

MTA SZTAKI

Egy általános információs rendszer kórházi morbiditási adatok
feldolgozására

Ruda Mihály

Az MTA Számítástechnikai és Automatizálási Kutató Intézetében a múlt év végén indult meg egy általános adatfeldolgozó rendszernek az ÁSZSZ HwB 66/20-as, ill. 66/60-as gépén történő megvalósítása. A rendszer fejlesztésének alapját a SZTAKI régebbi kutatásai és az Egészségügyi Minisztériummal kötött szerződés képezik. A rendszer tehát elsősorban orvosi, egészségügyi adatfeldolgozás céljaira készült. Az előzetes tapasztalatok és a rendszerrel szemben támasztott követelmények azonban kívánatosá tették egy igen széles körben alkalmazható általános adatfeldolgozó rendszer felépítését.

A következőkben elsőként a rendszerrel szemben támasztott igényekről, utána a megvalósítás módjáról beszélünk. Befejezésésként röviden áttekintjük az irodalomban található, vagy már hazai gépeken is hozzáférhető adatfeldolgozási lehetőségeket.

1. A rendszer célja, a rendszerrel szemben támasztott igények

A rendszer elsődleges célja az, hogy az évenként ismétlődő kórházi morbiditási adatfelvétel számítógépes feldolgozását biztosítsa.

A feldolgozás statisztikai táblázatok készítését jelenti. A kórházi morbiditási adatfeldolgozás nem egy zárt rendszer. Változhatnak az alapadatok (szerkezetükben és tartalmukban is), változhatnak az adatokra vonatkozó információs igények is.

Felsorolunk néhány, a jelen pillanatban folyamatban lévő változást. Most áll kidolgozás alatt az új kórházi adatszolgáltató rendszer (adatlap). Ezzel a teljes feldolgozandó adatrendszer megváltozik. Leglényegesebb változások a kórházak és szakmák besorolásában történnek, de megváltozik pl. a betegazonosító és a beteg foglalkozá-

sának kódolása is. Ezekről a változásokról függetlenül más külső tényezők miatti változások is fellépnek. Jelenleg tervezik az un. ANH személyazonosító bevezetését, és a betegségek nemzetközi osztályozását (BNO kód) is rendszeresen átdolgozzák. Éppen a közeljövőben kerül sorra az új BNO kód bevezetése.

Ezek a változások esetenként teljesen új ellenőrzési eljárásokat igényelnek, és megváltoztatják az adatok között fennálló logikai kapcsolatokat rendjét is.

Az információ-igények ugyancsak változnak, egyrészt az előbb vizsgált adatváltozások miatt, másrészt a rendszer működtetésével nyert tapasztalatok és más források alapján is változik a felhasználó igénye.

A rendszerrel szemben támasztott legfontosabb követelmény tehát a változó feltételekhez és igényekhez való alkalmazkodás képessége. Emellett természetesen gazdaságossági szempontokat is figyelembe kell venni.

Célunk egy olyan általános statisztikai adatheldolgozó rendszer kialakítása volt, amely a gépkapacitás, az emberi munkamennyiség és az átfutási idő minimalizálására törekszik, miközben a lehető legnagyobb mértékű rugalmasságot biztosítja.

2. A rendszer működésének lépései

2.1. Adatellenőrzések, kiegészítések

A rendszer gyakorlatilag tetszőleges fix rekordokból álló input file-t feldolgozhat. Mivel a rendszer statisztikai adatheldolgozást végez, ezért elsősorban numerikus adatok vizsgálatára használjuk. Szöveges információk csak megjelenítésre kerülnek (pl. táblázatok fejlécein stb.).

Az input rekord leírását a felhasználó adja meg. Az adatellenőrzésnek és új adatok előállításának a következő módjai között választhatunk:

2.1.1. az egyes adatok a megengedett korlátok között vannak-e,

2.1.2. vannak-e olyan adatok, amelyek a megengedett határokon belül tiltott értékeket vesznek fel,

2.1.3. egy rekordon belül vannak-e egymásnak ellentmondó adatok, adatcsoportok,

2.1.4. az eredeti adatokból, adatcsoportokból új adatokat lehet képezni,

2.1.5. a felhasználó a rendszerbe illeszthet egyéb adatellenőrzési és adatképzési eljárásokat is.

Az itt felsorolt lehetőségekkel a felhasználó gyakorlatilag tetszőleges módon élhet. Néhány példa a kórházi morbiditási vizsgálatból: gyakorlatilag tetszőlegesen változó kórházkódszámrendszer esetén kiválogathatók a hibás kórházkódok, megadhatók a beteg korával és nemével (elég bonyolult módon) összeférhetetlen diagnózisok, ugyanígy az eredetileg adott négyjegyű diagnózis kód összevont jegyzékre való leképezése is könnyen megvalósítható stb.

Az ellenőrzési eljárásoknál a felhasználó csak a mindenképpen szükséges paramétereket adja meg: az adatok közötti logikai kapcsolatok, a megengedett adatértékek, új adatok létrehozásakor az eredeti és új adatok megfeleltetése.

Az ellenőrzési és új adatképzési eljárások leírása sokszor hosszadalmas és bonyolult folyamat lehet. Éppen ezért a rendszer a megadott paramétereket részletesen ellenőrzi és hiba esetén hibajelzést ad. Ha a felhasználó szintaktikusan hibátlan feladatleírást ad, akkor egy jól áttekinthető listát kap a rendszertől az általa leírt feltételekről és adatmegfeleltetésekről (értéktáblázatokról). Ezen a listán ellenőrizheti, hogy tartalmilag helyesen írta-e le az egyes logikai feltételeket. Ezek a listák a későbbiekben is eligazítást biztosítanak az adatértelmezésekkel kapcsolatban.

2.2. Válogatás, konvertálás

Adatfeldolgozásnál általában szükség van a teljes adatrendszer leszűkítésére - egyrészt a felesleges adatmozgatások elkerülése érdekében, másrészt a feldolgozás egyik célja lehet éppen egy részpopuláció kiválasztása. Egy példa: ki akarjuk választani egy adott betegség típusban szenvedő, 40-50 év közötti nehézipari, fizikai férfidolgozók csoportját.

Rendszerünk a legbonyolultabb logikai feltételek alapján is képes elvégezni a kiválasztást.

A válogatási eljárással egy lépésben végezhető karakter formából bináris formába történő konverzió is. A bináris forma igen gyors feldolgozást biztosít a továbbiakban, ugyanakkor a rendszer egy igen tömör tárolási módot alkalmaz, lényegesen csökkentve a bináris tárolási mód egyébként nagy helyigényét.

A válogatási eljárásokról a felhasználó gyakorlatilag tetszőlegesen rendelkezhet, a rendszer szintaktikus szabályainak keretei között.

2.3. Speciális file-ok létrehozása

A többször ápolts személyek és a változó számban előforduló kísérőbetegségek vizsgálata speciális eljárások beépítését teszik szükségessé. Ezek az eljárások lehetővé teszik változó hosszúságú rekordok kezelését is. A rendszer a változó hosszúságú rekordokat egy már csak fix rekordokat tartalmazó file-ba viszi át.

2.4. Transzformált file-ok

A statisztikai adatfeldolgozásnál - az adatnyilvántartó rendszerekkel szemben - nem egyedekkel, hanem típusokkal foglalkozunk. Az adatfelvételek azonban általában egyedek adatait rögzítik. Az egyedeket tartalmazó file-ról a típusok file-jára való átmenetet nevezzük itt file-transzformációnak.

A rendszer az input adatokból nyerhető tetszőleges típusbesorolást megvalósít. Egy ilyen típus lehet pl. egy adott kórház egy kijelölt osztályán fekvő adott korcsoportba tartozó betegek összessége.

Ez a részrendszer tehát a kívánt statisztikai táblázatok adatait (vagyis az egyes típusokra vonatkozó adatokat) állítja elő. Ez a részrendszer lehetővé tesz még adatösszevonásokat - pl. korcsoportok, ápolási időcsoportok stb. kialakítása - is. Ezzel a típusalkotási lehetőségek még változatosabbak lesznek.

2.5. A táblázatok megjelenítése

Ennek a részrendszernek a következő feladatai vannak:

2.5.1. a már kialakított típusok kívántság szerinti tetszőleges összevonása, típuscsoportok kihagyása,

2.5.2. a táblázatokon belül tetszőleges részösszegek képzése,

2.5.3. komulativ eloszlások megadása,

2.5.4. az eloszlások grafikus megjelenítése,

2.5.5. a változó táblafejlécek szerkesztése.

A file-transzformáció és a táblázatmegjelenítés különválasztásának fő oka az, hogy a felhasználó sokszor több, közel azonos tartalmu táblázatot igényel. Ilyenkor egyetlen file-transzformáció után állithatunk elő több táblázatot. Egy másik lényeges szempont az, hogy így a felhasználó neki tetsző módon változtathatja a táblaformátumot, miközben a táblaadatokat (a transzformált file) továbbra is rendelkezésére állnak.

2.6. Általános megjegyzések

A rendszer működtetése az operációs rendszer vezérlőkártyáival és az aktivizált rendszerprogramok paraméterkártyáival történik. Mivel a paraméterkártyák egymáshoz kapcsolódó feladatokat és azok megoldási módját jelölik ki, ezért ezeknek a kártyáknak az összességét egy célnyelvnek tekinthetjük, amely statisztikai adatfeldolgozási folyamatok leírására szolgál.

A rendszerprogramok futásonként újra és újra generálódnak azért, hogy a változó feladatok optimális megoldását, és egy igen gazdag kombinációs lehetőséget biztosítsanak.

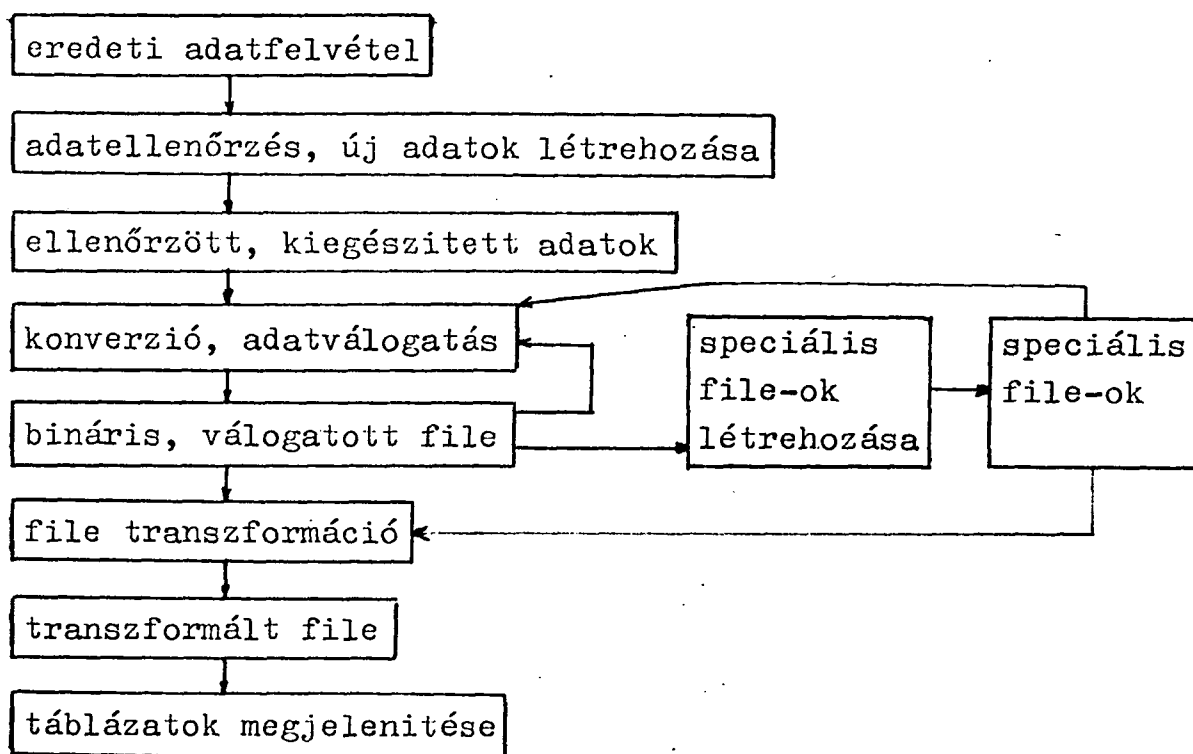
Mivel numerikus adatok kezelése a cél, ezért a rendszer nyelve FORTRAN. Ezen a nyelven írhatók a rendszerhez csatolt kiegészítések is.

Külön kell szólni a rendszer működéséhez szükséges kódtáblák feltöltésének módjáról.

A táblák feltöltésére a rendszer egy igen tömör és könnyen áttekinthető leírási módot biztosít. Egy példa: a 15, 252-317, 915, 33, 17 értékeket akarjuk rendre az 1-70 értékeknek megfeleltetni. Ezt a következő utasítással biztosíthatjuk:

lnnnln 15,252-317,915,33,17.

Befejezőként egy vázlatos ábrát adunk az egyes részrendszerek egymáshoz való kapcsolódásáról.



3. A rendszer viszonya más, hasonló célú számítástechnikai eszközökhöz

Az információs rendszerekkel kapcsolatos kutatások és fejlesztések fő vonala az "adatnyilvántartás" kérdéskörére irányul. A statisztikai adatfeldolgozás területén részfeladatok megoldása ismert. Ilyenek pl. a különböző programcsomagokban, programkönyvtárakban megtalálható, vagy gépi rendszerekbe beépített adatellenőrzési eljárások vagy adatkezelő célnyelvek (mint pl. az IBM 360 OS-ben működő Utility-Coder).

Ugyanigy bizonyos logikai kapcsolatokat realizáló gépi lehetőségek is léteznek - pl. a Honeywell IDS lehetősége. Ismertek továbbá általános táblaelőállítási eljárások is (mint pl. az ugyancsak IBM rendszerekben működő TPL).

Rendszerünk azonban egy egységbe integrált teljes és általános adatfeldolgozó rendszer, amely a feldolgozási folyamatot statisztikai szempontok alapján optimalizálja is.

Mivel a rendszer egy általánosan használt nyelven (FORTRAN) íródott, ezért más gépekre történő átvitele és kiterjesztése sem ütközik különösebb akadályba.

Külön felhívjuk a figyelmet arra, hogy az adatnyilvántartási feladatokra kifejlesztett rendszereket, nyelveket, (pl. az előbb említett IDS) nagyméretű statisztikai adatfeldolgozásra nem célszerű alkalmazni.

4. Az adatfeldolgozás statisztikai problémái

Információs rendszerünk kidolgozásakor a következő matematikai statisztikai kérdéseket vizsgáltuk még:

1. a mintavétel statisztikai szabályai,
2. optimális minta - megbízhatósági kérdések,
3. azonosító kódok statisztikai viselkedése,
4. az adattartalom szerepe a feldolgozási módszerek kiválasztásában,
5. optimális számítástechnikai megoldások,
6. a cluster-analízis alkalmazási lehetőségei.

Ezekkel a kérdésekkel egy külön tanulmányban foglalkozunk.

