

Reader-Friendly Media for the Documentation of Software Projects

Andreas Rüping

sd&m software design & management GmbH & Co.KG

Thomas-Dehler-Straße 27
D-81737 München, Germany

e-mail: rueping@acm.org

Introduction

When documenting software projects, authors regularly face the question of which media they should choose for their documents. Basically, authors have two different options:

- Print is the traditional medium; it's been available for centuries. Until the present day, print still offers numerous ergonomical advantages.
- On the other hand, online documents are becoming more and more popular. They offer navigation and search capabilities that can never be achieved for printed documents.

This decision can be quite intricate since documents produced for one medium cannot automatically be used for the other as well. On the one hand, an online document cannot be used as a print document since it doesn't print very well. On the other hand, producing online documents requires more effort than just putting a print document online; authors have to add hyperlinks and to adjust the document to screen layout.

However, the real question is not print vs. online, but how these two options can be combined. Authors can of course choose print for some documents and online for others. But they can also choose to make documents available for both media.

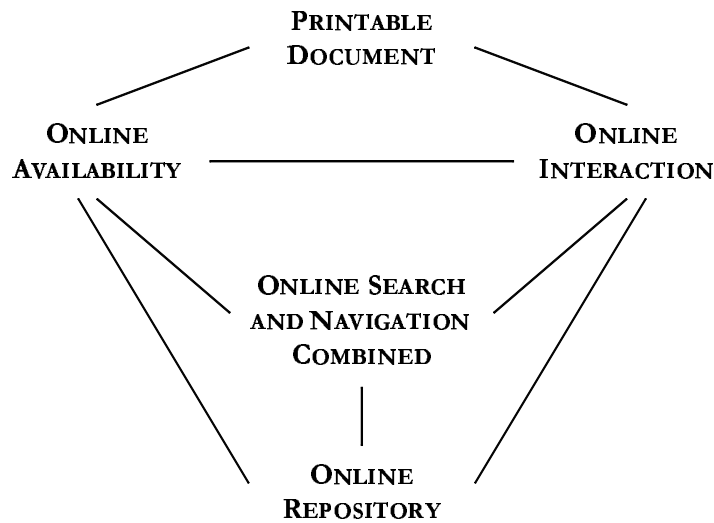
The choice of a medium should reflect each document's typical usage. It's the combination of media that makes documentation reader-friendly. The patterns in this paper explain the advantages of each medium, and give some guidelines on when to use which and on how to combine them.

Guidelines for the Readers

This paper is part of a series of pattern papers dealing with the documentation of software projects. Previous papers have presented collections of patterns dealing with the structure and layout of project documents [11], typography [14], project documentation management [13], and the process of writing and reviewing documents [12].

The pattern form is the same that was used in the previous papers. Each pattern includes a problem statement, a discussion of forces, a solution, and a discussion of related patterns and rationale. The problem section and the first paragraph of the solution section of each pattern form so-called pattlets — thumbnails (printed in boldface) that give a quick overview.

The following diagram gives an overview of the patterns in this paper.



1 Online Availability

Problem How can documents be made available to the readers in a structured way?

Forces Project documentation typically consists of a large number of documents. While search capabilities can help readers look up information, readers still need to get an overview of which documents exist and which documents might be important for them. Once readers have found the documents they need, they should be able to retrieve the documents easily.

Project documentation must often be available on different platforms. A platform-independent, portable availability is desirable.

Information can expire. Project documents are therefore often updated. It is important that the most recent versions are made available. It must be possible to add new documents and to include new versions of existing documents quickly, preferably in one central place.

Solution **An overview of the project documentation should be given online. Readers should be able to retrieve individual documents using this overview, whether an individual document is meant for print or online use.**

The overview should be structured in the following way:

- The online overview should consist of one or more web pages that mirror the project organisation. The overview should explain which documents are available in the project's different areas.
- The overview should make clear for each document whether it is available for print, for online use, or for both.
- Hyperlinks should be offered that act as a pathway to the project documents. All print documents should be reachable by hyperlinks, while for online documents a hyperlink to an entry point is sufficient.
- The overview should be kept up to date; new documents should be added and new versions should replace old versions.

Because web pages are used to present the project overview, the project documentation is available in a platform-independent way.

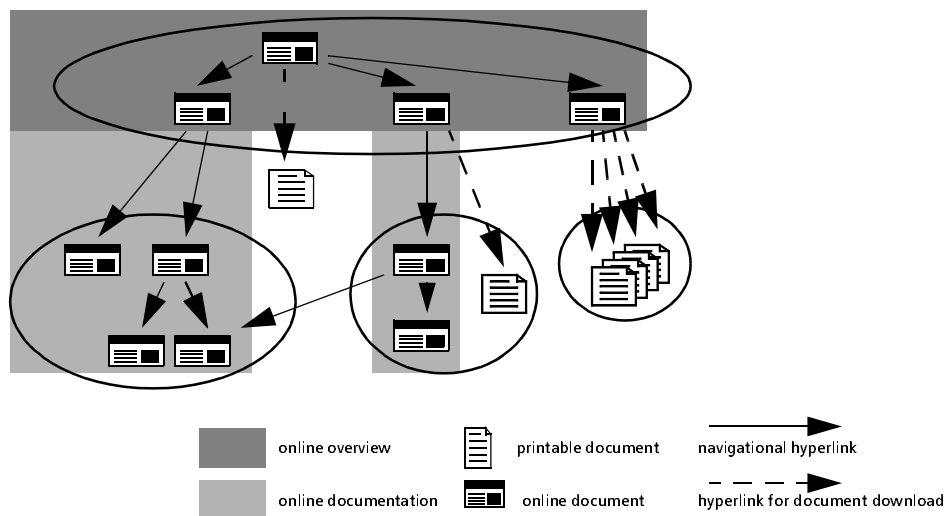


Figure 1 A document landscape implemented with navigational links

Discussion

This pattern demonstrates how A DOCUMENT LANDSCAPE [13] can be implemented using online navigation. Travelling through the project overview, readers are referred to the documents they may need. The document landscape in Figure 1 shows four overview web pages, from which an introductory document and three areas of documentation can be reached. One area consists of online documents only, one consists of a printable document that is also available as two online documents, and one consists of a set of printable documents.

If there is no online version available for a document, providing THUMBNAIL SKETCHES [11] in the online overview can be helpful to give readers an idea what the PRINTABLE DOCUMENT is about.

2 Printable Document

Problem

How can authors ensure comfortable reading?

Forces

Sometimes readers read a document, while sometimes they just browse through, look up a specific detail, or find a reference to another document.

The time readers spend on a document influences their preference for a medium. Most people prefer print when otherwise they would have to use a computer screen for reading for, say, more than 10 minutes. There are a number of reasons for this. Paper has better resolution and contrast than a computer screen does [6]. Paper doesn't need electrical power supply [6]. Paper allows readers to sit down, lay back, and concentrate.

In addition, readers can take printed documents with them without having to carry a computer [6]. Printed documents can easily be marked for making editorial comments. However, readers have to use a document online when they wish to use hyperlinks that the document may offer [6].

The choice of a medium is also influenced by the fact that sometimes documents can be generated, for instance from comments in source code. Generated documents often consist of textual building blocks that are connected with hyperlinks, and are normally intended for online use.

However, documents that consist of a steady flow of natural language — the ones that readers typically prefer to read as print — can hardly be generated, but are usually authored by humans.

Solution

Paper is for reading; online is for looking up information. A printable version should be offered for each document, unless the document can automatically be generated as an online document.

A version can be considered printable if it is formatted according to the layout of the printed page.

- The printable version should be made available in a print format that is widely available, such as Portable Document Format (PDF) or PostScript.
- Obviously all documents can be printed somehow, even documents that are originally meant for online use. Still, if they aren't properly formatted, documents don't print very well. For instance, page breaks are always awkward in printed HTML documents.

Examples of documents that require printing include specifications, design descriptions, research reports, management summaries, even budget plans. These documents cannot be generated. On the other hand, they will be studied in detail — reading them online is too much to ask.

The need for a printable version doesn't conflict with a possible online version. Sometimes both a printable and an online version of one document are useful, for instance if including hyperlinks turns out helpful.

Discussion

When preparing printable documents, authors should follow a few typographic guidelines: There should be TEXT ON 50% OF A PAGE [14] while the rest of the page should remain blank. The line width should allow for TWO ALPHABETS PER LINE [14]. Authors should use 120% LINE SPACING [14], normally no more than TWO TYPEFACES [14], and they should make CAREFUL USE OF TYPE VARIATIONS [14].

The requirement for a print format is motivated by the desire for a SEPARATION OF PROCESSING AND PRINTING [13], which allows authors to produce printable documents with well-defined layouts that readers cannot change, and documents that are portable in the sense that they can be printed without the need for a particular desktop publishing software.

If a document is necessary both online and on paper, it is sometimes tempting to set up the document for online use and also use this version for printing. Besides the reduced effort, this has the advantage that information isn't kept redundantly in two documents. Nonetheless, printing documents formatted for online use isn't a good idea. To get rid of the redundant information authors can consider to use TEXT GENERATORS [13], for instance for generating an HTML version from a printable version.

There is an ongoing discussion whether printed documents and books will eventually be replaced by electronic documents and books. There are technological advances as far as the ergonomics of electronic documents are concerned [10]. For the time being, however, most people prefer printed documents for reading.

3 Online Interaction

Problem

How can authors present material that involves their readers?

Forces

Project documents are sometimes characterized by a high degree of referentiality. There are cross-references within documents, as well as references to other documents. Readers often want to follow these references. However, following textual references in a printed document is tedious, in particular when there are many of them.

To the contrary, hyperlink structures feature a large expressive power as far as referentiality and navigation are concerned [5]. In addition, arranging information in a non-linear, more-dimensional fashion is considered to be quite natural [8].

In addition, online presentation allow readers to play a more active role. Readers become users who can take influence on which information they get, and on how it is presented.

However, if a network of hyperlinks becomes too complicated, readers can get lost. Experiences with hypertexts suggest that hyperlink networks are relatively easy to follow if they span a tree [2]. A tree represents a balance between structuring and comprehensibility. Most users prefer a broad tree over a narrow one; about 7 subnodes per node is fine while the tree depth should be kept as small as possible [7][16]. In most cases, a tree depth of 3 is sufficient [5]. Enriching trees by multiple entry points appears to be a possible extension [4].

Solution

Material should be put online if its presentation can profit from user interaction, most notably for navigation and for the selection of views.

User interaction can take on different forms, including these examples:

- Material with a high degree of referentiality can profit from the inclusion of hyperlinks for navigational purposes. Users are given the freedom to determine the order in which they access the information. For instance, an online tutorial allows users to find their own way through the material provided, and gives them the freedom to follow references to related issues as they see fit. The online tutorial could even include a simulation of the actual software which would require user interaction to demonstrate how the software works.
- Users can define views, that is, they can extract information and calculate the results they need. This is more common for numerical than for textual information. For instance, many projects use spreadsheets for their budget plans. With the spreadsheet available online, users can interactively calculate a more optimistic or a more conservative budget plan.

Online documents typically consist of a number of web pages connected by hyperlinks. Ideally these pages should be organised as a tree with a depth of 3 or smaller.

It is recommended to use hyperlinks for significant relationships between documents only. A relationship can be considered significant when it is likely that readers will stop reading in one place and continue reading somewhere else [5].

The drawback of online documents is that they can be costly to produce. Providing accurate hyperlinks always represents a fair amount of work which must be weighed against the expected advantages.

Discussion

In most cases an online document alone is not sufficient, but should be made available in addition to a PRINTABLE DOCUMENT, with the exception of online documents that can be generated from other sources such as source code. Generated documents are produced with TEXT GENERATORS [13] that extract information from their sources, and relate different pieces of information with hyperlinks. Generated documents are typical online documents that feature a high degree of referentiality.

Online documents should be integrated with other online documents. This way A DOCUMENT LANDSCAPE [13] with ONLINE AVAILABILITY evolves. It can be helpful to distinguish true navigational links, which readers need to follow for an overview of the entire documentation, from mere cross-references. Figure 2 gives an example.

When writing online documentation, it's important to follow general guidelines for organising web sites. Robert Orenstein has published a pattern language for an essay-based web site [9]. Among other things he recommends LOW-DEPTH DOCUMENT TREES [9] and an INTRODUCTORY SECTION [9] with an INTRODUCTORY PICTURE [9].

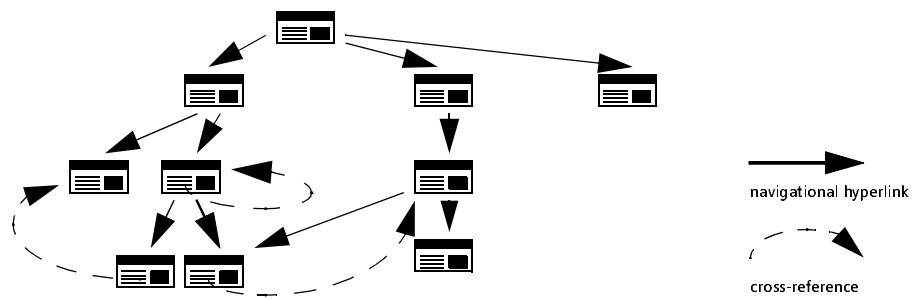


Figure 2 Online documents with navigational links and cross-references

It may sound interesting to make quickly changing documents online documents, in the hope that readers don't keep outdated printed versions, but read the documents online; they would then use the most recent versions. However, readers would often print the online documents. Frequent changes alone don't justify putting documents online. It is more promising to extract quickly changing information into separate documents, and to give users the opportunity to actively request the most recent versions of these documents online.

4 Online Search and Navigation Combined

Problem How can readers find some particular information?

Forces

It's common that readers look for a certain piece of information within the project documentation. They may, however, not know exactly where the information can be found. An overview document or an overview diagram will not help them a lot. In this case full text search or key word search can be helpful.

With all search systems, however, there is a trade-off between recall and precision. Recall describes how much of the relevant information is found, while precision describes how much of the information found is relevant. Recall and precision typically add up to 100 percent [15]. For instance, a study on full-text search showed that 80 percent of the information retrieved was useful, but only 20 percent of the desired information was found [1]. Search systems can be configured to find more relevant information, but then users are confronted with more irrelevant information among the search results. 50 percent recall and 50 percent precision can be considered a typical result [3].

However, even if a search doesn't immediately retrieve the desired information, the search results can be used as entry points into the documentation; users can retrieve more relevant information by navigation starting from these entry points.

The larger a project is, the more important is the need for automatic search support.

Solution **Online search should be combined with online navigation.**

An online search capability should be embedded into the online documentation:

- Ideally, the search capability should be a search engine as it is known from the web. Users should be able to navigate to the search engine within the online project documentation.

- The result of a search query should be an HTML document with hyperlinks to the documents found generated into it. If the documents found are online documents, readers can travel along the hyperlinks to look for more relevant information.
- In case the documents found aren't online documents, the search result should provide a link to where they are available for download.

In small projects, setting up a search engine can be too expensive, though. In this case the search capabilities offered by operating systems and desktop publishing systems can be considered a legitimate fall-back strategy.

Discussion

The result of a search query should be an online document; users can navigate to the documents they are looking for by following the hyperlinks (see ONLINE INTERACTION). A general ONLINE AVAILABILITY is necessary to make sure that all documents found can be retrieved.

An ONLINE REPOSITORY is a tool that stores information about various project artefacts and classifies both the artefacts and the relations between them. It is a way to implement the combination of online navigation and online search capabilities.

5 Online Repository

Problem

How can the relationships between many project documents be managed?

Forces

In a large project that produces many documents the number of references between documents can become huge. The number of references increases by the square of the number of documents.

Information isn't present in textual documents only. Often there are connections between textual documents and code. Such connections can be interesting to the readers. For instance, someone who reads the documentation of a particular class may want to travel through the class hierarchy to see what the superclasses and subclasses are.

While online techniques such as navigation and search capabilities are available for online documents, they are not always readily available for code or for documents that are not presented online.

Solution

In a large project with a large number of documents, an online repository can be introduced that integrates navigation and search capabilities for all kinds of documents.

An online repository is a tool that uses a database to store documents and to manage the relationships between them. It allows users to set up a meta model for all kinds of documents and the relationships between them. It offers the following features:

- Users can define types for all kinds of documents. This includes traditional documents such as analysis documents and design documents, but also other artefacts such as Smalltalk classes, Java classes, etc.
- Types also exist for parts of documents. For instance, a document can be defined as consisting of an abstract, a list of keywords, and the main text.
- Users can define types for all kinds of relationships between documents, such as *is subclass of* and *is described in* etc.

- Users can formulate queries to search for information of a certain kind along particular relationship paths.
- The results of such queries are documents or partial documents. The online repository presents these results as hypertext so that they can be used as the starting point for further queries.

An online repository is much more powerful than a simple search engine since it allows users to define the types of information they want to search for, in other words, users can define their own meta model for their documentation. Tagging pieces of information with certain meta information is what makes an online repository particularly effective.

Only a large project can justify such a powerful tool. However, an online repository can be useful to store and relate the documents of many projects.

Discussion

An online repository requires that the underlying information be available in a well-structured format which allows to tag the parts of a document with certain meta information, such as XML. Unless the documents are in such a format already, a tool is required that can extract the necessary information from the original document. An XML export is an example.

Although an online repository is more powerful than a search engine, it implements the idea of having ONLINE SEARCH AND NAVIGATION COMBINED. It therefore integrates the two most important mechanisms that the online medium offers.

Conclusions

Print and online media sometimes seem like different worlds, one traditional, the other modern and fashionable. Yet they aren't opposed to each other, but complement each other well. Each medium has its pros and cons; and it's fine to take the best of both worlds.

Projects can combine both print and online media. These patterns show when each should be used, and how they can be combined. This way, software projects can increase the reader-friendliness of their documentation, probably much to their readers' ease.

Acknowledgements

The patterns in this paper reflect a lot on how print and online documents are used at sd&m. I'd like to thank all colleagues at sd&m who have shared their insights into print and online documents with me.

Thanks are due to Mike Wu for his suggestions for improvement as the PLoP shepherd for this paper.

Last but not least, thanks are due to the participants of the PLoP 2000 workshop in which this paper was discussed. The feedback from the workshop helped improve this collection of patterns a lot.

References

- [1] David C. Blair, M. E. Maron. “An evaluation of retrieval effectiveness for a full-text document-retrieval system”, in: *Communications of the ACM*, Vol. 28, No. 3, March 1985.
- [2] Rodrigo A. Botafogo, Ehud Rivlin, Ben Shneiderman. “Structural analysis of hypertexts: identifying hierarchies and useful metrics”, in: *ACM Transactions on Information Systems*, Vol. 10, No. 2, April 1992.
- [3] Susan T. Dumais. “Textual Information Retrieval”, in: *Handbook of Human-Computer Interaction*, North-Holland, 1988.
- [4] George W. Furnas, Jeff Zacks. “Multitrees: enriching and reusing hierarchical structure”, in: *CHI '94 — Conference proceedings on Human factors in computing systems*, ACM, 1994.
- [5] William Horton. *Designing and Writing Online Documentation*. John Wiley & Sons, Second Edition, 1994.
- [6] Richard C. Hsu, William E. Mitchell. “After 400 years, print is still superior”, in: *Communications of the ACM*, Vol. 40, No. 10, October 1997.
- [7] J. L. Kiger. “The Depth/Breadth Tradeoff in the Design of Menu-Driven User Interfaces”, in: *International Journal of Man-Machine Studies*, Vol. 20, 1984.
- [8] Jakob Nielsen. *Hypertext and Hypermedia*. Academic Press, 1990.
- [9] Robert Orenstein. “A pattern language for an essay-based web site”, in: John M. Vlissides, James O. Coplien, Norman L. Kerth (Eds.), *Pattern Languages of Program Design*, Addison-Wesley, 1996.
- [10] Larry Press. “From P-books to E-books”, in: *Communications of the ACM*, Vol. 43, No. 5, May 2000.
- [11] Andreas Rüping. “The structure and layout of technical documents”, in: Jens Coldewey, Paul Dyson (Eds.), *Proceedings of the Third European Conference on Pattern Languages of Programming and Computing 1998*, Universitätsverlag Konstanz, 1999.
- [12] Andreas Rüping. “Writing and reviewing technical documents”, in: Jens Coldewey, Paul Dyson (Eds.), *Proceedings of the Third European Conference on Pattern Languages of Programming and Computing 1998*, Universitätsverlag Konstanz, 1999.
- [13] Andreas Rüping. “Project documentation management”, in: Paul Dyson, Martine Devos (Eds.), *Proceedings of the Fourth European Conference on Pattern Languages of Programming and Computing 1999*, Universitätsverlag Konstanz, 2000 (to appear).
- [14] Andreas Rüping. “Typography and desktop publishing”, in: Paul Dyson, Martine Devos (Eds.), *Proceedings of the Fourth European Conference on Pattern Languages of Programming and Computing 1999*, Universitätsverlag Konstanz, 2000 (to appear).
- [15] Gerard Salton. *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*. Addison-Wesley, 1989.
- [16] Thomas S. Tullis. “Designing a Menu-Based Interface to an Operating System”, in: *CHI '85 Proceedings*, ACM, 1985.