FH-Dozent Ing. Mag. Horst Greifeneder
Allgemein beeideter und gerichtlich zertifizierter Sachverständiger, Fachgebiet Informationstechnik;
Lehrbeauftragter an der FH Salzburg, Master-Studiengang: Informationstechnik & System-Management;
Computer Forensik Spezialist.

Forensische Metadatenanalyse bei Office Word-Dokumenten

Microsoft Office-Anwendungen bilden in verschiedenen Einsatzbereichen wie Textverarbeitung, Tabellenkalkulation oder Präsentationen den De-facto-Standard im Privatleben und Geschäftsalltag. So ist es nicht weiter verwunderlich, dass vor allem Word- oder Excel-Dokumente als Beweismittel in zivil- oder strafrechtlichen Rechtssachen immer wieder eine wichtige Rolle spielen. In der Regel sind es Fragen zur Urheberschaft, zu inhaltlichen Änderungen bzw zu zeitlichen Aspekten der sichergestellten Dokumente, die das Gericht und in der Folge den Sachverständigen beschäftigen.

Der vorliegende Artikel ist der Beginn einer zweiteiligen Artikelserie zum Thema. Im ersten Teil werden Aspekte der forensischen Metadatenanalyse für Microsoft Word-Dokumente des Dateityps Word 97-2003 behandelt. Im zweiten Teil folgen Dokumente des Dateityps Word 2007-2013.

1. Forensische Fragestellungen

Im Zusammenhang mit der Auswertung von sichergestellten Word-Dokumenten stehen in der Regel nachfolgende Fragestellungen im Mittelpunkt des richterlichen Interesses:

- Wer hat das Dokument ursprünglich erstellt?
- Wer hat das Dokument zuletzt bearbeitet?
- Wann wurde das Dokument erstmals erstellt?
- Wann wurde das Dokument zuletzt bearbeitet?
- Wurde das Dokument ausgedruckt?
- Wie häufig wurde das Dokument bearbeitet?

Eine wertvolle Grundlage zur Beantwortung der angeführten, richterlichen Sachfragen bilden dabei die im Zusammenhang mit dem Dokument gespeicherten Metadaten.

2. Dateiformat für Word 97-2003

In Microsoft Office Word 2003, Microsoft Word 2002, Microsoft Word 2000 und Microsoft Word 97 ist das Binärdateiformat MS-DOC (*.doc) das Standarddateiformat. Dieses Dateiformat gilt für alle Dateien mit den Erweiterungen DOC und DOT.

Aus technischer Sicht handelt es sich beim MS-DOC-Dateiformat um ein Object Linking und Embedding (OLE) Compound Dateiformat. Eine Compound-Datei enthält eine Sammlung von Storages und Datenströme und ist vergleichbar mit einem Dateisystem: Storages sind Verzeichnisse, Datenströme sind Dateien.[1] Das Compound-Dateiformat ist die Basis für die Speicherung von Meta-Attributen zu Dateien (zB Autor, Firma, Versionsnummer usw).

Die Basisdateneinheit in einem Word-Dokument ist ein Zeichen, wozu Formatierungs- und andere nicht sichtbare Zeichen sowie ANSI- und Unicode-Zeichen zählen. Sämtliche Zeichendaten befinden sich im Word-Dokument-Datenstrom, dem Hauptdatenstrom einer DOC-Datei, welcher alle Daten in der Datei – außer für Tabellen – enthält.[2]

Am Anfang dieses Datenstroms befindet sich eine File Information Block (FIB) genannte Struktur, die Verweise auf alle Daten in der Datei enthält.

Bei den ersten 32 Bytes im FIB handelt es sich um eine Struktur namens FibBase. Die FibBase-Struktur enthält unter anderem die Kennzeichnung als Word-Binärdatei (0xA5EC), Informationen zur Installationssprache des Programms zur Dokumentenerstellung, zur Existenz von Bildern im Dokument und Sicherheitseinstellungen im Dokument wie Passwortschutz.[3]

3. Metadaten

Metadaten sind Daten über Daten, das heißt, Metadaten beschreiben bestimmte Eigenschaften eines Dokuments. Bei einem Word-Dokument beinhalten Metadaten unter anderem Informationen über den Autor, den Zeitpunkt der Erstellung und letzten Bearbeitung des Dokuments.

Immer wenn ein Benutzer in einem MS Office-Programm ein Dokument erstellt, druckt oder speichert, werden im Dokument – neben dem eigentlichen Dateiinhalt – ereignisbezogene Metadaten automatisch gespeichert. Metadaten werden für verschiedene Zwecke eingesetzt, zB um das Bearbeiten, Anzeigen, Ablegen und Abrufen von Office-Dokumenten zu verbessern.[4]

3.1. Metadaten bei MS Office Word

Metadaten in Word-Dokumenten werden auf verschiedene Weisen und aus unterschiedlichen Quellen erstellt. Daher

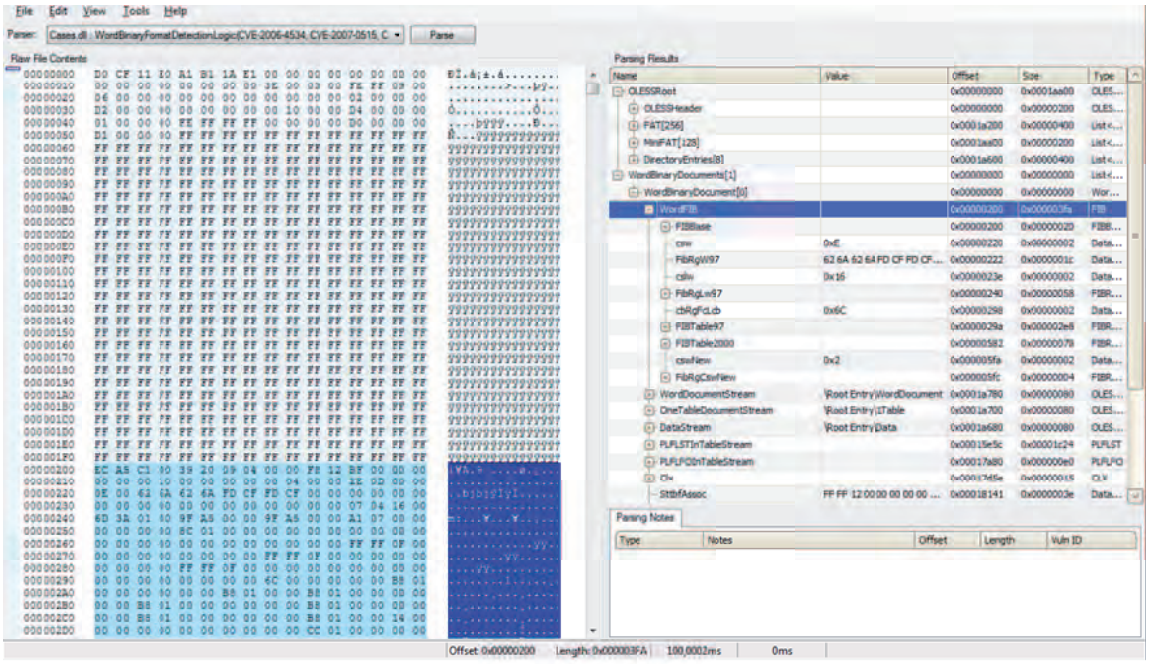

Abbildung 1: Struktur eines MS-DOC-Dokuments (mit hervorgehobenen FIB-Bereich)[5]

ist es nicht möglich, alle diese Inhalte mit einer einzigen Methode aus dem Dokument zu entfernen.[6]

Einen Teil der gespeicherten Metadaten kann man in Word bei geöffnetem Dokument über die Dokumenteneigenschaften einsehen und auch ändern.[7]

Zu den wichtigsten, allgemeinen Dokumenteigenschaften zählen:

- Standardeigenschaften, die vom Anwender im Anzeigedialog des Programms geändert werden können.[8] Hierunter fallen Autor (Benutzername, der bei der Installation des Programms angegeben wurde), Titel, Betreff, Schlüsselwörter, Kommentare.
- Dynamisch aktualisierte oder aktualisierbare Eigenschaften, die vom Anwender im Anzeigedialog nicht geändert werden können. Hierzu zählen Zeitstempel, Speicherpfad bzw statistische Daten wie die Anzahl der Seiten, Absätze, Zeilen, Wörter etc.

Neben den allgemeinen Dokumenteigenschaften können in Word-Dokumenten weitere forensisch relevante Metadaten gespeichert sein:[9]

- Sicherheitsoptionen wie Verschlüsselung und Kennwortschutz;
- Verweise zB Informationen über interne Verweise zwischen den Seiten eines Word-Dokuments und externe Links wie Internet-Adressen (URLs) oder E-Mail-Adressen;
- Informationen über Formularfelder, allgemeine Steuerelemente und Webtools;
- über die Auswahl hinzugefügte Kommentare zum Dokumenteninhalt;
- Verweise und Verknüpfungen, wie Fuß- oder Endnoten sowie Inhaltsverzeichnis (Table of Contents, TOC);
- falls die Funktion „Änderungen nachverfolgen“ aktiviert ist, Hinweise auf Änderungen, einschließlich Einfüge- und Löschvorgänge sowie Formatierungsänderungen;
- ausgeblendete Texte und (nicht sichtbare) Seitenhintergründe;
- Stichwortverzeichnis (Index) mit einer Liste von im Dokument enthaltenen Schlüsselwörter mit Verweis auf Seitenzahlen.

Die meisten Metadaten können im Anzeigedialog des Programms ausgelesen werden. Auf andere Metadaten kann nur mit besonderen Verfahren zugegriffen werden, etwa indem das Dokument in einem Hex-Editor geöffnet wird.

3.2. Speicherung der Metadaten

Das MS-DOC-Dateiformat speichert eine Reihe von dokumentenspezifischen Eigenschaften (Metadaten) im Summary Information Property Set des Dokuments.

Das Property Set wird im Word-Dokument-Datenstrom identifiziert durch die Formatkennzeichnung FMTID_SummaryInformation {F29F85E0-4FF9-1068-AB91-08002B27B3D9} und beinhaltet nachfolgende Metadaten.[10]

3.2.1. Titel

Verschiedene Office-Versionen speichern bei fehlenden Benutzerangaben im Feld „Titel“ standardmäßig bis zu 126 Zeichen der ersten Textzeile des Dokuments.

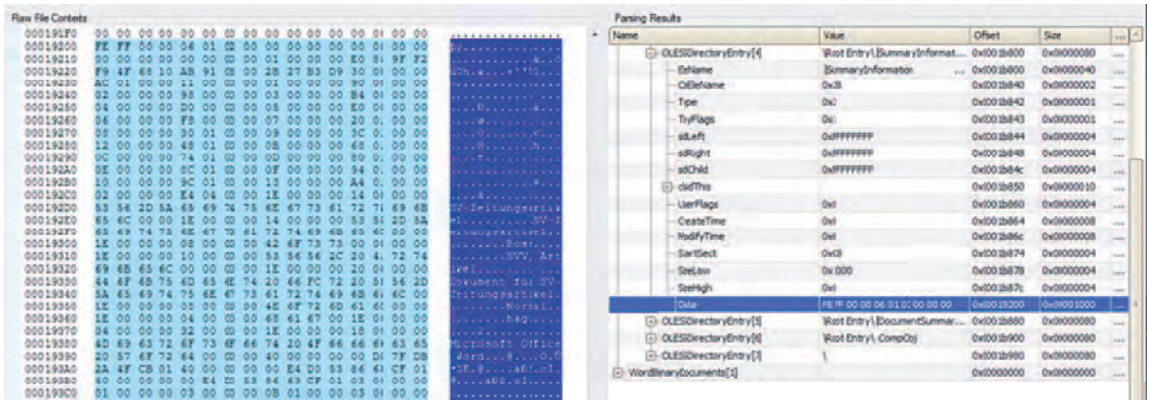

Abbildung 2: Summary Information Property Set im Word 97-2003-Dateiformat

Name	Property ID	Beschreibung
Title	0x00000002	Der Titel des Dokuments
Subject	0x00000003	Der Betreff des Dokuments
Author	0x00000004	Der Autor des Dokuments
Keywords	0x00000005	Schlüsselbegriffe zum Dokument
Comments	0x00000006	Kommentare zum Dokument
Template	0x00000007	Name der Dokumentenvorlage,
Last Saved By	0x00000008	Zuletzt geändert von <Benutzer>
Revision Number	0x00000009	Anzahl der Überarbeitungen des Dokuments
Total Editing Time	0x0000000A	Gesamtbearbeitungszeit des Dokuments
Last Printed	0x0000000B	Zuletzt gedruckt <Datum im FILETIME (UTC)-Format>
Create Time/Date	0x0000000C	Erstellt <Datum im FILETIME (UTC)-Format>
Last saved Time/Date	0x0000000D	Letzte Änderung <Datum wird im FILETIME (UTC)-Format>
Number of Pages Number of Words Number of Characters	0x0000000E 0x0000000F 0x00000010	Anzahl von Seiten, Wörter und Zeichen im Dokument
Name of Creating Application	0x00000012	Bezeichnung des Programms, mit dem das Dokument erzeugt worden ist
Security	0x00000013	Sicherheitseinstellungen des Dokuments, zB Passwortschutz

Abbildung 3: Felder im Summary Information Property Set

3.2.2. Autor und Initialen

Jedes neu erstellte Dokument enthält die bei der Office-Installation eingegebenen Werte für den Namen und die Initialen des Benutzers als Autorenangabe. Die Benutzerinformationen sind in der Registry abgespeichert.

Ab Office 2003 können die Werte für den Benutzernamen und die Initialen über die allgemeinen Optionen von Word geändert werden. Bereits vorhandene Dokumente enthalten jedoch möglicherweise bereits die bei der Office-Installation eingegebenen Werte.[11]

3.2.3. Vorherige Autoren

Das Programm Microsoft Office Word (2000, Anmerkung SV) speichert die Namen der letzten 10 Personen, die an einem Dokument gearbeitet haben. Dies ist eine automatische Funktion, die nicht deaktiviert werden kann.[12] Benutzer können jedoch die Namen der letzten 10 Autoren aus einem Dokument entfernen, indem sie das Dokument in einem Format speichern, das keine solchen Informationen enthält (zB RTF oder TXT).

Spätere Versionen von Word speichern diese Informationen zu den Bearbeitern des Dokuments aus forensischer Sicht leider nicht mehr.

3.2.4. Zuletzt geändert von

Bei einer Überarbeitung des Dokuments speichert das Programm in diesem Feld den Namen des letzten Benutzers. In diesem Zusammenhang ist zu beachten, dass mit dem Speichervorgang keine vorhergegangene Veränderung des Dokumenteninhalts verbunden sein muss.

3.2.5. Dokumentvorlage

Alle in Word erstellten Dokumente basieren auf einer Vorlage. In der Standardeinstellung ist dies die Vorlagedatei Normal.dot im Ordner Vorlagen. Sie können jedoch auch Dokumente erstellen, die auf anderen Vorlagen an ande-

Dateioperationen	Überarbeitungsnummer (REVNUM)
Erstellen und Speichern eines auf die Standarddokumentenvorlage beruhenden neuen Dokuments	REVUM = 1
Erneutes Öffnen und Speichern der Datei, unabhängig von einer tatsächlichen Überarbeitung des Dokuments	REVNUM wird um 1 erhöht
Öffnen des Dokuments und Schließen (ohne zu speichern)	REVNUM bleibt unverändert
Datei kopieren oder verschieben	REVNUM bleibt unverändert
Datei aus einem E-Mail Anhang speichern und öffnen	REVNUM bleibt unverändert Anmerkung: Beim Öffnen erscheint Hinweis auf Internet-Herkunft und die Bearbeitung ist gesperrt
Öffnen einer Datei (mit RevNum 5) und Speichern unter ... einen neuen Namen und öffnen	REVNUM = 2
Öffnen einer Datei im docx-Format und Speichern unter ... gleichen Namen im doc-Format	REVNUM = 2

Abbildung 4: REVNUM-Verhalten bei verschiedenen Dateioperationen

ren Speicherorten basieren. Pfad und Name dieser Vorlage werden in den Eigenschaften des Dokuments gespeichert.[13]

3.2.6. Überarbeitungsnummer

Die Überarbeitungsnummer {REVNUM} gibt an, wie oft ein Dokument gespeichert wurde.[14]

In einem Versuch hat der Sachverständige folgendes Verhalten von Microsoft Word bei der Verwaltung der Überarbeitungsnummer feststellen können.[15]

Anzumerken ist, dass sich die Überarbeitungsnummer bei jedem Speichern unabhängig von einer etwaigen Modifikation des Inhalts um den Wert 1 erhöht.

3.3. Datums- und Uhrzeitinformationen

Bei der forensischen Auswertung von Office-Dateien spielen meist Fragen nach dem Zeitpunkt der Erstellung oder letzten Speicherung eines Word-Dokuments eine wichtige Rolle. Zeitinformationen zu sichergestellten Word-Dokumenten können aus verschiedenen Quellen gewonnen werden.

3.3.1. Zeitstempel des Dateisystems

Windows Dateisysteme wie NTFS speichern in der Regel nachfolgende Daten und Uhrzeiten zu einer vom Dateisystem verwalteten Datei:

- Zeitpunkt der Erstellung der Datei;
- Zeitpunkt des letzten (lesenden) Zugriffs auf die Datei;
- Zeitpunkt der letzten Modifikation der Datei.

3.3.2. Zeitstempel von Microsoft Word

Unabhängig von den Zeitstempeln des Dateisystems speichert Microsoft Word nachfolgende Daten und Uhrzeiten in Zusammenhang mit einem Dokument:

- Erstellzeitpunkt des Dokuments;
- Zeitpunkt der letzten Speicherung des Dokuments;
- Zeitpunkt des letzten Ausdrucks des Dokuments.

Die Zeitstempel können über Microsoft-eigene Feldfunktionen ausgelesen werden und stehen auch bei wiederhergestellten Dateien zur Verfügung, bei denen Zeitstempelinformationen aus dem Dateisystem meist nicht mehr nutzbar sind.

4. Auswertung von Metadaten

Die nachfolgenden Zeilen beschäftigen sich mit Methoden zur Auswertung der Metadaten in Word-Dokumenten.

4.1. Eigenschaften in Word

Die nächstliegende Methode zur Auswertung der Metadaten in Word ist der Rückgriff auf die entsprechenden Funktionen des Programms selbst.

Die zu untersuchende Dokumentendatei ist dabei vor einem etwaigen Schreibzugriff zu schützen (zB durch Brennen der Datei auf eine CD). Im nächsten Schritt können über die Feldfunktionen die gewünschten Dokumenteninformationen ausgelesen bzw dokumentiert werden.

Die beschriebene Methode ist allerdings auf Metadaten, die über Feldfunktionen ausgelesen werden können, beschränkt.

4.2. Extraktion der Metadaten per Spezialprogramm

Eine weitere Möglichkeit ist die Extraktion der vorhandenen Metadaten per Spezialprogramm aus dem sichergestellten Word-Dokument. Die Palette der Programme reicht vom einfachen Textextrahierungsprogramm wie Strings,[16] welche alle ASCII- und UNICODE-Zeichen aus dem Dokument extrahieren, bis hin zu gängigen Forensik-Programmen wie EnCase, X-Ways Forensics oder dem Forensic

Toolkit, welche Metadaten aus den untersuchten Dokumenten filtern und anzeigen.[17]

4.3. Auswertung per Hex-Editor

In der Regel wird man bei der Auswertung der Metadaten auf programmeigene Funktionen oder Spezialprogramme zurückgreifen. Bei manipulierten oder nur teilweise rekonstruierten Dokumentdateien stellt der Einsatz eines Hex-Editors ein unverzichtbares Instrument dar, um vorhandene Metadaten sicherstellen und auswerten zu können. Die Analyse von Word-Dokumenten per Hex-Editor setzt ausreichende Kenntnisse über den strukturellen Aufbau und Inhalt des vorhandenen Dateiformats voraus.

5. Metadaten in Alternate Data Streams

Das Windows-Dateisystem NTFS nutzt sogenannte Alternate Data Streams (ADS), um Daten – für den Benutzer unsichtbar – fest an eine Datei zu binden. Betriebssysteme wie unterschiedliche Windows-Versionen verwenden Alternate Data Streams unter anderem zur Speicherung von dateispezifischen Metadaten.

So wird unter Windows ein als Zone Identifier bezeichnetes Datum gespeichert, das es ermöglicht, Dateien zu erkennen, die aus dem Internet heruntergeladen wurden. Die entsprechenden Informationen werden beim Herunterladen der Datei über den Browser (zB Internet Explorer oder Mozilla Firefox) dem ADS hinzugefügt.[18] Über den Zone Identifier mit dem Wert 3 ist es gegebenenfalls möglich, festzustellen, dass eine Datei ursprünglich über das Internet zB von einer Website bezogen oder per E-Mail empfangen wurde.

6. Fazit

Metadaten in Word-Dokumenten stellen eine wichtige Informationsquelle zur Beantwortung von forensischen Fragestellungen im Zusammenhang mit der Erstellung und Bearbeitung von Word-Dokumenten dar. Besonderes Augenmerk gilt jenen Daten, die durch das Programm automatisch erzeugt werden und durch den Benutzer oder die Benutzerin nur bedingt modifiziert werden können. Existenz und Beschaffenheit der Metadaten-Inhalte sind abhängig vom Dateityp und Version des jeweiligen Microsoft Word Programms.

Anmerkungen:

1 *Schwichtenberg*, COM-Komponenten-Handbuch: Systemprogrammierung und Scripting mit COM-Komponenten (2001) 102.

2 Seite „Grundlegendes zum MS-DOC-Binärdateiformat von Word“, URL: http://msdn.microsoft.com/de-de/library/office/gg615596%28v=office.14%29.aspx#UnderstandMS-DOC_Overview.

3 Seite „FibBase“, URL: http://msdn.microsoft.com/en-us/library/dd944620%28v=office.12%29.aspx.

4 Seite „Minimieren von Metadaten in Microsoft Word 2000-Dokumenten“, URL: http://support.microsoft.com/kb/237361/de.

5 Darstellung einer MS-DOC-Datei im Microsoft-Tool OffVis. Das Tool stellt die analysierte Datei in Hex-Code dar und bildet zusätzlich die innere Struktur der eingebundenen Elemente ab.

6 Seite „Minimieren von Metadaten in Microsoft Word 2000-Dokumenten“, URL: http://support.microsoft.com/kb/237361/de.

7 White Paper „Meta-Daten in Microsoft Office und in PDF-Dokumenten“, URL: http://soft-xpansion.eu/files/cc/Metadaten.pdf.

8 Bis einschließlich Word 2003 können Benutzer die Dokumenteigenschaften über Datei/Eigenschaften auf der Registerkarte Anpassen näher anschauen und modifizieren.

9 White Paper „Meta-Daten in Microsoft Office und in PDF-Dokumenten“, URL: http://soft-xpansion.eu/files/cc/Metadaten.pdf.

10 Seite: „PIDSI“, URL: http://msdn.microsoft.com/en-us/library/dd925819%28v=office.12%29.aspx.

11 Seite „Minimieren von Metadaten in Microsoft Word 2000-Dokumenten“, URL: http://support.microsoft.com/kb/237361/de.

12 Seite „Minimieren von Metadaten in Microsoft Word 2000-Dokumenten“, URL: http://support.microsoft.com/kb/237361/de.

13 Seite „Minimieren von Metadaten in Microsoft Word 2000-Dokumenten“, URL: http://support.microsoft.com/kb/237361/de.

14 Seite „Feldfunktionen: RevNum-Feld“, URL: http://office.microsoft.com/de-at/word-help/feldfunktionen-revnum-feld-HP005186188.aspx.

15 Der Versuch erfolgte mit MS Office 2010.

16 Das Programm Strings ist Bestandteil der Microsoft Sysinternals Tool Suite. Downlaod unter http://technet.microsoft.com/en-us/sysinternals/bb842062.aspx.

17 Neben den genannten Programmen existieren im Internet noch eine Reihe von Share- und Freeware-Tools zur Extraktion von Metadaten unter anderem DocSrubber, ExifTool oder Forensic FOCA.

18 Seite: „Alternativer Datenstrom“, URL: http://de.wikipedia.org/wiki/Alternativer_Datenstrom.

Korrespondenz:
Ing. Mag. Horst Greifeneder
Schenkelbachweg 32, A-4600 Wels
Tel.: 07242 / 77715
Fax: 07242 / 77716
E-Mail: office@fds.at

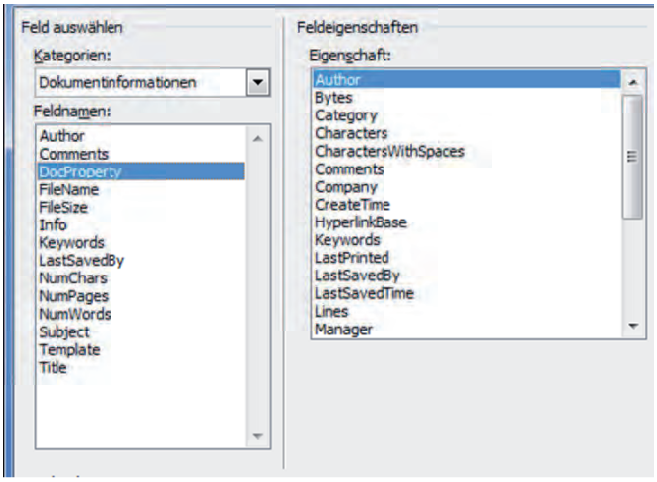

Abbildung 5: Feldfunktionen für Dokumentinformationen in Microsoft Word