

Download - und was dann? Dateiformate erkennen und handhaben

Autor:

Ulrich Babiak

Diplom-Dokumentar, Freier MA am FB Bibliotheks- und Informationswesen der FH Köln

Wittgensteinstr. 20, 50931 Köln

E-Mail: v9100055@athena.rrz.uni-koeln.de

©Ulrich Babiak 1995

Abstract

Die Verbreitung textbasierter Informationen auf elektronischem Wege erfordert auf Benutzerseite Kenntnisse über die Erkennung und Handhabung der verwendeten Dateiformate. Der Beitrag bespricht die wichtigsten Dateiformate, die derzeit verwendet werden, erläutert ihre Erkennung und stellt Hilfsprogramme für ihre Handhabung am PC vor.

The use of electronically published information requires a basic knowledge of the applied electronic file formats and their handling. The article covers the most commonly used electronic file formats for the distribution of character based information, explaining both their identification and the use of necessary software tools for personal computers.

Inhalt:

Download - und was dann ?	2
Dateiformate und Betriebssysteme.....	2
Erkennung von Dateiformaten	3
Viewer und Entpacker.....	3
Sonderformate für die ausschließliche Benutzung am PC	3
Komprimierte Dateien.....	4
Für den Mailversand codierte Dateien	4
ASCII-Dateien ("Textdateien").....	5
Textverarbeitungs-Dateien.....	5
TeX/LaTeX/DVI.....	5
Postscript	6
Der kommende Standard: Adobe Acrobat (PDF-Dateien)	7
Integration von Viewern für die Nutzung des World Wide Web.....	8
Tabelle 1: Dateiformate, Endungen, Programme	9
Tabelle 2: Viewer und Entpacker.....	10

Download - und was dann ?

Dateiformate identifizieren und handhaben

Am Ende einer Recherche im Internet oder in anderen Netzen ist das eigentliche Informationsproblem, das der Recherche zugrunde lag, noch nicht unbedingt gelöst.

Dies hängt nicht nur mit dem inhaltlichen Erfolg der Recherche zusammen: als Resultat der Bemühungen liegen vielmehr häufig Dateien vor, die aufgrund von Verweisen oder Inhaltsbeschreibungen für relevant gehalten und deshalb im Verlauf der Recherche auf den eigenen Rechner heruntergeladen werden. Die Formate solcher Dateien bereiten anschließend oftmals Probleme. Auch beim Versand von Dateien ohne eine adäquate Anleitung, sei es auf Datenträgern oder via E-Mail, stellen sich ähnliche Probleme.

Diese Übersicht behandelt die gängigsten Dateiformate, die derzeit zur Verbreitung textbasierter Informationen verwendet werden und erläutert ihre Erkennung und weitere Handhabung mit PC-Software. Eine online-Version mit Verweisen auf weitere Informationsquellen und Hilfen zur Beschaffung der besprochenen Programme ist im WWW unter <http://www.fh-koeln.de/fbi/pers/babiak/abi.html> verfügbar.

Dateiformate und Betriebssysteme

Die Bearbeitung oder Benutzung einer Datei setzt voraus, daß bekannt ist, in welchem Format sie vorliegt. Nur so kann ein entsprechendes Programm zur weiteren Bearbeitung ausgewählt werden.

Üblicherweise dient der Dateiname als Träger dieser Information, denn den meisten Dateiformaten ist eine entsprechende Dateinamenerweiterung (Extension) zugeordnet. Sie wird an den Dateinamen angehängt, um das Format zu kennzeichnen, so zum Beispiel die Endung ".txt" für ASCII-Textdateien. Eine Auflistung der Extensionen gängiger Formate findet sich in Tabelle 1.

Die Standardextensionen werden schon bei der Erstellung der Dateien von den Anwendungsprogrammen vergeben. Die Anwender können sich aber durch Vergabe anderer Endungen darüber hinwegsetzen. Oft geschieht dies, weil unter MS-DOS nur Dateinamen von 8 + 3 Zeichen Länge zuläßt - Unix, OS/2, Macintosh und auch das neue Windows95 ermöglichen dagegen Dateinamen mit bis zu 255 Zeichen Länge.

Bei der Dateiübertragung von einem dieser anderen Betriebssysteme auf einen MS-DOS-Rechner, z.B. via ftp, werden die Dateinamen oft so verstümmelt, daß die Extensionen verlorengehen. Ärgerlich ist dies vor allem beim Transfer von Dateien mit mehreren Extensionen. So werden unter Unix beispielsweise komprimierte Postscript-Dateien mit den Endungen ".ps.Z" versehen, um sowohl das Postscript-Format als auch die Komprimierung zu kennzeichnen. Beim Transfer auf DOS-Rechner wird dies auf die - dann nicht mehr zutreffende - Endung .ps reduziert.

Die zunehmende Vernetzung unterschiedlicher Rechnerwelten führt zudem zu Schwierigkeiten, wenn Dateien genutzt werden sollen, die unter anderen Betriebssystemen erstellt wurden. Oft sind solche Dateien nur mit dem entsprechenden Ursprungsbetriebssystem lesbar oder benötigen zumindest besondere Software oder sogar Hardware (z.B. Postscript-Drucker). Allerdings gibt es inzwischen Konvertierungs- und Hilfsprogramme für fast jede denkbare Kombination von Ursprungs- und Zielsystem¹.

¹Solche Programme finden sich auf allen gutsortierten ftp-Servern, meist in Kategorien wie "fileutils" oder "tools", z.B. ftp.uni-

Erkennung von Dateiformaten

Wie kann nun das Format von Dateien ohne eine kennzeichnende Dateinamenerweiterung erkannt werden?

Zunächst sollte versucht werden, eine unbekannte Datei mit einer Textverarbeitung zu öffnen. Moderne Textverarbeitungen besitzen Import- und Konvertierungsfunktionen, die man sich auf diese Weise zunutze machen kann. So wird bereits eine Vielzahl der gängigen Textformate abgedeckt.

Sollten bei diesem Versuch nur wirre Zeichen erscheinen, ist dennoch eine nähere Bestimmung möglich. Bei vielen Dateiformaten sind Informationen über den Dateityp auch in der Datei selbst enthalten, meist in einem sogenannten "Header"-Teil, also einem Bereich zu Beginn der Datei vor dem eigentlichen Inhalt. Neben mehr oder weniger kryptischen Zeichenkombinationen lassen sich oft auch lesbare Partien finden, die Auskunft über das Dateiformat geben. So enthalten beispielsweise Postscript-Dateien in der ersten Zeile den Hinweis "%!PS-ADOBE...", PDF-Dateien beginnen mit der Zeichenkombination "%PDF", und komprimierte Dateien listen häufig die Namen der in ihnen enthaltenen Dateien im Header auf (die genannten Formate werden im Anschluß besprochen). Auch bei anderen Formaten lassen sich in den lesbaren Header-Teilen Hinweise finden.

Einfacher geht es mit Hilfsprogrammen, die genau diese Detektivarbeit der Dateianalyse übernehmen. Besonders leistungsfähig ist das in Tabelle 2 aufgeführte Programm FILE, das zuverlässig fast alle hier besprochenen Dateiformate erkennt - sowie eine Vielzahl weiterer Formate wie z.B. Bild-, Audio- oder Animationsformate. Das Programm analysiert unbekannte Dateien und gibt im Handumdrehen deren Dateityp an.

Viewer und Entpacker

Ist das Dateiformat erst erkannt, folgt die Auswahl der geeigneten Werkzeuge für die Bearbeitung. Komprimierte Dateien müssen zunächst mit einem passenden "Entpacker" dekomprimiert werden. Hierfür gibt es eine Reihe frei verfügbarer Programme für alle im Anschluß besprochenen Formate (vgl. Tabelle 2.)

Für die "passive" Nutzung von Dateien gibt es spezielle Programme, sogenannte "Viewer" (Betrachter). Sie lassen eine weitere Bearbeitung zwar nicht mehr zu, sind aber zum einen weniger umfangreich als die zur Produktion der Dateien verwendeten Programme und vor allem meist kostenlos verfügbar. Solche Viewer werden teilweise von kommerziellen Anbietern zur Verfügung gestellt, um die Nutzung von Dateiformaten eigener Produkte zu fördern. Zum anderen Teil werden sie aber auch von engagierten Programmierern frei verfügbar gemacht. Auf diese Weise fanden beispielsweise viele Unix-Tools wie z.B. "tar" oder "uncompress" Eingang in die DOS-Welt und ermöglichen die Nutzung von ehemals Unix-spezifischen Dateien unter DOS.

Eine Auswahl wichtiger und zweckmäßiger Viewer und Entpacker soll nun im Zusammenhang mit den jeweiligen Dateiformaten besprochen werden; eine Übersicht findet sich in Tabelle 2.

Sonderformate für die ausschließliche Benutzung am PC

Viele Autoren wählen für die Verbreitung von Publikationen spezielle Formate, die ausschließlich für die Nutzung am Computer und nicht für eventuelle Ausdrucke gedacht sind. Vor allem Autorensysteme wie Toolbook, Norton Guide oder andere fallen in diese

Kategorie. Auf diese Weise können sehr ansprechende Dokumente gestaltet werden, die Grafiken, Töne, Filme und ausgeklügelte Benutzeroberflächen, Navigationshilfen oder Hypertextlinks enthalten, was vor allem für in sich geschlossene Präsentationen, Lernprogramme o.ä. genutzt wird. Solche Systeme enthalten stets Viewer-Software, die der Autor zusammen mit seinem Dokument mitliefern muß.

Ein Sonderfall in dieser Kategorie ist das Format der Windows-Hilfe-Dateien (Endung ".hlp"). Die Viewer-Software ist bereits im Lieferumfang von Windows enthalten - es handelt sich um das Programm winhelp.exe, das auch immer aufgerufen wird, wenn die Hilfe-Funktion eines Windows-Programms benutzt wird. Genauso wie Hilfedateien zu Windows-Programmen können auch beliebige andere Dokumente mit gleichen Eigenschaften erstellt werden: Stichwortliste, Inhaltsverzeichnis, Navigationsbuttons, klickbare Grafiken sind nützliche Hilfen für die Präsentation längerer Texte. Die weite Verbreitung der Viewer-Software macht dieses Format zum brauchbaren Medium für das Publizieren elektronischer Texte. Zur Nutzung von Winhelp-Dateien muß lediglich das Programm winhelp.exe gestartet und die .hlp-Datei geöffnet werden.

Komprimierte Dateien

Um Speicherplatz und Übertragungszeit zu sparen, werden Dateien komprimiert. Unter DOS wird meist das Komprimierungsprogramm "pkzip" verwendet; die Dateien erhalten dann die Endung ".zip". In einer pkzip-Datei lassen sich eine oder mehrere Dateien komprimiert archivieren und gebündelt weitergeben. Gleiches gilt für andere unter DOS verwendete Komprimierungsprogramme wie "arj" oder "lharc". Unter Unix wird das Archivieren und Komprimieren in zwei Schritten vorgenommen. Sind mehrere Dateien zu bündeln, werden diese zunächst mit "tar" in einer einzigen Datei zusammengefaßt (Endung ".tar") und anschließend mit "compress" (Endung ".Z") oder "gzip" (Endung ".z" oder ".gz") komprimiert. Solche Dateien tragen dann die doppelte Endung ".tar.Z" bzw. "tar.gz", wobei letzteres auch mit ".tgz" abgekürzt wird.

Prinzipiell benötigt man für jede Art der Komprimierung ein entsprechendes Programm zur Entkomprimierung. Solche Hilfsprogramme sind via ftp überall erhältlich (siehe Tabelle 2). Einfacher ist allerdings die Benutzung integrierter Tools wie z.B. "Winzip". Es kann nicht nur pkzip-, sondern auch tar-, compress- und gzip-Dateien entpacken und ist übersichtlich unter Windows bedienbar. Außerdem ermöglicht es die Untersuchung des Inhalts komprimierter Dateien ohne die Notwendigkeit, sie zu entpacken. Mit diesem Programm sind die am weitesten verbreiteten Kompressionsformate abgedeckt; zudem ist es möglich, zusätzliche Entpacker zur Behandlung weiterer Formate einzubinden.

Nicht ganz so komfortabel, aber dafür kostenlos ist der "StuffIt Expander". Dieses Windows-Programm entpackt nicht nur die DOS-Kompressionsformate zip, arc und arj sowie die Unix-Formate gz und Z, sondern auch die Macintosh-Formate sit und hqx. Daneben kann es auch die im nächsten Abschnitt besprochenen uu-codierten Dateien decodieren. Lediglich tar-Dateien können nicht entpackt werden.

Für den Mailversand codierte Dateien

Für den Versand von Dateien via E-Mail oder für die Verbreitung über die Usenet-News wird - allein oder zusätzlich zur Komprimierung - eine besondere Codierung verwendet. Grund hierfür ist die historische Beschränkung von E-Mail-Transportsoftware auf die 128 Zeichen des US-Ascii-Zeichensatzes. Binärdateien wie z.B. Programme, aber auch Textverarbeitungs- oder komprimierte Dateien, enthalten aber Zeichen außerhalb dieses Rahmens. Sie müssen deshalb auf den zulässigen Zeichensatz umcodiert werden, um unbeschädigt transportiert werden zu können. Hierfür wird das aus der Unix-Welt stammende Programm "uuencode" verwendet. Die standardmäßige Endung solcher

Dateien ist ".uue"; erkennen kann man sie an einem langen Block von Zeilen genau gleicher Länge (meist 60-61 Zeichen), die scheinbar zusammenhanglose Zeichen, jedoch immer ohne Leerzeichen, enthalten. Anfang und Ende der Codierung sind mit "begin" und "end" gekennzeichnet. Zur Dekodierung dient z.B. das Programm "uudecode" (vgl. Tabelle 2) oder der bereits besprochene "StuffIt Expander".

ASCII-Dateien ("Textdateien")

Ascii-Dateien, oft auch "Text-" oder "reine Textdateien" genannt, sind der kleinste gemeinsame Nenner für den Austausch von Texten. Der Grund hierfür ist die ausschließliche Verwendung des ASCII-Zeichensatzes, der wiederum von allen Editoren und Textverarbeitungen gelesen werden kann.

Allerdings steckt hier der Teufel oftmals im Detail, da lediglich die Verwendung der ersten 128 Zeichen dieses Zeichensatzes einheitlich erfolgt (sog. 7-bit- oder US-ASCII). Für landesspezifische Sonderzeichen wie die deutschen Umlaute werden unterschiedliche Codierungen verwendet. So basieren MS-DOS-Programme auf dem sogenannten "erweiterten" ASCII-Zeichensatz mit 256 Zeichen (auch PC-8, 8-bit-Ascii- oder IBM-PC-Zeichensatz genannt), während Windows-Programme den ANSI-Zeichensatz verwenden.

Bei der Bearbeitung englischer Texte spielt dies keine Rolle. Wird jedoch z.B. ein deutscher Text, der mit einem DOS-Editor erstellt wurde, mit dem Windows-Notizblock geöffnet, erscheinen anstelle der Umlaute andere Sonderzeichen.

Auch die Kennzeichnung des Zeilenendes stiftet Verwirrung, da das Vorbild der Zeilenschaltung aus alten Schreibmaschinen-Zeiten unterschiedlich umgesetzt wird. Im Ergebnis sehen Unix- und Macintosh-Dateien unter MS-DOS wie kontinuierlich geschriebene Texte ohne Absätze und Zeilenumbrüche aus, sind dafür aber gespickt mit eigenartigen Sonderzeichen (nämlich an den Stellen, an denen üblicherweise eine neue Zeile beginnt).

Zwar gibt es genügend Filter- und Konvertierungsprogramme, die solche Dateien auch unter DOS wieder lesbar machen, aber einfacher ist die Verwendung leistungsfähiger Textverarbeitungen. Diese haben entsprechende Filter bereits eingebaut, man muß ggf. nur den richtigen auswählen. Word für Windows verwendet beispielsweise die Bezeichnung "Nur Text" für die ANSI-Codierung, wie sie von allen Windows-Programmen benutzt wird, und die Bezeichnung "MS-DOS-Text" für die unter DOS übliche 8-bit-ASCII-Codierung.

Wird eine Datei mit dem richtigen Filter geöffnet, ist sie anschließend problemlos mit den Funktionen der Textverarbeitung zu betrachten, zu bearbeiten und zu drucken.

Wer nicht mit (Textverarbeitungs-) Kanonen auf Spatzen schießen will und ein kleines, schnelles Programm zur Textbetrachtung unter Windows sucht, dem sei der "JEC-Fileviewer" empfohlen. Er kann (im Gegensatz zum Windows-Notizblock) auch große Textdateien verarbeiten und erlaubt die Suche nach Textpassagen. Allerdings kann es bei deutschen Texten zu den genannten Problemen mit Umlauten und Sonderzeichen kommen.

Textverarbeitungs-Dateien

Dateien, die mit Textverarbeitungsprogrammen erstellt wurden, folgen herstellerspezifischen Standards. Idealerweise sollten sie mit dem gleichen Programm, mit dem sie erstellt wurden, auch wieder gelesen werden, um alle Formatierungs- und Gliederungsmerkmale garantiert zu erhalten. Ist dies nicht möglich, sollte zunächst versucht werden, solche Dateien mit einer anderen Textverarbeitung zu öffnen. Die modernen Textverarbeitungsprogramme haben Importfilter für die Formate anderer Hersteller, wenn auch oftmals nicht für die jeweils allerneueste Version. Sollte auch dies

nicht gelingen, gibt es noch die Möglichkeit, nach einem der bereits erwähnten Viewer Ausschau zu halten. Für das weit verbreitete Format von Word für Windows gibt es kostenlos einen Viewer zum Ansehen und Ausdrucken der entsprechenden Dateien (siehe Tabelle 2). Bei exotischen Formaten, für die keine Viewer oder Filter existieren, bleibt nur der Weg, den Absender/Autor um eine Version in einem anderen Format zu bitten.

TeX/LaTeX/DVI

TeX und LaTeX sind in der Unix-Welt weitverbreitete Systeme zur Erstellung vor allem wissenschaftlicher Publikationen, die wegen Ihrer Vorzüge bei der Vorbereitung drucksatzgerechter Layouts und bei der Einbindung von Grafiken und mathematischer Formeln den Textverarbeitungen der PC-Welt überlegen sind - allerdings erfordern sie einen erheblichen Einarbeitungsaufwand.

Der Quell- oder Rohtext im ASCII-Format wird vom Autor mit Steuerzeichen versehen, um z.B. bestimmte Formatierungen oder Gliederungen zu kennzeichnen. Anschließend wird der Rohtext vom TeX-System kompiliert, d.h. alle Formatierungs- und sonstigen Steueranweisungen werden umgesetzt. Das Ergebnis dieses Prozesses ist eine DVI- (Device Independent) Datei, die nicht mehr direkt lesbar ist, sondern mit spezieller Software in hoher Qualität angezeigt oder gedruckt werden kann.

Auch wenn diese DVI-Dateien eigentlich nur eine Vorstufe für die Postscriptausgabe sind, werden sie doch oft auch als Format zur Verbreitung von Informationen eingesetzt, weil im wissenschaftlichen Bereich die entsprechende Software weit verbreitet ist. Für den PC existieren zwar auch Programme zur Betrachtung solcher Dateien (vgl. Tabelle 2), diese sind aber relativ umfangreich - vor allem wegen der benötigten Fonts - und sollten daher nur eingesetzt werden, wenn die gesuchte Information nicht in einem anderen Format erhältlich ist.

Postscript

Postscript ist immer noch DER Standard zur Verbreitung von wissenschaftlichen Publikationen. Es ist unabhängig vom jeweiligen Betriebssystem verwendbar und kann von allen Textverarbeitungs- und Satzsystemen erzeugt werden.

Ein Grund für die weite Verbreitung von Postscript ist die optimale Ausnutzung der vorhandenen Ausgabegeräte für qualitativ hochwertige Ausdrücke bis hin zur Druckvorstufe und Satzbelichtung. Tatsächlich ist es auch als Format für die Druckausgabe entwickelt worden, nicht zum Austausch oder zur Weitergabe von Informationen. Weil aber das vom Autor gewählte Layout unverändert erhalten bleibt und ausgedruckt werden kann, hat es sich auch als Übertragungsformat für die elektronische Publizierung durchgesetzt.

Postscript ist eine "Seitenbeschreibungssprache", orientiert am Paradigma der Druckseite. Es wird jeweils eine komplette Seite mittels Steuercodes und - im Falle von Grafiken - mathematischen Zeichenanweisungen definiert. Das Ergebnis ist eine ASCII-Textdatei, die ähnlich einem Programmcode eine Vielzahl von Steueranweisungen für die Druckausgabe enthält. Im Prinzip wird die Formatierung und Platzierung von Buchstaben und Zeichen auf der Seite definiert, wobei jegliche syntaktische und gliederungsmäßige Information des Ursprungstextes verlorengeht.

Dieser Code wird schließlich vom Ausgabegerät, z.B. einem Postscript-Drucker, umgesetzt. Allerdings kann die Umsetzung auch mittels geeigneter Software erfolgen (vgl. Tabelle 2), so daß PS-Dateien auch auf dem Bildschirm oder auf nicht-PS-Druckern ausgegeben werden können.

Wer also einen Postscript-Drucker besitzt, kann eine PS-Datei einfach durch das Abschicken dieser Datei an den Drucker ausdrucken, beispielsweise mit dem copy-Befehl (etwa "copy [Dateiname] lpt1"). Komfortabler ist es allerdings, sich eine solche Datei erst am Bildschirm anzusehen, um dann zu entscheiden, ob und was ausgedruckt werden soll. Hierfür ist Software nötig, die die Postscript-Kommandos interpretieren und in eine Bildschirmanzeige umsetzen kann. Bei Unix-Systemen sind solche Programme meist im Lieferumfang enthalten. Für DOS-PCs gibt es das in Tabelle 2 aufgeführte Programm "ghostscript", das frei verfügbar ist. Zur leichteren Bedienung empfiehlt es sich, auch das dazugehörige "ghostview" zu installieren, denn damit kann ghostscript komfortabel unter Windows bedient werden. Der wohl größte Vorzug von ghostscript/ghostview ist die Möglichkeit, auch ohne Postscript-Drucker drucken zu können, und zwar nicht nur ganz, sondern auch auszugs- und seitenweise.

Allerdings muß man hierbei kleine Qualitätsabstriche hinnehmen, weil die Originalfonts urheberrechtlich geschützt sind und daher in der frei erhältlichen Version durch annähernd ähnliche Bitmap-Fonts ersetzt werden.

Der kommende Standard: Adobe Acrobat (PDF-Dateien)

Wie erläutert, wurde Postscript hauptsächlich im Hinblick auf eine hochwertige und betriebssystemunabhängige Druckausgabe konzipiert. Die Entwickler von Postscript, die Firma Adobe, haben nun ein neues Dateiformat vorgestellt, das die Schwächen von Postscript umgehen und gleichzeitig den Austausch und die Verbreitung von Dokumenten im Originallayout erleichtern soll.

Dieses neue Format, genannt PDF (Portable Document Format), basiert auf der Funktionsweise von Postscript, erweitert es aber um mehrere wichtige Funktionen:

Es ist nicht nur für die Druck-, sondern auch für die Bildschirmausgabe geeignet, es sind Hypertextlinks möglich (womit sich auch die Navigation im Dokument erheblich vereinfacht, z.B. durch anklickbare Inhaltverzeichnisse, Indizes, Querverweise etc.), und es können Zeichenketten gesucht werden. Darüberhinaus sind PDF-Dateien - mit Zusatzsoftware - indizierbar. Bei Ihrer Erstellung werden Kompressionsverfahren eingesetzt, so daß die Dateigröße in einem vertretbaren Rahmen bleibt.

Obwohl PDF ein proprietärer Standard ist, ist die Nutzung von PDF-Dateien dank des bereits erwähnten "Viewer"-Prinzips problemlos möglich: Während für die Erstellung dieser Dateien Software von Adobe erworben werden muß, stellt die Firma unter dem Namen "Acrobat Reader" zum Ansehen und Drucken von PDF-Dateien kostenlose Viewer für alle Betriebssysteme zur Verfügung - für Windows sogar in einer deutschen Version. Zwar können PDF-Dateien auch von der neusten Version 3.3 des Postscript-Viewers Ghostscript (s.o.) verarbeitet werden, allerdings können nur mit dem "Acrobat Reader" alle beschriebenen Leistungsmerkmale genutzt werden.

PDF erlaubt dem Autor die vollständige Kontrolle über das Erscheinungs- und Druckbild des Dokumentes - es wird vom Viewer so wiedergegeben, wie es erstellt wurde, natürlich immer abhängig von der Leistungsfähigkeit des eingesetzten Monitors bzw. Druckers.

Darüberhinaus kann der Benutzer in den Dateien auch nach Textteilen suchen, Annotationen vornehmen und ganz oder in Auszügen drucken. Diese im Gegensatz zu Postscript unkomplizierte und komfortable Handhabung trägt dazu bei, daß sich PDF als Format für die Verbreitung layoutabhängiger Publikationen etabliert. So bietet die amerikanische Bundes-Finanzbehörde Formulare im PDF-Format zum Download im Internet an, Firmen verteilen Prospekte und Informationsmaterial und Verlage bieten ihre Publikationen oder - als Appetithäppchen - Auszüge davon in diesem Format an.

Details zur genauen Funktionsweise von PDF finden sich in (2)², aber auch ohne die technischen Aspekte genauer zu beleuchten, läßt sich prognostizieren, daß die Vorzüge, die PDF gegenüber anderen Formaten aufweist, es zum vielversprechenden Kandidaten für einen neuen Standard machen. Insbesondere aus Sicht von Informationsanbietern (ob kommerziell oder nicht) bietet die Kombination aus absoluter Layoutkontrolle und leichter Bedienbarkeit auf Nutzerseite immense Vorteile - ganz zu schweigen von kommerziell nutzbaren Features wie z.B. Verschlüsselung und Passwortabfrage...

Vor allem aber etabliert sich PDF als nützliche Ergänzung zu HTML, dem Standard für die Erstellung von Dokumenten im World Wide Web. Um diese Entwicklung weiter zu fördern, bietet Adobe zusätzlich zum "Reader" ebenfalls kostenlos eine Erweiterung an, die den "Reader" mit WWW-Browsern wie Netscape oder Mosaic verbindet. So wird es ermöglicht, Hypertextlinks aus PDF-Dateien auf externe Ressourcen im Internet zu verfolgen.

Integration von Viewern für die Nutzung des World Wide Web

Die hier besprochenen Dateiformate werden insbesondere zur Verbreitung von Informationen im Internet genutzt. Dabei werden Beschreibungen oder Nachweise häufig über Mailinglisten, News oder im WWW veröffentlicht, während die eigentliche Datei über FTP oder ebenfalls im WWW abrufbar ist. Ebenso häufig kommt es vor, daß umfangreiche Texte für die Darstellung im WWW aufbereitet, gesplittet und konvertiert werden, und die Original- oder Komplettfassung dann in einem anderen Format, meist Postscript oder PDF, angeboten wird.

Die vorgestellten Viewer lassen sich einfach zur Zusammenarbeit mit WWW-Browsern wie Netscape und Mosaic einstellen. Es muß lediglich vorgegeben werden, welche Dateiformate (und -endungen) welchem Programm zugeordnet werden. Dies erfolgt in den jeweiligen Konfigurationsmenüs von Netscape oder Mosaic. Ist alles richtig eingestellt, so führt z.B. das Anwählen eines Verweises auf eine PDF-Datei dazu, daß diese Datei vom Zielrechner geholt wird, der Acrobat-Reader gestartet und die Datei gleich angezeigt wird. Gleiches gilt für Postscript-Dateien und ghostview sowie für komprimierte Dateien und Winzip. Mit den drei letztgenannten Programmen sowie einer leistungsfähigen Textverarbeitung ist man für die Nutzung elektronischer Dokumente aller Art gut gerüstet.

²Merz, Thomas: Acrobat schön.- In: iX (1995) 8, S. 130-135

Tabelle 1: Dateiformate, Endungen, Programme

Dateiname (Endung)	Dateiformat	Passendes DOS/Windows- Programm
<i>Dateien mit Kombinationen mehrerer Endungen (siehe jeweils dort) von rechts nach links abarbeiten</i> Beispiel: <i>"Text.ps.z" - erst mit gunzip/winzip auspacken, dann mit Ghostview betrachten</i>		
arj	mit arj komprimiert	unarj
doc	Ascii-Text oder mit Textverarbeitung erzeugte Datei	Textverarbeitung
dvi (vgl. "tex")	mit TEX kompiliertes Dokument	dviwin
eps (siehe "ps")	"encapsulated PostScript",	ghostscript/ghostview
gz	mit "gzip" komprimiert (Unix)	gunzip, winzip
hlp	evtl. eine Datei im Windows-Hilfe-Format	winhelp.exe (mit Windows mitgeliefert)
hqx (Mac)	Macintosh-Codierung entsprechend uue	StuffIt-Expander
pdf	Adobe "Portable Document Format"	Acrobat Reader, u.U. ghostscript/ghostview
ps	Postscript-Datei	ghostscript/ghostview
sit (Mac)	mit StuffIt komprimiert	StuffIt-Expander
tar	Unix-tar-Datei	winzip,
tex (vgl. "dvi")	Quelldatei für das Satzsystem TEX	-
tgz	Kombination aus "tar" und "gz" (siehe dort)	winzip
txt	wahrscheinlich Ascii-Text	Textverarbeitung, JEC-Fileview
uue	mit uuencode codierte Datei	uudecode, StuffIt-Exp.
Z	mit "compress" komprimiert (Unix)	gunzip, winzip
z (siehe "gz")		

zip	mit "pkzip" komprimiert	pkunzip, unzip, winzip
-----	----------------------------	---------------------------

Tabelle 1: Dateiformate, Endungen, Programme

Tabelle 2: Viewer und Entpacker

Programm (Dateiname der aktuellen Version)	Kurzbeschreibung	Preis
Acrobat Reader	Zeigt und druckt Adobe-pdf-Dateien	Kostenlos
File 2.0 (ffile20.zip)	erkennt eine Vielzahl von Dateiformaten	Shareware, 30 DM
Dviwin (dviwin29.zip)	Zeigt und druckt dvi-Dateien unter Windows	Kostenlos
Ghostscript / Ghostview (gs333.zip; gsview.zip)	Zeigt Postscript Dateien unter Windows an und ermöglicht Druckausgabe auf nicht-Postscript-Druckern	Kostenlos
J.E.C File Viewer (fview118.zip)	Zeigt und druckt Ascii-Dateien ohne Größenbeschränkung unter Windows an, ermöglicht Suche nach Textstellen	Kostenlos
Microsoft Word Viewer (wordvu.exe)	Zum Ansehen und Drucken von Winword-Dateien	Kostenlos
StuffIt Expander (sitex10.exe)	Entpackt arj, arc, gzip, sit, zip - Dateien, decodiert uue und hqx - Dateien	Kostenlos
unarj (unarj.exe)	Entpackt arj-gepackte Dateien	Kostenlos
uudecode (uudecode.com)	decodiert uucodierte Dateien	Kostenlos
WinZip 5.6 (winzip56.zip)	Entpackt komfortabel zip, tar, gz, Z Dateien, ermöglicht auch das Lesen oder Probeinstallieren gepackter Dateien	Shareware, 29 \$
<p><i>Bezugsquellen: mit den in Spalte 1 angegebenen Dateinamen können alle Programme im Internet mittelsarchie gefunden werden. Leser mit WWW-Zugang können eine Zusammenstellung der Fundstellen aller hier angesprochenen Programme unter http://www.fh-koeln.de/fbi/pers/babiak/abi.html nutzen</i></p> <p><i>Programme zur Dekomprimierung gepackter Dateien (z.B. pkunzip/unzip, unarj, gunzip, tar usw.) sind auf vielen FTP-Servern verfügbar, so zum Beispiel auf ftp.uni-stuttgart.de; Verzeichnis /pub/systems/pc/archiver. Sie sind meist kostenlos</i></p>		

Tabelle 2: Viewer und Entpacker