

Klinke, Sigbert; Witzel, Rodrigo

Working Paper

MD*Book online: A tool for creating interactive documents

SFB 373 Discussion Paper, No. 2002,21

Provided in Cooperation with:

Collaborative Research Center 373: Quantification and Simulation of Economic Processes, Humboldt University Berlin

Suggested Citation: Klinke, Sigbert; Witzel, Rodrigo (2002) : MD*Book online: A tool for creating interactive documents, SFB 373 Discussion Paper, No. 2002,21, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, Berlin, <https://nbn-resolving.de/urn:nbn:de:kobv:11-10048795>

This Version is available at:

<https://hdl.handle.net/10419/65356>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

MD*Book online - a tool for creating interactive documents¹

Sigbert Klinke and Rodrigo Witzel

Humboldt-Universität zu Berlin, School of Business and Economics, Institute of Statistics and Econometrics, Spandauer Str. 1, 10178 Berlin, Germany

Summary

Since the age of Gutenberg, publishing has meant to print a book readable by a person. Of course, interaction and dynamic illustration were not possible with this media. Even during the electronic age (till the mid of the nineties), publishing had not changed much. Although media, like diskettes and CDs, were often added to books, they often contained only additional material, e.g., lengthy computer programs.

With the breakthrough of the internet and the World Wide Web, the publishing of books has changed dramatically. Authors provide, beside the printed text, an electronic version of the book on the internet and on a CD. Interactive examples are sometimes embedded in books. Authoring systems appeared on the market, but their use usually they requires special knowledge of the programs. Moreover, the interactivity and the dynamics often depends on the authoring system and not on the author of a book.

We first define what we mean by an *e-book* and explain our choice of the document base. Since the MD*Book project bundles and extends techniques developed over the last several years, we consider the main roots of it: the XploRe helpsystem (APSS), the “Golden Solution”, and the MM*Stat project. Finally we describe MD*Book and the MD*Book online project. Please find at the end the internet addresses.

¹Financial support was received by the Deutsche Forschungsgemeinschaft, SFB 373 (“Quantifikation und Simulation Ökonomischer Prozesse”), Humboldt-Universität zu Berlin.

Keywords: internet, electronic publishing, electronic books

1 General considerations

For teaching statistics in basic courses as well as in advanced courses we use different types of documents, e.g.,

textbooks which are books, like the the XploRe learning guide (Härdle, Klinke, Müller, 2000a);

short books which are between books and slides, which are approximately represented in \LaTeX by the documentclass `report`;

slides which are used in the lectures itself.

In some sense, this classification represents the stages by which a textbook evolves. We usually start with some slides for the lectures which only contain key notes. Then we enhance these slides with examples, etc.. Thus we create a short book. Finally the text is enhanced with examples, exercises, extended informations, etc., and is structured in chapters, sections etc., and formatted to form the textbook.

Already during the development of XploRe, we recognized that different types of users (students, data analysts, researchers) require different types of interaction and access to statistical software. This has resulted in the development of menu-oriented software or teachware in XploRe 3 (see Bianchi 1995, Proenca 1995, Klinke 1997). The same holds for XploRe 4, e.g., see Klinke et al. (1998), Zanter (1998) and various authors in Härdle et al. (2000a, 2000b).

In the current state of development of MD*Book, it leads to three different types of interaction for the user:

Execute : Reproduce a graphic or output in the document;

Interactive : Reproduce a graphic or output and modify interactively the parameters, data sets , etc.;

Edit/Computing : Access the source code of the underlying program and modify it.

By an e-book we mean a document (slides, short book, or textbook) which is available on the intra/internet as well as on a CD. It should offer at least two of the three mentioned interaction types depending on the aim of the book. Additionally, we would like to generate all necessary output formats (PostScript, PDF, HTML) from one source, including the graphics.

When we first started doing e-books, they were written in HTML, but we soon needed another basis for the documents, for example, to integrate formulas. From our background, L^AT_EX was the natural choice, but immediately the question arises “Why not using Microsoft Word as a document basis?”. There are serious advantages for Word, but also, serious disadvantages:

+ commonly used

Teaching materials for basic studies and, to some extent, for the advanced studies in statistics are available in Word. But most of our teaching material for advanced studies is available in L^AT_EX.

+ graphic formats

Word supports the integration of a lot of different graphic formats. This is helpful for the Word user, but it requires a lot of translation work to generate graphics automatically in PostScript, PDF, or HTML.

- internal format

In the last 12 years we have seen at least *six* different Word versions (WinWord 2.0, WinWord 6.0, Word 95, Word 97, Word 2000, Word 2002 on the Windows platform alone). The internal format of Word has been modified, too. One could argue that Microsoft has now changed to a XML-based format, but somehow we doubt that Microsoft will use a standard format, e.g., like DocBook. Instead, they will use their own proprietary document format.

In contrast to that, the L^AT_EX format is stable. Moreover, entering formulas in Word is tedious and formulas are difficult to standardize across a document.

- translators

Although translators from Word to PostScript, PDF, and HTML have improved, they are still not good enough. For L^AT_EX we have good translation programs, e.g., `dvips`, `pdflatex`, `latex2html` and `ghostscript`.

Thus we decided to use L^AT_EX as a document basis.

2 The XploRe help system

Already, the XploRe 3 (although completely based on MS-DOS) helpsystem had hypertext capabilities. For example, by pressing <F10> XploRe 3 tried to open a help page, according to the word which was under the cursor. For XploRe 4, the helpsystem was based on HTML.

The cornerstone of the XploRe help system are: the “Function Groups”, the “Alphabetical Index”, the “Keywords”, and the “Tutorials”, e.g., see the

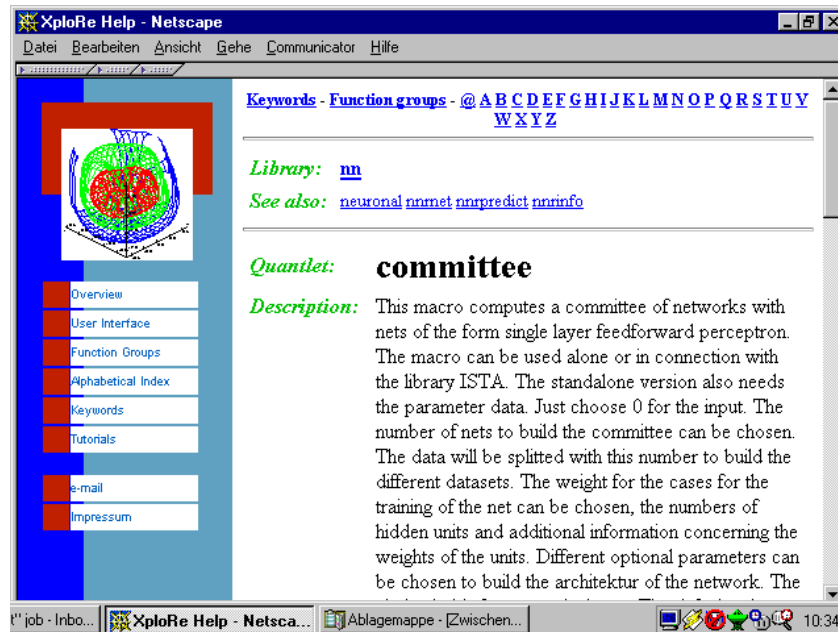


Figure 1: committee help file in the XploRe helpsystem.

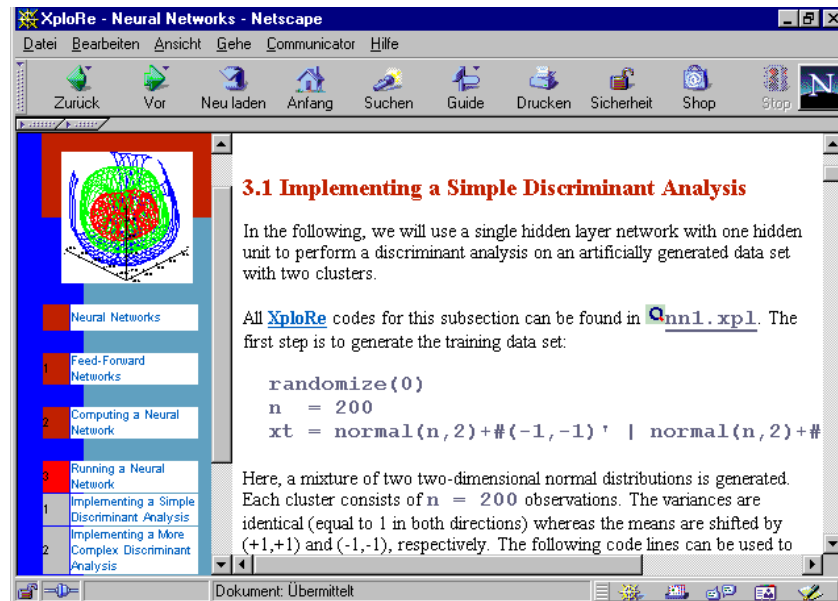


Figure 2: Neural network tutorial in the XploRe helpsystem.

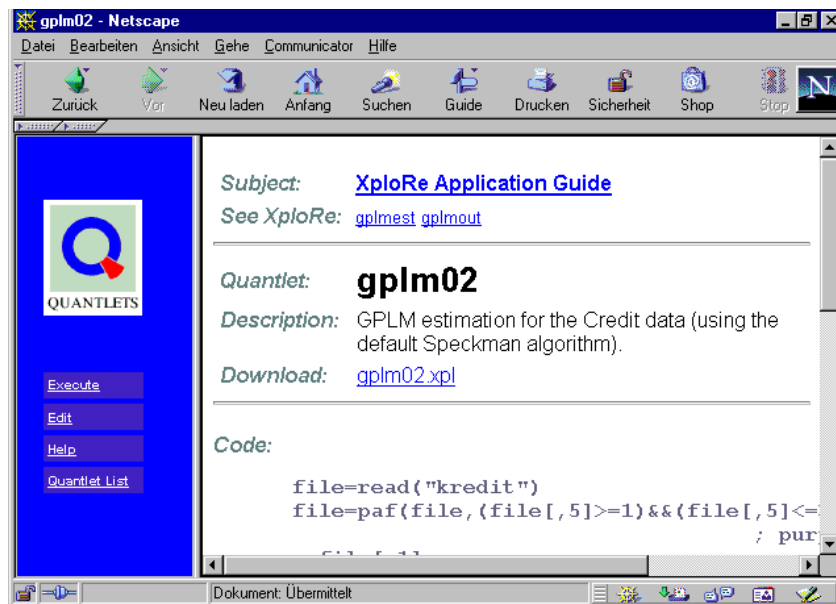


Figure 3: The “Golden Solution” for the quantlet `gplm02` which is part of the XploRe Application Guide (Härdle et al. 2000b). In the left frame we see the interaction types *Execute* and *Edit*.

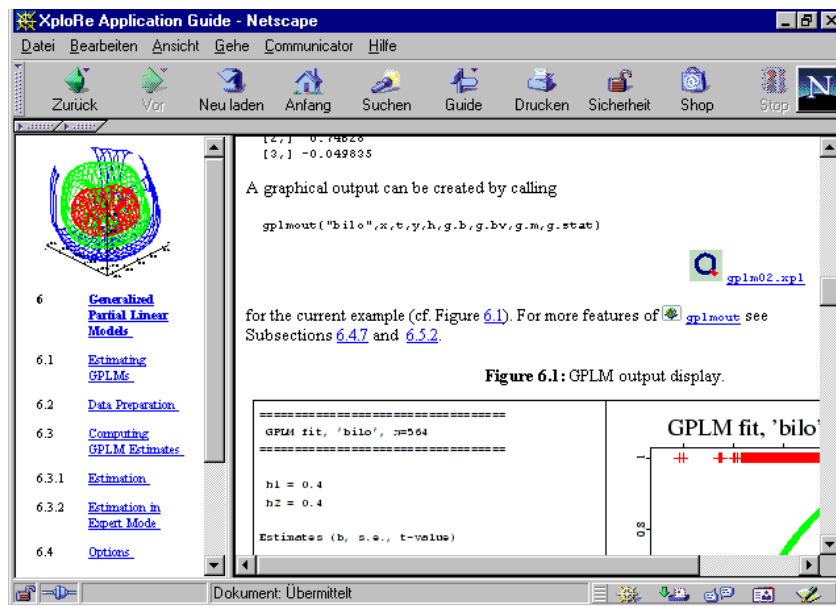


Figure 4: The link to the “Golden Solution” of `gplm02` in the XploRe Application Guide (Härdle et al. 2000b).

left frame in Figure 1. As you see in Figure 1, the text under the section *Description* is very long. To overcome this problem we introduced a section named *Notes*.

However, the problem of treating a topic which affects more than a quantlet remains (e.g., in Figure 1 `committee` is a quantlet based on neural networks). That leads us to introduce “Tutorials” for specific topics, like Neural Networks, Generalized Linear Models, etc. (see Figure 2). Formerly these tutorials were written in HTML, but now we use L^AT_EX.

From L^AT_EX it is simple to generate PostScript and PDF documents. The `latex2html` program, which is used to generate HTML, is fragile. Thus we translate every tutorial separately to generate the correct HTML pages. If this (automatic) translation process fails, then we only lose one tutorial. To generate the navigation in the left frame in Figure 2, we create a L^AT_EX-document containing all tutorials. We analyze the table of contents file, generated by L^AT_EX, with a self-written C program (`toc2html`) and generate the frames for the navigation in the document (see Figure 5).

3 The Golden Solution

Notice the link to `nn1.xpl` in the middle of Figure 2. In the help system as well as in the XploRe Learning Guide (Härdle et al. 2000a), this link goes to a simple ASCII page containing the source code of the quantlet `nn1.xpl`. It is used to generate the pictures in the neural network tutorial.

In the last few years, we started to develop XploRe as client/server software in addition to the standalone versions for Unix/Linux and Windows. The client is written in Java and can run as an application as well as an applet in a HTML page (Lehmann and Kleinow 2001).

Thus the link `gplm02` in Figure 4 goes to a HTML page which allows us to embed the XploRe client applet. This technique works for HTML documents (internet, CD) as well as for PDF documents (internet). In the left frame of Figure 3, we see that we can do two interaction types: *Edit* and *Execute*.

The generation of these pages consists of two steps: 1) generation of the HTML pages (the right frame in Figure 3) and the quantlet overview; 2) the generation of the HTML pages for navigation and the XploRe client applet. For the first step, we use a Perl script. It is a modification of a Perl script used previously to generate the XploRe helpsystem. Now, the help system is generated with the command `makehelp` within XploRe. For the second step, we use a makefile which generates the HTML pages and necessary files for the XploRe client applet from a set of master files, e.g., using `sed`.

4 The MM*Stat project

Another step in the development of MD*Book was the MM*Stat project, which supports the teaching of basic statistics (Rönz 2001). It is an example of a short book which may become a textbook in the future. The basic structure is given by the table of contents which divides up the different lectures. These lectures are supported by “More Information”, “Explained Example” and other enhancements. The two interaction enhancements are *Interactive* and *Computing*. The linking does not follow the book structure, e.g., we have links to and from the glossary or the direct jump from one “Multiple choice” enhancement to another.

Most of the outline of the MM*Stat project is done via Cascading Style Sheets, frames, and JScript. The interaction enhancements are again realized as XploRe client applets within a HTML page.

The german version of MM*Stat (Härdle, Rönz, 2001) is completely written in HTML. All other foreign language versions are written in \LaTeX and MD*Book is used to generate the HTML versions (see Figure 6). One of the reasons for using \LaTeX was that the students wanted to have a *printed* version of MM*Stat which turns out to be very difficult using HTML pages.

5 The MD*Book project

The MD*Book project emerged directly from our translation of the \LaTeX tutorials in the XploRe helpssystem to HTML. During an analysis of an `*.toc` file, generated by \LaTeX , we noticed some serious difficulties. Mainly, they result from the fact that the internal format used in this file differs slightly from standard \LaTeX . Moreover, the integration of formulas or special math symbols was not possible.

This leads to another approach for analyzing the \LaTeX files. We developed two programs: `tex2sk` and `mdbook1`. The program `texsk` decomposes a \LaTeX file set and generates a control file. This control file is used in `mdbook1` to generate all necessary information (see Figure 7).

`mdbook1` has five main options: `-ps`, `-pdf`, `-html`, `-java` and `-xpl`. `-ps` and `-pdf` generate a PostScript document or a PDF document, respectively. Since the translation from \LaTeX to HTML takes time, these options were planned to determine, if everything compiles under \LaTeX without problems.

The `-html` and `-java` options both generate HTML versions: `-html` without using JavaScript in the HTML pages and `-java` generates a MM*Stat-like version. A difference between the two versions is that the first one runs under Netscape and Internet Explorer whereas the intensive use of JScript in MM*Stat leads to the current situation that Internet Explorer 5.x is necessary to run the MM*Stat versions.

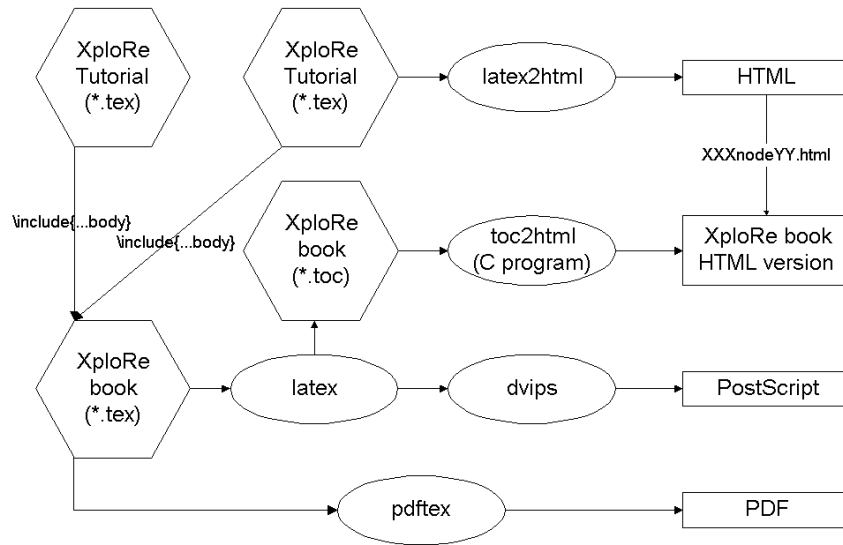


Figure 5: Overview of the generation of the XploRe tutorials.

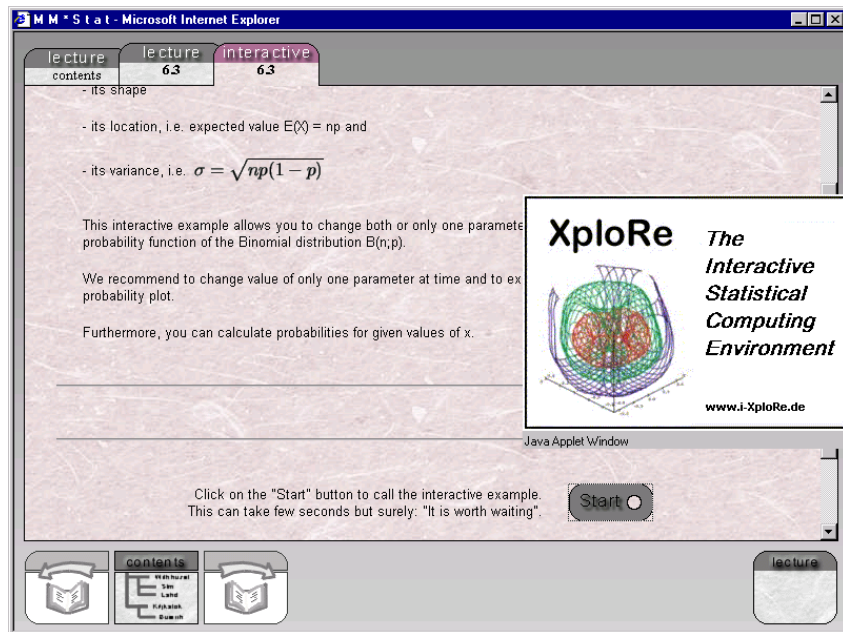


Figure 6: *Interactive* example in the english MM*Stat version generated with MD*Book; just starting the XploRe client. It contains an *computing* example as well.

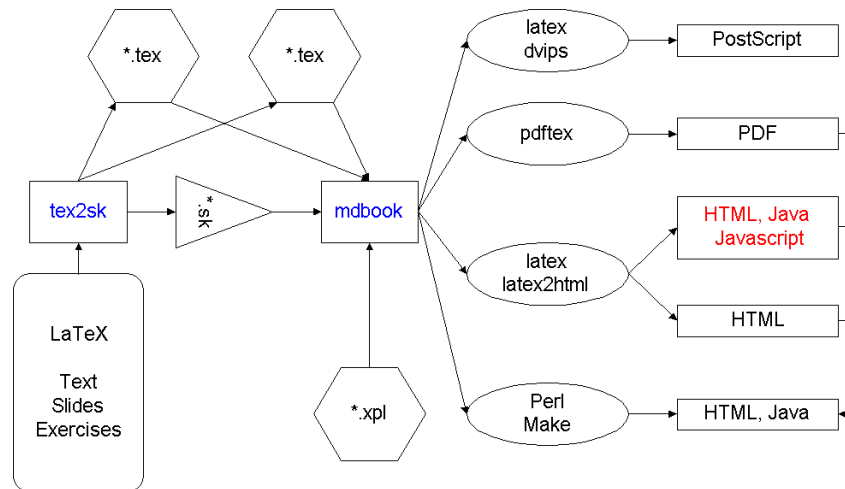


Figure 7: Generation of an *e-book* from L^AT_EX sources and XploRe quantlets with MD*Book.

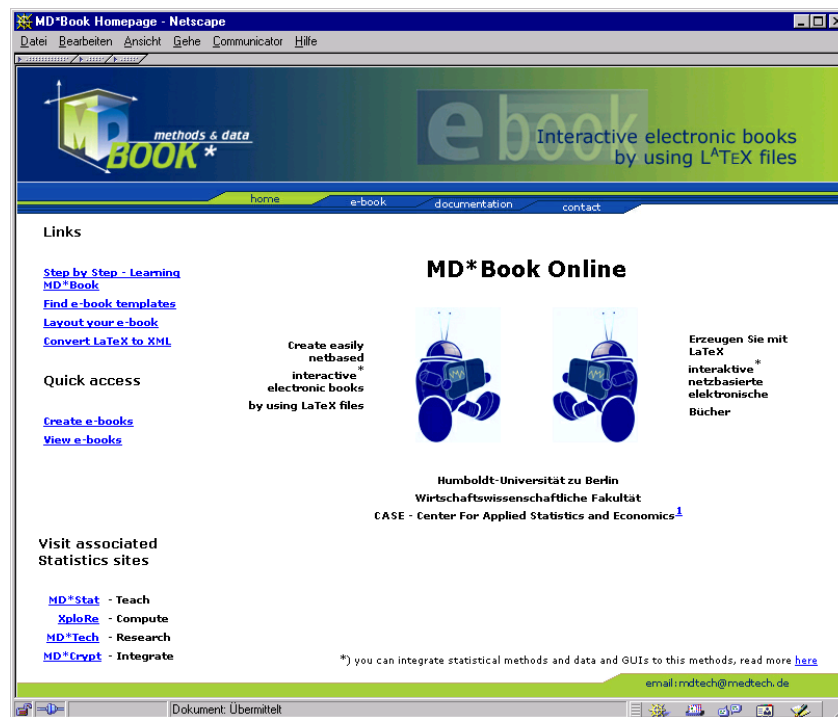


Figure 8: Entry page of the Web based interface of MD*Book: www.md-book.com

The last option, `-xp1`, generates the “Golden Solution” for XploRe quantlets.

The control file contains, beside the document structure, other informations, e.g., if we generate a CD or an internet version, where are the final directories that need to be copied etc.. Depending on the main option we can include or exclude blocks of \LaTeX files. It is even possible to handle several documents with one control file which allows the user to ensure uniformity of the structure in these documents.

A set of parameters in the control file and \LaTeX files, e.g., for the `MM*Stat` format (currently 170), allow the user to control various aspects of the outline and appearance of the document. On the `MD*Book` online pages you can find currently three different basic `MD*Booklet` formats (one of them is the `MM*Stat` format).

6 MD*Book online

`MD*Book Online` wraps `MD*Book` so that it can be used in internet browsers and via FTP clients. `MD*Book` itself can only be used with a valid UNIX account in our intranet and it is started by entering command lines in a UNIX shell. The desire to use `MD*Book`'s sophisticated abilities to create interactive books outside of our intranet leads to the development of such a wrapper.

`MD*Book` is a classical sequential algorithm (in contrast to distributed algorithms). Therefore you can consider its behaviour a function of its inputs and its environment.

`behaviour(inputs, environment(hardware, software, access rights))`

Behaviour includes all output (streams), allocations of resources (file system, memory etc.) over time, executability and program crashes.

From this simple model of behaviour results that the environment has to be hold constant or to be restrained at least in order to preserve the desired behavior for every possible input when you try to transfer an algorithm to another computer.

Therefore the easiest way to transfer an algorithm is transferring an installer and copies of the executable files. The installer adjusts the application software and the operating system and requests access rights. An additional list of requirements determines the set of usable hardware and operating systems. In terms of the prementioned model the complete environment of a program is stabilized by its installer. But keep in mind that in general the usable hardware and operating systems are highly restricted since the development of installers for several operating systems is too expensive.

Some algorithms are not associated with a particular operating system (e.g.

fore they are already available for a large variety of hardware and operating systems.

But client/server systems are always distributed algorithms which means that you have to care about synchronising clients and distributing resources among clients. Synchronising mechanisms have to work when MD*Book is started repeatedly in the same directory. The confidential nature of e-book sources requires authentic access to the server's resources. To solve those two problems we introduce work spaces and locks that prevent running more than one MD*Book at the same time.

Because we use only free widely available client/server systems in our implementation of MD*Book Online we also developed a general model and a general implementation of wrappers for command line software.

The workflow of an e-book author consists of storing texts, diagrams and interactive items via FTP in a MD*Book Online work space first, initializing the production of a desired e-book format by Java applet and finally downloading produced books via FTP or internet browser. For all other tasks, like viewing documentation, templates for e-books and e-book previews, it is sufficient to use an internet browser.

7 Conclusion and acknowledgements

With MD*Book and MD*Book online we have build a tool which allows us and others to easily generate e-books as seen by the number of e-books and MD*Booklets on our Web pages (see last two internet adresses at the end).

Finally, we would also like to thank our colleagues, especially Marlene Müller, Torsten Kleinow and Jörg Feuerhake for contributing knowledge and experience to the MD*Book program and Zdeněk Hlávka for extensive testing.

References

- Bianchi, M. (1995): Bandwidth Selection in Density Estimation. In *XploRe: An Interactive Statistical Computing Environment*, Härdle, W. et al. (eds.), 101–112, Springer Verlag, New York.
- Härdle, W., Klinke, S. and Turlach, B.A. (1995): *XploRe: An Interactive Statistical Computing Environment*. Springer Verlag, New York.
- Härdle, W., Klinke, S. and Müller, M. (2000a): *XploRe Learning Guide*. Springer Verlag, Heidelberg.
- Härdle, W., Hlávka, Z. and Klinke, S. (2000b): *XploRe Application Guide*. Springer Verlag, Heidelberg.

Härdle, W. and Rönz, B. (2001): *MM*Stat - Eine interaktive Einführung in die Welt der Statistik*. Springer Verlag, Heidelberg.

Klinke, S. (1997): *Data Structures for Computational Statistics*, Physica-Verlag, Heidelberg.

Klinke, S., Golubev, Y., Härdle, W. and Neumann, H. (1998): Teaching wavelets in XploRe. *Computational Statistics*, **13**, 141–151.

Lehmann, H. and Kleinow, T. (2001): *Client/Server based Statistical Computing*. Proc. of the ISM symposium “Statistical software in the Internet age”. ISM, Tokyo, 2001.

Proenca, I. (1995): Interactive Graphics for Teaching Simple Statistics. In *XploRe: An Interactive Statistical Computing Environment*, Härdle, W et al. (eds.), 113–140, Springer Verlag, New York.

Rönz, B. (2001): *The Multimedia-Project MM*Stat for Teaching Statistics*. Proc. of the ISM symposium “Statistical software in the Internet age”. ISM, Tokyo, 2001.

Zanter, K. (1998): *Untersuchung des Berliner Grundstückmarktes 1995–1997. Entwicklung statistischer Software und Anwendung zur Datenanalyse*. Diploma thesis at Humboldt-Universität zu Berlin (in German).

Internet addresses

XploRe 3: <http://ise.wiwi.hu-berlin.de/~sigbert/xplore3>
 XploRe 4: <http://www.xplore-stat.de>

E-Books: <http://www.xplore-stat.de/~ebooks>
 MM*Stat/
 MD*Booklet: <http://www.md-stat.com>
 MD*Book
 — online: <http://www.md-book.com>
 — Mailingliste: <mailto:mdbook@rz.hu-berlin.de>

Sigbert Klinke <http://ise.wiwi.hu-berlin.de/~sigbert>
<mailto:sigbert@wiwi.hu-berlin.de>
 Rodrigo Witzel <http://ise.wiwi.hu-berlin.de/~witzel>
<mailto:witzel@wiwi.hu-berlin.de>