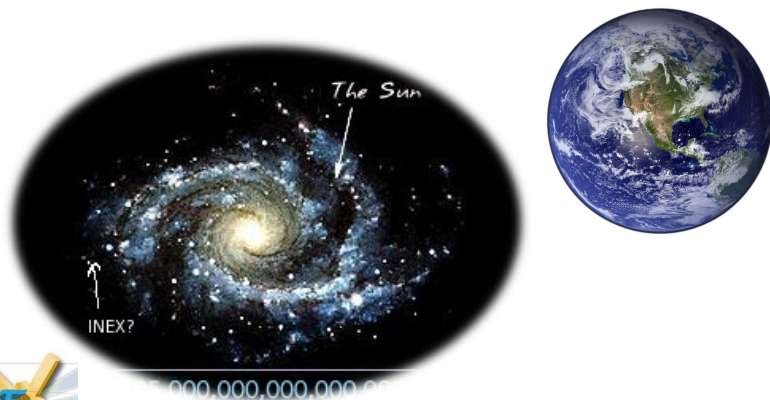# Element Retrieval in Digital Libraries: Reality Check

Philipp Dopichaj

August 10, 2006

XML Retrieval – Light Years from Earth?
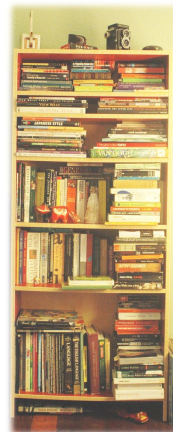
Online Library Services

Features

Both Books24x7 and Safari

- have business and IT books,
- are accessed with a web browser,
- are subscription-based, and
- provide basic keyword search.

Key difference

Safari subscribers have to deal with the "bookshelf" concept.

Google Book Search

Try your search on <u>NYTimes</u>, <u>ReviewScout</u>, <u>Amazon.com</u>, <u>Barnes&Noble</u>, <u>Blackwells</u>

<u>Advances in **XML** Information **Retrieval**: Third International Workshop of the Initiative for the... - Page 1</u>
Computers - 2005 - 438 pages
This means that an **XML retrieval** system needs not only to find relevant ...
Evaluating the effectiveness of **XML retrieval** systems requires a test ...
<u>Limited preview</u> - <u>Table of Contents</u> - <u>Index</u> - <u>About this book</u>

<u>Intelligent Search on **Xml** Data: Applications, Languages, Models, Implementations, and Benchmarks - Page 133</u>
Computers - 2003 - 319 pages
In this context, we describe and deploy a **retrieval** model that H. Blanken et al.
(Eds.): Intelligent Search on **XML** Data, LNCS 2818, pp.-143, 2003. ...
<u>Limited preview</u> - <u>Table of Contents</u> - <u>Index</u> - <u>About this book</u>

<u>Advances in Information **Retrieval**: 27th European Conference on IR Research, Ecir 2005, Santiago... - Page 110</u>
Computers - 2005 - 572 pages
Proceedings of the First Workshop of the INitiative for the Evaluation of **XML Retrieval** (INEX), 2002. 10. N. Fuhr and K. Großjohann. ...
<u>Limited preview</u> - <u>Table of Contents</u> - <u>Index</u> - <u>About this book</u>

Google Book Search

Book Search

Books **1** - **10** with **7050** pages on **xml retrieval**. (0.00 seconds)

Try your search on [NYTimes](#), [ReviewScout](#), [Amazon.com](#), [Barnes&Noble](#), [Blackwells](#)

Advances in XML Information ... Workshop of the Initiative for the... - Page
Computers - 2005 - 438 pages
This means that an **XML retrieval** ...
Evaluating the effectiveness of **XML**
Limited preview - Table of Contents

Results 1-5 of 30

Page v 🔑
... **XML retrieval** systems. ...

Page x
... of Contents Hybrid **XML Retrieval** ...

Page 1
This means that an **XML retrieval** ...

Page 7 🔑
... content-oriented **XML retrieval** ...

Page 14 🔑
... approaches for **XML retrieval** ...

Next »

Intelligent Search on Xml D... ...dels, Implementations, and Benc...
Computers - 2003 - 319 pages
In this context, we describe and dep
(Eds.): Intelligent Search on **XML** D
Limited preview - Table of Contents

Advances in Information Retrieval: 27th European Conference on IR Research, Ecir 2005, Santiago... - Page 110
Computers - 2005 - 572 pages
Proceedings of the First Workshop of the INitiative for the Evaluation of **XML Retrieval** (INEX), 2002. 10. N. Fuhr and K. Großjohann. ...
Limited preview - Table of Contents - Index - About this book
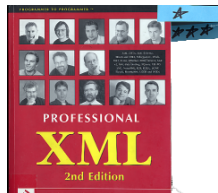
Results Grouped by Document

The result list consists of documents;
in each document, the most relevant sections
are highlighted.

INEX Tasks

Relevant in Context, Best in Context
(2005: Fetch and Browse)

Books24x7

XML Databases and the Semantic Web
by Bhavani Thuraisingham
Auerbach Publications © 2002 (306 pages)
ISBN:0849310318
A comprehensive view of critical technologies for the Web in general and XML in particular.

Top Section Hits (of 6 in this book)
- Chapter 4: Information Retrieval Systems and XML
- Chapter 1: Introduction (1.5 Organization of This Book)
- Chapter 4: Information Retrieval Systems and XML (4.7 Markup Languages and SGML)

Relevant Chapters in the Table of Contents

Web-Enabled Systems Integration: Practice and Challenges
by Ajantha Dahanayake and Waltraud Gerhardt (eds)
Idea Group Publishing © 2003 (317 pages)
ISBN:1591400414
A collection of quality research papers that describes original ideas and insights associated with the task of integration.

Top Section Hits (of 9 in this book)
- Chapter IX: Web Retrieval of XML Documents—Practice and Challenges
- Chapter IX: Web Retrieval of XML Documents—Practice and Challenges (INFORMATION RETRIEVAL APPROACHES TO XML DOCUMENT RETRIEVAL)
- Chapter X: Flexible Digital Library Search (POPULATING AND MAINTAINING THE INDEX)

Relevant Chapters in the Table of Contents

Books24x7

XML Databases and the Semantic Web
by Bhavani Thuraisingham
Auerbach Publications © 2002 (306 pages)
ISBN:0849310318

A comprehensive view of critical technologies for the Web

Top Section Hits (of 6 in this book)
- Chapter 4: Information Retrieval Systems and XML
- Chapter 1: Introduction (1.5 Organization of This Book)
- Chapter 4: Information Retrieval Systems and XML (4.7
Relevant Chapters in the Table of Contents

Web-Enabled Systems Integration: Practice and Challen
by Ajantha Dahanayake and Waltraud Gerhardt (eds)
Idea Group Publishing © 2003 (317 pages)
ISBN:1591400414
A collection of quality research papers that describes origi
integration.

Top Section Hits (of 9 in this book)
- Chapter IX: Web Retrieval of XML Documents—Practice and Challenges
- Chapter IX: Web Retrieval of XML Documents—Practice and Challenges (INFORMATION RETRIEVAL APPROACHES TO XML DOCUMENT RETRIEVAL)
- Chapter X: Flexible Digital Library Search (POPULATING AND MAINTAINING THE INDEX)
Relevant Chapters in the Table of Contents

Table of Contents

XML Databases and the Semantic Web
Preface
Chapter 1 - Introduction
Part I - Supporting Technologies for XML
Chapter 2 - The World Wide Web and XML
Chapter 3 - Web Data Management and XML
Chapter 4 - Information Retrieval Systems and XML
Chapter 5 - Information Management Technologies and XML
Chapter 6 - E-Commerce and XML
Chapter 7 - Metadata, Ontologies, and XML
Conclusion to Part I

Safari "View by Book"

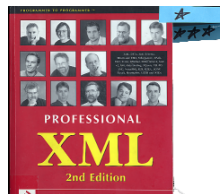| Title	Rank	Relevant Sections	Publisher	Pub. Date
Book **MCAD/MCSD Training Guide (70-310): Developing XML Web Services and Server Components with Microsoft® Visual Basic® .NET and the Microsoft .NET Framework** By Mike Gunderloy Slots: 2.0 Table of Contents	★★★	1. X–Z 2. Accessing and Manipulating XML Data 3. What the Developing XML Web Services and Server Components with Microsoft Visual Basic .NET and the Microsoft .NET Framework Exam (70-310) Covers More...	Que	2003/03/25
Book **Microsoft®C# Programming for the absolute beginner** By Andy Harris Slots: 1.0 Table of Contents	★★★	1. Chapter Basic XML: The Quiz Maker 2. Storing Entire Objects with Serialization 3. Examining the Quizzer Program More...	Premier Press	2002/01/01
Book **DB2® Universal Database™ v8 Application Development Certification Guide, 2nd Edition** By David Martineau, Kevin Gashyna, Steve Sanyal, Michael Kyprianou Slots: 2.0 Table of Contents	★★★	1. DB2 XML Extender 2. Summary 3. Usage More...	IBM Press	2003/06/30

INEX

The result list consists of
parts of documents.

INEX Tasks

- Thorough: overlaps are allowed
  (deemed to be system-oriented)

- Focused: no overlapping results
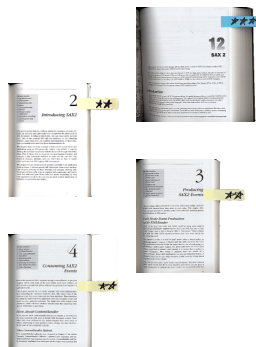
INEX

The result list consists of *parts* of documents.

INEX Tasks

- Thorough: overlaps are allowed (deemed to be system-oriented)
- Focused: no overlapping results

INEX

The result list consists of *parts* of documents.

INEX Tasks

- Thorough: overlaps are allowed (deemed to be system-oriented)
- Focused: no overlapping results

INEX

The result list consists of
parts of documents.

INEX Tasks

- Thorough: overlaps are allowed
  (deemed to be system-oriented)
- Focused: no overlapping results

INEX

The result list consists of
parts of documents.

INEX Tasks

- Thorough: overlaps are allowed
  (deemed to be system-oriented)
- Focused: no overlapping results

Safari "View by Section"

Title	Rank	Chapter/Section Title	Publisher	Pub. Date
Book MCAD/MCSD Training Guide (70-310): Developing XML Web Services and Server Components with Microsoft® Visual Basic® .NET and the Microsoft .NET Framework By Mike Gunderloy Slots: 2.0 Table of Contents	★★★	X–Z	Que	2003/03/25
		... retrieving sets of nodes from an XML document. XML attribute A piece of information describing an XML element. XML element An XML tag together with its contents. XML wire format The structure of the actual XML messages passed between Web services servers and clients. ...		
Book Microsoft®C# Programming for the absolute beginner By Andy Harris Slots: 1.0 Table of Contents	★★★	Chapter Basic XML: The Quiz Maker	Premier Press	2002/01/01
		... XML: The Quiz Maker Since the late 1990s, there has been a great deal of interest in a technology called XML (eXtended Markup Language) . XML has promised to transform many types of programming, and support of XML is one selling point of the .NET frame work. You will now learn how to incorporate XML in your ...		
Book DB2® Universal Database™ v8 Application Development Certification Guide, 2nd Edition By David Martineau, Kevin Gashyna, Steve Sanyal, Michael Kyprianou Slots: 2.0 Table of Contents	★★★	DB2 XML Extender	IBM Press	2003/06/30
		... XML Extender DB2 XML Extenders map XML documents to and from relational databases. This means the XML documents can be broken down into elements and inserted into a relational database, or a new XML document can be generated from existing relational database tables. XML Extender also supports validating ...		

Content-and-Structure Search

The query not only specifies the search terms *("what")*,
but also structural constraints *("where")*.

INEX: NEXI

```
//template[about(.//@name,book reference)]
   //*[about(.,architecture)]
```

Safari and Books24x7 provide limited support in their
advanced search interfaces.

Content-and-Structure Search

The query not only specifies the search terms *("what")*,
but also structural constraints *("where").*

INEX: NEXI

```
//template[about(.//@name,book reference)]
   //*[about(.,architecture)]
```

Safari and Books24x7 provide limited support in their
advanced search interfaces.

Safari Advanced Search

Search For:
- ● with **all** of these words
- ○ with **exactly** these words
- ○ with **at least one** of these words

In:
- ● the **full text**
- ○ **Code fragments** only
- ○ **Section title** words only
- ○ **tips and how-tos** only

Words in Book Titles:

Primary Category:

Author:

ISBN:

Year Published: **from** [   ] **to** [   ]

Publisher:
```
All

Addison Wesley
Addison Wesley Professional
Adobe Press
```

Books24x7 Advanced Search

Search: [Content: Full Text ▼]   [ ]   GO

Using: [Words in proximity ▼]

Sort Results By: [Relevance ▼]   ↕ reset form

Restrict the search to a subset of the library

Search In: ● All Books

                ○ My Bookshelf [All Folders ▼]

Specify additional constraints

Additional Fields: Include [all ▼] of the following criteria

[Content: Full Text ▼]	[Contains ▼]	[ ]	[Words in proximity ▼]
[Content: Code Listings ▼]	[Contains ▼]	[ ]	[Words in proximity ▼]
[Book Info: Titles ▼]	[Contains ▼]	[ ]	[Words in proximity ▼]
Copyright Year [▼]		Date [ ] e.g. 1999	

History of Safari

2002

2006

History of Safari

2002

$$\Rightarrow \text{ it can't be that bad!}$$

2006

Discuss!

Good News
Element retrieval has real-world uses!

- Element retrieval is most useful on longer documents.
- Documents should be independent.
- Both search interfaces are somewhat "thorough".

Wishful thinking: Can we cooperate with Safari or Books24x7?

Discuss!

Good News

Element retrieval has real-world uses!

- Element retrieval is most useful on longer documents.
- Documents should be independent.
- Both search interfaces are somewhat "thorough".

Wishful thinking: Can we cooperate with Safari or Books24x7?

Acknowledgements

- Screenshots from http://safari.oreilly.com, http://www.books24x7.com, http://acmsel.safaribooksonline.com, http://acm.books24x7.com, http://books.google.com
- Earth image by NASA/Visible Earth http://visibleearth.nasa.gov/
- Milky way image by NASA
- Bookshelf image by Anthony Georgeff, released under a CC licence http://flickr.com/photos/manthatcooks/12842921/