

A Párizsi Videotéka teljesen automatizált dokumentációs rendszere

Pierre Emmanuel költő ötlete és kitartó rábeszélése folytán Párizs polgármestere, Jacques Chirac kezdeményezte a Párizsi Videotéka (Videothèque de Paris) felállítását, amely 1988 februárjában nyitotta meg kapuit. A videotéka működésének egészét a francia főváros költségvetéséből finanszírozzák. Az egyik legfrekvenciáltabb kereskedelmi és kulturális negyedben, a Halles Forumban helyezték el (közel a világhírű Pompidou Kulturális Központ). A videotéka teljes mértékben nyilvános, alkalmakként 18 frankért valamennyi szolgáltatását igénybe lehet venni.

A Párizsi Videotéka gyűjtőköre a francia fővárosra mint filmtémára épül, egyrészt hogy megőrizze a Párizsra vonatkozó audiovizuális dokumentumok teljességét, másrészt hogy elősegítse a város létének és fejlődésének minél differenciáltabb (építészeti, kulturális, szociológiai, gazdaságpolitikai stb.), ugyanakkor minél komplexebb megértését. Az indulásnál az állomány 2500 címből állt, amely mintegy 1300 órás műsoranyagot tett ki. (Ez az állomány egy év múlva 2800 címre, 1800 műsorórára emelkedett, s évi 40 000 látogatója volt.)

A gyűjtésben nem tesznek különbséget a legkülönbözőbb műfajú és technológiával készült filmek között, így egymás mellett sorakoznak a játékfilmek, a hirdók és a reklámfilmek. Minden állományegységnek van filmszalagos és videokazettás változata. A gyűjtés kezdőideje a mozgófilm feltalálása.

A filmek megtekintése két alapvető szolgáltatástípus keretében lehetséges. Az első *menü* jellegű, azaz fix programokból áll. Ezek egy központi téma köré csoportosulnak (pl. az éneklő Párizs, Párizs külvárosai, Párizs a III. köztársaság idején). A második szolgáltatástípus *à la carte* jellegű, azaz a felhasználók a Pierre Emmanuel-olvasóterem 30 db filmkereső-olvasó berendezésének valamelyikével választhatják ki az őket érdeklő egy vagy több filmet.

A második szolgáltatástípus igénybevétele sincs korlátozva. Kifejlesztésénél a fő kihívást az jelentette, hogy egyszerűen minden felhasználó szabadon, természetes nyelven, a maga egyéni – esetenként laikus – kifejezőmódján tehesse fel a keresőkérdéseit, másrészt a keresés teljesen automatikusan, közvetlenül, azaz bármiféle emberi tényező (pl. tájékoztató könyvtáros, profilszerkesztő vagy online közvetítő) közbeiktatása nélkül valósuljon meg. Ezt úgy oldották meg, hogy egyesítették a videotexes szakirodalmi információkeresést a releváns találatokat képező audiovizuális dokumentumok azonnali, automatikus megjelenítésével, ill. lejátszásával. Ennek érdekében a számítástechnika, a robotika, az optika és a koaxiális kábeltechnológia, valamint a videotechnika legújabb eredményeinek igen új és eredeti kombinációját kellett megvalósítani. Mindez maga után vonta

a hagyományos könyvtári és dokumentációs munkafolyamatok, illetve a szakirodalmi információkeresés technológiájának alapvető modernizálását.

VIDEOTEX ÉS TERMÉSZETES NYELV: SZAKIRODALMI INFORMÁCIÓKERESÉS

Itt az volt a cél, hogy közvetlen, interaktív kommunikáció épüljön ki a természetes nyelven kérdező felhasználó és a filmek audiovizuális nyelvezetű világa között. E célt az információkereső rendszer három fő és nélkülözhetetlen alrendszere révén sikerült elérni:

- ▶ Videotex rendszer, amelynek használata annyira egyszerű és természetes, hogy semmiféle előképzettséget nem igényel.
- ▶ Természetes nyelvű szakirodalmi információkeresés, amelyben a felhasználó mentesül a szaknyelv kötöttségeitől, így akár a legegyszerűbb személyes kifejezőmóddal is gyorsan, eredményesen kereshet.
- ▶ Filmadatbázis, amelyben – műfajától s egyéb jellemzőitől (gyártási idő, nemzetiség stb.) függetlenül – minden egyes filmről azonos felépítésű, igen részletes dokumentumkép készül, s ez azonnali, közvetlenül bekapcsolható megjelenítést (elejétől végig történő megnézés) tesz lehetővé.

A filmek dokumentumképét tartalmazó fájl szerkezete

A fenti három alrendszer részletesebb ismertetése akkor történhet a leglogikusabb módon, ha a harmadikkal kezdjük.

A filmek dokumentumképét tartalmazó fájl – mint már utaltunk rá – teljes mértékben homogén szerkezetű. E szerkezet három fő részre tagolódik, egyes részei kb. egy videotex oldal hosszúságúak:

- ▶ részletesen annotált filmográfia;
- ▶ azoknak a párizsi vonatkozású részleteknek, képsoroknak a jegyzéke, amelyek a francia főváros életének legkülönbözőbb vonatkozásait (történelmi, művészeti, szocioprofessionális, mindennapi stb.) bármely műfaj (pl. dokumentumfilm, játékfilm) keretében ábrázolják;
- ▶ a filmkópián vagy a videokazettán található valamennyi adat (rendező, operatőr, gyártásvezető, zeneszerző, vágó, szereplők stb.), illetve a gyártástechnológiára, a szerzői és forgalmazási jogokra vonatkozó információk).

Ez a feldolgozási mód a filmet elsősorban műalkotásként kezeli, ugyanakkor azonban jól szolgálja a videotékának azt az eredeti funkcióját is, hogy általa fel lehessen tárni Párizs rendkívül változatos arcúlatának valamennyi filmnyelven történt bemutatását, még akkor is, ha az ilyen jellegű képsorok egy-egy adott műalkotásban csak futólag fordulnak elő.

Fontos hangsúlyozni, hogy miközben minden egyes filmről készül egy neki megfelelő dokumentumkép, ha azonban a szóban forgó filmet az alkotók több részről állónak tartják, akkor minden így minősített részéről külön-külön dokumentumképet is készítenek.

A dokumentumképben szereplő valamennyi szónak kulcsszó státusa van, azaz a dokumentumkép minden szava közvetlen információkeresésre használható. A felhasználó monitorján a dokumentumképnek csak egy része jelenik meg, de a benne szereplő kulcsszavak szisztematikusan egy sor általánosan szükséges adattal egészülnek ki, mint pl. az adott mű audiovizuális műfaja, fő témája, a forgatás, ill. a cselekmény időpontja, a gyártás ideje, helye, a film nemzetisége, forgalmazója. Ez utóbbi adatok a keresés kezdeténél, tehát a felhasználó előtt nem jelennek meg automatikusan, de megnövelik a keresés eredményességét, sokoldalúságát és gyorsaságát. Ha pedig a szükség úgy kívánja, hozzáíródnak a keresőkérdésekre adott válaszokhoz.

Az Eiffel-torony kulcsszóval végrehajtott keresésre adott válaszként pl. műfajra, korszakra vagy eredetre való tekintet nélkül megkapjuk mindazoknak a filmeknek a listáját, amelyekben az Eiffel-torony látható, akár mint a film fő témája, akár mellékesen, mint amikor pl. egy játékfilm cselekményének a háttérben jelenik meg. A lista minden egyes tétele szerint lehívható egy-egy dokumentumkép, amelynek információi alapján el lehet dönteni, kell-e a film vagy sem.

A dokumentumképeket tartalmazó fájl tehát két fő célt szolgál egyszerre:

- maximális számú megközelítési pontot biztosít az információkereséshez,
- olyan részletességgel írja le az adott filmet, hogy igen specifikus igények esetén is képes a felhasználó eldönteni: számára releváns vagy nem releváns.

A természetes nyelven folytatott információkeresés

A kereséshez, mint láttuk, nem szükséges semmiféle tezaurusz, osztályozási rendszer vagy egyéb mesterséges nyelv ismerete. A felhasználó minitelt billentyűzet segítségével gépeli be a kérdést, majd az ENTER billentyűvel (franciául: ENVOI) küldi el. Erre megjelenik a kérdésre adható válaszok száma. Ha most ismét lenyomja az ENTER billentyűt, akkor megjelenik az összes cím, majd az egyes címeknek megfelelő dokumentumkép. De ha a felhasználó túl soknak találja a kapott válaszok számát, és ezzel összefüggésben úgy véli, hogy kérdése további pontosításra, szűkítésre szorul, akkor párbeszédés üzemmódban további keresőszavakat, adatokat gépelhet be a rendszerbe. Például:

MEG AKAROK NÉZNI EGY BŰNÜGYI FILMET

49 találat

AMELY A MONTMARTRE-ON JÁTSZÓDIK

2 találat

JEAN GABINNEL

Mivel a legutolsó kérdésre csak egyetlen film vonatkozik, az ezt leíró dokumentumkép azonnal meg is jelenik a képernyőn. Ekkor a felhasználó megnyomhatja a VISIONNAGE billentyűt, amely néhány másodpercen belül kikeresteti és elindítja a kívánt filmet tartalmazó kazettát.

A keresések tehát az információs igények függvényében a dokumentumkép 3 fő részének bármelyik elemével végrehajthatók. E módszert létrehozói *kaleidoszkopikus megközelítésnek* nevezik. Ennek értelmében pl. a MARCEL CARNÉ nevével végrehajtott keresés nemcsak a rendező 3 játékfilmjét eredményezi, de felsorolja azokat a játékfilmeket is, amelyekben első asszisztensként tevékenykedett, továbbá a róla készült dokumentumfilmet és televíziós interjút, valamint egyéb filmeket, amelyek valamiféle kapcsolatban állnak a személyével.

A kaleidoszkopikus megközelítésnek a létrehozók alapvető fontosságot tulajdonítanak: ez tudja csak igazán közvetíteni a közönségnek azt a páratlan sokoldalúságot, amellyel az audiovizuális technika az általa feldolgozott témákat képes bemutatni. A keresés megkönnyítésének érdekében nem kényszeríti a felhasználót kérdésének átformálására, de maga a rendszer sem torzíttja el vagy egyszerűsíti le a felhasználótól kapott kérdést. Mindez azonban nem szab gátat a menet közben felmerülő, illetve szükségessé váló pontosító és differenciáló lépéseknek, sőt ezek hatékonyságát is növeli. Így mindenképpen a releváns információszolgáltatás eredményes eszközének bizonyul.

A keresőkérdéseket a fentieknél jóval "profibb" módon is fel lehet tenni. Erre szolgálnak a Boole-algebrára épülő logikai operátorok, amelyek logikai összeadást (halmazegyesítést, jele: OU), logikai szorzást (halmazmetszést, jele: ET), valamint a logikai negálttal történő szorzást (a komplementer halmazzal való halmazmetszést, jele: SAUF) tesznek lehetővé.

A keresés szabadságának és a válaszok minőségének további tökéletesítésére az információkereső rendszer megalkotói egy sor egyéb megoldást is bevezettek; közülük a legfontosabbak:

- a szavak végén előforduló s karaktert (a francia nyelvben a többes szám jele) a program nem veszi figyelembe;
- összeállították az ún. üres szavak listáját, ezek a szavak nem váltanak ki keresést;
- az ekvivalenciák révén (pl. a XI. kerület franciául háromféleképpen is kifejezhető: 19ÈME ARRONDISSEMENT, 19E ARRT vagy XIXE ARRT) a többféle változat bármelyike alkalmas a keresés végrehajtására;
- a szinonimák révén az előző pont mintájára működik a rendszer;
- egyes szócsoportok, szóösszetételek eleve nem bonthatók alkotóelemeikre, tehát az alkotóelemeknek megfelelő egyszerű szavakkal külön-külön nem hajt végre a gép keresést. Pl. a LA BELLE ÉPOQUE nem bontható szét a BELLE (= szép) és

az *ÉPOQUE* (= korszak) szavakkal történő keresésekre, a keresés eredménye így a fentiek értelmében *N. Vedrès Paris 1900* és *M. Allégret Exposition 1900* című filmje lesz.

A videotex rendszert nem a Párizsi Videotékában fejlesztették ki, hanem saját szükségleteikre adaptálták a minitel 1984. évi videotex verzióját. A minitel a nyolcvanas évek végére Franciaország legnépszerűbb lakossági információs rendszerévé, illetve hálózattá vált, és korszakváltást hozott a franciák millióinak információkereső szokásaiban, magatartásában. A köztudatban meggyökeresedett a számítógépes információszolgáltatás fogalma, valamint ennek előnyei a régebbi módzatokkal szemben. Eltűntek a minden különösebb előképzettség nélkül is jól kezelhető minitel terminálokkal szembeni kezdeti előítéletek. Nemcsak a vállalatok vagy hivatalok irodáiban, de a francia családok ezreinek otthonában is van már minitel terminál. Bölcsen választott tehát a Párizsi Videotéka akkor, amikor olvasótermébe minitel terminálokat telepített, hiszen ezeket majdnem mindenki ismeri.

A terminálokon egyetlen – igen egyszerű, de rendkívül hatékony – módosítást végeztek: a billentyűzetet kiegészítették a VISIONNAGE (= megtekintés) billentyűvel. Lenyomására annak a filmnek a lejátszása kezdődik el, amelynek dokumentumképe éppen a képernyőn látszik.

A lejátszás teljes mértékben automatikus. A VISIONNAGE billentyű lenyomására a 6000 db 3/4 BVU videokazettát befogadó MAGNUS típusú robot helyezi el a képmagnóba a kért filmet tartalmazó kazettát. Mivel a minitel terminálokhoz jó minőségű színes monitorok csatlakoznak, ezek kiválóan alkalmasak a videokazetták bemutatására is. A minitel terminálok bevezetésének további nagy előnye az is, hogy az érdeklődők az adatbázisban tárolt információkhoz nemcsak a Párizsi Videotéka 30 terminálján keresztül, de a minitel hálózat valamennyi csatlakozópontján keresztül is hozzáférhetnek (a filmeket azonban a videotékán kívüli terminálokon nem lehet megjeleníteni).

VIDEOTEX-SZOFTVER ADAPTÁCIÓ A VIDEO TÉKA SZÁMÁRA

A Párizsi Videotéka a *Vidiling 5* szoftver 1984. évi verziójából indult ki, amely már eredetileg is megfelelt a videotéka információkereső rendszere által támasztott követelmények többségének. Az adatok lehívása minden esetben 3 másodpercen belül történik.

Az eredeti szoftver azonban bizonyos kiegészítésekre szorult. Ki kellett fejleszteni az azonnali áttérést a megjelenített dokumentumképről a film lejátszására, a dokumentumképek memóriában tartását, valamint hogy egyetlen indexelés után is háromféle visszakeresési módot lehessen a felhasználóknak biztosítani. Ez a három módzat a következő:

- A teljes dokumentumkép megjelenítése, amelyben a 3 alrendszer a megszokott sorrendben követi egymást.
 - A filmről szóló referátum és a filmben közölt adatok megjelenítése azért, hogy a felhasználó eldönthesse, teljes mértékben vagy csak részlegesen, esetleg futólag foglalkozik a film a keresett témával.
 - Az ún. explicit több szempontú keresés, amelyben a felhasználóval a filmnek csak azokat az adatait közlik, amelyekre rákérdez (pl. ki rendezte a filmet, ki volt a zeneszerző, melyik évben forgatták a filmet).
- Az eredeti szoftver kiegészítését jelenti még:
- a fastruktúrájú tezaurusz, amelyet a felhasználó a keresések végrehajtása előtt tanulmányozhat (de ennek használata egyáltalán nem kötelező, mert mint láttuk, megengedett a legteljesebb szabad szöveges keresés is);
 - a szövegszerkesztő a dokumentumképek különböző formátumokban történő kinyomtatásához;
 - az adatbázis anyagából teljes vagy részleges filmográfiák készítése, alfabetikus és tematikus rendezéssel, indexekkel, jegyzetekkel, továbbá biztosítja a fentiek fényesítését és nyomtatását is.
- A feldolgozó munka hagyományos, 80 karakteres terminálok segítségével történik. A feldolgozó állomásokon bevitt változások azonnal érvénybe lépnek, függetlenül attól, hogy ez a friss anyag feldolgozása kapcsán történik-e, vagy pedig a már régebben rögzített adatállományban eszközölnék módosításokat.
- A *Vidiling 5* szoftvert nagy hajlékonysága miatt választották ki, és a Párizsi Videotéka azt állítja, hogy az adaptáció során ezt a hajlékonyságot még tovább lehet fokozni.

A FILMEK DOKUMENTÁCIÓS FELDOLGOZÁSA (INDEXELÉS)

Mint láttuk, a visszakeresésnél a dokumentumkép minden egyes eleme (szava) kulcsszóként működhet, tehát a keresőszóval minden dokumentumképet teljes egészében indexel. Ez azonban nem egyenlő azzal, hogy a dokumentumnak az indexelése is teljesen automatizáltan valósul meg.

Előzetes feldolgozás a gyarapítás szakaszában

A feldolgozási lánc első lépése a film beszerzésével kapcsolatos döntés. A feladattal megbízott dokumentátor megtekinti a filmet, és egy minden műfaj számára egységesen szabványosított, igen részletes táblázat segítségével felveszi a filmre vonatkozó valamennyi lényeges filmográfiai, technikai és esztétikai adatot. Ha a film beszerzése mellett döntenek, akkor a fenti módon kitöltött táblázat az adatbázisba véglegesen bevitt dokumentumkép alapja.

Az input adatok kiválasztása, a dokumentumkép megszerkesztése

Az input adatok kiválasztása és bevitele a dokumentumkép már részletezett hármastagoltasú szerkezetének megfelelően történik.

A filmográfiai tétel (címléírás) terjedelme maximumán három-négy 40 karakteres sor. Szerkezetét igyekeztek úgy kialakítani, hogy a felhasználók számára a legkedvezőbb legyen, azaz:

- ▶ cím,
- ▶ sorozati cím (ha van),
- ▶ a rendező neve,
- ▶ a film műfaja (dokumentumfilm, játékfilm, híradó, reklám, képfelvétel, varieté); a copyright éve; fekete-fehér, ill. színes; időtartam.

Példák (magyarra fordítva):

Az Eiffel-torony

sorozat, az FR3 tagja

rendezte: Fansten, Jacques

dokumentumfilm 1978 színes 1 óra 45 perc

Az alvó Párizs

rendezte: Clair, René

játékfilm 1923 néma fekete-fehér 34 perc

A filmográfiai tétel a legrövidebb, legtömörebb a dokumentumképet alkotó három szerkezeti egység közül. Főként a címjegyzékek megjelenítésénél használják. Ez a dokumentumkép egyetlen olyan része, amelyet nem indexelnek, tekintve, hogy az itt szereplő összes adat valamelyike előfordul a másik két szerkezeti egységben (vagy külön-külön, vagy pedig együtt).

A referátum a gyarapítás során eszközölt előzetes feldolgozás elemeire támaszkodik, feladata, hogy folyékonyan olvasható stílusban alapvető információkat szolgáltatson a film egészéről. Terjedelme legfeljebb 13–14 db 40 karakteres sor lehet. Mivel a referátumot a felhasználó a film megtekintése előtt olvassa el, tájékoztatást kell adnia a film tárgyáról, a cselekmény, ill. a tartalom helyéről, idejéről, valamint a mű egész szelleméről. A Párizsi Videotéka szolgáltatásainak nyilvánosságából következik, hogy a referátumokban szigorúan kerülendő a szaknyelv, ezzel összhangban alapvető követelmény az egyszerű, közérthető és természetes nyelvezet. Ez utóbbi követelmény azonban nem mehet az indexelhetőség rovására, amit két alapvető módszer segítségével védenek ki:

- ▶ az input adatok ellenőrzése, ami azt jelenti, hogy kiválasztják a mondatszerkesztés szempontjából nélkülözhetetlen, ám információt nem hordozó szavakat, és ezeket nem használják fel a referátum indexelésére;
- ▶ a referátum saját szövegével végzett "önindexelés" szisztematikus kiegészítése olyan indexelt információkkal, amelyek a film tartalmára, helyére, idejére stb. vonatkoznak (ezeket a felhasználó számára nem jelenítik meg a képernyőn, ám a tárgyszavas keresésnél ezek a megfelelő összefüggésekben azonnal aktivizálódnak).

A meg nem jelenített indexkifejezések általános, tág értelmű, vagy pedig rokon értelmű fogalmakat

jelölnek, és az indexelőnyelv egyetlen mesterséges, azaz kötött szavas részét alkotják. Ez a kötött szavas alrendszer három tárgyszójegyzékre tagolódik; minden tárgyszójegyzék két hierarchiaszintet foglal magában. Az első hierarchiaszint meglehetősen korlátozott számú tárgyszóból áll: 11 a filmtémák, 6 a műfajok és 2 a színhelyek információira vonatkozik.

Bár, mint láttuk, a kötött szavas nyelvvel indexelt információk nem jelennek meg a referátumban, maga a kötött szavas nyelv igen nagy mértékben segítheti a felhasználó tájékozódását. Különösen akkor, amikor bizonytalan abban, hogyan fogalmazza meg keresőkérdéseit, információs igényeit.

További kiegészítést és keresési pontokat eredményez még az is, hogy a filmeket abból a szempontból is feldolgozzák, milyen párizsi vonatkozású helyszíneket, személyeket és eseményeket tartalmaz. A párizsi vonatkozások összességét külön fájl tartalmazza, amely tovább gazdagítja az információkeresés lehetőségeit.

A film elején vagy végén megadott adatok összesen 16 mezőben közlik a filmben a film készítéséről és forgalmazásáról megadott valamennyi lényeges információt.

A tárgyszójegyzék karbantartása

Alapvetően a gyakorlati tapasztalatokon alapszik, elválaszthatatlanul kötődik az új szerzemények feldolgozásához. Az első ezer dokumentum feldolgozása után a tárgyszójegyzék gyarapodási aránya meredek visszaesést mutatott, ami általánosan ismert, természetes jelenség. A cikk megírásának időpontjában már túl voltak a háromezredik dokumentum feldolgozásán is, és a tapasztalatok szerint minden egyes új dokumentum feldolgozásánál átlagosan egy tétellel gyarapodott, illetve módosult a tárgyszójegyzék.

DOKUMENTÁCIÓS RENDSZERSZERVEZÉS

Az adatbázisok

A dokumentációs lánc minden egyes láncszeméhez hozzá van rendelve egy adatbázis, amely videotex módban tanulmányozható. Ez összesen négy adatbázist jelent, amelyekhez ugyanazt a szoftvert használják. A fenti szerkezet kialakítását a következő funkciók tették szükségessé:

- ▶ a felhasználók számára a filmek megtekintésével összekapcsolt dokumentációs keresés szolgáltatása;
- ▶ azoknak a fejleményeknek a felmérése, értékelése, amelyeket az újonnan feldolgozott tételek adatainak a bevitele, valamint ehhez kapcsolódva a tárgyszójegyzék különböző módosításai jelentenek;
- ▶ a vetítőterem műsorának a beütemezése.

A videotexes adatbázisok a következő elnevezések alatt működnek:

FLUCTUAT: A négy adatbázis közül ez az egyetlen, amelyet a felhasználók (is) lekérdezhetnek. Az olvasóteremben megtekinthető filmek adatait tartal-

mazza, havonként egészítik ki az új szerzemények adataival, ez utóbbiakat a NEC adatbázisból viszik át.

NEC: Belső adatbázis. Azokat a filmeket írja le, amelyeket már beszereztek, vagy a beszerzésük folyamatban van. Felhasználják a tárgyszójegyzék módosításai által kiváltott hatás tesztelésére és a vetítőtermek műsorának programozásában is.

MERGITUR: Belső adatbázis, a megismert, de be nem szerzett filmeket írja le. Nélkülözhetetlen ahhoz, hogy teljes képet kapjanak a Párizsi Videotéka gyarapítási tevékenységéről, segít abban is, ha az eredetileg mellőzött filmet később mégis be akarják szerezni.

JURIS: Belső adatbázis, a filmek bemutatásához szükséges szerzői jogi, forgalmazási és technikai adatokat tartalmazza. Havonként egészítik ki, az input bevitelére IBM PC-ket használnak.

Mivel mind a négy adatbázis ugyanazzal a szoftverrel működik, a fájlokat az egyik adatbázisból át lehet másolni a másikba. A videotéka dokumentátorai minitel termináljaik segítségével a négy adatbázis bármelyikét lekérdezhetik; az adatbázisok közti átlépésre a GUIDE billentyű szolgál.

A FLUCTUAT nyilvános adatbázis feltöltése. Ha egy-egy, az állományban még nem szereplő film megtekintése után a beszerzés mellett döntenek, a NEC-ben hoznak létre neki megfelelő fájlt. Ha elvetik a filmet, akkor a MERGITUR-ban. Az effektív beszerzés után a JURIS-ba beviszik a film forgalmazását meghatározó jogi, és a lejátszáshoz szükséges technikai információkat. Ezek egy része automatikusan átmásolódik a NEC-re (pl. a film pontos időtartama). Amikor a NEC-ben már olyan fájl szerepel, amely a raktárban is ténylegesen rendelkezésre álló dokumentumra vonatkozik, akkor ezt a fájlt – bizonyos ellenőrzések után – automatikusan átmásolják a FLUCTUAT-ra. A NEC-ben módosított tárgyszólistákat szintén átmásolják, természetesen az előzetes ellenőrzések után.

Önellenőrzés és a közönség megfigyelése

A dokumentációs feldolgozás sikerességének ellenőrzésére a fent megnevezett két tevékenység szolgál. Az önellenőrzés mindenekelőtt a NEC adatbázisra épül, mivel ez tartalmazza a legtöbb adatot, és belső használatra is ezt veszik a leggyakrabban igénybe.

A közönség megfigyelését a FLUCTUAT adatbázis segítségével végzik. Rögzítik és naponta kiadják a FLUCTUAT-nak feltett minden egyes felhasználói kérdést, valamint az ezekre adott válaszok számát, továbbá a tanulmányozott fájlokat is. E jegyzékek

tanulmányozása hozzájárul a tárgyszójegyzék tökéletesítéséhez, hiszen jól mutatják, mely kifejezőmódok vezetnek találatokhoz és melyek nem. A közönség "gépi" megfigyelését kiegészítik a felhasználókkal kialakított személyes kapcsolatok. E célból minden dokumentátor napi egy-két órát az olvasószolgálatban tölt el. A Párizsi Videotéka vezetősége meg van győződve róla, hogy a teljesen automatizált információs rendszerek sem működhetnek az emberi kapcsolatok ápolása nélkül.

A KÖZÖNSÉG

Az olvasóterem látogatóinak megoszlásáról a következő főbb adatokat hozták nyilvánosságra: sok a fiatal (67% a 35 év alattiak aránya), az egyetemi hallgató (42%), de a dolgozók aránya is magas (53%), több a férfi (64%), mint a nő. A statisztikák szerint a Párizsi Videotéka közönségének szokásai bizonyos fokig eltérnek az átlagfranciák szokásaitól: az előbbieket ugyanis kevesebb ideig nézik a televíziót, és többet járnak moziba. Legnagyobb részük szórakozni jön a videotékába, de 24% szakmai érdekből, 12% pedig tanulmányai kapcsán.

Mintegy 5000 személynél vizsgálták meg, milyen a viszonyuk a teljesen informatizált eszközökhöz. Ezek a fiatalokból nagy lelkesedést és aktivitást váltottak ki, ugyanakkor a 60 éven felüliek igen nagy féltékenységet tanúsítottak a modern informatikai eszközök használatában. Érdekes szót ejteni a felhasználók információkereső módszereiről is: 28% tárgy, 22% név, 15% cím, 10% műfaj, 10% helyszín, 4% pedig időpont alapján keresett. A felhasználóknak mindössze 11%-a keresett több szempont kombinációjával.

Egy 1988 júliusában végzett felmérés szerint a videotékában megnézhető öt műfaj közül a két legnépszerűbb – szinte egyenlő arányban – a dokumentumfilm (43%) és a játékfilm (42%) volt. Az 1988 végi kép ettől bizonyos fokig már eltér: a legkeresettebbeknek a nyolcvanas években készült televíziós reklámok és videoklipek, továbbá bizonyos játékfilmek (*J. L. Godard: Kifulladásig*, *R. Lester: Superman II.*, *L. Malles: Zazie a metróban*) bizonyultak. A dokumentumfilmek pozíciója a fentieknél gyengébb volt; közülük a legjobb helyezést *R. Depardon* Reporters c. munkája érte el (az összesített mezőnyben viszont ez az alkotás csupán a tizedik helyen végzett).

/FOURNIAL, C.: Vidéotheque et videotex, le système documentaire de la Vidéotheque de Paris. = Documentaliste, 26. köt. 1. sz. 1989. p. 3–10./

(Sebestyén György)

Az online keresés költségei a Data-Star, a Dialog, az ESA-IRS és az STN host esetében*

Az online hostok árainak összehasonlítása soha sem volt egyszerű, mostanában még bonyolultabbá lett. A hostok is megváltoztatták áraik szerkezetét, s szolgáltatásaik különböző távközlési hálózatokon keresztül érhetők el, amelyeknek szintén saját árrendszerük van.

A hostok kapcsolódási árai eltéréseinek megismerése céljából a Finnországi Műszaki Kutatóközpont (VTT) Információs Szolgálatát összehasonlító vizsgálatot végzett 1990 tavaszán. A következő négy hostot vizsgálták: *Data-Star*, *Dialog Information Service*, *ESA-IRS* és *STN International*. Az összehasonlításra hat olyan adatbázist használtak fel, amelyek mind a négy hostnál elérhetők voltak. Ezek a következők: *BIOSIS*, *Chemical Abstracts (CA)*, *COMPENDEX*, *FSTA*, *INSPEC* és *NTIS*. A kutatás célja volt, hogy megállapítsák, hogyan hatnak a különböző ár-szerkezetek az online keresés különböző típusaira, s hogyan lehet egyszerűen megbecsülni a keresés teljes költségét.

Hagyományosan az online keresés költségei a hostokhoz való kapcsolódási időből és az output árából tevődnek össze. Több nagy rendszer is csökkentette a kapcsolati idő árát, de az egyéb költségeket növelte. A legnagyobb változás az ESA-IRS árrendszerében következett be 1989 elején. Gyakorlatilag minimálisra csökkentették a kapcsolati idő díját, helyette egy ún. "terminálülési díjat" (session rate) vezettek be, ami azt jelenti, hogy ennek fejében tetszőleges ideig lehet egy adatbázisban maradni keresés céljából. Ezzel együtt az output díjait emelték meg. Az ESA-IRS megfogalmazása szerint az információt fizettetik meg, nem pedig annak keresését.

A hostok díjai mellett számolni kell a távközlési díjakkal is. Ezek országonként változnak. Az országos csomagkapcsolt hálózatok díjai általában két részből állnak: a kapcsolati idővel és az átvitt adatok mennyiségével arányos részből. Nagyobb sebesség esetén ez utóbbi a domináns. A most vizsgált négy host elérhető mind 1200, mind 2400 baud sebességgel.

A keresési fázis költségei

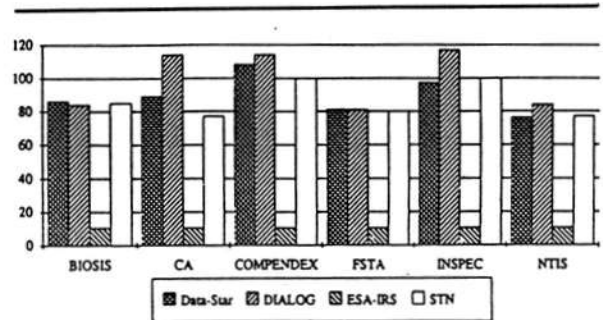
A keresési fázis költségei a kapcsolási idővel arányos díjakból (valamennyi rendszer), a "terminálülési díj"-ből (csak az ESA-IRS-nél), a mintarekordok kiíratási díjából, és bizonyos egyéb költségekből tevődnek össze. A költségek 1 perc keresési időre vonatkoznak.

* Az online keresés költségadatai 1990-ből valók. A mondanivaló lényege azonban azóta sem változott, az arányok is csak kismértékben, annak ellenére, hogy a tényleges árak és költségek növekedtek az egyes hostoknál. A költségek összetétele és az egyes hostok összehasonlításából adódó következtetések ma is helytállóak. — A szerk.

A keresési fázis távközlési díjait az alábbi óradij alapján állapítottuk meg (zárójelben az átviteli hálózat neve szerepel):

	1200 baud	2400 baud
Data-Star (DATAPAK)	12,50 \$	13,75 \$
Dialog (DIALNET)	25,50 \$	25,50 \$
ESA-IRS (ESANET)	00,00 \$	00,00 \$
STN (DATAPAK)	12,50 \$	13,75 \$

Ehhez adódik hozzá az a díj, amelyet a hostok állapítanak meg adatbázisonként. Az ESA-IRS-nél az egyetlen, a kapcsolati időtől függő díj egy óránként 10,47 dolláros "háztartási díj". A "terminálülési díj" a domináns költség rövid keresések esetén. Ha a keresési fázis hosszabb, a díjak az ESA-IRS esetében sokkal lassabban nőnek, mint a korábbi hostnál. Már egy tízperces keresés is feleannyiba kerül, mint az utána következő Data-Star esetében. Minden adatbázisnál az ESA-IRS a legolcsóbb, s általában a Dialog a legdrágább. Az ESA-IRS használata hosszabb keresések esetében egyre előnyösebb, ez az előny öt percnél rövidebb keresések esetében eltűnik. Az 1. ábra mutatja a kapcsolati díjak alakulását hostonként és adatbázisonként, a 2. ábra a keresési fázis átlagos költségeit hostonként és adatbázisonként.



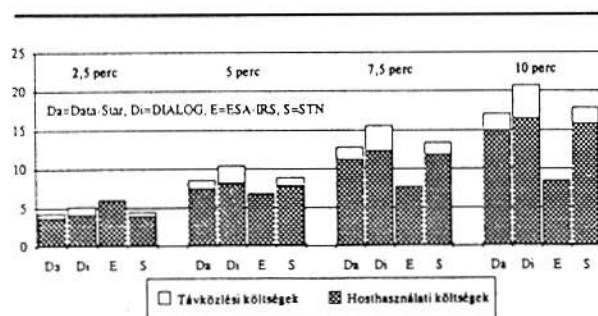
1. ábra Az online kapcsolati óradijak USD-ben (1990. május) a négy hostnál a vizsgált hat adatbázisra

Az output fázis díjai

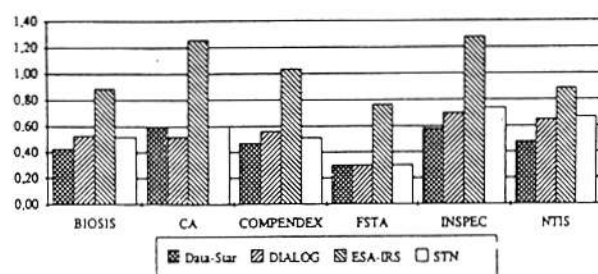
A kiírt vagy kinyomtatott rekordok árának a következő összetevői vannak: a használt formátum díjtétele, a kapcsolati idő díja és a távközlési díjak. 6–10 rekordra mérték az egy rekord kiírásához szükséges kapcsolási időt az összes adatbázisra.

Az output fázis távközlési díjait a következő óradij alapján állapították meg:

	1200 baud	2400 baud
Data-Star (DATAPAK)	53,50 \$	101,25 \$
Dialog (DIALNET)	25,50 \$	25,50 \$
ESA-IRS (ESANET)	0,00 \$	0,00 \$
STN (DATAPAK)	53,50 \$	101,25 \$



2. ábra A keresési fázis átlagos költségei USD-ben a kapcsolati idő függvényében. Az átlagérték hat adatbázisra vonatkozik: BIOSIS, CA, COMPENDEX, FSTA, INSPEC és NTIS



3. ábra Egy online megjelenített teljes rekord ára a hat adatbázisra, a hostok árjegyzéke szerint USD-ben (1990. május)

1. táblázat

A rekordonkénti megjelenítési (kiíratási) idő s-ban, teljes és közepes hosszúságú formátumban, 1200 baud sebességgel (A megfelelő értékek 2400 baudon a fenti adatok fele.)

Teljes formátum	Data-Star	Dialog	ESA-IRS	STN
BIOSIS	21,50	18,50	23,18	20,25
CA	15,20	11,30	18,00	14,40
COMPENDEX	11,60	13,00	17,00	13,40
FSTA	11,20	11,20	13,60	14,40
INSPEC	15,60	14,50	17,80	15,20
NTIS	15,00	13,50	14,73	14,70
Átlag	15,02	13,67	17,38	15,39
Közepes formátum	Data-Star	Dialog	ESA-IRS	STN
BIOSIS	13,40	12,50	23,18	14,60
CA	15,20	11,30	18,00	14,40
COMPENDEX	9,20	10,40	17,00	10,70
FSTA	9,00	9,00	13,60	11,50
INSPEC	10,40	9,75	17,80	10,20
NTIS	13,50	10,25	14,73	9,60
Átlag	11,78	10,53	17,38	11,83

Az átvitt adatok mennyisége jelentős tényezője az output fázis díjának. 1200 baud sebességen a DATAPAK-hálózat távközlési díjai több mint kétszeresei a DIALNET-ének, 2400 baud esetében ez már a négyszeresénél is több. Meg kell jegyezni, hogy a keresési fázisban a DATAPAK az olcsóbb.

A 3. ábra mutatja az egyes adatbázisok találati díjait a különböző hostoknál. Átlagosan a Dialog egy rekordra jutó költségei a legalacsonyabbak, különösen 2400 baud esetében, mert akkor az output ideje, s következésképpen a kapcsolati idő és a távközlési díj fele az 1200 baud esetén jelentkezőnek. A Data-Star és az STN esetében is olcsóbb egy rekord 2400 baud sebességen, de itt kisebb a különbség a magas távközlési díj miatt. Az ESA-IRS-nél majdnem azonos egy rekord költsége a két sebességen, mivel a kapcsolati idő olcsó, s nincsenek távközlési díjak.

A költségeket az egyes találatok díja is meghatározza, de az egy találatra jutó díjat az is befolyásolja, mennyi idő alatt lehet egy tételt kiíratni az adott rendszerben, s ez mibe kerül. Így a legtöbb adatbázisnál hiába a legkisebb a Data-Star találati díja, az egy találatra eső ár mégis kevesebb a Dialog esetében. Ennek fő oka a DIALNET olcsóbb ára. A 2400 baud sebesség árcsökkentő hatása különösen azoknál az adatbázisoknál érvényesül, amelyeknek hosszúak a rekordjai, és nagy a kapcsolati idő díja. Az 1. táblázatban láthatók az egyes adatbázisok rekordjainak megjelenítéséhez szükséges idők (s-ban) a különböző hostoknál, 1200 baudon, 2400 baudon ezek az idők feleződnek.

Ha az indexkifejezések nélküli (ún. középhosszúságú) rekordot használjuk a teljes helyett, az összköltséget a Data-Star esetében 15%-kal, a Dialognál és az STN-nél pedig 11%-kal lehet csökkenteni; ez a szám 2400 baudnál 8, illetve 7%. Ennél is jóval nagyobb megtakarítás érhető el, ha a találati halmazból a címek alapján kiválasztjuk a releváns tételeket, s csak ezeket nyomtatjuk ki teljes terjedelemben, vagy a hostok által ajánlott offline nyomtatás lehetőségét választjuk.

Az online keresés teljes költsége

Az online keresés teljes költsége a két fenti fázisból tevődik össze, a keresésből és az outputból. A teljes költséget mindkét sebesség esetében hat

különböző mintakeresésben vizsgáljuk meg: 10 és 30 perces keresésekkel, 25, 100 és 200 találattal. A 2. és 3. táblázatban minden adatbázisban és minden egyes mintakeresésben indexként a legalacsonyabb árat vettük 100-nak. (A tényleges minimális árakat az összes mintáról az alsó sorban tüntettük fel. A 110 alatti indexértékeket dőlt betűvel szedtük.) A viszonylag durva közelítések miatt úgy értékelhetjük, hogy az aláhúzott alternatívák egyformán gazdaságosak.

1200 baudnál a CA kivételével minden adatbázisra az ESA-IRS ajánlja a leg gazdaságosabb megoldást, ha a találatok száma alacsony (max. 25), s akkor is, ha a keresési fázis hosszú (30 perc). Ez különösen akkor érvényesül, ha ez a kettő egyszerre következik be, mint a 4-es típusú mintakeresésben. Rövid

2. táblázat

Mintakeresések összköltségének összehasonlítása 1200 baud sebességen

(A legkevesebb költségű keresést minden adatbázisban 100-ra mint indexre vettük, s ehhez viszonyítottuk a többi. A legkisebb költségeket USD-ben is megadtuk minden példa utolsó sorában.)

	BIOSIS	CA	COMPENDEX	FSTA	INSPIC	NTIS	Átlag
1. példa: 10 perces keresési fázis + 25 teljes rekord							
Data-Star	150	100	133	119	117	128	122
Dialog	142	117	146	119	133	142	132
ESA-IRS	100	105	100	100	100	100	100
STN	161	105	121	137	129	143	130
Minimum (USD)	32	42	36	28	42	32	35
2. példa: 10 perces keresési fázis + 100 teljes rekord							
Data-Star	138	101	111	108	100	115	105
Dialog	123	110	119	100	107	123	107
ESA-IRS	100	130	102	109	100	100	100
STN	143	100	100	134	113	133	112
Minimum (USD)	104	109	116	89	142	102	116
3. példa: 10 perces keresési fázis + 200 rekord							
Data-Star	135	103	111	111	100	112	101
Dialog	119	109	119	100	105	118	102
ESA-IRS	100	138	107	117	104	100	100
STN	140	100	100	140	114	131	108
Minimum (USD)	200	197	213	169	275	195	223
4. példa: 30 perces keresési fázis + 25 teljes rekord							
Data-Star	214	145	212	191	176	188	185
Dialog	217	184	239	203	212	218	211
ESA-IRS	100	100	100	100	100	100	100
STN	237	162	195	204	188	201	195
Minimum (USD)	38	51	41	34	49	37	42
5. példa: 30 perces keresési fázis + 100 teljes rekord							
Data-Star	160	100	137	126	120	137	127
Dialog	150	115	150	123	134	150	135
ESA-IRS	100	104	100	100	100	100	100
STN	171	103	124	148	133	154	135
Minimum (USD)	110	149	123	94	149	107	122
6. példa: 40 perces keresési fázis + 200 teljes rekord							
Data-Star	147	100	119	109	107	124	113
Dialog	133	110	129	103	116	133	117
ESA-IRS	100	118	100	100	100	100	100
STN	154	100	108	133	121	142	121
Minimum (USD)	205	237	233	175	281	201	229

keresés és közepes output esetében nincs lényeges különbség az átlagos költségekben a hostok között, míg sok találat esetén a helyzet adatbázisról adatbázisra változik. A CA-t az STN-ben a legérdekesebb keresni, míg a BIOSIS-t és az NTIS-t az ESA-IRS-ben.

2400 baudnál az ESA-IRS gazdaságossága kevésbé feltűnő. Hosszú keresési idő és kevés találat esetén ekkor is ez a leggazdaságosabb host, rövid keresési fázis és sok találat esetén viszont már a Dialog és a Data-Star. Az STN árai közelebb kerülnek ugyan a többi hosthoz, de ez sohasem lesz igazán alacsony.

Amint a 3. táblázatból is látható, minden egyes adatbázisra igaz, hogy nincs egyetlen host, amely mindenfajta keresésre a leggazdaságosabb választás

lehetne. A BIOSIS-ban való rövid keresés és sok output esetén a Dialog kínálja a legalacsonyabb árakat, de az összes többi esetben az ESA-IRS a leggazdaságosabb. A CA keresésére a Dialog a leggazdaságosabb, kivéve a hosszú kereséseket kevés találattal. Ha sok a találat, a COMPENDEX esetében sok kiíratott találat esetén, nincs lényeges különbség, de kevés találat esetén már az ESA-IRS a legcélszerűbb. Az FSTA keresése legolcsóbb a Dialogban és az ESA-IRS-ben. Az INSPEC adatbázisra sok találat esetén a Data-Star és a Dialog a leggazdaságosabb host, de minden hosszabb keresésben a legjobb választás az ESA-IRS. Az NTIS-hez is az ESA-IRS választása előnyös, de rövid keresés és sok találat esetén a Data-Star és Dialog is gazdaságos lehetőségek.

3. táblázat

Mintakeresések összköltségének összehasonlítása 2400 baud sebességen

(A legkevesebb költségű keresést minden adatbázisban 100-ra mint indexre vettük, s ehhez viszonyítottuk a többi.

A legkisebb költségeket USD-ben is megadtuk, minden példa utolsó sorában.)

	BIOSIS	CA	COMPENDEX	FSTA	INSPEC	NTIS	Átlag
<i>1. példa: 10 perces keresési fázis + 25 teljes rekord</i>							
Data-Star	133	101	117	110	106	117	113
Dialog	123	100	126	106	118	128	116
ESA-IRS	100	100	100	100	100	100	100
STN	145	106	118	116	117	133	121
Minimum (USD)	31	42	35	28	41	31	35
<i>2. példa: 10 perces keresési fázis + 100 teljes rekord</i>							
Data-Star	116	121	100	115	100	101	106
Dialog	100	100	103	100	103	105	100
ESA-IRS	101	142	113	135	116	100	114
STN	129	119	106	127	113	119	115
Minimum (USD)	99	97	101	64	120	99	99
<i>3. példa: 10 perces keresési fázis + 200 rekord</i>							
Data-Star	119	127	100	120	100	100	105
Dialog	100	100	102	100	100	102	100
ESA-IRS	107	155	121	149	122	103	123
STN	126	123	108	134	114	120	119
Minimum (USD)	181	172	182	111	221	185	177
<i>4. példa: 30 perces keresési fázis + 25 teljes rekord</i>							
Data-Star	202	152	201	185	168	180	179
Dialog	203	175	223	194	200	207	199
ESA-IRS	100	100	100	100	100	100	100
STN	227	168	195	189	179	194	190
Minimum (USD)	37	52	40	34	48	37	41
<i>5. példa: 30 perces keresési fázis + 100 teljes rekord</i>							
Data-Star	140	105	118	115	107	124	117
Dialog	128	100	126	108	117	134	117
ESA-IRS	100	102	100	100	100	100	100
STN	153	109	121	123	119	142	126
Minimum (USD)	106	144	120	92	146	105	119
<i>6. példa: 40 perces keresési fázis + 200 teljes rekord</i>							
Data-Star	125	116	100	112	100	110	104
Dialog	109	100	104	100	104	115	100
ESA-IRS	100	126	102	117	107	100	103
STN	134	116	105	123	113	129	113
Minimum (USD)	199	218	223	146	258	196	218

Következtetés

Az ESA-IRS új árpólitikájának legfőbb előnye, hogy felszabadít bennünket az idő szorítása alól, sokkal interaktívabb keresésre ösztönöz. Minden gyakorlott kereső tudja, hogy a bonyolult kérdések megválaszolása csak iterációs módszerrel oldható meg. A gyakorlatlan végfelhasználók számára is előnyös az ESA-IRS árstruktúrája. Ők ugyanis még az egyszerűbb keresésben is lassúak, s az alacsony kapcsolati díj nem bünteti őket lassúságukért.

A használók számára az az árrendszer előnyös, amelyben előre könnyen ki lehet számítani a keresés teljes költségét. Mivel az ESA-IRS-nél a "terminálulási" és a találati díj az alapvető összetevő, ez könnyen becsülhető. Hagyományos árstruktúra esetén, mint amilyen a Dialogé, a használónak mind a kapcsolati időt, mind a találatok számát figyelembe kell venni a költségbecsléshez. A Data-Starnál még nehezebb a számítás, mert a távközlési díjakat az átvitt adatcsomagok száma is befolyásolja, míg az STN-nél néhány adatbázisban (pl. a CA-ban) még speciális díjak is vannak.

A hagyományos árrendszerű hostok különféle megoldásokat ajánlanak az online keresés árának optimalizálására (nagyobb kapcsolati sebesség, rövidebb output formátumok, offline nyomtatás, szelektív output). Az ár akkor optimális, ha minimalizálják a

kapcsolati időt, s szelektív outputtal csökkentjük a találati díjat. Az ESA-IRS-nél csak ez utóbbi árcsökkentő tényező, itt az alacsonyabb ár sokkal inkább a keresési technikákon és stratégiákon múlik, mint az output formátumokon. Az alacsony kapcsolati díjú rendszerekben gazdaságos többlet keresési idővel finomítani a keresést, s csak a releváns tételeket kinyomtatni (erre ad módot az ESA-IRS ZOOM lehetősége).

A nemzetközi adatbázisok gyakorlott használói akkor oktatták az online keresésre, amikor a legfőbb szempont a keresési idő csökkentése volt. Ettől megszabadulva egy sokkal interaktívabb – és sokszor jobb – használati mód alakítható ki. Ahogyan D. T. Hawkins az információ-visszakeresés árstruktúrájáról megállapítja (In search of ideal information pricing. = Online, 13. köt. 2. sz. 1989. p. 15–30.), az árak kialakításának arra kell törekedniük, hogy ne időegységeket és számítógép-teljesítményt adjanak el, hanem információk biteket és bájtokat.

/ESKOLA, P. – SORMUNEN, E.: Cost comparison of online searching in four hosts: Data-Star, Dialog, ESA-IRS and STN. = Online Review, 14. köt. 5. sz. 1990. p. 303–316./

(Murányi Péter)

Vegyületekkel kapcsolatos információk számítógéppel segített keresésének stratégiája

A vegyületek pontos kémiai adatai alapvető fontosságúak, ha fel akarjuk becsülni a környezetre gyakorolt veszélyes hatásukat. Az adatgyűjtés nehéz feladat. Azonkívül, hogy a kémia alapos ismerete szükséges, a keresőnek tisztában kell lennie a számítógépes információforrásokkal, a bibliográfiai, numerikus vagy faktografikus és a teljes szövegű adatbázisokkal.

A probléma definiálása

Először is tisztázni kell, vajon közvetlenül visszakereshető-e az információ, vagy több lépés szükséges. A kémia területén a legtöbb számítógépes adatbázisban a CAS nyilvántartási száma (Chemical Abstracts Service Number) nem kereshető mező, a keresőnek a kémiai nevet kell használnia az adat visszakeresésére. Ez a keresést bonyolultabbá és nehezebbé teszi.

1988-as adatok szerint 3000 online adatbázis létezett, ezek közül körülbelül 200 foglalkozik vegyületekkel.

A probléma megoldására korábban több mint 100, vegyületekről információt tartalmazó adatbank és adatbázis elérésére alkalmas rendszert fejlesztettek ki. Ezek az adatbázisok hat központon keresztül

érhetőek el – CIS, Datacentralen, Data-Star, DIMDI, ESA, STN –, és mindegyiknek saját speciális keresőnyelve van.

Ezenkívül minden adatbázisnak megvan a maga struktúrája és több más sajátossága is. Ahogy a több adatbázisban való párhuzamos keresés mindinkább elterjed, az ehhez kapcsolódó számítógéppel segített stratégia kialakításának szükségessége is nyilvánvalóvá válik. Két megközelítést fontolhatunk meg. Vagy egy parancsnyelvfordítót fejlesztünk ki, amely a keresési parancsokat lefordítja a különböző központok parancsnyelvére, vagy előre definiált keresések halmazát állítjuk fel, majd megfelelő stratégia segítségével az optimális keresési utat azonosítjuk. Itt ez utóbbi megközelítés érvényesül.

Nézzünk egy olyan példát, ahol meghatározott kémiai anyag víztoxicitásáról kell információ, de csak a márkanév ismert. Az 1. ábra bemutat egy közvetlen keresést, amellyel kicsi a valószínűsége, hogy adatot találunk; mutat két alternatív keresési lépést, amellyel a CAS-nyilvántartási szám található meg; és három adatbázis kiválasztását, amelyek tartalmazzák az adatot.

Az ábra irányított gráfot mutat. A csúcsok azt reprezentálják, milyen jellegű információ szükséges