SPATIAL SEARCH

RICHARD GÖBEL EXPLAINS HOW THE RETRIEVAL OF SECURITY RELATED DATA FROM THE INTERNET VIA SPATIAL SEARCH ENGINES BECOMES MORE EFFECTIVE AS GEOGRAPHY IS INCLUDED.

Today the Internet is the starting point for finding information on almost every topic. Also security applications increasingly use the Internet to collect data for different themes. In many cases articles from international, national and local newspapers are of interest since these articles provide useful information about recent events even from remote places.

Traditional search engines like Google do not provide optimal support in this application context. The key problem is the missing support for the geographical dimension. This is already true for the definition of the search criteria. Although place or region names may be used in a search, the search engine does not recognise these words as geographical locations. As an example a search for "terror attacks" in "Iraq" will retrieve all documents containing the phrase "terror attack" and the word "Iraq". Documents with the word "Baghdad" or "Basra" will however not necessarily be retrieved since the relationship between these two towns and the country Iraq is not known by the search engine. A second issue is the geographical location of the search results, which is not known either.

At present a spatial search engine is under development by the Eurpean Satellite Centre in Torrejon and the University of Applied Sciences in Hof. This search engine supports the geographical search conditions and provides the geographical location of the found documents. With this approach it is for example possible to visualise search results on a map. This article describes the functionality of the new search engine and discusses the challenges for its development. It concludes with a summary of the potential and limitation of spatial search technologies.

Introduction

The Internet is increasingly used in security applications for providing information on natural disasters, technical accidents, terror attacks or conflicts. Of particular interest are articles from different publishers (e.g. newspaper) related to these events. International, national and regional news agencies publish already many of their articles in the Internet. Since these agencies compete in publication speed, articles are available in near real time for significant events.

Key tools for accessing any type on information in the Internet are search engines. Current search engines like Google support searches for documents defining words or phrases. ("full text search"). These search engines already use sophisticated methods like algorithms deriving word stems or methods normalising characters to overcome problems with alternative spellings or language specific characters (e.g. automatic mapping of the German letter 'ä' to 'ae').

A full text search is however not sufficient for security applications. In this context most searches refer to certain locations or regions. In general a geographical location could be specified by its name (e.g. Iraq). This approach however will not find documents which use

alternative names (Irak) or names of locations inside a region (e.g. Baghdad, Basra, etc.). Another problem arising from this approach are place names which may have also a different meaning in other contexts. An example is the term "BUSH" either indicating a person, a plant or a location.

The European Union Satellite Centre and the University of Applied Sciences are currently developing a search engine supporting geographical positions as search conditions and providing geographical positions for found documents. This search engine visits a predefined number of sites on a configurable schedule (typically from new agencies) and analyses all available documents for place names, words and phrases. Then the engine stores references to these documents in a database which contains geographical positions together with the contained words and phrases. A search interface accesses this database to retrieve all documents for a given set of words and a single location.

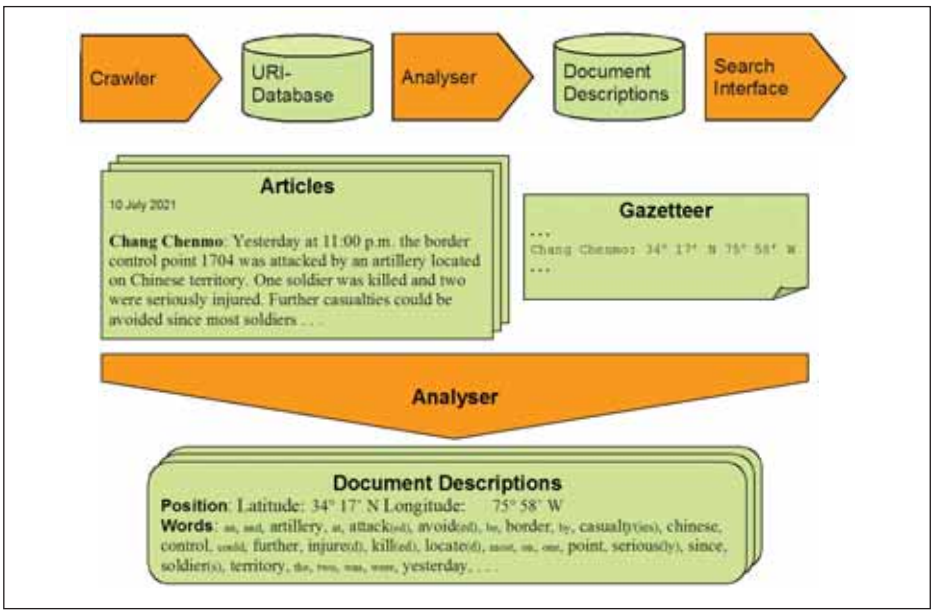
In the next section we will present the concept and functionality of a spatial search engine and discuss key challenges. The last chapter summarises current potential and future perspectives of spatial search engines.

Concept

Traditional search engines support the finding of documents from the Internet containing specific words or phrases. For this purpose search engines need to navigate the Internet on a regular basis, analyse the content of all found documents and generate an index from their content. This index is the core of all search engines since it facilitates efficient searching. Without this index search times would be unacceptable high. Accordingly search engines contain at least the following logical components:

The task of the *Crawler* is to find all documents in a specified part of the Internet. The Crawler starts with a given set of so called Uniform Resource Identifier (URI) and navigates the Internet by following the links in the found documents. The output of the Crawler is a set of URI's pointing to all found documents. The "journey" of the Crawler could be restricted by an upper bound for the number of links which the Crawler may follow from a start URI. In addition the search could be restricted to one or more (sub-) domains in the Internet. The search frequency of the Crawler is defined by a schedule which might different for every start URI.

The *Analyser* retrieves documents via their URI and analyses their content. A conventional search engine will extracts words and phrases from the document and stores this information in document descriptions (probably performing some processing as normalisation and stemming before storing this information). In our case the analyser needs to derive the geographical position as well.



Logical Components of a Search Engine, Analysis of Documents

The *Search Interface* supports searches of end users. Here the user may specify one or more terms or phrases. In our application context a user may specify in addition a geographical region. Then the search interface will return all relevant documents according to the specified geographical region and the given words or phrases.

Figure 1 shows the data flow between the logical components of a conventional search engine.

The key part of every search engine is the Analyser. Already analysers of conventional search engines face several challenges. As an example, analysers have to deal with different types of document formats and need to be sufficiently robust in case of format errors. An additional challenge in our context is the derivation of the geographical location.

The analysis of a document is visualised in Figure 2. The analyser will extract all words and phrases from a document (extraction of phrases is not shown here). During this extraction the analyser may remove so called stop words ("an", "and", "at", . . .) occurring in almost every document. Also the extracted words could be processed. In our figure the analyser derives word stems facilitating some flexibility in the writing of words.

A spatial search engine needs to identify also place names in a document. Candidates for these place names are all words from the document which are also contained in a so called *Gazetteer*. A Gazetteer provides all relevant place names together with their geographical positions. Using this information the Analyser will store positions of some or all place names in the document description.

Richard Göbel *is at the University of Applied Sciences Hof in Germany. Email:* Richard.Goebel@fh-hof.de

REFERENCES

- Lars Arge, Mark de Berg, Herman J. Haverkort, Ke Yi: The Priority R-tree: "A Practically Efficient and Worst-Case Optimal R-tree", Proceedings of the 2004 ACM SIGMOD International Conference on Management of Data, p. 347 - 358, Paris, France,2004
- R. Göbel: "Towards Logarithmic Search Time Complexity for Two-Dimensional R-Trees", Proceedings of Second International Conference on Systems, Computing Sciences and Software Engineering 2006, Springer, to appear in 2007
- R. Göbel, L. Colaiacomo, A. de la Cruz: A Spatial Search Engine Supporting Security Applications, Proceedings Fourth Conference Image Information Mining, ESA Workshop Proceedings, 2006
- A. Guttman: „R-Trees: A Dynamic Index Structure for Spatial Searching" Proc. ACM SIGMOD Conference, Boston, pages 47 - 57, 1984
- M. Kimler: „Geo-Coding: Recognition of geographical references in unstructured text and their visualisation", Diplomarbeit, Fachhochschule Hof, 2004
- K.V. Ravi Kanth, A. Singh: "Optimal Dynamic Range Searching in Non-Replicating Index Structures" Proc. International Conference on Database Theory, pages 257 - 276, Springer Verlag Berlin Heidelberg, 1999
- S. Rill: „SARA Semi Automatic Research Archiv - Entwicklung eines Konzepts zur ortsbezogenen Archivierung von Nachrichten und Fakten - Demonstration dieses Konzepts in Form eines Prototyps", Diplomarbeit, Fachhochschule Hof, 2005

KCS TraceME

GPS/GPRS/SMS Module

Very low price

www.TraceME.eu

Why buy TraceME products?

1. monitor your fleet with GPS/GPRS/SMS
2. monitor your staff in any country
3. send messages to/from your people
4. monitor your machines wireless via internet
5. update machines worldwide with your PC
6. sufficient I/O for sensors etc.
7. iButton auto start safety device
8. guard properties via mini camera
9. unlimited applications possible
10. Dutch design
11. Dutch production and support
12. Satisfied customers:

- IMTECH
- DUTCH POLICE FORCE
- PORTUGUESE FIRE DEPARTMENT
- AGGREKO GENERATORS
- SKYNET MOBILE
- UK GOVERNMENT
- MOTOROLA DUBAI
- AFRICAN OIL COMPANIES
- GERMAN ROADSERVICE
- and many others worldwide

A TraceME solar version can be supplied for situations where no fixed power source is available.

TraceME mini SPY camera watches over your belongings and transmits pictures via GPRS. Watch them via your web browser.

Communicate in your car with the office or with another vehicle via a.o. SMS with the TraceME LCD module. Secure your vehicle with the help of an iButton start.

TraceME application areas

Measurement & Control engineering:
Wireless distance control of generators, all equipment with an SPI or RS232-interface. Climate control.

Cold-storage engineering:
Detection of temperatures during cold-storage transport and alarm signals when deviations occur.

Medical:
Monitoring transport of organs, stem cells, blood, etc. Ambulances and other transports.

Services & Logistics:
Tracing transports: Busses, cars, lorries, taxis, fire engines, river craft, trailers (also on solar energy)

M2M (machine to machine)
Pilot/control your machines/pumps etc. via the internet without the help of mechanics.

Security/ safety:
Buildings with or without miniature cameras and/or sensors via wireless transmission.

Police:
Various police services have already been equipped with TraceME modules + cameras in various countries.

* All information in this advertisement subject to change. E. & O. E.

(c)KCS BV, Kuipershaven 22, NL-3311AL Dordrecht. Fax: +31 20-5248130 www.TraceME.eu kcs@kcs.tv