Projekt – zadanie

Napíšte program, ktorý bude pracovať s fragmentom DNA sekvencie, zapísaným v súbore `DNAsekvencia.dat`. Tento súbor obsahuje písmená označujúce nukleotidy z abecedy *X = {A, C, G, T}*, pričom tieto môžu byť zapísané malými alebo veľkými písmenami. Predpokladajte, že v súbore môže byť zapísaný fragment obsahujúci najviac 1000 nukleotidov (znakov).
Program bude vykonávať príkazy načítané zo štandardného vstupu. Každý príkaz bude predstavovať malé písmeno nasledované koncom riadku a to:

- `v` – na výpis sekvencie zo súboru na obrazovku. Po aktivovaní tejto voľby používateľ zadá celé číslo *p (1 ≤ p ≤ 20)* predstavujúce počet písmen (nukleotidov), ktoré sa vypíšu na riadok. Potom program otvorí súbor a vypíše celý fragment DNA na obrazovku tak, že na riadok vypíše *p* písmen. Napr. pre sekvenciu *AACGTGCC* a *p = 3* by bol výstup nasledovný:

```
AAC
GTG
CC
```

  V prípade ak sa súbor nepodarí otvoriť, vypíše správu `Neotvoreny subor` nasledovanú znakom konca riadku. V prípade zadania hodnoty *p* z mimo určeného intervalu, vypíše program správu `Nespravny pocet nukleotidov na riadok` nasledovanú znakom konca riadku.

- `n` – na načítanie sekvencie do statického jednorozmerného poľa, pričom skontroľuje, či môže naozaj ísť o DNA sekvenciu, t.j. či neobsahuje iné znaky ako písmená z abecedy *X*. Ak už bola nejaká sekvencia do poľa predtým zapísaná, je potrebné ju nanovo zapísať (medzičasom sa mohla zmeniť). Ak sekvencia v súbore obsahuje viac ako 1000 znakov, načíta sa len prvých 1000 znakov a zvyšok sa ignoruje. Ak sekvencia pozostáva z písmen abecedy *X*, výstupom bude správa `Sekvenciu sa podarilo nacitat` nasledovaná znakom konca riadku. V opačnom prípade sa sekvencia nenačíta a program vypíše správu `Sekvencia nesplna podmienky` nasledovanú znakom konca riadku. V prípade, že sa nepodarí otvoriť súbor, program vypíše správu `Neotvoreny subor` nasledovanú znakom konca riadku.

- `h` – na výpis histogramu, t.j. počtov výskytov jednotlivých nukleotidov. Výstupom sú 4 riadky, každý ukončený znakom konca riadku. Každý riadok začína veľkým písmenom označujúcim nukleotid – v poradí *A, C, G, T*. Toto písmeno je nasledované dvojbodkou, medzerou a počtom výskytov daného písmena v poli načítanom pomocou voľby `n`. Napr. pre sekvenciu *AACGTGCC* by bol výstup nasledovný:

```
A: 2
C: 3
G: 2
T: 1
```

  Ak do poľa ešte sekvencia nebola načítaná, program vypíše správu `Sekvencia nie je nacitana` nasledovanú znakom konca riadku.

- `p` – na vyhľadanie všetkých výskytov podsekvencií dĺžky najviac 10. Po zvolení tejto voľby program načíta sekvenciu najviac 10 znakov ukončenú znakom konca riadku. Ak by bolo načítaných viac znakov, ďalej bude program pracovať len s prvými 10 znakmi. Pre každý výskyt tejto podsekvencie v sekvencii DNA načítanej v poli, program vypíše (v poradí v akom sa vyskytujú v sekvencii) jeden riadok obsahujúci pozíciu prvého písmena nájdeného výskytu danej podsekvencie v sekvencii (pozícia sa počíta od 1), medzeru a časť podsekvencie obsahujúcu 3 nukleotidy pred výskytom podsekvencie, danú podsekvenciu a 3 nukleotidy za výskytom sekvencie. Ak sa nachádza výskyt podsekvencie na začiatku alebo

na konci sekvencie, namiesto chýbajúcich nukleotidov doplňte výstup pomlčkami (znakom mínus).
Napr. pre sekvenciu *ACTTGACGGACCCACG* a podsekvenciu *AC* (výskyty podsekvencie sú vyznačené šedou farbou) program vypíše:

```
1 ---ACTTG
6 TTGACGGA
10 CGGACCCA
14 CCCACG--
```

Ak do poľa ešte sekvencia nebola načítaná, program vypíše správu `Sekvencia nie je nacitana` nasledovanú znakom konca riadku. Ak je sekvencia načítaná, ale podsekvencia obsahuje znaky, ktoré nepatria do abecedy X, program vypíše správu `Neplatny vstup` nasledovanú znakom konca riadku.

- `c` – na výpis komplementárneho vlákna k DNA vláknu zapísanému v poli. K DNA sekvencii predstavujúcej jedno vlákno je možné vyrobiť komplementárne vlákno nasledujúcim spôsobom: nukleotid sa prepíše komplementárnym nukleotidom a zmení sa orientácia vlákna. Komplementárne sú si navzájom nukleotidy *A-T* a *C-G*. Zmena orientácie je prečítanie vlákna z opačného konca. Výstupom je výpis komplementárneho vlákna a znaku konca riadku na obrazovku. Napr. pre sekvenciu *AACGTGCC* by komplementárna sekvencia bola:

```
GGCACGTT
```

Ak do poľa ešte sekvencia nebola načítaná, program vypíše správu `Sekvencia nie je nacitana` nasledovanú znakom konca riadku.

- `d` – na výpočet priemernej vzdialenosti medzi dvoma zadanými nukleotidmi n_1, n_2 v danom poradí. Po zvolení tejto voľby program načíta 2 znaky n_1 a n_2 oddelené jednou alebo viacerými medzerami a nasledované znakom konca riadku. Ak načítané znaky patria do abecedy *X* (načítané môžu byť aj ako malé písmená), program vypočíta priemernú vzdialenosť medzi každou dvojicou n_1, n_2. Priemerná vzdialenosť sa vypíše na 2 desatinné miesta a ukončí znakom konca riadku.
  Napr. pre sekvenciu *CTACGTATTCGCA* a $n_1 = A$, $n_2 = C$ sa vypočítajú vzdialenosti medzi prvým označeným *A* a nukleotidmi *C*, ktoré sú za ním – *CTACGTATTCGCA* (1, 7, 9), a druhým označeným *A* a nukleotidmi *C*, ktoré sú za ním – *CTACGTATTCGCA* (3). Za posledným *A* sa nukleotid *C* nenachádza, tak sa do priemernej vzdialenosti nemôže započítať žiadna ďalšia vzdialenosť. Výstupom je priemer z čísel 1, 7, 9 a 3:

```
5.00
```

  Je dôležité brať do úvahy poradie nukleotidov. Priemerná vzdialenosť nie je rovnaká, ak n_1 a n_2 vymeníme, napr. pre $n_1 = C$, $n_2 = A$ je priemerná vzdialenosť 5.14.
  Ak do poľa ešte sekvencia nebola načítaná, program vypíše správu `Sekvencia nie je nacitana` nasledovanú znakom konca riadku. Ak je sekvencia načítaná, ale niektorý zo znakov n_1, n_2 nepatrí do abecedy *X*, program vypíše správu `Neplatny vstup` nasledovanú znakom konca riadku.

- `k` – na ukončenie programu.

Dôležité poznámky:

Príkazy (voľby od používateľa) načítavajte zo štandardného vstupu, sekvenciu zo súboru. Na výpis používajte štandardný výstup.
Nedodržanie presného formátu výpisu bude mať za následok zníženie hodnotenia.
Používajte funkcie, t.j. každý príkaz (prípadne okrem `k`) sa vykoná vo svojej funkcii, pričom použite prenos argumentov, nie globálne premenné.