Finger on the Pulse: The Value of the Activity Stream in the Enterprise

Ido Guy, Tal Steier, Maya Barnea, Inbal Ronen, and Tal Daniel

IBM Research-Haifa, Israel

{ido, talst, mayab, inbal, taldan}@il.ibm.com

Abstract. The activity stream, which syndicates user activities across social media, has been gaining popularity on the web. With social media infiltrating the enterprise and higher portions of the workforce becoming accustomed to consuming information through activity streams, it also has the potential to play a key role in shaping the workplace. This work provides a first comprehensive study of an enterprise activity stream. We analyze different characteristics of the stream, its usage through a faceted search-based application, and the way users search it compared to traditional enterprise search. We also discuss various use cases of the stream, both from an individual employee's perspective and from an organizational perspective, exposing the potential value and role of the activity stream in the enterprise of the future.

Keywords: Activity streams, collaboration, cscw, enterprise, enterprise search, real-time search, social analytics, social business, social media, social search, social software, social streams, web 2.0

1 Introduction

The *real-time web* [12] is characterized by highly intensive streams of updates and news generated by millions of users. Leading social media sites, such as Facebook, Twitter, LinkedIn, Myspace, and Google+, publish activity streams that include millions activities per day by users who write status updates, share links and photos, join groups, comment, and "like" others' activities.

Enterprise social media applications, enabling employees to share and interact behind the organization's firewall, have become popular in recent years and were extensively studied, from blogging [10] and microblogging [34], through wikis [4] and forums, to social bookmarking [25], social file sharing [31], collaborative task management [26], and people tagging [13]. Recently, several enterprise social media platforms have introduced enterprise activity streams, which syndicate employees' activities across the organization's social media. Examples include SalesForce Chatter(.com), SAP StreamWork(.com), Yammer(.com), and SocialCast(.com); however,

the use and potential value of activity streams in the enterprise has yet to be extensively studied.

The emergence of activity streams within the enterprise poses a great opportunity both from an individual (egocentric) and an organizational (sociocentric) perspective [14]. For the individual employee, this unique medium of highly intensive activities, concise in text and metadata, allows staying tuned with recent updates and discover new developments related to areas of interest. On the organizational level, the stream exposes timely information flowing at high pace from a wide variety of employees. As such, it can help increase awareness of projects and processes and expose recent trends and opinions to executives, administration people, or community owners.

In this work, we examine the potential value of the activity stream to the enterprise. We study our organization's activity stream, which syndicates activities across a wide variety of social media applications, including blogs, microblogs, wikis, files, bookmarks, and forums. All of these applications have been deployed in our organization for over three years and enjoy a large user base and frequent activity. Within our organization, an activity, defined as the basic unit of the stream, occurs at a frequency of less than four seconds during working hours. For instance, an activity in the stream can be: "*John Doe commented on the blog entry The Future of CRM in the Social Business blog*."

In our study, we used *Streamz*, an application that provides faceted search and analytics on top of an enterprise activity stream [18]. We analyzed the query logs of *Streamz* over a period of two months and interviewed 20 active users, in order to gain a better understanding of the activity stream usage scenarios from an egocentric perspective. Our analysis also focused on the comparison between search over the activity stream and traditional enterprise content search. We found that stream search is essentially different than "regular" enterprise search and serves for other complementary needs. Additionally, we augmented *Streamz* with a user interface that demonstrates the use of organizational analytics over the activity stream. Following the first set of interviews, we identified three lines of business (LOBs) that can benefit from such analysis: sales management, human resources, and IT. We conducted 8 additional interviews of users that belong to these LOBs to better understand their potential use of the stream from a sociocentric perspective. Overall, we believe our study provides a first comprehensive look at the potential use of the activity stream in the enterprise. Understanding this potential use can be fundamental to how employees collaborate and interact in the future workplace.

In the next section we discuss related work. We then describe the research setting, including a brief overview of the *Streamz* user interface and a description of the additional organizational analytics view. In the following section we describe a rich set of results, based on both quantitative and qualitative analysis, and then conclude.

2 Related Work

A large portion of the literature that relates to social streams studied the use of Twitter, the leading microblogging service (e.g., [22, 24]). Despite the vast popularity of

research on Twitter, its search functionality has thus far received relatively little attention. Busch et al. [5] describe the architecture of Earlybird, the core retrieval engine behind Twitter's real-time search service. Naveed et al. [27] also examined search on Twitter, focusing on two challenges—content sparsity and the need for a static document quality measure. In 2011, the Text Retrieval Conference (TREC) initiated a microblog track that exposed a reusable test collection for the "study of Twitter as a real-time information repository" [28]. Teevan et al. [33] examined the use of search on Twitter and how it differs from usual web search. In our analysis, we compare search on *Streamz* with enterprise content search and link our results to their findings.

Microblogging services have also emerged in enterprises, used by employees to share messages about their work, promote new ideas, and converse around topics of interest. The use of microblogging in the enterprise has been studied in several papers. Zhang et al. [34] presented a case study of the use of Yammer, an enterprise microblogging tool, in a fortune 500 company, and showed that employees use it for a variety of work-related purposes, despite its "noise-to-value ratio". Ehrlich and Shami [11] compared microblogging inside and outside the workplace and found that internal microblogs were mostly used to solicit technical help or as part of a conversation, while the external microblogs were used for status updates and sharing of general information. Zhao et al. [35] explored microblogging's impact on informal communication at work and found that it may help colleagues know each other better and facilitate collaboration at work.

The Twitter stream principally consists of "tweets"—short status update messages of up to 140 characters. Our work focuses on a heterogeneous stream, in which status updates account for less than 10% of all activity. Other heterogeneous activity streams, such as the Facebook newsfeed, have been studied to a much lesser extent than Twitter. Paek et al. [29] used learning techniques to predict the importance of posts and friends on Facebook's newsfeed. They found that importance ratings varied greatly from one user to another and suggested personalization as a promising direction to address this issue. Other studies about the newsfeed of Facebook focused on privacy issues [3, 20] and diffusion models [32]. FriendFeed is another example of a heterogeneous activity stream that received some attention in the literature [7, 16], aggregating friends' activities across other social media sites.

A few recent studies examined activity streams in the enterprise, focusing on personalization techniques. Freyne et al. [15] proposed a method for narrowing the stream of the SocialBlue enterprise social network site based on person and action relevance inferred from users' browsing behavior. Daly et al. [8] suggested viewing the activity stream through "social lenses", based on user-defined collections of people and entities. Guy et al. [17] studied personalization of the stream based on a user model that includes people, terms, and places of interest. Our paper does not focus on evaluating methods for stream personalization, but rather provides a broad analysis to better understand the potential value of the activity stream to the enterprise.

We analyze the stream through the *Streamz* application, which provides faceted search on top of an enterprise activity stream and is described in our previous work [18]. Other studies have also proposed the use of faceted search to slice and dice aggregated feeds. Visual Backchannel [9] applied a faceted search approach on microb-

log conversations to show related people, images, and topics. Similarly to *Streamz*, they also use a stacked graph to present topics over time. The Eddi tool [2] represented topics as facets and suggested an alternate Twitter interface that allowed topic-based browsing. Hong et al. [21] introduced FeedWinnower, an enhanced feed aggregator that allows knowledge workers to filter feed items by topic, person, source, and time. Only an initial evaluation was provided, based on interviewing 15 employees, which pointed out two findings: all participants saw the value in faceted browsing and the most important facet was found to be the topic facet. In this work, we explore an activity stream of social media behind the firewall, rather than external feeds of websites or microblogging services.

3 Research Setting

3.1 Enterprise Activity Stream

We experiment with the activity stream of IBM Connections (IC) [23]—a social media application suite for the enterprise that includes eight types of applications, all of which have been deployed in our organization for over three years. IC publishes an activity stream of all public actions occurring across its applications. Table 1 details the different types of activities in the IC stream, including the originating applications, all possible actions, and frequency of occurrence as measured during one month of activity (percentage out of the entire stream).

3.2 The *Streamz* Applications

Our research is based on the *Streamz* application [18], which was designed to help users consume an activity stream. *Streamz* provides faceted search over the enterprise activity stream, as well as different types of analytics, including topic extraction, sentiment analysis, activity grouping, and personalization based on an interest profile.

Table 1. Activity types included in the IC stream.

Application	Actions	% Occurrence
Blogs	create, edit, comment, like	4.09
Bookmarks	create	5.93
Files	create, edit, share, comment, like, download	14.53
Forums	create, reply	7.82
Microblogs	create, reply on own or another's board	9.45
Network	add to network, follow, tag self or another person	23.16
Tasks	create, edit, assign, comment, complete	8.25
Wikis	create, edit, comment, like	26.78

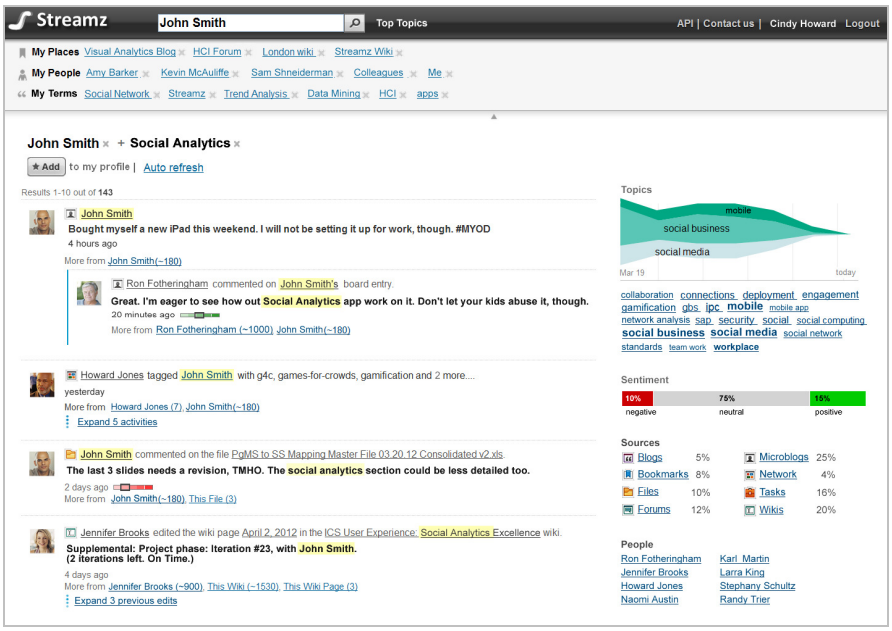

Fig. 1. *Streamz* main user interface.

Main View. The main user interface of *Streamz* is shown in Fig. 1. The user can search the activity stream in various ways, but ultimately one user interface is used to present any subset of the stream. The UI consists of three main parts. Its main component, the stream (on the left), displays the activities in the stream based on the search criteria currently applied. Activities are presented in reverse-chronological order, with indication of their freshness (e.g., "2 days ago"). Each underlined entity within the activity description is a link to its corresponding IC page. The "more from" line allows the user to re-filter the stream to any of the entities in the current activity. The second component appears on the right and includes the facets [1], which summarize various aspects of the stream, including key topics, sentiment, source application, and active people. By clicking a facet, the user can drill down (refine) the results to the clicked value. The third component is the user's interest profile, which appears on the upper section for authenticated users. The profile includes people, terms, and places (wikis, blogs, forums, etc.) of interest and serves for personalizing the stream to a mix of activities according to the interests of the individual user [17]. The user can add profile items by saving search queries that yield interesting results.

Overall, a user of *Streamz* has four different ways to search the activity stream: (1) through a manual search query in the search box at the top of the page; (2) by clicking a profile item; (3) by clicking one of the "more from" links; and (4) by clicking a facet. The first three initiate a new query, while the latter refines the current results.

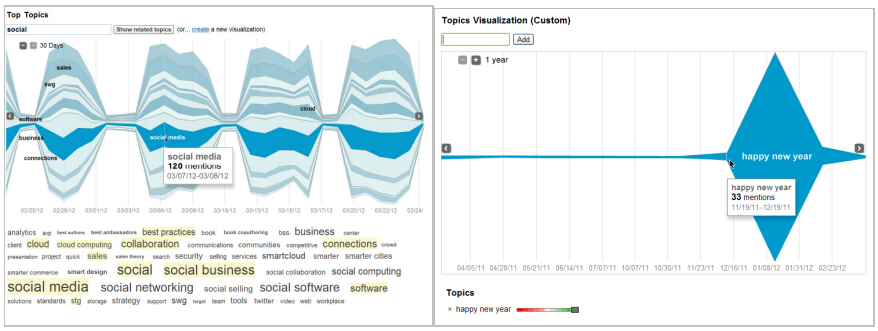

Fig. 2. Organizational views (Left:) Top topics view (Right:) Custom visualization view.

Streamz uses various analytics components: (1) topic extraction, which associates the key topics with each activity; (2) sentiment detection, which associates each activity with a one of three sentiment classes: negative, neutral, and positive; (3) activity grouping, which collapses together similar or related activities. *Streamz* implements three types of grouping: a microblog *thread*, which includes a message and all its replies (see the first activity in Fig. 1); a *duplicate* activity, which includes consecutive wiki or blog edits that are identical but have different timestamps (see the last activity in Fig. 1); and a *compound* person tagging in case a user has tagged the same person with multiple tags (see the second activity in Fig. 1).

A detailed description of *Streamz*, including its user interface and analytics, is provided in [18].

Organizational View. For this study, *Streamz* was extended with a user interface that enables to further gain organizational insights from the stream as a whole. To this end, we added the top topics page, which can be reached by clicking the corresponding link at the top of the main view. The page presents in detail the leading topics in the organization, as extracted from the entire activity stream. The user can also enter a query to view the top topics that relate to that specific query. Fig. 2 (left) illustrates the top topics view. The upper part is a stacked graph visualization [6] of the top 20 topics over time. The user can change the time range from 30 days to 3 months or to a year. The time range is partitioned into 12 sub-periods. Hovering over a layer in the graph displays a pop-up window showing the topic represented by this layer, the sub-period corresponding to the current cursor position, and the number of activities that mention the topic during that sub-period. The user can move the cursor vertically to move between layers or horizontally to move across sub-periods. The word cloud below the stacked graph shows the top 100 topics that relate to the search query (or to the entire stream, if no query was entered). Topics that are part of the user's interest profile are highlighted. Upon clicking a topic, the user moves back to the main view, presenting activities that match the topic.

Users are also allowed to create their own custom stacked graphs by clicking the 'crate a new visualization' link. In this view, illustrated in Fig. 2 (right), the user can

add and remove topics of her choice to an initially-empty stacked graph. A sentiment bar for each of the input topics appears below the stacked graph. This view is useful to track the evolvement of a single topic or compare a few topics over time.

3.3 Experimental Setup

Our evaluation is primarily based on an analysis of the *Streamz* query logs, which document every query issued to the system along with a timestamp, the user's IP address, and ID if they are authenticated. We analyzed the logs recorded during two months of activity. We also conducted interviews with 20 users who used *Streamz* during the selected time period to get an in-depth understanding of why and how they use *Streamz*. Interviews were semi-structured and covered general use of enterprise social media, usage scenarios of *Streamz*, and different aspects of search in *Streamz* (repeated queries, saved queries, and differences from traditional enterprise search). They were conducted by phone and lasted half an hour each. The interviewees originated from 11 countries and spanned different usage levels of *Streamz*. Three of the 20 interviewees were managers. Following these interviews, we identified three lines of business (LOBs) that can potentially benefit from the stream's organizational viewpoint, and followed up with 8 *additional Streamz* users who represented these LOBs. These interviews lasted 45 minutes, conducted by phone, and focused on the potential use of the stream to the LOB. Interviewees were presented with the organizational view user interface as a demonstration for how insight could be drawn from the stream as a whole.

4 Result Analysis

Our results are organized as follows. We open by a general descriptive analysis of the organization's activity stream along the inspected time period, to provide a sense of what the stream consists of. Since this is a new domain of research, we thought it is important to describe these characteristics as observed in our organization. We then describe the egocentric use cases of the enterprise activity stream as reflected in our initial interviews and through the usage analysis of *Streamz*' main UI. The following subsection inspects in detail the key functionality of searching the activity stream through analysis of the queries issued to *Streamz* along the time period. These are compared with the query log of a traditional enterprise search engine. The final subsection describes the organizational use cases as reflected through our follow-up interviews with LOB representatives.

4.1 Stream Statistics

Overall, the activity stream included 628,947 activities along the 60-day period. As expected in an enterprise environment, a sharp difference existed between the number of activities on a working day—13,285 on average (stdev: 3,937, median: 13,638,

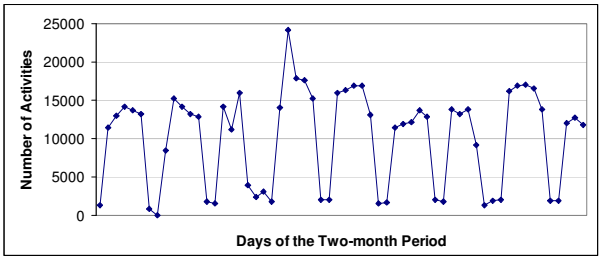

Fig. 3. Activity in the stream along the inspected time period.

max: 24,181), compared to 1,735 on a weekend day (stdev: 630, median: 1,833, max: 3,089). Fig. 3 shows the volume of stream activity over the inspected time period.

As mentioned before, the rightmost column of Table 1 displays the occurrence frequency of activities per source application (percentage of activities out of the entire stream). Wikis were the most common source of activity, producing over a quarter of the activities in the stream. The fact that each wiki edit saved generates another activity inflates the number of wiki activities. Network activities and files were the next most common activities. Microblogs constituted only 9.45% of the stream, and blogs were the least common with only 4.09%.

The activities of the stream originated from 56,866 distinct authors and 51,144 distinct places. The top 10 authors, all with over 750 activities along the time period, originated from 9 different countries and 4 organizational divisions. The average number of activities per author was 11.07 (stdev: 57.34, median: 3, max: 1,340). Fig. 4 (left plot) shows the distribution of activities by author, which follows a power law with slope $\alpha=-1.64$. The average number of activities per place was 9.65 (stdev: 36.43, median: 2, max: 1,964). The top 10 active places, all with over 1,000 activities, included 4 wikis, 2 blogs, 2 forums, and 2 tasks. Fig. 4 (right plot) shows the distribution of activities by place, which follows a power law with slope $\alpha=-1.58$.

The average number of words per activity as presented on *Streamz* was 28.46, and varied substantially among activity types, from 8.53 for network activities to over 45 for blogs, forums, and wikis. File and bookmark activities were also concise, while microblogs and tasks were close to the average. After applying sentiment classifica-

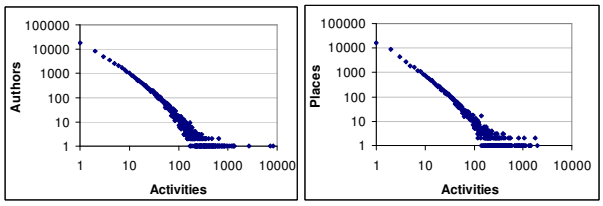

Fig. 4. Distribution of activities (Left:) over authors and (Right:) over places.

tion, 16.43% of the activities were identified as positive, whereas only 0.74% as negative. One of our interviewees reflected: "*We have a very 'positivish' organization, it's a cultural thing […] we are implicitly encouraged to say good things and avoid saying bad things*." Inspecting negative activities, we observed that they commonly referred to technical issues or bugs. A smaller portion expressed a negative opinion about a topic, a technology, a tool, or a process.

The highest portion of non-neutral activities was detected for the wordy sources—forums, blogs, and wikis (about a third were non-neutral)—and for microblogs (28%). The other sources had 10% or less non-neutral activities, with almost 100% of the network activities being neutral. Overall, we found the set of activities in the stream to be diverse in length and sentiment. This may affect the way it can be used by different tools or systems. For example, an application that is mostly interested in sentiment analysis may only focus on activities that stem from microblogs, forums, blogs, and wikis, disregarding activities referring to bookmarks or to the network.

After applying our three grouping methods, the number of activities in the stream shrank to 83.15% of the original. Table 2 summarizes the statistics, including the number of grouped activities per each type, their average length, and their impact on reducing the number of activities. Grouping of duplicate activities was the most common type, leading to a reduction of almost 40% in the overall number of wiki activities. A smaller reduction was associated with blogs. Threading substantially shrank the number of microblogs, while people-tagging activities shrank by almost 60% due to "compound" grouping. The latter produced the highest length of grouped activities—almost 5 activities per group on average. Overall, we observe that by several basic rules, grouping can substantially condense the stream. All three grouping methods we experimented with produced an average group length of 3.5 or higher. Further research is required to validate that grouping renders a more readable stream, and to extend grouping to additional sources and methods, such as grouping by topic or by place (e.g., 3 people edited a wiki). Grouping can also play a key role when filtering the stream, since a grouped activity may attract extra interest.

4.2 Egocentric Usage Scenarios

In this section, we analyze the personal (egocentric) use of the activity stream as reflected through our set of 20 interviews and the use of the *Streamz* main UI along the inspected two-month period of our study.

Table 2. Characteristics of grouped activities.

Type	Number	Avg Length	Source	Reduction
Duplicate	26,978	3.5	wikis	39.29%
			blogs	12.79%
Thread	8151	3.55	microblogs	35.39%
Compound	5263	4.91	person tags	59.93%

During the study period, 239 authenticated employees used *Streamz*. They originated from 26 countries and spanned the different divisions in our organization: 34.7% were from Software, 22.6% Services, 21.8% Sales, 9.2% Research, 7.1% Headquarters, and 4.6% others. Of these users, 14.2% were managers, which is similar to the percentage of managers in the entire organization. None of the top 10 users were managers. These results differed from those found for people search in the enterprise [19], where the majority of top users were managers. In contrast, it appears that activity stream search is as equally popular for employees as it is for managers. We note that this finding reflects the early adoption phase and may change with time.

Only 38.08% of the users used *Streamz* for more than one day, implying that the use of the activity stream in the enterprise is still in its early adoption phase. The average number of days per user was 2.46 (stdev: 3.08 max: 23). 28.57% of the *Streamz* users did not author any activity in the IC stream over the inspected time period. On the other hand, 18.07% were avid users of IC, who authored at least 100 activities during that period. The average number of authored activities by *Streamz* users was 79.43 (stdev: 172.77, median: 19.5, max: 1,340). Overall, users spanned a wide range of IC activity level, from non-active to highly active.

Many interviewees mentioned the term *dashboard* with regards to how they use *Streamz*. "*[Streamz is] useful as a dashboard view that provides a snapshot of the ongoing activities across a wide variety of applications in one place*," said one interviewee and another noted: "*It's an awareness system, monitoring what's going on, what are the things I need to know*, *what matters to me*." Another said: "*Similarly to Facebook, Streamz gives you a diverse feed of things you care about […] It really allows to keep your finger on the pulse of the organization*."

Another common usage scenario was getting a sense about what people say and think. "*I use Streamz to stay updated with what is being said about our projects*" said one interviewee. Another noted, "*Streamz gives me the 'talk' or 'vibe' about something going around […] understand who talks about it and where.*" This benefit was particularly highlighted by salespeople: "*I use Streamz to understand what people think about new material we create for a sales opportunity. For example, we created a presentation deck and want to understand how much traction it gets*" noted one interviewee. Another salesperson said: "*I usually look for competitive information […] what people write about our own and others' products, what feature they want to see*."

Some interviewees mentioned that their interest in the stream revolves around individuals or groups, for example, "*I use Streamz to get a glimpse of what a group of people I'm interested in is doing,*" or "*I mainly look for my peers, my department, and my report chain to stay up to date*." Another interviewee revealed, "*When a new CEO was appointed and said she would use social media to communicate, I searched to see how many people started to follow her*."

Using *Streamz* to track specific topics of interest was also frequently mentioned. "*I use Streamz to see what's new in the technologies and subjects my work involves*" said one interviewee. Another told us: "*I work in the finance industry and I use Streamz every day to see what's new in the finance sector […] usually using three keywords: finance, economics, and banking*."

4.3 Searching the Activity Stream

The following section focuses on analysis of the search operation on top of the activity stream, in an attempt to understand if and how it is different from traditional enterprise search. 145 authenticated users (60.67%) issued at least one query through one of the four mechanisms that enable doing so in *Streamz*. On average, each of them issued 8.55 queries (stdev: 14.11, median: 3, max: 97), 5.75 of which were manually-typed search queries, 1.41 were issued by clicking on a profile item, 1.08 by clicking on a facet, and 0.31 by clicking "more from".

On average, 7.35 of the 8.55 queries were unique (85.96%). Teevan et al. [33] reported a substantially higher portion of repeated queries for both Twitter (55.76%) and the web (34.71%). The low percentage of repeated queries may reflect the fact that *Streamz* is not yet used so frequently as Twitter and web search engines. However, it is also likely that *Streamz* users take advantage of the ability to save queries to their interest profile, thus sparing the need to repeatedly search for the same things. This point was made clear by several interviewees. For example, one said: "*When my query produces interesting results I save it to my profile so I can keep track*." Inspecting the profiles of *Streamz* users, 76.1% were non-empty, with an average of 11.2 saved queries per profile (stdev: 9.74, median: 9, max: 58), indicating that query saving was indeed in common use.

We next compare the queries issued to *Streamz* with queries issued to a more traditional enterprise content search engine, called *SaND* [30]. *SaND* is used to search for documents in the intranet, such as bookmarked pages, blogs, or wikis. It is not the official intranet search engine in our organization, but is used by several hundreds of employees every month. Alongside the document results, *SaND* presents related people, related tags, and source applications as facets, in a similar manner to *Streamz*. The similarity of both the usage levels and the user interfaces of *Streamz* and *SaND* enabled us to more neatly examine the differences stemming from the type of searched objects (activities versus documents) and the corresponding sorting of results (by recency versus by relevance).

To get a sense of the basic difference in result freshness between the two search applications, we inspected the overlap in the top 10 results for the top 50 queries common to both *Streamz* and *SaND* over one week. We found that the average overlap for *Streamz* across the 50 queries was only 14.2% (stdev: 8.6%, median: 10%, max: 60%), while for *SaND* it was as high as 92.4% (stdev: 11.6%, median: 100%). These differences demonstrate that within a week, most *Streamz* search results are replaced with new ones, while the vast majority of *SaND* results remain steady. This highlights the dynamic real-time nature of stream search, where results continuously change, and further motivates the comparison with traditional search.

Table 3 compares the characteristics of queries in *Streamz* versus *SaND*, based on the 50 most popular queries in each, as well as the top 50 queries common to both. Query popularity was determined by the number of unique individuals who issued the query at least once over the inspected time period. Teevan et al. [33] conducted a similar comparison between Twitter and web queries and found that Twitter queries were noticeably shorter than web queries (12 chars 1.64 words versus 18.8 chars 3.08

words, respectively). By contrast, in our results, *Streamz* queries were found to be somewhat longer than *SaND* queries. In general, queries in the enterprise appear to be shorter than queries on the web.

While a considerable portion of *Streamz* queries referred to individual people (18%), none of the *SaND* queries did. As already mentioned, one of the key usage scenarios for stream search is keeping track with the activities of a person of interest. Teevan et al. [33] also found a substantially higher portion of people searches in Twitter than on the web: 15.22% of the Twitter queries referred to celebrity names, compared to only 3.11% of the web queries. In our case, the people sought for are not celebrities, but rather coworkers or interesting employees in the organization.

On *SaND*, 36% of the queries referred to internal tools, such as a travel reservation system, a procurement system, a CRM system, and utilities such as an internal browser plugin, a survey tool, and a URL shortener. These queries are navigational in nature, i.e., they aim to get to a particular resource. The *Streamz* queries, in contrast, barely included references to internal tools. Teevan et al. [33] found a similar trend: while many of the web queries were navigational, Twitter queries were not.

A considerable portion of both the *Streamz* and *SaND* queries referred to internal products or projects (24% and 34%, respectively). Based on the comments we received in our interviews, it seems that the use of each application to search for projects or products is different. *SaND* is often used to find the main page of the product, or to locate its internal deployment. *Streamz*, on the other hand, is used to learn about latest features, opinions, plans, or issues that relate to the project or product. Another common query type to both *Streamz* and *SaND* referred to global topics or themes, such as "mobile", "user experience", "social media", "big data", and "travel". Again, the use of each application to search for such topics seems to be different, as demonstrated by one of our interviewees: "*For a query like 'cloud computing', I would search SaND to get to the main wiki or forum, while I would query Streamz to get a sense of the recent buzz around it and understand how it can be relevant to my job*." The list of popular common queries includes two main types: internal projects/products and topics/themes. As mentioned, the usage scenarios for these may often differ between content and stream search.

Table 3. Characteristics of *Streamz* queries versus *SaND* queries and queries common to both.

	Streamz	***SaND***	**Common**
Avg length (chars.)	**10.38**	8.72	8.4
Avg length (words)	**1.4**	1.26	1.24
Is person	**18%**	0	2%
Is internal tool	2%	**36%**	2%
Is internal proj./prod.	24%	34%	**50%**
Is topic/theme	**34%**	28%	**34%**
Starts with #	**16%**	0	2%

The hash symbol was used in quite a few of the popular *Streamz* queries (16%), although a smaller percentage than found for Twitter queries (21.28%) [33]. This is likely because microblogs occupy only a subset of the stream, and the rest of the activities do not typically include hash tags. The three most sought-for hash tags were: *#socbiz* (referring to a theme), *#ls12* (business event), *and #helpmeit* (a mark for the IT department). None of the popular *SaND* queries included hash tags.

Our interviewees shed more light on the differences between content search and stream search. Content search was said to be used for navigation to pages in the intranet of internal tools and projects and for finding the main projects or products related to a topic, or other 'stable' information, such as tutorial presentations, sales material, or marketing methods. On the other hand, *Streamz* was said to be searched for getting a sense of what people are saying right now about a topic, find something that was said before, get updated with recent news regarding a topic, a product, or a person, and understand the latest insights and buzz. One interviewee said: "*Streamz helps since it connects the content with the person who created it and the place it came from […] then, for example, I can locate people with similar problems.*" Another interviewee explained the difference when it comes to repeated queries: "*In SaND I repeat a query to re-find the same link, while in Streamz I repeat it to get the latest updates*." Another referred to the analogy of **searching** the content versus **filtering** the stream: "*[In Streamz] I don't search, I filter. Rather than trying to find something I need, I am narrowing the stream with different signals I remember such as people, keywords, and time range, in the hope to discover interesting updates*".

The value of the results' freshness in *Streamz* compared to content search was demonstrated by a few interviewees. One commented: "*I was looking for information on an annual organizational process and in SaND I found old pages since this process has been going on for many years. In Streamz I could instantly find the reference to the process taking place this year*". Another interviewee said: "*Streamz gives me the freshest information someone has bookmarked or wrote status about, while in SaND I often get old links sometimes even no longer valid*". Another stated: "*On SaND, the results are expected, while on Streamz I look to discover new things […]; often I find interesting stuff I did not even know I was looking for: a new interesting document, a new reference to my work, or even someone I know who got promoted and people congratulate him on his board.*"

4.4 Organizational Usage Scenarios

31.8% of *Streamz* users accessed the organizational view. 83.1% of the interactions with the custom visualization view referred to addition of new terms to the visualization, while 16.9% referred to term removal. The most popular terms inspected through that visualization included the organization's name, its leading brands and divisions, prominent product names, broad area names ("social media", "cloud computing"), bursty topics (e.g., a company acquired by the organization), events (both one-time and recurring), people names, and source names ('wikis', 'forums', 'bookmarks').

Some of our interviewees saw the value of the top topics view from an egocentric perspective, for example: "*It gives me a company-wide view of what's going on, like*

reading the headlines of a newspaper" and another described: "*I look at this page to see the main topics, what is hot and talked about […] I then select topics of importance that day to get a further look at the activities around those*". Another interviewee said: "*I use this view to search a person to see what he writes about the most and how his topics change over time*".

Through our interviews, we identified three specific lines of business in the organization that can benefit from the organizational view provided by the stream. We elaborate on these scenarios in the rest of this section and demonstrate each through quotes from follow-up interviews, conducted with four sales managers, two HR people, and two IT people, who used *Streamz* throughout the inspected period.

Sales Manager Scenario. Social media opens more channels for salespeople to document their activities, for example through blogs, microblogs, wikis, or shared files. One of our interviewees noted: "*Reporting on sales activity, such as customer meetings, through social media, especially microblogs, can serve as a good replacement for current CRM tools, since [the former] are much simpler and faster to use. This is especially important since sellers are usually on the road and need a simple interface they can use on their mobile device.*" Sales managers and executives now have the opportunity to learn more about the activities in their groups or divisions by analyzing the stream. Our follow-up interviews with four sales managers revealed several specific scenarios and requirements:

- "*I would like to be able to search for a customer name and see my group's volume of activity along a time-line, with a possible breakdown to each person in my group*"
- "*[…] use the 'top topics' view to see the activity in my department fragmented by key customers*"
- "*[...] know in real-time whether there is a sharp drop in activity or negative sentiment regarding a certain client*"
- "*[…] identify overlapping activities with the same customer […] We are a big company and often times we act in parallel, duplicating efforts and generating complexity without being aware*"
- "*[…] get actionable recommendations based on anomaly detection […] For example, create a sales opportunity around a specific trending topic that is relevant for a customer; or arrange a meeting with a client with whom activity has decreased*"

Human Resource Scenario. As more employees share their thoughts and opinions through social media, Human-Resource (HR) professionals can investigate the "employee voice" reflected in the activity stream to better understand the sentiment around certain policies, processes, or programs, and the main topics employees are concerned about. Activity stream analysis can replace traditional internal surveys and interviews, provide more authentic feedback, and allow taking action in near real-time

to address emerging issues. We interviewed two HR practitioners who shed more light on how a potential HR dashboard can be built on top of the activity stream:

- "*A key desired capability is segmentation of the stream based on person attributes originating from HR systems, such as country, business unit, job role, seniority, band, salary, or performance grade*"
- *"[…] slice and dice the stream to view the sentiment and top topics in different segments of the organization, for example the sentiment of people with less than five years seniority in the Services division*"
- "*More important than getting the current status across the organization is getting the change map: where are the most drastic changes taking place right now?"*
- "*Mix the internal stream data with what employees say on external social media, which may reflect an even more authentic opinion […] Allow comparing between the two*"
- "*Analyze the flow of information and collaboration between departments and countries […] who is talking to whom and where do gaps exist?*"
- "*Alerts […] As opposed to just searching, I would like the system to push me notifications when a change happens, such as a new trend in South America or a drop in sentiment around BCG process for managers"*

IT Scenario. Discussions about organization's internal tools are quite popular on enterprise social media. Internal IT teams can use the activity stream to track new issues and enhancement requests, and get a view of the strengths and weaknesses of the tools they deploy. As mentioned above, in our general stream analysis, we discovered that many of the activities with negative sentiment referred to issues, bugs, or error notifications from internal tools. Our interviews with two members of the IT department revealed the following needs:

- "*[…] get a sense of the 'popularity' of tools by the number of activities that refer to them […] view popularity over time using custom topic visualization*"
- "*[…] spotting changes in real-time is critical, to discover new bugs or issues, for example by identifying a new trending topic or a change in sentiment regarding a tool or a key functionality it provides*"
- "*Comparative view like the 'top topics' is especially important, but should also reflect sentiment […] for example a new tool we have deployed versus the legacy one; or an organic tool versus a third party; or several versions of a tool*"
- "*Compare the use by geography in a similar visualization to the 'top topics' view*"
- "*Get notifications and recommendations for action […] for example, handle the error in our CRM tool, with reference to activities that discuss this error*"
- "*Identify solutions for problems or useful tips to reuse*"

Overall, our eight follow-up interviews revealed many potential merits for the use of the activity stream from an organizational perspective. There were many common-

alities among the requirements of the three scenarios, such as segmentation of the data by person attributes (HR attributes, sales-related attributes such as industry or client, general attributes like country or business unit); tracking changes over time with regards to volume of activity or level of sentiment; getting alerts and actionable recommendations in a "push"; or using comparative views (e.g., across peers in sales management or tools in IT). Overall, the high intensity of the stream and its growing usage hold a lot of promise for changing the way LOBs track organizational trends and opinions. Gaining these types of insights from the stream can eliminate the need for solicited feedback via surveys, feedback forms, or other more traditional means.

5 Conclusions and Future Work

Our analysis provides an overview of the way *Streamz* is used within the enterprise. While *Streamz* is still in its early-adoption phase, dozens of employees already use it for a variety of purposes, such as to get a sense of what people say or think, to track the activities of individuals or groups, to search for recent news regarding projects or topics and understand who is involved in them, to search for past discussions, and to generally stay up-to-date.

We observed that stream search is different from traditional enterprise content search in many ways and poses a new complementary paradigm for searching in the enterprise. While content search is used mainly for navigation to pages of internal tools or projects or for finding the main resources related to a topic, stream search is used for other goals, such as getting updated with the recent news or developments, tracking the latest conversations, or understanding the recent buzz. We learned that users typically issue queries as a means to filter the stream rather than to search—as opposed to seeking something specific and known in advance, users often issue queries to render an interesting channel for news discovery.

While *Streamz* originally provided a view for individual employees to help them bisect and consume the activity stream in a personalized way, it was extended in this work with a few initial capabilities that allow better exploration of the stream from an organizational viewpoint. Our follow-up interviews with sales managers, HR, and IT professionals, reveal that this new form of public interaction holds a lot of potential for LOBs who are interested in better understanding what employees do or think with regards to a specific customer, an organizational process, or an internal tool. The UI provided in this work was used to demonstrate an initial set of stream analysis capabilities with an emphasis on the organizational (sociocentric) perspective. More advanced capabilities should allow bisecting the stream by criteria such as industry or person attributes (seniority, salary, etc.), provide more advanced means for tracking trends and sentiment over time, and support active alerts and recommendations for action when anomaly or change occur.

The main limitation of this study is the fact it was conducted in a single enterprise, using a system in its early adoption phase. The results may thus be affected by the characteristics of the specific enterprise and the early adoption. We note, however, that we focused on a rich set of common social media applications that are likely to be

relevant in many organizations. We hope future work can further extend and generalize the findings of our research, by inspecting activity streams in other companies and at later stages of adoption.

This paper provides a first glimpse into the use of an activity stream in the enterprise. Currently, the activity stream syndicates activities across social media; however, its potential use is even broader, as it can grow to encompass activities from other organizational systems, such as CRM, software development, or business process management. One of our interviewees even suggested "*I would love to see a Streamz-like unified view of both my social media and email.*" Indeed, email messages can be viewed as another type of a stream activity. With younger individuals joining the workforce, accustomed to using tweets and news feeds as their main means for interaction, an integrated activity stream, equipped with search and analytics capabilities, could potentially replace the traditional mailbox and become the future organizational dashboard for workplace collaboration.

References

1. Ben-Yitzhak, O., Golbandi, N., Har'El, N., Lempel, R., Neumann, A., Ofek-Koifman, S., Sheinwald, D., Shekita, E., Sznajder, B., & Yogev, S. 2008. Beyond basic faceted search. *Proc. WSDM '08*, 33-44.
2. Bernstein, M.S., Suh, B., Hong, L., Chen, J., Kairam, S., & Chi, E.H. 2010. Eddi: interactive topic-based browsing of social status streams. *Proc. UIST '10*, 303-312.
3. Boyd, D. 2008. Facebook's privacy trainwreck: exposure, invasion and social convergence. *Convergence* 14, 1 (Feb. 2008), 13-20.
4. Buffa, M. 2006. Intranet wikis. *IntraWebs Workshop*, *WWW '06*.
5. Busch, M., Gade, K., Larson, B., Lok, P., Luckenbill, S., & Lin, J. Earlybird: real-time search at Twitter. *Proc. ICDE '12*, 1360-1369.
6. Byron, L. & Wattenberg, M. 2008. Stacked Graphs - Geometry & Aesthetics. *IEEE Trans. Vis. and Comp. Graphics* 14, 6 (Nov. 2008), 1245-1252.
7. Celi, F., Di Lascio, F.M.L., Magnani, M., Pacelli, B., & Rossi, L. 2010. Social network data and practices: the case of Friendfeed. *Advances in Social computing*, 346-353.
8. Daly, E.M., Muller, M., Millen, D.R., & Gou, L. 2011. Social lens: personalization around user defined collections for filtering enterprise message streams. *Proc. ICWSM '11*.
9. Dork, M., Gruen D., Williamson, C., & Carpendale, S. 2010. A Visual Backchannel for Large-Scale Events. *IEEE Trans. Vis. and Comp. Graphics* 16, 6 (Nov. 2010), 1129-1138.
10. Efimova, L. & Grudin, J. 2007. Crossing boundaries: a case study of employee blogging. *Proc. HICSS '07*.
11. Ehrlich, K. & Shami, N.S. 2010. Microblogging inside and outside the workplace. *Proc. ICWSM '10*.
12. Esparza, S.G., O'Mahony, M.P., & Smyth, B. 2010. On the real-time web as a source of recommendation knowledge. *Proc. RecSys '10*, 305-308.
13. Farrell, S., Lau, T., Nusser, S., Wilcox, E., & Muller, M. 2007. Socially augmenting employee profiles with people-tagging. *Proc. UIST '07*, USA, 91-100.
14. Freeman, L.C. The Development of Social Network Analysis: A study in the Sociology of Science. *Empirical Press*, 2004.
15. Freyne, J., Berkovsky, S., Daly, E.M., & Geyer, W. 2010. Social networking feeds: recommending items of interest. *Proc. RecSys '10*, 277-280.

16. Gupta, T., Garg, S., Carlsson, N., Mahanti, A., & Arlitt, M. 2009. Characterization of Friendfeed: a web based social aggregation service. *Proc. ICWSM '09*.
17. Guy, I., Ronen, I., & Raviv, A. 2011. Personalized activity streams: sifting through the "river of news". *Proc. RecSys '11*, 181-188.
18. Guy, I., Steier, T., Barnea, M., Ronen, I., & Daniel, T. 2012. Swimming against the Streamz: search and analytics over the enterprise activity stream. *Proc. CIKM '12*, 1587-1591.
19. Guy I., Ur, S., Ronen I., Weber, S., & Oral, T. 2012. Best faces forward: a large-scale study of people search in the enterprise. *Proc. CHI '12*, 1775-1784.
20. Hoadley, C.M., Xu, H., Lee, J.J, & Rosson, M.B. 2010. Privacy as information access and illusory control: the case of the Facebook news feed privacy outcry. *E-Commerce Research and Applications* 9, 1 (Jan. 2010), 50-60.
21. Hong, L., Convertino, G., Suh, B., Chi, E.H., & Kairam, S. 2010. FeedWinnower: layering structures over collections of information streams. *Proc. CHI '10*, 947-950.
22. Huberman, B., Romero, D., & Wu, F. 2009. Social networks that matter: Twitter under the microscope. *First Mon.* 14, 1 (Jan 2009).
23. IBM Connections – Social Software for Business: http://www.ibm.com/software/lotus/products/connections
24. Kwak, H., Lee, C., Park, H., & Moon, S. 2010. What is twitter, a social network or a news media?. *Proc. WWW '10*, 591-600.
25. Millen, D.R., Feinberg, J., & Kerr, B. 2006. Dogear: social bookmarking in the enterprise. *Proc. CHI '06*, 111-120.
26. Moran, T.P., Cozzi, A., & Farrell, S.P. 2005. Unified activity management: supporting people in e-business. *Commun. ACM* 48, 12 (Dec. 2005), 67-70.
27. Naveed, N., Gottron T., Kunegis, J., & Alhadi, A.C. 2011. Searching microblogs: coping with sparsity and document quality. *Proc. CIKM '11*, 183-188.
28. Ounis, I., Macdonald, C., Lin, J., & Soboroff, I. 2012. Overview of the TREC 2011 microblog track. *Proc. Trec '11*.
29. Paek, T., Gamon, M., Counts, S., Chickering, D.M., & Dhesi, A. 2010. Predicting the importance of newsfeed posts and social network friends. *Proc. AAAI '10*.
30. Ronen, I., Shahar, E., Ur, S., Uziel, E., Yogev, S., Zwerdling, N., Carmel, D., Guy, I., Har'el, N., & Ofek-Koifman, S. 2009. Social networks and discovery in the enterprise (SaND). *Proc. SIGIR '09*, 836.
31. Shami, N.S., Muller M.J., & Millen, D.R. 2011. Browse and discover: social file sharing in the enterprise. *Proc. CSCW '11*, 295-304.
32. Sun, E., Rosenn, I., Marlow, C., & Lento, T. 2009. Gesundheit! Modeling contagion through Facebook News Feed. *Proc. ICWSM '09*.
33. Teevan, J., Ramage, D., & Morris, M.R. 2011. #TwitterSearch: a comparison of microblog search and web search. *Proc. WSDM '11*, 35-44.
34. Zhang, J., Qu, Y., Cody, J., & Wu, Y. 2010. A case study of micro-blogging in the enterprise: use, value, and related issues. *Proc. CHI '10*, 123-132.
35. Zhao, D. & Rosson, M.B. 2009. How and why people Twitter: the role that micro-blogging plays in informal communication at work. *Proc. GROUP '09*, 243-252.