

Crowd-sourcing tools within the PREPARE analytical platform

A. Ikonomopoulos¹ and S. Konstantopoulos²

¹ Institute of Nuclear and Radiological Sciences and Technology, Energy and Safety, National Center for Scientific Research “Demokritos”, Agia Paraskevi 15310, Greece.

² Institute of Informatics and Telecommunications, National Center for Scientific Research “Demokritos”, Agia Paraskevi 15310, Greece.

Abstract – A brief overview of the crowd-sourcing tools developed within the analytical platform of the PREPARE project is presented. Crowd sourcing relates to methods that: offer a Web Content Discovery Service that automatically acquires and analyses content from social networks and the Web to extract information during nuclear or radiological (NR) events as well as the public opinion and perspective regarding an event occurrence and development; support an Ask the Expert Service for the public to have access to information relevant to NR events. Crowd sourcing provides the means to enhance the relevance and quality of publicly available information that is to be distributed through the analytical platform. To that end, processes have been formulated to collect: Web content relevant to a NR event and analyse it for trend identification regarding public opinions and concerns expressed in Web forums. Large-scale content acquisition and text analytics for Web-forum monitoring within social networks provides results at a scale that cannot be manually achieved; feedback from the user interaction with the system and analyse it to identify documents of interest and relevance to the queries posed.

Keywords: nuclear or radiological emergency / Web crawling / PREPARE analytical platform

1 Introduction

In the General Safety Requirements Part 7 issued in the IAEA Safety Standards Series (IAEA, 2015) it is stated under Requirement 10 that: “...The government shall ensure that arrangements are in place to provide the public who are affected or are potentially affected by a nuclear or radiological emergency with information that is necessary for their protection...” “...The information shall be provided in the languages mainly spoken by the population residing within the emergency planning zones and emergency planning distances” and under Requirement 13 that: “Arrangements shall be made so that in a nuclear or radiological emergency information is provided to the public in plain and understandable language” “...Arrangements shall be made to identify and address, to the extent practicable, misconceptions, rumours and incorrect and misleading information that might be circulating widely in a nuclear or radiological emergency...”

This context highlights the relevance of the PREPARE crowd-sourcing tools. The PREPARE project developed methods and tools for an analytical platform, a collection and analysis center of information related to nuclear or radiological (NR) events. The platform aims to become the European point of reference for information related to NR events, their consequences and future development. Within this infrastructure, experts will have the means to collect, evaluate and make available information related to an NR-event development. Crowd-sourcing tools have been integrated within the PREPARE analytical platform to perform information aggregation from Web sources and facilitate the

retrieval of relevant and reliable information by the public. Fundamental requirement for the tool operation is the retrieval of information without expert knowledge on NR events.

The following sections present an introduction of the foundation technologies employed in the work described, followed by an overview of the crowd-sourcing functionality. The conclusions drawn from this work and future steps that could be taken are discussed in the last section.

2 Background

Semantic Web technologies and ontologies are specialized tools for tackling the issues encountered when navigating through, searching and analyzing information. Ontologies define and organize the terminological items used within a domain of discourse and axiomatize the way in which such terms can be meaningfully used. This has found several applications where it is important to abstract away from varying orthographies and alternative or descriptive ways of referring to the same entity, so that a single abstract entity can be linked to alternative wordings such as ‘nuclear power’, ‘atomic energy’ and ‘energy generated from nuclear fission’. Furthermore, ontological relations between entities establish the context within which text processing operates, so that in a ‘radioactive material transport’ context the term ‘containment system’ will be more likely disambiguated as ‘an assembly of packaging components’ rather than as ‘the physical barrier of a nuclear installation’ (IAEA, 2016).

The work performed in the framework of developing crowd-sourcing tools for the analytical platform is based on the

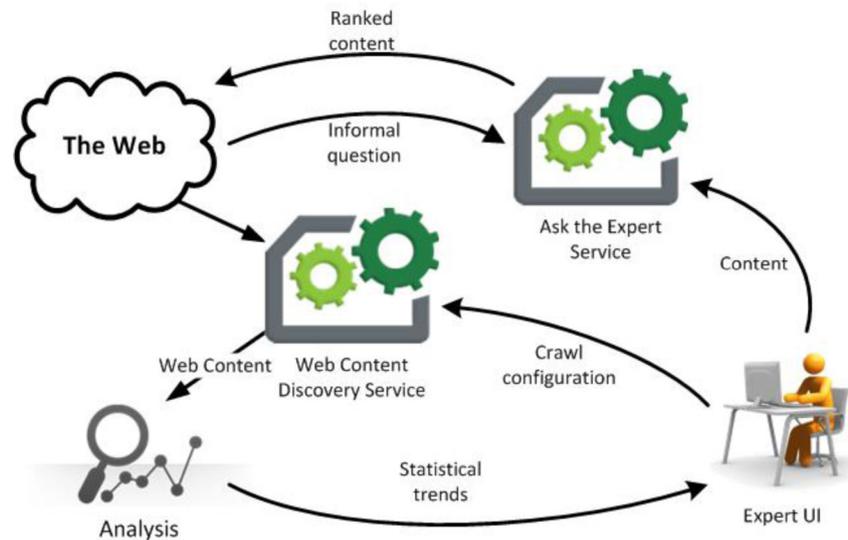


Figure 1. Overview of the PREPARE crowd-sourcing tools.

Nuclear or Radiological Emergency Ontology (NREO), a subject-specific ontology for the NR emergency field (Konstantopoulos and Ikonopoulos, 2015). Besides establishing the abstract model of the NR emergency field, Konstantopoulos and Ikonopoulos (2015) have also transferred three relevant glossaries (IAEA, 2007; Rojas-Palma *et al.*, 2009; Nisbet *et al.*, 2010) into NREO.

3 The PREPARE crowd-sourcing tools

The crowd-sourcing tools within the analytical platform have been designed to collect information from Web sources and manage personalized communication services of reliable information to the public. More specifically, the PREPARE Web portal during normal operation will be a source of reliable information that can be navigated, searched and understood without requiring specialized knowledge. A Web portal that will become known to the public as a source of accurate information will provide invaluable services during an NR event offering accurate information relevant to each visitor queries as opposed to broadcasting general guidelines. In this framework, NREO is used to index Web content and documents in a database that allows *semantic searching* that is, understanding a searcher's intent and the meaning of terms in the database to fetch more relevant results than keyword string matching.

The crowd-sourcing tools are organized in two services along with the corresponding user interfaces: the *Web Content Discovery Service* and the *Ask the Expert Service*. These services are integrated in the PREPARE analytical platform developed for the *Liferay Web Platform*.³ The *Web Content Discovery Service* crawls public forums, social media and, in general, the publicly accessible Web for content and subsequently tags it with NREO terms. During, or in the aftermath of an event, this content is used to collect real-time information about it. Outside of an event, the experts who

administer the Web portal employ statistical analysis regarding the occurrence and co-occurrence of NREO terms over time to identify emerging trends in the concerns and issues discussed by the public. The experts use these insights in order to maintain a collection of relevant, publicly available documents that can be read by non-experts. The *Ask the Expert Service* offers semantic searching functionality over this collection, exploiting NREO to fetch relevant results to queries expressed ambiguously or in everyday language rather than using formal terms. Figure 1 gives a graphical overview of the PREPARE crowd-sourcing tools.

The *Web Content Discovery Service* extends and adapts the Web-crawling platform developed in the context of the NOMAD project.⁴ The NOMAD platform implements Web-crawling services, search-engine clients and feed listeners as well as further processing services such as extracting clean text from HTML pages. Within the PREPARE project, the NOMAD platform is employed to retrieve content from the BingTM search engine as well as the Google+TM and TwitterTM social networks.⁵ The *Web Content Discovery Service* extends the NOMAD platform with functionality for: (a) browsing NREO terms to select crawling terms, (b) using NREO to expand the selected crawling terms with related terms and alternative linguistic realizations of these terms and (c) visualizing the frequency of term occurrence and co-occurrence in the retrieved documents.

The *Ask the Expert Service* is based on the document management service of the PREPARE analytical platform through which the expert administrator can add and annotate documents with formal terms. The *Ask the Expert Service* matches the annotations in the documents against queries in order to present content. This matching uses NREO to estimate

⁴ For more details, please cf. <http://www.nomad-project.eu>.

⁵ For more details, please cf. <https://datamarket.azure.com/dataset/bing/search> about BingTM, <https://developers.google.com/+web/api/rest/> about Google+TM and <https://dev.twitter.com/overview/documentation> about TwitterTM.

³ For more details, please cf. <https://www.liferay.com>.

the relevance of content to a query, so that relevant content can be recommended even if it does not match the exact search keywords. Furthermore, the system maintains usage logs of the user-selected document among those recommended as well as the time spent viewing a document before selecting another one.

4 Conclusions and future work

In the aftermath of the Fukushima event that proved, once again, the need for provision of accurate information to the public, the PREPARE analytical platform has been endowed with crowd-sourcing functionality. In this framework, the *Web Content Discovery Service* and the *Ask the Expert Service* have been developed and integrated. These services have been designed to index Web content in a flexible way that supports its accurate formal characterization despite the potential misalignment between the formal terminology used by the domain experts and the informal lay terms used by the public (or the informal usage of formal terms). This supports a content life cycle where experts are able to identify trends in the public concerns, provide relevant information and channel such information through a querying interface that does not require specialized terminological knowledge.

Further work involves automatically identifying situations where irrelevant content has been recommended and not selected – or only briefly examined and immediately rejected – by the user. These failure examples may be processed in order to fine-tune the relevance parameters and improve the relevance of the recommended content. Another direction would be to explore the mining of crawled Web data for term co-occurrences that indicate semantic relationships not currently captured by NREO. Naturally, NREO updates need to be validated by experts but such a mining tool can assist

in identifying newly emerged lay expressions that refer to concepts related to NR emergencies.

Acknowledgment. The research leading to these results has received funding from the European Atomic Energy Community Seventh Framework Programme FP7/2012-2013 under grant agreement 323287.

References

- IAEA (2007) Safety Glossary: Terminology Used in Nuclear Safety and Radiation Protection. International Atomic Energy Agency, Vienna.
- IAEA (2015) Preparedness and Response for a Nuclear or Radiological Emergency: General Safety Requirements. International Atomic Energy Agency, Vienna.
- IAEA (2016) Safety Glossary: Terminology Used in Nuclear Safety and Radiation Protection. International Atomic Energy Agency, Vienna. Draft 2016 Revision, <http://www-ns.iaea.org/standards/safety-glossary.asp>.
- Konstantopoulos S., Ikonopoulos A. (2015) A conceptualization of a nuclear or radiological emergency, *Nucl. Eng. Des.* **284**, 192-206.
- Nisbet A.F., Brown J., Cibanca T., Jones A.L., Andersson K.G., Hänninen R., Ikäheimonen T., Kirchner G., Bertsch V., Heite M. (2010) Generic Handbook for Assisting in the Management of Contaminated Inhabited Areas in Europe Following a Radiological Emergency, Version 2. Deliverable D18C1R4, European Commission Project EURANOS.
- Rojas-Palma C., Liland A., Jerstad A.N., Etherington G., del Rosario Pérez M., Rahola T., Smith K., Eds. (2009) *TMT Handbook: Triage, Monitoring and Treatment of People Exposed to Ionising Radiation following a Malevolent Act*. Deliverable, European Commission Project TMT Handbook.

Cite this article as: A. Ikonopoulos, S. Konstantopoulos. Crowd-sourcing tools within the PREPARE analytical platform. *Radioprotection* 51(HS2), S187-S189 (2016).