



Unerlaubtes Outsourcing von Studienleistungen entdecken: Was DOCX-Dateien verraten

Oliver Trevisiol

07.06.23

Outsourcing an Dritte, wie z.B. Contract-Cheating-Anbieter oder die nicht offen gelegte Textproduktion durch Künstliche Intelligenz, gelten als schwierig zu entdecken. Textdateien im DOCX- oder ODT-Format können dazu genutzt werden, gezielt nach Hinweisen auf unerlaubtes Outsourcing zu suchen. Selbstverständlich ist es Menschen mit entsprechendem technischen Wissen möglich, Informationen an den hier vorgestellten Stellen zu ändern oder zu löschen. In sehr vielen Fällen wird aber nicht von dieser Möglichkeit Gebrauch gemacht. Andererseits ist es auch möglich und vollkommen legitim, für die Abgabeversion eines Textes eine neue und saubere Datei zu erstellen. Ein einzelner Hinweis auf potentielles Outsourcing ist noch kein Beweis, aber mehrere können ein erhellendes Gesamtbild ergeben. Weitere Hinweise können z.B. die Informationen in den Eigenschaften von elektronisch eingereichten Dokumenten liefern.

Hinweis zur weiteren Verwendung:



Sie können dieses Material weiternutzen. Es wird unter der Creative Commons Lizenz CC-BY 4.0 zur Verfügung gestellt.

Das bedeutet unter anderem, dass Sie das Material weiterverbreiten, die Inhalte für eigene Materialien als Grundlage nehmen oder es weiterentwickeln dürfen. Bedingung dafür ist, dass Sie das Projekt in angemessener Weise nennen und deutlich machen, ob Sie Material verändert oder unverändert übernommen haben.

Detaillierte Informationen zur Lizenz finden Sie unter

<https://creativecommons.org/licenses/by/4.0>

Unerlaubtes Outsourcing von Studienleistungen entdecken: Was DOCX-Dateien verraten

DOCX mit ZIP öffnen

DOCX- oder ODT-Dateien sind eigentlich zip-archivierte Containerdateien, die XML Dateien enthalten. Solche Archivdateien können mit Packprogrammen wie z.B. 7-Zip geöffnet und entpackt werden. Oft wird diese Funktion direkt im Windows Dateimanager nach einem Klick mit der rechten Maustaste auf die DOCX-Datei angeboten. Die darin enthaltenen XML Dateien sind zwar für das menschliche Auge nicht besonders lesefreundlich, können aber mit jedem Texteditor angezeigt werden. Sie enthalten zusätzliche Informationen, die Hinweise auf unerlaubtes Outsourcing geben können.

Beispiel für den Inhalt des Unterordners Word einer entpackten DOCX-Datei :

Name	Größe	Gepackte Größe	Geändert am
 embeddings	89 889	89 889	
 media	4 287 479	362 633	
 theme	6 793	1 571	
 _rels	1 244	316	
 document.xml	19 999	2 729	1980-01-01 00:00
 fontTable.xml	1 340	441	1980-01-01 00:00
 settings.xml	2 822	1 037	1980-01-01 00:00
 styles.xml	28 769	2 887	1980-01-01 00:00
 webSettings.xml	576	280	1980-01-01 00:00

Embeddings

Der Ordner *embeddings* existiert, wenn das Word Dokument eingebettete Dateien wie z.B. Excel Tabellen beinhaltet. Diese lassen sich wiederum öffnen, sodass über Eigenschaften Informationen aus den Metadaten der XLSX-Datei entnommen werden können (vgl. hierzu auch das Informationsblatt zu Metadaten in Dateien).

Media

Im Ordner *media* finden Sie im Word Dokument enthaltene Bilder und Grafiken etc. Die Dateien können ebenfalls geöffnet und die Metadaten untersucht werden. Manchmal liegen Bilddateien in noch nicht zugeschnittenem Zustand vor, die weitere Informationen wie z.B. Links zu Fundstellen oder Speicherorten bereitstellen.

settings.xml

Word vergibt allen Bearbeitungen eine Session ID (w:rsid). Eine Session ist jeweils der Bearbeitungszeitraum zwischen zwei Speichervorgängen. Alle Bearbeitungen innerhalb einer Session erhalten die gleiche Session ID. Die IDs werden zufällig und in aufsteigender Reihenfolge vergeben. Die erste Bearbeitungssession eines Dokuments ist als w:rsidRoot zu erkennen. In der Datei *settings.xml* sind alle Session IDs aufgeführt. Haben zwei Dokumente eine Reihenfolge von Session IDs gemeinsam, dann kann auf einen gemeinsamen Ursprung der Dokumente geschlossen werden.

document.xml

Die Datei *document.xml* enthält den eigentlichen Text einer DOCX-Datei. In dieser Datei sind alle Bearbeitungen mit ihrer jeweiligen Session ID gekennzeichnet. Enthält ein längerer Text nur sehr wenige Session IDs, dann ist das ein Zeichen dafür, dass nur sehr wenig und kurze Zeit mit dem Text gearbeitet wurde, z.B. weil fertiger Text mit Copy&Paste übernommen wurde.

Literatur

Crockett, Robin (2022): Presentation, Properties and Provenance. The Three Ps of Identifying Evidence of Contract Cheating in Student Assignments. In: Sarah Elaine Eaton, Guy J. Curtis, Brenda M. Stoesz, Joseph Clare, Kiata Rundle und Josh Seeland (Hg.): Contract Cheating in Higher Education. Global Perspectives on Theory, Practice, and Policy. Cham: Palgrave Macmillan, S. 169–184.

Microsoft: Rsid Class,

<https://learn.microsoft.com/en-us/dotnet/api/documentformat.openxml.wordprocessing.rsid?view=openxml-2.8.1> (Zugriff am 6.6.2023)