

**Sveučilište u Zagrebu
Filozofski fakultet
Ivana Lučića 3
Zagreb**

Koraljka Golub

**PREDMETNO PRETRAŽIVANJE U KNJIŽNIČNIM
KATALOZIMA S WEB-SUČELJEM**

MAGISTARSKI RAD

Zagreb, 2003.

Najsrdajnije zahvaljujem svojoj voditeljici, profesorici dr. sc. Jadranki Lasić-Lazić, na svoj pomoći.

Također zahvaljujem profesorici dr. sc. Danieli Živković na pruženoj podršci.

Posebno zahvaljujem profesorici dr. sc. Tatjani Aparac-Jelušić na svim sugestijama i savjetima.

U Zagrebu, lipnja 2003.

Koraljka Golub

Uvod

U današnje doba globalnog informacijskog povezivanja javno dostupni online knjižnični katalogi imaju ključnu ulogu. Korisnicima s pristupom na Internet omogućuju stalan pristup informacijama o građi koju knjižnica posjeduje, kao i određenim jedinicama mjesno ili daljinski dostupne elektroničke građe. Istraživanja pokazuju da, kako bi ovim informacijama i građi pristupili, korisnici online knjižničnih kataloga najčešće obavljaju predmetna pretraživanja, kojima otkrivaju i pronalaze jedinice građe na određenu temu. Međutim, ova vrsta pretraživanja ujedno je i najmanje uspješna.

Korisnici često na upit vezan uz pretraživanje po predmetu dobiju previše ili premalo relevantnih rezultata i ne trude se preoblikovati upit i ponoviti pretraživanje. Stoga se ovim radom namjeravaju problematizirati dosadašnja sporna pitanja i postojeći odgovori na njih u obliku predloženih rješenja, te kritički promotriti obilježja predmetnog pretraživanja kataloga s web-sučeljem sveučilišnih knjižnica Velike Britanije u usporedbi s hrvatskima. U Hrvatskoj se mali broj stručnjaka bavi pretraživanjem, a posebice predmetnim pretraživanjem¹ pa se ovim radom žele podastrijeti određena rješenja te potaknuti daljnja istraživanja.

Budući da su izrasli iz kataloga na listićima, online katalogi zadržali su više ili manje osnovna svojstva tradicionalnih kataloga, istodobno ne uvažavajući prednosti računalnog povezivanja i hipertekstualnog medija. Knjižnični katalogi s web-sučeljem imaju velik potencijal za pružanje brojnih mogućnosti, od samog sadržaja kataloga (bibliografski zapisi, daljinski dostupna građa, potpuni tekstovi i dr.) do oblikovanja korisnički pristupačnih sučelja (prozori, sličice, povlačni izbornici i dr.) te stoga mogu utjecati na rješavanje nekih od problema na koje korisnici nailaze pri predmetnom pretraživanju. Katalogi s web-sučeljem ubrajaju se u tzv. treću generaciju online knjižničnih kataloga u čija obilježja spadaju mogućnost raznolikih pretraživanja, datoteke predmetnih preglednih kataložnih jedinica, tezaursi, mogućnosti prebiranja (browsing), automatskog kraćenja i ispravljanja pogrešaka u pisanju, oblikovanje upita prirodnim jezikom bez potrebe za korištenjem složene sintakse, podržavanje veće interakcije s korisnikom i osiguravanje pomoći pri prevođenju pojmova iz upita na rječnik korišten u katalogu,

¹ Usp. Slavić, A.; Turkulin, B. Predgovor. // Vjesnik bibliotekara Hrvatske, 1-4, 44(2001). Str. XX.

preoblikovanje upita pri neuspješnom pretraživanju i ispis rezultata prema sličnosti s upitom ili prema relevantnosti. Međutim, istraživanja pokazuju da su zapravo rijetki WebPAC-i online katalogi treće generacije u pravom smislu te riječi, sa svim njihovim obilježjima, već da se od prošle dvije generacije razlikuju jedino po tome što im se može pristupiti s Mreže.

Predmetno pretraživanje sintagma je koja se u kontekstu ovoga rada odnosi na načine na koje se podacima o jedinicama građe u katalogu može pristupiti na temelju njihovih sadržajnih obilježja. Temelji se na terminologiji kakvu rabi M. Mikačić (1990) govoreći o *predmetnoj obradi*, iako se u hrvatskom jeziku rabe i druge sintagme vezane uz sadržaj jedinica građe². Ključne sastavnice ovoga procesa u računalnom okruženju su pristupnice (access points) poput klasifikacijskih oznaka, predmetnih odrednica, deskriptora i riječi iz naslova. Takve pristupnice koje omogućavaju pronalaženje jedinica građe na temelju njihova sadržaja nazivaju se predmetnim pristupnicama (subject access points).

Sintagma *knjižnični katalogi s web-sučeljem* hrvatska je inačica dviju engleskih sintagmi, *web-based online public access catalog* i *web online public access catalog*, koje se odnose na knjižnične kataloge kojima se pristupa putem Mreže (World Wide Web). U literaturi se umjesto ovih sintagmi rabe i akronimi *Web OPAC*, *Web-OPAC* i *WebPAC*, a, kao što ističe A. Barbarić (2002, str. 1), *WebPAC* je najprimjereniji jer “Web-” podrazumijeva online pristup koji predstavlja slovo “O” u akronimu *Web OPAC*, pa je “O” suvišno. U engleskoj se literaturi zapravo najčešće nailazi na skraćene sintagme *online library catalog*, *web-based catalog*, *web catalog*. Sintagma *online knjižnični katalog* u ovome će se radu koristiti kao nadređeni pojam za online kataloge, bilo da se radi o web-sučelju ili ne.³

² “Ispitivanjem sadržaja bavi se više disciplina, naročito tekstologija i teorija informacija. U tim je disciplinama uveden naziv *analiza sadržaja*, (content analysis), ili analiza teksta, što znači dublju i svestraniju analizu smisla teksta. Zato je bolje da u bibliotekarskoj struci pored uopćenog naziva *sadržajna obradba dokumenta* više upotrebljavamo naziv *predmetna obradba dokumenta* ili *predmetna obradba teksta*, (subject analysis)” (Mikačić 1990, str. 25). Ovaj se navod odnosi na pojmovni sadržaj dokumenta, odražen u stvarnim knjižničnim katalogima. *Stvarni katalog* viši je rodni pojam za *stručni katalog* i *predmetni katalog*.

³ “Online Public Access Catalogue (online katalog s javnim pristupom). Katalog knjižnice ili informacijskog centra korisnicima dostupan online putem, koji većinom pruža niz dodatnih usluga poput informacije o statusu dostupnosti građe, online rezervacija, i knjižničnih novosti. S nestankom kataloga na listićima, nestala je potreba da se naglašava “online s javnim pristupom” pa se danas o ovoj vrsti kataloga jednostavno govori samo kao o ‘katalogima’.” (Online Public Access Catalogue. // Harrod's librarians' glossary and reference book, 2000)

U većem dijelu rada o predmetnom se pretraživanju govori u kontekstu online knjižničnih kataloga općenito, i to iz tri razloga. Prvo, o predmetnom pretraživanju u online katalogima puno se istraživalo i pisalo prije nastanka WebPAC-a. Osim toga, većina načela koja su se pokazala korisnima za funkcije predmetnog pretraživanja prvih online kataloga mogu se primijeniti i na WebPAC-e. Treći razlog odnosi se na to da se WebPAC-i u funkcijama predmetnog pretraživanja od prethodne dvije generacije razlikuju samo po tome što im se može pristupiti s Mreže.

Prvo poglavlje osvrnut će se na predmetno pretraživanje kao zadaću knjižničnog kataloga i na pretpostavke za predmetno pretraživanje u bazi podataka računalnog kataloga.

Razvojem predmetnog pretraživanja u online katalogima i prednostima u odnosu na tradicionalne kataloge bavit će se drugo poglavlje.

Iduće poglavlje osvrnut će se na problematiku predmetnog pretraživanja, na dosadašnje pristupe i rješenja postojećih pitanja.

Budući da se korisnički sustavi trebaju temeljiti na potrebama i ponašanju svojih korisnika, u četvrtom će se poglavlju prikazati kako se korisnici služe predmetnim pretraživanjem u online knjižničnim katalogima, koji su problemi s kojima se susreću te koja su njihova očekivanja.

U petom poglavlju opisat će se različite vrste predmetnih pristupnica i predstaviti obilježja prirodnih i kontroliranih jezika za označivanje i pretraživanje te utjecaj tih obilježja na predmetno pretraživanje. Predstavit će se i tehnike pretraživanja iz područja informacijskog pretraživanja (information retrieval⁴), koje se nedostavno primjenjuju u predmetnom pretraživanju WebPAC-a.

Šesto poglavlje obrađuje interakciju korisnika s katalogom s naglaskom na sintaksu postavljanja upita, što obuhvaća razne operatore, oblike kraćenja, oblikovanje upita, prebiranje i rangiranje rezultata.

U idućem poglavlju prikazat će se pitanja vezana uz oblikovanje sučelja online knjižničnog kataloga za predmetno pretraživanje, u koja se ubrajaju pitanja dizajna, pristupačnosti sučelja, osiguravanja pomoći i uputa te višejezičnosti.

⁴ "Information retrieval (informacijsko pretraživanje). Traženje i pronalaženje informacija u računalnom spremištu, prve metode obuhvaćale su klasificiranje i katalogiziranje i pretraživanje baza podataka raznim mehaničkim tehnikama. Elektroničke metode danas su zamijenile ove sustave pa se suvremeno pretraživanje temelji na traženju u bazama podataka s potpunim tekstom, bibliografskim bazama podataka i dostavi dokumenata putem mreže." (Information retrieval. // Harrod's librarians' glossary and reference book, 2000)

Osmo poglavlje izdvojit će specifičnosti predmetnog pretraživanja vezane uz WebPAC-e.

Predmetno pretraživanje vrlo je važno znanstvenom i nastavnom osoblju te studentima, koji provode istraživanja na određenu temu. Zbog toga je istraživanje u okviru ovoga rada usmjereno prvenstveno na sveučilišne knjižnice, a opisano je u devetom poglavlju. Na temelju istraživanja i evaluacije dosadašnjih pristupa izdvojit će se elementi i metodologija oblikovanja predmetnog pretraživanja primjerenog korisnicima WebPAC-a u hrvatskoj akademskoj zajednici.

1. Osnovne pretpostavke predmetnog pretraživanja u online knjižničnim katalozima

1.1. Predmetno pretraživanje kao zadaća knjižničnog kataloga

Predmetno pretraživanje kao zadaću knjižničnog kataloga navodi još 1876. godine C. Cutter u svojim Pravilima za ukršteni katalog i to u sljedeće dvije zadaće: 1. omogućiti pronalaženje knjige za koju korisnik poznaje autora, naslov ili predmet (subject), te 2. pokazati što knjižnica od određenog autora, o određenom predmetu i određenoj vrsti literature posjeduje (Cutter 1948, str. 12).

Budući da korisnici ne moraju uvijek znati ime autora ili naslov jedinice građe⁵ koju traže, potrebno je osigurati mogućnost pronalaženja građe i po sadržaju. Stvarni knjižnični katalozi na listićima pružaju informacije o građi iz određenog stručnog područja (stručni katalozi) te o određenom predmetu (predmetni katalozi⁶). Postoje i ukršteni ili rječnički katalozi (dictionary catalog), koje je u američke knjižnice uveo C. Cutter, a sastoje se od abecednog i predmetnog niza. U elektroničkom okruženju svi ovi katalozi, uključujući, dakako, i formalne i mjesne kataloge, srastaju u jedinstveni katalog pa se pretraživanje može obavljati po svim elementima, odnosno poljima kataložnog zapisa (po autoru, naslovu, predmetu, signaturi itd.).

Mnoga istraživanja pokazuju da predmetno pretraživanje u online katalozima dominira ostalim vrstama pretraživanja. Većina današnje literature o korištenju online kataloga temelji se na studiji američkoga Savjeta za knjižnične izvore (Council on Library Resources – CLR) iz ranih 1980-ih, u kojima je sudjelovalo više od 12.000 online korisnika iz dvadeset i devet knjižnica sa sedamnaest različitih online kataloga (Lombardo, Condic 2000; Olson, Boll 2001, str. 2). To je istraživanje pokazalo da najviše korisnika pretražuje po predmetnim pristupnicama, i to najviše po ključnim riječima. D. Slone (2000, str. 757), nakon što je proučila rezultate istraživanja raznih autora, ističe da su pretraživanja po autoru i naslovu

⁵ Termin *jedinica građe* koristit će se isključivo u kontekstu knjižničnih kataloga, dok će se sinonim *dokument* rabiti kada se govori općenito o informacijskim sustavima za pretraživanje, uključujući knjižnične kataloge.

⁶ Kao što je napomenuto u Uvodu, sintagma *predmetno pretraživanje* odnosi se na sve vrste pristupnica koje označuju sadržaj, a nije ograničena na pristupnice *predmetnog kataloga*.

najčešća među korisnicima kataloga na listićima i prvih online knjižničnih kataloga, a da korisnici provode više predmetnog pretraživanja u online katalogima nego u katalogima na listićima. Na osnovi tih rezultata moguće je zaključivati o tome da je predmetno pretraživanje nadasve korisno i poželjno.

Godine 1997. IFLA⁷-inom studijom Funkcionalni zadaci bibliografskih zapisa (Functional Requirements for Bibliographic Records – FRBR) zadaće se kataloga nastoje prilagoditi računalnom okruženju. Pritom, katalog se više ne promatra kao niz bibliografskih zapisa, već kao mreža povezanih informacija. Kako bi se informacije sadržane u kataložnim zapisima prilagodile potrebama njihovih korisnika, oblikovan je model bibliografske baze *entitet-odnos* i određene su tri skupine entiteta. Prva skupina obuhvaća djelo (work), izraz (expression), manifestaciju (manifestation) i pojedinačni primjerak (item), tj. podatke koji su tradicionalno dio bibliografskog zapisa; druga skupina sastoji se od osoba i korporativnih tijela odgovornih za intelektualni i umjetnički sadržaj, proizvodnju, diseminaciju ili vlasništvo entiteta iz prve skupine (tzv. bibliografski entiteti); a treću skupinu (tzv. predmetni entiteti) sačinjavaju elementi koji se odnose na sadržajnu obradu: pojam (concept), predmet (object), događaj (event) i mjesto (place). Svi ovi elementi proizašli su iz Međunarodnih standarda za bibliografski opis (International standard bibliographic description – ISBD), Priručnika za UNIMARC (UNIMARC manual), Smjernica za izradu preglednih kataložnih jedinica i uputnica (Guidelines for authority and reference entries – GARE) i Smjernica za izradu predmetnih preglednih kataložnih jedinica i uputnica (Guidelines for subject authority and reference entries – GSARE). (Buzina 2000; Functional requirements for bibliographic records 1998; Katić 2001; Riesthuis, Zumer 2002).

Valja istaknuti da i entiteti iz prve i druge skupine mogu biti predmet djela. Jedan od problema pretraživanja jest taj što često nije lako napraviti jasnu razliku između formalne i sadržajne obrade jer korisnik može koristiti iste pristupnice za traženje poznate mu manifestacije i manifestacije o određenoj temi. Poznati je primjer ime osobe kao ključna riječ, kojim se pretraživanjem dobivaju i manifestacije djela kojima je ta osoba autor i manifestacije djela o toj osobi (Riesthuis, Zumer 2002, str. 170).

⁷ Međunarodno udruženje knjižničnih društava i ustanova (International Federation of Library Associations and Institutions – IFLA).

Prema studiji FRBR, pretraživanje se ne ograničava samo na autora, naslov i na predmet. Bibliografske baze podataka trebale bi osigurati pristup po svim obilježjima važnim za pronalaženje, dakle po svim obilježjima bibliografskih entiteta djela, izraza, manifestacija i pojedinačnih primjeraka, uz pristup datotekama preglednih kataložnih jedinica i uputnica te datotekama predmetnih preglednih kataložnih jedinica i uputnica. Ove datoteke trebale bi omogućiti korisniku da pronađe ispravan oblik pristupnica posebno pri problematičnim slučajevima poput klasifikacijskih oznaka i pristupnica za autora i za naslov, za koje postoji dvije ili više varijanti. Kvaliteta modela predloženog u studiji FRBR moći će se bolje sagledati nakon što se izradi studija koja će ista načela primijeniti i na pregledne kataložne jedinice i uputnice, čime se danas bavi IFLA-ina Radna grupa za funkcionalne zadatke i obrojčavanje preglednih kataložnih jedinica (Working Group on Functional Requirements and Numbering for Authority Records – FRANAR) (Chen, Lin, Chen 2002; Riesthuis, Zumer 2002).

1.2. Organizacija i struktura baze podataka

Baza podataka online kataloga sastoji se od nekoliko datoteka, od kojih su za predmetno pretraživanje posebno važne bibliografska datoteka, datoteka predmetnih preglednih kataložnih jedinica i uputnica, te invertirana datoteka.

Bibliografska datoteka sastoji se od bibliografskih zapisa, surogata koji predstavljaju izvornu jedinicu građe, bilo da se ona nalazi u knjižnici ili negdje na Mreži. Svaki je zapis kodiran, a međunarodni format za kodiranje elektroničkih bibliografskih zapisa, odnosno za strojno čitljivo katalogiziranje jest format MARC (MAchine Readable Cataloging). Postoje različite inačice formata MARC⁸, kao, primjerice, UNIMARC, koji se koristi i u Hrvatskoj, ili američki MARC 21, koji je nedavno naslijedio USMARC i britanski UKMARC.

Polja u zapisima formata MARC koja sadrže podatke o sadržaju jesu sljedeća: 6XX, za termine iz kontroliranih rječnika; polja od 050 do 088, za klasifikacijske oznake i signature; te većina polja između 2XX i 5XX, predviđena za

⁸ Kada se u tekstu koristi samo kratica “MARC”, ona predstavlja generički naziv za sve njegove inačice.

potencijalne ključne riječi za pretraživanje (Olson, Boll 2001, str. 12-13). U poljima iz niza 6XX koja sadrže predmetne odrednice postoji kôd sustava, popisa predmetnih odrednica ili tezaurusa, iz kojeg je termin preuzet (npr. Library of Congress Subject Headings – LCSH). U UNIMARC-u⁹ su predmetne odrednice i klasifikacijske oznake u poljima 675-686, signatura je u 990, a interna lokacija dokumenta u 992. Od ostalih polja u zapisu formata MARC po kojima se može pretraživati po ključnim riječima najvažnija su ona za naslove i napomene. Često naslovi sadrže termine koji su vrlo specifični i nisu obuhvaćeni kontroliranim sustavom termina, a slično je i s napomenama, poput napomene o sadržaju (u UNIMARC-u polje 327, a u MARC-u 505) i napomene za kratak sadržaj ili sinopsis (u UNIMARC-u polje 330, a u MARC-u 520).

Polja koja sadrže klasifikacijske oznake i signature, tvrde H. Olson i J. Boll (2001, str. 14), dobit će na popularnosti kada se uvidi prednost knjižničnih klasifikacijskih sustava nad klasifikacijama koje koriste mrežne web-tražilice. No, valja imati na umu da oni govore s američkog stajališta, gdje kataložna tradicija, za razliku od europske, u pravilu nije rabila klasifikacijske sustave za pretraživanje, već prvenstveno za izradu signatura i smještaj građe na policama.

Datoteka predmetnih preglednih kataložnih jedinica i uputnica sadrži standardizirane oblike predmetnih odrednica, njihove sinonime i srodne, šire i uže termine pa je ova datoteka, može se kazati, elektronička inačica kontroliranog rječnika (Olson, Boll 2001, str. 117). Od 1998. postoji i format za klasifikacijske oznake u MARC-u 21 (MARC 21 concise format for classification data 2002), dok je format za klasifikacijske oznake u UNIMARC-u još u razvoju.

Invertirane datoteke su one koje sustav oblikuje automatski, tako da se pojedinačna polja, potpolja ili riječi stavljaju u posebnu datoteku. Invertirane datoteke pospješuju učinkovito pretraživanje, a pri oblikovanju knjižničnih sustava i invertiranih datoteka postavljaju se sljedeća pitanja: od kojih će se polja i potpolja oblikovati invertirane datoteke, odvajanje datoteka za pretraživanje po ključnim riječima i po predmetnim odrednicama, hoće li se invertirane datoteke sastojati samo od riječi ili i od izraza te što će sve uključivati stop-lista (Olson, Boll 2001, str. 24).

⁹ Usp. Mikačić, M. Teorijske osnove sustava za predmetno označivanje. Zagreb : Hrvatsko bibliotekarsko društvo, 1996. Str. 383-385.

Što se tiče uključivanja potpunog teksta jedinica građe u kataložnu bazu, postoje različita mišljenja. Obradujući pitanja vezana uz mogućnosti pretraživanja potpunih tekstova (full-text retrieval), A. Barbarić (2002, str. 43) ukazala je na to da je poželjno da takva pretraživanja budu primjenjiva samo na dio građe, dok bi ostala građa bila dostupna samo po uobičajenim pristupnicama, te da bi ipak primjerenije bilo da katalog samo osigurava pristup potpunom tekstu jedinica građe.

U globalnom mrežnom okruženju i distribuiranim digitalnim knjižnicama, gdje granice između online kataloga, baza podataka i mrežnih izvora nestaju, a koriste se različiti propisi za obradu informacijskih izvora, posljednjih se nekoliko godina u literaturi predstavljaju prijedlozi i pokušaji modela primjene novih formata zapisa, a samim time i nove strukture baza podataka (vidi, npr., Willer 2000). Tako se navodi da formati MARC onemogućavaju ostvarivanje potencijalnih mogućnosti koje nudi mrežno okruženje, koje se znatno razlikuje u usporedbi s mogućnostima kakve su postojale u katalogima dostupnim na lokalnim mrežama ili putem telnet. SGML (Standard Generalized Markup Language – standardni jezik za prikazivanje dokumenta) i njegova jednostavnija inačica XML (eXtensible Markup Language – proširljivi jezik za prikazivanje dokumenata), omogućavaju pretraživanje samih jedinica građe i njihovih dijelova, pregledavanje bibliografskih zapisa u web-preglednicima (browser), web-tražilicama (search engine) i knjižničnim sustavima bez potrebe za konverzijom formata (vidi, npr., Oritz-Repiso, Moscoso 1999). XML se u pojedinim sustavima koristi i kao format za kontrolirane rječnike, a njegova obilježja omogućuju da se putem jednog sučelja pristupi različitim kontroliranim rječnicima (vidi, npr., Chaudiron, Ihadjadene, Role 2002).

U stručnim se krugovima podosta raspravlja i o potrebi prilagodbe katalogizacijskih pravila i strukture kataloga, jer su postojeća pravila i struktura samo preslikani na računalni katalog, a nisu prilagođeni mrežnom okruženju (Fattahi 1995; Ayres 1999). Novi model bibliografske baze opisan je u studiji FRBR, koja se temelji na novom okruženju i na načelima oblikovanja relacijskih baza podataka. Bibliografska baza sastoji od mreže spojenih zapisa, npr. zapis za djelo povezan je sa zapisima za izraze toga djela i zapisima za manifestacije u kojem postoji jedan ili više izraza. Svaki od ovih zapisa može se spojiti s jednim ili više zapisa iz datoteka preglednih kataložnih jedinica i uputnica, iz kojih se svaki pak može povezati s nekoliko zapisa za bibliografske jedinice. Na ovaj način

moguća je navigacija od zapisa do zapisa, među različitim datotekama (Riesthuis, Zumer 2002; Chen, Lin, Chen 2002).

Govoreći o budućnosti MARC-a, H. Olson i J. Boll (2001, str. 15) tvrde da ranije navedena polja predstavljaju golemu fleksibilnost u predmetnom pretraživanju, te da je teško zamisliti da postoji neka vrsta informacije o sadržaju koja se ne bi mogla obuhvatiti ovim formatom. Poboljšanja koja oni smatraju potrebnima jesu kvalitetnije klasificiranje i mogućnost pretraživanja po pojedinim fasetama.

Metapodaci mogu jedini “ugroziti” format MARC jer se rabe i ugrađuju u sve veći broj elektroničkih informacijskih izvora dostupnih na Mreži. Međutim, nema velike konceptualne razlike između metapodataka i bibliografskih zapisa – i jedni i drugi su podaci o podacima¹⁰. Osnovni problem s metapodacima jest da su se oni do sada najmanje bavili sadržajnim pristupom informacijama. Ipak, nude mogućnosti za standardizirani prikaz raznolikijih informacijskih izvora, i to u puno širem kontekstu no što to može današnja katalogizacijska praksa. Postojeći kontrolirani rječnici mogu se primijeniti za poboljšanje sadržajnog pristupa informacijama metapodataka u budućnosti. Osim toga, veliki su izgledi da se granice između kataložnih podataka i metapodataka pomute zbog današnje tendencije povezivanja online kataloga s elektroničkim kazalima časopisa i bazama podataka te zbog sve većeg broja online kataloga na Mreži (Olson, Boll 2001, str. 15). Već postoje alati za prijenos iz metapodataka poput Dublinske jezgre (Dublin Core) u MARC i obrnuto, a razvijaju se u okviru projekta CORC (Cooperative Online Resource Catalog) američkog OCLC-a (Online Computer Library Center) – zapisi se mogu importirati ili eksportirati u nekom od postojećih formata. Budućnost će pokazati hoće li se format MARC samo mijenjati, ili će ga se napustiti i zamijeniti nekim drugim formatom.

¹⁰ MARC zapravo spada u istu kategoriju kao i SGML, HTML i XML, dakle u standarde za kodiranje elemenata u strojnočitljiv oblik, a od njih se razlikuje po tome što je specijaliziran za knjižničnu građu. Za razliku od ovih standarda, standardi metapodataka propisuju elemente koji će biti uključeni u surogat (kao što je kataložni zapis u katalogu).

2. Razvoj i pristupi problematici predmetnog pretraživanja online knjižničnih kataloga

2.1. Prednosti i nedostaci online kataloga u odnosu na kataloge na listićima u predmetnom pretraživanju

Još 1984. A. Seal uočava da je jedna od najvažnijih prednosti online kataloga nad tradicionalnima interaktivnost koja utječe na pretraživanje općenito, pa tako i na predmetno pretraživanje. Primjerice, katalog može korisniku pokazati koje mu mogućnosti pretraživanja stoje na raspolaganju, ispravljati pogreške, predlagati alternative jedinice građe koje mogu odgovarati kriterijima pretraživanja i voditi korisnika kroz dugo pretraživanje.

Dok se u katalogu na listićima okupljanje i pretraživanje jedinica građe na istu temu postiže u stvarnim katalogzima, u online okruženju za predmetno pretraživanje postoji više mogućnosti. Tako u online katalogu postoje dodatne pristupnice (ključne riječi po različitim poljima), manipuliranje i kombiniranje podataka na mnogo novih načina (operatori, kraćenje, istodobno pretraživanje po više polja zapisa te sužavanje opsega pretraživanja (jezik, vrsta građe, godina izdavanja itd.).

Online katalozi za razliku od kataloga na listićima omogućuju da se korisniku pruži pomoć na nekoliko razina i načina. Sustav može odgovoriti na korisnikove probleme i pomoći mu da donese odluku o tome što dalje učiniti da dobije bolje rezultate (Crawford 1992, str. 67). Međutim, u online okruženju korisnici su zapravo prepušteni sami sebi pa je kvalitetna pomoć u obliku uputa neophodna. Istraživanja, međutim, pokazuju da današnji sustavi za pomoć i upute nisu dovoljno usavršeni.

U odnosu na katalog na listićima, u online katalogu načelno je niži prag tolerancije pogrešaka, bilo da se radi o pogreški katalogizatora ili korisnika. Zbog toga je potrebno razvijati sustave koji omogućuju prebiranje, prepoznaju pogreške i korisniku nude pomoć.

Iako su prednosti online kataloga nad katalogzima na listićima u predmetnom pretraživanju mnogobrojne, ta su svojstva u današnjim katalogzima samo djelomično

iskorištena, što se posebno odnosi na interaktivnost, tehnike pretraživanja kakve se koriste u drugim sustavima za informacijsko pretraživanja, pa čak i na predviđena svojstva polja za predmetni pristup informacijama formata MARC koja se gotovo uopće ne koriste.

2.2. Razvoj predmetnog pretraživanja u online katalozima

U literaturi se razvoj online knjižničnih kataloga dijeli u tri osnovna razdoblja, s izraženim razlikama između prve, druge i treće generacije (Borgman 1996; Hildreth 1984; Mitev, Venner, Walker 1985).

Prva generacija online kataloga pojavila se sredinom 1970-ih, kada su se računala intenzivnije počela koristiti u knjižnicama. Slijedila je dva dotadašnja modela: katalog na listićima (za generaciju korisnika koja je poznavala kataloge na listićima oblikovana su njima bliska sučelja u računalnom okruženju) i Booleov model pretraživanja (na temelju baza podataka poput DIALOG-a ili Medline-a). Struktura zapisa, sadržaj i primarna polja za pretraživanje temeljila su se na katalogu na listićima, a funkcije pretraživanja na sustavima za pretraživanje. Od mogućnosti pretraživanja ti su katalozi nudili samo osnovne funkcije. Pristupnice su bile iste kao i u kataložnim listićima, dakle autor, naslov i predmet u Sjedinjenim Američkim Državama, odnosno autor, naslov i klasifikacijska oznaka u Europi (Borgman 1996, str. 497). A. O'Brien (1994) spominje i signaturu. C. Hildreth (1984) pristup po predmetnim odrednicama svrstava u drugu generaciju. U prvoj generaciji kataloga nizovi znakova iz korisničkog upita morali su se u potpunosti podudarati s onima u bibliografskom zapisu. Podržavali su se samo oni upiti koji su se sastojali od jedne riječi, a odnose među riječima u upitima s više riječi korisnik je morao definirati Booleovim operatorima. Prva generacija online kataloga se, stoga, naziva sustavima s prethodnim označivanjem (precoordinated) ili sustavima s izrazima (phrase-indexed), a dobri su za pronalaženje jedinica građe za koje korisnik zna da postoje i ima podatke potrebne za njihovo pronalaženje (known item searching) (Hildreth 1984; O'Brien 1994).

Druga generacija online kataloga dominirala je sredinom 1980-ih. Pružala je dodatne mogućnosti pretraživanja kao što su ključne riječi, implicitni Booleovi

operatori i prebiranje. Velika je značajka uvođenje ključnih riječi kao pristupnica jer je predmetno pretraživanje postalo moguće provoditi po ključnoj riječi po bilo kojem indeksu polja u bazi podataka. Druga generacija podržavala je unos više riječi u upitu, tako što je sustav automatski, sam postavljao Booleove operatore među riječima iz upita. O online katalogima druge generacije često se govori kao o katalogima s naknadnim označivanjem (postcoordinated) ili katalogima s ključnim riječima (keyword) jer su pristupnice riječi iz naslova, predmetnih odrednica, autora ili drugih imena, a upit se može oblikovati korištenjem Booleovih operatora (O'Brien 1994). Međutim, svaki je sustav imao vlastiti način provođenja odnosa među riječima, npr. između riječi se automatski stavlja operator OR ili operator AND, ili se niz riječi tretira kao izraz. Neki su sustavi poštivali hijerarhiju Booleove logike (NOT, AND, OR), a drugi su pak operacije obavljali jednostavno s lijeva na desno. Zgrade su neki sustavi dozvoljavali, a drugi nisu. C. Hildreth (1984) navodi i mogućnost prebiranja po predmetnim odrednicama, ali ističe probleme s rječnikom za predmetno pretraživanje te prevelike i loše organizirane skupove rezultata. Kao što tvrdi C. Borgman (1996), iako su nastupile promjene u smislu poboljšanja mogućnosti pretraživanja, funkcionalnost druge generacije online knjižničnih kataloga u odnosu na prvu nije značajno poboljšana. C. Hildreth (1984) smatra da druga generacija online kataloga nije uspješna jer katalogi automatski ne nude korisniku alternativu pri pretraživanju ako pretraživanje ne uspije, jer ne omogućuju fleksibilno pretraživanje, te korisnicima ne nude pristup sažecima i recenzijama knjiga. Isti je autor u svome radu (Hildreth 1984) predložio potrebna poboljšanja za treću generaciju: pomoć pri oblikovanju strategije pretraživanja, integrirani pristup riječima slobodnog teksta (free-text) i riječima iz kontroliranog rječnika, opsežnije i bolje kataložne zapise, istodoban pristup različitim bazama podataka, unos upita na prirodnom jeziku, individualizirani prikaz rezultata, i ispravljanje pogrešaka prilagođeno kontekstu.

Krajem 80-ih počelo se raspravljati o *Web-based OPAC-ima* odnosno *Web OPAC-ima (WebPAC-i)*, online knjižničnim katalogima treće generacije, kojima se može pristupiti putem sučelja World Wide Weba. U odnosu na svoje tekstualne prethodnike, ti katalogi pružaju niz novih mogućnosti. Hipertekstualni model organizacije, prebiranja i pronalaženja informacija pruža knjižničnim katalogima mogućnost da postanu jedan, kako to C. Long (2000, str. 161) naziva, "trgovinski centar" za bibliografske zapise, građu s potpunim tekstom i podatke s hipervezama,

osiguravajući tako iscrpnije korištenje informacija koje online katalogi već sadrže. Podržavaju dodatne tehnike pretraživanja (uz one koje se temelje na Booleovim operatorima, npr. tehnike za pronalaženje najbližih rezultata, tzv. *best match* tehnike), nude kvalitetnije zapise koji uključuju dodatne kontrolirane i nekontrolirane pristupnice (poput naslova poglavlja u knjigama), omogućuju oblikovanje upita na prirodnom jeziku, uz mogućnost korištenja rječnika za kratice, sinonime, ili pravopisne varijante. Također podržavaju kontekstualne automatizirane upute, korištenje termina iz dobivenih relevantnih zapisa za poboljšanje pretraživanja te prikaz najrelevantnijih zapisa prvih (O'Brien 1994). To što osiguravaju korisnicima poznate alate poput povlačnih izbornika (pull-down menus), prozora i sličica, kao i tražilice nalik onima na Internetu (search-engines), korisnicima znatno olakšava predmetno pretraživanje (Long 2000, str. 141).

Unatoč do danas ostvarenim poboljšanjima, istraživanja su pokazala (Borgman 1996; Drabenstott, Weller 1996; Slone 2000) da korisnici imaju velikih problema s predmetnim pretraživanjem, od toga da ne mogu prevesti svoje koncepte u jezik kataloga, do toga da pronalaze previše ili premalo zapisa te da se susreću s velikim brojem neuspjelih pretraživanja. Problematika treće generacije online knjižničnih kataloga složenija je od problematike prethodnih dviju generacija, jer oni osim bibliografskih zapisa osiguravaju pristup i online bazama podataka, udaljenim izvorima s Mreže, raznim vrstama datoteka vezanim uz bibliografske zapise te drugim knjižničnim katalogima (Matthews 1997, str. 115). Iako je dizajn jednostavniji i jasniji nego ranije, funkcionalnost online kataloga u osnovi se nije promijenila od ranih 1980-ih (Borgman 1996, Matthews 1997). Vjeruje se da većina problema koji su i nadalje vezani uz online kataloge postoji zbog zadržavanja strukture kataloga na listićima i nemogućnosti da se oblikuju sučelja koja se temelje na znanju i vještinama korisnika online kataloga. Važan iskorak, prema našem mišljenju, zasniva se na uvažavanju korisničkog ponašanja pri traženju informacija (usp. Borgman 1996).

3. Pristupi problematici predmetnog pretraživanja u online katalozima

Jedno od prvih istraživanja predmetnog pretraživanja u online katalozima opisuje K. Markey 1983. godine. Istraživanje je pokazalo da korisnici imaju problema s predmetnim pretraživanjem, te je predstavila i očekivanja od online kataloga, koja su, zajedno s osobljem, naveli korisnici. Većinu tih očekivanja C. Hildreth (1984) navodi kao obilježja treće generacije online kataloga.

U slučajevima kada ne znaju autora, naslov ili predmetnu odrednicu, korisnici smatraju korisnom mogućnost pretraživanja po ključnim riječima i po signaturi. Pri pretraživanju po ključnim riječima vrlo je korisno da sustav sam predloži sve moguće termine po kojima bi se upit također mogao provoditi. Kao i u predmetnom i stručnom katalogu na listićima, važno je da postoji mogućnost prebiranja i odabira. To je važno iz nekoliko razloga, od toga da se pri unošenju riječi i oblikovanju upita lako pogriješi pri pisanju (tipfeleri), ili učini pravopisna pogreška, do toga da problemi mogu nastati i zato što postoje riječi koje se mogu različito pisati (npr. sa i bez povlake). Sve ove slučajeve spominje i M. Yee (1991).

Najveći problem korisnicima predstavlja pronalaženje relevantne predmetne odrednice. Riječi iz upita za pretraživanje često su odviše specifične ili pak preopćenite. O tome govore i drugi autori (npr. Yee 1991). M. Bates (1989b) kao razlog nedovoljne specifičnosti popisa predmetnih odrednica Kongresne knjižnice (LCSH) za velike online kataloge ističe to što se predmetna odrednica tradicionalno birala tako da obuhvati cjelovitu knjigu; smatra da je potrebno povećati broj predmetnih odrednica po jedinici građe i pobrinuti se za utvrđivanje specifičnije teme djela, a ne samo opće teme. U pronalaženje relevantne predmetne odrednice ubraja se i problem prevođenja upita na jezik sustava. M. Bates (1989b) tvrdi da se vjerojatnost da će dvoje ljudi koristiti isti termin za koncept ili knjigu ili da će tražitelj i informacijski sustav koristiti isti termin za koncept, kreće između 10 i 20%. Stoga je važno osigurati vezu između termina koji odabere korisnik i termina koji koristi sustav. Istraživanje koje su proveli Wilkes i Nelson (1995) o ulozi datoteke predmetnih preglednih kataložnih jedinica i uputnica jasno je pokazalo je takva datoteka od velike vrijednosti – od 73.9% uspješnih predmetnih pretraživanja, u katalogu bez ovakve datoteke svega je 40.6% uspješnih pretraživanja. K. Markey

(1984) naglašava da su predmetna pretraživanja po prirodi apstraktnija i često zahtijevaju više mentalnog napora da bi se neki koncept izrazio u nekoliko izraza ili riječi, za razliku od korisniku poznatih drugih jedinica građe.

Osim oslanjanja na datoteku predmetnih preglednih kataložnih jedinica i uputnica, problem pronalaženja pravih termina nastojao se riješiti i uporabom metode analize korisničkog upita na prirodnom jeziku te usporedbom s kontroliranim rječnikom. P. Le Loarer (1993, str. 259) spominje takav sustav koji se uspješno primjenjivao u francuskim aplikacijama *Minitel Guide des Services* (Minitelov vodič usluga) i *Annuaire Electronique* (Elektronički telefonski imenik), i zalaže se za primjenu takvog sustava u online katalozima. Sustav se temelji na leksičkoj, sintaktičkoj i semantičkoj analizi, a zapravo predstavlja proširenje metode koja se koristila u eksperimentalnom online knjižničnom katalogu Okapi. Drugi pristup koji se primjenjivao temelji se na rezultatima dobivenim mjerenjem statističkih odnosa između riječi iz naslova i LCSH. Takav pristup primijenjen je u eksperimentalnom online knjižničnom katalogu Cheshire. U katalogu Cheshire za pretraživanje po riječima iz prirodnog jezika isprobana je i tehnika tzv. *classification clustering* (klasifikacijsko okupljanje). Dok se u online katalozima klasifikacijska oznaka općenito tretira kao samo još jedna ključna riječ u zapisu, u ovom projektu klasifikacijske su oznake ekstrahirane iz zapisa MARC da bi se povećao broj termina vezanih uz pojedini zapis, i dodani su termini analogni s uputnicama *use* (upotrijebi) ili *RT* (Related Term – srodni naziv)) u tezaursu, upućujući korisnika na kontrolirani rječnik pojmova prikazanih klasifikacijskim oznakama. Na taj način *class clusters* (klasifikacijske skupine) funkcioniraju poput tezaurusa (O'Brien 1994, str. 227, 230).

Od ostalih problema s predmetnim pretraživanjem, K. Markey (1983, str. 383) izdvaja one koji su vezani uz razumijevanje predmetnih odrednica. Naime, pokazuje se da korisnici ne razumiju i zbunjeni su kada trebaju koristiti predmetne odrednice iz LCSH jer ne znaju na što se, primjerice, odnosi predmetna odrednica *Drug abuse - Congresses*. Razočarani su kada u popisu predmetnih odrednica naiđu na predmetnu odrednicu, a u knjižnici ne postoji jedinica građe na tu temu. Osim toga, često se u literaturi navode problemi s LCSH zbog nekonzistentnosti tog sustava, njegove nedostatne specifičnosti i ažurnosti, nepostojanja uputnica *vidi* te pomalo nelogičnog oblikovanja odrednica. Međutim, neki autori smatraju da je jednostavnije poboljšati taj sustav nego izmisliti novi (Cherry, Williamson, Jones-

Simmons, Gu 1994, str. 182), tim više što ga koristi velik broj knjižnica, pa korisnici imaju iskustvo pri njegovoj uporabi. O kontroliranim rječnicima poput LCSH govori i A. O'Brien (1994, str. 223), a kao osnovni problem ističe unos dugačkog niza znakova umjetnoga jezika. Stoga predlaže, osim prebiranja i obogaćivanja zapisa drugim sadržajnim označiteljima, pretraživanje po ključnim riječima, uključujući i pretraživanje po riječima iz predmetnih odrednica (vidi i C. Borgman, str. 497).

Što se tiče stručnog kataloga u računalnom obliku, korisnici su izrazili želju da se uz klasifikacijsku oznaku navede i opis te oznake riječima, odnosno prijevod klasifikacijske oznake. Također su naveli da bi uz signature trebalo navesti i pridruženi popis knjiga na policama (Markey 1983, str. 384).

Korisnici su također naveli želju za obogaćivanjem zapisa predmetnim pristupnicama kao što su kazala iz knjiga, opisi s knjižnih omotnica, predgovori i pogovori, te veći broj predmetnih odrednica¹¹ (Markey 1983, str. 384).

Uz rezultate Markeyinog istraživanja, u literaturi se navode i drugi problemi pri predmetnom pretraživanju. M. Yee (1991, str. 81) navodi da korisnici ne razumiju razlike između pojedinih datoteka i polja, između kontroliranih rječnika i ključnih riječi. A. O'Brien (1994, str. 224) također ističe da korisnici često ne vide razliku između pretraživanja po ključnim riječima i po terminima iz kontroliranih rječnika. Rješenje za te probleme predložila je M. Bates (1989b): ukoliko se pretraživanje po ključnim riječima odnosi na riječi iz naslova, ta se vrsta pretraživanja može jednostavno nazvati *naslov*, a ne *ključna riječ*. Ili se može zvati *pretraživanje po sadržaju* (subject search), a da se pretraživanje obavlja po naslovu ili predmetnim odrednicama ili po oba ova polja istodobno. Ovakav pristup posebno je koristan na sučelju za jednostavno ili početno pretraživanje, kada u online katalogu postoji više sučelja za pretraživanje, dok bi u sučelju za složeno ili napredno pretraživanje trebalo precizno navesti koje se datoteke pretražuju pod određenim nazivima ili oznakama polja.

M. Yee (1991) nadalje navodi da korisnici dobivaju prevelik broj rezultata; da u upitu koriste riječi iz stop-liste jer ne znaju da takva lista postoji niti što obuhvaća, provode pretraživanje po ključnim riječima kada je pretraživanje po

¹¹ Usp. O'Brien, A. Online catalogs : enhancements and developments. // Annual review of information science and technology (ARIST). Vol. 29, 1994, str. 219-241. Autorica navodi određena istraživanja prema kojima povećanje broja pristupnica dovodi do većeg broja rezultata, a prevelik broj rezultata čest je problem predmetnog pretraživanja. No, nadalje se ističe, ta su istraživanja samo indikativna.

izrazu učinkovitije i ne koriste mogućnosti poput kraćenja. Korisnici često ne razumiju kako se koriste Booleovi operatori, ne znaju kako sustav obrađuje kratice i inicijale (npr. problem može biti da korisnik u svom upitu ignorira kraticu, dok je katalog obrađuje kao riječ), ne znaju što točno znače uputnice “vidi” i “vidi i”, i ne razumiju zaslon s uputama. Pokazalo se i da zbog rijetkog korištenja kataloga korisnici svaki put moraju iznova svladavati sustav (Yee 1991, str. 81).

Prema C. Borgman (1996), u kontekstu predmetnog pretraživanja, valja istaknuti i sljedeće. Korisnik mora imati mogućnost unošenja fragmenata pitanja, istraživanja strukture rječnika, zadržavanja djelomičnih rezultata, preoblikovanja pretraživanja uz pomoć različitih specijaliziranih programa, tzv. inteligentnih agenata, zadržavanja elemenata pretraživanja za buduća pretraživanja te prijenosa elemenata u druge sustave.

Spomenute ideje i pokušaji poboljšanja predmetnog pretraživanja u online katalogu, od uvođenja predmetnih preglednih kataložnih jedinica i uputnica, tezaurusu ili odnosa kakvi postoje u tezaurusu, osiguravanja pomoći pri prevođenju unesenih pojmova u rječnik korišten u katalogu i preoblikovanju upita pri neuspješnom pretraživanju, do osiguravanja mogućnosti prebiranja (Larson 1992; Hildreth 1989), prvi su omogućili eksperimentalni online katalogi treće generacije poput Okapia i CHESHIRE-a (Larson 1992). Njihove prednosti obuhvaćaju i mogućnost oblikovanja upita prirodnim jezikom bez potrebe za korištenjem složene sintakse, ispis rezultata prema sličnosti s upitom ili prema relevantnosti te podržavanje veće interakcije s korisnikom, i to i prilikom definiranja upita i prilikom korištenja povratne informacije o relevantnosti (relevance feedback) pri čemu se relevantni rezultati koriste za daljnje preoblikovanje upita i dobivanje sličnih jedinica građe. Istraživanja Okapia (Hancock-Beaulieu, Walker 1992) nadalje su pokazala prednosti dodatnih mogućnosti predmetnog pretraživanja koje se odnose na automatsko preoblikovanje upita (automatic query expansion), rangirani ispis rezultata prema ključnoj riječi (“best match” keyword searching), automatsko kraćenje, automatsko ispravljanje pravopisnih pogrešaka i pogrešaka u pisanju te sustav uputnica. Iduće istraživanje nove inačice, XOkapi, iz 1995. (Hancock-Beaulieu, Fieldhouse, Do 1995) obuhvatilo je i interaktivno preoblikovanje upita (interactive query expansion) na temelju povratne informacije o relevantnosti, jer su rezultati pokazali da automatsko preoblikovanje upita ostavlja korisniku nedovoljno nadzora nad preoblikovanjem upita i pretraživanja te da

implicitno korištenje tezaurusa ponekad ima nedostatke u odnosu na eksplicitno, kada korisnik sam odabire tražene pojmove.

4. Problematika korisničkih pristupa pri predmetnom pretraživanju

Svrha informacijskog sustava jest korisnicima omogućiti pristup relevantnim dokumentima. Korisnici se međusobno znatno razlikuju u pretraživanju informacija, pa su provedena mnoga istraživanja s ciljem boljeg razumijevanja načina na koji korisnici tragaju za informacijama. Tri su različite škole u informacijskoj i knjižničnoj znanosti koje povezuju korisnike s informacijom: fizička, kognitivna i konstruktivistička (Olson, Boll 2001, str. 265-273).

Fizička škola izrasla je iz empiricističke tradicije i usmjerena je na mjerenja informacije koje se prenesu iz dokumenata na korisnike te na mjerenje stupnja do kojeg informacija zadovoljava korisničke potrebe. Prema ovoj školi korisničke potrebe su konkretne i mogu se identificirati te se smatra da je moguće na njih odgovoriti potrebnim informacijama. Korisničke karakteristike poput, na primjer, računalnih vještina kojima vladaju, mogu se smatrati “šumom” kroz koju se može pronaći put ako se poduzmu primjerene mjere.

Kognitivna škola potječe iz racionalističke tradicije, prema kojoj je logika superiornija od iskustva. Vjeruje se da korisnici imaju individualne strukture znanja koje se temelje na pravilima i koje određuju ulazne i izlazne informacije na logičan način. Istraživanja se bave prepoznavanjem čimbenika koji utječu na interakciju korisnika i sustava, a posebno utjecajem obrazovanja i iskustva na oblikovanje strukture znanja pojedinaca. Ovoj školi pripada koncept ASK (Anomalous State of Knowledge – anomalijsko stanje znanja) N. Belkina (Belkin, Oddy, Brooks 1982), prema kojemu je korisnikovo znanje centralno pitanje u informacijskom sustavu. N. Belkin tvrdi da tzv. *best match* sustavi nisu prikladni jer od korisnika zahtijevaju da definira svoj informacijski problem. Informacijski sustav trebao bi nastojati korisnika sam ispitati i dokučiti njegovu informacijsku potrebu.

Konstruktivistička škola proizlazi iz tradicije historicizma, usmjerene na društvo i kulturu, a ne na pojedinca. Prema ovoj školi informacijske su potrebe pojedinca uvjetovane društvenim silama.

Dosadašnja istraživanja pokazala su da su karakteristike korisnika koje utječu na njegovo ponašanje vezano uz pretraživanje informacija dob i spol, obrazovanje i iskustvo te svrha istraživanja, istraživački stadij i slično. Međutim,

ovako široko postavljena problematika ne ulazi u opseg ovoga rada. Dio koji slijedi usmjeren je prvenstveno na ponašanje korisnika pri predmetnom pretraživanju i implikacijama za oblikovanje online kataloga za tu vrstu pretraživanja.

Većina današnjih kataloga zahtijeva od korisnika određeni napor. Naime, korisnik mora preoblikovati i ponovno unositi upit sve dok ne dobije zadovoljavajuće rezultate. Ovaj pristup temelji se na pretpostavci da korisnik zna što želi, da može opisati potrebnu informaciju u obliku koji je prihvatljiv računalnom sustavu, da može razumjeti poruke koje dobije od kataloga i riješiti probleme pri pretraživanju tako da na kraju dođe do željenih rezultata (Borgman 1996; Borgman 2000; Matthews 1997; Smeaton 1997). Pritom, kao što smo ranije istaknuli, većina problema u online katalogima postoji zbog ostataka strukture kataloga na listićima, koja se dobro ne preslikava u online okruženje, i nemogućnosti da se oblikuju sučelja koja se temelje na znanju i vještinama korisnika online kataloga. Online katalogi izrasli su iz kataloga na listićima, a istraživanja su pokazala da korisnici i u manualnim i automatiziranim sustavima dolaze do kataloga s nedovoljnom informacijom za bilo koju pristupnicu. Cutterova pravila ne odražavaju prikladno način na koji ljudi traže informacije. Iako je sam Cutter tvrdio da uvijek ima na umu korisnika, njegova perspektiva kao i perspektiva drugih teoretičara iz tog razdoblja odražava racionalan, pozitivistički pristup, a ne znanje o načinu na koji ljudi traže informacije.

Danas korisnici online kataloga imaju više mogućnosti pretraživanja no što im je to uopće poznato i no što ih koriste. U katalogima na listićima postojale su predodređene metode za pretraživanje – autor, naslov, predmet. Prvi online katalogi zahtijevali su od korisnika da znaju oblikovati strategiju pretraživanja. Druga generacija online kataloga približila se korisničkim potrebama, međutim, korisnici kataloga treće generacije još se uvijek susreću s problemom razvoja strategije za pretraživanje (Lombardo, Condic 2000, str. 141).

Prema C. Borgman (1996; 2000), postoje tri razine znanja potrebne za pretraživanje online kataloga. Konceptualno znanje odnosi se na prevođenje informacijske potrebe na upit za pretraživanje, npr. “Koje pojmove upotrijebiti ako tražim građu o osobama s posebnim potrebama”. Semantičko znanje je ono koje se primjenjuje pri odabiru mogućnosti sustava, npr. “Trebam li koristiti naredbu *Find* ili *Browse*?”. Sastoji se od tri parametra: akcije (pronaći, odabrati, pročitati, prebirati), pristupnice ili oznake polja (autor, naslov, predmet) i termina za

pretraživanja, i to pojedinačnih (npr. Grieg, fizika) ili u kombinaciji s Booleovim operatorima (npr. Barcelona OR Madrid). Tehničko znanje obuhvaća temeljne računalne vještine i sintaksu upita, npr. "Trebam li unijeti *Borges Jorge L.* ili *Jorge L. Borges*?". Slično tome, A. Large i J. Beheshti (1997) navode da korisnik mora prvo konceptualizirati potrebu za informacijom o određenom predmetu, zatim tu konceptualizaciju izraziti u određenim pojmovima za pretraživanje, te za te pojmove pronaći odgovarajuće termine, koji se koriste u online katalogima za označivanje predmeta.

Istraživanja ponašanja korisnika pri pretraživanju usmjerena su na semantičko znanje i tehničke vještine, a vrednovanje sustava najčešće se temelji samo na tehničkim vještinama (mjeri se uspješnost korisnika i pogreške u oblikovanju upita). Upute o korištenju online kataloga prečesto su usmjerene samo na tehničke vještine, bez da se te vještine stave u kontekst semantičkog znanja, a još manje konceptualnog (Borgman 1996).

C. Borgman (1996; 2000) također tvrdi da je korisnicima najteže odrediti termine za pretraživanje u predmetnom katalogu, jer ne znaju da su predmetne odrednice preuzete iz kontroliranog popisa ili tezaurusa koji je zasebno pretraživ. Umjesto toga, koriste ključne riječi koje najbolje poznaju, često metodom pokušaja i pogrešaka. Pretraživanje informacija problematično je jer je potrebno opisati informaciju koju nemate. Tražitelji moraju prevesti svoju informacijsku potrebu na opis informacije koju traže, oslanjajući se na njihovo znanje o problemu, poznavanje alata koje sustav pruža da pomogne u opisivanju problema, te, ako je moguće, na usluge vještih knjižničara ili posrednika koji vrše pretraživanje.

S jedne strane postoje sustavi koji zahtijevaju od korisnika da u potpunosti oblikuje upit za pretraživanje, pa takav sustav vraća skup zapisa koji odgovaraju upitu. S druge strane, postoje sustavi koji korisniku dopuštaju da unese bilo koje fragmente ideje kao početnu točku, osiguravajući razne alate kao pomoć u ispitivanju informacijske potrebe. Većina kataloga temelji se na modelu oblikovanja upita koji imaju određeni stupanj mogućnosti preoblikovanja pretraživanja, ali daleko su od idealnih sustava. Oblikovanje upita prema autoru i naslovu također je teško jer korisnik mora imati preciznu informaciju koja se u potpunosti poklapa s odrednicom u katalogu ili, pak, mora znati kako prepoznati druge moguće oblike. Predmetno pretraživanje još je teže jer korisnik treba iznaći način na koji će artikulirati namjeravano značenje korištenjem termina koji postoje u katalogu, bilo

da su to termini koje rabe autor, katalogizator ili osoba koja izrađuje predmetne pristupnice. O tome govori i D. Slone (2000, str. 761) i ističe da je odnos između onoga što korisnik upiše i načina na koji su jedinice označene, često slab ili nepouzdan, što dovodi do nezadovoljavajućih rezultata pretraživanja. Poseban problem u njenom istraživanju predstavljali su usko postavljeni, specifični upiti za koje je bilo potrebno utrošiti najviše vremena na pretraživanju. Kao uspješna tehnika pretraživanja pokazalo se postepeno sužavanje šire teme, uz služenje predmetnim odrednicama iz pronađenih zapisa kao termina za novo pretraživanje.

U online katalogima također se ne uzima u obzir to što korisnici upit oblikuju postepeno, već njihov dizajn odražava stajalište da korisnici oblikuju upit kojim je utvrđen cilj pretraživanja i da je svaka sesija pretraživanja nezavisna. Pretraživanje može obuhvatiti niz sesija s različitim izvorima, bilo online ili offline, pri čemu korisnik odabire neke od brojnih mogućnosti da bi došao do tražene informacije. Model postepenog pretraživanja koji bi se trebao primijeniti u online informacijskim sustavima prva je uvela M. Bates (1989a). Njezin model, tzv. *berrypicking* (sakupljanje bobica), u osnovi se temelji na ideji da rezultat pretraživanja nije odjednom pronađeni skup jedinica, već nekoliko izvršenih odabira dijelova informacije tijekom pojedinačnih stadija pretraživanja i stalnog preoblikovanja upita. Priroda upita je promjenjiva, ona se razvija, nije jedinstvena, pretraživanje ne vodi do jednog najboljeg skupa pronađenih rezultata, tehnika pretraživanja stalno se mijenja, kao što se i pronađeni izvori mijenjaju i u obliku i sadržaju. M. Bates (1989a) stoga sugerira da sustavi za pretraživanje trebaju osigurati što više različitih strategija za pretraživanje, što će rezultirati učinkovitijim pretraživanjem.

Pokazalo se da je Booleova logika teško razumljiv koncept (Yee 1991) te da čak i znanstvenici i inženjeri koji imaju iskustava u primjeni logike na drugim aplikacijama često pri pretraživanju koriste AND i OR obrnuto (Borgman, Case and Meadow 1989), a istraživanja iz kognitivne psihologije (Tversky, Kahneman 1974) pokazuju da ljudi slijede intuiciju pa AND i OR koriste u njihovom jezičnom značenju, tj. da je AND inkluzivno i da znači više, a OR da je ekskluzivno i da smanjuje skup – dakle, obrnuto od Booleove logike (Borgman 1996, str. 496). Za oblikovanje upita uz pomoć Booleovih operatora potrebno je konceptualno znanje o tome kako se termini mogu kombinirati i semantičko znanje o tome kako određeni sustav izvršava Booleove operatore. Online katalogi različito koriste operatore, posebno implicitne, što zbunjuje korisnike. Ako se unese više termina, neki sustavi

ih tretiraju kao AND, neki kao OR, a neki kao izraz. Neki poštuju logičku hijerarhiju, neki samo po redu izvršavaju, neki koriste zagrade, a neki ih ignoriraju. Samo najiskusniji tražitelji imaju dovoljno znanja o Booleovoj logici i o tome kako sustav tretira upite od više riječi. Krajnji korisnici uglavnom nemaju volje to ispitati, pa prenose znanje o sustavu koji su prije koristili (Borgman 1996, str. 498). U online katalozima s komandnim sučeljem (command-driven) postoje također velike razlike u načinu na koji se koriste naredbe i nazivi polja pa tako za autora nazivi u različitim sustavima mogu biti npr. *AU* i *AUT* (od *author*), *NA* (od *name*), *PA* (od *personal author*).

Kao što se pokazalo u prethodnom poglavlju, velik je broj neuspješnih predmetnih pretraživanja, kojima su razlozi raznoliki. Osim spomenutih, problem je i taj što korisnici nakon prvog neuspjeha najčešće odustaju od pretraživanja uopće. P. Le Loarer (1993, str. 251) navodi da korisnik, kada ne dobije rezultate, odustane od traženja jer smatra da informacija koju traži u katalogu ne postoji.

Važno je također istaknuti da često, čak i kad postoje mogućnosti za različite vrste predmetnog pretraživanja, korisnik toga ne mora biti ni svjestan. Zbog toga je potrebno osigurati upute o mogućnostima te ponuditi pomoć koja će korisnika uputiti na način kako da preoblikuje upit radi pronalaženja željenih informacija (Loarer 1993, str. 253; Hancock-Beaulieu 1993, str. 278).

Također valja upamtiti da je korisnička zajednica online kataloga heterogena po dobi, jeziku, kulturi, poznavanju područja, računalnim vještinama, i da će u njoj većina biti vječni početnici u informacijskom pretraživanju. Ako se promatraju samo korisnici visokoškolskih knjižnica, razvidno je da posjeduju različite razine znanja o korištenju knjižnica, od dodiplomskih studenata s malim ili nikakvim iskustvom, do znanstvenika koji su u pretraživanju knjižničnih kataloga vrlo iskusni (Rasmussen 1999, str. 410). Pokazalo se, također, da postoji snažna veza između akademskog smjera i provođenja pretraživanja. Tako, primjerice, osobe s tehničkim obrazovanjem imaju bolje rezultate pretraživanja – oni koji su studirali matematiku i prirodne znanosti bili su brži u pretraživanju nego oni iz humanističkih znanosti (Borgman 1989, str. 247-248).

Pokazalo se i da neki tražitelji imaju “stil” koji primjenjuju na većinu zadataka pretraživanja, dok drugi mijenjaju svoje ponašanje pri pretraživanju ovisno o zadatku. Neki tražitelji prvo dobro promisle o oblikovanju upita, dok drugi primjenjuju pristup pokušaja i pogrešaka, pri čemu je cilj pretraživanja bolje

oblikovanje upita. Neki tražitelji istražuju, a drugi se odlučuju za najizravniji put do cilja. Određene funkcije sustava problematične su većini, dok su druge problem samo nekolicini. Potrebno je provoditi još istraživanja o tome koji aspekti traženja informacija su univerzalni, a koji specifični za neku situaciju ako želimo oblikovati informacijske sustave koji će služiti heterogenoj skupini korisnika sa sve većim razlikama u korisničkim informacijskim potrebama (Borgman, Hirsh, Hiller 1996, str. 581).

S obzirom na različita znanja i potrebe korisnika jednog sustava, suvremena sučelja trebala bi biti prilagođena pojedincu. Takvo sučelje može se temeljiti na različitim vrstama informacija prikupljenih o korisniku, kao što su njegova znanja o određenom području, vještine pretraživanja kojima vlada, njegove tipične potrebe pretraživanja (preciznost (precision) ili odziv (recall); prebiranje ili oblikovanje upita, jezik, ako je sučelje višejezično i sl.), baze koje najčešće pretražuje s pripadajućim lozinkama i dr. Takvo bi sučelje trebalo sadržavati strategije pretraživanja i dobivene rezultate iz prethodnih pretraživanja. Relevantne informacije o korisniku mogu se prikupljati na razne načine. Na primjer, pri prvom ulasku u sustav može se izraditi korisnički profil na osnovu podataka koje korisnik ispuni u priređenom obrascu, ili se podaci mogu prikupljati praćenjem korisnikovih akcija za vrijeme pretraživanja te na temelju upita koje oblikuje (Vickery, Vickery 1993, str. 114-115).

C. Borgman (1996) zaključno navodi da bi dugoročan cilj trebao biti oblikovanje intuitivnih sustava u kojima se dodatna pomoć može svesti na najmanju moguću mjeru, dok se, kratkoročno gledano, može osigurati dodatno obučavanje i upute o pretraživanju, imajući na umu da to nije zamjena za dobro oblikovan sustav. Dugoročno gledano, informacijski stručnjaci trebali bi oblikovati sustave koji se temelje na bihevioralnim modelima postavljanja pitanja. Korisniku je pri oblikovanju upita potrebno osigurati pomoć, kao i pri pregledavanju rječnika, tim više što je korisnik na Internetu prepušten sam sebi, ali pomoć i upute ne mogu zamijeniti dobar intuitivan dizajn, već sustavi moraju biti tako oblikovani da predvide svaku moguću situaciju i potrebnu pomoć pri iznalaženju pravog pristupa. Također je potrebno osigurati mogućnost preoblikovanja pretraživanja uz pomoć specijaliziranih programa, tzv. inteligentnih agenata, zadržavanja djelomičnih rezultata za buduće pretraživanje, te prenošenje segmenata pretraživanja u druge sustave (Borgman 1996).

Osim sučelja i mehanizama predmetnog pretraživanja, i same je kontrolirane rječnike potrebno oblikovati prema potrebama korisnika i prema načinima na koje će korisnik pronalaziti informacije, o čemu će više govora biti u poglavlju koje slijedi.

5. Predmetni pristup informacijama u online knjižničnim katalozima

5.1. Vrste predmetnih pristupnica

Pronalaženje informacija temelji se na interakciji korisnika s bazom informacijskog sustava. Korisnik postavlja upit koji se tipično sastoji od riječi koje se uspoređuju s pristupnicama u bazi podataka. Osim ovoga načina, informacijski sustavi omogućuju odabiranje pristupnica iz ponuđenog popisa, koja se tehnika pretraživanja naziva *prebiranjem* ili *pregledavanjem* (browsing). Postoje i napredne tehnike pretraživanja uspoređivanjem riječi uz upita s pristupnicama u bazi podataka, poput tzv. *latentnog semantičkog označivanja* (latent semantic indexing), kojom se riječi iz upita s predmetnim pristupnicama u bazi podataka uspoređuju na temelju statističke semantičke analize upita i odnosa među riječima u upitu. Ovaj se način razlikuje od prethodnih po tome što omogućuje pronalaženje relevantnih jedinica građe i kada se riječi iz upita i pristupnice razlikuju morfološki, ali se preklapaju semantički (Hjorland, Nielsen 2001, str. 251).

U online knjižničnim katalozima pristupnica može biti bilo koji element iz zapisa. Danas su najčešće predmetne pristupnice one za predmetnu odrednicu, klasifikacijsku oznaku, deskriptor i ključnu riječ, koja može biti iz jednog, nekoliko ili svih polja iz zapisa odjednom, npr. iz naslova, predmetne odrednice, napomene o sadržaju ili, rjeđe, iz anotacije¹². Predmetne pristupnice u računalnom okruženju mogu biti i podaci o formalnim obilježjima građe, pa je podjela između predmetnih i ostalih vrsta pristupnica danas teško određiva. U literaturi se donedavno isticala i razlika između predmetnih pristupnica i opisnih podataka (descriptive data). Opisni podaci omogućavali su traženje poznatih jedinica, a predmetni podaci traženje poznatih ili nepoznatih jedinica na danu temu. Danas se, međutim, jasna podjela između predmetnih i opisnih podataka gubi (Hjorland, Nielsen 2001, str. 251). Tako su, na primjer, naslovi istodobno i opisni podaci i predmetne pristupnice. Ono što se može prihvatiti kao valjano polazište je to da su najvrednije predmetne pristupnice

¹² Vidi, npr., WebPAC Knjižnica grada Zagreba, URL: <http://www.kgz.hr>.

one koje korisniku omogućavaju da u bazi informacijskog sustava pronađe najrelevantnije dokumente.

U predmetnom su pretraživanju kao pristupnice vrlo korisne ključne riječi iz termina kontroliranih rječnika, poput predmetnih odrednica i deskriptora. Ovaj način pretraživanja korisnicima omogućava da pronađu odobrene kontrolirane termine koje ne poznaju. Za pronalaženje kontroliranih termina osigurava se i sustav uputnica, čija je izrada, međutim, u usporedbi s ključnim riječima, zahtjevna i skupa, ali svakako preporučena. Vrlo su korisne i ključne riječi iz ostalih polja za sadržaj, poput naslova, sažetka i napomena, čime se omogućava pronalaženje po drugim aspektima no što su oni pokriveni kontroliranim terminima. Primjerice, ključne riječi iz ovih polja mogu biti specifičnije od dvije ili tri predmetne odrednice koje služe opisu cjelovitog djela. Da bi se postigli najbolji rezultati pretraživanja po ključnim riječima, potrebno je koristiti operatore i kraćenje.

Kada se govori o ključnim riječima iz naslova kao predmetnim pristupnicama, valja naglasiti da ova vrsta predmetnog pretraživanja ima nekoliko nedostataka. Iako se nepunoznačne riječi (stop-riječi poput članova, prijedloga i veznika) ne uvrštavaju u indeks riječi iz naslova, rezultati ove vrste pretraživanja često su vrlo niske razine preciznosti. Za to postoji nekoliko razloga. Najveći je problem taj što autori naslove često izražavaju metaforičkim jezikom, kako bi pobudili zanimanje potencijalnih čitatelja. Usto, naslovi često izražavaju općenitije aspekte, no što su u radu zapravo obrađeni. Ako govorimo o knjižničnim katalogima koji popisuju različite vrste građe, još je jedan problem s naslovima taj što brojni dokumenti, poput slika i pisama ili poruka zapravo ni nemaju naslov (Voorbij 1998).

P. Hodges (1983), pak, smatra da je, iako se pretraživanjem po riječima iz naslova pronalazi manje od 50% relevantnih naslova iz prethodno navedenih razloga, takva preciznost često je dovoljna korisniku visokoškolskih knjižnica. Kao razlog tome navodi da studenti i nastavnici najčešće ne traže detaljan popis jedinica građe o određenoj temi, već samo žele steći uvid u količinu raspoložive građe o toj temi ili dobiti osnovne informacije.

Na predmetno pretraživanje po sažecima utječe njihova kvaliteta. Sažeci koje izrade autori smatraju se informativnima jer nisu objektivan prikaz rada, za razliku od tzv. indikativnih sažetaka koje izrađuju informacijski stručnjaci i koji su stoga kritični. Informacijski stručnjaci koji priređuju sažetke mogu koristiti

tezauruse, klasifikacijske sustave i slična pomagala te se na taj način sadržajna analiza i terminologija u sažecima može formalizirati (Fidel 1986). Sažeci imaju znatno veći utjecaj na povećanje razine odziva i preciznosti, nego naslovi i ključne riječi (Hjorland, Nielsen 2001, str. 271).

Iako knjižnični katalogi uglavnom ne sadrže sažetke, formati MARC osiguravaju polje za sažetak. Tako u UNIMARC-u postoji polje 330, predviđeno za kratki sadržaj ili sinopsis jedinice. Ovo je polje izborno, a može sadržavati svaku vrstu sažetka, bilo informativan ili indikativan sažetak, kritički osvrt ili procjenu.

Osim sažetka, u literaturi se često navodi obogaćivanje bibliografskih zapisa stranicama sa sadržajem. Zbog razmjerno visokih troškova ovako iscrpne obrade knjižnične građe, rijetki su katalogi koji ove predmetne pristupnice doista sadrže. Međutim, danas postoji mogućnost preuzimanja ovih podataka iz drugih izvora na Mreži, poput online knjižara, čiji su zapisi obogaćeni i drugim potencijalno korisnim podacima. T. Koch (2002) sugerira da bi možda bilo dovoljno samo povezati ISBN iz zapisa u knjižničnom katalogu s ISBN-om u zapisu u online knjižari i na taj način omogućiti pristup tim podacima, bez potrebe za preuzimanjem čitavog zapisa. Ukoliko se u tom slučaju želi omogućiti pretraživanje polja čijem sadržaju je samo omogućen pristup, potrebno je implementirati tražilicu ili osmisliti druge tehnike kojima će se takvo pretraživanje osigurati.

U ovome se smjeru na međunarodnoj razini i kreće suradnja knjižničarske zajednice s nakladničkim i knjižarskim zajednicama. Tako su Britanska knjižnica (British Library) i OCLC izradili tablice za konverziju nakladničkog i knjižarskog formata ONIX u knjižnične formate UNIMARC i MARC 21. Sam format ONIX razvijan je u suradnji s knjižničarskom zajednicom te se, između ostalih dokumenata, temelji na FRBR-u (Green 2001; Danskin 2001). Danas zapise ONIX-a već preuzima Kongresna knjižnica (Library of Congress), a isto su najavile i britanska, njemačka i švedska nacionalna knjižnica.

O uključivanju potpunog teksta u katalog postoje različita stajališta, a A. Barbarić (2002, str. 43) tvrdi da se potpuni tekst ne bi trebao uključivati u katalog jer ostale jedinice građe za koje potpuni tekst ne postoji, ne bi bile pretražive po tom polju. Stoga je, smatra, dovoljno iz kataloga osigurati pristup potpunom tekstu dokumenta. U ovome će se radu, s obzirom na opsežnost problematike pretraživanja

po potpunom tekstu, ograničiti samo na najosnovnije¹³. Istraživanja provedena na informacijskim sustavima s potpunim tekstom, poput komercijalnih baza podataka, najviše su usmjerena na dodane oznake (value-added tags), koje prema kontroliranim rječnicima za označivanje unose informacijski stručnjaci. Pokazalo se, naime, da pretraživanja po potpunom tekstu imaju nižu razinu preciznosti od pretraživanja po sažetku i kontroliranom rječniku pa su ove oznake vrlo korisne (Tenopir, Ro 1990). Također se puno istražuje o ulozi računalne obrade prirodnog jezika te o pretraživanju po određenim dijelovima potpunog teksta dokumenta, tj. po pojedinim odlomcima ili sadržajnim cjelinama kao što su uvod, metodologija i zaključak. Smatra se da su sažeci optimalni za potrebe pretraživanja jer pretraživanje po potpunom tekstu često dovodi do velike količine nerelevantnih rezultata, odnosno niskom preciznošću (Pinto, Lancaster 1999).

Današnje je stajalište struke da su kontrolirani jezici (s predmetnim pristupnicama poput predmetnih odrednica, deskriptora i klasifikacijskih oznaka) i prirodni jezici (ključne riječi) međusobno komplementarni u predmetnom pretraživanju, jer svaki ima svoje prednosti i nedostatke. Relativne snage i slabosti različitih vrsta jezika za označivanje i pretraživanje utvrđuju se istraživanjima na temelju svojstava i potreba za različite vrste dokumenata, građe, stručnih područja i korisničkih zajednica. O ovoj problematici više će biti govora u poglavlju 5.5.

5.2. Vrste sustava za označivanje i pretraživanje

Na uspješnost predmetnog pretraživanja utječu sustavi za označivanje i pretraživanje. Postoje različite podjele ovih sustava. U ovome radu rabiće se podjela H. Olson i J. Bolla (2001, str. 31-51), prema kojima se sustavi za označivanje i pretraživanje dijele u četiri osnovne skupine.

Prvu skupinu obilježava to što se pretraživanje, za razliku od ostale tri skupine, temelji na netekstualnoj osnovi (nonverbal, nontextual), poput pretraživanja po slikama, glazbenim melodijama, kemijskim formulama i geografskim koordinatama, iako se takva vrsta pretraživanja često kombinira s

¹³ Pretraživanjem potpunog teksta danas se intenzivno bave konferencije TREC (Text Retrieval Conference), URL: <http://trec.nist.gov>.

tekstualnim opisom objekta. Budući da današnji online katalogi ne osiguravaju pristup bazama podataka s netekstualnim informacijama i da za ovu vrstu netekstualnog pretraživanja još ne postoje standardi ni dovoljno razvijene tehnologije, u tu se problematiku u ovome radu neće detaljno zalaziti. U prvu skupinu ubraja se i tzv. citatno označivanje (citation indexing), a temelji se na citiranju jednog rada u drugim radovima. Ono zapravo nije sustav za označivanje i pretraživanje u pravom smislu te riječi, već je to sustav bibliografskih referenci, utemeljen na pretpostavci da se dokumenti koji citiraju druge dokumente bave istim predmetom. Ono omogućava i jednostavno praćenje razvoja nekog predmeta u stručnoj literaturi, a intenzivno se koristi u području prava, u svom osnovnom obliku još od dvanaestog stoljeća. U druge se znanosti proširilo dolaskom računala, preciznije 1961., kada je američki Institut za znanstvene informacije (Institute for Scientific Information – ISI) počeo izdavati Citatno kazalo za prirodoslovlje (Science Citation Indeks)¹⁴.

Druga vrsta sustava za označivanje i pretraživanje su nekontrolirani ili prirodni sustavi, a temelje se, dakle, na riječima prirodnog jezika, u kojima se jedini nadzor nad semantikom i sintaksom obavlja uz pomoć stop-liste. Preostale dvije vrste sustava za označivanje i pretraživanje čine kontrolirani sustavi, i to oni koji se temelje na kontroliranim terminima iz prirodnog jezika, tzv. abecedni sustavi (alphabetical systems) (treća vrsta), ili, pak, na klasifikacijskim oznakama koje se sastoje od brojeva i slova, tzv. klasifikacijski sustavi (classed systems) (četvrta vrsta).

Nekontrolirani sustavi, koji se temelje na riječima preuzetima iz dokumenta, počeli su se koristiti 1950-ih. Tada se M. Taube zalagao za tzv. sustav uniterma, odnosno termina koji se sastoje od samo jedne riječi, smatrajući da se višesložni termini mogu oblikovati korištenjem Booleovih operatora u trenutku pretraživanja. Ubrzo se pokazalo da unitermi dovode do niske preciznosti u rezultatima i da su neprimjereni za predstavljanje glavnih pojmova određene discipline i strukturiranje rječnika te discipline (Svenonius 2000, str. 134). Sustav uniterma, indeksi ključnih riječi iz naslova poznati pod nazivima KWAC (keyword-alongside-context), KWIC (keyword-in-context), i KWOC (keyword-out-of-context), s vremenom su prerasli u ključne riječi iz teksta čitavog dokumenta, uključujući naslov, sažetak, predmetne

¹⁴ Thomson ISI, URL: <http://www.isinet.com>

odrednice i ostala polja bibliografskog zapisa. Danas se u online katalozima za sve njih rabi samo termin *ključna riječ*.

U abecedne sustave spadaju popisi predmetnih odrednica ili predmetni sustavi, kojima u elektroničkom okruženju kao izvor termina služi datoteka predmetnih preglednih kataložnih jedinica i uputnica ili tezaursus, te deskriptori, kojima kao izvor termina služi tezaursus. Temeljna razlika između ove dvije vrste sustava za označivanje i pretraživanje jest u tome što se predmetnim odrednicama u pravilu opisuje glavni predmet dokumenta, a deskriptorima pak pojedini aspekti predmeta dokumenta. Deskriptori mogu biti samostalne, pojedinačne riječi ili prekoordinirani termini, koji se sastoje od dvije ili više riječi u predodređenom obliku. Obično se dodjeljuje jedna do tri predmetne odrednice, te između pet i dvadeset i pet deskriptora (Olson, Boll 2001, str. 37). No, Kongresna je knjižnica 1996. uvela praksu prema kojoj je dozvoljeno do deset predmetnih odrednica radi potpunijeg opisa sadržaja, čime se gubi razlika između deskriptora i predmetnih odrednica.

U klasifikacijske sustave ubrajaju se enumerativni i sintetički klasifikacijski sustavi. Izvor oznaka su im klasifikacijske tablice. Enumerativni klasifikacijski sustavi, poput Deweyeve decimalne klasifikacije – DDK (Dewey Decimal Classification – DDC) i Klasifikacije Kongresne knjižnice – KKK (Library of Congress Classification – LCC), u tablicama popisuju sve dozvoljene oznake, za razliku od sintetičkih klasifikacijskih sustava, poput Univerzalne decimalne klasifikacije – UDK (Universal Decimal Classification – UDC), čije tablice popisuju elemente koji se naknadno kombiniraju u klasifikacijske oznake. Najbolje sintetičke klasifikacijske tablice su fasetne, poput Klasifikacije s dvotočkom – KD (Colon Classification – CC), u kojima su predmeti podijeljeni na sastavne elemente sa pojedinačnih aspekata.

Osnovna razlika u strukturi predmetnih i klasifikacijskih sustava jest ta što su u predmetnom sustavu termini poredani abecednim slijedom po početnim slovima riječi, a u klasifikacijskom su sustavu okupljeni po skupinama i podskupinama te hijerarhijama po sličnosti i srodnosti. No, i u predmetnom sustavu donekle postoji okupljenost po sličnosti, što može biti “slučajno”, tj. da i termini okupljeni po abecednom načelu jesu semantički povezani (npr. *maloljetnost*, *maloljetničko prijestupništvo*), ili “namjerno”, tj. oblikovanjem predmetnih odrednica s predmetnim pododrednicama (npr. *knjižničari – obrazovanje*).

Postoje i djelomično kontrolirani sustavi za označivanje i pretraživanje (semicontrolled). Tu spadaju “obogaćeni” *KWIC* i *KWOC* sustavi (“enriched” *KWIC* or *KWOC*), koji su “obogaćeni” imenima ili drugim dodanim kontroliranim terminima; te sintaktički sustavi za označivanje i pretraživanje (string indexing systems). Sintaktički sustavi temelje se na ideji da je predmetna odrednica rečenica pa se označivanje obavlja tako što se analiziraju različiti aspekti predmeta o kojima se u dokumentu radi i bilježe se odobrenim terminima i tzv. operatorima uloga (role operators), na temelju kojih računalo generira jedan niz riječi koji prikazuje sadržaj dokumenta. Zatim računalo automatski preoblikuje ovaj niz tako da je i svaki značajan termin zasebno pretraživ preko indeksa. Zbog pokrivanja pojedinih aspekata koji su zasebno pretraživi, i ovi se sustavi ubrajaju u fasetne. Primjeri takvih sustava su britanski PRECIS (Preserved Context Indexing System - Sustav za predmetno označivanje sa sačuvanim smislom konteksta), indijski POPSI (Postulate-based Permuted Subject Indexing – Predmetno označivanje s preokretom, utemeljeno na postulatima) i hrvatski Sintaktički sustav za označivanje predmeta (SSZOP). Iako ovi sustavi do sada nisu zaživjeli, njihova svojstva vrlo su poželjna u elektroničkom okruženju.

U online katalozima i bazama podataka klasifikacijski, abecedni sustavi i ključne riječi danas se koriste integrirano, te je u pojedinim online katalozima i bazama podataka razlika između ova tri sustava korisniku nevidljiva, pružajući povećane mogućnosti pretraživanja. Integrirano korištenje različitih sustava poželjno je jer svaki od njih omogućava pretraživanje sa drukčijih aspekata.

5.2.1. Podjela sustava za označivanje i pretraživanje prema prethodnom i naknadnom označivanju

Prekoordinirani sustavi ili sustavi s prethodnim povezivanjem (precoordinated) su oni u kojima se termini oblikuju u trenutku označivanja. Postkoordinirani sustavi ili sustavi s naknadnim označivanjem (postcoordinated) zahtijevaju da korisnik u samome trenutku pretraživanja oblikuje jedan “termin” koji se sastoji od niza riječi i predstavlja traženi koncept. Ovaj se termin sastoji od termina koji su u sustavu međusobno nepovezani, iako sami mogu biti prekoordinirani (višečlani termini).

Postkoordinacija se pojavila 1940-ih, kada su nastali prvi sustavi utemeljeni na ključnim riječima poput uniterma i koristile se tehnologije preteče računalnim tehnologijama, poput sustava tzv. optičkog podudaranja (optical coincidence) ili *peek-a-boo* sustava. Mogućnosti postkoordinacije do izražaja su došle pojavom računala, a posebno razvojem online kataloga.

Sintetički i fasetni klasifikacijski sustavi te sintaktički sustavi za označivanje i pretraživanje su prekoordinirani sustavi, jer se elementi povezuje u trenutku označivanja. No, online katalogi imaju tu prednost da mogu osigurati postkoordinaciju, tj. povezivanje u trenutku pretraživanja nekoliko prekoordiniranih termina ili oznaka, kao i omogućiti pretraživanje samo po jednom elementu klasifikacijske oznake ili odrednice sintaktičkog sustava.

Enumerativni klasifikacijski sustavi i abecedni sustavi su sustavi s ograničenom ili djelomičnim prekoordinacijom. Određeni termini ili oznake mogu i ne moraju biti prethodno povezani, već ovisno o korištenim terminima odnosno oznakama i pravilima sustava.

Nekontrolirani sustavi koji se temelje na ključnim riječima su postkoordinirani sustavi. U njima se svaka ključna riječ tretira kao jedan termin, iako pojedine ključne riječi mogu biti prekoordinirane.

U trenutku traženja moguće je naknadno povezivati termine ili oznake (bilo da su oni kontrolirani ili nekontrolirani, prethodno povezani, abecedni ili enumerativni), ako online katalog to omogućava. Online katalogi često omogućuju pretraživanje po ključnim riječima iz predmetnih odrednica. Također je moguće pretraživati po pojedinim elementima kontroliranih termina ili oznaka, ukoliko to sam sustav za označivanje omogućava i sam online katalog podržava.

5.3. Kontrolirani sustavi u predmetnom pretraživanju

Kontrolirani sustavi omogućavaju označivanje i pretraživanje i po konceptima ili pojmovima¹⁵, a ne samo po riječima. Kontrolirani sustavi “kontrolirani” su do različitih stupnjeva. Nastali su iz potrebe za rješavanjem sinonimije, homonimije,

¹⁵ Pojam se određuje kao misao o bitnim karakteristikama onoga o čemu mislimo.

izraza i različitih morfoloških oblika riječi prirodnog jezika, s ciljem da se za pojedine koncepte ili pojmove iz dokumenta osigura jedinstven jezični oblik. U kontroliranom jeziku za svaki pojam postoji samo jedan termin ili oznaka, određen je opseg tog termina ili oznake, hijerarhijski odnosi i odnosi srodnosti mogu biti eksplicitno prikazani, uspostavljena je kontrola nad različitim načinima pisanja i eksplicitno su identificirani koncepti izraženi homonimima upotrebom izraza, pridjeva, kvalifikatora ili precizne terminologije. U računalnim katalogima korisnika se može uputiti s riječi koju je unio u upit na odobreni termin u kontroliranom sustavu, ili se pretraživanje može automatski preusmjeriti na termin iz kontroliranog sustava. Ova druga mogućnost nije preporučena jer korisnici, pokazalo se, žele i trebaju imati nadzor nad pretraživanjem. Prikaz semantičkih odnosa među terminima korisniku može pomoći pri izboru termina ili pri promjeni upita nakon nezadovoljstva dobivenim rezultatima. U svakom je slučaju preduvjet za uspješno predmetno pretraživanje u online katalogima datoteka predmetnih kataložnih jedinica i uputnica povezana s bibliografskom datotekom, s razrađenom strukturom uputnica.

Kontrolirani sustavi u online knjižničnim katalogima mogu biti klasifikacijski sustavi, popisi predmetnih odrednica, tezaursi i sintaktički sustavi. E. Svenonius (2000, str. 131-132) rabi podjelu kontroliranih jezika na temelju toga koristi li jezik sintaksu. Jezike koji imaju sintaksu naziva sintetičkima, a ostale enumerativnima. Većina sintetičkih jezika ima, dakle, nešto što odgovara rečeničnom subjektu i predikatu. Primjerice, kategorija *Key system* u PRECIS-u i *Personality* u Klasifikaciji s dvotočkom odnose se na ono o čemu se govori (subjekt). Oznaka koja određuje odnos različitih značenjskih elemenata i pobliže određuje značenje je operator odnosa (relator) (predikat). Operatori odnosa u PRECIS-u, sintaktičkom sustavu za označivanje, zovu se operatori uloga. Oni uklanjaju višeznačnost, određuju redosljed jezičnih elemenata i pokazuju odnose među njima. U Univerzalnoj decimalnoj klasifikaciji, sintetičkom klasifikacijskom sustavu, kao operatori odnosa koriste se matematički simboli i interpunkcija, koji razjašnjavaju značenje jezičnog elementa, određuju njihov redosljed i omogućuju razlikovanje inače identičnih oznaka. Operator odnosa u formatu MARC 21 osigurava prije nemoguće postkoordinirano pretraživanje pojedinih elemenata predmetne odrednice ili klasifikacijske oznake, iako danas gotovo da nema sustava koji ovu prednost doista koristi (Olson, Boll, 2001, str. 65). U poljima 6XX za

podatke o sadržaju, potpolja imaju funkciju da odijele elemente prekoordiniranog jednočlanog ili višečlanog termina, ali i da određuju kategoriju svakog potpolja. Tako je u poljima za predmetne odrednice potpolje \$x\$ za tematsku, \$y\$ za vremensku, a \$z\$ za zemljopisnu podjelu. U UNIMARC-u je potpolje \$x\$ tematska, \$y\$ zemljopisna, a \$z\$ vremenska podjela. Primjer za publikaciju *Podrška demokraciji* G. Sorosa glasi

606 \$aDemokracija\$x\$socijalizam, kriza\$yEuropa, istočna\$z20. st.

(606 je polje za tematsku predmetnu odrednicu, \$a\$ je potpolje za prvi element unosa, odnosno za pojam u obliku koji propisuje sustav predmetnih odrednica) (Priručnik za UNIMARC 1999, str. 606-2). Računalno okruženje, dakle, omogućava pretraživanje po svakom od ovih aspekata na način da pretražuje svako od ovih potpolja, a ne samo jednu odrednicu, kao što je to bio slučaj u papirnom okruženju. Na ovaj način prekoordinirani i djelomično prekoordinirani sustavi preuzimaju svojstva postkoordiniranih sustava. Ovi se aspekti također u literaturi nazivaju fasetama, pri čemu se naglasak stavlja na velik potencijal oblikovanja odrednica ili oznaka uz pomoć jasno definiranih faseta, međusobno neovisnih ili povezanih preciznim sintaktičkim pravilima, dostupnih pri računalnom pretraživanju po bilo kojem redosljedu, po hijerarhijskim ili abecednim kriterijima, po obliku, vremenskom razdoblju ili drugim kriterijima.

S potrebom za integriranjem i usklađivanjem pojedinih informacijskih sustava, koji koriste različite kontrolirane sustave za označivanje i pretraživanje, oblikuju se tezaursi kao alati za povezivanje ovih sustava. Osnovna prednost tezaurusa jest ta što su u njemu precizno određeni odnosi ekvivalentnosti, srodnosti i hijerarhije, za što se koriste simboli BT (broader term – širi naziv, ŠN), NT (narrower term – uža naziv, UN), RT (related term – srodni naziv, SN), UF (use for – koristiti umjesto nekog drugog naziva) i USE (umjesto ovoga, koristiti neki drugi, određeni naziv). Usporedbe radi, u sustavima predmetnih odrednica odnosi su određeni samo uputnicama *vidi*, koje se koriste za odnose ekvivalentnosti, i uputnicama *vidi i*, koje se koriste za hijerarhijske odnose i odnose srodnosti.

Proteklih nekoliko desetljeća veće knjižnice stavljale su primjerke tiskanih LCSH uz katalog na listićima, međutim, kao problem se pokazalo to što su takvi sustavi bili namijenjeni prvenstveno knjižničarima i svojim su kraticama,

objašnjenjima i stručnom terminologijom zbunjivali korisnike. U 1980-im godinama M. Bates i drugi počeli su predlagati oblikovanje tezaurusa za korisnike (end-user thesauri, user thesauri), za pretraživanje, a ne za označivanje, koji bi sadržavali puno više nedeskriptora po kojima korisnici mogu tražiti neki pojam (Osłon, Boll 2001). Ovi tezaursi povezuju riječi korisnika na temelju korisničkog jamstva ili jamstva informacijskoga zahtjeva s rječnikom u bazi, bilo kontroliranim ili nekontroliranim. Usporedbe radi, J. Aitchinson, A. Gilchirst i D. Bawden (2000, str. 1-3) razlikuju tri osnovne vrste tezaurusa: klasični tezaurs (classic thesaurus), koji se upotrebljava i za označavanje i za pretraživanje, tezaurs za označivanje (indexing thesaurus), koji se koristi samo za označivanje, čija je prednost ta što sadrži velik broj sinonima i užih naziva kojima se mogu obogatiti bibliografski zapisi da bi se povećao broj pristupnica, te tezaurs za pretraživanje (searching thesaurus), koji se koristi samo za pretraživanje, a njegova je specifičnost ta što pomaže u pretraživanju baza podataka predlažući dodatne termine po kojima se pretražuje, posebice sinonime i uže pojmove. Tezaurs za pretraživanje koristi se kada se zapisi ne obogaćuju dodatnim pristupnicama, a korisnika se na dodatne termine može uputiti interaktivno, tako da mu se ponudi dodatni popis termina na odabir, ili automatski (Bawden 2001, str. 185).

Godine 1994. godine sastavljen je za online okruženje Tezaurs za umjetnost i arhitekturu (Art and architecture thesaurus), na temelju određenog broja dotadašnjih, često uže specijaliziranih popisa termina. Američka Nacionalna medicinska knjižnica (National Library of Medicine) objavila je 1990. metatezaurs Jedinostveni medicinski jezični sustav (Unified Medical Language System) koji se temelji na više od trideset kontroliranih rječnika iz područja medicine. Metatezaursi omogućuju konverziju termina između različitih tezaurusa, a ovaj uključuje i druge rječnike kao što su Medicinske predmetne odrednice MeSH (Medical Subject Headings), tezaurs koji se koristi za označivanje baza podataka *Medline* i *SNOMED* (detaljna nomenklatura patologije), i Međunarodna klasifikacija bolesti (International Classification of Diseases) (Bawden 2001, str. 185). Sam MeSH ima integrirane termine iz sustava predmetnih odrednica i klasifikacijskih oznaka, što zbog neodgovarajućih odnosa između termina dvaju različitih sustava koji su se razvijali neovisno jedan o drugome inače nije jednostavno postići.

Razvoj tehnologija poput XML-a pospješuje pristup različitim kontroliranim rječnicima putem jednog sučelja (vidi, npr., Chaudiron, Ihadjadene, Role 2002;

Light, R. 2001), što je alternativa izradi skupih integriranih kontroliranih sustava poput MeSH-a i Jedinственog medicinskog jezičnog sustava.

Iako se sintaktički sustavi danas gotovo uopće ne koriste, njihova fleksibilnost i mogućnost pristupa svakom aspektu predmeta nekog dokumenta čine ga dobrim kandidatom za online katalog, što bi zahtijevalo određenu razinu prilagodbe (Olson, Boll 2001, str. 191).

5.3.1. Klasifikacijski sustavi

U klasifikacijskim sustavima predmeti dokumenata okupljeni su prema sličnosti i srodnosti po skupinama i podskupinama, odnosno hijerarhijama. U klasifikacijske sustave ubrajaju se enumerativni i sintetički klasifikacijski sustavi. Enumerativni klasifikacijski sustavi, poput Deweyeve decimalne klasifikacije i Klasifikacije Kongresne knjižnice, u tablicama imaju popisane sve dozvoljene elemente, za razliku od tablica sintetičkih klasifikacijskih sustava u kojima su popisani samo elementi koji se kombiniraju u klasifikacijske oznake. To omogućava znatno fleksibilnije izražavanje i prikaz različitih složenih koncepata. Najbolje sintetičke klasifikacijske tablice su one fasetne, u kojima su predmeti prikazani sa svih važnih stajališta, a svako stajalište je jedna faseta kojoj je pridružen jedan simbol, na temelju kojih se zatim slažu primjerene oznake. Svaka faseta ima vlastitu hijerarhiju, svojstva i oznake pa je tipična fasetna klasifikacija multihijerarhijska. U tome je očigledna razlika od ostalih sintetičkih klasifikacija, u kojima se hijerarhije ne mogu međusobno jednostavno razlikovati po prepoznatljivim oznakama. Što ima više faseta, to je precizniji prikaz sadržaja, ali je oznaka duža.

Knjižnični klasifikacijski sustavi desetljećima su se koristili za organizaciju bibliografskih izvora, a najviše za raspored građe na policama. Od nedavna se provode brojna istraživanja o tome kako poboljšati knjižnične klasifikacijske sheme radi usklađivanja s potrebama novih informacijskih okruženja. Od 1960-ih klasifikacijski sustavi počeli su se ponovno razmatrati kao sustavi za označavanje i pretraživanje u elektroničkom okruženju i online katalogima. U 1980-im i 1990-im online katalozi počeli su razvijati pristupe i metode za odabir termina iz abecednog

popisa koji bi ih zatim doveo do stručnih oznaka, po kojima bi mogli nastaviti pretraživanje (Olson, Boll 2001, str. 76).

Prva istraživanja klasifikacijskih sustava uglavnom su bila usmjerena na označivanje, a s pretraživačkim aspektom detaljno se prvi put bavio projekt OCLC-a pod vodstvom K. Markey, započet 1983. (Olson, Boll 2001, str. 76). Bibliografski zapisi obogaćeni su prijevodima klasifikacijskih oznaka iz Deweyeve decimalne klasifikacije, pri čemu je neke bilo potrebno prilagoditi krajnjim korisnicima. Pokazalo se da su rezultati predmetnog pretraživanja poboljšani. Godine 1993. objavljena je i elektroničkom okruženju prilagođena inačica Univerzalne decimalne klasifikacije, tzv. *Master Reference File* (MRF), a ubrzo i Klasifikacije Kongresne knjižnice i Deweyeve decimalne klasifikacije.

Prvi i jedini opći fasetni klasifikacijski sustav oblikovao je S. R. Ranganathan 1920-ih godina, pod nazivom Klasifikacija s dvotočkom, koja se temelji na pet faseta. Kasnije je razvijeno nekoliko specijaliziranih fasetnih klasifikacija poput Klasifikacije knjižnične i informacijske znanosti (Classification of Library and Information Science) i ostalih sustava koji se temelje na fasetama, poput sustava za lančano označivanje (chain indexing systems) i sintaktičkih sustava. Još je S. R. Ranganathan za fasetnu klasifikaciju razvio abecedno lančano kazalo (chain index) radi olakšavanja pristupa stručnim oznakama i fasetama koje oznaku sačinjavaju. Lančana kazala sastoje se od bibliografskih opisa poredanih u stručnom slijedu pod stručnom oznakom faseta iz stručne oznake djela koje opisuju, a ove su oznake putem uputnica izraženih riječima povezane s višim hijerarhijskim razinama složenog predmeta.

Tijekom 1990-ih godina eksperimentiralo se s primjenom lančanih kazala u online katalozima za enumerativne Deweyevu decimalnu klasifikaciju i Klasifikaciju Kongresne knjižnice, namijenjene korisnicima, a ne klasifikatorima (Olson, Boll 2001, str. 174). Tehnika se pokazala kvalitetnom, međutim, uvidjelo se i da je kao preduvjet za izradu lančanih kazala u enumerativnim klasifikacijama potrebno uskladiti nekonzistentne elemente postojećih tablica. Za razliku od lančanih kazala, sintaktički su sustavi potpuno neovisni o određenom klasifikacijskom sustavu. Stoga valja razmotriti mogućnost primjene sintaktičkih sustava u enumerativnim klasifikacijskim sustavima kao načinu pristupa pojedinim aspektima iz klasifikacijskih oznaka.

Fasete omogućavaju specifičnije i detaljnije označivanje, nego što je to slučaj s enumerativnim sustavima, te pretraživanje po pojedinim elementima složenih oznaka, a posebno putem riječi i prijevoda tih oznaka. Fasetnost se nastoji usavršiti i u Deweyevoj decimalnoj klasifikaciji, Univerzalnoj decimalnoj klasifikaciji (McIllwaine 2001), Blissovoj bibliografskoj klasifikaciji (Bliss Bibliographic Classification) (Broughton 2001), a P. Cochrane tvrdi da će Klasifikacija Kongresne knjižnice s vremenom postati fasetna (2002). Primjena faseta u elektroničkom okruženju posebno je korisna zbog jednostavne mogućnosti pretraživanja po svakoj pojedinoj faseti.

Tradicionalni klasifikacijski sustavi u online pretraživanju koriste se za pretraživanje po samoj klasifikacijskoj oznaci, ali najbolje je kada se u abecednom kazalu klasifikacijskog sustava može odabrati željeni termin (prijevod oznake), povezan s klasifikacijskom oznakom, uz pomoć koje se pronalaze bibliografski zapisi. Mogu se koristiti i kao osnovni oblik mehanizma povratne informacije, na temelju čega se zatim pronalaze i drugi zapisi s istom klasifikacijskom oznakom kao prethodno dobiveni relevantni zapisi.

U mrežnom okruženju klasifikacijski sustavi potencijalno imaju dodatna svojstva koja pospješuju mogućnosti pretraživanja poput kraćenja, prebiranja po hijerarhijskim strukturama, pretraživanja po svakom, a ne samo prvom elementu klasifikacijske oznake, uključujući i u enumerativnim klasifikacijskim sustavima (npr. pronaći samo udžbenike u ovom brojčanom nizu), ukoliko tablice osiguravaju dovoljno preciznosti za ovakvu vrstu pristupa (Olson, Boll 2001, str. 185). Klasifikacijski sustavi u WebPAC-ima lakše se mogu eksploatirati za bolji prikaz hijerarhijskih odnosa i faseta. U web-sučelju također nisu potrebne klasifikacijske oznake, već korisnicima puno prikladniji prijevodi, termini koji prirodnim jezikom opisuju oznaku u tablicama. Prevođenje naziva i opisa oznaka u klasifikacijskim tablicama na jezik krajnjih korisnika pokrenuo je OCLC za Deweyevu decimalnu klasifikaciju (Vizine-Goetz 1996), a u pojedinim katalozima koriste se inačice koje su izradile same knjižnice, kao na primjer Knjižnica Hrvatske akademije znanosti i umjetnosti na temelju UDK (Leščić, Cvitaš 2001).

Dok knjižnični softveri danas omogućavaju pretraživanje po klasifikacijskim oznakama, sve mogućnosti hijerarhijskog sustava i prebiranja po širim i užim oznakama rijetko su prisutne. Format MARC omogućava i pretraživanje po svim elementima oznake, a ne samo po prvim (što vrijedi i za predmetnu odrednicu), no

malo koji online knjižnični katalog to dozvoljava. Isto vrijedi i za kraćenje, na koje dodatno utječe i način na koji je klasifikacijska oznaka oblikovana te je i na to potrebno misliti pri izradi oznaka. Poseban je problem povezivanje popisa signatura s klasifikacijskim tablicama.

U mrežnom okruženju prikaz klasifikacijskih sustava je vizualiziran i korisnicima razumljiviji, iako postoje različiti tipovi ljudi (vizualni, verbalni, apstraktni) te je za sve njih potrebno osigurati primjereno sučelje. Skraćeni oblik klasifikacijskih tablica, tzv. pretraživačko stablo (search tree, subject tree) korisniku omogućava lociranje određenog predmeta i pregledavanje viših i nižih razina specifičnosti. Postoje različite vrste ovakvih sustava s raznolikim mogućnostima. Većina ih prikazuje broj zapisa i hipervezu na zapise koji sadrže određene oznake. Na Internetu se često koriste ovakva pretraživačka stabla utemeljena na Univerzalnoj decimalnoj klasifikaciji, Deweyevoj decimalnoj klasifikaciji i Klasifikaciji Kongresne knjižnice, obično u kombinaciji s web-tražilicom, no takvi su sustavi vrlo široko postavljeni i ne mogu se mjeriti sa specifičnošću i preciznošću pristupa kakav osiguravaju knjižnični klasifikacijski sustavi.

Važno je naglasiti da se u ovim sustavima ne koriste klasifikacijske oznake, već njihovi prijevodi, koje su korisnicima razumljivije i jednostavnije za korištenje. Tzv. predmetni *gatewayi* kontrolirane kvalitete (quality-controlled subject gateways) osiguravaju razrađeniji predmetni pristup građi putem klasifikacijskih sustava, najčešće zajedno sa sustavima predmetnih odrednica i tezaurusima. Tako, npr. SOSIG koristi Univerzalnu decimalnu klasifikaciju, BUBL Deweyevu decimalnu klasifikaciju, a OMNI klasifikacijski sustav američke Nacionalne medicinske knjižnice (Koch 2000). U okviru stvaranja globalne informacijske infrastrukture valja istaknuti međunarodni projekt DESIRE (Development of a European Service for Information on Research and Education), pokrenut s ciljem razvoja europske mreže međusobno pretraživih informacijskih izvora namijenjenih akademskoj zajednici, u okviru kojeg se ispituju tehnike automatske klasifikacije na temelju ugrađenih metapodataka, baze podataka koje "puni" robot i distribuirano mrežno označivanje.

Kada se govori o globalnoj informacijskoj infrastrukturi i prilagodbi klasifikacijskih sustava elektroničkom okruženju, za povezivanje i pretraživanje različitih klasifikacijskih sustava potrebno je uspostaviti zajedničke elemente. Danas su takva nastojanja u tijeku, ali na samome početku uspostavljanja osnovne

klasifikacijske strukture. I. Dahlberg (2001) uspostavila je tablicu od 81 jednakosti između vodećih klasa iz DDK, UDK, KKK, BK i KD sa stručnim skupinama u autoričnoj Klasifikaciji za kodiranje informacija (KKI) (Information Coding Classification – ICC). U matricu koja se sastojala od devet područja i devet faseta smjestila je sve skupine KKI-a i popunila je odgovarajućim oznakama klasifikacijskih sustava. Većina polja mogla se popuniti i sada ova matrica predstavlja temelj za proučavanje kompatibilnosti na užim razinama stručnih područja. KKI može biti važan alat i pokazatelj pri reviziji klasifikacijskih tablica.

Revizija klasifikacijskih tablica može se temeljiti i na istraživanju korisničkih potreba. Jedno takvo istraživanje provedeno za Deweyevu decimalnu klasifikaciju i Kinesku decimalnu klasifikaciju pokazalo je da neki usko povezani predmeti u tablicama nisu hijerarhijski grupirani (Pu 2002). Istraživanje je provedeno na temelju podataka o posudbi s ciljem utvrđivanja srodnih predmeta na temelju jedinica građe koje posuđuju slični korisnici, a kao zaključak se predložilo obogaćivanje klasifikacijskih tablica s odnosima među klasama utvrđenih na temelju korisničkih potreba.

H. Olson i J. Boll (2001, str. 193-196) opisuju kako bi trebalo oblikovati sučelje za predmetno pretraživanje po klasifikacijskim oznakama, pri čemu većina vrijedi i za pretraživanje po predmetnim odrednicama. Sami ističu kako je upitno jesu li sva ta obilježja isplativa s obzirom na sredstva potrebna za njihovu implementaciju, ali su ona svakako poželjna i tehnički izvediva.

Ističu da bi, općenito, upute trebale biti jasne i na laičkom jeziku. Predlažu otvaranje prozorčića s kratkim objašnjenjem u trenutku prelaska kursora nad određenim terminom, a pogotovo nad sličicom. Jasne upute o tome kako jednostavno isključiti ili uključiti ovu opciju neophodne su.

Što se tiče samog prikaza i oblikovanja sučelja, potrebno je osigurati nekoliko različitih sučelja za korisnike, od početničkog do naprednog. Osim na početničkom sučelju, korisnik treba imati mogućnost ući u željeni dio sustava bez da prolazi nepotrebne korake. Bez obzira na to na kojem se stadiju pretraživanja korisnik nalazi, uvijek treba biti u mogućnosti ograničiti pretraživanje po godini, jeziku, zemlji izdavanja i drugim kriterijima. Sustav treba dopustiti nekoliko tehnika pretraživanja poput komandnog, putem izbornika i rangiranja po relevantnosti te dopustiti prelazak s jedne tehnike na drugu u bilo kojem stadiju pretraživanja. Upute

o pretraživanju trebaju jasno opisati je li prekoordinirani skup termina obavezan, dozvoljen ili ne postoji. Isto se odnosi na kriterije ograničavanja pretraživanja.

Osim po klasifikacijskim oznakama, valja omogućiti i pretraživanje po signaturi. Korisniku treba biti jasno koja je razlika između ta dva popisa te treba postojati mogućnost jednostavnog prelaska s jednog popisa na drugi. Ukoliko tražena klasifikacijska oznaka ili signatura ne postoji, rezultat pretraživanja treba biti ispis najbližih oznaka ili signatura u hijerarhijskoj strukturi, s pojašnjenjem toga rezultata. Treba postojati mogućnost prebiranja po brojčanom nizu, pri čemu valja prikazati i pripadajuće prijevode uz svaki broj ili oznaku. U sustavu trebaju postojati dva popisa prijevoda pojedinih oznaka, jedan za korisnika, a drugi za knjižničara. Pristup popisu signatura ili klasifikacijskih oznaka treba osigurati i putem oznake i riječima kao što su predmetne odrednice, ključne riječi iz opisa, bilježaka, indeksa ili drugih termina korištenih za opis broja u tablicama. Kada god se prikaže klasifikacijska oznaka ili signatura, korisnik treba imati mogućnost da otvori prozor u kojem će se prikazati taj broj i njegove šire i uže razine, zajedno sa svim prijevodima. Pri pretraživanju po signaturama, prikaz zapisa unutar jedne skupine treba biti moguć prema popisu signatura, po kronološkom slijedu, godini izdavanja, relevantnosti i ostalim važnim kriterijima. Valja omogućiti kombinirano oblikovanje upita po predmetnim odrednicama ili ključnim riječima s klasifikacijskim oznakama korištenjem Booleovih operatora. Za vrijeme pretraživanja po ključnim riječima iz različitih dijelova tablica s Booleovim operatorima, treba omogućiti prikaz klasifikacijskih oznaka s prijevodima, zajedno s brojem bibliografskih zapisa sa svakom oznakom iz popisa signatura kao i s traženom kombinacijom. Također, uz svaku prikazanu oznaku valja prikazati bilješku o korištenju u tablicama, npr. kada je uvedena u tablice.

Prebiranje valja omogućiti po klasifikacijskim tablicama pri čemu uz sve oznake treba prikazati oznake i prijevode oznaka, a uz svaku oznaku treba omogućiti prijelaz na popis signatura, a uz signaturu broj dokumenata s osnovnim podacima (autor, naslov, godina izdavanja). Isto je tako potrebno omogućiti prebiranje po popisu signatura, te jednostavan prelazak s klasifikacijskih tablica na popis signatura i obrnuto, i sa skraćenih zapisa na potpune. Valja integrirati popise signatura knjižnica u skupnom katalogu i omogućiti ograničenje pregledavanja samo određenog popisa.

Prebiranje valja omogućiti i po hijerarhijskim odnosima, tj. treba imati mogućnost odlaska s glavnih skupina na podskupine do najuže oznake, s pridruženim prijevodima, kao i u obrnutom smjeru. Treba postojati mogućnost da se od struke dođe do konkretne jedinice i obrnuto (npr. od biologije do dijelova biljke). Koristan bi bio ekspertni sustav koji bi automatski za sintetiziranu signaturu pokazao sve hijerarhijske popise za pojedine elemente signature. To MARC omogućava i za predmetne odrednice. Ako pretraživanje počinje s riječju, sustav bi trebao prikazati i popis predmetnih odrednica i klasifikacijske tablice, ostavljajući korisniku da sam napravi izbor.

Predmetno pretraživanje po klasifikacijskim oznakama u današnjim se online katalogima ne koristi dovoljno, iako je već dugo poznato da se ovom vrstom pretraživanja dobivaju drukčiji rezultati nego pretraživanjem po riječima i da ta dva pristupa pretražuju dokumente koji se preklapaju, ali ne i udvostručavaju. Format MARC 21 omogućuje predmetno pretraživanje na puno sofisticiraniji način s više aspekata no što je to bilo moguće do tada. Iako ova vrsta pretraživanja ne spada u ono što bi svakodnevni, usputni korisnik trebao, trebalo bi ju omogućiti za potrebe iscrpnih i temeljitih predmetnih pretraživanja.

Istraživanja su pokazala da se najbolji rezultati pretraživanja dobiju kombiniranjem kontroliranog stručnog ili predmetnog sustava s nekontroliranim rječnikom te se sva tri pristupa trebaju koristiti ukoliko se želi najpotpuniji mogući popis rezultata. Međutim, samo je mali broj online kataloga uspješno implementiralo sindetičke strukture klasifikacijskih sustava ili sustava predmetnih odrednica u kojima se može prebirati po hijerarhijskim odnosima koji zbog toga korisnicima ostaju skriveni (Olson, Boll 2001, str. 185). Većina online kataloga pruža mogućnosti pretraživanja po predmetnoj odrednici ili ključnoj riječi, mnogi i po signaturi, neki i po početku signature, neki omogućavaju prebiranje po popisu signatura, ali gotovo da nema kataloga koji sustav omogućava proširenje ili suženje upita prebiranjem po stručnoj oznaci. Rijetki su i online katalogi koji dopuštaju oblikovanje kombiniranih upita od predmetnih odrednica, ključnih riječi i klasifikacijskih oznaka. Primjerice, korisno bi bilo za upit o *arhitekturi u Barceloni prije Gaudija* oblikovati upit na sljedeći način: *(klasifikacijska oznaka za arhitekturu grada Barcelone) OR (predmetna odrednica za arhitekturu grada Barcelone) OR (ključne riječi Barcelona AND arhitektura) NOT ((klasifikacijska oznaka za Antonia Gaudija) OR (predmetna odrednica za Antonia Gaudija))*.

5.3.2. Popisi predmetnih odrednica

Popisi predmetnih odrednica ili predmetni sustavi kontrolirani su rječnici za označivanje koji se sastoje od rječnika i sintakse. Rječnik je popis dozvoljenih ili odobrenih riječi, na temelju kojih se sastavljaju jednostavne ili složene predmetne odrednice. Sintaksu čine pravila koja propisuju određivanje predmeta dokumenta i oblikovanje predmetne odrednice, koja se može sastojati i od pododrednica (npr. vremenskih, geografskih i formalnih). Cilj sustava predmetnih odrednica jest sažeti čitav sadržaj dokumenta uz pomoć jednog ili više dozvoljenih termina koji su jednako specifični kao i predmet koji označuju. Neke je predmete nemoguće prikazati jednom predmetnom odrednicom, u kojim se slučajevima dodjeljuje njih nekoliko, obično jedna do tri (Olson, Boll 2001, str. 37), iako je Kongresna knjižnica 1996. uvela praksu prema kojoj je dozvoljeno do deset predmetnih odrednica radi potpunijeg opisa sadržaja.

Važno je omogućiti pretraživanje po pojedinim podpredmetnicama, čime se povećava broj potencijalnih predmetnih pristupnica. K. Drabenstott i M. Weller (1996b) predlažu tehniku koju nazivaju “stablo za odlučivanje” (decision tree), a temelji se na ideji o programiranju sučelja na način da se za svaki pojedini upit utvrdi najučinkovitiji način pretraživanja. Ukoliko se riječi iz upita potpuno preklapaju s predmetnom odrednicom, onda se pretraživanje provodi u datoteci predmetnih odrednica, a ako ne, pretraživanje se prvo provodi po ključnim riječima iz predmetnih odrednica, zatim iz naslova i tako dalje. Valja naglasiti da je korisnika uvijek potrebno uputiti u način na koji sustav obrađuje njegov upit.

Popis predmetnih odrednica može imati uputnice na kontrolirane termine ili integriran automatski sustav prevođenja na riječi sustava, kakav mehanizam ima MeSH, pri čemu treba postaviti uputu o provedenom postupku. Neki sustavi poput LCSH nemaju ovakva rješenja te pribjegavaju uključivanju ekvivalenata u svoju datoteku predmetnih preglednih kataložnih jedinica i uputnica. To se može ostvariti na dva načina, uključivanjem svakog termina u hijerarhijsku strukturu, ili izradom uputnica sa svih drugih ekvivalentnih termina za vrijeme izrade nove odrednice. Pokazalo se da korisnicima nije uvijek jasno kako uputnice funkcioniraju. Stoga je potrebno osigurati upute o tome što su npr. nedeskriptori, širi, už i srodni termini i što koja uputnica znači. U svakom slučaju, korisnici bi uvijek trebali znati i vidjeti gdje se u hijerarhiji nalaze, pa je u online katalogu potrebno uvijek prikazati jednu

ili dvije šire ili uže razine. Primjerice, problem u online inačici LCSH je taj što korisnik mora uvijek pritisnuti tipku *Enter* da bi od *Business* dobio *Business intelligence* (Wasylenko 2000).

LCSH je u Sjevernoj Americi najrašireniji sustav predmetnih odrednica te, premda nije prilagođen online okruženju, Kongresna knjižnica pozitivno reagira na kritike i provodi preinake. Kongresna knjižnica povećala je broj predmetnih odrednica koje se dodjeljuju dokumentu, a A. O'Brien (1994, str. 225-226) sugerira da bi, po uzoru na Nacionalnu medicinsku knjižnicu, trebala napraviti razliku između glavnih i sporednih predmetnih odrednica, kako bi se sačuvala mogućnost pristupa po glavnom predmetu, a ne bi se onemogućio pristup po sporednim predmetima.

Radi potpunijeg pristupa s različitih aspekata, OCLC nastoji povezati LCSH i Klasifikaciju Kongresne knjižnice (Olson, Boll 2001, str. 141). Budući da se klasifikacijski sustavi temelje na stručnom rasporedu, a predmetni na abecednom, integrirani bi sustav omogućio prikazivanje i pristupanje svakoj temi i građi u različitim kontekstima i sa različitih aspekata. Međutim, povezivanje ovih dvaju sustava u puno je slučajeva nemoguće jer ne postoje ekvivalenti. MeSH je uspješno integrirao termine iz sustava predmetnih odrednica i klasifikacijskih oznaka.

Jednostavnije rješenje za povezivanje popisa predmetnih odrednica s klasifikacijskim sustavima jesu tezaursi, gdje se prvo okupljaju pojmovi u fasete i hijerarhije, a zatim se svakom pojmu pridruži predmetna ili stručna oznaka u trenutku kada se sustav oblikuje. Primjeri takvih tezaurusa su Tezaurus ERIC-ovih deskriptora (Thesaurus of ERIC descriptors) i Tezaurus za umjetnost i arhitekturu, koji su specijalizirani pa ih je jednostavnije oblikovati nego opći tezaurus. Kao model za jedan opći tezaurus mogao bi poslužiti online tezaurus MeSH, u kojem je svaki predmet sadržan i u popisu predmetnih odrednica i deskriptora, a ima integrirane termine iz sustava predmetnih odrednica i klasifikacijskih oznaka. Ovi tezaursi pokazuju da se predmetne odrednice mogu strukturirati u abecedno i hijerarhijski koordinirane sustave. Što se tiče spajanja LCSH i Klasifikacije Kongresne knjižnice, možda je bolje, smatra se, stvoriti čvrstu sindetičku strukturu za sam sustav LCSH, kakvu ima MeSH (Olson, Boll 2001, str. 142).

5.3.3. Označivanje na temelju kontroliranih sustava

Jezik ima posebno važnu ulogu u pretraživanju. Korisnik postavlja upit riječima, a informacijski sustav sadrži informacije o dokumentima izražene najčešće riječima. Komunikacija između pošiljatelja i primatelja uspješna je ukoliko obje strane razumijevaju poruku na isti način. Stoga su pretraživanje i označivanje dva međusobno povezana procesa koja se moraju temeljiti na istoj terminologiji. Na pretraživanje utječe i kvalitetno označivanje jedinica građe prema kontroliranim sustavima, kao i kvaliteta samog kontroliranog sustava. Označivanje i kontrolirani sustav treba prilagoditi korisničkoj zajednici (Bates 1989b; Hjørland, Nielsen 2001; Pu 2002).

O oblikovanju kontroliranih sustava kao i o predmetnom označivanju postoji opsežna literatura, no detaljno obrađivanje te tematike nadilazi predviđeni opseg ovoga rada. Stoga ćemo se ovdje ograničiti na osnovnu problematiku.

Prema standardu ISO 5963-1985 za predmetno označivanje, tri su osnovna koraka u predmetnom označivanju: određivanje teme ili sadržaja jedinice, pojmovna analiza radi utvrđivanja koji se aspekti jedinice trebaju prikazati u bibliografskom zapisu, i prevođenje pojmova ili aspekata na kontrolirani rječnik, čime se oblikuju pristupnice. Svi ovi koraci trebaju biti usklađeni s pravilima određene knjižnice ili standardima poput popisa predmetnih odrednica ili klasifikacijskih tablica.

Na pretraživanje posebno utječe točnost, iscrpnost, specifičnost i konzistentnost označivanja (Olson, Boll 2001, str. 88). Točnost označivanja obuhvaća sposobnost osobe koja provodi označivanje da identificira pojmove koje je potrebno predstaviti i sposobnost da te pojmove prevede na termine kontroliranog rječnika. Nakon utvrđivanja sadržaja i predmeta dokumenta, potrebno je odrediti koliko će se aspekata prikazati predmetnim odrednicama, deskriptorima i/ili klasifikacijskim oznakama. Knjižnični katalozi tradicionalno imaju relativno nisku razinu iscrpnosti. Iscrpnost ili dubina spadaju u pojmovnu analizu, nakon koje slijedi prevođenje. U prevođenju je važna specifičnost koja se može opisati kao hijerarhijska razina prikaza predmeta, a može se podijeliti na tri čimbenika: specifičnost i prikladnost rječnika (koliko su hijerarhijske razine samog rječnika specifične), specifičnost primjene (pravilo je da se upotrijebi najuži postojeći termin) i specifičnost termina u kontekstu danog kataloga (visoka specifičnost znači da je termin u katalogu relativno rijedak; ova specifičnost treba biti čimbenik u

odabiru vrste kontroliranog rječnika prema katalogu, no, knjižnica često nema puno izbora). Svi ovi čimbenici utječu na preciznost i odziv u pretraživanju. Većina dokumenata ima glavni predmet koji će osobe koje vrše označivanje uglavnom isto odrediti; međutim, označivanje sporednih predmeta može znatno varirati. Konzistentnost se odnosi na potrebu da se isti predmeti označe istim terminima, kako bi u katalogu bili okupljeni slični predmeti, što se u praksi pokazalo teško izvedivim. Konzistentnost je obrnuto proporcionalna s iscrpnošću i specifičnošću jer visoka razina iscrpnosti ili visoka razina specifičnost dovode do niske razine konzistentnosti (Olson, Boll 2001, str. 89).

Kvaliteta samog kontroliranog sustava također utječe na uspješnost predmetnog pretraživanja. Za izbor termina koji ulaze u kontrolirani rječnik tradicionalno se primjenjuju kriteriji književnog jamstva (*literary warrant*), korisničkog jamstva (*use warrant*) i strukturalnog jamstva (*structural warrant*) (Svenonius 2000, str. 135-137). Književno jamstvo zahtijeva uvrštavanje termina na temelju građe koju je potrebno opisati, korisničko jamstvo odnosi se na potrebu da se uvrste oni termini ili nazivi koje koristi korisnik pa je važno da se uključe svi nazivi pod kojima se pojam može pronaći, a strukturalno jamstvo postoji zbog potrebe za uključivanjem onih termina koji nisu odabrani ni prema prvom, niti prema drugom jamstvu, ali su neophodne zbog strukturalne funkcije u rječniku, npr. zbog nadopunjavanja praznina u hijerarhiji ili okupljanja užih termina. Često se ističe da su najbolji izvor za kontrolirani sustav korisnici i načini na koje provode pretraživanja, a tek onda jedinice građe (Bates 2002; Cochrane 2002).

Međutim, oblikovanje sustava prema ovim jamstvima često nije u potpunosti moguće. Na primjer, postoje sustavi koji se koriste globalno pa je teško slijediti književno jamstvo. Postavlja se i pitanje treba li uvoditi nadzor nad riječima iz društvenih i humanističkih znanosti koje imaju jako puno značenja (npr. *kultura*) i ima li to uopće smisla. Slično tome je pitanje interdisciplinarnih rječnika jer mnoge riječi imaju različita značenja u različitim disciplinama. Osim toga, kontrolirani sustavi za građu u video ili audio formatu imaju posebne potrebe, pa je tako npr. za oblikovanje Tezaurusa za umjetnost i arhitekturu uvedeno predmetno jamstvo (*object warrant*), prema kojem su dozvoljeni svi nazivi koje umjetnički predmeti mogu imati.

U posljednje se vrijeme istraživanja sustava za označivanje i pretraživanje bave sučeljem (npr. kako pospješiti korisnikov odabir termina za pretraživanje),

poboljšanjem prirodnog jezika kao sredstva pretraživanja, ili hipertekstom (za povezivanje srodnih jedinica građe), a malo se govori o učinkovitosti termina za prikaz pojmova iz dokumenata ili samih dokumenata. Takva su istraživanja potrebna jer je u doba kada je nastala većina kontroliranih sustava koja se danas koristi okruženje uglavnom bilo papirno.

5.3.3.1. Automatizacija označivanja na temelju kontroliranih sustava

Kada se koriste kontrolirani jezici, označivanje radi čovjek, iako su dosadašnja istraživanja usmjerena ka automatizaciji ovoga procesa u pojedinim segmentima dala uspješne rezultate. Zbog ograničenog opsega ovoga rada, u ovu se tematiku neće dublje zalaziti.

E. Svenonius (2000, str. 132) navodi da su algoritmi pri preslikavanju riječi iz prirodnog jezika na riječi kontroliranog jezika potrebni za izdvajanje terminologije kontroliranog jezika iz teksta na prirodnom jeziku, za izdvajanje i formalizaciju semantičkih informacija u rječnicima prirodnog jezika radi uporabe u kontroliranom jeziku (homonimi, polisemi, sinonimi, hijerarhijski odnosi i odnosi srodnosti), za kategorizaciju terminologije kontroliranog jezika po semantičkim i gramatičkim kategorijama te za oblikovanje termina kontroliranog jezika (predmetnih odrednica).

U prirodnim i tehničkim znanostima jednostavno je razlikovati termine od riječi koje to nisu. Postoje i neki automatski načini na koje se to može na temelju statistike i leksičko-gramatičkih svojstava nizova znakova. No, u društvenim i humanističkim znanostima značenja se često iznova definiraju i mijenjaju (Svenonius 2000, str. 133), što dodatno komplicira oblikovanje kontroliranih sustava.

Nakon izdvajanja termina i utvrđivanja semantičke kategorije potrebno je utvrditi njihove odnose ekvivalentnosti i hijerarhičnosti. Što se tiče automatizacije utvrđivanja odnosa ekvivalentnosti, ono može biti jednostavno, ukoliko su termini morfološki srodni. Mnogi današnji sustavi za pretraživanje već osiguravaju ovu vrstu povezivanja, npr. automatsko utvrđivanje oblika u jednini i množini ili uklanjanje sufiksa radi okupljanja termina s istom osnovom. Vjeruje se da će s

vremenom većina sustava za pretraživanje ugraditi algoritme za ujednačavanje ortografskih, sintaktičkih i leksičkih oblika. Za riječi koje su semantički, ali ne i morfološki ekvivalentne, potrebni su sofisticiraniji postupci i rječnici, tezaursi, pojmovnici, ili baze termina za automatsko utvrđivanje odnosa sinonimije. Hijerarhijske odnose teško je automatski ustvrditi. Ono što jest moguće je okupiti termine na temelju njihovih sličnih statističkih distribucija i zatim ih ručno hijerarhijski poredati.

Kvalitetno označivanje, uspješno pretraživanje i učinkovito automatsko označivanje ovise o mogućnosti definiranja predmeta. Stupanj do kojeg se može reći da se u nekom dokumentu radi o nečemu ovisi o nekoliko čimbenika, od kojih su najvažnija obilježja jezika koji se koristi. To se može ilustrirati korištenjem modela predmetne analize koja se temelji na rečeničnoj gramatici (Svenonius 2000, str. 46-47).

Postupci za automatizaciju označivanja i dalje se intenzivno istražuju. Danas automatizacija omogućuje premošćivanje jaza između rječnika koji rabi korisnik i kontroliranog rječnika koji postoji u katalogu. Suvremeni razvoj automatizacije označivanja na temelju kontroliranih sustava temelji se na računalnoj obradi prirodnog jezika (Slavić 1998, str. 107) (vidi poglavlje 5.6.).

5.4. Prirodni sustavi u predmetnom pretraživanju

Prirodni sustavi za označivanje temelje se na riječima preuzetima iz same jedinice građe, a ne na terminima preuzetima iz nekog kontroliranog sustava za označivanje. Sve riječi koje se u dokumentu spominju potencijalni su termini za označivanje. U praksi označivanje prirodnim jezikom obuhvaća riječi iz naslova, iz ostalih polja kataložnog zapisa (npr. polje za kratki opis ili sinopsis), sažetka ili potpunog teksta dokumenta.

Prirodni sustavi za označivanje su postkoordinirani sustavi jer termine za pretraživanje “oblikuje” korisnik u trenutku postavljanja upita. O sustavima za označivanje riječima prirodnog jezika M. Mikačić (1996) govori kao o sustavima bez gramatike, a E. Svenonius (2000, str. 131) kao o enumerativnim predmetnim jezicima ili jezicima bez sintakse (za razliku od sintetičkih predmetnih jezika), jer se

riječi iz teksta dokumenta tom dokumentu pridružuju kao niz slobodnih, međusobno nepovezanih riječi. Drugim riječima, za razliku od kontroliranih sustava za označivanje, u prirodnim sustavima ne postoje propisani načini oblikovanja termina.

Prvi popisi ključnih riječi koji su se izrađivali računalno su unitermi i riječi iz naslova (KWAC, KWIC i KWOC). Označivanje prirodnim jezikom u većini današnjih sustava provodi se automatizirano, tj. računalom koje izrađuje invertirane datoteke. Invertirane datoteke popisuju i okupljaju sve riječi iz svakog pojedinog zapisa po određenim poljima iz kataložnih zapisa. Riječi koje nisu obuhvaćene invertiranim datotekama, poput članova, veznika i prijedloga, čine tzv. stop-listu. Pretraživanje se temelji na usporedbi riječi iz upita s riječima u invertiranoj datoteci.

Osnovni problem prirodnog jezika u funkciji pretraživanja čine njegova obilježja sinonimije, homonimije i polisemije (Lancaster 1986; Foskett 1997; Olson, Boll 2001). Korištenje neograničenog broja prirodnih riječi pri označivanju i pretraživanju dovest će do određenog broja pogrešaka u informacijskom pretraživanju. Jedina kontrola koja postoji nad terminima u prirodnom sustavu jest stop-lista.

T. Peters i M. Kurth (1991) preporučuju strategije pretraživanja kojima se iskorištava prednost korištenja ključnih riječi iz naslova: 1) ako korisnik zna za barem jednu jedinicu građe, unese riječ iz naslova, dobije zapis u kojem pronade predmetne odrednice, prema kojima provodi daljnje pretraživanje; 2) ako korisnik zna za barem jednu jedinicu građe, riječi iz naslova te jedinice koristi za pronalaženje drugih sličnih jedinica; 3) korisnik može pokušati pretraživati po ključnoj riječi iz naslova da bi dobio barem jednu relevantnu predmetnu odrednicu koja će mu pomoći za daljnje pretraživanje; te, 4) traženje po ključnim riječima iz naslova može se koristiti za predmetni pristup umjesto traženja po kontroliranom rječniku.

Uporaba prirodnog jezika povećava odziv, ali smanjuje preciznost, dok je s kontroliranim jezicima obrnuto (Rowley 1994). Pretraživanjem putem ključnih riječi ne pronalaze se sve relevantne jedinice, a broj nerelevantnih rezultata nadilazi relevantne. Pretraživanje po riječima iz odrednica (ne samo predmetnih, već i osobnih, korporativnih i stvarnih) predstavlja moćan alat pri predmetnom pretraživanju (Borgman 1996, str. 497). Metode i tehnike pretraživanja poput ključnih riječi, kraćenja, Booleovih i drugih operatora, automatskog označivanja i pretraživanja (algoritmi), rangiranja po relevantnosti i ekspertnih sustava, koji se

koriste sa sustavima prirodnih jezika, mogu djelomično zamijeniti kontrolirane jezike. Međutim, ono što se danas uočava kao najbolje rješenje jest istodobna mogućnost korištenja i pretraživanja po prirodnim i kontroliranim sustavima.

5.5. Prirodni sustavi naspram kontroliranih i mogućnosti pretraživanja

Rasprava o ulozi prirodnog i kontroliranog jezika u pretraživanju pokrenuta je prije više od jednoga stoljeća. E. Svenonius još je 1986. povijest ove debate podijelila u tri razdoblja, a J. Rowley (1994, str. 113-116) dodaje četvrto. Razdoblja o kojima govore jesu sljedeća: 1. uvođenje kontroliranog jezika, 2. usporedba različitih jezika za označivanje s ciljem određivanja najboljeg, 3. spoznaja da se najbolji rezultati dobivaju usporednim korištenjem obiju vrsta jezika, te, 4. naglasak je na oblikovanju sustava prema potrebama krajnjeg korisnika. Valja naglasiti da su se argumenti za i protiv određenih sustava često temeljili na tada dostupnoj tehnologiji. Tek razvojem online sustava omogućeno je istodobno korištenje različitih sustava, kada se postiže konsenzus o usporednoj uporabi različitih sustava (Olson, Boll 2001).

Prvo razdoblje, iz 19. stoljeća, obilježeno je kontroliranim jezicima, a rasprava započinje kada označivanje po riječima iz naslova postalo popularno. Klasifikacijski sustavi koristili su se za redanje bibliografskih zapisa u stručnim katalozima. Povećavanjem kataloga korisnici su sve teže pronalazili određenu jedinicu građe, a terminologija u katalogu nije bila korisnički pristupačna. Sredinom 1850-ih u Velikoj Britaniji predloženo je označivanje riječima iz naslova, no C. Cutter prevladao je svojom ustrajnom tvrdnjom da naslovi ne izražavaju istinski predmet dokumenta te da građa o istom predmetu u knjižnici nije okupljena ako se riječi u naslovima sličnih jedinica građe razlikuju. C. Cutter uveo je ukršteni katalog koji se sastojao od abecednog i predmetnog niza, koji je bio prototip sustava LCSH. Time je ova rasprava ostala zapečaćena na pola stoljeća, u korist kontroliranih rječnika.

1960-ih postavilo se pitanje same ocjene rezultata kao relevantnih. U početku se smatralo da je savršen onaj sustav koji kao rezultate daje sve dokumente

iz sustava koji odgovaraju upitu (ne samo neke), te samo one koji su relevantni (ne i nerelevantne). Kasnih 1960-ih T. Saračević i drugi počeli su uvoditi mjere poput relevantan, manje relevantan i nerelevantan, koje se temelje na matematičkoj višeznačnoj logici i o višeznačnoj teoriji skupova, koje je prvi opisao L. A. Zadeh sredinom 1960-ih (Olson, Boll 2001, str. 74). Formule za višeznačne skupove omogućavaju pronalaženje i djelomično odgovarajućih rezultata i njihov ispis po vjerojatnoj relevantnosti.

Drugo razdoblje, 1960-ih, obilježavaju istraživanja usmjerena na uspoređivanje različitih jezika za označivanje s ciljem utvrđivanja koji je najbolji. Ovo razdoblje vezano je uz pojavu računalno utemeljenih sustava za označivanje i pretraživanje, a prvi je 1959. uveo H. P. Luhn, kao ključnu riječ u kontekstu (KWIC). KWIC se temeljio na računalnom prikazu svake riječi iz naslova, okružene ostalim riječima iz naslova. Niski odziv rezultata pretraživanja ukazao je na činjenicu da je za sofisticiranije označivanje potrebno u obzir uzeti i semantička i sintaktička obilježja jezika. Najpoznatija istraživanja ovoga razdoblja provedena su u okviru cranfieldskih eksperimenata, koji su uspješnost mjerili na temelju preciznosti i odziva. Preciznost i odziv utvrđivali su se za trideset i četiri jezika. Pokazalo se, između ostalog, da se iscrpnijim i potpunijim označivanjem sadržaja dokumenta povećava odziv, a smanjuje preciznost, da specifičnost termina za označivanje pospješuje odziv i preciznost, te da uputnice u kontroliranom rječniku pridonose povećanju odziva i smanjenju preciznosti. Pokazalo se i da uspješnost sustava za označivanje ne ovisi o tome koristi li se klasifikacijski ili predmetni sustav za označivanje, već o mogućnostima jezika za označivanje i standarda za označivanje, što se danas još uvijek prihvaća. Tada je utvrđeno i da su preciznost i odziv u pravilu u obrnuto proporcionalnom odnosu, čemu su se drugi znanstvenici suprotstavili jer se, primjerice, odziv može povećati uključivanjem sinonima, što nikako ne smanjuje preciznost (Olson, Boll 2001, str. 75). Također su neki pokazali da zadovoljstvo korisnika više ovisi o odzivu nego o preciznosti. Drugi cranfieldski eksperiment pokazao je da prirodni jezik uz minimalnu kontrolu sufiksa i sinonima daju jednako dobre ili bolje rezultate nego sofisticirani jezik za označivanje. K. Markey, P. Atherton i C. Newton (1980) pokazale su da kontrolirani jezici poboljšavaju preciznost, a nekontrolirani jezici odziv. Istraživanja provedena u 1960-ima i 1970-ima također su pokazala da prirodni jezici daju jednako dobre rezultate, ako ne i bolje. No, ta su istraživanja bila usmjerena na to koji je jezik

bolji, a ne kako se mogu zajedno koristiti za dobivanje najboljih rezultata. Mnogi autori poput R. Fugmanna i E. Svenonius kasnije su pokazali komplementarnost prirodnog jezika i kontroliranih jezika. Dok su s jedne strane ovi eksperimenti kritizirani zbog laboratorijskih, a ne stvarnih uvjeta i činjenice da se određivanje preciznosti i odziva temeljilo na procjeni relevantnosti, koja je subjektivna, mnogi vjeruju u rezultate ovih eksperimenata jer drugih pokazatelja nema, a i teško bi bilo provesti testove koji bi bili pouzdaniji. Prvi i drugi cranfieldski eksperimenti nisu obuhvatili online kataloge.

Tijekom trećeg razdoblja, 1970-ih, i dalje su se provodila manja istraživanja. Više se nije težilo velikim eksperimentima već manjim studijama slučaja. Novi naglasak istraživanja jest na korisnički prilagođenim sučeljima i algoritmima za učinkovitije korištenje prirodnih jezika. Osnovno je obilježje ovoga razdoblja spoznaja da se najbolje pretraživanje može postići istodobnim korištenjem prirodnog i kontroliranog jezika za označivanje. Uspoređuju se prednosti i nedostaci jednog i drugog jezika, koje navode C. Dubois (1987) te T. Aitchinson i A. Gilchrist (1987). U osnovne prednosti prirodnog jezika C. Dubois ubraja niske troškove obrade, pojednostavljeno pretraživanje zbog riječi prirodnog jezika, mogućnost pretraživosti po svim sadržajnim elementima, svaka riječ ima istu vrijednost u pretraživanju, izostanak ljudskih pogrešaka pri označivanju i ažurnost u uključivanju novih termina. T. Aitchinson i A. Gilchrist dodaju da visoka specifičnost osigurava preciznost i izvrsnost u pronalaženju imena osoba i organizacija, a iscrpnost omogućava visoki odziv, te olakšanu razmjenu među bazama podataka jer izostaje problem jezične nekompatibilnosti. H. Olson i J. Boll (2001, str. 49) navode i da nije potrebno utrošiti vrijeme, novac i napor na izradu rječnika i složene sindetičke strukture, da korisnik koji traži pod bilo kojim od postojećih sinonima ima dobre šanse da nađe barem nekoliko odgovora, te da nekontrolirane riječi mogu poslužiti kao ulazne točke za kontrolirane termine i kao djelomična zamjena za tezaurus jer omogućuju istodobno ili slijedno pretraživanje nekoliko kontroliranih jezika u više baza podataka.

C. Dubois u nedostatke prirodnog jezika ubraja veće opterećenje korisnika zbog odabira termina, izostavljanje sadržajne informacije koja nije eksplicitno navedena riječima, nepostojanje veza sa specifičnih termina na opće i potrebu za poznavanjem stručnog područja. T. Aitchinson i A. Gilchrist spominju i probleme sa sintaksom i lažne veze ističu da iscrpnost može dovesti do smanjene preciznosti. H.

Olson i J. Boll (2001, str. 50) spominju da ažurnost može biti i nedostatak jer korisnici ne moraju znati novu terminologiju (npr. promjena naziva država) te naglašavaju da su lažne veze posebno moguće ako sustavi ne podržavaju operator bliskosti.

Što se tiče kontroliranih jezika, prema C. Dubois njima se rješavaju brojni semantički problemi, identificiraju se generički odnosi te prikazuju stručna područja, a T. Aitchinson i A. Gilchrist također ističu navođenje pojmova koji su u tekstu implicitno navedeni, zaobilaženje sintaktičkih problema korištenjem višečlanih termina i drugih alata, ublaženo snižavanje razine preciznosti zbog "pretjerane iscrpnosti" (pronalaženje manje značajnih pojmova), a posebno se ističe njihova vrijednost u numeričkim bazama podataka i višejezičnim sustavima. H. Olson i J. Boll (2001, str. 39) navode veću vjerojatnost da će i označitelj i korisnik na isti način izraziti jedan pojam, veću vjerojatnost da će i jedan i drugi lakše doći do željenog termina uz pomoć sindetičkih obilježja (uži/širi pojam, srodni pojam, vidi, vidi i), veću vjerojatnost da će isti termin koristiti različiti označitelji ili isti označitelji u različitim slučajevima čime se osigurava konzistentnost, ubrzava pretraživanje, ali i označavanje jer nije potrebno prisjećati se i tražiti moguće sinonime kada prvi traženi termin nije u bazi. Također se pomaže korisniku da lakše i brže osvijesti svoju informacijsku potrebu.

Nedostaci kontroliranog rječnika su visoki trošak, potencijalna neprimjerena pokrivenost područja, ljudska pogreška koja može dovesti do izostajanja rezultata, zastarjelost terminologije, problem sustavnog uspostavljanja svih relevantnih odnosa među terminima (C. Dubois), nedovoljna specifičnost, čak i u detaljno razrađenim sustavima, nedostatak iscrpnosti ili potpunosti, potreba da korisnik nauči umjetni jezik, te otežana razmjena zbog nekompatibilnosti (T. Aitchinson i A. Gilchrist).

U četvrtom razdoblju razvijanje i oblikovanje sustava usmjereno je ka potrebama krajnjeg korisnika. Započelo je u kasnim 1980-ima sa studijama koje su provodile istraživanja s pravim korisnicima i u novonastalim sustavima kao što su grafička korisnička sučelja. Potvrđuje se da se najbolji rezultati pretraživanja dobiju komplementarnim korištenjem prirodnog i kontroliranog jezika za označavanje. O tome svjedoči niz provedenih istraživanja, kao i ono S. Cousins (1992). Ona se bavila ispitivanjem mogućnosti predmetnog pretraživanja u online katalozima te je došla do zaključka da i prirodni jezik i PRECIS kao nadopuna zapisima MARC

pozitivno utječu na broj pronađenih relevantnih jedinica građe, s time da PRECIS dovodi do boljih rezultata, dok je u kraćenju prirodni jezik najbolji. Također navodi da osnovni problem s prirodnim jezikom leži u odabiru pojmova za pretraživanje, međutim, automatsko kraćenje i korištenje uputnica pospješuju uspješnost pretraživanja.

5.6. Računalna obrada prirodnog jezika u predmetnom pretraživanju

Zbog dominantnosti prirodnog jezika tijekom pretraživanja, računalna obrada prirodnog jezika (natural language processing – NLP) može imati značajnu ulogu u pospješivanju predmetnog pretraživanja. Prijedlozi da se računalna obrada prirodnog jezika iskoristi kao metoda za unapređenje uspješnog predmetnog pretraživanja temelje se na činjenici da riječi u korisničkom upitu pripadaju jeziku te da kao takve imaju određena morfološka ili sintaktička obilježja, na temelju kojih se mogu utvrditi i pronalaziti semantičke informacije. Tako bi, primjerice, izraz *European policy on money* trebalo biti moguće pronaći izrazima kao što su *European monetary policy*, *monetary policy of the common market*, *ecu policy in Europe* itd. (Le Loarer 1993, str. 254). Međutim, WebPAC-i koji bi na takav način rukovali korisničkim upitima, još su u istraživačkom stadiju. Ovo poglavlje nastojat će prikazati dosadašnja postignuća i moguće razvojne smjerove.

Tri su osnovne razine NLP-a: leksička, sintaktička i semantička (Smeaton 1997). Leksička razina obuhvaća riječi i vrste riječi. Svaka riječ obrađuje se morfološki na temelju različitih nastavaka, neovisno o kontekstu, pri čemu se koristi rječnik. U engleskom jeziku, primjerice, vrlo često različite vrste riječi imaju isti oblik pa su zbog toga potrebne i druge vrste jezične analize. Sintaktička razina odnosi se na rečeničnu gramatiku i službu riječi, a pokazalo se da je nemoguće u potpunosti ustvrditi sintaktička pravila i da takva sveobuhvatna gramatika ne postoji. Semantička razina odnosi se na analizu značenja.

Za računalnu obradu prirodnog jezika neophodan je kontrolirani rječnik, čija sofisticiranost, uz same algoritme za obradu prirodnog jezika, utječe na kvalitetu ovoga procesa. Rječnici su potrebni za rješavanje problema homonimičnosti u

jeziku, za iznimke u pravilima za obradu jezika (npr. nepravilne množine, idiomi), za određivanje semantičkih veza između riječi i kontroliranih termina, uključujući veze između riječi iz upita i deskriptora ili oznaka korištenih u bazi podataka, za prikaz i korištenje strukture predmetnog područja, za podatke o svakoj riječi potrebne pri oblikovanju upita, za povezivanje riječi u različitim jezicima, zbog čega je izrada rječnika vrlo zahtjevna (Vickery, Vickery 1993, str. 135).

Obrada prirodnog jezika u pretraživanju obično se odnosi na postupak kojim se korisnikov upit oblikovan na prirodnom jeziku, korisnikovim riječima, pretvara u upit s Booleovim operatorima. Ovaj proces opisuju B. Vickery i A. Vickery (1993, str. 121-125). Prvi korak sačinjava sintaktička i semantička analiza, kojom se uklanjaju suvišne riječi iz upita (npr. riječi iz stop-liste), utvrđuje se jedno, kontekstom određeno značenje riječi, te semantički odnosi među preostalim riječima. Za ove postupke potrebna je baza s pravilima za obradu jezika i rječnikom.

Samo prepoznavanje riječi složen je proces zbog različitih slučajeva¹⁶: zbog riječi koje se pišu s povlakom, apostrofom, kosom crtom, zbog (s) u zagradi, brojeva (mogu se pisati s decimalnom točkom, zarezom ili znakom minusa), nizova brojeva (brojevi mogu biti odijeljeni povlakom ili kojim drugim znakom), brojčanih oznaka, slovobrojčanih oznaka, kratica koje sadrže točku, zbog mjernih jedinica koje mogu biti kratice i sastojati se od znakova poput % i \$, ili od broja i naziva mjerne jedinice, zbog matematičkih formula koje sadrže znakove poput +, -, *, zbog kemijskih formula, te znakova interpunkcije poput zareza, točke i zgrade.

Za izdvajanje riječi ili skupina riječi iz upita i za određivanje njihovih međusobnih odnosa postoji nekoliko pristupa. Najjednostavnije je izdvojiti skupine riječi odijeljene funkcionalnim riječima iz stop-liste¹⁷ i svaku skupinu podvrći analizi. Drugi pristup je složeniji, ali vrlo raširen u informacijskom pretraživanju, a sastoji se u primjeni programa za parsiranje (sintaktička analiza). Ovaj program niz riječi iz upita pretvara u gramatičku strukturu, na temelju koje se izvode zaključci o odnosima među terminima. Nadopuna ovome pristupu može biti rječnik, u kojem uz svaku riječ stoji oznaka o vrsti te riječi. Najsloženiji pristup temelji se na semantičkim kategorijama, koje se definiraju za područje čiji se tekst analizira. Ove

¹⁶ Temelje se na engleskom jeziku.

¹⁷ Stop-riječi mogu biti veznici, prijedlozi, članovi i neke nepunoznačne riječi ili nizovi riječi poput "treba mi nešto o" ("I am interested in").

kategorije unaprijed su određene i ručno se unose uz svaku riječ u rječniku. Ovaj pristup najuspješniji je za usko specijalizirane domene.

Idući problem predstavlja izdvajanje višečlanih termina ili izraza (npr. *električna struja*). Prvi pristup sastojao se od određivanja apsolutne frekvencije svih skupina od po dvije i tri riječi koje postoje u dokumentu, a zatim bi se kao višečlani termini odabirale one skupine riječi koje prelaze određeni frekvencijski prag. Suvremeniji pristupi zamjenjuju statistička mjerenja s algoritmima koji se temelje na obradi prirodnog jezika. Pri tome je potrebna baza pravila za određivanje koje će se skupine riječi iz upita prihvatiti kao višečlani termini. Ova pravila mogu se temeljiti na gramatičkim ili semantičkim kategorijama pridruženima svakom terminu.

Jezična višeznačnost predstavlja najveći problem. Različite su vrste jezične višeznačnosti: sintaktička višeznačnost, semantička višeznačnost, homonimija (riječi istog morfološkog oblika, ali nepovezanih značenja), te višeznačnost specifična za pretraživanje, a odnosi se na riječi iz upita koje imaju značenje za pretraživanje i one koje nemaju (*Treba mi informacija o ulozi informacije u demokratizaciji društva; Uporaba Booleovih operatora AND i OR u bazama podataka*) (Svenonius 2000, str. 148); postoje i situacije u kojima se javlja više oblika višeznačnosti zajedno. Sintaktička višeznačnost odnosi se na slučajeve kada jedan oblik riječi predstavlja različite vrste riječi (npr. *igra* može biti imenica i glagol). Ova višeznačnost razrješava se parsiranjem (sintaktičkom analizom), tako da se odredi kategorija riječi prema sufiksu, a ako nema razlike, onda prema kategorijama u koje spadaju susjedne riječi. Semantička višeznačnost odnosi se na slučajeve u kojima postoje dva ili više semantički povezana značenja (“perilica” može biti osoba ili stroj). Ova višeznačnost također se često može razriješiti kontekstom, za što je preduvjet da je u rječniku uz svaku riječ navedena njena definicija i opseg pojma koji predstavlja. Automatske ili poluautomatske metode za utvrđivanje kontekstualnog značenja i rješavanje homonimije često se temelje na analizi susjednih riječi. Npr. riječ *inmate* (zatvorenik, stanar) u blizini riječi *cell* (ćelija, stanica) upućuje da se radi o *prison cell* (zatvorskoj ćeliji), a ne o *blood cell* (krvnoj stanici) (Svenonius 2000, str. 155). Ako upitom nije određen kontekst višeznačnih riječi, korisniku je potrebno postaviti pitanje, npr. “*luk* se odnosi na: (1) oružje, ili (2) povrće”. Također se za svaku višeznačnu riječ u rječniku mogu uvesti pravila specifična za pojedinu riječ, npr. ako iza riječi *informacija* stoji neka od

riječi *o, za, vezana uz, koja se odnosi na* i sl., onda se *informacija* ne uzima kao termin za pretraživanje. Postoje i nastojanja da se izrade općenitija pravila vezana uz stručno područje. Primjerice, ukoliko je na temelju korisničkog profila u bazi podataka poznato stručno područje upita, na temelju tog podatka može se odrediti kontekstualno značenje. Za to je potrebno osigurati da uz svaku riječ i njena značenja budu navedena sva stručna područja kojima može pripadati.

Malo je pomaka postignuto u automatskom razvrstavanju pojmova u semantičke kategorije, djelomično zbog toga što ne postoje dovoljno razvijene baze riječi, koje kombiniraju rječnik i tezaurus, odnosno preslikavanje prirodnog u kontrolirani jezik, metatezaurusi, semantičke mreže, banke termina i razne druge vrste alata za prikaz znanja. Ovakvi alati poboljšat će automatizaciju razvrstavanja termina u semantičke kategorije, što će pomoći i u automatskom označivanju.

Različiti su načini na koje sustavi provode obradu prirodnog jezika u predmetnom pretraživanju (Vickery, Vickery 1993; Smeaton 1997). Prvi je korak obično kraćenje sufiksa riječi iz upita, na temelju čega se pronalaze odgovarajući termini u rječničkoj bazi (to može biti i sustav predmetnih odrednica, npr. MeSH). Na primjer, upit *information retrieval* u kontroliranom rječniku ili u invertiranoj datoteci pronalazi sve pod *information retriev, information, retriev*. Najbolja vrsta kraćenja je svođenje na normirane osnove riječi. Zatim se obično na temelju konteksta utvrđuje značenje pojedinih termina. Pri tome se rabe kontekstualna pravila, ili se korisnik na temelju postavljenih pitanja odluči za neko od postojećih značenja. Daljnji postupak može biti automatska zamjena riječi iz upita terminima iz kontroliranog rječnika ili prikaz popisa relevantnih termina (uključujući, npr. šire, uže, srodne termine) iz kontroliranog rječnika, korisniku na odabir. Metoda utvrđivanja relevantnih termina koja se danas koristi je određivanje “težine” pojedinih termina na temelju broja dokumenata koji sadrže taj termin u bazi, te na temelju broja termina u samom dokumentu. Najpopularnija metoda za određivanje “težine” jest ona koja kombinira ove dvije frekvencije pojavljivanja. “Težina” će se, dakle, izvući statističkom analizom koja se temelji na frekvenciji pojavljivanja riječi u dokumentu, u odnosu na frekvenciju pojavljivanja termina u drugim dokumentima u bazi.

Katalog Nacionalne medicinske knjižnice podržava unos upita na prirodnom jeziku (Vickery, Vickery 1993, str. 161). Nakon što se upit unese, razlomi se na riječi koje se zatim usporede sa stop-listom od 600 riječi i krate se. Svaka osnova

riječi uspoređuje se s popisom riječi s istom osnovom. Zatim se svim pronađenim oblicima s istom osnovom dodijeli “težina”, koja se određuje na temelju njihove frekvencije u invertiranoj datoteci. Svim terminima s istom osnovom dodjeljuje se ista “težina”. “Težine” termina koriste se za izračunavanje “težine” dokumenata, a rezultati se prikazuju rangirano prema “težini” dokumenta. Osnove riječi koriste se i za pretraživanje termina, uputnica, i anotacija MeSH-a. Primjerice, upitom *biofeedback* pronaći će se predmetne odrednice *feedback* i *biofeedback (psychlogy)*. Identificirane predmetne odrednice MeSH nose kombiniranu “težinu” unesenih riječi koje odgovaraju tim predmetnim odrednicama. Unesene riječi iz upita i njihove varijante, predmetne odrednice i pododrednice MeSH rangiraju se po “težini” i prikazuju korisniku na odabir. Korisnik se poziva da odredi koji su termini relevantni, nakon čega sustav analizira ostale termine MeSH korištene za označavanje relevantnih jedinica, te prikazuje popis tih jedinica korisniku na odabir radi provođenja daljnjeg pretraživanja (Vickery, Vickery 1993, str. 107-108).

Slijedi još nekoliko primjera modela informacijskih sustava za pretraživanje koji koriste prethodno navedene metode za računalnu obradu prirodnog jezika, a temelje se na opisima koje su dali B. Vickery i A. Vickery (1993). Mogu poslužiti kao modeli primjene računalne obrade prirodnog jezika za predmetno pretraživanje u online katalozima.

Sustav FASIT ima stop-listu koja, osim uobičajenih riječi poput članova, veznika, zamjenica, svih oblika pomoćnih glagola *to be* i *to have*, sadrži neke opće riječi bez značenjske vrijednosti u označivanju (npr. *certain*, *other*, *basis*), te vlastita imena, iako autori tvrde da bi imena trebalo staviti u posebnu datoteku preglednih kataložnih jedinica i uputnica. Ukoliko riječ iz upita nije na stop-listi, utvrđuje se njezin sufiks na temelju baze od otprilike 450 sufiksa. U ovoj bazi postoji i otprilike 1.000 pravila za utvrđivanje višeznačnih sufiksa, a temelje se na riječima iz upita koje slijede i prethode analiziranoj riječi. Nakon ovog postupka rečenica iz upita pretvorena je u skup kategoriziranih osnova riječi, koje se zatim uspoređuju sa 160 kombinacija vrsta riječi (npr. imenica – prijedlog – imenica), radi utvrđivanja rečeničnih elemenata. Zatim se iz svakog elementa uklanjaju suvišne riječi (npr. *retriev of information* skraćuje se u *retriev information*). Pronalaze se svi oblici koji se u barem jednoj osnovi preklapaju (od *retriev of information* dobiva se *information retriev*, *information* i *retriev*).

Sustav CIRCE nakon izbacivanja suvišnih riječi prema stop-listi, preostale riječi uspoređuje s terminima iz tezaurusa, nakon čega se izradi popis korisnikovih riječi uz pripadajuće termine iz tezaurusa. Uz svaki termin stoji oznaka o njegovom statusu, koja označuje sljedeće: izvor termina (upit ili tezaurus, ako tezaurus, je li korisnik vidio i komentirao termin); ako je korisnik komentirao termin, njegovo razumijevanje termina i njegova procjena relevantnosti za upit; mjera koliko točno termin iz tezaurusa odgovara korisnikovom terminu iz upita; ima li termin više od jednog značenja; te, specifičnost termina (npr. ako je korisnikov termin pronađen u tezaurusu označen kao srednje specifičan, korisniku se predloži da odabere uži termin iz tezaurusa). Pitanja koja se postavljaju korisniku odabiru se na temelju pravila iz 60 rečeničnih predložaka u koje se ubacuju relevantni termini, npr. *Could you please clarify that input?* (Možete li pojasniti svoj upit?), *What is X related to?* (Na što se odnosi X?), *How is X related to Y?* (U kakvom su odnosu X i Y?), *Does X mean the same as Y?* (Znači li X isto što i Y?).

LEXIQUEST se temelji na postupcima sličnima sustavu CIRCE. Posebnost ovoga sustava jest u tome što se nakon izdvajanja riječi koje nisu na stop-listi oduzimaju njihovi pluralni sufiksi, a preostale osnove riječi uspoređuju se s terminima u kontroliranom rječniku.

Još jedna tehnika koja se može primijeniti u obradi prirodnog jezika je strojno učenje (Cunnigham, Witten, Littin 1999). Ono nalazi primjenu i na drugim razinama procesa označivanja i pretraživanja informacija, koje sve utječu na predmetno pretraživanje. Prema D. Lewisu (1991), postoje 4 faze u tom procesu: označivanje, oblikovanje upita, usporedba i povratna informacija. Prva faza, označivanje, obuhvaća utvrđivanje i odabir obilježja (feature extraction), grupiranje dokumenata (document clustering) i klasificiranje teksta (text classification). Najvažniji dio procesa je odabir obilježja i najbolji se rezultati postižu ako se shemama za strojno učenje da određeni broj distinktivnih obilježja na temelju kojih algoritmi uče. Grupiranje dokumenata pomaže u pretraživanju stvaranjem veza između sličnih dokumenata. Time se omogućava pronalaženje sličnih dokumenata, nakon što se jedan dokument procjeni relevantnim za upit. Veze se mogu uspostaviti na bilo kojoj razini, od samih ključnih riječi do sličnih podtema po odlomcima. Slijedeći ove veze, sustav pronalazi dokumente s određenim stupnjem sličnosti i prema tome ih rangira. Ostale tri faze obično tvore petlju koja se ponavlja sve dok korisnik nije zadovoljan pronađenim dokumentima. Kao što se pokazalo u četvrtom

poglavlju, odabir termina za upit korisnicima predstavlja veliki problem. Stoga sustavi koji koriste mehanizam povratne informacije o relevantnosti pronađenih dokumenata korisniku pomažu da odabere najprimjerenije termine, pri čemu se korisniku mogu predložiti i novi termini iz dobivenih relevantnih dokumenata.

Primjena strojnog učenja tek započinje. Tehnike strojnog učenja koje se danas primjenjuju u informacijskim sustavima za pretraživanje jednostavne su, a neki od preostalih problema mogu se riješiti korištenjem novijih i naprednijih tehnologija (Cunningham, Witten, Littin 1999, str. 373).

Ekspertni sustavi računalni su programi koji se sastoje od baze znanja (knowledge base), s informacijama iz vrlo uskog područja ljudskog znanja, i stroja za zaključivanje (inference engine), koji primjenjuje prikladne informacije iz baze znanja za rješavanje određenog problema iz tog uskog područja. Današnji ekspertni sustavi mogu rješavati probleme bolje nego čovjek laik, a koriste se uglavnom u bazama s potpunim tekstovima. Međutim, danas još uvijek postoje problemi za koje se ne vide rješenja za proces pretraživanja jer teško je predvidjeti sve moguće potrebe korisnika. Razvoj ekspertnih sustava s vremenom će možda korisniku omogućiti jednostavno postavljanje upita na prirodnom jeziku i dobivanje preciznih i potpunih rezultata, a do tada će se predmetno pretraživanje svoditi na usporedno korištenje prirodnih i kontroliranih jezika (Rowley 1994, str. 117).

Računalna obrada prirodnog jezika još uvijek ima problema sa sintaksom, a od rješavanja semantičke problematike još je daleko. Savršeno i sveobuhvatno izvođenje vjerojatno je nedosezivo u predvidljivoj budućnosti jer su korišteni modeli prirodnog jezika ograničeni. Ipak, primjenu NLP-a u predmetnom pretraživanju danas ne treba zanemariti jer se u određenim segmentima postižu dobri rezultati, kao što potvrđuju prikazani primjeri. Može se dodati primjer tražilice WebPAC-a knjižnica Filozofskog fakulteta, koja će na temelju tzv. datoteke afiksa za hrvatski jezik uskoro omogućiti automatsko pretraživanje po svim mogućim morfološkim oblicima riječi (Pavlinušić 2002).

6. Provođenje predmetnog pretraživanja

6.1. Operatori

Pri predmetnom pretraživanju po riječima prirodnog jezika operatori omogućuju određen stupanj rješavanja problema sinonimičnosti i homonimičnosti u prirodnom jeziku te problema različitih morfoloških oblika riječi. Omogućavaju pronalaženje pojmova ili aspekata pojmova za koje ne postoji, ili nije poznato da postoji, termin ili oznaka kontroliranog rječnika u bazi podataka. Njima se, također, u upitu mogu kombinirati i termini iz kontroliranog rječnika. Neophodni su u postkoordiniranom predmetnom pristupu, gdje se termini zajedno kombiniraju u trenutku pretraživanja, a ne u trenutku označivanja.

Booleovi operatori temelje se na logici skupova. Prilikom pretraživanja koriste se operacije logičkog zbroja (AND), logičkog produkta (OR) i logičke razlike (NOT), čime se često dobivaju i one jedinice građe u kojima se traženi predmeti ne nalaze, tj. uspostavljene su tzv. lažne veze (Mikačić, 1990, str. 46). U predmetnom pretraživanju operatori AND, OR i NOT stavljaju se između riječi ili skupina riječi ili drugih elemenata u upitu. Ovi elementi u suvremenim sustavima mogu biti ključne riječi, kontrolirani termini, skraćene riječi, klasifikacijske oznake, pa i neverbalni elementi.

Operatorom OR pronalaze se svi dokumenti koji sadrže bilo koji od elemenata koje operator povezuje. Ovaj operator utječe na povećavanje odziva. Koristan je za pronalaženje sinonima, npr. *knjižnično društvo OR knjižnično udruženje OR knjižnična udruga*, kao i varijanti u pisanju i kratica, ukoliko sustavi nemaju automatska rješenja ove slučajeve. Inačica XOR koristi se u sličnoj funkciji, s time što pronalazi bilo koji od dva termina iz upita, ali ne oba. Ukoliko se žele precizniji rezultati i manji odziv, koristi se operator AND. Operator AND pronalazi dokumente koji sadrže obje riječi iz upita povezane ovim operatorom. Operator NOT također povećava preciznost, a smanjuje odziv, te se upotrebljava ukoliko se neki termin želi isključiti iz traženog skupa rezultata. Tako će se, npr. upitom *zagađenje NOT zrak*, pronaći dokumenti o zagađenju općenito i dokumenti o zagađenju voda i tla, ali ne i zraka.

Problem pretraživanja s Booleovim operatorima jesu lažne veze, tj. neočekivani i nerelevantni rezultati koji se dobivaju kada se riječi iz upita nalaze u dokumentu, ali pripadaju drukčijem kontekstu no što je korisnik želio. Smanjivanju broja nerelevantnih rezultata mogu pridonijeti tzv. operatori blizine (proximity operators, positional operators). Ovi operatori korisniku omogućuju da definiira gdje se u polju ili poljima traženi termini nalaze u odnosu jednih na druge. Pretraživanje se najčešće provodi tako da se termini pojavljuju jedan uz drugi ili da su odvojeni s nekoliko riječi u istom polju ili u istom potpolju. Operator ADJ pronalazi riječi koje nisu međusobno udaljene više od određenog broja riječi (*riječX ADJn riječY*, gdje $n=1-99$). Postoji još nekoliko inačica ovoga operatora, npr. WITHIN n i % n , a sličan je i operator ! n koji, osim što određuje maksimalan broj riječi između riječi iz upita (n), određuje i da redoslijed riječi bude isti kao u upitu. Operator SAME pronalazi riječi koje se nalaze u istom odlomku (*riječX SAME riječY*), a operator NEAR riječi u istoj rečenici (*riječX NEAR riječY*). Pojedini informacijski sustavi koriste još neke podvrste operatora blizine, pod različitim nazivima. Iako su ovi operatori vrlo korisni, danas se, osim Booleovih, u online katalozima ostale vrste operatora rijetko primjenjuju (Ortiz-Repiso, Moscoso 1999).

Prisutnost operatora u sustavu može biti eksplicitna, tj. korisnici ih unose sami, ili implicitna, tj. sustav automatski među riječi iz upita postavlja određeni operator. Implicitno postavljanje operatora počelo se uvoditi u sustave jer su istraživanja pokazala da se korisnici često ne znaju služiti Booleovim operatorima. Najčešći implicitni operatori su operator AND, operator blizine i operator OR (Olson, Boll 2001, str. 57).

Online katalozi međusobno se razlikuju po nazivima i vrsti operatora (pogotovo operatora blizine), po vrsti implicitnih operatora te po redoslijedu izvršavanja operacija određenih operatora. Prva generacija online kataloga dozvoljavala je unos samo jedne riječi u upitu, a u upitima s više riječi odnose među riječima bilo je potrebno definirati Booleovim operatorima. Od druge generacije online kataloga dozvoljava se unos više riječi u upitu, a sustav sam određuje odnose među njima: jedni sustavi riječi povezuju operatorom OR, drugi operatorom AND, neki operatorom blizine, a ostali unesene riječi smatraju izrazom. Jedni poštuju hijerarhiju Booleove algebre pa prvo izvršavaju operacije s NOT, zatim s AND, te na kraju s OR, a drugi ne pa operacije izvršavaju s lijeva na desno ili nekim drugim

redosljedom, neki dozvoljavaju zagrade pa prednost imaju operacije u zagradama, a neki ne.

Pokazalo se da je Booleova logika teško razumljiv koncept (Yee 1991) te da čak i znanstvenici i inženjeri koji imaju iskustava u primjeni logike na drugim aplikacijama često pri pretraživanju koriste AND i OR obrnuto (Borgman, Case and Meadow 1989), a istraživanja iz kognitivne psihologije (Tversky, Kahneman 1974) pokazuju da ljudi slijede intuiciju pa AND i OR koriste u njihovom jezičnom značenju, tj. da je AND inkluzivno i da znači više, a OR da je ekskluzivno i da smanjuje skup – dakle, obrnuto od Booleove logike (Borgman 1996). C. Hildreth (1997) proveo je istraživanje druge generacije online kataloga i ustvrdio da korisnici nemaju nikakvog znanja o tome kako implicitni Booleovi operatori funkcioniraju.

C. Borgman (1996) tvrdi da je za oblikovanje uspješnih Booleovih upita potrebno konceptualno znanje o načinima na koje se pojmovi za pretraživanje mogu kombinirati, te semantičko znanje o tome kako neki sustav obrađuje ili provodi Booleove upite. Niska razina tih dviju vrsta znanja osnovni je čimbenik koji dovodi do nezadovoljavajućih rezultata pretraživanja. Budući da se online katalogi razlikuju u korištenju operatora, korisnici često ne dobivaju tražene rezultate jer uglavnom nemaju volje ispitati kako se u sustavu koriste operatori, već prenose znanje koje imaju o sustavu koje su prije koristili (Borgman 1996, str. 498). C. Borgman predlaže dva rješenja ovih problema online kataloga: poboljšanje metoda obučavanja korisnika i poboljšanje sučelja online kataloga koje odražava ponašanje vezano uz informacije. Drugo rješenje puno je praktičnije ni iz jednog drugog razloga no što će uvijek biti manje sustava koje treba usavršiti nego korisnika koje treba obučiti. Neovisno o tome, WebPAC-i u hipertekstualnom okruženju imaju potencijal za interaktivnu funkcionalnu pomoć tražiteljima (Hildreth 1997, str. 62), što svakako valja iskoristiti i u ovome kontekstu.

6.2. Kraćenje

Pri oblikovanju upita s operatorima često se koristi kraćenje. Ono omogućava okupljanje različitih morfoloških oblika iste riječi, npr. jednine i množine ili imenice, glagola i pridjeva.

Značenje termina *kraćenje* ili *krnjenje* u hrvatskome jeziku obuhvaća značenje triju engleskih termina: *truncation*, *stemming* i *wildcard*. Ovi termini razlikuju se u sljedećem: *truncation* se najčešće koristi u značenju kraćenja sufiksa i afiksa za određen broj znakova bez obzira na korijen riječi, *stemming* se koristi za pronalaženje svih riječi iz upita koje imaju isti korijen, a *wildcard* se najčešće odnosi na znakove koji se mogu staviti u sredinu riječi iz upita i zamijeniti jedan ili više znakova. Uobičajen primjer za korištenje znakova *wildcard* jest za različite varijante riječi, kao npr. u engleskom jeziku *color* (američka varijanta) i *colour* (britanska varijanta).

Ovisno o katalogu, kraćenje može biti eksplicitno, tj. određeno od strane korisnika, i automatsko ili implicitno, kada sustav sam krati određen broj znakova. Najčešće se krati sufiks nakon određenog znaka, npr. sedmog, kako bi se upitom obuhvatili svi oblici riječi koja može imati različite nastavke (npr. ovisno o rodu, broju, padežu). Kada je kraćenje eksplicitno, umjesto jednog ili više znakova obično se stavlja neki od sljedećih znakova: *, ?, \$, #, !, ili %. Eksplicitno kraćenje zahtijeva od korisnika da se prisjeti svih mogućnosti, a implicitno ne uzima u obzir kontekst već automatski krati po postojećem algoritmu ili postojećim pravilima, pa se postavlja pitanje koje je rješenje najbolje za sustav.

Jasno je da kraćenje nije pravo rješenje za sve probleme pretraživanja jer ne omogućava pristup širim i užim terminima, sinonimima, varijantama u pisanju ili inačicama na drugim jezicima. Za ove slučajeve sustav treba osigurati kontrolirani sustav, ili se korisnik sam mora dosjetiti svih mogućnosti. Međutim, kraćenje je potrebno osigurati i po terminima iz kontroliranog rječnika.

Oblikovanjem upita za pretraživanje korištenjem kraćenja mogu se dobiti i neočekivane i nerelevantne riječi. Primjerice, ako u katalogu na engleskom jeziku tražimo jedinice građe o čitanju i unesemo *read?*, možemo dobiti zapise koji sadržavaju riječi *readable*, *reader*, *readership*, *reading*, ali i *readapt*, *readiness*, *readdress*, *readies*, *readily*.

6.3. Oblikovanje upita

Pronalaženje informacija temelji se na interakciji korisnika s bazom informacijskog sustava. Korisnik postavlja upit koji se tipično sastoji od riječi koje se uspoređuju s pristupnicama u bazi podataka. Osim ovoga načina, informacijski sustavi pružaju mogućnost odabiranja pristupnica iz ponuđenog popisa, koja se tehnika pretraživanja naziva *prebiranjem* ili *pregledavanjem* (browsing). Postoje i napredne tehnike pretraživanja uspoređivanjem riječi uz upita s pristupnicama u bazi podataka, poput tzv. *latentnog semantičkog označivanja* (latent semantic indexing), kojom se riječi iz upita s predmetnim pristupnicama u bazi podataka uspoređuju na temelju statističke semantičke analize upita i odnosa među riječima u upitu. Ovaj se način razlikuje od prethodnih po tome što omogućuje pronalaženje relevantnih jedinica građe i kada se riječi iz upita i pristupnice razlikuju morfološki, ali se preklapaju semantički (Hjorland, Nielsen 2001, str. 251).

Osnovni problem s kojim se korisnik susreće pri pretraživanju jest prevođenje upita na jezik sustava. Za to je potrebno uspostaviti datoteku predmetnih kataložnih jedinica i uputnica i slične mehanizme koji će korisnika uputiti s nekorisćenih termina na korištene, istodobno im pokazujući šire, uže i srodne pojmove te bilješke o upotrebi. Pitanje koje se postavlja jest da li upit automatski prevesti na kontrolirani jezik ili korisniku pokazati mogućnosti i dozvoliti mu da sam napravi odabir. Termini za pretraživanje mogu se korisniku predložiti i nakon provedenog pretraživanja, na temelju povratne informacije o relevantnosti dobivenih dokumenata, kada se iz zapisa koje korisnik odredi kao njemu relevantne, ponude na odabir ostali termini iz tih zapisa. Osim toga, termini mogu biti slični jer se zajedno citiraju, jer su međusobno povezani hipervezom, ili su sličnih statističkih distribucija, na čemu se u nekim informacijskim sustavima također temelje određene tehnike za prikaz sličnih termina (Svenonius 2000, str. 162).

Veliki problem, dakle, pri pretraživanju predstavlja izbor riječi za upit. U tom kontekstu provode se istraživanja o načinu preoblikovanja upita, preklapanju riječi iz upita i termina u kontroliranom rječniku i pogreškama u pisanju. Rezultati ovih istraživanja upućuju da su upiti koje trebaju postavljati krajnji korisnici loši te da je potrebno poboljšati učinkovitost novim modelima koji će smanjiti kognitivni napor potreban za oblikovanje strategije pretraživanja koju informacijski sustav

razumije. Ove metode obuhvaćaju interaktivno pretraživanje preoblikovanjem upita (gore prikazano) i neku vrstu prebiranja (Bates 1989a; Marchionini 1997).

Osim upisivanja riječi koje se zatim uspoređuju s pristupnicama u bazi sustava, odnosno oblikovanja upita svojim riječima, korisniku online kataloga na raspolaganju može biti i mogućnost odabira termina putem izbornika iz pojedinih polja, tj. prebiranje. Prebiranje se u literaturi predlaže kao druga mogućnost kada korisnik ne zna što točno traži ili u sustavu ne postoji točan odgovor na njegov upit. Prebiranje po indeksima pojedinih polja korisno je i zbog toga što korisnici ne trebaju znati kako se točno neka riječ piše, a time se sprječavaju i pogreške u pisanju, odnosno pravopisne pogreške. Ono se može definirati i kao istraživanje baze podataka s idejom da se slijeđenjem veza može naići na nešto zanimljivo ili korisno (Olson, Boll 2001, str. 287). Posebna vrsta prebiranja je ona u kojoj se kao odgovor na upit iskazan korisnikovim riječima prikazu rezultati, ali i ostale susjedne jedinice građe, bilo da su poredane po abecedi naslova ili autora, relevantnosti ili nekom drugom kriteriju.

Novi modeli prikaza sadržaja jedinica građe zamjenjuju klasičan sustav u kojem se riječi iz upita koje daje korisnik uspoređuju s terminima u invertiranoj datoteci baze kataloga. Kada govorimo o prebiranju, tradicionalni kontrolirani rječnici prilagođeni online okruženju učinkoviti su alati, posebno u kombinaciji s naprednim tehnikama informacijskog pretraživanja poput onih koji koriste računalnu obradu prirodnog jezika. Tome je razlog to što ovakvi alati pospješuju konceptualno znanje korisnika o informacijskom sustavu, njegovom sadržaju i o načinu na koji je informacijski prostor organiziran, olakšavajući mogućnost odabira termina za proširenje ili preoblikovanje upita.

Danas mnogi online katalogi ne iskorištavaju prednosti mrežnog okruženja pa se prebiranje svodi na pristup terminima linearno, tj. svaki je termin povezan samo s terminom koji mu neposredno prethodi ili slijedi (Ortiz-Repiso, Moscoso 1999). Prebiranje bi se trebalo temeljiti na hipervezama, trebalo bi biti nelinearno i fleksibilno, tako da korisnik može jednostavno odlaziti s termina na termin, već prema tome kakve odnose među terminima žele slijediti. Kao dobar primjer iskorištavanja prednosti mrežnog okruženja može poslužiti portal *Highwire Press* Sveučilišta Stanford (University of Stanford), gdje se osigurava pristup elektroničkim publikacijama baza podataka, a podržavaju se napredne funkcije pretraživanja, uključujući po referencama, te dvije vrste prebiranja po predmetnim

područjima: linearno i po tematskoj mapi (TopicMap), gdje su popisane struke, predmeti unutar struke, a pod svakim predmetom nalazi se popis postojećih radova (Atkins 2002).

6.4. Rangiranje i prikaz rezultata

Korištenje prvih online kataloga zahtijevalo je precizan unos riječi u upit jer su rezultati bili samo oni koji su točno odgovarali upitu (to su tzv. *exact match* sustavi). Kasniji katalogi postali su tzv. probabilistički sustavi (ili *best match* sustavi), koji omogućuju pronalaženje i onih rezultata koji samo djelomično odgovaraju upitu. U ovim je sustavima svakako poželjno da zapisi ili dokumenti dobiveni pretraživanjem budu rangirani po relevantnosti, iz nekoliko razloga. Rangiranje prema procijenjenoj sličnosti s upitom omogućava korisniku da sam odluči u kojem trenutku da prestane s pregledavanjem rezultata jer nema potrebe da ih pregledava sve, jedan po jedan (Smeaton 1997). Rangiranje rezultata olakšava sustavu da na temelju povratne informacije ponovi pretraživanje prema onome što korisnik odabere kao relevantne dokumente, tako da doda ili izbacuje termine iz upita.

Skupni katalog Bopac2¹⁸, pokrenut 1996. godine s ciljem umrežavanja kataloga različitih knjižnica putem protokola Z39.50, posebno se bavio mogućnostima grupiranja rezultata po sličnim kriterijima. Kada se dobije više od pet rezultata, oni se prikazuju u tzv. *Full Cluster View*, a na temelju tzv. okupljajućeg (clustering) algoritma, pri čemu se prikazuje naslov, autor i dio zapisa koji sadrži termine iz upita. Na taj su način slični zapisi okupljeni, i to prema naslovu, autoru, jeziku, vrsti građe ili po knjižnici, a predviđeno je i okupljanje po predmetu, na temelju postojećih kontroliranih termina iz zapisa MARC i putem tehnika za određivanje relevantnosti kakve se koriste u informacijskom pretraživanju (Ayres, Nielsen, Ridley 1999, str. 28-34).

Rangiranje rezultata po relevantnosti je, međutim, još uvijek na početnom stupnju razvoja. Eksploatira se najviše u bazama s potpunim tekstom i na web-tražilicama, dok softveri za online kataloge više ili manje ove funkcije ne

¹⁸ URL: <http://www.bopac2.comp.brad.ac.uk/~bopac2/>

podržavaju. Rangiranje rezultata temelji se na pretpostavci da što pojedinih riječi u jednom dokumentu ima više, to je veća vjerojatnost da se u tom dokumentu radi o pojmu označenom tom riječju. Stoga se pretraživanje temelji na usporedbi riječi iz upita s riječima u invertiranoj datoteci¹⁹, pri čemu se svakoj riječi dodjeljuje “težina” (weight), vrijednost koja odražava vjerojatnost da ta riječ prikazuje sadržaj dokumenta. Ova vrijednost dobiva se statističkom analizom na temelju čestoće te riječi u zapisu ili dokumentu, obično u odnosu na čestoću riječi u drugim zapisima ili dokumentima u bazi i u odnosu na dužinu dokumenta. Ova tri postupka temelje se na sljedećim pretpostavkama: termini koji se javljaju samo u nekim zapisima ili dokumentima vjerojatno su korisniji od onih koji se javljaju u mnogima; što je riječ češća u dokumentu, to je vjerojatnije da je važnija za taj dokument; te, da je riječ koja se javlja jednak broj puta u kratkom dokumentu kao i u dugačkom vjerojatno važnija za kratak dokument nego za duži. Sve se ove “težine” kombiniraju da bi se dobila najpreciznija “težina” za svaki termin. B. i A. Vickery (1993, str. 151) navode sljedeće tehnike za određivanje “težine” termina: veća vrijednost pridaje se izvornim terminima (onima iz upita), nego naknadno dodanim sinonimima i ostalim terminima za preoblikovanje upita; subjektivno dodjeljivanje vrijednosti o važnosti termina od strane korisnika; te, određivanje obrnute vrijednosti u odnosu na frekventnost riječi u relevantnim i nerelevantnim jedinicama, pri čemu korisnik radi procjenu relevantnosti jednog uzorka rezultata. Rezultati se ispisuju prema stupnju u kojem odgovaraju upitu u padajućem slijedu.

Slijedi primjer kako je rangiranje po relevantnosti riješeno u jednoj bazi podataka, prema B. Vickeryu i A. Vickery (1993, str. 146). Vrijednost svakog termina izračuna se prema formuli $w = \log(N/n)$, gdje je N broj jedinica u zbirci dokumenata (unaprijed poznat), a n je broj pojavljivanja termina u dobivenim rezultatima. Vrijednost dobivenog skupa rezultata je ukupan zbroj vrijednosti pronađenih termina. Zatim se dobiveni skup rangira i prikazuje korisniku. Korisnik tada može pojedine dokumente odrediti kao relevantne, nakon čega se vrijednosti skupa ponovno izračunaju. Isto se tako može dodati novi termin, pa će iznova biti potrebno izračunati vrijednosti, ili se termin može izostaviti, pri čemu će njegova vrijednosti biti jednaka nuli.

¹⁹ Kao što je ranije istaknuto, u online katalozima i bazama podataka riječi iz zapisa i/ili dokumenata koje nisu na stop-listi nalaze se u invertiranoj datoteci.

Istraživanja su pokazala da različiti genetski algoritmi za povratnu informaciju o relevantnosti rezultata pretraživanja imaju dobru mogućnost primjene i kao takvi poboljšavaju mogućnosti sustava za pretraživanje (vidi, npr., Lopez-Pujalte, Guerrero Bote, de Moya-Anegon 2002).

Još jedna tehnika koja se može razmotriti za primjenu u predmetnom pretraživanju u bazama s hipertekstualnim dokumentima jest analiza strukture hiperveza, koja se sve više koristi u algoritmima za pretraživanje i rangiranje rezultata u web-tražilicama na Mreži (vidi, npr., Lempel, Moran 2002).

No, kao što tvrdi E. Svenonious (2000, str. 191), računalna automatska obrada može pomoći u rangiranju po relevantnosti na temelju statističkih i drugih postupaka, ali semantičko rangiranje bez ljudske pomoći nije moguće.

7. Korisničko sučelje za predmetno pretraživanje

7.1. Oblikovanje sučelja za predmetno pretraživanje

Što se tiče uključivanja korisnika pri pretraživanju, važno je odlučiti koji se stupanj fleksibilnosti želi pružiti korisniku – neki koraci mogu se automatizirati, ili ih može obavljati korisnik (Vickery, Vickery 1993, str. 110-111). Pojedini sustavi dozvoljavaju unos upita koji se zatim dalje obrađuje i prenosi u bazu, dok neki očekuju od korisnika da aktivno pridonosi oblikovanju upita. To se odnosi na funkcije poput sljedećih: da li koristiti implicitne ili eksplicitne Booleove operatore, da li koristiti automatsko kraćenje ili ne, automatsko prevođenje na kontrolirani rječnik ili ponuditi popis kontroliranih termina na izbor korisniku.

Istraživanja su pokazala da korisnici trebaju imati nadzor nad algoritmima, poput onoga za automatsko kraćenje, jer taj algoritam može tjerati korisnika da traži dalje i dalje, a da odgovarajuća jedinica u katalogu zapravo ne postoji (Mandel, 1986; McCarn, 1983). Okapi koristi složene algoritme koji pomažu korisniku u pronalaženju najbližih termina nakon što tzv. *exact matching* (potpuno preklapanje) nije dovelo do rezultata (Hancock-Beaulieu 1992, str. 33-34). No, korisnik bi nad njima trebao imati nadzor. Nije dobro činiti nešto što korisnik nije zatražio da se učini i bez njegova znanja. Npr. ako korisnik traži informaciju o temi koja nije korištena, izuzetno bi bilo korisno izbaciti popis predmetnih odrednica, prikazujući “Ova knjižnica koristi '“ i da se hipervezom odmah ode na tu predmetnu odrednicu ili se odabere u upit i kombinira s drugim terminima u upitu. Ukoliko se ne može prikazati popis, dobro je prikazati poruku s tekstom, npr. “U ovoj se knjižnici ne koristi izraz *osobe s invaliditetom*, već *osobe s posebnim potrebama*”. Međutim, automatski promijeniti pretraživanje nije u redu (Crawford 1992, str. 69).

Također je važno da u svakom trenutku rada sa sustavom korisnik zna gdje se nalazi u procesu, koji zadaci su završeni, i što je još potrebno riješiti. Istodobno na zaslonu trebaju postojati tekući rezultati dobiveni od sustava, poruka s uputama, informacije o mogućnostima koje korisniku stoje na raspolaganju (npr. pomoć, provođenje određenih zadataka) (Vickery, Vickery 1993, str. 152-153). Istraživanja su pokazala da se korisnici ne snalaze u katalozima i da je pri oblikovanju sustava

potrebno osigurati odgovore na pitanja poput *Gdje sam? Kako se mogu vratiti na popis zapisa? Što bih sada trebao učiniti?* (Matthews 1997, str. 112).

Predmetne odrednice u zapisu treba istaknuti te povezati hipervezama s drugima. Mora postojati mogućnost da se pronađu djela vezana uz prikazani zapis, ili djela vezana uz prikazanu predmetnu odrednicu ili signaturu (Hildreth 1995).

Za ispravljanje pogrešaka i ujednačavanje pravopisnih varijanti najbolje bi bilo razviti softvere. Softveri su poželjni i za američku i britansku varijantu engleskoga jezika. Međutim, važno je za funkcije koje će softveri podržavati razviti standarde, jer, npr. danas neki softveri pretvaraju povlaku u razmak pa se riječ indeksira kao dvije riječi, a drugi je brišu pa se riječ indeksira kao jedna riječ. Korisnici ne bi trebali pamtititi kako koji sustav radi do ovakvih detalja.

Važno je da korisnik može mijenjati redoslijed akcija u pretraživanju i vratiti se na prethodne upite i promijeniti ih. Istraživanja pokazuju da mnogi zadaci vezani uz pretraživanje informacija mogu imati koristi od automatski i/ili ručno snimljenih prethodnih pretraživanja (Komlodi, Soergel 2002, str. 152). Pokazalo se da korisnici često ponavljaju dijelove svojih upita pri provođenju pretraživanja, što upućuje da zaboravljaju što su već probali, i na potrebu za nekom vrstom mogućnosti pregleda prošlih pretraživanja. Poželjno je i da se prošla pretraživanja mogu provesti i na nekom drugom sustavu.

Na temelju istraživanja (Large, Beheshti 1997) zaključeno je da korisnici pri pretraživanju koriste i ključne riječi i kontrolirani rječnik, te bi obje opcije trebale biti ponuđene na prvoj stranici za pretraživanje. Također se pokazalo da čimbenici poput redoslijeda utječu na korisnikov izbor, i to tako da su češće odabrani oni prvi navedeni. Stoga C. Long (2000, str. 166) predlaže da se na vrh početne stranice za pretraživanje stave one mogućnosti za koje postoji najveća vjerojatnost da će biti najučinkovitije.

Istraživanja su pokazala da korisnici ne poznaju rad sustava i da ne razlikuju pretraživanje po ključnim riječima od pretraživanja po predmetnim odrednicama (Hildreth 1997; Borgman 2000). Stoga je potrebno uz svaku opciju za pretraživanje ponuditi detaljne upute da bi korisnici znali odabrati što im najbolje odgovara. U tom smislu, korisnici trebaju znati i koji se dio bibliografskog zapisa pretražuje kad se radi o ključnim riječima, kako bi dobili ideju u opsegu i ograničenjima svojeg pretraživanja. Korisnici trebaju znati čemu služi kontrolirani rječnik, koji se kontrolirani rječnik koristi, da se kontrolirani rječnik može nezavisno pretraživati i

gdje se nalazi. Budući da korisnici pregledavaju predmetne odrednice u dobivenim zapisima s ciljem preoblikovanja upita (Long 2000, str. 167; Slone 2000, str. 758), dodavanje hiperveza na ostale zapise koji sadrže relevantne predmetne odrednice, kao i na deskriptore kada se unese nedeskriptor, te na šire, uže i srodne deskriptore, jedno je od najkorisnijih svojstava koja sučelja WebPAC-a mogu pružati. Posebno označivanje traženih pojmova (npr. bojom) u dobivenom zapisu može pomoći pri određivanju relevantnosti dobivene jedinice (Long 2000, str. 167).

Za poboljšanje predmetnog pretraživanja, C. Long (2000, str. 170) priredio je sljedeće smjernice za oblikovanje sučelja WebPAC-a (pored ovih, priredio je i smjernice vezane uz pristupačnost i upute, o kojima će kasnije biti riječi):

1. Na početnoj stranici za pretraživanje trebale bi postojati obje mogućnosti pretraživanja, i za ključne riječi i za predmetne odrednice, te bi trebale biti dovoljno istaknute.
2. Upute uz pojedine mogućnosti pretraživanja trebale bi jasno opisati koja se vrsta pretraživanja provodi.
3. U opciji za ključne riječi potrebno je navesti koja će se polja iz bibliografskog zapisa pretraživati.
4. Korisnicima je pri pretraživanju po predmetnim odrednicama potrebno reći da pojmovi iz njihova upita za pretraživanje moraju odgovarati onima iz popisa predmetnih odrednica.
5. Potrebno je navesti podatak o vrsti korištenog popisa predmetnih odrednica.
6. Popis predmetnih odrednica korišten u katalogu trebao bi biti pretraživ online putem.
7. Korisnici bi trebali imati na raspolaganju alternativni način pretraživanja po predmetu, poput ključnih riječi po polju za predmetne odrednice ili pretraživanja po klasifikacijskoj oznaci.
8. Predmetni sustav kataloga trebao bi imati izgrađen sustav uputnica.
9. Dobiveni zapisi trebali bi sadržavati hiperveze s predmetne odrednice na ostale zapise s istom predmetnom odrednicom.
10. Korisnikovim pojmovi za pretraživanje trebali bi u dobivenim zapisima biti posebno istaknuti.

7.2. Pomoć i upute

WebPAC treba osigurati da novi korisnici upoznaju sustav s minimalnim potrebnim uputama, istodobno omogućujući iskusnijim korisnicima da provode složena istraživanja. Kao što je ranije spomenuto, za poboljšanje predmetnog pretraživanja mogu se oblikovati metode obučavanja korisnika i prilagoditi sučelja online kataloga koja će odražavati ponašanje korisnika pri pretraživanju (Borgman 1996). Ovo rješenje puno je praktičnije ni iz jednog drugog razloga no što će uvijek biti manje sustava koje treba usavršiti nego korisnika koje treba obučiti. Neovisno o tome, WebPAC-i, u hipertekstualnom okruženju, imaju potencijal za interaktivnu funkcionalnu pomoć tražiteljima (Hildreth 1997, str. 62), što svakako valja iskoristiti.

Ono što se predlaže kao općenita pomoć koju treba osigurati korisniku jest sljedeće: nakon ulaska u sustav, pokazati zaslon s osnovnim informacijama o korištenju kataloga, osigurati upute o mogućnostima pretraživanja drugih kataloga ako takve postoje, osigurati novosti o knjižnici, uslugama, katalogu, upute o povratku na prethodne korake i o tome kako izaći iz sustava ili kako se vratiti na početni zaslon. Sama "pomoć" može biti eksplicitna (na zahtjev) ili implicitna (sama se poziva ovisno o kontekstu, a za nove korisnike mogu se osigurati tutoriali²⁰).

Istraživanja su pokazala da su upute koje online knjižnični katalogi pružaju nedostatne te da ih je potrebno osigurati pri svakom koraku tijekom pretraživanja (Borgman 1996; Hancock-Beaulieu 1992; Hildreth 1995). Važno je da upute za pomoć budu stalno dostupne dok korisnik obavlja radnju, a ne da se prozor s uputama mora zatvoriti radi druge aktivnosti.

Pomoć od računalnog zaslona trebalo bi omogućiti pri preoblikovanju upita, npr. pri promjeni operatora ili termina za pretraživanje. Predmetne odrednice u dobivenim zapisima mogu biti korisne za daljnje pretraživanje, ali korisnike je na to potrebno uputiti, kao i osigurati pomoć o uputnicama u knjižničnim katalogima, koje su neophodne, ali kako one funkcioniraju nije uvijek jasno korisnicima (Wasylenko 2000). Istraživanja su pokazala (vidi, npr., Zumer, Lei 1994) da, iako većina online kataloga, podržava eksplicitno kraćenje sufiksa ključnih riječi, to nigdje ne navode u

²⁰ Vidi, npr., tutoriala Nacionalne medicinske knjižnice (NLM LocatorPlus help, tutorials and FAQs), URL: <http://locatorplus.gov/help.html>

uputama pa korisnik nije svjestan mogućnosti. Pokazalo se da bi korisnici voljeli da ih katalog uputi kada je poželjna koja vrsta pretraživanja i da uz svaku vrstu pretraživanja stoji uputa.

C. Borgman (1996) tvrdi da korisnici online kataloga trebaju više pomoći s konceptualnim aspektima pretraživanja, a da je danas pomoć pri korištenju online kataloga uglavnom usmjerena na tehničke vještine, ostavljajući sa strane one vještine u kontekstu semantičkog, a još manje konceptualnog znanja. Upute za korisnike trebale bi osigurati pomoć u smislu kako mogu prevesti svoje ideje na pojmove korištene u katalogu, kao i upute o tehničkim vještinama, npr. kako se provodi samo pretraživanje. Posebno je važno (Hildreth 1995) osigurati pomoć ako na upit nema odgovora, bilo da se korisnik usmjeri općenito na to kako oblikovati upit, ili kako proširiti upit. Čak i ako na upit postoji određen broj odgovora, dobro je osigurati pomoć jer dobiveni zapisi možda ne zadovoljavaju korisnikove informacijske potrebe.

U tom smislu C. Long (2000, str. 172) preporučuje sljedeće:

1. Potrebno je osigurati online kontekstualne upute o načinu provođenja pretraživanja po ključnim riječima i predmetnim odrednicama.
2. Upute za pretraživanje po ključnim riječima i predmetnim odrednicama trebale bi, osim postupaka pretraživanja, obuhvatiti konceptualne aspekte pronalaženja.
3. Pomoć je potrebno ponuditi kada nema pronađenih zapisa.
4. Pomoć je potrebno ponuditi i ako je određen broj zapisa pronađen.

7.3. Vrste korisničkih sučelja

Korisničko sučelje točka je komunikacije sustava s korisnikom. Općenito bi korisničko sučelje trebalo osigurati pristup informaciji na jednostavan, razumljiv i fleksibilan način, kako korisnici ne bi trebali trošiti vrijeme na upoznavanje sa sustavom.

Iako se današnji online katalogi znatno razlikuju po svojim mogućnostima, više ili manje osiguravaju barem dvije vrste sučelja, komandno i ono s izbornicima.

Komandno sučelje može se smatrati sučeljem u kojem je korisnik osnovni pokretač i voditelj, dok je u sučelju s izbornicima to sustav. Često postoje dvije inačice sučelja s izbornicima, jedno za početnike, a drugo za napredne korisnike. Komandno sučelje danas koriste uglavnom knjižničari i informacijski stručnjaci.

Komandno sučelje prva je vrsta sučelja korištena u online katalozima. Ovim sučeljem danas se služe uglavnom iskusni korisnici i informacijski stručnjaci jer je potrebno poznavati naredbe. Sustavu korisnik daje naredbu na koju sustav odgovara, a svaka se naredba sastoji od operatora, koji govori o kojoj se vrsti naredbe radi, i operanda, riječi koja se traži u zapisima ili dokumentima (npr. u upitu *SU rivers*, *SU* (SUBject) je operator koji označava da se radi o predmetnom pretraživanju, a *rivers* je operand, tj. upit traži riječ *rivers* u nekom od informacijskih sustava). Problem je taj što gotovo svaki sustav ima vlastite operatore, čak i u Sjedinjenim Američkim Državama, unatoč činjenici što je 1992. donesen standard o zajedničkom komandnom jeziku (ANSI/NISO Z39.58-1992).

Sučelje s izbornicima predočava korisniku mogućnosti koje mu stoje na raspolaganju i iz kojih može birati. U ovom su sučelju naredbe skrivene od korisnika pa korisnici ne moraju poznavati operatore za pojedine funkcije. Starije vrste sučelja s izbornicima zahtijevale su navigaciju uz pomoć tipkovnice i pritisak od dvije ili tri tipke istodobno radi odabira željenog izbora. Razvojem grafičkih korisničkih sučelja (graphical user interface – GUI), u kojima se u prozoru otvaraju izbornici, omogućeno je da se mišem odabere ili aktivira željeni izbor. Manje iskusni korisnici radije se služe ovim sučeljem, iako su mogućnosti često nedovoljno objašnjene i jasne (Olson, Bol 2001, str. 280). Korisnici često ne razlikuju predmetne odrednice od ključnih riječi i klasifikacijskih oznaka, primjerice. U izbornicima treba postojati mogućnost hijerarhijskog prebiranja, bez obzira kojom pristupnicom korisnik započinje pretraživanje. Za izradu ovakvih uputa i za jasan i pristupačan dizajn nedostaju standardi. Za iskusna korisnika u sučelju s izbornicima potrebno je osigurati “kratice”, odnosno mogućnosti koje mogu odmah odabrati, a ne moraju ići korak po korak.

Jedna inačica sučelja s izbornicima je sučelje s kućicama za unos riječi (fill-in form), kakvo je u WebPAC-ima najčešće. Sastoji se od kućica ili praznina u koje korisnik unosi traženu informaciju (operand), bilo da se ta kućica nalazi uz unaprijed određeno polje, npr. za naslov, ili se uz kućicu nalazi padajući izbornik u kojem korisnik određuje operator (naslov, ključna riječ i sl.), a ove se kućice često

moгу povezati Booleovim operatorima ili njihovim opisnim inačicama (npr. *sve riječi, bilo koja riječ, i, ili, je, nije*). Grafičko sučelje korisniku omogućava da vidi popis predmetnih odrednica ili tezaurus istodobno kao i rezultate pretraživanja, što je vrlo važno pri oblikovanju pretraživanja.

U današnjem sve globalnijem informacijom okruženju potrebno je osigurati sučelja koja će biti pristupačna svima, neovisno o fizičkim, kognitivnim, emocionalnim i kulturološkim različitostima, uključujući višejezična sučelja (Marchionini 1992, str. 162).

Pitanje koje se postavlja jest je li moguće definirati skup globalnih strategija za pretraživanje, a da je svaka prikladna za određeni stereotip korisnika? Sustav mora zadovoljiti golemu raznolikost osobnih karakteristika korisnika, njihove kognitivne sposobnosti i zadatke koje imaju. Za njihove zadatke nije potrebna samo prikladna informacija koja se traži i pronalazi, već i podrška za donošenje odluka pri pretraživanju. U stručnoj literaturi nalazi se na dva rješenja: oblikovanje različitih sučelja za različite razine znanja korisnika (npr. za osnovno i napredno pretraživanje), te oblikovanje individualiziranih sučelja za svakog korisnika.

Oblikovanje individualiziranih sučelja smanjuje korisnikov napor u odabiru relevantnih jedinica građe u promjenjivom informacijskom okruženju. Stalno ažuriranje sučelja, uz pomoć korisnikove povratne informacije, dozvoljava sustavu da prati promjene u interesima i informacijskoj potrebi (Cunningham, Witten, Littin 1999, str. 373). Idealno bi bilo individualizirano sučelje prilagođeno potrebama svakog korisnika i ekspertni sustav koji “uči” kako korisnik pretražuje i njemu se prilagođava. Teško je predvidjeti kada bi takvi sustavi mogli postati realnost, no sučelja s izbornicima možda će biti zamijenjena sučeljima koja se temelje na drugim vrstama veza, a izrastaju iz jezika za označavanje poput HTML-a, SGML-a i XML-a (Olson, Boll 2001, str. 203-204).

7.4. Višejezičnost

O višejezičnosti u online katalozima, kao o jednom od ključnih pitanja u globalnom informacijskom povezivanju, provode se mnogobrojna istraživanja, no u ovome će

se radu problematizirati samo o segmentu potrebnom za upotpunjavanje slike o predmetnom pretraživanju u WebPAC-ima.

E. Svenonius (2000, str. 197) tvrdi da je potrebno razvijati višejezična sučelja za pretraživanje, što je jedna od sastavnica u ostvarivanju univerzalne bibliografske kontrole. Osim višejezičnih sučelja, višejezičnost se može osigurati i pri postavljanju upita, prevođenjem korisničkog upita na jezik baze koja se pretražuje, za što je potrebno da sučelje omogući obradu upita na više od jednoga jezika i pristup višejezičnim rječnicima (Vickery, Vickery 1993, str. 137).

Kada se radi o višejezičnim sučeljima, problem je relativno jednostavan jer je oblikovati nekoliko istih sučelja na različitim jezicima prilično jednostavno. Današnje web-tražilice nastoje podržavati višejezičnost i to uspijevaju do različitih stupnjeva (vidi, npr. Chan, Lin, Zeng 2000). Ono što bi WebPAC-i od njih mogli preuzeti je prepoznavanje domene za državu iz koje korisnik pristupa katalogu i automatsko prikazivanje sučelja na jeziku te države.

Preduvjet za prihvaćanje upita na jednom jeziku i pretraživanje na drugom jest prevođenje termina uz pomoć višejezičnih rječnika, što povlači nekoliko problema: pojam predstavljen jednom riječju u jednom jeziku nema uvijek ekvivalent od jedne riječi u drugom jeziku (isto vrijedi i za pojmove izražene s dvije ili više riječi); jednoznačna riječ u jednom jeziku može u drugom biti višeznačna pa su potrebni posebni postupci za odabir prikladnog značenja; te, postoje i riječi iz upita koje nemaju ekvivalenta u drugom jeziku ili u rječniku koji koristi određeni sustav (Vickery, Vickery 1993, str. 137). Jedno od predloženih rješenja prevođenja upita jest prevođenje oznaka iz općeg klasifikacijskog sustava poput UDK, koji pokriva sva područja znanja, na termine iz tezaurusa na više jezika (Francu 2002).

Za pitanje višeznačnosti vezan je problem pisama. Ovaj se problem u knjižničnim katalogima rješava najčešće transliteracijom. Međutim, ne koriste svi knjižnični softveri iste standarde za transliteraciju, niti su standardi za sve vrste pisama doneseni. Primjerice, Kongresna knjižnica koristi vlastiti standard (Library of Congress Romanization), koji neka slova ćirilice prikazuje kao dva znaka, dok standardi ISO koriste jedan, pa korisnik koji poznaje standard ISO neće dobiti rezultate jer ne zna da Kongresna knjižnica transliterira drukčije (Chroust 2001, str. 47). Iako danas postoji međunarodni Unicode standard za kodiranje znakova ISO/IEC 10646, mnogi zapisi i dokumenti i dalje se izrađuju u standardima za jedan jezik ili pismo, pa čak i u lokalnim inačicama (Borgman 2000, str. 219). Što se tiče

pretraživanja na izvornom pismu, Britanska knjižnica, na primjer, omogućuje takvo pretraživanje na terminalima putem zaslona osjetljivog na dodir (touch screen) na kojem su prikazane tipkovnice s traženim pismima (Butcher 1993, str. 169).

7.5. Pristupačnost sučelja WebPAC-a

Osobe koje zbog potrebe ili želje koriste tekstualne web-preglednike (browsers), samo će djelomično moći koristiti sučelja s grafikom. Softveri za sintezu govora još uvijek dobro čitaju jedino tekst, te slijepi i slabovidne osobe mogu koristiti samo tekstualne informacije. Problem predstavljaju sučelja na kojima je sama grafika (npr. ikona) osnovni nositelj informacije, kada je potrebno u njoj neposrednoj blizini uvrstiti tekst s objašnjenjem (Web Content Accessibility Guidelines 1.0). U obzir treba uzeti potrebu korisnika s motoričkim oštećenjima koji znatno lakše barataju sučeljima koja u manjoj mjeri zahtijevaju pomicanje sadržaja na zaslonu (scroll) te bi stoga idealno bilo kada bi sadržaj jedne "stranice" stao na zaslon (Edwards, Van Mele, Verheust, Spaepen 1997). S obzirom na navedeno, C. Long (2000, str. 171) daje tri smjernice za dostupnost WebPAC-a:

1. Potrebno je osigurati i tekstualno sučelje.
2. Potrebno je izbjegavati grafiku u funkciji pretraživanja.
3. Web-sučelje potrebno je oblikovati i za osobe s posebnim potrebama.

U obzir svakako valja uzeti i specifičan način informacijskog pretraživanja osoba s posebnom potrebama. Projekt NoVA (Non-Visual Access to the Digital Library) (Craven, Brophy 2003) podnio je opsežno izvješće o načinu na koji slijepi i slabovidne osobe pretražuju i snalaze se u digitalnom svijetu, koje može poslužiti kao temelj za smjernice za oblikovanje sučelja digitalnih knjižnica. Osnovni problem slijepim i slabovidnim osobama je taj što pomoćne tehnologije kojima se one služe podržavaju jedino tzv. linearno pregledavanje digitalnih izvora, tj. čitanje riječ po riječ pa je navigacija po kompleksno organiziranim mrežnim stranicama, primjerice s puno tzv. *frameova* (okvira), ovim osobama otežana. Ovakva

istraživanja potrebno je provesti i za ostale skupine osoba s posebnim potrebama i prirediti smjernice za izradu svima pristupačnih sučelja za pretraživanje.

7.6. Dizajn sučelja

Dizajn treba osigurati da mogućnosti za pretraživanje na zaslonu budu jasne i da navode korisnike na korištenje funkcija koje će olakšati njihovo pretraživanje. Dobri sustavi mogu biti nefunkcionalni ukoliko im je dizajn loš.

Ovo je područje kojem je potrebno pristupiti interdisciplinarno, uključujući saznanja iz semantike koja se bavi istraživanjem komunikacije, iz grafičkog dizajna i web-dizajna, ponašanja korisnika pri pretraživanju i interakcije s računalom, i drugoga.

IFLA-ina Sekcija za informacijsku tehnologiju (IFLA Section on Information Technology) 1995. pokrenula je projekt izrade standardnog skupa ikona za bibliografske informacijske sustave, u okviru kojeg je napravljen prijedlog za standardizaciju sličica na sučeljima bibliografskih baza podataka²¹, koji, međutim, nije naišao na veći odjek. Godine 1999. IFLA je izradila Smjernice za sučelje OPAC-a (Guidelines for OPAC displays), ali su one povučene te je u tijeku izrada novih.

U literaturi se, vezano uz dizajn sučelja online katalog, preporučuje sljedeće: okupljanje srodnih bibliografskih polja, pozicioniranje oznaka s desne strane tako da je oznaka u blizini početka polja u kojem su prikazani podaci, prikazivanje oznaka koje u sebi sadrže razmake (npr. signatura) sa tim razmacima, uporaba punih riječi za nazive polja, a ne kratica, korištenje laičkih naziva polja, a ne stručnih, te uporaba jasnih i preciznih naziva polja (Cherry 1998, str. 132).

Postoji niz online kataloga koji su se svojim posebnim multimedijским dizajnom nastojali približiti korisnicima (Onion Patch(sm) 2000). *IRCAM Multimedia Library*²² je multimedijška knjižnica Instituta za istraživanje akustike i glazbe (Institut de Recherche et de Coordination Acoustique/Musique - IRCAM), u kojoj je omogućena navigacija i pretraživanje korištenjem tlocrta knjižnice.

²¹ Bibliographic standard GUI icon set, URL: <http://www.scran.ac.uk/iconstd/>

²² URL: <http://mediatheque.ircam.fr/index-e.html>

Korisnici mogu prebirati po policama i pridruženim klasifikacijskim oznakama, pronaći kataložni zapis za odabrane jedinice građe, pa i samu jedinicu ukoliko je dostupna online. Ovaj katalog također omogućuje navigaciju i prebiranje u trodimenzionalnoj virtualnoj stvarnosti. U Informatičkoj knjižnici Sveučilišta u Oslu (Informatikbiblioteket Universitetet i Oslo) razvijao se projekt za osmišljavanje pretraživanja online kataloga korištenjem tlocrta knjižnice. Njihov specijalni klasifikacijski sustav Udruženja za računalnu tehnologiju (Association of Computing Machinery – ACM) može se prebirati pritiskom na određeni dio na tlocrtu koji u opisu sadrži mjesto djela u toj klasifikacijskoj shemi. Uz pojedine skupine u klasifikacijskoj shemi prikazuju se postojeći naslovi iz knjižnice. Slično sučelje razvijano je na Windsorskom sveučilištu (Univeristy of Windsor) u Kanadi, pod nazivom *VRML to Navigate Information Space*, u kojem se na temelju dobivenih rezultata prikazuje građa na policama, a temelji se na Klasifikaciji Kongresne knjižnice. *Speakable OPAC*, razvijan u Sveučilišnoj knjižnici Sjeverne Karoline (North Carolina State University Library), prototip je sustava za čitanje odabranih sadržaja online kataloga. Uz prikladni softver, *Speakable OPAC* prihvaća određene izgovorene naredbe i na glas čita autora, naslov, signaturu i status dostupnosti jedinice građe.

8. Online katalogi treće generacije - WebPAC-i

8.1. Okapi

Okapi (Online Keyword Access to Public Information) razvija se od 1982. godine na današnjem Sveučilištu Westminster (Westminster University), a od 1989. na Gradskom sveučilištu (City University) u Londonu. Obilježja Okapia po kojima se ubraja u kataloge treće generacije uključuju korisnički pristupačno grafičko sučelje, mogućnost postavljanja upita na prirodnom jeziku, ispravljanje pogreški u pisanju, uputnice za sinonime, korištenje tezaurusa u pretraživanju, automatsko preoblikovanje upita (automatic query expansion - AQE), interaktivno preoblikovanje upita (interactive query expansion - IQE) i rangiranje rezultata po relevantnosti.

Cilj Okapia bio je načiniti sustav koji bi bio jednostavan za korištenje nekom tko uopće ne poznaje online kataloge, ali da istodobno pruža mogućnosti potrebne iskusnim korisnicima. Razvijao se u nekoliko stadija.

U Okapiu '84 omogućen je unos upita na prirodnom jeziku. Primijenjena je tehnika najboljeg rezultata (best match), što je predstavljalo značajan napredak u odnosu na tadašnje sustave utemeljene na Booleovim operatorima, u kojima su se rezultati dobivali jedino ako su u potpunosti odgovarali upitu. Rezultati pretraživanja rangirani su zatim prema stupnju u kojem odgovaraju upitu.

U Okapiu '86 omogućeno je ispravljanje pogreški u pisanju i kraćenje, te su uvedene uputnice za sinonime, sa ciljem smanjivanja broja neuspjelih pretraživanja. Budući da je vrednovanje Okapia '84 pokazalo da korisnici žele više predmetnih pretraživanja, bibliografskim datotekama dodano je polje MARC-a 651 za predmetne odrednice s potpoljima za tematske, zemljopisne i formalne pododrednice.

Pogreške u pisanju provjeravale su se tako što su se riječi koje sustav nije mogao provjeriti uspoređivale s indeksom iz kojeg su korisniku ponuđene alternative kao zamjena. Kasnije istraživanje pokazalo je da korisnici ne vole svojstvo Okapia da automatski ispravlja pogreške u pisanju.

Indeksirane ključne riječi i riječi iz upita se krata, što povećava vjerojatnost pronalazjenja traženih jedinica građe. Korišteno je tzv. slabo i jako kraćenje. Slabo

kraćenje je uklanjanje uobičajenih sufiksa za plural i drugih gramatičkih sufiksa poput *-ing* i *-ed*. Jako kraćenje obuhvaća uklanjanje derivacijskih sufiksa poput *-ent*, *-ence*, *-ision*, za koje se pokazalo da vraća previše neočekivanih rezultata te da se treba izbjegavati.

Za sinonime je korištena tablica koja je sadržavala istoznačnice i bliskoznačnice, kratice, akronime, parove imenica/pridjev (npr. *Wales/Welsh*), nepravilne množine (npr. *wife, wives*), i alternativne načine pisanja (npr. *tsar, czar, csar*). Pokazalo se da tablica za sinonime nije ni posebno korisna ni štetna za pretraživanje, ali je korisniku pomogla da se ne mora sam dosjećati drugih sinonima. Međutim, uvijek postoje troškovi održavanja ovakvog sustava uputnica koje je potrebno razmotriti u odnosu na korisnost određenog obilježja.

U Okapiu '88 uvedena je tehnika preoblikovanja upita (query expansion), koja se temelji na korisnikovom sudu o relevantnosti pronađenih jedinica građe. Tehnika se sastojala od prikupljanja termina iz dobivenih zapisa koje korisnik smatra relevantnima, dodavanja tih termina postojećim terminima iz upita, te izračunavanja “težine” svakog termina. Nakon toga, katalog se pretraživao na temelju šesnaest najrelevantnijih termina. Ispis rezultata bio je rangiran, a temeljio se na tri vrste pridruženih vrijednosti ili “težina”. “Težina” svakog zapisa određivala se u odnosu prema “maksimalnoj težini”, tj. onoj koja se pridružuje zapisu sa svim riječima iz upita. Za zapise koji su imali dvije trećine od maksimalne “težine” smatralo se da dobro odgovaraju korisničkom upitu, a za one koji su imali pola od maksimalne “težine” smatralo se da manje odgovaraju. Zapisi koji su imali manje od jedne trećine maksimalne “težine” određeni su kao oni koji odgovaraju upitu, ali ne dovoljno.

Vrednovanjem se pokazalo da je vrlo mali postotak pretraživanja bio neuspješan i da je preoblikovanje upita učinkovito u pronalaženju specifičnijih predmeta ili u mijenjanju naglaska pretraživanja. Također se pokazalo učinkovitim u slučajevima kada prvo pretraživanje nije bilo uspješno. U sljedećem istraživanju kojim se detaljnije željela utvrditi korisnost preoblikovanja upita provedeno je 1990. Pokazalo se da je vrlo mali broj neuspješnih pretraživanja, svega 3,9% u odnosu na prijašnjih 40% u sustavu za potpuno preklapanje rezultata s upitom (best match). Otkrilo se da su korisnici preoblikovanje upita najčešće koristili kada rezultati pretraživanja prvoga upita nisu bili dobri ili kada je Okapi pronašao samo dio rezultata. Pokazalo se da je samo jedna trećina korisnika rabila preoblikovanje upita,

ali gotovo u pola pretraživanja u kojima je korišteno preoblikovanje upita, pronađen je još barem jedan zapis.

Idući projekt, Okapi '93, uveo je dvije različite vrste preoblikovanja upita, automatsko i interaktivno. Automatsko preoblikovanje upita (automatic query expansion - AQE) je ono u kojem sustav sam izdvoji riječi iz relevantnih zapisa te ih zajedno s riječima iz upita prikaže korisniku. Interaktivno preoblikovanje upita (interactive query expansion - IQE) je ono u kojem sustav izdvoji termine iz relevantnih jedinica građe i prikaže ih korisniku koji sam odabere željene termine. Ove tehnike vrednovane su u eksperimentima Okapi '94, u kojima se pokazalo da davanje korisnicima više nadzora nad uključivanjem novih riječi i izraza nije uvijek bilo korisno te da su samo rijetki odabrali interaktivno preoblikovanje upita.

Idući projekt pod nazivom *Enquire Okapi* s novim sučeljem, ispitivao je kako najbolje kombinirati AQE i IQE s traženjem po tezaurusu, a da se dobije maksimalno od korisnikovih procjena o relevantnosti za poboljšanje pretraživanja. Pokazalo se da je AQE učinkovitiji alat za pretraživanje nego prebiranje po zapisima u blizini relevantnog zapisa u klasificiranom slijedu, te, ponovno, da je to vrlo uspješna tehnika pretraživanja. Ova istraživanja u suprotnosti su s negativnim rezultatima istraživanja tehnike preoblikovanja upita o kojima se ranije pisalo (Hancock-Beaulieu, Walker 1992).

Također se pokazalo da i eksplicitno i implicitno korištenje tezaurusa može biti korisno pri oblikovanju upita. Iako sustav može pronaći korisne termine iz tezaurusa za preoblikovanje upita, najvredniji su termini koje korisnici sami odaberu.

Glavno obilježje sučelja *Enquire Okapi* jest da pretraživanje uključuje procjenjivanje relevantnosti i termina i jedinica građe, čime je poboljšana i stupanj interakcije (Hancock-Beaulieu 1997).

Tzv. *Okapi-Pack* je inačica Okapia koja se može dobiti od Centra za istraživanje interaktivnih sustava (Centre for Interactive Systems Research) na Gradskom sveučilištu u Londonu, za potrebe istraživanja.

8.2. Cheshire II

Online katalog Cheshire II razvijen je sredinom 1990-ih na Kalifornijskom sveučilištu (University of California) u Berkeleyu, sa ciljem da postane sljedeća ili treća generacija online kataloga korištenjem naprednih tehnika za predmetno pretraživanje iz područja informacijskog pretraživanja.

Cheshire II ima mogućnost unosa upita na prirodnom jeziku. Podržava nekoliko tehnika prevođenja termina iz korisničkog upita na rječnik korišten u bazi kataloga: stop-listu, pretvorbu i raspored riječi iz upita po određenim poljima iz zapisa, algoritme za kraćenje sufiksa i pretvaranje, npr., iz oblika u množini u oblik u jednini, te preslikavanje riječi iz upita u standardizirani oblik kakav postoji u rječniku i tezaurusu.

Dok se u online katalozima klasifikacijska oznaka općenito tretira kao samo još jedna ključna riječ u zapisu, u ovom projektu klasifikacijske su oznake ekstrahirane iz MARC zapisa da bi se povećao broj termina vezanih uz pojedini zapis, i dodani su termini analogni s uputnicama *use* ili *RT* iz tezaurusa, upućujući korisnika na kontrolirani rječnik pojmova prikazanih klasifikacijskim oznakama. Na taj način tzv. *class clusters* (skupine klasifikacijskih oznaka) funkcioniraju poput tezaurusa (O'Brien 1994, str. 227, 230).

Korisnikov se upit, dakle, uspoređuje s određenim klasifikacijskim skupinama (classification clusters) iz kataloga, tako da su skupine (clusters) prikazuju u opadajućem slijedu prema vjerojatoj relevantnosti za korisnikov upit. To korisniku olakšava odluku o predmetu koji ga zanima. Probabilistička metoda rangiranja temelji se na algoritmu za koji se pokazalo da postiže izvrsne rezultate u pretraživanju potpunog teksta sustava za informacijskog pretraživanja.

Sustav je uveo metodu rangiranja rezultata koja se temelji na dvije paralelne tražilice, čiji su rezultati implementirani u jednom algoritmu. Ovako korisnici mogu iskoristiti prednosti eksplicitnog pretraživanja Booleovim operatorima, najkorisnijeg za usko definirana pretraživanja, i probabilističkog pronalaženja, korisnog u slučajevima kada korisnici ne poznaju rječnik za označivanje korišten za opis traženih jedinica. Osim toga, nije dokazano da je jedan od ovih algoritama za pretraživanje bolji za sve vrste pretraživanja (Larson, McDonough, O'Leary, Kuntz, Moon 1996).

8.3. WebPAC-i danas

Sve veća prisutnost Mreže, popularnost grafičkog okruženja i opća prihvaćenost standarda Z39.50 doveli su do radikalnih promjena u sučelju kataloga za korisnike. Nadalje, sve veća brzina, snaga, sofisticiranost i pouzdanost računalne tehnologije omogućili su načine pretraživanja kakvi su ranije bili dostizni jedino u eksperimentalnim katalogima. Katalogi druge generacije (Hildreth 1984) uveli su poboljšanja poput većeg broja pristupnica, višestrukog načina pristupa pretraživanju i puno više informacija kada su zapisi pronađeni. Svojstva koja je C. Hildreth naveo kao napredna svojstva treće generacije online kataloga, poput rangiranja po relevantnosti, danas postaju dostupna. Osobit potencijal online kataloga imaju u poboljšanju pretraživanja i prebiranja. WebPAC-i zbog korištenja HTML-a imaju posebnu mogućnost ostvarivanja veza u katalogu koje olakšavaju pretraživanje i otvaraju razne mogućnosti za prebiranje. Sami korisnici ovih kataloga brojniji su i raznolikiji jer se katalogima može pristupiti preko Mreže.

Kao što je ranije prikazano, mrežno okruženje online kataloga omogućava implementaciju mnogih funkcija koje poboljšavaju predmetno pretraživanje, počevši od pristupačnijeg sučelja i jednostavnijeg korištenja uz pomoć izbornika. Za razliku od komandnih sučelja ili funkcijskih tipki u prvim online katalogima, danas korisnici ne moraju pamtiti pravila složene sintakse i sučelja su intuitivna. Osim korisnički pristupačnog sučelja, hiperveze pretraživanje čine posebno fleksibilnim, omogućavajući odabir termina iz kontroliranog popisa za prebiranje, iz dobivenih zapisa i sl. Zatim su tu mogućnosti zadržavanja strategija i rezultata prethodnih pretraživanja i kombiniranja pojedinih rezultata, te rangiranje rezultata po relevantnosti. Posebno su važne mogućnosti računalne obrade prirodnog jezika u predmetnom pretraživanju, poput kraćenja, utvrđivanja sinonima te interaktivno preoblikovanje upita.

Međutim, istraživanja pokazuju da su današnji WebPAC-i negdje između prve i druge generacije online kataloga i da su čak njihove *telnet* ili mjesno dostupne inačice sofisticiranije ili fleksibilnije (Ortiz-Repiso, Moscoso 1999) te da uglavnom nisu uvrstili prednosti mrežnog okruženja i mogućnosti koje postoje za predmetno pretraživanje (Long 2000, str. 178). Neka druga istraživanja pokazala su optimističniju sliku, poput britanskih visokoškolskih knjižnica (Babu, O'Brien 2000), gdje postoji navigacija po hipervezama, razrađen sustav uputnica koji može

biti automatiziran i mogućnost pretraživanja po ključnim riječima iz predmetnih odrednica ukoliko postoje u zapisu.

WebPAC-i se često u literaturi uspoređuju s web-tražilicama, pri čemu se naglašava jednostavnost korištenja web-tražilica i upućuje se da bi knjižnični katalozi trebali biti jednako intuitivni. Web-tražilice prihvaćaju upite na prirodnom jeziku, temelje se na višeznačnoj logici i rangiraju rezultate po vjerojatnoj relevantnosti. Također neke web-tražilice podržavaju preoblikovanje upita na temelju dodatnih termina. Npr. AltaVista korisnicima nakon unesenog upita ponudi popis srodnih termina po kojima mogu pretraživati. Google, danas najpopularnija web-tražilica, omogućava dobivanje rezultata bez potrebe za oblikovanjem upita korištenjem složene sintakse. Osim toga, Google oponaša način na koji ljudi vrednuju tiskane i druge informacije u realnom svijetu: prvo evaluira strukturu hiperveza na određenoj stranici (što ima više hiperveza na određenu stranicu, to je veća važnost stranice), a zatim se vrednuje sadržaj dane stranice putem sofisticiranog algoritma. Većina korisnika vječni su početnici i dok su na Internetu nemaju na raspolaganju informacijskog stručnjaka, već su prepušteni sami sebi. Zato je važno i kataloge tako oblikovati. Većina online kataloga složena je za korištenje i ne podržava napredne funkcije poput preoblikovanja upita i postavljanja upita na prirodnom jeziku (Borgman 1996).

C. Khoo (1999) opisuje E-Referencer, posredni sustav za pretraživanje online kataloga na Mreži. E-Referencer pretvara online kataloge koji se temelje na Booleovim operatorima u sustave za pretraživanje s mogućnostima kakve imaju web-tražilice. Obrađuje upite koje korisnik unese u prirodnom jeziku, preslikava ih na predmetne odrednice Kongresne knjižnice i upit šalje knjižničnom sustavu. Na temelju korisnikove povratne informacije o dobivenim rezultatima odabire se strategija za poboljšanje rezultata pretraživanja. E-Referencer temelji se na protokolu Z39.50. Baza znanja preslikava ključne riječi na termine LCSH, sadrži strategije pretraživanja koje se koriste za pretvaranje upita korisnika na prirodnom jeziku u primjeren upit s Booleovim operatorima, strategije za preoblikovanje upita radi poboljšanja rezultata na temelju prethodno dobivenih rezultata i povratne informacije od korisnika, strategije za sužavanje ili preoblikovanje upita, te pravila za odabiranje prikladne strategije za pretraživanje. Baza znanja koja preslikava ključne riječi u LCSH sadrži informacije o tome koje su predmetne odrednice Kongresne knjižnice povezane s pojedinom ključnom riječju. Svakoj predmetnoj

odrednici tada je pridružena vrijednost jednaka broju naslova pridruženih predmetnoj odrednici. Ove vrijednosti zatim se normiraju na temelju dijeljenja s najvišom vrijednošću (vrijednost za najčešću predmetnu odrednicu). Takve normirane vrijednosti odražavaju koliko je svaka predmetna odrednica povezana s ključnom riječi.

Softveri za knjižnične sustave su komercijalni. Neke od njih, koji podržavaju napredne funkcije pretraživanja, navodi E. Rasmussen (1999, str. 410-412): *Endeavor Information Systems* ima tražilicu koja podržava upite na prirodnom jeziku i rangiranje rezultata po relevantnosti (korisnik u upitu naznači koji su termini esencijalni, a koji važni); *Innovative Interfaces* podržava rangiranje rezultata po relevantnosti, što vrijedi i za pretraživanje po potpunom tekstu na temelju AltaVista-inog algoritma; *EOS International* također omogućava postavljanje upita na prirodnom jeziku, preoblikovanje upita na temelju rječnika, pretraživanje na temelju višeznačne logike i rangiranje rezultata po relevantnosti. Neki sustavi za online kataloge razvijani su, međutim, radi istraživanja i implementacije u akademskim knjižnicama, poput Okapia i Cheshire II, kojima je cilj bio razviti online kataloge treće generacije.

9. Istraživanje WebPAC-a britanskih i hrvatskih sveučilišnih knjižnica

9.1. Uvod u istraživanje

Predmetno pretraživanje online knjižničnih kataloga vrlo je važno znanstvenom i nastavnom osoblju te studentima koji vrše istraživanja na određenu temu. Zbog toga je istraživanje u okviru ovoga rada usmjereno prvenstveno na sveučilišne knjižnice. Korisnici sveučilišnih knjižnica različitih su znanja i iskustava, od dodiplomskih studenata s malim ili nikakvim iskustvom u pretraživanju informacijskih izvora, do znanstvenika s bogatim iskustvom, pa sučelja trebaju podržavati ove različitosti.

Istraživanjem se željela utvrditi raznolikost i kvaliteta predmetnog pretraživanja informacija u WebPAC-ima sveučilišnih knjižnica Velike Britanije i Hrvatske. Knjižnice Velike Britanije odabrane su iz razloga što su se tamo implementirali i ispitivali eksperimentalni katalogi poput Okapia s naprednim funkcijama treće generacije online kataloga. Osim toga, istraživanja visokoškolskih knjižničnih kataloga u Velikoj Britaniji pokazala su postojanje naprednih svojstava kataloga treće generacije (Babu, O'Brien 2000). Istraživanje je obuhvatilo mogućnosti prebiranja i pretraživanja po riječima, odnosno oznakama, prirodnog i kontroliranog jezika i mogućnosti oblikovanja složenih upita.

Osnovna hipoteza istraživanja jest da sveučilišne knjižnice Velike Britanije i Hrvatske djelomično podržavaju predmetno pretraživanje svojstveno katalogima treće generacije online kataloga.

9.2. Uzorak istraživanja

Uzorak se sastoji od četrdeset engleskih sveučilišnih knjižnica odabranih prema slučajnom uzorku, te četiriju hrvatskih sveučilišnih knjižnica koje imaju online katalog (vidi Prilog 1). Izbor engleskih knjižnica izvršen je na temelju popisa P. Scotta (2003). Nacionalna i sveučilišna knjižnica u Zagrebu te Gradska i sveučilišna

knjižnica u Osijeku knjižnice su s dvojnomo funkcijom, koje, međutim, nisu jasno razdvojene u smislu kataloga pa jesu obuhvaćene istraživanjem, zajedno sa Sveučilišnom knjižnicom Rijeka i Sveučilišnom knjižnicom u Splitu.

9.3. Metodologija istraživanja

Istraživanje se temeljilo na popisu provjernih točaka (check list), koji je većim dijelom pripremljen na osnovi smjernica C. Longa (2000, str. 164-165), sličnih istraživanja (Babu, O'Brien 2000; Cherry, Williamson, Jones-Simmons, Gu 1994) i prijedloga H. Olson i J. Bolla (2001, str. 193-196). Pitanja obuhvaćaju prebiranje i pretraživanje po ključnim riječima, pretraživanje po predmetnim odrednicama te oblikovanje složenih upita, a sastoje se od strukturiranih i nestrukturiranih pitanja. U okviru seminarskog rada na Poslijediplomskom studiju iz informacijskih znanosti, kolegij Indeksni jezici, autorica je provela slično istraživanje na uzorku od petnaest europskih sveučilišnih knjižnica koje je ovome poslužilo kao pilot istraživanje, na temelju kojeg je popis provjernih točaka promijenjen i prilagođen. Popis se sastojao od ukupno 49 pitanja podijeljenih u pet osnovnih skupina: općenita pitanja o sučelju, prebiranje, oblikovanje upita, sintaksa i rezultati pretraživanja (vidi Prilog 2).

Iako se većina provjernih točaka mogla istražiti pregledavanjem različitih sučelja, otvaranjem izbornika i jednostavnih upita, za neke je slučajeve bilo potrebno oblikovati složene upite. To se odnosi na kataloge u kojima nije bilo moguće bez postavljanja upita razlučiti o tome kako se određuje prednost pojedinih operatora te da li se koristi implicitni operator ili se riječi iz upita obrađuju kao izraz. U prvom slučaju postavljali su se sljedeći upiti: *London AND Glasgow OR Edinburgh* i *London AND (Glasgow OR Edinburgh)*, a za drugo pitanje *London Glasgow* i *London AND Glasgow*, nakon čega je uslijedila usporedba rezultata i donošenje zaključka.

Istraživanje je provedeno u razdoblju od 1. do 15. ožujka 2003. godine korištenjem preglednika Microsoft Internet Explorer 6.0.

9.4. Analiza rezultata

9.4.1. Općenita pitanja o sučelju

Što se tiče općenitih pitanja o sučelju za pretraživanje, dvadeset i četiri engleska kataloga imaju dva ili više sučelja s različitim mogućnostima pretraživanja, prilagođenih različitim znanjima korisnika (npr. osnovno sučelje i napredno sučelje ili jednostavno sučelje i složeno sučelje) (pitanje 1.). Isto tako hrvatski katalogi²³ imaju dva različita sučelja, jedno za jednostavno, a drugo za složeno pretraživanje. Ostali engleski katalogi imaju nekoliko zasebnih sučelja za pretraživanje po pojedinom polju (npr. sučelje za predmetne odrednice, sučelje za naslov) i po dva polja istodobno (npr. sučelje za autora i naslov, sučelje za naslov i ključnu riječ).

Na početnoj stranici za pretraživanje obje mogućnosti pretraživanja, i po ključnim riječima i po predmetnim odrednicama, ima dvadeset i četiri engleske knjižnice, a hrvatski katalogi na početnom sučelju imaju mogućnost pretraživanja samo po ključnim riječima (pitanje 2.).

9.4.2. Prebiranje

Mogućnost prebiranja po deskriptorima iz tezaurusa ne postoji niti u jednom katalogu obuhvaćenom istraživanjem (pitanja 3., 3.1., 3.2.).

Prebiranje po predmetnim odrednicama koje je određeno kao takvo postoji u petnaest engleskih kataloga, međutim, u okviru pretraživanja po riječima ili ključnim riječima iz predmetnih odrednica na nekom drugom sučelju (a koje se izričito ne određuje kao prebiranje), postoji još petnaest drugih kataloga, koji omogućuju prebiranje po popisu predmetnih odrednica nakon upisa riječi (pitanja 4., 4.1., 9.2.). U većini kataloga nije bilo moguće utvrditi o kojem se popisu predmetnih odrednica radi (može se samo nagađati da se radilo o vlastitom popisu), a od onih u kojima se moglo ustvrditi, u tri kataloga bili su to popisi LCSH, u jednom LCSH i

²³ S obzirom da svi online katalogi hrvatskih sveučilišnih knjižnica imaju isto sučelje i mogućnosti, dovoljno je koristiti sintagmu *hrvatski katalogi*.

MeSH, a u jednom katalogu LCSH, MeSH i jedan popis specifičan za zbirku (East Midlands Collection). Od kataloga koji omogućuju prebiranje popisa predmetnih odrednica preko sučelja za pretraživanje, jedan katalog koristi prijevode klasifikacijskih oznaka, četiri ih ima LCSH, jedan ima LCSH i MeSH, a jedan LCSH i Predmetne odrednice iz paleografije (Paleography Subject Headings). Hrvatski katalozi ne omogućuju prebiranje po predmetnim odrednicama.

Svi katalozi koji na neki način omogućavaju pregledavanje popisa predmetnih odrednica, hipervezom povezuju predmetne odrednice sa zapisima koji tu predmetnu odrednicu sadrže (pitanje 4.2.).

Prebiranje po oznakama klasifikacijskog sustava ima deset kataloga, međutim, kao što je to slučaj i s predmetnim odrednicama, postoje katalozi u kojima se dozvoljava prebiranje nakon upisa oznake preko sučelja za pretraživanje (dodatnih 16) (pitanje 5). Također u većini kataloga nije navedeno koji se klasifikacijski sustav koristi, a od onih koji su navedeni, postoje dva kataloga s Klasifikacijom Kongresne knjižnice, jedan s Deweyevom decimalnom klasifikacijom, jedan s Blissovom bibliografskom klasifikacijom, a jedan s pet različitih klasifikacijskih sustava (s Klasifikacijom Kongresne knjižnice i Klasifikacijom Medicinske knjižnice (NLM), s dvije vlastite klasifikacije i prijevodima klasifikacijskih oznaka). Od klasifikacijskih oznaka koje se mogu prebirati preko sučelja za pretraživanje jesu jedna knjižnica s Deweyevom decimalnom klasifikacijom, jedna s Deweyevom decimalnom klasifikacijom i Klasifikacijom Kongresne knjižnice, jedna s Deweyevom decimalnom klasifikacijom i vlastitom klasifikacijom, jedna s Deweyevom decimalnom klasifikacijom, Klasifikacijom Kongresne knjižnice, Blissovom bibliografskom klasifikacijom i vlastitom klasifikacijom, jedna s Klasifikacijom Kongresne knjižnice i vlastitom klasifikacijom, a jedna s Deweyevom decimalnom klasifikacijom i Klasifikacijom Medicinske knjižnice (pitanje 5.1.). U hrvatskim katalozima mogućnost prebiranja po klasifikacijskim oznakama ne postoji.

Od kataloga koji kroz jedno od sučelja omogućuju prebiranje po oznakama klasifikacijskog sustava, svi osim jednog imaju izravnu hipervezu na zapise označene tim klasifikacijskim oznakama (pitanje 5.2.).

Četiri engleska kataloga omogućuju prebiranje po signaturi, a još tri imaju mogućnost prebiranja popisa signatura nakon unosa upita u sučelju koje nije posebno predviđeno za prebiranje (pitanje 6). Za dva je popisa signatura bilo

moguće sa sigurnošću utvrditi da se temelje na klasifikacijskim sustavima, i to na istima (Klasifikacija Kongresne knjižnice, Deweyeva decimalna klasifikacija, Nacionalna medicinska klasifikacija i nekoliko drugih) (pitanje 6.1.). Svi ovi katalozi imaju izravnu vezu iz popisa signatura sa zapisima kojima su te signature pridružene (pitanje 6.2.). U hrvatskim sveučilišnim knjižnicama prebiranje po signaturi nije moguće.

Prebiranje po ostalim poljima preko sučelja za prebiranje ima devetnaest engleskih kataloga i svi nude barem dva dodatna polja (pitanje 7.). Polja najčešće ponuđena na izbor za prebiranje su sljedeća: autor (osamnaest kataloga), naslov (osamnaest kataloga), nakladnička cjelina (četnaest kataloga), naslov časopisa (osam kataloga), izdavač (tri kataloga), mjesto izdavanja (dva kataloga) ili sva navedena polja odjednom (sedam kataloga) (pitanje 7.1.). Osim ovih kataloga, postoje tri kataloga obuhvaćena istraživanjem u kojima se pri pretraživanju oblikovanjem upita nakon upisane riječi ponudi popis s hipervezama (pitanje 7.2.). Od ovih kataloga njih osamnaest ima izravno hiperveze na zapise koji sadrže odabrane termine. Hrvatski katalozi nemaju mogućnost prebiranja.

9.4.3. Oblikovanje upita

Pretraživanje po deskriptorima iz tezaurusa ne postoji ni u jednom od kataloga obuhvaćenih istraživanjem (pitanja 8., 8.1., 8.2.).

Pretraživanje po predmetnim odrednicama postoji u trideset i dva engleska kataloga (pitanje 9.). Za većinu nije bilo moguće sa sigurnošću utvrditi koji se popis predmetnih odrednica koristi, a u ostalim je slučajevima situacija sljedeća: u pet kataloga LCSH, u jednom LCSH i MeSH odjednom, u jednom LCSH i Predmetne odrednice iz paleografije, a u jednom LCSH i MeSH (pitanje 9.1.). Nakon upisivanja riječi iz predmetne odrednice u okvir za pretraživanje, od ova trideset i dva kataloga, njih osamnaest prikazuje dio popisa predmetnih odrednica u kojem se nalazi tražena riječ (pitanje 9.2.). U hrvatskim katalozima postoji pretraživanje po riječima iz (nekoliko vrsta) predmetnih odrednica, pri čemu se koristi vlastiti sustav predmetnih odrednica (Gradska i sveučilišna knjižnica Osijek izrađuje vlastite predmetne odrednice za svoju lokalnu bazu, Sveučilišna knjižnica Rijeka izrađuje

vlastite normirane predmetne odrednice za svoju lokalnu bazu, Sveučilišna knjižnica Split preuzima normirane predmetne odrednice koje izrađuje Nacionalna i sveučilišna knjižnica prema vlastitome pravilniku). Nakon upisivanja riječi, dobiva se popis naslova, a ne popis obližnjih predmetnih odrednica.

Mogućnost pretraživanja po klasifikacijskim oznakama postoji u dvadeset i tri engleska kataloga (pitanje 10.). Od sustava koji su se sa sigurnošću mogli utvrditi, svi koriste barem dvije vrste sustava. Jedan katalog koristi Deweyevu decimalnu klasifikaciju i Klasifikaciju nacionalne medicinske knjižnice, jedan katalog koristi Klasifikaciju Kongresne knjižnice i vlastitu klasifikaciju, jedan Blissov bibliografsku klasifikaciju, Deweyevu decimalnu klasifikaciju, Klasifikaciju Kongresne knjižnice i vlastitu klasifikaciju, jedan Blissov bibliografsku klasifikaciju, a jedan Deweyevu decimalnu klasifikaciju, Klasifikaciju Kongresne knjižnice i vlastitu klasifikaciju (pitanje 10.1.). Nakon upisivanja klasifikacijske oznake u okvir za pretraživanje, od ova dvadeset i tri kataloga njih sedamnaest prikazuje dio popisa klasifikacijskih oznaka u kojem se nalazi tražena oznaka (pitanje 10.2.). U hrvatskim katalozima pretraživanje po klasifikacijskim oznakama ne postoji.

Pretraživanje po signaturama omogućeno je u četiri engleska kataloga, a ne postoji u hrvatskima (pitanje 11.)

Pretraživanje po riječima ili ključnim riječima postoji u svim katalozima obuhvaćenima istraživanjem (pitanje 12.). Polja koja stoje na raspolaganju navedena su u trideset i devet engleskih kataloga, bilo da su polja ponuđena na izbor ili su u Pomoći navedena da se automatski pretražuju.

Kada postoji samo pretraživanje po ključnim riječima (osam kataloga), u uputama ili Pomoći navedeno je koja se polja pretražuju: naslov (osam kataloga), u jednom katalogu naslov i predmetne odrednice, a u jednom naslov, napomena o sadržaju, napomena o disertaciji i predmetne odrednice.

Polja koje stoje na raspolaganju da se po njima može pretraživati po pojedinim riječima ili ključnim riječima najčešća su: po naslovu (dvadeset i osam kataloga), po autoru (dvadeset i sedam kataloga), po predmetu koji obuhvaća sve predmetne odrednice (dvadeset i šest kataloga), po nakladničkoj cjelini (petnaest kataloga), po naslovu časopisa (devet kataloga), po izdavaču (osam kataloga), po napomenama (šest kataloga), te sva navedena polja istodobno (dvadeset kataloga) (pitanje 12.1.).

U hrvatskim katalozima korisnik može birati u sučelju za jednostavno pretraživanje birati ključne riječi, autora, naslov, riječi iz naslova i jezik, standardne brojeve te specifičan tzv. NSK-UNIN broj, a u sučelju za složeno pretraživanje riječ iz naslova, riječ iz bilo kojeg imena, riječ iz bilo koje predmetnice, riječ iz vlastitog imena, riječ iz imena korporativnog tijela, riječ iz obiteljskog imena, riječ iz jedinstvenog stvarnog naslova, riječ iz tematske odrednice te riječ iz zemljopisne odrednice. U sučelju za jednostavno pretraživanje nije objašnjeno što spada pod ključne riječi.

Mogućnost istodobnog, kombiniranog pretraživanja po dva ili više polja postoji u trideset i devet engleskih kataloga obuhvaćenih istraživanjem te u hrvatskim katalozima (pitanje 13.).

Prebiranje ili pretraživanje po integriranim kontroliranim rječnicima omogućeno je u četiri kataloga, od čega su u tri kataloga pri pretraživanju po predmetnicama uz njih prikazane i odgovarajuće klasifikacijske oznake, a u jednom katalogu se pri odabiru naredbe *all subjects* pod prebiranjem dobije popis koji obuhvaća i MeSH i LCSH. U hrvatskim katalozima ovo svojstvo ne postoji (pitanje 14.).

9.4.4. Sintaksa

Od zahvaćenog uzorka trideset i devet kataloga koristi operatore (pitanje 15.). Svih trideset i devet kataloga koristi operatore AND i OR, trideset koriste NOT uz još tri koja koriste AND NOT, sedam ih koristi i XOR (bilo koji od dva termina, ali ne smiju oba), tri ih koriste NEAR (riječi jedna do druge), dva koriste WITHIN N (N= maksimalan broj riječi koliko mogu biti udaljene), a jedan koristi i %N pri čemu je N maksimalan broj riječi koji može stajati između riječi povezanih ovim operatorom, i !N, prema kojem redosljed riječi mora biti isti kao u upitu (pitanje 15.1.).

Slijed izvođenja operacija u četrnaest kataloga određen je zagradama, u četrnaest ovisi o redosljedu unosa upita i odabira padajućih izbornika, sedam ih ne “razumije” zagrade i ima operator OR jači od operatora AND (pitanje 15.2.).

Hrvatski katalogi koriste Booleove operatore, pri čemu se razmak rabi umjesto operatora AND, + umjesto OR, a – umjesto NOT. Operatori se izvršavaju prema redoslijedu kojim su uneseni. Ako se riječi u upitu posebno ne povežu operatorima, sustav ih automatski obrađuje kao da je između njih AND (jer ionako razmak stoji umjesto AND). Mogućnost pretraživanja po izrazu ne postoji. Hrvatski katalogi podržavaju kraćenje, i to tako što se znak * rabi za desno kraćenje, a ? se primjenjuje u sredini nizova znakova (povezuje i riječi, ne samo unutar riječi).

Ukupno trideset i šest kataloga riječi unesene bez operatora automatski provodi kao da između njih postoji neki operator ili ih obrađuje kao izraz (pitanje 16.). U trideset i dva kataloga koristi se operator AND, pri čemu se u dva kataloga AND povezuje samo riječi u istom polju, a u jednom od njih operator AND zamjenjuje se operatorom OR ukoliko operatorom AND nisu pronađeni nikakvi rezultati. Četiri kataloga upit koji se sastoji samo od riječi automatski prevode na pretraživanje po izrazu (pitanje 16.1.).

Pretraživanje po izrazu podržavaju dvadeset i dva kataloga (pitanje 17). Od toga četiri kataloga imaju pretraživanje po izrazu tako što se ne unose nikakvi drugi operatori, u tri kataloga izraz se određuje navodnicima, a u preostalim petnaest kataloga postoji mogućnost odabira na sučelju za pretraživanje po izrazu (*exact, search for a phrase*) (pitanje 17.1.).

U trideset i sedam kataloga postoji mogućnost kraćenja, od čega se u jednom automatski krati sufiks, a u ostalima je kraćenje eksplicitno (pitanje 18.). Trideset i šest kataloga primjenjuje kraćenje sufiksa, pri čemu pojedini katalogi koriste znakove *, ? i \$. Sedamnaest ih omogućuje zamjenu znakova u sredini riječi, za što pojedini katalogi također koriste znakove *, ? i \$. Jedanaest ih podržava kraćenje početka riječi, a znakovi koje katalogi koriste su ? i * (pitanje 17.1.). Od ostalih preciznijih mogućnosti pojedini sustavi koriste * za sufikse od jedan do pet znakova, a ** za neograničen broj znakova (tri kataloga), a neki i sljedeće: # traži jedan ili dva znaka, a ! jedan znak. U hrvatskim katalogima koristi se * za desno kraćenje, a ? u sredini nizova znakova, koji povezuje i riječi (nije da vrijedi samo unutar riječi).

9.4.5. Rezultati pretraživanja

Svojstvo da dobiveni zapisi sadrže predmetne odrednice koje su hipervezom povezane s ostalim zapisima s istim predmetnim odrednicama posjeduje dvadeset i dva engleska kataloga. Riječi iz korisničkog upita posebno su istaknute u petnaest kataloga. Rezultate prethodnih pretraživanja prikazuje osamnaest kataloga, a tijekom prethodnih pretraživanja dvadeset i jedan. Hrvatski katalogi od ova četiri svojstva podržavaju povezivanje predmetnih odrednica hipervezom iz zapisa (pitanja 19., 20., 22., 23.).

Mogućnost preoblikovanja upita poput proširivanja ili sužavanja ili drukčijeg mijenjanja bez ponovnog upisivanja čitavog upita omogućeno je u dvadeset i tri kataloga, a ne postoji u hrvatskim katalogima (pitanje 21.).

Mogućnost ograničavanja upita podržava trideset i šest kataloga (pitanje 24.). Najčešći kriteriji za ograničavanje su sljedeći: godina izdanja (trideset i šest kataloga), vrsta građe (trideset i četiri kataloga), knjižnica ili lokacija (trideset i jedan katalog), jezik (dvadeset i četiri kataloga), vrsta publikacije (osam kataloga), izdavač (pet kataloga) i način posudbe (četiri kataloga). U hrvatskim katalogima podržano je ograničavanje po sljedećim kriterijima: jezik, bibliografska razina, vrsta građe, vrsta publikacije i godina (pitanje 24.1.).

Sortiranje rezultata podržava dvadeset i sedam engleskih knjižnica obuhvaćenih istraživanjem, od čega većinom po sljedećim kriterijima: dvadeset i sedam knjižnica omogućuju sortiranje po godini, dvadeset i četiri po autoru, dvadeset i tri po naslovu, pet po relevantnosti, četiri po vrsti građe (pitanja 25., 25.1.). Hrvatske knjižnice ne podržavaju sortiranje rezultata.

Tablica 1. Prikaz rezultata za strukturirana pitanja (za ostalo vidi tekst pod analizom rezultata). Gornji rezultat odnosi se na engleske, a donji na hrvatske kataloge.

Redni broj pitanja	Pitanje	Broj kataloga	%
1.	Postoje li dva ili više sučelja s različitim mogućnostima predmetnog pretraživanja (vrste pristupnica, način postavljanja upita) namijenjenih različitim znanjima i vještinama korisnika?	24/40	60
		4/4	100
2.	Postoje li na početnoj stranici za pretraživanje obje mogućnosti pretraživanja, i po ključnim riječima i po predmetnim odrednicama?	24/40	60
		0/4	0
3.	Postoji li mogućnost prebiranja po deskriptorima iz tezaurusa?	0/40	0
		0/4	0
4.	Postoji li mogućnost prebiranja po riječima iz popisa predmetnih odrednica?	30/40	75
		0/4	0
4.2.	Postoje li hiperveze na bibliografske zapise koji sadrže odabrane predmetne odrednice?	30/40	75
		0/4	0
5.	Postoji li mogućnost prebiranja po oznakama klasifikacijskog sustava?	26/40	65
		0/4	0
5.2.	Postoje li hiperveze na bibliografske zapise koji sadrže odabrane oznake?	25/40	62,5
		0/4	0
6.	Postoji li mogućnost prebiranja po signaturama?	7/40	17,5
		0/4	0
6.2.	Postoje li hiperveze na bibliografske zapise koji sadrže odabrane signature?	7/40	17,5
		0/4	0
7.	Postoji li mogućnost prebiranja po drugim poljima?	22/40	55
		0/4	0
7.2.	Postoje li hiperveze na bibliografske zapise koji sadrže odabrane termine?	21/40	52,5
		0/4	0
8.	Postoji li mogućnost pretraživanja po deskriptorima iz tezaurusa?	0/40	0
		0/4	0
9.	Postoji li mogućnost pretraživanja po riječima iz popisa predmetnih odrednica?	32/40	80
		4/4	100

Redni broj pitanja	Pitanje	Broj kataloga	%
9.2.	Prikazuje li se dio popisa predmetnih odrednica u kojem se nalaze riječi iz upita nakon izvršenja upita?	18/40	45
		0/4	0
10.	Postoji li mogućnost pretraživanja po oznakama klasifikacijskog sustava?	23/40	57,5
		0/4	0
10.2.	Prikazuje li se dio klasifikacijskog sustava u kojem se nalazi oznaka iz upita?	17/40	42,5
		0/4	0
11.	Postoji li mogućnost pretraživanja po signaturama?	4/40	10
		0/4	0
12.	Postoji li mogućnost pretraživanja po ključnim riječima?	40/40	100
		4/4	100
13.	Postoji li mogućnost istodobnog, kombiniranog pretraživanja po različitim poljima?	39/40	97,5
		4/4	100
14.	Postoji li mogućnost prebiranja ili pretraživanja po integriranim kontroliranim rječnicima ako se u katalogu koristi više od jednog kontroliranog rječnika?	4/40	10
		0/4	0
15.	Postoji li mogućnost korištenja operatora u oblikovanju upita?	39/40	97,5
		4/4	100
16.	Koristi li se implicitni operator ili sustav riječi iz upita razumijeva kao izraz?	36/40	90
		4/4	100
17.	Postoji li mogućnost pretraživanja po izrazu?	22/40	55
		0/4	0
18.	Postoji li mogućnost kraćenja?	37/40	92,5
		4/4	100
19.	Sadrže li dobiveni zapisi hiperveze s predmetnim odrednicama na ostale zapise s istim predmetnim odrednicama?	22/40	55
		4/4	100
20.	Jesu li riječi iz korisnikovog upita posebno istaknute u dobivenim zapisima?	15/40	37,5
		0/4	0
21.	Postoji li mogućnost proširivanja i sužavanja rezultata dodavanjem novih termina bez ponovnog upisivanja čitavog upita?	23/40	57,5
		0/4	0
22.	Prikazuje li sustav rezultate prethodnih pretraživanja?	18/40	45
		0/4	0
23.	Prikazuje li sustav tijek prethodnih pretraživanja?	21/40	52,5
		0/4	0
24.	Postoji li mogućnost ograničavanja rezultata?	36/40	90
		4/4	100
25.	Postoji li mogućnost sortiranja rezultata?	27/40	67,5
		0/4	0

9.5. Zaključak istraživanja

Osnovna hipoteza istraživanja da sveučilišne knjižnice Velike Britanije i Hrvatske djelomično podržavaju predmetno pretraživanje svojstveno katalogima treće generacije online kataloga potvrđena je.

Na temelju analize rezultata istraživanja dolazi se do zaključka da online katalogi engleskih sveučilišnih knjižnica podržavaju veći broj naprednih funkcija predmetnog pretraživanja, no što je to u hrvatskim online katalogima sveučilišnih knjižnica.

Što se tiče općih obilježja sučelja, postojanje različitih sučelja prilagođenih potrebama i vještinama različitih skupina korisnika hrvatske sveučilišne knjižnice imaju, kao i većina engleskih knjižnica. Smjernica za koju se zalaže C. Long (2000, str. 164) da na početnom sučelju postoje obje opcije za predmetno pretraživanje, i po ključnim riječima i predmetnim odrednicama, hrvatske knjižnice nemaju (postoji samo opcija za ključne riječi), a većina engleskih knjižnica ima obje mogućnosti na početnom sučelju.

Velika razlika između hrvatskih i engleskih knjižnica jest u mogućnostima prebiranja. Dok većina engleske knjižnica omogućava prebiranje po predmetnim odrednicama, klasifikacijskim oznakama i drugim poljima iz katalognog zapisa, hrvatske knjižnice ne podržavaju prebiranje niti po jednom polju. Iako hrvatske knjižnice omogućavaju pretraživanje po predmetnim odrednicama, nakon provođenja upita dobije se, kao i pri pretraživanju po ostalim poljima, odmah popis naslova, za razliku od engleskih kataloga, gdje se, u većini, nakon postavljanja upita dobiva dio popisa ili indeksa termina ili oznaka u kojem se nalazi riječ ili oznaka iz upita, hipervezom povezana na naslove koji ih sadrže. Manje od 25% engleskih knjižnica ima prebiranje po signaturi. Niti jedna hrvatska ni engleska knjižnica od kontroliranih rječnika ne podržavaju pretraživanje po deskriptorima iz tezaurusa.

Ono u čemu su hrvatski i engleski katalogi vrlo slični jest to što velika većina engleskih i svi hrvatski katalogi podržavaju pretraživanje po ključnim riječima i predmetnim odrednicama i mogućnost kombiniranog pretraživanja po različitim poljima.

Slično je i s uporabom operatora. Većina engleskih i svi hrvatski katalogi zahvaćeni istraživanjem podržavaju oblikovanje upita uz pomoć Booleovih operatora, a ukoliko se riječi ne povezuju eksplicitnim operatorima, sustav ih

povezuje implicitnim operatorom AND. Mogućnosti kraćenja postoje također u većini engleskih kataloga i u hrvatskima.

Što se tiče rezultata pretraživanja, većina engleskih i svi hrvatski katalogi zahvaćeni istraživanjem omogućavaju izravno povezivanje predmetnih odrednica iz dobivenog zapisa na ostale zapise s istim predmetnim odrednicama te ograničavanje rezultata po određenim kriterijima. Većina engleskih kataloga za razliku od hrvatskih omogućava nakon dobivenih rezultata preoblikovanje upita bez potrebe za ponovnim upisivanjem čitavog upita, prikazuje tijek prethodnih pretraživanja i omogućava sortiranje rezultata po određenim kriterijima. Također, za razliku od hrvatskih kataloga, 37,5% engleskih u dobivenim zapisima ima posebno istaknute riječi iz upita, a 45% prikazuje i rezultate prethodnih pretraživanja.

Zaključak

Namjera ovoga rada bila je osvrnuti se na problematiku predmetnog pretraživanja u knjižničnim katalogima s web-sučeljem, koje je, prema istraživanjima, najčešći i ujedno najneuspješniji način pretraživanja ovih kataloga. Nastojali su se izdvojiti elementi i metodologija oblikovanja predmetnog pretraživanja primjerenog korisnicima WebPAC-a u hrvatskoj akademskoj zajednici.

U prvome poglavlju raspravljalo se o predmetnom pretraživanju kao zadaći knjižničnog kataloga, o važnosti predmetnog pretraživanja te o pretpostavkama za predmetno pretraživanje u bazi podataka računalnog kataloga. Predmetno pretraživanje je, kao zadaću knjižničnog kataloga, prvi u 19. stoljeću odredio C. Cutter, a brojna istraživanja potvrđuju važnost ove zadaće, posebice u online okruženju, gdje je, pokazalo se, većina pretraživanja koje korisnici provede upravo predmetna. Predmetno pretraživanje kao zadaću kataloga potrebno je, kao i ostale zadaće kataloga, prilagoditi online okruženju kako bi se iskoristili svi potencijali računalnog okruženja. Ova problematika razrađena je u IFLA-inoj studiji Funkcionalni zadaci bibliografskih zapisa (FRBR), prema kojoj se katalog više ne promatra kao niz bibliografskih zapisa, već kao mreža povezanih informacija, a model opisan u ovoj studiji temelji se na načelima oblikovanja relacijskih baza podataka.

Baza podataka online kataloga danas se sastoji od nekoliko datoteka, od kojih su za predmetno pretraživanje posebno važne bibliografska datoteka, datoteka predmetnih preglednih kataložnih jedinica i uputnica, te invertirana datoteka. Sam format MARC, iako sadrži dovoljan broj polja za predmetni pristup informacijama (za termine i oznake kontroliranog rječnika, za napomene i za naslov), linearne je i krute strukture, prenesene iz papirnog okruženja, i kao takav sprječava razvoj strukture čvorova i hiperveza svojstvene mrežnom okruženju. Budućnost će pokazati hoće li se format MARC samo mijenjati i prilagođavati, ili će ga knjižnična zajednica potpuno napustiti i zamijeniti nekim fleksibilnijim formatom, poput XML-a.

Na razvoj predmetnog pretraživanja u online katalogima i prednosti online kataloga u odnosu na tradicionalne kataloge osvrnuli smo se u drugome poglavlju.

Dok se u katalogu na listićima okupljanje i pronalaženje jedinica građe na istu temu postiže u stvarnim katalogima, u online okruženju za predmetno pretraživanje postoji još nekoliko mogućnosti, poput većeg broja pristupnica (npr. ključnih riječi iz svakog pojedinog polja), kombiniranja ili koordinacije podataka iz različitih polja u trenutku pretraživanja raznim operatorima, korištenja tehnika kraćenja i sužavanja opsega pretraživanja. Iako su prednosti online kataloga nad katalogima na listićima mnogobrojne, ta su svojstva u današnjim katalogima samo djelomično iskorištena, što se posebno odnosi na interaktivnost, tehnike pretraživanja kakve se koriste u drugim sustavima za informacijsko pretraživanje, pa čak i na predviđena svojstva polja za predmetni pristup informacijama formata MARC koja se gotovo uopće ne koriste. Za razliku od kataloga na listićima, postoji veća mogućnost pogreški u pisanju pri unošenju upita, ali isto tako i pri oblikovanju zapisa, zbog čega je potrebno osigurati mogućnosti prebiranja i mehanizam prepoznavanja pogrešaka. Isto tako, korisnici u online okruženju nemaju na raspolaganju pomoć knjižničara pa upute trebaju biti sofisticirane i interaktivne.

Razvoj online kataloga u literaturi se dijeli na tri osnovna razdoblja, odnosno na tri generacije, pri čemu su katalogi treće generacije oni s naprednim svojstvima, kakva bi trebali imati katalogi s web-sučeljem. Međutim, i današnji su katalogi, pokazalo se, utemeljeni na katalogima na listićima, odnosno nisu prilagođeni mrežnom okruženju, i ne uzimaju u obzir korisničko ponašanje pri pretraživanju informacija. Stoga se zaključuje da današnji katalogi više ili manje još uvijek pripadaju drugoj generaciji online kataloga.

U trećemu poglavlju problematizirala su se pitanja predmetnog pretraživanja identificirana u istraživanjima od 1983. godine do danas i njihova predložena rješenja. Zbog posebno prisutnog problema prevođenja upita na jezik sustava, odnosno pronalaženja prikladne predmetne odrednice ili drugog kontroliranog termina koji koristi sustav, predlaže se datoteka predmetnih preglednih kataložnih jedinica i uputnica, mogućnost pretraživanja po riječima iz predmetnih odrednica ili drugog abecednog sustava, mogućnost prebiranja, pomoć pri prevođenju riječi iz upita na jezik korišten u katalogu te pomoć i upute o preoblikovanju upita. Kao rješenje ovoga problema u eksperimentalnim katalogima Okapi i Cheshire predloženo je analiziranje korisničkog upita na prirodnom jeziku po leksičkim, sintaktičkim i semantičkim svojstvima, kao i eksploatiranje statističkih odnosa

između riječi iz naslova i predmetnih odrednica Kongresne knjižnice (Library of Congress Subject Headings). Što se tiče obogaćivanja zapisa predmetnim pristupnicama, pokazalo se da su, osim povećanja broja predmetnih odrednica radi povećanja specifičnosti u označivanju i pretraživanju, poželjne i pristupnice poput stranica sa sadržajem, tekstova s knjižnih ovitaka, predgovora i pogovora. Također je prebiranje i pretraživanje potrebno omogućiti i po nazivima ili opisima klasifikacijskih oznaka. Uz signature bi trebalo navesti i popis knjiga na policama radi mogućnosti prebiranja. Zbog čestih pogreški u pisanju i pravopisnih pogrešaka predlaže se uvođenje mogućnosti prebiranja po raznim poljima i mehanizam ispravljanja pogreški. Osim toga, pokazalo se da korisnici ne razumiju sučelje kataloga, od samih uputa o pretraživanju, korištenja operatora, do razlike između pojedinih datoteka i polja. Stoga je potrebno poboljšati izgled samoga sučelja, primjerice, na osnovnom sučelju pojednostaviti terminologiju, npr. koristiti *naslov* umjesto *ključna riječ iz naslova* te osigurati jasne i interaktivne upute.

Osnovni zaključak četvrtog poglavlja, u kojem se raspravlja o načinima na koje korisnici provode predmetno pretraživanje online kataloga, jest da online katalozi nisu oblikovani na znanjima o ponašanju korisnika pri predmetnom pretraživanju, već su zapravo preslikani model kataloga na listićima. C. Borgman razlikuje tri razine znanja potrebnih za pretraživanje online kataloga: konceptualno (prevođenje informacijske potrebe na upit za pretraživanje), semantičko (odluka o odabiru mogućnosti koje sustav nudi) i tehničko (poznavanje sintakse oblikovanja upita i temeljne računalne vještine). Današnji sustavi nedovoljno podržavaju ova znanja, a posebno se malo pažnje posvećuje oblikovanju uputa za konceptualno znanje. Osnovni problem predmetnog pretraživanja je taj što je za postavljanje upita potrebno opisati informaciju koja se ne posjeduje. Stoga je potrebno omogućiti mehanizme preoblikovanja upita i razne alate koji će korisniku pomoći da osvijesti i izrazi svoju informacijsku potrebu. U online katalozima također se ne uzima u obzir činjenica da korisnici upit oblikuju postepeno, već dizajnom odražavaju stajalište da korisnici oblikuju upit koji je fiksni cilj pretraživanja i da je svaka sesija pretraživanja samostalna. Međutim, online katalozi trebali bi osigurati što je više moguće različitih strategija za pretraživanje. S obzirom na raznolikost znanja i potreba korisnika jednoga sustava, katalozi bi trebali omogućiti personalizaciju sučelja, prilagođenih svakom pojedinome korisniku.

U petom poglavlju opisane su različite vrste predmetnih pristupnica i predstavljena obilježja prirodnih i kontroliranih jezika za označivanje i pretraživanje te njihov utjecaj na predmetno pretraživanje. Osvrnuli smo se i na tehnike pretraživanja iz područja informacijskog pretraživanja (information retrieval), poput računalne obrade prirodnog jezika, koje se nedostavno primjenjuju u predmetnom pretraživanju WebPAC-a.

Računalno okruženje omogućava da svaka riječ iz svakog polja pojedinog zapisa bude pristupnica. Zbog toga se danas i gubi razlika između predmetnih i ostalih vrsta pristupnica, kao i razlika između opisnih podataka, za traženje korisniku poznatih jedinica, i predmetnih podataka, za traženje korisniku poznatih ili nepoznatih jedinica na danu temu.

Posebno korisne u predmetnom pretraživanju su ključne riječi iz kontroliranih termina, pri čemu za najbolje rezultate valja koristiti operatore i kraćenje. Što se tiče dodatnih pristupnica poput stranica sa sadržajem i sažecima, čija je izrada skupa i dugotrajna, preporuča se iskoristiti prednost umreženog okruženja i ove podatke preuzimati iz postojećih zapisa online knjižara ili nakladnika.

Razlika između popisa predmetnih odrednica i tezaurusa u računalnim katalozima gubi se iz dva osnovna razloga. Prvi je taj što se svakoj jedinici građe dodjeljuje veći broj predmetnih odrednica koje ne označuju više samo glavni predmet dokumenta, već i pojedine aspekte, za što su ranije služili isključivo deskriptori. Drugi razlog je taj što se popisi predmetnih odrednica i tezaursi stapaju u jedan kontrolirani sustav, pri čemu do izražaja dolaze prednosti obiju ovih vrsta sustava za označivanje i pretraživanje. Osim toga, nastoje se integrirati i klasifikacijski sustavi. Razvoj tehnologija poput XML-a pospešuje pristup različitim kontroliranim rječnicima putem jednog sučelja, što predstavlja jednostavnije rješenje u odnosu na izradu integriranog kontroliranog sustava. U svakom je slučaju integrirano korištenje različitih sustava, poput klasifikacijskih, abecednih i prirodnih, poželjno jer svaki od njih omogućava predmetni pristup informaciji s različitih aspekata. Upravo je najveća prednost online okruženja ta što se pretraživanje može provoditi istodobno po različitim sustavima, npr. po abecednom i klasifikacijskom sustavu te po ključnim riječima. Sve se više govori i o

razvoju neverbalnih sustava za pretraživanje i omogućavanje pristupa neverbalnim vrstama građe ili njihovoj integraciji u online kataloge.

Kataloge je potrebno oblikovati i tako da se korisnike uputi s riječi koju unese na odobreni termin ili oznaku u kontrolirani termin, bilo da se osigura datoteka predmetnih preglednih kataložnih jedinica i uputnica, tezaurus ili/i prebiranje po kontroliranim sustavima. Za povezivanjem i usklađivanjem različitih informacijskih sustava, koji koriste različite kontrolirane sustave za označivanje i pretraživanje, tezaursi su se pokazali najboljim alatom jer su u njima precizno određeni odnosi ekvivalentnosti, srodnosti i hijerarhije.

Tzv. operatori odnosa u kontroliranim sustavima poput sintaktičkih, klasifikacijskih i predmetnih sustava omogućuju postkoordinirano pretraživanje prekoordiniranih termina i oznaka, iako se ove mogućnosti danas u online katalogima rijetko koriste. Mijenja se podjela postkoordiniranih i prekoordiniranih sustava jer se prekoordinirani termini i oznake u računalnom okruženju mogu pretraživati po svojim sastavnim elementima.

Poseban se naglasak u predmetnom pretraživanju i oblikovanju sustava za označivanje stavlja na fasetnost zbog jednostavnog pristupa pojedinim fasetama u računalnom okruženju. Fasete mogu biti međusobno neovisne ili povezane preciznim sintaktičkim pravilima, dostupne pri predmetnom pretraživanju po bilo kojem redoslijedu, po hijerarhijskim ili abecednim kriterijima, po obliku, vremenskom razdoblju ili drugim obilježjima. Primjeri fasetnih sustava su Klasifikacija s dvotočkom, Blissova bibliografska klasifikacija, MeSH, PRECIS i Tezaurus za umjetnost i arhitekturu.

Klasifikacijski sustavi u online katalogima trebali bi biti pretraživi po nazivima oznaka, a ne po oznakama samima. Ukoliko se pretraživanje provodi po samim klasifikacijskim oznakama, valja osigurati dodatne mogućnosti pretraživanja, poput kraćenja oznake, prebiranja po hijerarhijskim strukturama i pretraživanje po svakom, a ne samo po prvom elementu klasifikacijske oznake, što je potrebno osigurati i pri oblikovanju oznaka. Ove mogućnosti danas su rijetko iskorištene, iako formati MARC to dopuštaju. Iako ova vrsta pretraživanja ne spada u ono što bi svakodnevni, usputni korisnik trebao, trebalo bi je omogućiti za potrebe iscrpnih i temeljitih predmetnih pretraživanja.

Osim samih sustava za označivanje i pretraživanje, na pretraživanje posebno utječe točnost, iscrpnost, specifičnost i konzistentnost označivanja. Konzistentnost

je obrnuto proporcionalna s iscrpnošću i specifičnošću jer visoka razina iscrpnosti ili visoka razina specifičnosti dovode do niske razine konzistentnosti.

U posljednje vrijeme istraživanja sustava za označivanje i pretraživanje bave se sučeljem, poboljšanjem pretraživanja na temelju prirodnog jezika, ili hipertekstom, dok se nedovoljno raspravlja o učinkovitosti termina za prikaz pojmova iz dokumenata ili samih dokumenata. Takva su istraživanja potrebna jer je u doba kada je nastala većina kontroliranih sustava koja se danas koristi okruženje uglavnom bilo papirno.

Uporaba prirodnog jezika u predmetnom pretraživanju povećava odziv, ali smanjuje preciznost, dok je s kontroliranim jezicima obrnuto. Metode i tehnike pretraživanja poput kraćenja, Booleovih i drugih operatora, rangiranja po relevantnosti i ekspertnih sustava, koje se koriste s nekontroliranim sustavima mogu djelomično zamijeniti kontrolirane sustave. Prijedlozi da se računalna obrada prirodnog jezika iskoristi kao metoda za unapređenje uspješnog predmetnog pretraživanja temelje se na činjenici da riječi u korisničkom upitu pripadaju jeziku te da kao takve imaju određena morfološka ili sintaktička obilježja, na temelju kojih se mogu utvrditi i pronalaziti semantičke informacije. Stoga postoje tri osnovne razine računalne obrade prirodnog jezika, leksička, sintaktička i semantička. No, istraživači iz područja računalne obrade prirodnog jezika danas se više ili manje bave prvim dvjema razinama i još su daleko od semantičke razine. Ipak, određene tehnike iz ovoga područja, poput kraćenja, ispravljanja pogrešaka ili postavljanja upita na prirodnom jeziku, već se primjenjuju u WebPAC-ima.

Šesto poglavlje obrađuje interakciju korisnika s katalogom, pri čemu je naglasak na sintaksi postavljanja upita, što obuhvaća razne operatore, oblike kraćenja, oblikovanje upita, prebiranje i rangiranje rezultata. Uobičajeni postupci rješavanja problema prirodnog jezika, tj. raznih morfoloških oblika riječi, sinonimičnosti i homonimičnosti u jeziku pri predmetnom pretraživanju jesu kraćenje i operatori. Različiti operatori koriste se ovisno o vrsti upita i o tome je li naglasak na odzivu ili preciznosti. Najprecizniji je operator blizine i njegove inačice, a najveći odziv dobiva se operatorom OR. Operatori u online katalogima mogu biti eksplicitni i implicitni (najčešće AND i ADJ). Pokazalo se da korisnici imaju velikih problema s razumijevanjem operatora i nepoznavanjem sintakse oblikovanja upita.

Potrebno je oblikovati intuitivne sustave koji će zahtijevati minimalnu obuku korisnika.

Prije su postojali samo sustavi čije je korištenje zahtijevalo precizan unos riječi u upitu, jer su vraćali samo one rezultate koji u potpunosti odgovaraju upitu. Druga vrsta sustava koja se danas koristi jesu probabilistički sustavi, u kojima je poželjno da dokumenti dobiveni pretraživanjem budu rangirani. Rangiranje prema procijenjenoj sličnosti s upitom omogućava korisniku da sam odluči u kojem trenutku prestati s pregledavanjem rezultata, jer nema potrebe da ih pregledava sve. Rangiranje rezultata olakšava ponavljanje pretraživanja na temelju povratne informacije od korisnika, prema onome što korisnik odabere kao relevantne dokumente, te korisniku omogućava da promijeni dodavanjem ili izbacivanjem termina iz upita.

U sedmom poglavlju prikazana su pitanja vezana uz oblikovanje sučelja online knjižničkog kataloga za predmetno pretraživanje, u koja se ubrajaju pitanja oblikovanja sučelja za različite korisnike, pa tako i pristupačnosti i višejezičnosti, te pitanja dizajna i osiguravanja pomoći i uputa.

Korisnik mora imati nadzor nad postupcima koje sustav provodi, što se posebno odnosi na napredne funkcije WebPAC-a poput algoritma za prevođenje riječi iz upita na termine kontroliranog sustava, automatskog kraćenja ili implicitnih operatora. Za ispravljanje pogrešaka i ujednačavanje pravopisnih varijanti najbolje je osigurati softvere, i to tako da se temelje na standardima. Treba postojati mogućnost pregleda provedenih pretraživanja, kao i mogućnost ponavljanja istih upita na drugom sustavu. Budući da korisnici ne razumiju razliku između kontroliranog sustava i ključnih riječi, potrebno je osigurati detaljne upute uz svaku opciju za predmetno pretraživanje (npr. koja se polja pretražuju po ključnim riječima, pri čemu valja osigurati pretraživanje po elementima kontroliranih termina), kako bi korisnici mogli odabrati ono što im najbolje odgovara. Jedno je od najkorisnijih svojstava koja sučelja WebPAC-a mogu pružati jest dodavanje hiperveza u dobivenim zapisima na ostale zapise koji sadrže relevantne predmetne odrednice, kao i na deskriptore kada se unese nedeskriptor, te na šire, uže i srodne deskriptore. Posebno označivanje traženih pojmova (npr. bojom) u dobivenom zapisu može pomoći pri određivanju relevantnosti dobivene jedinice.

Računalno okruženje omogućava oblikovanje vrlo korisne interaktivne pomoći i tzv. *context-sensitive* uputa. Upute je potrebno osigurati kada provedeni upit nije dao rezultata, kao i o preoblikovanju upita općenito. Upute o predmetnom pretraživanju trebaju obuhvatiti i konceptualne aspekte pretraživanja.

Zbog velike raznolikosti korisnika u današnjem sve globalnijem informacijom okruženju, potrebno je osigurati sučelja koja će biti pristupačna svima, neovisno o fizičkim, kognitivnim, emocionalnim i kulturološkim različitostima, bilo da se oblikuju različita sučelja za različite razine znanja korisnika (npr. za osnovno i napredno pretraživanje), ili individualizirana sučelja za svakog korisnika. S obzirom na različita znanja korisnika, online katalogi više ili manje osiguravaju barem dva različita sučelja, jedno s izbornicima, a drugo komandno. Komandno sučelje danas koriste uglavnom knjižničari i informacijski stručnjaci. Često postoje dvije inačice sučelja s izbornicima, jedno za početnike, a drugo za napredne korisnike. U WebPAC-ima je najčešća ona inačica sučelja s izbornicima, na kojem postoje kućice za unos riječi (fill-in form). Potrebno je oblikovati sučelja koja će biti pristupačna osobama s posebnim potrebama, koje se pri korištenju računala i Interneta služe raznim pomoćnim tehnologijama s ograničenim mogućnostima pregledavanja digitalnih sadržaja.

Globalno informacijsko povezivanje i dostupnost kataloga na Mreži, osim raznolikosti korisnika, traži i osiguravanje višejezičnosti. Najjednostavnije je ostvariti višejezičnost sučelja, za razliku od prevođenja korisničkog upita na jezik baze koja se pretražuje, za što je potrebno da sučelje omogući obradu upita na više od jednoga jezika i pristup višejezičnim rječnicima. Uz višejezičnost veže se i problem transliteracije i različitih standarda koji se u tu svrhu danas koriste u svijetu.

Dizajn treba osigurati da mogućnosti za pretraživanje na zaslonu budu jasne i da navode korisnike na korištenje funkcija koje će olakšati njihovo pretraživanje. Dobri sustavi mogu biti nefunkcionalni ukoliko im je dizajn loš. U literaturi se, između ostalog, posebno preporučuje okupljanje srodnih bibliografskih polja, uporaba punih riječi za nazive polja (a ne kratica) te uporaba laičkih naziva polja (a ne stručnih).

U osmom poglavlju obrađeni su WebPAC-i, s naglaskom na eksperimentalne WebPAC-e Okapi i Cheshire II, koji su implementirali

najnaprednije funkcije predmetnog pretraživanja. Okapi (Online Keyword Access to Public Information) se razvija od 1982. godine u Londonu, a obilježja po kojima se ubraja u kataloge treće generacije jesu korisnički pristupačno grafičko sučelje, mogućnost postavljanja upita na prirodnom jeziku, ispravljanje pogreški u pisanju, uputnice za sinonime, korištenje tezaurusa u pretraživanju, automatsko preoblikovanje upita (automatic query expansion - AQE), interaktivno preoblikovanje upita (interactive query expansion - IQE) i rangiranje rezultata po relevantnosti. Cheshire II podržava nekoliko tehnika prevođenja termina iz korisničkog upita na rječnik korišten u bazi podataka, koristi probabilističke tehnike pronalaženja najboljeg rezultata koje se temelje na algoritmu za koji se pokazalo da postiže izvrsne rezultate u pretraživanju potpunog teksta sustava za informacijskog pretraživanja.

Mrežno okruženje online katalogizma omogućava implementaciju mnogih funkcija koje poboljšavaju predmetno pretraživanje, počevši od pristupačnijeg sučelja i jednostavnijeg korištenja uz pomoć izbornika. Osim korisnički pristupačnog sučelja, hiperveze pretraživanje čine posebno fleksibilnim, omogućavajući odabir termina iz kontroliranog popisa za prebiranje, iz dobivenih zapisa i sl. Zatim su tu mogućnosti zadržavanja strategija i rezultata pretraživanja i kombiniranja pojedinih rezultata, te rangiranje rezultata po relevantnosti. Posebno su važne mogućnosti računalne obrade prirodnog jezika u predmetnom pretraživanju, poput kraćenja i utvrđivanja sinonima, te interaktivno preoblikovanje upita.

Međutim, istraživanja pokazuju da su današnji WebPAC-i negdje između prve i druge generacije online kataloga te da više ili manje nisu uvrstili prednosti mrežnog okruženja i mogućnosti koje postoje za predmetno pretraživanje.

WebPAC-i se često u literaturi uspoređuju s web-tražilicama, pri čemu se naglašava jednostavnost korištenja web-tražilica i upućuje se da bi knjižnični katalogi trebali biti jednako intuitivni. Web-tražilice prihvaćaju upite na prirodnom jeziku, temelje se na višeznačnoj logici i rangiraju rezultate po vjerojatnoj relevantnosti. Također neke web-tražilice podržavaju preoblikovanje upita na temelju dodatnih termina.

U devetom poglavlju opisano je istraživanje svojstava predmetnog pretraživanja u WebPAC-ima britanskih i hrvatskih sveučilišnih knjižnica. Pokazalo

se da, dok i hrvatski i britanski katalozi podržavaju određena svojstva kataloga treće generacije, britanski katalozi imaju više takvih obilježja. Istraživanjem se nastojalo ukazati na nedostatke postojećih WebPAC-a u Hrvatskoj i podastrijeti smjernice oblikovanja obilježja predmetnog pretraživanja WebPAC-a za potrebe hrvatske akademske zajednice.

Bibliografija

1. Aitchinson, T. M.; Gilchrist, A. Thesaurus construction. 2nd ed. London : Aslib, 1987. Citirano prema: Rowley, J. The controlled versus natural indexing languages debate revisited : a perspective on information retrieval practice and research. // Journal of information science. 20, 2(1994), str. 108-119.
2. Aitchinson, J.; Gilchrist, A.; Bawden, D. Thesaurus construction and use : a practical manual. 4th ed. London : Aslib, 2000.
3. AltaVista.
<http://www.altavista.com> (13.02.2003)
4. ANSI/NISO Z39.58-1992 : Common command language for online interactive information retrieval.
5. Ayres, F. H. Time for change : a new approach to cataloguing concepts. // Cataloging and classification quarterly. 28, 2(1999), str. 3-16.
6. Ayres, F. H.; Nielsen, L. P. S.; Ridley, M. J. BOPAC2 : a new concept in OPAC design and bibliographic control. // Cataloging and classification quarterly. 28, 2(1999), str. 17-44.
7. Babu, B. R.; O'Brien, A. Web OPAC interfaces : an overview. // The electronic library. 18, 5(2000), str. 316-327.
8. Barbarić, A. Javno dostupni knjižnični računalni katalozi – nastanak i razvitak : magistarski rad. Zagreb, 2002.
9. Bates, M. J. Building controlled vocabulary. // 65th ASIS&T SIG/CR Classification Research Workshop, Philadelphia, 2002. Bilješke s predavanja.
10. Bates, M. J. The design of browsing and berrypicking techniques for the online search interface. // Online review. 13, 5(1989a), str. 407-424.
11. Bates, M. J. Rethinking subject cataloging in the online environment. // Library resources and technical services. 33, 4(1989b), str. 401-412.
12. Bawden, D. Tezaurusi : nova postignuća. // Vjesnik bibliotekara Hrvatske. 44, 1-4(2001), str. 182-187.
13. Belkin, N. J.; Oddy, R. N.; Brooks, H. M. ASK for information retrieval : part I : background and theory. // Journal of documentation, 38, 2(1982), str. 61-71.

14. Bibliographic standard GUI icon set. 2001.
<http://www.scran.ac.uk/iconstd/> (24.03.2003.)
15. BOPAC2.
<http://www.bopac2.comp.brad.ac.uk/~bopac2/> (12.01.2003.)
16. Borgman, C. L. All users of information retrieval systems are not created equal : an exploration into individual differences. // Information processing and management. 25, 3(1989), str. 237-251.
17. Borgman, C. L. From Gutenberg to the global information infrastructure : access to information in the networked world. Cambridge, Mass.; London : MIT Press, 2000.
18. Borgman, C. L. Why are online catalogs still hard to use? // Journal of the American Society for Information Science. 47(7), 1996, str. 493-503.
19. Borgman, C. L.; Case, D. O.; Meadow, C. T. The design and evaluation of a front end user interface for energy researchers. // Journal of the American Society for Information Science. 40, 1989, str. 86-98. Citirano prema:
Borgman, C. L. Why are online catalogs still hard to use? // Journal of the American Society for Information Science. 47(7), 1996, str. 493-503.
20. Borgman, C. L.; Hirsh, S. G.; Hiller, J. Rethinking online monitoring methods for information retrieval systems : from search product to search process. // Journal of the American Society for Information Science. 47, 1996, str. 568-583.
21. Broughton, V. Klasifikacija za 21. stoljeće : načela i struktura Blissove bibliografske klasifikacije. // Vjesnik bibliotekara Hrvatske. 44, 1-4(2001), str. 38-51.
22. BUBL information service.
<http://bubl.ac.uk/> (24.03.2003.)
23. Butcher, R. Multi-lingual OPAC developments in the British Library. // Program. 27, 2 (1993), str. 165-171.
24. Buzina, T. Zadaci bibliografskog zapisa prema modelu entitet-odnos. // Vjesnik bibliotekara Hrvatske. 43, 1-2(2000), str. 30-37.
25. Chan, L. M.; Lin, X.; Zeng, M. L. Structural and multilingual approaches to subject access on the Web. // IFLA journal. 26, 3(2000), str. 187-197.
26. Chaudiron, S.; Ihadjadene, M.; Role, F. Authorial index browsing in an XML digital library. // Proceedings of the Seventh International ISKO

- Conference, 10-13 July 2002, Granada, Spain / edited by Maria J. Lopez-Huertas. Str. 358-363.
27. Chen, Y.; Lin, S. C.; Chen, S. An application practice of the IFLA FRBR model : a metadata case study for the National Palace Museum in Taipei. // Proceedings of the 65th Annual Meeting of the American Society for Information Science and Technology, Philadelphia, 2002 / editor Elaine G. Toms. Str. 181-185.
 28. Cherry, J. M. Bibliographic displays in OPACs and Web catalogs : how well do they comply with display guidelines? // Information technology and libraries. September (1998), str. 124-137.
 29. Cherry, J. M.; Williamson, N. J.; Jones-Simmons, C. R.; Gu, X. OPACs in twelve Canadian academic libraries : an evaluation of functional capabilities and interface features. // Information technology and libraries. September (1994), str. 174-195.
 30. Cheshire II project home page.
<http://cheshire.lib.berkeley.edu/> (23.02.2003.)
 31. Chroust, D. Z. Slavic-language material in OCLC and the search for matching records : reconsidering and overlooked problem. // Slavic and East European information resources. 1, 4(2001), str. 41-67.
 32. Cochrane, P. The future of classification. // 65th ASIS&T SIG/CR Classification Research Workshop, Philadelphia, 2002. Bilješke s predavanja.
 33. Cousins, S. A. Enhancing subject access to OPACs : controlled vocabulary versus natural language. // Journal of documentation. 48, 3(1992), str. 291-309.
 34. Craven, J.; Brophy, P. Non-Visual Access to the Digital Library (NoVA) : the use of the digital library interfaces by blind and visually impaired people.
<http://www.mmu.ac.uk/h-ss/cerlim/projects/nova.html> (22.02.2003.)
 35. Crawford, W. Starting over : current issues in online catalog user interface design. // Information technology and libraries. 11, 1(1992), str. 62-76.
 36. Cunningham, S. J.; Witten, I. H.; Littin, J. Applications of machine learning in information retrieval. // Annual review of information science and technology. 34, 1999, str. 341-384.

37. Cutter, C. A. Rules for a dictionary catalog. 4th ed. London : Library Association, 1948. Citirano prema: Svenonius, E. The intellectual foundation of information organization. London ; Cambridge, Mass. : The MIT Press, 2001.
38. Dahlberg, I. Načela klasifikacijske strukture : istraživanja, iskustva, zaključci. // Vjesnik bibliotekara Hrvatske 44, 1-4(2001), str. 26-37.
39. Danskin, A. UNIMARC, ONIX and the future. // IFLA Council and General Conference, August 16-25, 2001.
<http://www.ifla.org/IV/ifla67/papers/008-188e.pdf> (11.10.2002.)
40. DESIRE : Development of a European Service for Information on Research and Education. 1999.
<http://www.lub.lu.se/desire> (22.02.2003.)
41. Drabenstott, K. M.; Weller, M. S. Failure analysis of subject searches in a test of a new design for subject access to online catalogs. // Journal of the American Society for Information Science. 47, 7(1996a), str. 519-537.
42. Drabenstott, K. M.; Weller, M. S. The exact display approach for online catalog subject searching. // Information processing and management. 32, 6(1996b), str. 719-745.
43. Dubois, C. P. R. Free text versus controlled vocabulary : a reassessment. // Online review. 11, 10(1987), str. 243-253. Citirano prema: Rowley, J. The controlled versus natural indexing languages debate revisited : a perspective on information retrieval practice and research. // Journal of information science. 20, 2(1994), str. 108-119.
44. Edwards, K.; Van Mele, I.; Verheust, M.; Spaepen, A. Evaluation of user interface design to optimize access to library databases for people who are motor-impaired. // Information technology and libraries. 16, 4(1997), str. 175-180. Citirano prema: Long, C. E. Improving subject searching in Web-based OPACs : evaluation of the problem and guidelines for design. // Internet searching and indexing : the subject approach / Alan R. Thomas, James R. Shearer, editors. New York ; London ; Oxford : Haworth Information Press, 2000. Str. 159-195.
45. Endeavor Information Systems.
<http://www.endinfosys.com> (23.02.2003.)

46. EOS International.
<http://www.eosintl.com> (23.02.2003.)
47. Fattahi, R. A comparison between the online catalog and the card catalog. // Library review. 44, 2(1995), str. 44-58.
48. Fidel, R. Writing abstracts for free-text searching. // Journal of documentation. 42, 1(1986), str. 11-21. Citirano prema: Hjørland, B.; Nielsen, L. K. Subject access points in electronic retrieval. // Annual review of information science and technology. 35, 2001, str. 249-298.
49. Foskett, A. C. The subject approach to information. 5th ed. London : Library Association Publishing, 1997.
50. Francu, V. Language-independent structures and multilingual information access. // Proceedings of the Seventh International ISKO Conference, 10-13 July 2002, Granada, Spain / edited by Maria J. Lopez-Huertas. Str. 404-411.
51. Functional requirements for bibliographic records : final report / IFLA Study Group on the Functional Requirements for Bibliographic Records. München : K. G. Saur, 1998.
52. Google.
<http://www.google.com> (20.02.2003)
53. Green, B. ONIX, a new product information standard. // IFLA Council and General Conference, August 16-25, 2001.
<http://www.ifla.org/IV/ifla67/papers/008-188e.pdf> (11.10.2002.)
54. Hancock-Beaulieu, M. A comparative transaction log analysis of browsing and search formulation in online catalogues. // Program. 27, 3(1993), str. 269-280.
55. Hancock-Beaulieu, M. Experiments on interfaces to support query expansion. // Journal of documentation. 53, 1(1997), str. 8-19.
56. Hancock-Beaulieu, M.; Fieldhouse, M.; Do, T. An evaluation of interactive query expansion in an online library catalogue with a graphical user interface. // Journal of documentation. 51, 3(1995), str. 225-243.
57. Hancock-Beaulieu, M. User friendliness and human-computer interaction in online library catalogs. // Program. 26, 1(1992), str. 29-37.
58. Hancock-Beaulieu, M.; Walker, S. An evaluation of automatic query expansion in an online library catalog. // Journal of documentation. 48, 4(1992), str. 406-421.

59. HighWire Press.
<http://www.highwire.org> (20.02.2003.)
60. Hildreth, C. R. OPAC research : laying the groundwork for future OPAC design. // The online catalogue : development and directions / C. R. Hildreth (ed.). London : The Library Association, 1989. Str. 1-24.
61. Hildreth, C. R. Pursuing the ideal : generations of online catalogs. // Online catalogs, online reference : converging trends : proceedings of a Library and Information Technology Association Preconference, June 23-24, 1983, Los Angeles / edited by Brian Aveney and Brett Butler. Chicago : American Library Association, 1984. Str. 31-56. Citirano prema: Rasmussen, E. M. Libraries and bibliographical systems. // Modern information retrieval / ed. Ricardo Baeza-Yates; Berthier Ribero-Neto. New York : ACM Press, 1999. Str. 397-414.
62. Hildreth, C. R. The GUI OPAC : approach with caution. // The public access computer systems review. 6, 5(1995).
<http://info.lib.uh.edu/pr/v6/n5/hild6n5.html> (12-01-2002).
63. Hildreth, C. R. The use and understanding of keyword searching in a university online catalog. // Information technology and libraries. June (1997), str. 52-62.
64. Hjørland, B.; Nielsen, L. K. Subject access points in electronic retrieval. // Annual review of information science and technology. 35, 2001, str. 249-298.
65. Hodges, P. R. Keyword in title indexes. // Special libraries. 74, 1(1983), str. 56-60. Citirano prema: Hjørland, B.; Nielsen, L. K. Subject access points in electronic retrieval. // Annual review of information science and technology. 35, 2001, str. 249-298.
66. Information retrieval. // Harrod's librarians' glossary and reference book : a directory of over 9600 terms. 9th ed. Aldershot ; Hants : Gower Publishing Company, 2000.
67. Innovative Interfaces.
<http://www.iii.com> (23.02.2003.)
68. IRCAM Multimedia Library.
<http://mediatheque.ircam.fr/index-e.html> (12.02.2003.)

69. ISO 5963-1985 : Documentation – methods for examining documents, determining their subjects, and selecting index terms.
70. Katić, T. Funkcionalni zadaci bibliografskih zapisa i UNIMARC format : strukturiranje strojno čitljivog zapisa za staru knjigu. // Mogućnosti suradnje u okruženje globalne informacijske infrastrukture : zbornik radova / 4. seminar Arhivi, knjižnice, muzeji ; uredile Mirna Willer i Tinka Katić. Zagreb : Hrvatsko knjižničarsko društvo, 2001. Str. 81-94.
71. Khoo, C. S. G. Transforming Boolean OPACs to Web search engines. // 65th IFLA Council and General Conference, Bangkok, August 20-28, 1999. <http://www.ifla.org/IV/ifla65/papers/010-143e.htm> (12.12.2002.)
72. Knjižnice grada Zagreba. <http://www.kgz.hr> (12.02.2003.)
73. Koch, T. Distributed digital library. Privatni razgovor. (16.11.2002.)
74. Koch, T. Quality-controlled subject gateways : definitions, typologies, empirical overview. 2000. <http://www.lub.lu.se/~traugott/OIR-SBIG.txt> (24.03.2003.)
75. Komlodi, A.; Soergel, D. Attorneys interacting with legal information systems : tools for mental model building and task integration. // Proceedings of the 65th Annual Meeting of the American Society for Information Science and Technology, Philadelphia, 2002 / editor Elaine G. Toms. Str. 152-163.
76. Lancaster, F. W. Vocabulary control for information retrieval. 2nd ed. Arlington : Information Resources Press, 1986.
77. Large, A.; Beheshti, J. OPACs : a research review. // Library and information science research. 19, 2(1997), str. 111-133.
78. Larson, R. L. Evaluation of advanced retrieval techniques in an experimental online catalog. // Journal of the American Society for Information Science. 43, 1(1992), str. 34-53.
79. Larson, R. R.; McDonough, J.; O'Leary, P.; Kuntz, L.; Moon, R. Cheshire II : design and evaluation of a next-generation online catalog system. 1996. http://sherlock.berkeley.edu/asis_paper/paper.html (21.02.2003)
80. Le Loarer, P. OPAC : opaque or open, public, accessible and co-operative? : some developments in natural language processing. // Program. 27, 3(1993), str. 251-268.

81. Lempel, R.; Moran, S. Introducing regulated bias into co-citation ranking schemes on the Web. // Proceedings of the 65th Annual Meeting of the American Society for Information Science and Technology, Philadelphia, 2002 / editor Elaine G. Toms. Str. 425-435.
82. Leščić, J.; Cvitaš, M. Sadržajna obrada dokumenata temeljem UDK-a u računalnom katalogu Knjižnice Hrvatske akademije znanosti i umjetnosti. // Vjesnik bibliotekara Hrvatske 44, 1-4(2001), str. 96-109.
83. Lewis, D. D. Learning in intelligent information retrieval. // Proceedings of the 8th International Workshop on Machine Learning, 1991 June, Evanston IL. San Mateo, Calif. : Morgan Kaufmann, 1991. Str. 235-239. Citirano prema: Cunningham, Sally Jo; Witten, Ian H.; Littin, James. Applications of machine learning in information retrieval. // Annual review of information science and technology. 34, 1999, str. 341-384.
84. Light, R. Online thesauri and information engineering. // UK museums and the Web, 2001.
<http://flood.nhm.ac.uk/mcg/talks/rl/> (06.03.2003.)
85. Lombardo, S. V.; Condic, K. Empowering users with a new online catalog. // Library hi tech. 18, 2(2000), str. 130-141.
86. Long, C. E. Improving subject searching in Web-based OPACs : evaluation of the problem and guidelines for design. // Internet searching and indexing : the subject approach / Alan R. Thomas, James R. Shearer, editors. New York ; London ; Oxford : Haworth Information Press, 2000. Str. 159-195.
87. Lopez-Pujalte, C.; Guerrero, V. P. B.; de Moya-Anegon, F. Evaluation of the applications of genetic algorithms to relevance feedback. // Proceedings of the Seventh International ISKO Conference, 10-13 July 2002, Granada, Spain / edited by Maria J. Lopez-Huertas. Str. 422-428.
88. MARC 21 concise format for classification data. 2002.
<http://www.loc.gov/marc/classification/eccdhome.html> (12.02.2003.)
89. Marchionini, G. Interfaces for end-user information seeking. // Journal of the American Society for Information Science. 43, 2(1992), str. 156-163.
90. Marchionini, G. Information seeking in electronic environments. 2nd ed. New York : Cambridge University Press, 1997. Citirano prema: Caro Castro, C.; Travieso, R. C. Ariadne's thread : knowledge structures for browsing in

- OPACs. // Proceedings of the Seventh International ISKO Conference, 10-13 July 2002, Granada, Spain / edited by Maria J. Lopez-Huertas. Str. 437-442.
91. Markey, K.; Atherton, P.; Newton, C. An analysis of controlled vocabulary and free text search statements in online searches. // Online review. 4, September (1980), str. 225-236. Citirano prema: Olson, H. A.; Boll, J. J. Subject analysis in online catalogs. Englewood : Libraries Unlimited, 2001.
92. Markey, K. Subject searching in library catalogs : before and after the introduction of online catalogs. Dublin, Ohio : OCLC, 1984. Citirano prema: Cherry, J. M. Bibliographic displays in OPACs and Web catalogs : how well do they comply with display guidelines? // Information technology and libraries. September (1998), str. 124-137.
93. Markey, K. Thus spake OPAC user. // Information technology and libraries. 2, 4 (1983), str. 381-87.
94. Matthews, J. R. Time for new OPAC initiatives : an overview of landmarks in the literature and introduction to WordFocus. // Library hi tech. 15, 1-2(1997), str. 111-122.
95. McIllwaine, I. C. Prilagodba UDK-a mrežnom svijetu. // Vjesnik bibliotekara Hrvatske 44, 1-4(2001), str. 52-67.
96. Mikačić, M. Teorijske osnove sustava za izradbu predmetnog kataloga : doktorska disertacija. Zagreb, 1990.
97. Mikačić, M. Teorijske osnove sustava za predmetno označivanje. Zagreb : Hrvatsko bibliotekarsko društvo, 1996.
98. Mitev, N. N.; Venner, G. M.; Walker, S. Designing an online public access catalogue : Okapi, a catalogue on a local area network. // Library and information report 39. London : The British Library, 1985.
99. NLM LocatorPlus help, tutorials and FAQs.
<http://locatorplus.gov/help.html> (15.03.2003.)
100. O'Brien, A. Online catalogs : enhancements and developments. // Annual review of information science and technology. 29, 1994, str. 219-241.
101. Okapi-Pack. 2001.
<http://www.soi.city.ac.uk/~andym/Okapi-PACK/index.html>
(23.02.2003.)

102. Olson, H. A.; Boll, J. J. Subject analysis in online catalogs. Englewood : Libraries Unlimited, 2001.
103. OMNI : the UK's gateway to high quality Internet resources in health and medicine.
<http://omni.ac.uk/> (24.03.2003.)
104. Onion Patch(sm) : New Age OPACs, innovative public access systems. 2000.
<http://www.public.iastate.edu/~CYBERSTACKS/Onion.htm> (21.02.2003.)
105. Online Public Access Catalog. // Harrod's librarians' glossary and reference book : a directory of over 9600 terms. 9th ed. Aldershot ; Hants: Gower Publishing Company, 2000.
106. Oritz-Repiso, V.; Moscoso, P. Web-based OPACs : between tradition and innovation. // Information technology and libraries. 18, 2(1999).
http://www.lita.org/ital/1802_moscoso.html (12.03.2002.)
107. Pavlinušić, D. NN : pretraživač za hrvatski jezik. 2002.
<http://nn.rot13.org/~dpavlin/nn.html>. (15.02.2003.)
108. Peters, T. A.; Kurth, M. Controlled and uncontrolled vocabulary searching in an academic library online catalog. // Information technology and libraries. 27, 1991, str. 201-211.
109. Pinto, M.; Lancaster, F. W. Abstracts and abstracting in knowledge discovery. // Library trends. 48, 1(1999), str. 234-248.
110. Priručnik za UNIMARC : bibliografski format. 2. hrv. izd. Zagreb : Nacionalna i sveučilišna knjižnica ; Hrvatsko knjižničarsko društvo, 1999.
111. Pu, H.-T. Discovery of user-oriented class associations for enriching library classification schemes. // Proceedings of the 65th Annual Meeting of the American Society for Information Science and Technology, Philadelphia, 2002 / editor Elaine G. Toms. Str. 194-205.
112. Rasmussen, E. M. Libraries and bibliographical systems. // Modern information retrieval / ed. Ricardo Baeza-Yates ; Berthier Ribero-Neto. New York : ACM Press, 1999. Str. 397-414.
113. Riesthuis, G. J. A.; Zumer, M. The functional requirements for bibliographic records and knowledge organization. // Proceedings of the Seventh International ISKO Conference, 10-13 July 2002, Granada, Spain / edited by Maria J. Lopez-Huertas. Str. 165-172.

114. Rowley, J. The controlled versus natural indexing languages debate revisited : a perspective on information retrieval practice and research. // Journal of information science. 20, 2(1994), str. 108-119.
115. Scott, P. LIBDEX : the library index.
<http://www.libdex.com/> (16-02-2003)
116. Seal, A. The development of online catalogs. // Introducing the online catalogue / edited by Alan Seal. Bath : Centre for Catalogue Research, Bath University Library, 1984. Str. 54-57. Citirano prema: O'Brien, A. Online catalogs : enhancements and developments. // Annual review of information science and technology. 29, 1994, str. 219-241.
117. Slavić, A. Automatsko predmetno označivanje : od računalno potpomognutog predmetnog označivanja do znalačkih sustava. // Predmetna obrada : ishodišta i smjernice : zbornik radova / uredile Jadranka Lasić-Lazić, Jelica Leščić, Jasna Petric. Zagreb : Hrvatsko knjižničarsko društvo, 1998. Str. 98-115.
118. Slavić, A.; Turkulin, B. Predgovor. // Vjesnik bibliotekara Hrvatske. 44, 1-4(2001). Str. xix-xxii.
119. Slone, D. J. Encounters with OPAC : on-line searching in public libraries. // Journal of the American Society for Information Science. 51, 8(2000), str. 757-773.
120. Smeaton, A. F. Information retrieval : still butting heads with natural language processing? 1997.
<http://www.math.utah.edu/pub/tex/bib/lncs1997a.ps.gz> (20.12.2002.)
121. SOSIG : social science information gateways.
<http://www.sosig.ac.uk/> (24.03.2003.)
122. Svenonius, E. The intellectual foundation of information organization. Cambridge, Mass. ; London : MIT Press, 2000.
123. Tenopir, C; Ro, J. S. Full text databases. New York : Greenwood Press, 1990. Citirano prema: Hjørland, B.; Nielsen, L. K. Subject access points in electronic retrieval. // Annual review of information science and technology. 35, 2001, str. 249-298.
124. Text Retrieval Conference (TREC).
<http://trec.nist.gov>. (23.01.2003.)
125. Thomson ISI.

- <http://www.isinet.com> (23.01.2003.)
126. Tversky, A.; Kahneman, D. Judgments under uncertainty : heuristics and biases. // Science. 185(1974), str. 1124-1131. Citirano prema: Borgman, C. L. Why are online catalogs still hard to use? // Journal of the American Society for Information Science. 47(7), 1996, str. 493-503.
127. Vickery, B.; Vickery, A. Online search interface design. // Journal of documentation. 49, 2 (1993), str. 103-187.
128. Vizine-Goetz, D. Online classification : implications for classifying and document[-like object] retrieval. // Knowledge organization and change : proceedings of the 4th International ISKO Conference, 15-18 July 1996, Washington, D.C., / Rebecca Green, ed. Frankfurt: INDEKS Verlag. Dostupno na: <http://staff.oclc.org/~vizine/isko/dvgisko.htm> (12.01.2003.)
129. Voorbij, H. J. Title keywords and subject descriptors : a comparison of subject search entries of books in the humanities and social sciences. // Journal of documentation. 54, 4(1998), str. 466-476. Citirano prema: Hjørland, B.; Nielsen, L. K. Subject access points in electronic retrieval. // Annual review of information science and technology. 35, 2001, str. 249-298.
130. Wasylenko, L. W. Oh say, can we see? See also? Subject referencing possibilities in OPACs : highlights of the ALCTS CCS Subject Analysis Committee program. // Library collections, acquisitions, and technical services. 24, 2000, str. 267-350.
131. Web Content Accessibility Guidelines 1.0 : W3C recommendation 5-May-1999.
<http://www.w3.org/TR/WCAG10/> (19.01.2002.).
132. Wilkes, A.; Nelson, A. Subject searching in two online catalogs : authority control vs. non-authority control. // Cataloging and classification quarterly. 20, 4(1995), str. 57-79.
133. Willer, M. Format i kataložna pravila : katalogiziranje elektroničke građe. // Vjesnik bibliotekara Hrvatske. 43, 1-2(2000), str. 9-22.
134. Yahoo!
<http://www.yahoo.com/> (20.01.2003.)

135. Yee, M. M. System design and cataloging meet the user : user interfaces to online public access catalogs. // Journal of the American Society for Information Science. 42, 2(1991), str. 78-89.
136. Zumer, M.; Zeng, L. Comparison and evaluation of OPAC end-user interfaces. // Cataloging and classification quarterly. 19, 2 (1994), str. 67-98.

Sažetak

Polazišta su ovome radu dosadašnja istraživanja, problemi i spoznaje koje upućuju na to da se korisnici knjižničnih kataloga najčešće oslanjaju na predmetna pretraživanja, odnosno ključne riječi, pojmove iz predmetnih sustava i tezaurusa, a da su takva pretraživanja, sa stajališta zadovoljstva korisnika dobivenim rezultatima, nedovoljno uspješna. Iako svojstva WebPAC-a (Web-based Online Public Access Catalogs – knjižnični katalozi s web-sučeljem), koji pripadaju tzv. trećoj generaciji online knjižničnih kataloga, obuhvaćaju velike mogućnosti pretraživanja, te su mogućnosti danas uglavnom neiskorištene pa se korisnici i nadalje, unatoč stalnom unaprjeđivanju sustava za predmetno pretraživanje, susreću s mnogim problemima, kao što su problemi vezani uz vještinu prevođenja zahtjeva odnosno misaonih koncepata pojedinca na jezik kataloga, sustave upućivanja koji se za njih priređuju i sl.

Predmetni pristup u online knjižničnim katalozima može se osigurati putem različitih pristupnica. U tu svrhu rabe se tzv. prirodni i umjetni jezici za označivanje, a svaki među njima ima određene prednosti i nedostatke. Označivanje prirodnim jezikom obavlja se računalom, tako da se iz određenih polja bibliografskog zapisa automatski preuzimaju sve riječi. Kontrolirani jezici za označivanje indeksni su jezici u kojima odabir i pridruživanje pojmova dokumentima nadzire čovjek. To mogu biti klasifikacijski sustavi, tezaursi, popisi predmetnih odrednica i drugi predmetni sustavi. Tijekom 1970-ih godina opće je prihvaćeno mišljenje da se najbolje pretraživanje postiže usporednim korištenjem jezika za označivanje. Osim jezika za pretraživanje, potrebno je poznavati ponašanje korisnika pri pretraživanju, a pri oblikovanju sučelja za pretraživanje uzeti u obzir znanje i vještine korisnika, osigurati upute, pomoć i povratnu informaciju na svakom koraku tijekom pretraživanja.

Cilj istraživanja bio je utvrditi raznolikost i kvalitetu predmetnog pristupa informacijama u WebPAC-ima britanskih sveučilišnih knjižnica, od mogućnosti pretraživanja po riječima, odnosno oznakama, prirodnog i kontroliranog jezika, mogućnosti prebiranja (browsing) do oblikovanja jednostavnih i složenih upita kako bi se zaključilo o dosadašnjim postignućima, ponuđenim modelima i korištenim metodama, te provela usporedna analiza na WebPAC-ima hrvatskih sveučilišnih knjižnica.

Ključne riječi: britanske sveučilišne knjižnice, hrvatske sveučilišne knjižnice, korisničko sučelje, online knjižnični katalozi, ponašanje korisnika pri pretraživanju, predmetne pristupnice, predmetno pretraživanje, sustavi za označivanje i pretraživanje, WebPAC-i.

Summary

Title: Subject retrieval in web-based library catalogs

This thesis has been motivated by past research, problems and realizations that online library catalog users frequently perform subject searches – using keywords, subject headings and descriptors – and these searches have yielded unsatisfactory results. Web-based catalogs or WebPACs (Web-based Online Public Access Catalogs), belonging to the so-called third generation of online catalogs and providing a wide variety of search options, remain largely underutilized despite the continuous advancement of information retrieval systems. Users still encounter a number of problems, such as those related to translating their concepts to the language of the catalog's system and cross-references prepared to this purpose.

Subject access in online library catalogs can be provided through different access points. To that purpose natural and controlled indexing and retrieval languages are used, and each among them has its advantages and downsides. Natural language indexing is performed by the computer, in which process words from defined fields are automatically extracted. Controlled indexing languages are those in which selection of terms to be assigned to documents is manually performed. These are, for example, classification systems, subject heading languages and thesauri. During the 1970s, a consensus was reached that the best retrieval results are gained when using both types of indexing languages simultaneously. Apart from indexing languages, it is necessary to take into account user search behavior; and while designing user interface one has to allow for the users' skills and knowledge - ensuring instruction, help and feedback information at every step of the retrieval process.

The aim of the research was to determine the variety and quality of subject access to information in WebPACs of British university libraries, including searching by words or classification marks, natural and controlled languages, browsing options, and forming simple and complex queries in order to conclude about existing advancements, offered models and employed methods and compare them to WebPACs of Croatian university libraries.

Keywords: British university libraries, Croatian university libraries, indexing and retrieval systems, online library catalogs, subject access points, subject retrieval, user interface, user search behavior, WebPACs

Životopis

Rođena sam 1975. godine u Čakovcu. U rodnom gradu završavam osnovnu i srednju školu. Godine 1994. upisujem studij engleskog jezika s književnošću i informatologije. Studij uspješno privodim kraju obranom diplomske radnje iz područja sintakse engleskoga jezika u srpnju 2000. godine.

Od 1999. do 2001. radim kao knjižničar – katalogizator u Nacionalnoj i sveučilišnoj knjižnici.

Godine 2000. upisujem poslijediplomski studij informacijskih znanosti – smjer bibliotekarstvo.

Od veljače 2001. godine radim kao znanstvena novakinja na Katedri za bibliotekarstvo Odsjeka za informacijske znanosti Filozofskog fakulteta u Zagrebu. U sklopu studija bibliotekarstva vodim *Vježbe iz Sustava za označivanje i pretraživanje*.

Prilog 1. Popis knjižnica i URL-a njihovih kataloga obuhvaćenih istraživanjem

a) engleske knjižnice i katalogi

1. Aston University Learning and Information Services
<http://library.aston.ac.uk/>
2. Bournemouth University Library
http://talis.bournemouth.ac.uk/www-bin/www_talis32
3. Brunel University Library
<http://library.brunel.ac.uk:8080/uhtbin/webcat>
4. Cambridge University Library
<http://newton.lib.cam.ac.uk:7703/>
5. Coventry University Library
<http://libwww.coventry.ac.uk:8000/>
6. Durham University Library
<http://library.dur.ac.uk/>
7. John Rylands University Library of Manchester
<http://rylibweb.man.ac.uk/catalogue/>
8. Keele University Information Services
<http://opac.keele.ac.uk/>
9. Kingston University Library
http://opac.kingston.ac.uk/www-bin/www_talis32
10. Lancaster University Library
<http://cat.lib.lancs.ac.uk/>
11. Leeds University Library
<http://lib.leeds.ac.uk/>
12. London Guildhall University Library
<http://library.lgu.ac.uk/search>
13. Loughborough University Library
<http://opac.lboro.ac.uk/>
14. Manchester Metropolitan University Library
<http://opac.library.mmu.ac.uk/talis>
15. Middlesex University Library and Information Services
<http://www.ilrs.mdx.ac.uk/lib/index.htm> (knjižnični URL)
16. Newcastle University Library
<http://fulltext.ncl.ac.uk:8000/>

17. Oxford University Libraries
<http://library.ox.ac.uk/>
18. Reading University Library
<http://vlsirsi.rdg.ac.uk/uhtbin/cgisirsi/9OocMm7O3x/0/57/49>
19. South Bank University Library
<http://136.148.1.170/webclient.html>
20. Staffordshire University Library and Learning Resources Services
<http://thomlib052.staffs.ac.uk/ipac-cgi/ipac.exe>
21. University College London Library Services
<http://library.ucl.ac.uk/>
22. University of Bath Library and Learning Centre
<http://www.bath.ac.uk/library/webcat/>
23. University of Bradford Learning Support Services
<http://www.brad.ac.uk/library/news/catalogue.php>
24. University of Bristol Information Services
<http://www.lib.bris.ac.uk/ALEPH>
25. University of East Anglia Library
<http://www.lib.uea.ac.uk/> (knjižnični URL)
26. University of Exeter Library and Information Service
<http://lib.ex.ac.uk/>
27. University of Hertfordshire Learning and Information Services
<http://voyager.herts.ac.uk/cgi-bin/Pwebrecon.cgi?DB=local&PAGE=First>
28. University of Huddersfield Computing and Library Services
<http://webcat.hud.ac.uk/>
29. University of Hull Library
<http://library.hull.ac.uk/>
30. University of Kent at Canterbury - Templeman Library
<http://opac.ukc.ac.uk/cgi-bin/Pwebrecon.cgi?DB=local&PAGE=First>
31. University of Leicester Library
<http://library.le.ac.uk/uhtbin/cgisirsi/F9AIXsRfGM/41930101/60/58/X>
32. University of Liverpool Library
<http://library.liv.ac.uk/>
33. University of London Library
<http://193.63.81.240/search~b1o1c1i1r1/>

34. University of Nottingham Library
<http://aleph.nottingham.ac.uk:4505/ALEPH>
35. University of Sheffield Library
http://library.shef.ac.uk/www-bin/www_talis32?
36. University of Southampton Library
<http://www-lib.soton.ac.uk/>
37. University of Sunderland Information Services
<http://webcatalogue.sunderland.ac.uk/>
38. University of Sussex Library
<http://catalogue.sussex.ac.uk/home>
39. University of Warwick Library
http://opac.warwick.ac.uk/www-bin/www_talis32
40. University of York Library
<http://libcat0.york.ac.uk/F>

b) hrvatske knjižnice i katalozi

1. Gradska i sveučilišna knjižnica Osijek
<http://baza.gskos.hr/anev/asearch.html>
2. Nacionalna i sveučilišna knjižnica
<http://www.nsk.hr/opac-crolist/uvod.html>
3. Sveučilišna knjižnica Rijeka
<http://161.53.149.10/>
4. Sveučilišna knjižnica u Splitu
<http://libar.svkst.hr/anev/asearch.html>

Prilog 2. Upitnik

I. Općenita pitanja o sučelju

1. Postoje li dva ili više sučelja s različitim mogućnostima predmetnog pretraživanja (vrste pristupnica, način postavljanja upita) namijenjenih različitim znanjima i vještinama korisnika?
2. Postoje li na početnoj stranici za pretraživanje obje mogućnosti pretraživanja, i po ključnim riječima i po predmetnim odrednicama?

II. Prebiranje

3. Postoji li mogućnost prebiranja po deskriptorima iz tezaurusa?
 - 3.1. kojeg tezaurusa?
 - 3.2. postoje li hiperveze na bibliografske zapise koji sadrže odabrane deskriptore?
4. Postoji li mogućnost prebiranja po riječima iz popisa predmetnih odrednica?
 - 4.1. kojeg sustava predmetnih odrednica?
 - 4.2. postoje li hiperveze na bibliografske zapise koji sadrže odabrane predmetne odrednice?
5. Postoji li mogućnost prebiranja po oznakama klasifikacijskog sustava?
 - 5.1. kojeg klasifikacijskog sustava?
 - 5.2. postoje li hiperveze na bibliografske zapise koji sadrže odabrane oznake?
6. Postoji li mogućnost prebiranja po signaturama?
 - 6.1. na kojem se sustavu temelje signature?
 - 6.2. postoje li hiperveze na bibliografske zapise koji sadrže odabrane signature?
7. Postoji li mogućnost prebiranja po drugim poljima?
 - 7.1. po kojim poljima?
 - 7.2. postoje li hiperveze na bibliografske zapise koji sadrže odabrane termine?

III. Oblikovanje upita

8. Postoji li mogućnost pretraživanja po deskriptorima iz tezaurusa?
 - 8.1. kojeg tezaurusa?
 - 8.2. prikazuje li se dio tezaurusa u kojem se nalazi deskriptor iz upita?
9. Postoji li mogućnost pretraživanja po riječima iz popisa predmetnih odrednica?
 - 9.1. kojeg sustava predmetnih odrednica?
 - 9.2. prikazuje li se dio popisa predmetnih odrednica u kojem se nalaze riječi iz upita nakon izvršenja upita?

10. Postoji li mogućnost pretraživanja po oznakama klasifikacijskog sustava?
 - 10.1. kojeg klasifikacijskog sustava?
 - 10.2. prikazuje li se dio klasifikacijskog sustava u kojem se nalazi oznaka iz upita?
11. Postoji li mogućnost pretraživanja po signaturama?
12. Postoji li mogućnost pretraživanja po ključnim riječima?
 - 12.1. koja su polja ponuđena na izbor?
13. Postoji li mogućnost istodobnog, kombiniranog pretraživanja po različitim poljima?
14. Postoji li mogućnost prebiranja ili pretraživanja po integriranim kontroliranim rječnicima ako se u katalogu koristi više od jednog kontroliranog rječnika?

IV. Sintaksa

15. Postoji li mogućnost korištenja operatora u oblikovanju upita?
 - 15.1. kojih operatora?
 - 15.2. o čemu ovisi slijed izvođenja operacija?
16. Koristi li se implicitni operator ili sustav riječi iz upita razumijeva kao izraz?
 - 16.1. koji operator ili izraz?
17. Postoji li mogućnost pretraživanja po izrazu?
 - 17.1. kako se određuje da se pretražuje po izrazu?
18. Postoji li mogućnost kraćenja?
 - 18.1. koje vrste?

V. Rezultati pretraživanja

19. Sadrže li dobiveni zapisi hiperveze s predmetnim odrednicama na ostale zapise s istim predmetnim odrednicama?
20. Jesu li riječi iz korisnikovog upita posebno istaknute u dobivenim zapisima?
21. Postoji li mogućnost proširivanja i sužavanja rezultata dodavanjem novih termina bez ponovnog upisivanja čitavog upita?
22. Prikazuje li sustav rezultate prethodnih pretraživanja?
23. Prikazuje li sustav tijekom prethodnih pretraživanja?
24. Postoji li mogućnost ograničavanja rezultata?
 - 24.1. po kojim kriterijima?
25. Postoji li mogućnost sortiranja rezultata?
 - 25.1. po kojim kriterijima?

Prilog 3. Primjeri naprednih svojstava predmetnog pretraživanja u WebPAC-ima

Primjer 1. Lančano kazalo u WebPAC-u.

Address <http://lib.ex.ac.uk/search/ara> Go

University of Exeter Library Catalogue

[Next Page](#)
[Brief Display](#)
[New Search](#)
[Another Search](#)
[Limit This Search](#)
[Search As Words](#)

CHAIN SUBJECT

CHAIN SUBJECTS (1-12 of 173)

Num	Mark	Year
1	Civil Arbitration -- See -- KN 385	
2	Civil Aviation Aden -- See -- 387.7095335	
3	Civil Aviation Arabian Gulf States -- See -- 387.709536	
4	Civil Aviation Bahrain -- See -- 387.7095365	
5	Civil Aviation North Yemen -- See -- 387.7095332	
6	Civil Aviation Oman -- See -- 387.709535	
7	Civil Aviation Qatar -- See -- 387.7095363	
8	Civil Aviation South Yemen -- See -- 387.7095335	
9	Civil Aviation United Arab Emirates -- See -- 387.7095362	
10	Civil Aviation Yemen Arab Republic -- See -- 387.7095332	
11	Civil Aviation Yemen People's Democratic Republic -- See -- 387.7095335	
12	Civil Code Iran -- See -- KT 1337	

[Next Page](#)
[Brief Display](#)
[New Search](#)
[Another Search](#)
[Limit This Search](#)
[Search As Words](#)

Primjer 2. Istodobno prebiranje po popisima predmetnih odrednica LCSH i MeSH.

Address <http://library.ox.ac.uk/>

Databases
Keyword Searching
Browse Searching

Patron Functions
Help
Logout

Browse in Oxford OLIS




Browse for in

su= [See] Kidner, E. Martin (Egerton Martin) (LCSH) thru su= Kidney Calculi (MeSH)

[Previous](#) [Next](#)

- ...[See] Kidner, E. Martin (Egerton Martin) (LCSH) (1)
- Kidner, Michael, 1917. -- Exhibitions (LCSH) (1)
- Kidner, Michael -- Exhibitions (LCSH) (1)
- Kidner, Pauline (LCSH) (3)
- Kidney (MeSH) (16)**
- Kidney (LCSH) (3)
- Kidney -- abnormalities (MeSH) (2)
- Kidney -- abnormalities -- congresses (MeSH) (1)
- Kidney -- anatomy & histology (MeSH) (4)
- Kidney -- anatomy & histology -- atlases (MeSH) (1)
- Kidney, Artificial (0)
- ...[See] Artificial kidney (LCSH) (17)
- Kidney, Artificial (MeSH) (8)
- Kidney, Artificial -- congresses (MeSH) (13)
- Kidney, Artificial -- popular works (MeSH) (1)
- Kidney bean (LCSH) (1)
- Kidney -- blood supply (MeSH) (4)
- Kidney -- blood supply -- congresses (MeSH) (2)
- Kidney calculi (0)
- ...[See] Kidneys -- Calculi (LCSH) (4)
- Kidney Calculi (MeSH) (5)

Primjer 3. Pronalaženje riječi iz naslova korištenjem automatskog kraćenja.

Address  <http://library.ox.ac.uk/>   Go

Databases

Keyword Searching

Browse Searching

Patron Functions



Help

Logout

Browse in Both Oxford OLIS and Bodleian Pre-1920 Catalogues

Browse for in

ti= Les Croates et la civilisation du livre : actes du Ier Symposium International d'études sur l'aire culturelle croate : tenu à l'occasion du 500e anniversaire de la parution du premier livre croate imprimé en caractères glagolitiques (Paris, 3 décembre 1983) thru ti= Croatia at the United Nations : compilation of statements issued by the Permanent Mission of Croatia to the United Nations, October 21, 1993 - January 16, 1998

 **Previous**  **Next**

Les Croates et la civilisation du livre : actes du Ier Symposium International d'études sur l'aire culturelle croate : tenu à l'occasion du 500e anniversaire de la parution du premier livre croate imprimé en caractères glagolitiques (Paris, 3 décembre 1983) (1)

Les Croates et les Slovènes ont été les amis de l'Entente pendant la guerre : quelques documents officiels tirés des archives militaires d'Autriche-Hongrie, présentés par F. Barac (Pre20 Title) (1)

Les Croates sous le joug magyar. (Confé. du foyer) (Pre20 Title) (1)

Les Croates vus par les Français (1)

Croatia (5)

Croatia (Journal Title) (4)

Croatia, 1941-1945 : before and after (1)

Croatia, 1998 (1)

Croatia & Bosnia (1)

Croatia : a history (1)




Croatia : a nation forged in war (2)

Croatia - a new European state : proceedings of the symposium held in Zagreb and Cakovec, September 22-25, 1993 (1)

Croatia - Adriatic Coast (1)

Croatia - Adriatic Coast, Dubrovnik (1)

Primjer 4. Hipertekstualna povezanost podataka iz zapisa s ostalim zapisima koji sadrže iste podatke.

Address  <http://aleph.nottingham.ac.uk/ALEPH>  

University of Nottingham Library Online Catalogue

[Log in](#) [Reset](#) [Search](#) [Results List](#) [Previous Searches](#) [User](#) [ILL](#) [Central Store](#) [Help](#)

Display format: [Standard](#) | [Citation](#) | [MARC tags](#)

Standard format allows you to:

Click "All items" to see all copies held and to make a reservation or click the sub-library name for copies in that library only.

Click on underlined entries (*) to find related references.

Click on an External link to access an electronic resource.

Record 4 out of 81 [Save/mail](#) [previous](#) [next](#)

All holdings [All items](#)

Sub-Lib Holdings [Hallward Library](#)

Main author * [Great Britain. Ordnance Survey](#)

Title * [South Derbyshire and NW Leicestershire 1896.](#)

Cartogr. data Scale 1:63,360. 1 in. to 1 mile.

Imprint * [Consett. Alan Godfrey Maps. 2002.](#)

Description 1 map : .31 x 46 cm. folded to 21 x 13 cm.
* [\[Old Ordnance Survey maps. England & Wales sheet 141\]](#)

Series [Surveyed in 1879-83. Revised in 1895-96 and published 1896. Boundaries revised to December 1898.](#)

General note With text on reverse by John Gough; an extract from the 1:2,500 map of Melbourne (1899); and an extract from Kelly's directory of Derbyshire 1891 (covering Melbourne).

General note "The Godfrey edition".

East Midlands sub. * [Maps -- Derbyshire](#)

East Midlands sub. * [Maps -- Leicestershire](#)

Added entry * [Gough, John, 1943-](#)

ISBN 1841513660

Primjer 5. Pregled prethodno provedenih pretraživanja.

Address <http://aleph.nottingham.ac.uk/ALEPH>

University of Nottingham Library Online Catalogue

[Login](#) [Reset](#) [Search](#) [Results List](#) [Previous Searches](#) [User](#) [ILL](#) [Central Store](#) [Help](#)

Previous Searches

[View](#) [Delete](#) [Combine](#)

View previous searches to work with them again
Delete searches from the history to simplify the list
Combine searches to narrow or widen your results

<input type="checkbox"/>	Catalogue	Search Request	No. of Records
<input type="checkbox"/>	UoN	Subject= (zagreb) OR Subject= (rijeka) OR Words= (osijek)	24
<input type="checkbox"/>	UoN	Authors=Bryson, Alex.	12
<input type="checkbox"/>	UoN	Subject= (gaudi) AND Subject= (barcelona)	2
<input type="checkbox"/>	UoN	Title word(s)= (south derbyshire)	81

[Information Services, University of Nottingham](#)

Primjer 6. Nakon što katalog nije pronašao traženu riječ u naslovu, dobiva se hipertekstualni popis sličnih riječi.

Address: <http://lib.ex.ac.uk/search/w?SEARCH=lbz>

University of Exeter Library Catalogue

LBZ is not in any titles.
Therefore "LBZ" is discarded.

[Previous Page](#) [Next Page](#) [New Search](#)




WORD: [View Entire Collection](#)

Mark	No matches found; nearby WORDS are:	Year	Entries
	lazurus		1
	lazy		11
	lazzara		1
	lazzaro		2
	lazzi		1
	lbi		1
	laboratories		1
Your entry lbz would be here			
	lca		3
	lcare		1
	lcc		3
	lcf		1


[Save Marked Records](#)

[Previous Page](#) [Next Page](#) [New Search](#)

Primjer 7. Filtriranje dobivenog skupa rezultata na temelju različitih kriterija.

Address  <http://cat.lib.lancs.ac.uk/FXATKJDPXCFCKRWDSPTLZL5M8UWCKV28XMR3R7NDXPCVFSGK-66575?func=short-filter>  

Lancaster University Library: Online Catalogue

	Log On	My Library Card	System News	Feedback	Help
Search by Collection	New Search	Results List	Previous Searches	Basket	

Filter a set of records :
w-titles= (croatia)

Choose a filter to view a subset of the List of Records

- **Alphabetical Range on data** (Example: Title field, from A to B, will show all titles in the list beginning with A through all titles beginning with B.)
- **Text** (Example: Title field, and poems for text, will show all records in the list that have the word *poems* in the title.)
- **Year** (Example: 1992 to 1995 will include all records in the list published from 1992 to the end of 1995.)
- **New records** (Example: Records that were added to the database in the last 7 days.)
- **Available records** (Records with copies available on the shelf.)


[Lancaster University Home Page](#) | [Lancaster University Library Home Page](#)

© 2002 Ex Libris & Lancaster University

Primjer 8. Sortiranje rezultata po različitim kriterijima.

Address <http://cat.lib.lancs.ac.uk/FXA/TKJDPXPFCKRWQZSP1ZL5NEW/MCKWV28XMR3R7NDJKPCVFS9K-65675?func=short>

Lancaster University Library: Online Catalogue

	Log On	My Library Card	System News	Feedback	Help
Search by Collection	New Search	Results List	Previous Searches	Basket	
View Selected	Email	Create Subset	Add to Basket		
Select All	Deselect	Refine	Filter		

Brief List

Sort options: [Auth/Title \(A\)](#) [Auth/Year \(D\)](#) [Auth/Year \(A\)](#) [Title/Year \(D\)](#) [Title/Year \(A\)](#) [Year/Title \(A\)](#) [Format/Auth \(A\)](#) [Format/Title \(A\)](#)




Click on the Title to display a catalogue record, or click on All Copies to display Classmark/Loans Information.



To mark records, click in the box and then click on

Records 1 - 11 of 11

#	Author	Title	Year	Format	Classmark/Loans
1	<input type="checkbox"/> Andreis, Josip.	Music in Croatia.	1974	Book	All Copies
2	<input type="checkbox"/> Eterovich, Francis H.	Croatia.		Book	All Copies
3	<input type="checkbox"/> Franolic, Branko.	Books on Croatia and Croatians recorded in the British Library general catalogue.		Book	All Copies
4	<input type="checkbox"/> Goldstein, Ivo.	Croatia.	1999	Book	All Copies
5	<input type="checkbox"/> Omrcanin, Ivo.	Diplomatic and political history of Croatia.	1972	Book	All Copies
6	<input type="checkbox"/> Rothenberg, Gunther E.	The military border in Croatia, 1740-1881.	1966	Book	All Copies
7	<input type="checkbox"/> Seton-Watson, Robert William.	Absolutism in Croatia.	1912	Book	All Copies
8	<input type="checkbox"/> Smith, James D.D.	Canada in Croatia.	1995	Pamphlet	All Copies
9	<input type="checkbox"/> Thompson, Mark.	Forging war.	1999	Book	All Copies
10	<input type="checkbox"/> United Nations. Institute for Disarmament Research	Managing arms in peace processes.	1996	Book	All Copies

Primjer 9. Razradene mogućnosti pretraživanja i prebiranja MeSH-a.

Address  <http://www.nlm.nih.gov/mesh/nbrowser.html>   Go

 **NATIONAL LIBRARY OF MEDICINE**  **MEDICAL SUBJECT HEADINGS**

[Health Information](#) | [Library Services](#) | [Research Programs](#) | [New & Noteworthy](#) | [General Information](#)

[MeSH Home](#) | [Contact NLM](#) | [Site Index](#) | [Search Our Web Site](#) | [NLM Home](#)

MeSH Browser.
The files are updated every week on Sunday.
[Browse 2002 MeSH](#)

Enter term or the beginning of any root fragments: or

Search for these record types: Main Headings Qualifiers Supplementary Concepts All of the Above


Search in these fields of chemicals: Search as text words in Annotation & Scope Note

Heading Mapped To (HM) (Supplementary List)
 Indexing Information (II) (Supplementary List)
 Pharmacological Action (PA)
 CAS Registry/EC Number (RN)
 Related CAS Registry Number (RR)

Search as Mesh Unique ID Search as text words in Annotation & Scope Note

[About MeSH Browser](#) | [MeSH Home Page](#) | [Questions or Comments](#)
[NLM Classification, the scheme used to categorize and organize books, audiovisuals, and similar materials.](#)

Primjer 10. Razradeni sustav uputnica i ostalih informacija o jednom deskriptoru u MeSH-u.

Address 	
MESH Descriptor Data	
Return to Entry Page	
MESH Heading	Kidney
Tree Number	A05.810.453
Annotation	kidney tissue or cells in cultures: Manual 18.6.15, 18.7.2, 26.23+; fetal or embryonic kidney tissue or cells in culture: do not use / embryol with KIDNEY; / blood supply: consider also RENAL CIRCULATION; / radionuclide: consider RADIOISOTOPE RENOGRAPHY, a specific kidney function test (see note there); / surg: consider NEPHRECTOMY; / transpl = KIDNEY TRANSPLANTATION: do not coord with TRANSPLANTATION; HOMOLOGOUS unless particularly discussed; cyst = KIDNEY, CYSTIC but KIDNEY, POLYCYSTIC is also available; inflammation = NEPHRITIS; Goldblatt kidney = HYPERTENSION, GOLDBLATT see HYPERTENSION, RENOVASCULAR; nephrostomy: consider NEPHROSTOMY, PERCUTANEOUS (see Mesh definition)
Entry Term	Kidneys
Entry Term	Renal Effects
See Also	Diuresis
See Also	Nephrectomy
See Also	Renal Circulation
Consider Also	consider also terms at <u>GLOMERUL-</u> , <u>NEPHR-</u> , <u>PYEL-</u> , and <u>RENAL</u>
Allowable Qualifiers	AB AH BS CH CY DE EM EN GD IM IN IR ME MI PA PH PP PS RA RE RI SE SU UL US VI
History Note	/transplantation was KIDNEY TRANSPLANTATION 1963-65; was KIDNEY/transplantation 1966-89
Entry Combination	transplantation:Kidney Transplantation
Unique ID	D007668

**Prilog 4. Prijedlog smjernica za osiguravanje mogućnosti
predmetnog pretraživanja u knjižničnim katalozima
visokoškolskih knjižnica s web-sučeljem**

Vrste i obilježja predmetnih pristupnica

1. Osigurati mogućnost pretraživanja i prebiranja po svakom polju iz zapisa.
2. Osigurati mogućnost pretraživanja po pojedinim riječima iz svakog polja u zapisu.
3. Osigurati mogućnost istodobnog pretraživanja i prebiranja po svim postojećim poljima i po pojedinim riječima iz svakog polja u zapisu.
4. Osigurati mogućnost pretraživanja i prebiranja po predmetnim odrednicama i riječima iz predmetnih odrednica.
5. Osigurati mogućnost pretraživanja i prebiranja po klasifikacijskim oznakama i pojedinim elementima klasifikacijske oznake, uz prikaz njihovih naziva (opisa).
6. Osigurati mogućnost pretraživanja i prebiranja po prijevodima klasifikacijskih oznaka.
7. Osigurati mogućnost pretraživanja i prebiranja po signaturi.
8. Osigurati mogućnost istodobnog pretraživanja i prebiranja po svim popisima predmetnih odrednica koji se koriste u knjižnici.
9. Pri označivanju predmetnim odrednicama ili deskriptorima, potrebno je napraviti razliku između glavnih i sporednih predmetnih odrednica, kako bi se sačuvala mogućnost pristupa po glavnom predmetu, a ne bi se onemogućio pristup po sporednim predmetima.
10. Osigurati mogućnost istodobnog pretraživanja i prebiranja po klasifikacijskim sustavima koji se koriste u knjižnici.
11. Osigurati povezivanje kontroliranih sustava i mogućnost integriranog pretraživanja i prebiranja po kontroliranim sustavima koji se koriste u knjižnici, uključujući signature.
12. Koristiti kontrolirane sustave koji se temelje na fasetama i među terminima odnosno oznakama imaju izgrađene odnose svojstvene tezaurusima.
13. U kontroliranim sustavima prebiranje valja omogućiti i po hijerarhijskim odnosima. U klasifikacijskom sustavu treba imati mogućnost odlaska s glavnih skupina na podskupine do najuže oznake, s pridruženim prijevodima oznaka, kao i u obrnutom smjeru.

14. Koristan bi bio ekspertni sustav koji bi automatski za sintetiziranu signaturu pokazao sve hijerarhijske popise za pojedine elemente signature.
15. Za kontrolirane sustave osigurati datoteku predmetnih preglednih kataložnih jedinica i uputnica, povezanu s bibliografskom datotekom.
16. Pri pretraživanju pomoću složenog upita po riječima iz termina ili po dijelovima oznaka kontroliranih sustava, potrebno je omogućiti prikaz pojedinačnih termina odnosno oznaka s pripadajućim prijevodima, zajedno s brojem bibliografskih zapisa uz svaki pojedinačan termin ili oznaku.
17. Ako pretraživanje počinje s riječju, sustav bi trebao prikazati i popis predmetnih odrednica i klasifikacijske tablice i tezaursus, ostavljajući korisniku da sam napravi izbor.
18. Predmetno pretraživanje pospješiti obogaćivanjem zapisa dodatnim predmetnim pristupnicama poput sažetka iz online knjižara i nakladničkih kuća.
19. Bez obzira na to na kojem se stadiju pretraživanja nalazi, korisnik uvijek treba biti u mogućnosti ograničiti pretraživanje po godini, jeziku, zemlji izdavanja i ostalim važnim kriterijima.
20. Potrebno je implementirati algoritme za računalnu obradu prirodnog jezika koji na temelju morfoloških ili sintaktičkih obilježja riječi iz upita omogućuju pronalaženje semantički relevantnih jedinica građe. U tom slučaju osigurati uputu o tome kako sustav obrađuje upit.
21. Za ispravljanje pogrešaka u pisanju i ujednačavanje pravopisnih varijanti potrebno je implementirati softver, koji će korisnika uputiti na ispravnu odnosno korištenu varijantu, ili će se automatski provesti pretraživanje, pri čemu treba osigurati uputu o provedenom postupku.
22. Ukoliko postoji mogućnost pretraživanja po potpunom tekstu jedinica građe, primjenjivati tehnike citatnog označivanja i pretraživanja.
23. Ukoliko katalog obuhvaća digitalne netekstualne jedinice građe poput slika, glazbenih melodija, kemijskih formula i geografskih koordinata, osigurati mogućnost predmetnog pretraživanja i ovih vrsta građe.

Rezultati pretraživanja

24. Potrebno je osigurati rangiran prikaz rezultata pretraživanja, i to po relevantnosti, po godini odnosno datumu izdavanja ili nastanka, abecedno po autoru, abecedno po naslovu i po ostalim važnim kriterijima.
25. Termine ili oznake kontroliranih rječnika u dobivenim zapisima potrebno je hipertekstualno povezati s ostalim zapisima koji sadrže iste termine ili oznake.
26. Ukoliko traženi termin ili oznaka ne postoji, rezultat pretraživanja treba biti ispis najbližih termina ili oznaka u abecednoj i hijerarhijskoj strukturi, s pojašnjenjem toga rezultata.
27. Tražene riječi, termine ili oznake iz upita u dobivenim zapisima treba istaknuti.
28. Potrebno je uspostaviti mehanizme za povratnu informaciju o relevantnosti dobivenih jedinica građe radi preoblikovanja upita.
29. Nakon dobivenog skupa rezultata, potrebno je ponuditi hipertekstualni popis kontroliranih termina ili prijevode oznaka koji odgovaraju riječima iz upita, i to iz svih kontroliranih sustava koji se koriste u knjižnici.
30. Potrebno je osigurati mogućnost pregledavanja prošlih pretraživanja.
31. Potrebno je osigurati mogućnost iskorištavanja dijelova prošlih pretraživanja pri oblikovanju novog upita.

Sintaksa predmetnog pretraživanja

32. Potrebno je osigurati mogućnost kraćenja, i to implicitnog, ali tako da se može isključiti, te eksplicitnog. Ovo vrijedi i za termine kontroliranog sustava, i prirodnog jezika.
33. Potrebno je osigurati korištenje Booleovih operatora i operatora blizine, i to implicitnih, ali tako da se mogu promijeniti, te eksplicitnih. Ovo vrijedi i za termine kontroliranog sustava, i prirodnog jezika.

Sučelje

34. Potrebno je osigurati barem dvije vrste sučelja, jedno za početnike, a drugo za napredne korisnike i informacijske stručnjake. Na naprednom sučelju potrebno je osigurati razrađene mogućnosti pretraživanja, a na početničkom one najvažnije i najjednostavnije.
35. Potrebno je oblikovati individualizirana sučelja prilagođena potrebama svakog korisnika.
36. Potrebno je osigurati višejezična sučelja i osigurati mehanizam prepoznavanja domene za državu iz koje korisnik pristupa katalogu te automatskog prikazivanja sučelja na jeziku te države.
37. Poželjno je osigurati i mogućnost prevođenja korisničkog upita na jezik baze koja se pretražuje.
38. Potrebno je koristiti međunarodni standard za kodiranje znakova Unicode ISO/IEC 10646.
39. Sučelje WebPAC-a potrebno je oblikovati na način da bude pristupačno osobama s posebnim potrebama i osobama koji koriste starije tehnologije.

Pomoć i upute

40. Potrebno je osigurati pomoć pri predmetnom pretraživanju, i to s konceptualnim aspektima pretraživanja, poput prevođenja korisnikovih pojmova na termine korištene u katalogu, semantičkim, poput odabira najprikladnijih mogućnosti, kao i s tehničkim aspektima pretraživanja, poput sintakse upita.
41. Upute je potrebno osigurati pri svakom koraku u pretraživanju (tzv. kontekstualne upute). Potrebno je osigurati i pomoć koja se poziva na zahtjev.
42. Potrebno je osigurati upute o tome kada je poželjna koja vrsta pretraživanja, a uz svaku opciju pretraživanja treba stajati uputa.
43. Upute uz pojedine mogućnosti pretraživanja trebaju jasno opisati koja se vrsta pretraživanja provodi.

44. Upute je potrebno osigurati kada na upit nema odgovora, bilo da se korisnik usmjeri općenito na to kako oblikovati upit, ili kako preoblikovati postojeći upit.
45. Upute je potrebno osigurati i ako na upit postoji određen broj odgovora, jer dobiveni zapisi možda ne zadovoljavaju korisnikove informacijske potrebe.
46. Upute trebaju biti detaljne, precizne, jasne i na laičkom jeziku.

Sadržaj

Uvod	1
1. Osnovne pretpostavke predmetnog pretraživanja u online knjižničnim katalozima.....	7
1.1. Predmetno pretraživanje kao zadaća knjižničnog kataloga.....	7
1.2. Organizacija i struktura baze podataka.....	9
2. Razvoj i pristupi problematici predmetnog pretraživanja online knjižničnih kataloga.....	13
2.1. Prednosti i nedostaci online kataloga u odnosu na kataloge na listićima u predmetnom pretraživanju	13
2.2. Razvoj predmetnog pretraživanja u online katalozima	14
3. Pristupi problematici predmetnog pretraživanja u online katalozima.....	17
4. Problematika korisničkih pristupa pri predmetnom pretraživanju.....	22
5. Predmetni pristup informacijama u online knjižničnim katalozima	29
5.1. Vrste predmetnih pristupnica.....	29
5.2. Vrste sustava za označivanje i pretraživanje	32
5.2.1. Podjela sustava za označivanje i pretraživanje prema prethodnom i naknadnom označivanju.....	35
5.3. Kontrolirani sustavi u predmetnom pretraživanju	36
5.3.1. Klasifikacijski sustavi	40
5.3.2. Popisi predmetnih odrednica.....	47
5.3.3. Označivanje na temelju kontroliranih sustava	49

5.3.3.1. Automatizacija označivanja na temelju kontroliranih sustava	51
5.4. Prirodni sustavi u predmetnom pretraživanju	52
5.5. Prirodni sustavi naspram kontroliranih i mogućnosti pretraživanja	54
5.6. Računalna obrada prirodnog jezika u predmetnom pretraživanju	58
6. Provođenje predmetnog pretraživanja	65
6.1. Operatori	65
6.2. Kraćenje	67
6.3. Oblikovanje upita	69
6.4. Rangiranje i prikaz rezultata	71
7. Korisničko sučelje za predmetno pretraživanje	74
7.1. Oblikovanje sučelja za predmetno pretraživanje	74
7.2. Pomoć i upute	77
7.3. Vrste korisničkih sučelja	78
7.4. Višejezičnost	80
7.5. Pristupačnost sučelja WebPAC-a	82
7.6. Dizajn sučelja	83
8. Online katalogi treće generacije - WebPAC-i	85
8.1. Okapi	85
8.2. Chesire II	88
8.3. WebPAC-i danas	89
9. Istraživanje WebPAC-a britanskih i hrvatskih sveučilišnih knjižnica	92
9.1. Uvod u istraživanje	92
9.2. Uzorak istraživanja	92
9.3. Metodologija istraživanja	93
9.4. Analiza rezultata	94
9.4.1. Općenita pitanja o sučelju	94

9.4.2. Prebiranje	94
9.4.3. Oblikovanje upita.....	96
9.4.4. Sintaksa	98
9.4.5. Rezultati pretraživanja	100
9.5. Zaključak istraživanja.....	103
Zaključak	105
Bibliografija.....	115
Sažetak.....	128
Summary.....	130
Životopis.....	132
Prilog 1. Popis knjižnica i URL-a njihovih kataloga obuhvaćenih istraživanjem.....	133
Prilog 2. Upitnik	137
Prilog 3. Primjeri naprednih svojstava predmetnog pretraživanja u WebPAC-ima	140
Prilog 4. Prijedlog smjernica za osiguravanje mogućnosti predmetnog pretraživanja u knjižničnim katalogima visokoškolskih knjižnica s web-sučeljem.....	151