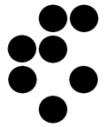
Using Personalized PageRank for Keyword Based Sensor Retrieval

Lorand Dali
Alexandra Moraru
Dunja Mladenić

Jožef Stefan Institue
Slovenia

ailab.ijs.si

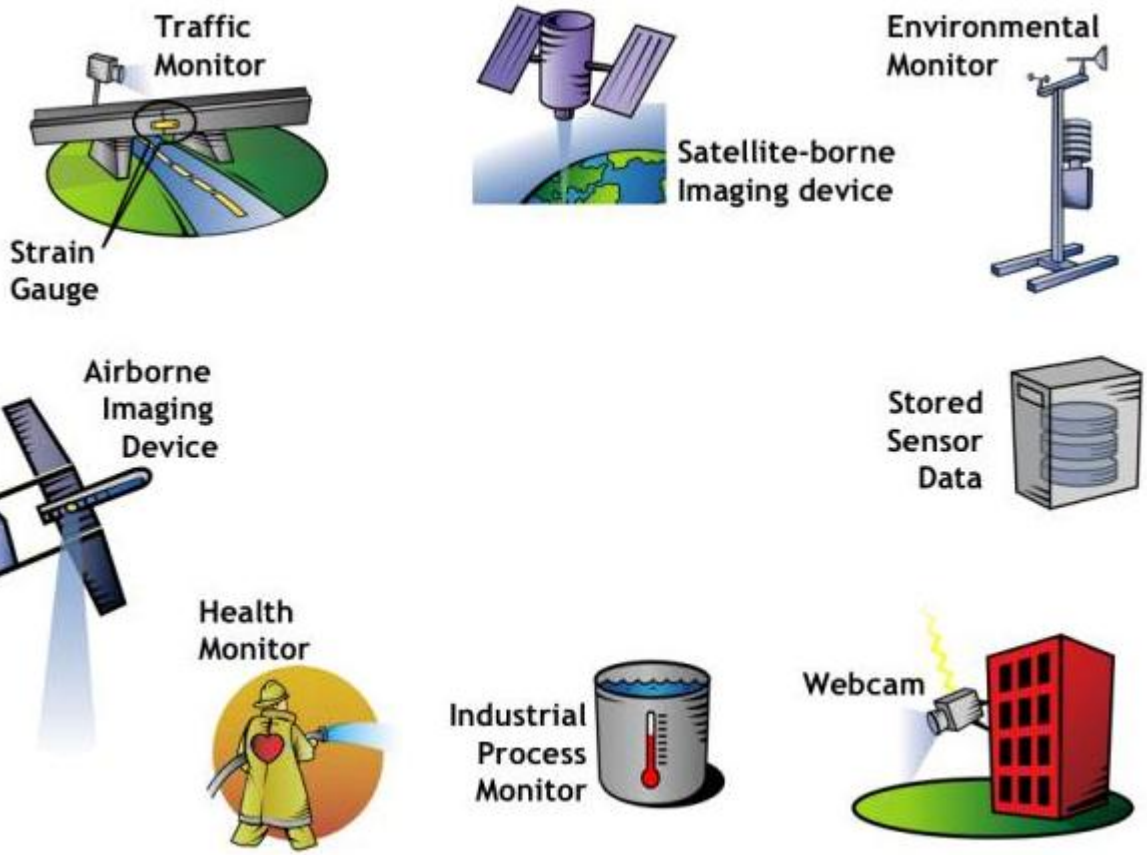


Outline

- Motivation
- Problem Description
- Data Description
- System Architecture
- Search and Ranking
- Demo
- Conclusions

Motivation

Sensors are everywhere!



Source: M. Botts, G. Percivall, C.Reed, J. Davidson, OGC SWE: Overview And High Level Architecture

Motivation

- Sensor Web – OGC Vision

  - *Web accessible sensor networks and archived sensor data that can be discovered and accessed using standard protocols and application program interfaces*

  - Sensors will be able to
    - report position
    - connected to the web
    - register metadata
    - be readable and controllable remotely

Source: M. Botts, G. Percivall, C.Reed, J. Davidson, OGC SWE: Overview And High Level Architecture

Motivation

- *We want a way of doing sensing that can make the data available to any application that needs that specific data*[1]

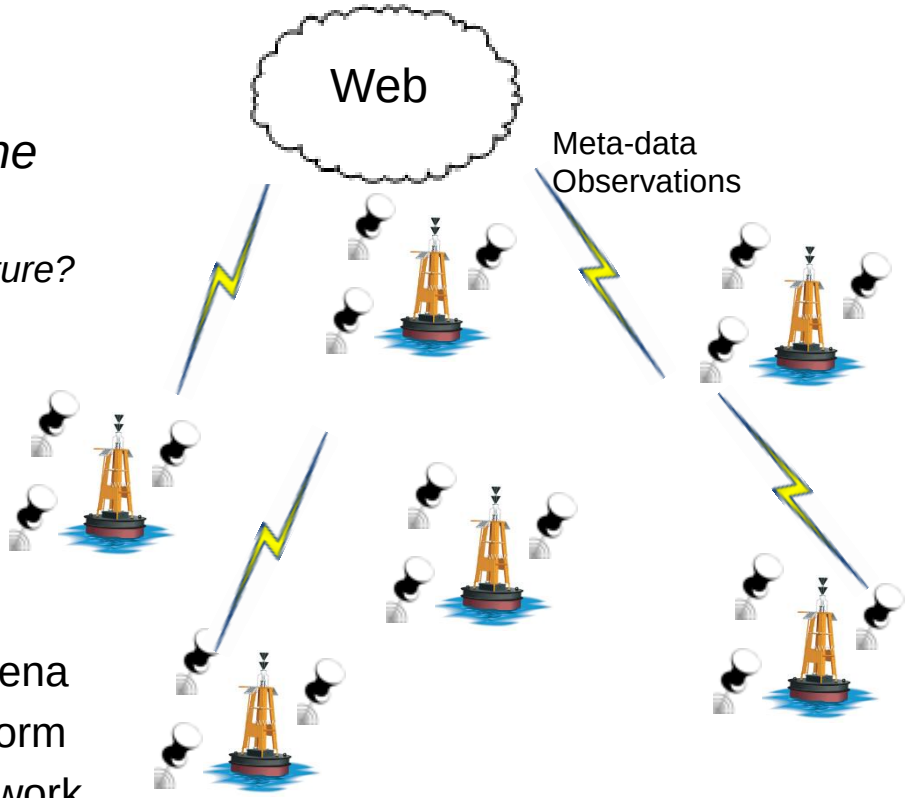
- How do we search for these data?

[1]John Cox, Turning the world into a sensor network, NetworkWorld, August 11, 2010.

Problem Description

- System for **keyword based sensor search**, apply the Personalized PageRank algorithm for **ranking**, and filtered results based on **geo-location**.

- *What is the water temperature in the costal region of Goa?*
  - *What about wind, currents, air temperature?*



- Search and ranking criteria:
  - Textual description extracted from sensor's metadata
  - Sensors measuring same phenomena
  - Sensors located on the same platform
  - Sensors deployed in the same network

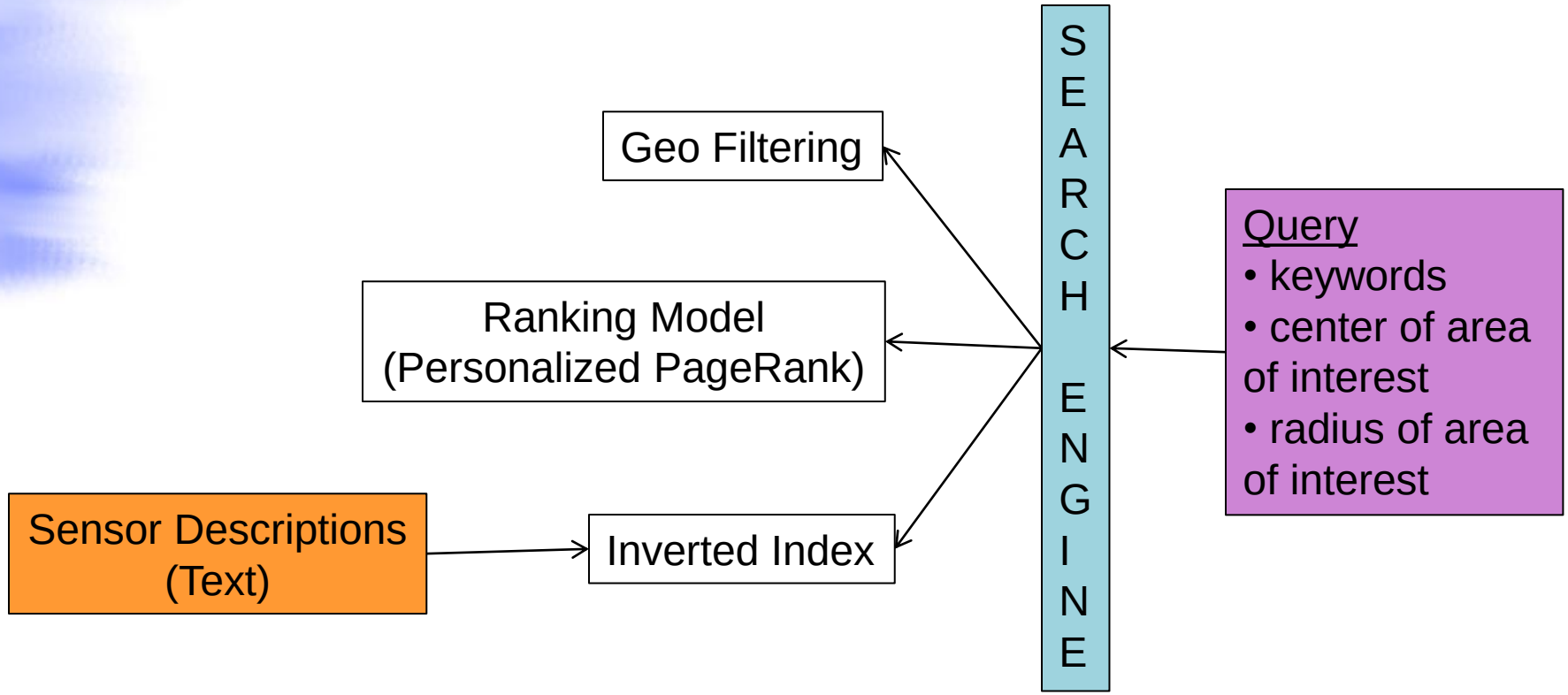
Data Description

- Sensors in the area of ocean tides and currents, defined by[1]:
  - networks
  - platforms
  - sensors
  - observed property

- Large number of standardized sensor descriptions
  - the representation format is SensorML, facilitating parsing and extraction of relevant metadata.
  - each sensor can observe one property (i.e. air temperature, water salinity, etc.) and is attached to one platform;
  - each platform is deployed in one network and can have one or more sensors attached.
  - each platform is given the latitude and longitude for its location.

[1]Center for Operational Oceanographic Products and Services, http://tidesandcurrents.noaa.gov/index.shtml

System Architecture



Search and Ranking

- The goal of the search
  - retrieve and rank a list of sensors based on the user's request
  - Input:
    - keyword query
    - geographic location (given by latitude and longitude coordinates)
    - distance (interpreted as a radius around the location)
  - Output:
    - list of ranked sensors

- Text descriptions taken into consideration for keyword search:
  - platform, sensor and property **names**, given by system owners
  - standard name and definition of the **property observed**
    - From Climate and Forecast standard names parameter vocabulary. (MMI ontology)
  - sensor **description given by owner**

Search and Ranking

- PageRank algorithm
  - query independent ranking of web pages
  - from a directed graph => scores for each of the nodes
  - based on random walk model

- Personalized PageRank
  - Query dependent
  - subset Q of nodes matched by the keyword search are important apriori
  - constraint on the jumps in the random walk model

Search and Ranking

- Equation for computing score:

$$\boldsymbol{p} = d \cdot \boldsymbol{M} \cdot \boldsymbol{p} + (1 - d) \cdot \boldsymbol{u}, \qquad \boldsymbol{p}, \boldsymbol{u} \in \mathbb{R}^n, \boldsymbol{M} \in \boldsymbol{\mathcal{M}}(n)$$

  - n is the number of nodes in the graph
  - $\boldsymbol{p}$ is the PageRank vector containing the score for each node and is initialized with 0
  - $\boldsymbol{M}$ is the transition matrix constructed in the following way:

$$\boldsymbol{M}[\boldsymbol{i}, \boldsymbol{j}] = \begin{cases} 5, & \boldsymbol{i}\ and\ \boldsymbol{j}\ measure\ the\ same\ thing \\ 4, & \boldsymbol{i}\ and\ \boldsymbol{j}\ are\ on\ the\ same\ platform \\ 1, & \boldsymbol{i}\ and\ \boldsymbol{j}\ are\ on\ the\ same\ deployment \\ 0, & otherwise \end{cases}$$

  - d is the damping parameter
  - $\boldsymbol{u}$ is the jump vector and its entries are $\boldsymbol{u}[i] = \frac{1}{n}, \forall i.$
    - constraint: $\boldsymbol{u}[i] = \frac{1}{|Q|}$ if $i \in Q$ and 0 otherwise.

Search and Ranking

- Geo-Filtering of search results
  - sensor scores are added to calculate platform scores
  - platform score is adjusted by dividing with it with the number of radiuses it is away from the location which the user has specified
  - Small radius => Very strict about location

Demo

Search Example

water level Search

Search results

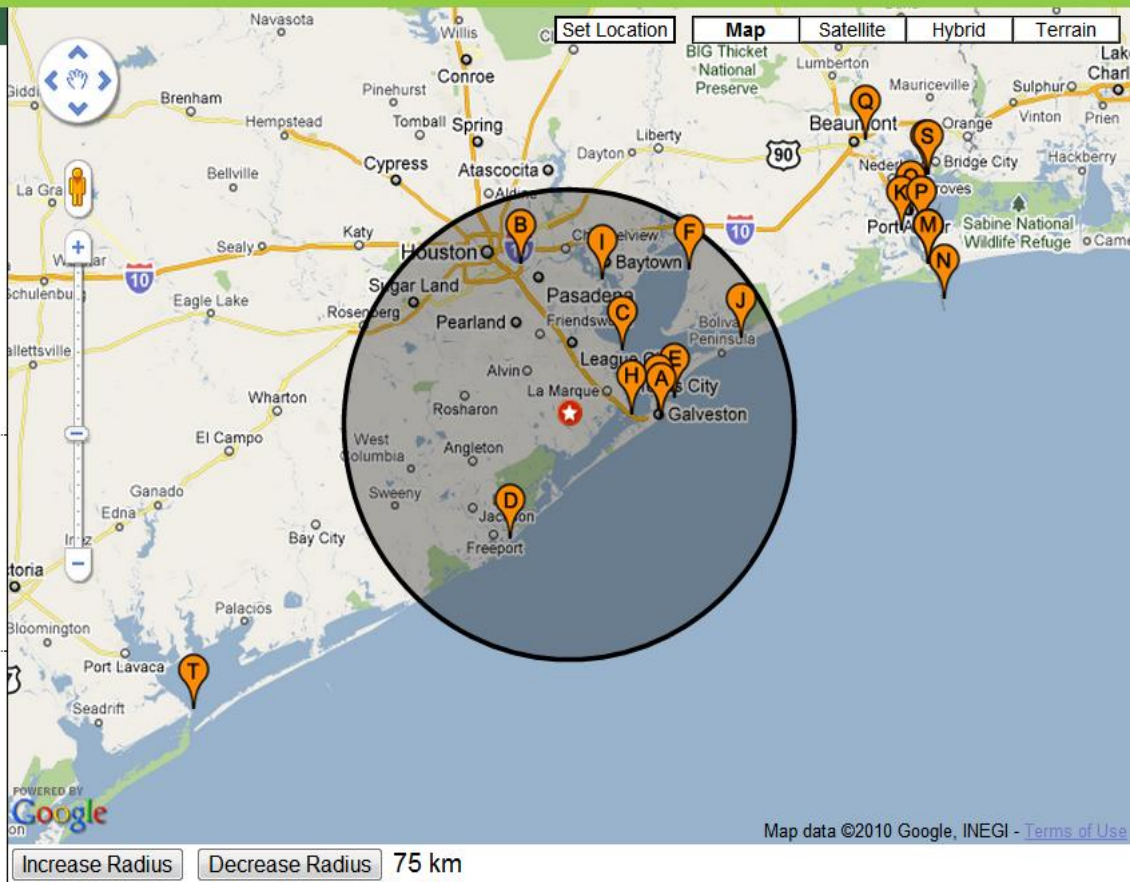
A **Galveston Pleasure Pier**
Station information for Galveston Pleasure Pier (8771510). Observed data: WaterLevel, WaterLevelPredictions, Winds, AirTemperature, WaterTemperature, BarometricPressure.
- sensor-WaterLevel - WaterLevel instrument for station 8771510
- sensor-WaterLevelPredictions - WaterLevelPredictions instrument for station 8771510
- sensor-Winds - Winds instrument for station 8771510
- sensor-Winds - Winds instrument for station 8771510
- sensor-Winds - Winds instrument for station 8771510
- sensor-AirTemperature - AirTemperature instrument for station 8771510
- sensor-WaterTemperature - WaterTemperature instrument for station 8771510
- sensor-BarometricPressure - BarometricPressure instrument for station 8771510

urn:x-noaa:def:station:NOAA.NOS.CO-OPS::8771510

B **Manchester**
Station information for Manchester (8770777). Observed data: WaterLevel, WaterLevelPredictions, WaterTemperature.
- sensor-WaterLevel - WaterLevel instrument for station 8770777
- sensor-WaterLevelPredictions - WaterLevelPredictions instrument for station 8770777
- sensor-WaterTemperature - WaterTemperature instrument for station 8770777

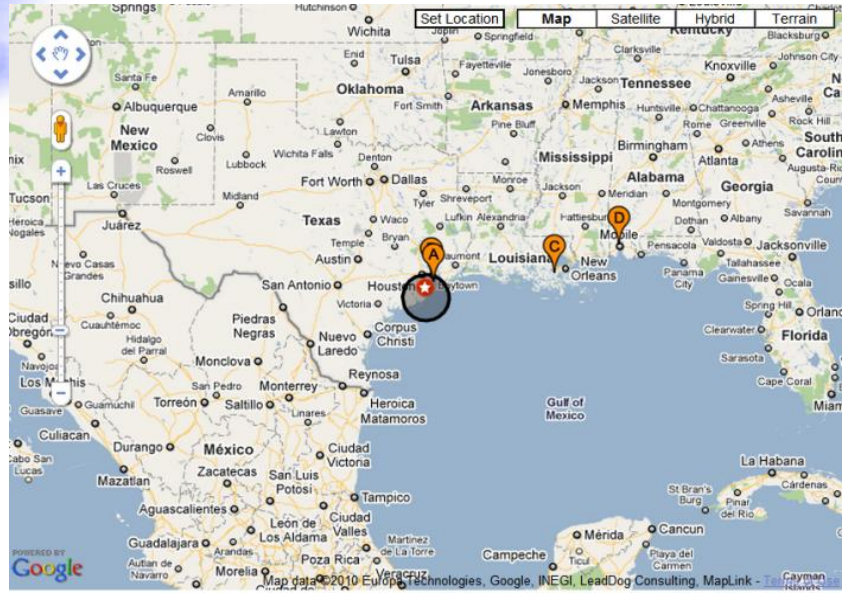
urn:x-noaa:def:station:NOAA.NOS.CO-OPS::8770777

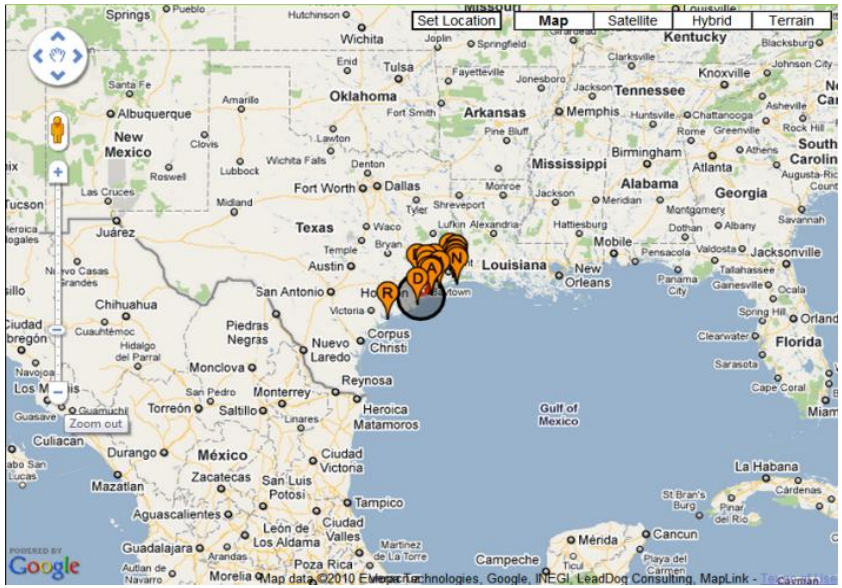
C **Eagle Point**
Station information for Eagle Point (8771013). Observed data: WaterLevel, WaterLevelPredictions, Winds, AirTemperature, WaterTemperature, BarometricPressure, Conductivity, Salinity.
- sensor-WaterLevel - WaterLevel instrument for station 8771013
- sensor-WaterLevelPredictions - WaterLevelPredictions instrument for station 8771013
- sensor-Winds - Winds instrument for station 8771013
- sensor-Winds - Winds instrument for station 8771013



Search Example

- Performing the proposed ranking results in obtaining more platforms closer to the area of interest
  - we consider relevant also sensors located on the same platform or those that are in the same deployment





Conclusions

- We need sensor search and ranking
- Personalized PageRank can be a solution to obtain the most relevant sensors

- Future work
  - Integrate more datasets
  - Evaluation
    - Obtain relevance feedback data
    - Determine the parameters in an empirical way
  - Considering measurements for search


THE END!