

1

CONTEXT

The Belgian job market is characterized by its multilingual job offers, especially in ICT.

EN

NL

FR

5 functions were targeted for data collection :

</>

Developer

cloud

DevOps

network

Network Engineer

shield

Security Engineer

monitor

System Engineer

2

REQUESTS

The training centre who ordered the study had three main questions/requests:

1

To describe the **job market** in terms of estimating number of OJA per function, what types of contracts were offered, where jobs were located.

2

To list the **job-related skills and certifications** asked by employers for each function.

3

To determine if **emerging or new functions** are **in demand** and not covered by the current training programme.

3

LEXICOMETRY PIPELINE

1

Data collection

Web scraping with Power Automate on LinkedIn and ICTjob.be

9,991 OJA collected

2

Data processing

Cleaning unrelated OJAS and duplicates, language detection, with Excel

3

Data analysis

Through lexicometry, with TXM and Excel

T

X

M

Lexicometry consists in identifying lexical terms within a corpus of texts in order to quantify the number of times a word is mentioned - namely, its occurrences - throughout a body of text.

4

RESULTS

1

1st question: **fully answered**, but needed manual observations on Excel.

2

2nd question: **answered only when specific questions were provided**. Skills intelligence proved difficult with lexicometry.

3

3rd question: **could not be answered** with lexicometry.

5

DISCUSSION

Advantages of lexicometry

- Analyze large amounts of text
- Ease to use, open-source
- Programmability to use for multilingual corpora

Disadvantages of lexicometry

- Lexical polymorphism
- Difficult to target skills, especially emerging skills, and in multilingual environments

6

MACHINE LEARNING PIPELINE

1

Data collection

Eurostat OJA database

2

Data processing

Classification models for relevance and duplicates. 3 options for multilingual OJA:

- language detection and split per language
- use multilingual model
- translate to English

3

Data analysis

Descriptive analysis: TF-IDF and bigrams

Skills analysis: NER, transformer-based NER, clustering for emerging jobs/skills.