A General E-Commerce Data Model For Strategic Advantage: Mapping Site Structure To Site Visit Behavior

David H. Olsen, (E-mail: dolsen@b202.usu.edu),Utah State University
Nicole Forsgren, (Email: nicolef@cc.usu.edu), Utah State University

Abstract

*This paper includes a general data model that supports e-commerce activity. Numerous data models have been presented which support traditional business activities, but none has yet addressed the specific needs of e-commerce systems. An e-commerce system has unique opportunities to capture information about the entire online experience of each customer. Our **g**eneral **e**-commerce data **m**odel (GEM) captures information such as the length of time a customer spends on a given web page, pictures that are examined or are "clicked on", and the characteristics of visits which lead to customer purchases. GEM will be extremely useful to many e-commerce personnel including marketing strategists as they will be able to determine which parts of a web page visit lead to successful and unsuccessful visits. Additionally, GEM is designed to facilitate business intelligence (BI), which concerns the area of electronically monitoring competitors' web presence. GEM can be used to monitor those accessing a given website and understanding their site behavior. GEM is presented using the CASE dialect of the entity-relationship (E/R) modeling paradigm.*

I. Introduction

Data modeling is one of the most critical tasks in building an information system. A poorly developed data model will never lead to an implementation that yields accurate, timely information no matter the talent level of the implementation specialists. It seems that though there are relatively numerous database implementation specialists, there are few that are accomplished data modelers and even fewer that understand the needs of e-commerce systems.

In this paper we present a **g**eneral **e**-commerce data **m**odel (GEM) using the entity/relationship (E/R) data modeling method (Chen, 1976). While some believe that e-commerce systems are merely online extensions of traditional business systems and thus do not require any new data models, the nature of buying and selling online presents an organization with the opportunity of following customers through the selection and purchasing process. GEM then is designed to capture information about the entire customer experience including things such as time spent on a web page, goods that were viewed, and whether or not they were not purchased. Additionally, GEM is designed to facilitate and combat business intelligence (BI), which concerns organizations monitoring one another's web presence (Sullivan, 2001). GEM can be used track the site behavior of any visitor, including automated agents.

Readers with comments or questions are encouraged to contact the authors via email.

In the end, the purpose of a data model is to yield a robust database that is properly constructed, eliminates anomalies, and contains business logic. This paper continues with a literature review of data modeling, universal data model concepts, and e-commerce in Section II while Section III reviews the data modeling conventions used in this paper. Section IV includes a description of GEM for e-commerce applications using the conventions described in Section III. Section V concludes the paper and provides direction for future research.

II. Literature Review

Data modeling is the act of capturing the abstract world of interest within a predefined language or methodology. The present state of data modeling practice has its roots in the entity/relationship (E/R) model (Chen, 1976). Some alternative current data modeling approaches include the E/R model approach, semantic modeling approaches, the accounting REA model, and UML. The E/R model is presented in (Chen, 1976) and the REA model, which is a specialized accounting data modeling paradigm, is presented in (McCarthy, 1982).

The importance of generalized data models is becoming increasingly apparent as more and more companies are having to redesign their information support systems. With an abstract and generalized set of tables as the foundation for a database, organizational and definitional aspects of the business can be changed while only affecting the content of the tables without impacting the structure of the tables (Hay, 1995). These data models can save significant amounts of time in the system design process when used as a starting point and modified to meet specific business needs. When used for this purpose, modifications are expected and encouraged but should be carefully analyzed by the systems development team to ensure that changes will not result in denormalization of any existing structures (Silverston, 1997).

Temporal data is essential to the analysis of a website visit. The three fundamental temporal data types are: 1) instant, which is defined as something that happens at an instant in time; 2) interval, or a length of time; and 3) period, which is an anchored period of time (Snodgrass, 2000). All three types are addressed in GEM and will provide strategists and analysts with the ability to find average time spent on a page, average time spent on the site before a successful sale, and other useful time-centered queries.

The movement to online retail business presents an organization with many different opportunities and challenges, with the most important being the design and structure of a website which leads a customer to a purchase. Website design researched by (Jennings, 2000) outlines the importance of: 1) aesthetic experience, 2) flow, 3) habitat selection, and 4) an aesthetic framework. With a data model that supports the tracking of site visit behavior, unsuccessful visits (as defined by the organization) will flag web pages and their contents that may be analyzed to discover problems with aesthetic experience, flow, and aesthetic framework, which are listed above as items one, two, and four.

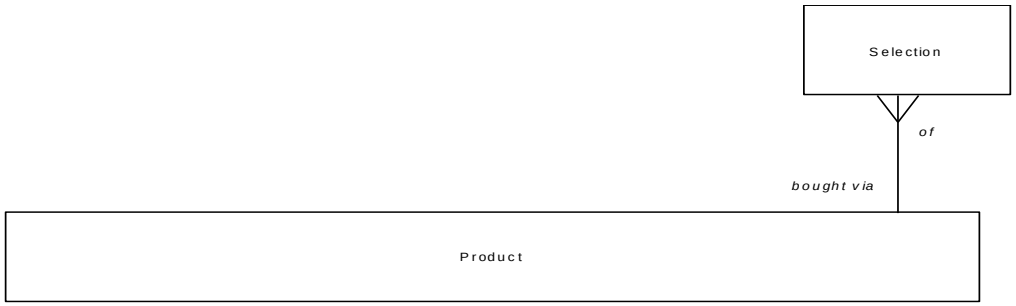
Another important issue in recording web site traffic is the surge of dynamically-generated page content. With more and more pages being composed by query strings and profile criteria, static HTML pages cannot be assumed to be the norm. With website structural complexity growing exponentially, the ability to track user activity must keep pace (Richebacher, 2001). GEM supports the tracking and analyzing of dynamically-generated web page contents.

III. Modeling Conventions

This paper uses the CASE*Method™ notation, which is used throughout the Oracle corporation. The syntax for our model utilizes entities, supertypes and subtypes, attributes, and relationships.

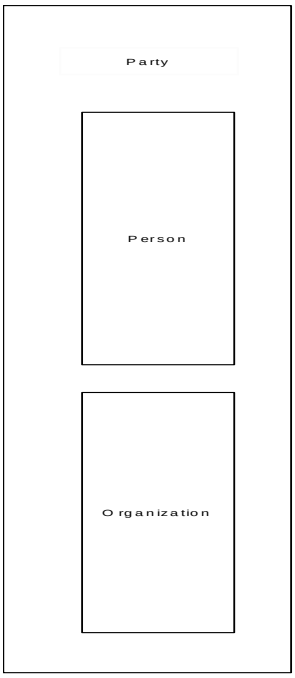
An entity is a thing of significance that an organization wishes to maintain and manipulate information about. Figure 1 shows the entities PRODUCT and SELECTION from GEM. Entities are represented by rectangles in our data model and are written in the singular. Entity names will be written in small upper case letters throughout this paper, as noted above with PRODUCT.

Figure 1



Supertypes and subtypes are kinds of entities that allow for object-oriented notation within the E/R model and appear as rectangles (entities) contained within another rectangle (entity). A supertype is a parent class while a subtype is an occurrence of a supertype that inherits all attributes of its parent class while further distinguishing itself with some of its own. An example from our model appears as the superclass PARTY with its subclasses being PERSON and ORGANIZATION as shown in Figure 2. PERSON and ORGANIZATION both contain attributes common to PARTY, such as ID and Address (through the PARTYADDRESS table), while containing some unique attributes such as FirstName (in the case of PERSON) and Contact Name (in the case of ORGANIZATION). In this case, PERSON and ORGANIZATION are mutually exclusive; PARTY can be either a PERSON or an ORGANIZATION, but each PERSON must be a PARTY and each ORGANIZATION must be a PARTY.

Figure 2



Attributes are items used to describe an entity, with unique combinations of attributes (usually expressed as a primary key) being used to describe unique occurrences of an entity. As an example, attributes of PRODUCT may include ProductID, Price, and Description. Our data model diagram does not include a listing of attributes, but possible attributes are used in the description of the data model in Section IV and are listed in Appendix A.

Relationships between entities are shown with solid or dashed lines and crow's feet. Solid lines are used to show that the relationship is required, where a dashed line shows optionality. In our model, only solid lines were

used, but optionality can be assumed throughout, depending on documented business rules. A crow's foot is used to describe the cardinality of a relationship and appears attached to the entity with the "many" relationship. Relationship descriptors appear next to the entity they describe and should be read clockwise. Refer to the relationship between PRODUCT and SELECTION above in Figure 1: Each SELECTION is of one and only one PRODUCT, and each PRODUCT is bought via one or more SELECTIONS.

IV. GEM

While the business model for e-commerce is quite similar to any general retail business model at first glance, the nature of buying and selling online presents an organization with the challenge and opportunity to not only record sales figures but to follow their customer through the selection and purchasing process. Such information is a dream come true for marketing and sales analysts, who can use this information to more accurately determine what website components and layouts lead – or don't lead – to a successful sale. Figure 3 represents a **g**eneralized **e**-commerce data **m**odel (GEM) that can be used by online retailers to correlate site structure with a customer's website visit. GEM addresses the foundations of retail business and addresses issues unique to e-commerce, but does not include accounting or human resource modeling for simplicity.

The PARTY entity deals with the people involved in the purchasing process and contains two subtypes: PERSON and ORGANIZATION. Both of these subtypes have similar attributes and are often modeled as customer, but many companies need to distinguish between a purchase made by an organization and one made by a person.

PARTY ADDRESS holds telephone and address information for a PARTY and allows each party to have one or more addresses. While the address type (for example, billing or shipping) could be modeled as a separate entity with each PARTY ADDRESS being of one and only one address type and each address type embodied in one or more PARTY ADDRESSES, this model captures the same information in the PARTY ADDRESS field type. Normalization is maintained through a constraint on the field with a limited domain. This implementation simplifies the model and is also used in the entities ORGANIZATION, ENTRY, PAGE, and PAGE ITEM.

Two activities that PARTY participates in are a VISIT to the website and a SALE. In most business rules, a PERSON visits a site; that PERSON may act on his or her own behalf or on the behalf of another organization, but the VISIT is made by a PERSON. A PERSON makes one or more VISITS to the site, while a particular VISIT is made by one and only one PERSON. In this case, a VISIT is defined as any visit made to the site up until the time of purchase or until the site is left and may include attributes such as StartTime, EndTime, and PriorURL. Business rules may also specify a given length of time a page remains idle before the VISIT is terminated. If a VISIT is terminated and site navigation is later resumed, a new VISIT is recorded.

A SALE is similar to a purchase order and contains all items sold to a particular PARTY and information such as DateOrdered and TotalOrder. A SALE can take place between either an ORGANIZATION or a PERSON, and so is shown with a relationship to PARTY as a whole. A particular SALE is made to one and only one PARTY, while a PARTY is the recipient of one or more SALES. A particular SALE is composed of one or more SELECTIONS, which are analogous to a line item on a purchase order. Each SELECTION is part of one and only one SALE. A SELECTION is defined as a purchase choice made by the PARTY and bought at the conclusion of the site visit.

The selection of a PRODUCT is represented in the SALE through a SELECTION, and a PRODUCT can be bought via one or more SELECTIONS. A PRODUCT is an inventory item that can be represented in one or more PRODUCT REPRESENTATIONS, which in turn can represent one and only one PRODUCT. A PRODUCT REPRESENTATION can be a text description, a sound wave, a video clip, or any other representation appropriate for that PRODUCT. Attributes in this table include a Description of the representation, the date and time LastUpdated, and the Representation itself.

Because any one PRODUCT REPRESENTATION can appear on many PAGES and each PAGE can contain many PRODUCT REPRESENTATIONS, the bridge entity PAGE ITEM was created. A PRODUCT REPRESENTATION appears as one or more PAGE ITEMS, which in turn can be of one and only one PRODUCT REPRESENTATION. PAGE ITEM contains attributes such as ScreenPosition (expressed in pixel coordinates), Height and Width, most likely measured in pixels.

Because a PAGE ITEM can be any kind of multimedia representation, the field Type becomes important. Item type distinguishes between a text description, picture, sound clip, video, etc. and is implemented as a constraint on the field with a limited domain. (While a video clip does contain graphic information similar to that of a picture, that information is inferred and it is more important to note the presence of animation.)

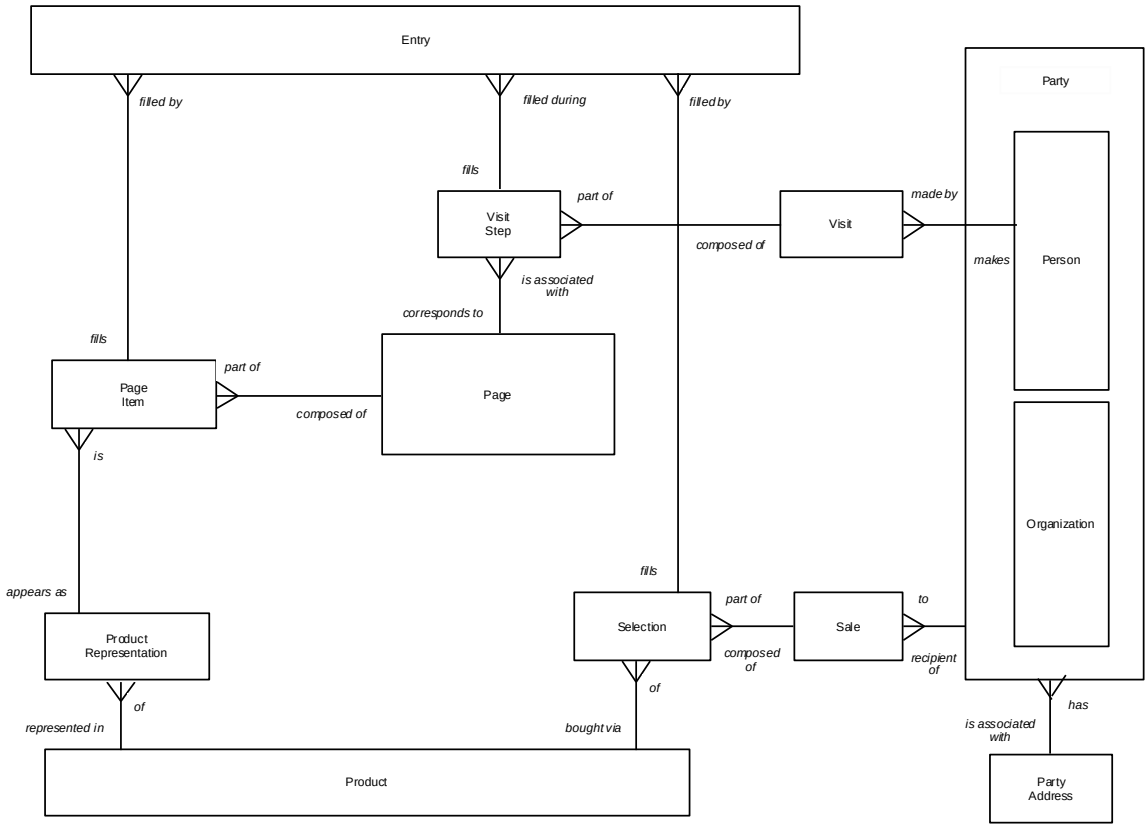
A PAGE ITEM can appear on one and only one page, while a PAGE can contain one or more PAGE ITEMS. A PAGE is defined as "a block of data available on the World-Wide Web, identified by a URL" (Online Computing Dictionary). In the most basic form, a web page is composed of static HTML, which is recorded as a PAGE. In more elaborate pages, content is generated dynamically. This distinction is made in the attribute Dynamic, which holds a Boolean value. The ability to record everything seen by a user on a dynamic page is captured through the relationship between PAGE ITEM (which contains content, size, and position information) and PAGE (which represents the whole). If the value in Generated is true, PAGE will have more than one PAGE ITEM associated with it; if it is false, only one PAGE ITEM will correlate. A company may wish to classify page types, i.e., the distinction between customer and supplier Internet pages. This information is captured in the field Type, which is implemented as a constraint with a limited domain.

The entities PAGE and VISIT meet in VISIT STEP. A PAGE corresponds to one or more VISIT STEPS, while one particular VISIT STEP is associated with one and only one PAGE. A VISIT is composed of one or more VISIT STEPS, while a particular VISIT STEP corresponds to one and only one VISIT. Attributes of VISIT STEP include StartTime and EndTime.

The final entity, ENTRY, represents an action taken by a website visitor. An entry is made here by VISIT STEP, PAGE ITEM, and SELECTION. When a user takes an action on a PAGE ITEM through a VISIT STEP or a SELECTION is made, information is recorded in ENTRY. This entity is the key to the correlation between site structure and visit behavior. Possible attributes include Type (e.g., radio button, check box, etc.) and SubmitTime. Type is implemented as a constraint on the field with a limited domain.

It is important to note the statistical and analytical capabilities available with the temporal data being collected. Possibilities include, but are not limited to average time spent on a page, average time spent on the site before a successful sale, and number of pages accessed before a successful sale.

Figure 3



V. Conclusion

A general data model for e-commerce applications is important because business strategists would be better able to track the online movements of customers. Among other things, this should lead to better service to customers because strategists could see which styles, page setups, colors, pictures, etc. lead to more successful sales. Conversely, understanding that slow speeds or awkward models lead to customers leaving a website should also lead to a change in the website structure. It is then imperative that e-commerce specialists have the data to track such movements, which is, of course, the purpose of GEM. Additionally, GEM can be used for a business information (BI) counterintelligence purpose which means that an organization can use GEM to track other organization's as they access the e-commerce website.

While a GEM prototype is currently being produced, future research in this area includes validating and refining the model and then empirically testing the model to demonstrate that it adds value to an e-commerce system.

Special thanks to David C. Hay for his guidance on the initial version of GEM.

References

1. Chen, P.P., "The Entity-Relationship Model: Toward a Unified View of Data." *ACM Transactions on Database Systems* (January 1976): 9-36.
2. Hay, D.C., *Data Model Patterns: Conventions of Thought,* New York, NY, Dorset House Publishing (1995): 6-7.
3. Jennings, M., "Theory and Models for Creating Engaging and Immersive Ecommerce Websites," *Proceedings of the 2000 ACM SIGCPR Conference* (2000): 77-85.
4. Online Computing Dictionary, http://www.InstantWeb.com/foldoc.cgi?query=webpage (March 2001).
5. Richebacher, T.F., "The Model Customer." *Intelligent Enterprise,* Volume 4 Number 2 (January 2001).
6. Silverston, L., Inmon W.H., and Kent Graziano, *The Data Model Resource Book: A Library of Logical Data Models and Data Warehouse Designs*, New York, NY, John Wiley & Sons, Inc. (1997).
7. Snodgrass, R.T., Developing Time-Oriented Database Applications in SQL, San Francisco, CA, Morgan Kaufmann (2000).
8. Sullivan, T., Business Intelligence Keeps Tabs on the Net, *Info World,* Volume 23, Issue 10 (March 5, 2001).

Appendix A – Entities and Attributes

Party Table:
{PartyID, *PersonID, OrganizationID*}

PartyAddress Table:
{PartyAddressID, *PartyID,* Street, City, State, ZipCode, Country, Phone1, Phone2, Phone3, Fax, Email, Type}

Person Table:
{PersonID, FirstName, LastName}

Organization Table:
{OrganizationID, ContactFirstName, ContactLastName, Type}

Visit Table:
{VisitID, *PersonID,* PriorURL, StartTime, EndTime}

Sale Table:
{SaleID, *PartyID,* DateOrdered, EstimatedDeliveryDate, TotalOrder, PaidInFull}

Selection Table:
{SelectionID, *SaleID, ProductID,* Quantity, ExtendedPrice, OnBackOrder, ActualDeliveryDate}

Product Table:
{ProductID, Description, Price, Cost}

Product Representation Table:
{ProductRepID, *ProductID,* Description, Representation, LastUpdated}

Page Item Table:
{PageItemID, *ProductRepID, PageID,* ScreenPosition, Height, Width, Type}

Page Table:
{PageID, URL, Dynamic, LastUpdated, Type}
Visit Step Table:
{VisitStepID, *VisitID, PageID,* StartTime, EndTime}

Entry Table:
{EntryID, *PageItemID*, *VisitStepID*, *SelectionID*, SubmitTime, Type}

Notes