Web Publishing 2.0
Jason Hunter
Introductions

• Jason Hunter
  • Principal Technologist at Mark Logic
  • jhunter@marklogic.com
  • http://developer.marklogic.com
  • http://www.marklogic.com
Mark Logic

- Enterprise grade XML Content Server
- Load, Query, Manipulate, Render
If we're moving toward Web 2.0, what does that mean for web publishing?
The Trends

- Answers Not Links
- Sweat the Content
- Content in Context
- Emphasis on Google
- User Participation
- Personalized Access
- Leveraging the Structure
- Enrichment
- Content Analytics
- Agility
The Business Situation

• Increasing content sizes
  • More *public* content, more *private* content, and more *government* content
• Increasing user expectations
  • Easy, immediate, searchable access to all content
CHAPTER I
VARIATION UNDER DOMESTICATION

Causes of variability—Effects of habit and the use or disuse of parts—Correlated variation—Inheritance—Character of domestic varieties—Difficulty of distinguishing between varieties and species—Origin of domestic varieties from one or more species—Domestic pigeons, their differences and origin—Principles of selection, anciently followed, and their effects—Methodical and unconscious selection—Unknown origin of our domestic productions—Circumstances favourable to man’s power of selection.

Causes of Variability

When we compare the individuals of the same variety or sub-variety of our older cultivated plants and animals, one of the first points which strikes us is, that they generally differ more from each other than do the individuals of any one species or variety in a state of nature. And if we reflect on the vast diversity of the plants and animals which have been cultivated, and which have varied during all ages under the most different climates and treatment, we are driven to conclude that this great variability is due to our domestic productions having been raised under conditions of life not so uniform as, and somewhat different from, those to which the parent species had been exposed under nature. There is, also, some probability in the view propounded by Andrew Knight, that this variability may be partly connected with excess of food. It seems clear that organic beings must be exposed during several generations to new conditions of life to cause any great amount of variation; and that, when the organisation has once begun to vary, it generally continues varying for many generations. No case is on record of a variable organism ceasing to vary under cultivation. Our oldest cultivated plants, such as wheat, still yield new varieties; our oldest domesticated animals are still capable of rapid improvement or modification.

As far as I am able to judge, after long attending to the subject, the conditions of life appear to act in two ways,—directly on the whole organisation or on certain parts alone, and indirectly by affecting the reproductive system. With respect to the direct action, we must bear in mind that in every case, as Professor Weismann has lately insisted, and as I have...
ON THE
ORIGIN OF SPECIES.

"But with regard to the material world, we can at least go so far as this—we can perceive that events are brought about not by insulated interpositions of Divine power, exerted in each particular case, but by the establishment of general laws."

W. WHEWELL: Bridgewater Treatise.

"To conclude, therefore, let no man out of a weak conceit of sobriety, or an ill-applied moderation, think or maintain, that a man can search too far or be too well studied in the book of God's word, or in the book of God's works; divinity or philosophy; but rather let men endeavour an endless progress or proficience in both."

BACON: Advancement of Learning.

Down, Bromley, Kent,
October 1st, 1859.
Technology Situation

- Standardization on XML
- Web, blogs, books, articles, citations
- Expanding XML processing capabilities
- Why XML?
  - Ordered, hierarchical, textual, irregular, not naturally relational
Major Trends
Answers Not Links

- People want to be one step away from answers, not two
Answers Not Links

- People want to be one step away from answers, not two
Answers Not Links

• People want to be one step away from answers, not two
Answers Not Links

- People want to be one step away from answers, not two
Answers Not Links

• As a publisher, you want to do this as well
• Leveraging your XML structure
From **Greasemonkey Hacks**
Chapter 5: Developer Tools
By Mark Pilgrim  ·  Published November 01, 2005  ·  Statistics

Save the following user script as `xhrRequest-tracing.user.js`:

```javascript
// ==UserScript==
// @name XMLHttpRequest Tracing
// @namespace  http://blog.monstuff.com/archives/cat_greasemonkey.html
// @description Trace XMLHttpRequest XMLHttpRequestlogging calls into a JavaScript console
calls
// @include http://pick.some.domain

XMLHttpRequest.prototype.uniqueID = function() {
    if (!this.uniqueIDMemo) {

```
### Differential Diagnosis

#### Follicular Carcinoma

<table>
<thead>
<tr>
<th>Name</th>
<th>+/-</th>
<th>% positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thyroglobulin</td>
<td>+</td>
<td>96</td>
</tr>
<tr>
<td>TTF-1</td>
<td>+</td>
<td>100</td>
</tr>
</tbody>
</table>

**Diagnostic Pearls**
- TTF-1 positive
- Thyroglobulin positive

**Additional stain information for Follicular Carcinoma**

#### Papillary Carcinoma

<table>
<thead>
<tr>
<th>Name</th>
<th>+/-</th>
<th>% positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>CK7</td>
<td>+</td>
<td>100</td>
</tr>
<tr>
<td>TTF-1</td>
<td>+</td>
<td>100</td>
</tr>
<tr>
<td>CK19</td>
<td>+</td>
<td>100</td>
</tr>
</tbody>
</table>

**Diagnostic Pearls**
- Nuclei are TTF positive

**Additional stain information for Papillary Carcinoma**

### Add a diagnosis

- Adenomatoid Goiter
- Adenomatous Hyperplasia
- Anaplastic Carcinoma
- Angiosarcoma [Liver]
- Autoimmune Thyroiditis
- Basedow’s Disease
- Benign Bile Duct Proliferations
- Benign Bile Duct Proliferations
- Benign Infantile Hemangioendothelioma
- Bile Duct Adenoma
- Bile Duct Carcinoma
- Bile Duct Hamartoma
- Biliary Adenofibroma
- Biliary Cystadenoma and Cystadenocarcinoma
Sweat the Content

- You have two ways to make more money
  - 1. Create more content
  - 2. Do more with the content you have
Sweat the Content:

Microproducts
Sweat the Content:

Microproducts

AT A GLANCE
Tubman, Harriet Ross

- Sex: Female
- Born: Dorchester County, Maryland c.1820
- Died: Auburn, New York 10 March 1913
- Activity/Profession: Abolitionist, Slave

6 articles on Tubman, Harriet Ross

Tubman, Harriet

Source: Encyclopedia of African American History, 1619-1895: From the Colonial Period to the Age of Frederick Douglass

Word Count: 732 Includes: Bibliography

(b. c. 1821; d. 10 March 1913), legendary Underground Railroad conductor. Born Araminta "Minty" Ross in Dorchester County, Maryland, to the slaves Harriet Greene and Benjamin Ross, Harriet Tubman spent her early childhood as a housekeeper on the large plantation of Edward Brodas in Maryland. Disabled by physical abuse and starvation, she moved into the fields at the age of thirteen. Two years later she nearly died when an overseer struck her on the head after she blocked his pursuit of a fleeing slave. The attack left the slight, five-foot-tall Ross with a permanently convex forehead and a lifelong semonolence to accompany her missing upper front teeth. By 1844 she had recovered enough to be forced to marry the freeman John Tubman. In 1849, upon her owner's death, Tubman faced separation from her family ... Read full article

Tubman, Harriet Ross


Word Count: 1051

1822–1913

American abolitionist: who escaped from slavery and returned repeatedly to the South to lead other...
Sweat the Content:

Custom Publishing
Sweat the Content:

Custom Publishing

Table of Contents

Current Estimated Totals:
Pages: 205
Cost: $32.80

Custom Book Content

*Chapter 1* Advanced XML Hacks

- Hacks #91–100
  - by Michael Fitzgerald; *XML Hacks*
- Pipeline XML with Ant
  - by Michael Fitzgerald; *XML Hacks*
- Use Elements Instead of Entities to Avoid the "amp
  Explosion Problem"
  - by Michael Fitzgerald; *XML Hacks*
- Use Cocoon to Create a Well-Formed View of a Web Page,
  Then Scrape It for Data
  - by Michael Fitzgerald; *XML Hacks*
- From Wiki to XML, Through SGML
  - by Michael Fitzgerald; *XML Hacks*
- Create Well-Formed XML with JavaScript
  - by Michael Fitzgerald; *XML Hacks*
- Inspect and Edit XML Documents with the Document
  Object Model
  - by Michael Fitzgerald; *XML Hacks*
- Processing XML with SAX
  - by Michael Fitzgerald; *XML Hacks*
Sweat the Content:

Amusement
Sweat the Content:

Amusement
Content in Context

- Location of text, situation of the user, historical context
Content in Context

- Location of text, situation of the user, historical context

Important discussion of "milrinone"...

From article:
Medical Treatment of Peripheral Arterial Disease and Claudication (May 24, 2001)

... Because of the experience with milrinone, the cilostazol label includes a black-box warning that cilostazol should not be given to patients with claudication who also have heart failure. [Jump]

From article:
Heart Failure (May 15, 2003)

... Inotropes such as dobutamine and milrinone increase myocardial contractility. [Jump]
Content in Context

- Location of text, situation of the user, historical context
The language of this bill would change 20 laws

- PL 109-59 — SAFETEA-LU (5 changes)
- PL 109-53 — Dominican Republic-Central America-United States Free Trade Agreement Implementation Act (1 change)
  - 19 USC 4034 (1 change)
- PL 108-429 — Miscellaneous Trade and Technical Corrections Act of 2004 (4 changes)
  - 7 USC 7101 (4 changes)
- PL 108-218 — Pension Funding Equity Act of 2004 (1 change)
- PL 107-210 — Trade Act of 2002 (2 changes)
  - 19 USC 3807 (2 changes)
- PL 106-200 — Trade and Development Act of 2000 (1 change)
- PL 105-34 — Taxpayer Relief Act of 1997 (2 changes)
- PL 103-465 — Uruguay Round Agreements Act (2 changes)
- PL 99-272 — COBRA (1 change)
  - 19 USC 58c (1 change)
- PL 93-618 — Trade Act of 1974 (4 changes)
  - 19 USC 2317 (1 change)
Emphasis on Google

- Fear: Owning the content, the user
- Opportunity: Better search
- Opportunity: Instant AdWords registration
- Opportunity: Personalized landing pages
User Participation

- Direct: blogging, reviewing, tagging
- Indirect: Search and guidance based on collective intelligence
Leveraging the Structure

• For search of course
• For browse
• For offline access
• For special views
• For historical analysis
• For printing actual books
Novel Risk Markers and Clinical Practice

October 23, 2003

Teri Manolio, M.D., Ph.D.

A growing understanding of the importance of the rupture of atherosclerotic plaque in the pathogenesis of coronary events has led to the identification of an expanding array of markers of plaque vulnerability. Reports of markers of inflammation and coronary risk are appearing with increasing frequency; such markers include adhesion molecules, metalloproteinases, and as described in this issue of the Journal, myeloperoxidase (reported by Brennan et al., pages 1595–1604) and glutathione peroxidase 1 (reported by Blankenberg et al., pages 1605–1613). What should the clinician look for in these reports to determine their implications for patient care?

First, the measure should add independent information about the risk or prognosis. Risk can be assessed with the use of case–control and prospective cohort designs (see Figure), both of which are valid if they are carefully designed. These approaches produce estimates in the form of relative risks, odds ratios, or hazard ratios, comparing the incidence of disease among persons who have a given risk factor with the risk among those who do not have that risk factor. These measures should be adjusted mathematically for other factors that may account for some of the observed risk. Significant risk estimates, fully adjusted for other known confounders (not just age and sex), indicate that the marker adds independent information and thus has potential clinical importance. Applying this criterion to the study of myeloperoxidase reported by Brennan et al. and the study of glutathione peroxidase 1 reported by Blankenberg et al. shows that both markers provide independent information about risk.
Prognostic Value of Myeloperoxidase in Acute Coronary Syndromes

Inflammation appears to have a key role in acute coronary syndromes. Myeloperoxidase, an enzyme that generates reactive oxygen species, is released from leukocytes on activation, and plasma levels of myeloperoxidase may serve as a marker of inflammation. In this study, plasma myeloperoxidase levels were found to be predictive of subsequent coronary events in patients with chest pain, even when patients were initially negative for troponin T.

Plasma myeloperoxidase levels may be of use in risk stratification among patients with suspected acute coronary syndromes.

Extracted Tables and Figures

First, the measure should add independent information about the risk or prognosis. Risk can be assessed with the use of case–control and prospective cohort designs (see Figure), both of which are valid if they are carefully designed. These approaches produce estimates in the form of relative risks, odds ratios, or hazard ratios, comparing the incidence of disease among persons who have a given risk factor with the risk among those who do not have that risk factor. These measures should be adjusted mathematically for other factors that may account for some of the observed risk. Significant risk estimates, fully adjusted for other known confounders (not just age and sex), indicate that the marker adds independent information and thus has potential clinical importance. Applying this criterion to the study of myeloperoxidase reported by Brennan et al. and the study of glutathione peroxidase 1 reported by Blankenberg et al. shows that both markers provide independent information about risk.
Enrichment

- Adding Structure:
  - People, Places, Companies, Facts, ...
Build Query:

paris [delete]  [As person: Paris Hilton]  [As city: Paris]

By Source:
www.infoplease.com (22)  bbc.co.uk (7)  cnn.com (2)  aljazeera.net (1)  reuters.co.uk (1)

By Date:
2006 (26)  2005 (7)

33 Results (Showing top 10):

**BBC - Music / Features - Paris in the 1920s**
Hello Paris The City!

See entities
bbc.co.uk/news/358716615-33746-16860-1243492.xml - Nov 11, 2005

**Jacobins**
Maximilien Marie Isidore Robespierre: Early Life - Early Life A poor youth, he was enabled to study law in Paris through a scholarship. He won ...

See entities
Build Query:

napoleon [delete]  [As person: Napoleon Bonaparte; Napoleon I; Napoleon II; Napoleon III]
City: Paris [delete]

Require Term Proximity:
100 words (8) 20 words (6) 5 words (4)

By Source:
www.infoplease.com (9)

By Date:
2006 (8) 2005 (1)

9 Results:

August 15 Birthdays: Napoleon I
entered Paris (March 20), beginning his ephemeral rule of the Hundred Days. Attempt Bibliography: See Pieter Geyl, Napoleon: For and Against (1949); studies of Napoleon and his era by J. C. Herold (1955), Louis Bergeron (1981), Jean Tulard (1971), Owen...

People: Napoleon (39), Bonaparte (13), Napoleon I (2), Josephine (2), Nelson (2)
Cities: Paris (5), Vienna (4), Amiens (2), Moscow (2), Toulon (1)
Countries: France (19), Russia (8), Italy (7), Austria (6), Britain (5)
Organizations: Congress (3), Second Coalition (2), Homework Center (1), Fifty States (1), al-Qaeda (1)

Content Analytics

- Find the Haystack
Agility

- We're feeling our way forward
- People have dreams, but aren't sure what will work and what won't
- Must be agile in design, in development, and in business
- *The Red Queen Paradox:* You must run as fast as you can just to stay in place
The Trends

- Answers Not Links
- Sweat the Content
- Content in Context
- Emphasis on Google
- User Participation
- Personalized Access
- Leveraging the Structure
- Enrichment
- Content Analytics
- Agility
Contact Information

- Jason Hunter
  - Principal Technologist at Mark Logic
  - jhunter@marklogic.com
  - http://developer.marklogic.com
  - http://www.marklogic.com