

Harvesting Geospatial Knowledge from Social Metadata

Suradej Intagorn
Anon Plangprasopchok
Kristina Lerman

Motivation



Tags:

station

fire

los

angeles

california

satellite

image

digitalglobe

Los Angeles

USA

Problem

- How to extract **geospatial knowledge** from user contributed annotations (metadata)?

Advantages

- Cheaper and Up-to-date than knowledge produced by a group of experts
- Supplement to exist knowledge created by experts



Combine geospatial annotations from many users to learn more complete representation of places & their relations

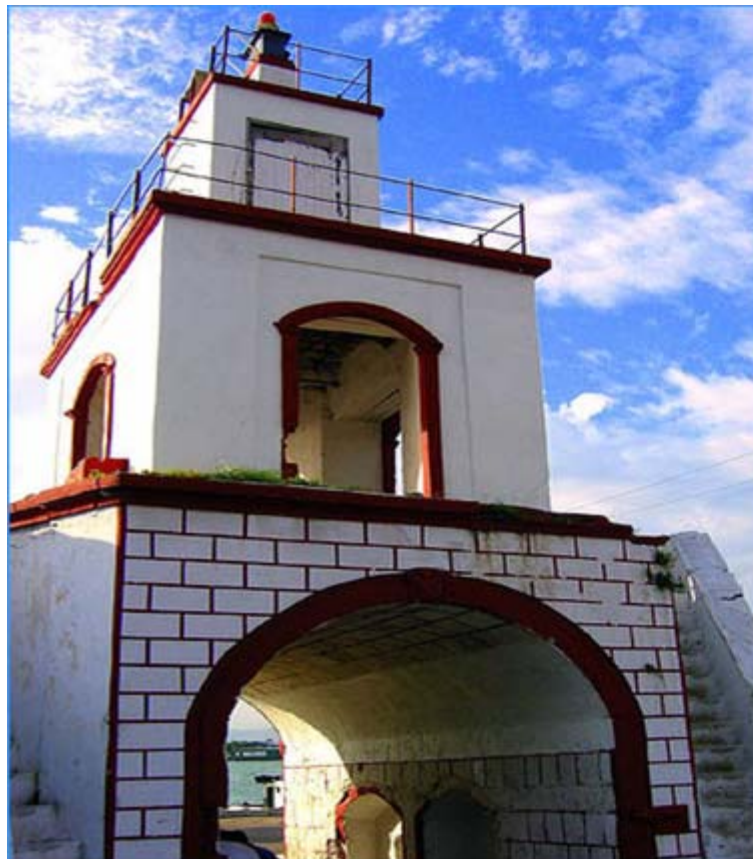
- tags,
- geo-tags
- set names (to be explained)

Challenges:

1. Place names are often ambiguous
2. Noisy data points
3. Sparseness



Types of Social Meta Data



The old parola of Opon probably dates back to the 50s. It is no longer a common sight, but it is a reminder of the past. I remember playing in this area...

Tags

- Opon
- Mactan
- Lapu-Lapu
- Church
- Cebu
- Philippines
- archway
- jetty
- port
- dock
- red
- white
- colors
- parola
- lighthouse
- faro
- Cebu-Sugbo

Additional Information

- All rights reserved
- Anyone can see this photo
- Taken in Lapu-Lapu City** (map)
- Taken with a Canon
- More properties
- Taken on August 4, 2005
- 3 people call this photo a faro
- Viewed 767 times


+ Farl's photostream


+ **Cebu (Set)**

This photo also belongs to:


- + aPeeling Paint (Pool)
- + Pinoys Galore (Pool)
- + Philippines (Pool)
- + LIGHTHOUSE LOVERS (Pool)
- + Southeast Asia Images (Pool)
- + Philippines Images (Pool)
- + Historic Preservation (Pool)
- + Catchy Colors (Pool)
- + "Only in the Philippines!" (Pool)
- + visit the world - the travel guide (Pool)
- + Intense and Vibrant colors (Pool)
- + Travel Photography (Pool)

Types of Social Meta Data

 **Cebu**
Thumbnails | Detail | Map | 19 comments



my home
210 photos | 16,409 views
Items are from between 14 May 2001 & 26 Sep 2005.



Types of Social meta data



Summary of Social Metadata

Photo

Photo ID

User ID

Set ID

Set Name

Latitude

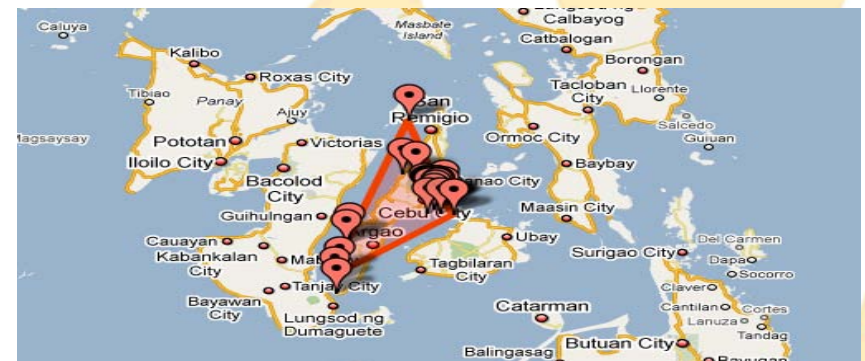
Longitude



- Each photo is just a point in spatial domain.

- Each set is a collection of points representing a place

- They can be aggregated to get better representation of the place.



Recognizing Places



Disambiguating
Places



Noise Filtering



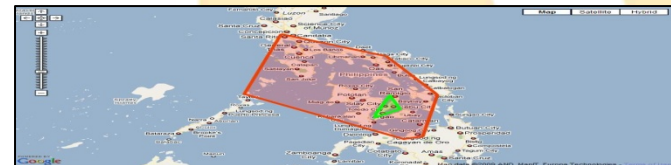
Representing
Places



Geographic
subsumption



GeoNames



Recognizing Places



Disambiguating Places



Noise Filtering



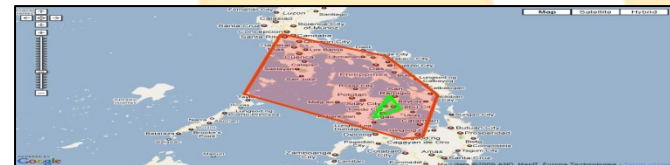
Representing Places



Geographic subsumption



GeoNames



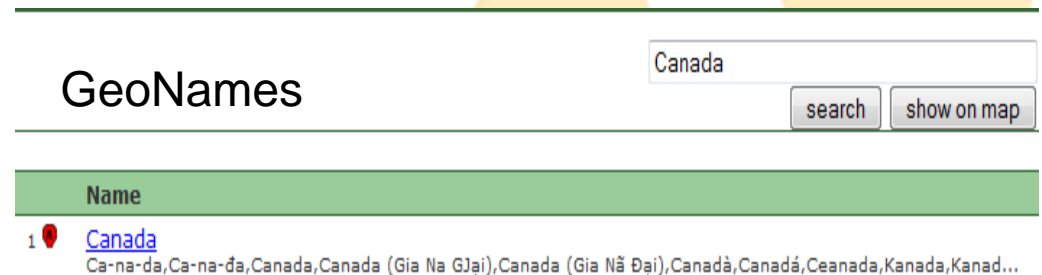
Recognizing Place-related Set Names

Assumption

- Set names refer to places
- But, need to filter out non-place set name

Current approach

- Use reference set from GeoNames.org
- Simple name matching with set name



Recognizing Places



Disambiguating Places



Noise Filtering



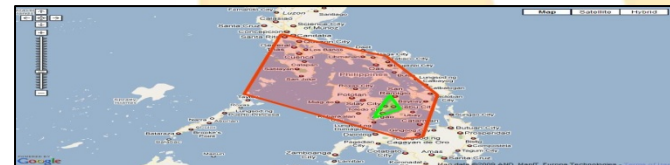
Representing Places



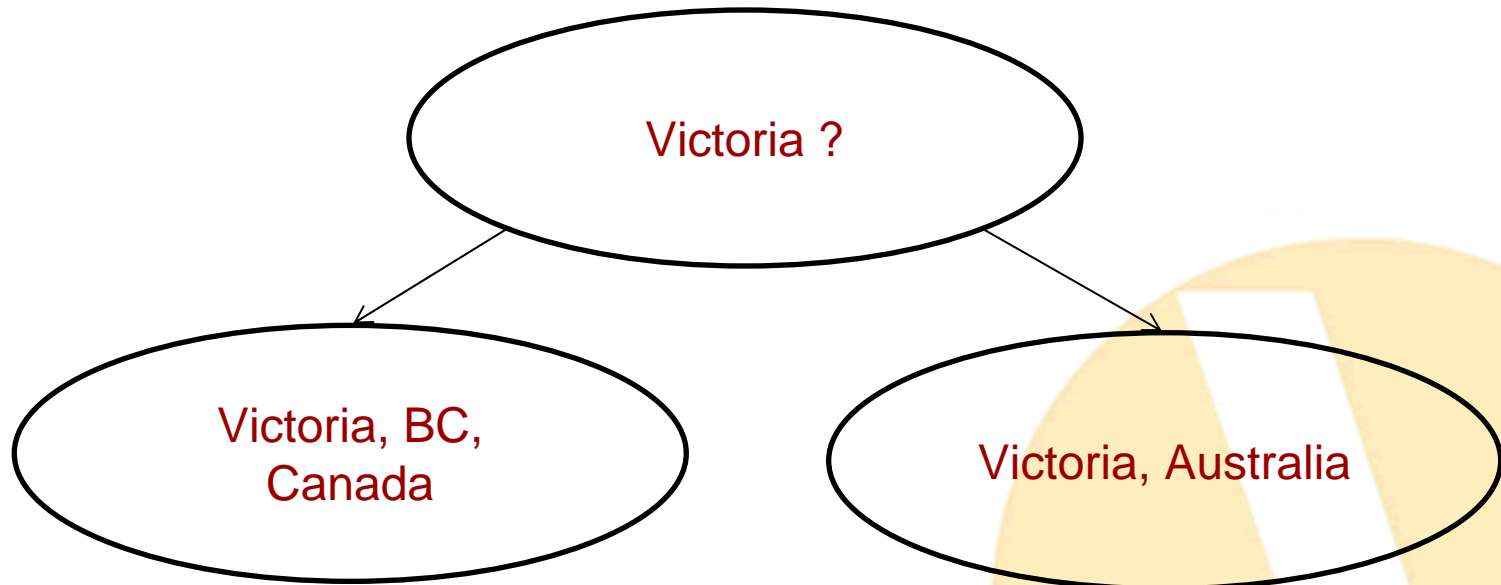
Geographic subsumption



GeoNames

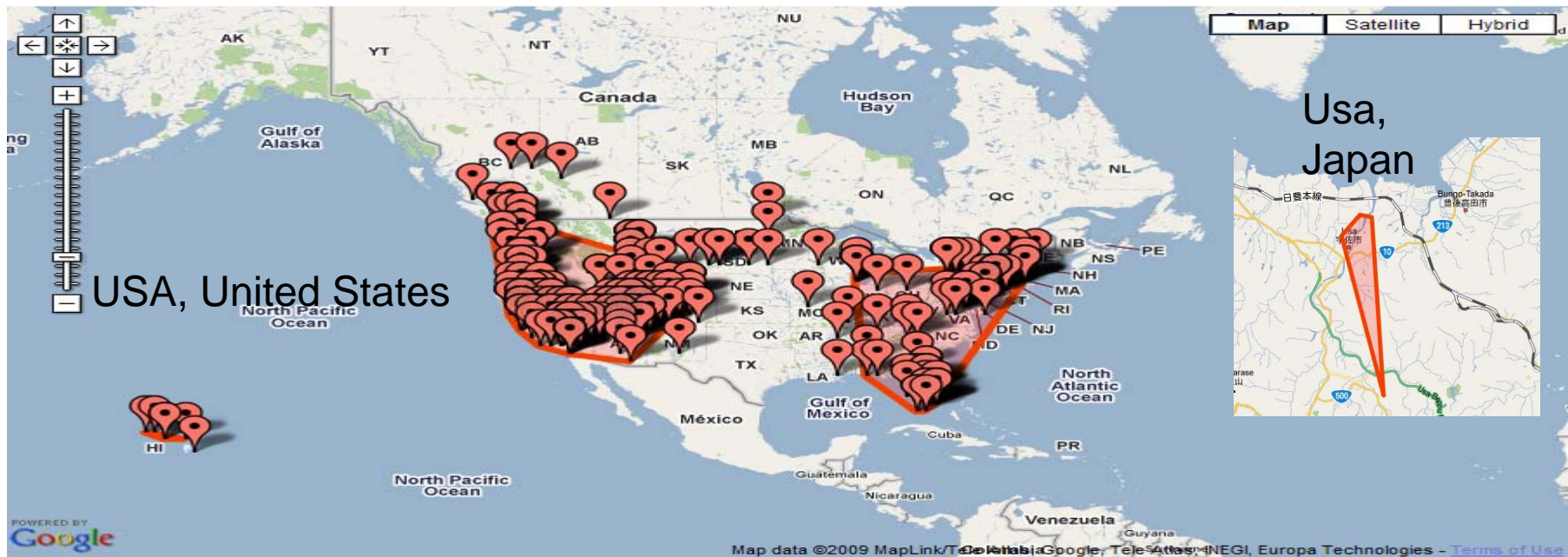


Place names are often ambiguous



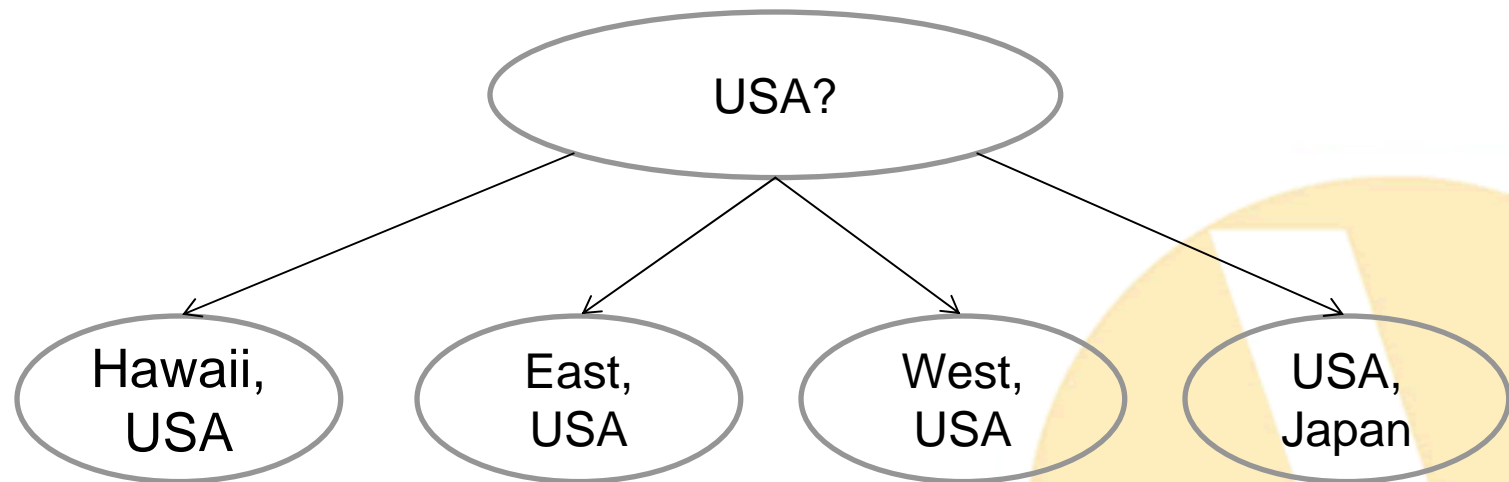
Disambiguating Places

Problem: Place may be also non-contiguous



Disambiguating Places

Two “USA” places: (1) United States; (2) Usa, Japan



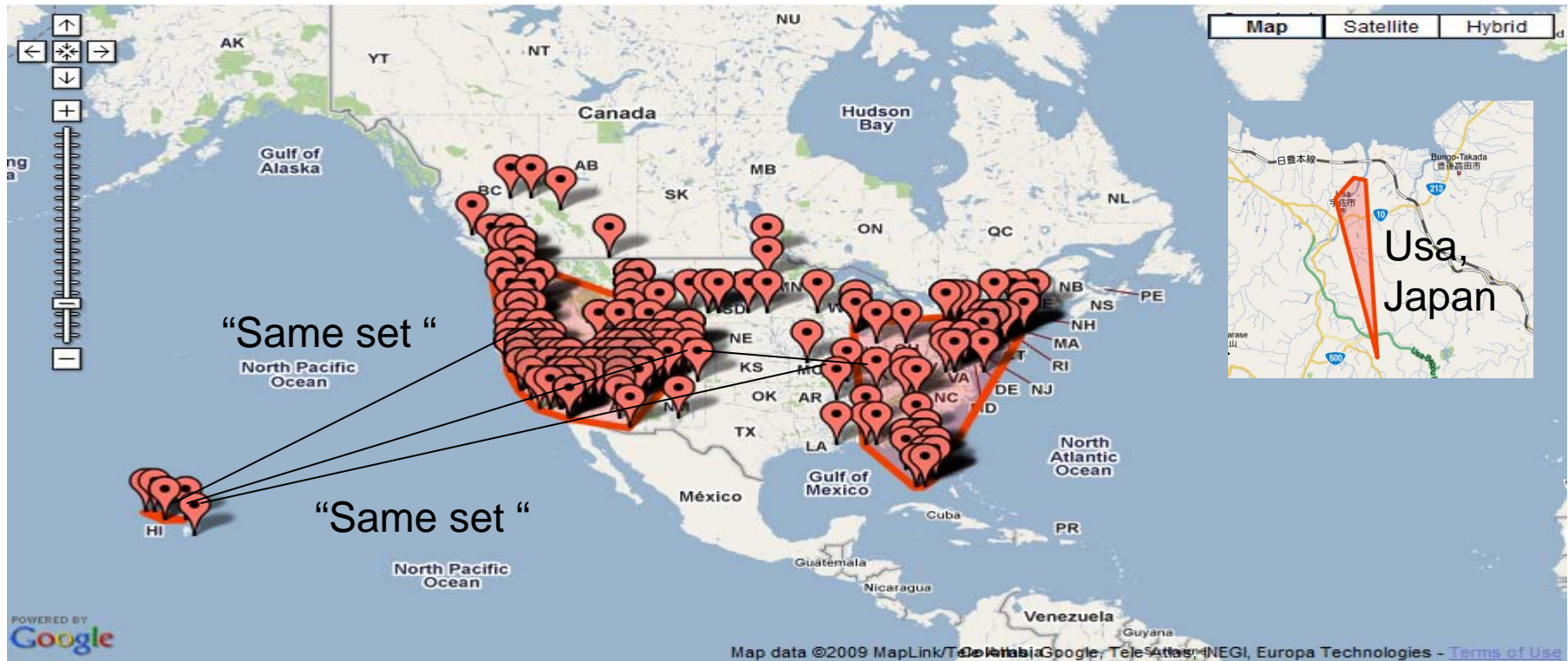
Assumption: ambiguous places can be disambiguated by their locations

How?

1. Construct graph of geo points
2. Then, finding connected components



Learning Non-Contiguous Places



Construct Graph $G1 = (V, E)$,

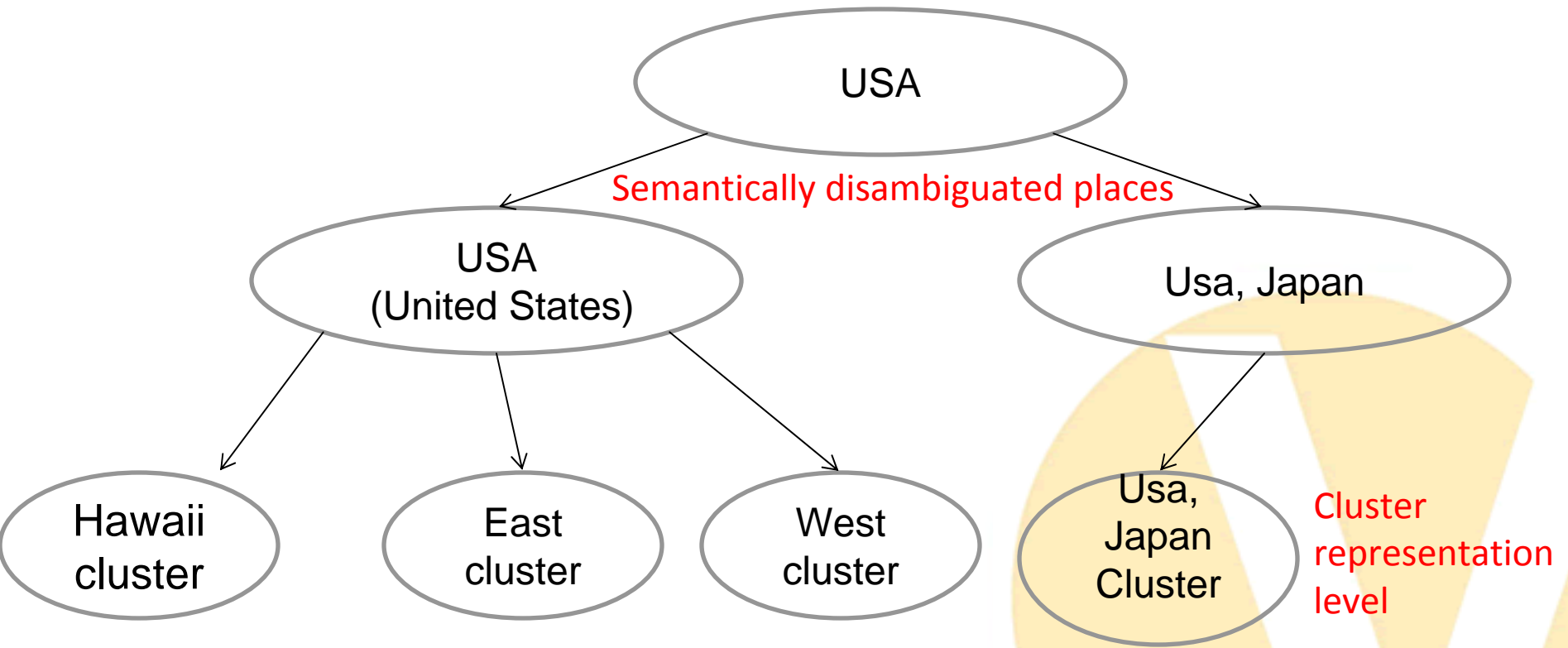
Vertices, V , are points corresponding to geo-tagged photos

Edges, E , are the edges between vertices

When is the edge between two photos created?

$$\text{dist}(v_i, v_j) < \tau \text{ or } s_{v_i} = s_{v_j}.$$

Multi Cluster Representation



Recognizing Places



Disambiguating
Places



Noise Filtering



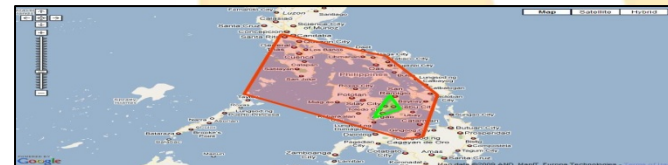
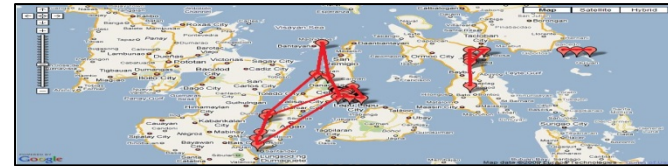
Representing
Places



Geographic
subsumption



GeoNames



Noise Filtering



Heuristic: Cluster that contains photos from $< k$ users is noise

Recognizing Places



Disambiguating
Places



Noise Filtering



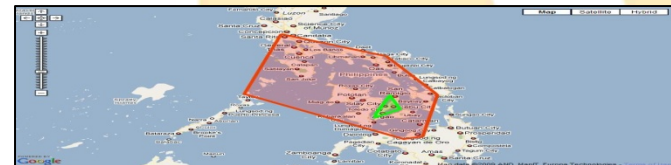
Representing
Places



Geographic
subsumption



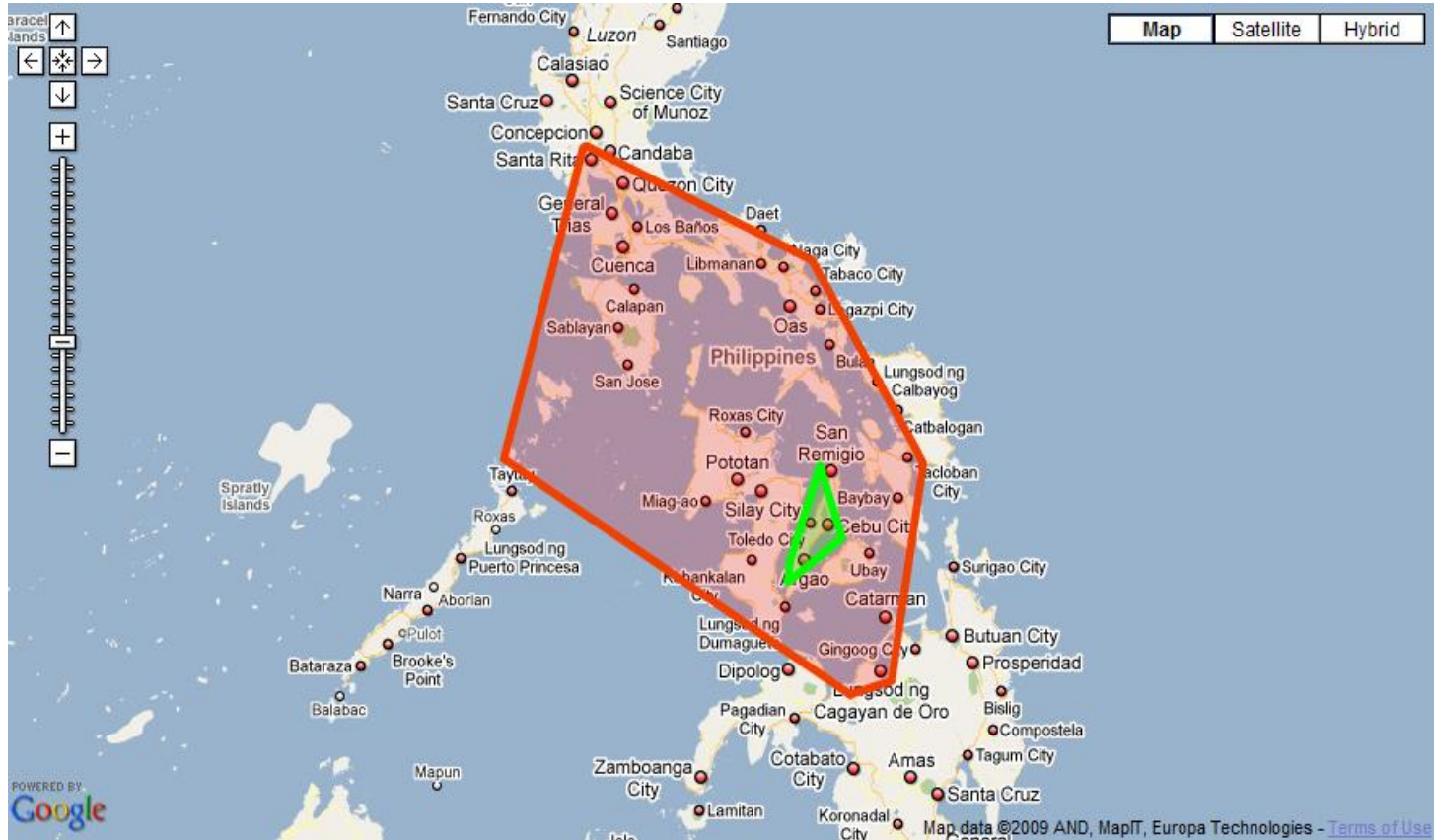
GeoNames



Multiple Convex Hulls



Learning relations between places



Based on probabilistic subsumption
(Sanderson and Croft, 1999; Schmitz, 2006)

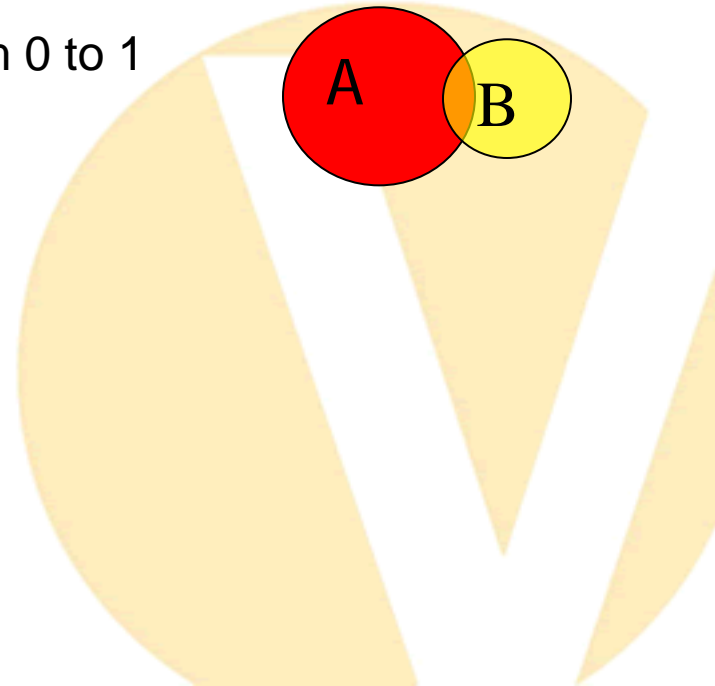
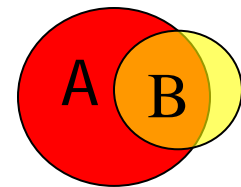
place A subsumes place B if

$$p(A|B) \geq t \text{ and } p(B|A) < t$$

, where t is a predefined threshold ranging from 0 to 1

$$p(A|B) = \text{Area}(A \cap B) / \text{Area}(B)$$

$$p(B|A) = \text{Area}(A \cap B) / \text{Area}(A)$$



Probabilistic Subsumption on Tags

Probabilistic subsumption on tags (Schmitz, 2006)

place A subsumes place B if

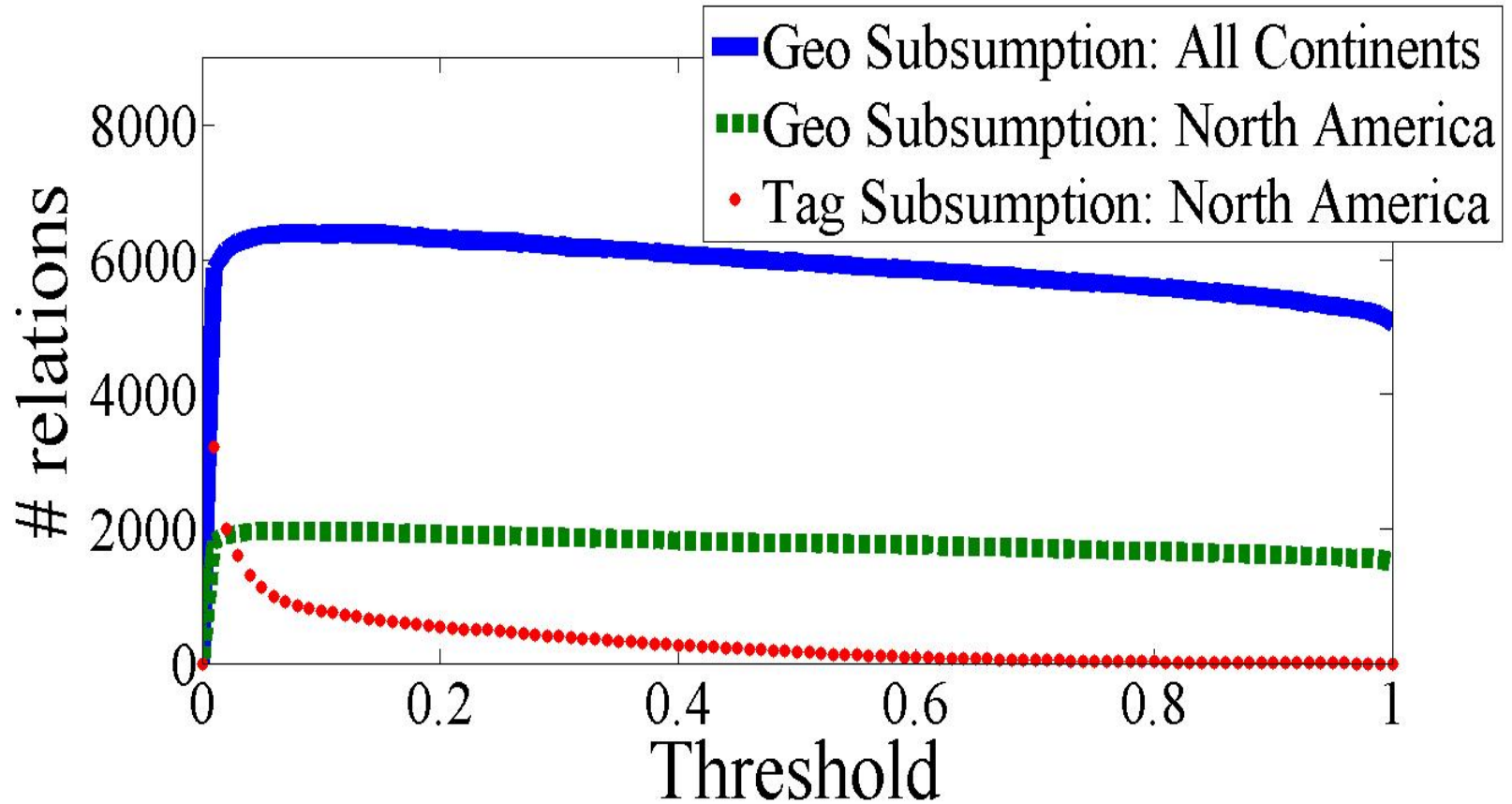
$$p(A|B) \geq t \text{ and } p(B|A) < t$$

$$p(A|B) = \text{Frequency}(A,B) / \text{Frequency}(B)$$

$$p(B|A) = \text{Frequency}(A,B) / \text{Frequency}(A)$$

Frequency(A,B): how many times tags A and B co-appear in the data set;
Frequency(A): how many times tag A appears in the data set.

Result: # of Induced Relations



Evaluate the quality of
learned relations against
the reference set from
GeoNames.org



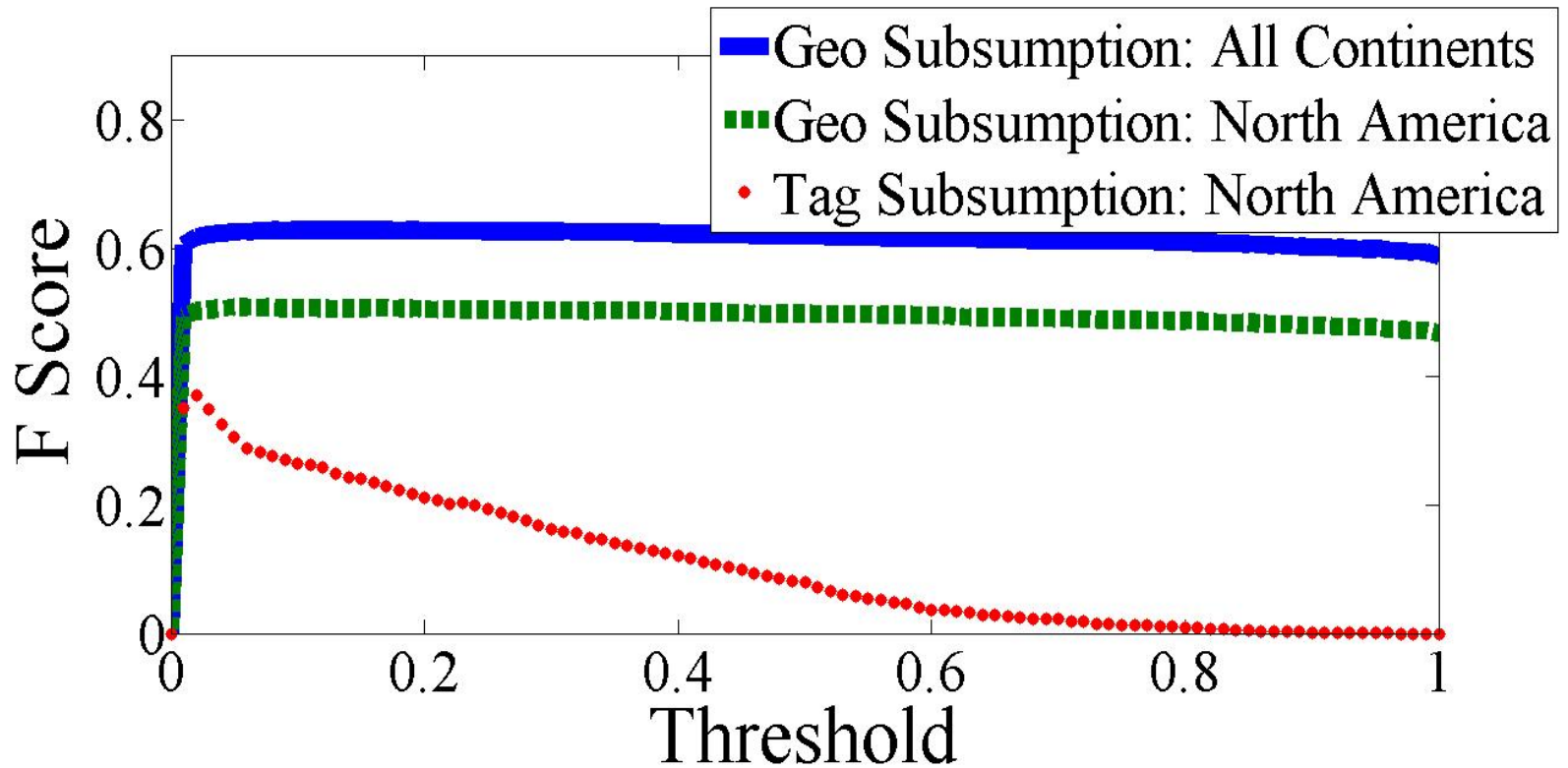
GeoNames

Earth

- Africa
 - Algeria (DZ)
 - Angola (AO)
 - Ascension (SH)
 - Benin (BJ)
 - Botswana (BW)
 - Burkina Faso (BF)
 - Burundi (BI)
 - Cameroon (CM)
 - Cape Verde (CV)
 - Central African Republic (CF)
 - Chad (TD)
 - Préfecture du Batha (01)
 - Préfecture du Biltine (02)
 - Préfecture du Borkou-Ennedi-Tibesti (03)
 - Préfecture du Chari-Baguirmi (04)
 - Préfecture du Guéra (05)
 - Préfecture du Kanem (06)
 - Préfecture du Lac (07)
 - Préfecture du Logone Occidental (08)
 - Préfecture du Logone Oriental (09)
 - Préfecture du Mayo-Kébbi (10)
 - Préfecture du Moyen-Chari (11)
 - Préfecture du Ouaddaï (12)
 - Préfecture du Salamat (13)
 - Préfecture du Tandjilé (14)
 - Comoros (KM)
 - Congo-Brazzaville (CG)

Result: Quality of the Learned Relations

$$\text{F-score} = (2 * \text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$$



Examples of Novel Relations

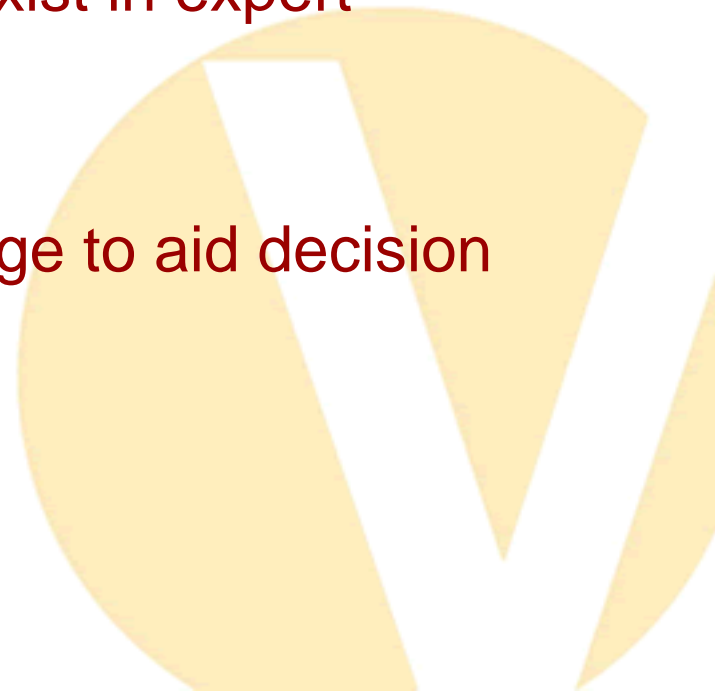
Child	Parent		Child	Parent
anaheim	la		disneyland resort	disneyland
ballard	puget sound		disneyland	la
brandywine park	wilmington		golden gate bridge	san francisco bay
bronx	new york city		pearl harbor	oahu
bronx zoo	new york city		times square	new york city
coney island	new york city		university of washington	puget sound

- Inducing conceptual hierarchies from tags' frequencies [Schmitz, 2006; Brooks & Montanez, 2006, Mika, 2007]
- Inducing conceptual hierarchies from collection – set relations [Plangprasopchok & Lerman, 2009]
- Identifying “place tags” from geopoints [Rattenbury & Naaman, 2009]
- Web-A-Where [Toni: please shortly mention it]
- Li et al. [Toni: please shortly mention it]
- Spirit tagger?



Our approach can:

- learn more high quality relations than the baseline
- learn novel relations that do not exist in expert-created knowledge bases
- leverage user generated knowledge to aid decision makers



Thank you

