Disaster Monitoring with Wikipedia and Online Social Networking Sites: Structured Data and Linked Data Fragments to the Rescue?

Thomas Steiner*
Google Germany GmbH
ABC Str. 19
D-20355 Hamburg, Germany
tomac@google.com

Ruben Verborgh
Multimedia Lab – Ghent University – iMinds
Gaston Crommenlaan 8 bus 201
B-9050 Ledeberg-Ghent, Belgium
ruben.verborgh@ugent.be

Abstract

In this paper, we present the first results of our ongoing early-stage research on a realtime disaster detection and monitoring tool. Based on Wikipedia, it is language-agnostic and leverages user-generated multimedia content shared on online social networking sites to help disaster responders prioritize their efforts. We make the tool and its source code publicly available as we make progress on it. Furthermore, we strive to publish detected disasters and accompanying multimedia content following the Linked Data principles to facilitate its wide consumption, redistribution, and evaluation of its usefulness.

1 Introduction

1.1 Disaster Monitoring: A Global Challenge

According to a study (Laframboise and Loko 2012) published by the *International Monetary Fund* (IMF), about 700 disasters were registered worldwide between 2010 and 2012, affecting more than 450 million people. According to the study, "[d]amages have risen from an estimated US$20 billion on average per year in the 1990s to about US$100 billion per year during 2000–10." The authors expect this upward trend to continue "as a result of the rising concentration of people living in areas more exposed to disasters, and climate change." In consequence, disaster monitoring will become more and more crucial in the future.

National agencies like the *Federal Emergency Management Agency* (FEMA)[1] in the United States of America or the *Bundesamt für Bevölkerungsschutz und Katastrophenhilfe* (BBK,[2] "Federal Office of Civil Protection and Disaster Assistance") in Germany work to ensure the safety of the population on a national level, combining and providing relevant tasks and information in a single place. The *United Nations Office for the Coordination of Humanitarian Affairs* (OCHA)[3] is a United Nations (UN) body formed to strengthen the UN's response to complex emergencies and disasters. The *Global Disaster Alert and Coordination System* (GDACS)[4] is "a cooperation framework between the United Nations, the European Commission, and disaster managers worldwide to improve alerts, information exchange, and coordination in the first phase after major sudden-onset disasters." Global companies like Facebook,[5] Airbnb,[6] or Google[7] have dedicated crisis response teams that work on making critical emergency information accessible in times of disaster. As can be seen from the (incomprehensive) list above, disaster detection and response is a problem tackled on national, international, and global levels; both from the public and private sectors.

1.2 Hypotheses and Research Questions

In this paper, we present the first results of our ongoing early-stage research on a realtime comprehensive Wikipedia-based monitoring system for the detection of disasters around the globe. This system is *language-agnostic* and leverages *multimedia content* shared on online social networking sites, striving to help disaster responders prioritize their efforts. Structured data about detected disasters is made available in the form of Linked Data to facilitate its consumption. An earlier version of this paper without the focus on multimedia content from online social networking sites and Linked Data was published in (Steiner 2014b). For the present and further extended work, we are steered by the following hypotheses.

$\mathbb{H}1$ Content about disasters gets added very fast to Wikipedia and online social networking sites by people in the neighborhood of the event.

$\mathbb{H}2$ Disasters being geographically constrained, textual and multimedia content about them on Wikipedia and social networking sites appear first in local language, perhaps only later in English.

*Second affiliation: CNRS, Université de Lyon, LIRIS – UMR5205, Université Lyon 1, France

[1] FEMA: http://www.fema.gov/

[2] BBK: http://www.bbk.bund.de/

[3] OCHA: http://www.unocha.org/

[4] GDACS: http://www.gdacs.org/

[5] Facebook Disaster Relief: https://www.facebook.com/DisasterRelief

[6] Airbnb Disaster Response: https://www.airbnb.com/disaster-response

[7] Google Crisis Response: https://www.google.org/crisisresponse/

$\mathbb{H}3$ Link structure dynamics of Wikipedia provide for a meaningful way to detect future disasters, *i.e.*, disasters unknown at system creation time.

These hypotheses lead us to the following research questions that we strive to answer in the near future.

$\mathbb{Q}1$ How timely and accurate is content from Wikipedia and online social networking sites for the purpose of disaster detection and ongoing monitoring, compared to content from authoritative and government sources?

$\mathbb{Q}2$ To what extent can the disambiguated nature of Wikipedia (things identified by URIs) improve on keyword-based disaster detection approaches, *e.g.*, via online social network sites or search logs?

$\mathbb{Q}3$ How much noise is introduced by full-text searches (which are not based on disambiguated URIs) for multimedia content on online social networking sites?

The remainder of the article is structured as follows. First we discuss related work and enabling technologies in the next section, followed by our methodology in Section 3. We describe an evaluation strategy in Section 4, and finally conclude with an outlook on future work in Section 5.

2 Related Work and Enabling Technologies

2.1 Disaster Detection

Digitally crowdsourced data for disaster detection and response has gained momentum in recent years, as the Internet has proven resilient in times of crises, compared to other infrastructure. Ryan Falor, Crisis Response Product Manager at Google in 2011, remarks in (Falor 2011) that "a substantial [...] proportion of searches are directly related to the crises; and people continue to search and access information online even while traffic and search levels drop temporarily during and immediately following the crises." In the following, we provide a non-exhaustive list of related work on digitally crowdsourced disaster detection and response. Sakaki, Okazaki, and Matsuo (2010) consider each user of the online social networking (OSN) site Twitter[8] a sensor for the purpose of earthquake detection in Japan. Goodchild and Glennon (2010) show how crowdsourced geodata from Wikipedia and Wikimapia,[9] "a multilingual open-content collaborative map", can help complete authoritative data about disasters. Abel et al. (2012) describe a crisis monitoring system that extracts relevant content about known disasters from Twitter. Liu et al. (2008) examine common patterns and norms of disaster coverage on the photo sharing site Flickr.[10] Ortmann et al. (2011) propose to crowdsource Linked Open Data for disaster management and also provide a good overview on well-known crowdsourcing tools like Google Map Maker,[11] OpenStreetMap,[12] and Ushahidi (Okolloh 2009). We have developed a monitoring system (Steiner 2014c) that detects news events from concurrent Wikipedia edits and auto-generates related multimedia galleries based on content from various OSN sites and Wikimedia Commons.[13] Finally, Lin and Mishne (2012) examine realtime search query churn on Twitter, including in the context of disasters.

2.2 The *Common Alerting Protocol*

To facilitate collaboration, a common protocol is essential. The *Common Alerting Protocol* (CAP) (Westfall 2010) is an XML-based general data format for exchanging public warnings and emergencies between alerting technologies. CAP allows a warning message to be consistently disseminated simultaneously over many warning systems to many applications. The protocol increases warning effectiveness and simplifies the task of activating a warning for officials. CAP also provides the capability to include multimedia data, such as photos, maps, or videos. Alerts can be geographically targeted to a defined warning area. An exemplary flood warning CAP feed stemming from GDACS is shown in Listing 1. The step from trees to graphs can be taken through Linked Data, which we introduce in the next section.

2.3 Linked Data and Linked Data Principles

Linked Data (Berners-Lee 2006) defines a set of agreed-on best practices and principles for interconnecting and publishing structured data on the Web. It uses Web technologies like the Hypertext Transfer Protocol (HTTP, Fielding et al., 1999) and Unique Resource Identifiers (URIs, Berners-Lee, Fielding, and Masinter, 2005) to create typed links between different sources. The portal http://linkeddata.org/ defines Linked Data as being *"about using the Web to connect related data that wasn't previously linked, or using the Web to lower the barriers to linking data currently linked using other methods."* Tim Berners-Lee (2006) defined the four rules for Linked Data in a W3C Design Issue as follows:

1. Use URIs as names for things.
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL).
4. Include links to other URIs, so that they can discover more things.

Linked Data uses RDF (Klyne and Carroll 2004) to create typed links between things in the world. The result is oftentimes referred to as the *Web of Data*. RDF encodes statements about things in the form of `(subject, predicate, object)` triples. Heath and Bizer (2011) speak of *RDF links*.

2.4 Linked Data Fragments

Various access mechanisms to Linked Data exist on the Web, each of which comes with its own trade-offs regarding query performance, freshness of data, and server cost/availability. To retrieve information about a specific subject, you can dereference its URL. SPARQL endpoints allow to execute

[8] Twitter: https://twitter.com/

[9] Wikimapia: http://wikimapia.org/

[10] Flickr: https://www.flickr.com/

[11] Google Map Maker: http://www.google.com/mapmaker

[12] OpenStreetMap: http://www.openstreetmap.org/

[13] Wikimedia Commons: https://commons.wikimedia.org/

```
<alert xmlns="urn:oasis:names:tc:emergency:cap:1.2">
  <identifier>GDACS_FL_4159_1</identifier>
  <sender>info@gdacs.org</sender> <sent>2014-07-14T23:59:59-00:00</sent>
  <status>Actual</status> <msgType>Alert</msgType>
  <scope>Public</scope> <incidents>4159</incidents>
  <info>
    <category>Geo</category><event>Flood</event>
    <urgency>Past</urgency><severity>Moderate</severity>
    <certainty>Unknown</certainty>
    <senderName>Global Disaster Alert and Coordination System</senderName>
    <headline /><description />
    <web>http://www.gdacs.org/reports.aspx?eventype=FL&amp;amp;eventid=4159</web>
    <parameter><valueName>eventid</valueName><value>4159</value></parameter>
    <parameter><valueName>currentepisodeid</valueName><value>1</value></parameter>
    <parameter><valueName>glide</valueName><value /></parameter>
    <parameter><valueName>version</valueName><value>1</value></parameter>
    <parameter><valueName>fromdate</valueName><value>Wed, 21 May 2014 22:00:00 GMT</value></parameter>
    <parameter><valueName>todate</valueName><value>Mon, 14 Jul 2014 21:59:59 GMT</value></parameter>
    <parameter><valueName>eventtype</valueName><value>FL</value></parameter>
    <parameter><valueName>alertlevel</valueName><value>Green</value></parameter>
    <parameter><valueName>alerttype</valueName><value>automatic</value></parameter>
    <parameter><valueName>link</valueName><value>http://www.gdacs.org/report.aspx?eventtype=FL&amp;amp;eventid=4159</value></parameter>
    <parameter><valueName>country</valueName><value>Brazil</value></parameter>
    <parameter><valueName>eventname</valueName><value /></parameter>
    <parameter><valueName>severity</valueName><value>Magnitude 7.44</value></parameter>
    <parameter><valueName>population</valueName><value>0 killed and 0 displaced</value></parameter>
    <parameter><valueName>vulnerability</valueName><value /></parameter>
    <parameter><valueName>sourceid</valueName><value>DFO</value></parameter>
    <parameter><valueName>iso3</valueName><value /></parameter>
    <parameter>
      <valueName>hazardcomponents</valueName><value>FL,dead=0,displaced=0,main_cause=Heavy Rain,severity=2,sqkm=256564.57</value>
    </parameter>
    <parameter><valueName>datemodified</valueName><value>Mon, 01 Jan 0001 00:00:00 GMT</value></parameter>
    <area><areaDesc>Polygon</areaDesc><polygon>,,100</polygon></area>
  </info>
</alert>
```

Listing 1: *Common Alerting Protocol* feed via the *Global Disaster Alert and Coordination System* (http://www.gdacs.org/xml/gdacs_cap.xml, 2014-07-16)

complex queries on RDF data, but they are not always available. While endpoints are more convenient for clients, individual requests are considerably more expensive for servers. Alternatively, a data dump allows you to query locally.

Linked Data Fragments (Verborgh et al. 2014) provide a uniform view on all such possible interfaces to Linked Data, by describing each specific type of interface by the kind of *fragments* through which it allows access to the dataset. Each fragment consists of three parts:

data all triples of this dataset that match a specific selector;

metadata triples that describe the dataset and/or the Linked Data Fragment;

controls hypermedia links and/or forms that lead to other Linked Data Fragments.

This view allows to describe new interfaces with different trade-off combinations. One such interface is *triple pattern fragments* (Verborgh et al. 2014), which enables users to host Linked Data on low-cost servers with higher availability than public SPARQL endpoints. Such a light-weight mechanism is ideal to expose live disaster monitoring data.

3 Proposed Methodology

3.1 Leveraging Wikipedia Link Structure

Wikipedia is an international online encyclopedia currently available in 287 languages[14] with these characteristics:

1. Articles in one language are interlinked with versions of the same article in other languages, *e.g.*, the article "Natural disaster" on the English Wikipedia (http://en.wikipedia.org/wiki/Natural_disaster) links to 74 versions of this article in different languages.[15] We note that there exist similarities and differences among Wikipedias with "salient information" that is unique to each language as well as more widely shared facts (Bao et al. 2012).
2. Each article can have redirects, *i.e.*, alternative URLs that point to the article. For the English "Natural disaster" ar-

[14] All Wikipedias: http://meta.wikimedia.org/wiki/List_of_Wikipedias

[15] Article language links: http://en.wikipedia.org/w/api.php?action=query&prop=langlinks&lllimit=max&titles=Natural_disaster

ticle, there are eight redirects,[16] *e.g.*, "Natural Hazard" (synonym), "Examples of natural disaster" (refinement), or "Natural disasters" (plural).

3. For each article, the list of back links that link to the current article is available, *i.e.*, inbound links other than redirects. The article "Natural disaster" has more than 500 articles that link to it.[17] Likewise, the list of outbound links, *i.e.*, other articles that the current article links to, is available.[18]

By combining an article's in- and outbound links, we determine the set of mutual links, *i.e.*, the set of articles that the current article links to (outbound links) and at the same time receives links from (inbound links).

3.2 Identification of Wikipedia Articles for Monitoring

Starting with the well-curated English seed article "Natural disaster", we programmatically follow each of the therein contained links of type "Main article:", which leads to an exhaustive list of English articles of concrete types of disasters, *e.g.*, "Tsunami" (http://en.wikipedia.org/wiki/Tsunami), "Flood" (http://en.wikipedia.org/wiki/Flood), "Earthquake" (http://en.wikipedia.org/wiki/Earthquake), *etc.* In total, we obtain links to 20 English articles about different types of disasters.[19] For each of these English disasters articles, we obtain all versions of each article in different languages [step *(i)* above], and of the resulting list of international articles in turn all their redirect URLs [step *(ii)* above]. The intermediate result is a complete list of all (currently 1,270) articles in all Wikipedia languages and all their redirects that have any type of disaster as their subject. We call this list the "disasters list" and make it publicly available in different formats (`.txt`, `.tsv`, and `.json`), where the JSON version is the most flexible and recommended one.[20] Finally, we obtain for each of the 1,270 articles in the "disasters list" all their back links, *i.e.*, their inbound links [step *(iii)* above], which serves to detect instances of disasters unknown at system creation time. For example, the article "Typhoon Rammasun (2014)" (http://en.wikipedia.org/wiki/Typhoon_Rammasun_(2014))—which, as a concrete *instance of* a disaster of type tropical cyclone, is *not* contained in our "disasters list"—links back to "Tropical cyclone" (http://en.wikipedia.org/wiki/Tropical_cyclone), so we can identify "Typhoon Rammasun (2014)" as *related to* tropical cyclones (but not necessarily *identify as* a tropical cyclone), even if at the system's creation time the typhoon did not exist yet. Analog to the inbound links, we obtain all outbound links of all articles in the "disasters list", *e.g.*, "Tropical cyclone" has an outbound link to "2014 Pacific typhoon season" (http://en.wikipedia.org/wiki/2014_Pacific_typhoon_season), which also happens to be an inbound link of "Tropical cyclone", so we have detected a mutual, circular link structure. Figure 1 shows the example in its entirety, starting from the seed level, to the disaster type level, to the in-/outbound link level. The end result is a large list called the "monitoring list" of all articles in all Wikipedia languages that are somehow—via a redirect, inbound, or outbound link (or resulting mutual link)—related to any of the articles in the "disasters list". We make a snapshot of this dynamic "monitoring list" available for reference,[21] but note that it will be out-of-date soon and should be regenerated on a regular basis. The current version holds 141,001 different articles.

3.3 Monitoring Process

In the past, we have worked on a Server-Sent Events (SSE) API (Steiner 2014a) capable of monitoring realtime editing activity on all language versions of Wikipedia. This API allows us to easily analyze Wikipedia edits by reacting on events fired by the API. Whenever an edit event occurs, we check if it is for one of the articles on our "monitoring list". We keep track of the historic one-day-window editing activity for each article on the "monitoring list" including their versions in other languages, and, upon a sudden spike of editing activity, trigger an alert about a potential new instance of a disaster type that the spiking article is an inbound or outbound link of (or both). To illustrate this, if, *e.g.*, the German article "Pazifische Taifunsaison 2014" including all of its language links is spiking, we can infer that this is related to a disaster of type "Tropical cyclone" due to the detected mutual link structure mentioned earlier (Figure 1).

In order to detect spikes, we apply exponential smoothing to the last n edit intervals (we require $n \geq 5$) that occurred in the past 24 hours with a smoothing factor $\alpha = 0.5$. The therefore required edit events are retrieved programmatically via the Wikipedia API.[22] As a spike occurs when an edit interval gets "short enough" compared to historic editing activity, we report a spike whenever the latest edit interval is shorter than half a standard deviation $0.5 \times \sigma$.

A subset of all Wikipedia articles are geo-referenced,[23] so when we detect a spiking article, we try to obtain geo co-

[16] Article redirects: http://en.wikipedia.org/w/api.php?action=query&list=backlinks&blfilterredir=redirects&bllimit=max&bltitle=Natural_disaster

[17] Article inbound links: http://en.wikipedia.org/w/api.php?action=query&list=backlinks&bllimit=max&blnamespace=0&bltitle=Natural_disaster

[18] Article outbound links: http://en.wikipedia.org/w/api.php?action=query&prop=links&plnamespace=0&format=json&pllimit=max&titles=Natural_disaster

[19] "Avalanche", "Blizzard", "Cyclone", "Drought", "Earthquake", "Epidemic", "Extratropical cyclone", "Flood", "Gamma-ray burst", "Hail", "Heat wave", "Impact event", "Limnic eruption", "Meteorological disaster", "Solar flare", "Tornado", "Tropical cyclone", "Tsunami", "Volcanic eruption", "Wildfire"

[20] "Disasters list": https://github.com/tomayac/postdoc/blob/master/papers/comprehensive-wikipedia-monitoring-for-global-and-realtime-natural-disaster-detection/data/disasters-list.json

[21] "Monitoring list": https://github.com/tomayac/postdoc/blob/master/papers/comprehensive-wikipedia-monitoring-for-global-and-realtime-disaster-detection/data/monitoring-list.json

[22] Wikipedia last revisions: http://en.wikipedia.org/w/api.php?action=query&prop=revisions&rvlimit=6&rvprop=timestamp|user&titles=Typhoon_Rammasun_(2014)

[23] Article geo coordinates: http://en.wikipedia.org/w/api.php?action=query&prop=coordinates&format=json&colimit=

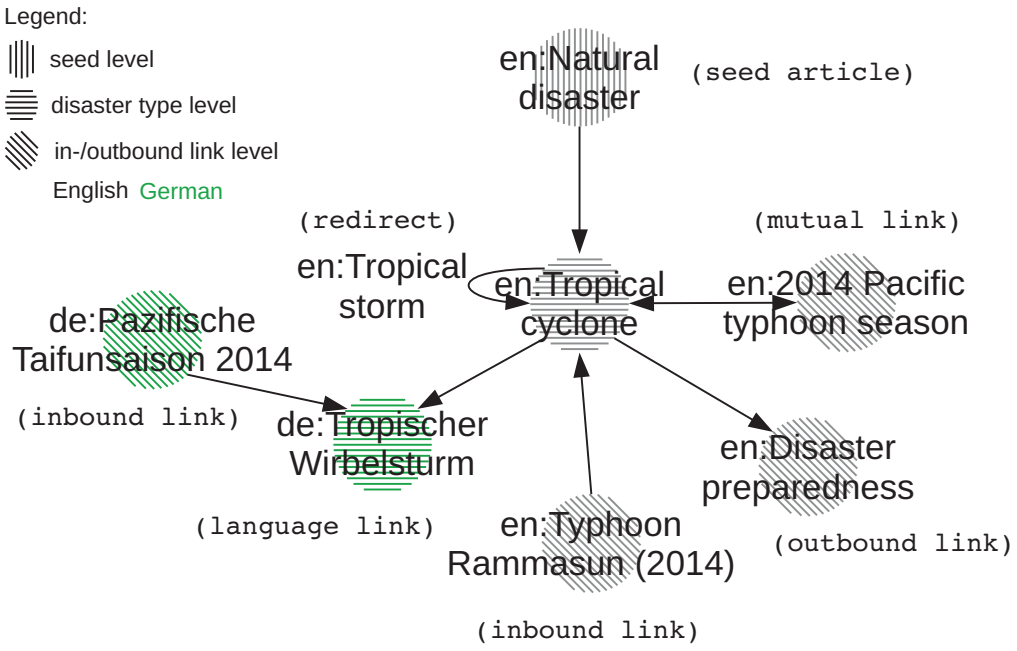

Figure 1: Extracted Wikipedia link structure (tiny excerpt) starting from the seed article "Natural disaster"

ordinates for the article itself (*e.g.*, "Pazifische Taifunsaison 2014") or any of its language links that—as a consequence of the assumption in ℍ2—may provide more local details (*e.g.*, "2014 Pacific typhoon season" in English or "2014年太平洋季" in Chinese). We then calculate the center point of all obtained latitude/longitude pairs.

3.4 Multimedia Content from Online Social Networking Sites

In the past, we have worked on an application called *Social Media Illustrator* (Steiner 2014c) that provides a social multimedia search framework that enables searching for and extraction of multimedia data from the online social networking sites Google+,[24] Facebook,[25] Twitter,[26] Instagram,[27] YouTube,[28] Flickr,[29] MobyPicture,[30] TwitPic,[31] and Wikimedia Commons.[32] In a first step, it deduplicates exact- and near-duplicate social multimedia data based on a previously describe algorithm (Steiner et al. 2013). It then ranks social multimedia data by social signals (Steiner 2014c) based on an abstraction layer on top of the online social networking sites mentioned above and, in a final step, allows for the creation of media galleries following aesthetic principles (Steiner 2014c) of the two kinds *Strict Order, Equal Size* and *Loose Order, Varying Size*, defined in (Steiner 2014c). We have ported crucial parts of the code of *Social Media Illustrator* from the client-side to the server-side, enabling us now to create media galleries at scale and on demand, based on the titles of spiking Wikipedia articles that are used as separate search terms for each language. The social media content therefore does not have to link to Wikipedia. One exemplary media gallery can be seen in Figure 2, each individual media item in the gallery is clickable and links back to the original post on the particular online social networking site, allowing crisis responders to monitor the media gallery as a whole, and to investigate interesting media items at the source and potentially get in contact with the originator.

3.5 Linked Data Publication

In a final step, once a given confidence threshold has been reached and upon human inspection, we plan to send out a notification according to the *Common Alerting Protocol* following the format that (for GDACS) can be seen in Listing 1. While *Common Alerting Protocol* messages are generally well understood, additional synergies can be unlocked by leveraging Linked Data sources like DBpedia, Wikidata, and Freebase, and interlinking them with detected potentially relevant multimedia data from online social networking sites. Listing 2 shows an early-stage proposal for doing so. The alerts can be exposed as triple pattern fragments to enable live querying at low cost. This can also include push, pull, and streaming models, as Linked Data Fragments (Verborgh et al. 2014) allow for all. A further approach consists in converting CAP messages to Linked Data by transform-

max&coprop=dim|country|region|globe&coprimary=all&titles=September_11_attacks

[24]Google+: https://plus.google.com/

[25]Facebook: https://www.facebook.com/

[26]Twitter: https://twitter.com/

[27]Instagram: http://instagram.com/

[28]YouTube: http://www.youtube.com/

[29]Flickr: http://www.flickr.com/

[30]MobyPicture: http://www.mobypicture.com/

[31]TwitPic: http://twitpic.com/

[32]Wikimedia Commons: http://commons.wikimedia.org/wiki/Main_Page

ing the CAP eXtensible Markup Language (XML) format to Resource Description Format (RDF) and publishing it.

3.6 Implementation Details

We have created a publicly available prototypal demo application deployed[33] at http://disaster-monitor.herokuapp.com/ that internally connects to the SSE API from (Steiner 2014a). It is implemented in Node.js on the server, and as a JavaScript Web application on the client. This application uses an hourly refreshed version of the "monitoring list" from Section 3.2 and whenever an edit event sent through the SSE API matches any of the articles in the list, it checks if, given this article's and its language links' edit history of the past 24 hours, the current edit event shows spiking behavior, as outlined in Section 3.3. The core source code of the monitoring loop can be seen in Section 3, a screenshot of the application is shown in Figure 2.

4 Proposed Steps Toward an Evaluation

We recall our core research questions that were ℚ1 *How timely and accurate for the purpose of disaster detection and ongoing monitoring is content from Wikipedia, compared to authoritative sources mentioned above?* and ℚ2 *Does the disambiguated nature of Wikipedia surpass keyword-based disaster detection approaches,* e.g.*, via online social networking sites or search logs?* Regarding ℚ1, only a manual comparison covering several months worth of disaster data of the relevant authoritative data sources mentioned in Sec-

[33] Source code: https://github.com/tomayac/postdoc/tree/master/demos/disaster-monitor

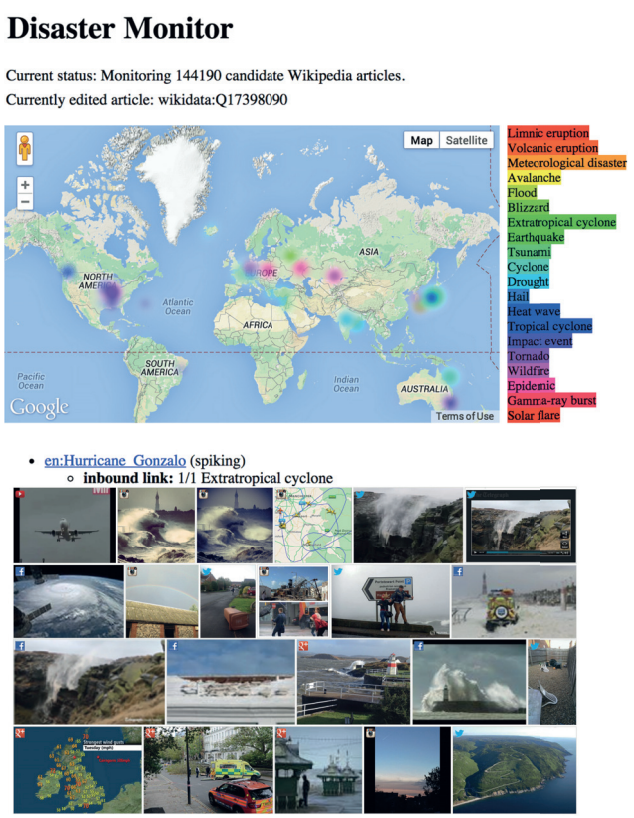

Figure 2: Screenshot of the *Disaster Monitor* application prototype available at http://disaster-monitor.herokuapp.com/ showing detected past disasters on a heatmap and a media gallery for a currently spiking disaster around "Hurricane Gonzalo"

```
(function() {
  // fired whenever an edit event happens on any Wikipedia
  var parseWikipediaEdit = function(data) {
    var article = data.language + ':' + data.article;
    var disasterObj = monitoringList[article];
    // the article is on the monitoring list
    if (disasterObj) {
      showCandidateArticle(data.article, data.language,
          disasterObj);
    }
  };

  // fired whenever an article is on the monitoring list
  var showCandidateArticle = function(article, language, roles) {
    getGeoData(article, language, function(err, geoData) {
      getRevisionsData(article, language, function(err,
          revisionsData) {
        if (revisionsData.spiking) {
          // spiking article
        }
        if (geoData.averageCoordinates.lat) {
          // geo-referenced article, create map
        }
        // trigger alert if article is spiking
      });
    });
  };

  getMonitoringList(seedArticle, function(err, data) {
    // get the initial monitoring list
    if (err) return console.log('Error␣initializing␣the␣app.');
    monitoringList = data;
    console.log('Monitoring␣' + Object.keys(monitoringList).length
        + '␣candidate␣Wikipedia␣articles.');

    // start monitoring process once we have a monitoring list
    var wikiSource = new EventSource(wikipediaEdits);
    wikiSource.addEventListener('message', function(e) {
      return parseWikipediaEdit(JSON.parse(e.data));
    });

    // auto-refresh monitoring list every hour
    setInterval(function() {
      getMonitoringList(seedArticle, function(err, data) {
        if (err) return console.log('Error␣refreshing␣monitoring␣
            list.');
        monitoringList = data;
        console.log('Monitoring␣' + Object.keys(monitoringList).
            length +
            '␣candidate␣Wikipedia␣articles.');
      });
    }, 1000 * 60 * 60);
  });
})();
```

Listing 3: Monitoring loop of the disaster monitor

```
<http://ex.org/disaster/en:Hurricane_Gonzalo> owl:sameAs "http://en.wikipedia.org/wiki/Hurricane_Gonzalo",
                                                         "http://live.dbpedia.org/page/Hurricane_Gonzalo",
                                                         "http://www.freebase.com/m/0123kcg5";
                                              ex:relatedMediaItems _:video1;
                                              ex:relatedMediaItems _:photo1;
_:video1 ex:mediaUrl "https://mtc.cdn.vine.co/r/videos/827962270911343031733232517l2_2ca88ba5444.5.1.1669873818247419980 4.mp4";
         ex:micropostUrl "http://twitter.com/gpessoao/status/527603540860997632";
         ex:posterUrl "https://v.cdn.vine.co/r/thumbs/231E0009CF1134303174572797952_2.5.1.1669873818247419980 4.mp4.jpg";
         ex:publicationDate "2014-10-30T03:15:01Z";
         ex:socialInteractions [ ex:likes 1; ex:shares 0 ];
         ex:timestamp 1414638901000;
         ex:type "video";
         ex:userProfileUrl "http://twitter.com/alejandroriano";
         ex:micropost [
             ex:html "Here's Hurricane #Gonzalo as seen from the @Space_Station as it orbited above today https://t.co/RpJt0P2bXa";
             ex:plainText "Here's Hurricane Gonzalo as seen from the Space Station as it orbited above today" ].
_:photo1 ex:mediaUrl "https://upload.wikimedia.org/wikipedia/commons/b/bb/Schiffsanleger_Wittenbergen_-_Orkan_Gonzalo.jpg";
         ex:micropostUrl "https://commons.wikimedia.org/wiki/File:Schiffsanleger_Wittenbergen_-_Orkan_Gonzalo_(22.10.2014)_01.jpg";
         ex:posterUrl "https://upload.wikimedia.org/wikipedia/commons/thumb/b/bb/Schiffsanleger_Wittenbergen_-_Orkan_Gonzalo_
              %2822.10.2014%29_01.jpg/500px-Schiffsanleger_Wittenbergen_-_Orkan_Gonzalo_(22.10.2014)_01.jpg" .
         ex:publicationDate "2014-10-24T08:40:16Z";
         ex:socialInteractions [ ex:shares 0 ];
         ex:timestamp 1414140016000;
         ex:type "photo";
         ex:userProfileUrl "https://commons.wikimedia.org/wiki/User:Huhu_Uet";
         ex:micropost [
             ex:html "Schiffsanleger_Wittenbergen_-_Orkan_Gonzalo_(22.10.2014)_01";
             ex:plainText "Schiffsanleger_Wittenbergen_-_Orkan_Gonzalo_(22.10.2014)_01" ].
```

Listing 2: Exemplary Linked Data for *Hurricane Gonzalo* using a yet to-be-defined vocabulary (potentially HXL http://hxl.humanitarianresponse.info/ns/index.html or MOAC http://observedchange.com/moac/ns/) that interlinks the disaster with several other Linked Data sources and relates it to multimedia content on online social networking sites

tion 1.1 with the output of our system can help respond to the question. Regarding $\mathbb{Q}2$, we propose an evaluation strategy for the OSN site Twitter, loosely inspired by the approach of Sakaki *et al.* in (Sakaki, Okazaki, and Matsuo 2010). We choose Twitter as a data source due to the publicly available user data through its streaming APIs,[34] which would be considerably harder, if not impossible, with other OSNs or search logs due to privacy concerns and API limitations. Based on the articles in the "monitoring list", we put forward using article titles as search terms, but without disambiguation hints in parentheses, *e.g.*, instead of the complete article title "Typhoon Rammasun (2014)", we suggest using "Typhoon Rammasun" alone. We advise monitoring the sample stream[35] for the appearance of any of the search terms, as the filtered stream[36] is too limited regarding the number of supported search terms. In order to avoid ambiguity issues with the international multi-language tweet stream, we recommend matching search terms only if the Twitter-detected tweet language equals the search term's language, *e.g.*, English, as in "Typhoon Rammasun".

[34]Twitter streaming APIs: https://dev.twitter.com/docs/streaming-apis/streams/public

[35]Twitter sample stream: https://dev.twitter.com/docs/api/1.1/get/statuses/sample

[36]Twitter filtered stream: https://dev.twitter.com/docs/api/1.1/post/statuses/filter

5 Conclusions and Future Work

In this paper, we have presented the first steps of our ongoing research on the creation of a Wikipedia-based disaster monitoring system. In particular, we finished its underlying code scaffolding and connected the system to several online social networking sites allowing for the automatic generation of media galleries. Further, we propose to publish data about detected and monitored disasters as live queryable Linked Data, which can be made accessible in a scalable and *ad hoc* manner using triple pattern fragments (Verborgh et al. 2014) by leveraging free cloud hosting offers (Matteis and Verborgh 2014). While the system itself already functions, a good chunk of work still lies ahead with the fine-tuning of its parameters. A first examples are the exponential smoothing parameters of the revision intervals, responsible for determining whether an article is spiking, and thus a potential new disaster, or not. A second example is the role that disasters play with articles: they can be inbound, outbound, or mutual links, and their importance for actual occurrences of disasters will vary. Future work will mainly focus on finding answers to our research questions $\mathbb{Q}1$ and $\mathbb{Q}2$ and the verification of the hypotheses $\mathbb{H}1$–$\mathbb{H}3$. We will focus on the evaluation of the system's usefulness, accuracy, and timeliness in comparison to other keyword-based approaches. An interesting aspect of our work is that the monitoring system is not limited to disasters. Using an analogous approach, we can monitor for human-made disasters (called "Anthropogenic

hazard" on Wikipedia) like terrorism, war, power outages, air disasters, *etc.* We have created an exemplary "monitoring list" and made it available.[37]

Concluding, we are excited about this research and look forward to putting the final system into operational practice in the weeks and months to come. Be safe!

References

Abel, F.; Hauff, C.; Houben, G.-J.; Stronkman, R.; and Tao, K. 2012. Twitcident: Fighting Fire with Information from Social Web Streams. In *Proceedings of the 21st International Conference Companion on World Wide Web*, WWW '12 Companion, 305–308. New York, NY, USA: ACM.

Bao, P.; Hecht, B.; Carton, S.; Quaderi, M.; Horn, M.; and Gergle, D. 2012. Omnipedia: Bridging the wikipedia language gap. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, 1075–1084. New York, NY, USA: ACM.

Berners-Lee, T.; Fielding, R. T.; and Masinter, L. 2005. Uniform Resource Identifier (URI): Generic Syntax. RFC 3986, IETF.

Berners-Lee, T. 2006. Linked Data. http://www.w3.org/DesignIssues/LinkedData.html.

Falor, R. 2011. Search data reveals people turn to the internet in crises. http://blog.google.org/2011/08/search-data-reveals-people-turn-to.html.

Fielding, R.; Gettys, J.; Mogul, J.; Frystyk, H.; Masinter, L.; Leach, P.; and Berners-Lee, T. 1999. Hypertext Transfer Protocol – HTTP/1.1. RFC 2616, IETF.

Goodchild, M. F., and Glennon, J. A. 2010. Crowdsourcing Geographic Information for Disaster Response: A Research Frontier. *International Journal of Digital Earth* 3(3):231–241.

Heath, T., and Bizer, C. 2011. *Linked Data: Evolving the Web into a Global Data Space*. Synthesis Lectures on the Semantic Web: Theory and Technology. Morgan & Claypool.

Klyne, G., and Carroll, J. J. 2004. Resource Description Framework (RDF): Concepts and Abstract Syntax. Recommendation, W3C.

Laframboise, N., and Loko, B. 2012. Natural Disasters: Mitigating Impact, Managing Risks. IMF Working Paper, International Monetary Fund. http://www.imf.org/external/pubs/ft/wp/2012/wp12245.pdf.

Lin, J., and Mishne, G. 2012. A Study of "Churn" in Tweets and Real-Time Search Queries (Extended Version). *CoRR* abs/1205.6855.

Liu, S. B.; Palen, L.; Sutton, J.; Hughes, A. L.; and Vieweg, S. 2008. In Search of the Bigger Picture: The Emergent Role of On-line Photo Sharing in Times of Disaster. In *Proceedings of the Information Systems for Crisis Response and Management Conference (ISCRAM)*.

[37] Anthropogenic hazard "monitoring list": https://github.com/tomayac/postdoc/blob/master/papers/comprehensive-wikipedia-monitoring-for-global-and-realtime-disaster-detection/data/monitoring-list-anthropogenic-hazard.json

Matteis, L., and Verborgh, R. 2014. Hosting queryable and highly available Linked Data for free. In *Proceedings of the ISWC Developers Workshop 2014, co-located with the 13th International Semantic Web Conference (ISWC 2014), Riva del Garda, Italy, October 19, 2014.*, 13–18.

Okolloh, O. 2009. Ushahidi, or "testimony": Web 2.0 Tools for Crowdsourcing Crisis Information. *Participatory learning and action* 59(1):65–70.

Ortmann, J.; Limbu, M.; Wang, D.; and Kauppinen, T. 2011. Crowdsourcing Linked Open Data for Disaster Management. In Grütter, R.; Kolas, D.; Koubarakis, M.; and Pfoser, D., eds., *Proceedings of the Terra Cognita Workshop on Foundations, Technologies and Applications of the Geospatial Web, In conjunction with the International Semantic Web Conference (ISWC2011)*, volume 798, 11–22. Bonn, Germany: CEUR Workshop Proceedings.

Sakaki, T.; Okazaki, M.; and Matsuo, Y. 2010. Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, 851–860. New York, NY, USA: ACM.

Steiner, T.; Verborgh, R.; Gabarro, J.; Mannens, E.; and Van de Walle, R. 2013. Clustering Media Items Stemming from Multiple Social Networks. *The Computer Journal*.

Steiner, T. 2014a. Bots vs. Wikipedians, Anons vs. Logged-Ins (Redux): A Global Study of Edit Activity on Wikipedia and Wikidata. In *Proceedings of The International Symposium on Open Collaboration*, OpenSym '14, 25:1–25:7. New York, NY, USA: ACM.

Steiner, T. 2014b. Comprehensive Wikipedia monitoring for global and realtime natural disaster detection. In *Proceedings of the ISWC Developers Workshop 2014, co-located with the 13th International Semantic Web Conference (ISWC 2014), Riva del Garda, Italy, October 19, 2014.*, 86–95.

Steiner, T. 2014c. *Enriching Unstructured Media Content About Events to Enable Semi-Automated Summaries, Compilations, and Improved Search by Leveraging Social Networks*. Ph.D. Dissertation, Universitat Politècnica de Catalunya.

Verborgh, R.; Hartig, O.; De Meester, B.; Haesendonck, G.; De Vocht, L.; Vander Sande, M.; Cyganiak, R.; Colpaert, P.; Mannens, E.; and Van de Walle, R. 2014. Querying datasets on the web with high availability. In Mika, P.; Tudorache, T.; Bernstein, A.; Welty, C.; Knoblock, C.; Vrandečić, D.; Groth, P.; Noy, N.; Janowicz, K.; and Goble, C., eds., *The Semantic Web – ISWC 2014*, volume 8796 of *Lecture Notes in Computer Science*. Springer International Publishing. 180–196.

Westfall, J. 2010. Common Alerting Protocol Version 1.2. Standard, OASIS. http://docs.oasis-open.org/emergency/cap/v1.2/CAP-v1.2-os.doc.