Semantic Scraping:
Semantic Scraping:
How to beat Sindice in 2 days for $20

Charalampos Chelmis, Gustavo Rocha, Luigi Selmi, Joshua Wulf
ESWC Summer School 2012
Google REST API
Google REST API
Sameas.org
Google  REST API

Sameas.org

Any23.org
Google REST API

Sameas.org

Any23.org
Cut and

Google  PASTE API

Some of

Sameas.org

Any23.org
Crete
Crete

Olympic Games
Crete
Olympic Games
Lady Gaga
Crete
Olympic Games
Lady Gaga
Barry Norton
Crete
Olympic Games
Lady Gaga
Barry Norton
Dan Brickley
Crete
Olympic Games
Lady Gaga
Barry Norton
Dan Brickley
Crete
Olympic Games
Lady Gaga
Barry Norton
Dan Brickley
Crete
Olympic Games
Lady Gaga
Barry Norton
Dan Brickley
How many (semantic) facts
How many (semantic) facts can we find
How many (semantic) facts can we find in Google Search results?
Scrape Google Search Results
Scrape Google Search Results

Extract Triples
Get URI
Get URI

Sameas.org
Get URI

Sameas.org

owl:sameAs on store
Is URI?

- yes:
  - Acquire URIs set from sameas.org
  - Google Search for URIs
- no:
  - Invoke Google Search
  - Triplify Results Using Any23.org

- Local Repository
- yes:
  - Semantic Data Present?
  - Local Repository

- Semantic Data Present?
  - yes:
    - Local Repository
Fact Extraction
URI as Subject

\[ \text{select ?x {URI ?p ?x.}} \]

\[ \text{select ?x {?x ?p URI.}} \]
URI as Subject

URI as Object
URI as Subject

select ?x {URI ?p ?x.}

URI as Object
URI as Subject

```
select ?x {URI ?p ?x.}
```
Crete
Lady Gaga
Lady Gaga
Barry Norton
Dan Brickley
Dan Brickley

![Bar Chart]

- **Subject**
- **Object**
- **[Legend](#)**

- **Axes:**
  - **X-axis:** 100, 200, 300, 400, 500
  - **Y-axis:** 0, 20, 40, 60, 80, 100, 120, 140, 160

- **Bars:**
  - **100:** Subject: 120, Object: 50
  - **200:** Subject: 140, Object: 60
  - **300:** Subject: 150, Object: 80
Thanks:  
Denny  
Barry  
Aidan