

FIGYELŐ  
SZOLGÁLAT

ISMERTETÉSEK

681.3.048

DÜRR, H.G.:

ANIS - ein allgemeines nichtnumerisches Informations-System  
für das IBM System/360.

/ANIS - általános nem-numerikus információrendszer az IBM  
System/360 adatfeldolgozó berendezés részére/

IBM Nachrichten, 18.k. 190.sz. 1968. aug.  
p.284-290.

A tanulmány bevezetése néhány érdekes adattal illusztrálja napjaink információlavináját. Vácsolja a numerikus és nem-numerikus információk feldolgozásának mibenlétét, rámutatva a kettő közötti eltérésekre, s azokra a nehézségekre, melyek a nem-numerikus információk feldolgozásánál jelentkeznek. Ezután a nem-numerikus információrendszerekről, ezek két típusáról szól: a dokumentumonkénti és a deskriptoronkénti adattárolást alkalmazókról, melyeket a szerző a computer perifériális memóriáinak tárolási módzatai nyomán szekvenciális /egymásutáni/ és közvetlen elérésű tárolást alkalmazó rendszereknek nevez.

Közvetlen elérésű tárolást alkalmaz a tanulmányban ismertetett rendszer, az ANIS /Allgemeines Nichtnumerisches Informations-System - Általános nem-numerikus információrendszer/ is. A rendszer jelentősebb sajátosságai: a dokumentumok feltárását magának a rendszernek a keretei között kialakított tezausz alapján végzi; információkereső programja a tezauszban szereplő fogalmak tetszés szerinti logikai kombinációja nyomán történő információ-visszakeresést tesz lehetővé; a gépi közegen tárolt információállományból a kívánt információ dialogizáló eljárással nyerhető ki, ami egyrészt a fölösleges információk leválasztását, másrészt a keresés tetszés szerinti finomítását biztosítja, tehát hatékony tájékoztatást valósít meg.

Az ANIS, mint a legtöbb tájékoztatási rendszer, több különböző típusú adatféleséget - deskriptorokat, dokumentumok azonosságai szignaturáit, címadatokat stb. - ölel fel. Ezek összességét a tanulmány

adatbanknak nevezi. Az adatbank létesítésére, gyarapítására és hasznosítására kidolgozták a computer programokat, mégpedig több programnyelven /pl. FORTRAN, COBOL/. Az ANIS rendszerét tehát az IBM 360-as computerok üzemmodjától függetlenül alakították ki. Az adatbank kiépítését és karbantartását úgy oldották meg, hogy az minél kevesebb emberi közreműködést igényeljen. Ily módon igyekeztek kiiktatni a legjelentősebb hibaforrást.

A feldolgozandó szövegekből /címleírás, referátum/ a computer a program nyomán végzett automatikus elemzéssel válogatja ki a releváns szavakat, melyekből a deskriptorokat alakítják ki. E művelettel kapcsolatban a következők méltók említésre:

hogy meddig terjed egy szó, azt a gép a szóköznek, vagy a szavak után következő interpunkciós jeleknek figyelembe vételével állapítja meg;

a releváns szavakat az irreleváns szavak szótárának segítségével határozza meg a gép. Ezt a szótárt manuális uton előre össze kell állítani. Amelyik szót a gép a szótárban nem találja meg, relevánsként kezeli;

ha a téma kifejezésére a szöveg nem tartalmaz elegendő számú deskriptort, illetve megfelelő deskriptorokat, a gépi feldolgozás előtt emberi közreműködéssel /a hiányzó deskriptorok betoldásával/ a kívánt szintre kell emelni a szöveg információtartalmát;

a deskriptoroknak a szövegekben mindig azonos írásmódban kell szerepelniük, hogy a gép kiválasztásuk során következetesen tudjon eljárni. Ha nem így lenne, emberi munkával korrigálni kell a szöveget.

A feldolgozás során érvényesülő másik szempont az, hogy a dokumentumok adataiból /azonossági szignatura; cím adatok; referátum/ annyit tárolnak, amennyire a tájékoztatás során szükség lesz. Ez a volumen elég tág határok között mozoghat. Minimális, ha csupán az azonossági szignaturából áll, maximumát pedig a dokumentum teljes szövege képezi.

Az információkat bármilyen input-eljárással be lehet juttatni a rendszerbe: lyukkártyáról, lyukszalagról, mágnesszalagról való beolvasással, szövegolvasógép alkalmazásával stb. A gépi közegre való rávitel szabályai egyszerűek és könnyen elsajátíthatók. A gépi feldolgozás során a közvetlen elérésű tárolóba /mágneslemezre vagy mágnescsik-tárolóba/ kerülnek azok az adatok, amelyeket az információkeresés eredményeként a gépnek ki kell adni. A dokumentum témáját jellemző deskriptorokat a vonatkozó szignaturákkal a másik külső memórián, mágnesszalagon tárolja a gép.

A deskriptorok összessége képezi a rendszer tezauszát. A tezausz alapját képező szókészletet maga a computer válogatja ki a szövegekből. Ebből a deskriptor-állomány kialakítása emberi közreműködést igényel. A feldolgozott szövegekből a gép feliratozott kártyákon adja ki a szavakat. Ezt a kártyaanyagot szakértők vizsgálják fe-

túl, s döntenek arról, hogy a szavak milyen formában kerüljenek deszkriptorként a tezauszba. A deszkriptorok különféle variánsairól /pl. deklinációs formáiról, szinonimáiról/ utalnak a deszkriptorra. Ez gyakorlatilag rendkívül egyszerű művelettel történik: a deszkriptor kártyája mögé helyezik a variánsok kártyáit. Az így átrendezett kártyaállományt újra beolvassák a computerbe, mely jegyzék formájában írja ki a tezauszot. Ebben a variánsokról a deszkriptorra való utalás már szöveges formában jelenik meg. A deszkriptorok mellé pedig kiteszik az előfordulásuk multiplicitását mutató számot is.

A releváns szavaknak ilyen kezelése lehetővé teszi más nyelvű anyagok feldolgozását is. Ehhez két művelet szükséges. Az egyik: összeállítani a másik nyelv irreleváns szavainak szótárát, s azt csatolni a már a gépi memóriában tárolt szótárhoz. A másik: az idegen nyelvű releváns szavakat variánsokként kezelni, azaz kártyájukat az eredeti nyelvű deszkriptor kártyája mögé helyezni.

A tezausz nyomán a gép ún. címjegyzéket állít elő. A címjegyzék betűrendbe sorolt deszkriptorok mellett közli azoknak a dokumentumoknak azonosítási szignaturáit, amely dokumentumoknak a vonatkozó deszkriptor jellemzője. A címjegyzéket közvetlen elérési memóriában tárolják oly módon, hogy az könnyen bővíthető legyen. A címjegyzék, s annak említett tárolása biztosítja az ANIS rendszernek azt az alapvető célkitűzését, hogy a feltett kérdésre a lehető legrövidebb időn belül tud kimerítő információt szolgáltatni.

Az információkeresés során a deszkriptorok között "és", "vagy", "nem" logikai kapcsolatok érvényesíthetők. Lehetővé teszi a rendszer, hogy az információt igénylő személy a géppel folytatott dialógus útján szerezhesse be a számára szükséges információkat a gép memóriából. Ez úgy történik, hogy a kérdező egy külön e célra szolgáló készüléken lebillentyűzi a kérdést, mely távközlési vonalon bejut a computerbe. A gép pillanatok alatt előkeresi a kért információt, s azt szintén távközlési vonalon juttatja vissza a kérdést feltevő készülékbe, mely azt szöveges formában adja ki. Ha a válasz nem kielégítő, a kérdező újabb szempontok közlésével módosíthatja, pontosíthatja kérdését, melyre a gép újra és újra - szintén megfelelően módosított - válaszokat ad. Párbeszédesszerű kapcsolat alakul így ki a kérdező személy és a computer között.

A dialogizáló eljárást úgy alakították ki, hogy a kérdés feltételekor különleges feltételek is megadhatók; mint

az információ keletkezési /pl. a dokumentumok megjelenési/ időpontjai; és pedig az alsó határ /...évtől/, vagy a felső határ /...évig/, illetve mindkettő;

a válaszba felvehető dokumentumok számszerű mennyisége /ha esetleg túlzott mérvű eredmény várható/;

"és" logikájú kérdések esetén annak a rögzítése, hogy a válaszban csak olyan dokumentumok lehetnek, melyekben a kérdésnek meghatározott számú deszkriptora szerepel;

"vagy" logikája kérdések esetén annak a rögzítése, hogy válaszként csak azok a dokumentumok adhatók ki, melyekben a kérdés deskriptorai közül a legtöbb fordul elő.

Az információk illetéktelen hozzáférése ellen két biztosító tényezőt iktattak a rendszerbe. Az egyik a kérdésre jogosult egyén azonossági kódja. Kérdésfeltevéskor ezt a kódot közölni kell a géppel, különben nem szolgáltat információt. A másik a bizalmasan kezelendő információk titkos zároló-kódja. Az ilyen kóddal ellátott információkat csak azokra a kérdésekre szolgáltatja ki a gép, amelyekbe a zároló-kódot előírt formában beiktatták.

Az ANIS-rendszer az IBM 360-as gépcsalád bármelyik computeréhez használható a 30-as típustól felfelé. Kétféle üzemmód lehetséges: központositott információszolgáltatás vagy egyéni önkiszolgálással történő információbeszerzés. Az előbbi esetben a kérdések a computerrel rendelkező központba futnak be, ott összegyűjtik azokat és meghatározott időközökben /pl. a nap kijelölt órájában/ egyetlen gépmenetben megválaszoltatják valamennyit. A másik esetben a computerhez távközlési vonalakon az információt igénylők hálózata csatlakozik. Mindegyiknek saját adatállomása van, amellyel a computernek közvetlenül tehetnek fel kérdéseket, illetve vehetnek át attól információkat. Ilyenkor alkalmazzák a már vázolt dialogizáló eljárást. Az első üzemmód feltétele a gép részéről a legalább 64 K /= 65 536 byte/ memória-kapacitás, a másik üzemmódhoz pedig legalább 128 K /= 131 072 byte/ volumenű memória kell, s természetesen a szükséges perifériális készülékek.

A továbbiakban röviden ismerteti még a cikk az IBM cég adatállomásainak különféle típusait, melyek az ANIS-rendszerhez felhasználhatók. Befejezésül az ANIS-rendszer alkalmazási területeiről szól. Ezek között felemlíti a szakirodalmi dokumentációt, a szabadalmi tájékoztatást, valamint a kutatási nyilvántartásokkal, jelentésekkel kapcsolatos információszolgáltatást.

Orosz Gábor