
BrainstorML

Stefania La Vattiata

Jun 04, 2021

OVERVIEW

1	Description	3
1.1	About	4
1.2	Prototype Pipeline	5
1.3	Dataset	5
1.4	Prototype Evaluation	6
2	Indices and tables	9

Welcome to the BrainstorML documentation page.

DESCRIPTION

Web app that allows for automatic clustering and generation of ideas based on word embeddings, using Natural Language Processing models.

build/html/_static/documentation/Home.png

1.1 About

This project started as my thesis project, after exploring Machine Learning pre-trained models throughout my Masters program at Carnegie Mellon University.

I wanted to focus on the Generative phase synthesis as a space to apply ML models to optimize work and assist Designers.

During this analysis, I asked myself **how do designers generate ideas?**

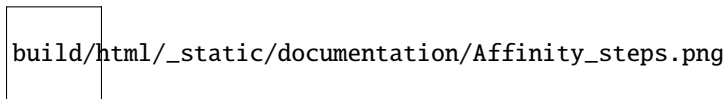
- What kind of methods/processes do they go through?
- Might there be an opportunity to use machine intelligence to make the process undemanding and focused on the creative instead of the time-consuming small tasks?

During the generative research phase, many participatory activities allow for collaboration between the designer and the user. During this collaboration, some of the methods may include the following: creative toolkits, design workshops, card sorting, collages, cognitive mapping, affinity diagrams, and flexible modeling, to name a few (Hanington & Martin, 2019).

Based on my experience as an interaction designer and the interviews with other professional designers, I noticed that affinity diagrams were a standard method used by all of us. Although the collaborative methods mentioned before are valuable, I wanted to concentrate on a frequent generative approach for designers to engage with (Lucero, 2015).

Thinking, of course, the AI integration with this method could enhance their sense-making and reduce time-consuming tasks, like the clustering component, and potentially generate more ideas.

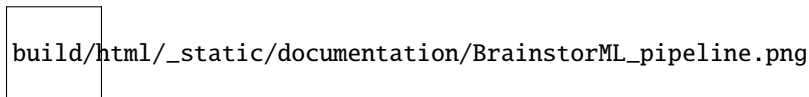
“In design practice, affinity diagrams are used to analyze a design problem or create first design solutions. With a given design problem in mind, designers write down words, short sentences or make small sketches on sticky notes to stimulate diversity in ideation phases. Ideas are then clustered to identify common issues and potential solutions” (Lucero, 2015).



1.2 Prototype Pipeline

I divided the main components into two. The front-end covered all the interactions for the interface and user's input and backend, referring to the NLP and ML elements and integration with the interface.

Based on the Affinity Diagram structure showed in the About page, I modeled the prototype's pipeline and how the experience would be.



1.3 Dataset

Word vectors are the magic component making the clustering work.

1.3.1 GloVe

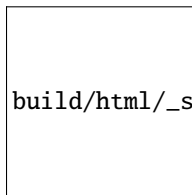
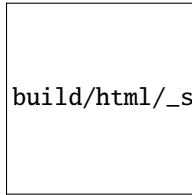
To summarize, word vectors are the numeric representation of the meaning of a word. They represent words as multi-dimensional numbers that group together to semantically similar words (Ahire, 2018).

With this figured out, I just needed a word vector dataset to work with. Fortunately, Stanford scientists developed an open-source ML model that could quickly get semantically similar words of any text in the vector space (Pennington, Socher, & Manning, 2014). The datasets come in different sizes. The one I worked on had 400,000 English words.

To be specific, the meaning that these words are given in a mathematical interpretation is obtained from a corpus of text, which was given to train the Machine Learning model.

Gender bias

I must admit that I am not aware of what are the texts that build this corpus. They could be any type of literature and writings, like websites, articles, books, blogs, news, or maybe all of them combined. This may cause several defaulted assumptions from the texts that don't follow current ethical standards, like gender bias roles.



1.4 Prototype Evaluation

To evaluate the prototype, I made sure to have a working prototype that covered the satisficing principle from Herbert Simon (Simon, 2013), a composition of the words satisfy and suffice. He mentions that through decision making, the Design should quickly meet the user's needs instead of being the "optimal" solution that covers all the features intended.

Based on this term, I prototyped only the primary clustering of the words by their semantical proximity instead of having all the ideal characteristics in the application. For example, grouping sentences, categorizing automatically, generating more ideas for a group, and modify the sticky notes accordingly to the user's interaction. So I just covered the clustering, which was the primary need of the designer through this process.

1.4.1 Questions

Once I had that, I discussed the tool's functionality and opportunities with three professional designers. The discussion took around 30 minutes, where I asked fundamental questions about the UI of the landing page and content, like:

What is your (first) impression of this product/feature?

They mentioned that as they were reading the first sentences of the landing page. They expected the synthesis to be fast.

Once we entered the main workspace, and after writing a brainstorming prompt to guide the discussion, in this case, was the question what makes you happy? They started sharing information for me to work on within the sticky notes.

After we had enough sticky notes with different words, I clicked the group button, and the algorithm grouped the ideas in specific clusters. As mentioned before, some words were so similar that some sticky notes overlapped, so I had to move them around a little bit to see the groups better.

After the groups were clear, I asked the following:

Did the clustering make the groups as you expected?

When I asked this question, they analyzed the groups and started realizing some patterns between some of the words, but others did not fit so well with the rest. So they were mentioning how they could see relations between some but that some could be integrated with other terms better.

Do you think you can see relationships between the clustered groups?

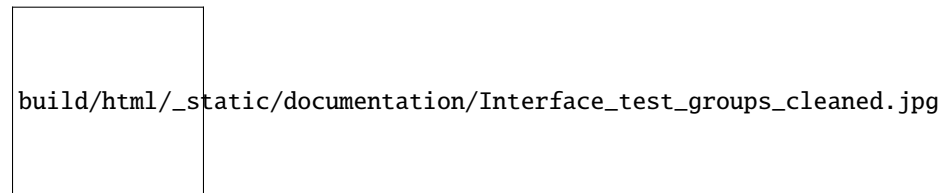
When I asked this, they said they were looking at some of the words that could fit better together and that considering the context and other words existent in the board is essential when clustering information.

What do you think would be best to make the relationships clearer between the groups?

For this question, they said that having overarching categories in the groups would make it easier to understand the context or the thread in the groups. This feature would provide the designer with a clearer idea of the patterns existent throughout the board.

Would this be useful for you?

They mentioned that it would be for the initial stages of ideation. Even sorting it with a margin of error was enough to start finding patterns in the information, which would reduce the cognitive load of doing this by hand. Additionally that it saves them time to do this by hand from the beginning.



1.4.2 Insights

This session helped me understand how potential users thought about this intervention and how useful it could be. It uncovered certain aspects that I did not consider, like reducing the cognitive burden of having so much information cluttered by just organizing it by semantic approximation and other features like filtering by relevance. The most relevant insights from the session are the following:

It is helpful for the initial round of synthesis and seeing what patterns are existent.

“if I had a massive amount of post-its on this board, even if the initial grouping that ml thing does is like kind of off, it would still be helpful, because of how it creates that initial round of seeing what patterns are there, it’s just like understanding the board.”

It reduces time significantly and allows for clarity in identifying the cluster outliers.

“If we had to do this manually, I think we would spend at least 15 minutes to reach this stage by ourselves, but because the clustering just happened, the only thing that we had to do was identify the outliers or things that did not make sense”

It would be beneficial to spark creativity by throwing random ideas.

“I think this might be helpful in very early stages of ideation, if you have to come up with 50 ideas, this might be a tool that assists you by throwing random words that you think of and it might help us spark some creativity with the pattern recognition.”

INDICES AND TABLES

- `genindex`
- `modindex`
- `search`