USC Viterbi School of Engineering

Information Sciences Institute

Rapidly Integrating Services into the Linked Data Cloud

Mohsen Taheriyan, Craig Knoblock,
Pedro Szekely, Jose Luis Ambite

11th International Semantic Web Conference (ISWC 2012)
Boston, USA

OPENLINK SOFTWARE

dbpedia.org

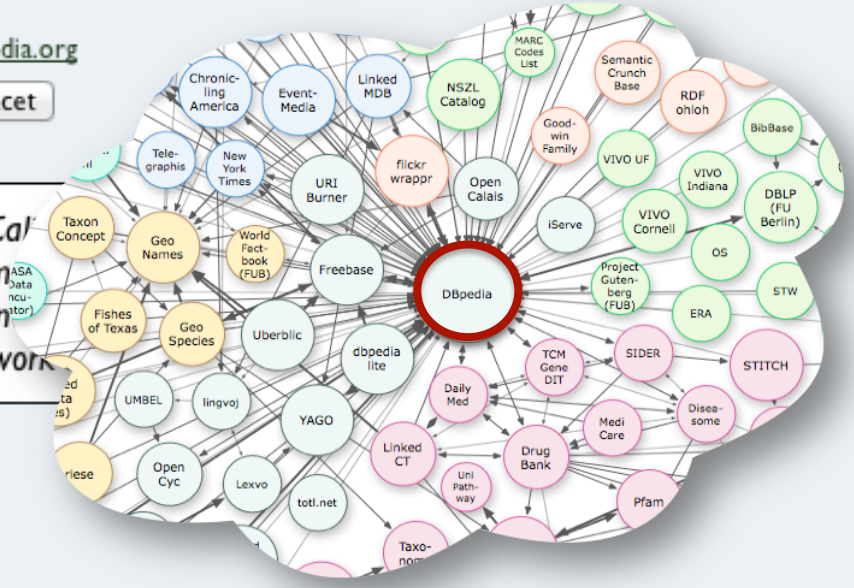
About: Downtown Los Angeles

Sponge Permalink

An Entity of Type : Neighborhoods in Los Angeles, California, within Data Space : dbpedia.org

Type: yago:NeighborhoodsInLosAngeles,California Constrain facet

Downtown Los Angeles is the central business district of Los Angeles, Cal
center of the metropolitan area. The area features many of the city's m
opportunities, a variety of skyscrapers and associated large multination
shopping opportunities. Downtown is the hub of the city's freeway networ



leader title	City Council U.S. House State Senate State Assembly
population total	39537(xsd:integer)
area total (km2)	8.538 13.73988692533248
foaf:homepage	http://www.downtownla.com
foaf:depiction	

OPENLINK SOFTWARE

Not in dbpedia.org

About: Downtown Los Angeles

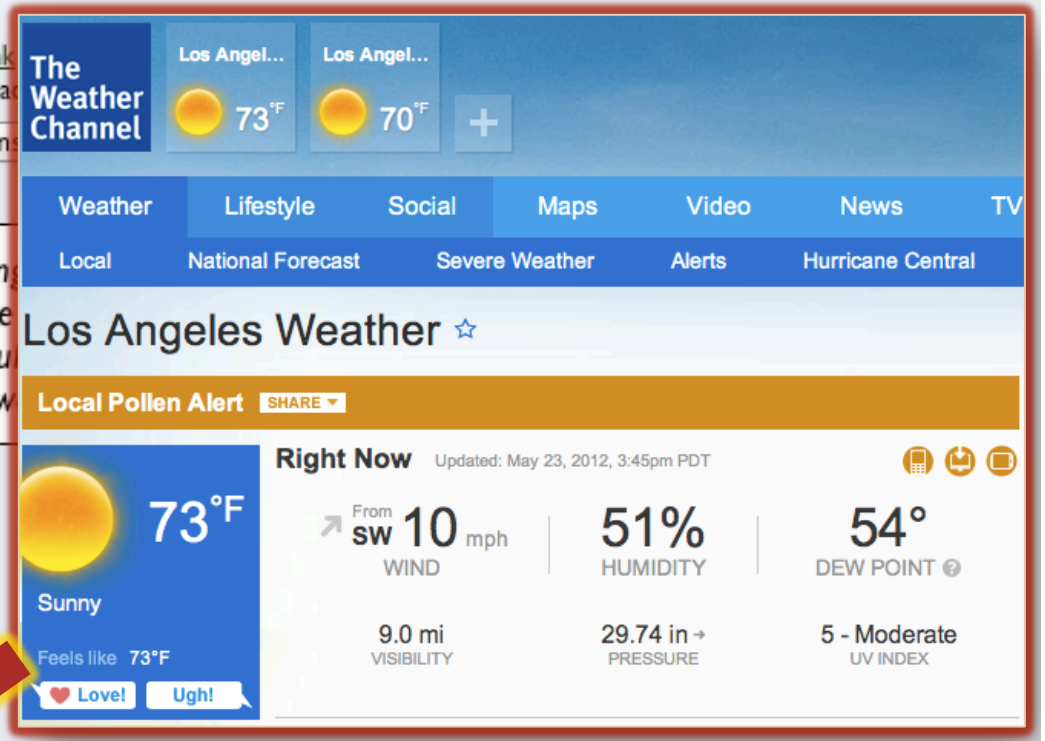
Sponge Permalink

An Entity of Type : Neighborhoods in Los Angeles, California, within Data Spa

Type: yago:NeighborhoodsInLosAngeles,California

Downtown Los Angeles is the central business district of Los An
center of the metropolitan area. The area features many of the
opportunities, a variety of skyscrapers and associated large mu
shopping opportunities. Downtown is the hub of the city's freew

leader title	City Council U.S. House State Senate State Assembly
population total	39537(xsd:integer)
area total (km2)	8.538 13.73988692533248



Current temperature	**73°F**
Wind	**10 mph SW**
Humidity	**51%**
Dew Point	**54°**

Live weather

OPENLINK SOFTWARE

Not in dbpedia.org

About: Downtown Los Angeles

Sponge Permalink

An Entity of Type : Neighborhoods in Los Angeles, California, within Data Space : dbpedia.org

Type: yago:NeighborhoodsInLosAngeles,California

Downtown Los Angeles is the central business district c
center of the metropolitan area. The area features ma
opportunities, a variety of skyscrapers and associated l
shopping opportunities. Downtown is the hub of the cit

leader title	City Council U.S. House State Senate State Assembly
population total	39537(xsd:integer)
area total (km2)	8.538 13.73988692533248
Current temperature	73°F
Wind	10 mph SW
Humidity	51%
Dew Point	54°
Events	Transparent Cities 5/1/2012 REDCAT Theater Saving Our Sons : A Community Conversatio… …

Events for May. 2012

Film
Transparent Cities
5/1/2012
REDCAT Theater

Meeting
Saving Our Sons: A Community Conversation
5/2/2012
Los Angeles Trade Tech College

Special Event
Autumn Lights Night
5/3/2012
401 s Main Street

Concert
Martin Hasselbock Leads LACO's Baroque Conversations
5/3/2012
Colburn School of Performing Arts

Ongoing Events

Exhibit
A Nation Emerges: The Mexican Revolution Revealed
9/8/2011 - 6/3/2012
LA Public Library - Central

Museum
Justice, Balance and Achievement: African Americans in the California Courts
1/12/2012 - 5/27/2012
California African American Museum

Exhibit
Trouble In Paradise: Music and Los Angeles, 1945-1975
2/22/2012 - 6/3/2012
GRAMMY Museum

Exhibit
Xploration Lab 2012
3/17/2012 - 6/17/2012

Events

OPENLINK SOFTWARE

Not in dbpedia.org

About: Downtown Los Angeles

Sponge Permalink

An Entity of Type : Neighborhoods in Los Angeles, California, within Data Space : dbpedia.org

Type: yago:NeighborhoodsInLosAngeles,California | Constrain facet

Downtown Los Angeles is the central business district of Los Angel
center of the metropolitan area. The area features many of the cit
opportunities, a variety of skyscrapers and associated large multin
shopping opportunities. Downtown is the hub of the city's freeway

leader title	City Council U.S. House State Senate State Assembly
population total	39537(xsd:integer)
area total (km2)	8.538 13.73988692533248

Current temperature	73°F
Wind	10 mph SW
Humidity	51%
Dew Point	54°
Events	Transparent Cities 5/1/2012 RED Saving Our Sons : A Community …
Twitter Feed	Los Angeles becomes the largest U.S. city to ban.. Plastic grocery bags will now be banned in … …

Yale Environment 360 @YaleE360 18m
Los Angeles becomes the largest U.S. city to ban **plastic bags** bit.ly/JA0PdX #LosAngeles #plasticbagban
Expand

DestinationTampaBay @DestTampaBay 48m
Plastic grocery **bags** will now be banned in #LosAngeles. What do you think of that concept for #Florida?
Expand

annmarie anderson @AnnmarieCloud 55m
"@HuffingtonPost: The nation's second largest city has just banned **plastic bags** in grocery stores huff.to/Jygs5N" #LosAngeles #green
Expand

gomaam @gomaam 1h
#LosAngeles is now the largest U.S. city to #ban #**plastic bags** dlvr.it/1cX0bW
Expand

Ozzyopolis @Ozzyopolis 3h
Los Angeles Becomes Largest U.S. City to Ban **Plastic Bags**: bit.ly/Jpl9xw via @TreeHugger #LosAngeles #plasticbagban
Expand Reply Retweet Favorite

Live Twitter Feed

OPENLINK SOFTWARE

The information is available in Web APIs

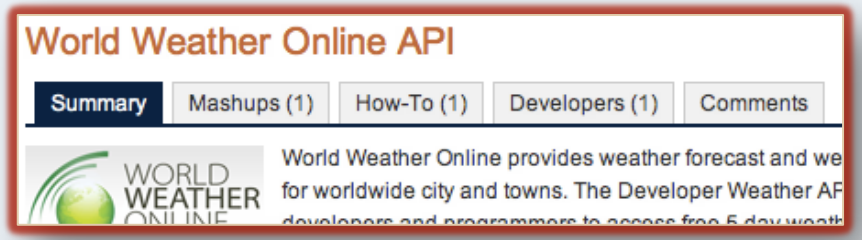
About: Downtown Los Angeles
An Entity of Type : Neighborhoods in Los Angeles, California

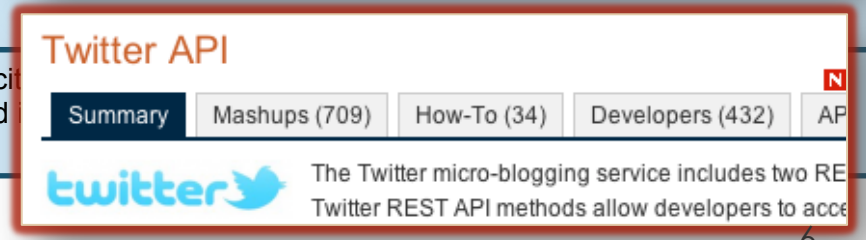
Type: yago:NeighborhoodsInLosAngeles,California

Downtown Los Angeles is the central business district of Los Angeles, California, United States, located close to the geographic center of the metropolitan area. The area features many of the city's major arts institutions and sports facilities, sightseeing opportunities, a variety of skyscrapers and associated large multinational corporations and an array of public art and unique shopping opportunities. Downtown is the hub of the city's freeway network and growing Metro rapid transit system.

leader title	City Council U.S. House State Senate State Assembly
population total	39537(xsd:integer)
area total (km2)	8.538 13.73988692533248

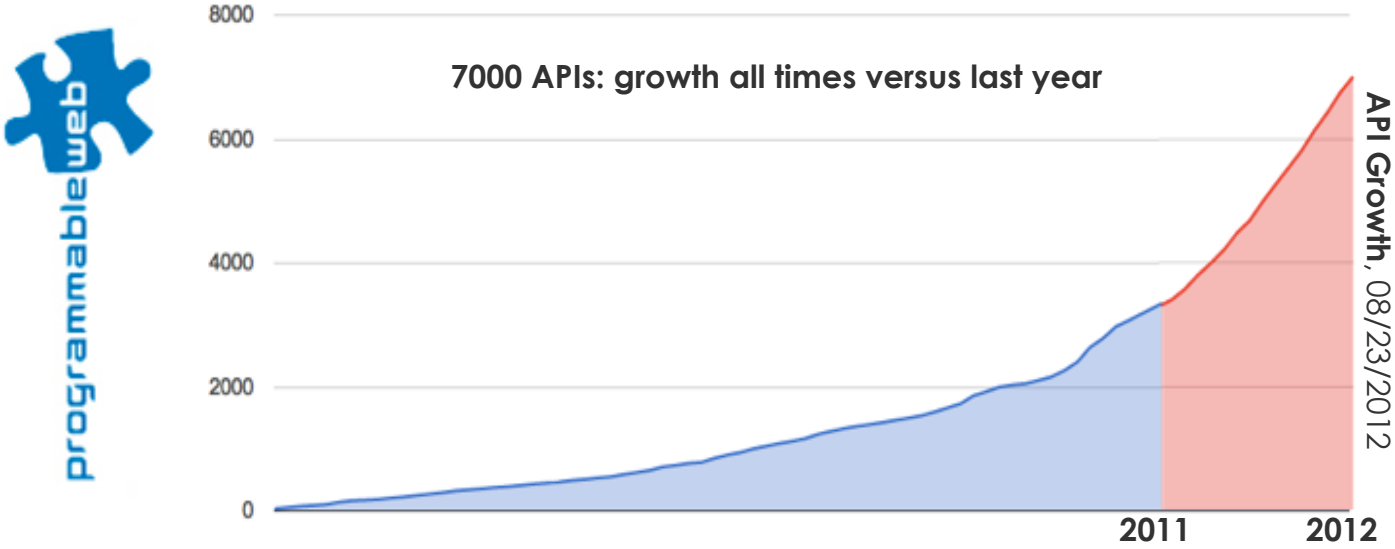
Current temperature	73°F
Wind	10 mph SW
Humidity	51%
Dew Point	54°
Events	Transparent Cities 5/1/2012 REDCAT Th… Saving Our Sons : A Community Conversatio… …
Twitter Feed	Los Angeles becomes the largest U.S. ci… Plastic grocery bags will now be banned … …

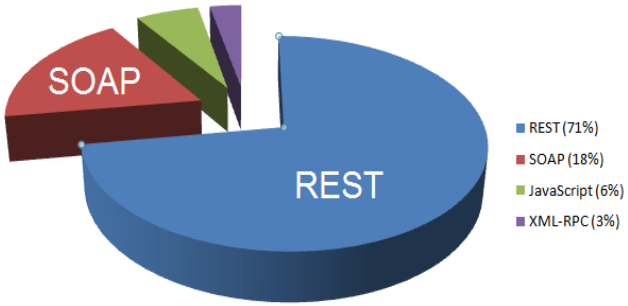




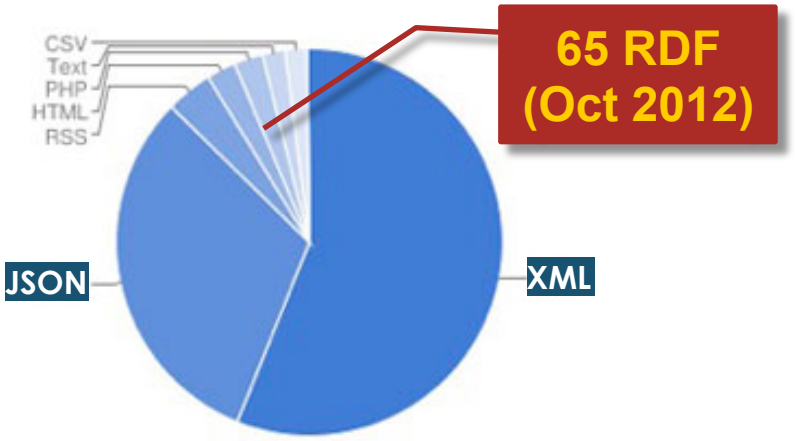
Web APIs

USC Viterbi School of Engineering





API Protocols, 02/10/2012

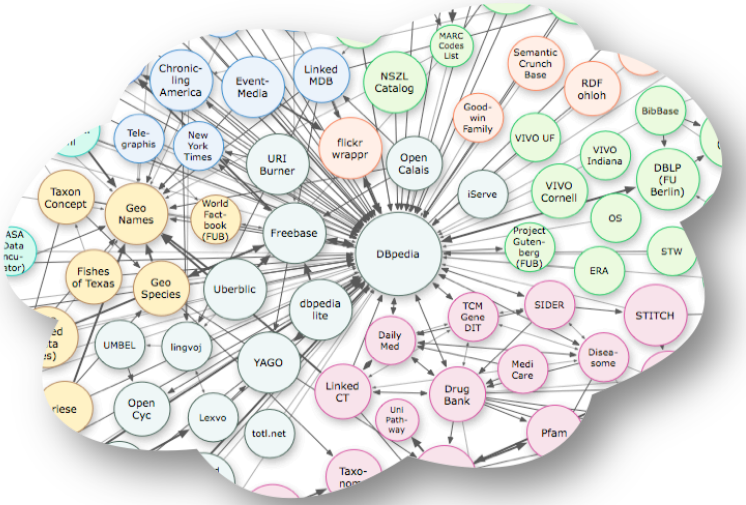


API Data Formats, 03/08/2011

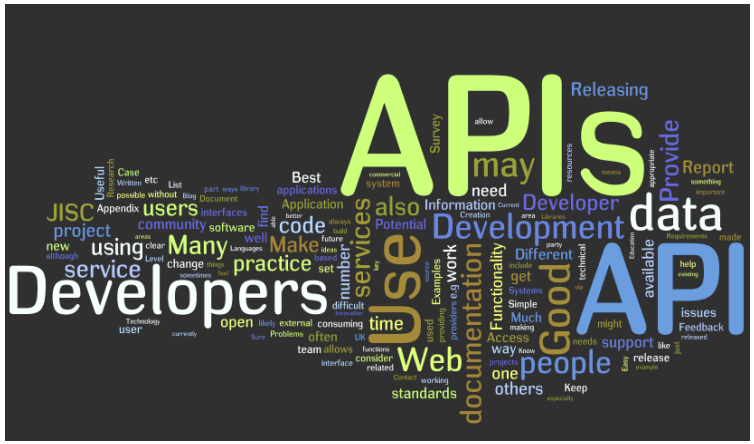
The Problem

USC Viterbi School of Engineering

RDF



RDF



API Invocation

JSON/XML

Motivation Example

USC Viterbi School of Engineering

For most of the Web APIs, only a textual description is available

Most of the Web APIs do not consume or produce linked data (RDF)

RDF Input

```
<http://sws.geonames.org/5145067/>
    a gn:Feature ;
    wgs84:lat "40.78343" ;
    wgs84:long "-73.96625" .
```

Lowering

```
<http://sws.geonames.org/5145067/>
   gn:neighbour [
     a gn:Feature ; gn:name "Woodside";
      gn:nearby [
         a gn:Feature ; gn:name "...-Queens";
         gn:parentCountry [ ...
```

Linked RDF

Lifting

Neighbourhood / reverse geocoding

The neighbourhood for US cities. Data provided by Zillow under c

Webservice Type : REST
Url : api.geonames.org/neighbourhood?
Parameters : lat,lng
Result : returns the neighbourhood for the given latitude/longitude

Example http://api.geonames.org/neighbourhood?lat=40.78343&lng=-73.96625&username=demo

Invocation Example

```
<geonames>
 <neighbourhood>
    <countryCode>US</countryCode>
    <countryName>United States</countryName>
    <adminCode1>NY</adminCode1>
    <adminName1>New York</adminName1>
    <adminCode2>061</adminCode2>
    <adminName2>New York County</adminName2>
    <city>New York City-Manhattan</city>
    <name>Central Park</name>
  </neighbourhood>
</geonames>
```

XML Response

Outline

USC Viterbi School of Engineering

- Rapidly Integrating APIs with the LD
  - Building API Semantic Model
  - Representing API Descriptions
  - Building Linked APIs
- Evaluation
- Related Work
- Conclusion & Future Work

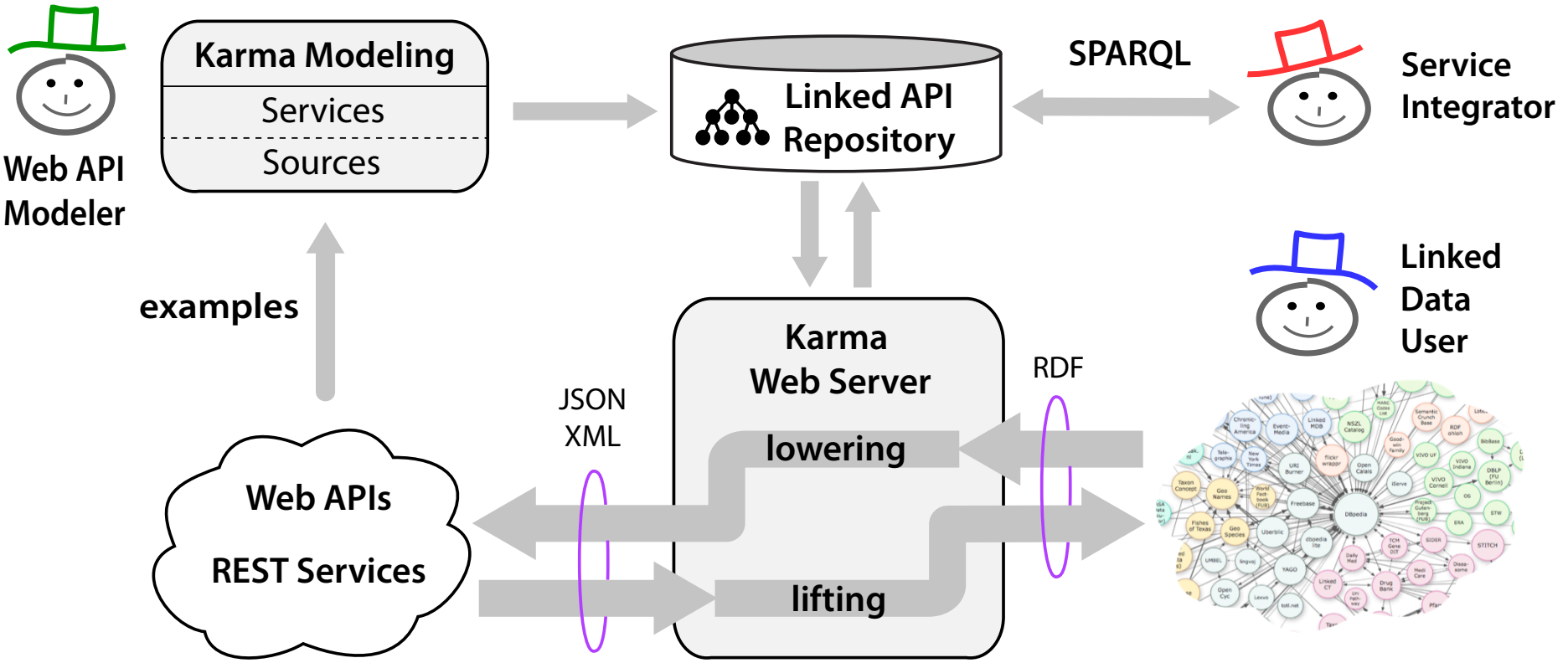
Outline

USC Viterbi School of Engineering

- Rapidly Integrating APIs with the LD
  - **Building API Semantic Model**
  - Representing API Descriptions
  - Building Linked APIs
- Evaluation
- Related Work
- Conclusion & Future Work

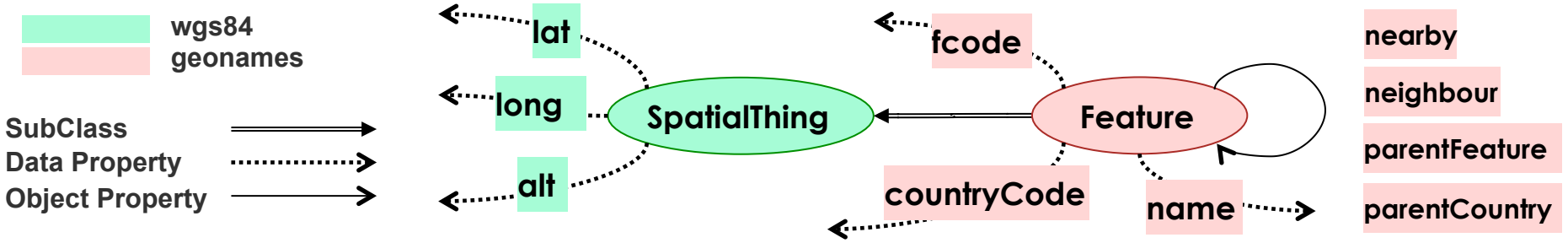
Overview

USC Viterbi School of Engineering

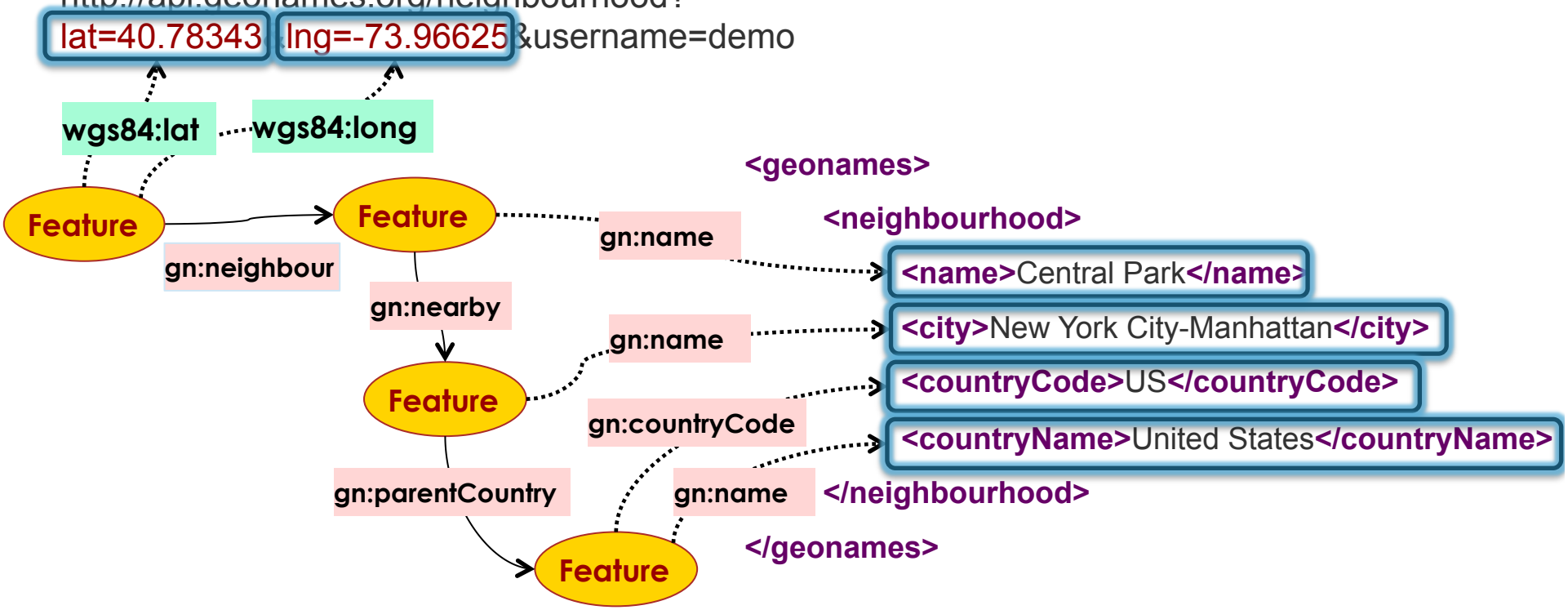


The Semantic Model

USC Viterbi School of Engineering



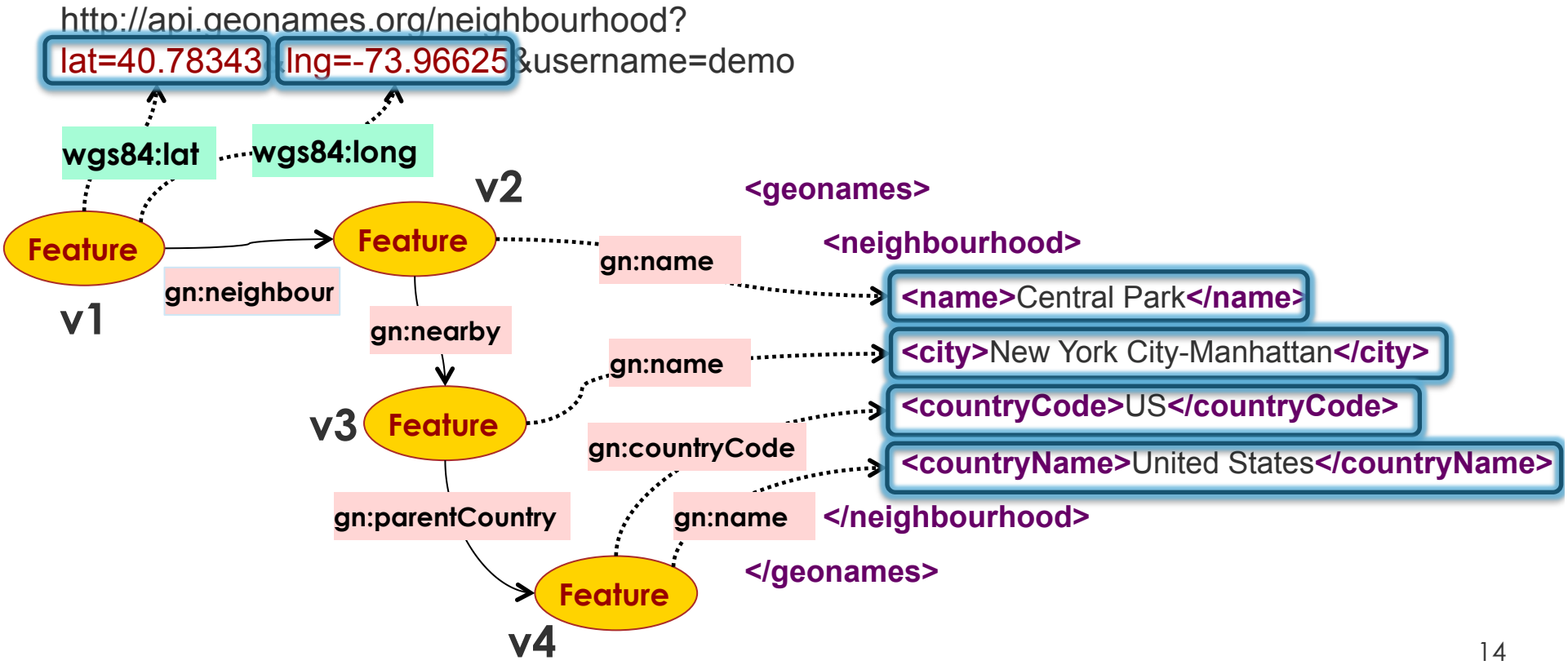
http://api.geonames.org/neighbourhood?
lat=40.78343 lng=-73.96625 &username=demo



The Semantic Model

USC Viterbi School of Engineering

$$\textbf{neighbourhood(\$lat, \$long, @countryCode, @countryName, @city, @name)} \rightarrow$$
$$\text{gn:Feature(v1)} \wedge \text{wgs84:lat(v1, \$lat)} \wedge \text{wgs84:long(v1, \$long)} \wedge \text{gn:}\textit{neighbourhood}\text{(v1, v2)} \wedge$$
$$\text{gn:Feature(v2)} \wedge \text{gn:name(v2, @name)} \wedge \text{gn:}\textit{nearby}\text{(v2, v3)} \wedge$$
$$\text{gn:Feature(v3)} \wedge \text{gn:name(v3, @city)} \wedge \text{gn:}\textit{parentCountry}\text{(v3, v4)} \wedge$$
$$\text{gn:Feature(v4)} \wedge \text{gn:countryCode(v4, @countryCode)} \wedge \text{gn:name(v4, @countryName)}$$

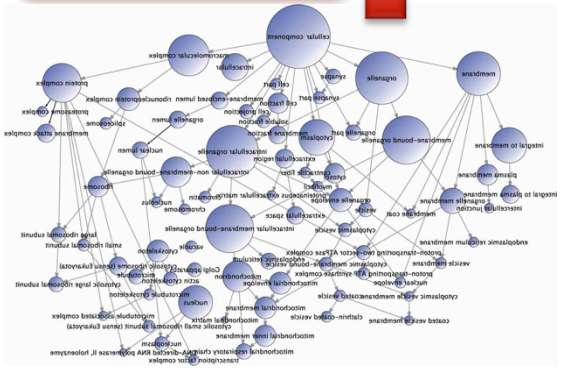


Semi-Automatically Modeling APIs

USC Viterbi School of Engineering

Examples of API Request URLs

Domain Ontology



Karma Modeling

- Invoking APIs
- Annotating Attributes
- Extracting Relationships

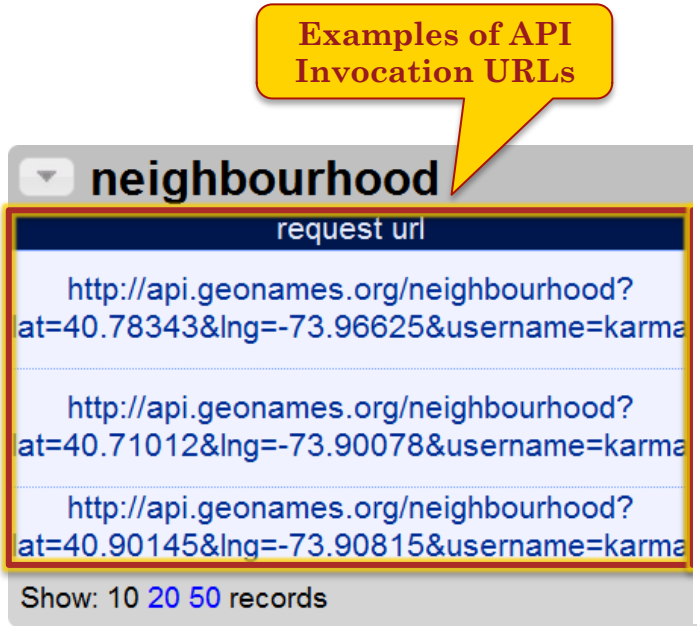
Previous Work on Modeling Static Sources

API Semantic Model

Invoking APIs

USC Viterbi School of Engineering

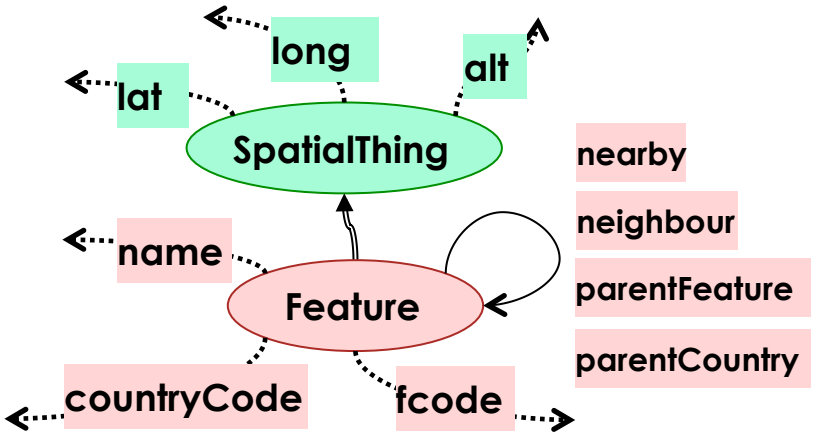
- User provides examples of the API request URLs
- Karma extracts the input values from the sample request URLs
- Karma invokes the API and extracts the output attributes from the API response (XML/JSON)



Annotating Inputs and Outputs
[Goel, Knoblock, Lerman, 2012]

USC Viterbi School of Engineering

- A CRF-based model to assign a **Semantic Type** *to* each column from its data
- Semantic Type
  - Ontology Class
  - Data Property + Domain



(wgs84:lat, gn:Feature)

lat	lng	username	countryCode	countryName	city	name
40.78343	-73.96625	karma	US	United States	New York City-Manhattan	Central Park
40.71012	-73.90078	karma	US	United States	New York City-Queens	Ridgewood

(wgs84:long, gn:Feature)

(gn:countryCode, gn:Feature)

(gn:name, gn:Feature)

SubClass
Data Property
Object Property

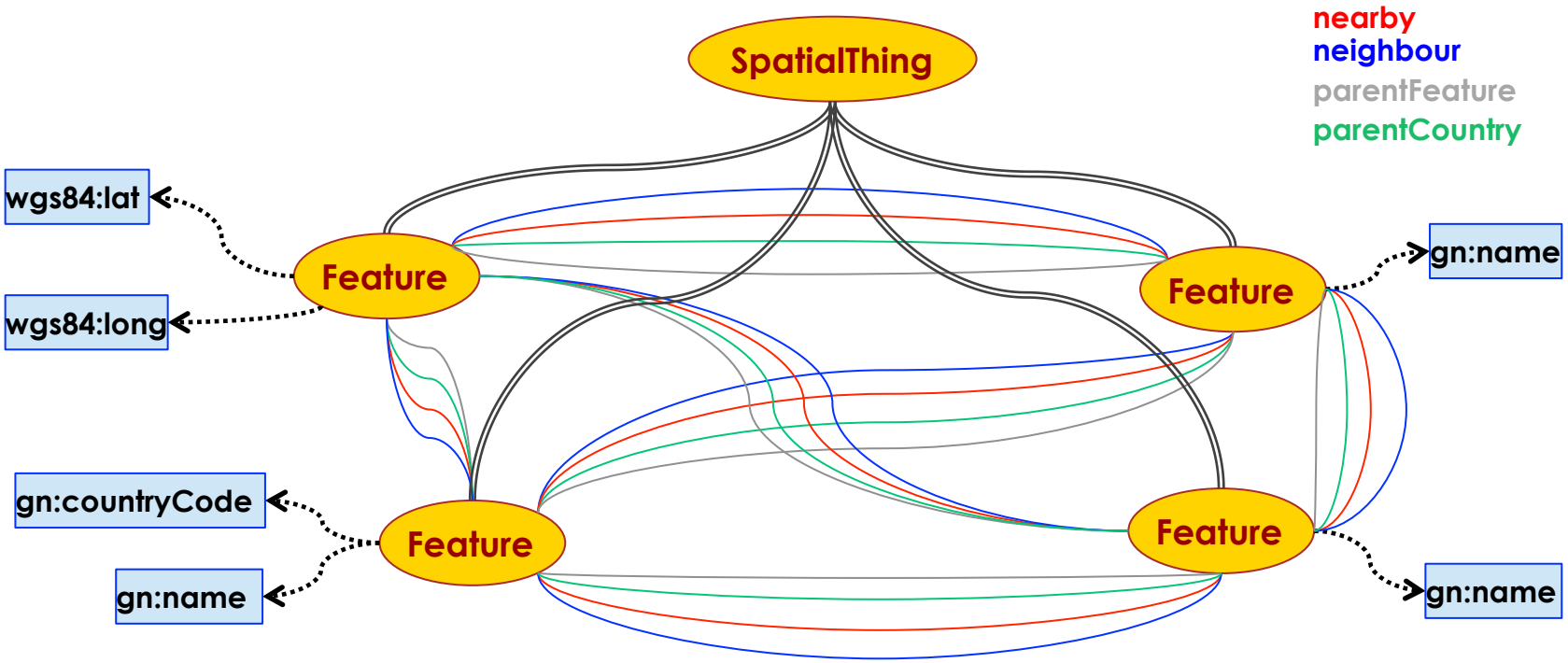
wgs84
geonames

Extracting Relationships

[Knoblock et al., ESWC 2012]

USC Viterbi School of Engineering

- Construct a graph from semantic types and ontology graph
- Select minimal tree that connects all semantic types
  - A customized **Steiner tree algorithm**

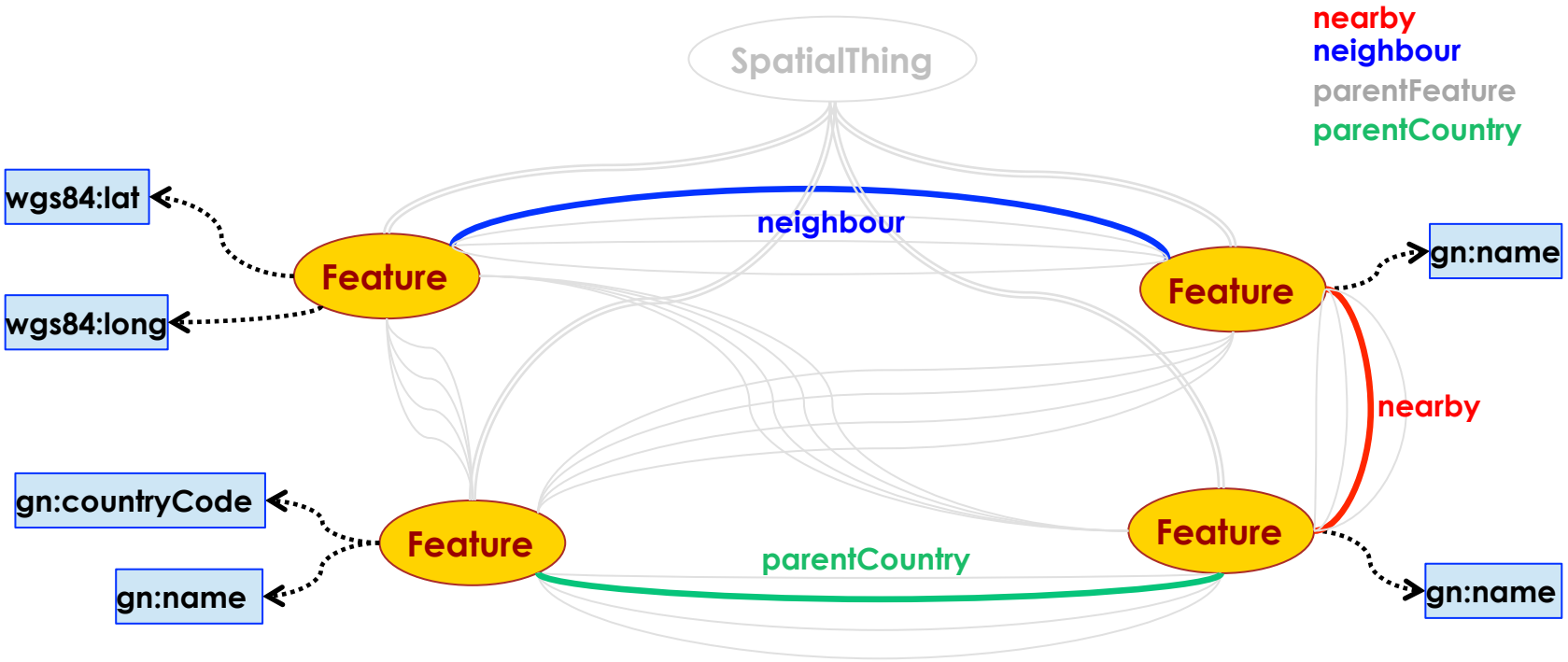


Extracting Relationships

[Knoblock et al., ESWC 2012]

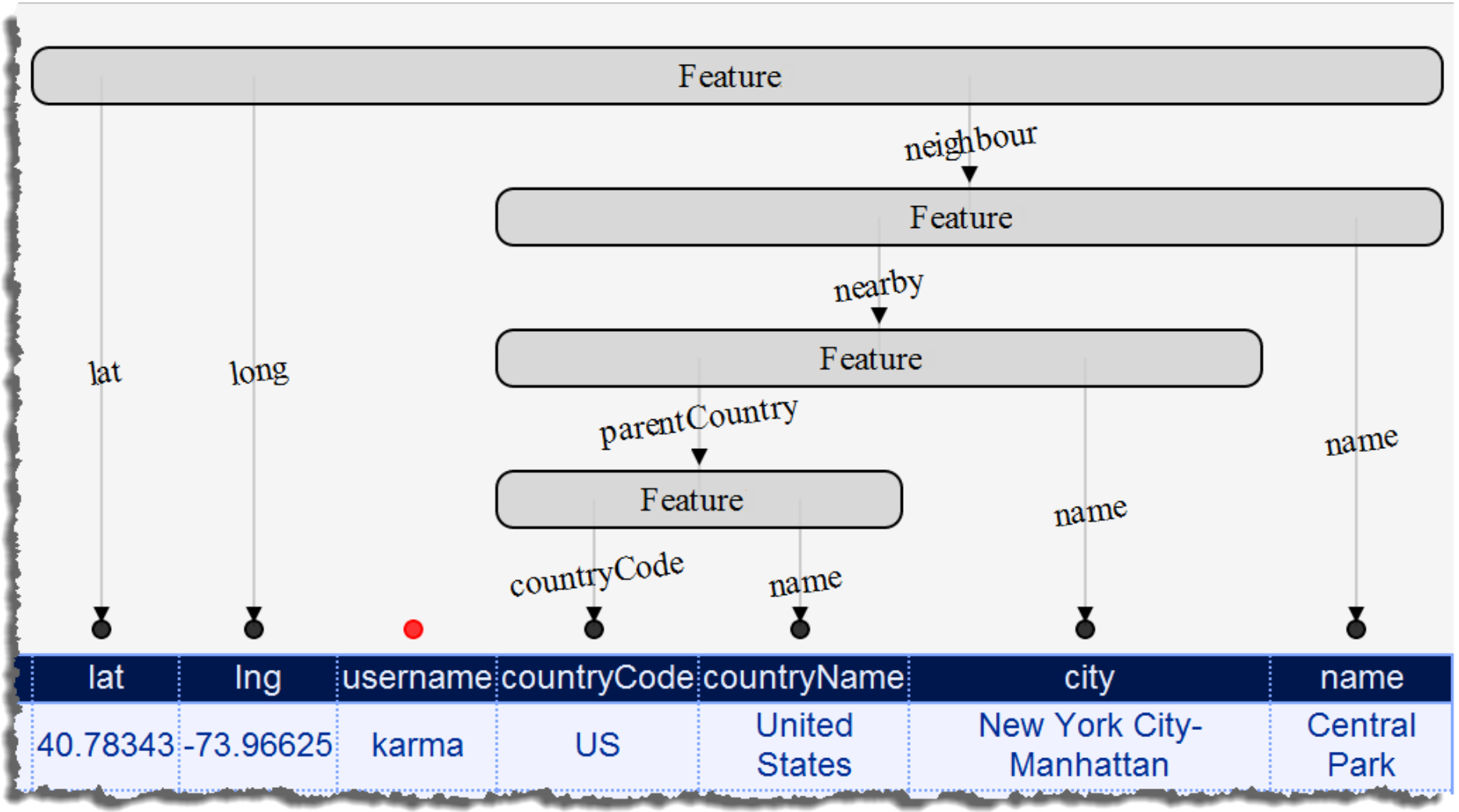
USC Viterbi School of Engineering

- Construct a graph from semantic types and ontology graph
- Select minimal tree that connects all semantic types
  - A customized **Steiner tree algorithm**



API Semantic Model in Karma

USC Viterbi School of Engineering



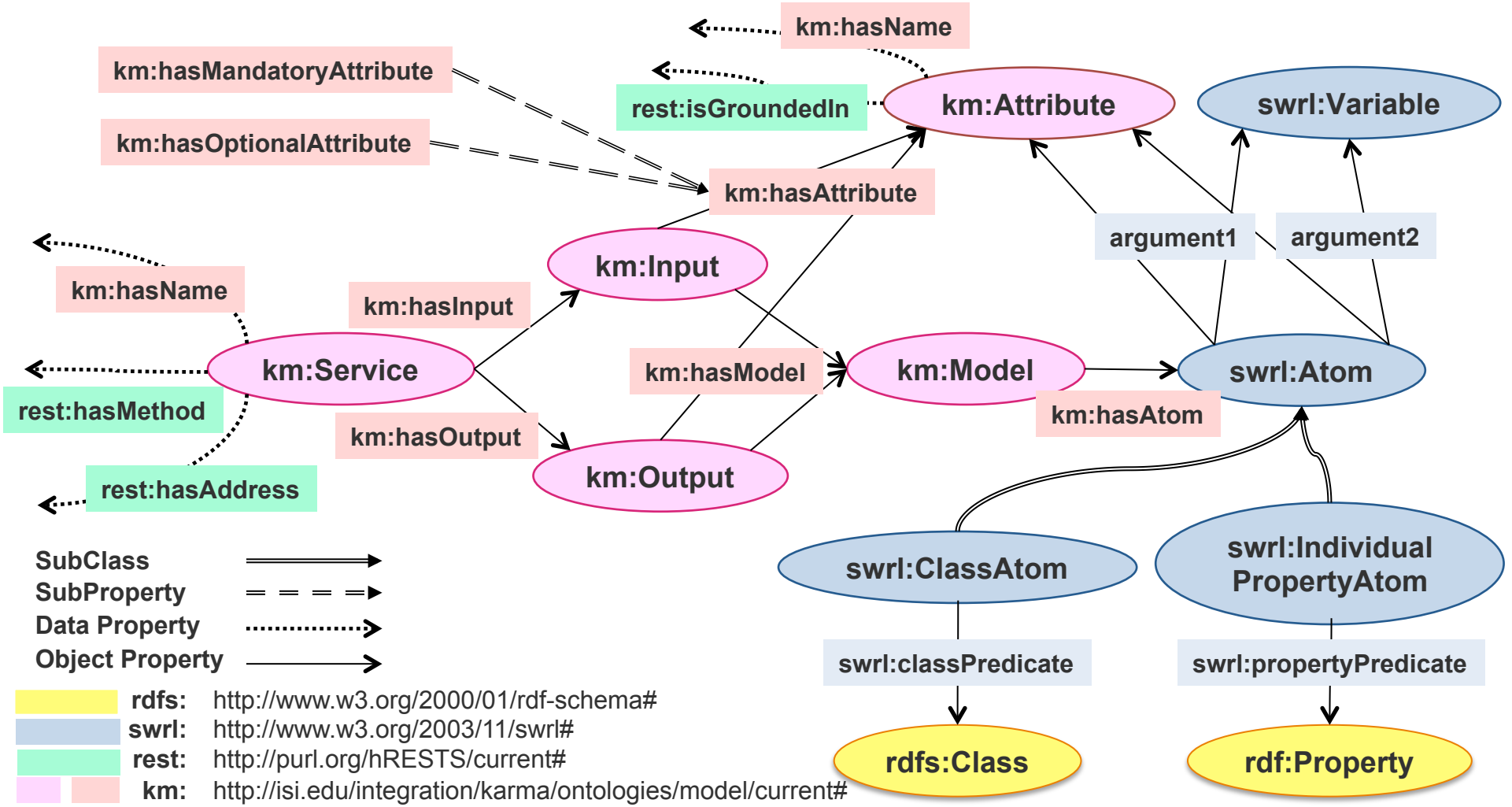
Outline

USC Viterbi School of Engineering

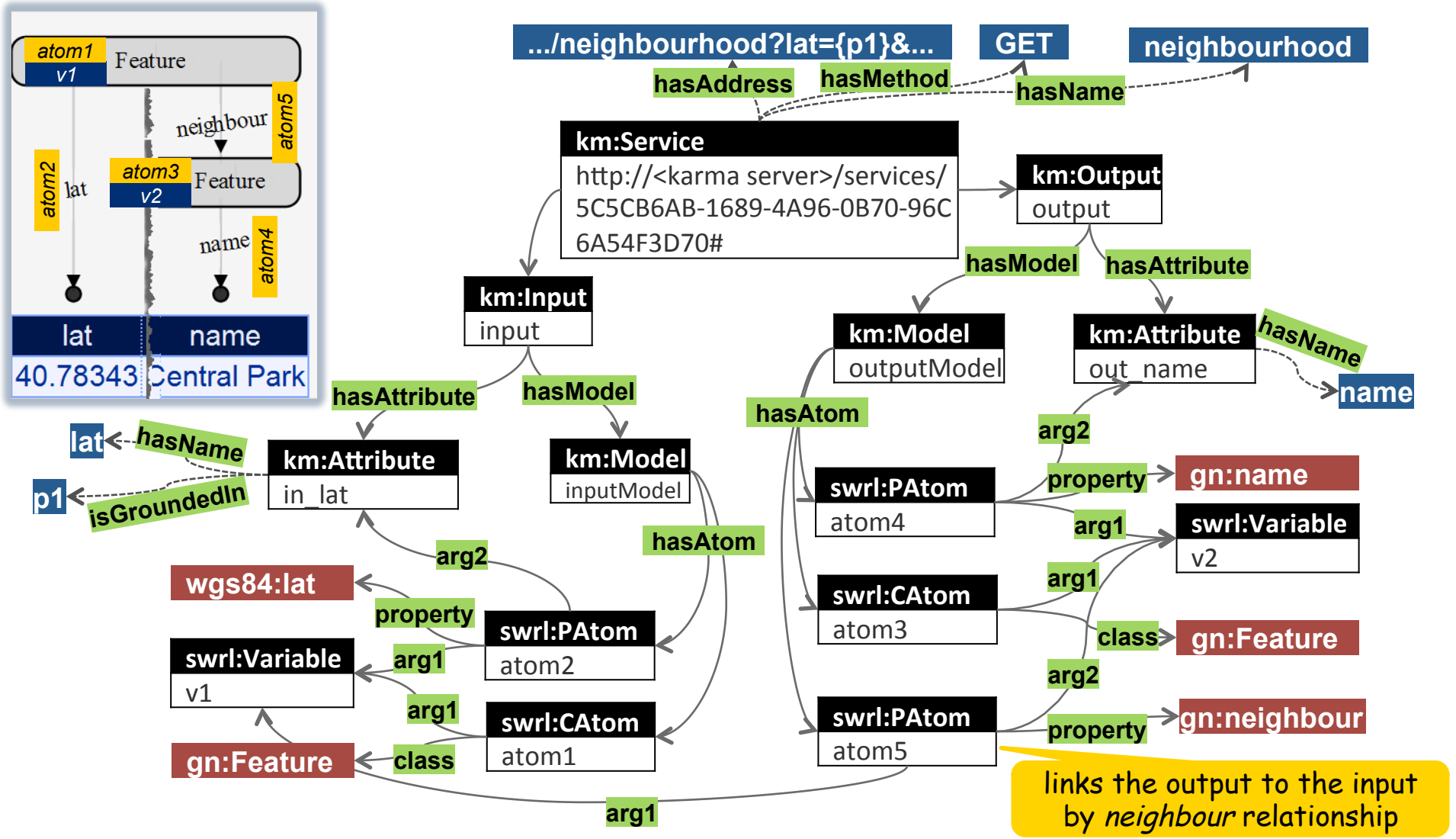
- Rapidly Integrating APIs with the LD
  - Building API Semantic Model
  - **Representing API Descriptions**
  - Building Linked APIs
- Evaluation
- Related Work
- Conclusion & Future Work

Service Modeling Ontology

USC Viterbi School of Engineering



Example: Service Description



Example: Service Description (N3)

USC Viterbi School of Engineering

```
@prefix : <http://<karma server>/services/5C5CB6AB-1689-4A96-0B70-96C6A54F3D70#> .
...
: a km:Service; km:hasName "neighbourhood" ;
        hrests:hasAddress "http://api.geonames.org/neighbourhood?lat={p1}&lng={p2}&
                           username={p3}" ^^ hrests:URITemplate ;
        hrests:hasMethod "GET"; km:hasInput :input; km:hasOutput :output.
```

```
:input a km:Input;
km:hasAttribute :in lat, ... ;
km:hasModel :inputModel.
:in lat a km:Attribute; km:hasName "lat" ;
          hrests:isGroundedIn "p1"^^rdf:PlainLiteral.
...
:feature1 a swrl:Variable .
:inputModel a km:Model;

km:hasAtom [ a swrl:ClassAtom ;
               swrl:classPredicate gn:Feature;
               swrl:argument1 :feature1 ];
km:hasAtom [ a swrl:IndividualPropertyAtom;
               swrl:propertyPredicate wgs84:lat;
               swrl:argument1 :feature1;
               swrl:argument2 :in lat];
```

```
:output a km:Output;
km:hasAttribute :out name, ... ; km:hasModel :outputModel.
:out name a km:Attribute; km:hasName "name" .
...
:feature2 a swrl:Variable . :outputModel a km:Model;
km:hasAtom [ a swrl:ClassAtom ;
               swrl:classPredicate gn:Feature;
               swrl:argument1 :feature2] ;
km:hasAtom [ a swrl:IndividualPropertyAtom ;
               swrl:propertyPredicate gn:neighbour;
               swrl:argument1 :feature1 ;
               swrl:argument2 :feature2];
km:hasAtom [ a swrl:IndividualPropertyAtom ;
               swrl:propertyPredicate gn:name ;
               swrl:argument1 :feature2 ;
               swrl:argument2 :out name];
```

Service Discovery

USC Viterbi School of Engineering

- Finds all services that take latitude and longitude as inputs

```
SELECT ?service WHERE {
    ?service km:hasInput [km:hasAttribute ?i1, ?i2].
    ?service km:hasInput [km:hasModel [km:hasAtom
        [swrl:propertyPredicate wgs84:lat; swrl:argument2 ?i1],
        [swrl:propertyPredicate wgs84:long; swrl:argument2 ?i2]]]
}
```

Service Discovery

USC Viterbi School of Engineering

- Find services that return the neighbor feature given the latitude and longitude

```
SELECT ?s WHERE {
    ?s km:hasInput [km:hasAttribute ?i1, ?i2].
    ?s km:hasOutput [km:hasAttribute ?o1].
    ?s km:hasInput [km:hasModel [km:hasAtom
        [swrl:classPredicate gn:Feature; swrl:arg1 ?f1],
        [swrl:propertyPredicate wgs84:lat; swrl:arg1 ?f1; swrl:arg2 ?i1],
        [swrl:propertyPredicate wgs84:long; swrl:arg1 ?f1; swrl:arg2 ?i2]]].
    ?s km:hasOutput [km:hasModel [km:hasAtom
        [swrl:classPredicate gn:Feature; swrl:arg1 ?f2],
        [swrl:propertyPredicate gn:neighbour; swrl:arg1 ?f1; swrl:arg2 ?f2]]]}
```

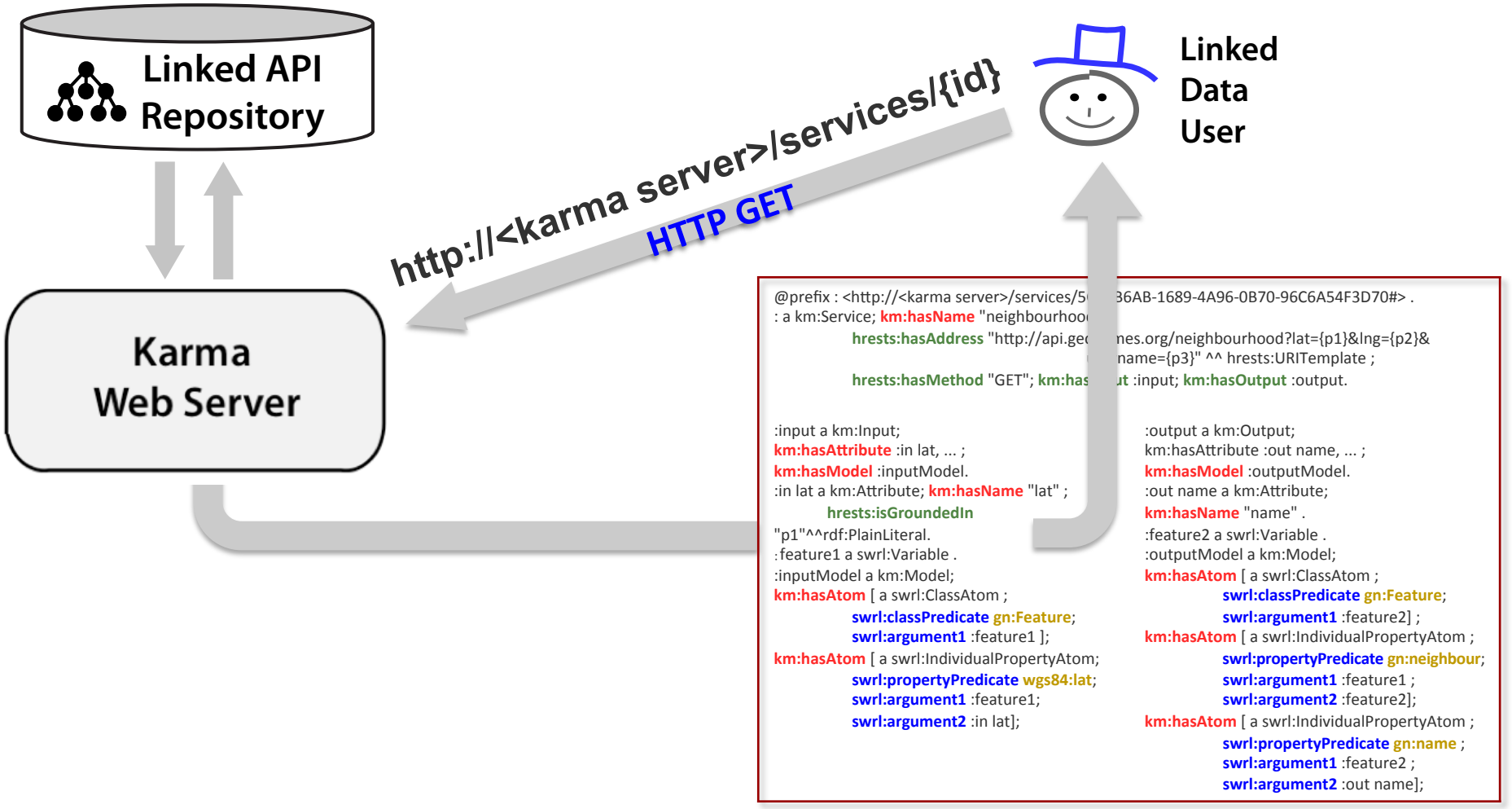
Outline

USC Viterbi School of Engineering

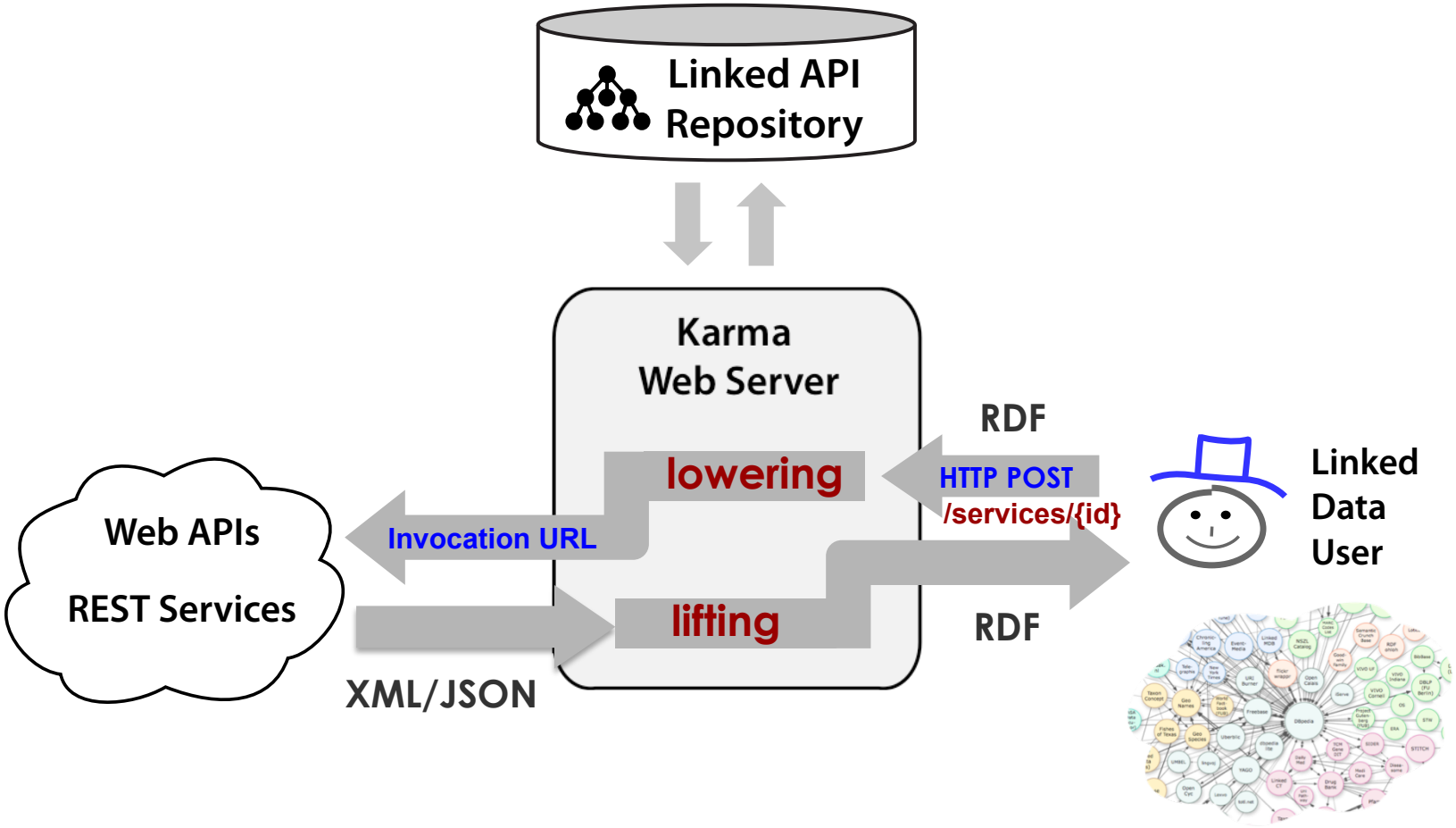
- Rapidly Integrating APIs with the LD
  - Building API Semantic Model
  - Representing API Descriptions
  - **Building Linked APIs**
- Evaluation
- Related Work
- Conclusion & Future Work

Invoking Linked APIs (GET)

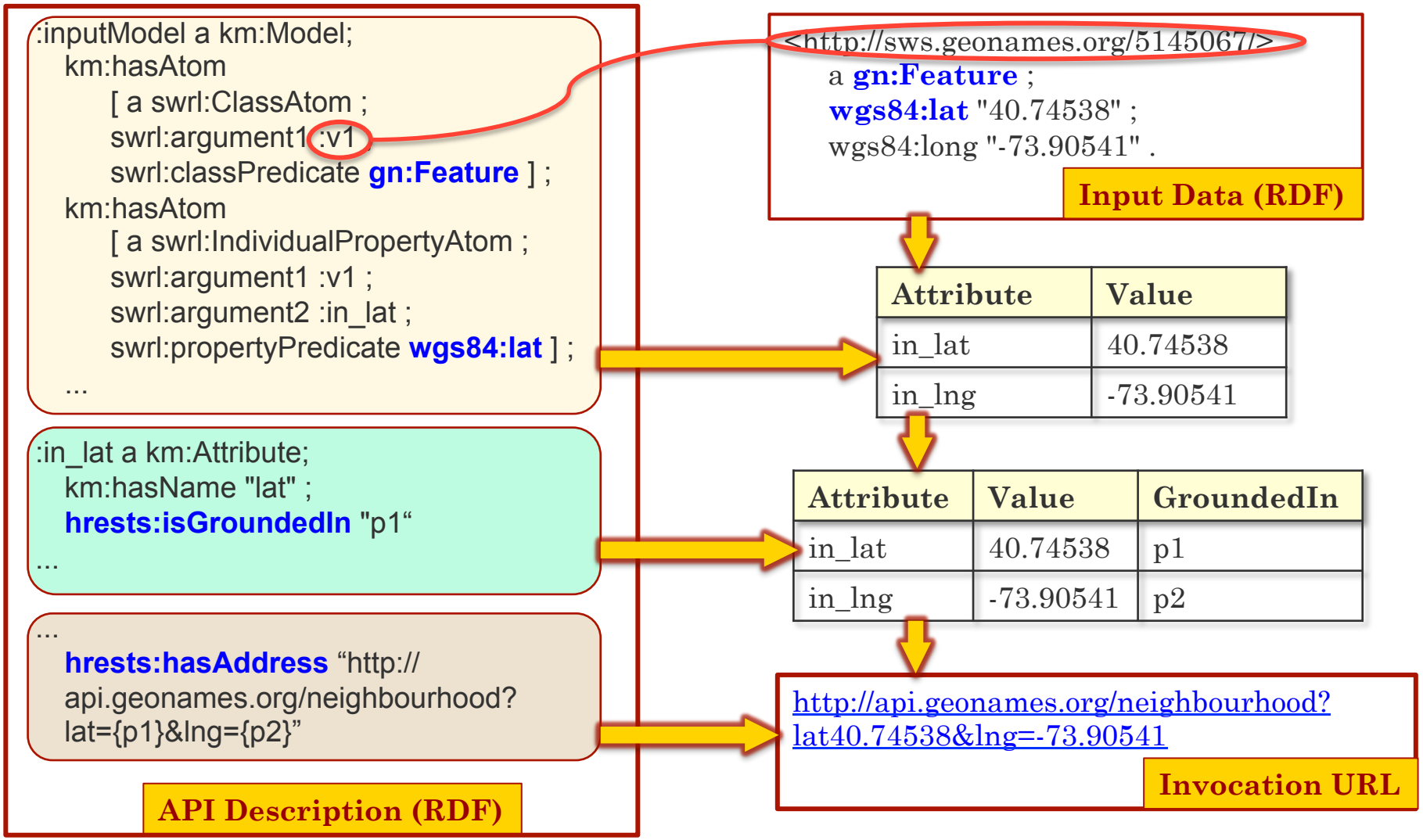
USC Viterbi School of Engineering



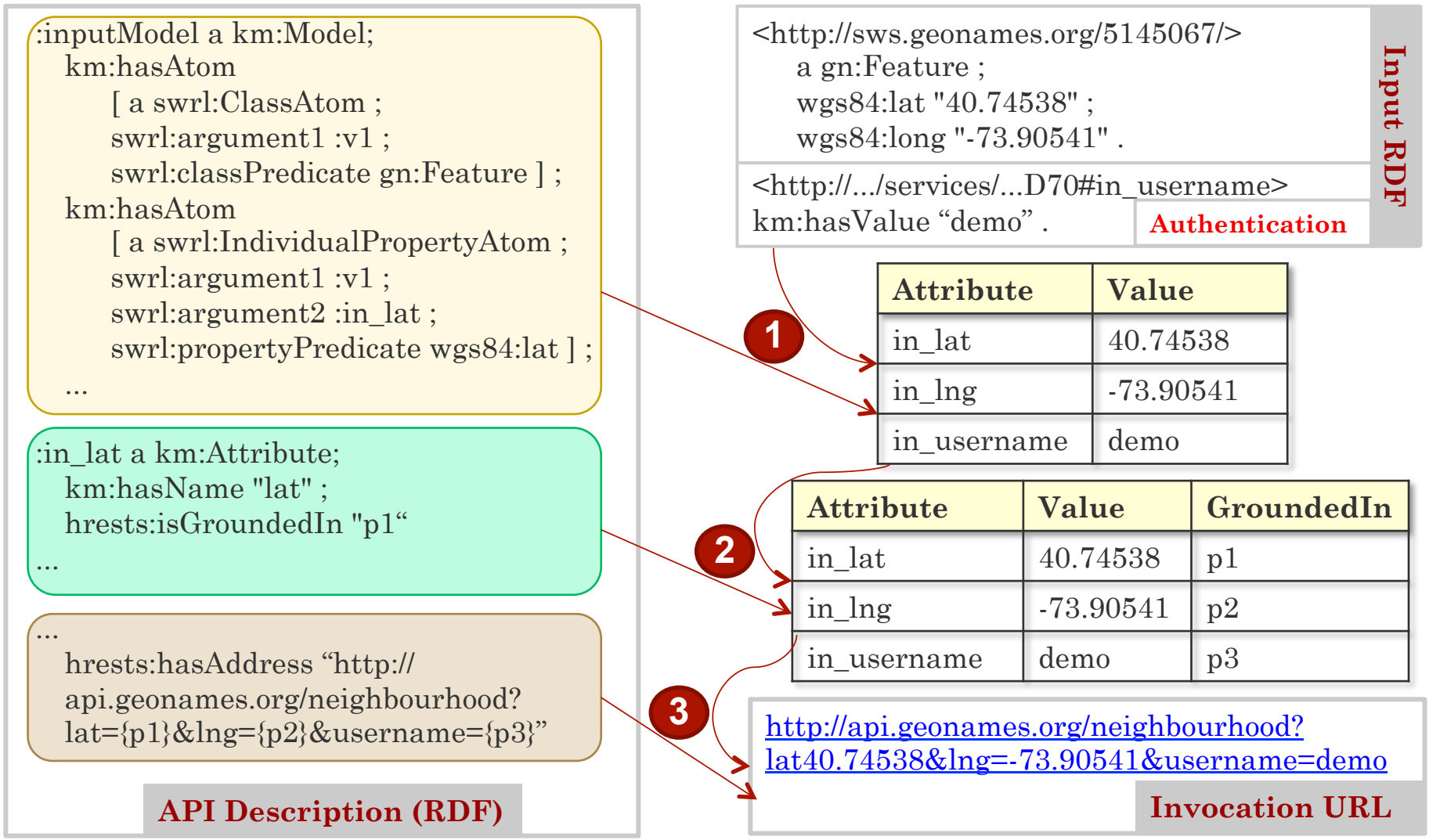
Invoking Linked APIs (POST)



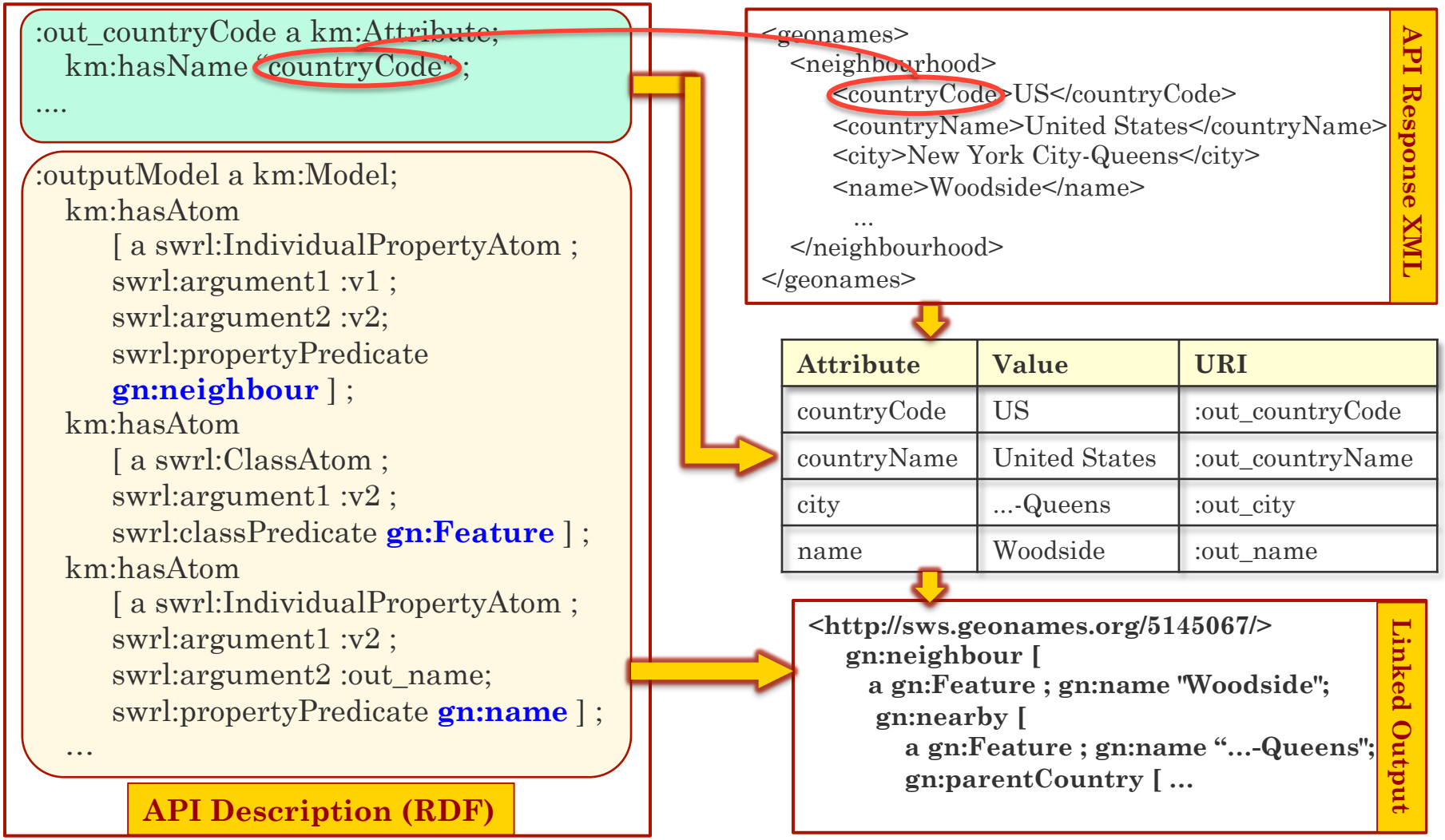
Lowering


:inputModel a km:Model;
km:hasAtom
[a swrl:ClassAtom ;
swrl:argument1 :v1
swrl:classPredicate gn:Feature] ;
km:hasAtom
[a swrl:IndividualPropertyAtom ;
swrl:argument1 :v1 ;
swrl:argument2 :in_lat ;
swrl:propertyPredicate wgs84:lat] ;
...
:in_lat a km:Attribute;
km:hasName "lat" ;
hrests:isGroundedIn "p1"
...
...
hrests:hasAddress "http://
api.geonames.org/neighbourhood?
lat={p1}&lng={p2}"
API Description (RDF)
<http://sws.geonames.org/5145067/>
a gn:Feature ;
wgs84:lat "40.74538" ;
wgs84:long "-73.90541" .
Input Data (RDF)
Attribute
Value
in_lat
40.74538
in_lng
-73.90541
Attribute
Value
GroundedIn
in_lat
40.74538
p1
in_lng
-73.90541
p2
http://api.geonames.org/neighbourhood?
lat40.74538&lng=-73.90541
Invocation URL

Lowering


:inputModel a km:Model;
km:hasAtom
[a swrl:ClassAtom ;
swrl:argument1 :v1 ;
swrl:classPredicate gn:Feature] ;
km:hasAtom
[a swrl:IndividualPropertyAtom ;
swrl:argument1 :v1 ;
swrl:argument2 :in_lat ;
swrl:propertyPredicate wgs84:lat] ;
...
:in_lat a km:Attribute;
km:hasName "lat" ;
hrests:isGroundedIn "p1"
...
...
hrests:hasAddress "http://
api.geonames.org/neighbourhood?
lat={p1}&lng={p2}&username={p3}"
API Description (RDF)
<http://sws.geonames.org/5145067/>
a gn:Feature ;
wgs84:lat "40.74538" ;
wgs84:long "-73.90541" .
<http://.../services/...D70#in_username>
km:hasValue "demo" .
Authentication
Input RDF
1
Attribute
Value
in_lat
40.74538
in_lng
-73.90541
in_username
demo
2
Attribute
Value
GroundedIn
in_lat
40.74538
p1
in_lng
-73.90541
p2
in_username
demo
p3
3
http://api.geonames.org/neighbourhood?
lat40.74538&lng=-73.90541&username=demo
Invocation URL

Lifting


:out_countryCode a km:Attribute;
km:hasName "countryCode";
....
:outputModel a km:Model;
km:hasAtom
[a swrl:IndividualPropertyAtom ;
swrl:argument1 :v1 ;
swrl:argument2 :v2;
swrl:propertyPredicate
gn:neighbour] ;
km:hasAtom
[a swrl:ClassAtom ;
swrl:argument1 :v2 ;
swrl:classPredicate gn:Feature] ;
km:hasAtom
[a swrl:IndividualPropertyAtom ;
swrl:argument1 :v2 ;
swrl:argument2 :out_name;
swrl:propertyPredicate gn:name] ;
…
API Description (RDF)
<geonames>
<neighbourhood>
<countryCode>US</countryCode>
<countryName>United States</countryName>
<city>New York City-Queens</city>
<name>Woodside</name>
...
</neighbourhood>
</geonames>
API Response XML
Attribute
Value
URI
countryCode
US
:out_countryCode
countryName
United States
:out_countryName
city
...-Queens
:out_city
name
Woodside
:out_name
<http://sws.geonames.org/5145067/>
gn:neighbour [
a gn:Feature ; gn:name "Woodside";
gn:nearby [
a gn:Feature ; gn:name "...-Queens";
gn:parentCountry [...
Linked Output

Outline

USC Viterbi School of Engineering

- Rapidly Integrating APIs with the LD
  - Building API Semantic Model
  - Representing API Descriptions
  - Building Linked APIs
- Evaluation
- Related Work
- Conclusion & Future Work

Evaluation

USC Viterbi School of Engineering

Geonames API	#Examples	#Columns	Time (min)
neighbourhood	3	10	6
neighbours	2	9	5
children	2	10	3
sibling	1	9	3
ocean	2	3	1
findNearby	3	11	3
findNearbyPostalCodes	3	11	7
findNearbyPOIsOSM	3	7	3
findNearestAddress	3	14	6
findNearestIntersectionOSM	3	8	3
postalCodeCountryInfo	1	5	2
Total	**26**	**97**	**42**

Average 4 minutes to build a linked API

Related Work

USC Viterbi School of Engineering

- Linked Services [Pedrinaci & Domingue, 2010]
  - Annotates inputs and outputs by concepts from ontologies
  - Publishes service descriptions into the LOD cloud using Minimal Service Model (MSM)
  - Cannot represent relationships between service attributes

- Linked Open Services (LOS) [Krummenacher & Norton & Marte, 2010] , Linked Data Services (LIDS) [Speiser & Harth, 2010]
  - SPARQL graph patterns to describe inputs and outputs
  - Service discovery is not straightforward

- RESTdesc [Verborgh et at, 2012]
  - N3 logical rules to capture API functionality

Related Work

USC Viterbi School of Engineering

- Linked Services [Pedrinaci & Domingue, 2010]
  - Annotates inputs and outputs by concepts from ontologies
  - Publishes service descriptions into the LOD cloud using
  - SPARQL graph patterns to describe inputs and outputs
  - Service discovery is not straightforward

Karma semi-automatically builds service descriptions and the modeling process does not require expertise in SW technologies

- RESTdesc [Verborgh et at, 2012]
  - N3 logical rules to capture API functionality

Discussion

USC Viterbi School of Engineering

- Rapidly build rich semantic models of services
- Publish service descriptions into the LD cloud
- Provide strong support for service discovery and composition
- Build linked APIs that consume and produce linked data

Future Work

USC Viterbi School of Engineering

- Apply our approach on REST-like URLs
  - http://www.ex.com/weather/CA/Los Angeles
- Compose data and services in Karma

S1(address → street, city, state, zipcode,)

S2(city, state→ temperature, windspeed, …)

S3(zipcode, distance→ hotel, ranking)

Composition

events

name	address
ACM GIS 2012	300 N. Harbor Drive Redondo Beach, CA 90277
ISWC 2012	50 Park Plaza at Arlington Street, Boston, MA 02116

Show: 10 20 50 records Previous Next

More Information

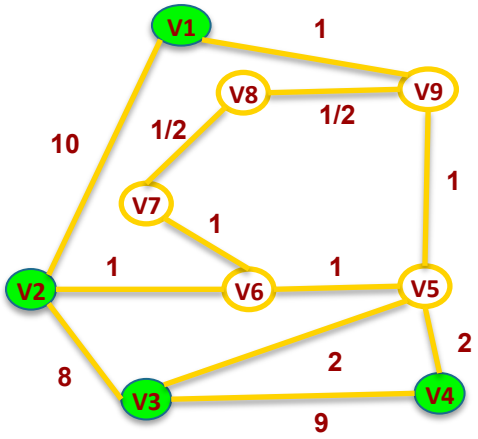
USC Viterbi School of Engineering

- More information/papers/software/demos:
  - http://www.isi.edu/integration/karma/
- Contact
  - Mohsen Taheriyan: mohsen@isi.edu
  - Craig Knoblock: knoblock@isi.edu
  - Pedro Szekely: pszekely@isi.edu
  - Jose Luis Ambite: ambite@isi.edu

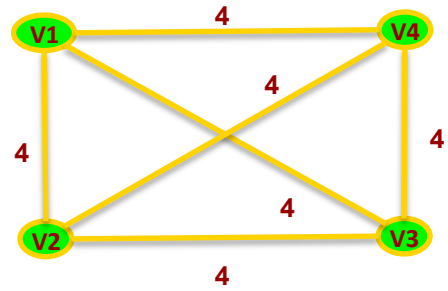
Backup Slides

Steiner Tree Algorithm

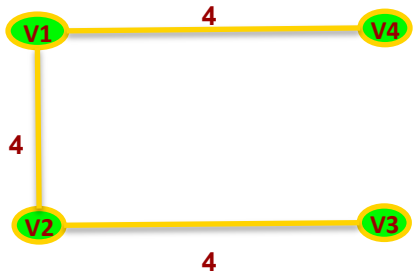
USC Viterbi School of Engineering



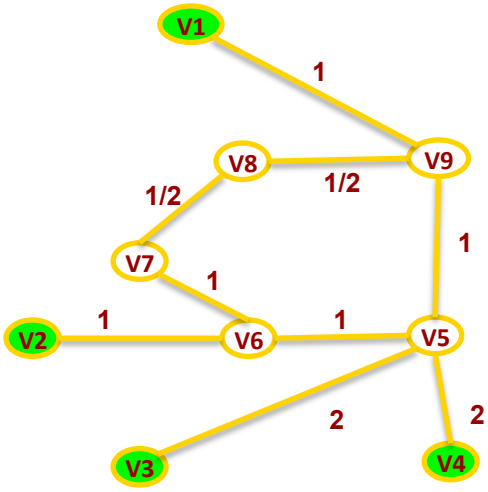
Steiner nodes: {V1, V2, V3, V4}



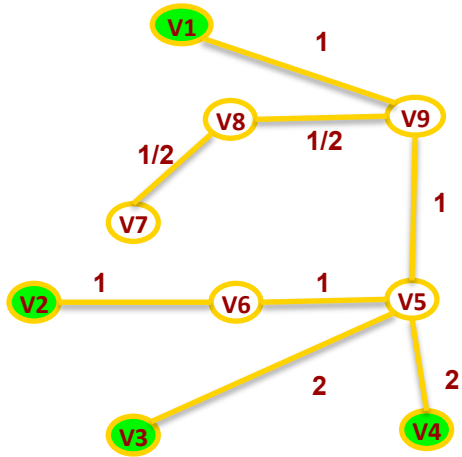
1. construct the complete graph (Nodes: Steiner Nodes, Links Weights: shortest path from each pair in original G)



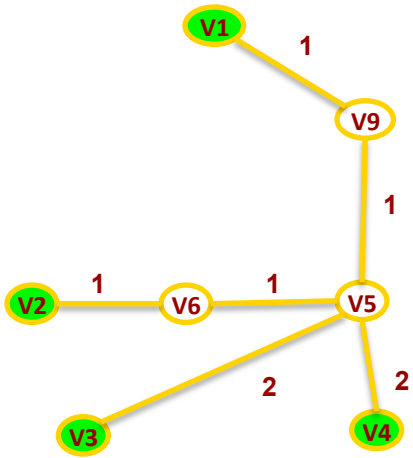
2. Compute MST



3. replace each link with the corresponding shortest path in original G



4. Compute MST



5. remove extra links until all leaves are Steiner nodes