A Search and Query Engine for Educational Research

A field specific search-engine and a multilingual search-engine

Thomas Severiens, ISN
severiens@isn-oldenburg.de

Berlin, September 15, 2011
Why a Search-Engine?

- Provide consistent interface to documents from various sources
- Detect educational research in the WWW
- Multilingual document base and multilingual interface
Focussed crawling

- Ask experts for good starting points
Focussed crawling

- Ask experts for good starting points
- Crawl all the links and collect more objects
Focussed crawling

- Ask experts for good starting points
- Crawl all the links and collect more objects
- and more, more, more…
Focussed crawling

- Ask experts for good starting points
- Crawl all the links and collect more objects
- and more, more, more…
- But: Are all of them still in the focus?
Focussed crawling

- Ask experts for good starting points
- Crawl all the links and collect more objects
- and more, more, more…
- But: Are all of them still in the focus?
- We need an algorithm to decide, if an object is in the focus or out of…
Focussed crawling (using Nutch)

- Extended crawl cycle
- Calculate score
- Index only relevant documents
- Decider-algorithm: metadata + classifier

Thanks to Sybille Peters for this slide!
How the Classifier works:
1. Calculate the fingerprint
How the Classifier works

- Machine based learning: Training data required to train the algorithm (positive and negative data space)
- Use of the trained algorithm via API (REST-based)
- Computing done in background on a cloud of computers (6)
How good does a Classifier separate relevant from irrelevant objects?

- Degree of “separation” depends on:
  - Coverage of relevant data space by training data,
    - the more training objects, the more likely this is fulfilled,
    - many journals,
    - many authors,
    - mixture of all genres of publications,
  - no multilingual documents & separation by language,
  - typical language in negative training data, but no overlap in field specific wording
students perceptions of selected aspects of primary school teaching. This information can be useful ... and coding methods. Results showed that their perceptions about primary school teaching were positive. ... & Feldman 2002 Student teachers perceptions of primary school teaching. A few studies investigated student AUH06592 Perceptions of Primary School Teaching by First Year Education Students Myung sook AUH University of New England ABSTRACT The purpose of the study was to investigate first year Education students perceptions of selected aspects of primary school teaching. ... Also few studies examine what reasons primary student teachers provide for their perceptions about primary school teaching and which are more frequently mentioned reasons for their perceptions. ... This study intends to examine selected issues.
Multilingual processing

- Add cross-language search to search engine, using
  - Term networks
  - Linguistic processing

Thanks to XEROX and DIPF for this slide!

www.eerqi.eu

A Project funded under the Socio-Economic Sciences and Humanities Theme.
Term networks: objective

- Enhancing the multilingual search facilities in the EERQI content base:
  - Provide related terms for user queries as suggestions for automatic or interactive „query expansion“
  - Provide translations of relevant terms in the topic area

- Problem:
  - Thesaurus terms as source for query expansion do not cover many possible query terms
  - Dictionaries as source for query expansion do not contain relations between terms and do not cover a topic area closely enough

  -> Construction of multilingual „term networks“

Thanks to XEROX and DIPF for this slide!