



Improving Government Services with Social Media Feedback

Stephen Wan and Cécile Paris

The CSIRO Computational Informatics Division

Sydney, Australia

firstname.lastname@csiro.au

ABSTRACT

Social media is an invaluable source of feedback not just about consumer products and services but also about the effectiveness of government services. Our aim is to help analysts identify how government services can be improved based on citizen-contributed feedback found in publicly available social media. We present ongoing research for a social media monitoring interactive prototype with federated search and text analysis functionality. The prototype, developed to fit the workflow of social media monitors in the government sector, collects, analyses, and provides overviews of social media content. It facilitates relevance judgements on specific social media posts to decide whether or not to engage online. Our user log analysis validates the original design requirements and indicates ongoing utility to our federated search approach.

Author Keywords

Social Media Monitoring, Federated Search, Natural Language Processing, eGovernment

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI)

INTRODUCTION

Social media is fast becoming an important means for community engagement and public communication. It empowers the ordinary web user to become not just a consumer but also a producer of information, potentially encouraging further contributions from other users. For example, microblogs (e.g., Twitter¹), blogs and discussion forums are often used to share advice, opinion, and commentary. As a result, social media is an invaluable source of feedback not just about consumer products and services but also about the effectiveness of government services.

In the government sector, it has been recognised that social media tools can facilitate the delivery of government services [6]. We are working with various government departments to

¹www.twitter.com

find feedback in publically available social media that suggests how existing services can be improved, and to identify opportunities to improve information dissemination about the services to which the public are entitled. Research-wise, we aim to support social media monitoring activities through the design and development of tools that couple text analysis technology with flexible, easy-to-use interfaces. Ultimately our hope is that the tools can be deployed seamlessly within existing social media monitoring workflows.

Our user study of social media monitoring for government services [8] defines some of the core tasks involved. In short, the monitoring activity requires sifting through social media search results (for example, a comment, microblog or blog), judging it for relevance, identifying the underlying issues, and potentially acting upon the content to engage with an online community. These issues may reveal how existing services are performing, highlighting potential areas of improvement for both the services themselves and the departments' communications about them. Through social media monitoring, media and communication teams aim to identify both the reach of their own communication, and the forums/media to target so that information dissemination about service entitlements reaches the appropriate audiences. Finally, they also look for opportunities to engage directly with online communities to make sure that information about government services being shared by citizens is correct.

The study also found that dealing with multiple applications, each corresponding to a separate social media platform, is time-consuming and can introduce problems in consistent engagement with online communities. Ideally, within a single application, social media monitors would perform the core tasks above and gain insights as to how content from one social media platform interacts with another. As an example of how such associations arise, consider the case of a government press release which triggers subsequent news articles. These articles can in turn evoke discussions and comments on news sites, opinion blogs and social networking sites, to name but a few. Each of these items may require the attention of a social media monitor, but judging how to respond to these posts may be made more difficult if these are presented in isolation, as is currently done in most search engine results.

We introduce a social media monitoring prototype, *Vizie*, that presents 1) a single federated search interface for different social media platforms that allows users to understand why information is collected and where the content came from, and 2) a unified analysis of the aggregated data to support data exploration. The analyses, using methods from Natural Lan-

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).

IUI'14, February 24–27, 2014, Haifa, Israel.

ACM 978-1-4503-2184-6/14/02.

<http://dx.doi.org/10.1145/2557500.2557513>

guage Processing (NLP), serve to help media monitors find commonly discussed issues across different types of social media and assist in identifying opportunities for engagement with online communities.

In this paper, we present Vizie and describe the technical and real world challenges involved in producing a tool that is used by 17 government departments. We analyse the aggregate user log statistics to examine the effect of introducing a federated search interface on query reformulations. Our analysis reveals the following:

1. Our prototype design, a novel combination of interactive interfaces, NLP and information retrieval capabilities, has demonstrated utility, as indicated by stable usage patterns spanning over 18 months.
2. The usage patterns validate the design of the prototype in facilitating overview and engagement activities.
3. Stable usage is associated with well-defined monitoring goals, as reflected by maintenance and refinement of queries queries.

RELATED WORK

To our knowledge, federated search approaches to social media monitoring, where a uniform query interface and a corresponding analysis is presented to the user, have not been explored in-depth within the research community. Whilst there are a number of tools available online to perform social media monitoring (for example, GoogleAlerts² and Socialmention³), these generally focus on a retrieval task that does not necessarily include data exploration. For example, although SocialMention retrieves content from various social media platforms, it does not easily allow content from one platform to be compared to another, nor is there an analysis to guide the user on salient issues and topics within the search results. Existing tools for general search and analysis, including SocialMention, are generally limited to showing trending words. This kind of analysis is typically presented as a ranked list. Our interface provides the user with a sense of the volume associated with major clusters of data, and how this changes across time.

There are a variety of tools dedicated to a specific social media platforms. Tools like TwitterMap⁴ can present twitter content with geotag data on maps to highlight what is being discussed in a particular region. Software such as Tweetdeck⁵ can be used to monitor microblog conversations on a topic. Analytics about the rise and fall of particular terms can be presented as graphs using tools like Hootsuite⁶, which offers a microblog and social network dashboard for Facebook⁷ and Twitter. Whilst Hootsuite and Tweetdeck are popular, this forces social media monitors interested in a diverse range of

social media platforms to use a number of tools and websites in an *ad hoc* manner to obtain an overall impression of the data, resulting in multiple views of collected social media.

Within the academic community, there are a number of systems and tools that help identify what in Twitter might be of importance. For example, TweetStand classifies Tweets into news or junk [15], while the system by Sriram *et al.* [16] classifies Tweets into five generic categories: news, events, opinions, deals, and private messages. These works differ from ours in that they are not geared towards facilitating specific social media tasks, and they handle only Twitter content. In contrast, our work considers many social media channels, and the data analysis is intended to help the user with specific monitoring tasks.

Some related research work is aimed at providing an overview of content by applying Bayesian topic modelling approaches to social media content, for example, on Twitter [10], on discussion forums [13], and on bookmarking sites [11]. However, these have not yet been deployed widely for general use in an end-user application.

There are some notable systems that present contentful overview analyses of online text. The Newssift tool⁸, by the Financial times, performed topic analysis on news articles matching a query. Unfortunately, this site was closed down in February 2010. Bernstein *et al.* [3] describe a social media tool that highlights related discussions across Twitter. By treating Tweets as queries, a set of related documents can be retrieved from search engines to overcome data sparseness issues. Although the work of [3] shares a similarity in application and in the use of topic-based clustering, our research differs in examining multiple social media types.

Within the context of government-based social media monitoring, there is work that exploits social media for emergency situation management, in particular to identify Tweets that might be useful to detect an emergency or help support a response to an emergency (for example, see [14, 18, 1, 20]). In our work, we are not looking specifically at emergency crisis, and, again, we are interested in monitoring a wide range of social media channels.

Finally, the TweetGathering prototype [21], which focuses solely on Twitter data, is designed to help journalists gather news, using an interactive interface and techniques from NLP. The prototype was shown to journalists for evaluation, but, to our knowledge, it was not used in their jobs. In contrast, our prototype monitors all social media information, not just news. Furthermore, our evaluation is based on user log data.

GOVERNMENT SERVICES AND SOCIAL MEDIA

Although government social media monitors often use a number of the tools described above, their monitoring task involves more than simple brand monitoring or identifying the relevant buzzwords to engage with an intended audience. Often there is a desire to improve existing government services, by identifying service feedback discussed in online discussions and potentially engage with the community.

²www.google.com/alerts. At the time of initial design, Google Alerts search results were available as an RSS feed that could be used for data analysis. As of July 2013, this is no longer the case.

³www.socialmention.com

⁴www.twittermap.com

⁵www.tweetdeck.com

⁶www.hootsuite.com

⁷www.facebook.com

⁸www.newssift.com

Collecting social media content is not straightforward, however. Queries pertaining to government services may be ambiguous in meaning, resulting in the collection of non-relevant content. For example, the department in charge of social services cannot simply use a query such as “social welfare”. This would retrieve content from all across the world. Using the official name of the benefit may help narrow the focus to a country, if the name is unique. However, official names for services are also subject to misspellings and colloquial references. For example, “bereavement benefit” may simply be referred to as “payment” for the “death” of a close relative.

This ambiguity is exacerbated by the problem that the collected social media is heterogeneous in nature, given that each social media platform is marketed towards a niche in online communication. Consequently, a search query designed to retrieve specific content on one platform may perform poorly on another. For example, the query “school kids bonus”, the name of a benefit to help Australian families with the cost of schooling, retrieves Facebook statuses requesting information about the social benefit, whereas retrieved Twitter statuses includes political commentary.

Finding online discussions with useful feedback is a daunting task given the volume of social media content. In addition, the user requirements identified in [8] suggests that it is problematic for social media monitors to juggle together the existing tools. Ideally, users would have access to a single content overview, regardless of media type (blogs, microblogs, news, forums, etc) to help them identify what issues are being currently discussed online. From that overview, they should be able to drill in to access relevant content.

Furthermore, social media monitors need to identify when to engage with online communities. For example, someone may post a question about a government service on a site like Yahoo!Answers⁹. Ideally, other online community members would help to answer each other’s questions. Occasionally, however, incorrect answers may be posted. These require the social media monitor to engage with the relevant community to address the problem, an engagement which may then be retrieved in future web searches. This is particularly relevant in a government scenario, as policy changes may affect the eligibility rules of various services for citizens, rendering some information outdated [8]. With the advent of microblogs and ubiquitous computing devices, relevant content grows rapidly, making social media monitoring increasingly difficult. Without any dedicated social media tools for analysis and aggregation of data, dealing with this burgeoning data is difficult.

Our prototype is designed to address these problems. However, the design of tools to support such tasks is not immediately straightforward. The real world constraints are challenging and range from 1) designing an affordable but scalable architecture to handle hundreds of users spread across a number of government departments; 2) compliance with the terms of service of social media platforms; and 3) choosing

analysis methods that are appropriate for the data collected, in terms of scalability of volume and time to process.

The first constraint influences how we design the system to ensure that the load placed by one department does not negatively impact the user experience of another. It also has a bearing on how data is stored to keep each department’s collected data separate. The second constraint affects how much data we collect from a specific platform. Finally, the third constraint forces us to use text analysis methods that are domain independent given that the services and activities of the participating departments span many areas of public life. In addition, the system must be able to analyse large quantities of textual data on limited computing resources.

AN OVERVIEW OF VIZIE’S KEY FEATURES

The prototype Vizie, designed to be used by a social media monitoring team, provides a single point-of-entry to both data collection and data analysis interfaces. These are realised as a novel combination of interactive and collaborative interfaces which rely on NLP and information retrieval techniques, including language identification¹⁰, clustering algorithms (based on keywords or Latent Dirichlet Allocation (LDA) [4]), discussion summarisation [19], keyword and keyphrase detection (based on [7]), and extractive summarisation [9].

The various pages of the User Interface (UI) are implemented as a combination of several web components including navigation menus, graphs, search interfaces, etc. Each page also has options for the user to interact with in order to examine the data. Viewing these pages and interacting with the prototype are user events which are logged for further study. We return to this analysis below. We now describe how this functionality is exposed in the prototype’s interfaces, particularly with respect to the requirements outlined above.

A Social Media Federated-Search Interface

To make the problem of collecting data of different social media types easier, we created a single federated search interface where the same query is issued to multiple social media platforms. In this way, we retrieve only publically available content. We use the Application Programming Interfaces (APIs) of the various platforms like Twitter, Facebook, YouTube, LinkedIn, and Instagram. For image and video data, our prototype only processes the textual content found in descriptions and comments. In addition, we utilise the API of the SocialMention social media search engine to collect data. This is particularly helpful in collecting data in the recent past if there are technical difficulties, either within our system or with a specific API. Finally, data can be collected by specifying RSS or ATOM feeds. These feeds are checked hourly for new content.

In addition to these data collection methods, the prototype collects data across from different social media platforms by downloading all linked content. The prototype reconstructs discussion threads based on the “in-reply-to” associations found in the metadata returned by each platform for retrieved

⁹<http://answers.yahoo.com>

¹⁰<https://code.google.com/p/language-detection>

The screenshot shows a web application interface for query registration. On the left, there are navigation links: 'Welcome', 'Home', 'Logout', 'Overview', 'View Posts', 'Engagements', 'Report', 'Config', and 'Help'. Below these are sections for 'User Accounts' (Add User, Manage Users, Edit Your Profile), 'Data Sources' (Manage monitoring activities, Initial, Set-up, Create a Monitoring Activity, Add Query, Add Feed), and 'Experimental Functionality' (Import Queries, Export Queries, Enable your account with: Twitter, Google, LinkedIn, Facebook, Instagram, Yammer). The main 'Add Query' section has a text input for 'Query' with the value 'social welfare' and a 'Preview' button. Below this is a note: 'Data collected will contain ALL of the query words you enter. That is, Vizie assumes an "AND" between query words. Additional search operators (e.g., "OR", "NEAR") are currently NOT supported.' There are fields for 'Display Name' (social welfare) and 'Ignore results with these words'. A 'Note on queries and collected data' section lists two bullet points: 'To specify a phrase, you can add double quotation marks: for example, "my phrase".' and 'We limit collection to English data. In the future, you will be able to turn language filters on or off.' An 'Optional configuration' section states 'The following items are optional. To finalise your changes, use the buttons at the bottom of this form.' Below this is a 'Select web search engines to collect from:' section with checkboxes for 'Collect live Twitter data', 'Collect recent Twitter data', 'Collect data from Facebook', 'Yammer', 'Youtube', 'GooglePlus', 'Instagram', and 'Collect data from the Social Mention search engine'. There are also checkboxes for 'Add results feed' and 'Copy-paste your RSS/ATOM feed URL here'. At the bottom is a button 'add another results feed'. On the right, the 'Preview Results' section shows a 'Feed Title: Twitter Preview' and a 'Published on: Wed, 09 Oct 2013 02:51:07 GMT'. It lists several tweets related to 'social welfare' and 'Social Security', including links to tweets and a 'Search for Other Data' button.

Figure 1. The query registration interface. The form allows the specification of complex queries which are then issued to a number of social media platforms. Previews for collected results are presented on the right to allow the user to refine the query through the use of exclusion terms and additional contextual query terms.

posts. For example, if a Tweet mentioning the query term is part of a discussion, the system will attempt to download the conversational context for that post.

Similarly, if a data has a shortened URL, the link is resolved and the linked web content is captured since it provides the context for some online commentary. Given the transient nature of the web, the system stores snapshots of the webpage, if possible, to archive the content, a legal requirement for many government departments. The data archived by the prototype is not available for public dissemination, and the user takes responsibility for the use of any data collected by the prototype, which is available to them only.

As system designers, one problem we find is that users do not always spend a great deal of time in manually refining queries. Often, our users are extremely time-poor. Some users, such as those involved in government policy, are not in fact part of a media and communications group, and they perform their social media monitoring activities as an extra task in addition to their normal daily activity. For example, the user may *a priori* have a query in mind, such as the official name of a service. In a naïve user interface, this might then be entered into the system without verifying that the query is appropriate (e.g., as in our previous “bereavement” example). In some cases, the query term may simply not be one that is adopted by the online public. The user must then wait for a period of time 1) for new content to be created by the online public, and 2) for the system to collect data. The user can then be frustrated if, after this period, no relevant data is collected.

To alleviate this problem, our prototype provides previews of data to indicate what content is likely to be collected. In Figure 1, we present our interface for registering a query within the prototype. To the right, a series of panels presents previews for each of the platforms. The previews give the user an idea of what information from each platform will be col-

lected. If this does not match their expectations, they can then reformulate the queries as appropriate. This is necessary as a query issued to one platform (like Twitter) may not suit another (for example, Facebook) and may result in very different content being collected. The system is designed so that the user is forced to vet the preview results before data collection and analysis. This is important as the more non-relevant data that is collected, the harder it is for subsequent automated data analyses to provide useful results.

The problem of finding a query that will work in federated search is further complicated by the fact that each platform’s API has a different query language. For example, the search-based streaming API for Twitter does not support phrasal queries. In addition, boolean operators such as AND and OR are not available in a uniform manner across each platform, making it difficult for users to issue complex queries with multiple terms.

As shown in Figure 1, the prototype’s interface uses a form to collect the parameters of the complex query. The top text field can be used to provide one or more words, any of which may be scoped within phrasal markers (double quotation marks). In the interface, we indicate that a conjunction is used for each query term or phrase. At this stage, we do not allow disjunction, which can, in any case, be achieved by registering variations within the prototype as a separate query. We adopt this position because a preview interface would otherwise be overly complicated; the user would not know which part of the disjunction is contributing to the results. Finally, the form also provides fields for the NOT operator which can be used to exclude content to refine the query. To the right, the previews of data to be collected from each social media platform are presented.

The previews and the data collected are passed through a series of filters designed to enforce the query constraints cap-

tured through the form in Figure 1. These filters are necessary to uniformly impose query constraints on the data returned by the various APIs, given that each API provides a different query language. In addition, we presently use an existing language filter to restrict data collection to English content¹¹.

Overviews for Exploring Collected Data

Once the data is collected, the prototype presents a single overview of the aggregate data, based on analyses providing the major data clusters. This can be used to sift and explore the data, either to find feedback or opportunities for engagement, or else to further refine queries by understanding what data is eventually collected. The overview is available in both a textual and visual interactive interfaces.

The analysis includes a categorisation of each social media post to specify its media type. For example, YouTube and Vimeo content would be labelled as video. The microblog category includes predominantly Twitter data but also contains some Identica posts. Given that Facebook is the only social network that we currently support, it is given its own label. The other categories currently in use include blogs, pictures, comments, government, press releases, social network profiles, questions, bookmarks, and forums. We also use a miscellaneous category for content that cannot be identified.

The overviews are created by grouping posts according to keywords, where a salient keyword is chosen for each post. To select a single keyword to represent a post, candidate keywords are first generated using various cues such as word length, capitalisation, and membership within an indicative phrase. Phrases are provided by a rule-based keyphrase identifier that delimits the post into phrases based on stopwords. To select the single keyword to assign the post to a group, the system first checks if any of the keywords has a match in a list of specific terms that are of interest to the user. If not, the system chooses the one that is most popular (as measured by frequency) within a one week window. In this way, groupings can capture the streaming nature of social media data.

This data grouping approach, a surrogate for a clustering algorithm, was chosen as it is 1) fast and 2) easily deployed on the low powered virtual machines that we dedicate per department. More in-depth but computationally expensive analyses, for example a clustering provided by LDA [4], are available on-demand.

The overviews are then presented for a series of days, typically the preceding week. Figure 2 shows the interactive textual version of this overview, which can be displayed on devices without Java or Flash, including government computers and handheld smart devices. The overview is in a mode designed to maximise screen to show the top eight to ten groupings of data, indicated with a keyword, and to show the proportions between groupings for a single day and *within* a specific social media type.

In this figure, data about countryside fires was collected using the queries such as “#ausheatwave”, “#bushfire”, “#heatwave”, “#nswfires”, “#safires”, “#sydneyheatwave”, “#tas-



Figure 2. A textual overview of retrieved social media content. The overview is designed to make full use of the screen to highlight the cluster labels. The large panel to the right shows the cluster currently selected. The small pop-up window presents a preview of an individual post. The user can drill down to a specific post from that panel.

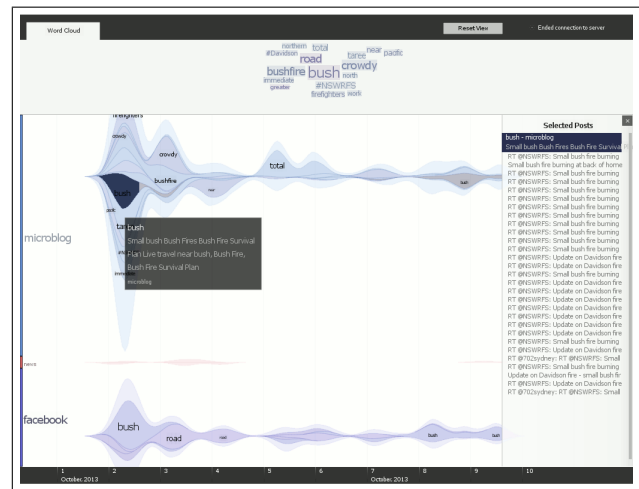


Figure 3. A visual overview of social media about rural fire services.

fires”, “#vicfires”, and “@NSWRFS”. A different colour is assigned to each mediatype — for example, green for blogs and blue for microblogs. In this figure, the users has selected one of the keywords “grass”. The interface highlights all mentions of that string throughout the week to allow the user to gauge its prominence across the week. To the right of the overview, a listing shows the posts included in one of the groupings for the keyword “grass”, each represented by a snippet of text. Hovering over the post will show more of the text and clicking on it will take the user to the archived content for that post.

The visual overview, presented in Figure 3, allows the cluster sizes to be compared across days. It is inspired by the Stacked Graphs visualisation [5] but differs from it in terms of the smoothing methods used to produce curves, and in our use of the multiple stacked graphs to show different channels of data.

Although the prototype generates both versions of the overview from the same data, the visual overview is able to

¹¹<https://code.google.com/p/language-detection>



Figure 4. A summary of the news article that is linked to from the shortened URL is presented to the user as context to the Twitter discussion.

show some of the smaller clusters, as our interactive visualisation affords a zoom-in-zoom-out interaction. Figure 3 shows the visualisation zoomed to a level to show discussions about rural fires on Twitter and Facebook. This allows users to inspect some of the smaller clusters, which can be useful when the social media monitor is trying to find problems and issues as they are beginning to trend. As in the textual overview, the user can drill down to the actual post content if they deem it relevant and worth reading further.

In both the textual and visual overviews, multi-document summaries are presented that include the indicative phrases for each keyword. The pop-up in Figure 3 presents a summary for the cluster under the mouse pointer. In this case, the summary provides more context for the cluster label “bush” containing key phrases like “Bush fire” and “Bush Fire Prevention Plan”. This might indicate to the user that a prevention plan for rural fires is the key concept for the cluster.

Supporting Online Engagements and Reporting

The user requirements [8] discussed above indicated that engagement with the online community was important. To support this activity, the prototype assists with relevance judgements for each post, in which the user must decide if the content is such that an engagement is required (for example, to provide a correction to user-contributed information). The interface also supports record keeping to archive the engagements that have been made, since the archival of correspondence is usually required by government departments.

To support relevance judgements, where possible, we provide the context that triggers online discussions. For example, shortened URLs might indicate the web content that led to a Twitter discussion. In Figure 4, we present a Tweet in the top half of the figure with such a link embedded within it. To help the social media monitor decide how to react to the post, we provide as much context to the Tweet as possible, including a summary of the news article that is linked to, as this might be quite lengthy to read. For our summarisation of the news article, we employ a generic sentence extraction summariser which utilises centroid-based methods (for example, [9]).

Although not shown in this figure, if the post is part of a discussion, the prototype presents the number of discussion turns and, if possible, the extracted date of posting. These help the user to appraise the discussions potential viewership and thus impact. Using thread summarisation methods, the system also shows sentences related to the key issues, as selected by vector space methods that rely on contentful word repetition across various discussion turns (for example, [19]). We are currently researching question-answer summarisation methods for discussion summaries as in [12].

The prototype also facilitates other tasks such as report generation and search within the collected data. To make it easier to manage the collected data, we provide a simple filing system to the user. The data collected is automatically tagged with a series of user categories that organise the queries used with the social media platforms. These categories, which we call Monitoring Activities, typically specify the different foci of a data collection exercise. For example, a monitoring activity for “rural fires” might be created in addition to one entitled “fire prevention public education”, each housing different queries for the social media platforms. These groupings help divide the data collected from the queries into different subsets, which can be analysed and reported on separately. These subsets of data can also be exported for the user to analyse with other tools.

USER LOG ANALYSIS

We originally developed our prototype to support one specific government department and worked with the communication team in that department to design the tool, ensuring it would fit their work environment. However, since 2012, we released the prototype to 17 government departments and have about 200 registered users.

Our users are all busy professionals with heavy workloads. As a result, we cannot perform a formal task-based evaluation of our tool, for example, to see whether it helps them cover a larger space of social media posts they would cover otherwise, or to do this faster. Instead, we consider the fact that some have continuously used the tool in their daily activities for about 18 months now as evidence of the utility and effectiveness of the tool. We present here a user log analysis which looks at the amount of data that was collected and the average number of times a user logs in the tool. We then examine general patterns in the actions performed by users, looking at the types of functionality users employ on average.

General Usage Patterns

Our user logs for the prototype span a total of 18 months and include the usage data of about 200 registered users spread across around 17 government departments. Each department is given an instance of the prototype that is independent from the others, allowing us to examine usage per department. During this period, we collected 45 million posts. Of these, over half were from Twitter.

The departments started using the prototype at different times, with the majority starting in June 2012. The rapid growth around March 2013 in the number of posts collected

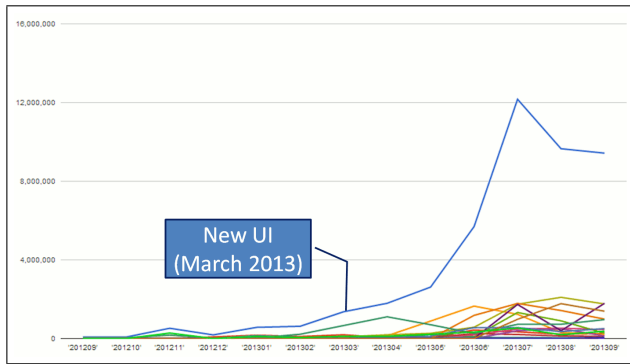


Figure 5. Number of posts collected each month in a 18 month window.

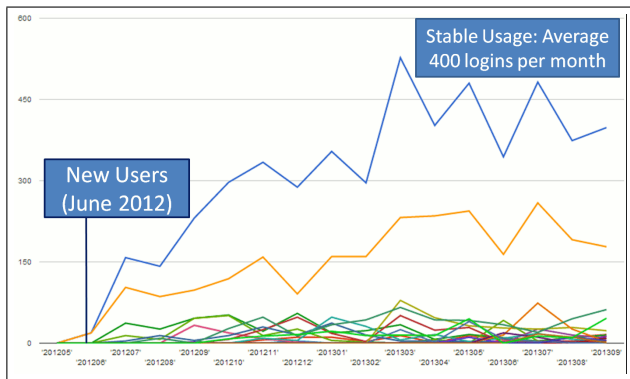


Figure 6. Number of logins over 18 months.

shown in Figure 5 is attributed to the introduction of the federated search interface. The top blue curve in Figures 5 to 7 represents the aggregate over the departments. The other lines represent a specific department.

Prior to March 2013, users had to enter each site to monitor separately. The new interface essentially automates the collection from multiple platforms. Since users typically would not exhaustively provide URLs for data collection, this resulted in more posts being collected. It also has an impact on query refinement, as we shall see. To handle this growth in data collection, we also redesigned our infrastructure to deal with the collection, storage and analysis of this growth of data.

The main indicator of stable usage by a department is the number of logins to the system, as shown in Figure 6. Most of the government departments started using the prototype in mid-2012, explaining the rapid growth then. The aggregate logins reach a plateau of about 400 logins a month, indicating that in general we have a stable user base. This is despite some login patterns indicating that a handful of departments have tried the prototype once but decided not to use it. The stable usage suggested by the graph indicates that the users find value in the prototype to the extent that they logged back in regularly throughout the 18 months to use its functionality.

Similarly, the number of prototype UI components (recall these are combined to form pages in the web application) viewed by users reaches a stable level of around 4000 components per month, as shown in Figure 7. On average, each user

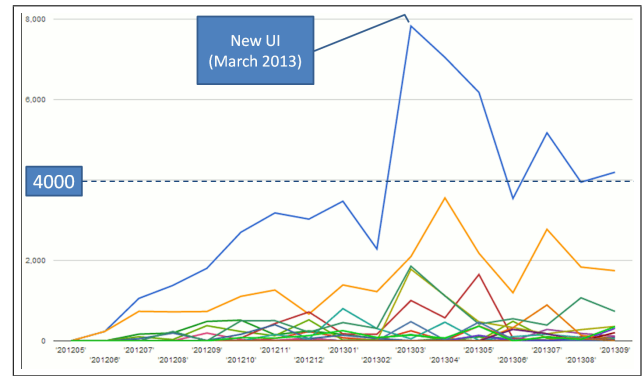


Figure 7. Number of prototype UI components viewed.

session with the prototype involves viewing 10 components, which corresponds to viewing a number of pages. We interpret this to indicate that the user is doing more than logging in; rather they are examining the data collected.

The system's federated search interface, which allows users to preview data from different platforms, was released gradually (with one API added after another) from March to May. We note that shortly after the time that the federated search interface was introduced there is an initial drop in components viewed around June 2013. However, this stabilises around August 2013.

We interpret this graph as showing how the users adapt to the new behaviour of the system, which, by automating the issuing of queries to the various platforms via the federated search interface, dramatically increases the number of posts collected. This interface can save the user time in collecting content, but it can also increase the number of non-relevant posts collected. Shortly before usage stabilisation, we observed that users returned to the configuration pages of the prototype to refine their queries, as presented in Figure 8. The stable usage suggests that users were able to deal with the complex task of query formulation, using the preview facility in the federated interface to vet the data collected from each platform.

Indeed, qualitative feedback from the departments using our prototype are encouraging. Yatu Widders, Communications Advisor at the National Mental Health Commission (NMHC) has said “With the Commission having a mandate and commitment to speaking to the broader Australian community on issues of mental health and suicide prevention, Vizie has provided unique insights into what topics are being discussed and are of interest to a diverse range of groups, beyond just the health and Government sectors.”¹²

Similarly, Amanda Dennett, Senior Social Media Adviser, Department of Human Services (DHS), said “VIZIE has really helped increase our exposure and credibility in the social media space. With student forums for example, it's gone from students thinking ‘why are you here’, to them now waiting to receive the official response because they know we are on-line. It has certainly helped us address misinformation out

¹²Public remarks in a letter, 2013.

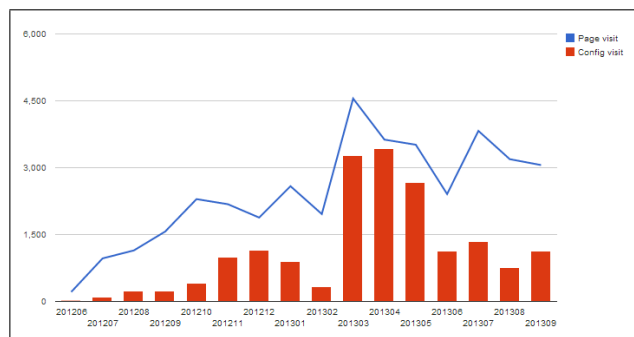


Figure 8. The period from March to May 2013 was associated with query adjustments following the gradual deployment of the federated search interface. The blue line indicates non-config prototype usage (as measured by interface page views) and the red bars indicates the number of config pages viewed.

there and proactively provide the information that people are looking for.”¹³

Interaction Patterns within the Prototype

We can investigate the primary types of interactions further using a server log inspection tool [17] that shows usage patterns for a website. In Figure 9, we present a visualisation from this tool showing the top 100 ways that users interact with the prototype. These interaction patterns are aggregations of individual user sessions.

After users log in, they are presented with a *home* screen containing a dashboard showing a summary of content for the last 24 hours. The main use of the system after viewing this *home* screen is to make relevance judgements on social media posts (see the *listPost* box) and also to record engagements with the online community (see the *listEngagement* box) which validates part of the original intended use of the prototype for managing online engagement.

Users also drill down to individual posts, search across the collected data (including using the advanced search interface within the prototype), and use the overview visualisation. Interestingly, we observe alternations between *config* and *home* screens, in which the user configures the system with queries for federated search, and then returns to the home page within the prototype. The repeated visits to the config page suggests that query refinement for federated search to multiple social media platforms is not a trivial task.

Three Case Studies of Query Maintenance

The user groups of each of the departments now covers a diverse set of interests, ranging from media and communications units to data archivists, to name but a few. This represents a range of monitoring information needs relating to the government services associated with the department, and we find that users have appropriated the prototype for their purposes. Some of the services are citizen focused, for example, the applications for and delivery of social benefits. Others are public services like documentation or knowledge management with other government and professional bodies as the key stakeholders.

¹³Excerpt taken from a CSIRO and DHS brochure, 2013.

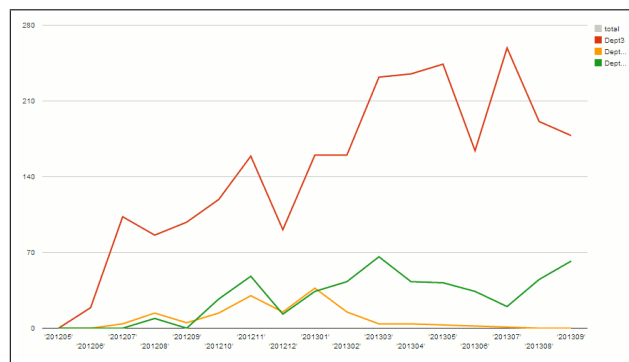


Figure 10. Login patterns for three user groups. The usage for the NMHC (yellow) is event-driven and peaks in late 2012 but then drops away in 2013. For the DHS (red) and SLNSW (green) the usage is on-going.

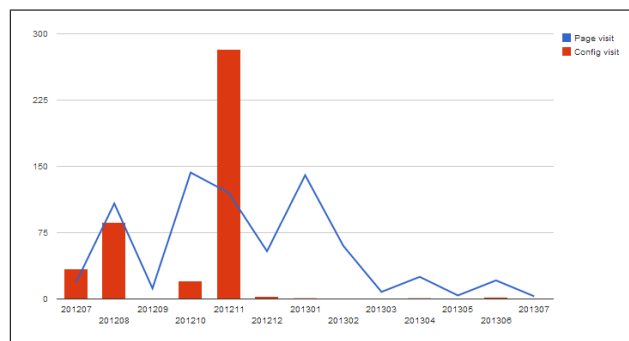


Figure 11. An example of query configuration and monitoring for events.

We can divide our user groups into categories depending on the type of social media monitoring being performed. Monitoring tasks can be event-driven or on-going. Figure 10 compares the login patterns of three organisations: the National Mental Health Commission (NMHC) is an example of event-driven monitoring whilst the other two examples, the Department of Human Services (DHS) and the State Library of New South Wales (SLNSW), perform on-going social media monitoring.

The NMHC is a government body set up to assess and document the state of mental health services in Australia. They used Vizie to monitor the social media feedback to an annual report from journalists, health professionals and the public alike. In particular, Vizie provided insights about what vocabulary the community used to discuss health issues, since professional vocabulary is not always adopted by the public. Figure 11 shows the interaction between configuration of Vizie and usage, with two main configuration events in August and November 2012 in preparation for the release of the report in late November, followed by a monitoring period into early 2013.

As an example of an on-going monitoring task, SLNSW has a mandate to collect and preserve documents that reflect lives of citizens, for use by researchers now and in the future. SLNSW staff are currently trialling Vizie to collect social media content that is significant to major events in a specific region. This complements existing web archival projects

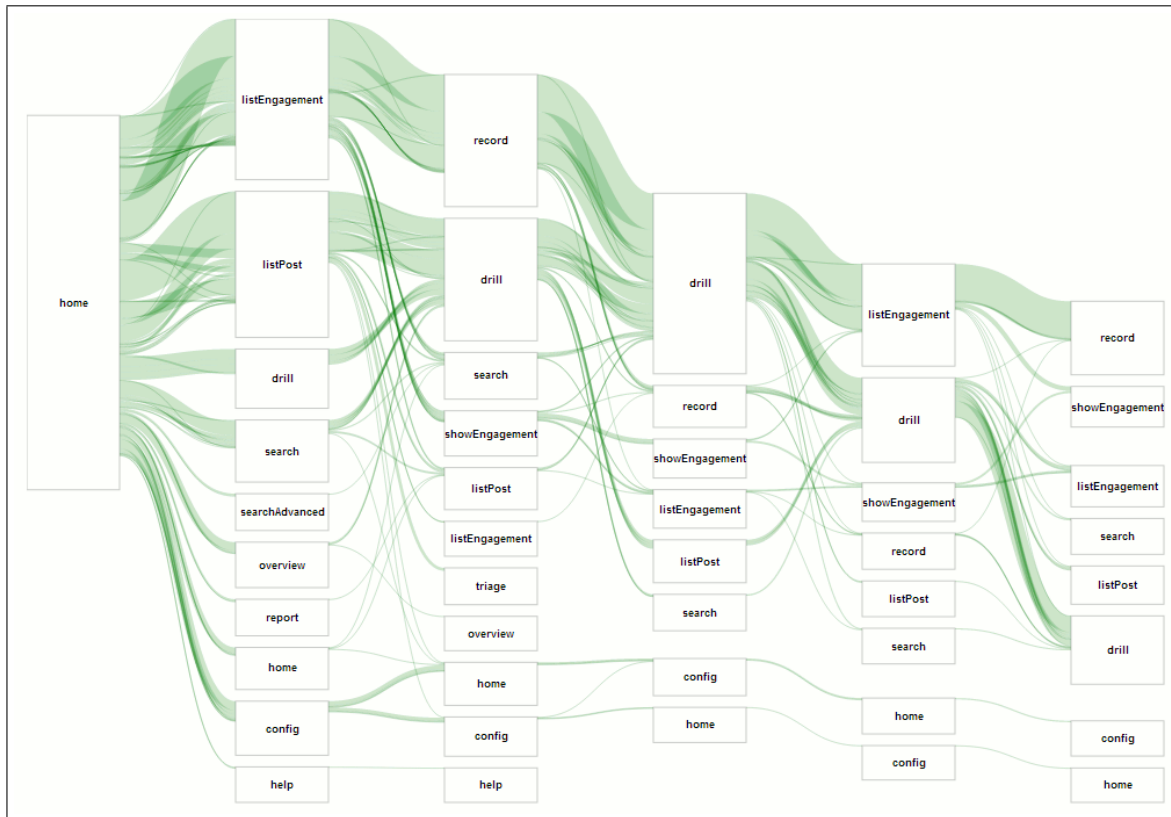


Figure 9. Interaction patterns within the prototype. Most of the components of the UI are self-explanatory. The remaining items are defined as follows: *home* is the user's landing page within the web application; *listEngagement* shows a list of archived engagements; *listPost* displays the most recent posts in reverse chronological order; and *drill* is the action of viewing the actual content of a post.

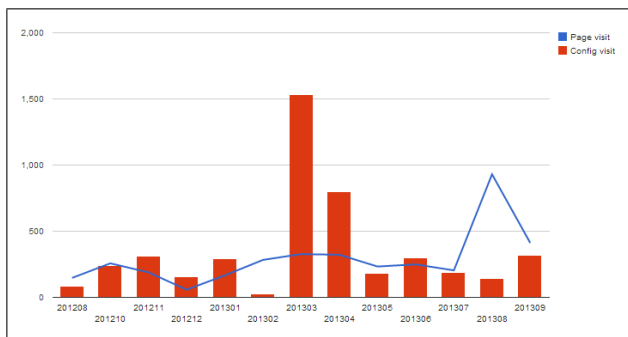


Figure 12. An example of query configuration and monitoring to archive social media descriptions of life unfolds with respect to major event in a region.

which were started before the advent of social media. In addition to the query adjustments following the introduction of the federated search interface, Figure 12 shows how this user group redefines its queries regularly for this task reflecting the rapidly adaptation to major events as they unfold. The SLNSW framework for defining queries to capture this data for prosperity's sake is introduced in [2].

The DHS has a social media monitoring group with one central coordinator and a team of media monitors. The team has an interest in providing accurate information about social welfare services. As such, they regularly examine online

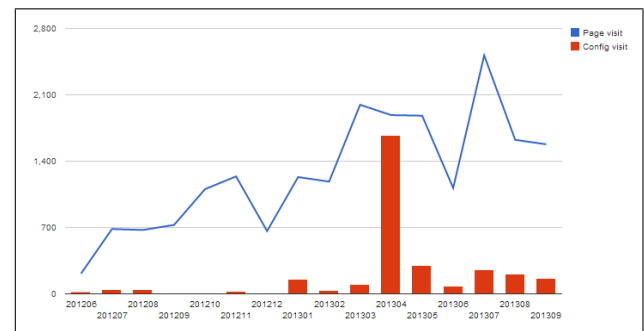


Figure 13. An example of query configuration and monitoring to capture social media feedback about a stable set of government services.

discussions to answer questions and to ensure that the information provided by other members of the public is accurate. Figure 13 shows how refinement takes place for a stable set of queries about a government service, with one major refinement taking place in April 2013 as a result of the new interface change.

CONCLUSION

In this paper, we described a social media monitoring prototype, Vizie, for improving government services based on published user requirements. The prototype uses single federated search interface for users to submit queries to different social media platforms, and provides an overview of the aggregated

results. From an analysis of users logs from nearly 200 registered users across 17 government departments, we are able to show utility. The analysis and qualitative feedback suggests that Vizie is able to support data exploration and to facilitate relevance judgments of online discussions that might lead to engagement with an online community. We investigated the interaction between complex query formulation and federated search for social media platforms. The analysis revealed encouraging changes in behaviour suggesting that the federated search interface is useful in managing how queries are issued to different platforms. Finally, we presented case studies illustrating query maintenance strategies for three social media monitoring scenarios.

REFERENCES

1. Abel, F., Hauff, C., Houben, G.-J., Stronkman, R., and Tao, K. Semantics + filtering + search = twitcident. exploring information in social web streams. In *Proc. of the 23rd ACM Conference on Hypertext and social media*, ACM (New York, NY, USA, 2012), 285–294.
2. Barwick, K., Joseph, M., Paris, C., and Wan, S. Hunters and collectors: seeking social media content for cultural heritage collections. In *VALA 2014: Streaming With Possibilities* (2014).
3. Bernstein, M. S., Suh, B., Hong, L., Chen, J., Kairam, S., and Chi, E. H. Eddi: interactive topic-based browsing of social status streams. In *Proc. of the 23rd annual ACM symposium on User Interface Software and Technology*, ACM (New York, NY, USA, 2010), 303–312.
4. Blei, D. M., Ng, A. Y., and Jordan, M. I. Latent dirichlet allocation. *J. Mach. Learn. Res.* 3 (March 2003), 993–1022.
5. Byron, L., and Wattenberg, M. Stacked Graphs – Geometry & Aesthetics. *IEEE Transactions on Visualization and Computer Graphics* 14, 6 (2008), 1245–1252.
6. Government 2.0 Taskforce. Engage: Getting on with government 2.0. World Wide Web: <http://gov2.net.au/files/2009/12/Draft-Government-2-0-Report-release.pdf>, December 22 2009. Last viewed 29th January 2010.
7. Kupiec, J., Pedersen, J., and Chen, F. A trainable document summarizer. In *Proc. of the 18th Annual International ACM SIGIR, Text Summarization* (1995), 68–73.
8. Paris, C., and Wan, S. Listening to the community: social media monitoring tasks for improving government services. In *Proc. of the 2011 CHI Extended Abstracts*, ACM (New York, NY, USA, 2011), 2095–2100.
9. Radev, D., Otterbacher, J., and H. Qi, D. T. MEAD ReDUCs: Michigan at DUC 2003. In *Document Understanding Conference 2003: Workshop on Text Summarization* (Edmonton, Canada, May 2003).
10. Ramage, D., Dumais, S., and Liebling, D. Characterizing microblogs with topic models. In *ICWSM* (2010).
11. Ramage, D., Hall, D., Nallapati, R., and Manning, C. D. Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora. In *Proc. of the 2009 Conference on EMNLP* (Singapore, August 2009), 248–256.
12. Rambow, O., Shrestha, L., Chen, J., and Laurdisen, C. Summarizing email threads. In *HLT-NAACL 2004: Short Papers*, D. M. Susan Dumais and S. Roukos, Eds. (Boston, Massachusetts, USA, May 2 - May 7 2004), 105–108.
13. Rossi, R., and Neville, J. Modeling the evolution of discussion topics and communication to improve relational classification. In *SOMA 2010: Workshop on Social Media Analytics* (Washington, DC, July 2010).
14. Sakaki, T., Okazaki, M., and Matsuo, Y. Earthquake shakes twitter users: real-time event detection by social sensors. In *Proc. of the 19th International Conference on World Wide Web, WWW '10*, ACM (New York, NY, USA, 2010), 851–860.
15. Sankaranarayanan, J., Samet, H., Teitler, B. E., Lieberman, M. D., and Sperling, J. Twitterstand: news in tweets. In *Proc. of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, GIS '09*, ACM (New York, NY, USA, 2009), 42–51.
16. Sriram, B., Fuhry, D., Demir, E., Ferhatosmanoglu, H., and Demirbas, M. Short text classification in twitter to improve information filtering. In *Proc. of the 33rd Annual International ACM SIGIR, SIGIR '10*, ACM (New York, NY, USA, 2010), 841–842.
17. Thomas, P. Explaining difficulty navigating a website using page view data. In *ADCS, A. Trotman, S. J. Cunningham, and L. Sitbon, Eds.*, ACM (2012), 31–38.
18. Verma, S., Vieweg, S., Corvey, W., Palen, L., Martin, J., Palmer, M., Schram, A., and Anderson, K. Natural language processing to the rescue? extracting "situational awareness" tweets during mass emergency. In *ICWSM* (2011).
19. Wan, S., and McKeown, K. Generating overview summaries of ongoing email thread discussions. In *COLING '04* (Morristown, NJ, USA, 2004), 549.
20. Yin, J., Lampert, A., Cameron, M., Robinson, B., and Power, R. Using social media to enhance emergency situation awareness. *IEEE Intelligent Systems* 27, 6 (2012), 52–59.
21. Zubiaga, A., Ji, H., and Knight, K. Curating and contextualizing twitter stories to assist with social newsgathering. In *Proc. of the 2013 IUI, IUI '13*, ACM (New York, NY, USA, 2013), 213–224.