

Real-time labour market information: Skills requirements analysis

Validation of working prototype

30 November – 1 December 2015, CEDEFOP, Thessaloniki, Greece

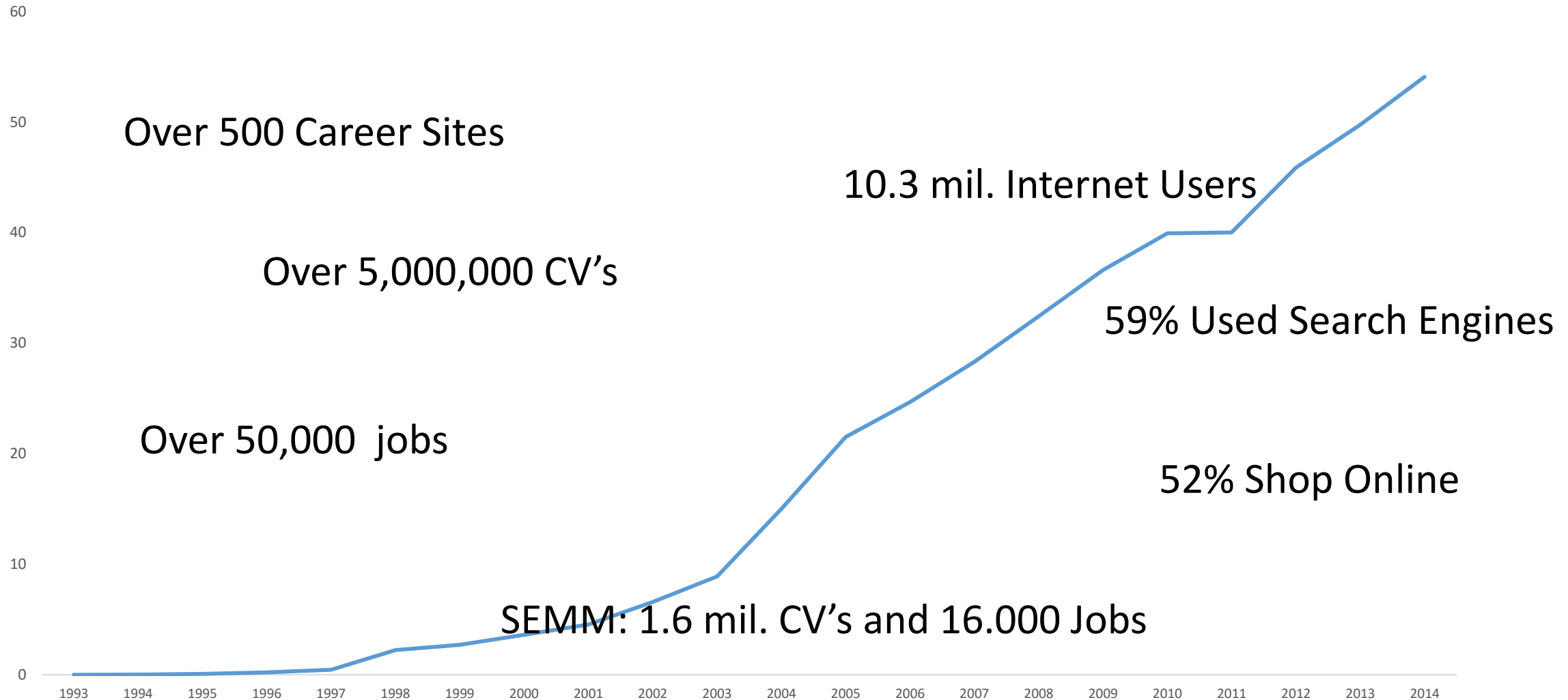
Analysing Labour Demand and Supply Using Web Mining and Data Mining

A case study on Romanian Labour Market

Ciprian Panzaru and Claudiu Brandas

West University of Timisoara, Romania

Context



Why Data Mining?

Because:

- There is a huge (virtually infinite) amount of data;
- Present technological development allow the automatic extraction of knowledge from large databases.

Allow us to:

- Classification;
- Clustering;
- Discovering the Hidden Patterns;
- Modeling;
- Build forecasting models.

Labor Market Analysis

in the Big Data Era

Big Data is an important source for information about the Labour Market.

Labour Market Analysis in the Big Data context using **Web Mining** and **Data Mining**:

- Labor demand profile based on the characteristics of the employment ads;
- Labor supply profile based on the characteristics of the users (visitors) and job searchers.

Web Mining

- **Web content mining**
- Web structure mining
- Web usage mining

Research Framework

- Quantitative analysis;
- Based on employment ads published by the National Agency for Employment (ANOFM) via the Electronic Labour Mediation Service (SEMM);
- Volume of analysed data:
 - ✓ 16,827 job ads (representing labour demand);
 - ✓ 18,511 CVs (representing labour supply);
- Web content mining for extraction and structuring of data (using import.io);
- Data mining clustering techniques (Berkhin 2006) using the simple k-means algorithm (using WEKA);
- Data Spatialisation in GIS (using Google Table Fusion).

Research Model



Web Content Mining – Process 1

Web Content Mining process using **Import.io**

- **job_demand_data_set** table with the following attributes (*location, job_demand_description*): table contain data about job demand.
- **job_supply_data_set** table with the following attributes (*location, job_supply_description*): table contain data about job supply.

Web Content Mining – Process 1

Web Content Mining process using Import.io

The screenshot shows a web browser window with the Import.io extension installed. The browser is displaying a page titled "ANOFM - ofertele de locuri de munca disponibile la data de". The page content includes a search bar, a list of job offers, and a sidebar with a list of counties (judete).

Cautare dupa ocupatia

Numar total oferte : 5171 Total locuri vacante : 16827

Cod	Judet	Ocupatie	Nr. posturi	Tip contract
01/03/1150030002/1	ALBA	muncitor necalificat la asamblarea, montarea pieselor	2	Contract de munca pe perioada nedeterminata cu norma intreaga
01/07/1150030015/1	ALBA	ajutor ospatar	2	Contract de munca pe perioada determinata cu norma intreaga
01/02/1150030062/1	ALBA	lacatus mecanic	1	Contract de munca pe perioada nedeterminata cu norma intreaga
01/02/1150030062/2	ALBA	electrician de intretinere si reparatii	1	Contract de munca pe perioada nedeterminata cu norma intreaga
01/02/1150030062/3	ALBA	inginer mecanic	1	Contract de munca pe perioada nedeterminata cu norma intreaga
01/02/1150030062/4	ALBA	consilier/expert/inspector/referent/economist in gestiunea economica	1	Contract de munca pe perioada nedeterminata cu norma intreaga
01/03/1150030111/1	ALBA	pedagog de recuperare	1	Contract de munca pe perioada nedeterminata cu norma intreaga
01/03/1150030115/2	ALBA	ospatar (chelnar)	3	Contract de munca pe perioada nedeterminata cu norma intreaga
01/03/1150030119/1	ALBA	manipulant marfuri	1	Contract de munca pe perioada nedeterminata cu norma intreaga
01/03/1150030120/1	ALBA	sofer de autoturisme si camioane	1	Contract de munca pe perioada nedeterminata cu norma intreaga
01/06/1150030125/1	ALBA	sofer autocamion/masina de mare tonaj	1	Contract de munca pe perioada determinata cu norma intreaga
01/06/1150030125/2	ALBA	confactioner produse pirotehnice	2	Contract de munca pe perioada determinata cu norma intreaga
01/06/1150030125/3	ALBA	presator metale la rece	2	Contract de munca pe perioada determinata cu norma intreaga
01/06/1150030125/4	ALBA	lacatus mecanic	1	Contract de munca pe perioada determinata cu norma intreaga
01/06/1150030126/1	ALBA	muncitor necalificat in metalurgie	20	Contract de munca pe perioada nedeterminata cu norma intreaga
01/06/1150030126/2	ALBA	lacatus mecanic	4	Contract de munca pe perioada nedeterminata cu norma intreaga
01/06/1150030126/4	ALBA	operator la masini-unelte cu comanda numerica	5	Contract de munca pe perioada nedeterminata cu norma intreaga
01/06/1150030127/1	ALBA	muncitor necalificat in metalurgie	10	Contract de munca pe perioada nedeterminata cu norma intreaga
01/06/1150030127/2	ALBA	lacatus mecanic	4	Contract de munca pe perioada nedeterminata cu norma intreaga
01/06/1150030127/4	ALBA	operator la masini-unelte cu comanda numerica	10	Contract de munca pe perioada nedeterminata cu norma intreaga

Cauta dupa judet

- BUCURESTI
- ALBA
- ARAD
- ARGES
- BACAU
- BIHOR
- BISTRITA-NASAUD
- BOTOSANI
- BRASOV
- BRAILA
- BUZAU
- CARAS-SEVERIN
- CALARASI
- CLUJ

Record a query

Hit stop when you are done.
If the website doesn't work, hit the "Not Working" button.

Not working? Recording

LMV ANOFM

Cauta

LMV ANOFM

Data Mining Analyses – Process 2

WEKA and Simple K Means clustering algorithm

For **job_demand_data_set**:

- six clusters with one instance for each cluster representing the cluster centroid (mean vectors for each cluster).
- numbers and percentage of instances for each cluster.

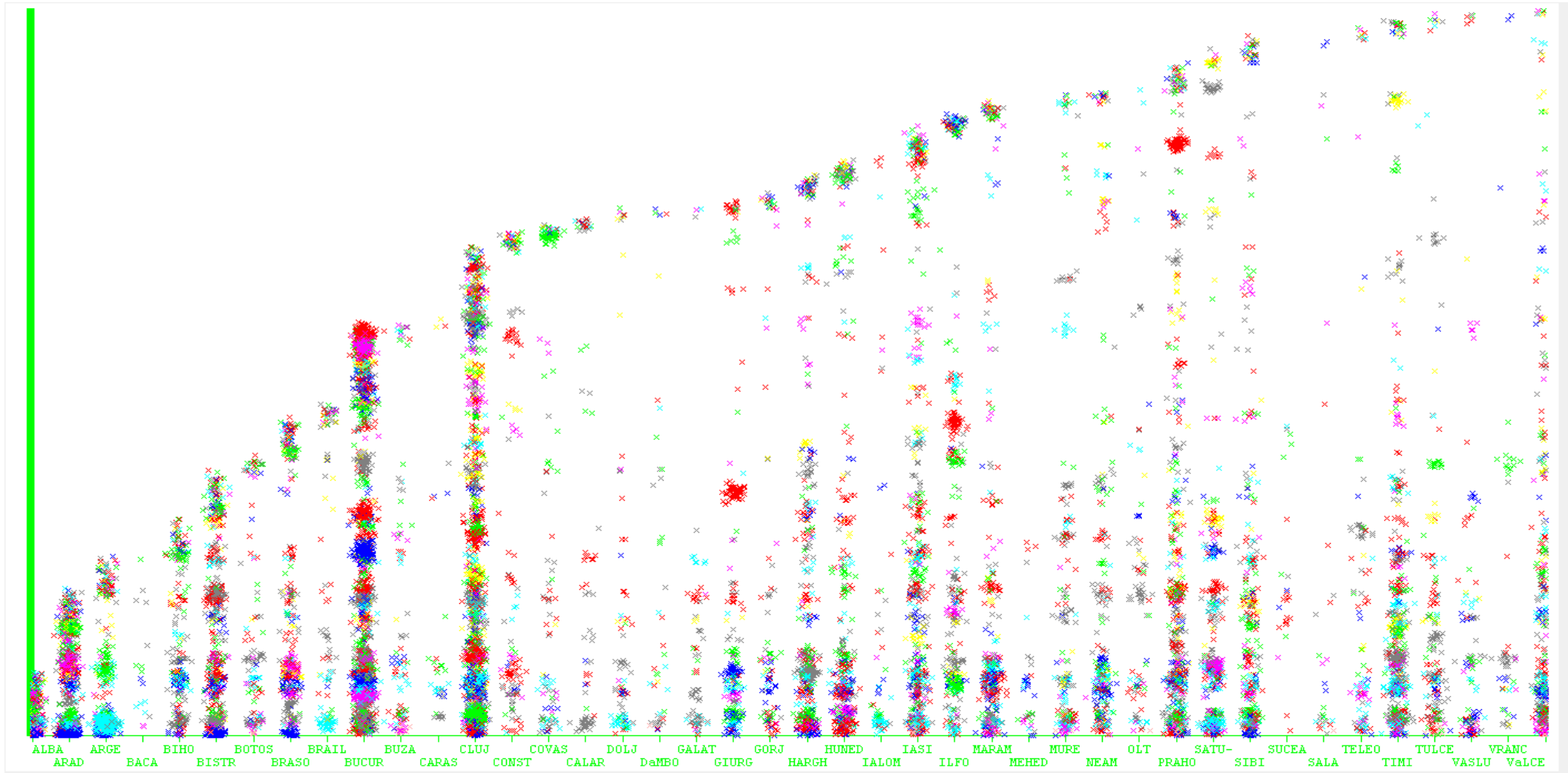
Data Mining Analyses – Process 2

WEKA and Simple K Means clustering algorithm

For the **job_supply_data_set**:

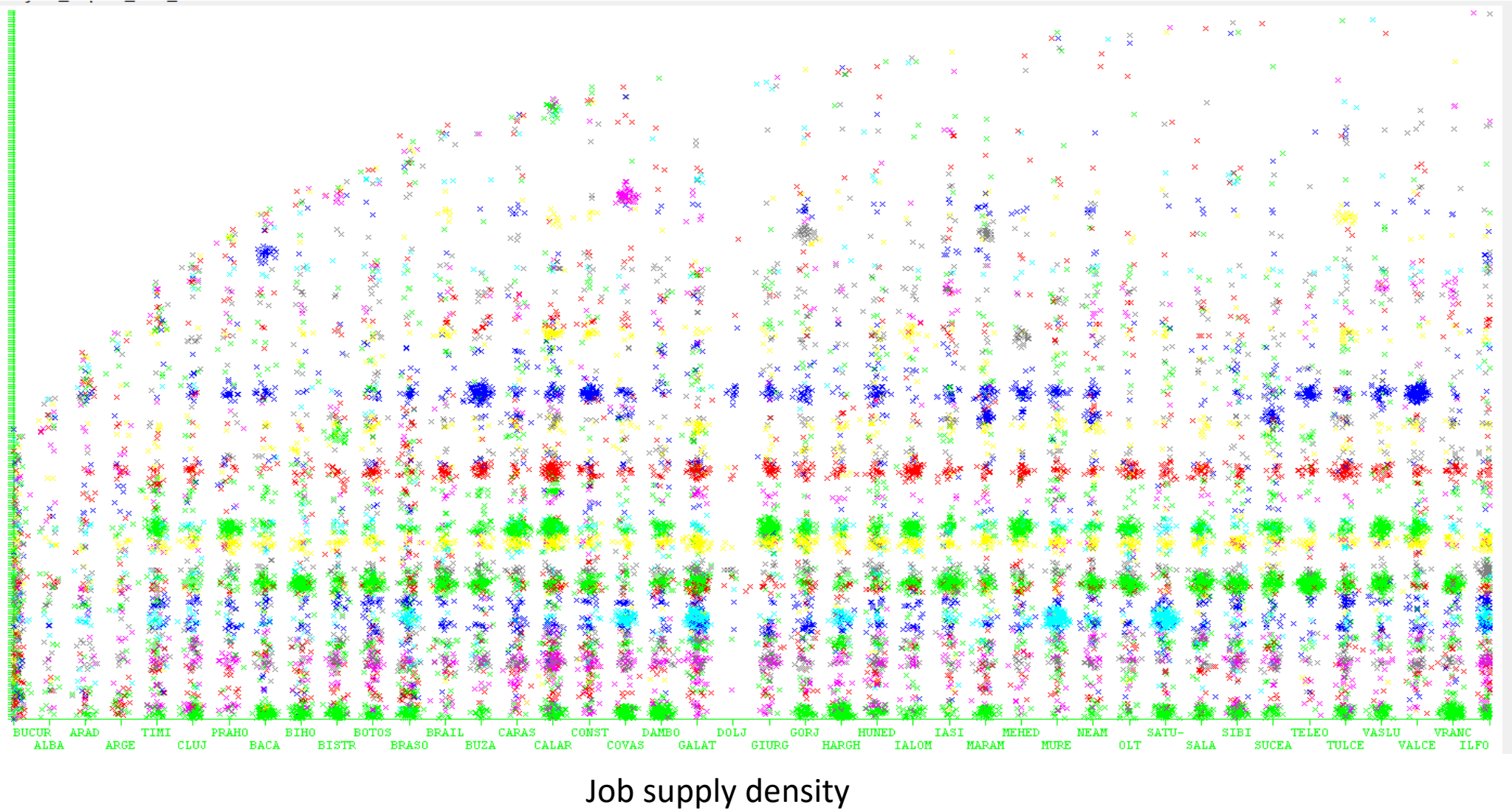
- six clusters with one instance for each cluster
- numbers and percentage of instances for each cluster.

Results – Process 2

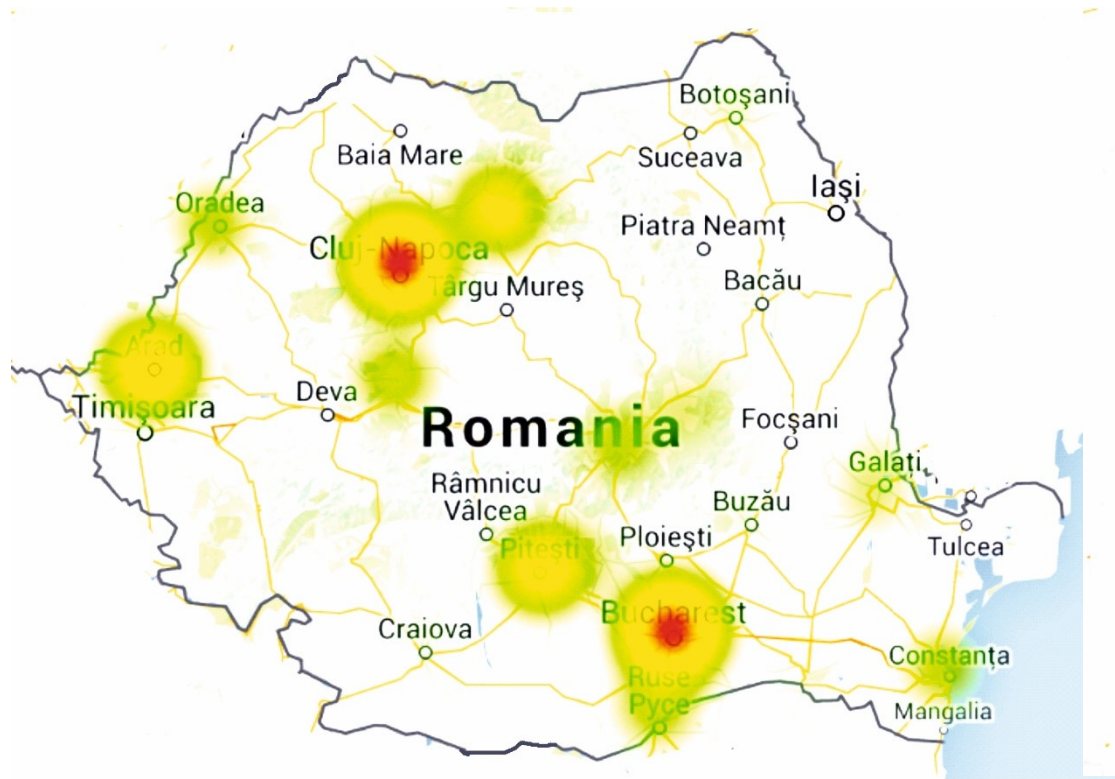


Job demand density

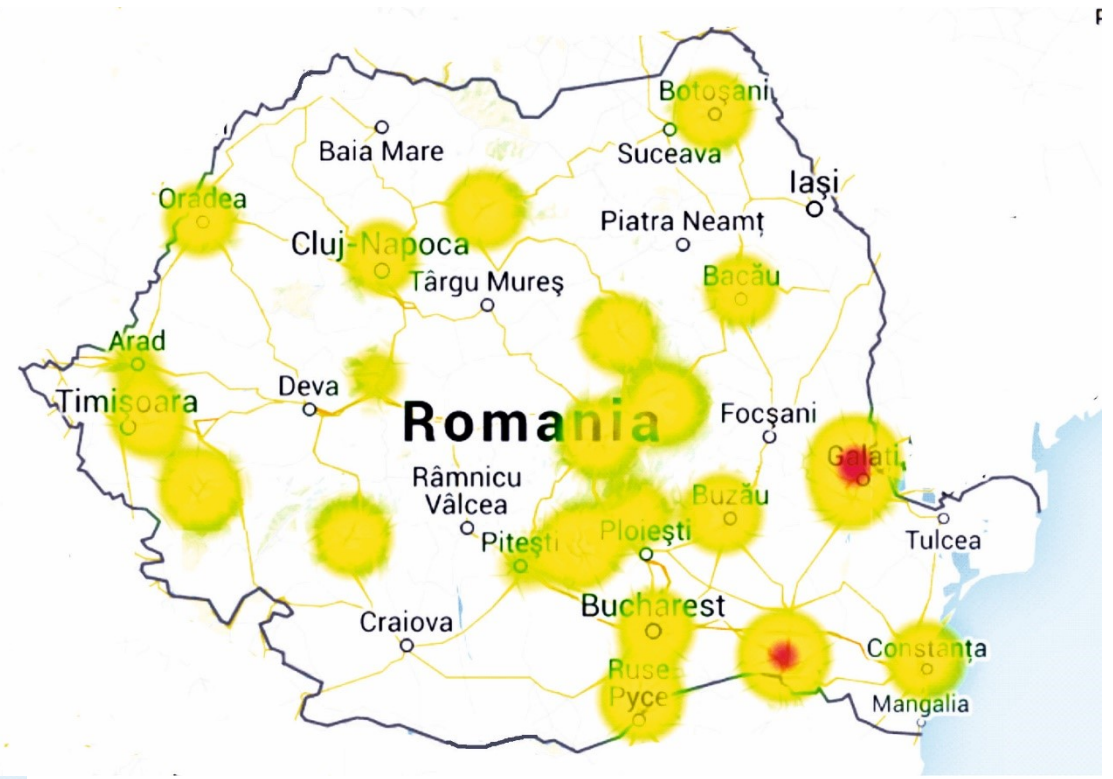
Results – Process 2



Spatialisation – Process 3



Job demand density



Job supply density

Conclusions

Job Demand

In Cluster #0, we have the most instances (64%), followed by the Cluster #5 (15%). In Cluster #4 we have the fewest instances (2%). For the representative Cluster #0 we have the cluster centroid formed by location “CLUJ” and job_demand_description “confectioner-asamblor articole din textile”. For the least representative Cluster #4 we have the cluster centroid formed by location „BRASOV” and job_demand_description “montator subansamble”.

Job Supply

In Cluster #0, we have the most instances (67%) followed by Cluster #5 with 11% and Cluster #3 with 10%. In Cluster #1 and Cluster #4 we have the fewest instances (3%). For the representative Cluster #0 we have the cluster centroid formed by location „GALATI” and job_supply_description “muncitori necalificati in ferme mixte”. For the least representative Cluster #1 we have the cluster centroid formed by location “COVASNA” and job_supply_description “dulgheri si tamplari”.

Conclusions

The Internet based technologies has generated in recent years huge volumes of diverse data both as source and as content. In the particular case of this research, web mining and data mining have helped to determine clusters for types of careers. These results can be further used to identify decision trees for modeling career paths.

We tried to show that web mining and data mining techniques can be used to obtain extremely valuable information on the labor market.

Advantages of using Data Mining in Analyzing Labor Market

Data mining solves a widespread paradox in classical statistical analysis: the more data there is to analyse, the more difficult it is to analyse and retrieve information.

If the traditional methods only modelling data (looking for a model, function), data mining involves the discovery of intelligible patterns expressed as trees, rules and data associations.

Data Mining is a technique by which huge amounts of data can be interpreted in such a way as to be easily read and understood.

Problems

- Lack of structural and semantic compatibility of data occurs frequently;
- Diversity of sources of information and attribute values.

Recommendations

- Using: Ontologies, Web Semantics and Web 3.0

Thank you!