Meeting 1 –

Idea in detail

Mean idea is that are a lot of jobs online on different website. We need to collect these jobs or develop a system to catogorise them.

But the goal is to expend it to

Catogorize skills on the jobs marketeer.

- Soft or hard skill
- Highly requested

Compare it to other programs and the HU. (Hogeschool Amsterdam, Utrecht enzv.

Mainly is it bacholer over masters studying.

Job scraping 2 different levels. Top 100 Nederland because they are the frontrunners

Look at trends compare the results.

Collect jobs – make a system that does that automatically – marketing jobs specifically – main skills for these jobs. Categorize them. Scrape them off the websites. HU, HA, HR. Bachelor over masters study.

What do you expect from our group

- o That we are gonna work in a team and be creative. Develop a tool that is sustainable and can be expanded to other schools.

Who is person in charge of this project and who is going to be our contact person.

- o Hemd. And the assistant manager: Grieta van den ven.

What would be your preferred form of communication when we run into problems? (email, Teams, Whatsapp etc)

- o Email. I fits urgent than Teams.

Where can we find the data

- o Excel file and python code. (Send by Email)

What would you consider the minimal viable product (MVP)?

- o Into a software and if its user friendly into a website. But it needs to be expandable.

Data

Are there legal or ethical guidelines we need to consider when scraping?

- GD PR guidelines. Terms of services of a website.
- Ethical guidelines of job scraping.

Which metadata from job postings are most crucial for you (title, skills, location, sector, salary, etc.)?

Netherlands (Location)

Citys or universities

Sector

Salary

Language

Experiences (upscaling) (Add on)

Dashboard

Who is the intended audience for the dashboard (policy makers, lecturers, students, researchers)?

- Education organisations
- Students
- Policy makers

Which KPIs and visualizations are most important (e.g., top 10 required skills, curriculum–market mismatch, regional trends)?

- Heat maps
- What is easiest for us
- Pick a chart that is easy with different colors.
- Skill demand

Should the dashboard be interactive, or mainly generate reports (PDF/Excel export)?

-

How frequently should the data be updated (real-time, daily, monthly)?

- Just yearly

o   If there is a spike if its possible to make a alert out of it.

Constraints & Maintenance

- How should the solution be scalable for the future?
    - The big 3 schools and in the Netherlands.
    - 5 year program. This year only Netherlands hogescholen

- Who will be responsible for maintaining the system after delivery?
    - Lecturaat, Hemd.
    - ICT kan het overnemen.

- Are there budget or time limitations we should be aware of?
    - Limited budget.
    - Het is mogelijk om bij it te vragen om te helpen. Als het geld kost..

Can we feed this website or system. Can we predict the 4 scenarios that can happen. Or do we need to do our research before and put the input in the website or system.

What are the tools that we planning to use. (For now python)

Meeting 9-10

We have 3 programs to try within marketing.
But first, choose one bachelor's program (not a master's) to focus your study on.

*Use the keyword matching method from the old Python file:
competency.analysis.py
*sub-analyses of the big data about the top 100 companies.
*document everything

Step 1: Data Collection
We will collect data from job websites.
It's easier to start collecting data from HU (Hogeschool Utrecht).

Step 2: Tasks to Do
Create a list of matching keywords to feed into the algorithm that pulls job data.

The broader and more flexible your keyword list is, the more job results you'll retrieve.

Document everything carefully.

Make sure all data collection is done legally.

Step 3: Filters for Job Data

We need to set the following filters, for example:

Experience level (e.g., entry, mid, senior)

Posting date — select "any time" so we get all available jobs, not only those posted in the last week.

Step 4: Sub-Analyses

Conduct sub-analyses of the big data focusing on the top 100 companies.

Step 5: How to Start

Get the curriculum of your chosen bachelor's program.

Extract the keywords (skills, subjects, competencies).

Make a list of competencies.

When analyzing job descriptions — for example, "social media marketing" — look for matching skills that appear in multiple job postings. This helps refine the keyword matching.

Also check the "What are you bringing?" section in job ads for more keyword matches.

Step 6: Questions to Analyze

How many times is years of experience mentioned?

How often is Dutch language required?

Compare these findings with the curriculum and course outcomes.

Step 7: End Result

Identify which competencies match and which do not.

It's not our responsibility to change the school curriculum.

Our job is to collect the data and present the results.

The matching competencies that align well can be shared on the website that we are building

Hi, Just to confirm these are the data mining goals after your feedback. Do you agree with the data mining goals?:

1. The automated data collection and processing pipeline will ensure that all relevant data is consistently prepared and available for mining tasks with at least 90% data completeness and no more than 6% duplicate entries. This automated pipeline not only reduces manual intervention but also ensures high data integrity, consistency, and readiness for downstream analytical and machine learning tasks.

2. An automated web scraping system will continuously extract and standardize job posting data, achieving at least 90% extraction accuracy and 90% temporal consistency across weekly samples, to provide a robust and reliable foundation for mining job–skill relationships.

3. A working matching algorithm will be developed that correctly matches job postings to relevant competenties taught by HU programs based on required versus learned skills, achieving a precision score of at least 90%.

👍

Laatst gelezen

Vandaag

Hend Elsayed 09:58

good morning Fatima
yes that is correct now

10:00

good morning, Thanks! :)

Typ een bericht