**Domácí úkol č. 3**

Karel Vrabec

**Zadání**

Obsah obrázku text

Popis byl vytvořen automaticky

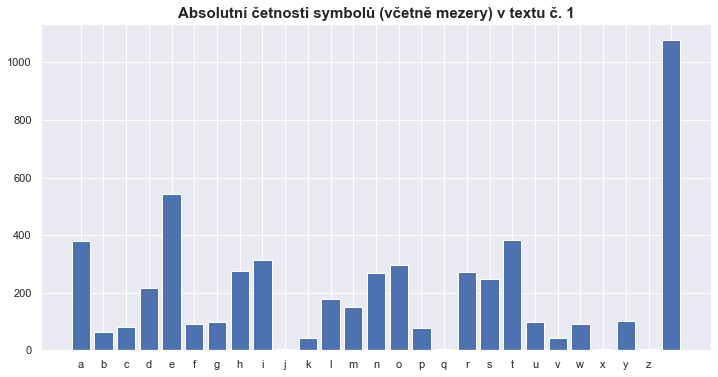
**Řešení**

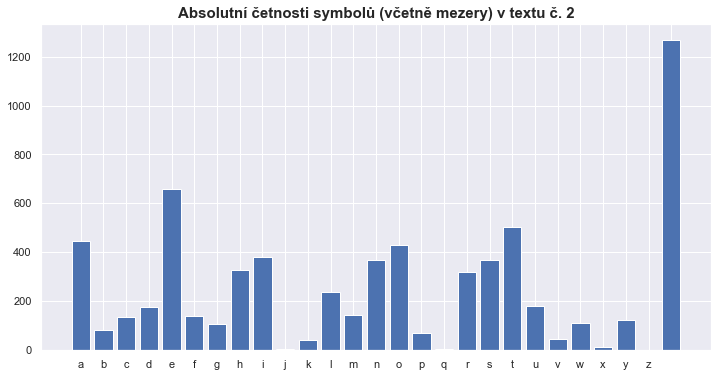
Testované soubory: **005.txt** (jako **text č. 1**), **017.txt** (jako **text č. 2**)

Programovací jazyk: **Python**

Kódy k dispozici v přiloženém archivu **source.zip** (soubor **assignment.ipynb**)

Nejdříve jsem načetl texty z obou datových souborů. Pro každý z nich jsem vypočítal **absolutní četnosti jednotlivých znaků** (včetně mezery). Tyto četnosti jsem také graficky znázornil v následujících grafech.

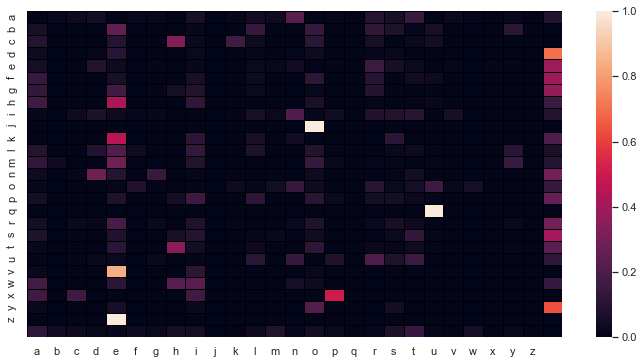




Obsah obrázku stůl

Popis byl vytvořen automaticky

**1.** Nejdříve jsem vytvořil matici o rozměru **počet znaků x počet znaků** (tedy ) vyplněnou nulami. Následně jsem do této matice doplnil **četnosti jednotlivých přechodů** získaných z **textu č. 1**. Postupně jsem procházel tento text a podle **aktuálního** (např. „a“) a **následujícího** (např. „b“) znaku jsem **inkrementoval** hodnotu . Získané četnosti jsem poté vydělil příslušnými řádkovými součty, čímž jsem provedl **normalizaci hodnot matice**. Získanou **odhadnutou matici přechodu** jsem zobrazil v následující **heat mapě**. V ní lze např. zpozorovat, že za znaky často následuje písmeno **e**, příp. **mezera**.



Příslušná ukázka kódu (tvorba matice přechodu):

Obsah obrázku text

Popis byl vytvořen automaticky

**2.** Odhadnutou **matici přechodu** z předchozího bodu jsem použil při výpočtu **stacionárního rozdělení** pro **text č. 1**. Vycházel jsem z rovnice , kterou jsem maticovými úpravami postupně převedl na rovnici (pro výpočet jsem totiž použil knihovnu **NumPy**, která vyžadovala rovnici ve tvaru ). Do matice jsem navíc přidal **nový řádek jedniček** a do nulového vektoru pravých stran **jedničku** tak, aby byl splněn vztah (tj. součet složek vektoru je roven jedné). Jednotlivé složky stacionárního rozdělení a jejich hodnoty jsem zapsal do následující tabulky.

Obsah obrázku stůl

Popis byl vytvořen automaticky

Příslušná ukázka kódu (výpočet stacionárního rozdělení):

Obsah obrázku text

Popis byl vytvořen automaticky

**3.** Provedl jsem **test při známých parametrech**, kdy testuji **nulovou** hypotézu (rozdělení znaků textu č. 2 se rovná rozdělení z předchozího bodu) vůči **alternativní** (nerovnost zmíněných rozdělení) na hladině významnosti .

Obsah obrázku text

Popis byl vytvořen automaticky

Pro tento test jsem provedl příslušné výpočty v nástroji **Excel**. Znaky **j**, **q** a **z** jsem sloučil, aby jejich **teoretické četnosti** dosahovaly hodnoty **alespoň 5**.

Obsah obrázku stůl

Popis byl vytvořen automaticky

Na základě těchto výpočtů jsem získal **testovou statistiku** () a **kritickou hodnotu** (). Vzhledem k tomu, že platí (tzn. že se nacházíme v **kritickém oboru**), **zamítám** nulovou hypotézu o rovnosti rozdělení znaků textu č. 2 a rozdělení ve prospěch hypotézy alternativní.