



Search engine dependency and its influence on data quality

By Ronan CHARDONNEAU

Intermediate report presentation –
Final Master Thesis



Table of contents

- Structure of the thesis
- Summary of selected chapters
- Further research and proceeding plan
 - Conclusion

Structure

Chapter 1: Introduction of the topic background;

Chapter 2: Concept of data quality;

Chapter 3: Search engines dependency;

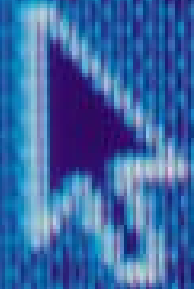
- Search engine market configuration;
- Search engines dependency aspect;
- Search engines dependency problems;

Chapter 4: Risks of search engines dependency and its influence on data quality;

- Measuring the problematic and trying to answer it;

Chapter 5: The Google example;

- Emphazing the report on Google

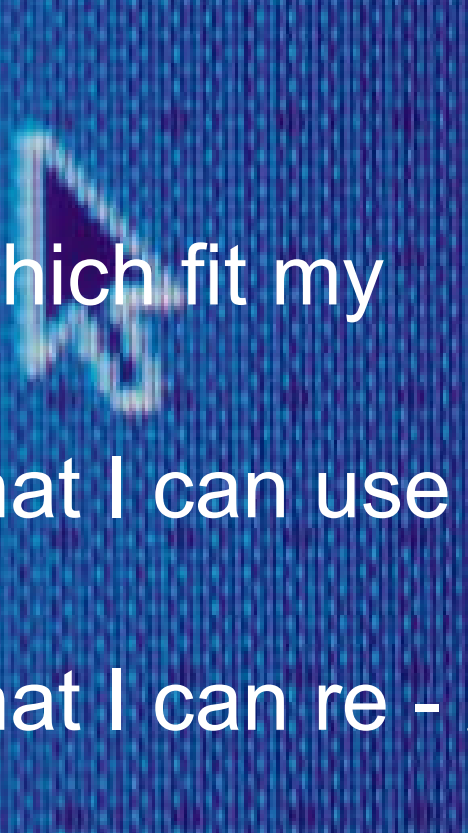


Chapter 1: Introduction

- Internet is everyday getting bigger and bigger;
- We are using more and more Internet to find information;
- Search engines are until now the best way to crawl the web;
- Search engine dependency is quite relevant;



Chapter 2: Data quality

- I have no results;
 - I expect to have some results;
 - I expect to have some results which fit my expectations;
 - I expect to have some results that I can use afterward;
 - I expect to have some results that I can re-use afterward;
- 

Chapter 3: Search engine dependency

- 3 different categories of search engine;
- 1 leader (Google), 1 follower (Yahoo), 2 champions (MS & Baidu) and very few local heroes;
- 1 winner/country, no revolution since 1998;
- SED phenomenon 3 categories
- Cons of addiction (one can think that the information does not exist, Privacy, disappearance, awareness about others SE, bad addiction);

Chapter 4: Topic

- Search engine dependency is proved;
- Search engines are using different technology and are not browsing the Internet at the same place;
- How to browse the Web properly;
- Measurement of the gap of information;

Chapter 5: The Google example

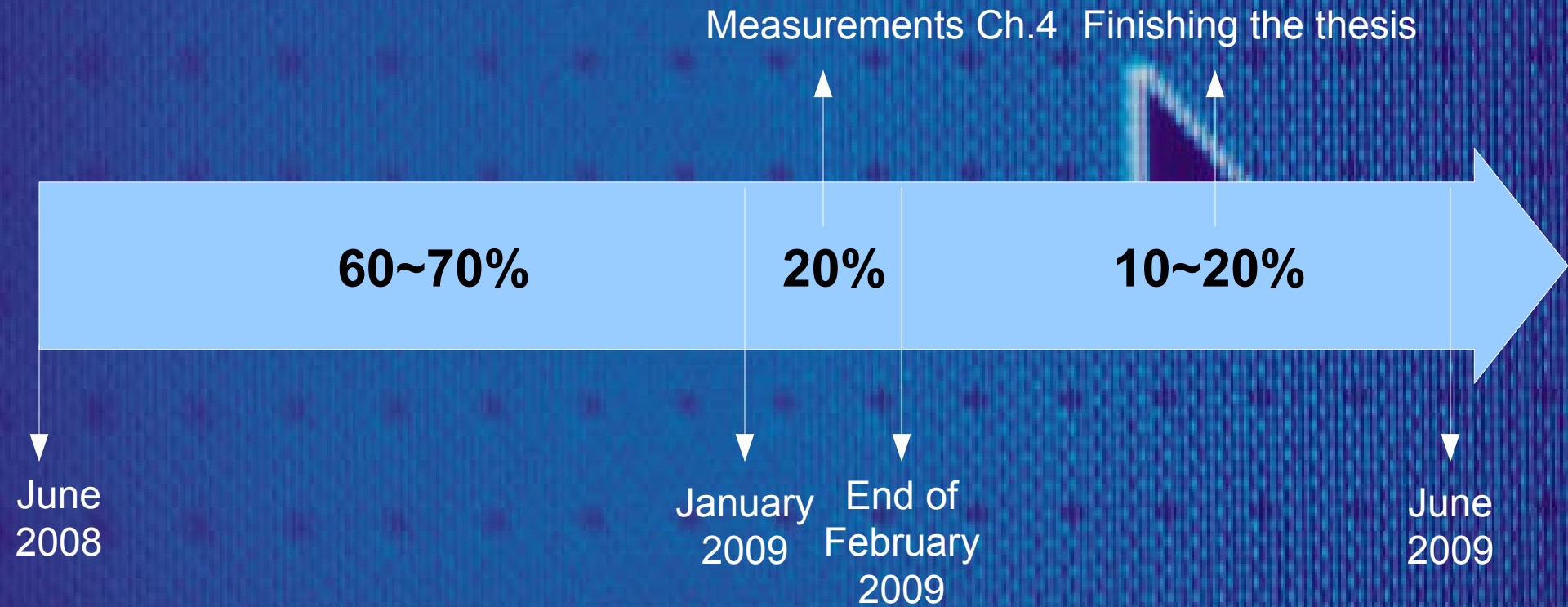
- An anglo-Saxon phenomenon (West vs East);
- Blogger, Gmail, Adsense, Google News...
- Banalization of Google;
- Google success weakness;
- Google sub search engines;



Further research and proceeding plan

- Chapter 4: the measurement of the gap information; (3 weeks of work)
- Coming back on Chapter 5 and make it better;
- Rewriting the report in an impersonal way;

Milestones



Conclusion

- On time;
- No particular stress for the coming events;
- A final thesis which until so far gave me great opportunities to learn and to do something I am really enjoying;
- Still some work to go on but quite insignificant regarding all the work I did so far;

Thank you for your attention

