Designing for ROI: Toward a Value-Driven Discipline for E-Commerce Systems Design

Jungpil Hahn
Doctoral Program
jhahn@csom.umn.edu

Robert J. Kauffman
Professor and Chair
rkauffman@csom.umn.edu

Jinsoo Park
Assistant Professor
park@umn.edu

Information and Decision Sciences, Carlson School of Management,
University of Minnesota, Minneapolis, MN 55455

Abstract

*E-business management is an on-going process of understanding consumer needs and developing online solutions to meet those needs To be effective in today's competitive environment of e-commerce, e-businesses cannot afford to neglect justifying the return on investment (**ROI**) of their online operations. Ideally, companies would like to understand how users are using their Web site and how this translates into value creation (or value leakage) so that online solutions can be implemented to increase business performance. However, the current situation is that such value-driven management is not possible due to the lack of tools and methods that can provide a clear link between systems design and business performance. The purpose of this paper is to present a value-driven systems design methodology for e-commerce Web sites. The focus of our methodology is on identifying value-creating opportunities and value-diminishing problems through the analysis of actual customer Web site usage behaviors through Web usage mining. The identification of value-creating opportunities and value-diminishing problems enables us to plan for designing additional Web site features and for redesigning existing functionalities to increase business performance and maximize design ROI.*

KEYWORDS: Data mining, electronic commerce, information foraging theory, IT value, return on investment, systems design, Web-based applications.

1. Introduction

The industry has progressed to the second phase of electronic commerce. The first phase of e-commerce was relatively simple: the goal for most companies was to secure a share of the virtual market space with a strong online presence by attracting as many visitors as possible to their Web site. Whereas now, as e-commerce begins to "grow up,"the ability to conduct online operations justified by return on investment (**ROI**) is the only way e-businesses can survive.

Recent industry analyses, however, point out that e-commerce companies are earning low scores on ROI, by failing to meet consumers' purchase needs with the poor usability and errant designs of their Web-based storefronts [1]. For example, a study conducted by A.T. Kearney showed that 80% of experienced online shoppers gave up shopping on e-commerce Web site sites due to problems that they encountered while interacting with the Web site [2]. Yet another study conducted by Creative Good showed that 43% of purchase attempts ended in failure due to poor usability of the Web sites [3]. This shortfall in realized value compared to the potential value that Web-based selling approaches offer is dramatic. The Creative Good study points out that this level of failed purchase attempts is consistent with an estimated loss of $14 billion in sales for e-tailers in the 2000 Christmas-New Year's holiday shopping season alone. Recent academic research reinforces the picture that emerges. Apparently the quality of the online customer experience that effectively-designed Web sites create not only has a positive effect on the financial performance of a firm, but it also possesses the potential to create unique and sustainable competitive advantage for Internet-based sellers and other e-commerce firms [4].

Developing, launching and maintaining a Web site is a significant investment for e-commerce firms. E-commerce firms typically budget $1-2 million per year for Web site setup and maintenance, whereas leading Web sites require annual investments nearing $8 million [2]. Despite the importance of Web site development requiring such significant investments, the process of designing high quality Web sites for e-commerce is still more of an art than a science. E-commerce companies are still relying largely on intuition when it comes to designing their Web sites. To make matters worse, design changes and their impacts are not tracked, making it impossible to measure the benefits of site design [5]. It is clear that e-commerce companies need disciplined methods for justifying the business value of IT investments in systems design.

In this paper, we propose a value-driven approach to systems design for Web-based e-commerce applications. In order to effectively manage the activities related to systems design, we apply the concept of *return on investment* — just as the performance of financial investments can be evaluated, the value of systems design, which we regard as an IT *investment*, also needs to be measured, assessed and managed via an *ROI trajectory*.

The purpose of this paper is to present a value-driven systems design methodology for e-commerce Web sites. The major focus of our methodology is on measuring the value of systems design. We measure the impact of site design on business performance by analyzing customers' patterns of Web site usage with Web usage mining techniques. The impact of systems design on business performance can then be estimated by relating Web site usage patterns with business performance metrics.

2. Literature review

The effective design of an information system depends on a deep understanding of how users are using (and should use) the information system. However, one of the major difficulties for systems design for e-commerce arises due to the difficulty in acquiring a clear understanding of how the e-commerce Web site is being used. Unlike with physical stores, e-commerce managers cannot *directly* observe the customers' behavior within the electronic storefront. Fortunately, Web servers log low-level system requests (i.e., HTTP requests) that can be potentially used to *indirectly* show how users are using the Web site. However, making sense of such low-level log data is not without problems and presents important research challenges. In this section, we review the key literature on Web usage mining to investigate the issues related to using Web log data for extracting usage patterns and then turn to the literature on online consumer behavior to gain an understanding of how to make sense of online usage patterns.

2.1. Web usage mining

Web usage mining can be defined as the application of data mining techniques to discover patterns from Web data (e.g., Web server logs) to understand Web usage behaviors to better serve the needs of the users of Web-based applications [6]. Early Web usage mining techniques (e.g., Open Market's Web Reporter, www.opemmarket.com, and NetIQ's WebTrends, www.webtrends.com) have focused on generating insights related to general site usage. For example, Web server logs can be compiled to generate descriptive statistics on Web site usage for questions such as: What are the most requested web pages within the site? Where are our users coming from? And how many times was a particular banner advertisement clicked on?

Current directions in research on Web usage mining have broadened the scope of analysis to perform even more sophisticated pattern extraction related to site usage. Examples of such applications include:

- **Association rules** enable an analyst to correlate the set of Web pages within a site that are accessed within single or multiple specific user or customer sessions [e.g., 7].
- **User clustering capabilities** enable inferences to be made about customer demographics for market segmentation based on observed click stream behavior on a Web site [e.g., 8].
- **Classification gauges** support the development of individual and group-member user profiles [e.g., 9].
- **Sequential pattern identification** has the power to predict future site navigation patterns based on current and prior observed click stream data and user patterns [e.g., 10].
- **Dependency modeling** develops preliminary descriptive models representing dependencies that occur among variables describing Web site use and the Web site user [e.g., 11].

Even though Web usage mining techniques may be capable of extracting interesting Web site usage patterns to deepen our understanding of how customers are interacting with the Web site, the vast majority of the Web usage mining literature has not been tailored towards e-commerce purposes. For example, the general literature on Web usage mining has focused on extracting information only from Web server log data. Hence, the proposed Web usage mining techniques may be used to uncover interesting data patterns in terms of association rules, user clusters and navigation patterns of users of the Web site. However, the uncovered knowledge from the mining process cannot be linked to business performance of e-commerce Web sites such as customer acquisition, retention, and sales effectiveness. Recent development in the Web usage mining literature has addressed this lack of integration of Web log data with marketing data and knowledge. For example, Büchner, Baumgarten, Anand, Mulvenna and Hughes [11] propose the MIMIC (Mining the Internet for Marketing IntelligenCe) architecture that integrates server data (i.e., Web server logs), marketing data (e.g., customer, product and transaction databases) and Web meta data (i.e., data about the e-commerce Web site such as site architecture and topology) in order to discover marketing intelligence related to customer attraction (i.e., what are common characteristics of existing visitors and customers and how can we use this knowledge to identify profitable potential customers?), customer retention (i.e., can we identify Web browsing patterns and associations across time and use this knowledge to display special offers and promotions dynamically to keep a customer interested in the Web site?), and cross-sales (i.e., infer customers' tastes from their prior purchase activities to recommend related complementary products). Similarly, Berry and Linoff [12] propose market basket analysis. This technique examines the content of Web shopping carts to infer patterns of product co-occurrence so that cross-sells and up-sells may be targeted on the fly via recommendation systems [13].

Even though the integration of Web usage mining techniques with marketing data provides important insights into the performance of e-commerce Web sites, there are still limitations when we consider the overall process of managing the e-business. E-business management is an on-going process of understanding consumer needs and developing online solutions to meet those needs. Hence, e-commerce Web sites need to be constantly redesigned and fine-tuned to meet the needs of online consumers. The purpose of current Web usage mining techniques for e-commerce is to identify marketing knowledge (e.g.: Who are our customers? What are they likely to purchase given prior purchase / browsing behaviors? And what are they likely to purchase given their demographic segment?). This can help determine effective promotion, advertising and / or personalization of the Web site.

An implicit assumption here is that the Web site under consideration performs well and the problem is to identify the right marketing mix so that customers will be effectively targeted. However, as outlined in the Introduction, Web sites cannot be assumed to perform at high business value standards. Customers frequently drop out of the purchasing process, not because a particular product is not available at the online storefront, but because they cannot find the product sought due to poor designs of the Web site. To make the analysis of Web site usage managerially-actionable, the focus of our systems design methodology is to make use of user behavioral data from Web site logs to identify problems and areas for improvement to increase business performance.

2.2. Online consumer behavior

Even though Web usage mining techniques provide useful ways to extract user behavior sequences, the interpretation of the extracted paths is still problematic because these techniques lack an underlying model of user actions and goals. In other words, to discover ways to improve a Web site, we need to understand in detail how users are using the Web site (e.g.: What were the user's goals? Did they succeed? If not, why did they fail?) This way, we may identify the reasons why a particular design is effective (or ineffective).

Information foraging theory [14], provides a useful lens for understanding user behavior in online environments. This theory seeks to explain human information-seeking and sense-making behaviors given the structure of the interface between people and information repositories. On the World Wide Web, users typically forage for information by navigating through pages of a Web site via hyperlinks. The actual content of the pages associated with these hyperlinks can only be inferred by snippets of text or graphics that are presented to the user as hyperlinks. Foragers use these proximal cues to assess the distal content and follow the link with the strongest *information scent* given their information need. Information scent is the user's imperfect, subjective perception of the value of information obtained from proximal cues. So, given an information goal of a user and knowledge of the structure and content of a Web site, we may predict how a user will navigate through the site in order to satisfy her goals: the *inference problem*. Conversely, given a user's navigation path and knowledge of the structure and content of a Web site, we may infer what the information goal of the user was: the *prediction problem*.

Based on the theoretical foundation of information foraging theory, two algorithms have been developed for solving the inference and prediction problems [15]. The IUNIS (Inferring User Needs by Information Scent) takes as input a user's navigation path and Web site topology, linkage and content information to output a set of weighted user information need keywords. The WUFIS (Web User Flow by Information Scent) algorithm is the exact reverse. WUFIS takes as input a set of user information need keywords and Web site topology, linkage and content information to output a predicted navigation path through simulation.

Even though the information foraging model and algorithms have been successfully applied in analyzing and predicting browsing behaviors on the World Wide Web [e.g., 16], one limitation of these information-seeking models is that they have been developed for general information-seeking tasks. Hence, they may not capture the uniqueness of the e-commerce environment [17]. The information foraging model and accompanying algorithms (i.e., IUNIS and WUFIS) are limited for application in the e-commerce domain in that they assume a unique goal (i.e., seeking information content) for a Web browsing session. The *goal uniqueness assumption* may not apply in the e-commerce context since a customer's interaction with the Web site typically involves several goals depending on the stage of the consumer behavior process [18; 19] or on the phase of transaction [20]. For example, Schubert and Selz [20] characterize the three phases of e-commerce transactions consisting of information (i.e., information gathering on potential products and services), agreement (i.e., reaching agreement when negotiations between customers and suppliers take place), and settlement (i.e., settling on the payments and logistics when the products and services are actually delivered). The goal of information-seeking may only be relevant to the information phase of the e-commerce transaction. For the agreement and settlement phases, the goal shifts to *task completion*. The customer's goal is not to find information relevant to price or payment, but to complete the task of filling forms with adequate information so that the price may be agreed upon and checkout may proceed. Even the information-seeking goal may not be at the appropriate level of aggregation in e-commerce contexts. Customers typically shop for multiple products, implying that that there may be several repetitions of the information-seeking goal with different information needs.

All the phases of the transaction process need to be effectively supported by the e-commerce Web site. Abandoned shopping carts are illustrative examples where the customer was successful at finding a product worth purchasing (i.e., the information seeking goal was achieved), but could not complete the purchasing transaction due to problems in the checkout process (i.e., the task completion goal for settlement was not achieved). If the Web site does not fully support all phases of the transaction process *value leakages* may occur. Value leakages occur when the *realized value* of IT in some applied setting compares unfavorably with the *potential value* that was envisioned when an investment was made [21].

We extend IUNIS and WUFIS and propose two new approaches: E-IUN (E-commerce Inferring User Needs) and E-WUF (E-commerce Web User Flow) that are more appropriate for e-commerce contexts. The first extension we make is to allow partitioning of the user sessions into several atomic tasks. In other words, if a customer's Web site interaction behavior involves several tasks such as log-in and authentication, several product searches, providing delivery information and checkout, E-IUN must be able to partition the session into these smaller subtasks, and infer the respective information-seeking or task completion goals. Accordingly, the E-WUF must be able to simulate a customer's entire Web interaction session when given a list of tasks to complete at the Web site. This is achieved by incorporating additional task goal related meta-data to the page topology, linkage and content information. As a result, we may detect a new atomic task by estimating significant deviations in the information need or task goal.

3. E-commerce systems design

To provide a realistic context and motivation for our proposed value-driven systems design methodology for e-commerce, this section provides a brief description of the systems evaluation and design practices at an actual e-commerce company. The rationale for this section is to verify the assumptions underlying our value-driven approach, as well as gauge the validity and feasibility of our proposed approach.[1]

With the aim of understanding the current practice of e-commerce systems design, we conducted an exploratory case study of an Internet-based seller of groceries. OnlineGrocery.com[2] is a pure-play e-tailer that delivers groceries directly to the customer's doorstep with the mission of "taking the dread out of grocery shopping." The company made its first delivery in April 1999, and by mid-July 2000, it had over 9000 customers generating more than $16 million in revenue. Currently, OnlineGrocery.com operates only in one metropolitan area, where it is the only online service provider within its regional market.

3.1. Web site design at OnlineGrocery.com

A significant source of ideas for redesign is OnlineGrocery.com's personnel. All employees are also regular customers of the online service and, hence, have intimate knowledge of the strengths and weaknesses of the Web site's design. The employees identify and suggest areas where the site may be redesigned to improve the overall shopping experience for OnlineGrocery.com's shoppers. Other sources of ideas to improve the Web site include the online feedback forms that permit customers to post questions and suggestions, the customer service call center, and occasional customer focus groups. These sources produce many more proposals for development projects than the software development staff and the IS budget at OnlineGrocery.com can handle.

To prioritize and decide on what development projects to pursue, OnlineGrocery.com holds monthly "Web Board" meetings where senior executives discuss potential development projects and set high-level priorities and business goals. A "Web Team" meeting is also held weekly, and the staff members discuss details of the development projects that are planned and underway. This process results in development projects being prioritized based on estimated impact on key business performance metrics (e.g., customer acquisition, customer conversion, dollar ring etc.), which act as loose proxies for ROI. However, the estimation is *ad hoc*, and intuition plays a more significant role than it should, given the critical importance of ensuring that the firm has a high-performance Web site as the front-end presence for interacting with its customers.

---

[1] Due to the limited sample size, we make no claims as to the generalizability of this case description. However, our discussions with several other e-commerce companies, as well as recent articles in the popular press [e.g., 5], suggest that the following description of the systems design and evaluation practices is typical and representative of many e-commerce companies today.

[2] The name of the organization has been disguised for confidentiality.

Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS-35'02)
0-7695-1435-9/02 $17.00 © 2002 IEEE

3.2. Web site performance evaluation at OnlineGrocery.com

Currently, the data for estimating business metrics are derived from two separate systems: (1) the marketing data warehouse and (2) a Web site analysis tool that is provided as a bundled service by the application service provider that hosts the firm's Web site. The data warehouse, which contains customer, product and transaction data, is used to conduct market basket analysis. For example, final sales statistics are used to answer questions such as: "What are our best selling products?" "What are the demographics of our customer segments?" And "What is the average profitability for each customer segment?" This analysis is valuable for assessing the overall performance of the online service. However, it provides very little managerially-actionable information about how to improve OnlineGrocery.com's Web site.

The Web site analysis tool is employed towards this second goal. The Web site analysis tool, *WebTrends*™, is a first-generation Web usage mining tool that compiles Web server logs to generate Web site hit statistics. The analysis tool offers a series of pre-packaged reports that show various aspects of online activity. For example, some reports list the most requested pages, whether the page hits come to OnlineGrocery.com through external "referring" Web sites, the browsers that are used by people who visit the site, the number of hits and visits for a given date range on different parts of the Web site, and the most frequently occurring HTTP errors. These reports are used to answer very questions such as: What are the most popular product categories? What are the most popular products? When do customers shop? "And how long are the typical shopping sessions?

3.3. Lessons and insights

We see that the analysis tools and techniques that OnlineGrocery.com currently employ are limited in that they provide two extreme views of performance. The analysis performed with the customer data warehouse is only able to convey information about high-level business performance (e.g., merchandising and marketing effectiveness). In contrast, the Web site analysis tools are only capable of depicting inflexible, low-level raw statistics of site usage. It is difficult to bridge the gap between high-level business performance and low-level site usage.

Table 1 presents OnlineGrocery.com's "wish list" of analysis capabilities that we distilled from our field interviews with the firm's senior management team and its Web operations staff. The reader should note that in contrast to the kinds of data that the tools we discussed can deliver, the "wish list" requires data that provide a richer understanding of detailed online consumer behavior. For example, OnlineGrocery.com's management asks (row 1): Where are customers actually putting items in the shopping carts? Do they use the specials page, the order history page, browse through product catalogs or conduct keyword searches?

The same product may be presented at any or all of the areas mentioned. However, the data warehouse can only suggest whether a particular product was bought or not, whereas the Web site analysis tool can only tell how many hits were generated at the catalog pages or the order history page or the specials page, etc. To effectively answer management's original question, we would need to understand the customers' purchasing process: Where did she come from? Where did she go to first? Did she look at other similar products? What were the prices of the other products? Was the product the same brand as the one she had previously purchased? And when and where was the "Add-to-Cart" button finally clicked?

This data and analysis required would need to bridge the gap between high-level business performance metrics and low-level site usage. Moreover, it is important to recognize the critical issue. Substantially all of the most actionable information that Web site evaluation and usage mining tools need to deliver to improve management is in this unmapped middle ground. It simply is not being made available at OnlineGrocery.com, and at many other firms that have adopted similar performance evaluation approaches based on the tools and techniques that currently exist in the marketplace.

Our exploratory case study of OnlineGrocery.com reveals several interesting insights into the requirements for measuring the value of systems design in an e-commerce setting:

- **Importance of business value.** OnlineGrocery.com's Web site analysis tool provides useful information to infer customer interaction with the Web site. However, it is difficult to associate behavior with ROI-related outcomes.
- **Clear understanding of consumer behavior.** Information needed to improve the Web site design should paint a detailed description of actual consumer behavior. For example, knowing that product X is the best selling product for this week is good to know, but there is nothing that can be done about it. Instead, it would be more appropriate to have specific information that indicates the effectiveness of the placement of product promotions, and the extent to which different kinds of placement create marginal impacts in sales. With this kind of information in hand, management would have more information about how to improve the design of the Web site so as to maximize the ROI associated with the redesign of the screen real estate.

Table 1. OnlineGrocery.com's "Wish List" of site performance analysis capabilities

MANAGEMENT'S QUESTIONS	ROI-ENHANCING ANSWERS	APPLICABLE BUSINESS LEVER
❑ Where are people hitting the buy button: order history, specials, categories, search results, etc.?	❑ Understand where sales are generated to estimate value of screen estate, and to guide which areas of the site to improve	❑ Dollar ring ❑ Margin
❑ How many clicks does it take people to get items in their carts?	❑ Guide design to reduce clicks	❑ Frequency
❑ What is the conversion rate of promotions? ❑ How many people saw / clicked on / bought promotion items?	❑ Understand the best way to promote	❑ Dollar ring ❑ Margin
❑ What are the top X searches with results?	❑ Infer customer needs	❑ Frequency ❑ Dollar ring
❑ What are the top X searches that yield 0 results?	❑ Populate product synonym list ❑ Improve product descriptions ❑ Infer customer needs	❑ Frequency ❑ Dollar ring
❑ Where are people aborting? ❑ How many abandoned carts are there by age of cart?	❑ Identify technical problems ❑ Guide design to reduce technical problems	❑ Frequency
❑ Do people shop in one trip? Or several? ❑ Does this vary by market / customer segment? ❑ What are average times per trip? ❑ What is average time to fill a cart, by combining multiple trips?	❑ Understand how people shop to redesign Web site for ease of use ❑ "Dial in" ROI in use-specific contexts ❑ Find a basis for identifying how to redesign to create greatest leverage for maximizing ROI by market / customer segment	❑ Frequency
❑ Are people browsing / looking for specific items when they shop?	❑ Improve information architecture	❑ Frequency ❑ Dollar ring
❑ What features are customers asking for on our site?	❑ Prioritize development projects	❑ Frequency
❑ Web availability problems? - System up 24 x 7? - Peak hours?	❑ Identify technical problems ❑ Explain variance in order volume ❑ Design to service maximum usage	❑ Frequency
❑ Web response time? - During shopping? - During checkout?	❑ Identify technical problems ❑ Identify situations where unacceptable site performance is most likely to create conditions for lost revenues	❑ Frequency
❑ State of the "digital shelves"? - number of items in the store - number of items in stock - number of missing photos ❑ Number of products out of stock during check out	❑ Understand customer shopping experience ❑ Identify dollar ring issues to determine product mix inefficiencies that may diminish site potential for generating revenues	❑ Frequency ❑ Dollar ring
❑ Customer's client technology? - What browsers are people using? - What platforms? - How many AOL users?	❑ Develop style guides ❑ Understand technology design space ❑ Define the range of "necessary standards" with which to comply	❑ Conversion
❑ Are credit cards being rejected?	❑ Understand customer shopping experience	❑ Conversion

- **Linking consumer behavior to business performance.** OnlineGrocery.com's technology support for performance evaluation was either too high-level (i.e., customer data warehouse) or too low-level (i.e., hit analysis). However, the most pressing questions were at the middle ground where low-level click stream information is mapped with high-level site performance.
- **Integrating Web site evaluation and design.** The goal of performance evaluation is ultimately to devise ways to improve business performance. The results of Web site evaluation should guide the generation of ideas for improving the Web site in this way.

4. Value-driven e-commerce systems design methodology

In this section we present our value-driven systems design methodology for e-commerce Web sites. We distinguish between systems *development* methodologies and systems *design* methodologies. Systems development methodologies focus on the "how" of systems development: the processes of systems analysis, design, implementation and maintenance of an information system and the effective management of such processes. We focus on systems *design* methodologies where the focus is on the "what" of systems development: what type of system functionalities do we need to implement in order to maximize business performance.

An important input into the design of an information system is the understanding of how users are actually using the system. Furthermore, to identify systems design initiatives that augment business performance we need to relate usage and business value. However, this is not a trivial matter. The integration of several heterogeneous data sources is necessary to provide this comprehensive view of usage and performance. Hence, an important precursor to our proposed systems design methodology is the ability to recreate an accurate depiction of Web site usage behavior and performance. We start by describing the data requirements, pre-processing procedures, before we discuss the details of our proposed systems design methodology.

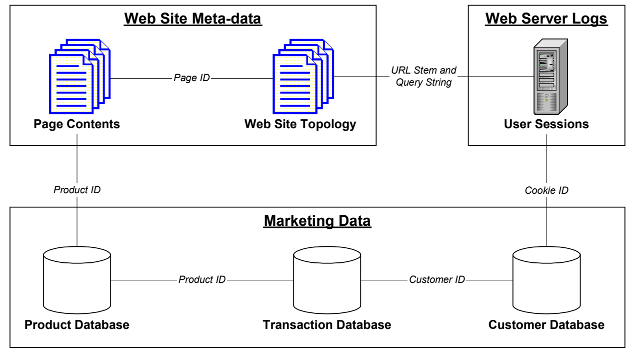

Figure 1. Data integration

4.1. Data sources

Consistent with recent advances in Web usage mining techniques for e-commerce [13; 22], we make use of three data sources for analyzing the value of e-commerce Web site designs: (1) Web server logs, (2) marketing data, and (3) Web site meta-data.

Web servers automatically log all client-server interactions in several log files: access logs, error logs and cookie logs. Error logs store data on failed requests (e.g., missing pages or images and HTTP authentication failures). Although error logs may be used to detect incorrect links or server capacity problems, its use has been proven rather limited for evaluating e-commerce performance. Cookie logs store customized information related to cookies which can be used to identify and track individual users even though they may have used different IP addresses between sessions. Finally, access logs are the most commonly used log files for Web usage mining. Access logs store detailed information about HTTP requests (i.e., date and time of request, client IP address, HTTP auth username, number of bytes transferred, server IP address, the URL stem of request, query string of request, HTTP status, HTTP version, user agent, cookie ID and referrer URL). Web usage mining techniques are performed on the access logs to extract user sessions and patterns of usage behaviors.

Marketing data are stored in information systems that e-commerce companies use to operate the e-business. Even though the level of sophistication of such information systems varies widely across companies, we may assume that to operate an e-business, e-commerce information systems typically store information about customers, products and transactions. The marketing data is used to track who the customers are and what products the customers have purchased. More sophisticated marketing information systems could also store promotional event data which tracks what marketing events (e.g., promotions, advertising, etc.) have taken place, which customers were targets for those events and how the customers have responded to those events.

Finally, Web site meta-data need to be defined to adequately represent the structure and content of the e-commerce Web site. Web site structural meta-data provide the topology of the Web site (e.g., entry points, the product hierarchy, link structures, etc.) that can be represented as a directed graph structure. Web site meta-data also need to provide semantic information related to the content on the individual pages, which can be represented as keyword matrices. Web site meta-data are used to make sense of the usage behavior patterns by identifying the nature and content of the pages the customers have gone through to browse and purchase products on the Web site.

The various sources of data can be integrated by joining the heterogeneous data with the appropriate fields. The mapping of the various data sources is presented in Figure 1. For example, cookie IDs from the user sessions extracted from the Web server logs are used to identify the respective customer in the customer database of the marketing information system. Likewise, the URL stem and query string information from the user sessions can be used to join the user sessions and the Web site topology to identify what pages a particular customer was viewing at a particular time. The actual content of the pages viewed can be inferred by joining the Web site topology and the keyword matrices by the page ID.

The integrated data can now be drilled with Web usage mining techniques to portray an accurate picture of customer Web site usage behavior. For example, we may extract individual customers' Web site navigation paths to infer the customer's purchase decision-making process.

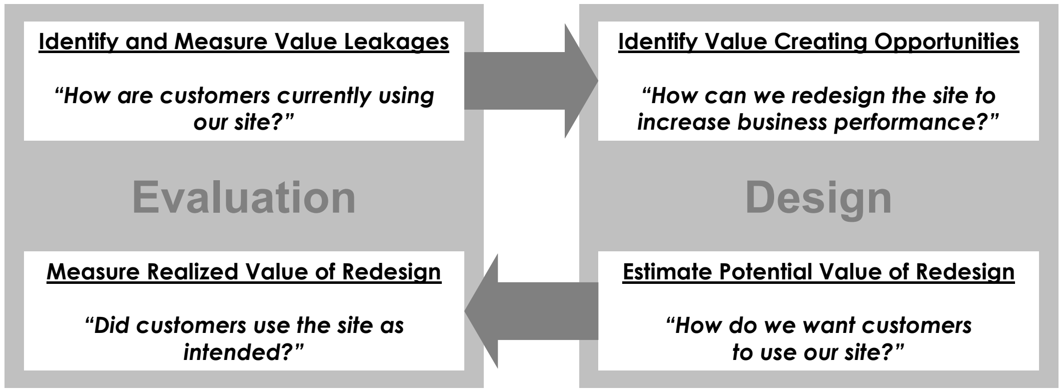

Figure 2. Value-driven system design methodology (high-level overview)

4.2. Systems design process overview

A high-level overview of our systems design methodology for e-commerce is presented in Figure 2. The systems design process is an iterative process of Web site evaluation and design, which is guided by the analysis of customer-Web site interaction. Web usage mining techniques are employed to understand how the Web site customers are currently using the Web site. We use our understanding of online consumer behavior to identify problem areas where value leakages occur. Subsequently, we may identify design opportunities that rectify the value leakages so as to increase Web site effectiveness in terms of desired business performance objectives. The goal of Web site improvement is to design affordances that are expected to bring about the desired business performance. Expected patterns of usage are then used to estimate the potential ROI benefits of the various design initiatives. For example, an often-cited problem with e-commerce Web sites is the problem of abandoned shopping carts due to extensive delays in the transaction processing. If a new design of the checkout process were to reduce the duration of the final transaction processing from 20 seconds to 10 seconds, we may be able to estimate through probabilistic models of delay tolerance the reduced number of abandoned shopping carts. This would enable us to estimate the value-added by the new design and compare it to the cost of redesigning and implementing the checkout process to produce an estimate of expected ROI. The expected ROI can be used to prioritize development initiatives. Finally, once the new designs are implemented, actual and expected usage patterns are compared in order to evaluate the realized effectiveness of the design initiative.

A more detailed depiction of the systems design process is presented in Figure 3. To prevent visual clutter, marketing data is not presented in the figure. The step-by-step process is as follows. Server data (i.e., access logs) are cleaned and prepared to extract user sessions and the constituent atomic task paths. In parallel, Web meta-data are extracted from the current Web site to define the content and topology of the Web site. These Web meta-data definitions, along with the atomic task paths, are inputs into the E-IUN algorithm to extract customer information needs. The inferred information needs are mapped to the marketing outcomes (e.g., success or failure purchase) to infer the effectiveness of the Web site in meeting customer needs. This understanding allows us to plan for design and redesign initiatives to rectify Web site problems or increase business value. There may be several ideas for design / redesign. For each design / redesign initiative, the expected Web site design is re-analyzed to extract the content and site topology definitions. The Web meta-data for the expected new design, along with the information needs of the customers inferred previously, are inputs into the E-WUF algorithm to produce predicted usage paths for various atomic tasks, which can be analyzed to predict the performance of the new design. The predicted performance is mapped to the marketing data to estimate the potential ROI benefits for the design / redesign initiative. Based on the estimated ROI for each design initiative, we may select the one with the greatest potential ROI. To effectively manage the ROI trajectory, we continually refine the parameters for estimating ROI by comparing the potential ROI estimated in the design phase with the realized ROI once the design is in production mode.

This process is iterative in that the effectiveness of the selected design initiative is evaluated by starting the process all over again with the newly designed Web site as the current Web site. The advantage of our approach over previous approaches to Web usage mining is that we take into consideration the dynamic nature of e-commerce Web sites that are constantly changing, whereas previous Web usage mining techniques have implicitly assumed a more design that is relatively stable over time.

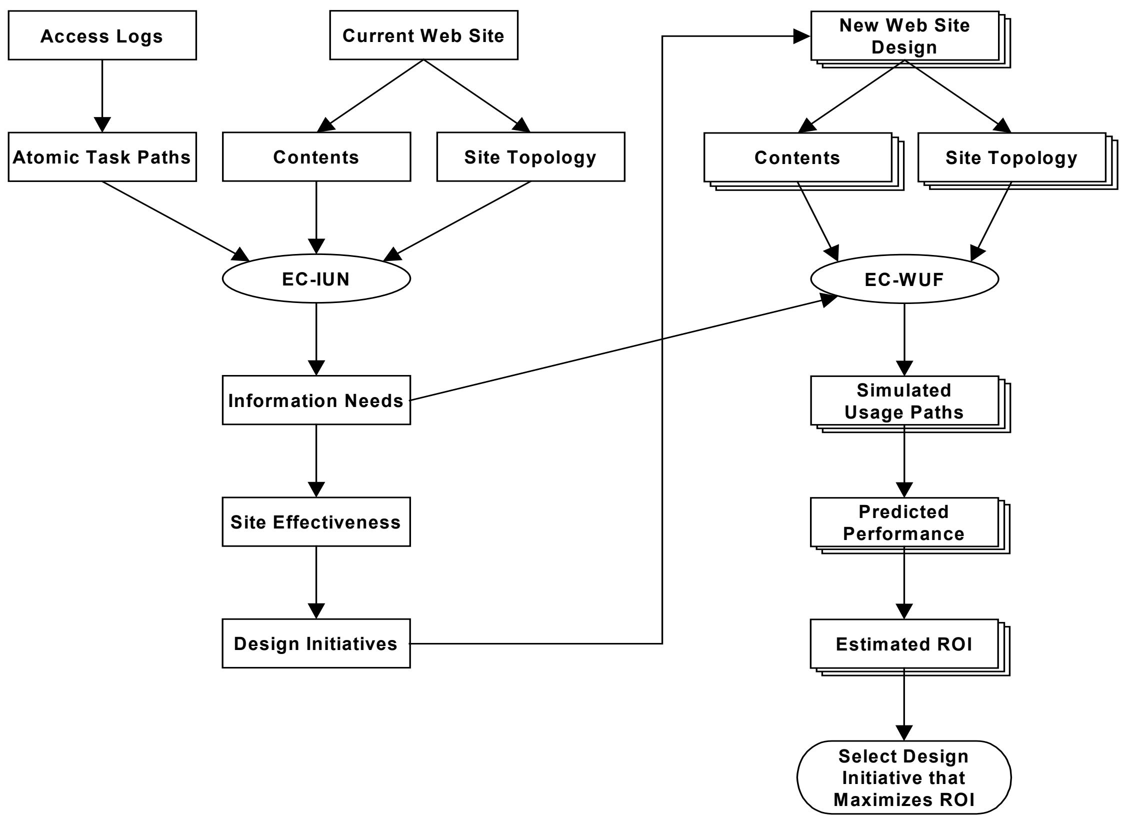

Figure 3. Systems design process

5. Conclusion

To remain competitive, e-commerce companies need to conduct online operations justified by the financial discipline of ROI. This situation brings to the foreground the importance of business value assessment of systems design activities. However, e-commerce companies still are facing difficulties due to the lack of tools and methods that link design with business performance. In this paper, we propose a value-driven systems design methodology for e-commerce. Our proposed methodology for value-driven e-commerce systems design strives to achieve a higher level of measurement rigor than traditional systems design methodologies (e.g., usability engineering [23], scenario-based design [24] and user-centered design [25]). Traditional methodologies put more focus on the process with the assumption that if you faithfully follow the prescribed procedure, then the result of the design process will be successful. In contrast, our methodology focuses on the measurement and value assessment of designs so that success is no longer assumed but managed. This is not to say that traditional methodologies are no longer useful. Our proposed value-driven design methodology should be used in conjunction with such methodologies to generate better outcomes.

A potential limitation of our proposed methodology is that the methodology may only be applicable to the dynamic redesign of an existing Web site and not to the design of a completely new one, due to the extensive data requirements needed for analyzing existing site usage patterns. However, we believe that our methodology would maintain its usefulness, even in the case of designing a new site. This is because our methodology explicitly requires designers to predict usage behaviors as well as track actual usage patterns so that design value may be assessed and managed.

We are currently conducting a multi-year longitudinal research project with OnlineGrocery.com to test the efficacy of an extended version of this value-driven framework for the performance evaluation and design of Internet-based selling Web sites. Our goal for this research stream is to develop a formal basis for a value-driven discipline for Web-based systems design and to extend Web usage mining techniques to better fit various kinds of selling environments on the World Wide Web.

0-7695-1435-9/02 $17.00 (c) 2002 IEEE

Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS-35'02)
0-7695-1435-9/02 $17.00 © 2002 IEEE

6. References

[1] Souza, R., Manning, H., Sonderegger, P., Roshan, S. and Dorsey. (2001). “Get ROI From Design” (Forrester Research Report). Cambridge, MA: Forrester Research Inc., June 2001.

[2] Rizzuti, K. and Dickinson, J. (2000). “Satisfying the Experienced On-Line Shopper: Global E-Shopping Survey” (Research Report). London: A.T. Kearney 2000.

[3] Rehman, A. (2000). “Holiday 2000 E-Commerce: Avoiding $14 Billion in ’Silent Losses‘” (Research Report). New York, NY: Creative Good, October 2000.

[4] Rajgopal, S., Venkatachalam, M. and Kotha, S. (2001). “Does the Quality of Online Customer Experience Create a Sustainable Competitive Advantage for E-commerce Firms?” (Working Paper). Seattle, WA: School of Business Administration, University of Washington, April 2001.

[5] Wallach, S. L. (2001, 7/9/2001). *Usability Improvements Payoff for Web Site*, ITworld.com. Available: www.itworld.com/nl/sup_mgr/07092001/

[6] Srivastava, J., Cooley, R., Deshpande, M. and Tan, P.-N. (2000). "Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data". *SIGKDD Explorations, 1*(2), pp. 12-23.

[7] Borges, J. and Levene, M. (1998). "Mining Association Rules in Hypertext Databases". *Proceedings of the 4th International Conference on Knowledge Discovery and Data Mining (KDD '98)*, New York, NY, August 27-31, Menlo Park: CA: AAAI Press, pp. 149-153.

[8] Heer, J. and Chi, E. H. (2001). "Identification of Web User Traffic Composition using Multi-Modal Clustering and Information Scent". *Proceedings of the 2001 SIAM Conference on Data Mining (Workshop on Web Mining) (SDM 2001)*, Chicago, IL, April 7, Philadelphia, PA: SIAM Press, pp. 51-58.

[9] Murray, D. and Durrell, K. (1999). "Inferring Demographic Attributes of Anonymus Internet Users". *Proceedings of the 5th ACM International Conference on Knowledge Discovery and Data Mining (WEBKDD '99 Workshop) (SIGKDD '99)*, San Diego, CA, August 15, New York, NY: Springer, pp. 7-20.

[10] Pitkow, J., E. and Pirolli, P. L. T. (1999). "Mining Longest Repeated Subsequences to Predict World Wide Web Surfing". *Proceedings of the 2nd USENIX Symposium on Internet Technologies and Systems (USITS '99)*, Boulder, CO, October 11–14, Los Alamitos, CA: IEEE Computer Society.

[11] Büchner, A. G., Baumgarten, M., Anand, S. S., Mulvenna, M. D. and Hughes, J. G. (1999). "Navigation Pattern Discovery from Internet Data". *Proceedings of the 5th ACM International Conference on Knowledge Discovery and Data Mining (WEBKDD '99 Workshop) (SIGKDD '99)*, San Diego, CA, August 15, New York, NY: Springer, pp. 25-30.

[12] Berry, M. J. A. and Linoff, G. (1997). *Data Mining Techniques for Marketing, Sales, and Customer Support*. New York, NY: John Wiley and Sons.

[13] Ansari, S., Kohavi, R., Mason, L. and Zheng, Z. (2000). "Integrating E-Commerce and Data Mining: Architecture and Challenges". *Proceedings of the 6th ACM International Conference on Knowledge Discovery and Data Mining (WEBKDD 2000 Workshop) (SIGKDD 2000)*, Boston, MA, August 20-23, New York, NY: ACM Press.

[14] Pirolli, P. L. T. and Card, S. K. (1999). "Information Foraging". *Psychological Review, 106*(4), pp. 643-675.

[15] Chi, E. H., Pirolli, P. L. T., Chen, K. and Pitkow, J. E. (2001). "Using Information Scent to Model User Information Needs and Actions on the Web". *Proceedings of the 2001 ACM Conference on Human Factors in Computing Systems (CHI 2001)*, Seattle, WA, March 31-April 5, ACM Press, pp. 490-497.

[16] Huberman, B. A., Pirolli, P. L. T., Pitkow, J. E. and Lukose, R. M. (1998). "Strong Regularities in World Wide Web Surfing". *Science, 280*(3), pp. 95-97.

[17] Miles, G. E., Howes, A. and Davies, A. (2000). "A Framework for Understanding Human Factors in Web-Based Electronic Commerce". *International Journal of Human-Computer Studies, 52*(1), pp. 131-163.

[18] Maes, P., Guttman, R. H. and Moukas, A. G. (1999). "Agents that Buy and Sell". *Communications of the ACM, 42*(3), pp. 81-91.

[19] O'Keefe, R. M. and McEachern, T. (1998). "Web-Based Customer Decision Support Systems". *Communications of the ACM, 41*(3), pp. 71-78.

[20] Schubert, P. and Selz, D. (1999). "Web Assessment: Measuring the Effectiveness of Electronic Commerce Sites Going Beyond Traditional Marketing Paradigms". In R. Sprague (Ed.), *Proceedings of the 32nd Hawaii International Conference on System Sciences (HICSS 32)*, Maui, HI, January 5-8, Los Alamitos, CA: IEEE Computer Society Press.

[21] Davern, M. J. and Kauffman, R. J. (2000). "Discovering Potential and Realized Value from Information Technology Investments". *Journal of Management Information Systems, 16*(4), pp. 121-143.

[22] Büchner, A. G. and Mulvenna, M. D. (1998). "Discovering Internet Marketing Intelligence through Online Analytical Web Usage Mining". *SIGMOD Record*, 27(4), pp. 54-61.

[23] Nielsen, J. (1993). *Usability Engineering*. San Francisco, CA: Morgan Kaufmann Publishers.

[24] Carroll, J. M. (1995). *Scenario-Based Design*. New York, NY: Wiley.

[25] Cooper, A. (1999). *The Inmates are Running the Asylum: Why High-Tech Products Drive Us Crazy and How to Restore the Sanity*. Indianapolis, IN: SAMS.

0-7695-1435-9/02 $17.00 (c) 2002 IEEE