ERASMUS+: Employability in Programme Development (EPD) Project

Erasmus+ Programme, Key Action 2: Strategic Partnerships (agreement number 2020-1-UK01-KA203-079171)

Intellectual Output: IO3

Intellectual Output title: Prototype Employability Dashboard

Document title: Report setting out the scope and functionality description for the prototype and graphical mock up

Contributing authors: [||]Johan Loeckx, [||]Corneliu Cofaru

Author affiliation: [||]Vrije Universiteit Brussel (VUB)

Co-funded by the Erasmus+ Programme of the European Union

Report setting out the scope and functionality description for the prototype and graphical mock up

Johan Loeckx, Corneliu Cofaru
Artificial Intelligence Laboratory
Vrije Universiteit Brussel

Introduction

One of the main outputs within the EPD project was the development of so-called "employability dashboards". The concrete scope and functionality of these dashboards was intentionally undefined in the proposal, as it needed to reflect real needs identified by stakeholders in the hackathons.
These hackathons – however difficult to implement and execute – indeed led to crucial insights. They exposed the enormous amount of work still to be done, before we can reach the holy grail of integrated dashboards that connect the different actors in employability in real-time. More importantly, they revealed two "first steps" on the "employability ladder". These will be the subject of the prototypes discussed below.

1. Prototype 1 – Insight engine to explore employability literature

1.1. Context

One of the most important learnings from the first hackathon was that the gap between 1/ employers and curriculum designers, on the one hand, and 2/ AI system implementors and people active in the field of employability was wider than expected.

As a consequence, the original aspiration of connecting the datasets relevant to students, lecturers, curriculum designers, career professionals, employability officers and employers was deemed too difficult to pull off. The main reason was that the organizational and legal hurdles to get access to the data were too high. Furthermore, the preparatory work to extract sensible information from the data sets requires high investment from the stakeholders, with an uncertain return.

1.2. Rationale

However, the need for and enthusiasm from the partners to better understand each other's business became very clear. As such, we identified another way to connect these different actors.
During the many discussions between the partners of the EPD project, as well as during the focus groups, multiplier events and other dissemination activities, we noticed that the gap does not only exist at a data level, but also at an epistemological level. One concrete example is the observation that **literature is highly fragmented**.

1.3. Impact we wish to achieve

The ultimate ambition of prototype #1 is to **increase the cohesion and mutual understanding** between the different actors within the employability spectrum.

1.4. High-level (AI-based) approach

We propose a so-called **"insight engine"** called LEDA, developed at the Vrije Universiteit Brussel, to make the best and newest **employability-related documents discoverable.** An "insight engine" is a kind of combination between a web search engine like Bing or Google, and structured search within databases.

A crucial aspect of a well-functioning engine is its ability **to bridge different domains.** Often, an end-user is an expert in his/her own domain (e.g., teaching, hiring or curriculum design), but is not familiar with another (like employability).

The role and challenge of an insight engine is to provide relevant documents in domain Y (employability) based on answers to questions in domain X (the domain that is mastered by the end user).

1.5. Target audience

The primary target audiences are **professionals** involved in **employability** of **higher education students**:

1. Teachers
2. Curriculum designers
3. Career services
4. Quality assurance agencies
5. Employers
6. Researchers

In the coming paragraphs, “end-users” will refer to individuals from all above audiences.

1.6. Objectives

The goals of the first employability dashboard are to increase (for the end-users):

1. the **“breadth”** of documents consulted;
2. the uptake of **up-to-date** information;
3. the **relevance** of provided information;
4. the **productivity** of finding this information.

1.7. Description of the prototype

Mock-ups

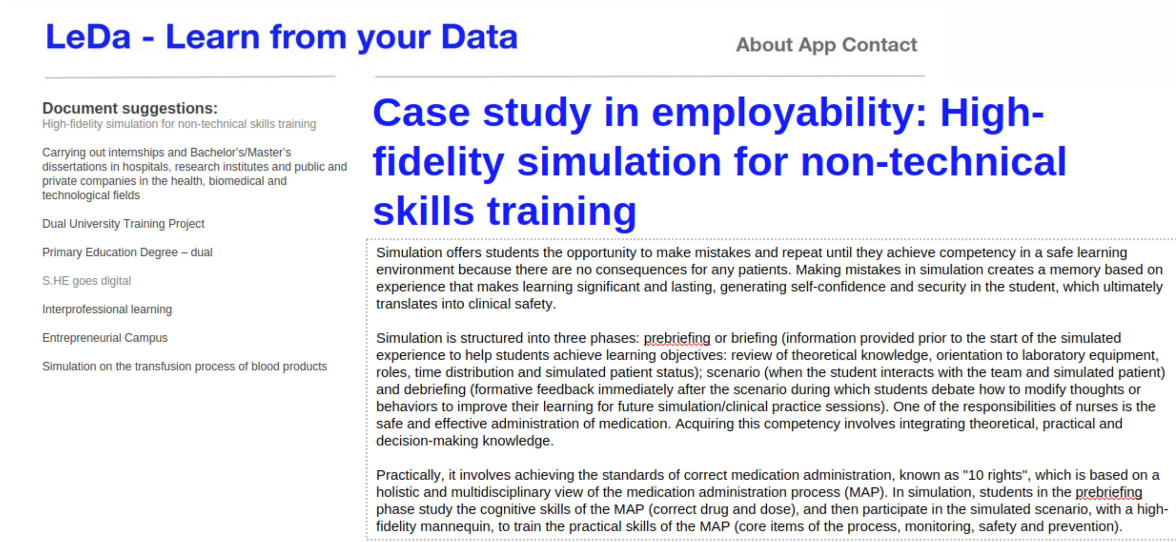
The functionality of the application is most easily explained by so-called “mock-ups”, design templates that will serve as guidelines for the real software development. In principle, the software is “responsive” which means that it can be consulted both on smartphone (not recommended), tablet or computer.

1. When launching the app, the user is informed that (s)he needs to answer **three compulsory questions**. Later, the user will be able to change the answers.

   As the user is required to answer the questions, a so-called “modal” dialogue is displayed that grabs the attention and cannot be escaped.

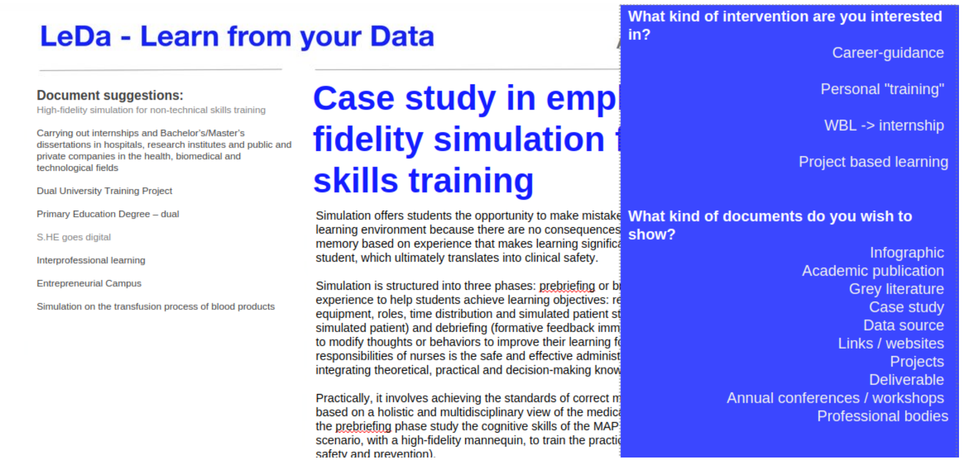
   Optionally, one can remember the user’s answers and give him/her the possibility to move fast forward.

2. Once these three questions have been answered, a first ranking of relevant documents is shown on the left, with the document itself on the right. Documents that were already visited, are greyed out.



3. On the bottom of the screen, one can revisit important past pages, either bookmarks, history or related documents identified by an AI recommendation system.

4. At any time, the end-user can update its search criteria by pulling a sidebar from the right:



Questions to be asked

There will be three compulsory questions:

1. The **kind of documents** one wants to retrieve: *[sector]*
    a. Case studies in employability (selected in IO4 of the EPD project)
    b. Deliverables from the project
    c. Grey literature
    d. Data sources
    e. Infographic
    f. Academic publications (identified in the project)
    g. Links / websites
    h. Projects (E+ but also others)
    i. Annual conferences / workshops
    j. Professional bodies
2. **Type of intervention** (how to integrate into the curriculum) *[role]*
    a. Career-guidance
    b. Personal "training"
    c. Work Based Learning (like internships)
    d. Project based learning
    e. Curriculum design
3. **Granularity** of (curriculum) intervention *[data_types]*
    a. Regional / national / city
    b. Institution
    c. Classroom

Next, there are optional questions:

4. Year of publication
5. Country
6. Sector (folksonomy)
7. Student target group

Finally, the user can specify some open-ended, unstructured text query to which the content of the documents is matched.

Artificial intelligence core

At the heart of the dashboard, we place a so-called **“learning-to-rank" algorithm** whose purpose is to learn to rank existing documents as to optimize a particular criterion like “engagement” with the documents (e.g., clicking and opening a document).

It uses machine learning techniques to predict the expected chance of opening a document, based on the profile of the user and previous interactions with the system. Though this seems easy, it is far from straightforward. At a basic level, learning could mean prioritizing documents of a specific type like data sources because the user showed this preference before. At a more advanced level, it also should attempt to learn preferences based on the content of the documents, trade-offs between different preferences (for example, newer information is more important than the type of intervention). Finally, one hopes the system to learn *across* users: two social scientists in the same research domain can be expected to have similar preferences.

It is important to note that these techniques need a lot of finetuning, careful engineering and sufficient data to get them working. This knowledge often results from real usage, and this data will not be available at the start, and maybe never will be (a phenomenon called “cold start”).

2. Prototype 2: Prompt-based Analytics Dashboard to query and connect graduate outcomes and employers’ surveys

2.1. Context

Employees in professional organizations like employability offices, universities or quality insurance institutions all take actions that are based on the best understanding of domain experts. New university programmes are developed based on (real or perceived) needs and skill gaps, specific audiences are targeted because they are believed to be disadvantaged, new initiatives are supported by policy makers based on the assumed added value they bring to employers, etc.

This knowledge comes from their own experience, of course, but also from interaction with peers, (textual) reports, and data sets that are collected by different parties, at different levels of granularity.

2.2. Rationale

We will focus on this last source of intelligence, because they are best suited for machine consumption by computerized techniques (algorithms). At the same time, they are also the most difficult to interpret by experts who have good knowledge of the domain of employability, but do not necessarily have the necessary technical skills to process this data and extract robust evidence to support or discourage a particular policy or practice.

2.3. Impact we wish to achieve

We aim to lower the barrier for employability experts to extract robust evidence from Labour Market Intelligence data sources.

2.4. High-level (AI-based) approach

We propose a **knowledge-driven prompt-based system** that a (trained) domain expert can interact with using a formal language to semi-automate.

During a specific hackathon

Examples of questions

2.5. Target audience

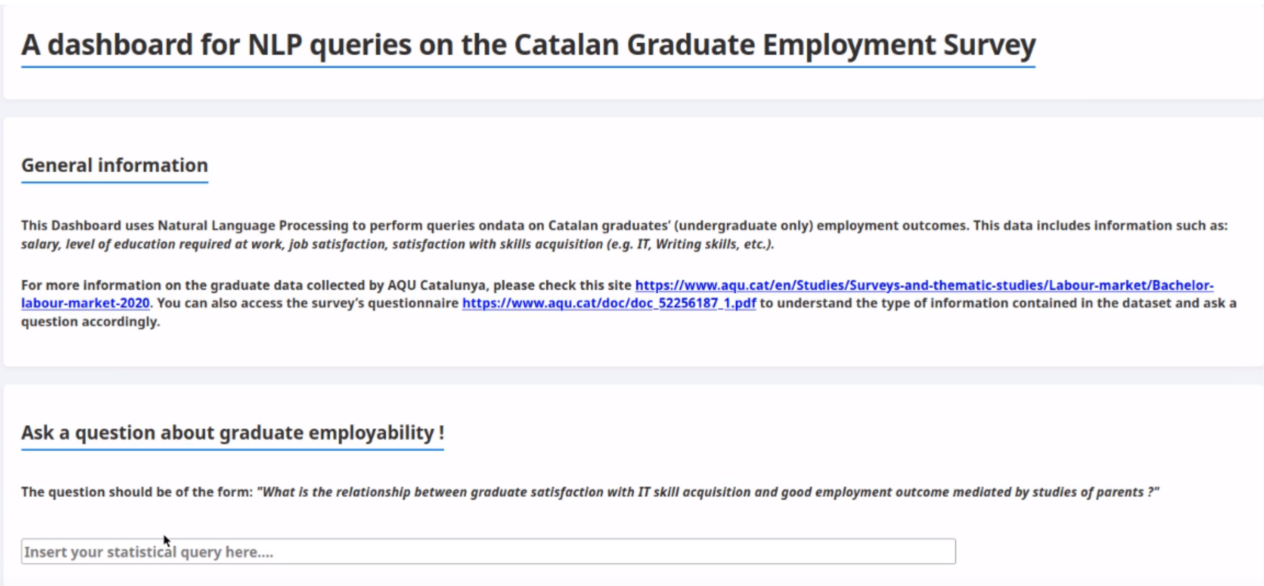
Employability experts with a good knowledge of statistics and/or data analysis, and very basic programming skills.

Within the EPD project, we chose to collaborate with a project partner, because creating such a dashboard requires significant effort and commitment, which cannot be reasonably expected from external partners. Furthermore, there are legal hurdles (data sharing, GDPR, NDAs) to be taken that are easier solved within the consortium as there is sufficient trust.

2.6. Description of the prototype

The prototype consists of a "prompt engine" tailored to employability, and more specifically, to the datasets of a partner, in this case AQU.

The user can ask questions to the database in natural language.



The engine then parses this text and interprets the vocabulary and grammar in function of the defined data sources. For example, "IT skills" is translated to one variable or a set of variables in the data set.

Question parsing results...

```
Parsed query: [
  [
    "what is",
    "UNK"
  ],
  [
    "IT skill",
    "INDEP_VARS"
  ],
  [
    "acquisition and good",
    "UNK"
  ],
  [
    "mediated",
    "MEDIATED"
  ],
  [
    "employment outcome",
    "DEP_VARS"
  ],
```

The “autoML” engine that is employed, converts this natural language query into an intermediate structured data source. This is then parsed by an existing framework, implemented by VUB in Julia, and soon to be open sourced.   It is **knowledge-driven** (i.e. not using a large language model or anything like it) so that the results can be **audited and explained for optimal transparency.**

Finally, the results are displayed.

AutoML progress...

Start AutoML!

```
•-[root]
└ •-[LoadCSV]
  └ •-[ExtractTargetsOnly]
    └ •-[DirectFeatureSelection=Union{Nothing, Int64}[143, 144, 150, 145, 146, 147, 39, 149, 43, 141, 140, 130, 94, 95, 88, 93, 2, 54, 151]]
      └ •-[GroupByFeaturesDirect=["ngestio_10", "nivest"]]
```

Analysis results