

Identifying geographic locations using OS based side channels

D. Sai Pavan CS14B041
S. Sai Teja Reddy CS14B051



Aim of the project

The aim of this project is to predict the city rendered in Google Maps using the process' memory footprint.

The memory footprint is measured using the browser's DRS (data resident size) aka RSS (resident set size). DRS value reflects the processes' total size of its heap, stack, and mmap-allocated memory.



Background

Consider two websites having very different layouts. They produce vastly different memory footprints.

They can be classified using a matching algorithm based on a similarity index like Jaccard Index.

In Suman Jana and Vitaly Shmatikov's 2012 paper, in a dataset of about 100,000 websites, they reported that about 30% of the websites are distinguishable.

The Challenge

These have very low similarity while...

Experiments - CC Networks Stoke v Leicester Avicii - Waitl Case study - Go Course Project premierleague

Secure | <https://www.google.co.in/search?dcr=0&source=hp&ei=ub9WcmNczH0ASckGYBw&q=premierleague...>

Seedr: Torrent D ISPUnblock Yesmovies - Wal Premier League Flash Scores: Liv soccerstreams.n Exploit Exercise Prime Video FMovies

Google premierleague

All News Books Images Videos More Settings Tools

About 18,50,00,000 results (1.05 seconds)

Premier League

Scores & Schedule

< 8 Oct	Sun, 29 Oct	Tue, 31 Oct	Today	Tomorrow	Sat, 18 Nov	Sun, 1 >
Stoke City	2			Southampton	0	Live - 56'
Leicester City	2	FT		Burnley FC	0	
Newcastle	0			Huddersfield	1	Live - 55'
Bournemouth	0	Live - 55'		West Brom	0	
Swansea City	0			West Ham		11:00 PM
Brighton	1	Live - 55'		Liverpool		

All times are in India Standard Time

premierleague on Twitter
<https://twitter.com/search/premierleague>

Premier League
(@premierleague)

Time for more wonder hits...?

Follow all the action

SportsJOE
(@SportsJOE_UK)

Very clever of the Premier League to arrange the four

Bridge News
(@cfc_wale)

Morata: "They demonstrated to me that they really wanted

Teams

View 20+ more

Experiments - CC Networks Stoke v Leicester Avicii - Waitl Case study - Go Course Project Yahoo

Secure | <https://in.yahoo.com>

Seedr: Torrent D ISPUnblock Yesmovies - Wal Premier League Flash Scores: Liv soccerstreams.n Exploit Exercise Prime Video FMovies

Home Mail News Cricket Celebrity Movies Lifestyle Flickr Mobile More

YAHOO!

Sign in Mail

Mail Cricket News Finance Lifestyle Movies Celebrity Shopping More...

Live Commentary: India falter in 197 chase

Live ball-by-ball action from the 2nd T20I in Rajkot.

Scorecard »

1. Masood Azhar 6. India vs NZ T20I
2. Trump Twitter account 7. Thor Ragnarok
3. NTPC blast toll 8. World Food India 2...
4. Chennai rains 9. Kamal R Khan Twit...
5. House cleaning 10. Ranji Trophy 2017/...

Chennai, Tamil Nadu

Today	Sun	Mon	Tue
27° 25°	28° 24°	27° 25°	26° 24°

Featured Picks

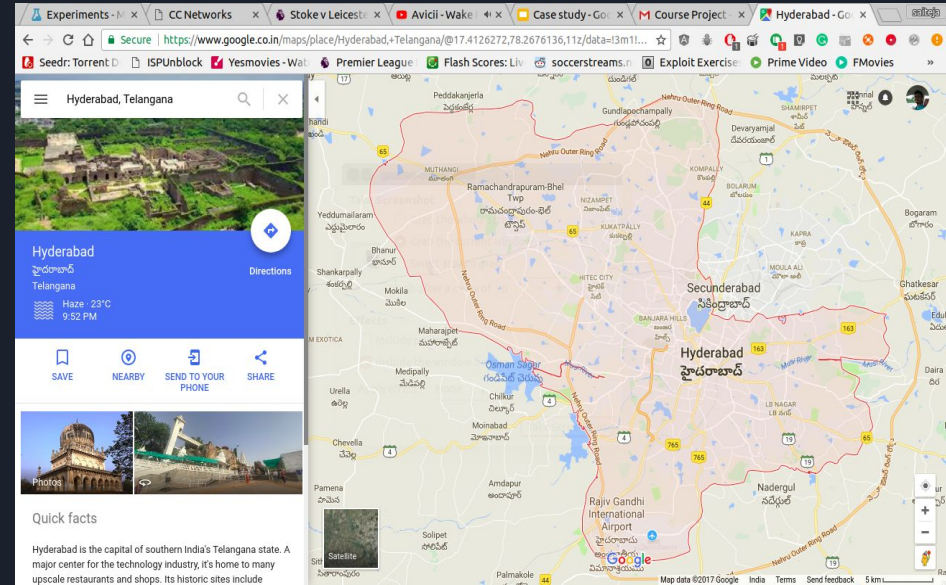
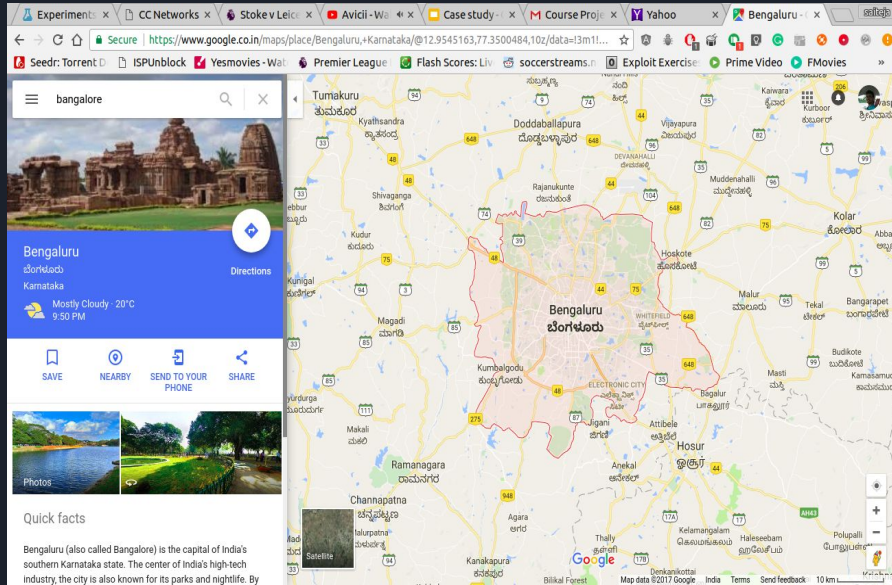
Match Photos: India vs New Zealand, 2nd T20I
Yahoo Cricket

Rs 2.11 lakh crore infusion into

<https://cricket.yahoo.com/cricket-live-score-india-vs-new-zealand-197343>

The Challenge

these have relatively high similarity.



Implementation





Collecting data

We used chrome to open the web sites.

Chrome uses a multi process architecture

- One main process
- One GPU process
- One process per tab, extension

We are interested in the Google map tab's process.

Reqd DRS value is in the 2nd column of
`/proc/<pid>/statm`



Collecting data

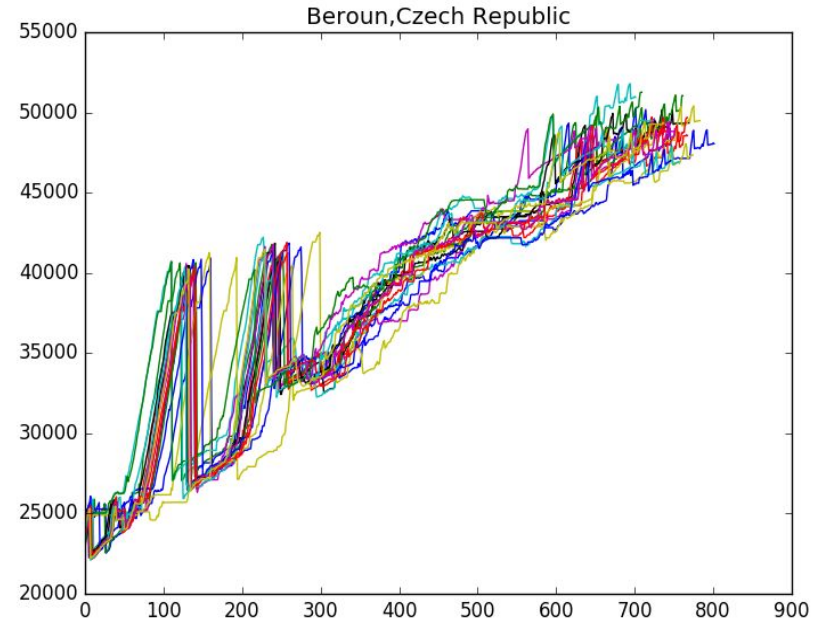
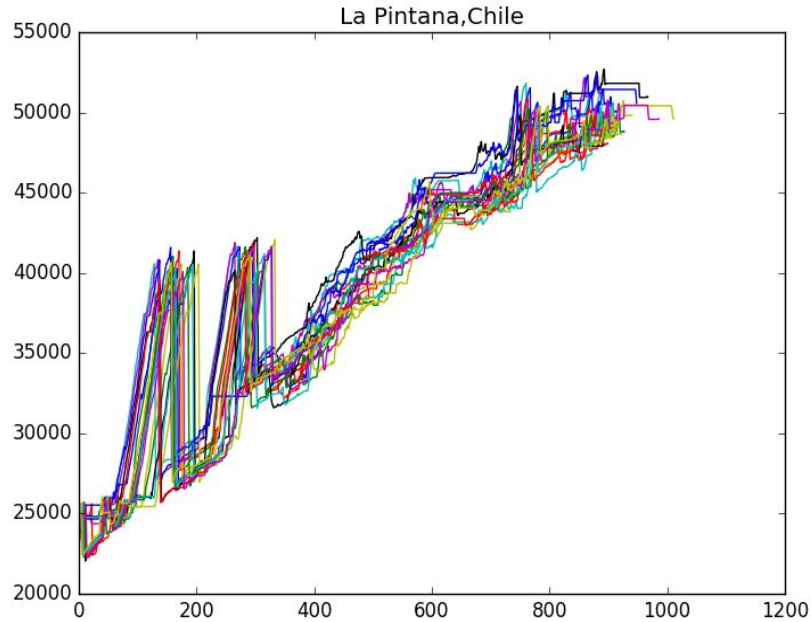
From a dataset of 23,000 cities we chose 50 cities randomly.

This was done to avoid selecting a biased metropolitan dataset.

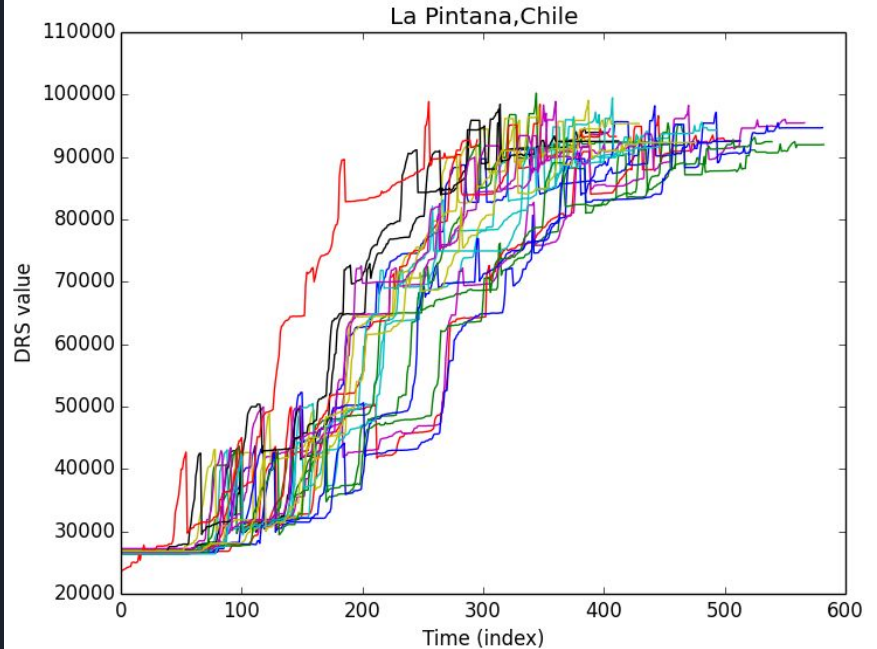
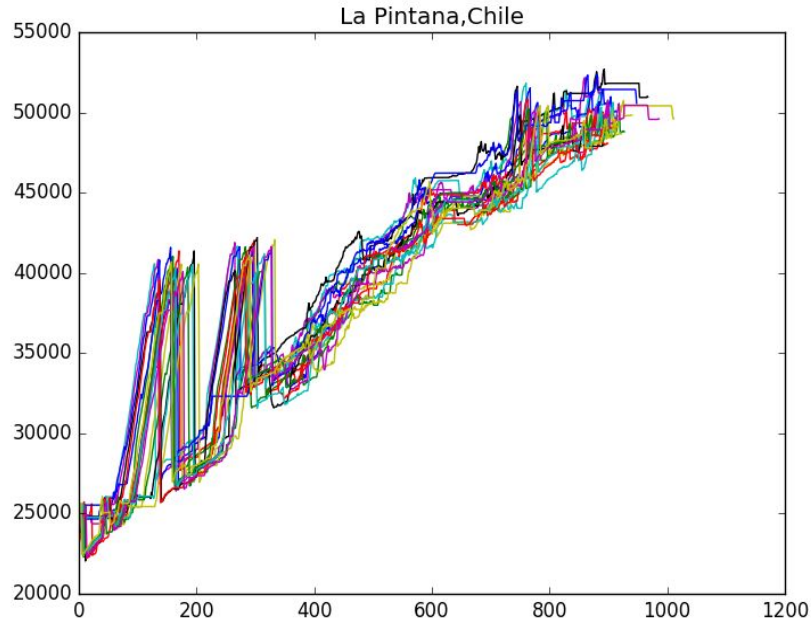
We then proceeded to collect the DRS value for each city 20 times using an automated script written in python.

The DRS value was collected continuously until the Google Maps page finished loading.

DRS value vs time for two different cities (on the same machine)



DRS value vs time for the same city on different machines





Analyzing data

We used the collected data to build a Random Forests machine learning model.

The feature vector for a city consisted of all the DRS values collected.

Cities with smaller features vectors were padded with their last obtained DRS values so as to have equal number of features.

We used Azure ML studio for this purpose.



Results

Using the obtained trained models we obtained an accuracy of 30.24% on validation data.

On the corresponding test data we obtained an accuracy of 29.59% .

Thus we were able to match the accuracy reported in Suman Jana and Vitaly Shmatikov's 2012 paper for a relatively more difficult problem.



Improvements

The obtained results can be improved by

- Better data collection techniques
- Further fine tuning the Random Forest parameters.
- Using other techniques like RNNs.



References

Suman Jana and Vitaly Shmatikov's 2012
paper:

Memento: Learning Secrets from Process
Footprints

<http://ieeexplore.ieee.org/document/6234410/>



Thank You