



**Zadatak** Koristeći Apache Lucene biblioteku indeksirati kolekciju tekstualnih fajlova i odraditi pretraživanje nad kreiranim indeksom.

1. Kreirati Java aplikaciju u sa nazivom **<broj\_indeksa>\_<ime>\_<prezime>\_lab1** u Eclipse okruženju i referencirati potrebne JAR fajlove iz Apache Lucene biblioteke.
2. Pronaći najmanje 4 obična tekstualna fajla veličine od 30KB do 500KB. Fajlove je npr. moguće pronaći npr. na sajtu Project Gutenberg ([www.gutenberg.org](http://www.gutenberg.org)) koji sadrži preko 50000 besplatnih knjiga **Obavezno je da svi studenti imaju različite kolekcije fajlova sa kojima rade.**
3. Ako želite da indeksirate knjigu koja je veća od 500KB odbacite deo teksta tako da **fajl koji indeksirate nije veći od 500KB.**
4. Kreirati indeks nad tim fajlovima tako da svaki fajl predstavlja poseban dokument u indeksu i sačuvati ga negde na fajl sistemu. Indeks treba da sadrži najmanje sledeća polja: sadržaj fajla, naziv fajla, kompletnu putanju do fajla na fajl sistemu i veličinu fajla u bajtovima.
5. Od prethodne kolekcije fajlova napraviti novu kolekciju od najmanje 400 fajlova istog sadržaja tako što se svaki od originalna 4 fajla deli na po 100 približno jednakih delova. Za podelu napisati posebnu aplikaciju u Javi ili koristiti neki gotov program koji to radi. Fajlove je moguće podeliti na bilo kom mestu tj. ne treba da se vodi računa o deljenju na kraju rečenice ili paragrafa.
6. Kreirati i drugi indeks sa novom kolekcijom fajlova i istim poljima i uporediti veličinu oba kreirana indeksa i vreme potrebno za njihovo kreiranje.
7. Kreirati bar jedan logički upit od najmanje 3 termina gde će se primeniti sve 3 logičke operacije (**NOT**, **AND** i **OR**) i izvršiti ga nad jednim i nad drugim indeksom. **Identični upit kreirati na 2 načina** – direktno kreirati objektni model bez korišćenja parsera i isti objektni model upita kreirati parsiranjem tekstualnog upita. Kreirani upit izvršiti nad jednim i nad drugim indeksom.
8. Kreirati još jedan upit. Tip upita odredite na osnovu **poslednje cifre svog broja indeksa**:
  - 0 ili 5 – **TermRangeQuery**,
  - 1 ili 6 – **PointRangeQuery**,
  - 2 ili 7 – **PrefixQuery**,
  - 3 ili 8 – **WildcardQuery**,
  - 4 ili 9 – **PhraseQuery**.

**Identični upit kreirati na 2 načina** – direktno kreirati objektni model bez korišćenja parsera i isti objektni model upita kreirati parsiranjem tekstualnog upita. Kreirani upit izvršiti nad jednim i nad drugim indeksom.