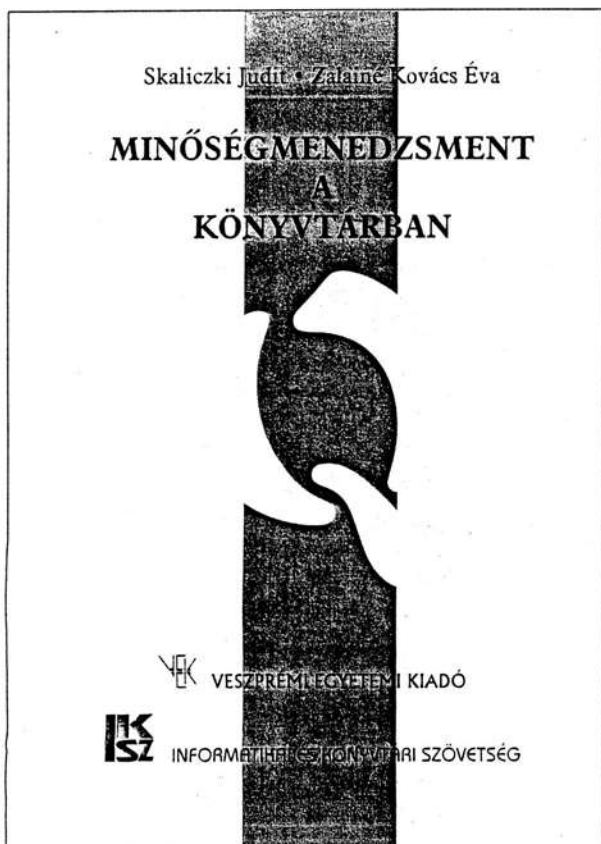


Csubák Antoaneta

BME OMIKK

Minőségmenedzsment a könyvtárakban

A minőség fontosságáról és növekvő gazdasági szerepéről sok tanulmány látott napvilágot hazánkban, különösen az elmúlt 10 esztendőben, amióta a piaci versenyképesség növelése szükségessé vált. A minőség-ellenőrzés, minőségbiztosítás és minőségirányítás fogalmak tisztázására, értelmezésére és alkalmazására az iparban széles körű szakirodalom áll rendelkezésre. Más a helyzet a szolgáltatások területén. Ez egy heterogén ágazat, ezért bizonyos szolgáltatásoknál – kereskedelem, bank, informatika stb. – a minőségbiztosítás már jól ismert és bevezetett rendszerként működik. Vannak azonban olyan területek – például a könyvtárak – ahol a minőségirányítással való megismerkedés és alkalmazása a gyakorlatban úttörő munkának számít. Skaliczki Judit és Zalainé Kovács Éva „Minőségmenedzsment a könyvtárban” című, 2001-ben megjelent könyve a könyvtárszakmában a könyvtári minőségirányítás ábécéjeként használható.



Minőségmenedzsment a könyvtárban / Skaliczki Judit, Zalainé Kovács Éva ; az „Esettanulmány” című (9.) fejezetet Ramháb Mária írta. - Veszprém : Veszprémi Egyetemi K., Budapest : Informatikai és Könyvtári Szövetség, 2001. - 193 p. ; 24 cm
ISBN 963 9220 70 1

Tart.j. és előszó angol nyelven is. Bibliogr.: 189-193.

A minőségügy tág értelemben a társadalom életminőségének ügye, a társadalom minőségkultúrájának szintje. Szűkebb értelemben a minőségügy a termékek, a szolgáltatások minőségével foglalkozik, vagyis a nemzetgazdaság versenyképességét és a nemzeti kultúrát meghatározó stratégiai kérdés. A fejlett országokban a minőségügyet már évtizedek óta kulcsfontosságú kérdésnek tartják. A minőség ellenőrzésére, biztosítására és irányítására különböző modelleket alakítottak ki és alkalmaznak (pl. Japánban – TQC = Total Quality Control, az USA-ban – TQM = Total Quality Management, Európában – a TQM eszközökkel támogatott ISO 9000-es szabványon alapuló minőségirányítási rendszer).

Az elmúlt években Magyarországon is körvonalazódott a minőségügy szabályozási rendszere. Létrejötték a minőségügy alapvető jogszabályi, intézményi és ösztönzési feltételei (a szabványosítási és akkreditálási törvény, a vizsgáló és tanúsító intézmények hatósági és szolgáltató tevékenységének, feladatainak szétválasztása, a Nemzeti Minőségi Díj alapítása, a minőségbiztosítási rendszerek széles körű bevezetésének állami támogatása, EU-jogszabályok átvétele stb.). Megfogalmazódott a Nemzeti Minőségösztönző Politika és a Nemzeti Minőségfejlesztési Program. Az utóbbiban részt vesz a kormány, a minisztériumok, az oktatási intézmények és a közsféra minden területe. Összefoglalva, Magyarországon a minőségirányítás kiemelt szerepet kap a közélet minden szektorában, beleértve a könyvtárakat is.

A *Minőségmenedzsment a könyvtárban* című könyv szerzői merész feladatra vállalkoztak: 190 oldal terjedelemben ismertetni a minőségügy alapfogalmait, valamint a könyvtári minőségirányítási rendszer lényegét és bevezetésének lépéseit, felvázolni a hazai könyvtárügy stratégiai tervének elkészítési módját, bemutatni a TQM módszereket és a minőségügyi dokumentációs rendszer felépítését, sőt az EFQM Kiválóság Modell sem maradt ki.

A szerzők arra törekedtek, hogy a „száraz definíciók” egyszerű példákkal „fűszerezve”, a könyvtár nyelvére lefordítva közérthetővé váljanak. A *Katona József Könyvtárban* megvalósított projektről szóló esettanulmány is hozzájárult ahhoz, hogy a könyv ne csak elméleti ismereteket nyújtson, hanem gyakorlati oldalról is alátámassza a minőségirányítás működőképességét a könyvtárakban.

A hallgatók számára ez tipikus tankönyv: tömören, lényegretörően, felsorolásszerűen és szinte teljeskörűen ismerteti a minőségirányítás fogalmait, módszereit, modelljeit, a hazai és a külföldi trendeket (kivéve az ISO 9000-es minőségirányítási rendszert).

A gyakorló könyvtárosok számára újszerű gondolkodásmódot és munkamódszert mutat be, amelynek elsajátításával és alkalmazásával a nemzetközi követelményeknek is megfelelő modern könyvtárat lehet kialakítani és működtetni.

Végül, a laikus (nem könyvtáros) számára is hasznos olvasmány, logikus felépítése és az értékes, nagy munka árán összegyűjtött információ-tartalom miatt.

A mű legnagyobb érdeme – az, hogy olvasmányosan, világos gondolatmenettel, jól áttekinthetően, a témába csak most bekapcsolódni szándékozók számára is közérthetően, gyakorlati példákkal gazdagon illusztrálva mutatja be a könyvtári minőségirányítás lényegét – a szerzők (könyvtári és minőségügyi) szakmai felkészültségének, tapasztaltságának, profizmusának köszönhető.

A „Minőségmenedzsment a könyvtárban” c. könyv alapján összeállítást készítettünk a minőségirányítás célkitűzéseiről, a rendszer felépítésének lépéseiről, a megvalósítás lehetséges módszereiről és az elért eredmények értékelésének metodikájáról.

Milyen a jó könyvtár a használó szemével nézve?

Klasszikus értelemben, a 19. századbeli könyvtárkép alapján az olvasó a mai napig is elvárja, hogy a könyvtárban

- esztétikus, rendezett, tiszta környezet fogadja;
- a könyvtárosok legyenek udvariasak, készségesek, a keresés legyen eredményes;
- kényelmes helyet kapjon az irodalomkutatáshoz;
- a szakirodalom legfrissebb kiadásai, és széles választéka álljon rendelkezésére.

Az évtizedek során ezek az igények tovább bővültek. A könyvtáros – aki a minőségirányítás középpontjában áll – feladata az ügyfelek, használók folyamatosan növekedő, változó igényeit figyelni, és ennek megfelelően intézkedni. Ide kapcsolódik a szerzők által is ismertetett minőség kérdése, miszerint a minőség piackutatással, felmérésekkel kezdődik. Nemcsak a meghatározott igényeket (pl. helyben használat, kölcsönzés, tájékoztatás) szükséges megismerni, hanem célszerű a látens igények (amelyeknek az ügyfél nincs tudatában) felméréséhez meghatározott kutatásokat a fejlesztés, tervezés fázisában is végezni. A használók igényeinek felméréséhez jól bevált módszer a megfelelően összeállított kérdőíven történő kérdezés, és ami még lényegesebb – a kapott válaszok elemzése, értékelése és közzététele. Nem maradhatnak el a javító intézkedések, sőt a jól megtervezett megelőző intézkedésekkel megakadályozható, hogy ugyanaz a probléma újra előfordulhasson. Jó megoldás a szerzők által javasolt „minden egyes szolgáltatási pontnál elhelyezni az olvasói javaslatok könyvét, és minden névvel és címmel ellátott javaslatra, válaszolni”. Más szóval, törődni kell az olvasóval.

Az utóbbi években jelentősen megváltoztak az olvasói szokások: a könyvtárhasználók létszáma megnövekedett, ugyanakkor más jellegű könyvtári szolgáltatásokra (elektronikus) is igény van. Ennek megvalósítására pedig a kommunikációs és információs technológiákkal a technika fejlődése tág teret nyitott arra, hogy „az éppen adott könyvtár szolgáltatásain kívül a használók egy virtuális könyvtárban barangolhassanak a világ legjelentősebb könyvtári állományai és információi között”.

A külső körülmények (globalizáció, információs társadalom kialakulása, jogi szabályozások, szabványok, irányelvek egységesedése stb.) változása, a technika fejlődése adta lehetőségek (virtuális könyvtár) és a használók megváltozott igényei a

könyvtárakat új szerepkörhöz juttatták, a könyvtár stratégiai intézménnyé változott, amely a kormányprogramokban is helyet kapott.

Visszatérve a kiinduló kérdéshez – *Milyen a jó könyvtár?* –, a válasz: *a mai könyvtárhasználó a korszerű virtuális könyvtárat tartja megfelelőnek, ahol egy helyen gyorsan és pontosan hozzájuthat a keresett információhoz, és mindez kellemes környezetben és udvarias könyvtárosi segítséggel történik.*

Ennek az alapgondolatnak kell alávetni az egész könyvtári tevékenységet, a jövőkép kialakítását, a küldetésnyilatkozat megfogalmazását, a stratégiai terv készítését, a minőségirányítási rendszer bevezetését.

A könyvtári minőségirányítási rendszer felépítésének lépései

1. A könyvtár minden dolgozója (a vezetőséggel az élen) ismerkedjen meg a minőségirányítás alapfogalmaival, és azonosuljon a minőség-szemlélettel. (A könyvtáros higgyen benne, hogy a minőségirányítási rendszerben való működés a könyvtár színvonalát emeli.) Ezt megfelelő oktatással lehet elérni.
2. Minőségirányítási munkacsoport létrehozása. Ajánlott létszám: 5–10 fő.
3. Meg kell fogalmazni és közzé kell tenni a könyvtár jövőképét, küldetésnyilatkozatát és minőségpolitikáját. Ezek tartalmára vonatkozóan segítséget ad a „Minőségmenedzsment a könyvtárban” c. könyv.
4. Kiinduló helyzetfelmérést kell végezni, amely kiterjed a minőségirányítási rendszer minden elemére: vezetés (szemléleti kérdés), minőségügyi stratégiai terv, humán és tárgyi erőforrások biztosítása, a szolgáltatási folyamatok áttekintése – ellenőrzési metodikák, teljesítménymérés és -értékelés, teljesítménymutatók alkalmazása, használói elégedettség felmérése, értékelése, javító intézkedések, a dolgozók elégedettségének mérése, statisztikák készítése, értékelése, szervezeti kultúra, a könyvtár társadalmi hatásának vizsgálata.
5. A feltárt „gyenge pontok” elemzése. Intézkedési terv készítése az ún. SWOT-analízis alkalmazásával.
6. Minőségügyi stratégiai terv összeállítása.
7. A szolgáltatások fejlesztésével kapcsolatos feladatok megfogalmazása a „PDCA ciklusmodell” alapján. Célszerű minden egyes szol-

gáltatást külön elemezni: milyen munkafázisokból áll, hogyan kapcsolódnak egymáshoz a különböző munkafázisok, hol vannak az ellenőrzési pontok, hogyan regisztráljuk a nem-megfelelőségeket (hibákat), milyen módon történik a hibák kiküszöbölése. Megvizsgálni, mi okozta a hibát. A hibakövek feltárására, kiküszöbölésére és megelőzésére alkalmazhatjuk a TQM módszerek közül bármelyiket, pl. a Pareto-elemzést, a halszálkadiagramot vagy a hisztogramot.

8. A minőségfejlesztési feladatok végrehajtása.
9. Teljesítménymérés és értékelés:
 - statisztikák készítése – mennyiségi megközelítés (pl. kötetek száma, beiratkozott olvasók száma);
 - értékelés a különböző teljesítménymutatók alkalmazásával – minőségi megközelítés az MSZ ISO 11 620:1998 sz. szabvány alapján.
10. A dolgozók elégedettségének mérésére megfelelő kérdőívek összeállítása, kitöltetése, javaslatok elemzése, intézkedések.
11. A használók elégedettségének mérésére alkalmas módszerek kiválasztása és alkalmazása (kérdőív, interjú, panaszláda stb.).
12. A minőségügyi rendszer dokumentációjának aktualizálása (pl. Olvasói tájékoztató, űrlapok, Ügyrend, Belső szabályzatok).
13. A minőségirányítási rendszer hatékonyságának értékelésére szolgáló ún. „Vezetőségi felülvizsgálat”, évente legalább egyszer.

A fenti 13 lépésből álló folyamat nem egy köbe vésett eljárás, csak egy lehetőség. Szabad a feladatok sorrendjét felcserélni, az egyes lépéseket leegyszerűsíteni vagy kibővíteni, ez mindig a konkrét környezettől függ.

A Katona József Könyvtár minőségbiztosítási projektet bemutató Esettanulmányából például látható, hogy a megvalósítást külső angol és magyar szakemberek által tartott tréningekhez kötötték, ami azzal magyarázható, hogy az egész munka egy pályázat keretén belül zajlott. Ez a megvalósítási forma mintaszerűnek bizonyult, és módszere modellként szolgálhat más könyvtáraknak.

A TQM szemlélet megvalósítására szolgáló modellt rugalmasan lehet építeni. Többféle út vezet az új könyvtárkép kialakításához. A „Minőségmenedzsment a könyvtárban” c. könyvben például kevés szó esett az ISO 9000-es szabvány szerint felépített minőségirányítási rendszerekről, amelyek szintén a minőségbiztosítás, minőségfejlesztés témakörébe tartoznak.

Az ISO 9000-es szabvány követelményei szerint felállított rendszer jellemzői

2000 decemberében megjelent az ISO 9000-es szabványsorozat új verziója az alábbiak szerint:

A szabvány száma	A szabvány megnevezése
ISO 9000: 2000	Minőségirányítási rendszerek Alapok és szakszótár
ISO 9001: 2000	Minőségirányítási rendszerek Követelmények
ISO 9004: 2000	Minőségirányítási rendszerek Irányelvek a teljesítőképesség továbbfejlesztéséhez

A nyolc alapelv

- Folyamatszempléltű megközelítés:** a kívánt eredményt hatékonyabban lehet elérni, ha a tevékenységeket és a velük kapcsolatos erőforrásokat folyamatként kezelik.
- Rendszerszemlélet az irányításban:** az egymással összefüggő folyamatok rendszerként való azonosítása, megértése és irányítása hozzájárul ahhoz, hogy a szervezet eredményesen és hatékonyan valósítsa meg céljait.
- Tényeken alapuló döntéshozatal:** az eredményes döntések az adatok és egyéb információ elemzésén alapulnak.
- Vezetés:** a vezetők megteremtik a célokra és a szervezet vezetésének egységét. A vezetők felelőssége létrehozni és fenntartani olyan belső környezetet, amelyben a munkatársak teljes mértékig részt vehetnek a szervezet céljainak elérésében.
- Vevő(használó)-központúság:** a szervezetek vevőiktől függenek, ezért fontos, hogy megértsék a jelenlegi és a jövőbeli vevői szükségleteket, teljesítsék a vevők követelményeit, sőt igyekezzenek felülmúlni a vevők elvárásait.
- Folyamatos fejlesztés:** a szervezet teljes működésének folyamatos fejlesztése legyen a szervezet állandó célja.
- A munkatársak bevonása:** a szervezet lényegét minden szinten a munkatársak jelentik, és teljes bevonásuk teszi lehetővé képességeik kihasználását a szervezet javára.

- Kölcsönösen előnyös kapcsolatok a beszállítókkal:** a szervezet és beszállítói kölcsönösen függenek egymástól, és kölcsönösen előnyös kapcsolatuk fokozza mindkettőjük értékteremtő képességét.

Az új szabvány a vevői igényeket helyezi a középpontba, és a szervezet felső vezetésének feladat-körébe sorolja a gondoskodást arról, hogy a vevői igényeket követelményekké alakítsa át, és teljesítse ezeket a követelményeket a vevő elégedettségnek elérése érdekében. Ezáltal a szabvány az eddigi minőségbiztosítási rendszer helyett a minőségirányítási rendszer követelményeit, a korszerű minőségügyi filozófiát (TQM) foglalja magában.

Nincs merev határ a TQM és az új verziójú ISO minőségirányítási rendszermodellek között. Az alapelvek lényegében azonosak, a minőségfejlesztési eszközök, módszerek is azonosak. A különbség csak a rendszer szerkezetében és a dokumentálási formákban van.

Bármelyik rendszert vezetik be egy könyvtárban, azonos a cél – minden intézkedés a könyvtárhasználó elégedettségének növelése érdekében történik.

Kihívás a könyvtárak számára

A könyvtári menedzsment területén is az európai normákhoz kell igazodni. Az új minőségszemléltű kultúrát be kell vezetni a könyvtárakban is, ez a jövő útja. A felzárkózás elkerülhetetlen, minden könyvtár érdeke időben hozzáfogni a munkához, gondoljunk a Katona József Könyvtárra, ahol a felkészülési folyamat három évet vett igénybe. De megérte! Csak a kezdet nehéz: meghozni a döntést a minőségirányítási rendszer építésére. A folytatáshoz *szemléletváltásra, új gondolkodásmódra és a minőségügy iránti elkötelezettségre* van szükség.

Beérkezett: 2002. II. 21-én.

Tóth Erzsébet

Nyíregyházi Főiskola Matematika és Informatika Intézet

Az automatikus osztályozással kapcsolatos kísérleti kutatások eredményei

A digitális dokumentumok rugalmas és rendszerezett formában történő elérése a világhálón alapvető elvárás a felhasználók részéről. A kérdés megoldásában nagy feladat hárul a könyvtárosokra is, mert ott van lehetőség a korábbi felhalmozott tudásanyag megosztására, és szorosabb együttműködés kialakítására a fejlesztőkkel. Ez az írás áttekinti azokat a nemzetközi kutatási programokat, amelyek a HTML dokumentumok automatikus indexelését és osztályozását oldják meg.

A kísérleti kutatások jelenlegi helyzete

Az automatikus osztályozás lehetőségeit vizsgáló kísérleti kezdeményezések kiemelt kutatási tevékenységnek tekinthetők a könyvtári osztályozás területén. Korábban az automatikus osztályozás klaszterálási és statisztikai módszerekre épült, azonban ezeknek a módszereknek a konkrét megvalósításait akadályozta a számítógépesítés hiánya és a korlátozott tárkapacitás. Napjainkban viszont egyre több olyan kutatási programot indítanak, ahol a korábbi indexelési technikákat alkalmazzák, és a különböző klaszterálási módszerek hatékonyságát vizsgálják.

Jelenleg nincs tudomásunk olyan gyakorlati megvalósításról, ahol a hagyományos osztályozási rendszerek szerepét teljes mértékben átvennék az automatikus módszerek, habár fokozott érdeklődés mutatkozik ilyen jellegű rendszerek fejlesztése iránt. A kísérleti kutatásokat többnyire multinacionális cégek támogatják. Ez egyrészt azzal az igénnyel magyarázható, hogy szükség van olyan gyakorlati megoldásra, amely biztosítja a digitális dokumentumok rugalmas elérhetőségét rendszerezett formában a hálózaton. Másrészt ez a törekvés annak is tulajdonítható, hogy az internetes keresők egyszerű indexelési technikái nem nyújtanak kielégítő találatokat a felhasználóknak.

A tematikus weboldalakra alkalmazott osztályozási rendszerek egy általános böngésző struktúrát nyújtanak a HTML dokumentumok számára [1]. Az önállóan kialakított osztályozási rendszerek legnagyobb hátránya az, hogy nem zárják ki a következtetéseket struktúrájukból, és ezáltal könnyen ellentmondásokhoz vezethetnek. A kísérleti kuta-

tásokban ezért a hagyományos könyvtári osztályozási rendszereket próbálták ki különböző módszerekkel.

Az ETO-t általában hasznos rendszerező eszköznek tekintik a weblapoknál, hiszen géppel értelmezhető formátuma nagymértékben elterjedt a fejlesztők körében. Az ETO-t használó internetes szolgáltatásoknál megfigyelhető a dokumentum témájának kötöttebb meghatározása, ami ellentétben a Dewey Tizedes Osztályozást és a Kongresszusi Könyvtár osztályozását alkalmazó rendszerek gyakorlatával. Az utóbbi két osztályozásnál sokkal egyszerűbb jelölési rendszert alkalmaznak, és a tartalmi részletezés is hiányzik [1].

A skandináv WAIS/World Wide Web program

1993 nyarán kezdődött el a lundini egyetemi könyvtár és a dán műszaki könyvtár együttműködésével, amely egy évig tartott. 660 WAIS adatbázis automatikus osztályozását valósította meg, kizárólag az ETO 51-es osztályára épült. Az ETO géppel olvasható formátumát teljes mértékben kiaknázta, ahol az osztályozási kategóriák és a jelzetek között automatikus megfeleltetés van [1, 2]. Technikájára jellemző, hogy minden egyes WAIS adatbázis számára kulcsszavas listákat állítottak elő, ahol a kifejezéseket az adatbázis-leírások kulcsszavas, tárgyszavas és szabadszöveges mezőjéből vették. Ezután az adatbázis-leírásokból nyert kulcsszavas kifejezéseket megfeleltették az ETO osztályozási kategóriáinak. A két kifejezés pontos megegyezésénél különböző osztályozási jelzeteket kaptak eredményül. A jelzeteket egy algoritmus segítsé-

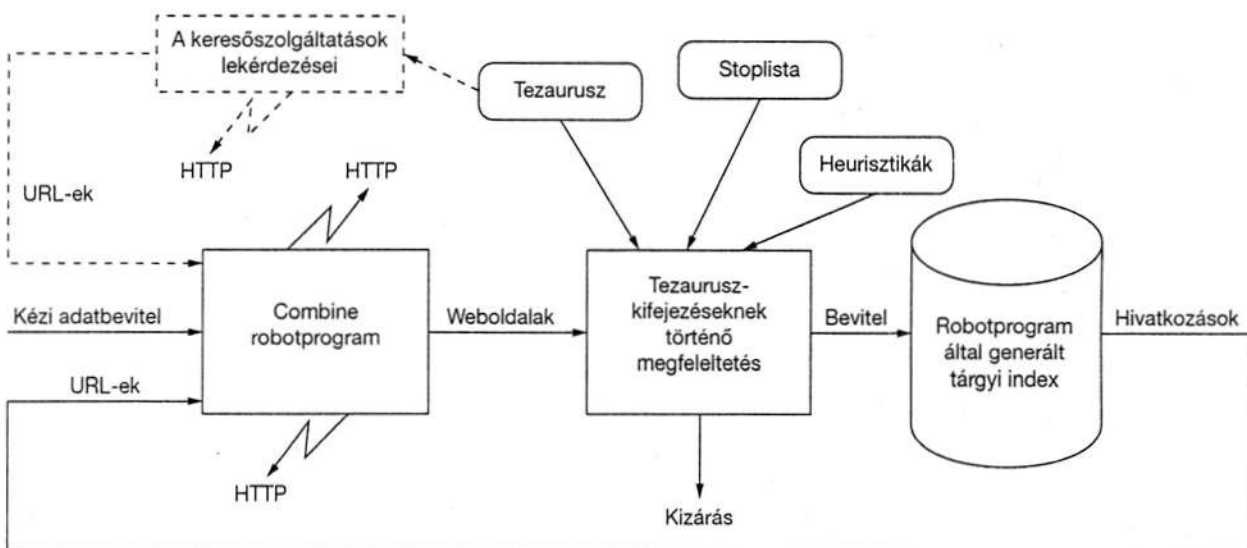
gével súlyozták a hozzájuk rendelt kulcsszavas kifejezések származási helye szerint. Az osztályozási jelzetek súlyozása révén nyert eredményeket végül összehasonlították, és a legmegfelelőbb jelzeteket kiválasztották [3, 2].

A DESIRE program második része

Az Európai Unió *DESIRE* (*Development of a European Service for Information on Research and Education*) nevű programjának általános célkitűzése az európai kutatást támogató információs szolgáltatások fejlesztése volt. A program második része 1998 júliusában kezdődött el tíz partnerintézmény közreműködésével, amely két évig tartott [4]. Vizsgálták annak lehetőségét, hogy hogyan lehet összekapcsolni egy szakértők által kiválogatott szakterületi hivatkozásgyűjteményt egy robotprogram által generált tárgyi indexszel. A vizsgálatokat a lundí egyetem *EELS* (*Engineering Electronic Library System*) nevű szakterületi „gateway” szolgáltatására és az „All Engineering” nevű tárgyi indexre alapozták. A tárgyi index böngészőfelületét egy korábban használatos, szakterületre szabott *EI* (*Engineering Information Inc.*) osztályozási rendszer segítségével alakították ki.

ségével valósították meg. Gyűjtési stratégiájukban kétféle módszert alkalmaztak és teszteltek. Az első módszernél néhány manuálisan szerkesztett szakterületi hivatkozási gyűjteményből indultak ki, és azoknak a további hivatkozásait a földaltól számított harmadik szintig, míg a második módszernél a második szintig követték, ily módon párhuzamosan az összes idézett weboldalt begyűjtötték az indexadatbázisba. Minimális átfedést találtak az idézett források között, hiszen többségük csak egy szolgáltatásban fordult elő. Egy fejlettebb megoldásnál már tezaurust használtak a számításba vehető weblapok tartalmi szűrésére, ahol a weboldalak szövegét összehasonlították a tezauszban használatos fogalmakkal. A kapott eredményeket súlyozták, ami alapján eldönt, hogy egy weblap bekerült-e az indexadatbázisba avagy sem.

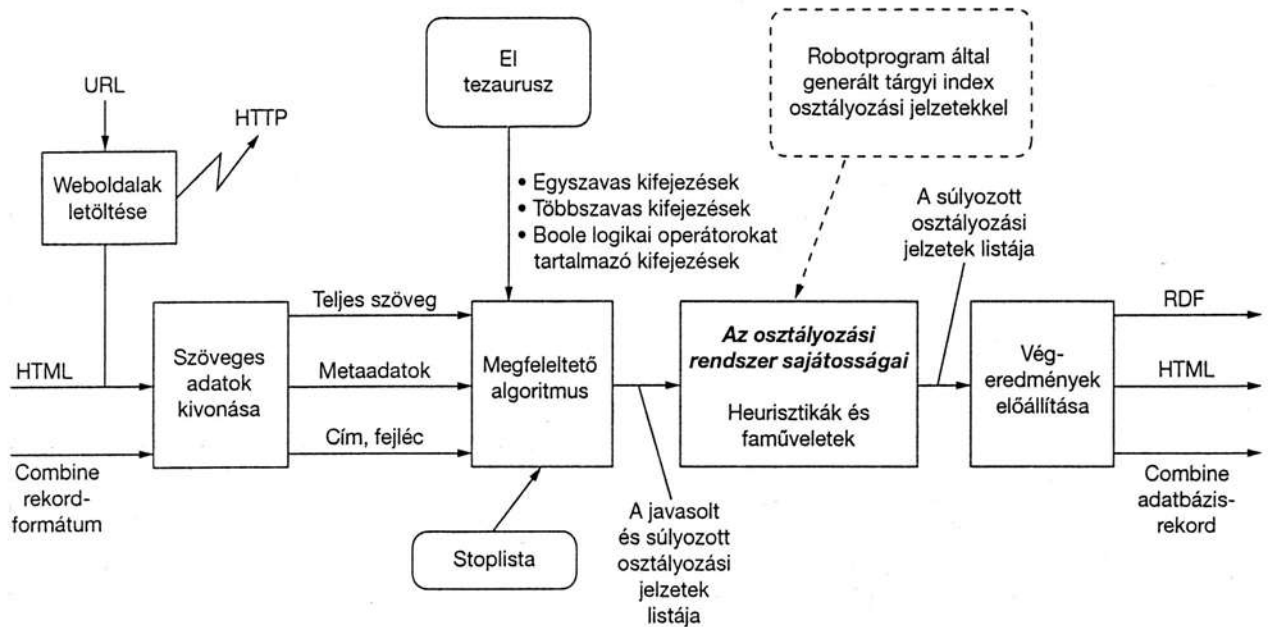
A weblapok automatikus osztályozását az EI tezausz segítségével végezték el, aminek nyomtatott változata tartalmazza a benne használatos fogalmak és az EI osztályozási jelzetek közötti összefüggéseket. Ennek a két információnak az összekapcsolása rendkívül fontos az automatikus osztályozás szempontjából. Az EI tezausz 17 458 fogalmat tartalmaz, ezek közül 8273 fogalom pre-



1. ábra A robotprogrammal működő tárgyi index begyűjtési stratégiája

Az 1. ábrán a Combine nevű robotprogram begyűjtési módszere látható, amely egy tezausz integrál működésébe a weblapok tartalmi szűrésére. Először is meg kellett oldaniuk azt a problémát, hogy minél relevánsabb dokumentumok kerüljenek be az indexadatbázisba, amelyet a Combine segít-

ferált kifejezésként jelenik meg. A belső tezausz formátumba való betöltés után az EI tezausz 3000 egyszavas fogalmat és körülbelül 18 000 összetett fogalmat tartalmazott. Az utóbbi kategóriába tartoznak a Boole logikai operátorokat tartalmazó kifejezések és a többszavas kifejezések.



2. ábra Az automatikus osztályozás folyamata

A 2. ábra részletezi a DESIRE programban alkalmazott automatikus osztályozás különböző lépéseit. Az első lépésben metaadatok kerülnek az indexadatbázisba, amelyeket a HTML dokumentum metaadatokot tartalmazó TAG-jéből, fejlécéből és törzséből vesznek. Ezután a létrehozott indexkifejezéseket megfeleltetik az EI tezaurusz fogalmainak, ahol a stopszavakat is figyelembe veszik. Ha a két kifejezés pontosan megegyezik egymással, akkor egy osztályozási jelzetekből álló lista generálódik a dokumentumhoz. Az eredményül kapott indexkifejezés és osztályozási jelzet hozzárendeléseit súlyozzák egy algoritmus segítségével.

Erre a súlyozásra azért van szükség, hogy megkülönböztessék az indexkifejezések relevanciáját egymástól, hiszen egy Boole logikai operátorokat tartalmazó kifejezés és egy többszavas kifejezés pontosabban behatárolja a dokumentum témáját, mint az egyszavas kifejezés. Sokszor az egyszavas kifejezések hamis találatokat eredményeznek a téma meghatározásánál, ilyenek például a homonimák. Az indexkifejezések súlyozásánál többféle szempontot vesznek figyelembe a végső súlyozási érték meghatározásához:

- az indexkifejezés típusát (egyszavas, többszavas vagy Boole logikai operátorokat tartalmazó kifejezésről van-e szó),
- a metaadatok származási helyét rangsorolják,
- az indexkifejezés előfordulási gyakoriságát is megvizsgálják az eredeti dokumentum szövegében.

Minden dokumentumhoz egy osztályozási jelzetekből álló listát rendelnek, ahol a jelzetek súlyozási értéke szerint csökkenő sorrendben jelennek meg. Általános osztályozási gyakorlatnak tekinthető, hogy a dokumentumhoz mindig a legspecifikusabb osztályozási jelzetet kapcsolják. A javasolt jelzetek számát csonkolással csökkentik, azaz minden jelzet végleges súlyozási értékét egy küszöbértékhez hasonlítják. Tehát a küszöbérték alatti súlyozási értékeket figyelmen kívül hagyják. A küszöbértéket egy heurisztika segítségével állapítják meg. Az eredményül kapott jelzetek kétféleképpen jelennek meg a böngésző szolgáltatásban HTML vagy RDF formátumban.

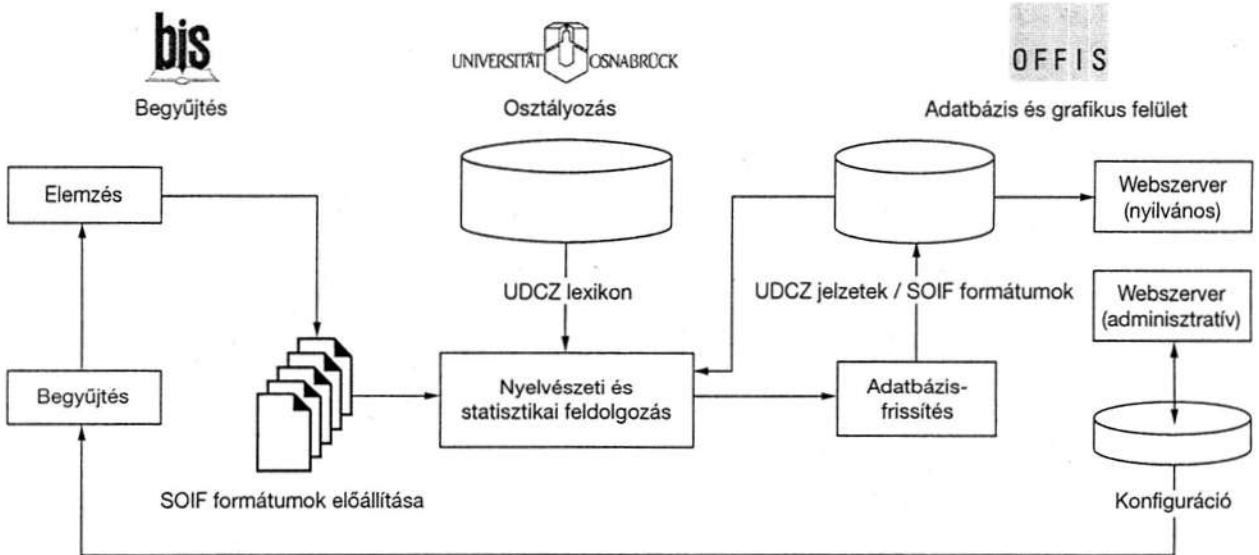
A kísérlet eredményei a következők voltak: 86 468 dokumentumot osztályoztak automatikusan, egy weblaphoz átlagosan 6 osztályozási jelzetet rendeltek. 923 weboldalt mintaként megvizsgáltak, és megállapították, hogy az automatikusan generált osztályozási jelzetek 57 és 66% között feleltek meg az intellektuális osztályozással nyert jelzeteknek [5].

A GERHARD program

Német nyelvű weboldalak automatikus indexelését és osztályozását oldja meg. Kifejlesztett szolgáltatása együttesen kínálja fel a keresés és a böngészés lehetőségét a felhasználók számára. Az ETO-nak egy háromnyelvű kiadását – az ETH-

Zürich szerinti változatát – használják, amit UDCZ-nek neveznek. Ez a rendszer 60 000 osztályozási kategóriából áll, amelyek között 15 különböző reláció lehetséges. Minden egyes osztályozási kategóriája tartalmaz egy UDCZ jelzetet, egy szöveges leírást és a lehetséges szinonimákat németül, angolul és franciául. Ha szükséges, az osztályozási kategóriánál egyértelműen hivatkoznak a korábban használt és a kapcsolódó jelzetekre.

szövegfeldolgozásra fordított minimális időnek a biztosítása. E célok elérése érdekében különböző nyelvészeti és statisztikai módszereket alkalmaztak. Alapvető elképzelésük az volt, hogy az UDCZ osztályozási rendszert konvertálják egy olyan lexikon formátumba, amely tartalmazza az osztályozási kategóriák és az UDCZ jelzetek közötti megfeleltetéseket. Ezután a dokumentum szabadszövegéből vett kifejezéseket megfeleltették a lexi-



3. ábra Rendszerarchitektúra

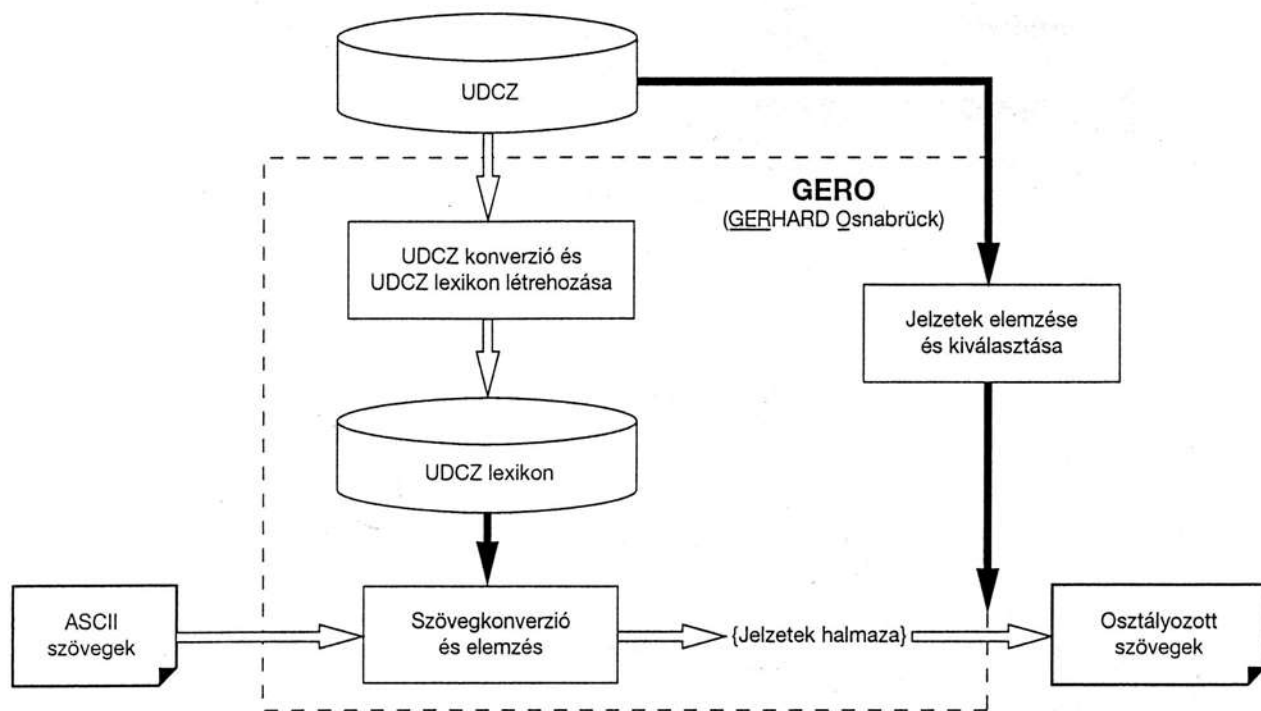
A 3. ábrán egy összetett rendszerarchitektúra látható, amely a következő elemeket tartalmazza: a weboldalak begyűjtését és szöveges elemzését, a SOIF (Summary Object Interchange Format) formátumok előállítását, a nyelvészeti és a statisztikai feldolgozást, amely az UDCZ lexikon segítségével történik. Ennek a folyamatnak a végén az UDCZ jelzetek és a SOIF formátumok egy Oracle relációs adatbázisba kerülnek. A GERHARD robotprogramja a tudományos szintű és releváns dokumentumokat keresi és gyűjti az indexadatbázisba különböző szűrési szabályok segítségével. A releváns weboldalak helyének a meghatározásához az összes konfigurációs adatot egy relációs adatbázisban tárolják. Ezek a konfigurációs adatok elérhetők és módosíthatók a weben keresztül. Begyűjtés után a HTML dokumentumok szövegét alaposan kielemezzik, és felkészítik a további feldolgozásra. A dokumentum lényeges tartalma egy strukturált SOIF formátumba kerül át egy másik szoftver segítségével.

Két szükséglet befolyásolta nagymértékben az automatikus osztályozás megvalósítását: az osztályozási jelzetek maximális pontosságának és a

konban használatos osztályozási kategóriáknak. Ha a két fogalom pontosan megegyezett egymással, akkor a dokumentumhoz osztályozási jelzetek generálódtak, amelyeket a pontosság érdekében súlyoztak.

A 4. ábra alapján megállapítható, hogy a nyelvészeti alapú osztályozási rendszer három fő alkotóelemből áll: az UDCZ osztályozási rendszer konverziójából és az UDCZ lexikon létrehozásából; a dokumentum szövegének a konverziójából és annak elemzéséből; az UDCZ jelzetek vizsgálatából és a megfelelő jelzetek kiválasztásából. Az ábrán megfigyelhető, hogy a szövegkonverzió és az elemzés bemenetét alkotják az ASCII szöveges dokumentumok, a megfelelő jelzetekkel ellátott dokumentumok pedig végeredményként jelennek meg.

Az osztályozási rendszer konverziójának elsődleges célja az, hogy az UDCZ ömlesztett szövegéből kiválogassák azokat a természetes nyelvű kifejezéseket, amelyek az osztályozási kategóriákat írják le. Az UDCZ konverziót három lépésben hajtották végre:



4. ábra Nyelvészeti alapú osztályozás a GERHARD-ban

1. Az osztályozási kategóriákban előforduló szavakat morfológiai szempontból a szótővéig elemezték, és megállapították, hogy milyen szófajba tartoznak. Erre a célra nyelvészeti szoftvereket használtak.
2. Különböző szabályokat alkalmaztak arra vonatkozóan, hogy hogyan lehetséges a természetes nyelvű kifejezéseket azonnal felismerni, és kiválogatni az elemzett szövegekből.
3. Törölték a rövidítéseket, stopszavakat és az annotációkat az osztályozási kategóriák leírásából.

Az UDCZ strukturálatlan szövegében minden tétel tartalmaz egy osztályozási jelzetet, és a hozzárendelt osztályozási kategória természetes nyelvű leírását három nyelven (lásd 1. táblázat). A 2. táblázat első három sorában az elemzett kifejezések szótővei és szófajai láthatók. Ugyanennek a táblázatnak az utolsó három sorában az eredményül kapott természetes nyelvű kifejezések fedezhetők fel.

1. táblázat
Példa az UDCZ-ben található strukturálatlan szöveges adatokra

```

001Z~03
002DDUEBERSETZUNGEN/TECHNISCHE U.
NATURWISSENSCHAFTLICHE
003DETRANSLATIONS/TECHNICAL AND SCIENTIFIC
004DFTRADUCTION/SCIENTIFIQUE ET TECHNIQUE
    
```

2. táblázat
Az UDCZ rendszer konverziójának eredményei

```

translation~N/technical~Adj and~Conj scientific~Adj
uebersetzung~N/technisch~Adj u.~Conj
naturwissenschaftlich~Adj
technical translation; scientific translation
technisch uebersetzung; naturwissenschaftlich
uebersetzung
    
```

A 3. táblázatban az osztályozási kategóriák és a hozzájuk rendelt UDCZ osztályozási jelzetek jelennek meg.

3. táblázat
Az eredményül kapott UDCZ lexikon egyik tétele

```

technical translation:— :~03
gene:xxx s:575.113.1
    
```

Az osztályozásra szánt dokumentum szövegét az UDCZ lexikon formális elvárásaihoz igazítják (stopszavak és umlaut törlése a szövegből). A lexikonból egy olyan felismerő eszközt hoznak létre, amely képes megfeleltetni az UDCZ lexikonban lévő osztályozási kategóriákat a dokumentum szövegében előforduló természetes nyelvű kifejezéseknek. Az alkalmazott módszer lényege a következő: a szoftver a szótővekhez különböző csonkolási variációkat told hozzá, hogy elősegítse a szótő illeszkedését a speciális szóalakokkal, például: *technische uebersetzungen, technischer ueber-*

setzungsvorschriften. Ezek a csonkolási változatok általában hasznosak, mert a szóalakokat rugalmasan feleltetik meg. Rövid szavak esetében azonban gyakran hamis találatokhoz vezetnek, például: a *gene* kifejezés illeszkedik a *general* és a *generic* szóalakokkal, ezért a szavak összes lehetséges morfológiai végződését kilistázzák a megkülönböztetés érdekében. A 3. táblázatban látunk erre egy példát, ahol a – jel jelzi a tetszőleges szóvégződéseket, és az xxx jelöli azt, hogy a szótó megegyezik a szóalakkal.

A két összehasonlítandó kifejezés pontos megegyezésénél egy osztályozási jelzetekből álló lista generálódik a dokumentumhoz. A releváns jelzetek kiválasztása két fő lépésben történik:

- Felhasználják a jelzetek által kifejezett információt, és vizsgálják előfordulási gyakoriságukat. A jelzetek relatív fontosságát abszolút gyakoriságuk alapján határozzák meg. A jelzet szöveges megfeleltetése alapján keresik meg azt a jelzet-klasztert, amelyhez tartozik.
- A kapott jelzeteket statisztikai módszerekkel súlyozzák, és számukat csökkentik. Megfelelő algoritmus alkalmazásával szűrik ki a releváns osztályozási jelzeteket. A nyelvészeti alapú osztályozás minden dokumentumhoz átlagosan 14 jelzetet rendel. Ezt a számot 6-8 jelzetre csökkentik.

Súlyozásnál figyelembe veszik a jelzetek közötti hierarchikus relációkat, azaz a jobb oldali csonkolás elvével mindig a fölérendelt osztály jelzetét kapják meg. Az automatikusan generált jelzeteket egy karakterekből álló fastruktúrába rendezik, ahol minden egyes pont a fában implicit információt közöl a jelzet prefixumáról. Így érvényesül az az alapelv is, hogy minél hosszabb egy jelzet, annál speciálisabb osztályozási kategóriát jelöl a fastruktúrában belül.

A rendszerben történő navigációt tulajdonképpen egy irányított gráf segíti elő, amely külön tartalmazza az osztályozási kategóriákat és a közöttük lévő relációkat. A gráf több ciklusból áll, ahol a csomópontokban (NODES) található az osztályozási kategóriák, a gráf szélein (EDGES) pedig az implicit relációk jelennek meg. A gráfban külön csoportokban tárolhatók: az ismétlődő rekordok, kereszthivatkozások a korábbi jelzetekre, a navigálással el nem érhető új kategóriák, és a böngészés kezdőoldalai. Ez a tulajdonság rugalmas adatkezelést biztosít. A statisztikai feldolgozás után a dokumentumok strukturált SOIF formátumai és az osztályozási jelzetek az Oracle relációs adatbázis tábláiban tárolódnak [6]. A rugalmas böngészés lehetőségeit az 5. és a 6. ábra szemlélteti.

The screenshot shows a web browser window titled "GERHARD - German Harvest Automated Retrieval and Directory - Microsoft Internet Explorer". The main content area is titled "NAVIGATION IN DIRECTORY" and displays a list of categories under the "GERHARD" heading. The categories are listed in a table-like format with document counts and expandable icons.

Category	Count
NUMERICAL MATHEMATICS AND COMPUTER APPLICATIONS IN MATHEMATICS (446509)	23922
COMPUTER SCIENCE (413051)	36322
COMPUTER APPLICATIONS IN / STANDARDS(0)	
COMPUTER APPLICATIONS IN / ROBOTICS(0)	
COMPUTER APPLICATIONS IN / HUMANITIES AND SOCIAL SCIENCES(0)	
COMPUTER APPLICATIONS IN / MUSEUMS(0)	
COMPUTER APPLICATIONS IN / SOCIAL SCIENCES(409)	409
COMPUTER APPLICATIONS IN / FURTHER EDUCATION AND RETRAINING(4)	5
COMPUTER APPLICATIONS IN / FINANCE AND BANKING(24)	1
COMPUTER APPLICATIONS IN / LAW(93)	100
DATA PROTECTION + SAFEGUARDING OF DATA(208)	301
COMPUTER APPLICATIONS IN / CRIMINOLOGY(0)	
DATA PROCESSING LAW(0)	
COMPUTER APPLICATIONS IN / PUBLIC ADMINISTRATION(74)	80
COMPUTER APPLICATIONS IN / INSURANCE(0)	
COMPUTER APPLICATIONS IN / EDUCATION(0)	
COMPUTER APPLICATIONS IN / TOURISM(0)	
INFORMATICS AND COMPUTER APPLICATIONS IN ENVIRONMENTAL PROTECTION AND RESEARCH(0)	
COMPUTER APPLICATIONS IN / MATHEMATICAL STATISTICS(17)	17
STATISTICAL SOFTWARE(2127)	2240
DATA PROCESSING MODELS(0)	
SIMULATION / APPLICATION TO COMPUTER SCIENCE(0)	

5. ábra Böngészés az UDCZ osztályozási kategóriák között

GERHARD - German Harvest Automated Retrieval and Directory - Microsoft Internet Explorer

File Edit View Favorites Tools Help

OVERVIEW OF DOCUMENTS

GERHARD

↑ **COMPUTER APPLICATIONS IN / LAW (100)** [continue to subordinate terms](#)
↓

navigation in directory [return to navigation](#)

search in directory

help

Feedback into preferences

ORACLE digital

attributed documents 1 on to 25

- Ⓛ [lehreveranstaltungen von prof. dr. detterbeck](http://www.uni-marburg.de/jura/oeffr/detterbeck/info/lehre.html)
- Ⓛ [forschungsstelle fuer rechtsinformatik](http://www.uni-marburg.de/jura/straf/meurer/rechtsinformatik/welcome.html)
- Ⓛ [f. krueger - hyperjuris - 314 rechtsinformatik](http://www.fask.uni-mainz.de/user/krueger/dissweb/diss-314.html)
- Ⓛ [institut fuer rechtsinformatik](http://www.tt.uni-hannover.de/forkat/0147.htm)
- Ⓛ [forschungsstelle fuer rechtsinformatik](http://www.uni-marburg.de/jura/fs/rechtsin/welcome.html)
- Ⓛ [lehreangebot von dr. edgar weiler](#)
- Ⓛ [zugriffe sommersemester1997](#)
- Ⓛ [mitarbeiter von prof. dr. georgios gounalakis](#)
- Ⓛ [prof. dr. ralph backhaus](#)
- Ⓛ [prof. dr. stephan buchholz](#)
- Ⓛ [interaktiver spaziergang durch das juristische seminar](#)
- Ⓛ [prof. dr. winfried mummenhoff](#)
- Ⓛ [www-service](#)
- Ⓛ [ausstattung des pc-saals](#)

6. ábra Egy adott osztályozási kategóriába tartozó dokumentumok megtekintése

A SCORPION program

Az OCLC (Online Computer Library Center) 1998-ban indította el a SCORPION kutatási programot, amely az automatikus osztályozás különböző módszereit vizsgálja. A program összekapcsolja az indexelést a katalogizálással. A GERHARD programhoz hasonlóan nyelvészeti és statisztikai módszereket használ [6]. A SCORPION szoftver a Dewey Tizedes Osztályozás adatállományából létrehozott kereshető adatbázisra épül, ahol a Dewey-jelzethez a megfelelő osztályozási kategóriák vannak hozzárendelve [7]. Itt tehát a dokumentum egy olyan keresőkérdésnek tekinthető, amellyel az adatbázisban keresést hajtanak végre, végeredményül pedig a dokumentum Dewey-jelzeteit rangsorolva kapják meg [3]. Szövegfeldolgozásnál a dokumentum szövegéből kiválogatják azokat a kulcsszavas kifejezéseket, amelyeket a jelzetek visszakeresésénél és rangsorolásánál figyelembe vesznek. Jelenleg is keresik azokat a kifinomult automatikus módszereket, amelyek révén a dokumentum szövegében előforduló kulcsszavas kifejezések azonnal felismerhetők [7]. A jelzetek súlyozásánál vizsgálják a kulcsszavas kifejezések előfordulási gyakoriságát az adatbázisrekordok-

ban. Koszinusznormalizálással mérik a különbséget az adatbázisrekord és az osztályozandó dokumentum vektoros reprezentációja között [8].

Az OCLC másik kutatási programja a CORC (Cooperative Online Resource Catalog) a weblapok metaadataihoz biztosít ellenőrzött hozzáférést egy rugalmas navigációs eszköz segítségével. Ez a szolgáltatás sikeresen integrálja működéséhez a SCORPION szoftvert Dewey-jelzettek előállításánál [7].

Klaszterálási lehetőségek és a SCORPION program kapcsolata

A SCORPION program keretében vizsgálják a klaszterálási módszerek felhasználását az eredményül kapott Dewey-jelzettek szűrésében. A SCORPION előre definiált klasztereket hoz létre az adatbázisrekordokból, amelyeket a Dewey-jelzettek finomítására használ. A jelzetekből képezett klasztereket súlyozzák, és az alacsony súlyozási értékeket törlik a találati halmazból. Az adatbázisrekordok előzetes klaszterálása olyan mátrix használatát igényli, amely tartalmazza a rekordpárok

közötti előre megszabott távolságértékeket. Tulajdonképpen ez a táblázat sorolja be az adatbázis-rekordokat a legközelebbi klaszterekbe. Egy új rekord abba a klaszterbe kerül, amelynek a legközelebbi klaszterhez mért távolsága alacsonyabb a megállapított küszöbértéknél [9].

Összehasonlították a hagyományos klaszterálási és a SCORPION klaszterálási teljesítményét egymással, ahol a klaszterálást mindkét esetben ugyanazon a dokumentumhalmazon végezték el. Új, objektív mérőszámokat kellett bevezetniük a hatékony összehasonlítás céljából. Egy adott klaszter hasonlóságának az értékét úgy határozták meg, hogy a benne lévő jelzetszámok közötti átlagos távolságot vették figyelembe. Ezt követően több klaszter hasonlóságának az átlagát vették, és megkapták a *klaszterek közötti hasonlóság átlagértékét*. Jó minőségű klaszterek esetében ez az érték magas, mivel a nagyon hasonló jelzetszámok kerülnek egy adott klaszterbe. Két tetszőleges klaszter maximális hasonlóságának az értékét úgy állapították meg, hogy a bennük található, egymáshoz nagyon közel álló jelzetszámok távolságát vették alapul. Ezután az összes klaszterpár maximális hasonlóságának az átlagát vették, és megkapták a *klaszterek közötti maximális hasonlóság átlagértékét*. Jó minőségű klaszterek esetében ez az érték alacsony, mivel a klaszterek jól elkülönülnek egymástól.

A SCORPION szoftverrel előállított klaszterekben a dokumentumoknak csak az első jelzetét vették figyelembe, hogy nagyságrendi eltérések ne legyenek a vizsgált klaszterek között. A vizsgálat eredményei a következők voltak: a dokumentumok jelzeteit 485 SCORPION klaszterbe és 386 hagyományos klaszterbe sorolták be. Hagyományos klaszterálásnál azonos méretű klaszterek jöttek létre. Ezzel szemben a SCORPION program néhány nagyobb klasztert és sok egy jelzetszámú álló klasztert állított elő. A klaszterek minőségi vizsgálatánál kiderült, hogy a SCORPION-nal létrehozott és a hagyományos klaszterek minősége között nincs különbség. A SCORPION klaszterek hasonlóságának átlagértéke magasabb volt, mint a hagyományosaké. Maximális hasonlóságuk átlagértéke pedig alacsonyabb volt, mint a hagyományos klasztereké. A SCORPION szoftver algoritmus gyorsabban lefutott, és kevésbé volt erőforrásigényes. Tehát a kísérleti eredmények rámutatnak arra, hogy a SCORPION szoftver hatékony és eredményes klaszterálási eszköznek minősül [9].

* * *

A jövőben alternatív fejlesztési megoldásokat kell keresni a különböző internetes szolgáltatások számára, azonban eredményeik összehasonlítása nélkülözhetetlen feladat. A hagyományos osztályozási rendszereket át kell alakítani, hogy megfelelő böngészőrendszerként funkcionáljanak ezekenél a szolgáltatásoknál. Testre szabott vizuális és navigációs technikák kifejlesztése egyaránt szükséges a rugalmas böngészéshez. Ezeknél a fejlesztéseknél valószínűleg a tartalomelemzésre, a használatra és a hivatkozásra épülő módszerek, valamint a klaszterálási lehetőségek kerülnek majd előtérbe. A neurális hálózati vagy más egyéb mesterségesintelligencia-technikák alkalmazása igéretesnek tűnik ezen a téren annak ellenére, hogy korábban más területeken alkalmazták őket [5].

Irodalom

- [1] NEWTON, Robert: Information technology and new directions. = Marcella, R.–Maltby, A.: The future of classification. Aldershot Brookfield, Vt., Gower, 2000. p. 43–57. ISBN 0-566-07992-5
- [2] ARDÓ, Anders–FALCOZ, Franck–KOCH, Traugott et al.: Improving resource discovery and retrieval on the Internet: the Nordic WAIS/World Wide Web Project – Summary Report. = Nordinfo-NYTT, 17. köt. 4. sz. 1994. p.13–28.
- [3] MACLENNAN, Alan: Classification and the Internet. = Marcella, R.–Maltby, A.: The future of classification. Aldershot Brookfield, Vt., Gower, 2000. p. 59–68. ISBN 0-566-07992-5
- [4] DESIRE (Development of a European Service for Information on Research and Education). EU Project, 1999. <http://www.lub.lu.se/desire>
- [5] ARDÓ, Anders–KOCH, Traugott: Automatic classification applied to full-text Internet documents in a robot-generated subject index. = Online Information 1999: the proceedings of the 23rd International Online Information Meeting. London, 7–9 December 1999. p. 239–246. <http://www.lub.lu.se/~anders/online99>
- [6] MÖLLER, Gerhard–CARSTENSEN, Kai-Uwe–DIEKMANN, Bernd et al.: Automatic classification of the World Wide Web using Universal Decimal Classification. = Online Information 1999: the proceedings of the 23rd International Online Information Meeting. London, 7–9 December 1999. p. 231–237.
- [7] HICKEY, Thomas B.–VIZINE-GOETZ, Diane: The role of classification in CORC. = Online Information 1999: the proceedings of the 23rd International Online Information Meeting. London, 7–9 December 1999. p. 247–250.
- [8] SHAFER, Keith–THOMPSON, Roger: Scorpion: SMART Weighting Schemes. 1997. http://orc.rsch.oclc.org:6109/smart_weight.html
- [9] SUBRAMANIAN, Srividhya–SHAFER, Keith: Clustering. 1998. <http://orc.rsch.oclc.org:6109/clustering.html>

Beérkezett: 2002. III. 12-én.