

## 14-132-R "Zur Erstellung von Meinungsprofilen auf der Basis von Daten aus sozialen Medien: Eine Fallstudie über Restaurantbewertungen"

Prof. Hatem Ghorbel, Haute Ecole Arc, Martin Hilpert, Universität Neuenburg

CHF 48'500.-

Das Ziel dieses Projekts war es, aus online verfügbaren Restaurantbewertungen Meinungsprofile zu erstellen, welche die Qualität von Restaurants im Hinblick auf mehrere Kriterien widerspiegeln. Die Analyse stützt sich auf die linguistische Theorie der Rahmensemantik und die automatische Sentiment-Analyse. Basierend auf dem "Yelp Academic Dataset" haben wir eine semantische Klassifizierung von Restaurantbewertungen vorgenommen, welche die Qualität eines Restaurants nicht auf eine einfache lineare Skala herunterbricht, sondern die Stärken und Schwächen nuanciert abbildet. Es wurde eine web-basierte Anwendung erstellt, mit welcher der Nutzer Restaurants mit spezifischen Profilen finden und sich so informieren kann.

Was erwartet man von einem Restaurantbesuch? Was sind die Kriterien, nach denen ein solcher Besuch positiv oder negativ bewertet wird? Internetportale wie Yelp komprimieren dieses Wissen typischerweise auf eine Anzahl von Punkten oder Sternen und als weiteres Mass die preisliche Kategorie des Restaurants. Diese Informationen sind sehr nützlich, aber sie sind nur die Spitze eines Eisbergs von viel umfangreichem Wissen, das in die Bewertung von Restaurants eingeht. Es gibt verschiedene Arten von Restaurants, die auf verschiedene Zielgruppen abzielen und unterschiedliche Erwartungen bedienen. Je nach Art des Restaurants, das beurteilt wird, werden unterschiedliche Kriterien bewertet. Der erste Schritt der Analyse war somit die Identifizierung wiederkehrender Konzepte, die in Restaurantbegutachtungen vorkommen. Diese Konzepte beinhalten die Qualität des Angebots, die Atmosphäre des Restaurants, das Preisniveau, die sanitären Einrichtungen, die Freundlichkeit des Personals etc.. Insgesamt wurden 22 Kriterien berücksichtigt.

Eine offensichtliche Schwierigkeit bei der Analyse der Bewertungen ist es, automatisch zu erkennen, über welche Kategorie ein Rezensent in einem bestimmten Satz spricht. Um diesen Schritt zu ermöglichen, wurde für jede Kategorie eine Wortliste erstellt, so dass z.B. Wörter wie "Auto" oder "Garage" als Indikatoren für die Kategorie "Parkplätze" ausgewählt wurden.

Die Wortlisten wurden auf der Basis verschiedener Arten von Schlüsselwörtern zusammengesetzt, so dass jeder Kategorie eine Liste von etwa 30 Wörtern zugeordnet war. Anhand dieser Wörter ordnet ein eigens erstellter Algorithmus jeden Satz in der Datenbank einer semantischen Kategorie zu und ermittelt die Polarität der Bewertung. Dies ist in Abb. 1 schematisch dargestellt.

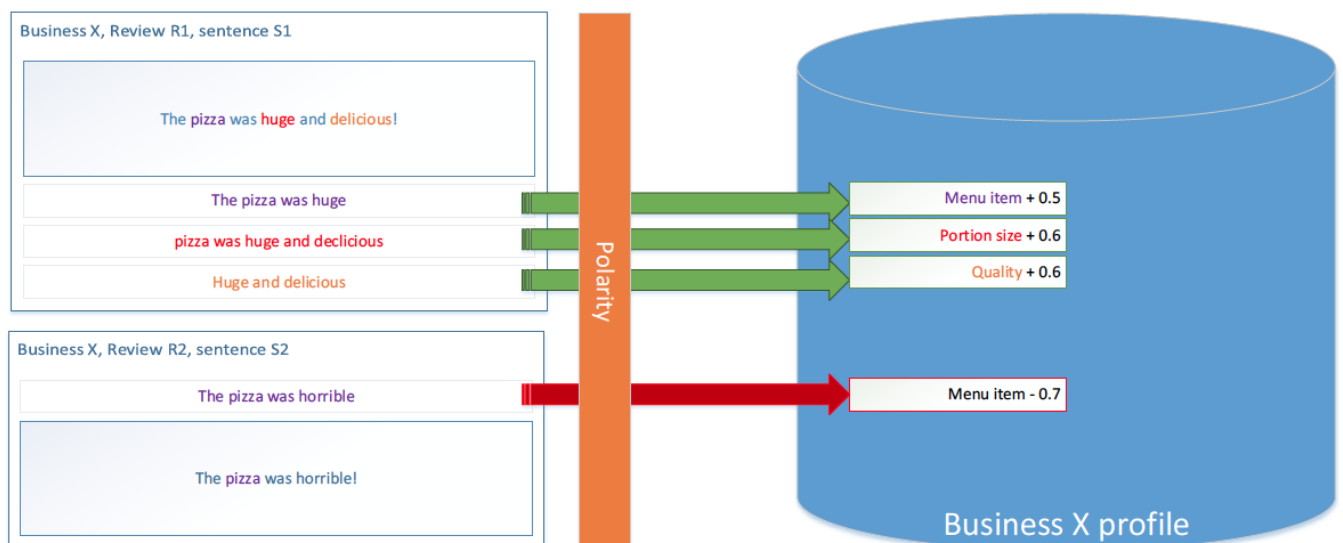


Abb.1

# the **cogito** foundation

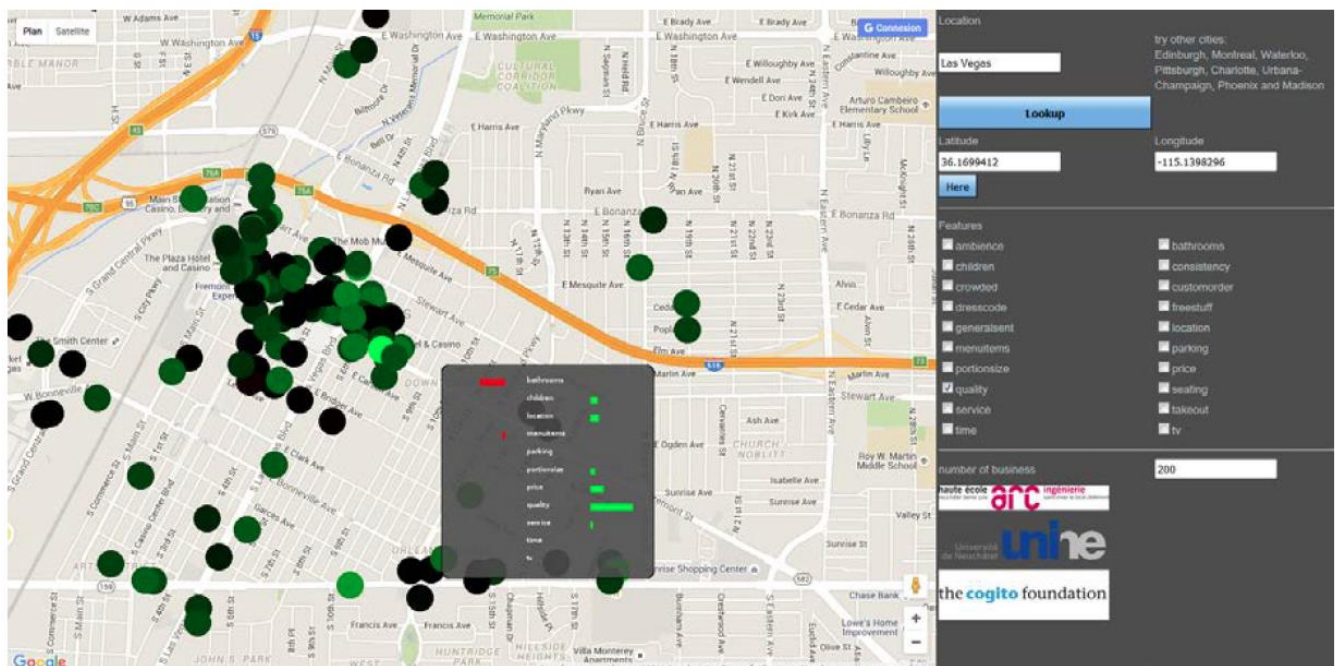


Abb. 2 zeigt, auf welche Weise die Information in der Datenbank für den Nutzer verfügbar ist. In einem Menü können verschiedene Kriterien ausgewählt werden. Der Nutzer wählt die für seine Suche relevanten Kriterien, woraufhin auf dem Bildschirm Restaurants angezeigt werden, die für diese Kriterien positive Bewertungen erhalten haben.

Um die Qualität des Algorithmus zu testen, wurden die automatisch erstellten Ergebnisse mit manuell erstellten Daten verglichen. Die Ergebnisse zeigen unter anderem, dass positive Bewertungen einfacher automatisch zu identifizieren sind als negative Bewertungen.

Das Projekt hat Informatiker und Geisteswissenschaftler in Neuenburg enger zusammen gebracht. Die Ergebnisse sind sowohl für eine breite Öffentlichkeit als auch für den weiterführenden Austausch von Geistes- und Naturwissenschaften relevant.