Künstliche Intelligenz

Robert Fugmann*

Über die Möglichkeiten und Grenzen der programmierten Informationsbereitstellung

Gedanken eines Informationsveteranen

DOI 10.1515/iwp-2016-0028

Zusammenfassung: In der Vergangenheit haben Fehleinschätzungen über die Programmierbarkeit intellektueller Arbeiten zum Abbau von Personal in Informationsabteilungen geführt. Es existiert jedoch eine Grenze, welche die determinierten und damit programmierbaren Prozesse von den nicht determinierten und damit nicht zufriedenstellend programmierbaren Prozessen trennt. Der Autor ruft dazu auf, nicht alle klassischen Arbeiten der Auswertung und Bereitstellung von Informationen in die altertümliche Ecke zu verbannen.

Deskriptoren: Künstliche Intelligenz, Software, Auswertung, Verarbeitung Natürlicher Sprache, Wissensrepräsentation, Rationalisierung, Innerbetriebliche Information

Possibilities and limitations of computerized information procurement. An information veteran looks back.

Abstract: In recent decades the possibilities of computerization of intellectual work has been overestimated. As a result information units in companies were closed. But there is a huge difference between determined processes that can be coded and undetermined processes, which cannot be programmed in a satisfying way. The author claims, not to stigmatize any classical documentary and information provision work as antiquated.

Descriptors: Artificial intelligence, Software, Analysis, Natural language processing, Knowledge representation, Rationalization, Inhouse information

Les possibilités et les limites de la mise à disposition informatisée de l'information. La rétrospective d'un vétéran de l'information

Résumé: Dans le passé, les possibilités d'informatisation du travail intellectuel ont été surévaluées et cette erreur a

conduit à des réductions de personnel dans des services d'information. Il existe cependant une frontière qui sépare les processus déterminés, et par conséquent programmables, des processus non déterminés, qui ne peuvent donc pas être informatisés de façon satisfaisante. L'auteur lance un appel afin que tous les travaux classiques du traitement et de la mise à disposition de l'information ne soient pas bannis dans le coin des antiquités.

Describeurs: Intelligence artificielle, logiciel, évaluation, traitement des langues naturelles, représentation des connaissances, rationalisation, information interne

1 Einführung

Vor vierzig Jahren begann eine neuartige Infektionskrankheit sich auszubreiten. Sie äußerte sich durch erhöhte Temperatur, körperliche Schwäche, Schmerzen im Kopf und in den Gelenken, Übelkeit und Durchfall. Schon gegen leichte Infektionen konnten die Infizierten keinen Widerstand leisten und verstarben bald. Unter einer Vielzahl von Beschreibungen der unterschiedlichsten Art erschienen Berichte über dieses Syndrom. Um die Erforschung dieser Krankheit in Gang zu bringen und um über sie kommunizieren zu können, gab man ihr den Namen „AIDS“ (Acquired Immune Deficiency Syndrome). Fortan erschienen (fast) alle weiteren Publikationen unter dieser *lexikalischen* Bezeichnung, wodurch sie nun leicht auffindbar wurden.

Vor diesem Zeitpunkt hatten alle diesbezüglichen Publikationen mit definitionsartigen, *nichtlexikalischen*, *paraphrasierenden* Umschreibungen der vorgenannten Art erscheinen müssen, von denen es in der freien Textformulierung für diesen Begriff in einem Volltextspeicher *eine unbegrenzt große Zahl* gibt. Wollte man in einem Volltext Speicher der medizinischen Literatur beim Suchen nach einschlägigen Dokumenten vollständig fündig werden, dann hätte man sich eine fast unendlich große Zahl der unterschiedlichsten Formulierungen überlegen müssen

*Kontaktperson: Dr. rer. nat. Robert Fugmann, Auf der Au 4, 65510 Idstein, E-Mail: r.fugmann@amweg.de

und zur Suchbedingung machen müssen, ein Ding der Unmöglichkeit.

Mit so großen Lücken in der Zugänglichkeit der Fachliteratur wollte man sich bei diesem Syndrom nicht abfinden. Deshalb bildete man in den USA eine Arbeitsgruppe von Fachleuten, die sich vornahm, zunächst erst einmal die medizinische Fachliteratur der zurückliegenden zehn Jahre durcharbeiten. Jedem einschlägigen Dokument, das hierbei gefunden wurde, wurde nachträglich das nun auch zum Suchen geeignete Schlagwort¹ „AIDS“ **zugeteilt**, ein zeitraubendes und kostspieliges Vorgehen. Aber so wurde die Zugänglichkeit auch der älteren Literatur zu diesem Thema rückwirkend gesichert, und nun auch für die Zukunft unter der Voraussetzung, dass nun *in allen* einschlägigen Dokumenten fortan der Fachausdruck „AIDS“ verwendet wurde.

Was hier praktiziert wurde, ist eine Variante vom traditionellen Indexieren, so alt wie das gesamte Bibliothekswesen, denn der Bedarf nach gesicherter und schneller Zugänglichkeit von publiziertem Wissen besteht seit Jahrhunderten.

In einer ähnlich misslichen Lage befand man sich bei einem literarischen Großprojekt zu den soziologischen Ursachen und Folgen des Alkoholismus. Das Wort „Alkohol“ tritt in unterschiedlichen Bedeutungen und Zusammenhängen auf, als Methylalkohol, Propylalkohol, Alkohol als Brennstoff, zur Thermometerfüllung, zu Desinfektionszwecken, zur Reinigung, zur Verdünnung und Entfernung von Lacken, usw. Ob es (aus der Perspektive dieses Projektes) auch zur inhaltlichen Essenz eines Dokuments gezählt werden kann, kann nur durch sachverständiges und sorgfältiges Lesen des Dokuments festgestellt werden, und entsprechend wurde in diesem Projekt dann auch verfahren.

¹ Zu den Schlagwörtern werden hier auch die Notationen gezählt, weil auch sie einem Dokument von einem Fachmann zugeteilt werden. Die hierarchische Struktur, welche einem solchen System von Notationen innewohnt, dient dem Zweck, einen guten Überblick über das gesamte Vokabular zu gewährleisten. Diesen Überblick braucht der Fragesteller, ebenso wie auch der Indexierer, damit beide die jeweils bestpassenden Schlagwörter finden. Nur unter dieser Bedingung ist der genügend zuverlässige Gebrauch eines großen Vokabulars gewährleistet.

2 Die Illusion der unbegrenzten Programmierbarkeit von menschlicher Aktivität

Die Vertreter der sogenannten „harten“ Künstlichen Intelligenz stehen auf dem Standpunkt: „Was der Mensch kann, kann auch der Computer, und zwar schneller und besser“ und insbesondere: „Der Mensch macht auch nichts anderes, als einen Text zu lesen. Das kann auch der Computer“.

Der „Computer der Zukunft“ wird angeblich den Bedarf der Menschheit selbst ermitteln und sich selbst konstruieren. Das blinde Vertrauen in „den Computer“ ging schon so weit, dass es zum computergesteuerten Abschuss eines Passagierflugzeuges geführt hat, welches sich einem Kriegsschiff näherte und welches nicht schnell genug auf die Warnung des Kriegsschiffes reagierte.

Man steht auf dem Standpunkt, dass fast jegliche Art von menschlicher Tätigkeit auch von Computern ausgeführt werden kann, wenn auch nicht sogleich, so doch zumindest in der absehbaren Zukunft. In der Neuzeit scheint diese Devise sogar an Berechtigung gewonnen zu haben, in Anbetracht der großen Fortschritte, die inzwischen erzielt worden sind. Alle Welt spricht vom vollautomatischen Kraftfahrer, der den Menschen am Lenkrad entbehrlich machen wird, mit dem Hinweis auf die gelungene automatisierte Fahrt auf der Autobahn. Das Auto der Zukunft fährt nicht nur von ganz allein, es weiß angeblich sogar schon ohne Auftrag, wohin der Fahrgast will. Es gibt den automatischen Rasenmäher, der die Grundstücksgrenzen beachtet. Es gibt den Haushaltsroboter für das Briefkastenleeren und für das Teekochen, und „der Computer“ besiegt schon den Schachweltmeister. Mit dem Hinweis hierauf wird sogar die prinzipielle Überlegenheit „des Computers“ gegenüber dem Menschen konstatiert. Es melden sich nun Menschen zu Wort, die sich einen freundlichen Roboter als Lebensgefährten vorstellen können, weil man ihm ein Programm einverleibt hat, welches seinem Gesicht ein freundliches Lächeln aufprägen kann.

Nichts scheint sich der Programmierbarkeit zu entziehen, wenn nur für die Erreichung dieses Ziels genügend Zeit und Geld zur Verfügung steht. Denn vieles hatte man schon für unmöglich gehalten und hat es dennoch verwirklicht. Soll man das „Unmögliche“ nicht wenigstens einmal probieren? Dies hat zur Folge, dass mit hohem Aufwand an Arbeitszeit und Kosten **illusorische Ziele** verfolgt werden und dass in der Erwartung der baldigen Erreichbarkeit derselben die traditionelle Arbeitsweise voreilig gedrosselt oder eingestellt wird, ohne dass der erwartete Ersatz zur Verfügung steht, sehr zum Nachteil und Schaden der bisherigen Nutzer dieser Tätigkeit. So hat

man in Vertrauen auf die volle Funktionsfähigkeit der vollautomatischen Fremdsprachenübersetzung schon geglaubt, auf den professionellen Dolmetscher verzichten zu können, ein Irrtum mit schlimmen Konsequenzen, denn viele Internet-Botschaften in orientalischen Sprachen sind deswegen schon unübersetzt und unbeachtet geblieben.

Und doch gibt es Grenzen der Erreichbarkeit und doch gibt es die Aussichtslosigkeit manch eines Vorhabens. Definitiv unerreichbar ist es beispielsweise, einen massiven Körper auf Lichtgeschwindigkeit zu beschleunigen, ein verlockendes Ziel, wenn man die Erkundung des Universums vor Augen hat, mag man auch schon große Fortschritte bei der Annäherung an ein solches Ziel erreicht haben. Man müsste nämlich für die Erreichung dieses Ziels einen unendlich großen Energieaufwand treiben, ein einfaches Gesetz der Physik.

Die Vertreter der harten Künstlichen Intelligenz stehen seit Jahrzehnten auf dem Standpunkt, dass auch alle Formen der Verarbeitung von Text und Bild programmierbar und automatisierbar seien, ein verlockendes Ziel, in Anbetracht der hohen Kosten und des hohen Zeitaufwandes, welcher mit der Verarbeitung solcher Dokumente auf die traditionelle Art durch sachverständige und sorgfältig arbeitende Fachleute verbunden sind.

Großer Schaden ist schon dadurch entstanden, dass im Vertrauen auf diese Versprechungen manch ein Projekt der traditionellen Textverarbeitung voreilig eingestellt worden ist, ohne dass der versprochene Ersatz zur Verfügung gestanden hat, weil die Entwicklung eines praxistauglichen, automatisierten Ersatzes auf der Strecke geblieben ist². Wenn der Fehlschlag offenkundig wird, müsste eigentlich die gesamte bisherige automatisierte Arbeit auf die traditionelle Art wiederholt werden, was kaum zu schaffen ist. Dann muss man sich mit den Folgen des Verzichtes auf eine traditionelle Arbeitsweise abfinden, und der eingetretene Schaden ist groß.

Es ist das Ziel dieses Aufsatzes, den Nachweis dafür zu führen, dass die automatisierte Textverarbeitung kein Ergebnis von der Qualität erbringen kann, wie es auf dem traditionellen Weg durch die sachkundige und sorgfältige Arbeit eines Fachmannes möglich ist.

Wohl aber lässt sich eine **zufriedenstellend erscheinende Automatisierung** dieser Tätigkeiten erreichen,

- wenn man sich auf die Verarbeitung von solchen Texten beschränkt, welche für die automatisierte besonders gut geeignet sind (Kapitel 5.1)

² Manch eine Stätte der bibliothekarischen oder informationswissenschaftlichen Ausbildung ist in jüngerer Vergangenheit schon geschlossen worden.

und

- wenn man die Ansprüche an die Qualität des Ergebnisses herabsetzt (Kapitel 5.2).

Der Fehlschlag droht dort, wo es auf die **größtmögliche Vollständigkeit der Recherche** in gespeicherten Texten ankommt, wie es der Fall ist

- im privaten Archiv der Familienakten und der beruflichen Akten
- im privaten Archiv der gesammelten Dokumente auf dem eigenen Interessengebiet
- bei der Patentliteratur³
- bei den unternehmensinternen Archiven
- bei der Erfassung der Nebenwirkung von Arzneimitteln⁴

Es genügt auch nicht, dass ein *wichtiges und sofort benötigtes Dokument* erst nach langem Suchen wiedergefunden wird.⁵

Wie groß sind die realistischen Aussichten auf Erfolg bei der programmierten Textverarbeitung einzuschätzen?

Einen Überblick über die Möglichkeiten und Grenzen einer vollautomatisierten Informationsbereitstellung gewinnt man dann, wenn man eine **Grenze zwischen den determinierten und den indeterminierten Prozessen** zieht (z. B. Mater 1990). Die indeterminierten Prozesse werden sich als unprogrammierbar erweisen, *wenn man sich als Ziel ein Ergebnis setzt, das der Arbeit des fach- und sachkundigen, sorgfältig arbeitenden Fachmannes gleichkommt*.

In dem Bemühen um eine verlockende Automatisierung sind in der Informationstechnologie allerlei obskure Algorithmen entwickelt worden (und werden dort noch entwickelt), deren undurchsichtige Ergebnisse unter wirklichkeitsfernen Umständen produziert wurden und die

³ Wenn die für eine Erfindung bestehenden Schutzrechte nicht rechtzeitig beachtet werden, dann kann es zu Forderungen des Patentbesitzers in unternehmensgefährdender Höhe kommen.

⁴ Aus Gründen der Arzneimittelsicherheit muss immer ein vollständiger Überblick über alle bisher beobachteten Nebenwirkungen eines Arzneimittels herrschen.

⁵ In den USA musste ein Kernkraftwerk für längere Zeit abgeschaltet werden, weil ein wichtiges Dokument zu Fragen der Betriebssicherheit unauffindbar gewesen ist. Treffend ist eine solche Situation durch eine Anekdote charakterisiert, welche in Fachkreisen zirkuliert: In einem Verkehrsministerium wurde das Manuskript zu einer Rede gesucht, welche zur Einweihung einer Straßenbrücke über einen Fluss gehalten worden war. Mit keinem der zahlreichen Stichwörter, unter denen im Archiv gesucht wurde, war das Dokument zu finden. Es musste der zuständige Archivar aus dem Urlaub zurückgerufen werden. Dieser fand es sofort mit dem Suchwort „halb“. Es war unter „H“ abgelegt: Die Brücke war **halb** zwei Uhr eingeweiht worden.

sich der Überprüfung durch den Nutzer entziehen. Wo liegt die Grenze zwischen Illusion und Realität?

3 Der indeterminierte Prozess und seine Automatisierung

Will man einen Prozess irgendwelcher Art automatisieren, dann muss der Programmierer für einen jeden Schritt in diesem Prozess einen entsprechenden Befehl im Computerprogramm festlegen, denn **„der Computer“ führt nur das aus, was ihm vom Programmierer in seinem Programm vorgegeben worden ist.**

Kann man nicht *im Voraus* festlegen, was alles vom Programm verarbeitet werden soll und wie es verarbeitet werden soll, dann gibt es auch keine brauchbare Automatisierung. Auch der eindrucksvollste Computer kann ohne solche vorherige Festlegungen nicht arbeiten, mag dies bei naiver Betrachtungsweise auch möglich erscheinen, weil man von „dem Computer“ schon so manche Überraschung erlebt hat. Auch wäre es dann nicht der Computer, sondern der Programmierer, der ein solches Wunderding vollbracht hätte.

Das Wesen eines indeterminierten Prozesses liegt in seiner **Unvorhersehbarkeit**. Dem Urtyp eines indeterminierten Prozesses begegnet man in der Physik beim Zerfall radioaktiver Atome. Es ist unvorhersehbar, wann und ob ein bestimmtes Atom zerfällt und das daneben liegende Atom der gleichen Art erst tausend Jahre später. Macht man nun einen Vorgang in der Makrowelt von dem Zerfall eines Atoms abhängig, wie es Pascal Jordan im Gedankenexperiment getan hat und Erwin Schrödinger mit seiner legendär gewordenen Katze, dann ist das Eintreten eines solchen Prozesses ebenfalls unvorhersehbar.

Ein anderes Beispiel der Indeterminiertheit aus der Physik ist das Doppelpendel, ein Pendel, an welchem ein zweites Pendel befestigt ist: Jenseits eines bestimmten Schwingungsausschlages vollführt dieses Pendel „chaotische“, d. h. unvorhersehbare Bewegungen, die keiner Gesetzmäßigkeit folgen.

Prozesse können auch dadurch indeterminiert sein, dass sie in einer **unendlich großen Zahl von unterschiedlichen Varianten ablaufen** können und dass es **unvorhersehbar** ist (oder nicht mehr rekonstruierbar ist oder nicht erraten werden kann), in welcher Variante der Prozess ablaufen wird (oder abgelaufen sein könnte).

Zumindest teilweise indeterminiert ist zum Beispiel der Verlauf einer Reise. Zwar kann man den Flug buchen und auch das Hotelzimmer. Aber es werden immer unvorhersehbare Ereignisse eintreten, auf welche man vor Ort

und zum gegebenen Zeitpunkt richtig reagieren muss. Würde man den Verlauf seiner Reise ausschließlich nach seinen *vorherigen* Festlegungen gestalten, wie man es tun müsste, wenn man deren Verlauf von einem Computerprogramm steuern lassen würde, dann würde man bald Schiffbruch erleiden.

Auch hat der Schachweltmeister keineswegs „gegen den Computer“ verloren, sondern gegen eine ganze Gruppe von hervorragenden Schachspielern, welche das Programm entwickelt haben. Sie haben im Gegensatz zum Weltmeister für die Festlegung ihrer Züge monatelang Zeit gehabt, konnten sich untereinander beraten und die Schachbibliotheken zu Rate ziehen. Sie haben den Computer lediglich als ein gutes Werkzeug benutzt. Und das Einparken in eine Lücke ist lediglich eine Aufgabe für die Mathematik, nachdem die Lücke ausgemessen und mit der Lenkgeometrie abgestimmt worden ist. Es ist voreilig und unseriös, die Besonderheit dieser Sachlage außer Betracht zu lassen und die baldige Entbehrlichkeit des denkenden Menschen am Lenkrad zu behaupten. Wie oft ist eine kritische Situation im Straßenverkehr (z. B. vier Autos gleichzeitig an der Kreuzung) nur dadurch folgenlos geblieben, dass man gegen die Regeln des Straßenverkehrs verstoßen hatte, gegen Regeln, wie sie in einem künftigen Computerprogramm ihren Niederschlag finden würden!

Indeterminiert ist auch der Prozess, wie der Mensch seine Ideen mit den Wörtern und Sätzen seiner Fach- und Umgangssprache ausdrückt.

Die natürliche Fach- und Umgangssprache verfügt nämlich über eine Unendlichkeit von sprachlichen Ausdrucksmöglichkeiten für einen Gegenstand des Interesses, einschließlich der definitionsartig, satzartig umschreibenden, nichtlexikalischen („paraphrasierenden“) Ausdrucksmöglichkeiten (z. B. Blair 2002, Mater 1990).

Hiervon kann man sich leicht überzeugen, wenn zwei verschiedene Personen von ein und demselben Aufsatz ihre Zusammenfassungen schreiben. Diese werden sich immer voneinander unterscheiden. Von dieser Sachlage kann man sich auch im Selbstversuch überzeugen, wenn man diese Arbeit in zeitlichem Abstand wiederholt.

Es ist deswegen **unvorhersehbar** (und auch nicht rekonstruierbar), für welche Wortwahl sich ein Verfasser entscheiden wird (oder entschieden haben könnte). Indeterminiert ist deswegen auch eine jegliche Bearbeitung, welche an einem solchen Text vorgenommen wird, z. B. dessen Indexierung, Übersetzung, Zusammenfassung (siehe Kapitel 4).

Soll nun der indeterminierte Prozess der Sprachbearbeitung auf dem programmierten Weg abgewickelt werden, dann steht der Programmierer vor einem Problem: Er müsste nämlich die unendlich zahlreichen Möglichkeiten

der natur- und fachsprachlichen Ausdrucksweise in seinem Programm berücksichtigen. Es wird also auf seiner Seite die Kenntnis aller denkbaren Ausdrucksweisen für einen Gegenstand von Interesse vorausgesetzt, auch der zukünftigen! Er muss sozusagen *a priori* arbeiten (nach Kant). Und er müsste für jede dieser Ausdrucksweisen die entsprechenden Befehle zu ihrer Verarbeitung niederlegen. Aber niemand kann in einem Computerprogramm unendlich viele Möglichkeiten berücksichtigen und dafür unendlich viele Befehle niederlegen.

Deswegen basiert die programmierte Textverarbeitung immer auf einer mehr oder minder großen *Auswahl von Einzelfällen von Ausdrucksweisen*, und diese Auswahl wird immer mangelhaft sein, nämlich mehr oder minder lückenhaft. Für diejenigen Ausdrucksweisen, welche in dieser Auswahl unberücksichtigt geblieben sind, findet das Programm nun keine Verarbeitungsbefehle vor, und diese Fälle werden vom Programm nicht verarbeitet werden. Bestenfalls bemerkt das Programm diese Lücke und weist diese Fälle einem Fachmann zur Bearbeitung zu. Von einer vollautomatischen Textverarbeitung kann dann aber keine Rede mehr sein.

Hingegen wird bei der traditionellen Abwicklung eines indeterminierten Prozesses dem Bearbeiter die so problematische (*a priori*) **Voraussicht** der zur Bearbeitung anstehenden Einzelfälle **nicht abverlangt**. Der Bearbeiter hat die zu bearbeitenden Fälle auf seinem Schreibtisch, nimmt sie zur Kenntnis und in seine Bearbeitung. Er arbeitet sozusagen ganz unproblematisch *a posteriori* (nach Kant). Dies ist eine Aufgabe, welche für den sachkundig und sorgfältig arbeitenden Fachmann leicht lösbar ist und auch immer gelöst worden ist. Das einzige Problem besteht in dem damit verbundenen hohen Aufwand an Zeit und Kosten. Aber das ist eher eine Frage der Prioritäten beim Einsatz der vorhandenen personellen und finanziellen Ressourcen.

Aber auch der traditionelle Weg ist nicht fehlerfrei: Es kann Mangel an Sachkundigkeit und Aufmerksamkeit herrschen. Jedoch gibt es probate Mittel, um diesen Fehlerquellen zu begegnen: Man kann z. B. jede Indexierung von mindestens zwei unabhängig voneinander arbeitenden Fachleuten bearbeiten lassen, wie es im Dokumentationsunternehmen der IDC (Internationale Dokumentationsgesellschaft für Chemie mbH) Routine gewesen ist und auch bei der firmeninternen Chemiedokumentation der ehemaligen Hoechst AG. Auf diese Weise lässt sich ein Maximum an Zuverlässigkeit bei der traditionellen Abwicklung eines indeterminierten Prozesses erreichen.

Der Straßenverkehr ist ein Gebiet, auf welchem der Mangel an Aufmerksamkeit und Sachkundigkeit die Regel ist und auch unausrottbar ist. Die Folge ist, dass hier-

zulande mehr als tausend Tote jährlich im Straßenverkehr zu verzeichnen sind. Hier hat es der Automat relativ leicht, selbst mit seinem notorisch lückenhaften Programm im Ergebnis gleichzuziehen mit der traditionellen Fahrweise, selbst wenn er sich auf das Einhalten eines ausreichenden Sicherheitsabstandes vom vorausfahrenden Fahrzeug beschränkt und auf die Feststellung von Ermüdungserscheinungen beim Fahrer. Auch kann man dem Kraftfahrer Erholungspausen auferlegt oder dass man technische Hilfsmittel anwendet, welche Ermüdungserscheinungen auf seiner Seite feststellen und melden. Auch ein Mangel an Sachkundigkeit ist leicht feststellbar. Er kann durch die Auferlegung eines Auffrischkurses behoben werden.

Natürlich kann auch die Programmierung der Kraftfahrt fortgesetzt verbessert werden, indem die Auswahl derjenigen Einzelfälle fortgesetzt erweitert wird, auf welche das Programm dann richtig reagiert. So wird auch die Annäherung an die traditionelle, wohlüberlegte Fahrweise fortgesetzt verbessert. Aber immer wird diese Auswahl lückenhaft bleiben, denn die Gesamtmenge derjenigen Einzelfälle, welche im Straßenverkehr eintreten können und auf welche das Programm richtig reagieren sollte, ist praktisch unbegrenzt groß. Es ist eine Illusion, dass man auf dem programmierten Weg gleichziehen könne mit dem genügend sachkundigen und genügend aufmerksamen Kraftfahrer. Von der baldigen völligen Entbehrlichkeit des Menschen am Lenkrad in tagtäglichen komplexen Verkehrssituationen, wie sie fortgesetzt aus Kreisen der Informationstechnologie und der Künstlichen Intelligenz postuliert wird, kann keine Rede sein (siehe auch Kapitel 5.5: Das übertölpelte Management). Eine vollautomatische Demonstrationsfahrt von Deutschland nach China, veranstaltet von einem Autohersteller, sollte die Praxis-tauglichkeit der eingebauten Automatik beweisen. Diese Fahrt musste wegen des Versagens des Programms bald abgebrochen werden. Sie wurde auf die traditionelle Art fortgesetzt.

Dem Begriff der Indeterminiertheit wird man in Kreisen der Künstlichen Intelligenz kaum begegnen, verweist er doch auf die natürlichen Grenzen dieses Gebiets, über welche dort nicht gern gesprochen wird. Auch bestätigt man dort nicht gern die Existenz solcher Grenzen. Wir werden aber hoffentlich nie den vollautomatischen Arzt, Friseur, Reiseplaner haben, obwohl man sie uns wegen ihrer Billigkeit und Schnelligkeit schmackhaft zu machen versuchen wird und obwohl damit mancherorts Geld gespart werden kann.

4 Die programmierte Verarbeitung von Texten zur vervollkommenen Informationsbereitstellung

Es genügt weder im Alltag, noch im Berufsleben, dass man die täglichen *Nachrichten* an sich vorüberziehen lässt und dass man sich dasjenige zu merken versucht, was *momentan* von Interesse und was demzufolge zur *Information* zu rechnen ist⁶. Vielmehr muss man sich den Zugriff auch auf all dasjenige bewahren, was *später einmal* von Interesse sein könnte. Dies gilt nicht nur für das, was einem in seiner Muttersprache begegnet, sondern auch für Fremdsprachliches. Der Bedarf an einer vervollkommenen Informationsbereitstellung herrscht

- bei der **Fremdsprachen-Übersetzung** (Kapitel 4.1),
- beim **Indexieren** zum Erstellen von Registern und zum **Recherchieren in einem indexierten Speicher** (Kapitel 4.2),
- beim **Recherchieren in einem natursprachlichen Dokumentenspeicher** (Kapitel 4.3),
- beim **Erstellen von Zusammenfassungen** (Kapitel 4.4).

Die hohen Kosten und der hohe Zeitaufwand bei der traditionellen Arbeitsweise haben schon frühzeitig dazu angeregt zu versuchen, diese Arten der Informationsbereitstellung zu automatisieren, dies in Anbetracht der großen Erfolge, die man andernorts mit der Automatisierung schon erzielt hat. Fortlaufend erreichen uns die entsprechenden Erfolgsmeldungen aus den Kreisen der Künstlichen Intelligenz. Aber viele von ihnen haben sich als voreilig herausgestellt und sind vom Wunschdenken der betreffenden Arbeitsgruppen geprägt gewesen und von ihrer Fehleinschätzung vom Wesen dieser Prozesse.

Es hängt viel davon ab, ob man als Nutzer eines solchen Programms erkennen kann, wie weit man sich mit der Automatisierung dem hohen Ziel genähert hat oder noch von ihm entfernt ist. Der Anbieter eines solchen Programms wird gerne eine übertrieben große Annäherung an das gesteckte Ziel konstatieren.

Kennt man den Wortlaut eines solchen Textes nicht, weil man ihn nicht gelesen (und ihn sich nicht gemerkt)

hat, dann kann man auch keine brauchbaren Befehle zur programmierten Ausführung einer solchen Verarbeitung geben. Man wird dann auch keine gute Automatisierung erreichen können. Man kann dann nur *vermuten*, mit welchem Wortlaut die gesuchten Texte formuliert sein könnten, und nur aufgrund solcher Vermutungen kann man dann die entsprechenden Befehle ins Programm aufnehmen. Mit den entsprechenden Unsicherheiten ist dann auch jegliche Automatisierung der Textverarbeitung verbunden.

4.1 Die programmierte Fremdsprachenübersetzung

Am deutlichsten sind Erfolg und Fehlschlag der Automatisierung bei der Fremdsprachenübersetzung zu erkennen. Hier hat der Nutzer eines solchen Programms einen vollkommenen Überblick über die Leistung und die Fehlschläge des Programms.

Nachfolgend wird die automatische Übersetzung eines englischsprachigen Textes ins Deutsche wiedergegeben, erzielt mit <https://translate.google.com/#en/de> im Stand vom Oktober 2015.

Es handelt sich um das Abstract zum Aufsatz: The five-axiom theory of indexing and information supply (Robert Fugmann) in Journal of the American Society of Information Science 1985, 36, No. 2, 116–129 (Best JASIS paper 1985):

Ausgangstext (Lauber):

The paper is an attempt to work out theoretical foundations of indexing and information processing from the work experience of information systems in chemistry and related disciplines. The proposed theory rests on five axioms:

Definability: The composition of responses relevant to a topic can be delegated only to the extent to which the inquirer can define the topic in terms of concepts and concept relations.

Order: Any compilation of responses relevant to a topic is an order-creating process.

Sufficient degree of order: The demands made on the degree of order increase as the size of the collection and/or the frequency of the searches increases.

Representational predictability: The completeness of any directed search (especially the recall ratio) depends on the predictability of the modes of expression for concepts and concept relations in the search file.

Representational fidelity: The accuracy of any directed search for texts (especially the precision ratio) depends on the fidelity with which concepts and concept relations are expressed in the search file.

The observation of the axioms assists in the design of effective IR systems.

⁶ In Kreisen der Informationstechnologie und der Künstlichen Intelligenz ist die Unterscheidung zwischen Information und Nachricht verloren gegangen (Rechenberg in seinem Buch Was ist Informatik? „Informatik hat nur wenig mit Information zu tun“). Die Informatik sollte sich besser und weniger irreführend als Nachrichtentechnik bezeichnen, zumindest so lange sie sagt „Die Einheit für die Information ist das Bit“.

Die automatische Übersetzung ins Deutsche lautet buchstäblich:

Das Papier ist ein Versuch, theoretische Grundlagen der Indexierung und Informationsversorgung ausgehend von der Arbeitserfahrung von Informations Systemen in der Chemie und verwandten Disziplinen. Die vorgeschlagene Theorie ruht auf fünf Axiomen.

Definierbarkeit: Die Zusammenstellung der relevanten Antworten zu einem Thema kann nur in einem Umfang übertragen auf die der Anfrager definiert werden kann das Thema in Bezug auf Konzepte und Konzeptarbeit.

Bestellen Sie: Jede Compilation relevanter zu einem Thema ist ein Befehl-Erzeugungsprozess.

Ausreichende Ordnungsgrad: Die Anforderungen an den Grad der Ordnung Anstieg als die Größe hergestellt der Sammlung und/oder Häufigkeit der Suchfragen zu.

Representational Berechenbarkeit: Die Vollständigkeit der gezielten Suche nach relevanten Texten (insbesondere erinnern Verhältnis) abhängig von der Voraussagbarkeit der Ausdrucks für Konzepte und Konzeptarbeit bei der Suche Datei.

Representational Treue: Die Genauigkeit jeder gezielten Suche nach relevanten Texten (insbesondere Genauigkeit Verhältnis) abhängig von der Genauigkeit, mit der Konzepte und Konzeptarbeit werden in der Suchdatei ausgedrückt.

Die Einhaltung dieser Axiome unterstützt in der Gestaltung der effektiven IR-Systeme.

Als Ergebnis ist festzuhalten:

- Das Übersetzungsprogramm hat bei der Aufgabe versagt, die Bedeutung der Textwörter zuverlässig zu erkennen: „Order“ ist in „Befehl“ und „bestellen Sie“ übersetzt worden. In „order-creating“ hingegen ist die Bedeutung von „order“ richtig erkannt worden.
- „Concept relations“ ist in „Konzeptarbeit“ übersetzt worden.
- Die Bedeutung vom Wort „compilation“ ist nicht erkannt worden. Es ist unübersetzt geblieben.
- „Recall ratio“ hätte als „Vollständigkeitsverhältnis“ übersetzt werden müssen und nicht als „erinnern Verhältnis“.

Das Urteil darüber, ob der englische Text zufriedenstellend übersetzt worden ist, bleibt dem Leser überlassen.

Herrschen problemmildernde Umstände, wie sie in den Kapiteln 5.1 bis 5.2 beschrieben sind, dann können **durch eine automatische Übersetzung brauchbare Ergebnisse** erzielt werden. Zumindest kann auf diesem Weg schnell und preiswert eine Ahnung vom Thema eines Dokuments vermittelt werden.

Aber davon, dass das Problem der automatischen Fremdsprachenübersetzung gelöst sei, wie aus den Kreisen der Künstlichen Intelligenz und verbreitet aus den zeitgenössischen Medien verlautet, dürfte jedoch keine

Rede sein, nach mehr als einem halben Jahrhundert intensiver Forschung auf diesem Gebiet.

Bei der automatischen Übersetzung begegnen wir der Möglichkeit, dass sich der Nutzer eines Computerprogramms einen echten Einblick in die Möglichkeiten und Grenzen seines Programms verschaffen kann. Sie treten hier offen zutage.

4.2 Das programmierte Indexieren von Dokumenten

Schon immer hat der Bedarf bestanden, in einer Sammlung von Dokumenten (oder in einem Buch) etwas Gesuchtes effektiver wiederfinden zu können, als durch das bloße Durchblättern einer Sammlung von Dokumenten oder eines Buches. Das schnelle und treffsichere Wiederfinden wird durch das **Indexieren** der gesammelten Dokumente gewährleistet. Es ist so alt wie der Buchdruck.

Unter „Indexieren“ versteht man unter manchem anderen

- erstens das Erkennen der fachlichen Essenz eines Dokuments
- und
- zweitens das Beschreiben dieser Essenz in einer hinreichend *wiedergabetreuen* und hinreichend sicher *voraussehbaren* Form (laut FID, Fédération Internationale de la Documentation).

Erstens wird bei der **klassischen Indexierung** erkannt, was vom Inhalt eines Dokuments wahrscheinlich für einen Fachmann des betreffenden Gebiets wesentlich ist, was als Essenz zu betrachten ist und was deswegen wiederauffindbar gemacht werden sollte. Wird das Unwesentliche nicht abgetrennt, dann droht bei der Recherche eine Überflutung mit Ballastantworten. Keinem Information Suchenden ist damit gedient, wenn ihm auf seine Fragestellung Hunderttausend Antworten vorgelegt werden, unter welchen allenfalls jede Hundertste von wirklichem Interesse für ihn ist, eine Kalamität, die vom Recherchieren im Internet her wohlbekannt ist. Aus gutem Grund wird man bei der Volltextrecherche im Internet mit „Auf gut Glück“ begrüßt. Unter den Informationsanbietern im Internet ist ein ständiges Wettrennen im Gang, mit ihren Angeboten auf die ersten Plätze unter den Antworten bei einer Fragestellung zu gelangen.

Der professionelle Indexer nimmt Aufträge zum Registerstellen nur auf seinem Fachgebiet an, wo er sich für die Entscheidung zwischen Wesentlichem und Unwesentlichem kompetent fühlt. Es ist nicht zu erkennen, wie man einem Computerprogramm die Intuition und das Fach-

wissen hierfür einverleiben kann. Der Schachcomputer ist hierfür kein Gegenbeispiel, denn die Zahl der in Betracht zu ziehende Züge ist nicht unbegrenzt groß. Demzufolge handelt es sich bei der Wahl der Züge im Schachspiel nicht um einen indeterminierten Prozess.

Zweitens ist beim Indexieren die Essenz eines Dokuments in einer Sprache auszudrücken und einzuspeichern, die für einen Fragesteller voraussehbar (oder rekonstruierbar) ist.

Den *Stichwörtern*, die man den Texten entnimmt, haftet meistens die Unvoraussehbarkeit der natursprachlichen Ausdrucksweise an (Ausnahmen werden in Kapitel 5.1 erörtert). So kann zum Beispiel der Vorgang der Korrosion oder die gefährliche Wasserglätte auf der Fahrstraße auf unbegrenzt viele verschiedene Arten natursprachlich ausgedrückt werden. Entnimmt man solche Stichwörter (oder Kombinationen von ihnen) unbesehen aus einem Text zur Einspeicherung, dann kann man keine zuverlässigen Suchbedingungen an einen solchen Speicher richten, weil man nicht wissen kann, mit welchen Wörtern man suchen muss.

Die *Schlagwörter* hingegen, die der Indexer eines Textes festlegt und benutzt und die er dem Informationssuchenden für dessen Fragestellung anbietet, erfüllen die Forderung der Voraussehbarkeit. Sie gewährleisten mit hoher Sicherheit eine vollständige Literaturrecherche.

Allerdings muss dieses Angebot an Schlagwörtern auch übersichtlich und gut strukturiert sein, damit sowohl der Indexer, als auch der Fragesteller die für sein Thema **bestpassenden** Schlagwörter auch genügend leicht findet⁷.

Somit ist jede Art von Indexierung, die diesen Namen verdient (und welche nicht auf die bloße Extraktion von Stichwörtern aus den gespeicherten Texten hinausläuft) zugleich auch **eine Übersetzung**, nämlich die Übersetzung in die Sprache der **Schlagwörter**, welche für die

betreffenden Texte zu deren inhaltlichen Erschließung angelegt worden sind⁸.

Einer jeglicher Automatisierung des Indexierens haften demzufolge unausweichlich die Mängel der automatischen Übersetzung an.

Es handelt sich um Mängel, die gravierend sein können, die aber unter besonderen Umständen auch hingenommen werden können, wie in den Kapiteln 5.1 und 5.2 ausgeführt wird.

An alldem können auch allerlei undurchsichtige wörter- und buchstabenstatistische Algorithmen nichts ändern, durch welche man zuweilen erfolgreiches automatisches Indexieren vorgetäuscht bekommt. Was hier zuweilen den Entscheidungsträgern in den Betrieben vorgeführt wird, kann man mit Fug und Recht als faules Ei des Kolumbus bezeichnen (Fugmann 2005). Es hat mit einem solchen Algorithmus die schwere Erkennbarkeit seines Mangels gemeinsam.

Bei der Belobigung dieser Algorithmen wird die Tatsache verkannt, dass das geschriebene oder gesprochene Wort meistens die Bedeutungsklärung und Interpretation erfordert, wenn es seinen Zweck erfüllen soll (siehe auch Kapitel 4.1: Die programmierte Fremdsprachenübersetzung). Aber auch hier handelt es sich um einen Prozess, welcher wegen seiner Indeterminiertheit nur mangelhaft programmierbar ist.

Nun kann man aber bei der Indexierung eines Textes nicht für *alles* von mutmaßlichem Interesse die passenden Schlagwörter einrichten. Ein solches Schlagwortvokabular würde viel zu schnell wachsen und würde bald viel zu unübersichtlich werden. Dem Zweck, auch diesen Teil des gespeicherten Wissens für die Wiederauffindung zu erschließen, kann ein **natursprachlicher Speicher für die Zusammenfassungen** (oder Annotationen) zu den gespeicherten Dokumenten dienen und die **Volltextrecherche** in diesem Speicher (siehe Kapitel 4.3).

Zwar begegnet man in diesem natursprachlichen Speicher wieder der Unvoraussehbarkeit der Ausdrucksweise. Aber man trifft wenigstens Vorsorge für Dokumente, wel-

⁷ Auch der Indexierer muss einen guten Überblick über sein Schlagwort-Vokabular haben, denn er muss immer bemüht sein, die **bestpassenden** Schlagwörter für das jeweilige Thema zu verwenden (Cutter's rule), so wie es ja auch der Fragesteller tun wird, damit er seiner Fragestellung die gewünschte Treffsicherheit verleiht. So ist allein ein vielversprechendes Schlagwortvokabular mit sehr vielen spezifischen Schlagwörtern noch keine Gewähr für gutes Indexieren und treffsicheres Recherchieren. Es muss vom Indexierer auch zuverlässig angewandt worden sein. Die Dewey Decimal Classification ist ein gutes Beispiel für ein zwar hochspezifisches, aber zugleich auch unzuverlässig angewandtes Vokabular von Notationen. Bei der Suche mit einer dieser Notationen findet man nämlich nur einen Bruchteil dessen, was in dieser Klassifikation erfasst ist. Der Rest bleibt unter anderen Notationen weit verstreut.

⁸ Höher entwickelte Klassifikationen benutzen zusätzlich zu ihrem Vokabular auch eine Grammatik. Hierdurch wird auch die Art der Verknüpfung dargestellt, welche zwischen den Schlagwörtern in einem einzuspeichernden Text herrscht. Dies gewährleistet eine hochgradige Wiedergabetreue der indexsprachigen Ausdrucksweise, mit entsprechend hoher Treffsicherheit der Recherche, bei erhalten bleibender Übersichtlichkeit des Vokabulars, allerdings auch bei entsprechend höherem Einspeicherungsaufwand. Es erfordert viel unternehmerischen Weitblick, ein derartig kostspieliges Dokumentationsvorhaben schon zu Zeiten in Gang zu bringen, zu denen der Bedarf an hoher Treffsicherheit und Wiedergabetreue noch klein ist, wie es im Anfangsstadium immer der Fall ist.

che sich sonst der Wiederauffindbarkeit gänzlich entziehen würden. Ein Speicher für natursprachliche *Abstracts*, den man sich auch aus den in Kapitel 4.4 genannten Gründen anlegen sollte, ist für diesen Zweck besser geeignet als die *vollen Texte* der Dokumente.

4.3 Die Recherche im natursprachlichen Speicher

Zuweilen wird in der unbesehenen Einspeicherung von natursprachlichen Volltexten die Lösung aller Archivierungsprobleme erblickt („Das macht man heute alles mit dem Computer“). Und leicht lässt sich auch ein Erfolg im Anfangsstadium des Wachstums einer Sammlung vorführen. Hierauf wird im Kapitel 5.3 näher eingegangen. Aber leicht kann man in die Klemme geraten, wenn man in einem solchen Speicher dringend ein wichtiges Dokument sucht und wenn man dann mit keinem der naheliegenden Stichwörter fündig wird.

Im Archiv der Haushaltsakten des Verfassers wurde einmal ein Zeugnis gesucht, aber es wäre mit dem Stichwort „Zeugnis“ in einem nichtindexierten natursprachlichen Speicher unauffindbar gewesen. Die gespeicherten Zeugnisse hätte man dort suchen müssen unter

„Erlaubnis zur Führung der Berufsbezeichnung“,
 „Gebührenbefreiung wegen sehr guter Leistung“,
 „Erreichung des Hauptschulabschlusses“,
 „Ergebnis der Abschlussprüfung“,
 „Bestätigung der mittleren Reife“,
 „Approbationsurkunde“, usw.,

um sich das Gewünschte herausuchen zu können, alles Suchbedingungen, die in ihrer Gesamtheit dem Suchenden in seiner Bedrängnis sicher nicht eingefallen wären.

Ein großer Teil der informationswissenschaftlichen Profession hatte sich in den 1960er Jahren durch einen Großversuch (genannt ASLIB Cranfield II)⁹ übertölpeln lassen. Hier wurde scheinbar nachgewiesen, dass man auf die Indexierung von Texten verzichten könne, weil man in nichtindexierten Texten praktisch genau so gut fündig werden könne.

Was hatte man gemacht? Die Versuchspersonen bekamen Dokumente vorgelegt, die sich im Speicher befanden und nach denen sie anschließend versuchsweise suchen

sollten („source documents“). So konnten sie zur Formulierung ihrer Suchfragen Einblick in den Wortlaut der Dokumente nehmen, die sie wiederfinden sollten. Es wurden ihnen für ihre Suchfragen also die genau passenden Stichwörter vorgelegt, mit denen sie dann natürlich auch fündig wurden.

Es kann keine Rede davon sein, dass hier ein Test von vier unterschiedlichen Recherchenmöglichkeiten durchgeführt wurde, wie behauptet wurde. Es hat sich vielmehr lediglich um einen Funktionstest des benutzten Wörtersuchprogramms gehandelt. Man hatte die Versuche unter wirklichkeitsfernen, Problem mildernden Umständen durchgeführt, von denen weitere in den Kapiteln 5.1 und 5.2 erörtert sind. Noch lange Zeit ist an diesen fehlerhaften und irreführenden Versuchsergebnissen fest gehalten worden, trotz der massiven Kritik, die hieran geübt worden ist (z. B. Swanson 1965 „They turned titles into questions, a trivial exercise“)¹⁰.

4.4 Das programmierte Erstellen von Zusammenfassungen

Wenn der Speicher der gesammelten Dokumente einige Jahre lang gewachsen ist und schon Tausende von Dokumenten umfasst, dann wird man beim Recherchieren in diesem Speicher auf zahlreiche Fundstellen stoßen. Dann wird man es sehr schätzen, wenn man über die zahlreich gewordenen Fundstellen einen guten Überblick vermittelt bekommt, wie er mit Zusammenfassungen vermittelt wird. Dies ist ein Vorteil, den man im Kleinzustand eines Informationssystems noch nicht zu schätzen weiß (siehe Kapitel 5.3: Das Kleinsystemsyndrom und die Gedächtnisfalle).

Steht schon die programmierte *Erkennung der Bedeutung der Wörter* in einem Text auf schwachen Füßen, dann gilt dies logischerweise auch für die programmierte Erkennung von der Wichtigkeit der Wörter (und Wortkombinationen zu Sätzen) in einem Text. Solche Entscheidungen sind beim *Erstellen einer Zusammenfassung* zu einem Text fortgesetzt zu treffen. Deswegen darf man auf die *programmierte* Erstellung von Zusammenfassungen keine großen Hoffnungen setzen.

⁹ In ASLIB Cranfield I hatte man sich nicht darauf einigen können, welche gefundenen Dokumente als zutreffende Antworten zu zählen sind. Man hatte versäumt, die Grenze zwischen *relevant* (d. h. zutreffend für die gestellte Frage) und *pertinent* (d. h. auch anderweitig interessant für den Fragesteller) zu ziehen.

¹⁰ In Analogie zu der Versuchsgestaltung bei ASLIB Cranfield II gilt: Wer die Ostereier selbst versteckt hat, ist beim Suchen danach immer im Vorteil, selbst wenn er halbblind ist. Nur darf er dann nicht behaupten, dass beim Suchen gutes Sehvermögen entbehrlich ist, ein Fehler, den man ganz in Analogie zu ASLIB Cranfield II begehen würde.

Es ist ratsam, dass die Zusammenfassungen zu den gesammelten Dokumenten vom Nutzer selbst erstellt werden oder von einem Fachmann des betreffenden Gebietes, denn die Erkennung der Essenz eines Textes ist hochgradig subjektiv. Ein solcher Volltextspeicher von Zusammenfassungen dient auch als wertvolle Ergänzung eines indexierten Speichers für die Volltextrecherche in demselben, wie bereits in den Kapiteln 4.2 und 4.3 beschrieben.

5 Automatisierung fördernde Umstände

Nun können aber in der Praxis Umstände herrschen, in denen die programmierte Verarbeitung von Texten zur verbesserten Informationsbereitstellung vereinfacht und erleichtert ist. Auch kann es sein, dass der Nutzer eines Informationssystems nur bescheidene Ansprüche hat. Dann kann auch das programmierte Verarbeiten von Texten zufriedenstellende Ergebnisse erbringen. Wenn die Besonderheit der Sachlage nicht erkannt wird, in der man sich hier befindet, dann kann leicht der Eindruck entstehen (oder erweckt werden), dass die Programmierung *generell* zufrieden stellend arbeitet und *auf allen ihren Gebieten* bedenkenlos eingesetzt werden könne. Aber wer gegenüber den Entscheidungsträgern seine Zufriedenheit mit einem billigen, anspruchlosen Informationssystem äußert, das er nur unter Problem mildernden Umständen kennen gelernt hat, der behindert die Entwicklung und Unterhaltung von leistungsstarken, teureren Systemen.

5.1 Problem mildernde Umstände aufseiten der zu verarbeitenden Texte

Besonders verarbeitungsfreundlich sind Texte, in denen die dort verwendeten Wörter keiner **Bedeutungskklärung** bedürfen. Kommt noch hinzu, dass das Gemeinte immer **lexikalisch ausgedrückt** ist, das heißt nicht in der paraphrasierend-umschreibenden Form, dann steht man nicht vor der Unendlichkeit von denkbaren Ausdrucksweisen, wenn man nach einem Thema seines Interesses auf der Suche ist. Herrscht obendrein auch ein relativ *einfacher Satzbau*, dann steht ein Übersetzungs- oder Indexierungsprogramm vor einer leichten Aufgabe.

Zum Beispiel braucht man bei den Wetterberichten nicht damit zu rechnen, dass die Himmelsrichtungen in der nichtlexikalischen, paraphrasierenden, unvorhersehbaren Weise ausgedrückt sind. Auch begegnet man dort nicht der sonst weit verbreiteten *Mehrdeutigkeit* der Wör-

ter, ein Problem, welches bis heute in der Programmierung ungelöst ist, wie aus dem Übersetzungsbeispiel in Kapitel 4.1 hervorgeht. Auch ist es eine fast elementare Aufgabe, ein Programm für die *Liste der Autoren* von gesammelten Texten anzufertigen, denn Autorennamen sind von Natur aus lexikalisch. Obendrein stehen sie immer an maschinell leicht zu ermittelnder Stelle.

Auch herrschen drastisch vereinfachende Umstände, wenn sich die Recherche ausschließlich auf Individualbegriffe beschränkt, wie Personen, Städte, Länder, geographische Begriffe usw. Es ist typisch für dieselben, dass auch sie fast immer in der lexikalischen, gut voraussehbaren Weise ausgedrückt sind. Der Umgang mit nur einigen Hunderttausend Wörtern, auf welche sich ein natur-sprachlicher Wortschatz beschränkt, ist kein Problem für ein Computerprogramm.

Der Verfasser einer Schrift, die zur maschinellen Übersetzung bestimmt ist, kann viel zur Milderung der Probleme beitragen, welchen man hier begegnet. Entsprechende Ratschläge findet man unter „übersetzungsgerechtes Schreiben“ im Internet. Beispiele hierfür sind das Vermeiden von mehrdeutigen Wörtern, von komplexen Satzstrukturen, von Füllwörtern. Man sollte die Verben nur in der Gegenwartsform benutzen und manches andere mehr. Unter diesen vereinfachenden Umständen lassen sich maschinelle Übersetzungen von erstaunlicher Qualität erzielen.

5.2 Problem mildernder Umstände aufseiten der Nutzer der Informationsbereitstellung

Verbreitet herrscht im Alltag kein Bedarf an einer absolut vollständigen Informationsbereitstellung. Sucht man nach einem Stadtplan von München, dann braucht man keinen Hinweis auf *sämtliche* Stadtpläne. Möchte man sich über die Symptome einer bestimmten Krankheit informieren, dann braucht man nicht mit der gesamten medizinischen Literatur konfrontiert zu werden, wenn man sich nicht gerade als Arzt auf der Suche befindet. Es ist leicht, hier beim Nutzer auch mit lückenhaft arbeitenden Programmen Zufriedenheit zu wecken. Nur sollte nicht deswegen die *generelle Entbehrlichkeit* einer vollständig arbeitenden Informationsbereitstellung konstatiert werden. Andernorts wird sie nämlich dringend benötigt, wie in Kapitel 2 ausgeführt.

5.3 Das Kleinsystemsyndrom und die Gedächtnisfalle bei der „modernen“ Archivierung

Weit verbreitet sind die fehlerhaften Schlussfolgerungen, die man voreilig aus dem scheinbar guten Funktionieren eines (noch!) kleinen Informationssystems zieht. In diesem Frühstadium der Entwicklung vertraut man leicht darauf, dass die eigene Gedächtnisleistung zum Finden der passenden Suchwörter auch in der fernerer Zukunft noch ausreichen wird, und dass man zum Suchen in der Sammlung ausreichend Zeit zur Verfügung haben wird. Aber im fortgeschrittenen Entwicklungsstadium wird man immer häufiger wichtige und wichtig gewordene Dokumente nicht mehr wiederfinden können oder nur noch mit untragbar großem Aufwand an Zeit und Geduld und mit unsicherer Erfolgsaussicht. Man ist in vielen kleinen Schritten zum Opfer des Kleinsystem Syndroms („small system's syndrome“) geworden. Der Schaden ist groß. Entweder man kapituliert und verzichtet immer häufiger auf die Nutzung seiner Dokumentensammlung oder man beginnt auf besser gesicherter Grundlage noch einmal ganz von vorn.

Es ist auch die irrige Meinung weit verbreitet worden, dass man mit der *Digitalisierung* seines Aktenbestandes, unternommen mit dem Ziel der Reduzierung des Platzbedarfes für das Archiv, bereits zugleich auch das bestmögliche für dessen *Erschließung zur Recherche* getan habe. Der Computer könne ja dann Wort für Wort in einem solchen Archiv wiederfinden, und der Mensch mache ja auch nichts anderes, wenn er sich in einem solchen Archiv auf die Suche begibt. So lautet die Argumentation der Programmverkäufer. Wie sehr man sich mit dieser Meinung im Irrtum befindet, wenn man mehr wiederfinden können möchte als lediglich die Namen von Personen, Institutionen, Städten und Ländern, geht aus den vorangegangenen Ausführungen hervor. Auch auf diese Weise ist das sachkundige Management schon oft übertölpelt worden (siehe auch Kapitel 5.5).

5.4 Die Verborgtheit der Fehlschläge des Programms

Manchmal ist man mit seinem Suchprogramm nur deswegen zufrieden, weil man nicht weiß, was einem bei der Recherche alles entgeht. Dann wird der Standpunkt vertreten, dass man keine professionelle Informationsbereitstellung benötigt. Dann kommt es zur Überraschung, wenn man sein Forschungsergebnis zum Patent anmelden will: Alles schon bekannt, jahrelange Arbeit umsonst. Das Management ist gut beraten, wenn es die Mitarbeiter ver-

pflichtet, vor dem Beginn einer jeden Arbeit in Forschung und Entwicklung eine professionelle Literaturrecherche durchführen zu lassen. Und die betriebsinterne Dokumentationsabteilung hat den Auftrag, solche Literaturrecherchen auch *unaufgefordert* durchzuführen und die Ergebnisse den Mitarbeitern auch *unaufgefordert* zuzuleiten, dies auf der Grundlage ihrer Kenntnis von den Arbeitsgebieten der Mitarbeiter.

5.5 Das übertölpelte Management

Holt sich ein Betrieb einen Unternehmensberater ins Haus, dann kommt es oftmals auch zu einer wenig kompetenten Beratung in Fragen des unternehmensinternen Informationsflusses¹¹. Hier wird oft eine Möglichkeit zur drastischen und schnellen Kosteneinsparung gesehen („Das macht man heutzutage alles mit dem Computer“). Der Dr. Wichtigmann (nach Jörg Staute) kommt mit mokantem Lächeln gern auf das Postkutschenzeitalter zu sprechen und dass man „nicht die Frösche fragen darf, wenn man den Sumpf trocken legen will“. Das professionelle Personal für die Informationsvermittlung wird in die altertümliche Ecke gestellt. Man beeindruckt den Forschungsleiter, indem man ihm eine komplette Liste seiner Veröffentlichungen ausdrückt, und das ganz allein auf bloßen Knopfdruck. Der Informationsprofessional wird dann zu den entscheidenden Sitzungen gar nicht mehr hinzu gezogen, weil sein Arbeitsgebiet auf eine andere Ebene verlagert worden ist und weil sich hier jetzt andere Leute zu Wort melden. Manch eine unkorrigierbare Fehlentscheidung ist auf diesem Weg schon entstanden (siehe auch Kapitel 5.3).

6 Schlussfolgerung

Die hier in den Vordergrund gestellte Art der Informationsbereitstellung basiert auf der Zuweisung von inhaltskennzeichnenden Schlagwörtern, „Indexierung“ genannt. Dieses Verfahren ist mit hohem Aufwand verbunden. Der späte Lohn hierfür besteht im schnellen und weitestgehend vollständigen Wiederfinden. Diese Investition lohnt sich aber nur dort, wo es wirklich auf schnelles und gesichertes Wiederfinden des Gespeicherten ankommt. Eine derartige Sachlage ist im Informationsalltag nur selten anzutreffen. Stattdessen hat sich die unbesehene elek-

¹¹ Der Anti-Consulter: Ein Unternehmensberater attackiert seine Zunft. Geldschinderei und Wortgeklingel. Consulting als Hochstaplei. Focus 35/1996, 148.

tronische Einspeicherung von Dokumententexten weit ausgebreitet. Sie findet wegen ihrer Schnelligkeit und Billigkeit großen Zuspruch, dies aber größtenteils auch deswegen, weil die dort eintretenden Suchfehlschläge meistens verborgen bleiben. Sie werden nämlich nur dann offenbar, wenn man vergeblich versucht hat, ein Dokument wiederzufinden, von welchem man weiß, dass es im Speicher enthalten ist.

Dr. rer. nat. Robert Fugmann

Auf der Au 4

65510 Idstein

r.fugmann@amweg.de

Dr. rer. nat. Robert Fugmann, Diplomchemiker, (geboren 1927 als Sohn des Lehrers Alfred Fugmann, Meldereiter im ersten Weltkrieg und seiner Ehefrau Elly, geborene Thomas), mit einem schon in die Wiege gelegten Faible für Nachricht und Information, Soldat in der Nachrichtentruppe der Deutschen Luftwaffe im Zweiten Weltkrieg, 1952 bis 1959 tätig im Pharma-Bereich der Farbwerke Hoechst AG, Frankfurt am Main Höchst, von 1960 bis zur Pensionierung 1987 Leiter der Abteilung für Wissenschaftliche Dokumentation bei der Farbwerke Hoechst AG, Urheber der computerisierten Chemie-Informationssysteme GREMAS und TOSAR, eingeladen zum Vortrag der Keynote Address bei der 47. Konferenz der Fédération Internationale de la Documentation (FID), Tokio, 2. bis 9. Oktober 1994, Empfänger des Ranganathan Award vom Sarada Ranganathan Endowment und Empfänger der Gmelin-Beilstein Gedenkmünze der Gesellschaft Deutscher Chemiker, Empfänger des Skolnik Award der American Chemical Society, Autor des Best Paper 1985 im Journal of the American Society for Information Science, tätig gewesen als Lehrkraft an mehreren Universitäten und Fachhochschulen in Deutschland und in den USA, Ehrenmitglied in zwei informationswissenschaftlichen Gesellschaften.