

Žodžių dažnių skaičiavimo programa

PROGRAMA

Parašykite Shell programą, kuri skaičiuotų žodžių skaičius (dažnius) įvesties tekste. Žodžiais laikysime bent kokius raidžių derinius, susidedančius tik iš mažųjų ar didžiųjų raidžių, atskirtus tarpais ar skyrybos ženklais. Didžiąsias ir mažąsias raides laikykite skirtingomis (t.y. skirkite didžiąsias ir mažąsias raides).

Programos vardas: word-frequencies-cs

Programos iškvietimas: word-frequencies-cs < input.txt word-frequencies-cs inputs*.txt

Įvesties failai 'inputs*.txt'; jei jų nėra, programa turi skaityti iš savo standartinės įvesties.

DUOMENŲ FORMATAI

Įvesties formatas:

Unix teksto failas UTF-8 koduote.

Išvesties formatas:

Duomenys turi būti išvedami dviejų stulpelių CSV [1] faile. Pirma eilutė turi būti stulpelių antraštės (stulpelių vardai. Toliau pateikiami žodžių skaičiai ir patys žodžiai. Pvz. (parodyta tik failo pradžia):

```
Count,Word
5,turi
4,programa
2,ar
2,failai
2,ir
2,klaida
2,programos
1,and
1,apdorojamo
1,apie
```

DIAGNOSTIKA

Programa turi pranešti apie šias klaidas:

- nurodyti failai neegzistuoja, negali būti atidaryti, įvyko skaitymo klaida.

Klaidos pranešimo formatas – toks, kokį išveda Linux įrankiai. Pranešimuose turi būti nurodytas apdorojamo failo, kuriame rasta klaida, vardas ir klaidinga eilutė.

NUORODOS

1. Y. Shafranovich (2005) Common Format and MIME Type for Comma-Separated Values (CSV) Files. URL:<https://tools.ietf.org/html/rfc4180>