Wanted: Element Retrieval Users

Andrew Trotman
University of Otago
Dunedin, New Zealand

Outline

- I'll try and demonstrate that we don't know what we're doing
  - Document collection
    - Inappropriate
  - Querying methods
    - Inappropriate
  - Measurements
    - Not in agreement
  - Judgments
    - Inconsistent
- Of course, it might just be that *I* don't know what *I'm* doing!

Interpretation of Element Retrieval

- The same as document retrieval except:
  - The fundamental unit of retrieval is an element
  - In INEX it is an XML element
- A result list
  - Is a list of elements
  - Is not a list of documents
- Example
  - Collection of books
  - Result is “a few relevant pages”

Documents

- 12,107 XML documents from IEEE taken between 1995-2002
  - 12 magazines and 6 transactions
  - Academic documents
- Academic documents
  - Different types
    - Posters, conference papers, journal articles, books
    - Written to stand alone
    - Cited in their entirety
  - Are atomic
- Element retrieval of academic documents
  - Plucks document pieces, and presents them out of context

Nature of a Document Collection

- Elements must:
  - Make sense individually
  - Make sense within a larger contest
- Suitable documents must:
  - Be made of disparate parts
  - Have low coupling with their elements
  - Be marked up with “suitable” elements
    - IEEE collection, 29 terms
    - Relevant element over 1000 terms

Possible Collections

- Newspapers / magazines
  - Extraction of sub-atomic stories from atomic newspapers
  - Consists of stories
- Radio broadcast / magazine television
  - Extraction from a random mix
  - Combination of stories, music, dialogue, and advertising
- Plays
  - Extraction of dialogue, scenes and acts
  - Shakespeare (hasn’t this been done?)
- Lonely Planet Guide
  - Not yet examined

This Looks OK

- We've just been using the wrong document collection

Querying

- Tags are used for multiple purposes
  - Presentation purposes
  - Identify document structures
- IEEE collection
  - 192 tags in DTD
  - 11 (6%) used as targets in 2003 / 2004

Element	Percent
sec	41%
article	27%
Not specific	8%

- Are only <article> and <sec> suitable target elements?

The Existing Queries

- INEX 2003 topics
  - 63% contained errors (19 of 30)
- INEX 2004 topics
  - NEXI introduced
  - Error rate dropped to 12%
  - Parser downloaded by 13 IP addresses
  - Online parser used 635 times
    - For 84 CAS topics
    - Thats 7.5 times per topic!
  - Error rate declined, but it was very hard to write queries!

The Interactive Queries

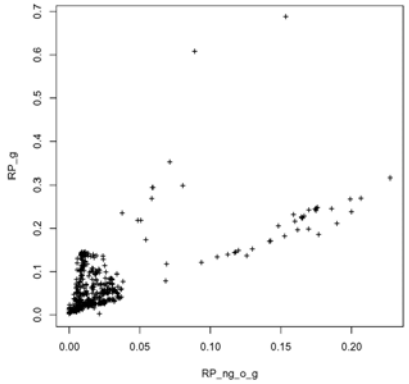
- Tombros et al., and Kim & Son
  - + and – used in no queries
  - Phrases used in less than 10% of queries
  - Average between 3.0 and 3.4 terms per query
- Summary:
  - We're going to get queries that look just like any other query: 3.0 terms and no structure!
- Consequence:
  - More effort on CO
  - Less effort should on CAS (remove it?)

This Doesn't Look Good

- We've been using the wrong document collection
- We don't know how to ask it questions

Measurement

- The existing metrics measure different things



- Essentially no agreement between XCG and some others
- What should we measure? Unless we know what users want, our experiments are simply thought experiments.

This Looks Bad

- We've been using the wrong document collection
- We don't know how to ask it questions
- We can't measure it

- At least we agree on what it is – don't we?
  - Given a query do we really agree on the answers?

Non-Zero Agreement

- Results taken from 12 double-judged topics at INEX 2004

Evaluation	Agreement ($\cap/\cup$)
TREC 4 P/B	0.49
TREC 4 A/B	0.43
TREC 4 P/A	0.42
TREC 6	0.33
INEX 2004 documents	0.27
INEX 2004 elements	0.16

- We agree on which documents are relevant
- We do not agree on which elements are relevant

E3S3 Agreement

- Results taken from 12 double-judged topics at INEX 2004
- Necessary to determine validity of strict quantization

Evaluation	Agreement ($\cap/\cup$)
Document contains E3S3	0.12
Element is E3S3	0.05

- That is
  - We do not agree which documents contain E3S3 elements
  - We totally disagree which elements are E3S3

- In other words, we can't see it, even if we're given it!

This Looks Pretty Bad

- We've been using the wrong document collection
- We don't know how to ask it questions
- We can't measure it
- We can't even spot it when we see it

Thesis

- All methodological problems stem from one cause
  - No one uses element retrieval (yet)
- Corollary
  - If we can identify some users, and build a system for them, then we'll have some way to know what they want
  - Consequents
    - We'll have a suitable document collection
    - We'll know how to ask it questions
    - We'll be able to measure it
    - We'll be able to make real advances in element retrieval

Conclusions

- Stop the thought experiments
  - No "thought user model" will substitute for a user base
- Find an application of element retrieval
  - Fetch & Browse?
    - Is this really element retrieval
    - Must compare with passage retrieval
- More interactive experiments are needed – this is the best we have in the absence of users