New Review of Hypermedia and Multimedia

Applying measurement principles to improve hypermedia authoring
M. Emilia*; X. Mendes*; Wendy Hall*; Rachel Harrison*

* Computer Science Department, The University of Auckland, Auckland, New Zealand
* Multimedia Research Group, Department of Electronics and Computer Science, University of Southampton, Highfield, Southampton, UK
* School of Computer Science, Cybernetics and Electronic Engineering, The University of Reading, Whiteknights, UK

Online publication date: 20 July 2010

To cite this Article Emilia, M., Mendes, X., Hall, Wendy and Harrison, Rachel (1999) 'Applying measurement principles to improve hypermedia authoring', New Review of Hypermedia and Multimedia, 5: 1, 105 — 132

To link to this Article DOI: 10.1080/13614569908914710

URL: http://dx.doi.org/10.1080/13614569908914710

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
Applying measurement principles to improve hypermedia authoring

M. Emilia X. Mendes
Computer Science Department, The University of Auckland, Private Bag 92019, Auckland, New Zealand
e-mail: emilia@cs.auckland.ac.nz

Wendy Hall
Multimedia Research Group, Department of Electronics and Computer Science, University of Southampton, Highfield, Southampton S017 1BJ, UK
e-mail: wh@ecs.soton.ac.uk

Rachel Harrison
School of Computer Science, Cybernetics and Electronic Engineering, The University of Reading, Whiteknights, PO Box 217 Reading, RG6 6AH, UK
e-mail: Rachel.Harrison@reading.ac.uk

Employing scientific investigation is essential to understanding hypermedia processes and products, making hypermedia authoring a science rather than an art. This paper describes a case study aimed at validating empirically metrics proposed to measure the development effort involved in authoring a hypermedia application. The case study is one of the empirical evaluations that we have conducted and it has been planned according to a scientific method of research. Our objective is to improve hypermedia authoring by measuring some quality characteristics of hypermedia applications and the processes involved in developing hypermedia applications. The metrics proposed here adhere to the representational theory of measurement and all the measurement activities have been developed according to Fenton and Pfleeger’s Conceptual Framework for Software Measurement and the guidelines from the DESMET project. The theoretical validation of the metrics proposed is also described. Results have shown that several of our metrics have been confirmed as possible measures of development effort.

1. INTRODUCTION

The last decade has seen the proposal of a great number of hypermedia methods, methodologies and systems, each one offering to improve some characteristic of hypermedia applications, their development, or their maintenance. Although we all seek hypermedia methods, methodologies and systems that will make us more productive and improve the quality of our applications, rigorous experimentation is needed to evaluate new technologies and their effects on our organisations, processes, and products (1). Using a scientific investigation is essential to understanding hypermedia processes and products and to making hypermedia authoring a science rather than an art. As engineering involves the analysis of measurements (2), hypermedia engineering will only
become a true engineering discipline when a solid foundation of measurement-based theories is built.

We regard measurement as important for three basic activities: (i) understanding the development, reuse and maintenance processes; (ii) controlling software projects; (iii) improving processes and products. We have proposed metrics to measure some quality characteristics of hypermedia applications as a way of improving hypermedia authoring. The quality characteristics considered were the maintainability of hypermedia applications, the reusability of hypermedia information and the hypermedia application's development quality. The principles of the metrics we developed are based on Fenton and Pfleeger's Conceptual Framework for Software Measurement (3) and on the DESMET Methodology (4), in the field of software engineering. We have drawn analogies between the way that such methods and processes have been applied in software engineering to evolve well understood techniques for improving the quality of computer programs, and the need for such rigorous techniques to be applied to hypermedia authoring (5).

All our metrics adhere to the representational theory of measurement and have been validated theoretically and empirically according to the validation framework proposed by Kitchenham et al. (6).

Two empirical evaluations were conducted in order to validate our choice of metrics: a quantitative survey and a quantitative case study. The survey evaluated whether or not our metrics could be used to measure the maintainability of hypermedia applications and the reusability of hypermedia documents and the case study evaluated whether or not our metrics could be used to measure the development effort involved in authoring a hypermedia application. The reader is referred to (7, 8, 9) for a detailed description of the survey, as it is not the focus of this paper. The case study is described in depth in section 3.

In section 2 we describe and analyse previous work in hypermedia metrics, noting that little of this work had been validated using scientific methods of research. We then present in more detail, in section 3, our research project and describe the case study. Finally, in section 4, we give our conclusions and comments on future work.

2. BACKGROUND
2.1 Hypermedia metrics proposals
The definition of metrics to improve hypermedia authoring and evaluation has stimulated considerable interest. This section presents briefly several
proposals of hypermedia metrics and analyses them based on the validation principles suggested by Briand et al. (10). The reader is referred to (9) for a more detailed description of those proposals.

Proposal 1: Botafogo et al. (11) defined metrics based on the assumption that giving authors an idea of the complexity and cross-referencing of the hypertext application could help them to reduce undesired structural complexity and create applications that readers can traverse more easily. They proposed node metrics (depth and inbalance) and global metrics (stratum and compactness), the latter two calculated considering only the structure of the hyperdocument. They are described below:

- The depth metric indicates a node's distance from the root.
- The inbalance metric indicates whether the hyperdocument looks like a balanced tree and what are the nodes that cause the tree to be unbalanced.
- The compactness metric indicates the level of cross-referencing of a hypertext application. Its value varies between 0 and 1, and they correspond respectively to a completely disconnected application and a completely connected application.
- The stratum metric indicates to what degree the application is organised so that some nodes must be read before others. Its value also varies between 0 and 1, 1 corresponding to a linear application and 0 corresponding to no imposed reading order.

Proposal 2: Yamada et al. (12) proposed three metrics: interface shallowness, downward compactness and downward navigability. Each one is described below:

- The interface shallowness metric represents the ‘heaviness’ of the cognitive load on users. The essential idea is that links can lead to documents in the same layer or in a different one.
- The downward compactness metric denotes the structural complexity of reaching the nodes from the root.
- The downward navigability metric measures hypermedia navigability. This assumes that an easily navigable hypermedia application (i) has a shallow interface layer from the root to the nodes (light cognitive load) and (ii) is compact from the root (that is, it is structurally simple to reach the nodes from the root).

Proposal 3: Hatzimanikatis et al. (13) define a hypertext quality model using the Factor-Criteria-Metric model (14) and then propose structure metrics, i.e., metrics that can be computed using only the structure of the hypertext application (the hypertext graph). From the hypertext quality model proposed (see figure 1), the two factors considered were readability and maintainability.

The New Review of Hypermedia and Multimedia 1999
The authors defined metrics for path complexity, tree impurity, modularity and individual node complexity. Each will be described below:

- **Path complexity** can be measured using either compactness, stratum (11), cyclomatic complexity number (assuming that the hyperdocument can be compared to a computer program) or the data structure complexity metric.
- **Tree impurity** is measured using a metric of tree impurity described in Fenton and Pfleeger (3).

**FIG. 1: Hypertext Quality Model (13)**

**FIG. 2: Factor, Criteria, Metric for Usability (15)**

The New Review of Hypermedia and Multimedia 1999
• Modularity is measured using an information flow complexity metric.
• The individual node complexity also uses the same metric used to measure modularity.

Proposal 4: Garzotto et al. (15) provide a framework for design-oriented hypermedia evaluation, using both a set of design objects, based on HDM (Hypermedia Design Model), and a set of fine-grained attributes of these objects that can impact on hypermedia usability. Their main hypothesis is that assessing the design quality is one of the steps towards evaluating the overall quality of a hypermedia application. They concentrated on a specific product attribute, usability, which was decomposed as shown (see figure 2). In figure 2 we can see that criteria proposed have no corresponding metrics defined.

More recently, the authors have published a review of this model (16), and a new evaluation model which is a specialisation for hypermedia of a general methodology for usability evaluation of interactive systems called SUE (Systematic Usability Evaluation). SUE is based on the use of a hypermedia model (HDM), a set of hypermedia-specific usability attributes (see table 1), and a set of patterns of inspection activities, called abstract tasks.

<table>
<thead>
<tr>
<th>General Principles</th>
<th>Criteria</th>
<th>Attributes</th>
<th>HDM Dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>S  N  B  UC</td>
</tr>
<tr>
<td>Efficiency</td>
<td>Accessibility</td>
<td>Access Layer Soundness</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Navigational Richness</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>Orientation</td>
<td>Session History Soundness</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Context Observability</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reuse Soundness</td>
<td>*  *  *  *</td>
</tr>
<tr>
<td></td>
<td>User Control Availability</td>
<td>Media Control Availability</td>
<td>*  *</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Navigation Control Availability</td>
<td>*</td>
</tr>
<tr>
<td>Learnability</td>
<td>Consistency</td>
<td>Structural Consistency</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dynamic Consistency</td>
<td>*  *  *</td>
</tr>
<tr>
<td></td>
<td>Predictability</td>
<td>Regularity</td>
<td>*  *  *</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Media Interference Soundness</td>
<td>*  *</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Navigation and Media Interference Soundness</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Collection Ordering Coherence</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Users' Knowledge Conformance</td>
<td>*  *  *</td>
</tr>
</tbody>
</table>

Legend: S = Structure; N = Navigation; B = Behaviour; UC = User Control

Table 1: Summary of hypermedia-specific usability attributes (16)

Proposal 5: Fletcher et al. (17) propose an approach for assessing multimedia component and system characteristics and also use these features to estimate the associated development effort. Their approach is based on the assertion that Multimedia Information Systems (MMIS) development effort is a function of: i) building the application’s content; and ii) authoring the application. The metrics considered to measure the
process of building the application's content are: filename, media type (graphic, audio, video, animation, photograph, scan), whether the media is original or pre-existing, what was the effort involved in creating the media (in case it is original), what was the effort involved in digitising scans, video and audio, what was the effort involved in editing the files, and what was the duration of each component that is an animation, sound or video.

The metrics proposed to measure the process of authoring the application are: the screen name, the effort involved in authoring each screen, the complexity of each type of media, the complexity of each screen. For each screen the attributes are: the number of objects on the screen (including sounds), the number of links between that screen and other screens, the number of events on a screen and the average number of actions per event. The media complexity considers only graphics data: the number of objects on the component, whether it had been reused elsewhere in the project, the form that it took (button, toolbar, screen, background, etc).

A follow-up publication (18) shows that the authors have decided to change their previous approach and to focus on the development of a literature- and industry-based metrics framework for multimedia systems effort modelling, followed by a wider industry verification.

Proposal 6: Fortes and Nicoletti (19) propose link based metrics as a way of improving reliability and maintainability in hypermedia authoring. Their hypothesis is that a tool which can provide feedback on the hyperdocument structure (concerning its links), can contribute to a successful management of the hyperdocument authoring process and maintenance of Web applications. Their metrics are based on a taxonomy of possible reuse cases for links (see table 2). They have identified eight different reuse cases, representing three categories: general links (cases 3 and 4), unusual links (cases 0 and 7) and contextual links (cases 1,2,5,6).

<table>
<thead>
<tr>
<th>Reuse case number</th>
<th>Links having</th>
<th>Same source nodes</th>
<th>Same source anchors</th>
<th>Same destination nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>(000)_{n} = 0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>(001)_{n} = 1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>(010)_{n} = 2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>(011)_{n} = 3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>(100)_{n} = 4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>(101)_{n} = 5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>(110)_{n} = 6</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>(111)_{n} = 7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Legend: 0 = no; 1 = yes.

Table 2: Possible reuse cases of link components (19)
2.2 Analysis of the Proposals

The six proposals presented above are compared and analysed here (see table 3) based on the four questions that, according to Briand et al. (10), should be asked when validating a measure:

i) Is the measure adequately capturing the attribute it purports to measure (i.e., construct validity)?

ii) Is the attribute itself well defined based on an explicit empirical model (i.e., empirical relational system)?

iii) Is there any empirical evidence supporting the underlying hypotheses of the empirical model?

iv) Is the measure useful from a practical perspective?

<table>
<thead>
<tr>
<th>Questions</th>
<th>Metrics</th>
<th>Proposals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Y</td>
</tr>
<tr>
<td>(1)</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>(2)</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>(3)</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>(4)</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

B = Botafogo et al. (1992); G = Garzotto et al. (1994); Y = Yamada et al. (1995); F = Fletcher et al. (1997); H = Hatzimanikakis et al. (1995); N = Fortes and Nicoletti (1997).

No = the characteristic has not been fulfilled by the proposal;
Yes = the characteristic has been fulfilled by the proposal.

Table 3: Comparison of the proposals

All the quality factors presented by the six proposals can be useful from a practical perspective, although their metrics lack a corresponding empirical model and do not capture the attributes they were supposed to measure. None of the metrics proposed had been validated theoretically.

Botafogo et al. (11) proposed metrics that, although interesting from the practical perspective, do not have an explicit empirical model and were not validated empirically.

Yamada et al. (12) proposed three metrics which are based on an empirical model. However, they made several assumptions for their hypotheses which have not been validated.

Hatzimanikakis et al. (13) has used mainly metrics from the software engineering field and applied them directly to hypermedia. Although some of these metrics may be useful for software engineering, they are not necessarily adequate when applied to the hypermedia context. The authors suggest that path complexity can be measured by compactness, stratum, cyclomatic complexity and data structure. Stratum is the only metric,
among the four proposed, that could be used to measure the complexity of a hypermedia path and, consequently, readability. Cyclomatic complexity measures the number of linearly independent paths and the hypothesis is that the higher the number of links on a document, the more difficult its readability will be. However, a document may have many links and still be quite easy to read, e.g., table of contents. The metric tree impurity assumes that hypermedia applications organised in a hierarchy make readability easier. However, what about applications that have been structured sequentially or in a network style, where there are enough interface cues to help navigation? Modularity and node complexity use an information flow principle. However, in many situations, following a link represents an action where no data is passed between the entities involved (as in programming languages).

Garzotto et al. (15) and Garzotto and Matera (16) do not define any direct metrics or empirical model for their product attributes. The authors also use only HDM as a reference model to describe the hypermedia applications, narrowing the external validity of their proposal.

Fletcher et al. (17) and McDonell and Fletcher (18) proposed a set of direct attributes to measure development effort. These attributes were defined according to an empirical model. They considered the development of an application as a result of building the content and designing the interface, leaving aside other phases where metrics would be very useful, e.g. application design. Their metrics, despite being related to multimedia applications rather hypermedia applications, have given us insights to measuring development effort.

Fortes and Nicoletti (19) proposed a metric, measured at a nominal scale, to measure the type of link reuse. However, they did not describe explicitly an empirical model for their attribute nor evaluate their metric empirically.

The metrics proposed by Botafogo et al. (11) and Fletcher et al. (17) are particularly relevant to our work, as they relate to some of the factors identified when developing our metrics.

3. APPLYING MEASUREMENT PRINCIPLES TO HYPERMEDIA AUTHORING

Using as inspiration the published literature in the field of software measurement, we have defined, in the scope of our project, a methodology aimed at improving hypermedia authoring by using a scientific method of research. The main objectives of our project are as follows:

i) The proposal of metrics to measure the maintainability of hypermedia
applications, the reusability of hypermedia documents and the development effort involved in authoring a hypermedia application;

ii) The proposal of metrics which adhere to the representational theory of measurement;

iii) The use of Fenton and Pfleeger's (3) conceptual framework for software measurement and the guidelines from the DESMET project (4) to plan the hypermedia measurement activities that we undertook;

iv) The empirical and theoretical validation of our metrics according to the validation framework proposed by Kitchenham et al. (6).

The metrics that we have proposed have been validated empirically using two evaluations: a quantitative survey and a quantitative case study. The next section will describe in depth the case study, presenting the theoretical and empirical validation of the metrics and the results obtained.

4. THE CASE STUDY EVALUATION

4.1 Introduction

The case study evaluation measured some characteristics of hypermedia applications and systems (independent variables), and the effort involved in developing a hypermedia application (dependent variable), analysing statistically whether or not the independent variables had any association with the dependent variable. A questionnaire was used to collect the necessary data. The independent variables, our metrics, used in the case study were: hyperdocument size, connectivity, compactness, stratum, link generality and link representation. Each one is described as follows:

- **Hyperdocument size** refers to the number of documents that the hypermedia application has. Documents are considered here to be either HTML files or any kind of file that is defined as a document in the hypermedia systems used in the evaluations. The hyperdocument represents a collection of documents designed with a specific aim. In our case, the domain considered to be representative was education. The applications here are considered static, where the number of dynamically generated documents is quite small, if not absent. Although we have measured hyperdocument size by counting the number of documents we are aware that other metrics can also be used (for example, the length of the application, the complexity of documents and the functionality offered) and they will be the focus of our next evaluation.

- **Connectivity** refers to the number of links that the hypermedia application has. The links considered here are either structural or referential (20), leaving apart dynamically computed links. Based on the number of documents and the number of links, the average number of links per document can also be computed. Although this information
is useful to measure of the quality of hypermedia usability, its use is outside the scope of this paper.

- **Compactness** (11) indicates how inter-connected the documents are, i.e., the level of cross-referencing in a hypermedia application. Here subjects were asked to estimate the compactness of the hypermedia application, from 1 (completely disconnected) to 5 (completely connected). Although the automatic computation of compactness can be difficult as a Web site can have several links leading to external nodes, it can still be a useful metric to measure more self-contained Web sites. The increase in the number of links may influence the compactness of an application, however, this relationship is not always clear-cut.

- **Stratum** (11) indicates to what degree the hypermedia application is organised for directed reading, i.e., the level of linearity of a hypermedia application. Here subjects were asked to estimate the stratum of the hypermedia application, from 1 (completely non-linear) to 5 (completely linear).

- **Link Generality** refers to whether the link applies at a single anchor only, as for example, point-to-point links, or whether it applies (or can be applied) at multiple anchors (at the same document or in different documents), as for example, generic links in Webcosm (21). In order to check this hypothesis we have used two different hypermedia systems in the case study, the Web (22) and Webcosm, to compare links with different levels of generality. In relation to the link generality: i) In the Webcosm model, a link associates a particular source selection (anchor) with its destination and can be recognised by the system within three different scopes (limits): specific (point-to-point), local or generic. A point-to-point link has as its scope the anchor associated to it; a local link can be followed from any occurrence of the source selection in a particular document which means that the document is its scope. A generic link can be followed from any occurrence of that source selection in any document, which means that its scope is the whole hypermedia application. Generic links have a higher generality than local links, which have a higher generality than specific or point-to-point links. For the case study evaluation, link generality was measured considering only generic and point-to-point links, as these links are in practice the ones most frequently used by most Webcosm authors. ii) In the standard Web model there are only point-to-point links.

Compactness and stratum were estimated by subjects, rather than directly measured in the applications. It can be argued that estimated values are not ideal and may weaken the results obtained. However, it would not be feasible to ask subjects to calculate themselves those measures as it would
be very time-consuming. Compactness, as proposed by Botafogo et al. also uses an estimated value on its calculation (the conversion constant), leading to subjective measures as well. Finally, stratum is described by the authors as "not a perfect metric, where some problems are evident". Therefore, we decided to re-use these concepts and measure them in a different way.

The dependent variable collected to measure the development quality was time, representing the estimated elapsed time spent by the author in developing the hypermedia application.

Although according to the scientific method of research the hypotheses used in any evaluation exercise should be the null hypotheses, we decided to present here the alternative hypotheses as they indicate our assumptions about the direction of the relationships investigated (for hypotheses H1 to H4):

H1 - The bigger the hyperdocument size the more time-consuming the hypermedia application's development becomes.
H2 - The greater the connectivity the more time-consuming the hypermedia application's development becomes.
H3 - The higher the compactness the more time-consuming the hypermedia application's development becomes.
H4 - The smaller the stratum the more time-consuming the hypermedia application's development becomes.
H5 - Link generality: generic links allow the development of a hypermedia application with less effort than point-to-point links.

Hypothesis H5 should only be considered whenever hypermedia authors see the use of generic links as being appropriate.

Each metric used in one of these hypotheses represents a different perspective of the development effort. Consequently, the hypotheses and their corresponding metrics should not be considered separately from one another.

Intuition is the starting point for all measurement (3). Consequently, the metrics and the hypotheses proposed here are based on our intuition of which attributes of hypermedia applications and systems are related to the development effort of a hypermedia application and what the direction of that relationship is.

Another point is that the measurement proposed here is a measurement for assessment rather than a measurement for prediction (3). This means that we are interested in understanding hypermedia authoring at this stage, rather than trying to produce a mathematical model to be used as a predictor of development effort. In addition, as some of the metrics were
measured using an ordinal scale, having them as parameters of a predictive equation to calculate the development effort would not be adequate as it would violate the permissible transformations for that scale.

4.2 The theoretical validation of the metrics

This sub section presents a theoretical validation of our metrics according to the validation framework proposed by Kitchenham et al. (6).

For the validation framework, to decide whether a measurement is valid it is necessary to confirm:

- **Attribute validity**, i.e., whether the attribute is actually exhibited by the entity being measured

- **Unit validity**, i.e., whether the measurement unit being used is an appropriate means of measuring the attribute.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Scale</th>
<th>Unit</th>
<th>Properties that apply</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hyperdocument size</td>
<td>absolute</td>
<td>documents</td>
<td>(1) Different applications may have different hyperdocument sizes. (2) The same application with different names will lead to the same hyperdocument size if a software program is used as an instrument. (3) Has one unit. (4) Different applications can have the same hyperdocument size.</td>
</tr>
<tr>
<td>Connectivity</td>
<td>absolute</td>
<td>links</td>
<td>(2) Different applications may have different connectivity. (3) The same application with different names will lead to the same connectivity if a software program is used as an instrument. (4) Has one unit. (5) Different applications can have the same connectivity.</td>
</tr>
<tr>
<td>Compactness</td>
<td>ordinal</td>
<td>Ranking of cross-referencing</td>
<td>(1) Different applications may have different compactness. (2) The same application with different names may lead to the same compactness if the respondents share a common knowledge about the application’s structure. (3) Has one unit. (4) Different applications can have the same connectivity.</td>
</tr>
<tr>
<td>Stratum</td>
<td>ordinal</td>
<td>Ranking of linearity</td>
<td>(1) Different applications may have different stratus. (2) The same application with different names may lead to the same stratum if the respondents share a common knowledge about the application’s structure. (3) Has one unit. (4) Different applications can have the same stratum.</td>
</tr>
<tr>
<td>Link generality</td>
<td>Nominal</td>
<td>Boolean (anchor/not anchor)</td>
<td>(1) Different systems may have different link generalities. (2) The same system with different names will lead to the same link generality. (3) Has one unit. (4) Different systems can have the same link generality.</td>
</tr>
<tr>
<td>Development effort for a hypermedia application</td>
<td>Absolute</td>
<td>Estimated elapsed time in hours</td>
<td>(1) Different applications may have different development efforts (duration). (2) The same hypermedia application with different names may lead to the same duration, if answered by the same author. (3) Has one unit. (4) Different hypermedia applications can require the same development effort.</td>
</tr>
</tbody>
</table>

Table 4: Theoretical validation of the direct measures
• **Instrument validity**, i.e., whether any model underlying a measuring instrument is valid and the measuring instrument is properly calibrated.

• **Protocol validity**, i.e., whether an acceptable measurement protocol is adopted.

The properties of measures used in the validation are:

1. For an attribute to be measurable, it must allow different entities to be distinguished from one another.
2. A valid measure must obey the representation condition, preserving our intuitive notions about the attribute and the way in which it distinguishes different entities.
3. Each unit of an attribute contributing to a valid measure is equivalent.
4. Different entities can have the same attribute value.

The theoretical validation for our metrics is presented in table 4:

**Validation of Instrument Models**

Validating a measurement instrument represents the confirmation that the instrument accurately measures attribute values in a given unit. The instrument used in the quantitative case study measured all the attributes using appropriate scales and units.

**Validation of Measurement Protocols**

The protocol applied to the instrument used was validated by peer acceptance. In addition, when asking for volunteers, the questionnaire and evaluation's objectives were always explained.

4.3 Applying the conceptual framework to the case study

All the measurement activities that we have carried out were based on Fenton and Pfleeger's conceptual framework for software measurement (3) and on the guidelines from the DESMET project (4).

Fenton and Pfleeger's conceptual framework can be applied to the diverse software-measurement activities that contribute to an organisation's software practices. These practices can be not only the development and maintenance activities but also any experiments and case studies performed in order to investigate new techniques and tools. The framework is based on three principles:

1. Classifying the entities to be examined;
2. Determining relevant measurement goals; and
3. Identifying the level of maturity that the organisation has reached.

The DESMET project was a collaborative project part-funded by the U.K.
Department of Trade and Industry. Its goal was to develop a comprehensive methodology for evaluating software engineering methods and tools.

Our classification of entities for the case study is presented in table 5 and described as follows:

<table>
<thead>
<tr>
<th>ENTITIES</th>
<th>ATTRIBUTES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Products</td>
<td>Internal</td>
</tr>
<tr>
<td>Hypermedia Application</td>
<td>Hyperdocument size, Connectivity, Compactness, Stratum,</td>
</tr>
<tr>
<td>Hypermedia System</td>
<td>Link generality, Link representation</td>
</tr>
<tr>
<td>Processes</td>
<td></td>
</tr>
<tr>
<td>Development of a hypermedia application</td>
<td>Effort (time)</td>
</tr>
</tbody>
</table>

Table 5: Classification of Products and Processes for the case study

The classification of entities shows that we wish to measure the entities hypermedia application, hypermedia system and the development of a hypermedia application. The attributes of the entity hypermedia application that we wish to measure are: *hyperdocument size, connectivity, compactness* and *stratum*. The attribute of the entity hypermedia system that we wish to measure is *link generality*. Finally, the attribute of the entity development of a hypermedia application that we wish to measure is effort, which represents the elapsed time used to author a hypermedia application.

The Goals, Questions and Metrics used in the case study, which were proposed based on the GQM model (23), are presented in table 6 and described as follows:

<table>
<thead>
<tr>
<th>Goal</th>
<th>Question</th>
<th>Metric</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purpose: to assess</td>
<td>what is the influence of some hypermedia application's attributes on the developability of the application?</td>
<td>Hyperdocument size, Connectivity, Compactness, Stratum, Effort</td>
</tr>
<tr>
<td>Issue: the hypermedia application's development process</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Object: process</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Viewpoint: author's viewpoint</td>
<td>what is the influence of the hypermedia system on the developability of the application?</td>
<td>Link generality, Effort</td>
</tr>
</tbody>
</table>

Table 6: The case study's goals, questions and metrics
Our measurement goal for the case study is to assess the development process for a hypermedia application, from the author's point of view. In order to achieve this goal it is necessary to know the influence of hypermedia system's and application's attributes on the development of hypermedia applications. Finally, metrics are proposed in order to measure these influences and give the necessary feedback towards the goal.

The level of maturity of the organisation considered in the survey was identified using the Capability Maturity Model (24) and is at level 1. The organisation, which is the University, did not give the students any strategies for project management. They were only given general guidelines for structuring their applications.

4.4 The empirical validation of the proposed metrics

The case study consisted of developing a hypermedia application to teach about Human-Computer Interaction concepts, where the application should be structured using the principles from the Cognitive Flexibility Theory (25) and should have at least 50 documents. The subjects of our evaluation were second year Computer Science students attending the Human-Computer Interaction course. They were split into two groups: group A, which developed the hypermedia applications using only the Web; and group B, which developed the hypermedia applications using both the Web and Webcosm. As it was important to guarantee that the levels of experience in Web authoring would be similar in both groups, stratified random samples (26) were used.

Before developing the applications, subjects were given an initial task, in order to avoid any confounding factors (27). The task involved the development of their personal Web pages. They were also pointed to two Web tutorials so that it would be possible for subjects with no Web authoring experience to learn how to develop Web pages. In addition, they all received training on the CFT authoring principles (90 minutes) and the set of subjects that would also use Webcosm received training in the use of that hypermedia system (90 minutes).

Two questionnaires were used in the case study: the first one asked subjects about their Web and Webcosm authoring experiences; the second one (see Appendix) gathered data about their applications (hyperdocument size, connectivty, compactness and stratrum) and also asked them how long, in hours, it had taken to develop the application. The development phases considered on the second questionnaire were adapted from existing literature (17, 28 - 34). Both questionnaires were checked by members of the research group for ambiguous questions, unusual tasks, definitions in the questionnaire's appendix and the number of questions.
We received 45 questionnaires from group A and 36 from group B. Five questionnaires from each group were incomplete and therefore invalid.

4.4.1 Validity of the Case study

There are three types of threats to the validity of an empirical study (27):

- Construct validity: Represents to what extent the independent and dependent variables precisely measure the concepts they claim to measure.
- Internal validity: Represents to what extent conclusions can be drawn about the casual effect of the independent variables on the dependent variables.
- External validity: Represents the domain to which a study's findings can be generalised.

The following problems might threaten the construct validity of the experiment:

1) There is no agreement on what constitutes the development process of a hypermedia application and, therefore, the development activities used in this study may not be representative of the majority in the hypermedia community. On the other hand, recent empirical work tends to support the development activities which were used (17,18). Therefore, the choice of development activities (subtasks) seems to be reasonable.

2) Although the subjects had to estimate the amount of hours used to author the applications, we believe that those estimates are reliable since they received the questionnaires at a very early stage and they were also asked to give estimates as accurate as possible.

The following problems might threaten the internal validity of the experiment:

1) There were three possible confounding factors in this evaluation: i) subjects experience; ii) maturation effects, which are learning effects caused by subjects learning as an experiment proceeds; and iii) tools used to help develop/manage the hypermedia application. The data collected showed that: i) although the authoring experiences of both groups did vary greatly, statistical analysis of variance (Student-t test, α=0.01) showed that there were no differences of population within either of the groups. Consequently, it was decided not to split any of the groups into sub-groups; ii) subjects were asked to develop their Web pages before developing the hypermedia applications and they all received training in the Cognitive Flexibility Theory (CFT) principles. Group B also received training in Webcosm. In addition, all the subjects were given a general structure for their applications, using the CFT. Consequently, it seems that the maturation effects were
controlled; and iii) the number of tools used by both groups did not vary greatly. All the results were confirmed statistically;

2) The selection effects, which are due to natural variations in subject performance, were controlled in both groups;

3) The instrumentation effects in general did not occur in this evaluation. The questionnaires used were nearly the same and both groups had to use essentially the same hypermedia system to develop their hypermedia applications – the Web.

The following problems might limit the external validity of the experiment:

1) The results may be domain dependent as all the subjects answered the questionnaires based on their experience in developing a hypermedia application for education. This evaluation should therefore be repeated in domains other than education if the results are to be generalised to other domains.

2) The experimental scale (35) is a threat when the experimental setting or the materials are not representative of industrial practice. The hypermedia applications developed presented good quality interface and contents, when compared to other hypermedia applications found in industry. Therefore, from this perspective it seems that the applications developed are representative of industrial practice.

3) The hypermedia applications developed did not have more than 100 documents and 300 links, on average, as the subjects were supposed to use only three working days to develop them. These numbers represent small hypermedia applications and are probably not representative of the large hypermedia applications for education developed by some organisations, in general. On the other hand, the hypermedia applications developed had a similar to or even better interface and contents quality than hypermedia applications developed by professionals, which seems to indicate that the results of the case study are likely to scale-up;

4) A threat to subject generalizability (35) may exist when the subject population is not drawn from the wider population. The subjects that participated in the case study may not be representative of hypermedia professionals, since second-year Computer Science students cannot be categorised as experienced hypermedia authors. The use of students as subjects, while sometimes considered unrealistic, is justified for two reasons: firstly, empirical evidence by Boehm-Davis and Ross (36) indicates that students are equal to professionals in many quantifiable measures, including their approach to developing software; secondly, for pragmatic considerations, having students as subjects was the only viable option for this case study. The efficacy of students as subjects is also supported by (37, 38).
4.4.2 Statistical results and analysis

To check the hypothesis related to the metric link generality the Student-t test was used, with a 2-tailed significance test at the 5% level. Point-to-point links were compared against generic links (number of links will usually in fact represent more than one point-to-point link in the hypermedia application). Consequently, to compare point-to-point and generic links, for each Web/Webcosm application the actual number of point-to-point links that each generic link lead to was used (see table 7).

<table>
<thead>
<tr>
<th></th>
<th>point-to-point links</th>
<th>generic links</th>
<th>Level of significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean/Median for number of links</td>
<td>94.93/50.00</td>
<td>263.20/98.00</td>
<td>0.03*</td>
</tr>
<tr>
<td>Mean/Median for number of hours</td>
<td>2.10/2.00</td>
<td>1.70/1.00</td>
<td>0.46</td>
</tr>
</tbody>
</table>

* the result is statistically significant at the 5% level

Table 7: Comparison between point-to-point and generic links

The results showed that the number of generic links created, compared to point-to-point links, for a similar amount of time, was a lot higher, suggesting that the use of generic links saves time and, consequently, confirming hypotheses H5. This result should be interpreted with caution as it does not imply that generic links should always be used. As noted earlier, hypothesis H5 only applies whenever the use of generic links is appropriate.

In order to check the hypotheses concerning the independent variables hyperdocument size, connectivity, stratum and compactness the non-parametric Spearman's rank correlation test was used, with a 1-tailed significance test at the 1% and 5% levels. Statistically

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Create</td>
<td>Create the text contents</td>
</tr>
<tr>
<td>Createg</td>
<td>Create the graphics contents</td>
</tr>
<tr>
<td>Createv</td>
<td>Create the video contents</td>
</tr>
<tr>
<td>Createa</td>
<td>Create the audio contents</td>
</tr>
<tr>
<td>Createc</td>
<td>Total time creating text, graphics, video, audio</td>
</tr>
<tr>
<td>Createp</td>
<td>Create the point-to-point links</td>
</tr>
<tr>
<td>Createg</td>
<td>Create the generic links (for Webcosm students)</td>
</tr>
<tr>
<td>Tcreate</td>
<td>Total time creating the links (point-to-point and generic)</td>
</tr>
<tr>
<td>Structap</td>
<td>Structure the application</td>
</tr>
<tr>
<td>Designip</td>
<td>Design the Interface: planning of the interface</td>
</tr>
<tr>
<td>Designii</td>
<td>Design the Interface: implementing of the interface</td>
</tr>
<tr>
<td>Testl</td>
<td>Test the application: testing the links in the application</td>
</tr>
<tr>
<td>Testm</td>
<td>Test the application: testing the media.</td>
</tr>
</tbody>
</table>

FIG. 3: Legend for the subtasks involved in the development of a hypermedia application
significant negative correlations for Stratum and statistically significant positive correlations for Hyperdocument size, Connectivity and Compactness helped to confirm our hypotheses.

The legend for the subtasks presented in tables 8 and 9 is described in figure 3.

Results obtained for Web subjects

<table>
<thead>
<tr>
<th>Tasks</th>
<th>Hyperdoc. size</th>
<th>Connectivity</th>
<th>Compactness</th>
<th>Stratum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Createt</td>
<td>0.52**</td>
<td>0.34*</td>
<td>0.31*</td>
<td>0.13</td>
</tr>
<tr>
<td>Createg</td>
<td>0.09</td>
<td>0.27*</td>
<td>0.20</td>
<td>-0.11</td>
</tr>
<tr>
<td>Createv</td>
<td>nc</td>
<td>nc</td>
<td>nc</td>
<td>nc</td>
</tr>
<tr>
<td>Createa</td>
<td>-1.00</td>
<td>-1.00</td>
<td>-1.00</td>
<td>-1.00</td>
</tr>
<tr>
<td>Createc</td>
<td>0.52**</td>
<td>0.33*</td>
<td>0.33*</td>
<td>0.05</td>
</tr>
<tr>
<td>Createl</td>
<td>0.31*</td>
<td>0.25</td>
<td>0.11</td>
<td>0.20</td>
</tr>
<tr>
<td>Structap</td>
<td>0.22</td>
<td>0.16</td>
<td>0.03</td>
<td>0.15</td>
</tr>
<tr>
<td>Designp</td>
<td>0.44**</td>
<td>0.30*</td>
<td>0.35*</td>
<td>0.11</td>
</tr>
<tr>
<td>Designii</td>
<td>0.26</td>
<td>0.34*</td>
<td>0.42**</td>
<td>0.31*</td>
</tr>
<tr>
<td>Test1</td>
<td>0.52**</td>
<td>0.42**</td>
<td>0.44**</td>
<td>0.03</td>
</tr>
<tr>
<td>Testm</td>
<td>0.05</td>
<td>0.20</td>
<td>0.23</td>
<td>0.62</td>
</tr>
</tbody>
</table>

* denotes that the result is statistically significant at the 1% level
** denotes that the result is statistically significant at the 5% level
nc – could not be computed

Table 8: Spearman’s correlation coefficient for metrics vs. time for Web subjects

The results obtained (see table 8) show several statistically significant correlations between effort and hyperdocument size, connectivity and compactness, confirming hypotheses H1, H2 and H3.

The elapsed time spent in implementing the interface showed a statistically significant positive correlation with stratum, suggesting that it took longer to implement the interface of a hypermedia application developed with a more linear style (concerning its navigation). However, this result was not sufficient to reject hypothesis H4. Although the correlation between stratum and Testm was 0.62, the result was not statistically valid due to the small sample size used in the calculations.

Results obtained for Web/Webcosm subjects

The results obtained (see table 9) show several statistically significant correlations between effort, hyperdocument size and connectivity, confirming hypotheses H1 and H2.

Summary of the Results

Hyperdocument size and connectivity were both confirmed as possible metrics by A and B groups, suggesting that the number of documents and...
<table>
<thead>
<tr>
<th>Tasks</th>
<th>Hyperdoc. size</th>
<th>Connectivity</th>
<th>Compactness</th>
<th>Stratum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Create</td>
<td>0.62**</td>
<td>0.29</td>
<td>-0.15</td>
<td>0.05</td>
</tr>
<tr>
<td>Createg</td>
<td>0.42**</td>
<td>0.22</td>
<td>0.11</td>
<td>-0.25</td>
</tr>
<tr>
<td>Createv</td>
<td>nc</td>
<td>nc</td>
<td>nc</td>
<td>nc</td>
</tr>
<tr>
<td>Createa</td>
<td>0.06</td>
<td>1.00**</td>
<td>0.97</td>
<td>0.96</td>
</tr>
<tr>
<td>Createc</td>
<td>0.60**</td>
<td>0.32*</td>
<td>-0.05</td>
<td>-0.03</td>
</tr>
<tr>
<td>Createl</td>
<td>0.15</td>
<td>0.25</td>
<td>0.06</td>
<td>-0.29</td>
</tr>
<tr>
<td>Creategl</td>
<td>0.32</td>
<td>0.35</td>
<td>0.33</td>
<td>0.15</td>
</tr>
<tr>
<td>Toreatel</td>
<td>0.42**</td>
<td>0.52**</td>
<td>0.19</td>
<td>-0.13</td>
</tr>
<tr>
<td>Structap</td>
<td>0.29</td>
<td>0.68**</td>
<td>0.24</td>
<td>0.26</td>
</tr>
<tr>
<td>Designp</td>
<td>0.28</td>
<td>0.52**</td>
<td>0.25</td>
<td>0.15</td>
</tr>
<tr>
<td>Designi</td>
<td>0.27</td>
<td>0.54**</td>
<td>0.12</td>
<td>-0.21</td>
</tr>
<tr>
<td>Testl</td>
<td>0.16</td>
<td>-0.10</td>
<td>0.00</td>
<td>-0.03</td>
</tr>
<tr>
<td>Testm</td>
<td>0.25</td>
<td>0.37</td>
<td>0.32</td>
<td>-0.56</td>
</tr>
</tbody>
</table>

** denotes that the result is statistically significant at the 1% level
* denotes that the result is statistically significant at the 5% level
nc - could not be computed

Table 9: Spearman's correlation coefficients for metrics vs. time

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Metric</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1</td>
<td>Hyperdoc. size</td>
<td>Confirmed by Web and Web/Websom subjects</td>
</tr>
<tr>
<td>H2</td>
<td>Connectivity</td>
<td>Confirmed by Web and Web/Websom subjects</td>
</tr>
<tr>
<td>H3</td>
<td>Compactness</td>
<td>Confirmed by Web subjects</td>
</tr>
<tr>
<td>H4</td>
<td>Stratum</td>
<td>Not enough evidence to either confirm or reject the hypothesis</td>
</tr>
<tr>
<td>H5</td>
<td>Link generality</td>
<td>Confirmed</td>
</tr>
</tbody>
</table>

Table 10: Results for the hypermedia system's independent variable

Links are related to the effort involved in developing Web applications. Although group A confirmed compactness as a metric, it was not confirmed by group B. This result may suggest that compactness may not be useful to measure development effort when the hypermedia system used stores the links in linkbases, as Websom does. A replication of the case study using another hypermedia system, e.g. HyperWave (39), where links are stored separately from the documents, would help to confirm this hypothesis.

The lack of enough evidence to confirm hypothesis H4 on both groups may be related to the fact that all the applications were structured according to the CFT principles, which could have determined a similar stratum for all the applications. A replication of the case study without structuring the applications the same way would help to confirm whether or not stratum can be used to measure development effort.
4.4.3 Recommendations to Improve Hypermedia Authoring

There are some recommendations for improving hypermedia authoring, based on the results obtained from the case study. It seems that both Webcosm and Web subjects have, in general, confirmed that hyperdocument size, connectivity and compactness are related to the time spent in authoring a hypermedia application. In practice these results may suggest that:

- Hypermedia applications that have larger number of documents will tend to take longer to author. Our recommendation is that planning the authoring process in advance may help reduce the time spent in authoring the documents. By planning authors can identify, for example, similar nodes in which contents and/or interface components could be reused. In addition, if authors organize authoring into different phases (interface design, structural knowledge design and authoring of the contents of the documents), different groups of authors can tackle different parts of the project, which may reduce the time spent in authoring.

- Hypermedia applications that have higher number of links will tend to take longer to author. Our recommendation is that certain information about the links could be used to improve authoring. Examples of useful information are types and context added to links. Links within contexts may also reduce the time spent in authoring as they do not need to be updated in every document in which they are referenced. This has been confirmed by the hypothesis link generality.

- Hypermedia applications that have higher compactness will tend to take longer to author. Trying to reduce the coupling among documents is not the way to improve authoring as one of the most important aspects of hypermedia is its rich linking capabilities. However, using links within contexts can help not only to document the application, but also to organize the cross-referencing within the application. Consequently, the time spent in authoring links can be reduced.

5. CONCLUSIONS

This paper has reported the findings of a quantitative case study aimed at measuring the development effort involved in authoring Web applications. The case study involved the development of Web applications using the principles of the Cognitive Flexibility Theory. The data was collected using questionnaires, which were answered by second-year Computer Science students.

Results have shown that, for the applications developed using a standard Web environment, apart from the stratum metric, all the other metrics were
confirmed. For the applications developed using the Web/Webcosm, the metrics hyperdocument size and connectivity were confirmed. These results have confirmed that there are associations between some of the metrics and development effort. However, further experimentation is necessary in order to predict the development effort by proposing a formula that, using the metrics as input, gives as output a number which represents the amount of time necessary to develop the application.

An important lesson learnt from the use of measurement is that it is better to start from a set of metrics addressing important improvement areas, and evolve these metrics over time, instead of debating forever, trying to find perfect metrics (40). The metrics proposed in this paper, if considered in isolation from one another, may look intuitively obvious and possibly meaningless as measures. However, they should be considered together, as a vector of points in a multidimensional space. In addition, the emphasis of our work was on the methodology used in determining and validating the metrics. We believe that the steps used to propose and validate the metrics were as important as, or even more important than, the metrics themselves.

From a practical point of view, the results presented here can be used to: i) understand better the relationship between some attributes of hypermedia applications and the development effort and consequently, improve authoring; ii) help establish a baseline against which other projects can be compared; iii) increase the maturity level of the development process, and, consequently, increase the range of possible measurements; iv) use the data obtained to propose a development effort estimation model.

We have identified and measured in our project some important quality characteristics of hypermedia applications. However, a complete framework is useful for specifying and evaluating hypermedia applications objectively and quantitatively. A hypermedia quality model will help determine what other quality characteristics, and corresponding metrics, are important in order to assess and improve hypermedia authoring. The work reported here is therefore just the first step on what may prove to be a long, but we hope very fruitful, road.

REFERENCES

126

The New Review of Hypermedia and Multimedia 1999


The New Review of Hypermedia and Multimedia 1999
APPENDIX

Questionnaire given to Web/Webcosm subjects. The questionnaire used with Web students was basically the same, apart from questions 7 and 12.f.

Name: 
Surname: 

The terms underlined are defined in the appendix.

For all time-related questions, please supply your estimate of elapsed time, in hours, considering only the mechanical aspects of the hypermedia design tasks and not the time it took you to research the contents.

1) How many web pages does your application have? ____________
2) How many web pages did you develop from scratch? ____________
3) How many web pages did you reuse from other sites? ____________

4) If you reused a web page, did you do this by (tick more than one if necessary):
   making a link to it from your application ( )
   making a local copy of that page and using it ( )

5) When reusing a web page did you (tick more than one if necessary):
   add new links to it ( )
   not add new links to it ( )

6) Approximately how many point-to-point links did you create? ____________
7) Approximately how many generic links did you create? ____________

8) What is the compactness of the application?

   Completely Connected 5 4 3 2 1 Completely Disconnected

9) What is the stratum of the application?

   High stratum 5 4 3 2 1 No stratum

10) Please circle the structure which best describes how the core documents of the application are organised:
    Sequence  Hierarchy  Network

    The New Review of Hypermedia and Multimedia 1999

129
11) For each of the following tasks, please estimate the time it took, to:

Create all:
  a) text contents
  b) graphics contents
  c) video contents
  d) audio contents

Create all:
  e) point-to-point links
  f) generic links

Structure the application

Design the Interface:
  g) planning of the interface
  h) implementing of the interface

Test the application:
  i) testing the links in the application
  j) testing media, i.e., testing if
    the video works, the sound
    works, etc.

12) What tools did you use to develop the application and for what purpose? Tick more than one if appropriate

  Purpose

  ( ) An HTML editor

  ( ) An application generator

  ( ) Software to organise and manage the HTML files

  ( ) if other, which one?
APPENDIX

Compactness - the connectedness of the hyperdocument. A high compactness indicates that from any document a user can easily reach, using links, any other document in the application, suggesting a large amount of cross-referencing.

Generic link - is a link defined using the link editor and, once created, is valid within any web document where there is text matching the generic link. All the generic links are stored in a linkbase.

Hierarchy - a structure of a hyperdocument where the documents are organised as a tree (see figure 1).

Hyperdocument - set of documents and respective links.

Network - hyperdocument where its structure is arranged as a net of documents and links (see figure 2).

Point-to-point link - is the link created in HTML using a HREF tag. It is not stored in a linkbase, but stays embedded within the document.

Sequence - a structure of a hyperdocument where the documents are linearly linked (see figure 3).

The New Review of Hypermedia and Multimedia 1999
**Stratum** - suggests if there is an order for reading the hypertext document. Minimum stratum means that structurally it makes no difference from what document one starts reading. Maximum stratum is achieved in a linear hypertext document (the topology is a *sequence*).

**Structure** - the arrangement of the core documents of the application. They can be arranged as a *hierarchy*, a *sequence*, or a *network*. 