

műanyagok kémiájának mind mélyebb megértése ezt a veszélyt elhárították.

Elfogulatlanabb vélemények szerint a lemezek várható élettartama a minőségtől, a tárolás és a kezelés gondosságától függően tíz–harminc év.

Harminc év hosszú időnek tűnik, különösen, mivel ez a technika alig több mint egy évtizedes. Ha azonban megnézzük könyvtárunk mikrofilm- és mikrofilm-lap-gyűjteményét, ott bőven találunk tíz évnél öregebb példányokat. Ha a mikrofilmes tárolást CD-ROM technikával kívánjuk felváltani, az élettartam lényeges kérdés.

Mit tegyünk, hogy CD-ROM gyűjteményünk legáltalában saját nyugdíjba menetelünkig ép maradjon? Íme néhány teendő:

1. A kompaktlemezt mindig csak a szélénél fogjuk meg.

2. Ne hajlítsuk meg a lemezt, amikor kiemeljük a dobozából. A hajlítás mikropedéseket okoz a lemez felületén.

3. Használaton kívül mindig a dobozában tartuk a lemezt. Nagyon egyszerű az adatbázisváltások során egyre növekvő oszlopot rakni a lemezekből, ettől azonban azok súlyosan karcosodhatnak. Karcosítást okozhat az is, ha durva felületre letesszük a lemezt.

Előnyös az olyan CD-ROM olvasó, amelybe a lemezt kazettában kell betenni. Ha minden lemeznek saját kazettát veszünk, akkor a lemezt kézbe vennünk csak kétszer kell: egyszer akkor, amikor áttesszük a kazettába, másodszor pedig akkor, amikor lecseréljük.

4. Az olvasóba való minden behelyezés előtt ellenőrizzük a lemez felületét, ha a lemezt nem kazettá-

ban tartjuk. Tisztítására nagyon bevált a pamutpelenka. Olcsó, hatékony, és még a beszélgetést is holtbiztosan megindítja. Tisztítószert, oldószert soha ne használjunk! Tönkretelheti a lemez műanyag felületét*.

5. Soha ne írjunk, ne ragasszunk címkét a lemez felületére. A tinta vagy a ragasztó vegyi anyagai behatolhatnak a védőborításba.

6. Időszakonként ellenőrizzük lemezeink felületét, hogy nincs-e rajtuk repedés, bemetszés vagy karcosítás. Jegyezzük fel a hibákat, hogy lássuk, mely lemezeken térnek azok vissza.

7. A sérült lemezeket azonnal cseréljük le. A legtöbb szállító nagylelkű cserepolitikát folytat. Használjuk ki ezt a nagylelkűséget, amíg tart.

/KING, A.: The care and feeding of your CD-ROM disk. = Database, 14. köt. 6. sz. 1991. p. 105–107./

(Válasz György)

* Más szerzők hangsúlyozzák, hogy a lemezt mindig csak sugárirányban szabad törölni, soha nem az adatsávval párhuzamosan. Sugárirányú töréskor az esetleg keletkező karcosítás egy-egy adatblokkban csak néhány bitet tesz tönkre, így a CD-ROM technika hatékony hibajavító mechanizmusa még rekonstruálhatja az adatokat. Az adatsávval párhuzamos törés esetleges karcosításából hosszú darab eshet egyazon adatblokkba, így a hiba helyrehozhatatlan. A kompakt hanglemezekhez készült lemeztisztító szerkezetek körkörösén törölnek (a hanglemez kevésbé kényes a sérülésre, néhány bit hibát a zenében a legélesebb fűlű hozzáértő sem hall meg), ezért a használatuk adatlemezekhez tilos. – A ref.

A rangsorolási információkeresés hatékonysága különböző súlyozási eljárások esetén

Bár közel harminc éve folynak kísérletek olyan információkeresési eljárások kidolgozására, amelyek nem teszik szükségessé, hogy a kereső Boole-algebrai kifejezéseket állítson össze, kevés ilyen típusú keresőrendszer működik, és egyikük sem terjedt el szélesebb körben.

Az explicit Boole-algebrai összefüggéseket mellőző rangsorolási keresőrendszerek közül a legegyszerűbb a *Quorum match* (a határozatképességhez szükséges legkisebb létszám egyeztetése) módszer, amelyet a Cranfield-2 indexelési nyelv kísérletei során dolgoztak ki [1]. Ez a rendszer egyszerűen a keresőkérdés szavai és az adatrekord szavai közötti egyezések száma szerint értékeli az adatrekordokat, majd az ilyen egyezések csökkenő száma szerinti sorrendbe rendezi őket. Már a Cranfield-2 program során felmerült az eljárás tökéletesítésének igénye [1, 2].

A javasolt tökéletesítési módok többsége az egyezések előfordulási gyakoriság szerinti súlyozásával operál [3]. Az itt leírt összehasonlítási kísérletben ezek közül két eljárást vontunk be. A tökéletesítési kísérletek másik része párokba rendezi a keresőkifejezéseket, és a pár két tagjának a rekordon belüli egymás közötti távolsága szerint súlyozza a találatokat [4]. Az ilyen eljárások közül egyet vontunk be kísérletünkbe. Az így kapott három összehasonlítandó eljárás mellé negyediknek egy szubjektív súlyozásos módszert választottunk.

Egyszerű súlyozásos eljárások

Az eljárások összehasonlítása céljából végzett kísérlet során a *Library and Information Science Abstracts* 1982-es évfolyamának 6004 adatrekordjában végeztünk 35 keresést. A keresőprofilok meglehetősen hosszúak voltak, 7–26 keresőkifejezést tartal-

maztak. A keresés négy adatmezőben történt: a cím, a kivonat, a fő tárgykör és a deskriptor mezőkben, ezek rekordonkénti átlagos hossza 63 kereshető szót tett ki. Az eljárások hatékonyságát a "Quorum match" eljárásával hasonlítottuk össze, amely az egyezéseket 1, a nem egyezéseket 0 faktorral súlyozza.

A négy vizsgált súlyozási eljárás közül az első az *összesített gyakoriság* módszere. Ez az eljárás annál nagyobb súlyt ad egy szónak, minél ritkábban fordul elő az összes adatrekordban együttesen (tehát minél karakterisztikusabb keresőszó), mégpedig az előfordulási szám *logaritmusával* fordított arányban osztja ki a súlyokat [5]. Ha például a "search services in polytechnics" kifejezéssel kerestünk a fenti mintában (az *in* előjáró nem kereshető szó), akkor a *search* szó 691-szer fordult elő, ezért 5,72 súlyt kapott, a *service(s)* szó 3594-szer fordult elő, 1,74 súlyt kapott, a *polytechnic(s)* szó 50-szer fordult elő, 7,91 súlyt kapott. A teljes kifejezés súlya a három súly összege, 15,37. Az összesített gyakoriság módszerét szinte valamennyi rangsorolós keresőrendszer alkalmazza vagy önmagában, vagy más módszerrel kombinálva. Ez az eljárás azokat az adatrekordokat emeli ki, amelyekre az egyezés ritka, tehát karakterisztikus keresőszóval áll fenn.

A másik vizsgált súlyozási módszer a *rekordonkénti gyakoriság* módszere. Ez akkora súlyt ad az egyes szavaknak, ahányszor előfordulnak a találatrekordban. Megint az előbbi példához visszatérve, a vizsgált minta 400. adatrekordjában a *search* szó hat-szor fordul elő, a *service(s)* szó kétszer, a *polytechnic(s)* szó egyszer sem, így a rekordnak adott súly $6 + 2 + 0 = 8$. Ez az eljárás azokat az adatrekordokat emeli ki, amelyekben a keresőszó sokszor ismétlődik, ami azt jelzi, hogy a referált dokumentum tárgya szempontjából ez a szó fontos. A rekordonkénti gyakoriság módszerét a rangsorolós keresőrendszerek önmagában sohasem alkalmazzák, viszont gyakran alkalmazzák más eljárással kombinálva.

A harmadik vizsgált súlyozási eljárás, a *páronkénti távolság* módszere megint csak az adatrekord tartalmán alapszik. A pozicionális információt hasznosító számos módszer [4] közül azt választottuk, amely a keresőszavakból párokat képez, és minden ilyen párnak aszerint adja a súlyát, hogy a két szó egymástól milyen távolságban található az adatrekord egy-egy mondatán, adatmezőjén belül. 10-es súlyt kapnak a szomszédos párok, 9-est azok, amelyeket egy közbülső szó választ el, és így tovább, 1-es súlyt azok, amelyeket kilenc vagy több szó választ el. A páronkénti súlyok összegét 10-zel osztva kapjuk a rekord súlyát [5]. Ez az eljárás azokat az adatrekordokat emeli ki, amelyekben a keresőszavak egymáshoz közel helyezkednek el, tehát egyazon kifejezés szavait képezik.

A negyedik vizsgált eljárásban a kereső *szubjektív súlyokat* ad az egyes keresőszavaknak, aszerint, hogy mekkora jelentőséget tulajdonít nekik. A súlyozás alapelvei a következők. A súlyok 1-től 6-ig

adhatók. A keresés szempontjából központi jelentőségű keresőszavak ezen a skálán belül nagy súlyt kapnak. Azok az alternatív kifejezések, amelyekről feltételezhető, hogy mindegyikük előfordul az adatrekordban, egyforma súlyt kapnak, mégpedig elég kicsit ahhoz, hogy együttesen se váljanak túlonként dominánssá. Azok az alternatív kifejezések, amelyekről valószínűtlen, hogy egyazon adatrekordban előfordulnak, akkora súlyt kapnak, hogy bármelyikük külön-külön is elegendő súlyt adjon a rekordnak. Például a 32. keresés tárgya a következő volt: "online searches in polytechnics or universities". Itt az *online* szó 5-ös súlyt kapott, ugyanennyit kapott a kétszavas *on line* kifejezés is. A *polytechnic(s)* és a *university(ies)* szavak súlya külön-külön 1-1. Egy másik keresésben az *SDI* rövidítés súlya 4, a vele egyenértékű *selective dissemination (of) information* kifejezés szavainak súlya 2-1-1. Az egyes kereső-kifejezésekre képezett súlyokat a találatrekordra összegeztük. Ez az eljárás azokat az adatrekordokat emeli ki, amelyek a kérdésfeltevő által központiak ítélt szemponttal foglalkoznak. Nem ismerünk olyan rangsorolós keresőrendszert, amely ezt az eljárást alkalmazná.

Kombinált súlyozási eljárások

Az egyes súlyozási eljárásokkal külön-külön végzett kísérletek után a négy eljárás minden lehetséges kombinációjával is elvégeztük a kísérletet: a 6 párral, 4 hármassal és az összessel együtt, vagyis végeredményben 15 kombinációval. A legkézenfekvőbb a párok vizsgálata volt, mivel a legtöbb kísérleti vagy üzemszerű rangsorolós keresőrendszer két-két eljárás kombinációját alkalmazza. Csak egy olyan rendszert ismerünk, amely mind a négy módszert kombinálja.

Néhány kombináció logikusan indokolható. Az összesített gyakoriság és a rekordonkénti gyakoriság módszerének kombinációja például azokat a rekordokat emeli ki, amelyekben kis összegű gyakorisággal rendelkező kifejezések sokszor fordulnak elő, ami nagy relevanciára ad reményt. Ugyanígy reményt keltő az összesített gyakoriság és a páronkénti távolság módszerének kombinációja, viszont a rekordonkénti gyakoriság módszerének kombinálása a páronkénti távolság módszerével nem látszik indokoltnak.

A különféle módszerekkel kapott súlyok egyesítésére nyilván számos algoritmus képzelhető el. Az egyik lehetőség például az, hogy az egyik módszert elsődleges módszerként kezeljük, a másikkal csak az így egyazon helyre rangsorolt rekordokat rendezzük tovább egymás között.

Mi három összegzési algoritmussal kísérleteztünk:

- összegeztük az egyes módszerekkel az egyes egyezésekre kapott súlyokat;
- egy rekordon belül az egy-egy módszerrel kapott súlyokat összegeztük valamennyi egyezésre, majd az így módszerenként kapott összegeket összeszorzottuk;

c) egy rekordon belül az egy-egy módszerrel kapott súlyokat valamennyi egyezésre összeszoroztuk, majd az így módszerenként kapott szorzatokat összeadtuk.

A három algoritmus összehasonlítását csak néhány kiválasztott esetre végeztük el. Ezek a próbálkozások azt mutatták, hogy az eredmény nem nagyon függ a választott algoritmustól. Végül a b) algoritmus mellett kötöttünk ki, mert az egyes esetekben a legjobbnak bizonyult, emellett egyszerűbben alkalmazható, mint a c) algoritmus.

A kapott (néha több százat elérő) súlyokat keresésenként normáltuk, a legjobb egyezést elért rekordok súlyát véve 100%-nak.

A kísérletek eredménye

A kísérletek eredményét az 1. táblázat foglalja össze. Ebben három teljességi arány szerepel. Összeszámoltuk, hogy az egyes keresési stratégiákra hány releváns találatot tartalmaz a teljes (6004

rekordból álló) minta. A rangsorolt rekordsort ott vágtuk el, ahol a releváns találatok száma elérte az összes releváns találat 25 százalékát (kis teljesség), 50 százalékát (közepes teljesség), illetve 75 százalékát (nagy teljesség), majd meghatároztuk a rangsorolt rekordsor tetejéből ilyen vágással kapott halmaz pontosságát, végül az eredményt a 35 keresésre átlagoltuk. A pontosságra két mérőszámot képeztünk. A százalékos arány azt fejezi ki, hogy a vágással kapott részalmaz rekordjai között milyen arányban vannak a relevánsok, a rangsorpozíció pedig azt, hogy hányadik rekord után történt a vágás. A tizenöt kísérleti módszer mellett összehasonlításként felvettük a táblázatba a Quorum match módszerrel kapott eredményt, és a lehetséges legjobb eredményt is. Ez utóbbit úgy kaptuk, hogy minden releváns rekordot kikerestünk, és ezeket a rangsor elejére tettük. Ekkor tehát csak releváns rekordok kerültek a vágás fölé, a pontosság 100%. Átlagosan 8 releváns találat volt keresésenként,

1. táblázat

Kísérleti eredmények a LISA adatbázis 6004 adatrekordjából végzett 35 kereséssel

Sorszám	Súlyozási eljárás	Kis teljesség Küszöb: 25%		Közepes teljesség Küszöb: 50%		Nagy teljesség Küszöb: 75%	
		P. a.	Rs. p.	P. a.	Rs. p.	P. a.	Rs. p.
Ö. h.	Leh. lj.	100%	2	100%	4	100%	6
Ö. h.	Quorum	45,4%	9	31,1%	28	22,0%	79
1	A	62,4%	4	51,3%	14	34,5%	44
2	B	56,8%	5	42,5%	25	29,8%	53
3	C	49,1%	8	39,6%	22	27,0%	54
4	D	41,2%	10	33,2%	34	24,8%	79
5	AB	70,6%	4	54,0%	12	38,2%	44
6	AC	68,7%	4	54,7%	12	39,4%	32
7	BD	65,3%	5	51,1%	16	36,8%	48
8	AD	59,6%	6	52,0%	15	33,5%	53
9	CD	58,8%	7	47,1%	16	34,3%	47
10	BC	56,6%	5	41,7%	21	29,8%	57
11	ABC	68,4%	4	52,8%	13	39,1%	34
12	ABD	68,4%	5	57,4%	11	37,9%	44
13	ACD	68,4%	4	53,3%	12	37,0%	40
14	BCD	66,9%	5	50,9%	15	36,9%	50
15	ABCD	71,4%	4	56,5%	12	39,4%	38

Súlyozási eljárások:

- Leh. lj.: az elvileg lehetséges legjobb eljárás
 Quorum: Quorum match eljárás
 A: a páronkénti távolság módszere
 B: az összesített gyakoriság módszere
 C: szubjektív súlyozás
 D: a rekordonkénti gyakoriság módszere
 Betűkombináció: kombinált rendszer

Fejlécben:

- P. a.: pontossági arány
 Rs. p.: rangsorpozíció
 Ö. h.: összehasonlító eljárás

tehát a három kiválasztott teljességi aránynak a 2, 4 és 6 rangsorpozíció felel meg. A táblázatban az egyszerű súlyozási eljárásoknak betűjelet adtunk:

- A – a páronkénti távolság módszere,
- B – az összesített gyakoriság módszere,
- C – a szubjektív súlyozás,
- D – a rekordonkénti gyakoriság módszere.

A kombinált súlyozási módszereket betűkombinációk jelzik.

Ha először a négy egyszerű súlyozási eljárást hasonlítjuk össze, ezek közül valamennyi teljességi szinten a páronkénti távolság módszere adja a legjobb eredményt, ezt az összesített gyakoriság módszere követi, majd a szubjektív súlyozás (kivéve közepes teljességen a rangsorpozíció értékét). A rekordonkénti gyakoriság módszere adja a leggyengébb eredményt. A Quorum match módszernél ezek mindegyike és minden kombinált módszer jobb eredményt ad, kivéve a rekordonkénti gyakoriság módszerét kis teljességnél.

Az ötödiktől a tizedikig a két módszerből kombinált eljárások találhatók. Ezek körében nagyon leszűkül a távolság a legjobb és a leggyengébb között.

A páronkénti távolság és az összesített gyakoriság módszerének kombinációja adja kis teljességen a legjobb eredményt, ezt szorosan követi a páronkénti távolság kombinálása a szubjektív súlyozással, sőt, közepes és nagy teljességen ez a legjobb. Szorosan követi az első kettőt a klasszikus kombináció: az összesített gyakoriság és a rekordonkénti gyakoriság módszeré. 5% körüli eltéréseket jelentett a pontossági arányban az, hogy a módszereket melyik algoritmussal kombináltuk. Mint várható volt, a rekordonkénti gyakoriság módszere nem javítja a páronkénti távolság módszerét, a 9. és 10. módszer pedig értelmetlen, mert ezeknél az egyszerű módszerek között is akad jobb.

A hármas kombinációk körében már 3% alá csökken a legjobb és a leggyengébb közötti különbség a pontossági arányban. A három objektív módszer kombinációja különösen közepes teljességnél emelkedik ki, itt ez az összes verzió közül a legjobb.

Ha a három teljességi szintre átlagolunk, akkor a négyes kombináció adja a legjobb eredményt.

Végezetül összehasonlítottuk a vizsgált módszereket a Boole-logika szerinti kereséssel. E hagyományos módszerrel egy kezdő kereső mindössze 25% teljességet és 20% pontosságot ért el, míg egy tapasztalt kereső 66% teljességig és 35% pontossáig jutott el, de a keresés közbülső, kisebb teljességet elérő szakaszaiban sem haladta meg egyszer sem az átlagosan 44% pontossági arányt. Ez az eredmény azt sugallja a szerzőnek, hogy a Boole-logika szerinti keresés eredményessége elérhető, sőt meghaladható a rangsorolós kereséssel.

Rendszertechnikai és felhasználói értékelés

Bár a hatékonyság döntő szempont, a rendszer megvalósíthatóságát és a felhasználói fogadtatást is

figyelembe kell venni az értékelésben. Figyelembe kell venni például olyan számítástechnikai tényezőket, mint a válaszidő, a CPU-, lemezterület- és tárigény, a szükséges állományszerkezet és aktualizálási megoldás stb.

Az egyszerű súlyozási eljárások közül a szubjektív súlyozás és az összesített gyakoriság módszere számítástechnikai igénye nem nagy, mert ezekhez használhatók az invertált állományok, és előre tárolhatók az előfordulási gyakoriságok. A rekordonkénti gyakoriság módszere bonyolultabb állományszerkezetet és több számítást igényel, a páronkénti távolság módszerének a számítástechnikai igénye még ennél is nagyobb. Ez utóbbihoz tárolni kell a helyzeti információt is, de ez az igény a helyzeti operátorokat megengedő logikai keresőrendszerek esetén is fennáll.

Ami a felhasználói fogadtatást illeti, a legkisebb számítástechnikai igényt jelentő módszer, a szubjektív súlyozás igényli a legtöbb erőfeszítést a felhasználótól, ezért elfogadhatósága kétséges. Bár a páronkénti távolság módszere mostani formájában automatikus eljárás, bizonyos felhasználói beavatkozás a "jó" és a "rossz" párok szétválasztásában előnyös lehet. Legalább a szükséges kizárásokat meg kell határozni a felhasználónak (a logikai ÉSNEM művelet megfelelőjét). Van olyan megvalósítás, amelyik arra is módot ad, hogy egyes párokat a felhasználó elengedhetetlenül szükségesnek jelöljön ki. Ha a keresés végén megjelenik az egyezések táblázata, az segíti a felhasználót abban, hogy szükség esetén finomítsa a keresést.

A keresőrendszer használhatóságának és felhasználóbarát voltának vannak olyan vonásai is, amelyek laboratóriumi körülmények között nem vizsgálhatók, szélesebb körű tesztet igényelnek.

Következtetések

A rangsorolós keresőrendszerekben használatos súlyozási eljárásoknak a szokásosnál szélesebb körét vizsgáltuk. Úgy találtuk, hogy a páronkénti távolság módszere meglehetősen hatékony, ezért még további vizsgálatokat, és alkalmazásnak látszik az összesített gyakoriság és a rekordonkénti gyakoriság módszerével kombinálva a rangsorolós keresőrendszerekben történő felhasználásra. A szubjektív súlyozásnak is megvan a maga haszna, de csak másodlagos eszközként. A kombinált módszerekről nehéz határozott következtetést levonni. Ezek hoznak némi javulást a hatékonyságban, de a számítástechnikai igényt is megnövelik. Tervbe vettük a vizsgálatok folytatását más témakörökben, különböző típusú keresésekre.

Irodalom

- [1] CLEVERDON, C. – KEEN, M.: Factors determining the performance of indexing systems: Volume 2, Test results. = ASLIB Cranfield Research Project, Cranfield, 1966.

- [2] CLEVERDON, C.: The Cranfield tests on index language devices. = ASLIB Proceedings, 19. köt. 6. sz. 1967. p. 173–192.
- [3] WILLET, P. (szerk.): Document retrieval systems. Taylor Graham. = Institute of Information Scientists, London, 1988.
- [4] KEEN, E. M.: The use of term position devices in ranked output experiments. = Journal of Documentation, 47. köt. 1. sz. 1991. p. 1–22.
- [5] KEEN, E. M.: The effectiveness of term position and frequency for output ranking. = Proceedings of the British Computer Society 13th Research Colloquium on Information Retrieval, Lancaster, 1991. április.

/KEEN, E. M.: Query term weighting schemes for effective ranked output retrieval. = Proceedings (of) 15th International Online Information Meeting, 10–12 December 1991, London, England; Learned Information, Oxford and New Jersey, 1991. p. 135–142./

(Válasz György)

Az ISBN rendszer alkalmazása számítógépes termékekre

Sok éve vannak piacon különböző számítógépes információs kiadványok, például adatbázisok. Kiadjuk nemegyszer nyomtatott termékeket is kiad. Mindezek ellenére a számítógépes kiadványok mind bibliográfiai leírásuk szabályozottságában, mind szerzői jogvédelmük szintjében jócskán elmaradnak a nyomtatott kiadványoktól.

Mind több az olyan kiadvány, amely kizárólag számítógépes formában jelenik meg, és ezek egyre többféle hordozóközegen látnak napvilágot, például online adatbázisként, CD-ROM formában, mágneslemezen. Ezért nyilvánvaló, hogy szükség van szabványos, egységes azonosításukra és számozásukra.

Új számozási rendszer kifejlesztése helyett célszerűbbnek látszik a meglévő szabványos nemzetközi rendszerek valamelyikének adaptálása.

A megoldandó problémák

A hagyományos adatbázisokkal kapcsolatban a következő azonosítási problémák merülnek fel:

- ▶ Megegyező adatbázisok eltérő neveken szerepelhetnek.
- ▶ Eltérő adatbázisoknak azonos nevük lehet.
- ▶ Azonos nevű adatbázisok eltérő időszakot fedhetnek le.
- ▶ Egyazon adatbázis több különböző hordozóközegen jelenhet meg.

Éppen a legnépszerűbb adatbázisok, a *Chemical Abstracts*, az *ERIC*, a *MEDLINE*, az *AGRICOLA* több különböző változatban léteznek, több különböző hordozóközegen, eltérő időbeli lefedéssel.

A felhasználó könnyen tudomást szerezne arról, hogy két azonos nevű kiadvány között eltérés van, ha egységes számozási rendszerük lenne ezeknek a kiadványoknak, és az előállító az állomány eltérő változatainak eltérő számokat adhatna. Emellett az egységes számozási rendszerben kiosztott számok a bibliográfiai leírásnak is részét képezhetnék. Így a felhasználó több problémáját megoldaná az egységes számozás.

Hasznos lenne az egységes számozás a kiadó számára is, mégpedig nemcsak azért, mert így meg-

különböztetné egymástól a kiadvány különböző változatait, de nem utolsósorban azért is, mert ez jelentős lépés lenne a szerzői jogvédelem megvalósítása felé.

Követelmények

A fenti problémák megoldhatóvá tétele megköveteli néhány követelmény kielégítését:

1. Nemzetközi legyen a számozási rendszer.
2. Szabványos legyen a számozási rendszer.
3. A számozási rendszert fogadják el az előállítók, és ők osszák ki a számokat.
4. A számok kiosztására álljon rendelkezésre nemzetközi infrastruktúra.

Az első két követelmény kézenfekvő. Ha a számozási rendszer nem nemzetközi vagy nem szabványos, akkor a haszna csekély.

Ami a harmadik követelményt illeti, fontos, hogy az előállító "számkiosztási egység" legyen. Egyrészt azért, mert az előállítók tartják valójában kézben a kiadványt, másrészt mert ők rendelkeznek a számkiosztáshoz szükséges információval. Emellett később, ha megszületik a bibliográfiai leírás szabványa, és megvalósul a szerzői jogvédelem, ők lesznek leginkább érdekelték a számozási rendszerben.

Végül a számkiosztás nemzetközi infrastruktúrájára azért van szükség, mert enélkül soha meg nem valósulhat a nemzetközi szabvány.

Sok fejlesztési munka megtakarítható, ha a meglévő számozási rendszerek valamelyikét adaptáljuk a számítógépes kiadványokra. A két meglévő szabványos nemzetközi rendszer közül az ISSN nem elégíti ki a harmadik követelményt, nem tartalmazza az előállító azonosítóját, ezért csak az ISBN-re gondolhatunk.

ISBN a számítógépes kiadványokra

Az ISBN rendszerben az azonosító számot az előállító vagy a kiadó osztja ki. Ez az adatbázisok esetében az a szervezet, amely az adatbázist előállítja és kiadja, nem pedig az, amely azt online elérhetővé teszi, szolgáltatja. Ez utóbbit inkább a könyv terjesztőjéhez hasonlíthatjuk.