

## INDEXELÉS – OSZTÁLYOZÁS

## A kiemeléses indexelés

Az információs rendszer felhasználóját e rendszeren belüli kommunikációs csatornának tekinthetjük. Mivel a csatorna kapacitása véges, ezért egy felhasználó korlátozott mennyiségű információt képes befogadni. A feldolgozás határfoka akkor maximális, ha a kapott információ teljes, releváns, és nem redundáns.

A gyakorlatban azonban az információs rendszerek nem felelnek meg a fenti követelményeknek. A költségek csökkentése például megköveteli, hogy az egyes dokumentumokat a lehető legszélesebb felhasználói igények szerint indexeljék. Ez viszont a visszakeresés költségeit növeli meg. Ezen alapszik az ún. *dokumentumra orientált indexelési elv*, amelynek lényege, hogy a deskriptorokat a dokumentumban tárgyalt fogalmak egyszerű azonosítása alapján választják ki, figyelmen kívül hagyva a dokumentum mélyebb tartalmi összefüggéseit, környezetét. Az ilyen indexelési eljárás jól alkalmazható bizonyos gépi módszereknél, pl. a KWIC-indexeknél.

Egy másik indexelési elv az ún. *kérdésre orientált módszer*. Ennek lényege: ha az indexelők előre ismerik a rendszer felhasználóinak igényeit, akkor a dokumentumokat ezeknek megfelelően elemzik és indexelik. Ehhez alaposan ismerni kell a felhasználói tématerületet, és nem lehet azonos indexeléssel többféle felhasználói csoport igényeinek eleget tenni. Ezért ez a módszer jóval drágább a dokumentumra orientált indexelésnél.

E két módszer hátrányait részben kiküszöböli a javasolt új indexelési elv, amelyet *kiemeléses indexelésnek* neveztek el. Lényege, hogy a deskriptorok pontos, de kritikátlan kiválasztása helyett a *deskriptorokat* egy adott dokumentumban *azok viszonylagos fontosságának sorrendjében határozzák meg*, vagyis a kiemelt deskriptorok jegyzéke fogalmi összefüggéseket is tartalmaz.

## Összehasonlítás a súlyozott indexelési módszerekkel

Az ismert deskriptor súlyozási eljárások – pl. a szavak előfordulási gyakorisága szerinti súlyozás, a kapcsolatok elemzése, a szerepjelölő és a kapcsolatjelölő segédesszközök alapján végzett indexelés stb. – általában arra irányulnak, hogy az adott dokumentum tartalma, összefüggései szerinti indexelésre nyíljon mód, ill. a *dokumentum kulcselemeit logikailag lehessen megállapítani*. Mindez azonban dokumentumra orientált eljárás, és nincs meg az a tulajdonsága, hogy a felhasználó a visszakeresett dokumentum potenciális értékét előzetesen megbecsülhesse; további bonyolult kereső stratégiára van szükség, ahhoz hogy a releváns információ mennyiséget elfogadható mértékre korlátozzuk.

## Az új módszer alapfeltevései

A kiemeléses indexelés a következő három fő hipotézisen alapul:

*A fogalmi indexelés stabilitása.* Ha egy adott dokumentumot elemző indexelők csoportjának olyan megbízást adunk, hogy a dokumentum fogalmainak deskriptorokkal való jellemzése helyett rangsorolják a fogalmakat fontosságuk szerint, következetesebb eredményt kapunk.

*A fogalom, definiálásának stabilitása.* A témakörben járatos szakember a rendelkezésére bocsátott deskriptorokat úgy tudja rangsorolni, hogy azok megfelelnek érdeklődési körének akkor is, ha a deskriptorok nem fedik pontosan a keresőkérdést.

*A fogalom visszakeresésének stabilitása.* A témakört ismerő szakember az érdeklődési körében megjelent dokumentumnak megfelelő deskriptorok jegyzékének birtokában nagyobb eséllyel határozhatja meg a dokumentum potenciális értékét akkor, ha a deskriptorok fontosságuk sorrendjében vannak rangsorolva.

## A kiemeléses indexelés ismertetése

A javasolt új módszer *kétlépéses indexelési műveletet* kíván. Az első lépésben a dokumentumot mélységében indexeljük; a hagyományos eljárással kapott deskriptorokat *változóknak* nevezzük. Ezt követően a változók jegyzékét a dokumentumban elfoglalt szerepük hangsúlyossága, vagyis kiemelkedő jelentőségük szerinti rangsorba állítjuk. A változóknak ezt a láncát a dokumentum *nyomjelzőjének (trace)* nevezzük. Az ebből készült *mátrix-ábrázolást* – a következő pontban bemutatott példa szerint készítve – a *dokumentum karakterisztikájának* nevezzük.

## A módszer fő jellemzői:

azonos szakterületet azonos perspektíva szerint tárgyaló két dokumentum megegyező nyomjelzőt és karakterisztikát szolgáltat. Ezt a helyzetet *tökéletes illeszkedésnek* tekintjük;

azonos szakterületet különböző perspektíva szerint tárgyaló két dokumentum változói megegyeznek, nyomjelzője és karakterisztikája azonban eltérő. Ez a helyzet a *kevert illeszkedés*;

részben átfedő területeket tárgyaló két dokumentum néhány közös változóval jellemezhető, ezek közös nyomonvonala és karakterisztikája aszerint ad teljes vagy kevert illeszkedést, hogy a közös területet azonos vagy különböző perspektíva szerint tárgyalja a dokumentum.

## Példa a kiemelés indexelésre

Legyen az indexelendő dokumentum tárgya lasersurgázzal aranytargetből kiváltott fotoelektromos emisszió vizsgálata elektronográfiai kamerával. A dokumentumot a következő deskriptorokkal lehet jellemezni (angol ábécé szerinti sorrendben):

1. elektronográfia, 2. arany, 3. műszertechnika,
4. laser, 5. fotoelektromosság, 6. piko-másodperc,
7. felvevőberendezés.

E deskriptorokat a következőképpen lehet rangsorolni, a dokumentumban elfoglalt hangsúlyos szerepük szerinti sorrendben kiemelve:

1. Az 5. sz. deskriptor: fotoelektromosság
2. A 4. sz. deskriptor: laser
3. A 7. sz. deskriptor: felvevőberendezés
4. A 3. sz. deskriptor: műszertechnika
5. Az 1. sz. deskriptor: elektronográfia
6. A 2. sz. deskriptor: arany
7. A 6. sz. deskriptor: piko-másodperc.

A dokumentum nyomjelzője a deskriptorok ilyen sorrendben való felírása, vagy kódolva: 5, 4, 7, 3, 1, 2, 6. A karakterisztika mátrix alakja, ha 1 jelöli a megfelelő deskriptort a megfelelő rangsorban:

Rangsor \ Deskriptor	1	2	3	4	5	6	7
1	0	0	0	0	1	0	0
2	0	0	0	0	0	1	0
3	0	0	0	1	0	0	0
4	0	1	0	0	0	0	0
5	1	0	0	0	0	0	0
6	0	0	0	0	0	0	1
7	0	0	1	0	0	0	0

## A módszer előnyei

**Flexibilitás.** Legfontosabb előnye az, hogy a felhasználókat nem készíti találgatásokra a dokumentumot illetően, és feloldja a költségek és az indexelés specifikussága közötti ellentmondást azzal, hogy a hangsúlyt átteszi a megfelelő kifejezések kiválasztásáról azok kombinációjára. A deskriptorok korrekt megválasztása ugyanis nem annyira jelentős, mint azok összefüggései, tehát viszonylag általános deskriptorok is precíz indexelést tesznek lehetővé.

**Általánosság.** Nincs szükség a felhasználó perspektíva meghatározására, elegendő érdeklődési körének ismerete.

Ezért azonos output mélységben és/vagy perspektívában eltérő igényeknek is eleget tehet.

**Az alkalmazás egyszerűsége.** Mivel az új módszer nem használja a bonyolult Boole-algebrát, kis visszakereső rendszerekhez is jól használható. A keresőkérdések és a rendszer output-jának kifejezései egyszerűek, így a felhasználó közvetlen kapcsolatba léphet az információs rendszerrel.

## Alkalmazások

A kiemelés indexelésnek három fő alkalmazási területét különböztethetjük meg:

az első az irodalom felmérése, ugyanis az adatbázisból azok a dokumentumok kereshetők ki, amelyek a keresőkérdés nyomjelzőjével és karakterisztikájával bizonyos mértékben illeszkednek;

a második fő alkalmazási terület a szelektív információterjesztés, mivel a perspektíva megadása lehetővé teszi a nagy adathalmazból való közvetlen szelekciót, megfelelő pontossággal;

a harmadik terület az irodalmi anyag felduzzadásának kézbentartása. A nyomjelzőkön és a karakterisztikákon alapuló illesztés jó gyakorlati lehetőség a redundáns információk kizárására.

A módszert sikerrel alkalmazzák a Queens College Könyvtártudományi részlegében.

/D'ALLEYRAND, M. R.: *Emphasis indexing = The Indexer*, 10. köt. 2. sz. 1976. p. 70-72./

(Roboz Péter)



## Tezaurusz egy szerteágazó szakterület számára

A kisebb szakkönyvtárak legjellemzőbb ismérve a felhasználók speciális igényeihez szabott gyűjtemény. Ez a felhasználónak kellemes tény viszont a könyvtárost nehéz feladat elé állítja, mert az általános gyűjtemények számára készült tárgymutató-rendszer az adott célra rendszerint túlságosan nagyvonalú és pontatlan. Ebből a szempontból a nem-természettudományos szakkönyvtárak helyzete különösen nehéz.

Az amerikai CCL (Center for Creative Leadership = Vezetőképző Központ) könyvtárában úgy döntöttek, hogy egy saját igényekre szabott tezauruszt dolgoznak ki a könyvtárépítés és információkeresés céljaira.

A könyvtár gyűjtőköre ugyan szűk területre, a szervezés és vezetéstudomány tárgykörére irányul, de a tudományok igen széles köréből kell az idevágó dokumentumokat összegyűjteni. Ezek: matematika, pszichológia, szociológia, történelem, közgazdaságtan, oktatás stb.