

A Search Log Analysis of a Portuguese Web Search Engine

Miguel Costa, Mário J. Silva

LaSIGE @ Faculty of Sciences, University of Lisbon

Foundation for National Scientific Computing

INFORUM 2010, Braga, Portugal

Do Portuguese users search
 in the same way as other users?

Does search behavior influence
 web search engine design?



- Speed
 - e.g. special indexes, cache
- Quality of results
 - e.g. better ranking
- Web design
 - e.g. stand out most used functionalities

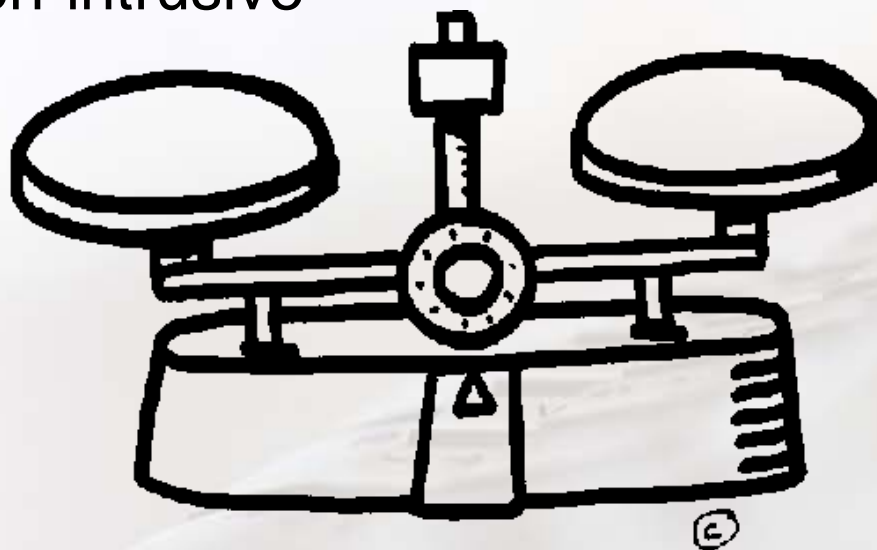
- Introduction
- Methodology & Dataset
- Results
- Conclusions

PROS:

- Large and varied
- Less bias
- Cheap
- Non-intrusive

CONS:

- Lack of context
- Lack of control



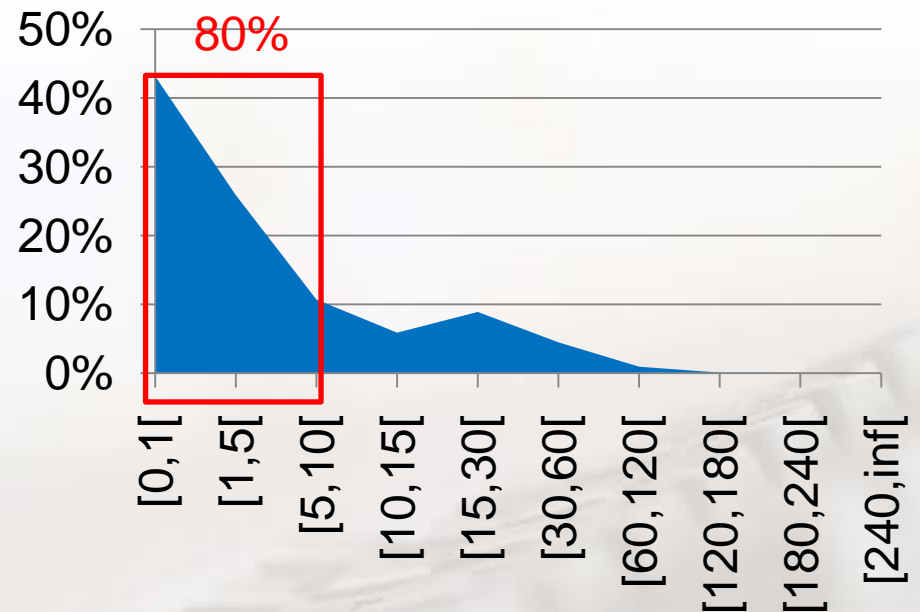
- Tumba – <http://www.tumba.pt>
- 2 full years – 2003 & 2004
 - several studies from the same period
 - baseline for future works
- 90% of the IP addresses → Portugal
- 98% of the interactions → Portuguese interface

How do users search?

Fast and short sessions

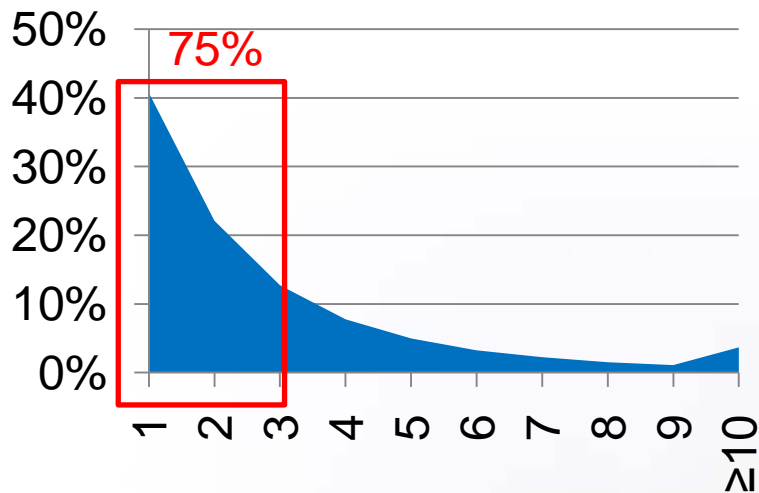
- Fast
- Few queries
- Few terms
- Few result pages
- Few clicks

Session Duration (min)

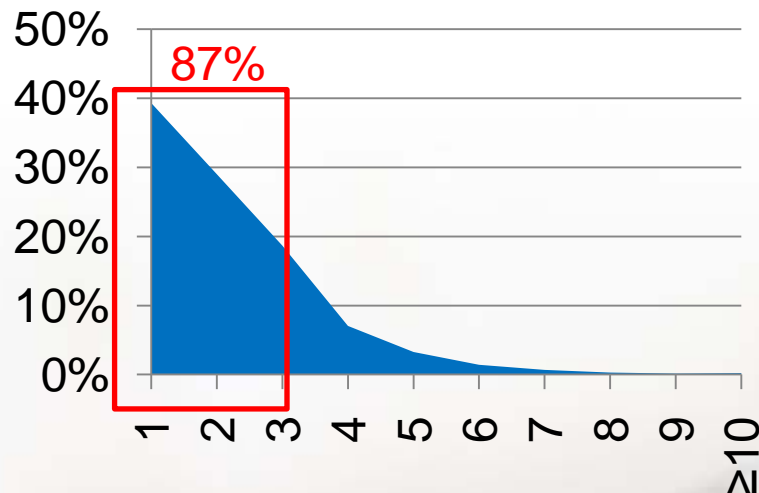


How do users search

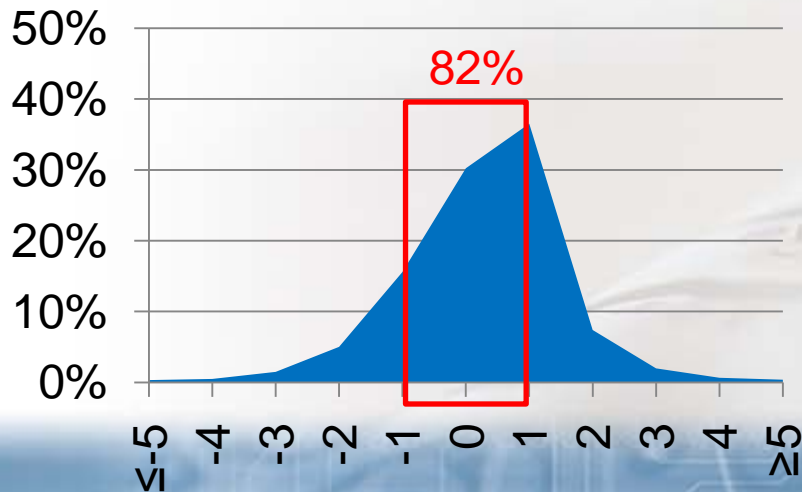
Queries per Session



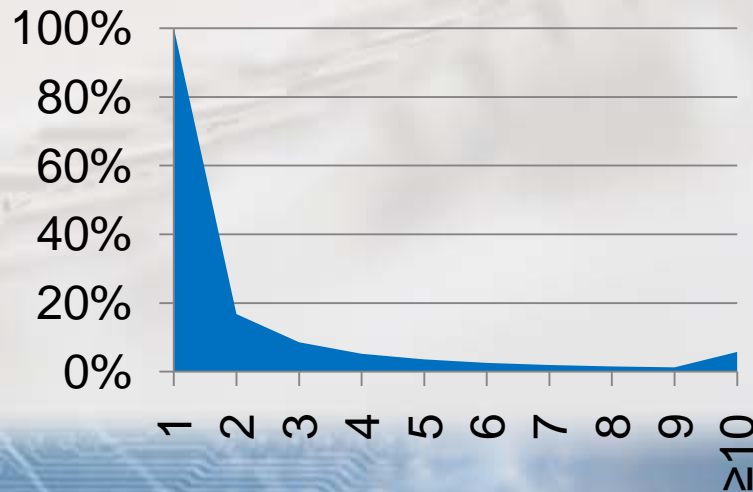
Terms per Query



Terms Changed



Result Page Viewed



Evolution from 2003 to 2004

- -1/2 term of query length
- +10% of sessions with less than 1 minute
- +9% of sessions with only one query
- +8% of sessions where only the first result page was viewed

Evolution from 2003 to 2004

- -1/2 term of query length
- +10% of sessions with less than 1 minute
- +9% of sessions with only one query
- +8% of sessions where only the first result page was viewed

Less data submitted, less results seen

What do users search for?

	Categories	2003	2004	Δ%
		% queries	% queries	
1	Commerce, Travel, Employment or Economy	22.4	20.3	- 2.1
2	People, Places or Things	14.8	17.7	2.9
3	Health or Sciences	10.5	11.8	1.3
4	Education or Humanities	7.2	10.5	3.3
5	Society, Culture, Ethnicity or Religion	5.6	6.1	0.5
6	Computers or Internet	6.4	5.9	- 0.5
7	Sex or Pornography	4.9	5.8	0.9
8	Entertainment or Recreation	8.7	5.1	- 3.6
9	Government	7.0	4.2	- 2.8
10	Performing or Fine arts	1.6	1.6	0.0
11	Unknown or Other	11.2	11.3	0.1

world region search engine	U.S. Excite	Europe FAST	Portugal Tumba!
single term queries	20% -30%	25% -35%	40%
terms per query	2.6	2.3	2.2
result pages viewed	1.7	2.2	1.4
queries per session	2.3	2.9	2.49 -2.94
topic most seen	Commerce, Travel	People, Places	Commerce, Travel

Less data submitted, less results seen

Conclusions

- Portuguese users
 - spend little time and effort on individual searches
 - tend to submit less data and see less results
 - search differently than other users
 - specificities can be used to tune web search engines

- Updated characterization of Portuguese users
- Characterization of Portuguese users from web archives



<http://archive.pt>

**80% of the web documents are
unavailable after 1 year**

Thank you.

