

# ***SPIRIT:***

## **Spatially-aware Information Retrieval on the Internet**

<http://www.geo-spirit.org/>

**Roelof van Zwol**

**roelof@cs.uu.nl**

<http://www.cs.uu.nl/~roelof>

**Marc van Kreveld**

**Avi Arampatzis**

**Iris Reinbacher**

**Remco Veltkamp**

**Utrecht University**

**Institute of Information and Computing Sciences**

**GIVE: Center for Geometry, Imaging and Virtual Environments**

# Overview

- ▶ **About SPIRIT.**
- ▶ **Motivation for SPIRIT.**
- ▶ **SPIRIT architecture in a nutshell.**
- ▶ **How to build a spatial footprint?**
- ▶ **How to build a spatial index?**
- ▶ **Relevance ranking in SPIRIT.**
- ▶ **Work in progress.**

# About *SPIRIT*

## ▶ EC fifth framework programme

### ▪ Project partners (involvement):

- \* Cardiff University (ontology design)
- \* University of Hannover (meta-data extraction, sketching)
- \* Institut Geographique National, IGN (meta-data storage, evaluation)
- \* University of Sheffield (core search engine)
- \* University of Zurich (user Interface)
- \* Utrecht University (relevance ranking)

# ***Motivation for SPIRIT***

## **▶ Searching the Internet:**

- Type keywords.
- Lack all spatial orientation/knowledge.

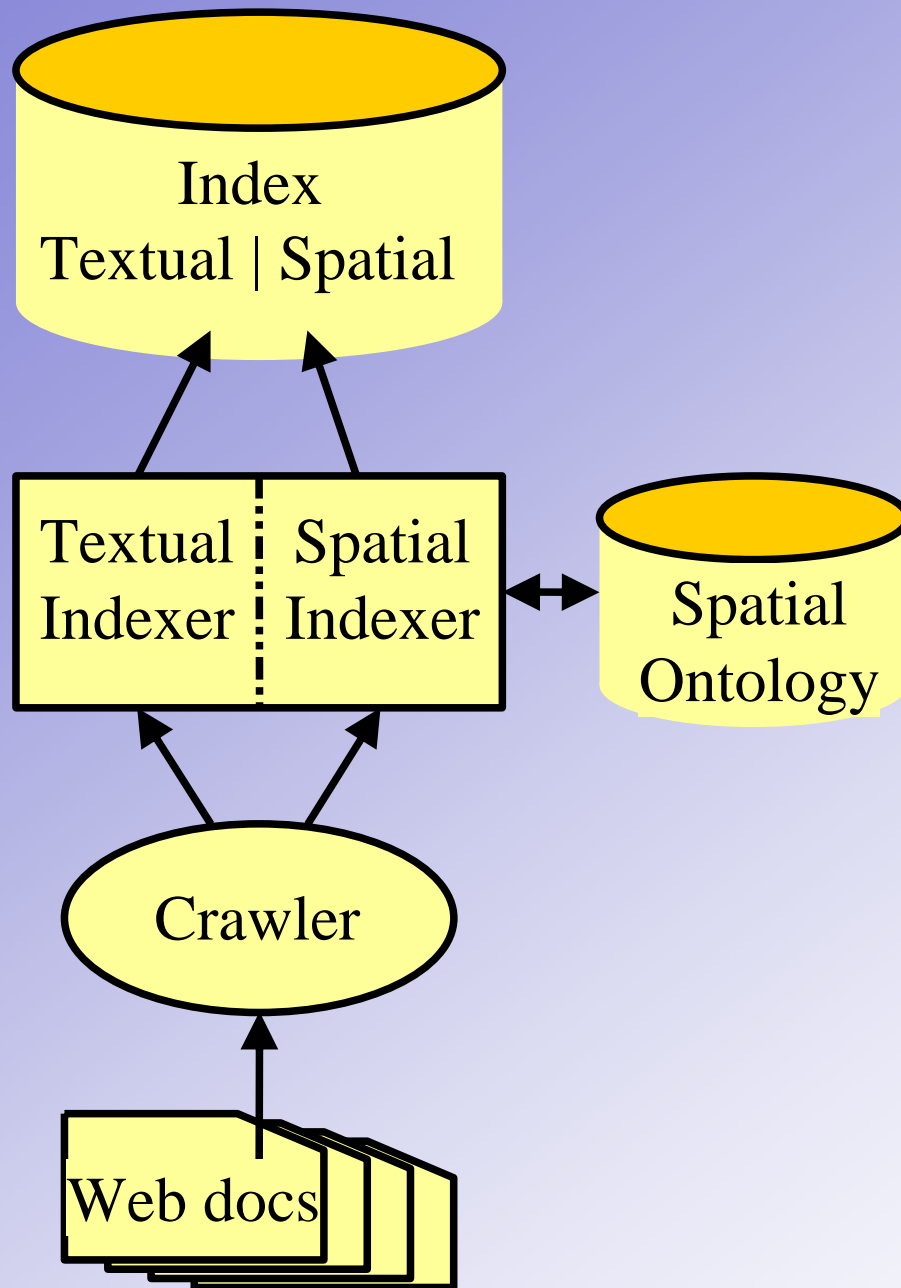
## **▶ The goal in SPIRIT:**

- Use spatial information contained in Web-documents, to improve the retrieval process.

## **▶ Research issues:**

- Extraction of spatial information (meta-data handling).
- Spatial reasoning, using ontologies.
- Spatial indexing.
- Query formulation.
- Relevance ranking over multiple indices.

# ***SPIRIT Architecture in a nutshell***



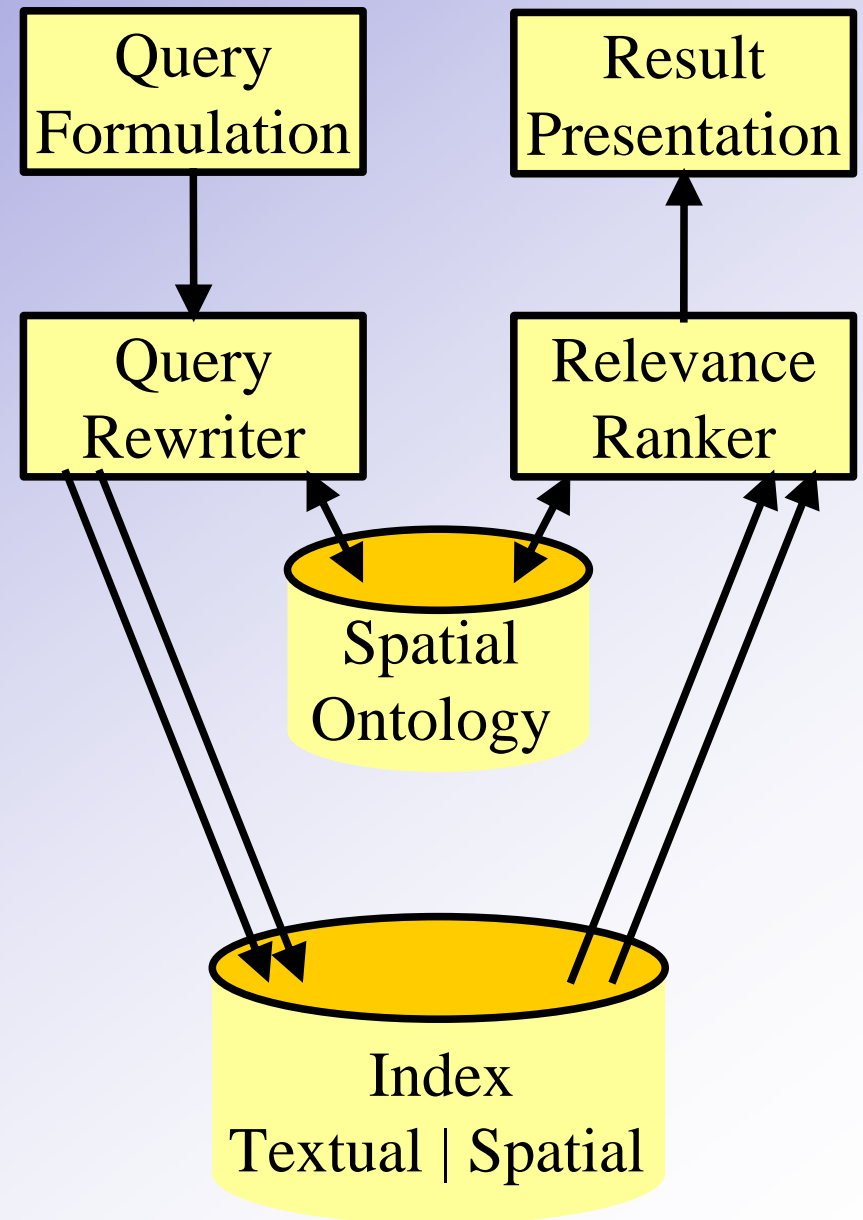
## ► **Spatial indexing:**

- Crawl for documents.
- Pre-process the web documents.
- Derive a set of spatial footprints for each web document.
- Build the spatial index.

# ***SPIRIT Architecture in a nutshell***

## ► **Spatial Information Retrieval Engine (SPIRE):**

- Query formulation (information request).
- Query rewriter.
- Determine spatial footprint for query.
- Spatial search & textual search.
- Relevance ranking.
- Result presentation.



# How to build a spatial footprint???

Netherlands - Off the beaten track

## Wadden Islands

...~~Texel~~ is the largest and most populated island. It has **24km of beaches** and hosts the world's largest **catamaran race** in June.

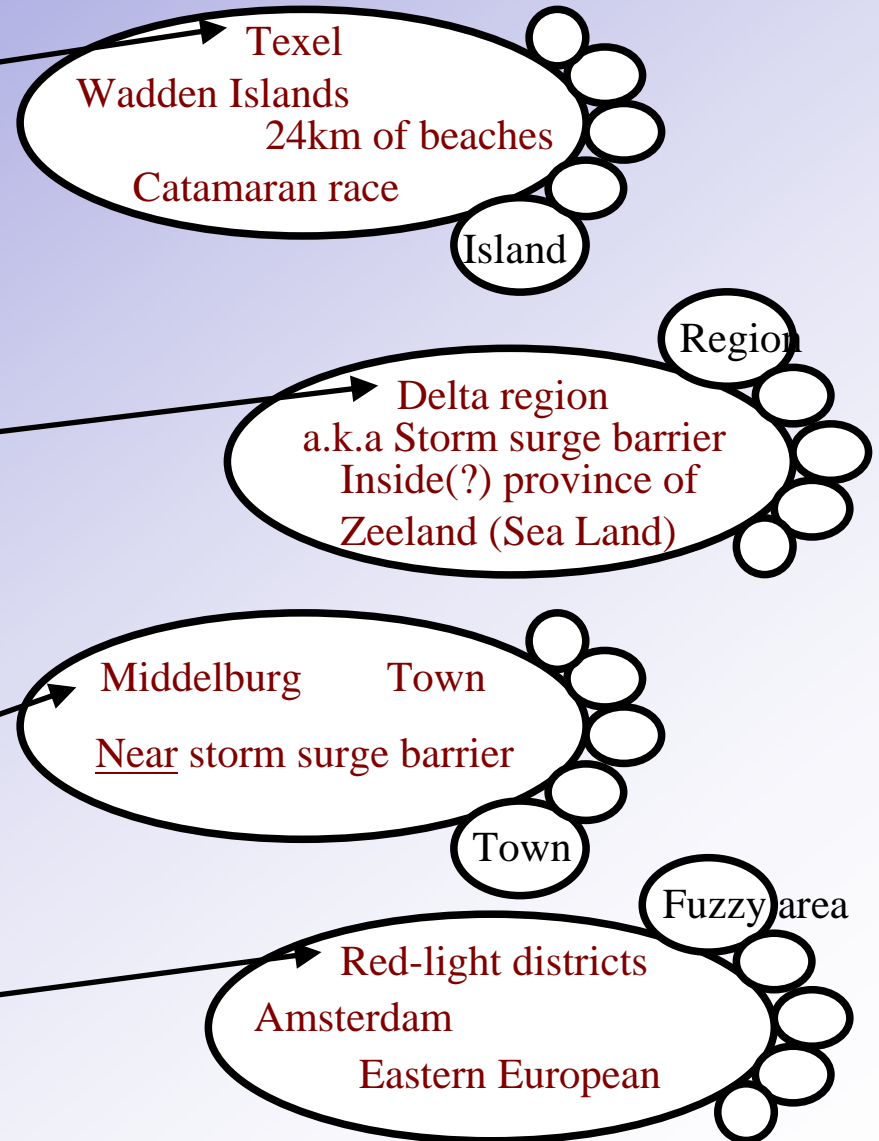
## Delta region

...The Netherland's aptly named province of **Zeeland (Sea Land)** makes up most of the **Delta region**.

...The **Delta Expo** sits steadfastly on top of the remarkable **3.2km storm surge barrier near** the quaint **town** of **Middelburg**.

## Red-light districts

Amsterdam's red-light area is (in) famous, ...young **Eastern European**.



# How to build a spatial index???

## Netherlands

Off the beaten track

## Wadden Islands

...Texel is the largest and most populated island - it has 24km of beaches and hosts the world's largest catamaran race in June. ...

## Delta region

The Netherlands aptly named province of Zeeland (Sea Land) makes up most of the Delta region. ...

...The Delta Expo sits steadfastly on top of the remarkable 3.2km storm surge barrier near the quaint town of Middelburg.

## Red-light districts

Amsterdam's red-light area is (in) famous, ...young Eastern European.





# Relevance ranking in SPIRIT

## ▶ Example:

- cheap *hotel*
- *Near Amsterdam*

## ▶ Questions:

- How to combine the spatial results with the textual results?
- What about the distribution of term relevance (either spatial or textual...)

## ▶ Just an idea - Produce a combined ranking that:

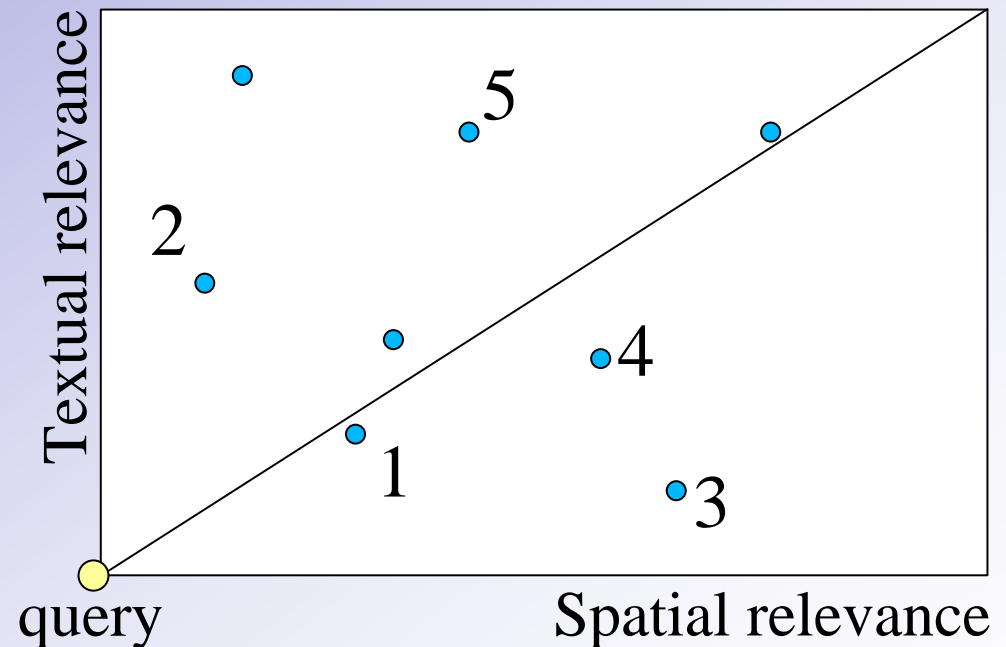
- minimizes the distance to the query,
- but maximizes the relative angle between documents.

# Relevance ranking (cont'd)

## ► Using the relative angle:

- Visualize spatial relevance vs. textual relevance.
- 'Force' a diversity in the ranking of documents given a set of query terms (both spatial and textual).

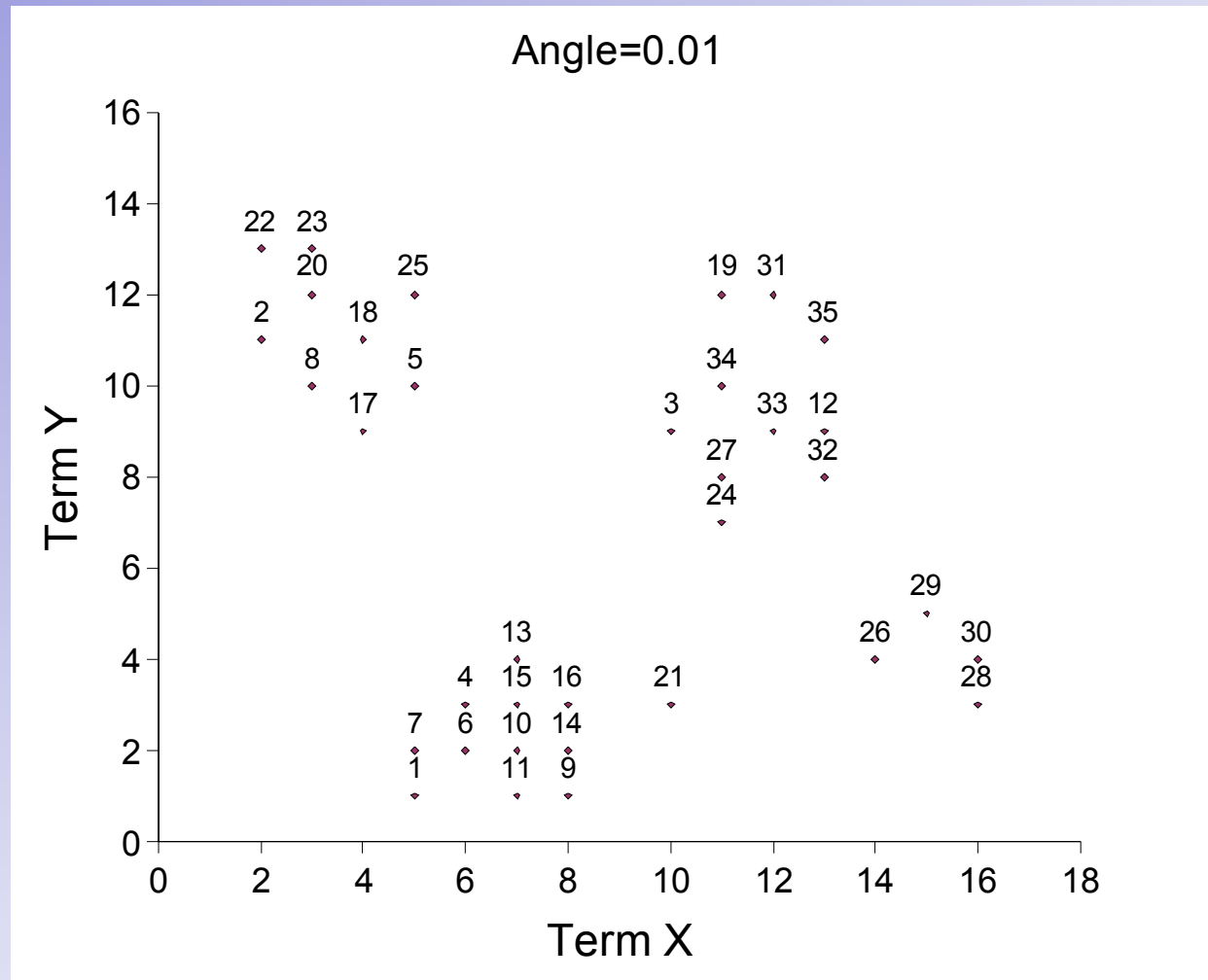
## ► Is it reasonable to combine the results of two different indices in a single ranking?



# Relevance Ranking (cont'd)

## ► Rel. Angle Spreadings-function:

- $(\text{Rel\_Angle} + \text{Ca}) / (\text{Dist} + \text{Cd})$



# *Relevance Ranking (cont'd)*

**Experimental Search Engine:**

# SPIRE

---

cheap hotel

Standard Ranking  Spreaded ranking

SPIRE search

# Relevance Ranking (cont'd)

## Results (cheap hotel; using relative angle):

1. [http://huins/lonelyplanet/letters/eur/mal\\_pc.htm](http://huins/lonelyplanet/letters/eur/mal_pc.htm)  
(1) [*cheap*=15.7358888416812] [*hotel*=32.8547491539902]
2. [http://huins/lonelyplanet/destinations/north\\_america/bermuda/money.htm](http://huins/lonelyplanet/destinations/north_america/bermuda/money.htm)  
(6) [*cheap*=81.1227387036586] [*hotel*=20.3252032520325]
3. <http://huins/lonelyplanet/destinations/caribbean/aruba/money.htm>  
(19) [*cheap*=43.7387919345668] [*hotel*=36.5283459964933]
4. [http://huins/lonelyplanet/destinations/caribbean/puerto\\_rico/money.htm](http://huins/lonelyplanet/destinations/caribbean/puerto_rico/money.htm)  
(5) [*cheap*=46.9153178512784] [*hotel*=23.5083924961211]
5. <http://huins/lonelyplanet/destinations/europe/slovakia/money.htm>  
(8) [*cheap*=30.4618009016693] [*hotel*=38.1606563632894]
6. <http://huins/lonelyplanet/destinations/africa/lesotho/money.htm>  
(28) [*cheap*=27.8559291345163] [*hotel*=104.690117252931]

# ***Work in progress***

- ▶ **Realize the integration of the spatial component.**
- ▶ **Retrieval Performance Experiment:**
  - **Lonely Planet document collection.**
  - **Compare two search engines:**
    - \* **Basic search engine.**
    - \* **SPIRE, using min. dist / max. spreading.**
- ▶ **Another Retrieval Performance Experiment:**
  - **1 Tera-byte Web-collection**
  - **Compare two search engines:**
    - \* **Basic search engine.**
    - \* **SPIRE, using an intelligent algorithm for multimedia relevance ranking. (Combines the result of the independent indices available).**