

Seminar
Datenreinigung
LaTeX Einführung

12.6.2007
Jens Bleiholder

Überblick

2

- TeX und LaTeX Geschichte
- Motivation
- Die Basics
- Die Seminar-Vorlage
- Online und offline Hilfe
- Software:
 - Distributionen
 - Editoren
 - Installationen am Institut

- Aussprache:
 - English: Tec oder Tech
 - Deutsch: Tech (Griechisches Chi)
- Textformatierungssprache
 - Entwickelt für Wissenschaftler

```
\proclaim(Theorem) The value of
\int_{-\infty}^{\infty} e^{-x^2}\,dx is
\int_{-\infty}^{\infty} e^{-x^2}\,dx=\sqrt{\pi}.
```

THEOREM. The value of $\int_{-\infty}^{\infty} e^{-x^2} dx$

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}.$$

- Entwickelt von Donald Knuth, späte 70er
 - Motivation: 2ter Band „The Art of Computer Programming“
 - 10 Jahre Entwicklung
- Frei verfügbar, Version 3.141592

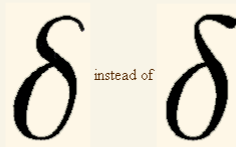
Donald Knuth

- <http://www-cs-faculty.stanford.edu/~knuth/>



Important Message to all Users of TeX

If you see that your system produces the symbol



for the Greek lowercase delta, you should tell your system administrator immediately to *upgrade your obsolete version of the Computer Modern fonts*.

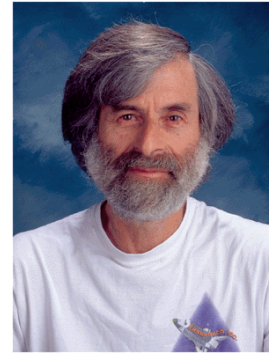
I made important corrections to all those fonts in the spring of 1992, but alas, I still see many books, journals, and preprints using the old versions. Please help me abolish the old forms from the typefaces of the earth.

- Erweiterung von TeX
- Erste Version: LaTeX 2.09, 1985
- Aktuelle Version: LaTeX 2 ϵ , 1994
- Ständige Erweiterung durch Pakete

Leslie Lamport

7

- Microsoft Research
- Andere Forschung
 - Synchronisation verteilter Uhren
 - Grundlegende Theorien
 - Wichtig bei Transaktionen



Motivation

8

- Aussprache
 - Lahtech oder Laytech oder Laytec
- Warum sollte ich LaTeX benutzen?
 - Logisches Design statt visuelles Design
 - Design Makros
 - Z.B. mathematische Umgebungen
- Trennung von Inhalt und Design beim Schreiben
 - Konzentration auf die Aussage des Textes

Motivation gegenüber Word et al.

9

Vorteile

- Umsonst
- Erweiterbar
 - Hunderte Pakete verfügbar
- Programmierbar
- Portierbar (ASCII)
- Schön
- Skalierbar
- Formeln

Nachteile

- Kein WYSIWYG
- Schwieriger zu lernen
- Programmiersprache statt Desktopanwendung
- ...nichts für Anfänger

LaTeX – Das Grundprinzip

10

- Erstellung eines ASCII Textes mit beliebigem Editor
 - Unter Verwendung von LaTeX Befehlen
- Kompilierung durch LaTeX
- Ansicht oder Ausdruck

\documentclass{article} % Der Input-Datei muss mindestens diese beiden
\begin{document} % Zeilen und den \end{document} Befehl am Ende
 enthalten.

\section{Einfacher Text} % Dieser Befehl erzeugt eine Abschnittüberschrift.

Wörter werden durch ein oder mehr Leerzeichen getrennt. Absätze werden durch eine oder mehr Leerzeilen getrennt. Der output wird durch extra white-space in der Input Datei nicht beeinflusst.

Doppelte Anführungszeichen werden mit zwei ` ` einzelnen Anführungszeichen" gesetzt.

Einfache Anführungszeichen werden ` direkt gesetzt'.

Lange Bindestriche werden als drei einzelne Striche gesetzt---so.

Betonter text wird so gesetzt: **\emph{Dies ist betont}**.

Fetter Text wird so gesetzt: **\textbf{Dies ist fett}**.

\end{document} % Die Inputdatei endet mit diesem Befehl.
 Jens Bleiholder | Seminar: Datenreinigung | SS 2007

Freier Text

- Whitespace (tab, space) wird ignoriert.
- Alleinige Absatzzeichen werden ignoriert.
 - Absätze durch Leerzeile
- Satzzeichen werden als solche erkannt.
 - Längerer Space
 - Abhilfe mit „dies ist nur eine Abk.\ und der...“
- Reservierte Symbole:
 - \$ & # % _ { } ~ ^ " < > | und \
 - Abhilfe:
 - \\$ \& \# \% _ \{ \}
 - bzw. \backslash
 - bzw. `` ` `
 - bzw. in Formeln: \$a < b\$

LaTeX – The Basics

13

- Befehle beginnen mit „\“
 - Z.B. `\maketitle`
- Umgebungen werden mit „{...}“ umschlossen
 - Z.B. `\emph{Dies wird betont.}`
- Parameter werden mit „[]“ umschlossen
 - Z.B. `\documentclass[a4paper,11pt]{book}`
- Kommentare beginnen mit % und enden am Zeilenumbruch.

LaTeX – Frequent Symbols

14

- -, --, ---
- Großbuchstaben am Satzende: Ich verwende UNIX\@.
- `\latex`, `\tex`
- `\dots`: ...
- Dieser~Text~soll~nicht~umgebrochen~werden.
- `\verb+` Dieser Text erscheint unverändert.+

LaTeX – Frequent Commands

15

- Dies ist `\emph{wichtig}`. (meist kursiv)
- Dies ist `\textbf{fett}`.
- Dies ist `\textit{kursiv}`.
- Dies ist `\underline{unterstrichen}`.
- Dies ist eine `\footnote{Fussnote}`.
 - Nummerierung und Satz automatisch.

LaTeX – Deutsche Sprache

16

- Umlaute
 - Entweder direkt: ä ö ü ß
 - Oder mit Anführungszeichen: "a "o "u "s bzw. `\ss`
- Rechtschreibung
 - Je nach Editor
 - "ck für korrekte Trennung in k-k
- Anführungsstriche
 - Automatisch durch `\usepackage{ngerman}`

Dokumentklassen

17

- Article
 - Für kurze Texte
- Report
 - Für längere Texte
 - Kaum Unterschied zu article
- Book
 - Erlaubt Kapitel (`\chapter{}`), nicht nur Abschnitte (`\section{}`)
 - Erlaubt `\frontmatter`, `\mainmatter` und `\backmatter`
- Letter
 - Erlaubt `\address`, `\signature`, etc.
- Slides
 - Nicht besonders empfehlenswert
 - Außer bei vielen Formeln

Jens Bleiholder | Seminar: Datenreinigung | SS 2007

LaTeX – Abbildungen

18

- .eps (encapsulated postscript)
 - Jpeg2ps
 - Adobe Acrobat
- .jpg, .gif etc.
 - Nur in pdflatex
- `\includegraphics{arch.eps}`
- `\includegraphics[width=5cm,height=1cm]{arch.eps}`
- `\includegraphics[width=\textwidth]{arch.eps}`
- `\includegraphics[width=0.5\textwidth]{arch.eps}`

```

\begin{picture}(100,120)(0,0)
\thicklines
\setlength{\unitlength}{2pt}

\put(10,10){\makebox(0,0){\Join}} % DEPNUM = DEPNO
\put(60,110){\makebox(0,0){\Join}} % PNO = PNUMBER
\put(95,60){\makebox(0,0){\Join}} % SSN = ESSN

\put(73,44){\makebox(0,0)[r]{\tiny EMPLOYEE}}
\put(73,40){\vector(1,1){10}} % Works_On
\put(97,44){\makebox(0,0)[l]{\tiny PROJECT}}
\put(97,40){\vector(-1,1){10}} % Project
\put(90,62){\makebox(0,0)[l]{DEPNUM \neq DEPNO}}

\put(-2,-6){\makebox(0,0)[r]{\tiny WORKS_ON}}
\put(-2,-2){\vector(1,1){10}} % Employee
\put(22,-6){\makebox(0,0)[l]{\tiny EMPLOYEE}}
\put(22,-2){\vector(-1,1){10}} % Works_On
\put(5,12){\makebox(0,0)[r]{SSN = ESSN}} %

\put(60,110){\makebox(0,0){PNO = PNUMBER}} %

\put(13,15){\vector(1,2){45}} % SSN = ESSN -> DEPNUM = DEPNO
\put(82,55){\vector(-1,2){20}} % PNO = PNUMBER -> DEPNUM = DEPNO

\end{picture}

```

Jens Bleiholder | Seminar: Datenreinigung | SS 2007

- `\begin{figure}[ht]`
 - `\centering`
 - `\includegraphics...`
 - `\caption{Dies ist eine Abbildung}`
 - `\label{abbildung1}`
- `\end{figure}`
- Floating, d.h. die Abbildung kann im Text verschoben auftauchen

- Windows
 - Powerpoint
 - Jfig:
 - <http://tech-www.informatik.uni-hamburg.de/applets/jfig/download.html>
 - Malprogramme
 - Umwandlung:
 - Jpeg2ps:
 - <http://www.pdfplib.com/products/more/jpeg2ps.html>
 - Adobe Acrobat
- Xwindows
 - Xfig
 - gimp

- `\begin{tabular}{c|l} \hline`
`Name & Jahr \\ \hline`
`Schultz & 1971 \\`
`Meyer & 2001 \\ \hline`
`\end{tabular}`
- Positionierungen: lcr
 - `\begin{tabular}{c|l|r}`
- `\multicolumn{2}{c}{Text}`

Name	Jahr
Schultz	1971
Meyer	2001

- `\begin{table}[ht]`
`\centering`
`\begin{tabular}{l|c}`
`...`
`\end{tabular}`
`\caption{Dies ist eine Tabelle}`
`\label{tabelle1}`
`\end{table}`
- Floating

- Im Text erscheinen Formeln so: $3 - 2 = 1$.
- Innerhalb der $\$...\$$ Umgebung gelten andere Regeln:
 - Sonderzeichen: $+$, $-$, $<$, $>$, etc.
 - Buchstabentrennung (Text innerhalb von Formeln mit `\mbox{...}` umschliessen)
 - In der Formel $5 \times 6 = \mbox{dreissig}$ kommt Text vor.

- `\leq`, `\geq`
- `\frac{3}{4}`, `\sqrt{25}`
- `\ldots`, `\cdots`, `\vdots`, `\ddots`
- Hoch- und Tiefstellung
 - n^2 ergibt n^2 ,
 - n_i ergibt n_i ,
 - n_{i+1}^{11} ergibt n_{i+1}^{11} (bzw. n_{i+1}^{11})
- `\alpha`, `\beta`, `\omega`, `\Omega`
- `\cap`, `\cup`, `\leftarrow`, `\rightarrow`
- `\sum`, `\prod`, `\int`, `\log`, `\sin`

LaTeX – Formeln

25

- Abgesetzte Formeln
- `\begin{equation}`
...
`\end{equation}` (mit Nummern)
- `\[... \]` (ohne Nummern)
- `\begin{eqnarray}`

$$x = 17y$$

$$y = 125x$$
`\end{eqnarray}`

LaTeX – Listen

26

- | | |
|--------------------------------|------------|
| <code>\begin{itemize}</code> | ■ Schultz |
| <code>\item Schultz</code> | ■ Meyer |
| <code>\item Meyer</code> | |
| <code>\end{itemize}</code> | |
| <code>\begin{enumerate}</code> | |
| <code>\item Schultz</code> | 1. Schultz |
| <code>\item Meyer</code> | 2. Meyer |
| <code>\end{enumerate}</code> | |
| <code>\begin{itemize}</code> | |
| <code>\item Schultz</code> | ■ Schultz |
| <code>\begin{enumerate}</code> | 1. Schultz |
| <code>\item Schultz</code> | 2. Meyer |
| <code>\item Meyer</code> | |
| <code>\end{enumerate}</code> | ■ Meyer |
| <code>\item Meyer</code> | |
| <code>\end{itemize}</code> | |

LaTeX – Eigene Befehle

27

- `\newcommand{\be}{\begin{enumerate}}`
- `\newcommand{\betonen}[1]{\emph{#1}}`
- `\newenvironment{\emphit}`
`{\begin{itemize}\em}{\end{itemize}}`
- `\newtheorem{bspl}{Beispiel}`
`\begin{bspl}Test \end{bspl}` ergibt:
 Beispiel 4: Test

LaTeX – Labels und Referenzen

28

- `\section{Einleitung}`
`\label{sec: einleitung}`
 ...
 - `\subsection{Struktur der Arbeit}`
`\label{subsec: struktur}`
 ...
 - `\section{Zusammenfassung}`
`\label{sec: zusammenf}`
- Wie schon in Abschnitt~`\ref{sec: einleitung}` auf
 Seite~`\pageref{sec: einleitung}` erwähnt, ...

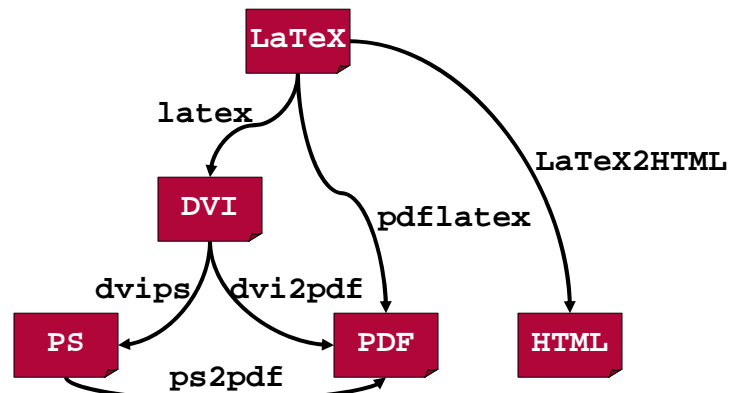
LaTeX – Labels und Referenzen

29

- `\begin{theorem}`
`\label{the:gauss}`
`...`
`\end{theorem}`
- Für figure und table, siehe vorher.
 - Wichtig: Immer erst nach der caption!
- `\begin{equation}\label{eqn:test}`
`...`
`\end{equation}`
- Referenzen immer gleich: `\ref{eqn:test}`
 - Meist ~ voranstellen

LaTeX – Output

30



- **LaTeX Warning:** Label(s) may have changed. Rerun to get cross-references right.
 - Einfach nochmals latex laufen lassen
- **LaTeX Warning:** There were multiply-defined labels.
- **Runaway argument?**

```
{test
! Paragraph ended before \label was complete.
<to be read again>          \par 1.33
```

 - In Zeile 33 nachsehen (oder 32, oder 34)
 - Abbruch mit q, ignorieren mit r, alles ignorieren mit R
- **Underfull \hbox ...**
 - Nicht so schlimm
- **Overfull \hbox....**
 - Etwas schlimmer: Es ragt etwas über den Rand hinaus.
 - Trennvorschläge machen: Da\~ten\~bank

- dvi = device independent
- Standard Output für LaTeX
- Previewer
 - Windows: YaP
 - Xwindows: xdvi
 - binden eps Dateien ein

LaTeX - postscript

33

- Viewer
 - Windows: Ghostscript und Ghostview
 - <http://www.cs.wisc.edu/~ghost/>
 - Adobe Acrobat
 - Xwindows: gv und kghostview (KDE)
- Kann direkt an postscript Drucker geschickt werden.

LaTeX – Output (Hilfsdateien)

34

- | | |
|----------------------|--------------------|
| ■ .aux | ■ .toc |
| □ Auxiliary file | □ Table of Content |
| □ Referenzen, Zähler | ■ .lot |
| ■ .log | □ List of tables |
| □ Fehlermeldungen | ■ .lof |
| ■ .bbl | □ List of Figures |
| □ Bibliographie | ■ .idx |
| ■ .blg | □ Index |
| □ BibTeX log | ■ ... |

- Einbindung durch `\usepackage[optionen]{paketname}`
 - Noch vor `\begin{document}`
- <http://www.ctan.org/tex-archive/help/Catalogue/catalogue.html>
- Tipp: Konservativ mit Paketen umgehen.
- MikTeX: Automatisches Herunterladen und Installieren von Paketen
 - Sonst: Manuell entpacken und in die richtigen Verzeichnisse stellen

- XWindows
 - DeTeX
 - LaCheck
 - wc (word count)
- Windows
 - LaCheck
 - WC, z.B. mit WinEdt
- MS Word
 - Grammatik und Rechtschreibung

- Pakete
 - seminar
 - pdfslide
 - foiltex
 - ...
- Dokumentklasse
 - slides
- Vorführung
 - Ausdruck auf Folien und OH Projektor
 - Acrobat oder Ghostview und Beamer

Referenzen in .bib Datei schreiben

- @ARTICLE{Hernandez98,
 author = {Mauricio A.\ Hern\andez and Salvatore J.\ Stolfo},
 title = {Real-world Data is Dirty: Data Cleansing and The Merge/Purge Problem},
 journal = {Data Mining and Knowledge Discovery},
 year = {1998},
 volume = {2(1)},
 pages = {9-37}}
- @STRING{ICDE = {Proceedings of the Int.\ Conference on Data Engineering (ICDE)}}
 @InProceedings{Galhardas00,
 AUTHOR={Helena Galhardas and Daniela Florescu and Dennis Shasha and Eric Simon},
 TITLE={An Extensible Framework for Data Cleaning},
 BookTitle=ICDE,
 ADDRESS={San Diego, CA},
 PAGES = {312},
 YEAR=2000}
- @BOOK{Ceri84,
 AUTHOR = {Stefano Ceri and Giuseppe Pelagatti},
 ADDRESS = {New York},
 PUBLISHER = {McGraw-Hill Book Company},
 TITLE = {Distributed Databases: Principles and Systems},
 YEAR = {1984}}

- Zitieren im Text mit ... `\cite{Hernandez98}`
- Aufruf-Reihenfolge:
 - latex sem (Finden der neuen `\cite{}` Befehle)
 - bibtex sem (Erzeugung der bbl Datei mit formatierten Referenzen)
 - latex sem (Einbinden der Referenzen)
 - latex sem (Gegenenefalls re-formatieren)

Fernuni Hagen:

- PDF, 122 Seiten: http://www.fernuni-hagen.de/URZ/urzbib/ls_broschueren.html#TuG

Uni Giessen: Kochbuch

- <http://www.uni-giessen.de/hrz/tex/cookbook/cookbook.html>

Newsgroups

- de.comp.text.tex mit 11-teiliger FAQ
- <http://www.dante.de/faq/de-tex-faq/>
- comp.text.tex

Dante: <http://www.dante.de/>

- mit downloadserver: <http://www.dante.de/software/ctan/>

Bücher

41

- Leslie Lamport
- Das LaTeX Handbuch



Jens Bleiholder | Seminar: Datenreinigung | SS 2007

Bücher

42

- Goossens, Mittelbach & Samarin
- Der LaTeX Begleiter
- Beschreibung von 100+ Paketen

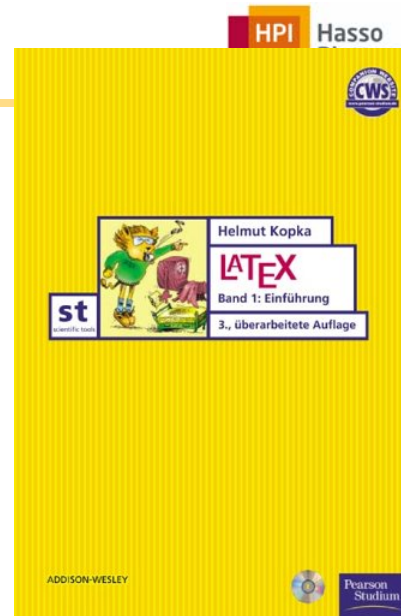


Jens Bleiholder | Seminar: Datenreinigung | SS 2007

Bücher

43

- Helmut Kopka
- LaTeX I. Einführung
- LaTeX II. Ergänzungen
- LaTeX III. Erweiterungen

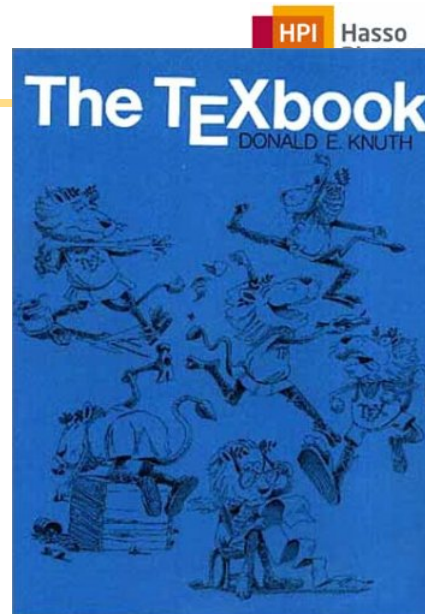


Jens Bleiholder | Seminar: Datenreinigung | SS 2007

Bücher

44

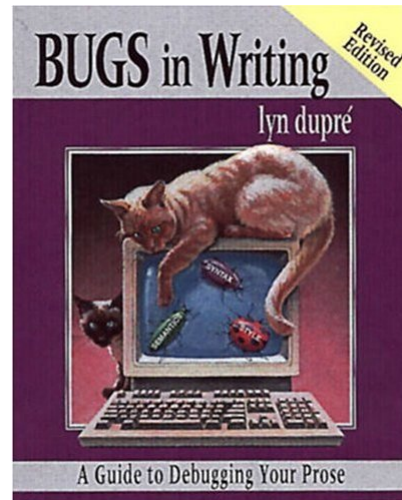
- Donald Knuth
- The TeXbook



Jens Bleiholder | Seminar: Datenreinigung | SS 2007

45

- Bugs in Writing
- Lyn Dupré
- 20.90 Euro bei Amazon



Jens Bleiholder | Seminar: Datenreinigung | SS 2007

46

- Für Windows
 - MikTeX: www.miktex.org
 - emTeX, BaKoMa, fpTeX usw.
- Für Linux
 - Redhat: teTeX
 - Suse etc. ?

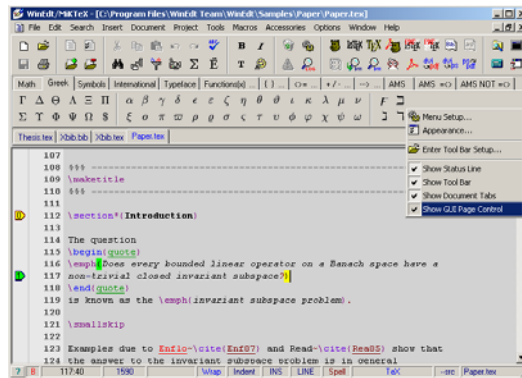
Jens Bleiholder | Seminar: Datenreinigung | SS 2007

Windows Editoren

47

WinEDT:

- www.winedt.com
- 31 Tage trial, dann 40\$



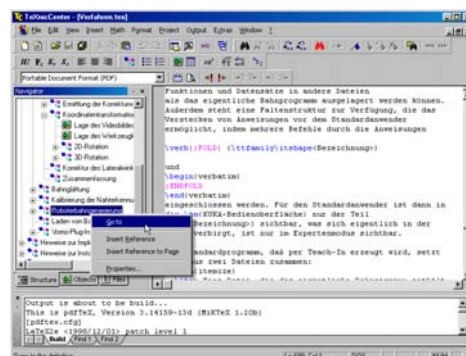
Jens Bleiholder | Seminar: Datenreinigung | SS 2007

Windows Editoren

48

TeXnic Center

- <http://www.toolscenter.org/products/texniccenter/index.htm>

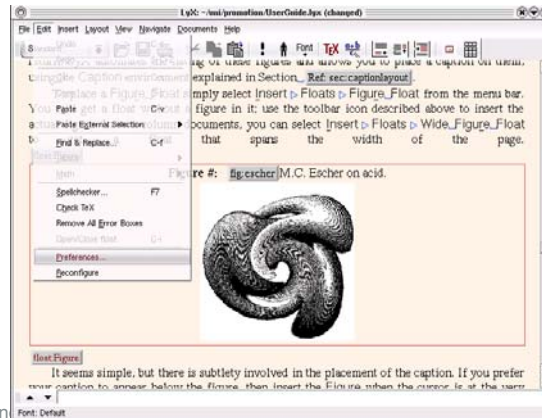


Jens Bleiholder | Seminar: Datenreinigung | SS 2007

Windows Editoren

49

- LyX: <http://www.lyx.org/>
- WYSIWYM



Jens Bleiholder | Seminar: Datenreinigung

Linux Editoren

50

- Xemacs
 - LaTeX-mode ist standard
 - Syntax-Highlighting aktivieren!
 - BibTeX - mode

Jens Bleiholder | Seminar: Datenreinigung | SS 2007

Viel Spaß beim TeXen!

51

LATEX