Towards Effective Web Site Designs: A Framework for Modeling, Design Evaluation and Enhancement

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

Paul Hu
School of Business
University of Utah
paulhu@ust.hk

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

Abstract

Effective Web site design is critical to the success of e-commerce. Therefore, the evaluation and enhancement of a Web site design is of great importance. In this vein, accessibility is important and has been examined by a lot of researchers from different points of views. By and large, Web site accessibility is a structural problem and may be analytically investigated using mathematical approach. We propose a framework for representing real-world design problems as generic Web site designs, which then can be mapped into accessibility models analyzable or solvable using established analytical techniques. The framework consists of generic design and graph models, with the necessary mapping. We describe a generic Web site design using its objective and constraints, which correspond to important design requirements. By representing design problems using well-defined structures and rigorous analysis methods, this framework measures Web site accessibility using systematic and quantifiable approaches rather than qualitative ad-hoc practice. Hence, the framework facilitates the overall Web site design process, enhances design quality, and increases ease of analysis, implementation and continuous improvement.

1. Introduction

Effective Web site design is critical to the success of e-commerce. Considerable efforts have been undertaken to evaluate and improve Web site designs. In this vein, accessibility is very important and has been examined by prior research, for example, Web content accessibility guideline by W3C, which aims to explain how to make Web content accessible to people with disabilities and to define target levels of accessibility. Hull suggested expanding the scope of accessibility to all conceivable users, while recognizing the lofty costs and near impossibility [6].

We propose a framework which consists of 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 various aspects of Web site design, including user interface, structure, and navigation support. Shneiderman summarized key theories, principles and guidelines for designing user interfaces for Web sites [9]. Ivory surveyed prevailing (automated) evaluation methods for Web sites and graphical designs, highlighting the importance of such design elements as architecture, text, link, graphics, and page. Central to the design of a Web site is structure, which significantly affects on its accessibility [7].

Typically, a Web site consists of a large collection of pages, each linked to others. Brin and Page modeled the structural design of a Web site [1] in graph. Kumar et al. 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 [8]. Cooley developed the WebPersonlizer System, leveraging Web usage mining for personalized product messages on a massive scale [2].

Design evaluation also has been studied. Ivory and Hearst summarized common tools for Web site design evaluation, which include simulation-based or quantitative techniques, and guideline reviews [7]. Dhyani et al. surveyed common metrics for Web design and suggested several key categories, including Web graph properties, Web page significance, usage characterization, Web page similarity, Web page search and retrieval, and information theoretic [4]. Some metrics target link-structure or focus on user interaction, while others address both simultaneously.

A review of extant literature suggests a lack of systematic approaches for analyzing, evaluating and improving accessibility of Web sites. Of particular alarming is the limited discussion of using well-defined structure and analytical techniques in Web site design evaluation and enhancement. In response, we propose a framework for representing a real-world design problem as a generic Web site design transformable to an appropriate graph problem to be analyzed for enhanced design.

3. A Framework for Web Site Design Evaluation and Enhancement

A website embraces both structural and functional aspects. 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. As shown in Figure 1, applications are the topmost layer of our framework and can be described by purpose, targeted users, industry sector, and others.

Regardless of the diversity or heterogeneity, different design problems can be characterized by a set of generic Web site designs, each of which has its own primary *objective* and key *constraints*. In general, users value effective search and efficient retrieval of target (relevant) contents. Thus, important objectives include minimized search time, number of click and content access time. Additional objectives may include maximized relevant information search-space, choice set or retrieved contents. On the other hand, designers may be motivated by additional and different objectives; e.g., maximized site traffic, content availability, visitors' exposure to advertisements (static or dynamic), transaction volume, or conversion rate. To a large extent, designers' objectives often are realized only when a design satisfies users' objectives. This suggests a design should address the user's objectives prior to or in conjunction with designers' objectives. Web site design can be modeled as an optimization problem; i.e., maximizing or minimizing an objective, subject to specified constraints. Examples of Objectives and Constraints are listed in Table 1. We then can classify generic Web site designs using their respective objectives and constraints. Common objectives include minimized user path required for reaching target content(s) or maximized the number of interesting (relevant) products for consideration. Examples of key constraints include provision of explicit links for returning to a proper page on a Web site, or restricting the distance between/among pages for related products (e.g., similar or complementary) within a specified number of clicks. In addition, "optimal" Web site designs need to address additional challenges than conventional optimization problems, including frequent (continuous) content updates/changes and visitors' dynamic access behaviors which in part co-determine "optimality".

3.1 Mapping from the application requirement to analytical models

Prior research has modeled Web site design problems using the graph theory. Modeling a generic design as a graph-construction configuration problem allows the search of an optimal design or evaluating an existing design analytically. Understandably, a Web site has constraints pertinent to structure (e.g., directed tree or directed acyclic graph), page sequence (e.g., immediate versus general sequence), outgoing degree (e.g., constraints on upper or lower bound), and cost of arc (e.g., identical versus non-identical). Then the generic design can be transformed to a particular graph problem. Let's consider a Web site designed for customers' self-served trouble shooting for a specific laptop model. Assume the primary objective be minimizing the number of click needed to access target trouble-shooting materials. This problem can be mapped into a shortest path problem for minimized user path, subject to key constraints of structure, sequence, outgoing degree and cost consideration. If all the pages of the Web site are connected, without allowing circles in the graph representing the Web site, the structure constraint then is a directed tree. A sequence constraint exists when a visitor must brose through some pages before reaching his or her target content(s). When a page by design cannot have more than 30 hyperlinks pointing to other pages, there exists a upper bound for outgoing degree. Page loading time can be an important constraint as well and therefore should be taken into consideration when estimating the cost of a page. A Web site presents search results in a systematic manner (e.g., in alphabetic order) can be modeled as a searching tree in which a visitor wants to access each page once and then return to starting page; i.e., a Travel Sales Problem (TSP).

As illustrated, many Web site designs can be modeled as graph problems. Some graph problems have already found their applications in Web site design analysis/evaluation, while others still need further explorations. On the other hand, some Web site design problems may not be examined using graph-based modeling, either. Further investigations are needed and can shed light on the potential boundary of graph-based modeling for Web site designs and, at the same time, point to some areas where graph theory might expand. An array of classic problem categories have been discovered and analyzed in graph modeling. Examples include spanning tree, routing, flow and cut, covering and partitioning, vertex ordering, isomorphic, sub-graph, and others. Different spanning tree problems have been examined, including minimal spanning tree (MST), capacitated spanning tree, degree constrained spanning tree, and maximum leaf spanning tree. Similarly, routing problems include shortest/longest path between two nodes, K^{th} shortest path, shortest path between one and all the other nodes, and traveling salesman problem (TSP). Some graph problems can be solved in polynomial time, while others are NP-hard. Heuristic methods can be developed for problems that cannot be solved analytically.

3.2. Generic Web Site Models

To describe and illustrate our framework, let us assume that a financial institution offers clients with Web access to information about various investment funds. The security requirement mandates that each access cannot have a back hyperlink, and that page caching is not allowed. Also observed is a constraint on the number of hyper links branching out from a page so that the page is crammed with too many links. Some pages have to be accessed following an immediate order. The primary objective for both the designer and the user's is the user's quick information access/retrieval. In this case, we can represent the problem as the following. The primary objective is minimized average access time. The described security requirement and disabled caching 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 the page access signals an immediate sequence relationship. We can represent the described problem using a 4-parameter (constraint) tuple (C,O,R,S), where C is the cost of each arc, O denotes out-going degree constraint on vertices, R represents sequence relationship between some vertices, and S describes structure constraint . Table 2 lists possible values of the constraints described.

Cost of Each Arc (C) represent the download time of a page, which is related to page size and network traffic. In a simplistic scenario, we assume each page require the same loading time; i.e., identical cost of arc. We further assume the cost of each arc is independent of direction. In the last scenario, each page has a different loading time; thus, a directed graph should be considered and backward link should also be taken into account. There exist other constraints worth considering. For instance, the cost of each arc should be constrained to present page from growing overly bulky and therefore taking much time to load. *Out-Going Degree (O)* shows the number of pages to which a page connects. Imaginably, there should be constraints on the upper and the lower bound of out-going degree. *Sequence Relationship (R)* describes the relationship between of two nodes, which can be immediately adjacent to each other, or be within some distance. When the distance is within 2, there exists at most one node between the two. *Structure Constraint (S)* depicts the structure of a graph. Tree structure is simple and easy to maintain; e.g., updating its links. A directed acyclic structure allows the user to reach the target content following different navigation paths. A directed cyclic graph allows the user to return to any of the previously visited pages.

3.3 Exploration and Integration

The incompleteness of the mapping, especially, between Generic Web Site Design Layer and Graph Modeling Layer is expected. For example, cache features can result in the "conditional" back link that once the link is traversed then the corresponding back link is generated. It is not straightforward to map the problem instance above to a matching or suitable graph models. On the other hand, similarly, some graph models may not have the counterpart in generic Web site models because most of the designers and users are more concerned about some fundamental or obvious requirements in the Web design and have not deemed or explored other possibilities in the design goal or constraints. The incompleteness of the mapping might create the potential side dish topics for exploration during the research.

The 3-layer framework provides the generalized site modeling and analysis. The 3-layer framework can be extended and linked with three major Web design issues – navigation guidance [14], assessment and improvement models [10], and design guideline as shown in Figure 1. The graph models derived from the Graph Model Layer can be applied to generate the optimal navigation guidance for both deterministic and stochastic problem instances. The assessment and improvement model based on the accessibility is to balance the accessibility and popularity (A-P) analysis [10]. Four accessibility models are introduced - expected link numbers, accumulated accessibility, sum of distance reciprocal, and sum of expected distance reciprocal [10]. The graph models derived from the Graph Model Layer can be extended as the selection criteria of appropriate accessibility model for Web different applications with specific requirements. Finally, 3-layer framework functions as the verification and validation toolkit for the structure-based Web design guideline for requirement analysis, performance evaluation and improvement criteria.

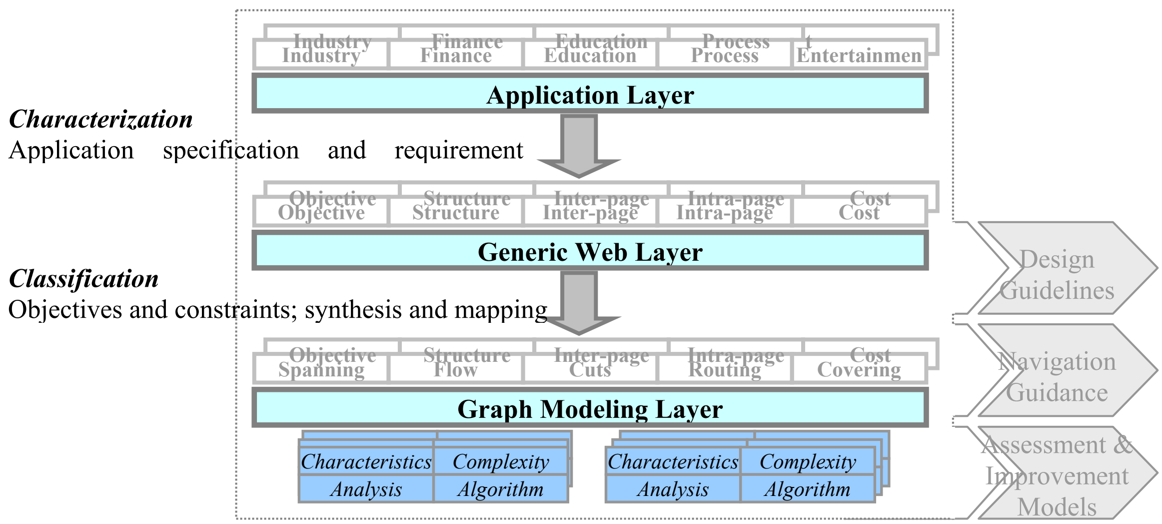

Figure 1. A 3-Layer Framework for Evaluating and Enhancing Web Site Designs

Table 1. Examples of Important Objectives and Key Constraints

			Objectives	Constraints
Designer	Structure	Site Level	Longest path for products ; #/cost of links minimization for maintaining & efficiency; etc.	The # of out links is constrained; constraints on back links and visiting precedence on order/distance; adapt to dynamic access behaviors; etc.
		Page Level	Contents can be minimized / updated automatically; etc.	Promotional product should be centered in the middle of the page; etc.
	Human Computer	Text	Readability should be acceptable [13] ; etc.	Text should be broken up into smaller units on multiple pages [11] ; etc.
		Link	Expose multiple levels of the architecture (link clustering with headings) [12] ; etc.	Use different forms for repeated links, e.g. text and image [7]. The number of links branching out from each page is constrained; etc.

		Graph	The download speed of the graph should be minimized.	Graph animation is irritating to users and it impedes scanning [13]; etc.
		Page	Text density should facilitate page scanning [12, 13]; etc.	The download speed should be no more than 10 seconds [11]; screen size should be specified [11]; etc.
User	Structure	Site Level	Minimize searching time, # of clicks and content access time to target contents; maximize accessibility of the page; etc.	The size of the web site (the # of pages or documents) should be appropriate for different applications, e.g. newspaper or different kinds of books [7]; related pages should be connected; constraints on height and width; etc.
		Page Level	Contents can be minimized and be updated frequently; etc.	The layout of pages should be consistent; the download speed of one page should be limited; the download speed of graphs should be minimized; etc
	Human Computer Interface	Text	Readability should be acceptable [13].	Text content is related between corresponding Web pages; etc.
		Link	Maximize the accessibility; Related links can be found in portal page of the site; etc	Use different forms for repeated links, e.g. text and image [7]. Use traditional methods to identify links, e.g. underline works, etc.
		Graph	Download speed of the graph should be minimized; etc.	The graph, especially for products should be attractive; etc.
		Page	The download speed should be no more than 10 seconds [11] ; etc.	The page should be accessible to people with disabilities; etc.

Table 2. Listing of 4-Parameter Constraint Tuple (C,O,R,S) – Example Values

	Cost of Each Arc (C)	**Out-Going degree (O)**	**Sequence Relationship (R)**	**Structure Constraint (S)**
0	Identical	No Constraint	No Constraint	No Constraint
1	Symmetrical	Constraint on Upper Bound	Immediate Sequence Relationship	Directed Tree Structure
2	Nonsymmetrical	Constraint on Lower Bound	At most one node between	Directed Acyclic Graph Structure
3		Constraint on Upper and Lower Bounds	At least one node between	Directed Cyclic Graph Structure

4. Illustration of Using Framework for 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 proposed an iterative refinement method for computing reasonably good sub-optimal solutions for large instances of constrained spanning tree problems [3]. 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:

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

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

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

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

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

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

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

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

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. We solve this example problem using LINGO. With respect to constraints (7), we set m at 2, the upper bound of outgoing degree of each node.

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 transform 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. However, 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 graph model or other mathematical models in addition to accessibility 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. To address this and other limitations, several areas need our continued investigations. First, the completeness of the accessibility model should be justified and extended in future. Second, empirical justification should be shown in classification of objectives and constraints. More real and complex example should be shown for illustration of using the accessibility model.

Reference

[1] S. Brin and L. Page. The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Proceedings of the Seventh International World Wide Web Conference, Australia,* April 1998

[2] R. Cooley. Automatic Personalization Based on Web Usage Mining. *Communications of ACM, Vol. 43, No. 8,* August 2000

[3] N. Deo and N. Kumar, Capacitated MST and Related Problems: Challenges in a Unified Approach, http://citeseer.ist.psu.edu/61074.html, April 09 1996.

[4] D. Dhyani, W. K. Ng, and S. S. Bhowmick. A Survey of Web Metrics, *ACM Computing Surveys,* Vol. 34, No. 4, pp. 469-503, December 2002

[5] Garey, M. R. and D. S. Johnson. *Computers and Intractability - A Guide to the Theory of NP-Completeness.* W. H. Freeman and Company, New York 1979

[6] L. Hull. Accessibility: It's Not Just for Disabilities Any More. *ACM Interactions*. Volume11, Issue 2; Column: The whiteboard Page:36-41, March/April 2004

[7] M. Y. Ivory and M. A. Hearst. Improving Web Site Design, *IEEE Internet Computing, Special Issue on Usability and the WWW 6,* no. 2: 56-63. March and April, 2002

[8] S. R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, and A. Tomkins. Random Graph Models for the Web Graph. *Proceedings of 41st IEEE Symposium on Foundations of Computer Science.* 2000

[9] Shneiderman, B. *Designing the user interface: strategies for effective human-computer-interaction*. Addison Wesleys Longman, 1998

[10] B. P.-C. Yen. Structure-Based Analysis of Web Sites, *Proceedings of IEEE International Conference on e-Technology, e-Commerce, and e-Service*, March 29-31, 2004

[11] J. Nielsen. *Designing Web Usability: The Practice of Simplicity.* Indianapolis, New Riders Publishing, 2000.

[12] Sawyer, Paul, R. Danca, and W. Schroeder. Report 6: Myths of page layout. *Designing Information – Rich Web Sites.* Bradford, MA: User Interface Engineering. 2000.

[13] Spool, T. Scanlon, W. Schroeder, C. Snyder, and T. Deangelo. *Web Site Usability: A Designer's Guide.* San Francisco, CA: Morgan Kaufmann Publishers, Inc. 1999

[14] B.P.-C. Yen and Y.W. Wan. 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.