Towards Semantically Aware Word Cloud Shape Generation

K. J. Kevin Feng*
kjfeng@uw.edu
University of Washington
Seattle, USA

Alice Gao*
atgao@cs.washington.edu
University of Washington
Seattle, USA

Johanna S. Karras*
jskarras@cs.washington.edu
University of Washington
Seattle, USA

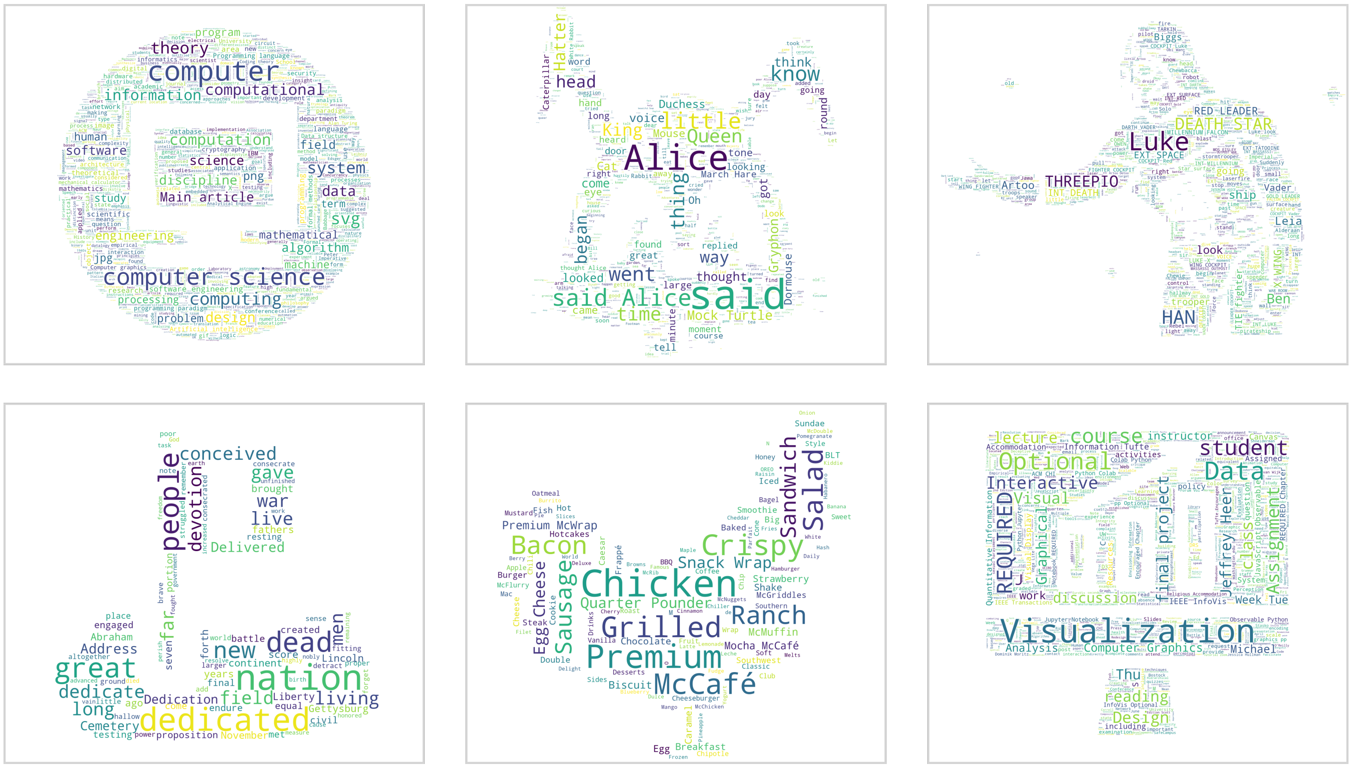


Figure 1: Word clouds generated by our tool. The input texts, from left to right, are as follows: the Wikipedia page for Computer Science, full text of Alice in Wonderland, the movie script for Star Wars Episode IV: A New Hope, speech text of Abraham Lincoln's Gettysburg Address, the McDonald's menu, and syllabus for a graduate data visualization class.

ABSTRACT

Word clouds are a data visualization technique that showcases a subset of words from a body of text in a cluster form, where a word's font size encodes some measure of its relative importance—typically frequency—in the text. This technique is primarily used to help viewers glean the most pertinent information from long text documents and to compare and contrast different pieces of text. Despite their popularity, previous research has shown that word cloud designs are often not optimally suited for analytical tasks such as summarization or topic understanding. We propose a solution for generating more effective visualization technique that shapes the word cloud to reflect the key topic(s) of the text. Our method automates the processes of manual image selection and masking required from current word cloud tools to generate shaped word clouds, better allowing for quick summarization. We showcase two approaches using classical and state-of-the-art methods. Upon successfully generating semantically shaped word clouds using both methods, we performed preliminary evaluations with 5 participants. We found that although most participants preferred shaped word clouds over regular ones, the shape can be distracting and detrimental to information extraction if it is not directly relevant to the text or contains graphical imperfections. Our work has implications on future semantically-aware word cloud generation tools as well as efforts to balance visual appeal of word clouds with their effectiveness in textual comprehension.

*Authors contributed equally to this work.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
UIST '22 Adjunct, October 29-November 2, 2022, Bend, OR, USA
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9321-8/22/10.
https://doi.org/10.1145/3526114.3558724

CCS CONCEPTS

• **Human-centered computing → Visualization techniques**.

KEYWORDS

Text visualization, word clouds, multimodal computer vision

ACM Reference Format:

K. J. Kevin Feng, Alice Gao, and Johanna S. Karras. 2022. Towards Semantically Aware Word Cloud Shape Generation. In *The Adjunct Publication of the 35th Annual ACM Symposium on User Interface Software and Technology (UIST '22 Adjunct), October 29-November 2, 2022, Bend, OR, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3526114.3558724

1 INTRODUCTION

Some of the world's most interesting data lie in unstructured text: news articles, film manuscripts, journal publications, and much more. We consume text primarily through reading and skimming, but our reading speed may be a bottleneck when we try to quickly capture the essence of a piece of text at a glance. Text visualizations can help us in such scenarios; however, text is inherently difficult to visualize due to its nominal nature [7].

Word clouds are a popular data visualization tool used to summarize textual information in a visually appealing manner. The first digitally typeset word clouds are believed to have been conceived by Douglas Coupland for his 1995 novel *Microserfs*, in which he tried to artistically depict what a "computer's dream would look like" [3]. To do so, he selected a series of words and phrases from a body of text, weighted them according to their frequency of appearance, and varied their font size based on that weighting. Coupland then went on to work at technology magazines such as *Wired*, where word clouds were popularized for advertising purposes in the 2000s [3].

Word clouds, or "wordles", are still popular today due to their visual appeal. Plus, there are several benefits of word clouds: they allow viewers to quickly find relevant words and themes in a very large text corpus; they are visually appealing and are understandable to viewers who do not know anything about the original text; and they are easy to share with others through digital channels such as web articles and social media.

However, the amount of information a viewer can obtain from a quick glance at a word cloud remains limited due to the fact that it takes time to parse each individual word. Additionally, viewers must also reason about the visual encodings in the word cloud: *What do the highlighted words reveal about the main topic of the document? What kind of text does this word cloud correspond to?* In fact, formal usability studies on word clouds showed that irregular font sizes and Wordle-style layout can be detrimental to text understanding [6, 19]. In cases of tag-finding tasks, studies showed that users preferred a plain alphabetically ordered list with no font size variation over word clouds [15]. For word lookup tasks, users tended to prefer a search box for finding specific keywords rather than trying to locate them in a word cloud [9, 13].

Our work seeks to improve upon the effectiveness of the word cloud design in conveying key ideas in the text, while enhancing the aesthetic appeal of the word cloud for more creative expression. Our method improves the ease of topic recognition and effectiveness of word clouds by shaping the word cloud to reflect the key topics of the text. Unlike the standard word cloud that comes in an amorphous cloud-like shape (hence the name), our word clouds immediately give viewers the overall topic of the text. While it is possible to generate shaped word clouds using current tools, it is a manual process that requires users to choose their own image to use as a mask. This may be burdensome as (1) not all images are well-suited to be masks and may require substantial image editing, and (2) users have to leave their word cloud generation workflow to find suitable mask images. Our work focuses on streamlining this process to allow users to easily create semantically-aware, shaped word clouds.

In summary, this work contributes the following:

(1) A pipeline by which one can automatically generate a word cloud in the shape of a relevant idea in the text.
(2) Two approaches to obtain a mask for creating the shaped word cloud.
(3) A brief evaluation of the generated word clouds and its implications for future work in this area.

2 RELATED WORKS

2.1 Alternative Word Cloud Designs

One group of previous work seeks to improve upon the effectiveness of word clouds using word bubbles. Works related to word bubbles represent the top n most frequent words in their own circle, or "bubble", whose size indicates the word's relative significance in the text. However, bubbles do not convey new information about the text's context, thus facing similar limitations to the original word cloud design.

Other existing research seeks to draw semantic connections between words. For example, Parallel Tag Clouds [2] puts words of interest into their own semantically similar column then creates links between similar words of different columns. WordBridge [11] uses a composite word cloud representation with a node link diagram representing the context and relationship between entities. These links can also consist of their own miniature word cloud as well. Some works such as [8] look at the improved word cloud design for topic recognition tasks while maintaining the aesthetic visualization of the word cloud by grouping semantically similar words together. While these design alternatives make it easy for the careful viewer to parse semantic relationships between words in the word cloud, it remains difficult for the casual viewer to do so at a glance. Additionally, a large-scale study of the Wordle word cloud creation tool [**?**] showed that most users were not interested in studying semantic connections, but wanted to create an artifact for creative expression using a piece of text they were already familiar with [22]. We design our tool with this demographic of users in mind.

2.2 Word Clouds for Creative Expression

Jonathan Feinberg released the the popular word cloud creator Wordle in 2008 [**?**], and since then, it has been used over 7 million times to create word clouds. However, Feinberg himself stated that the Wordle-style layout may be ineffective for intensive analytical tasks, but its value lay in its simple, playful nature and typographic liveliness [20]. While at IBM research, where he created Wordle's layout algorithm, Feinberg and colleagues conducted a survey of

4306 Wordle users and found that most users created word clouds for pleasure with texts that were already familiar to them [22]. In other words, Wordle word clouds were not for exploring and analyzing new texts, but for creating artistic artifacts for enhanced personal expression of familiar text. Another major use case surfaced in the study is in education: teachers used word clouds as a way to engage students in the subject material at hand, using words to pique students' interest and curiosity.

Other researchers have since focused on the potential of word clouds as an informal, creative tool. Mueller created a word cloud generator for Python that lets users shape word clouds into a mask of their choice [16]. ManiWordle is a tool that provides users with custom visual manipulations over standard Wordle word clouds, including options to change typography, colour, and composition for not just the overall layout, but also individual words [12]. Wordle-Plus supports direct manipulation of Wordle development through pen and touch interactions [10]. There has also been some work on how to improve the comprehensibility of word clouds through modifying layout based on semantic relationships. Hearst et al. [8] ran a series of controlled experiments and showed that arranging words into visually distinct zones based on semantic similarity allowed for more effective understanding of underlying topics compared to traditional word cloud layouts. EdWordle [24] allows the user to keep certain groups of words together as they manipulate the Wordle layout. Wang et al. showed that a semantically organized word cloud can lower mental demand in time-constrained theme identification tasks [23].

We further enhance the ability of word clouds to be a tool for creative expression while improving its communication of key topics by morphing the word cloud into a semantically relevant, visually engaging form.

3 DESIGN GOALS

Our primary design goals are:

- To accurately convey the main idea of the input text using the shape of the word cloud. This way, a viewer can glean the main idea of a text without having to read into the word cloud, and the word cloud itself may be more expressive and engaging.
- To convey the relative importance of words inside the word cloud using font size, in order for the viewer to grasp the core sub-topics of the text and their relative importance.
- To create a fully automatic pipeline to generate semantically shaped word clouds. This will make it easy to use for summarization and artistic purposes. We explore how state-of-the-art computer vision techniques can help us achieve this goal, as well as more traditional tools.

4 METHODS

Our method consists of (1) generating a topic search query from a text document, (2) selecting an image that captures the essence of the text using the query and text-to-image search, (3) converting the image into a mask for the word cloud, and (4) creating a shaped word cloud from the text document and mask. See Figure 2 for our pipeline diagram and an example.

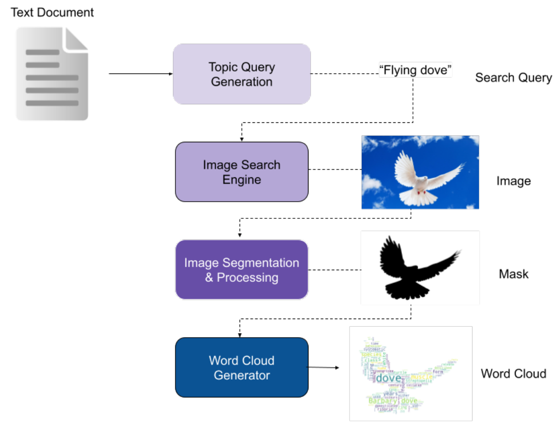


Figure 2: Pipeline overview for automatically generating shaped word clouds from a text document.

4.1 Generate Search Query for the Mask

To generate the search query for the mask, we first take in the desired text we wish to generate the word cloud for. We use Python's `spaCY` NLP library to tokenize the text. We then count the frequency of each word in the text and return the top 10 words. These words combined, along with "mask" or "silhouette" appended to the end, serve as a query for CLIP (one of our image search approaches), while we use only the top word as a query in `The Noun Project` API (our other image search approach).

4.2 Image Search and Processing

As indicated earlier, we implemented two approaches to image search: The first uses `The Noun Project` API [**?**] and the second uses a pre-trained, multi-modal transformer model, CLIP [18]. We then also process the images into masks with a technique depending on which image search approach was used. If the mask found by either method is particularly poor (e.g. a speck of maskable area on a blank canvas), an unshaped word cloud is created.

4.2.1 Approach 1: The Noun Project. `The Noun Project` is an icon and image library contributed to by a global community of designers. Their API allows for the querying of icons similar to how one would search for an icon in a search engine. `The Noun Project` stood out from other icon library APIs with its powerful text-to-image inference and association: by querying on the term "Luke", we were able to obtain Star Wars related icons, whereas other APIs returned nothing or an irrelevant icon of a generic man.

An effective mask for the word cloud should have large areas of black for the word cloud to mask onto. Additionally, the mask should not have too many "holes"; otherwise the word cloud would not look cohesive. To account for this, we took the top 3 darkest icons from the response to our API call using average pixel value calculations and then out of the 3, we selected the image with the fewest connected components, computed using OpenCV's `connectedComponents()` function. We then scaled up the image and re-applied a black-and-white filter on it to create clean boundaries for the mask.

4.2.2 Approach 2: CLIP. CLIP is a state-of-the-art multimodal transformer model, originally released by OpenAI, that jointly learns to represent texts and images [18]. CLIP achieves state-of-the-art results on a variety of computer vision and natural language processing tasks with minimal additional fine-tuning. These tasks include image search, classification, and generation. In our work, we apply CLIP to image search with a natural-language query. Given time and computational resource constraints, we take advantage of a pre-trained CLIP model trained on the Unsplash dataset, an open-source library containing over 250,000 images [1, 5].

After we obtain the image, we segment the foreground object to create a mask. Our segmentation approach uses MaskRCNN, a recurrent convolutional neural network that identifies object shapes in images, pre-trained on the COCO (Common Objects in Context) dataset [14]. COCO is a state-of-the-art computer vision dataset, consisting of over 330,000 images of objects within 80 common categories, including humans, animals, traffic objects, foods, and more. We process the output from the MaskRCNN into a mask image where a black segmented object lies against a white background.

4.3 Word Cloud Generation

Finally, we generate a shaped word cloud using the input text document and the mask created from the previous section. We do so using the existing `WordCloud` library in Python [17]. This library provides a simple function call to generate a word cloud given a text and also provides options for custom shapes and colour encodings. We wrap our entire pipeline in a singular Python function, `generate_shaped_wc`, where the user specifies a text file and an image search method (either `api` or `clip`) and get a word cloud in return as an image.

5 RESULTS AND EVALUATION

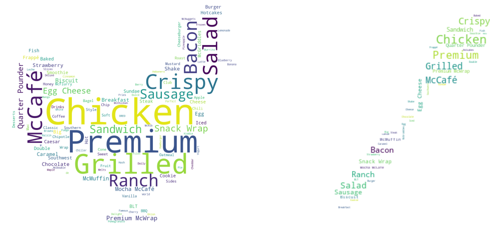
We successfully generated shaped word clouds (e.g. Fig. 1) with our end-to-end pipeline. The majority of word cloud shapes corresponded to a key idea or theme in the text, but a select few did not. We compare and discuss the results from our two approaches below.

5.1 Comparing the Two Approaches

(a) (b)

Figure 3: Word cloud for *Alice in Wonderland* using (a) `The Noun Project` and (b) CLIP.

`The Noun Project` was able to create effective masks as it returned icons that usually displayed an object in a clear, faithful way. This is important in a word cloud mask as the shape of the word cloud may break the strict graphical borders of a recognizable object. To achieve clarity in graphical form with a word cloud, the shape of the object should be strikingly clear as a regular image, which it often was with this approach.



(a) (b)

Figure 4: Word cloud for the McDonald's menu using (a) `The Noun Project` and (b) CLIP.

(a) (b)

Figure 5: Word cloud for the script of *Star Wars Episode IV: A New Hope* using (a) `The Noun Project` and (b) CLIP.

Overall, using CLIP as an image search query engine followed by MaskRCNN to segment the image to create a mask produced acceptable results. However, the resulting image is not always perfect. This is due to the fact that the Unsplash dataset powering CLIP mainly uses professional stock-like photos, which unlike icons, are not always ideal for mask generation. Additionally, MaskRCNN is sometimes unable to detect the desired class (e.g. hamburger for the McDonald's menu in Fig. 4 (b)) and returned a spoon instead since it is only trained on the COCO classes. In other cases, such as the script for *A New Hope*, the resulting mask from CLIP (Fig. 5 (b)) has a final mask shape that does not convey much meaning. There are also some cases in which the automatically generated search query fails for CLIP, such as for the Computer Science Wikipedia entry, and thus the word cloud generated was not a shaped one.

5.2 User Feedback

We informally evaluated our word cloud designs with 5 participants. Our participants were all between the ages of 18 and 26, based in the United States, and were familiar with the concept of word clouds. We gave participants 3 sets of word clouds, where each set contained a regular one in the shape of a rectangle, and one generated by our tool. The 3 input texts we used were *Alice in Wonderland*, the script of *A New Hope*, and the Gettysburg Address. Participants were asked to state which word cloud in the set they preferred along with a brief explanation why.

In general, participants preferred the ones generated by our tool over the regular ones. One participant mentioned that the shaped word clouds were useful for the two fictional stories because *"the shape directly reflects a representative character from the movie or the novel.* Another participant mentioned that *"the [regular ones] are dull and outdated—I've seen them being used since 2010 [...] you need variations"*.

On the other hand, some participants mentioned that the shapes can be distracting, especially when they do not obviously align with the text content. Referring to the Gettysburg Address word cloud, one participant said that *"[the] shape is like a small island with a flag on it, which is irrelevant to the text's topic that is related to a 'nation.'"* One participant preferred the regular ones for all 3 examples, with the reason being that in addition to the distracting shapes, *"it's also harder to interpret the relative size of words in the novel shapes. Some of the obvious words [in the regular one] become hard to recognize [in the shaped one]."*

This evaluation showed that in order for shaped word clouds to be effective, the shape must be tightly associated with the text content. There also seems to be a tradeoff when it comes to analytical clarity (e.g. clearly displaying font size variation) and visual novelty (e.g. masking the word cloud into a sophisticated shape). This may be an interesting avenue for future work.

6 CONCLUSION

6.1 Contribution

Our work provides a fully-automated pipeline for generating word clouds with semantically-aware shapes from arbitrary text documents. Compared to ordinary word clouds, our shaped word clouds provide a big-picture overview of the context of a text at a single glance.

6.2 Limitations

Our method is fundamentally limited to text documents with non-abstract topics that are easy to represent with a shape. It is unlikely to find an informative shape for highly complex or abstract topics, such as politics documents or poetry.

6.3 Future Work

Future work should employ colour encodings to enhance engagement and comprehension of shaped word clouds. For example, different word sentiments can be represented with a different colour in the cloud to capture emotions present in the text. Further investigations can include generating multiple word masks for a single text document, capturing multiple topics at once. Lastly, we suggest researchers to consider new approaches to portraying abstract topic classes, in order to make shaped word clouds more applicable to a larger range of areas and audiences.

ACKNOWLEDGMENTS

We thank professors Cecilia Aragon and Leilani Battle for their feedback and guidance, as well as the course staff of CSE 512: Data Visualizaton (Spring 2022) at the University of Washington for general discussion and advice.

REFERENCES

[1] Luke Chesser. 2020. UNSPLASH. https://unsplash.com/blog/the-unsplash-dataset/
[2] Christopher Collins, Fernanda B. Viegas, and Martin Wattenberg. 2009. Parallel Tag Clouds to explore and analyze faceted text corpora. , 91-98 pages. https://doi.org/10.1109/VAST.2009.5333443
[3] Google Arts & Culture. 2013. Word Clouds. https://artsandculture.google.com/asset/word-clouds/zAE9ZeRSirwvWw
[21]]feinberg Jonathan Feinberg. [n. d.]. Jonathan Feinberg. https://www.mrfeinberg.com/
[5] Vladimir Haltakov. 2021. Unsplash Image Search. https://doi.org/10.5281/zenodo.1234
[6] Martin J Halvey and Mark T Keane. 2007. An assessment of tag presentation techniques. In *Proceedings of the 16th international conference on World Wide Web*. 1313–1314.
[7] Marti Hearst. 2009. *Search user interfaces*. Cambridge university press.
[8] Marti A. Hearst, Emily Pedersen, Lekha Patil, Elsie Lee, Paul Laskowski, and Steven Franconeri. 2020. An Evaluation of Semantically Grouped Word Cloud Designs. *IEEE Transactions on Visualization and Computer Graphics* 26, 9 (2020), 2748–2761. https://doi.org/10.1109/TVCG.2019.2904683
[9] Florian Heimerl, Steffen Lohmann, Simon Lange, and Thomas Ertl. 2014. Word cloud explorer: Text analytics based on word clouds. In *2014 47th Hawaii international conference on system sciences*. IEEE, 1833–1842.
[10] Jaemin Jo, Bongshin Lee, and Jinwook Seo. 2015. WordlePlus: expanding wordle's use through natural interaction and animation. *IEEE computer graphics and applications* 35, 6 (2015), 20–28.
[11] KyungTae Kim, Sungahn Ko, Niklas Elmqvist, and David S. Ebert. 2011. WordBridge: Using Composite Tag Clouds in Node-Link Diagrams for Visualizing Content and Relations in Text Corpora. In *2011 44th Hawaii International Conference on System Sciences*. 1–8. https://doi.org/10.1109/HICSS.2011.499
[12] Kyle Koh, Bongshin Lee, Bohyoung Kim, and Jinwook Seo. 2010. Maniwordle: Providing flexible control over wordle. *IEEE Transactions on Visualization and Computer Graphics* 16, 6 (2010), 1190–1197.
[13] Byron YL Kuo, Thomas Hentrich, Benjamin M Good, and Mark D Wilkinson. 2007. Tag clouds for summarizing web search results. In *Proceedings of the 16th international conference on World Wide Web*. 1203–1204.
[14] Tsung-Yi Lin, Michael Maire, Serge Belongie, Lubomir Bourdev, Ross Girshick, James Hays, Pietro Perona, Deva Ramanan, C. Lawrence Zitnick, and Piotr Dollár. 2014. Microsoft COCO: Common Objects in Context. http://arxiv.org/abs/1405.0312 cite arxiv:1405.0312Comment: 1) updated annotation pipeline description and figures; 2) added new section describing datasets splits; 3) updated author list.
[15] Steffen Lohmann, Jürgen Ziegler, and Lena Tetzlaff. 2009. Comparison of tag cloud layouts: Task-related performance and visual exploration. In *IFIP Conference on Human-Computer Interaction*. Springer, 392–404.
[16] Andreas Mueller. 2012. A Word Cloud in Python. https://peekaboo-vision.blogspot.com/2012/11/a-wordcloud-in-python.html
[17] Andreas Mueller. 2018. WordCloud. https://github.com/amueller/word_cloud.
[18] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning Transferable Visual Models From Natural Language Supervision. arXiv:arXiv:2103.00020
[19] Anna W Rivadeneira, Daniel M Gruen, Michael J Muller, and David R Millen. 2007. Getting our head in the clouds: toward evaluation studies of tagclouds. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 995–998.
[20] Julie Steele and Noah Iliinsky. 2010. *Beautiful visualization: Looking at data through the eyes of experts*. " O'Reilly Media, Inc.".
[21]]the-noun-project The Noun Project. [n. d.]. Build with Noun Project's Icon API. https://thenounproject.com/api/
[22] Fernanda B Viegas, Martin Wattenberg, and Jonathan Feinberg. 2009. Participatory visualization with wordle. *IEEE transactions on visualization and computer graphics* 15, 6 (2009), 1137–1144.
[23] Ji Wang, Jian Zhao, Sheng Guo, Chris North, and Naren Ramakrishnan. 2014. ReCloud: semantics-based word cloud visualization of user reviews.. In *Graphics Interface*. 151–158.
[24] Yunhai Wang. 2017. EdWordle. https://www.edwordle.net/