

EGY PERMUTÁCIÓS INDEX-MODELL DESKRIPTORALKOTÁSI SZABÁLYAI

Schiff Ervin

Országos Műszaki Könyvtár és Dokumentációs Központ

Az OMKEK Fejlesztési és Hálósatszervezési Igazgatósága azt a feladatot kapta, hogy egy kisebb dokumentum gyűjteményre /kb. 80-1000 egység/ mély tartalmi feltárás segítségével /kötetenként 40-50 deskriptor/ a dokumentumok manuális visszakeresésére alkalmas /évi mintegy 50 egységgel bővíthető, kumulálható/ produktum modelljét állítsa elő. Az elvi /és gyakorlati/ modell 1968 februárjában készült el az Igazgatóság író-szervező automatái által nyújtott lehetőségek kihasználásával.

E közleményben a modell elvi szervezése közben felmerült néhány gondolatot szeretnénk összefoglalni, míg a műszaki-technikai problémákkal és megoldásokkal itt nem foglalkozunk.

A visszakeresésre alkalmas végső produktum a deskriptorok láncából kialakított permutált index, illetve a megbízó kérésére egy kartotéktár, amelyben a betűrendben tárolt céaulák egy-egy indexsört és a dokumentum azonosítási jelét tartalmazzák. Az indexet kumulálásával nagyobb időközönként hozzuk naprakész állapotba, míg a kartotéktár állandóan naprakészen tartható.

A deskriptorokat tartalmazó permutált indexek készítésére a világszerte kialakult különböző módszerek az elvben lehetséges két szélsőséges lehetőség, a közlemény tartalmát leíró

szabad deskriptorok mindenféle rendszer nélkül egy vagy több láncba írása és a deskriptorláncok megfelelő feldolgozása /ciklikus permutálás, betűrendbe rendezés stb./, vagy

kötött deskriptorok szabályozott rendezése alapján kialakított deskriptorláncok feldolgozása

közötti változatokat valósítják meg.

OROSZ Gábor és PATAKY Ernő javaslata alapján - mely szerint e feladatra az első típusba sorolható index készítését tartották célszerűnek - indultunk el a modell kidolgozásánál és kisebb változta-

tásokkal ezt a gondolatot valósítottuk meg. A munka közben felmerült problémák megoldását célzó változtatások azonban, úgy érezzük, bizonyos általános jelleggel is bírnak. Az egyes problémák felmerülésének és megoldásának történeti sorrendjét mellőzve, a modell szervezését kialakító megfontolásokat az alábbiakban foglaljuk össze.

Már a deskriptorok megalkotásánál különböző szabályokat kellett létrehozunk, hogy a magyar nyelv által okozott egyes, sajátos problémákat és technikai jellegű problémákat meg tudjuk oldani.

1. szabály. A permutálás /akár kézi, akár félig vagy teljesen automatizált/ szempontjából minden, két szóköz között elhelyezkedő fogalom vagy fogalomsor egy elemnek tekintendő.

Igy például a következő kifejezések:

hengerművi-kikéssítés

hengerművi-gyártmány-struktúra

a permutálás közben együtt, de a következők:

nehéz- közép- finom- hengersor

acél- rud-sajtolás

a permutálás közben külön-külön mozognak. Ez azt jelenti, hogy míg az elsőként említett kifejezések csak az első szó szerint kereshetők, a másodikként említettek minden - szóközzel megelőzött - szavuk szerint az indexben új sort indítanak. Ebből következik, hogy a

melegen-hengerelt-abroncsacél

kifejezés az m és az a betűvel kezdődő deskriptorláncok között található majd meg.

Ezt foglalja össze a

2. szabály. A magyarban oly gyakori összetett szavaknál és kifejezéseknél az összekapcsolást egybeírás vagy kötőjel segítségével kell megoldanunk, ha a hátrább lévő szóra vagy kifejezésre nem akarunk keresni, tehát nem akarjuk, hogy az új sort indítson. De kötőjelet és utána elhelyezett szóközt együttesen kell alkalmaznunk, ha a hátrább lévő összetevő/k/ is kereshető elem/ek/.

Igy például az acélrud sajtolás és az acélosó hengerlés gazdasági egybevetésével foglalkozó közleményt, ha csak az egybevetés előtt nem tartjuk külön kiemelendőnek, a következő deskriptorláncsal jellemezzük:

acél- rud- sajtolás- cső- hengerlés- gazdasági-egybevetése.

Az egy sorban elférő betűjelek korlátozott száma miatt fontos kérdés a hellyel való takarékoskodás. A fenti két szabály ezt is elősegíti. Így például, ha egy közlemény az acéllemezrel, acélszalaggal,

acélhuzallal, acélruddal, acélcsővel és műanyagcsővel foglalkozik, a következő deszkriptorláncot rendeljük hozzá:

acél- lemez szalag huzal cső műanyag- cső

Ugyancsak az összevonást célozza a

3. szabály. Ha egy deszkriptorláncban előfordul egy adott fogalom önálló elemként és egy másik, nem önálló /tehát nem sorkezdő/ elem jelzőjeként, akkor az adott fogalmat + jellel csatoljuk a nem önálló fogalomhoz és az egészet a permutálás szempontjából egy elemként kezeljük.

Példaképpen megemlítjük a következő esetet. Az adott dokumentum tárgyalja a rudsajtolás általános kérdéseit, de foglalkozik speciálisan a rudsajtolás szerszámaival is. Ha a szerszám fogalmat túl általános jellege miatt nem tekintjük önálló elemnek, akkor a 3. szabály értelmében a következő deszkriptort adjuk meg:

rudsajtolás+szerszám

A deszkriptorok megalkotásával kapcsolatban az a még nem teljesen lezárt nézetünk alakult ki /mely figyelembeveszi a feldolgozandó dokumentumok kis egységszámát és témaköri jellegét/, hogy mivel a szakemberek az egyes dokumentumokhoz hozzárendelték az azokra jellemző fogalmakat, a fenti szabályok /és egyéb, a gyakorlatban kialakult - a tolmáclékleválasztásra, szinonimákra stb. vonatkozó - szabályok/ szerint egységesített deszkriptorokat alakítsunk ki ezekből a fogalmakból az információtartalom gondos megtartása mellett.

Ezután következik a deszkriptorláncok kialakítása. Különböző szabályok bevezetése itt is célszerűnek látszott.

A fenti 40-50 deszkriptor átlag 4-5 sorban fér el, ezért a deszkriptorok halmazát szükségszerűen részhalmazokra kell bontanunk. Ezt szabályozza a

4. szabály. A deszkriptorláncokat úgy alakítjuk ki, hogy az egyes láncok csak egy adott általánosabb fogalomkörbe tartozó deszkriptorokat tartalmazzanak /ilyen fogalomkör lehet: gyártmány, alapanyag, technológia stb./ úgy, hogy ezeket az általánosabb fogalomköröket egy-egy dokumentum deszkriptorai határozzák meg.

Igy természetesen megtörténhet az, hogy egy-egy sort nem teljesen tölt ki a szöveg, ez azonban nem vezet a sorok számának gyarapodásához, mert a sorok száma a permutálás során önálló elemként kezelendő deszkriptorok láncba rendezésével nem változik. Erre a kérdésre még visszatérünk.

A deszkriptoroknak a szükségszerűen fellépő, de bármilyen szervezésű különálló láncokba való osztásával azonban az a hátrányos jelenség lép fel, hogy a dokumentumot leíró teljes deszkriptor-halmaz egyes részhalmazai a permutált láncok betürendbe rendezésekor egymástól elválhatnak, rendszerint már nem is tekinthetők át. Ezt a problémát

őhajtottuk megoldani az 5.^{ik} szabály alkalmazásával, melyet azonban később módosítanunk kellett /5. szabály/.

5.^{ik} szabály. Minden kialakítandó deszkriptorlánc azonos, a dokumentum tartalmára általánosan jellemző fogalommal kezdődjék.

Igy például az acélgéártás egyes vonatkozásait tárgyaló dokumentum deszkriptoraiként kialakított minden deszkriptorlánc a deszkriptorok között amugy is szereplő acélgéártás szóval kezdődjék, azaz az összes deszkriptorlánc az acélgéártás deszkriptornál összesítve megtalálható legyen. Ezzel, ahányszor "felelegetesen" használtuk az acélgéártás szót /azaz az összes sorok száma -1/, annyival megnövekszik a sorok száma.

Már a kevés számú kösleményt feldolgozó modell megalkotása közben is a fenti szabály alkalmazása két hibalehetőséget hozott magával.

Az első: előfordult az, hogy két vagy több dokumentum ugyanazzal az átfogó témakörrel foglalkozott, tehát a deszkriptorláncok azonosnak tartott fogalma így több dokumentumra azonos volt. Ezért a dokumentumok deszkriptorláncai a betűrendbe rendezéskor - melyet itt a második deszkriptor betűrendje determinált - összekeveredtek és így nem sikerült a szabály alkalmazásával a kívánt célt elérnünk.

A második: a hosszabb deszkriptor kiemelése például az acélgéártás és nyersvasgéártás fejlesztésével foglalkozó dokumentumnál az

acél- nyersvas-géártás-fejlesztése

formájában, ismétlődésével már sok helyet foglal el, ezenkívül a fenti példa a permutáció során két önálló elemet jelent, tehát a sorok számát $2/n-1$ -el növeli, ahol n az ismétlődés nélküli elemekből megalkotott permutált sorok száma.

Ezért volt szükség az 5.^{ik} szabály módosítására.

5. szabály. Minden dokumentum ötjegyű ugró sorszámot kap. Ez a sorszám az adott dokumentumra jellemző összes deszkriptorláncban permutálás közben is önálló elemként szerepel.

A fenti szabály alkalmazásával az eredeti célkitűzés megvalósul, a dokumentumra egyértelműen jellemző sorszám alapján a rendezéskor a dokumentum tartalmát leíró összes deszkriptor egy helyen is megtalálható lesz, a bármely permutált deszkriptorláncban is jelenlévő szám alapján /az index sorai ezenkívül természetesen tartalmazzák a külső rendszerben kialakított hivatkozási kódot is/.

A fenti ugró sorszám egyben /tekintettel a kis egységsszámra és szigorúan korlátozott témakörre/ a tárolás rendszerének alapját is képezhetné.

Végezetül közöljük a fenti módon kialakított modell egy sorát:

AB 12/1 acél- rudsaajtolás rudsaajtó+szerszámok 0009

SCHIFF, E.: Rules of the selection of descriptors for a model of permuted index

A model of a Hungarian-language permuted index was prepared by constructing chains of coordinated free descriptors. The specific problems arising in the course of the preparations may be divided into two groups:

1. problems pertaining to forming descriptors;
2. problems pertaining to forming descriptor-chains.

From the viewpoint of permutation, the basic element has been defined as a single concept consisting of one or more words, being not separated from each other by spacing but by special signs. This facilitated the compression of words, irrelevant by themselves into concepts, as well as the decomposition /by hyphen or other special symbol/ of a complex concept the components of which by themselves were, however, relevant.

Each word-chain, composed of free descriptors, represents a logical unit out of the content of the document. It seemed appropriate to organize the model in a form that all word-chains referring to a given document should appear in one group, in subsequent lines. This has been achieved by coordinating to each document, a five-digit number.



ИИФФ.3.: Правила составления дескрипторов модели пермутационного указателя

Создана модель пермутационного указателя на венгерском языке из цепей слов, образованных путем постановки самостоятельных дескрипторов друг за другом. Специфические проблемы, возникающие в ходе создания модели, разделены на две группы:

- проблемы в связи с образованием отдельных дескрипторов;
- проблемы в связи с образованием цепей дескрипторов.

Под одним элементом с точки зрения пермутации понимается понятие, состоящее из одного или больше слов между двумя пробелами, но друг от друга не разделенных пробелом. Тем самым открывалась возможность для слияния отдельно нерелевантных слов и для разделения и по своим компонентам релевантного слова (с применением дефиса или других специальных знаков).

Отдельная цепь слов - состоящая из самостоятельных дескрипторов - отражает часть содержания документа, объединяемую в логическую единицу. Оказалось целесообразным такое построение модели, в котором все цепи слов, относящиеся к документу, появляются в одном месте, в рядах друг за другом. Это было достигнуто путем присвоения пятизначного числа каждому документу, с использованием которых желаемое упорядочение автоматически осуществлялось.



SCHIFF, E.: Regeln der Deskriptorbildung für das Modell eines permutierten Indexes

Das Modell des ungarischen permutierten Indexes besteht aus Wortketten, welche durch Nebeneinanderstellung von freien Deskriptoren hergestellt wurden. Die sich im Laufe der Bearbeitung des Modells ergebenden spezifischen Probleme können in zwei Gruppen geteilt werden:

1. Probleme im Zusammenhange mit der Bildung der einzelnen Deskriptoren;
2. Probleme in Hinsicht auf die Bildung von Deskriptorenketten.

Vom Gesichtspunkt der Permutation wurden die sich zwischen zwei Wortabständen befindlichen, aus einem Wort oder aus mehreren - jedoch nicht durch Wortabstand, sondern durch besondere Zeichen getrennten - Wörtern bestehende Begriff als ein Element betrachtet. Das ermöglichte die Zusammenziehung der an und für sich irrelevanten Wörter, bzw. die Zergliederung der von komplexen Begriffen in ihre Komponenten aus relevanten Wörtern, mittels Anwendung von Bindestriichen oder anderen besonderen Zeichen.

Die einzelnen - aus freien Deskriptoren zusammengestellten - Wortketten widerspiegeln je eine logische Einheit aus dem Inhalt des Dokumentes. Es schien zweckmässig die sich auf ein Dokument beziehenden sämtlichen Wortketten an einer Stelle, in aufeinanderfolgenden Zeilen erscheinen zu lassen. Dies wurde dadurch erreicht, dass jedem Dokument eine fünfstellige Nummer zugeordnet wurde.

