

Call for Challenges für NLP-Hackathon

Texte lassen sich mittels Methoden des Natural Language Processing (NLP) auf sprachliche und inhaltliche Informationen untersuchen und auswerten. Um neue Blickwinkel auf die unterschiedlichen NLP-Problemstellungen zu erhalten und neue Arbeitsweisen kennenzulernen, veranstaltet die Universität Bern am 24. und 25. März 2021 einen NLP-Hackathon. Eigene Challenges können jetzt eingereicht werden.

Erfolgreiche NLP-Anwendungen zeichnen sich durch den Einsatz von maschinellen Lernverfahren aus, die einerseits auf Sprachmodellen basieren und andererseits über genügend Trainingsmaterial verfügen. Die Herausforderungen sind entsprechend vielseitig und betreffen von der textuellen Aufbereitung über den Einsatz von *deep learning* Algorithmen bis zu Visualisierungsformen unterschiedliche digitale Anwendungen. Im Rahmen des ersten NLP-Hackathon der Universität Bern am **Mittwoch und Donnerstag, 24. und 25. März 2021** sollen Teilnehmende konkrete Challenges rund um NLP lösen und diese anschliessend einander vorstellen. Vorgängig können NLP-Challenges von Uni-internen Stellen oder von Extern eingereicht werden, für die sich die Teilnehmenden anschliessend registrieren.

Zielsetzungen

Bei einem Hackathon finden Personen zusammen, die gemeinsam in kurzer Zeit eine kreative, technische Lösung auf bestimmte Problemstellungen (Challenges) entwickeln wollen. Das übergeordnete Ziel des NLP Hackathon ist der Wissensaufbau, der Erfahrungsaustausch und die verbesserte Vernetzung von Interessierten betreffend NLP. Ausserdem sollen mit der Bearbeitung von konkreten Challenges die Möglichkeiten und Grenzen von heutigen NLP-Technologien getestet und mit neuen Tools experimentiert werden. Zudem sollen die einreichenden Stellen von Lösungsansätzen für ihre Challenges profitieren und bei Interesse die Zusammenarbeit mit den Teilnehmenden fortsetzen.

Berücksichtigung COVID-19

Normalerweise finden Hackathons an einem Veranstaltungsort statt, wo sich alle treffen, austauschen und auch verpflegen können. Da es in der aktuellen Pandemie-Lage jedoch schwierig ist, grössere Anlässe vor Ort zu organisieren, soll der NLP Hackathon im März 2021 dezentral durchgeführt werden. Die teilnehmenden Teams arbeiten im Homeoffice oder in ihrer jeweiligen Arbeitsumgebung an der ausgewählten Challenge. Zu Beginn des Hackathons gibt es eine kurze virtuelle Vorstellung aller Teams und gewählten Challenges und zum Abschluss des Tages präsentieren die Teams ihre Resultate virtuell.

Challenges

Gesucht werden NLP-Challenges beispielsweise zu folgenden Themen:

- Spezifische Erkennformen (Part-of-Speech, Named Entity Recognition etc.)
- Information Extraction und Open Knowledge Graph Canonicalization
- Relationship Extraction
- Fachsprachen
- historische Sprachformen
- etc.

Eine NLP-Challenge kann entweder A) kompetitiven oder B) kreativen Charakter haben:

- A) Bei den **kompetitiven Challenges** müssen die Problemstellung und die zu analysierenden Daten exakt vorgegeben werden, damit die Ergebnisse messbar und untereinander vergleichbar sind. Es geht also darum, dass die teilnehmenden Teams versuchen, eine möglichst gute Lösung für die NLP-Challenge zu entwickeln und damit die Challenge zu gewinnen. Dazu muss eine entsprechende Competition auf Kaggle (www.kaggle.com) erstellt und für den Hackathon eingereicht werden.
Beispiel: «Zur Verfügung steht ein Trainings-Datenset von deutschsprachigen Schweizer Bundesgerichtsentscheiden annotiert mit Named Entities. Erziele eine möglichst hohe F1-Score auf dem Testset bei der Vorhersage von Named Entities.»
- B) Bei den **kreativen Challenges** können NLP-bezogene Problemstellungen eingereicht werden, deren Ergebnisse sich nicht einfach quantifizieren lassen (bspw. offene Fragestellungen, Datenvisualisierungen, kreativer Umgang mit NLP-Ausgaben etc.). Dennoch braucht es eine klare Beschreibung der Aufgabenstellung und der erwarteten Resultate, damit die Arbeiten fokussiert ausgeführt werden können.
Beispiel: «Zur Verfügung steht ein Knowledge Graph aufgebaut auf einem Schweizer Gerichtsurteil. Reduziere diesen auf die wesentlichsten Fakten und stelle deine Erkenntnisse ansprechend dar.»
Alternativ: «Visualisiere ein Sprachmodell (*language model*) mit dem Ziel einzelne Tokens zu finden oder Wortcluster zu identifizieren und interpretierbar zu machen.»

Für beide Arten von NLP-Challenges können **bis Montag, 22. Februar 2021** per Email die jeweiligen Kaggle-Links bzw. die Beschreibung der kreativen Challenges an dh@wbkolleg.unibe.ch eingereicht werden. Bei Fragen zu den Challenges, dem Ablauf, dem Aufwand etc. können die untenstehenden Personen kontaktiert werden.

Ablauf

- Mitte Januar 2021: Veröffentlichung des Call for NLP-Challenges
- 22. Februar 2021: Eingabeschluss für kompetitive und kreative Challenges
- 1. März 2021: Veröffentlichung der eingereichten Challenges
- 15. März 2021: Anmeldeschluss für teilnehmende Teams
- 24. und 25. März 2021: Durchführung des NLP-Hackathons

Veranstaltende

Der Anlass ist eine Fortsetzung des [NLP-Workshop im August 2020](#) und wird durch die Digital Humanities der Phil.-hist. Fakultät, das Phil.-nat. Kompetenznetzwerk Digitalisierung (CND), die Forschungsstelle Digitale Nachhaltigkeit des Instituts für Informatik und die Science IT Support Unit der Universität Bern organisiert. Der NLP-Hackathon ist Teil der [vDHD2021 virtual Digital Humanities Community im deutschsprachiger Raum](#).

Hinweis für Doktorierende

Doktorierende der Graduate School of the Arts and Humanities (GSAH) der Universität Bern erhalten 1 ECTS für die aktive Teilnahme an diesem NLP-Hackathon

Kontaktpersonen

[Prof. Dr. Tobias Hodel](mailto:tobias.hodel@wbkolleg.unibe.ch)
Digital Humanities, Phil.-hist. Fakultät
tobias.hodel@wbkolleg.unibe.ch

[PD Dr. Matthias Stürmer](mailto:matthias.stuermer@inf.unibe.ch)
Leiter Forschungsstelle Digitale Nachhaltigkeit
Institut für Informatik, Phil.-nat. Fakultät
matthias.stuermer@inf.unibe.ch