

ISTRAŽIVANJE INTELIGENTNIH SISTEMA ZA PRETRAČIVANJE INFORMACIJA

1 UVOD

Poslednjih decenija informacioni sistemi (IS) beleže brzi razvoj u kojem se, pored drugih značajnih karakteristika, nameće i konstatacija da je, uprkos sve većoj lakoći formiranja informacionih kolekcija i njihovog čuvanja, čoveku potreban i sve veći napor da dođe do relevantne informacije. Ovo važi naročito za tekstualne baze podataka (BP), čiji pokazatelji korišćenja, a pre svega kvantitativni pokazatelji performansi u pretraživanju informacija (PI), imaju nezadovoljavajuće niske vrednosti. Jedan od pokazatelja jeste i činjenica što je za postizanje boljih rezultata u PI, neophodno da korisnik poseduje i razna znanja. Od korisnika se zahteva veliki intelektualni napor jer, ne samo što su mu u procesu PI potrebna znanja, na primer, o predmetnoj oblasti, o sistemu za pretraživanje informacija (SPI), o predmetnoj klasifikaciji u BP koje pretražuje, već i zato što on treba da ih primeni u dinamičnoj i interaktivnoj komunikaciji sa SPI (Fidel, 1986, 1991; Ofori-Dnjumfuo, 1984; Sparck Jones, 1991; Chen, Dhar, 1991).

Proučavanje SPI i traganje za rešenjima problema koji utiču na PI traju koliko i njihov razvoj te se može uočiti oviše generacija tehnika za rešavanja tih problema. Prva generacija tehnika za poboljšavanje performansi PI, tehnike PI koje su se tokom sedamdesetih razvijale na bazi tehnika teorije verovatnoće (Bookstein, Snjanson, 1975), donose prednosti, ali i svoje značajne slabosti. Iako ne smenjuju tradicionalne SPI, prisutne su i danas, zajedno sa sledećom generacijom tehnika, sa tehnikama veštačke inteligencije (VI) koje rešenja traže u znanju potrebnom za PI i njegovom ugrađivanju u SPI. Prvi talas primene tehnika VI u PI, već oko sredine osamdesetih (Shoval, 1985), ugrađuje znanja čoveka-eksperta koji rešava probleme u PI, ali i danas ima prepreka u prikupljanju tog znanja. Najnovija istraživanja, takođe u krilu VI, nastoje da sistemi mašinskog učenja potrebno znanje stiču učenjem iz primera, automatski, iz raspoloživih izvora znanja. Ono što je uočljivo to je da komercijalni sistemi SPI koriste tekovine istraživanja uglavnom prihvatajući novu tehnologiju i nove pristupe PI, uz kontinualna poboljšavanja tradicionalnog pristupa, a ne revolucionarnom smenom paradigme. U istraživanjima i razvoju SPI danas, prisutne su i primenjuju se sve tehnike poboljšavanja.

U ovom radu se daje pregled istraživanja tehnika koje poboljšanja SPI baziraju na œprimeni inteligencije, od početka 80-tih, kada su se znanja za rešavanje određenih problema ugrađivala u komponente ekspertnih sistema (ES). Prvi nazivi tehnika i realizovanih sistema sadržali su češće sintagmu œekspertni sistemi za PI (Vickery, Brooks, 1987). U istraživanjima se tokom vremena dolazi do saznanja da je ekspertiza u PI prilično neuhvatljiva, pre svega što problem koji ES treba da reši nije jasno omeđen. Domen znanja koji treba da se razvije za PI prilično je razuđen, što ne odgovara jednom od osnovnih zahteva uspešnog ES. Moguće je da je sve ređa pojava termina ES u blizini termina PI svojevrsno korigovanje istraživača. Međutim, istraživanja ovih, uglavnom prototipova i sistema u razvoju, nastavljaju se nesmanjenim intenzitetom i, kako pokazuju izvori informacija, pa i oni na Internetu, pored naziva œES za PICE i naziva œinteligentni informacioni sistemi, česti su i nazivi œsistemi sa bazama znanja (knowledge-based systems) (Monarch, 1987; Smith, 1989; Sormunen, 1989). U vezi sa više domena znanja kojima, radi primene u PI, mora da se bavi istraživanje i razvoj ovih sistema, tokom vremena se uvidelo da se u rešavanju problema u PI pomoću baza znanja čovek - korisnik ili posrednik - ne može potpuno izostaviti. Stoga se razvijani sistemi sa epitetom œinteligentni, često nazivaju œkonsultantima (npr. za PI, za indeksiranje), ili œposrednicima (npr., za pretraživanje) (Goldsmith, 1985; Fidel, 1995; Humphrey, 1989; Sormunen, 1989), ali i œinteligentnim interfejsima (Vickery, 1989). Uz osvrt na PI i probleme u PI koji su inspirisali istraživanja ovih sistema, u odeljcima koji slede izdvojena su njihova važnija obeležja i rezultati istraživanja koja su se, radi iznalaženja efikasnijeg načina predmetnog pristupa tekstualnim informacijama, bavila integrisanja SPI i sistema sa bazama znanja (SBZ).

1.1 PROBLEMI U PRETRAŽIVANJU INFORMACIJA I NJEHOVA SEMANTIČKA PRIRODA

U procesu PI, korisnik svoju informacionu potrebu iskazuje upitom - skupom termina koji predstavlja semantiku te potrebe. Formulisući ga pomoću termina i na način koji određuje korišćeni SPI, korisnik traži informacije koje ispunjavaju određene uslove, npr. koje se odnose na pojmove ključne za njegov informacioni problem. U procesu PI po predmetu, u SPI koji su danas preovlađujući, informacije se pronalaze u BP na osnovu opisa dokumenata koji su formirani pri ažuriranju BP, te PI zavisi i od kvaliteta tog opisa. Predmetni opis dokumenta, odnosno njegova karakteristika za PI, dobijen je dodeljivanjem termina iz teksta dokumenta ili dodeljivanjem termina iz kontrolisanog rečnika odrednica u procesu predmetnog označavanja (indeksiranja). Danas se indeksiranje izvodi manuelno, uz pomoć računara ili potpuno automatski (Milstead, 1990). U sva tri oblika prepoznaju se slični mehanizmi i instrumenti, na primer, pravila za indeksiranje i liste odrednica za dodelu dokumentima. Dok je u manuelnom indeksiranju problem nekonzistentnosti gotovo nemoguće izbeći, on se ipak ublažava primenom kontrolisanog rečnika i postkoordinatnog indeksiranja. Automatsko indeksiranje se bazira na statističkoj analizi reči i sintagmi korišćenih u tekstovima dokumenata, preko izračunavanja frekvencija pojavljivanja svake pojedine reči u svakom dokumentu i ukupnih frekvencija svake reči u svim dokumentima u BP. Za dodeljivanje predmetnih oznaka dokumentima zadržavaju se reči sa srednjom frekvencijom, a u praksi se primenjuje niz modifikacija ove metode.

U novije vreme se za poboljšanje efikasnosti indeksiranja kombinuju manuelne i automatske metode i istražuju primene baza znanja u nastojanju da se ostvari interaktivno i poimovno indeksiranje (Milstead, 1990, Salton, 1983, Humphrey, 1989; Sparck Jones, 1991).

PI po predmetu se izvodi poređenjem instrukcije za PI - reprezentacije uslova iz korisnikovog zahteva za informacijama - i reprezentacije karakteristika dokumenata koje se za taj upit pretražuju. Kriterijum upoređivanja jeste sličnost između tih reprezentacija. Problem nastaje upravo zato što relevantna informacija, kao odgovor PI, treba da proistekne iz semantičke podudarnosti, a ne bilo kakve podudarnosti termina iz pomenutih reprezentacija.

Najčešći problemi korisnika koji se javljaju pri formulisanju upita u SPI proizilaze iz nedostajućih znanja poput: 1. poznavanja i tačne primene komandi jezika upita; 2. odgovarajućeg izbora Bulovih i drugih operatora i 3. terminološke kompatibilnosti korisnikovog upita i opisa dokumenata u BP. Prva dva problema je moguće umanjiti, budući da je reč o sintaksnim teškoćama i proceduralnom znanju vezanom za konkretni SPI. Iako svi problemi u PI ostaju do naših dana, problem opisivanja, kako predmeta nekog dokumenta, tako i upita, budući da je semantičke prirode, ostaje glavni uzrok niskim performansama SPI (Fugmann, 1982; Ingnjersen, 1987; Rada, 1987; Furnas, 1987; Fidel, 1986; 1991; Chen, 1992). Ove teškoće uvećava i problem nekonzistentnosti indeksera u izboru predmetnih termina za dodelu dokumentima. Zbog nedovoljne preciznosti, koja može narasti u susretu sa korisnikovim upitom, PI u tekstualnim BP praćeno je većim faktorom neizvesnosti nego što je to slučaj kod PI u drugim vrstama BP (Salton, 1983; Milstead, 1994; Reich, 1990). Zbog velike uključenosti i uticaja znanja za pitanja predstavljanja semantike upita i tekstualnih informacija, od značaja su inteligentne tehnike, tj. tehnike koje dodatnim znanjem povećavaju efikasnost semantičkog pristupa.

1.2 ZNAWA POTREBNA U PI I U SBZ ZA PI

Ukupno znanje za PI se može grubo podeliti na dve značajne celine - ekspertizu vezanu za sistem i ekspertizu u predmetnoj oblasti (Ingnjersen, 1984). Učesnici u SPI se, prema pojednostavljenoj predstavi posedovanja ovih znanja, mogu podeliti u četiri grupe: 1. oni koji imaju obe ekspertize, 2. posrednici koji imaju ekspertizu o sistemu, 3. krajnji korisnici koji imaju ekspertizu o predmetnoj oblasti i 4. korisnici koji nemaju nijednu od ovih ekspertiza. Zbog potrebe poznavanja jezika indeksiranja i postupaka za formulisanje upita u radu sa SPI, korisnicima je potrebna pomoć iskusnijeg stručnjaka za PI. I to ne samo pri izboru baze podataka iz ponude komercijalnih informacionih servisa - potreba za njim naročito je vidljiva u online SPI i pri istovremenom PI više baza podataka. Specijalisti za pojedine grupe BP i predmetne oblasti su obavezni posrednici za povremene i manje iskusne korisnike. Često radeći zajedno sa korisnikom, takav posrednik je, polazeći od razumevanja korisnikove potrebe za informacijama, a zatim koristeći znanja o predmetnom području, specifičnostima BP i jeziku predmetnih oznaka u BP, izvodio prevođenje korisnikovog zahteva, odnosno - formulisanje upita, na taj jezik one BP koja bi u PI dala najbolji odgovor za

korisnika. Pomenute ekspertize - za posrednika i krajnjeg korisnika - mogu se sa više detalja predstaviti kao niz znanja čiji opis sledi.

1. Znanje o predmetnom domenu kojem upit pripada - pojmovima i vezama između njih i, posebno, o terminologiji kojom se to znanje, pojmovi i veze iskazuju u tom domenu: identifikovanje širih i užih pojmova/termina za predmet od interesa ili vrste njihove asocijativne veze sa nekim drugim pojmom/terminom, jeste znanje kojim i korisnik-ekspert može doprineti uspehu u PI. Zbog prirode procesa PI, u kojima se intenzivno koristi pamćenje i zaključivanje, za uvek razučeni domen znanja date baze podataka potrebne su što bogatije reprezentacije tih znanja.

2. Znanje o izvođenju pretraživanja, strategijama - opštim postupcima za rešavanje ovog zadatka ili specifičnim postupcima vezanim za konkretnu BP, smatra se suštinskim delom znanja posrednika u PI. Kao minimum su potrebni poznavanje dejstva logičkih operatora i lakoća njihovog kombinovanja sa terminima jezika pretraživanja i komandama jezika upita.

3. Znanje o BP, organizaciji predmetnih kategorija i predmetnih oznaka kao posledici primenjenog načina njihovog dodeljivanja jedinicama u toj BP, kao i znanje o komandama za pretraživanje u toj BP - najčešće su u posedu posrednika kojim on utiče na efikasnost u PI.

4. Proceduralno znanje o primeni računara i softvera sa kojim se radi; ono obuhvata i razvijanje interakcije korisnika i sistema i predviđanje ponašanja sistema zavisno od nekih okolnosti; online SPI treba da bude takav da se smanjuje potreba za proceduralnim znanjem, kako bi se pretraživač bavio samo (za njega) relevantnim elementima zadatka, odnosno pojmovima iz svoga zahteva; znanje koje treba dodati radi ublažavanja nedostataka čovekovog pamćenja doslednosti, sistematičnog povezivanja postupaka/izbora i dobijanih rezultata.

Kao i kod drugih ekspertnih sistema, i za PI se pošlo od pretpostavke da bi se, primenom predstavljanja znanja eksperta (u PI) i primenom pravila zaključivanja koja pokazuju kako ekspert primenjuje to znanje, moglo automatski određivati šta u datoj situaciji u PI treba učiniti. Pritom bi to određivanje moglo biti u komunikaciji sa korisnikom, na primer, u obliku dijaloga/konsultacije, ili bi, nakon dobijenih podataka od korisnika, odlučivanje bilo prepušteno sistemu. Većina ES za PI obuhvata uobičajene komponente ES sa raspodelom određenih klasa znanja posrednika u SPI ugrađenih u više modula:

1. baze znanja koje obuhvataju najmanje dve vrste znanja: znanje eksperta za određeni domen (nauku) i znanje o strategijama i pravilima za izbor termina i formulisanje upita tokom PI;

2. mehanizam zaključivanja ili rezonovanja koji omogućava sistemu da iz baze znanja pronađe relevantno znanje i primeni ga na rešavanje problema do zaključka o rešenju ili odgovarajućeg saveta;

3. baza činjenica koja opisuje tekući status problema koji se rešava, a koje uglavnom daje korisnik u interakciji sa sistemom;

4. korisnički interfejs koji omogućava komunikaciju/interakciju sistema i korisnika kojim je često obuhvaćen i modul za obrazlaganje preporuka ili zaključaka sistema.

Od komponenti u SBZ za PI, pored baza znanja i mehanizma za zaključivanje, korisnički interfejs bi, npr. u konsultativnom sistemu, imao odlučujuću ulogu, s obzirom na potrebnu interaktivnost sistema sa korisnikom (Reich, 1991). U većini razvijanih ekspertnih sistema za PI, prihvaćeno je da za navedene probleme u PI, minimalno rešenje može dati predstavljanje i primena znanja čoveka-posrednika u PI zbog pretpostavke da se u njemu akumuliraju obe ekspertize. Stoga je, skoro u svakom od inteligentnih sistema iz proučene literature, prisutna baza znanja (BZ), najpre ona koja je nastala sistematskim prikupljanjem i organizovanjem znanja za datu naučnu oblast. Kako je veliki broj prototipova zahvatio rešavanje zadatka formulisanja upita, BZ koja mu je odgovarala najčešće je implementirana u obliku pravila zaključivanja (IF - THEN pravila) kako za izbor strategije tako i za izbor termina kojima će se formulisati upit. Ova obeležja su u skladu sa definicijom koju je dala Karen Spark Xouns, prema kojoj se pod *inteligentnim sistemom za PI* podrazumeva računarski sistem sa mogućnostima *zaključivanja nad znanjem* i rezonovanja kojima uspostavlja vezu između korisnikovog, moguće nedovoljno preciznog, zahteva za PI i skupa kandidata relevantnih dokumenata. Definiciju je dopunio H. M. Bruks (1987) precizirajući domen ugrađenog znanja (o dokumentima, korisnicima, temama...), kao i upotrebu informacija o korisnikovom problemu u zaključivanju o tome koji će dokumenti doprineti razrešavanju korisnikovog problema. U odnosu na druge tehnike pristup sa inteligentnim PI dalo je korisniku centralnu ulogu, odnosno, posredno i korisničkom interfejsu.

Komunikacioni aspekti procesa PI mogu se vezati za različita znanja učesnika. Npr., nisu sva znanja, makar i u uskom predmetnom domenu, dovoljno ažurna u posredniku (ili u BZ) u odnosu na nekog korisnika koji dolazi sa upitom, i terminologijom, npr. iz najnovijih istraživačkih laboratorija. Taj problem se u SBZ nadovezuje na problem prikupljanja znanja. S druge strane, nije svaki korisnik u mogućnosti da tačno iskaže svoj informacioni problem. Za ove probleme su vezane teorijske postavke semantičkog i kognitivnog pristupa u razvoju SBZ za PI od početka do danas.

1.3 KOGNITIVNI PRISTUP U PI

Za razumevanja značaja jednog od glavnih obeležja istraživanih i primenjenih pristupa u razvoju SBZ za PI, treba poći od pojma relevantnosti informacija i uloge koju je taj pojam imao tokom razvoja discipline PI. U ocenjivanju rezultata PI, kao kriterijumi merenja korišćeni su koeficijent potpunosti (broj relevantnih dokumenata pronađenih upitom podeljen brojem svih relevantnih dokumenata) i koeficijent tačnosti (broj relevantnih dokumenata pronađenih upitom podeljen brojem svih pronađenih dokumenata). Ove mere baziraju se na subjektivnoj oceni korisnika koji, po pregledanju rezultata pretraživanja dobijenih od sistema na svoj upit, određuje broj dokumenata koji su relevantni. Drugi naziv za koeficijent tačnosti je koeficijent relevantnosti. Koeficijent potpunosti je teže odrediti, jer se za neki upit obično ne zna koliko u kolekciji ima ukupno dokumenata koji su relevantni za taj upit. Niske vrednosti koeficijenata inspirisala su, i danas inspirišu, traganja za poboljšanjima performansi IS. Posle niza teorijskih i eksperimentalnih istraživanja efikasnosti PI

tokom razvoja ove discipline, zaključeno je da je primena čovekove ocene relevantnosti u PI kao kriterijuma merenja pokušaj formalizovanja, u svrhe merenja, jednog vida čovekove intelektualne moći, a problemi koji pritom nastaju su problemi koji proističu iz ocpimene ljudi kao mernih instrumenata (Ellis, 1984).

1.3.1 Povratna informacija o relevantnosti i statističke metode u PI

Proces pretraživanja podrazumeva inicijalno pretraživanje i modifikaciju upita za iteraciju pretraživanja u drugoj fazi, na osnovu povratne informacije o relevantnosti informacija (relevance feedback) dobijenih u prvoj fazi (Salton, 1983, 1989). Jedna od ispitivanih metoda koja bi automatski generisala upit u drugoj fazi, uzima početni upit kao vektor termina kojem se dodaju termini iz dobijenih relevantnih informacija, a oduzimaju termini iz onih koje se ocene nerelevantnim. Druga metoda se, na sličan način, bazira na izmeni reprezentacije dokumenta. Ono što je ovde važno istaći, jeste to da su modifikacije u obe metode, koje se sastoje u približavanju relevantnog dokumenta upitu i udaljavanjem nerelevantnih od njega, nastale na osnovu statistike. One se izvode bez ikakvog uvida u karakteristike pridruženim relevantnim ili nerelevantnim dokumentima, kako bi se otkrilo koji su to, npr., termini koji ih takvima čine, tačnije, kako bi se našlo koji su termini najpogodniji za predstavljanje upita.

Neka istraživanja polaze od povratne informacije o relevantnosti u PI i primenjuju je u SPI, kao sistemu kojim se može automatski upravljati, sa ciljem da se dođe do mogućnosti za izgradnju adaptivnog SPI (Frants, V., 1991). Pri tome se automatsko upravljanje odnosi na sam proces pretraživanja, a adaptivnost se odnosi na sistem, a ne na prilagođavanje korisnika sistemu koje se, u opštem slučaju, podrazumeva u tradicionalnim SPI. Iako se i u ovom primeru SPI, kao sistem upravljanja, istražuje kao mogući odgovor na problem (u velikom broju istraživanja inače navođen kao „anomalija” prirodnog jezika) koji se manifestuje kao posledica nekompatibilnosti jezika korisnika i sistema, i ovde se poboljšanje postiže preko algoritama baziranih na statistici. Uviđa se da problem nastaje kao posledica semioze - prevođenja mentalnih predstava korisnika pri formulaciji upita, ali i svih prevođenja i preslikavanja iskaza na korisnikovom jeziku u jezik sistema. Korisnik ima na umu svoju stvarnu informacionu potrebu (npr. S) i želi da je sistemu postavi kao upit, ali je zbog anomalije jezika i semioze stvarno iskazuje (nevoljno) izmenjenu, npr. kao U (a ne kao S). Samo jedan deo odgovora sistema (budući da je to odgovor na upit U) zadovoljava korisnikovu stvarnu potrebu i korisnik uglavnom nije svestan onog dela odgovora koji nije pronađen. U redukovanoj interakciji, koja se postiže primenom statističkih metoda, ne dolazi do otvaranja strukture podataka sistema kojom bi se korisniku omogućilo da sagleda semantiku tih podataka, postojeću u sistemu u vezi sa aktuelnim upitom. Najboljem rezultatu koji bi bio moguć doprinelo bi, pored adaptiranja sistema, i korisnikovo prilagođavanje jeziku sistema kao posledica saznavanja i učenja tokom upita.

Situacija, u kojoj se obe strane prilagođavaju, najbolje se opisuje terminom „usaglašavanje”, tačnije, „dogovaranje (negotiation)” kako bi se došlo do što boljeg rezultata. Međutim, ma koliko se problem nekompatibilnosti jezika

nametao kao glavni, on nije i jedini - on je u tesnoj vezi sa nivoom znanja prisutnog kod oba učesnika - korisnika i sistema. Ako se njihovi kompleksi znanja predstave odgovarajućim tezaurusima, za efikasnu komunikaciju bilo bi poželjno da tezaursi budu usaglašeni. Kako potpuno usaglašeni ne mogu biti - zbog promenljivosti prirodnog jezika kojim se iskazuju tekstualne informacije (upita ili dokumenata) i, takođe, zbog nemogućnosti tačnog iskazivanja problema, potrebno je da se u toku sesije razlika među njima smanjuje u nekom stepenu. Inteligentni SPI bi olakšavao komunikaciju pokazivanjem svoga znanja koje je relevantno za dati trenutak rešavanja problema, a njegova adaptivnost bi se mogla ostvarivati obogaćivanjem BZ znanjem i elementima iz korisnikovog upita i iz zaključivanja i razmene znanja, i time učenja sistema, tokom interakcije u sesiji pretraživanja (Guntzer, 1989).

1.3.2 Kognitivna priroda procesa u PI

Istraživanja vezana za performanse i merenja u PI, ukazala su i na novosagledana svojstva ocene relevantnosti. Dobijanje korisnikove relevantnosti, uvedene u prvoj fazi razvoja SPI, temeljilo se na iskazanom upitu korisnika i pretpostavci da su informaciona potreba i ocena relevantnosti korisnika statični. Ova polazišta su se pokazala nerealnim: iskazani upit ne mora da bude pouzdan prenosilac informacione potrebe, npr. kada je ona tako složena da se ne može tačno iskazati; s druge strane, ta potreba se menja tokom pretraživanja. Za ove nalaze vezala se grupa istraživača početkom osamdesetih koja u svoje projekte ugrađuje tehnike i rezultate iz istraživanja VI. Novom pravcu istraživanja doprinosi prisutnost kognitivnog stanovišta, čiji su začetnici naučnik Bruks (Brookes B. C.) sa svojim radovima u drugoj polovini sedamdesetih i Belkin s početka osamdesetih. Belkin je ove nalaze razradio preko pojma ASK (Anomalous State of Knowledge) koji u integraciji sa informacijom predstavlja kognitivni nivo (sistema) uporedo sa kojim stoji lingvistički nivo istraživanih sistema (Belkin, 1980, 1982). Razmatranje sazajne prirode procesa u PI dovelo je do Bruksove fundamentalne jednačine informacione nauke - $I + K_{SS} \rightarrow K_S + S_{SS}$, kojom se iskazuje da informacija modifikuje strukturu znanja od K_{SS} na $K_S + S_{SS}$. Struktura znanja (SZ) tokom vremena postaje značajni pojam u informacionoj nauci, a po svom tvorcu ona oesubjektivno interpretira o podatke koje čovek dobija iz spoljnjeg sveta i transformiše ih u informaciju. Ovo je u skladu sa opšte prihvaćenom - doduše, jednom od mnogih - definicija informacije koja kaže: oinformacija je protumačeni podatak o. Po drugoj definiciji informacije, ona je mera povećanja znanja čoveka nastalog po njenom prijemu i, budući da ta mera zavisi od prethodnog znanja primaoca, reč je o relativnom određenju. Belkin na sličan način definiše informaciju, ističući da ona transformiše ne samo doda-vanjem već i podešavanjem sazajne strukture primaoca. U kognitivnoj psiho-logiji se to pripisuje shemama koje su u čovekovoj svesti uopšteni slučaj nekog objekta/događaja, ali koji dopušta i njihove specifičnosti (objekata/događaja) pošto se izmene neki skupovi vrednosti (očekivanja na osnovu prošlih iskustava) oepodrazumevanih karakteristika o (slots). Ovo je ilustrovala još teorija okvira (frame theory) (Minsky, 1975) po kojoj se perceptualna shema menja preko interfejsa između interne sheme pojedinca i podataka koje prima iz spoljnjeg sveta. Tim podacima je dopušteno da utiču na okvir (jedan tip sheme) preko oeslotova o na nižim nivoima okvira, za razliku od

onih na višim nivoima koji predstavljaju nepromenljive entitete za te okvire (Cole, 1994).

Iz Bruksove jednačine proizilazi da se struktura znanja meri istim jedinicama kao i informacija, a što se tiče ocena relevantnosti kojim bi se ona merila, budući da zavise od mentalnog stanja čoveka tokom čitanja informacije (dokumenta) koju treba da oceni, ova ocena se menja, odnosno kognitivne je prirode (Brookes, 1981). Otuda i pristup istraživača tehnika ES a priori polazi od uključenosti znanja i saznanog u procesima PI i jedan od glavnih zadataka istraživanja primene tih tehnika u PI jeste da uključivanjem znanja u sistem rasterete čoveka u kognitivnom naporu, pre svega olakšavanjem interakcije između čoveka i znanja (Buckland, 1991). Po mišljenju Ingversena (Ingjersen) „Zadatak PI je da dovede kognitivne strukture autora (informacija primarnih dokumenata, prim. M. M.), projektanta SPI i indeksera u sklad sa kognitivnim strukturama korisnika, da bi se suočio i nosio sa aktuelnim potrebama” (Ingversen 1982).

U istraživanjima se još od osamdesetih godina anticipira da se u budućim SPI neće oslanjati samo na jednoj strategiji pretraživanja, već na ugrađivanju većeg broja tehnika, koje se pozivaju i kojima se upravlja preko dijaloga baziranog na kognitivnom modelu korisnika. Prvenstveno se misli na mehanizam odlučivanja u određivanju koja je teorija primenljiva za neko određeno ASK, kojih, naravno, ima više (Belkin, 1982). S druge strane, sve broj-nija istraživanja user-friendliness-a, lakoće rada sa sistemom, stepena korišćenja sistema i drugih pokazatelja uspešnosti sistema u dobijanju informacija koje su korisniku potrebne, uzimajući u obzir i njegovu informacionu potrebu viđenu kao ASK, ukazuju da se relevantnost možda neće više kvantifikovati kao ranije. Naime, drugačiju kvantifikaciju nalaže problem neodređenosti (uncertainty) koja se u kognitivnoj paradigmi vezuje za strukture znanja svih učesnika/komponenti: korisnika, posrednika, indeksera, primarnih izvora informacija... (Bookstein, 1985; Ingjersen, 1987). U vezi sa neizvesnošću, u istraživanjima je razmatrana teorija rasplnutih (fuzzy) skupova u PI i primena na njoj zasnovanog računa lingvistički kvantifikovanih propozicija u SPI za obradu nepreciznih upita (Kacprzyk, 1986).

2 ISTRAŽIVAWA RAZVOJA SISTEMA SA BAZAMA ZNAWA

2.1 PRIMERI ISTRAŽIVAWA I RAZVOJA SPI SA BAZAMA ZNAWA

Istraživanja pristupa sa obuhvatom znanja kojima se informacioni stručnjak - posrednik u PI - služi za građenje upita i heuristika za njegova poboljšanja, najpre su imali za cilj razvoj sistema koji su u donošenju odluka u PI bili samostalni. Međutim, oni SPI koji su takođe koristili znanje u računaru, ali u interakciji sa korisnikom, pokazali su se vitalnijim i za takve postoji mnoštvo primera u prikupljenoj literaturi. U njoj je opisano više desetina ES ili SBZ za izvođenje zadataka u oblasti PI od kojih, ipak, najveći broj jesu prototipovi ili sistemi u razvoju. Specifičnosti preko kojih se mogu analizirati odnose se na njihovu svrhu, domen znanja, strukture predstavljenog znanja, baze znanja, komponente arhitekture, korisnički interfejs, obradu prirodnog jezika, jezik

programiranja ili implementacioni softver sa kojim su razvijeni... Radi ilustrovanja glavnih pojmova, ovde su ukratko dati prikazi nekoliko takvih sistema.

Sistem OASIS (Goldsmith et al., 1986) kombinuje funkcionisanje nekonsultantaCE u offline i online modu i obuhvata intervju korisnika, online PI i pokazivanje rezultata. Određivanje relevantnosti rezultata i preformulacija upita izvode se u interakciji sa korisnikom u offline modu. BZ sa reprezentacijom znanja u obliku produkcionih pravila nezavisna je od programa i može se ažurirati. Interpreter pravila čeka na poruku ili od korisnika ili od BP, pretražuje bazu pravila sve dok ne nađe ono koje se podudara sa tekućim stanjem činjenica i uslovima sistema kada izvršava akciju iz zaključka pravila.

CANSEARCH (Pollitt, 1987), ekspertni sistem, koji koristi pravila, zamenjuje informacionog stručnjaka posrednika i pretražuje u BP MEDLINE reference dokumenata na upit kojim se traži odgovarajuća terapija raka. Krajnjem korisniku, koji uz pomoć menija bira termine, sistem pomaže u formulaciji upita. U predstavljanju znanja primenjena je hijerarhijska organizacija okvira (frames) za termine iz tezaurusa Medical Subject Headings (MeSH). ES je napisan na jeziku Prolog.

MICROARRAS (Gauch et al., 1989) je online pomoćnik za PI u full-text bazi koji locira relevantne pasuse u dokumentima. U cilju poboljšanja rezultata pretraživanja automatski reformuliše Bulov upit, pošto je prethodno pokazao pronađene pasuse i rangirao ih po relevantnosti. Sistem radi sa invertovanim datotekama i tezausom, ali prihvata i slobodni tekst za pretraživanje. ES ima BZ o strategijama pretraživanja kao BZ koja je nezavisna od predmetnog domena. Programi su pisani na jeziku C.

EP-X (Parsaye et al., 1989), posrednik u PI sa BZ, razvijen je u okruženju BP Chemical Abstracts. Cilj interakcije sa EP-X jeste preslikavanja teme od interesa u skup odgovarajućih termina. Ono se izvodi preko iterativnog unosa ključnih reči i sintagmi i interpretacije, postupaka za postizanje dogovora između korisnika i sistema u cilju dobijanja što optimalnije tačnosti i potpunosti PI. Sistemu je za ovakvu interakciju potrebno znanje o: semantičkim primitivama (pojmovima), članstvu u klasi (hijerarhiji pojmova), odnosu između klasa (temi), kao i o drugim relacijama.

Inteligentni pomoćnik SAFIR (Reiter, 1989) je ES koji bira host/BP, ključne reči i sintagme, i izvodi PI uzimajući u obzir koeficijente potpunosti i tačnosti. Znanje o hostovima/BP, predmetnim područjima, strategijama i korisnicima, predstavljeno je logikom predikata i produkcionim pravilima, semantičkom mrežom i hijerarhijom okvira. Za izgradnju ES je korišćena ljuska ES Nexpert Object Shell.

U NLM (National Library of Medicine, SAD) koristi se pojmovno indeksiranje koje se bazira na predstavljanju pojmova iz dokumenta preko okvira. Sistem omogućava nasleđivanje, obavezivanje na ograničenja i druge funkcije koje se implementiraju dodatim procedurama koje su deo sistema okvira (frames). U ovom sistemu za indeksiranje u NLM, indeksirer unosi indeksne termine, a sistem sa BZ izgrađenom polazeći od tezaurusa MeSH, vodi korisnika kroz mrežu kategorija i relacija u domenu. Za svaki dokument u BP, tokom sesije proizvodi se preko baze znanja skup okvira specifičnih za dokument, skup indeksnih

okvira. Oni se povezuju kao instance odgovarajućih okvira iz baze znanja okvira koja predstavlja opšte znanje domena. Isti tip pojmovnog indeksiranja (na bazi okvira) potrebno je u SPI sa BZ, odnosno u PI sa SBZ (Humphrey, S. M., 1989).

2.2 VAŽNIJI REZULTATI ISTRAŽIVAWA SBZ ZA PI

Preklapanja istraživanja na problemu obrade prirodnog jezika, u više naučnih disciplina - pretraživanju informacija, računarskoj lingvistici, veštačkoj inteligenciji, dovela su i do jednog broja operatora (algoritama) - od onih, davno traženih (Sparck, 1971), iz automatskog indeksiranja, do onih za morfološku, sintaksnu i semantičku analizu i mašinsko razumevanje teksta. Obrada prirodnog jezika u PI, međutim, ostaje i dalje nezaobilazni i teško rešivi problem u PI. Neki istraživački primeri koristili su relacije BP kao model za predstavljanje znanja za rešavanje prirodnojezičkih problema tekstualnih informacija (Parsaye, 1989). Dok su reči koje su njihove atomske komponente, u zadatku pretraživanja informacija, najčešće tretirane kao niske simbola uz zanemarivanje njihovog semantičkog sadržaja, u ovoj grupi istraživanja nastojalo se da se inteligentnijom obradom upravo taj sadržaj uključi u rešavanje problema. Npr., u pristupu koji je uzet kao primer (Pavlović-Lažetić, 1987), za razrešavanje poznatog problema u PI - problema višeznačnosti termina, polazi se od teksta na prirodnom jeziku, za koji se, najpre, dobije reprezentacija preko leksičkog tipa podataka. Ovaj tip podataka se definiše kao reč okarakterisana bitnim morfološkim, sintaksnim i semantičkim svojstvima. Sa minimalnim procesiranjem teksta za reprezentaciju leksičkim tipom i uz primenu metoda relacionih baza podataka, ostvaruje se baza znanja koja omogućava dobijanje informacija iz teksta kao odgovor na upit. Relacija relacije šeme predstavlja i objekte i veze među objektima koji su u izvornom tekstu predstavljeni rečenicom ili delom rečenice koji daju attribute od značaja za objekat/vezu. Ovo istraživanje koje je bazirano na eksperimentima leksičke obrade, nalazi da je relacioni sistem i leksički pristup u tretiranju teksta izuzetno pogodan za leksičke obrade teksta, pa i u operacijama u PI kakve su indeksiranje i pretraživanje tekstualnih informacija.

Primer razvijanja tehnika (Chen, 1994) kojima će se u inteligentnom IS moći da ekstrahuju i čuvaju informacije iz tekstova ilustruje takođe primenu relacionog modela (sa nestrukturiranim podacima), kao metode predstavljanje znanja i manipulisanja znanjem. Istraživana tehnika može imati praktičan značaj za automatsku izgradnju BZ iz dokumenata, jer smanjuje poznati problem prikupljanja znanja, a takođe je dobar primer uloge tezaurusa kao dela BZ; u njemu se komponente tezaurusa kao pojmovne memorije eksplicitno povezuju sa inteligencijom SPI. U sledećim odeljcima je sa više detalja prikazano rešavanje zadataka u PI sa bazom znanja građenoj na principu tezaurusa.

2.2.1 Istraživanja tezaurusa kao baze znanja za SPI

Tradicionalni alat u PI - tezaurus - obuhvata skup termina, uglavnom iz leksike prirodnog jezika, i skup semantičkih i generičkih veza među terminima kao kontrolisani jezik za indeksiranje, služi za prevođenje sa prirodnog jezika dokumenata, indeksera ili korisnika informacija na podjezik informacionog sistema. Još je Vilfred Lankester (F. Wilfred Lancaster) 1968. godine definisao

tezaurus kao ulazno/izlazno sredstvo čija je svrha da oedovede jezik pretraživača u podudaranje sa jezikom indeksera (Reich, 1991). Iako su istraživači SBZ u PI obuhvatili i razvoj niza komponenti i sistema, kao što su komponente za obradu prirodnog jezika (Pavlović-Lažetić, 1987) i modelovanje korisnika (Brajnik, 1990), izgleda da je tezaurus, kao osnova za komponentu koja treba da obuhvati domen znanja, najzahvalniji sa istraživačko-razvojnog aspekta. Većina prototipova ES za PI koriste neki oblik tezaurusa, uglavnom kao BZ o domenu - predmetnoj oblasti BP, koja se kao BZ koristi i u pretraživanju i u indeksiranju. Postoje brojni primeri inteligentnih sistema u kojima je tezaurus jedan od modula, u većini slučajeva, baza znanja (Pollitt, 1987; Monarch, 1987; Croft, 1987; Bruandet, 1989; Humphrey, 1989;). O tome i u novijoj literaturi ima eksplicitnih stavova: "A thesaurus is a kind of knowledge base, which provides controlled vocabularies for describing the subject matter" (Lee, J. H. et al., 1994). Znanje u takvoj BZ je često predstavljeno kao mreža koja je slična mreži deskriptora (pojmovna) u tezaurusu. U implementaciji BZ primenjene su strukture znanja - semantička mreža (SM), okviri, produkciona pravila, pojmovni grafovi, objekti i neuronska mreža (Cohen, 1986; Kjeldsen, 1987; Meghabhab, 1994; Chen, 1995).

U najnovije vreme se tezaurus - BZ, sa uvođenjem težina i algoritama širenja aktivacije, često tretira analogno neuronskoj mreži. Dodela težina je proistekla iz proširivanja ideje o nalaženju rastojanja ili putanje između tačaka (termina) u strukturi tezaurusa na kojima se, u PI, lociraju dokumenti i upiti. Smatra se da se sa upitom najbolje podudaraju oni dokumenti koji su, preko tezaurusnih tačaka, na najkraćoj putanji od upita. Težine se dodeljuju odnosima među terminima prema pretpostavljenim jačinama njihovih veza i udaljenost upita i dokumenta izračunava se sabiranjem težina veza duž puta koji ih povezuje.

2.2.2 Integracija tezaurusa za bogatiju bazu znanja

U nekoliko projekata BZ se izgrađuje integracijom tezaurusa srodnih BP (Rada, 1987). Tako je jedan od najvećih projekata koji je imao za cilj iskorišćenje znanja u postojećim tezaurusima vezan za tezauruse u medicini i BP MEDLINE, počeo krajem osamdesetih sa dodavanjem pojmova u MeSH (Medical Subject Headings) uključivanjem tezaurusa CMIT (Current Medical Information and Terminology) i SNOMED (Systematized Nomenclature of Medicine). U ovom projektu razvija se dugoročno inteligentni automatizovani sistem UMLS (Unified Medical Language System) koji treba da oerazume biomedicinske termine i njihove međusobne veze, a sa ciljem da se korisnicima pomogne u PI. Glavne komponente UMLS su: 1. Metatezaurus, koji sadrži informacije o pojmovima (u biomedicini) i njihovoj reprezentaciji u više od deset tezaurusa i rečnika koji su integri-sani, 2. Semantička mreža sa tezaurusnim informacijama o tipu termina i, 3. Mapa izvora informacija o opsegu, lokaciji, rečniku i uslovima pristupa bazama podataka u ovom sistemu. Projekat UMLS u NLM (National Library of Medicine) je najveće ostvareno integrisanje tezaurusa i drugih izvora znanja (Rada, 1988, 1990).

U drugom primeru integracije tezaurusa, deo odrednica iz normative kontrole u Kongresnoj biblioteci - LCSH (Library of Congress Subject Headings) ugrađen je kao BZ za inteligentni SPI (Chen, 1991). U cilju proširivanja obuhvata znanja u BZ Public KB uključena su i dva spoljna izvora u obliku tezaurusa: celi ACM CRCS (Association for Computing Machinery Computing Review Classification

System), koji uvodi opšte kategorije računarstva u ACM za klasifikovanje literature o računarstvu, i deo odrednica iz LCSH, koji predstavlja opšte računarske termine za izbor u klasifikovanju knjiga iz računarstva u Kongresnoj biblioteci SAD. Za svaki od ovih izvora znanja postojali su razlozi za integrisanje. ACM CRCS ima hijerarhijsku strukturu sa četiri nivoa specifičnosti i predmete koji su manje specifični, ali predstavlja pogodno opšte termine iz računarstva i njihove veze. Identifikovano je 18 opštih kategorija i 1.141 specifični termin u ACM tezaurusu. Identifikovano je 5 vrsta veza u oko 3.000 pojavljivanja: BT/NT - između specifičnih (širih - BT i užih - NT) termina, zatim veze RT (razne semantičke veze između termina) i ISA/INST (veze is-a i instance-of) koje pokazuju veze između specifičnih termina i opštih kategorija. LCSH je mreža predmetnih odrednica koja sadrži termine i unakrsne uputnice između termina koje ukazuju na pet odnosa: USE/UF, koji pokazuju sinonimni odnos, RT - pokazuje asocijativni odnos, a BT/NT - hijerarhijske odnose. Podskup ugrađen za ovu integraciju sadrži preko 11.000 termina i 33.000 odnosa.

2.2.3 Istraživanje pojma u konsultaciji sa BZ za izbor termina

Ispitivanje načina korišćenja tezaurusa za inteligentno PI dovelo je do interaktivne navigacije u tezaurusu preko ugrađenih algoritama zasnovanih na pravilima ili dodeli težina. Za kognitivni pristup u PI od značaja su ponašanja korisnika u interakciji sa pojmovima za koje mu tezaurus pokazuje značenje, a sa ciljem da se identifikuju uspešne strategije za inkorporisanje u automatske procedure nalaženja odgovarajućeg termina za upit. Rezultati istraživanja se često razlikuju u stavu da neki rezultati opravdavaju korisnost uvođenja takvih algoritama (Chen, 1995), dok drugi ukazuju da je za korisnike najpotrebnije da imaju tezauruse koji dobro pokrivaju predmetni domen i da im user-friendly interfejs omogući lako pregledanje deskriptora kako bi pravili fine diskriminacije u svom izboru. Preliminarna navigacija korisnika kroz tezaurus pokazuje se uvek korisnom za poboljšanje rezultata u PI (Jones, 1995).

Inteligentnom PI potrebne su metode koje omogućavaju korisniku efikasno korišćenje znanja u interakciji sa BZ za izvođenje zadataka kao što su: pregledanje deskriptora i veza, istraživanje pojmova, izbora deskriptora za upite, upoređivanja i manipulacije pojmovima. Razvijene metode za istraživanje pojmova odlikuju se interaktivnošću koja je od velike pomoći korisnicima u prolasku kroz BZ sa više hiljada do miliona pojmova kako bi identifikovanja relevantnih pojmova i poboljšavanja upita postala manje kognitivno zahtevna. Istraživanjem razvijanja sistemske pomoći u istraživanju pojma i tezaurusnoj konsultaciji došlo se do algoritma za heuristički bazirano širenje aktivacije, koje je kao mehanizam asocijativnog pamćenja poznato iz istraživanja pamćenja kod čoveka (Pejtersen, 1987). Da bi se omogućilo širenje aktivacije kroz BZ, razvijeni algoritmi pretraživanja prostora znanja obuhvataju zaključivanje koje se izvodi prolaskom preko označenih veza NT, RT, BT - aktiviranjem veza - i čvorova povezanih sa inicijalnim čvorovima (pojmovima) korisnikovog upita.

Analogno modelu formulacije upita i njegove modifikacije, usvojenom u većini prototipova, korisnikovo pretraživanje BZ dodeljivanjem težina i šema širenja se sastoji iz dve faze: U prvoj fazi korisnik iskazuje dva kriterijuma

aktivacije: težine koje dodjeljuje pojedinim delovima baze znanja i težine koje dodjeljuje raznim tipovima veza. Tipovi veza su predstavljeni klasičnim oznakama BT, NT i RT i, na primer, dodela više bodova (sa skale od 0 do 10) za NT u odnosu na BT pokazuje korisnikov nameru za ispitivanje specifičnijih termina. Date težine se koriste preko algoritma širenja aktivacije za određivanje smera aktivacije, čime se omogućava korisniku da prolazi kroz izvore znanja i tipove veza efikasnije i inteligentnije. U drugoj fazi se kriterijumi za aktivaciju veza dobijeni u prethodnom koraku mogu modifikovati da odraze korisnikove kriterijume u PI.

Algoritam Branch-and-bound (BAB), ima u odnosu na druge algoritme optimalne karakteristike u pogledu nalaženja najkraćeg puta u pretraživanju prostora stanja, tokom kojeg se prvo aktiviraju najbolji čvorovi, odnosno identifikuje skup najrelevantnijih pojmova u BZ. Heuristiku dodeljivanja većih težina čvoru, do kojeg se može stići iz raznih početnih čvorova u mreži, koristili su i drugi sistemi sa širenjem aktivacije (Shoval, 1985; Cohen&Kjeldsen, 1987; Chen&Dhar, 1991).

2.3 NAJNOVIJA ISTRAŽIVAWA O INTELIGENTNOM SPI I UČEWU SISTEMA

U Uvodu pomenuta najnovija faza razvoja i poboljšavanja SPI, pomoću tehnika VI, odvija se preko sistema mašinskog učenja CE koji se baziraju na algoritmima za ekstrakovanje znanja, ili identifikaciju modela u primerima ili podacima. Automatsko generisanje tezaurusa je zasnovano na aktivnom istraživačkom području - otkrivanju znanja u bazama podataka - koje uključuje tehnike iz statističke analize, automatske obrade teksta, projektovanja BZ, najčešće sa ciljem prikupljanja znanja i izgradnje inteligentnih SPI. O njima se saznaje tokom poslednje decenije, preko istraživanja neuronskih mreža, simboličkog učenja i genetičkih algoritama (Chen, H., 1995). Tako o neuronskim mrežama, koje se nadovezuju na model vektorskog prostora i probabilistički model u PI, izveštavaju Dožkoš (Doszko) i drugi (1990), ukazujući na primene konekcionističkih pristupa u PI. Ovi pristupi uključuju neuronske mreže, širenje aktivacije, asocijativne mreže, u kojima se procesi odvijaju bez spoljašnjih programa i koji se mogu primeniti i na tezaurus. Prve realizacije konekcionističkog modela u PI su u eksperimentalnoj fazi, a dalji razvoji se kreću ka grupisanju dokumenata, automatskom kreiranju mreže, integrisanju tezaurusa, razvoju pojmovnog PI uz pomoć algoritama za pregledanje i istraživanje pojmova (Chen, 1995).

Neuronske mreže omogućavaju istraživanje i primenu pojmovnih povezivanja između dokumenata, indeksnih termina, klasifikacija i upita u automatskoj izgradnji tezaurusa, automatskoj klasifikaciji i inteligentnom PI. Za neuronske mreže se naročito ističe korisnost u procesu dogovaranja pri formulisanju upita (lquery negotiation) u inteligentnom PI, kao obliku komunikacije koja podrazumeva razmenu informacija iz konteksta između učesnika u komunikaciji (Meghabghab, 1994). Posebno se podvlači moć neuronskih mreža u savlađivanju rasplnutosti (fuzziness) - nekompletnih i nekorektnih upita u PI, u

prilagođavanju koje proizilazi iz sposobnosti učenja, u inteligentnom pretraživanju informacija koje je posledica paralelnog računanja i analoškog rezonovanja, inače karakterističnog za čoveka u rešavanju problema adaptiranjem sličnih problema iz memorije (Chen, 1995).

2.4 OCEWIVAWE SISTEMA SA BAZAMA ZNAWA ZA PRETRAŽIVAWE INFORMACIJA

Izostanak testiranja i ocenjivanja SBZ za PI razvijanih do polovine 90-tih godina je obeležje ovih istraživanja koje se može dovesti u vezu sa, u tom periodu, relativno malim brojem implementiranih sistema u odnosu na broj prototipova i sistema u razvoju. Jedan od retkih izveštaja koji se eksplicitno posvećuje ocenjivanju ovih sistema, dat je u radu Lankastera i dr. (1996), u kojem su analizirane tekuće metode testiranja SBZ, njihovi atributi, parametri, ciljevi i komponente, i opisan projekat pripreme eksperimenta za testiranje SBZ u PI. Razvijeni model za testiranje ES, koji obuhvata poređenje procedura u realizovanju zadatka sa ES i procedura bez ES, primenjen je za ocenjivanja SBZ MedIndEx-a, sistema za pomoć u indeksiranju sa bazom znanja.

Analizom skupa atributa ovih sistema utvrđeno je da su u praktičnom ocenjivanju uvek prisutna ograničenja i da se za pojedinačne sisteme primenjuju ispitivanja samo nekih kombinacija atributa koje, kao meritorne, ocenjivač mora da izabere. U tom smislu se moraju uzimati u obzir i različiti aspekti, kao što su: odlike zadatka od kritičnog značaja, stepen automatizacije, ekspertiza korisnika, stepen distribuiranja sistema i dr. Uvažavanje uticaja, npr., ekspertize korisnika na uspeh rada SBZ je dovelo do modelovanja korisnika za razvoj aplikacija kao što su sistemi za obuku korisnika u PI, sistemi za konsultovanje korisnika i korisnički interfejsi kojima se oni vode u izvođenju zadatka PI (Chiarabella, Defude, 1987; Brajnik et al. 1990; Reiter, 1989). U modelovanju korisnika je najčešća podela na: 1. korisnike eksperte i 2. neiskusne korisnike (Ingjersen, 1984, 1987; Buckland, 1991). U slučaju ocenjivanja SBZ za konsultovanje povremenih i neiskusnih korisnika u PI, potrebno je insistirati na vrhunskom kvalitetu elemenata kao što su: 1. kvalitet odgovora sistema; 2. kvalitet BZ; 3. lakoća korišćenja; 4. obuhvat sistema. Empirijsko testiranje i ocenjivanje, koje je predloženo istraživanjem, zasniva se na poređenju kvaliteta zadataka koje su korisnici izveli sa SBZ sa kvalitetom izvođenja tih zadataka bez SBZ (Lancaster, 1996).

3 ZAKQUČCI IZ ISTRAŽIVAWA PRETRAŽIVAWA INFORMACIJA POMOCU SISTEMA SA BAZAMA ZNAWA

Odnos korisnik - SPI prisutan je u istraživanjima od početka razvoja inteligentnih sistema i on je i dalje neizbežna tema koja se povezuje sa poboljšavanjem PI, usavršavanjem softvera za PI, pre svega, korisničkih interfejsa. Komponente SPI u koje se tokom poslednje decenije ugrađuju tehnike VI, projektuju se i razvijaju pod raznim imenima: pretprocesori, user friendly interfejsi, interfejsi sa bazama znanja, inteligentni informacioni sistemi, ES za PI i dr. U najvećem broju ovih sistema interakcija korisnik - sistem je, kao faza u kojoj

se znanje prikazuje, koristi, razmenjuje - otvaranjem struktura znanja - u centru pažnje, kako u razvoju sistema, tako i u njihovoj evaluaciji.

Osnovni problem u PI - definisanje preslikavanja između skupa dokumenata i skupa mogućih upita - u istraživanjima PI sa SBZ posmatra se kao implikacija između dokumenata i upita. Wu karakteriše neizvesnost koja se vezuje za pojam relevantnosti a ovaj zavisi od semantičkog modela predstavljanja sadržaja dokumenata i upita. Glavni deo dokazivanja logičke implikacije između dokumenata i upita izvodi se semantičkim zaključivanjem, dok se semantički princip ogleda u potrebi za većom BZ sa predstavljenim pojmovima domena BP. U PI SBZ imaju glavnu ulogu u obradi upita - izboru termina, formulisanju upita i njegovoj modifikaciji; drugu značajnu ulogu SBZ imaju u predmetnom indeksiranju.

U velikom broju SBZ semantički i kognitivni pristup realizuje se pomoću BZ - bogatog tezaurusa koji omogućava onoliko veza među semantičkim primitivama koliko je to potrebno za njihovo opisivanje. Uzimanje u obzir svih vrsta odnosa koji prate atomske elemente u tekstu (izbor zavisnosti među terminima) - od sintagmatskih do paradigmatičkih, semantičkih i pragmatičkih odnosa u susretu sa čovekom - je glavno obeležje SBZ i semantičkog zaključivanja u PI/indeksiranju. Kao posledica uvažavanja kognitivnog pristupa u procesima PI, suštinska potreba za razmenom znanja između SBZ i korisnika nameće ulogu korisničkog interfejsa u otvaranju struktura znanja SBZ i olakšavanju dogovaranja između korisnika i SBZ.

Zaključci o istraživanjima inteligentnih sistema za PI svode se na to da su ona i danas aktuelna. A da su i dalje neophodna, poslednjih godina nas u to uveravaju i informacije o istraživanjima o još izraženijim (a u PI poznatim) problemima u PI na Internetu . Istraživanja ukazuju da proučavane tehnike iz svih korišćenih pravaca razvoja i preovlađujućih SPI - sa ključnim rečima i punim tekstom, sistema baziranih na verovatnoći, bazama znanja i na učenju, mogu u komplementarnoj primeni doprineti projektovanju i razvoju adaptivnih i inteligentnih SPI.

Inženjerstvo znanja doživljava buran razvoj u poslednjoj deceniji i nesumnjivo značajni prostor u tom razvoju posvećuje bazama znanja. U radu opisani UMLS je primer sistema sa BZ specijalno razvijenim za PI: koristi se preko Interneta i inspirisao je druge SBZ na mreži. Bez obzira na neke postignute rezultate u inženjerstvu znanja u PI, ono nailazi na teškoće, npr. one u primeni razvijanih mehanizama za zaključivanje nad domenom koji je nužno veliki i nije jasno omeđen, ili, pak, za teško rešive probleme mašinskog razumevanja prirodnog jezika. Iako naučnici - pioniri u PI, koji se često citiraju u istraživanjima o primeni tehnika VI u PI, ističu (Sparck Jones, 1991) da ne treba precenjivati potencijal tih tehnika (što je često slučaj) i iznose argumente da u PI u nekim zadacima ljude neće skoro zameniti sistemi koji se nazivaju inteligentnim, oni ipak smatraju da će ove tehnike doprineti značajnom poboljšanju rezultata rada sa sistemima u kojima se ljudi i mašina dopunjuju.

LITERATURA

- Belkin, N. J. (1980). Anomalous states of knowledge as a basis for information retrieval. *Canadian Journal of Information Science*, 5, 133-143.
- Belkin, N. J., Oddy, R. N., Brooks, H. M. (1982). ASK for IR: Part II. Results of a design study. *J. Documentation*, 38(3), 145-164.
- Bookstein, A., Swanson, D. R. (1975). Probabilistic models for automatic indexing. *JASIS*, 26, 45-50.
- Bokstein, A. (1985). Probability and fuzzy-set applications to information retrieval. In M. Williams (ed.), *Annual Review of Inf. Sci. and Technol.*, vol. 20, 117-151.
- Brajnik, G., Guida, G., Tasso, C. (1990). User modeling in expert man-machine interfaces: a case study in intelligent information retrieval. *IEEE Trans. on Systems, man, and Cybernetics*, 20 (10), 166-185.
- Brookes, B. C. (1981). The foundation of information science. Part IV: Information Science. The changing paradigm. *J. of Inf. sci.*, 3, 3-12.
- Brooks, H. M. (1987). Expert systems and intelligent information retrieval. *Inf. Process. and Manag.*, 23(4), 367-382.
- Bruandet, M.-F. (1989). Outline of a knowledge-base model for an intelligent information retrieval system. *Inf. Process and Manag.*, 25(1), 89-115.
- Buckland, K. M., Florian, D. (1991). Expertise, task complexity, and artificial intelligence: a conceptual framework. *JASIS*, 42(9), 635-643.
- Chen, H., Dhar, V. (1991). Cognitive process as a basis for intelligent retrieval system design. *Inf. Process. Manag.*, 27, 405-432.
- Chen, P. S. (1994). On inference rules of logic-based information retrieval systems. *Inf. Process. Manag.*, 30(1), 43-59.
- Chen, H. (1995). Machine learning for IR: Neural networks, symbolic learning and genetic algorithms. *JASIS*, 46(3), 194-216.
- Chen, H., Ng, T. (1995). An algorithmic approach to concept exploration in a large knowledge network (automatic thesaurus consultation): symbolic branch-and-bound search vs. connectionist Hopfield net activation. *JASIS*, 46(5), 348-369.
- Chen, H., Yim, T., Fye, D. (1995). Automatic Thesaurus Generation for an Electronic Community System. *46, 3, 175-193.*
- Chiaramella, Y., Defude, B. (1987). A prototype of an intelligent system for information retrieval: IOTA. *Inf. Process. and Manag.*, 23, 285-303.
- Cohen, P., Kjeldsen, R. (1987). Information retrieval by constrained spreading activation in semantic networks. *Inf. Process and Manag.*, 23(4), 255-268.
- Cole, C. (1994). Operationalizing the notion of information as subjective construct. *JASIS*, 45(7), 465-476.
- Croft, W. B. (1987). Approaches to intelligent information retrieval. *Inf. Process and Manag.*, 23(4), 249-254.
- Doszkoocs, T. E., Reggia, J., Lin, X. (1990). Connectionist models and information retrieval. *Ann. Review of Information Science and Technology (ARIST)*, 25, 209-260.
- Ellis, D. (1984). Theory and explanation in information retrieval research. *J. of Inf. Sci.*, 10(8), 25-38.
- Fidel, R. (1986). Towards expert systems for the selection of search keys. *JASIS*, 37, 1, 37-44.
- Fidel, R. (1991). Searchers' selection of search keys: I. The selection routine; II. Controlled vocabulary or free-text searching; III. Searching Styles. *JASIS*, 42, 7, pp. 490-527.
- Fidel, R., Eftimiadis, E. N. (1995). Terminology knowledge structure for intermediary expert systems. *Inf. Process. Manag.*, 31(1)15-27.
- Fugmann, R. (1982). The complementarity of natural language and indexing languages. *Int'l Classification*, 9, 140-144.
- Furnas, G. W., Landauer, T. K., Gomez, L., Dumais, S. T. (1987). The vocabulary problem in human-system communication. *Comm. of the ACM*, 30, 964-971.
- Gauch, S., Smith, J. B. (1989). An expert system for searching in full-text. *Inform. Process. Manag.* 25(3), 253-263.
- Goldsmith, G., Williams, P. W. (1986). Online searching made simple: a microcomputer interface for inexperienced users, (LIR, no. 41).
- Gross, D. (1988). Applications of semantic networks in database search and retrieval. *Expert Systems*, May, 5(2), 152-154.
- Guntzer, U., Juttner, G., et al. (1988). Automatic thesaurus construction by machine learning from retrieval sessions. *Inf. Process. Manag.*, 25(3), 265-273.

- Hsinchun, C. (1992) Knowledge-based document retrieval framework and design. *J. of Inf. Sci.*, 18(4), 293-314.
- Humphrey, S. M. (1989). A Knowledge-based expert system for computer-assisted indexing. *IEEE EXPERT*, Fall, pp. 25-38.
- Ingwersen, P. (1982). Search procedures in the library - analysed from the cognitive point of view. *J. documentation*, 38, 165-191.
- Ingwersen, P. (1984). On-line man-machine interaction facilities: a cognitive view. 325-358. *U: Representation and Exchange of Knowledge as a Basis of Information Processes*, Elsevier Sci. Publishers, North-Holland.
- Ingwersen, P. (1987). Towards a new research paradigm in information retrieval. 150-168. In Wormell, I. (ed.): *Knowledge engineering: expert systems and information retrieval*. Taylor Graham, London, 1987.
- Ingwersen, P., Wormell, I. (1989). Modern indexing and retrieval techniques matching different types of information needs, *Int. Forum Inf. and Docum.*, 14, (3), 17-22.
- Jones, S. Gatford, M. et al. (1995). Interactive thesaurus navigation: intelligence rules OK?. *JASIS*, 46(1), 52-59.
- Kacprzyk, J., Ziolkowski, A. (1986). Database queries with fuzzy linguistic quantifiers. *IEEE Trans. on Systems, Man and Cybernetics*, SMC-16 (3), 474-479.
- Kjeldsen, R., Cohen, P. R. (1987). The evolution and performance of the GRANT system. *IEEE EXPERT*, Summer 1987, 73-79.
- Lancaster, F. W., Jacob, W. U. et al. (1996). Evaluation of interactive knowledge-based systems: overview and design for empirical testing. *JASIS*, 47(1), 57-69.
- Lee, J. H., Kim, M.H., Lee, Y. J., (1994). Ranking documents in thesaurus-based boolean retrieval systems, *Inf. process. manag.*, 30 (1) 79-91.
- Meghabghab, George V. Meghabghab, Dania B. (1994). INN: an intelligent negotiation neural network for information systems: a design model, *Inf. Proc. & Manag.*, 30(5)663-685.
- Milstead, J. (1990). Methodologies for subject analysis in bibliographic databases: background paper and report of meeting sponsored by IAEA and ETDE.
- Milstead, J. L. (1994). Needs for research in indexing. *JASIS*. 45(8), 577-582.
- Minsky, M. (1975). A framework for representing knowledge. In P. H. Winston (ed.), *The Psychology of computer vision*. New York: McGraw-Hill.
- Monarch, I., Carbonell, J. G. (1987). CoalSORT: A knowledge-based interface. *IEEE Expert*, 39-53.
- Ofori-Dwumfuo, G. O. (1984). Using a cognitive model of dialogue for reference retrieval. *J. of Inform. Sci.*, (9), 19-28.
- Paice, C. (1986). Expert systems for information retrieval? *Aslib Proceed.* 38(10), 343-353.
- Parsaye, K., Chignell, M., Khosafian, S., Wong, H. (1989). *Intelligent databases: object-oriented, deductive hypermedia technologies*. New York: Wiley.
- Pavlović, G. (1987). *Baze podataka i ekspertni sistemi u upravljanju tekstem*. Doktorska disertacija, Univerzitet u Beogradu, PMF, Beograd.
- Pejtersen, A. M. (1987). Development of a term association interface for browsing bibliographic databases based on end users' word associations. 92-112. In Wormell, I. (ed.): *Knowledge engineering: expert systems and information retrieval*. Taylor Graham, London, 1987.
- Pollitt, S. (1987). CANSEARCH: An expert system approach to document retrieval. *Inf. Process. and Manag.*, 23, 119-138.
- Rada, R. (1987). Connecting and evaluating thesauri. *Issues and cases. Int'l Classif.* 14(2), 63-69.
- Rada, R. (1987). Knowledge-sparse and knowledge-rich learning in information retrieval. *Inf. Process. Manag.*, 23(3), 195-210.
- Rada, R. (1990). Maintaining thesauri and metathesauri. *Int'l Classif.*, 17(3/40), 158-164.
- Rau, L. (1987). Knowledge organization and access in a conceptual information system. *Inf. Process and Manag.*, 23(4), 269-283.
- Reich, Ph. (1991). Indexing Consistency: The input/output function of thesauri. *Coll./Res. Libraries*, 52(4)
- Reiter, M. (1989). Improving online information retrieval with an intelligent front-end system. *Online Information 89*, 13th Int'l Online Information Meeting, London 12-14 Dec. 1989 Proceedings, pp. 597-604.
- Salton, G., McGill, M. J. (1983). *An introduction to modern information retrieval*. New York: McGraw-Hill.
- Shoval, P. (1985). Principles, procedures and rules in an expert system for information retrieval. *Inf. Process. Manag.*, 21(6), 475-487.

- Silvester, J. P., Genuardi, M. T., Klingbiel, P. H. (1994). Machine-aided indexing at NASA, *Inf. Proc. & Manag.*, 30(5), 631-645.
- Smith, L. (1989). Expert systems. 17th Consultative Meeting of INIS Liaison Officers, Vienna, 16-19 May 1989 (Saopštewe).
- Sormunen, E. Alkula, R. (1988). A prototype for a knowledge-based intermediary system. 291-300. In: 12th Int'l Online Information Meeting, London.
- Sparck Jones, K. (1971). Automatic keyword classification for information retrieval, London, Butterworths.
- Sparck Jones, K. (1987). Architecture problems in the construction of expert systems for document retrieval, 7-35. In Wormell, I. (ed.): Knowledge engineering: expert systems and information retrieval. Taylor Graham, London.
- Sparck Jones, K. (1987). Natural language processing, U: Intelligent knowledge-based systems: an introduction, ed. by Tim O'Shea et al., str. 175-196.
- Sparck Jones, K. (1991). The role of artificial intelligence in information retrieval. *JASIS*, 42(8), 558-565.
- Vickery, A., Brooks, H. M. (1987). PLEXUS - the expert system for referral, *Inf. Process. and Manag.*, 23, 99-117.
- Vickery, B. (1989). Intelligent interfaces for user friendly access to databases, 17th Consultative Meeting of INIS Liaison Officers, Vienna, 16-19 May 1989 (Saopštewe).

Rezime

Polazeći od problema u pretraživanju informacija (PI) i znanja potrebnih za njihovo rešavanje, u pregledu istraživanja i razvoja inteligentnih sistema za PI ističu se neka njihova obeležja i rezultati. Pod inteligentnim sistemom za PI podrazumeva se računarski sistem sa mogućnostima zaključivanja nad znanjem i rezonovanja kojima uspostavlja vezu između korisnikovog, moguće nedovoljno preciznog, zahteva za PI i skupa kandidata relevantnih dokumenata. Glavni deo dokazivanja logičke implikacije između dokumenata i upita jeste semantičko zaključivanje. U najvećem broju raznovrsnih pristupa u razradi modela ovih sistema, reč je o modelovanju reprezentacija informacija i upita sa kojima se u procesu PI može izvoditi semantičko zaključivanje. Otuda predstavljanje znanja u bazama znanja (BZ) o domenu i strategijama pretraživanja zauzima značajno mesto u istraživanjima ovih sistema koji, u početku razvoja najčešće sa nazivom ekspertni sistemi (ES) sve više postaju sistemi sa bazama znanja (SBZ). BZ o domenu - oblasti baze podataka (BP) - nastaje uglavnom iz tezaurusa, a najčešće implementirane strukture znanja u njoj su: semantička mreža, okviri, produkciona pravila, pojmovni grafovi, objekti, neuronska mreža... Jedan broj naučnih radova nužno se bavi aspektima indeksiranja, ali i mogućnošću koju indeksiranje daje za razvoj i ažuriranje BZ o domenu, npr. u fazi prikupljanja znanja tokom formulisanja upita, odnosno učenja SBZ. Istraživanja su pošla od kognitivnog karaktera procesa tokom PI i pokazala neophodnost otkrivanja struktura znanja, kao i razmene znanja između pretraživača i SBZ. Za ove kooperativne procese u inteligentnim konsultacijama korisnički interfejs ima značajnu ulogu.

Mr Milena Matić

Issledovanie intelligentnâh sistem dlâ poiskov informaciè

Rezâme

Ishodâ iz problem v poiskah informaciè (PI) i znaniè neobhodimâh dlâ ih razrešeniâ, v obzore issledovaniâ i razvitiâ intelligentnâh sistem dlâ PI, vâdelâetsâ nektorâe ih primetâ i rezulâtatâ. Pod intelligentnoè sistemoè dlâ PI podrazumevaetsâ kompâternââ sistema s vozmo`nostââ zaklâ~eniâ nad znaniem i rassu`deniâ, kotorâmi ona ustanavlivaet svâzâ me`du trebovaniem PI polâzovatelâ, mo`et bâtâ, nedostato~no to~nogo, i sovokupnostââ kandidatov va`nâh dokumentov. Glavnoè ~astââ dokazvaniâ logi~eskogo vovle~eniâ me`du dokumentami i zaprosami âvliâetsâ semanti~eskoe zaklâ~enie. V bolâ`om ~isle raznoobraznâh podhodov v razrabotke modeleè âtîh sistem, re~â idet o modelirovanii predstavleniè informaciè i zaprosov, pri pomoæi kotorâh v processe PI mo`no vâvoditâ semanti~eskoe zaklâ~enie. Poâtomu predstavlenie znaniè v bazah znaniè (BZ) o oblasti i strategiâh poiskov zanimaet zna~itelânoe

mesto v issledovaniãh ãtih sistem, kotorãe v na~ale svoego razvitiã ~asto pod nazvanie~m "ãkspertnãe sistemã (ES)", vse bolãe stanovãtsã "sistemami s bazami znaniã" (SBZ). BZ o sfere – oblasti bazã dãnnãh (BD) – ro`daetsã obãknovenno iz tezaurusov, a samãe ~astãe v neã vãpolnennãe strukturã znaniã, sleduããie: semanti~eskaã setã, ramki, produkcionnãe pravila, ponãtiennãe grafiki, obãektã, neãroni~eskaã setã... Opredelennoe ~islo nau~nãh trudov obãzatelãno zanimãetsã aspektami indeksacii, no i vozmo`nostãã, kotoruã indeksaciã predostavlãt razvitiã i zãblagovremennomu popolneniã BZ o oblasti, na pr. v faze nakopleniã znaniã v te~enie formirovaniã zaprosa, t.e. u~eniã SBZ. Issledovaniã na~ali s poznavatelãnogo haraktera processã v te~enie PI i oni pokazali neobhodimostã otkrãtiã strukturã znaniã, kak i neobhodimostã obmena znaniãmi me`du zanimãããimisã poiskami i SBZ. Dliã ãtih kooperativnãh processãh v intelligentnãh konsulãtantãh interfeãs potrebitelã imeet zna~itelãnuã rolã.

**Milena Mati}, MA
Research in Intelligent Systems for Information Search**

Summary

Starting out from problems of information search (IS) and knowledge necessary for solution of such problems, this paper gives a review of research and development of intelligent systems for IS, stressing some of their characteristics and results. An intelligent system for IS is a computer system which is capable of concluding from knowledge and through reasoning by which it establishes the connection between user's, maybe not precise enough, demand for IS and the collection of candidates of relevant documents". The main bulk of proof of the logical implication between documents and research request is semantic deduction. In most of the various approaches in the elaboration of such systems, we dealt with the modeling of the representations of information and requests with which the process of IS can perform semantic concluding. Therefore, representation of knowledge in knowledge bases (KB) on domain and searching strategies takes important place in research of these systems which, in the initial phases of development were often called "expert systems (ES)" then evolved to "knowledge bases systems (KBS)" KB on domain – parts of data bases (DB) originate mostly from thesaurus, and most frequent implemented structures of knowledge are: semantic networks, frames, production rules, notion graphs, objects, neural networks, ... A number of scientific papers inherently deal with aspects of indexing, but also with possibility which indexing gives to development and speeding up of KB on domain, e.g. in the phase of collecting knowledge during the formulating of the quest, that is learning the KBS. The researches have started from the cognitive character of the process during IS and showed the necessity of disclosure of knowledge structure, as well as the interchange of knowledge between the searcher and KBS. For these cooperative processes in intelligent consultants, user interface plays important role.