

## Motivation

- Web data is lost everyday.
- Web archives have been harvesting since 1996.
- URL search is limited. Users want full-text search.
- Web search engines do not support time.

## Why a new evaluation?

- **Larger scale:** web archiving is incremental.
- **Poor ranking:** sparse and versioned web graphs. Web design trends evolve. Time is ignored.
- **Versions:** which to return and how to fuse them?



Procurar por: FCCN

entre 1 de Janeiro de 1996 e 13 de Julho de 2009

Localizer

Ajuda | Sobre o Arquivo

Resultados 1 - 10 de cerca de 163.412

### Portal FCCN - 28 de Março, 2002

Portal FCCN FCCN RCTS RCTS2 DNS GigaPix Projectos Localização Software FCCN Prod. Informat... sequência do sucesso das iniciativas anteriores, a FCCN e a Universidade de Coimbra, com o impres ... SELF-SERMCE ©2001 FCCN - Fundação para a Computação Científica Nacional Todos os direitos ...

<http://www.fccn.pt/> - [outras versões](#)

### Portal FCCN - 20 de Dezembro, 2001

Portal FCCN FCCN RCTS Contactos Serviços Projectos Directório Documentos Ligação das EB1 Software Infraestrutura Agenda RCTS2 DNS GigaPix Projectos Localização Software FCCN Prod. Informat ... reinserção social; etc. Com base em protocolos estabelecidos com o MCT, a FCCN tem dotado estas ...

<http://www.fccn.pt/RCTS/escolas/> - [outras versões](#)

### Speedmeter FCCN - 19 de Maio, 2007

Speedmeter FCCN CONDIÇÕES DE UTILIZAÇÃO O speed meter é uma funcionalidade desenvolvida internamente pela FCCN que permite efectuar medições de débito. Este sistema é disponibilizado "as is" não se garantindo a fiabilidade total dos resultados apresentados, pelo que, em caso algum, a FCCN será ...

<http://speedmeter.fccn.pt/> - [outras versões](#)

## What are users' intentions?

- ✓ **Navigational:** specific pages and archived versions.
- ✓ **Informational:** pages of topics written in the past.
- ✗ **Transactional:** do not access resources, such as old files.

## How are users searching?

- Are they "Googlefied"?
- Short queries.
- Dates and temporal expressions not provided.
- Use a date range filter to narrow the search.

## Conclusions

- Search engines technology cannot be directly applied in web archives.
- Time is the main feature of web archive collections, but is ignored for ranking results.
- The creation of a new evaluation is essential to foster research and produce sustainable knowledge.