

Hotel Benchmarking in der Bodenseeregion: Eine Text-Mining Analyse am Beispiel der Online Reiseagentur Booking.com

Dominic Regitz, Jona Uekert,
 Forschungsgruppe Data Science, Institut für Digitalen Wandel (IDW)

Einführung

- Online-Tourismusportale bieten in Form von Kundenbewertungen umfangreiche Informationen für Hoteliers
- Durch Text-Mining-Verfahren können aus Kundenbewertungen Stärken und Schwächen einzelner Hotels, in unterschiedlichen Bereichen, aufgedeckt werden

Zielsetzung

- Durchführung eines Hotel-Benchmarkings für ausgewählte Hotels in der Bodenseeregion mit konkreter Auflistung der Verteilung von positiven und negativen Bewertungen
- Vergleich der Hotels in den Bereichen Location & Property, Food & Beverages, Service & Staff sowie Room & Bathroom

Zusammenfassung

- Analyse von Online-Reviews mittels Verfahren des Text-Mining ermöglichen einen vereinfachten Vergleich einzelner Hotels
- Die Umsetzung der Verfahren in Deutsch ist anspruchsvoller als in Englisch, da Nuancen der deutschen Sprache die automatische Textanalyse erschweren

Methodik

Datenbeschaffung



RAPIDMINER

- Einmalige Datenextraktion von ca. 20 000 Kundenbewertungen von 16 Hotels in der Bodenseeregion mit Octoparse
- Weiterverarbeitung durch Rapidminer

Datenbereinigung

- Entfernen von Duplikaten
- Entfernen von Kundenbewertungen ohne textlichen Inhalt

Datenverarbeitung

- Identifizierung von vier Hotels mit den meisten Bewertungen
- Bewertungsinhalte bereits in positive und negative Sentiments kategorisiert
- Satzweise Aufspaltung der Texte mittels Text-Mining-Verfahren
- Lexikonbasierte Topic Detection für die Bereiche
 1. Location & Property
 2. Food & Beverages
 3. Service & Staff
 4. Room & Bathroom
- Vergleich der Hotels basierend auf identifizierten positiven und negativen Sätzen für die einzelnen Topics

Ergebnisse

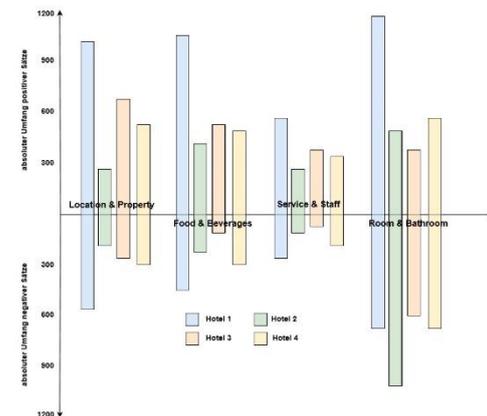


Abb. 1: Absolute Anteile positiver und negativer Sätze zu den Topics der Hotels

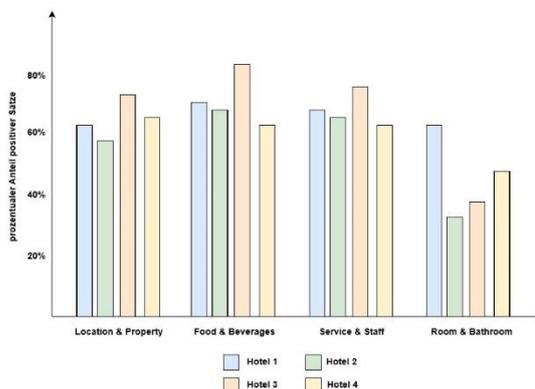


Abb. 2: Prozentuale Anteile positiver Beiträge zu den Topics für die Hotels

Allgemeine Erkenntnisse

- Überwiegend positive Berichte zu Location & Property, Food & Beverages und Service & Staff
- Room & Bathroom weist den größten Anteil an identifizierten negativen Sätzen auf
- Allgemein geringe Erwähnung des Topics Service & Staff

Hotel 1

Erzielt die besten Ergebnisse in der Kategorie Room & Bathroom.

Hotel 2

Weist vergleichsweise wenig negative Sätze in der Kategorie Room & Bathroom auf.

Hotel 3

Führt in den Bereichen Location & Property, Food & Beverages und Service & Staff.

Hotel 4

Liegt insgesamt im Mittelfeld.

Kontakt

Prof. Dr. Wolfram Höpken
 Leiter IDW / Forschungsgruppe Data Science
 +49 751 501 9764
 wolfram.hoepken@rwu.de

Literatur

- Menner, T., Höpken, W., Fuchs, M. and Lexhagen, M. 2016. "Topic detection – Identifying relevant topics in tourism reviews", in Inversini, A. and Schegg, R. (Ed.s.), Information and Communication Technologies in Tourism 2016, Springer, New York: 411-423
- Schmunk, S., Höpken, W., Fuchs, M. and Lexhagen, M. 2014. "Sentiment Analysis – Implementation and Evaluation of Methods for Sentiment Analysis with Rapid-Miner®", in Xiang, Ph. and Tussyadiah, I.(Ed.s.), Information and Communication Technologies in Tourism 2014, Springer, New York: 253-265