

*promoting access to White Rose research papers*



**Universities of Leeds, Sheffield and York**  
**<http://eprints.whiterose.ac.uk/>**

---

This is an author produced version of a paper published in **Advances in Information Retrieval**.

White Rose Research Online URL for this paper:  
<http://eprints.whiterose.ac.uk/78507>

---

#### **Published paper**

Clough, P., Joho, H. and Purves, R. (2006) *Judging the Spatial Relevance of Documents for GIR*. In: Lalmas, M., MacFarlane, A., Rüger, S.M., Tombros, A., Tsirikas, T. and Yavlinsky, A., (eds.) *Advances in Information Retrieval*. 28th European Conference on IR Research, 10th - 12th April 2006, London, UK. Springer Berlin Heidelberg, 548 - 552.  
[http://dx.doi.org/10.1007/11735106\\_62](http://dx.doi.org/10.1007/11735106_62)

---

# Judging the spatial relevance of documents for GIR<sup>1</sup>

Paul D. Clough<sup>1</sup>, Hideo Joho<sup>2</sup> and Ross Purves<sup>3</sup>

<sup>1</sup>Department of Information Studies, University of Sheffield, UK  
[p.d.clough@sheffield.ac.uk](mailto:p.d.clough@sheffield.ac.uk)

<sup>2</sup>Department of Computer Science, University of Glasgow, UK  
[hideo@dcs.gla.ac.uk](mailto:hideo@dcs.gla.ac.uk)

<sup>3</sup>Department of Geography, University of Zurich, Switzerland  
[rsp@geo.unizh.ch](mailto:rsp@geo.unizh.ch)

**Abstract.** Geographic Information Retrieval (GIR) is concerned with the retrieval of documents based on both thematic and geographic content. An important issue in GIR, as for all IR, is relevance. In this paper we argue that spatial relevance should be considered independently from thematic relevance, and propose an initial scheme. A pilot study to assess this relevance scheme is presented, with initial results suggesting that users can distinguish between these two relevance dimensions, and that furthermore they have different properties. We suggest that spatial relevance requires greater assessor effort and more localised geographic knowledge than judging thematic relevance.

## 1 Introduction

Geographic Information Retrieval (GIR) is a relatively new research area, concerned with the retrieval and ranking of documents from collections based on queries that specify both thematic and geographic scopes [1,2]. As with any new form of retrieval appropriate methodologies and resources are required to evaluate GIR systems [3][4]. In IR, test collections are often used to benchmark system performance (e.g. TREC, CLEF and INEX), however existing resources do not necessarily distinguish spatial aspects of information, an important point to consider when evaluating GIR [5].

Assessing the relevance of documents across multiple dimensions is not a new problem in IR. For example, INEX has implemented such a scheme to assess both structural and conceptual relevance of XML documents. Cai [2] suggests two subspaces for GIR which represent two different cognitive aspects of relevance: geographic and thematic. In the former, relevance is judged based upon spatial relationships (e.g. overlap and adjacency) between the query location and spatial footprints identified within a document. It is therefore necessary to assess whether different relevance schemes are required to evaluate GIR systems based upon the specifically geographic aspects of documents.

This paper investigates the usability of an assessment scheme which takes into account both thematic and spatial relevance for the evaluation of GIR systems. The following sections present our proposed relevance scheme and a pilot test evaluating the usability of such a scheme, before briefly discussing some aspects of our results and considering the implications.

---

<sup>1</sup> Research part-funded by EU-IST Projects IST-2001-35047 (SPIRIT) and IST-2002-2.3.1.12 (BRICKS).

## 2. Relevance scheme

Our initial proposed scheme (shown in Table 1) was based on assessing the relevance of Web documents and used a three-point scale to indicate the degree of relevance (we intentionally did not label the scales as highly relevant, partially relevant, or relevant to avoid the influence of preconceived ideas about relevance).

**Table 1.** Relevance scheme

<b>Thematic relevance</b>	
Score 1	A document which contains relevant information about the concept queried AND on its own allows you to form a judgment about the document (i.e. requires no external knowledge).
Score 2	A document is relevant, since it points to a resource MENTIONING the concept, but you must consult further pages referenced by the document to perform a judgment.
Score 3	A document does not provide information about the concept provided.
<b>Spatial relevance</b>	
Score 1	A document refers to a location that is/near the query location AND you think that the location in the document has sufficient detail for you to find it on a local map of the area.
Score 2	A document refers to a location that is in/near the query location BUT you think that there is insufficient information for you to find that location on a local map of the area.
Score 3	A document does not fall within the query location.

## 3 Experiment and results

A pilot user study was carried out to investigate the effectiveness of the proposed scheme. Subjects were asked to make relevance judgments regarding the thematic and spatial relevance of 10 documents per topic using the proposed schemes. Each subject was given five topics to judge. The judged scores were analysed in relation to their distribution across the topics, inter-assessor agreement, ease of assessment, confidence in judgments and any difficulties they faced in assessing topics through the use of pre-topic, post-topic and post-session questionnaires.

Documents were retrieved using SPIRIT, a prototype spatially-aware search engine [6], based on a set of approximately 20,000 web pages. A set of 10 documents were retrieved based for each of the following five topics: 1) Caving in Derbyshire (UK), 2) Castles in Wales (UK), 3) Skiing near Glencoe (UK), 4) Art festivals in Edinburgh (UK), and 5) Music in Montreux (Switzerland).

In addition, a paragraph-length description of each topic was provided to subjects in order to help them make their judgments. Subjects were also allowed to use the Internet as a source of geographical knowledge and could also select “Not sure” when unable to make an appropriate decision for a particular document. A total of 11 subjects participated in the experiment giving a total of 1,100 judgments (550 judgments for each type of relevance). The results of our experiment are as follows.

### 3.1 Perception of subject assessments

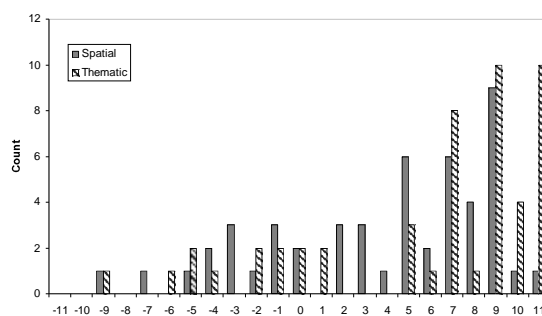
Subjects were asked three questions: (Q1) Was the three-point scale suitable, (Q2) Were the schemes easy to understand, and (Q3) Did you make a judgment confidently? and results are shown in Table 2. Assessors agreed that a ternary scheme was suitable for judging and most found the scheme easy to understand, especially for thematic relevance. Assessors appeared to be more confident in judging thematic than spatial relevance.

**Table 2.** Participants' perception of the assessment scheme (T: Thematic S: Spatial).

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Q1 (T)	0	2	0	7	2
Q1 (S)	0	1	0	7	3
Q2 (T)	0	0	1	5	5
Q2 (S)	1	1	0	6	2
Q3 (T)	0	0	4	6	1
Q3 (S)	1	2	3	3	2

### 3.2 Relevance assessments

Although our subjects appeared to have been somewhat confident about making relevance judgements in both the thematic and spatial cases, on initial investigation of our data we found that inter-annotator agreement was in many cases relatively poor (a multi-rater Kappa test gave  $k=0.1886$ ,  $p<.05$  for thematic relevance and  $k=0.1388$ ,  $p<.05$  for spatial relevance). We decided to investigate this contradiction in more detail and, to ease analysis, reduced both the thematic and spatial relevance judgements to a binary scale.

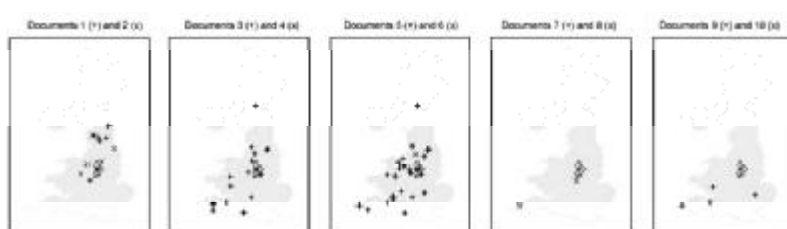


**Figure 1a.** Sum of relevant and non-relevant judgments per document

Figure 1 shows a histogram illustrating summed relevance judgements for each document, where a single judgement of not relevant scored -1, and a judgement of relevant scored 1. Thus, 10 documents were judged to be thematically relevant by all 11 of our subjects. The histogram is biased towards relevant documents since the judgements were made on documents retrieved by a GIR system which we expect to retrieve at least some relevant documents. However, it is clear that there is considerably more inter-annotator agreement for thematic judgements than spatial judgements. Furthermore, thematic relevance judgements are only moderately correlated with spatial relevance ( $r=0.63$ ).

Since it appeared that, although our subjects were happy with the scheme for making spatial judgements, in many cases they disagreed about the spatial relevance of individual documents, we decided to investigate such judgements in more detail. We parsed location names from documents, and where possible geocoded them for each of the 10 documents from each topic. Figure 2 shows the results for the 10 documents retrieved for Topic 1 (caving in Derbyshire). The spatial relevance judgement for each of these documents was as follows: (doc:score): 1:2, 2:9, 3:-5, 4:7,

5:-1, 6:0, 7:5, 8:0, 9:-1 and 10:-7. Therefore, documents 2, 4 and 7 were judged to be spatially relevant by more than 2/3 of our subjects, and documents 3 and 10 was judged to be spatially irrelevant by more than 2/3 of our subjects. Documents 1, 5, 6, 8 and 9 are all more or less ambiguous in terms of judgements of spatial relevance.



**Figure 2.** Distribution of spatial references in documents for topic 1 (mapped as pairs).

#### 4. Discussion

The spatially relevant documents (2,4,7) all appear to have a small set of locations tightly focussed around Derbyshire (indicated in the centre of Figure 2). However, much less of a pattern is evident in the case of both the ambiguous or irrelevant documents. We suggest that this is because documents which have a well-defined geographic focus and are centred on the query location are probably much easier to assess, in particular for subjects with less geographic knowledge of the query region. Documents referring to many locations appear difficult to assess according to the subject's background knowledge of the area resulting in poor inter-annotator agreement.

These initial qualitative results show that, in general, spatial relevance appears more difficult than thematic relevance to judge. However, they also suggest that use of a "spatially-aware" relevance scheme is appropriate. Further work is required with larger subject and topic groups to refine the scheme and to propose guidelines for making appropriate judgements of spatial relevance.

#### References

- [1] McCurley, S.K. 2001. Geospatial mapping and navigation of the web. In Proceedings of the Tenth International WWW Conference, Hong Kong, 221-229.
- [2] Cai, G. (2002) GeoVSM: An Integrated Retrieval Model for Geographic Information. LNCS 2489, 70-85.
- [3] Bucher, B., Clough, P., Joho, H., Purves, R., and Syed, A. K. (2005) Geographic IR Systems: Requirements and Evaluation. In Proceedings of the 22nd ICC, A Coruña, Spain.
- [4] Martins, B. Silva, M.J. and Chaves, M. (2005) Challenges and Resources for Evaluating Geographical IR, In Proceedings of GIR'05, Bremen, Germany.
- [5] András Kornai (2005) MetaCarta at GeoCLEF 2005, In "The working notes of the CLEF workshop", Vienna, Austria, 21-23 September 2005.
- [6] Jones, C.B, A.I. Abdelmoty, D. Finch, G. Fu and S. Vaid, 2004. The SPIRIT Spatial Search Engine :Architecture, Ontologies and Spatial Indexing. LNCS 3234, 125-39.