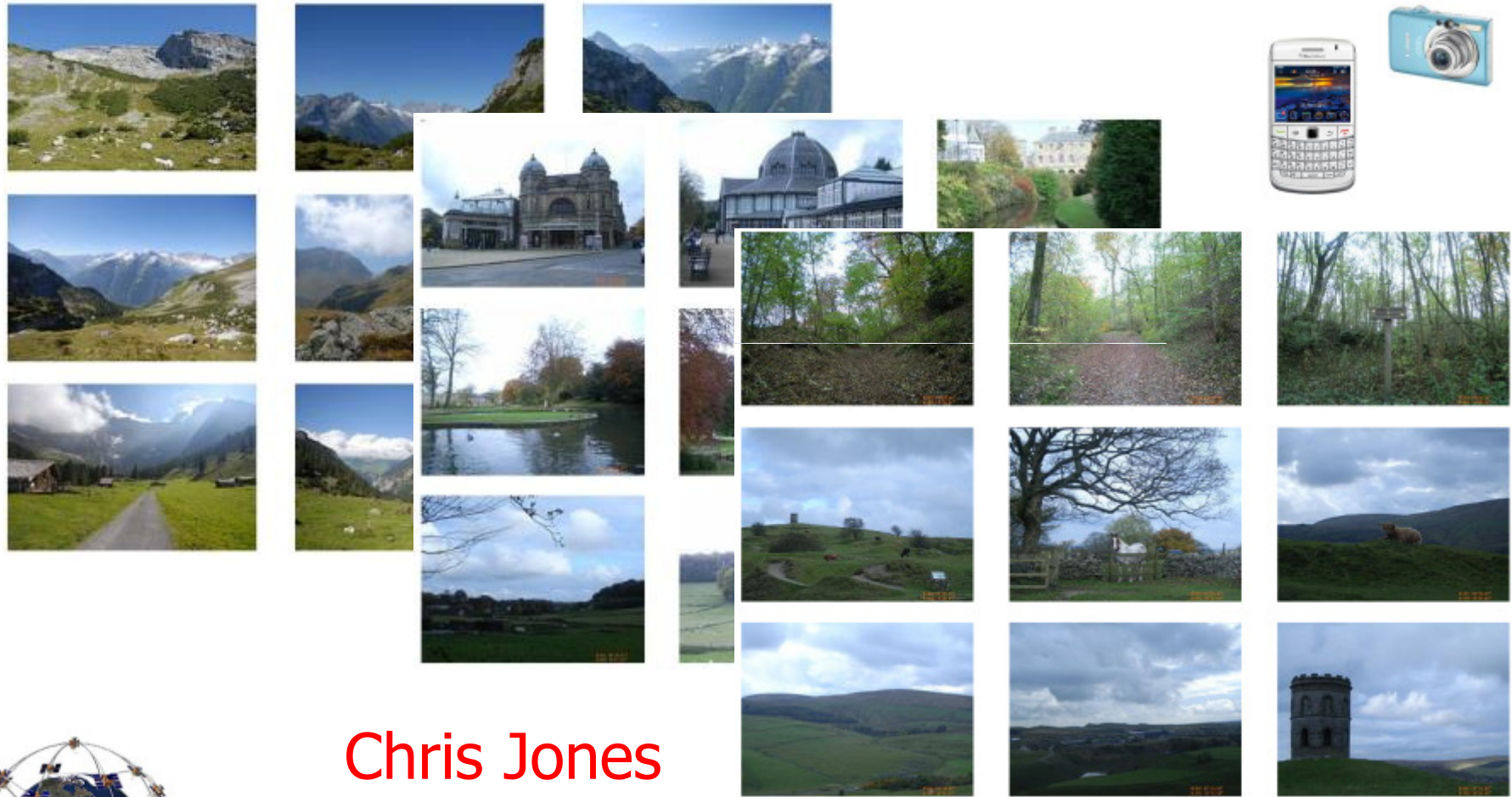


Natural Language Photo Captioning with Geo-referenced Data



Chris Jones
Mark Hall, Phil Smart, F. Twaroch, Vlad Tanasescu
Cardiff School of Computer Science & Informatics

GeoCaption

Version 2.1

Mobile App (Android)

View Gallery

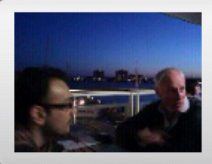
Capture Image

View Gallery

About GeoCaption



University of Portsmouth photographed in the early afternoon near Portland Street in Portsmouth (Region), United Kingdom.



Spink Tower and Portsmouth Harbour photographed in the evening near Portsmouth Harbour in Portsmouth (Region), United Kingdom.



Spink Tower in Portsmouth (Region), United Kingdom.



Adamsdown photographed at sunset on Clifton Street in the United Kingdom.



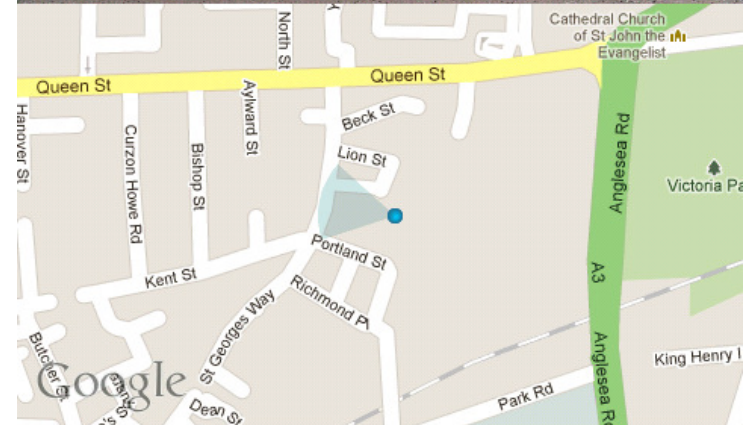
Adamsdown photographed at sunset near Clifton Street in the United Kingdom.



Adamsdown photographed at sunset on Clifton Street in the United Kingdom.

University of Portsmouth photographed in the early afternoon near Portland Street in Portsmouth (Region), United Kingdom.

28 Apr 2011 11:29:34 GMT



Two Broad Categories of Caption

From 2D Geo-Data

Camera Location only

Photo taken in the afternoon near Chatsworth House in the Peak District National Park, United Kingdom



Camera Location + Direction

Ladybower Reservoir photographed in the early afternoon near Snake Pass in the Peak District National Park, United Kingdom

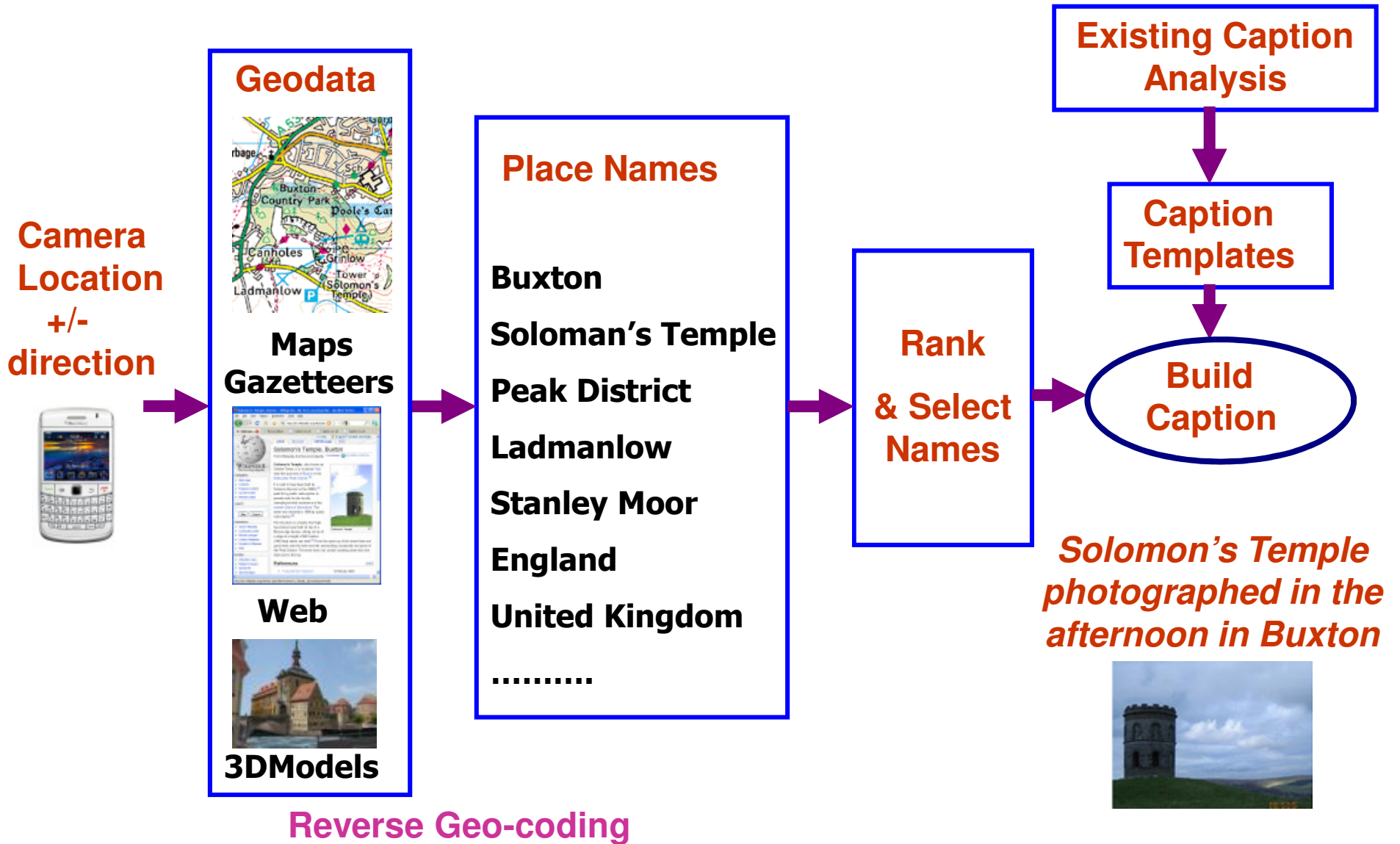


From 3D and 2D Geo-Data with Camera Orientation

Nieuwe Kerk and Koninklijk Paleis photographed in the afternoon, Amsterdam, Netherlands.
The Nieuwe Kerk is the building in the middle. The Koninklijk Paleis is the building on the left.



Overview: How is it done?

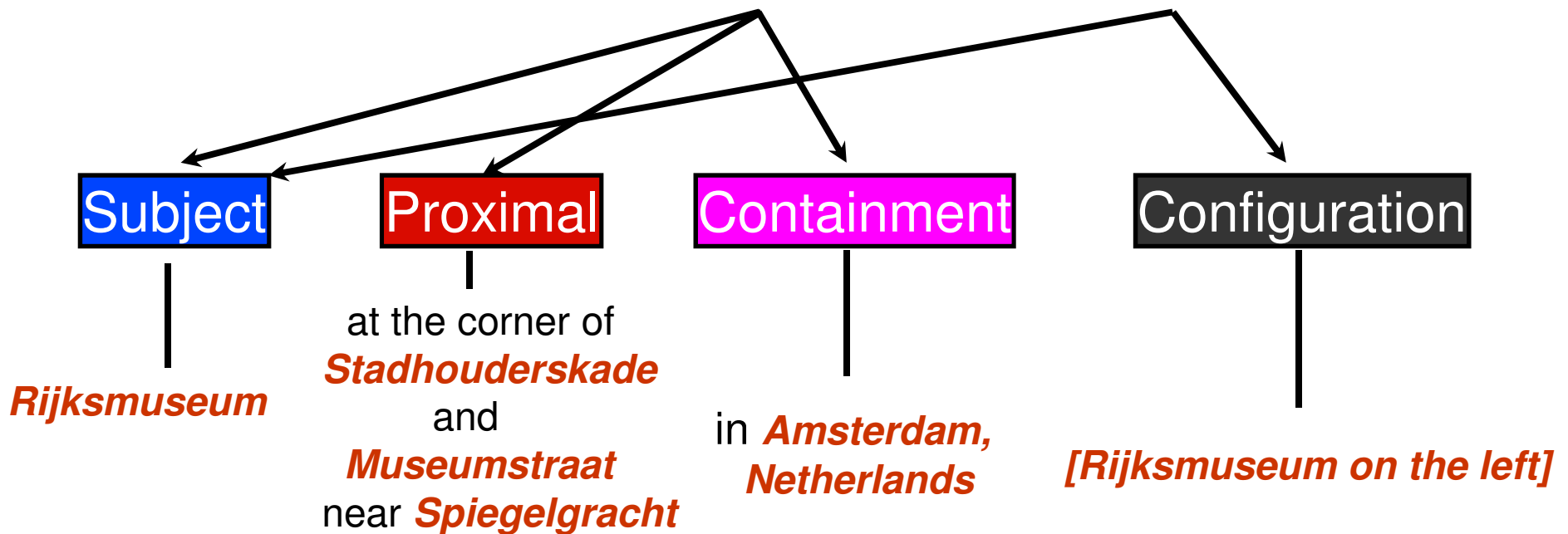


Caption Generation from Geo-data: Natural Language Templates



Rijksmuseum photographed at 2.15 pm at the corner of Stadhouderskade and Museumstraat near Spiegelgracht in Amsterdam, Netherlands

Reverse Geo-coding from 2D data 3D models



Depending on the data available not all elements may be generated

Understanding spatial language from analysis of Geograph captions

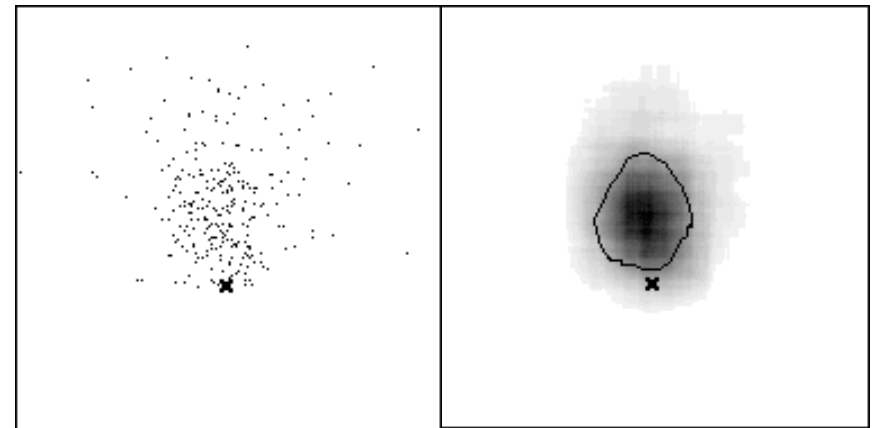
Need functional models of the usage of common prepositions such as near, north, at, etc

Most frequent spatial prepositions in Geograph
at, near, to, on, from, in, north, west, south, east

Plot distribution of camera coordinates relative to target (here Mosedale) using gazetteer coordinates



Boulders on moorland north of Mosedale



“North of”

X is target location

Caption language patterns from analysis of existing captions

Analysis of co-occurrence of words [POST], i.e. bi-grams, and of groups of words

→ Most Frequent patterns

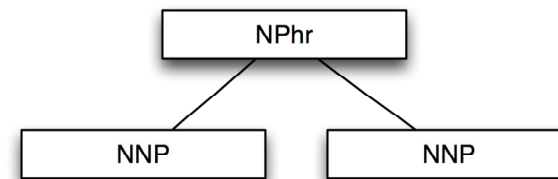
→ Templates, e.g.

Noun phrase

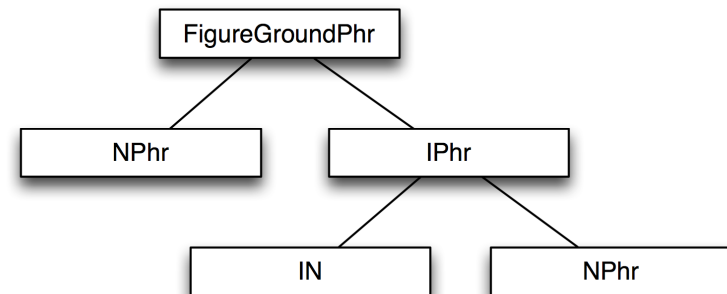
Figure ground phrase

Containment phrase

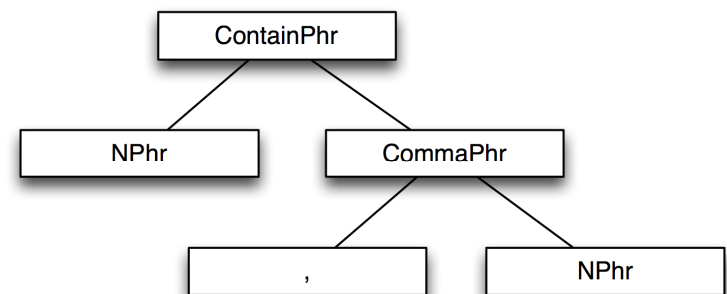
Given a photo location and nearby toponym, use field models to decide which is best preposition



"Merthyr Tydfil"



"Pontsticill reservoir near Merthyr Tydfil"



"Roath Park, Cardiff"

Data sources used in TRIPOD to find place names (Reverse Geocoding)



**Commercial digital maps & gazetteers
(Ordnance Survey, SwissTopo, TeleAtlas)**



Geo-referenced Wikipedia articles



Open Street Map

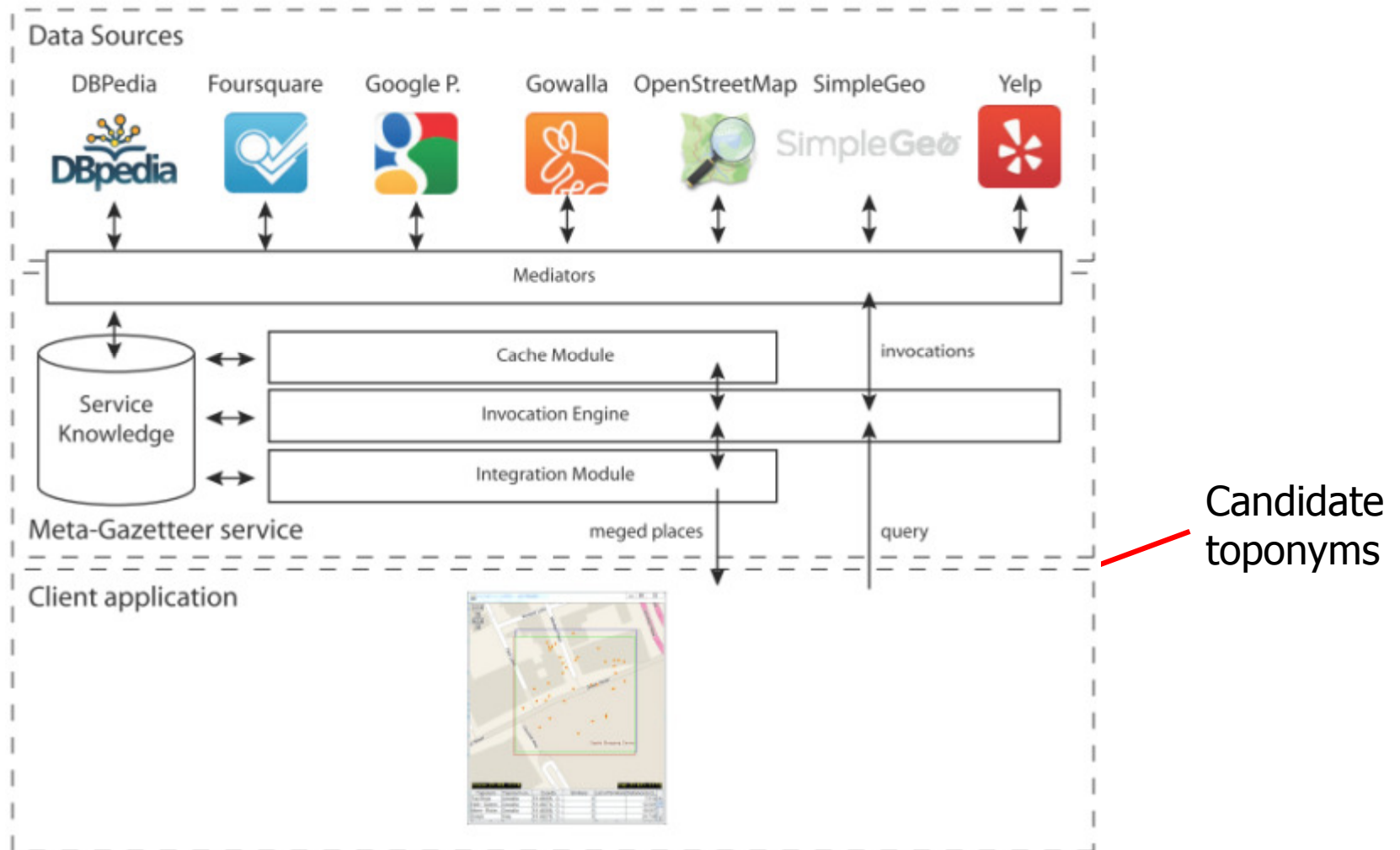
**Other Web sources:
Geonames Gazetteer
Yahoo WhereOnEarth**



3D Models of Cities

Reverse Geocoding with Meta-Gazetteer

Mediation Architecture for Selection and Integration of Multiple Sources



Mining the Web for Vernacular (Colloquial) Names

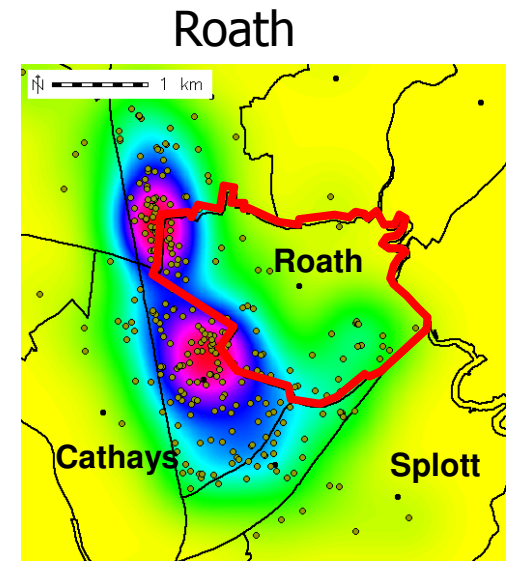
The Gumtree web site has adverts with place names and their postcodes



The screenshot shows several listings from Gumtree. The following vernacular names are circled in red:

- Caerau Cardiff
- Atlantic Wharf
- Roath
- Cardiff Bay
- Riverside
- Rumney

A blue box with the text "Neighbourhood name + Postcode" is overlaid on the listings, with an arrow pointing to the circled "Roath" listing.



Discover informal names

Understand the way
“formal” names are
actually used in practice



Photo taken on
Queen Street near
the Thistle Parc
Hotel in **Cardiff City
Centre**, Cardiff,
United Kingdom

Selecting and Ranking Names (Semantic Saliience)

Determine relative importance of the reverse geocoded toponyms

- Distance from the camera
- Web popularity from search of numbers of hits
- Flickr popularity use of the toponym at this location
- Boost for Wikipedia presence



**Derwent Reservoir
photographed in the early
afternoon in Derwent Moors,
United Kingdom**

Creating a simulated scene from 3D city models



Need to know:

location of camera (from GPS)

and its orientation (which direction it is pointing)

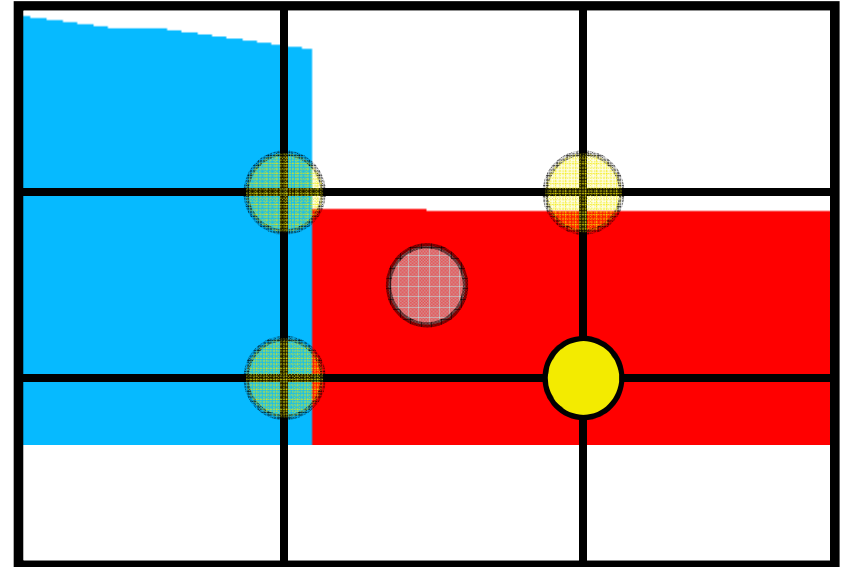
and camera characteristics (focal length,)

3D Urban Captioning with Spatial Relations



Nieuwe Kerk and Koninklijk Paleis
photographed in the afternoon, Amsterdam,
Netherlands.

The Nieuwe Kerk is the building in the middle.
The Koninklijk Paleis is the building on the left.



Derive spatial
configuration from
relations between
object regions and
the 9 sectors

Future

- Update Android captioning app with additional data sources e.g. foursquare, Gowalla, Google Places, Yelp etc
- Enrichment of 3D city models to provide additional information about places
- Augmented reality applications (Recognition project)
 - Adapt to user interests / preferences
 - Natural language scene description