

Anhang C1: RTFtoHTML-Konvertierer 3.x¹

Dr. Frank Krüger

Fachbereich 23, Johannes Gutenberg-Universität Mainz

e-mail: krueger@acm.org

- [Vorbemerkungen](#)

1. Word-Dokumentvorlagen

- [Zeichenformate](#) vs. [Absatzformate](#)
- [Symbolleisten](#) und [Makros](#)

2. [Optionen des Konvertierprogrammes](#)

3. [Tips und Tricks für die Konvertierung](#)

4. [Fazit: Für welche Zwecke eignet sich das Konvertierprogramm](#)

5. [Webseiten Schritt-für-Schritt mit RTFtoHTML-Konverter](#)

Konvertierung von WORD/RTF nach HTML

Vorbemerkungen

Eine der Möglichkeiten für die komfortable Erstellung von Web-Seiten ist die Bearbeitung der Seiten mit der gewohnten Textverarbeitung (am FASK bevorzugt WORD 6.0 FÜR WINDOWS oder MACINTOSH) und die anschließende Konvertierung in eine HTML-Datei. Auf diese Weise kann die Erstellung von (gedruckten) Texten und ihre Formatierung für das WWW in einem Arbeitsgang erledigt werden, da zusätzlich zu der elektronischen Version eine ansprechend formatierte "Papierausgabe" erstellt werden kann. So stehen am Fachbereich für die Arbeit mit dem Konvertierungsprogramm und WORD FÜR WINDOWS/MACINTOSH 6.0 folgende zwei Dokumentvorlagen zur Verfügung:

- FASK_WWW.DOT enthält alle notwendigen Formatvorlagen, einige Makros und die wichtigsten Formatvorlagen als Symbole in zwei speziellen Symbolleisten.
- FASK_DRUC.DOT definiert einige der Formatvorlagen so um, daß sich der Dokumentinhalt in einer ansprechenden Form ausdrucken läßt (vgl. dieses Skript) und erlaubt Text bzw. Elemente mit aufzunehmen, die nicht in der elektronischen Version erscheinen sollen (vgl. Bemerkungen zu Graphiken unter [Abschnitt 4, Tips und Tricks](#)).

1. Word-Dokumentvorlage WWW_FASK.DOT

A. Zeichenformate

(# = Zahl, Optional = Zusatzspezifikation für Netscape, siehe auch nachfolgende !)

Hinweis zur ersten Spalte: Hinter der tatsächlichen Formatierung steht in Klammern die Formatvorlage. Die Symbole sind im [Abschnitt C](#) zusammengefaßt.

Eine Formatierung als ..	erzeugt eine Auszeichnung als ...	HTML-Markup und Parameter	im Graphikbrowser dargestellt als
a) Hypertext/Verweise			
doppelt unterstrichen/ verborgen (link)	Verweiseintrag (Verknüpfung)		unsichtbar
doppelt unterstrichen/ nicht verborgen (linkmark)	Textanker (Aktiver Text) 	Farbig/unterstrichen
roter Text (neu mit 3.0, URL)	Verweis mit genau diesem Text	 URL 	Kombination der beiden vorherigen
verborgen/ Konturschrift bzw. Hochstellung (linkname)	Verknüpfungsziel (Name)		unsichtbar
(sog. Inline-Grafik)*	eingebette Grafik		entsprechendes Bild eingebett im Text
(andere Grafiken)*	Verweis auf Grafik	, wobei Name=Datei, #= lfd. Nr., ext=Typ	Bild erscheint als eigenes Dokument, wobei GIF für Zeich- nungen und JPEG für Fotos geeignet ist.

***Hinweis:** Es wird empfohlen, die Grafiken aus technischen Gründen mit der Formatvorlage PRINT formatieren, die bewirkt, daß diese Elemente (und etwaig begleitender Text) bei der Konvertierung wegfallen. So können die Grafiken manuell angelegt, bearbeitet und [aussagekräftig benannt](#) werden. Für das HTML-Dokument wird ein entsprechender Verweis auf die Grafik als HTML-Code eingebettet und mit der Formatvorlage HTML oder CODE formatiert.

¹ : Eine elektronische Version dieses Dokumentes befindet sich auf dem WWW-Server des Fachbereiches unter der Adresse <http://www.fask.uni-mainz.de/cafl/kurse/komm/komm-c1.html> (Kleinschreibung beachten!).

Eine Formatierung als ..	erzeugt eine Auszeichnung als ...	HTML-Markup und Parameter	im Graphikbrowser dargestellt als
b) logische Auszeichnung			
Courier oder Monaco	Schreibmaschinenschrift	<tt> bzw.	Courier 12
Konturschrift/schattiert (em)	Betonung		Fett
fett/schattiert (strong)	starke Betonung		fett, größer
kursiv/ unterstrichen (cite)	Zitat	<cite>	
fett/einfach doppelt etc. unterstrichen	Def.begriff versch. Ebenen im Text link zu Glossar	<dfn>, in Glossar als DT	
Courier/Kursiv (samp)	Beispiel	<samp>	Courier
Courier/Unterstr (kbd)	Tastatureingabe	<kbd>	Courier
Courier/Fett (var)	Variable	<var>	Courier
c) physikalische, direkte Auszeichnung			
verborgen/schattiert(Code)	unveränderter HTML-Code	(literal)	entsprechender HTML-Markup
fett	fett		fett
kursiv/n.unterstrichen	Kursiv	<i>	kursiv
hochgestellt	Hochstellung	<sup>	Hochgestellt

B. Absatzformatvorlagen

a) Allgemein			
Überschrift # (Ü1, Ü2, Ü3)	Überschriftsebenen	<h#>	Größer, fett
Address	Autorenangaben mit Institution	<address>	Courier (üblicherweise am Textende)
Pre (pre)	In Schreibmaschinenschrift (z.B. Tabellen)	<pre>	Courier

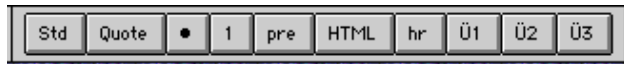
Eine Formatierung als ..	erzeugt eine Auszeichnung als ...	HTML-Markup und Parameter	im Graphikbrowser dargestellt als
hr (hr)	Waagerechte Linie (horizontal ruler)	<hr>	waagerechte Linie (vgl. Netscape-Erweiterungen)
blockquote (quote)	längeres Zitat	<blockquote>	eingerrückt, Courier
html (html)	nicht umgesetzte Zeilen	(_literal)	entsprechender Markup
b) Schachtelbare Formatvorlagen			
bullet list, # (#="" u. 1-2; Symbol •)	Unnummerierte Liste der #.ten Ebene Aufzählungszeichen plus Tab wird in Vorlage entfernt!	 TYPE = disc circle square	Eingerückt, mit versch. Zeichen (• etc.)
numbered list, # (#="" u. 1-2; Symbol 1)	Numerierte Liste der #.ten Ebene Nummer plus Tab in Vorlage wird entfernt!	 TYPE = a A i i1, START = #	Eingerückt, mit Numerierung
Term, Term1, Term2	Termini zur Beschreibung	<dl><dt>	Aufzählung
Def, Def1, Def2	Beschreibungen	<dd>	Aufzählung
glossary, glossary 1-2 (Glossary)	Glossare (Teile getrennt durch Tabulator)	<dl>	eingerrückt
menu, menu1, menu2	Menüs (ein Eintrag pro Zeile)	<menu>	untereinander
dir, dir1, dir2	Verzeichnisse (sehr kurze Einträge)	<dir>	untereinander

C. Sonstiges

a) Zusätzliche Umsetzungen			
Tabellen mit oder ohne Rand	Umformatierung von Tabellarischem Text	<Table> plus entsprechende Untermarkierung	Tabellen mit bzw. ohne Rand
Fußnotentext			verknüpfte Datei
weitere (optionale) verknüpfte Dateien:			
<ul style="list-style-type: none"> Inhaltsverzeichnis und Index: Dazu jeweilige Optionen im Konvertierprogramm (Prefs.) aktivieren! 			

b) Symbolleisten

Für die Dokumentvorlagen WWW_FASK und WWW_DRUC sind folgende beiden Symbolleisten mit Symbolen für die am häufigsten verwendeten Formatvorlagen definiert:



- WWW-Absatzformatierung, mit Symbolen für folgende Absatzformatvorlagen:

- STD = Standard, Quote= Blockquote
- 1 = numbered list, • = bullet list
- Ü1 bis Ü3 = Überschrift 1 bis Überschrift 3



- WWW-Zeichenformatierung, wobei nur folgende Symbole nicht identisch zu den Zeichenformatvorlagen (dünne Schrift) benannt sind:

- = link, also Verweisadresse (verborgen/doppelt unterstrichen)
- = linktext, also der blau hervorgehobene Verweistext
- = linkmark, also die Zieladresse innerhalb eines Dokumentes
- URL = Markierung eines Textteiles als Dokumentadresse (durch rote Textfarbe), erspart eine mehrfache Eingabe als LINK und LINKTEXT

Alle anderen, nicht in den Symbolleisten enthaltenen Formatvorlagen können über das Pull-Down-Menü Formatvorlagen in der Formatierungs-Symbolleiste abgerufen werden

c) Makros und Textbausteine

Neben Formatvorlagen und Symbolleisten sind in den beiden Dokumentvorlagen auch Makros definiert, die bestimmte Befehlsfolgen zusammenfassen:

- Das wichtigste Makro ist die **Umlautkonvertierung**, die immer bei der Verwendung von unter Windows oder DOS erzeugten RTF-Dateien auf dem Macintosh ausgeführt werden sollte. Mac und Windows definieren nämlich Umlaute (und andere nationale Sonderzeichen) unterschiedlich, auch wenn sie in Word selbst richtig dargestellt werden. Das Konvertierungsprogramm jedenfalls wandelt ohne dieses Makro die Umlaute falsch um. Alternativ dazu kann die RTF-Datei mit der Konvertierungsoption "NUR TEXT" in Word geladen werden und dort am Datei-anfang der Eintrag von "ANSI" (Windows-Zeichensatz) in "MAC" geändert werden.

- Ein anderes Makro, **Absatzkonvertierung**, ist für die Anwendung bei der dritten vordefinierten Formatvorlage, FASK_HTML.dot gedacht und setzt alle Absatzmarken in HTML-konforme <p> Markierungen um.
- Alle anderen Makros sind derzeit ohne Bedeutung für die Praxis.
- **Textbausteine** können verwendet werden, um längere Textpassagen zu speichern und nach Vorgabe eines Kürzels auf Tastendruck (Mac: Apfel-Alt-V, Windows: F3) einzufügen. Beispiele, die aber teilweise nicht in der Dokumentvorlage WWW_FASK.DOT enthalten sind:

www dem WWW-Team

kurs [Frank Krüger](#) im Rahmen des [Kommunikationskurses](#) von [Frank Krüger](#) im Sommersemester 1996

image

Hinweis: alle Beispiele werden mit der Formatvorlage HTML eingefügt!

2. Optionen für die Konvertierung

Die meisten Optionen für das Konvertierprogramm können vor dem Ablauf der Konvertierung in den verschiedenen Konfigurationsdateien eingestellt werden. Auf diese soll hier nicht näher eingegangen werden, da diese - für Spezialisten - ausreichend in der Originaldokument beschrieben sind.

Für den einzelnen Ablauf der Konvertierung können folgende Einstellungen verändert werden (Macintosh: Menü EDIT PREFERENCES, andere Versionen: Optionen der Befehlszeile beim Programmaufruf):

- -h level bzw. Eingabefeld "Split Level": Überschriftsebene, bei der das Dokument in einzelne Dateien getrennt wird (0=eine Datei)
- -c bzw. Ankreuzfeld "Table of Contents": Erzeugen einer Datei mit dem Inhaltsverzeichnis (generell hier nicht notwendig, da Inhalt eingebettet in Dateien)
- -x bzw. Ankreuzfeld "Index": Erzeugen eines Index, falls Einträge vorhanden
- Um - wie empfohlen (*) - die Graphiken manuell verweisen zu können sollten die Optionen "SKIP GRAPHICS" bzw. -g eingeschaltet sein!
- Von den anderen Optionen ist vor allem die Generierung einer Navigationszeile interessant, die entsprechend einer in dem Feld "Panel" (Option -N Datei) definierten Datei erzeugt wird, wenn eine Aufspaltung der Datei mit der Option -h angegeben wurde. Näheres dazu findet sich in der [Dokumentation zum Konvertierungsprogramm](#).

3. Tips und Tricks für die Konvertierung

Die folgende Liste ist eine (ungeordnete) und bei weitem nicht vollständige Aufzählung von bekannten Problemen bei der Erstellung von Seiten und speziell mit diesem Konvertierprogramm:

- Einige Fehler können durch das direkte Aufeinanderfolgen verschiedener Formatvorlagen entstehen, etwa BULLET LIST und HEADER #. Am besten also vor Überschriften immer eine Leerzeile im Format STANDARD einfügen.
- Tabellen lassen sich auch mit Microsoft Excel erstellen und mit dem dafür vorhandenen Internet Assitant in die fertig konvertierte HTML-Datei einfügen.
- Wie erwähnt, kann bei dem Konvertierprogramm eingestellt werden, daß ein Inhaltsverzeichnis erstellt und die enthaltenen Grafiken herausgezogen werden. Die Grafiken werden dann entsprechend der HTML-Dateien benannt und durchnumeriert (bei NAME.html entsprechend name1.pict, name2.pict ... bzw. name1.bmp etc.) in eigenen Dateien geschrieben. Da die Grafiken von dem Konverter in dem Macintosh Grafik-Format PICT abgespeichert werden, daß die meisten Browser (wie auch Netscape) nicht lesen können, müssen sie mit einem Grafikprogramm (z.B. Grafikkonverter) in das GIF oder JPEG-Format konvertiert werden.
- Dennoch sollten Bilder immer mit der Formatvorlage PRINT formatiert werden und parallel dazu ein Verweis `` für das entsprechende Bild angegeben werden. Auf diese Weise können zentrale Verzeichnisse mit Bildern angegeben und so die gleichen Bilder mehrfach genutzt werden. Handelt es sich um größere Graphiken, so sollte aus Gründen der Netzwerkökonomie - lange Übertragungszeiten für Bilder - in den Text lediglich ein **kleines Symbol** eingefügt werden, mit dem auf das eigentliche Bild verwiesen wird.
Syntax (Dateipfade und -namen in Kleinschreibung):
` `
Auf diese Weise kann auch die neue HTML-Erweiterung USEMAP für **benutzerseitig definierte Karten** verwendet werden.
- Fertige HTML-Dateien, die z.B. von anderen Servern heruntergeladen und gespeichert wurden (bei dem Befehl SAVE AS die Option SOURCE nicht vergessen!) können auch mit Word bearbeitet werden, z.B. um die Kopf- und Fußteile aus der Dokumentvorlage hinzuzufügen. Dabei müssen Zeilenumbrüche, Umlaute und andere Sonderzeichen aber von Hand kodiert werden (siehe Liste). Über eine spezielle Dokumentvorlage [WWW_HTML.dot](#) stehen dafür auch Makros zur Verfügung.
- Besonders bei Aufzählungen, aber auch Formatvorlagen allgemein ist darauf zu achten, daß nur die Absätze bzw. Zeichen mit der Formatvorlage formatiert werden, die dieses Format auch bekommen sollen. Sonst entstehen z.B. leere Zeilen mit Punkten, wie diese hier:
- Grundsätzlich können zwar Dokumente für die Konvertierung auch mit allen anderen Textverarbeitungssystemen erstellt und bearbeitet werden, die das Rich Text Format lesen und schreiben können (also z.B. frühere Word-Versionen, wie Word 6.0 für DOS, Word 2.0 für Windows, Word 5.1. für Macintosh oder andere Systeme wie

WordPerfect (ab 5.1), AmiPro, ClarisWorks etc.), dabei geht dann aber u.a. der Komfort der individuellen Symbolleisten und Zeichenformatvorlagen verloren bzw. müßte in den anderen Programmen mit den dortigen Möglichkeiten nachgebildet werden.

4. Fazit: Für welche Zwecke eignet sich das Konvertierprogramm am besten?

Das Konvertierungsprogramm RTFtoHTML bietet gegenüber den Editoren grundsätzlich den Vorteil, bestehende und ggf. veränderliche "Druck"-Dokumente, die also eigentlich für eine Papierform bestimmt sind, gleichzeitig in elektronischer Form für das WWW umwandeln zu können. Durch die parallele Dokumentvorlage *WWW_Druc* (gleiche Namen der Formatvorlagen, aber angepasste Formatierung) lassen sich dabei beide Versionen in einer Datei eingeben und einheitlich pflegen, aber durch geänderte Zuweisung der Dokumentvorlage in der jeweils gewünschten Form ausgeben.

Ferner lassen sich mit dem Konvertierprogramm Inhaltsverzeichnis und Index (falls eingestellt bzw. wenn entsprechende Einträge vorgegeben sind) automatisch erstellen und verweisen. Auch Fußnoten werden von dem Konvertierungsprogramm erkannt und als Verweise umgesetzt.

In der Version 3.0 hat das Konvertierungsprogramm zusätzlich den Vorteil, längere Dokumente durch entsprechende Einstellungen in dem "Preferences"-Menü (Mac) bzw. durch Angabe von Parametern (DOS/Unix) **automatisch** in kleinere, webgerechte Hypertextdokumente aufsplitten zu können und untereinander durch eine graphische Symbolleiste zu verknüpfen.

Ein Nachteil der "Konvertierungsmethode" ist allerdings der gegenüber **Editoren** erhöhte Aufwand, da die Dokumente nicht immer direkt im WYSIWYG-Modus (What you see is what you get) bearbeitet werden können, sondern das wirkliche Ergebnis immer erst nach der Konvertierung sichtbar ist. Entsprechend müssen Dateistruktur und -inhalte vorher gut überlegt sein (ggf. an einer kleinen Testdatei ausprobieren...). Allerdings kann das Aussehen in Word weitgehend "simuliert" werden, indem die Formatvorlagen an das Aussehen in Netscape als graphischem Browser angepaßt werden.

Auf der anderen Seite ist damit dieses System für den didaktischen Einsatz im Unterricht besonders geeignet, da es so zu einer systematischen Beschäftigung mit HTML und dem Design von WWW-Seiten anregt.

Und auch die Erweiterbarkeit für zukünftige Markierungen und Attribute in HTML - das ja schneller fortentwickelt wird, als Unterlagen wie diese hier geschrieben werden können - ist grenzenlos, da alle Elemente im Zweifelsfall als "reine" HTML-Markierungen (sog. **literal text**) eingegeben und formatiert werden können.

5. Webseiten Schritt-für-Schritt mit dem RTFtoHTML-Konverter Version 3.0 (Mac, Windows, Unix)

1. In WORD FÜR WINDOWS oder MACINTOSH 6.0 das Menü DATEI/NEU öffnen und eine der Dokumentvorlagen *WWW_FAS(K).dot* bzw. *WWW_Druc.dot* auswählen (Zum Unterschied vgl. [Vorbemerkungen](#); zu anderen Programmen [Abschnitt 3](#)).
2. Der Text kann nun mit Hilfe der Symbolleisten und/oder der Liste der "Formatvorlagen" bearbeitet werden. Neben einer Anpassung an die Konventionen der Web-Seiten (z.B. Aufzählungen und Listen, Tabellen) gehört dazu v.a. die Ergänzung um Verweise verschiedener "Reichweite" und Typen.
3. Das Dokument im sog. **Rich Text Format** (in der Liste unter DATEI/SPEICHERN UNTER/DATEIFORMAT auswählen) abspeichern.
Die enthaltenen Grafiken müssen in eines der Internet-Grafikformate GIF oder JPEG konvertiert werden und können bei dieser Gelegenheit auch nachbearbeitet werden (z.B. transparent, um sich dem Hintergrund anzupassen oder ergänzt um die Bildbeschriftung).
4. Die RTF-Datei wird mit dem Programm RTFtoHTML auf dem Macintosh (im Alias-Ordner Kommunikation), unter DOS/Windows (*Batch-Datei* RTFKONV auf *NFASK2*) oder UNIX (auf Anfrage) in HTML umgewandelt. Fußnoten werden in eigene Dateien geschrieben bzw. an das Ende der Datei gesetzt (vgl. Tips in [Abschnitt 2](#)).
5. Die in den vorherigen Schritten erzeugten HTML--Dateien (sowie etwaige Grafik-, Ton- und Videodateien) werden mit einem FTP-Programm (z.B. Fetch auf dem Macintosh, FTP unter Windows) auf den Web-Site in das entsprechende Unterverzeichnis übertragen.
 - Hinweis:
Dieser Schritt kann zur Zeit nur mit besonderen Zugriffsrechten auf den UNIX-Rechnern ausgeführt werden!!. Diese werden von Manfred Orschel in seiner Sprechstunde vergeben. Sollte dies momentan nicht nötig sein (z.B. da es sich um ein begrenztes "Projekt" handelt), kann die Datei auf dem Server in ein entsprechendes Verzeichnis abgelegt oder [per e-mail an einen Mitarbeiter des WWW-Team](#) geschickt werden.