



It's All in the DATA

ALAR 2003 Annual Report

Table of Contents

Steering Committee	1
Staff	2
Research Investigators	3
Introduction	4
Laboratory Expands University Partnerships	6
ALAR Designated Arkansas Center for Applied Technology in Information Technology	7
ALAR Develops Process for Proposal Submissions	7
Research Projects Started in 2003	9
An Architecture for Capturing, Storing, and Connecting Event Information to Facilitate Customer Data Integration (Phase II)	9
Research Projects Completed in 2003	10
Mining Hidden Information From Large Datasets	10
An Architecture for Capturing, Storing, and Connecting Event Information to Facilitate Customer Data Integration (Phase I)	10
Research Projects Completed in 2002	11
Script, Semi-script, and Feature based HTML Document Filtering (WebGrabber Phase I)	11
Script, Semi-script, and Feature based HTML Document Filtering (WebGrabber Phase II)	11

(Table of Contents, continued)

Structured and Semi-structured Data Extraction from HTML Documents (Browser Harness)_____	12
Other ALAR Activities _____	12
Proposal submitted to the National Science Foundation_____	12
Conference on Applied Research in Information Technology_____	15
Expanded Website_____	15
Publications _____	16
Year Four Objectives _____	17

Preface

By Dr. John R. Talburt, Director, Acxiom Laboratory for Applied Research

There is no doubt that 2003 has seen the most exciting developments for the Laboratory since its inception. Three milestones for the Laboratory that you will read about in more detail in this report are:

- o An expanded membership and research scope;
- o A designation as an Arkansas Center for Applied Technology; and
- o A more inclusive and well-defined proposal solicitation process.

These have all clearly been important developments for the Laboratory and its stakeholders. However, having been associated with research at Acxiom since 1993, I believe that the new proposal solicitation process will have, and to some extent already has had, the most impact.

It is the first time in my tenure at Acxiom that the company has, in a systematic way, asked its technology leaders to identify and define a set of strategic customer data management problems as the basis for soliciting proposals from its academic research partners. Even though the primary motivation for the new process was to assure that research dollars are invested in the most strategic projects, it has also produced other positive side effects. Perhaps the most important is that it is beginning to engender a much stronger sense of "research community" than has been experienced before--not just between Acxiom associates and research faculty members but also among the

Acxiom technical leaders who participated. The process has already revealed that several Acxiom teams share the same problems, and they can benefit by a common solution.

Let me also take this opportunity to welcome the new participants from the University of Arkansas, Fayetteville. I hope that we will see even more research partners added in 2004 and the years to come.

Finally, on behalf of the Steering Committee and Laboratory Staff, I want to thank all of the Laboratory stakeholders whose support has made the accomplishment of these milestones possible. Special thanks go to Jerry Adams and Alex Dietz of Acxiom Corporation and to Dr. John Ahlen and Dr. Herb Monoson of the Arkansas Science and Technology Authority for their enthusiastic support of the Laboratory's mission and activities.

Steering Committee

Walter Anderson
Acxiom European Technology
501-342-2496 wander@acxiom.com

Yupo Chan, Ph.D.
UALR Department of Systems Engineering
501-569-8926 yxchan@ualr.edu

Fred Davis, Ph.D.
UAF Department of Computer Information Systems and Quantitative Analysis
479-575-4500 fdavis@walton.uark.edu

Chad Fitz
Acxiom Solutions Development
501-252-4031 cfitz@acxiom.com

Charles Wesley Ford, Jr., Ph.D.
UALR Department of Computer Science
501-569-8130 cwford@ualr.edu

Mary L. Good, Ph.D.
UALR Donaghey College of Information Science and Systems Engineering
(CyberCollege)
501-569-8188 mlgood@ualr.edu

Donald Hinman, Ph.D.
Acxiom InfoBase Global Products
501-252-2752 dhinma@acxiom.com

Charles Howland
Acxiom Data and Grid Technology
501-252-1123 chowla@acxiom.com

Robert Ludwig
Acxiom InfoBase and Customer Information Infrastructure Delivery Center
501-252-0316 rludwi@acxiom.com

David Roland
Acxiom Fayetteville Development Unit
501-342-1549 drolan@acxiom.com

Kazem Sohraby, Ph.D.
UAF Department of Computer Science and Computer Engineering
479-575-6036 sohraby@uark.edu

Steering Committee (continued)

John Talburt, Ph.D.
Acxiom Global Data Development
501-252-2436 jtalbu@acxiom.com

Terry Talley, Ph.D.
Acxiom InfoBase and Customer Information Infrastructure Technology
501-342-2887 ttalle@acxiom.com

Craig Thompson, Ph.D.
UAF Department of Computer Science and Computer Engineering
479-575-6519 cwt@enr.uark.edu

Staff

Director
John Talburt, Ph.D.
Acxiom Global Data Development
252-2436 jtalbu@acxiom.com

Associate Director
Bob Crisp, Ph.D.
UAF Department of Computer Science and Computer Engineering
479- 575-6416 crisp@enr.uark.edu

Associate Director
Joe Swaty, APR
UALR Donaghey College of Information Science and Systems Engineering
(CyberCollege)
501-537-9864 jwsваты@ualr.edu

Program Manager
Judy Camp, M.A.
UALR Donaghey College of Information Science and Systems Engineering
(CyberCollege)
501-683-7050 jacamp@ualr.edu

Research Investigators

Principal Investigators:

Coskun Bayrak, Ph.D.
UALR Dept. of Computer Science
501-569-8137 cxbayrak@ualr.edu

Russel Bruhn, Ph.D.
Chair, UALR Dept. of Information Science
501-683-7056 rebruhn@ualr.edu

Yupo Chan, Ph.D.
UALR Dept. of Systems Engineering
501-569-8926 yxchan@ualr.edu

Chia-Chu Chiang, Ph.D.
UALR Dept. of Computer Science
501-569-8142 cxchiang@ualr.edu

Yijun Ding, Ph.D.
Acxiom Global Data Development
501- 252-1249 yijun.ding@acxiom.com

Charles Wesley Ford, Jr., Ph.D.
UALR Dept. of Computer Science
501-569-8130 cwford@ualr.edu

Ray Hashemi, Ph.D.
UALR Dept. of Computer Science
501-569-8135 rmhashemi@ualr.edu

Rasesh Patel
InfoBase Technology and Architecture
501-342-1013 rasesh.patel@acxiom.com

Steve Sieloff
Acxiom Global Data Development
407-321-7103 steve.sieloff@acxiom.com

Tom Schweiger, Ph.D.
Acxiom Fayetteville Development Unit
501-342-6294 tom.schweiger@acxiom.com

John Talburt, Ph.D.
Acxiom Global Data Development
501-252-4236 john.talburt@acxiom.com

Ningning Wu, Ph.D.
UALR Dept. of Information Science
501-683-7051 nxwu@ualr.edu

Student Research Assistants:

Ashish Bansal
UALR Graduate Student
Computer Science

Radhika Chilka
UALR Graduate Student
Computer Science

Linda Ning
UALR Graduate Student
Information Science

Nan Tang
UALR Graduate Student
Computer Science

Hao Wu
UALR Graduate Student
Computer Science

Yanbin Ye
UALR Graduate Student
Applied Science

Introduction

The Acxiom Laboratory for Applied Research (or ALAR) is collaboration between Acxiom Corporation and its university research partners, including the University Arkansas at Little Rock (UALR) and the University of Arkansas, Fayetteville (UAF). Its primary purpose is to facilitate Acxiom's engagement in collaborative, applied research with universities to solve key problems related to customer data management. ALAR, formerly known as the Acxiom Data Engineering Laboratory (ADEL), began with help from the Arkansas Science and Technology Authority (ASTA) in March 2001 as a research partnership between Acxiom Corporation and the UALR CyberCollege, which hosts the ALAR administrative office.

Now in its third year of operation, the Laboratory has an established record that includes a number of successfully completed research projects, the development of a comprehensive model to guide Acxiom's applied research engagements, and Joint Research Agreements in place with two of its principal university partners to cover intellectual property and product transfer issues. The Laboratory has proved to be an efficient means to marshal the research resources that help Acxiom maintain its global leadership in customer data management.

In the Laboratory's third-year of operation, 2003, it expanded its scope, both in terms of its membership and areas of research. In adopting a new governance agreement in September 2003, the Laboratory created a framework for adding new members and affiliates. The University of Arkansas, Fayetteville,

became the first new member to be added under the new agreement, joining UALR, the Laboratory's founding member. In addition, the new agreement changed the Laboratory's name from the "Acxiom Data Engineering Laboratory" to the "Acxiom Laboratory for Applied Research" to reflect its expanded scope of sponsoring research in all areas related to customer data management.

In 2003, the Laboratory applied for recognition as an Arkansas Center for Applied Technology in Information Technology by the Arkansas Science and Technology Authority (granted in January 2004). In addition, it also developed and began executing a new process for soliciting proposals to assure that they address Acxiom's strategic research needs. The key element of the new process is the increased involvement of Acxiom technical associates at each step from the problem development through the final funding decision.

In its first and second years, the Laboratory sponsored five, successful applied research projects with ASTA's support. In the third year, 2003, the Laboratory sponsored an additional project, which will be completed in August 2004, and received nine "pre-proposals" that were refined into three, full proposals under consideration for funding in 2004.

The Laboratory also organized its annual conference for the dissemination and sharing of research between regional university and business professionals—the third annual "ALAR Conference on Applied Research in Information Technology" will be held February 27, 2004, on the UALR campus. An expanded ALAR website (<http://alar.ualr.edu>) includes on-line registration for the

conference as well as updated information about the Laboratory's projects and members.

In 2004, in addition to organizing its next regional conference, the Laboratory expects to sponsor several new projects and continue to seek federal grant support. It will also continue to identify member and affiliate schools for inclusion in ALAR and invite Problems/Solutions from Acxiom associates and university partners for future proposal funding.

Laboratory Expands University Partnerships

In September 2003, the Acxiom Laboratory's Steering Committee revised its Governance Agreement to include both the University of Arkansas at Little Rock and the University of Arkansas, Fayetteville, as member schools. It also made provisions for adding other member and affiliate schools and partners that have an interest in collaborative research with Acxiom. As previously mentioned, the Steering Committee also changed the Laboratory's name to the Acxiom Laboratory for Applied Research (ALAR) from its previous title, "Acxiom Data Engineering Laboratory (ADEL)," in order increase both its scope and outreach. Initially focused on data engineering problems in partnership with the UALR Donaghey CyberCollege, the Laboratory now pursues projects in all areas of research related to customer data management in concert with member schools.

ALAR Designated Arkansas Center for Applied Technology in Information Technology

In November 2003, the Laboratory, through the sponsorship of UALR, submitted a proposal to the Arkansas Science and Technology Authority to be designated as an Arkansas Center for Applied Technology in Information Technology (ASTA). The proposal was approved at the January 2004 meeting of the ASTA Board of Directors, and ALAR is the only organization in Arkansas to be so designated at this time. The new five-year Center designation, beginning January 19, 2004, will allow the Laboratory to streamline its funding of proposals while allowing Acxiom Corporation to receive tax credits for research and development. The Laboratory will be able to sponsor identified projects at UALR, UAF, and other ALAR affiliated universities under the new Center.

ALAR Develops Process for Proposal Submissions

In September 2003, the Laboratory began developing a process for soliciting research proposals from member schools to assure that Acxiom's most critical research needs were being met. The new process first calls for "Problem Statements" from Acxiom business and technical leaders. The statements are then distributed to the member schools, which invite the research faculty to respond with "pre-proposals." An Acxiom technical committee then reviews the pre-proposals and makes its recommendations to the Laboratory Steering

Committee. The Steering Committee then decides which pre-proposals should be developed into full proposals.

In the second phase of this process, ALAR organizes a "Proposal Workshop" for university researchers and Acxiom associates to exchange information and refine the endorsed pre-proposals. After receiving subsequent full proposals from researchers, the Acxiom review committee then recommends a final set of projects for funding action by the ALAR Steering Committee. Winning proposals of the first round of the new process will be announced and project plans presented at the ALAR Conference on Applied Research in Information Technology on February 27, 2004.

This year's process included five Acxiom problem statements to which ALAR received nine solution pre-proposals. The final result of the Acxiom reviews and Proposal Workshop was that three full proposals should go forward for funding consideration in 2004. These are:

1. *Grid Node Failover and Partitioning* - Dr. Dale R. Thompson (UAF), Dr. Amy Apon (UAF), and Dr. Craig Thompson (UAF)
2. *A Framework to Automate the Generation of BCDI Process Flows* - Dr. Craig Thompson (UAF), Dr. Wing Li (UAF), and Dr. Coskun Bayrak (UALR)
3. *Unified Adaptive Model of the Acxiom Grid Architecture* - Dr. Craig Thompson (UAF), Dr. Amy Apon (UAF), Dr. Russell Deaton (UAF), Dr.

Henry Hexmoor (UAF), Dr. Dale R. Thompson (UAF), Dr. Mihail Tudoreanu (UALR), and Dr. Xiaowei Xu (UALR)

Research Projects Started in 2003

1. An Architecture for Capturing, Storing, and Connecting Event Information to Facilitate Customer Data Integration (Phase II) – Dr. Charles Ford (UALR, Computer Science); Dr. Chia-Chu Chiang (UALR, Computer Science); Dr. John Talburt (Acxiom); Dr. Tom Schweiger (Acxiom); Steve Sieloff (Acxiom); Hao Wu (UALR, Graduate Student); and Radhika Chilka (UALR, Graduate Student). September 2003 – August 2004.

This project continues investigation of a previously supported ASTA/Acxiom/UALR project, “An Architecture for Capturing, Storing, and Connecting Event Information to Customer Data Integration” by Dr. Ray Hashemi. This continuation now focuses on techniques for evaluating Web data quality, which authenticates the retrieved information against a repository of known information. Its primary goals are to develop a prototype framework for implementing queries against a distributed database, business rules for authenticating records; an algorithm for calculating the confidence level of the result; and a middleware architecture with defined interfaces among principal components. The project is expected to greatly improve Acxiom’s techniques for data acquisition and management and to provide this knowledge-based company’s clients with useful data to increase their market shares.

Research Projects Completed in 2003

1. Mining Hidden Information From Large Datasets--Dr. Ningning Wu (UALR-Information Science); Dr. Russel Bruhn (UALR-Information Science); Rasesh Patel (Acxiom Corporation); Linda Ning (UALR Graduate Student) and Yanbin Ye (UALR Graduate Student). May 2002-October 2003.

This project aims at investigating and developing data mining techniques to efficiently discover new and hidden knowledge from huge datasets. Acxiom expects to use the resultant knowledge obtained from this project to improve existing business rules.

2. An Architecture for Capturing, Storing, and Connecting Event Information to Facilitate Customer Data Integration (Phase I)--Dr. Ray Hashemi (UALR - Computer Science); Dr. Charles Ford (UALR - Computer Science); Steve Sieloff (Acxiom Corporation); Ashish Bansal (UALR Graduate Student); Nan Tang (UALR Graduate Student). May 2002-August 2003.

The initial work focuses on two areas: the development of an ontology for extracting features from narrative event announcements in online newspapers and a schema for building a related geographic index.

Research Projects Completed in 2002

1. Script, Semi-script, and Feature based HTML Document Filtering

(WebGrabber Phase I)--Dr. Ray Hashemi (UALR - Computer Science); Dr. Charles Ford (UALR - Computer Science); and Dr. Yeong-Tae Song (UALR - Computer Science). March 2001-May 2002.

The WebGrabber prototype delivered by the research team became production-ready at Acxiom in December 2001. Researchers report an extremely high overall accuracy for name recovery. Several system enhancements have improved the system's performance and automated the auditing of the compiled records.

2. Script, Semi-script, and Feature based HTML Document Filtering

(WebGrabber Phase II) --Dr. Ray Hashemi (UALR - Computer Science); Dr. John Talburt (Acxiom Global Data Development Business Unit) September 2001-May 2002.

This project is an extension of the original project designed to increase efficiency of the original prototype. It generalizes the architecture of the original prototype so that the same basic "web crawler engine" extracts a variety of text-based features beyond the basic name and title elements targeted by the original research.

3. Structured and Semi-structured Data Extraction from HTML Documents (Browser Harness)--Dr. Coskun Bayrak (UALR - Computer Science). April 2001-May 2002.

The original Browser Harness Version 1.0, delivered in May 2002, has had several added features, mostly related to usability. The current production version is now 1.4.0. One of the most important features added lets the operator select among three different options for the output format of the captured data, MS Access, CSV, or XML. Another helpful feature is that it can now automatically follow the "next results" links for sites that return a fixed number of items per page and display overflow items on separate pages.

Browser Harness is being used to support a new Acxiom product, the InfoBase Patriot Act Solution.

Other ALAR Activities

1. Proposal submitted to the National Science Foundation after receiving favorable white paper review.

After receiving word from the Director of the Division of Information and Intelligent Systems at the National Science Foundation that his project concept for data visualization is appropriate, Dr. Mihail Edi Tudoreanu (UALR) has submitted a NSF proposal for consideration in the Human Computer Interaction Program. The following summarizes this research, which if funded, would begin in August 2004:

Animated Visualization for Extracting the Underlying Processes that Affected Business and Social Data over Time – Dr. Mihail Edi Tudoreanu (UALR, Information Science); Dr. John Pittenger (UALR, Department of Psychology); and Don Hinman (Acxiom Corporation):

Business and government information systems have created enormous amounts of information that span decades. This data contains, hidden among trivia and random effects, a mark left by the processes that shaped social and business environments in the past. Similar processes are likely to affect current data, and analysts who understand their effect should be able to extract more precise actionable information. One straightforward application of such knowledge is in the prediction of future conditions, such as market share or urban development, based on the current data and on the understanding gathered from historical information. Other applications include judging the coverage and useful information content of a new dataset or data source.

Understanding processes from a stream of snapshot data requires sifting through large amounts of data and, in fact, adds a new dimension to the already large datasets. A brute force approach is to gain a deep knowledge of the datasets at each point in time and then to assemble all this knowledge in hope of discovering the underlying mechanisms that transformed the data. Because effects of processes span long periods and involve branching or repetition, brute force is not a viable alternative. Huge quantities of data would have to be manipulated, leading to a slow or error prone analysis. This proposal addresses the problems of a) finding techniques that are capable of presenting processes

rather than data; b) determining an appropriate manner of introducing such techniques in the work environment to benefit performance; and c) understanding the tools and information currently at the disposal of analysts as well as the analysts' performance in traditional environments.

The research will proceed by expanding results of program visualization that convey the processes taking place within a running computation, by considering the cognitive science on the role of graphics, and by obtaining access to study real-life data analysts and data sources at a world-class data management corporation. This proposal seeks to develop an information space beneficial to the extraction of underlying processes from historical data. The technology development is the background on which the proposed work pursues empirical observations and evaluation of user activities and animated visualization. Central to this approach is the use of animated views of the data transformations over time. These views will be adapted from the PI's work in program visualization and will be designed for improving cognition via external media and adherence to ecological perception concepts.

2. Conference on Applied Research in Information Technology

The UALR CyberCollege; University of Arkansas, Fayetteville; Acxiom; and ALAR will host the "Conference on Applied Research in Information Technology" on the UALR campus on February 27, 2004. Last year's event attracted approximately 150 students, faculty members, business people, and Acxiom associates from around Arkansas and the neighboring states of Alabama, Mississippi, Oklahoma, and Texas.

This year's participants were able to register for the conference on the ALAR website. In addition to ten selected research papers and three reports on Acxiom-sponsored research, they will also hear presentations from keynote speakers Bob Thome, Senior Manager, Oracle Corporation, and Dr. Bhavani Thursaisingham, Program Director, National Science Foundation. Thome will speak on "Enterprise Grid Computing," and Thursaisingham, on "Data Mining and Information Assurance."

In all, conference attendees will benefit from more than 18 hours of informative presentations on information technology research and curricula. They will also hear presentations of selected research papers from universities such as UALR, UA-Fayetteville, the University of Central Arkansas, and Arkansas State University.

3. Expanded Website

A full-featured website (<http://alar.ualr.edu>), developed in 2001, continued to expand in 2003, adding new members, researchers, and projects. The site was used for online registration for this year's ALAR Conference on Applied

Research in Information Technology. In addition to supplying highlights of the conference, it also includes a posting of "Working Papers" from participants' presentations. Additionally, the site includes information and vitae on principal investigators as well as news about research projects and other ALAR activities. The site also features direct links to websites at UALR and the Donaghey CyberCollege; UAF and the College of Engineering; the Sam Walton College of Business and Information Technology Research Center; Acxiom Corporation, and the Arkansas Science and Technology Authority.

4. Publications

The following papers and presentations were supported in part through the Acxiom Laboratory for Applied Research:

Bayrak, C., Kolukisaoglu, H., & Sieloff, S. (2003, September). Data extraction from repositories on the web: A semi-automatic approach. *Transactions of Design and Process Science*, 7(4), 13-23, Available from <http://www.sdpsnet.org/vol7-4.htm>

Bayrak, C., Kolukisaoglu, H., Chung, H., & Talburt, J. (2002, June 23-27) Information harnessing on the World Wide Web. *Proceedings of the 6th Biennial World Conference on Integrated Design and Process Technology*, 1, Pasadena, CA.

Bayrak, C. (2002, July 22). Structured and semi-structured information extraction from the World Wide Web. Presented to the faculty and students of the Electrical and Computer Engineering Department at Karadeniz Technical University, Trabzon-Turkey.

Bayrak, C. (2002, July 6). Supervised information extraction. Presented to the faculty and students of the Electrical and Computer Engineering Department at Bogazici University, Istanbul, Turkey.

Hashemi, R. , Ford, C. W., Vamprooyan, T. & Talburt, J. (2002, November). Extraction of features with unstructured representation from html documents. In P. Isaias (Ed.), *Proceedings of the IADIS International Conference on WWW/ Internet 2002* (pp. 47-53).

Lisbon, Portugal: *International Association for Development of Information Society*

Wu, N. (2003, June). Factor analysis based intrusion detection. *Proceedings of IEEE Workshop on Information Assurance*, New York, 108-115.

Wu, N. (2002, November 1). Mining association rules from very large datasets. Acxiom Data Engineering Laboratory Conference on Applied Research in Data Engineering, University of Arkansas at Little Rock.

Year Four Objectives

During 2004, the Acxiom Laboratory for Applied Research has among its objectives:

1. Successful conclusion to the "first round" of the new proposal solicitation and funding process and the start of a second round.
2. Identification of new schools for membership or affiliation with the Laboratory.
3. Successful conclusion and hand-off to Acxiom stakeholders in August 2004 of the deliverables from the project:

An Architecture for Capturing, Storing, and Connecting Event Information to Facilitate Customer Data Integration

Dr. Charles Ford (UALR, Computer Science); Dr. Chia-Chu Chiang (UALR, Computer Science); Dr. John Talburt (Acxiom); Dr. Tom Schweiger (Acxiom); Steve Sieloff (Acxiom); Hao Wu (UALR, Graduate Student); and Radhika Chilka (UALR, Graduate Student).

4. Continued support of proposals seeking funding from the National Science Foundation and other national funding institutions for projects related to the Laboratory's research objectives, including:

Animated Visualization for Extracting the Underlying Processes that Affected Business and Social Data over Time

Dr. Mihail Edi Tudoreanu (UALR, Information Science); Dr. John Pittenger (UALR, Department of Psychology); and Don Hinman (Acxiom Corporation)

5. Advance planning for the 2005 Conference on Applied Research in Information Technology for Spring 2005 with a format similar to the previous three years for the dissemination of applied research conducted through ALAR and through other state and regional universities.