

Question Answering: Das Retrieval der Zukunft?

Der Artikel geht der Frage nach, ob und inwieweit Informations- und Recherchesysteme von der Technologie natürlichsprachlicher Frage-Antwortsysteme, so genannter Question Answering-Systeme, profitieren können. Nach einer allgemeinen Einführung in die Zielsetzung und die historische Entwicklung dieses Sonderzweigs der maschinellen Sprachverarbeitung werden dessen Abgrenzung von herkömmlichen Retrieval- und Extraktionsverfahren erläutert und die besondere Struktur von Question Answering-Systemen sowie einzelne Evaluierungsinitiativen aufgezeichnet. Zudem werden konkrete Anwendungsfelder im Bibliotheksweisen vorgestellt.

To what extent can information and research systems profit from the technology of natural language queries, or so-called question-answering systems? After a general introduction to the goals and the historical development of this special area of machine language processing, the author elucidates how it can be distinguished from traditional information retrieval and extraction procedures and elaborates on the particular structure of question answering systems (QA) as well as various evaluative endeavours. Furthermore, the author presents concrete areas of application in the library world.

Wie werden die Benutzerschnittstellen zukünftiger Auskunft- und Retrievalsysteme aussehen? Werden wir weiterhin einzelne Suchbegriffe eingeben, in der Hoffnung, dass diese ausreichen, um die gewünschte Information zu erhalten? Oder werden wir in der Lage sein, natürlichsprachliche Fragen zu formulieren, die das System eindeutig versteht oder auf die es entsprechende Rückfragen stellt, und werden wir als Resultat dieser Anfrage eine einzige und richtige Antwort zu erhalten?

Diese Fragen sind von großer Relevanz für die Welt der Bibliotheken und Archive, hält man sich vor Augen, dass die zunehmende Digitalisierung der Bibliotheken sowie die Digitalisierungsinitiativen führender Suchmaschinen längerfristig dazu führen werden, dass Recherchen weniger auf das Aufsuchen relevanter Dokumente und deren Ranking beschränkt bleiben, sondern dass gleichsam auch die Suche nach den relevanten Informationsbestandteilen *innerhalb* der Dokumente Aufgabe des Recherche- und Retrievalprozesses sein wird.

Dies verlangt vom Benutzer, dass er sein Informationsbedürfnis genauer spezifiziert, beispielsweise in einer oder mehreren vollständig ausformulierten natürlichsprachlichen Fragen. Gleichzeitig steigen die Anforderungen an die maschinelle Verarbeitung der Benutzereingabe, den Retrievalvorgang und die Präsentation der Antwort, die idealerweise ein neues Aggregat aus unterschiedlichen Textfragmenten darstellt. Stellt man der Suchmaschine Ask (www.ask.com) die Frage »Where was Rousseau born?« (siehe Abbildung 1)

erhält man nicht nur die Antwort »Jean-Jacques Rousseau was born in Geneva, Switzerland.«, sondern gleichfalls ein Bild Rousseaus und Hinweise auf weiterführende Informationsquellen.

Question Answering, ein Sonderzweig der maschinellen Sprachverarbeitung, der gelegentlich auch als Answer Mining bezeichnet wird, beschäftigt sich mit dieser Aufgabe und weitet die Suche nach dem relevanten Informationsträger durch eine Suche *im* Informationsträger aus. Es geht folglich um die Erstellung offener Frage-Antwortsysteme, die in der Lage sind, Fragen jeglicher Art mit genau *einer* präzisen und ausreichend kurzen Antwort zu beantworten, die aus einer unübersichtlich großen, elektronischen Datensammlung extrahiert wird.

»You ought to be able to type in ›Why is the sky blue?‹, and get an answer – not just a bunch of people who want to sell you blue paint.« (Bender 2004). Mit diesen Worten wies niemand anderes als Bill Gates seinen Entwicklern den Weg und postulierte im gleichen Atemzug die Lösung dieser Aufgabe als eine der großen Grenzen, die die Informatik noch zu überschreiten hat. Dabei sind bei genauerem Hinsehen nicht allein die allgemeine Informatik, sondern vielmehr ihr verwandte oder beigeordnete Disziplinen wie etwa die Computerlinguistik oder die Künstliche Intelligenz (KI) an der Beschreitung dieses Wegs beteiligt. Gleichzeitig messen auch die Marktführer der Suchmaschinentechnologie dem Bereich des Question Answering immer größere Bedeutung bei, zumal das Benutzerinteresse an Online-Enzyklopädien, die eine Hauptquelle bei der Beantwortung natürlichsprachlicher Fragen sind, immer weiter zunimmt. Da es sich hierbei auch um einen Bereich der Wissensvermittlung handelt, der traditionell den Bibliotheken vorbehalten war, soll zunächst vorgestellt werden, welches Potenzial diese Technologie darstellt und in welchen Bereichen der Informationsvermittlung reelle Einsatzmöglichkeiten bestehen.



René Schneider

offene Frage-Antwort-Systeme



Jean-Jacques Rousseau was born in Geneva, Switzerland
Mostly self-educated in Switzerland, Jean-Jacques Rousseau ended up in Paris, France in the 1740s and became acquainted with Voltaire and Denis Diderot. Rousseau published *Discourse on the Origin and Foundations of Inequality* in 1754, arguing that the natural... [More »](#)

Abb. 1: Where was Rousseau born?
(Quelle: Ask.com 24.10.2006)

EINSATZMÖGLICHKEITEN

Suchmaschinen

Aufgrund der großen Nähe des Question Answering zum Information Retrieval stellen Suchmaschinen das Hauptanwendungsfeld dar. Sowohl die Marktführer als auch Newcomer wie etwa Ask entwickeln seit beträchtlicher Zeit eine ganze Reihe von Algorithmen für den Bereich des Question Answering.

Der Marktführer Google bietet zudem unter www.google.com/answers einen kostenpflichtigen Informationsdienst an, bei dem Benutzer sehr spezifische Fragen stellen können. Die Antworten werden jedoch nicht durch Maschinen, sondern durch Informationsspezialisten gefunden. Einen ähnlichen und zudem kostenfreien Dienst bietet das Internetportal Yahoo! unter der URL <http://ask.yahoo.com>. Die Tatsache, dass beide Dienste offenbar regen Zuspruch finden, liefert Kritikern des Question Answering, die davon ausgehen, dass eine sorgfältige Auswahl von Suchbegriffen natürlichsprachliche Anfragen obsolet macht, wohl den stärksten Gegenbeweis. Eine kurze Lektüre der aktuell eingereichten Fragen zeigt, dass es sehr wohl Informationsbedürfnisse gibt, die sich natürlichsprachlich besser ausdrücken lassen als mit einigen wenigen Suchbegriffen. Noch interessanter wäre es, aktuelle Systeme zum Question Answering anhand dieser Fragen zu testen.

Im Frühjahr 2005 integrierte Google gleichzeitig ein Question Answering-Tool zur Beantwortung faktischer Fragen (Betz 2005). Faktische Fragen erwarten, wie die Bezeichnung unschwer erkennen lässt, als Antworten einfache Fakten, d.h. die Maschine ist nicht genötigt, schwierige Kontextinformation abzuleiten. Diese Fragen sind kurz und beginnen in der Regel mit einem klassischen Fragewort, beispielsweise »Wer ist der Präsident der USA?«, »Wann ist der französische Nationalfeiertag?«, »An welchem Fluss liegt Genf?« und dergleichen. Die Fragen können auch in rudimentärer Form, also nicht unbedingt fehlerfrei eingegeben werden; findet die Suchmaschine eine Antwort, wird diese vor den herkömmlichen Suchergebnissen ausgegeben. Dabei handelt es sich häufig um das Textschnipsel, das die Suchmaschine aus Wikipedia oder einer anderen digitalen Enzyklopädie entnommen hat. Ein statistischer Wert bildet die Grundlage für die Auswahl der Antwort.

Die Suchmaschine Ask (früher: AskJeeves) stellt womöglich die ausgereifteste Version der Kombination aus Suchmaschine und Frage-Antwort-System dar, zumindest handelt es sich um den Suchdienst, der für sich reklamiert, sich am längstem auf dem Gebiet des Question Answering zu bewegen. Auch hier werden im Eingabefeld konkret ausformulierte Fra-

gen erwartet, wenngleich auch hier die herkömmliche Eingabe von Suchbegriffen nicht ohne Antwort bleibt. Findet das System eine Antwort, wird diese in einem gesonderten Ausgabefeld ausgegeben und durch Retrievalergebnisse, also die übliche Auflistung von Suchergebnissen, erweitert. Stellt man bei Ask.com etwa die Frage: »What is the capital of Bolivia?«, erhält man vor den üblichen Links, die auf die relevanten Dokumente verweisen, die Antwort: »The capital of Bolivia is La Paz (seat of government), Sucre (legal capital and seat of judiciary)«.

Chatbots

Wie der Rückblick auf die Entstehungsgeschichte von Systemen zum Question Answering zeigen wird, bestand in den Anfangsjahren noch eine große Nähe zwischen eher dialogorientierten Systemen (etwa das berühmte System ELIZA, Weizenbaum 1966) und eher wissensorientierten Systemen (etwa das weniger bekannte SHRDLU, Winograd 1977) und erstere entwickelten sich schnell zu dem, was heute im allgemeinen unter Chatbots verstanden wird. Die Aufgabe von Chatbots besteht hauptsächlich darin, allgemeine Fragen zu beantworten, die man in etwa am Empfang oder im Servicecenter einer Bibliothek stellen würde. Im Bibliotheksbereich wird mittlerweile eine ganze Reihe von Chatbots für die Kommunikation mit Online-Benutzern verwendet. Diese Anwendung ist dabei, sich als festes Produkt in der Kundenkommunikation zu etablieren, und es wurde mittlerweile eine eigene Beschreibungssprache (AIML = Artificial Intelligence Markup Language) für sie entwickelt. Eine repräsentative Auswahl englischsprachiger Bots findet sich unter der URL www.pandorabots.com/botmaster/en/mostactive.

Wirft man allerdings einen Blick in die XML-Dateien, die den Chatbots zugrunde liegen, oder rekonstruiert man deren Konzeption, wird man sehr schnell ernüchtert feststellen, dass die wenigsten dieser Chatbots über eine ausgereifte natürlichsprachliche Flexibilität verfügen und immer noch dem rein behavioristischen Paradigma des stimulus-response verpflichtet sind. Größtenteils bestehen diese Dateien aus einfachen Frage-Antwortmustern, deren Algorithmen weder mit Tippfehlern noch mit einer gewissen syntaktischen Varianz umgehen können. Die daraus resultierenden starren Dialoge führen somit in der Regel sehr schnell zu frustrierten bis gereizten Benutzern (De Angeli et al. 2001).

An dieser Stelle könnte die Integration von Question Answering-Technologie dazu führen, die formale und inhaltliche Beschränktheit von Chatbots aufzubrechen, um deren Wissen zu erweitern und sie in ih-

Informationsspezialisten

Textschnipsel

rer Dialogfähigkeit flexibler und benutzerfreundlicher zu gestalten.

Digitale Bibliotheken

Wie wenige andere Begriffe ist die Digitalisierung von Bibliotheken zum Thema für Anbieter und Nutzer bibliothekarischer Dienstleistungen geworden. Dabei sind sowohl informationsethische als auch rechtliche Argumente Treibstoff für die derzeit geführte Diskussion, allerdings gerät die Interaktion im erweiterten Informationsangebot, insbesondere die Suche und Navigation, leicht in den Hintergrund. So führt eine kurze Evaluation der Beta-Version von Google Book Search und weiteren Evaluierungsinitiativen sehr schnell vor Augen, dass die bislang entwickelten Benutzerschnittstellen für den Umgang mit digitalisierten Bibliotheken weder effektiv noch effizient sind (Schneider 2006). Gerade hier könnte eine intelligente Fusion von einerseits bibliothekarischem Wissen in Form der Metadaten (Kataloginformation, Taxonomien) und andererseits von Suchmaschinen- und Chatbottechnologie auf unterschiedlichen Ebenen der Informationsvermittlung eine erhebliche Verbesserung darstellen.

Dies kann zuvorderst dazu dienen, allgemeine Informationen über Bibliotheksbestände zu vermitteln, etwa auf Fragen wie: »Ich suche das Buch zum Film ›Das Parfum‹«, oder ganz allgemeine Fragen wie etwa: »Gibt es das Gesamtwerk von Shakespeare auch zweisprachig?«. Darüber hinaus gilt es, sich vor Augen zu halten, dass viele zukünftige ›Besucher‹ digitaler Bibliotheken immer weniger ein Interesse an digitaler Lektüre haben, sondern gezielt nach suchrelevanten Textpassagen oder multimedialer Information suchen. Dies kann häufig in Fragebündeln geschehen, etwa nach folgendem Muster: »Wann wurde der Halleysche Komet entdeckt?«, »In welchem zeitlichen Abstand passiert er die Erde?«, »Welche weiteren Kometen kommen zyklisch an der Erde vorbei?« usw.

Gerade Experten, die häufig eine langjährige Erfahrung im Umgang mit Suchmaschinen gewonnen haben, mag diese Abkehr von wohldefinierten Termen und deren Kombination ungewohnt erscheinen; gleichzeitig zeigt die alltägliche Praxis der Internetbenutzer, dass mit einem oder zwei Suchbegriffen meistens gefunden wird, was man sucht. Andererseits sollte man die Konditionierung, die durch die Suchmaschinen vorgenommen wird, nicht unterschätzen und gerade das anfangs besprochene große Engagement der Marktführer zeigt, dass das Potenzial dieser Technologie beträchtlich ist und sich nicht allein auf die in diesem Abschnitt beschriebenen Anwendungsbereiche beschränkt. Um diese Technologie genauer zu verste-

hen, sollen im Folgenden die Entstehungsgeschichte und der generelle Aufbau von Systemen zum Question Answering beschrieben werden.

ENTWICKLUNGSGESCHICHTE VON QUESTION ANSWERING SYSTEMEN

Die ersten Systeme sind nahezu so alt wie die Computer selbst und wurden in den 50er Jahren realisiert, um Datenbankabfragen durch natürlichsprachliche Anfragen flexibler zu gestalten. Als erster Prototyp gilt *The Conversational Machine*, ein System, das Auskünfte über das aktuelle Wetter ermöglichte. Eine erste Evaluierung bestehender Systeme fand in den sechziger Jahren statt (Simmons 1965) und umfasste 15 Systeme, die jeweils unterschiedliche Ansätze aus dem Bereich der sich damals etablierenden Künstlichen Intelligenz verfolgten. Zu diesem Zweck wurden Texte elektronisch indexiert und allgemeines Wissen in Attribut-Wert-Paaren (nach dem Muster: [Land: Schweiz]; [Hauptstadt: Bern]; [Nationalfeiertag: 1. August]) dargestellt. Gleichzeitig wurde damit begonnen, natürlichsprachliche Sätze in ihrem Aufbau (Syntax) und ihrer Bedeutung (Semantik) zu erfassen.

Allen Systemen gemein war die Beschränkung auf eine einzelne Domäne, d.h. ein spezielles Wissensgebiet. Innerhalb dieser Domäne wiesen die Systeme allerdings bereits damals eine hohe Performanz auf, so konnte beispielsweise LUNAR (Woods 1977), ein System zur Beantwortung von Fragen zur geologischen Gesteinsanalyse der Apollo-Missionen, 90 % der Fragen korrekt beantworten. Weitere Systeme wie SHRD-LU, das sich mit Computeroperationen innerhalb der für die KI typischen Blockwelten beschäftigte, verfeinerten diesen Ansatz, wohingegen der Schwerpunkt von Dialogsystemen wie ELIZA im Führen und Aufrechterhalten von Konversationen bestand. Dieser Ansatz koppelte sich von herkömmlichen Question Answering-Systemen ab und wurde – wie bereits erwähnt – in Chatbots weiterverfolgt.

Parallel dazu wurden in den 70er und 80er Jahren sog. Expertensysteme sowie erste Fragetaxonomien (Lehnert 1977) entwickelt. Der im Bereich der maschinellen Sprachverarbeitung zu dieser Zeit vorherrschende hohe theoretische Anspruch und die damit einhergehende Fokussierung auf sprachliche Genauigkeit führten jedoch gleichzeitig zu einem Mangel an Robustheit und Flexibilität der realisierten Systeme.

Eine pragmatische Neuorientierung entstand erst im Rahmen der Evaluierungsinitiativen zum Information Retrieval sowie der Informationsextraktion: der TREC (= Text REtrieval and Evaluation Conference, <http://trec.nist.gov>) und MUC (= Message Understanding Conference, www-nlpir.nist.gov/related_projects/)

Fusion von bibliothekarischem Wissen und Suchmaschinenteknologie

Mangel an Robustheit und Flexibilität

muc/), auf die im weiteren Verlauf des Artikels genauer eingegangen wird. Nur auf diesem Hintergrund sind der gegenwärtige Stand der Forschung sowie die große Nähe des Question Answering zum Information Retrieval einerseits und zur Informationsextraktion andererseits zu verstehen. Hauptziel aller Anstrengung seit den 90er Jahren ist das Aufbrechen der Bindung an spezielle Wissensgebiete (oder Domänen), man spricht von daher häufig auch von Open Domain Question Answering.

Der allgemeine Verlauf der Entwicklung von Systemen zum Question Answering zeigt, dass hier ähnlich zur allgemeinen Entwicklung der Informatik eine Hinwendung von festen Datenstrukturen, die in einer Datenbank abgelegt wurden, zu offenen und kaum strukturierten Datenkollektionen etwa in Nachricht-

tenkorpora oder dem Web stattfand. Einige Systeme haben sich auf die Wissensextraktion in semi-strukturierten Datenkollektionen, beispielsweise in digitalen Enzyklopädien, spezialisiert. Diese besitzen den Vorteil, dass sie in der Regel eine mehr oder weniger fest vorgegebene Struktur in Form von mehr oder weniger fest belegten Rubriken mit entsprechenden Überschriften besitzen, die Rückschluss auf die Inhalte und somit wertvolle Information für die Erschließung des Fließtexts, also der unstrukturierten Information, zulässt.

QUESTION ANSWERING VS. INFORMATION RETRIEVAL & INFORMATION EXTRACTION

Question Answering wird häufig auch mit den Begriffen des ›short passage information retrieval‹ und der ›open domain information extraction‹ charakterisiert, um sowohl eine Abgrenzung von diesen beiden Disziplinen zu erreichen als auch die große verwandtschaftliche Nähe dieser Gebiete zu unterstreichen. Diese Erweiterungen des herkömmlichen Retrieval- und Extraktionsbegriffs lassen sich folgendermaßen erklären:

Das klassische Dokumentenretrieval (siehe Abbildung 2) geht von einem Dokumentenkörper aus, der – ausgehend von einigen frei zu wählenden bzw. miteinander verknüpften Suchbegriffen – nach relevanten Dokumenten durchforstet wird. Als Resultat dieses Prozesses wird eine Liste dieser Dokumente, idealerweise mit einem Verweis, wo diese zu finden sind, sowie einer Rangfolge, die die Relevanz aufzeigt, erwartet. Jede Suchmaschine folgt noch heute diesem Wechselspiel aus Suchbegriffen und Dokumentverweisen.

Die erste Erweiterung zum herkömmlichen Retrieval besteht nun darin, dass – nach Eingabe einer vollständigen natürlichsprachlichen Frage – in den Texten selbst weiter gesucht wird und aus dem Fließtext ein einzelnes Textschnipsel isoliert wird, das mit hoher Wahrscheinlichkeit die Antwort auf die gestellte Frage enthält (siehe Abbildung 3). Dieser Vorgang wird häufig auch als ›passage retrieval‹, ›short answer extraction‹ oder ›answer mining‹ bezeichnet und entspricht genau der Aufgabenstellung, die in den ersten Evaluierungen zum Question Answering gestellt wurde. Auch Suchmaschinen können diese Fragen bereits recht gut beantworten. Gibt man beispielsweise die Frage »When was Rousseau born?« in dieser Form bei Google ein, geben sowohl die Textpassagen, die auf die Dokumente verweisen oder die darin befindlichen Textpassagen die Antwort. Das Weiterklicken erübrigt sich somit streng genommen (siehe Abbildung 4).

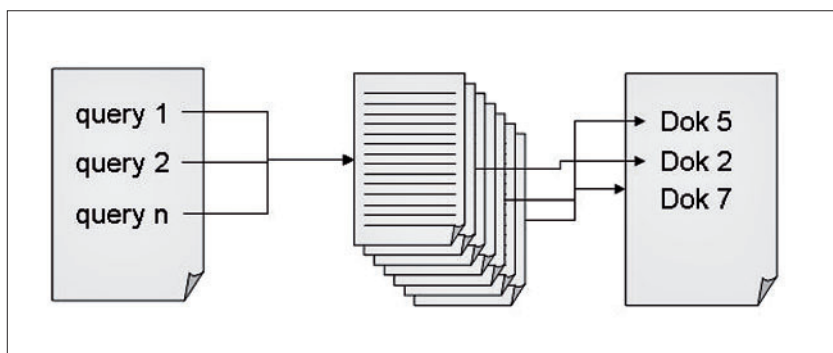


Abb. 2: Klassisches Dokumentenretrieval

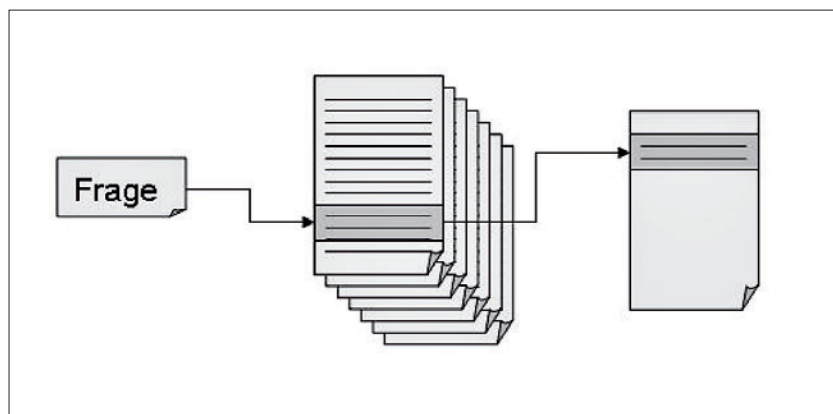


Abb. 3: Passage Retrieval

Jean Jacques Rousseau (1712-78).

Jean Jacques **Rousseau**, born at Geneva, was deserted by his family at age 10. As a young man he ran away from his caretakers and was to be referred, ...
www.blupete.com/Literature/Biographies/Philosophy/Rousseau.htm

Abb. 4: When was Rousseau born?
Quelle: Google.com 24.10.2006

Diese spezielle Erweiterung des Retrieval, das sog. passage retrieval, stellt gleichzeitig eine Verbindung zur Informationsextraktion dar.

Bei der Informationsextraktion (siehe Abbildung 5) werden allerdings weder Suchbegriffe noch eine ausformulierte Frage erwartet, sondern vielmehr eine komplexe Aufgabenstellung, für die vorher ein spezielles ›Template‹, also eine Art formale Schablone, erwartet wird, welche die jeweilige Domäne der Aufgabenstellung und die zu identifizierenden Textelemente abbildet. Diese Templates bestehen in der Regel aus einer Reihe von Attribut- und Wertpaaren, die in einer ›slot-and-filler‹ Architektur resultieren: entsprechend der vorgegebenen Aufgabe werden feste Einträge (slots) definiert, die nach Beendigung des Suchalgorithmus durch korrespondierende Textbausteine aufzufüllen sind (filler).

Das Ziel der Informationsextraktion besteht demnach darin, eine Textsammlung (in der Regel Nachrichtentexte) nach für die Aufgabenstellung relevanten Ereignissen zu durchsuchen und die dabei gefundenen Textschnipsel im Template als Bündel von Attribut-Wertpaaren abzulegen. Abbildung 6 zeigt ein solches Template mit den zugehörigen slots (beispielsweise Datum, Ort, Ziel, Effekt) und den im Text aufgefundenen Entsprechungen. Unschwer lässt sich erkennen, dass es sich bei der Suchmenge um Texte mit einem terroristischen Hintergrund in Mittelamerika handelt.

Zur weiteren Illustration sei auf andere Aufgabenstellungen verwiesen, die im Rahmen der MUC-Initiative an die Teilnehmer gestellt wurden. Diese hatten – auch aufgrund der Förderung durch die amerikanische Regierung – häufig einen politischen Hintergrund und betrafen neben dem o.g. Beispiel aus den Jahren 1991 und 1992 (siehe Abbildung 6), Unternehmensfusionen (1993) oder Personalveränderungen auf der Managementebene großer Unternehmen (1995).

Wie im passage retrieval stehen auch im Fall der Informationsextraktion Textpassagen im Zentrum des Interesses, allerdings sind es nun mehrere Bestandteile eines oder mehrerer Texte, die in einem übergeordneten Zusammenhang zueinander stehen.

Daraus ergeben sich auch spezielle Aufgabenstellungen der Informationssuche, die im Zeitraum der MUC-Konferenzen gestellt wurden, beispielsweise zur Ko-Referenzresolution (also der Aufdeckung von Personalpronomen und weiteren Referenzen) und vor allen Dingen zur named entity recognition, also der Erkennung von Eigennamen. Hierbei wird der Begriff der Eigennamen jedoch sowohl sehr weit gefasst als auch sehr fein unterschieden. In diesen Bereichen wie auch in anderen Aufgaben aus der morphologischen und syntaktischen Informationsverarbeitung wurden

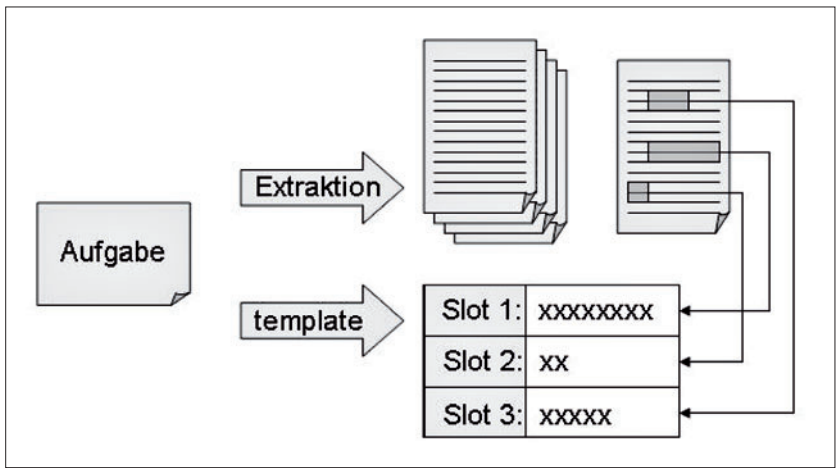


Abb. 5: (Closed-domain) information extraction

INCIDENT TYPE	bombing
DATE	March 19
LOCATION	El Salvador: San Salvador
PERPETRATOR	urban guerilla commandos
PHYSICAL TARGET	power tower
HUMAN TARGET	-
EFFECT ON PHYSICAL TARGET	destroyed
EFFECT ON HUMAN TARGET	no injury or death
INSTRUMENTATION	bomb

Abb. 6: Template für Terroranschläge (Quelle: Grishman 1997)

respektable Ergebnisse erzielt und Erkenntnisse gewonnen, die heutzutage zum soliden Fundament des Question Answering gehören.

Größere Probleme bereitet nach wie vor jedoch die semantische Ebene. Dieses Problem zeigt sich vor allen Dingen darin, dass für jede neue Aufgabenstellung, d.h. jede neue Anwendungsdomäne nicht nur ein neues Template, sondern gleichfalls eine neue Wissensbasis erstellt werden muss. In diesen Zusammenhang passt auch die Tatsache, dass die MUC-Evaluationen 1997 nach siebenmaligem Stattfinden nicht fortgeführt wurden. Ausschlaggebend war einerseits der Umstand, dass die meisten Teilnehmer aufgrund der Erfahrungen vorangegangener Konferenzen dieselben oder ähnliche Algorithmen für die jeweiligen Einzelmodule verwendeten und somit auch vergleichbare Resultate bezüglich der Performanz der einzelnen Systeme erreichten. Zum anderen wiesen sämtliche Systeme einen Reifegrad auf, der erkennen ließ, dass weit reichende Verbesserungen nicht unbedingt zu erwarten waren.

Dennoch sind gerade die im Bereich der Informationsextraktion erreichten Ergebnisse als nicht gering einzuschätzen. Viele Experten gehen sogar davon aus,

Textpassagen im Zentrum des Interesses

dass die maschinelle Sprachverarbeitung in der Informationsextraktion erstmals einen pragmatischen Ansatz fand, der zu robusten und effizienten Systemen ohne theoretische Überfrachtung führte (Grishman 2003).

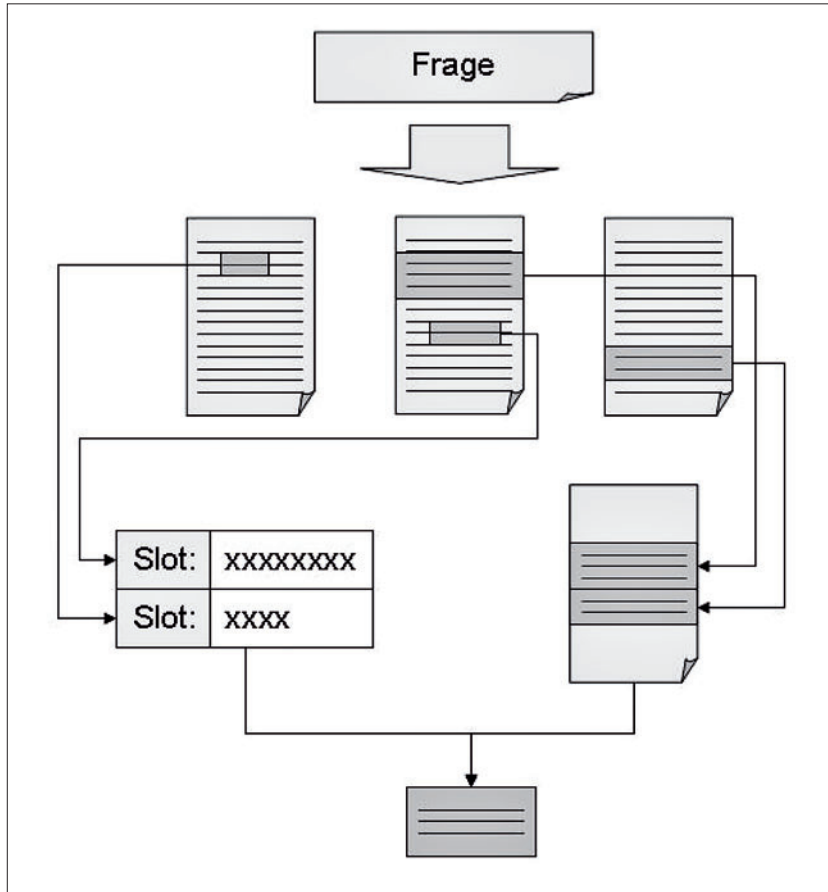


Abb. 7: Question Answering als Fusion von passage retrieval und open domain information extraction

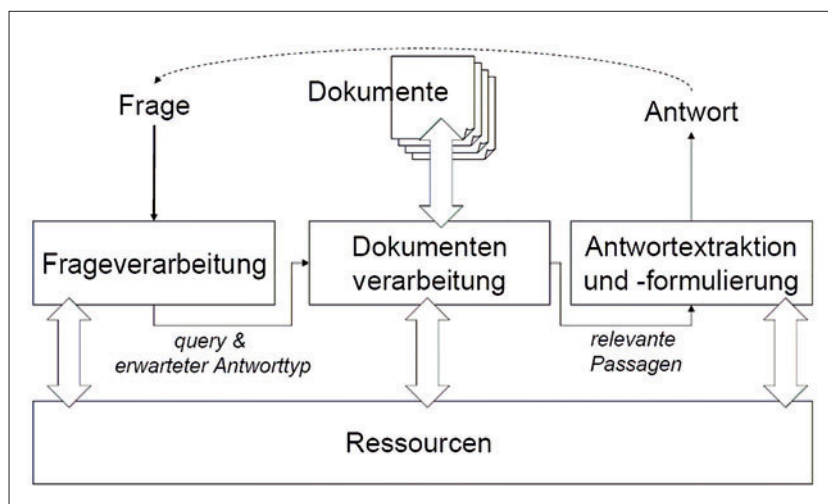


Abb. 8: Architektur eines Question Answering-Systems

QUESTION ANSWERING SYSTEME: ANSPRUCH UND REALISIERUNG

Im Question Answering (siehe Abbildung 7) werden nunmehr die Such- und Extraktionsstrategien der beiden verwandten Teildisziplinen passage retrieval und open domain information extraction miteinander verbunden und um entsprechende Algorithmen erweitert.

Der größte Anspruch des Question Answering ist es, sowohl die Grenzen des Retrievals (in der Beschränkung auf die Dokumente an sich) und der Informationsextraktion (in der Beschränkung auf eine Domäne) aufzuheben beziehungsweise neu abzustechen. Die Umsetzung dieses Anspruchs in die Tat besteht nicht nur in einer Fusion und Weiterentwicklung der bestehenden Algorithmen, sondern darüber hinaus in einem Rückgriff auf die Erkenntnisse der Künstlichen Intelligenz. Eine zunehmend größere Bedeutung gewinnen zudem die Verfahren aus dem Bereich des Maschinellen Lernens.

Die Architektur eines Systems zum Question Answering (siehe Abbildung 8, vgl. Harabagiu/Moldovan 2003, S. 567) ist in ihrem Kern eher einfach und mit vergleichsweise geringem Aufwand lässt sich ein einfaches System erstellen (Strötgen et al. 2005), auch wenn dessen Leistung nicht optimal ist. Elementar ist von daher die Verfeinerung der einzelnen Module zur Erhöhung der Performanz. In vielen Fällen bedarf dieser Prozess nicht nur der Entwicklung elaborierter Algorithmen, sondern aufwändiger Handarbeit zur Verarbeitung der unterschiedlichen Wissensbasen.

Jedes System besteht prinzipiell aus drei Hauptbestandteilen: zunächst wird die eingegebene Frage – in der Regel mit Hilfe einer entsprechenden Frage-taxonomie – maschinell analysiert. Nachdem die Frage klassifiziert und der zu erwartende Antworttyp generiert wurde, findet ein herkömmliches Retrieval statt; der vollständige Fragestring wird dazu in der Regel in mehrere Queries (Suchbegriffe) zerlegt. Häufig werden an dieser Stelle verschiedene Algorithmen zur Expansion oder Reduktion der Queries unter Zugriff auf entsprechende Ressourcen verwendet. Anschließend werden die relevanten Dokumente, die mit höchster Wahrscheinlichkeit die Antwort oder Teile der Antwort enthalten, ausgewählt und an das Modul zum Passagenretrieval übergeben. Hier wird nunmehr aus einer oder mehreren Passagen der Antwortstring extrahiert bzw. zusammengesetzt.

Insbesondere in den Modulen zur Frageverarbeitung sowie der Antwortextraktion finden sich die größten Abweichungen beziehungsweise Erweiterungen vom herkömmlichem Information Retrieval und der Informationsextraktion: die Antizipation eines

Antworttyps sowie die Suche nach der richtigen Passage und die Auswahl der wahrscheinlichsten Antwort. Häufig wird erst hierbei wieder auf das im ersten Modul, der Frageverarbeitung, abgeleitete Wissen zurückgegriffen. So kann die Ableitung des Fragetyps genaueren Aufschluss über den Antworttyp geben; beispielsweise wird in »Wer«-Fragen in der Regel nach einer Person gesucht. Dieses Wissen hilft dabei, Antworten, die keinen Personennamen enthalten, als Antwortkandidaten auszuschließen. Ein weiteres Hilfsmittel sind unterschiedliche Heuristiken zur statistischen Gewichtung der Passagen, die zum Beispiel die Anzahl, Dichte und Verteilung der Suchbegriffe berücksichtigen und in entsprechenden Formeln miteinander verrechnen. Von nicht zu unterschätzender Bedeutung sind in jedem Fall die unterschiedlichen Wissensressourcen, die ins Gesamtsystem integriert werden: dies sind insbesondere elektronische Wörterbücher und digitale Enzyklopädien.

Einen profunden Überblick über den derzeitigen Anspruch und die Vorhaben des Question Answering geben Burger et al. (2002) in ihrer Roadmap für die weitere Entwicklung dieser Technologie. Hier spielen neben der maschinellen Sprachverarbeitung, das heißt der elaborierten Verarbeitung von Fragen und der Extraktion einer Antwort aus heterogenen Datenquellen, weit reichendere Ansätze eine Rolle. Diese beschäftigen sich mit besonderen Problemen, etwa der Multilingualität und einer weit reichenden Kontextanalyse. Des Weiteren werden die Fusion multipler und heterogener Datenquellen sowie multimediales und benutzeradaptives Question Answering thematisiert. Anzumerken ist hierbei jedoch, dass diese Ansätze, je weiter sie sich vom Kerngebiet des Question Answering entfernen, auch mit allgemeinen Neuerungenstendenzen des Retrieval bzw. der Mensch-Maschine-Interaktion im Allgemeinen konvergieren. Einen weiteren Überblick über Entwicklungstendenzen gibt Maybury (2004).

EVALUIERUNGEN

Stellenwert und Entwicklungsstand des Question Answering lassen sich hauptsächlich an der Evaluierungsinitiative des NIST (= National Institute for Standards and Technology), einer der amerikanischen Regierung untergeordneten Behörde, ablesen. Seit 1999 findet in jährlichem Abstand ein eigener Question Answering-Track im Rahmen der Text Retrieval and Evaluation Conference (TREC) statt. Die sich daran anschließenden Veröffentlichungen der Ergebnisse sowie die ausgewählten Darstellungen einzelner Systeme dokumentieren jährlich die Herausforderungen, Ergebnisse und Grenzen dieser Disziplin.

So wurden in den Anfangsjahren hauptsächlich die bereits vorgestellten faktischen Fragen nach dem Muster »Wer ist ____?« bzw. »Wo liegt ____?« usw. gestellt mit der Aufgabe, Textschnipsel, welche die jeweilige Information beinhalten, als Antwort zu liefern. Dieser Fragetyp macht auch heute noch den Hauptbestandteil aller Fragen aus; erwartet wird mittlerweile *eine* genaue oder eben keine Antwort. Die Ergebnisse dieses Fragetyps werden durch Menschen evaluiert und dabei entweder als korrekt, ungenau oder falsch klassifiziert. Die Erfolgsrate des Systems berechnet sich aus dem Prozentsatz aller korrekten Antworten. Dieser Fragetyp wurde durch so genannte Listen-Fragen nach dem Muster »In welchen Ländern ____?«, »Finden sie n Beispiele für ____?« erweitert. Die Antwort ist hierbei eine ungeordnete Liste von korrekten Begriffen, die mittels des im Information Retrieval verwendeten F-Werts (bei einer Gleichgewichtung der Werte für Precision und Recall) evaluiert werden.

Alle weiteren Fragen werden unter der Typbezeichnung »Other« zusammengefasst, sie spielen in den aktuellen Evaluierungen jedoch quantitativ betrachtet eine untergeordnete Rolle.

Die unterschiedlichen Ergebnisse werden dann zu einem gemeinsamen Wert fusioniert. Dabei variiert die Performanz stark zwischen den einzelnen Teilnehmern. Die allgemeine Erfolgsrate liegt beispielsweise im Fall der faktischen Fragen zwischen 60 und 20 %, allerdings liegen mehr als zwei Drittel der Teilnehmer unter 30 %. Genaueren Aufschluss geben die jeweiligen Veröffentlichungen der nach den Tracks abgehaltenen Konferenzen (<http://trec.nist.gov/pubs.html>).

Gegenwärtig werden unterschiedliche Fragen zu einem Thema gruppiert. Wie das in Abbildung 9 wiedergegebene Beispiel verdeutlicht, referieren die Fragen dabei unterschiedlich stark aufeinander. Einen umfassenden Überblick über sämtliche Fragen der einzelnen Evaluierungen gibt die Website der NIST unter der URL http://trec.nist.gov/data/qa/t2004_qadata.html; ein allgemeiner Überblick über die Ergebnisse der jeweiligen Teilnehmer findet sich unter der URL http://trec.nist.gov/presentations/t2004_presentations.html.

Das in Abbildung 9 wiedergegebene Beispiel eines Fragesets, das sich aus einem zunächst unter »target« angeführten Schwerpunkt ergibt, spiegelt zudem eine Weiterentwicklung des herkömmlichen QA-Ansatzes wider: Es geht nicht mehr darum, einzelne Fragen zu beantworten, sondern einen allgemeinen Rechercheprozess abzuwickeln, wie er in unterschiedlichen Anwendungen der Informationswirtschaft vorstellbar ist.

Parallel zu den Evaluierungen der TREC existiert in Europa seit 2003 eine multilinguale Variante dieses

Roadmap

Erfolgsrate zwischen 60 % und 20 %

Herausforderungen, Ergebnisse und Grenzen

Abb. 9: Beispielfragen aus dem TREC QA-Track (Quelle: http://trec.nist.gov/data/qa/2005_qadata/QA2005_testset.xml)

```
<target id="75" text="Merck & Co.">
  <q id="75.1" type="FACTOID">
    Where is the company headquartered?
  </q>
  <q id="75.2" type="FACTOID">
    What does the company make?
  </q>
  <q id="75.3" type="FACTOID">
    What is their symbol on the New York Stock Exchange?
  </q>
  <q id="75.4" type="FACTOID">
    What is the company's web address?
  </q>
  <q id="75.5" type="LIST">
    Name companies that are business competitors.
  </q>
  <q id="75.6" type="FACTOID">
    Who was a chairman of the company in 1996?
  </q>
  <q id="75.7" type="LIST">
    Name products manufactured by Merck.
  </q>
  <q id="75.8" type="OTHER">
    Other
  </q>
</target>
```

Tracks innerhalb des Cross Lingual Evaluation Forum (CLEF), die jährlich um diverse Sprachen erweitert wird (<http://clef-qa.itc.it/>). Hierbei kommt dem Englischen, das sowohl als Quell- als auch als Zielsprache in den unterschiedlichen Sprachpaarungen fungiert, eine besondere Rolle zu. Aufgrund des ungleich schwierigeren Hintergrunds der Multilingualität ist die Erfolgsquote der teilnehmenden Systeme um einiges geringer.

Insgesamt betrachtet stellen beide Evaluierungsinitiativen mit dem Charakter eines kollegialen Wettbewerbs einen ungemeinen Vorteil für das Gebiet des Question Answering dar: aufgrund der jährlich gestellten Aufgaben und den von Jahr zu Jahr steigenden Schwierigkeitsgraden begrenzen sie die theoretische Forschung und lenken die Entwicklung in konkrete Anwendungszusammenhänge. Der Nachteil mag darin bestehen, dass die einzelnen Teilnehmer durch Antizipation der Aufgaben und Adaption an die jeweilige ›Evaluierungstradition‹ ihre Systeme zu stark an die jeweiligen Tracks anpassen. Spontanevaluationen, bei denen die Fragen nicht eine geraume Zeit vorher den Teilnehmern zugesandt werden, sondern die vor Ort oder via Internet getestet werden, könnten hier eine allgemeinere bzw. kritischere Evaluation ermöglichen.

FAZIT

Der Artikel widmete sich dem Gebiet des Question Answering, einer speziellen Form der Informationssuche, die eine Erweiterung des herkömmlichen Retrievalbegriffs darstellt. Es ist festzustellen, dass viele grundsätzliche Probleme dieser Technologie als gelöst gelten können, auch wenn von einer allgemeinen Marktreife nicht unbedingt zu sprechen ist.

Die Frage, ob Question Answering dauerhaft den herkömmlichen Retrievalansatz ablösen wird, so wie dies im Titel des Artikels angesprochen wird, kann von daher derzeit nicht eindeutig beantwortet werden. Die Performanz aktueller Systeme wird sich sicherlich noch verbessern lassen, nur wird hier ähnlich wie im Fall der Informationsextraktion in näherer Zeit eine obere Marge erreicht werden, die nur mit größten Anstrengungen zu überwinden sein wird. Andererseits bietet die Informationswirtschaft allgemein ein großes Betätigungsfeld für diese Technologie, auch in multimedialen bzw. multimodalen Erweiterungen wie etwa im Bereich des Spoken Question Answering oder der Integration in Smart-Phones und mobile digitale Assistenten, um Question Answering zu einem festen Modul in Auskunftssystemen jeder Art werden zu lassen.

Zu klären bleibt darüber hinaus, welche Form der Eingabe Benutzer bevorzugen: ob sie weiterhin suchmaschinenkonform eine Handvoll Wörter eingeben möchten oder für tiefer gehende Recherchen natürlichsprachliche Fragen bevorzugen. Entsprechende Studien zur Rezeption natürlichsprachlicher Schnittstellen vs. termorientierten Anfragen stehen nach Kenntnis des Verfassers noch aus.

Allerdings ist es äußerst wahrscheinlich, dass sich die im Titel geäußerte Frage in einigen Jahren nicht weiter oder nicht mehr in dieser Form stellen wird. Mit hoher Wahrscheinlichkeit werden die Ansätze und Ergebnisse des Question Answering – ähnlich wie der Bereich der automatischen Textzusammenfassungen – integrative Bestandteile der kommenden Generation von Suchmaschinen und Auskunftssystemen sein und, so bleibt zu hoffen, benutzerfreundliche digitale Bibliotheken gleichfalls von dieser Technologie Gebrauch machen.

LITERATUR

- Bender, Eric:** The Ultimate Change Agent. Technology Review, (2004-02.27). www.technologyreview.com/read_article.aspx?id=13469&ch=infotech. – [Stand 26.07.2006] Deutsche Übersetzung unter www.heise.de/tr/artikel/45276. – [Stand 26.07.2006]
- Betz, Jonathan:** Just the facts, fast. <http://googleblog.blogspot.com/2005/04/just-facts-fast.html>. – [Stand 26.07.2006]

- Burger, John; Cardie, Claire; Chaudhri, Vinay; Gaizauskas, Robert; Harabagiu, Sanda; Israel, David; Jacquemin, Christian; Lin, Chin-Yew; Maiorano, Steve; Miller, George; Moldovan, Dan; Ogden, Bill; Prager, John; Riloff, Ellen; Singhal, Amit; Shrihari, Rohini; Strzalkowski, Tomek; Voorhees, Ellen; Weishedel, Ralph:** Issues, Tasks and Program Structures to Roadmap Research in Question & Answering (Q&A) 2002. www-nlpir.nist.gov/projects/duc/papers/qa.Roadmap-paper_v2.doc. – [Stand 26.07.2006]
- De Angeli, Antonella; Johnson, Graham I.; Coventry, Lynne:** The unfriendly user: exploring social reactions to chatterbots. In: Helander, M.G. et al. (Hrsg.): *Proceedings of The International Conference on Affective Human Factors Design*. Ondond: Asean Academic Press, 2001. S. 467–474.
- Grishman, Ralph:** Information Extraction. In: Mitkov, Ruslan et al. (Hrsg.): *The Oxford Handbook of Computational Linguistics*. Oxford: Oxford University Press, 2003. S. 545–559.
- Grishman, Ralph:** Information Extraction: Techniques and Challenges. In: Maria Teresa Pazienza (Hrsg.): *Information Extraction. Lecture Notes in Artificial Intelligence*. Rome: Springer, 1997. S. 10–27.
- Harabagiu, Sanda; Moldovan, Dan:** Question Answering. In: Mitkov, Ruslan et al. (Hrsg.): *The Oxford Handbook of Computational Linguistics*. Oxford: Oxford University Press, 2003. S. 560–582.
- Hirschman, Lynette; Gaizauskas, Robert:** Natural Language Question Answering: The View from Here. In: *Natural Language Engineering* 7,4 (2001), S. 275–300.
- Lehnert, Wendy:** A Conceptual Theory of Question Answering. In: *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*, 1977. S. 158–164.
- Maybury, Mark T.:** *New Directions in Question Answering*. Boston: AAAI Press, 2004.
- Schiffhauer, Nils:** Das Wissen der Welt aus dem Netz und von der Scheibe. In: *Frankfurter Allgemeine Zeitung*, (2005–10–18), S. T 1.
- Schneider, René:** Virtuelle Eselsohren – Über das Lesen im Virtuellen Zeitalter. In: *arbido print, E-Books*, 4/2006, S. 7–10.
- Simmons, Robert F.:** Answering English Questions by Computer: A Survey. In: *Communications of the ACM*, 8,1 (1965), S. 53–70.
- Strötgen, Robert; Mandl, Thomas; Schneider, René:** A Fast Forward Approach to Cross-lingual Question Answering for English and German. In: *Working Notes of the 6th Workshop of the Cross-Language Evaluation Forum, CLEF 2005*. Wien: Sept. 2005.
- Turing, Alan M.:** Computing machinery and intelligence. In: *Mind*, 59 (1950), S. 433–560.
- Voorhees, Ellen M.; Buckland, Lori P.:** The Fourteenth Text Retrieval Conference Proceedings (TREC 2005). http://trec.nist.gov/pubs/trec14/t14_proceedings.html. – [Stand 26.07.2006]
- Weizenbaum, Joseph:** Eliza – A computer program for the study of natural language communication between man and machine. In: *Communications of the ACM*, 9,1 (1966), S. 36–45.
- Winograd, Terry:** Five lectures on artificial intelligence. In: Antonio Zampolli (Hrsg.): *Linguistic Structures Processing*. New York: North Holland, 1977. S. 399–520.
- Woods, W.A.:** Lunar rocks in natural English: Explorations in natural language question answering. In: Antonio Zampolli (Hrsg.): *Linguistic Structures Processing*. New York: North Holland, 1977. S. 521–569.

DER VERFASSER

Prof. Dr. René Schneider ist Professor für Informationswissenschaft an der Genfer Haute Ecole de Gestion, Campus de Battelle, Bât. F, Route de Drize 7, CH–1227 Carouge, Schweiz, und koordiniert den dortigen zweisprachigen Studiengang für Information und Dokumentation.
rene.schneider@hesge.ch