Collecting Social Media for the 2015 NSW State Election

Brendan Somes

State Library of New South Wales

Stephen Wan and Cécile Paris

CSIRO Data61

Abstract

Introduction The State Library of New South Wales (the Library) has a mandate to document life in New South Wales. This has resulted in an extensive collection of materials covering all aspects of New South Wales life from the time of the coming of the Europeans to the present day. In 2015, the Library extended this activity to collect public social media discussions about significant state events such as the NSW State Election (28 March 2015).

The collection of social media content relating to elections raises new methodological and technical challenges. Firstly, one must decide upon a

systematic process for defining query terms to be used with social media search engines; these will collect public discussions from all the electorates and all the election topics. Secondly, monitoring the effectiveness of these terms and the topical relevance of the collected data is a time-consuming task that can quickly overwhelm Library staff.

Method The Library and the CSIRO collaborated on these challenges, using the social media monitoring tool Vizie to select, archive and analyse public digital material documenting the candidates, parties, interest groups and election issues. Specifically, the Library developed a new collection framework to collect digital material for elections, identifying the query terms, digital presences and sites representing the candidates, parties, interest groups, and election issues. These included Twitter accounts and hashtags, Facebook pages, websites and blogs which were utilised by the Vizie tool to capture digital posts.

The CSIRO designed new data organisation tools and analyses to help Library staff gauge the effectiveness of the collection framework and the collected data. One key new development was a data labelling tool for attributing content to each of the 93 electorates, ensuring that each electorate was represented in the data set. Analyses revealed commonalities in public discussions and provided feedback on which query terms accounted for the collected data.

Results Sourced primarily from Twitter and Facebook, over 500,000 posts were collected between December 2014 and April 2015, however additional data was also sourced from websites, blogs, and other social media platforms. Post-election analysis of the collection revealed some interesting insights: for example, election issues shared via online sources correlated moderately with the major election issues of the general population. Furthermore, the volume of posts per electorate indicated where the election battles were hardest fought.

Conclusion This paper details a new election-specific collection framework, including the process for identifying and collecting the material, as well as novel Vizie extensions implemented to provide ongoing feedback on the collection framework. This contribution has the potential to benefit other institutions wishing to capture meaningful collections of social media posts around specific public events, such as elections. The paper will thus also include lessons learnt and thoughts for future election digital collections.

Relevance This paper is relevant to the Create theme of the Conference. The paper details how the Library collected a new form of archived content, social media, using innovative technology.

1 Introduction: the challenge of social media archival for elections

The State Library of New South Wales (the Library) has a mandate to collect and preserve documentary heritage about life in New South Wales (NSW) for future generations. For many decades, the Library has collected documentation about contemporary NSW events, focusing on the traditional media of newspapers, books, serials, manuscripts, pictures and photographs. In recent years, the Library has also turned its gaze towards the ephemeral realm of social media (Barwick et al., 2014), instrumenting a framework to preserve this public documentation of life. Public social media platforms, like Twitter[1], blogs, and public Facebook[2] pages, provide an opportunity for citizens to engage in commentary, political debate, information sharing, humour and perhaps most importantly the expression of unfiltered opinions.

The Library and the Commonwealth Scientific and Industrial Research Organisation (CSIRO) have been working together for a number of years to collect this ephemeral social media. Within the Data61 wing of the CSIRO,

[1] twitter.com

[2] facebook.com

researchers have developed the Vizie system (Wan and Paris, 2014), a social media analytics tool which is currently in use at the Library for this purpose.

While the framework and the Vizie system described in (Barwick et al., 2014) worked well for capturing public discussions on a variety of topics about life in NSW, it had certain limitations for other data collection strategies, for example one that focuses on a long-running public event. One example pertinent to any Australian state is the collecting of documentation about an election. This kind of collecting activity presents additional challenges as one needs to consider not only topic coverage but other aspects such as candidate and geographical coverage in terms of electorates.

To understand the issues better and to develop new tools to address this limitation in the data collection process, the Library and the CSIRO worked together to archive public discussions on social media regarding the 2015 NSW State Election, which took place on the 28th of March, 2015. Whilst the Library has for a long time collected election material - for example, how to vote leaflets, posters, and handbills - the 2015 Election was the first election where we sought to collect social media and its discussions.

The collection of social media content relating to elections raises new methodological and technical challenges. Firstly, one must decide upon a systematic process for defining query terms to be used with social media search engines; these should collect public discussions from all the

electorates and provide good coverage of election topics. Secondly, one needs to monitor the effectiveness of these terms and the topical relevance of the collected data; a time-consuming task that can quickly overwhelm Library staff.

In this paper, we describe changes to the collection framework protocol and extensions to the Vizie system to tackle these challenges. In particular, we describe how the new user interface provides feedback on which electorates are represented in the data set and the discussion topics covered therein, thereby allowing Library staff to more effectively curate the set of queries used for data collection.

In the remainder of this paper, we present related work on social media data collection and analysis for studies about political elections (Section 2). In Section 3, we describe the collection framework developed specifically for political elections. In Section 4, we describe the new Vizie user interface, designed to support the query curation process. We discuss insights possible from the collected data in Section 5. Finally, we summarise our findings in Section 6.

2 Related Work

In this section, we summarise the related work discussing the collection and analyses of social media discussions about political elections.

The partnership of the Library and the CSIRO to tackle the collecting of social media about life in NSW complements existing approaches to the collecting of online material. One such approach is the National Library of Australia's PANDORA Web Archive project (Cathro et al., 2001). The National Library has been collecting election related websites using Pandora since 1996. For the 2013 Australian Federal Election, the National Library archived the websites of candidates, parties, and interest groups and collected data from Youtube[3], MySpace[4] and a selection of Twitter accounts of candidates and parties.[5] For the 2015 NSW Election, the Library undertook archiving of similar material: its results can be found at Pandora.[6] However, what we were able to do as well, using Vizie, was to collect social media relating to specific queries not just accounts.

There is a growing body of work in developing natural language processing tools to help answer research questions about political elections. For example, Scharl and Weichselbraun (2006) and Ahmad et al. (2011) have studied the effects of media biases in social media. Researchers pursuing these paths have documented some of the procedures they have used to systematically collect data. Although often their focus is not archival, the collection

[3] Youtube.com
[4] Myspace.com
[5] For more details, see http://pandora.nla.gov.au/col/12283
[6] For more details, see http://pandora.nla.gov.au/col/13262

mechanisms they use are relevant to our work, and so we focus on a few examples from which we can generalise data collection best practices.

One approach is to collect a sample of social media data corresponding to the population who will vote, and then to search within that data for mentions of party names. For example, Tjong Kim Sang and Bos (2012) first filtered a Twitter stream to localise data to Dutch speakers for the context of elections in the Netherlands. To do this, they used well-known Dutch hashtags and keywords. A language filter was then applied on the filtered data. Within this general data set, election-related content was found by searching for references to political parties. Both the full name and the abbreviation of the party name were used. The authors collected a data set of approximately 7,000 Twitter posts.

Similarly, Lampos et al. (2013) employ a similar method to collect social media data for UK and Austrian elections. However, the collection focused on Twitter users rather than posts, as Twitter profiles for users can include some descriptions of location. Within the Twitter posts authored by these users, names and abbreviations for key political parties were then searched for to derive observation counts.

Instead of first collecting data about a region to then narrow the data selection to be about an election, one can use social media Application

Programming Interfaces (APIs) [7] to directly collect data using election-specific keywords. In the context of the 2011 Singaporean presidential elections (Choy et al., 2011), candidate names of the four presidential candidates were used. Twitter content was collected indirectly using the Google API resulting in approximately 16,000 Twitter posts.

The approach of Vizie combines elements of these preceding works. It subscribes to social media accounts for candidates and uses keywords (based on candidate names, political party names, electorate names and variants) to collect data. Data is collected from a number of different social media platforms, including Twitter, discussions on public Facebook pages, Instagram, blogs and news websites. Consequently, our data collection mechanism allows for the collection of very large data sets. In this work, approximately half a million social media posts were collected and archived.

Finally, we note that there are number of examples of research that seeks to either (i) predict an election outcome (Tumasjan et al., 2010) ; or (ii) determine the public sentiment towards a candidate in terms of positive or negative reactions (for examples, see Wang et al. (2012), Diakopoulos and Shamma (2010), and Tumasjan et al. (2010)). In those works, the focus is the prediction of a metric, whereas, in this work, we focus on the quality of the

[7] APIs are web services maintained by a social media platform that offering access to information services such as search engines.

data collected. For example, an incomplete but representative sample may predict the winner of an election, but it might not serve as a good representation of the election topics discussed to be preserved as a record for future research.

3 A collection framework for the NSW state election

The NSW State Election was held on the 28 March 2015. The Parliament of New South Wales has two democratically elected houses - the Legislative Assembly and Legislative Council. The Legislative Assembly is the 'Lower House' - comparable to the Federal Parliament's House of Representatives - and the party with majority support in this house forms government. The Legislative Council is the 'Upper House' or 'House of Review' and is comparable to the Federal Parliament's Senate.

For the election, all of the seats - 93 seats, one for each electorate - in the Lower House and half of the 42 seats in the Upper House were contested. A total of 504 candidates nominated for the 93 Lower House electorates, and 394 candidates nominated for the 21 Upper House seats. Four major parties contested the Election, the Liberal Party, the Labor Party, the National Party and the Greens. There were also a number of small parties and independents. The total number of voters was 5,044,562.

As the starting point for the Election collection framework, the Library utilised the existing Pandora election collecting classification. These subjects include *Candidates*, *Parties*, *Interest Groups*, and *Media*. To these primary subjects, the Library added appropriate secondary subjects: to *Candidates* and *Parties*, we added the name of the parties, and to *Interest Groups*, we added the area of their interest. For example, *Candidates-Australian Labor Party* and *Interest Group-Rural* were two secondary subjects we used.

To classify the discussions of the social media hashtags, the Library used the primary classification *Topic* and then refined it with the topic subject. Where the subject covered all political topics, the classification *Topic-General* was used. So, for example, the most popular social media hashtag for NSW politics is *#nswpol*. This was classified as *Topic-General*. Similarly, the popular election hashtags of *#nswvotes* and *#nswelection* were also classified as *Topic-General*. For more specific topics we used the specific subject area - for example, *Health, Indigenous, Infrastructure and Mining*. The *Topic-Mining* classification included the hashtags *#CSG, #LiverpoolPlans* and *#nocsg* as queries.

Where hashtags were instigated by a political party, they were classified under the appropriate *Parties* heading. For example, the Labor Party used the hashtags *#newapproach* and *#noplanBaird* and the Liberal Party used *#FoleyFail, #RebuildNSW* and *#KeepNSWWorking*.

With this broad framework in place, the initial focus was to identify the candidates, the parties and their digital sites. These sites could be, most likely, a website, Twitter, Facebook, or less likely, YouTube, Instagram[8], or GooglePlus[9]. This work was primarily undertaken four months before the Election, in December 2014 and January 2015. It was a resource intensive task; whilst many of the candidates for the major parties were listed on the parties' websites, there was no reference source that listed the candidates and parties and their digital sites. As candidate nominations did not close until two weeks prior to the Election, the NSW Electoral Commission did not produce the final confirmed list of candidates until March.

Once candidates, parties and their digital sites had been identified, the Library entered a range of queries into Vizie. These can be uploaded to Vizie in bulk using a spreadsheet import mechanism. The system allows queries and the corresponding data collected with those queries to be grouped into a larger unit called a *Monitoring Activity*. Consequently, the first step was to use these to represent the classifications of the collection framework. For example, there were monitoring activities for *Candidates-Australian Labor Party*, *Party-Liberal Party*, *Election Day*, *Topic-Mining*, *Interest Group-Unions*. Using

[8] Instagram.com
[9] Googleplus.com

monitoring activities allows access to predefined subsets of data throughout the Vizie tool and can be used as data for various analytic methods.

After monitoring activities were defined, queries were added to them. These queries consisted of combinations made up from the Twitter account name, the candidate name, the electorate and the party. Similar work was undertaken for interest groups and media organisations. Twitter and Facebook accounts were subscribed to, as were RSS feeds of relevant websites. Each of the queries and subscriptions were assigned a classification. Vizie would then collect posts that satisfied the queries and subscriptions. While queries were first entered into Vizie in December 2014, regular updates were performed as more information became available from various sources, such as the Australian Electoral Commission and the different political party websites.

With this initial 'groundwork' performed, there was the ongoing monitoring of what was happening in the election campaign, what was happening on social media, and what was being collected. There were a number of specific campaign events that were documented, each with related hashtags and keywords that were added to Vizie. For example, the debates between the party leaders - #*leadersdebate*; and the March 4 trade union march on Parliament House - #*March4*and #*SolidaritySelfie*. For the day of the election, a special classification, *Election Day,* was created with a number of queries. It

included “*below the line*”, “*polling place*”, “*vote 1*”, *#electionday*, *#fourmoreyears*, *#newspoll*, “*NSW election day”, “vote nsw”, “vote today”, “voting nsw”* and “*voting today”*.

The collecting started in December 2014 and ended in April 2015 following the declaration of results by the Electoral Commission. A total of 520,000 posts were collected. Each post equates to a Twitter post, Facebook post, blog post, etc. The posts were collected from a total of approximately 3,800 queries and subscriptions. The most popular queries/subscriptions were *#nswpol* (98,557 posts), *#nswvotes* (65,913), *@mikebairdMP* (30,667), *#csg* (20,248). These queries and subscriptions were assigned to a total of 28 classifications - 9 relating to the *Candidates*, 7 for *Parties*, 6 for *Topics*, 3 for *Interest Groups*, and 1 each for *Media*, *Election Day* and the *Electoral Commission*.

The range of the creators of the posts was wide ranging from political candidates, parties, interest groups, media organisations, government users and public users of social media. 13,000 unique Twitter users are represented in the collection. Interest groups collected covered the spectrum of opinions on issues such as mining, electricity poles and wires, and infrastructure development.

4 A co-designed user interface for monitoring data collection

As mentioned above, in previous work, the Library has used the Vizie social media system to collect and archive public social media relating to public life in NSW (Barwick et al., 2014). The Vizie system provides a federated search interface so that queries can be issued to the search engines of the various social media platforms. With user curation of queries in mind, Vizie provides feedback on the type of social media content that would be collected by presenting previews of search results from each of the search engines (Wan and Paris, 2014).

While configuring and refining queries *by topics* was possible with the federated search interface, it was of interest for the Library to perform refinements by considering the geographical range of the content being collected. This was important information from a collecting quality assurance perspective - for example, this helped to check if the queries entered were only collecting material from one region and missing the discussions in another.

To address this extra geographical consideration, the CSIRO and the Library co-designed a new interface to obtain feedback on which electorates were covered by the existing queries. The co-design process spanned several

months and involved regular fortnightly meetings between Library and the CSIRO to discuss progress on the interface and obtain feedback on early design sketches and prototypes.

4.1 Gauging coverage of electorate: a rule-based categorisation scheme

To determine which electorates in NSW were covered by the queries used for this project, the CSIRO implemented a rule-based categorisation system to assign posts, where it could be determined, to electorate labels.

The aim was not to determine how representative the data was of the entire electorate, as this requires validation data from external sources that is not always readily available. Rather, the aim was to quickly determine whether content for each electorate existed, using a series of reliable rules whose application could be easily interpreted by Library staff.

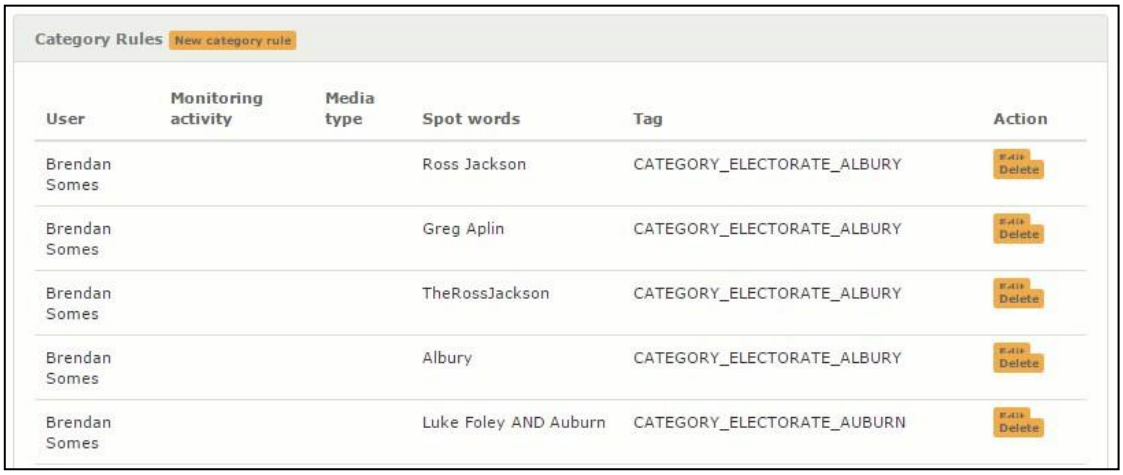

Figure 1. An example of the Vizie user interface to manage a rule-based categorisation system.

In conjunction with the Library, a user interface was designed that would allow the Library's Vizie users to define or refine one or more rules per electorate. Figure 1 shows a screen shot of the interface. As illustrated in the figure, rules contain words to identify (spot words), based on the candidate names, Twitter profile names, and the electorate name. In this case, four of the rules presented would assign the category *CATEGORY-ELECTORATE-ALBURY to* a post based on different words and phrases (e.g., "Albury", "Ross Jackson"). The last rule would assign a post to the category *CATEGORY-ELECTORATE-AUBURN* if the words "Luke Foley" and "Auburn" were detected.

The rules can be applied to a specific subset of posts filtered by the monitoring activity, or additionally to data from specific platforms. In the example above, such filters are not used, and so the rules apply to all posts.

As metadata, the interface lists the creator of the rule (e.g., “Brendan” in the figure), since Vizie is a collaborative environment and different rules may be created by different Vizie users. The rules could always be refined further or deleted.

For this work, the CSIRO Vizie team also provided data management services to the Library to import a large set of rules. For example, from the list of candidates, their social media accounts, and the list of the electorate names, it was possible for the Library to populate a spreadsheet with candidate rules and vet them manually for soundness using external tools like Excel. Once these rules were ready, they were imported into Vizie, which then automatically used them on any newly collected data. Retrospective application of the rules could also be triggered as a once-off process.

Aggregate statistics are collected in the background to keep track of how many posts are assigned to each category. Such counts are used to populate an interactive display as in Figure 2. Here the user has selected the electorate of “Manly” for the period from the 1st of January, 2015 to the 25th of March, 2015. The bar representing the volume of data has subsequently been highlighted blue. This changes the descriptive statistics in the rest of the

interface which shows: (i) which query terms and account subscriptions have contributed to that selection of data (for “Manly”), see Figure 3; (ii) which social media platforms are represented in that data subset (predominantly “microblog”, the Vizie system’s label for Twitter), see Figure 4; and (iii) the keywords associated with that data subset, see Figure 5. Finally, in Figure 6, we show how the selected posts are presented in reverse chronological order in the interface. Within Vizie, Figures 2-6 are presented together in the same web interface. Using this interface, Library staff monitored the application of the rules to refine the queries if necessary.

Interestingly, the top ranked electorates include the electorate of Manly, the seat of the incumbent Premier for the state of NSW (who was recognised as having a strong social media presence), and electorates where the election battles were hardest fought.

Key words are automatically detected from each post based on a combination of the temporal context of what is trending and heuristics about the importance of each word in isolation. What is interesting from a query curation point of view is that the key words can be used to suggest new query terms related to a subset of data. This list can reveal other hashtags that may be prominent in the data set but which are not being used to collect data.

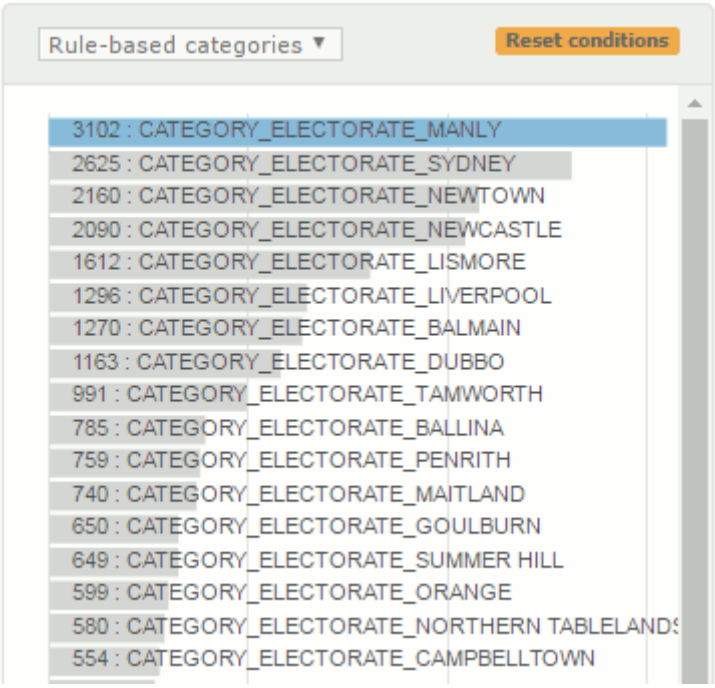

Figure 2. An example of an interactive visualisation showing electorate counts.

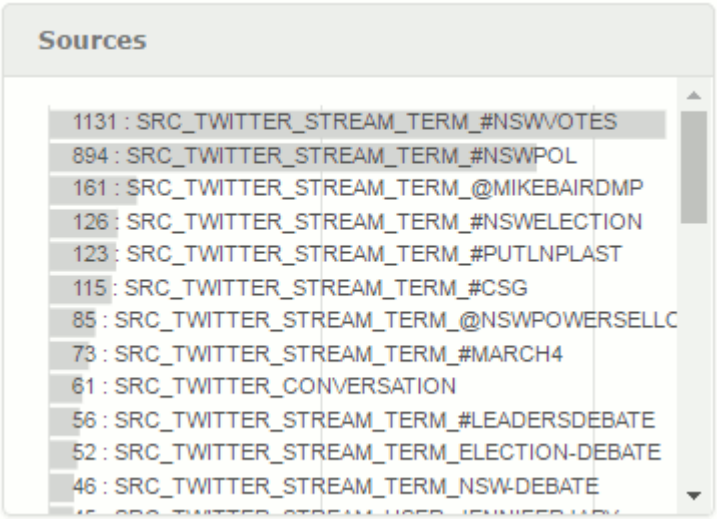

Figure 3. The queries and data sources responsible for the content related to "Manly"

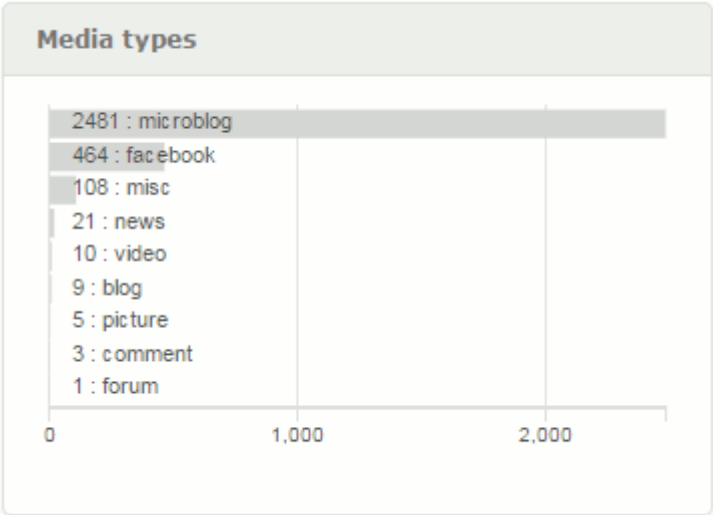

Figure 4. A summary of the distribution of content across different media types.

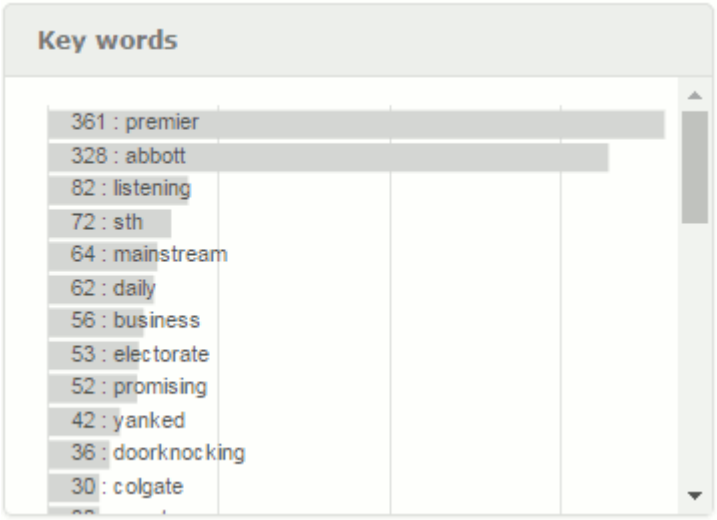

Figure 5. The keywords associated with the data from "Manly".

Figure 6. A reverse chronological presentation of the selected data.

4.2 Measuring shared news articles: a surrogate for an event summary

To help vet that the collected content is part of a state-wide public discussion on election topics, we designed a new visualisation that highlights shared news content. In this view, a news article can be ranked based on the number of times it was shared. At present, shared news content is mined from the collected Twitter content, given that one of its primary uses is for news dissemination.

By examining which election news articles are popular, we can characterise the popular issues that weigh on the collective mind of the electorate. This acts as a kind of summary of the public discussions during that time period.

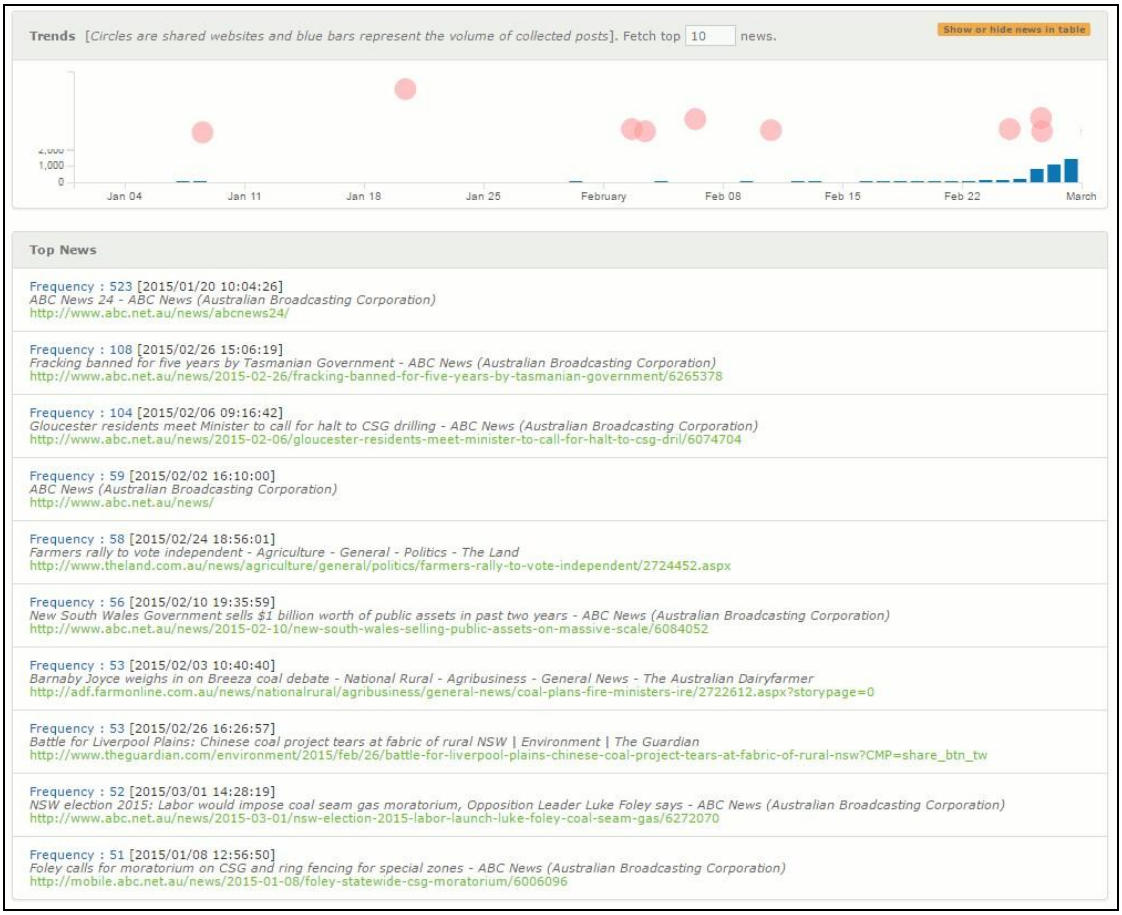

Figure 7. An example of an interactive visualisation showing shared news content.

To compute these counts, shortened URLs detected in Twitter content were automatically followed, and the linked web resources are archived, if possible. Each web resource was represented explicitly in the Vizie system to keep track of how many Twitter posts link to it. The web resources were then filtered to contain only news articles for display in the visualisation.

The visualisation for the period from the 1st of January, 2015 to the 22nd of March, 2015 (the beginning of the final week of the election period) is displayed in Figure 7. The number of shared news articles can be adjusted. By default, the interface shows up to 20 articles, but in this screenshot, we have adjusted it to display up to 10 articles. Each article is represented as a circle, with its height on the y-axis indicating the number of posts that point to it. The x-axis indicates time. The volume of social media content is displayed as blue bars on the same graph.

The visualisation includes a second display showing all the headlines from the articles. These are then displayed in order of most to least shared. In this example, the top shared news articles are selected from all data about all electorates, so the view represents the entire state. As one can see, the shared news articles are either about the election or else a link to the ABC website, which at the time would have focused on the election. This indicates that the queries are reasonably well-curated; had there been news articles on a different topic that would have meant the queries needed further refinement.

The visualisation is interactive. Clicking on a circle will show the sharing pattern for that news article, where the number of shares per day are graphed. By clicking on the headline, the user can “drill down” to see the full news article.

5 Discussion: Lesson learnt

The new interface allows Library staff to provide some quality assurances for the election-specific data collected framework. The categories also serve as an index into the data and showed that each electorate was represented in the data collected. Ideally, these would assist researchers in undertaking sociological studies. We sketch how this is possible within the current interface.

Mining was a prominent issue in the Election. Over 43,000 posts were collected in the *Topic-Mining* classification. This is supported by the shared news articles, as evidenced by Figure 7, and in other statistics such as the keyword analysis for the period.

We also explored the correlation between rankings of election issues represented in the shared news articles and the ranking determined by a survey conducted by VoteCompass[10] on behalf of the ABC (Wan and Paris, 2015). In the VoteCompass survey, feedback on the prevalence of different election issue categories were obtained through an online form and represents the views of the participants. Our data had a moderate correlation in rankings with that of the VoteCompass survey, with the Kendall's Tau of 0.55 (2-sided p = 0.047), which is statistically significant at α = 0.05. Our data suggests that

[10] VoteCompass.com

the environment was the number one ranked category. Interestingly, we were able to show that rankings with this level of correlation were possible using only data collected up to mid-February 2015. This highlights the potential to use these data insights to understand key voter issues in the lead up to an election event.

6 Conclusion

In this work, we presented a case study in which we introduce and demonstrate the use of a new data collection framework for the purposes of collecting public discussions on social media about the 2015 NSW State Election. The study demonstrated that social media can be collected for an election, and that the queries could be curated in an effective manner. The collection of over 500,000 posts comprises posts collected from subscribing to specific accounts, for example Mike Baird's Twitter account, and those satisfying specific hashtags like *#nswvotes* and other election-related keywords.

The new user interface enabled Library staff to perform quality assurance validations on the posts being collected in order to update the collection framework when it was determined that a collection gap existed. Using the new categorisation feature in Vizie, Library staff were able to determine that content for each electorate was collected. Furthermore, when examining the

top-ranked electorates (based on the categorisation rules), these were observed to be the hotly contested election battles, providing validation that the collection framework was indeed collecting data as intended. From a usability point-of-view, the categories also provide for dynamic indexing of material that can provide future entry points to the collection for future researchers.

The combination of the new collection framework and the new user interface in Vizie means that collecting social media for elections (and other events) can be done on a large scale. The number of subscriptions/queries entered into Vizie totalled 3,800 - this is far above what has been done in the past, as outlined in related work. The Vizie functionality of being able to bulk import lists of queries and subscriptions meant a large number of queries/subscriptions could be entered quickly.

We observed that the most resource intensive work from the Library's perspective was the identification of the candidates and their digital presences. It may be that this cannot be avoided; however, for future elections the capacity to share information across government and political parties should be considered. In future work, we would like to consider data integration approaches to capitalise on existing data repositories that can help fast-track the configuration of Vizie. For example, automatically mining

resources maintained by the Electoral Commission is an obvious candidate where information of this kind could be utilised.

To finish, an aspiration: we hope that the data collected can be made available (in a variety of forms including anonymised aggregated statistics) to researchers in real-time to explore and study. Our hope is that this can lead to new information tools that help us to vote and understand better the complexities of elections and life in NSW.

References

Khurshid Ahmad, Nicholas Daly, and Vanessa Liston. (2011). What is new? News media, General Elections, Sentiment, and Named Entities. In *Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2011)*, pages 80–88, Chiang Mai, Thailand, November. Asian Federation of Natural Language Processing.

Kathryn Barwick, Mylee Joseph, Cecile Paris, and Stephen Wan. (2014). Hunters and collectors: seeking social media content for cultural heritage collections. In *VALA2014, 17th Biennial Conference and Exhibition*, pages 3–6.

Warwick Cathro, Colin Webb, and Julie Whiting. (2001). Archiving the Web: The PANDORA Archive at the National Library of Australia. In the *Proceedings of the Preserving the Present for the Future Web Archiving Conference*, pages 18– 19.

Murphy Choy, Michelle L F Cheong, Ma Nang Laik, and Koo Ping Shung. (2011). A sentiment analysis of Singapore Presidential Election 2011 using Twitter data with census correction. Number abs/1108.5520. CoRR.

Nicholas A. Diakopoulos and David A. Shamma. (2010). Characterizing Debate Performance via Aggregated Twitter Sentiment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 1195–1198, New York, NY, USA. ACM.

Vasileios Lampos, Daniel Preoiuc-Pietro, and Trevor Cohn. (2013). A user-centric model of voting intention from social media. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 993–1003, Sofia, Bulgaria, August. Association for Computational Linguistics.

Arno Scharl and Albert Weichselbraun. (2006). Web Coverage of the 2004 US Presidential Election. In *Proceedings of the 2nd International Workshop on Web As Corpus*, WAC '06, pages 35–42, Stroudsburg, PA, USA. Association for Computational Linguistics.

Erik Tjong Kim Sang and Johan Bos. (2012). Predicting the 2011 Dutch Senate Election Results with Twitter. In *Proceedings of the Workshop on Semantic Analysis in Social Media*, pages 53–60, Avignon, France, April. Association for Computational Linguistics.

A. Tumasjan, T.O. Sprenger, P.G. Sandner, and I.M. Welpe. (2010). Predicting elections with Twitter: What 140 characters reveal about political sentiment. In *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, pages 178–185.

Stephen Wan and Cecile Paris. (2014). Improving government services with social media feedback. In the *Proceedings of the IUI'14 19th International Conference on Intelligent User Interfaces, IUI'14, Haifa, Israel, February 24-27, 2014*, pages 27–36.

Stephen Wan and Cecile Paris. (2015). Ranking election issues through the lens of social media. In the *Proceedings of the 9th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH)*, pages 48–52, Beijing, China, July. Association for Computational Linguistics.

Hao Wang, Dogan Can, Abe Kazemzadeh, François Bar, and Shrikanth Narayanan. (2012). A System for Real-time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle. In *Proceedings of the ACL 2012 System Demonstrations*, pages 115–120, Jeju Island, Korea, July. Association for Computational Linguistics.