Ulisse

Soft Skills for Employability

IO3 - REPORT

Student CV analysis

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

ULISSE is a Strategic Partnership for Higher Education project (2018-1-IT02-KA203-048286). The European Commission support for the production of this website does not constitute an endorsement of the contents which reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Intellectual Output 3 – Student CV analysis

Authors:

Donata Gabelloni, Riccardo Apreda, Tommaso Pavanelo, Giovanni De Santis, Dario Brugnoli, Andrea Mazzoni, Giovanni De Santis (ERRE QUADRO), Antonella Magliocchi, Gualtiero Fantoni, Filippo Chiarello, Rossano Massai, Chiara Pasca, Mariangela Barbarito, Alessandro Guadagni (UNIPI)

Contributors:

Domingo Galiana Lapera, Dolores Lopez Martinez, Abel Torrecillas Moreno, José Juan López Espín, María José López Sánchez (UMH); Anda Paegle, Lasma Saimena (LU), Manuel Salvador Araújo, Isabel Ardions, Paula Carvalho, Diana Aguiar Vieira, Viviana Meirinhos (PPORTO).

Statement of originality:

This deliverable contains original unpublished work, except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Disclaimer

INTRODUCTION

The aim of the activity described in the present report is to identify how students/graduates describe and self-assess their own soft skills in their CVs. This process has been carried out using Natural Language Processing tools supported by the lexicon of soft skills previously developed in the project (IO1). The dataset is composed of a sample of 4479 student CVs gathered by the partner Universities.

Based on the results obtained from the employers' survey conducted in the IO2, together with a revisited literature review, the final soft skills list is composed of 17 soft skills. However, a remark should be done for the skill "Orientation to results". This skill was not mentioned in the IO1 ("The Soft Skills Lexicon") and emerged only later, when conducting the employers' survey. This means that it is not possible to find this skill in "The Soft Skills Lexicon". As a result, the CV analysis process has been conducted using 16 Soft Skills.

This analysis of students' CV belongs to IO3, whose ultimate aim is to conduct a skill gap analysis, comparing the educational offer of the partner Universities with the needs expressed by the employers (IO2) and the self-assessment of the students, as declared in their CVs. The output of IO3 will be the ULISSE training course for soft skills.

METHODOLOGY

The goal of this methodology is to detect and extract textual expressions of soft skills from student CVs. This has been done using Natural Language Processing (NLP) together with a lexicon of soft skills (i.e. the expression of soft skills collected in the *Soft Skills Lexicon - IO1*).

We can divide our process into three macro phases: (1) CV Collection, (2) Preprocessing, and (3) Skill Extraction.

Figure 1 – Student CVs analysis

CV Collection

During this first phase, each partner collected a set of CVs belonging to students and graduates. Each CV was anonymized in compliance with privacy laws and respecting sensible information, by removing any personal data. Only information about students career (or professional experience), studies (or academic experience) and skills were present on the CVs.

Preprocessing

The CVs were pre-processed in order to store them in a constituent way. In fact, student CVs can be written using different formats and languages: this makes impossible to make a comparison and a statistical analysis of their content. Even though the most common format is the *Europass*, each university and each student are free to choose their own format for their CVs. Furthermore, since some of the CVs are provided in the native language of students, it has been necessary to translate the documents into English.

For these reasons the following pre-processing steps have been carried out:

- *Free Text Extraction*: All the meta information contained in the CVs (tables structure, page layout etc.) were removed and only the text was considered for the analysis
- *Translation*: CVs were collected by partners from their countries (Italy, Portugal, Spain, Latvia) in their own National languages. For this reason, they were all translated automatically into English
- *Sentence splitting:* The text was split in sentences
- *Tokenization:* The text was split in single words (also called “tokens”)
- *Lemmatization*: For each word, its root form was computed (e.g. men → man, played → play) in order to make it easier for the phase of skill extraction to detect identical soft skills that were written in different forms (e.g. problem solving → problems solving)

The output of this phase is a structured dataset of CVs belonging to the four countries, ready to be analyzed by the NLP tools in order to detect the Soft Skills according to the 16 labels identified in the Soft Skill Lexicon, as revised after IO2.

Skill Extraction

The aim of this phase is to extract those sentences containing a soft skill. The process was performed automatically thanks to the use of the Technimeter® and to the lexicon of soft skills collected in the IO1. Following its specific algorithm, Technimeter® is able to extract relevant entities according to the definition (initial list of expressions) of these entities. In this case, this definition is represented by the soft skills lexicon.

After the soft skills were identified in CVs, we counted every time that a student mentions a specific soft skill. Furthermore, thanks to the preprocessing phase and to the versatility of the Technimeter®, the method is able to reconduct different expression of the same soft skill to a unique label (decided by the partners and encoded in the Soft Skills Lexicon). For example, if a student mentions “proven ability to solve problems”, the Technimeter® reports the presence of the soft skill “Problem Solving”.

RESULTS

The set of documents collected by the partners is composed of a total of 4479 CVs distributed across the four countries involved in the project. The total number of CVs per country is shown in *Figure 1*.

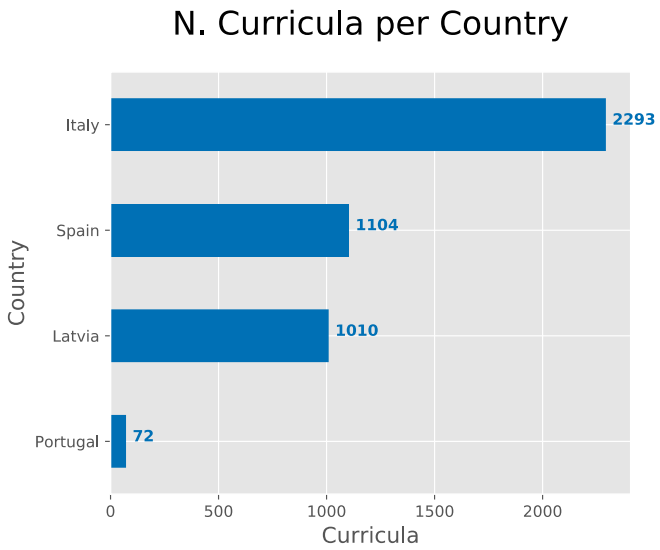

Figure 1 - Number of CVs processed per Country

The soft skills extraction phase gave as a result a total number of 1731 soft skills, of which 90 unique soft skills, without counting the repetitions. These 90 skills (wordings) are referred to all the 16 soft skills (labels) contained in the Soft Skills Lexicon.

Figure 2 shows the Soft Skills (as expressed by students) ranked by the number of times each one appears in the CVs analyzed; similarly, Figure 3 shows the same count for the Soft Skills Labels.

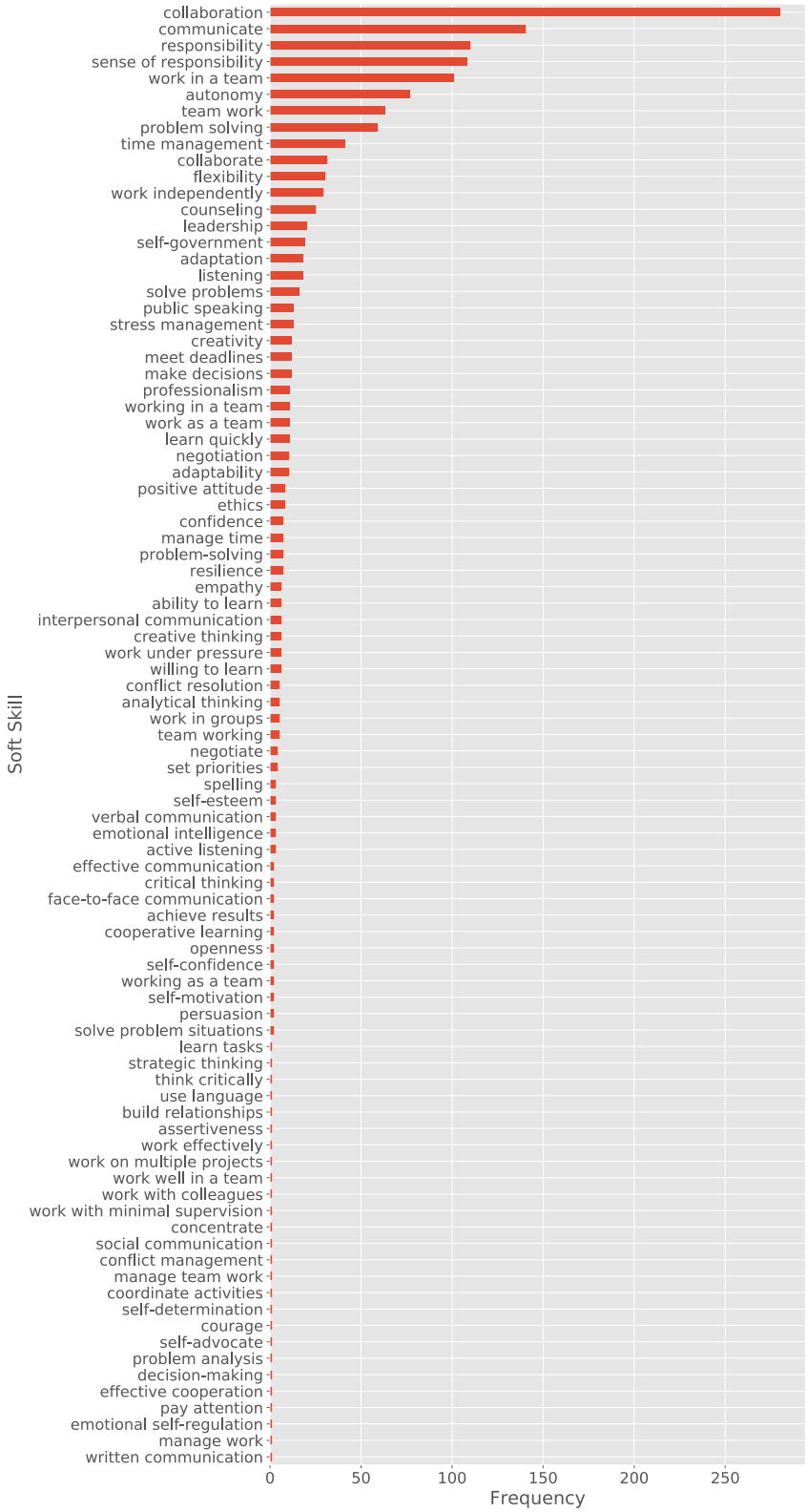

Figure 2 – Ranking of the 1731 Soft Skills (wordings) appearing in the CVs analyzed

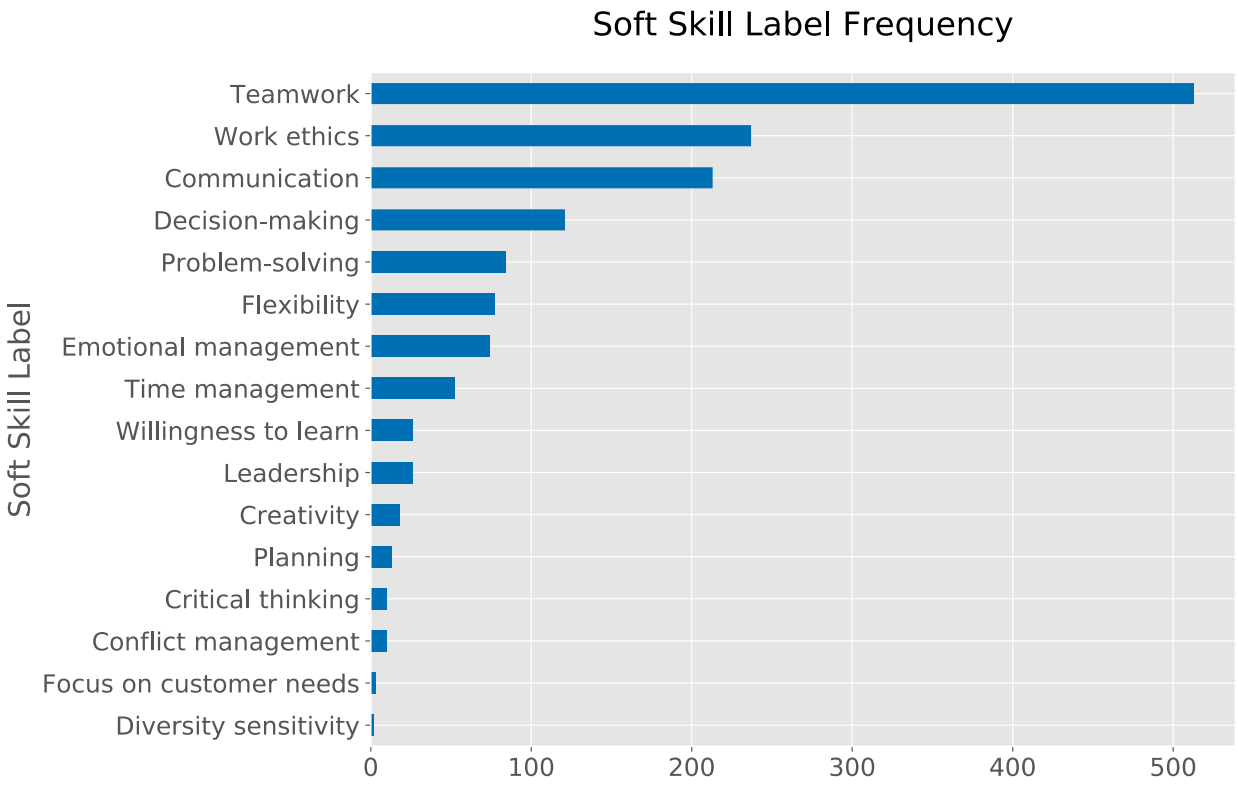

Figure 3 - Ranking of 1731 Soft Skills (labels) appearing in the CVs analyzed

Distribution among the countries

Further analysis can be done studying the differences among countries. Figure 4 shows the percentage of Soft Skills per Country. For example, the value of *75%* for Portugal means that 75% of the Portuguese CVs contain at least one soft skill.

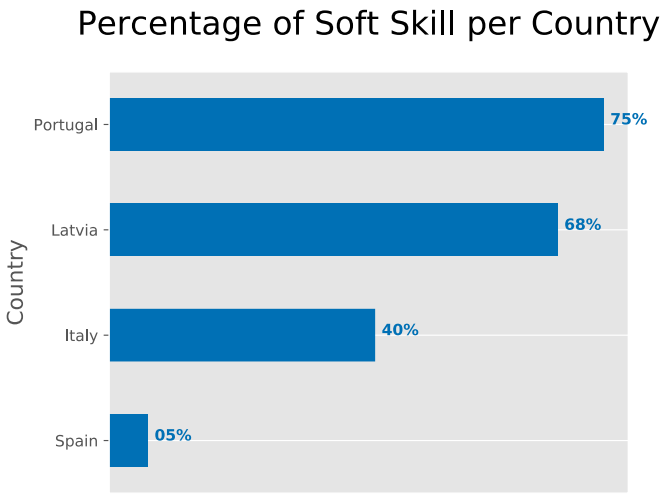

Figure 4 - Percentage of Soft Skills per Country

Figure 5 shows the percentage of unique soft skills per Country (i.e. excluding the repetitions of the skills). For example, the meaning of Latvia's 86% value is that the students mentioned in their CVs the 86% of the unique soft skills present in the Lexicon.

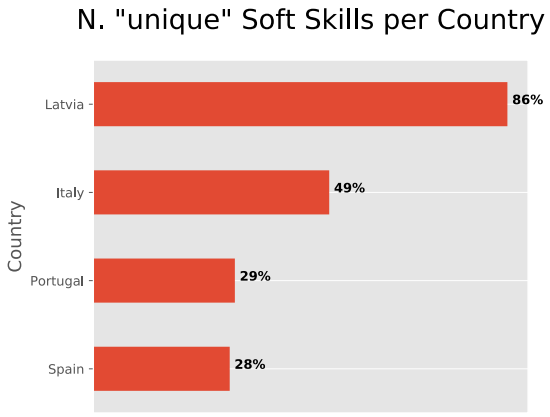

Figure 5 - Distribution of "unique" Soft Skills among the countries

Figure 6 shows the percentage of Soft Skill labels in each country. For example, the meaning of Latvia's 100% value is that the Latvian students mentioned in their CVs all the 16 soft skills labels.

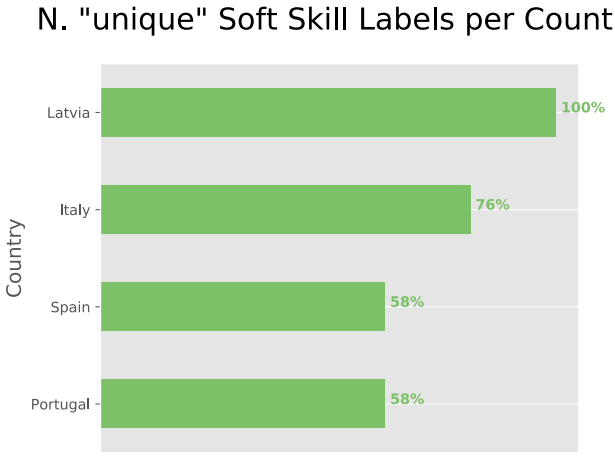

Figure 6 - Distribution of Soft Skills Labels among the countries

Finally, figure 7 shows the top 5 Soft Skills and the top 5 Soft Skill Labels per Country.

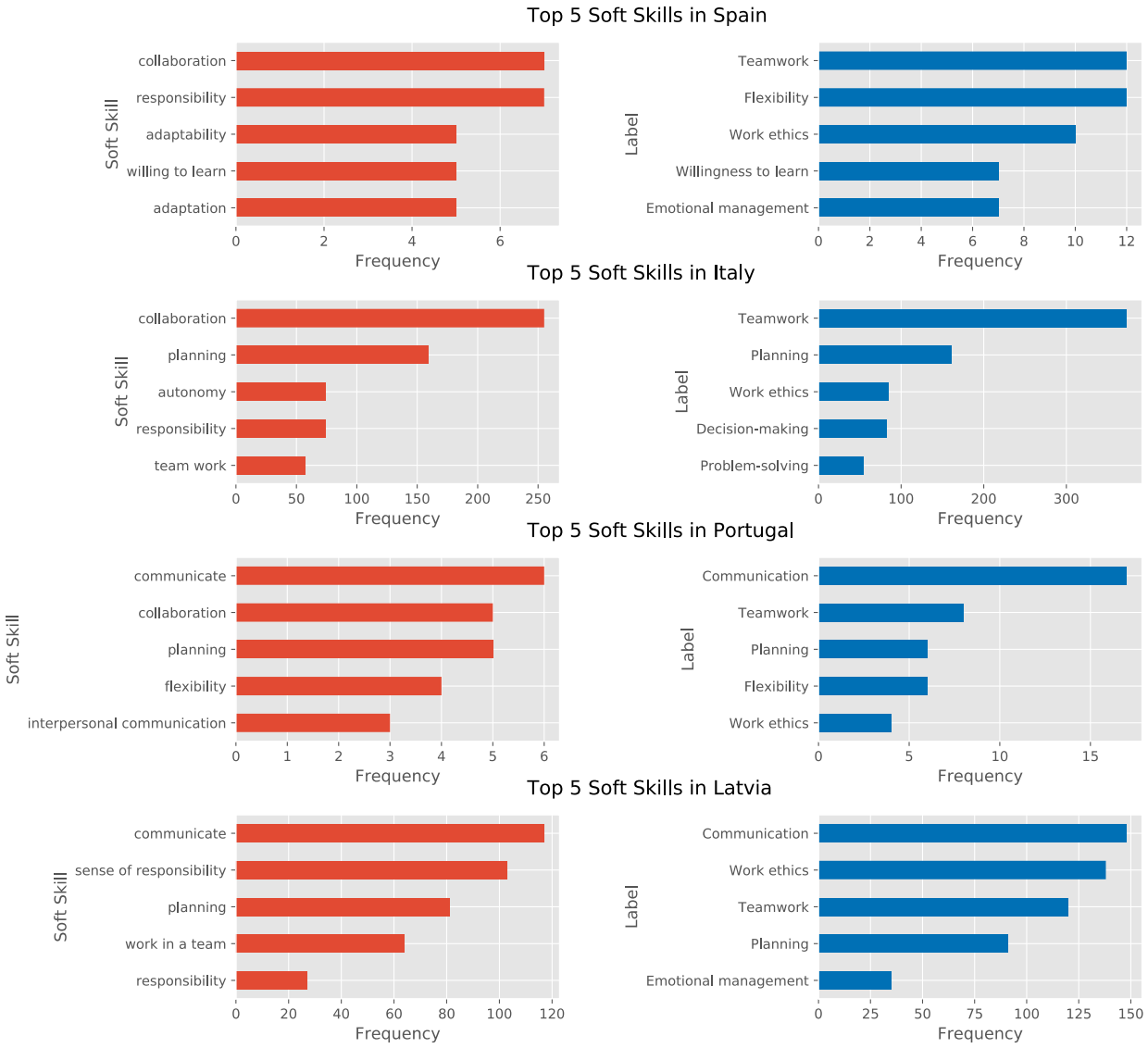

Figure 7 - Top 5 Soft Skills per Country