

Manuskript (aus Schlussvortrag Frankfurt 20. Mai 2009)

MAX – eine Maschine, mit der man sprechen kann

Ipke Wachsmuth (Universität Bielefeld)

Der Beitrag geht der Frage nach, ob wir eines Tages mit »künstlichen Intelligenzen« als gleichwertige Kommunikationspartner sprechen können, die verstehen, was wir von ihnen wollen und die die Rolle eines sozialen Gegenübers einnehmen. Im Internet existieren schon virtuelle Helfer, sog. Avatare, die mit dem Menschen kommunizieren und ihm den Weg durch digitale Welten weisen. Max, der in der Dauerausstellung des Heinz Nixdorf MuseumsForums (HNF) in Paderborn zu sehen ist, ist ein solcher Avatar, eine virtuelle Person. Er kann in seiner virtuellen Welt auf menschliche Anweisung spezifische Aufträge erledigen. Er kommuniziert nicht nur sprachlich, sondern auch mit Gestik und Mimik; sein Verhalten ist dem des Menschen nachgebildet. Entwickelt wurde Max an der Universität Bielefeld im Zusammenhang mit der Erforschung der menschlichen Kommunikation und ihrer Modellierung in technischen Systemen. Im letzten Teil des Beitrags wird auch ein Bogen zu künftigen Anwendungsfeldern geschlagen, insbesondere zur Anbindung von Max an eine maschinenlesbare Repräsentation enzyklopädischen Wissens der Wikipedia.

Einleitung

In diesem Vortrag soll ein Blick über den Tellerrand geboten werden. Es geht auch etwas um das Thema „Web“ und was wir in Zukunft noch erwarten können. Damit wird sicher Neuland betreten, und so sollte mein Thema recht gut zu dem Thema der Tagung passen. Ich komme von der Universität Bielefeld. Die Arbeiten, die wir dort zur Kommunikationsforschung unternehmen, zum Beispiel mit Robotern, oder auch mit virtuellen Wesen – um die wird es gleich gehen –, erfolgen in den Kontexten eines „Center of Excellence“, mit dem Titel „Kognitive Interaktionstechnologie“ (kurz CITEC), und eines Sonderforschungsbereichs der Deutschen Forschungsgemeinschaft, in dem es um Kommunikation geht.

Mein Gebiet ist die Künstliche Intelligenz, ein Gebiet, was durch den schillernden Namen oft falsche Vorstellungen oder gar Ängste hervorruft. Ich möchte dazu einfach Folgendes sagen: Als akademische Disziplin befasst sich die Künstliche Intelligenz mit der Konstruktion informationsverarbeitender Systeme, die Wissensleistungen des Menschen modellieren und sie in technischen Anwendungen verwerten. Was sind denn Wissensleistungen des Menschen – kognitive Leistungen? Wahrnehmen gehört dazu, wir können erkennen, was wir sehen und hören, daraus Schlussfolgerungen ziehen und dann auch überlegen, was getan werden kann, also handeln. Durch unsere Forschung wollen wir verstehen, wie der Mensch in dieser Hinsicht funktioniert, und die erlangten Einsichten technisch anwenden können.

Sie haben vielleicht schon einmal den Begriff „Agent“ in diesen Kontexten gehört. Schaut man zu dem Begriff ins Wörterbuch, dann ist ein Agent zunächst einmal eine Person, die autorisiert ist für jemand anderen zu handeln – so wie ein Beauftragter einer Versicherungsagentur. In der technischen Welt wird unter „Agent“ aber auch oft ein Computerprogramm verstanden, das Aktivitäten in seiner digitalen Umwelt beobachten oder initiieren und das mit anderen Agenten kommunizieren kann. Technisch sind das Softwareprozesse, die in einem Computersystem unsichtbar ablaufen, bei denen einzelne Beteiligte miteinander technische Nachrichten austauschen und dabei wie ein Team zusammenarbeiten.

Wenn wir von einem „virtuellen Agenten“ sprechen, so versteht man darunter einen digitalen Agenten mit einer sichtbaren Erscheinung. Unser Max ist ein solcher virtueller Agent. Manchmal wird so etwas auch ein „Avatar“ genannt: eine sichtbare Erscheinung des technischen Systems, die dazu dient, mit dem Menschen besser Kontakt aufnehmen zu können. Hier möchte ich bitte verstanden wissen, dass es nicht darum geht, solche

Maschinen *a/s* Menschen, sondern *wie* Menschen zu bauen. Das ist ein großer Unterschied. Dadurch, dass sie aussehen wie ein Mensch, weiß jeder, wie man mit ihnen in Verbindung tritt, denn wir wissen ja, wie wir mit einer Person sprechen, dazu benötigt man keine Bedienungsanleitung.



Bildunterschrift: Begrüßung mit Max in der virtuellen Welt

In unserem Labor in der Universität Bielefeld können Sie Max begegnen. Er begrüßt Sie dann etwa: „Hallo ich bin Max. Willkommen im Labor für Künstliche Intelligenz.“ Max ist ein virtueller Humanoide. Er besteht aus 68 Einzelteilen, die durch 57 Gelenke miteinander verbunden sind. Eine der Ausgangspunkte für unsere Forschung ist, dass man ein intelligent interaktionsfähiges „Wesen“ schaffen möchte. In dem ersten Filmausschnitt, den ich Ihnen gezeigt habe, konnten Sie sich in unserer Laborumgebung umsehen. Dort setzt man eine Brille auf, durch die man dreidimensional sieht. An der Brille sind kleine Reflektorkugeln angebracht, die vermitteln, wo der Mensch gerade hinschaut. Das kann Max erkennen und dann zum Beispiel an die gleiche Stelle schauen. Max ist also ein Maschinenwesen, dem wir menschliche Kommunikation beibringen. In unserem Labor existiert er nur in einer Projektion – für die, die das kennen, in etwa so wie der holografische Doktor in Star Trek Voyager.

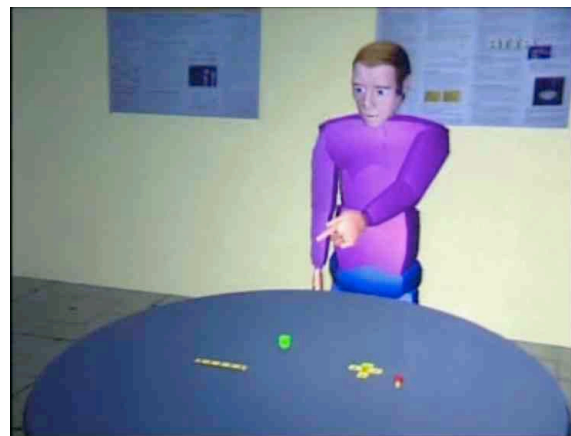
Die digitale Kluft

Ich möchte auf die „digitale Kluft“ zu sprechen kommen. In der Informationsgesellschaft wird die Kluft zwischen kundigen Teilnehmern und denen, die der digitalen Welt nicht oder nicht mehr gewachsen sind, zunehmend größer. Unsere Frage ist, wie geht man damit um? Könnten wir nicht einfach mit einem System in natürlicher Weise reden, ihm unsere Fragen stellen und darauf Antworten erhalten? An solchen Ansätzen forschen wir an der Universität Bielefeld mit einer großen Zahl von Projekten, aus denen ich im Folgenden einiges vorstelle. Also ist meine Frage nun: können wir vielleicht eines Tages mit „künstlichen Intelligenzen“ sprechen, die unsere Ansprechpartner in der digitalen Welt sind?

Vor sieben Jahren, als Max zum ersten Mal der Öffentlichkeit vorgestellt wurde, stand in der Zeitung „Marc und Max sind begeistert – beim Handschlag sprühen die Funken...“ Die Funken sprühten (Marc Latoschik hatte da ein paar Tricks eingebaut), damit man in der virtuellen Realität überhaupt glaubte, dass man Kontakt mit dem künstlichen Wesen hat. Am Beispiel von Max wollen wir die Grundlagen der kommunikativen Intelligenz des Menschen erforschen, und das besonders im Dialog von Angesicht zu Angesicht: wo man sich gegenseitig sieht, wo man die Gestik und Mimik des Anderen wahrnimmt. Also müssen wir

uns erst einmal für die Physis – das Körpersystem – interessieren, speziell für die Gestik und Mimik. Aber auch das Wissenssystem ist von Belang, wenn es um die Sprache und das Verstehen der Bedeutung geht. Wir arbeiten schließlich an einem Bewertungssystem, mit dem unser Max Emotionen erhält, so dass man ihn zum Beispiel beleidigen kann. Das geschieht etwa, wenn man ihn mit politisch unkorrekten Worten anspricht – dann kann er mit beleidigtem Gesichtsausdruck sogar weggehen. Das kann hier nicht alles angesprochen werden.

Entsprechend ausgerüstet kann man über Sprachbefehle gemeinsam mit Max zum Beispiel einen Propeller bauen. In unserer Forschung haben wir uns lange mit der Konstruktion von Aggregaten aus Baukästen beschäftigt, weil man damit schon viele kooperative Situationen erproben kann. Max sagt dann zum Beispiel „Lass uns zusammen einen Propeller bauen.“ Wir können zum Beispiel anfangen und sagen „Stecke diese Schraube in diese Leiste“ und dabei auf die betreffenden Teile zeigen. Max sagt dann „Ok, dann mach ich das.“ Er führt den Schritt aus und sagt dann „Wir haben soeben einen Propeller richtig zusammengebaut“. Wie Sie in meinem zweiten Filmbeispiel sehen konnten, können Mensch wie Maschine Sprache und Gestik simultan verwenden, sich multimodal äußern, wie wir sagen. Auch wenn dieser Bausatz noch kein Wunderwerk der Technik ist, so verdeutlicht er doch den künftigen Einsatz von Max. Max soll Ansprechpartner des Menschen in einer künstlichen Welt sein.



Bildunterschrift: Mensch und Max können Sprache und Gestik verwenden

Zunächst sei noch einmal erklärt, warum wir das überhaupt machen. Zum einen ist es in meiner Disziplin – Künstliche Intelligenz – der Gedanke, dass man ein Verhalten des Menschen im Detail verstehen lernt, wenn es gelingt, es mit einem künstlichen System zu erzeugen. Es geht also um Erklärungsmodelle in Form von Verhaltenssimulationen. Aber auch die Prozesse, die zu dem Verhalten führen, sollen in dem System in vergleichbarer Weise wie beim Menschen ablaufen. Speziell wollen wir die kognitiven und die sozialen Faktoren von Kommunikation besser verstehen. Zu den sozialen Faktoren könnte zum Beispiel Folgendes gehören: Wenn ich jemanden freundlich anschau, dann spürt das der andere und ist vielleicht freundlicher gestimmt gegenüber dem, was ich als nächstes sagen werde.

Auf der anderen Seite erforschen wir mit der künstlichen Kommunikation auch neue Formen der Mensch-Maschine-Schnittstelle. Dort finden „verkörperte“ Systeme heute großes Interesse: Es geht also nicht mehr nur um Textsysteme, die mit der Computertastatur bedient werden, sondern um einen menschlich anmutenden Ansprechpartner, der vielleicht eines Tages in Form eines sprachfähigen Roboters zur Verfügung stehen könnte. Und hier ist der Gedanke, dass sprechende Maschinen die digitale Kluft auf diese Weise überwinden helfen könnten.

Die Technik hinter Max

Nun möchte ich wenigstens einige Dinge darüber sagen, wie der Max funktioniert. Das kann hier längst nicht alles beschrieben werden. Aber im Darwin-Jahr sei doch einmal das Buch „Der Ausdruck der Gemütsbewegungen bei dem Menschen und den Tieren“ erwähnt, in dem vor 137 Jahren Charles Darwin im Detail beschrieben hat, durch welche Aktivitäten von Gesichtsmuskeln es zustande kommt, dass unser Gesicht Ausdruck erhält. Da kommen so spannend klingende Muskeln vor wie der Augenbrauenrunzler, mit dem man finster schauen kann, oder der Augenringmuskel, der unsere Augen zusammenzieht, wenn unser Lächeln „von Herzen“ kommt und so unserem Gegenüber einen entsprechenden Eindruck vermittelt. Oder vielleicht auch der ganz spezielle Muskel „Heber der Oberlippe und des Nasenflügels“, der eingesetzt wird, wenn man die Nase rümpft und sich über etwas mokiert.

So etwas haben wir im Ansatz mit einem Studententeam nachgebaut. Durch die Simulation von Gesichtsmuskelleffekten hat Max ein animierbares Gesicht erhalten. Seine Gesichter kommen also durch Einstellungen der Muskeln zustande; technisch verwenden wir dazu sogenannte Aktionseinheiten. Die Muskeln lassen sich durch die „Stimmungen“ von Max antreiben, die sein Emotionssystem in der Simulation erzeugt – abhängig davon, was ihm in der Kommunikation widerfährt. Die Abbildung zeigt einige Beispiele.



Bildunterschrift: Mit seiner Gesichtsmimik kann Max unterschiedliche Emotionen ausdrücken

Der Körper von Max hat ein bewegliches Skelett mit 103 Freiheitsgraden. Freiheitsgrade betreffen die Art und Weise, wie man seinen Körper bewegen kann, durch Drehen oder Anwinkeln der Gelenke. Der menschliche Körper hat knapp 300 Freiheitsgrade; Max hat also ein gutes Drittel davon. Er ist besonders gelenkig in den Händen und Armen, weil wir speziell die Gestik untersuchen wollen. Weniger Freiheitsgrade hat Max im Rücken, in der Wirbelsäule, da hat er nur drei Wirbel, ist dort also eher zu etwas steifen „roboterhaften“ Bewegungen fähig.

Mit einer synthetischen Stimme kann Max sich sprachlich und gestisch gleichzeitig äußern, indem er zum Beispiel sagt „Und jetzt nimm diese Leiste und mach sie so groß“. Dabei kann er auf ein Bauteil zeigen und die gewünschte Größe mit den Händen andeuten. Ein solcher Verhaltensausschnitt in Form einer sprachlich begleiteten Animation wird mit wenigen Zeilen Computercode erzeugt. Für Interessierte sei gesagt, dass dahinter eine XML-basierte Markup-Sprache für multimodale Äußerungsrepräsentationen steht (und die beachtliche Leistung eines ehemaligen Doktoranden, Stefan Kopp).

Wie sag ich es denn nun der Maschine – wie kann Max die Eingaben des Menschen verstehen? Auch das will ich an einem Beispiel kurz andeuten. Stellen wir uns vor, dass der Mensch sagt: „Steck die gelbe Schraube in die lange Leiste.“ Zunächst einmal muss Max die gesprochenen Wörter erkennen. Im nächsten Schritt muss er den Satz in seiner Bedeutung verstehen, also das „steck“ als Befehl, eine Verbindung herzustellen, das Wort „die“ als bestimmter Artikel, das „gelbe“ als eine Farbe, „Schraube“ als ein Objekttyp usw. Wenn Max dies so repräsentiert hat, muss er schließlich das Gesagte auf das zu Sehende beziehen und die entsprechenden Teile verbinden.

Max im Nixdorf-Museum

Jetzt möchte ich darüber berichten, wie sich Max als Museumsführer macht. Seit Januar 2004 arbeitet er im Heinz Nixdorf MuseumsForum in Paderborn. Das HNF ist nicht nur ein aufbewahrendes Museum, sondern man will dort auch zeigen, wie die Technik der Zukunft aussehen könnte. Dort ist Max zugleich ein Exponat, an dem man sich vorstellen kann, so wird vielleicht unsere Zukunft, und gleichzeitig ein Museumsführer, denn er erklärt dort auch die Ausstellungsstücke in der Ausstellung für künstliche Intelligenz und Robotik. Oder er erklärt den Computerpionier Heinz Nixdorf, oder sich selbst: „Max, das bin ich. Ich bin eine künstliche Person, die sprechen und gestikulieren kann. Ich selber bin künstlich, kann mich aber genauso ausdrücken wie du.“ Im Paderborner Heinz-Nixdorf-Museum hat also die Zukunft der Kommunikation schon begonnen mit dem künstlichen Museumsführer Max.



Bildunterschrift: Max im Nixdorf-Museum

Wie sieht das denn vor Ort aus? Hier ist Max nicht in einer Stereoprojektion, sondern auf einer großen Projektionswand dargestellt. Sie sehen rechts im Bild eine Kamera, die die Leute aufnimmt, die davor stehen. Für Max ist dann die Tatsache, dass ein hautfarbener Fleck in Gesichtgröße gefunden wird, einfach der Hinweis, da steht jetzt ein Mensch. Und er kann auf diese Weise zwischen mehreren Gesichtern hin und her schauen und seinen Blick auf das lebhafteste ausrichten. Aber auch nur so lange, wie es höflich ist: Das haben wir als Regel eingebaut. Man kann Max dort etwa nach dem Wetter fragen und er antwortet dann zum Beispiel: „Moment ich frag mal einen Wettersatelliten... Bis zum Mittag regnet es teilweise recht stark, zum Nachmittag ist es dann heiter bis wolkig. Maximal 4 Grad. In der Nacht kühlt es sich bei wechselnder Bewölkung auf -1 Grad ab.“

Wie haben wir das gemacht? Wir haben Max mit dem Internet verbunden, und wenn nach dem Wetter gefragt wird, dann wird auf der Internetseite der Wetterbericht nachgesehen, und der wird dann von Max vorgelesen. Max versteht natürlich nicht, was er da sagt, aber der, der es hört, versteht das. So ähnlich wie wenn ich jemanden aufsuche, der sich im Internet gut auskennt, und sage: „Such mir da mal etwas raus und lies mir vor, was du gefunden hast.“ Vielleicht versteht der, der es mir raussucht, gar nicht, worum es geht, aber mir sagt das dann etwas.

Max ist dabei auch recht prominent geworden. Er war auch schon mal bei M^{TV}; in dem kurzen Filmclip, den ich Ihnen gezeigt habe, konnten Sie sehen, dass die Stimmung gut war, und die Interviewerin sagte schließlich „Ich will den mit nach Hause nehmen.“

„Max goes Wiki“

Unser nächstes Vorhaben ist „Max goes Wiki“. Wir wollen aktuelle Forschung zu künftigen Anwendungsfeldern hier auch erwähnen, insbesondere zur Anbindung von Max an eine maschinenlesbare Repräsentation enzyklopädischen Wissens der Wikipedia. Mit der Wikipedia ist ein Projekt unterwegs, in dem freiwillige Autoren Artikel schreiben, mit denen potenziell das Wissen über unsere ganze Welt im Internet für jeden verfügbar gemacht wird. Die Pflege dieses Wissens liegt in den Händen der Autoren; das hat natürlich auch Nachteile, denn Fehler können sich einschleichen, die von der Einschätzung der einzelnen Autorin oder des einzelnen Autors herrühren. Der Vorteil der Wikipedia ist aber, dass sie strukturiert ist. Es gibt dort ein umfangreiches Kategoriensystem, in das die Texte einsortiert sind. Es gibt ein Verbindungssystem aus Links, das zeigt, wie zum Beispiel „Pudel“ mit der Kategorie „Hunderasse“ verbunden ist. Es gibt Einträge zu vielen bekannten Personen und Verweise auf weiterführende Seiten, so wie es in einer Enzyklopädie üblich ist.

Zusammen mit dem Computerlinguisten Alexander Mehler betreiben wir an der Universität Bielefeld ein neues Projekt, „KnowCIT“ heißt es, dort geht es um „knowledge enhanced embodied cognitive interaction technologies“. Eine „Wissensverstärkung“ für den verkörperten Agenten Max soll vorgenommen werden. Sie müssen sich vorstellen, dass es ausgesprochen mühsam war, Max von Hand das Viele beizubringen, was er jetzt schon kann. All das Wissen, das er zum Beispiel über den Computerpionier Heinz Nixdorf hat, oder über die Städte Paderborn und Bielefeld oder über den neuen Hauptbahnhof in Berlin oder über die Verwandtschaftsverhältnisse von Horst und Eva Köhler, ist ihm von Hand beigebracht worden. Das ist auf Dauer nicht zu leisten. Was liegt also näher, als die globale Wissensressource, die von vielen Menschen kooperativ erstellt wird, nämlich die im Internet verfügbaren Wissensquellen, zu nutzen?

Ein erstes Vorhaben, das wir angehen, ist das Ermitteln des Gesprächsthemas. Wenn wir mit jemandem über etwas reden und ein Dritter hört zu, dann ist es dem in der Regel ziemlich schnell möglich herauszufinden, worum es überhaupt geht. Dazu reicht schon das Hören einzelner Wörter, wie „Empore“, „Orgel“, „Choral“ usw. Wir merken daran auch schnell, ob uns das interessiert oder nicht. Wir bauen jetzt also einen Topic-Agenten – das ist wieder ein unsichtbares Softwaresystem –, der durch Reizwörter, die in den Texten zu finden sind, herausfiltern soll, worum es eigentlich geht. Damit könnte Max, wenn ein Besucher irgendeine neue Frage stellt, auch bemerken, dass das Thema gewechselt hat. Im ersten geplanten Test wollen wir ihn einfach fragen: „Worüber sprechen wir überhaupt gerade?“ Und er soll antworten: „Wir sprechen über ...“ und nennt dann den jeweiligen Topic, das Gesprächsthema: „Wir sprechen gerade über die neue Bundeskanzlerin.“ Nach dem Wechsel im Bundeskanzleramt war das System von „Schröder“ auf „Merkel“ umstellen, das geschah noch von Hand. So etwas sollte Max in Zukunft vielleicht selbst tun können.

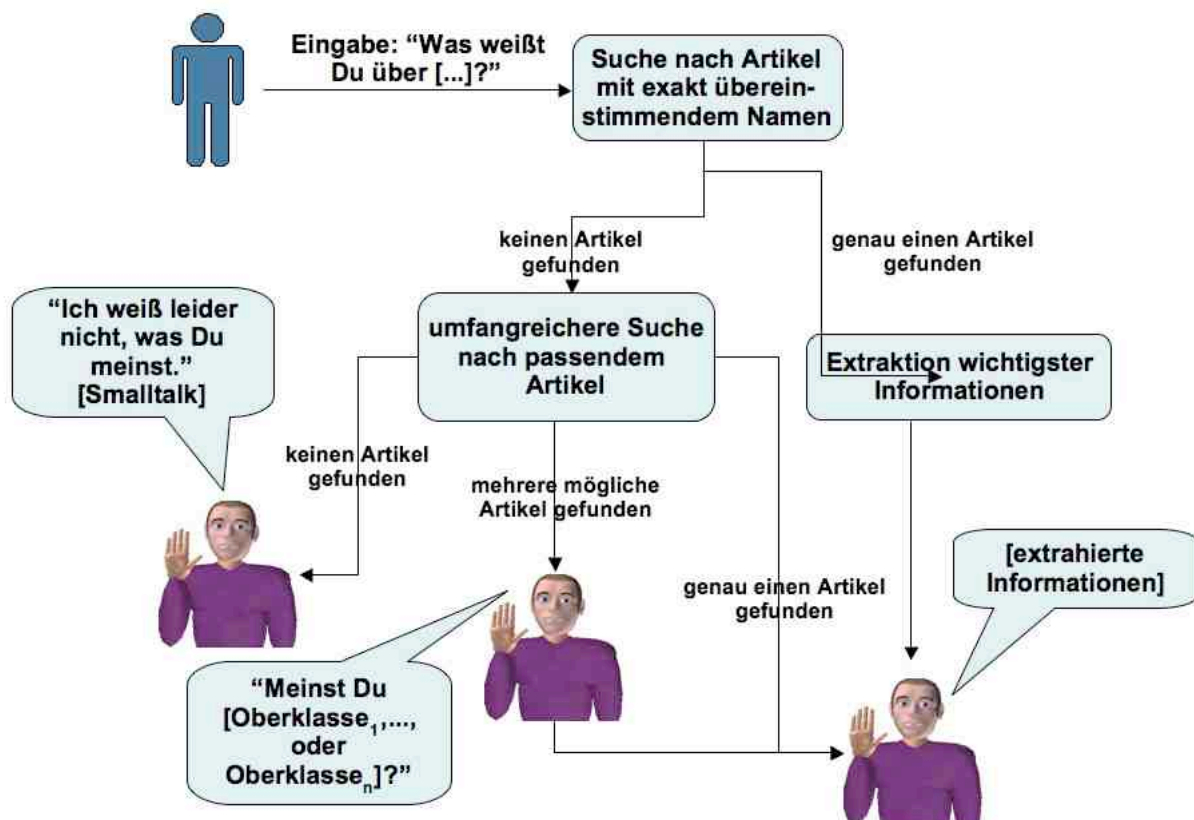
So haben wir schon exploriert, wie das möglich ist. Dazu wurde zunächst ein Artikelindex automatisch aus einem sogenannten SQL-Dump der Wikipedia erstellt. SQL – „Structured Query Language“ – ist eine Datenbanksprache. Der SQL-Dump enthält alle Artikel- und Kategoriennamen. Jeder Eintrag ist strukturiert aufgebaut, es gibt einen Titel mit einem Artikelnamen, es ist ein Typ zugewiesen worden usw. – wer sich schon einmal in der Wikipedia umgesehen hat, dem sagt das etwas. Wir haben zunächst mit der DBpedia gearbeitet, das ist eine datenbankartige Repräsentation der Wikipedia. Allerdings haben wir festgestellt, dass die DBpedia seltener aktualisiert wird als die Wikipedia – zuletzt im Dezember letzten Jahres – und dass sie von der englischen Wikipedia ausgeht: So fehlen dort gerade Fakten aus dem deutschen Bereich, die für unser Vorhaben wichtig sind.

Deshalb wollen wir nun direkt die deutsche Wikipedia verwenden. Nun sind die dort zu findenden Informationsmengen immer noch so riesig, dass ein weiterer Index benötigt wird, mit dem man schnell in einen Artikelindex hineinfinden kann. Dieser Index wird aus einem sogenannten XML-Dump der Wikipedia erstellt. „Dump“ heißt hier: eine Kopie der gesamten

Information, die man aktuell findet. Diese wird im Computer aufgearbeitet – das dauert einige Stunden und geht über Nacht – und daraus wird ein Index erstellt, der alle Artikelnamen und Texte enthält. Das ergibt fast eine Million Einträge, und zu jedem Eintrag gehören weitere Informationen in großer Zahl wie: Bezeichnungen der eingehenden Links; die Namen von Begriffsklärungsseiten, die auf den Artikel verweisen; die Namen der Listen, in denen der Artikel enthalten ist; die Namen der Kategorien, denen der Artikel zugeordnet ist, und so fort. Das sind nun die Treffermöglichkeiten für Max, um relevante Informationen aufzufinden.

Unsere Vision ist, dass wir Max eines Tages einfach irgendetwas fragen können und er erteilt uns immer eine Auskunft. Das geht zwar jetzt auch schon, aber Max kann bislang nur mit dem antworten, was seinem Wissenssystem von Hand eingegeben wurde. Wenn er nichts zu einer Frage weiß, ist er dennoch nie um ein Wort verlegen, denn wir haben ihn auf raffinierte Weise mit „Smalltalk“-Regeln ausgestattet. Im Notfall würde er antworten: „Ich weiß leider nicht, was du meinst.“ Oder er würde sagen: „Ich hoffe, du meinst diese Frage nicht ernst“ oder „Das willst du doch nicht wirklich wissen.“ Dadurch ist Max ein System, das im Prinzip „24 hours a day, seven days a week“, also Tag und Nacht laufen kann und immer auch eine Antwort geben kann, selbst wenn es eine ausweichende Antwort ist. Aber wer von uns macht das nicht auch schon einmal?

Unser Ziel ist es aber, dass Max in Zukunft von dem aus der Wikipedia erhältlichen Wissen Gebrauch machen kann. Fragt man ihn dann „Was weißt Du über [...]?“, sucht er zunächst nach einem Artikel mit exakt übereinstimmendem Namen. Wird ein solcher gefunden, dann soll Max die Informationen daraus verbalisieren. Allein dies ist allerdings schon eine nicht ganz einfache Aufgabe. Wenn es ein längerer Artikel ist, könnte es für den Menschen ermüdend sein, den im Ganzen vorgelesen zu bekommen; es müssten gegebenenfalls die wichtigsten Informationen extrahiert werden. Nun kann es aber sein, dass kein passender Artikel gefunden wird, dann würde innerhalb des Fragetopics eine umfangreichere Suche nach einem passenden Artikel angestoßen, wie es in der Abbildung veranschaulicht ist.



Bildunterschrift: Artikelsuche in der Wikipedia

Wird überhaupt kein Artikel gefunden, dann gibt Max eine seiner „Smalltalk“-Antworten – die haben sich oft bewährt. Das war zum Beispiel der Fall, als im Juli 2005 der damalige Bundeskanzler Schröder das Nixdorf-Museum besuchte und Max ein Gedicht von Goethe aufsagen sollte. Bei der Eingabe prellte die Tastatur und es erschien „Goethe“ mit zwei „e“. Dazu wusste Max dann gar nichts und antwortete: „Jetzt hast du mich erwischt.“ Darauf musste der Kanzler herzlich lachen und die Stimmung war spontan fröhlich. Dann allerdings sagte Max: „Ich kann ein Gedicht von Durs Grünbein aufsagen“ und darauf der Kanzler: „Das beeindruckt mich jetzt aber. Das ist nämlich ein ganz Großer, der ist gerade in Amerika.“ „Ja,“ sagte der Museumsdirektor „und letzten Sonntag war er hier im Museum“ – da war nämlich bei der Kulturveranstaltung „Wege durch das Land“ der Dichter dabei, und das war auch der Grund, warum wir Max noch schnell ein Gedicht von Durs Grünbein beigebracht hatten.

Ich möchte hier keinen falschen Eindruck vermitteln: es bleibt noch viel zu tun, wenn man das alles so hinbekommen möchte. Aber wäre es nicht eine nette Sache, wenn man sich über alles, was einen gerade interessiert, an einen Ansprechpartner wie Max wenden könnte und zum Beispiel fragen, was sind eigentlich Triptychons? Und bekäme dazu dann eine Auskunft, die Max aus dem Internet besorgt. Ich könnte mir das gut vorstellen. Selbst ein technisch versierter Mensch hat nicht immer die Zeit, sich auf die Computer-Recherche einzulassen, und nicht jeder kann oder möchte das. Es ist doch so viel einfacher, jemand anderen zu fragen.

Ansprechpartner in der digitalen Welt

Max ist eine Maschine, mit der man sprechen kann. Max ist in gewissem Sinne autonom: arbeitet nach eigenen Regeln und ist nicht etwa auf jede Situation vorprogrammiert. Er hat gewisse soziale Fähigkeiten: kann mit dem Menschen sprachlich und auch nicht-verbal – mit Gestik und Mimik – interagieren. Er nimmt seine Umgebung wahr und kann darauf reagieren: zum Beispiel einen neu in sein Gesichtsfeld tretenden Besucher mit „Hallo“ begrüßen und ihm zuwinken. Er ist aber auch proaktiv, kann von sich aus die Initiative ergreifen und den Menschen ansprechen, zum Beispiel: „Wollen wir Tiere raten spielen?“

Wenn man so will, könnten wir von einer Evolution maschineller Kommunikation sprechen. Früher sah man Maschinen als Werkzeuge, die man benutzt. Das wird sicher für bestimmte Typen von Maschinen auch immer so bleiben. Man baut sie ja gerade, damit man einen Nutzen davon hat und sie benutzt. Dann kam der Gedanke auf, Maschinen könnten auch intelligente Werkzeuge sein, die man über ein Mikrofon instruiert; so etwas gibt es teilweise schon in Textverarbeitungssystemen: „Öffne File – Name“ usw. Eine nächste Stufe wäre es, Maschinen zu bauen, die dem Menschen verkörpert gegenüber stehen, mit denen man eine unterhaltsame Konversation führen kann, und schließlich sogar Maschinen zu konzipieren, die mit dem Menschen gemeinsam komplexe Aufgaben lösen, also kooperieren.

Sprechende Maschinen könnten in Zukunft somit Partner des Menschen werden, statt nur Werkzeuge sein. Eingebettet in die digitale Welt, in der schon heute eine unglaubliche Fülle von Informationen verfügbar ist, könnte „ein Max“ Auskunft geben über den Wetterbericht, über Busabfahrtszeiten und vieles mehr – grundsätzlich über alles, was man im Internet finden kann. Damit komme ich zum Schluss: Sie können sich sicher vorstellen, dass ein solches Vorhaben nur möglich ist mit einem Team, dem für engagierte Mitarbeit zu danken ist; und dass es auch Spaß macht, an dieser doch sehr anstrengenden und fordernden Forschung zu arbeiten, möchte ich ebenfalls vermittelt haben.

Anmerkung:

Dieser Beitrag wurde nach einem Mitschnitt des Schlussvortrages auf der Frühjahrstagung der Mediendokumentare und -archivare in Frankfurt, Mai 2009, rekonstruiert. Eigenheiten des mündlichen Vortrags blieben zum Teil erhalten. Der Deutschen Forschungsgemeinschaft (DFG) und dem Heinz Nixdorf MuseumsForum (HNF) sei für langjährige Förderung gedankt.

Professor Ipke Wachsmuth ist Mathematiker und Informatiker. Er ist seit 1989 Professor für Künstliche Intelligenz an der Universität Bielefeld, war dort von 2002 bis 2009 Direktor des Zentrums für interdisziplinäre Forschung (ZiF) und ist seit 2009 Sprecher des DFG-Sonderforschungsbereichs „Alignment in Communication“. Forschungsschwerpunkte: Sprachverstehen, Agentensysteme, Virtuelle Realität, Gestik, multimodale Interaktion, Emotion.

Neuere Veröffentlichungen u.a.: Der Körper spricht mit, Gehirn & Geist, 2006; Embodied communication in humans and machines (Hg. zus. mit M. Lenzen u. G. Knoblich), Oxford, 2008; Modeling communication with robots and virtual humans (Hg. zus. mit G. Knoblich), Springer, 2008; Affective computing with primary and secondary emotions in a virtual human (zus. mit C. Becker-Asano), Journal of Autonomous Agents and Multi-Agent Systems, 2009.