Real-time Labour Market Information on Skill Requirements

Setting up the infrastructure for EU system

Text processing and information extraction

Alessandro Vaccarino – Mauro Pelucchi

Expert Workshop
Milan, 20-21 March 2018
Topics

1. Context
2. Text processing and multi-language environment
3. Classification processes, feature extraction and machine learning
4. Focus on occupation’s categorization
5. Focus on skill’s categorization
6. Preliminary results
Topics

1. Context
2. Text processing and multi-language environment
3. Classification processes, feature extraction and machine learning
4. Focus on occupation’s categorization
5. Focus on skill’s categorization
6. Preliminary results
Context

• Language
  • 25 official languages + 5 semi official languages

• High volumes of vacancies
  Millions of potential vacancies
Junior Software Developer

As Junior Software Developer, you will develop excellent software for use in field mapping, data collection, sensor networks, street navigation, and more. You will collaborate with other programmers and developers to autonomously design and implement high-quality web-based applications, restful API’s, and third party integration.

We’re looking for a passionate, committed developer that is able to solve and articulate complex problems with application design, development and user experiences. The position is based in our offices in Harwell, United Kingdom.

2512 – Software Developer

Skills: develop software, implement web based applications, problem solving, develop user experiences

Harwell, UK
Structured Data

You are the right one, if

- you are interested in designing, implementing and integrating specialized applications and you like to work with high-end development tools and infrastructures
- you like using state-of-the-art web technologies to realize customized software solutions while being able to expertly use standard products / SDKs as well as libraries
- You have a solid computer science education (HTL, FH or University) and solid programming skills with Java and Eclipse
- Ideally, you have already gained experience with SOAP, WSDL and Web Services, as well as object-oriented modeling (UML) skills

We offer you

- working in a dedicated, highly motivated team in a professional environment
- Consistent training and further education as well as broadening horizons by attending international conferences
- interesting development opportunities for Certified Project Manager, Certified Scrum Master etc.
- numerous benefits, benefits and cool events as well
- above-average payment (at least EUR 35,000 gross annual salary, depending on your experience and qualifications also more) and a performance-based bonus system

Contact Person: Mag. Dipl.-Ing. Christian Distelberger
E-mail: job@fabasoft.com
Tel: +43 732 606162-0
Topics

1. Context

2. Text processing and multi-language environment

3. Classification processes, feature extraction and machine learning

4. Focus on occupation’s categorization

5. Focus on skill’s categorization

6. Preliminary results
Preliminary results from ~1 million job ads

Structured data: 75%

Record linkage with NACE: 96%

Record not linked (unclassified): 4%
**Contract**

Preliminary results from ~1milions job ads

---

**StepStone**

**Fabasoft International Services GmbH**

**Application Developer (w / m)**

- Linz, Vienna
- Solid employment
- full-time
- Erschienen: vor 12 Tagen

**You are the right one, if**

- you are interested in designing, implementing and integrating specialized applications and you like to work with high-end development tools and infrastructures
- you like using state-of-the-art web technologies to realize customized software solutions while being able to expertly use standard products / SDKs as well as libraries
- You have a solid computer science education (HTL, FH or University) and solid programming skills with Java and Eclipse
- Ideally, you have already gained experience with SOAP, WSDL and Web Services, as well as object-oriented modeling (UML) skills

**We offer you**

- working in a dedicated, highly motivated team in a professional environment
- Consistent training and further education as well as broadening horizons by attending international conferences
- interesting development opportunities for Certified Project Manager, Certified Scrum Master etc.
- numerous benefits, benefits and cool events as well
- above-average payment (at least EUR 35,000 gross annual salary, depending on your experience and qualifications also more) and a performance-based bonus system

**Contact Person:** Mag. Dipl.-Ing. Christian Distelberger
**E-mail:** job@fabasoft.com
**Tel:** +43 732 86162-0

---

**Custom taxonomy**

**Reference**

Permanent
Self Employment
Temporary

**Structured data**

75%

**Record linkage with taxonomy**

83%

**Record not linked (unclassified)**

17%

---

**CEDEFOP**

European Centre for the Development of Vocational Training
Working Hours

Preliminary results from ~1 million job ads

- Custom taxonomy: 61%
- Structured data: 78%
- Record linkage with taxonomy: 78%
- Record not linked (unclassified): 22%
Educational Level

Preliminary results from ~1 million job ads

Reference

ISCED 2011

Structured data

48%

Record linkage with ISCED 2011

66%

Record not linked (unclassified)

34%
Preliminary results from ~1 million job ads

**Reference**
- 13 levels

**Structured data**
- 9%

**Custom taxonomy**
- 12%

**Record linkage with taxonomy**
- 88%

**Record not linked (unclassified)**
- 88%
Preliminary results from ~1 millon job ads

Experience

Structured data 45%

Reference 8 levels

Record linkage with taxonomy 45%

Record not linked (unclassified) 55%
Place

Preliminary results from ~1 million job ads

Reference

Structured data

48%

Record linkage with NUTS & LAU

99%

Record not linked (unclassified)

1%
Occupation

Preliminary results from ~1 million job ads

Reference
(ESCO 4th level)

ESCO v1
ISCO

Structured data

19%

Record linkage with ESCO/ISCO

100%

StepStone

Fabasoft International Services GmbH
Application Developer (w/m)

You are the right one, if

- you are interested in designing, implementing and integrating specialized applications and you like to work with high-end development tools and infrastructures
- you like using state-of-the-art web technologies to realize customized software solutions while being able to expertly use standard products / SDKs as well as libraries
- You have a solid computer science education (HTL, FH or University) and solid programming skills with Java and Eclipse
- Ideally, you have already gained experience with SOAP, WSDL and Web Services, as well as object-oriented modeling (UML) skills

We offer you

- working in a dedicated, highly motivated team in a professional environment
- Consistent training and further education as well as broadening horizons by attending international conferences
- interesting development opportunities for Certified Project Manager, Certified Scrum Master etc.
- numerous benefits, benefits and cool events as well
- above-average payment (at least EUR 35,000 gross annual salary, depending on your experience and qualifications also more) and a performance-based bonus system

Contact Person: Mag. Dipl.-Ing. Christian Distelberger
E-mail: job@fabasoft.com
Tel: +43 732 606162-0

CEDEFOP

European Centre for the Development of Vocational Training
Skill

Preliminary results from ~1milions job ads

ESCO + Custom

Structured data

10%

Record linkage with ESCO/custom taxonomy

92%

Record not linked (unclassified)

8%
Topics

1. Context
2. Text processing and multi-language environment
3. Classification processes, feature extraction and machine learning
4. Focus on occupation’s categorization
5. Focus on skill’s categorization
6. Preliminary results
Overall Data Flow - Recap

Data Ingestion → Pre-Processing → Information Extraction → ETL → Presentation Area
Information Extraction is an area of natural language processing that deals with finding factual information in free text.

This task uses machine learning techniques (ontology based learning, supervised learning and unsupervised learning) to match job ads with standard classifications.

Machine Learning → Ontology based learning, supervised learning and unsupervised learning, etc.
Information Extraction: analyse an unstructured document with the scope of extracting specific information.
Classification

Language Detector → Pre Processing → IT – Ontology-based Models → ES – Ontology-based Models → EN – Ontology-based Models
Classification

Language Detector → Pre Processing → IT – Ontology-based Models → IT – Machine Learning Models → Training set By ESCO Corpus IT → Classified Items and explanation

Language Detector → Pre Processing → ES – Ontology-based Models → ES – Machine Learning Models → Training set By ESCO Corpus ES → Classified Items and explanation

Language Detector → Pre Processing → EN – Ontology-based Models → EN – Machine Learning Models → Training set By ESCO Corpus EN → Classified Items and explanation
Topics

1. Context
2. Text processing and multi-language environment
3. Classification processes, feature extraction and machine learning
4. Focus on occupation’s categorization
5. Focus on skill’s categorization
6. Preliminary results
Goals

- **Classify occupation from title of job ads**
- Develop a generic approach valid for all 25 official languages of European Union (but also the co-official languages)
- Reduce the use of gold datasets with the scope to minimize the impact of human errors and ambiguities
- Design a system that is easily controllable and can be improved in case of misclassification: importance of explainable of outputs
- Create an occupations ontology

**Requirements**
- High precision
- High comprehensibility
Data and training set

- Multilanguage
- ~430 possible targets (ESCO Occupation Level 4)
- Large Gold Dataset for training ML models
  - For each language
    - 430 classes x 300 job ads ~ 120,000 manually labelled job ads
    - 300 most common classes x 300 job ads ~ 90,000 manually labelled job ads
First test

- Supervised learning
  - One model for each language (IT, DE, EN)
- Data labelled by expert from each country
  - ~35k job ads
  - 436 possible targets
- Evaluating set ~5,000 job ads
  - Weighted Precision ~30%
  - ~250 detected professions
Data Powers AI and Machine Learning

**AI models** can only learn from **diverse, high quality data**

More sophisticated models require even **more data** (high quality data!)

* e.g. **deep learning requires millions of documents!**
Chefs are culinary professionals with a flair for creativity and innovation to provide a unique gastronomic experience.
Ontology to improve classification

System

Engineer
ITC
Develop
Robotics

System

Plastic
Ceramic
Waste
Develop

2511 – System Engineer

Noise
Second test

- Ontology based learning + Supervised learning
  - Esco Ontology
  - One model for each language (IT, DE, EN)
- Data labelled by expert from each country
  - ~25k job ads (cleaned train set using ESCO Ontology)
  - 436 possible targets
- Evaluating set ~5.000 job ads
  - Weighted Precision ~50%
  - ~400 detected professions
Improve ontology
Topic modelling and unsupervised learning

SAP Consultant
Posted 4 days ago by Ascendant Recruitment

- Milton Keynes, Buckinghamshire
- £50,000 - £60,000 per annum
- Contract, full-time
- Be one of the first ten applicants

SAP Consultant
12 month contract
£50,000 - £60,000

Milton Keynes

We are delighted to be working with a prestigious client based in Milton Keynes for the recruitment of a qualified SAP consultant.

This is an exciting and challenging opportunity for an experienced SAP Consultant to join the e-business division of this large organisation.

The role will involve analysing business requirements as well as taking the lead in the design and delivery of solutions for various IT projects and business change programmes.
Final considerations

- Ontology based learning + Supervised learning
  - Esco Ontology
  - New labels from Topic modelling
- One model for each language (IT, DE, EN)
- Data labelled by expert from each country
  - ~25k job ads (cleaned train set using ESCO Ontology)
  - 436 possible targets
- Evaluating set ~5,000 job ads
  - Weighted Precision ~80%
  - ~423 detected professions
Topics

1. Context
2. Text processing and multi-language environment
3. Classification processes, feature extraction and machine learning
4. Focus on occupation’s categorization
5. Focus on skill’s categorization
6. Preliminary results
Fabasoft International Services GmbH

Application Developer (w / m)

- Linz, Vienna
- Solid employment
- full-time
- Erschienen: vor 12 Tagen

You are the right one, if

- you are interested in designing, implementing and integrating specialized applications and you like to work with high-end development tools and infrastructures
- you like using state-of-the-art web technologies to realize customized software solutions while being able to expertly use standard products / SDKs as well as libraries
- You have a solid computer science education (HTL, FH or University) and solid programming skills with Java and Eclipse
- Ideally, you have already gained experience with SOAP, WSDL and Web Services, as well as object-oriented modeling (UML) skills

We offer you

- working in a dedicated, highly motivated team in a professional environment
- Consistent training and further education as well as broadening horizons by attending international conferences
- interesting development opportunities for Certified Project Manager, Certified Scrum Master etc.
- numerous benefits, benefits and cool events as well
- above-average payment (at least EUR 35,000 gross annual salary, depending on your experience and qualifications also more) and a performance-based bonus system

Contact Person: Mag. Dipl.-Ing. Christian Distelberger
E-mail: job@fabasoft.com
Tel: +43 732 606162-0
You are the right one, if

- you are interested in designing, implementing and integrating specialized applications and you like to work with high-end development tools and infrastructures
- you like using state-of-the-art web technologies to realize customized software solutions while being able to expertly use standard products / SDKs as well as libraries
- You have a solid computer science education (HTL, FH or University) and solid programming skills with Java and Eclipse
- Ideally, you have already gained experience with SOAP, WSDL and Web Services, as well as object-oriented modeling (UML) skills

We offer you

- working in a dedicated, highly motivated team in a professional environment
- Consistent training and further education as well as broadening horizons by attending international conferences
- interesting development opportunities for Certified Project Manager, Certified Scrum Master etc.
- numerous benefits, benefits and cool events as well
- above-average payment (at least EUR 35,000 gross annual salary, depending on your experience and qualifications also more) and a performance-based bonus system

Contact Person: Mag. Dipl.-Ing. Christian Distelberger
E-mail: job@fabasoft.com
Tel: +43 732 606162-0
You are the right one if:

- you are interested in designing, implementing and integrating specialized applications and you like to work with high-end development tools and infrastructures
- you like using state-of-the-art web technologies to realize customized software solutions while being able to expertly use standard products / SDKs as well as libraries
- You have a solid computer science education (HTL, FH or University) and solid programming skills with Java and Eclipse
- Ideally, you have already gained experience with SOAP, WSDL and Web Services, as well as object-oriented modeling (UML) skills

We offer you

- working in a dedicated, highly motivated team in a professional environment
- Consistent training and further education as well as broadening horizons by attending international conferences
- interesting development opportunities for Certified Project Manager, Certified Scrum Master etc.
- numerous benefits, benefits and cool events as well
- above-average payment (at least EUR 35,000 gross annual salary, depending on your experience and qualifications also more) and a performance-based bonus system

Contact Person: Mag. Dipl.-Ing. Christian Distelberger
E-mail: job@fabasoft.com
Tel: +43 732 606162-0
You are the right one if:

- you are interested in designing, implementing and integrating specialized applications and you like to work with high-end development tools and infrastructures
- you like using state-of-the-art web technologies to realize customized software solutions while being able to expertly use standard products / SDKs as well as libraries
- You have a solid computer science education (HTL, FH or University) and solid programming skills with Java and Eclipse
- Ideally, you have already gained experience with SOAP, WSDL and Web Services, as well as object-oriented modeling (UML) skills

We offer you:

- working in a dedicated, highly motivated team in a professional environment
- consistent training and further education as well as broadening horizons by attending international conferences
- interesting development opportunities for Certified Project Manager, Certified Scrum Master etc.
- numerous benefits, benefits and cool events as well
- above-average payment (at least EUR 35,000 gross annual salary, depending on your experience and qualifications also more) and a performance-based bonus system

Contact Person: Mag. Dipl.-Ing. Christian Distelberger
E-mail: job@fabasoft.com
Tel: +43 732 606162-0
You are the right one, if

- you are interested in designing, implementing and integrating specialized applications and you like to work with high-end development tools and infrastructures
- you like using state-of-the-art web technologies to realize customized software solutions while being able to expertly use standard products / SDKs as well as libraries
- You have a solid computer science education (HTL, FH or University) and solid programming skills with Java and PHP
- Ideally, you have already gained experience with SOAP, WSDL and Web Services, as well as object-oriented modeling (UML) skills

We offer you

- working in a dedicated, highly motivated team in a professional environment
- Consistent training and further education as well as broadening horizons by attending international conferences
- interesting development opportunities for Certified Project Manager, Certified Scrum Master etc.
- numerous benefits, benefits and cool events as well
- above-average payment (at least EUR 35,000 gross annual salary, depending on your experience and qualifications also more) and a performance-based bonus system

Contact Person: Mag. Dipl.-Ing. Christian Distelberger
E-mail: job@fabasoft.com
Tel: +43 732 606162-0
You are the right one, if

- you are interested in designing, implementing and integrating specialized applications and you like to work with high-end development tools and infrastructures
- you like using state-of-the-art web technologies to realize customized software solutions while being able to expertly use standard products / SDKs as well as libraries
- You have a solid computer science education (HTL, FH or University) and solid programming skills with Java and Eclipse
- Ideally, you have already gained experience with SOAP, WSDL and Web Services, as well as object-oriented modeling (UML) skills

We offer you

- working in a dedicated, highly motivated team in a professional environment
- Consistent training and further education as well as broadening horizons by attending international conferences
- interesting development opportunities for Certified Project Manager, Certified Scrum Master etc.
- numerous benefits, benefits and cool events as well
- above-average payment (at least EUR 35,000 gross annual salary, depending on your experience and qualifications also more) and a performance-based bonus system

Contact Person: Mag. Dipl.-Ing. Christian Distelberger
E-mail: job@fabasoft.com
Tel: +43 732 606162-0
You are the right one, if

- you are interested in designing, implementing and integrating specialized applications and you like to work with high-end development tools and infrastructures
- you like using state-of-the-art web technologies to realize customized software solutions while being able to expertly use standard products / SDKs as well as libraries
- You have a solid computer science education (HTL, FH or University) and solid programming skills with Java and Eclipse
- Ideally, you have already gained experience with SOAP, WSDL and Web Services, as well as object-oriented modeling (UML) skills

We offer you

- working in a dedicated, highly motivated team in a professional environment
- Consistent training and further education as well as broadening horizons by attending international conferences
- interesting development opportunities for Certified Project Manager, Certified Scrum Master etc.
- numerous benefits, benefits and cool events as well
- above-average payment (at least EUR 35,000 gross annual salary, depending on your experience and qualifications also more) and a performance-based bonus system

Contact Person: Mag. Dipl.-Ing. Christian Distelberger
E-mail: job@fabasoft.com
Tel: +43 732 606162-0
Goals

- Extract skills from description of job ads (and other metadata fields, e.g. skills, drive licensing, languages, …)
  - Develop a generic approach valid for all 25 official languages of European Union (but also the co-official languages)
- Identify noise
- Use alternative labels from common job ads languages (and industry de-facto standard)
- Record linkage between extracted skills and ESCO Ontology
- Design a system that is easily controllable and can be improved in case of misclassification: importance of explainable of outputs
Text Similarity Approaches

String based
String similarity measures operate on string sequences and character composition.
Jaro-Winkler, Jaccard, Cosine similarity

Corpus based
Corpus-Based similarity is a semantic similarity measure that determines the similarity between words according to information gained from large corpora.
Latent Semantic Analysis, Explicit Semantic Analysis, DIStributionally similar words using CO-occurrences

Knowledge based
Knowledge-Based Similarity is based on identifying the degree of similarity between words using information derived from semantic networks
Text Similarity Approaches

Example

You are the right one, if

- you are interested in designing, implementing and integrating specialized applications and you like to work with high-end development tools and infrastructures
- you like using state-of-the-art web technologies to realize customized software solutions while being able to expertly use standard products / SDKs as well as libraries
- You have a solid computer science education (HTL, FH or University) and solid programming skills with Java and Eclipse
- Ideally, you have already gained experience with SOAP, WSDL and Web Services, as well as object-oriented modeling (UML) skills

object-oriented modelling

use object-oriented programming

unified modelling language
Text Similarity Approaches

Example

String based

String similarity measures operate on string sequences and character composition.

Jaro-Winkler, Jaccard, Cosine similarity

You are the right one, if

- you are interested in designing, implementing and integrating specialized applications and you like to work with high-end development tools and infrastructures
- you like using state-of-the-art web technologies to realize customized software solutions while being able to expertly use standard products / SDKs as well as libraries
- You have a solid computer science education (HTL, FH or University) and solid programming skills with Java and Eclipse
- Ideally, you have already gained experience with SOAP, WSDL and Web Services, as well as object-oriented modeling (UML) skills

Object-oriented modelling

Jaro Winkler

~0.26

use object-oriented programming

Jaro Winkler

~0.5

unified modelling language
Corpus-based similarity is a semantic similarity measure that determines the similarity between words according to information gained from large corpora.

Latent Semantic Analysis, Explicit Semantic Analysis, DIStributionally similar words using CO-occurrences

You are the right one, if

- you are interested in designing, implementing and integrating specialized applications and you like to work with high-end development tools and infrastructures
- you like using state-of-the-art web technologies to realize customized software solutions while being able to expertly use standard products / SDKs as well as libraries
- You have a solid computer science education (HTL, FH or University) and solid programming skills with Java and Eclipse
- Ideally, you have already gained experience with SOAP, WSDL and Web Services, as well as object-oriented modeling (UML) skills

Object-oriented modelling

Use object-oriented programming

Unified modelling language

CEDEFOP
European Centre for the Development of Vocational Training

AO/DSL/VKVET-GRUSSO/Real-time LMI 2/009/16
Knowledge-based Similarity is based on identifying the degree of similarity between words using information derived from semantic networks.

Text Similarity Approaches

Example

You are the right one, if

- you are interested in designing, implementing and integrating specialized applications and you like to work with high-end development tools and infrastructures
- you like using state-of-the-art web technologies to realize customized software solutions while being able to expertly use standard products / SDKs as well as libraries
- You have a solid computer science education (HTL, FH or University) and solid programming skills with Java and Eclipse
- Ideally, you have already gained experience with SOAP, WSDL and Web Services, as well as object-oriented modeling (UML) skills

- use object-oriented programming
- object-oriented modelling
- alternative labels
- unified modelling language
Topics

1. Context
2. Text processing and multi-language environment
3. Classification processes, feature extraction and machine learning
4. Focus on occupation’s categorization
5. Focus on skill’s categorization
6. Preliminary results
Multi sources, single country, single language

Multi sources, single country, multi languages
Skills categorization

**Multi sources, single country, single language**

Skills Distribution by Esco Level 0

Num Skill Distribution by Esco Level 0

Num Skill Vs Num Occupation

Num Skill vs Job Ads

- problem solving
- provide information
- technical drawings
- create solutions to problems
- foreign languages
- use a computer
- English
- adapt to change
- assist customers
- customer service

~1884 Skill
Skills categorization

Multi sources, single country, multi languages

Skills Distribution by Esco Level 0

Num Skill Vs Num Occupation

Num Skill Vs Job Ads

Spanish
quality standards
technical drawings
use office systems
use microsoft office
customer service
use a computer
foreign languages
English
adapt to change
Thanks for your attention