IR Course Project

Prof. Fabio Crestani & Prof. Monica Landoni

Universitá della Svizzera Italiana (USI)

Switzerland

A Reminder …

- Marking:
  - 2 theoretical course tests, on different parts of the course content
  - 1 practical course test (or mini project)

- Marking scheme:
  - 70% theoretical course tests (35% each)
  - 30% practical course test (a.k.a. project)

- Remember not miss any of them, otherwise you will lose that part of the final mark!

Course tests

- More specifically, there are 3 course tests:
  1. On textual IR, evaluation, and interfaces (ch. 1, 5-9, user interfaces & user eval)
  2. On clustering, recommendation, categorization, and web search (ch. 10-12, 14, 15, and web search interfaces)
  3. **Small practical course project to show that you "understood" the theoretical part!**

- The course project count for 30% of your final course marking!

The course project

- The course project consists in the design, implementation and evaluation of a search engine, *given some simple user and system requirements*

- Notice the steps:
  1. Design (with justification of the design choices)
  2. Search engine implementation
  3. Search interface implementation
  4. Evaluation (user evaluation)

- Notice also: it is a *joint* project!

  - One project, two students!
  - Each project has a different set of system requirements
  - Each project will be user evaluated by at least 3 users (of your choice)

Example of Course Project

- Example from last year:

Notice:

- The specific description
- The additional features
- The evaluation

Information Retrieval
Academic Year 2022-23
Course Project
Prof. Monica Landoni

For this assignment you will work in pairs to implement a working prototype of an information retrieval system for a specific task and user needs. The goal of this assignment is to apply the concepts and tools you have learned in the class to a practical, real world, application.

Description

Each group was assigned a topic, together with an example website you need to crawl your data from. You need to build a system that gathers a large collection of samples associated with the topic and enable the search over this collection. To build the collections, you need to crawl multiple sources (websites), similar in content and topic to the website suggested. The system must provide an interface for searching, browsing, and presentation of the data to the user.

The system should also have two additional features. Additional features are described in the document on iCorsi and categorized into two groups: complex and simple features. You can choose which features you want to implement yourself, provided there is at least one feature from each of the groups (two in total).

Submission procedure and evaluation

An important step is the evaluation of your tool. Three students of the class, of your choice, will act as test users of your system and help you in evaluating your system from the point of view of the user experience. System evaluation (i.e. recall and precision) is not required. Please, coordinate with the other groups in order to meet the deadline.

You have to produce a report (max. 10 pages, including cover) of your work and of the evaluation. It will contain a concise explanation of how you tackled the design and implementation of the system. The code of the project and the report need to be submitted via iCorsi by end of day 7th December.

Additional Features

Two types of features:

You should pick one simple and one complex feature

Simple Features

Results presentation: results should be presented in a tabular format, so that many results could be seen at the same time. Each table cell should contain appropriate information for your project.

Filtering: in addition to being able to search by title, an user should be able to filter the results based on at least 3 relevant attributes for your project.

Results Snippets: present result snippets of each retrieved result (maximum 2-3 lines) in a kind of "Google style", with query terms highlighted.

Complex Features

Automatic Recommendation: In addition to the relevant search results pertaining to the user query, the user should also be suggested "similar" products based on, say, category, description, price etc. The ordering among the recommended items is not important. However, you should mention how did you arrive at the recommendations. For this purpose, you can use any open-source recommenders available.

User Relevance Feedback: after presenting the search results to an user, the user may provide a positive or negative feedback on the results (i.e. mark relevant and irrelevant documents). Based on this feedback the search results have to be updated and presented again.

Results clustering: the system should group results into topics. There should be a possibility for expanding/shrinking a topic to show all the results related to the topic. In addition, topics should be sorted in a descending order by the number of relevant results under them.

The report and the code

- What do you have to submit?
  - Project report, code, and slides
1. The report:
   - A concise explanation (max 10 pages) of how you tackled the design, implementation and evaluation
     - Explain design choices, implementation choices, evaluation design, evaluation results
   - Just imagine this is a report for a client who does not know much about IR, but wants to know what you are delivering him/her
2. The code:
   - A Zip file with all the code, sufficiently documented to make it possible to evaluate it

Submission …

- Submission:
  - Submission is via iCorsi
    - Or anywhere else, if you provide us with a link in time!
  - Progress report (one page max) by **November 13**th (end of day)
  - Report and code must be submitted by **December 11**th (end of day)
    - Submission site will be open from a few days earlier
  - Each day of delay in submission cut your mark by 10%

… and marking

- The marking is based on the content of the report and the code (in this order)
  - Remember: each day of delay in submitting the report or the code results in a 10% reduction of the mark
  - Some of you might be called to give a demo of your project, to check if everything works

- The marking will be available via iCorsi a few weeks later!

Questions

The Projects ...

A sample:

N.	Project	Guidelines	Example Website
1	Rent a car	Choose at least 3 different websites about Renting a car	https://www.centauro.net/it/
2	Basketball Players	Choose at least 3 different websites about Basketball Players	https://www.basketball-reference.com
3	Footbal Clothing	Choose at least 3 different websites that sell football clothes	https://www.prodirectsport.it/l/adulto/categorie-scarpe-da-calcio/attività-calcio/?.
4	Online Florist	Choose at least 3 different websites for buying plants and flowers	https://www.interflora.it/?utm_source=google&utm_medium=cpc&utm_campaign
5	Fitness Exercises	Choose at least 3 websites listing fitness exercises or routines.	https://www.bodybuilding.com/
6	Music Albums	Choose at least 3 websites listing music albums or reviews.	https://www.rollingstone.com/music/
7	Recipes	Choose at least 3 websites listing recipes.	https://www.allrecipes.com/
8	Board Games	Choose at least 3 websites listing board games or reviews.	https://boardgamegeek.com/
9	Science Experiments	Choose at least 3 websites listing science experiments or tutorials.	https://www.sciencebuddies.org/
10	Tattoo Designs	Choose at least 3 websites showcasing tattoo designs or artists.	https://www.tattoodo.com/
11	Photography Tips	Choose at least 3 websites offering photography tips or tutorials.	https://www.dpreview.com/
12	Hiking in Switzerland	Choose at least 3 different websites about outdoor activities/hiking in Switzerland	https://www.myswitzerland.com/it-ch/scoprire-la-svizzera/estate-autunno/escursi
13	Movies	Choose at least 3 websites about movies or movie reviews.	https://www.imdb.com/
14	Art for Sale	Choose at least 3 websites that help artists to sell their artworks.	https://www.catawiki.com/
15	Crypt0Currenc1es	Choose at least 3 websites that show informations about the trends of crypto-currer	https://coinmarketcap.com/https://www.coingecko.com/
16	Nerd Gadgets	Choose at least 3 websites selling only gadgets that nerds (or also geeks) would buy	https://www.dottorgadget.it/

and … Project Assignment!

- This will be available next Monday!

- Please fill the form as soon as possible or …

- You have until Wednesday 18th for "group assignment complaints"

12

Questions