

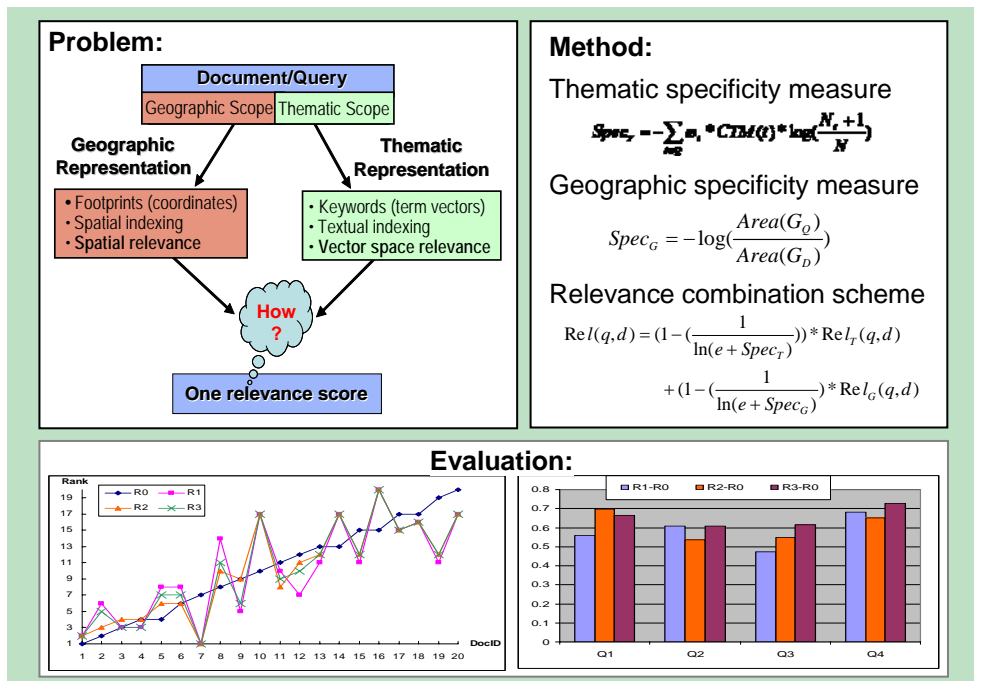
Geographic Information Retrieval

A Dynamic Document Ranking Scheme Based on an Integrated geo- and thematic- relevance model

Benefit: Homeland security often requires assembly of diverse information relevant to particular places (e.g., developing sound mitigation plans to increase resilience of communities to natural disasters). The required geographic information often exists in documents having no explicit specification of the places they are about. Geographic Information Retrieval (GIR) methods and tools being developed by NEVAC directly address the challenge of retrieving documents that are relevant to a task in terms of both document theme and geographic focus.

When analysts need information about a *topic* (e.g., community preparedness for forest fires) in a *place* (e.g., Southern California), existing information is often inaccessible – scattered in reports and other documents whose titles do not make their relevance obvious. In traditional information retrieval models, relevance is represented by the similarity metric between a document and a query. A single relevance score based on all the measurable evidence is calculated to rank documents. Geographic Information Retrieval (GIR) goes beyond this to focus on both topic and place. This presents a unique challenge, because relevance ranking must consider more than topic, but also include the spatial relationships (such as contain, overlap, intersect, connect, near etc.) between documents and queries in geographical space. NEVAC is focusing on the key GIR question of how to address the relative independence and complex interactions between the two document subspaces (geographic and thematic) in order to retrieve only the most relevant documents.

In our recent work, we developed a dynamic document ranking scheme to combine the thematic and geographic relevance measures on a per-query basis. Unlike previous work based on fixed functions, our method uses a functional form that is dynamically determined by the relative importance of thematic and geographic relevance through analyzing *query specificity*. Our assumption is that users tend to apply more specific thematic terms if thematic relevance is paramount, while they tend to use more specific geographic scopes if the place is critical. A preliminary evaluation comparing the computed relevance measure with human judgment shows that our method is potentially effective. To improve success, we are exploring the application of Dempster-Shafer's theory to develop additional strategies for combining the two different sources of ranking evidence.



Funded by: Department of Homeland Security

Early Development

Lab Prototype

Commercial Product

Jan 2008

For more information, contact:
Alan MacEachren, (814) 865-7491, maceachren@psu.edu
www.geovista.psu.edu/NEVAC