

Semantic Scraping:

Charalampos Chelmis, Gustavo Rocha, Luigi Selmi, Joshua
Wulf
ESWC Summer School 2012

Semantic Scraping:

How to beat Sindice in 2 days for \$20

Charalampos Chelmiss, Gustavo Rocha, Luigi Selmi, Joshua
Wulf
ESWC Summer School 2012



lady gaga



Search

About 415,000,000 results (0.25 seconds)

Everything

Images

Maps

Videos

News

Shopping

Blogs

More

Search near...

Enter location

Set

The web

Pages from Australia

More search tools

News for lady gaga



[Lady Gaga facing Thailand backlash after fake Rolex joke](#)

Telegraph.co.uk - 3 hours ago

Lady Gaga's provocative lyrics and costumes have angered Catholic groups in South Korea and the Philippines and Islamists in Indonesia, ...

[Lady Gaga refuses to tone down her shows](#)

Capital FM Kenya - 5 hours ago

[Lady Gaga angers Thai fans with fake Rolex comment](#)

Fox News - 5 hours ago

[Lady Gaga - Wikipedia, the free encyclopedia](#)

en.wikipedia.org/wiki/Lady_Gaga - Cached

Stefani Joanne Angelina Germanotta known by her stage name **Lady Gaga**, is an American singer and songwriter. Born and raised in New York City, she ...

↳ [Discography](#) - [Born This Way](#) - [The Fame](#) - [Born This Way Ball](#)

You visited this page on 23/05/12.

[Lady Gaga : Merchandise](#)

www.ladygaga.com.au/ - Cached

LADY GAGA - THE BORN THIS WAY BALL RECORDS MULTIPLE SELL OUT SHOWS IN AUCKLAND, BRISBANE, SYDNEY, AND MELBOURNE! DUE TO ...

[LADY GAGA | BORN THIS WAY BALL](#)

www.ladygaga.com/ - Cached

American pop singer and songwriter. Includes biography, tour information, videos, ringtones, photos and a forum.

[Lady Gaga : News](#)

www.ladygaga.com/news/ - Cached



by Lady Gaga · in 2,057,025 Google+ circles

6 days ago – Lady Gaga will guest star on The Simpson's season finale this Sunday, May 20th at 8/7c on Fox. When Gaga discovers that the city of ...

Images for lady gaga · Report images

Google REST API

Google REST API

Sameas.org

Google REST API

Sameas.org

Any23.org

~~Google REST API~~

Sameas.org

Any23.org

~~Google REST API~~

~~Sameas.org~~

Any23.org

Equivalent URIs for http://dbpedia.org/resource/Lady_Gaga –

1. http://dbpedia.org/resource/Ga_Ga
2. <http://dbpedia.org/resource/Hitmixes>
3. http://dbpedia.org/resource/LADY_GAGA
4. http://dbpedia.org/resource/Lady_GaGa
5. http://dbpedia.org/resource/Lady_Gaga
6. http://dbpedia.org/resource/Lady_gaga
7. <http://dbpedia.org/resource/Germanotta>
8. http://dbpedia.org/resource/Lady_Ga_Ga
9. http://dbpedia.org/resource/Haus_of_GaGa
10. http://dbpedia.org/resource/Haus_of_Gaga
11. http://dbpedia.org/resource/Haus_of_gaga
12. http://dbpedia.org/resource/Electric_Kiss
13. http://dbpedia.org/resource/Mother_Monster
14. http://dbpedia.org/resource/Glitter_&_Grease
15. http://dbpedia.org/resource/Stefani_Germanotta
16. [http://dbpedia.org/resource/Lady_GaGa_\(band\)](http://dbpedia.org/resource/Lady_GaGa_(band))
17. [http://dbpedia.org/resource/Fancy_pants\(song\)](http://dbpedia.org/resource/Fancy_pants(song))
18. [http://dbpedia.org/resource/Disco_Heaven_\(album\)](http://dbpedia.org/resource/Disco_Heaven_(album))
19. http://dbpedia.org/resource/Joanne_Stefani_Germanotta
20. http://dbpedia.org/resource/Stefani_Joanne_Germanotta
21. [http://dbpedia.org/resource/Hitmixes_\(Lady_Gaga_EP\)](http://dbpedia.org/resource/Hitmixes_(Lady_Gaga_EP))
22. [http://dbpedia.org/resource/Red_and_Blue_\(Lady_Gaga_EP\)](http://dbpedia.org/resource/Red_and_Blue_(Lady_Gaga_EP))
23. [http://dbpedia.org/resource/Red_and_Blue_\(Lady_Gaga_ep\)](http://dbpedia.org/resource/Red_and_Blue_(Lady_Gaga_ep))
24. [http://dbpedia.org/resource/Born_This_Way_\(Lady_Gaga_song\)](http://dbpedia.org/resource/Born_This_Way_(Lady_Gaga_song))
25. http://dbpedia.org/resource/Stefani_Joanne_Angelina_Germanotta
26. <http://bnb.data.bl.uk/id/person/LadyGaga>
27. <http://d-nb.info/gnd/135898838>
28. <http://data.bibsys.no/data/notrbib/authorityentry/x09060029>
29. <http://dbpedialite.org/things/17782843#id>
30. <http://dbpedialite.org/things/17854475#id>
31. <http://dbpedialite.org/things/19777829#id>
32. <http://dbpedialite.org/things/20086410#id>
33. <http://dbpedialite.org/things/21177182#id>
34. <http://dbpedialite.org/things/21780425#id>
35. <http://dbpedialite.org/things/21867575#id>
36. <http://dbpedialite.org/things/22090093#id>
37. <http://dbpedialite.org/things/24370817#id>
38. <http://dbpedialite.org/things/25776596#id>
39. <http://dbpedialite.org/things/26192320#id>
40. <http://dbpedialite.org/things/26194136#id>

<sameAs>

~~Google REST API~~

Sameas.org

Any23.org

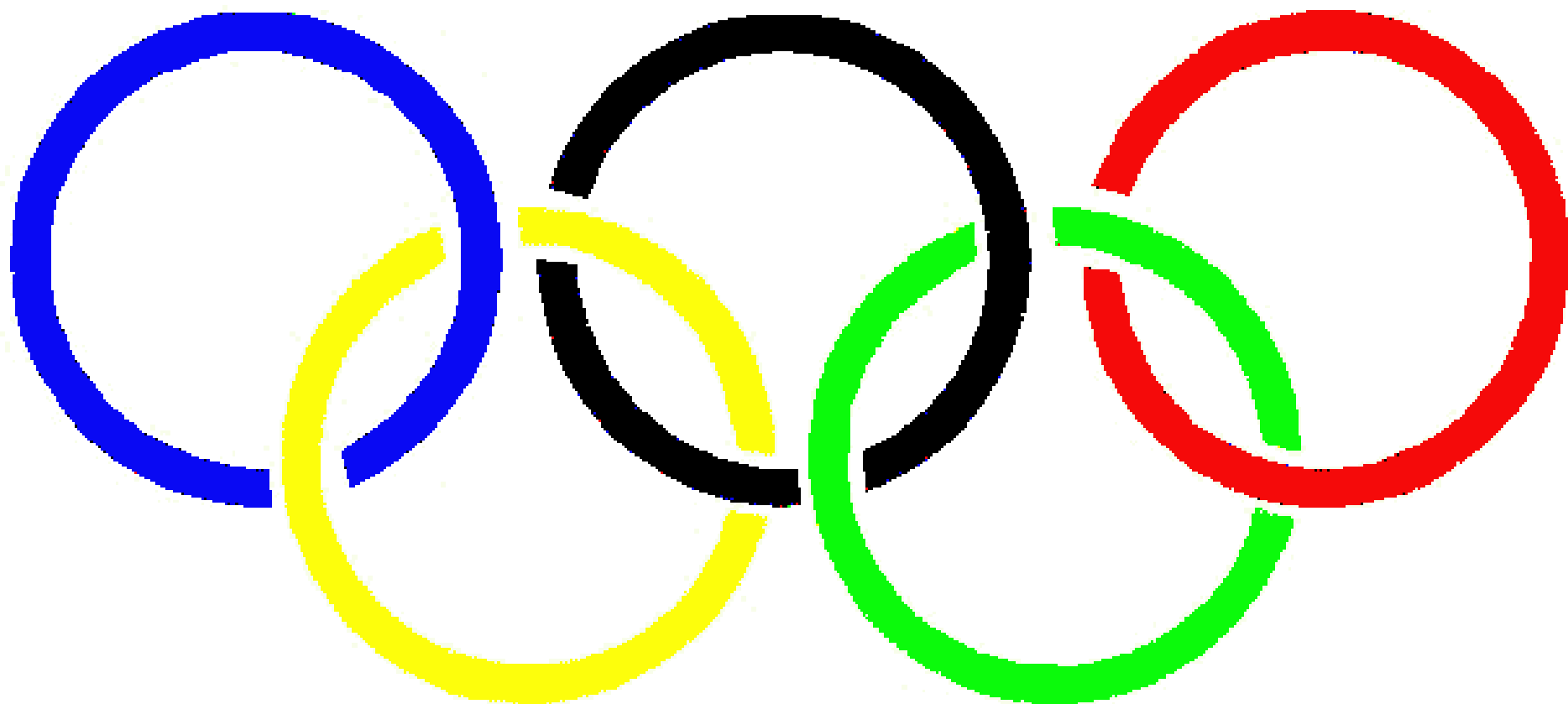
Google ^{Cut and} PASTE API

Some of

Sameas.org

Any23.org





GAGA

JUST DANCE

NOTHING
ELSE I
CAN SAY

LOVE GAME

POKERFACE

GAGA



LADY GAGA









Crete

Crete

Olympic Games

Crete

Olympic Games

Lady Gaga

Crete

Olympic Games

Lady Gaga

Barry Norton

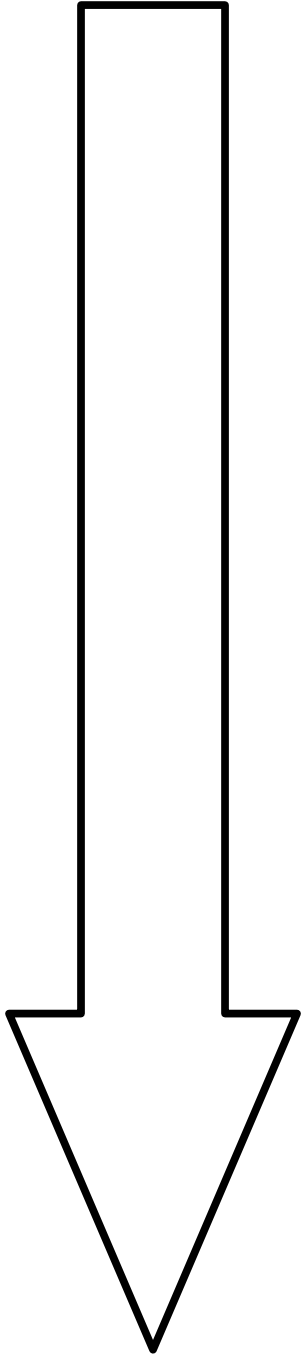
Crete

Olympic Games

Lady Gaga

Barry Norton

Dan Brickley



Crete

Olympic Games

Lady Gaga

Barry Norton

Dan Brickley



100

Crete

Olympic Games

Lady Gaga

Barry Norton

Dan Brickley



100

Crete

200

Olympic Games

Lady Gaga

Barry Norton

Dan Brickley



100

Crete

200

Olympic Games

300

Lady Gaga

Barry Norton

Dan Brickley



100

Crete

200

Olympic Games

300

Lady Gaga

400

Barry Norton

Dan Brickley



100

Crete

200

Olympic Games

300

Lady Gaga

400

Barry Norton

500

Dan Brickley

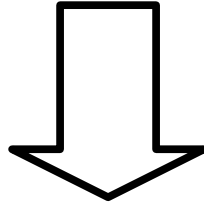
How many (semantic) facts

How many (semantic) facts
can we find

How many (semantic) facts
can we find
in Google Search results?

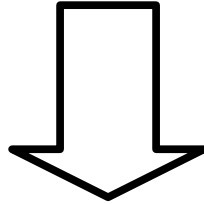
Scrape Google Search Results

Scrape Google Search Results

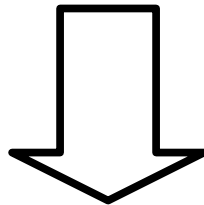


Extract Triples

Scrape Google Search Results



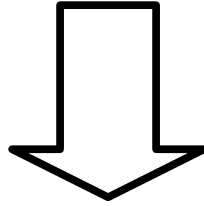
Extract Triples



Upload to Store

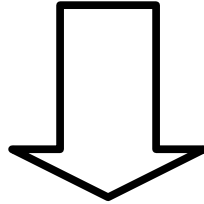
Get URI

Get URI

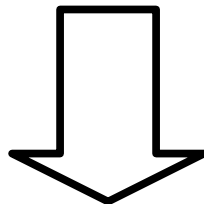


Sameas.org

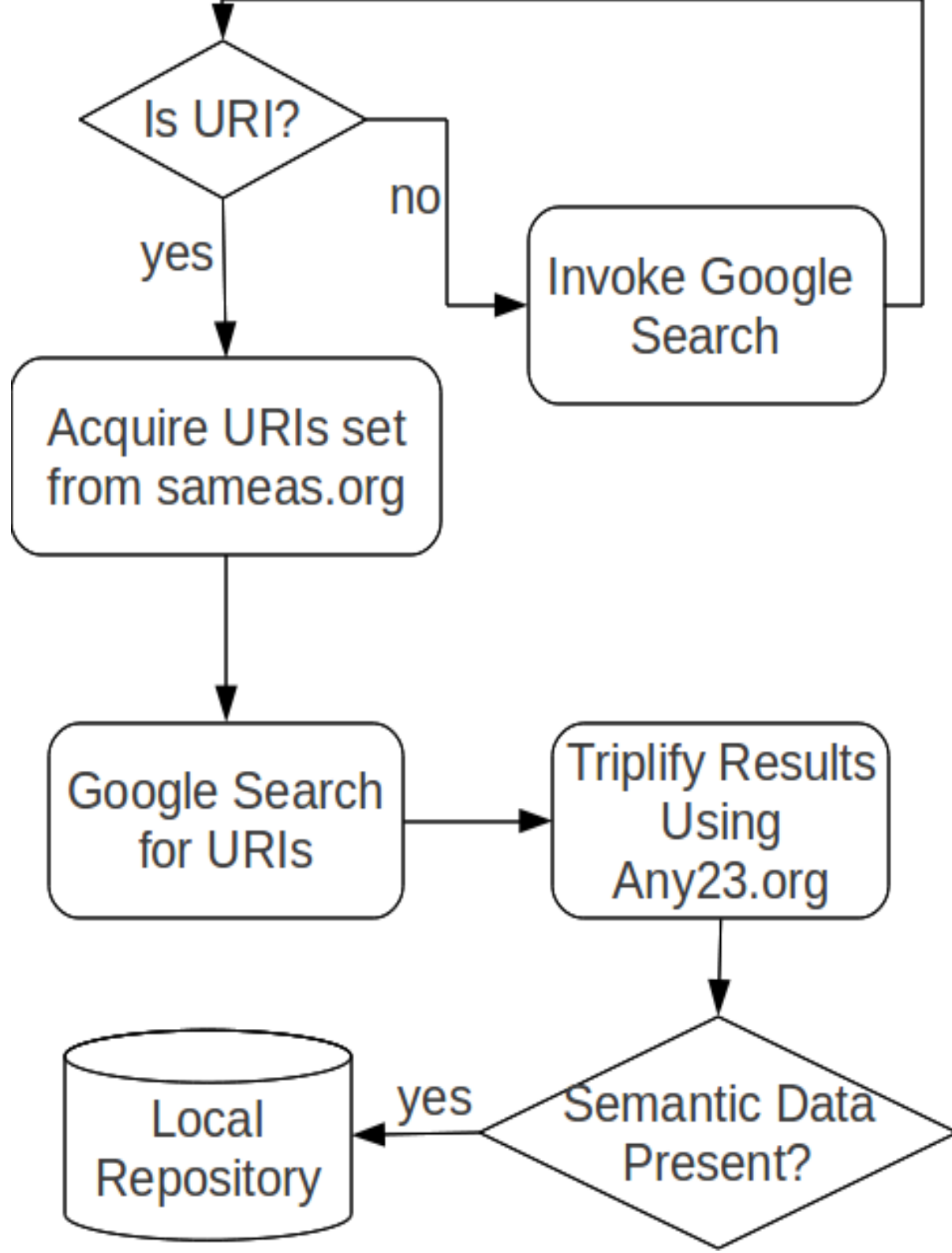
Get URI



Sameas.org



owl:sameAs on store



Fact Extraction

URI as Subject

URI as Subject

URI as Object

URI as Subject

```
select ?x {URI ?p ?x.}
```

URI as Object

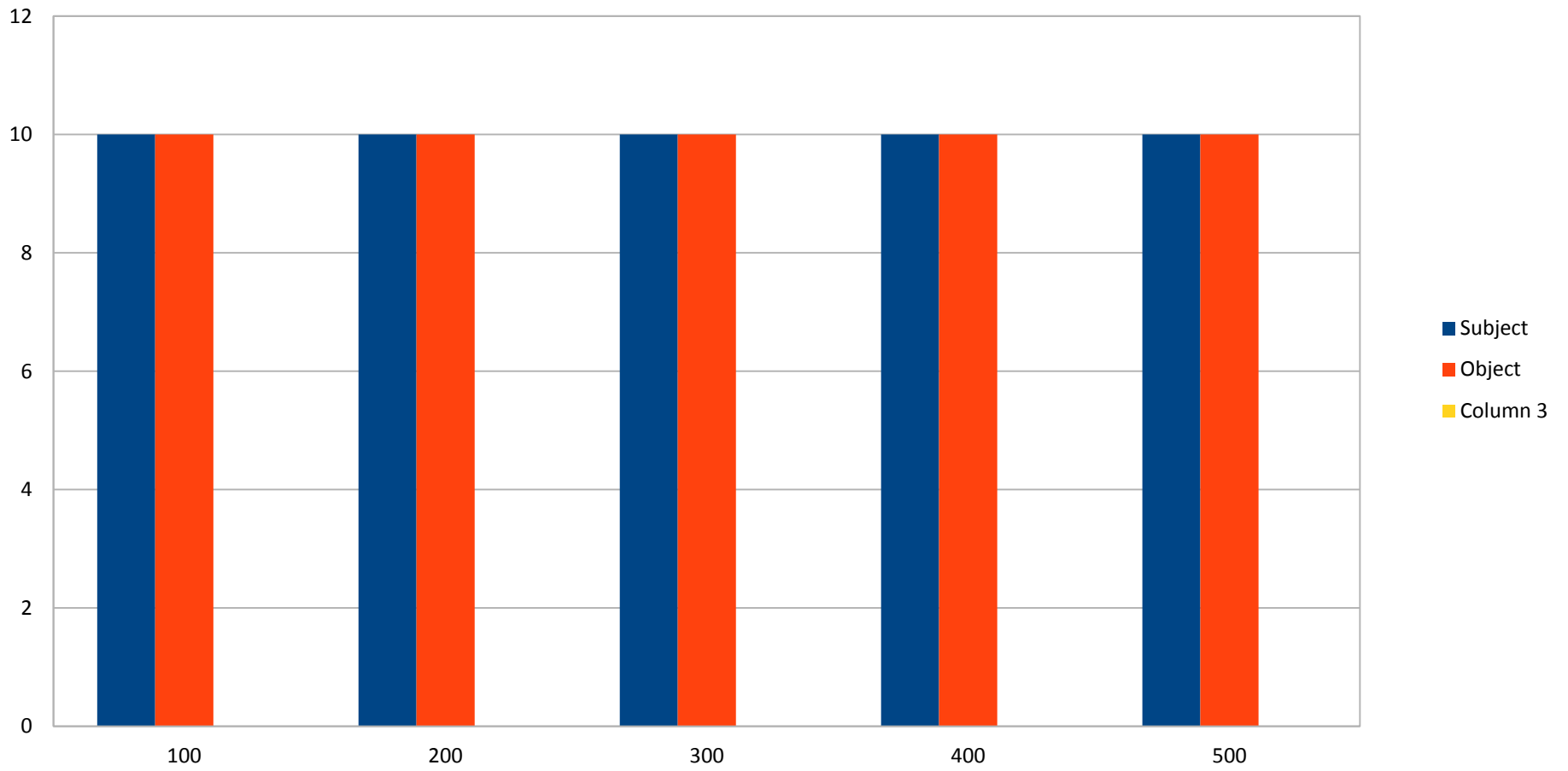
URI as Subject

```
select ?x {URI ?p ?x.}
```

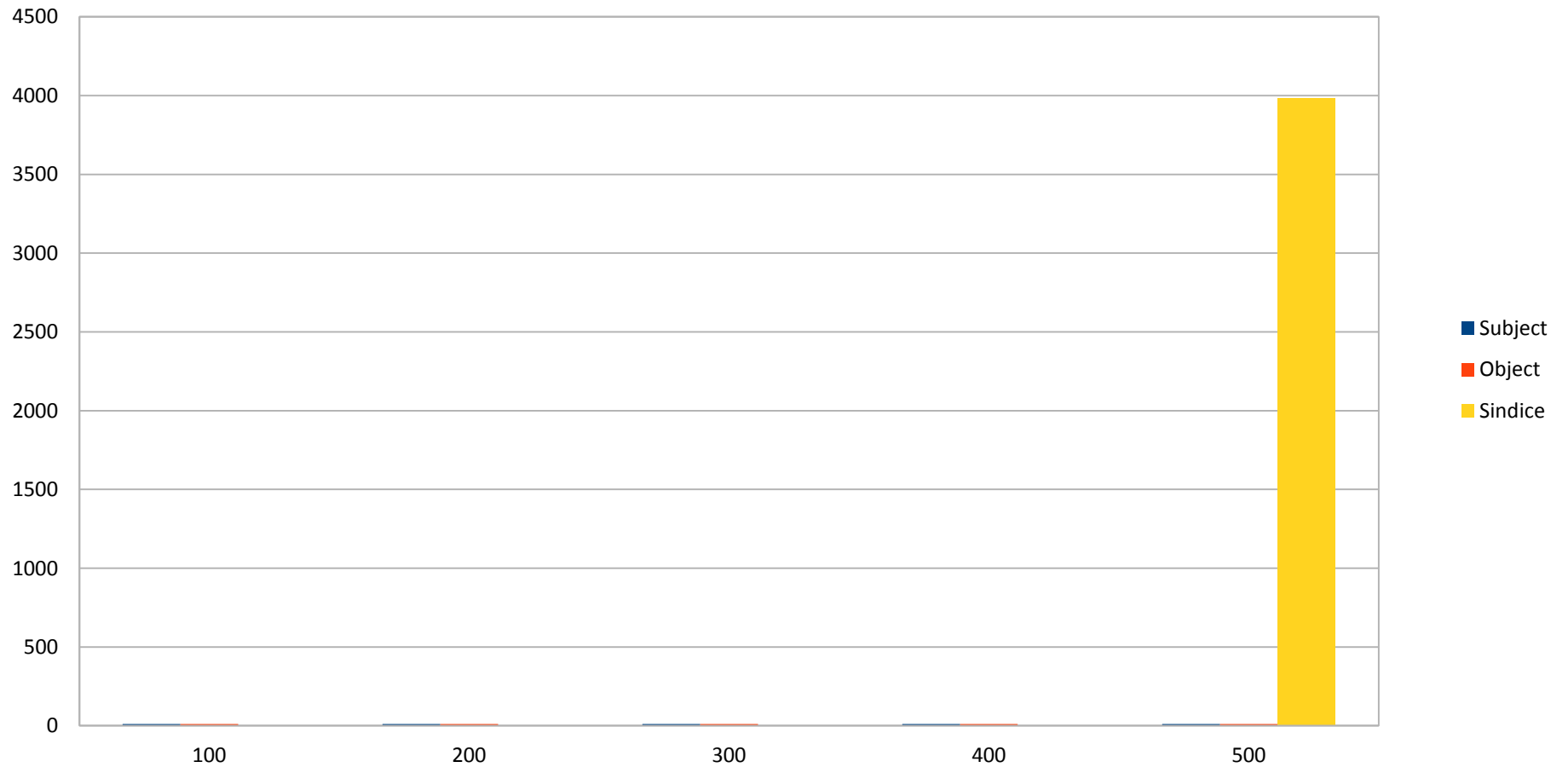
URI as Object

```
select ?x {?x ?p URI.}
```

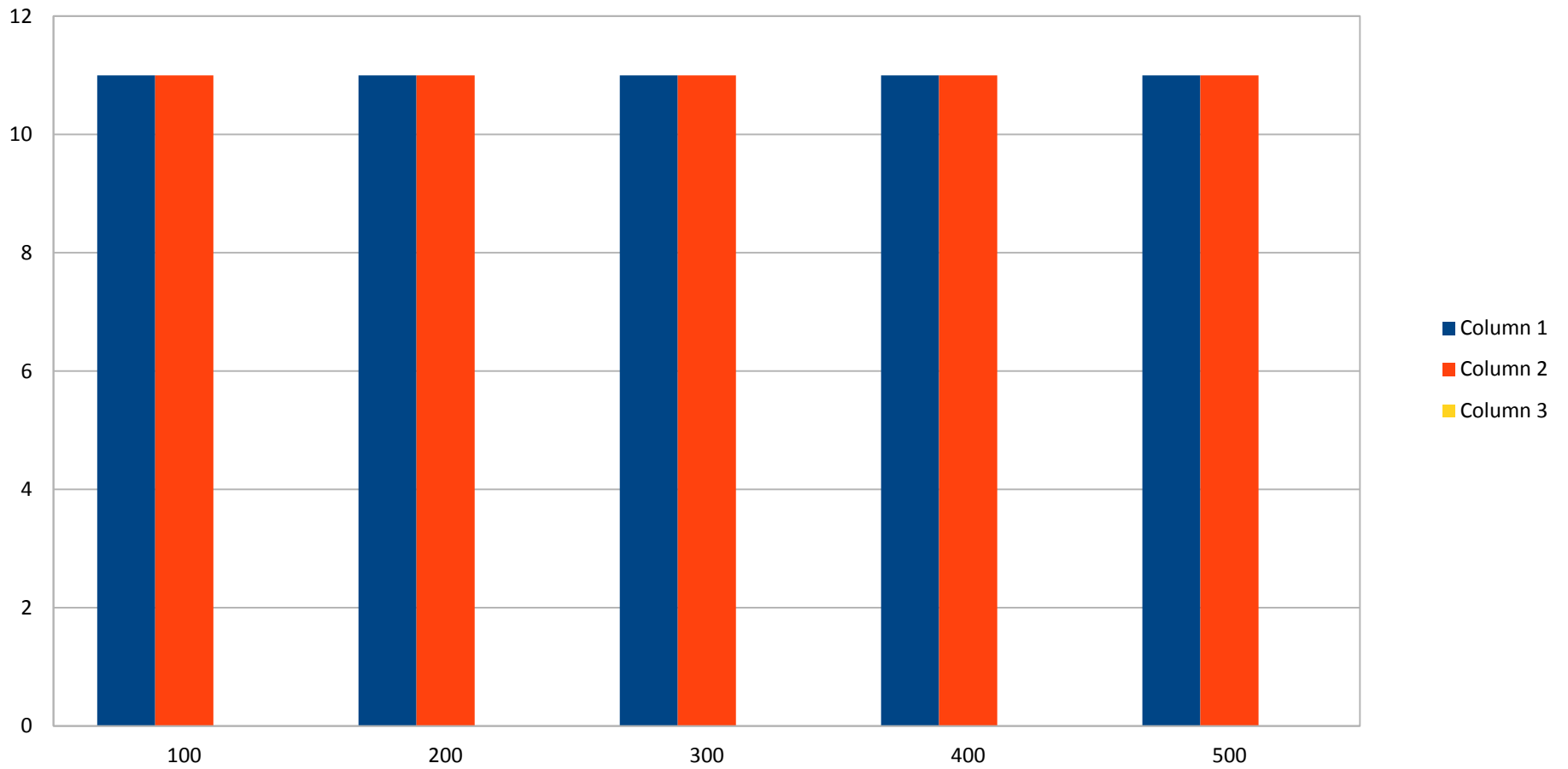
Crete



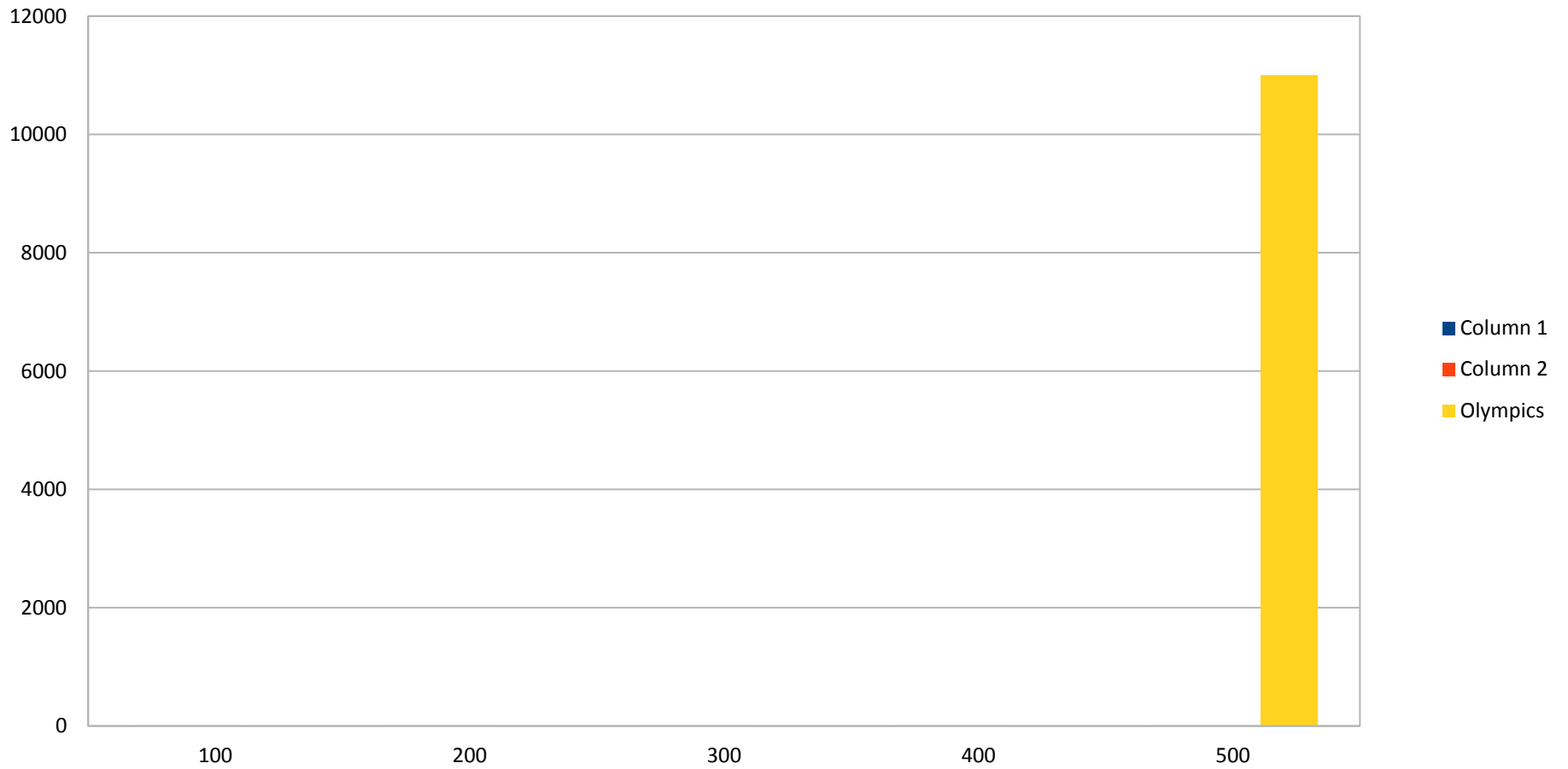
Crete



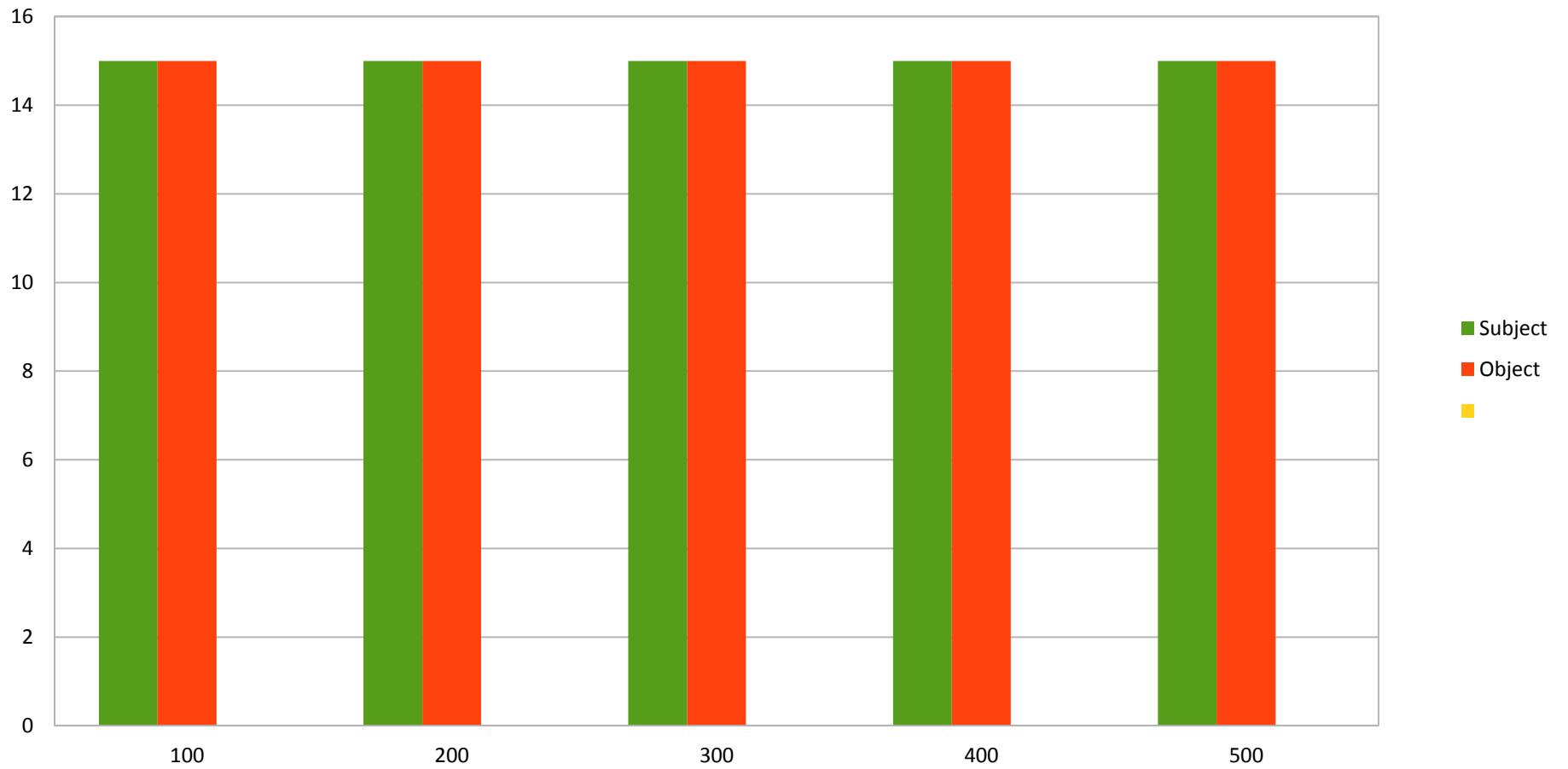
Olympics



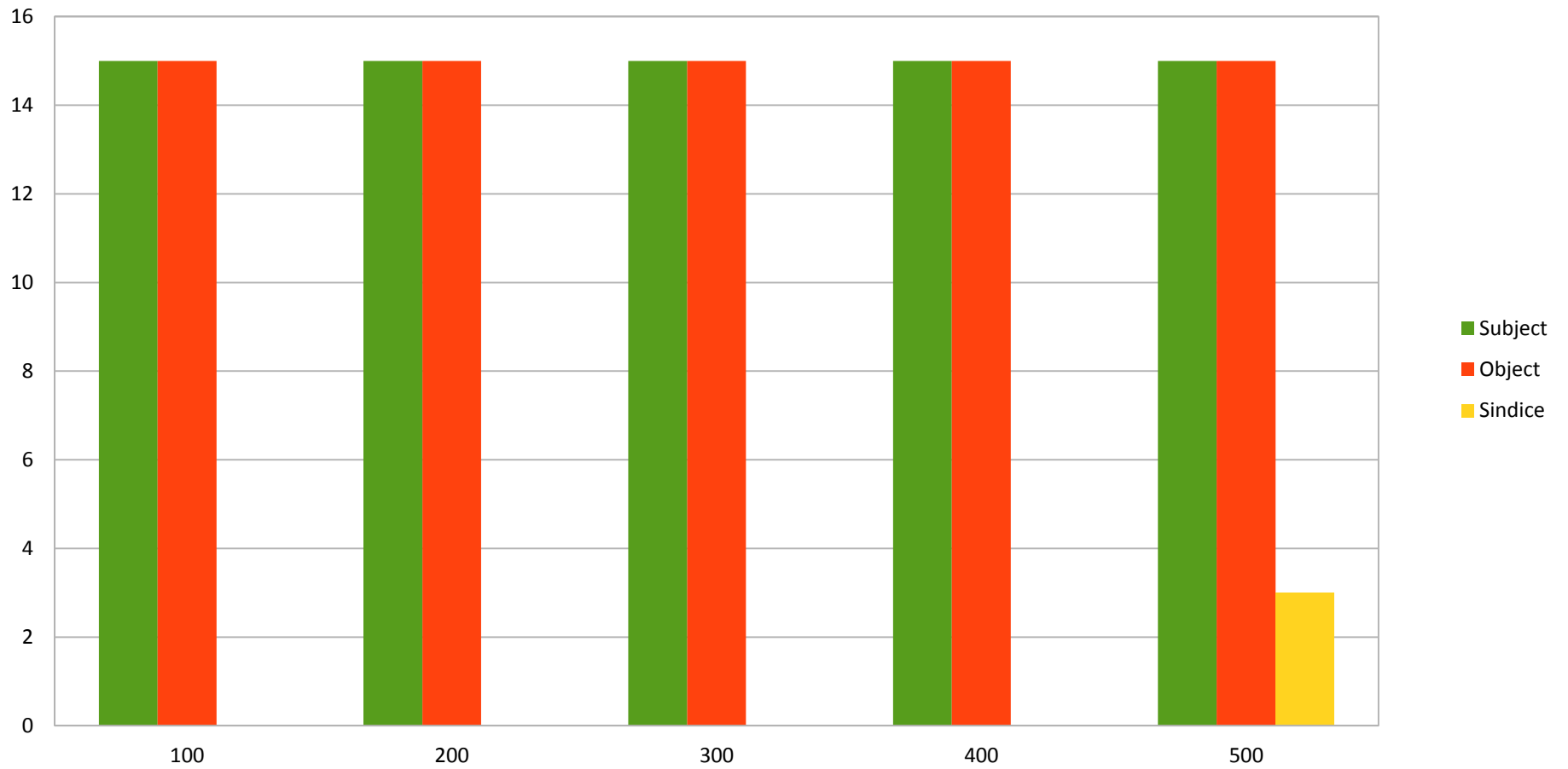
Olympics



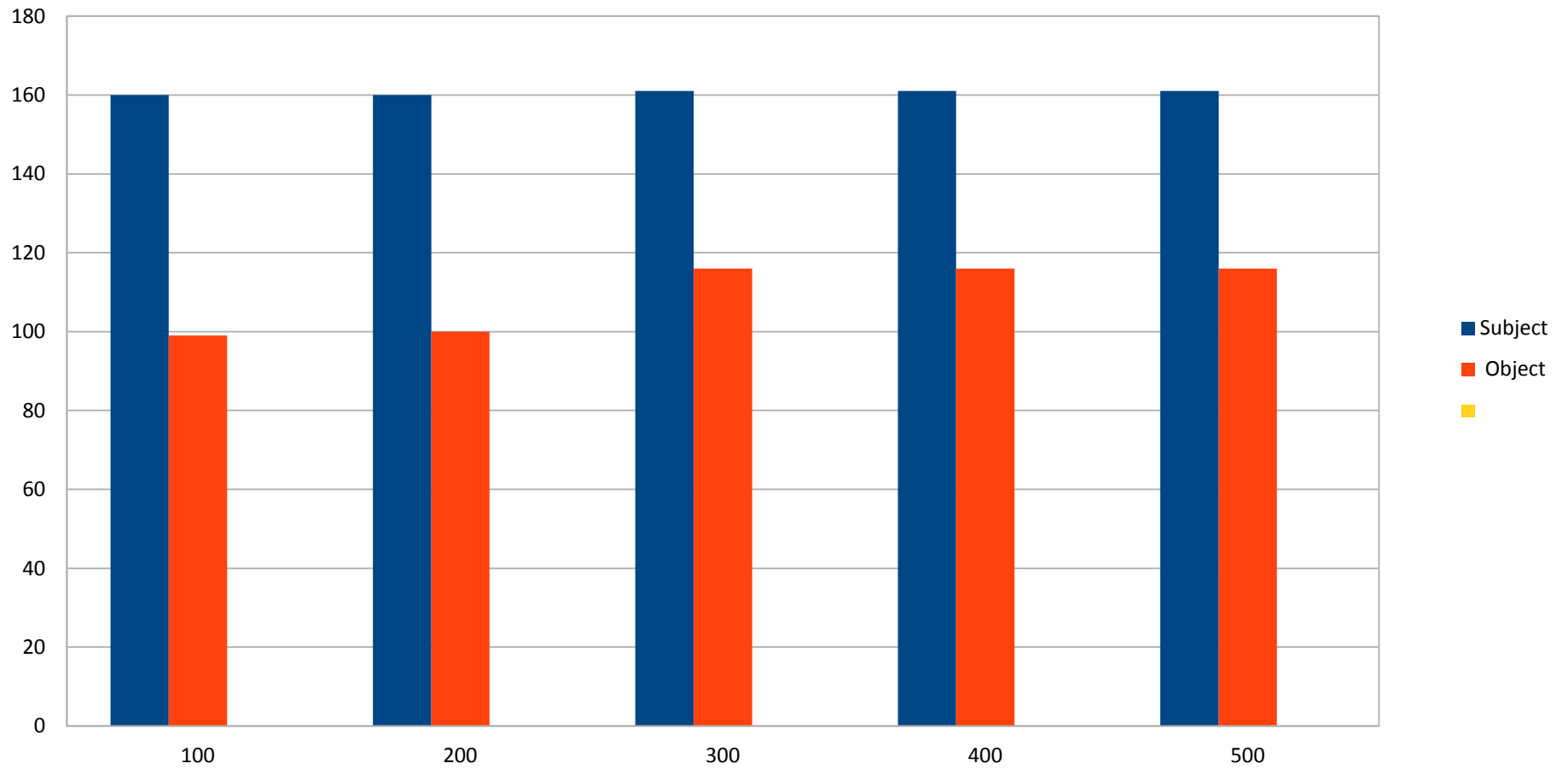
Lady Gaga



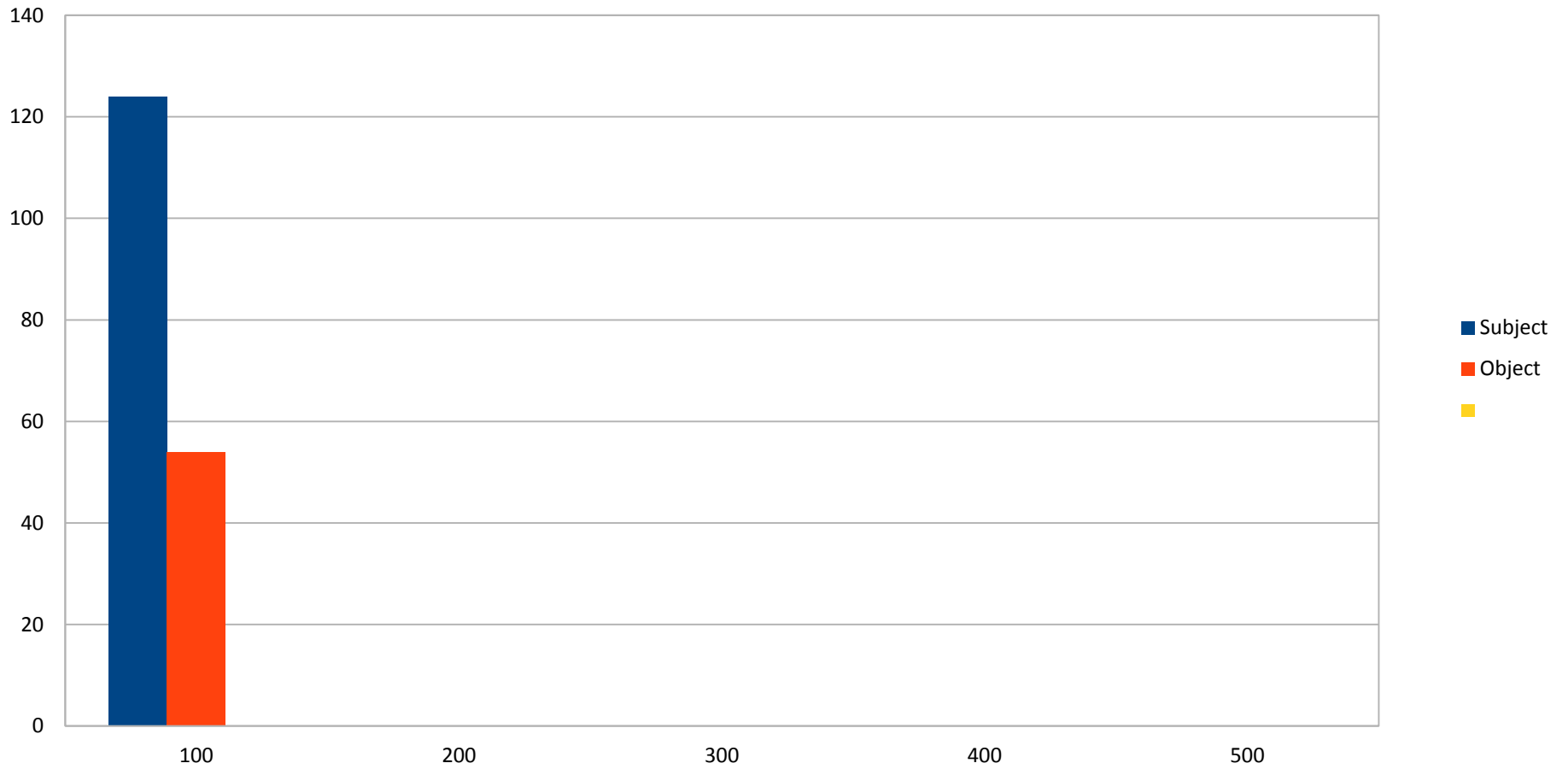
Lady Gaga



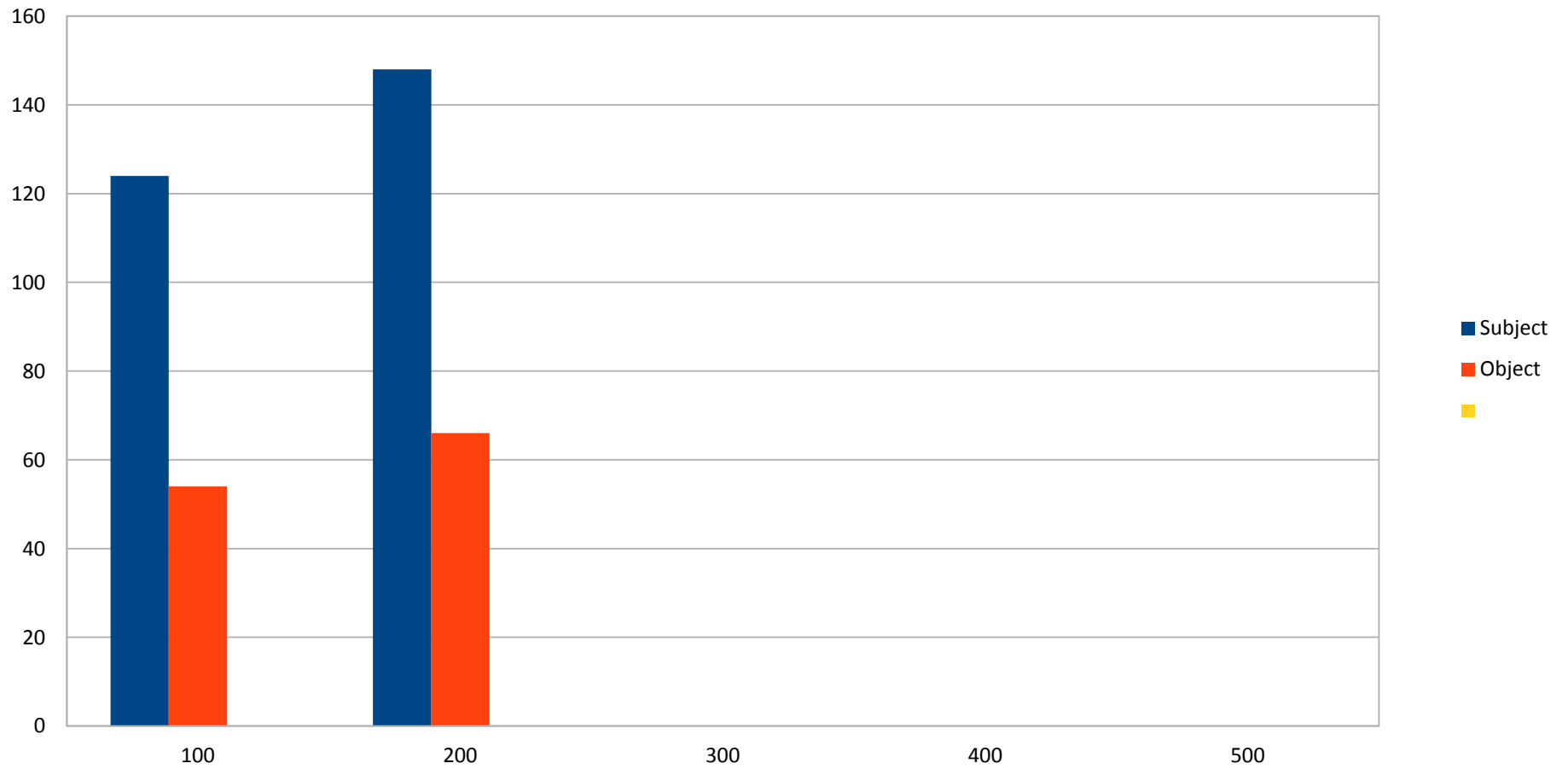
Barry Norton



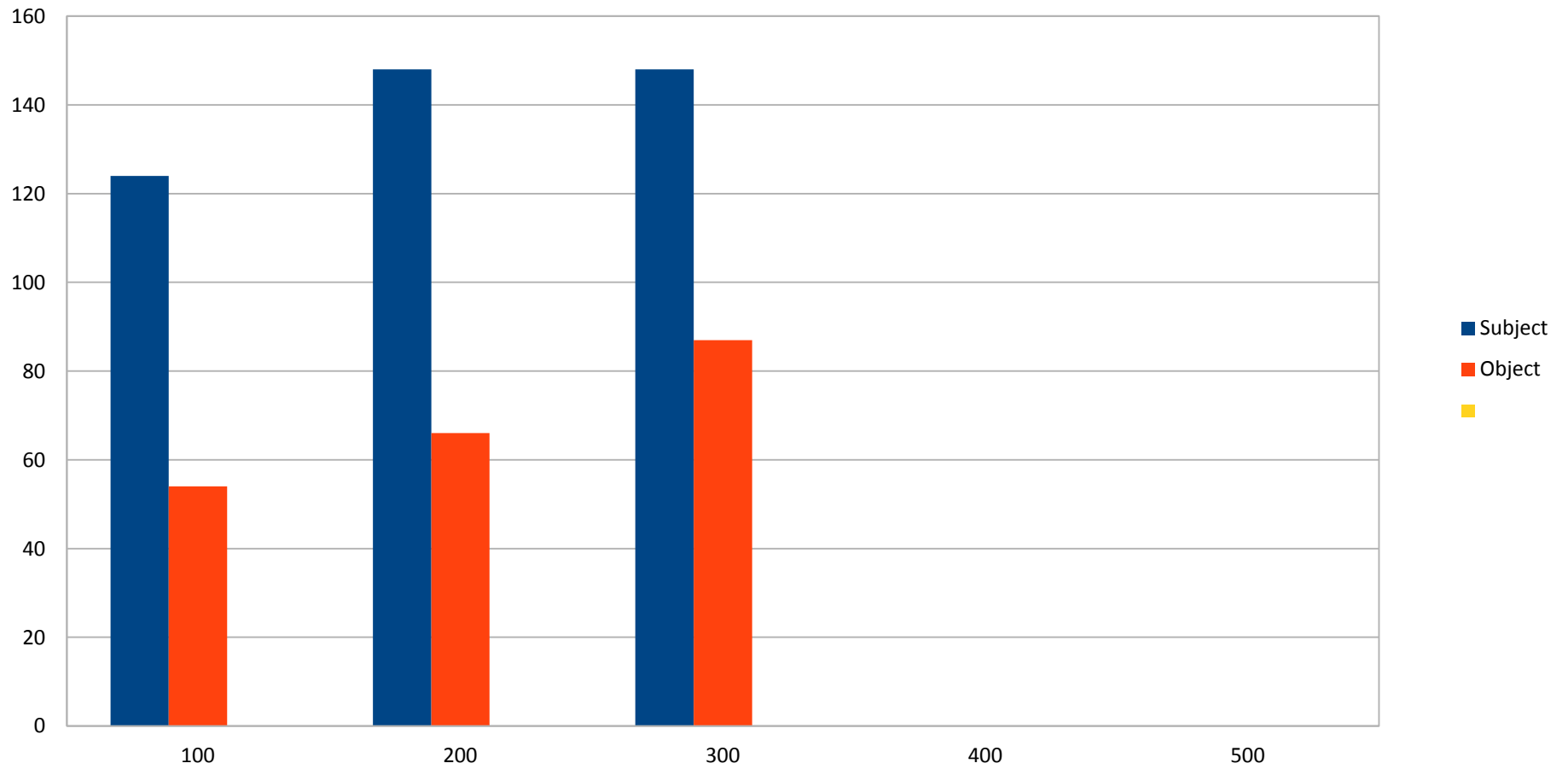
Dan Brickley



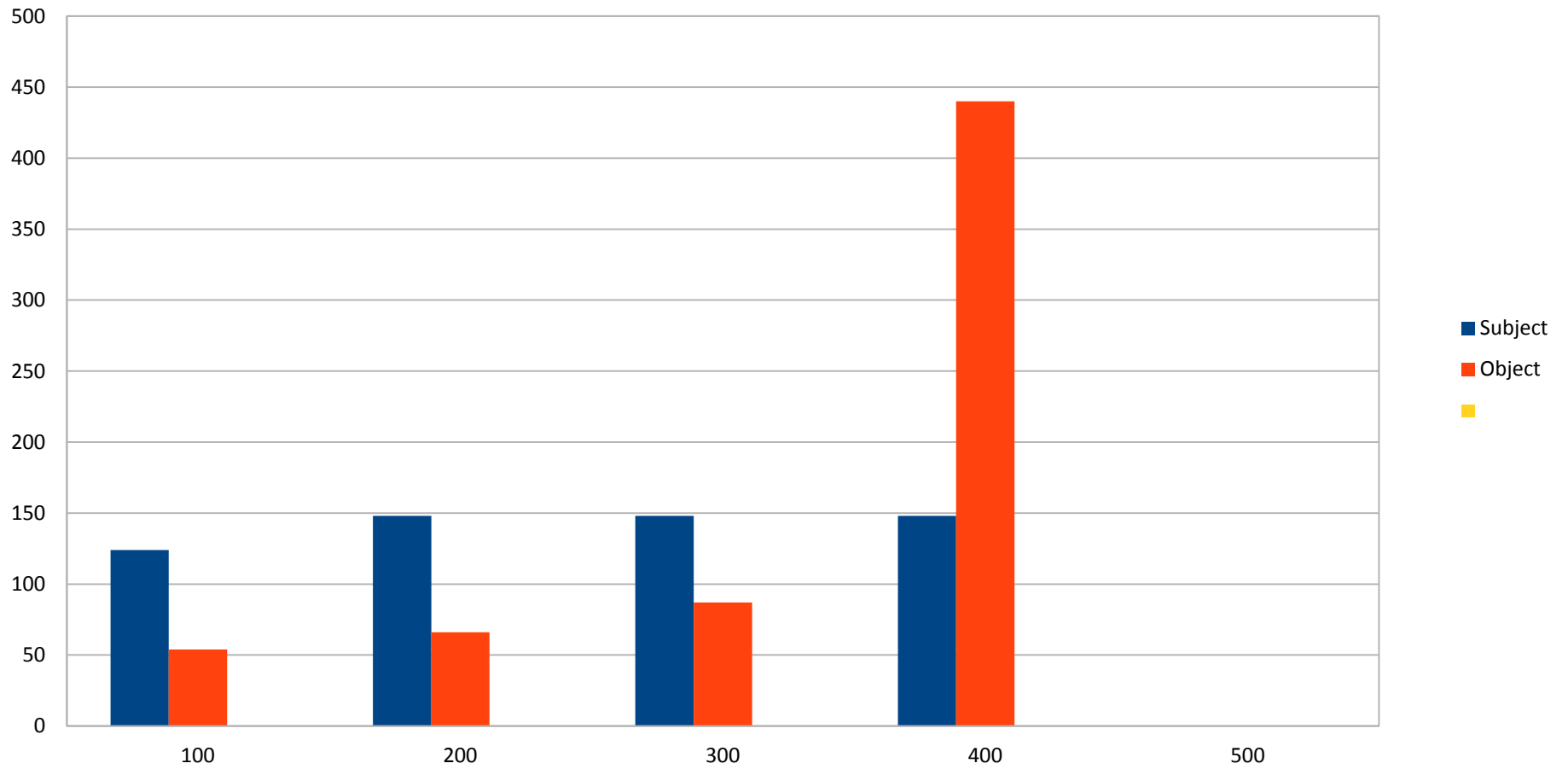
Dan Brickley



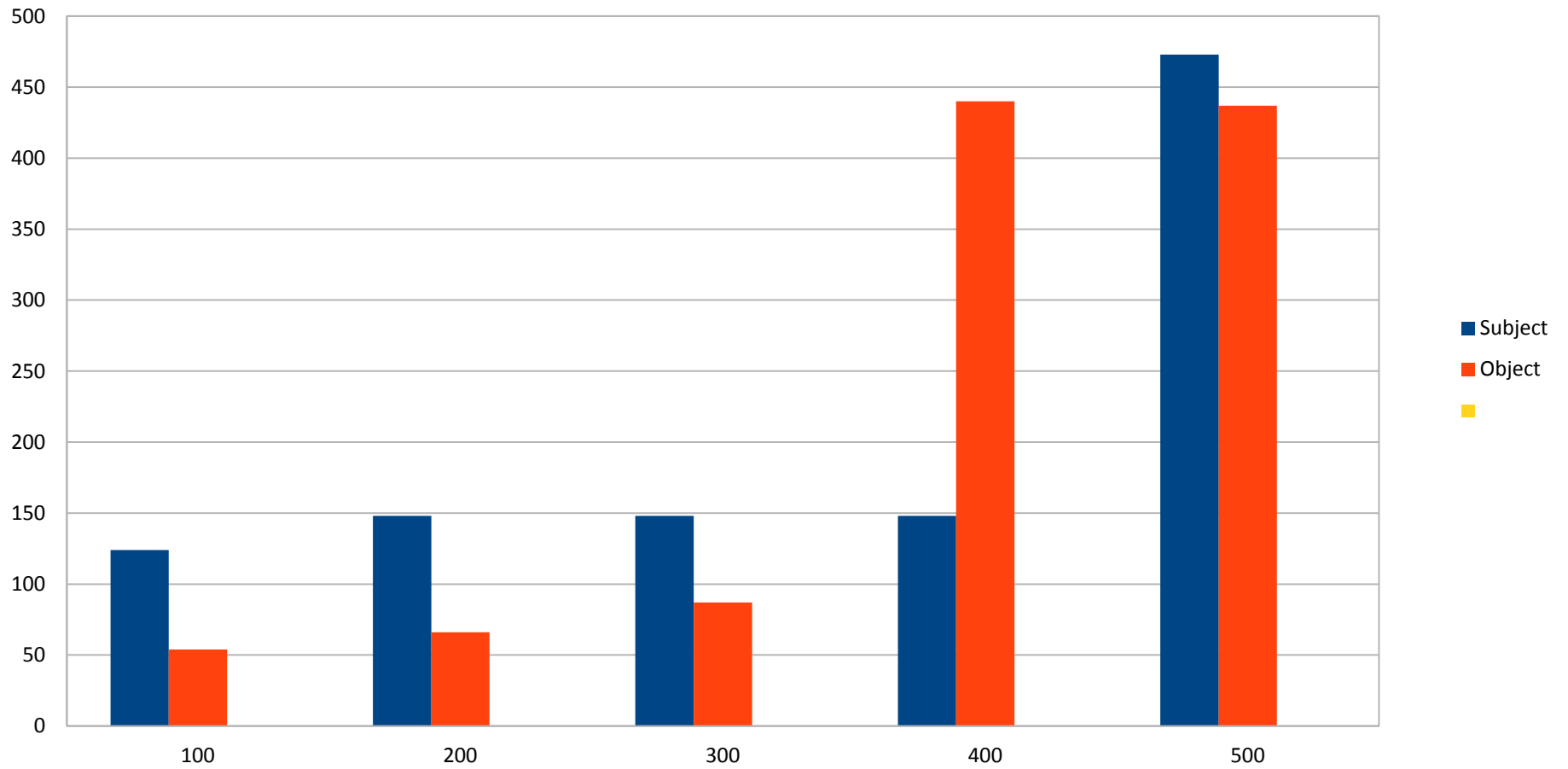
Dan Brickley



Dan Brickley



Dan Brickley





Thanks:
Denny
Barry
Aidan