

Naloga 7

Jaka Čop

8. januar 2025

1. V sistemu COBISS v vzajemnem katalogu opravite poizvedbo na določeno temo.

Opišite potek poizvedovanja, preglejte relevantnost zadetkov ter izračunajte natančnost (preciznost, precision) in navedite, na kakšen način bi ocenili odziv (priklic, recall).

Če boste pri poizvedbi dobili več kot 30 zadetkov, zaradi lažjega dela upoštevajte le prvih 30 zadetkov (tako pri presoji relevantnosti kot pri izračunu natančnosti).

2. V neki zbirki je 1000 dokumentov, od katerih jih je 100 relevantnih za našo informacijsko zahtevo. Naredili smo poizvedbo, kjer smo dobili 250 zadetkov, od katerih je 50 relevantnih. Kakšna sta odziv in natančnost?

1. V sistemu COBISS smo sestavili sledečo poizvedbo (po ključnih besedah): umetna inteligenca IN znanstveno komuniciranje. Iz našega iskanja pa smo izločili (IN NE) diplomska, magistrska in doktorska dela, ker nas ta že iz začetka niso zanimala. Skupaj smo na to poizvedbo dobili 66 rezultatov. Morda bi v tem primeru res lahko pregledali vse, a smo se odločili omejiti na 25 rezultatov (to je ena COBISS-ovih možnosti, ki nam bo po našem mnenju še vedno dala dovolj dober približek). Od teh rezultatov je bilo za nas relevantnih 14. To nam, da natančnost:

$$N = \frac{RZ}{Z} = \frac{14}{25} = 0,56 = 56\%$$

Za oceno odzivnosti je verjetno najboljši način „Pooling“. Poskusili bi nekoliko spremeniti našo prvotno iskalno poizvedbo. Na primer s krajšanjem ali iskanjem z narekovaji. Preko takšnega iskanja bi potem lahko ocenili zgornjo mejo (kar se nam zdi najbolj smiseln način ocene). Pri vsakem iskanju bo verjetno nekaj nerelevantnih zadetkov in nekaj relevantnih – pri vsakem nekoliko različnih med seboj (različni najdeni zadetki in izpuščeni zadetki pri vsakem iskanju). Seštevanje teh rezultatov verjetno nima smisla (za oceno vseh zadetkov), ker bi vsakič prišteli veliko nerelevantnih zadetkov. Iz enakega razloga tudi povprečevanje teh rezultatov nima veliko smisla. Zato smo omenili zgornjo mejo vseh iskanj kot morda najboljši način ocene vseh relevantnih dokumentov. Vprašanje pa ostaja kako natančen je tudi ta način – verjetno ne zelo.

2. Podatki:

$$RD = 100$$

$$Z = 250$$

$$RZ = 50$$

Natančnost je tako:

$$N = \frac{RZ}{Z} = \frac{50}{250} = 0,2 = 20\%$$

odzivnost pa:

$$O = \frac{RZ}{RD} = \frac{50}{100} = 0,5 = 50\%$$

Komentar. Natančnost naše proizvodbe je bila sicer bistveno boljša od danega primera. Verjetno predvsem zaradi tega, ker je bila naša proizvodba dokaj specifična in je imela razmeroma malo zadetkov (66).