



**Newsletter - August, 1999**

**In This Issue**

- WF Workshops
- XML Extender in DB2
- WF for Resumes and Jobs
- Seminars on WF
- Searchable Public Records
- Intarka Prospect-Miner
- WF Book Review
- Globalization To The Max
- WF at Siebel
- WebData Databases
- Focused Crawling
- Latest Web Assessment
- Copernic Server

**R. Hackathorn**  
editor

[Take a Quick Survey about this Newsletter](#)

[Comments?](#)

[Subscribe To This Newsletter](#)

[Previous Issues](#)

[PDF Version](#)

[Home Page for WebFarming.com](#)

[The Web Farming Book](#)

**WF Workshop Series**

A Web Farming workshop series starts on September 9 with a one-day treatment of **Global Discovery Services: Use and Misuse**. Frustrated with HotBot, AltaVista and the like. This is a MUST for you!

Sponsored by **US West Advanced Technologies**, it is conducted at their Boulder facilities. [more ▶](#)

**Seminars on Web Farming**

Three-day seminars on **Building Web Farming Systems: Methods & Tools** are offered on:

- **September 15-17** in San Francisco
- **November 10-12** in Dallas

For full details and registration, see the **DCI online brochure**. Sign up NOW. Space is limited. Content is awesome!

**Intarka ProspectMiner**

Intarka is a startup focused on using web farming technology in various settings. Their first product is ProspectMiner that discovers and filters sales prospects using a combination of keyword searching, relevance analysis, and learning feedback. Funded by New Enterprise Associates, they have 38 employees at their San Jose and West Bengal offices. Read their white paper on ProspectMiner. It is one of the better business justifications for a topic-specific web farming application.



**WF at Siebel**

**Siebel** is following in **SAP** footsteps. (see the May issue.) As a major vendor of sales automation and similar systems, Siebel offers InterActive Briefings which gathers information on company profiles, business news, subsidiaries, and affiliates from web-based sources. In addition, Siebel just signed an agreement with **Dun & Bradstreet** to enhance their external data. [more ▶](#)



**Focused Crawling**

Another **IBM Almaden** project is leading the way to better web farming technology. As the Best Paper Prize at the recent **WWW8 Conference**, this paper describes the use of custom crawling and link

**XML Extender in DB2**

In the June issue, a brief mention was made about the new **XML Extender in IBM DB2 UDB**. Here is a longer description of the functionality supported by this extender, as it is going through beta testing. [more ▶](#)

**WF for Resumes/Jobs**

The management recruiting industry is hot! With a tight labor market for high tech personnel, these firms are under pressure to find the right people quickly. Guess what information resource they are using - the Web! [more ▶](#)

**Searchable Public Records**

**Search Systems** offers a directory of 982 searchable databases containing land records, court cases, licensing registrations, and much more for most of the U.S. state governments.

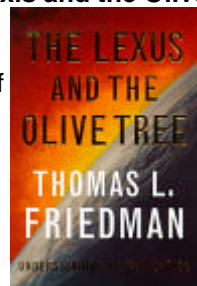


**WF Book Review**

The June issue of **Enterprise Systems Journal** contained a review by Prof. **Elliot King** at Loyola College of the book "Web Farming for the Data Warehouse." He has captured the spirit, along with the content, of the book. [more ▶](#)

**Globalization To The Max**

A recent book "**The Lexis and the Olive Tree**" by **Thomas Friedman** has created quite a stir. Regardless of your political orientation, the book contains insights into the instant global world in which we now live. Great airplane reading. And, be sure to read the **Amazon.com** reader reviews.



**WebData Databases**

**ExperTelligence** offers **WebData**, another comprehensive guide to searchable databases. Excellent Yahoo-style organization. They specialize in "finding, categorizing and



[An Introduction to Web Farming](#)

the use of custom crawling and link analysis to generate lists of topic-specific websites. [more ▶](#)

### Copernic Server

**Copernic** has repackaged their search tool into a server for businesses wishing to offer their own specialized meta-search site. See the [MetNets](#) site as an example.

organizing online databases."

### Latest Web Assessment

Ever wonder how big the Web really is? The latest assessment comes from **Steve Lawrence** and **Lee Giles** of the **NEC Research Labs**. Based on a random sampling of IP addresses, they estimated that there are 2.8 million active websites containing 800 million indexable webpages in 15 TB of text (only 6 TB if you strip out the HTML junk). More importantly, they estimated that the major search engines are covering only 16% of the indexable pages, with a strong bias towards popular, U.S. and commercial sites. Request the free report. [more ▶](#)



## Newsletter - August, 1999

### XML Extenders for IBM DB2 UDB

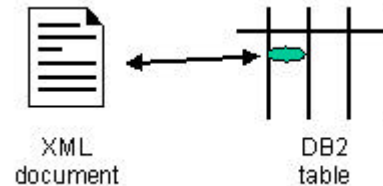
XML is one of the emerging technologies that have the potential of causing fundamental changes in whole industries. XML is becoming a standard and reliable way to exchanging structured information for B2B (business-to-business) e-commerce and for advanced information retrieval.

The conversion of existing databases to XML servers is at the heart of these changes. The introduction of the XML Extender for DB2 UDB will provide a quick and easy way to accomplish this conversion.

As another member of the DB2 Extender family, the XML Extender embeds XML functions directly into the DB2 engine. The XML Extender enables a wide range of new applications through the following functions:

1. Store an entire XML document within a column value
2. Extract and update within a XML document
3. Query within a XML document
4. Create XML documents from one or more tables
5. Update one or more tables from a XML document

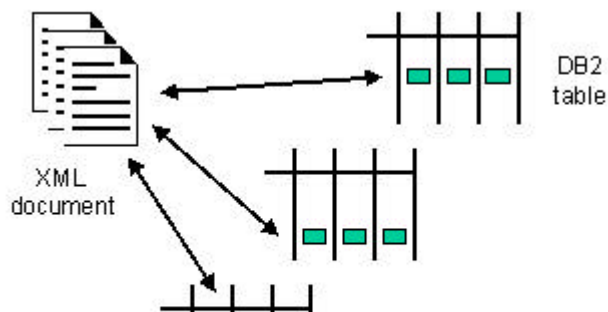
First, a XML document can be stored within a DB2 column value. A set of new UDTs (user defined types) for XML columns allows an entire XML document to be stored as a varchar, CLOB, or even a double-byte CLOB. The XML document can also use a datalink type, meaning that the XML document can reside on the local or remote file system and still be managed by DB2. Further, an embedded XML parser will validate the XML document during its initial load and subsequent updates.




Second, values within a XML document can be extracted or updated (in place) from SQL. A set of UDF (user defined functions) retrieves any portion of the XML document within an SQL SELECT statement. Likewise, an UPDATE statement against one or more Xcolumn values can update any element or attribute value within the XML document.

Third, values within a XML document can be efficiently used within a query. By declaring specific XML elements as indices, special side tables are created and automatically maintained to enable these indices. If thousands of XML documents were stored within a DB2 table, quick retrieval of specific documents would be insured. Further, the XML Extender is compatible with the Text Extender to support complex text searching.

Fourth, XML documents can be created from one or more DB2 tables. A mapping is specified from the XML structure to various DB2 tables, which can be either newly created or existing from legacy applications.



Finally, a XML document can update one or more DB2 tables, which is the reverse of the previous function.



With the support of these five functions, the XML Extender for DB2 UDB will be a major advance for implementing XML servers as extensions of legacy applications.

- *Richard Hackathorn*  
[dick@webfarming.com](mailto:dick@webfarming.com)



**Newsletter - August, 1999**

**Web Farming for Resumes and Job Openings**

## Challenge

The management recruiting industry is being reshaped by Web technology. As web-based resume databases have facilitated candidate discovery, employers expect job searches to be completed more rapidly and less expensively. With more companies taking the 'do it yourself' route by posting open jobs on careers sites, recruiters are challenged to demonstrate the value of their services by locating skilled and gainfully employed candidates who are not actively trying to change jobs. In recruiting lingo, these individuals are termed 'under the radar'.

This summer, WebFarming.com was retained by a Colorado management recruiting firm to evaluate how to systematize the detective work of recruiting. While the Web is also being leveraged within the recruiting industry to advertise positions, accept employment applications, and to market recruiting services, the primary focus of our effort was to improve the online discovery process. The objectives were to identify new sources of qualified candidates and to integrate information about candidates in the firm's contact management system.

## Approach

Because the ultimate goal was to match possible candidates with open jobs, we began by analyzing positions to be filled. A key question to be answered was, "How variable are the search assignments?" If there was a high degree of similarity between jobs, it would be considerably easier to automate aspects of the search process than if each job were unique. Since our client specialized in filling a specific type of high-tech sales position, there was substantial commonality among the assignments. For each position, we identified target companies, skills key words, and job location.

A central element of our strategy was to probe target companies for possible candidates. Using different search engines, we identified Internet domains associated with target companies and attempted to access pages with information about employees and user group members. We also identified sites linked to target company sites, implemented 'power searches' using Boolean logic to scan for candidates, and searched electronic communities like GeoCities for candidate homepages. Consistent with our goal of locating passive candidates (i.e., not actively looking for a new job), we did not scan resume databases like Monster.com. Operating on the assumption that different search engines produce different results, each step of the research process was implemented with more than one search engine.

## Experience

The initial research was extremely time consuming. On average, each search required about six hours research time. We were sure that the process could be faster. A key factor affecting research time required was the consistency of information checked. While all the documents in the contact management system and resume databases were resumes, a search can match product announcements and company job postings, as well as candidate profiles. While we tried to structure our search terms to exclude irrelevant data by using 'AND NOT', we continued to retrieve irrelevant items.


Variable quality of information was also problematic. Duplicate listings were a persistent issue. It was considerably easier to locate some types of candidates than others. For example, information about the employees of large companies was more widely available than information about the employees of small companies. And, technical candidates were more prevalent than sales candidates.

On a positive note, there was very little overlap between the candidates identified through the new Internet research process and the candidates sourced via resume databases. Our methods were producing new blood for the database!

## Hit Rate

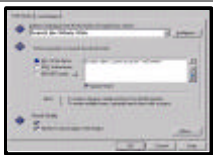

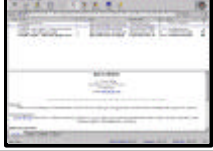
After performing our pilot searches, we evaluated our 'hit rate' or percentage of relevant to total matches. Analysis was limited to searches that worked effectively and might be considered for repetition. As a result of omitting non-productive searches, we calculated the best case hit rate. The hit rate for the best 29 searches was 3%. From a systematization perspective, the low hit rate meant that considerable manual intervention was required to sort out the high quality matches. Improving the hit rate was the next big challenge!

## Search Tools

To determine if it was possible to improve search precision, we tested a commercial tool, BullsEye from [IntelliSeek](#). BullsEye is a meta-search engine that is designed to access and combine results from many different search engines. Using BullsEye, it was possible to achieve hit rates ranging from 15% to 41%, a substantial improvement over the hit rates achieved with free search engines. Several features contributed to the improved accuracy. 

BullsEye does not rely on different search engines that operate differently to execute Boolean logic correctly and consistently. Instead, it performs searches in two successive steps. In the first step, BullsEye downloads all matches to a more inclusive query from web sites to local storage. In the second step, BullsEye allows the user to make the match criteria more specific by adding terms to exclude or synonyms to match. In addition to checking that matched sites are active, BullsEye verifies that the text of matched documents actually meets the more restrictive search criteria.

The following figures illustrate how you would use BullsEye to set the query parameters separately for the *web query* (external search) in contrast to the *local query* (internal search).

	Shows the parameters for a 'web query' that will be submitted to 20 or more search engines. Also note that the results will be checked for bad links and will a 'second-stage' assessment with the local query.
	Shows the parameters for the search within the locally compiled results (from above). Note that the local query can consist of a complex Boolean expression, which the various search engines may not be able to process correctly.
	When the two-stage query was processed, 26 search engines returned 349 items, of which 346 had valid links. Of those, only 5 passed the constraints of the local query.

An important feature of BullsEye is the ability to schedule effective searches to rerun automatically. BullsEye monitors matches already reported and presents new matches only. Taking advantage of this capability, we set up regular queries for each of the target companies.

## Next Steps

We are continuing to focus on how we can continue to improve productivity of our searches. Because manual review of matches will almost always be required to verify resume quality, we do not think it will be practical to completely automate the process. At the same time, we believe that we can make the discovery and acquisition of candidates even more efficient. We believe that streamlining how data is transferred from the Web to internal systems represents a big opportunity. In a time consuming procedure, each document is now copied, pasted, and indexed individually. We are investigating different tools to identify specific data elements within differently formatted documents and methods for importing the documents as a group.

We are also evaluating how to apply more indirect Web research methods. Other sources of names, like searches of discussion groups, alumnae lists and ISP subscriber homepages, tend to produce more variable lists. We are currently testing use of lists of this type in email campaigns to brand and market our recruiting firm's services.

Of course, the true gauge of the project's effectiveness will ultimately be its return on investment. With a new employee placement generating \$15,000 revenue on average, only one incremental employee placement per month will be required to justify the new research investment. We are monitoring number of resumes generated, number of interviews scheduled, and number of placements closed as an outcome of the new process. While no placements have occurred during the first eight weeks, our Internet candidates are being actively presented to employers. And, the fact that our client's database has expanded with many more qualified candidates has improved his ability to sell the value of his service!

- Phyllis Rheiner  
[phyllis@webfarming.com](mailto:phyllis@webfarming.com)



## Workshop Series - September, 1999

### Global Discovery Services: Use and Misuse

Sponsored By:

**USWEST** ADVANCED TECHNOLOGIES

**Description:** Survey of the popular 10-15 global search engines with an in-depth examination of Yahoo!, AltaVista, HotBot, Northern Light, and Google. Explanation of how these services discover content, what they index, and how frequently. Use of both the simple and advanced interfaces (field searches, non-HTML objects). Useful start or launch pages. Emphasis on construction of Boolean expressions and their differing interpretations. A brief look at meta-discovery services (DogPile, Mamma), specialized discovery services (xCentral, USGovtSearch), and personal tools (BullsEye, Copernic). [more ▶](#)

**Why Attend?** Do you need to locate information about customers, suppliers, competitors, technology and the like? Are you trying to use the Web for this research? Do your searches generate too much junk and too little information? Are you just plain frustrated with HotBot, AltaVista, etc? Then, this workshop is for you! It is a MUST for the web-enabled professional.

**Instructor:** Dr. Richard Hackathorn, president of WebFarming.Com. [Bio](#).

**Registration:** Simple! Submit the registration form [below](#).

**Time:** Thursday, September 9, from 8:30 am to 4:00 pm. Registration starts at 8:00 with a continental breakfast. So, come early and get acquainted.

**Place:** US West Advanced Technologies facilities in Boulder. [Directions & Map](#)

**Fee:** \$195 which covers materials, breakfast, and breaks. Lunch is on-your-own at the US West cafeteria. Checks only are accepted and should be payable to Bolder Technology, Inc. Payment must be received prior to or at the workshop registration.

**Future Workshops:** On roughly a monthly basis, we plan to offer workshops on the following topics. Suggestions for your favorite topic are welcome via email to [WF Workshops](#).

- **Dealing with Hidden Web Content.** Hidden content is web content that is not indexed by the global discovery services, which is probably larger than the visible Web. Focus on web databases for news, discussion forums, patents, SEC financials, trademarks, etc. Examination of Beaucoup directory and Lycos Searchable Reference Databases (from IntelliSeek's Invisible Web Catalog).
- **Survey of Personal Discovery Tools.** Survey of browser enhancements (Alexa, InfoSeek Express) and separate personal tools (BullsEye, Copernic, WebCompass). Discussion of the limitations of using these tools in a group setting.
- **Acquiring Web Content.** Now that you found some good stuff, what do you do with it? Capturing links or content or both. Dealing with non-English content. Conversion of an HTML table to Excel spreadsheet. How to maintain content currency. Using MS-Office Web Queries. Developing manual web scrapers in Visual Basic and Perl. Dexter from Liaison Technology.
- **Working with Commercial Content Providers.** Survey of 15-20 popular vendors, with a

more detailed examination of the Electric Library, Hoover's, and Dow Jones Interactive.

- **Experiments with Hubs and Authorities.** Based on Klienberg's work (as utilized in the IBM Clever Project), we will manually construct virtual communities for several topics. See the June 1999 issue of Scientific American for a good description of this technique. Discussion of the practical application of this technique and a comparison with new Google website.
- **Legal and Ethical Issues with Web Espionage.** Any information accessible on the Web by any means is fair game, usable by anyone in any way for any purpose. Is this statement true? If your corporation is farming the Web, then you need to understand the answer to this question.
- **HyperText Analysis Techniques and Tools.**
- **XML For Fun and Profit.**
- **Information Visualization Techniques and Tools.**
- **Product Reviews.** The workshop will feature one or two key products for web farming systems.

Please register me for the workshop. . .

<b>First:*</b> (* Indicates required field)	<b>Last:*</b>
<input type="text"/>	<input type="text"/>
<b>Title:</b>	<b>Company:</b>
<input type="text"/>	<input type="text"/>
<b>Address:*</b>	<b>Address2:</b>
<input type="text"/>	<input type="text"/>
<b>Floor/Suite:</b>	<b>City:*</b>
<input type="text"/>	<input type="text"/>
<b>State or Province:*</b>	
<input type="text" value="Make a selection"/>	
<b>Postal Code:</b>	<b>Country:*</b>
<input type="text"/>	<input type="text" value="Make a selection"/>
<b>Telephone Number:</b>	<b>Fax Number:</b>
<input type="text"/>	<input type="text"/>
<b>Email:*</b>	
<input type="text"/>	
<b>Comments or Requests:</b>	
<input type="text"/>	

Submit Registration

Clear Form


[Attend an Event](#)
[Exhibit at an Event](#)
[Other DCI Services](#)

## ◀ [Attend an Event](#)

[Register](#) for this seminar

[Who](#) should attend

Seminar [outline](#)

About your [instructor](#)

[Meeting site](#) and hotel information

## ◀ [IT News & Features](#)

◀ [Register Now](#)

◀ [Customer Service](#)

◀ [Corporate Info](#)

◀ [Search Our Site](#)

## Building Web Farming Systems: Tools and Techniques

Enhanced Data Warehousing for your Enterprise

San Francisco - September 15-17, 1999

Dallas - November 10-12, 1999

### About This Seminar

The goal of this seminar is "to organize outside data" as Drucker recommends. It is to move your company from an information refining process that is haphazard and intermittent to one that is systematic and continuous. By blending with your existing warehousing systems, web farming can enable your company to adapt and even thrive upon the sudden changes happening in your industry.

This unique seminar provide you with the practical skills for evaluating and implementing web farming systems for your company. Covering both the technical and management aspects, this seminar captures the excitement and substance of this emerging area. Web farming is a practical step toward knowledge management-a way of pushing the curve by leveraging your investment in data warehousing. Following an exclusive four-stage methodology, this seminar will minimize the risk of an unsuccessful implementation while maximizing the resulting benefits. Web farming is a 'must-have' component of enterprise systems! Don't learn about the business potential of web farming from your competitors. This is one IT wave that you must lead!

### What's Happening

Doing business as usual is now a risky strategy. Current wisdom within your company is no longer valid. The market in which you were successful disappeared last night. This is the bad news. The good news is that the same happened to your competitors.

The focus of data warehousing and other business intelligence systems has been on data from operational systems and other internal sources. As management guru Peter Drucker argues, the challenge is "to organize outside data because change occurs from the outside."

Amid the chaos of the Web is a massive diversity of

ever-changing information, some of which is critical to your future success. The Web is the mother of all data warehouses! However, can your company realize this enormous value?.

### **What Makes This Seminar Unique**

Based on a recently published book, this is the only seminar that fully covers this emerging area of web farming, balancing a solid business perspective with the key technology issues. In particular, this seminar is uniquely at the intersection of data warehousing, information analysis, library science, and web technology. It is also unique in its focus on the input side of data warehousing. Most web technology is being applied to the delivery of information, rather than its discovery and acquisition.

[\[top\]](#)

### **Who Should Attend**

- Project Managers who are responsible for justifying and managing business intelligence and data warehousing systems.
- Project Leaders who need to know proven steps for selecting tools and services necessary for implementing web farming systems.
- Data Warehouse Designers who are looking for ways to enhance their existing systems for faster responsiveness to changing business conditions.
- Business Analysts who translate business requirements into requests for specific information.
- System Architects who must understand the proper architectures and infrastructures required for advanced business intelligence.
- Technology Planners who need a survey of the set of emerging Web technologies to establish priorities for their company.
- Web-literate IT professionals who wish to expand their knowledge and skills in this fast-moving area

### **Benefits of Attending**

- EXAMINE the tools available to maximize your data warehouse with Web Farming
- TAKE HOME in-depth product information of OVER 50 tools and services and save yourself and your company months of research
- LEARN how to apply a practical methodology for implementing business intelligence systems
- LEARN to build a production infrastructure for supporting this system

## Benefits to Your Company

Your company will benefit from:

- Rapid deployment of critical intelligence function, with minimal risk
- Strict focus on business objectives, with fast return on investment
- An ability to respond to sudden shifts in your market or technology

[\[top\]](#)

## Seminar Outline

1. What is Web Farming?
  - a. A Definition
  - b. Examples of Web Farming
  - c. Why Farming and not Mining?
  - d. Covering Your Info-Rear
  - e. Pivotal Role of Data Warehousing
  - f. Toward Knowledge Management
  - g. Information Refining
  - h. Four-Stage Methodology
2. Why Web Farming?
  - a. The Problem
  - b. Where is the pain?
  - c. Business Justification
  - d. The Business Drivers
3. Web Technology 101
  - a. Web Stuff That You Should Know
  - b. IP Addressing, Domain Naming
  - c. URL, URI, and URN
  - d. MIME Data Types
  - e. HTML and Variations
  - f. XML and its Promises
  - g. Web-Enabled Databases
  - h. Web Searching and Indexing
4. Stage 1 - Getting Started
  - a. Steps in Getting Started
  - b. Architecture for Getting Started
  - c. The Business Drivers in Detail
    - Customer Relationships
    - Supply Chain Relationships
    - Demand Chain Relationships
    - Competitive Awareness
    - Technology Trends
    - Market Micro-Dynamics

- Deregulation and Privatization
- Global Politics and Economics
- d. Compiling the Business Case
- e. Selling to a Skeptical Management
- f. Legitimizing the WF Function
- g. Secrets of Information Detectives
- h. The Terrain of WebScape
- i. The Hidden Web
- j. The Detective Process
- k. Search Query Formulation
- l. Survey: General Discovery Services
  - About.com (MiningCo)
  - Alta Vista
  - Excite
  - Google
  - Hotbot
  - Infoseek
  - Magellan
  - Northern Light
  - Web Crawler
  - Yahoo
  - plus Lycos, Snap, Fast, MSN
- m. Joys of Web Surfing
- n. Survey: Meta-Discovery Services
  - BigHub (formerly iSleuth)
  - DogPile (plus MetaFind)
  - Go2Net (Metacrawler)
  - InferenceFind
  - Mamma.com
  - Savvysearch
  - Daily Diffs from InGenius
  - CompanySleuth, Infonautics
  - Xcentral, jCentral from IBM
  - TwURL Analysis from ROI
- o. Refining Your Search
- p. Recall versus Precision
- q. Survey: Discovery Tools
  - Alexa from Amazon/Alexa
  - BullsEye from IntelliSeek
  - Copernic from Agents Tech
  - Enfish Tracker Pro, Enfish
  - Liveagent Pro from Agentsoft
  - Odyssey from General Magic
  - Webcompass, Quarterdeck
  - Webforia Organizer
- r. Survey: Content Providers
  - Dialog Information Services
  - Dow Jones Business Info
  - Electric Library

- Galenet
  - Hoover's, Inc.
  - . . . and many more
5. Stage 2 - Getting Serious
- a. Steps in Getting Serious
  - b. Architecture for Getting Serious
  - c. Selling to a Skeptical Management
  - d. Legitimizing the WF Activity
  - e. Role of Information Analyst
  - f. Competitive Intelligence
  - g. Survey: Acquisition Tools
    - Dexter from Liaison Tech
    - WebL from Compaq/Digital
    - WIDL from webMethods
    - W4F from Univ of Penn
  - h. Web Crawling and Indexing
  - i. Survey: Web Search Engines
    - Compass Server, Netscape
    - Excite for Web Servers
    - Index Server from Microsoft
    - Search'97 from Verity
    - Search Intranet from AltaVista
    - Searchserver from Fulcrum
    - SmartCrawl from Inktomi
    - Ultraseek Server, Infoseek
6. Stage 3 - Getting Smart
- a. Steps in Getting Smart
  - b. Architecture for Getting Smart
  - c. Hypertext Analysis Techniques
  - d. Survey: Hypertext Analysis Tools
    - Integrity from Vality Inc.
    - Intelligent Miner/Text, IBM
    - Linguistx from Inxight
    - Netowl Intelligence, Isoquest
    - Retrievalware from Excalibur
    - Webinator from Thunderstone
  - e. Information Structuring
  - f. Survey: Info. Visualization Tools
    - SemioMap from Semio Corp
    - SmartContent from Perspecta
    - ThemeScape from Cartia
    - UMAP from TriVium
    - VizControls from Inxight
  - g. Dissemination and Collaboration
  - h. Survey: Ent. Information Portals
    - Deliverymanager from VIT
    - Information Advantage
    - Plumtree Software
    - Powerize Server

- SageWave from SageMaker
    - ServicePort from Portera
    - SQRIBE (acquired by Brio)
    - Viador Inc.
    - X-Ray from InfoRay
  - i. XML In-Depth
  - j. Meta Data Chaos on the Web
  - k. Up the Meta-Foodchain
    - l. XML Documents and Namespaces
  - m. Survey: XML Authoring Tools
    - Visual-XML from Bluestone
    - XML Authority, Extensibility
    - XML Notepad from Microsoft
  - n. Resource Definition Framework
  - o. XML/RDF Vocabularies
    - Dublin Core
    - BizTalk
  - p. Survey: XML Servers
- 7. Stage 4 - Getting Tough
  - a. Steps in Getting Tough
  - b. Architecture for Getting Tough
  - c. Rendezvous with the Warehouse
  - d. Star Schema Design
  - e. The ERP Connection
  - f. Data Quality and Reliability
  - g. Creating Info-Markets
  - h. Survey: Knowledge Mgt. Systems
    - Agentware I3 from Autonomy
    - Dataware II from Dataware
    - Folio Suite, Open Markets
    - Infomagnet, Compassware
    - Knowledge Server, Intraspect
    - Knowledgex from IBM
    - Livelink Intranet, Open Text
    - Wisdombuilder
  - i. Deeper Issues
  - j. Privacy and Confidentiality
  - k. Intellectual Property Rights
    - l. Industrial Espionage
  - m. Information Warfare
  - n. Suggested Code of Ethics

[\[top\]](#)

### About Your Instructor

Dr. Richard Hackathorn is a well-known technology innovator and international educator with over 30 years of experience in the IT industry. As a university professor at Wharton School and at

the University of Colorado, he pioneered the early foundations for enterprise systems, database management, decision support, and data warehousing. He founded MicroDecisionware, an early middleware vendor which was acquired by Sybase. He also founded Bolder Technology, Inc. to provide premier consulting and education services related to data warehousing systems. Recently, he founded WebFarming.com (<http://webfarming.com/>) to provide system integration services for corporate clients. He has published three books, written numerous articles, organized major conferences, and conducted professional seminars in sixteen countries.

### **What Attendees Have Said**

Previous attendees have described the seminars conducted by Dr. Hackathorn as:

- "I gathered a lot of information on tools and techniques. I was pleased!"
- " I can follow a progression once I leave here. It gives me a solid roadmap"
- " You can walk away with a complete picture... An excellent course!"
- "Richard's casual style was excellent!"

### **Bonus Material**

Receive Your FREE Copy of Web Farming for the Data Warehouse, a professional textbook by Richard Hackathorn  
Other Bonus Materials:

- Detailed Seminar Notes
- List of Key Information Resources on the Web
- Recommended Readings
- An up-to-date glossary

### **Meeting Site & Hotel Information**

San Francisco - September 15-17, 1999  
SF Fisherman's Wharf (7479)

Dallas - November 10-12, 1999  
Wyndham Gardens (7481)

To receive discounted reservations or further information:  
Call CATS: 1-800-767-2755 or 978-470-3933  
Fax CATS: 978-470-2506  
Email CATS: [cats@dci.com](mailto:cats@dci.com)  
Web: [www.CATStravel.com](http://www.CATStravel.com)

**Want to bring this seminar to your site?**  
**Click here to find out how.**

[\[top\]](#)

**Register Now!**

**Building Web Farming Systems: Tools and Techniques**

**San Francisco: \$1195**

**Dallas: \$1195**

**Building Web Farming Systems: Tools and Techniques**  
**San Francisco - September 15-17, 1999**  
**SF Fisherman's Wharf**  
**\$1195**

[Data Mining: Evaluating Methods and Tools](#)  
**San Francisco - September 13-14, 1999**  
**SF Fisherman's Wharf**  
**\$995**

**Sign up for both seminars for \$1795 -- You Save \$395**

---

**Building Web Farming Systems: Tools and Techniques**  
**Dallas - November 10-12, 1999**  
**Wyndham Gardens**  
**\$1195**

[Data Mining: Evaluating Methods and Tools](#)  
**Dallas - November 8-9, 1999**  
**Wyndham Gardens**  
**\$995**

**Sign up for both seminars for \$1795 -- You Save \$395**

---

**Send me more info on this event in San Francisco**

**Send me more info on this event in Dallas**

[Attend an Event](#) | [Exhibit at an Event](#) | [DCI Services](#) | [IT News & Features](#) | [Register Now](#) | [Customer Service](#)  
[Corporate Info](#) | [Search Our Site](#) | [Home](#)

Affiliated Sites: [Conference and Travel Services \(CATS\)](#) | [IT Media Group](#)

© Copyright 1999 by DCI  
204 Andover Street, Andover, MA 01810, (978) 470-3880  
All event names are trademarks of DCI or its clients.  
DCI is a Massachusetts Business Trust  
Comments? [webmaster@dci.com](mailto:webmaster@dci.com)