RESOLVED DATA-CHALLENGE

BACKGROUND

Below you will find a typical data-challenge that we deal with at Resolved on a daily basis. There is no right or wrong answer to this problem, it is an open-ended challenge. Also, this is not a test. This challenge is meant to ensure you get a good flavour of the types of data problems we handle at Resolved, and to get your hands on the toolset we apply daily.

THE CHALLENGE

When organisations want to recruit for a position, they typically list a set of required skills. Some companies use competency models for this, however, nowadays most companies either use a variety of competency models, while others use phrasings entirely in the company language. For analytical approaches to recruitment, this poses difficulties, for instance, to compare an applicant across different positions when the wording of the skills slightly differ. Is this the same skill or not?

As a first step to overcome some of these difficulties, we could let the machine help to build skill sets in a single, standard frame: a competency model. Competency models have the advantage to act as a well-understood standard, and good models are linked to underlying psychometric properties. However, their rigidity is part of the reason that they are not used as often by organisations, or at least not in the same way. If we could allow any organisation to frame skills in their own words, but in the background, translate this into a competency model, we could do better analytics without forcing the organisation into a rigid structure.

THE DATA

For data, we will provide you with a competency model (competency model 2 dimensional.xlsx). We will also give you some examples of actual skill profiles for real jobs (jobprofiles.xlsx). Both data-sets can be downloaded for the purpose of this challenge from google drive.

Competency model 2 dimensional.xlsx contains 1 sheet with data. The data columns are:
- Competence: The label for the competency involved
- Skilled: Looks like: a behavioural description for someone highly skilled in this competency
- Unskilled looks like: a behavioural description for someone not skilled in this competency

jobprofiles.xlsx contains 1 sheet with data. The data columns are:
- Skill_id: unique identifier for the skill
- Crit: a criticality factor for that skill within the job
- Weight: a weight of importance within the skills of the same criticality for the job
- Vacancy_id: what vacancy does this skill belong to
- Type: how this skill will be scored, BI=binary, DI=dichotomous, CO=continuous
- Skillname: This is the skill definition in the user's term

- Related: this is the competency this skill relates to. Currently, these have been filled in manually by an expert judge. But, these use a different competency model!! Nevertheless, they may provide you with clues on what would be correct.

First, please be aware, to keep things realistic, that these are **uncleaned**, unprocessed data-sets, in atypical formats. You will have to deal with that yourself.

Second, clearly these data-sets are not **big**. Likely, they are even insufficient to solve the task at hand to a satisfactory level of accuracy. Again, this is real-life. It is up to you to find a solution to the sparsity of the data. There are many available options.

THE TASK

Given the problem, the data, data-cleaning and the sparsity of the data, can you find a way to build a classifier that matches any written skill description by a recruiter to the competency model provided. That is, for instance, if I type: 'can lead a team to successes', the classifier suggests back to me **directing**. Or **directing & coaching**.

EXPECTATIONS

We would like you to implement your solution in a web-based application that we can test. As a platform, you are free to choose between either a small django application or a shiny application. We do not expect you to provide an actual online version, but a code base that should run out-of-the-box on a local machine.

Do not spend too much time on the interface, an input field that accepts a long string, and some output of the classifier (textual or graphical, your choice) will be fine.

We are aware that this is not a 1 week project. So we do not expect industry ready solutions. What we do expect is to see an understanding of the problem, ideas for a solution, and some tests of how that solution might look like and work. That is, if you come up with a solution that works well on the given data, but not beyond, than make sure that you include input checks. Remember, the input part is a huge space (a subset of all possible sentences in any given language). The competency part is extremely limited. So while we do not expect you to 'solve' this problem, we do expect your solution to work from a client point of view. Assume the client will not read your limitations, but will try to break your solution. So from a quality and user point of view, built in restrictions where you need them. It is fine to have them, but make sure you check.

It is perfectly ok to come up with a solution that rephrases the problem statement. If you find a clever way to keep the open input, but are able to cross-compare applicants and jobs on similarity, please do so. However, do note that we include a competency model as a standard in the background to be able to link to psychometric tests. We do not include this link here, but if you find anything on that and exclude competencies, by all means feel free to do so. Again, it is an open ended problem and solution space.

The language used in this challenge is English, and English only. So no need to think about multiple languages at this point.

ADDITIONAL INFORMATION

Think about the problem at hand carefully. There are many forms of a solution. Do not only think about the technical side of the matter, but also take in mind the client needs (as outlined above) and the economic viability of the matter at hand. Some solutions will require human effort combined with machine learning, these sort of solutions are acceptable, provided the above factors are reasonable.

On the sparsity of the data, it is ok to think of this as an ongoing learning system that starts of with an initial set of parameters, and learns from user feedback. This would be an ideal scenario, but not mandatory.

On the skill profiles, please note that some skills, like degrees, or languages, or programming languages and so on or not in the competency model. Competency models typically only include so-called soft-skills. How you deal with that is up to you.

Some skill descriptions will contain a combination of two or more competencies, like the example given above. We leave it up to you to deal with this.

Make sure to approach the problem from all sides, that includes psychology. There is an extensive literature to be found around competency modelling which might contain insights to tackle this issue.

Finally, bear in mind, that once we can translate skill descriptions to competencies, we would like to use the underlying competency model to compare jobs, for instance for similarity. As you understand, if skills are related to more than a single competency, we will need to be able to take this into account.