



Die Wahrheitssucher

Bozen – War früher noch der fehlende Zugang zu Informationen das größte Problem, ist es heute der Überfluss. Zu allem und jedem liefert das Internet Informationen zuhauf. Für den Einzelnen ist es unmöglich geworden, alle Dokumente zu einem Thema zu lesen, zu analysieren, abzugleichen, Relevantes herauszufiltern und Unwichtiges oder Unsinniges zu eliminieren. Selektion ist demnach zur wichtigsten Aufgabe des Internetnutzers geworden, und trotzdem bleibt immer die Frage: Habe ich auch wirklich alle wichtigen Informationen gefunden?

Um die Aufgabe des Selektionierens und Extrahierens von Informationen im Internet zu erleichtern oder überhaupt erst ein Scannen aller Quellen zu ermöglichen, setzt man heute mehr und mehr auf die Technik. Sogenannte Extraktoren durchforsten selbstständig die Fülle vorhandener Texte zu einem Thema und fassen die darin erfassten Informationen zusammen. „Damit wir aber wissen, ob die Extraktoren wirklich alle Informationen erfasst haben, die wichtig und sinnvoll sind, braucht es Prüfverfahren“, erklärt Werner Nutt, Professor an der Informatik-Fakultät der unibz.

Maschinen als Qualitätsprüfer

Was es also braucht, ist eine Qualitätskontrolle, damit sichergestellt ist, dass die gefundenen Informationen auch halten, was sie versprechen. An der unibz werden daher einerseits Extraktoren entwickelt, etwa solche, die relevante Infos aus Hotelbewertungen im Internet herausfischen und zugänglich machen. So spart man sich das Scrollen durch alle Nutzer-Kommentare und kann diese auf relevante Informationen hin durchsuchen: Ist das Frühstück gut? Ist es im Preis inbegriffen? Stört der Straßenlärm?

Andererseits arbeitet man an der unibz auch an Verfahren zur Prüfung der Informationen auf Vollständigkeit und Wichtigkeit. Da gilt es zuallererst, Grundlagenarbeit zu leisten: „Das Problem der Vollständigkeit und Wichtigkeit der Daten in Wissensbanken taucht zwar immer wieder auf, trotzdem fehlen bis dato grundsätzliche Untersuchungen dazu“, erklärt Nutt. Einen ersten Schritt zur Schließung dieser Lücke hat Professor Nutt mit seinem ehemaligen unibz-Kollegen Simon Razniewski und Vevake Baralaman von der Uni Trient gesetzt, indem sie viel genutzt

WISSENSCHAFT – Das Internet hält Informationen in Überfülle bereit. Nur: In Zeiten von Fake News & Co. wird es immer schwieriger, die richtigen und wichtigen Fakten zu finden. Abhilfe können sogenannte Extraktoren schaffen. Die unibz hat sie auf dem Schirm.

te Wissensbanken – etwa Wikidata – und die darin enthaltenen Informationen unter die Lupe genommen haben. „Wissensbanken haben das Ziel, das gesamte Wissen der Welt zu erfassen“, sagt Nutt, „weil dies aber unmöglich ist, werden sie immer unvollständig sein“. Das Problem sei demnach weniger, für Vollständigkeit zu sorgen, als vielmehr jenes, die wichtigsten Informationen bereitzustellen.

Was ist überhaupt wichtig?

Die Frage ist dann allerdings auch: Was ist wichtig? Welche Informationen müssen etwa zu einem Menschen in einer Wissensbank gesammelt und bereitgestellt werden, damit die Daten als „vollständig“ gelten, also alles Wichtige greifbar ist? Und ist jede Information über jeden Menschen von

gleicher Bedeutung? In ihrem Paper „Doctoral Advisor or Medical Condition: Towards Entity-specific Rankings of Knowledge Base Properties“ führen die drei Autoren ein Beispiel dafür an, dass dies nicht der Fall ist. Während nämlich für einen Politiker die Information über die Parteizugehörigkeit wichtig sei, sei sie für einen Musiker marginal. Über diesen sei es wichtiger zu wissen, welche Instrumente er spiele. Und ist die Zahl der geschossenen Tore in der Bewertung von Lionel Messi oder Cristiano Ronaldo zentral, ist sie im Fall von Papst Franziskus eher nebensächlich – auch wenn er in der Jugend Fußball gespielt hat.

Welche Information wichtig ist oder nicht, variiert demnach von Person und Person. „Was es braucht, um Wichtiges von Unwichtigem zu unterscheiden, ist Hausverstand“, sagt Professor

Nutt. Und genau darin liegt das Problem: Maschinen fehlt dieser Hausverstand, er muss durch andere Verfahren ersetzt werden. Derzeit sind unterschiedliche in Gebrauch, rein statistische etwa oder solche, die die Anzahl von Google-Treffern als Messwert für wichtig oder unwichtig heranziehen. Alle diese Verfahren lieferten in den Tests von Razniewski, Baralaman und Nutt aber nur ungenügende Ergebnisse. Während menschliche Testpersonen bei der Unterscheidung von wichtig und unwichtig Übereinstimmungen von 80 bis 90 Prozent erzielten, kamen die getesteten bekannten Verfahren auf Prozentsätze zwischen 50 und 65 Prozent. „Unsere Frage war deshalb: Kann man das besser machen?“, so Nutt.

Maschinen Hausverstand beibringen

Die Zielsetzung für Nutt und seine Kollegen war deshalb klar: „Wir wollten versuchen, mehr Intelligenz in dieses System zu bringen“, so der unibz-Professor, „und zwar mit maschinellem Lernen“. Wenn er also schon keinen Hausverstand hat, will man dem Computer etwas Ähnliches, möglichst Gleichwertiges beibringen, und dafür gibt es zwei Möglichkeiten. Die erste ist das überwachte Lernen, bei dem dem Computer möglichst viele Beispiele gegeben werden, wie Menschen in bestimmten Fällen Entscheidungen treffen. „Daraus erstellt der Computer dann Regeln, denen er selbst bei der Entscheidungsfindung folgt“, so Nutt. Im vorliegenden Fall kam dies aufgrund der Komplexität der Informationen nicht infrage.

Zweite Möglichkeit des maschinellen Lernens ist das „transfer learning“. „Man zieht die Erkenntnisse aus einem verwandten Bereich heran und versucht, diese auf den Untersuchungsbereich anzuwenden“, erklärt der Professor, der diesen Weg mit seinen Kollegen eingeschlagen hat. So wurden Wikipedia-Einträge semantisch analysiert, um zu eruierten, welche Begriffe häufig verwendet werden, um ein bestimmtes Thema zu beschreiben. Kommen demnach in einem Artikel über eine Person Begriffe wie „Wahlen“, „gewählt“ oder „Partei“ vor, hat diese Person höchstwahrscheinlich mit Politik zu tun. „Danach gleicht der Computer den Wissensbank-Eintrag mit den für den Bereich Politik als wichtig definierten Informationen ab und kann eruierten, inwieweit alle relevanten Informationen vorhanden sind“,

so Nutt. Mit dem semantischen Ansatz konnte die Zuverlässigkeit zwar gesteigert werden, aber noch immer waren die Ergebnisse nicht die, die man sich erwartet hatte.

Der Mix macht's

Die besten Ergebnisse lieferten nicht einzelne Verfahren allein, sondern eine Kombination aus unterschiedlichen bekannten und neuen. So konnte

STOP KAUFLE ALLE AUTOS
... zu Höchstpreisen,
zahle in bar, inkl. Abmeldung!

PKW's und LKW's, Busse, Wohnmobile, usw., in gepflegtem Zustand. Auch Unfallautos mit Motorschaden, vielen Kilometern oder reparaturbedürftig.

 Tel. 347 37 92 000
kfzankauf@me.com

ten „Trefferquoten“ von rund 75 Prozent erreicht werden. „Aber auch diese Ergebnisse sind noch nicht ideal“, erklärt der Professor, der weiß: Das Problem der Prüfung der Informationen in Wissensbanken auf Vollständigkeit und Wichtigkeit ist noch nicht gelöst. Die Lösung war allerdings auch nicht zu erwarten: „Wir wollten das Problem umfassend darstellen, Interesse wecken und Wege zur Lösung andeuten“, so Nutt.

Das ist dem Forscherteam auf der Achse Bozen-Trient in jedem Fall gelungen, es wurde weitere Grundlagenarbeit für die sinnvolle Nutzung der Informationsfülle im Internet geleistet. Bis allerdings ein valides Instrument zur Qualitätskontrolle fehlt, kann auch die für den Nutzer wichtige Frage nicht beantwortet werden, ob die aus dem Internet extrahierten Informationen auch wirklich alles Wichtige wiedergeben und dies in einer sinnvollen Art und Weise tun. Und bis dahin wird man auch an der unibz weiter den Fragen nachgehen: Wie kann man Informationen aus Texten im Internet extrahieren? Wie kann man überprüfen, ob diese Informationen vollständig sind? Und wie kann man sicherstellen, ob die automatisch generierte Antwort auf eine Frage im Internet vollständig, sinnvoll und richtig beantwortet wurde?

Christian Rainer

**BAUGRUNDSTÜCK
IM ÜBERETSCH**

GENEHMIGTES PROJEKT FÜR EINE WOHNANLAGE MIT MEHREREN HÄUSERN

- Möglichkeit der Errichtung einer Hofstelle – bestehendes Weingut
- Hotel in ruhiger Lage
- Privater Luxuswohnsitz
- ...andere Wünsche und Vorstellungen?

Interessiert?
natur3home@gmail.com