

ORIENTIERUNG IN DER DATENFLUT

Usern unter die Arme greifen: Die Informatikerin Eva Zangerle arbeitet an der Verbesserung von Empfehlungssystemen am Beispiel des Social-Media-Portals Twitter.



EVA ZANGERLE: Das Team der Informatikerin sammelte seit 2011 vier Milliarden Tweets.

Sie sind zwar nur 140 Zeichen lang, enthalten aber dennoch zahlreiche Informationen: Kurznachrichten, so genannte „Tweets“, sind das Herzstück des Microblogging-Dienstes Twitter, der aktuell von mehr als 300 Millionen Menschen weltweit aktiv genutzt wird. Für Eva Zangerle von der Arbeitsgruppe für Datenbanken und Informationssysteme am Institut für Informatik steht das Social-Media-Portal schon seit mehreren Jahren im Mittelpunkt ihrer Forschung. „Twitter ist eine diverse Plattform, die sich durch eine heterogene Usergruppe auszeichnet und mit dementsprechend vielfältigem Datenmaterial aufwarten kann“, erklärt Zangerle. Die Herausforderung liegt nun darin, diese riesigen Datenmengen möglichst sinnvoll zu filtern und in Hinblick auf verschiedenste Aspekte durchsuchbar zu machen.

Dass die Wahl in der mittlerweile großen Vielfalt von sozialen Netzwerken auf Twitter fiel, hat aber auch einen pragmatischen Grund, wie Zangerle deutlich macht: „Twitter stellt etwa ein Prozent seiner Tweets über eine Schnittstelle gratis zur Verfügung. Je nach Forschungsfrage holen wir uns die für uns relevanten Daten und speichern das Material hier bei uns ab.“ Dieser Vorgang, der als „crawlen“ bezeichnet wird, setzt große Speicherkapazitäten und den Einsatz von Hochleistungsrechnern voraus: Seit 2011 sammelte das Team am Institut für Informatik vier Milliarden Tweets, die dank der enormen Rechenkraft am Supercomputer LEO III der Uni Innsbruck analysiert werden können.

Empfehlungen

Zentrales Element in der Kommunikation auf Twitter sind sogenannte Hashtags. Diese mit einem Rautezeichen (#) versehenen Schlagworte werden in einen Tweet eingebaut und ordnen den Inhalt einem gewissen Themenbereich zu. „Hashtags zählen zu den wenigen Möglichkeiten, um auf Twitter Information

eine Struktur zu geben, sie durchsuchbar zu machen“, sagt Eva Zangerle. Dass es in diesem Bereich Aufholbedarf gibt, war für die Informatikerin schon früh klar, weshalb sie sich bereits in ihrer Doktorarbeit Empfehlungssystemen widmete. „Egal ob bei Konferenzen oder politischen Ereignissen: Wie lautet der richtige Hashtag? Diese Frage war für viele Nutzerinnen und Nutzer oft schwierig oder kaum zu beantworten und machte viele Informationen später unauffindbar.“

Aus diesem unbefriedigenden Zustand entwickelte sich das Bedürfnis nach mehr Unterstützung. Auch Twitter selbst reagiert darauf und arbeitet laufend an Verbesserungen: Bereits beim Tippen eines Tweets werden die hoffentlich „richtigen“ Hashtags

„Gerade dort, wo wir riesige Ansammlung von Wissen haben, möchte ich einen Beitrag dazu leisten, es besser greifbar zu machen.“

Eva Zangerle, Institut für Informatik

vorgeschlagen. Der Nutzer oder die Nutzerin erhält unmittelbar passende Vorschläge. Was im ersten Moment relativ banal klingen mag, macht umfassende Berechnungen und Analysen erforderlich. Die Arbeit an diesen Empfehlungsdiensten, die in der Fachsprache Recommender-Systeme genannt werden, hat in der Informatik in den letzten Jahren zusehends an Bedeutung gewonnen. Was in Onlineshops mit Sätzen wie „Das könnte Ihnen auch gefallen“ oder personalisierten Werbeeinschaltungen schon längst zum Alltag gehört, findet auch in Social-Media-Anwendung – wenn auch weniger offensichtlich. Recommender-Systeme sind ständige „Begleiter“ für alle, die sich im Internet bewegen. An ihrer Verbesserung arbeiten nicht nur Internet-Giganten wie Amazon, auch die Wissenschaft hat großes Interesse an ihrer Funktionsweise.

Hörtipps

Automatisierte Empfehlungstechnologien verfolgen das Ziel, den oder die Nutzerin möglichst gut kennenzulernen, um Prognosen darüber abzugeben, was in verschiedenen Situationen von Interesse sein könnte. Die analytische Basis dafür und gleichzeitig die größte Herausforderung ist der Vergleich.

Eva Zangerle ist auch zu Gast in unserem Podcast „Zeit für Wissenschaft“ unter www.uibk.ac.at/podcast/zeit/2016/zfw025.html



Welche Tweets sind sich ähnlich? Wo gibt es Überschneidungen in der Verwendung von Hashtags und in welchen Situationen wurden sie benutzt? Das System ermittelt Ähnlichkeiten zwischen Elementen und Personen und lernt nach und nach, passende Empfehlungen abzugeben. „Die Berechnungen erfolgen bei uns an 30 bis 40 Computern gleichzeitig“, erzählt Zangerle. „Angesichts der Datenmengen ist das nicht anders möglich, wenn wir einen realistischen Zeitplan für unsere Ergebnisse einhalten wollen“.

In einem Projekt hat sich die Informatikerin in den letzten Jahren gemeinsam mit ihren Kollegen der Etablierung von Musikempfehlungen auf Twitter gewidmet. „Sehr viele Menschen informieren ihre Follower immer wieder darüber, welche Musik sie gerade hören. Wir haben uns dazu auf die Hashtags ‚#nowplaying‘ und ‚#listeningto‘ konzentriert und Datenmaterial mit diesen Stichworten gecrawlt“, beschreibt Zangerle den Projektansatz. „Mit unseren Berechnungen verfolgten wir das Ziel, mehr über den Musikgeschmack und das Hörverhalten der Twitter-User zu erfahren, um dann in einem weiteren Schritt ein Empfehlungssystem ableiten zu können.“ Dieses Datenset aus mehreren Millionen Tweets bereinigten die Wissenschaftler zunächst. „Die Daten, die wir über die Schnittstelle erhalten, beinhalten bei Weitem nicht nur den Inhalt des Tweets, sondern bis hin zur Profil-Hintergrundfarbe auch zahlreiche, für uns zu einem großen Teil irrelevante Metadaten“, so Zangerle. Erst nach diesem Filtervorgang kann die vergleichende Analyse beginnen. Die Informatikerin entwickelte auf Basis der so ermittelten Informationen ein Datenset, das den Grundstein für individuelle Musikempfehlungen legt und gleichermaßen in anderen Themenbereichen Anwendung finden könnte.

Von Musikempfehlungen bis zu Hashtag-Vorschlägen – für Zangerle steht die Perspektive des Nutzers im Vordergrund: „Gerade im Social-Media-Bereich finde ich die Frage sehr spannend, wie Menschen diese Plattformen nutzen. Und natürlich auch: Wie kann man sie dabei unterstützen?“ Filtern, Datenselektieren und Navigieren wurde in den letzten Jahren ein immer größeres Thema in der Wissenschaft, die Informatikerin sieht auf dem Gebiet der Recommender-Algorithmen auch künftig viele Verbesserungsmöglichkeiten. „Gerade dort, wo wir riesige Ansammlung von Wissen haben, möchte ich einen Beitrag dazu leisten, es besser greifbar zu machen.“

mb

TWITTER & WIKIPEDIA:

Wikipedia gilt als eines der größten Wissensportale im Internet und wurde bereits aus der Perspektive zahlreicher wissenschaftlicher Disziplinen untersucht. Meist allerdings aus einer internen Sichtweise. Hier setzt ein aktuelles Projekt der Informatikern Eva Zangerle an, das die Nutzung von Wikipedia aus der Perspektive von Twitter-Nutzerinnen und -Nutzern betrachtet: „Durch diese Außenansicht erhoffen wir uns Qualitätsmerkmale für Wikipedia-Einträge ableiten zu können.“ Die Analyse von mehr als sechs Millionen Tweets aus dem Zeitraum zwischen Oktober 2014 und April 2015 mit dem Stichwort „wikipedia“ ergab, dass etwa 20 Prozent aller Hinweise zu Wikipedia-Einträgen auf anderssprachige Einträge verweisen, sich also nicht mit der Sprache des ursprünglichen Tweets decken. „Wenn ich zum Beispiel als deutschsprachige Person auf einen englischen Eintrag der Wikipedia verlinke, dann mache ich das vermutlich, weil mir das deutsche Pendant nicht ausführlich oder gut genug erschien – oder weil es den Eintrag in meiner Muttersprache gar nicht gibt. Für die Bewertung der Qualität von Wikipedia-Einträgen können das sehr wertvolle Informationen sein.“