

Naloga 4

Jaka Čop

20. november 2024

V vzajemni zbirki COBIB, zbirki LISC (na seznamu kliknite na rdeče obarvano ime zbirke) ter v DiKUL v osnovnem iskanju ločeno iščite s frazo „sistemi za poizvedovanje“ ter z angleško ustreznico „information retrieval systems“. Koliko rezultatov ste dobili pri vsakem iskanju? Kakšne rezultate ste dobili, npr. z vidika tipa dokumentov, jezika,...? Zakaj ste dobili v različnih zbirkah različne rezultate?

Komentirajte!

Pri iskanju po navedenih bazah smo dobili naslednja števila rezultatov:

jezik	COBIB	LISC	DiKUL
SLO	928	50	0
ENG	1992	53660	312293

Tabela 1: Št. rezultatov iskanja s frazo sistemi za poizvedovanje/information retrieval systems.

Že pred začetkom iskanja smo domnevali, da bo rezultatov v angleščini več, saj je angleški jezik veliko bolj razširjen in uporabljan. Poleg tega je angleščina svetovno uveljavljen jezik znanosti. Toda rezultati so nas vseeno nekoliko presenetili. V COBIB-u (slovensko omrežje knjižnic) je rezultatov približno dvakrat več v tujem jeziku, kar je bilo v okviru naših predhodnih pričakovanj. Že v zbirki LISC pa je rezultatov več kot $1000\times$ več. Ob krajšem premisleku pa ta rezultat morda ni tako presenetljiv. Slovensko govori optimisično tri milijone ljudi. Angleščino pa gotovo, vsaj kot drugi jezik, več kot tri milijarde ljudi (angleščina je eden od uradnih jezikov Indije). Kar je enak faktor (1000) kot pri primerjavi rezultatov med zbirkama COBIB in LISC. Ne vemo kako primerjati rezultate v zbirki e-virov Univerze v Ljubljani, DiKUL, med jezikoma, saj v slovenščini nismo dobili niti enega rezultata. Je pa številka, ki smo jo dobili z iskanjem angleške fraze ogromna, tudi v primerjavi z ostalima zbirkama. Digitalna Knjižnica Univerze v Ljubljani ima očitno na tem področju zakupljeno največjo zbirko virov.

V COBIB-u glede tipov gradiva v slovenščini prevladujejo knjige – večinoma diplomska in magistrska dela, medtem pa je v LISC-u 49 od 50 zadetkov klasificiranih kot znanstveni članek (scholarly journal). V angleščini v LISC-u prav tako dominirajo znanstveni članki (45427), v COBIB-u pa je največ knjig (1616). Od tega največ zbornikov (531). V DiKUL-u pa je skoraj polovica vseh zadetkov označenih kot znanstveni članek (156014). Glede jezika rezultatov pričakovano močno prevladujeta slovenščina in angleščina (vsak

v svojem primeru). Ni pa tu niti malo presenetljivo, da smo našli tudi nekaj drugih sorodnih in manj sorodnih jezikov, saj smo iskali po celotnem besedilu in ne po npr. predmetnih oznakah. Pri iskanju s slovensko frazo smo tako našli tudi nekaj rezultatov v hrvaščini, srbsčini, slovaščini (vsi slovanski – podobni jeziki), pa tudi nekaj rezultatov v ne tako zelo podobnih jezikih – celo v madžarščini. Podobno je bilo tudi pri iskanju z angleško frazo majhen del rezultatov v neangleškem jeziku. Rezultate v neiskanih jezikih smo našli zaradi dveh glavnih razlogov. Nekatera dela so dvojezična (najdeno madžarsko delo na primer), nekatera dela pa imajo predmetne oznake kot so: „informacijski sistemi“ in „poizvedovanje informacij“ in jih je naše iskanje, ki je iskalo po vseh možnih kriterijih vključilo v skupni nabor zadetkov.

Enega od razlogov za različno število rezultatov po različnih zbirkah smo že omenili. To je nekonsistentno dodeljevanje predmetnih oznak. Morda je nekonsistentno napačna beseda, toda za veliko starejših tekstov podatki verjetno enostavno še niso bili vnešeni in tako je pač to eden od razlogov za primanjkljaj rezultatov. Morda bi bilo na tej točki zanimivo pogledati na listkovni katalog, toda štetje rezultatov ročno bi bilo naravnost mučno. Po drugi strani pa je velika večina zadetkov na iskano temo nastala po letu 2000 (ne vemo sicer kdaj točno se je digitalizacija virov začela toda več kot polovica iskanih zadetkov je vsaj v zbirki DiKUL nastala celo po letu 2010). Zato je verjetno glavni razlog za razlike v številih zadetkov ta, da smo preprosto iskali po različnih zbirkah oziroma bazah podatkov. Verjetno nobena zbirka nima dostopa do vsega gradiva. Tudi NUK-ov mEga iskalnik nam, da različno število rezultatov kot katerakoli od, pri tej nalogi uporabljenih zbirk. Knjižnice in ustanove se zato povezujejo v razne konzorcije in skupno zakupujejo raznovrstne informacijske vire. To seveda ni poceni in posledično imamo celo študentje znotraj Univerze v Ljubljani dostop do različnih virov glede na to na katero fakulteto smo vpisani. Odprta znanost zares ne more priti prekmalu.