E-Commerce Intelligence: Measuring, Analyzing, and Reporting on Merchandising Effectiveness of Online Stores

Stephen Gomory, Robert Hoch, Juhnyoung Lee, Mark Podlaseck, Edith Schonberg

IBM T.J. Watson Research Center
P.O. Box 704
Yorktown Heights, NY 10598
email: {gomory, rhoch, jyl, podlasec, ediths}@us.ibm.com

Abstract

This paper contributes to several areas of Web usage analysis for e-commerce. First, we interviewed a number of real-world online retailers for collecting and categorizing usage analysis requirements for e-commerce Web sites. We present the analysis requirements as a set of business questions and metrics for Web traffic and sales. Second, we introduce a novel way, referred to as *micro-conversion rate*, of tracking and analyzing the effectiveness of various business efforts in online stores such as marketing, merchandising, product assortment, and site design. Third, we describe a data management architecture for e-commerce data analysis that incorporates these new metrics. The architecture consists of a number of components including a Web server instrumentation module that dynamically gathers relevant data from Web pages, a data model that combines Web traffic and sales data, and an OLAP (On-line Analytical Processing) system that provides exploratory analysis and reporting capabilities.

1. Introduction

In just a few years, Web sites have moved from displaying electronic brochures to providing a channel for sales, customer service, and information gathering for companies large and small. More and more business processes have moved to the Web, and the Internet is turning into a virtual marketplace. As this trend continues, return on investment is becoming the major concern of Web merchants. Because rapid surfing from site to site is the norm, it is increasingly critical for Web merchants to understand the effectiveness of their Web sites, and to gather, analyze, and act on Web usage data for competitive advantage. The Web channel provides new opportunities and challenges for analysis: on the one hand, it is possible to collect large amounts of detailed information on every user action. On the other hand, the problem of extracting useful information that is meaningful across relevant business areas in an organization is not well understood. There is little experience and knowledge of the end-to-end process from identifying what usage analysis is interesting to the organization, what needs to be tracked and measured, to acting on the analysis for revising Web content, updating advertising and promotion strategies.

Web traffic analysis through mining usage data recorded in Web server logs is an actively studied area [5, 7, 13, 19], and there are a number of commercial Web traffic analysis tools available [1, 3, 8, 9, 11, 14, 15, 16, 17, 18] which extract data from Web logs. Web traffic analysis tools typically keep count of hits, rank the most popular pages requested, and tell where the user came from, the length of time each page was viewed and the page from which the user exited the site. Beyond basic metrics, session and other traversal path data can be derived from Web logs. A limitation of general usage analysis based on Web logs in the e-commerce arena is that it does not give sufficient insight into sales and business results.

Among the important e-commerce usage metrics are the banner ad *click-through rate*, which is the percentage of viewers who click on a banner ad, and the *conversion rate*, which is the percentage of visitors who purchase from the store. More recently, banner ad return on investment (ROI) has become the significant metric for effective marketing. None of these metrics are reliably computable from Web logs alone. Measuring the click-through requires obtaining the banner ad *impression* count, which must be obtained from external sources. Measuring ROI requires knowledge of sales and products, in addition to traffic.

This paper presents novel concepts for e-commerce data analysis, and shows how they can be incorporated into a data management system architecture, referred to as *E-Commerce Intelligence* (ECI). The architecture includes a Web server instrumentation module that dynamically gathers relevant data from Web pages, a data model that combines Web traffic and sales data from diverse sources, and an OLAP (On-line Analytical Processing) engine that provides for exploratory analysis and reporting. The highlights of our contribution include:

- **Integrated analysis of traffic and sales**

We interviewed a number of real-world online retailers for collecting and categorizing usage analysis requirements for e-commerce Web sites. We present the analysis requirements as a set of business questions and metrics for Web traffic and sales. Our multi-dimensional data model is designed to support the analysis for these questions.

- **Analysis of merchandising effectiveness**

We focus on the merchandising effectiveness of an online store, an area where there are few existing analysis tools. Specifically, we view an e-commerce site as a collection of banner ads for individual products. From this perspective, we define metrics for the effectiveness of various product merchandising implementations which are similar to banner ad metrics, and measure conversion rates and ROI for internal Web site features.

- **Semantic view of commerce site**

It is important to present reports in a language that business users are familiar with and can act on. Therefore, we describe Web pages and hyperlinks in a site using semantic names instead of URLs according to their merchandising purpose. Semantic information can include a product label, a cross-sell or promotion label, or a tag indicating where the product is being displayed. To provide such semantic labels for Web pages and hyperlinks in the site, we use a site dictionary which stores meta-data about the site structure and operation, and data about categories of Web pages and hyperlinks.

- **Micro-conversion rate**

We introduce a novel way, referred to as the *micro-conversion rate*, of tracking and analyzing the effectiveness of various business efforts in online stores such as marketing, merchandising, product assortment, and site design. These areas of business efforts coincide with the areas of usage analysis required by online retailers in our interviews.

- **Tracking dynamic Web page content**

The standard Web server logs report Web usage by URLs which indicate only the location of served Web pages and often very little about the content. ECI uses enhanced Web server logging as a way to collect extra data about served Web pages, most importantly, information about product impressions and design features of hyperlinks which is necessary for computing the micro-conversion rates. Enhanced logging is performed by a Web server plug-in, which dynamically parses HTML source code of every Web page served in order to collect the useful content of the page. The ability to dynamically scan Web pages as they are served is critical for tracking Web usage, because more and more Web pages are dynamically created from databases and contain personalized and adaptive content. A simple example of a dynamically created Web page is a search result page commonly found in online stores. For a search result page, the ECI plug-in logs the dynamically selected set of product links. Subsequent analysis can compare these product impressions with sales in order to evaluate the effectiveness of the search engine.

The rest of this paper is structured as follows: Section 2 presents a high level set of usage analysis areas and related business questions. Section 3 defines a new set of metrics for merchandising, referred to as *micro-conversion rate*. The system architecture for data management is described in Section 4, and an illustration of using the multi-dimensional data model to answer business questions is presented in

Section 5. In Section 6, related work is evaluated and compared against our approach. Finally, in Section 8, conclusions are drawn and further work is outlined.

2. Areas of Web Usage Analysis

For collecting and categorizing Web usage analysis requirements, we interviewed a number of business users in several real-world online stores. In order to cover different analysis objectives and business questions from different business areas, we interviewed a mix of professionals, including those from marketing, advertising, merchandising, and Web design and development. In this section, the results from the interviews are presented in the form of sample business questions categorized by area of usage analysis. Answering these questions requires measuring both Web traffic and sales. From the interviews we have identified nine different areas which are of interest to different business users. Table 1 summarizes the areas of analysis with the sample business questions for each area.

Areas of Analysis	Business Questions
overall store performance	What is the sales value for a specific period of time, say, week? What is the number of customer visits for the day? What is the store conversion rate for the week? What is the sales value index for the week?
advertising	Which banner ads are pulling in the most traffic? How many sales are driven by each banner ad? What products do shoppers from a particular banner ad purchase? What is the conversion rate for each banner ad?
external referrals	Which portal sites are pulling in the most traffic? Which are generating the most sales? How many sales are generated by each referral site/search engine? What products do shoppers from a particular portal site purchase?
shopper segmentation	How many visitors are from a specific domain? What is the distribution of first-time vs. repeat shoppers? What characterizes the shoppers of a particular set of products? What characterizes the shoppers who abandon shopping baskets?
product grouping	How much do cross-sells/up-sells contribute to gross revenue? What are the best performing cross-sell pairs? Worst? What is the overall conversion rate for cross-sells/up-sells?
promotions and recommendations	How much do promotions contribute to gross revenue? Which promotions are generating the most sales? What is the overall conversion rate for promotions? What is the overall conversion rate for recommendations? Which levels in site hierarchy are the best promotions located at?
shopping metaphor	What generates the most sales value: search or browsing? How much does search contribute to gross revenue? What is the conversion rate for search?
design features	What are the features of links customers most frequently click? What are the features of links customers most frequently buy from? What are the parts of page customers most frequently buy from? Do products sell better in the upper left corner?
product assortment	What are the top sellers for the week? What is the conversion rate for a particular department? How is a product purchased: purchase frequency and quantity? What characterizes the products that end up being abandoned? How much of the sales of each product are driven by search?

Table 1: Areas of usage analysis and sample business questions for online stores

The first area in Table 1, the overall store performance, is directed towards upper management, as well as to all business users. The next three areas of analysis, advertising, external referrals and shopper segmentation, are of interest to marketers of online stores. Marketing is broadly defined as activities used to acquire customers to the online store and retain them. Shopper segmentation is important in marketing because it enables the store to better understand its shoppers and their needs. Clustering can use different sets of variables, such as demographic characteristics and shopping behavior, that are selected by the characteristics of the business problem to be solved. In the Web, a few new clustering variables, such as the domain name where customers are coming from, the external portal sites or banner ads that customers are referred by, and shopping basket behavior, may be interesting.

The rest of the Web usage analysis areas in Table 1, product grouping, promotions and recommendations, Web design features, shopping metaphor, and product assortment are of primary interest to business users who implement merchandising and Web design strategies for the site. Merchandising includes activities used to present merchandise within the online store.

Merchandising cues in an online store are the different ways Web merchants present and/or group their products to motivate purchase. Examples of merchandising cues include cross-sells, up-sells, promotions, and recommendations[1]. Shopping metaphors in an online store are different ways that shoppers can use to find products of interest. Examples include browsing, search, possibly in various forms, and configuration for "build-to-order" products. In online stores, both merchandising cues and shopping metaphors are associated with hyperlinks on Web pages. This fact means that it is possible to categorize and group together hyperlinks in an online store by their types of merchandising cue and shopping metaphor. Other merchandising aspects that can be used to categorize hyperlinks are their design features such as media type (e.g., image or text), font (if text), size, color, location and so on.

Just as online marketing uses banner ads and/or referral sites to attract customers from external sites to an online store, online merchandising uses hyperlinks and image links within the store to lead customers to click to Web pages selling products. Web merchants employ a diversity of merchandising schemes associated with hyperlinks. From this perspective, the problem of tracking and measuring the effectiveness of different merchandising strategies in an online store turns into three sub-problems: (1) classifying each hyperlink by its merchandising purposes, (2) tracking and measuring traffic on the hyperlinks and analyzing their effectiveness (e.g., profitability), and (3) attributing the profit of the hyperlinks to their merchandising cue type, shopping metaphor type, and design features. The analysis of the effectiveness of marketing strategies may be conducted in a similar way. The only difference is that the originating hyperlinks in marketing efforts are presented and controlled in external sites. This online store effectiveness analysis based on hyperlink effectiveness analysis is a core idea of this paper. How we measure this effectiveness is described in more detail in the next section.

3. Micro-Conversion Rate

The conversion rate of an online store is the percentage of visitors who purchase from the store. While this measure is useful for evaluating the overall effectiveness of the store, it does not help understand the possible factors affecting the sales performance. The notion of micro-conversion rate extends the traditional measure by considering the four general shopping steps in an online store, which are:

1. product impression: the view of hyperlink to a Web page presenting a product,
2. click-through: a click on the hyperlink and view the Web page of the product,
3. basket insertion: placement of the item in the shopping basket if interested, and
4. purchase: the purchase of the item - completion of a transaction.

[1] In online stores, the merchandising schemes are presented in forms of hyperlinks on Web pages. A cross-sell link refers the visitor to a Web page marketing an item complementary in function to the item marketed on the current Web page. An up-sell link refers the visitor to a Web page presenting a similar but more upscale item. A recommendation link highlights product pages that are likely to be of interest to the shopper based on knowledge of the shopper and the behavior of a larger population. A promotion link refers a visitor to a product page from a "What's Hot" page or a high traffic area such as the "Home" page for informing, persuading and/or reminding the shoppers about a product and/or other aspects of the site.

Micro-conversion rates are computed for each adjacent pair of measures resulting in the following three rates:

1. *look-to-click rate*: how many product impressions are converted to click-throughs
2. *click-to-basket rate*: how many click-throughs are converted to basket placement
3. *basket-to-buy rate*: how many basket placements are converted to purchases

Note that the first one, look-to-click rate, is similar to the click-through rate used for measuring the amount of traffic on banner ads, and also that the micro-conversion rates relate the traffic-related measure to sales that happen later in the shopping process. By tracking precisely the shopping steps in this way, it is possible to spot exactly where the store loses the most customers and to figure out what might cause the loss. In addition to the three above, another useful metric is *look-to-buy rate* which considers the first and last shopping steps and accounts for what percent of product impressions are eventually converted to purchases. By looking at this rate, online merchants can tell if a product is over-exposed or under-exposed and take action to change the presentation of the product.

The notion of micro-conversion rate also extends the traditional measure by considering the marketing or merchandising purposes associated with hyperlinks viewed in the first shopping step described above. In this way, the micro-conversion rate is related to strategies for marketing and merchandising, and can be used for evaluating the effectiveness of different merchandising aspects of the store. Unlike the traditional conversion rate that gives just one number for the entire site, the micro-conversion rates can be calculated for individual products, individual shopping metaphor, individual merchandising cue types, individual design features, and individual banner ads, i.e., basically all the individual hyperlinks pointing to product pages in various forms and purposes, internal or external to the site. Table 2 summarizes different types of micro-conversion rate by the merchandising or marketing aspects associated with the hyperlink.

Note that the notion of micro-conversion rate is based on the availability of product impression data. In the next section, we will explain how the data on product impressions through hyperlinks can be collected by extended server logging. Also note that it is possible to derive a variation of the look-to-buy rate that is useful for evaluating the effectiveness of a hyperlinks by measuring the amount of sales value the hyperlink is driving. This measure, which we call *hyperlink effectiveness* is similar to the ad banner ROI. While the ad banner ROI metric focuses on marketing-related questions, this metric is used to evaluate the effectiveness of merchandising strategies within a store such as merchandising schemes and shopping metaphor.

Type	Description
products	Look-to-click, click-to-basket, basket-to-buy and/or look-to-buy rates can be computed for individual products in an online store to measure the product performance in the site. The rates for individual products can be rolled up to give the micro-conversion rates for categories of the products and then all the way up to the entire site, if an appropriate taxonomy of products is given.
merchandising cues	The four types of micro-conversion rates and the hyperlink effectiveness can be calculated for hyperlinks having merchandising schemes such as cross-sell, up-sell, promotion, or recommendation to show how effectiveness those schemes work in an online store.
shopping metaphor	The four types of micro-conversion rates and the hyperlink effectiveness can be computed for hyperlinks which, for example, appear on search result pages to measure the effectiveness of the search engine in a site.
design features	Hyperlinks to Web pages presenting products can be classified by their design features such as media type (e.g., text or image), color, font (if text), size, location and so on. The four types of micro-conversion rates and the hyperlink effectiveness can be computed for the hyperlinks to show the effectiveness of the different design features in a store.

Table 2: Reports on micro-conversion rate

Figure 1 illustrates sample micro-conversion rates for three different merchandising cues, i.e., promotions, recommendations, and cross-sells. Visitors are seeing twice as many impressions for promotions (40,000) than cross-sell impressions (20,000), and twice as many cross-sell impressions than impressions for recommendations (10,000). However, the look-to-click rate for recommendations (18%) is twice as high as for either promotions or cross-sells. Additionally, recommendations are resulting in a relatively high look-to-buy rate (2%). This means that the recommendation engine is relatively effective at personalization. On the other hand, this example shows that promotions are not working effectively. Of the visitors who place a promoted item in a shopping basket, 30% of them purchase the product. However, the click-through rate for promotions is 10% and the click-to-basket rate is only 2.5%, so the look-to-buy rate is 0.075%, which shows poor overall performance, and an over-exposure of the promoted items. Finally, the look-to-buy rate for cross-sells in this example is about 0.5%.

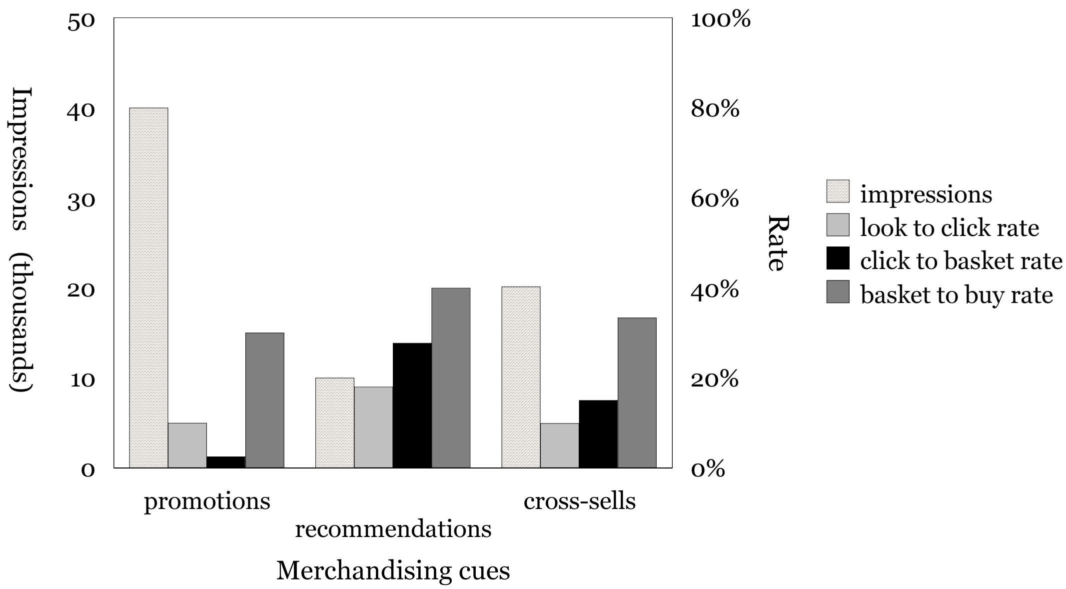


Figure 1: Micro-conversion example

4. ECI Data Management System

The architectural goal of the ECI solution is to provide an analysis environment that is rich in data expressed in terms that are comfortable for the business analyst to answer the business questions from Section 2. The data employed to achieve this goal comes from Web servers, commerce servers, and the merchant's enterprise data systems. Combining these disparate data sources into a single analysis environment puts electronic commerce activities into their business context. In order to provide this environment, three distinct components are necessary. The first is an enhancement to the Web server log which records detailed information about the user view of the Web site. The second is a set of descriptive meta data the ties the Web site design to business terms and product information. The third is the data transformations necessary to tie this information together with additional merchant data in order to produce our multi-dimensional data model for analysis. These components are described below, and followed by an overview of the data model. The ECI system architecture is illustrated in Figure 2.

4.1. Plug-in and Semantic Log

Since existing Web server logs do not contain enough information for reporting on Web page content, we have implemented a plug-in to run in the commerce server environment. The plug-in parses the content of each Web page as it is served and generates an enhanced Web server log that contains hyperlink information. More specifically, the plug-in uses a configuration file generated from the site dictionary which allows it to find each product link on a page. In addition, the plug-in classifies the link by extracting relevant parameters such as product identifier or merchandising cue type. This information is combined with referral information and the URL into a single log record. Lastly, cookies are used in order to group the log records into individual sessions. The result is a log file, referred to as *semantic log*, that contains detailed information about the customer's interaction with the commerce Web site.

4.2. Site Dictionary

Semantic labeling is accomplished via Site Dictionary which maps URLs to merchandising labels. For example, there can be a table in Site Dictionary that contains information for mapping URLs in the site to product identifiers, which are the numbers the online merchant use to catalog products in the store. There can be another table for mapping URLs to various types determined by their merchandising purposes, i.e., merchandising cues such as cross-sells and promotions, and shopping metaphor such as search. Also, there can be a table for mapping URLs to their design features, i.e., media types, color, size, location and so on. The information in Site Dictionary is used when the Web log is processed and during data cleansing, to build the database.

The information stored in Site Dictionary is site-specific: URL patterns, Web page types and layout, and site structure vary from one site to another in many ways. We find however, that most large scale commerce sites rely on templates for page design and navigation. As a result, we can classify URL patterns and parameters and map them to product displays, shopping metaphors, and merchandising cues. Since online stores will change frequently, it is important to automate this process of site meta-data collection and provide for versioning in order to maintain historical comparisons.

4.3. Merchant Data Sources

Additional inputs to the ECI data architecture include information from the merchant about product hierarchies. This gives the analyst the ability to view product related facts either within a roll-up of product categories or across products by attributes or supplier. This product information is extensible without altering the overall model. The merchant may also bring in information about cost of goods, external product promotions, direct mail campaigns or customer profile information.

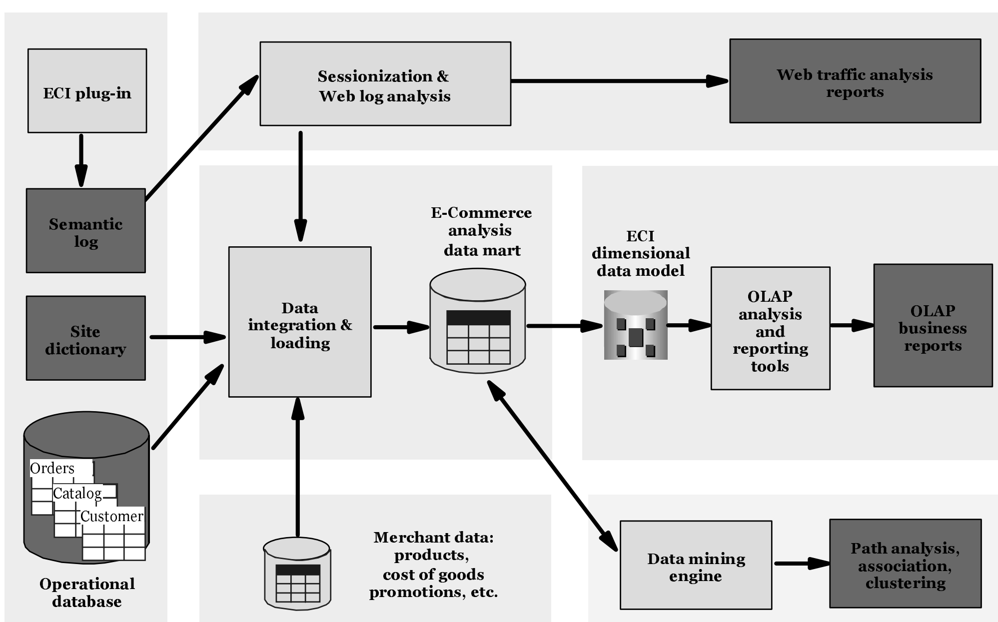


Figure 2: ECI system architecture

4.4. Data Cleansing

Since the Web server activity has no inherent persistence, server log files need to be processed and then stored into a staging area prior to merging with merchant business data and commerce data. The purpose of the staging area is to allow for data to be cleansed, normalized and then loaded into a dimensional model for use by analysts. Here, individual versions of the site dictionary can be tied to the current Web site design, current product hierarchies can be used and the relationship between external marketing campaigns can be tied to the Web site. Note that in this design, additional derived dimensions such as customer segmentation can be added or developed over time as patterns emerge from the accumulation of data.

4.5. Data Model

The end goal of the data preparation step is to produce a data mart consisting of data about Web traffic, product impressions and product purchases. In this environment, the business analyst can answer complex and useful questions like those outlined in section 2. The warehouse is implemented as three related star schemata providing OLAP capabilities. Figure 3 represents the three main areas of analysis and demonstrates the areas of overlap. Each oval contains the area name, its basic facts and the dimensions available. This provides a convenient way to diagram the common dimensions across these areas of analysis. The individual areas of analysis are illustrated and described in more detail below.

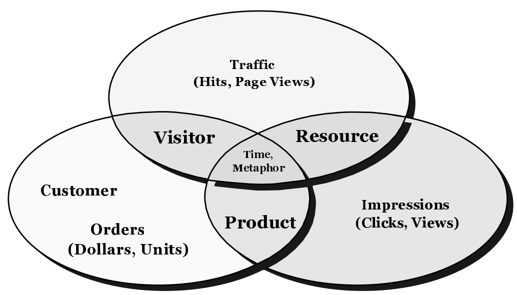


Figure 3: Multi-dimensional data model

Traffic analysis is well defined, and forms the basis of existing analysis and reporting tools. Traffic analysis data comes directly out of server log information via a session building tool. In this context it is interesting because it provides the half of the data necessary to compute metrics like conversion rate. In addition, it provides enough information about the origin of the visitor to be useful for measuring effectiveness of web-based ad campaigns.

The orders dimension has at its center the basic facts, item price and units ordered, about customer purchases through the Web site. In addition, through Web log analysis, using information provided by the semantic labeling, it is possible to associate a dimension describing the product impression associated with the sale (creative) and to identify the path that brought the user to the sale (shopping metaphor). The traffic analysis also contributes information about the "Web-demographics" of the customer, in other words, the sub-domain, platform, browser, referral-source for this session. Lastly, using enterprise data from the merchant, the product identifier can be tied to a product taxonomy as well as other product characteristics like color, size and brand.

The third area of analysis is the product impressions, which are recorded in the server log by the ECI plug-in. This includes a record of how frequently a product was seen and/or clicked on by a visitor. The impression is characterized by its placement within the site, its role as a merchandising cue, and the type of visual presentation on a page. For example, the product view may occur on a promotion page as a featured element with a large graphic. Knowing the number of visitors that see, and consequently click on this graphic gives an indication of its effectiveness and is the first step in calculating the micro-conversion rate. To illustrate one of the star schemata in our model, the orders dimension is represented in Figure 4.

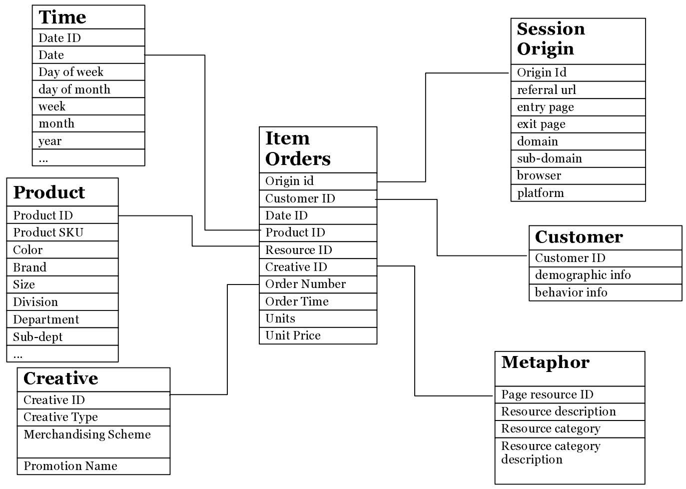
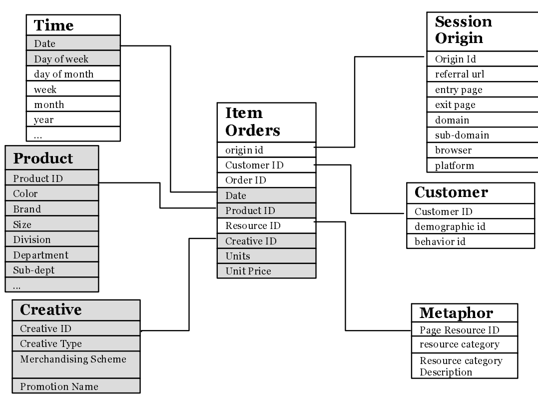


Figure 4: Orders dimension model

5. Sample Analysis

For the purpose of illustration, we walk through a business analysis by focusing on the orders data model presented above. Beginning with a variation on a question from Section 2 above, "What design features drive the sales of the five most popular products over the past week," we can define a strategy for using the ECI data mart. Note that the operations described here can be done with commercially available OLAP tools. The first step would be identifying the "five most popular products". This could be defined as those five products with the most unit sales. The second half of the question: "What design features drive the sales..." revolves around the merchandising cues used to present the product. These may be sketches, images, or textual descriptions. As with any analysis, different breakdowns of the question at hand would result in different queries against the data model. The results of this example are shown in Figure 5 with individual constraints grayed.



Sales $	**Photo**	**Sketch**	**Text**
Wine Glasses	**80%**	**5%**	**15%**
Waffle Iron	**15%**		**85%**
Knee Pads			**100%**
Dish Towels	**100%**		
Teak Bench		**100%**	
Cutlery	**20%**		**80%**

Figure 5: Sample business analysis result

6. Related Work

There are several commercial Web traffic analysis tools available [1, 3, 9, 11, 14, 16]. Most of these Web traffic analysis tools generate reports mainly on Web traffic and system measurements. It is difficult, though not impossible, to report on the effectiveness of specific marketing and merchandising efforts because those tools primarily rely on information in Web server logs which hardly include data useful for measuring the business efforts. They typically report their findings on traffic in technical terms such as URLs which convey little meaning to the business analyst. Most of these traffic analysis tools are not adequate for answering the business questions in Table 1. There are several commercial services and software tools for evaluating the effectiveness of Web advertising such as banner ads, in terms of traffic and sales driven by them [8, 15, 17, 18]. They use metrics such as click-through rates and ad banner ROI. The objective of these tools is different from our work; they focus on understanding the effectiveness of advertising, while ECI focuses on the effectiveness of merchandising. The techniques used for Web advertising tracking tools are not directly applicable to tracking and measuring the merchandising effectiveness within an online store.

Recently, Web traffic analysis by applying data mining techniques to Web server logs has been actively studied [5, 7, 13, 19]. Some of the data mining algorithms that are commonly used in the studies are association rule generation, sequential pattern generation, and clustering. Association rule mining techniques [2] discover unordered correlations between items found in a database transactions. In the context of Web usage mining, a transaction is a group of Web page accesses and an item is a single page access. Sequential pattern generation [13] discovers inter-transaction patterns such that the presence of a set of items is followed by another item in the time stamp ordered transaction set. By analyzing this information, a Web usage mining system can determine temporal relationships among data items. Clustering analysis [12] allows one to group together users or data items that have similar characteristics. Clustering of user information or data from Web server logs can facilitate the development and execution of marketing strategies. While ECI and these Web mining techniques share similar objectives of finding interesting aspects of Web usage of an online store which are potentially useful for improving marketing and merchandising strategies on the site, they address different types of business questions and may be used in a complementary way.

There has been some work done on loading Web usage data into a data cube structure [10] in order to perform both data mining and traditional OLAP activities such as roll-up and drill-down of the data [4, 6, 20]. These studies formally defined generic data hypercubes for Web usage data and provided designs for exploratory analysis and reporting. Also they showed how data mining techniques are used on the data model in electronic commerce scenarios. Büchner et. al. [4] described a way of combining various types of data, such as Web server logs, site meta-data, and marketing data, for discovering actionable marketing strategies in an electronic commerce environment. Zaiane et. al. [20] proposed a way of using meta-data about Web site design, specifically the meaning behind CGI (Common Gateway Interface) scripts and parameters to aid in Web log analysis. This approach provides a high-level way of categorizing Web pages that is similar to annotating Web pages with semantic labels in ECI. However, this work focuses on general Web usage analysis, while ECI concentrates on the analysis of merchandising effectiveness. Also, this work addresses different types of questions that are more appropriately answered by using data mining techniques.

7. Concluding Remarks

This paper has presented a collection of Web usage analysis requirements in the areas of marketing and merchandising in the form of business questions and metrics for Web traffic and sales. This paper also presented the details of tracking and analyzing the effectiveness of merchandising in online stores by using the notion of micro-conversion rate. We are currently in the process of implementing the system architecture for Web usage data management described in this paper by using a real-world online store site duplicated in the laboratory. We plan to experiment with industrial data extracted from the online store to illustrate how different merchandising efforts are tracked and measured. The area of understanding the effectiveness of business efforts in online stores is relatively new. There has not been much work done on the subject, although it poses an important and imminent challenge for electronic commerce across almost every industry.

Our work presented can be extended and varied in many ways to address different questions on both business and system effectiveness of online stores. Some of the possibilities of extending the work presented in this paper for future study are as follows: First, the idea of classifying hyperlinks and labeling them with semantic vocabularies can be generalized and applied to other types of business questions in other areas such as marketing and operation. One example is clustering customers by their shopping behavior such as types of hyperlinks they click. In this case, hyperlinks in an online store need to be categorized and labeled to distinguish characteristics of shoppers' behavior. Second, the approach of OLAP-based exploratory analysis used in this work may be combined with a different, but complementary approach, i.e., data mining, that can find useful information on navigation paths, association, and clustering, which cannot be directly obtained by a pure OLAP approach. Finally, the metrics for merchandising effectiveness presented in this work need to be adjusted and extended for the use in new shopping paradigms in the Internet such as online auction.

References

1. Accrue Software Inc's Insight, http://www.accrue.com/.
2. R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules," *Proc. of the 20th VLDB Conference*, Santiago, Chile, 1994, pp. 487-499.
3. Andromedia Inc's Aria, http://www.andromedia.com/.
4. A. G. Büchner and M. Mulvenna, "Discovering Internet Marketing Intelligence through Online Analytical Web Usage Mining," *SIGMOD Record*, 27(4):54-61, December 1998.
5. R. Cooley, B. Mobasher, and J. Srivastava, "Web Mining: Information and Pattern Discovery on the World Wide Web," *Proc. of the 9th IEEE International Conference on Tools with Artificial Intelligence*, 1997.
6. R. Cooley, B. Mobasher, and J. Srivastava, "Data Preparation for Mining World Wide Web Browsing Patterns," *Journal of Knowledge and Information Systems*, 1(1), 1999.
7. M. S. Chen, J. S. Park, P. S. Yu, "Data Mining for Traversal Patterns in a Web Environment," *Proc. of the 16th International Conference on Distributed Computing Systems*, 1996.
8. DoubleClick Inc, http://www.doubleclick.com/.
9. Engage Technologies Inc's I/Pro, http://www.engagetechnologies.com/.
10. J. Gray, A. Bosworth, A. Layman, and H. Pirahesh, "Data Cube: a Relational Aggregation Operator Generalizing Group-By, Cross-Tab, and Sub-Totals," *Proc. of IEEE 12th International Conference on Data Engineering*, 1996, pp. 152-159.
11. IBM Corp's SurfAid, http://surfaid.dfw.ibm.com/.
12. L. Kaufman, and P. J. Rousseeuw, *Finding Groups in Data: an Introduction to Cluster Analysis*, John Wiley & Sons, 1990.
13. H. Mannila, and H. Toivonen Portland, "Discovering Generalized Episodes Using Minimal Occurrences," *Proc. of the 2nd International Conference on Knowledge Discovery and Data Mining*, Portland, Oregon, 1996, pp. 146-151.
14. Marketwave Corp's Hit List, http://www.marketwave.com/.
15. Media Metrix, http://www.mediametrix.com/.
16. net.Genesis' net.Analysis Pro, http://www.netgenesis.com/.
17. NetRating Inc, http://www.netratings.com/.
18. Straight UP!, http://www.straightup.com/.
19. T. Yan, M. Jacobsen, H. Garcia-Molina, U. Dayal, "From User Access Patterns to Dynamic Hypertext Linking," *Proc. of the 5th International Conference on the World Wide Web*, 1996.
20. O. R. Zaiane, M. Xin, J. Han, "Discovering Web Access Patterns and Trends by Applying OLAP and Data Mining Technology on Web Logs," *Proc. of Advances in Digital Libraries Conference*, 1998.