Journal of Universal Computer Science, vol. 4, no. 4 (1998), 382-403
submitted: 10/12/97, accepted: 28/12/97, appeared: 28/4/98 *© Springer Pub. Co.*

Applying Metrics to the Evaluation of Educational Hypermedia Applications

Emilia Mendes
(Multimedia Research Group, University of Southampton, UK
mexm95r@ecs.soton.ac.uk)

Wendy Hall
(Multimedia Research Group, University of Southampton, UK
wh@ecs.soton.ac.uk)

Rachel Harrison
(Declarative Systems and Software Engineering Group, University of Southampton, UK
rh@ecs.soton.ac.uk)

Abstract: This paper reports the results of applying metrics to hypermedia authoring under the SHAPE research project. The aim of SHAPE is to help authors develop high quality large hypermedia applications for education. The quality characteristics considered are the reusability of information, the maintainability of applications and the development effort. Although a number of metrics for hypertext systems have been proposed, we believe that many of the measures proposed in the past lack the necessary mathematical and/or empirical justification. The metrics proposed in this paper have been developed using the Goal-Question-Metric approach, and adhere to the representational theory of measurement. We describe the development of the metrics and the results of a quantitative empirical study which compares two different hypermedia authoring systems.

Key Words: Hypermedia, reusability, maintainability, hypermedia authoring, metrics, evaluation.

1 Introduction

Hypermedia authoring has been a wide area of research and interest in the last few years. There are many ways in which the meaning of authoring can be understood and, consequently, many corresponding solutions offered to hypermedia authors. This is reflected in the literature by the many different proposals of hypermedia authoring models [Garzotto et al. 91], [Rossi et al. 95], methodologies [Balasubramanian et al. 94], model oriented environments [Türing et al. 95], [Nanard and Nanard 95], [Jordan et al. 89], [Andrews et al. 1995a], [Marshall et al. 91], [Marshall et al. 95], [Marmann et al. 92], [Duval and Olivié 95], [Catlin and Garrett 91] and general-purpose environments [Davis et al. 92], [Meyrowitz 86], [Bernstein et al. 91], [Goldberg et al. 96], [Thimbleby 96], [Andrews et al. 95b].

The process of hypermedia authoring leads to various products, for example, a specification, a design and a hypermedia application. Understanding the process

contributes to its control and improvement, and one way of doing that is using measurement.

Measurement can be used to [Basili et al. 94]:

i) Support project planning.
ii) Determine the strengths and weaknesses of the current processes and products.
iii) Provide a rationale for adopting/refining techniques.
iv) Evaluate the quality of specific processes and products.
v) Assess the progress of a project during its course.
vi) Take corrective action based on this assessment.
vii) Evaluate the impact of such action.

Although the application of metrics to hypermedia has already stimulated considerable interest [Botafogo et al. 92], [Rivlin et al. 94], [Garzotto et al. 94], [Garzotto et al. 95], [Hatzimanikatis et al. 95], [Yamada et al. 95], they have been developed in an ad-hoc fashion, expressing measures in an ambiguous manner and thereby limiting their application. For example, there are many different decisions that have to be made when defining a usability measure or maintainability measure. These decisions have to be made with respect to the goal of the measure and by defining an empirical model based on a hypotheses. Unfortunately, many measures proposed in the literature do not have the motivation behind these decisions documented, making it difficult to understand the underlying assumptions of the measure.

This paper presents the metrics developed so far under the SHAPE research project. In Section 1, we give an overview of the research project SHAPE and the first empirical study within it. In Section 2, we give a brief overview of hypermedia metrics proposed in the literature, followed by, in Section 3, the definition of our metrics. Section 4 explains our first evaluation of the metrics proposed, and in Section 5 we present the results. Finally, in Section 6 we give our conclusions and comments on future work.

2 The SHAPE Project

Southampton **H**ypermedia **A**uthoring **P**aradigm for **E**ducation (SHAPE) [Mendes and Hall 97a], [Mendes and Hall 97b] is a research project being carried out at the University of Southampton. The aim of SHAPE is to help authors in the development of high quality large hypermedia applications for education. The quality characteristics considered are reusability of information, maintainability of applications, and development effort.

2.1 The First Empirical Study within SHAPE

Our first idea towards SHAPE was to create a hypermedia authoring methodology to be used with Microcosm [Davis et al. 92]. Consequently, apart from a literature review, we conducted an empirical study [Mendes and Hall 97a], [Mendes and Hall 97b] aimed at analyzing the cognitive processes involved in the authoring of hypermedia

applications for education. The empirical study was carried out using interviews, and the participants were either researchers or lecturers involved in the authoring of hypermedia applications for education: thirteen people volunteered. The main conclusions of the empirical study were:

2.1.1 In relation to a general methodology

We could not identify any general methodology, but the top-down approach was widely used as a way of planning, organizing thoughts and structuring knowledge. Links were not considered essential when structuring the knowledge domain. Authors planned the knowledge structure beforehand rather than considering links in order to know how to structure the information.

The authors in our study do not think that there was a common way of structuring all the domains. This reinforces our conclusion that one single hypermedia authoring methodology is not what hypermedia authors need.

Our study led us to deduce that authors' cognitive writing processes remain the same when considering hypermedia authorship. They do not seem to need to know what they are going to link before structuring their knowledge: they prepare the overall structure first and only as a second stage plan the links. So, what is necessary to complement this process is a rhetoric of links, rather than a general rhetoric, as discussed by Landow [Landow 87].

The authors answered that the development of a hypermedia application can be divided into different phases and amongst groups of authors, as long as they interact among themselves. The phases considered were interface design, structural knowledge design, and authoring of the contents of the nodes. This gives some insight into co-operative authoring and what kind of activities different groups could be involved in to develop an application co-operatively.

2.1.2 In relation to the type of use

All the authors in the study used hypermedia applications as a complement to their classes, which means that they see technology as an aid to the lecturer rather than as a substitute. The instructional approach mostly used was problem-solving which reflects the teaching methodology of the volunteer authors in our study.

2.1.3 In relation to hypermedia design

Cognitive maps, contextual links, and typed links are important for knowledge structuring and for conveying local as well as global coherence to learners. This implies that authors should seek to make the knowledge as explicit as possible in order to give both coherence and context to the learners. Text was the resource mostly used for conveying information to learners, followed by images. If text is still widely and heavily used all the research that already exists about good writing can be immediately applied.

2.2 Re-thinking SHAPE after the Empirical Study

According to Thüring every hypermedia application can be developed within one of the following scenarios [Thüring 94], [Schuler and Thüring 94]:

1) Turning text into hypertext. All the material is already available in a pre-structured format: for example, as electronic documents, lecture notes, in a book, as a manual, etc.
2) Synthesizing hypertext from heterogeneous sources. All or nearly all of the material is available but does not form a coherent entity: for example, different books, articles, and video clips that have to be aggregated into a hypermedia application.
3) Designing from scratch. None or only very few materials are available. The designer has to create most of the content as well as build a complete new structure on her/his own.
4) Re-engineering an application. There is already a hypermedia application which has to be updated or thoroughly revised.

The applications considered in the empirical study were examples of scenarios 1,2, and 3. From the results of that study it seems that, at least for those three scenarios, there is little point in trying to improve the authoring of hypermedia applications for education by proposing an authoring methodology. Authors do not seem to change their cognitive writing processes because they are using hypermedia, and they often have their own usually pre-determined way of organizing the application's structure.

But, whatever the scenario, there are four important issues that arise which can influence both the quality of the application and the quality of the development process:

i) Development effort.
ii) Reusability of information.
iii) Maintainability of the application.
iv) Contents.

In relation to the contents, which - in the case of hypermedia applications for education - reflects a particular theory of learning, there are two theoretical views that have particular relevance to the design and use of hypermedia applications: The Cognitive Flexibility Theory (CFT) [Jacobson and Spiro 95a], [Jacobson and Spiro 95b], [Spiro et al. 95] and the Situated Cognition Theory (SCT) [Jacobson et al. 96], [Jacobson and Archodidou 97]. They are both theories of learning and incorporate not only the planning principles, but they allow deep learning of complex domains and the transfer of knowledge to new situations. They have been tested by empirical evaluations and have had very positive results so far [Rana and Bieber 97], [Carvalho and Dias 97], [Jacobson 97], [Moreira 96].

As the CFT and SCT seem to be adequate in helping authors in the design of the contents for hypermedia applications for education, we decided to focus the SHAPE

project on the reusability of information, development effort, and maintainability of hypermedia applications. We believe that the results of SHAPE will be both a complement to the work being done with the CFT and SCT and a help to authors in the development of high quality large-scale hypermedia applications for education. The relationship between the three issues is shown in Figure 1.

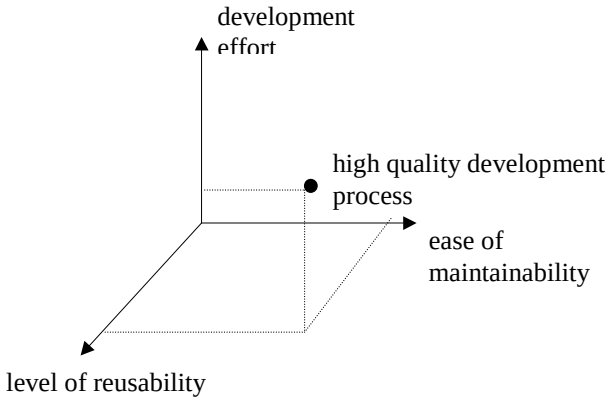

Figure 1: Important Issues for High Quality Development

Our choice was to apply a scientific approach to hypermedia authoring for education. So, rather than defining improvements to be applied to Microcosm and later verifying if they are adequate, we have decided to use a more consistent and systematic approach, which is to develop metrics in order to measure how adequate Microcosm is for the maintainability of applications, information reuse in applications, and the level of development effort required.

As we had no baseline to compare our results with, we chose to compare Microcosm with the Web [Berners-Lee et al. 94]. The two systems propose different ways of representing and managing links, and this seems to have a big influence on authoring [Hill et al. 95]. Microcosm is an open environment, characterized by the separation of link structures from the information being linked [Hill et al. 95]. The WWW, on the other hand, provides a simple point-to-point linking model based upon embedded links.

3 Metrics in Hypermedia

The application of metrics to hypermedia has already stimulated considerable interest [Botafogo et al. 92], [Rivlin et al. 94], [Garzotto et al. 94], [Garzotto et al. 95], [Hatzimanikatis et al. 95], [Yamada et al. 95]. However, little corresponding empirical validation of these metrics has been published.

Table 1 compares the four proposed hypertext metrics considering the four questions that should be asked when validating a measure [Briand et al. 97]:

1) Is the measure adequately capturing the attribute it purports to measure (i.e., construct validity)?
2) Is the attribute itself well-defined based on an explicit empirical model (i.e., empirical relational system) ?
3) Is there any empirical evidence supporting the underlying hypotheses of the empirical model?
4) Is the measure useful from a practical perspective?

'No' means that the characteristic has not been fulfilled by the proposal and 'Yes' means that it has been fulfilled.

	Metrics		Proposals	
Questions	[Botafogo et al. 92]	[Garzotto et al. 94]	[Hatzimanikatis et al. 95]	[Yamada et al. 95]
(1)	Yes	No	No	Yes
(2)	No	No	No	Yes
(3)	No	No	No	Yes
(4)	Yes	Yes	Yes	Yes

Table 1: Comparison of the proposals

Garzotto at al. [Garzotto et al. 94], [Garzotto et al. 95] did not define internal attributes that could be measured in an empirical evaluation. They considered the involvement of end users to be unnecessary since they see their work as complementary to Human Computer Interface methods that evaluate quality factors (such as usability). A relevant point to consider is that without empirical evaluation there is no real data to prove the usefulness of the metrics proposed.

Yamada et al. [Yamada et al. 95] defined a set of metrics based on some assumptions concerning navigation and cognitive load. In order to validate their metrics, they developed three hypertext applications: two were card-type interfaces and the other one was scene-selection-type interface. They wanted to compare two different menu styles using those three applications, But they structured all the three applications using the same menu style. Consequently, the results obtained are not useful to test their hypothesis.

The metrics proposed by Botafogo et al. [Botafogo et al. 92], [Rivlin et al. 94] were neither based on an empirical model nor tested by empirical evaluation. Although they have already been mentioned in the literature [Adams and Jr 97], [Calvi and DeBra 97], their usefulness has not yet been evaluated. According to Calvi & DeBra [Calvi and DeBra 97] "while these methods are able to find link structures which are likely to be unusable, they cannot guarantee that link structures having all suggested values for different metrics will actually belong to highly usable hyperdocuments".

Hatzimanikatis et al. [Hatzimanikatis et al. 95] did not define any empirical relationships for their proposed metrics. They were not able to present precise limits or ranges for acceptable values of the metrics because they believed that acceptable

values could vary according to the application, the authoring tools used and the production environment. Empirical evidence would help to provide baselines for these metrics.

4 Applying Metrics to SHAPE

The principles of the metrics we have developed [Mendes 97] are based on the goal-based framework for software measurement proposed by Fenton and Pfleeger [Fenton and Pfleeger 96], and on the guidelines from the DESMET project [Kitchenham 96], [Kitchenham 93]. Both have been extensively used in experiments in the software engineering field [Harrison et al. 95], [Daly 96], [Briand et al. 96], [Basili and Rombach 88], [MacDonell 91].

We have planned two evaluations for SHAPE. The first was a quantitative evaluation and the second is both quantitative and qualitative. In the next sub-section, we describe our metrics for the first evaluation.

4.1 Entities to be Examined

The conceptual framework proposed by Fenton and Pfleeger [Fenton and Pfleeger 96] can be applied to the diverse software-measurement activities that contribute to an organization's software practices. The 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. It is based on three principles:

i) Classifying the entities to be examined.
ii) Determining relevant measurement goals.
iii) Identifying the level of maturity that an organization has reached.

The entities considered for our first evaluation were:

i) Application.
ii) Tool.
iii) Maintenance.
iv) Reuse.
v) Authors.

To measure the Maintainability and Reusability of a hypermedia application, we proposed the following independent variables:

i) Size of the application.
ii) Connectivity.
iii) Structure (topology).
iv) Compactness [Botafogo et al. 92], [Vocht 94].
v) Stratum [Botafogo et al. 92], [Vocht 94].

Compactness indicates "the intrinsic connectedness of the hypertext" [20,54], and stratum reveals "to what degree the hypertext is organized so that some nodes must be read before the others" [Botafogo et al. 92], [Vocht 94].

To measure the entity Tool we proposed the following independent variables:

i) Link representation.
ii) Link type.
iii) Highlighting of anchors.

Link Representation means whether the links are "embedded" within the document or not. The highlighting of anchors refers to whether or not the anchors are explicitly presented to the readers (using a different color, for example). Our hypothesis is that being able to see an anchor can influence both the maintenance and reuse of the corresponding link.

To measure the entity Author we proposed the independent variables:

i) Role.
ii) Experience.

To measure the entities Maintenance of the application and Reuse of information we used the following dependent variables:

i) Time.
ii) Difficulty.

4.2 Relevant Measurement Goals

The relevant measurement goals were determined using the Goal-Question-Metric (GQM) approach [Basili et al. 94], which is based upon the assumption that any measurement must be defined in a top-down fashion. The result of applying the GQM approach is a model that has three levels: i) The conceptual level - Goal; ii) The operational level - Question; and iii) The quantitative level - Metric. The goal is refined into several questions and each question is then refined into metrics, either objective or subjective. Shape's corresponding GQM is presented below:

Goal: To evaluate the quality of the hypermedia application, from the authors' viewpoint

Question: What is the quality of the hypermedia application?

Metrics: Maintainability, reusability

Goal: To improve the maintenance of hypermedia applications and the reuse of information

Question: What is the influence of the tool on the maintainability/ reusability?

Metrics: Highlighting of Anchors, representation of links, type of links

Question: What is the influence of the application on the maintainability/ reusability?

Metrics: Size of the application, connectivity, structure of the application, compactness, stratum

Question: What is the influence of the author on the maintainability/reusability?

Metrics: Role, experience

4.3 The Maturity Level

The level of maturity within the hypermedia application development community that is considered for SHAPE is either level 1 or 2. Level 1 typically means that an organization does not provide a stable environment for developing and maintaining software. Level 2 means that there are policies for managing a software project and procedures to implement those policies are established.

5 The First Evaluation of Metrics within SHAPE

5.1 The Design

For the first evaluation the stated hypothesis was:

H1- Microcosm applications are more maintainable and their information more reusable than applications built using a standard WWW environment.

Using the data collected we also wanted to evaluate if :

i) The use of a link service allows both a better maintainability of applications and reusability of information than embedded ones.
ii) Generic links allow a better maintainability and reusability of information than the equivalent set of point-to-point links.

In the Microcosm model, a link associates a particular source selection with its destination and can be specific (point-to-point), local or generic. A local link can be followed from any occurrence of the source selection in a particular document [Davis et al. 92]. In the standard implementation of Microcosm local link anchors are not highlighted. A generic link can be followed from any occurrence of that source selection in any document [Davis et al. 92]. In the standard implementation of Microcosm they are also not highlighted

5.2 The Method

The survey involved the use of questionnaires completed by either Microcosm or Web authors.

A survey approach was chosen because it offers the following advantages [Kitchenham 96]: i) reaches many users; ii) makes use of existing experience; iii) makes use of standard statistical analysis techniques; and iv) confirms that an effect generalizes to many projects/organizations.

Both questionnaires had three sections: reusability, maintainability, and experience. Our understanding of reusability and maintainability is presented in the next sub-sections.

5.3 Reusability applied to SHAPE

Reuse is "the use of everything associated with a software project including knowledge" [Basili and Rombach 88], and reusability is the "degree to which a thing can be reused" [Frakes and Terry 95]. Reusability metrics indicate "the likelihood that an artifact is reusable" [Frakes and Terry 96].

We prepared the questionnaire considering four different, but complementary, classifications for reusability/reuse. These classifications represent the work done by Ruben Prieto-Diaz [Prieto-Díaz 93], Frakes & Terry [Frakes and Terry 95], [Frakes and Terry 96], Bieman & Karunanithi [Karunanithi and Bieman 93]and [Garzotto et al. 96].

We adapted the four classifications mentioned above and the resultant classification is presented in Table 2. This classification represents what it is important to consider for SHAPE, concerning reusability.

Facets of Reuse	Type chosen
Perspective	Server
Development Scope	Private, Public
Implementation	by value, by reference
Reused Entity	document, link
Domain Scope	vertical
Modification	white-box, black-box
Management	ad hoc

Table 2: Classification of reusability applied to SHAPE

The server perspective is similar to a software library or a particular software library component [Karunanithi and Bieman 93]. We chose this perspective since our scenario considers that the hypermedia application or any of its components (the server) will be reused by other applications (the clients).

Our scope of reuse is vertical. Vertical reuse is the reuse of software within the same domain or application area, and its goal is to derive generic models for families of systems that can be used as standard templates for assembling new systems.

5.4 Maintainability applied to SHAPE

For SHAPE we are considering the definition and classification of maintainability as it is presented in the ISO/IEC 9126. Maintainability is defined as "a set of attributes that bear on the effort needed to make specified modifications" [ISO/IEC 91]. Its sub-characteristics are:

1) Analyzability: attributes of software that bear on the effort needed for diagnosis of deficiencies or causes of failures, or for identification of parts to be modified.
2) Changeability: attributes of software that bear on the effort needed for modification, fault removal or for environmental change.
3) Stability: attributes of software that bear on the risk of unexpected effect of modifications.
4) Testability: attributes of software that bear on the effort needed for validating the modified software.

In order to prepare the maintainability and the reusability sections we had also to consider common tasks accomplished by authors in the development of hypermedia applications for education.

5.5 The Pilot Study

Before sending the questionnaires to both Microcosm and Web authors, we carried out a pilot study where the questionnaires were answered by a group of five people. They all had previous experience in the development of hypermedia applications for education, using either Microcosm or the Web. Their feedback concerned:

i) Ambiguous questions.
ii) Unusual tasks.
iii) Definitions in the appendix.
iv) Number of questions.

6 The Results

The survey results were analyzed using standard statistical techniques. To determine whether the two sets of questionnaires (from Microcosm and Web authors) were from different populations, we generated all the levels of significance using the Kruskal-Wallis one-way analysis of variance, with a level of significance of 5% and 10%. The Kruskal-Wallis one-way analysis of variance is an extremely useful test for deciding whether k independent samples are from different populations or whether they represent merely chance variations among random samples from the same population.

To identify the correlation between the independent and dependent variables we used Gamma as a measure of correlation, with a level of significance of 10%. Gamma gives in a single number a summary measure of the existence, strength, and direction of the relationship [Healey 93].

We analyzed 44 questionnaires - 16 from Microcosm authors and 28 from Web authors. Both groups shared similar experiences and levels of involvement in the development of the applications. No statistically significant differences were found. The median for the experiences of authors was, in an interval of 1 to 5, 4 and 3 for respectively, the Web and Microcosm.

The applications developed by either Web authors or Microcosm authors shared similar compactness, stratum, size of the application, connectivity, and structure of the applications. No statistically significant differences were found.

Both groups made use of various planning methods for the development of their applications.

The structure that was used the most was the hierarchical, as we can see from the data in Table 3:

Structure	Microcosm percentage %	Web percentage %
Sequential	5.5	04
Hierarchical	67	64
Network	22	25
No answer	5.5	07
	100	100

Table 3: Type of structure used by both groups

There was a statistically significant difference at the 5% level between the number of tools used by Web authors and Microcosm authors. Web authors used a higher number of tools than Microcosm authors. The tools mentioned in the Web questionnaire were: An HTML editor, an application generator, and software to organize and manage the HTML files. The tools mentioned in the Microcosm questionnaire were: a link editor, a document management system, and a word processor.

We measured the two dependent variables - time and level of difficulty - using a questionnaire with 15 questions. These questions are presented in the appendix. Thirteen questions were based on the usual tasks concerning maintenance and reuse. As we did not want to bias the evaluation, only two questions were developed where the tasks involved might be more effectively accomplished using generic or local links. These were questions 12 and 13 respectively .

When comparing tasks involving point-to-point links in both Microcosm and the Web, we found that in 33% of the answers the medians for the level of difficulty were lower for Microcosm than for the Web and in 46% of the answers the time was shorter.

In 46% of the answers the time spent in both Microcosm and the Web was the same. But Web authors needed to use an auxiliary set of tools in order to accomplish the tasks in a reasonable time and with a low level of difficulty. This was not necessary using Microcosm.

Even with 7 answers where the level of difficulty was higher for Microcosm than for the Web, there was no corresponding increase in the time spent to accomplish the

tasks. As Microcosm is an open system, the author has to edit the linkbase many times in order to maintain links. This task can be considered more difficult than changing links on the Web, but, as shown by the data, there is no overhead on the time spent.

When comparing tasks involving point-to-point links in both Microcosm and the Web, we also found 8 answers with a statistically significant difference. Four showed advantages for the Web and four showed advantages for Microcosm. The Medians for tasks involving Microcosm point-to-point links (Median point-to-point Microc.), Web point-to-point links (Median point-to-point Web) and the corresponding level of significance (Level Sig.) are presented in Table 4.

Quest.	Attribute	Median point-to-point Microc.	Median point-to-point Web	Level Sig.
02	Time	1	2.5	0.04*
05	Difficulty	2	1	0.00*
06	Difficulty	2	1	0.03*
08	Time	1	3	0.03*
12	Difficulty	1	2	0.04*
13	Difficulty	1	2	0.00*
14	Difficulty	3	1.5	0.03*
15	Difficulty	2	1	0.00*
*denotes that the result is statistically significant at 5% level				

Table 4: Medians for tasks involving point-to-point links in Microcosm and the Web, with corresponding level of significance.

Questions 5,6,14 and 15 (presented in the appendix) represent simple tasks, but for Microcosm authors they involve the editing of the linkbase in order to update the information about the links. We understand that this was the reason for a higher level of difficulty using Microcosm. But, even with a higher level of difficulty, no statistically significant differences were found when comparing the time involved in the same tasks.

Questions 2, 8 (presented in the appendix) showed a statistically significant difference in the time spent in accomplishing the tasks. The time was higher using the Web. Questions 12 and 13 (presented in the appendix) also showed a statistically significant difference in the level of difficulty spent in accomplishing the tasks. The level of difficulty was higher using the Web. Questions 12 and 13 would be easily accomplished (in Microcosm) using generic links for the former question and local links for the latter question.

For 13 questions that were not specifically designed considering tasks that would be better suited for generic or local links, Microcosm authors were asked to estimate the time and level of difficulty in accomplishing the tasks if the links were either point-to-point or generic.

When comparing the answers given by Microcosm authors for tasks involving generic links to the same tasks involving point-to-point links on Web, we found 10 answers with a statistically significant difference. All the 10 answers showed advantages for generic links. The medians for generic links (Median Generic

Microc.), medians for point-to-point links on the Web (Median point-to-point Web) and the corresponding level of significance (Level Sig.) are presented in Table 5.

Quest.	Attribute	Median Generic Mircroc.	Median point-to-point Web	Level Sig.
03	Time	1.0	1.5	0.00*
04	Time	0.5	1.0	0.04*
05	Difficulty	1.0	1.0	0.00*
08	Time	1.0	3.0	0.00*
09	Time	1.0	2.0	0.03*
10	Time	1.0	2.0	0.07**
12	Time Difficulty	2.0 1.0	3.0 2.0	0.07** 0.00*
13	Time	1.0	2.0	0.08**
	Difficulty	1.0	2.0	0.00*
*denotes that the result is statistically significant at 5% level **denotes that the result is statistically significant at 10% level				

Table 5: Medians for tasks involving generic links and point-to-point links, with corresponding level of significance.

We can see that in 62% of the questions considered, generic links enabled either a shorter time or lower level of difficulty, when compared to accomplishing the same tasks involving point-to-point links on the Web.

The only question (question 13, in the appendix), that compared tasks involving local links to point-to-point links showed a statistically significant difference with advantage to local links. The median for local links (Median Local Microc.), median for point-to-point links on the Web (Median Point-to-point Web), and the corresponding level of significance are presented in Table 6:

Quest	Attribute	Median Local Microc.	Median Point-to-point Web	Level Sig.
13	Time Difficulty	1 1	2 2	0.00* 0.08**
*denotes that the result is statistically significant at 5% level **denotes that the result is statistically significant at 10% level				

Table 6: Medians for tasks involving local links and point-to-point links, with corresponding level of significance.

We can see that when the applications require the definition of links to be valid within the whole application or within a particular document, the use of point-to-point links on the Web increases either the time involved or the level of difficulty in accomplishing the task.

For the independent variables size of the application, compactness, stratum, and experience we found significant Z values for questions 9,14 and 15. The results are presented in Table 7:

Questions		Number Docum.	Compactness	Stratum	Experience
09	Time	1.85*			
	Diffic.	2.04*			
14	Diffic.		1.67*	2.06*	2.31*
15	Diffic.				1.98*
* A Z critical of 1.64, denoting that the result is statistically significant at 10% level					

Table 7: Significant association between independent and dependent variables

We found values of Gamma higher than 0.50 not only for the four independent variables presented in table 7, but also for the connectivity and the structure of the application. Values for Gamma equal or higher than 0.50 show that there exists an association between the variables compared.

In relation to the influence of the highlighting of anchors on maintainability/reusability the results are presented in Table 8:

	Median Microcosm	Median the Web
Highlighting for Maintenance	4	3
Highlighting for Reuse	4	2

Table 8: Influence of the highlighting of anchors on Maintainability/Reusability

The medians for Microcosm are higher than the medians for the Web, and this is probably caused by the fact that links on the Web are generally highlighted, which is not the case in Microcosm. We can see from the median that there is an influence of the highlighting on the maintainability/reusability.

The questionnaire did not consider reuse of links, since it does not make sense to reuse point-to-point links, and these are the only types of links available on the Web. But the reuse of links is an important issue in the reuse process as a whole, and we believe that one of the advantages of Microcosm is that it allows the reuse of local and generic links, as they are stored separately from the documents - in linkbases. Any linkbase can be ‘plugged into’ any application, giving a high level of flexibility to authors.

7 Conclusions and Future Work

We have presented our approach to the development of metrics within the SHAPE research project and how they were evaluated.

The metrics were proposed to measure the maintainability and reusability of hypermedia applications for education, so that we could evaluate whether a particular hypermedia application for education was more or less maintainable or reusable than another hypermedia application for education. Therefore, the metrics proposed are not restricted to a particular hypermedia system since they can be used to measure the maintainability and reusability of any hypermedia applications for education.

In order to evaluate the metrics proposed, we collected the data using applications developed in either Microcosm or the Web.

The data collected showed strong evidence that the link representation, link type, highlighting of anchors, structure of the application, and the author's experience can strongly influence the maintainability of the application and the reusability of information.

We also found some evidence that the size of the application, compactness, and stratum can also influence the maintainability of the application and the reusability of information.

Our next evaluation will be a quantitative/qualitative evaluation and will measure the development effort involved in developing a hypermedia application for education. The evaluation will consist in developing the same application using both Microcosm and the Web. The application will be designed using the principles from the Cognitive Flexibility Theory and the authors will be undergraduate students from the Human-Computer Interaction discipline.

References

[Adams and Jr 97] Adams, W. J., Jr, Curtis A. Carver: "The Effects of Structure on Hypertext Design"; Proceedings of ED-MEDIA'97, Calgary (1997).

[Andrews et al. 1995a] Andrews, Keith, Nedoumov, Andrew, and Scherbakov, Nick: "Embedding Courseware into the Internet: Problems and Solutions"; Proceedings of ED-MEDIA 95, Graz (1995), 69-74.

[Andrews et al. 95b] Andrews, Keith, Kappe, Frank, Maurer, Hermann, and Schmaranz, Klaus: "On Second Generation Network Hypermedia Systems"; Proceedings of ED-MEDIA 95, Graz (1995), 75-80.

[Balasubramanian et al. 94] Balasubramanian, P., Isakowitz, Tomás, and Stohr, Edward A.: "Designing Hypermedia Applications"; Proceedings of the Twenty-Seventh Annual Hawaii International Conference on System Sciences, Hawaii (1994), 354-365.

[Basili and Rombach 88] Basili, V. R. and Rombach, H. D.: "Towards a Comprehensive Framework for Reuse: A Reuse-Enabling Software Evolution Environment"; Technical Report CS-TR-2158, Dept. of Computer Science, University of Maryland, College Park, MD 20742 (December, 1988).

[Basili et al. 94] Basili, V., Caldiera G., and Rombach, D.: "The Goal Question Metric Approach", Encyclopedia of Software Engineering, Wiley (1994).

[Berners-Lee et al. 94] Berners-Lee, T., Cailliau, R., Luotonen, A., Nielsen, H. Frystyk, and Secret, A.: "The World Wide Web"; Communications of the ACM, 37, 8 (August, 1994), 76-82.

[Bernstein et al. 91] Bernstein, Mark, Brown, Peter J., Fisse, Mark, Glushko, Robert, Landow, George, Zellweger, Polle: "Structure, Navigation, and Hypertext: The Status of the Navigation Problem"; Proceedings of Hypertext'91, ACM Press, San Antonio (1991), 363-367.

[Botafogo et al. 92] Botafogo, Rodrigo A., Rivlin, Ehud, and Shneiderman, Ben: "Structural Analysis of Hypertexts: Identifying Hierarchies and Useful Metrics", ACM TOIS, 10, 2 (1992), 143-179.

[Briand et al. 96] Briand, L., Bunse, C, Daly, J, Differding, C.: "An experimental comparison of the maintainability of OO and structured design documents"; Proceedings of EASE (1996).

[Briand et al. 97] Briand, L., Devandu, P. and Melo, M.: "An Investigation into Coupling Measures for C++"; Proceedings of ICSE'97, Boston (1997), 412-421.

[Calvi and DeBra 97] Calvi, Licia and DeBra, Paul: "Using Dynamic Hypertext to Create Multi-Purpose Textbooks"; Proceedings of ED-MEDIA'97, Calgary (1997).

[Carvalho and Dias 97] Carvalho, Ana Amélia Amorim and Dias, Paulo: "Hypermedia Environment using a Case-Based Approach to Foster the Acquisition of Complex Knowledge"; Proceedings of ED-MEDIA'97, Calgary (1997).

[Catlin and Garrett 91] Catlin, Karen Smith, and Garrett, L. Nancy: "Hypermedia Templates: An Author's Tool"; Proceedings of Hypertext'91, ACM Press, San Antonio (1991), 147-160.

[Daly 96] Daly, J.: "Replication and a Multi-Method Approach to Empitical Software Engineering Research", PhD thesis, Department of Compyter Science, University of Strathclyde, Glasgow, (1996).

[Davis et al. 92] Davis, Hugh, Hall, Wendy, Heath, Ian, Hill, Gary, and Wilkings, Rob: "Towards an Integrated Information Environment With Open Hypermedia Systems"; Proceedings of the ACM Conference on Hypertext, ACM Press, Milan (1992), 181-190.

[Duval and Olivié 95] Duval, Erik, and Olivié, Henk: "A Home for Networked Hypermedia"; Proceedings of ED-MEDIA 95, Graz (1995), 193-198.

[Fenton and Pfleeger 96] Fenton, Norman E., and Pfleeger, Shari Lawrence: "Software Metrics, A Rigorous & Practical Approach", Second Edition, PWS Publishing Company and International Thomson Computer Press (1996).

[Frakes and Terry 95] Frakes, William and Terry, Carol: "Software Reuse and Reusability Metrics and Models"; Technical report TR-95-07, Virginia Polytechnic Inst. and State University (1995).

[Frakes and Terry 96] Frakes, William and Terry, Carol: "Software Reuse: Metrics and Models"; ACM Computing Surveys, 28, 2 (1996), 415-435.

[Garzotto et al. 91] Garzotto, Franca, Paolini, Paolo, and Schwabe, Daniel: "HDM - A Model for the Design of Hypertext Applications"; Proceedings of Hypertext'91, ACM Press, San Antonio (1991), 313-328.

[Garzotto et al. 94] Garzotto, Franca, Mainetti, Luca, and Paolini, Paolo: "Analysing the Quality of Hypermedia Applications: A Design-Oriented Framework", Workshop on hypermedia design and development, Edinburgh (1994).

[Garzotto et al. 95] Garzotto, Franca, Mainetti, Luca, and Paolini, Paolo: "Hypermedia Design, Analysis, and Evaluation Issues", Communications of the ACM, Special Issue on Hypermedia Design, August (1995).

[Garzotto et al. 96] Garzotto, Franca, Mainetti, Luca, and Paolini, Paolo: "Information Reuse in Hypermedia Applications"; Proceedings of the ACM Conference on Hypertext'96, ACM Press, Washington DC (1996), 93-101.

[Goldberg et al. 96] Goldberg, M. W., Salari, S., and Swoboda, P.: "World Wide Web - Course Tool: An environment for building WWW-based courses"; Proceedings of the Fifth International World Wide Web Conference, Paris (1996), 1219-1232.

[Harrison et al. 95] Harrison, R., Samaraweera, L. G., Dobie, M. R., and Lewis, P. H.: "Estimating the quality of functional programs: an empirical investigation", Inf. Softw. Technol., 37, 12 (1995), 701-707.

[Hatzimanikatis et al. 95] Hatzimanikatis, A. E., Tsalidis, C. T., and Christodoulakis, D.: "Measuring the Readability and Maintainability of Hyperdocuments"; J. of Software Maintenance, Research and Practice, 7 (1995), 77-90.

[Healey 93] Healey, J. F.: "Statistics, a tool for social research"; Wadsworth Publ. (1993).

[Hill et al. 95] Hill, Gary, Hall, Wendy, De Roure, D., and Carr, L.: "Applying Open Hypertext Principles to the WWW", Proceedings of the International Workshop on Hypermedia Design '95, Montpelier (1995).

[ISO/IEC 91] ISO/IEC: " International Standard: Information Technology - Software product evaluation - Quality characteristics and guidelines for their use"; ISO/IEC 9126 (1991).

[Jacobson 97] Jacobson, Michael J.: "The Evolution Thematic Investigator: Research and the Design of Hypermedia Learning Environments", Proceedings of ED-MEDIA'97, Calgary (1997).

[Jacobson and Spiro 95a] Jacobson, Michael J., and Rand J. Spiro: "Hypertext Learning Environments, Cognitive Flexibility, and the Transfer of Complex Knowledge: An Empirical Investigation", J. Educational Computing Research, 12,4 (1995), 301-333.

[Jacobson and Spiro 95b] Jacobson, M. J., and Rand J. Spiro: "Hypertext learning environments and epistemic beliefs: A preliminary investigation"; Technology-based learning environments: Psychological and educational foundations, S. Vosniadou, E. DeCorte, & Mandl (Eds.), Springer-Verlag, (1995), 290-295.

[Jacobson and Archodidou 97] Jacobson, M. J., and Archodidou, A.: "Case-based hypermedia and learning neo-Darwinian evoluationary biology: Promoting conceptual change of complex

scientific knowledge", Manuscript submitted for publication (1997), http://lpsl.coe.uga.edu/Jacobson/papers/JacobsonEvo97.pdf.

[Jacobson et al. 96] Jacobson, M. J., Mishra, Maouri, C. P., and Kolar, C.: "Learning with hypertext learning environments:Theory, design, and research"; Journal of Educational Multimedia and Hypermedia, 5, 3/4, (1996), 239-281.

[Jordan et al. 89] Jordan, Daniel S., Russell, Daniel M., Jensen, Anne-Marie S., and Rogers, Russell A.: "Facilitating the Development of representations in Hypertext with IDE"; Proceedings of Hypertext'89, ACM Press, Pittsburgh (1989), 93-104.

[Karunanithi and Bieman 93] Karunanithi, Santhi and Bieman, James M.: "Measuring Software Reuse in Object Oriented Systems and Ada Software"; Technical Report CS-93-125, Colorado State University (1993).

[Kitchenham 96] Kitchenham, Barbara Ann: "Evaluating Software Engineering Methods and Tool, Part 1: The Evaluation Context and Evaluation Methods"; Software Engineering Notes, 21, 1 (Jan. 1996), 11-15.

[Kitchenham 93] Kitchenham, Barbara: "DESMET METHODOLOGY: Guidelines for Evaluation Method Selection", DESMET Project Deliverable D2.3.1, The National Computing Centre Ltd. (1993).

[Landow 87] Landow, George P.: "Relationally Encoded Links and the Rhetoric of Hypertext"; Proceedings of Hypertext'87, ACM Press, (1987), 331-343.

[Marmann et al. 92] Marmann, Michael, and Schlageter, Gunter: "Towards a Better Support for the Hypermedia Structuring: The HYDESIGN Model"; Proceedings of the ACM European Conference on Hypertext, ACM Press, Milano (1992), 11-22.

[Marshall et al. 91] Marshall, Catherine C., Halasz, Frank G., Rogers, Russell A., and Jr., William C. Janssen: "Aquanet: a hypertext tool to hlod your knowledge in place"; Proceedings of Hypertext'91, ACM Press, San Antonio (1991), 261-275.

[Marshall et al. 95] Marshall, Catherine C., and Shipman III, Frank M.: "Spatial Hypertext: designing for Change"; Communications of the ACM, Special Issue on Hypermedia Design, August (1995).

[MacDonell 91] MacDonell, S. G.: "Rigor in Sofware Complexity Measurement Experimentation"; J. Systems Software, 16 (1991), 141-149.

[Mendes 97] Mendes, M. E. X.: "SHAPE - Southampton Hypermedia Authoring Paradigm for Education", transfer Thesis from MPhil to Ph.D., Department of Electronics and Computer Science, University of Southampton (1997).

[Mendes and Hall 97a] Mendes, M. Emilia X. and Hall, Wendy: "An empirical study of hypermedia authoring for education"; Proceedings of the CAL97 Conference, Exeter (1997).

[Mendes and Hall 97b] Mendes, M. Emilia X. and Hall, Wendy: "The SHAPE of Hypermedia Authoring for Education"; Proceedings of ED-MEDIA & ED-TELECOM 97, Calgary (1997).

[Meyrowitz 86] Meyrowitz, N.: "Intermedia: The Architecture and Construction of an Object-oriented Hypermedia System and Applications framework"; Proceedings of the OOPSLA'86, (1986), 186-201.

[Moreira 96] Moreira, A. A. de F. G.: "Desenvolvimento da flexibilidade cognitiva dos alunos-futuros-professores: uma experiencia em Didatica do Ingles"; PhD thesis (1996), University of Aveiro, Portugal.

[Nanard and Nanard 95] Nanard, Jocelyne, and Nanard, Marc: "Hypertext Design Environments and the Hypertext Design Process"; Communications of the ACM, Special Issue on Hypermedia Design, August (1995), 49-56.

[Prieto-Díaz 93] Prieto-Díaz, R.: "Status Report: Software Reusability"; IEEE Software, 10, 3 (1993), 61-66.

[Rana and Bieber 97] Rana, Ajaz R., Bieber, Michael: "Towards a Collaborative Hypermedia Educational Framework"; Proceedings of Thirtieth Annual Hawaii International Conference on System Science, Maui (1997), 610-619.

[Rivlin et al. 94] Rivlin, Ehud, Botafogo, Rodrigo, and Schneiderman, Ben: "Navigating in Hyperspace: designing a structure-based toolbox"; Communications of the ACM, 37, 2 (1994), 87-96.

[Rossi et al. 95] Rossi, G., Schwabe, D., C. Lucena, J. P., and Cowan, D. D.: "An Object-Oriented Model for Designing Human-Computer Interface of Hypermedia Applications"; Proceedings of the IWHD'95, Montpellier (1995), 131-152.

[Schuler and Thüring 94] Schuler, Wolfgang, and Thüring, Manfred: "Pragmatical Hypertext Design (PHD)", Technical Report, GMD Institute, Germany (1994), 3-58.

[Spiro et al. 95] Spiro, Rand J., Feltovich, Paul J., Jacobson, Michael J., and Coulson, Richard L.: "Cognitive Flexibility, Constructivism, and Hypertext: Random Access Instruction for Advanced Knowledge Acquisition in Ill-Structured Domains"; Constructivism, L. Steffe & J. Gale (Eds.), Hillsdale, N.J.:Erlbaum (1995).

[Thimbleby 96] Thimbleby, Harold: "Systematic web authoring", The British HCI Symposium The Missing Link: Hypermedia Usability Research & The Web, UK (1996), http://www.cs.mdx.ac.uk/harold/webpaper/.

[Thüring 94] Thüring, Manfred A: "Conceptual Framework for Hypermedia Design Methodologies"; Workshop on Hypermedia Design and development, Edinburgh (1994).

[Türing et al. 95] Türing, Manfred, Hannemann, Jörg, and Haake, Jörg: "Hypermedia and Cognition: Designing for Compreension", Communications of the ACM, Special Issue on Hypermedia Design, August (1995).

[Vocht 94] Vocht, J. W.: "Experiments for the Characterization of Hypertext-structures"; Master's Thesis, Eindhoven University Technology, Department of Mathematics and Computing Science (1994).

[Yamada et al. 95] Yamada, Shoji, Hong, Jung-Kook, and Sugita, Shigeharu: "Development and Evaluation of Hypermedia for Museum Education: Validation of Metrics"; ACM Transactions on Computer-Human Interaction, 2, 4 (December, 1995), 284-307.

Appendix

Questions asked in the questionnaire

1) Finding dangling links within a document that has 5 links to other documents

2) Finding whether a document is part of an island, if we do not consider generic links.

3) Deleting a document, that has 5 links to other documents, without leaving dangling links

4) Adding a new paragraph to the beginning of a text document, that has 5 links to other documents, keeping the links intact.

5) Modifying the source anchor of a link

6) Modifying the destination of a link

7) Deleting a link

8) Moving 5 documents, each with 5 links to other documents, from one directory to another, keeping their links valid.

9) Moving 5 documents, each with 5 links to other documents, from machine A to machine B, keeping their links valid, where both machines have the same operating system.

Assume for the tasks 10 and 11 that you have just deleted a document that has 2 links and 3 different links from other documents.

10) Checking for dangling links caused by the deletion of that document

11) Checking for islands caused by the deletion of those 3 links

12) Suppose that you want to link 10 terms used in your application to descriptions defined in a glossary. The glossary already exists and is comprised of one text document in its docuverse. This document contains the definition of those 10 terms and some others as well.

13) Suppose that 5 text documents in your application use the word ‘rose’ within their contents. You want to link this word, wherever it occurs within those 5 documents, to a destination document that shows the picture of a rose.

14) Suppose that you have a set of 5 documents in your application, each with 2 links to other documents, that you want to copy (duplicate) to another application, keeping all the links already defined. Both applications share the same docuverse

15) Suppose that you have a document in your application, with 2 links to other documents, that you want to duplicate within your application (for future enhancement), keeping all the links already defined.