

# An excursus to Google Scholar

## How deep does this search engine dig?

### Background

- Launch November 2004, Beta
- Interesting approach: Autonomous Citation Indexing (ACI) and citation analysis
- Size, coverage and actuality are unknown

### Study

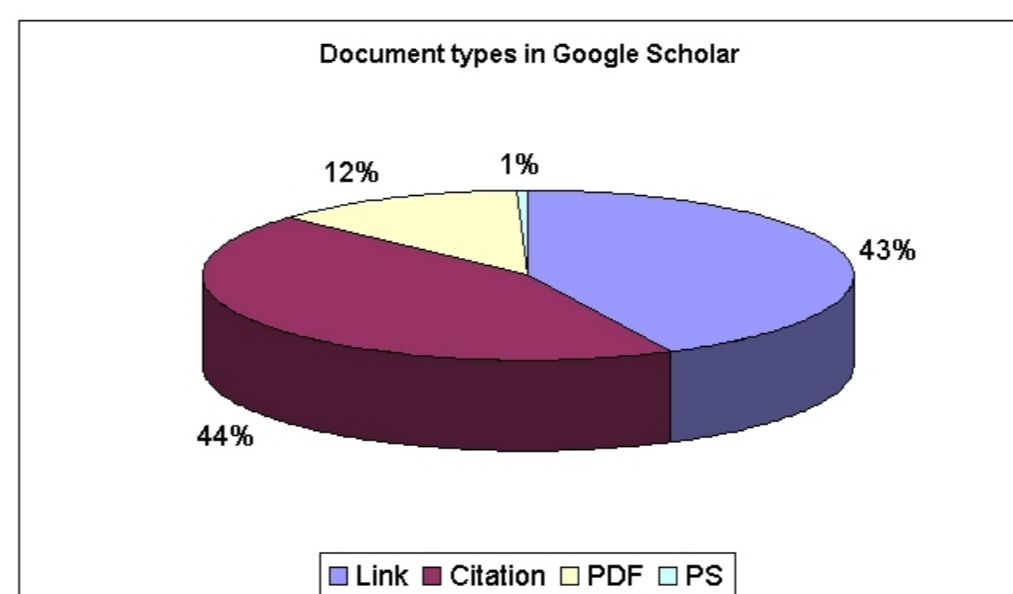
Querying complete journals lists

- ISI Journals (10.684 titles)
- DOAJ Journals (1.423 titles)
- IZ / SOLIS Journals (324 titles)

### Analyses

- Journal coverage
- Distribution of document types
- Webserver coverage
- Size estimation
- Recall test

List	Title	Match & Hit	no Match	no Hit
IZ (Solis)	324	228 (0.70)	89 (0.27)	20 (0.06)
DOAJ	1423	1078 (0.76)	337 (0.24)	221 (0.16)
ISI	10684	8931 (0.84)	1714 (0.16)	401 (0.04)



Webserver	Type	Frequency
ncbi.nlm.nih.gov	Digital Library	150616
ingenta.com	Publisher	68925
csa.com	Publisher	54652
ingentaconnect.com	Publisher	52051
springerlink.com	Publisher	21114
doi.wiley.com	Publisher	19280
kluweronline.com	Publisher	18196
adsabs.harvard.edu	Digital Library	16381
portal.acm.org	Digital Library	15280
blackwell-synergy.com	Publisher	14216
dx.doi.org	Linkresolver	13697
taylorandfrancis.metapress.com	Publisher	13221
ideas.repec.org	Digital Library	7681
ieeexplore.ieee.org	Digital Library	6405
journals.cambridge.org	Publisher	5379
nature.com	Publisher	4630
content.karger.com	Publisher	4219
muse.jhu.edu	Digital Library	3944
link.aip.org	Digital Library	3602
pubmedcentral.nih.gov	OA fulltext	3377
extenza-eps.com	Publisher	3303
papers.ssm.com	Digital Library	3271
iop.org	Digital Library	2259
arxiv.org	OA fulltext	2076
leonline.com	Publisher	1838

### Results

- Majority of journal title matches
- Bad Open Access coverage
- About 8.000.000 documents
- No actualization, no broad coverage of large servers
- Non scientific content

### Conclusions

- A good prototype with some annoying characteristics
- No substitute for traditional databases
- Potential in ACI and citations analysis
- OA search engine and WoS alternative

