Szita Balázs
SZBRAGI.ELTE, CVDW69
szitab@ik.elte.hu

Feladat

Egy nemzetközi cég egy szöveges fájlban tartja számon, hogy mely országok mely városaiban van üzlete.

A szöveges fájl minden sora egy országot reprezentál. A sor elején az ország neve (egy szó) szerepel, majd városnév-szám párok. A városnév lehet több szavas is, a szám az adott város lakosságát jelenti (1000 főben).

A fájlban ezek és csak ezek az adatok találhatók meg. Lehet hogy szerepel a fájlban olyan ország, amihez nincs város, és lehet, hogy a fájl üres. Feltehetjük, hogy a szöveges állomány helyesen van kitöltve.

Válaszoljuk meg az alábbi kérdéseket:

- Listázzuk ki országokra lebontva, hogy hány városában van jelen a vállalat
- A cég potenciálisan hány emberhez jut el világszerte? Azaz mennyi az érintett városok összlakosságszáma?
- Melyik országban tudja a cég a legtöbb embert megszólítani?

Csak egyszer menjünk végig a fájlon.

Specifikáció

Definiáljuk a felsorolandó adatok típusát ekképpen:

$$Data \coloneqq record(orszag\colon String, varosok\colon (String \times \mathbb{N})^*)$$

Tehát olyan *rekord*ok, melyeknek egyik *mező*je az adott ország neve, a másik mezője pedig egy szöveg-szám párosokból álló lista. Ezek lesznek a városok: a szöveg a város neve, a szám pedig a lélekszáma.

Amink van:

$$A = (x\colon SeqInFile(String), lista\colon (String \times \mathbb{N})^*, osszlakok\colon \mathbb{N}, maxnev\colon String)$$

De mennyivel jobb lenne, ha egy ilyenünk lenne:

$$A = (t\colon enor(Data), lista\colon (String \times \mathbb{N})^*, osszlakok\colon \mathbb{N}, maxnev\colon String)$$

$$ef = (t = t' \wedge |t| > 0)$$

A második feltételre a *maximumkiválasztás* tétel miatt van szükség.

$$uf = \Bigg(lista = \bigoplus_{e \in t'} (e.orszag, |e.varosok|) \wedge s = \sum_{e \in t'} osszlakossag(e) \wedge osszlakok = s \cdot 1000 \wedge maxorszag = \underset{e \in t'}{MAX_2}\, osszlakossag(e) \wedge maxnev = maxorszag.orszag \Bigg)$$

ahol:

$osszlakossag\colon Data \to \mathbb{N}$, úgy hogy:

$$osszlakossag(e) \coloneqq \sum_{v \in e.varosok} v_2$$

Szita Balázs
SZBRAGI.ELTE, CVDW69
szitab@ik.elte.hu

Tehát úgy kapjuk egy adott e ország összlakosságát, ha a városainak listáján végigmenve, azoknak rendre a második mezőjét (azaz a lélekszámát) összegezzük.

Visszavezetés és algoritmus

Felsoroló:

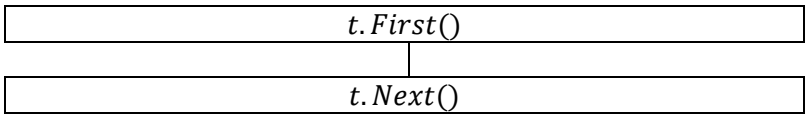
egyedi felsoroló, mely $Data$ típusú elemeket képes felsorolni

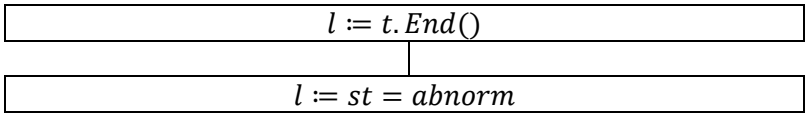
E ~ $Data$
$enor(e)$ ~ $enor(Data)$

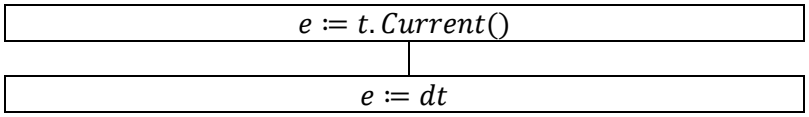
Ezt a felsorolót egy stringeket tartalmazó fájlfelsorolóra építjük. Jelöljük ezt x-szel.

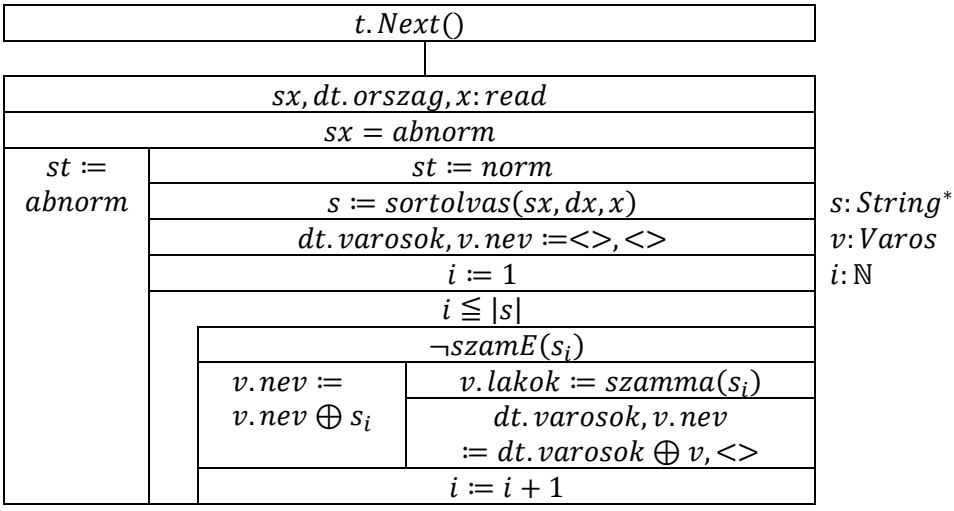
Továbbá a felsoroló osztályunk tartalmazzon egy $\{norm, abnorm\}$ értékű st és egy $Data$ típusú dt nevű változót is.

Műveletei:









Szita Balázs
SZBRAGI.ELTE, CVDW69
szitab@ik.elte.hu

A $Next()$ érdemel egyedül kis magyarázatot. Úgy kapjuk meg a következő felsorolandó országot, hogy először a mögöttes fájlfelsorolóból megkíséreljük kiolvasni a következő adatot (történetesen ez az ország neve), ha ez sikerült, akkor pedig a sor végéig minden mást is feldolgozhatunk (tegyük fel, hogy rendelkezésünkre áll egy sort végigolvasni és szóközök mentén string-listába menteni képes művelet).

A sort miután elmentettük a string-listába, már csak ezen kell végigmennünk, és minden olyan alkalommal, amikor számként értelmezhető adatot találunk (tegyük fel, hogy ilyen művelet is van), akkor úgy vesszük, hogy egy városnév-lakosságszám párral kész vagyunk, beszúrjuk hát azt a listába, majd a „globális” v segédváltozónkat újraindítjuk, ha pedig szöveget olvastunk ki, akkor csak hozzáfűzzük azt az aktuális város nevéhez.

A városok sorozatát is minden körben üríteni kell, különben végül az összes ország összes városát beszúrnánk abba.

Elvét tekintve ez egy *lineáris keresés* és egy *összegzés* tétel egy intervallumon.

Külső tételek:

<u>összegzés (listázás)</u>

s	~	$lista$
H	~	$(String \times \mathbb{N})^*$
$f(e)$	~	$< (e.orszag, \lvert e.varosok \rvert) >$
$+$,0	~	$\oplus$, <>

<u>maximumkiválasztás</u>

$elem$	~	$maxorszag$
max elhagyva		
H	~	$\mathbb{N}$
$f(e)$	~	$összlakosság(e)$

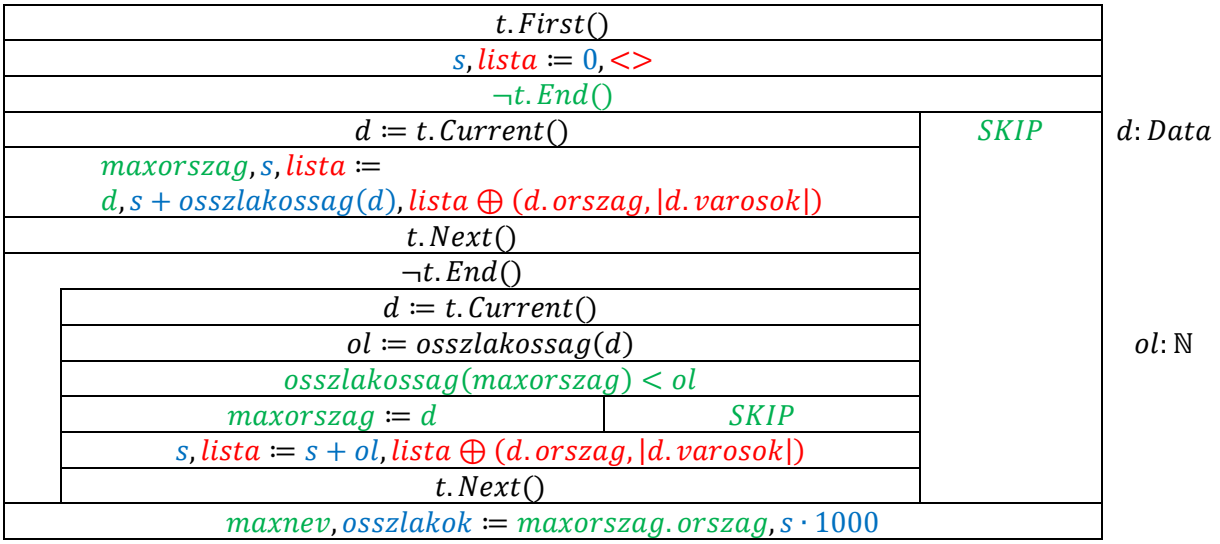
<u>összegzés (lélekszám)</u>

H	~	$\mathbb{N}$
$f(e)$	~	$összlakosság(e)$

Mivel a felsorolón csak *egyszer* mehetünk végig, és az is tiltva van, hogy *elmentsük* a felsorolt adatokat egy segédtömbbe, ezért a feladat három tételét kénytelenek vagyunk „*párhuzamosan*”, egy ciklusban számolni.

Amúgy a feladathoz tartozik még két *összetett függvény kiszámítása* konstrukció is, hiszen a végső lélekszámot úgy kapjuk, hogy s-t megszorozzuk 1000-rel, a legnagyobb lélekszámú ország nevét pedig úgy, hogy lekérjük $maxorszag$ nevét.

Szita Balázs
SZBRAGI.ELTE, CVDW69
szitab@ik.elte.hu



- feketével jelöltem a *felsorolás* műveleteit, és az *FHV*-t,
- pirossal a *listázós összegzés tételt*,
- kékkel az *összlakosságszámos összegzés* tételt (és a hozzá tartozó *összetett függvény*t),
- zölddel a *maximumkiválasztás*t (és a hozzá tartozó *összetett függvény*t).

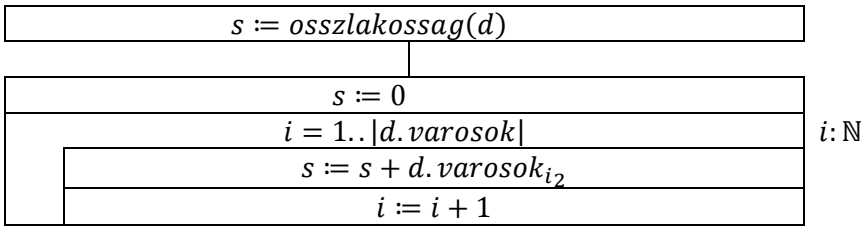
Jegyezzük meg, hogy mivel a két számlálás ciklusának „intervalluma” (teljes felsorolás) eltér a maximumkiválasztásétól (2. elemtől kezdődően a felsorolás), ezért az utóbbihoz igazodtunk és hasonlóan a maximumkiválasztáshoz, a két összegzés első lépését is kidelegáltuk a ciklus elé.

Szóval ezek miatt ilyen bitang bonyolult ez a stuki ☺

Belső tétel:

<u>összegzés + sorozat felsorolása</u>

H	~	$\mathbb{N}$	E	~	$String \times \mathbb{N}$
$f(e)$	~	e_2	$enor(E)$	~	$(String \times \mathbb{N})^*$
			t	~	$(d.varosok, i)$



Tehát minden i-re az i. város második mezőjét, a lakosságszámot összegezzük.

Szita Balázs
SZBRAGI.ELTE, CVDW69
szitab@ik.elte.hu

Implementáció

A programot `C++` nyelven valósítottuk meg.

Adattípusok

A feladat megfogalmazásakor megadott *lista* nevű sorozat szerepét a konzolra kiírás (azaz a `cout`) veszi át.

A városok listája egy `vector`, aminek az elemtípusa egy rekord (`struct`). Ezt a rekordot `Varos`nak nevezzük és két adattagja van: egy `nev` nevű string és egy `lakok` nevű egész szám.

Magát a felsorolót egy osztállyal valósítottuk meg, a négy művelet kódja a mellékelt struktogramoknak megfelel, mögöttes nevezetes fájlfelsorolót használnak.

A $First()$ műveletnél még azt is ellenőrizzük, egyáltalán sikerült-e megnyitni a fájlt. Ha nem, akkor úgy tekintünk rá, mintha üres lett volna.

A $Next()$ műveletnél található „megelőlegezett” műveleteket feloldása a következő:

- $sortolvas()$ ~ `getline` művelet, bár ez nem string-sorozatot, hanem csak egy stringet ad vissza. A bejárását se indexeléssel oldottuk meg, hanem `stringstream`mé alakítottuk és szóközönként olvastunk ebből, akár egy fájlból
- $szamE()$ ~ `atoi` konverzió, amennyiben sikeres volt
- $szamma()$ ~ sikeres `atoi` utáni int típusú változó értéke

Adatok formátuma

A mögöttes fájlt soronként dolgozzuk fel, a memóriában a követelményeknek megfelelően mindig egy sornyi adatot tárolunk csak.

A bemenet formátuma *kötött*, és helyességét NEM ellenőrizzük. A program ugyanakkor MŰKÖDIK üres fájlra is, ekkor a maximumkiválasztás tétel nem fut le, hiszen sérül az EF-e. Ezen kívül értelmetlenül kitöltött fájlokra se „fagy ki” (lásd tesztesetek).

A bemenet egy sora ezt a formátumot követi:

`országnév (1 szó) városnév`$_1$ `(valahány szó) lakosok`$_1$ … `városnév`$_n$ `lakosok`$_n$

Példa:

```
Japan Szapporo 1900 Fukuoka 1400 Oszaka 2600
USA New York 8300 Seattle 600
Romania
Mexiko Ciudad de Mexico 8800 Guadalajara 1450 Monterrey 1300 Tijuana 1300
```

Szita Balázs
SZBRAGI.ELTE, CVDW69
szitab@ik.elte.hu

A kimenet

A listázás a feldolgozás során (annak ciklusában), a két összetett függvényes tételre adott válasz pedig a teljes feldolgozás után kerül kiírásra. A formátum így alakul:

```
listázás (országnév városszám párok)
összlakosságszám
max. lakosságszámú ország neve
```

A bemenetnél megadott példára:

```
Japan 3
USA 2
Romania 0
Mexiko 4
27650000
Mexiko
```

A projekt felépítése

A következő modulokat használjuk:

- *main* – a `main.cpp` forrásfájl, a program belépési pontja (`main` függvény, benne a három külső tétellel), valamint a belső tétel implementációja található itt.
- *enor* – `enor.h` fejlécállomány valamint `enor.cpp` forrásállomány. Előbbi adja meg a felsoroló típusdefinícióját, illetve a felsorolandó rekord típusát, míg utóbbi a műveletek definícióit.
- *iostream* – külső könyvtár, mely a *konzolos kommunikáció* eszközeit teszi elérhetővé.
- *fstream* – külső könyvtár, mely a *fájlkezelés*ben nélkülözhetetlen, a beolvasáshoz használjuk
- *sstream* – külső könyvtár, amelynek segítségével a bemenet egy sorát tudjuk könnyen kezelni és feldolgozni
- *cstdlib* – a program megírásához szükséges volt `atoi()` függvény lelőhelye
- *vector* – külső könyvtár, az `std::vector` típushoz

Tesztelés

Az alábbiakban megadok néhány érvényes és érvénytelen tesztesetet

- Nem létező fájl [**faf.txt**] – úgy kezeli, mintha üres lenne
- Üres fájl [**f0.txt**] – erdemény: 0 (a listázás üres string, a maximum értelmetlen, ez a 0 az összlakosságszám eredménye)
- Hibás formátumú fájl, lakosságszámok nélkül [**f2.txt**] – nem sorol fel semmit, mert nem megy abba az ágba sose a vezérlés, ahol integert talált volna
- Hibás formátumú fájl, üres sorral [**f3.txt**] – az üres sort figyelmen kívül hagyja, jól működik[1]
- Hibás formátumú fájl, egy ország több sorba kerül [**f4.txt**] – a külön sorokat külön országként kezeli
- Hibás fájl, negatív lakosságszámmal [**f5.txt**] – logikusan működik

[1] ez egy eltérés a nevezetes felsorolós megoldáshoz képest. Itt is csak azért van így, mert nem egyből soronként olvastam be, hanem először külön << operátorral az országnevet. Ez ugrotta át az üres sort.

Szita Balázs
SZBRAGI.ELTE, CVDW69
szitab@ik.elte.hu

Egy elem:

- nem létezik hozzá város [**f1.txt**] – Ő a max, 0 város, 0 a lakosság
- egy város létezik hozzá [**f6.txt**] – Ő a max, 1 város, annak lakossága a lakosság
- több város létezik hozzá [**f7.txt**] – Ő a max, n város, az összlakosságszámuk a lakosság

Több elem:

- van köztük 0 városos [**f.txt**] – lásd a dokumentáció példája
- az első a legnagyobb lakosságú [**f8.txt**] – azt is írja ki
- nem az első a legnagyobb lakosságú [**f.txt**] – lásd a dokumentáció példája