Mashup-based Linked Data Integration

Instructor: Vũ Tuyết Trinh

1

Agenda

1. Introduction
2. Linked Widgets Framework
3. Mashup Models
4. Conclusion and Future Work

2

1. Introduction

3

Motivation

CSV(lat, long, name, etc.)

980 parks in Vienna

XML(lat, long, Image URL)

Flickr/Google Image Search

JSON(lat, long, CO, NO_2, etc.)

Air quality data

CSV(lat, long, name, etc.)

51 public swimming pools in Vienna

?

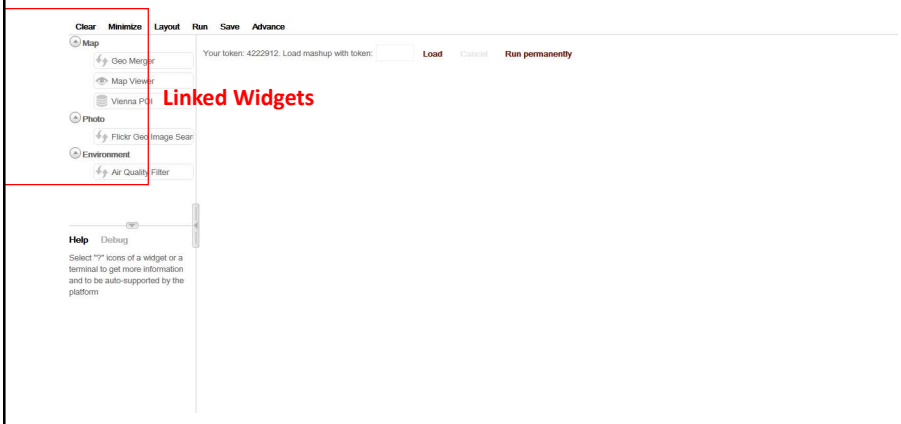
4

Motivation

5

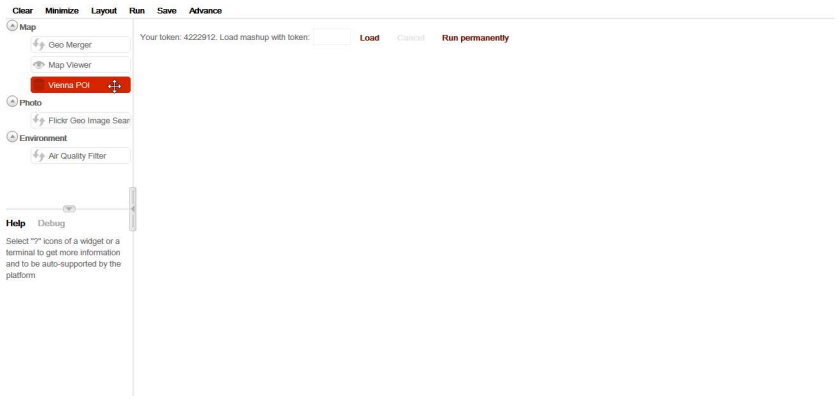
5

Modular approach for data integration



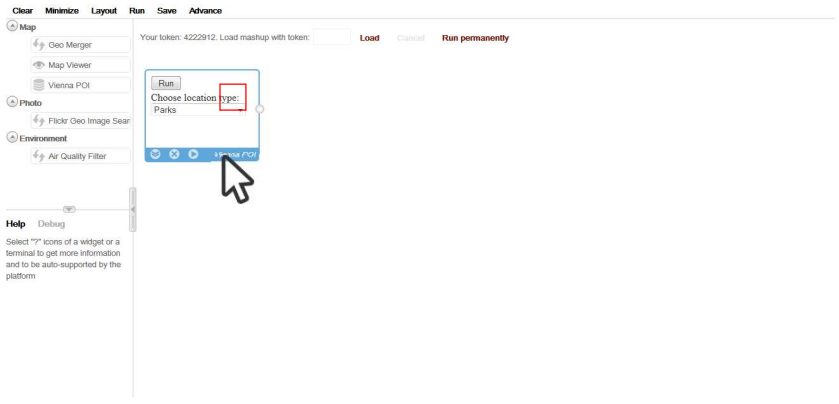
6

Modular approach for data integration



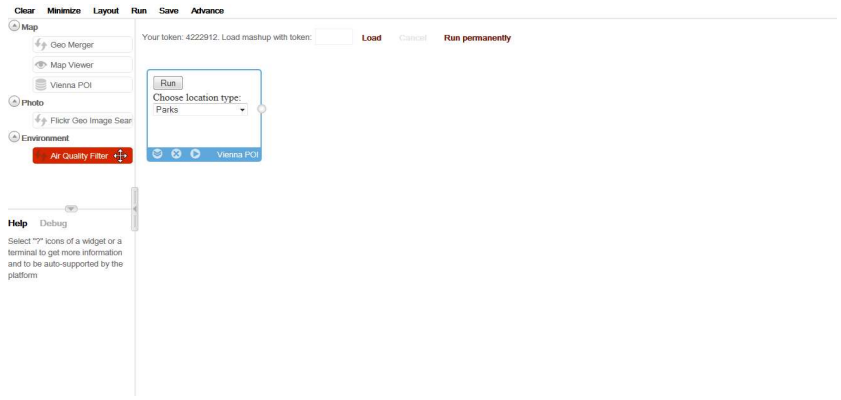
7

Modular approach for data integration



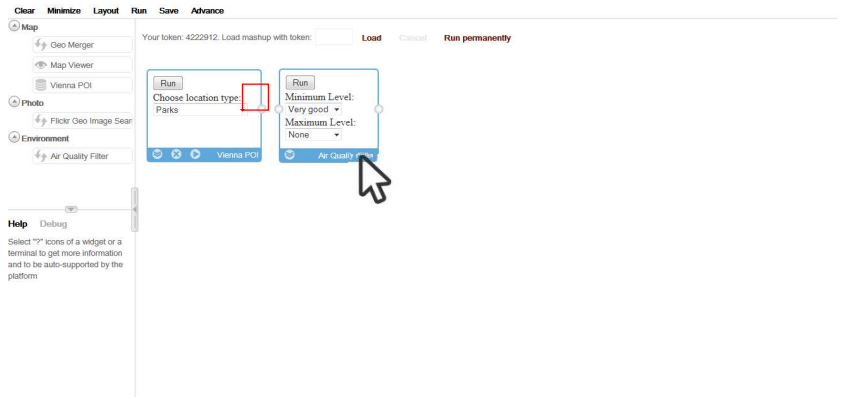
8

Modular approach for data integration



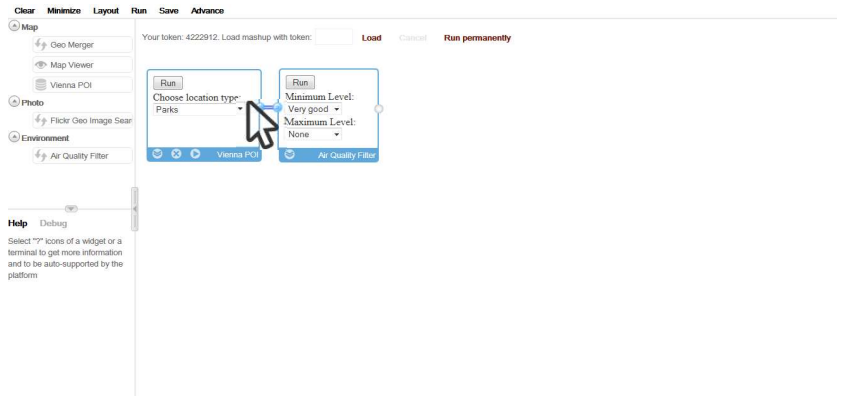
9

Modular approach for data integration



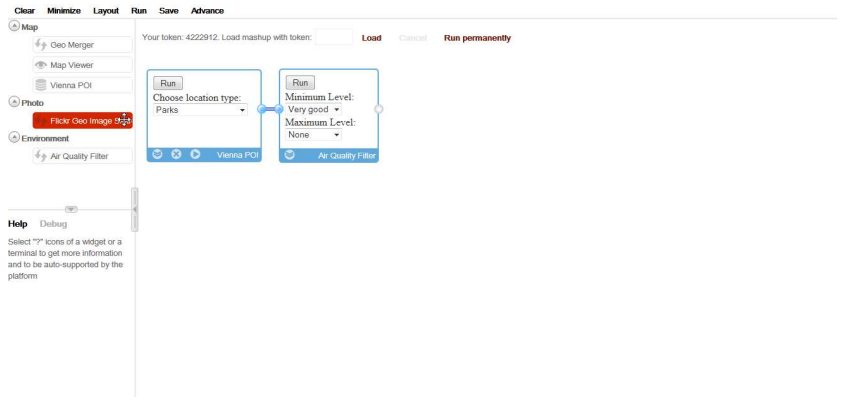
10

Modular approach for data integration



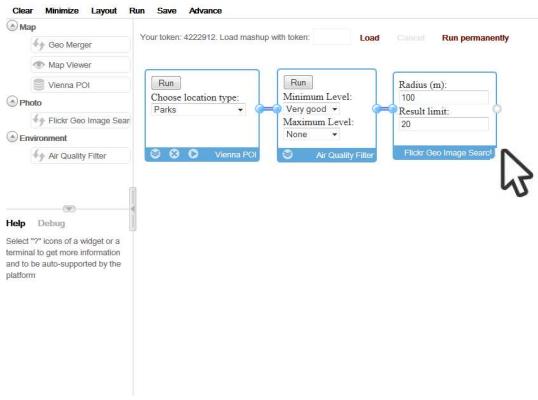
11

Modular approach for data integration



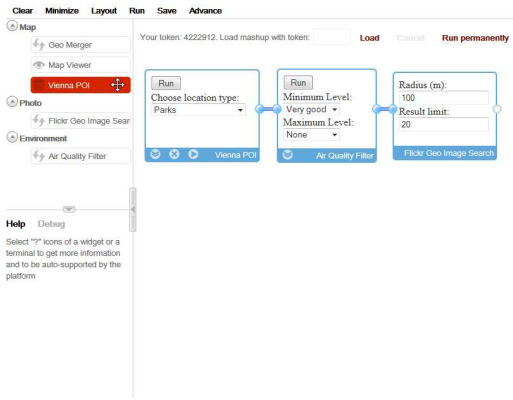
12

Modular approach for data integration



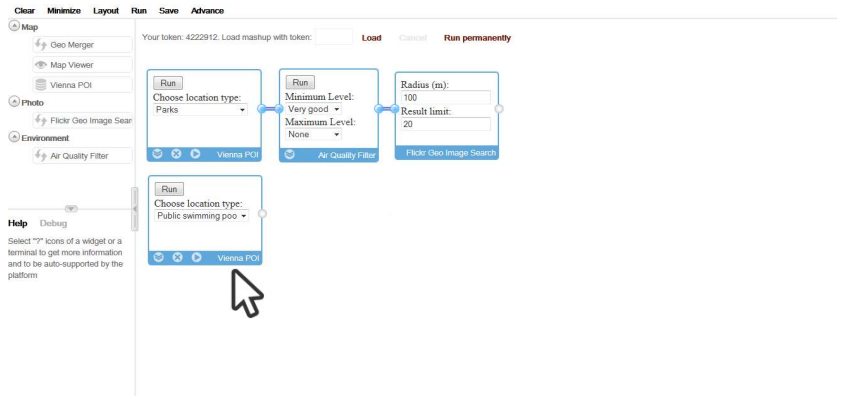
13

Modular approach for data integration



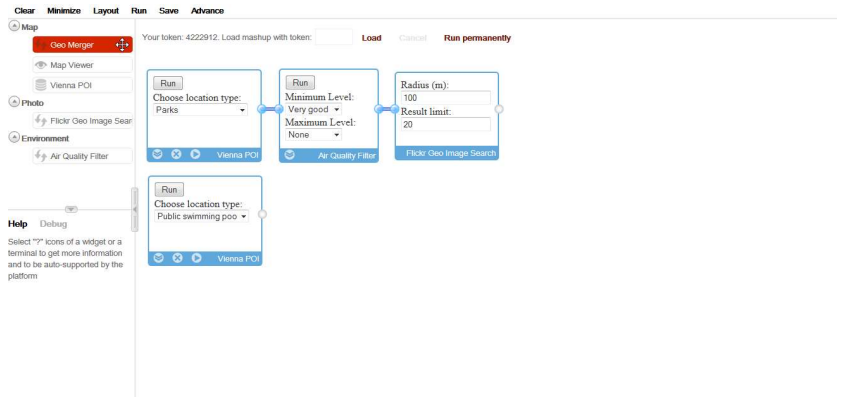
14

Modular approach for data integration



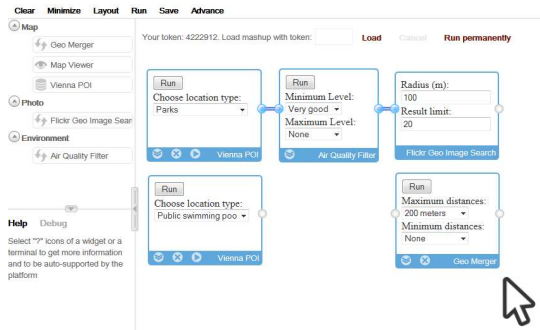
15

Modular approach for data integration



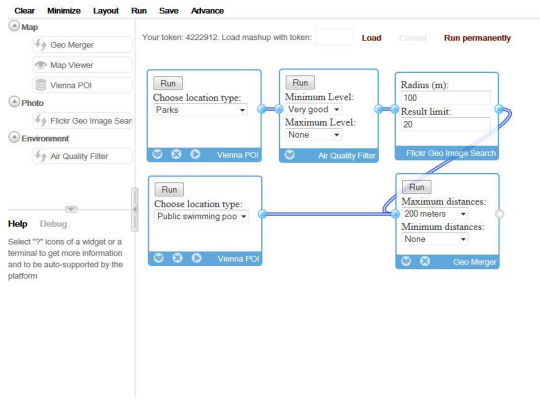
16

Modular approach for data integration



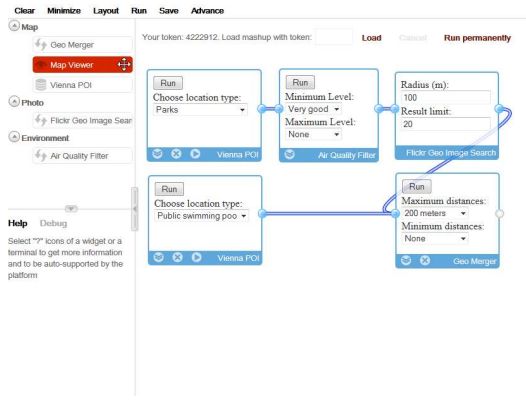
17

Modular approach for data integration



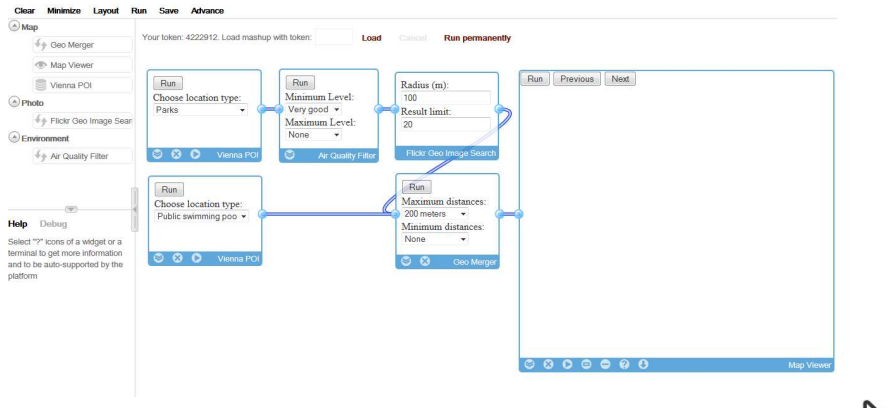
18

Modular approach for data integration



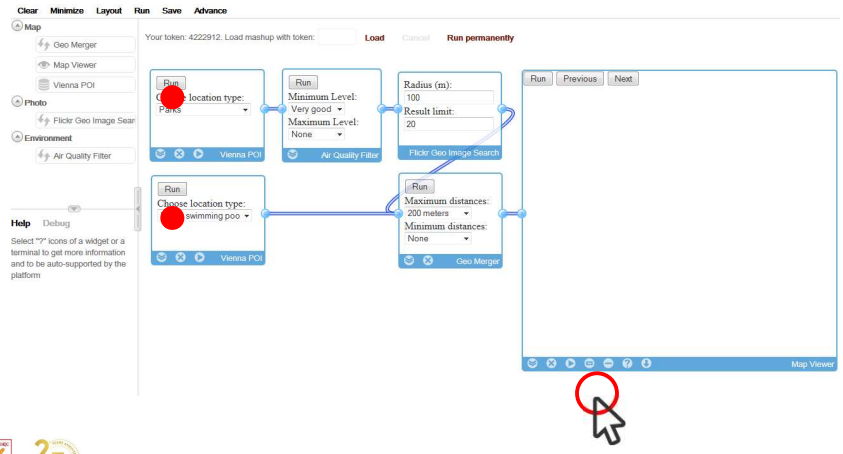
19

Modular approach for data integration



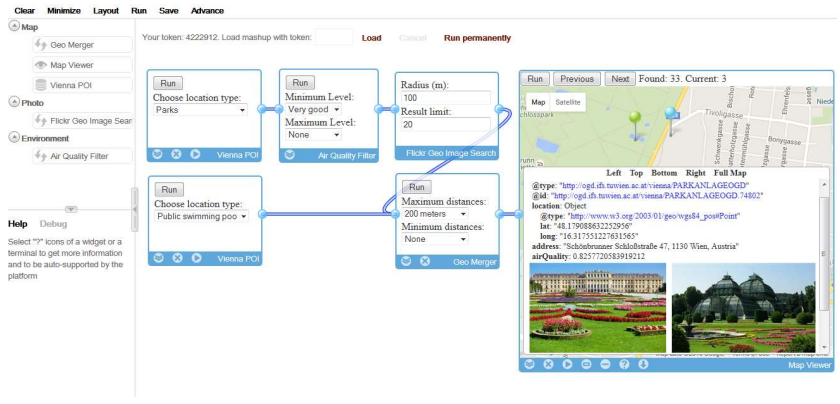
20

Modular approach for data integration



21

Modular approach for data integration



22

Mashup-based data integration

- Mashups: innovative paradigm that "*combines data from multiple sources into an integrated and single graphical interface*" [1]
  - are quick, flexible, and cost-effective
  - turn users from content consumers to content providers

[1] Yu, J., Benatallah, B., Casati, F., Daniel, F.: *Understanding mashup development*. IEEE Internet Computing (2008)

Challenges of mashup-based data integration research

1. Collaborative work among end-user, data publisher and developer communities is not considered
2. Integrating data that is distributed in different devices and not available on the web is not considered
3. Semantic mashup – "*A data mashup using RDF(S) as data model*" [2] – is still in its early stage

[2] Endres-Niggemeyer, Brigitte. *Semantic mashups: Intelligent Reuse of Web Resources* (2013)

Research Question

How can non-expert users be enabled to explore and integrate heterogeneous data sources?

RQ1: *How is it possible to support non-expert users in addressing data heterogeneity?*

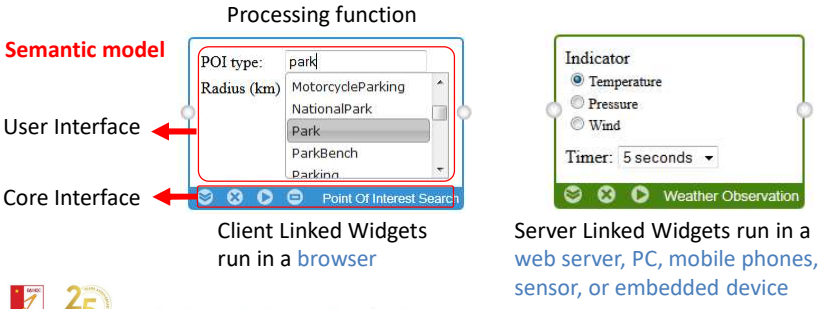
RQ2: *How can non-expert users be enabled to collaboratively integrate data?*

***RQ3**: How is it possible to automate the data exploration and integration process?*

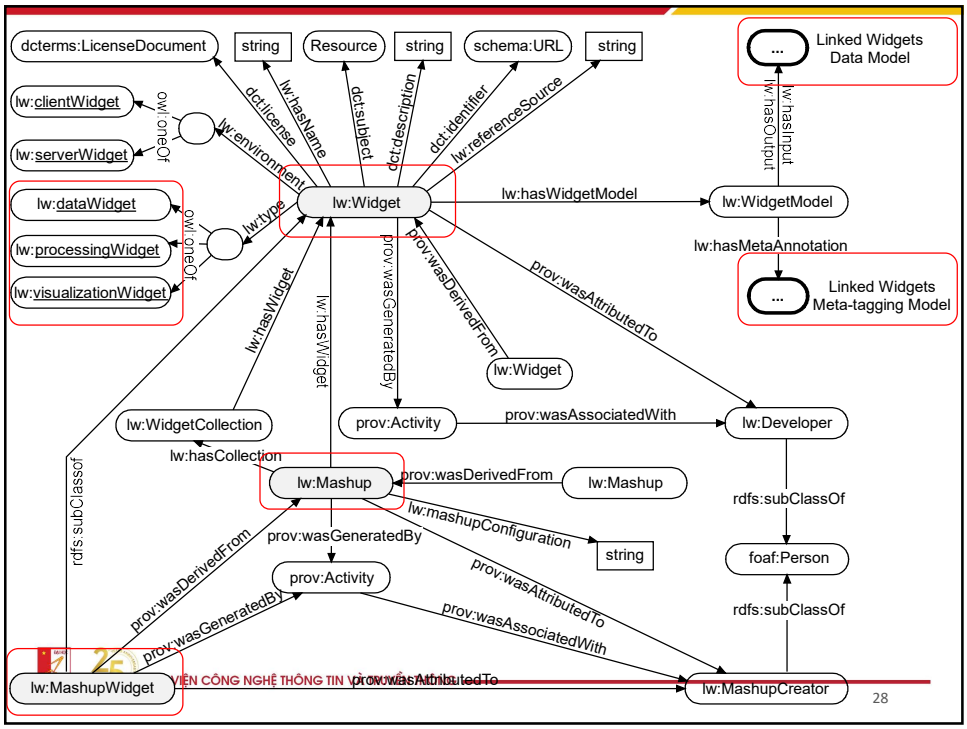
2. Linked Widgets Framework

Linked Widgets

- extends W3C standard web widgets with a semantic model
- lift data sources to a semantic level to facilitate data processing
- can be created by independent developers, hosted on different servers

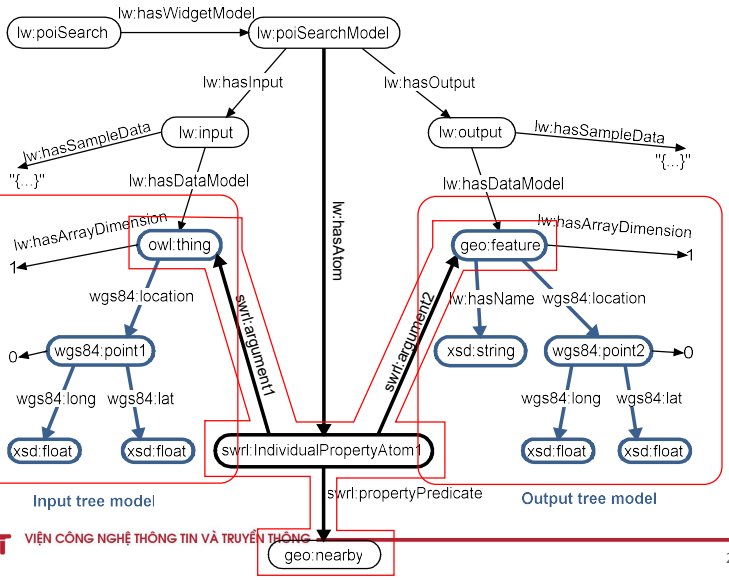


27



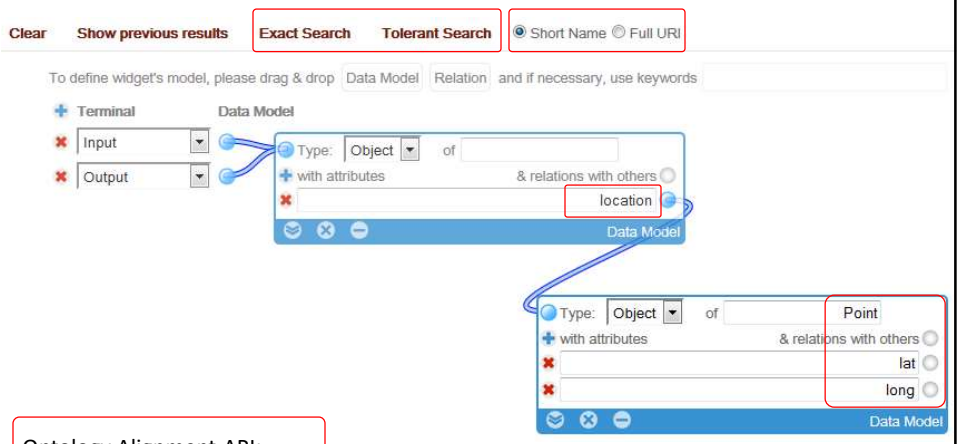
28

Example: Data model of the POI search widget



29

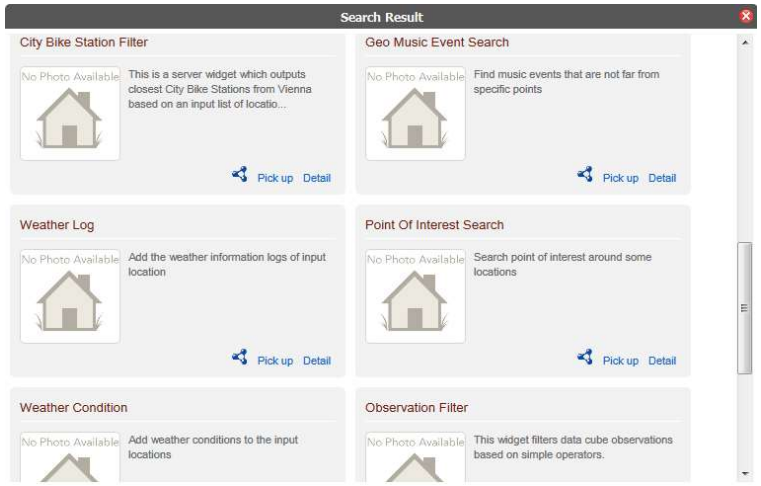
M1: Semantic Widget Discovery



Ontology Alignment API:
http://alignapi.gforge.inria.fr/

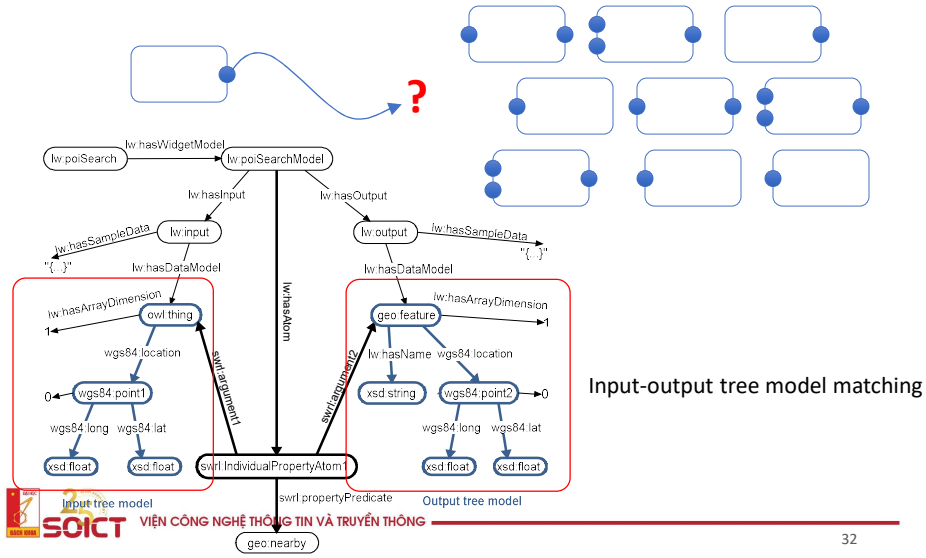
30

M1: Semantic Widget Discovery



31

M2: Terminal matching



Input-output tree model matching

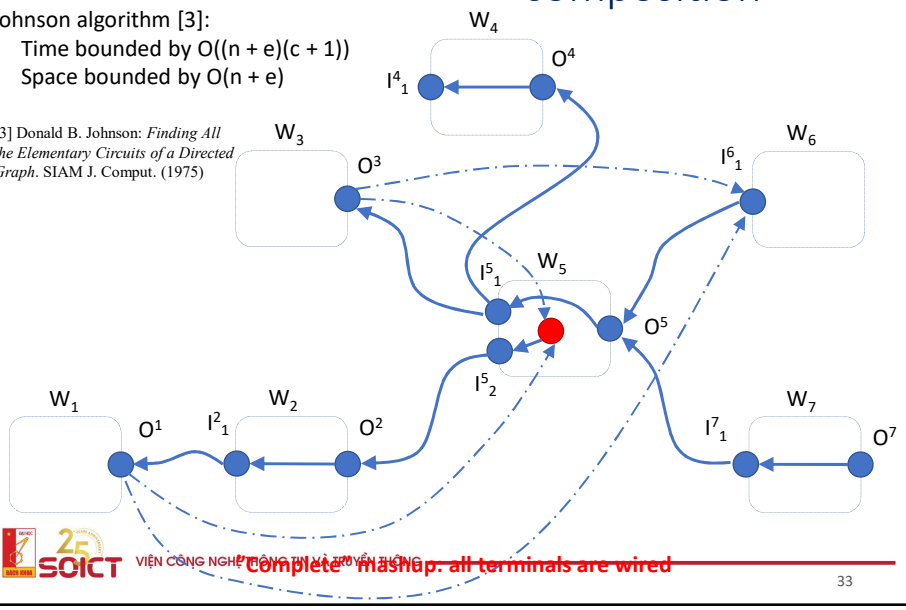
32

M3: Automatic mashup composition

Reduce "*automatic composition*" to "*list all cycles*" algorithm

Johnson algorithm [3]:
- Time bounded by O((n + e)(c + 1))
- Space bounded by O(n + e)

[3] Donald B. Johnson: *Finding All the Elementary Circuits of a Directed Graph*. SIAM J. Comput. (1975)



Complete mashup: all terminals are wired

33

M4: Interactive tag-based composer

1. Define requirement → 2. Locate related widgets → 3. Execute automatic composition algorithm → 4. Rank and return mashups

Enter text + tie words to DBpedia (Wikipedia) resources

Restaurant nearby Bank

http://dbpedia.org/resource/Restaurant

http://www.geonames.org/ontology#nearby

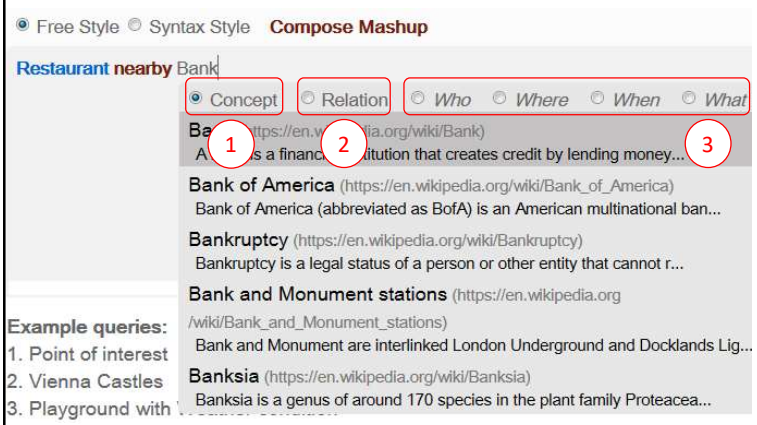
http://dbpedia.org/resource/Bank

A bank is a financial intermediary that creates credit by lending money to a borrower, thereby creating a corresponding deposit on the bank's balance sheet etc.

Benefits of the semantic tagging techniques
- No ambiguity
- Multiple language

34

Define mashup requirement



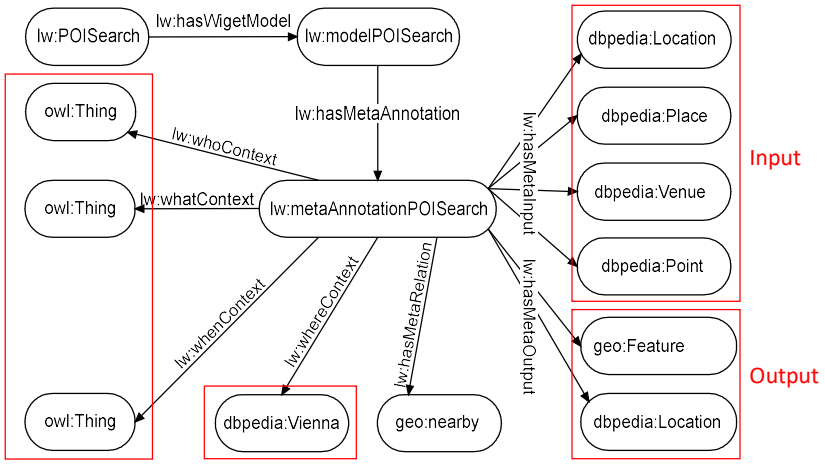
1. Concept tagging (DBpedia concepts)

2. Relation tagging (LOD properties)

3. Entity tagging DBpedia entities

35

Meta-tagging model



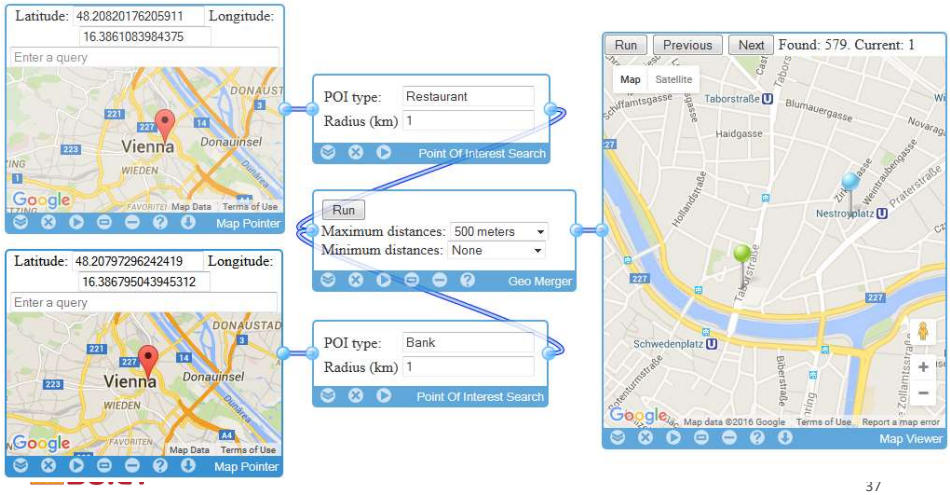
lw http://linkedwidgets.org/ontologies/ **dbpedia** http://dbpedia.org/resource/
owl http://www.w3.org/2002/07/owl# **geo** http://www.geonames.org/ontology#

Example: Meta-tagging model of the Vienna POI Search widget

36

Example

4 widgets: Map pointer, Geo merger, POI search, Map viewer



37

37

3. Mashup Models

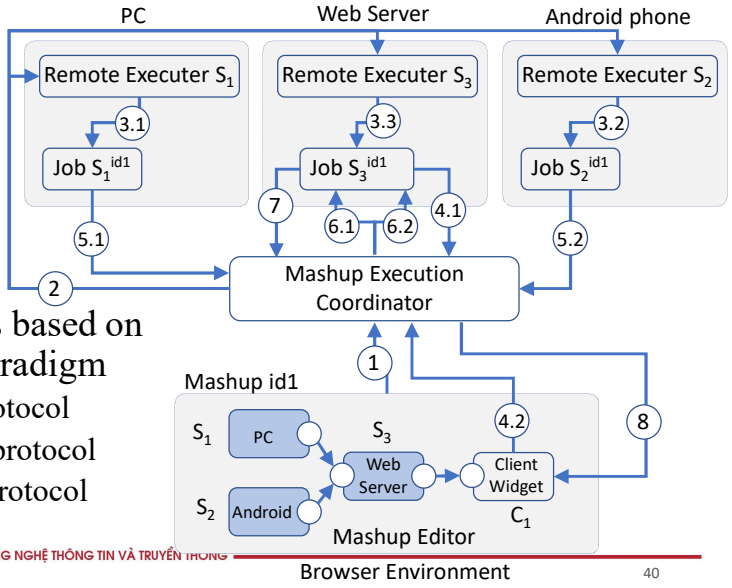
38

3. Mashup models

- Model 1: collaborative mashups
- Model 2: persistent mashups
- Model 3: streaming mashups
- Model 4: distributed mashups

39

Linked Widget communication protocols

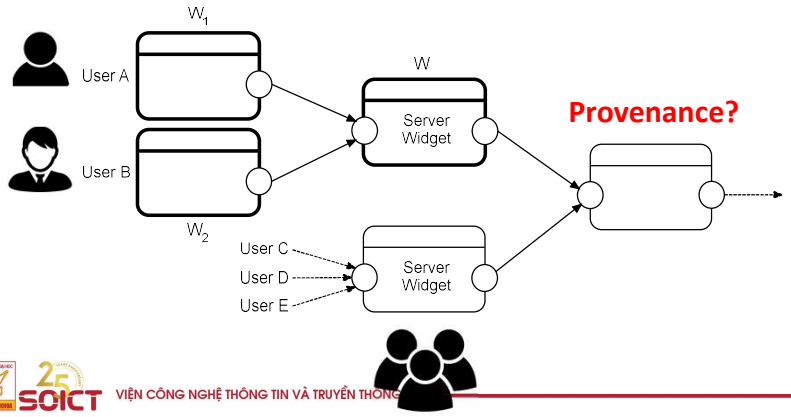


- 3 protocols based on Pub/sub paradigm
  - Local protocol
  - Remote protocol
  - Hybrid protocol

40

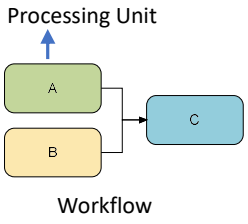
Model 1 – Collaborative mashups

- Mashups are created and/or operated by multiple users at the same time

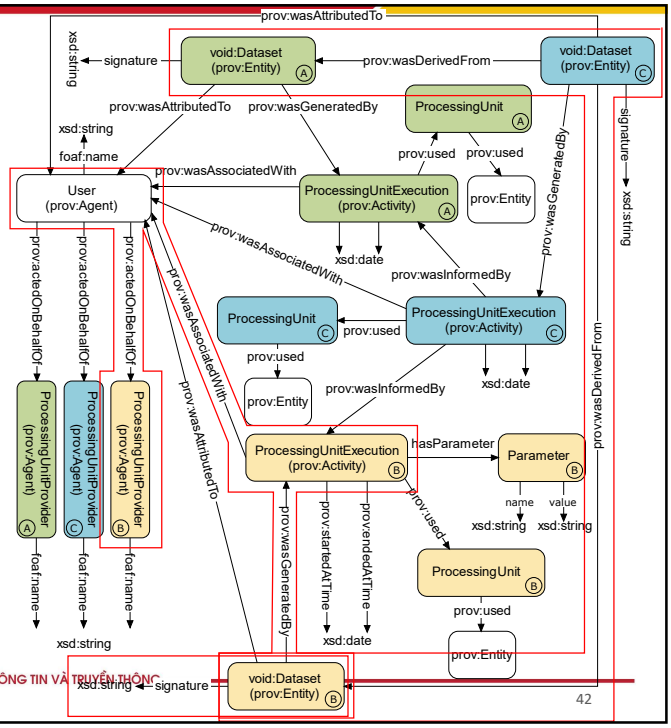


41

Provenance Model



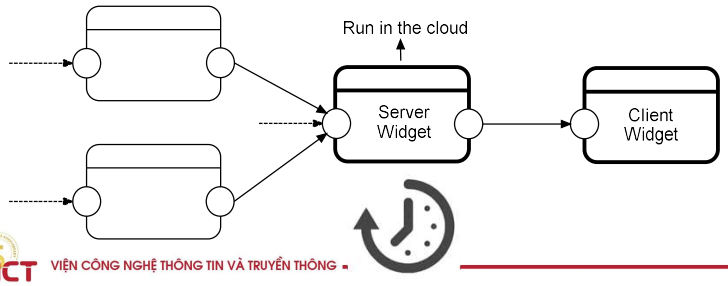
The model is based on **PROV-O ontology**



42

Model 2 – Persistent mashups

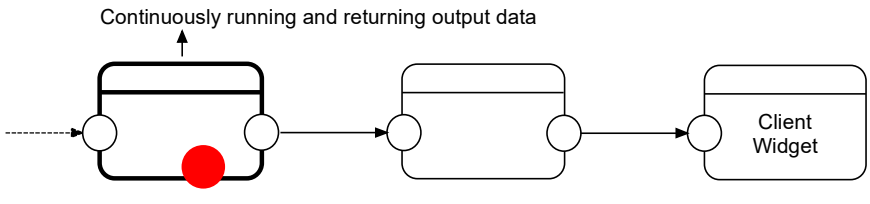
- run continuously in the background
- are useful for time-consuming data integration tasks



43

Model 3 – Streaming mashups

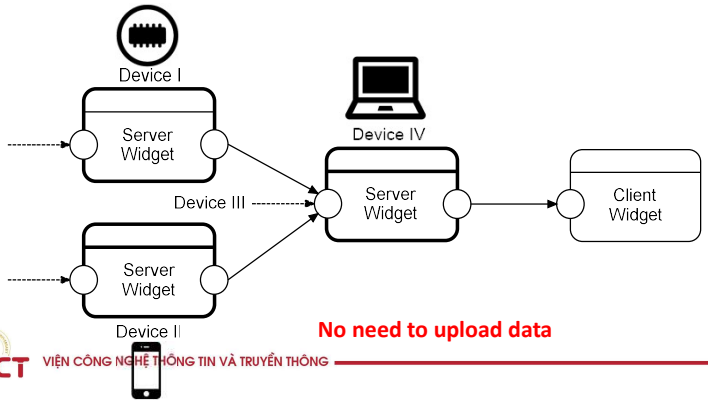
- Data is continuously flowing from a widget to others
- Streaming mashups are useful for data monitoring use cases.



44

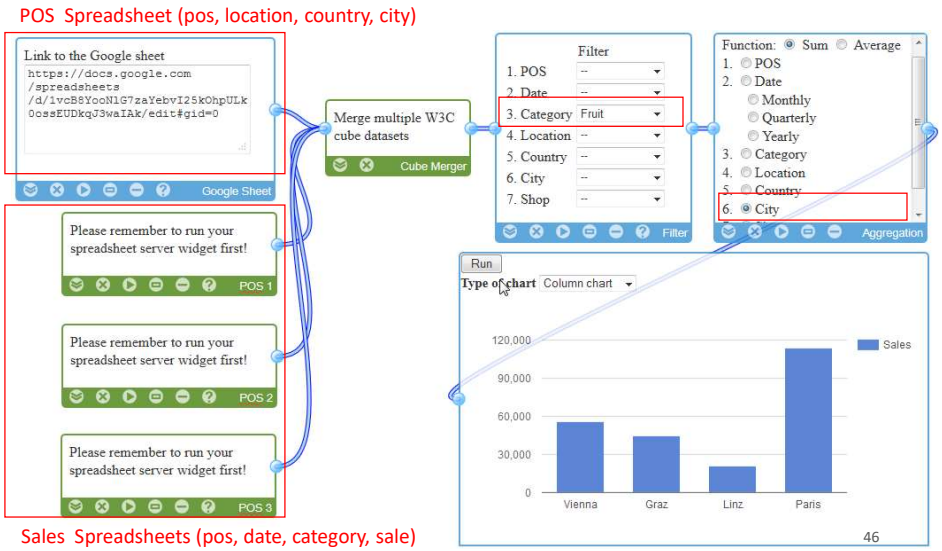
Model 4 – Distributed mashups

- Widgets are hosted by distributed nodes and devices
- Useful for integrating sensor data and data from embedded devices



45

Distributed & collaborative mashup example: Combine and visualize sales data for a series of retail points of sale (POS)



46

Conclusions

- Development of concepts combining semantic web and mashups
- Semantic model for widgets
- Facilitating collaborative work among end-user, data publisher and developer communities
- Introducing a new model of *semantic*, *distributed*, and *collaborative* mashups
- Running prototype: http://linkedwidgets.org