UCLA
Posters

Title

Republishing Sensor Data: users publishing transformations of existing sensor data

Permalink

https://escholarship.org/uc/item/1br5d5x1

Authors

Park, Unkyu
Heidemann, John

Publication Date

2007-10-10

Peer reviewed

Republishing Sensor Data: users publishing transformations of existing sensor data

Unkyu Park and John Heidemann

working with Brian Fulkerson, Sasank Reddy, Nathan Yau, Sung Jin Kim, Mark Hansen, Junghoo Cho

ISI Laboratory for Embedded Networked Sensor Experimentation - http://www.isi.edu/ilense/

Introduction: Sensor-Internet Sharing and Search, Transformation of Sensor Data
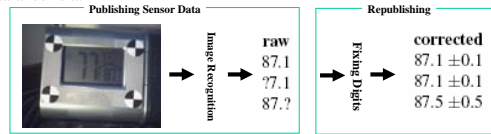
Sensor-Internet Sharing and Search

- **Sensor-Internet** [S. Reddy, G. Chen, B. Fulkerson, S. J. Kim, U. Park, N. Yau, J. Cho, M. Hansen, and J. Heideman. Sensor-Internet Share and Search: Enabling Collaboration of Citizen Scientists. in Data Sharing and Interoperability on the World-wide Sensor Web, IPSN 2007, April 2007]
    - Goal: Share and Search across many independently run sensor networks
    - Basic Components:
        - **sensor:** sense the environments
        - **sensor publisher:** sends collected data to sensor store
        - **sensor store:** a data repository
        - **sensor search engine:** searches sensors or data
        - **republisher:** transforms the existing data

Basic deployment: sensors pass data through publisher to store, republisher aggregate data, search engine indexes data, users access data

- **SensorBase.org** [http://sensorbase.org] ongoing work by Reddy, Yau, Hansen
    - A centralized sensor data repository: our sensor store implementation
    - Store the sensor data and support queries and visualization
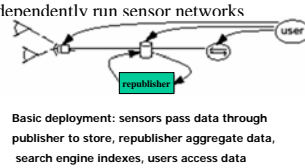
Republishing: Transformation of Sensor Data

- **Beyond just sharing the sensor data**
- **Usually require further processing on the data to user needs**
    - aggregation, filtering, statistical estimation, vetting and error suppression, etc.
- **Example: Temperature data collection**
    - Read of-the-shelf wireless temperature sensors with standard web cameras (image processing done by Brian Fulkerson)
    - Publishing: Web Camera, Wireless Temperature Sensors, USB LED Light
    - Republishing: Fix unrecognized digits of published temperatures based on temporal-spatial correlation

Publishing Sensor Data		Republishing	
Image Recognition	raw 87.1 ?7.1 87.?	Fixing Digits	corrected 87.1 ±0.1 87.1 ±0.1 87.5 ±0.5

Challenges: Software Infrastructure for Sensor Data Republishing

What kind of software infrastructure is required for sensor data republishing?

- **Source Discovery :** How to discover the source data? (search over the sensor data as well as sensor information) – research by Sung Jin Kim and Junghoo Cho
- **Data Management:** Support to take source data and publish back the transformed data. Unkyu Park and John Heidemann
- **Access Control & Security:** What kind of access control and security is needed for sharing sensor data? - research by Nathan Yau and Mark Hansen
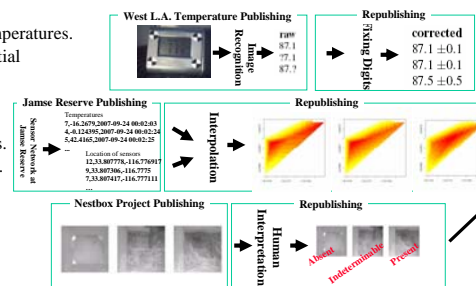
Data Management in Republishing

- Data Providence: How to trace republished data back to its source data.
- Robustness: Coping with network outages when publishing and republishing
- Synchronization: Matching republishers with variable or different publication rates of sources

Preliminary Results: Republishing Examples and New SensorPress Framework

Examples of Republishing

- **Digit Repair in Temperature Capture**
    - Sources: An image-captured temperature and preceding temperatures.
    - Process: Fix the unrecognized digits based on temporal-spatial correlation.
    - Output: Corrected temperature (if possible)
- **Temperature Contour Map at James Reserve**
    - Sources: Temperatures and geographical location of sensors.
    - Process: Interpolate point data into a complete contour map.
    - Output: A temperature contour map
- **Human Interpretation of Nestbox Images**
    - Source: An image of nestbox
    - Process: Human interpretation on the indeterminable image
    - Output: An annotation on the image

West L.A. Temperature Publishing — Image Recognition — raw 87.1 ?7.1 87.? / **Republishing** — Fixing Digits — corrected 87.1 ±0.1 87.1 ±0.1 87.5 ±0.5

Jamse Reserve Publishing — Sensor Network at James Reserve — Temperatures 7,-16.2679,2007-09-24 00:02:03 / -0.124395,2007-09-24 00:02:24 / 5.02.4165,2007-09-24 00:02:25 — Location of sensors 12.33.807778,-116.776917 / 9.33.807306,-116.7775 / 7.33.807417,-116.777111 — **Republishing** — Interpolation

Nestbox Project Publishing — **Republishing** — Human Interpretation — Absent Indeterminable Present

SensorPress:

A new Sensor Data Framework that provides common components required for republishing composition.

Republisher saves time and effort to build a republishing.

[Work in progress]

### Tracking Sources: Sensor-data Link

- **Problem:** How to trace back the source of transformed data?
- **Solution:** Use sensor-data links to locate the source data
    - Sensor-data Link Template:
        - sb://<location of wsdl file>?s=<service name>&a1=<arg1>&a2=<arg2>…@TIMESTAMP
    - e.g.) Sensor-data Link of Digit Repair Republishing:
        - sb://sensorbase.org/soap/sensorbase2.wsdl?s=getData&a1=datetime,sensorid,rawtemperature,temperature&a2=p_97_temperature&a3=sensorid="sum-in" and datetime > "2007-10-10 00:00:00"&a4=0&a5=2
    - Easy to extend the capability by defining new webservice.
- **Problem:** The sensor-data link may not reproduce same sources after the republishing.
- **Solution:** Embedded the timestamp when the republishing happens.

Incremental Compression Algorithm to Store Links Efficiently

- **Problem:** A simple solution of data providence is just storing source links. But it is storage-intensive to provide tuple-level data providence
- **Solution:** Use an incremental compression to store sensor data links efficiently, exploiting patterns in the previous data.
- Reduces storage cost of sensor-data links to 15% of uncompressed.

no compression
per-link gzip compression
cross-link incremental gzip compression

size of sensor-data (byte)
sensor-data link
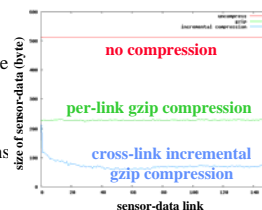
Tolerate the network outage

- **Problem:** Network occasionally goes down.
- **Solution:** Use Delay Tolerant Networking techniques to add robustness to network outage
- **Work in progress**

Optimizing Republishing Rate

- **Problem:** Ideally, republishers should synchronize their analysis with publishers of source data, but publisher rates may vary, or two source data streams may publish at different rates
- **Solution:** Ideally, publishers export publish rates; alternatively: republishers estimate and predict future publishing rates
- **Work in progress**

More Information

Project webpage: http://www.isi.edu/ilense/siss/