SwissText 2026 · Oral Presentation

Skill Extraction from Resumes and Job Offers

across Six Languages

Laura Vásquez-Rodríguez (Doodle AG)*, Bertrand Audrin (EHL, HES-SO), Samuel Michel (Idiap), Samuele Galli (Arca24), Julneth Rogenhofer (EHL, HES-SO), Jacopo Negro Cusa (Arca24), Lonneke van der Plas (USi)

*Work done while at the Idiap Research Institute.







AGENDA

What we'll cover

- **01 Motivation:** The recruitment gap
- **02 Data:** 1,200 documents, six languages
- **03 End-to-end system:** methods, evaluation schema
- **04 Results:** what worked well? What was difficult?
- **05 Evaluation:** manual analysis
- **06 Takeaways:** what did we learned?

Global recruitment is multilingual

Innovation project supported by

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Innosuisse – Swiss Innovation Agency

Most methods in academic research are limited to English job offers, failing to capture the complexity of real, multilingual hiring.

- We evaluate multiple extraction methods **across six languages.**
  English, French, German, Italian, Spanish, Portuguese.
- We compare **resumes and job offers** across ~45 professional domains.
- The goal: a more comprehensive, **industry-relevant** framework.
- **Collaboration**: 2-year **Innosuisse project** working together with Arca24, EHL, Idiap Research Institute, and USI.

Human-annotated by HR specialists

1,200

TOTAL DOCUMENTS

100 job offers + 100 resumes per language

~45

PROFESSIONAL DOMAINS

10,379

SKILLS

IN-HOUSE taxonomy (previously collected)

SIX LANGUAGES

EN FR DE IT ES PT

HUMAN-IN-THE-LOOP

Manually annotated by HR specialists with the Docanno tool, mapped via the **KSAO**[1] framework, Knowledge, **Skills**, Abilities & Other attributes.

Conflicts were identified and analyzed together after a first round of annotation to ensure full **alignment** among the **team**.

[1]Campion et al. (2011). Doing competencies well: Best practices in competency modeling. *Personnel Psychology*, 64(1).

Publicly available datasets

~10,000

SENTENCES

Green Dataset (English UK jobs)

5+

PROFESSIONAL DOMAINS: IT, Finance, Healthcare, Sales

13,939

SKILLS

ESCO Taxonomy, 28 official languages

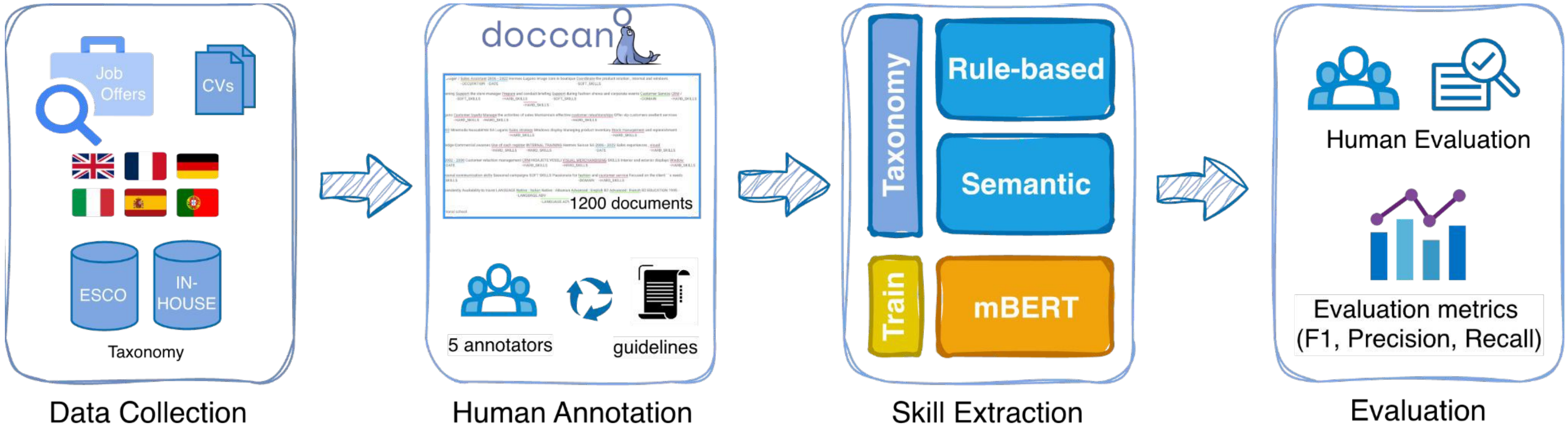
ONE LANGUAGE

EN

HUMAN-IN-THE-LOOP

Sentences annotated via Amazon Mechanical Turk (AMT) using 39 highly qualified workers + automatic post processing.

From documents to evaluation

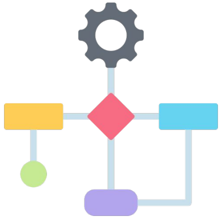


Three extraction methods

Rule-based

Taxonomy matching

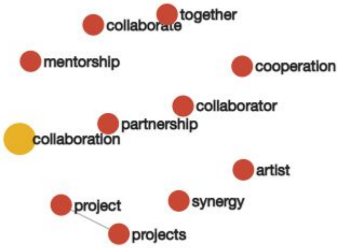
Strict matching with in-house & ESCO taxonomies. Fully transparent and easy to update, but poor at variation.



Semantic

Sentence embeddings

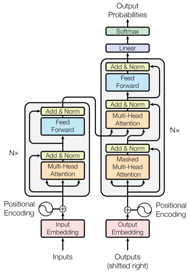
Finds closely related terms. Better generalisation, still traceable to the taxonomy, but can't learn new concepts.



Supervised mBERT

Fine-tuned model

Multilingual BERT trained on manual annotations. Captures context across languages, but is a black box.



03 EVALUATION

Exact and partial schema

Gold

Data analysis using Python and Pandas

Exact

Data analysis using Python and Pandas

Partial

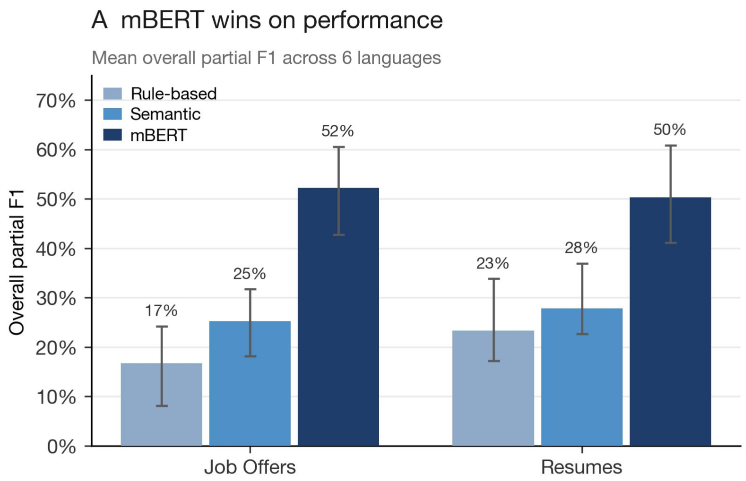
Data analysis using Python and Pandas

In a partial evaluation we give credit for partially detected skills such as in "Data" for "Data analysis"

Gold — human annotation

System — model extraction

mBERT wins, skill boundaries are difficult



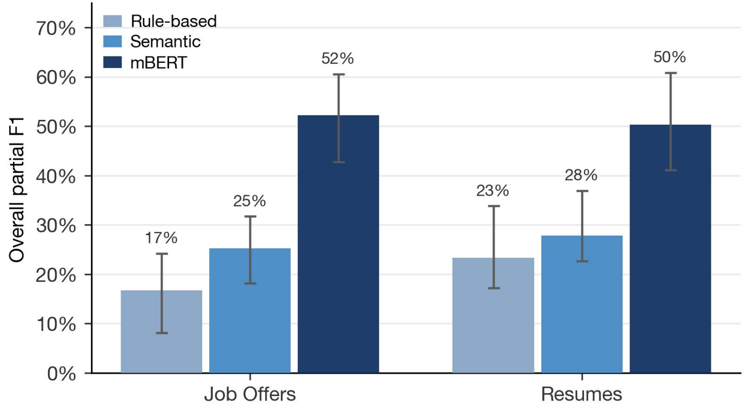
A · ON PERFORMANCE

Fine-tuned **mBERT** beats the rule-based and semantic systems by a wide margin (**52%** mean partial F1 on job offers and **50%** in resumes).

mBERT wins, skill boundaries are difficult

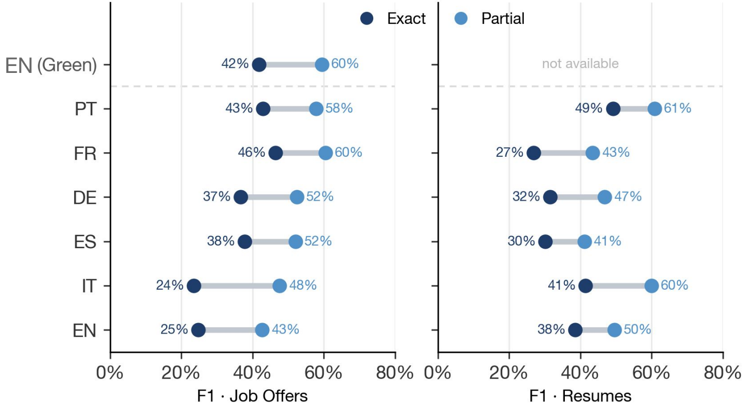
A mBERT wins on performance

Mean overall partial F1 across 6 languages



B Exact vs. partial scores

mBERT, per language — gap = boundary uncertainty



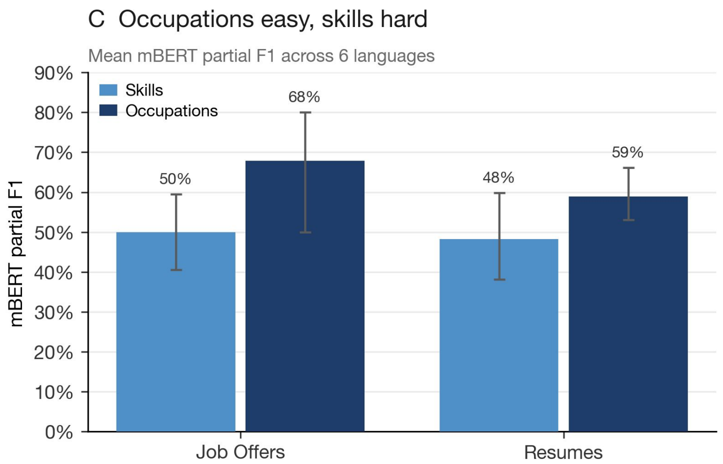
A · ON PERFORMANCE

Fine-tuned **mBERT** beats the rule-based and semantic systems by a wide margin (**52%** mean partial F1 on job offers and **50%** in resumes).

B · EXACT VS. PARTIAL

Partial scores exceed exact. The gap is shows **boundary uncertainty**. The model finds the skill, but not exactly where it begins and ends.

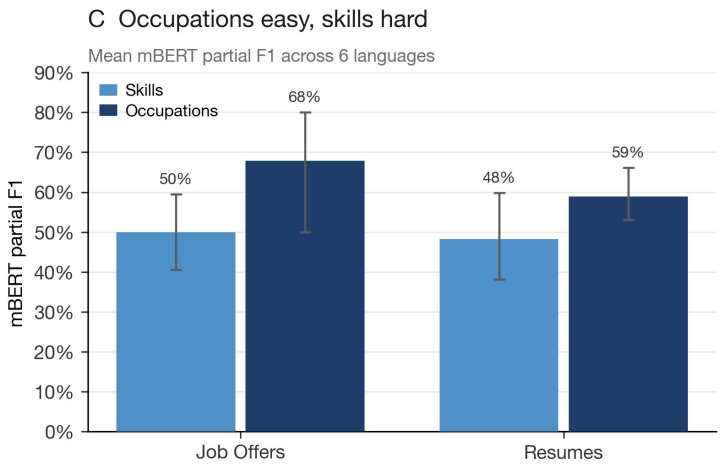
Easy occupations, hard skills

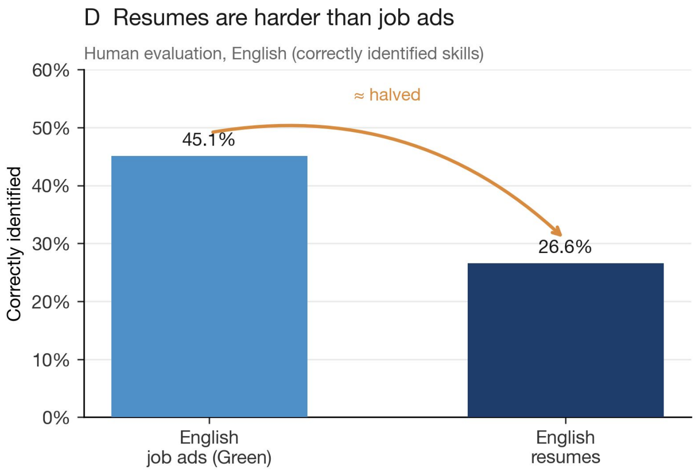


C · OCCUPATIONS VS. SKILLS

Job titles are easier, partial F1 up to **68%**. Finding the the exact **skill span** is still the bottleneck.

Easy occupations, hard skills





C · OCCUPATIONS VS. SKILLS

Job titles are easier, partial F1 up to **68%**. Finding the the exact **skill span** is still the bottleneck.

D · RESUMES VS. JOB ADS

Correctly-identified skills roughly **halve** from clean job ads to variable resumes **45.1%** → **26.6%** in EN.

Manual analysis

Job Offers (EN) **Semantic (in-house)**

Gold — An **analytical** approach to **problem solving** excellent **team working skills**

System — An **analytical approach** to **problem solving** excellent team working skills

Resumes (EN) **Rule-based (in-house)**

Gold — …different languages PHP HTML CSS Javascript jQuery SQL Visual Basic Linux Bash FileMaker Script C

System — …different languages PHP HTML CSS Javascript jQuery SQL Visual Basic Linux Bash FileMaker Script C

Gold — human annotation

System — model extraction

Three takeaways

01 · PERFORMANCE

mBERT beats rule-based and semantic systems by a wide margin, in exchange of a **"black box" model** that is harder to manipulate.

Alternative: map back to the taxonomy, curate new concepts.

02 · SKILLS ARE HARD

Job titles are easy to find. Knowing exactly where a **skill begins** and **ends** is still very **hard** for AI.

03 · RESUMES ≠ JOBS

Job offers are straight-forward; resumes are personal and non-standardized across many dimensions: **people**, **domains**, **language**.

Thank you!

Questions?

PROJECT & PARTNERS

Innovation project supported by

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Innosuisse – Swiss Innovation Agency