



**NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS**

**DEPARTMENT OF INFORMATICS AND TELECOMMUNICATIONS  
INTERDISCIPLINARY PROGRAM OF POSTGRADUATE STUDIES IN LANGUAGE  
TECHNOLOGY**

**MASTER THESIS**

**The Greek Labour Market And The European System Of  
Classification Of Skills And Occupations ESCO:  
A First Approach.**

**Maria C. Pateli**

**Supervisor: Dr. Papageorgiou Haris**, Research Director at the Institute for Language  
and Speech Processing

**ATHENS**

**JULY 2022**



**ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ**

**ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ**

**ΔΙΔΡΥΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ ΣΤΗΝ ΓΛΩΣΣΙΚΗ  
ΤΕΧΝΟΛΟΓΙΑ**

**ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

**Η Ελληνική Αγορά Εργασίας Και Το Ευρωπαϊκό Σύστημα  
Ταξινόμησης Δεξιοτήτων Και Επαγγελμάτων (ESCO): Μια  
Πρώτη Προσέγγιση.**

**Μαρία Χ. Πατέλη**

**Επιβλέπων: Δρ. Παπαγεωργίου Χάρης**, Ερευνητής Α' Ινστιτούτο Επεξεργασίας  
Λόγου

**ΑΘΗΝΑ**

**ΙΟΥΛΙΟΣ 2022**

**MASTER THESIS**

The Greek Labour Market and The European System Of Classification Of Skills And  
Occupations ESCO:  
A First Approach.

**Maria C. Pateli**

**SN:** Lt1200029

**SUPERVISOR:** Dr. Papageorgiou Haris, Research Director at the Institute for  
Language and Speech Processing

**THESIS COMMITTEE:** Markantonatou Stella, Researcher  
Dr. Papageorgiou Haris, Research Director at the Institute for  
Language and Speech Processing  
Papavasiliou Vasilis, Researcher

## **ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ**

Η Ελληνική Αγορά Εργασίας Και Το Ευρωπαϊκό Σύστημα Ταξινόμησης Δεξιοτήτων Και  
Επαγγελμάτων (ESCO): Μια Πρώτη Προσέγγιση.

**Μαρία Χ. Πατέλη**

**ΑΜ:** LT1200029

**ΕΠΙΒΛΕΠΩΝ:** : Δρ. Παπαγεωργίου Χάρης, Ερευνητής Α' Ινστιτούτο Επεξεργασίας  
Λόγου

**ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ:** Μαρκαντωνάτου Στέλλα, Ερευνήτρια Α'  
Παπαγεωργίου Χάρης, Ερευνητής Α'  
Παπαβασιλίου Βασίλης, Ερευνητής

## **ABSTRACT**

The lack of standardisation in the linguistic description of jobs and the skills of candidates makes it difficult to link them properly. In this context, the European system of classification of skills/competences and occupations (ESCO) provides a wealth of information on occupations, but also on the skills/competences, qualifications and knowledge that these occupations require. This information is uniformly coded for 27 languages and organised in an ontological knowledge base. This work aims at creating a mechanism for the processing and automatic display of the linguistic descriptions of jobs in the ESCO ontology, facilitating the interconnection of the Greek labour market with European occupational standards.

**SUBJECT AREA:** Semantic Text Matching

**KEYWORDS:** Natural Language Processing, Semantics, Concept Based Mining, Text Similarity

## ΠΕΡΙΛΗΨΗ

Η έλλειψη τυποποίησης στη γλωσσική περιγραφή των θέσεων εργασίας και των δεξιοτήτων των υποψηφίων δυσχεραίνει τη σωστή σύνδεσή τους. Στο πλαίσιο αυτό, το ευρωπαϊκό σύστημα ταξινόμησης δεξιοτήτων/ικανοτήτων και επαγγελμάτων (ESCO) παρέχει πληθώρα πληροφοριών για τα επαγγέλματα, αλλά και για τις δεξιότητες/ικανότητες, τα προσόντα και τις γνώσεις που απαιτούν τα επαγγέλματα αυτά. Οι πληροφορίες αυτές είναι ομοιόμορφα κωδικοποιημένες για 27 γλώσσες και οργανωμένες σε μια οντολογική βάση γνώσεων. Η παρούσα εργασία στοχεύει στη δημιουργία ενός μηχανισμού επεξεργασίας και αυτόματης προβολής των γλωσσικών περιγραφών των επαγγελμάτων στην οντολογία ESCO, διευκολύνοντας τη διασύνδεση της ελληνικής αγοράς εργασίας με τα ευρωπαϊκά επαγγελματικά πρότυπα.

**ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ:** Σημασιολογική Αντιστοίχιση Κειμένου

**ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ:** Επεξεργασία Φυσικής Γλώσσας, Σημασιολογία, Εξόρυξη Με Βάση Έννοιες, Ομοιότητα Κειμένου

# CONTENTS

<b>Abstract</b> .....	<b>5</b>
<b>1 ONTOLOGY OF PROFESSIONS</b> .....	<b>8</b>
<b>1.1 Issues To Be Explored</b> .....	<b>8</b>
1.1.1 Advertisements Focusing On Social And Emotional Skills .....	8
1.1.2 Advertisements With Discriminatory Practices Against Individuals Or Groups .....	9
1.1.3 Incomplete Advertisements .....	9
<b>1.2 Thesis Objective</b> .....	<b>10</b>
<b>2 ESCO</b> .....	<b>11</b>
<b>2.1 Introduction To Esco</b> .....	<b>11</b>
2.1.1 Concepts And Terms .....	11
<b>2.2 The Occupations Pillar (The Occupations Pillar)</b> .....	<b>12</b>
<b>2.3 The Knowledge, Skills And Competences Pillar</b> .....	<b>14</b>
<b>2.4 Why Is Esco Useful?</b> .....	<b>17</b>
<b>2.5 Esco Database</b> .....	<b>17</b>
<b>2.6 Esco Db Blueprint</b> .....	<b>17</b>
<b>3 COLLECTION OF TEXTUAL JOB ADVERTISEMENTS (TJAD)</b> .....	<b>19</b>
<b>3.1 Greek Market</b> .....	<b>19</b>
<b>3.2 Data Collection</b> .....	<b>19</b>
<b>3.3 Data Organization</b> .....	<b>19</b>
<b>3.4 Problems During The Procedure</b> .....	<b>20</b>
<b>4 AUTOMATED MATCHING MECHANISM</b> .....	<b>22</b>
<b>4.1 Preprocessing Stage</b> .....	<b>22</b>
<b>4.2 Matching Stage</b> .....	<b>23</b>
<b>4.3 Sorting Stage</b> .....	<b>23</b>
<b>4.4 Filtering Stage</b> .....	<b>24</b>
<b>4.5 Esco-Tjad Database Blueprint</b> .....	<b>24</b>
<b>4.5.1 Processing Esco-Tjad Database</b> .....	<b>25</b>
<b>4.6 Solutions To Improve Matches</b> .....	<b>26</b>
<b>5 EVALUATION</b> .....	<b>31</b>
<b>5.1 Quantitative Results</b> .....	<b>31</b>
<b>5.2 Qualitative Results</b> .....	<b>32</b>
<b>6 CONCLUSION - FUTURE WORK</b> .....	<b>34</b>
<b>6.1 Conclusion</b> .....	<b>34</b>
<b>6.2 Future Work</b> .....	<b>34</b>
6.2.1 Skills To Occupation Title Relation .....	34
6.2.2 Profession To Education Linking .....	34
6.2.3 Identification Of Not Politically Correct Terms .....	35

## LIST OF TABLES

Table 1.....	8
Table 2.....	9
Table 3.....	13
Table 4.....	17
Table 5.....	23
Table 6.....	24
Table 7.....	25
Table 8.....	25
Table 9.....	25
Table 10.....	26
Table 11.....	26
Table 12.....	26
Table 13.....	27
Table 14.....	27
Table 15.....	28
Table 16.....	28
Table 17.....	28
Table 18.....	29
Table 19.....	29
Table 20.....	31
Table 21.....	31
Table 22.....	32
Table 23.....	33



**LIST OF FIGURES**

Figure 1 ..... 12  
Figure 2 ..... 14  
Figure 3 ..... 18  
Figure 4 ..... 18  
Figure 5 ..... 20  
Figure 6 ..... 22  
Figure 7 ..... 22

# 1. ONTOLOGY OF PROFESSIONS

## 1.1 Issues to be explored

The labour market is an area that concerns the majority of the population. The process of finding a job has been modernized, with the help of the internet and e-services. What is lacking in the labour market in Greece is standardisation based on an ontology covering the labour market and including the occupations and the skills they require, so that there are common principles. The standardisation of these entities would also serve to create a standards-compliant format to produce job advertisements and CVs. Job advertisements and CVs in the labour market are created by the individual employer/employee, with the data or requirements that each considers important. Because of this process, the labour market is unevenly shaped, with a variety of characterisations of the same entities, misunderstandings and misconceptions that cause problems in the proper functioning of the labour market.

The process of creating a job advertisement should typically include the title of the position, its description, and the skills/competencies that the employer requires for each position. The advertisements found on the labour market often lack the formal format mentioned above, thus causing consistency problems and misunderstandings. The difficulties encountered in relation to job advertisements can be divided into three main categories: advertisements with a strong emphasis on the social and emotional skills (soft skills) of candidates; advertisements with discriminatory practices (gender, use of sexist language, etc.); incomplete advertisements.

### 1.1.1 Advertisements focusing on social and emotional skills

The main problem that job advertisements suffer from is the emphasis on the "soft skills" of the worker, while the technical skills (hard skills) related to the occupation are often not given the same importance. Searching for candidates based on soft skills creates difficulties in rationally matching a suitable person with the position in question.

Soft skills are an important part of the characteristics that an employee should have. Characteristics such as politeness, the ability to work in a team, and organisational skills are essential for the proper functioning of a workplace. However, if not accompanied by skills relevant to the profession, these elements give a partial picture, making it problematic to find a suitable candidate. Table 1.1 shows the frequency of demand for certain soft skills in a sample of 3 000 advertisements.

Table 1: Social And Emotional Skills Lemmas, Greek Language

Lemma	Counter
Επικοινωνων- <i>communic-</i>	1017
Ομαδικ- <i>Teamwork-</i>	577
Οργανωτ- <i>Organis-</i>	484
Επαγγελματική εμφάνιση/ συμπεριφορά	297

<i>Professional behavior/ appearance</i>	
Δυναμικ- <i>Dynamic</i>	197
Πίεση <i>Pressure</i>	151
Κριτική σκέψη <i>Critical thinking</i>	142
Ευχαρ- <i>Pleasant</i>	120
Ευγεν- <i>Polite</i>	55

### 1.1.2 Advertisements with discriminatory practices against individuals or groups

Discrimination of employees based on criteria related to their personal characteristics, such as age, gender, nationality, is also a characteristic of job advertisements that appear in Greece.

E.g., «επιθυμητή ηλικία άνω των 30», «ηλικία 18-30», «επιθυμητή ηλικία 35-45», «υπηκοότητα ΕΕ»,

«ζητούνται νεαρές κοπέλες...», «νέοι με καλή φυσική κατάσταση»

"Desirable age up to 30", "age 18-30", "desirable age 35-45", "EU nationality", "young girls wanted...", "young man in good physical condition"

### 1.1.3 Incomplete advertisements

The possibility of finding a correlation between an advertisement and a job becomes more difficult when the advertisement is incomplete. The advertisement in these cases may be deficient in terms of working conditions, type of job, location, qualifications. In this way there is a possibility that suitable candidates may not apply, considering themselves not to be suitable for the post, and that unsuitable candidates, meeting the criteria set out in the advertisement, may be faced with a job they cannot support.

Θέση: υπάλληλος γραφείου

Table 2: Example of incomplete advertisement, Greek language

Αναγραφόμενες δεξιότητες	Ενδεικτικές δεξιότητες
Χαμογελαστή, ευχάριστη προσωπικότητα, εξυπηρέτηση πελατών <i>Smiley, pleasant personality, customer service</i>	Πενθήμερη οκτάωρη εργασία σε εξυπηρέτηση πελατών, στα γραφεία μας στο Μαρούσι. Απαραίτητες δεξιότητες: χρήση ηλεκτρονικού υπολογιστή, γνώση Microsoft office suite, τηλεφωνική εξυπηρέτηση πελατών Χαρακτηριστικά: ομαδική εργασία, οργανωτικός χαρακτήρας, ευγένεια κατά την εξυπηρέτηση

	<p><i>Five-day, eight-hour work in customer service at our offices in Marousi.</i></p> <p><i>Required skills: computer usage, knowledge of Microsoft office suite, telephone customer service</i></p> <p><i>Characteristics: teamwork, organizational character, courtesy in service</i></p>
--	--

## **1.2 Thesis objective**

The issue being explored in the context of the thesis is the lack of a comprehensive formulation of occupations and their skills. By creating a mechanism to match occupations and skills with the ESCO ontology, an attempt is made to automatically match job advertisements and ESCO, with the aim of applying the tool to the Greek labour market to improve the job search process.

## 2 ESCO

### 2.1 Introduction to ESCO

ESCO is the multilingual European system for classifying skills, competences, qualifications, and occupations. The ESCO classification identifies and categorises skills, competences, qualifications, and occupations relevant to the EU labour market, education, and training. It systematically presents the relationships between the different concepts.

ESCO is organised in three main areas:

- Occupations pillar
- Knowledge, skills and competences pillar
- Qualifications pillar

The skills pillar of the ESCO classification distinguishes between the concepts of skills/competences on the one hand and knowledge on the other, identifying the type of skills. However, it does not make any distinction between skills and competences. Each of these concepts is accompanied by a preferred term and several alternative (non-preferred) terms in each of the 27 ESCO languages. Each concept also includes an explanation in the form of a description. The ESCO skills domain contains 13 485 concepts hierarchically structured into four sub-classifications. Each sub-classification refers to different types of knowledge and skill/competence concepts: knowledge, skills, professional principles & values, language skills and knowledge.

Vocational qualifications are the formal outcome of an assessment and validation process, achieved when the competent body certifies that an individual has achieved learning outcomes that meet specific standards. One of ESCO's main missions is to build stronger links between the world of education and training and the world of work, helping to reduce skills mismatches and supporting a better functioning labour market.

Overall, this distinction allows ESCO to organise the terminology/ontology for the European labour market and education sector in a useful way.

#### 2.1.1 Concepts and terms

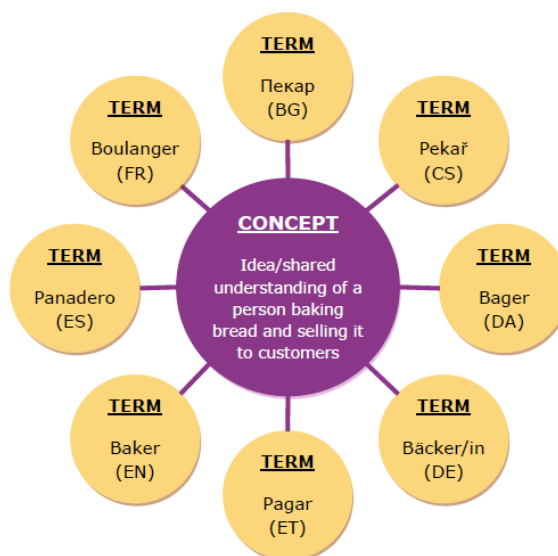
ESCO provides for a distinction between "concept" and "term". A 'concept' is defined as a valid idea (with its corresponding description) in the workplace, irrespective of its linguistic realisation. Consequently, these concepts are not linguistically dependent. A 'term' is the linguistic realisation or description of a concept and is therefore language dependent.

E.g., concept: a man who bakes bread and sells it.

Term: φούρναρης (GR) , baker (EN)

Each concept is associated with at least one term in each language included in ESCO. In many cases, the language has more than one term to refer to the same or a very close concept. ESCO provides for multiple terms per concept.

Within the ESCO data model, each term is a discrete unit, and all terms are always related to a concept. This is illustrated in the following diagram, which shows a concept-term relationship for eight languages:



**Figure 1: ESCO concept-term relationship for eight languages**

Three types of terms are used in ESCO: preferred terms, non-preferred terms, and hidden terms. Each ESCO occupation and skill concept has at least one term for each of the 27 ESCO languages: the preferred term. This term is not reused for another occupation or skill in the same language and is therefore unique. From a group of terms with similar meanings, the preferred term is the one that best represents linguistically the profession or skill it describes.

As mentioned earlier, a language may contain more than one term to refer to the same/similar concept and therefore ESCO may contain multiple terms per concept in each language. Non-preferred terms may be synonyms (words with similar or close meanings), but may also be spelling variations, conjugations, abbreviations, etc. They are regularly used by jobseekers, employers, or educational institutions to refer to concepts described in the classification by the preferred term.

ESCO also lists terms that are commonly used in the labour market to refer to an occupation, but are also considered outdated, misspelled or politically incorrect. These are referred to as 'hidden terms' as they are useful for indexing, search and text mining purposes, but are not accessible to end users. When searching for a hidden term in the ESCO portal, the user is automatically redirected to the profession with the preferred term. The hidden term is not displayed.

## 2.2 The occupations pillar

The occupations pillar aims to describe all occupations related to the European labour market. ESCO includes 3008 occupations in v.1.0.3. Each occupation concept describes the core meaning and essence of the occupation and provides some useful information about it (metadata).

The key element that defines an ESCO occupation is the main idea or understanding of what the occupation is about and how it differs from other occupations. These are captured in the description and scope note.

The description in the ESCO is a text field that provides a brief explanation of the concept of the occupation and how it is to be understood. For this reason, a description is always provided for each ESCO occupation.

The scope in the ESCO is sometimes used to make things less ambiguous and to clarify its semantic boundaries. For example, it may list the professions that are considered in scope and those that are considered out of scope.

### Knowledge, skills and competences in professional profiles

Each ESCO occupation is associated with basic and optional concepts of knowledge, skills and competences:

Essential knowledge, skills and competences are those knowledge, skills and competences that are commonly relevant to an occupation, regardless of the work context, employer, or country.

Optional knowledge, skills and competences are those knowledge, skills and competences that may be relevant or occur in the exercise of an occupation, depending on the employer, work context or country. Optional knowledge, skills and competences are very important for job matching because they reflect the diversity of jobs within the same occupation.

**Table 3: ESCO essential and optional skills examples**

Waiter/waitress

Essential skill	Σερβίρισμα ποτών <i>Serve beverages</i>
Optional skill	Προετοιμασία φλαμπέ πιάτων <i>Prepare flambeed dishes</i>

### The structure of the professions sector

Occupations in ESCO are structured by mapping them to the international standard classification of occupations (ISCO-08), which has been developed by the international labour organisation (ILO). The ESCO occupations and the isco-08 hierarchy make up the ESCO occupational domain. Isco-08 provides the first four levels, while the ESCO occupations provide the fifth and higher levels. Each ESCO occupation is assigned to a group of isco-08 units (even if they are not directly related to it, e.g. If they are at level six or seven).

### Hierarchy of occupations

Occupations in ESCO are described at different levels, depending on the language used and the requirements of the labour market. However, these may differ between groups of countries (e.g., Some states may need to cover different types of sommeliers, while in others, the distinction in this occupation may have little relevance). For ESCO to cover both cases, some professions have broader-closer links between them. However, only occupations that are relevant to the European labour market are included. Therefore, the more detailed occupations do not necessarily cover the full scope of the more general occupation. (e.g., Not all types of sommeliers are covered under sommelier).



Figure 2 2.4 ESCO hierarchy of occupations example

### 2.3 The knowledge, skills and competences pillar

The knowledge, skills and competences pillar, provides a comprehensive list of skills relevant to the European labour market. ESCO includes 13890 skills in v.1.0.3.

#### Knowledge, skills and competence concepts

The skills domain includes knowledge, skills and competences defined as follows:

**Knowledge:** the set of principles, theories and practices related to a field of work or study. Knowledge is described as theoretical and/or practical and is the result of assimilating information through learning.

**Skills:** the ability to apply knowledge and use expertise to complete tasks and solve problems. Skills are described as cognitive (involving the use of logical, intuitive and creative thinking) or practical (involving manual dexterity and the use of methods, materials, tools and instruments).

**Competencies:** the demonstrated ability to use knowledge, skills and personal, social and/or methodological competencies in work or study situations, and in professional and personal development.

Although sometimes used as synonyms, the scope of the terms 'skill' and 'competence' can be distinguished. 'Skill' refers to the use of methods or means in a specific context and in relation to defined tasks. 'Competence' is broader and refers to the ability of an



individual, faced with new situations and unforeseen challenges, to use and apply knowledge and skills in an independent and self-directed manner. However, there is no distinction between skills and competences recorded in the esco skills domain.

### **Contents of the skills domain**

As for occupations, ESCO provides metadata for each skill domain concept, including:

A **preferred term** used for the linguistic realization of the knowledge, skill/competence.

**Non-preferred terms** (synonyms, spelling variations, conjugations, abbreviations, etc.).

**Hidden terms** (e.g. Obsolete, misspelled or politically incorrect terms).

A **description** that explains in greater depth what the skill is about according to the action verb and the level of detail used in the title.

A **scope note** that clarifies the semantic boundaries of the concept.

The **type of skill**: i) skill/ability concepts or ii) knowledge concepts.

The **relationship to ESCO occupations**. This field indicates for which occupations the knowledge, skill or competence is usually relevant, including those for which it is necessary and those for which it is optional. In some cases, the relationship will show how the knowledge, skill or competence is relevant to other knowledge, skills and competences. The relationship shall also include a distinction between essential and optional.

The **level of reusability**, which indicates the breadth of application of a knowledge, skill or competence concept. This is crucial to support occupational mobility. ESCO distinguishes four levels of skills reuse:

**Horizontal knowledge**, skills and competences are relevant to a wide range of occupations and sectors,

**Cross-sectoral knowledge**, skills and competences are relevant to occupations in different economic sectors,

**Sectoral knowledge**, skills and competences are specific to one sector but relate to more than one occupation in that sector,

**Occupation-specific knowledge**, skills and competences are usually only applicable to one occupation or specialisation.

### **Horizontal knowledge, skills and competences**

As mentioned above, horizontal knowledge, skills and competences relate to a wide range of occupations and economic sectors. They are often referred to as basic, core or soft skills and are the cornerstone of an individual's intellectual development.

Within the skills domain, transversal skills and competences are organised in a hierarchical structure under the following five headings:

- thinking
- language
- application of knowledge
- social interaction
- attitudes and values

### **Adaptation of skills**

Skills contextualisation is a method for creating concepts of knowledge or skills and competences through the analysis of how transversal skills, competences or knowledge are applied in the specific context of a domain or occupation. In this way, transversal knowledge, skills and competences that are rather abstract can be transferred to a more detailed level so that they can be directly applied to occupational profiles.

*Example: the skill "measure" is too abstract to be directly linked to the occupation "metal furnace operator". This relationship would produce too many results if it were used in skill-based job matching, since measurement is relevant for a large number of occupations and sectors. Through the framing of skills, the skill can be made more specific. A skill called 'measuring oven temperature' could, for example, be used in the occupational profile of 'metal oven operator'.*

### **Structure of the skills domain**

The ESCO skill domain does not include a complete hierarchical structure. Instead, the elements of the domain are structured in four different ways:

- Through their relationship to occupations, using the occupation outlines as an entry point,
- Through a hierarchy (only for transversal knowledge, skills and competences),
- Through relationships showing how knowledge, skills and competences are related to other knowledge, skills and competences (in particular in cases where skills are framed),
- Through functional collections that allow the selection of subsections of the skill domain, depending on the purpose for which they are to be used.

### **The certification (qualification) domain**

The certification domain aims to collect existing information on qualifications.

### **Content of the qualifications domain**

The qualifications in ESCO are derived from the national qualification databases of the states. These qualifications are included in national qualifications frameworks that have

been referred to ESCO. Since 2014, the commission has been financially supporting member states and other partner countries (EFTA, EEA and candidate countries) to develop national qualifications databases and link them to the learning opportunities and qualifications in Europe (loq)24 portal and to ESCO.

In contrast to the occupations and skills domains, the qualifications domain is complemented exclusively by external sources and not by data generated by the commission.

## 2.4 Why is ESCO useful?

ESCO was created to serve several purposes. Firstly, through the database, it improves the communication between the education and training sector and the European labour market. By creating a commonly accepted code of definition of occupations, citizens can more easily move professionally between countries. Similarly, the process of identifying suitable professionals by large companies and organisations is simplified. This gives the process of data exchange between employee and employer a common conceptual - ontological basis.

## 2.5 ESCO database

For the analysis all ESCO occupations and skills were collected in Greek and English language. Several meta-information (metadata) were included, such as the ESCO specific code (URI) for unique identification of each entry, preferred, alternative and hidden tags for each occupation and skill.

All this information was merged into two databases: ESCO occupations DB, ESCO skills DB. All data went through pre-processing and normalization processes to be in a computational format that can be used for workplace applications (e.g., Application for automatic generation of suggestions and recommendations for employers/candidates).

Table 4: ESCO databases size

<b>ESCO occupations DB</b>	<b>3008 occupations</b>
<b>ESCO skills DB</b>	<b>13890 skills</b>

## 2.6 ESCO DB blueprint

The databases created were in the following format. As already mentioned, each entry is accompanied by meta-information. The meta-information for the occupations includes the conceptual type (occupation), the unique website (link), the ISCO group to which the occupation belongs, the preferred, alternative and hidden terminological annotations, the current status of the occupation description in the ESCO database where released are considered all entries that are in use in the database, and obsolete those that have been withdrawn and replaced. There is also the date when each profession was created or edited, and the categorization of regulated professions. Regulated professions are those professions that require a certification or meet certain minimum requirements to practice them. Finally, a brief description of each profession and its specific code is included.

The Greek Labour Market And The European System Of Classification Of Skills And Occupations ESCO: A First Approach.

	A	B	C	D	E	F	G	H	I	M	N
1	conceptType	conceptU	isocGr	preferredLabel	altLabels	iddenLi	statu	modified	regulatedProfessionN	description	code
2992	Occupation	<a href="http://data.e">http://data.e</a>	4222	χειριστής ζωντανών συζητήσεων/χειριστρια ζωντανών συζητήσεων	live chat/live chat o	released	2016-07-05	<a href="http://data.europa.eu/es">http://data.europa.eu/es</a>	Οι χειριστές ζωντανών συζητήσεων	4222.1.1	
2993	Occupation	<a href="http://data.e">http://data.e</a>	2422	υπεύθυνος για θέματα αθλητισμού/υπεύθυνη για θέματα αθλητισμού	released	2021-03-16	<a href="http://data.europa.eu/es">http://data.europa.eu/es</a>	Οι υπεύθυνοι για θέματα αθλητισμού	2422.12.13		
2994	Occupation	<a href="http://data.e">http://data.e</a>	2611	νομικός σύμβουλος επιχείρησης/νομική σύμβουλος	δικηγόρος με ειδίκευση στο δικαίωμα	released	2020-12-09	<a href="http://data.europa.eu/es">http://data.europa.eu/es</a>	Οι νομικοί σύμβουλοι	2611.1.1	
2995	Occupation	<a href="http://data.e">http://data.e</a>	2511	αναλυτής επιχειρησιακών αναγκών και διαχειριστής επιχειρησιακών αναγκών	released	2017-02-02	<a href="http://data.europa.eu/es">http://data.europa.eu/es</a>	Οι αναλυτές επιχειρησιακών αναγκών	2511.9		
2996	Occupation	<a href="http://data.e">http://data.e</a>	3521	τεχνικός μεταδόσεων	τεχνικός συστημάτων μετάδοσης	released	2017-01-17	<a href="http://data.europa.eu/es">http://data.europa.eu/es</a>	Οι τεχνικοί μεταδόσεων	3521.2	
2997	Occupation	<a href="http://data.e">http://data.e</a>	3111	τεχνικός μέτρησης ατμοσφαιρικής ρύπανσης	τεχνικός μέτρησης σωματιδίων/τεχνικός μέτρησης ατμοσφαιρικής ρύπανσης	released	2020-09-02	<a href="http://data.europa.eu/es">http://data.europa.eu/es</a>	Οι τεχνικοί μέτρησης ατμοσφαιρικής ρύπανσης	3111.2.1	
2998	Occupation	<a href="http://data.e">http://data.e</a>	2330	καθηγητής χημείας δευτεροβάθμιας εκπαίδευσης/καθηγήτρια χημείας δευτεροβάθμιας εκπαίδευσης	released	2016-12-22	<a href="http://data.europa.eu/es">http://data.europa.eu/es</a>	Οι καθηγητές χημείας	2330.1.4		
2999	Occupation	<a href="http://data.e">http://data.e</a>	2635	κοινωνικός λειτουργός με ειδίκευση στην ανάλυση κοινωνικής ανάπτυξης	released	2022-01-13	<a href="http://data.europa.eu/es">http://data.europa.eu/es</a>	Οι κοινωνικοί λειτουργοί με ειδίκευση στην ανάλυση κοινωνικής ανάπτυξης	2635.3.5		
3000	Occupation	<a href="http://data.e">http://data.e</a>	3322	εμπορικός αντιπρόσωπος ανανέωσης ενέργειας/εμπορική αντιπρόσωπος ανανέωσης ενέργειας	released	2016-07-05	<a href="http://data.europa.eu/es">http://data.europa.eu/es</a>	Οι εμπορικοί αντιπρόσωποι ανανέωσης ενέργειας	3322.1.2		
3001	Occupation	<a href="http://data.e">http://data.e</a>	7512	τεχνίτης ζαχαροπλαστικής/τεχνίτρια ζαχαροπλαστικής	ζαχαροποι/τεχνίτρια ζαχαροπλαστικής/ζαχαροποι	released	2016-07-05	<a href="http://data.europa.eu/es">http://data.europa.eu/es</a>	Οι τεχνίτες ζαχαροπλαστικής	7512.5	

Figure 3: ESCO occupations database sample

The structure is similar for skills. It includes the concept of each skill, the unique link, the type of skill, which, as already mentioned, are divided into knowledge, skill and competence. The reuse level is also defined: skills are divided into cross-sector, occupation-specific, sector-specific. Cross-sectoral knowledge, skills and competences relate to occupations in different sectors. Sector-specific knowledge, skills and competences are specific to an industrial sector, but relate to more than one occupation in that sector. Finally, occupation-specific knowledge, skills and competences are usually applicable to only one occupation or specialisation. The following meta-information concerns the preferred, alternative and hidden terminological annotations that capture the linguistic realisations in use of the concept under consideration. As for occupations, the status of the skills is defined as released or obsolete, the date of their modification, and the description of each skill.

	A	B	C	D	E	F	G	H	I	M
1	conceptType	conceptUr	skillType	reuseLe	preferredLabel	altLabels	iddenLabel	status	modifiedDate	description
13565	Knowledge	<a href="http://data.europa.eu/es">http://data.europa.eu/es</a>	skill/comp	sector-spe	δημιουργία σχεδίου πτήσης	released	2016-12-2	Ανάπτυξη		
13568	Knowledge	<a href="http://data.europa.eu/es">http://data.europa.eu/es</a>	knowledge	sector-spe	πρότυπα για την ασφάλεια των	released	2017-01-0	Πρότυπα γ		
13569	Knowledge	<a href="http://data.europa.eu/es">http://data.europa.eu/es</a>	skill/comp	sector-spe	διενέργεια καρδιαγγειακών επε	released	2021-12-1	Διενέργεια		
13570	Knowledge	<a href="http://data.europa.eu/es">http://data.europa.eu/es</a>	knowledge	sector-spe	τεχνολογία τεχνολογία φινιρίσμ	released	2017-01-0	Διεργασίες		
13573	Knowledge	<a href="http://data.europa.eu/es">http://data.europa.eu/es</a>	knowledge	sector-spe	γεωγραφία τουριστικές περιοχές	released	2016-09-1	Ο τομέας τ		
13581	Knowledge	<a href="http://data.europa.eu/es">http://data.europa.eu/es</a>	skill/comp	sector-spe	αξιολόγηση της ποιότητας της ι	released	2021-12-1	Αξιολόγησ		
13582	Knowledge	<a href="http://data.europa.eu/es">http://data.europa.eu/es</a>	skill/comp	sector-spe	τήρηση διοικητικών αρχείων στ	released	2021-12-1	Δημιουργί		

Figure 4: ESCO skills database sample

### **3 COLLECTION OF TEXTUAL JOB ADVERTISEMENTS (TJAD)**

#### **3.1 Greek market**

The existence of ESCO ensures the information and data needed at European level. For the research it was necessary to collect data exclusively related to the Greek labour market. In this context, a relevant, structured collection of advertisements from the Greek labour market (TJAD) was created. The TJAD was fed by job clippings as posted on the relevant website (kariera.gr). Kariera.gr is a recognized job search website in Greece.

#### **3.2 Data collection**

Job advertisements were collected from kariera.gr using a crawler. Specifically, the python library "beautiful soup" was used for the crawling process. The crawler repeated this process monthly. A total of 7 months of advertisements were collected from the website, corresponding to 31074 advertisements including 23579 skills.

#### **3.3 Data organization**

The data was divided into two sub-collections to facilitate the research process. The first has all job titles and descriptions as derived from the job advertisements, while the second collection contained all the skills included in each of them.

The pre-processing that this information received was carried out for both English and Greek. In particular, the data went through the lemmatization process. Also, at this stage all special characters that could create noise in the following procedures were removed.

To distinguish between title, job description and skills, qualifications and constraints, identifiers were created. The title is distinct from the structure of the advertisement, while the description in most cases is in the form of text. Key words or key phrases are used to aggregate skills, applied within the context of the advertisement description. The key words are words or phrases such as "skills:", "essential qualifications:", "for this position you need:". All the elements under the words/phrases are counted as skills. The structure of the skills within the description is also provided for, so that they can be identified in all cases. Usually, skills are either displayed in a list or in the form of a single text.

The ads had the following format:

Job Title	Πωλητής/Πωλήτρια <i>Salesman/Saleswoman</i>
Job Description	Η εταιρεία μας αναζητά για το κατάστημα Ναυπλίου Πωλήτρια-η καταστήματος part time <i>Our company is looking for a part time saleswoman for our store in Nauplio</i>
Skills/Qualifications/Constraints	<ul style="list-style-type: none"> <li>- Γνώση του χώρου της μόδας, <i>Knowledge of the fashion industry</i></li> <li>- Επικοινωνιακός χαρακτήρας και ομαδικό πνεύμα, <i>Communicative character and team spirit</i></li> <li>- Πτυχίο Αγγλικής γλώσσας, <i>Degree in English language</i></li> <li>- Χρήση Η/Υ, <i>Computer skills</i></li> <li>- Τουλάχιστον 2 χρόνια προϋπηρεσία, <i>At least 2 years of experience</i></li> </ul>

Soft Skills

Figure 5: Example of the data collecting process

### 3.4 Difficulties of the procedure

During the data collection process, difficulties were observed in terms of the subsequent need to match the data with es ESCO co standards.

1. **Locality:** in Greece some occupations require skills that are country-specific. They are skills that would not be required for the same occupation in another country.

Π.χ. Επάγγελμα: λογιστής

Απαιτούμενα προσόντα: γνώση ελληνικών λογιστικών προτύπων

*Occupation: accountant*

*Skills required: knowledge of Greek accounting standards*

2. **Evolution over time - dynamics of occupations:** as humanity and technology progresses, the skills of occupations evolve and change gradually, which complicates the process of identifying the basic skills that characterise an occupation.

Π.χ. (2000)

Επάγγελμα: σερβιτόρος/σερβιτόρα

Απαιτούμενα προσόντα: εμπειρία στο σέρβις, γνώση ελληνικών κρασιών

*Occupation: waiter/waitress*

*Skills required: experience in service, knowledge of Greek wines*

(2022)

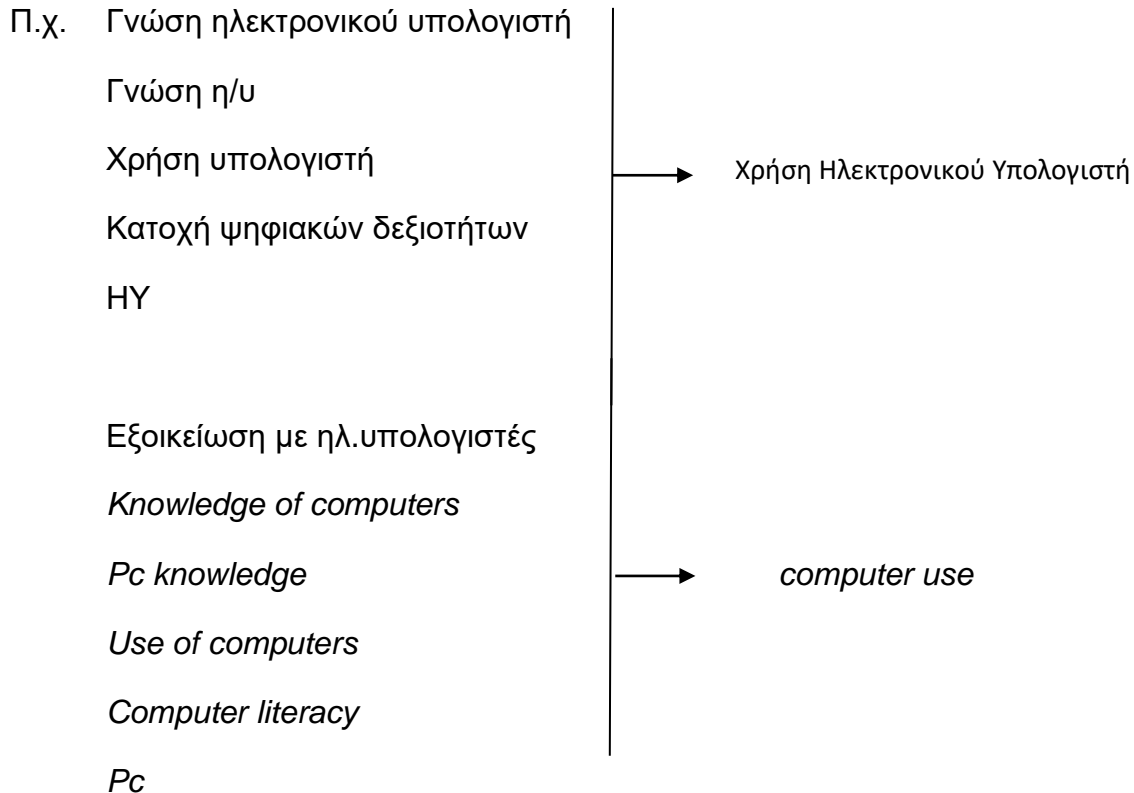
Επάγγελμα: σερβιτόρος/σερβιτόρα

Απαιτούμενα προσόντα: γνώση PDA, χρήση ηλεκτρονικού υπολογιστή, εμπειρία στο σέρβις

*Occupation: waiter/waitress*

*Skills required: PDA knowledge, computer skills, experience in service*

**3. Specificity of the linguistic expression of the Greek language:** the Greek language has a great richness and variety of expressions. This results in several difficulties in grouping and standardising concepts.



From the considerations above, arises the need to normalise and standardise professions and their skills to improve CVS and job advertisements, modernizing the job search process.

## 4 AUTOMATED MATCHING MECHANISM

The matching mechanism uses the ESCO ontology to match occupations and skills for the Greek labour market. It pre-processes the data coming from TJAD and match them with ESCO data. Then, the collected data are ranked according to their score, and separated based on their reliability.



Figure 6 : Stages of the matching mechanism

### 4.1 Pre-processing stage

In the pre-processing stage, the data are put in their optimal form in order to proceed to the matching stage. First, the titles written in mixed Greek English are corrected using python code for language processing. Then, the verbal analysis of both the descriptions in ESCO and the job advertisements in TJAD follows. Using code, punctuation marks, sentences, lists are separated and items that are identified as verbal units are lemmatized.

E.g., normalization - lemmatization

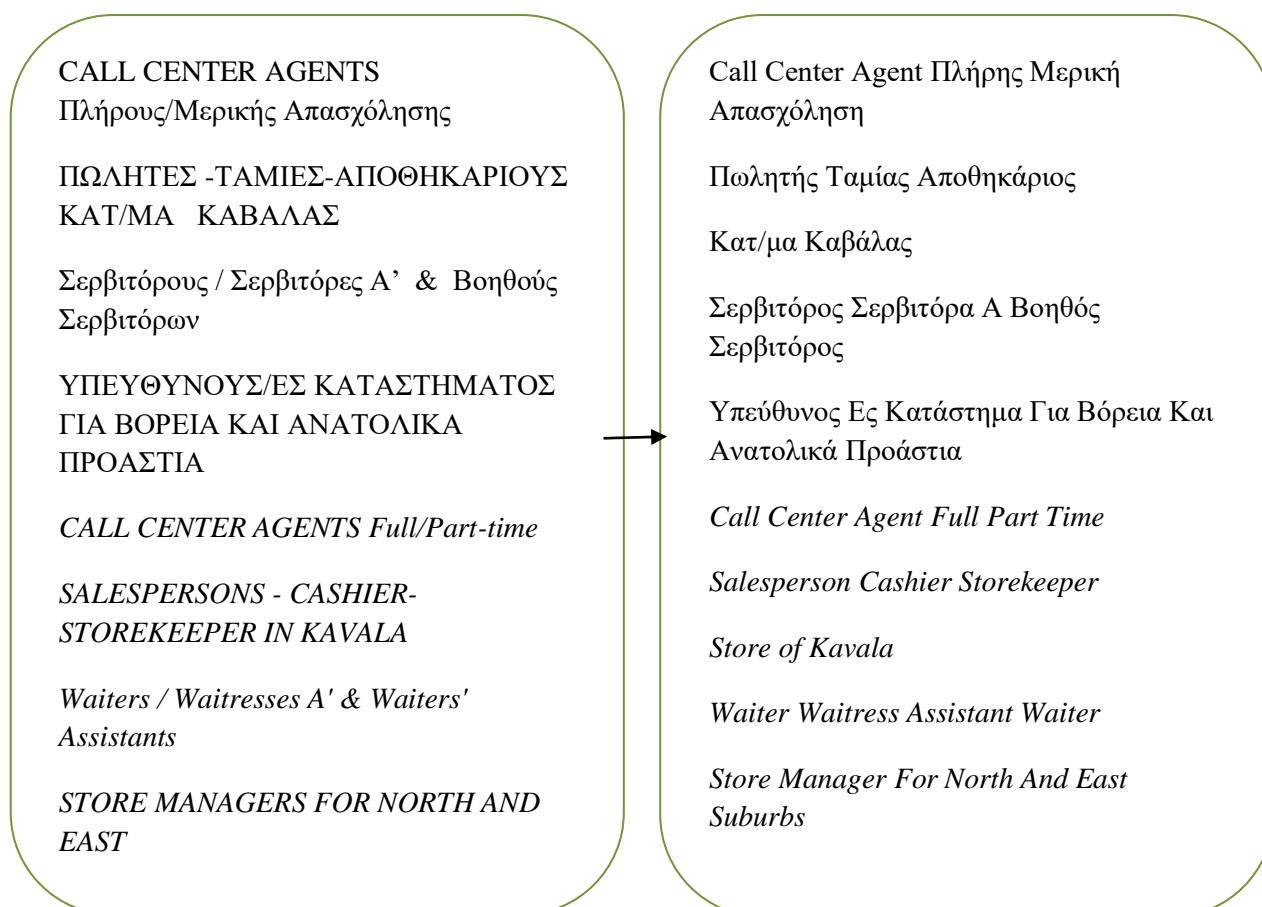


Figure 7: Example of normalization- lemmatization



## 4.2 Matching stage

In the next stage, the set of entries from the advertisements is compared with the entries in ESCO (as obtained from the entry process of the descriptions found in the esco titles). For each possible pair, a set of scores is calculated. The scores also depend on the level at which an ad tag was successfully matched with the ESCO possible pairs. From the matching process, all occupations/skills that managed to obtain a total score > 0 are kept.

The scores compute the common elements of TJAD occupation and ESCO occupation, in terms of:

(a) the size of the TJAD title being matched  $score =$

$$\frac{TJAD\ occupation + ESCO\ occupation\ common\ elements}{size\ of\ TJAD\ title}$$

(b) the size of the ESCO title being matched  $score =$

$$\frac{TJAD\ occupation + ESCO\ occupation\ common\ elements}{size\ of\ ESCO\ title}$$

The final score is derived:  $score = \frac{a+\beta}{2}$

**Table 5: Example of ESCO matches scores**

	ESCO match	Score
Απόδοση προσοχής στην λεπτομέρεια <i>Attention to detail</i>	Απόδοση προσοχής στην λεπτομέρεια <i>Attention to detail</i>	1.0
Γνώση της αγγλικής γλώσσας <i>Knowledge of the english language</i>	Αγγλικά English	0.87
Επιθυμία για γνώση και αντιμετώπιση νέων προκλήσεων <i>Willingness to learn and tackle new challenges</i>	Επίδειξη βούλησης για μάθηση <i>Eager to learn</i>	0.58
Εμπειρία με φορείς κινητής τηλεφωνίας θα θεωρηθεί προσόν <i>Experience with mobile operators will be considered as a plus</i>	Μάρκετινγκ μέσω κινητών συσκευών <i>Mobile phone marketing</i>	0.45

## 4.3 Sorting stage

This stage involves the ranking of the entries that have been found at the matching stage. For some TJAD entries there are matches corresponding to ESCO preferred label, but also alternative and hidden. If TJAD entries have been matched to the preferred label of the ESCO entry, the score will be higher than that of a match to the alternative label. A lower score is given to cases where the matching was made with the hidden label of ESCO. The ranking therefore goes as follows:

Matched to ESCO preferred label → matched to ESCO alternative label → matched to ESCO hidden label.

#### 4.4 Filtering stage

In the final stage, the data are filtered to verify the priority order given in the sorting stage. In particular, it is checked whether the matches that are first in the hierarchy have score = 1.0. If it is true that score = 1.0, then the match is accurate and has priority over the other matches. There is one more category of matches, those with scores >0.75. These matches pass to a lower ranking, but remain in the process. These items are considered less reliable than the 1.0 matches, but seem useful in cases of ambiguity. Matches with a score lower than 0.75 are not considered reliable and are removed from the process for accuracy reasons.

#### 4.5 ESCO-TJAD database blueprint

The matching process results in a new database, which includes the TJAD ads with their ESCO matches. The database is divided into two parts. The first is about the occupations, while the second is about the skills of each occupation.

##### Occupations

The database on occupations brings together the TJAD titles with their accompanying descriptions. For their ESCO counterpart, the preferred, alternative, and hidden labels and occupation description are recorded for both English and Greek. Finally, the match score is written, to separate the exact matches (1.0) from the less accurate (>0.75).

**Table 6: ESCO-TJAD occupations blueprint**

TJAD Title	ESCO Preferred	ESCO Alternative	ESCO Hidden	ESCO Description	Match Score
ΟΔΗΓΟΣ ΑΣΘΕΝΟΦΟΡΟΥ	οδηγός ασθενοφόρου <i>emergency ambulance driver</i>	βοηθός ασθενοφόρου / νοσηλεύτρια ασθενοφόρου / τεχνικός ασθενοφόρου / νοσηλεύτης ασθενοφόρου  <i>technician in emergency medical care / ambulance emergency medical technician / assistant in emergency care / ambulance technician</i>		Οι οδηγοί ασθενοφόρου χρησιμοποιούν οχήματα έκτακτης ανάγκης για την αντιμετώπιση έκτακτων ιατρικών περιστατικών, υποστηρίζουν το έργο των νοσηλευτών, μετακινούν ασθενείς με ασφάλεια, καταγράφουν τις αλλαγές στις ζωτικές παραμέτρους του ασθενούς και αναφέρουν σχετικά στους επικεφαλής νοσηλευτές, διασφαλίζοντας παράλληλα τη σωστή αποθήκευση, μεταφορά και λειτουργία του ιατρικού εξοπλισμού, υπό την επίβλεψη και κατά παραγγελία του ιατρού. <i>Emergency ambulance drivers use emergency vehicles to respond to medical emergencies and support the work of paramedics, move patients safely, take note of changes in the patient's vital signs and report to the paramedics in charge, ensuring the medical equipment is well stored, transported and functional, under supervision and on order of a doctor of medicine.</i>	1.0

##### Skills

The skills database follows the same pattern. Specifically, for each TJAD skill there is the corresponding ESCO, with preferred, alternative and hidden labels in Greek and English, description of the ESCO skill and match score.

**Table 7: ESCO-TJAD skills blueprint**

TJAD Title	ESCO Preferred	ESCO Alternative	ESCO Hidden	ESCO Description	Match Score
Ταυτόχρονη εκτέλεση πολλαπλών καθηκόντων <i>Ability to multitask</i>	ταυτόχρονη εκτέλεση πολλαπλών καθηκόντων <i>perform multiple tasks at the same time</i>	ταυτόχρονη πραγματοποίηση πολλαπλών εργασιών / ταυτόχρονη διεκπεραίωση πολλαπλών εργασιών / ταυτόχρονη εκτέλεση πολλαπλών καθηκόντων / ταυτόχρονη διεκπεραίωση πολλαπλών παραγγελιών / multi-tasking <i>perform multitasking activities / carry out multitasking / perform multi tasks at the same time / multi-tasking / multitask /</i>		Ταυτόχρονη εκτέλεση πολλαπλών καθηκόντων, έχοντας επίγνωση των βασικών προτεραιοτήτων. <i>Execute multiple tasks at the same time, being aware of key priorities.</i>	1.0

#### 4.5.1 Processing ESCO-TJAD database

The next step was editing and checking on all occupation and skill records. Each entry was evaluated based on whether a match was found in the ESCO database. Through the evaluation process, the data was separated into two categories:

##### I. ESCO match found

If there was a match between the TJAD occupation/skill with ESCO occupation/skill, then the respective entry was passed as correct. For those cases where the matching tool had failed to match an ESCO occupation/skill which did exist, it was manually placed and thus the entry was passed to the category of found matches.

##### II. ESCO match not found

All entries that cannot be linked to any of the ESCO entities end up in this category. This can happen for several reasons:

Very general reference: in some cases, the advertisement was written in such a way that it was not possible to understand the profession/skill being referred to from the title alone.

**Table 8: Example for no match found due to very general reference**

Job title	Ζητείται άτομο για πρωινή βάρδια <i>Person for the morning shift needed</i>
Description	Ζητείται άτομο για την κάλυψη πρωινής βάρδιας σε κατάστημα λιανικής πώλησης στο Μαρούσι αττικής. <i>We are looking for a person to cover the morning shift in a retail store in Maroussi, Attica.</i>

Very specific reference: there are cases where an entry is so specific that it cannot be matched with an ESCO entity, either by the mechanism or by the person.

**Table 9: Example for no match found due to very specific reference**

Kariera skill	Καθαριστής πισίνας <i>Pool cleaner</i>
Closest ESCO skill match	Καθαριστής/ καθαρίστρια <i>Cleaner</i>

Occupation/skill mainly found in Greece: some advertisements refer to professions and skills that are not so widespread in the rest of Europe.

**Table 10: Example for no match found due to localization**

Kariera skill	Atlantis ERP
---------------	--------------

Entries that cannot be considered skills: one of the most difficult categories to manage is that of entries that cannot be classified as skills, but provide information relevant to the job.

**Table 11: Example for no match found due to entry not being a skill**

Kariera ≠skill	Εμπειρία τουλάχιστον δύο χρόνια <i>At least two years of experience</i>
Kariera ≠skill	Πτυχίο πολυτεχνείου <i>Engineering degree</i>
Kariera ≠skill	Ευγένεια κατά την επικοινωνία <i>Politeness in communication</i>

"noise" resulting from machine errors: the code that was used attempted to reduce the noise present within the ads, through the removal of punctuation marks and with the help of keywords. There were, however, several cases where such entries failed to generalise in a way that could be avoided.

**Table 12: Example for no match found due to entry being noisy**

Kariera noise	Για αποστολή βιογραφικών πατήστε εδώ! <i>Send your cv here!</i>
---------------	--

#### **4.6 Solutions to improve matches**

In order to improve the process of matching advertisements in greece with the esco ontology, a number of solutions have emerged. During the manual evaluation, two categories of entries were identified for further processing. Those entries that were regularly observed not to find a match, but which could be manually assigned to one of the ESCO categories, and those that, while having many occurrences, did not match any ESCO entry. Thus, the following changes are made to optimize the matching process:

##### **TJAD alternative labels**

The alternative labels used by ESCO have already been mentioned. To these, alternative labels derived from the TJAD data are added. The need for more alternative labels is usually due to language/locality issues. While ESCO has provided for the translation of occupations and skills into many languages, it lacks the knowledge of the language as it is used in everyday speech.

**Table 13: TJAD alternative labels, occupations**

Preferred Label	Alternative Labels	TJAD Alternative Labels
πωλητής προϊόντων και υπηρεσιών μέσω επίδειξης/πωλήτρια προϊόντων και υπηρεσιών μέσω επίδειξης	πωλήτρια προϊόντων και υπηρεσιών μέσω επίδειξης πωλητής προϊόντων και υπηρεσιών μέσω επίδειξης	promoter σημείων πώλησης promoter προϊόντων στο καταναλωτικό κοινό προωθητής προωθήτρια
νοσηλεύτης γενικής περίθαλψης/νοσηλεύτρια γενικής περίθαλψης	γενική νοσοκόμα νοσηλεύτης γενικής περίθαλψης γενικός νοσοκόμος νοσηλεύτρια γενικής περίθαλψης	επισκέπτης υγείας επισκέπτρια υγείας
executive assistant	personal executive assistant senior administrative secretary executive secretary executive administrative assistant personal assistant multilingual secretary foreign language assistant assistant manager foreign language correspondent executive assistant	PA CEO P.A.

**Table 14: TJAD alternative labels, skills**

Preferred Label	Alternative Labels	TJAD Alternative Labels
create solutions to problems	create solution to problems conceive solutions to problems creating solutions to problems devise solutions to problems formulate solutions to problems come up with solutions to problems create solutions	problem solving problem solver
operate electronic payment terminals	use electronic payment devices perform transactions using electronic payment terminals operate electronic payment devices take payments using electronic payment devices utilise electronic payment terminals collect payments using electronic payment devices use electronic payment terminals conduct transactions using electronic payment terminals utilise electronic payment devices collect debit of credit card payments using electronic payment terminals	PDA
επικοινωνία		επικοινωνιακές δεξιότητες επικοινωνιακός επικοινωνιακές ικανότητες επικοινωνιακός χαρακτήρας προφορική επικοινωνία γραφική επικοινωνία διαπροσωπικές σχέσεις

## TJAD new labels

In some cases, it was decided that new concepts additions need to be made to ESCO. In the occupations part, the additions are new end nodes in large occupational trees, defining an occupation very specifically. The new skills added are either country-specific or modern, digital skills that have not yet been added.

**Table 15: TJAD new labels, occupations**

Preferred Label	Alternative Labels	TJAD Alternative Labels
καταμετρητής/ καταμετρήτρια	καταμετρητής καταμετρήτρια υπεύθυνος καταμέτρησης υπεύθυνη καταμέτρησης	Οι καταμετρητές καταμετρούν τις αξίες (χαρτονομίσματα, κέρματα ή άλλα) που διακινεί μία εταιρεία, τηρώντας τους προβλεπόμενους κανόνες ασφαλείας.
counter	cash counter	Counters count the valuable items such as banknotes, coins or other that a company is transporting, whilst taking security precautions.
επόπτης εξέτασης/ επόπτρια εξέτασης	επόπτης εξέτασης επόπτρια εξέτασης εξεταστής εξετάστρια	Οι επόπτες εξέτασης συμβάλλουν στην αξιολόγηση και πιστοποίηση επαγγελματικών ή άλλων δεξιοτήτων μέσω εξέτασης. Παρακολουθούν τους υποψηφίους καθ'όλη τη διάρκεια της εξέτασης, απαντούν σε απορίες των υποψηφίων, επιλύουν τεχνικά προβλήματα όταν αυτά προκύπτουν και διασφαλίζουν ότι η εξέταση λαμβάνει χώρα σύμφωνα με τα καθορισμένα πρότυπα και οδηγίες.

Table 16

#### 4.14 TJAD new labels, skills

Preferred Label	Alternative Labels	TJAD Alternative Labels
Adobe Creative Suite		Το Adobe Creative Suite είναι μια διακοπείσα σουίτα λογισμικού με εφαρμογές γραφικού σχεδιασμού, επεξεργασίας βίντεο και ανάπτυξης ιστού που αναπτύχθηκε από την Adobe Systems.
Adobe Creative Suite	Creative Suite Adobe	Adobe Creative Suite (CS) is a discontinued software suite of graphic design, video editing, and web development applications developed by Adobe Systems.
Πληροφοριακό Σύστημα Εργάνη	YEKA Εργάνη	Το πληροφοριακό σύστημα Εργάνη φροντίζει για την δήλωση και την ασφάλιση της εργασίας, ενώ παράλληλα ασχολείται με τις γραφειοκρατικές και διοικητικές διαδικασίες των επιχειρήσεων αλλά και την δημιουργία νέων θέσεων εργασίας.
ERGANI		Ergani is an online system that allows employers to register themselves and their employees with Social Insurance Services electronically.

### TJAD qualifications & constraints

The process of collecting and matching skills resulted in a fairly significant class of entries that could not be identified as skills, but was important information for the advertisement. This results in two new distinctions:

#### Qualifications

This category brings together all requirements that need a certification, document, diploma. These are considered necessary in the context of a job search, but cannot be included in the skills category.

**Table 17: Qualifications**

Qualifications
Απολυτήριο λυκείου <i>High school diploma</i>
Εκπλήρωση στρατιωτικών υποχρεώσεων <i>Army obligations fulfilled</i>
Προπτυχιακός/ μεταπτυχιακός τίτλος σπουδών <i>Bachelors / masters degree</i>
Πιστοποίηση γλωσσομάθειας IELTS

<i>IELTS language certification</i>
Άδεια άσκησης επαγγέλματος <i>Professional licence</i>
Δικαίωμα υπογραφής <i>Right to sign</i>
Κάρτα ανεργίας <i>Unemployment card</i>

#### Constraints

The constraints set by the employer in the context of looking for an employee for a job are grouped under the category of constraints. Similar to qualifications, they are not skills, but elements important for the recruitment process.

**Table 18: Constraints**

<b>Constraints</b>
Λευκό ποινικό μητρώο <i>Clean criminal record</i>
Κάτοικος (Αττικής) <i>Resident (Attica)</i>
Διαθεσιμότητα για επαγγελματικά ταξίδια <i>Availability for business travel</i>
Εμπειρία χ χρόνια <i>Experience x years</i>
Μερικής απασχόλησης, ολικής απασχόλησης <i>Full-time, part-time</i>
Εργασία από το σπίτι <i>Work from home</i>
Βάρδιες <i>Working in shifts</i>

#### TJAD soft skills & hard skills

Soft skills are character traits and interpersonal skills that characterize a person's relationships with other people. In the workplace, soft skills are considered to be a complement to hard skills, which refer to a person's knowledge and occupational skills. It appears that the Greek labour market considers soft skills to be really important, thus the number of soft skills in an advertisement usually exceeds the number of hard skills. To improve the matching mechanism, it is important to separate soft and hard skills, as soft skills are transversal and can complicate the matching process.

**Table 19: Soft skills**

<b>Soft skills</b>
Επικοινωνία <i>Communication</i>
Επίλυση προβλημάτων <i>Problem solving</i>
Αναλυτική σκέψη <i>Analytical thinking</i>

Προσοχή στην λεπτομέρεια  
*Attention to detail*



## 5 EVALUATION

### 5.1 Quantitative results

The data obtained from the matching process was passed through the human cross examination process. In particular, 5420 occupations were manually checked, with their 25365 skills. The following results are thus obtained:

Regarding the occupations, in the 5420 advertisements checked, for 5007 of them a match with the ESCO ontology was found. There were also 111 additions of occupations to ESCO. Finally, of the total number of occupations examined, 302 were rejected. These were unclear jobs, jobs where there was an inability to understand the occupation, poor job description, very specific occupations not worth adding to ESCO, jobs in other languages, educational programs/scholarships/internships that are not occupations.

**Table 20: Quantitative results, occupations**

	<b>Job titles</b>
<b>Kariera occupation descriptions examined</b>	5420
<b>Found in ESCO</b>	5007
<b>Added in ESCO</b>	111
<b>Kariera occupation descriptions rejected</b>	302

The 25365 skills corresponding to the 5420 ads examined are divided into the following categories. The 13376 skills were successfully matched to the ESCO ontology. 575 new skills were added that were considered worthy of addition due to their frequency of occurrence. The entries rejected by the matching process are 10783. Finally, 161 qualifications and 288 constraints were added from the improvements made.

**Table 21: Quantitative results, skills**

	<b>Skill titles</b>
<b>Kariera skill descriptions examined</b>	25365
<b>Found in ESCO</b>	13376
<b>Added in ESCO</b>	575
<b>Kariera skill descriptions rejected</b>	10783
<b>Added as qualifications</b>	161
<b>Added as constraints</b>	288

### Evaluation of matching tool results

To calculate the performance of the matching mechanism, we use the mean reciprocal rank (MRR) and mean average precision (MAP) metrics. These metrics are used because the result of the application of the matching mechanism is a calibrated ranking of matches, and at any given time, especially in the skills, there is more than one correct match with ESCO.

## Mean reciprocal rank (MRR)

The mean reciprocal rank is a statistical measure for evaluating any procedure that produces a list of possible answers to a sample of questions, ranked by probability of correctness. The mutual ranking of a question answer is the multiplicative inverse of the ranking of the first correct answer: 1 for first position, 1/2 for second position, 1/3 for third position, and so on. The mean mutual ranking is the average of the mutual rankings of the results for a sample of questions  $q$ :

$$\text{MRR} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{\text{rank}_i}$$

Where  $\text{rank}_i$  refers to the ranking position of the first relevant document for the  $i$ -th query. The reciprocal value of the mean reciprocal ranking corresponds to the harmonic mean of the rankings.

## Mean average precision (MAP)

The mean average precision (MAP) for a set of queries is the average of the average precision scores for each query.

$$\text{MAP} = \frac{\sum_{q=1}^Q \text{AveP}(q)}{Q}$$

From our evaluation of the system, we draw conclusions. The results for the exact matches are quite high both in the case of occupations and skills, which gives a good chance that the results obtained are correct.

Table 22: MRR and MAP scores

Occupations	MRR	MAP	Support
All	0.403	0.427	5019
Exact matches	0.674	<b>0.716</b>	2178

Skills	MRR	MAP	Support
All	0.488	0.539	13222
Exact matches	0.716	<b>0.794</b>	7373

## 5.2 Qualitative results

For the quality check of the matching tool, it is tested on evaluation data of TJAD. The aim is to show the reliability of matching on new data.

Results of 1524 job advertisements were examined. The results were divided into correct, errors related to the matching tool and errors related to the job ads.

**Table 23: Qualitative results**

	Count
Total job posts	1523
Correct matching	1089
Matching tool errors	314
Job post errors	120

The category of correct matching includes all the ads that brought correct results regarding the occupation title and its skills. Matching tool errors stem from shortcomings in the matching mechanism, which have to do with specific entries that the system has not been able to predict. The job post errors are errors that are either too generic or too specific to match to the data.

E.g., correct matching

Βασική γνώση εργατικής νομοθεσίας (TJAD evaluation data) → εργατική νομοθεσία (ESCO data)

*Basic knowledge of labour law (TJAD evaluation data) → labor law (ESCO data)*

Matching tool errors

Χρήση η.υ., use of p.c. → hy, h/y (pc,p/c) are existing alternative labels. In the TJAD data there was no apparent need to create the h.y.(p.c.) Tag, so the machine failed to recognize it.

Job post errors

Good designing skills → *designing could mean architectural design, web design etc. When there is not a specification the machine can not match with a correct skill.*

The problems related to our matching tool can be fixed by continuously updating alternative labels and keywords to extract “noise” from the ads. In particular, out of the 314 incorrect matches, 95 can be corrected by using new alternative labels. The remaining 220 are noisy entries, i.e. Things that the extraction should not bring about.

## 6 CONCLUSION - FUTURE WORK

### 6.1 Conclusion

The conclusion that emerges from all the research is that the work of matching ads in an occupational and skills ontology is a continuous and evolving task. The work carried out as part of this thesis covers the data that exists today, in 2022. The potential of such a tool is important for the evolution of the labour market space and could greatly simplify the job search process. There are of course improvements that would help to refine the tool at a later stage. As mentioned above, this process is not static but constantly evolving, depending on the needs of the labour market. Thus, this project is only the beginning of a larger concept that could be formed.

### 6.2 Future work

#### 6.2.1 Skills to occupation title relation

One idea for improving the matching mechanism suggests linking the occupational title to the skills of the occupation. In the ESCO ontology, as already mentioned, each occupation includes its necessary and optional skills. Using this feature, the proposed mechanism takes the skills of a job post of TJAD, and from the set of them, identifies the occupations that include these skills as necessary skills. This implementation can significantly help with observed problems such as language ambiguities.

Occupation: software architect

Skills: computer programming, debug software, develop software prototypes

Current matching tool:

Software architect → architect, (100% match to ESCO)

Software developer, altlabel software architect,  
(50% match to ESCO altlabel)

Proposal:

Computer programming		→	software developer
Debug software			
Develop software prototypes			

#### 6.2.2 Profession to education linking

The pillars of the cv are usually those of education, work experience and skills that characterise each worker. In the work so far, an attempt has been made to organize

and correlate the occupations and skills of an employee. It follows, therefore, that a subsequent step would be to add the education domain to the matching effort.

Correlating occupations with education is a development that will enrich the overall system by bringing about improvements in individual digital applications in the workplace. More specifically, we refer to the separation of the education sectors, firstly into secondary and tertiary education and then to individual distinctions of each category. In the secondary education sector, a distinction is made between general high school education and vocational high school education, with the corresponding categories. Higher education has a greater separation, since there are several directions. The aim is to concentrate the skills acquired after completing an undergraduate/postgraduate course.

Then each degree with its corresponding skills, the process of matching a job with a particular employee is simplified, since in addition to the information entered from the skills and experience reported in the employee's cv, the information from the education received by the employee will be added.

### **6.2.3 Identification of not politically correct terms**

One of the major problems that could not be addressed in the context of the work is the use of non-politically correct terms. This could be improved by creating a suitable digital service that would complement the matching tool. Its aim would be to create a mechanism to identify these terms using keywords. Thus in the event of such an entry existing within the skill set of an advertisement, the non-politically correct entry could be flagged and the author of the advertisement could be notified that it is not considered acceptable.

## **BIBLIOGRAPHY**

- [1] DIRECTORATE-GENERAL FOR EMPLOYMENT, Social Affairs And Inclusion (European Commission), ESCO Handbook, ESCO Handbook, European Skills, Competences, Qualifications And Occupations.
- [2] DIRECTORATE-GENERAL FOR EMPLOYMENT, SOCIAL AFFAIRS AND INCLUSION (European Commission), ESCO Services Data Model.