Equator: Faster Decision Making for Geoscientists

Lars Döhling, Heiko Woith, Dirk Fahland, Ulf Leser
(doehling|leser)@informatik.hu-berlin.de,
heiko.woith@gfz-potsdam.de, d.fahland@tue.nl

Abstract: Earthquakes form a permanent threat to humankind, causing death and injuries each year. For this reason, studying earthquakes has become an important area of research in geoscience. Responding to an earthquake event, for mitigation or for research purposes, demands quick decision processes which require timely and precise situational information. In this paper, we present Equator, a web-based content management system that automatically collects, integrates, and visualizes earthquake related information from various agencies. Equator was developed to meet specific requirements of the German Task Force Earthquake. It is in productive use by the Task Force since 2008 and proved to significantly reduce time required for acquiring situational awareness.

1 Introduction

Natural events like earthquakes form a permanent threat to humankind, causing death and injuries each year. Understanding their formation and action will likely help to reduce their disastrous effects in the future. Therefore studying earthquakes is an important area of research in geoscience. One group of scientists dealing with earthquakes is the German Task Force for Earthquakes[1] located at GFZ Potsdam.

Founded in 1993, the major purpose of the Task Force is to coordinate the allocation of an interdisciplinary scientific-technical expert team after catastrophic earthquakes worldwide. The Task Force is activated immediately after a major earthquake has occurred, collects necessary information about the event and its location, decides about, and prepares for an on-site mission. In case of a mission, Task Force members begin collecting and analyzing data in the disaster area as soon as possible. This includes seismic data of aftershocks, post-seismic deformation, hydro geological data, and the damage distribution as well as structural conditions of buildings. Their findings support local decisions and provide a scientific basis for an improved intermediate and long term mitigation of earthquake effects, and improve existing theories and their application [FW08].

As analyzing earthquakes heavily depends on being on-site promptly, the decision about a Task Force mission needs to be reached within 10 to 14 hours after the event, based on precise and timely situational information. This information is

[1] http://www.gfz-potsdam.de/portal/gfz/Struktur/Departments/Department+2/sec21/TaskForce?locale=en

provided by several data sources and comes in various forms such as texts, maps or diagrams. Prior to this case study, the information was pulled manually from respective data sources and managed as a pile of printouts, which has several drawbacks:

- It neither allows distributed access nor searching.
- Relations between data sources printed out on different sheets cannot be expressed or get lost.
- It is not possible to connect further electronic resources, e.g. PDF files or web pages, although most of them are already electronic.
- Keeping information up-to-date is hard as they change over time, which is a typical characteristic of crisis events.

In this paper, we present Equator (EarthQUake dAta collecTOR), a web-based content management system (CMS) developed to overcome some of the described drawbacks. Besides basic CMS functionalities (i.e. creating, reading, updating and deleting content elements) Equator (1) supports automatic collecting of information from four earthquake agencies, and (2) generates an integrated view on earthquake facts on Google Maps. The four agencies constitute vital information sources and are recurrently polled at each event. By automating this step, Equator reduces the time required for acquiring situational awareness significantly.

2 Equator

Equator stores a list of earthquake articles. Each article describes one event. It consists of a list of paragraphs that are all rendered on one page as shown in Fig. 1. Each paragraph contains information of one data source, consisting of text, images, and links, all represented in HTML.

2.1 User-Friendly Content Editing and Versioning

The content of a paragraph is maintained by an integrated WYSIWYG HTML editor, offering a word-processing-like interface and functionality to users. The editor supports content copy & paste from other web pages (i.e. data sources). As web sites replace or remove linked content over time, the editor's back end implements a deep copying of web page data. Additionally, uploaded documents or web links can be attached to a paragraph.

As knowledge about an earthquake event evolves, a data source may update its information. For this reason, Equator manages several versions of each paragraph in an article, reflecting updates of the data source. The most recent version is shown by default, other versions are accessible via menu controls. A printer-friendly view on the latest versions of an earthquake article can be exported to PDF.

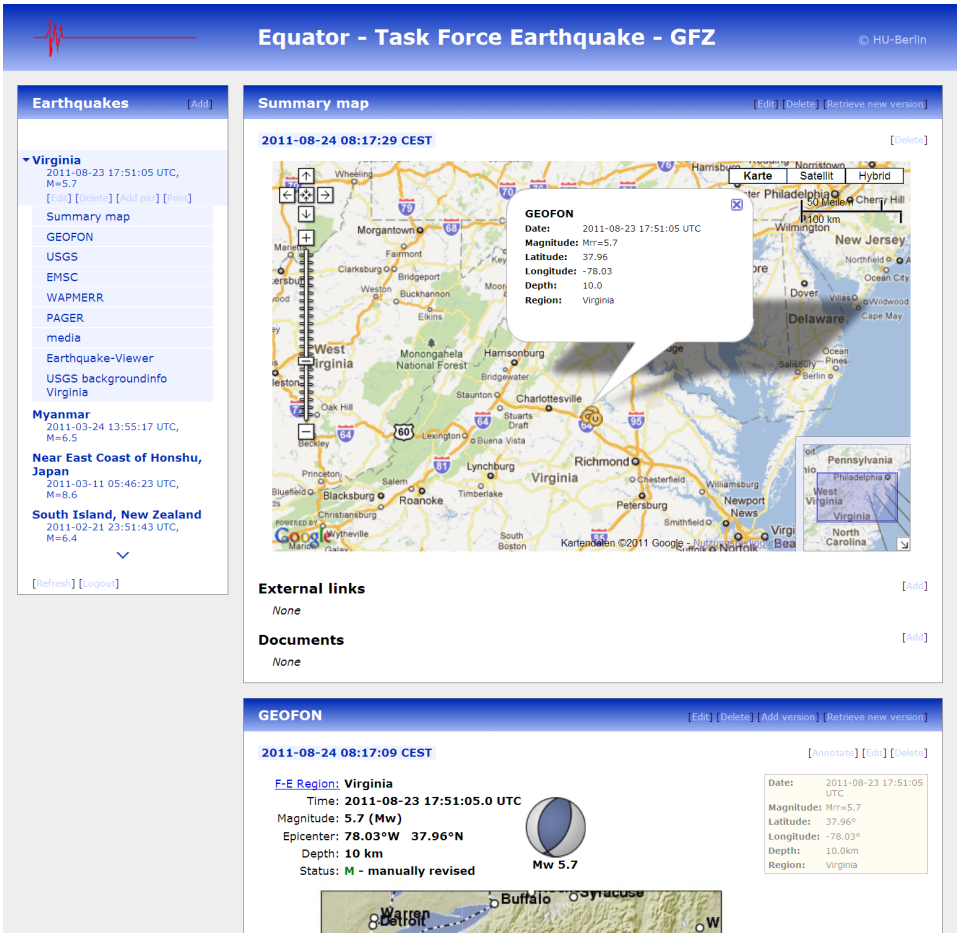

Figure 1: Equator screenshot showing the data collected for the August 2011 Virginia earthquake. The navigation bar (left) contains the list of articles and their data source paragraphs. The maps visualizes the facts extracted automatically from the earthquake agencies, each linked to the reported epicenter respectively.

2.2 Automatic Article Creation

As time is critical and many earthquake information sources provide data on simple HTML pages whose structure remain stable over time, Equator supports semi-automatic earthquake article creation, executed as follows:

1. A user selects an earthquake event from a list captured from GEOFON[2].
2. Equator determines matching reports at three other earthquake agencies

[2] `http://geofon.gfz-potsdam.de/geofon`, primary information source on seismic events for the Task Force Earthquakes.

(USGS[3], EMSC[4], WAPMERR[5]) via rule-based near-duplicate detection. Because each agency uses their own array of seismographs, their reported measures of date, location, etc. differ for identical events. Thus, no real inter-source duplicates exist, requiring near-duplicate detection. For this, we defined a similarity score between earthquake events, based on the differences in time, location, magnitude, depth and region name. To determine the near duplicate, we calculate the similarity score of the selected GEOFON event with all events of the other agencies, which occurred within a limited time frame. We consider the best scoring event whose score is higher than a particular threshold as a match. This simple near-duplicate detection algorithm performs quite well in practice, as natural events like earthquakes can be distinguished perfectly by combining time and location.

3. Equator deep copies the contained data from all four agencies into a new earthquake article applying Web scraping techniques. The copy process is controlled by a set of XPath expressions matching the desired parts of the web page. As this set is web-site-specific, it may require adjustments in case of structural changes.
4. Additionally, it extracts core facts from each source, i.e. date, location, magnitude and depth, and generates an integrated view on Google Maps (Figure 1). This extraction process is based on regular expressions which may require adjustments, if the format used changes.

3 Conclusions and Future Work

In this paper, we presented Equator, a custom CMS to simplify collecting and managing information about earthquake events. It was developed to meet the specific requirements of the Task Force Earthquakes where it is in productive use since 2008. Equator proved its effectiveness, and is constantly receiving further development [DL11].

References

[DL11] Lars Döhling and Ulf Leser. EquatorNLP: Pattern-based Information Extraction for Disaster Response. In *Proceedings of Terra Cognita 2011 Workshop*, 2011. (to appear).

[FW08] Dirk Fahland and Heiko Woith. Towards Process Models for Disaster Response. In *PM4HDPS'08*, volume 17 of *LNBIP*, pages 254–265, Milan, Italy, September 2008. Springer.

[3] `http://earthquake.usgs.gov/earthquakes`
[4] `http://www.emsc-csem.org`
[5] `http://www.wapmerr.org`