

# Kódování znaků v Javě

## Table of Contents

Kódování znaků a národní prostředí .....	1
Volba kódování .....	1
Implicitní kódování .....	1
Řešení: implicitně UTF-8 .....	2
Zjištění kódování [1/2] .....	2
Zjištění kódování [2/2] .....	2

## Kódování znaků a národní prostředí

### Locale

národní prostředí, zahrnuje abecedu, kódování, zápis času/data, měny a národní jazyk

### Charset

množina dostupných znaků, které program (typicky v řetězcích) může využít

### Encoding

kódování znaků na vnějších zařízeních

## Volba kódování

- U mnoha (většiny) I/O operací a proudů lze při vytvoření navolit, které kódování se předpokládá nebo chce.
- Nebo ponechat kódování výchozí (system-default) převzaté ze systému, kde aplikace (JVM) běží.

## Implicitní kódování

- V některých případech je výchozí kódování nastaveno napevno bez ohledu na systémová nastavení.
- Některé metody JDK neumožňují zadání znakové sady a vždy předpokládají "výchozí" znakovou sadu UTF-8 pouze pro danou metodu a bez ohledu na lokalizaci nebo konfiguraci systému.
- Většina moderních systémů má UTF-8 stejně jako výchozí.
- Např. metody utilitní třídy `Files` vytvořené kvůli jednoduchosti použití:
  - `Files.newBufferedReader(Path)`,
  - `Files.newBufferedWriter(Path, OpenOption...)`,
  - `Files.readAllLines(Path)`,
  - `Files.write(Path, Iterable, OpenOption...)` či

- `Files.lines(Path)`.

## Řešení: implicitně UTF-8

- Od Javy 18 je UTF-8 výchozí znakovou sadu standardních rozhraní API Javy:

Díky této změně se budou rozhraní API závislá na výchozí znakové sadě chovat konzistentně ve všech implementacích, operačních systémech, místních jazycích a konfiguracích.

-- <https://openjdk.org/jeps/400>

## Zjištění kódování [1/2]

```
out.println("Default Locale: " + Locale.getDefault());
out.println("Default Charset: " + Charset.defaultCharset());
out.println("file.encoding; " + System.getProperty("file.encoding"));
out.println("sun.jnu.encoding: " + System.getProperty("sun.jnu.encoding"));
out.println("Default Encoding: " + getEncoding());
```



Výše uvedený kód bude fungovat, deklarujeme-li na začátku import objektu `out`:  
`import static java.lang.System.out;`

## Zjištění kódování [2/2]

Vypíše se:

```
Default Locale: cs_CZ
Default Charset: UTF-8
file.encoding; UTF-8
sun.jnu.encoding: UTF-8
Default Encoding: UTF8
```