INSTRUCT - EKSPERIMENTALNI SISTEM ZA ISKANJE INFORMACIJ

mag. Mirko Popovič; Narodna in univerzitetna knjižnica, Ljubljana

UDK 519.683.5

POPOVIČ, Mirko: INSTRUCT - eksperimentalni sistem za iskanje informacij. Knjižnica, Ljubljana, 34 (1990), št. 4, str. 29-51

Članek opisuje programski paket INSTRUCT, ki je bil zasnovan kot učni pripomoček v okviru visokošolskega študija bibliotekarstva in informacijskih ved. INSTRUCT vsebuje številne metode, ki so nastale na podlagi raziskovalnega dela na področju iskanja dokumentov. Poseben poudarek je na metodah iskanja po načelu optimalnega primerjanja, ki so alternativa konvencionalnemu, Boolovemu modelu iskanja informacij. V članku so najprej opisani osnovni moduli originalne verzije INSTRUCT-a. Temu sledi opis dodatnih modulov, ki so vgrajeni v novo inačico INSTRUCT-a, ki je nato predstavljena kot učni pripomoček in kot delovno orodje v številnih raziskovalnih projektih. V zaključku so opisani glavni postopki, ki so bili potrebni za izdelavo slovenske verzije programskega paketa INSTRUCT.

UDC 519.683.5

POPOVIČ, Mirko: INSTRUCT - Experimental System for Information Retrieval. Knjižnica, Ljubljana, 34 (1990), no. 4, p. 29-51

This paper describes INSTRUCT, an interactive computer program which has been developed as a teaching aid for use within schools of librarianship and information science. INSTRUCT is a text retrieval program which incorporates a large number of techniques that have been suggested by research in document retrieval; in particular, it makes extensive use of the best match model of document retrieval, rather than the Boolean model which underlies the great majority of current text retrieval systems. A brief outline of the main components of the original version of INSTRUCT is given, which is followed by an explanation of modules added to INSTRUCT. This forms the basis for the description of INSTRUCT both

as a teaching resource and as a test-bed for research problems. The paper is concluded by a brief summary of the main modifications to the original version of INSTRUCT which were needed to develop a new, Slovene version of INSTRUCT.

Uvod

Vedno vidnejša spoznanja o slabostih in omejitvah konvencionalnega Boolovega načina iskanja informacij spodbujajo v zadnjem času interes za uporabo metod in tehnik iskanja po načelu optimalnega primerjanja. Ta način iskanja informacij odpravlja vse težave, povezane z uporabo Boolovih logičnih operatorjev, omogoča rangiranje in s tem kontrolo obsega zadetkov ter vključuje povratno relevantno informacijo s strani uporabnika. Z drugimi besedami, neizkušeni uporabniki se lahko popolnoma samostojno, brez pomoči informacijskega posrednika, vključujejo v procese iskanja informacij.

Uspešni rezultati, ki jih prinaša ta alternativni način iskanja dokumentov, so vidni v številnih novih programskih paketih, kot npr. MASQUERADE (Brzozowski, 1983), CITE NLM (Doszkocs, 1983), OKAPI (Mitev et al., 1985), MUSCAT (Porter, Galpin, 1988), STATUS/IQ (Pape, Jones, 1988) in PERSONAL LIBRARIAN (Cringean, 1989). Zanimivo je, da se v svetu vključujejo v razvijanje alternativnih oblik iskanja informacij tudi številne visokošolske ustanove na področju bibliotekarstva in informacijskih ved. V začetku so te šole sodelovale predvsem pri oblikovanju različnih simulacijskih modelov (Wood, 1984), ki so študentom služili kot učni pripomoček pri seznanjanju s konvencionalnim, Boolovim, načinom iskanja informacij. Ravno prodor nekonvencionalnih tehnik iskanja dokumentov pa je spodbudil potrebo po razvoju novih učnih pripomočkov, ki bi vsebovali nove metode ter tako študente že med študijem pripravili na soočenje s sodobnimi trendi razvoja.

S tem namenom je British Library spodbudila razvoj programskega paketa INSTRUCT (INteractive System for Teaching Retrieval Using Computitional Techniques), ki je nastal na Oddelku za informacijske vede Univerze v Sheffieldu, Velika Britanija. Študenti bibliotekarstva in informacijskih ved imajo tako možnost, da se s pomočjo tega sistema seznanijo s sodobnimi statističnimi metodami iskanja informacij (Willett, Wood, 1989). INSTRUCT se trenutno uporablja v številnih izobraževalnih institucijah ne le v Veliki Britaniji, temveč tudi v drugih državah. Ob tem pa je zanimivo poudariti, da je INSTRUCT v zadnjem obdobju postal izredno koristno orodje za testiranje številnih raziskovalnih problemov na področju iskanja dokumentov. Eno izmed teh področij je tudi procesiranje dokumentov in poizvedb v slovenskem jeziku.

Zato bom v tem članku, ki je nadaljevanje prispevka "Sodobni trendi v iskanju dokumentov", objavljenega v reviji Knjižnica, (Popovič, 1990), natančno opisal

programski paket INSTRUCT. Najprej bom predstavil dve osnovni verziji INSTRUCT-a, sledila bo razlaga uporabe INSTRUCT-a tako v izobraževalnem kot tudi v raziskovalnem programu, v zaključku pa bom opisal slovensko inačico INSTRUCT-a in nakazal smernice bodočega razvoja na področju iskanja informacij v Sloveniji.

2. Originalna verzija INSTRUCT-a programske zmogljivosti

Prva, originalna verzija INSTRUCT-a je nastala že leta 1985. Napisana je bila v programskem jeziku PASCAL in prilagojena operacijskemu sistemu PRIMOS, ki je implementiran na miniračunalniku PRIME 9950. Ta verzija INSTRUCT-a vsebuje naslednje module: procesiranje poizvedbe v naravnem jeziku (blokiranje funkcijskih besed, avtomatsko krnjenje besed in razširitev iskalnih izrazov), iskanje po načelu optimalnega primerjanja, povratno relevantno iskanje in Boolov način iskanja dokumentov.

Osnovo za iskanje predstavlja podatkovna zbirka LISA (Library and Information Science Abstracts) iz leta 1982, ki vsebuje 6.004 dokumente. Vsak zapis v tej zbirki sestavljajo inventarna številka, naslov in povzetek dokumenta. Težave z diskovnimi kapacitetami so namreč preprečile vključitev dodatnih bibliografskih podatkov. Iskanje dokumentov temelji na prisotnosti besed v naslovih in povzetkih; osnovo iskanja pa predstavlja invertirana organizacija zbirke v INSTRUCT-u.

2.1 Oblikovanje poizvedbe

INSTRUCT omogoča uporabniku vnos poizvedbe v naravnem jeziku. To pomeni, da uporabnik ne potrebuje predhodnega znanja o aplikaciji Boolovih logičnih operatorjev (čeprav ima uporabnik v kasnejši fazi iskanja na voljo tudi Boolovo opcijo). Ko uporabnik zaključi tipkanje poizvedbe, INSTRUCT najprej izloči t.i. funkcijske besede oz. besede brez predmetne vsebine (v angleščini so to npr.: AND, BUT, ARE itd.). Vse ostale besede gredo skozi proces avtomatskega krnjenja na osnovi algoritma, ki ga je za angleški jezik razvil Porter (1980). Rezultat tega procesiranja je seznam besednih krnov z odgovarjajočo frekvenco pojavljanja v dokumentih v podatkovni zbirki LISA 1982. Na tej stopnji lahko uporabnik spremeni poizvedbo z naslednjimi operacijami: dodajanje, izločanje in razširitev iskalnih izrazov. Dodajanje in brisanje besed sta zelo preprosti operaciji, ki ju dosežemo z ažuriranjem podatkovne strukture.

Postopek razširitve iskalnih izrazov pa temelji na metodi merjenja podobnosti med posameznimi nizi v tekstu. Ta metoda temelji na hipotezi, da znakovna podobnost dveh besed pomeni morebiti tudi njuno vsebinsko sorodnost. Metoda podobnosti niza, ki je implementirana v INSTRUCT-u, temelji na številu t.i. trigramov (tj.čtroz-

nakovnih podnizov), ki so skupni izbrani besedi iz poizvedbe in krnom iz slovarske datoteke. Za izračun podobnosti se uporablja algoritem, ki ga je razvil Noreault s svojimi sodelavci (1977). Rezultat tega procesiranja je seznam desetih najbolj podobnih izrazov, ki jih lahko uporabnik posamično dodaja v poizvedbo.

2.2 Iskanje informacij

Ko je uporabnik zadovoljen z nizom ključnih besed v svoji poizvedbi, lahko sproži proces iskanja dokumentov v podatkovni zbirki. V originalni verziji INSTRUCT-a ima uporabnik na voljo tri osnovne oblike iskanja: iskanje po načelu optimalnega primerjanja, relevantno povratno iskanje, Boolovo iskanje.

2.2.1 Iskanje informacij po načelu optimalnega primerjanja

Iskanje po načelu optimalnega primerjanja (t.i. "best-match" oz. "nearest-neighbour searching") predstavlja osnovni iskalni mehanizem v INSTRUCT-u. Ta način iskanja dokumentov temelji na naslednjih korakih:

- ugotovitev prisotnosti oz. odsotnosti iskalnih izrazov v posameznih dokumentih v podatkovni zbirki (dokumenti, ki ne vsebujejo nobenega od iskalnih izrazov, so izločeni iz nadaljnjega procesiranja),
- izračun vsote uteži za te izraze,
- sortiranje teh vsot kot temelj za rangiranje dokumentov.

Iskanje po načelu optimalnega primerjanja v INSTRUCT-u temelji na uporabi algoritma, ki ga je najprej predlagal Noreault s svojimi sodelavci (1977) in potem natančno analizirala še Perry in Willett (1983). Ponderiranje iskalnih izrazov pa izhaja iz koncepta inverzne frekvence dokumentov, ki sta ga razvila Croft in Harper (1979). Ta oblika iskanja informacij omogoča tudi vključevanje podatkov o relevantnosti posameznih gesel.

2.2.2 Relevantno povratno iskanje informacij

Ko se posamezni zadetki pojavljajo na zaslonu v rangiranem zaporedju, ima uporabnik možnost določitve relevantnosti posameznih zadetkov. Ta podatek tvori osnovo za modifikacijo uteži posameznih gesel, kar omogoča še kvalitetnejše rangiranje preostalih dokumentov v podatkovni zbirki. Povratno relevantno iskanje v INSTRUCT-u temelji na algoritmu, ki sta ga izdelala Robertson in Sparck Jones (1976). Ta algoritem upošteva prisotnost in odsotnost iskalnih izrazov tako v relevantnih kot tudi v nerelevantnih dokumentih.

2.2.3 Boolovo iskanje informacij

INSTRUCT ima vključen tudi modul za konvencionalno Boolovo iskanje informacij. Ta modul temelji na uporabi Boolovih operatorjev AND, OR in NOT ter je nastal predvsem z namenom, da omogoči primerjalno analizo z nekonvencionalnim iskanjem (tj. iskanjem po načelu optimalnega primerjanja).

Vsekakor pa je potrebno poudariti, da INSTRUCT omogoča tudi t.i. "hibridno iskanje". Tu gre v bistvu za proceduro rangiranja zadetkov, vendar brez tistih dokumentov, ki ne ustrezajo pogojem, definiranim z Boolovimi logičnimi operatorji. Ta način iskanja postaja vedno bolj popularen tudi v komercialnih sistemih.

3. Dodatni moduli v INSTRUCT-u

V preteklem letu je nastala nova verzija INSTRUCT-a, ki poleg obstoječih vključuje še tri dodatne module:

- razširitev iskalnih izrazov na osnovi skupne navzočnosti v bazi,
- iskanje po načelu razvrščanja v skupine,
- prelistavanje ("browsing").

Razvoj teh modulov je bil v skladu z osnovim izhodiščem INSTRUCT-a, tj.čseznaniti študente s sodobnimi trendi v iskanju informacij. Tudi ta verzija je napisana v programskem jeziku PASCAL, vendar tokrat prilagojena operacijskemu sistemu VM/CMS za računalnik IBM 3083. Dodatna razlika glede na originalno verzijo je tudi v obsegu podatkovne zbirke, saj nova inačica INSTRUCT-a obsega kar 20.280 zapisov iz baze LISA za leta 1982-1985. Podobnost z originalno verzijo pa je vidna v implementaciji invertirane datoteke.

3.1 Uporabniški vmesnik

Originalno verzijo INSTRUCT-a sestavljata dva osnovna vmesnika, namenjena komuniciranju z uporabnikom: vmesnik, prilagojen neizkušenemu uporabniku (pomoč je zagotovljena v obliki pojasnjevalnega teksta) in vmesnik, namenjen uporabniku, ki že obvlada delovanje INSTRUCT-a (omejena količina razlagalnega teksta). Tema vmesnikoma je v novi verziji dodan še t.i. ekspertni vmesnik, namenjen tistim uporabnikom, ki natančno poznajo značilnosti in zmogljivosti INSTRUCT-a. S pomočjo tega modula naj bi se INSTRUCT uporabljal predvsem za iskanje referenc in ne kot orodje za učenje sodobnih metod iskanja informacij.

3.2 Razširitev iskalnih izrazov na temelju skupne navzočnosti v bazi

Poleg razširitvenega modula, ki temelji na metodi podobnosti niza (tj.čskupno število trigramov), vsebuje nova inačica INSTRUCT-a tudi razširitev izrazov s pomočjo metode razvrščanja v skupine. Identifikacija najbolj podobnih besed temelji na pogostosti skupnega pojavljanja iskalnega izraza in določenega korena besede v podatkovni zbirki.

Algoritem za izračun te podobnosti je razvil Willett (1981). Čeprav je to trenutno eden najučinkovitejših algoritmov, pa je zanj še vedno značilna prevelika uporaba računalniških kapacitet. Na srečo je podatkovna zbirka v INSTRUCT-u statične narave in omogoča paketno procesiranje 20-tih najbližjih izrazov za vsak krn v slovarski datoteki. Ti seznami predstavljajo tudi osnovo za dodatno vključitev posameznih izrazov v uporabnikovo poizvedbo.

3.3 Iskanje po načelu razvrščanja v skupine

Poleg Boolovega iskanja in iskanja po načelu optimalnega primerjanja vsebuje nova verzija INSTRUCT-a tudi metode iskanja, ki izhajajo iz koncepta razvrščanja dokumentov v skupine ("clustering"). Ta način iskanja temelji na t.i. algoritmu "najbližje skupine", ki ga je natančno opisal Griffiths s svojimi sodelavci (1984).

3.4 Dodatne možnosti iskanja informacij

Ob osnovnih treh vrstah iskanja dokumentov omogoča INSTRUCT tudi dodatno izboljšanje kvalitete zadetkov z implementacijo naslednjih modulov:

- hibridno iskanje (tj.čuporaba Boolovih logičnih operatorjev pri vplivanju na rezultate iskanja po načelu optimalnega primerjanja in razvrščanja v skupine)
- prelistavanje
- povratno iskanje

3.4.1 Hibridno iskanje

Nova inačica INSTRUCT-a vsebuje možnost odstranitve vseh tistih dokumentov iz rangiranega seznama - oblikovanega na osnovi iskanja po načelu optimalnega primerjanja ali razvrščanja v skupine - ki ne ustrezajo zahtevam Boolovih logičnih operatorjev. Ko ima uporabnik na voljo seznam rangiranih dokumentov, lahko ekperimentira z različnimi Boolovimi nizi.

3.2.2 Prelistavanje

Prelistavanje ("browsing") dobiva v zadnjem času vedno večjo vlogo tudi na področju razvijanja sistemov za iskanje informacij (Hildreth, 1982; Bawden, 1986). S pomočjo prelistavanja lahko uporabnik spremlja določen dokument v podatkovni zbirki, ne da bi pri tem izgubil "rdečo nit" poizvedbe. Prelistavanje ima še poseben pomen pri odkrivanju t.i. "mejnih" dokumentov, ki so lahko popolnoma nepričakovano relevantni za poizvedbo (Wade in Willett, 1988). Možnost prelistavanja je vključena tudi v novo verzijo INSTRUCT-a, kjer se lahko uporabnik na osnovi enega ali več relevantnih dokumentov odloči za:

- verižno iskanje (chain search)
- sorodstveno iskanje (seed search)

Verižno iskanje temelji na spremljanju verige dokumentov, ki so med seboj tesno vsebinsko povezani. Ta povezava se oblikuje s pomočjo algoritma "najbližje skupine". Ko uporabnik izbere določen dokument kot izhodišče za prelistavanje, se na zaslonu prikaže temu dokumentu vsebinsko najbližji dokument. To verižno iskanje poteka vse dotlej, dokler se veriga dokumentov ne zaključi (Murtagh, 1983).

Sorodstveno iskanje pa temelji na iskanju po načelu optimalnega primerjanja, kjer se iskalni izrazi iz poizvedbe nadomestijo z besedami iz dokumenta, ki ga uporabnik izbere kot osnovo za prelistavanje. Rezultat tega iskanja je potemtakem rangirani seznam dokumentov glede na njihovo podobnost z izbranim dokumentom (t.i. "seed"). Zaradi potrebe po hitrem odzivnem času, uporablja INSTRUCT le 25 izrazov iz vsakega dokumenta. To so tisti izrazi, ki se pojavljajo z najnižjo frekvenco; s tem se zagotavlja preciznost pri iskanju.

3.4.3 Povratno iskanje

Povratna informacija o relevantnih dokumentih, ki jo uporabnik posreduje po zaključku začetnega iskanja ali prelistavanja, služi INSTRUCT-u za modifikacijo poizvedbe. Spreminjanje originalne poizvedbe dosežemo na dva načina:

- s spreminjanjem uteži posameznim besednim krnom,
- z razširitvijo poizvedbe z novimi besedami iz relevantnih dokumentov.

Prva opcija temelji na določanju uteži posameznim iskalnim izrazom v poizvedbi; ta opcija je opisana v sekciji o relevantnem povratnem iskanju (glej tudi Wade, Willett, 1988). Vsekakor je zanimiva tudi druga opcija, kjer pa ne določamo uteži iskalnim izrazom, temveč vsem besedam iz relevantnih dokumentov. Na tej osnovi dobimo rangirani seznam novih iskalnih izrazov, ki jih potem uporabnik lahko vključi v svojo poizvedbo.

Iz zgoraj opisanih sekcij lahko torej povzamemo, da INSTRUCT vsebuje naslednje module, ki demonstrirajo uporabo sodobnih tehnik in metod iskanja informacij v tekstovnih podatkovnih zbirkah:

- Uporabniški vmesnik:
  - ☐ začetnik
  - ☐ izkušeni uporabnik
  - ☐ ekspert
- Oblikovanje poizvedbe:
  - ☐ vnos poizvedbe v naravnem jeziku
  - ☐ blokiranje besed brez predmetne vsebine
  - ☐ avtomatsko krnjenje besed
  - ☐ določanje začetnih uteži iskalnim izrazom
- Preoblikovanje poizvedbe:
  - ☐ dodajanje iskalnih izrazov
  - ☐ izločanje iskalnih izrazov
  - ☐ razširitev iskalnih izrazov:
    - koncept podobnosti niza (t.i. trigrami)
    - koncept skupne navzočnosti v bazi
- Iskanje
  - ☐ iskanje po načelu optimalnega primerjanja
  - ☐ Boolovo iskanje
  - ☐ iskanje po načelu razvrščanja v skupine
- Dodatne možnosti iskanja:
  - ☐ hibridno iskanje (z uporabo Boolovih operatorjev)
  - ☐ prelistavanje:
    - verižno iskanje
    - rastoče iskanje
  - ☐ povratno iskanje:
    - spreminjanje uteži iskalnim izrazom
    - določanje uteži besedam iz relevantnih dokumentov

Poudaril sem že, da se INSTRUCT intenzivno uporablja kot učni pripomoček na številnih šolah bibliotekarstva in informacijskih ved v svetu. Hkrati pa postaja INSTRUCT tudi pomembno raziskovalno orodje za proučevanje problemov na področju iskanja informacij. V naslednjem sklopu bomo opisali, kako INSTRUCT

opravlja ti dve funkciji na Oddelku za informacijske vede Univerze v Sheffieldu, Velika Britanija.

4. Uporaba INSTRUCT-a na Univerzi v Sheffieldu

4.1 INSTRUCT kot učni pripomoček

Uporaba INSTRUCT-a v učnem programu Oddelka za informacijske vede Univerze v Sheffieldu je še posebej pomembna v okviru predmeta Shranjevanje in iskanje informacij. Ker ta predmet vsebuje številne praktične vaje, si je oddelek vedno prizadeval za razvijanje različnih simulacijskih modelov za online iskanje informacij. S pomočjo teh modelov (npr.čDIALOG-SOS) so si študenti pridobili koristne izkušnje, ki so jih potem s pridom uporabili pri resničnem online iskanju podatkov. Osnovna slabost teh modelov je v tem, da sicer omogočajo demonstracijo konvencionalnih, Boolovih metod iskanja podatkov, ne pa tudi testiranja sodobnih, alternativnih trendov.

Ravno zaradi tega ob spodbudnih rezultatih raziskovanja na področju sodobnih metod iskanja informacij se je oddelek odločil za razvoj informacijskega sistema INSTRUCT, ki naj bi bil v prvi vrsti namenjen izobraževanju študentov. Osnovni cilji INSTRUCT-a so bili potemtakem naslednji:

- omogočiti študentom seznanitev s sodobnimi, nekonvencionalnimi metodami iskanja informacij,
- omogočiti oblikovanje poizvedb in iskanje v podatkovni zbirki z zadostnim številom dokumentov,
- omogočiti iskanje dokumentov v podatkovni zbirki LISA, tj.čpodatkovni zbirki, ki je vsebinsko zelo blizu študentom knjižničarstva in informacijskih ved,
- omogočiti spoznavanje razlik med Boolovim in nekonvencionalnim pristopom do iskanja informacij.

4.2 INSTRUCT v raziskovalnem programu

Kljub temu, da je INSTRUCT nastal v prvi vrsti kot učni pripomoček, pa se v zadnjem obdobju intenzivira njegova vloga v različnih raziskovalnih projektih. Nekateri izmed teh projektov so opisani v nadaljevanju.

4.3.1 Boolovo iskanje in iskanje po načelu optimalnega primerjanja primerjalna analiza

Prvi raziskovalni eksperiment, ki je nastal na osnovi uporabe INSTRUCT-a, je obsegal primerjalno evalvacijo dveh vrst iskanja informacij, tj.čBoolovo iskanje in iskanje po načelu optimalnega primerjanja. Čeprav so bile v preteklosti opravljene številne primerjalne analize teh dveh vrst iskanja informacij, pa je bila njihova osnovna slabost v tem, da so potekale v t.i. laboratorijskih pogojih (tj.čuporaba standardnih zbirk dokumentov z vnaprej pripravljenim nizom poizvedb in relevantih dokumentov). Prednost testiranja s pomočjo INSTRUCT-a pa so bili rezultati, ki so nastali na osnovi resničnih uporabnikov in njihovih dejanskih informacijskih potreb (tj. študenti, ki so potrebovali reference za svoje magistrske disertacije). Evalvacija teh dveh vrst iskanja informacij je potekala tako na kvalitativni kot tudi na kvantitativni ravni.

Kvalitativna evalvacija, ki je potekala leta 1985, je temeljila na osebnih mnenjih študentov o razlikah med Boolovim iskanjem in iskanjem po načelu optimalnega primerjanja v okviru INSTRUCT-a. Študenti so bili predvsem navdušeni nad preprostim načinom iskanja, ki ga prinaša metoda optimalnega primerjanja. Kljub temu pa so pokazali dokaj veliko stopnjo nezaupanja v to metodo, kajti velik del iskanja poteka mimo kakršne koli kontrole s strani uporabnika.

Kvantitativna primerjava obeh načinov iskanja dokumentov s pomočjo INSTRUCT-a je potekala v šolskih letih 1986/87 in 1987/88. Rezultati te primerjave so pokazali, da sta oba načina iskanja dokumentov približno enako uspešna, saj sta več ali manj priklicala podobne nize dokumentov. Prednost metode optimalnega primerjanja pa je seveda v tem, da zahteva manj znanja in truda s strani uporabnika.

4.2.2 Primerjava med ekspertnim in statističnim pristopom do iskanja informacij

INSTRUCT je tipičen primer informacijskega sistema, ki uporablja statistične metode in tehnike v procesu iskanja dokumentov. Zanimivo pa je poudariti, da ob teh sistemih nastajajo v svetu tudi številni t.i. ekspertni posredniški sistemi, ki uporabljajo metode iz arzenala umetne inteligence. Ti sistemi naj bi zagotovili predvsem uporabniško prijazen in razumljiv dostop do bibliografskih podatkovnih zbirk, ki temeljijo na Boolovem iskanju informacij. Eden izmed predstavnikov teh sistemov se imenuje PLEXUS in ga razvija Informacijski center Univerze v Londonu (Vickery et al., 1987).

Tako PLEXUS kot INSTRUCT opravljata naslednjo temeljno funkcijo: zmanjšati težave in ovire, ki jih povzroča Boolov način iskanja informacij. Medtem ko

PLEXUS teži k doseganju tega cilja s kodiranjem ekspertize informacijskega posrednika, pa delovanje INSTRUCT-a temelji na preprostem vnosu poizvedbe v naravnem jeziku ter na uporabi statističnih iskalnih metod. Čeprav se oba pristopa temeljito razlikujeta med seboj, pa je njun skupni cilj omogočiti uporabniku izvedbo kvalitetne poizvedbe brez kakršnegakoli vmešavanja informacijskega posrednika.

Primerjalna evalvacija učinkovitosti obeh sistemov je potekala na osnovi raziskovalnega projekta, ki so ga opravili Wade in sodelavci (1989). Zanimivo je poudariti, da so bili najboljši rezultati doseženi takrat, ko se je PLEXUS uporabljal za dokončno oblikovanje poizvedbe, katero je potem procesiral INSTRUCT z uporabo metod optimalnega primerjanja. Ti rezultati so bili spodbuda za nov projekt, ki je trenutno v teku na Univerzi v Sheffieldu in proučuje možnosti integracije obeh pristopov.

4.2.3 Iskanje po načelu optimalnega primerjanja s pomočjo tekstovnih signatur

INSTRUCT tako kot velika večina tekstovnih informacijskih sistemov temelji na uporabi invertirane zbirke. Čeprav omogoča invertirana zbirka izredno hitro iskanje in jo je enostavno implementirati v velikih bazah podatkov, pa sta njena glavna problema velike pomnilniške zahteve in stroški ažuriranja. Vključevanje sekvenčne zbirke kot alternative invertirani zbirki doslej ni bilo resno proučevano, in sicer predvsem zaradi počasnega procesiranja. Kljub temu pa je v zadnjem času raziskovalno delo na področju tekstovnih signatur (glej npr. Faloutsos, 1985) spodbudilo interes tudi za implementacijo sekvenčne organizacije zbirk.

Tekstovno signaturo lahko opišemo kot bitno predstavitev niza, kjer definiramo posamezne bite, če obstajajo določeni nizi (npr. besedni krni) v tekstu (Wade et al., 1989). Bitne nize oblikujemo s pomočjo algoritma za razprševanje. Niz signatur, ki označujejo zbirko dokumentov, lahko potemtakem definiramo kot razpored bitov. Zanimivo je, da je takšen pristop do iskanja informacij učinkovit predvsem na področju iskanja kemijskih struktur (Willett et al., 1986)

S pomočjo programskega sistema SIBRIS (Sandwich Interactive Browsing and Ranking Information System) Wade et al., 1989 poteka trenutno na Univerzi v Sheffieldu eksperiment ocenjevanja učinkovitosti metode optimalnega rangiranja s pomočjo strukture razporejanja bitov. Ta sistem prav tako omogoča vnos poizvedbe v naravnem jeziku ter rangiranje zadetkov. Prvi rezultati testiranja so pokazali, da so uporabniki s precej velikim navdušenjem sprejeli ta sistem.

4.2.4 Iskanje s pomočjo INSTRUCT-a po polnem tekstu

INSTRUCT je bil v začetni fazi oblikovan kot tipični predstavnik bibliografskih sistemov za iskanje informacij. Ali z drugimi besedami: s pomočjo INSTRUCT-a je uporabnik našel le reference o določenem dokumentu (avtor, naslov, itd.).

Vedno večja uporaba elektronskega publiciranja pa danes v veliki meri povečuje število dokumentov, ki so dosegljivi v strojno čitljivi obliki (Tenopir, 1985). Zanimivo je, da večina sistemov za iskanje informacij iz teh dokumentov zopet temelji na klasičnih in konvencionalnih Boolovih metodah. Ravno zaradi tega je toliko bolj privlačen projekt Al-Hawamdeha in Willetta (1989), ki proučujeta, v kolikšni meri lahko metode in tehnike optimalnega primerjanja apliciramo v procesu iskanja po polnem tekstu (full- text searching).

Izhodiše tega projekta je predpostavka, da je uporabnik že našel dokumente, ki jih potrebuje pri svojem delu. S pomočjo metod optimalnega primerjanja želi najti le tiste segmente v dokumentih, ki ga najbolj zanimajo oz. so najbolj relevantni. Če odstavke v posameznih tekstih definiramo kot osnovne enote zapisa (namesto povzetkov), lahko s pomočjo INSTRUCT-a dosežemo rangiranje odstavkov glede na njihovo podobnost s poizvedbo. Prva eksperimentalna testiranja tega pristopa so prinesla izredno spodbudne rezultate, kajti večina relevantih odstavkov je bila razvrščenih v enotno skupino v samem vrhu rangirnega seznama (Al- Hawamdeh, 1989).

Možnost rangiranja odstavkov je še posebej privlačna opcija za implementacijo metod prelistavanja v prostem tekstu (Wood, Willett, 1989). Na tej osnovi bi lahko v prihodnje nastal alternativni koncept sedanjemu modelu iskanja informacij, ki temelji na t.i. hypertextu.

5. Procesiranje dokumentov in poizvedb v slovenskem jeziku

Številne module, ki sestavljajo INSTRUCT, lahko označimo kot jezikovno neodvisne module. Edini izjemi sta postopka za blokiranje besed in avtomatsko krnjenje besed. Če temu dodamo dejstvo, da je bil INSTRUCT napisan v standardnem programskem jeziku PASCAL, potem sledi, da je dokaj enostavno prilagoditi INSTRUCT kateremukoli jeziku. Izdelava slovenskega modela iskanja informacij po prostem tekstu je potemtakem zahtevala dvoje:

- oblikovanje seznama blokiranih besed in izdelava algoritma za avtomatsko krnjenje slovenskih besed
- prikrojitev posameznih modulov v INSTRUCT-u za procesiranje slovenskega teksta.

Delo na oblikovanju seznama blokiranih besed in razvoj slovenskega algoritma za krnjenje besed sta natančno opisana v reviji *Literary and Linguistic Computing* (Popovič, Willett, 1990). Zato bom v nadaljevanju opisal temeljne modifikacije INSTRUCT-a, ki so bile potrebne za izdelavo slovenskega modela.

5.1 Slovenska verzija INSTRUCT-a

Osnovni pomen oblikovanja seznama slovenskih blokiranih besed in razvoja algoritma za krnjenje besed v slovenščini je viden v njuni kompatibilnosti z metodami optimalnega primerjanja; s tem naj bi imel uporabnik tudi v slovenskem prostoru omogočen enostaven in učinkovit dostop do bibliografskih podatkovnih zbirk.

Zato je za implementacijo slovenske verzije INSTRUCT-a zadostovala njegova originalna inačica, tj.čverzija, oblikovana za PRIME računalnik. Ta verzija vsebuje enega najpomembnejših modulov, tj. iskanje po načelu optimalnega primerjanja in rangiranje zadetkov. Odločitev za preoblikovanje originalne verzije je narekovalo tudi dejstvo, da je slovenski model nastajal na mikroračunalniku PC AT. Vključitev ostalih modulov iz nove verzije INSTRUCT-a bi nedvomno vodila k pomnilniškim problemom in bi potemtakem zahtevala popolno preoblikovanje celotnega programa.

Slovenska verzija INSTRUCT-a, napisana v programskem jeziku TURBO PASCAL 5.5, vsebuje naslednje module:

- vnos poizvedbe v naravnem jeziku (v slovenščini)
- izločanje besed brez predmetne vsebine iz poizvedbe (seznam 1.593 slovenskih blokiranih besed kot so npr.: IN, PA, TER, ALI)
- avtomatsko krnjenje besed (algoritem, ki ga je razvil Popovič; glej Popovič, Willett, 1990)
- razširitev iskalnih izrazov s pomočjo metode podobnosti nizov
- iskanje po načelu optimalnega primerjanja (z možnostjo dodatne vključitve Boolovih operatorjev)
- relevantno povratno iskanje
- Boolovo iskanje

Kot sem že poudaril, je originalna verzija INSTRUCT-a doživela glavne spremembe v dveh jezikovno odvisnih modulih (tj.čblokiranje in krnjenje besed). S tem sem želel zagotoviti predvsem učinkovito procesiranje dokumentov in poizvedb v slovenskem jeziku. Ker je prva verzija INSTRUCT-a nastala na velikem računalniku, bi bilo potrebno izboljšati številne ostale komponente npr. atraktivnejši uporabniški vmesnik itd. Ker se ta projekt ni ukvarjal s tovrstnimi dodatnimi

problemi, mora vsaka kvalitativna evalvacija INSTRUCT-a upoštevati tudi te pomanjkljivosti.

5.1.1 Oblikovanje in procesiranje zbirke slovenskih dokumentov

Poudaril sem že, da je temelj prve verzije INSTRUCT-a predstavljala podatkovna zbirka LISA iz leta 1982, ki jo sestavljajo 6.004 zapisi. Predpogoj oblikovanja slovenske verzije INSTRUCT-a je bila zagotovitev ekvivalentne podatkovne zbirke, sestavljene iz zapisov v slovenskem jeziku. Temeljna komponenta teh zapisov pa so vsekakor povzetki.

Glede na intenziven razvoj slovenskega sistema znanstvenega in tehničnega informiranja, ki obsega gradnjo številnih specializiranih podatkovnih zbirk, je bila za projekt načrtovana uporaba ene izmed teh zbirk. Ugotovitev, da med številnimi specialnimi podatkovnimi zbirkami niti ena ne vsebuje bibliografskih zapisov, ki vključujejo tudi povzetke v slovenskem jeziku, je bila dokaj hud udarec za nadaljevanje projekta. Edina izjema je bila zbirka biografij in bibliografij visokošolskih učiteljev in sodelavcev Univerze v Mariboru, ki pa zaradi pokrivanja različnih strokovnih področij (in tudi kasnejše evalvacije) ni bila primerna za vključitev v INSTRUCT.

Zato je bila za nadaljnji obstoj projekta edina rešitev gradnja nove podatkovne zbirke, ki vsebuje zapise tako z osnovnimi bibliografskimi podatki kot tudi s povzetki. Za gradnjo te podatkovne zbirke sem uporabil članke iz revij *Knjižnica* (1972- 1990) in *Informatologia Yugoslavica* (1969-1989). Povzetke iz *Informatologie Yugoslavice* je bilo seveda potrebno prevesti v slovenski jezik. Ta podatkovna zbirka obsega trenutno 504 enote, kar je zadostna velikost tako za implementacijo INSTRUCT-a kot tudi za eksperimentalno testiranje kvalitete posameznih modulov v INSTRUCT-u.

Za procesiranje te podatkovne zbirke sem oblikoval nekaj samostojnih podprogramov. Rezultat tega procesiranja je bila izgradnja invertirane zbirke.

5.1.2 Modifikacija izvirne kode INSTRUCT-a

Originalna verzija INSTRUCT-a je napisana v modularni obliki s pomočjo standardnega programskega jezika PASCAL. S tem sta bili omogočeni tako prenosljivost INSTRUCT-a na druge sisteme kot tudi enostavnost pri bodočih spremembah in dograditvah. Edino izjemo predstavljajo diskovni postopki, ki so seveda neizogibno povezani z določenim operacijskim sistemom.

Zaradi tega je bilo dokaj preprosto prenesti izvirno kodo INSTRUCT-a na mikroračunalnik IBM PC (operacijski sistem MS-DOS) ter dograditi oz. spremeniti

nekatere module s pomočjo programskega jezika TURBO PASCAL 5.5. Ta proces je obsegal naslednje operacije:

- sprememba diskovno odvisnih postopkov (tj. prilagoditev postopkov za odpiranje in zapiranje zbirk v MS-DOS),
- razdelitev obsežne izvirne kode INSTRUCT-a v manjše enote, kar je pogojeno z uporabo programskega jezika TURBO PASCAL,
- aplikacija variable STRING v celotnem programu.

Ob tem so bile potrebne tudi nekatere dodatne modifikacije modulov. Globalne spremembe sta doživela predvsem modula, ki sta odgovorna za avtomatsko blokiranje in krnjenje besed. Kot sem že poudaril, temelji slovenski algoritem za krnjenje besed (Popovič, Willett, 1990) na seznamu 2.093 končnic, katerim je dodanih tudi osem kontekstno odvisnih pravil. Postopek odbijanja končnic deluje na principu najdaljšega primerjanja ter dodatne kontrole morfoloških značilnosti slovenskega jezika. To pomeni, da se slovenski algoritem krepko razlikuje od angleškega algoritma, ki ga je razvil Porter (1980) in temelji na principu iteracije.

Slovenski algoritem zahteva potencialno mnogo večje računalniške kapacitete kot Porterjev algoritem, saj krnjenje besed poteka v dveh korakih na osnovi obsežnega seznama končajev. Vendar pa je prva evalvacija tega algoritma znotraj INSTRUCT-a potrdila, da nam je uspelo doseči ravnotežje med kvaliteto algoritma in preprostostjo ter učinkovitostjo procesiranja. Še več, primerjalna evalvacija na primeru procesiranja 48 poizvedb v slovenskem jeziku (glej Popovič, 1990) je pokazala, da s pomočjo avtomatskega krnjenja dobimo popolnoma enakovredne rezultate "ročnemu" desnemu odrezu.

6. Ilustracija iskanja po načelu optimalnega primerjanja s pomočjo slovenske verzije programskega paketa INSTRUCT

Pred zaključkom želim na konkretnem primeru ilustrirati delovanje slovenskega algoritma za odbijanje končajev ter prikazati nekatere druge zmogljivosti INSTRUCT-a. Za ilustracijo bomo uporabili poizvedbo VISOKOŠOLSKO IZOBRAŽEVANJE IN BIBLIOTEKARSKI KADRI. Dialog uporabnika z INSTRUCT-om je označen z ležečimi tiskanimi črkami.

> *INSTRUCT*

* * * * INSTRUCT * * * *
(An INteractive System for Teaching Retrieval Using Computional Techniques)

Če ste:

- seznanjeni z INSTRUCT-ovim načinom iskanja, vtipkajte : 1

- neizkušeni in želite dodatna pojasnila tekom iskanja, vtipkajte : 2

Prosim za odgovor:
> *1*

Vtipkajte poizvedbo in jo zaključite z znakom #:
> *Visokošolsko izobraževanje in bibliotekarski kadri #*

ISKANJE SE JE ZAČELO
prosim za kanček potrpljenja

Trenutni seznam iskalnih izrazov:

1 VISOKOŠOL (18)	2 IZOBRAZ (73)
3 BIBLIOT (66)	4 KADER (24)

Če želite:
- dodati nove iskalne izraze, vtipkajte : 1
- zbrisati nekatere izraze, vtipkajte : 2
- najti sorodne izraze, vtipkajte : 3
- iskati po načelu OPTIMALNEGA PRIMERJANJA, vtipkajte : 4
- aplicirati BOOLOVO iskanje, vtipkajte : 5
- zapustiti INSTRUCT, vtipkajte : 6

Prosim za odgovor:
> *3*

ISKANJE PODOBNIH IZRAZOV:

Sedaj lahko razširite iskalni izraz
- Če izberete enega izmed krnov v poizvedbi, boste dobili 10 novih izrazov, ki so sestavljeni iz podobnih znakovnih nizov
- Uporabne izraze lahko dodate svoji poizvedbi

Sledi seznam krnov iz vaše poizvedbe ...

1 VISOKOŠOL (18)	2 IZOBRAZ (73)
3 BIBLIOT (66)	4 KADER (24)

Če želite iz zgornjega seznama razširiti kakšen izraz, vtipkajte njegovo zaporedno številko in pritisnite ENTER; v nasprotnem primeru pa vtipkajte le # in zaključite s tipko ENTER
> *2*

ISKANJE SE JE ZAČELO
prosim za kanček potrpljenja

Iskali ste besedo IZOBRAZ
in sistem je našel naslednje termine:

1 PREDIZOBRAZ (1)	2 SAMOIZOBRAZ (1)
3 SPLOŠNOIZOBRAZ (36)	4 VISOKOIZOBRAZ (1)
5 OBRAZ (5)	6 PREOBRAZ (4)
7 IZRAZ (6)	8 FRAZ (1)
9 IZKAZ (1)	10 OBRAZLOŽ (1)

Če želite vključiti katere od novih terminov v vašo poizvedbo, vtipkajte njihove zaporedne številke -- zaključite z # in ENTER, npr. 1 2 3 #.
V nasprotnem primeru pa vtipkajte le # in pritisnite tipko ENTER
> *4#*

Trenutni seznam iskalnih izrazov:

1 VISOKOŠOL (18)	2 IZOBRAZ (73)
3 BIBLIOT (66)	4 KADER (24)
5 VISOKOIZOBRAZ (1)	

Če želite razširiti iskalni izraz, vtipkajte njegovo ŠTEVILKO in zaključite s tipko ENTER.
Če želite zapustiti to sekcijo in se vrniti v glavni MENU, vtipkajte znak # in pritisnite na tipko ENTER.
> *#*

Če želite:

- dodati nove iskalne izraze, vtipkajte : 1
- zbrisati nekatere izraze, vtipkajte : 2
- najti sorodne izraze, vtipkajte : 3
- iskati po načelu OPTIMALNEGA PRIMERJANJA, vtipkajte : 4

- aplicirati BOOLOVO iskanje, vtipkajte : 5
- zapustiti INSTRUCT, vtipkajte : 6

Prosim za odgovor:
> *4*

ISKANJE PO NAČELU OPTIMALNEGA PRIMERJANJA

ISKANJE SE JE ZAČELO
prosim za kanček potrpljenja

Rezultat iskanja je 142 zadetkov, ki imajo najmanj en skupni termin s poizvedbo.

Ali si pri rangiranju zadetkov želite pomagati z Boolovimi operatorji? (Y/N)
> *N*

Koliko dokumentov želite videti?
S pritiskom na tipko ENTER boste videli 5 dokumentov
> *<ENTER>*

Če želite videti:
- BIBLIOGRAFSKE PODATKE, vtipkajte : 1
- BIBLIOGRAFSKE PODATKE in POVZETEK, vtipkajte : 2

> *1*

Če se v katerikoli fazi odločite, da želite zaključiti pregledovanje dokumentov, vtipkajte # in pritisnite na tipko ENTER.

1/163
Kadri in znanje: Primer splošnoizobraževalnih knjižnic v Ljubljani
* Kolenc, J.: Knjižnica, 32(1988)1/2, str. 23-45

Ali je ta dokument relevanten za vašo poizvedbo?(Y/N)
Če želite prej videti še abstrakt, vtipkajte 'A'
> *A*

1/163
Kadri in znanje: Primer splošnoizobraževalnih knjižnic v Ljubljani
* Kolenc, J.: Knjižnica, 32(1988)1/2, str. 23-45

Poleg ustreznih prostorov in sodobne opreme so kakovostni kadri (strokovni delavci) bistveni pogoj za učinkovito izvajanje knjižnično-informacijske dejavnosti. Tudi KIS v SRS in SFRJ namenjena temu vprašanju poseben pomen. Kadrovska funkcija v splošnoizobraževalnih knjižnicah pa ni le potentia temveč tudi agentia (agens) vsakršnega razvoja. Zahtevani prehod v informacijsko družbo postavlja tudi pred nas imperativ: profesionalizacija! Znanje se v knjižničarstvu lahko pretaka le skozi visokoizobražen in kvalificiran kader.Razvoj višjih oblik delitve dela je pogoj za dvig družbenega ugleda (statusa) knjižničarskega poklica.

Ali je ta dokument relevanten za vašo poizvedbo? (Y/N)
> *Y*

2/155
Izobraževanje knjižničarskih delavcev v Jugoslaviji
* Berčič, B.: Knjižnica, 31(1987)1, str. 110-129

Ali je ta dokument relevanten za vašo poizvedbo? (Y/N)
Če želite prej videti še abstrakt, vtipkajte 'A'
> *Y*

3/185
Začetek visokošolskega študija bibliotekarstva v Sloveniji
* Berčič, B.: Strokovni posvet ZBDS, Bled 1987, 13 str.

Ali je ta dokument relevanten za vašo poizvedbo? (Y/N)
Če želite prej videti še abstrakt, vtipkajte 'A'
> *Y*

4/79
Kako je pri nas z razvojem bibliotekarske vede?
* Sepe, M.: Knjižnica, 26(1982)1/2, str. 76-80

Ali je ta dokument relevanten za vašo poizvedbo? (Y/N)
Če želite prej videti še abstrakt, vtipkajte 'A'
> *A*

4/79
Kako je pri nas z razvojem bibliotekarske vede?
* Sepe, M.: Knjižnica, 26(1982)1/2, str. 76-80

Razvoj knjižnične stroke naj bi bil preplet treh dejavnikov: Narodne in univerzitetne knjižnice, ki s svojo matično službo kot organizatorjem dejavnosti, z raziskovalnim centrom kot središčem raziskovalnega dela in ob sodelovanju strokovnjakov z drugih specialnih področij bibliotekarske stroke predstavlja nosilca razvoja bibliotekarstva (to NUK-u nalaga tudi Zakon o knjižničarstvu). Skrb za preučevanje in razvoj posameznih strokovnih področij naj še naprej ostane pri Društvu bibliotekarjev. Tretji dejavnik, ki prispeva k razvoju stroke, pa naj bi bile tudi knjižnice same, predvsem s strokovnim izobraževanjem svojih kadrov.

Ali je ta dokument relevanten za vašo poizvedbo? (Y/N)
> *Y*

5/204
Trženje v visokošolski knjižnici
* Češnovar, N.: Knjižnica, 33(1989)3/4, str. 123-128

Ali je ta dokument relevanten za vašo poizvedbo? (Y/N)
Če želite prej videti še abstrakt, vtipkajte 'A'
> *A*

5/204
Trženje v visokošolski knjižnici
* Češnovar, N.: Knjižnica, 33(1989)3/4, str. 123-128

Marketing oziroma trženje je proces upravljanja, ki ga enačimo s pričakovanji, zahtevami in zadovoljitvijo strank, pri tem pa imamo dobiček. Knjižnica, ki se tržno obnaša, mora: zagotoviti hiter dostop do svojega gradiva; izvajati medknjižnično izposojo, kreirati informacijske zbirke, ki še ne obstajajo; slediti mora modernemu razvoju informacijskih služb. Da lahko sledi tem zahtevam, morajo biti kadri v visokošolskih knjižnicah primerno usposobljeni, da lahko ugotovijo in strokovno presodijo, kaj uporabnik potrebuje. Na tej stopnji pa informacijsko delo že začne preraščati v raziskovalno dejavnost.

Ali je ta dokument relevanten za vašo poizvedbo? (Y/N)
> *N*

Če na tej točki iskanja uporabnik meni, da ni prejel zadostnega števila relevantnih dokumentov, se lahko odloči med naslednjimi možnostmi: dodatno pregledovanje rangiranih dokumentov, povratno relevantno iskanje ali Boolovo iskanje.

7. Zaključek

V tem članku sem opisal programski paket INSTRUCT, katerega glavna funkcija je demonstracija sodobnih metod iskanja informacij. Ker so v dosedanjih eksperimentih te metode prinesle izredno kvalitetne rezultate, najdemo danes vedno večje število online informacijskih sistemov, ki prevzemajo nekonvencionalni pristop do iskanja bibliografskih informacij.

Čeprav je bil INSTRUCT najprej oblikovan kot učni pripomoček pri študiju, pa danes postaja izredno pomembno orodje za raziskovanje številnih problemov na področju iskanja informacij. Tako je v okviru INSTRUCT-a nastal tudi slovenski algoritem za avtomatsko krnjenje besed, kar je pomemben prispevek v procesu uvajanja nekonvencionalnih metod iskanja informacij v slovenski prostor. Če hočemo v prihodnje še spodbuditi te procese, pa bi kazalo uresničiti naslednje predloge:

- uporaba slovenske verzije INSTRUCT-a v okviru visokošolskega študija bibliotekarstva na Filozofski fakulteti v Ljubljani,
- namestitev INSTRUCT-a kot demonstracijskega paketa na omrežje; s tem bi postal dosegljiv za večje število uporabnikov,
- vgraditev nekaterih modulov INSTRUCT-a v obstoječe, konvencionalne informacijske sisteme v Sloveniji.

Ker je evalvacija slovenske verzije INSTRUCT-a (opis evalvacije sledi v eni izmed prihodnjih številk revije Knjižnica) prinesla zelo kvalitetne rezultate, dejansko ni večjih ovir, da ne bi začeli implementirati sodobnih metod iskanja informacij tudi v Sloveniji.

Zahvala

Zahvaljujem se Narodni in univerzitetni knjižnici v Ljubljani, Republiškemu sekretariatu za kulturo, Republiškemu sekretariatu za raziskovalno dejavnost in tehnologijo ter Britanskem svetu za njihovo finančno podporo pri doktorskem študiju na Univerzi v Sheffieldu, Velika Britanija. Zahvaliti se želim tudi svojemu mentorju, dr. Petru Willettu, ki mi nudi dragoceno pomoč in nasvete pri raziskovalnem delu in pisanju doktorske disertacije.

Literatura:

AL-HAWAMDEH, S.: Paragraph-based Retrieval in Full Text Documents. PhD thesis. - Sheffield : University of Sheffield, 1989.

AL-HAWAMDEH, S., WILLETT, P.: Paragraph-based nearest neighbour searching in full-text documents. - Electronic Publishing, 4(1989), str. 14-25.

BAWDEN, D.: Information systems and the stimulation of creativity. - Journal of Information Science, 12(1986), str. 203- 216.

BRZOZOWSKI, J.P.: MASQUERADE: searching the full text of abstracts using automatic indexing. - Journal of Information Science, 6(1983), str. 67-73.

CRINGEAN, J.: Personal Librarian at the online exhibition. - IT Link, 7(1989), str. 9.

CROFT, W.B., HARPER, D.J.: Using probabilistic models of document retrieval without relevance information. - Journal of Documentation, 35(1979), str. 285-295.

DOSZKOCS, T.E.: CITE NLM: natural language searching in an online catalog. - Information Technology and Libraries, 2(1983), str. 364-380.

FALOUTSOS, C.: Access methods for text. - Computing Surveys 17(1985), str. 49-74.

GRIFFITHS, A. et al: Hierarchic agglomerative clustering methods for automatic document classification. - Journal of Documentation, 40(1984), str. 175-205.

HILDRETH, C.R.: Online browsing support capabilities. - Proceedings of the ASIS Annual Meeting 19. White Plains, New York : Knowledge Industry Publications Inc., 1982, str. 127-132.

MITEV, N. et al: Designing an online public access catalogue: OKAPI, a catalogue on a local area network. - London : British Library Research and Development Department, 1985.

MURTAGH, F.: A survey of recent advances in hierarchical clustering algorithms. - The Computer Journal, 26(1983), str. 354-359.

NOREAULT, T. et al: Automatic ranked output from Boolean searches in SIRE. - Journal of the American Society for Information Science, 28(1977), str. 333-339.

PAPE, D.L., JONES, R.L.: STATUS with IQ - escaping from the Boolean straitjacket. - Program, 22(1988), str. 32-43.

PERRY, S.A., WILLETT, P.: A review of the use of inverted files for best match searching in information retrieval systems. - Journal of Information Science, 6(1983), str. 59-66.

POPOVIČ, M.: Sodobni trendi v iskanju dokumentov. - Knjižnica, 34(1990), str. 9-31.

POPOVIČ, M., WILLETT, P.: Processing of documents and queries in a Slovene language free text retrieval system. - Literary and Linguistic Computing, 5(1990), v tisku.

PORTER, M.F.: An algorithm for suffix stripping. - Program, 14(1980), str. 130-137.

PORTER, M.F., GALPIN, V.: Relevance feedback in a public access catalogue for a research library: Muscat at the Scott Polar Research Institute. - Program, 22(1988), str. 1-20.

ROBERTSON, S.E., SPARCK JONES, K.: Relevance weighting of search terms. - Journal of the American Society for Information Science, 27(1976), str. 129-146.

TENOPIR, C.: Full text databases. - Annual Review of Information Science and Technology, 19(1984), New York : Elsevier Science Publishers, str. 215-246.

VICKERY, A. et al: A reference and referral system using expert system techniques. - Journal of Documentation, 43(1987), str. 1- 23.

WADE, S.J., WILLETT, P.: INSTRUCT: a teaching package for experimental methods in information retrieval. Part 3. Browsing, clustering and query expansion. - Program, 22(1988), str. 44-61.

WADE, S.J. et al: A comparison of knowledge-based and statistically- based approaches to reference retrieval. - Online Review, 12(1988), str. 91-108.

WADE, S.J. et al: SIBRIS: the Sandwich Interactive Browsing and Ranking Information System. - Journal of Information Science, 15(1989), str. 249-260.

WILLETT, P.: A fast procedure for the calculation of similarity coefficients in automatic classification. - Information Processing and Management, 17(1981), str. 53-60.

WILLETT, P. et al: Implementation of nearest-neighbour searching in an online chemical structure search system. - Journal of Chemical Information and Computer Science, 26(1986), str. 36-41.

WILLETT, P., WOOD, F.E.: Use of the INSTRUCT text retrieval program at the Department of Information Studies, University of Sheffield. - Education for Information, 7(1989), str. 133-141.

WOOD, F.E.: Online teaching aids from UK library schools. - Journal of the American Society for Information Science, 35(1984), str. 53-55.