# Spatially-Aware Information Retrieval on the Internet

**SPIRIT is funded by EU IST Programme**
**Contract Number: IST-2001-35047**

# *Aspects of Spatial Similarity Measures*

**Deliverable number:** D8:5101

**Deliverable type:** R

**Contributing WP:** WP 5

**Contractual date of delivery:** 1 July 2003

**Actual date of delivery:** 1 July 2003

**Authors:** Roelof van Zwol, Marc van Kreveld, Iris Reinbacher, Remco Veltkamp, Cristopher B Jones

**Keywords:** spatial similarity measures, spatial object.

**Abstract:**

The aim of this document is to provide an overview of available similarity measures in the context of spatial information retrieval. A summary of possible representations of geographic objects is given, and a discussion of relevant (spatial) similarity measures is held, based on the available literature in the areas of Geographic Information Systems (GIS) and Information Retrieval.

# Contents

# D8 5101

# *Aspects of Spatial Similarity Measures*

## 1. Introduction

Information on the Internet is often related to geographic positions, or has a spatial character that is neglected by current search engine technology. Taking this spatial information into account when searching for information on the Internet might increase the performance in terms of recall and precision [10]. To realize this, it is necessary to investigate what spatial similarity measures are necessary to realize a successful integration of the spatial retrieval component into a web-based search engine.

Whereas Geographic Information Systems (GIS) can use complex query formulation techniques, allowing complex spatial relations to be resolved in the query, the limitation of a search engine targeting a wide variety of users and an immense collection of data allows only a subset of these spatial relationships to be used. This automatically influences the set of spatial similarity measures that can be successfully determined.

In literature a wide variety of similarity measures are described, which are not only limited to the spatial domain. In the area of Information Retrieval, the relevance of text fragments and images to an information request is calculated using similarity measures. When focusing on the spatial domain again, a number of applications can be distinguished that compare spatial objects, using spatial similarity measures: map generalization [3], cartogram construction [1, 2], terrain analysis [4], and image analysis of satellite images [7].

In SPIRIT the spatial aspects of information retrieval on the Web are examined to realize a spatially-aware search engine for the Internet [28]. To achieve this, several components have been identified that interoperate to successfully perform the spatial retrieval task. A description of these components can be found in [19]. Among these components, a spatial ontology, a meta-data component, a spatial index, and a relevance ranking component are used that require spatial similarity measures to be available to complete their task. In this document the focus will be on similarity measures that may turn out to be relevant for those components. Contrary to the user-driven approach followed for the user requirements specification [18], a data-driven approach will be followed here. The main motivation for this is that during the indexing process no intervention by the user is possible. Of course one should take precautions to avoid a gap between the user requirements and the spatial similarity measures that are used within the SPIRIT search engine.

Both qualitative and quantitative measures are needed within SPIRIT for the evaluation of the spatial character of a document and its relevance to a given information request. Furthermore, there is a special interest in fuzzy matching of geographic objects is expressed [26, 27], since a user may not always be able to pinpoint the exact location or occurrence of a geographic object. Secondly, it is likely that it is not always feasible to exactly identify all geographic objects found in a document. This is partially the task of the spatial ontology and the meta-data component, whereas the spatial index and relevance ranking component will focus on finding matches between the spatial information derived from the information request and a document. The result of this match can be precise, if the similarity measures allow exact identification of the similarity between two geographic objects. But, it is also likely that the result expresses a certain degree of relevance between two objects based on the spatial similarities that are used. The relevance ranking component primarily depends on spatial similarity measures to calculate a ranking of documents. Therefore the similarity between spatial objects found in the information request, and the spatial objects derived from the Web-documents has to be calculated.

### 1.1. Organization

The remainder of this document is organized as follows. Section 2 starts with a discussion of spatial objects and the forms in which they can occur. Next, in Section 3 this discussion is continued with an introduction to (spatial) similarity measures. In Sections 4 and 5 we will go into more detail on similarity measures in the areas of GIS and IR. We will finish off with a discussion of the usability of the identified similarity measures in the case of SPIRIT in Section 6, and present our conclusions in Section 7.

## 2. Spatial Objects

Spatial objects are called 'spatial' because they exist inside 'space', also called the embedding space [8]. The following typology of embedded space is often useful:

- **Euclidean**. This type of space allows distances and bearing between spatial objects to be calculated. The spatial object is then represented by a set of coordinates.
- **Metric**. Again distances between spatial objects can be measured. For example the travel time between points x, and y in space, assuming that the travel time is symmetric.
- **Topological**. Topological relationships differ from those measured in Euclidean and metric space, since they are preserved under `rubber sheet' transformations. This allows relationships such as connectivity and adjacency to be calculated between two objects in the topological space.
- **Set-oriented**. In the set-oriented space relationships between spatial objects such as membership, containment, union and intersection can be computed. Usually a hierarchical structure is used to preserve the set-oriented relationships.

In the most common situations the underlying space is Euclidean when a spatial object is represented by a set of coordinates. In the topological space spatial objects are traditionally represented (in maps) using three types of representations in a two dimensional model: points, lines, and areas [31, 33]. A particular spatial object can often be visualized by two of the three types, depending on the scale that is used for the map. Consider for instant a city, which in one map is represented by an area, while in another map the same geographic object is given as a point in space. With the arrival of computer cartography a differentiation between vector-based and raster-based representation of geographic objects can be made [31, 32, 33]. Although in some applications the raster-based representation can be preferred above vector-based representation, the focus in this document will be on the presentation of spatial objects used a vector-based approach.

Based on the available representations of a spatial object different measures need to be invoked to calculate the similarity between geographic objects, as is discussed in Section 4. Apart from the representation of spatial objects, the thematic characteristics of spatial objects are of importance to the SPIRIT project, since the majority of data of the Web is given in a textual form, instead of using a more sophisticated two-dimensional model. The following thematic characteristics of spatial objects have been identified [5]:

- **Nominal**. Spatial objects in a textual representation, on which no particular mathematical order can be applied. For example the name of a country.
- **Ordinal**. A spatial object, which is represented by a bounded number of classes that are ordered. For example the ranking of a resort.
- **Interval**. Some objects are relative to each other, but have not real origin. The temperature in degrees Celsius is such an example.
- **Ratio**. When an interval spatial object does have a real origin, its characteristic belongs to the group of ratio themes. In those cases the ratio between the values characterizing the object is of importance.

In many of the text-based fragments that are found on the Internet geographic objects can be identified based on their thematic characteristics. This provides the means for a first identification of the spatial information contained in a document. However, geographic objects in that form are not very useful, when it comes to identifying similarity measures in the spatial context. The domain of Information Retrieval, as discussed in Section 5 provides some approaches to measure the similarity of spatial objects based on their thematic characteristics.

## 2.1. Spatial representation in SPIRIT and the geographical ontology

The geographical ontology in SPIRIT (Figure 1 from [20]) will play a fundamental role in determining the data that will be available for computing similarity measures. The ontology will maintain spatial and thematic information about all places, i.e. geographic features, that can be recognised when analysing the content of documents and when processing user queries that refer to places [20]. Figure 1b summarises the information content associated with a geographical feature, on the basis of the initial proposal for a geographical ontology. With regard to thematic data, geographical features have a preferred name and alternative names associated with date and language attributes. A feature has one or more feature types, derived from a feature type thesaurus. The location of a feature is represented both quantitatively with Euclidean geometry and qualitatively with topological relations. The quantitative representation, referred to as a footprint, may consist of one or more of a point, a polyline or a simple polygon. The stored topological relations are those of containment, overlap and adjacency (connectivity).

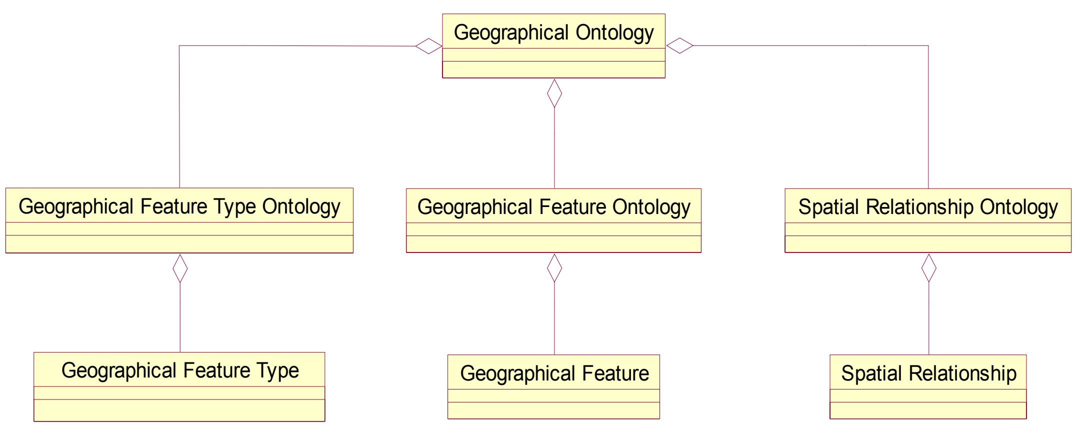


Figure 1a Overview of Geographical Ontology in SPIRIT

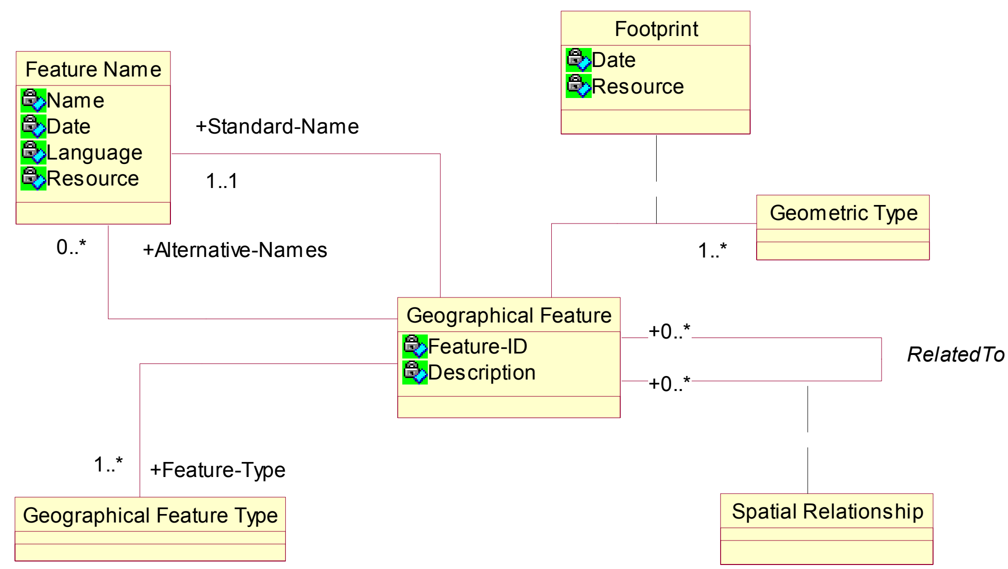


Figure 1b Base Schema of Geographical Feature Ontology in SPIRIT geographical ontology

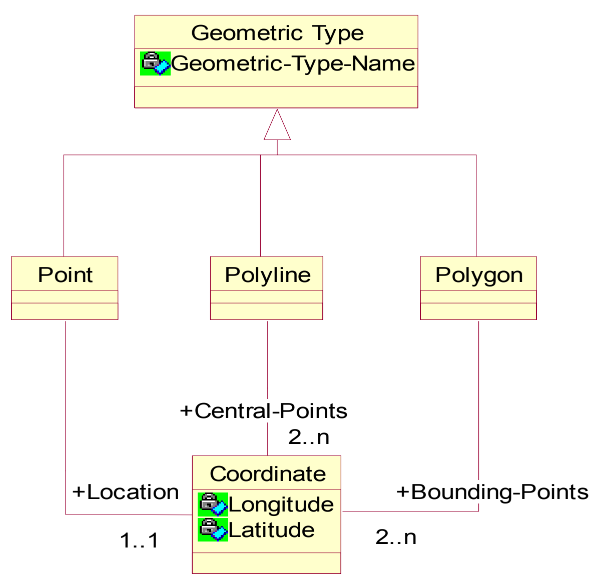


Figure 1c Geometric Feature Types in the SPIRIT geographical ontology

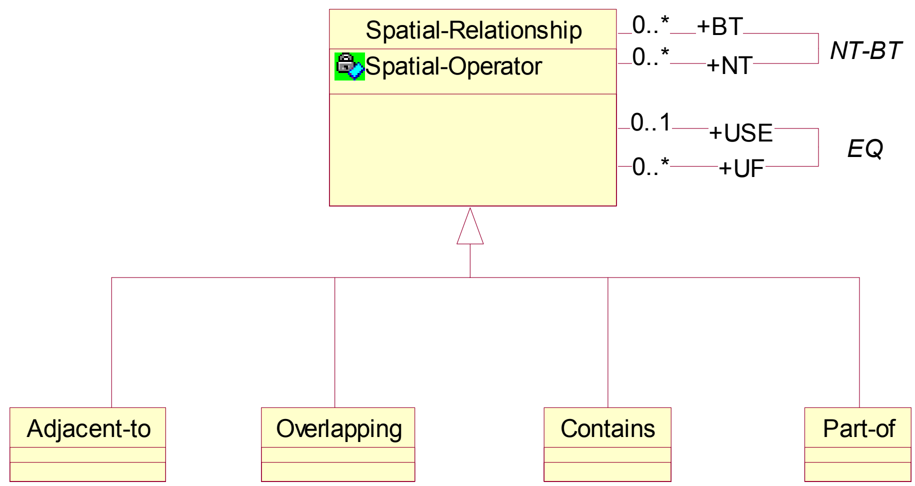


Figure 1d Spatial Relationships in the SPIRIT geographical ontology

## 3. Similarity Measures

The following definition of a similarity measure can be found in [10]:

> "*A **similarity measure** is a function that associates a numeric value with (a pair of) sequences, with the idea that a higher value indicates greater similarity.*"

It is also defined there that a similarity measure is a type of scoring function, where a scoring function maps an abstract concept to a numeric value. However, a large number of operations on spatial objects do not provide a numeric value. To overcome this problem the following approaches can be followed, to obtain a numeric value:

- **Equality measure**: A similarity measure based on the equality operation (see Section 4, Table 1) can be defined as the number of occurrences of a spatial object in a document, where the objects are judged similar, based on their textual equality.
- **Hamming distance measure.** Another measure for textual equality that can also be applied to the spatial domain is the Hamming distance. The Hamming distance of two sequences of equal length is calculated by counting the character positions in which they differ.
- **Hausdorff distance**. An essential problem in the symbol-matching approach to document image compression is to decide whether two spatial objects or images are sufficiently close in shape that they can be adequately represented by the same prototype shape selected from a limited dictionary or *library*. One way of measuring the degree to which two spatial objects or images differ in shape is by first aligning the two marks to coincide spatially as nearly as possible, then determining which point in either shape after alignment lies farthest from the other shape [34, 35].
- **Area of overlap**. Alternatively, the maximum area of overlap, between two objects can be computed, instead of using the Hausdorff distance. This applies only to areal features.

Numerous other similarity measures can be found in literature. In the next two sections an extended discussion on similarity measures in the areas of GIS and Information Retrieval can be found. Other related areas of interest include pattern recognition and shape matching [36, 37].

## 4. Spatial relationships in GIS

Based on the embedded space of a spatial object different spatial relationships (operations) can be invoked. A catalogue of spatial operations on spatial objects is given in Table 1 [4].

Table 1: Catalogue of (static) spatial operations.

| Group | Operation | Operand | Operand | Result |
|---|---|---|---|---|
| General | equals | spatial | spatial | Boolean |
| Set-oriented | equals | extent | extent | Boolean |
| | member of | point | extent | Boolean |
| | Is empty | extent | | Boolean |
| | subset of | extent | extent | Boolean |
| | Disjoint from | extent | extent | Boolean |
| | intersection | extent | extent | extent |
| | union | extent | extent | extent |
| | difference | extent | extent | extent |
| | cardinality | extent | extent | cardinal |
| Topological | boundary | area | | set(loop) |
| | interior | area | | Open area |
| | closure | area | | Closed area |
| | meets | area | area | Boolean |
| | overlaps | area | area | Boolean |
| | Is inside | area | area | Boolean |
| | covers | area | area | Boolean |
| | connected | area | | Boolean |
| | components | area | | set(region) |
| | extremes | arc | | set(point) |
| | Is within | point | simple loop | Boolean |
| Euclidean | distance | point | point | Real |
| | bearing / angle | point | point | $[0,2\pi]$ |
| | length | arc / loop | | Real |
| | area | area | | Real |
| | perimeter | area | | Real |
| | centroid | area | | Point |

The catalogue assumes that the spatial objects are defined in a continuous plane. Each of the operations discussed is grouped into one of the embedded spaces general, set-oriented, topological, or Euclidean. The operations can be applied to either one or two spatial objects, based on the Operands defined in the catalogue. This means that the union can be applied to for instance a spatial object that is represented as a line, and another spatial object, using the area representation.

In the general space two spatial objects can be equal, without any knowledge of the underlying structures. In the set-oriented space several operations are available, such as equals, member of, is empty, etc. The operands of the operation form the extent of the spatial objects that can be `reduced' to a point-set on which the operation is performed [22]. Operations on spatial objects in the topological space are considered to be more complex than in any of the other spaces. A well defined overview on those operations can be found in [4].

The catalogue contains six operations that can be performed on spatial objects in the Euclidean space. The outcome of these operations on either one or two spatial objects do provide numeric values. This kind of measure is of great importance to the relevance ranking component, where a degree of relevance of a document to an information request can be based directly on the outcome of these operations and the number of occurrences of the spatial objects/relationships in the document. Boolean measures can also be used, but in that case the relevance ranking of the document is only determined by the number of occurrences of the spatial objects/relationships in the document, as explained in Section 3.

# 5. Information Retrieval

Contrary to databases, information retrieval is devoted to finding relevant information rather than finding exact matches to data-patterns. With today's popularity of the WWW, information retrieval techniques have a large impact on the way a user's information need is satisfied. In this section a short introduction into information retrieval is given, to provide more insight into the way search engines work. Furthermore, it illustrates the parallel between textual and spatial similarity measures that are of interest to the SPIRIT project.

## 5.1. Text retrieval

An important part of information retrieval is dedicated to text retrieval. A large number of retrieval strategies have been developed for text retrieval. According to [9, 10], the common goal of these strategies is to assign a measure of similarity between an ad-hoc query and a document. The more frequent a term is found in both the query and the document, the more relevant that document will be for the given query. In [9] the following definition for a retrieval strategy is given:

> "*A **retrieval strategy** is an algorithm, that takes a query Q, and a set of documents $D_1, D_2, \ldots, D_n$ and identifies the similarity coefficient $SC(Q,D_i)$, for each document $1 \le i \le n$.*"

In the area of text retrieval many IR models have been proposed, implementing a retrieval strategy, such as the *Boolean* model, the *probabilistic* model, the *fuzzy set* model, the *Bayesian network* model, and the popular *vector-space* model. More detailed information on the different topics can be found in [10, 9, 11, 12].

The parallel between text retrieval strategies and an approach that may also be successful for spatial information retrieval is that the relevance of a document is based on the occurrences of query terms. For spatial information retrieval a similar approach can be followed. Here a spatial footprint is derived for each of the spatial objects identified in the information request. The relevance of a document is then determined by searching for corresponding footprints in the documents. However, this view would be a little too simplistic, since a match in text retrieval is based on equality, while for spatial information retrieval the match can be based on a number of different spatial similarity measures (operations).

## 5.2. Image retrieval

Image retrieval systems [7] form a typical example of IR systems, where, based on relevance feedback, the user's information need is specified. Although some image retrieval systems use annotation-based queries to start up the query process, relevance feedback is often used to refine an initial query. Following a content-based approach the visual content of images is extracted using color distributions (for example HSB- or RGB-histograms), textures, shapes, and region detection [13, 14].

Similar images are retrieved by the image retrieval system, using "query-by-example". The initial query is given one or more `relevant' images, which are used to find close matches of images. Images are judged similar, if they share the same characteristics in their features. Well-known examples of image retrieval systems are VisualSEEK [15], IBM's QBIC [16], and Blobworld [17].

Again a parallel can be drawn with spatial information retrieval. The different features that are extracted from an image can be seen as different similarity measures. Whereas text retrieval uses only the equality measure to determine the relevance of a document, image retrieval is already using a combination of similarity measures to determine the relevance. Another interesting aspect is the relevance feedback. The next section discusses the problem of uniquely identifying the location of a spatial object. Spatial information retrieval can benefit from relevance feedback, to avoid ambiguity at the query formulation side of the spatially-aware search engine [18].

### 5.3. Spatial ontologies and gazetteers

A commonly used approach in Information Retrieval to enhance the recall of a retrieval system is to use ontologies [29]. These large encyclopedias are based on a structured knowledge base containing information about a certain domain of expertise. A gazetteer [30] is a specialized ontology in the spatial domain. Not only could it contain synonyms of spatial objects and locations, it also contains the geographical position of the spatial object in question, usually represented by a coordinate of the earth's surface. Gazetteers provide the crucial connection between spatial objects in their different thematic representations. The geographical ontology in SPIRIT, as described in 2.1, can be regarded as a form of gazetteer, and in the case of the SPIRIT ontology it provides the potential to implement similarity measures, developed in the context of conventional (non-spatial) ontologies, as well as spatial similarity measures that could employ both Euclidean and topological data. In [30] some similarity measures based on a geographical ontology were presented. These included a simple Euclidean (great circle) distance metric in combination with measures based on spatial containment hierarchies (counting non-common parents) and path distances within a thematic semantic net (thesaurus).

## 6. Usability in spatially-aware search engines

Search engines for the Internet, including the spatially-aware search engine that is to be built in the SPIRIT project, basically consist of two parts: an indexer, and a retrieval engine. In the case of SPIRIT both parts become more complex, because we are not only focusing on the textual analysis, but also on the spatial information contained in the Web-documents. Figure 1 depicts a simplified overview of the process that is followed for the SPIRIT search engine. More detailed overviews of the architecture of the SPIRIT search engine can be found in [19].

The indexing process as shown here, uses the Web-crawler to fetch the documents, and initiate the indexing process. For text-indexing a standard procedure is followed, while for the spatial indexing a more complex procedure is followed as discussed below. The indexer extracts meta-data from the web-documents and uses a spatial ontology [20] to identify, enrich and validate the spatial information that is extracted and used for building the spatial index.

The right-hand side of the picture shows the retrieval process. An information request is formulated and is send to the query rewriter, which is responsible for formulating the textual and spatial query. This time, the ontology is used to derive the spatial content of the query. The two sub-queries are executed, and the results sent to the relevance ranking module. There the result sets are combined and sent back to the user interface for presentation purposes.

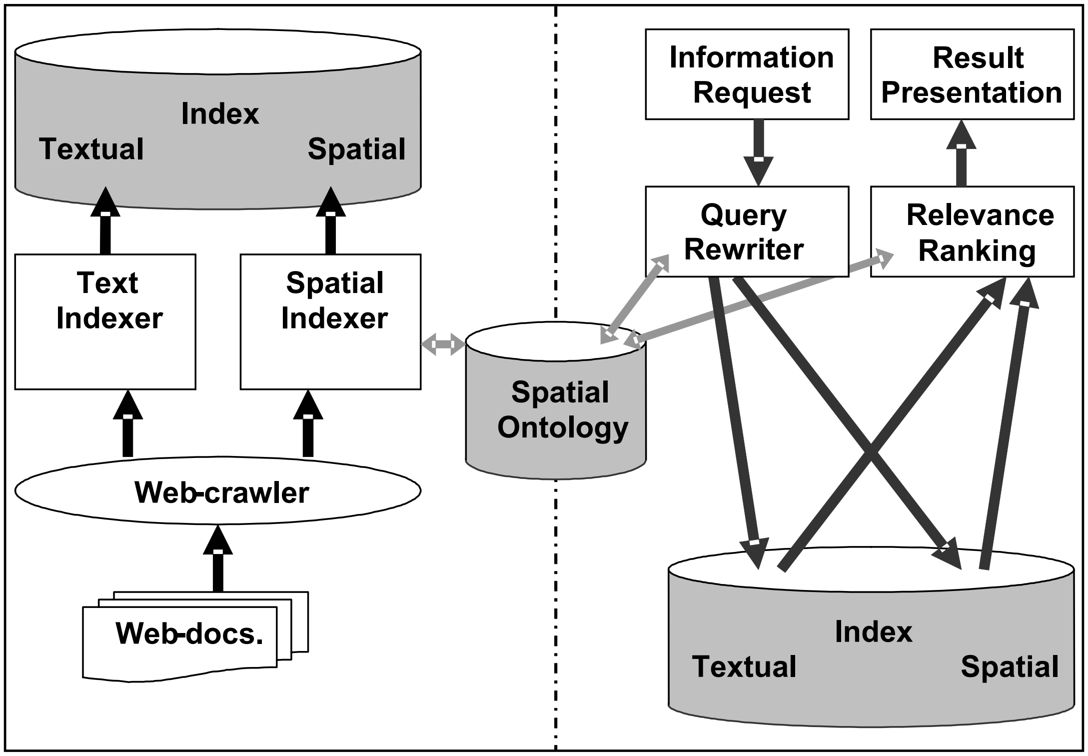


*Figure 1: Simplified architecture for spatially-aware search engines. (Left: indexing process; Right: retrieval process)*

The following aspects should not be neglected when building the index of a spatially-aware search engine:

- **Recognition of spatial objects**. To be able to identify spatial objects from a text-fragment, the indexer should have access to a spatial ontology [20, 19, 21], or gazetteer [24]. A match between a term, or more likely a sequence of terms, derived from the Web-document, and an entry in the spatial ontology should lead to recognition of spatial objects from Web-documents.
- **Recognition of spatial relationships**. The task of recognizing spatial relationships between spatial objects in a text document is even harder, and requires more GIS functionality to be integrated in the indexing process [23]. Furthermore, full-text analysis techniques are needed to identify spatial relationships from for instance a sentence, or a paragraph in the document. These kind of techniques are far more expensive, in terms of processing time and used resources, than the indexing techniques commonly used for text retrieval. There a Web-document can be seen as a bag of terms, without having to bother about the structure of the document.
- **Incompleteness/semi-structuredness of Web-data**. Contrary to the spatial information normally used in a GIS, spatial information obtained from Web-documents can be considered to be semi-structured [25, 12]. It does not have a single or explicit structure, has a weak notion of typing and is often incomplete.
- **Avoiding spatial misinterpretations/ambiguity**. Spatial misinterpretation at the object level can be easily introduced due to ambiguous naming of spatial objects. Consider the name “San Francisco”, while searching the Getty-gazetteer [24]. There turn out to be over a hundred entries in the gazetteer. It is therefore crucial to pinpoint potential locations to exactly one place. One can imagine that it is rather awkward if one intends to go to Sydney, Australia, but ends up in Sydney, Canada.
- **Costs of resolving spatial objects and relationships.** At present the size of the WWW is still growing rapidly. But only 20 to 60 percent of the WWW is indexed by search engines, using simple text-indexing techniques that are already time

consuming. To be able to build a spatial index for a reasonable part of the Internet, it is required that the techniques used are not too expensive in terms of processing time and resources.

- **Lack of focus in the spatial domain**. Another problem arising is the lack of focus when building a spatially-oriented index for searching the Internet. To be successful, access to a spatial ontology or other GIS resources is necessary. These resources are often limited to a certain part of the world, or cover only a few thematic topics. Information on the Internet however, targets every possible topic, or spatial domain.

When focusing on the retrieval side of a spatially-aware search engine other aspects come to light that influence the usability of spatial similarity measures. In general, a query follows the same path as a document, to derive one or more spatial footprints [20] from the query terms. A document is judged similar in the spatial context, if the set of spatial footprints of a query match the spatial footprints derived from a document. This matching is based on the available similarity measures that again depend on the spatial information extracted from the Web-documents, the underlying spatial data-structure and the spatial index. Also at the retrieval side, time and resources should be dealt with in a careful manner. The average Web-user is not accustomed to wait for `ages' for an answer, while in GIS the analysis of spatial information can be a time consuming process.

Based on the above mentioned aspects that play a role in building a spatially-aware search engine the following remarks can be made with respect to the usability of the spatial operations and measures:

- Similarity measures that depend on a complex underlying data-structure such as needed for the operations qualified under topology are unlikely to be usable in the Web-context, since such a structure is not available at a world-wide scale.
- Operations in the Euclidean space are most likely to be useful, since they can be calculated in a relatively small amount of time, and provide good measures.
- Assuming that a limited set of topological relations are stored in the geographical ontology then there is also the potential to develop and employ similarity (or distance) measures based on hierarchical and lattice structures composed by the topological relations.
- Measures in the General and Set-oriented space can also be useful. However these operations often provide an exact (Boolean) answer to the similarity between two spatial objects, instead of a relevance measure. This makes them less useful, when it comes to determining a relevance ranking for the set of documents matching the query based on their spatial content. These measures can be adopted to obtain a non-Boolean score.

## 7. Conclusions

The aim of this report is to investigate the aspects of spatial similarity measures that are useful when building a spatially-aware search engine for the Internet [28]. Two approaches can be followed for identifying the necessary spatial similarity measures and aspects that play a role. Contrary to the approach followed in [18], where a user-driven approach is followed to specify the user requirements for the SPIRIT search engine, the angle followed here is that of a data-driven one. The task of indexing large quantities of the Internet and the internals of the retrieval engine, do not allow a user (end-user or administrator) to intervene.

Based on the classification of spatial objects and operations, a first selection of usable similarity measures is made. Spatial objects and the operations embedded in the Euclidean space provide for the basic support of spatial similarity measures during the indexing and retrieval process.

Objects and operations defined in the General and Set-oriented spaces are also useful. Especially during the extraction of spatial information from Web-documents similarity measures defined in the general space will prove crucial, in combination with the spatial ontology and other GIS-related systems.

Due to the immense size of the Internet, the complexity of the spatial similarity measures should be minimized in terms of processing time and used resources. At this stage of the project, it is hard to predict which similarity measures should be incorporated into the search engine. In Section 6 the usability of similarity measures is discussed, to give a first indication. Two major factors can be distinguished that influence the usability. First of al it depends on the richness of the data contained in the spatial ontology. Second, it depends on the structure of footprints and the spatial index that is built during the indexing stage.

## 8. References

[1] B. D. Dent. Cartography Thematic Map Design. C. Brown Publishers, Dubuque, IA, 1996.

[2] Christopher J. Kocmoud Donald H. House. Constructing continuous cartograms. In proceedings of IEEE Visualization, Research Park, NC, USA, October 1998.

[3] R. Weibel J.-C. Muller, J.-P. Lagrange, editor. GIS and Generalization: Methodology and Practice. Taylor and Francis, London, 1995.

[4] John C. Gallant John P. Wilson, editor. Terrain Analysis: Principles and Applications. Wiley europe, August 2000. ISBN: 0-471-32188-5.

[5] R. Laurini and D. Thompson. Fundamentals of Spatial Information Systems. Taylor and Francis, Londen Great Brittain, 1992.

[6] Christopher B. Jones Robert Weibel. Computational perspectives on map generalization. GeoInformatica, 2(4):307-314, 1998.

[7] Remco C. Veltkamp and Mirela Tanase. Content-Based Image and Video Retrieval, chapter A Survey of Content-Based Image Retrieval Systems, pages 47-101. Kluwer, 2002.

[8] Micheal F. Worboys. GIS: a computing perspective. Taylor & Francis, London, United Kingdom, 1995. ISBN 0-7484-0064-8.

[9] D.A.Grossman and O. Frieder. Information Retrieval: Algorithms and Heuristics. Kluwer international series in engineering and computer science. Kluwer Academic Publishers, 1998. ISSN ISBN: 0-7923-8271-4.

[10] R. Baeza-Yates and B. Ribeiro-Neto. Modern Information Retrieval. ACM Press, 1999. ISBN ISSN: 0-201-39829-X.

[11] D. Hiemstra. Using Language Models for Information Retrieval. PhD thesis, Centre for Telematics and Information Technology, Enschede, the Netherlands, Januari 2001.

[12] R. van Zwol. Modelling and search web-based document collections. PhD thesis, Centre for Telematics and Information Technology, Enschede, the Netherlands. April 2002. ISBN: 90-365-1721-4.

[13] P. Aigrain, H. Zhang, and D. Petkovic. Content based representation and retrieval of visual media: A state-of-the-art review. Multimedia Tools and Applications, 3(3):179-202, 1996.

[14] Venkat N. Gudivada and Vijay V. Raghavan. Content-based image retrieval systems. IEEE Computer, 28(9):18-22, 1995.

[15] John R. Smith and Shih-Fu Chang. VisualSEEk: A fully automated content-based image query system. In proceedings of ACM Multimedia, pages 87-98, 1996.

[16] M. Flickner, H. Sawhney, and W. Niblack. Query by image and video content: The QBIC system. IEEE Computer Magazine, 28:23-32, September 1995.

[17] C. Carson, M. Thomas, S. Belongie, J.M. Hellerstein, and J. Malik. Blobworld: a system for region-based image indexing and retrieval. In proceedings of the 3rd International Conference on Visual Information and Information Systems, pages 509-516, Amsterdam, the Netherlands, 1999.

[18] S. Balley, B. Bucher, D. Petrelli, A. Ruas, M. van Kreveld, M. Sanderson, M. Sester. User requirements Specification. SPIRIT project, E.U.: IST-2001-35047 (D3:7101).

[19] Dave Finch and SPIRIT Technical Group. Specification of System Functionality. SPIRIT project, E.U.: IST-2001-35047 (D4:1101).

[20] Gaihua Fu, Alia Abdelmoty and Christopher Jones. Design of Geographical Ontology. SPIRIT project, E.U.: IST-2001-35047 (D5:3101).

[21] H. Alani, C.B. Jones, D. Tudhope. Associative and Spatial Relationships in Thesaurus-based Retrieval. In Research and Advanced Technology for Digital Libraries, 4th European Conference, ECDL., pages 45-58. Springer. 2000. ISBN: 3-540-41023-6.

[22] M. Egenhofer and R. Franzosa. Point-Set Topological Spatial Relations. In International Journal of Geographical Information Systems. 5(2):161-174, 1991

[23] Yohsuke Inoue , Ryong Lee, Hiroki Takakura, Yahiko Kambayashi. Web Page Ranking Based on Geographical Location and Link Structure. In proceedings of the International Conference on Internet Information Retrieval. November 2002.

[24] The Getty Thesaurus of Geographic Names. The Getty Research Institute. LA, USA. http://www.getty.edu/ research/tools/vocabulary/tgn/.

[25] S. Abiteboul. Querying semi-structured data. In proceedings of the International Conference on Database Theory (ICDT), 1997.

[26] Vincent B. Robinson. A Perspective on Geographic Information Systems and Fuzzy Sets. In proceedings of North American Fuzzy Information Processing Society, IEEE, New Orleans, LA, 2002, p. 1-6.

[27] Max J. Egenhofer. Query Processing in Spatial-Query-by-Sketch. Journal of Visual Languages and Computing. 8(4):403-424. 1997.

[28] Jones C.B. et al. Spatial Information Retrieval and Geographical Ontologies: An Overview of the SPIRIT project in SIGIR 2002: Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval August 11-15, 2002, Tampere, Finland, ACM Press, pp.387 - 388.

[29] F. Fonseca, M. Egenhofer, and C. Davis. Ontology-Driven Information Integration. In proceedings of AAAI-2000 Workshop on Spatial and Temporal Granularity. Austin, Texas, USA. Pages 61-64. August 2000.

[30] Christopher B. Jones, Harith Alani, Douglas Tudhope. Geographical Information Retrieval with Ontologies of Place. In proceedings of COSIT: Spatial Information Theory: Foundations of Geographic Information Science, pages 322 – 335. Lecture Notes in Computer Science. 2001.

[31] Ian Heywood, Sarah Cornelius, Steve Carver. An Introduction to Geographical Information Systems. Prentice Hall. Edinburgh Gate, Harlow, England. ISBN: 0130 611980. 1998.

[32] Micheal F. Worboys. GIS: a computing perspective. Taylor & Francis. London, United Kingdom. ISBN 0-7484-0064-8. 1995

[33] Cristopher Jones. Geographical Information Systems and Computer Cartography. Addison Wesley Longman Limited. Edinburgh Gate, Harlow, England. ISBN: 0 582 04439 1. 1997.

[34] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge. Comparing images using the Hausdorff distance. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, pp. 850-863, September 1993.

[35] W. J. Rucklidge. Efficiently locating objects using the Hausdorff distance. International Journal of Computer Vision, vol. 24, no. 3, pp. 251-270, 1997.

[36] Remco C. Veltkamp, Michiel Hagedoorn. State-of-the-art in shape matching. Principles of Visual Information Retrieval, M. Lew (ed.), Springer, 2001, ISBN 1-85233-381-2, 87-119.

[37] Remco C. Veltkamp, Mirela Tanase, Danialle Sent. Features in Content-Based Image Retrieval Systems: A Survey. Remco C. Veltkamp and Hans Burkhardt and Hans-Peter Kriegel (eds.), State-of-the-Art in Content-Based Image and Video Retrieval. Kluwer, 2001, pages 97-124.