Method for the creation of databases of events having a mediatic echo in the internet

Eventi di frana nel 2014 intercettati sulla rete internet

Patent Status

- Italian patent application submitted
- International patent application submitted
- Ongoing evaluation.
Abstract:
The invention concerns a method to create databases of events that have an echo in the media and Internet. In particular hazardous geological events such as landslides, earthquakes, floods are updated constantly and automatically. The method collect the data with several information such as location, dating, pertinence and intensity.

The Technology and its advantage
Geohazards are a common cause of casualties and economical losses therefore, significant efforts are always ongoing to mitigate related risks. When studying geohazards, especially at a regional or a national scale, the availability of archives and geodatabases, which can provide information about past and recent events (including magnitude, timing and location), is of primary importance. In particular, the availability of updated and complete inventories is of paramount importance for hazard and risk assessment and for the creation of early warning models. Unfortunately, one of the primary limitations of the existing archives and databases (especially for landslides and floods) is their update rate and methodology. They are usually compiled manually using field surveys and, sometimes, remote sensing. Systems using automated or real-time updates are still uncommon and only used for a few types of geohazards. The data mining methodologies allow to extract with analytical techniques, information implicit, hidden, from data already structured. However, methods that apply such techniques to obtain specific information from geological events risky are far from trivial. The same technology that permit us to scan Internet and acquire news via feeds is improved to allows cataloging with characteristic parameters, including at least the location of the event, the date of the event, the intensity of the event and the inherence to the topic. These parameters are determined by data mining techniques. The event is geo-located even in the absence of placenames in the text, using alternative procedures of search, looking for of the location of the news broadcaster, looking for adjectives, geographical indications or equivalencies not directly expressible as a placename. The entire process takes place in real time thanks to continuous scans of the Internet.

Database creation via data mining of online news bulletins
Examples of data mining for geolocation (left) and rejecting news using pertinence scoring (right)

**Market opportunity**

The main scope is the creation of structured databases of events to use in prevention and civil protection purposes. This database can be used in several scientific researches, using the characteristics associated with each events, in particular position and intensity. From the databases can be obtained maps representing the occurrence of an event on a particular date. This can be used for statistical analysis aimed at identifying the relationship between other phenomena measured, such as the relationship between rainfall and landslides. This allows to evolve from simple feedforward forecasting systems to forecasting systems that use a feedback loop. The algorithm used is also scalable and extendable to the study of other phenomena that have a mediatic echo in the internet, scientific or not. The procedure can be also viewed as an engine to geolocate news. This make it attractive to web bulletins that wish to provide location-based information, taking advantage of the increased interest towards near news.

From feedforward forecasting to forecasting with feedback loop
In five years the system has already collected about 18,000 news of floods, 25,000 news of landslides and 15,000 news of earthquakes.

Inventors

Alessandro Battistini, Nicola Casagli, Sandro Moretti

Contacts

Ufficio brevetti proprietà intellettuale Unifi
brevetti@csavri.unifi.it