

Кодировки на текстови файлове в Linux

1. Unicode кодировки (препоръчителни)

Поддържат широк набор от символи и езици и се използват във всички съвременни софтуерни системи.

- **UTF-8** (най-често използваната, стандарт в Linux)
 - Това е **най-препоръчваната и най-често използваната** кодировка в съвременните Linux дистрибуции.
 - Поддържа всички езици, включително кирилица.
 - Компактен за текстове на латиница (ASCII символите остават 1 байт).
 - Динамична дължина на символите (1-4 байта).
 - Повечето Linux дистрибуции използват **UTF-8 по подразбиране**.
 - Командата **locale** може да покаже текущата кодировка на системата.
- **UTF-16 (LE/BE) и UTF-32 (LE/BE)**
 - Използват фиксирана дължина на символите (2 или 4 байта).
 - По-рядко срещани в Linux, защото **UTF-8 е по-ефективен**.
 - Често се използват в Windows и някои специфични приложения.

2. Кодировки, използвани в стари системи

Ако работите със стари файлове или системи, може да срещнете следните кодировки:

- **Windows-1251 (CP1251)** – Използвана в Windows за кирилица.
 - Кодировка, използвана в Windows, но поддържана и в Linux.
 - Подходяща за стари текстови файлове, създадени в Windows.
 - Може да се конвертира към UTF-8 с командата:

```
iconv -f CP1251 -t UTF-8 файл.txt -o нов_файл.txt
```
- **KOI8-R / KOI8-U** – По-стари руски и украински кодировки.
 - **KOI8-R** е използвана предимно за руски език.
 - **KOI8-U** е разширена версия за украински език.
 - По-рядко срещани днес, но може да се наложи работа с тях при стари системи.
- **ISO 8859-5** – Част от ISO 8859 семейството, поддържащо кирилица.
 - Част от ISO 8859 семейството, разработено за кирилица.
 - Рядко използвана днес, защото UTF-8 я е изместил напълно.

3. ASCII (ANSI, ISO 646)

- Стандартната 7-битова кодировка за латиница.
- Не поддържа кирилица или специални символи.

Проверка и смяна на кодировката

1. Проверка на кодировката на терминала

```
locale
```

2. Проверка на кодировката на файл

```
file -i файл.txt
```

или

```
enca файл.txt
```

3. Преобразуване на кодировка

```
iconv -f CP1251 -t UTF-8 файл.txt -o нов_файл.txt
```