Published in: International Journal of Web Engineering Technology, Vol 1, No. 1 (2003)

Exploiting Link Types during the Conceptual Design of Web Sites

Olga De Troyer and Sven Casteleyn
Department of Computer Science, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussel, Belgium
Olga.DeTroyer@vub.ac.be, Sven.Casteleyn@vub.ac.be

Abstract. 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 definition 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. Next, four different types of links are defined, with a description of how each of them can be exploited to enhance the design process as well as the usability of the web site. Explicitly modeling the links as belonging to one of these four categories during the design phase allows to separate concerns during the modeling process; provide for clearer design guidelines; allows for better usability, multiple customized presentations, and other automatically generated enhancements to web sites (e.g. site map, in-page navigation structures, …). The approach is presented in the context of WSDM, an audience driven web site design method, but the principles are generally applicable.

Keywords: link types, web site design, 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 hyper*media* system, in contrast to a hyper*text* 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[1].

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.

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 [3] 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 web sites 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) [19] is a W3C attempt to describe the resources on the web, by providing a framework to create, exchange and use meta data. As XML[2], 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 [13], DAML [20], Ontobroker [12], and others).

The quiet recent[3] XML Linking Language (XLink) [9] from W3C aims on explicitly representing the relationship between (fragments of) resources, along with its characteristics (attributes) as

[1] For example, Microsoft Internet Explorer 5.0 and higher uses (or should 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.
[2] XML provides the default syntax to encode RDF descriptions
[3] Version 1.0 became available in July 2000

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.

Here, we will focus our attention on different types of link. In the current WWW, links are the primary user interface elements, and they are used for different reasons and purposes. They may lead the user to other (internal or external) pages, provide in page navigation, invoke (internal or external) functionality, trigger database queries, cause file download, cause information to be sent, etc. Clearly, "links" are heavily overloaded, and it is left largely to the user to derive *intent* and *meaning* of links from the context the link appears in. Using different types of “Links” when designing a web site and exploiting these 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, it 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.

In this paper we will investigate the role different types of links can play in the design process of a web site to improve the design process as well as the usability of the web site. 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.

As a framework to present our work, we will use WSDM (Web Site Design Method) [5] [6] [10] [11], an audience driven web design method. Nevertheless, the ideas presented are in principal generally valid, and could be applied to any web design method.

The rest of the paper is organized as follows. In the section 2, we will review some of the link types and link taxonomies that have already been proposed in the literature (mainly in the field of Hypermedia Systems). In section 3 we will provide the reader with a short overview of WSDM. In section 4, we define the concept of link and some of its properties that will be used in the rest of the paper. In section 5, we explain the link types that WSDM distinguishes during its conceptual design phase and in section 6 we explain how these link types may influence the design process and the usability of the web site. Section 7 presents conclusions.

2 Existing Link Taxonomies

One of the earliest link taxonomies was made in the context of Textnet [22], 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 [22]. 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 denote another ignored work supersedes this work) and are thus not applicable for general use.

Baron [2] argued in her 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 [18] 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 [8] 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 [8].

Several attempts have been made to extract links and link types automatically from sources. From this viewpoint, James Allen [1] 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 [1]. 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 [21]. Support and analysis for other link types was future work [21]. Other researchers are committed to the same goal.

In the digital library community, Hansen et al [17] attempt to combine the WWW, digital libraries and Open Hypermedia. The Webvise 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 [14], 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 [7] 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 [4] into six link types: I-Links, T-Links, R-Links, X-Links, S-Links and their corresponding R-Link[4]. 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 (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

[4] Although confusing, this corresponding R-Link (Response Link) is *not* the same as the R-link mentioned before, which is a Requirement Links)

with a certain Navigation Access Model, and the corresponding R-links (Recovery Link) when the interface recovers the control from the application.

3 WSDM: An Overview

Here, we give a short overview of WSDM, more information can be found in [5] [6] [10] [11]. The two basic characteristics of WSDM are the audience driven approach, and the explicit conceptual design phase. The first characteristic puts the emphasis of the design method on the different kind of users, and their different needs and requirements. The design process takes into account these different audiences from the start, and let their requirements drive the entire process. The second important characteristic is the explicit conceptual design phase, free of any presentation or implementation detail, yet being expressive enough to model different information requirements and functionality for the different users. Moreover, different presentations for different users and/or platforms are possible.

In figure 1 an overview of the WSDM method is given. The first step is to define t*he Mission Statement.* The Mission Statement should express the purpose and the subject of the web site and declare the target audience. Based on this Mission Statement a two-step Audience Modeling phase is performed. In the first step, Audience Classification, the different kinds of users are identified and classified. Members of the same Audience Class have the same information and functional requirements. In the next step, Audience Class Characterization, the characteristics of the different Audience Classes are given. The result of the Audience Modeling is a set of *Audience Classes*, ordered into an *Audience Class Hierarchy*, together with an informal description of all their requirements: information- and functional - as well as navigational- and the usability requirements, and their characteristics. For a formal definition of Audience Classes, and a method to automatically derive them, we refer to [5].

Next, we perform a Conceptual Design of the site. The Conceptual Design phase is divided in three steps: Information Modeling, Functional Modeling and Navigational Design. These steps don't need to be performed sequentially. During Information & Functional Modeling, the information & functional requirements of the different Audience Classes are modeled. Each requirement is modeled by first decomposing the requirement into so-called *elementary requirements*; next for each elementary requirement a *Chunk* is created that models the information needs described in an elementary information requirement or the task step needed to support an elementary functional requirement. To control the redundancy that we may introduce in this way (different Chunks may deal with the same information), all Chunks are integrated later on into a single information model, called the *Business Information Model.*

During Navigation Design we design the conceptual structure of the web site and model how the members of the different Audience Classes will be able to navigate through the site. For each Audience Class an *Audience Track* is created. Navigational requirements are taken into consideration in this step. All Audience Tracks together form the *Conceptual Structural Model* of the site. The integration of the Information Chunks and Functional Chunks in the Conceptual Structural Model, together with all other models developed during Navigation Design (see later on) is called the *Conceptual Schema* of the web site.

During Implementation Design we essentially design the page structure as well as the 'look and feel' of the web site. The aim is to create a consistent, pleasing and efficient look and feel for the conceptual design made in the previous phase by taking into consideration the usability requirements and characteristics of the Audience Classes. The design of the page structure starts from the Conceptual Structural Model. The Implementation design should also provide the specification for the logical data design. This could be a logical database design, an XML DTD, RDF definitions, or any other suitable data definition format. Whatever format chosen, the data design can be derived from the Business Information Model.

The last phase, Implementation, is the actual realization of the web site using the chosen implementation environment.

4 Definition of Link

We first define the concept “link” as we will use it further on. We realize that the term “link” is heavily overloaded in the context of the WWW, but we use it here in it pure sense as “means of connection” (from The Oxford Dictionary). We give a very generic definition. In this way the concept is generally applicable. A link connects two or more nodes. A node is also a generic concept, it may be any identifiable piece of information of any granularity (i.e. a complete web site, a document within a web site, a fragment within a document, a multimedia element, a service, …) as well as an abstract concept used during modeling. Note that some people think of "links" as pure navigational links, and thus do not consider a link to invoke functionality or a menu item as a "link". Others also consider a link to invoke functionality or a menu item as a "link". In our opinion, this confusion is the result of the concept of "link" being heavily overloaded and used with different meanings by different people on one hand, and the usage of (hyper)links for different purposes (e.g. navigation, functionality, semantic relationships, …) on the other hand. In this paper, we will try to shed some light onto these issues by pointing out the different uses of (hyper)links.

Definition
Let N be a set of nodes, then a *link* l is a 2-tuple (S_1, S_2) where

$S_1 \subseteq N, S_1 \neq \emptyset$ and $S_2 \subseteq N, S_2 \neq \emptyset$

A link is either *uni-directional* or *bi-directional*.

- A uni-directional link points from one set of nodes, called the *source* to one set of nodes, called the *target*. The source of a link l is noted source(l) and the target is noted target(l).
- A bi-directional links points from one set of nodes to one other set of 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 to the notion of link used in XLink. XLink allows “extended links”, links that are not limited to one source and one target node, 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. In the following definition the symbol # is used to denote the cardinality of a set (as in standard set theory).

Definition
Let N be a set of nodes
A *link* $l = (S_1, S_2)$ where $S_1 \subseteq N$ and $S_2 \subseteq N$ is

- *One-to-one*: if $\# S_1 = 1$ and $\# S_2 = 1$
- O*ne-to-many*: if $\# S_1 = 1$ and $\# S_2 > 1$
- M*any-to-one*: if $\# S_1 > 1$ and $\# S_2 = 1$
- *Many-to-many*: if $\# S_1 > 1$ and $\# S_2 > 1$

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. The first one is a conditional link; the second one is an adaptive link.

Definition
A conditional link is defined as a link where each node is associated with a condition.
Let N be a set of nodes, C a set of Boolean conditions then a *conditional link* l is a 2-tuple (SC_1, SC_2) where

$SC_1 \subseteq N \times C, S_1 \neq \emptyset$ and $SC_2 \subseteq N \times C, S_2 \neq \emptyset$

The link from a source node or to a target node of a conditional link is only available if the condition associated with the node is satisfied (i.e. its truth value is TRUE).
The condition TRUE can be used if no condition is needed.

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 [10]).

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 web site.

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

- *Temporal* link: conditional link where the conditions are 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.
- P*ersonalized* link: link where the conditions allow 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. See [6] for a more elaborated explanation on this subject.

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.

In the rest of the paper we frequently need to represent links graphically. Therefore we introduce a graphical representation for a link (see figure 2). No graphical representation for a node is given because this is a generic concept. Its representation is dependent on the concept used as node.

5 Links in the WSDM Conceptual Modeling Phase

We are now ready to investigate the link types WSDM considers during its Conceptual Design phase. The examples used to illustrate the concepts are mostly taken from the Conference Review system. See [10] for a description of this case study. In section 6 we will investigate how the link types introduced here improve the design process and how they can be used to enhance the usability of the later web site.

5.1 Link Types

5.1.1 Structural Links

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. In WSDM this structuring of information is first done at a conceptual level and results in the Conceptual Structural Model. In this model, the *conceptual structure* of a web site is expressed by means of so-called *structural links* between *components*. Components represent units of information or functionality (tasks that can be done). In the Conceptual Structural Model units of information and functionality are represented by means of components. The components are linked according to a certain structure[2]

[2] In WSDM this structure is defined by the structure of the Audience Class Hierarchy and is therefore a hierarchical structure.

(e.g. a hierarchical structure or a linear structure). An example of this is given in figure 3. This figure shows a possible structure for the Conference Review System web site. Components are graphically represented as rectangles.

Please note that components do not necessarily correspond with the actual pages in a web site. It is only in a later phase (during Page Design) that components are assigned to pages. Depending on the size of a component, different components can be grouped on one page or a component can be distributed among pages. The choice of this grouping is left to the designer, and can be different depending on the targeted device. As a consequence of this, some of the conceptual structural links may in fact disappear at the implementation or become in page links.

The main purpose of the structural links in WSDM is to define a conceptual structure for the web site. In [5], 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. In section 5.2.1 we will give this structure for the Conference Review System web site. As you will see this is a different structure than the one defined in figure 3. This is a typical characteristic of structural links; they only define a possible structure; different structures for the same web site may be possible. 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 by Baron in [2].

The structural links may be uni- or bi-directional, conditional and/or adaptive. Figure 3 shows an example of a conditional link.

5.1.2 Semantic Relationship Links

Next to the structural links, WSDM allows to express a second kind of link during Conceptual Modeling, the so-called *semantic relationship link*. Semantic relationship links are based on relationships that exist between concepts in the universe of discourse. E.g. in the Conference Review System there exists a (semantic) relationship between papers and reviews (i.e. 'Paper with/for Review' relationship) and as a modeller we may wish to express that it must be possible to navigate from a paper to its reviews and from a review to the paper. In WSDM, we can model this by means of links between Chunks. As already explained, Chunks model pieces of information or functionality and are defined during Information and Functional Modeling. Chunks model the elementary information and functional requirements of the users. In WSDM, this type of modeling is done using an extended version of ORM (Object Role Modeling) (see [16] for ORM and [10] for the extensions defined for WSDM). ORM uses object types; relationship between object types; and IS-A subtypes to model information in the Universe of Discourse. Because Chunks model the semantic relationships between concepts in the Universe of Discourse, they can be used to express the semantic relationship kind of links. In figure 4 two Chunks are given. Chunks are graphical represented as rounded rectangles. The left hand Chunk represents the Chunk that models the information need by a PC-chair about a paper, and the right hand one represents the Chunk that models the information needed by a PC-chair about a review. One of the usability requirements for a PC-chair is that it must be easy to navigate from a paper to its reviews and back. This is modeled by means of a bi-directional semantic relationship link between the two Chunks. In the figure, the link is labeled with the name of the semantic relationship used for this link. A more detailed picture is given in figure 5. In this figure the object types and relationships inside the Chunks are shown and the semantic relationship link is connected to the specific relationship used for this link.
Semantic relationship links inside a Chunk are not very useful. A Chunk represents an elementary requirement or functionality and therefore all the information modeled in a Chunk will always be available together to a web user, there is no need for a semantic relationship link.

Because semantic relationship links are based on semantic 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 relationship link is a bi-directional link but the modeler can restrict this.

5.1.3 Process Logic Links

Web sites that provide 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 card details, finally you may receive an order number to be able to trace your order later on. In WSDM, the individual steps are modeled by means of functional Chunks. The workflow or process logic is added by connecting these functional Chunks by means of so-called *process logic links*. However to allow for reuse of the functional Chunks, the Chunks are not connected directly but through components. E.g. in the Conference Review System, adding a co-author to the author list of a paper can be done in two different tasks. Firstly, co-authors can be added when a new paper is registered and secondly co-authors can be added when the information about a registered paper is updated. The elementary task of adding a co-author to a co-author list of a paper is modeled only once by means of the functional Chunk 'AddCo-Author' and used twice: once in the task 'Register New Paper" and once in the task "Update Submission'. These two tasks are graphically represented in figure 6 and 7. The connection between a component and a Chunk is graphically represented by means of a dotted line.

To represent choices in the process logic we can use (conditional) one-to-many links. Also a reflective link may be useful in this context as for the 'Add Co-Author' case in figure 6 or for a "retry" link.

5.1.4 Navigational Aid Links

Next to the structural links, the semantic links and the process logic links, WSDM also allows to define *navigational aid links* in the conceptual model. Links of this type are added on top of the conceptual structure defined by means of the structural links to ease the navigation through the web site and to enhance the usability of the web site. From the viewpoint of being able to reach information, they are strictly speaking, not needed; all information and functionality should also be reachable without the navigational aid links. They could be compared to adding an index and post-it pointers to chapters in a book: the information in and the structure of the book stays the same, but the user is provided with shortcuts for more easily accessing the content of the book. Typical for navigational aid links is that usually we are not a-prior 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 aid link is the home link. Some navigational aid 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 aid links can be compared to shortcuts or accelerators used in most graphical user interfaces. Other examples of navigational links are the "landmark" links; the links used inside a site map to directly navigate to a part of the web site; the link to a site map itself. The home link and the landmark links are typical many-to-one links.

In WSDM, the home link and the landmark links are graphical represented by means of a symbol in the target component (being respectively the home-component and the landmark-component). The home-component is indicated with the letter H in a circle; a landmark-component with the letter L in a circle. This is illustrated in figure 8.

5.2 The link Types in the Conceptual Schema

In this section we discuss the different link types explained in the previous section in the context of the Conceptual Schema, which is the output of the Conceptual Design Phase. First we give an informal explanation and then we define the concepts more formally.

5.2.1 Building the Conceptual Schema

As explained in section 2, the Conceptual Design phase of WSDM consists of the sub phases Information Modeling, Functional Modeling and Navigational Design. During Information & Functional Modeling, the requirements of the different Audience Classes are modeled using

Chunks. Chunks are used to model elementary information requirements and elementary tasks (requirements are first decomposed into elementary requirements). Next, components and project logic links are used to construct an *Audience Task Model* for each requirement. Because each Audience Class has its own set of requirements, a set of Audience Task Models is made for each Audience Class.

During Navigational Design, the Conceptual Structural Model is made. Using the audience driven approach of WSDM, this is done by first creating for each Audience Class a so-called *Audience Track*, which is in fact a Conceptual Structural Model for a particular Audience Class. Next, all Audience Tracks are combined using the same hierarchical structure as the one defined in the Audience Class Hierarchy. This defines the final Conceptual Structural Model of the web site. In figure 9 the Conceptual Structural Model for the Conference Review System web site is represented graphically. Note that a shorthand notation is used for an Audience Track (double rectangle). This is done to hide the details of each Track. One of the Tracks (the author track) is given in figure 10. Also here a shorthand notation is used: Audience Task Models are represented as a dotted rectangle.

Also during Navigational Design, the semantic relationship links are defined. Also this is done per Audience Class. An overview of all Chunks of an Audience Class and their semantic relationship links is given in the *Audience Chunks Model*. Figure 11 gives the Audience Chunk Model for the PC-Chair Audience Class.

Navigational aid links are added where appropriate.

The combination of all these models: Conceptual Structural Model, Audience Tracks, Audience Task Models, Audience Chunks Models, and the individual Chunks themselves, is called the *Conceptual Schema*.

5.2.2 Formal Description

In this section, we formally define the concepts used during the Conceptual Design in WSDM. Chunks and Audience Classes are not defined formally because this it not directly relevant for this paper.

Definition 1 We define a *Conceptual Schema* CS for a web site as a 4-tuple (N, H, C, L), where:

- N is a finite non-empty set of components
- H is a finite non-empty set of Chunks
- $C \subseteq N \times H$ is the set of connections between nodes and Chunks
- L is a 4-tuple (L_S, L_R, L_P, L_A) representing the set of links where:
  - $L_S \subseteq 2^N \times 2^N$, is the set of structural links between the components
  - $L_R \subseteq 2^H \times 2^H$, is the set of semantic relationship links between Chunks
  - $L_P \subseteq 2^N \times 2^N$, is the set of process logic links between components
  - $L_A \subseteq 2^N \times 2^N$, is the set of navigational aid links between components

The node of a connection c is denoted *node(c);* the Chunk is denoted *chunk(c)*.

Definition 2 Let CS = (N, H, C, L) be a Conceptual Schema for a web site W where L = (L_S, L_R, L_P, L_A), then (N', H', C', L') is the *sub model* of CS if

$N' \subseteq N, N' \neq \emptyset$
$H' \subseteq H,$
$C' \subseteq C$
$L' \subseteq L$

where

$\forall l \in L'$: source(l) $\subseteq N'$ and target(l) $\subseteq N'$ and
$\forall c \in C'$: node(c) $\in N'$ and chunk(c) $\in H'$ and
$\forall h \in H': \exists n \in N' \wedge \exists c \in C'$: node(c) = n $\wedge$ chunk(c) = h and
$\forall n \in N': \exists l \in L'$: n $\in$ source(l) $\vee$ n $\in$ target(l)

The first two conditions ensure that all Chunks and nodes used in L' exist within the sub model, the last two ensure that no "un-used" (not connected, unreachable) Chunks and nodes exist.

By definition, the following lemma holds:

Lemma 1 A sub model of a Conceptual Schema is a Conceptual Schema

Definition 3 Let CS = (N, H, C, L) be a Conceptual Schema for a web site W where L = (L_S, L_R, L_P, L_A), then CSM=(N', H', C', L_S) is a *Conceptual Structural Model* if CSM is a sub model of CS.

Theorem 1 A Conceptual Structural Model of a Conceptual Schema is a Conceptual Schema

Proof: This is a consequence of Lemma 1

Definition 4 Let CS = (N, H, C, L) be a Conceptual Schema for a web site W where L = (L_S, L_R, L_P, L_A); let $\mathbf{A}$={A_1, ..., A_n} be the set of Audience Classes for W, then

M_{Am} = (H_m, L_{Rm}) is an *Audience Chunks Model* for the Audience Class $A_m \in \mathbf{A}$ if

$H_m \subseteq H$, $H_m \neq \emptyset$

$L_{Rm} \subseteq L_R$

where

$\forall\, l \in L_{Rm}$: source(l) $\subseteq H_m$ and target(l) $\subseteq H_m$ and

H_m contains all and only the Chunks that resulted from the information and functional requirements of A_m

Intuitively, an Audience Chunk Model ensures that all Chunks resulting from the requirements for every visitor (Audience Class) are in the model, and that all semantic links originate and end in existing Chunks.

Definition 5 Let CS = (N, H, C, L) be a Conceptual Schema for a web site W where L = (L_S, L_R, L_P, L_A); let $\mathbf{A}$={A_1, ..., A_n} be the set of Audience Classes for W, then

$T_{Am\text{-}R}$ = ($N_{Am\text{-}R}$, $H_{Am\text{-}R}$, $C_{Am\text{-}R}$, $L_{Am\text{-}R}$) is an *Audience Task Model* for the Audience Class $A_m \in \mathbf{A}$ and a requirement R if $T_{Am\text{-}R}$ is a sub model of CS

where

$L_{Am\text{-}R} \subseteq L_P$

$H_{Am\text{-}R}$ contains all and only the Chunks that resulted from the requirements R of A_m

An Audience Task Model for a certain requirement thus contains all and only the Chunks that resulted from this requirement and the process logic link between them.

Theorem 2 An Audience Task Model of a Conceptual Schema is a Conceptual Schema

Proof This is a consequence of Lemma 1

Finally, we are able to formally define an Audience Track. Intuitively, an Audience Track in an audience driven approach is a path for a particular visitor throughout the site that contains all information concerning him.

Definition 6 Let CS = (N, H, C, L) be a Conceptual Schema for a web site W where L = (L_S, L_R, L_P, L_A); let $\mathbf{A}$={A_1, ..., A_n} be the set of Audience Classes for W; and let CSM = (N', H', C', L_S) be the Conceptual Structural Model, then

T_{Am} = (N_m, H_m, C_m, Ls_m) is an *Audience Track* for the Audience Class $A_m \in \mathbf{A}$ if T_{Am} is a sub model of CSM

where

$Ls_m \subseteq L_S$

H_m contains all and only the Chunks that resulted from the information and functional requirements of A_m

Theorem 3 An Audience Track of a Conceptual Schema is a Conceptual Schema and a Conceptual Structural Model

Proof This is a consequence of Lemma 1 and Definition 3

6 Impact on the Design Process and the Usability of the Web Site

6.1 Impact on the Design Process

Distinguishing between the different types of links allows the modeler of a web site to focus on different aspects of the design separately and prohibit that diagrams are cluttered. In addition, the link concept as defined here is more powerful that the usual one.

During Information & functional modeling, the modeler will concentrate on the creation of the Chunks and on modeling the requirements using process logic links.

During navigational design the main focus is on structural links. The main question to be answer by the modeler in this design phase is: "How do I structure the available information and functionality?" WSDM provides some help by using an audience driven approach to this problem. This means that the information and functionality is structured following the requirements of the different Audience Classes. For each Audience Class, a navigational track is constructed that makes all information and functionality required by the members of the Audience Class available. This is done by combining the Audience Task Models built in the Information & Functional Modeling phase. Next, the different Audience Tracks are combined following the Audience Class Hierarchy built during the Audience Modeling phase.

Separately from the structure, the modeler can specify the navigation needs derived from semantic relationships present in the Universe of Discourse.

Only in a later stage, navigational aid links are added on top of the structure defined by the structural links. The navigation aid links put a second navigation layer on top of the navigation layer defined by the structural links. Mixing both types of links may be confusing and may hinder the designer to build a transparent structure for the web site. A clear conceptual structure is important for the usability of the later web site. An unclear structure will prevent the future user to create a correct mental model of the structure of the web site and induce the "lost-in-hyperspace" syndrome. Separating the structural links from the other types of links will allow the modeler to consciously reflect upon an appropriate structure (e.g. hierarchical, linear, grid) rather than to ensure that everything is reachable in some way.

During modeling, different diagrams can be used to represent the different links. Structural links are represented in the *Conceptual Structural Model* as well as some of the navigational aid links (e.g. the home link and the landscape link) (see figure 8 for an example). The other navigational aid links are represented where appropriate. An overview of the semantic relationship links for an Audience Class is given in the *Audience Chunks Model* (see figure 11 for an example). The process logic links are represented in an *Audience Task Model* (see figure 6 and 7 for an example). One Model per task is made. Please note that semantic relationship links and process logic links are always given in the context of an Audience Class. This is because Chunks are always related to an Audience Class. This has the advantage that it is possible to hide a semantic relationship for the members of one Audience Class and make it available for the members of an another, e.g. the Authors are not allowed to navigate from their paper to the reviewers of this paper while the PC-chair can do so. In addition, re-use of Chunks for different Audience Classes is possible, and reduces redundancy. Also the conceptual structure is given in the context of an Audience Class because in WSDM an *Audience Track* is created for each Audience Class. All Audience Tracks together form the Conceptual Structural Model. Although the type of a link can be derived from the diagram in which appears, it is better to represent them differently: colors are most appropriate for this.

The concept of link that we define here is much more powerful that the simple one-to-one uni-directional link currently used in the Web. The definition is in line with the notion of link used in Xlink. The one-to-many, many-to-one and many-to-many types of links allow to group logically related links and the concept of conditional link allows to constraint the use of the link on each branch of the link. This eases the modelling of temporal available links, and the personalisation and customisation of links. The advantage of bi-directional links is obvious.

6.2 Impact on the Usability of the Web Site

Because we distinguish between different types of links at an early stage, we will also be able to represent them differently in the eventual web site, e.g. different colors or concepts (menu, button, …) can be used. 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. From a user point of view, a web site in which for example navigation aid links are systematically presented in a certain way, will help the user to immediately understand why a link is there and what purpose is serves.

The 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, and the actual browsing strategy defined by this structure in the implementation.

The navigation aid 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. Therefore, navigational aid links should have a visual presentation that is different from the visual representation of the structural links (e.g. separate navigation pane, extra frame, …). 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).

Structural links can also be used to automatically generate a site map; semantic relationship links can be used to automatically generate a kind of "semantic web map". Opposed to the usual web map (which reflects (or should reflect) 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 [11]).

Identifying process logic links allows us to keep track of user transactions and to visualize this towards the user of the web site, i.e. we can allow the PC-Chair of our conference system user to suspend the marking of papers for a while (e.g. for asking an additional review) and later on allow him to return to the marking process. Looking for the details of a review or a paper is in principle not part of the process logic and therefore they should not be modeled by means of process logic links (but as semantic relationship 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.

7 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 sets of nodes. A node is a generic concept and can in principle be anything. A link can uni- or bi-directional. We also defined the arity of a link: one-to-one, one-to-many, many-to-one or many-to-many. Conditional links are defined as a special case and turned out to be very useful in modeling temporal links and for the personalization and customization of links. Also adaptive links are defined as a special case. The power of this type of link is not explored in this paper, but in [6] we use it in the modeling of the adaptive behavior of web site.

Next, we have defined four types of links that are useful to distinguish during the design phase of a web site: structural links, semantic relationship links, process logic links and navigational aid links. Structural links are used to structure the information and functionality in a web site; semantic relationship links are used to denote the need for navigation based on semantic relationships that exist between concepts in the universe of discourse; process logic links express parts of workflows; and navigational aid links are introduced to provide navigational aids to the user of the web site. For each of these types of links we have discussed the role that they play in the design process; how they enhance the modeling capabilities of the method and how they may influence the usability of the web site.

References

1. Allan, J: "Automatic Hypertext Link Typing", In Proceedings of ACM Hypertext96, Washington DC, 42-52, 1996.

2. Baron, L., Tague-Sutcliffe, J., Kinnucan, M.T.: "Labeled, Typed Links as Cues when Reading Hypertext Documents", In Journal of the American Association for Information Science, 47 (12), 896-908, 1996.

3. Bray, T., Paoli, J., Sperberg-McQueen, C.M., Maler, E.: "Extendible Markup Language (XML) 1.0 (Second Edition)", W3C Recommendation 6th October 2000 (http://www.w3.org/TR/REC-xml), 2000

4. Cachero, C., Gomez, J., Parraga, A., Pastor, O.: "Conference review system: a case study", In Proceedings of IWWOST 2001 (http://www.dsic.upv.es/~west2001/iwwost01/), Valencia, Spain, 2001

5. Casteleyn, S., De Troyer, O.: "Structuring Web Sites using Audience Class Hierarchies", Proceedings of DASWIS 2001 workshop (attached to the ER 2001 conference), Yokohama, Japan, 2001

6. Casteleyn, S., De Troyer, O., Brockmans, S.: "Design Time Support for Adaptive Behaviour in Web Sites", Proceedings of the 18th ACM Symposium on Applied Computing (SAC2003), Florida, USA, 2003

7. Ceri, S., Fraternali, P., Bongio, A., Maurino A.: "Modeling data entry and operations in WebML", In Proceedings of WebDB 2000, Dallas, USA, 2000

8. Cleary, C., Bareiss, B.: "Practical Methods for Automating Linking in Structured Hypermedia Systems", In Proceedings of ACM Hypertext96, 31-41, 1996

9. DeRose, S., Maler, E., Orchard, D.: "XML Linking Language (XLink) Version 1.0", W3C Recommendation 27 June 2001 (http://www.w3.org/TR/xlink/), 2001

10. De Troyer, O., Casteleyn, S.: "The Conference Review System with WSDM", In Proceedings of IWWOST 2001 (http://www.dsic.upv.es/~west2001/iwwost01/), Valencia, Spain, 2001

11. De Troyer, O.M.F. and T. Decruyenare (2000), "Conceptual Modelling of Web Sites for End-Users," In: WWW Journal, Volume 3, Issue 1, Baltzer Science Publishers, 2000.

12. Fensel, D., Angele, J., Decker, S., Erdmann, M., Schnurr H., Staab, S., Studer, R., Witt, A.: "On2broker: Semantic-Based Access to Information Sources at the WWW". In Proceedings of the World Conference on the WWW and Internet (WebNet 99), Honolulu, Hawai, USA, 1999.

13. Fensel, D., Horrocks, I., Van Harmelen, F., Decker, S, Erdmann, M., Klein, M.: "OIL in a nutshell", In Knowledge Acquisition, Modeling, and Management, Proceedings of the European Knowledge Acquisition Conference (EKAW-2000), R. Dieng et al. (eds.), Lecture Notes in Artificial Intelligence, LNAI, Springer-Verlag, 2000.

14. Garzotto, F. Paolini, P, Schwabe, D.: "HDM A Model-Based Approach to Hypertext Application Design", ACM Transaction on Information Systems, vol. 11, nº 1, pages 1-26, 1993

15. Halasz, F. G., Schwartz M.: "The Dexter Hypertext Reference Model", Proceedings of NIST Hypertext Standardization Workshop, Gaithersburg, Maryland (1990)

16. Halpin, T.: "Conceptual Schema and Relational Database Design, second edition", Prentice Hall, Australia (1995)

17. Hansen, K.M., Yndigegn, C., Grønbæk, K.: "Dynamic Use of Digital Library Material - Supporting Users with Typed Links in Open Hypermedia", In proceedings of ECDL'99, Paris, 1999

18. Kopak, R.W.:"A proposal for a taxonomy of functional link types", In ACM Computing Surveys 31(4), December 1999

19. Lassila, O., Swick, R.R.: "Resource Description Framework (RDF) Model and Syntax Specification", W3C Recommendation 22nd February 1999 (http://www.w3.org/TR/REC-rdf-syntax/), 1999

20. Scott, C.R, Finin, T, , Joshi, A., Peng, Y., Nicholas, C., Soboroff, I., Chen, H., Kagal, L., Perich, F., Zou, Y.,Tolia, S.:."ITtalks: A Case Study in the Semantic Web and DAML+OIL", In IEEE Intelligent Systems, Jan/Feb Vol. 17, No 1, 2002

21. Tebbutt, J.M.: "User evaluation of automatically generated semantic hypertext links in a heavily used procedural manual", Interim report, http://www.itl.nist.gov/iad/894.02/works/papers/user_eval.html

22. Trigg, R.: "Network-Based Approach to Text Handling for the Online Scientific Community", Ph.D. dissertation, Departement of Computer Science, University of Maryland, technical report TR-1346, 1983

Figures

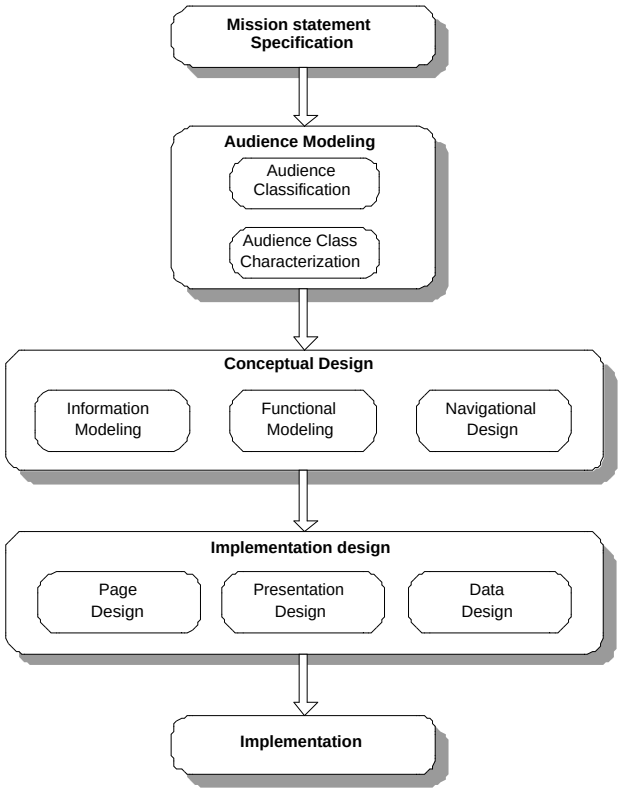

Figure 1: WSDM overview

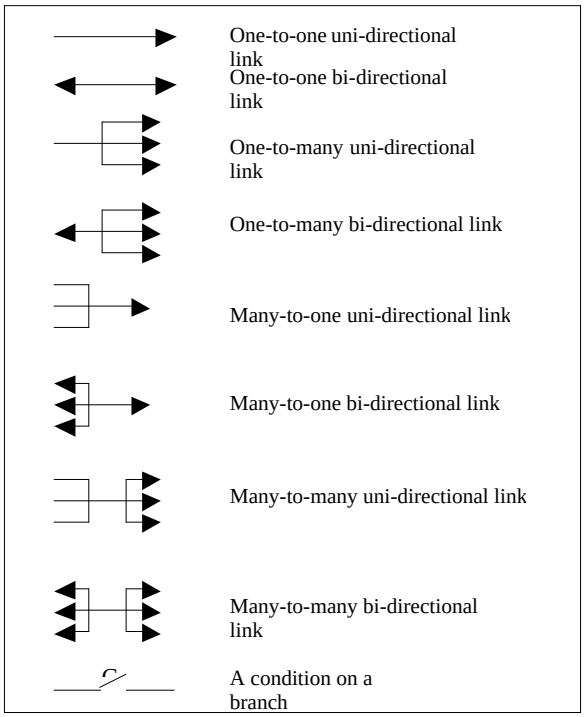

Figure 2: Graphical Representation of Links

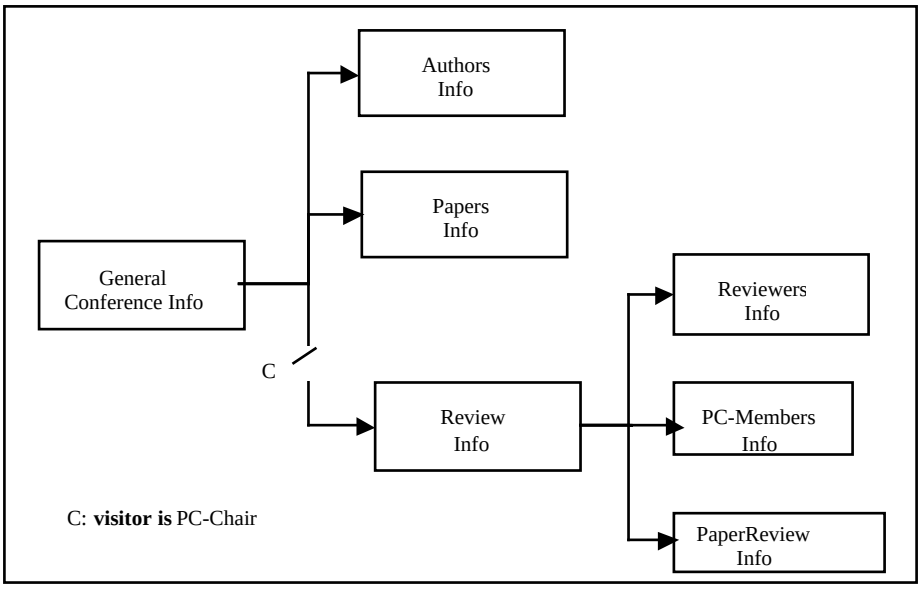

Figure 3: An Example Conceptual Structure for the (simplified) Conference Review Web Site

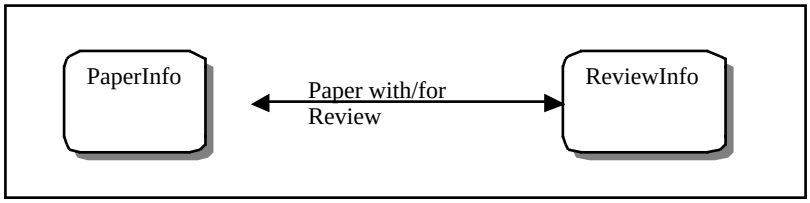

Figure 4: A semantic Relationship Link

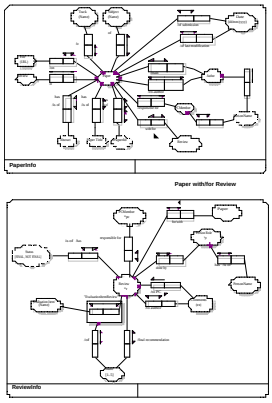

Figure 5: More Detailed Specification of a Semantic Relationship Link

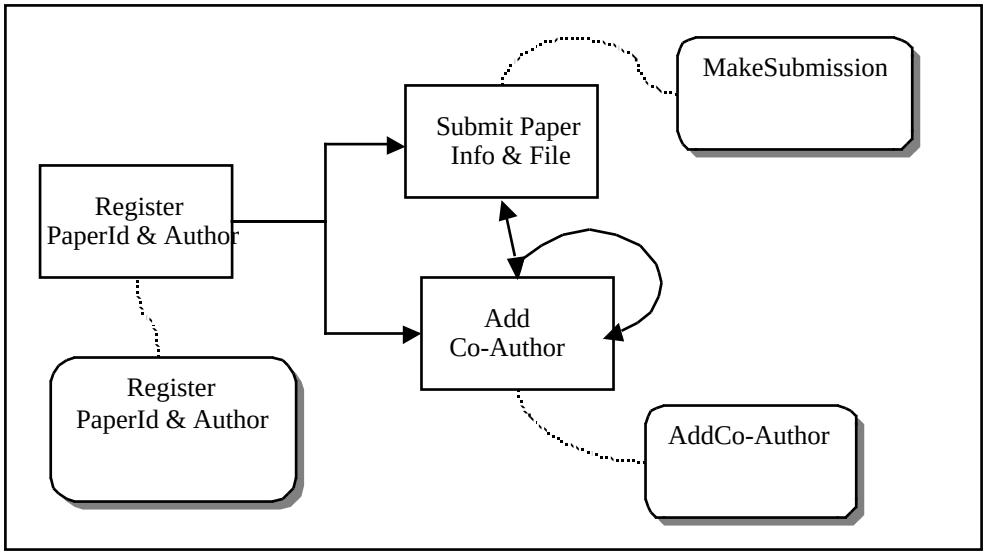

Figure 6: Modeling of the task 'Register New Paper'

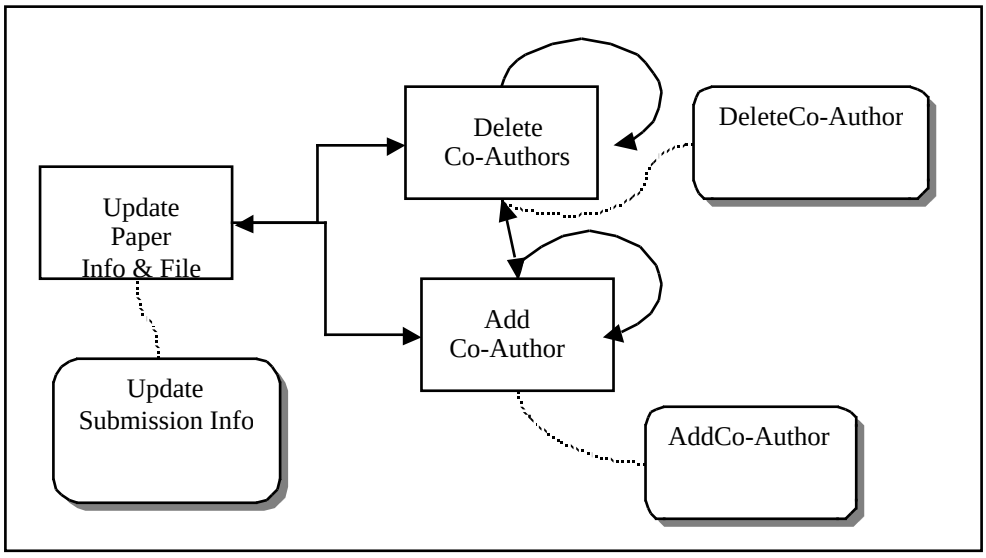

Figure 7: Modeling of the Task 'Update Submission'

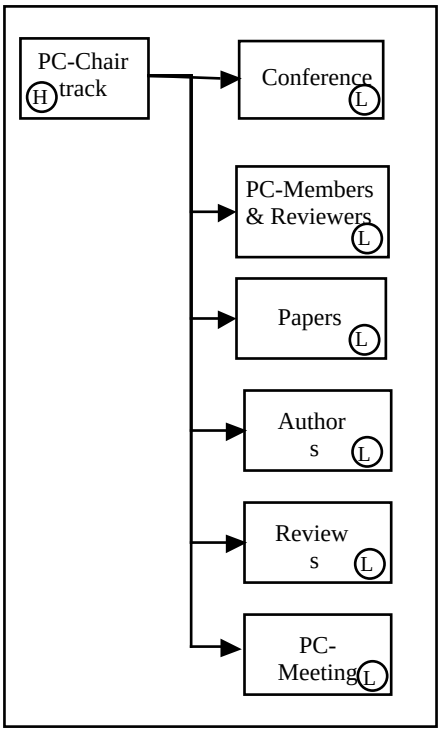

Figure 8: Example of home and landmark links

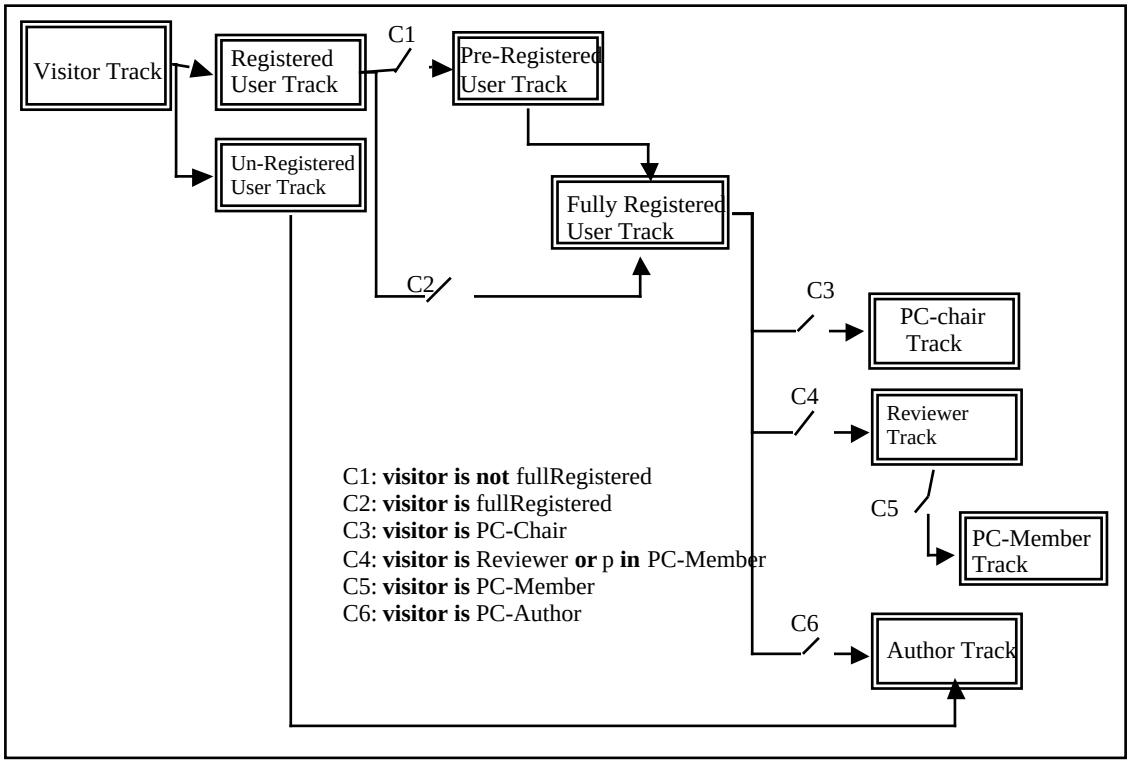

Figure 9: Conceptual Structural Model for the Conference Review System Web Site

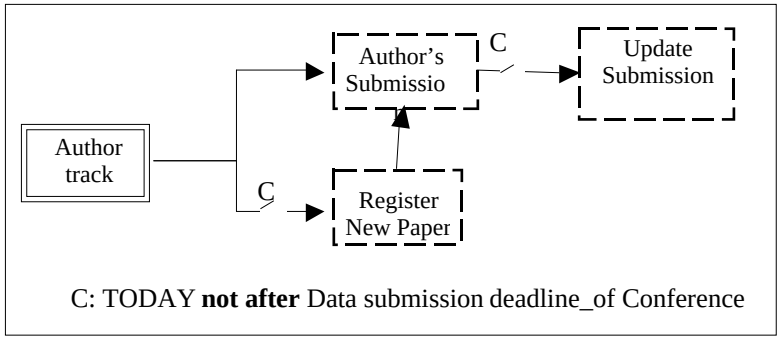

Figure 10: Details of Audience Track 'Author track'

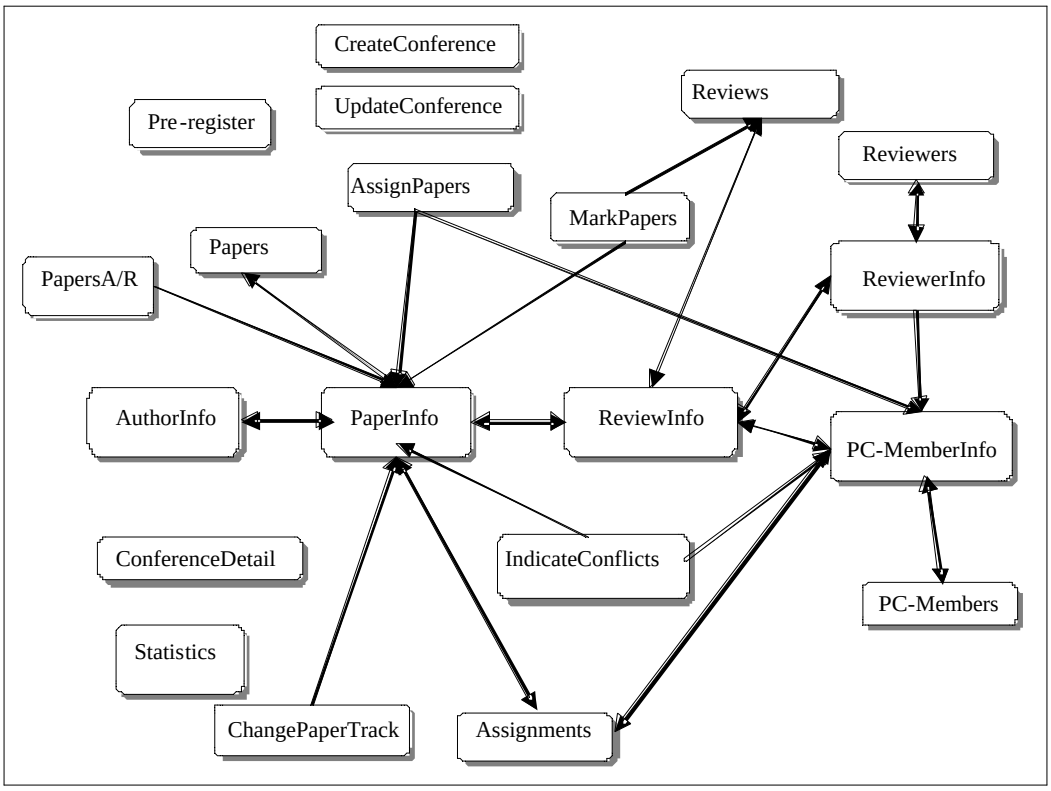

Figure 11 Audience Chunk Model for the PC-Chair Audience Class