



Analysis and Retrieval of Multimodal Environmental Information for Personalized Decision Support

Stefanos Vrochidis



Informatics and Telematics Institute
Centre for Research and Technology Hellas

Outline of the Talk

- Introduction
- PESCADO project
- Use Case
- Discovery and Extraction of Environmental Information
- User Interaction Modes
- Conclusions

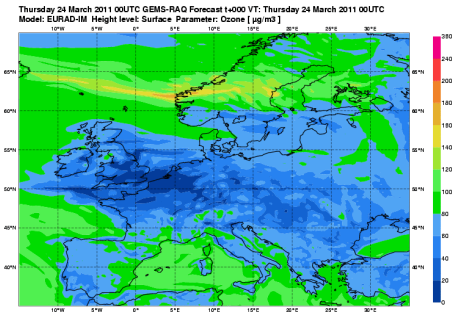
Introduction

- Nowadays:
 - well-established national air quality and meteorological networks offer a variety of environmental services (i.e. air pollution, weather forecast, pollen).
 - Only a few of the data providers are giving access to their raw forecast and observation data.
 - A number of initiatives has been undertaken towards this direction, however the issue of harmonized access to environmental information remains open.
- The situation is not expected to change in the near future..
- The need today..
 - Citizens are increasingly aware of the influence of **air quality** and **meteorological conditions** on the **quality of their life** and demand for high quality environmental information that is tailored to one's specific context, background and needs.
 - E.g. People are highly interested in Pollen forecast.
 - More than 15% of the worldwide population suffer from pollen allergy.



Environmental Nodes and Data

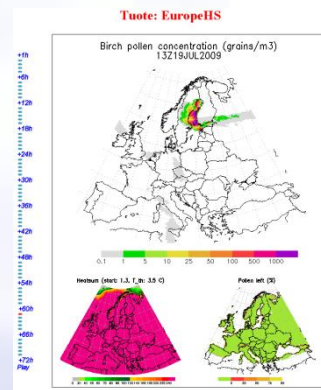
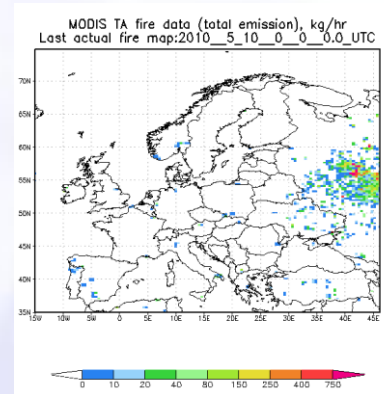
- Environmental Nodes
 - Websites with environmental information
 - Variety of encoding and representations
 - Several hundreds of websites
 - Usually free access
 - No detailed information of the covered area and stations employed
 - Multimodal Data
 - Free text
 - Text in tabular format
 - Images (heatmaps/graphs/icons)
 - Flash objects
 - Services with environmental information
 - Usually not freely accessible
 - They provide quality data
 - Detailed description of the stations is provided
 - Numerical structured data



Detailed forecast
Wed Thu Fri Sat Sun Mon Tue Wed Thu

WEDNESDAY May 19

06:00	+16°	19 km/h	Moderately cloudy Precip. chance 5% Humidity 74%
09:00	+16°	21 km/h	Moderately cloudy Precip. chance 5% Humidity 83%
12:00	+17°	28 km/h	Partly cloudy Precip. chance 5% Humidity 83%
15:00	+20°	28 km/h	Moderately cloudy Precip. chance 5% Humidity 83%
18:00	+19°	21 km/h	Partly cloudy Precip. chance 5% Humidity 84%
21:00	+17°	18 km/h	Cloudy Precip. chance 5% Humidity 83%



Intermeteo.com
Helsinki 00-03
+7, +9
+13, +15

10 day forecast

Today	Tomorrow	Fri May 21	Sat May 22	Sun May 23
Hi: +20° Lo: +14° ← 24 km/h	Hi: +20° Lo: +11° ← 12 km/h	Hi: +25° Lo: +9° ↗ 19 km/h	Hi: +19° Lo: +15° ↗ 13 km/h	Hi: +18° Lo: +11° ↗ 19 km/h
Mon May 24	Tue May 25	Wed May 26	Thu May 27	Fri May 28
Hi: +19° Lo: +11° ↘ 22 km/h	Hi: +16° Lo: +8° ↘ 22 km/h	Hi: +15° Lo: +10° ↘ 27 km/h	Hi: +15° Lo: +10° ↘ 21 km/h	Hi: +15° Lo: +6° ↗ 30 km/h

AIR QUALITY IN FINLAND

Measured data
Measured data
12.02.2010 15:00



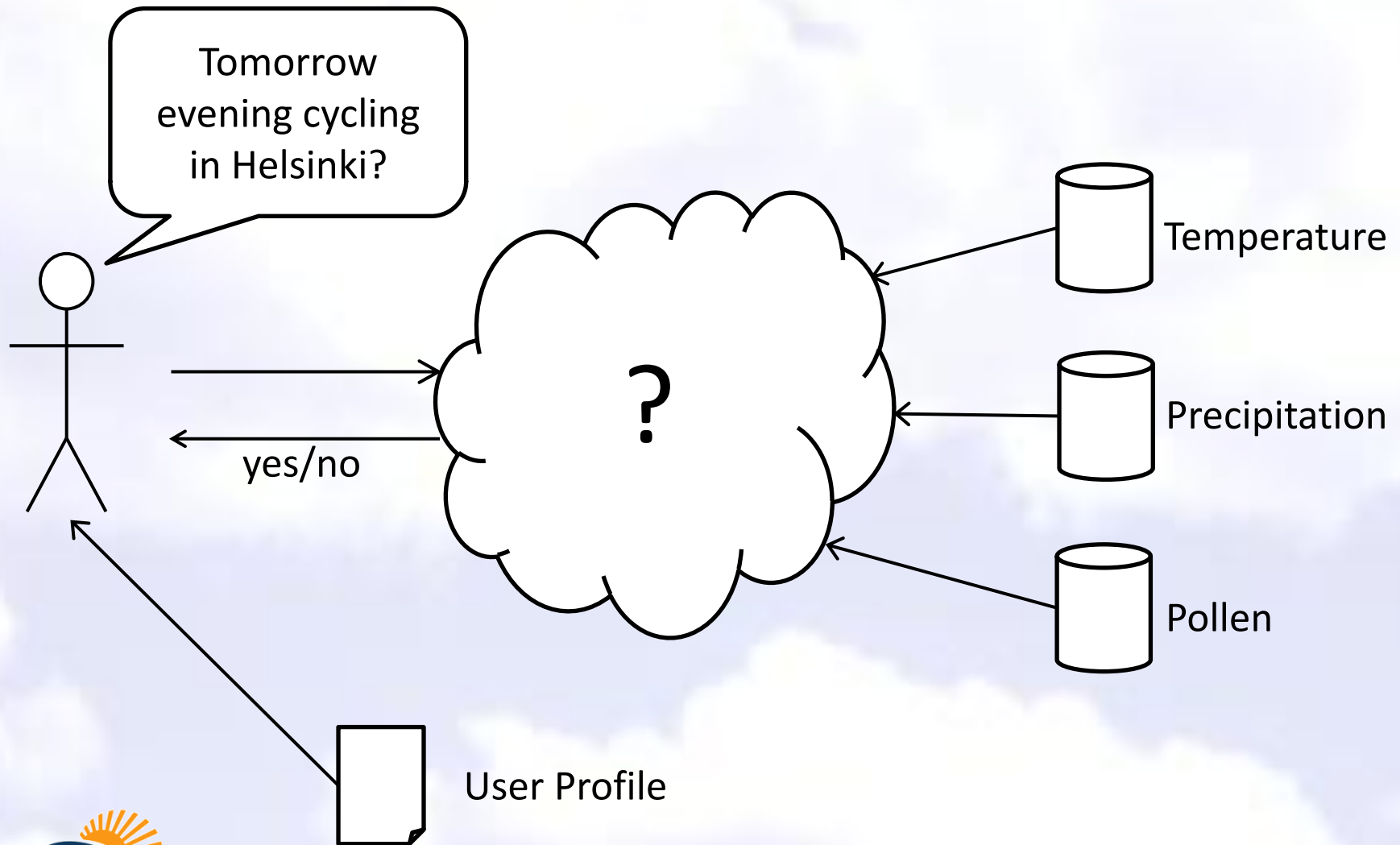
PESCaDO Project

- There is an increasing need for the orchestration of environmental nodes to provide users with personalized services supporting recommendations for environment and health conditioned tasks.

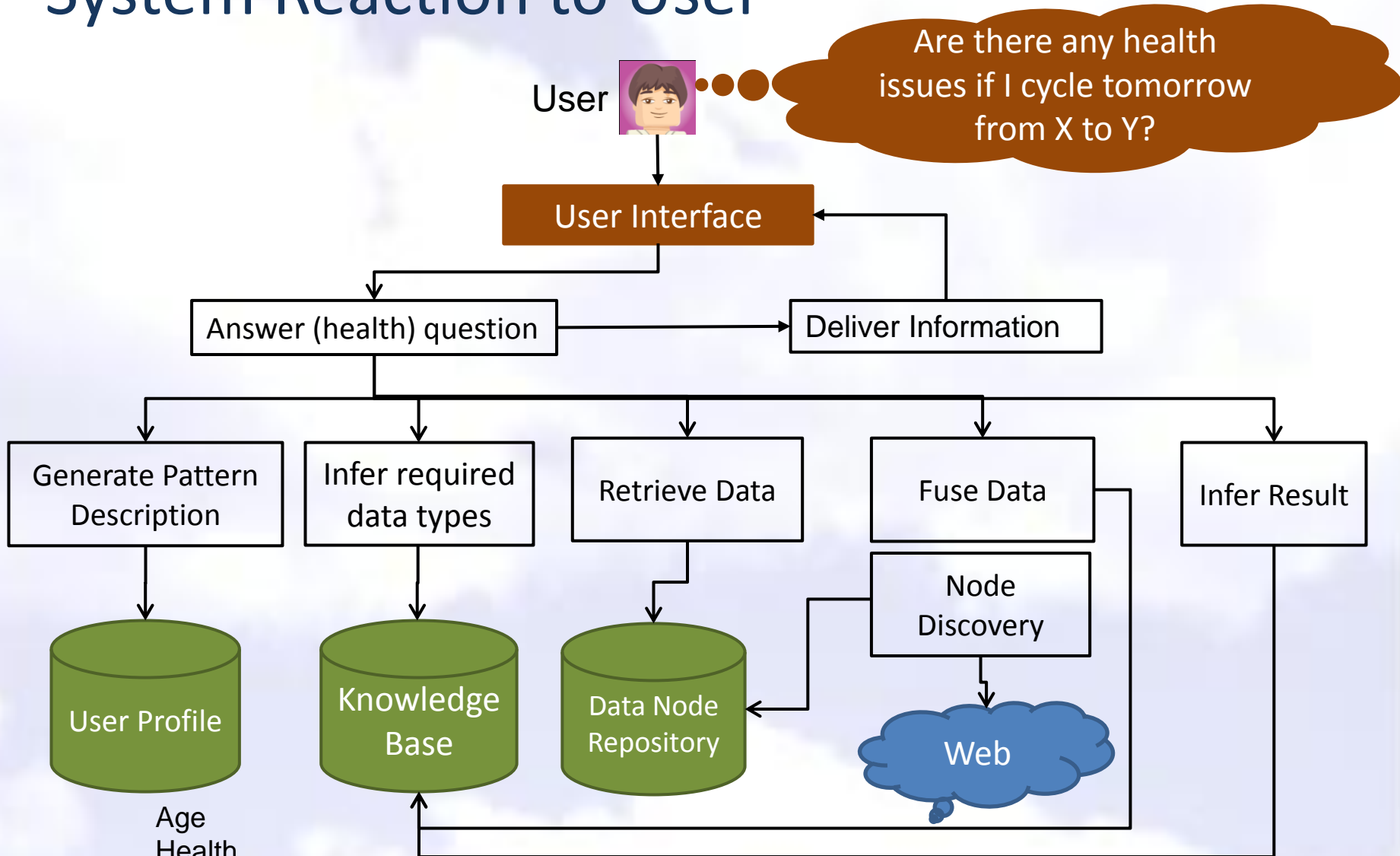


Personalized Environmental Service Configuration and Delivery Orchestration (PESCaDO) project addresses this challenge!

A Use Case ...

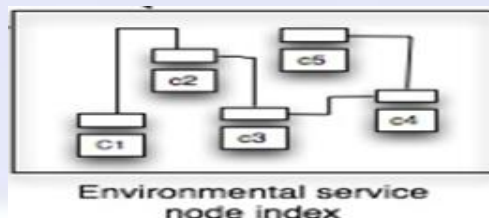


System Reaction to User

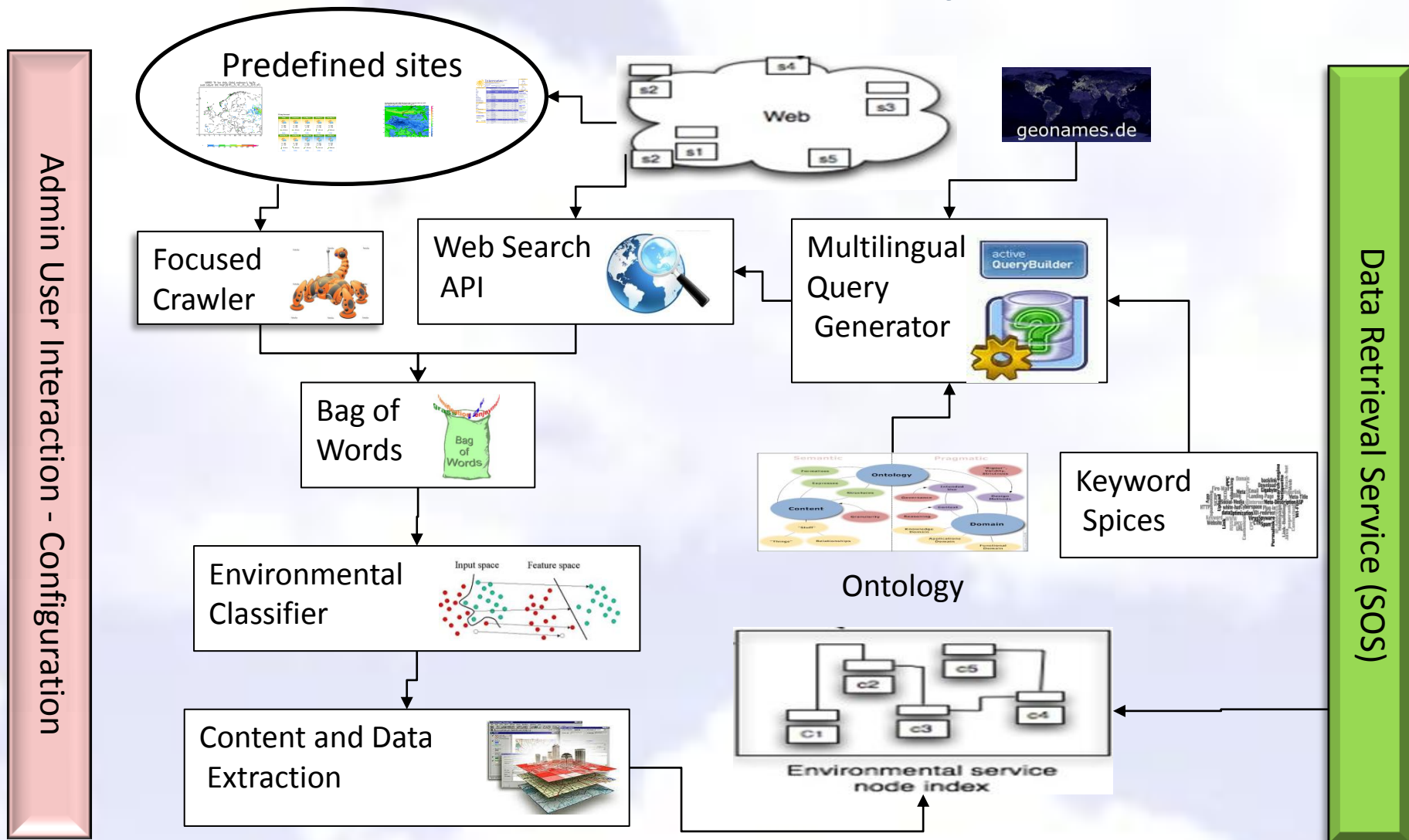


Environmental Nodes and Data

- A prerequisite for offering PESCaDO services are:
 - Discovery of Environmental Nodes
 - A problem of domain specific search
 - Content Extraction
 - Apply web text extraction and image analysis techniques



Environmental Node Discovery



Techniques for Domain Specific Search

- Discovery of Environmental Nodes

- Use a general purpose search engine

- E.g. Yahoo, Google, etc.
- Multilingual query generation using
 - Ontology
 - Keyword spice
 - Resources of geographical information (e.g. GeoNames, Foreca, etc.)
 - E.g. Helsinki + ozone + “air quality”
- Application of post processing techniques
 - Classification
 - Filtering



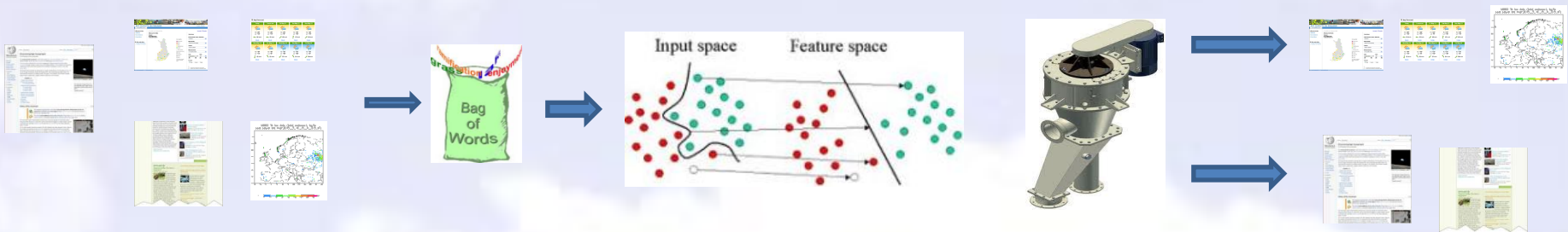
- Employ a focused crawler

- Crawls the web in a directed way based on machine learning.
- Based on a predefined set of websites (seed).
- The crawler employs classification to collect only those Web pages belonging to a certain topic.



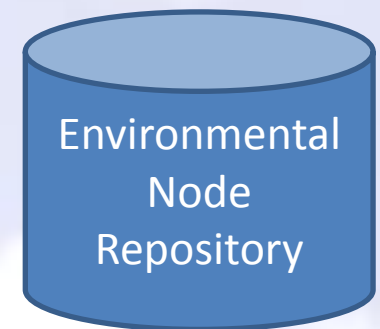
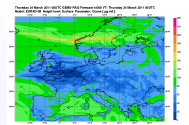
Techniques for Domain Specific Search

- Post processing based on Classification
 - Support Vector Machines are used for classification.
 - A model is trained using relevant and non relevant websites.
 - Classification can be employed also at lower level (e.g. weather classifier)
- Bag of words technique
 - Textual features generation using key-phrase extraction tools (KX)
 - Tokenization and N-grams extraction
 - Filtering of N-grams to select multiword expressions using morphological analysis
 - Ranking key-concepts based on normalized frequency and relevance



Empirical study of Environmental Nodes

- Environmental Website Statistics
 - Weather information: 80% of all information are reported in textual format or tables.
 - Air quality: 70% of the information about air quality is conveyed through images.
 - Pollen: 80% of the information about pollen is conveyed by images and maps.
 - No textual information at all for pollen and air quality forecast!
- Node Types
 - **Weather forecast:** temperature, pressure, wind, relative humidity
 - **Air Quality:** Pollutants: Sulphur Oxides (e.g. SO₂), Nitrogen Oxides (NO₂), Carbon Monoxide (CO), Dioxide (CO₂), etc.
 - **Pollen:** Grass pollen, birch pollen, alder pollen, etc.
 - **Alerts** (extreme weather conditions or phenomena): Extreme temperature, avalanches, coastal events, thunderstorms and snow/ice.
- The important information to be extracted
 - Type of measurement (e.g. temperature)
 - Type of data (historical, observed, statistical, forecast)
 - Measurement and Unit (e.g. 19°C)
 - Geographic Location of measurement (e.g. Helsinki)
 - Date/time (e.g. 14:00, 3/9/2011)



Techniques for Textual Data Extraction

- Processing webpage content
 - Process content and metadata
 - Parsing the HTML structure of the retrieved pages
 - Extraction of the relevant information from either the textual content or the HTML structure.
 - Filtering out irrelevant sections (menus, boilerplate, advertisement, etc.)
 - Regular expressions
 - Tree representation to capture the webpage structure
 - Combination

The screenshot shows a weather forecast page from Intermeteo.com. It features a search bar at the top, a navigation menu, and a main content area with a table of weather forecasts for various locations. The table includes columns for location, weather condition, temperature, and wind speed. The page also contains advertisements and a sidebar with additional weather information.

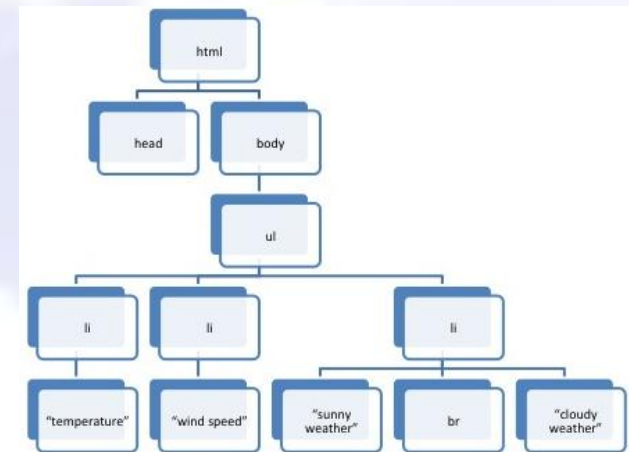
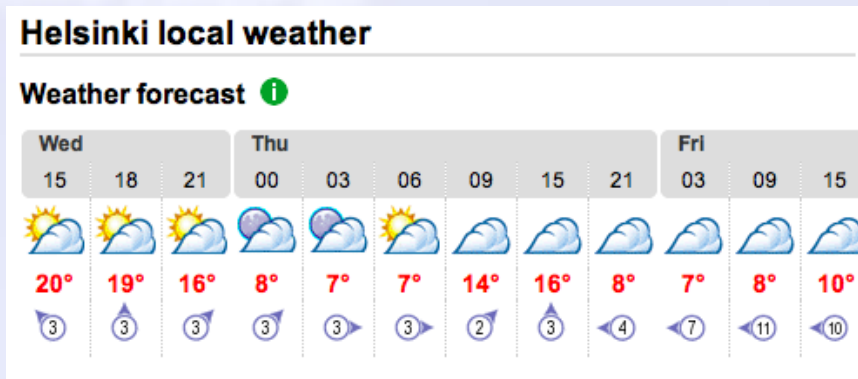
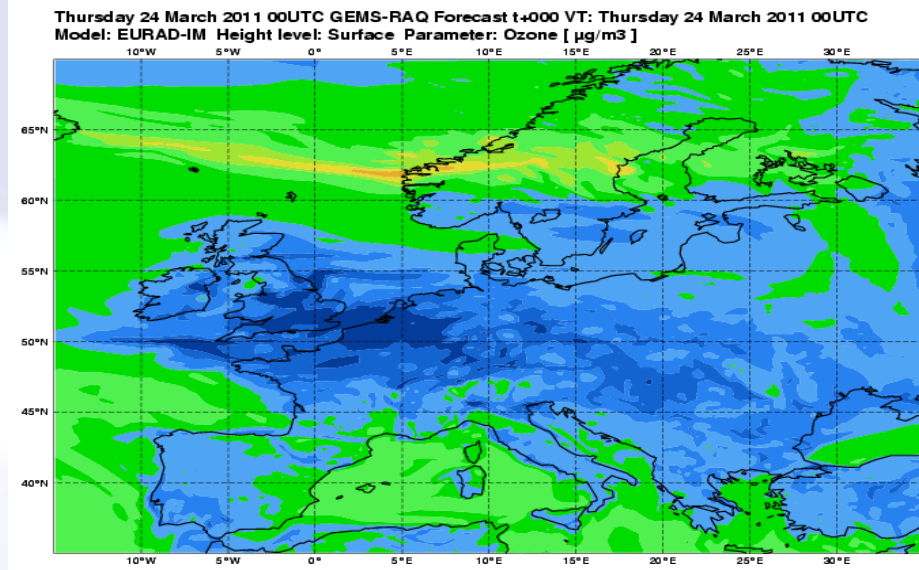
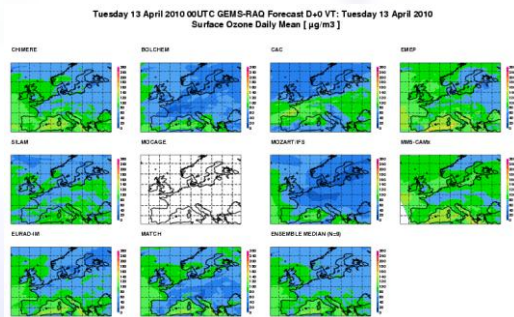


Image Types

- Heat Maps



- Variety of Diagrams

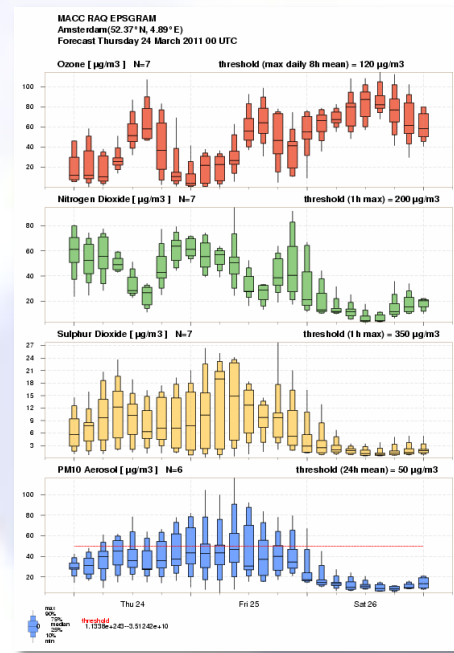
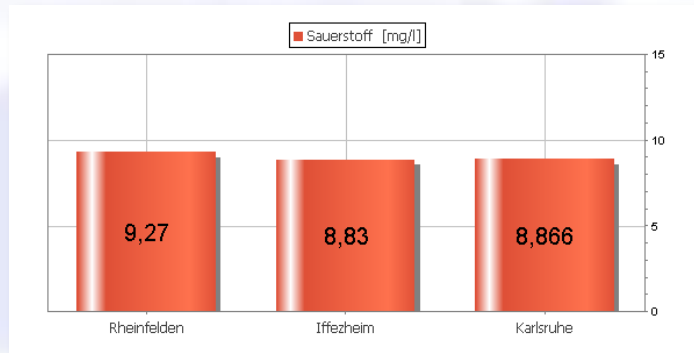
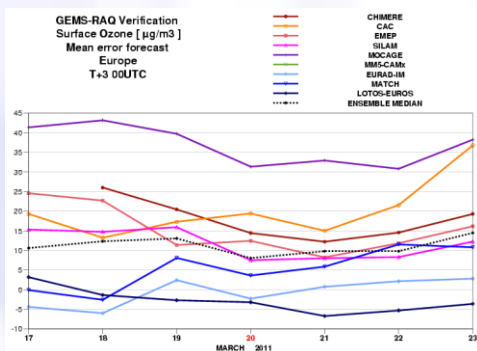
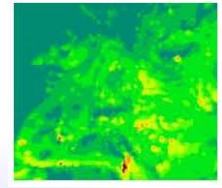
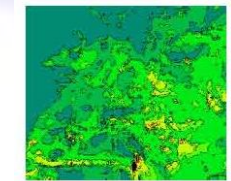
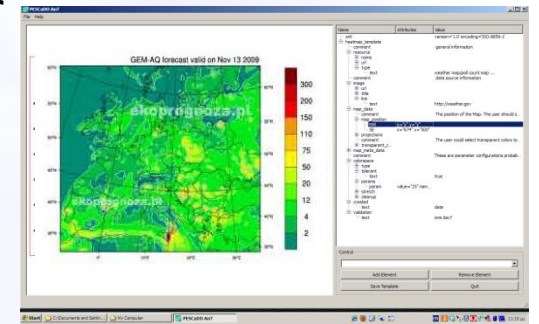


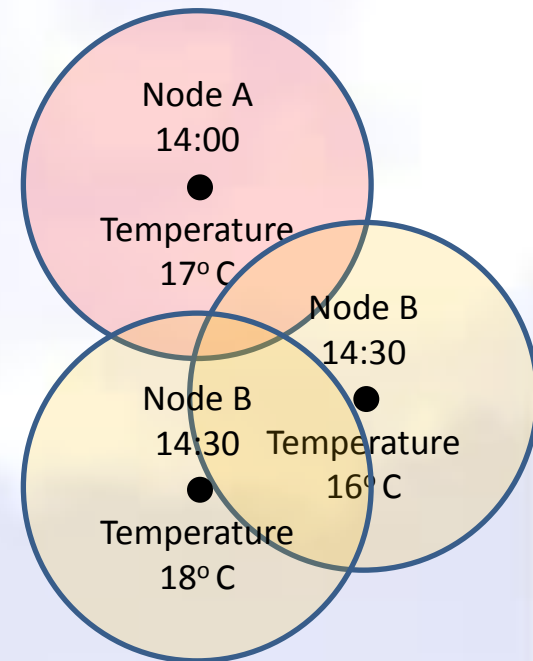
Image Content Distillation

- The large variety of images leads us to a user-assisted approach.
- We adopt a template-based representation (XML) for each image category (e.g. heatmap, line graph).
- We employ an annotation tool for template configuration
 - User assisted segmentation
 - User assisted geopinning
- OCR-based service to identify axis information, title, time and date.
- Mapping the index values to specific coordinates.
- Reconstruction of missing values and data gaps using interpolation.

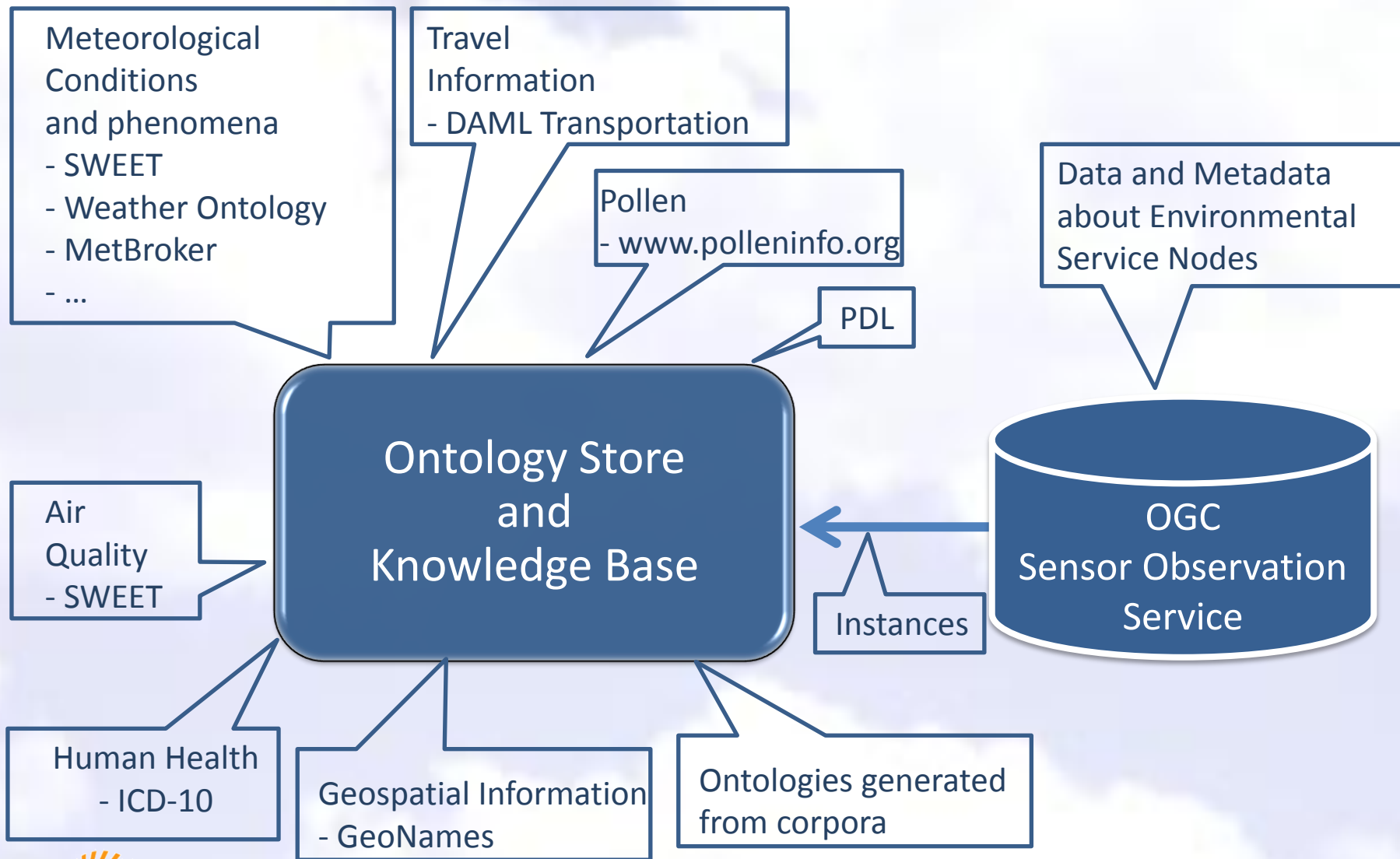


Environmental Data Fusion

- Fusion of environmental data is very complicated task
 - Environmental service nodes may provide on the same aspect for the same or the neighboring location:
 - Competing data
 - Complementary data
 - Different types of measurements (averages or exact)
 - For different time ranges
 - To ensure the availability of a most reliable and comprehensive content the content proceeding from these nodes must be fused.
 - This implies an assessment of the quality of the contributing services and data.
 - Quality metrics will be based on the evaluation of environmental nodes in time
 - Comparison of forecast data with observed data
 - Comparison of the fluctuation of a certain forecast over time



Knowledge Base



User Interaction – Select Activity

The screenshot shows the PESCaDO web application interface. The browser address bar displays "PESCaDO" and "cadoui/Wo". The page features a navigation menu with icons for user profile, search, and accessibility, along with a "Start Query" button and a "Streetmap" dropdown. A map of Helsinki is displayed, showing various districts and roads. A dialog box titled "Select your activity type" is open, listing four options: "Attending an Open Air Event", "Doing Some Physical Outdoor Activity", "Long Term Staying", and "Travelling". The "Doing Some Physical Outdoor Activity" option is highlighted. A callout box points to this option with the text "Activity = PhysicalOutdoorActivity". Another callout box points to the user profile icon with the text "Your interest". A third callout box points to the "Start Query" button with the text "Login for custom user profile or modify default user profile without account".

Your interest

Your activity

Login for custom user profile or modify default user profile without account

Activity = PhysicalOutdoorActivity

Select your activity type

- Attending an Open Air Event
- Doing Some Physical Outdoor Activity
- Long Term Staying
- Travelling

OK Cancel

Manual
Please go through the buttons in the top row to set

- (balloon) the type of request: Warning - Any Health Issues
- (golf club) the type of activity: Doing some physical outdoor activity
- (green flag) the are of the activity: draw a polygon near helsinki
- (timtable) the time of interesst: ?

press Start Query and wait for the result to appear near the selected region

POWERED BY Google

Map data ©2011 Geocentre Consulting, Tele Atlas - Terms of Use

SEVENTH FRAMEWORK PROGRAMME

User Interaction – Build Question

Automatically filled from

- User Profile
- Geolocation API of browser
- IP Address of request

User input via

- Map or
- Text (automatic extension)

Provide route or area of travel

Area	Route
	From Address: Helsinki
	To Address: Nuukio
Select Waypoint:	done
OK	Cancel

Query Overview:

- User: Fiona Fit
- Request: AnyHealthIssue
- Activity: PhysicalOutdoorActivity
- Start Date: Mon Jun 20 17:56:54 GMT+200 2011
- End Date: Mon Jun 20 17:56:54 GMT+200 2011
- Selected Route: From ? to ?
- Selected Area: ?
- Health Issues: sensitive to Alder pollen

User Interaction – Result Presentation

The screenshot displays the PESCaDO Workbench interface. The main window shows a weather overview for Helsinki from 18.03.2011 to 20.03.2011. The interface includes a search bar, navigation buttons, and a detailed weather table. A callout box highlights the detailed weather information, and another callout points to a zoomed-in map view showing pollen concentrations and weather symbols.

Weather Overview 18.03.2011 to 20.03.2011 (Daily Average)

	Friday	Saturday	Sunday
Temp.	5°C 1°C	11°C 1°C	12°C 3°C
Cloudiness (0-8)	~5	~1	~1
WindSpeed	18 km/h	16 km/h	19 km/h

Overview and Recommendation:
The weather will be warm and the AQ index will be satisfactory. The wind will be moderate and there will be a lot of alder pollen in the air, so the probability of experiencing allergy symptoms is quite high. We recommend that you take the necessary measures or postpone your excursion until the risk has gone down.

Zoomed out view shows why the map is coloured red

The map displays the area with additional information e.g. weather symbols and pollen concentrations

Conclusions

- People are increasingly interested in environmental information and therefore environmental applications are of great interest to the public.
- Some of the important challenges are:
 - The discovery of environmental nodes
 - The extraction of multimodal environmental information
 - The fusion of environmental information
- The variety of representations and lack of standards in the presentation of environmental information make these tasks very complicated.
- Several techniques from text mining , image analysis and information retrieval areas could be applied in the environmental domain
 - These techniques need to be optimized to deal with environmental data.
 - They could be of added value for the development of environmental related services



- **Interdisciplinary** research in the **Environmental** domain and **Information Retrieval** area could be beneficial for humanity in the context of offering services and solutions for **healthcare** issues related to **environmental** conditions.

PESCaDO Partners



Universitat Pompeu Fabra (ES), (Coordinator)

UPF



Fraunhofer IOSB (D)

IOSB



Finnish Meteorological Institute (FIN)

FMI



Institute for Visualization and Interactive Systems, University of Stuttgart (D)

USTUTT



Fondazione Bruno Kessler (IT)

FBK



Centre for Research and Technology Hellas, Informatics and Telematics Institute (GR)

CERTH



Helsinki Region Environmental Services Authority (FIN)

HSY

External Cooperations



Informatics Systems and Application Group, AUTH (GR) ISAG

Contacts

Leo Wanner (Coordinator), UPF, leo.wanner@upf.edu

Stefanos Vrochidis, ITI-CERTH, stefanos@iti.gr

Ioannis Kompatsiaris, ITI-CERTH, ikom@iti.gr

THANK YOU FOR YOUR ATTENTION

Questions?

<http://www.pescado-project.eu>

