December 2005

# Towards Analytical Approach to Effective Website Designs: A Framework for Modeling, Evaluation and Enhancement

Benjamin Yen
*University of Hong Kong*

Paul Hu
*University of Utah*

May Wang
*University of Hong Kong*

# Towards Analytical Approach to Effective Website Designs: A Framework for Modeling, Evaluation and Enhancement


Benjamin Yen
School of Business and Economics,
University of Hong Kong
benyen@business.hku.hk

Paul Jen-Hwa Hu
Accounting and Information Systems,
University of Utah
paulhu@ust.hk

May Wang
School of Business and Economics,
University of Hong Kong
amaywy@business.hku.hk



**Abstract**

*Effective website design is critical to the success of electronic commerce and digital government. Most prior website design research has taken a computational or cognitive/behavioral approach which may not yield optimal designs demanded by specific requirements. We consider website design as a structural problem which can be examined using analytical approach, such as mathematical optimization. Specifically, we propose a framework which classifies real-world design problems into generic website design categories and maps each resulting category into a graph model which can be analyzable or solved using appropriate analytical techniques. Our framework consists of generic designs and graph models, together with the necessary mapping. We classify the Web site applications and review their features proposed by previous research. We describe a generic website design category using its objective and key constraints that correspond to important design requirements. By modeling website design problems using well-defined structures and rigorous analysis methods, this framework is able to measure website accessibility in a systematic and quantifiable manner, arguably more desirable than existing qualitative ad-hoc practices. Overall, our framework can facilitate the website design process, enhance design quality, and increase ease of analysis, implementation and continuous improvement.*



**Keywords:** Web Site Design, Accessibility


## 1. Introduction

Effective website design is critical to the success of electronic commerce and digital government. Considerable efforts have been undertaken to evaluate and improve website designs. Both structure and accessibility are critical to website design and often have reciprocal causal effects. Conceivably, the structure of a website affects its accessibility of which analysis results (using visitors' browsing behaviors) in turn should guide the website structure design. There exist different definitions for accessibility in literature. According to the guideline by W3C, Web content accessibility should define target accessibility levels as well as making content accessible to people with disability. Lee et al. (2002) mentioned 35% of user concerned about accessibility issues and summarized accessibility as ease of operations, system availability, transaction availability, privileges, usableness, quantitativeness, convenience of access, locatability, security, privacy, etc.

Yang (2005) takes accessibility as availability and responsiveness, including being available all the time and desired speedy log-on, access, search and Web page download. In this study, we quantify the accessibility and map it into graph models to get an optimal structure given specific requirements of designers and users.

Navigation guidance, assessment and improvement models, and design guidelines are three common Web design issues in most e-commerce applications. Navigation guidance provides the visitors with the best directions to the destinations, and is dependent on both the navigation patterns and the analytic examination of site structures. The assessment and improvement models of Web sites function as critical building blocks for sustainable Web applications. The performance of evaluation criteria is profoundly influenced by Website structure. Besides the human factors, the determination of the Web structure and analysis is the cornerstone for the design guideline. The common essential part for the above is the systematic procedure and analytical models for analyzing, evaluating and enhancing website designs based on the interpretation of user requirements from various Web applications. Functionally, Web sites differ in processing capability, transaction complexity, information flow, security services, and range of supported applications. In this research, we focus on the structural design of Web sites, which can affect the actual functional utility of a Web site considerably.

We propose a framework which consists of application analysis, generic Web site design and graph modeling, together with the necessary mapping. Using this framework, we can represent a real-world design problem as a generic Web site design, which then can be mapped into accessibility model analyzable or solvable with established analytical techniques. Moreover, this framework can be applied to different domain, such as e-commerce, e-libraries, etc. We use formal modeling to analytically evaluate and improve Web site design accessibility, which have been mostly ad-hoc and qualitative. We also make contributions by characterizing Web site design problems using a set of generic designs, each of which can be described in terms of quantifiable objectives and constraints solvable for optimality. We illustrate the use of the proposed framework to map a generic design problem to accessibility model and evaluate Web site designs. For practices, we provide designers with a systematical process for guiding and enhancing their design and accessibility evaluation tasks, together with associated analytical techniques. Section 2 presents a review of related work and bring forward our motivation. Section 3 describes our proposed framework. Section 4 gives examples to illustrate how this framework is used to evaluate Web site design and make enhancement. Section 5 draws the conclusions and discusses the future work

## 2. Literature Review and Motivation

Previous research has examined different aspects of website design, including analysis, modeling and evaluation. Chevaliera and Ivory (2003) studied the cognitive activities of Web site designers and illustrated different Web site design constraints from investor's and user's points' of view. Shneiderman (1998) summarized prevailing theories, principles and guidelines for effective user interface designs for various websites. Ivory and Hearst (2002) surveyed popular (automated) website evaluation methods and graphical design practices, highlighting the importance of such design elements as architecture, page, link, text and graphics. Dhyani et al. (2002) classified and discussed a wide range of Web metrics and quantifies various attributes of Web sites, including Web graph properties, Web page significance (e.g. PageRank), Web page similarity, search and retrieval, usage characterization and information theoretic properties. Typically, a website contains a large collection of interconnected pages; this suggests the criticality of page placement and hyperlink in website structure design. The placements of individual pages on a website and the hyperlink structure connecting these pages have profound impacts on the effectiveness of a website design and often determine its usability and utilization. Web mining has been applied to improve website structure designs. Cooley et al. (2000) made an explicit distinction between content mining and usage mining. Content mining targets automated retrievals, filtering and categorization of Web documents and other resources, whereas usage mining extracts from Web logs prominent and potentially useful access behaviors of visitors. Prior Web mining research targeting website structure has focused on data that describe the (static) structural organization of a website; e.g., hyperlinks. In general, the Web mining approach is data-driven and may yield improved designs but not "optimal." In this connection, the use of defined data structure and analytical methods to evaluate and enhance website designs is appealing and has been explored. For instance, Brin and Page (1998) used a graph to model the structural design of a website. Kumar et al. (2000) proposed a stochastic model for building a Web graph of which edges are statistically dependent on each other and new vertices can be dynamically created over time.

The navigation aspect of website design has also been studied. Chakrabarti et al. (1999) analyzed the hyperlink structure of the Web and developed algorithms for improving information discovery and categorization. Anchored from a navigation aspect, Chakrabarti et al. (1999) as well as Gibson et al. (1998) differentiated pages on a website as "authority" and "hub." Based on visitors' content-access patterns, Sarukkai (2000) developed a Markov Chain model for link prediction and path analysis. The model can be used to predict the HTTP requests by the visitor and is capable of adapting to Web navigation and tour generation supported by personalized hubs and authorities.

## 3. A Framework for Website Design Modeling, Evaluation and Enhancement

The proposed framework consists of three layers, with a necessary mapping between the adjacent layers. As shown in Figure 1, the topmost layer is Application Layer which provides a high-level description of the website design under analysis and evaluation, such as industry (such finance, information technology, or entertainment), intended website purpose (such as informational or transaction-centric), and performance requirements and expectations (such as page loading time less than 3 seconds, allowing more than half of the visitors to access the targeted contents using 5 clicks or fewer). A website design described at the Application Layer is then transformed to an equivalent instance in the Generic Website Design Layer. We use objective and 4 constraint parameters to design generic website design categories. Exemplar objectives are minimal search time or number of clicks. The constraint parameters included in Generic Website

Design Layer are pertinent to cost of accessing a page, number of hyperlinks included on a page, desired order for content accesses, overall website structure. The framework includes a mapping between Application Layer and Generic Website Design Layer, which facilitates the transformation of a website design problem in the format of primary objective and specified constraints that correspond to its purpose and requirements (described in Application layer). A website design represented by its primary objective and 4 constraint parameters can be modeled using a corresponding graph and then analytically solved or evaluated by appropriate methods. In turn, such analyses and evaluations will shed light on "optimal designs" and general guidelines for website (navigation) design. Details of each layer in the framework are as follow.

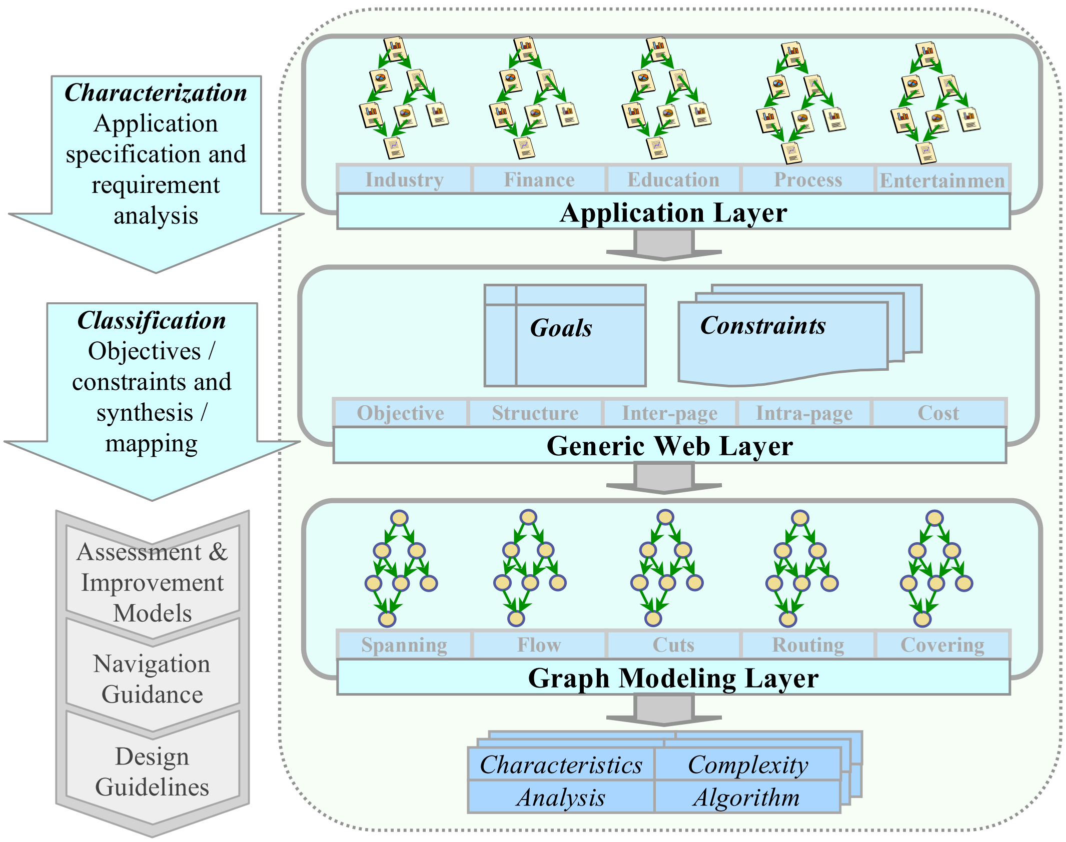


**Figure 1.** A 3-Layer Framework for Evaluating and Enhancing Web Site Designs

### *3.1 Application Layer*

Organizations vary in their intended use of websites. While some build website to disseminate product/service information, others depend on websites to executing transactions online. Websites also differ considerably in requirements at both industry and organization levels. For instance, websites pertinent to the finance sector may have more stringent caching requirements than education websites. Similarly, organizations in the same sector may also vary in their website design requirements, such as overall structure, page design, and navigation support. In this section, we will classify Web site applications and describe the features of each group.

### *3.1.1 Classification of Applications*

Based on business function, Hoffman et al. (1995) classified commercial Web sites into to two types: “Destination Sites” and “Web Traffic Control Sites”. The “Destination Site” integrates a firm’s virtual counterpart. The “Web Traffic Control Site” will direct consumers to various “Destination Sites”. Huizingh (2000) carried out an empirical study to identify five kinds of industries: computers, information, finance & insurance, services and products industries. The categories they included are listed in the Table 1.

| **Classification** | | | | | | **Supported Reference** |
|---|---|---|---|---|---|---|
| **Destination Site** | | | | | **Web Traffic Control** | Hoffman et al. 1995 |
| **Computers** | **Information** | **Finance & Insurance** | **Services** | **Products** | **Mall, Incentive Site, Search Agent** | Huizingh 2000, Hoffman et al. 1995 |
| Computer hardware, Software | Audio, books, communication & media services, music, telecom, broadcast stations, newspapers & magazines | Financial services, insurance | Advertising, consulting, corporate services, travel, PR, DM | Apparel, automotive, breweries, electronics, games, gifts, home & garden, industrial suppliers, jewelry, photography, toys, clothing, industry | Collection of online storefronts, search agent to identify other site through key word (e.g. Yahoo) | Huizingh 2000 |

Table 1: Classifications and categories of industries from Web Site

Zhang et al. (2000) also clustered company Web home pages into five groups and identified their different features. There are limitations in the existing classification. For example, most of the companies included in Huizingh 2000 are from the U.S and Netherlands. Moreover, a non-standard classification was reported (Robbins and Stylianou 2003). Robbins and Stylianou (2003) reported that content features were found to be significantly different across various cultural groups. However, this is not the case for design features. In the following section we study the features of different groups.

### *3.1.2 Features of Industries’ Web Sites*

Though designers create Web sites, two other actors do influence the design process (Chevaliera and Ivory 2003). One is the site’s clients (i.e. persons who fund the Web site), and the other is the site’s future users (i.e. future customers of the site’s clients). Both of the two actors will set some constraints on the Web site design. The client constraints are the requirements that clients explicitly prescribe or inferred by designers from prior interactions with other clients. The user constraints are that designers infer from their prior experiences as Web site users. User constraints usually address general interests to users and usability (e.g. ease of navigation). Clients may propose superficial requirements and simple features of Web site, such as e-mail address of the company, color of the background, etc. Table 2 consists of several features on Web home page.

| **Web Home Page Characteristic** | **Feature on Web Home Page** | **Supporting Reference** |
|---|---|---|
| Service | e-mail, on-line help, frequently asked questions | Zhang et al. 2000 |
| Index | search, site index, listbox | Zhang et al. 2000 |
| Commercial | e-commerce, log-ins, advertisement | Zhang et al. 2000 |
| Legal | legal, privacy, copyright | Zhang et al. 2000 |
| Financial | news, investor information, stock price | Zhang et al. 2000 |
| Number of window dressing features | colors, links, images (e.g. use different forms for repeated links, use conventional means for highlighting links, e.g. underline words/label) | Zhang et al. 2000 Ivory and Hearst, 2002 |
| Animation | animation, (e.g. graph animation is irritating to users and impede their scanning of contents) | Spool, et al. 1997 Zhang et al. 2000 |
| Miscellaneous | Jobs, career information, etc. | Zhang et al. 2000 |

Table 2: Web Home Page Characteristic (Zhang et al. 2000)

However, users may require more usage features. According to previous research there are several classifications of Web site criteria from user's perception. We can classify features from different point of view. For example, from user's satisfaction's point of view, there are features like relevancy, accuracy, comprehensibility and comprehensiveness of information, ease of use, entry guidance, Web site structure, hyperlink connotation, Web site speed, page layout and language customization [Muylle et al. 2004]. From content & design's point of view, there are features like information, transaction, entertainment, perception content, navigation structure, search function, protected content, quality of structure, and image and presentation style [Huizingh 2000]. There are overlaps in different frameworks proposed in previous research and the definitions are not the same in different work. For our later use for structure analysis, we classify them into two groups in Table 3. One is content related and the other is structure/navigation related.

| **Content** | | | **Structure/Navigation** | **Supporting Reference** |
|---|---|---|---|---|
| **Intrinsic IQ** | **Contextual IQ** | **Representational IQ** | **Accessibility IQ** | Lee *et al* 2002 |
| Accuracy, believability, reputation, objectivity | Value-added, relevance, completeness, timeliness, appropriate amount | Understandability, interpretability, concise and consistent representation | Accessibility, ease of operations, security | Wang and Strong 1996 |
| **Product Quality** | | | **Service Quality** | |
| Free-of error, concise representation, completeness, consistent representation, appropriate amount relevancy, | | | Timeliness, security, believability, | Lee *et al* 2002 |

| | | | |
|---|---|---|---|
| understandability, interpretability, objectivity | | accessibility, ease of manipulation, reputation, value-added | |
| **Quality** | **Presentation** | **Navigation** | |
| Believable information | Information location, distracting elements, graphics usage | Quantity, functionality and relevance of hyperlinks, layout, navigation structures | Zhang *et al* 2000 |
| **Information Quality** | | **System Quality** | |
| Content usefulness, information adequacy | | Usability, accessibility, security, interaction | Yang et al. 2005 |
| Accuracy, content relevance and completeness, format, timeliness | | Navigation, ease of use, response time, security in | Cheung and Lee 2005 |

Table 3: User's Requirements

The definition of information quality varies in different research for different purpose. Lee et al. 2002 includes security and operations as accessibility information quality, while Yang et al. 2005 includes only content related criteria in information quality and put ease of use and security into system quality. To summarize, the features can be view from two aspects. One aspect is content related. The content is very important, which have features shown in Table 4. The other is structure/navigation related. In this study we will focus on structure, the features of which were shown in Table 5.

| Dimensions | Brief Description | Importance & Examples | Supporting References |
|---|---|---|---|
| **Accuracy** | The accuracy of information on the Web site. The readability should be acceptable | The reliability of the information affects consumer evaluation of the Web site and purchasing decision. | Ho and Wu 1999 Spool et al. 1997 |
| **Relevance and Completeness** | The relevance and completeness of information on the Web site content. | Content is the most important thing, providing relevant information to help user surf and shop online. Complete information can help user make competent and informed decision about product or service. | Madu and Madu 2002 |
| **Media Format** | The way information is presented on the Web site | The media richness of the Web facilitates the provision of graphics, text, sound, and video, making information attractive as well as useful. | Madu and Madu 2002 |
| **Timeliness** | The timeliness of the information on the Web site | The Web site should be frequently updated to avoid outdated information | Madu and Madu 2002 |

Table 4: Content Related Features

| Dimensions | Brief Description | Examples | Supporting References |
|---|---|---|---|
| **Size** | The number of pages of the Web site. Size is | Provide textual information in smaller units or using multiple pages. | Huizingh 2000 |

| | | | |
|---|---|---|---|
| | significantly related with almost all content and design aspects. | Size of website be appropriate and can vary with applications. | Nielsen, 2000 Ivory and Hearst, 2002 |
| **Page Layout** | The format of a single Web page. | To facilitate marketing purposes or aesthetic requirements e.g. the promotion product can be placed in the center. | |
| **Structure/Navigation** | There are tree structure and general graph structures. Different structure provides different kinds of access or navigation features. | Expose users to multi-level information architecture (i.e., link clustering with headings). The sequence of pages should follow application requirements. The number of links going out of a page should be constrained. | Saywer, et al. 2000 |
| **Response Time** | The speed of access and availability of the Web site at all time. The download speed of files and graphs. The searching time and the number of clicks to access the destination page should be as short as possible. | The slow speed will cause the user to abandon the transaction. Download speed should not exceed 10 seconds, screen size should be explicitly defined | Kim and Lim 2001 Nielsen, 2000 |
| **Security** | The website's ability in protecting consumer personal information collected from its electronic transactions from unauthorized use or disclosure. | Privacy and security of online transaction are important to build trust and long-term relationship. Some page must be accessed after log-ins. | Madu and Madu 2002 |

Table 5: Structure/Navigation Related Features

Different industries have difference features of their Web sites. An empirical study has revealed some of the features in different industries (Huizingh 2000). Table 6 shows the features of different industries in "Destination Sites". Search agent Web sites will direct to "Destination Sites" and will not be discussed here.

| **Industry** | | **Computers** | **Information** | **Finance & Insurance** | **Services** | **Products** | **Supporting References** |
|---|---|---|---|---|---|---|---|
| **Content** | **Accuracy** | Medium | High | High | Medium | Low | Partially Supported by Zhang et al. 2000 Huizingh 2000 |
| | **Content** | High | High | Medium | Medium | Low | |
| | **Media Format** | Complex | Complex | Simple | Simple | Complex | |
| | **Timeliness** | High | High | High | Medium | High | |
| **Structure** | **Size** | Large | Large | Small | Small | Small | |
| | **Layout** | Complex | Complex | Simple | Medium | Simple | |
| | **Structure** | Complex | Complex | Simple | Medium | Simple | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | **Response Time** | High | Medium | High | Medium | Low | |
| | **Security** | Medium | High | High | Low | Low | |

Table 6: Features of Industries

The features may not be exactly the same even for the same industry. The website scale and structure vary in different website. Table 6 only shows some of the common features in some of website in the industry studied in previous work. For example, the security requirement of financial website is high and user cannot access certain page without going through the log-in page. For illustration, let's assume a financial institution plans to build a website which offers clients online access to their accounts and information of different investment funds. This website supports online transactions (such as purchasing or selling financial assets) and therefore its design prohibits an explicit return (back) hyperlink for any page access and, at the same time, disallows content caching. For increased usability, all pages on this website cannot contain more than 15 hyperlinks pointing to other pages, internal or external. For general marketing or promotion purposes, page accesses by the visitor has to follow some specified sequence, such as exposing the visitor to a page highlighting a particular investment fund before his or her access to a targeted content (such as his or her account). In light of their respective objectives and requirements, website designs can be considered as optimization problems; that is, maximizing or minimizing an objective while subject to particular constraints. In the next sub-section, we describe the representation of website designs using a generic objective-constrain structure that can be directly translated to the corresponding mathematical formulation.

### *3.2 Generic Website Design Layer*

In spite of the heterogeneous objectives and diverse requirements, websites can be classified into generic categories. In Generic Website Design layer, we define a set of generic design categories that differ in objectives and constraints. The objective of a website conceivably can be analyzed from a perspective of the user or the designer. In this study, we consider both perspectives in defining generic website objectives but place a conscious emphasis on the user's objective because its realization is necessary for accomplishing the designer's objective. Users typically value effective searches, such as accessing more relevant documents/resources or assembling a large choice set for their purchase consideration. Users may also value search efficiency, such as retrieving targeted or interesting contents using less time or fewer numbers of clicks. Table 1 lists example objectives and constraints.

| | | | **Objectives** | **Constraints** |
|---|---|---|---|---|
| **Designer** | **Structure** | **Website Level** | Users given the longest path to destination; more products will be seen; number/cost of links should be minimized for ease of maintaining and efficiency, etc. | Number of links branching out from a page is constrained; some pages cannot be accessed using back links for security; some pages have to be accessed in particular order or within specified distance; adapt website to dynamic access behaviors, etc. |

| | | | | |
|---|---|---|---|---|
| | | Page Level | Contents can be minimized and updated automatically, etc. | Promotional product placed in the center of a page, etc. |
| | Human Computer Interface | Text | Acceptable readability (Spool, et al., 1997) , etc. | Provide textual information in smaller units or using multiple pages (Nielsen, 2000), etc. |
| | | Link | Expose users to multi-level information architecture (i.e., link clustering with headings) (Saywer, et al., 2000), etc. | Use different forms for repeated links, e.g. text and image (Ivory and Hearst, 2002), number of links branching out from a page is confined, etc. |
| | | Graph | Download speed of graphs be minimized. | Graph animation is irritating to users and impede their scanning of contents (Spool, et al., 1997), etc. |
| | | Page | Text density should facilitate page scanning by users (Saywer, et al., 2000; Spool, et al., 1997), etc. | Download speed should not exceed 10 seconds, screen size should be explicitly defined (Nielsen, 2000), etc. |
| User | Structure | Site Level | Search time, number of clicks and content access time be minimized; maximize page accessibility, etc. | Size of website (such as number of pages or documents) be appropriate and can vary with applications, e.g. newspaper or different kinds of books (Ivory and Hearst, 2002), related pages be connected; entire website is within a certain height and width for ease of search, etc. |
| | | Page Level | Contents can be minimized and be updated frequently, etc. | Page layout be consistent; page download speed be constrained; download speed of graphs be minimized, etc. |
| | Human Computer Interface | Text | Acceptable readability (Spool, et al., 1997). | Textual content on inter-connected pages should be related, etc. |
| | | Link | Maximize accessibility of links; related links can be found on portal page, etc | Use different forms for repeated links, e.g. text and image (Ivory and Hearst, 2002), use conventional means for highlighting links, e.g. underline words/labels, etc. |
| | | Graph | Download speed of graphs be minimized, etc. | Attractive graphic designs, particularly those describing products and services, etc. |
| | | Page | Download time should not exceed 10 seconds (Nielsen, 2000), etc. | Pages be accessible to people with disability, etc. |

Table 7: Examples of General Website Design Objectives and Constraints

We represent generic website design categories using a 4-constraint-parameter tuple (*C*,*O*,*R*,*S*), where *C* is the cost of an arc, *O* denotes out-going degree constraint on vertices, *R* represents sequencing relationship between some vertices, and *S* describes structure constraint. Cost of an arc is the time required for downloading a page, which is affected by page size and network traffic. For simplicity, we can assume identical page loading time; i.e., symmetric cost of arc. Casual observations show pages considerably vary in loading time, thus suggesting the use of a directed graph that embraces backward links. We also consider other constraints. To provide easy access to other pages and prevent pages from growing overwhelmingly, a website should limit the out-going degree of its pages, such as confining the number of hyperlinks on a page to specified upper and lower bounds. Sequencing relationships depict the order in which the visitor access pages on a website. By design, some pages are shown to the visitor before his or her access particular pages. This suggests a relationship between pages which can be immediately adjacent or away by a specified distance. For instance, two pages distant within 2 nodes have one intermediate page between them. Structure constraints describe the structure of a graph. Tree is a common structure and is advantageous in simplicity and maintainability (such as hyperlink updates). Directed acyclic structure is effective for allowing the user to reach a targeted content by following different navigation paths. A directed cyclic graph allows the user to return to any of pages previously visited. When N constraints are taken into consideration, N-constraint-parameter tuple will be used to represent the generic website design. These four constraints are chosen for preliminary study and for their importance to represent structural and security aspect of a website.

Continued with the financial website example, let's assume the primary objective be quick information access (retrieval). Using a 4-constraint-parameter tuple, we can model this website design as the followings. The primary objective is minimized average access time. The security and disabled caching requirements imply non-identical and asymmetric costs of arc. The restriction on the branching links suggests an upper-bound constraint on out-going degree of a page. Order of page access signals immediate sequencing relationships desired or planned. Table 2 lists possible values of the constraints described.

| | **Cost of Arc (C)** | **Out-Going Degree (O)** | **Sequencing Relationship (R)** | **Structure Constraint (S)** |
|---|---|---|---|---|
| 0 | Identical | No Constraint | No Constraint | No Constraint |
| 1 | Symmetrical | Upper Bound Constraint | Immediate Sequencing | Directed Tree Structure |
| 2 | Nonsymmetrica l | Lower Bound Constraint | At most one node in-between | Directed Acyclic Graph Structure |
| 3 | | Upper and Lower Bound Constraints | At least one node in-between | Directed Cyclic Graph Structure |

Table 8: Listing of Possible Values for 4-Parameter Constraint Tuples (C,O,R,S)

### *3.3 Mapping between Application Layer and Generic Website Design Layer*

To illustrate the mapping between the Application Layer and Generic Website Design Layer, let's consider a website for supporting self-served trouble shooting by customers of a particular laptop made and model. Assume the primary design objective be minimizing the number of clicks required to access targeted trouble-shooting contents (such as technical documents or online instructions). The website can be modeled using a directed tree if all of its pages by design are connected and do not allow circles in its graph representation. Alternatively, the website can be consider as a search tree in which the visitor wants to access each page once and then return to starting page when it presents search results systematically, such as in alphabetic order. A sequencing constraint exists to ensure the visitor's browsing through particular pages before reaching a targeted content. To balance connectivity and usability, there is a constraint on the upper and lower bounds for outgoing degree, such as pages containing (inclusively) between 5 and 20 hyperlinks pointing to other pages (internal or external). Page loading time is critical and is used to approximate the cost of a page. In this case, we can model this website design using tuple 3 of the generic design defined by the constraints on cost of arc, outgoing degree, sequencing relationship, and structure (see Table 2).

Many but not all website designs can be modeled as graph problems. At the same time, some graph-based methods are directly applicable to analyzing or evaluating website designs while others require further extensions to be so. From a research aspect, it is interesting to scrutinize the intersections between website design and graph-based analysis (such as how a website design can be analyzed or evaluated using a particular graph-based method) while identifying those not readily for such analyses and specifying the additional requirements to be addressed. In this study, we propose a framework that allows us to classify generic categories of website designs and maps each category to a graph structure to be analyzed and evaluated using established analytical methods. In addition, we also explore website designs not directly analyzable using graph-based modeling, thus shedding light on the potential boundary of its applications in website designs and evaluations, together with the necessary extensions.

### *3.4 Graph-based Modeling and Analysis Layer*

The use of graph theory to model and analyze website designs has been investigated. Modeling a generic design as a graph-construction configuration problem allows searches of optimal designs as well as evaluating existing designs analytically. Websites usually have constraints germane to structure (such as directed tree versus directed acyclic graph), page sequence (such as immediate versus general sequence), outgoing degree (such as upper or lower bound), and cost of arc (such as identical versus non-identical). Accordingly, we can use primary objectives and key constraints to define generic website design categories and model each category as a particular graph problem.

Mathematical notations of the generic design will be provided to facilitate design analysis mathematically. Some of the objectives and constraints can be formulated as mathematical programming problems, while some can not. For example, no typo error on website is one of the constraints, but it cannot be formulated as a mathematical programming problem. Some of the objectives and constraints can only be achieved through user interface design, such as the preference for color, etc. In this research, we focus on the structural design of Web site, and will discuss the objectives and constraints that can be quantified. We described some of them in mathematical notation. First, the objective of accessibility will be illustrated as example to show the transformation from description to mathematical notation. Second, several constraints will be presents in mathematical form.

The mathematical formulation of objective to achieve accessibilities is shown below. The accessibility of a link is the attractiveness or priority in a Web page. It can be denoted as a low value or a high value. Accessibility of Web pages can be determined by the expected link number (Yen, 2004).

$$AP_i = f(EN_i) = aEN_i + b,$$
$$EN_i = \sum_{j=1}^{N} w_i AL_i,$$

where $EN_i$ is the expected number of the links pointing to page i. The $j^{th}$ link pointing to page i is denoted as $i_j$. $w_i$ is the level weight of the page that the $j^{th}$ link located and $N_i$ is the total number of the links pointing to page i. $AL_i$ is the accessibility of $j^{th}$ link pointing to page i. Each AL here is set as 1 for simplicity. AL can also be assigned other value to represent the link's attractiveness.

We try to illustrate some constraints that can be formulated as mathematical programming problems in the following. Several constraints are shown below.

1) All the pages in the Web site should be connected.

   This constraint commands the structure of the website link structure to be a connected graph. We assume that the node 1 is the root. For each node, except the root, there is at least one in-going link. There must be at lease one out-going link from the root. Here are some general constraints:

$$\sum_{i}^{n} x_{ij} \geq 1, i \in (2,...,n), \qquad \text{(E1)}$$

$$\sum_{j}^{n} x_{1j} \geq 1, \qquad \text{(E2)}$$

$$x_{ii} = 0, i \in (1,...,n), \qquad \text{(E3)}$$

$$x_{ij} \text{ 0-1 integer} \qquad \text{(E4)}$$

Where $x_{ij}$ is a zero-one variable indicating whether arc (i,j) belongs to the spanning tree. The $c_{ij}$ denotes the cost of arc (i,j), i.e. the loading time of page j, or depending on the settings in the experiments. Constraint (E1) indicates that each vertex should be pointed to by some other vertex, except the root node, so that the graph is connected. Constraint (E2) shows that the root node must point to some other vertex/vertices. Constrain (E3) and (E4) shows that $x_{ij}$ is a zero-one variable and there is no self-loop for each vertex.

2) All the pages are of the same size.

We assume the loading time of each page is the same. In another word, all the costs on the arcs are the same.

$$C_{ij} = 1, i, j \in (1,...,n), i \neq j \qquad \text{(E5)}$$

3) If the out-degree is specified, the following constraint exists:

$$\sum_{j}^{n} x_{ij} \leq m, j \neq i, \text{ integer.} \qquad \text{(E6)}$$

Different constraints may have different values. The cost of each link can be taken as the same when we assume the loading time of the pages are similar. The degree of the out-going degree can be changed when we define the number of links that one page can point to. We will define sequence relationship for purpose of security and define the structure as well. Moreover, different objectives are corresponding to a set of constraints.

Prior research has defined a set of classic problems that can be modeled using graphs. Examples include spanning tree, routing, flow and cut, covering and partitioning, vertex ordering, isomorphic, and sub-graph. Each problem has been further examined and applied in different contexts. For instance, different spanning-tree problems have been discussed, such as minimal spanning tree (MST), capacitated spanning tree, degree constrained spanning tree, and maximum leaf spanning tree. Similarly, routing problems can be further classified as shortest/longest path between two nodes, $K^{th}$ shortest path, shortest path between one and all other nodes, and traveling salesman problem (TSP). Some graph modeling problems can be solved in polynomial time, while others are NP-hard and therefore may be examined using heuristic, computational methods.

We expect incomplete mapping between Generic Website Design Layer and Graph-based Modeling Layer. For example, caching can provide "conditional" back links, i.e., generating the corresponding back link when a link is traversed. Mapping such design scenarios to adequate graph models is challenging. On the other hand, some graph models cannot be directly applicable to generic website designs which may focus on fundamental or obvious requirements rather than exploring alternatives for in design objective or constraints. The incompleteness is not addressed in this study but requires further investigations.

The proposed 3-layer framework represents an analytical approach towards website design, evaluation and enhancement and can be extended to support other important issues surrounding website design, such as navigation guidance (Yen and Wan, 2003), assessments and improvements (Yen, 2004), and general design guideline. The graph models derived from Graph-based Modeling and Analysis Layer can be applied to generate "optimal" navigation for deterministic as well as stochastic instances. Assessment and improvement models that take page accessibility into account can produce designs with a desired balance between page accessibility and popularity. Yen (2004) defined four accessibility models: expected link numbers, accumulated accessibility, sum of distance reciprocal, and sum of expected distance reciprocal. Towards this, graph-based models derived from Graph-based Modeling and Analysis Layer can be extended to yield criteria for selecting accessibility models appropriate for different website applications and requirements. In addition, the proposed framework can be used as a verification and validation toolkit for structure-based guidelines for website design requirements analysis, performance evaluation and continuous enhancements.

## 4. Illustrating Use of Framework - Accessibility Analysis and Evaluation

Using different values of these 4 parameters as shown in Table 2 that represent key constraints, a total of 192 scenarios are possible. The scenario described at the beginning of the section indeed represents case (1,1,1,0). To solve this problem, we first transform it into case (1,1,0,0), for which several existing theories have been developed. Based on important problem characteristics, case (1,1,0,0) can be considered to be as a typical degree-constrained MST problem, which is NP-complete. Deo and Kumar (1996) proposed an iterative refinement method for computing reasonably good sub-optimal solutions for large instances of constrained spanning tree problems. The refinement repetitively compute an MST and penalize edges incident to nodes with a degree greater than a specified upper bound, terminated when the spanning tree obtained has node with a degree exceeding the upper bound. This method yields encouraging results for degree-constrained MST problems. In the following, we provide an instance of such scenarios. We assume that there are only 8 pages in the Web site, page 1 is the root, and page 2 immediately follows page 3. The problem can be transformed to an optimization problem as the following:

$$\text{Minimize } CX^{\mathrm{T}}$$

$$\sum_{i}^{n} x_{ij} \geq 1, i \in (2,\ldots,n), \qquad (1)$$

$$\sum_{j}^{n} x_{1j} \geq 1, \qquad (2)$$

$$x_{ii} = 0, i \in (1,\ldots,n), \qquad (3)$$

$$x_{ij} \quad \text{0-1 integer} \qquad (4)$$

$$Level(1)=0;\ Level(j)=Level(i)+1 \text{ if } x_{ij}=1 \qquad (5)$$

$$x_{23}=1, \qquad (6)$$

$$\sum_{j}^{n} x_{ij} \leq m, j \neq i, \quad m=2. \qquad (7)$$

$C$ is a matrix that represents the cost from page $i$ to page $j$. $x_{ij}$ is a zero-one variable indicating whether arc $(i,j)$ belongs to the spanning tree. $c_{ij}$ denotes the cost of arc $(i,j)$; i.e. the loading time of page $j$ or dependent on the experimental design. Constraint (1) states that each vertex should be pointed to by some other vertex, except the root node; thus the graph is connected. Constraint (2) shows that the root node must point to some other vertex/vertices. Constraint (3) and (4) suggest that $x_{ij}$ is a zero-one variable, and that there exists no self-loop for each vertex. If the link structure of the Web site is a spanning tree, we observe characteristic of constraint (5). In addition, constraint (6) implies page 2 to be immediately following page 3. With respect to constraints (7), the upper bound of outgoing degree of each node is set to 2. This example problem can be solved by using some software (such as LINGO) for linear or integer programming problems. The result can be used as an alternate design solution for the designer and compared with initial design.

## 5. Future Research Directions

To better work out and evaluate accessibility of Web site designs, we propose a framework consisting of generic design and accessibility modeling. Using this framework, we can map a Web site design to a generic design problem, which is then mapped into accessibility model with established techniques. Our framework facilitates the Web site design process by using measurements by quantifiable rather than qualitative approaches. In this paper, we focus on the structural design of Web sites. This framework can be applied into the content management system by forming a logical content hierarchy. The application analysis indicates a need to perfect the classification of industries in according to their standard features instead of evaluation criteria. The framework can be further applied to other aspects, such as functionality, to improve information flow, transaction complexity and security services. It also may be extended to other mathematical models in addition to accessibility or graph model. Furthermore, stochastic models can also be applied in the framework as dynamic behaviors are considered. The framework can evolve as the design constraints change or extend. However, the proposed framework cannot be applied to support all design objectives and constraints, especially some human interface related objectives and constraints. To address this and other limitations, several areas need our continued investigations. First, the classification of the application industries will be verified by qualitative methods. Second, empirical justification should be shown in classification of objectives and constraints. Third, the completeness of the graph model should be justified and extended in future. More real and complex example should be shown for illustration of using the framework.

**Acknowledgment of PACIS paper -**

This research is partially sponsored by The University of Hong Kong under Research Initiation Grant 02-04 and Seed Funding for Basic Research 04-05.

**Reference**


- Brin, S. and Page, L. The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Proceedings of the Seventh International World Wide Web Conference, Australia,* April 1998
- Chakrabarti, S., Dom, B., Kumar, S.R., Raghavan, P., Rajagopalan, S., Tomkins, A., Gibson, D. and Kleinberg, J.M. Mining the Web's Link Structure. *IEEE Computer*, 32(8), Pages 60-67, 1999.
- Chakrabarti, S., Dom, B., Kumar, S.R., Raghavan, P., Rajagopalan, S., Tomkins, A., Kleinberg, J.M. and Gibson, D. Hypersearching the Web, *Scientific American*, June, 1999.
- Cheung, M.K. and Lee, K.O., Research Framework for Consumer Satisfaction with Internet Shopping, working paper, (access date: 31/5/05; *http://www.is.cityu.edu.hk/Research/WorkingPapers/working_paper05.htm* )
- Chevaliera, A. and Ivory, M. Y. Web site designs: Influence's of designer's expertise and design constraints *Int. J. Human-Computer Studies 58 (2003) Pages: 57–87*
- Cooley, R. Automatic Personalization Based on Web Usage Mining. *Communications of ACM*, 43(8), August 2000
- Deo, N. and Kumar, N. Capacitated MST and Related Problems: Challenges in a Unified Approach, http://citeseer.ist.psu.edu/61074.html, April 09 1996.
- Dhyani, D., Ng, W.K. and Bhowmick, S.S. A Survey of Web Metrics, *ACM Computing Surveys,* 34(4), Pages 469-503, December 2002

- Garey, M. R. and Johnson, D.S. *Computers and Intractability - A Guide to the Theory of NP-Completeness.* W. H. Freeman and Company, New York 1979
- Gibson, D., Kleinberg, J. and Raghavan, P. Structural Analysis of the World Wide Web. Position paper at the *WWW Consortium Web Characterization Workshop*, November, 1998.
- Ho, C.F., and Wu, W.H. Antecedents of customer satisfaction on the Internet: an empirical study of online shopping, *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences, IEEE Computer. Soc, Los Alamitos, CA, Maui, HI, USA, 1999, Page 9.*
- Hoffman, D. L., Novak, T. P. Chatterjee, P. Commercial Scenarios for the Web: Oppportunities and Challenges, *Journal of Computer Mediated Communication 1 (3) 1995*
- Huizingh, K. R. E. The content and design of web sites: an empirical study, *Information & Management*, 37(3), 1 April 2000, Pages: 123-134
- Hull, L. Accessibility: It's Not Just for Disabilities Any More. *ACM Interactions*. 11(2), Column: The whiteboard Page:36-41, March/April 2004
- Ivory, M.Y. and Hearst, M.A. Improving Web Site Design, *IEEE Internet Computing, Special Issue on Usability and the WWW ,* 6(2), Pages 56-63. March and April, 2002
- Kahn, B. K., Strong, D. M., Wang Y. R., Information quality benchmarks: product and service performance, *Communications of the ACM,* 45(4), April 2002, *Supporting community and building social capital SPECIAL ISSUE on Virtual extension* Pages: 184 - 192
- Keng Siau, P. K. Measuring information quality of web sites: development of an instrument, *International Conference on Information Systems Proceeding of the 20th international conference on Information Systems Charlotte, North Carolina, United States* 1999, Pages: 279 - 285
- Kim, S.Y., and Lim, Y.J. Consumers' Perceived Importance of and Satisfaction with Internet Shopping, *Electronic Markets 1(3,) 2001,* Pages: 148-154.
- Kumar, S.R., Raghavan, P., Rajagopalan, S., Sivakumar, D. and Tomkins, A. Random Graph Models for the Web Graph. *Proceedings of 41st IEEE Symposium on Foundations of Computer Science.* 2000
- Lee, Y. W., Strong, D. M. Beverly K. Kahn and Richard Y. Wang, IMQ: a methodology for information quality assessment, *Information and Management,* 40(2), (December 2002) Pages: 133 - 146
- Madu, C.N., and Madu, A.A. Dimensions of E-quality, *International Journal of Quality and Reliability Management* (19:3) 2002, Pages: 246-258.
- Muylle, S., Moenaert, R. and Despontin, M. The conceptualization and empirical validation of web site user satisfaction, *Information & Management*, 41(5), May 2004, Pages 543-560
- Nielsen, J. *Designing Web Usability: The Practice of Simplicity.* Indianapolis, New Riders Publishing, 2000.
- Sarukkai, R.R. Link Prediction and Path Analysis using Markov Chains. *Computer Network*, 33, Pages 377-386, 2000.

- Sawyer, Paul, Richard Danca, and Will Schroeder. Report 6: Myths of page layout. In *Designing Information-Rich Web Sites*. Bradford, MA: User Interface Engineering, 2000.
- Shneiderman, B. *Designing the User Interface: Strategies for Effective Human-Computer-Interaction*. Addison Wesley Longman, 1998
- Spool, J.M., Scanlon, T., Schroeder, W., Snyder, C. and DeAngelo, T. Web Site Usability: A Designer's Guide*, North Andover, MA User Interface Engineering*, 1997.
- Robbins, S. S, Stylianou, A. S., Global Corporate Web Sites: an Empirical investigation of content and design *Information and Management 40 (2003), Pages 205-212*
- Yang, Z., Cai, S., Zhou, Z. and Zhou, N. Development and validation of an instrument to measure user perceived service quality of information presenting Web portals, *Information & Management,* 42(4), May 2005, Pages 575-589
- Yen, B. P.-C. Structure-Based Analysis of Web Sites, *Proceedings of IEEE International Conference on e-Technology, e-Commerce, and e-Service*, March 29-31, 2004
- Yen, B.P.-C. and Wan, Y.W. Design and Evaluation of Improvement method on the Web Information Navigation - A Stochastic Search Approach, *Proceedings of IEEE 36th Hawaii International Conference of System Science (HICSS-36),* Hawaii, January 2003.
- Zeithaml, V. A., Parasuraman, A., Malhotra, A. Service Quality Delivery through Web Sites: a Critical Review of Extant Knowledge, *Journal of the Academic of Marketing Science* 30(4), 2002, Pages 362-374
- Zhang, X.,Keeling, K. B., Pavur, J. R. 2000 Information Quality of Commercial Web Site Home Pages: an explorative analysis, *Proceedings of the twenty first international conference on Information systems* Brisbane, Queensland, Australia Pages: 164 - 175