

Szegedi Orvostudományi Egyetem Számítástechnikai Központ

Kérdőívek feldolgozására szolgáló általános programcsomag

/FREGOLI/ - az orvosi alkalmazások tükrében

Török Rozália tudományos munkatárs, Kasza Ferenc tudományos
segédmunkatárs

A Szegedi Orvostudományi Egyetem Számítástechnikai Központjában az utóbbi években sok epidemiológiai, pszichológiai, szociológiai kérdőíves felmérés kiértékeléséhez nyújtottunk segítséget. A különböző klinikákról, intézetektől és gazdasági egységektől is több adatfeldolgozási feladatot kaptunk. Ezek a felhasználói szempontból különböző munkák számunkra hasonló problémákat vetettek fel. A sok - lényegében azonos - programozási munka elkerülése érdekében célszerűnek látszott a feladat általános megfogalmazása:

a./ Adott a feldolgozandó adathalmaz, amely tetszőleges számu, tetszőleges, de feladatonként rögzített hosszúságú adatrekordból áll. Ezen rekordok strukturája feladatonként azonos, vagyis az egyes rekordok megfelelő elemei azonos típusuak /egész, valós, karakter sorozat/, és azonos vonatkozásu információs is

hordoznak.

b./ Adott egy tevékenységlista, amelyben a felhasználó szövegszerűen megadja az adathalmaz feldolgozásának módját.

Már 1973-ban a CII-10010 típusu számítógépre kidolgoztunk egy ASTROL nyelvű programot, amely ennek a feladatnak néhány részproblémáját oldotta meg.

Az azóta nyert tapasztalatok alapján az R-10-es számítógép nagyobb hardware és software lehetőségeinek figyelembevételével terveztük meg ennek az általános feladatnak a megoldását. Célunk egy olyan programcsomag kidolgozása volt, amely a feldolgozandó adathalmazra vonatkozóan viszonylag kevés megszorítást tartalmaz, és tág lehetőséget nyújt a szövegszerű tevékenységlista elemeinek a gép számára értelmezhető formába való átirására.

A programcsomag R-10-re íródott assembly nyelven, foregroundban és backgroundban egyaránt futtatható.

Hardware igény: 32 Kbyte központi memória, kártya-, ill. szslegolvasó egység, legalább egy mágnesszalagos egység, 132 pozíciós sornyomtató és egy diszk egység.

Software igény: FMS-10 file-kezelő rendszer a monitorban.

A programcsomag részletes ismertetésénél gyakran használjuk a kérdőívfeldolgozás köznapi fogalmait /kérdőív, kérdések száma, kérdések sorszama, válasz, a válasz lehetséges értéke/, de hangsúlyozzuk, hogy a programcsomag más uton előállt azonos strukturájú adatok feldolgozására is használható.

Az első problémát az input file felépítése jelentette. Mivel egymástól eltérő jellegű adathalmazokat kívánunk kezelni, célszerűnek láttuk az adatrekordokat lényegében változtatás nélkül tárolni, és az értelmezéshez, módosításhoz szükséges információkat segédfile-okban elhelyezni.

Az első segédfile: KIFIL.

Az input file rekordjainak szerkezetét a KIFIL diszkes segédfile-lal definiáljuk. Felépítéséhez meg kell adni a rekordban szereplő adatok számát, "N"-kérdések száma a kérdőíven-, és specifikálni lehet az egyes adatokat a következő szintaxis szerint:

<SORSZÁMJELZŐ> ::= <KÉRDÉS TIPUSA>, <VÁLASZ TIPUSA>;

[<ALSÓ - FELSŐ>]≠

<SORSZÁMJELZŐ> ::= <EGÉSZ> | : <EGÉSZ> / <BETŰ>

<KÉRDÉS TIPUSA> ::= A | K | V

<VÁLASZ TIPUSA> ::= E | V | EGÉSZ

<ALSÓ> ::= <EGÉSZ>

<FELSŐ> ::= <EGÉSZ>.

Ha N-nél kevesebb adatspecifikációt adunk meg, akkor az alapfeltételezés az, hogy a sorszámjelzők egyesével növekednek, a kérdések kódolt típusúak, a válaszok egészek, alsó és felső határt nem adunk meg.

A második segédfile: EFILE.

Az EFILE diszkes segédfile az input rekordok ellenőrzésére szolgáló feltételeket tartalmazza. Az ellenőrző feltételek logikai kifejezések, amelyeket úgy fogalmazunk meg, hogy hibátlan rekord esetén hamis értékűek legyenek. Az EFILE üres is lehet.

A harmadik segédfile: KODFIL.

Ha a feldolgozás során szükség lesz bizonyos válaszok kódolására vagy átkódolására, illetve, ha az eredmény-megjelenítéskor bizonyos kódszámok helyett azok eredeti jelentését kívánjuk kiíratni, akkor a KODFIL diszkes segédfile-t felépítjük.

A KODFIL kódolási és dekódolási utasításokból áll, ezek megadási módja a következő:

⟨SORSZÁMJELZŐ⟩:⟨KÓDUTASÍTÁSOK⟩

⟨KÓDUTASÍTÁSOK⟩:=⟨KÓDUTASÍTÁS⟩|⟨KÓDUTASÍTÁS⟩;⟨KÓDUTASÍTÁSOK⟩

⟨KÓDUTASÍTÁSOK⟩::=⟨KÓDOLANDÓK⟩-⟨KÓDSZÁM⟩|⟨KÓDSZÁM⟩=⟨KÓDOLANDÓK⟩

⟨KÓDOLANDÓK⟩::=⟨KÓDOLANDÓK⟩|⟨KÓDOLANDÓ⟩,⟨KÓDOLANDÓK⟩

⟨KÓDOLANDÓ⟩::=⟨EGÉSZ⟩|⟨VALÓS⟩|⟨KARAKTERSOROZAT⟩|⟨EGÉSZ⟩-
⟨EGÉSZ⟩|⟨VALÓS⟩-⟨VALÓS⟩.

Az elsődleges adathordozón szereplő válaszokat az F input file-ba tároljuk. Mivel a feldolgozandó adathalmaz igen nagy méretű lehet, az input file elhelyezése mágnesszalagon történik - így tulajdonképpen korlátlan számú input rekordot kezelhetünk. A nálunk előforduló feladatok nagy részénél nem jelent megszorítást, ha az input rekordok méretét felülről korlátozzuk, így egy-egy input rekordban /kérdőíven/ maximálisan 256 kérdést engedünk meg.

Az adatok bevitele szalag- vagy kártyaolvasóról történik. Egy maximum 30 karakteres karaktersorozat jelzi minden rekord végét. A program az adatokon a következő ellenőrzéseket hajtja végre:

- megvizsgálja, hogy a KIFIL-ben tárolt választípusoknak megfelelnek-e az adatok;
- egész és valós válasz esetén, ha alsó és felső határt is megadunk, akkor ellenőrzi, hogy a válasz értéke ezen két érték közé esik-e;
- végjel olvasásakor megvizsgálja, hogy a rekordban lévő

adatok száma megegyezik-e N-nel, amit KIFIL-ben adtunk meg;

- ellenőrzést végez az EFILE-ban lévő - esetleges - logikai feltételek alapján.

Ha a program az ellenőrzések során valamilyen hibát észlel, akkor hibajelzést ad:

- megadja a hibás rekord sorszámát,
- a rekordon belüli hibás adat sorszámát,
- és utal a hiba jellegére.

Mivel az input rekordok sorrendje tetszőleges, a hibátlan rekordokat megfelelő konverzió után mágnesszalagra írjuk, a hibás rekordok javítva a hibátlanok után kerülhetnek.

A második probléma jól paraméterezhető tevékenységek definiálása volt, amelyek segítségével a felhasználó szöveges tevékenységkéréseit viszonylag egyszerűen adhatjuk meg a gép számára.

A definiált tevékenységek általános alakja:

- /1/ <TEVÉKENYSÉGNÉV>(<PARAMÉTERLISTA>)*
- /2/ [*OUT(<PERIFÉRIA>,<FORMÁTUMLEÍRÁS>)*],

ahol /1/ a tevékenységet írja le, /2/ pedig az eredmények megjelenítésével kapcsolatos speciális információkat.

Ha a /2/ hiányzik, akkor az eredményt az /1/-hez általunk hozzárendelt periférián, egy általunk rögzített standard formátumban kapjuk meg. A tevékenységnevek 4 betűs azonosítók, az első betű a tevékenység típusára utal. Három típusu tevékenységkérést különböztetünk meg:

C típusu a tevékenység, ha bizonyos input-rekordok számát kell meghatározni,

F típusu tevékenység, ha bizonyos input-rekordokból új file-t kell létrehozni.

P típusu tevékenység, ha minden egyes input-rekordot a tevékenységben meghatározott módon át kell alakítani.

Néhány definiált tevékenységet bemutatunk példaként:

a./ PKOD - a kódolások végrehajtásának része.

Formátuma a következő:

PKOD(<LISTA>)*

<LISTA> ::= <SORSZÁMJELZŐ> | <SORSZÁMJELZŐ>, <LISTA>.

A PKOD paraméterlistájában olyan sorszámjelzők vesszővel elválasztott sorozata szerepel, amelyeknek megfelelő változtatható típusu kérdéseket kódolni szeretnénk a KODFIL segédfile kódutasításai alapján.

b./ CSTA - standard tevékenység kérése.

Formátuma a következő:

CSTA [(<LISTA>)] \neq .

A CSTA paraméterlistája vagy üres, vagy kódolt típusu kérdések sorszámjelzőinek vesszővel elválasztott sorozatából áll. A CSTA tevékenység a listában fel nem sorolt kódolt típusu kérdésekre adott válasz összes lehetséges kódértékének gyakoriságát határozza meg.

A "standard" elnevezés azzal indokolható, hogy minden feldolgozásnál a felhasználó igényétől függetlenül célszerűnek tartanánk egy ilyen kigyűjtést előzetes tájékoztató érdeklődés érdekében.

c./ CTAB: különböző - maximum 10 dimenziós - táblázat elkészítése.

Formátuma a következő:

CTAB(<EGÉSZ₁>, <EGÉSZ₂>/<FELTÉTELEK>) \neq
<FELTÉTELEK> ::= <FELTÉTEL> | <FELTÉTEL> <FELTÉTELEK>
<FELTÉTEL> ::= <SORSZÁMJELZŐ>: <KÓDLISTA> <FELTÉTEL VÉG>
<KÓDLISTA> ::= <EGYSZERŰ KÓDLISTA> | <EGYSZERŰ KÓDLISTA>,
<KÓDLISTA>
<EGYSZERŰ KÓDLISTA> ::= <KÓDSZÁM>-<KÓDSZÁM> | <KÓDSZÁM>.
<KÓDSZÁM> | <ÖSSZEG>

$\langle \text{ÖSSZEG} \rangle ::= \langle \text{KÓDSZÁM} \rangle | \langle \text{KÓDSZÁM} \rangle + \langle \text{ÖSSZEG} \rangle$ $\langle \text{FELTÉTELVÉG} \rangle ::= \text{YN}$.

A CTAB paraméterlistájában $\langle \text{EGÉSZ}_1 \rangle + \langle \text{EGÉSZ}_2 \rangle$ feltétel szerepel. Az $\langle \text{EGÉSZ}_1 \rangle$ az un. szükítő feltételek száma, az $1. \langle \text{EGÉSZ}_1 \rangle$ db feltétellel jelöljük ki az input file egy részfile-ját. Az $\langle \text{EGÉSZ}_2 \rangle$ dimenzióméret. A $\langle \text{KÓDLISTA} \rangle$ -ban a $\langle \text{KÓDSZÁM} \rangle - \langle \text{KÓDSZÁM} \rangle$ ekvivalens a két kód szám közötti egész számok - mint kódszámok - vesszővel elválasztott sorozatával; a $\langle \text{KÓDSZÁM} \rangle . \langle \text{KÓDSZÁM} \rangle$ esetén a két kódszám által határolt zárt intervallumba tartozást kell vizsgálni, az $\langle \text{ÖSSZEG} \rangle$ -ben szereplő kódszámokat nem kell különbözőknek tekinteni. A $\langle \text{FELTÉTELVÉG} \rangle$ a feltétel végének jelzésén túl azt is mutatja, hogy a táblázat megfelelő "sor"-ában kell-e összeget képezni. Ha egy vagy több feltételt zárójelbe teszünk, és névvel látunk el, akkor más CTAB-ok leírásánál ezt a nevet használhatjuk a feltétel, illetve feltételek kiírása helyett /cimkedefiníció, címke-hivatkozás/.

A programcsomag a megfelelően paraméterezett tevékenységlistát egy diszkes LEKERD nevű file-ba helyezi. Ennek rekordjai változó hosszúságúak, három részből állnak:

- FM - a tevékenység leírását tartalmazó mező,
- EM - az eredménymegjelenítés módját tartalmazó mező,
- TM - az eredmények tárolására szolgáló mező.

A mágnesszalagos és diszkes file-ok kialakítása után következhet a tevékenységek tényleges végrehajtása. A mágnesszalagos input file minden egyes rekordján elvégzi a LEKERD file-ban tárolt összes tevékenységet.

P típus esetén az aktuális inputrekordot megfelelően módosítja;

C típus esetén az aktuális inputrekordtól függően a LEKERD rekord TM-jét módosíthatja;

F típus esetén az aktuális input rekordot új file-ba teheti át.

Az input file összes rekordjának feldolgozásával a P és F típusu tevékenységek eredményeit megkaptuk. A C típusu tevékenységek esetén a LEKERD file rekordjainak EM-jében meghatározott periférián és az ott megadott formátumban megjeleníti a TM-ben tárolt eredményeket.

A rendszer februárban készült el, ettől kezdve rendszeresen használjuk. Az első feldolgozás a SZOTE Nőgyógyászati Klinikáján végzett demográfiai vizsgálat kiértékelése volt. Ekkor 6000 kérdőívet, kérdőivenként 67 adatot /azaz közel félmillió adatot/ dolgoztunk fel, 90 táblázatot készítettünk. A tényleges feldolgozáshoz kb. 15 óra gépidő volt szükséges. Az egyetem tanulmányi osztályának kérésére félévenként elvégezzük az egyetemi hall-

gatók összesen kb. 60 ezer adatának kiértékelését. További feldolgozásaink voltak: az Ideggyógyászati Klinikán az alkoholistákról felvett adatlapok és a Sebészeti Klinika porckorong-sérves betegadatainak kiértékelése. Látható, hogy teljesen eltérő adathalmazokra használható a programrendszer, s külön kiemeljük, hogy a tevékenységkérések felírását a felhasználó igényeinek megfelelően középfoku végzettséggel rendelkező betanított munkakerő is el tudja végezni.

Szeptemberben a programcsomagot felhasználásra átadtuk a Szekszárdi Megyei Kórháznak, ahol információink szerint rendszeresen használják.