# Assignment 2: Algorithm Design (Project Work)

**2024-11-01**

Försök 1 Pågår

**NÄSTA: Lämna in uppgift**

Lägg till kommentar

**Obegränsat antal försök tillåts**

2024-11-01

**Information**

# General Information

## The Gist

The general idea for this assignment is that you work together in a group to curate a dataset, and design an algorithm to process that dataset in a meaningful way. You will publish both your dataset and your algorithm design on GitHub, alongside a simple website with some basic information on the project, dataset and algorithm, that will be hosted with GitHub Pages.

## Forming Groups

This is a group exercise, that is intended to be completed in groups of 3 or 4 students. Completing the assignment individually is not allowed. Students are expected to organise themselves into groups **before the Getting Started (https://hb.instructure.com/courses/8313/pages/getting-started) info session on Tuesday 22 October at the latest.** For more information on how to organise yourselves into groups, please refer to our page on **Forming Project Groups (https://hb.instructure.com/courses/8313/pages/forming-project-groups)** . If you are unable to form a group of the required size, please contact your teacher.

## Grading

This assignment will be graded on a scale from F to A, where any grade from E upwards constitutes a pass. To obtain a passing grade, all the 'minimal requirements' for each step in the project (described below) need to be met.

While in practice students in the same group are likely to obtain the same grade for the assignment, this is not necessarily the case. To determine whether or not we should adjust the grade of individual group members (typically: 1 grade up or down from the overall achievement), we will take additional information into account, such as (but not limited to):

- Your description of the division of labour (see STEP 4 below)
- Individual contributions in the `commit` history of your GitHub repositories

# Project Workflow

## Step 1: Conceptualisation

The first step for this project would be to brainstorm with your teammates to decide what kind of data you would be interested in, and what you would like to be able to do with that data. This includes:

1. Finding or conceptualising an interesting dataset
2. Defining a specific problem or need for the data, that can be resolved algorithmically
3. Modelling (or fine-tuning the pre-existing model of) your data in a way that it becomes relevant for your algorithm

This process involves discussing questions such as: Where would you find the data you want to work with? Or would you need to collect or fabricate the data yourself? In what format is the data available? Can we work with this format in our algorithm? Or do we need to translate the data to another format? Do we need to clean the data in any way (e.g.: delete irrelevant or duplicate information; structure the data differently; correct errors in the data; add additional information to the data that would be useful for the algorithm)? How do we intend to use the data, and do we need to make any changes to the (way the) data (is structured) to facilitate this?

It is important for this step that you **know your data**. This requires you to have a good understanding of how the data you want to work with is modelled. E.g. for XML data: Which tags and attributes contain which information? Or, in the case of a database: Which attributes do the tables use, and how are the relations between different tables linked? At the same time, it is important to try to **anticipate the needs of your algorithm**. What kind of information would our algorithm require, and how is this information recorded in the dataset? Is the dataset modelled in a way that makes it easy for my algorithm to obtain the relevant data?

For example: if you were designing an algorithm that maps out a citation network (as in: who is citing who?) based on a selection of academic papers, your data would have to be modelled in a way that makes it easy for your algorithm to find a) the author names of each of the papers in your database; and b) the author names of each of the references cited in each of those papers.

If you are starting from pre-existing data, this means: understanding what kind of information it contains, and how it is structured. If you are developing your own dataset, this means conceptualising your own model to map your data onto (as you did, for example, in Assignment 1). You may also want to do a little of both: use pre-existing data, but make some changes to the way it is structured (or the information it contains) that makes it easier for your algorithm to access the information you need.

Grading

| | Minimal Requirements |
|---|---|
| 1 | **Your dataset should be modelled in a way that is relevant for your algorithm.** Meeting this requirement should come quite naturally if steps 2 and 3 are executed correctly. |

## Step 2: Curate Your Dataset

Once you know how your data should be modelled, you can put it into practice. If you are collecting your own data: map it onto your data model. If you are starting from a dataset developed by others: clean the dataset and make relevant changes to the way it is modelled. The result of this step is that you will have a dataset that your algorithm can work with.

Grading

| | Minimal Requirements |
|---|---|
| 2 | **Your dataset needs to contain at least 50 items.** For example, if your dataset is a catalog of books, it needs to contain 50 books. If your dataset is an archive of (metadata pertaining to) newspaper articles, it needs to describe 50 articles. If your dataset is a full-text archive of academic papers, it needs to contain 50 papers. |
| 3 | **Your dataset needs to be well-formed.** For example, your `.xml` or `.csv` file(s) should not contain any syntactical errors. |

Tip

| Validate Your Dataset |
|---|
| You can use a validator (like **this one by the World Wide Web Consortium (https://www.w3schools.com/xml/xml_validator.asp)**) to check whether or not your XML is well-formed. And you can open your `.csv` file in any spreadsheet software (e.g. MS Excel, or Numbers) to check if they visualise your tables (rows / columns) correctly (to make sure you did not miss any comma's, etc.). |

## Step 3: Design Your Algorithm

Now that you have your dataset, you can start to process it. This is where your algorithm comes in. Using pseudocode, write out the different steps that are necessary to accomplish the task you have set out.

**Note:** For students who would like more of a challenge, we also allow you to write your algorithm in a full-fledged Turing-complete programming language, like python. If successful, we would of course consider this a strong demonstration of your skills and understanding of the course materials — which would have a positive effect on our determination of your

grade. That said, however, we feel the need to stress that this is in no way a requirement, and that it will be possible to obtain a perfect grade (A) for assignments where the algorithm is written in pseudocode.

Grading

| | Minimal Requirements |
|---|---|
| 4 | Your algorithm should be written in a plain **text** (`.txt`) or **markdown** (`.md`) file when writing pseudocode, or a relevant programming language specific file format (e.g. `.py` for python) |
| 5 | Your algorithm should include **30-50 lines of (pseudo)code.** |
| 6 | Your algorithm should include at least **one of each of the following**:<br>• **a variable** (defined by you)<br>• a **selection** (e.g. if-else statement) *OR* an **iteration** (e.g. while loop) |

Tip

| Flexible Workflows |
|---|
| Keep in mind that algorithm design is an iterative process, and try to maintain a flexible attitude towards your data model throughout the development process. You may have found an interesting dataset, and identified a relevant problem for your algorithm to solve (Step 1). But while you are designing your algorithm, you realise that the way your dataset is structured (Step 2) prevents your algorithm from accessing the information you need (Step 3). This may lead you to make some small changes to the model behind your dataset (Step 1), and to make changes to your dataset accordingly (Step 2), before you can continue fine-tuning your algorithm (Step 3). While it is important to anticipate the needs of your algorithm while you are conceptualising your data model, know that it is perfectly fine (and often necessary) to make changes along the way, and revisit previous steps. |

## Step 4: Design a Small Website for Your Project

Using HTML and CSS, design a small website, distributed over a few relevant hyperlinked web pages, to contextualise your algorithm. This website should contain some general information about the project/assignment, a description of the dataset, a description of the algorithm, and more information about your team, and the division of labor.

Grading

| Minimal Requirements | |
|---|---|
| 7 | Your website should include **at least four web pages**, that are linked together in a **navigation bar**. For example (but feel free to be creative here):<br>• a home page called `index.html`<br>• a page on the dataset called `dataset.html`<br>• a page on the purpose and design of the algorithm called `algorithm.html`<br>• a page on the people involved in the assignment, and the way you divided the work called `people.html` |
| 8 | Your website should contain **a motivation for the choices you made** when you were conceptualising your dataset and algorithm. This motivation can be written on a separate page, or added to other pages where relevant. (e.g. choices pertaining to the dataset on the page describing the dataset; choices pertaining to the algorithm design on the page describing the algorithm; choices pertaining to the division of labor on the page describing the contributors). |

Tip

| Sharing Personal Information |
|---|

If you would prefer not to share personal information (e.g. your name) on this website, you are free to anonymise this information **provided that you supply the teachers with a key upon submitting your assignment** (e.g.: 'Contributor 1' = John Doe). You can do this in the submission module, or by sending an email to **wout.dillen@hb.se (mailto:wout.dillen@hb.se)** .

Tip

| **Sharing Responsibilities** |
|---|
| It could be useful to designate 1 person as the 'responsible' person for each of the steps (data curation, algorithm design, website design, GitHub publication), even if you all contribute to each of the steps in some way or other. |

## Step 5: Publish Your Dataset, Algorithm, and Website on GitHub

Bundle all of the work you did for Steps 2 through 4 together in a single repository that you post on GitHub. A template for this repository will be provided during the class on GitHub. Make your website publicly accessible by publishing it using GitHub Pages. Information on how to do this will also be provided during the class on GitHub.

Grading

| | **Minimal Requirements** |
|---|---|
| 9 | Your assignment should be submitted as a **GitHub repository** |
| 10 | Your website should be published using **GitHub Pages** |

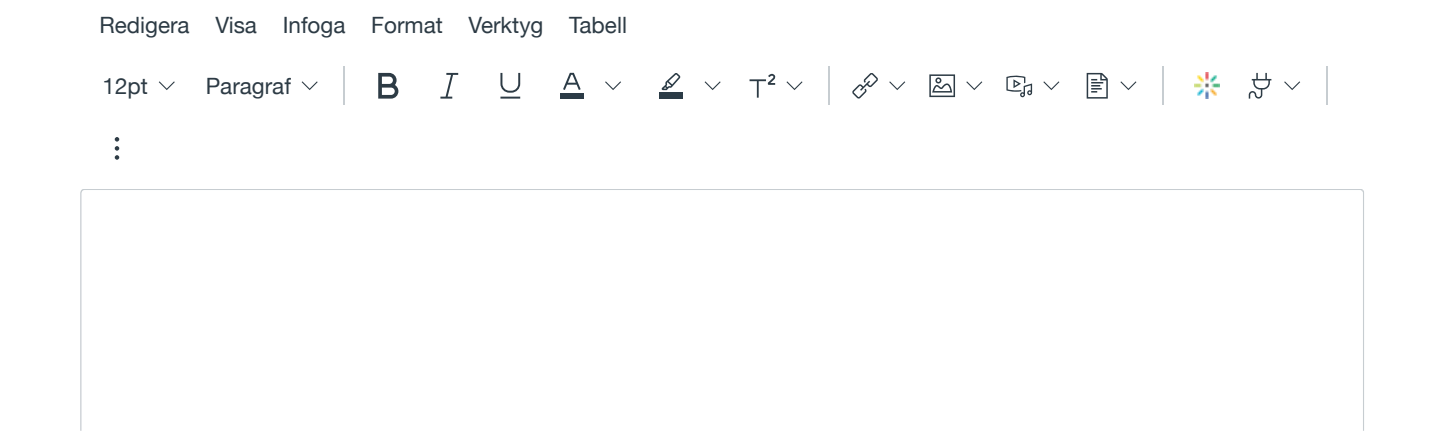
Tip

| **Collaborating in GitHub** |
|---|
| GitHub is a collaborative development environment. We would encourage you to use the functionalities offered by GitHub to your advantage here. Rather than each working on individual files, and emailing those files back and forth, you can use GitHub as the main hub for your project work, to work on the same files, collaboratively, from a distance. |

**Note:** Please keep in mind that you are expected to complete the different tasks of this assignments yourselves, without the aid of AI tools or similar. This is to help you reach the learning outcomes of this course and prepare you for future courses.

# How to submit

One member of the group should submit through this module. Post the URL to the GitHub repository in the text field. Group number and names of all group members should be included.



p

0 ord

(https://hb.instructure.com/courses/8313/modules/items/183335)

(https://hb.instructure.com/courses/8313/modules/items/1