

# Google Web APIs an Instrument for Webometric Analyses?

## Basics

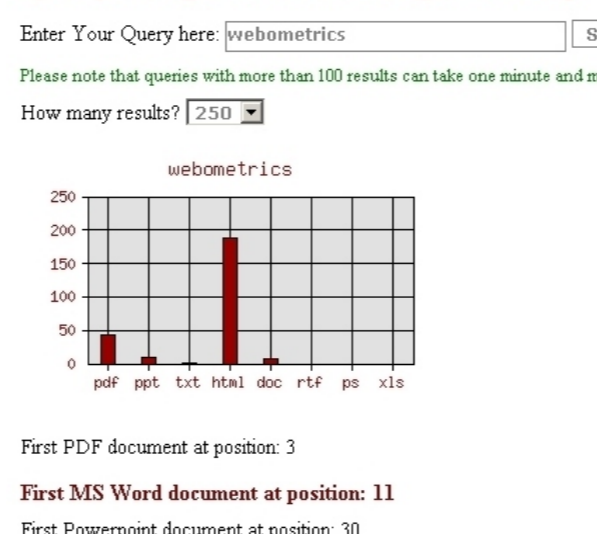
- Launch spring 2002, but still Beta
- Googlekey with 10.000 hits per day
- Up to 1.000 hits in one automated query
- Implemented as a SOAP Web Service / WSDL
- Supported by all common programming lang.

## Live Demos

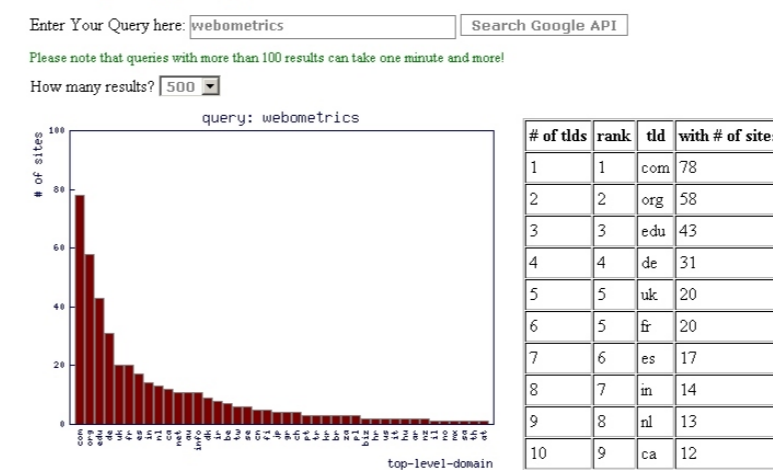
- File Format analysis
- TLD distribution
- Google vs. Google

### Analysing document file types

#### Query Google with the Google Web API



### Analysing Top Level Domains

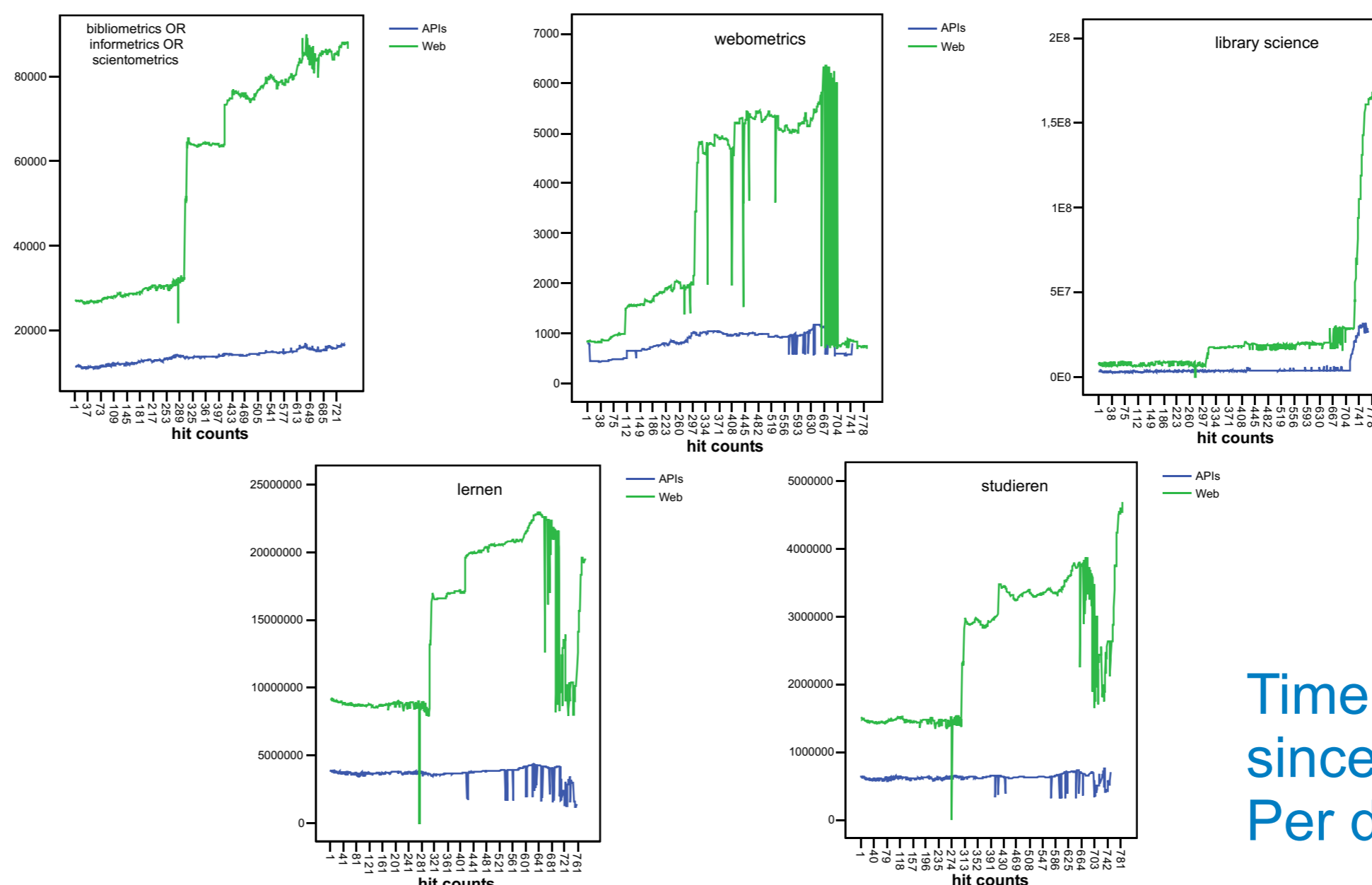


### Yahoo! Search WS vs. Google Web APIs

Yahoo! Search Web Services	Google Web APIs
Total: 1870	Total: 1170
1. Webometrics and the SOERs <a href="http://www.soei.com/webometrics/">http://www.soei.com/webometrics/</a>	1. Webometrics and the SOERs <a href="http://www.soei.com/webometrics/">http://www.soei.com/webometrics/</a>
2. webometrics <a href="http://www.de.de/2007/07/20/webometrics-43272.doc">http://www.de.de/2007/07/20/webometrics-43272.doc</a>	2. Webometrics PhD thesis [excerpt chapter 2: webometrics] Sm... <a href="http://www.de.de/2007/07/20/webometrics-43272.doc">http://www.de.de/2007/07/20/webometrics-43272.doc</a>
3. Webometrics <a href="http://webometrics.blogspot.com/">http://webometrics.blogspot.com/</a>	3. Webometrics - Wikipedia, the free encyclopedia <a href="http://de.wikipedia.org/wiki/Webometrics">http://de.wikipedia.org/wiki/Webometrics</a>

## Analyses

- Time Series
- Hit results are Juxtaposed



Time series are running since July 2004, 3 times Per day.

## Results

- Hit results from APIs and Web interface vary in range, structure and availability; 2 different indexes
- Google Web Search comes up with more results, larger fluctuations and different URLs
- APIs data show ups and downs later / less updated index
- Document ranking is not the same / APIs is slower

## Conclusions

- Few obstacles, but quick implementation
- Interesting for webometric prototyping
- It has to be clear that Google Web Search is very different to Google APIs

### Google Web APIs - Project Page

The project page has been set up after our presentation at the Laugen Nacht der Wissenschaften (12 July 2004). The site introduces Google Web APIs (Google API) as an instrument and program for webometric analysis. First results of our work with the Google Web APIs appear in a special issue of Informationswissenschaft Forum (IFF) in 2005. Some examples and demo implementations of the Google APIs are available below.

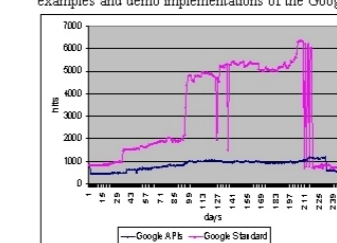


Figure 1: The query webometrics. Hit results in a time series from Google.com and Google Web APIs.

### Google Web APIs demo programs

1. Single API Query (Simple) - a very simple program to demonstrate the Google Web APIs [author: Paul Dondos - DCS-Vertrag-Fabrik-Transfer]
2. Google Versus - the following Google searches show how inadequately Web Services results can be adapted for an output.
  - = Google (URL) - only URL output
  - = Google (URL + Title) - URL output + title-tag
  - = Google (URL + Title + Support) - URL output + title-tag + support
  - = Google (URL + Title + Support + Date) - URL output + title-tag + support + the size of the hit
3. Analysis of file formats - a program that analyzes the file format (html, ppt, pdf, doc, rtf, ps, xls) in result sets from the Google APIs results with a graphic and tabular output. The positions of the first pdf, ppt, doc, rtf, ps, xls documents (alternative file types) are displayed. [concept: Philipp Meyer - programming: Fabian Trappert]
4. Analysis of Top Level Domains (TLD) - a program that analyzes top-level domains (e.g., com, edu, org, de, ca, ...) in result sets from the Google APIs results with a graphic and tabular output. [concept: Philipp Meyer - programming: Fabian Trappert]
5. Calculation of ranking scores - a program that performs a permutation of rank. 4 query terms and calculates afterwards a ranking score for the results of the query combinations (first).