

White Paper

Meta-Daten in Microsoft Office und in PDF-Dokumenten


Typen, Übernahme, Anzeige und Löschen


Inhaltsverzeichnis

Begriffsdefinition und Typen.....	2
Definitionen.....	2
Einige Typen	2
Meta-Daten in Office-Dateien.....	2
Word	3
Excel	4
PowerPoint	4
Meta-Daten in PDF-Dokumenten.....	5
Allgemeine Meta-Daten	5
Meta-Daten aus Office-Dokumenten.....	5
Darstellung und Anzeige	6
Entfernen von Meta-Daten.....	6

Begriffsdefinition und Typen

Definitionen

Unter Metadaten  nach einer ersten, ganz allgemeinen Definition Daten zu verstehen, die andere Daten beschreiben beziehungsweise Informationen über diese Daten geben. So sind beispielsweise die Buchstaben oder Zeichen in einem Text Daten. Die Anzahl der Buchstaben in dem Text sind Metadaten, die auch als Zusatzinformationen über die „eigentlichen“ Daten hinaus bezeichnet werden können.

Wenn man den Begriff etwas enger fasst und im Zusammenhang mit Dateitypen verwendet, sind zum Beispiel Informationen wie der Name und Titel einer Datei, ihres Verfassers, Stichwörter zum Inhalt der Datei oder das Datum der Speicherung Metadaten. Dann können Meta-Daten oder Meta-Informationen so charakterisiert werden, dass sie in einer Datei abgespeichert werden, aber in der Regel nicht auf den ersten Blick sichtbar sind, wenn man die Datei in einer Anwendungssoftware öffnet. Auch geht beim Ausdruck eines Dateiinhalts (zum Beispiel eines Dokuments) auf Papier die eine große Zahl Meta-Informationen verloren. Da dies bei der Weitergabe von Dateien in der Regel nicht passiert, können Dateien mehr Informationen enthalten, als der Ersteller eigentlich preisgeben will.¹ 

Ein zweifellos positiver Effekt von Metadaten in Dateien ist andererseits: Sie können dazu dienen, Datenbestände nach bestimmten, allgemeinen Kriterien leichter zu katalogisieren und gezielter durchsuchen zu können.

Einige Typen

Neben den Metadaten in Textdokumenten enthalten zum Beispiel auch Musik- und Videodateien, Fotos oder html-Seiten Meta-Informationen. In Musikdateien sind dies unter anderem Informationen über den Interpreten und den Titel oder spezielle Tags. Bei Videos zum Beispiel die Wiedergabedauer, Größe und Bildrate. In Fotos sind Meta-Daten in Form von Bild-Informationen häufig im Exif- oder IPTC-Format abgespeichert. Metadaten in HTML-Seiten (Seitentitel, Stichwörter zum Seiteninhalt etc.) werden angelegt, damit die Seiten in den Trefferlisten von Suchmaschinen angezeigt werden.

Die folgenden Ausführungen behandeln ausschließlich die Verwendung von Meta-Daten in Microsoft-Office-Dateien und in PDF-Dokumenten. Dabei werden die Office-Anwendungen Word, Excel und PowerPoint betrachtet.

Meta-Daten in Office-Dateien

Office-Dokumente allgemein

In Office-Dateien wird beim Speichern des Dokuments eine Vielzahl an Meta-Informationen abgelegt. Dies sind zum einen Informationen, die in Word, Excel und PowerPoint als „Dokumenteigenschaften“ bezeichnet werden und bei allen drei Dokumenttypen an

¹Siehe Rost, Martin; Wallisch, Arnold, Was Office-Dateien verraten können, c't Heft 2-2003, Seite 172ff.
Hinweis: Alle in diesem Dokument verwendeten Namen und Bezeichnungen von Produkten und Firmen sind u. U. eingetragene Waren- oder Markenzeichen der jeweiligen Hersteller oder Rechteinhaber.

derselben Stelle der Anwendung eingesehen (und gegebenenfalls geändert) werden können. Zum anderen sind es Informationen, die nur in der jeweiligen Anwendung vorkommen.

Einen Teil dieser Informationen (die Dokumenteigenschaften) kann man ansehen, wenn in den Versionen bis Office 2003 im Menü „Datei“ auf „Eigenschaften“ geklickt wird. In Word, Excel und PowerPoint 2007 wird zunächst auf die Microsoft Office-Schaltfläche, dann auf „Vorbereiten“ und schließlich auf „Eigenschaften“ geklickt. Alternativ können in Office 2007 die Eigenschaften auch beim Öffnen oder Speichern einer Datei angezeigt werden, und zwar im Dialogfeld „Öffnen“ oder im Dialogfeld „Speichern unter“. Die meisten der Dokumenteigenschaften können für das aktuell geöffnete Dokument nicht nur angesehen, sondern auch geändert werden. Zudem lassen sich neben den standardmäßig vorgegebenen auch benutzerdefinierte Eigenschaften erstellen²



Zu den wichtigsten, allgemeinen Dokumenteigenschaften gehören (am Beispiel Word 2007 erläutert):

- ❖ Standardeigenschaften, die vom Anwender im Anzeigedialog geändert werden können. Hierunter fallen Titel, Thema, Verfasser (automatisch der Name, den der Anwender bei der Installation angegeben hat), Firma, Stichwörter, Kommentare.

>Bild aus Word 2007>

- ❖ Dynamisch aktualisierte oder aktualisierbare Eigenschaften, die vom Anwender im Anzeigedialog nicht geändert werden können. Hierzu zählen Speicher- und Bearbeitungszeiten, Speicherpfad, statistische Daten (Anzahl Seiten, Absätze, Zeilen, Wörter etc.).

Zu den dynamischen Metainformationen gehören ganz allgemein alle Autotextfelder, also „Speicherorte“ für mehrfach verwendeten Text oder Grafiken, die wieder verwendet werden sollen.

>Bild aus Word 2007>

- ❖ Benutzerdefinierte Eigenschaften, die mit Zahlen, Texten, Zeitangaben oder dem Wert „Ja“/ „Nein“ versehen werden können.

>Bild aus Word 2007>

Neben den allgemeinen Dokumenteigenschaften sind weitere Meta-Informationen in Office-Dateien gespeichert, die jeweils für Word, Excel und PowerPoint spezifisch sind.

Word

Anwendungsspezifische Metadaten in Word-Dateien sind zum Beispiel

- ❖ Sicherheitsoptionen (Informationen über Verschlüsselung und Kennwortschutz)

² Zu Details siehe die Hilfe und Anleitungen auf <http://office.microsoft.com/de-de/help/HA100475241031.aspx>
Hinweis: Alle in diesem Dokument verwendeten Namen und Bezeichnungen von Produkten und Firmen sind u. U. eingetragene Waren- oder Markenzeichen der jeweiligen Hersteller oder Rechteinhaber.

- ❖ Verweise (Links): Navigation-Informationen über Sprungstellen von internen Verweisen zwischen den Seiten eines Word-Dokuments und externe Links beziehungsweise Internet-Links
- ❖ Inhaltsverzeichnis (Table of Contents, TOC) mit Sprung-Informationen, welche Seite beim Anklicken einer Überschrift im Inhaltsverzeichnis geöffnet werden soll
- ❖ Informationen über vorhandene Formularfelder, allgemeine Steuerelemente und Webtools
- ❖ (Ausgeblendete) Anmerkungen und Kommentare
- ❖ Fußnoten, Endnoten
- ❖ Hinweisen auf Änderungen
- ❖ (Ausgeblendete) Markup-Hinweise
- ❖ Ausgeblendete Texte und (nicht sichtbare) Seitenhintergründe
- ❖ Das Stichwortverzeichnis (der Index)

Excel

Anwendungsspezifische Metadaten in Excel-Arbeitsmappen sind

- ❖ Sicherheitsoptionen (Informationen über Verschlüsselung und Kennwortschutz)
- ❖ Verweise (Links): Navigation-Informationen über Sprungstellen von internen Verweisen zwischen Blättern einer Excel-Arbeitsmappe und externe Links beziehungsweise Internet-Links
- ❖ Informationen über vorhandene Formularfelder oder allgemeine Steuerelemente
- ❖ Kommentare
- ❖ Gitternetzlinien
- ❖ Zeilen- und Spaltenköpfe

PowerPoint

Anwendungsspezifische Metadaten in PowerPoint-Präsentationen sind

- ❖ Sicherheitsoptionen (Informationen über Verschlüsselung und Kennwortschutz)
- ❖ Verweise (Links): Navigation-Informationen über Sprungstellen von internen Verweisen zwischen den Folien einer Präsentation und externe Links beziehungsweise Internet-Links
- ❖ Informationen über vorhandene Formularfelder oder allgemeine Steuerelemente
- ❖ Notizen/Kommentare
- ❖ Gitternetzlinien
- ❖ Zeilen- und Spaltenköpfe
- ❖ (Ausgeblendete) Folien

Hinweis: Alle in diesem Dokument verwendeten Namen und Bezeichnungen von Produkten und Firmen sind u. U. eingetragene Waren- oder Markenzeichen der jeweiligen Hersteller oder Rechteinhaber.

- ❖ Folienrahmen
- ❖ Angepasste Graustufen-Darstellung

Meta-Daten in PDF-Dokumenten

*Allgemeine Meta-Daten*³

PDF-Dateien enthalten unabhängig vom Dateiformat der Ursprungsanwendung üblicherweise allgemeine Metadaten wie den Titel und Verfasser des Dokuments sowie Informationen zu seiner Erstellung und zu Änderungen. In PDF-Dateien können Metadaten durch den Anwender oder durch Plugin-Erweiterungen geändert oder neu hinzugefügt werden, wobei die Darstellung in PDF-Viewern von Adobe bei den Dokumenteigenschaften auf maximal 255 Bytes begrenzt ist.⁴ Mit der Version 1.4 der PDF-Spezifikation können Metadaten auch für einzelne Komponenten eines Dokuments definiert werden. In einem PDF-Dokument werden die Metadaten entweder in einem Dokumentinformationswörterbuch (Document Information Directory) oder in Metadaten-Strömen (Metadata Streams) gespeichert.

Meta-Daten aus Office-Dokumenten

Sofern es sich bei der Ausgangsdatei um eine Office-Datei aus Word, Excel oder PowerPoint handelt, können mit PDF-Convertern, die die Übernahme von Meta-Daten aus Office-Dateien beherrschen, in der Regel darüber hinaus die oben für die drei Programme aufgelisteten anwendungsspezifischen Metainformationen in die PDF-Datei übernommen werden. Solche Converter sind zum Beispiel [Office to PDF Premium](#), Word to PDF oder Excel to PDF von soft Xpansion.

Da es sich bei den Office-Dokumenten einerseits und bei den PDF-Dateien andererseits um separat voneinander entwickelte Dateitypen handelt, ist darauf hinzuweisen, dass nicht alle Metadatentypen aus Office-Dateien in PDF-Dokumente importiert werden können. So fehlen für einige Metadaten im Objektmodell des PDF-Formats die technischen Voraussetzungen für eine Übernahme, zum Beispiel bei den Autotextfeldern (nicht für deren im Dokument angezeigten Inhalt!).

Umgekehrt ist es technisch zum Beispiel möglich, ein Stichwortverzeichnis (Index) aus einer Word-Datei so in eine PDF-Datei zu importieren, dass die einzelnen Einträge im Verzeichnis in der PDF-Datei als Verweise angezeigt werden. Auf diese Verweise kann man dann klicken, um zu der Seite zu springen, die im Index angegeben ist.

Problematisch kann hier sein, dass im PDF-Format, anders als im Word-Dokument, aber keine Möglichkeit besteht, nach dem Hinzufügen oder Löschen von PDF-Seiten

³ Siehe Adobe Systems Incorporated, PDF Reference, Version 1.7, Seiten 843 - 847, zu den Grundlagen von Metadaten in PDF-Dateien

⁴ Ebenda, Seite 1125, Anmerkung 160.

Hinweis: Alle in diesem Dokument verwendeten Namen und Bezeichnungen von Produkten und Firmen sind u. U. eingetragene Waren- oder Markenzeichen der jeweiligen Hersteller oder Rechteinhaber.

den Index in PDF-Datei mit einem „Felder aktualisieren“-Befehl automatisch auf den neuesten Stand zu bringen.

Darstellung und Anzeige

Meta-Informationen wie Verfasser, Titel, Thema oder Schlüsselwörter werden, sofern eine Übernahme beim Erstellen der Datei bestimmt wurde, in PDF-Dateien üblicherweise in Dialogfenstern angezeigt, die mit „Dokumenteigenschaften“ oder ähnlich bezeichnet sind. Andere Metadaten wie zum Beispiel im Originaldokument ausgeblendete Anmerkungen, Notizen und Kommentare, Gitternetzlinien, Folien oder Folienrahmen werden in das PDF-Dokument übernommen und dort auch sichtbar angezeigt, wenn die Übernahme beim Erstellen der PDF-Datei festgelegt beziehungsweise erlaubt wurde.

Entfernen von Meta-Daten

Bisher wurde erläutert, welche Metadaten aus Office-Dateien in PDFs übernommen werden können und wie dies erfolgen kann. In bestimmten Fällen ist es allerdings sinnvoll, vor der Konvertierung in PDF und Weitergabe einer Datei alle oder bestimmte in ihr enthaltenen Meta-Daten zu entfernen. Nämlich dann, wenn der Verfasser des Dokuments sie eigentlich nicht bekannt geben möchte. Hier gibt es prominente Fälle, in denen dies aus Sicht der Verfasser versäumt wurde, wie den des Dokuments über die im Irak vermuteten Massenvernichtungswaffen, die den Einstieg Großbritanniens in den Irak-Krieg rechtfertigen sollten⁵. Eine Vielzahl an Metadaten aus Office-Dateien komfortabel entfernt werden kann, wird im Folgenden erläutert.

Für das Löschen von Metadaten bereits in der Originalanwendung Word, Excel oder PowerPoint selbst stehen zur Verfügung:

1. Zunächst besteht die Möglichkeit, Programm-Einstellungen innerhalb von Word, Excel und PowerPoint zu nutzen, um als sensibel erachtete Daten aus den Datei-Eigenschaften zu entfernen.⁶
2. Zudem können spezielle von Microsoft zur Verfügung gestellte Tools verwendet werden, um einige Metadaten zu löschen:

Bei Office bis Version 2003 erfolgt dies, wie auf der Seite <http://office-2003-xp-add-in-zum-entfernen-von-metadaten.shareware.de/> erläutert.

Bei Office 2007 erfolgt das Entfernen mit Hilfe einer *Dokumentinspektor* genannten Funktion, siehe

<http://office.microsoft.com/de-de/help/HA100375931031.aspx>

⁵ http://www.chip.de/artikel/Verraeterisches-Office-2_12875097.html

⁶ http://www.chip.de/artikel/Verraeterisches-Office-4_12875104.html

Hinweis: Alle in diesem Dokument verwendeten Namen und Bezeichnungen von Produkten und Firmen sind u. U. eingetragene Waren- oder Markenzeichen der jeweiligen Hersteller oder Rechteinhaber.

Darüber hinaus gibt es Freeware und ein kommerzielles Tool von einem anderen Hersteller, den *Document Trace Remover* und den *Metadata Analyzer*:

http://www.smartpctools.com/de/trace_removal/index.html (Kaufsoftware)

http://www.smartpctools.com/de/metadata_analyzer/index.html (Freeware)

Hinweis: Alle in diesem Dokument verwendeten Namen und Bezeichnungen von Produkten und Firmen sind u. U. eingetragene Waren- oder Markenzeichen der jeweiligen Hersteller oder Rechteinhaber.