# Datové formáty

**Datový formát** je způsob, jakým jsou (zpravidla v nějakém souboru) organizována data, tj. veškeré obsažené informace a způsob reprezentace dat a jejich následná interpretace

## Bit

* Základní jednotka informace
* Nabývá hodnot 0 a 1 (true/false, ano/ne, vede/nevede…)

## Byte

* „Bajt“
* Byte je 8 bitů pohromadě – v mnohých standardech se proto používá označení **oktet**
* Rozsah je od 00000000-11111111, takže má kombinací
* Může tedy reprezentovat **celé číslo od 0 – 255** nebo **jeden znak**

## Znakové sady

### ASCII (American Standard Code for Information Interchange)

* Jeden bajt umožňuje uložit jeden znak
* Základní znaková sada může tedy mít **256 znaků**
* Původně byl sedmibitový (128 znaků), ale byl rozšířen na 8 bitů kvůli potřebám dalších jazyků
* Písmena, číslice, jiné tisknutelné znaky a netisknutelné řídící kódy

### Unicode

* Standard, který reprezentuje, sjednocuje, kóduje většinu jazyků v textové podobě na světě
* **Obsahuje zhruba 120000 znaků**
* **Základem je ASCII**
* **Znaky Unicode** – Jednotnost, Univerzálnost, Jednoznačnost
* Poslední verze Unicode 8.0 je z června 2015
* Na Unicode je založen kancelářský balík Microsoft Office od verze 97
* Unicode je znakovou sadou pro HTML od verze 4.0 a XML dokumenty

## Kódování

Počítač chápe každé písmenko jako číslo od 0 do 255 (jeden bajt), proto je potřeba znaky kódovat

Čeština jako jazyk je zakódovaná primárně skrze – **UTF8, Windows1250** a **ISO 8859-2**

### UTF-8

* Způsob kódování řetězců **znaků Unicode** do sekvencí bajtů.
* Navrženo pro zpětnou kompatibilitu s ASCII
* Většina distribucí Linuxu má nativně UTF-8
* Také většina WWW prohlížečů má defaultní UTF-8

### Windows 1250

* Použit v operačních systémech Windows pro střední Evropu (Čeština, Polština, Slovenština, …)
* Velmi podobný **standardu ISO8859-2**
* Písmo latinka

**ISO 8859-2**

* Kódování na Unixu a na Linuxu, ale i v mnoha windows programech
* Někdy se označuje jako **Latin 2, ISO Latin 2.**
* Microsoft jej nazývá takto: **"Středoevropské jazyky (ISO)".**

### Kód Kamenických – staré kódování

* Pro Češtinu a Slovenštinu
* **Použit v MS-DOS**
* Upravená sada CP 437
* Název KEYBCS2

## Datové formáty souborů

|  |  |  |
| --- | --- | --- |
| **Typ** | **Přípona** | **Funkce** |
| Spustitelný | EXE, COM, DEB, JAR | Slouží ke spuštění aplikace |
| Objektový | OBJ, O | Přeložený nelinkovaný program |
| Zdrojový | C, ASM, JAVA | Zdrojový kód v programovacím jazyce |
| Dávkový | BAT, SH | Příkazy pro interpretr |
| Text | TXT | Textová data |
| Knihovny | LIB, DLL | Knihovny OS |
| Textový procesor | RTF, DOCX, ODT | Různé formáty textových procesorů |
| Grafický | GIF, JPG, PNG | Uchovávání obrázků |
| Zvukový | FLAC, MP3, WAV | Uchovávání zvukové stopy |
| Video | MKV, AVI, FLV | Uchovávání videostopy |
| Archivní | ZIP, RAR, TAR | Archivující soubory |

## PostScript

* 1985 Adobe System Incorporated
* Programovací jazyk určený ke grafickému popisu tisknutelných dokumentů
* Nezávislý na tiskovém zařízení
* Standard pro dražší tiskárny
* Přípona .ps

## TrueType

* Je standard pro popis vektorových počítačových písem
* Vyvinutý koncem 80. let **společností Apple**
* Konkurent fontů Adobe Type 1, v jazyce PostScript
* Písma TrueType v operačním systému Windows mají příponu TTF

## Open Type

* Standard pro popis vektorových počítačových písem – Fontů
* Vyvinutý **společností Microsoft**
* **Nástupce standardu TrueType**, ze kterého bere jenom ty nejlepší vlastnosti
* Později se k práci na standardu připojil i největší rival – společnost Adobe Systems
* Znaky mohou být popsány pomocí zvyklostí TrueType nebo PostScript
* Je založen na Unicode, z čehož vyplývá méně problémů s lokalizací písem
* Jeden soubor může popisovat až 65 536 znaků
* Přenositelný mezi systémy Windows, Apple a Unix
* Má lepší podporu pro typografické speciality

## Portable Document Format (PDF)

* Vyvinutý firmou Adobe pro ukládání dokumentů nezávisle na SW a HW
* PDF může obsahovat obrázky i texty
* Největší výhodou je zajištění stejného zobrazení na všech zařízeních
* Je založen na jazyce PostScript
* Schopnost vkládat do dokumentu fonty

## Open Document Format (ODF)

* Formát pro ukládání textových dokumentů, prezentací, tabulek, grafů a databází
* Využívám převážně aplikacemi typu OpenOffice a LibreOffice
* Založen na XML
* Alternativa pro uzavřené formáty X Microsoft Office (\*.doc, \*.xls a \*.ppt)

## Extensible Markup Langueage (XML)

* Vyvinut a standardizován konsorciem W3C
* Umožňuje vytváření konkrétních značkovacích jazyků určených pro účely a různé typy dat
* Využívá tzv. tagy a atributy ve tvaru (jméno=„hodnota")

## Souborový systém

* Způsob uchovávání a organizace souborů na paměťovém médiu
* Nejznámější zástupci - NTFS, FAT(32), exFAT, EXT3, EXT4, HFS+ (Mac OS X), UDF (DVD)

## Komprese

* Jedná se o zpracování dat za účelem zmenšit jejich objem, při zachování informací
* Ztrátová komprese - některé informace jsou vypuštěny a nelze je zpět zrekonstruovat (MP3)
* Bezztrátová komprese - není tak účinná jako ztrátová, ale lze z ní zrekonstruovat původní soubor

## Kibi

* ISO/IEC standard,
* Jedná se o předponu pro skupinu digitálních informací (1 Kibibajt má 1024 Bajtů)

## Atributy souborů

* A -Archivace, H -Skrytý, S -Systémový, R - Pouze pro čtení

## Vlastnosti souborů

* Název, autor, velikost souboru, datum vytvoření, ...