Videk: A Mash-up for Environmental Intelligence

Klemen Kenda[1], Carolina Fortuna[2], Blaz Fortuna[1], Marko Grobelnik[1]

[1] Artificial Intelligence Laboratory,
[2] Communication Systems Department
Jozef Stefan Institute, Jamova 39, 1000 Ljubljana, Slovenia
{klemen.kenda, carolina.fortuna, blaz.fortuna, marko.grobelnik}@ijs.si

Abstract. In this paper we introduce Videk – a mash-up for environmental intelligence. Videk currently uses four sources of sensor and linked data and relies on StreamSense engine for storage and processing. In this paper we present the architecture and current implementation of Videk as well as the lines along which we plan to extend and improve it.

Keywords: Mash-up, sensors, web of things, real-time, data mining, semantic web.

1 Introduction

Cities of the future will be equipped with sensing and actuating devices embedded in physical objects referred to as Smart Objects or Things. The presence of Smart Objects is increasing, but they currently form isolated networks controlled by different entities and most often the data remains closed and not used to its full potential. Connecting (or federating) the islands of things using web standards is also referred to as Web of Things. This development will provide new services and will enable new directions for communication and new types of agents involved in this process (i.e. "things-to-persons" or "thing-to-thing") [1][2].

In this paper we present Videk (see Fig. 1), a mash-up based on several sources of data, including data coming from Smart Objects. On the server side Videk uses StreamSense, a sensor stream processing system based on tightly integrated and scalable custom software modules. StreamSense provides interfaces and means of information collection from a set of Smart Objects and generic APIs for data feeds on one hand; and interfaces to application developers on the other. Videk provides functionality such as, finding illuminance measurements around a given location or, showing all the locations in some region that measure illuminance.

In the next section we present the architecture and implementation of the system while in Section three we discuss the road ahead, particularly functionality that will be added to Videk as we continue developing StreamSense.

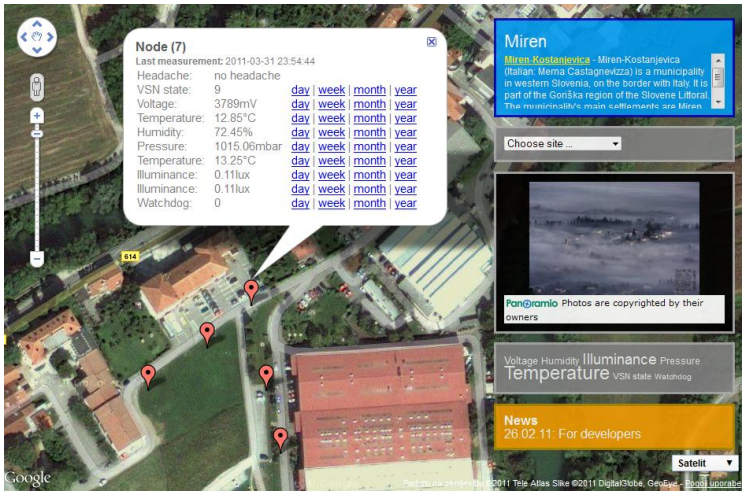

Fig. 1. Screenshot of Videk[1]

2 System architecture and implementation

The architecture of this system is depicted in Fig. 2 and consists of the following components: sources of data, processing server and front end. These three components are connected in a pipeline resulting in the provision of environmental intelligence in real time through the Videk mash-up and open API.

2.1 The sources of data

Videk currently uses four sources of data: sensor measurements, Geonames, Wikipedia and Panoramio.

The sensor measurements are coming for the Smart Objects from the three deployments in Slovenia: two deployments of sensors on public light poles (16 sensor nodes, ~80 streams of measurements) and one deployment inside stables on a farm (4 sensor nodes, ~16 streams of measurements)[2]. The public light poles deployment will gradually increase in size over the next year while deployments of with different settings are going to be added. Currently deployments for sports fishing and bee monitoring are underway.

For each sensor node we store the several meta-data fields: GPS coordinates, types of on board sensors, ZigBee address, etc. These meta-data is stored in an internal triple-store where it is linked, when possible, to several Linked Open Data[3][3] and other data providers. At the moment we fully integrate with

[1] http://sensors.ijs.si
[2] http://sensorlab.ijs.si/en/demos.html
[3] http://linkeddata.org/

Geonames, Wikipedia and Panoramio,. Linking to LOD cloud is also implemented[4] but the integration into the mesh-up's user interface is still ongoing.

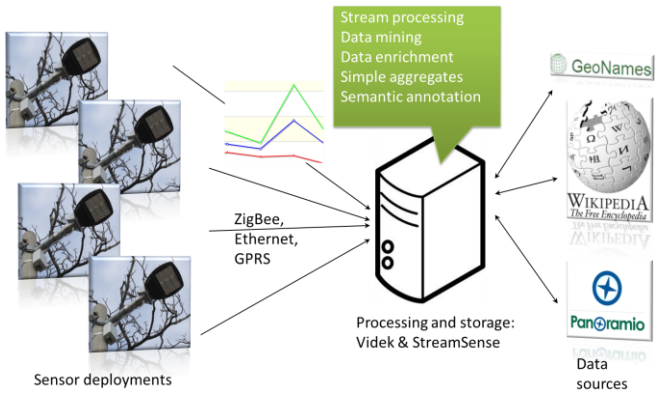

Fig. 2. Architecture of Videk's back-end.

Geonames[5] is a web service which provides interface to geographical data, such as name of the place, elevation, topographic hierarchy, Wikipedia entries etc. We use the reverse geocoding and the API for retrieving geocoded data entries from Wikipedia.

Panoramio[6] is a geolocation-oriented photo sharing community. API provides access to the published photographs, which can be selected by geolocation, topic etc. In Videk's current development status, the geolocation based retrieval is used.

2.2 The storage and processing engine

The underlying storage and processing engine is based on in house software libraries and we'll refer to it in this paper as StreamSense. The implementation has four stores, two for meta-data (sensor node and sensor type), one for measurements (sensor measurements) and one which essentially connects them (sensor). The indexing of the data is application-oriented, specific and therefore extremely efficient. As such it can perform real-time indexing of streams. Several clustering algorithms are available; Videk uses for now K-Means clustering of sensor nodes based on their location. Furthermore, aggregates and detected events (headache occurrence based on atmospheric pressure) are computed and

[4] http://sensors.ijs.si:2020/
[5] http://www.geonames.org/
[6] http://www.panoramio.com/

available via public API. Work on interfacing StreamSense with ETALIS[7] for complex event detection is in progress.

StreamSense also features search and ranking strategies. The existing search on structured data [4] demo[8] will be ported on the sensors dataset and integrated into Videk.

2.3 The presentation

Videk's graphical user interface (GUI) is based on Google Maps API[9]. For data manipulation, event handling and Ajax interactions jQuery[10] library is used. GUI receives data through API layer, which is based on a PHP/MySQL custom made content management system (CMS).

The API layer provides an additional aggregation and safety layer in the system. It can validate input parameters and safely access StreamSense. It can enrich sensor data with additional sources (such as Geonames or Wikipedia). The API layer by itself is a REST based web service for enriched sensor data. It accepts HTTP requests and returns data in XML format[11]. The API layer also uses the XML interface to communicate with StreamSense via its embedded HTTP server and Geonames API (XML), which is cached internally to minimize number of requests.

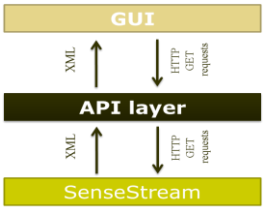

Fig. 3. Role of API layer.

Videk initializes with the data of the current state of the sensor nodes in the network: geolocation, number and type of real and virtual sensors on the node, last measurements. For display and navigation purposes nodes are clustered by their location.

There are 5 widgets in Videk. First contains node cluster location (nearest place from Geonames) and nearest Wikipedia entry with short description and link to the actual web page. Next is navigation widget, where the user can select the cluster to navigate to. Third is Panoramio widget which reinitializes with the photos, contained in the area of the map. Fourth widget displays tag cloud of the features of measurements, weighted by their importance (i. e. number of the sensors). Last widget shows news regarding system maintenance and API.

---

[7] http://code.google.com/p/etalis/
[8] http://sensors.ijs.si/static/index.html
[9] http://code.google.com/intl/sl/apis/maps/index.html
[10] http://jquery.com
[11] http://sensors.ijs.si/sl/api/index.html

Location and Panoramio widgets are interactive. Their contents changes according to the center and area of the shown map.

Clicking the cluster will zoom in the current cluster. Clicking the node will open an info window, containing node data, including sensors, their last measurements and option to click and retrieve last day, week, month or year charts for any measurement.

3 Related Work and Road ahead[12]

Other mash-ups involving data from real objects are Google maps[13], Pachube[14], EPCMashup[15] etc. The Google maps mash-up combines several sources of data including real time traffic monitoring and public data. It is able to provide a view of the city (i.e. New York) including work on public roads, traffic congestion, images of locations, etc. Pachube provides a platform for crowdsourced sensor streams and offers simple services on top of these. EPCMash-up is a research prototype based on simulated data in the context of supply chain management. Videk is different in the fact that is uses live data from real testbeds and LOD. We have complete control over the entire vertical and there is on going work to add structured search and complex event processing functionality.

We intend to extend and improve Videk and StreamSense to provide services for environmental intelligence. In this line, work on detecting complex events is ongoing. Next would be to focus on automatic monitoring, including event detection and intelligent control for a green planet.

References

1. Internet of Things — An action plan for Europe. *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions* (Brussels, June 18, 2009).
2. It's a smart world. *The Economist*. The Economist (London, Nov 4 2010).
3. A. Moraru, M. Vucnik, M. Porcius, C. Fortuna, M. Mohorcic, D. Mladenic, Exposing Real World Information for the Web of Things, IIWeb (WWW2011), Hyderabad, India.
4. L. Dali, A. Moraru, D. Mladenic , Using Personalized PageRank for Keyword Based Sensor Retrieval, SemSearch (WWW2011), Hyderabad, India.

[12] The authors would like to thank all our colleagues, particularly SensorLab members for their support and valuable contributions. Our acknowledgements also go to Miren-Kostanjevica Municipality and Envigence Ltd. for hosting the testbed. This work was partially supported by the Slovenian Research Agency and the ICT Programme of the EC under PASCAL2 (ICT-NoE-216886), ENVISION (ICT-2009-249120) and PlanetData (ICT-NoE-257641).

[13] http://maps.google.com/?layer=t&z=10&ll=40.714997,-74.006653

[14] http://www.pachube.com/

[15] http://epcmashup.appspot.com/