Discovery, Analysis, and Retrieval of Multimodal Environmental Information

Anastasia Moumtzidou
Information Technologies Institute, Greece

Stefanos Vrochidis
Information Technologies Institute, Greece

Ioannis Kompatsiaris
Information Technologies Institute, Centre for Research and Technology Hellas, Greece

INTRODUCTION

Environmental conditions are considered of utmost importance for human life. Citizens are increasingly aware of the important role that environmental data (i.e. weather forecast, air quality and pollen concentration) play on health issues (e.g. allergies), as well as to a variety of outdoor activities (e.g. agriculture, trip planning). Given the fact that ensembling information from several environmental providers can generate more reliable measurements, there is a need to combine environmental data from multiple resources, in order to facilitate retrieval of environmental information and support personalized services (Wanner et al., 2012).

In this context, this article analyzes the aforementioned needs and challenges (Figure 1) by discussing the application of techniques from the information technologies domain on environmental data. First, we address the discovery of environmental web resources (referred to as environmental nodes) as a domain-specific search problem. Then, we provide insights into the presentation formats of the environmental resources, as well as information extraction techniques that could be applied. Finally, we discuss indexing and retrieval of environmental information.

The article is structured as follows. First we present the background and basic definitions regarding the environmental information. Then, an empirical study on the presentation of environmental data is realized. In the following sections, the approaches for environmental data discovery, content extraction, as well as indexing and retrieval are reported. Finally, we present future trends and conclusions.

BACKGROUND

In general, the environmental domain covers a broad variety of data. However, in this article we focus on the environmental information that covers the following three aspects, which strongly affect both humans and vegetation: air pollutants, meteorological conditions, and pollen.

As an example of responses to air pollutant concerns, the European Parliament and Council (European Parliament, Council, 2008) wrote a number of directives that would ensure ambient air quality for Europe by setting limits to the concentration values of several air pollutants, such as sulphur dioxide, nitrogen dioxide and oxides of nitrogen (Karatzas and Moussiopoulos, 2000). Some of the pollutants that affect air quality are: Sulphur dioxide (SO_2), Nitrogen oxides ($NO+NO_2$), Thoracic particles (PM_{10}), Fine particles ($PM_{2.5}$) and Ozone (O3).

Weather (or meteorological) conditions comprise one of the most well-known environmental aspects. They are used for describing the state of the atmosphere in terms of temperature, wind (i.e. wind direction and speed), pressure and moisture (i.e. clouds and precipitation). In case these properties achieve extreme values, warnings are being produced to protect humans, flora and fauna. Examples of such extreme weather phenomena are tornadoes, hurricanes, typhoons and fires.

Finally, pollen consists of particles released from trees, weeds, and grasses that spread through the air to fertilize plants. Pollen allergies can trigger allergic rhinitis, worsen asthma and lead to other problems such as sinus and ear infections. Since the number of

Figure 1. Challenges in the environmental domain
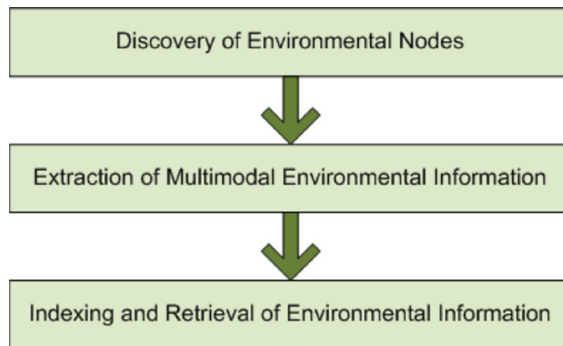
known pollen types is large, we will cite only some indicative ones: grass, birch, alder, ragweed, olive tree and cypress. Pollen concentration is a measure of how much pollen is in the air in a certain area at a specific time and it is expressed in grains of pollen per square meter of air collected over 24 hours.

The main sources of environmental information for the everyday user are web portals and sites, maintained usually by national meteorological and air quality agencies populated with data provided by distributed stations after analysis based on relevant models.

EMPIRICAL ANALYSIS OF ENVIRONMENTAL RESOURCES

In order to discover and extract the environmental information available on the web, it is necessary to study the structure and the encoding format of the environmental resources.

To perform the empirical study, we have selected a dataset of 45 of the most common environmental websites based on the suggestions of environmental experts (i.e. scientists working in the Finnish Meteorological Institute[1] and in the Helsinki Region Environmental Services Authority[2], who were involved in PESCaDO project[3] (FP7-248594)). More details on this empirical study are available in (Moumtzidou et al., 2010). Table 1 contains indicative sites from the original dataset. Table 2 contains the numerical statistics regarding the aspect measured, while Table 3 reports the area covered.

The study was conducted into two iterations based on the well-established technique of Empirical Cycle of A.D. de Groot (1961). Each iteration includes the following steps: observation, induction and testing. In the first iteration, conclusions about the type of information that is of importance were drawn, while the second one focused on identifying where the important information is located and how it is encoded. Figure 2 illustrates the overview of the conducted empirical study. The outcome of the study was that the information that needs to be detected includes: the measurement (type and data), the location, the time and date and the language. Finally, the statistics of the sites regarding the encoding of information are shown Table 4.

As it is shown in Table 4, a considerable amount of information is encoded in images. The most common image types that are found in the environmental nodes are:

- **Maps:** Maps (including heatmaps) of specific areas, countries and Europe in different scales and coordinate systems (Figure 3).

Table 1. Indicative sites from the original dataset with environmental information

URL	Content
http://www.ilmanlaatu.fi/ilmanyt/nyt/ilmanyt.php	Air Quality Data
http://www.polleninfo.org/	Pollen Data
http://pollen.fmi.fi/pics/EuropeHS.html	Pollen Data
http://www.fmi.fi/ilmanlaatu/	Air quality & Weather Data
http://www.ymparisto.fi/default.asp?node=4882&lan=fi	Air quality Data
http://www.foreca.com	Weather Data
http://www.intermeteo.com/europe/finland/helsinki/	Weather Data
http://www.intellicast.com/	Weather Data
http://gems.ecmwf.int/d/products/raq/	Air Quality Data

Related Content

Steel Surface Defect Detection Based on SSAM-YOLO

Tianle Yangand Jinghui Li (2023). *International Journal of Information Technologies and Systems Approach (pp. 1-13).*

www.irma-international.org/article/steel-surface-defect-detection-based-on-ssam-yolo/328091

Trust in Computer Mediated Communication

Ardis Hansonand Sheila Gobes-Ryan (2015). *Encyclopedia of Information Science and Technology, Third Edition (pp. 2122-2130).*

www.irma-international.org/chapter/trust-in-computer-mediated-communication/112620

Chaotic Map for Securing Digital Content: A Progressive Visual Cryptography Approach

Dhiraj Pandeyand U. S. Rawat (2016). *International Journal of Rough Sets and Data Analysis (pp. 20-35).*

www.irma-international.org/article/chaotic-map-for-securing-digital-content/144704

Digital Archives for Preserving and Communicating Architectural Drawings

Roberta Spalloneand Francesca Paluan (2018). *Encyclopedia of Information Science and Technology, Fourth Edition (pp. 5213-5225).*

www.irma-international.org/chapter/digital-archives-for-preserving-and-communicating-architectural-drawings/184226

Usability and User Experience: What Should We Care About?

Cristian Rusu, Virginica Rusu, Silvana Roncaglioloand Carina González (2015). *International Journal of Information Technologies and Systems Approach (pp. 1-12).*

www.irma-international.org/article/usability-and-user-experience/128824