Finding Business Information by Visualizing Enterprise Document Activity

Xiaoyu Wang
UNC Charlotte, USA
xwang25@uncc.edu

Bill Janssen
Palo Alto Research Center, USA
jasssen@parc.com

Eric Bier
Palo Alto Research Center, USA
eric@parc.com

ABSTRACT
In an enterprise environment, business information, such as project proposals and product discussions, is dynamic and often embedded in documents and document activities (e.g., emails, Web pages and office documents). Because this information is essential to business processes, corporate employees need an effective means to retrieve it. Some commercial products, including Google Desktop, provide keyword searches for finding some of this information. However, this approach is not always effective as successful keyword searches can be difficult to construct, and even the best queries may fail to find some important materials. In this paper, we present Taste (Temporal Activities and Story Telling), an interactive visual analytics system that enhances the enterprise employee’s capabilities in searching and sharing diverse and dynamic business information. Taste was designed, after interviews with corporate employees, to follow their information retrieval cues and help them manage, review and share the business information embedded in their document activities. Results of our lab and field studies validate that Taste provides employees the confidence and necessary features to more efficiently and effectively retrieve business information from their documents and activities.

Categories and Subject Descriptors
H5.2 [User Interface]: Graphical user interfaces (GUI)

Keywords
Temporal information visualization, personal information management, document management, collaboration.

1. INTRODUCTION
In the enterprise environment, employees’ document activities result in the creation of many information streams, including email threads, calendar entries, Web browsing histories, and versions of office documents. Many of these documents contain information essential to the operation of the business, such as project proposals and emails capturing product discussions. Thus, it is crucial for enterprise employees to have an effective means to manage these information streams and retrieve desired business information from them.

However, due to the dynamic and diverse nature of document activities, finding the desired information can be an exhausting task. Recent reports from Interactive Data Corp. (IDC) show that employees typically spend 3.7 hours per week searching but not finding information, and 2.5 hours per week recreating content that couldn’t be found [8].

One challenge in finding such information is coping with its diversity. Finding desired information may require using different tools or looking in different places as a result of the different behaviors and conventions of the many desktop applications, such as individual office document suites and email clients. This challenge becomes even more pronounced in projects that take place over long periods of time or that involve many people; both of these factors tend to increase the amount of information to be managed and the number of places in which the information is stored. It is observed that the complexity of finding such essential information can result in reduced productivity in an enterprise environment [8]. Therefore, there appears to be an urgent need for an information management system that can facilitate information search and retrieval in the enterprise environment.

Some commercial products, such as Google Desktop [7] and Apple Spotlight [1] have built-in document indexing that enable users to search for information with keywords. However, keyword search can be difficult and is often insufficient in an enterprise environment [19]. For example, if one can’t remember the name of a document or can’t think of any distinctive words or phrases in a document, a keyword-based search can be doomed to failure. In addition, if one only remembers a vague time frame of the occurrence of certain document, then searching through temporally-sorted results can be unacceptably slow.

Alternatively, as described in Thomson et al’s [4] theory, people remember and recall things through associations with other clues. A recent observatory study by Teevan et al [19] supports the theory that, instead of trying to directly locate the targeted information by keywords, people usually follow a chain of clues in finding the desired information. For example, users may not remember the particular title of a proposal or any text in it, but can find that proposal through some retrieval cues, such as, the person they have communicated with, the application they used, and even the rough time frame. In practice, the Feldspar system [3], has demonstrated the effectiveness of utilizing such retrieval strategy in managing document activities. Yet, research in applying such retrieval strategies in an enterprise environment is still in a preliminary stage. Our work on Taste is an investigation of some next steps.

In this paper, we focus first on identifying the specific retrieval cues that are typically used in analyzing document activities; next, we describe the design of visual abstractions, based on these retrieval cues, that enhance the user’s ability to search for and share business information.
To identify such retrieval cues, we conducted pre-design interviews with thirty employees in an international corporation. We found three retrieval cues that were rated highly among the employees: 1) The date and time when events happened. 2) Content keywords that are associated with document activities. 3) Document Types or Applications that are linked to a particular type of activity.

Based on our findings, we designed Taste, an interactive visual analytics system that enhances the employees’ capabilities in searching and sharing business information. Taste, as shown in Figure 1., is structured to embed the retrieval cues into a coordinated multiple-level visualization system. Seen at a high-level, Taste encodes these cues with a set of three visualizations, each of which presents a particular aspect of document activity information. In lower level views, Taste presents a visualization that integrates related activity information for a single document. Using this multi-level structure, Taste helps users to cohesively depict document activities from different points of view and to effectively find the desired information. Additionally, Taste extends the information seeking process from an individual effort to a group effort. Finally, Taste includes an interactive Story Telling view, which helps employees to collaboratively find, organize and share information between collaborators.

![Figure 1. System overview for Taste.](image)

To evaluate our system, we conducted a lab study and a field study. We found that most participants believed Taste to be useful and effective in helping them retrieve desired information.

The remainder of this paper is structured as follows: Section 2 describes previous work on tools for managing personal activities. Section 3 describes domain specific retrieval cues used by corporate employees. Section 4 details the design of the Taste system. Section 5 presents our laboratory and field studies, with detailed results. Section 6 provides a discussion and our future research directions, and Section 7 presents our conclusions.

2. RELATED WORK

In the psychology literature, Davies and Thomson [4] have shown that people often remember things in association with other things. Studies by Tulving et al [19] have demonstrated that the use of contextual information to connect and to recall individual activities is indeed essential in information seeking processes.

Based on these theories, many observation studies have been conducted to understand natural search behavior, from various kinds of activity information. For example, Whittaker et al [23] focused on analyzing email systems, while Jones et al [10] examined Web activities. Much of this research supports the theory that associations are important to human memory. More generally, Teevan et al [19] have conducted an observation study that focused on people’s search behavior across a broad class of electronic types, including email, files and the Web. This study confirmed the theories on a broader basis and suggested that, instead of trying to directly locate the targeted information by keywords, people often follow a chain of clues in their search for desired information. They also observe that the support of such information retrieval strategies is current insufficient.

Many computer science research and commercial products have provided support for various forms of searches. For example, desktop search engines, such as Apple Spotlight, Microsoft Desktop Search and Google Desktop, have been effective in helping people retrieve information based on keywords. Many systems have focused on enhancing keyword-based document search in various ways, including systems that utilize timeline-based visualizations [11,16,17], interfaces representing documents with spatial and facet information [13,14], and systems with interactive tagging capabilities that iteratively refine the search results [5]. However, while all of these systems support certain retrieval aspects, their support for the multi-level retrieval strategies are limited.

Four systems [3,6,12,18] are similar to Taste, in that they focus on integrating a broad range of electronic document types. However, there are significant differences between these systems and Taste. Compared to Haystack [12] and SIS [6], both of which present activity information using integrated interfaces, our system is customized to support the domain specific retrieval cues in an enterprise environment. While both Document Card [18] and Taste can present detailed information for a single document, Taste also provides a high-level overview that helps user to interactively navigate through large collections of documents. Although both the Feldspar [3] system and Taste support a multi-faceted retrieval strategy, Taste extends the information-seeking process from an individual effort to a group effort.

In the following sections, we describe our findings relative to a set of three retrieval cues that are highly rated by enterprise employees, and report on our visualization system, which is designed to take advantage of these retrieval cues and to enhance employee capabilities for information search and sharing.

3. COMMON RETRIEVAL CUES

To identify those information retrieval cues that are most used in an enterprise environment, we conducted 30 semi-structured interviews with employees from Xerox Corporation. These
interviewees held a broad range of positions, including product researchers who need to write proposals and research papers, managers who were in charge of business planning and marketing, and administrative staff members who oversee hiring. Our interview focused on examining three substantive questions:

### Table 1. Three substantive questions

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>How often do you need to search for document activity information?</td>
</tr>
<tr>
<td>2</td>
<td>What are the typical approaches you use in searching for desired information? Including both clues and applications.</td>
</tr>
<tr>
<td>3</td>
<td>How well do the applications that you use support your retrieval tasks?</td>
</tr>
</tbody>
</table>

#### 3.1 The limits in current practices

The results of our interviews add support to previous research (see Section 2) that suggests that, while some keyword-based approaches are used in workspaces, these applications are less tailored to support information seeking processes in enterprise environments. Many of the interviewees mentioned that they usually find it hard to describe things they want to find with keywords; some of them also told us that often the keywords they used were not precise enough for such search tools to have useful results. For example, one interviewee asked “How can I tell the search bar that, even though I don’t remember the title of a document I am trying to find, I do remember I have copied it to Alice in an email two weeks ago?” During our interviews, many employees expressed similar frustrations: they put in the effort but could not find any useful information.

From these interviews, we concluded that there is need for a tool that could enable users to be more effective at analyzing their activities and retrieving desired information.

#### 3.2 Identifying the essential retrieval cues

Many interviewees provided us with information about how they recall and search for document activity information in their daily routines. Combining their practical experiences and our review of the research literature [3,4,19], we categorized the retrieval cues into three major categories:

- **Temporal hints** contain significant information relating to document activities. Specifically, the interviewees mentioned that both the exact date and time when events occurred can be very helpful (e.g., when documents were received, read, created, or modified), and also the relative sequences of events.

- **Content keywords** (e.g., the title of a document or the name of a person or company) are often used in filtering down to the areas of interests during the initial stage of the retrieval process. In addition, it is observed that certain keywords can also help employees to associate and connect document activities together [3], and therefore lead them to recall events more precisely.

- **Document Types or Particular Applications** are also considered important clues for employees to locate desired information [13].

Many employees report that the first thing they think of in searching for document activities is often the applications they used.

In addition, many interviewees mentioned that these three retrieval cues generally coexist during their search for the desired information. Therefore, it is advantageous to integrate them into one system and present the cues to users cohesively.

### 4. INTRODUCING TASTE

In response to the identified needs for an integrated, efficient, information retrieval tool tuned to the enterprise environment, we designed Taste, an interactive visual analytics system that aims to enhance corporate employee capabilities for finding and sharing the business information that is embedded in their daily document activities.

In the following subsections, we first explain the data capturing process in Taste and then describe how each retrieval cue is depicted in our system.

#### 4.1 Data Capture and Storage

At the heart of Taste is an automatic and transparently real-time contextual data capturer. As part of the UbiDocs project [2], Taste runs on a single user’s computer, and captures user activities around office documents, calendars, emails, Web pages, etc.

Like a desktop search engine [7], Taste creates an index of documents on a user’s machine. However, unlike a desktop search engine, Taste also logs information about a user’s activities with documents. In particular, it collects information about which documents, emails, and web pages were read, and how long they were open. It also collects document metadata, such as information about the senders and recipients of email messages.

Taste stores this information, along with copies of the documents, into a unified document repository, UpLib [9]. As shown in Figure 2, Taste utilizes UpLib (Data Capture and Storage) to extract information from and about each document, including its title and authors, its text, the people and other entities that it mentions, its paragraphs and its images. All of the captured information is indexed and grouped with its related documents. So all documents sent to Alice are associated together, and are further interactively presented to the user through Taste’s visualization interface.

#### 4.2 Visualization Interface

Instead of presenting the diverse document activities through a keyword search interface, Taste embeds the retrieval cues into a coordinated multi-level visualization system.
At a high level, Taste encodes the three cues with a set of three visualizations, each of which presents a particular aspect of the document activity information. To provide a lower level detailed view, Taste also presents a visualization that integrates related activity information for a single document. Using this multi-level structure, Taste helps users to cohesively depict document activities from different points of view and to effectively find desired information.

4.2.1 High-level Visualization Overviews

4.2.1.1 The Temporal View presents temporal hints

The Temporal View shows how a user’s activities unfold over time. It presents both the number of documents a user interacted with in different time periods, and the types of those documents. This view is created as an interactive ThemeRiver [10], and it shows the temporal trends and patterns of a user’s document activities. In this view, each vertical axis represents a period of time, while horizontal ribbons indicate both the format of documents (i.e., email, Microsoft Word, etc.), and the time spent on each. For example, Figure 1(B) depicts the history of a day in which a user spent a significant time browsing Web pages and took a quick break around 3:00 pm.

Besides showing general trends and patterns, the temporal view also allows the user to drill down into time periods. When the user selects a time period on the horizontal axis in the center of the view, Taste zooms into that period of time. At the same time, a time period summary window appears on the desktop (see Figure 1(E)) that presents the highest ranked N documents for that time period and provides the user with a quick way to return to the original time scale.

4.2.1.2 Facet view shows Document Types

To help corporate employees efficiently retrieve specific documents of interest, we designed the Facet View to aggregate both the documents and the people that a user has interacted with during a particular period of time. Like Lee et al [14], our facet view allows the user to filter and sort information based on automatically-extracted data facets, including type (person or document) and format (email, slide presentation, text document, etc.). The user can query this view to filter its output based on a set of facets. For example, the user can choose to see or hide activities with email, with office documents, Web pages, or with people. Each of these facets can be turned on or off by pressing an associated button.

In order to fit the activity information into a reasonable amount of screen space and in order to draw the user’s attention to the most important activities in each time period, Taste sorts document activities by importance, and displays the most important documents at the top and with the most salient presentation by computing the importance value as:

\[ D_{importance} = F_{appearance} \times \sum_{T_{cur} \text{out}}^{T_{cur} \text{in}} T_{dwell} \]

In this equation, the importance of each document (\( D_{importance} \)) is set to the number of times each document appears in the repository (\( F_{appearance} \)) multiplied by the sum of the amount of time the document was open on the display (\( T_{dwell} \), measured in milliseconds). Document importance is computed in a particular time frame, which is between \( T_{cur} \text{start} \) to \( T_{cur} \text{end} \).

Therefore, based on this equation, Taste considers the document that a user spent the most time on to represent the most important activity the user performed during that time period. To give users a sense of the importance of each document, Taste displays scale bars next to each document, where the length of the bar denotes its importance in that time period (shown in light grey as the background of each document in Figure 1(A)).

To enable fast exploration, a summarized information panel (see Figure 1(D)) is shown when the mouse hovers over a visual element representing an activity. Like the Document Card [18], this panel includes a readable thumbnail and aggregated information about that visual element. If the user needs more details, the user can double click on the visual element to bring up a Detail view window (See section 4.2.1.5).

4.2.1.3 The Entity Tag view for Content Keywords

Since it is observed that content keywords from documents are helpful for information seeking, Taste extracts entities [2], such as company name, contacts, date/time, etc., from all of the documents the user has interacted with and displays them in the Entity Tag view.

To enable fast entry browsing and to emphasize the most frequently encountered entities from a selected time period, this view uses a TagCloud visualization of the entities. The size of each entity in the TagCloud is calculated based on the equation:

\[ \text{TagSize} = F_{Entity\_appearance} \times F_{Doc\_appearance} \times \sum_{T_{cur} \text{out}}^{T_{cur} \text{in}} T_{dwell} \]

In this equation, the entity appearance frequency (\( F_{Entity\_appearance} \)) is the number of times that the entity is mentioned in a particular document and the document appearance frequency (\( F_{Doc\_appearance} \)) is the number of times a document was used in a given time period.

As shown in Figure 1(C), the Entity Tag view colors the extracted entities based on their categories, i.e. company or person. When the mouse hovers over an entity tag, a summarized information panel appears that shows the top ranked six documents that mention that entity. The user can double-click on an entity name to open a persistent window with information related to that entity, including a larger selection of documents that mention it.

4.2.1.4 View Coordination

Since information retrieval can involve the utilization of all three retrieval cues, Taste is structured to encapsulate these visualizations within a coordinated system. All the visualizations in Taste are coordinated so that updates in one view are immediately reflected in the others. For example, if the user zooms in on a particular time period from the temporal view, the facet view responds by creating a new aggregated panel, and the entity tag view updates its displayed entities.

Therefore, a user can start the process of recalling document activities beginning from any retrieval cue that they remember. Updates in a coordinated view will often display information that is more similar to the desired information, allowing the user to follow a path through the visualizations and converge on the desired information quickly.

4.2.1.5 Low-level Detail View

While the above visualizations focus on presenting overviews of the entire collection, the Detail view depicts a single document
from multiple perspectives, showing temporal information, related document information (i.e. how many versions of this document are available), and other information. The Detail view can be invoked from any view in Taste to learn more about a document appearing in that view.

As shown in Figure 3, the detail view contains four panels. Figure 3A shows the preview panel that allows the user to view the document without having to reopen the corresponding application. By comparing paragraph, image, and page layout similarities [9], the related documents panel (Figure 3B) recommends documents that are similar in content to the selected one. The temporal information panel (Figure 3C) shows how much time the user has spent on that document and when activity with it occurred. Finally, Figure 3D presents the entity information panel in which the user can browse all of the entities that were extracted from the document that is presented in the preview panel.

![Figure 3. The Detail view for a selected document](image)

4.3 Seeking and sharing collaboratively

According to Pike et al. [13], the typical goal of information analytics is to create new understanding and communicate it to others. This sharing and collaborating factor is especially valuable in a large enterprise environment, where information can be spread out among multiple employees. In such workplaces, seeking information may require an organized effort by collaborating employees. Therefore, Taste supports the information seeking process for both individuals and groups.

By utilizing an interactive Story Telling view, Taste allows users to collaboratively find information, organize it and share it. In this view, the user takes a more active role in information tracking. Like the Detail view, this view is universally supported throughout Taste. Whenever a user comes across an interesting information object in Taste, they can right click on that object to add it to a new or existing Story view.

Once an element is in a Story view, as shown in Figure 4, the user can annotate or tag it [22]. The user can also perform basic grouping and sorting on the story elements. In addition, the user can drill down to more information about a given document or person by clicking on its icon.

The story created by one user around a collection of people and documents may be of interest to other users as well, so Taste allows stories created on one instance of the system to be shared with users on another instance. Other employees, who receive this shared story, are able to modify it based on their understanding of the topic and add or suggest removal of story elements. By sharing their stories about document activities, groups of employees can now understand those activities better and improve information retrieval for all members of the group.

While the story feature in Taste is quite new, we can already see that it may be useful for a number of applications, including capturing a related set of ideas, building an annotated bibliography, recording the history of a project, or sharing information about recent exchanges with a customer. We are still testing the usability and utilities of this feature in enterprise environments.

![Figure 4. A story that has been organized into 3 Story views.](image)

5. SYSTEM EVALUATION

To evaluate Taste, we conducted two studies: (i) a controlled laboratory study to assess the usability and utility of Taste’s interface with a fixed set of data and (ii) a field study to evaluate the effectiveness of Taste, which embedded with the tree retrieval cues, in aiding people’s information seeking process in users’ own work environments.

5.1 Study Design

5.1.1 Study Goals and Experimental Setups

For both studies, two conditions were examined: (i) the use of Taste; and (ii) the use of regular Mac OS X tools [1], hereafter referred as RT (regular tools), including Microsoft Office for Mac and Google Desktop. All participants experienced both conditions in a counter-balanced order.

For the field study, we first deployed the data capture module to each participant. Document logging for Taste was unobtrusive: there were no reported interruptions to the user, but each day we checked that the logger was active on all participant machines. Taste was able to sample around 200 documents during an 8-hour working day. The total number of documents sampled over the period of 1.5 weeks per participant was in the range of 1000 - 1200. Participants were instructed not to look at their data during collection. The Taste interface was not installed on any of their machines during this period.
In preparation for the lab study, we used Taste to collect one-month’s worth of workstation activity data. The resulting data became our test corpus. This data was diverse, containing different types of documents, web pages, images and emails. In total, there were 7,419 data objects, with over 24,275 pages. None of the participants had any prior exposure to this data. No new additions were made to the log during the study; so all participants accessed the same data set. For purposes of comparison, this data was also indexed by both Apple Spotlight and Google Desktop search. We ensured that all participants knew about the nature of the data set and locations (folders) of documents before the study began.

5.1.2 Participants
All of our participants from our hosting corporation were volunteers consisting of researchers, administrative staff, managers and business development staff. For our short-term controlled lab evaluation, we recruited 12 participants (5 female and 7 male) through email lists and by acquaintance. For the field study, six participants (1 female and 5 males) volunteered to use Taste on their personal data for 1.5 weeks during which their document activity was continuously logged. All were proficient users of the Mac OS X operating system and its applications. None of the participants had any prior knowledge of this project nor any prior experience with Taste.

5.1.3 Procedures
Since our participants would inevitably get more familiar with Taste and the sample dataset during the process of our studies [14], we took steps to isolate learning factors. In particular, we asked each participant to attend two sessions, separated by a two day time period. We hoped that over this time period participants would forget most of the details of the data set.

To balance performance gain, we randomized the sequence of conditions (using Taste vs. Regular Tools) across participants. In addition, the allocation of time to tool was balanced across participants. Each participant used either Taste or RT at the first session and switched to the other tool two days later.

To ensure that the study results were comparable across sessions, participants were given the same task questions for each tool; participants were asked to find information on the same time frame or about the same person.

5.1.4 Retrieval Tasks and Study Measures
In total, we asked each participant to carry out 12 tasks (6 tasks in 2 conditions). These tasks were identified based on our pre-design interviews, all of which are at least common in corporate daily tasks, if not predominant in all jobs. As shown in Table 1, we select these most representative tasks to evaluate the capability of Taste.

Table 2: Task questions for both conditions in the laboratory study (the word ‘tool(s)’ was replaced to reflect the condition)

<table>
<thead>
<tr>
<th>Task</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 1:</td>
<td>Please use tool(s) to find the most mentioned contacts within a particular time frame.</td>
</tr>
<tr>
<td>Task 2:</td>
<td>Please use tool(s) to find the most visited information sources within a particular time frame</td>
</tr>
<tr>
<td>Task 3:</td>
<td>Please use tool(s) to find and list the user's recent activities with Mr. Manager X</td>
</tr>
</tbody>
</table>

For each task, we collected and analyzed the following factors to measure the utility and usability of Taste:

**Accuracy**: the percentage of accurately retrieved items in the participant’s answers. In the lab study where standard answers were known, we measured this factor by comparing participants’ answers to the standard answers; in the field study, we analyzed this factor together with each participant after the two study sessions.

**Efficiency**: the time (in seconds) spent on answering each question.

**Confidence**: a 5-point Likert scale score measuring how confident users were in the accuracy of their answers.

When participants completed their tasks, we asked them two open-ended question about their experience using the assigned tool, and also asked them to score their answers on a 5-point Likert scale: 1) “How well do you think the provided tool covers your retrieval cues?”; and 2) “How do you like the design of the tools provided?”

![Figure 5](image-url) Study results for the three measured factors: efficiency, accuracy, and confidence. Lab study results are on top. Field results are on the bottom.

5.1.5 Results
Due to page limits, we only report the most significant results of our evaluations. In general, Taste has significant advantages over regular tools in the following four aspects:

**Taste provided better retrieval accuracy.**

Results from both studies suggest that Taste provided participants with more accurate retrievals. In our lab study where participants had no prior knowledge about the data, the results suggested a
Taste significantly improved retrieval efficiency.

As shown in Figure 5A and 5D, both studies suggested significant reductions in participants’ information retrieval time, with 36.8% ((176.3-111.3)/176.3) reduction in the lab study and 31.1% ((166-114.3)/166) in the field study. The ANOVA results (lab: (F(1,82) = 6.13 p < 0.015) and field: (F(1,78) = 10.84, p < 0.001)) show that Taste helps participants perform their tasks more efficiently.

Taste largely increased participants’ confidence

Initially, since participants had prior knowledge about the location of their own data, we were expecting the confidence value to be similar between Taste and RT in the field study; on the other hand, we expected Taste to increase user confidence in the lab study. While the lab results (F(1, 78) = 17.12, p < 0.001) partly supported our expectation, the field results yielded a significant positive confidence increase (ANOVA (F(1, 82) = 16.16, p < 0.0001)) when using Taste. Therefore, as shown in Figure 5C and 5F, participants trusted Taste more than the regular tools for retrieving their document activities.

Taste provided a more advanced interface.

In response to our post-task questions, Taste received an average 4.57 out of 5 in supporting users’ retrieval cues, even though participants had interacted with Taste for less than an hour. By contrast, the regular tools scored a 3.36 on these questions. Taste also received an overall score of 4.34 out of 5 for usability, suggesting that participants felt comfortable using Taste to perform the tasks. One participant noted, “This interface [Taste] provides me more control over the data. … it makes me feel more confident in searching for information”.

5.2 Summary

Given that participants had prior knowledge of the collected data, it seems likely that the accuracy of their answers depended largely on the tools provided. Since Taste provides a cohesive visual interface that incorporated highly rated retrieval cues, many participants positively rated Taste as an aid for information seeking. Our field study supported this result by suggesting that Taste performed better than regular tools on our three measured factors. One participant commented that, “‘By looking at all the [Taste] interface, I can easily relate all the information together and effectively examine my activities from different aspects.’”

Even if there is no previous knowledge about the data, our laboratory study suggests that Taste’s interface is designed sufficiently enough for users to follow important clues in the tasks and outperform the results from using regular tools. One participant noted that, “[Taste] is very good at giving a quick impression of data across the board. I like how you can mix different types of files and people together and represent them interactively.”

In addition, many participants found the idea of collaboratively searching for information interesting and the way we approached feature practical and useful. Although we didn’t set up a collaborative environment for participants due to privacy concerns, participants were still interested in utilizing the Story view and tried to share findings between different instances of Taste.

While Taste has only been used by a small number of people so far and for a short period of time, it appears to be a promising technology. Participants enjoyed using Taste to retrieve information and shared comments like “I can definitely see myself using it regularly”; and “…it can be quite helpful when I need to quickly put together some research reports”. Based on feedback we have received, we believe this kind of visualization interface can help corporate employees

6. DISCUSSION AND FUTURE WORK

6.1 Parameterize the document changes

While seeking business information embedded in multiple document activities is important, understanding the changes of different version of a single document is also crucial in dynamic corporate environments. For example, the ability to track changes in different versions of legal documents can have a tremendous impact on how a company operates and how it adjusts to new rules and regulations. Unfortunately, such tracking is not generally supported in current personal information management tools.

Taste has begun to embed this function into the Detail view, so users can visually depict progress and changes among different versions of a document. Although the current interface is still in a preliminary design, many participants have noticed the potential impacts of this feature and commented on the usefulness of it to help them better track and update documents.

6.2 Embedding personal knowledge in the document repository

Employees generally have their own methods for organizing and relating business information, based on their job requirements and individual experience. It is therefore interesting to consider whether externalizing such knowledge and embedding it into customized visualizations would be feasible for enhancing their information retrieval processes.

Although there is no definitive way to achieve complete knowledge transfer, our previous research [21] suggests the possibility of integrating a knowledge structure with visual analytics systems that can help users capture, store and reuse domain knowledge. Specifically, we are interest in analyzing the story elements a user has interactively collected using Taste, and translating them into knowledge artifacts that can be externalized to document repositories and then used as input to personalized search methods.

6.3 Collaborative information retrieval

In an enterprise environment where documents and information are generally decentralized, efficient information retrieval processes may require an organized effort including collaboration with other employees. While Taste has touched on collaborative information retrieval by providing Story views, we intend to extend this collaboration support to more effectively support groups as they seek to understand the history of their activities and find the information they need.
Finally, we plan to deploy and test our approach in enterprises with different activity patterns, social networks, and document types to see how well it supports these environments and what modifications are needed to make it most effective in each.

7. CONCLUSION
In an enterprise environment, documents are of great importance to business operations and information flows. It is therefore essential for corporate employees to have an effective means to retrieve this information and use it in their work. Although current commercial products present efficient methods for keyword-based searches, they are not as effective in an enterprise environment, where information is hard to find by keywords alone. In this paper, we present Taste, an interactive visual analytics system that presents an integrated document retrieval interface. As shown in Figure 1, Taste has been implemented specifically to help corporate employees manage, review and share information about their document activities. In both our lab and field studies, statistical results suggest that Taste is more effective in retrieving document activities than some traditional tools. Participants have expressed an interest in using Taste in their daily tasks.

8. REFERENCES