

Einführung in das Textsatzsystem



L^AT_EX



02 – Schriften und Kodierungen

THE LION'S WRITING SKILLS

25. Oktober 2013

Inhalt

- 1 Kurze Geschichte des Schreibens
- 2 Kodierungen, Zeichensätze
- 3 Was ist eine Schrift?
- 4 Schriften in TeX
- 5 Moderne Schrifttechnologien
- 6 Exkurs: Neo-Tastaturlayout
- 7 LuaLaTeX, Unicode, OpenType
- 8 Tips und Tricks für die Arbeit mit TeXworks

“Throughout the millennia humans have developed and adapted methods for storing facts and thoughts on a variety of different media.”

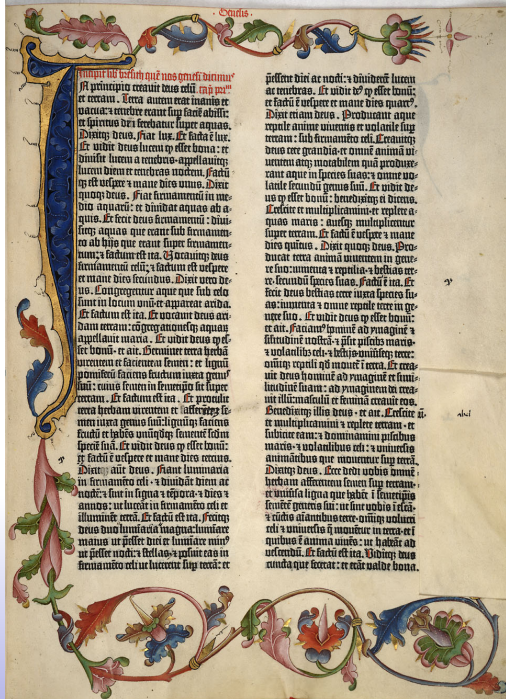
Schreiben ohne Automatisierung

- Am Anfang stand die Handschrift:
Feder (führt Handbewegung direkt aus), Keil (wird in Ton gedrückt),
Meißel (für monumentale Inschriften) u. ä. ...
⇒ flexibel, schnell, dauerhaft – aber jeweils unikat
- Diskus von Phaistos vermutlich erstes Druckwerk *mit beweglichen Lettern*
- Trajanssäule in Rom stellt bis heute die ästhetische Grundform der lateinischen Großbuchstaben!

Schreiben ohne Automatisierung

- Am Anfang stand die Handschrift:
Feder (führt Handbewegung direkt aus), Keil (wird in Ton gedrückt), Meißel (für monumentale Inschriften) u. ä. ...
⇒ flexibel, schnell, dauerhaft – aber jeweils unikat
- Diskus von Phaistos vermutlich erstes Druckwerk *mit beweglichen Lettern*
- Trajanssäule in Rom stellt bis heute die ästhetische Grundform der lateinischen Großbuchstaben!
- Druck mit beweglichen Lettern in China lange bekannt (wenn auch recht unpraktisch bei tausenden von Schriftzeichen)
- Gutenberg übertrug das Verfahren auf lateinische Schrift
- Ziel: höchste Qualität der Schriftstücke (Behauptung gegen Handschrift)
- Handsatz ist ein extrem aufwändiges, zeitraubendes und technisch anspruchsvolles Verfahren, aber für höchste Schriftqualität geeignet

42-zeilige Gutenbergbibel
in zweispaltigem Satz
mit aufwändigen
Verzierungen



Analoge Setzmaschinen: Die Linotype

- Name von „Line-of-Type“
 - halbautomatisches Druckverfahren:
 - Schreibprozess ähnlich einer Schreibmaschine
 - einzelne Zeilen werden direkt in Blei gegossen
- ⇒ schnelle Anfertigung für Massenproduktion möglich

Digitaler Satz

- Digitale Textverarbeitung:
- Schnelle, einfache Eingabe
- Automatisierte Anpassung des Designs
- Blocksatz und Mikrotypographie in einer Qualität, die nur von aufwändigem Bleisatz übertroffen werden kann
- Beliebig schnelle Vervielfältigung, oft ohne materielles Medium
- eBooks als neues Standardformat haben gar keine materielle Vorlage

Digitaler Satz

- Digitale Textverarbeitung:
- Schnelle, einfache Eingabe
- Automatisierte Anpassung des Designs
- Blocksatz und Mikrotypographie in einer Qualität, die nur von aufwändigem Bleisatz übertroffen werden kann
- Beliebig schnelle Vervielfältigung, oft ohne materielles Medium
- eBooks als neues Standardformat haben gar keine materielle Vorlage

- auch T_EX wurde im Laufe der Jahre weiterentwickelt
 - siehe `texdoc tex-overview` (auch zur Klärung vieler T_EX-bezogener Begriffe)

Wie kommt das „A“ ins pdf?

Der lange Weg einer Information vom Tastenanschlag bis zum Papier ...

- Tastendruck erzeugt eine Zahl (40 für die Taste, auf der ein „a“ ist, 62 bzw. 50 für die Shift-Tasten)
- Betriebssystem wandelt diese Zahl oder Kombination aus Zahlen (40 und 50 gleichzeitig gedrückt) in eine andere Zahl
- Editor bzw. graphische Oberfläche zeigt gemäß dieser Zahl eine Glyph
- Im Druck kann die Zuordnung Zahl \Leftrightarrow Glyph wiederum anders aussehen ...

Die ASCII-Kodierung

- Kodierungen sind abstrakte Listen, die einer Zahl eine Bedeutung zuordnen
- ASCII ist eine 7-bit-Kodierung
- Kodiert 128 Zeichen
- 95 druckbare Zeichen
- 33 nichtdruckbare Zeichen (Steuerzeichen, z. B. escape)
- Ausgelegt auf englische Sprache und Programmieren

Zeichenvorrat:

!"#\$%&'()*+,-./0123456789:;<=>?
@ABCDEFGHIJKLMN O PQRSTUVWXYZ[\]^_
`abcdefghijklmnopqrstuvwxyz{|}~

T_EX-Kodierungen

- T_EX1 konnte nur 7-bit-ASCII verstehen.
- T_EX3 kann auch 8-bit-Kodierungen verarbeiten.
- Zusätzlicher Zeichenvorrat durch spezielle Befehle (`\"a`)
- In pdfL^AT_EX zugänglich mittels `\usepackage[kodierung]{inputenc}`
- Verschiedene 8-bit-Kodierungen für verschiedene Sprachumgebungen

latin1: Westeuropäisch

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	nicht	belegt	(Steuerzeichen)													
1	nicht	belegt	(Steuerzeichen)													
2	!	"	#	\$	%	&	'	()	*	+	,	-	.	/	
3	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
6	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7	p	q	r	s	t	u	v	w	x	y	z	{		}	~	DEL
8	nicht	belegt	(Steuerzeichen)													
9	nicht	belegt	(Steuerzeichen)													
A		ı	¢	£	¤	¥	¦	§	¨	©	ª	«	¬	-	®	-
B	°	±	²	³	´	µ	¶	·	¹	º	»	¼	½	¾	¿	
C	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
D	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	Ý	Þ	ß
E	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
F	ð	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ý	þ	ÿ

latin2: Mitteleuropäisch

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0																
1																
2		!	"	#	\$	%	&	'	()	*	+	,	-	.	/
3	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
6	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7	p	q	r	s	t	u	v	w	x	y	z	{		}	~	DEL
8																
9																
A		Ą	˜	Ł	◌	Ł	Ś	Ś	˝	Š	Ş	Ť	Ž		Ž	Ž
B	◌	ą	˘	ł	◌	ł	ś	˘	˘	š	ş	ť	ž	˝	ž	ž
C	Ř	Á	Â	Ă	Ä	Ĺ	Č	Ç	Č	É	Ę	Ě	Ě	Í	Î	Ď
D	Đ	Ń	Ñ	Ó	Ô	Õ	Ö	×	Ř	Ů	Ú	Ů	Ü	Ý	Ť	ß
E	ř	á	â	ă	ä	ĺ	č	ç	č	é	ę	ě	ě	í	î	ď
F	đ	ń	ñ	ó	ô	õ	ö	÷	ř	ů	ú	ů	ü	ý	ț	·

latin5/9: Türkisch

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0																
1																
2		!	"	#	\$	%	&	'	()	*	+	,	-	.	/
3	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	_
6	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7	p	q	r	s	t	u	v	w	x	y	z	{		}	~	DEL
8																
9																
A		ı	¢	£	¤	¥	¦	§	¨	©	ª	«	¬		®	¯
B	°	±	²	³	´	µ	¶	·	¸	¹	º	»	¼	½	¾	¿
C	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï
D	Ğ	Ñ	Ò	Ó	Ô	Õ	Ö	×	Ø	Ù	Ú	Û	Ü	İ	Ş	ß
E	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï
F	ğ	ñ	ò	ó	ô	õ	ö	÷	ø	ù	ú	û	ü	ı	ş	ÿ

So viele Kodierungen ...

- Auf Dauer keine praktikable Lösung:
- Internationale Kommunikation wird erschwert
- Programme nicht lauffähig bei unterschiedlicher Kodierung
- Viele Sprachen haben Sonderzeichen, die berücksichtigt werden müssen
- Chinesisch oder Japanisch passen nicht in 256 Zeichen!

So viele Kodierungen ...

- Auf Dauer keine praktikable Lösung:
- Internationale Kommunikation wird erschwert
- Programme nicht lauffähig bei unterschiedlicher Kodierung
- Viele Sprachen haben Sonderzeichen, die berücksichtigt werden müssen
- Chinesisch oder Japanisch passen nicht in 256 Zeichen!

⇒ Unicode!

Unicode

The Unicode Standard is a character coding system designed to support the worldwide interchange, processing, and display of the written texts of the diverse languages and technical disciplines of the modern world. In addition, it supports classical and historical texts of many written languages.

<http://www.unicode.org/>

Unicode

The Unicode Standard is a character coding system designed to support the worldwide interchange, processing, and display of the written texts of the diverse languages and technical disciplines of the modern world. In addition, it supports classical and historical texts of many written languages.

<http://www.unicode.org/>

- Unicode bietet theoretisch eine immens große Zahl von Zeichen.
 - Für die Kodierung ist viel Speicher nötig (bis zu 32bit pro Zeichen!)
- ⇒ Bei utf8 werden variable Bitlängen verwendet
- ⇒ Lateinsprachige Texte sind klein, dennoch sind alle Zeichen verfügbar.

Was ist eine Schrift?

- Liste, die einer Zahl eine Glyphe zuordnet
- Eingabe des Codes 0061 muss nicht in einer Glyphe resultieren, die nach einem „a“ aussieht!
- Je nach Bedarf können Codepunkte völlig frei missbraucht werden (Schachfiguren, Feynmangraphen, Länderumrisse, ...)

Was ist eine Schrift?

- Liste, die einer Zahl eine Glyphe zuordnet
- Eingabe des Codes 0061 muss nicht in einer Glyphe resultieren, die nach einem „a“ aussieht!
- Je nach Bedarf können Codepunkte völlig frei missbraucht werden (Schachfiguren, Feynmangraphen, Länderumrisse, ...)

- z. B. Paket `CountriesOfEurope`:



- Schriftdateien können mit speziellen Programmen eingesehen werden
- [fontforge](#) ist eines der mächtigsten Schriftprogramme (freie Software!)

L^AT_EX-Schriftladen „in a nutshell“

- LuaL^AT_EX ermöglicht Verwenden beliebiger Systemschriften
- Paket `fontspec` stellt Nutzerinterface, Paket `luaotfload` macht die eigentliche Arbeit
- `\setmainfont` setzt eine Schrift als Hauptschrift, `\fontspec` kann innerhalb des Dokuments für kurze Passagen verwendet werden (weiteres später)
- Schriftnamen als Argument angeben
- Fette und kursive Varianten werden, sofern vorhanden, mitgeladen (mittels `\textbf`, `\textit` etc. verfügbar)
- Serifenlose Schnitte sind eigene Schrift und müssen separat geladen werden

L^AT_EX-Schriftladen „in a nutshell“

```
\setmainfont{Linux Libertine O}  
\begin{document}  
Text in Libertine ...  
{\fontspec{DejaVu Sans Mono} Text in DejaVu ...}  
Text in Libertine ...  
\end{document}
```

Kerning, Ligaturen

- Zeichen können nicht einfach hintereinander gesetzt werden:
- manche Zeichenkombinationen sehen „unschön“ aus
- Abstände können zu groß oder zu klein sein



box box file file

Kerning, Ligaturen

- Zeichen können nicht einfach hintereinander gesetzt werden:
- manche Zeichenkombinationen sehen „unschön“ aus
- Abstände können zu groß oder zu klein sein



Kerning, Ligaturen

- Zeichen können nicht einfach hintereinander gesetzt werden:
- manche Zeichenkombinationen sehen „unschön“ aus
- Abstände können zu groß oder zu klein sein
- Buchstabenformen können kollidieren (Lösung durch Ligatur):

box
box

file
file

Kerning und Ligaturen

- Kerning (Unterschneidung) und Ligaturen sind Schrifteigenschaften!
- anpassbar in *Sonderfällen* mittels `\kern`
- Im deutschen Satz: max. 3 Buchstaben bilden Ligatur (ffi, ffl, st)
- Im traditionellen arabischen Satz bis zu 7 Buchstaben!

Schriften


- Heutige Schriftformen durch römisches Alphabet bestimmt
- durch Kunsteinflüsse vielfältige Abänderungen der Grundformen
- wichtige Merkmale: Form der Bögen, Serifen, Neigung, Stärke

Schrifttypen

- Antiqua: „normale“ Schrift nach römischem Vorbild
- Grotesk: serifenlose Schrift
- gebrochene Schriften: Textura, Rotunda, Schwabacher, Fraktur
- gebrochene Grotesk, auch „Schaftstiefelgrotesk“: selten und hässlich ...

Schrifttypen – Bogenformen

Antiqua



keine Bogenbrechung → **o** **m** ← keine Bogenbrechung

Textura
(Gotisch)



Bogenbrechungen → **o** **m** ← Bogenbrechungen

Rotunda
(Rundgotisch)




Bogenbrechung nur angedeutet → **o** **m** ← Bogenbrechung nur angedeutet

Schwabacher



Bogenbrechung → **o** **m** ← Bogenbrechungen

Fraktur



Bogenbrechungen → **o** **m** ← Bogenbrechungen

T_EX-Schriften

- T_EX benötigt für eine Schrift mindestens zwei Dateien:
 - tfm (tex font metric)
enthält alle Informationen über die *Maße* der Glyphen (Breite, Höhe, Tiefe, Kerning)
 - Erscheinungsbild der Schrift wird erst im Viewer/Ausdruck festgelegt (ttf-Schrift, pk-Schrift o. a.)
- ⇒ Für Portabilität müssen Schriften ins Dokument *eingebunden* werden
- Pakete laden automatisch benötigte Dateien, Kodierungen etc.

Latin Modern

- Paket lmodern lädt Latin Modern-Schriften
- Verbesserung der ursprünglichen T_EX-Schriften (computer modern/cm-super)
- Sollte man verwenden, wenn mit pdfL^AT_EX gearbeitet werden muss

Schrifttypen

- Bitmap-Fonts: Kleine Bilder der Buchstaben, Form pixelweise angegeben
⇒ nicht beliebig vergrößerbar!

Schrifttypen

- Bitmap-Fonts: Kleine Bilder der Buchstaben, Form pixelweise angegeben
⇒ nicht beliebig vergrößerbar!
- Outline-Schriften:

Schrifttypen

- Bitmap-Fonts: Kleine Bilder der Buchstaben, Form pixelweise angegeben
⇒ nicht beliebig vergrößerbar!
- Outline-Schriften:
- Postscript Type1 (pfb/pfm): Beschreibt Buchstabenformen durch Bézierkurven

Schrifttypen

- Bitmap-Fonts: Kleine Bilder der Buchstaben, Form pixelweise angegeben
⇒ nicht beliebig vergrößerbar!
- Outline-Schriften:
- Postscript Type1 (pfb/pfm): Beschreibt Buchstabenformen durch Bézierkurven
- TrueType (ttf,ttc): dominierende Schrifttechnologie am Computer (quadratische B-Splines)

Schrifttypen

- Bitmap-Fonts: Kleine Bilder der Buchstaben, Form pixelweise angegeben
⇒ nicht beliebig vergrößerbar!
- Outline-Schriften:
- Postscript Type1 (pfb/pfm): Beschreibt Buchstabenformen durch Bézierkurven
- TrueType (ttf,ttc): dominierende Schrifttechnologie am Computer (quadratische B-Splines)
- smartfonts: Weiterentwicklungen von TrueType

Schrifttypen

- Bitmap-Fonts: Kleine Bilder der Buchstaben, Form pixelweise angegeben
⇒ nicht beliebig vergrößerbar!
- Outline-Schriften:
- Postscript Type1 (pfb/pfm): Beschreibt Buchstabenformen durch Bézierkurven
- TrueType (ttf,ttc): dominierende Schrifttechnologie am Computer (quadratische B-Splines)
- smartfonts: Weiterentwicklungen von TrueType
- OpenType (otf): Weiterentwicklung von Type1 und TrueType mit sehr vielen möglichen Features

Schrifttypen

- Bitmap-Fonts: Kleine Bilder der Buchstaben, Form pixelweise angegeben
⇒ nicht beliebig vergrößerbar!
- Outline-Schriften:
- Postscript Type1 (pfb/pfm): Beschreibt Buchstabenformen durch Bézierkurven
- TrueType (ttf,ttc): dominierende Schrifttechnologie am Computer (quadratische B-Splines)
- smartfonts: Weiterentwicklungen von TrueType
- OpenType (otf): Weiterentwicklung von Type1 und TrueType mit sehr vielen möglichen Features
- Apple Advanced Typography (aat): Entwicklung von Apple für höchste typographische Ansprüche

OpenType

- Verfügen über eine große Zahl von schriftspezifischen Features:
- (echte) Kapitälchen, Mediävalziffern, besondere Ligaturen, Schmuckschriften, verschiedene Varianten von Buchstaben, optische Skalierung, ...

OpenType

- Verfügen über eine große Zahl von schriftspezifischen Features:
- (echte) Kapitälchen, Mediävalziffern, besondere Ligaturen, Schmuckschriften, verschiedene Varianten von Buchstaben, optische Skalierung, ...

pdfT_EX kann direkt nicht mit OpenType-Schriften umgehen

- Metriken können nicht extrahiert werden
 - tfm-Datei nötig
 - spezielle Features nicht verwendbar
- ⇒ Pakete können Schriften verfügbar machen

Installation von Systemschriften

Windows

Schriftdatei (name.otf/ttf) in Ordner `C:\WINDOWS\Fonts` (o. ä.) schieben, wird automatisch installiert.

Linux

Datei in den Ordner `/usr/share/fonts` speichern und (als root) `fc-cache -vf` ausführen.

Apple

Dateien in den Ordner `/Library/Fonts` oder `/Users/Nutzername/Library/Fonts` speichern.

Die Tastatur

Warum sind die Tasten auf der Tastatur so seltsam angeordnet?

Die Tastatur

Warum sind die Tasten auf der Tastatur so seltsam angeordnet?

- Verteilung für Zehn-Finger-Schreibsystem
- Häufigste Buchstaben sollen auf kräftigen, schnellen Fingern in der Grundstellung sein
- Finger sollten sich möglichst wenig bewegen, um schnelles angenehmes Schreiben zu ermöglichen

Die Tastatur

Warum sind die Tasten auf der Tastatur so seltsam angeordnet?

- Verteilung für Zehn-Finger-Schreibsystem
 - Häufigste Buchstaben sollen auf kräftigen, schnellen Fingern in der Grundstellung sein
 - Finger sollten sich möglichst wenig bewegen, um schnelles angenehmes Schreiben zu ermöglichen
- ⇒ das ist bei QWERTZ nicht der Fall!
- ⇒ **QWERTZ ist historisch bedingt und hat (fast) nichts mit Ergonomie zu tun**

Neo

- Ein ergonomisches Tastaturlayout
- Häufigste Buchstaben auf kräftigen Fingern
- Möglichst häufige Handwechsel
- Berücksichtigung häufiger Buchstabenkombinationen
- Informationen, Treiber und Lernhilfen unter <http://neo-layout.org>
- Tastaturbelegungen betreffen nur die Software und können ohne neue Hardware verwendet werden!

Zwischenbilanz

- utf8: Vielzahl von Zeichen kodierbar
- otf/aat: Vielzahl von Zeichen/zusätzlichen Features darstellbar
- gute Tastaturlayouts: Vielzahl von Zeichen eingebbar
- Programm, damit umzugehen?

LuaT_EX Unicode, OpenType

- LuaT_EX arbeitet nativ mit utf8
- ⇒ kann problemlos mit Unicode umgehen
- LuaT_EX kann alle modernen Schrifttechnologien verwenden!
 - *alle* Schriftfeatures können verwendet werden
falls nicht: „leicht“ anpass- und erweiterbar

LuaT_EX Unicode, OpenType

- LuaT_EX arbeitet nativ mit utf8
- ⇒ kann problemlos mit Unicode umgehen
- LuaT_EX kann alle modernen Schrifttechnologien verwenden!
 - *alle* Schriftfeatures können verwendet werden
falls nicht: „leicht“ anpass- und erweiterbar
 - LuaT_EX baut auf pdfT_EX auf und kann daher auch Mikrotypographie!

Schriften laden (**low level**, nur wenn unbedingt nötig, ansonsten mit **fontspec** arbeiten!)

```
\font\warnock="Warnock Pro/I/S=5:+smcp" at 12pt
```

fontspec

Schriften laden mit fontspec:

high level:

```
\fontspec{Arno Pro}
```

(schaltet im aktuellen Dokument direkt auf die angegebene Schrift)

fontspec

Schriften laden mit fontspec:

high level:

```
\fontspec{Arno Pro}
```

(schaltet im aktuellen Dokument direkt auf die angegebene Schrift)

document level:

```
\setmainfont[Ligatures=tex-text]{Arno Pro}  
\setsansfont[ExternalLocation]{Linux Biolinum}  
\setmonofont{DejaVu Sans Mono}
```

(setzt die angegebene Schrift als Brotschrift, Serifenlose bzw. Dicktengleiche für das ganze Dokument)

Ligatures=tex-text sorgt für T_EX-Ligaturen (-- für –)

fontspec

Falls mehrere Schriften im Dokument verwendet werden, zwischen denen man wechseln will: `\newfontfamily` definiert Schalter dafür:

```
\newfontfamily\dejavu{DejaVu Sans Mono}  
...  
\begin{document}  
...  
\dejavu Text in DejaVu Sans Mono ...
```

Übung Viele Befehle sind bekannt und können vervollständigt werden:

`\doc <tab> ⇒ \documentclass{}`

`bit <tab> ⇒ \begin{itemize} \end{itemize}`

- bei Mehrdeutigkeit: mehrmals <tab> drücken