Can Geography Rescue Text Search? Where is Your Information About?

Authors: Randy Ridley, MetaCarta General Manager of Federal Systems
Jeffrey Sauder, MetaCarta Senior Systems Engineer
John Frank, MetaCarta President & CTO

Abstract

Decisionmakers depend on information buried in text documents. Millions of staff weeks are spent reading text. As the volume of textual information increases across an organization, users need new methods for understanding this critical information. When we communicate, we reference locations and events. We understand situations by interpreting geographic references like "20 miles east of Média." Until recently, software had trouble interpreting ambiguous names like Media. Recent technology advances change this. Sorting information geographically is creating a paradigm shift in organizational workflow. Latitude and longitude provide a grounded reference model for seeing patterns and correlating information. This new technology extends the power of geography to text documents and text search/retrieval.

Introduction

Emails, HUMINT and SIGINT reports, news feeds, and reports contain critical intelligence for mission support. Analysts spend millions of hours understanding these documents. Most documents contain at least one reference to a specific geographical location. Geography offers a powerful method of searching through these documents. Combining text search and geographic search allows analysts to see and find relevant information faster.

Geography offers a new framework for identifying and analyzing the overwhelming volume of text documents available today. Technology that automatically plots text documents on maps based on spatial references and addresses that appear within the text accelerates our ability to find relevant information. Open Source Intelligence (OSINT) gathering must cope with more sources than any team could possibly comprehend. Combining geographic search with full-text keyword search, document categorization, and temporal filtering accelerates the delivery of intelligence to decision makers in the field. Instead of missing essential information, analysts can see geographic trends as they emerge. Viewing a map with text documents represented as geographic features improves situational awareness and enables instant corroboration between disparate sources that discuss proximate locations.

Mission Critical Application - Minesweeping

Analysts must find and assess documents that relate to specific mission-support applications. For example, the United Nations Mine Action Center (MAC) in Sarajevo supervises the demining program and coordinates all mine awareness and training with local governments and international interests. Analysts at the National Ground Intelligence Center (NGIC) provide general military intelligence (GMI) in support of operational

forces and other services. Among the duties performed at NGIC are analysis of foreign military technology and development of strategies to deactivate landmines. [i] In this capacity, NGIC supports field commanders and demining teams with information that aids UN and NATO peacekeeping missions.

Field commanders rely heavily on geography to understand where unexploded landmines and other ordnance exist. Specifically, in Sarajevo, 1,818 minefields have been registered around the outskirts of the city. The former Yugoslavia has even more landmines. Bosnia has an estimated 600,000 to 1 million landmines within 115 square miles of land. In Kosovo, as many as 1 million mines occupy 2,000 minefields planted by Yugoslav forces and the rebel Kosovo Liberation Army during the conflict there. [ii]

Using map displays to aid communication, analysts provide specific information to officials in the field to help limit troop and civilian casualties. Commanders and UN officials need to know everything about a particular geographical location before risking lives. Troops gather intelligence from local people and farmers regarding the location of suspected minefields. Analysts provide additional intelligence by drilling down into historical reports and previous analyses.

Extending Beyond a Text Search

To support field commanders and aid peacekeeping missions such as minesweeping efforts, analysts use text searches to manually read articles and documents to provide mission support. In the case of demining, every piece of information on a specific location provides clues that could reveal unexploded landmines. To save lives, analyst must read every document that could identify the location and extent of a minefield, and they must do it as fast as possible.

However, without a way to search for documents by using a specific geographical location, analysts must expend many hours looking for direct and indirect references. For example, different documents can refer to locations within the same area, without calling out the location explicitly:

Regular latitude and longitude: 43°52'N 18°26'E

UTM or Military grid reference: 34TBP950601

Natural language relative reference: 8mi north west of Pale
289 Kilometres south east of Zagreb

Phone number: +387 (71) 667 313

Traditional keyword search techniques are unable to identify these locations. Searching for the location “Zagreb” illustrates a fundamental limitation: the results contain phrases like, “a spokesman said Mr. Zagreb would not be available for comment …”

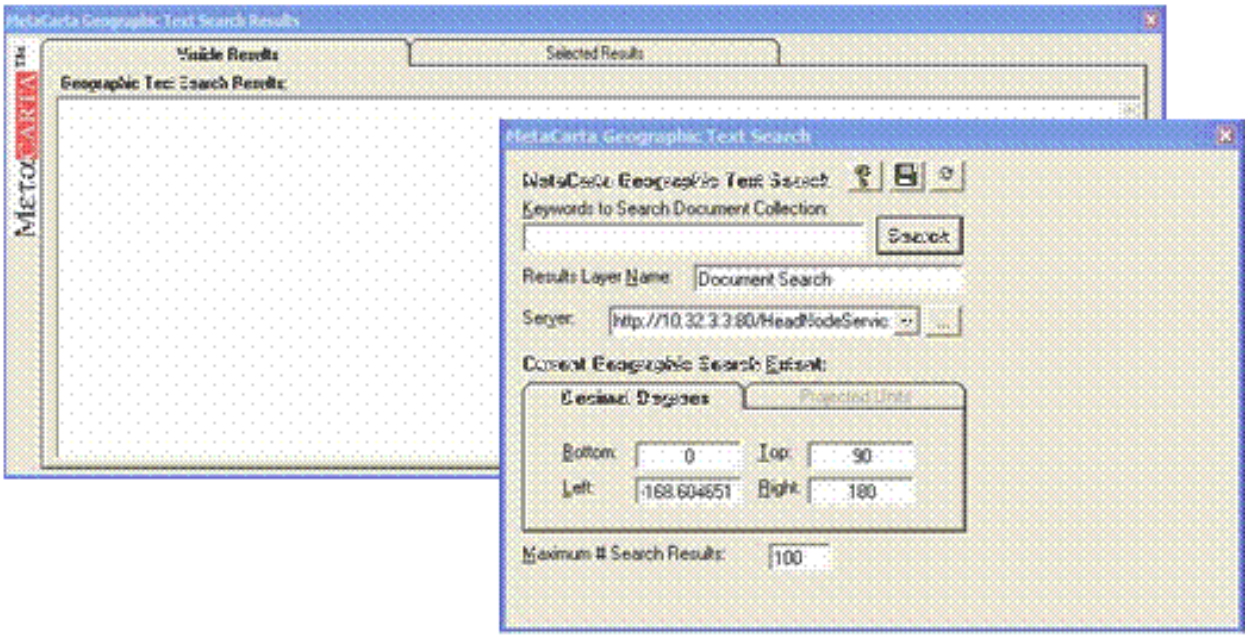
The Use of Maps in Conjunction with Text Search

To plot documents on maps, systems must extract text strings that identify locations and compute latitude-longitude coordinates for those references. This is referred to as geographic text search (GTS). This approach uses natural language processing to disambiguate the meaning intended by the author, and assigns confidence scores that estimate the probability that the system chose the correct latitude-longitude. Using these confidence scores, a GTS system sorts the most probable results to the top. This filtering improves analysts' efficiency by boosting the signal to noise ratio.

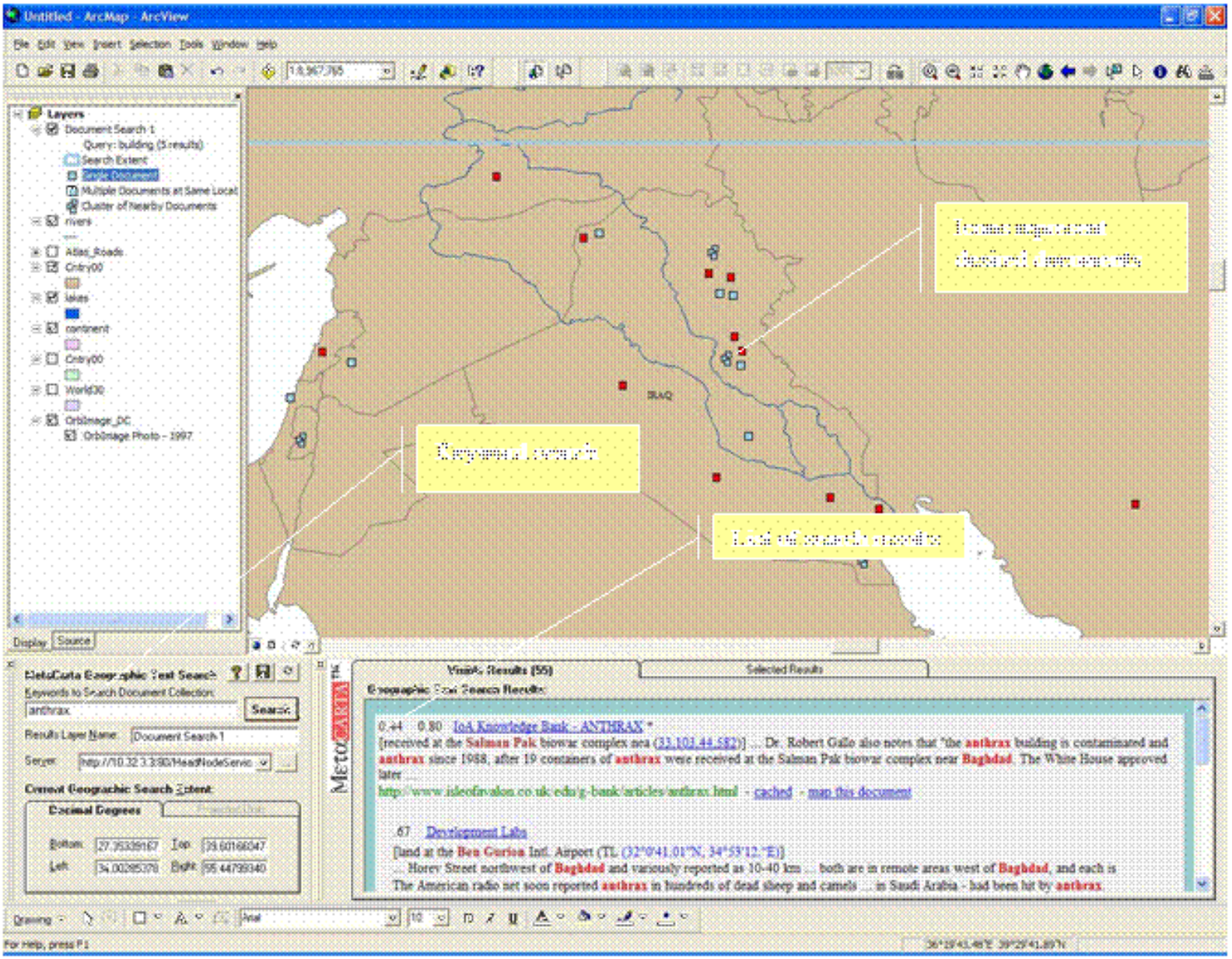
ESRI ArcGIS and MetaCarta Geographic Text Search

Through a cooperative effort, ESRI's ArcGIS software provides a powerful presentation layer for seamlessly merging geography and text searches served by the MetaCarta Geographic Text Search™ (GTS) system. For years, analysts have relied on outstanding ESRI tools to help sift through geographic materials, but geographic information buried in text documents has been a challenge. Analysts can now efficiently discover and comprehend geographical references within documents. Analysts leverage MetaCarta's geographic text searching capabilities through ESRI's ArcGIS interface courtesy of a freely available extension linking the two technologies. The following provides instructional information for performing a geographic text search via the ArcMap interface.

MetaCarta, Inc. provides a free extension for ESRI ArcMap that links the two technologies via a Web Services SOAP interface. After downloading the extension, an analyst using ArcMap will notice a new icon within the ArcMap interface (*Figure 1*). By clicking on the MetaCarta icons, two MetaCarta window boxes appear in the ArcMap interface (*Figure 2).* The smaller window presents an input field for entering a keyword. The larger window displays the results of the geographic text search.



The MetaCarta GTS™ automatically extracts geographic references from text documents and plots documents on a map. This allows analysts to search text archives using both keywords and geographic extents as filters. Small colored square icons appear in the map to represent the documents that contain desired keywords and geographic references to the location below the icon. These icons represent a direct hyperlink to the documents (*Figure 3*).



ESRI ArcGIS user interface with MetaCarta GTS plug-in maintains a consistent look and feel to minimize the learning curve.

Analysts can perform multiple text searches to narrow down the field of results. This iterative process of adjusting the input keywords has become commonplace on the Web. It will soon become commonplace inside GIS applications accessing documents, like the Web, to use MetaCarta GTS for ArcMap. ArcMap users can load a wide variety of data layers to create customized maps. This extension opens an entire new area of geographic data assets to the GIS user. Each text search creates a new data layer.

ESRI ArcIMS and Geographic Text Search

MetaCarta GTS also integrates with map servers, such as ESRI ArcIMS. This allows analysts to search using

a web browser displaying a backdrop map derived from private map data. ArcIMS can hold a central copy of the team's most current data. As this data is modified or updated, the most current version is immediately visible to analysts. By integrating the team's most current geospatial data with text search, analysts can identify documents that are specific to important areas and correlate them with other spatial data layers.

Again, analysts need to provide field commanders and peace keepers with geographically sensitive intelligence to understand where unexploded landmines and other ordnance exist. The use of ESRI ArcIMS and MetaCarta GTS provides a web-based solution to obtain up-to-date documents referring to such operational interests.

Conclusion

Geography offers a new framework for identifying and analyzing the overwhelming volume of text documents available today. Most documents contain at least one reference to a specific geographical location, so maps offer a new way to search text documents. Technology that automatically plots text documents on a map based on geographic references accelerates our ability to create immediate intelligence. Analysts can see geographic trends as they emerge, improving the quality of the information known about a given situation.

MetaCarta was founded by a team of MIT researchers in 1999 to create powerful new tools to help analysts gain more intelligence from data by synthesizing text and geography. With early support from DARPA and private investors, MetaCarta released MetaCarta Geographic Text Search (GTS) as a regular COTS product in 2002. MetaCarta GTS is the only geographic text search solution on the market today. MetaCarta is an In-Q-Tel portfolio company.

About the Authors

Randy Ridley is general manager of Federal systems. He is a seasoned, results-oriented sales executive with an extensive background in marketing, product, and program management for software, hardware, and services solutions. Viewed by the industry and his peers as an insightful and knowledgeable resource, Mr. Ridley has been published in the OP-ED pages of the Washington Post, Defense News, and has been interviewed on Federal matters by such publications as Washington Technology, Federal Computer Week, and Government Computer News. He has been an invited speaker on Information Technology (IT) matters in national and international theaters, and has provided his first-hand IT knowledge and experiences on television media via various outlets. Prior to MetaCarta, Randy held a senior project management position at TASC, Inc. - a business unit of Northrop Grumman Information Technology. Mr. Ridley is a U.S. Naval Officer, surface warfare, and holds a B.A. in history, University of Iowa, and a M.A. in national security studies from George Washington University.

Jeff Sauder brings more than 18 years experience to his role as senior systems engineer. Mr. Sauder handles all aspects of technical product integration, sales support, and application support for MetaCarta. He developed his skills writing software for the National Institute of Standards and Technologies and then later for the Department of Defense while working for several integration firms. Jeff honed his database skills at Sybase while performing professional services consulting for numerous clients. At IBM, he provided

technical support to the DB2 sales force. Jeff holds a B.S. in Computer Science from the University of Maryland, and a M.S. in Computer Science from Johns Hopkins Whiting School of Engineering, Laurel, MD.

John Frank is founder and president of MetaCarta, Inc. While working on his PhD in physics as a Hertz Fellow at Massachusetts Institute of Technology (MIT) in 1999, John conceived of a new way to view collections of text: geographically. His strategic vision and team-building leadership has carried MetaCarta to its current leadership position. Prior to MetaCarta, John refined his product development skills at IDEO, Palo Alto, CA. Focused on strategic planning, partner/customer relations, and business development, he has U.S. Government Secret-level security clearance and spends time in the field with users and decision makers to understand the community's new, evolving requirements. John holds a B.S. in physics, with distinction, from Yale University.

References

[i]
National Ground Intelligence Center (NGIC) Chapter 8 FM 34-37 Strategic, Departmental, and Operational IEW Operations Preliminary Draft; http://www.fas.org/irp/agency/inscom/ngic

[ii]
Chicago Tribune "Mines still render Bosnians helpless"; Alexandar S. Dragicevic, Associated Press , April 12, 2000; http://www.chicago.tribune.com/version1/article/0,1575,ART- 44246,00.html