Università
della
Svizzera
italiana

**Facoltà
di scienze
informatiche**

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.

Projects

N.	Project	Guidelines	Example Website
1	**IKEA hacks**	Choose at least 3 different websites with IKEA hacks \ articles about IKEA furniture \ reviews of IKEA products.	https://ikeahackers.net/category/hacks
2	**TV Series Episodes**	Choose at least 3 different websites about TV series'	https://www.tvmaze.com/
3	**Volunteer camps**	Choose at least 3 websites listing volunteer camps or opportunities.	https://www.workcamps.sci.ngo/icamps/welcome.html or https://www.volunteerworld.com/en
4	**Hairdressers**	Choose at least 3 websites and then focus on a specific city, any other kind of professional (plumbers, lawyers etc) is also ok. If you want to mix and match (like plumbers, electricians and painters, everything for the home, or hairdressers + aestheticians) it's also fine.	https://www.fresha.com/lp/en/bt/hair-salons/in/gb-london
5	**Clothing**	Choose at least 3 websites listing clothing (Asos, Zalando, Shein...)	https://www.asos.com/
6	**Short-term rentals**	Choose at least 3 websites, then focus on a specific city\area.	https://www.airbnb.co.uk/
7	**Textbooks**	Choose at least 3 websites liting textbooks (any subject)	https://open.umn.edu/opentextbooks
8	**Charities**	Choose at least 3 websites listing charities.	https://www.guidestar.org/NonprofitDirectory.aspx
9	**Indie Games**	Choose at least 3 websites listing indie videogames (commercial platforms like Steam are ok but only crawl indie games)	https://itch.io/games
10	**Stocks**	Choose at least 3 different stock exchanges.	https://www.borsaitaliana.it/borsa/azioni/listino-a-z.html?initial=A&lang=en
11	**Travel experiences**	Choose at least 3 websites listing travel experiences/tour, then focus on a specific city\area\country.	https://www.getyourguide.co.uk/london-l57/
12	**PhD positions**	Choose at least 3 websites listing PhD positions, any subject.	https://academicpositions.com/jobs/position/phd
13	**Content creators**	Choose at least 3 websites for Content Creators (like Patreon, Ko-fi...but please keep it safe for work!)	https://www.patreon.com/en-GB
14	**Concerts**	Choose at least 3 websites, then focus on a specific country, or 3	https://www.ticketmaster.ch/music/concerts/1221/events?language=en-us

		smaller countries (Switzerland is a small country!)	
15	**Pokemon**	Choose at least 3 websites listing pokemons.	https://www.pokemon.com/us/pokedex/
16	**Pets for adoption**	Choose at least 3 websites listing pets for adoption.	https://www.petfinder.com/
17	**Nannies/babysitters**	Choose at least 3 websites listing nannies\babysitters, then focus on a specific city\areas.	https://www.childcare.co.uk/find/Babysitters
18	**Motorcycles**	Choose at least 3 websites listing motorcycles.	https://www.motorcycle.com/
19	**Mangas**	Choose at least 3 websites listing Mangas. Manwha or Manhua (Korean or Chinese comics) are also ok!	https://mangatx.com/manga/
20	**Second-hand items**	Choose at least 3 websites offering second-hand items.	https://www.secondhand.org.uk/
21	**Football teams**	Choose at least 3 websites listing football teams, not necessarily from the same country.	https://footballdatabase.com/
22	**Medications**	Choose at least 3 websites listing medications.	https://www.drugs.com/drug_information.html
23	**Knitting patterns**	Choose at least 3 websites listing knitting\crochet patterns.	https://www.garnstudio.com/search.php?action=browse&c=home&lang=en

Groups

	Partner 1	Partner 2	Project number
1	Agostino Monti	Marco Farace	2
2	Albert Cerfeda	Alessandro Gobbetti	13
3	Alen Sugimoto	Mattia Monari	7
4	Alessandro Cagnani	Diell Kryeziu	5
5	Alessandro Cravioglio	Alfio Vavassori	19
6	Andrea Cardia	Fabrizio De Castelli	16
7	Andrea Gualandris	Alan Copa	6
8	Andrea Prato	Samuel Corecco	14
9	Anton Tanev	Manuele Jelmini	1
10	Alessio Cordivani	Roberto Palmieri	4
11	Bojan Lazarevski	Daniela Gjorgjieva	20
12	Catarina Morais	Michele Zucchi	8
13	Enrico Di Pietro	Giacomo Solaro	17
14	Francesco Casarella	Marco Biasion	15
15	Francesco Costa	Arnaud Fauconnet	21
16	Johan Jacob	Dylan Reid Ramelli	11
17	Kelvin Likollari	Gerald Prendi	18
18	Kevin Spahiu	Fabio Zampiello	10
19	Mak Fazlic	Harkeerat Sawhney	3
29	Ramazan Tafa	Daniel Klimas	9
21	Razvan Petrica Onciu	Elisa Spinelli	12
22	Zhang Xufeng	Arash Mowdoudi	23
23	Bahram Ismayilov	Hoormazd Pirayeshfar	22

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.