Java Developer Test Assignment

The aim of this test assignment is to write a program in Java which counts top Javascript libraries used in web pages found on Google.

Prerequisites

Please create an easily compilable project (e.g. Maven/Gradle/Intellij/Eclipse project).

Preferred time spent on the task is 3 hours. We expect incomplete solutions - the task is designed that way. In any case please provide notes on how would you finish or improve the solution if more time is provided.

Task - Web crawler

The command-line program should do the following:

0) Read a string (search term) from standard input

1) Get a Google result page for the search term

2) Extract main result links from the page

3) Download the respective pages and extract the names of Javascript libraries used in them

4) Print top 5 most used libraries to standard output

Bonus steps

- write tests or think about the approach for testing your code

- think about / implement Java concurrency utilities to speed up certain tasks

- think about / implement deduplication algorithms for the same Javascript libraries with different names

Notes

- use whatever approach you think is the best and most efficient, you don't need to create elaborate or complex parsing algorithms

- you can skip a step if it's too hard (you can mock data for the next step and provide some notes)

- if something is not clear or can be done in multiple ways, describe why you chose your approach

- use a minimum of 3rd party libraries