Integrating and Interpreting Social Data from Heterogeneous Sources

Matthew Rowe[1] and Suvodeep Mazumdar[2]

[1] OAK Group, Department of Computer Science,
[2] Department of Information Studies, University of Sheffield, Regent Court, 211
Portobello Street, S1 4DP Sheffield , United Kingdom
m.rowe@dcs.shef.ac.uk, s.mazumdar@shef.ac.uk

Abstract. Social data is now being published at a never seen before scale. The provision of functionalities and features on a wide range of platforms from microblogging services to photo sharing platforms empowers users to generate content. However, such is the rate of publication, and the wide range of available platforms to facilitate the creation of social data, that interpreting this data is limited. In this paper we present an approach to interlink social data from multiple Social Web platforms by using Semantic Web technologies to achieve a consistent interpretation of the data. We present a web application to demonstrate the effectiveness of this approach, using the Cumbrian Floods in the UK as a use-case for anomaly detection within published social data.

1 Introduction

Social Web platforms such as Twittter[3], Facebook[4], and Flickr[5], have seen widespread uptake and adoption across the Web. In each platform, and indeed across the Social Web in general, the focal point of usage is the end-user, empowering the individual with functionality and feature sets which make creating content and participating online easy. Whether it is the publication of a microblog on Twitter or uploading and sharing a photo on Flickr, the technical barrier is reduced to the click of a button. Web users are now content creators, sharing data with a multitude of distributed and disparate platforms and services. The motivation behind publishing such data is, in general, social: i.e. to share with friends, or to receive critique from a community. Therefore we denote any single user generated content item (image, video, message) shared with a community as a *social data fragment*.

Social data is now published at a never before seen scale. For instance on Twitter alone, 50 million microblogs are published every day with an average of 600 Tweets per second[6]. This scale of publication leads to *information overload*,

---

[3] http://www.twitter.com
[4] http://www.facebook.com
[5] http://www.flickr.com
[6] http://blog.twitter.com/

where making sense and interpreting social data becomes a problem. Current efforts to address this issue, such as trend services, require a user to *listen in* on a particular topic or subject in order to filter the relevant material. Furthermore such services only concentrate on a single source for social data at a given term (i.e. one single Social Web platform such as Flickr). Fusing social data from heterogeneous sources would provide web users with an overview, and a clear consensus of information, rather than a single portion. One of the interesting qualities of social data is its multi-faceted nature which can be broken down into three facets:

- *Provenance*: Who published the data? From what source? And when was the data published?
- *Topic*: What is the social data about? What is it tagged with?
- *Geo*: Where was the social data published?

At present existing trend services which analyse social data, such as Trendistic[7] and Blog Pulse[8], do not take into consideration the *geo* facet of social data. Given the increased use of smart phones and mobile applications, such as Four Square[9], social data is now being published which is geotagged, therefore requiring approaches to incorporate this *geo* facet into future analysis. Furthermore a multi-faceted perspective of social data would provide extra dynamics of the data to the end user, and would allow comparisons and suggestions based on the user's profile by a) considering where the person lives and showing social data which is relevant to that area, and b) showing social data published in the past which may be of relevance to the user.

In this paper we present an approach to interlink and interpret social data from heterogeneous sources. The approach is grounded in the use of Semantic Web technologies in order to provide a consistent interpretation of information from distinct sources. Our approach allows analyses to be performed over data distributed across the Social Web based on its multi-faceted nature. To ground our approach we use the scenario of anomaly detection and a dataset containing social data collected from Twitter and Flickr, for the year 2009 and the county of Cumbria in the UK. During this time the area experienced heavy flooding, the effects of which were reflected in the surge in social data production around that time. We provide a usable web application that exploits the facets of social data enabling an end-user to interpret a large amount of social data and therefore discover anomalies.

We have structured the paper as follows: section 2 presents our approach to combining social data from multiple sources, describing the process by which metadata is generated for social data and interlinking is achieved. Section 3 describes how we utilise the interlinked social data to discover anomalies within the data. Section 4 presents related work within the area of interlinking social data and current trend services. Section 5 finishes the paper with conclusions learnt from this work and our plans for future work.

---

[7] http://trendistic.com/
[8] http://www.blogpulse.com/
[9] http://www.foursquare.com

2 Interlinking Social Data

In order to interlink social data and allow it to be analysed we must first overcome the problem of social data being provided in proprietary formats. This is a common issue when interfacing with the APIs of Social Web platforms as the data from one source will be provided using a different data schema to data from another source. To address this limitation we use the approach shown in Figure 1. We first export social data from multiple platforms in their own format. For each platform we convert the returned data into RDF providing metadata descriptions using concepts from Web accessible ontologies. We then store the RDF collected from each platform in a central repository, this allows SPARQL queries to be processed over social data from heterogeneous sources. We now explain our process of building metadata from various sources before moving on to explain the implicit interlinking which is provided and several queries we are able to process over the data.
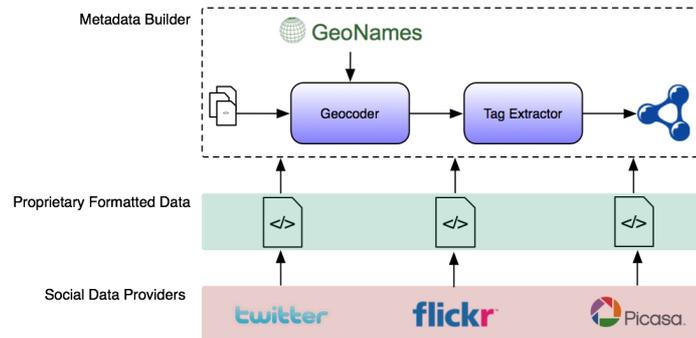

Fig. 1. Approach to extract and interlink social data

2.1 Metadata Generation

Social Web platforms and services provide access to data using APIs and data feeds. In the majority of cases the response of API calls is returned as XML according to the XML schema of the platform. For instance, when querying the Twitter API[10] for the user profile of one of the authors of this paper[11] we are returned the following response:

```
<user>
  <id>13092722</id>
  <name>Matthew Rowe</name>
  <screen_name>mattroweshow</screen_name>
  <location>Sheffield, UK</location>
```

---

[10] http://apiwiki.twitter.com/
[11] http://twitter.com/users/show.xml?screen_name=mattroweshow

```
  <description>PhD Student / Semantic Web / Web 2.0 Enthusiast</description>
  <url>http://www.dcs.shef.ac.uk/~mrowe</url>
</user>
```

We wish to interlink social data from distinct sources distributed across the Social Web. In terms of Twitter we define a single *social data fragment* as being a *Tweet*, more commonly known as a Microblog post. In terms of Flickr and Picassa a single *social data fragment* is an image. When querying Twitter for all the social data fragments that a users has produced we are provided with an XML response of the microblogs in descending chronological order. A single *social data fragment*[12] from the above user is provided in the following form:

```
<status>
  <created_at>Sun Feb 28 12:22:47 +0000 2010</created_at>
  <id>9774519667</id>
  <text>Writing up our Geovation work for #lupas2010.</text>
  <truncated>false</truncated>
  <in_reply_to_status_id></in_reply_to_status_id>
  <in_reply_to_user_id></in_reply_to_user_id>
  <favorited>false</favorited>
  <in_reply_to_screen_name></in_reply_to_screen_name>
  <geo xmlns:georss="http://www.georss.org/georss">
    <georss:point>53.3833,-1.4722</georss:point>
  </geo>
</status>
```

As mentioned previously, social data consists of three facets: *provenance*, *topic* and *geo*. In the above response snippet a single social data fragment contains each of these facets: the `<created_at>` element contains the provenance information (time of the fragment's creation) and the `<text>` element provides information describing the topic of the fragment, the `<geo>` element contains information about the geo facet of the data fragment. Using this information we build an RDF representation of the data fragment and represent the relevant information in a machine-readable and reusable way as follows:

To begin with, we create a URI for the data fragment using the derefenceable URL describing, in this case, the microblog post. We define this as an instance of *sioc:Post* from the SIOC (Semantically Interlinked Online Community) Ontology [2] and also as an instance of *itr:LocalizedResource* from the WeKnowIt Interaction Ontology[13]. This latter concept allows the instance to be defined as localized resource in the sense that the data was published at a given geographical location - this defines the *geo* facet of the social data fragment. We then associate the data fragment with the person who created it using the URI of the Twitter user. This allows queries to be performed which gather all the microblogs published by that user. The content of the microblog is then associated with *sioc:Post* instance using the *sioc:content* property. This forms the full description of the *topic* of the social data fragment. To enable easier discovery of social data for a given topic we extract all the tags from a given social data fragment. In terms of a Microblog these are the *hashtags* from within the content of the post - a given term preceded by a # symbol. For each extracted tag

---

[12] http://twitter.com/statuses/user_timeline.xml?screen_name=mattroweshow
[13] http://www.dcs.shef.ac.uk/~gregoire/interaction/ns#

we associate this with the social data fragment using the *dc:subject* property. To attribute geographical information to the instance of *sioc:Post* instance we create an instance of *gml:Geometry*, and assign to it the longitude and latitude of the social data fragment - this is extracted from the `<geo>` element in the above response code.

We create an instance of *foaf:Person* for the user who published the *social data fragment* and assign this user their name - using *foaf:name* - together with the posts they have published. A given user is assigned a URI based on their twitter username, this can be dereferenced to gather information about the user who published the social data. This forms our initial piece of provenance information. For the timeliness of the *social data fragment* we use the timestamp found within the `created_at` element of the above XML and assign this to the RDF representation of the data fragment using the *dcterms:created* property. An example of a given Twitter user with a single Microblog in RDF using Notation 3 (N3) syntax [1] looks as follows:

```
<http://twitter.com/mattroweshow>
        rdf:type foaf:Person ;
        rdf:type itr:LocalizedResource ;
        foaf:name "Matthew Rowe" ;
        foaf:homepage <http://www.dcs.shef.ac.uk/~mrowe> ;
        itr:has_Localization _:a1 .
<http://twitter.com/mattroweshow/13092722>
        rdf:type sioc:Post ;
        rdf:type itr:LocalizedResource ;
        sioc:hasCreator <http://twitter.com/mattroweshow> ;
        sioc:content "Writing up our Geovation work for #lupas2010." ;
        dcterms:created "2010-2-28 12:22:47.0" ;
        dcterms:subject "lupas2010" ;
        itr:has_Localization _:a2 .
_:a1
        rdf:type gml:Geometry ;
        gml:pos "53.3833,-1.4722" .
_:a2
        rdf:type gml:Geometry ;
        gml:pos "53.3833,-1.4722" .
```

So far we have only considered social data from the microblogging platform Twitter. Our goal is to combine social data from multiple sources, thereby interlinking it together. Another source for our social data is Flickr - a photo sharing site. In this instance we define a *social data fragment* as constituting a single photo which is shared on the site. When querying Flickr's API[14] for photos about a specific topic or posted by a given user we are returned an XML response, a single *social data fragment* - representing an image - is contained within the `<photo>` element as follows:

```
<photo id="949406913" media="photo">
  <owner nsid="54948696@N00" username="mattroweshow" location="England" />
  <title>DSC00171.JPG</title>
  <description></description>
  <visibility ispublic="1" isfriend="0" isfamily="0" />
  <dates posted="1205398307" taken="2009-01-09 09:16:31" lastupdate="1257421561" />
  <editability cancomment="1" canaddmeta="0" />
  <usage candownload="0" canblog="0" canprint="0" />
```

---

[14] http://www.flickr.com/services/api/

```
<comments>0</comments>
<notes />
<tags>
  <tag id="24539622-2330113101-400" author="54948696@N00" raw="arctic">arctic</tag>
  <tag id="24539622-2330113101-401" author="54948696@N00" raw="monkeys">monkeys</tag>
</tags>
<location latitude="53.4813" longitude="-2.2392" place_id="R8vDw_abBpSzUA">
  <locality place_id="R8vDw_abBpSzUA" woeid="27872">Manchester</locality>
  <region place_id="pn4MsiGbBZlXeplyXg" woeid="24554868">England</region>
  <country place_id="DevLebebApj4RVbtaQ" woeid="23424975">United Kingdom</country>
</location>
<geoperms ispublic="1" iscontact="0" isfriend="0" isfamily="0" />
<urls>
  <url type="photopage">http://www.flickr.com/photos/54948696@N00/949406913/</url>
</urls>
</photo>
```

We use a similar process for generating RDF for microblogs when creating RDF for images. We begin by creating an instance of *sioc:Item* to represent the *social data fragment*, which in this instance is an image. The semantics of using this class definition encapsulates any piece of content that is published within an online community space. We provide a URI for this instance using the URI of the image in Flickr - this is the URL which can be accessed to view the image - found within the `<url>` element of the above XML response. We create an instance of *foaf:Person* to represent the user on Flickr who created the photos and assign this instance a URI corresponding to their URI within the Flickr platform. This provides our first piece of *provenance* information about the *social data fragment*, the second piece of information is created from the date and time when the image was taken, which is found within the `taken` attribute of the `<dates>` element in the XML response. We assign this information to the photo instance using the *dcterms:created* property.

For the *topic* facet of the data fragment we use the tags assigned to the photo which are provided as individual elements of `<tag>` for each tag. As with the Twitter metadata, we assign the tags to the *social data fragment* using *dcterms:subject*. For the *geo* facet of the data we use the values from the latitude and longitude attributes within the `<location>` element. An instance of *gml:Geometry* is created to represent the geo location of the data fragment - in this case where the photo was taken - and is attributed the latitude and longitude using the *gml:pos* property. The location instance is related to the data fragment using the *itr:has_ Localization* predicate. Triples built from the above XML response would look as follows (using n3 syntax):

```
<http://www.flickr.com/people/54948696@N00>
        rdf:type foaf:Person ;
<http://www.flickr.com/photos/54948696@N00/949406913>
        rdf:type sioc:Item ;
        rdf:type itr:LocalizedResource ;
        sioc:hasCreator <http://www.flickr.com/people/54948696@N00> ;
        dcterms:created "2009-01-09 09:16:31.0" ;
        dcterms:subject "arctic" ;
        dcterms:subject "monkeys" ;
        itr:has_Localization _:a3 .
_:a3
        rdf:type gml:_Geometry ;
        gml:pos "53.4813,-2.2392" .
```

We can perform the same process for other Social Web sites such as Facebook and Picasa[15]. If the social data fragment is text-based then we create an instance of *sioc:Post* and assign the available information to it, otherwise, i.e. it is a video/image, we create an instance of sioc:Item. There are cases when handling social data, both from Twitter and Flickr, where no geocoded information is supplied - i.e. latitude and longitude of a location. In such instances we must build the geo information from location names. To do this we query the Geonames web service[16] using the location details - i.e. place name and country. The service returns a list of candidate URIs and geo information for the place ranked by popularity. We choose the top geo information from the list and use this as the geocoded representation of the data fragment. Of course in an ideal world everything would be geocoded, thus alleviating our need for geocoding.

2.2 Intergrated Social Data

As Figure 1 shows our approach functions by compiling a single RDF dataset containing social data fragments from multiple sources. As we have used common semantics to describe social data our interlinking functions in an implicit manner. We do not attempt to match content explicitly, instead we rely on the consistent metadata descriptions to facilitate SPARQL queries[17] across social data from heterogeneous sources. Using the following query we are able to gather all the data items which are associated with the "iranelections" and return them ordered by their date of publication. This would return all the images taken and the microblogs posted about the elections in descending chronological order.

```
PREFIX dcterms:<http://purl.org/dc/terms>
SELECT ?item
WHERE {
        ?item dcterms:subject "iranelections" .
        ?item dcterms:created ?date
}
ORDER BY DESC(?date)
```

Using the *geo* facet of the data we are able to perform a SPARQL query that retrieves all data fragments associated with a given location. For example, we can perform the following query which gets all the data fragments and their accompanying tags associated with the University of Sheffield's Department of Computer Science:

```
PREFIX dcterms:<http://purl.org/dc/terms>
PREFIX itr:<http://www.dcs.shef.ac.uk/~gregoire/interaction/ns#>
PREFIX gml:<http://www.opengis.net/gml/>
SELECT DISTINCT ?post ?tag
WHERE {
        ?post dcterms:subject ?tag .
        ?post itr:has_Localization ?geo .
        ?geo gml:pos "53.38091,-1.48067"
}
```

---

[15] http://www.picasa.com
[16] http://www.geonames.org/export/web-services.html
[17] http://www.w3.org/TR/rdf-sparql-query/

3 Interpreting Social Data: Cumbrian Floods Use Case

To demonstrate the effectiveness of our approach to integrating social data from multiple sources we now present a web application which presents data describing the region of Cumbria in the United Kingdom. In November 2009 this region suffered some of the worst flooding in its history[18]. Our intuition was that such a phenomena would be reflected in the publication of social data on the World Wide Web. Furthermore visualising this social data in a meaningful way would allow it to be interpreted and analysed more closely. To explore this hypothesis we extracted all microblogs published on Twitter from the year 2009 by users who lived in Cumbria. We first gathered a list of 200 Twitter users who lived in the region and extracted each person's tweets published throughout the year, this produced 3513 data fragments from Twitter. We then extracted all images from Flickr which had been taken within that area which produced 6663 data fragments. For both social datasets we used the above approach and generated an RDF dataset using consistent semantics, this generated 475,043 triples from Twitter data fragments and 182,304 from Flickr data fragments. Although we collected more data fragments from Flickr, more triples were created for the Twitter data due to the widespread use of hastags. This system is available at the following address:

> http://www.dcs.shef.ac.uk/∼suvodeep/ViziSocial

The data is visualised on Google Maps[19] based on the geocoded location of the *social data fragments*. The end user is able to zoom in or out altering the focus of the map, thereby increasing or decreasing the visible social data. Along the map, the user is presented with a slider, text box and a tag cloud. The user can then use the slider and text box to alter the visualisation. Zooming or panning in the map, dragging the slider or typing into the text box creates dynamic queries that are passed into the visualisation module to display the filtered results. Figure 2 shows the visualisation that has been developed.

3.1 Interactions

The slider in the Figure 2 represents individual days in the year 2009, starting from 01/01/2009 on the left, up to 31/12/2009 on the right. Dragging the slider to any particular day would select all the social data that has been posted on that particular day, and then display it on the map according to their associated geo-locations. On the right hand side, the tag cloud displays the *topic* facet of the selected social data, weighted according to how many times a given subject has occurred. The tag cloud is, in a sense representative of all the data displayed on the visible area of the map. At times, there are certain topics that the user might find interesting, and would only like to visualise the social data that has been associated with that topic. The user can then make use of the text box

---

[18] http://news.bbc.co.uk/local/cumbria/hi/people_and_places/newsid_8378000/8378388.stm
[19] http://maps.google.co.uk/

provided above the tag cloud and type in the query they would like to visualise e.g typing 'job' in the text box will look for all data fragments that have been tagged with 'job' on the day or any other day the user chooses to view. The data fragments from Twitter are displayed as blue markers in the map and the fragments from Flickr are displayed as pink markers. On clicking the markers, the users is shown either the tweets at that location or thumbnails of the photos from Flickr.

This visualisation implementation is designed to give the end user maximum control over what they can view and the ability to alter the facets of the data in a bespoke manner. A user can, quickly and effectively, view social data based on location, time and subject, the three facets of social data, thereby easing the process of detecting anomalies and analyzing trends relevant to a given user. This process follows the well known Shneiderman's approach of "*overview first, zoom and filter, then details-on- demand*" [6].

3.2 Observations

When loaded, the interface shows how social data has been shared and published in Cumbria over the past year, and many interesting trends, which normally would be very difficult to identify. As discussed previously, this study has been concentrated on social data associated with Cumbria - i.e. user's publishing microblogs from that region or photos taken there - and have been labelled with a given tag. Figure 2 shows the distribution of social data in Cumbria on 23/11/2009.
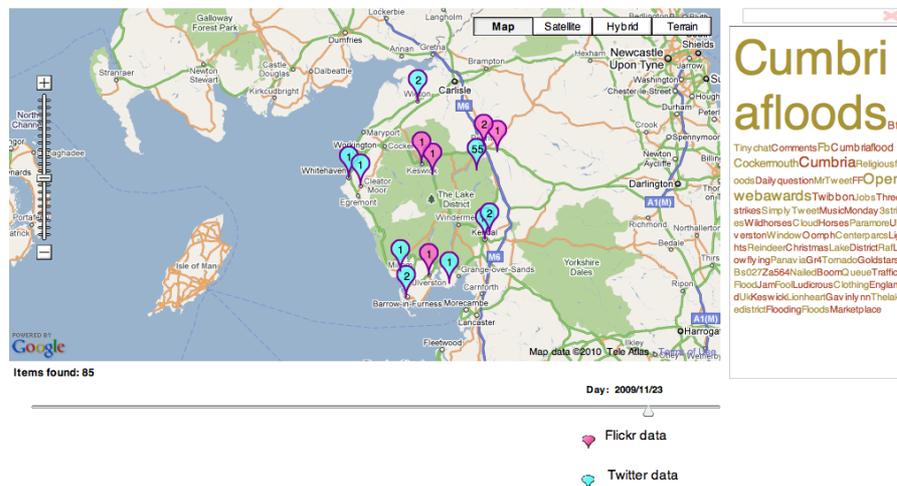

Fig. 2. Visualizing social data for users of Cumbria on 23/11/2009, showing floods at Cumbria and Cockermouth as a major topic of microblogging in the area on the day

The tag cloud that displays the topics of all the tweets show that 'Cumbriaflood', 'Cockermouth' and 'Cumbria' have been a major topic of discussion on the day. The benefit of this type of visualisation is that the users can immediately identify the trending topics of the day, thereby getting an idea of what people were talking about on the day. This can provide insights of what people talk about during major disasters or immediately before and after them. Clicking on the individual markers within the display provides further details of the *social data fragment*. Twitter users were posting updates about the condition of their locality or asking questions if a route is advisable to take and so on. For example, setting the slider to 19/11/2009, textbox query as "*flood*" and zooming into Kendal, there are 5 tweets shown. These tweets point to pictures of the flood and also provide information about the status of the floods and their localities e.g "*By pass out of Kendal to motorway now fully closed. Situation worsening*" or "*Has the Duddon Bridge collapsed?*". With the same filters, if we zoom into Windermere, we can see 9 flickr images clustered in the area. Clicking on the marker, we can see thumbnails of the images, and can immediately assess the level of flooding in that area.

Dragging the slider further on after the days of the disaster, the effect caused by the disaster is noticeable for a long time. Immediately after the floods, the effects are evident with the communities working towards helping people affected by the disaster and microblogs like "*GP cover available for URGENT home visits .. *", "*GPs: Drop-in clinics available all week..*" and so on appear. Tweets like "*Wath Brow Bridge closed after cracks found in structure*", "*FIRST grant given to flood victim Thanks for £345,000 already given PLEASE donate*" are aimed at providing further updates or request for support from the Twitter community. Looking at this kind of data is very helpful as it shows how people and communities interact with each other, help and provide support for the distressed, and build platforms for improvement of local services and so on. Information like this can prove to be invaluable to rescue services to find which areas are the most affected or even which routes are best to take.

4 Related Work

Attributing consistent semantics to social data has been explored in work by [3] in order align tags from videos with the concepts they represent., where the ambiguity of tags hinders the derivation of important information. Aligning tags with distinct dereferenceable concepts, from DBPedia, provides interpretation of social data, focusing on the *topic* facet of the *social data fragment*, which in this case is a video. Another approach to *semantify* social data is presented in [4]. In this instance IRC chat logs are converted into a machine-processable form using Linked Data principles and the SIOC ontology. In the same manner as our approach, each message posted within a designated chat room is denoted as an instance of *sioc:Post* and is associated with its author using the *sioc:hasCreator* predicate. The author is then identified using his/her WebID, which is defined as his/her URI, such that all the IRC message posted by the user can be retrieved.

Similar to our work involving Twitter, [5] introduce SMOB (Semantic Microblogging), an application which creates semantically enriched microblogs as Linked Data. The SIOC ontology is once again used to provide metadata descriptions for the data fragments. The Flickr Wrapper[20] allows DBPedia concepts to be searched, and using their URIs, retrieves photos from Flickr. The data fragments of the photos are represented using utilises Semantic Web technologies. Correlations exist between our approach and the intentions of the SIOC project[21] in general. The ethos behind SIOC is to bridge the gap between online communities, such that the data produced by a given person in multiple spaces could be leveraged and linked together. We believe that our approach provides an extension of this work, by considering additional facets of social data and providing a means by which social data can be interpreted based on these facets.

As we alluded to within the introduction of this paper, several trend services are available for social data vendors: Flickr trends[22], Trendistic[23] and Blog Pulse[24]. While these services provide tag-based trend information (i.e. *topic* facet) to an end user coupled with chronological information (i.e. *provenance* facet), any *geo* information is ignored. Moreover anomalies of social data as a whole may not be relevant to end users, instead they may only be interested in events within their region during a given time period. Additionally when performing manual analysis of social data based on known keywords this task is restricted by the lack of current efforts to visualise social data in a meaningful way. Instead users must search for a given topic e.g. "G20 protests" and then go back to the date when the summit was held. Our approach to overcome this burden on the end user collects social data and presents it in a logical manner. Presentation of social data in this way allows the end user to browse the data to discover trends and anomalies based on its different facets - focussing on a given location, looking at a given time period, searching for a given tag. This we believe presents the future in social data analysis by reducing the explicit prerequisites imposed on data - i.e. knowing what topic to search for - and allowing implicit anomalies to be perceived which are relevant to the user and not merely the general user base.

5   Conclusions

In this paper we have presented an approach to interlink and interpret social data from disparate Social Web platforms. The involvement of web users as content generators has seen an explosion in the rate of social data production, either in the frequent publication of microblogs or the uploading of images onto a photo sharing site, web users are now sharing more information than ever before. We believe that our approach to generating metadata description of *social data*

---

[20] http://www4.wiwiss.fu-berlin.de/flickrwrappr/
[21] http://www.sioc-project.org
[22] http://flickrtrends.appspot.com/
[23] http://trendistic.com/
[24] http://www.blogpulse.com/

fragments provides a means by which social data from heterogeneous sources can be interpreted consistently. Furthermore it allows end users to analyse social data based on its multi-faceted nature, something which is not currently possible using available trend services.

To provide an insight into how our approach functions, we have presented a web application which consumes social data following its metadata generation. The application is designed to exploit the dynamics of the data to allow the end user to delve deeper into the web of social data and discover anomalies, trends and idiosyncrasies which were not apparent from merely scratching the surface of the data. Such an application, in our view, acts as a proof of concept, by demonstrating the effects of interlinking social data. This work is the beginning of a study into the geographical facet of not only social data, but Linked Data[25] in general. The present implementation only focuses on static data. Our future work plans to use real-time visualisation of data based on its multi-faceted nature, thus allowing real-time anomalies to be identified as they occur.

6   Acknowledgements

References

1. T. Berners-Lee. Notation 3: A readable language for data on the web, March 2006.
2. J. G. Breslin, A. Harth, U. Bojars, and S. Decker. Towards semantically-interlinked online communities. In *2nd European Semantic Web Conference*, pages 500–514, 2005.
3. S. Choudhury, J. Breslin, and A. Passant. Enrichment and ranking of the youtube tag space and integration with the linked data cloud. In *Proceedings of the th International Semantic Web Conference (ISWC2009)*, LNCS. Springer, 10 2009.
4. T. Hastrup, U. Bojars, and J. Breslin. Sioclog: Providing irc discussion logs as linked data. In *Social Data on the Web (SDoW2009)*, 2009.
5. A. Passant, T. Hastrup, U. Bojars, and J. Breslin. Microblogging: A semantic web and distributed approach. In *4th Workshop on Scripting for the Semantic Web (SFSW2008)*, 2008.
6. B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *IEEE Visual Languages*, 1996.

---

[25] http://www.linkeddata.org