AGORA Research Initiative

Analysis of Online Job Vacancies

An initial approach – Status & Next Steps

01

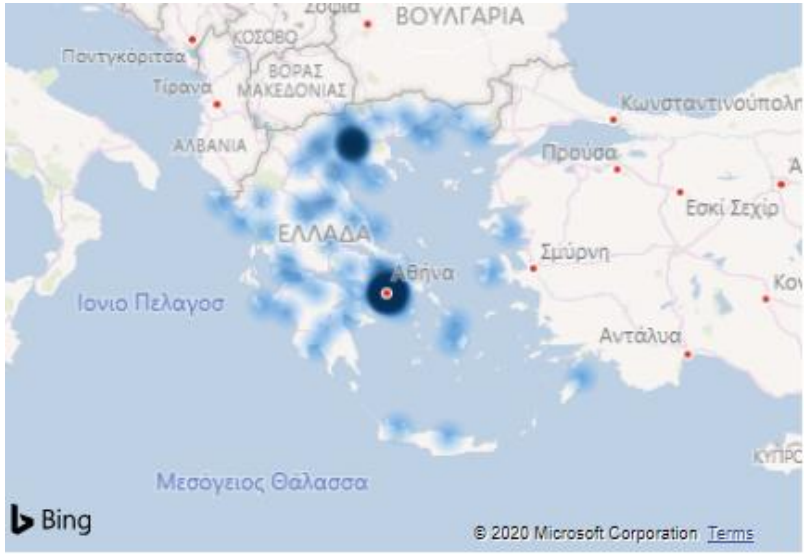
Objectives

Collect and Analyze Data from Online Job Portals to

- Diagnose market trends near real-time
- Identify occupations in demand
- Identify skills in demand

- Support evidence –based policy making
- Bridge the gap between universities and the private sector
- Inform curricula design
- Much more……

Current Setting



Data

- Data Collection from 4 major job vacancies Greek portals since July 2019.
- About 72K postings till now.

Analyze and Visualize Data

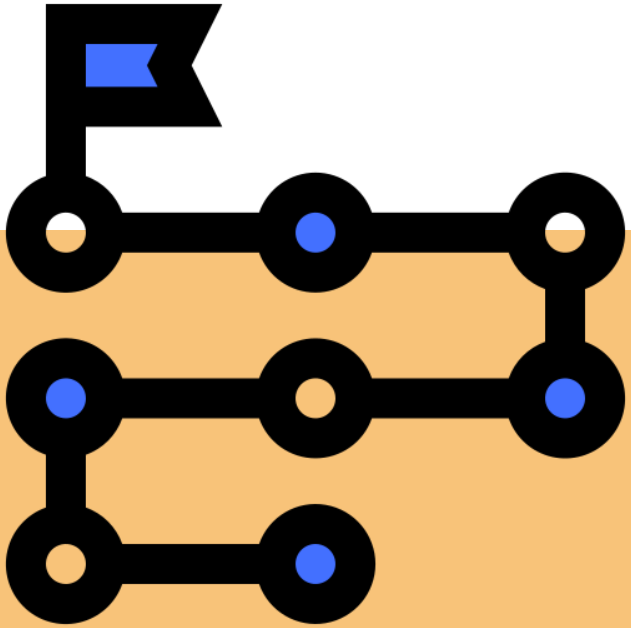
- Use of Machine Learning throughout the data analysis and Business Intelligence to visualize the results.

Obstacles

- Job vacancies duplicates.
- Machine learning model to be adapted to various languages needs and to project specifics.
- Differences between the structure of data in the various job portals.
- Setup of a common data model easily expandable and effective in relation to the various queries.
- Skills extraction without the use of an existing standard (e.g. ESCO).

02

Roadmap



A High-level RoadMap

- Python
- Scrapy FrameWork
- Portia

- Linux Server
- MariaDB

- Regular Expressions
- Machine Learning (Spacy Framework)
- ESCO API

- Regular Expressions
- Machine Learning (Spacy Framework)
- Store to MariaDB

- Machine Learning
- Microsoft Power Bi
- MariaDB

1\2 Useful Data Extraction

Creation of a Database Table, with the use of storing model key-value.

#	Όνομα	Τύπος	Σύνθεση	Χαρακτηριστικά	Κενό	Προεπιλογή	Σχόλια	Πρόσθετα
1	id	int(11)			Όχι	*Καμία*		AUTO_INCREMENT
2	**ulr**	text	utf8_general_ci		Όχι	*Καμία*		
3	**meta_key**	text	utf8_general_ci		Όχι	*Καμία*		
4	**meta_value**	text	utf8_general_ci		Όχι	*Καμία*		
5	**date**	datetime			Όχι	*Καμία*		
6	**Updated**	tinyint(4)			Όχι	0		

Cleansing of the Data

Machine Learning

- **Duplicate Detection** (Machine Learning to identify duplicates)
- **Location Validation** (Extraction of information from the job description with Entity Extraction)
- **Date Validation** (Update of all missing values from the original crawled date)
- **Employment Type Characterization** (With the use of machine learning from the job description)

06 Data Analysis and Visualization

- **Data Analysis** with the use of machine learning techniques
- **Data Visualization** with of Microsoft Power BI

Thank you!

For more information please contact
Giannis Tzimas
Giannis (at) Tzimas (dot) Info