

Digital humanities

Cvičení: Digitální materialita

Jindřich Marek

Přehled

- toto cvičení seznamuje s konceptem digitální materiality - studiem fyzických vlastností digitálních objektů
- naučíme se dekonstruovat a analyzovat strukturu dokumentu DOCX (MS Word), pochopit jeho reprezentaci v XML a prozkoumat skrytá metadata a kódování textu

Cíle cvičení

- po dokončení tohoto cvičení budete schopni:
 - pochopit strukturu moderních dokumentů (DOCX, EPUB)
 - dekonstruovat DOCX na XML komponenty
 - analyzovat metadata a vlastnosti dokumentů
 - porozumět konceptu digitální materiality
 - provádět forenzní analýzu dokumentů

Požadované nástroje

- kancelářský program ukládající do formátu DOCX (MS Word, Libre Office nebo jiný)
- program umožňující vytvářet a rozbalovat archivy ve formátu ZIP (integrovaný v operačním systému, nebo třeba 7Zip)
- textový editor (Notepad, Visual Studio Code či jiný)
- editor pro formát EPUB (Calibre)

Dataset

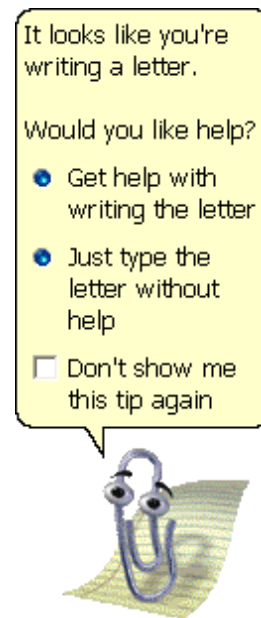
- soubor ve formátu DOCX

Postup

- rozbalení dokumentu ve formátu DOCX (unzip)
- prozkoumání struktury
- otevření document.xml v textovém editoru
- změna textového obsahu
- zabalení dokumentu zpět do formátu DOCX
- otevření změněného dokumentu
- export do formátu EPUB
 - vlastně import z DOCX v Calibre
- úprava obsahu ve formátu EPUB

Digitální materialita

- Formát DOCX, jeho struktura a obsah
 - Microsoft Office Word



文 28 languages

Read Edit View history Tools ▾

Office Open XML

- Office Open XML file formats
- Open Packaging Conventions
- Open Specification Promise
- Office Open XML software
- Comparison of Office Open XML software

V • T • E

[[edit](#)]

The OOXML Document icon, as appears on the [Microsoft OneDrive](#) web service

.docx, .docm

```
application/vnd.  
openxmlformats-  
officedocument.  
wordprocessingml.  
document[1]
```

Microsoft, Ecma, ISO, IEC

7 December 2006; 17
years ago

4th edition
26 October 2016; 7 years
ago

Document file format

XML, DOC,
WordProcessingML

ECMA-376, ISO/IEC 29500

Yes [\[4\]](#)

[edit]

Main article: [Standardization of Office Open XML](#)

Microsoft submitted initial material to [Ecma International](#) Technical Committee TC45, where it was standardized to become ECMA-376, approved in December 2006.^[12]

This standard was then fast-tracked in the [Joint Technical Committee 1](#) of ISO and IEC. After initially [failing to pass](#), an amended version of the format received the necessary votes for approval as an ISO/IEC Standard as the result of a [JTC 1](#) fast-tracking standardization process that concluded in April 2008. The resulting four-part International Standard (designated ISO/IEC 29500:2008) was published in November 2008^[14] and can be downloaded from the [ITTF](#).^[15] A technically equivalent set of texts is



 Domů

 Nový

 Otevřít

 Získat doplňky

Informace

Uložit kopii

Tisk

Sdílet

Exportovat

Převést

Zavřít

Účet

Možnosti

Informace

Můj dokument

OneDrive – Filozofická fakulta, Univerzita Karlova » Vyuka » nmgr_kurzy » digital_humanities_2024 » materialy » materialy_10



Sdílet



Kopírovat cestu



Kopírovat místní cestu



Otevřít umístění souboru



Zamknout dokument ▾

Zamknout dokument

Umožňuje určit, jaké typy změn mohou uživatelé v tomto dokumentu dělat.



Zjistit možné problémy ▾

Zkontrolovat dokument

Před publikováním tohoto souboru si uvědomte, že obsahuje:

- Vlastnosti dokumentu a jméno autora



Historie verzí

Historie verzí

Zobrazit a obnovit předchozí verze.



Spravovat dokument ▾

Spravovat dokument

 Nejsou žádné neuložené změny.

Vlastnosti ▾

Velikost	13,1 kB
Stránky	1
Slova	6
Celková doba úprav	4 minut
Název	Přidat název
Značky	Přidat značku
Komentáře	Přidat komentáře

Související data

Naposledy upraveno	Dnes, 13:37
Vytvořeno	Dnes, 13:35
Naposledy vytištěno	

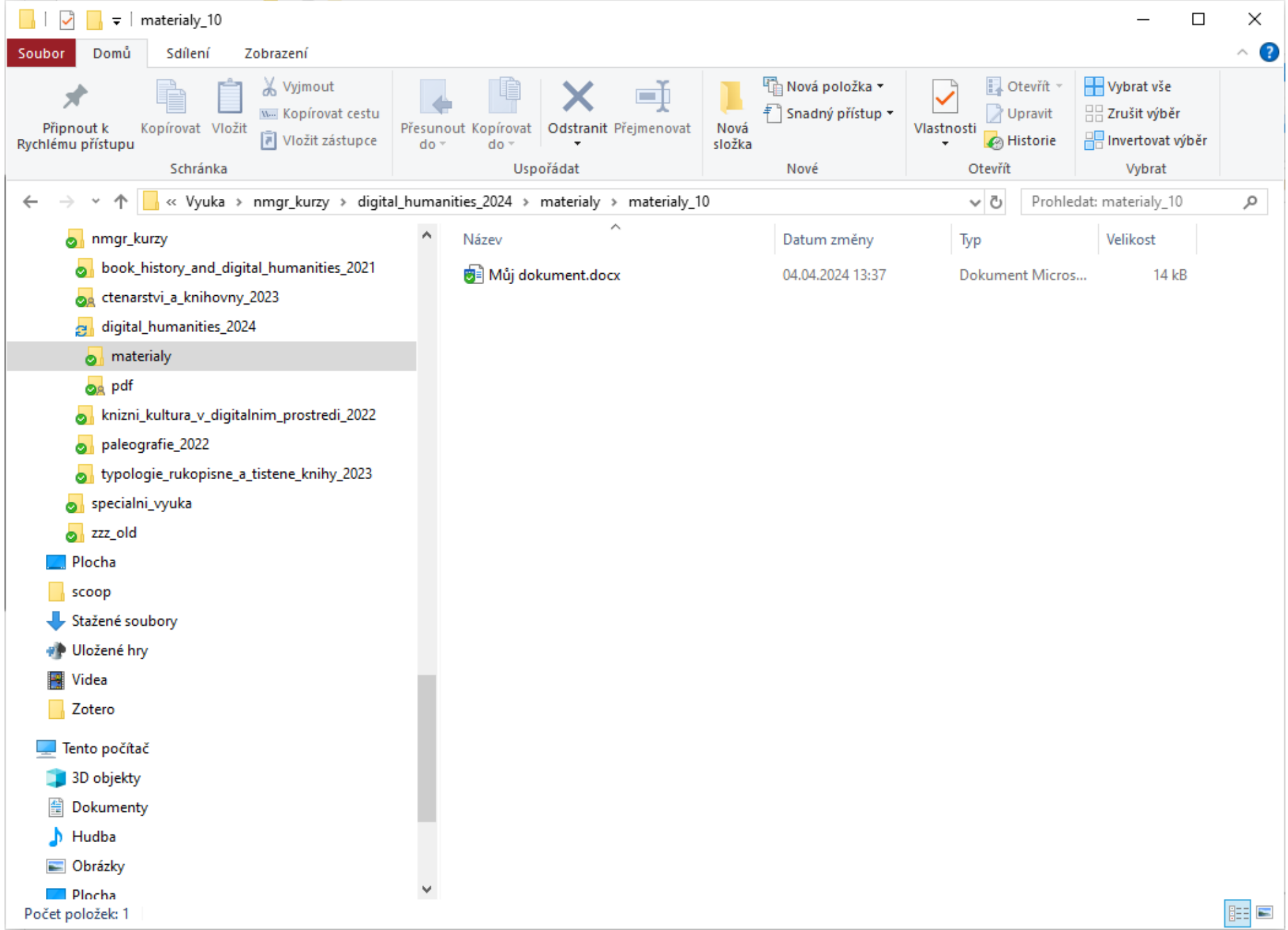
Související uživatelé

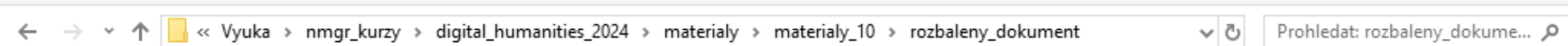
Autor	 Marek, Jindřich
	Přidat autora
Autor poslední změny	 Marek, Jindřich

Související dokumenty

 [Otevřít umístění souboru](#)

[Zobrazit všechny vlastnosti](#)





Počet položek: 4

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
```

```
<w:document xmlns:wpc="http://schemas.microsoft.com/office/word/2010/wordprocessingCanvas" xmlns:cx="http://schemas.microsoft.com/office/drawing/2014/chartex" xmlns:cx1="http://schemas.microsoft.com/office/drawing/2015/9/8/chartex"
xmlns:cx2="http://schemas.microsoft.com/office/drawing/2015/10/21/chartex" xmlns:cx3="http://schemas.microsoft.com/office/drawing/2016/5/9/chartex" xmlns:cx4="http://schemas.microsoft.com/office/drawing/2016/5/10/chartex"
xmlns:cx5="http://schemas.microsoft.com/office/drawing/2016/5/11/chartex" xmlns:cx6="http://schemas.microsoft.com/office/drawing/2016/5/12/chartex" xmlns:cx7="http://schemas.microsoft.com/office/drawing/2016/5/13/chartex"
xmlns:cx8="http://schemas.microsoft.com/office/drawing/2016/5/14/chartex" xmlns:mc="http://schemas.openxmlformats.org/markup-compatibility/2006" xmlns:aink="http://schemas.microsoft.com/office/drawing/2016/ink"
xmlns:am3d="http://schemas.microsoft.com/office/drawing/2017/model3d" xmlns:o="urn:schemas-microsoft-com:office:office" xmlns:otel="http://schemas.microsoft.com/office/2019/extlst"
xmlns:r="http://schemas.openxmlformats.org/officeDocument/2006/relationships" xmlns:m="http://schemas.openxmlformats.org/officeDocument/2006/math" xmlns:v="urn:schemas-microsoft-com:vml"
xmlns:wp14="http://schemas.microsoft.com/office/word/2010/wordprocessingDrawing" xmlns:wp="http://schemas.openxmlformats.org/drawingml/2006/wordprocessingDrawing" xmlns:w10="urn:schemas-microsoft-com:office:word"
xmlns:w="http://schemas.openxmlformats.org/wordprocessingml/2006/main" xmlns:w14="http://schemas.microsoft.com/office/word/2010/wordml" xmlns:w15="http://schemas.microsoft.com/office/word/2012/wordml"
xmlns:w16cex="http://schemas.microsoft.com/office/word/2018/wordml/cex" xmlns:w16cid="http://schemas.microsoft.com/office/word/2016/wordml/cid" xmlns:w16="http://schemas.microsoft.com/office/word/2018/wordml"
xmlns:w16sdtid="http://schemas.microsoft.com/office/word/2020/wordml/sdtdatahash" xmlns:w16se="http://schemas.microsoft.com/office/word/2015/wordml/symex" xmlns:wpg="http://schemas.microsoft.com/office/word/2010/wordprocessingGroup"
xmlns:wpi="http://schemas.microsoft.com/office/word/2010/wordprocessingInk" xmlns:wne="http://schemas.microsoft.com/office/word/2006/wordml" xmlns:wps="http://schemas.microsoft.com/office/word/2010/wordprocessingShape" mc:Ignorable="w14
w15 w16se w16cid w16 w16cex w16sdtid wp14"><w:body><w:p w14:paraId="3EEE053" w14:textId="0339DB11" w:rsidR="006279B2" w:rsidRDefault="00531798" w:rsidP="00531798"><w:pPr><w:pStyle w:val="Nadpis1"/></w:pPr><w:r><w:t>Nadpis mého
dokumentu</w:t></w:r></w:p><w:p w14:paraId="5570C200" w14:textId="56196BC2" w:rsidR="00531798" w:rsidRDefault="00531798"><w:r><w:t>Text mého dokumentu.</w:t></w:r></w:p><w:sectPr w:rsidR="00531798"><w:pgSz w:w="11906"
w:h="16838"/><w:pgMar w:top="1417" w:right="1417" w:bottom="1417" w:left="1417" w:header="708" w:footer="708" w:gutter="0"/><w:cols w:space="708"/><w:docGrid w:linePitch="360"/></w:sectPr></w:body></w:document>
```



```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<w:fonts xmlns:mc="http://schemas.openxmlformats.org/markup-compatibility/2006" xmlns:r="http://schemas.openxmlformats.org/officeDocument/2006/relationships" xmlns:w="http://schemas.openxmlformats.org/wordprocessingml/2006/main"
xmlns:w14="http://schemas.microsoft.com/office/word/2010/wordml" xmlns:w15="http://schemas.microsoft.com/office/word/2012/wordml" xmlns:w16cex="http://schemas.microsoft.com/office/word/2018/wordml/cex"
xmlns:w16cid="http://schemas.microsoft.com/office/word/2016/wordml/cid" xmlns:w16="http://schemas.microsoft.com/office/word/2018/wordml" xmlns:w16sdttdh="http://schemas.microsoft.com/office/word/2020/wordml/sdtdatahash"
xmlns:w16se="http://schemas.microsoft.com/office/word/2015/wordml/symex" mc:Ignorable="w14 w15 w16se w16cid w16 w16cex w16sdttdh"><w:font w:name="Aptos"><w:charset w:val="00"/><w:family w:val="swiss"/><w:pitch w:val="variable"/><w:sig
w:usb0="20000287" w:usb1="00000003" w:usb2="00000000" w:usb3="00000000" w:csb0="0000019F" w:csb1="00000000"/></w:font><w:font w:name="Arial"><w:panose1 w:val="020B06040202020204"/><w:charset w:val="EE"/><w:family
w:val="swiss"/><w:pitch w:val="variable"/><w:sig w:usb0="E0002EFF" w:usb1="C000785B" w:usb2="00000009" w:usb3="00000000" w:csb0="000001FF" w:csb1="00000000"/></w:font><w:font w:name="Aptos Display"><w:charset w:val="00"/><w:family
w:val="swiss"/><w:pitch w:val="variable"/><w:sig w:usb0="20000287" w:usb1="00000003" w:usb2="00000000" w:usb3="00000000" w:csb0="0000019F" w:csb1="00000000"/></w:font><w:font w:name="Times New Roman"><w:panose1
w:val="02020603050405020304"/><w:charset w:val="EE"/><w:family w:val="roman"/><w:pitch w:val="variable"/><w:sig w:usb0="E0002EFF" w:usb1="C000785B" w:usb2="00000009" w:usb3="00000000" w:csb0="000001FF"
w:csb1="00000000"/></w:font></w:fonts>
```

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
```

```
<w:document xmlns:wpc="http://schemas.microsoft.com/office/word/2010/wordprocessingCanvas" xmlns:cx="http://schemas.microsoft.com/office/drawing/2014/chartex" xmlns:cx1="http://schemas.microsoft.com/office/drawing/2015/9/8/chartex"
```

```
xmlns:cx2="http://schemas.microsoft.com/office/drawing/2015/10/21/chartex" xmlns:cx3="http://schemas.microsoft.com/office/drawing/2016/5/9/chartex" xmlns:cx4="http://schemas.microsoft.com/office/drawing/2016/5/10/chartex"
```

```
xmlns:cx5="http://schemas.microsoft.com/office/drawing/2016/5/11/chartex" xmlns:cx6="http://schemas.microsoft.com/office/drawing/2016/5/12/chartex" xmlns:cx7="http://schemas.microsoft.com/office/drawing/2016/5/13/chartex"
```

```
xmlns:cx8="http://schemas.microsoft.com/office/drawing/2016/5/14/chartex" xmlns:mc="http://schemas.openxmlformats.org/markup-compatibility/2006" xmlns:aink="http://schemas.microsoft.com/office/drawing/2016/ink"
```

```
xmlns:am3d="http://schemas.microsoft.com/office/drawing/2017/model3d" xmlns:o="urn:schemas-microsoft-com:office:office" xmlns:oel="http://schemas.microsoft.com/office/2019/extlst"
```

```
xmlns:r="http://schemas.openxmlformats.org/officeDocument/2006/relationships" xmlns:m="http://schemas.openxmlformats.org/officeDocument/2006/math" xmlns:v="urn:schemas-microsoft-com:vml"
```

```
xmlns:wp14="http://schemas.microsoft.com/office/word/2010/wordprocessingDrawing" xmlns:wp="http://schemas.openxmlformats.org/drawingml/2006/wordprocessingDrawing" xmlns:w10="urn:schemas-microsoft-com:office:word"
```

```
xmlns:w="http://schemas.openxmlformats.org/wordprocessingml/2006/main" xmlns:w14="http://schemas.microsoft.com/office/word/2010/wordml" xmlns:w15="http://schemas.microsoft.com/office/word/2012/wordml"
```

```
xmlns:w16cex="http://schemas.microsoft.com/office/word/2018/wordml/cex" xmlns:w16cid="http://schemas.microsoft.com/office/word/2016/wordml/cid" xmlns:w16="http://schemas.microsoft.com/office/word/2018/wordml"
```

```
xmlns:w16sdtid="http://schemas.microsoft.com/office/word/2020/wordml/sdtdatahash" xmlns:w16se="http://schemas.microsoft.com/office/word/2015/wordml/symex" xmlns:wpg="http://schemas.microsoft.com/office/word/2010/wordprocessingGroup"
```

```
xmlns:wpi="http://schemas.microsoft.com/office/word/2010/wordprocessingInk" xmlns:wne="http://schemas.microsoft.com/office/word/2006/wordml" xmlns:wps="http://schemas.microsoft.com/office/word/2010/wordprocessingShape" mc:Ignorable="w14
```

```
w15 w16se w16cid w16 w16cex w16sdtid wp14"><w:body><w:p w14:paraId="3EEE053" w14:textId="0339DB11" w:rsidR="006279B2" w:rsidRDefault="00531798" w:rsidP="00531798"><w:pPr><w:pStyle w:val="Nadpis1"/></w:pPr><w:r><w:t>Nadpis mého
```

```
dokumentu změněný</w:t></w:r></w:p><w:p w14:paraId="5570C200" w14:textId="56196BC2" w:rsidR="00531798" w:rsidRDefault="00531798"><w:r><w:t>Text mého dokumentu.</w:t></w:r></w:p><w:sectPr w:rsidR="00531798"><w:pgSz w:w="11906"
```

```
w:h="16838"/><w:pgMar w:top="1417" w:right="1417" w:bottom="1417" w:left="1417" w:header="708" w:footer="708" w:gutter="0"/><w:cols w:space="708"/><w:docGrid w:linePitch="360"/></w:sectPr></w:body></w:document>
```

rozbaleny_dokument

Soubor

Domů

Sdílení

Zobrazení

Připnout k Rychlému přístupu

Kopírovat

Vložit

Vymout

Kopírovat cestu

Vložit zástupce

Přesunout do

Kopírovat do

Odstranit

Přejmenovat

Nová složka

Nová položka

Snadný přístup

Vlastnosti

Otevřít

Upravit

Historie

Vybrat vše

Zrušit výběr

Invertovat výběr

Schránka

Uspořádat

Nové

Otevřít

Vybrat

Prohledat: rozbaleny_dokume...

materialy

cetba_04

cetba_09

cviceni_02

cviceni_03

cviceni_04

cviceni_05

cviceni_06

cviceni_07

cviceni_08

cviceni_09

materialy_01

materialy_02

materialy_03

materialy_04

materialy_05

materialy_06

materialy_08

materialy_10

rozbaleny_dokument

_rels

docProps

word

Název

Datum změny

Typ

Velikost

_rels

04.04.2024 13:43

Složka souborů

docProps

04.04.2024 13:43

Složka souborů

word

04.04.2024 13:43

Složka souborů

[Content_Types].xml

Microsoft Edge H...

2 kB

Můj dokument změněný.zip

04.04.2024 13:52

Komprimovaná sl...

12 kB

Počet položek: 5

rozbaleny_dokument

Soubor

Domů

Sdílení

Zobrazení

Připnout k Rychlému přístupu

Kopírovat

Vložit

Vyjmout

Kopírovat cestu

Vložit zástupce

Přesunout do

Kopírovat do

Odstranit

Přejmenovat

Nová složka

Nová položka

Snadný přístup

Vlastnosti

Otevřít

Upravit

Historie

Vybrat vše

Zrušit výběr

Invertovat výběr

Schránka

Uspořádat

Nové

Otevřít

Vybrat

Prohledat: rozbaleny_dokume...

materialy

cetba_04

cetba_09

cviceni_02

cviceni_03

cviceni_04

cviceni_05

cviceni_06

cviceni_07

cviceni_08

cviceni_09

materialy_01

materialy_02

materialy_03

materialy_04

materialy_05

materialy_06

materialy_08

materialy_10

rozbaleny_dokument

_rels

docProps

word

Název

Datum změny

Typ

Velikost

_rels

04.04.2024 13:43

Složka souborů

docProps

04.04.2024 13:43

Složka souborů

word

04.04.2024 13:43

Složka souborů

[Content_Types].xml

Microsoft Edge H...

2 kB

Můj dokument změněný.docx

04.04.2024 13:52

Dokument Micros...

12 kB

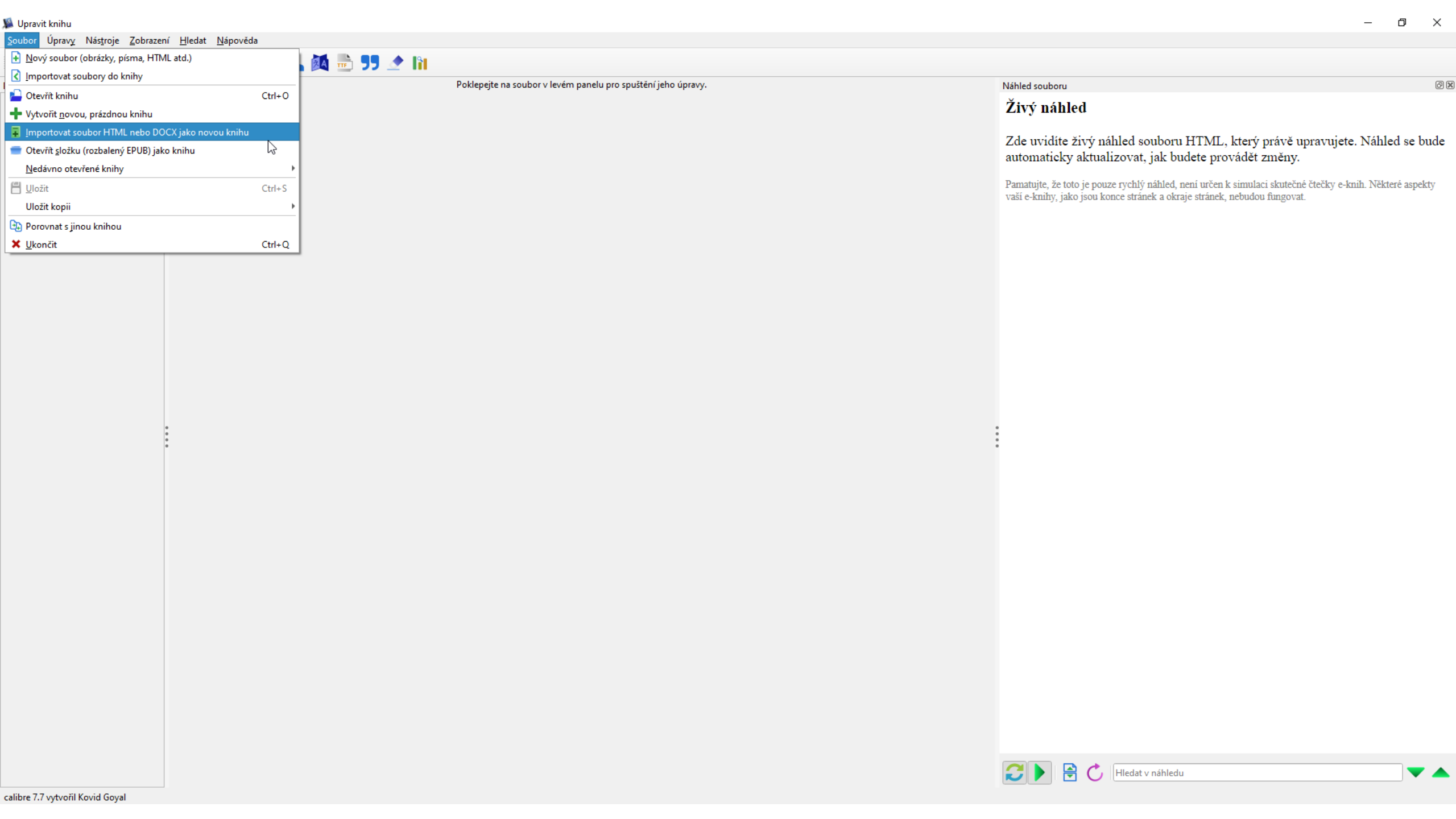
Počet položek: 5

Nadpis mého dokumentu změněný

Text mého dokumentu.

Vrstvy analýzy

- formát - DOCX je vlastně ZIP archiv
- metadata - autor, čas vytvoření, úpravy
- obsah - vlastní text strukturovaný v XML
- styly - formátování a vzhled
- vztahy - propojení mezi komponentami
- kódování - UTF-8



- Nový soubor (obrázky, písmo, HTML atd.)
- Importovat soubory do knihy
- Otevřít knihu Ctrl+O
- Vytvořit novou, prázdnou knihu
- Importovat soubor HTML nebo DOCX jako novou knihu
- Otevřít složku (rozbalený EPUB) jako knihu
- Nedávno otevřené knihy
- Uložit Ctrl+S
- Uložit kopii
- Porovnat s jinou knihou
- Ukončit Ctrl+Q

Poklepejte na soubor v levém panelu pro spuštění jeho úpravy.

Náhled souboru

Živý náhled

Zde uvidíte živý náhled souboru HTML, který právě upravujete. Náhled se bude automaticky aktualizovat, jak budete provádět změny.

Pamatujte, že toto je pouze rychlý náhled, není určen k simulaci skutečné čtečky e-knih. Některé aspekty vaší e-knihy, jako jsou konce stránek a okraje stránek, nebudou fungovat.

```
1 <?xml version='1.0' encoding='utf-8'?>
2 <html xmlns="http://www.w3.org/1999/xhtml" lang="cs" xml:lang="cs">
3   <head>
4     <title>Neznámý</title>
5     <link rel="stylesheet" type="text/css" href="docx.css"/>
6     <meta http-equiv="Content-Type" content="text/html; charset=utf-8"/>
7   </head>
8   <body>
9     <h1 class="block_1" id="toc_id_1">Nadpis mého dokumentu</h1>
10    <p class="block_2">Text mého dokumentu.</p>
11
12  </body></html>
```

Text mého dokumentu.

