

# Task 2: Abbreviation disambiguation



**Michał Marcińczuk**

(Wrocław University of Technology)

**Łukasz Kobyliński**

(Instytut Podstaw Informatyki PAN / Sages)

# Motivation

## How to find the full, expanded dictionary form of an abbreviation?

- In many cases, this may be done by a simple dictionary lookup, but:
  - the use of abbreviations is often unconventional,
  - there is no complete list of all possible abbreviation uses,
  - many of the abbreviations are ambiguous, translating to possibly different expanded forms.
- The disambiguation of abbreviations needs to include context and additional language knowledge to be feasible.

# Task description

## What about grammatical forms of expansions?

- We need to produce the expanded form of an abbreviation in correct grammatical form, in concordance with the rest of the sentence.

### Thus, the challenge is to:

- propose a method of disambiguating Polish abbreviations. The method should recognize if a given phrase is an abbreviation and, if so, produce its expanded form, both base, and inflected ones.

# Training data

Abbr	Expanded form	Base form	Context
s.	sobota	sobota	Karpaty Siepraw (n. 16). IV liga, grupa wschodnia: Olimpia Wojnicz - Grybovia (16), Orkan Szczyrzyc - Wolania Wola Rzędzińska (s. 16), Sandecja II Nowy Sącz -
d.	dawniej	dawniej	poinformowała w piątek na swej stronie internetowej rosyjska korporacja państwowa Rostech ( Rostechnologii). Nie podano daty przechwycenia amerykańskiego drona. „Dron MQ-5B,”
n.	niedziela	niedziela	11) Gościbia - Piast (s. 16) Wiślanka - Sęp (s. 16); Skawinka - pauza Sęp - Gościbia (16.30) Piast - Hejnał (n. 17) Orzeł - Jordan (n. 17) Czarni - Nadwiślanka (s.
pkt. proc.	punktu procentowego	punkt procentowy	proc. Kolejne 0,12 pkt. proc. wynika ze spadku popytu na polski eksport, a 0,08 z zaburzeń na rynku wewnętrznym” - oszacowali.
rp.pl.	rp.pl.	rp.pl.	Jutro rozpocznie się proces posła ruchu Palikota - dowiedziała się Biedroń został oskarżony o naruszenie nietykalności cielesnej funkcjonariusza policji

# Evaluation

We calculate two measures of accuracy for each provided submission:

- **Af** — the accuracy of provided expanded forms of abbreviations (case insensitive string match)
- **Ab** — the accuracy of provided base forms of abbreviations (case insensitive string match).

Based on these measures, the final score will be calculated using a weighted average:

- **$Acc = 0.25 * Af + 0.75 * Ab$**

# Results

Submission	Weighted Test-A	Weighted Test-B
Krzysztof Wróbel	92.76	92.01
Jakub Karbowski	91.75	91.27
Marek Kozłowski	89.00	88.73
Jakub Pokrywka	65.48	66.25
Rafał Prońko	n/a	19.09

**Thank you!**

