

Buying bestsellers online: A case study in Search and Searchability

Trystan Upstill

Australian National University

Nick Craswell and David Hawking

CSIRO Mathematical and Information Sciences

The Plan

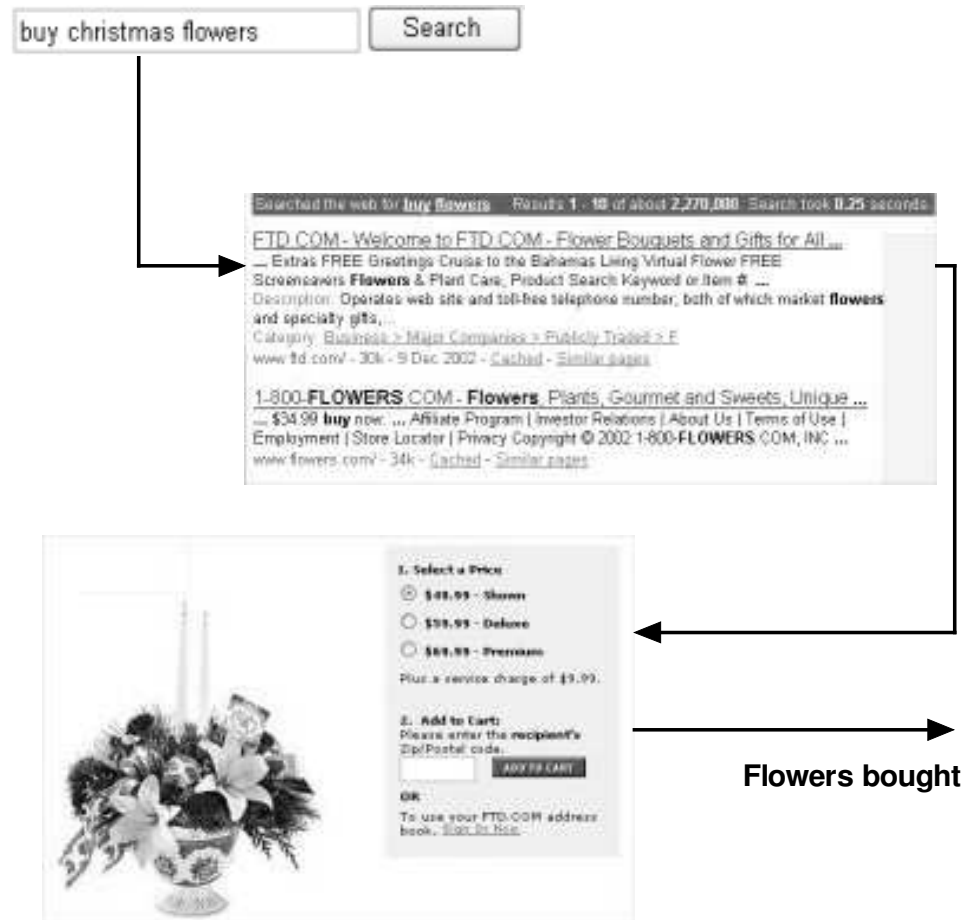
- Why study book buying on the Web?
- The Experiment
- Results
- Bookstore Properties
- Recommendations and Conclusion

Why study book buying on the WWW?

- People buy books
- Many bookstores use their online presence to boost their sales
 - Some depend on online sales
- Search engines should help people buy books
- (We hope to be able to generalise our results across all similar products sold on the WWW)

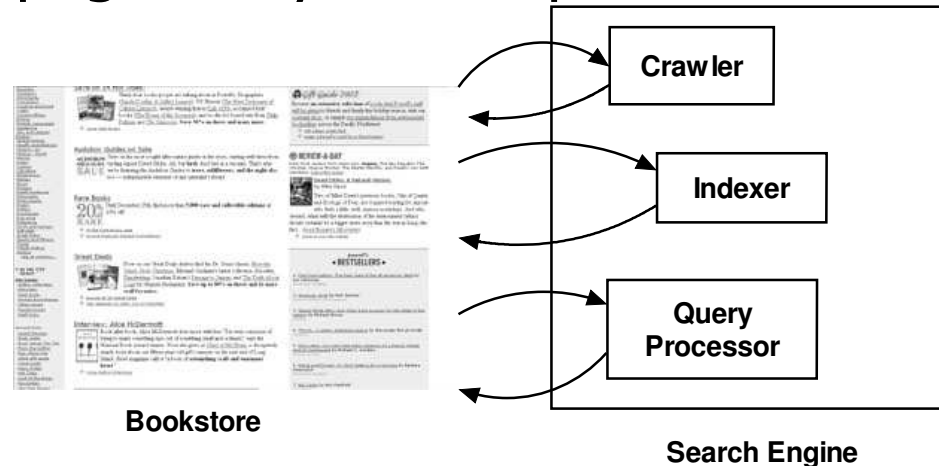
Question I - Search

- How well can search engines process search requests that result in transactions?
 - **Transactional search** (Travis and Broder, 2001)



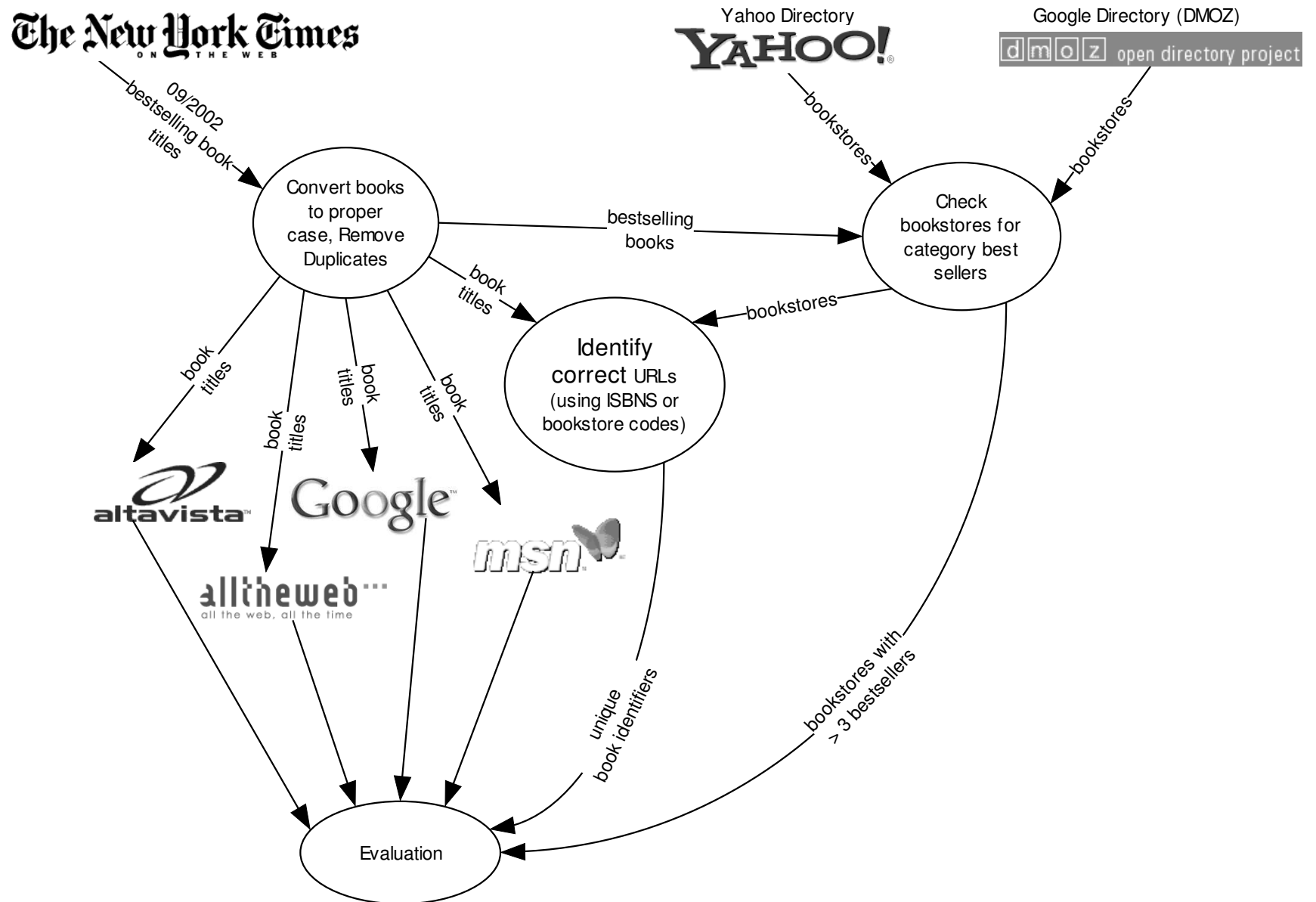
Question II - Searchability

- How compatible are the different service/product providers (bookstores) with existing search engines?
 - Can the pages be crawled/indexed?
 - Will pages be thrown out as duplicates?
 - How well will the pages be matched and ranked?
 - Is the best page returned?
- Providers that cannot be searched and/or return irrelevant pages may cut-off potential customers



ADCS 2002 - Buying bestsellers online: A case study in search and searchability

The Experiment



Manually Collected Data

Book information

Book	ISBN	AllDirect ID	Cat
A Painted House	044023722X, 038550120X	2993	8
Artemis Fowl	0786808012, 0786817070	8883	1
Olivia Saves the Circus	068982954X	9928	1
Red Rabbit	0399148701	3774	4

Bookstore information

Bookstores	Uses ISBN	Categories
1BookStreet.com	Yes	1,2,3,4,6,7,8,9
A1Books.com	Yes	3,5,6,7,8,9
AllDirect.com	No	1,2,3,4,5,6,7,8,9
Amazon.com	Yes	1,2,3,4,5,6,7,8,9

Amazon.com: Books: **A Painted House**

... Learn more. - or -. Don't have one? We'll set one up for you. **A Painted House** by John Grisham ... With **A Painted House**, however, he strikes out in a new direction. ...

www.amazon.com/exec/obidos/tg/detail/-/038550120X?vi=glance - 67k - [Cached](#) - [Similar pages](#)

Amazon.com: Books: **A Painted House**

... or -. Don't have one? We'll set one up for you. **A Painted House** by John Grisham ... With **A Painted House**, however, he strikes out in a new direction. ...

www.amazon.com/exec/obidos/tg/detail/-/044023722X?vi=glance - 58k - [Cached](#) - [Similar pages](#)

[[More results from www.amazon.com](#)]

CNN.com - Books - Excerpt: '**A Painted House**' - January 25, 2001

... Excerpt: '**A Painted House**'. "**A Painted House**" By John Grisham Doubleday Fiction. An open letter from John Grisham: "**A Painted House**" is not a legal thriller. ...

www.cnn.com/2001/books/beginnings/01/25/grisham.excerpt/ - 44k - 7 Dec 2002 - [Cached](#) - [Similar pages](#)

John Grisham | The Official Web Site

... **A Painted House** Doubleday 0-385-50120-X February 2001 \$27.95 ... **A Painted House** is a moving story of one boy's journey from innocence to experience. ...

www.randomhouse.com/features/grisham/books/paintedhouse/ - 11k - 7 Dec 2002 - [Cached](#) - [Similar pages](#)

Barnes & Noble.com - **A Painted House**

... Writers: **A Painted House** John Grisham, ... forever. **A Painted House** is a moving story of one boy's journey from innocence to experience. ...

search.barnesandnoble.com/booksearch/isbnInquiry.asp?isbn=0375431012&pwb=1 - 67k - [Cached](#) - [Similar pages](#)

Free book excerpt from **A Painted House** by John Grisham, plus ...

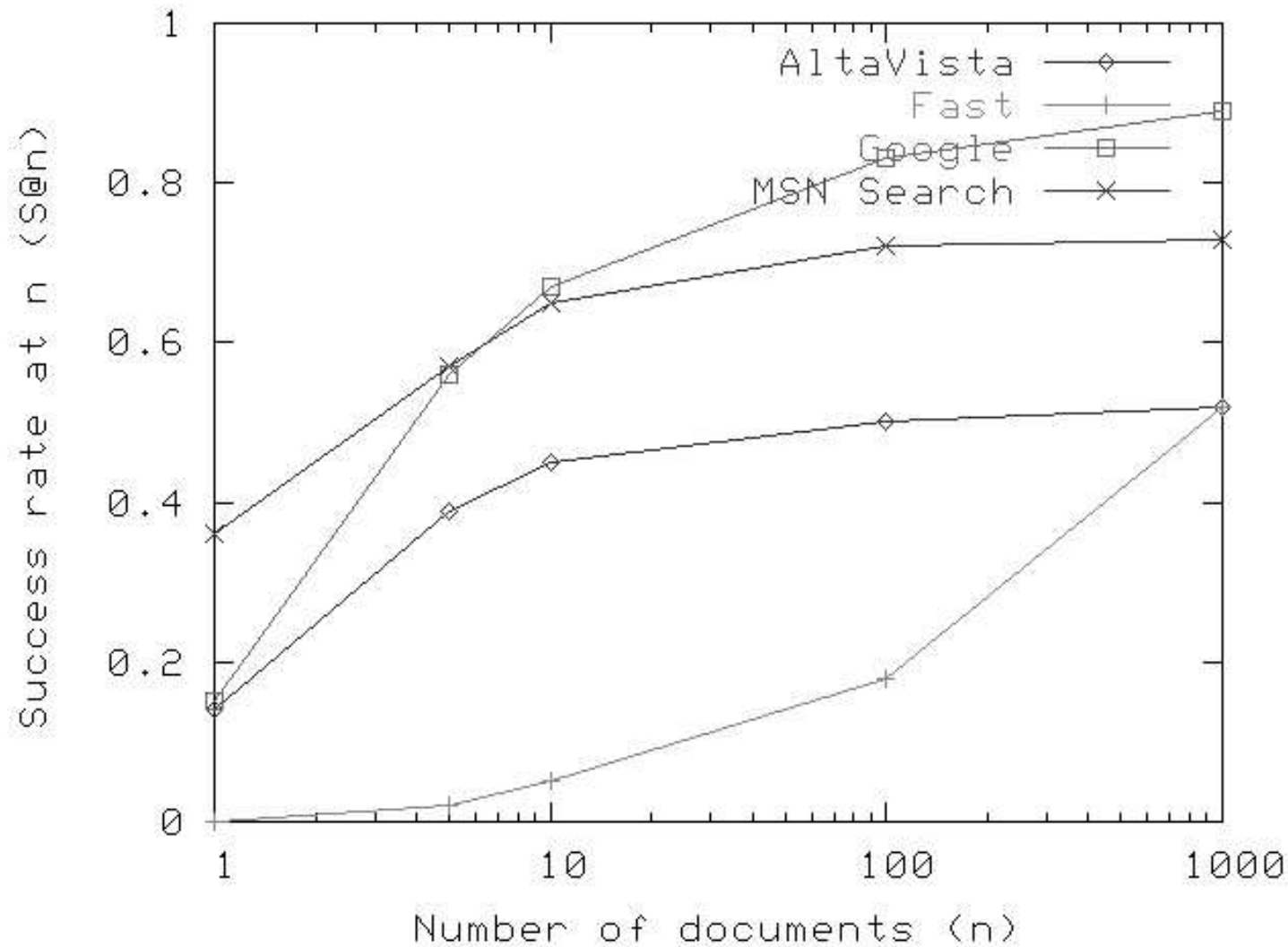
On this page: The book jacket summary and reviews of **A Painted House**

by John Grisham. Plus a substantial excerpt and author bio.

Evaluation Measures

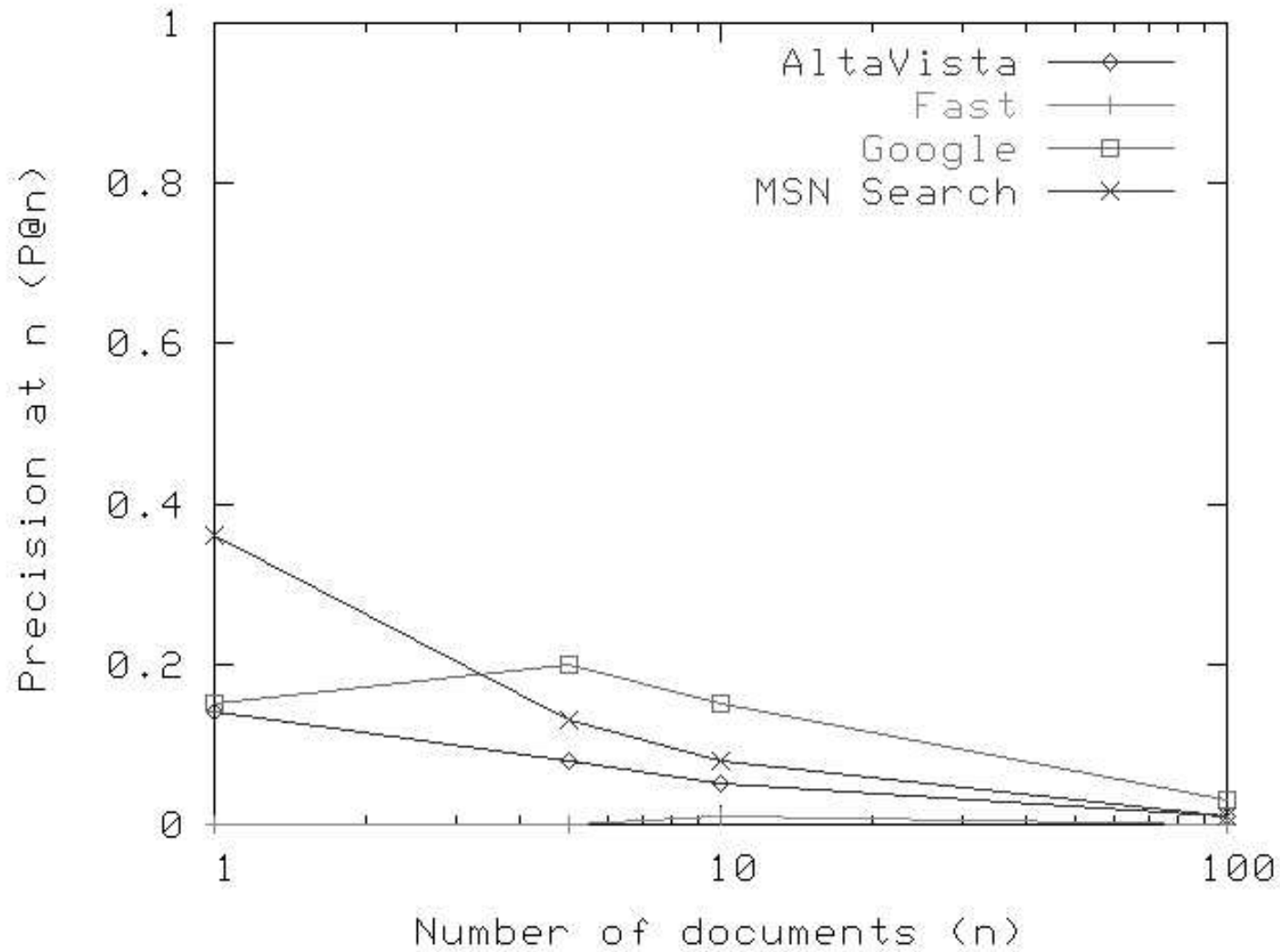
- Success Rate:
 - Represents how often search engines are returning relevant transactional pages at a certain cutoff
 - Represented by $S@n$
 - In the example $n=1$ (and 5, 10 ...) is a success
- Precision
 - Represents the amount of choice a search engine is offering (high precision indicates more book buying alternatives). What proportion of the results at a cutoff are relevant transactional pages.
 - In the example $P@5$ is $2/5$
 - $P@1 = S@1$

Search Engine Success Rates



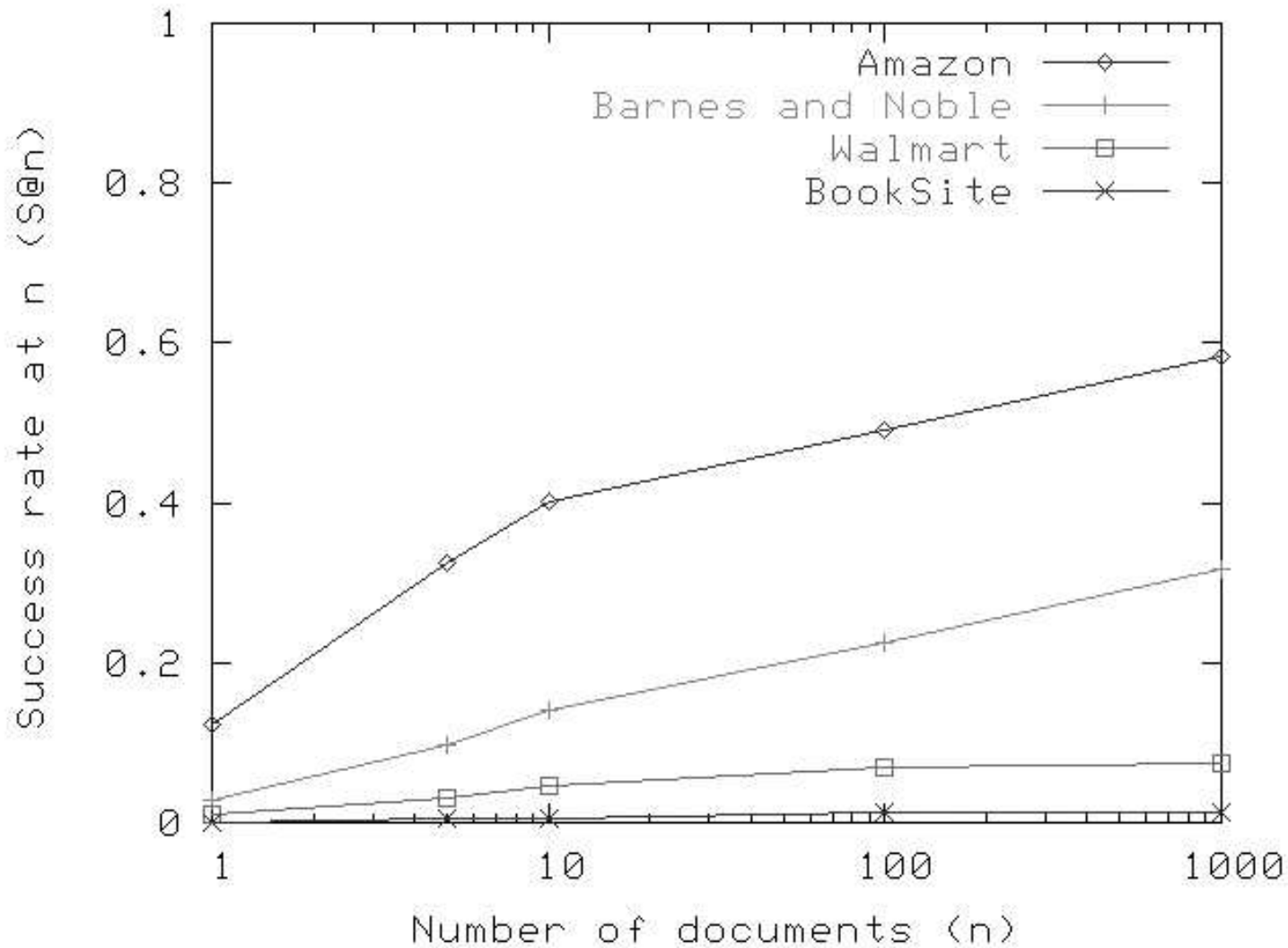
ADCS 2002 - Buying bestsellers online: A case study in search and searchability

Search Engine Precision



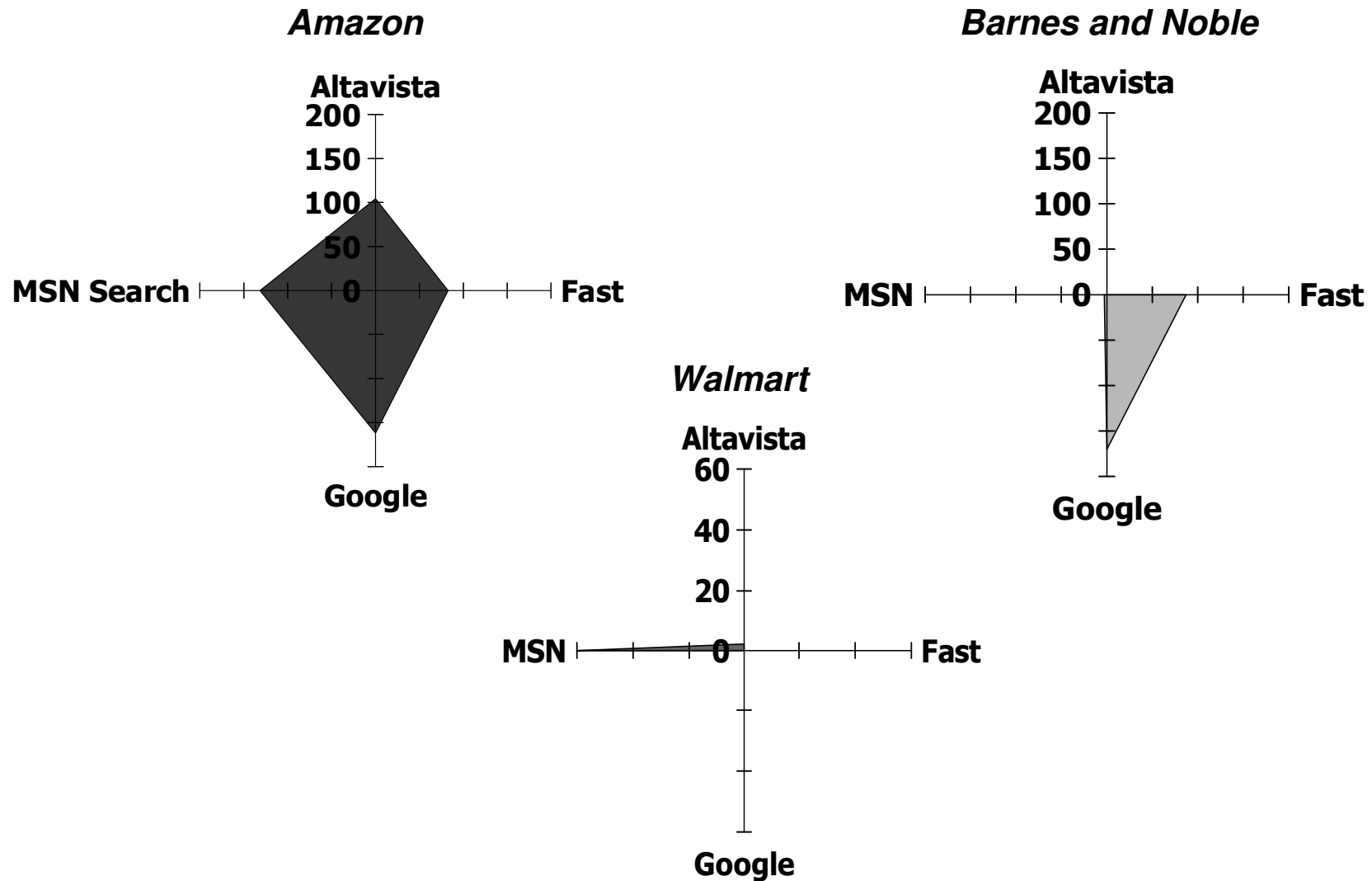
ADCS 2002 - Buying bestsellers online: A case study in search and searchability

Bookstore Success Rates



ADCS 2002 - Buying bestsellers online: A case study in search and searchability

Success Rate Distribution



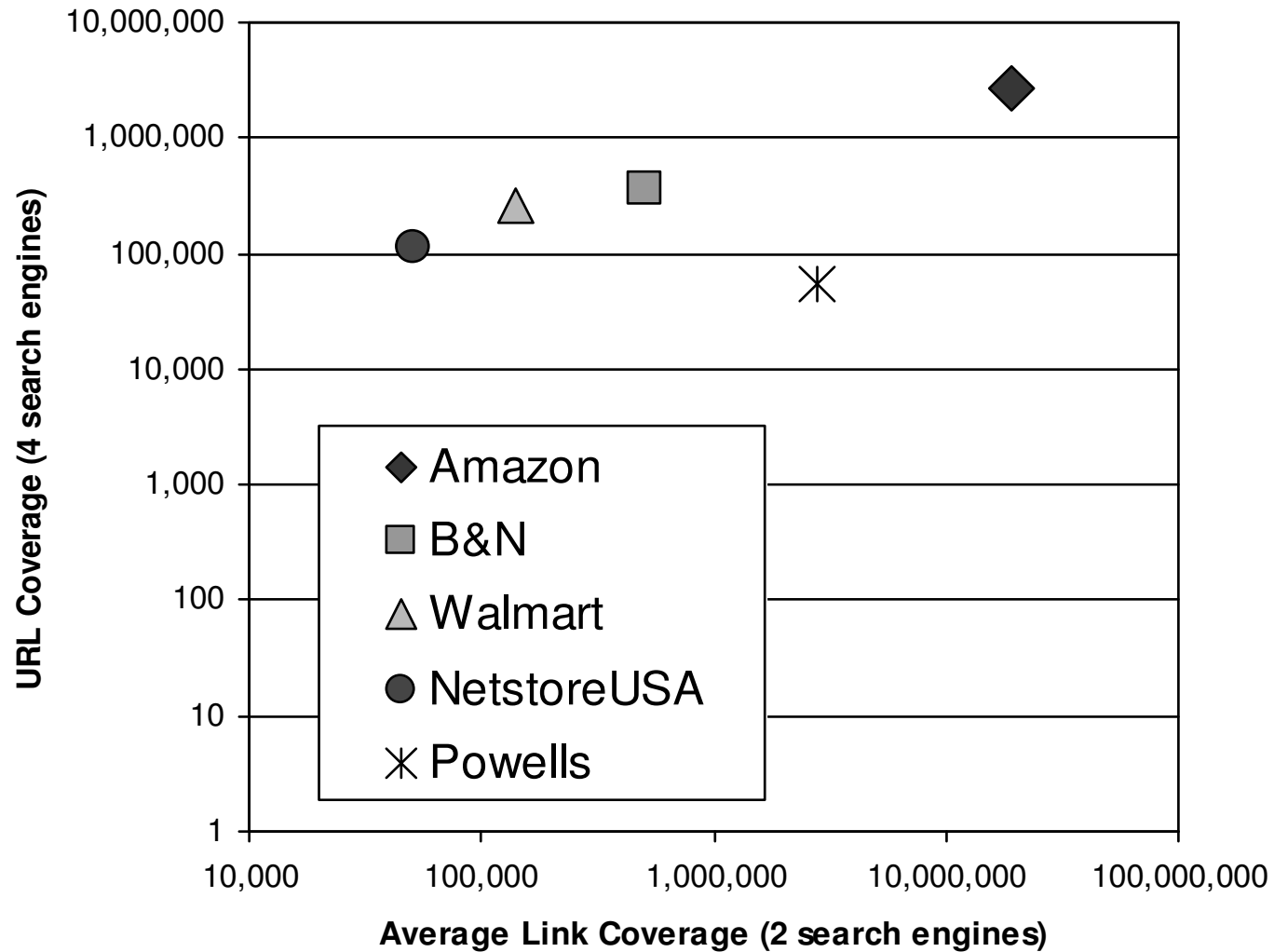
The Winners

- The Winners
 - Bookstore:
 - Amazon
 - Success Rate @ 1 was 0.125 (or 1/8 queries)
 - Barnes and Noble second best: S@1 was 0.028 (1/35 queries)
 - Search engine
 - MSN Search and Google
 - MSN Search returns more books as their first result, Google gives you more book buying choices.
 - » MSN S@1 was 0.36 (~ 1/3 queries) P@5 was 0.13 (~ 1/8 correct)
 - » Google's S@1 was 0.15 (~ 1/7 queries) P@5 was 0.20 (~ 1/5 correct)
 - Google's success rate betters MSN Search after 10 results
 - However...
 - Results still relatively poor.
 - Navigational Search S@1: 0.74 (Google & Fast, Airline HP Search)

Bookstore characteristics

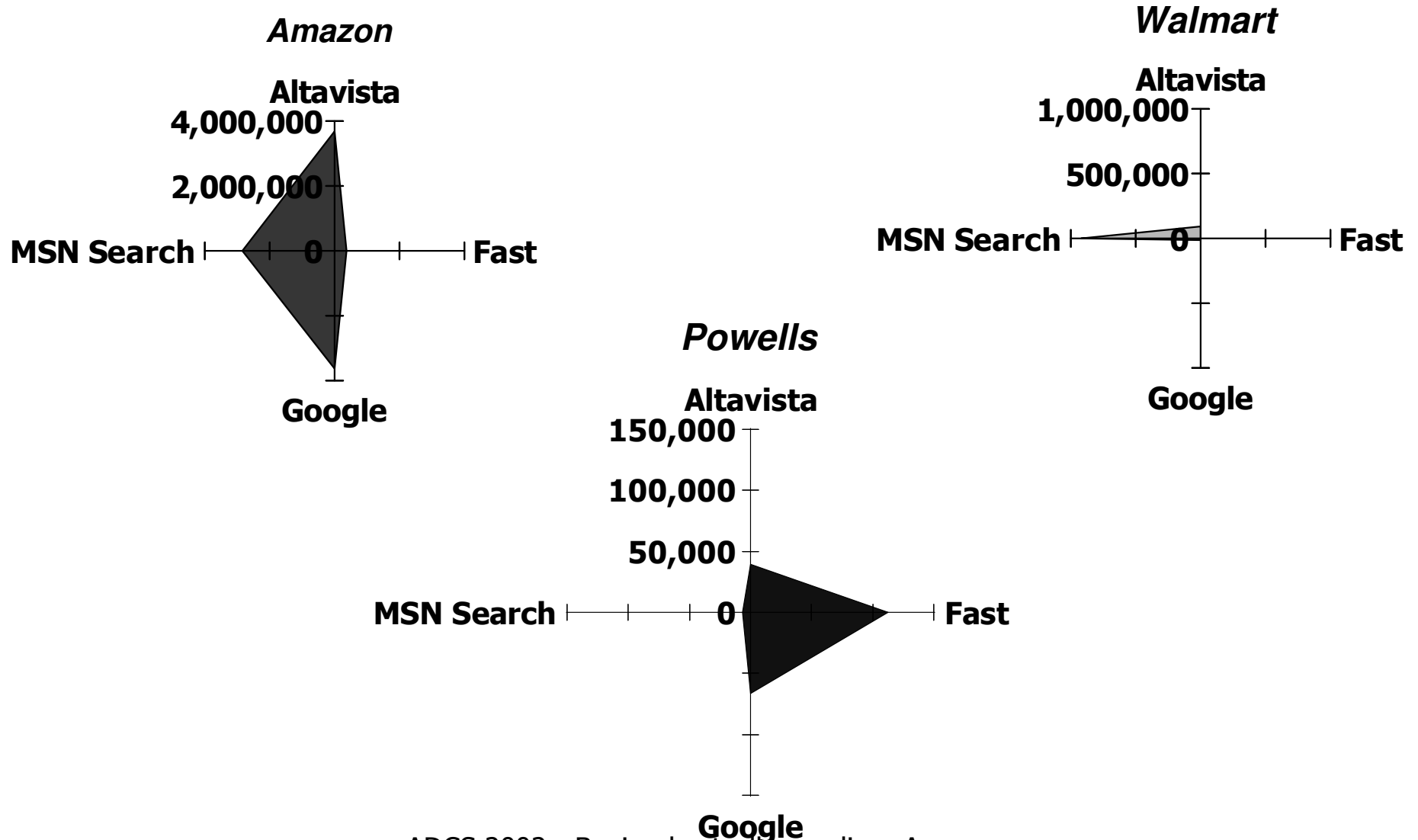
- We measured:
 - Bookstore URL Coverage
 - How many URLs from a bookstore are indexed by a search engine
 - If few URLs are indexed there is little chance that deep transactional pages would have been crawled/indexed.
 - Bookstore Link Coverage
 - How many links to a bookstore has a search engine observed
 - If few links are observed it is less likely that a bookstore will be crawled. Link ranking algorithms also favour sites with a large number of in-links.
- Coverage is only one aspect of searchability

Link/URL Coverage



ADCS 2002 - Buying bestsellers online: A case study in search and searchability

URL Coverage Distribution

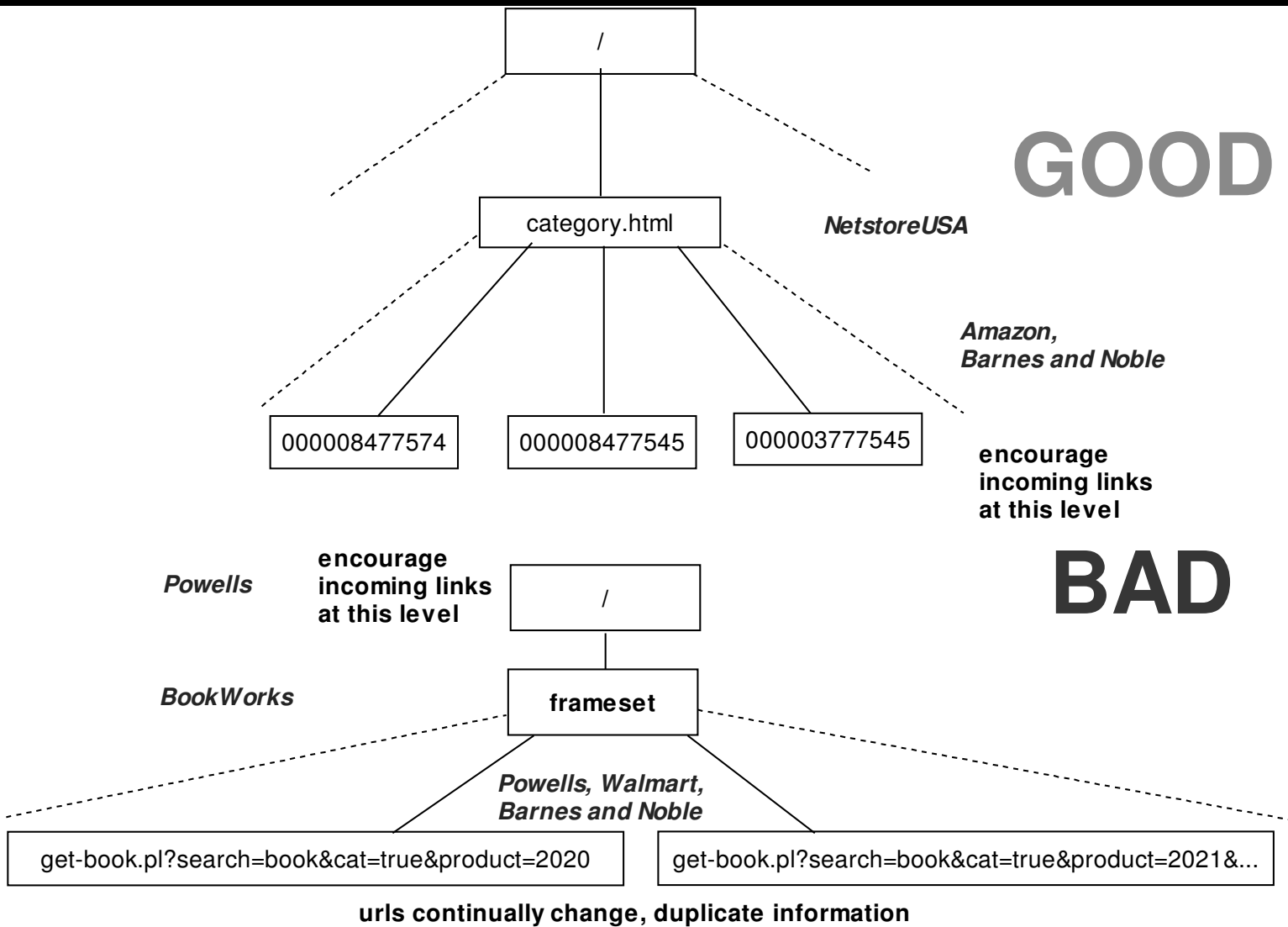


ADCS 2002 - Buying bestsellers online: A case study in search and searchability

Transactional search suggestions

- Perform deep crawls of at least one bookstore
- Be prepared to crawl dynamic sites
- Consider service provider partnerships
- Make sure you get your referral credits

Observed Good vs. Bad Site Structure



Conclusions

- Most bookstores have low searchability
 - Amazon > Barnes and Noble > Walmart > Other
- Transactional search is not performed very well by current search engines
 - MSN Search ~ Google > AltaVista > Fast
- People should be aware of these issues and consider them when designing Web sites

Future Work

- Find out what the other retrieved pages are
 - Are the informational, navigational, different media type?
- Attempt to classify transactional pages to improve search performance
- Expansion/Generalisation of experiments using CDs, Movies, Electronics search

Questions?

- Presentation available at:
<http://cs.anu.edu.au/~Trystan.Upstill/pubs/adcs-2002.pdf>
- Email:
Trystan Upstill: trystan@cs.anu.edu.au
Nick Craswell: Nick.Craswell@csiro.au
David Hawking: David.Hawking@csiro.au