

Semantische Anreicherung durch Definitiondownload

Automatisches Anreichern von Textdokumenten durch Definitionen aus externen Quellen

Research Studios Austria
DSc. Dialog Based Search

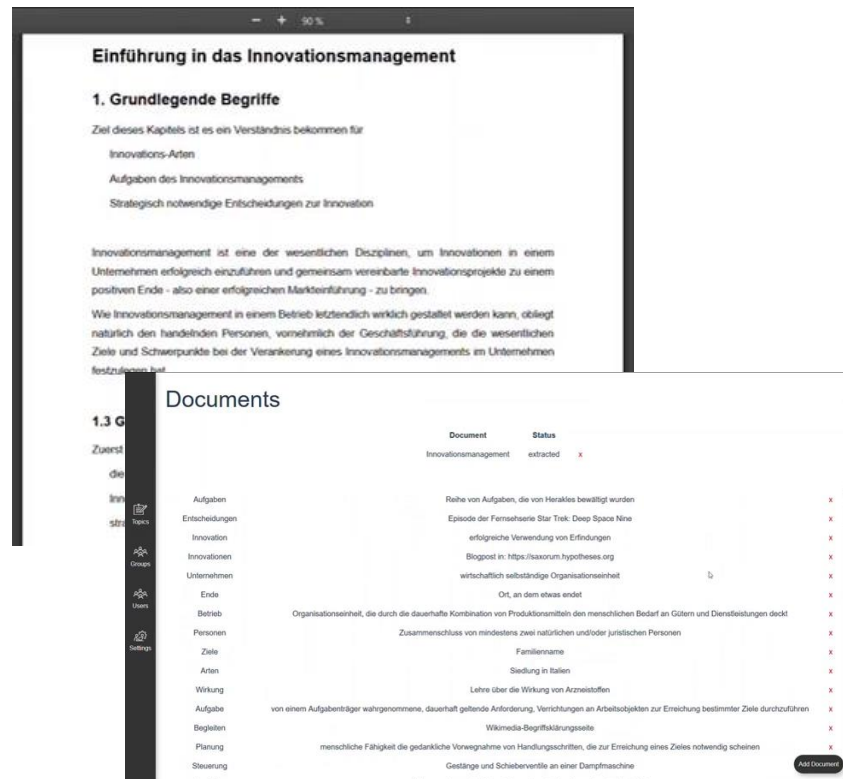


Abb. 1: Automatisierte Extraktion von Begriffsdefinitionen aus einem PDF-Dokument

Hintergrund

Acadybot extrahiert Dokumente und bereitet sie für einen Lern-dialog-Chatbot auf. Das Studio DSc untersuchte für den Kunden unterschiedliche Möglichkeiten der Semantischen Anreicherung dieser Dokumente.

Da die Untersuchung des Chatverhaltens von Nutzern ergab, dass diese ein besonderes Interesse an normativen Definitionen bestimmter Begriffe zeigen, wurde ein Prototyp erstellt, der das Dokument analysiert und zu relevanten Begriffen Definitionen aus externen Quellen lädt.

Ziel

Ziel war es, für den Kunden eine technologische Machbarkeits- abschätzung zur semi-automatisierten Anreicherung von Doku- menten mit Definitionsinformationen durchzuführen und einen Prototyp zur Demonstration der Verarbeitungspipeline zu entwik- keln.

Als externe Quellen wurden insbesondere semantische Graph- Quellen wie etwa WordNet, DBPedia oder WikiData untersucht.

Methode

Zur Dokument-Vorverarbeitung wurde die spaCy Bibliothek ver- wendet um durch verschiedene Verarbeitungsschritte (Tokeni- zer, Lemmatizer, POS-Tagger) aus den Volltexten einzelne Begriffe zu extrahieren.

Der entwickelte Prototyp extrahiert eine vollständige Liste der Substantive aus dem Text, normalisiert diese (stemming) und eliminiert Duplikate. Abschließend werden textuelle Beschrei- bungen über die WikiData Media API abgerufen und in der Da- tenbank gespeichert.

Der Prototyp bietet dazu eine Schnittstelle die extrahierten De- finitionen manuell zu verwalten.

Innovation

- Begriffsextraktion durch Dokumentanalyse
- Automatischer Download von Definitionen
- Benutzerschnittstelle zur Definitionsverwaltung

Nutzen

- Einfachere Erstellung von Lernmaterial
- Aufwertung von bestehenden Lernunterla- gen (Skripten, Präsentationen, etc.)
- Bessere Antworten des Chatbots für Dialog- basiertes Lernen

Demonstration

- Demo-Video für Acadybot GmbH
- iiWAS 2019

Publikation

- Göschlberger B., Brandstetter C. (2019) Conversational AI for Corporate e-Learning, 21st International Conference on Infor- mation Integration and Web-based Applica- tions & Services