

Beszámolók ◦ Szemlék ◦ Közlemények

Egy adatbázis nem elég –
egynémely azonban túl sok!

Egy vizsgálat és eredményei

A számítógépi adatbázisok lekérdezése jól előkészített keresőstratégiát igényel, a sikeres válasz érdekében. Az alig átgondolt, kapásból odavetett, szakszerűtlen keresőkérésre az adatbázistól sem várható értelmes felelet. Az ilyen online keresést nevezi a szakirodalom „gyors és piszkos” (quick and dirty) keresésnek. De vajon a gondosan megfogalmazott kérdésre adhat-e egyetlen adatbázis kielégítő, teljes, átfogó választ?

Ennek kiderítésére összehasonlító vizsgálatot végeztek, melynek során az amerikai Lockheed rendszerhez tartozó adatbázisokat egymással és néhány NSZK-beli adatbázissal vetették össze. A vizsgálatba a következő adatbázisokat vonták be:

• német adatbázisok

BGR/AIM	Bundesanstalt für Geowissenschaften und Rohstoffe (FIZ 6) Arbeitsgemeinschaft Information Meeresforschung und Meerestechnik A. I. M.
BAM(S)	Bundesanstalt für Materialprüfung Dokumentation Schweistechnik (FIZ 5)
BAM(T)	Bundesanstalt für Materialprüfung, Dokumentation Tribologie, Verschleiss, Reibung und Schmierung (FIZ 5)
NUCLear	Fachinformationszentrum Energie, Physik, Mathematik (FIZ 4)
SDIM	Système de Documentation et d'Information pour la Métallurgie (FIZ 5)

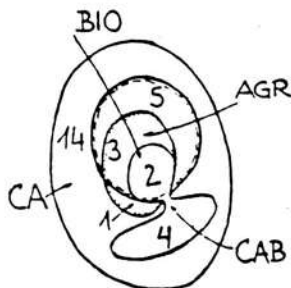
• nem-német adatbázisok

AGRicola	Agricultural Online Access
BHRA Fluid	Engineering Abstracts
BIOsis	Biological Abstracts
CAsearch	Chemical Abstracts
CAB	Commonwealth Agricultural Bureau Abstracts
COMPendex	Engineering Index
CONFerence	Papers Index
FSTA	Food Science and Technology Abstracts
INSPEC	Information Services in Physics, Electronics, Computers and Control
ISMec	Information Service in Mechanical Engineering
METadex	Metals Abstracts
METGEO	Meteorological and Geostrophysical Abstracts
NTIS	National Technical Information System
OCEANic	Abstracts (Data Courier)
POLLUTION	Abstracts (Data Courier)
SCISearch	Science Citation Index
WELDasearch	Database of The Welding Institute
World TEXTiles	Shirley Institute

A kapott eredményeket halmaz-diagramokon ábrázolták.

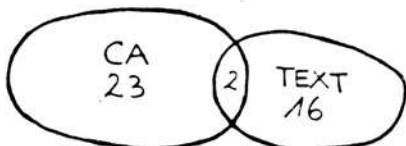
Az 1. ábrából kitűnik a *Chemical Abstracts* régóta tapasztalt minőségi fölénye: a bromoxynil gyomirtóra vonatkozó kérdésre a *Chemical Abstracts* 29 találatával szemben a másik 3 adatbázis összesen csak 15 találatot szolgáltatott. Ezek is a *Chemical Abstracts* információinak, sőt részben egymás tételeinek is átfedései.

Hasonló eredményre vezetett egy adott szerző műveire vonatkozó keresőkérdés. A 7 adatbázisban összesen 226 volt a releváns találatok száma. Ebből 80 a Chemical Abstract-ben szerepelt, a többi 6 adatbázis ehhez összesen csupán további 8 információs tételt szolgáltatott – a fennmaradó 138 találat mind átfedés volt.



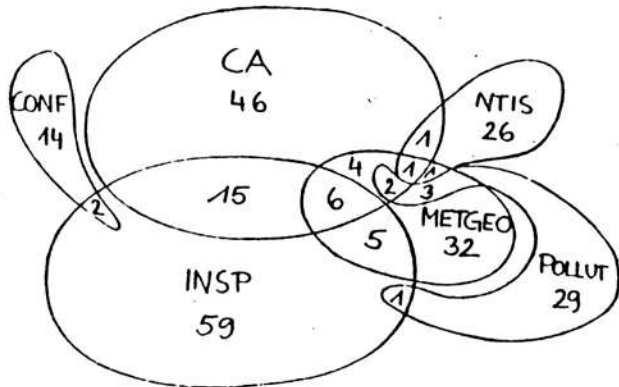
1. ábra

Amint azonban a kérdések a vegyi anyagok területéről távolodnak, mind kisebb lesz a Chemical Abstracts által szolgáltatott információk részaránya. Ez világosan kitűnik a 2. ábrából. A titán-dioxidos szulfénytelenítésre vonatkozó válaszok közül a Chemical Abstracts 25 és a World TEXTiles 18 találatából csupán 2 fedt egymást.



2. ábra

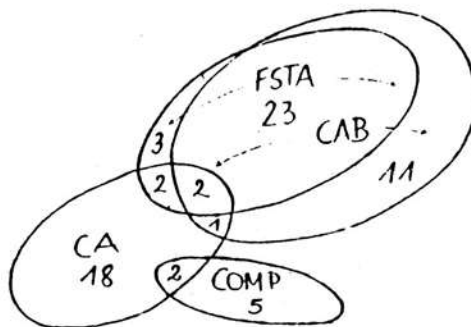
A 3. ábra egy lidaros (lézer-radaros) légszennyezés-mérésre, a 4. ábra az ultraszűrővel foglalkozó négy szerző műveire keresett válaszok megoszlását mutatja. Bár a 4. ábrán két adatbázis viszonylag nagy – 23 találatot felölelő – átfedést mutat, kitűnik, hogy minden további adatbázis újabakkal növeli az elsőből kapott információk mennyiségét.



3. ábra

Fentiek csak az amerikai adatbázisokra vonatkoztak. Ezeket az NSZK egyes szakinformációs központjai (Fachinformationszentren – jelölésük: FIZ 4, 5, 6) adatbázisaival egybevetve (5–8. ábra) megállapítható, hogy az átfedés viszonylag kevés, az adatbázisok tehát mintegy kiegészítik egymást, és önmagukban egyik sem tekinthető teljesnek.

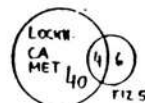
Különösen meglepő a 8. ábrán látható eredmény. A kérdés a FIZ 6 szakterületéhez tartozó vizalatti hegesztésre vonatkozott, a találatok számát tekintve mégis az amerikai szolgáltatások vannak túlsúlyban. A FIZ 6 mellett szól, hogy az amerikaiaktól eltérő forrásokat dolgozott fel, míg az utóbbi találatai viszont egymás között nagymértékű – 40%-os – átfedést mutatnak.



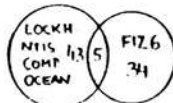
4. ábra



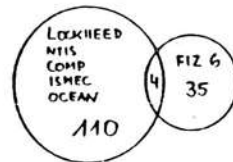
5. ábra



6. ábra



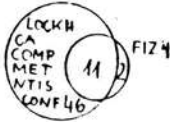
7. ábra



8. ábra

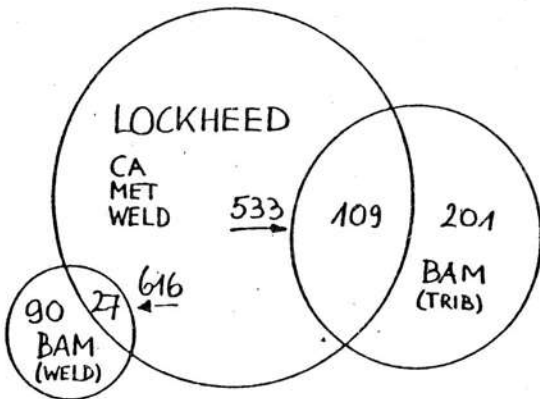
A 9. ábrán 5 amerikai adatbázis szerepel. Ismét anyagra vonatkozó kérdésről lévén szó, a találatok zömét a Chemical Abstracts szolgáltatta, a többi adatbázis lekérdezése szinte meg sem érte a fáradságot. A FIZ 4 13 releváns tételéből 11 a Lockheed adatbázisokban is szerepelt. Ha ezt a 13 találatot a Lockheed adatbázisokból kizárólagosan kapott összes találatához, a 46-hoz

viszonyítjuk (leszámítva az átfedést), kiderül, hogy a felhasználó a FIZ 4-től a létező információknak csupán 27%-át kapta meg, ez pedig a szakember számára nagyon kevés. Ilyen szempontból azonban hasonló bírálat érheti az egyes Lockheed-adatbázisokat is, melyek egyenként ugyancsak kevés információt szolgáltatnak.



9. ábra

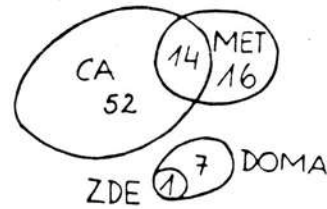
Egy, a hegesztés és tribológia területén végzett irodalomkutatásban 2 német és 4 amerikai adatbázis vett részt (10. ábra). A 6 adatbázisból külön-külön egy sem nyújtott többet a teljes információmennyiség 36%-ánál. Az átfedések mennyisége meglepően kevés: a találatok összmenyiségének csupán 22%-a szerepel két- vagy több szolgáltatásban. Az egyik német adatbázis a szolgáltatott 310 információs tétel ellenére sem adott a kérdésre kimerítő választ. Még inkább elgondolkodásra készítet a speciális metallurgiai adatbázisnak számító METadex 147 találat, amelyeknek több mint a fele szovjet folyóiratokból származik, így ezeknek angol fordításai természetesen találati duplumnak tekintendők. Figyelemre méltó a Chemical Abstracts viszonylag magas, 30%-os részaránya, ami egy technológiai vonatkozású téma esetén eléggé meglepő, és nyilván a hegesztésnél felhasználandó anyagokkal magyarázható.



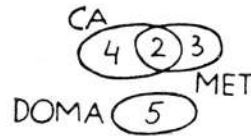
10. ábra

Az ólom ultrahangos kezelésének tárgyában az „ultrahang” és „ólom” keresőstratégiára a Chemical Abstracts és a METadex válaszai a két fogalomnak legalább 8-féle összefüggésére vonatkoznak, ezek közül kettő a DOMA és ZDE nevű német szolgáltatások válaszaiban is szerepelt. A 11. és a 11/a. ábra tanúsága szerint mind a különböző összefüggésekre, mind az erősen leszűkített fogalomra vonatkozó információs tételekben a két amerikai szolgáltatásban átfedések vannak, ilyenek azonban a német és az amerikai források

között egyáltalában nem tapasztalhatók. A valószínű magyarázat az, hogy a német adatbázisok elsősorban a német és a szovjet szakirodalomra támaszkodnak, míg amerikai, francia, kanadai és japán források csak elvétve találhatók bennük.



11. ábra



11/a ábra

A felhasználó hátránya

A vizsgálat csupán 14 keresőkérdésre (12 szakmai témára, egy adott szerzőre és egy témából 4 szerző műveire) szorítkozott, így eredményei, azaz egyes példái természetesen csak önmagukért beszélhetnek, az adat-szolgáltatások alapvető értékeléséhez nem elegendők mint statisztikai mennyiség. A mindennapi gyakorlat ismétlődő tapasztalatai azonban *negatívan befolyásolhatják* a felhasználói magatartást, különösen akkor, ha a felhasználó az adatbázisokból kevés releváns adatot kap. Tovább nő a bosszúsága akkor, amikor kézhez kapja a keresés eredményét, és ebből ki kell válogatnia a duplumokat, melyeket nyilván kétszeresen vagy többszörösen fizetett meg. Ha viszont csupán egyetlen információs központhoz fordul, sok esetben csak korlátozott válasza hagyatkozhat. Ha kutatásaiba bevonja az amerikai adatbázisokat is, észre kell vennie, hogy egyetlen szolgáltatás önmagában nem elég a teljes tájékozódáshoz.

Az NSZK szakinformációs központjainak (FIZ) hátránya

Az elektronikus adatfeldolgozás előretörésével mindinkább figyelembe kell venni a *gazdaságosság* kérdését. Törekedni kell az információs rendszerek egyszerűsítésére, egységesítésére, a találati biztonságra. Néhány többé-kevésbé releváns adatból álló hevenyészett címjegyzék a hagyományos eszközökkel is gyorsan összeszedhető. Az automatizálás lényege éppen a nagyon gyors, lehetőség szerint teljességre törekvő és pontos informá-

ciószolgáltatás. Több adatbázis igénybevétele esetén azonban nemcsak a kapcsolati órák száma és a munkaidő költsége növekszik, de nehézséget okoz az eltérő parancsnyelvekhez, a szabályozott szókinccsekhez és kódokhoz való alkalmazkodás szükségessége is. A duplumok pedig gépi úton csak nehezen szűrhetők ki.

Az európai országos információshálózatoknak számolniuk kell az amerikai versenytársakkal, erre a versenyre azonban az NSZK FIZ-jei, adatbázisai *nincsenek kellően felkészülve*. Anyaguk korlátozott, főleg a német nyelvre épül, működtetésük nem olyan egyszerű és kényelmes, mint az amerikaiaké. A hardver alulméretezett, keresőrendszerei (pl. a DIRS 3) pedig jó minőségük ellenére sem érik el az amerikaiak színvonalát (pl. a DIALOG, STAIRS).

Kielégítő információszolgáltatásért ma még elsősorban az amerikai adatbázisokhoz kell fordulni. A Pergamon Infoline megvásárlásával és az amerikai támogatással kiépült svájci (Radio Suisse) szolgáltatás bevezetésével az NSZK információshálózatai még inkább háttérbe szorulnak. Ebből az előnytelen helyzetből csak gyökeres változtatásokkal lehet kilábalni.

A bemutatott halmaz-diagramok tanúsága szerint arra van szükség, hogy a keresést több adatbázisban végezzük, és a kapott választ kisebb adatbázisokból keresett információkkal egészítsük ki. Ezek a kiegészítések azonban viszonylag drágák, különösen akkor, ha adott esetben mindig újabb adatközpontokhoz kell kapcsolódni, és új kérdésfeltevésével kell az információkat keresni. Az európai adatbázis-központok szövetsége nem pótolhat egyetlen nagy és jól szervezett adatbázis-központot, melyben a feltett kérdés automatikusan kerül tovább az adatbázisok online jegyzékéből előzetesen vagy menetközben kiválasztott megfelelő adatbankhoz.

Az egy-egy tudományág vagy fölérrendelt szakterület teljes szakirodalmához viszonyítva kevés anyaggal rendelkező adatbázisok nem szolgálják a felhasználó érdekeit. Arra kell törekedni, hogy a felhasználónak *szakterületek szerint kumulált adatbázisok álljanak rendelkezésére*. A kisebb, szűkösen ellátott és ezért csak ritkán használt adatbázisoknak a nagyobb adatbázisokba kell beépülniük. Megközelítőleg hasonló szolgáltatások, mint pl. az SDIM és a METadex az információszolgáltatás gazdaságossága érdekében nagyobb egységgé egyesülhetnének, főleg mivel a SDIM háromnyelvű indexszel dolgozik, így angolul is kereshető.

Európa és az NSZK egy sor értékes dokumentációs szolgáltatással rendelkezik, ezek többsége azonban önálló létezésre túlságosan kicsi, de nagyobb szolgáltatásokat jól kiegészíthetnének.

/BECHTEL, H.: Eine Datenbasis ist nicht genug – aber mensch eine ist zuviel! = Nachrichten für Dokumentation, 32. köt. 2. sz. 1981. p. 78–84.)

(Dezső Zsigmondné)

Egy vállalati könyvtár online információkereső szolgáltatásának értékelése a felhasználó szemével

A General Electric vállalat Kutatási és Fejlesztési Részlege Könyvtárának (Corporate Research and Development Whitney Library) kezelésében a következő online információkereső szolgáltatások állnak a vállalat tudományos, műszaki és vezető munkatársainak rendelkezésére:

a Bibliographic Retrieval Services (BRS),
az Information Bank,
a Lockheed,
a National Library of Medicine (NLM),
a NIH/EPA Chemical Information System (CIS),
a Systems Development Corporation (SDC) és az
US Department of Energy (DOE) RECON adatbázisai.

A vállalati K+F részleg könyvtára az 1978-ban és 1979-ben igénybe vett online szolgáltatások felhasználói körében *kérdőíves felmérést* hajtott végre. A megkérdezettek főként a Részlegtől távoli munkahelyeken dolgozó felhasználók és az első online felhasználók közül kerültek ki, de voltak köztük „állandó” ügyfelek is.

A vizsgálat célja az alábbiak megállapítása volt:

- az online keresőszolgáltatás milyen hatással van a felhasználók fejlesztő munkájára;
- a keresésben a felhasználó részvétele milyen hatással van a keresés eredményének minőségére;
- a visszakeresett dokumentumok hány százalékát olvassák el ill. rendelik meg;
- milyen módszerek alkalmasak a szolgáltatás népszerűsítésére és terjesztésére;
- milyen problémák merülnek fel a felhasználóknál, ill. mennyire kedvező számukra a szolgáltatás.

A kérdőívet 233 vállalati felhasználónak küldték meg, közülük 78-an adtak értékelhető választ (33%). Mivel azok közül is sokan, akik nem küldték vissza a kitöltött kérdőívet, további szolgáltatást igényeltek, a felhasználók elégedettségét az *ismételt szolgáltatási igényléssel* is értékelték: 1978-ban a keresések igénylőinek 55%-a két vagy több alkalommal kért keresést.

Ki honnan ismeri a szolgáltatást?

A kérdőív egyik pontja arra keresett felvilágosítást, hogy a válaszoló hogyan értesült az online szolgáltatásokról. 46%-uk azt válaszolta, hogy kollégáitól; további 45% a könyvtári értesítőkből, házi prospektusokból és a könyvtár munkatársaitól szerzett tudomást az ilyen információkeresés és szolgáltatás lehetőségéről.

A visszakeresés értékelése

A relevancia értékelését az 1. táblázat mutatja meg, a keresésben részt vevő ill. a benne részt nem vevő felhasználók szerinti bontásban. Az előbbi esetben a találatok túlnyomó részét vagy többségét relevánsnak tartotta a válaszolók 62%-a, míg az utóbbi esetben a válaszolóknak csak 56%-a nyilatkozott így. Ebből arra lehet következtetni, hogy a helyszíni, interaktív visszacsatolás a felhasználók részéről valamelyest javítja a keresés minőségét.

A kereséssel a felhasználó tudomására hozott dokumentumok megrendelésére vagy elolvasására vonatkozó felmérési adatokat mutatja be a 2. táblázat. Nem látható kölcsönös összefüggés a felhasználónak a keresésben való részvétele és az elolvasott dokumentumok aránya között. Látható viszont a táblázatból, hogy viszonylag nagy azoknak az aránya, akik a keresés eredményei alapján elolvasták a könyvtárban fellelhető dokumentumokat, ill. megrendeltek olyanokat, amelyek az állományból hiányoznak.

1. táblázat

A felhasználók relevancia-értékelése

A releváns találatok aránya	A terminálnál jelen nem lévő felhasználók szerint (%)	A terminálnál jelen lévő felhasználók szerint (%)	Az összes felhasználó értékelése szerint (%)
Legtöbb	13	10	12
Sok	42	52	45
Kevés	31	31	30
Egy sem	2	0	1
Nem tudja megítélni	9	3	7

2. táblázat

Az online keresésben részt vevő ill. a benne részt nem vevő felhasználók által megtekintett, elolvasott vagy megrendelt primer dokumentumok aránya

A visszakeresett találatok közül megkeresett eredeti dokumentumok százalékaránya

	0–25% NJ* J**		25–50% NJ J		50–75% NJ J		75–100% NJ J		Az összes felhasználó értékelése szerint
A könyvtárban megtekintett vagy elolvasott dokumentumok	47	41	9	14	7	7	2	0	62
Későbbi tanulmányozásra elvitt példány vagy másolat	29	17	2	7	4	0	2	0	32
Megrendelve a könyvtáron keresztül	56	41	11	10	2	0	0	3	62
Csak a dokumentum kivonatának megtekintése	22	17	16	14	27	14	20	21	75
Egyéb	7	0	–	–	–	–	–	–	4

* A terminálnál a kereséskor jelen nem lévő felhasználó szerint.

** A terminálnál a kereséskor jelenlévő felhasználó szerint.

A keresésből eredő előnyök

Arra a kérdésre, hogy mennyi időt takarítanak meg a felhasználók a manuális információkereséshez képest, összesen 84% válaszolta azt, hogy napokat venne igénybe az online kereséssel egyenértékű információkutatás, illetve hogy nem is tudnák manuálisan elvégezni. Ez jól megegyezik a szakemberek értékelésével, akik 20–22 órára becsülik a szakszerű irodalomkutatás átlagos idejét hagyományos módszerek esetén.

A megkérdezettek 84%-a szerint a keresés *megérte a ráfordított költségeket*. Következzen néhány kiragadott felhasználói vélemény arra vonatkozólag, hogy a vállalatnak és munkatársainak milyen előnyöket jelent az online információkeresés:

„... a *Chemical Abstracts*-ből származó információk háromról egy hónapra csökkentették laboratóriumi munkánkat”,

„... sok időt takarított meg számunkra”,

„... több információt kaptam néhány perc alatt, mint saját kereséssel több hónap alatt”,

„... évekkal visz előbbre versenytársainkhoz képest” stb.

A felmérésből látható, hogy az online keresőszolgáltatásnak a vállalati kutatásra–fejlesztésre gyakorolt hatása jól értékelhető ezzel a módszerrel, kvalitatív és kvantitatív értelemben is. Az online keresés előnyeinek egy további mérési módszere az lehetne, ha valamilyen vizsgálat a keresési kérések és a fejlesztés termelékenységének néhány mutatója (új publikációk, szabadalmak, megrendelések, díjak) közötti összefüggések kimutatására vállalkoznánk.

/WARDEN, C. L.: User evaluation of a corporate library online search service = Special Libraries, 72. köt. 2. sz. 1981. p. 113–117./

(Roboz Péter)

Az online információszolgáltatások megkezdése Portugáliában

A portugál Tudományos és Műszaki Dokumentációs Központot (*Centro de Documentação Científica e Técnica, CDCT*) 1976-ban szervezték át, azzal a céllal, hogy az addiginál hatékonyabb információszolgáltatást biztosítson a kutatás és fejlesztés számára. Az átszervezés közben szoros együttműködést alakítottak ki a svéd Királyi Műszaki Könyvtárral (*Royal Institute of Technology Library, RITL*), aminek eredményeként a CDCT műszaki és pénzügyi támogatást kapott a stockholmi intézettől.

A CDCT egyik kiemelt munkaterülete lett a számítógépes információszolgáltatások bevezetése Portugáliában. Két évi munka után jelentős tapasztalatokra tett szert, különösen az online információkeresésben.

Online információkeresési tevékenység

1977-ben a CDCT-ben a munka két szeminárium megtartásával indult. Azóta folyamatosan sokasodnak az online szolgáltatások igénybevevői, főként a tudományos kutatók, orvosok, tanárok, iparvállalati mérnökök stb. köréből.

A CDCT online szolgáltatásokkal foglalkozó részlegének hat munkatársa van, közülük négy felsőfokú végzettséggel rendelkezik. A svéd segítségnek köszönhető, hogy az első évben mindössze névleges díj ellenében nyújtottak online szolgáltatásokat a felhasználóknak. A következő szakaszban azonban új díjrendszert vezettek be: egy keresés ára kb. 15 angol fontnak megfelelő portugál escudo, amely maximum 30 perces online kapcsolati időt és maximum 200 kinyomtatott találatot foglal magában. 30 percnél hosszabb keresés esetén negyedóránként további mintegy 5 fontot, 200-nál több találat esetén pedig darabonként további 0,05 fontot kell fizetni. A fizetendő díj fedezi a tényleges költségeknek az egyharmadát, a többi kiadást a CDCT és a svéd–portugál együttműködési program együttesen „állja”.

A CDCT négy nagy külföldi online szolgáltató központhoz képes hozzáférést biztosítani. Az olaszországi központú ESA-IRS hálózathoz annak valamelyik csomópontjának tárcsázásával lehet kapcsolatot teremteni (1979 óta erre a madridi csomópontot használják). A brit BLAISE központhoz a birminghami csomópont tárcsázásával férnek hozzá. Az amerikai Lockheed és SDC rendszerek adatbázisait a TYMNET hálózat madridi csomópontján keresztül érik el.

Az 1. táblázat mutatja be az egyes adatbázis-központokkal létesített kapcsolati órák számát. Látható, hogy a 2. évben kb. 3-szor annyi időt fordítottak online keresésre, mint az elsőben. A második munkaévben összesen 53 adatbázisban végeztek online keresést. A legtöbbet használt adatbázisok – a csökkenő használat sorrendjében – a következők: Chemical Abstracts, Biosis Previews, MEDLINE, CAB, INSPEC, Excerpta Medica, ERIC, Psychological Abstracts, PASCAL, COMPENDEX, SCISEARC.

Az adatbázisok sokféleségét az magyarázza, hogy a CDCT az egyetlen online központ az országban, ezért a keresések témái igen széles spektrumon oszlanak el. Érdekes, hogy a gazdasági, üzleti jellegű adatbázisok egyáltalán nem szerepelnek az online szolgáltatások iránti igények között.