

A Social Semantic Web based Conceptual Architecture of Disaster Trail Management System

Ashfaq Ahmad^{1,2}, Roslina Othman¹, Mohamad Fauzan¹

¹Kulliyyah of Information and Communication Technology
International Islamic University Malaysia
Kuala Lumpur, Malaysia

²Faculty of Computer Science & Information Systems
Jazan University
Jazan, Saudi Arabia

Abstract—Disasters affect human lives severely. Due to these disasters, hundreds and thousands of human beings lost their lives and gracious properties. Government agencies, non-government organization and individual volunteers act to rescue the affected people and to mitigate the disaster effects. These teams require real time information about the nature, severity, area and number of affectees. Their efforts can be supported by providing timely, effective and specific information so that the rescuers can get better idea about the available routes to reach the affectees, urgency and mass of loss. People share huge amount of data through blogs and social media that can be utilized to help rescue operations. This information can electronically be filtered, arranged and formatted in a proper manner. Thus, semantic web technologies can play a vital role in providing timeliness information. Purpose of this research is to capture explicit knowledge of the domain in form of ontologies, automatic information extraction, generation of implicit knowledge and then disseminating this information to various stakeholders. Collection of implicit and explicit knowledge will help improve decision making for disaster trail management.

Keywords—Ontology; Disaster Trail Management; Information Extraction; Knowledge Management

I. INTRODUCTION

The word ‘disaster’ comes from ancient Greek words ‘dis’ means bad and ‘aster’ means star. The astrological sense of this word bases on calamity blamed on star positions. [1]

Disasters can be categorized mainly into two types; natural and man-made. Furthermore, according to their impact, these can be subdivided into major and minor disasters. Natural major disasters include Cyclones, Tornadoes, Drought, Flood, Earthquake and Tsunami whereas Epidemic, Deforestation, Chemical Pollution, Wars and Terrorism can be considered as man-made major disasters. Similarly, Cold/Heat Waves, Thunderstorms are natural minor disasters but Road/Train Accident, Riots, Food Poisoning and Industrial mishaps can be treated as minor man-made disasters. [2, 3]

Pakistan lies on the Indo-Australian plate, which is propelling northward and edging under the Eurasian plate. This specific geographic location of Pakistan results in frequent earthquake happenings in this area [5]. Since Pakistan lies in a seismic belt therefore, it is at a risk of earthquakes of high magnitude. Mountain ranges of the northern areas, Khyber

Pakhtunkhwa and Baluchistan provinces, and important cities like Islamabad, Karachi and Peshawar are located on the edges of high risk areas [6]. Pakistan is prone to disasters and has experienced a variety of major natural disasters including earthquakes and floods. These disasters cause large number of life losses, internally displaced persons (IDPs), loss of crops and agricultural lands, infra-structure and property damages. [7, 8]

Disasters always have rigorous effects on human lives [4]. Due to these disasters, hundreds and thousands of human beings lost their lives and their valued properties. Government agencies, non-government organization and individual volunteers act to rescue the affected people and to mitigate the disaster effects. These teams require real time information about the nature, severity and area of disaster and number of affectees. Their efforts can be supported by providing timely, effective and specific information so that the rescuers can get better idea about the available routes to reach the affectees, urgency and mass of loss.

Disaster management efforts affect due to multiple barriers. The most prominent factors lacking in some areas are community awareness, effective communication among the concerned departments, deployed human resources and budgeting, technology awareness and utilization and competency of management. [9]

A number of efforts have made in the field of Information and Communication Technology (ICT) to assist rescue operations by providing required information effectively and timely. The major task of ICT solutions is managing data starting from data gathering, organizing, integrating, analyzing and then generating information for decision making. Semantic Web technology can play better role in improving capability and usefulness of ICT solutions for Disaster and Trail Management using standard and mature semantic languages and tools like Web Ontology Language (OWL) [10] and General Architecture of Text Engineering (GATE) [11]. A number of ICT solutions are in local use in different countries whereas a few of them benefiting many countries. SAHANA [12] and Disaster Management Information System (DMIS) [13] are effective and the most widely used ICT solutions which have been used in various countries for natural disasters management. But these systems rely on manual data entry. In

disaster management, availability of latest data facts improves the efficiency and effectiveness of rescue operations.

People share huge amount of data through blogs and social media that can be utilized (along with the manual data entry) to help rescue operations. This information can be electronically filtered, arranged and formatted in a proper manner. Thus, semantic web technologies can play a vital role in providing timeliness information. People generally share information via social media not only in English but also in their native languages. If the focus is to grasp information in English only, then a huge amount of useful information (in native language) can be missing. Urdu is the national language of Pakistan. So, majority of the population prefer to use Urdu or 'Roman Urdu' to share information on social media. Purpose of this research is to develop a system to effectively capture explicit knowledge of the domain of the disaster whether shared in English or Urdu in the form of ontologies and disseminating this information to various stakeholders.

The paper is organized in different sections, each elaborates a particular segment of discussion. Selective projects, closely related to the proposed system, are discussed with their strengths and weaknesses in section II. Last paragraph of this section briefly describes the limitations of the projects to highlight the research gap. Section III provides an overview of the proposed system whereas tools and techniques required for system development are discussed in section IV. Section V identifies the potential research challenges. Possible solutions to these challenges have also been suggested in the same section. Conclusion is given in section VI.

II. LITRATURE REVIEW

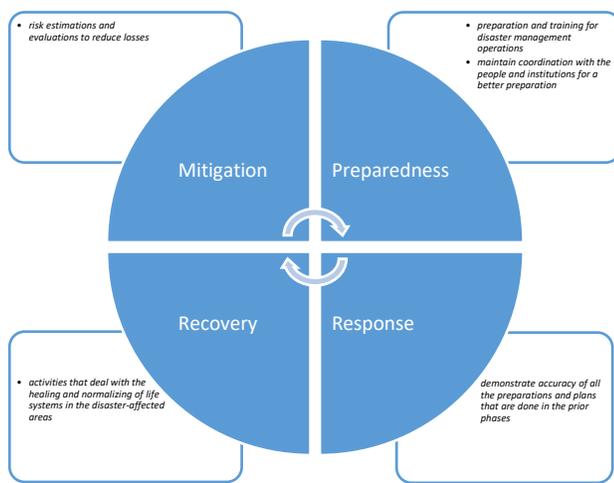


Fig. 1. The Disaster Management Cycle

Fig. 1 is self-explanatory and explains an overview of disaster management cycle with a brief introduction to phases of disaster management and activities to perform during each phase. The proposed system is related to *Recovery* Phase.

Disasters cannot be avoided but efforts can be made to minimize their effects. Disaster management is so important that although a lot of work has already been done, yet more

work is required to be done to rescue the affected people. ICT is playing a constructive and vital role in this regard. Following is a brief view of ICT efforts.

A. Organizational Solution - National Level Disaster Risk Management Framework of Pakistan

Realizing the need of disaster management, Pakistan has established a National Disaster Management Commission (NDMC). NDMC, headed by the Prime Minister of Pakistan, is the national policy making body for managing disasters and executes its policies and manages Disaster Risk Management (DRM) activities through National Disaster Management Authority (NDMA). At provincial level there exist Provincial Disaster Management Commissions (PDMCs) responsible to formulate provincial level policies and plans regarding DRM. PDMC act through Provincial Disaster Management Authorities (PDMAs) which further share responsibilities at grass root level by District Disaster Management Authorities (DDMAs). Effectiveness and scope of these agencies majorly depends on their position within national government, size, allocated budget, and staff training. NDMA, is the Federal level agency which supervises all Disaster Management Activities in Pakistan. For mitigation during disaster, all stakeholders take part in rescue operations as a unit of NDMA. [7, 14]

B. Conventional Disaster Management Systems

However, efficiency and effectiveness of disaster management activities can be improved by using ICT, the most adopting and growing technology. ICT is contributing its predominant part in increasing efficiency and accuracy in almost all walks of life. It is also playing a vital role in boosting operations of all phases of disaster management cycle. Various ICT based applications and services are available for disaster management. Among these, SAHANA EDEN is the most popular disaster management system that has been deployed by over 50 nonprofit and government agencies in different countries including USA, UK, France, Italy, Japan, India, Pakistan, Sri Lanka, Indonesia, Philippines, Bangladesh, Thailand, China, Australia, Taiwan and New Zealand.

1) SAHANA

SAHANA [12] is an open source disaster management platform that helps organizations and communities prepare for and respond to disasters. The word 'Sahana' means relief in Sinhalese, one of the national languages of Sri Lanka [12]. The software was originally developed by members of the Sri Lankan IT community in 2004 for Indian Ocean Tsunami. SAHANA is an open source software project, the development of which is contributed by a large number of volunteer programmers belonging to various countries of the world. It was successfully deployed in over 25 countries including Pakistan for 2005 earthquake and 2010 floods. Its free and open source format, excellent structure and well defined modules made it the most widely accepted and popular disaster management software.

The main modules of SAHANA EDEN include:

- Organization Registry – allows an organization to register itself in the SAHANA database and links it to

other modules like Human Resources, Assets and Inventory

- Project Tracking – provides situational data to and coordination among organizations
- Human Resources – keeps track of engaged staff and volunteers
- Inventory – manages inventory and donations
- Assets – keeps track of assets and their utilization
- Shelter Management – temporary shelters, required resources and shelters population information management
- Scenarios and Events – planning effective deployments of resources
- Mapping – provides situational awareness via maps
- Messaging – group messaging facility via email, SMS, Twitter and Google Talk. Receiving short queries via interactive messaging and automatic responses [12]

Sahana Eden, which is developed in Python, can be installed on any environment like Windows or OSX, which can run Python. This project includes organization registry module to help coordinate and collaborate with organizations taking part in helping affected people. Other strengths include its multi-lingual interface and its interoperability with other applications and datasets. Sahana available translation languages include Chinese, Hindi, Spanish, English, Arabic, Portuguese, Russian, German, Tamil, Burmese, Indonesian and Sinhalese [15]. During deployment in Pakistan, some weaknesses were observed in SAHANA including lack of locations related data which is the baseline data, non-availability of Urdu translation and poor SMS infrastructure. Above all, the key weakness is that it requires manual data entry [16]. The laborious task of manual data entry is time taking and less feasible in case of an emergency.

2) Disaster Management Information System (DMIS)

DMIS [13] was launched by the “Civil Society for Research & Initiatives for Sustainable Technologies and Institutions (SRISTI)”, on January 2002 in India. DMIS shares its services and data voluntarily with NGOs and especially to civil society institutions across the world. DMIS provides a platform for documenting and sharing disaster related experiences which can be utilized in developing strategies for future disaster management. It also provides coordination among individuals and disaster rescue organizations. It also provides an online database to share experiences and volunteer services but it does not provide any facility to decision makers for retrieving and processing real-time data.

C. Disaster Ontologies

The Semantic Web Community is contributing in addressing the disaster management issue by developing ontologies. Following are some notable disaster management ontologies:

In 2015, Zavarella Vanni et al. [17] developed a multilingual event extraction engine for information extraction

from online print media sources. It filtered correct information via the Linked Open Data principles. They outlined a procedure which propagates the information to larger scope ontologies for Crisis Management. The base of the procedure is ontology lexicalization method for text mining from social media that does not depend on language. The team also brought a keyword based query into work to get the tweets, related to events, for language normalization but this work needs more to be done. They need some validation event data in order to measure the performance of their system.

Chen-Huei Chou et al. [18] developed an ontology structure of elements for “web-based disaster management systems (WB-DMS)” in 2011. This structure identifies web elements using the grounded-theory approach. In this, selected semi-structured data representation approaches are used to organize the resulting ontology structure. Next the ontology structure is coded into a Web-based system. This ontology structure provides a valuable guideline in creating new WB-DMS web sites. This work can be enriched on bases of opinions of domain experts and analyzing online forums and actual documents generated for crises management. Also the work can be extended to other areas like business crises management.

In 2014 Juan Li et al. [19] developed a social semantic search engine for fetching emergency related information from twitter and also proposed Emergency Information System for disaster position identification. However, the focus is on authoritative information from emergency related organizations. Tool is designed to fetch disaster related information from twitter but in future they aim to integrate other social networking sites. The authors say that trusted courses of information can be preferably ranked. Prior to storing, they emphasized on information validation by format checking.

In 2014, Z. Ashktorab et al. [20] developed a twitter-mining tool to extract meaningful and actionable information for disaster relief workers during a disastrous situation. The ontology focuses only on twitter and using a small set of labelled data to extract relevant data. In future, the authors will focus on exploration of conflicting labels and determining more accurate labels.

In 2015, F. Morstatter et al. [21] introduced social media posts analysis techniques based on location information to help first responders provide relief to disaster affectees. They discussed machine learning techniques application in location finding. This team presented an algorithm that helps find users in the region affected by a crisis. Their future work consists of finding the “event location” and location privacy.

The above-discussed efforts have their own limitations. Most of the work done so far focuses on particular aspects of disaster management domain. The ICT solutions lack in locations related data for Pakistan, which is the baseline data. Above all, the key weakness is the dependency of these solutions on laborious task of manual data entry which is not feasible during emergency situations. Moreover, the constructed ontologies do not target to integrate with existing disaster management systems. Also, the information extraction from messages shared in Urdu is completely missing. Urdu

being the national language of Pakistan, is also used in sharing information on social media during and after the disaster. Ignoring the information shared in Urdu may cause missing a lot of information.

III. PROPOSED SYSTEM

Fig. 2, describes the proposed social semantic web based disaster trail management system. By collecting nearly real time data from social media, the system adequately solves the problems discussed in the previous sections. The system will be capable to integrate with existing disaster management systems. It will enhance the working effectiveness of these DMSs by providing latest situational data related to disaster.

Focus of research is to design a disaster trail management system for Pakistan. While designing the conceptual architecture of the system, it is kept in mind that people of Pakistan also use their national language to share information on social media. The proposed model can however, easily be generalized by replacing the translator. The proposed system will only capture textual information.

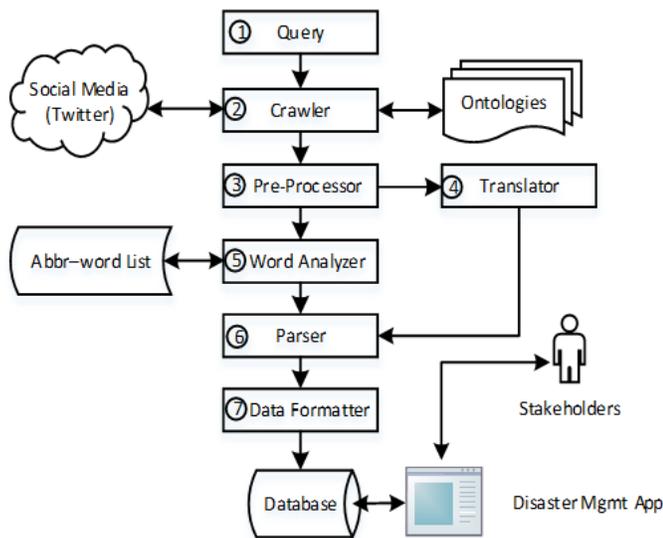


Fig. 2. Conceptual Architecture of Social Semantic Web based Disaster Trail Management System

The conceptual architecture of the proposed system has been illustrated in Fig. 2. It also clearly mentions data flow between various system components. Following is the description of the system's functionality.

Query is passed in the form of disaster related keyword(s). These keywords either can be provided by the user as command line arguments or can be stored in a text file that can be used as input file. Query is then passes to crawler.

1) The crawler performs tweets capturing from the social network, Twitter. Only text based tweets written in English or Urdu language are captured and irrelevant tweets will be ignored. Crawler makes use of semantic web ontologies for crawling the most relevant data. These filters (keyword and language) are required to use by crawler to collect only

relevant set of text from a huge stream of data shared on twitter.

2) The crawled tweets are then forwarded to pre-processor. As the objective is to capture maximum information shared by Pakistani people, that's why all concerned tweets are collected whether written in English or in the national language Urdu. Information extraction is comparatively easy on text written in English language. This module identifies the tweets written in Urdu language and forwards these Urdu Tweets to Translator. The text captured in English is forwarded to Word analyzer.

3) Translator translates Urdu text into English language and forward translated text to parser.

4) Due to restriction of tweet's length, people generally use abbreviations in their tweets. Word Analyzer identifies abbreviations and replaces them with respective complete words getting from Abbr.-Word List and then passes the document to parser.

5) Parser plays an important role in separating data and metadata (if any). Information Extraction is done by this module. Key entities are identified and duplicate entities are suppressed. Ultimately, the processed data set is forwarded to data formatter for dumping it into the database.

6) Data Formatter formats and stores data in accordance with the defined database rules. These data form a knowledge base that can be utilized in decision making process and by the disaster trail management software.

IV. TOOLS AND TECHNIQUES

The system can be subdivided into the following major sections:

- Tweets Collection
- Translation of Urdu to English
- Processing and Information Extraction
- Data storing for future use

Section 1 needs to collect textual data shared by public on social media, Twitter, which is a huge stream of data shared by people. Twitter provides streaming APIs which will be used for the data collection purpose. The objective is to collect only the relevant tweets excluding multimedia tweets. From this huge data world, to focus on relevant tweets, filters will be set for the purposes like searching keywords and setting of language only to collect tweets in English and Urdu language. Twitter has limitation of message length, that's why people use abbreviations, which increases complexity of information extraction process. For example, acdnt can be used instead of accident or abt instead of about. To make the collected messages more meaningful, it is required to identify and replace abbreviations with complete words.

The proposed model can be customized by translator. The focus is to translate tweets written into Urdu language to English. A translator like Google [22], Yandex [23], Bing [24] or SDL [25] can be embedded by using API.

Information Extraction (IE) is backbone of the system. For this purpose, Natural Language Processing (NLP) tools like GATE [11] or Stanford CoreNLP [26] can be used. GATE is provided with an IE pipeline ANNIE (A Nearly-New IE) System [11]. ANNIE relies on JAPE (Java Annotation Pattern Engine) language [11]. For Named Entity Recognition, the available ANNIE gazetteer lists will not be enough. Additional gazetteer lists will be required to develop, to make it more specific to disaster. TwitIE is a customized open-source IE of GATE ANNIE which is customized for Twitter [27].

Generally, disaster management systems have to store a huge amount of data. Instead of using traditional Relational Database Management Systems (RDBMS), NoSQL databases like MongoDB [28] are preferred due to more flexibility, performance and scalability. MongoDB is a free and open-source cross-platform document-oriented database. As the ultimate objective of data storage is to utilize this data by existing disaster management systems, so, NoSQL database will be used.

V. CHALLENGES AND SUGGESTED SOLUTIONS

All the major sections of the proposed system have their own issues and research challenges. First of all, relevant data collection from data floating on public data stream is a challenge. Shared data may be textual or multimedia. Each and every data cannot be collected and analyzed. So, the crawler should be intelligent enough to collect only relevant text data. For this purpose, the crawler should have been provided with reasonable filters.

The next challenge to cope with is the use of self-generated acronyms or short forms. They hinder the way of understanding the gist of the message. People do this to save their time and space for the message as some social media like twitter has limit for message length. People use a lot of unregistered language words to keep their message short and to the limits. There is no universal standard format to write the words in short forms. Some short form words are commonly used but not all. People define their own abbreviations to short. This makes their message short and complete but quite unintelligible for many readers. For example, people may use “f9” instead of writing “fine” just to save two letters space. Some efforts have been made to document most frequently used short forms, which can be used to prepare abbr.-word list (consisting abbreviations and corresponding complete word).

The system aims to collect the tweets not only in English but also written in Urdu, the national language of Pakistan. So, in the translation phase, it is required to embed a translator like Google [22], Yandex [23], Bing [24] or SDL [25].

The IE phase of the system is quite complicated and is the backbone of the system. Entities and their relationships are required to be identified. As the target is to collect disaster related information, so, the general purpose tools are needed to enhance their functionality above the recognition of general entities. In this section, Information will be extracted from the collected data. NLP tools like GATE or Stanford CoreNLP will play its role in identifying entities and their relationships. The general purpose NLP tools will provide general entities which are not able to solve the purpose. For specialized disaster

domain related entities, it requires to introduce another layer of exactitude, to be used by the tools for specific and targeted information extraction.

The final issue is to store extracted information in such a form that can easily be integrated with the existing disaster management systems. For this purpose, a schema-less or NoSQL open source database can be a good choice. MongoDB can be preferred for its scalability, performance and high availability [28].

VI. CONCLUSION

The proposed social semantic web based conceptual architecture of disaster trail management system has been elaborated with all its necessary components, their interaction and data flow. While considering the potential challenges and research issues, their possible solutions along with the tools that can be used in materializing the proposed system has also been suggested. The authors would develop the disaster trail management based on the purposed architecture and present as their future work.

REFERENCES

- [1] English Words of (Unexpected) Greek Origin, (2011), <https://ewonago.wordpress.com/2011/02/19/845/>
- [2] Raza and S. Haq, (2015), National Disaster Management Authority in Pakistan: Role of Pakistan Army in Disaster Management, *Journal of Social and Administrative Sciences*, vol. 2, issue 1
- [3] V. de Walle and T. Comes, (2015), On the Nature of Information Management in Complex and Natural Disasters, *Procedia Engineering*, Volume 107, Pages 403-411
- [4] N. F. Noy and D. L. McGuinness, (2001), *Ontology Development 101: A Guide to Creating Your First Ontology*, Knowledge Systems, AI Laboratory, Stanford University (KSL-01-05)
- [5] H. Ullah and G. Gungor, (2014), Comparative analysis of immediate response by national disaster management systems of the U.S., Pakistan, and Turkey, DTIC Document
- [6] NDMA Annual Report 2011, Disaster Management System in Pakistan, www.ndma.gov.pk.
- [7] H. Khan and A. Khan, (2008), Natural hazards and disaster management in Pakistan, Munich Personal RePEc Archive
- [8] S. A. Sayed and P. A. González, (2014), Flood disaster profile of Pakistan: A review, *Science Journal of Public Health*, 2(3): 144-149
- [9] F. Shafiq and K. Ahsan, (2014), An ICT based Early Warning System for Flood Disasters in Pakistan, *Research Journal of Recent Sciences*, ISSN 2277-2502, Vol. 3(9), 108-118
- [10] S. Bechhofer, (2009), *Encyclopedia of Database Systems*, Springer
- [11] General Architecture for Text engineering (GATE), <http://gate.ac.uk>
- [12] SAHANA Software Foundation, <http://sahanafoundation.org>
- [13] SRISTI: Disaster Management Information System, http://www.sristi.org/dmis/dmi_system
- [14] National Disaster Management Authority (NDMA): Government of Pakistan, <http://www.ndma.gov.pk>
- [15] <http://www.slideshare.net/TalkSahana/sahana-internationalisation-languages-and-beyond>
- [16] F. Boon and M. Howden, (2010), *Sahana: Case Studies*, Sahana Camp
- [17] Zavarella Vanni, Tanev Hristo, Steinberger Ralf, Van Der Goot Erik, (2015), An Ontology-Based Approach to Social Media Mining for Crisis Management, *CEUR Workshop Proceedings (CEUR-WS.org)*, ISSN: 1613-0073
- [18] Chen-Huei Chou, Fatemeh Mariam Zahedi, Huimin Zhao, (2011), Ontology for Developing Web Sites for Natural Disaster Management: Methodology and Implementation, *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, Volume: 41, Issue: 1, Jan. 2011

- [19] J. Li, Q. Li, A. Ginjala, N. Zaman, (2014), eSMS - a Semantics-assisted Emergency Information System Based on Social Media, *IJ. Information Technology and Computer Science*, 08, 18-24
- [20] Z. Ashktorab et al., (2014), Tweedr: Mining Twitter to Inform Disaster Response, *iscram.org*
- [21] F. Morstatter, H. Gao and H. Liu, (2015), Discovering Location Information in Social Media, *public.asu.edu*
- [22] Google Cloud Translation API <https://cloud.google.com/translate>
- [23] Yandex Translate API <https://tech.yandex.com/translate>
- [24] Microsoft Translator <http://www.bing.com/dev/en-us/translator>
- [25] SDL Language Cloud [https://languagecloud.sdl.com/translation-](https://languagecloud.sdl.com/translation-toolkit/api-documentation)
[toolkit/api-documentation](https://languagecloud.sdl.com/translation-toolkit/api-documentation)
- [26] Manning, Christopher D., Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky, (2014), The Stanford CoreNLP Natural Language Processing Toolkit, *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55-60.
- [27] K Bontcheva, L Derczynski, A Funk, MA. Greenwood, D Maynard, N Aswani, (2013), TwitIE: An Open-Source Information Extraction Pipeline for Microblog Text, In *Proceedings of the International Conference on Recent Advances in Natural Language Processing, RANLP*, pages 83–90, Hissar, Bulgaria, 2013. INCOMA Ltd.
- [28] MongoDB, <https://www.mongodb.com>