Social, Geographical and Temporal Relevance for Recreational Queries in Web Search

Omar Alonso

Disclaimer

The views, opinions, positions, or strategies expressed in this talk are mine and do not necessarily reflect the official policy or position of Microsoft.

Work done with a great team

Outline

Research

LBSN data for IR

Practice

How to ship

Part I: Research

Introduction

LBSN

LBSN data for IR

Information needs

Recreational intent with a location preference

Search results

POIs and not web pages

Research questions

For what type of recreational queries are users trying to find answers?

Can a retrieval model be built to offer relevant places for recreational queries, based on LBSN data?

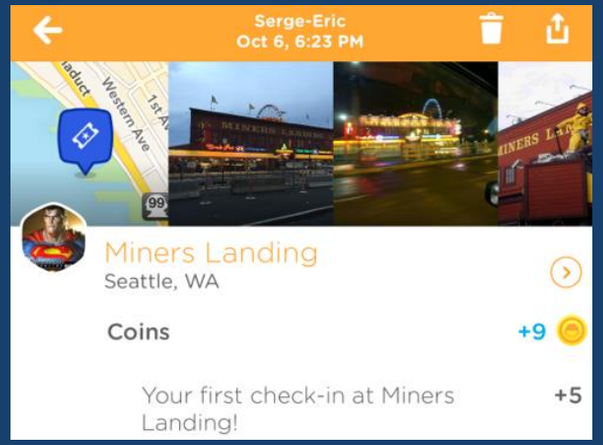
What's in a check-in?

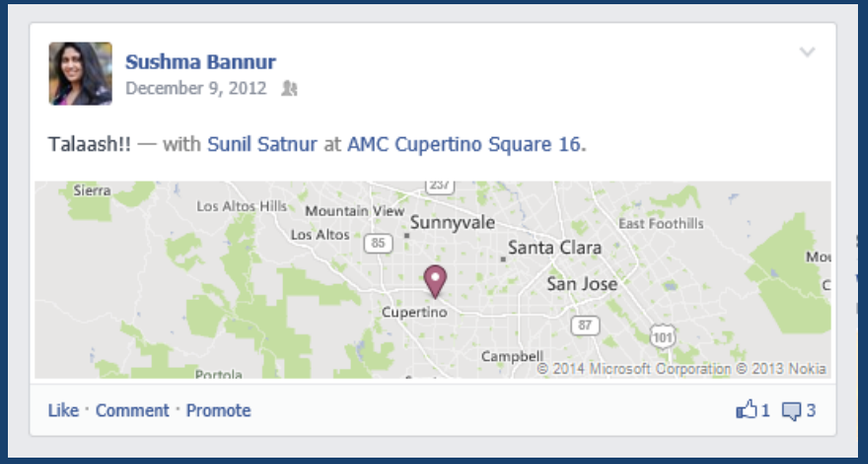
People checking-in at different places

The temporal view

Get-away check-ins peak during the summer months followed by Holiday seasons in winter

Patterns of human mobility correlated with the seasons in the geographic regions

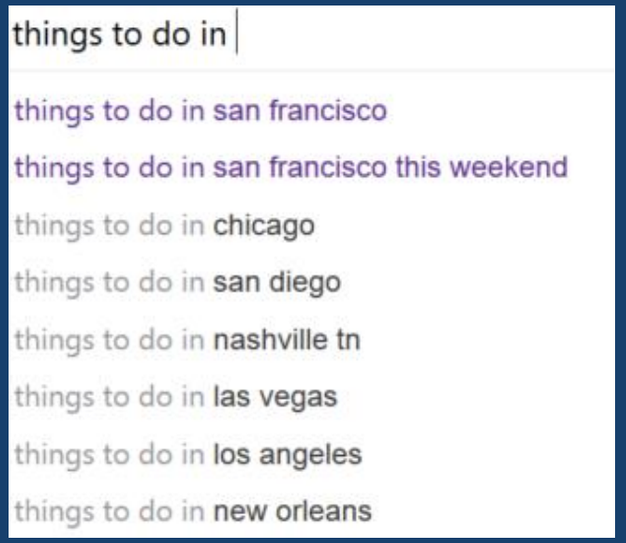


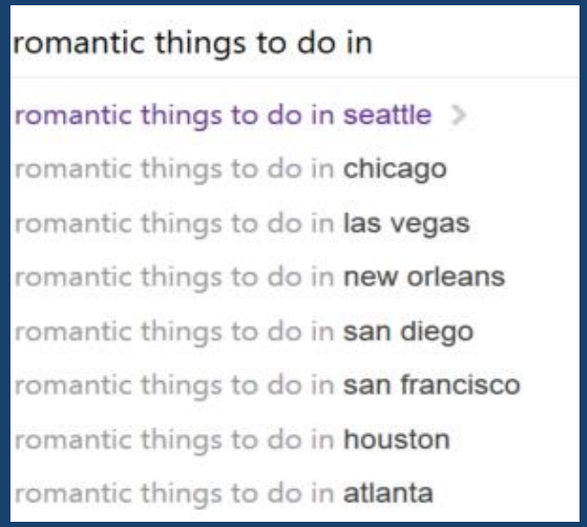


Connecting two crowds

People checking-in at different places

People looking for things to do in cities





Approach

Use LSBN data for search

Foursquare and Facebook check-ins to suggest points of interests (POIs)

Recreational queries

Social, geographical and temporal relevance

Recreational queries

Target segment: a common query intent

User is looking to perform an activity
… anchored around a place (a city)

Future trip planning or, real-time (now/soon)

Both tourist and local users

Desktop and mobile

Extracting location information

With insight from data sources

Query log analysis

Exploit venue check-in and tip popularity patterns

Data sources

Bing search query logs

IE behavioral data

Google queries, YouTube queries, Quora questions

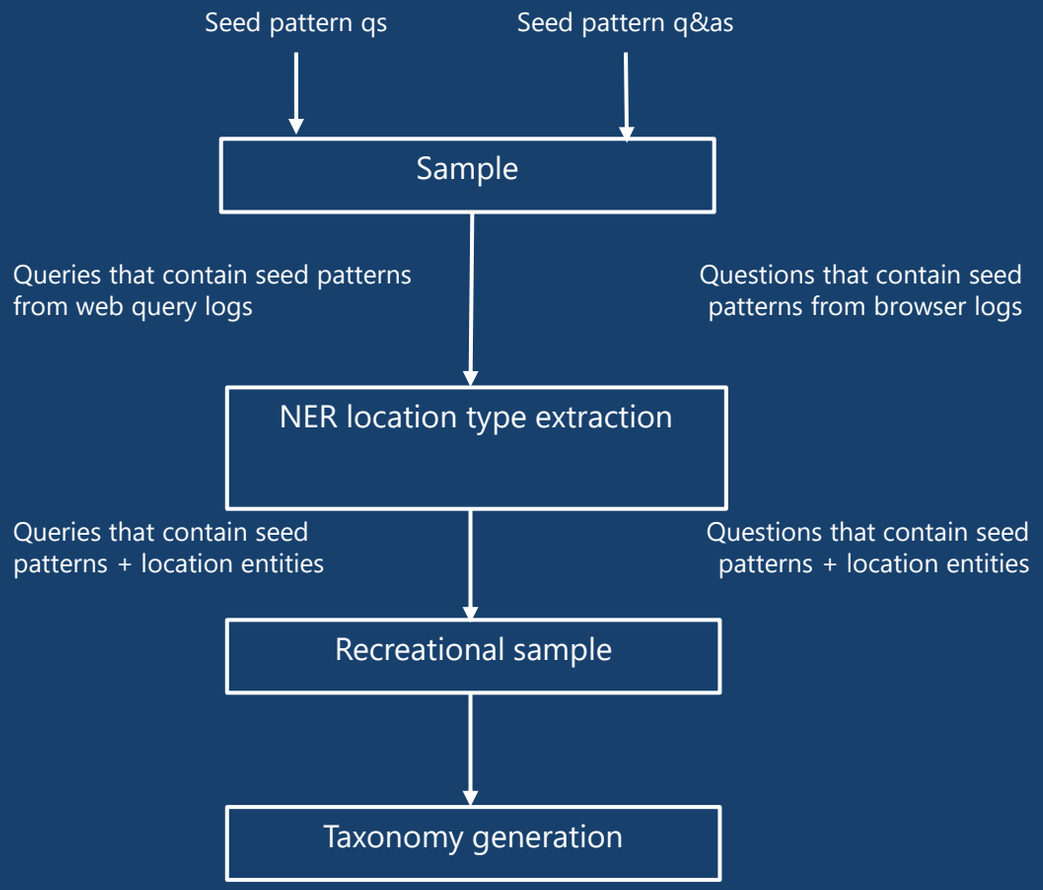
Foursquare check-ins + tips

Facebook check-ins + reviews

Recreational queries

Queries are composed of activity and venue needs

Taxonomy of recreational intents



A look at the query logs

Some patterns

Places to (p2)
Things to (t2)
What to (w2)

Pattern	Query	Location	Aspects
p2	places to visit in louisiana	Louisiana	visit
p2	unique places to stay in montana	Montana	unique
p2	places to hold a baby shower in colorado	Colorado	baby shower
p2	best places to snorkel in antigua	Antigua	best, snorkel
p2	best places to eat in modesto ca	Modesto	best, eat
t2	things to see in montana	Montana	see
t2	things to do in anchorage al	Anchorage	do
t2	things to do in stockholm sweden in august	Stockholm	do, august
t2	things to do in upstate ny in the winter	New York	do, winter
w2	what to do in houston this weekend	Houston	do, weekend
w2	what to do in gunnison with kids	Gunnison	do, kids
w2	what to do in madrid	Madrid	do

Browser logs

URLs from Q&A sites

https://quora.com/What-are-the-best-places-to-visit-on-a-family-vacation-to-St-Louis
https://quora.com/What-are-some-fun-things-to-do-in-London-with-kids

YouTube shows more queries with p2 pattern

Recreational query taxonomy

Taxonomy aspects

Geographical constraint

Absolute and relative (near, here)

Temporal constraint

Absolute (events/calendar) and relative (tonight)

Activity preference

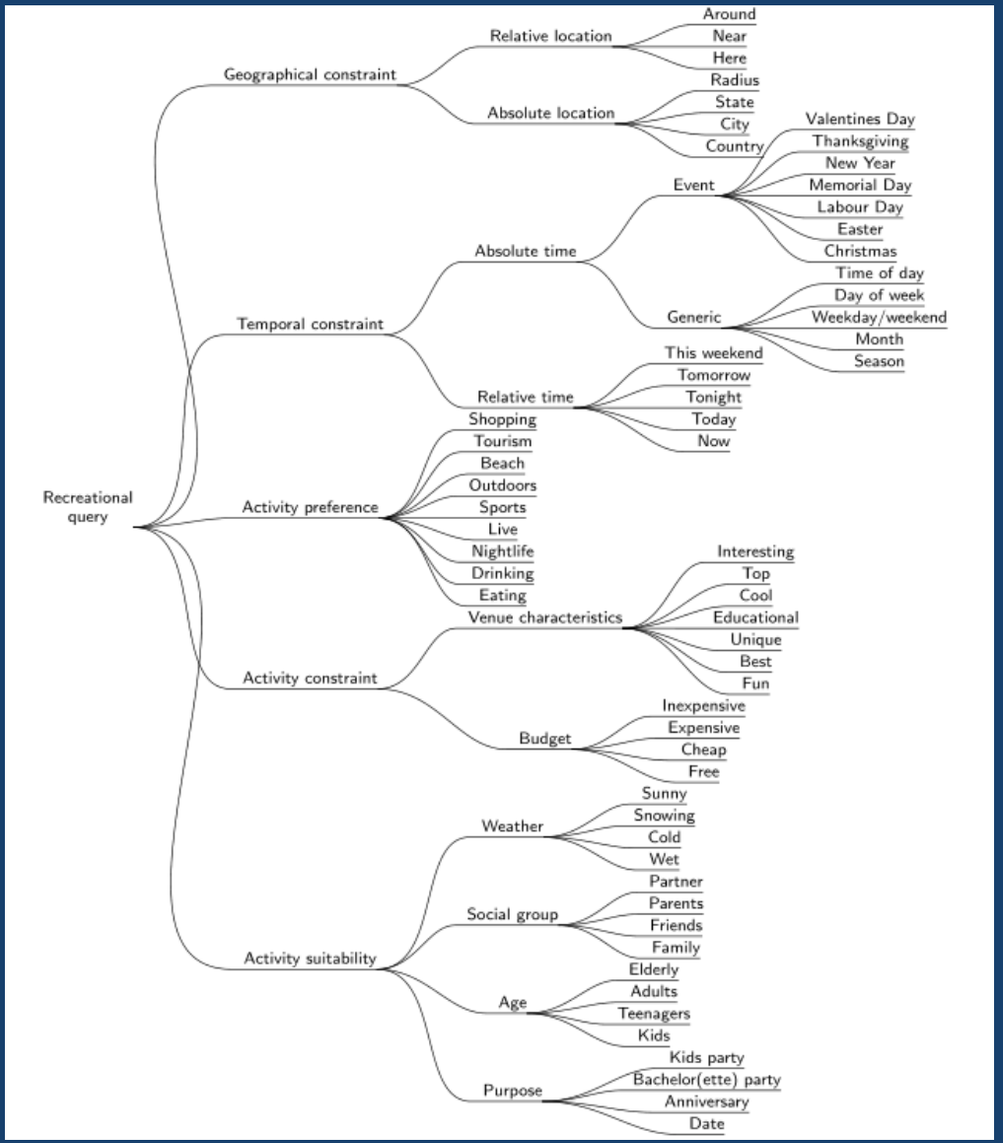
Eat, drink, shop

Activity suitability

Romantic, kids, age group

Activity constraint

Free, cheap, unique, educational



Query classification

Simple keyword-based classifier

Popularity of different taxonomy aspects

Changes between city and at different times

Miami	Paris	Rome
activity_type: tourism	activity_type: tourism	activity_type: tourism
activity_sentiment: best	activity_sentiment: best	activity_sentiment: best
age_group: child	activity_type: eating	activity_type: eating
relative_temporal: weekend	activity_sentiment: fun	activity_sentiment: fun
activity_type: beach	absolute_temporal: may	relative_location: near
activity_sentiment: fun	age_group: child	absolute_temporal: may
activity_type: eating	absolute_temporal: june	absolute_temporal: april
budget: free	budget: cheap	age_group: child
relative_location: near	budget: free	budget:cheap
activity_type: live	relative_location: near	budget: free

POI ranking

First by geographical constraints

Temporal

Check-in data tell us *what* places are popular

Check-in times tell us *when* places are popular

Aspects

Tips and reviews offer clues on the utility of the POI

Modeling POI relevance

Maximum likelihood estimation

$$MLE(q_{rec}) = \arg\max_{poi_i \in poi_c} P(poi_i|q_t, q_a)$$

Temporal ranking

Observe probability of check-in during week and seasons

$$P(poi_i|q_t) = P(poi_i|City, PartOfDay, DayOfWeek, WeekdayWeekend, Season)$$

Aspect ranking

Filtering by eating, drinking etc.: easy with category keyword matching

Match tips to the taxonomy using same aspect classifier for queries

POI – “my kids loved the great food here and it was cheap”

3 aspects (“for kids”, “cheap”, “great”)

$$aspect_{rel}(poi_i, q_a) = (1 + P(q_a|poi_i)) \times category_{rel}(q_i, poi_{category})$$

Bundles

Idea from composite retrieval

Aspects that people are looking for are covered in the taxonomy

User don’t always express their needs well

Bundles offer an exploratory entry option

Summarize of venues to provide in SERP

Ranking

Group together cohesive bundles of venues for each aspect, with tips

Building trails

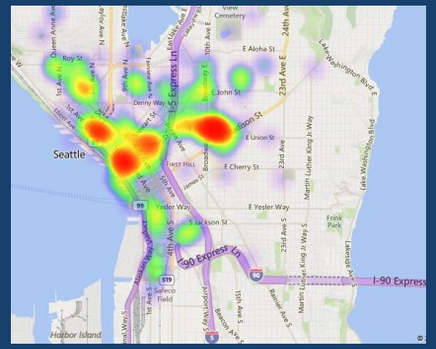
Where might a user go next? How far will they travel?

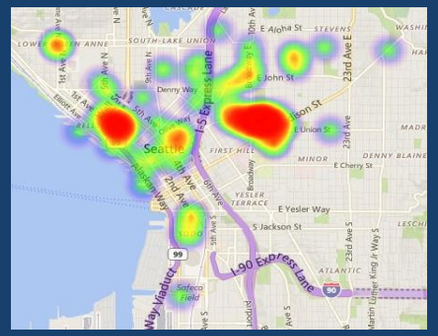
Moving beyond individual and categorized results

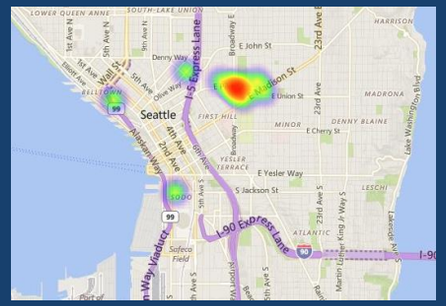
Assist activity ideas and planning

Recommend sequences of things to do

Dynamic heatmaps







Pike Place Market

The Crab Pot (0.2m)

see also:

- Kells Irish Restaurant & Pub (0.1m)
- Piroshky Piroshky (0.1m)
- Japonessa (0.1m)
- Beecher's Handmade Cheese (0.1m)
- The Cheesecake Factory (0.4m)

Starbucks (0.3m)

Downtown Seattle (0.3m)

Hard Rock Cafe Seattle (0.2m)

Von Trapp's

Unicorn (0.2m)

Garage Billiards (0.2m)

Q Nightclub (0.1m)

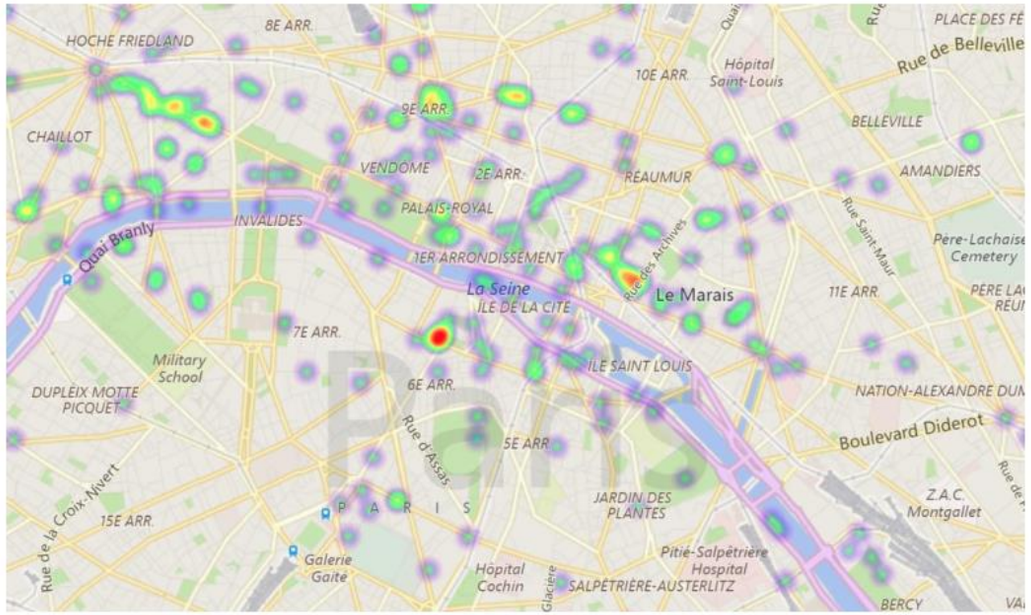
Tallulah's (1.0m)

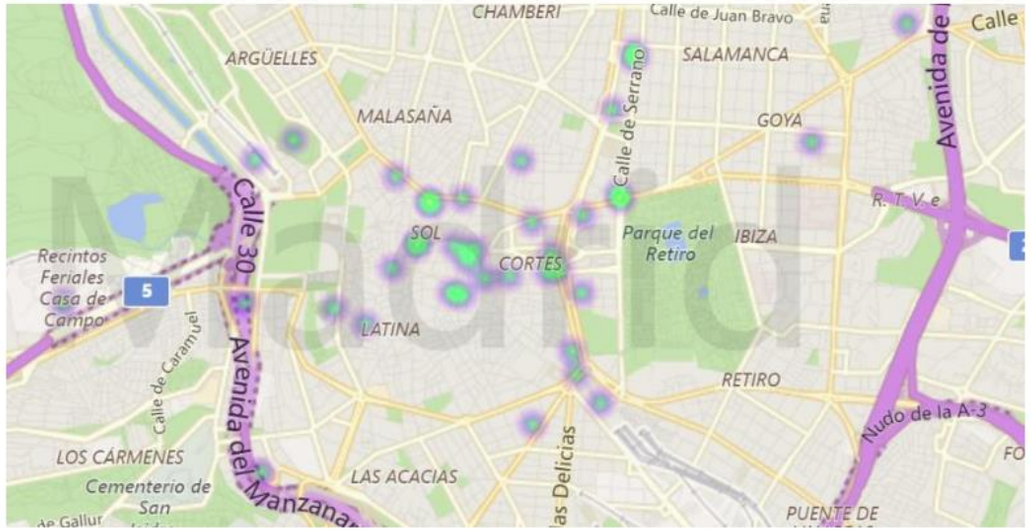
Neighbours

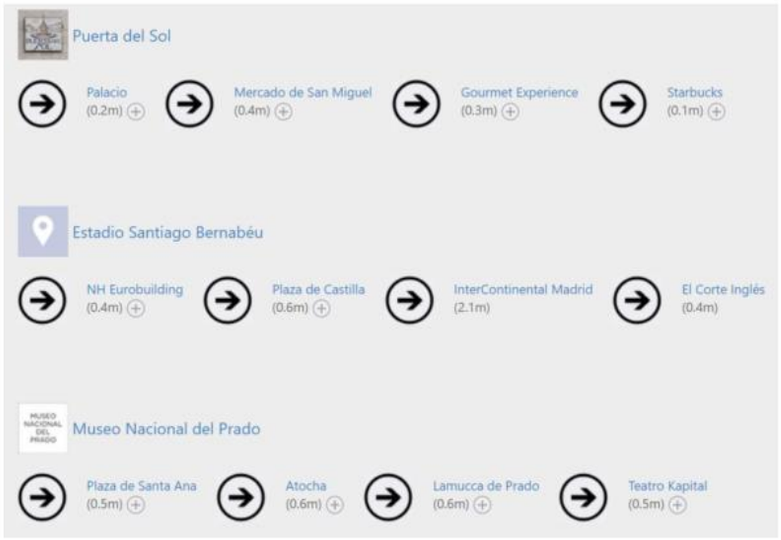
Re-Bar (0.5m)

Foundation Nightclub (0.8m)

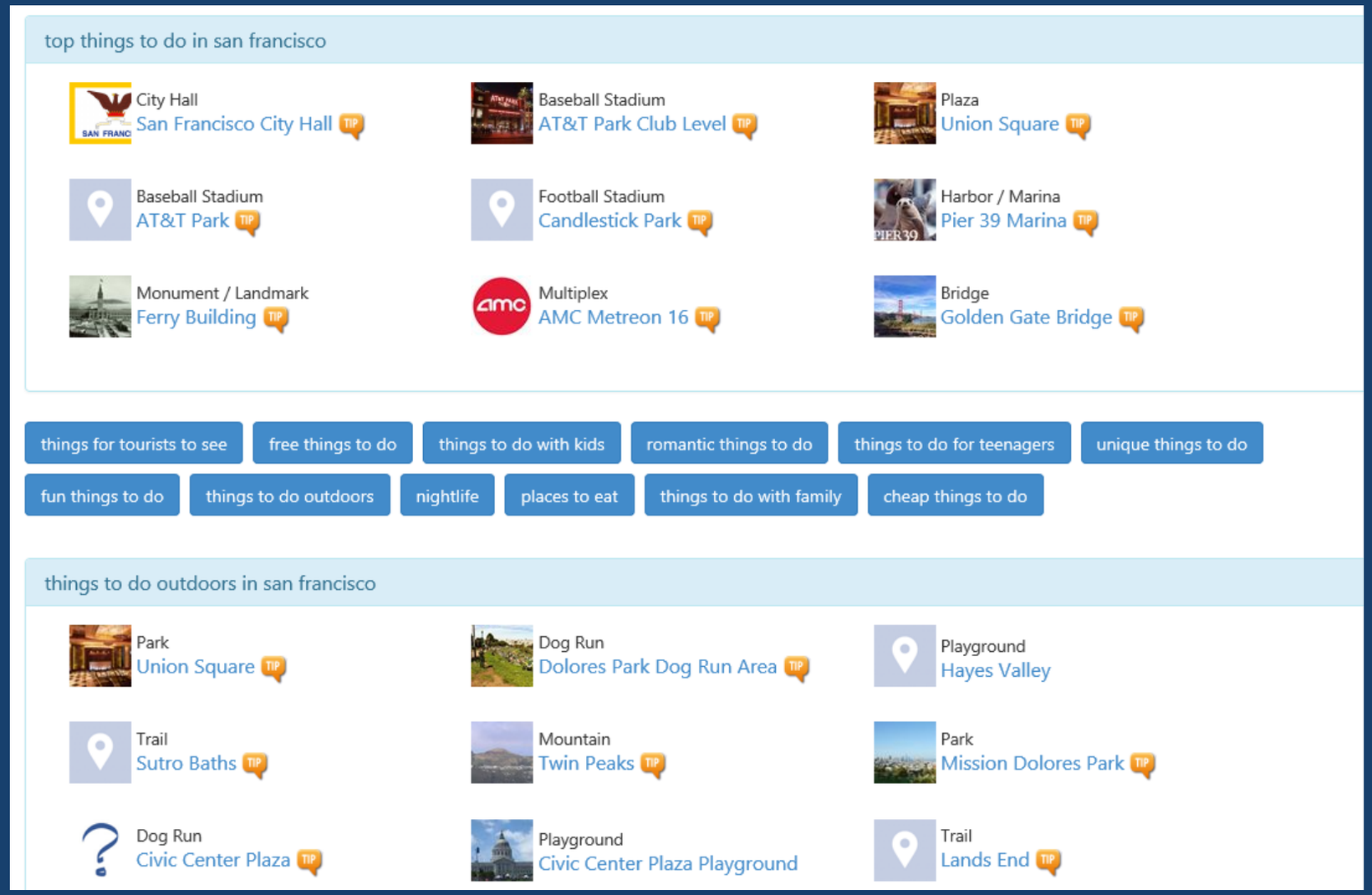
Dynamic heatmaps



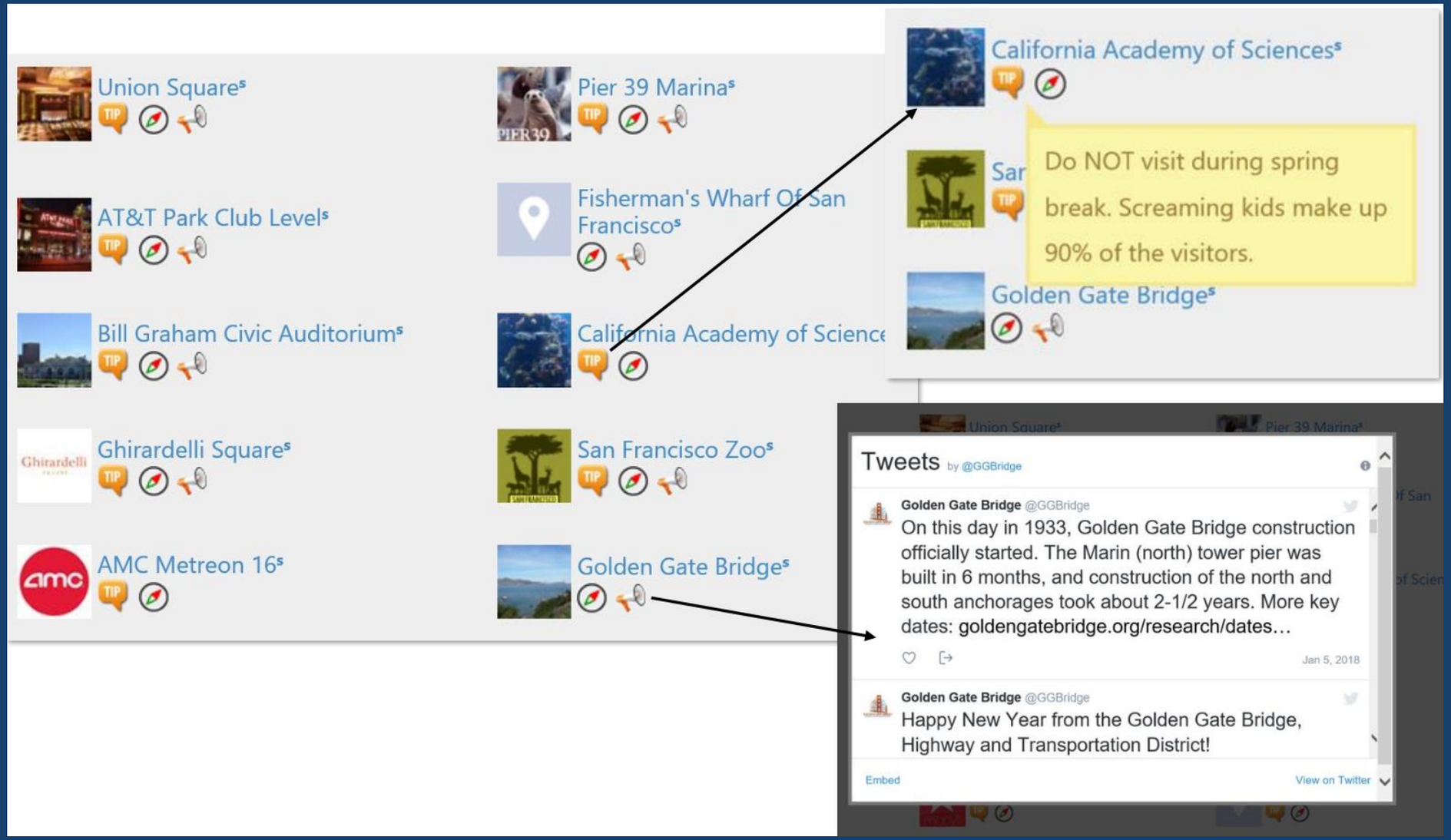




Bundles and aspects



POIs conflation + annotations



Dataset and initial results

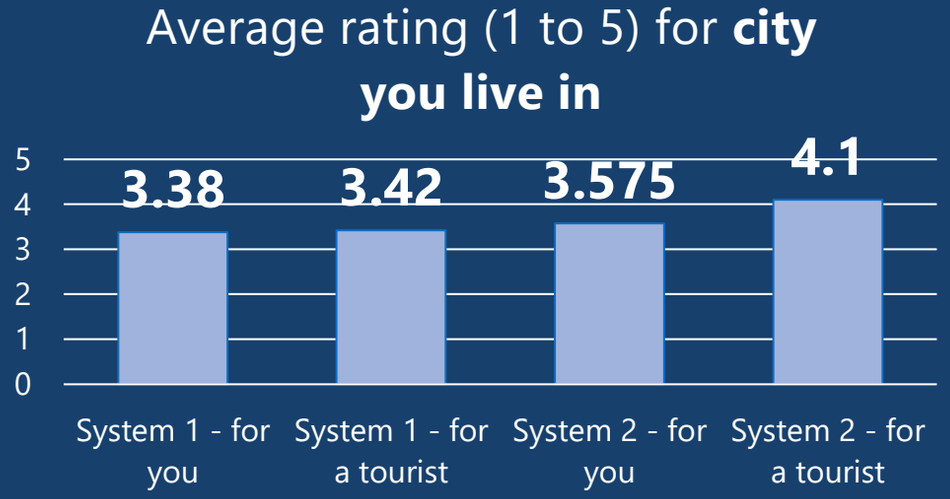
1B check-ins

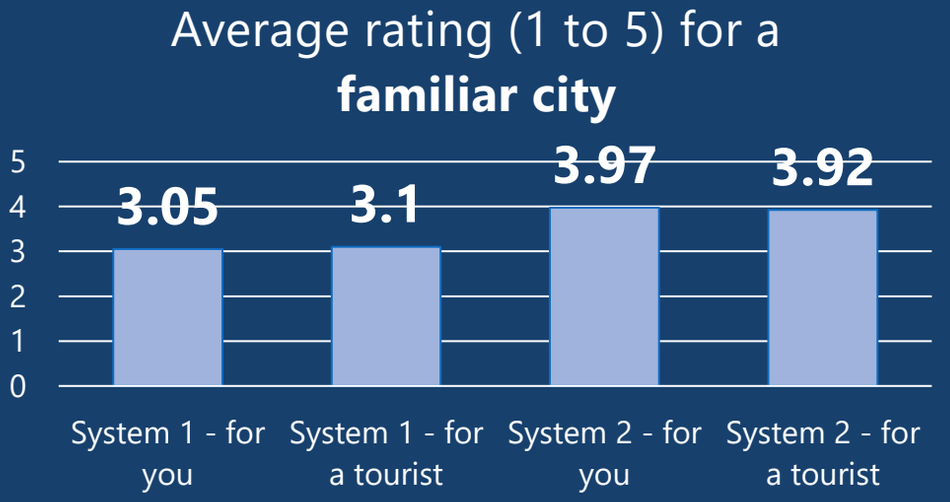
POI annotation

Check-ins
Tips and reviews
Photos
Categories
Social presence

User study for bundles

System 1 (list) vs System 2 (bundles)
Users prefer bundles, especially for familiar cities
Tourists could benefit more from bundles





Part II: How to ship

What we have

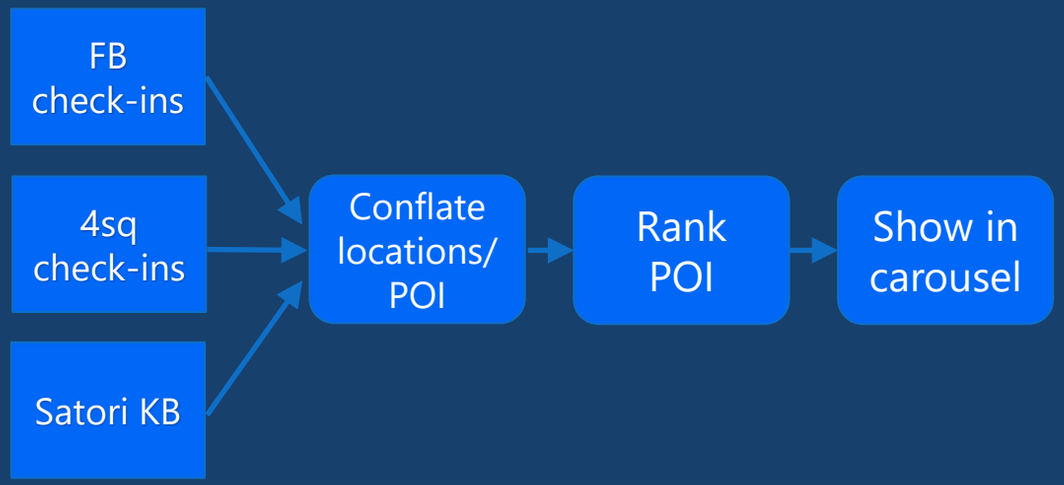
Several working prototypes

Scope/Cosmos

Lots of data

Different avenues to pursue

The plan



Ranking attractions and POIs

Not all POIs are suitable for recreational queries

Bad: bus stations, high schools, office buildings, prisons, gas stations
Ambiguous: bridges, train stations
Potentially good: restaurants, churches, parks
Good: famous landmarks, tourist attractions

Training with TLC

Collect labels; 3 ratings (Good, Fair, Bad)
Fastrank; standard TLC ranking gains

Features

Scaled by city query impressions and POI category averages
Higher order features to distinguish repeated vs. new visits (e.g., train station vs. tourist attraction)

Annotation classifier

Need to classify POIs using annotations

High quality -> "get chowder at pike place chowder - its yummy"
Low quality -> "what's not to like?"

Processing tips (4sq) and reviews (Facebook)

Quality classifier
Language detection
Sentiment analysis

Quality classification

Boosted decision tree model using TLC
Proper English, punctuation
Number of conjunctions
Contains phone number
Other NLP features

Evaluation

Human evaluation

10K POIs

POI rated Good, Fair, Bad

Average quality 2.77

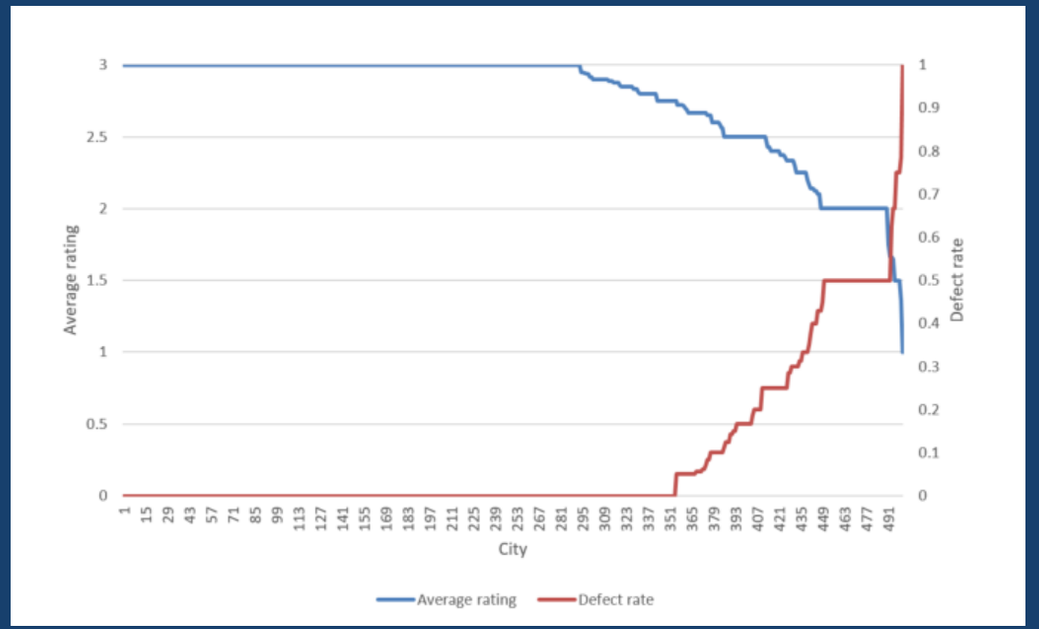
Defect rate about 17%

SGT ranker vs. check-in ranker

Online metrics

SBS

Can’t report behavioral data, I’m afraid



Position N	baseline NDCG@N	NDCG@N
1	0.877	0.977
2	0.889	0.978
3	0.890	0.977
4	0.895	0.976
5	0.900	0.977

Data management

Focus on high quality data

Own the process

Processing pipeline

POI conflation/entity disambiguation

Merge using a kd-tree to find nearby POIs

things to do in san francisco

Sign in 10

Web Images Videos Maps News Explore

San Francisco - Points of interest


Golden Gate Bridge


Alcatraz Island


Chinatown


Lombard Street


Transamerica Pyramid


Fisherman's Wharf


San Francisco cable car system


The Castro

things to do in san francisco with kids

Sign in 10

Web Images Videos Maps News Explore

San Francisco - Points of interest


California Academy of Sciences

San Francisco Zoo

Aquarium of the Bay

The Walt Disney Family Museum

Musée Mécanique


Pier 39


Lands End


The LAB

things to do in california

Sign in 10

Web Images Videos Maps News Explore

California - Points of interest


Mount Whitney


Golden Gate Bridge


Yosemite National Park


Disneyland


California State Capitol

Hollywood Sign


Lake Tahoe

University of California, Berkeley


Alcatraz Island

things to do in united states

Sign in 10

Web Images Videos Maps News Explore

United States - Points of interest


Statue of Liberty


White House


World Trade Center


United States Capitol


Mount Rushmore National Memorial


Empire State Building


Grand Canyon


United Nations Secretariat Building


Times Square

things to do in paris

Sign in 11

Web Images Videos Maps News Explore

Paris - Points of interest


Eiffel Tower


Louvre


Arc de Triomphe


Notre Dame de Paris


Champs-Élysées


Palace of Versailles


Disneyland Paris


Montmartre

things to do in europe

Sign in 12

Web Images Videos Maps News Explore

Europe - Points of interest

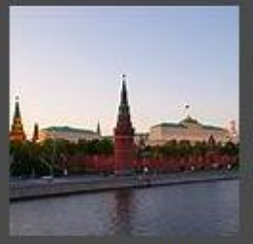

Eiffel Tower


Berlin Wall


Stonehenge


Colosseum


Auschwitz concentration camp


Moscow Kremlin


Leaning Tower of Pisa


Big Ben

The product side

Startup mentality

Your research idea won’t ship as is

Be ready to pivot

Focus on the business problem that you are trying to solve

Always have a working prototype ready

Working code wins over PPT

Produce valuable assets

Datasets

Code

Metric

Process

Conclusions - research

Summary of work

Characterization of recreational queries

Taxonomy

SGT framework

Dynamic heatmaps

Aspect bundles

Main publications

O. Alonso, V. Kandylas, S. Tremblay, S. Whiting “Answering recreational web searches with relevant things to do results” Inf. Process. Management, 2020

S. Whiting, O. Alonso, V. Kandylas, S. Tremblay “Urban Maps of Social Activity.” ICWSM 2018

S. Whiting, O. Alonso “SGT Framework: Social, Geographical and Temporal Relevance for Recreational Queries in Web Search” SIGIR 2016

Conclusions - product

From research to production

Influenced how to extract value out of social data (MS)

Reusable data set

Data generation pipeline
Relevance evaluation
Publishing
Privacy

US granted patents