

Semester assignment INFO345

Alain Starke

This document contains more details about the different assignment possibilities. The goal the semester assignment is write a research report of around 15 pages about your recommender system method. This can be an online evaluation in which you re-rank search results based on a certain metric (Option 1), two offline evaluation methods (1 collaborative, 1 content-based) on a recipe database, creating a food recommender system (Option 2), or a suggestion of your own (please contact me for this ASAP).

You will work in groups of 1-3. Please notify me (alain.starke@uib.no) of your group composition by September 16, 16:00. The deadline of the assignment is December 7, 14:00 – for which you will need to submit your report and share the materials you’ve used (web code, Python scripts, etc.).

Assignment options

Initially, I had planned a single assignment possibility for you. However, since you have such different backgrounds, I’ll allow for a few options. You need to pick one of them. This document will provide some more details on what each assignment option entails. All options will be evaluated according to the reporting quality, which should contain a short introduction of the problem, some reference to related scientific work, a research question, an overview of the used materials (database composition, methods/algorithms, procedure of web tool), research results, and discussion of ‘what it all means’.

Options 1 (online) and 2 (offline) will both use data from the AllRecipes database. You can find the relevant data for this on Mitt Uib. The database is quite expansive and for the online evaluation, you might not need to use all of the recipes (the website will become slow). In the current prototype, I’ve selected 32 recipes, but you can of course choose more and/or different ones!

In addition to the criteria described above, assignment 1 will be assessed on its research originality (does it address a new problem? This requires a little literature search!), assignment 2 will be assessed on the use of relevant offline evaluation metrics. For example, collaborative methods could be evaluated on precision & recall of relevant recommendation lists.

General report structure

The document should be around 15 pages long. It should be in the ACM proceedings template, which can be found here: <https://www.acm.org/publications/proceedings-template>. As soon as you have sent me your group composition, I will send you a link to an Overleaf document, so that you can collaborate there in real-life. Hence, the document should be processed in the LaTeX language, but Overleaf is a little more user-friendly.

- **Abstract:** a short summary of your
- **Introduction:** this includes a sketch of the problem at hand and introduces relevant research questions
- **Background:** Summarize the state-of-the-art of recommender research in what you are investigating. The last part of the background should include a section what your methods introduced in the paper do potentially better than the current state-of-the-art (offline evaluation) or what is the novelty of your approach (offline / online)
- **Methods:** Here you first describe the dataset and then the different recommender approaches, evaluations protocols, and possible baselines.
- **Results:** A latex table that including your approaches and the baselines (offline), or a short statistical analysis of your online method.
- **Discussion** (including conclusions, limitations and future work)
- **References** (in ACM style)

Titles of possible relevant work (By Christoph Trattner; this list might be updated):

- Exploiting Food Choice Biases for Healthier Recipe Recommendation
- Investigating the Healthiness of Internet-Sourced Recipes: Implications for Meal Planning and Recommender Systems

Options for assignments

Option 1: the web app (probably the most feasible)

Integrate a recommender component into a web search app. You will be provided a search prototype that actually works! At the moment, it is designed to serve in an experiment where we try to prioritize healthy food by re-ranking the search results and showing pretty images. Test it on your local server and you might see something like this below:

Recipe Search Tool





Task 2/4

Use the search bar below to search for a recipe. If you do so, eight recipes will appear on your screen. Please inspect them carefully and click on the recipe that you like the most. Afterwards we will ask you a few questions about the chosen recipe and the list of search results.

Please use the search bar below to type in the following search term:
Pasta. It is important that you use **exactly** this search term.

Please click on the recipe that you are most likely to make

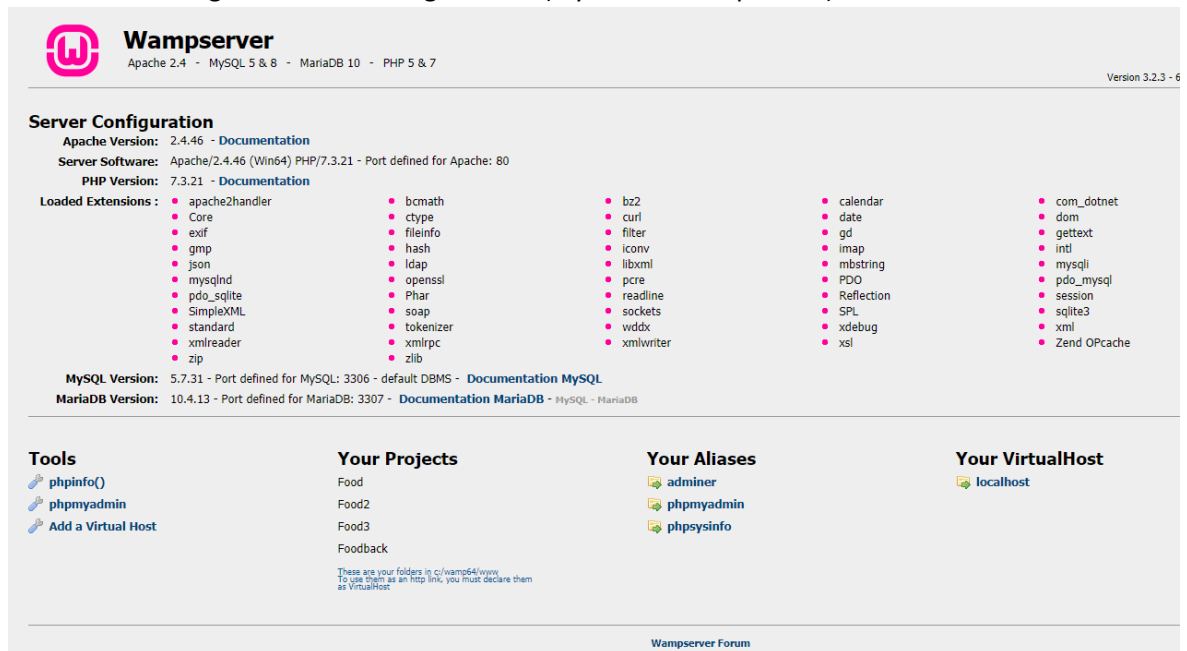
× 🔍

	Pasta with Spinach and Smoked Sausage
	Pasta with Lentil Soup Sauce
	Pasta Pomodoro
	Fusili pasta with Brocoli

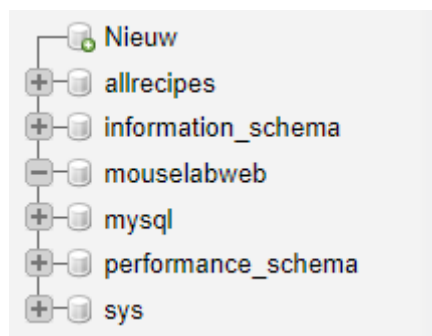
Installation etc. works as follows:

1. Go to Mitt UiB and download Food.zip. Thereafter, you can start to adapt the code to make something on your own.
2. However, to do that, you would first need to install a local server so that you can actually test whether the thing you are programming actually works. There are a few options for this: I have installed WampServer: <https://www.wampserver.com/en/>. You can just select the default options for installation.
3. After Wampserver is installed, you need to unpack Food.zip in the folder of your local server. For me, it is located at C:\wamp64\www. This will give you C:\wamp64\www\Food\ (The www folder is the root folder of your local server).
4. Then you can start your server by loading wampserver.exe from your start menu. As soon as the icon in the bottom right of your screen turns green, it's working!

5. If you want to go to the root folder of your server, load an internet browser and go to: “localhost”. It might look something like this (if you use Wampserver)



6. The main tool for storing server data is phpMyAdmin. You load it by going to: <http://localhost/phpmyadmin/index.php>. You will need to provide a username and password. The username is by default root, the password is [empty] (no password). If successful, you'll see a number of existing databases. A database contains a number of columns and rows in terms of data – such as a data matrix.
- For this project, you will need to create a database that has the same name as a local file (in Food.zip) that's called “mlwebdb.inc.php” (this file contains the credentials to communicate with your database). For now, call the database “mouselabweb”.



These are the database. Click 'new' or 'ny' or 'Nieuw' (if you are Dutch) to create a new database

7. To make your prototype work, you will need to run two .php queries. (That's what they are called). Go to your browser, open a new tab and run: http://localhost/Food/create_exps_table.php, and also run: http://localhost/Food/create_table.php. They should feedback to your something like 'Table created' if successful.

If you have taken all of these steps, then you should be able to test run the current version of the search prototype. So, each web project has a certain URL with which you can start the entire thing.

Each of the pages work separately, but this prototype has been built an experiment in which we expect each user to take the same steps.

8. To start the search prototype, go to: <http://localhost/Food/Informedconsent.php>. If successful, you will see this:

Recipe Search Tool

Informed consent

Welcome to our study 'Recipe Search'! The aim of this study is to understand how people search for online recipes and make food decisions. This study is designed by students of the Human-Technology Interaction Group at the Eindhoven University of Technology, NL, under supervision of dr.ir. Alain Starke & dr.ir. Martijn Willemsen. Before clicking 'start' to proceed, please carefully read the procedure followed in this study and give your informed consent for voluntary participation*. You can stop your participation at any time during the study.

Procedure

You will be asked to search for four different recipes using different keywords. For each list of search results, you will be asked to choose the recipe that you like the most. In addition, you will be asked to answer a few questions after each list of search results, as well as to fill out a short questionnaire at the end of the study. The study does not involve any risks, detrimental side effects, or cause discomfort.

Duration & Compensation

The study takes about 7 minutes to complete. You will be compensated with £1,- or \$1.25 for your participation.

By clicking the start button, you agree to have read and understood this consent form and that you agree to voluntarily participate in this research study of the Human-Technology Interaction Group.

Do you accept the informed consent?*

☐ No ☒ Yes, Accept

*All research conducted at the Human-Technology Interaction Group adheres to the Code of Ethics of the NIP (Nederlands Instituut voor Psychologen – Dutch Institute for Psychologists). We will not share personal information about you to anyone outside of the research team. The information that we collect from this research project is used for writing scientific publications and will be reported at group level. It will be completely anonymous and it cannot be traced back to you. For questions about the experiment, please contact the experimenter, Dr.ir. Alain Starke (a.d.starke@tue.nl). If you have any complaints about this experiment, please contact the supervisor, Dr.ir. Martijn Willemsen (M.C.Willemsen@tue.nl).

Each web study you will run in your life at a university start with what is called an Informed Consent. On this page, you will tell a prospective user of your experiment what can be expected from the study, what data is stored, who are the project owner, what is the reward, etc. etc. This experiment was initially run at a platform called 'Prolific', so the language is geared towards that audience.

If you wish, you can adapt the text in this screen. You can just change the text in the file C:\wamp64\www\Food\Informedconsent.php on your computer! You would first need to decide the programming software you are going to use.

9. Install programming software. I suppose that quite a lot of you have already many their own choices here. That's alright. In case you haven't, software I like a lot is called Sublime Text. You can download it here: <https://www.sublimetext.com/3>
10. So, now everything should work and you can use the tool. The choices that you make in the tool are actually stored! You can find the data via <http://localhost/phpmyadmin/index.php> and your mouselabweb database. Alternatively, you can also run what we have called 'the Datalyser'. It visualized the data in a somewhat nicer format. You can load it via <http://localhost/Food/datalyser.php>.

The search prototype has a lot of files that are relevant, so it's a bit of a hassle to go through it. If you just test-run the application, you can see the URL changing. This is also the order of the files that are being used:

- Informedconsent.php
- Intro.php
- Index1.php
- SurveyC1.php
- Index2.php
- SurveyC2.php]
- Index3.php

- SurveyC3.php
- Index4.php
- SurveyC4.php
- Thanks.html

Other files that are relevant and you should change at some point are:

- Visual_data.csv → this is the main data file. It contains a number of parameters: the rating of a recipe, its healthiness (by means of the FSA score), a location of its photo that is shown in the search prototype (which is also stored in the project folder) and its name.
- The queries: QueryBase, QueryRanking, QueryVisual, QueryVisualRanking → as you might have noticed, there are four search screens in this experiment. These query files each have different conditions that determine what images are shown in the list (pretty ones for healthy foods or ugly ones; queryvisual), the order of the list (ranked on popularity (querybase) or on health (queryranking), or both (queryvisualranking)). The order in which these queries are activated is randomized in Intro.php!
- Save.php → this helps to store the responses in the study correctly (in the database)
- mlwebdb.inc.php → this file establishes the database connection. In here, you need to indicate the username for your server, the password, the database name, etc. etc.

A lot of the other files are irrelevant or you simply should not touch 😊. For context: 'js' contains the core of JQuery, 'Zend' contains an optimized search protocol, 'json_files' are from a different project, 'css' can be updated if necessary, but this can also be hardcoded. The folders with images and imageselection you actually can touch: these are used to load the images in the search prototype!

So, how can I add something to this search prototype?

- The data file that is currently used is very limited! You can expand this by considering the larger database of recipes. However, the larger the .csv file you use, the longer the loading time will be!
- As you can see, there is a 'first screen' (intro.php). You could use this screen to ask actual users questions about their preferences and update something in the search that is personalized to this. You can, e.g., ask a question about 'do you like fat food' and then later on rank the fattiest burgers the highest up the list if the user searches for that!
- You can do some offline evaluation of data to come up with a new recommender approach that can be matched to some simple input in the first screen: you could ask users to rate a few recipes, for example.
- You could also think of ranking the search results on any other parameter to nudge users towards different foods. For example, you might be able to obtain information on food sustainability and re-rank the search results based on sustainability! Try to come up with new ideas of nudging users towards different options in the list.

!! Update: resolving problems of search results that are not updated

After you do some testing with the original file set 'Food.zip', you suddenly have the folders index1, index2, index3, and index4 in your Food folder. This is normal: these are more easily accessible indexations of the data file visual_data.csv, which speed up the search for future queries. **HOWEVER!** This means that if you make changes in visual_data.csv (e.g., add data), they won't show up in your search results. So, in order to see the changes after you have edited the data file (visual_data.csv), you would need to delete the index1-4 folders and try again.

This is of course a bit cumbersome. The workaround for not deleting the folders after every change in the data file is to include the line `"deleteDir($index_file)"` in each of the query files (QueryBase, QueryVisual, etc.). These commands are already in there, but I commented them out. For example, in QueryBase.php, there is one at line 70.

Finally: the code is coded to read tab-separated input. If you decide to edit data or to add data by hand, please be careful to do so in Excel, as it might store the data in a tab-separated format. Notepad typically works better. Alternatively, compile a datafile in your SQL-server, R, Python, wherever, and export the data columns as a tab-separated csv-file.

Option 2: offline evaluation of a food recommender system

For this assignment, you will need to think of recommendation method in food recommender systems. It should be novel in the sense that you can't go to GitHub, download an existing project, and submit it as a research project. However, if one paper or one person has already had a similar idea as you, it's okay to pursue that as well.

You will implement 2 different types of recommender algorithms. Either:

- One (or multiple) collaborative filtering (CF) method
- One content-based (CB) method
- One hybrid method (that should be a method that included CF, CB, or a knowledge-based method)

The code for the method can already exist in Python, but the type of analysis and the dataset should be different.

Evaluation criteria:

- You will perform offline evaluation (5-fold cross validation)
- Metrics: P@10, R@10, NDCG@10 AND another metric. This could be Healthiness of the list in terms of FSA and WHO health standards, but you could also think of another metric – and even find data for that!

Option 3: you own idea? Contact me!