# Hack #13 - Changing the underlying concept recognizer of the AgroPortal Annotator

AgroPortal

Annotator

mgrep
mallet
**alvisnlp**
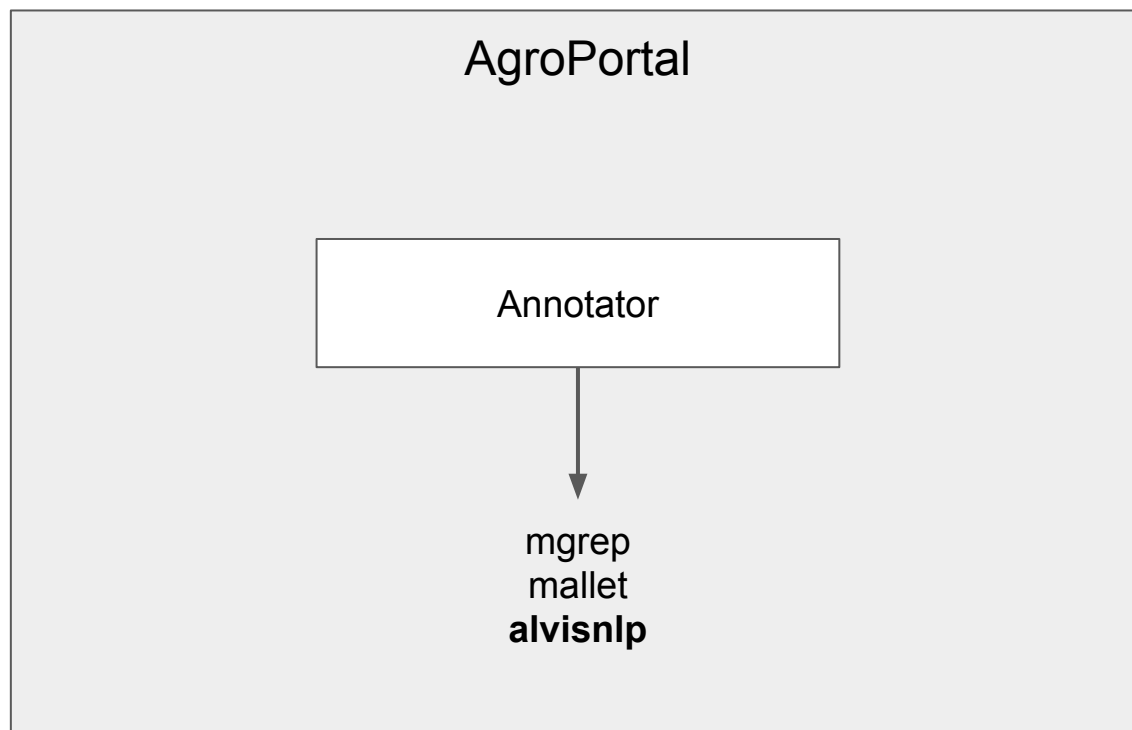
**Robert Bossy**
Amine Abdaoui
Mouhamadou Ba
Pierre Larmande
Nordine El Hassouni
Sara Remini
Antoine Affouard
Vincent Emonet

agroportal / ncbo_annotator / hack13

# **Before** Hack #13

Exact string match, misses a lot of variations:

- typographic (spaces and dashes);
- abbreviations;
- morphological (inflections);
- syntactic ("colon cancer" / "cancer of colon").

# **After** Hack #13

NLP processing, currently does:

- segmentation (words and sentences);
- lemmatization (morphological variations).

After the after, what can be considered:

- chunking/parsing and IR techniques (syntactic variations);
- coreference resolution (abbreviations);
- discourse analysis (salience, document topics);
- information extraction (properties, relations).

# Work done

- Development of a new recognizer for the Annotator
- Installation of AlvisNLP/ML (and dependencies)
- Configuration of AlvisNLP/ML
- Lemmatization of the ontologies dictionary.

children                child

🚫        ——lemmatization——→        ⬆⬇

child                child
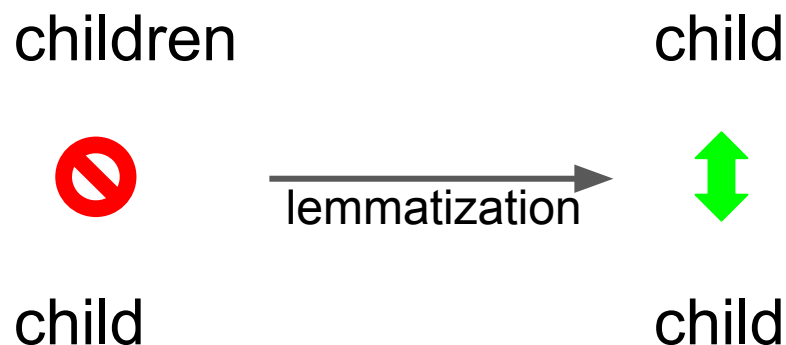
# The near future of Hack #13

- In production
  - scheduled preparation of the dictionary (lemmatization)


- ToMap (recognizer with syntactic variations)
  - really needs the scheduled preparation of the dictionary
  - for better results, requires ontology-specific lexical resources (ambiguous words and terms)