Exploiting Link Types during the Web Site Design Process to Enhance Usability of Web Sites

SVEN CASTELEYN
Department of Computer Science, Vrije Universiteit Brussel
and
OLGA DE TROYER
Department of Computer Science, Vrije Universiteit Brussel

With recent developments in World Wide Web technology, expressiveness of web sites and web applications is starting to converge to the functionality offered by Open Hypermedia Systems. In this paper, we examine how link typing, and an appropriate link taxonomy can be used during the web site design process. Different existing approaches to link categorization are reviewed before going into closer detail to the concept of link and its properties. Finally, a useful taxonomy from a web design point of view is proposed. The presented taxonomy defines four categories of links, with a description of how each of them can be exploited to enhance the usability of the web site. Explicitly modeling the types of the links as belonging to one of these four categories during the design phase allows for better usability, multiple customized presentations, and other automatically generated enhancements to web sites (e.g. site map, in-page navigation structures, …).

Categories and Subject Descriptors: D2.2 [Software] Software Engineering - Design Tools and Techniques; H5.4 [Information Interfaces and Presentation]: User Interfaces - Hypertext/Hypermedia
General Terms: Design, Human Factors
Additional Key Words and Phrases: Link Taxonomy, Link Classification, Web Site Design Process, Usability

1. INTRODUCTION

Although the World Wide Web is undoubtedly the world's best-known application of a hypermedia system, it offers only restricted functionality of what a full hypermedia system could offer. The term "hypermedia" is commonly used to denote a collection of information elements interconnected by links, which can be browsed by the user to access (in some way) relevant information concerning the current link. The World Wide Web, for now still mainly a collection of html pages, offers the following features which make it worthy of the name "Hypermedia":

- **Interconnectivity**: interconnectivity between sources via an embedded link in the source document to the target document
- **Heterogeneous links**: interconnectivity not only between text fragments, but also between other elements (e.g. image linking to text fragment); this is a feature of a hypermedia system, in contrast to a hypertext system, which only provides connectivity between text fragments.
- **Non linearity**: there is no a priori set order of traversal, the choice of navigation path lies completely with the user

Shortcoming of the World Wide Web regarded as a hypermedia system include:
• **No content-navigation separation**: in classical web sites the data and the navigation through that data are entangled and represented in one source. Even when underlying databases or other data sources are used, the client receives database content, document data and navigation all merged into one (html) document. Links are not considered first class objects; they are embedded in the data source.

• **No bi-directional links**: links are currently one-directional: they are embedded in the source, and make a connection from the source to a target. The target is not aware of the existence of the link, the designer of the target link is not aware of the consequences when he manages and changes his web pages. Apart from the uni- or bi-directionality, there is also no possibility to define more complex link structures, such as links with multiple endpoints.

• **No link typing**: links can only link (a fragment of) the source to another source (URL). There is no possibility to qualify the link with any additional information (e.g. semantic information, type information). Link type information could be useful to enhance usability of web sites in various ways; we will investigate this notion further in this paper.

• **No link constraints**: as a result of the lack of separate link management, and the lack of typing possibility, it is also difficult to specify constraints for certain links. Constraints could add to the ability for web site designers to customize the link being presented to the visitors.

• **No customizability**: although users do have control over the order of traversal, they do not have a lot of control over the hyperlink structures themselves: they cannot add or modify existing links, nor manipulate their presentation or its effect on frames or windows.

Note that html actually does provide some support for link typing, namely the “rel” and “rev” attributes of a link. The “rel” attribute permits to name the associated relationship for the current document A to some source B, and a “rev” attribute denotes the reverse relationship (i.e. from B to A). The Mozaic browser exploited these attributes by recognizing certain values specified in the HTML standard (e.g. home, ToC, Up, Next, …) and use this information to build a navigation sidebar. However, these tags seem to be widely unknown by the general public, and subsequently they are rarely used, and there is no standard support for it.

There are systems that do implement the features the World Wide Web currently lacks. Open Hypermedia systems are designed to manage the links separate from the data. As databases successfully separate data and programming logic, “link bases” seek to separate the data, and the connections between that data, the links. Such a link base may provide automatic support to manage the resources for the Hypermedia System, and the links between them, and it provides the ability for third parties to offer their own link bases for this and other Hypermedia Systems. Furthermore, these systems do offer bi-directional linking, and some Open Hypermedia Systems also implement simple constraint systems on links. For example, temporal information can be specified as a link constraint. Adaptive Hypermedia Systems may exploit link types and resource identification to customize the navigation for the end user.

---

1 For example, Microsoft Internet Explorer 5.0 and higher uses (or may we call it abuses) the “rel” attribute when storing favourites to denote a particular favourite is a locally stored cached version, by using the "offline" value.
Currently, several efforts are being made to enhance the World Wide Web, with the purpose of adding some of the missing Hypermedia functionality mentioned above.

The Extendible Markup Language (XML) was a W3C effort (Bray et al. [2000]) to provide (standard) extendable markup abilities for web pages, thus offering enhanced functionality and content. Although XML does provide easier means of syntactic automated document analysis, it is still insufficient for the simple fact that it does not provide any semantics, and subsequently automated use of the XML tags is still difficult. Most web servers on which websites run that are using XML as underlying document description language, translates these pages into HTML before sending them to the client (as opposed to sending the XML source itself), so the advantage of using markup is lost for the client.

The Resource Description Framework (RDF) (Lassila et al. [1999]) is a W3C attempt to describe the resources on the web, by providing a framework to create, exchange and use meta data. As XML, RDF merely provides syntactical markup; it contains no semantic information. The Dublin Core however, is a semantically rich standardized set of elements, describing several aspects of a range of network resources. RDF provides a perfect framework to express the Dublin Data Core elements. Next to the Dublin Core, there are several ontology projects ongoing, attempting to express semantics using RDF (e.g. OIL, Fensel et al. [1999], DAML, Scot et al. [2002], Ontobroker, Fensel et al. 2000, and others).

The quiet recent XLink language (XLink) (Derose et al. [2001]) from W3C aims on explicitly representing the relationship between (fragments of) resources, along with its characteristics (attributes) as metadata. The XLink language also supports the ability to store links separated from the resources it links, and means to add behavior to the links themselves (instead of leaving the interpretation of behavior to the user client). We could say the XLink language promotes links to first class object, thus alleviating the problems associated with this (see above).

With these efforts, the expressiveness of World Wide Web is slowly converging towards the functionality offered by more general (Open) Hypermedia Systems. It is time to investigate if our current web design methods are able to exploit this expressiveness. We will focus our attention on link types and link taxonomies. In the current WWW, intent and meaning of links are left largely to the user to derive from the context the link appears in. Link typing, and exploitation of link types when presenting the web site, might help users to better grasp why a link is there, where it goes to and what purpose it serves (e.g. reduce the "lost-in-hyperspace syndrome"). Furthermore, a robust and purpose oriented link taxonomy would enable the possibility of (automatic) filtering of links (for example, for certain users), selectively presenting certain links, adapting presentation of certain link types, etc.

Most web design methods are not able, or only have limited capabilities to express different types of links during the design of a web site. Therefore, it is important to study

---

2 XML provides the default syntax to encode RDF descriptions
3 Version 1.0 became available in July 2000
will investigate the role different types of links can play in the design process of a website to improve the usability of the website. We will give a first taxonomy of link types that may be of importance for the web designer. Note that although some research has been done in link taxonomies (see section 2), no specific research has been done to explore how such taxonomies can be exploited during the web site design, or how specialized taxonomies can be build for the web site designers use.

The rest of the paper is organized as follows. In section 2, we will review some of the link taxonomies that have already been proposed in the literature (in the field of Hypermedia Systems). In section 3, we define the concept of link and some of its properties that will be used in the rest of the paper. In section 4, we propose a first taxonomy of link types that may be of importance for the web designer and explain how these link types may influence the design process and the usability of the web site. Section 5 presents conclusions.

2. EXISTING LINK TAXONOMIES

One of the earliest link taxonomies was made in the context of Textnet (Trigg [1983]), a hypertext system to support the online scientific community in text creation and annotation, created by Randall Trigg in 1983 as a part of his Ph.D. dissertation (Trigg [1983]). Trigg makes a first subdivision of the links in two categories: normal links and commentary links. Normal links serve to connect nodes making up scientific work(s), while commentary links connect statements about a node to the node in question. These two categories are further subdivided into subcategories, of which some are again subdivided. Although some of his link types are generally applicable, most of them are mainly focused on the particularities of scientific writing (e.g. the "P-ill-posed" type is used to denote that the author badly stated his problem; the E-Isupersede is used to denote another ignored work supersedes this work) and are thus not applicable for general use.

Baron et al. [1996] argue in their study of online hypertext manuals that the use of typed links increases the usability of the hypertext manual in terms of speed of information discovery. At top level, she also subdivides the link types in two categories: organizational and content-based links. Organizational links mainly serve navigational purposes; content-based links deal more directly with the relationships between nodes. The latter category is subdivided into semantic, rhetorical and pragmatic links. Semantic links describe semantic relationships between nodes (Baron describes three semantic link types: similar, contrast and part/kind of); rhetorical links are meant to lead the reader through a series of information elements to achieve a learning goal supporting a task; pragmatic links link to practical results (e.g. a database error).

Richard Kopak (Kopak [1999]) aims to develop a taxonomy of link types that describe the function that characterizes the formation of links. With his research, he also wants to argument that a link taxonomy should be empirically derived, rather than set a priori.

Chip Cleary and Ray Bareiss (Cleary et al. [1996]) use a set of link types inspired by a simple theory of conversation. The set of link types used in their ASK system is based on so called conversational associative categories, general classes of questions a person will
probably ask in a conversation. The eight link types most commonly used are: context, specifies, analogies, alternatives, causes, results, opportunities, warnings. Some methods for automatic linking are described in Cleary et al. [1996].

Several attempts have been made to extract links and link types automatically from sources. From this viewpoint, James Allan (Allan [1996]) identifies three link categories, purely on basis of the ease to extract links automatically from documents: manual, pattern matching and automatic. Manual links are links that cannot be extracted without human intervention; pattern-matching links are links that are easily found using pattern-matching techniques; automatic links are links that cannot be found using pattern matching, but are automatically derivable using more advanced techniques (see Allan [1996]). He further specifies a subdivision of this latter category. John Tebbutt reported automatic (to some extend) generation of some "automatic" links (structural and referential links) in the LEIDIR (Link-Enhanced Information Discovery using Information Retrieval) System (see Tebbutt). Support and analysis for other link types was future work (see Tebbutt). Other researchers are committed to the same goal.

In the digital library community, Hansen et al (see Hansen et al. [1999]) attempt to combine the WWW, digital libraries and Open Hypermedia. The Webwise prototype implements a semantic link type system, where the semantics of the link consist of a set of methods and attributes guaranteed to exist for this type. Users are able to add new link types.

Given the explosive amount of (unstructured) information already present on the WWW, the extraction of links and link types from existing sources is a vital activity. But as already explained in the introduction it is of equal importance to study how link types and link taxonomies can be incorporate in the design process of web sites, thus providing improved links from the moment the web site is created. Despite this importance, few publications can be found on this topic.

In the HDM method (Garzotto et al. [1993], three different kind of links are distinguished: structural links (connect together components belonging to the same HDM-entity), perspective links (connect the different units corresponding to the same component) and application links (arbitrary, domain dependent relationships connecting components and entities); the latter category is grouped in link types. Advantages include automatic generation of perspective links and part of the structural links, by exploiting the implicit link information already present in the HDM entities.

WebML (Ceri et al. [2000]) only distinguishes between contextual and non-contextual links. The former are used when the content of the target depends upon the source of the link, while the latter is used to connect semantically independent pages.

We also note the attempt to categorize navigation links in the OO-H method (Cachero et al. [2001]) into six link types: I-Links, T-Links, R-Links, X-Links, S-Links and their corresponding R-Link4. Definition of these link types are closely coupled with the OO-H method: I-links (Internal Link) defines a navigation path inside a Navigation Target (NT); T-links (Traversal Link) are defined between navigation classes of different NT’s; R-links

4 Although confusing, this corresponding R-Link (Response Link) is not the same as the R-link mentioned before, which is a Requirement Links)
(Requirement Link) denote a starting navigation point inside an NT; X-links (Exit Links) are placed outside the boundaries of the application; S-links (Service Link) show the services available to the users associated with a certain Navigation Access Diagram, and the corresponding R-links (Recovery Link) when the interface recovers the control from the application.

3. LINKS & LINK PROPERTIES

Before we are able to classify links, we first have to define the concept “Link” as we will use it further on. We give a very generic definition, so that the concept is generally applicable. A link will connect nodes. A node is a generic name for any identifiable piece of information of any granularity, i.e. it might be a complete website, a document within a website, a fragment within a document, a sound sample, … any piece of information that can be uniquely identified.

Definition
A link connects two or more nodes. A link is either uni-directional or bi-directional.

- A uni-directional link points from one or more source nodes to one or more target nodes.
- A bi-directional links points from one or more nodes to one or more nodes and back. The role of source and target is interchangeable in this case.

For the sake of simplicity, we will also talk about source and target in case of a bi-directional link. Which of the two sets of nodes is source and which is target can be chosen arbitrarily.

Note that this definition corresponds with the notion of link used in XLink. XLink allows “extended links. links that are not limited to one source and one target, but possibly connect any number of resources. Subsequently, as we speak of ‘arity’ of relations in data modeling, it also makes sense to speak about ‘arity’ of links:

- One-to-one: the link connects exactly two nodes.
- One-to-many: the link connects exactly one source node to more than one target node
- Many-to-one: the link connects more than one source node to exactly one target node.
- Many-to-many: the link connects more than one source node to more than one target node.

We note one special case, the reflective link:
- A link is reflective if it is one-to-one and if the source and target node are the same.

In addition to the general definition of a link, we also define two special cases that will give more expressive power to our link concept.

Definition
A conditional link is defined as a link with an associated condition C. The link of a conditional link is only available if the condition C associated with the link is satisfied (i.e. its truth value is TRUE).

Conditional links are very useful to model certain dynamic aspects in web sites. Using conditional links it is possible to indicate when a link must be generated or not (for some elaborated examples of the use of conditional links see De Troyer et al. [2001]).

Note that in Open Hypermedia Systems using links externalized from the actual data (i.e. link bases), all links can be conditioned, not only by the web site designer, but also by third party clients. This allows for both client and server side customization of the links offered on a website.

According to the type of condition that is used, we can subdivide the conditional links into subcategories. The most commonly used subcategories are:

- **Temporal** link: conditional link where the condition is a time constraint. Temporal links can be used to constraint the availability of a link to a time frame. E.g. a link is only available for three weeks, or after a certain date a link becomes unavailable.

- **Personalized** link: link where the condition allows testing on the identity of visitor of the web site. The identity might correspond to one single user, or to a group of users with similar requirements.

From a technical point of view, next to being able to condition a link, it may also be useful to indicate if a link is adaptive or not. Adaptive links are needed in adaptive web site where we need to be able to specify which links are adaptive and which are not. For example, if some node deeply nested in the web site hierarchy is consulted frequently, it may be decided to place the link directly on the home page, in which case the source of the link has to be modified. Similarly, it might be necessary to change the target of a link.

**Definition**

A link is called an adaptive link if it is allowed to replace the source and/or the target of the link by other nodes.

### 4. A PURPOSE ORIENTED LINK TAXONOMY

We are now ready to investigate which link types are useful to distinguish during the web site design process and how these link types can be used to improve the design process and to enhance the usability of the later web site.

Links can be classified from different viewpoints resulting into different taxonomies. E.g. in the previous section we have given a classification based on the “technical” properties of the concept: one or more target or source nodes, target and source equal…The classification that we propose here is based on the purpose of the link. Links can be introduced during the design process of a web site for different purposes or reasons: to structure the information in the web site into pieces, to ease navigation (e.g. a “home” link); to express the existence of a (semantic) relationship between two concepts; to call an application; … Based on this criteria, we distinguish the following types of links:
structural links, navigational links, semantic links and process logic links. Each type of link is described into more detail in the following sections.

### 4.1 Structural links

**Definition**

A *structural* link connects two or more nodes in order to define a structure between the nodes.

A web site usually contains a lot of information and possibly also functionality. All this must be structured in some way before it can be presented to the users. In a web environment this is done by grouping information and functionality into pieces and then linking these pieces together, e.g. in a hierarchical structure. The totality of all the pieces and the links between them is a structured representation of the information and functionality being offered. We call these links structural links. The main purpose of these links is to define a structure for the web site. As an example figure 1 defines a possible structure for a web site for a research group. All links shown in this figure are structural links.

![Figure 1: Simple example structure for a research group web site](image)

In Casteleyn et al. [2001], an algorithm to derive the main (hierarchical) structure of an audience driven web site on basis of a simple step-by-step matrix construction is presented.

Typical for structural links is that they only define a possible structure; different structures for the same web site may be possible. In an adaptive web site this can be realized by making the structural links adaptive. Well-known hypertext organizational structures are linear, hierarchical, pure web and grid (dual linear structure). A linear structure can be expressed by means of one-to-one structural links, a hierarchical
structure by e.g. one-to-many structural links. Structural links can be compared with the “organizational” links introduced in Baron et al. [1996].

Special cases
A special case of a structural link is the well-known next/previous link used to navigate into a large collection. E.g. the result of a search action (in a search engine) may be far too large to be display in a single window (also the user is probably not interested in all results); therefore the result is broken down into pieces and structured in a linear way. In this case the structural links are one-to-one bi-directional links.

Impact on design process and the usability of the web site
Explicit description of structural links in the design process enables the possibility to generate the linking structures defined by the structural links according to default templates, or even customized at runtime according to type of user or type of browsing device. In case the structural links define a linear structure “next” and “previous” buttons can be generated; in case of a hierarchical structure, appropriate (GUI) menus for each level in the hierarchy can be created automatically. In the latter case, one might as well opt for a graphical tree representation for certain users. I.e. there is a better abstraction of the definition of structure between the data, and the actual browsing strategy defined by this structure in the implementation. Structural links can also be used to automatically generate a site map.

4.2 Navigational links
Definition
A navigational link connects two or more nodes in order to define a navigation path between the nodes.

Links of this type are usually added on top of the structure defined (by means of the structural links) for a web site to ease the navigation and to enhance the usability of the web site. Of course structural links are ultimately also used to navigate through the web site, but this is merely a side effect of the fact that in the Web hyperlinks are used to implement the structure of a web site.

Typical for navigational links is that usually we are not a-priori concerned with this type of links during conceptual design. Usually, they are added later on to ease the navigation and to enhance the usability of the web site. An example of a navigational link is the home link. Some navigational links may even only be introduced during implementation (design), e.g. a “top of page” link is only introduced if it turns out that a page contain much more information than viewable on one screen. Navigational links can be compared to shortcuts or accelerators used in most graphical user interfaces.

Special cases
Typical examples of navigational links are the “home” link, “landmark” links, links used inside a site map to directly navigate to a page, the link to a site map. A home link is a typical many-to-one link.

Impact on design process and the usability of the web site
The navigation links put a second navigation layer on top of the navigation layer defined by the structural links. Mixing both types of links in the visual presentation of a web site
may confuse the visitor; there may be different ways to reach the same information. This can hinder the user in the creation of a correct mental model of the structure of the web site and induce the “lost-in-hyperspace” syndrome. If we have the possibility to distinguish between structural and navigational links at an early stage, we will also be able to represent them differently in the web site (e.g. separate navigation pane, extra frame, in-page links, …). In this way, it will be easier for the user to build a correct mental model of the structure of the web site. As for structural links, their early identification may also lead to runtime customization (for example, a web site on a PDA will typically require more in-page navigation).

4.3 Semantic links

Definition
A semantic link connects two or more nodes to express a semantic relationship that exist between the concepts represented by the nodes.

Next to structural and pure navigational links we also have links in a web site that are derived from relationships that exist between concepts in the universe of discourse. E.g. in a university web site we have links that allow the visitor to navigate from a course page to a lecturer page because there exist a (semantic) relationship between courses and lectures, i.e. “is taught by – is teaching” relationship. Similar as for structural links, semantic links will also be used to navigate through the web site, but again this is a side effect of the fact that hyperlinks are used to implement these semantic relationships in a web environment.

Because semantic links are based on relationships between concepts in the universe of discourse, there is not much freedom in defining them. Except for deciding to make the link available or not, the rest is fixed; i.e. they cannot be adaptive (we cannot change the source and the target of the link freely without changing the underlying semantics of the link). Usually a semantic link is also a bi-directional link.

Special cases
Special cases of semantic links include explanation links, example links, and reference links. An explanation link associates a more detailed explanation with a concept (the underlying semantic relationship is “is explained by – explains”); an example link associates one or more examples with a concept (the underlying relationship is “with example - is example for”); and a reference links associates one or more references to a concept (the underlying relationship is “with reference - is reference for”). More of these types of links are of course possible.

Another special case of a semantic link is the link between a node that represents a collection of concepts (e.g. all courses) and the node that represents one instance of this collection. For example, it is common to allow the user of a web site to click on each of the instances of a collection to obtain more information about an instance. The hidden (implicit) underlying relationship here is the “contains – is element of” relationship.

Impact on design process and the usability of the web site
As nodes are uniquely identifiable pieces of information, it does not really matter how they are stored internally (underlying database, XML elements, RDF resources, …), and how the relationships between them are stored (conceptual database schema, RDF statements, …). As long as we are able to link these pieces of information to the
modeling entities used in the conceptual web site design, the underlying semantic relationship(s) may be exploited (e.g. added automatically).

Semantic links can also be used to automatically generate a kind of “semantic web map” of a web site. Opposed to the usual web map (which reflects the structure of the web site) a semantic web map reveals the semantic relationship between the concepts in a web site (more on this issue can be found in De Troyer et al. [2000]).

4.4 Process logic links

Definition

A process logic link connects two or more nodes to express a part of a workflow or an invocation of an (external) functionality.

Web sites that provide some functionality usually need to express some workflow. For example to order some book first you have to select the books and add them to the shopping basket, then you enter your coordinates, then you have to select the payment method and depending of the select payment method you have to provide some information like your credit count details, finally you may receive an order number to be able to trace your order later on. Usually, each individual step is represented as a node and the process logic is indicated by means of links. As these links represent a part of a workflow; we call them process logic links. To represent choices in the process logic we can use conditional (one-to-many) links. Also the links hidden behind buttons like “submit,” “login” or “send e-mail” are process logic links; they invoke some functionality and are in fact links between two nodes of which the target represents some (external) functionality.

Special cases

As special cases of process logic links we can mention: a retry link, an action link (e.g. submit/login), next step/back to previous step link, undo link, restart link, exit/abort link, error/warning link. A retry link is an example of a reflective link.

Impact on design process and the usability of the web site

Identifying process logic links will allow us to keep track of user transactions and to visualize this towards the user of the web site, i.e. we can allow the user of our bookshop web site to suspend the ordering of books for a moment (e.g. for looking back to an abstract or compare the book with some other book) and later on allow him to return to the ordering process. Looking for abstracts of books and comparing books are not part of the process logic and therefore they should not be modeled by means of process logic links. If the user starts following process logic links this could be visualized in the web site remembering him that he has some unfinished workflow when he takes some sidetracks. For this reason it may also be useful to label a process logic link with a process-id and to indicate the beginning and (possible) end(s) of the workflow. Also a more consistent user interface for a web site can be (automatically) generated if the meaning of links is known. E.g. it would be sufficient to specify once that a retry link will be represented as a specific button, to have all retry links in the web site represented in the same way.
5. CONCLUSIONS

In this paper we have motivated the need to distinguish between different types of links during the design process of a web site. Important factors for this are the growing expressiveness of the web technology and the possibilities typing offers to enhance the usability of web sites. First we have defined a link as a connection between two or more nodes. A node can be anything. This very generic definition of a link allows us to discuss and classify links without being limited to a certain method or context. Next we have given some of its properties and defined two special types of links: conditional links and adaptive links. These types of links have more expressive power and are useful to ease the modeling of complex web application such as dynamic web sites and adaptive web sites. Based on this general link definition, a useful taxonomy from a web design point of view is proposed. The classification is done from the viewpoint of the purpose of the link. If we investigate the reasons for defining links during a web site process, we distinguish four different types of links: structural links, navigational links, semantic links and process logic links. Structural links are used to structure the information and functionality in a web site; navigational links are introduced to provide navigational aids to the user of the web site; semantic links express semantic relationships that exist between concepts in the universe of discourse of the web site; and process logic links express parts of workflows. For each of these types of links we have discussed the role that can play in the design process and how they may influence the usability of the web site.

REFERENCES


