Workshop

Es geht auch ohne Formeln

Der Einsatz von TEX in den Digital Humanities am Beispiel kritischer Editionen

Martin Sievers

20. Januar 2015

1 Einleitung

Die Diskussion rund um digitale Editionen als Ergänzung oder sogar Ersatz für die klassische Buchausgabe ist in vollem Gange. Grund dafür ist auch der Siegeszug der plattform- und implementationsunabhängigen Metasprache XML in den „Digital Humanities“. Insbesondere der TEI-Standard[1] hat dafür gesorgt, dass Informationen aller Art in einem Sammelformat vorliegen, das von vielen neu entwickelten Werkzeugen als Ausgabe- und Austauschformat verwendet wird.

Der weitere Bearbeitungsprozess bis hin zur Fertigstellung einer Edition fußt daher heutzutage stark auf XML. Gleichwohl bleibt das Buch als notwendiges Ergebnis eines Editionsprojekts weiterhin die Regel. Somit stehen viele Wissenschaftler zu Beginn eines solchen Projekts vor dem Problem, einen Workflow auf XML-Basis zu definieren, der am Ende möglichst komfortabel – mit oder ohne Zutun eines Verlags – auch einen hochwertigen Buchdruck erlaubt.

Projekte wie XML-Print[2] oder Apache FOP[3] setzen hier an und wollen das Textsatzproblem innerhalb der „X-Technologien“[4] lösen. Es ist in den vergangenen Jahren jedoch deutlich geworden, dass der wissenschaftliche Textsatz von diesen Werkzeugen (noch) nicht in all seinen Facetten erfasst werden kann. Daher greifen aktuelle Editionsprojekte in der Regel auf etablierte

1 Vgl. Burnard und Bauman 2007.
2 Siehe dazu *XML-Print: typesetting arbitrary XML documents in high quality* 2014.
3 Siehe dazu *Apache(tm) FOP – a print formatter driven by XSL formatting objects (XSL-FO) and an output independent formatter* 2014.
4 Unter den X-Technologien versteht man die W3C-Standards XML, XSL und XPath sowie je nach Kontext weitere im XML-Umfeld entstandene Sprachen und Formate wie XQuery oder XLink.

Werkzeuge wie TUSTEP oder andere nicht notwendigerweise XML-basierte Ansätze zurück, indem der XML-Eingabetext mittels XSLT in die entsprechenden Zwischenformate transformiert wird.

Genau hier möchte der Workshop ansetzen und die Möglichkeiten des Textsatzsystems TEX[5] im Bezug auf die Erstellung einer historisch-kritischen Ausgabe (kritische Edition) vorstellen. Leider wird für dieser Weg aus Unwissenheit bzw. wegen falscher oder schlicht veralteter Informationen bzgl. des Funktionsumfangs viel zu selten beschritten. Umso erfreulicher sind Forschungs- und Arbeitsumgebungen wie FuD[6] oder ediarum[7], die am Ende des editorischen XML-basierten Workflows eine mit TEX erzeugte PDF-Datei zur Kontrolle bzw. als Vorstufe des Druckergebnisses ausgeben.

2 Funktionsweise von LATEX

2.1 Allgemein

Das quelloffene Textsatzsystem und die zugehörige Sprache TEX wurden Ende der Siebziger Jahre vom amerikanischen Mathematikprofessor Donald E. Knuth für den Druck seiner eigener Bücher entwickelt. Das Problem des Textsatzes „Wie bringe ich unter Beachtung verschiedener Regeln möglichst schön Zeichen aller Art aufs Papier?“ wurde von ihm als mathematisches Optimierungsproblem definiert und mit neuartigen Algorithmen gelöst. Die *subjektive* Schönheit wurde dadurch von Knuth auf Basis typographischer Traditionen und Methoden *objektiviert*.

Die so entstandenen Algorithmen, z. B. derjenige für den Zeilenumbruch[8] waren bahnbrechend und sind bis heute „State of the Art“. Entsprechend werden sie auch in jüngerer Software wie Adobe InDesign oder Apache FOP nahezu unverändert verwendet. Für den Autor eines Texts bedeutet dies, dass er sich vollständig auf die inhaltliche bzw. strukturelle Gestaltung konzentrieren kann. Dies entspricht der Arbeit mit XML-Quelldaten, die in der Regel keinerlei typographische Anweisungen enthalten.

Somit lebt die klassische Trennung zwischen Autor und Setzer wieder auf, die durch Text*verarbeitungs*programme stückweise aufgeweicht worden ist – mit negativen Folgen für die Druckqualität. In heutigen digitalen Arbeitsumgebungen entspricht der Setzer einem Satzprogramm, das eine Druckvorlage auf die Quelldokumente eines Autors anwendet.

5 TEX wird seit langem schon nur noch selten direkt angewendet, sondern in der Regel über eine Makrosprache wie LATEX (siehe z. B. http://www.latex-project.org/) oder ConTEXt (siehe z. B. http://wiki.contextgarden.net/What_is_ConTeXt) angesprochen. Der Workshop konzentriert sich auf LATEX.

6 Siehe dazu *FuD – Eine virtuelle Forschungsumgebung für die Geistenwissenschaften* 2014.

7 Siehe dazu *ediarum – eine digitale Arbeitsumgebung für Editionsvorhaben* 2014.

8 Vgl. dazu Knuth und Plass 1981.

2.2 Anwendungsfall Kritische Edition

Eine kritische Edition stellt besondere Anforderungen an ein Textsatzwerkzeug. Daher eignet sich dieser Dokumenttyp besonders gut, um die Qualität von LaTeX zu demonstrieren. Plachta definiert für eine kritische Edition zehn elementare Bestandteile.[9] Diese lassen sich zu den folgenden drei Themenkomplexen zusammenfassen:

- Edierter Text
- Apparate
- Verzeichnisse

Zu all diesen Bereichen liefert LaTeX entweder direkt oder über Erweiterungen Möglichkeiten, hochwertige Ergebnisse zu erzielen. Sie stehen über das zentrale Paketarchiv CTAN[10] allen Nutzern kostenfrei zur Verfügung.

3 Inhalte des Workshops

Entlang der in Abschnitt 2.2 genannten Bestandteile einer kritischen Edition wird der Workshop den Teilnehmern die Gelegenheit geben, LaTeX als Satzprogramm kennenzulernen. Im Detail werden die folgenden Inhalte vermittelt:

Edierter Text: Neben den grundlegenden Algorithmen werden mikrotypographische Fragen thematisiert. Dazu gehören neben der Nutzung typographisch korrekter Zeichen (z. B. bei Anführungszeichen, Gedankenstrich oder Ellipse) auch der automatische optische Randausgleich oder die Laufweitenänderung, die beide durch das Paket `microtype`[11] bereitgestellt werden.

Apparate: Die Anforderungen an den Satz von Apparaten gehen weit über diejenigen „normaler“ Fußnoten hinaus. Es wird das Paket `eledmac`[12] vorgestellt und neben verschiedenen Anpassungen auch Lösungen für Probleme wie z. B. überlappende Lemmata erarbeitet.

Verzeichnisse: Eine kritische Edition enthält neben einem Quellenverzeichnis verschiedene Register. Diese sollen im Idealfall direkt aus den Druckdaten dynamisch erzeugt werden. Dazu werden die Erweiterungen `biblatex`[13] sowie `splitindex`[14] vorgestellt, die genau dies bewerkstelligen.

9 Plachta 2006, S. 14 f.
10 Das Comprehensive TeX Archive Network stellt über zwei zentrale Server und mehr als hundert Spiegelserver (Mirrors) weltweit unter http://www.ctan.org insgesamt über 4500 Erweiterungen bereit.
11 Vgl. dazu http://www.ctan.org/pkg/microtype.
12 Vgl. dazu http://www.ctan.org/pkg/eledmac.
13 Vgl. dazu http://www.ctan.org/pkg/biblatex.
14 Vgl. dazu http://www.ctan.org/pkg/splitindex.

4 Teilnehmerkreis / Technische Ausstattung

Der Workshop richtet sich an alle interessierten Wissenschaftler, die Wert auf einen hochwertigen Druck ihrer Edition legen. Auch Entscheidungsträger, die ein Textsatzsystem für ihr Editionsprojekt suchen, sind ausdrücklich angesprochen. Für die praktischen Beispiele werden grundlegende Kenntnisse der Textsatzsprache LaTeX vorausgesetzt.

Da es sich um eine „Hands-on"-Sitzung handelt, sollte die Teilnehmerzahl 15 nicht übersteigen. Die Teilnehmer benötigen einen Laptop mit einer (möglichst aktuellen) TeX-Distribution (MiKTeX oder TeX Live). Für die Präsentation selbst wird ein Beamer benötigt.

Literatur

LaTeX – A document preparation system (2014). http://www.latex-project.org/ (besucht am 10. 11. 2014).

Apache(tm) FOP – a print formatter driven by XSL formatting objects (XSL-FO) and an output independent formatter (2014). http://xmlgraphics.apache.org/fop/ (besucht am 10. 11. 2014).

Burnard, Lou und Syd Bauman (2007). *TEI P5. Guidelines for Electronic Text Encoding and Interchange.* Text Encoding Initiative.

ediarum – eine digitale Arbeitsumgebung für Editionsvorhaben (2014). http://www.bbaw.de/telota/software/ediarum (besucht am 10. 11. 2014).

FuD – Eine virtuelle Forschungsumgebung für die Geistenwissenschaften (2014). http://fud.uni-trier.de/de/ (besucht am 10. 11. 2014).

Knuth, Donald E. und Michael F. Plass (1981). "Breaking paragraphs into lines". In: *Journal Software: Practice and Experience* 11.11, S. 1119–1184. DOI: 10.1002/spe.4380111102.

Plachta, Bodo (2006). *Editionswissenschaft. Eine Einführung in Methode und Praxis der Edition neuerer Texte.* 2. Aufl. Reclams Universal-Bibliothek 17603. Stuttgart: Philipp Reclam jun.

What is ConTeXt (2014). http://wiki.contextgarden.net/What_is_ConTeXt (besucht am 10. 11. 2014).

XML-Print: typesetting arbitrary XML documents in high quality (2014). https://sourceforge.net/projects/xml-print/files/.