

Universität Bamberg
Angewandte Informatik

Seminar KI: gestern, heute, morgen

Problematik des GPS in paradigmatischer Hinsicht und
historische Lösungsansätze

von

David Hollinger

15. Februar 2017

betreut von
Prof. Dr. Ute Schmid

In dieser Studie werden Paradigmen des Forschungsbereichs der Künstlichen Intelligenz (KI) erläutert und zueinander in Verbindung gesetzt. Paradigmen strukturieren die Art, wie über ein Thema gedacht wird. Zur Erläuterung der Paradigmen wird ein frühes Produkt der Forschung zur Künstlichen Intelligenz näher beleuchtet: Der General Problem Solver (GPS). Es werden dem GPS vier Paradigmen (Thinking Humanly, Symbolisch, Power-based/General Paradigma, Starke KI) zugeordnet. Dabei ergeben sich diverse Schwierigkeiten. In einem weiteren Schritt werden spätere Paradigmen (Kognitive Modellierung, Knowledge-based/Domain-specific Paradigma, Subsymbolisches Paradigma, Schwache KI) eingeführt, die häufig kontrastiv zu den Paradigmen des GPS stehen. Diese Paradigmen bieten Lösungen für Schwierigkeiten, die durch die Paradigmen des GPS hervorgerufen werden. Zum Abschluss wird der Forschungsgemeinschaft nahegelegt, sich mit den Paradigmen der eigenen Forschung auseinanderzusetzen.

1 Einleitung

Menschen erforschen bereits seit Jahrhunderten ihr eigenes Wesen, ihren Geist, ihre Kognition und ihre Intelligenz. Die Erfindung des Computers brachte die Möglichkeit mit sich, einer Maschine Intelligenz zu verleihen. Der Forschungsbereich, der sich diesem Anliegen annimmt, heißt **Künstliche Intelligenz**.

Künstliche Intelligenz ist ein Forschungsbereich, der in den 50er Jahren gegründet wurde. Die heutigen Fortschritte, Methoden und Ansichten in diesem Bereich beruhen zu großen Teilen auf Forschung aus dieser Zeit.

Es ist jedoch umstritten, was diesen Forschungsbereich ausmacht, denn die Frage danach, was Künstliche Intelligenz ist, lässt sich ganz unterschiedlich beantworten. Geht es darum, Systeme zu schaffen, die so intelligent wie möglich sind? Und was ist Intelligenz? Oder sollen die Systeme den Menschen ähnlich sein, oder sogar ihre kognitiven Prozesse möglichst korrekt abbilden? Diese und weitere Fragen stellen sich der Forschung schon seit Jahrzehnten. Je nachdem wie die Fragen beantwortet werden, ergeben sich unterschiedliche Forschungsergebnisse. So wird die Künstliche Intelligenz zu einem sehr breiten Forschungsgebiet, das sich aus mehreren nebeneinander existierende Unterbereichen zusammensetzt. Diese haben jeweils ihre eigenen Vorstellungen darüber, welche Annahmen für ihre Arbeit zugrunde gelegt werden können, und welche Ziele zu erreichen sind. Im Verlaufe der Arbeit wird auch deutlich, dass es sogar unterschiedliche Meinungen darüber gibt, welche Ergebnisse überhaupt zu erwarten sind. Selbstverständlich unterscheiden sich diese Unterbereiche auch durch die genutzten Werkzeuge. Die Gesamtheit aus Annahme, Ziel und Werkzeug, wird in dieser Arbeit als **Paradigma** bezeichnet. Ein Paradigma formt die Sicht auf den Forschungsgegenstand. In der Künstlichen Intelligenz gibt es mehrere, zum Teil kontrastive, zum Teil miteinander in Beziehung stehende Paradigmen.

In dieser Arbeit werden die Paradigmen der KI anhand eines sehr frühen Produktes der KI-Forschung näher beleuchtet: Der **General Problem Solver** (auch: GPS) von Newell und Simon aus dem Ende der 50er Jahre (Newell und Simon, 1959). Es handelt sich dabei um ein Programm, das den Anspruch hat, Probleme in der gleichen Weise wie der Mensch zu lösen. Kapitel zwei wird sich mit der Funktionsweise des GPS eingehend beschäftigen. Dabei soll deutlich werden, dass der GPS verschiedenen Paradigmen der KI-Forschung zugeordnet werden kann. Anhand dieser Paradigmen soll auf die Schwierigkeiten¹ des GPS eingegangen werden.

In Kapitel drei wird aufgezeigt, welchen anderen Paradigmen sich die spätere KI-Forschung zugewandt hat, wie diese im Gegensatz zu den Paradigmen des GPS stehen und wie versucht wird, die Schwierigkeiten zu lösen bzw. zu umgehen. Zur Struktur dieser Studie siehe auch Abbildung 1.

¹In dieser Studie werden Probleme im mathematischen Sinne verstanden. Um Missverständnisse zu vermeiden, werden die problematischen Aspekte des GPS, die alltagssprachlich ebenfalls als Problem bezeichnet werden könnten, als Schwierigkeiten bezeichnet.

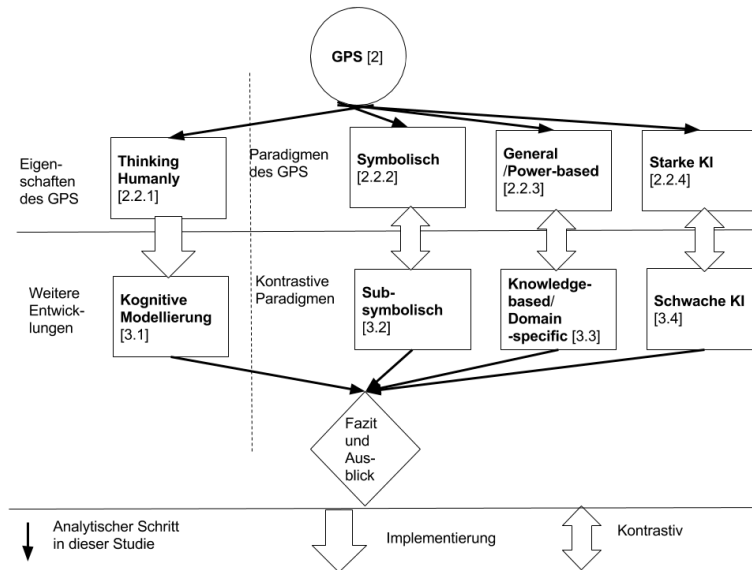


Abbildung 1: Die Struktur dieser Studie

2 General Problem Solver

2.1 Funktionweise

Die Funktionsweise des General Problem Solvers basiert auf vier Hauptbestandteilen: Rekursion, Trennung zwischen Problem und Problemlösungstechnik, Heuristiken sowie Speicher- und Programmorganisation (Newell und Simon, 1959, S.1). Auf die Speicher- und Programmorganisation wird in dieser Arbeit nicht weiter eingegangen, da sie für die Forschungsfrage nicht relevant ist. Die drei anderen Hauptbestandteile werden an geeigneten Stellen eingeführt.

Im Folgenden wird zunächst definiert, was unter einem Problem zu verstehen ist, danach wird erläutert, wie die Daten für den GPS gewonnen wurden und welche Heuristiken daraus abgeleitet wurden. Als nächstes wird das Prinzip der Problemreduktion dargelegt. Danach wird die konkrete Anwendung des GPS auf bestimmte Probleme demonstriert.

2.1.1 Problem und Problemlösung

Nach Newell und Simon (1959, S.1) liegt ein **Problem** vor, wenn ein Problemlöser (Mensch oder Maschine) ein Ziel erreichen möchte, dies jedoch nicht unmittelbar möglich ist. Der Problemlöseprozess besteht darin, solange bereits verfügbare Information zu nutzen, um mehr Information zu erlangen, bis ein Lösungsweg gefunden wurde.

Die durch den GPS zu lösenden Probleme lassen sich durch **Objekte** und **Operatoren** formulieren (Newell und Simon, 1959, S.1). Ein Objekt wird durch seine **Eigenschaften** und durch seine **Unterschiede** zu anderen Objekten definiert.² Ein Operator wird auf

²Unter Task Environment wird anhand von Beispielen deutlicher, was ein Objekt sein kann.

Objekte angewandt, um neue Objekte zu erzeugen. Ein Ziel kann z.B. ein Objekt mit bestimmten Eigenschaften sein.

Task Environment bezeichnet die Menge der Objekte und Operatoren für ein bestimmtes Anwendungsgebiet (Newell und Simon, 1959, S. 5). Eine mögliche Task Environment ist die symbolische Logik (Newell und Simon, 1959, S.13). Objekte sind dort Variablen (z.B. P, Q), Junktoren ($\wedge, \vee, \rightarrow$) und \neg . Die Operatoren sind hier 12 Regeln, die beschreiben, wie ein gegebener logische Ausdruck in einen anderen logischen Ausdruck überführt werden kann.

Bsp.: Regel R1³

$$A \vee B \Rightarrow B \vee A$$

Ein Ziel in dieser Task Environment wäre z.B. die Überführung eines gegebenen logischen Ausdrucks in einen anderen, ebenfalls gegebenen, Ausdruck.

Im Folgendem soll das Problemkonzept nach Schmid und Kindsmüller (1996, S.142) erweitert werden. Ein **Zustandsraum** bzw. Problemraum besteht aus der Menge aller Zustände eines Problems, der Menge der Anfangszustände, der Menge der Zielzustände und der Menge der Operatoren, die einen Zustand in einen anderen Zustand überführen. Zustandsräume können als Graphen dargestellt werden, in denen alle Zustände als Knoten repräsentiert sind und zwei Zustände durch Kanten miteinander verbunden sind. Diese Kanten geben an, dass einer der miteinander verbundenen Zustände in den anderen Zustand überführt werden kann. So kann die Problemlösung als die Suche nach einem Weg vom Anfangszustand zu einem Zielzustand in diesem Graphen betrachtet werden. Für einfache Probleme ist schlichtes Ausprobieren verschiedener Wege durch den Graphen prinzipiell möglich, für komplexere Probleme mit größeren Zustandsräumen ist dies nicht sinnvoll.

Eine mögliche Suchstrategie ist die Tiefensuche. Hier wird jeder Knoten maximal einmal besucht und alle ausprobierten Wege werden gespeichert. Dabei wird ein Weg so lange weiter verfolgt, bis der Zielzustand erreicht ist. Bei der Breitensuche hingegen werden alle Knoten, die vom aktuellen Knoten aus erreichbar sind, „erkundet“. Dies ist zwar aufwändiger, hilft aber den kürzesten Weg (= die beste Lösung eines Problems) zum Ziel zu finden. Tiefen- und Breitensuche sind grundlegende und wichtige Suchstrategien. Allerdings sind ihnen in der Anwendung durch Speicherplatzbeschränkungen Grenzen gesetzt.⁴ Der GPS nutzt jedoch eine andere Suchstrategie: Die Mittel-Ziel-Analyse. Diese wird im Abschnitt Heuristiken behandelt. Zunächst muss jedoch erläutert werden, was zur Wahl dieser Strategie geführt hat.

³Es sei darauf hingewiesen, dass in dieser Arbeit die logischen Symbole aus der Literatur an die modernen Konventionen angepasst wurden (z.B. - zu \neg).

⁴Beispiel Schach: Es gibt schlicht zu viele Zustände, um eine der beiden Suchstrategien in ihrer Reinform einsetzen zu können.

2.1.2 Informationsgewinnung durch Think-Aloud-Test

Statt Überlegungen zu einer möglichst effektiven und effizienten Lösung eines Problems, bildet ein psychologisches Experiment die Grundlage des GPS. In diesem Experiment wurden Versuchspersonen dazu aufgefordert, Probleme aus der symbolischen Logik zu lösen (Newell und Simon, 1959, S.2). Dabei sollte die Gleichheit zweier logischer Ausdrücke durch Umformung bewiesen werden.

Beispiel:

$$(R \rightarrow \neg P) \wedge (\neg R \rightarrow Q)$$

überführen in

$$\neg(\neg Q \wedge P)$$

Den Versuchspersonen wurden dabei 12 Regeln zur Umformung der Ausdrücke vorgelegt (Newell und Simon, 1961, S.111). Pro Schritt konnte eine Regel auf den ersten Ausdruck angewandt werden. Wenn die Versuchsperson eine Regeln anwenden wollte, hat sie dies dem Versuchsleiter mitgeteilt, welcher dann das Ergebnis der Anwendung (wenn möglich) auf eine Tafel schrieb. Es handelte sich um einen Think-Aloud-Test, d.h. die Versuchspersonen sollten während der Problemlösung erläutern, warum sie Schritte durchführen wollen. Somit sollte ein Einblick in die Suchstrategie der Versuchspersonen erlangt werden.

Es hat sich im Experiment ergeben, dass die Versuchspersonen häufig die Mittel-Ziel-Analyse zur Lösung eines Problems einsetzten. Das Programm soll diese Problemlösungsstrategie der Versuchspersonen simulieren (Newell und Simon, 1959).

2.1.3 Heuristiken

Der General Problem Solver nutzt die **Mittel-Ziel-Analyse** als Suchstrategie, erweitert durch **Planning** (Newell und Simon, 1959). Die Mittel-Ziel-Analyse basiert auf folgenden Annahmen (Newell und Simon, 1959, S.9):

1. Wenn ein gegebenes Objekt nicht das gewünschte Objekt ist, dann werden Unterschiede zwischen dem gegebenen und dem gewünschten Objekt beobachtbar sein.
2. Operatoren können benutzt werden, um Unterschiede zwischen dem Objekt, auf das der Operator angewandt wird, und gewünschtem Objekt zu beseitigen.
3. Es lohnt sich, schwierigere Unterschiede zwischen Objekten zu beseitigen, wodurch leichtere Unterschiede entstehen.

Newell und Simon (1959, S.8) zeigen, zusätzlich zur psychologischen Fundierung im Experiment, anhand des folgenden Beispiels, dass die Mittel-Ziel-Analyse eine alltäglich genutzte Heuristik ist. Im Beispiel nutzt die Person in einer Alltagssituation die Mittel-Ziel-Analyse, um ein Ziel zu erreichen:

I want to take my son to nursery school. What's the difference between what I have and what I want? One of distance. What changes distance? My automobile. My automobile won't work. What's needed to make it work? A new battery. [...]

Schmid und Kindsmüller (1996, S.182) beschreiben die Mittel-Ziel-Analyse als ein heuristisches Suchverfahren. Heuristische Suchverfahren (im Allgemeinen) suchen, ebenso wie Tiefen- oder Breitensuchen, von einem Anfangszustand aus nach einem Zielzustand, welcher die Lösung des Problems darstellt. Allerdings können heuristische Suchverfahren zusätzlich anhand einer Bewertungsfunktion abschätzen, ob sich der Übergang zu einem Folgezustand (also ein Knoten, zu dem eine Kante vom aktuellen Zustand her führt) lohnen wird, d.h näher an die Lösung heranbringen wird. Der Weg vom Anfangszustand zum Zielzustand wird in heuristischen Suchverfahren also mit Wissen (und zwar darüber, welche Aktionen sich lohnen) gesteuert. Statt der Bewertungsfunktion nutzt die Mittel-Ziel-Analyse, und damit auch der GPS, jedoch eine Differenz-Operator-Tabelle. Diese Tabelle gibt zu auftretenden Differenzen einen anwendbaren Operator an und muss im GPS für jede Task Environment eigens manuell (vom Programmierer) angelegt werden. Eine Differenz in der Task Environment zur Logik im GPS ist z.B. zwischen $P \subset Q$ und $P \vee Q$ zu finden, da hier die Konnektoren nicht identisch sind.

Da die Mittel-Ziel-Analyse immer nur einen Schritt „vorausdenken“ kann, wird Planning verwendet, um zunächst eine abstrakte Lösung zu generieren, bevor die eigentliche Lösung des Problems beginnt (Newell und Simon, 1959, S.18). Dabei werden folgende Schritte vorgenommen:

1. Abstrahieren, indem Details des Ursprungsobjekts weggelassen werden.
2. In der abstrakten Task Environment wird ein entsprechendes Problem formuliert.
3. Wenn das abstrakte Problem gelöst wurde, wird die Lösung benutzt, um das Ursprungsproblem zu lösen.
4. Der Plan wird zurück in die Ursprungs-Task Environment übersetzt und dort ausgeführt.

Der Vorteil von Planning ist, dass Probleme in Subprobleme zerlegt werden können. Diese Subprobleme sind einfacher zu lösen als das ursprüngliche Problem (Newell und Simon, 1959, S.20). Somit arbeitet das Suchverfahren des GPS nach dem Prinzip der **subgoal reduction**:

Make Progress by substituting for the achievement of a goal the achievement of a set of easier goals (Newell und Simon, 1959, S.8)

2.1.4 Weitere Task Environments

Wie bereits erwähnt ist eine der Task Environments des GPS die symbolische Logik. Eine andere Task Environment ist die Trigonometrie (Newell und Simon, 1959, S.16).

Hier sind die Objekte algebraische und trigonometrische Ausdrücke und die Operatoren sind Faktorisierung, algebraische Vereinfachung und trigonometrische Transformation. Ein Problem wäre hier z.B.:

$$(\tan x + \cot x) \sin x \cos x = 1$$

Hier muss der linke Ausdruck in den rechten Ausdruck überführt werden. Weitere Task Environments sind denkbar, so etwa im Schach.

2.2 Paradigmen

Im Folgenden sollen die Paradigmen, die dem GPS zugrunde liegen, erläutert werden. Dabei werden auch Schwierigkeiten benannt, die sich aus diesen Paradigmen ergeben. Im darauffolgenden Kapitel wird anhand komplementärer oder erweiternder Paradigmen aufgezeigt, wie die spätere Forschung auf die Schwierigkeiten reagiert hat.

2.2.1 Einordnung in Definition

Im folgenden werden vier Definitionen der Künstlichen Intelligenz von (Russell und Norvig, 2010, S.2) erläutert:

- **Acting Humanly:** Dieser Ansatz macht das Bestehen des Turing-Testes zur Bedingung dafür, dass ein System als intelligent angesehen werden kann. Der **Turing-Test** besteht daraus, dass ein Mensch schriftliche Fragen stellt, welche von einem Computer beantwortet werden. Wenn der Mensch nicht weiß, ob die Antworten von einem Menschen oder einem Computer kommen, dann hat der Computer den Turing-Test bestanden.
Dieser Ansatz wird jedoch kaum praktisch verfolgt, da die Ergründung der Intelligenz selbst als wichtiger als die Anfertigung einer Kopie des Menschen erachtet wird.
- **Thinking Humanly:** Die Idee des Ansatzes ist, dass durch eine hinreichend präzise Theorie des menschlichen Verstandes (Mind) es möglich wird, die Theorie als Computerprogramm auszudrücken. Um diese Theorie zu erstellen kann man entweder Introspektion, psychologische Experimente oder Untersuchungen des Gehirns anwenden. Dieser Ansatz wird von der **Kognitionswissenschaft** (KW) verfolgt.
- **Thinking Rationally:** In diesem Ansatz wird versucht, auf Grundlage logischer Regeln intelligente Systeme zu entwickeln. Dabei wird besonders darauf geachtet, dass die richtigen Inferenzen gezogen werden.
- **Acting Rationally:** Agenten handeln autonom, nehmen ihre Umgebung wahr, passen sich an und verfolgen Ziele. Ein **rationaler Agent** versucht, das bestmögliche Ergebnis zu erreichen. Dieser Ansatz ist dem Thinking Rationally-Ansatz sehr ähnlich, aber geht über diesen hinaus, z.B. kann ein rationaler Agent auch dann handeln, wenn es keine richtige Entscheidung gibt.

Russell und Norvig (2010, S.2) verfolgen in ihrem Einführungswerk zur Künstlichen Intelligenz ausschließlich den Ansatz Acting Rationally weiter und geben damit Anlass zur folgenden Einordnung: Künstliche Intelligenz ist zum Einen ein Oberbegriff für alle genannten Definitionen, gleichzeitig aber auch ein Unterbegriff dieses Oberbegriffs.⁵ Russell und Norvig (Russell und Norvig, 2010, S.3) ordnen den GPS in den Ansatz Thinking Humanly ein, da bei der Entwicklung der Fokus darauf gelegt wurde, dass der Ablauf (Trace) der Problemlösung mit den Arbeitsschritten der Versuchspersonen übereinstimmt, und nicht etwa, dass das Programm die Probleme richtig löst. Die Kognitionswissenschaft hat ihre Methode zur Verfolgung des Ansatzes Thinking Humanly unter dem Titel Kognitive Modellierung weiterentwickelt. Auf die Kognitive Modellierung wird im nächsten Überkapitel detaillierter eingegangen.

2.2.2 Symbolisches Paradigma

Ein Symbol ist eine Entität, die ein physikalisches Muster ist und als Komponente einer anderen Art von Entität auftreten kann, nämlich eines Ausdrucks (Newell und Simon, 1976, S.116).

Computer manipulieren durch ihre Berechnungen nicht nur Zahlen, sondern manipulieren auch Symbole (Newell und Simon, 1961, S.5), da die Zahlen Symbole repräsentieren, welche wiederum Dinge referenzieren.

Der GPS soll Symbole in der gleichen Weise manipulieren wie der Mensch. Im zuvor beschriebenen Task sieht das so aus: Das Programm erhält als Input die logischen Ausdrücke und produziert als Output eine Reihe von Regelanwendungen. Diese Reihe von Regelanwendungen des Programms soll mit der Reihenfolge der Regelanwendungen der Versuchsperson möglichst übereinstimmen (Newell und Simon, 1961, S.5).

Es handelt sich hierbei um einen symbolischen Ansatz, der später von Newell und Simon weiter ausgearbeitet wurde und als **Physical Symbol Systems Hypothesis** (Newell, 1980) bekannt wurde.⁶

Die Physical Symbol Systems Hypothesis lautet:

The necessary and sufficient condition for a physical system to exhibit general intelligent action is that it be a physical symbol system (Newell, 1980, S.170).

Notwendige Bedingung (necessary condition) bedeutet, dass jedes physikalische (bzw. physische) System, das generelle Intelligenz hervorbringt, ein Physical Symbol System ist. Mit genereller Intelligenz werden Systeme ausgeschlossen, die nur in einer beschränkten Domäne arbeiten können (Newell, 1980, S.171).

Hinreichende Bedingung (sufficient condition) bedeutet, dass jedes Physical Symbol System so strukturiert werden kann, dass es generell intelligente Handlungen hervorbringen kann (Newell, 1980, S.170). Der Ausdruck generell intelligente Handlungen bezieht

⁵Andernfalls wäre ein Widerspruch vorhanden, wenn sich KI und Kognitionswissenschaft voneinander abgrenzen. So sind KI und KW beide Teil der KI.

⁶Zwar ist der Abstand zwischen der Publikation des GPS und der Physical Symbol Hypothesis nicht unerheblich und daher eine Herstellung einer Verbindung zwischen diesen beiden nicht unproblematisch, jedoch geben die beiden Forscher an, dass die Physical Symbol Hypothesis in einer durch dieselben Forscher initiierten Forschungstradition steht (Newell, 1980, S.137).

sich auf ein Maß an Intelligenz, wie es beim Menschen zu beobachten ist („[...]that in real situations behavior appropriate to the ends of the system and adaptive to the demands of the environment can occur, within some physical limits“ (Newell, 1980, S.170)).

Russell und Norvig (2010, S.18) bringen die Hypothese auf den Punkt: „What they [Newell und Simon] meant is that any system (human or machine) exhibiting intelligence must operate by manipulating data structures composed of symbols.“

Die Physical Symbol Systems Hypothese besagt zusätzlich, dass die „technologische“ Grundlage eines Physical Symbol Systems prinzipiell irrelevant für das intelligente Verhalten eines Systems ist (Newell, 1980, S.175). Sowohl eine Computer-Architektur als auch das menschliche neuronale System dienen damit nur als Träger des Physical Symbol Systems.

Computer arbeiten nur mit diskreten Werten („Nullen und Einsen“), doch in der Natur sind kontinuierliche Phänomene zu beobachten. Physical Symbol Systems sind nach Newell und Simon ebenfalls diskret. Zwar steht dies in Widerspruch zur kontinuierlichen Natur, jedoch wird darauf hingewiesen, dass Physical Symbol Systems trotzdem in einer kontinuierlichen Natur agieren können (z.B. in der Sprachverarbeitung) (Newell, 1980, S.179).

Es sind die letzten beiden Ansätze, also die Irrelevanz der physikalischen Grundlage und Reduzierung auf diskrete Werte, die besonders kritisiert wurden. Es entwickelte sich daher ein Gegenparadigma, das insbesondere die Irrelevanz des neuronalen Systems infrage stellte, und sogar das symbolische Paradigma insgesamt angriff. Es handelt sich dabei um das subsymbolische Paradigma, welches später noch genauer erläutert wird.

2.2.3 Power-based/ General Paradigma

Der Namensbestandteil „General“ des General Problem Solvers geht auf die Behauptung zurück, dass dieses Programm theoretisch auf jedes Problem anwendbar wäre, da Problemlöseheuristiken und Task Environment getrennt sind und so dieselben Heuristiken auf unterschiedliche Task Environments angewandt werden können (Newell und Simon, 1959, S.6). Damit wird ein genereller Ansatz verfolgt, da ein Programm alle Probleme lösen soll (bzw. ein Programm alle Probleme so lösen soll wie der Mensch). Letztendlich konnte der GPS nur auf einfache Probleme angewandt werden, da die genutzten Heuristiken in ihrer generativen Ausführung mit der Komplexität anspruchsvollerer Probleme nicht zurechtkamen (Duda und Shortliffe, 1983, S.1). Später entwickelte sich mit Expertensystemen ein Paradigma, das nicht den Anspruch der Generalität stellte, sondern sich auf stark begrenzte Anwendungsgebiete fokussierte (siehe Knowledge-based Approach).

2.2.4 Starke KI

Die Idee, dass Maschinen intelligent sein können, ist schon lange Teil der Kultur. Mit dem Aufkommen des Science-Fiction-Genres wurde die Idee auch im modernen kulturellen Diskurs fest verankert. Eine mit diesem Gedanken verbundene Frage befasst sich mit der Möglichkeit, dass eine Maschine „Geist“ (**Mind**) besitzt. Wie die Antwort darauf lauten muss und ob diese Frage überhaupt zu beantworten ist, ist philosophisch stark

umstritten. An dieser Stelle sollen zwei Paradigmen in der Forschung zur Künstlichen Intelligenz aufgezeigt werden, welche die Frage unterschiedlich beantworten: Das Paradigma **Starke KI** sagt ja, die **Schwache KI** sagt nein. Im Folgenden wird auf das erste Paradigma eingegangen und argumentiert, dass der GPS dieses Paradigma annimmt. Die Ausführungen beziehen sich auf Searle (1980).

Nach (Searle, 1980, S.2) behauptet das Paradigma Starke KI, dass entsprechend programmierte Computer tatsächlich **verstehen** können und damit ein Mind besitzen. Searle argumentiert, dass Starke KI nicht möglich ist. Demgegenüber steht das Paradigma Schwache KI, das den Computer nur als Werkzeug begreift, um etwa eine Theorie (über die menschliche Kognition) zu testen. Russell und Norvig (Russell und Norvig, 2010, S.1020) führen diese beiden Definitionen ähnlich ein, jedoch mit einem Fokus auf Intelligenz: Demnach behauptet die Schwache KI, dass Maschinen so tun können, als ob sie intelligent wären, während die Starke KI argumentiert, dass ein solches Verhalten bereits intelligent ist. Es hat sich ein eigener, zwischen Philosophie und Informatik zu verortender, Diskurs entwickelt, der vor allem die Möglichkeit der Ansprüche der Starken KI diskutiert. Es soll hier nur kurz auf die Grundlage dieses Themas eingegangen werden, denn eine umfassende Betrachtung, die der Komplexität des Themas gerecht wird, muss Aufgabe einer eigenen Studie sein. Stattdessen wird hier stärker auf die wissenschaftstheoretischen Implikationen, die die Annahme des Paradigmas beinhaltet, eingegangen.

Searle (1980) veranschaulicht seine Position gegen die Starke KI mit seinem berühmten Gedankenexperiment „Das Chinesische Zimmer“: Eine englischsprachige Person ist in einem Raum eingesperrt. In dem Raum gibt es ein Regelbuch, das Ersetzungsregeln für chinesische Zeichen beinhaltet. Außerdem gibt es im Raum eine kleine Öffnung zur Außenwelt, durch die die Person Papiere mit chinesischen Zeichen erhält. Die Person konsultiert nun das Regelbuch, worin sie die Zeichen findet, die in diesen Papieren angegeben sind. Zu jedem Zeichen kann sie nun die im Regelbuch angegebenen Ersetzungen durchführen und das Ergebnis dieser Umformungen durch die Öffnung an die Außenwelt ausgeben. Für die außenstehenden Personen kann das Input- und Outputverhalten des chinesischen Zimmers so wirken, als ob der Raum bzw. die Person im Raum chinesisch verstünde. Searle argumentiert, dass in Wirklichkeit jedoch kein Verständnis vorliegt. Ebenso verhalte es sich mit dem Computer: Selbst wenn von außen das Verhalten des Computers so wirke, als ob dieser verstehe, so ist es intern nicht der Fall. Nach Searle (1980) ist es das Verstehen, das dem Computer fehle, um ein Mind zu sein. Vertreter der Starken KI weisen dieses Argument zurück.

Russell und Norvig kritisieren, dass Searles Gedankenexperiment nur auf Intuition basiere:

In the case of the Chinese Room, Searle relies on intuition, not proof: just look at the room; what's there to be a mind. But one could make the same argument about the brain: just look at this collection of cells (or of atoms), blindly operating according to the laws of biochemistry (or of physics)—what's there to be a mind? Why can a hunk of brain be a mind while a hunk of liver

cannot? That remains the great mystery. (Russell und Norvig, 2010, S.1033)

Da die Grundzüge der Debatte und ihre Unentscheidbarkeit nun grob umrissen sind, sollen die Implikationen, die das Paradigma Starke KI auf die Forschung hat, ausgeführt werden.

Der GPS wird hier als Produkt der Starken KI gewertet. Das scheint zunächst nicht offensichtlich zu sein, da das Programm in dieser Studie als Implementierung des Ansatzes Thinking Humanly – ein kognitionswissenschaftlicher Ansatz, der Computermodelle (Kognitive Modellierung) als Werkzeug nutzt, um Theorien der menschlichen Kognition zu testen – eingeordnet wurde. Allerdings ist der GPS symbolisch und Pionierarbeit für die Physical Symbol Systems Hypothesis. Searle ordnet die Arbeiten von Newell und Simon – explizit auch (Newell und Simon, 1961) – zur Starken KI, da diese behaupten, dass der Computer zur selben Kognition fähig sei wie der Mensch (Searle, 1980, S.5).

Um jedoch starke KI wirklich angemessen zu verkörpern, muss eben genau das Verständnis implementiert sein. Das ist natürlich beim GPS nicht der Fall. Der GPS stellt lediglich die Simulation einer einzigen Problemlösestrategie (Mittel-Ziel-Analyse), die nachweislich von Menschen eingesetzt wird, dar. Der GPS versteht nicht, was er tut. Jetzt könnte argumentiert werden, dass damit der GPS eben die Schwache KI als Paradigma vertritt. Doch dabei bleibt außer Acht, dass die Annahme des symbolischen Paradigmas eine Gleichheit zwischen menschlicher und Computer-Kognition erfordert. Um diesen Widerspruch der Paradigmen aufzulösen, gibt es zwei Möglichkeiten: Entweder soll das System menschliche Kognition nur simulieren (Schwache KI), oder es wird das symbolische Paradigma aufgegeben. Das nächste Kapitel wird beide Möglichkeiten erläutern.

3 Weiterführende Paradigmen

Im vorherigen Kapitel wurden Paradigmen der Künstlichen Intelligenz vorgestellt und es wurde ermittelt, inwieweit der GPS diese Paradigmen implementiert.

In diesem Kapitel werden weitere Paradigmen vorgestellt, die den Paradigmen des GPS kontrastiv gegenüberstehen. Dabei wird festgestellt, welche Lösungsansätze diese Paradigmen für Schwierigkeiten, die im GPS festgestellt wurden, bereithalten.

3.1 Kognitive Modellierung

Im letzten Kapitel wurde der GPS in die Definition Thinking Humanly eingeordnet. Es handelt sich dabei um einen Ansatz, der versucht eine Theorie des menschlichen Verstandes zu entwickeln und diese Theorie als Computerprogramm auszudrücken. An dieser Stelle soll ausgeführt werden, dass die Implementierung des Ansatzes später erweitert wurde.

Um prozedurales Wissen⁷ zu repräsentieren, werden Produktionssysteme eingesetzt (Schmid und Kindsmüller, 1996, S.174). Produktionssysteme beinhalten Produktionsregeln (bzw. Produktionen), die eine gegebene Datenstruktur auf eine in der Regel spezifizierte Weise transformieren, wenn die Bedingung zur Regelausführung erfüllt ist. Neben den Produktionsregeln stellt ein Produktionssystem außerdem einen Speicher zu Verfügung, der einzelne Datenmuster enthält, sowie einen Interpreter, der die Produktionen auf die Datenmuster anwendet (Schmid und Kindsmüller, 1996, S.175). Diese drei Komponenten sind auch im GPS zu finden. Der GPS ist ein solches Produktionssystem (Schmid und Kindsmüller, 1996, S.174). Er ist das erste Produktionssystem, das zur Beschreibung menschlicher Problemlöseprozesse verwendet wurde.

Der GPS wurde in den 80er Jahren zu Soar⁸ weiterentwickelt. Soar ist ebenfalls ein Produktionssystem. Im Gegensatz zum GPS ist Soar jedoch auch eine Kognitive Architektur und bietet damit die Möglichkeit zur Kognitiven Modellierung. Das bedeutet, dass das Programm Grundannahmen über die menschliche Kognition als Basis bereitstellt. Anwender des Programms können dann ihre Theorie über bestimmte Teilaspekte der menschlichen Kognition (z.B. über das Lösen einer bestimmten Aufgabe) als Kognitives Modell an die Kognitive Architektur übergeben.

Soar arbeitet wie der GPS mit Produktionsregeln, Problemzuständen und Teilzielen. Auch die Verwendung von Heuristiken ist in Soar wichtig für die Suche nach Lösungen eines Teilziels, um damit das eigentliche Ziel zu erreichen. Allerdings nutzt Soar das Konzept der Problemräume (Zustandsraum) intensiver als der GPS, indem es alle Produktionsregeln gleichzeitig in einem Auswertungsproblemraum anwendet und dann in einem Auswahlproblemraum die Produktionsregel mit einem Preference Value bewertet. Die Regel mit dem höchsten Wert wird dann tatsächlich angewendet (Schmid und

⁷Wissen darüber, wie etwas getan wird. Steht im Gegensatz zu deklarativem Wissen, dem „Was“.

⁸früher: SOAR für „State, Operator and Result“

Kindsmüller, 1996, S.186). Damit unterscheidet sich Soar vom GPS, da dieser eine Differenz-Operator-Tabelle nutzt, um die am besten passende Regel zu ermitteln.

Andere Kognitive Architekturen sind ACT* und ACT-R. Diese unterscheiden zwischen Faktenwissen (deklarativem Wissen) und Problemlösefähigkeiten (prozedurales Wissen). Damit rücken diese Produktionssysteme näher an eine differenzierte Betrachtung des menschlichen Gedächtnisses.

Schmid und Kindsmüller (1996) heben hervor, dass Architekturen wie Soar und ACT vor allem dazu dienen, eine komfortable und standardisierte Formulierung von Annahmen über kognitive Prozesse zu überprüfen. Eine direkte Implementierung wäre jedoch ebenfalls möglich.

3.2 Knowledge-based/ Domain-specific Paradigma

Wie bereits dargelegt, ist das General/Power-based Paradigma des GPS gescheitert. Später entwickelte sich ein **domänenspezifisches**, also auf ein bestimmtes Anwendungsgebiet festgelegtes, Paradigma, das großes Expertenwissen im jeweiligen Anwendungsgebiet nutzt, womit es sich um einen **wissensbasierten** Ansatz handelt (Duda und Shortliffe, 1983). Die daraus resultierenden Programme sind die **Expertensysteme**. Duda charakterisiert diesen Paradigmenwechsel folgendermaßen:

Paradoxically, it has been proved much easier to emulate the problem-solving methods of some kinds of specialists than to write programs that approach a child's ability to perceive, to understand language, or to make „common-sense“ deductions. (Duda und Shortliffe, 1983, S.2)

Allerdings stellen sich auch diesem Ansatz einige Schwierigkeiten, z.B. bei Wissensakquisition, Wissensrepräsentation, Inferenzen und Unsicherheit (Duda und Shortliffe, 1983, S.2). Schließlich hat der domänenspezifische Ansatz auch eine andere Zielsetzung als das General/Power-based Paradigma:

The goal of expert systems research is to provide tools that exploit new ways to encode and use knowledge to solve problems, not to duplicate intelligent human behavior in all its aspects. (Duda und Shortliffe, 1983, S.2)

3.3 Subsymbolisches Paradigma

Wie im Abschnitt zum symbolischen Paradigma bereits erwähnt, hat sich das subsymbolische Paradigma als Gegenbewegung zum symbolischen Paradigma entwickelt. Ein wichtiger Kritikpunkt der Vertreter des subsymbolischen Paradigmas am symbolischen Paradigma ist dessen Annahme, dass das Physical Symbol System unabhängig von seiner konkreten Implementierung (Computer-Architektur bzw. neuronales System beim

Menschen) ist. Smolensky (1987, S.98) spricht hier von verschiedenen kognitiven Levels. Das symbolische Paradigma geht laut Smolensky von zwei Levels aus: Das symbolische Level, auf dem das Physical Symbol System operiert, und das Implementierungslevel, auf dem die Computer-Architektur bzw. das neuronale System des Menschen zu finden ist. Das Implementierungslevel ist unwichtig für die korrekte Beschreibung von Kognition, da das symbolische Level unabhängig von der Implementierung arbeitet. Die Neurowissenschaft kritisiert an diesem Ansatz, dass mit ihm nur wenig Aufschluss über die Organisation der Berechnungsvorgänge im Gehirn möglich ist (Smolensky, 1987, S.98). Smolensky behauptet, dass nicht-symbolische Ansätze die Struktur der Kognition auf eine natürlichere Weise beschreiben können. Daher führt er das subsymbolische Paradigma ein, das ein weiteres Level zwischen dem symbolischen Level und dem neuronalen Level annimmt: Das **subsymbologische Level** (Smolensky, 1987, S.99). Dieses Level soll der Formalisierung der Prozesse im Nervensystem dienen. Dabei bildet die neuronale Struktur die Basis für die Formalisierung (Smolensky, 1987, S.100). Zur Realisierung des subsymbolischen Levels sollen konnektionistische Systeme eingesetzt werden, da diese deutlich näher am neuronalen System sind als symbolische Systeme (Smolensky, 1987, S.99).

Das subsymbolische Paradigma greift jedoch nicht nur die vom symbolischen Paradigma ausgewiesenen Ebenen der Kognition an, sondern auch den Kern dieses Paradigmas selbst: Die Symbole (Smolensky, 1987, S.100). Während im symbolischen Paradigma die gleichen Symbole benutzt werden, um einerseits auf semantisch interpretierbare Entitäten zu verweisen und andererseits durch Symbolmanipulation die Regeln des Systems zu definieren, werden im subsymbolischen Paradigma die semantisch interpretierten Entitäten als Aktivationsmuster einer großen Anzahl von Einheiten des Systems realisiert, wobei die Regeln für das System durch Activation Passing Rules (genauer: statistische Inferenzregeln) gebildet werden (Smolensky, 1987, S.100). Es geht dabei um einen Typ konnektionistischer Systeme, in dem Aktivationsmuster (und nicht etwa individuelle Elemente des Netzwerks) Konzepte repräsentieren.

Konnektionistische Systeme sind wie folgt aufgebaut: Grundlage sind hier viele kleine Prozessoren, die miteinander verknüpft sind und parallel arbeiten. Jeder dieser Prozessoren hat ein Aktivierungslevel (Activation Level). Die Stärke der Verknüpfungen (Connection Strength) zwischen den Prozessoren bestimmen, wie ein Input zu einem Output überführt wird. Damit wird ein System aufgebaut, das dem menschlichen Gehirn mit seinen Neuronen ähneln soll. Die Variablentypen Activation Level und Connection Strength sind kontinuierlich. Symbolische Systeme hingegen arbeiten nur in der diskreten Kategorie.

3.4 Schwache KI

Schwache KI ist das Gegenparadigma zur Starken KI. Es behauptet lediglich, dass Maschinen so tun können, also ob sie intelligent wären (Russell und Norvig, 2010, S.1020) und es nutzt Computer nur als Werkzeug zur Überprüfung von Theorien über die menschliche Kognition (Searle, 1980, S.2). Im Abschnitt zur Starken KI wurde gefordert, dass Starke KI und das symbolische Paradigma nicht in einem System vereint

werden sollen.⁹ Ersetzt man nun das Paradigma der Starken KI durch die Schwache KI, so kann auf alle Annahmen über die Gleichheit von Computer und menschlicher Kognition verzichtet werden. Searles Gedankenexperiment (Searle, 1980) ist für die Schwache KI irrelevant: Es spielt keine Rolle, ob die Person im chinesischen Zimmer versteht, was sie tut. Es zählt nur, dass das Output-Verhalten möglichst korrekt ist und dass mit dem Regelwerk neue Erkenntnisse über die Funktionsweise der chinesischen Sprache deutlich werden. Weniger metaphorisch formuliert: Computermodelle müssen nicht verstehen, was sie tun. Sie sollten aber menschliche Kognition präzise und konsistent beschreiben, um eine Überprüfbarkeit zu gewährleisten.

Als Systeme, die den GPS als Simulation menschlicher Problemlösung weiterentwickeln, dabei aber das Schwache Paradigma anstatt das Starke Paradigma annimmt, können alle Produktionssysteme, darunter auch Kognitive Architekturen, gesehen werden, die nicht beanspruchen, dass ihre Funktionsweise (z.B. Symbolmanipulation) genau der Funktionsweise der menschlichen Kognition entspricht.

⁹Eine Forderung, die gerne zurückgenommen wird, falls es in Zukunft Evidenz für Verstehen in symbolischen Systemen geben wird.

4 Fazit

In dieser Studie wurde festgestellt, dass der General Problem Solver ein Produktionssystem ist, das versucht, menschliche Problemlösung als heuristische Suche abzubilden. Dabei ist die Anwendung der Mittel-Ziel-Analyse zentral. Es wurde durch eine Think-Aloud-Studie gezeigt, dass die Mittel-Ziel-Analyse eine Problemlösestrategie ist, die von Menschen angewandt wird. Damit eignet sie sich auch als Strategie, um menschliche Problemlösung darzustellen.

In der paradigmatischen Analyse des GPS hat sich ergeben, dass dieser in die Definition Thinking Humanly von Russell und Norvig (2010, S.2) eingeordnet werden kann. In den folgenden Jahrzehnten wurde die Implementierung des Ansatzes Thinking Humanly zur Kognitiven Modellierung weiterentwickelt.

Des Weiteren wurde festgestellt, dass der GPS das symbolische Paradigma vertritt. Dieses Paradigma steht in Zusammenhang mit der Physical Symbol Systems Hypothesis. Die von diesem Paradigma vertretene Auffassung, dass sich Kognition unabhängig von der „Hardware“ (Computer-Architektur oder menschliches neuronales System) betrachten lasse, führte zum Aufkommen eines Gegenparadigmas: Das subsymbolische Paradigma. Dieses Paradigma ist mit dem Konnektionismus verbunden und versucht, eine Computer-Architektur zu nutzen, die eine größere technische Nähe zum menschlichen neuronalen System aufweist. Dabei wird jedoch auch die Bedeutsamkeit der Symbole selbst in Frage gestellt.

In dieser Studie wurde außerdem festgestellt, dass der GPS das Power-based, bzw. General Paradigma vertritt. Damit ist der Anspruch gemeint, dass ein Programm dazu fähig sein soll, alle (wohldefinierten) Probleme zu lösen. Dabei soll die gleiche Problemlösestrategie für alle Probleme angewandt werden (der „Kern“ des Programms bleibt also gleich), nur die entsprechende Task Environment muss an das Problem angepasst werden (andere Differenz-Operator-Tabelle für Logik als für Geometrie). Das Gegenparadigma hierzu lautet Knowledge-based bzw. Domain-specific. Dieses verzichtet auf den Anspruch, alle Probleme zu lösen, und beschränkt sich stattdessen auf klar abgegrenzte Bereiche (z.B. Medizin). Statt sich nur auf heuristische Suche zu verlassen, werden hier zusätzlich große Wissensdatenbanken über den entsprechenden Bereich angelegt.

Das letzte Paradigma, das in dieser Studie dem GPS zugeordnet wird, lautet Starke KI. Dieses Paradigma besagt, dass Computer verstehen und wirklich intelligent sein können (und nicht nur so tun). Es ist zugegebenermaßen die gewagteste Zuordnung in dieser Arbeit, doch setzt das symbolische Paradigma die Starke KI voraus: Wenn behauptet wird, dass Mensch und Maschine die gleichen Symbole nutzen, dann stehen Computer potenziell dem Menschen in ihren kognitiven Möglichkeiten in nichts nach. Die Schwache KI jedoch verzichtet auf diesen Anspruch und nutzt Computer nur zur Überprüfung von Hypothesen über die Kognition. Schmid und Kindsmüller (1996, S.333) hierzu:

Die Formulierung komputationaler Theorien setzt nicht die Annahme voraus, dass menschliche Informationsverarbeitung ein Prozess der Symbolmanipulation ist. Es genügt stattdessen, davon überzeugt zu sein, dass wesentliche

Aspekte menschlicher Informationsverarbeitung als Prozess der Symbolmanipulation beschreibbar sind.

Durch die vorliegende Arbeit wurde deutlich, dass eine differenzierte Betrachtung der Paradigmen, in der die eigene Forschung eingeordnet wird, fruchtbar sein kann, damit die Forschung wissenschaftstheoretischen Anfechtungen standhalten kann.

Literatur

- Richard O. Duda und Edward H. Shortliffe. 1983. Expert systems research. 220, 4594 (1983), 261–268.
- Allen Newell. 1980. Physical symbol systems. 4, 2 (1980), 135–183.
- Allen Newell und Herbert A. Simon. 1959. Report on a General Problem Solving Program. (1959).
- Allen Newell und Herbert A. Simon. 1961. GPS, A Program that Simulates Human Thought. (1961).
- Allen Newell und Herbert A. Simon. 1976. Computer science as empirical inquiry: Symbols and search. 19, 3 (1976), 113–126.
- Stuart J. Russell und Peter Norvig. 2010. *Artificial intelligence* (3. ed., internat. ed. ed.). Pearson.
- Ute Schmid und Martin Kindsmüller. 1996. *Kognitive Modellierung*. Spektrum, Akad. Verl.
- John R. Searle. 1980. Minds, brains, and programs. 3, 3 (1980), 417–424.
- Paul Smolensky. 1987. Connectionist AI, symbolic AI, and the brain. 1, 2 (1987), 95–109.