# Using Sentiment-Based Coloring to Improve Text Entry Retrieval Speed

An Analysis of the *Jot* Mobile Application

Taylor K. Beck
Vanderbilt University
Nashville, TN, USA
taylor.k.beck@vanderbilt.edu

James T. Kirven
Vanderbilt University
Nashville, TN, USA
james.t.kirven@vanderbilt.edu

***Abstract*—The mobile application being tested in the following experiments, *Jot*, is a diary app that uses sentiment analysis to provide meaningful user interface elements. The goal of this paper is to determine whether these sentiment-based UI elements provide a quantifiable advantage to the user over traditional methods of representing text entries. Participants performed time-on-task trials, in which they retrieved a provided text entry using various sentiment tools. We were unable to conclusively prove or disprove our hypothesis, but still analyze the test results and suggest future user interface adjustments.**

***Keywords*— diary; sentiment analysis; user interface.**

## I. Introduction

People write in diaries or journals in order to save information about their personal lives, often with the ultimate goal of reflecting on that information later. The inspiration for *Jot* stemmed from the desire to improve the self-reflection phase of diary use. Despite providing a plethora of features for inputting data, current diary apps still tend to lack truly observant methods of retrospection. With a large number of entries, it is difficult to display a summary to the user from a macroscopic point of view. For example, popular note-taking app *Evernote* displays entries in a list, in order from most recent to oldest, and provides a search functionality. Besides combing through individual entries, the user cannot quickly glean information about themselves. This problem remains in more journal-focused apps like *Journey*, which provide a calendar view, but do not provide further meaning other than whether or not an entry was made each day. If a tool was able to accomplish this expanded scope of retrospection, users would be more aware about their behaviors, ultimately giving them more power within their own lives through knowledge.

To assist the user in understanding themselves through their diary entries, *Jot* uses Stanford's natural language processing library, CoreNLP [2]. Sentiment analysis is a subset of natural language processing that aims to identify the subjective emotional characteristics of a textual statement. In *Jot*'s use case, sentiment analysis is performed by ranking statements along a scale from negative (angry, sad) to positive (happy). By analyzing the sentiment of a user's diary entries, it is possible to represent the user's sentiment visually over time. By representing sentiment trends in a graph, the application allows the user to quickly recognize patterns in their diary entries. A user can now be more aware of their emotional states over a longer period of time. In addition, users can quickly determine whether an individual entry contains the information that they are searching for, simply from a colored label indicating sentiment.

Note-taking applications, including *Jot*, revolve around their text-based entries. The application interface should at minimum provide functionality for the user to: i) enter new entries, ii) browse entries, and iii) find specific entries. Of these three base functionalities, finding specific entries (iii) was the most quantifiable. Although *Jot* has more features than the ones tested in the experiment trials, they did not provide clear numerical results for the sake of performance evaluation. To determine whether certain sentiment-based features had a quantifiable positive impact on the user, the analysis begins with the axiom that faster completion of a task is better. In the trials, test participants would be given a specific text entry to find, and must retrieve it using a given set of features. The initial hypothesis was that participants would be able to more quickly retrieve entries with sentiment-based features than without.

## II. System Description

### A. User-Facing Application Interface

*Jot*'s primary component is a mobile application for the Android phone platform. Upon opening the app, the user is presented with the main view (Fig. 4). This view consists of two primary elements: the list and the graph.

#### 1. User Interfaces Used in Testing

The list contains every entry in the diary. Each entry is represented on its own card, with the date and time of entry listed along the top. Additionally, the corners of the card are shaded either green, gray, or red. Green represents an overall positive sentiment within the entry, while gray is neutral and red is negative. There are multiple shades of green and gray as well, with vibrant green being the most positive and vibrant red being the most negative.

The graph represents the user's diary entries over time. Each node on the graph represents a day for which there was at least one entry. The horizontal axis represents time, while the vertical axis represents sentiment. The graph is also interactive: clicking on a node scrolls the list to the first entry on that day, and the user is able to pinch to zoom as well as scroll in order to navigate. Cross-hairs indicate the currently selected node, which moves when the list is scrolled manually.

Another view was added specifically for testing purposes. It consists of a slider with which the user can estimate the sentiment of a given prompt entry and a start button, which begins a timer. The timer ends when they select an entry. This process will be described in more detail in Methods (Section IV).

Participants in test trials each use four different versions of *Jot* in a random order. They have the following characteristics:

| ***Jot* App Versions** | **Features Available** | | | |
|---|---|---|---|---|
| | ***List*** | ***Graph*** | ***Colors on List*** | ***Colors on Graph*** |
| *Colored Graph (Fig. 4)* | X | X | X | X |
| *Colored List (Fig. 2)* | X | | X | |
| *Colorless Graph (Fig. 5)* | X | X | | |
| *Colorless List (Fig. 3)* | X | | | |

Fig. 1. Table describing each app version

### 2. *User Interfaces Absent from Testing*

The graph and the list are the only primary components of the application that are used in the testing described below, but there are other important features to the full version of the app. Clicking on the floating action button reveals a text input box for writing entries. The input also includes two emoticons to be used in entries, which are taken into account by the sentiment estimation algorithm on the server. There is also a standard search functionality for finding entries.

Another major feature of the app that is not tested in this paper's experiments is entity recognition. The server also analyzes each diary entry for people or important objects or places, hereafter referred to as entities. By associating the sentiment of a sentence in which an entity appears with that entity itself, the app can then display entries for a specific entity. This allows users to reflect on how a specific person has made them feel over time, for example. The display for specific entities is identical to the main view but with the list of entries limited to those that mention the entity.

## B. *Server-side Operations*

*Jot* connects to its servers over the web in order to analyze the sentiment of diary entries. The server is running in a Java Apache HTTP service which handles the text of each entry as well logging feedback from the test trials. When a new entry comes in the server parses the body of the entry and passes it through CoreNLP, a Stanford language processing library. The server returns XML to the app containing the sentiment for each sentence in the entry as well as information such as entities recognized in the entry. The decision to perform sentiment analysis on a remote server was made because the necessary libraries were too large to reasonably fit on a consumer phone. Also, entry data is stored on the phone rather than the server. The reasons for this were twofold: it simplifies the implementation greatly and preserves the privacy of the user's diary. The calculations for each entry's sentiment is done on each device by averaging the sentiment for each sentence. We also provided the user with the ability to use emoticons to force the sentence containing the emoticon to have that sentiment.

## C. *App Version Screenshots*

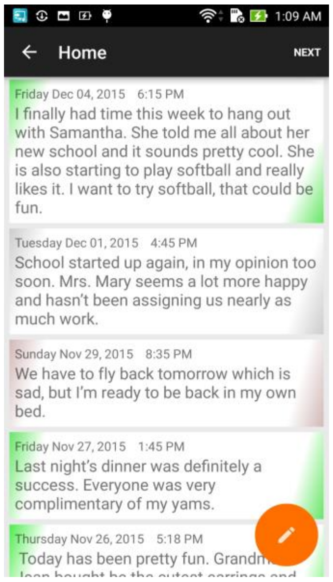

1:09 AM
Home NEXT
Friday Dec 04, 2015 6:15 PM
I finally had time this week to hang out with Samantha. She told me all about her new school and it sounds pretty cool. She is also starting to play softball and really likes it. I want to try softball, that could be fun.
Tuesday Dec 01, 2015 4:45 PM
School started up again, in my opinion too soon. Mrs. Mary seems a lot more happy and hasn't been assigning us nearly as much work.
Sunday Nov 29, 2015 8:35 PM
We have to fly back tomorrow which is sad, but I'm ready to be back in my own bed.
Friday Nov 27, 2015 1:45 PM
Last night's dinner was definitely a success. Everyone was very complimentary of my yams.
Thursday Nov 26, 2015 5:18 PM
Today has been pretty fun. Grandm

Fig. 2. Colored List

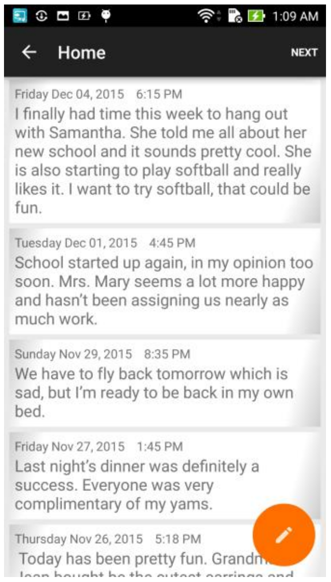
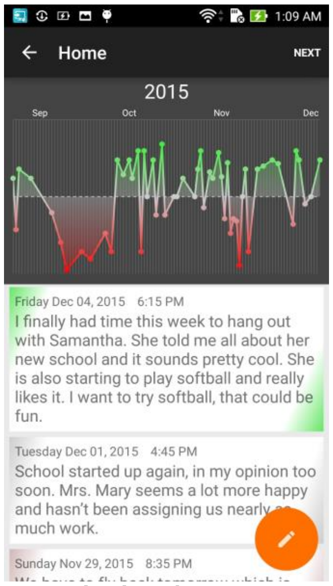
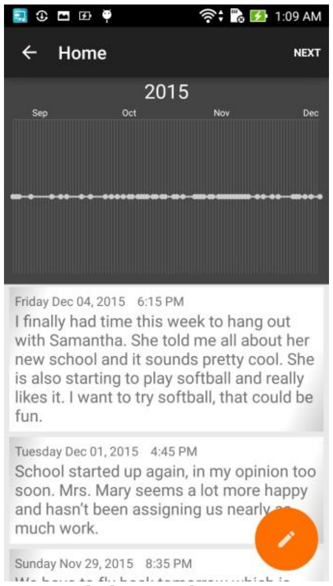

1:09 AM
Home NEXT
Friday Dec 04, 2015 6:15 PM
I finally had time this week to hang out with Samantha. She told me all about her new school and it sounds pretty cool. She is also starting to play softball and really likes it. I want to try softball, that could be fun.
Tuesday Dec 01, 2015 4:45 PM
School started up again, in my opinion too soon. Mrs. Mary seems a lot more happy and hasn't been assigning us nearly as much work.
Sunday Nov 29, 2015 8:35 PM
We have to fly back tomorrow which is sad, but I'm ready to be back in my own bed.
Friday Nov 27, 2015 1:45 PM
Last night's dinner was definitely a success. Everyone was very complimentary of my yams.
Thursday Nov 26, 2015 5:18 PM
Today has been pretty fun. Grandm

Fig 3. Colorless List



1:09 AM
Home NEXT
Friday Dec 04, 2015 6:15 PM
I finally had time this week to hang out with Samantha. She told me all about her new school and it sounds pretty cool. She is also starting to play softball and really likes it. I want to try softball, that could be fun.
Tuesday Dec 01, 2015 4:45 PM
School started up again, in my opinion too soon. Mrs. Mary seems a lot more happy and hasn't been assigning us nearly as much work.
Sunday Nov 29, 2015 8:35 PM

Fig 4. Colored Graph



1:09 AM
Home NEXT
Friday Dec 04, 2015 6:15 PM
I finally had time this week to hang out with Samantha. She told me all about her new school and it sounds pretty cool. She is also starting to play softball and really likes it. I want to try softball, that could be fun.
Tuesday Dec 01, 2015 4:45 PM
School started up again, in my opinion too soon. Mrs. Mary seems a lot more happy and hasn't been assigning us nearly as much work.
Sunday Nov 29, 2015 8:35 PM

Fig 5. Colorless Graph

## III. Methods

During the design and development phase of our app we created a wide range of features and requirements to provide the user with the best possible diary experience. In order to effectively test certain features of our app under a time constrained schedule, we chose to test only two main features: the sentiment color tags on each entry and the sentiment vs. time interactive graph tool. We hypothesized that individually each of these features would decrease the time it takes a user to find a certain entry and that together these features would provide an even faster look-up experience.

After a pilot study of 8 test subjects we noticed trends in our data and received feedback from our participants that indicated that our test may not have been providing statistically significant data. We revised our test plan in order to allow for more training and to reduce the amount of variance in our data set, while maintaining the integrity of our data.

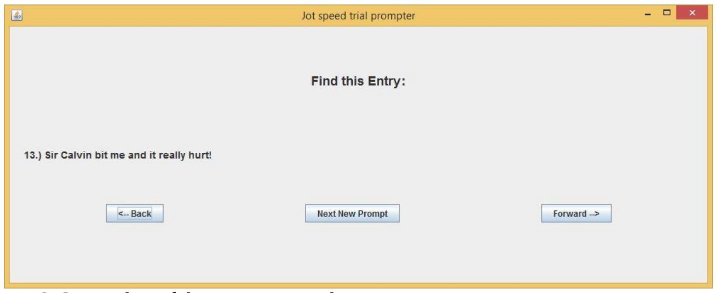
### A. *Participants*

Our participant base consisted primarily of convenience samples and was not, demographically, representative of our primary target market for this app. Of the 18 total participants, 10 participated in the second version of our trials, which will main source of our analysis. The median age of these participants was 21. From this set 60% of participants were male and 40% were female. 90% of the participants were right-hand dominant. 50% of participants had at one point owned an Android device but only 10% currently used android devices on a regular basis. The remaining 90% used either iOS or Windows smart phones on a regular basis.

### B. *Apparatus*

Our experimental design relied on a set of automated programs to provide users with a consistent and streamlined experimentation experience. The users first read a story consisting of 58 diary entries and was then introduced to the full version of the app in order to train. During this training period we showed the participant both the sentiment color tags on each entry card as can be seen in Figure 4 and explained how the color on the card ranging from red to gray to green represented the sentiment of the overall entry from very negative to neutral to very positive, respectively. We also showed the participant the sentiment vs. time interactive graph tool, which features the same color scheme as the cards, and instructed them on its functionality. The user was then allowed to train on this full version of the app to find 4 preselected entries from the corpus. We prompted them to use the features we had just taught them and asked them to focus on the relation between their interpretation of the sentiment of each entry and the color on the card. We made sure during this training to avoid telling them that these features were included to decrease their entry look-up times.

Once training was complete the prompt generation program shown in Figure 6 chose a random build from the following four test builds 1.) Plain entry list, containing no graph and no colors, 2.) Colored List only, containing sentiment colors on cards, but no graph, 3.) Colorless graph with colorless entry card list, and 4.) Colored graph with sentiment colors on the entry list.

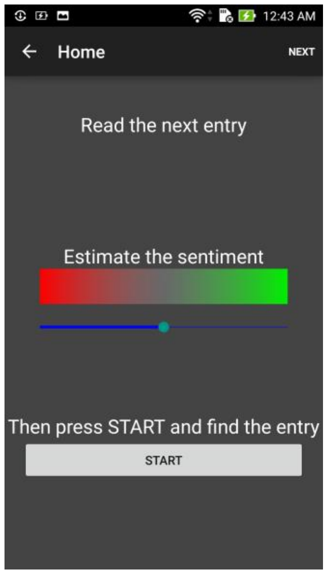

Jot speed trial prompter

Find this Entry:

13.) Sir Calvin bit me and it really hurt!

Fig 6. Screenshot of the prompter used in testing

The entry prompt program would then generate a randomized ordering to prompt the user to find 4 of 16 preselected entries. As the user went to find each entry on each build of the app they would be met with a start screen as shown in Figure 7. This would prompt them to estimate the sentiment of the entry they were about to find, forcing them to consider the color that may be on the card. This feature was on all builds, including the colorless builds for consistency. After 4 entry look-ups for a certain build the prompt program would instruct a new build of the app to test on next.



Read the next entry

Estimate the sentiment

Then press START and find the entry

Fig 7. The start screen shown prior to each prompt

At the end of each build trial we asked the participant to select what they used in order to find the entries in that trial from the following list 1.) The text in the entry, 2.) The date on the entry, 3.) The relative position in the list, 4.) The sentiment colors on the entry, and 5.) The sentiment vs. time graph. At the end of all trials they were also asked on a scale from 1-5 how the availability of each feature, the sentiment colors and the graph increased or decreased their look-up speed.

### C. *Experimental Design*

Our experimental design used a with-in subject setup. Each subject was tested on the same set of four different app builds. The dependent variables consisted of quantitative performance measures including the time required to complete each task based on the entries location in the list, the number of incorrect selections made. The independent variables consisted of the demographic information collected at the beginning of the experiment, the build type for each trial, and the percent difference in the participants estimated entry sentiment color value and the actual color value on the card. We additionally collected follow up data including how the different features affected the participants perceived look-up speed.

## IV. Results

The results we will discuss in this paper come entirely from our second trial described in section 3B. Since we made significant changes to the app and the experiment between the pilot study and the second trial the pilot data was left out. The demographic information is listed in section 3A and was not found to have a significant effect on look-up times and will not be the focus of this analysis. Figure 8 shows the mean look-up time and standard deviation for each build.

Using R, the statistical programming language, we were able to run Student's t-tests and determine p-values. The results of this analysis are shown in Figure 9 which shows a pairwise analysis of the effect of each additional feature from the baseline plain gray list.

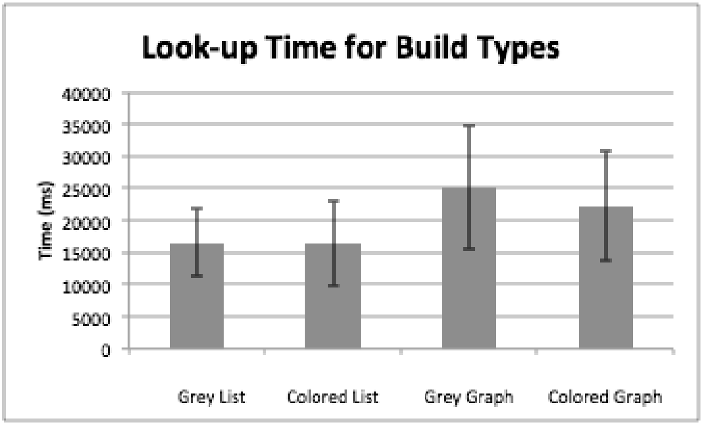


Fig 8.

| ***Statistical Analysis Results*** | *estimated coefficient* | *STD Error* | *t value* | *Pr(>\|t\|)* |
|---|---|---|---|---|
| *Baseline - Gray List* | 16567.8 | 2699.3 | 6.138 | 8.60E-09 |
| *Colored List* | -332.4 | 3789.2 | -0.088 | 0.9302 |
| *Gray Graph* | 9152.5 | 3737 | 2.449 | 0.0156 |
| *Colored Graph* | 6045.1 | 3712.8 | 1.628 | 0.1058 |

Fig 9.

The results for our follow-up questionnaire are shown in figure 10.

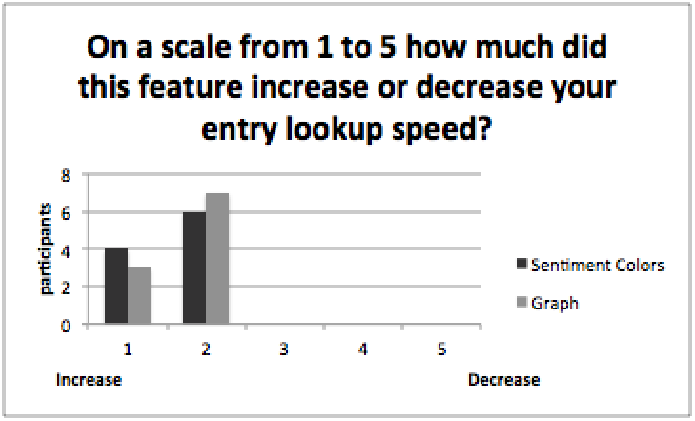


Fig 10.

## V. Discussion

The test results regarding the initially stated hypothesis were inconclusive, as the only statistically significant claim that can be made is that using the Colorless Graph version of *Jot* has an impact on time-on-task performance when compared against the Colorless List as well as the Colored List. Since our null hypothesis initially was that the app with the aforementioned features would not increase look-up speed we cannot accept that hypothesis. The only conclusion we can come to on this topic is that the hypothesis that adding the graph without colors causes a statistically significant slow-down with a P-value of <=0.0156 (alpha = 0.05). One possible observation that can be made from this result is that the graph slows the user down when it does not also represent sentiment. This phenomenon can be feasibly explained by the limited real estate on the phone screen for the list when the graph is enabled. A user scrolling through the list can only see half as many entries at any given time. A user interface adjustment to remedy this issue might be to minimize the graph when the user is scrolling. Another possible explanation is that the test participants are very accustomed to scanning text in scrollable list format; most mobile apps contain some sort of list. The graph, however, is a more novel interface with which the participants did not have nearly as much experience using. The user feedback provided an interesting result: with a P-value of $< 0.001$ users believed that these additional features helped them find entries faster.

## VI. Conclusions

Further testing is necessary to determine if the features added to our app could be beneficial in the right test environment under the right conditions. We hypothesize that in such a short test window users were unable to get accustomed to the new features, but operating under the belief that they would help them, users tended to waste time trying to use these unfamiliar features. One such environment could be created within a longitudinal study in which a user is able to generate their own content and have multiple days to familiarize themselves with the features. Additional features included in our app, but not in this study have the potential of increases a user's look-up speed such as entity based filtering and a direct word search.

## Acknowledgment

We would like to acknowledge Julie Adams for teaching the Human-Computer Interaction class, which gave us the opportunity to pursue this app in a systematic way.

## References

1. T. Beck, J. Kirven, "Personal Diary Entry Retrieval Time on Task Test Plan," unpublished.
2. Manning, Christopher D., Surdeanu, Mihai, Bauer, John, Finkel, Jenny, Bethard, Steven J., and McClosky, David. 2014. The Stanford CoreNLP Natural Language Processing Toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 55-60.