

Unlock Content™



XML and XQuery: Unlocking the Value of Content

- Welcome and introduction to Mark Logic
- XML and XQuery: unlocking the value of content
 - Stephen Buxton, director of product management for Mark Logic
 - Co-author, “Querying XML: XQuery, XPath, and SQL/XML in Context”
- Q&A

A Coming Generation of Content Apps

IDC
Analyze the Future

CUSTOMER NEEDS AND STRATEGIES

**The Future for Content Applications:
A Survey of Market Readiness and Technology Trends**

Geoffrey Bock, Contract Writer Melissa Webster

IDC OPINION

IDC sees a significant new trend in the marketplace for content technologies: the rise of content applications — applications that make intelligent use of unstructured information in order to automate a business process. Ever cheaper bandwidth, ubiquitous connectivity, the growth of mobile and wireless computing, cheaper storage, greater computing power at lower costs, higher-quality imaging, graphics boards, faster audio and video streaming, and others all contribute to the market receptivity for the next generation of content applications. In order for content applications to effectively make use of content in the context of automated business processes, the content itself must be prepared in such a way that it is intelligible to a computerized process. It is for this reason that IDC believes these next-generation solutions will be fueled by the adoption of XML-related tools and technologies. A recent survey of InfoWorld readers responsible for recommending or purchasing content technologies shows that:

- ☑ 76% need to develop content once and then reuse it in many different business situations.
- ☑ 64% need to store content for print publications and electronic distribution in a single repository.
- ☑ 61% need defined records management policies and procedures for archiving the content of their Web sites.
- ☑ 55% need to use well-tagged content to substantially improve how they manage and deliver unstructured information.
- ☑ XML adopters score significantly higher than nonadopters on these metrics and in their adoption of a long list of established and emerging content technologies.

File # 23451, Volume 1, Tab: Users

Out of Headquarters, 5 Spear Street, Framingham, MA 01701, USA P: 508.872.8200 F: 508.825.4015 www.idc.com

Business managers are overwhelmed by the sheer volume of information available to them, as they are forced to search for, cut/paste, or recreate content that could have been served to them *in context* .



Everything about content is hard: its hard to integrate, hard to find, hard to repurpose, and hard to deliver

It's All About the Infrastructure



	Data	Content
Data model	Relational	XML
Access method	SQL	XQuery
Platform	Oracle, DB2, SQL Server, etc.	MarkLogic Server, ???

Mark Logic provides of the industry's leading *XML content server*

- 30+ top-tier customers: *DISA, U.S. Air Force, DTIC, U.S. Army, Elsevier, JetBlue, Oxford University Press*
- 45+ employees
- Headquarters in San Mateo, California (Silicon Valley)
- Strong financial backing: *Sequoia Capital, Lehman Brothers*

Recognized Technology Leader



Gartner



- Named to Gartner's list of "Cool Vendors"
- Named a "Company to Watch" by Intelligent Enterprise magazine
- Named to KMWorld magazine's "100 Companies that Matter in Knowledge Management"
- Won 2005 Software and Information Industry Association Codie Award for Best Vertical Market Product or Service - Publishing
- Named a "Trend-Setting Product" by KMWorld magazine
- Named to prestigious list of companies that matter most in the digital content industry

Selected Mark Logic Customers



“Big, Fast, XML”

- **Big** – from gigabytes to terabytes ... and beyond
- **Fast** – capable of millisecond response times ... leveraging search-engine indexing techniques
- **XML** – designed specifically for XML content ... implementing the W3C-standard XQuery language

Mark Logic accelerates the creation and delivery of the specific information needed to accomplish vital missions

- Integrate content from diverse sources
- Repurpose content for multiple applications
- Enable dynamic custom views of content
- Deliver information through any channel or device
- Search-and-discover previously unknown information



Why Choose MarkLogic Server?



Only MarkLogic Server delivers

- “As-is” content loading
 - Does not require DTDs or schemas
 - No need to “shred” or “chunk”
 - Automatic conversion to XML
- Industry’s most powerful and complete XQuery engine
 - W3C standard for XML query
 - Flexible programming language designed for content
 - More powerful queries, more precise results
- High scalability and performance
 - Millisecond response times – multi-terabyte contentbases



- Welcome and introduction to Mark Logic
- XML and XQuery: unlocking the value of content
 - Stephen Buxton, director of product management for Mark Logic
 - Co-author, “Querying XML: XQuery, XPath, and SQL/XML in Context”
- Q&A

Content Applications

Physician Resource
tennis elbow
Sunday March 12th, 2006

ABSTRACT
Tennis elbow is a term first introduced in the 1880s to describe an inflammatory process that involves the radioulnar joint or lateral epicondyle of the humerus. It is a common exercise-related syndrome, and the mechanism is thought to be repetitive pronation and supination of the forearm. The actual pathologic nature of this syndrome is unclear. Repetitive torsion or overuse, tendinitis of the common extensor tendon, periostitis of the lateral epicondyle, and entrapment by scar tissue of the radial nerve have all been suggested as causes in this syndrome. Histologically, the abnormality has been described as angiolymphoid hyperplasia, a term subsequently modified to angiolymphoid tendinitis. Recently, the cause has been believed to be a degenerative process because of the paucity of acute inflammatory cells seen histologically. Medial tennis elbow and posterior tennis elbow have been reported, the former involving the posterior flexor and flexor carpi radialis insertions and the latter involving the triceps tendon. The following discussion pertains to lateral tennis elbow. **ROSEN'S EMERGENCY CONCEPTS AND CLINICAL PRACTICE**

Diagnosis

Message Traffic Analysis

Kenneth Lay [kenneth.lay@enron.com]

This person has been marked as a person of interest.

FREQUENTLY MENTIONED TOPICS

- Computer Security (1 mention)
- Reinforcement Theory (1 mention)
- Abstract: External Control Systems (1 mention)
- Federal Communications Commission (1 mention)
- Strategic Management Model (1 mention)
- Spaulding (1 mention)
- South Coast (1 mention)
- Knowledge Management (1 mention)
- Accounting Systems (1 mention)
- Research Results (1 mention)

INSIDER TRADES

Date	Action	Shares	@Price	Total
11/28/01	Sale	500,000	\$2.38	\$1,190,000
02/01/01	Sale	1,000,000	\$90.00	\$90,000,000
02/01/01	Exercise	95,630	\$21.76	\$1,999,760
07/01/01	Exercise	150,000	\$80.50	\$12,075,000
02/01/01	Exercise	120,000	\$20.00	\$2,400,000
01/01/01	Exercise	100,000	\$10.00	\$1,000,000
01/01/01	Sale	284,000	\$82.41	\$23,679,840
03/01/01	Exercise	60,000	\$18.75	\$1,125,000
02/01/01	Exercise	50,000	\$17.45	\$1,147,750
02/01/01	Exercise	15,000	\$17.45	\$261,750
07/01/01	Sale	185,000	\$76.79	\$14,185,570
03/01/01	Exercise	75,118	\$17.30	\$1,300,541
11/01/01	Exercise	120,111	\$17.30	\$2,077,911

Executive Pay Check

Executive Detail for "Steven P. Jobs":

Steven P. Jobs
Apple Computer, Inc.
Chief Executive Officer

Compensation Detail

Year	Salary	Bonus	All Other Compensation	Stock Options	Total Cash Compensation	Percentile
2002	\$1	\$2,248,498	\$1,302,795	7,500,000	\$3,071,494	98th
2001	\$1	\$43,511,534	\$40,484,594	-	\$83,996,129	99th
2000	\$1	-	-	20,000,000	\$1	1st

Executive Comparison

Company	Executive	Position	Total Cash Compensation	Percent Difference
Dell	Michael S. Dell	Chairman of the Board, Chief Executive Officer	\$3,436,473	-3.8%
Hewlett-Packard	Carleton S. Fiorina	Chairman and Chief Executive Officer	\$3,518,460	-1.5%
Microsoft	Steven A. Ballmer	Chief Executive Officer, Director	\$871,114	-75.6%

safariU v2

Organize The Content Of Your Book

Chapter	Title	Pages	Cost
Chapter 1	Pieces of the Puzzle	19	\$2.68
Chapter 2	Introduction	21	\$3.36
Chapter 3	Implementation	38	\$6.60
Chapter 4	Troubleshooting	10	\$1.60
Chapter 5	Security	09	\$4.00
Chapter 6	Applications	29	\$4.64
Chapter 7	Advanced Topics	18	\$2.88
Chapter 8	Case Study	19	\$3.04
Chapter 9	Performance Features	11	\$1.76

More, richer content



3 years

3 months

"locked-in" value

China - Google Search - Mozilla Firefox

File Edit View Go Bookmarks Yahoo! Tools Help

http://www.google.com/search?hl=en&lr=&q=China&btnG=Search

Google Search [Advanced Search](#) [Preferences](#)

Web Results 1 - 10 of about 1,800,000,000 for **China** [definition]. (0.11 seconds)

[News results for China](#) - [View today's top stories](#)

[Detroit uses top technology to win over China](#) - [International Herald Tribune](#) - 30 minutes ago
[Inhabitants make way for ecological restoration of China's top ...](#) - [Xinhua](#) - 54 minutes ago

[China General Information, China Information](#)
 An extensive source of information on **China**, its politics and society.
[www.chinatoday.com/](#) - 39k - [Cached](#) - [Similar pages](#)

[CIA - The World Factbook -- China](#)
 Features a map and brief descriptions of geography, economy, government, and people.
[www.cia.gov/cia/publications/factbook/geos/ch.html](#) - 131k - [Cached](#) - [Similar pages](#)

[Welcome to chinadaily.com.cn, your gateway to China](#)
 Chinadaily.com.cn is largest English portal in **China**, providing news, business info, BBS, learning materials and 中文国际新闻.
[www.chinadaily.com.cn/](#) - 108k - [Cached](#) - [Similar pages](#)

[China the Beautiful - Chinese Art and Literature](#)
 Primary resources of Chinese Classical Arts Literature Poetry Calligraphy Painting History Philosophy.
[www.chinapage.com/china.html](#) - 23k - [Cached](#) - [Similar pages](#)

[China Airlines](#)
 Country's international airline flight schedules, promotions, mileage program and other information about the airline.
[www.china-airlines.com/](#) - 3k - [Cached](#) - [Similar pages](#)

[ChinaOnline.com - The Information Network for China](#)
 Business news and analysis for Chinese marketplace

Done

Sponsored Links

[China](#)
 It may happen tomorrow get ready
 Don't lose your family or friends
[www.nokr.org](#)

[China children](#)
 Help change a child's life
 Sponsor a child who is waiting
[www.fosterparentsplan.ca](#)

[If you like The History Channel](#), you'll love this thriller about the "Clash of Civilizations"
[www.AWomanfromCairo.com](#)

[China](#)
 Find pics, news, movies, interviews filmography and more at Moviefone
[Moviefone.com](#)

[AMG Porcelathun](#)
 Fine **China**/Porcelain Dinner, Coffee Tea, Signature Sets from Bohemia
[www.porcelathun.com](#)
 San Francisco-Oakland-San Jose, CA

[CHINA](#)

database + search engine

=



or



??

XML
+ XQuery
+ XML content server

platform for content applications

Demos

Medbook

World Bank

Message Traffic Analysis

XML
+ XQuery
+ XML content server

platform for content applications

XML

+

XQuery

+

XML content server

platform for content applications

XML data, XML content

```
<?xml version="1.0"?>
<ipo:purchaseOrder xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:ipo="http://www.example.com/IPO" orderDate="1999-12-01">
  <shipTo exportCode="1" xsi:type="ipo:UKAddress">
    <name>Helen Zoe</name>
    <street>47 Eden Street</street>
    <city>Cambridge</city>
    <postcode>CB1 1JR</postcode>
  </shipTo>
  <billTo xsi:type="ipo:USAddress">
    <name>Robert Smith</name>
    <street>8 Oak Avenue</street>
    <city>Old Town</city>
    <state>PA</state>
    <zip>95819</zip>
  </billTo>
  <items>
    <item partNum="833-AA">
      <productName>Lapis necklace</productName>
      <quantity>1</quantity>
      <USPrice>99.95</USPrice>
      <ipo:shipComment>
        Use gold wrap if possible
      </ipo:shipComment>
      <ipo:customerComment>
        Want this for the holidays!
      </ipo:customerComment>
      <shipDate>1999-12-05</shipDate>
    </item>
  </items>
</ipo:purchaseOrder>
```

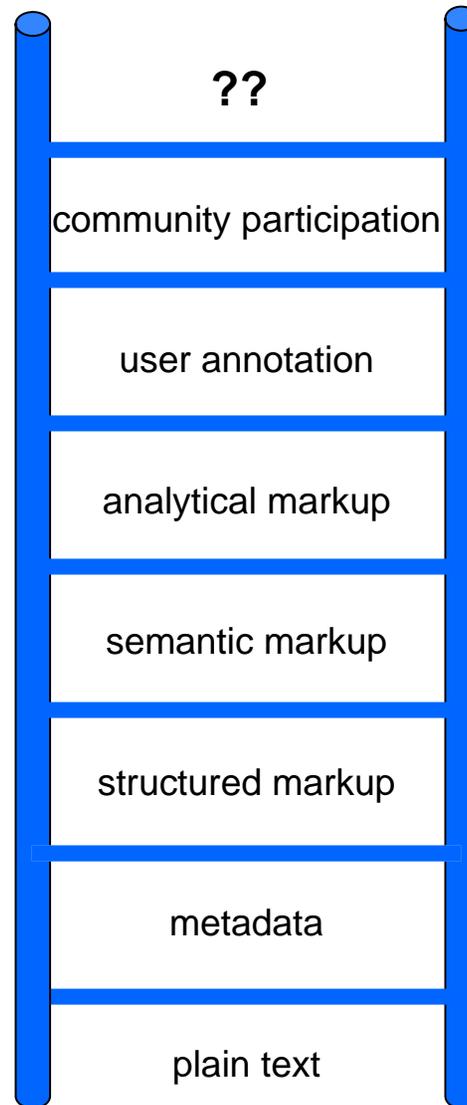
XML data, XML content

```
<!DOCTYPE book PUBLIC "-//OASIS//DTD DocBook XML V4.1.2//EN" "http://www.oasis-open.org/docbook/xml/4.1.2/docbookx.dtd">
<book>
  <bookinfo>
    <title>DocBook document example</title>
    <author>
      <firstname>Ovidiu</firstname>
      <surname>Predescu</surname>
    </author>
    <copyright>
      <year>2001</year>
      <holder>Ovidiu Predescu</holder>
    </copyright>
  </bookinfo>
  <preface>
    <title>Example document</title>
    <para>A simple DocBook example document.</para>
  </preface>
  <chapter>
    <title>DocBook and XSLT-process mode for GNU Emacs/XEmacs</title>
    <para>The <emphasis>XSLT-process</emphasis> mode, when used with the DocBook-XSL stylesheets makes GNU Emacs/XEmacs
the ideal DocBook processor.</para>
    <para>To find out more on how to use the <emphasis>XSLT-process</emphasis> mode for DocBook processing, point your browser
to <ulink url="http://xslt-process.sourceforge.net/docbook.php">http://xsltprocess.sourceforge.net/docbook.php</ulink>.</para>
  </chapter>
</book>
```

XML data, XML content

- Constraints
- Shredding
- Richness

XML content gets better and better



The XML eco-system

Authoring



Enrichment



Visualization/PDF



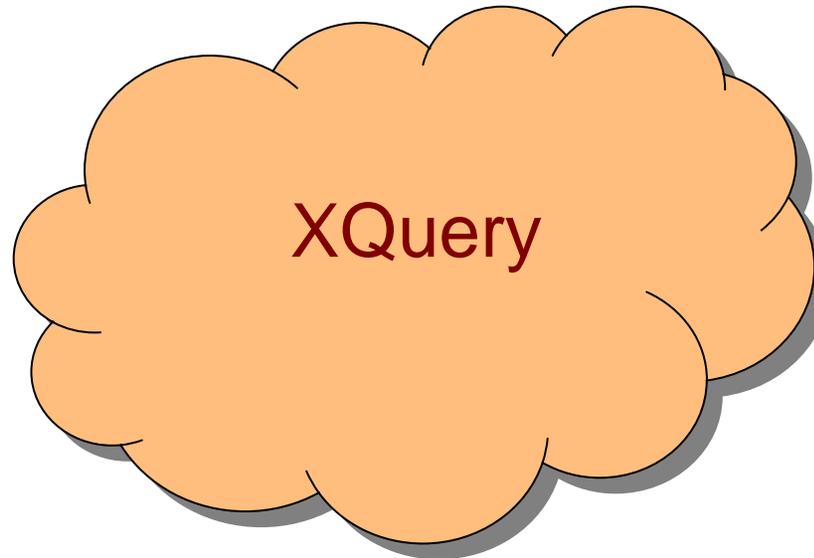
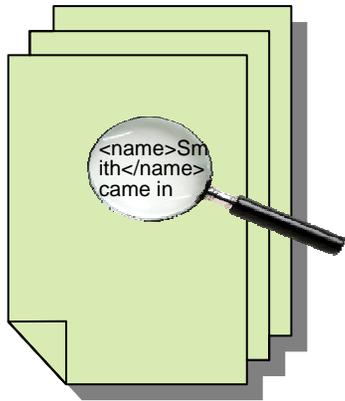
Services



XML
+ XQuery
+ XML content server

platform for content applications

Query, manipulate, and transform XML with XQuery



XML

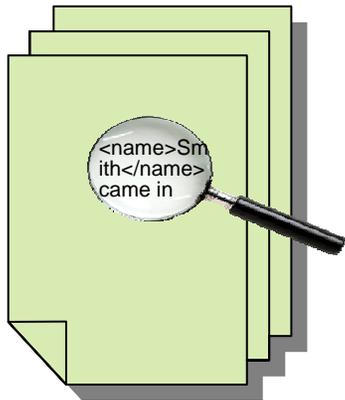
XHTML

RSS

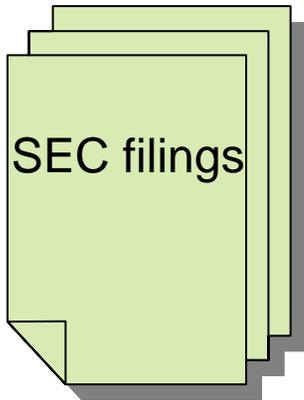
Web
Service

XSL:FO to
PDF

Mobile



Query, manipulate, and transform XML with XQuery



Convert:

- Get DEF 14A filings from SEC
- Get executive, competitor listings from Hoover's
- Load, automatically converting HTML → XML

Query:

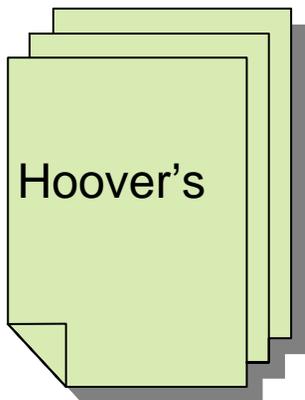
1. Identify and extract position title
2. Identify and extract paragraphs about Jobs
3. Identify and extract compensation table
8. Use Hoover's to identify top 3 competitors
9. Identify DEF-14A documents for competitors
10. Identify comparable executives
11. Extract their compensation information

Manipulate:

4. Identify table rows concerning Steven Jobs
5. Correlate table columns to generic standard
6. Calculate total cash compensation
7. Compare compensation to aggregate
12. Calculate comparative pay

Render:

13. Generate automatic cross-linkages
14. Generate output as HTML



Executive Pay Check, Powered by Mark Logic - Mozilla Firefox

http://paycheck.demo.marklogic.com/detailed-exec-comp.viz?company=Apple+Computer,+Inc.&exec=Steven+P.+Jobs

M A R K **Executive Pay Check**
L O G I C

Executive Detail for "Steven P. Jobs":

Steven P. Jobs
Apple Computer, Inc.
Chief Executive Officer

Executive Detail

Steven P. Jobs is one of the Company's co-founders and currently serves as its Chief Executive Officer. Mr. Jobs is also the Chairman and Chief Executive Officer of Pixar Animation Studios. In addition, Mr. Jobs co-founded NeXT Software, Inc. ("NeXT") and served as the Chairman and Chief Executive Officer of NeXT from 1985 until 1999 when NeXT was acquired by the Company.

Compensation Detail

Year	Salary	Bonus	All Other Compensation	Stock Options	Total Cash Compensation	Percentile
2002	\$1	\$2,268,698	\$1,302,798	7,500,000	\$3,571,494	98th
2001	\$1	\$43,511,534	\$40,484,594	-	\$83,996,129	99th
2000	\$1	-	-	20,000,000	\$1	1st

Executive Comparison

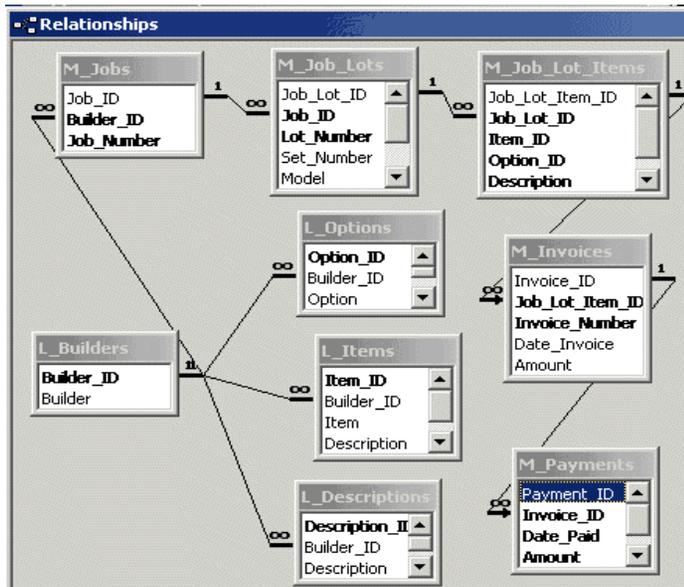
Compare and contrast Steven P. Jobs's compensation with that of executives at Apple Computer, Inc.'s top competitors:

Company	Executive	Position	Total Cash Compensation	Percent Difference
Dell	Michael S. Dell	Chairman of the Board, Chief Executive Officer	\$3,436,473	-3.8%
Hewlett-Packard	Carlton S. Fiorina	Chairman and Chief Executive Officer	\$3,518,460	-1.5%
Microsoft	Steven A. Ballmer	Chief Executive Officer, Director	\$871,114	-75.4%

Done

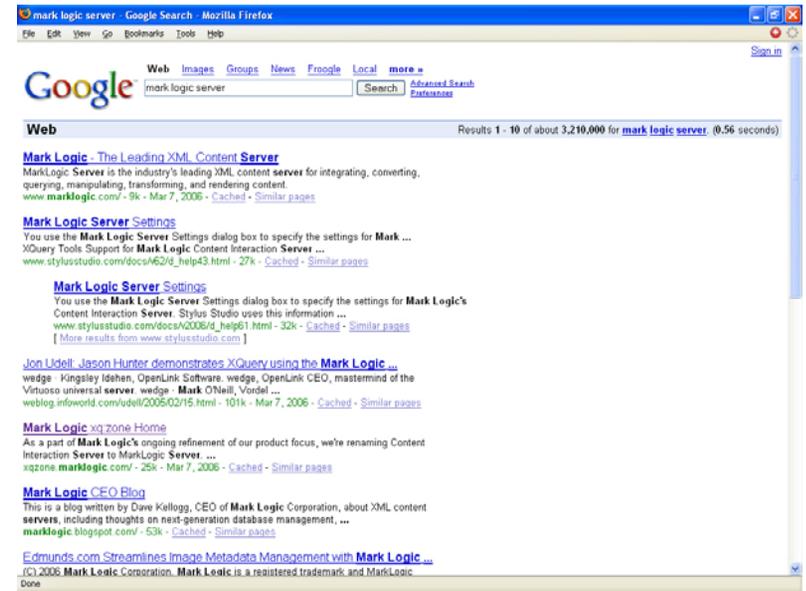
Databases, search engines, combinations fall short

Relational databases



Built for data

Search engines



Built for documents

Let's get specific

More than simple “XML-aware” full-text query:

- XML structure and phrases
- XML structure and keywords
- Fine-grained search and retrieval

Need more than “XML-aware” keyword search

Show me sections that say “retract the landing gear”

XML indicates a phrase break:

```
<step number="3">  
  Wait for the automatic pilot control to  
  retract.  
</step>  
<step number="4">  
  The landing gear should now be ready  
  to deploy.  
</step>
```

Phrase should not match, because of change in context:

```
<warning>  
  While waiting for the flaps to fully  
  retract<footnote>The landing gear need  
  to be stowed or else the flaps will  
  remain deployed</footnote>, ensure  
  that the automatic pilot control is at  
  hand.  
</warning>
```

XML does not indicate a phrase break:

```
<p>  
  It is vital that you retract <b>the landing  
  gear</b>, but leave the flaps fully  
  deployed, before attempting this type  
  of approach.  
</p>
```

Phrase should match *around* the change in context:

```
<step>  
  Partially retract<footnote>It is not  
  recommended to fully retract at this  
  point.</footnote> the landing gear.  
</step>
```

Need more than "fielded search"

Show me titles of books recommended by Oprah in 2006

```

<book>
  <name>Querying XML</name>
  <awards>
    <award>
      <name>NY Times best seller</name>
      <year>2006</year>
    </award>
    <award>
      <name>Recommended by Oprah</name>
      <year>2007</year>
    </award>
  </awards>
</book>
...

```

You need to be able to search this element

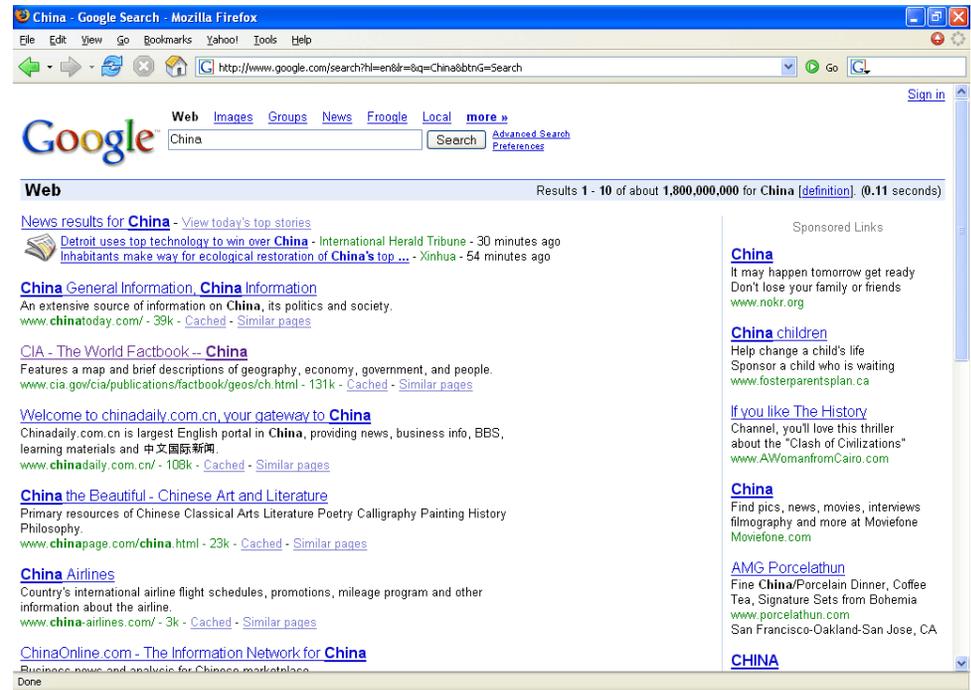
... and relate it to this parent

... and extract this ancestor (or not)

... and not get confused by this sibling

Arbitrary granularity

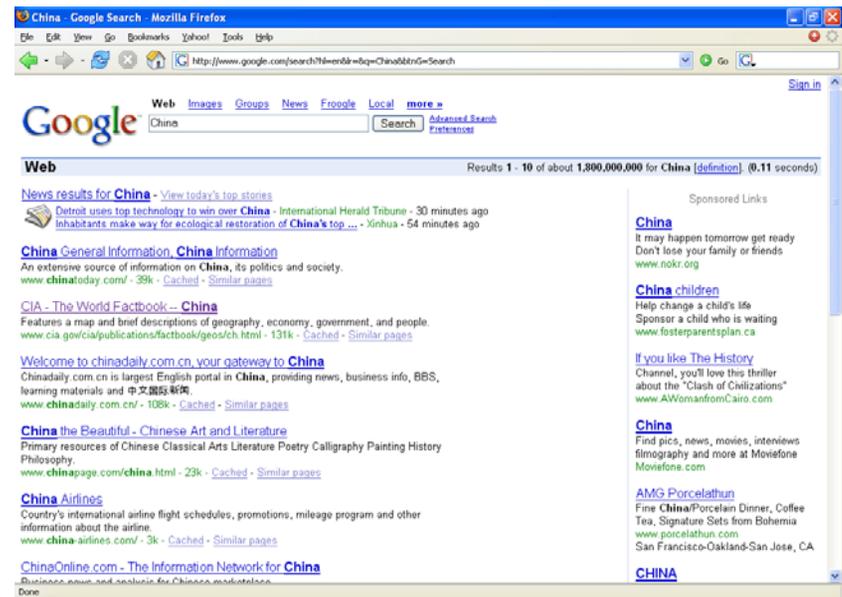
Show me all the documents that contain the word "China"



Fine-grained query

- Pre-process as metadata (search engine or database)
- “XML-aware full-text index”

Show me all the documents that contain the word “China” in the title.



Fine-grained query *and* retrieval

Show me all the sections that contain the word “China”.

```
<book>
  <title>Asian History</title>
  <section title="Introduction">
This book explains the history
of the ...
  </section>
  <section title="Middle Ages">
For centuries China stood as a
leading ...
  </section>
  <section title="Modern Day">
Today's China is ...
  </section>
</book>
```

```
<journal>
  <title>Business Today</title>
  <section title="Background">
The world economy is ...
  </section>
  <section title="Corporations">
Most American corporations
are still ...
  </section>
  <section title="Global">
The biggest companies in
China are ...
  </section>
</journal>
```

```
<news>
  <title>Baseball Update</title>
  <author>Tom China</author>
  <section>Today, the Red Sox
won their first ....
  </section>
  <section>Baseball to 'review'
Bonds allegations ...
  </section>
  <section>Canada stuns U.S.
with World Baseball challenge
match win ...
  </section>
</news>
```

Document-level query

Show me all the sections that contain the word “China”.

```
<book>
  <title>Asian History</title>
  <section title="Introduction">
This book explains the history
of the ...
  </section>
  <section title="Middle Ages">
For centuries China stood as a
leading ...
  </section>
  <section title="Modern Day">
Today's China is ...
  </section>
</book>
```

```
<journal>
  <title>Business Today</title>
  <section title="Background">
The world economy is ...
  </section>
  <section title="Corporations">
Most American corporations
are still ...
  </section>
  <section title="Global">
The biggest companies in
China are ...
  </section>
</journal>
```

```
<news>
  <title>Baseball Update</title>
  <author>Tom China</author>
  <section>Today, the Red Sox
won their first ....
  </section>
  <section>Baseball to 'review'
Bonds allegations ...
  </section>
  <section>Canada stuns U.S.
with World Baseball challenge
match win ...
  </section>
</news>
```

Classical search engine solution.

Fine-grained query

Show me all the sections that contain the word “China”.

```
<book>
  <title>Asian History</title>
  <section title="Introduction">
This book explains the history
of the ...
  </section>
  <section title="Middle Ages">
For centuries China stood as a
leading ...
  </section>
  <section title="Modern Day">
Today's China is ...
  </section>
</book>
```

```
<journal>
  <title>Business Today</title>
  <section title="Background">
The world economy is ...
  </section>
  <section title="Corporations">
Most American corporations
are still ...
  </section>
  <section title="Global">
The biggest companies in
China are ...
  </section>
</journal>
```

```
<news>
  <title>Baseball Update</title>
  <author>Tom China</author>
  <section>Today, the Red Sox
won their first ....
  </section>
  <section>Baseball to 'review'
Bonds allegations ...
  </section>
  <section>Canada stuns U.S.
with World Baseball challenge
match win ...
  </section>
</news>
```

Database or search engine with “XML-aware” full-text.

Fine-grained query and retrieval, separate

Show me all the sections that contain the word "China".

```
<book>
  <title>Asian History</title>
  <section title="Introduction">
    This book explains the history
    of the ...
  </section>
  <section title="Middle Ages">
    For centuries China stood as a
    leading ...
  </section>
  <section title="Modern Day">
    Today's China is ...
  </section>
</book>
```

```
<journal>
  <title>Business Today</title>
  <section title="Background">
    The world economy is ...
  </section>
  <section title="Corporations">
    Most American corporations
    are still ...
  </section>
  <section title="Global">
    The biggest companies in
    China are ...
  </section>
</journal>
```

```
<news>
  <title>Baseball Update</title>
  <author>Tom China</author>
  <section>Today, the Red Sox
  won their first ....
  </section>
  <section>Baseball to 'review'
  Bonds allegations ...
  </section>
  <section>Canada stuns U.S.
  with World Baseball challenge
  match win ...
  </section>
</news>
```

Database with XQuery + separate full-text.

Fine-grained query and retrieval, integrated

Show me all the sections that contain the word "China".

```
<book>
  <title>Asian History</title>
  <section title="Introduction">
    This book explains the history
    of the ...
  </section>
```

```
<section title="Middle Ages">
  For centuries China stood as a
  leading ...
</section>
<section title="Modern Day">
  Today's China is ...
</section>
```

```
</book>
```

```
<journal>
  <title>Business Today</title>
  <section title="Background">
    The world economy is ...
  </section>
  <section title="Corporations">
    Most American corporations
    are still ...
  </section>
```

```
<section title="Global">
  The biggest companies in
  China are ...
</section>
```

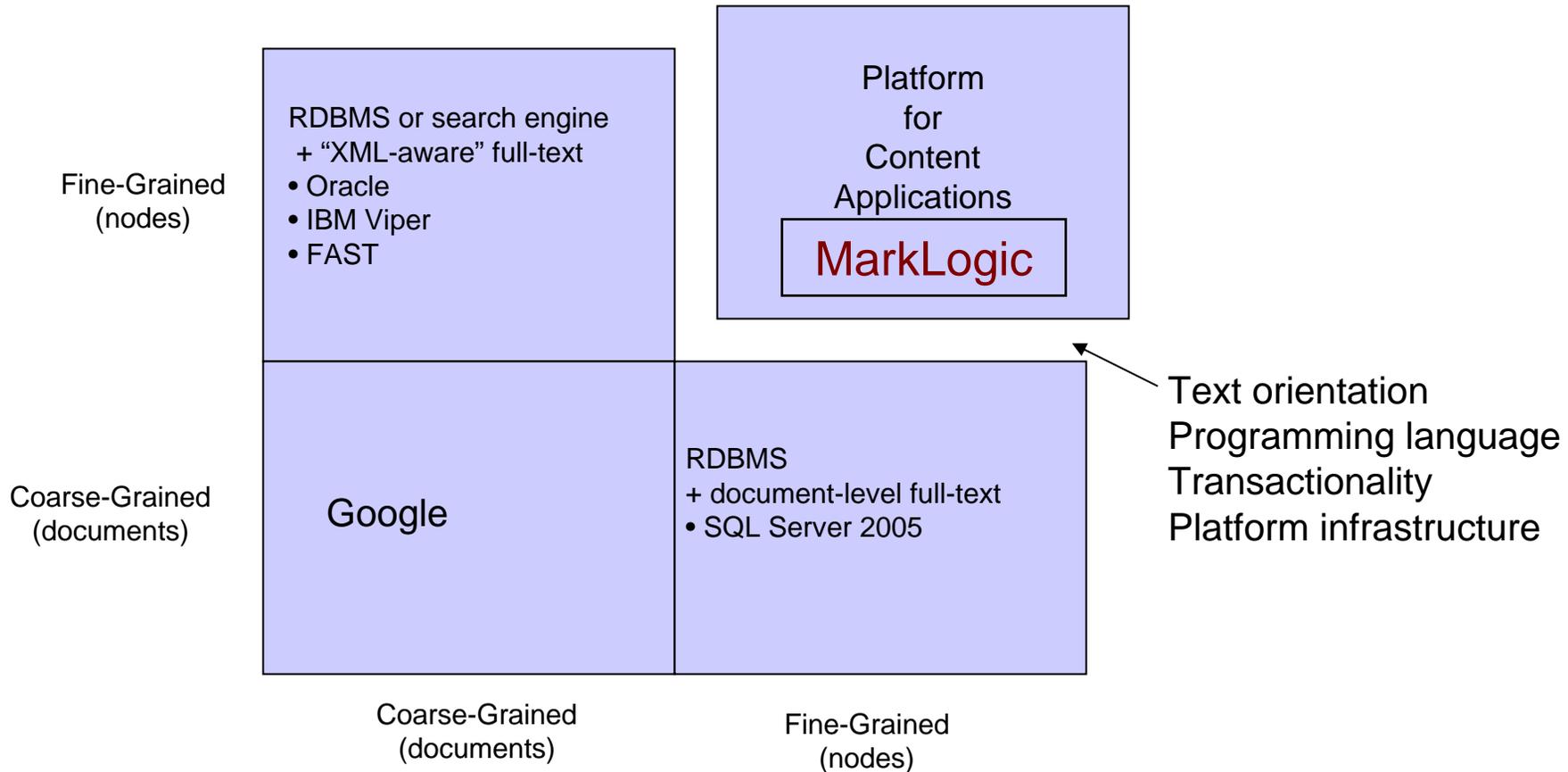
```
</journal>
```

```
<news>
  <title>Baseball Update</title>
  <author>Tom China</author>
  <section>Today, the Red Sox
  won their first ....
  </section>
  <section>Baseball to 'review'
  Bonds allegations ...
  </section>
  <section>Canada stuns U.S.
  with World Baseball challenge
  match win ...
  </section>
</news>
```

Contentbase with XQuery + integrated full-text.

Combinations

Granular Search



Granular Retrieval

XML
+ XQuery
+ XML content server

platform for content applications

MarkLogic is architected for Big, Fast, XML content



Content applications in the wild

Current Project Title: **Kerberos**

1 Gather Content **2 Organize Content** 3 Choose Options 4 Review 5 Order Books

Organize The Content Of Your Book

[Upload Your Own Materials](#)

Current Estimated Totals*: Pages: 213 Cost: \$34.08

Table of Contents		Add Chapter or Section Title	
Custom Book Content		Pages	Cost
▶[Chapter 1] Pieces of the Puzzle	edit remove	13	\$2.08
▶[Chapter 2] Protocols	edit remove	21	\$3.36
▶[Chapter 3] Implementation	edit remove	38	\$6.08
▶[Chapter 4] Troubleshooting	edit remove	10	\$1.60
▶[Chapter 5] Security	edit remove	30	\$4.80
▶[Chapter 6] Applications	edit remove	29	\$4.64
▶[Chapter 7] Advanced Topics	edit remove	18	\$2.88
▶[Chapter 8] Case Study	edit remove	19	\$3.04
▶[Chapter 9] Kerberos Futures	edit remove	11	\$1.76

Your Selections

Empty

No Content has been added.

- Welcome and introduction to Mark Logic
 - XML and XQuery: unlocking the value of content
 - Stephen Buxton, director of product management for Mark Logic
 - Co-author, “Querying XML: XQuery, XPath, and SQL/XML in Context”
- Q&A

Thank You

www.marklogic.com

