

Named Entity Recognition (NER) von Warndienstmeldungen im Gartenbau:

Eine empirische Studie zu Design, Entwicklung und Bewertung der statistischen und Deep-Learning benutzerdefinierten NER-Modelle

He-Bleinagel, Xia¹; Jung, Jascha Danil²; Golla, Burkhard¹;

¹Julius Kühn-Institut (JKI) Kleinmachow, ²Kuratorium für Technik und Bauwesen in der Landwirtschaft e.V. (KTBL)



Einleitung

Die Anforderungen an die gartenbauliche Beratung und Praxis hinsichtlich des Pflanzenschutzes sind in den Jahren immer komplexer geworden. Insgesamt werden mit hier vorgestellten Projekt die Voraussetzungen geschaffen, die Inhalte der Vielzahl von Warndienstmeldungen automatisiert zu erschließen und über gezielte Abfragen und Suchvorgänge zugänglich zu machen. Der Beitrag stellt das methodische Vorgehen und einige Analyseergebnisse beispielhaft dar.

Anhand gesammelter Pflanzenschutzhinweise und Warnmeldungen wurden Named Entity Recognition (NER) Modelle zur automatischen Erkennung und Klassifizierung von relevanten Begriffen des Gartenbaus (Kulturen, Schaderreger, Pflanzenschutzmittel, BBCH Stadium) erstellt. NER ist eine Teilaufgabe der Informationsextraktion, die darauf abzielt, benannte Entitäten, die in unstrukturiertem Text erwähnt werden, zu finden und in vordefinierte Kategorien einzuordnen.

Daten und Methoden

- Historische Warndienste (1994-2021)
- Methodik: Named Entity Recognition
- Maschinelles Lernen Methoden:
 - SpaCy
 - Flair
 - CRF

Kultur



Erreger

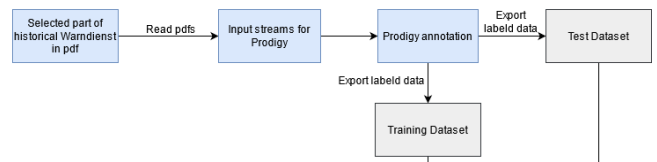


Mittel

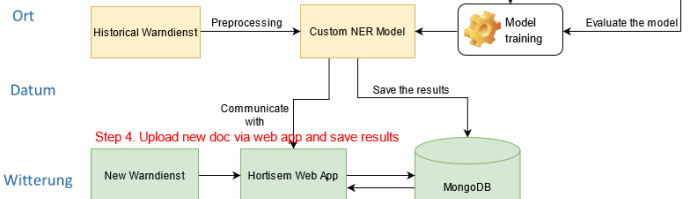


Ort
Datum
Witterung
Kultur
Erreger
Mittel

Step 1. Data collection, annotation with labelling tool Prodigy



Step 3. Feed historical data to the model, extract and save infos.



Step 4. Upload new doc via web app and save results

Step 5. User can search in the database via app

Abbildung Flowchart der Entwicklung eines NER-Modells

Ergebnisse

Auswertung

Bei NER handelt es sich um ein Klassifizierungsproblem.
 - Präzision ist das Verhältnis der korrekt identifizierten positiven Fälle zu allen vorhergesagten positiven Fällen.
 - Recall dagegen, ist das Verhältnis der korrekt identifizierten positiven Fälle zu allen tatsächlichen positiven Fällen.
 - Der F1-Wert bezeichnet das harmonische Mittel von Präzision und Recall und beschreibt so die Gesamtgüte des Modells.

Model	Precision	Recall	F1
Spacy	0.8947	0.9048	0.8997
Flair	0.8524	0.8978	0.8745
CRF	0.905	0.87	0.887

Model	Tag	Precision	Recall	F1
Spacy3	Erreger	0.8697	0.9368	0.902
	Kultur	0.9194	0.9344	0.9268
	Mittel	0.9189	0.8395	0.8774
	BBCH_Stadium	0.7895	0.6522	0.7143
Flair	Erreger	0.8182	0.9158	0.8642
	Kultur	0.8539	0.9344	0.8924
	Mittel	0.931	0.8333	0.8795
CRF	Erreger	0.905	0.8145	0.8488
	Kultur	0.9783	0.7493	0.8418
	Mittel	0.904	0.8405	0.8693
	BBCH_Stadium	0.8345	0.8713	0.8483

Tabelle Zusammenfassung der gemittelten Ergebnisse aller drei Modelle

Tabelle Detailzusammenfassung der Ergebnisse der drei Modelle bei den vier Klassen

Fazit

Mit Hilfe dieses NER Modells werden pflanzenschutzrelevante Information in Zukunft automatisch aus textlichen Quellen gewonnen.

