

## Personalization: Definition, Status, and Challenges Ahead

**Won Kim**, Cyber Database Solutions, Austin, Texas

### **Abstract**

As is often the case with a good marketing buzzword, the term personalization is used rather loosely today. It has come to stand for an ultimate goal of customer relationship management by businesses. It has also come to mean delivery of information of high relevance to an individual. In any case, given the huge and rapidly growing amounts of computerized information, and the unprecedented level of competition for customers, personalization is one of the most important trends in data processing and businesses. This article aims to offer a reasonably encompassing definition of personalization, examine techniques in use today to support personalization, and provide directions for future research and development for realizing the full potential of personalization.

### **1 INTRODUCTION**

The term “personalization” is widely and loosely used today [CACM 2000]. It is used in the context of receiving from a large body of information only the part that is of interest to an individual or a group of individuals. Examples include receiving only information about comedy and sports programs in television program guides, stock trading histories of only the past 3 months of 30 selected companies, announcements on the business wire of only XML-related software products, updates to websites of companies providing wireless communication services, etc. The term personalization is also used in the context of supporting one-to-one marketing, both in conventional and electronic commerce. Examples include displaying certain products or services or information on a web page that may be of potential interest to a particular website visitor as soon as the website visitor visits the web page; identifying potential customers of a new product from all existing customers, and sending them promotional materials or offering promotional deals; etc.

The objectives of this article are as follows:

1. Provide a reasonably clear definition of the term personalization on the basis of meaningful applications and objectives of such applications.
2. Summarize techniques used to support personalization.
3. Identify challenges, and therefore, directions for future research and development, for personalization.

## 2 DEFINITION OF PERSONALIZATION

As discussed earlier, the term personalization has at least two distinct origins. In fact, this is the cause for confusion about the meaning of the term in the market today. One is the vast sea of information that has been or can be stored in computers and disseminated through communication networks, including the World Wide Web. The amounts and types of information available for access through computers today have simply overwhelmed the capacity of people to sift only information that is relevant and useful to them. The huge number of websites around the world (currently estimated to be 3 billion) and the huge volumes of digitized information, ranging from personal files on desktop computers to data warehouses, to satellite transmissions, digital libraries, digital newspapers, digitized books, magazines, and articles, etc. have brought about a veritable information overload for people. The “push” technology was introduced several years ago by startups such as Marimba and Castanet as a means of gathering and delivering only the information that an individual user specified. Since then similar concepts have been incorporated into many information portal sites, such as Yahoo (myYahoo).

The objective of personalization for the purpose of delivery of personalized information is fairly straightforward. It is to deliver information that is relevant to an individual or a group of individuals in the format and layout specified and in time intervals specified. When information sources are updated, it is important that updated information be delivered to individuals. Updated information may be delivered immediately upon updates to the information sources or based on a schedule specified by the individuals (e.g., once a day, once a week, once a month) or by a system default.

The information delivered may be the result of a direct retrieval from an information source, or the result of transformations on information from one or more information sources. Examples of direct retrieval from an information source include an image of Thomas Edison from an image file, a WORD file describing the Wright Brothers' flight of Kitty Hawk, a PDF file of a research paper submitted to ACM Transactions on Internet Technology, etc. Examples of information delivered that requires transformations include the Web pages displayed by a Web browser of a document stored in an HTML or XML format in a website; a summary (produced, for example, by IBM's Intelligent Miner for Text text-mining software) of an article about Florida recount of the 2000 US Presidential Election ballots stored in the Wall Street Journal online edition; stock quotes (laid out in a spreadsheet form) of 30 Internet companies at yesterday's closing time of NASDAQ, etc.

Further, delivery of information includes consideration of the format (e.g., HTML, PDF, WORD, ASCII, LaTex, etc.) and layout in which information requested by an individual is delivered.

A second origin of the term personalization is the concept of one-to-one marketing in which a business does marketing tailored to a group of individual customers rather than to the entire population of its geographical marketing territory. The motivation for one-to-one marketing is to increase the revenue and decrease the loss for a business by understanding the needs, habits and lifestyle, preferences, likes and dislikes of its customers, and addressing (or at least giving the illusion of satisfying) customers' individual needs and preferences. The idea is that by understanding the needs, attitudes, and preferences of its customers, a business may tailor different marketing campaigns, and pricing and distribution strategies for different categories of customers, thereby becoming more successful in acquiring new customers, retaining existing customers, and selling additional goods and services to existing customers. Further, a business may be able to reduce financial losses and expenses by cutting the cost of acquiring new customers, preventing significant defection of existing customers, and detecting and preventing risky business transactions (e.g., fraudulent insurance claims, defaulting on a loan, fraudulent uses of credit cards, etc.) with some of the customers.

Personalization in the context of one-to-one marketing is by no means a new concept. In fact, this is as old as trade and business. The owner of a neighborhood store remembers his/her regular customers, engages in social chats with them, informs them of the arrival of certain new products; a waiter/waitress in a restaurant remembers his/her regular customers and their preferences, and suggests that "tonight really fresh Salmon is available", etc. In a sense, current one-to-one marketing based on personalized information is a rather "impersonal" attempt to reach customers they do not "personally" know. However, given the need to reach hundreds of thousand or tens of millions of customers for a line of products, this "impersonal" personalization is unavoidable.

The objective of personalization for one-to-one marketing should be considered from two different perspectives: business and customers. For businesses, as observed above, the objective is to increase revenue and/or decrease costs and losses. For customers, the objective is to receive useful and timely recommendations to purchase goods or services in the most favorable terms. Of course, the idea is that a customer's preferences and needs must be determined precisely and in a timely manner. It is not helpful to anyone if a customer's preferences and needs are erroneously determined (e.g., a Neil Diamond fan is offered a suggestion to buy a new Metallica CD; an employee laid off by a failed dot com company is asked to buy a Mercedes Benz, etc.) or determined after the window of marketing opportunities has closed (e.g., high-school text books are suggested to a family after the student graduates).

In this context, personalization does not refer to delivery of information, but rather to marketing efforts to individuals using information about the individuals. The information about the individuals includes direct information about the individuals (Domingo Chavez's census data, Louise Jefferson's purchase history at the Macy's, Arch Bunker's

motor vehicle registration and house mortgage applications, Frank Tarkenton's lifestyle data, etc.). It may be a large set of profiles or rules that may be deduced from a large database of demographic, lifestyle, commerce transaction records of hundreds of thousands or tens of millions of individuals over several years. Sometimes the profiles or rules are deduced offline, and other times they are deduced in real-time. In any case, on the basis of information about certain characteristics of an individual (e.g., income level, section of residence, age, ethnicity, sex, occupation, number of children, number and types of automobile owned, real estates owned, etc.), the individual is placed into a category of individuals who share some of the same characteristics. On the basis of the recorded history of commerce behavior of an individual or the category to which an individual belongs, the types of one-to-one marketing efforts appropriate for the individual are determined. Promotional materials are sent to targeted individuals; discounts are offered on certain goods and services for a certain duration; branch offices and stores are opened in a certain area of a city; certain goods and services are discontinued or newly offered; recommendations are made to visitors of e-commerce sites to suggest purchases of additional goods and services or more expensive goods and services; applications for loan or approval of purchase on credit are denied, etc.

The target of personalization, for both one-to-one marketing and delivery of selected information, is a group of individuals who share common interests or characteristics (e.g., Neil Diamond fans, college basketball fans, Italian Americans, etc.). The size of a group may range from one individual to a very large number of individuals.

On the basis of the above discussion, we may now define personalization as delivering to a group of individuals relevant information that is retrieved, transformed, and/or deduced from information sources.

### 3 PERSONALIZATION TECHNIQUES

Let us examine techniques used for personalization. The discussion will expose capabilities and limitations of personalization today.

The delivery of personalized information relies largely on push technology, Web search engines and Web crawlers (or Web robots) [Sullivan 2000], and document format conversion software/facilities. The push technology is used to gather specified information on the Web and deliver it in user-specified format and layout. Web search engines, such as Alta Vista and Google, and Web crawlers support keyword- and URL-based searches of the Web by maintaining indexes on keywords in web pages. Web directories, such as Yahoo, organize websites into a classification hierarchy to help more precise searches of companies and organizations in particular markets or business sectors. Meta-search engines, such as Metacrawler, query multiple search engines and merge the results. Once a reliable source of information has been found, Web crawlers, which are agents, may be used to retrieve Web pages on a regular basis.



Document format conversion software/facilities exist to convert documents between formats; from WORD to PDF, from WORD to ASCII, from WORD to HTML, etc. Web browsers convert and display Web pages stored in HTML, XML, WML, etc. formats.

The personalization for one-to-one marketing uses a range of techniques on a variety of data sources. Data sources include relational databases holding web logs, data warehouses of customer demographic, lifestyle, and transaction data, rule bases or profile databases that hold summarized profiles of general population, etc. Personalization techniques currently in use include the following:

### **Lookup of personal records in a database/file**

A database of personal records of customers is maintained, and keywords (e.g., name or customer ID) for a given customer are used to identify the customer's record in the database, and relevant information about the customer is extracted from the record. Relevant information about the customer may be stored in a single record, or in many records in several tables/files. Relevant information would be different from application to application. In general, it would include demographic data, lifestyle data, history of product purchases, history of customer support inquiries, credit data, etc.

### **Lookup of a rule base or a profile database**

A rule base or a profile database is maintained and is searched based on given characteristics (e.g., any combination of income level, sex, age, area of residence, ethnicity, etc.) of a customer. The rule base or profile database may have been created through use of data mining, weblog mining, text mining, and OLAP techniques (see below). Such techniques may be applied either offline or online against customer or transaction records stored in data warehouses [Inmon 96][Kimball et al. 98], depending on application requirements and performance and scalability limitations of the algorithms and available computing resources.

### **Data mining (of numerical and string data in formatted tables/files)**

Most existing data mining techniques are designed to process numerical (both continuous and categorical) data, and short-string data [Berry and Linoff 97][Han and Kamber 2000]. The techniques serve a variety of purposes, including clustering (grouping of data based on similar characteristics), classification or categorization (placing new data into one of the existing categories), association rules generation (determining the likelihood of events following certain other events, such as a customers' purchasing of shoes after purchasing suits), etc. Data mining algorithms may be run on a large number of customer records to segment customers or classify customers. They may be run on a large number of store transaction records to determine association rules.

### **Weblog mining**

Weblog data is stored in formatted tables/files. Weblog data is simple in that it contains only 10-15 fields; however, it can grow very large very fast, because every visit to a web page is logged. Weblog mining is a special case of data mining [Mena 99]. The objective is to determine a variety of access patterns to a website and web pages, such as the peak periods of access, average visit durations for different Web pages, common navigational patterns across web pages, repeat visits, identities of the websites that frequently lead to the website in question, etc. Various types of information that can be determined from analyses of weblog data are clearly useful to owners of websites. However, the absence of a visitor's name or personal identity (such as the social security number) limits usefulness of weblog data in one-to-one marketing.

### Text mining

Text mining aims to automatically determine various attributes of a free-form text (e.g., a news article, office memo, technical report, etc.) including key features, frequently occurring words, summary, category, etc [Sullivan 2000][IBM]. Key features include, for example, names of people or organizations or products, locations, dates, prices, etc. Frequently occurring words and certain phrases, such as "in conclusion", are used to help automatically generate a summary of a text. Key features and keywords are compared against keywords maintained in a knowledgebase to determine the category for a text (e.g., Internet taxation, professional basketball, luxury automobiles, social impact of the Internet, etc.).

### OLAP

Online analytical processing (OLAP) techniques are used to compute (numerical) summaries (or aggregates, such as total, subtotal, average, maximum, minimum) of data (known as measures, such as sales amount and quantity) from a number of dimensions (e.g., sales quarter, sales region, product lines, sales teams, etc.) [Thomsen 97][Berson and Smith 97]. OLAP usually involves grouping of records by certain fields (dimensions) and computing summaries of data for each group. For example, summaries of sales may be obtained by first grouping all sales records by product lines, and then computing summaries of sales for each group of records for the same product line.

## 4 TECHNICAL CHALLENGES

Computerized personalization is beginning to be applied to customer relationship management, electronic commerce and information portal services. As these application areas are fairly new, there are lots of technical challenges facing personalization. Let us examine them.

1. It is necessary to substantially reduce the number of irrelevant "hits" returned in Web searches. One of the most serious problems in delivering personal information today is the inability of the technology to allow people to easily and precisely specify the

information they want, or to understand the semantics of the contents of multimedia documents. The result is a large number of totally irrelevant “hits” by today’s Web search engines (e.g., causing people to think the book “Naked in Cyber Space” a pornographic website), the retrieval and manual scanning of large textual documents that contain a small amount of desired information or none at all, etc.

A solution lies in making search conditions much more versatile than are recognized by today’s Web search engines. It should be possible to specify search conditions that include keywords connected by the Boolean operators (AND, OR, NOT), comparators (=, >, <, etc.), set operators (CONTAINS, IN, Intersection, Difference, Union), temporal operators (before, after, etc.), and spatial operators (near, within, intersects, contains, etc.). Experiences in the use of relational database systems, and results of research into temporal data management and spatial data management should be leveraged.

2. It is necessary to generalize the “search keyword” to a “general object”, including a free-form text, an HTML or XML document, and eventually a multimedia object such as an image, video clip, or a speech fragment. Web search engines will then retrieve “general objects” on the Internet that exactly match a query object (e.g., a sample text, or a sample image) or that are “similar” to the query object. Measures of “similarity” between a query object and objects on the Internet differ depending on the types of object. For free-form texts, measures of similarity may include the subject category for the text, weighted sum of the matching keywords in the texts, etc. For XML documents, measures of similarity may include similarity of the structure of the documents, weighted sum of the matching terms, etc.

Once further advances have been made in understanding free-form text or such texts in XML documents, it is desirable to deliver texts in multiple levels of detail in accordance with what people desire.

There is little hope that technology will advance sufficiently in the foreseeable future to allow content matching of sound and video. However, there have been reasonable advances in speech recognition and image analysis, and it may be possible to apply them to Web searches in the foreseeable future.

3. It is necessary to further enrich the weblog data by linking a weblog record with related demographic, lifestyle, or transaction history data stored in a data warehouse (or a database) and/or to the actual web pages involved. The types of information captured in a weblog are rather limited, and the usefulness of the results of analyzing weblog data too is limited. Once weblog data can be linked to related records in a data warehouse, and/or the contents of the web pages that are referenced in a weblog can be meaningfully understood, a complete base for segmenting website visitors, and predicting their behavior will have been formed.

4. One of the most important enabling technologies for personalization is data mining and text mining. Despite some successes it has demonstrated (e.g., in detecting frauds in credit card or phone card uses, medical insurance claims, even discovering stars in a galaxy and new chemicals, etc.), data mining technology needs significant advances in ease of use, predictive accuracy, and performance and scalability. Further, technology

for creating a subject hierarchy remains a key challenge to classifying texts by text mining tools. The depth, breadth, and structure of a subject hierarchy are difficult even for humans to consistently and appropriately define. For example, where would one place a story about “a farmers’ march on the capital to demand forgiveness of debts” – economy, politics, civil unrest? –, “Bill Gates makes a \$100 million donation to promote arts and public education” – business, education, arts?

5. Further advances in detecting and correcting dirty data are necessary. If personalization for one-to-one marketing is done using dirty data, the result can range from wrong recommendations that may alienate customers to real damages to people. For example, if Pedro Gonzales’ income is recorded as \$200,000, when it is actually \$20,000, a credit card application system may approve a wrong credit limit, an electronic commerce website may recommend a lot more expensive goods or services than what Pedro Gonzales can afford, etc. Dirty data includes wrong data, missing data, and data that does not conform to a standard form [Kim et al 2002]. Dirty data arises from errors in data entry (mistyping, misspelling, missing information), incomplete consolidation of multiple independently created databases (different units of measure, redundant or conflicting entries for the same data, etc.), etc. Despite the importance of high-quality data, today inadequate attention is paid to data quality [English 99]. There are a few data-cleansing software tools on the market [FirstLogic][Validity][Trillium 98]. Such tools should be used and better techniques should be developed to detect and cleanse dirty data, and assess the impact of dirty data on applications.

6. Performance and scalability issues are always key technical challenges when processing a large volume of data that may be accessed by a large number of simultaneous users. The need for real-time recommendations in certain commerce situations (e.g., when a visitor visits a website, display several products of potential interest to that visitor) and the need to process very large volumes of customer data to determine a categorization for an individual are two of the factors that make timely delivery of personalization information difficult. Customer demographic and lifestyle data, and customer transaction data are typically large. In order to be able to offer personalized recommendations in real-time to electronic commerce website visitors, or to analyze huge volumes of data in time to reflect the results to business decisions, all of the techniques developed to address performance and scalability issues in relational database systems and other data-processing systems must be brought to bear. The techniques include parallel processing, indexing and hashing, use of a fast sorting package, prefetching of data from secondary storage, tuning of performance parameters, etc. In particular, parallel processing includes pipelined processing, partitioned database, symmetric multiprocessing, etc. In pipelined processing, application processing is organized into multiple processing steps and one step starts as soon as the results of the previous step become available. In partitioned database processing, a database is divided into multiple partitions, and each partition of the database is assigned to a different computer, so that multiple computers may process the same application on different parts of the database simultaneously. Indexes (often B-tree indexes) and hash tables are used to



allow quick identification of only the records in a database that satisfy certain search conditions.

## 5 NON-TECHNICAL CHALLENGES

Personalization for one-to-one marketing by businesses may sometimes backfire on the businesses. This may happen if personalized marketing efforts end up turning customers off, or if the use of customer data violates or appears to violate customers' privacy. Let us examine these two non-technical challenges to personalization in turn.

There is an often over-looked aspect of one-to-one marketing based on personalization. It is the psychology of the individual who is the target of one-to-one marketing. Unless done properly, all the efforts to sell, "cross-sell", and "up-sell" to a customer (e.g., a visitor to an e-commerce website), to have the customer defect from a competitor (e.g., to switch from AT&T to MCI WorldCom for long-distance phone services), etc. run the risk of irritating and alienating the customer with a feeling that he/she is "bombed and being squeezed for the last dollar". In today's highly competitive business world, the risk is indeed high.

Personalized marketing efforts may turn off customers if they are excessive or irrelevant. Excessive marketing includes offering too many recommendations (e.g., too many items suggested for purchase), "bombarding" customers (every week, or every time a customer visits a website), and "encumbering" customers (e.g., requiring too many questions to answer, poorly designed voice-based directed-dialogs for access to the Internet). These are the result of personalization done poorly, due to lack of consideration for human reaction to marketing. Irrelevant marketing includes inaccurate or irrelevant recommendations. These are often the result of recommendations based on insufficient or erroneous data, or inadequate training of the data mining algorithms.

A great deal of concern has been voiced over increased opportunities for invasion of privacy that increased use of the Internet and spread of electronic commerce may cause. The Internet has already proven to be a great means of disseminating information instantly worldwide and obtaining all sorts of information instantly from anywhere in the world. Sensitive information about an individual (e.g., bank records, medical records, school records, employment records, court records, income and tax records, purchase records of big-ticket items, even compromising photographs etc.), stored in lots of places (e.g., banks, hospitals, schools, employers, courts, accounting firms, department stores, etc.), may be placed on a website for instant distribution worldwide. Unauthorized dissemination of sensitive information about individuals may be the result of hackers' pilfering such information, unscrupulous persons (e.g., employees in a credit card company, banks, government agencies, hospitals, etc.) having access to such information, an accident or breakdown in internal procedures for issuing such information, or businesses' selling such information for commercial benefits, legally or otherwise. Of course, unauthorized dissemination of sensitive information can happen (and has

(happened) without the Internet. The instant, worldwide reach of the Internet simply makes the damage it can cause substantially greater than ever before.

Although concerns about privacy are to some extent legitimate, it is also to some extent irrational. For someone who buys a series of Tom Clancy and Frederick Forsythe novels, or shop for ordinary clothes and shoes at the Macy's, visiting IBM and Cisco's websites, there is no reason to be concerned that his/her purchase records or weblog data may circulate around the world. Further, there is in general no cause for concern if information about an individual, even sensitive information, is used merely as a part of broad statistical information (e.g., the number of people in Dallas who purchased a BMW 528i in 2000, a percentage of men who use escort services, a percentage of people who under-report their incomes to the Internal Revenue Service, etc.).

There is cause for concern, however, for someone who subscribes to a pornography website, who engages in banking transactions for the purpose of money laundering, under-reports income for the purpose of evading tax, goes to a hospital to be treated for some sexually transmitted disease, lies to his/her employer about his/her employment history or educational background, etc. Of course, there is the danger that a perfectly innocent person may suffer from wrong information about him/her; it often takes major efforts, expenses, and a long time to correct damages inflicted by such misinformation.

Concern over privacy will (and should) curve unbridled efforts for personalization for one-to-one marketing by businesses. There should be laws regarding whether businesses may distribute (for money or for free) their customer data (to other businesses, government agencies, individuals, etc.) without authorization by the customers, and, where they may, what types of customer data they may distribute. The lawsuits involving DoubleClick's selling customer data, and recent attempts by failing dot com companies to sell their customer data as part of bankruptcy liquidation point to a clear need for such laws.

However, there should be no problem with a merchant (online or offline) maintaining a history of purchases and demographic data that each customer has filled out, and using such information to offer better services to that particular customer (e.g., suggesting new products or services to purchase, even sending a birthday greeting card every year). There should be no problem with a merchant using customer data to understand various general trends among customers, and using such information to revise business strategies from time to time.

## 6 CONCLUDING REMARKS

This article examined two different origins of the term personalization, and provided a definition of personalization that encompasses both. Then it reviewed techniques in use to support personalization, and examined both technological and non-technological challenges that face personalization. In view of the huge and growing volumes of computerized information, on the World Wide Web and in corporate and government



data warehouses, the need to be able to deliver only information of direct relevance to an individual for a specific purpose at any point in time is clear. Further, from the need for a business to reach a large customer base across a large geographical territory, and the need to tailor marketing and business practices to different categories of customers are also clear. For these two reasons, personalization is one of the clear trends in data processing in this era of the Internet and electronic commerce. If the challenges reviewed in this article can be met, the full potential of personalization can be realized for the benefit of businesses and consumers alike.

## BIBLIOGRAPHY

- [Berry and Linoff 97] M. Berry and G. Linoff, *Data Mining Techniques for Marketing, Sales and Customer Support*, John Wiley and Sons, 1997.
- [Berson and Smith 97] A. Berson and S. Smith, *Data Warehousing, Data Mining, and OLAP (Data Warehousing/Data Management)*, Computing McGraw-Hill, 1997.
- [CACM 2000] *Communications of the ACM*, August 2000 Special Issue on Personalization.
- [English 99] L. English, *Improving Data Warehouse and Business Information Quality-Method for Reducing Costs and Increasing Profits*, Wiley & Sons, 1999.
- [First Logic] First Logic Inc., “Customer Data Quality- Building the Foundation for a One-to-One Customer Relationship,” White Paper, <http://www.firstlogic.com/>.
- [IBM] IBM Corporation
- [Han and Kamber 2000] J. Han and M. Kamber. *Data Mining Concepts and Techniques*, Morgan Kaufmann.
- [Inmon 96] W.H. Inmon, *Building the Data Warehouse*, John Wiley & Sons, 1996.
- [Kim et al 2002] W. Kim, et al. “A Taxonomy of Dirty Data”, to appear in *Journal of Data Mining and Knowledge Discovery*, June 2002, Kluwer Publishers.
- [Kimball et al 98] R. Kimball, et al., *The Data Warehouse Lifecycle Toolkit : Expert Methods for Designing, Developing, and Deploying Data Warehouses*, John Wiley & Sons, 1998.
- [Mena 99] J. Mena. *Data Mining Your Website*, Digital Press, 1998.
- [Sullivan 2000] D. Sullivan. “Eye on the Competition”, *Intelligent Enterprise*, September 8, 2000.
- [Thomsen 97] E. Thomsen. *OLAP Solutions*, John Wiley and Sons, 1997.

- [Trillium 98] Trillium Software System, "A Practical Guide to Achieving Enterprise Data Quality," White Paper, <http://www.trilliumsoft.com/>, Trillium Software, 1998.
- [Vality] Vality Technology Inc., "The Five Legacy Data Contaminants You Will Encounter in Your Warehouse Migration," White Paper, <http://www.vality.com/>.

### About the author



**Won Kim** is President and CEO of Cyber Database Solutions ([www.cyberdb.com](http://www.cyberdb.com)) and MaxScan ([www.maxscan.com](http://www.maxscan.com)) in Austin, Texas, USA. He is also Dean of Ewha Institute of Science and Technology, Ewha Women's University, Seoul, Korea. He is Editor-in-Chief of ACM Transactions on Internet Technology ([www.acm.org/toit](http://www.acm.org/toit)), and Chair of ACM Special Interest Group on Knowledge Discovery and Data Mining ([www.acm.org/sigkdd](http://www.acm.org/sigkdd)). He is the recipient of the ACM 2001 Distinguished Service Award.