


## WageIndicator Foundation

- Founded in the Netherlands in 2000 by Paulien Osse and prof. Kea Tijdens.
- Provides and collects information on wages, labour law and career. It's a free global online library.
- Now in 196 countries.
- Databases: minimum wages, prices, living wages, actual wages, labour law, **collective agreements (CBAs)**.

 = documents containing conditions of employment that result from negotiations between independent unions and employers

[WageIndicator.org/cbadatabase](http://WageIndicator.org/cbadatabase)

 **1600+ agreements**

**61 countries**

**27 languages**

## THE ANNOTATION (supported by Clarin)

For every CBA, the annotators answer to 249 labour rights related questions on **9 topics** and select the relevant clauses.

- Training / apprenticeship
- Individual employment contracts / job security
- Social security and pensions
- Sickness and disability
- Health and medical assistance
- Work-life balance arrangements
- Equality and violence in the workplace
- Working hours, schedules, holidays and days of leave
- Wages

**1** Dataset **TEXTS**  
(.csv dump with all CBA texts in html)

### National websites

**2** Dataset **CLAUSES**  
(.csv dump with all clauses assigned to a question (= 'bind'))

## ML - BASED DEVELOPMENTS

(supported by Clarin)

**Aim:** to ease future CBA texts annotation by finding the parts of texts where a question is answered = assign a 'bind' to paragraphs in new CBA texts

### TEST OF MODELS:

1. **Relative frequency (RF) model** (incl. cleaning with tokenisation and lemmatisation). Trained on the existing clauses dataset.
2. **Universal Sentence Encoder (USE) model**, scored in two ways: using the rank of the model output answers (USE-RANK) and using the Euclidean distance of each paragraph to the question embeddings (USE-ED).

### EVALUATION OF MODELS USING CROSS VALIDATION

**TOOL:** a new CBA in html, word or xml format can be uploaded, and the language has to be selected. For each language and bind the model giving a better CV result is applied. The tool returns the portion of text where the answer of each question can be found.

## SSHOC Partners

