

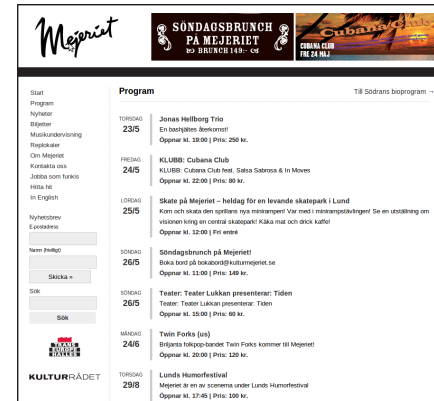
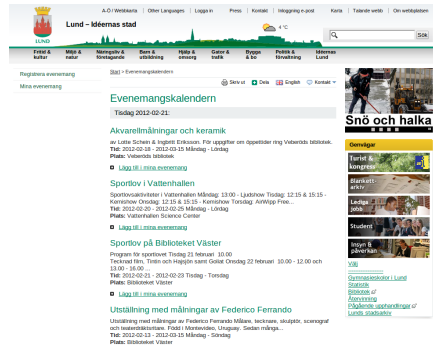
# **Automatic extraction of local events from web sites**

By:

Anton Risberg Alaküla

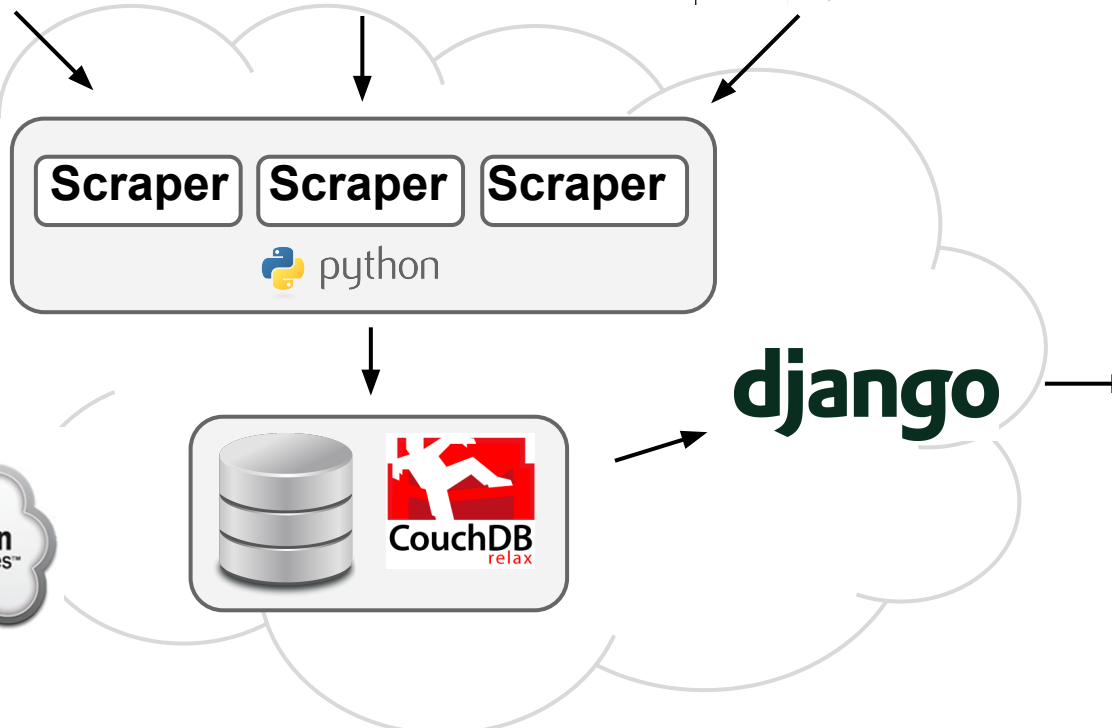
Karl Hedin Sånemyr

# Project Background

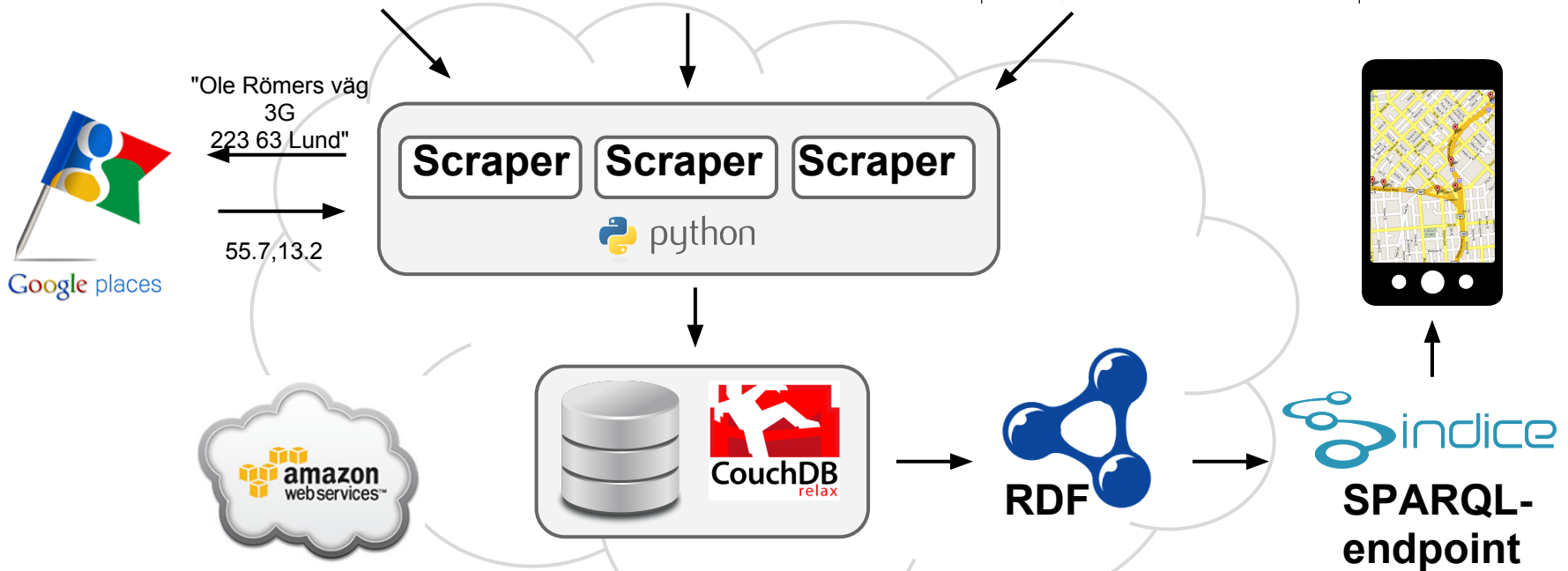
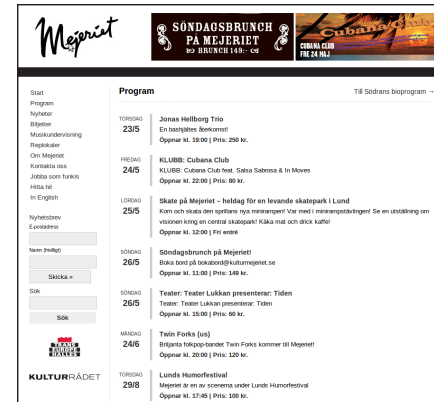
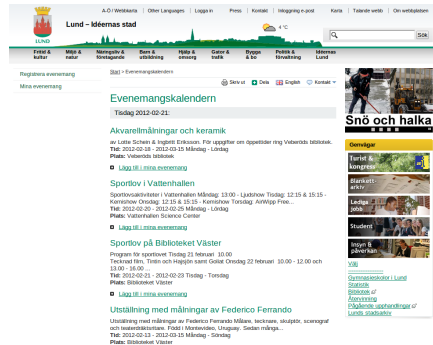


"Ole Römers väg  
3G  
223 63 Lund"

55.7, 13.2



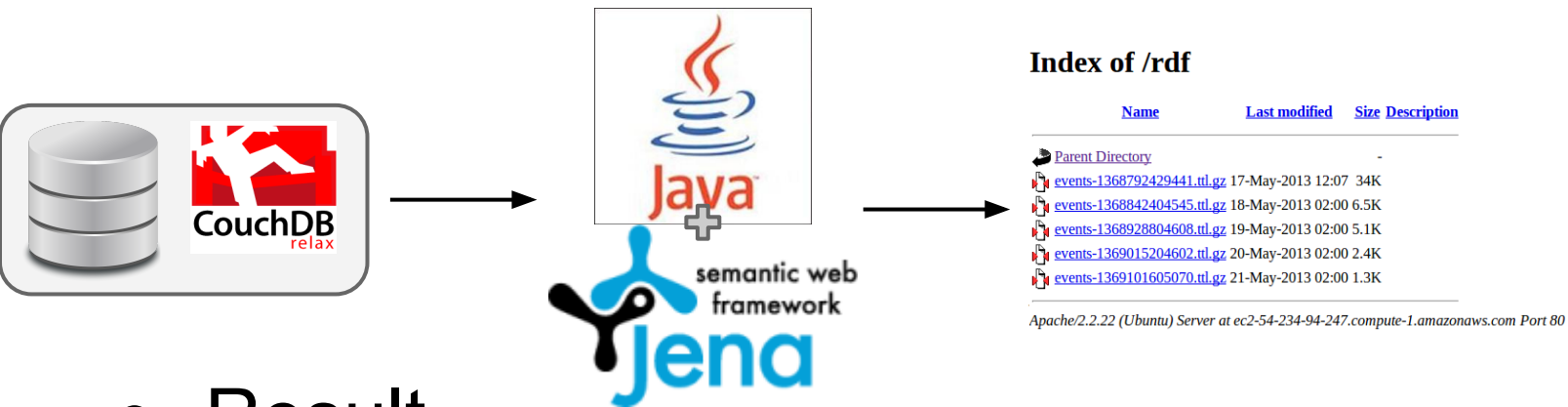
# Project Goal



# RDF Converter

- Plan:

- a. Dump scraper data
- b. Write a java program to format -> RDF
- c. Give RDF output to Sindice



- Result

- Learned a lot about rdf
- No real issues

# Scraper improvements

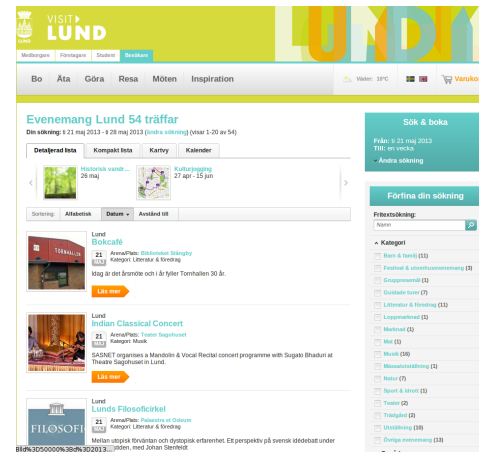
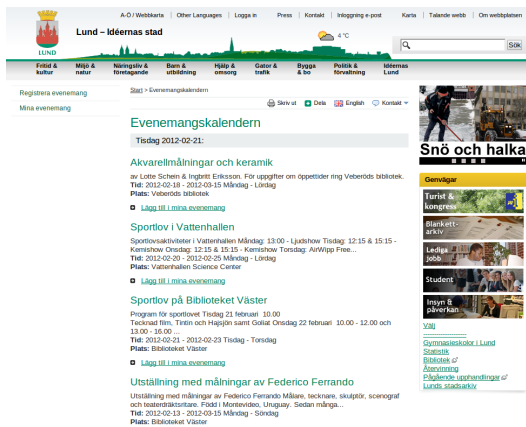
- Problems
  - Event descriptions short/missing
  - Geocoding not always successful (<30% success)

The image displays three screenshots of the 'DYGNET RUNT SE' website, illustrating the scraper's process. The first screenshot shows the main event listing for May 2013, with a blue circle highlighting the event 'MATHIEU PESQUE & ROLL PIGNALET'. A blue arrow points from this event to the second screenshot, which shows the detailed event description for 'MATHIEU PESQUE & ROLL PIGNALET'. Another blue arrow points from the event title in the second screenshot to the third screenshot, which shows the scraper's output for the event, including the event title, date, time, and location. The third screenshot also shows a map of the event location and a list of other events in the area.

- Results
  - Dygnetrunt now parses 100%, geocodes >80%

# Scraper improvements

- Problems
  - Lund.se complete redesign



- Results
  - Learned some Python!
  - New visitlund.se scraper

# SPARQL Endpoint

- Goal
  - Perform SPARQL queries on Sindice's servers.
- Problem
  - Delays @ Sindice



- Solution
  - Made our own SPARQL "endpoint" + "event search engine"

# Demo

[SPARQL endpoint](#)

[Search engine](#)