



A Spatial Dimension for Searching the World Wide Web

M. Andrea Rodríguez T.

Centro de Investigación de la Web - CIW

Universidad de Concepción

andrea@udec.cl

December -2002

Outline

- Search based on spatial operators
- Reasoning about spatial relations
- Modeling the geographic space in the Web
- Mapping Web documents onto a Geographic Map
- The Geographic Map as Topic Map in the Web
- Conclusions and future work

Diversity in Type of Information

The screenshot shows the CNN.com homepage with a main article titled "11 dead in Jerusalem terror". The article text reads: "A suicide bomber killed 11 people and injured 50 -- including many schoolchildren -- aboard a crowded Israeli commuter bus in Jerusalem this morning, according to Israeli police. The militant Islamic group Hamas has claimed responsibility for the terror attack, which Jerusalem's mayor calls 'something that you have to see to understand how terrible it is.'" Below the article is a "FULL STORY" link and a "Survivor: Everything was bloody" video link. The page also features a "MORE TOP STORIES" section with links to "American missionary shot to death in Lebanon", "Indonesia nabs alleged Bali bomb mastermind", "At least 19 dead as quake hits northern Pakistan", "Experimental vaccine could wipe out cervical cancer", and "Law suits: McDonalds to blame for kids' obesity". A "BUSINESS at CNN/MONEY" section is visible at the bottom left. The top navigation bar includes "Home Page", "World", "U.S.", "Weather", "Business at CNN.com", "Sports at CNN.com", "Politics", "Law", "Technology", "Science & Space", "Health", "Entertainment", "Travel", "Education", and "Special Reports".

Text

Images

Video

Audio

Diversity in Type of Information

The screenshot shows the CNN.com homepage with a main article titled "11 dead in Jerusalem terror". The article text reads: "A suicide bomber killed 11 people and injured 50 -- including many schoolchildren -- aboard a crowded Israeli commuter bus in Jerusalem this morning, according to Israeli police. The militant Islamic group Hamas has claimed responsibility for the terror attack, which Jerusalem's mayor calls 'something that you have to see to understand how terrible it is.'" Below the article is a "FULL STORY" link and a "Video" link with a thumbnail of a man in a "POLICE" cap. To the right, there are sections for "ON CNN TV" featuring "Lou Dobbs: The Blue Chip of Financial News" and "Two U.S. Soldiers Shot in Kuwait". A "MORE TOP STORIES" section lists various international news items. At the bottom, there are "BUSINESS at CNNMoney" and "STOCK/FUND QUOTES" sections.

Text

Images

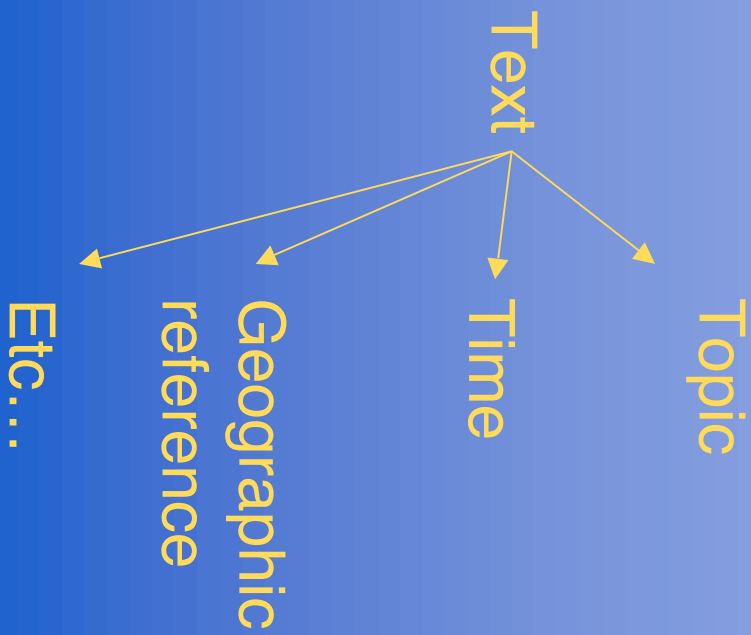
Video

Audio

Diversity in Information Content

The screenshot shows the CNN.com homepage with the following content:

- Navigation:** Home Page, World, U.S., Weather, Business, Sports, Law, Technology, Science & Space, Health, Entertainment, Travel, Education, Special Reports.
- Search:** SEARCH CNN.COM with a search bar and a "GO" button.
- Language:** Updated: 10:44 a.m. EST (10:44 GMT) November 21, 2002. Language: English on CNN TV. Transcription: CNN International. About CNN.com. Preferences.
- Main Headline:** **11 dead in Jerusalem terror**. A suicide bomber killed 11 people and injured 50 -- including many schoolchildren -- aboard a morning bus in Jerusalem, according to Israeli police. The militant Islamic group Hamas has claimed responsibility for the terror attack, which Jerusalem's mayor calls "something that you have to see to understand how terrible it is."
- Other News Items:**
 - ON CNN TV:** Come, China Tomlin! (8 p.m. ET). What is life like for sniper suspect John Lee Malvo? Step inside the adult jail where this juvenile awaits trial.
 - TWO U.S. SOLDIERS SHOT IN KUWAIT:** Kuwait says a junior police officer shot the pair -- now in serious condition -- then fled across the border to Saudi Arabia. Full Story
 - MORE TOP STORIES:**
 - American missionary struck to death in Lebanon
 - Bush wins NATO support on Iraq
 - Indonesia man's alleged Bali bombing mastermind
 - At least 19 dead as quake hits northern Pakistan
 - Experimental vaccine could wipe out cervical cancer
 - Law suit: McDonald's to blame for kids' obesity
 - CNN Money: Drug companies plan male contraceptive pill
 - BUSINESS at MONEY:** STOCK/FUND QUOTES: MARKETS: 10:42 a.m. ET 1121.89. DOW: +10.43 8724.43. NASDAQ: +1.68.



Diversity in Information Content

The screenshot shows the CNN.com homepage with a navigation bar at the top. The main content area features a large article titled "11 dead in Jerusalem terror" with a photograph of a young boy being held by a police officer. To the right, there are several smaller news items under the heading "ON CNN TV" and "TWO U.S. SOLDIERS SHOT IN KUWAIT". Below these, a "MORE TOP STORIES" section lists various international news items. At the bottom, there is a "BUSINESS" section with market quotes for the S&P 500 and Dow Jones.

Text

Topic

Time

Geographic
reference

Etc...

Queries based on spatial operators

- Retrieve hotels in Pucón or adjacent towns
- Retrieve ski resorts located within 100 Km from Santiago
- Retrieve airports and harbors in Chile that are within 100 km from each other

Retrieve hotels in Pucón and adjacent towns

Keyword: Hotel and (Pucón or Villarrica or Caburga or...)

Retrieve hotels in Pucón and adjacent towns

Keyword: Hotel and (Pucón or Villarrica or Caburga or...)

```
Query: Select      d.url
From              Document d
Where             SUCH THAT d.MENTIONS "hotels"
                  (d.G_reference IN "Pucón" or
                   d.G_reference EQUAL "Pucón" or
                   d.G_reference ADJACENT "Pucón" )
```

Retrieve ski resorts located within 100 kms
from Santiago

Keyword: Ski resorts and (Santiago or Rancagua or

Retrieve ski resorts located within 100 kms from Santiago

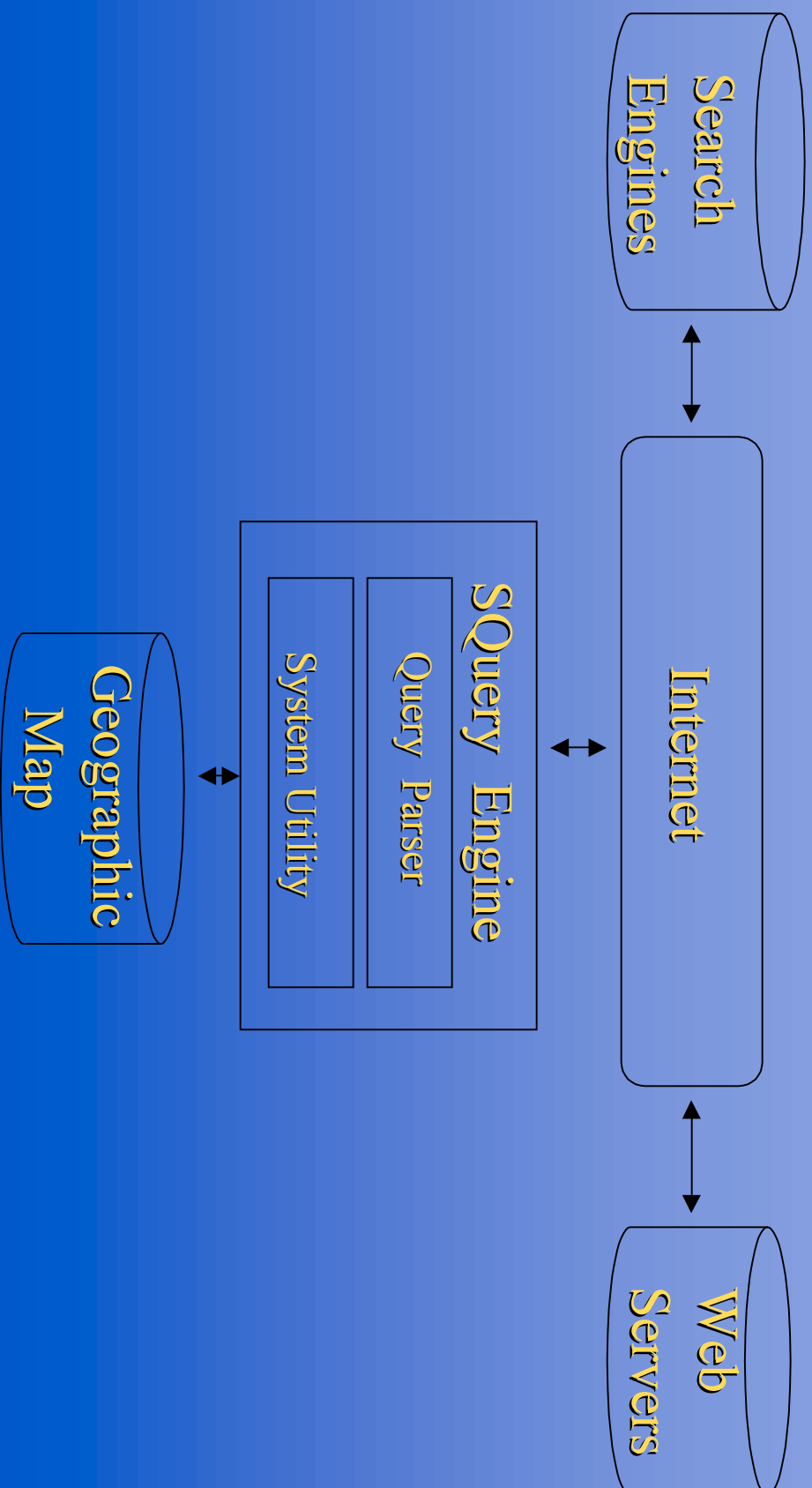
Keyword: Ski resorts and (Santiago or Rancagua or

```
Query: Select  d.url
          From  Document d
                SUCH THAT d.MENTIONS "ski resort"
          Where  Distance(d.G_reference, "Santiago") ≤ 100
```

Retrieve airports and harbors in Chile
that are within 100 kms from each other

```
Query: Select  d1.url, d2.url
          From    Document d1
                SUCH THAT d1 MENTIONS "airport",
                Document d2
                SUCH THAT d2 MENTIONS "harbor",
          Where  Distance(d1.G_reference, d2.G_reference) <= 100
                and (d1.G_reference IN "Chile")
                and (d2.G_reference IN "Chile")
```

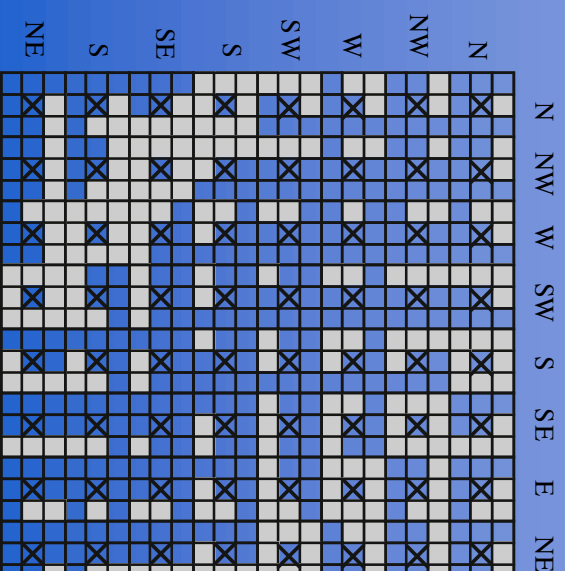
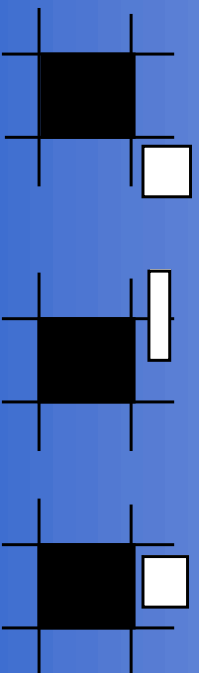
Seacrch System Architecture



Reasoning about Spatial Relations

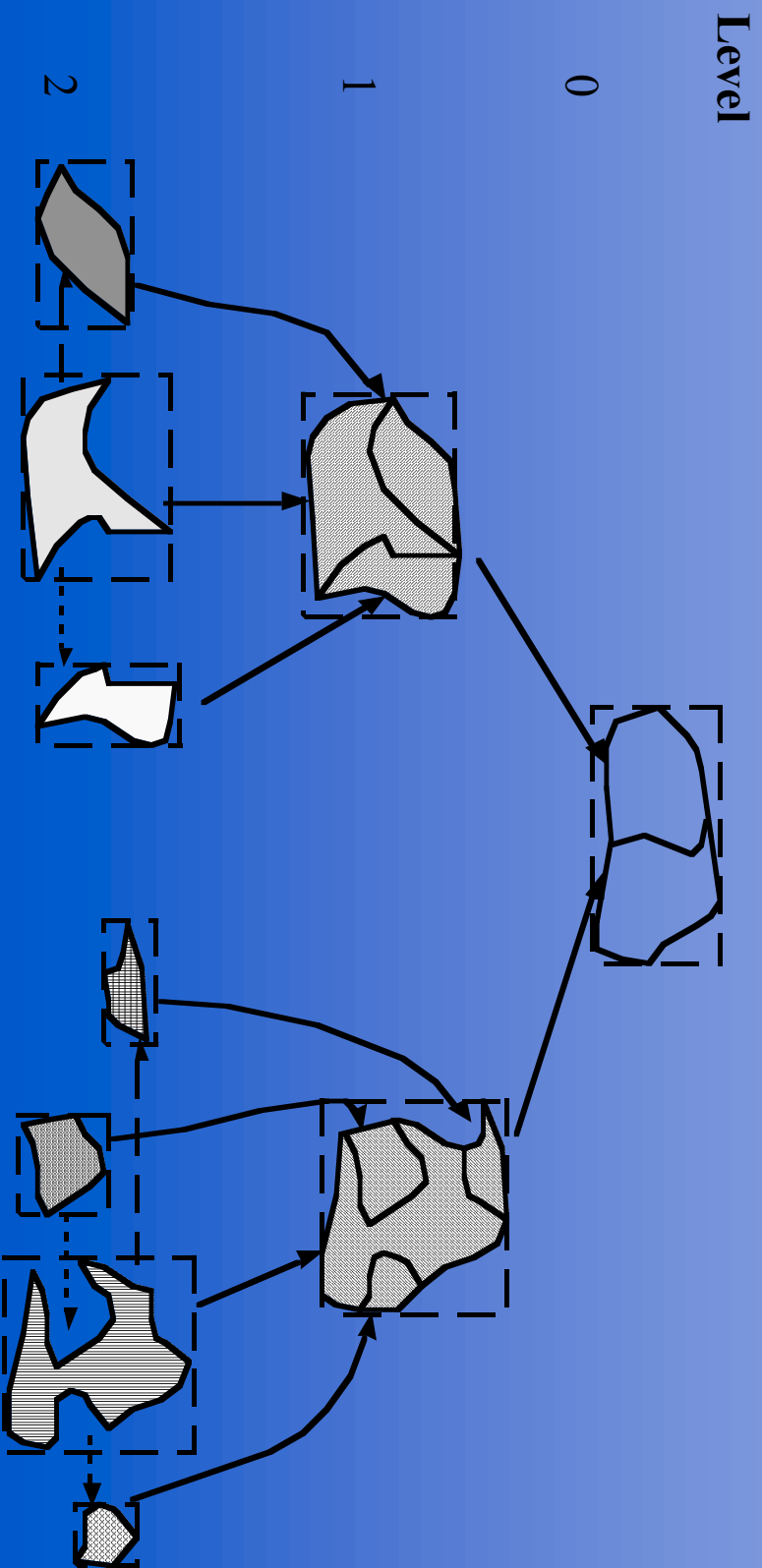
- Spatial relations define constraints in structural queries
- They are usually not stored in the database
- Their derivation and manipulation lay in the area of spatial reasoning

Orientation

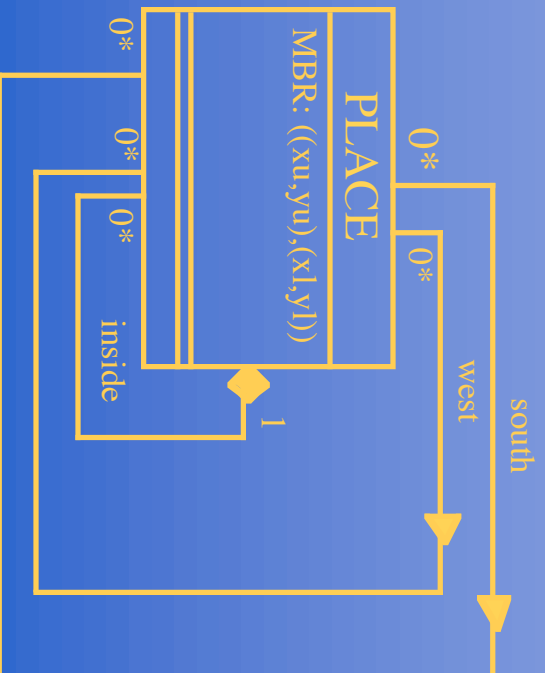


| | | |
|----|---|----|
| NW | N | NE |
| W | X | E |
| SW | S | SE |

Modeling Regions and Spatial Relations



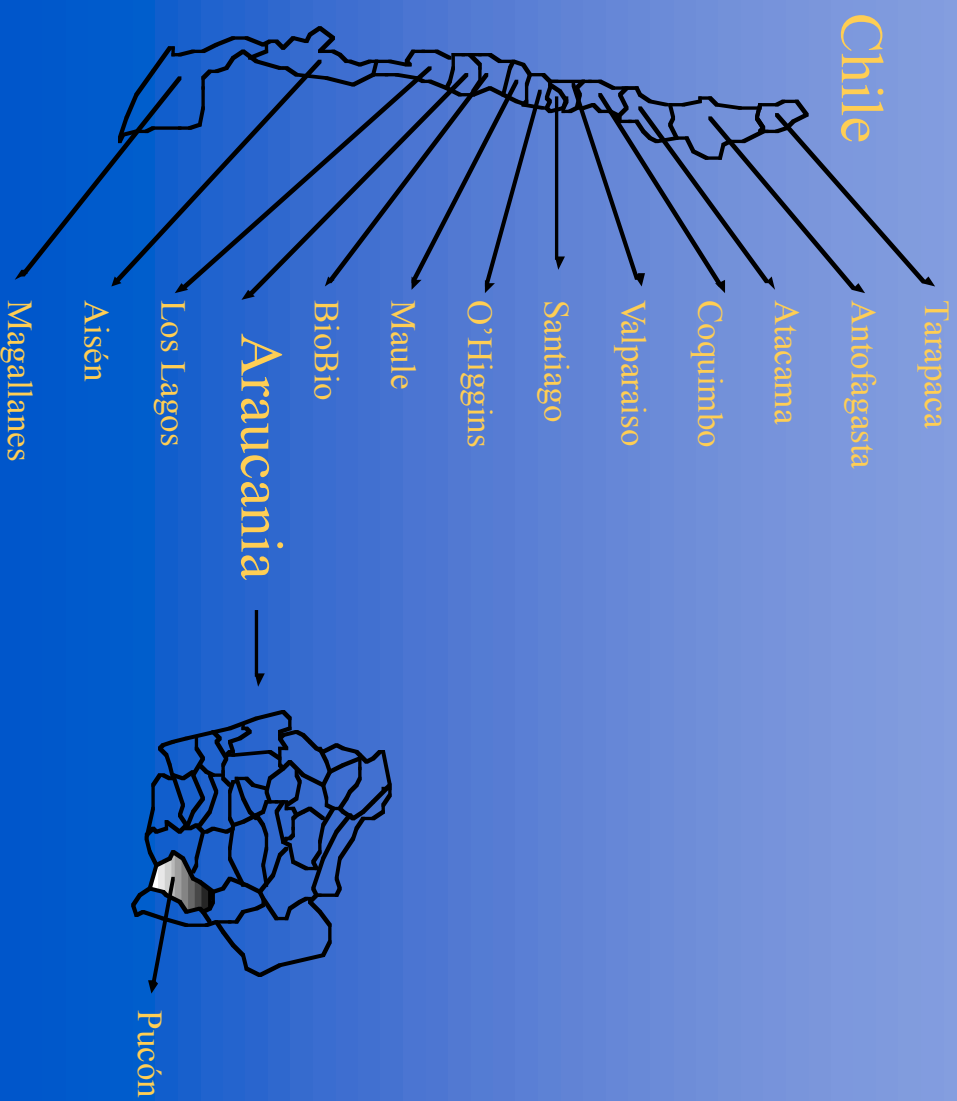
Modelling Regions and Spatial Relations



$(\square r_1 \neq r_2) \textit{inside}(r_1, r_2) \equiv \textit{contains}(r_2, r_1)$

$(\square r_1 \neq r_2) \textit{meet}(r_1, r_2) \equiv (\square x_1, x_2) (\textit{meet}(x_1, x_2) \square \textit{inside}(x_1, r_1) \square \textit{inside}(x_2, r_2))$

A Geographic Map of Chile



Mapping Web Documents onto the Geographic Map

- Metadata-based mapping (ISO/TC 211)

```
<spatReplInfo>  
  <geodDec> Name </geodDesc>  
</geobox>  
  <GeoBndBox>  
    <westBL> </westBL>  
    <eastBL> </eastBL>  
    <northBL> </northBL>  
    <southBL> </southBL>  
  </GeoBndBox>  
</geobox>  
</spatReplInfo>
```

Mapping Web Documents onto the Geographic Map

- Keyword-based mapping
 - Occurrence and co-occurrence of toponymies
 - Occurrence within the geographic hierarchy
 - Co-occurrence of toponimies and geographic entities

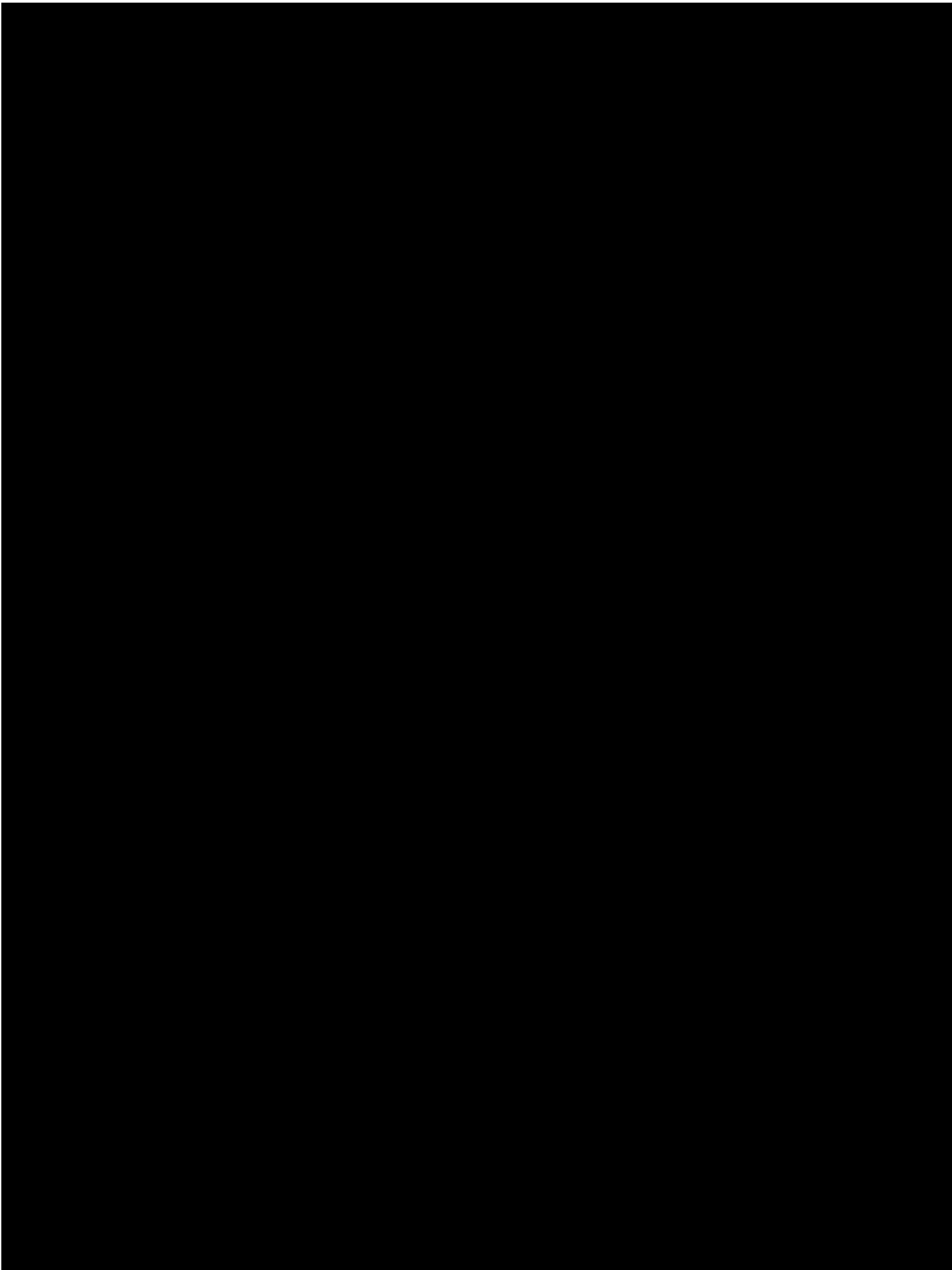
Geographic Map as Topic Map

- Topic Maps model the content of Web documents independently of Web occurrences
- It exists as an standard ISO/IEC 13250:1999
- Basic components:
 - Topics and types of topics
 - Associations and types of associations
 - Occurrences

Geographic Map as Topic Map

```
<topic id="chiletm-topic">
  <subjectIdentity>
    <subjectIndicatorRef xlink:href="#chiletm"/>
  </subjectIdentity>
  <baseName>
    <baseNameString>Chilean Map</baseNameString>
  </baseName>
</topic>

<association>
  <instanceOf><topicRef xlink:href="ontopsi.xtmm#superclass-subclass"/></instanceOf>
  <scope><topicRef xlink:href="ontopsi.xtmm#ontology"/></scope>
  <member>
    <roleSpec><topicRef xlink:href="ontopsi.xtmm#superclass"/></roleSpec>
    <topicRef xlink:href="#place"/>
  </member>
  <member>
    <roleSpec><topicRef xlink:href="ontopsi.xtmm#subclass"/></roleSpec>
    <topicRef xlink:href="#country"/>
  </member>
</association>
```



Conclusions and Future Work

- Queries based on spatial operations from textual information
- Extensions of current spatial query languages
- Indexing and search algorithms
- Spatial and Semantic Constraints