

## IR Homework: List Intersection

### Exercise 1 (80%)

1. Make sure that each inverted list contains a particular record id **at most once**, even if the respective word occurs multiple time in the same record. Make sure that the whole construction algorithm still runs in **linear time** in the number of words in the input. (16%)
2. Write a function **intersect** that computes the intersection of two inverted lists. The function must run in **linear time** in the total number of elements in the two lists. (24%)
3. Write a function process query that, given a **keyword** query, fetches the inverted lists for **each of the keywords** and computes their intersection (empty, if there is no inverted list for a keyword). (24%)
4. Write a function **main function** that constructs an inverted index from a given **text file** (one record per line, file name given as first argument on the command line) and then, in an infinite loop, asks the user for keyword queries and **outputs 3 matching records**. Optionally highlight the query words in the output, e.g., using ANSI escape codes. (16%)

### Exercise 2 (20 %)

Try your code on the file movies.txt on the Wiki. Find a query that gives good results (the records shown meet your expectations) and one that does not. Write them in your experiences.txt (see below), and very concisely (in one or two sentences) explain why one works and the other doesn't.