing bivariate geographic data. Vis Inform 3(1):27–37. https://doi.org/10.1016/j.visinf.2019.03.004

Li C, Dong X, Yuan X (2018) Metro-Wordle: an interactive visualization for urban text distributions based on wordle. Vis Inform 2(1):50–59. https://doi.org/10.1016/j.visinf.2018.04.006

Lu J, Wang J, Ye H, Gu Y, Ding Z, Xu M, Chen W (2020) Illustrating changes in time-series data with data video. IEEE Comput Graph Appl 40(2):18–31

Mascelli JV (2005) The Five C's of cinematography: motion picture filming techniques, 1st edn. Silman-James Press, Hollywood

Max W (1923) Untersuchungen zur lehre von der gestalt. Psycol Forsch 3:301–350

Mei H, Chen W, Ma Y, Guan H, Hu W (2018a) VisComposer: a visual programmable composition environment for information visualization. Vis Inform 2(1):71–81. https://doi.org/10.1016/j.visinf.2018.04.008

Mei H, Ma Y, Wei Y, Chen W (2018b) The design space of construction tools for information visualization: a survey. J Vis Lang Comput 44:120–132. https://doi.org/10.1016/j.jvlc.2017.10.001

Onoue Y, Kukimoto N, Sakamoto N, Koyamada K (2016) E-Grid: a visual analytics system for evaluation structures. J Vis 19(4):753–768. https://doi.org/10.1007/s12650-015-0342-6

Palmer SE (1992) Common region: a new principle of perceptual grouping. Cogn Psychol 24(3):436–447. https://doi.org/10.1016/0010-0285(92)90014-S

Segel E, Heer J (2010) Narrative visualization: telling stories with data. IEEE Trans Vis Comput Graph 16(6):1139–1148. https://doi.org/10.1109/TVCG.2010.179

Tang T, Tang J, Hong J, Yu L, Ren P, Wu Y (2020) Design guidelines for augmenting short-form videos using animated data visualizations. J Vis. https://doi.org/10.1007/s12650-020-00644-z

Thomas F, Johnston O (1995) The illusion of life: Disney animation. Hyperion, New York

Tversky B, Morrison JB, Bétrancourt M (2002) Animation: can it facilitate? Int J Hum Comput Stud 57(4):247–262. https://doi.org/10.1006/ijhc.2002.1017

Viégas FB, Wattenberg M (2008) Timelines: tag clouds and the case for vernacular visualization. Interactions 15(4):49–52. https://doi.org/10.1145/1374489.1374501

Viégas FB, Wattenberg M, Feinberg J (2009) Participatory visualization with wordle. IEEE Trans Vis Comput Graph 15(6):1137–1144. https://doi.org/10.1109/TVCG.2009.171

Wang Y, Chu X, Bao C, Zhu L, Deussen O, Chen B, Sedlmair M (2018a) EdWordle: consistency-preserving word cloud editing. IEEE Trans Vis Comput Graph 24(1):647–656. https://doi.org/10.1109/TVCG.2017.2745859

Wang Y, Zhang H, Huang H, Chen X, Yin Q, Hou Z, Zhang D, Luo Q, Qu H (2018b) InfoNice: easy creation of information graphics. In: Proceedings of the ACM conference on human factors in computing systems, pp 335:1–335:12. https://doi.org/10.1145/3173574.3173909

Wang Y, Chu X, Zhang K, Bao C, Li X, Zhang J, Fu C, Hurter C, Deussen O, Lee B (2020) ShapeWordle: tailoring wordles using shape-aware archimedean spirals. IEEE Trans Vis Comput Graph 26(1):991–1000

Wattenberg M, Vigas FB (2008) The word tree, an interactive visual concordance. IEEE Trans Vis Comput Graph 14(6):1221–1228

Wong J, Zhang X (2018) MessageLens: a visual analytics system to support multifaceted exploration of MOOC forum discussions. Vis Inform 2(1):37–49. https://doi.org/10.1016/j.visinf.2018.04.005

Wor (2020) Word art website. https://wordart.com. Accessed 1 May 2020

Wu Y, Provan T, Wei F, Liu S, Ma K (2011) Semantic-preserving word clouds by seam carving. Comput Graph Forum 30(3):741–750. https://doi.org/10.1111/j.1467-8659.2011.01923.x

Xia J, Chen W, Hou Y, Hu W, Huang X, Ebert DS (2016) DimScanner: a relation-based visual exploration approach towards data dimension inspection. In: IEEE conference on visual analytics science and technology (VAST), pp 81–90. https://doi.org/10.1109/VAST.2016.7883514

Yang Y, Yao Q, Qu H (2017) VISTopic: a visual analytics system for making sense of large document collections using hierarchical topic modeling. Vis Inform 1(1):40–47. https://doi.org/10.1016/j.visinf.2017.01.005

Zhou Z (2019) Abductive learning: towards bridging machine learning and logical reasoning. Sci China Inf Sci 62(7):76101:1–76101:3. https://doi.org/10.1007/s11432-018-9801-4