

## Kódovanie textu

Text je ustálený jazykový prejav; výsledok rečovej činnosti, t. j. jazykový útvar, ktorý vznikol konkrétnou realizáciou jazykového systému.

Každý text pozostáva zo znakov, ktoré sú pospájané za sebou. Pod znakmi sa myslia všetky písmená, čísla, medzera, interpunkčné znamienka ako bodka, čiarka, otáznik, atď., ale i mnoho ďalších symbolov ako @, &, #.

Pre počítač je text ľubovoľná kombinácia všetkých spomínaných znakov. Aby sa však jednalo o textovú informáciu, musia byť znaky pospájané tak, aby mali význam.

### Kódovanie textu pomocou kódu ASCII.

Tento štandard hovorí, že na zakódovanie každého znaku sa použije 8 bitov, čo umožňuje definovať kód pre 256 znakov. Pre Slovensko sa by sa mala na kódovanie znakov textu používať medzinárodná norma **ISO 8859-2**, ktorá sa tiež nazýva **Latin 2**. Firma Microsoft však tento štandard nepoužíva, používa štandard schválený americkým úradom ANSI (American National Standards Institute), označovaný tiež **windows-1250**.

Napríklad veta "Ema má mamu." je v počítači uložená nasledovne:

```
69 109 97   32  109 225   32   109 97 109 117
E  m  a medzera m  á  medzera  m  a  m  u
```

### Kódovanie Unicode

Toto kódovanie používa 16 bitov na zakódovanie jedného znaku, čo umožňuje zakódovať 65536 možných znakov. Tento počet znakov umožňuje zakódovať znaky všetkých abecied pomocou jednej medzinárodnej tabuľky. Tento spôsob kódovania používa i kancelársky balík MS Office. Toto kódovanie zabezpečuje, že ten istý znak má rovnaký kód v každej krajine i na každom type počítača.

Veta "Ema má mamu" je pomocou kódovania UNICODE v počítači uložená nasledovne:

```
69 00 109 00 97 00 32 00 109 00 225 00 32 00 109 00 97 00 109 00 117 00
E    m    a  medzera  m    á  medzera  m    a    m    u
```

### Textový dokument

Text sa do počítača zapisuje pomocou textového editora. Text, ktorý do tohto editora napíšeme, môžeme uložiť ako textový dokument. Okrem textových znakov však textový dokument obsahuje i formátovacie znaky, pomocou ktorých môžeme v dokumente nastaviť napríklad veľkosť, typ, hrúbku, farbu a ďalšie atribúty písma. Podľa schopností textových editorov môžeme text uložiť dvoma spôsobmi:

1. Jednoduchý textový dokument
2. Formátovaný textový dokument

## Jednoduchý dokument

Jednoduchý dokument pozostáva z textových znakov, v ktorých môžu byť uložené znaky konca riadka a znaky tabulátora a ďalšie znaky, ktoré majú kód 0 až 31 v ASCII tabuľke. Na konci dokumentu je vložený špeciálny znak, ktorý dokument ukončuje. Tieto znaky do dokumentu vkladá textový editor. Typickým príkladom takéhoto editora je Poznámkový blok, ktorý takýto jednoduchý textový dokument ukladá ako súbor s príponou "txt".

## Formátovaný dokument

Formátovaný textový dokument môže obsahovať viacero typov písma, znaky rôznej veľkosti, hrúbky, šírky či farby. Do textu môžu byť vložené rôzne ďalšie objekty ako obrázky, tabuľky, rovnice a podobne. Textový editor, ktorý umožňuje uloženie formátovaného dokumentu musí zabezpečiť, aby sa v dokumente všetky tieto atribúty textu uložili spolu s textom. Tento problém každý textový editor rieši iným spôsobom. Typickým príkladom tohto editora je editor MS Word, ktorý je súčasťou kancelárskeho balíka MS Office. Word svoje formátované dokumenty ukladá do súboru s príponou "doc". Ďalším editorom, ktorý je čím ďalej tým viac používaný, je textový editor balíka Open Office. Tento editor ukladá svoje textové dokumenty do súboru s koncovkou "sxw". Oba tieto spôsoby uloženia sú odlišné. Aby bolo možné prenášať textové dokumenty z jedného programu do druhého, vznikla snaha o vytvorenie štandardu i pre uloženie formátovaného textového dokumentu. Tento formát sa volá XML a mali by ho začať podporovať všetci výrobcovia textových editorov.

Veta "Ema má mamu.", napísaná písmom Times New Roman, veľkosti 12 je vo formáte XML uložená takto:

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<?mso-application progid="Word.Document"?>
<w:wordDocument xmlns:w="http://schemas.microsoft.com/office/word/2003/wordml"
  xmlns:v="urn:schemas-microsoft-com:vml" xmlns:w10="urn:schemas-microsoft-com:office:word"
  xmlns:sl="http://schemas.microsoft.com/schemaLibrary/2003/core"
  xmlns:aml="http://schemas.microsoft.com/aml/2001/core"
  xmlns:wx="http://schemas.microsoft.com/office/word/2003/auxHint"
  xmlns:o="urn:schemas-microsoft-com:office:office"
  xmlns:dt="uuid:C2F41010-65B3-11d1-A29F-00AA00C14882"
  w:macrosPresent="no" w:embeddedObjPresent="no" w:ocxPresent="no" xml:space="preserve">
<o:DocumentProperties>
  <o:Title>Ema má mamu</o:Title>
  <o:Author>Peter Mráz</o:Author>
  <o:LastAuthor>Peter Mráz</o:LastAuthor>
  <o:Revision>1</o:Revision>
  <o:TotalTime>0</o:TotalTime>
  <o:Created>2006-04-05T20:18:00Z</o:Created>
  <o:LastSaved>2006-04-05T20:18:00Z</o:LastSaved>
  <o:Pages>1</o:Pages>
  <o:Words>1</o:Words>
  <o:Characters>12</o:Characters>
  <o:Lines>1</o:Lines>
  <o:Paragraphs>1</o:Paragraphs>
  <o:CharactersWithSpaces>12</o:CharactersWithSpaces>
  <o:Version>11.6568</o:Version>
```

```
</o:DocumentProperties>
<w:fonts>
  <w:defaultFonts w:ascii="Times New Roman" w:fareast="Times New Roman"
    w:h-ansi="Times New Roman" w:cs="Times New Roman"/>
</w:fonts>
<w:styles>
  <w:versionOfBuiltInStylenames w:val="4"/>
  <w:latentStyles w:defLockedState="off" w:latentStyleCount="156"/>
  <w:style w:type="paragraph" w:default="on" w:styleId="Normlny"><w:name w:val="Normal"/>
  <wx:uiName wx:val="Normálny"/>
  <w:rPr>
    <wx:font wx:val="Times New Roman"/>
    <w:sz w:val="24"/><w:sz-cs w:val="24"/>
    <w:lang w:val="SK" w:fareast="SK" w:bidirectional="AR-SA"/>
  </w:rPr>
</w:style>
  <w:style w:type="character" w:default="on" w:styleId="Predvolenpismoodseku">
  <w:name w:val="Default Paragraph Font"/>
  <wx:uiName wx:val="Predvolené písmo odseku"/>
  <w:semiHidden/>
</w:style>
  <w:style w:type="table" w:default="on" w:styleId="Normlnatabuka">
  <w:name w:val="Normal Table"/>
  <wx:uiName wx:val="Normálna tabuľka"/>
  <w:semiHidden/>
  <w:rPr>
    <wx:font wx:val="Times New Roman"/>
  </w:rPr>
  <w:tblPr>
    <w:tblInd w:w="0" w:type="dxa"/>
    <w:tblCellMar>
      <w:top w:w="0" w:type="dxa"/>
      <w:left w:w="108" w:type="dxa"/>
      <w:bottom w:w="0" w:type="dxa"/>
      <w:right w:w="108" w:type="dxa"/>
    </w:tblCellMar>
  </w:tblPr>
</w:style>
  <w:style w:type="list" w:default="on" w:styleId="Bezzoznamu">
  <w:name w:val="No List"/>
  <wx:uiName wx:val="Bez zoznamu"/>
  <w:semiHidden/>
</w:style>
</w:styles>
<w:docPr>
  <w:view w:val="print"/>
  <w:zoom w:percent="75"/>
  <w:doNotEmbedSystemFonts/>
  <w:proofState w:spelling="clean" w:grammar="clean"/>
  <w:attachedTemplate w:val=""/><w:defaultTabStop w:val="708"/>
  <w:hyphenationZone w:val="425"/>
  <w:punctuationKerning/>
```

```

<w:characterSpacingControl w:val="DontCompress" />
<w:optimizeForBrowser />
<w:validateAgainstSchema />
<w:saveInvalidXML w:val="off" />
<w:ignoreMixedContent w:val="off" />
<w:alwaysShowPlaceholderText w:val="off" />
<w:compat>
  <w:breakWrappedTables />
  <w:snapToGridInCell />
  <w:wrapTextWithPunct />
  <w:useAsianBreakRules />
  <w:dontGrowAutofit />
</w:compat>
</w:docPr>
<w:body>
  <wx:sect>
    <w:p>
      <w:r>
        <w:t>Ema má mamu.</w:t>
      </w:r>
    </w:p>
  <w:sectPr>
    <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:line-pitch="360" />
  </w:sectPr>
</wx:sect>
</w:body>
</w:wordDocument>

```

Na príklade je vidno, že dokument obsahuje oveľa viac informácií ako len jednoduchý text "Ema má mamu." Preto sa dokumenty v tomto formáte budú komprimovať (zmenšovať) pomocou kompresie Zip, aby dokument nezaberal tak veľa pamäte počítača.

**Zdroj:** <http://maturitazinf.mrazovci.eu/>