

Generating Company Profiles Using Natural Language Processing

Thema:

Generating Company Profiles Using Natural Language Processing

Art:

MA

BetreuerIn:

Christian Wolff, Julia Wemken (Bosch), Tobias Kallenbach (Bosch)

BearbeiterIn:

Christoph Tögel

ErstgutachterIn:

Christian Wolff

ZweitgutachterIn:

Udo Kruschwitz

Status:

abgeschlossen

Stichworte:

NLP, Automatisierung, Datenanalyse, natural language processing, named entity recognition, deep learning, BERT

angelegt:

2020-10-19

Antrittsvortrag:

2020-11-09

Abschlussvortrag:

2021-04-26

Hintergrund

Damit Unternehmen ihre Position am Markt behaupten und auf Trends reagieren können, müssen aktuelle Entwicklungen und besonders die Aktivitäten von Wettbewerbern überwacht werden. Aufgrund der Vielzahl von Firmen und Größe des Marktes, ist dies ein tiefgehender und meist individueller Prozess für jede zu analysierende Firma. Mit modernen Textanalysetechniken (NLP) kann dieser Prozess jedoch automatisiert werden, indem Informationen aus frei verfügbaren Textquellen extrahiert werden. Im Forschungsprojekt, in dessen Rahmen diese Masterarbeit erstellt wird, sollen die Möglichkeiten eines solchen Systems erfasst werden. Dazu wurden verschiedene Anwendungsfälle festgelegt, von denen in dieser Arbeit die High-Cost/Low-Cost-Standortanalyse behandelt wird. Ziel dieser ist die Erfassung der Wirtschaftlichkeit eines Wettbewerbers auf globaler Ebene.

Zielsetzung der Arbeit

Im Laufe dieser Arbeit sollen mit Deep-Learning-Verfahren, wie BERT, die automatische Standortanalyse erarbeitet werden. Dabei sollen einzelne Standorte nach Land, Stadt und wenn möglich Adresse aufgeschlüsselt werden können. Als Grundlage werden Nachrichten und Websites genutzt. Ziel ist eine mit händischen Analysen gleichwertige Erfassung von Wettbewerberstandorten.

Konkrete Aufgaben

Im Laufe des Projekts soll eine Pipeline zur automatischen Erkennung von Standorten erstellt werden. Dazu müssen in mehreren Schritten unstrukturierte Textdaten erfasst und angebunden werden. Über unterschiedliche NLP-Techniken werden dann Texte klassifiziert, Firmen und Orte erkannt, sowie letztendlich in eine Datenbank mit strukturierten Daten überführt. Hierbei soll auch die automatische Verbesserung der „Deep Learning“-Verfahren mittels strukturierter Daten evaluiert werden. Im letzten Teil werden die erfassten Ergebnisse noch anderen Datenquellen zu Standortdaten verglichen, um die Performanz des Systems zu testen.

Erwartete Vorkenntnisse

Python, NLP, Datenvisualisierungstechniken, SQL-& NoSQL-Datenbanken

Weiterführende Quellen

Zuo, Z., Loster, M., Krestel, R., & Naumann, F. (2017). Uncovering Business Relationships: Context-sensitive Relationship Extraction for Difficult Relationship Types. In LWDA (p. 271).

Wright, J. M., & Mansingh, G. (2014). An Algorithm to Extract Jamaican Geographic Locations from News Articles-Using NLP Techniques. In CONF-IRM (p. 24).

Malmasi, S., & Dras, M. (2015, May). Location mention detection in tweets and microblogs. In Conference of the Pacific Association for Computational Linguistics (pp. 123-134). Springer, Singapore.

Liang, C., Yu, Y., Jiang, H., Er, S., Wang, R., Zhao, T., & Zhang, C. (2020, August). Bond: Bert-assisted open-domain named entity recognition with distant supervision. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 1054-1064).

From:
<https://wiki.mi.ur.de/> - MI Wiki

Permanent link:
https://wiki.mi.ur.de/arbeiten/automatisierung_der_erstellung_von_firmenprofilen_mithilfe_von_nlp_am_fallbeispiel_der_high_cost_low_cost-standortanalyse

Last update: 14.07.2021 16:08

