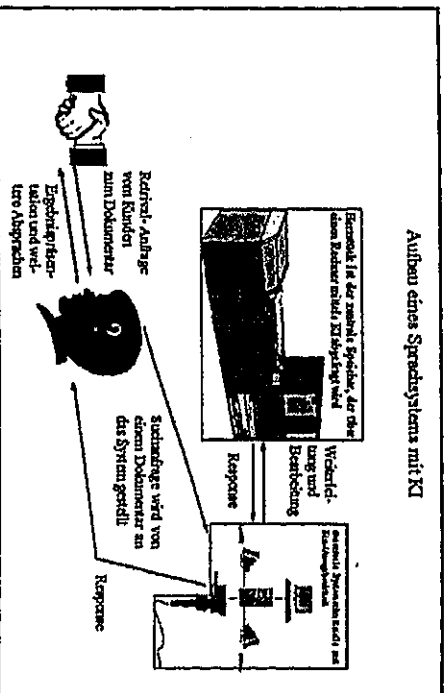
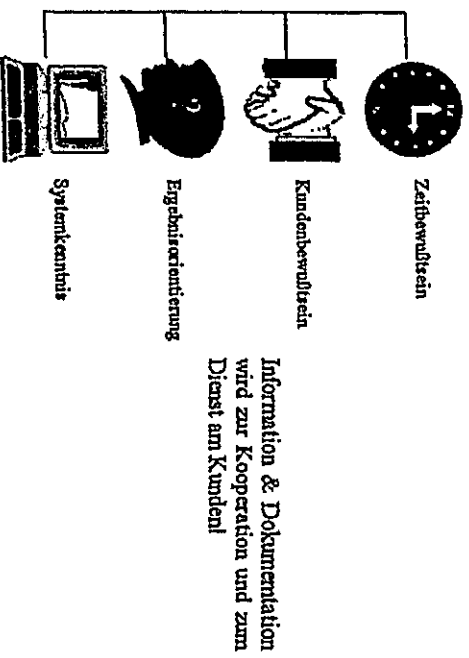


Aufbau eines Sprachsystems mit KI



### Veränderte Anforderungen an Dokumentarinnen und Dokumentare



## Speak & work statt click & do! – (Sprach) Steuerung in (Sprach) Archiven

Dirk Tunger

Fachhochschule Hamburg, Fachbereich Bibliothek und Information

Grindelhof 30 • D-20146 Hamburg

dirk\_tunger@yahoo.de

Special III: Information und Ausbildung I – 20. Mai 1999

**Abstract:** Vier Elemente sind wesentlich: Die Frage nach dem Stand der Technik in der Sprachsteuerung, die Zukunftsentwicklungen, das Zusammenwirken von Sprache und Dokumentation sowie die Frage nach neuen Anforderungen für Bedienstete im Bereich Dokumentation/Archiv. Diese Elemente zwingen gewisse Fragen regelrecht auf, beispielsweise nach den Möglichkeiten verbaler Retrievals in der Zukunft. In welchem Umfang kann in Archiven auf Sprachdateien in digitalisierter Spracheingabe zugegriffen werden? Als ebenfalls interessant kann die Frage betrachtet werden, ob die Zukunft der Spracharchivierung in der Schlagwortindexierung oder der Volltextsuche zu finden sein wird.

Noch nicht alzulange liegt die Zeit zurück, als die meisten Informationsbestände über Zettelkästen organisiert waren. Per Karte hatte man die Möglichkeit, mit der nötigen Portion Glück, die gewünschten Daten zu erhalten. Diese Methode der Archivsteuerung fiel aber durch den überproportionalen Arbeitsaufwand, den ein derartiges System für sich in Anspruch nimmt, negativ auf. Die Folge dessen: Es wurde ausrangiert und an den meisten Stellen durch ein Computersystem ersetzt, als die technischen Voraussetzungen hierfür erfüllt waren.

Die Vorteile dessen sind deutlich zu erkennen: es wird deutlich weniger Zeit zur Pflege des Systems und seiner Komponenten benötigt, das Retrievalergebnis liegt für den Benutzer wesentlich schneller vor, es können optional wechselnde Statistiken eines Mediums abgerufen werden; beispielsweise, ob eine Verfügbarkeit derzeit besteht.

Doch dies ist noch nicht alles: Mit zunehmender Größe werdenden Speichern und schnelleren Systemen wird die Kunst der Stunde genutzt: Etliche Archive legen nicht nur verschlagwortete Faksimiles im Rechner ab, sondern nutzen die Möglichkeit zur Volltextrecherche. An dieser Stelle ist dann auch der Schnittpunkt von neuen Problemen, aber auch von weitaus größeren Chancen zu sehen.

Zum einen ist der Benutzer zur Zeit noch in die Pflicht genommen, alle Suchanfragen so zu formulieren, wie sie im Datenbestand vorkommen. Abweichungen führen zu fehlerhaften Ergebnissen. Diese drücken sich dann meist in Extremen aus: entweder werden zu viele Ergebnisse freigesetzt, was zu einem Chaos führen kann oder es werden zu wenig Ergebnisse herausgegeben, was ein Desaster bedeuten kann. Hieran ändert auch eine Sprachengabe nichts.

Küpper 1994, S. 40: »Bei reinen Abfragesystemen bleibt dem Benutzer durch die nativsprachliche Schnittstelle zwar das Erlernen einer formalen Sprache erspart und er kann die vielfältigen Möglichkeiten der natürlichen Sprache nutzen. Er muss aber genau wissen, wonach er fragen kann, d. h. er muss die angeschlossene Datenbank, ihre Struktur und die enthaltene Information gut kennen, um eine zufriedenstellende Antwort zu erhalten. Daher scheint uns ein reines Abfragesystem [...] nicht ausreichend zu sein.«

Der Knackpunkt liegt demnach in der Auswertung der Eingabe: Hier müssen wir erkennen, dass es nicht nur das Ziel sein darf, eine Anfrage formal korrekt zu erfassen, sondern dass ebenfalls ihre korrekte inhaltliche Erschließung Ziel eines Gesamtsystems sein muss.

Um jetzt aber nicht zu versuchen, zwei Stufen auf einmal zu erklären, zunächst die Formalfassung mittels Sprache: Das Problem hier besteht darin, Sprache, eigentlich ein analoges Signal, zu digitalisieren und dann über Referenzen die richtige Auswahl zu treffen: d. h. das Ursprungswort zu erkennen oder zu ermitteln. Was so einfach klingt wird in der Praxis aber gleich durch mehrere Ursachen erschwert

- ein Mensch spricht ein Wort niemals zwei- oder mehrfach identisch aus;
- unterschiedliche Benutzer eines Systems sprechen ein Wort zum Teil mit gravierenden Unterschieden in Tonhöhe, Sprechgeschwindigkeit und Betonung aus;
- es gibt Wörter, die in der Aussprache sehr ähnlich klingen;
- Benutzer eines Systems wechseln sehr oft, was unterschiedlichste Dialekte zur Folge hat.

Es ist führenden Softwarefirmen aber trotz dieser oben genannten Widrigkeiten gelungen, im vergangenen Jahr gleich mehrere Produkte auf dem Markt zu platzieren, die es zum einen ermöglichen, Sprache zu digitalisieren und in ein gängiges Textformat umzuwandeln. Zum zweiten ist es u. a. mit dem verbmbil Projekt vom bmbf möglich, der Rechner bestimmen zu lassen, wie der spezielle Aufbau des digitalisierten Textes gestaltet. Mit dieser Form der didaktischen Analyse kann man Auskunft über Sinn und Zweck jedes einzelnen Wortes in einem Satzgefüge erhalten. An dieser Stelle muss erwähnt werden, dass bei Fehlern in der Erkennung, die zwangsläufig nie gänzlich auszuschließen sind, natürlich auch Fehler in der Analyse zur Folge haben. Da aber die Erkennungsrate bei 98% liegt, vernachlässigen wir die restlichen Prozente und fassen unsere bisherigen Erkenntnisse noch einmal zusammen:

**Unser Ziel:** Einen neuen Steuerungsmechanismus für Archive zu diskutieren, der Suchanfragen in digitalisierter Spracheingabe entgegennimmt, verarbeitet und das bestmögliche, präziseste, Ergebnis auf die Fragestellung liefert.

**Unser Standpunkt:** Wir nehmen Software zu Hilfe, die zu einem sehr hohen Prozentsatz fehlerfrei Sprache A/D-wandelt und das Ergebnis analysieren kann, um Sinn und Zweck zu erkennen.

**Unser Weg:** Um das Ziel von unserem derzeitigen Standpunkt zu erreichen, müssen wir unser Augenmerk verstärkt auf die Entwicklungen im Bereich der künstlichen Intelligenz (KI) richten, denn ihre Entwicklungen sind die Basis für weitere Bestandteile unseres Systems.

Geht man also eine Stufe weiter: Denken wir nicht mehr nur an Computersysteme, die auf der Nutzung von heutigen Betriebssystemen beruhen, sondern an weitaus komplexere Einheiten, mit denen man mittels Sprache kommunizieren kann.

Was fehlt demnach noch? Herzstück unserer Entwicklung, quasi lebenswichtig für den Dokumentar, ist, wie oben bereits erwähnt, die korrekte inhaltliche Erfassung der Anfrage. Hier muss KI zum Zuge kommen, um...

- Fehlerquellen frühest möglich zu erkennen,
- erkannte Fehlerquellen, also Zweideutigkeiten, zu entfernen,
- die richtige Entscheidung zu treffen, was Fragen der Relevanz angeht.

Zwei Beispiele, die unabhängig von einander, demonstrieren sollen, bei welchen Typen von Eingaben eine inhaltliche Erschließung unumgänglich ist

1. »Jedes A ist ein B.« »Jedes B ist beinahe ein C.« »Jedes A ist also beinahe ein C«, lautet die korrekte Schlussfolgerung. Setzen wir für die Buchstaben aber Begriffe ein, würde sich grammatikalisch nichts verändern, inhaltlich aber ein völlig anderer Sinn entstehen, wenn A ein Physiker sei, B ein Nobelpreisträger und C für »alk. stunde. Hier wäre die Schlussfolgerung einzuschänken und zu verifizieren.

2. Eine Suchanfrage zum Thema »Golf« bringt 60 Treffer: 20 mit dem Inhalt Golf als Auto, 30 zum Thema Golf und Sport und 10 zum Thema geographischer Golf. An dieser Stelle wäre also eine Spezialisierung zu erfragen.

Warum gehen wir im zweiten Beispiel auf das Synonym-Problem ein? Warum befassen wir uns mit logischen »Spielereien«? Die Antwort auf diese Frage mag simpel und unbedeutend klingen, ist es aber bei weitem nicht, weil ein System, das einen immensen Datenbestand verwaltet, dessen Ergebnisvielfalt sinnvoll genutzt werden kann, die aber auch zeitraubend sein kann, so präzise arbeiten muss, dass eine große Verlässlichkeit auszurechnen ist, was die nahezu unbedingte Richtigkeit vom Ergebnis angeht. Wäre dies nicht der Fall, wäre es nicht für das System möglich, in irgendeiner Form Sympathien zu gewinnen. Stellen wir uns folgende Szene einmal vor: Der erste Umstellungsprozess ist immer noch dabei, vollzogen zu werden, ein zweiter könnte mit der Umstellung zu sprachgesteuerten Systemen in der Zukunft folgen. Dies könnte auf der einen Seite ein weiterer Gewinn sein, die Produktivität in der Dokumentation zu steigern. Es könnte aber auch ein Flop werden, wenn das System nicht angenommen wird, weil es fehlerhafter ist als bisherige. c't im Gespräch mit Bill Gates, (c't 4/99, S. 16) der auf die Frage, was MS tue, um mehr Sicherheit und Einfachheit in die Betriebspraxis zu bringen, angibt: »Wir haben damit begonnen, die Fehlerbedingungen von MS-Software zu katalogisieren und uns gefragt: Kann der Computer nicht die Korrektur übernehmen? Welche Art Nachricht oder Diagnose muss laufen, wenn sie einen Fehler meldet? Wie können wir per Software einen menschlichen Assistenten simulieren?«

Wo ist der Zusammenhang zu unseren vorhergegangenen Aspekten? Im Grunde liegt er in der Formulierung WYSMUBEWYG (*What you say must be what you get*): Erst wenn Fehlererkennung und Behebung optimiert sind, erst wenn die Kapazitäten von verfügbarem Speicher und verfügbarer Geschwindigkeit erhöht sind, erst wenn DfÜ drastisch schneller von staten geht als bisher, erst wenn die KI-Forschung mit leistungsstarken, komplexen Lösungen aufwarten kann, erst dann kann an eine Einführung von sprachgesteuerten Gesamtsystemen in der Dokumentations- und Informationsbranche gedacht werden. Dies alles muss aber bei der derzeitigen Geschwindigkeit der technischen Entwicklung nicht allzu lange dauern.

Patrick Bock (IW 1/99, S. 54): »Ohne künstliche Intelligenz – und da ist den Herstellern der Spracherkennungssoftware kein Vorwurf zu machen – fehlt der Erkennung ein

Kernfaktor, eben der, den jeweiligen Kontext zu erkennen.« Dieses Problem wurde weiter oben schon angesprochen und verliert auch nichts an Bedeutung, wenn wir uns jetzt Spracharchiven zuwenden. Wenn in der Zukunft der Durchbruch in der KI-Technologie gemacht ist, viele Datenbestände in digitaler Form in Volltext vorliegen, wird es auch nicht mehr lange dauern, bis Audio-Datenbestände, vor allem Sprache, mittels Sprache recherchiert werden kann. Wie kann man sich dies für die Praxis vorstellen? Durch die Vernetzung im Datenverkehr würde bundesweit ein zentrales Archiv, in dem beispielsweise Interviews recherchiert werden können, ausreichen, alle Anfrager zufriedenzustellen. Es könnte nach dem Vorkommen bestimmter Wörter, Verknüpfungen von Wörtern, nach Passagen, nach Schlagwörtern oder Themen gesucht werden. Die Anfragen werden in humanspeech durchs Netz übertragen, von einem Dokumentar entgegengenommen und dem System übergeben. Durch ein Herz, das von KI gesteuert wird, würden nur die wirklich relevanten und im direkten Zusammenhang mit der Anfrage stehenden Dokumente ermittelt und zurückübertragen.

Zukunftsmusik? Im Moment sicherlich noch, aber wenn die Geschwindigkeit der technischen Revolution sich weiter erhöht, könnten die Visionen Wirklichkeit werden und realisierbar sein.

Was würde sich dann verändern, wenn KI fester Bestandteil unseres Lebens wäre? Welche Anforderungen an zukünftige Dokumentare, Informationsbeschaffer und IT-Agents würden sich zwangsläufig ergeben? Auf diese Fragen wollen wir im letzten Abschnitt eingehen, beginnend mit einem Zitat von Reinhard Golecki (Golecki 1998, S. 90f): »Erst mit der Entdeckung, dass sich Information nicht auf Materie reduzieren lässt, verschiebt sich das Thema vom lebenden Kunstwesen zur denkenden Maschine. Nun erst kann der alte Traum sich erfüllen: [Durch die] postulierte Unabhängigkeit der Intelligenz von ihrer materiellen Realisierung [besteht] erstmals die Chance, etwas ganz Eigenes herzustellen, wirklich Schöpfer zu sein.«

Lesen wir aus diesem Zitat etwa etwas Düsteres, Verdunkelndes heraus? Neid, Missgunst oder Panikmache? Oder das genaue Gegenteil: Aufbruchstimmung?

Sicherlich, wenn wir auf unser Problem schauen und uns vorstellen, eine Maschine verarbeitet eine Suchanfrage und selektiert zu einem gewissen Teil für uns das Ergebnis, mag dies befremdlich klingen. Wir müssen daher zwei Thesen zugestimmt haben, um mit dem Gedanken umgehen zu können:

- Daten sind eine Ware & Geschäfte mit großen Transfervolumen sind möglich;
- Die IT-Branche kann für Wirtschaftsstandorte von entscheidender Bedeutung sein, ähnlich befügend wie einst die Dampfmaschine von James Watt.

Uns wird deutlich, dass Zeit ein immer entscheidenderer Faktor sein wird, dass es sowohl auf Präzision, als auch Geschwindigkeit ankommt, mit der ein Ergebnis vorliegt. Die Tiefgründigkeit, wie sie teilweise noch immer verbreitet ist, ist dabei nicht mehr von entscheidender Bedeutung. Informiert sein bezieht sich in unserer Zeit nämlich immer weniger auf ein grundlegendes, fundiertes, allumfassendes Wissen als vielmehr auf ein *Wissen light, just in time*.

Was schließen wir daraus? Es wird sich nicht umgehen lassen, ein neues Anforderungsprofil an Bedenker im IuD-Bereich zu entwickeln, das die oben genannten Faktoren beinhaltet und berücksichtigt: Wenn wir von der Annahme ausgehen, dass die Zeit eine wichtige Rolle einnimmt, müssen wir einsehen, dass der Gedanke des Servicedienst-

leisters weiterverbreitet werden muss. Coy 1994: »Auf der EG-Ebene wird zusehends erkannt, dass wir für die kreative Nutzung neuer Technologien eine kulturelle und industrielle Renaissance brauchen.« Diese Gedanken sind quasi die Grundlage, die ein professionelles Arbeiten erst ermöglichen. Die Aufgabe des Dokumentars wäre dann, das Ergebnis quersuchen und in Absprache mit dem Auftraggeber zu entscheiden, ob es ausreichend ist. Sollte es nicht den Erwartungen entsprechen, besteht nun die Pflicht des Dokumentars, nachzuhaken, was unkorrekt ist und wie dies beseitigt werden kann.

Mag diese Tätigkeit auch als nicht intellektuell anspruchsvoll angesehen werden und die Frage auftauchen, warum ein Dokumentar mit dieser Tätigkeit ein Studium benötigt, so sei darauf geantwortet: Die Kunst besteht darin, zum richtigen Zeitpunkt das richtige Fragenschema zu entwickeln, um die Tätigkeit des Systems zu verringern. Von vornherein sollen Fehlerquellen ausgeschlossen werden: Um diesen Anforderungen gerecht zu werden, genauso wie der Aufgabe der Nachrecherche, wenn z. B. Widersprüche oder Halbheiten auftauchen, ist eine gewisse Qualifikation erforderlich, die am effektivsten in einem Studium vermittelt werden kann.

An weiterer Qualifikation sei noch erwähnt, dass die perfekte Beherrschung des Systems nicht vergessen werden darf: Es mag sich einfach anhören, man mag lediglich an das Übertragen von Anfragen in eine Eingabemaske mittels Sprache denken, doch diese Annahme ist nicht zu halten: Es kommen Erweiterungen dazu, das korrekte Geben von Befehlen zur Selektion. An dieser Stelle darf auch nicht vergessen werden, dass im Falle eines Versagens des Systems das Können vorhanden sein muss, schnellstmöglich eigene Recherchen anzustellen.

Fazit: Spracherkennungssysteme könnten in Zukunft in Zusammenwirkung mit KI in der Lage sein, Suchanfragen auszuführen und zu gewichten. Hierbei kommt es zu einer Zeitersparnis, die anderen Projekten oder Tätigkeiten gewidmet werden kann. Entscheidend für das Überleben einer Dokumentation wird sein, in welcher Zeit ein konkretes, aussagefähiges und der Anfrage entsprechendes Ergebnis vorliegt.

## Bücher

Golecki, Reinhard: *Können Computer denken?* Hamburg: Vlg. Dr. Kovac, 1998

Oranek, Günther; Cox, Wolfgang: *Die maschinelle Kunst des Denkens*. Braunschweig; Wiesbaden: Vieweg, 1994

## Zeitschriftenbeiträge

Bock, Patrik: *Speechbrowsing*. In: *Internet World* (1999), Nr. 1, S. 52-54

Kurl, Jürgen; Meyer, Egbert; Siegan, Peter: *Ich bin auf unsere Rolle stolz*.: Bill Gates über Windows, Wettbewerb und den Anti-Trust Prozess. In: *c't* (1999), Nr. 4, S. 16-20

Köpper, Delfé; Rössner, Dietmar; Striegl-Scherer, Andrea: *Natürlichsprachliche Zugangssysteme...* In: *Sprache und Datenverarbeitung* (1992), Nr. 1: S. 29-66

## Links

[http://www.dfki.uni-sb.de/twrbimobil/Vm\\_Info\\_Phase2.html](http://www.dfki.uni-sb.de/twrbimobil/Vm_Info_Phase2.html)

<http://www.philips-sp.de>

<http://www.naturalspeech.de>

<http://www.coling.uni-freiburg.de/workshop97/WNS-06.html>