1 Project Overview

EDITED
NOTE: All highlights have same meaning, different colors used for aiding differentiation when pairing an edit with its comment.

1.1 Project Summary

In today's world, information[a] comes in many forms and in an overwhelming abundance. People have access to a seemingly endless supply of news, but their news feed could use some improvement. On one end, news feeds fail to deliver relevant, interesting articles suited to its audience. On the other end, news feeds attempting to deliver news suited to its audience frequently over-personalize, leaving the audience with a narrow range of topics, repeated information, and biased outlooks. What's more, most people do not recognize the shortcomings of their news feed. People fail to realize more interesting news is out there for them, or that their news is biased because the news feed does not inform them of this debacle.

The basis of our project is to create a web service for delivering curated news, selected personally for each individual user. Two methods are used to establish a personal 'taste', as it were, for each user. To establish baseline interests, one will be able to subscribe a variety of topics and news outlets. As time goes on, each user's interests can be further refined based on periodic feedback and a user's viewing history. Where we differ from other news curation applications is in the means by which we select news. Currently, web search is often limited to content matching - if you search "Minnesota Vikings", you may only be able to see articles containing those words. Various machine learning algorithms and web-search techniques will allow us to not only perform content matching, but also perform searches based on news topics and interconnectedness between articles, such as matching "Minnesota Vikings" with "Mike Zimmer", the head coach of the team. In addition to this, existing platforms with personalization features fail to allow a user to "unlearn" the taste developed for their profile. As it currently stands, if I, for example, look at news about a Supreme Court appointment one day, there's no way to tell the program that it was a one-time occurence, and I, in fact, am not as interested in politics as the algorithm may think I am. We plan on implementing a mechanism with which which users can manually override the taste determined by the algorithm, and get more news that they actually want. [b]These methods of personalization, search techniques, and the ability to "unlearn" - all built upon a fast and reliable web application- make our proposed news curation service superior to any other service available today. [c]

1.2 Project Objectives

The first business objective for this project is to create a proof of concept, with which we can then gather feedback from focus groups and potential users. Our

project is the first step in a continuous cycle of prototyping, gathering feedback, and making improvements to our product.[d] In addition to this, another objective is to implement a feature pertaining to “unlearning” the topics associated with a user’s taste. As far as we’ve seen, this isn’t available in any other news curation product, or anywhere else - proving that this is both doable, and desirable from a user standpoint, is another objective of our project.[e]

1.3 Project Stakeholders

The main stakeholders for this project include developers Katie Wraith, Jason Conci, Chris Delaney, Kevin Tran, and Daniel Gallab, who will be creating the product and the deliverables that support it. They are joined by faculty advisor, Dr. David Schroeder, and Design Advisory Board member, Dan Lenz, both of whom will be helping the developers keep on task and apply their expertise where needed. Sponsor Scott Broder proposed the project, and will work alongside the developers to make sure his vision is fulfilled. At the conclusion of this project, Scott Broder will potentially take the product through focus groups to gather feedback, at which point the product can be further refined. The target users are people who desire a better personalized news feed based on their interests, or those who are interested in finding news from sources they may have yet to explore.[f]

1.4 Project Deliverables

The main project deliverable would be a web-application, containing both complete front and backend services deployed on an AWS or Azure-like hosting platform. The functioning application would allow users to search for curated news, and be delivered news via a machine learning based recommendation system. The platform would make use of various NLP and web semantics libraries to establish the recommendation and curated news system. This library usage also involves delivering documentation, in order to allow users to make RESTful calls to our backend [g]to search for specific new articles without the need for interacting on our client-side. This documentation will also allow future developers to understand how we made this application, and enable [h]them to build upon it to improve functionality. The documentation will be hosted on the same platform.

1.5 Project Scope

Provide a brief description of the project scope that states what aspects of the project already exist (out of scope) versus what aspects you will be developing from scratch (in scope). Your description must be accompanied by a high-level context diagram, highlighting the components that are outside of the scope of your project and how they will generally interact with your system.

<insert figure here>

Figure 1: Context diagram where ...

1.6 Related Work

The application that is the most similar product to what we are building is the updated Google News[1]. It uses an artificial intelligence algorithm to take the news as it comes in, and creates brief summaries of each story [i]by collecting data from several news sites and creating storylines as the coverage continues to be released. The app learns the user's interests through tracking what the user reads or clicks; however, this approach is slightly flawed, in that the user may not particularly like the article they viewed. In this case, the user wouldn't want the machine learning algorithm to be skewed to favor and display articles similar to those that [j]the user didn't like. Our product would give the reader the option to rate the article they are reading, in order to both learn their interests, and to "unlearn" certain preferences, rather than force the user to mask the unwanted preferences[k]. Google News also does not provide a tab dedicated to the user's reading history. If a user wants to come back to an article later for any reason, he or she would have had to make the conscious effort to "star" the article while reading it. Often,[l] users don't decide until a later point in time that they want to show their friends an article they found. Our app gives users the ability to access previously-read articles, just in case the user wants to show their friends later on. Another example of a similar application is Alexa's Flash Briefing Skill[2]. It provides a quick daily update on the user's preferred content such as news, lists, and even jokes. Developers can build Flash Briefing Skills in order to provide their own content for users' consumption through Amazon's developer kit on AWS using lambda. Instead of using Machine Learning to tailor the feed's topics to the user, though, Flash Briefing relies on[m] the user to pick exactly which feeds to follow. This doesn't allow the user to easily expand their horizons, nor explore other options that might interest them. The approach used in Flash Briefing[n] also doesn't allow the user to give feedback on the content it provides. Our product would determine the user's potential interests based on their clicks, set preferences, and feedback, and, given this information,[o] provide news from sources the user might not be familiar with. Our product will also focus exclusively on news, so that the project remains within a realistic scope.[p]

1 Project Overview

ORIGINAL

1.1 Project Summary

In today's world, news comes in many forms and in an overwhelming abundance. People have access to a seemingly endless supply of news, but their news feed could use some improvement. On one end, news feeds fail to deliver relevant, interesting articles suited to its audience. On the other end, news feeds attempting to deliver news suited to its audience frequently over-personalize, leaving the audience with a narrow range of topics, repeated information, and biased outlooks. What's more, most people do not recognize the shortcomings of their news feed. People fail to realize more interesting news is out there for them or that their news is biased because the news feed does not inform them of this debacle.

The basis of our project is to create a web service for delivering curated news, selected personally for each individual user. Two methods are used to establish a personal 'taste', as it were, for each user. To establish baseline interests, one will be able to select from a variety of topics and news outlets, for instance, "The Minnesota Vikings" and "SportsCenter". As time goes on, each user's interests can be further refined based on periodic feedback, search history, and articles related to one's viewing history. Where we differ from other news curation applications is in the means by which we select news. Currently, web search is often limited to content matching - if you search "Minnesota Vikings", you may only be able to see articles containing those words. Various machine learning algorithms and web-search techniques will allow us to not only perform content matching, but also perform searches based on news topics and interconnectedness between articles, such as matching "Minnesota Vikings" with "Mike Zimmer", the head coach of the team. These methods of searching will not only allow us to perform the same functions as our competitors, but also provide users with news on a fresher and more personal level than ever before. Our proposed methods, built upon a flexible web service, will allow us to port this product to a number of devices, creating perhaps the most personal and accessible news curation application on the market today.

1.2 Project Objectives

The business objectives for this project is to create a proof of concept to gain feedback from potential target users so that it can be improved upon for more feedback as the cycle continues. Our project won't produce the final project that we can release immediately to consumers. Rather, it will be the beginning of a long development process that we can show to potential users and receive feedback on the user experience, general approach, and outcome. This way, a new and improved version can be made and the cycle can start again and repeat until a final, user-friendly version is released that people will want to use.

1.3 Project Stakeholders

The main stakeholders for this project include developers Katie Wraith, Jason Conci, Chris Delaney, Kevin Tran, and Daniel Gallab, who will be creating the product and the deliverables that support it. They are joined by faculty

advisor, Dr. David Schroeder, and Design Advisory Board member, Dan Lenz, both of whom will be helping the developers keep on task and apply their expertise where needed. Sponsor Scott Broder proposed the project and will work alongside the developers to make sure his vision is fulfilled. The target users are people who desire a personalized news feed based on their interests or who are interested in finding news from sources they may have yet to explore in order to expand their horizons.

1.4 Project Deliverables

The main project deliverable would be a web-application with complete front and backend services deployed on an AWS or Azure-like hosting platform. The functioning application would allow users to search for curated news and machine learning recommendation system. The platform would make use of various NLP and web semantics libraries to establish the recommendation and curated news system. The project would also deliver a documentation to allow users to make RESTful calls to our backend to search for specific new articles without the need for interacting on our client-side. It will also allow future developers to be able to understand how we made this application and be able to build upon it to improve functionality. The documentation will be hosted on the same platform.

1.5 Project Scope

Provide a brief description of the project scope that states what aspects of the project already exist (out of scope) versus what aspects you will be developing from scratch (in scope). Your description must be accompanied by a high-level context diagram, highlighting the components that are outside of the scope of your project and how they will generally interact with your system.

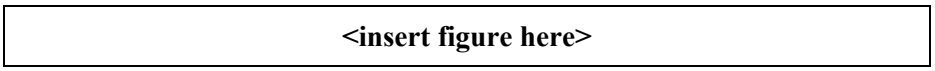

Figure 1: Context diagram where ...

1.6 Related Work

The application that is the most similar product to what we are building is the updated Google News[3]. It uses an artificial intelligence algorithm in order to take the news as it comes in and create a brief summary of news stories by collecting data from several news sites and creating storylines as the coverage continues to be released. The app learns the user's interests through tracking what the user reads or clicks. However, sometimes the user doesn't particularly like the article he or she clicked. In this case, the user wouldn't want the machine learning algorithm to be skewed to favor and display articles similar to the one the user didn't like. Our product would give the reader the option to rate the article they are reading so that the algorithm is also able to "unlearn" certain preferences instead of making the user taking the unnecessary time to mask the unwanted preferences. Google News also does not provide a tab dedicated to the user's reading history. If a user wants to come back to an article later for any reason, he or she would have had to make the conscious effort to "star" the article while reading it. Many times users don't decide until a much later conversation that

they want to be able to show their friends an article they found. Our app gives users the ability to access previously-read articles again in case the user wants to show his or her friends later on. Another example of a similar application is Alexa's Flash Briefing Skill[4]. It provides a quick daily update on the user's preferred content such as news, lists, and even jokes. Developers can build Flash Briefing Skills in order to provide their own content for users' consumption through Amazon's developer kit on AWS using lambda. However, instead of using Machine Learning to tailor the feed's topics to the user, Flash Briefing allows the user to pick exactly which feeds to follow. This doesn't allow the user to easily expand their horizons and explore other options that might interest them. It also doesn't allow the user to give feedback on the content it provides. Our product would determine the user's potential interests based on their clicks, set preferences, and feedback and provide news from sources the user might not be familiar with. It will also solely focus on news so that its scope doesn't spread too thin.

---

[1] https://www.blog.google/products/news/new-google-news-ai-meets-human-intelligence/

[2] https://developer.amazon.com/docs/flashbriefing/understand-the-flash-briefing-skill-api.html
[3] https://www.blog.google/products/news/new-google-news-ai-meets-human-intelligence/

[4] https://developer.amazon.com/docs/flashbriefing/understand-the-flash-briefing-skill-api.html

[a]Provide general context, then scope in on news.
[b]Added this. After meeting with our sponsor, we decided this is a more important feature than we initially made apparent.
[c]Rewritten to avoid weak phrasing - specifically, "perhaps superior..." to "superior".
[d]Rewritten - was difficult to follow.
[e]Added this. Again, we didn't convey this feature's importance to our project well before.
[f]Added. After discussion with our sponsor about his intentions with the project, I think that focus groups were relevant stakeholders as well.
[g]Modified the wording, seemed imprecise before.
[h]Reworded, unclear subject matter
[i]Reworded for clarity
[j]All of this pertains to phrasing
[k]Changed slightly to better convey that the rating system is dual-purpose - learn + unlearn
[l]Wording ("Often" + "Later point in time")
[m]"Allows" implies optional, "relies on" implies required (more accurate)
[n]Change "it" to "the approach used in flash briefing", unclear antecedent

[o]convey a more clear "stop" between getting information + using information

[p]Reworded - unclear antecedent, more formal terminology.