

Exploiting Semantics of Web Services for Geospatial Data Fusion

**Pedro Szekely, Craig A. Knoblock, Shubham Gupta,
Mohsen Taheriyani, Bo Wu**

University of Southern California

- **Decision makers have lots of data available**
 - Satellite imagery
 - Street maps
 - Structured online sources (e.g., phone books)
 - Cyber data (e.g., domain registration sites)
 - Social network data (e.g., facebook)
- **Difficult to fuse this information into an integrated view**
 - Even harder to apply various reasoning techniques
- **Our goal**
 - An integration framework where users can interactively fuse geospatial and other types of data

- Karma [R. Tuchinda, C. A. Knoblock, P. Szekely, Building mashups by demonstration, 2011]
- A fusion-by-example approach for extracting, modeling, cleaning and integrating geospatial sources
 - Does not require any programming or widget knowledge.
 - Focus on data, not on the process
 - Users specify fusion tasks by examples
 - Fusion results automatically displayed on a map

The screenshot displays the KARMA application interface, which is divided into two main sections: a data management interface on the left and a map view on the right.

Data Management Interface (Left):

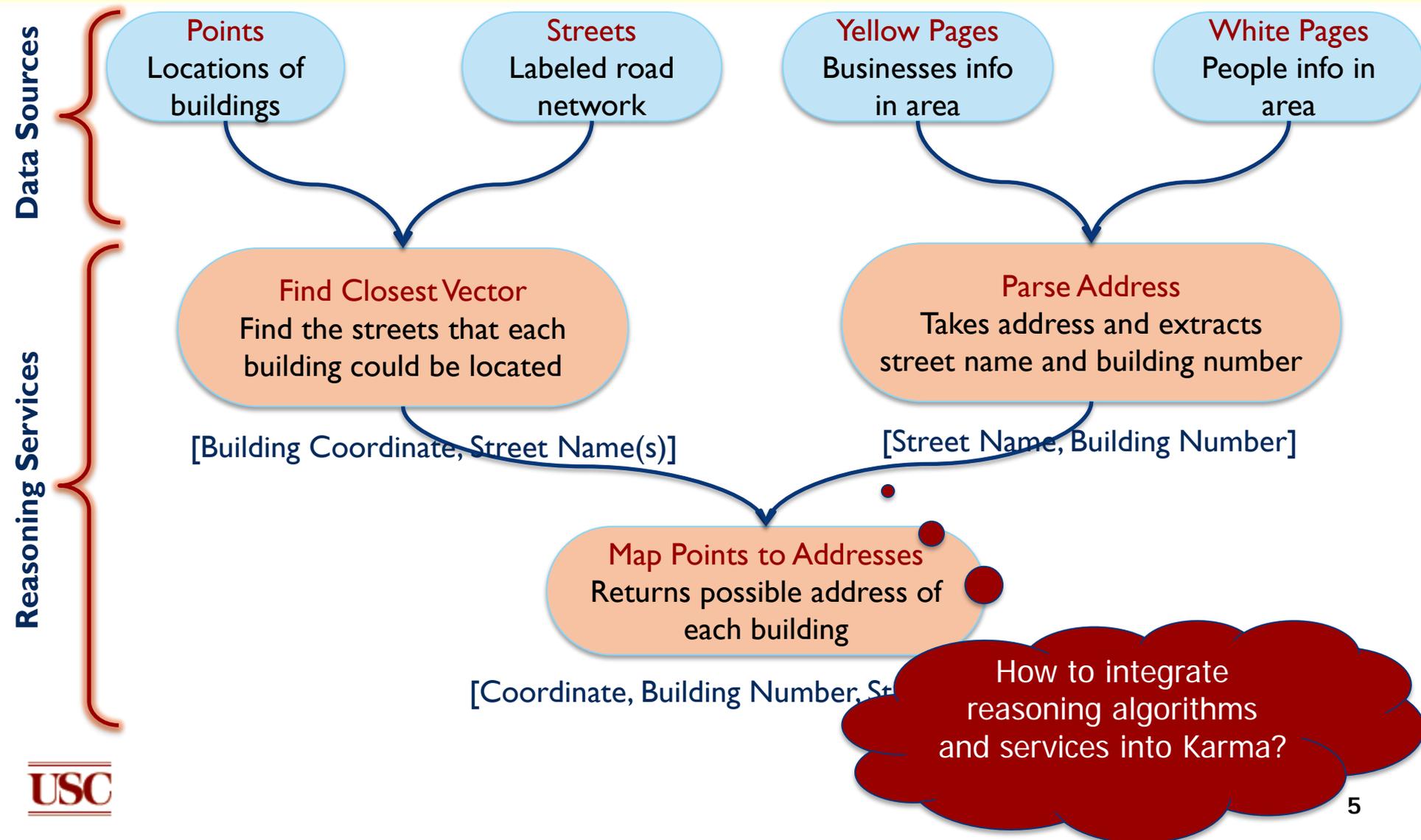
- Table View:** A table with columns for PR-String, PR-String, PR-String, PR-String, and Data Type. The data includes coordinates and placemark names like CARA LAZARA, NIKOLE SPASICA, and USKOČKA.
- Buttons:** Import, Clean, Integrate, Publish.
- Wrapper Selection:** Database, Excel, CSV, KML, WebService.
- File Selection:** Choose KML File.

Map View (Right):

- Map:** A satellite map showing a city street grid with red placemark markers and green lines representing data paths.
- Map Controls:** Map, Satellite, Hybrid.
- Browser Address Bar:** http://localhost:8080/infuse/tDemo.html
- Map Title Bar:** kmlnetworklink_example.html

Annotations:

- Spreadsheet Type Interface:** Points to the table view.
- Information Integration Operations:** Points to the Integrate button.
- Data Types Supported:** Points to the Wrapper selection area.



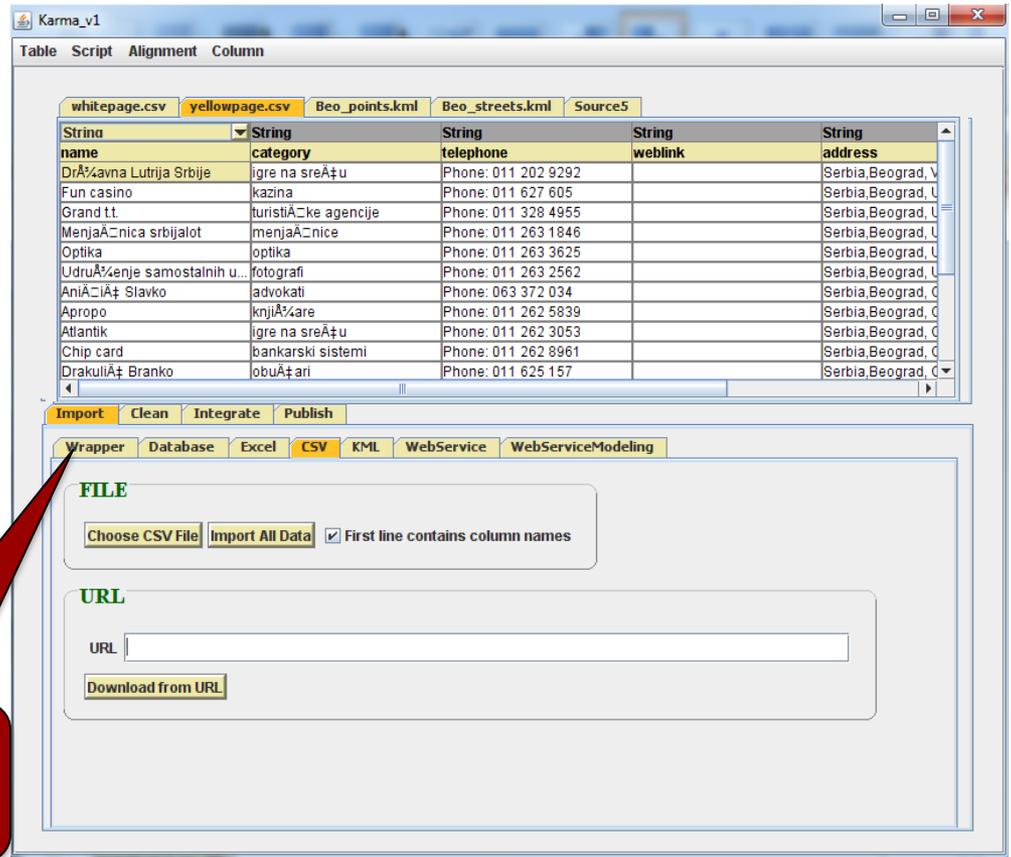
- Build a semantic model of reasoning services based on provided ontology
 - Data types of inputs and outputs, **plus** relationships between them
- Interactively invoke services using semantic model of sources and services
 - Which services can be invoked using available data?
 - Which sources can satisfy service inputs?
- Integrate outputs of service invocation with the other data

Importing Sources



The screenshot shows a web browser with a map interface. A search bar contains '988info.rs' and 'YellowPages.rs'. Below the search bar, there are filters for 'Advanced search' (By address, By brand, By phone) and a list of search results for 'Chinese restaurant Peking' and 'Bank Erste bank'. A red callout bubble points from the text below to the 'Wrapper' tab in the Karma interface.

Karma uses wrappers to extract web pages information



The screenshot shows the Karma_v1 interface. At the top, there is a table with columns 'name', 'category', 'telephone', 'weblink', and 'address'. Below the table, there are tabs for 'Wrapper', 'Database', 'Excel', 'CSV', 'KML', 'WebService', and 'WebServiceModeling'. The 'Wrapper' tab is selected, showing options for 'Choose CSV File', 'Import All Data', and 'First line contains column names'.

String	String	String	String	String
name	category	telephone	weblink	address
Državna Lutrija Srbije	igre na sreću	Phone: 011 202 9292		Serbia, Beograd, V
Fun casino	kazina	Phone: 011 627 605		Serbia, Beograd, U
Grand tt	turističke agencije	Phone: 011 328 4955		Serbia, Beograd, U
Menjačnica srbijalot	menjačnice	Phone: 011 263 1846		Serbia, Beograd, U
Optika	optika	Phone: 011 263 3625		Serbia, Beograd, U
Udruženje samostalnih u...	fotografi	Phone: 011 263 2562		Serbia, Beograd, U
Aničić Slavko	advokati	Phone: 063 372 034		Serbia, Beograd, C
Apropo	knjižare	Phone: 011 262 5839		Serbia, Beograd, C
Atlantik	igre na sreću	Phone: 011 262 3053		Serbia, Beograd, C
Chip card	bankarski sistemi	Phone: 011 262 8961		Serbia, Beograd, C
Drakulić Branko	obuštari	Phone: 011 625 157		Serbia, Beograd, C

Data Cleaning



whitepage.csv Source2 Cleaning Table		
String	Data Type	Data
Column Name	User Defined Values	Final
CARA LAZARA 15/2,BEOGRAD	CARA LAZARA 15/2.BEOGRAD	
CARA LAZARA 13/21,BEOGRAD		
NIKOLE SPASIĆA 2,BEOGRAD		

Karma uses learned transformation rules to remove all instances of

User provides examples of address without

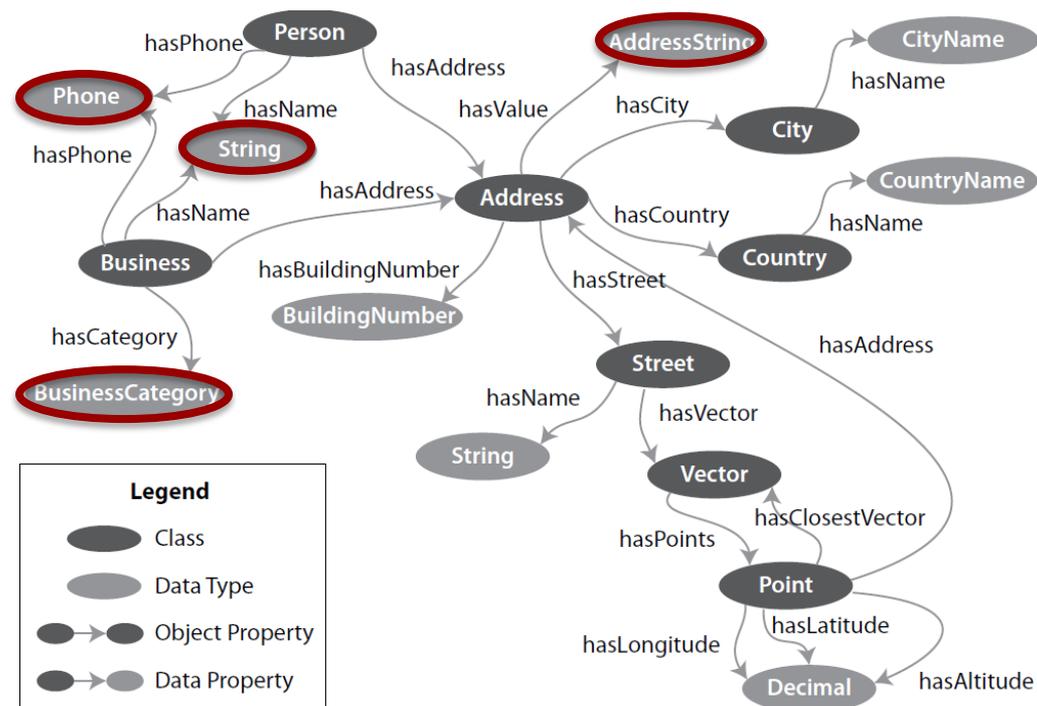
whitepage.csv Source2 Cleaning Table		
String	Data Type	Data
Column Name	User Defined Values	Final
CARA LAZARA 15/2,BEOGRAD	CARA LAZARA 15/2,BEOGRAD	
CARA LAZARA 13/21,BEOGRAD	CARA LAZARA 13/21,BEOGRAD	
NIKOLE SPASIĆA 2,BEOGRAD	NIKOLE SPASIĆA 2.BEOGRAD	

Source Modeling

- Karma automatically builds models of data according to provided ontology
 - Models help user to process data and integrate them
- Identify the semantic types
 - Supervised machine learning technique (CRF Model)
 - A. Goel, C. A. Knoblock, K. Lerman, Using conditional random fields to exploit token structure and labels for accurate semantic annotation, 2011
- Identify relationships among the data columns
 - Find the minimal tree that connects the semantic types
 - C. A. Knoblock, P. Szekely, J. L. Ambite, S. Gupta, A. Goel, M. Muslea, K. Lerman, Interactively Mapping Data Sources into the Semantic Web, 2011

Modeling YellowPages Source

I. Karma uses CRF technique to assign labels to each data column

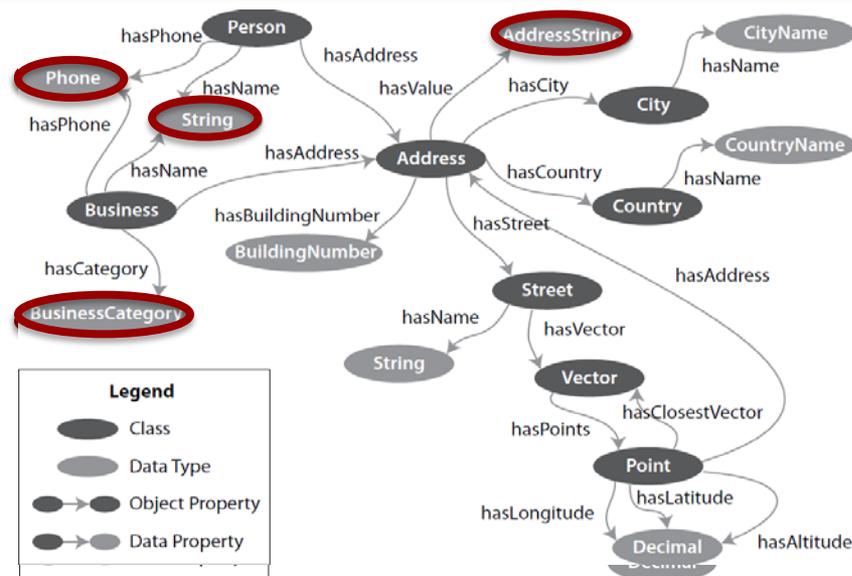


Semantic Types

String	BusinessCategory	AddressString	Phone
name	category	address	telephone
Državna Lutrija Srbije	igre na sreću	Serbia,Beograd, Vračar, Uskoč...	Phone: 011 202 9292
Fun casino	kazina	Serbia,Beograd, Uskočka 4	Phone: 011 627 605
Grand t.t.	turističke agencije	Serbia,Beograd, Uskočka 7	Phone: 011 328 4955
Menjačnica srbijalot	menjačnice	Serbia,Beograd, Uskočka 4	Phone: 011 263 1846

Modeling YellowPages Source

II. Karma selects the smallest tree that connects these semantic types and shows it at the top of the data worksheet.



Relationships between columns

yellowpage.csv Source2

Business			
(hasName) String	(hasCategory) BusinessCategory	(hasAddress) Address	(hasPhone) Phone
		(hasValue) AddressString	
String	BusinessCategory	AddressString	Phone
name	category	address	telephone
Državna Lutrija Srbije	igre na sreću	Serbia,Beograd, Vračar, Uskoč...	Phone: 011 202 9292
Fun casino	kazina	Serbia,Beograd, Uskočka 4	Phone: 011 627 605
Grand t.t.	turističke agencije	Serbia,Beograd, Uskočka 7	Phone: 011 328 4955
Menjačnica srbijalot	menjačnice	Serbia,Beograd, Uskočka 4	Phone: 011 263 1846

Modeling Web Services

- Semantic models of web services facilitate service invocation, discovery, and composition
- Karma allow the user to interactively build a model
 - User provides examples of service input and output
 - Modeling services can be done like data sources

Modeling of Parse Address Service in Karma

Parse Address

```

classDiagram
    class AddressString
    class Address
    class BuildingNumber
    class Street
    AddressString --> Address : hasRawString
    Address --> BuildingNumber : hasBuildingNumber
    Address --> Street : hasStreet
    BuildingNumber --> Street : hasName
    
```

AddressString	String
CARA LAZARA 15/2, BEOGRAD	CARA LAZARA
CARA LAZARA 13/21, BEOGRAD	CARA LAZARA
NIKOLE SPASIĆA 2, BEOGRAD	NIKOLE SPASIĆA

AddressString	String	BuildingNumber
CARA LAZARA 15/2, BEOGRAD	CARA LAZARA	15
CARA LAZARA 13/21, BEOGRAD	CARA LAZARA	13
NIKOLE SPASIĆA 2, BEOGRAD	NIKOLE SPASIĆA	2

AddressString	String	BuildingNumber
CARA LAZARA 15/2, BEOGRAD	CARA LAZARA	15
CARA LAZARA 13/21, BEOGRAD	CARA LAZARA	13
NIKOLE SPASIĆA 2, BEOGRAD	NIKOLE SPASIĆA	2

Choose S Class:

Address

Business

City

hasBuildingNumber (Domain: Address)

hasCategory (Domain: Business)

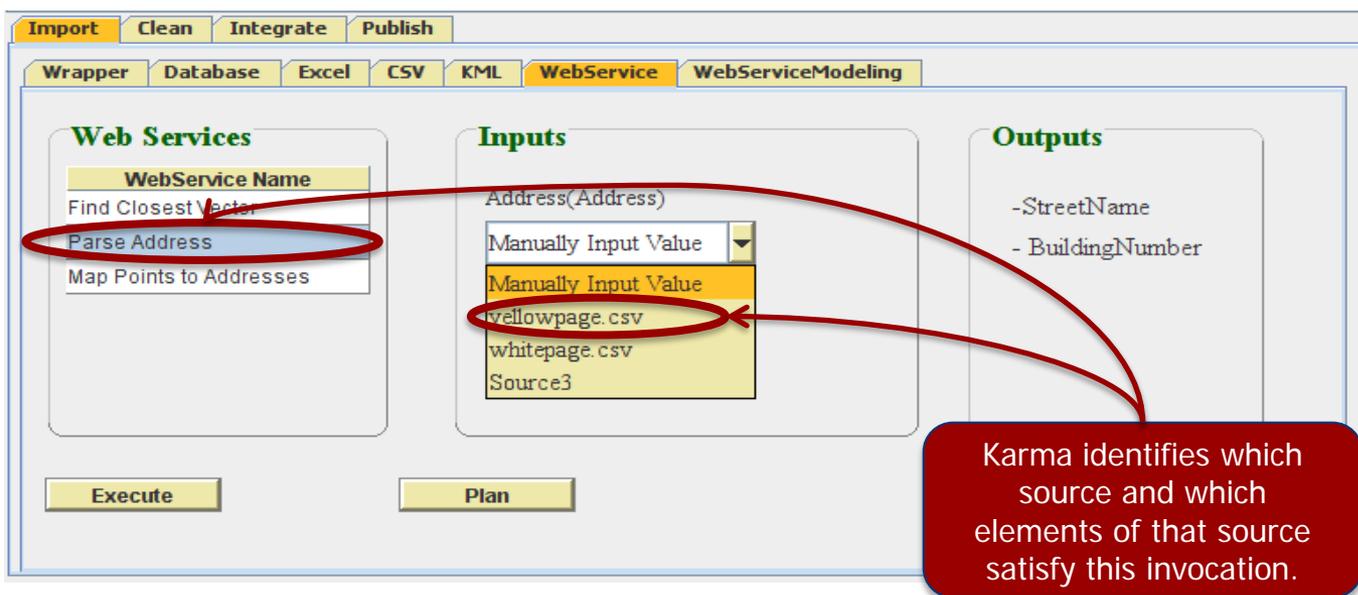
AddressString	String	BuildingNumber
CARA LAZARA 15/2, BEOGRAD	CARA LAZARA	15
CARA LAZARA 13/21, BEOGRAD	CARA LAZARA	13
NIKOLE SPASIĆA 2, BEOGRAD	NIKOLE SPASIĆA	2

Create new one:

3 Final service model

Data Fusion

- Ability for users to interactively invoke services on other data sources
- Semantic models make it possible to:
 - Automatically determine which services apply to the available data
 - Perform automatic transformations on data to get it into the required format to apply a service
 - Automatically compose services and sources to generate required data



The screenshot shows the Karma web service interface. At the top, there are tabs for 'Import', 'Clean', 'Integrate', and 'Publish'. Below these are sub-tabs for 'Wrapper', 'Database', 'Excel', 'CSV', 'KML', 'WebService', and 'WebServiceModeling'. The main area is divided into three panels: 'Web Services', 'Inputs', and 'Outputs'. In the 'Web Services' panel, 'Parse Address' is selected. In the 'Inputs' panel, 'yellowpage.csv' is selected from a dropdown menu. In the 'Outputs' panel, '- StreetName' and '- BuildingNumber' are listed. A red callout box at the bottom right contains the text: 'Karma identifies which source and which elements of that source satisfy this invocation.' Red arrows point from the callout box to the 'Parse Address' service and the 'yellowpage.csv' input.

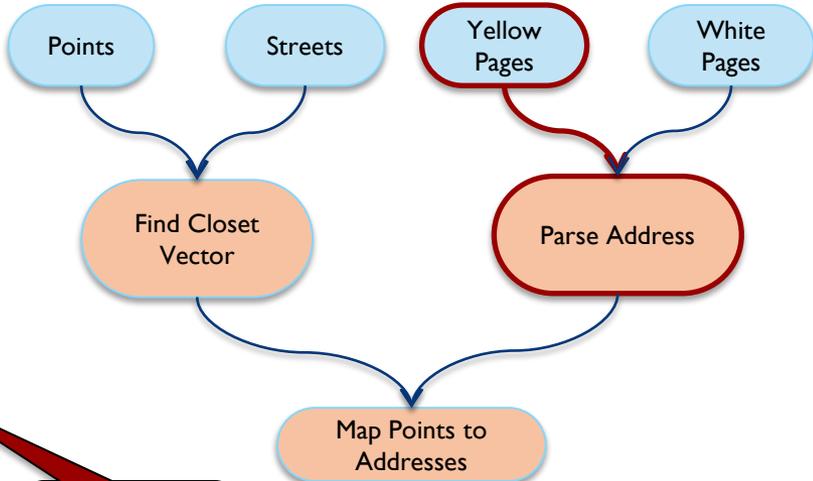
Matching Sources and Services

yellowpage.csv Source2

Business

(hasName) String	(hasCategory) BusinessCategory	(hasAddress) Address (hasValue) AddressString	(hasPhone) Phone
String	BusinessCategory	AddressString	Phone
name	category	address	telephone
Državna Lutrija Srbije	igre na sreću	Serbia,Beograd, Vračar, Uskoč...	Phone: 011 202 9292
Fun casino	kazina	Serbia,Beograd, Uskočka 4	Phone: 011 627 605
Grand t.t.	turističke agencije	Serbia,Beograd, Uskočka 7	Phone: 011 328 4955
Menjačnica srbijalot	menjačnice	Serbia,Beograd, Uskočka 4	Phone: 011 263 1846

Yellow Pages



Input

Output

Parse Address

Address

(hasValue) AddressString	(hasStreet) Street (hasName) String	(hasBuildingNumber) BuildingNumber
AddressString	String	BuildingNumber
Address	Street	Number
CARA LAZARA 15/2, BEOGRAD	CARA LAZARA	15
CARA LAZARA 13/21, BEOGR...	CARA LAZARA	13
NIKOLE SPASIĆA 2, BEOGRAD	NIKOLE SPASIĆA	2

Invocation Results

- Results of invocation are returned as another source that can be refined, integrated with other sources, visualized or published

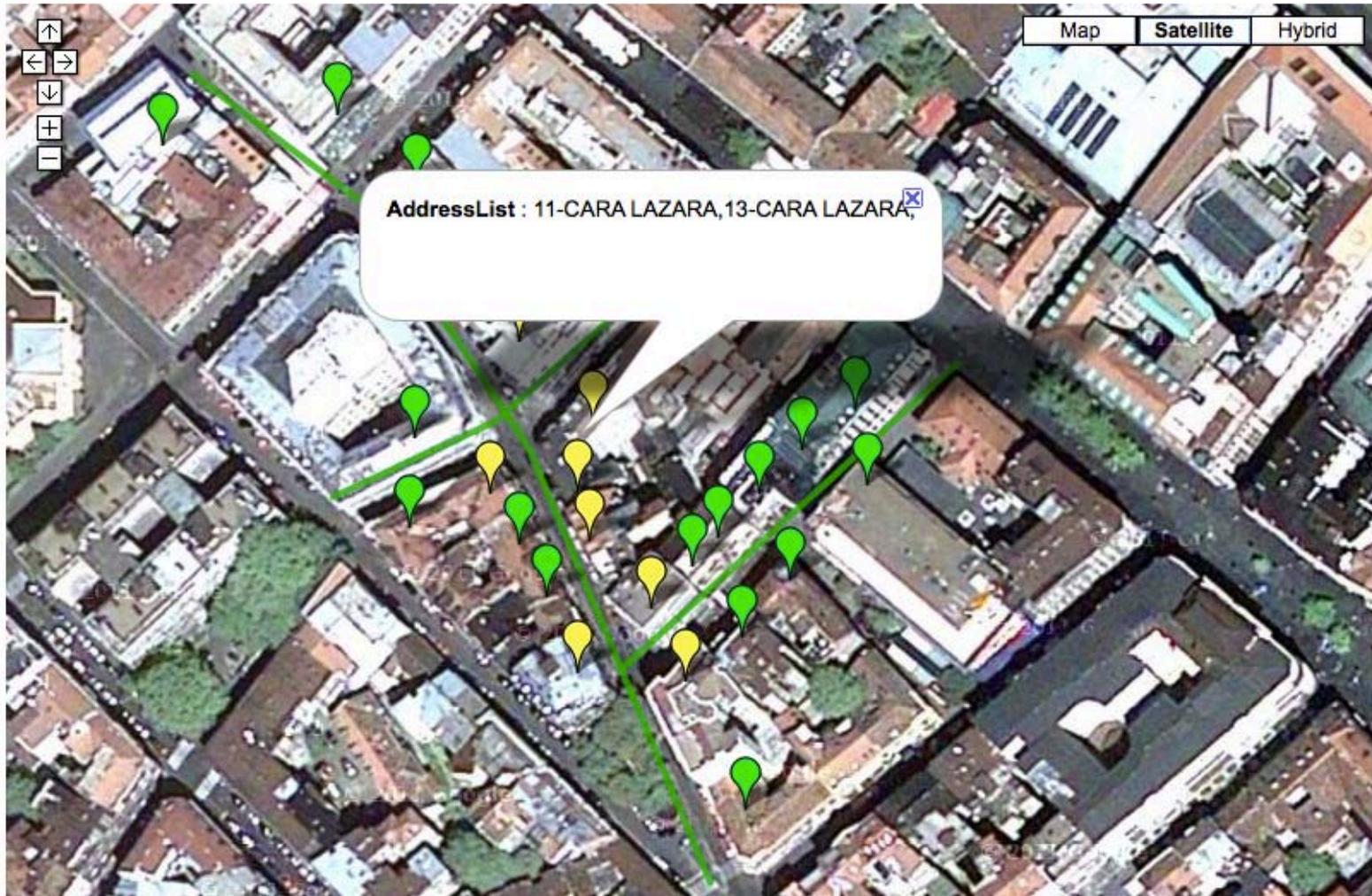
GeoFusionResult Source2		
✎ Point		
✎ (hasLongitude) Decimal	✎ (hasLatitude) Decimal	✎ (hasAddress) Address
		✎ (hasValue) AddressString
Decimal	Decimal	AddressString
Longitude	Latitude	AddressList
20.45329213142395	44.81892469068898	4-CARA LAZARA
20.45383930206299	44.8189931839843	1-CARA LAZARA
20.45408606529236	44.81883336616863	3-CARA LAZARA
20.454246997833252	44.81861266464735	5-CARA LAZARA
20.454407930374146	44.81848328750073	7-CARA LAZARA, 2A-NIKOLE...
20.45408070087433	44.81826638986837	4-NIKOLE SPASICA
20.45455013007800	44.81854036568043	2A-NIKOLE SPASICA

Visualize Final output



Information Sciences Institute

yellowpage.csv whitepage.csv Beo_streets.kml Beo_points.kml Find Closest vector Parse Address Map Points to Addresses



- **Exploit ontologies to attach semantics to geospatial services**
 - [L. Di, et. al., 2006], [P. Yue, et. al. , 2010]
 - User has to manually annotate the services according to an ontology like OWL-S
 - They model input and output types but not relationship among them
- **Linked Open Services (LOS)**
 - [B. Norton, R. Krummenacher, 2010]
 - Services that consume linked data as input and also return linked data as output
 - Use SPARQL to describe service inputs and outputs
 - Describing services might be easy for Linked Data community, but not for average Internet users
- **Google Fusion Tables**
 - [H. Gonzalez, A. Halevy, et al. 2010]
 - Import data from various source types and invoke web services
 - Allows advanced visualization
 - Integrating data from different sources is possible but without exploiting semantics

- Karma allows users to quickly and easily dynamically fuse a wide variety of geospatial data sources
- Modeling geospatial services is a big step in geospatial data fusion
- Based on provided ontology, Karma semi-automatically builds a semantic model of reasoning services including both input/output datatypes and their relationships
- Semantic descriptions enable user to easily find the desired service and invoke it using available data sources

- Applying the service modeling techniques to available REST web services
 - Create the service model just based on service invocation samples
- Answer queries like “Can I have the street names of the cities whose distance to Los Angeles is less than 50 miles?”
 - Automatically compose available web services using loaded data sources
- Publishing semantic description of web services in formats such as LOS

- More information available on Karma:
 - <http://www.isi.edu/~knoblock>
- Contact:
 - pszekely@isi.edu
 - knoblock@isi.edu
 - shubhamg@isi.edu
 - mohsen@isi.edu
 - bowu@isi.edu
- Software:
 - Software will be available as open source under the Apache license as soon as we complete the next version