



## e-Newsletter for October, 1999

588 Subscribers

### In This Issue

WF Seminars & Workshops

How Connected is the Web?

XML Vocabularies

XFRML Working Group

SAP SEM

Dublin Core goes Standard

Invisible Worlds

Internet Spy Software

Another twURL for WF

HTML goes XML

The Web in Perspective

R. Hackathorn  
editor

Comments?

Subscribe To  
This Newsletter

Previous Issues

PDF Version

Home Page for  
WebFarming.com

The Web  
Farming Book

An Introduction  
to Web Farming

### WF Seminars in Europe

Several web farming seminars are scheduled in Europe during October. First, IT Works will sponsor a one-day seminar entitled **Web Farming for the Data Warehouse** in Brussels on October 19.

[more ►](#) Second, **Technology Transer** will sponsor a two-day seminar entitled **Web Farming for Systematic Business Intelligence** on October 21-22 in Rome.

[more ►](#) Finally, TT and **Database Associates** are sponsoring a three-day Web Development 2000 Conference in Rome on October 26-28. As part of this conference, there is a half-day tutorial on **XML: Taming the Data Chaos on the Web.** [more ►](#)

### WF Workshop Series

The WF workshop series recently offered a one-day treatment of **Visible versus Hidden Web: Practices and Resources**, sponsored by **US West Advanced Technologies**. Check for similar offerings in the future. [more ►](#)

### SAP Business Info Collector

**SAP** again leads major vendors in the adoption of web farming techniques. Their latest version of the **Strategic Enterprise Management (SEM)** product enhances the Business Information Warehouse with a module called the **Business Information Collection**. The BIC provides "automated and semi-automated collection of structured and especially unstructured business information from internal and external sources. This includes an automatic search for relevant business information in the Internet, and the structuring of any information found." Sure sounds like web farming! [more ►](#)

### Invisible Worlds

A visit to **Carl Malamud's Invisible Worlds** site is a MUST for web farmers. Browse the Mappa Mundi magazine. The articles on *The Importance of Being EDGAR* and *Been There, Done That* give insights into the difficulties of making public information publicly available on the Web. Dig into their definition of Rocket Science and its [Space.cgi](#) interface. Play with the Visual Search of [EDGARspace](#). Download Danny Goodman's [SpaceKit Viewer](#) and even hack the JavaScript code! This is one

### How Connected is the Web?

A recent study by **Prof. Barabasi**, a physicist at **Notre Dame**, indicates that the Web is highly connected. On the average, about 19 links were needed to surf from one page to another. The study was based upon the *nd.edu* domain (consisting of 325,729 pages and 1,469,680 links), along with *whitehouse.gov*, *yahoo.com*, and *snu.ac.kr*. The results are encouraging for focused crawling techniques as the next generation of search engines. [more ►](#)

### XML Vocabularies

Ever wonder about tracking the exploding number of XML vocabularies? Bookmark the [XML.org Catalog](#). It is a precursor to their repository to be operational next year. There are currently 68 different organizations working on or publishing an XML vocabulary. Amazing!

### XFRML Working Group

Yet another XML vocabulary! **XFRML** (or the XML Financial Reporting Markup Language) is a standard way to publish and analyze corporate financial reports. Its supporters are impressive. Born within the **American Institute of Certified Public Accountants**, they are: **Arthur Andersen, Deloitte & Touche, Ernst & Young, KPMG, PricewaterhouseCoopers, FRx Software, Great Plains, Interleaf e-content, FreeEDGAR.com, EDGAR Online, Woodburn Group, and Microsoft.**

Take the tour. They have a well designed website that should serve as a model for other XML communities. [more ►](#)

### Dublin Core goes Standard

The **Dublin Core Metadata Initiative** has made version 1.1 a *recommendation* implying that the standard is stable and ready for adoption. The Dublin Core is the best XML/RDF vocabulary for describing web-based resources, both for publishing and discovery. The Dublin Core has widespread support from libraries, museums, government agencies, and commercial content providers. [more ►](#)

### Internet Spy Software

Have you recently received unsolicited email pushing Internet spy software? The claims are outrageous (but maybe accurate), such as: "Find that girl you met in traffic! Find out secrets about your

site to watch frequently.

### Another twURL for WF

**Susan Gerhart** of **twURLed World** has updated the WebFarming analysis. She used the set of global search engines to generate 5,000 references related to WebFarming, refined it to 1,400 and finally down to 800 'useful' ones. Review the resulting creative analyses. [more ▶](#)

### The Web in Perspective

An article by **Peter Lewis** in the **New York Times** on September 30 opens with the following perspective of the Web: *"Image New York's Greenwich Village with millions of jumbled streets, add Tokyo's chaotic street number system, have it grow faster than Las Vegas with a million new addresses a day, add all the languages of the United Nations, and make sure that every map is hopelessly outdated. That is today's World Wide Web."*

relatives, friends, enemies, co-workers, neighbors, and anyone else...even your spouse! Learn all about your mysterious neighbors! Get my FBI file straight from the FBI for free!" Please [email me](#) about your experiences with any of these products. I will summarize in a later issue.

### HTML goes XML

In August, **W3C** released **XHTML** 1.0 for review. XHTML is a redefinition of HTML 4.0 using **XML** 1.0, allowing users to blend HTML within XML. XHTML brings the rigor of XML to provide "richer Web pages on an ever increasing range of browser platforms including cell phones, televisions, cars, wallet sized wireless communicators, kiosks, and desktops." [more ▶](#)