

Koordinatno indeksiranje dokumenata

Dagica Terzić, dipl.inž.¹⁾

Dati su teorijski i praktični aspekti koordinatnog indeksiranja dokumenata koje se intenzivno koristi u automatizovanim sistemima za pretraživanje naučnotehnoških informacija. Ukratko je dat odnos deskriptorskog jezika za pretraživanje informacija, koji se koristi u procesu koordinatnog indeksiranja, i ostalih jezika za pretraživanje informacija. Navedene su definicije pojmova: ključna reč, deskriptor i identifikator, pomoćna sredstva koja se koriste u procesu koordinatnog indeksiranja, kao i uputstvo za koordinatno indeksiranje.

Ključne reči: Indeksiranje, koordinatno indeksiranje, klasifikacija, ključna reč, naučnotehnoške informacije, pretraživanje informacija.

Uvod

U sistemima za pretraživanje naučnotehnoških informacija (SPNTI) prema određenom predmetu, koriste se različiti jezici za pretraživanje informacija (JPI) [1]. Radi toga se, na osnovu teksta koji predstavlja kratku sadržinu dokumenta (naslovi, apstrakti, referati, anotacije itd.) ili celog teksta dokumenta, vrši indeksiranje dokumenta. Pod **indeksiranjem** dokumenata se podrazumeva izražavanje centralne teme ili predmeta iz teksta dokumenta ili opisivanje nekog fizičkog objekta pomoću JPI-a. Indeksiranje se može vršiti ručno ili automatski primenom računara. Indeksiranjem dokumenta formira se **karakteristika traženog dokumenta** [2]. Pretraživanje informacija, odnosno dokumenata, vrši se poređenjem karakteristika traženih dokumenata i instrukcije za pretraživanje. **Instrukcija za pretraživanje** predstavlja tekst koji, na korišćenom JPI, izražava smislaoni sadržaj informacionog zahteva i koji je specijalno namenjen za pronalaženje (poređenjem tog teksta sa karakteristikama traženih dokumenata) u nekom skupu onih dokumenata, koji odgovaraju datom zahtevu. Instrukcija za pretraživanje sadrži i logičke operacije koje treba izvršiti tokom pretraživanja informacija.

U savremenim automatizovanim SPNTI za indeksiranje dokumenata najviše se koriste deskriptorski JPI, pomoću kojih se vrši koordinatno indeksiranje dokumenata (u daljem tekstu: koordinatno indeksiranje) upotrebom ključnih reči, deskriptora i identifikatora.

Teorijski aspekti koordinatnog indeksiranja

Da bi se lakše shvatio princip koordinatnog indeksiranja, ukatko je dat odnos deskriptorskog JPI, koji se koristi za koordinatno indeksiranje, i ostalih JPI.

JPI je specijalizovani veštački jezik namenjen za opisivanje centralnih tema ili predmeta i formalnih karakteristika dokumenata radi kasnijeg pronalaženja potrebnih dokumenata u skupu drugih, i/ili za izražavanje sadržine informacionih zahteva i pronalaženja potrebnih dokumenata [1].

Svaki JPI predstavlja sistem klasifikacije pošto svaka reč, ili kombinacija reči kojom se označava neki pojam, predstavlja naziv neke klase. Pod klasom se podrazumeva skup predmeta od kojih svaki ima zajedničke karakteristike prisutne u odgovarajućim pojmovima. Klase mogu biti proste i složene.

Svi članovi proste klase imaju samo jednu opštu karakteristiku označenu nazivom te klase. Naziv proste klase se označava prostom rečju ili postojanom kombinacijom nekoliko reči.

Članovi složene klase imaju, pored jedne opšte karakteristike, i neku kombinaciju opštih karakteristika. Kao nazivi složenih klasa služe različite kombinacije reči ili čak cele fraze. Složene klase je moguće raščlaniti na proste, a da se pri tom ne izgubi mogućnost njihovog jednoznačnog obnavljanja putem logičkog množenja dobijenih prostih klasa.

Složena klasa se može obrazovati iz prostih klasa njihovim logičkim množenjem (presecanja, koordinacije) pre indeksiranja dokumenata, pri indeksiranju dokumenata ili u momentu njihovog pronalaženja, što zavisi od vrste primenjenog JPI.

Ako JPI ima strukturu i pravila na osnovu kojih se indeksiranje dokumenata vrši pomoću prethodno obrazovanih složenih klasa, a nazivi složenih klasa postaju karakteristike traženih dokumenata, takav JPI se naziva **prekoordinatni**. Ako se za indeksiranje dokumenata koristi JPI pomoću koje se složene klase obrazuju iz prostih u početku pri indeksiranju dokumenata (potencijalno), a zatim stvarno pri njihovom pronalaženju, takav JPI se naziva **postkoordinatni**.

Kod prekoordinatnih JPI sve reči u nazivima složenih klasa su međusobno logički povezane. Zbog toga, za pronalaženje informacija, odnosno dokumenata indeksiranih pomoću prekoordinatnih JPI, u instrukciji za pretraživanje treba koristiti iste one kombinacije reči (ako se radi o nazivima složenih klasa) koje je koristio stručnjak za indeksiranje (indekser) pri indeksiranju. U procesu pronalaženja dokumenata ne dozvoljava se nikakva zamena ili izmena reči i numeričkih oznaka korišćenih pri indeksiranju.

¹⁾ Vojnotehnički institut VJ, 11000 Beograd, Katanićeva 15

Za potrebe indeksiranja postoje mnogobrojni prekoordinatni i postkoordinatni JPI [1,3], među kojima su najvažniji:

Prekoordinatni JPI:

- Đuijeva decimalna klasifikacija,
- Blisova bibliografska klasifikacija,
- Klasifikacija Kongresne biblioteke SAD,
- Univerzalna decimalna klasifikacija,
- Međunarodna klasifikacija patenata,
- COSATI (Committee on Scientific and Technical Information – SAD) klasifikacija,
- hibridne klasifikacije
- alfabetsko-predmetne klasifikacije i
- fasetne klasifikacije.

Postkoordinatni JPI:

- deskriptorski jezici
- semantički kodovi i
- sintagmatski jezici

Osnovni i najveći nedostatak prekoordinatnih JPI je u tome što nazivi složenih klasa moraju biti unapred dati i samo takvi mogu biti korišćeni pri indeksiranju i pri pretraživanju informacija (pronalaženju dokumenata). Time se znatno povećava broj leksičkih jedinica takvih JPI, a samim tim je teže sastavljanje i primena ovih JPI. Složenost sastavljanja kompatibilnih prekoordinatnih JPI je ilustrovana na primeru razrade klasifikacije dobara i usluga u SRJ [4].

Od postkoordinatnih JPI najveću primenu imaju deskriptorski JPI. U osnovu deskriptorskih JPI postavljen je princip **koordinatnog indeksiranja**, koje se sastoji u tome da se centralna tema dokumenata, ili tema informacionog zahteva, izražava u obliku skupa reči i kombinacija reči prirodnog jezika koje, obično, predstavljaju nazive prostih klasa [1,5].

Kod deskriptorskih JPI reči i grupe reči (nazivi klasa) u karakteristikama traženih dokumenata i instrukcijama za pretraživanje funkcionišu nezavisno jedna od druge, tačnije rečeno, nalaze se među sobom u odnosu koordinatne veze [1].

Koordinacija (logičko množenje) naziva prostih klasa daje široke mogućnosti obrazovanja naziva složenih klasa iz naziva prostih klasa.

Zahvaljujući svojim širokim mogućnostima obrazovanja naziva složenih klasa iz naziva prostih klasa, deskriptorski JPI bilo iz koje oblasti mogu imati relativno mali broj leksičkih jedinica. Na primer, klase 5 i 6 (matematika, prirodne nauke, primenjene nauke, medicina, tehnika) u Univerzalnoj decimalnoj klasifikaciji imaju ≈ 100000 podela (tj. leksičkih jedinica JPI); Alfabetsko-predmetna klasifikacija, ekvivalentna ovim klasama, ima ≈ 70000 odrednica i pododrednica, a deskriptorski JPI svega 5000-7000 leksičkih jedinica (reči i grupa reči).

U osnovni rečnički sastav deskriptorskog JPI treba uključiti samo takve reči i grupe reči koje ne mogu jednoznačno da se obrazuju iz već postojećih njihovim logičkim množenjem, čime se znatno smanjuje broj leksičkih jedinica deskriptorskog JPI.

Za koordinatno indeksiranje dokumenata ili informacionih zahteva (tj. u svojstvu leksičkih jedinica deskriptorskog JPI) mogu se koristiti **ključne reči** (reči ili grupe reči odabrane neposredno iz tekstova koji se indeksiraju), **deskriptori** (ključne reči kod kojih su odstranjene sinonimija, homonimija i polisemija) i **identifikatori**.

Ključne reči, deskriptori i identifikatori

Pri koordinatnom indeksiranju predmet dokumenta (ili informacionog zahteva) se izražava pomoću skupa tzv. **ključnih reči** iz dokumenta (odnosno iz informacionog zahteva) [1,5]. **Ključna reč** i **termin** se, uslovno, smatraju sinonimima.

Ključne reči su najvažnije reči (ili grupa reči) koje imaju nominalnu (nazivnu) funkciju. To su nazivi predmeta, pojava, procesa, osobina, odnosa. Ključne reči se ne odnose na konkretne predmete već na pojmove koji su izraženi tim rečima. Izuzetak su one ključne reči koje se odnose na neke određene predmete, kao što su nazivi institucija, projekata, geografski nazivi, brojevi patentnih specifikacija i nacionalnih standarda i sl. Takve ključne reči se nazivaju **identifikatori**. Identifikatori su mnogobrojni nazivi koji neprekidno nastaju i iščezavaju kao rezultat ljudske delatnosti i koji, zbog svoje raznovrsnosti i velikog broja, ne mogu biti ranije dati nabranjem. U karakteristikama traženih dokumenata i instrukcijama za pretraživanje identifikatori funkcionišu kao elementi pretraživanja, ali ne ulaze u osnovni rečnički sastav deskriptorskog jezika.

Odnosi između ključnih reči

Detaljnije informacije u vezi s različitim odnosima i karakteristikama ključnih reči su date u [1,5]. Ovom prilikom će biti ukratko dati najvažniji odnosi i karakteristike ključnih reči, koji su značajni za formiranje kontrolisanih rečnika ključnih reči, deskriptorskih rečnika i tezaurusa koji se koriste u procesu koordinatnog indeksiranja dokumenata.

Jedan od najvažnijih uslova za povećanje efikasnosti pretraživanja informacija u koordinatno indeksiranim dokumentima pomoću deskriptorskih JPI jeste određivanje semantičkih odnosa između ključnih reči, odnosno obrazovanje deskriptora [1,5]. Semantički odnosi su odnosi između reči (naziva) i predmeta (pojava, procesa, osobina, odnosa) koje te reči označavaju, odnosno pojmova koji su izraženi tim rečima. Otkrivanjem i fiksiranjem semantičkih odnosa otklanja se sinonimija, homonimija i polisemija u deskriptorskim JPI. Na taj način se obezbeđuje korišćenje istih reči ili grupa reči za indeksiranje dokumenata i informacionih zahteva sa istim predmetom (sadržinom), što direktno utiče na efikasnost pretraživanja informacija.

Radi otklanjanja sinonimije u spisku ključnih reči, odabranih za dati SPNTI, vrši se grupisanje sinonima u klase (semantičke nizove). Iz svake klase bira se jedna ključna reč – semantička dominantna, kojom se ta klasa označava, tj. koja reprezentuje tu klasu. U odnosu na dominantu, sve ostale ključne reči iz neke klase označavaju se kao sinonimi. Svaka reč iz klase mora da bude sinonim ne samo dominante, već i svih ostalih reči te klase.

Semantički nizovi u kojima su grupisane ključne reči – sinonimi nazivaju se **klase ekvivalentnosti** [1,5].

Pri obrazovanju klasa ekvivalentnosti u jezicima za koordinatno indeksiranje ne vrši se grupisanje samo pravih sinonima (kojih u prirodnim jezicima, a naročito u naučnoj terminologiji nema mnogo) već i ključnih reči koje izražavaju pojmove bliske po sadržaju (intenziji). Pri tome se obimi (ekstenzije) pojmova ili preklapaju (deo jednog – pojma ulazi u obim drugog pojma i obrnuto) ili podređuju (obim jednog pojma čini deo obima drugog pojma). To su tzv. uslovni sinonimi (kvazisinonimi) – reči koje se usvajaju kao sinonimi za ciljeve određenog SPNTI-a, iako se one po značenju u stručnoj komunikaciji razlikuju.

Grupisanjem sinonima i kvazisinonima u klase dobijaju se **klase uslovne ekvivalentnosti** čija se semantička dominanta naziva **deskriptor** [1,5]. Deskriptor može da se definiše kao standardizovana ili formalizovana ključna reč (termin) koja se koristi za jednoznačnu reprezentaciju sadržine dokumenata i informacionih zahteva [1,2,5]. Ovo podrazumeva da se, osim uklanjanja sinonimije, mora odgovarajućim oznakama uz određene ključne reči odstraniti njihova homonimija i polisemija.

Od ostalih odnosa između ključnih reči najvažniji su paradigmatički odnosi koji obuhvataju odnose rod-vrsta i povezane odnose [1,5]. Utvrđivanje ovih odnosa između ključnih reči je moguće samo uz izradu tezaurusa ključnih reči (termina), čijom se primenom u procesu koordinatnog indeksiranja i pretraživanja informacija znatno poboljšava efikasnost pretraživanja informacija.

Pored utvrđivanja pojedinih odnosa između ključnih reči, za proces koordinatnog indeksiranja potrebno je pridržavati se određenih pravila koja se odnose na primenu jednine i množine ključnih reči, znakova interpunkcije, skraćenica i akronima [6].

Pošto se među ključnim rečima sreću sinonimi, homonimi a moguće je i različito pisanje jedne iste ključne reči, direktno poređenje karakteristika traženih dokumenata i instrukcija za pretraživanje postaje teško, a pronalaženje informacija manje efikasno.

Da bi se savladala ta teškoća, u deskriptorskim JPI se primenjuje leksikografska kontrola korišćenih ključnih reči [1,5], koja se sastoji u:

- dovođenju svih ključnih reči, korišćenih za indeksiranje, na ujednačen oblik (isto pisanje, jedinstveni morfološki oblik);
- potpunom odstranjivanju sinonimije, homonimije i polisemije ključnih reči.

Normativni rečnik u kojem su alfabetskim redosledom date sve ključne reči i deskriptori, u kojem su ključne reči i deskriptori snabdeveni uputnim, ograničavajućim i objašnjavajućim primedbama pomoću kojih se odstranjuje njihova sinonimija, polisimija i homonimija, naziva se **deskriptorski rečnik**.

Pomoćna sredstva za koordinatno indeksiranje

Koordinatno indeksiranje se može obavljati ručno i automatski pomoću računara. Ručno indeksiranje obavljaju indeksari koji su stručnjaci za određene predmetne oblasti. Pri ručnom indeksiranju koriste se različita pomoćna sredstva: rečnici ključnih reči – termina, permutacioni indeksi ključnih reči, rečnici identifikatora, permutacioni indeksi identifikatora, deskriptorski rečnici i tezaursi.

Indeksni termini (ključne reči koje se koriste za indeksiranje) mogu biti kontrolisani i nekontrolisani. Većina stručnjaka smatra da nekontrolisani rečnici za indeksiranje, koji mogu sadržavati svu raznolikost prirodnog jezika, doprinose unošenju mnogih dvosmislenosti i grešaka [7]. Zbog toga se, najčešće, koristi ograničeni JPI kod kojeg se indeksni termini strogo kontrolišu.

Postoje različiti nivoi kontrole indeksnih termina u kontrolisanim rečnicima:

- kontrolisani rečnici abecedno sređenih ključnih reči kod kojih je izvršena delimična leksikografska kontrola (isto pisanje, isti morfološki oblik) ključnih reči koje, u ovom rečniku, istovremeno predstavljaju indeksne termine. Ovaj rečnik može biti organizovan i kao permutacioni indeks ključnih reči;

- deskriptorski rečnik – rečnik ključnih reči u kojem su, osim dovođenja svih ključnih reči, korišćenih za indeksiranje, na jedinstveni oblik, odstranjeni sinonimija, homonimija i polisemija ključnih reči. U deskriptorskom rečniku sve ključne reči nisu indeksni termini, odnosno indeksni termini su samo deskriptori. Deskriptorski rečnik predstavlja prvu fazu u struktuiranju tezaurusa [1];

- tezaursus – rečnik ključnih reči u kojem su odstranjeni sinonimija, polisemija i homonimija ključnih reči i u kojem su dati odgovarajući uputi od užih indeksnih termina (deskriptora) na šire indeksne termine, paradigmatički odnosi (odnosi rod – vrsta i povezani odnosi) između indeksnih termina, objašnjenja obuhvatnosti indeksnih termina, načini primene indeksnih termina za indeksiranje složenih pojmova i dr.

Navedena pomoćna sredstva mogu biti unapred sastavljena za potrebe odgovarajućih SPNTI, a rečnik ključnih reči i deskriptorski rečnik se mogu sastavljati i u procesu formiranja i ažuriranja baza podataka u okviru odgovarajućeg SPNTI.

Najveća efikasnost u procesu pretraživanja informacija koordinatno indeksiranih dokumenata se postiže ako postoje odgovarajući tezaursi. Ograničavajuća okolnost, pritom, je to što je za sastavljanje tezaurusa potreban ogroman intelektualni rad, a samim tim, i troškovi su veliki. Tako, na primer, u sastavljanju tezaurusa [6] učestvovalo je više od 300 naučnika, stručnjaka, bibliotekara itd.

S obzirom na ovo, mnogi SPNTI se zasnivaju na primeni, za potrebe indeksiranja, kontrolisanih rečnika ključnih reči i deskriptorskih rečnika.

Uputstvo za koordinatno indeksiranje

Na efikasnost SPNTI, odnosno potpunost i preciznost pretraživanja informacija, znatno utiče indeksiranje dokumenata [1,9-11] i to iscrpnost, tj. dubina indeksiranja i specifičnost JPI-a, odnosno indeksnih termina, kao i individualni faktori indeksera. S obzirom na ovo, kao i na postojeća uputstva za indeksiranje [12,13] u daljem tekstu su data neka uputstva i informacije kojih bi trebalo da se pridržavaju indeksari pri ručnom indeksiranju dokumenata radi postizanja što veće efikasnosti SPNTI.

Indeksiranje dokumenata, uz primenu već pomenutih pomoćnih sredstava, vrše indeksari koji su stručnjaci za pojedine predmetne oblasti. Raspodela, odnosno klasifikacija dokumenata po predmetnim oblastima se vrši prema nekoj od grubih prekoordiniranih klasifikacija, kao što su Univerzalna decimalna klasifikacija, COSATI klasifikacija itd.

Analiza sadržine dokumenata

Pošto je vrlo teško definisati šta je to što predstavlja sadržinu dokumenta, u definiciji indeksiranja [1] se daje **centralna tema** ili **predmet dokumenta** pod kojima se podrazumeva skup uzajamno povezanih osnovnih pojmova teksta dokumenta [14]. Radi toga se analiza sadržine dokumenta ponekad naziva i pojmovna ili konceptualna analiza.

Određivanje sadržine dokumenta se vrši u tri faze:

- razumevanje ukupne sadržine dokumenta, namere autora itd.,
- identifikovanje pojmova koji predstavljaju ovu sadržinu i
- izbor osnovnih pojmova koji će se indeksirati.

Zbog ekonomskih, ali i praktičnih razloga, ne preporučuje se čitanje celog teksta dokumenta: Bez obzira na ovo,

indekser pri indeksiranju ne sme prevideti važne informacije iz dokumenta [1,5]. Pri kratkoj analizi sadržine dokumenta naročitu pažnju treba da posveti:

- naslovu,
- apstraktu, uvodu i uvodnim rečenicama poglavlja i odeljaka,
- ilustracijama, tabelama, dijagramima i njihovim nazivima,
- zaključku i
- rečima ili grupama reči koje su podvučene ili štampane neuobičajenom štamptom, kao i akronimima.

Namere autora su, obično, navedene u uvodnim poglavljima, dok se u završnim poglavljima najčešće konstatuje da koje su mere ove namere postignute.

Sve ove elemente indekser mora uzeti u obzir prilikom analize sadržine dokumenta.

Indeksiranje direktno iz naslova se ne preporučuje. Isto tako i apstrakt ne može poslužiti kao zadovoljavajuća zamena za čitanje teksta dokumenta. Naslovi mogu biti varljivi, naslovi i apstrakti mogu biti neadekvatni i, u većini slučajeva, ni jedni ni drugi ne mogu biti pouzdan izvor informacija potrebnih indekseru za indeksiranje.

Posle pregledanja dokumenta indekser bira one pojmove koji najbolje odražavaju sadržinu tog dokumenta [1,5].

Od pojmova, identifikovanih za vreme pregledanja dokumenta, za indeksiranje se biraju samo oni koji predstavljaju centralnu temu ili predmet dokumenta.

Indeksiranje izabranih pojmova iz dokumenta

Izabrani pojmovi se indeksiraju, odnosno njihovi nazivi se pomoću odgovarajućeg pomoćnog sredstva (rečnik ključnih reči, deskriptorski rečnik itd.) prevode u odgovarajuće ključne reči ili deskriptore i identifikatore.

Specifičnost ključnih reči (indeksnih termina)

Pojmove koji predstavljaju osnovnu sadržinu ili predmet dokumenta treba identifikovati što određenije, odnosno s obzirom na odnos rod-vrsta odabrati najspecifičnije pojmove za indeksiranje i prevesti u odgovarajuće ključne reči.

Išcrpnost indeksiranja

Broj ključnih reči, koje se pri indeksiranju mogu dodeliti jednom dokumentu, teorijski nije ograničen. Eksperimentalno je utvrđeno [15] da se povećanjem broja ključnih reči (indeksnih termina) sa 3 na 8 efikasnost (koeficijent potpunosti - %) pretraživanja informacija povećava sa 65% na 81%, dok se pri povećanju broja indeksnih termina sa 8 na 12 efikasnost neznatno povećava sa 81% na 83%. Znači da primena velikog broja indeksnih termina (iznad određene granice) ne poboljšava znatno efikasnost pretraživanja informacija i može dovesti do lažne koordinacije indeksnih termina pa se preporučuje da se za indeksiranje ne koristi više od 10 ključnih reči.

Indeksiranje složenih pojmova

Za indeksiranje složenih pojmova treba upotrebljavati dve ili više ključnih reči, što je glavna prednost koordinatnog indeksiranja.

Na primer, složeni pojam *oklopno točkaško borbena vozilo* se indeksira pomoću ključnih reči: *oklopno vozilo, točkaško vozilo, borbena vozilo*.

Doslednost pri indeksiranju

Idealni JPI ne postoje, odnosno ne postoje JPI koji bi omogućili savršen opis predmeta dokumenta, a samim tim i obrazovanje savršene karakteristike traženog dokumenta. Idealni slučaj je nemoguć zato što je ponašanje pri indeksiranju i pretraživanju delimično neodređeno i ima probablistički karakter. Praktično je nemoguće definisati šta je to što je predmet dokumenta, odnosno nemoguće je definisati šta je to što treba tražiti u jednom dokumentu da bi se identifikovala njegova sadržina [16]. Zbog toga je nemoguće obučiti indeksera kako da pronađu predmet nekog dokumenta pa se, u definiciji za indeksiranje, umesto predmet dokumenta koristi pojam centralna tema ili predmet dokumenta.

Baš zbog toga što se ne može dati sasvim jasna definicija predmeta dokumenta, i pored činjenice da je indeksiranje dokumenta obavljeno tačno i sa dovoljnom podrobnosti, u stvarnosti je nemoguće dosledno indeksirati dokument, jer u procesu indeksiranja obično učestvuju više indeksera. Čak i jedan isti indekser u različitim vremenskim periodima različito će indeksirati isti dokument.

Doslednost pri indeksiranju označava meru do koje jedan isti indekser u različitim vremenskim periodima dodeljuje iste indeksne termine određenom dokumentu, ili meru do koje različiti indekseri određenom dokumentu dodeljuju iste indeksne termine. Prema [16] doslednost pri indeksiranju, kada indeksiranje vrši jedan indekser, iznosi 50%, a kada indeksiranje vrše dva ili više indeksera 20%.

Jasno je da doslednost pri indeksiranju utiče na efikasnost SPNTI, tim pre što će i informacioni zahtevi, najverovatnije, sadržavati indeksne termine različite od onih koje je dokument dobio pri indeksiranju, za istu sadržinu dokumenta [17].

Nepriistrasnost pri indeksiranju

Pri analizi sadržine dokumenta radi indeksiranja ne sme se unositi nikakvo lično viđenje ili ocena sadržine dokumenta, odnosno sadržinu dokumenta se mora oceniti nepriistrasno.

Istinitost pri indeksiranju

Ključne reči, odabrane za indeksiranje jednog dokumenta, moraju istinito odražavati sadržinu dokumenta, jer će korisnik tražiti dokument koji je indeksiran određenom ključnom rečju, pretpostavljajući da taj dokument sadrži znatan deo informacija koje njega interesuju. Drugim rečima, upotreba npr. ključne reči *elektronski rat* za indeksiranje implicitno podrazumeva da će se u dokumentu naći znatan deo informacija o elektronskom ratu, odnosno da elektronski rat u dokumentu nije samo uzgred pomenut.

Redosled navođenja ključnih reči

Ključne reči se navode tako da se među prvim nađu one koje predstavljaju centralnu temu ili predmet dokumenta, a zatim ostale, koje imaju neku vezu sa centralnom temom ili predmetom dokumenta.

Ukoliko se u dokumentu radi o primeni nečega (sredstva, uređaja, procesa, metode, materijala i sl.) u nekoj predmetnoj oblasti, među prve ključne reči se stavljaju one koje se odnose na predmetnu oblast, a zatim one koje se odnose na sredstva, uređaje, procese, metode, materijale i sl. koji se primenjuju u određenoj predmetnoj oblasti.

Indeksiranje pomoću identifikatora

Kao što je već rečeno, identifikatori su specifične ključne reči koje se odnose na neke individualne predmete kao što su: nazivi institucija, projekata, geografski nazivi, brojevi patentnih specifikacija i nacionalnih standarda i sl. Primeri identifikatora su: *eksploziv LX-IV*, *haubica FX-77*, *raketa ROLAND* itd. Identifikatori se sastoje, najčešće, iz dela koji se odnosi na ključnu reč i dela koji se odnosi na identifikacioni faktor. Na primer: identifikator *tenk Leopard-2* se sastoji iz ključne reči *tenk* i identifikacionog faktora *Leopard-2*.

Za indeksiranje dokumenata pomoću identifikatora važi isto pravilo kao i za indeksiranje dokumenata pomoću ključnih reči, odnosno **pomoću identifikatora se indeksiraju samo oni pojmovi koji predstavljaju centralnu temu ili predmet dokumenta**, a ne i oni pojmovi koji su samo pomenuti u dokumentu.

Zaključak

Na osnovu izloženog može se konstatovati da je osnovna prednost postkoordinatnih JPI, u odnosu na prekoordinatne JPI, široka mogućnost obrazovanja naziva složenih klasa iz naziva prostih klasa čime se znatno smanjuje broj leksičkih jedinica (ključnih reči – deskriptora) potrebnih za indeksiranje i pretraživanje informacija, odnosno dokumenata. Primenom preporučeneog uputstva za koordinatno indeksiranje dokumenata povećava se efikasnost SPNTI. S obzirom na veliki intelektualni rad i sredstva, koja je potrebno utrošiti za izradu odgovarajućih tezaurusa kao pomoćnih sredstava za koordinatno indeksiranje, preporučuje se upotreba permutacionog indeksa ključnih reči, odnosno identifikatora.

Literatura

[1] ČERNYJ, A.I. *Vvedenie v teoriju informacionnogo poiska*. Nauka, Moskva, 1975.

- [2] BELOZEROV, V.N., ANTOŠKOVA, O.A., GRIŠINA, F.E. *Novye standarty na terminologiju informacionnogo poiska*. *Naučno-tehničeskaja informacija*, 1997, ser.1, no.1, p.14-21.
- [3] BUKUMIROVIĆ, S. *Uvod u informatiku*. Narodna biblioteka Srbije, Beograd, 1995.
- [4] BUKUMIROVIĆ, S. *Harmonization of Classifications, Nomenclatures and Identifications for Goods and Services, and Activities to Solve these Problems in Yugoslavia*. International Forum on Information and Documentation, 1991, vol.16, no.1, p.18-20.
- [5] MILJKOVIĆ, M. *Metodologija razrade tezaurusa*. magistarski rad. Sveučilište u Zagrebu, Zagreb, 1976.
- [6] *...Thesaurus of Engineering and Scientific Terms*. Engineers Joint Council, New York, 1967.
- [7] SALTON, G., MCGILL, M.J. *Introduction to Modern Information Retrieval*. McGraw-Hill Book Company, New York, 1983.
- [8] *Uputstvo za sastavljanje i razvoj jednojezičkih tezaurusa*. JUS standard, 1994.
- [9] LANCASTER, F.W. *Information Retrieval Systems. Characteristics, Testing and Evaluation*. John Wiley, New York, 1968.
- [10] MIHAJLOV, A.I., GILJAREVSKIJ, R.S. *Naučna informacija – izvori, pronalaženje, korišćenje* (prevod sa ruskog). Jugoslovenski centar za tehničku i naučnu dokumentaciju, Beograd, 1976.
- [11] TERZIĆ, D. *Efikasnost dokumentacionih sistema za pronalaženje informacija i faktori koji na nju utiču*. *Naučno-tehnički pregled*, 1978, vol. XXVIII, no.5, p.35-42.
- [12] *...Manuel de l'indexation*. CEDOCAR, Paris, 1970.
- [13] *...Principles of Indexing*. UNESCO/UNISTST, SC.75/WS/58, Paris, sept.1975.
- [14] KONYŠČEV, V.N. *Psiholingvističeskoe issledovanie indeksirovanija kak intelektual'nogo processa*. *Voprosy informacionnoj teorij i praktiki*, 1983, no.50, p.78-85.
- [15] VICKERY, R. *Vocabularies for Co-ordinate Systems*. *Aslib Proceedings*, 1963, vol.15, no.6, p.170-176.
- [16] BATES, M.J. *Subject Access in Online Catalogs: a Design Model*. *Journal of the American Society for Information Science*, 1986, vol.37, no.6, p.357-376.
- [17] SALTON, G. *Dynamic Information and Library processing*. Prentice-Hall Inc., New Jersey, 1975.

Rad primljen: 19.10.1998.god.