

Methoden der KI

Portfolioprüfung

Studienrichtung

Technische Informatik

Muhammad Aman Bin Ahmad Tifli

Matrikelnummer: 2042550

Prüfer: Prof. Dr. Thomas Rist

Abgabedatum: xx.xx.2021



**Hochschule
Augsburg** University of
Applied Sciences

**Fakultät für
Informatik**

Hochschule für angewandte
Wissenschaften Augsburg

An der Hochschule 1
D-86161 Augsburg

Telefon +49 821 55 86-0
Fax +49 821 55 86-3222
www.hs-augsburg.de
[info\(at\)hs-augsburg.de](mailto:info(at)hs-augsburg.de)

Fakultät für Informatik
Telefon +49 821 55 86-3450
Fax +49 821 55 86-3499

Verfasser der Diplomarbeit
Max Mustermann
Beispielstraße 31
86150 Augsburg
Telefon +49 821 55 86-3450
max@hs-augsburg.de

Inhaltsverzeichnis

1	Introduction	3
2	Formulierung von Problemen und Lösungen in der Symbolischen Informationsverarbeitung	5
2.1	Typische KI-Problemstellungen	5
2.2	Problemlösung mit KI	5
2.2.1	Schritte um Probleme zu lösen	5
2.2.2	Performanzmaß berechnen	6
2.3	Beispielformulierungen von Zielen und Problemen	6
2.3.1	8er Puzzle (Sliding block puzzle)	6
2.3.2	Staubsauger-Roboter	7
2.4	Klassifikation von Problemen	8
2.5	Rationaler autonomer Agent als Problemlöser	9
2.5.1	Einfaches Beispiel für einen rationalen autonomen Agenten	10
2.5.2	Arten von rationalen Agenten	11
3	Problemlösung als Suchaufgabe	15
3.1	Wegsuche ohne Karte	15
3.1.1	Bug Algorithmen Beispiel	15
3.1.2	Problem mit dem Bug-Algorithmus	16
3.2	Repräsentation von Suchräumen	16
3.2.1	Suchraum als Karte	16
3.2.2	Charakterisierung von Suchproblemen in Graphen	17
3.3	Wegsuche als systematisches Ablaufen von Graphen	17
3.3.1	Expliziter Graphen und sukzessive Expansion	17
3.3.2	Generelle Wegfindungsstrategie	17
3.3.3	Generelle Bewertung von Suchverfahren	17
3.4	Arten von Suchverfahren	18
3.4.1	Breitensuche in expliziten Graphen	18

3.4.2	Uniforme Kostensuche	18
3.4.3	Modifizierte Uniforme Kostensuche / Dijkstra	19
3.4.4	Ablauf eines Graphen mit Tiefensuche	19
3.4.5	Tiefensuche und Backtracking-Algorithmen	20
3.4.6	Limitierte Tiefensuche	20
3.4.7	Iterative Tiefensuche	20
3.4.8	Bidirektionale Suche	20
3.5	KI-Suchverfahren	21
3.5.1	Greedy Search	21
3.5.2	Der A* Algorithmus	21
3.5.3	Pfadplanung in Computerspielen	22
4	Suchverfahren für Strategiespiele	23
4.1	MiniMax-Algorithmus	23
4.1.1	Ablauf des MiniMax-Algorithmus	23
4.1.2	Eigenschaften des MiniMax-Algorithmus	24
4.1.3	Verbesserung des MiniMax-Algorithmus durch Pruning	24
4.1.4	Komplexität von MiniMax-Pruning	25
4.1.5	MiniMax in Spielen mit Zufälligkeit	25
4.2	Monte Carlo Tree-Search	25
4.2.1	Ablauf des MCTS-Algorithmus	25
5	Constraints	27
5.1	Lösen von Constrain-Problemen	27
5.1.1	Ansätze zur Lösung von Constraint Solving Problemen	27
5.1.2	Optimierung der Lösung von Constraint Problemen	28
5.1.3	Constraint Propagation	30
6	Wissensbasierte Systeme	31
6.1	Wissensdarstellung in wissensbasierten Systemen	31
6.2	Schnittstelle zur Interaktion mit einem Wissenbasierten System	32
6.3	Ansätze zur Wissensverarbeitung	33
6.4	Aufgaben beim Entwurf eines Wissenbasierten Systems	34
7	Regelbasierte Wissensverarbeitung und Expertensysteme	35
7.1	Aufbau eines regelbasierten Wissensbasierten Systems	35
7.2	Zustandsraumdarstellung eines Wissensverarbeitungsproblems	36
7.3	Kompilierung von Regelbasen	37
7.4	Konfliktauflösung von Regelsystemen	37
7.5	Aufgaben bei der Entwicklung eines Expertensystems	38

8	Formaler Logik	39
8.1	Wissensrepräsentation und wissensbasierte Verarbeitung auf der Grundlage formaler Logik	39
8.1.1	Syntax und Semantik einer Logiksprache	39
8.1.2	Begriffe der Aussagenlogik-Syntax	40
8.1.3	Belegung, Modell und Erfüllbarkeit	40
8.2	Lösen von SAT-Problemen mit SAT-Solver	40
8.3	Ausdrucksstärke Logiksprachen	41
8.3.1	Prädikatenlogik 1. Stufe	41
8.3.2	Semantik der Prädikatenlogik	42

1. Introduction

2. Formulierung von Problemen und Lösungen in der Symbolischen Informationsverarbeitung

Um Probleme mit Hilfe von KI zu lösen, müssen sie zunächst in einer Weise dargestellt werden, die von Computern verarbeitet werden kann. Dies kann mit herkömmlichen Programmiersprachen über symbolische Informationsverarbeitung geschehen

2.1 Typische KI-Problemstellungen

Viele Probleme können mit Hilfe von KI gelöst werden. Die häufigsten sind:

- **Navigation** z.B: Labyrinth/Navigationsspiele, autonomer Staubsauger, Wegplanung
- **Strategiespiele** z.B: Brettspiele, Puzzle
- **Komplexe Aufgaben** z.B: Robocup (Navigation + Strategie)

2.2 Problemlösung mit KI

2.2.1 Schritte um Probleme zu lösen

1. Zielformulierung:

- Soweit möglich, Plausibilitäts-Check dabei durchführen: Ist das Ziel machbar?
- **Beispiel:** Hans will von A nach B, kennt aber den Weg nicht.

2. Problemformulierung

- Ausgangssituation formulieren.
- feststellen welche Operationen möglich sind (z.B Spielregeln).
- **Beispiel:** Durch ausführen von Fahr-Aktionen von A über verbundene Nachbarorte nach B kommen. Mögliche Operationen wären: in die benachbarten Städte zu fahren.

3. Konstruktion einer Lösung

- bewerte Güte einer Lösung
- wähle effektiven Lösungsweg
- **Beispiel:** Ein möglicher Weg zur Lösung des Problems wäre die Erstellung eines Suchbaums.

4. Ausführung

- Läuft alles wie geplant?

2.2.2 Performanzmaß berechnen

- Oft gibt es mehrere zulässige Lösungswege zu einem Probleme
- Wie findet man die optimalste Lösung?
- Zur Bewertung der **Güte** einer Lösung berechnet man die Gesamtkosten

$$\text{Gesamtkosten} = \text{Suchkosten} + \text{Pfadkosten}$$

- Es ist oft schwierig, die Güte einer Lösung zu verrechnen, da es oft viele mögliche Aspekte gibt, die beobachtet und gemessen werden können.
- Manchmal ist es besser, die weniger optimale Lösung zu wählen, die schneller berechnet werden kann: Genauere Planung kann mehr Zeit kosten als sie erspart!

2.3 Beispielformulierungen von Zielen und Problemen

2.3.1 8er Puzzle (Sliding block puzzle)

- Hochgradig kombinatorisches NP-vollständiges Problem. Oft genutzt als Standardtest für neue Suchalgorithmen.

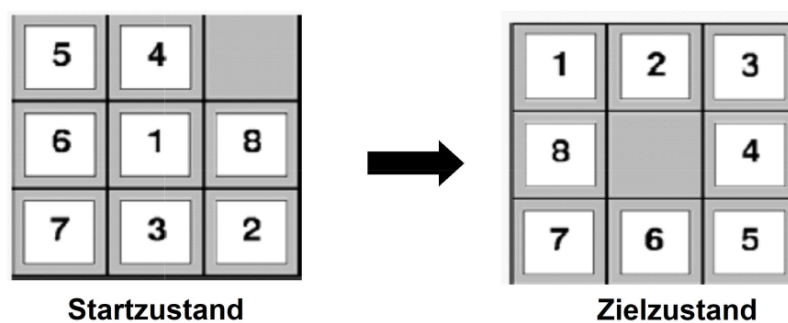


Abbildung 2.1: 8er puzzle Start- und Zielzustand

- Zustände: Lokalität der 8 Fliesen in eine der 9 Flächen plus eine freie Kachel
- Operatoren: Blank nach Links, Rechts, Auf, Ab
- Ziel-test: Blank-Kachel in der Mitte
- Pfadkosten: jeder Schritt kostet eine Einheit

2.3.2 Staubsauger-Roboter

Vieles an der Implementierung dieses Roboters muss abstrahiert werden:

- World States: Umfassen alle Aspekte der realen Welt
- Problem States: Nur Aspekte, die relevant für das Problem sind. Die Modellierung von diesen Aspekten erfolgt meist in Form **symbolischer** Beschreibungen.
- Zunächst müssen die möglichen World States als Problem States dargestellt werden:

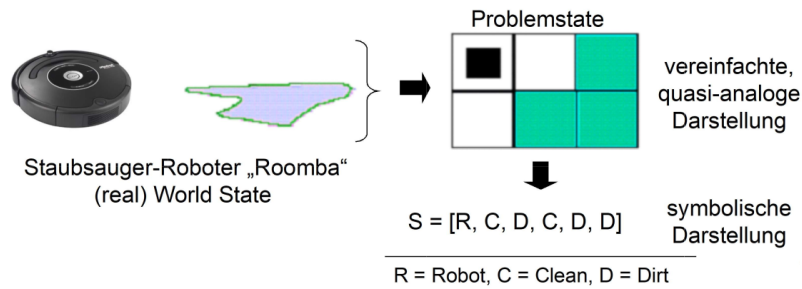


Abbildung 2.2: Abbildung World States auf Problem States

2.3.2.1 Stark vereinfachte Staubsaugerwelt

Eine sehr einfache Darstellung von einer Staubsaugerwelt hat zwei Orte. Jeder Ort kann entweder Staub enthalten oder nicht. Es gibt also 8 mögliche Zustände:



Abbildung 2.3: Staubsauger Zustände

Der Staubsauger kennt in diesem Fall 3 Operationen: Links, Rechts, Saugen und hat das Ziel, die Zustände S7 oder S8 zu erreichen, wo es keinen Staub mehr gibt.

In diesem einfachen Fall können die Lösung mit Hilfe eines endlichen Automaten gefunden werden:

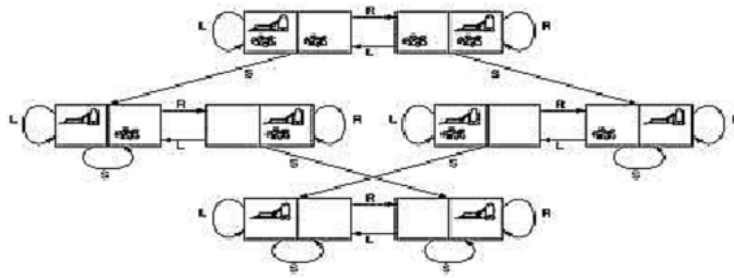


Abbildung 2.4: Staubsauger Pfadsucheautomat

2.4 Klassifikation von Problemen

Es gibt viele Arten von Problemen und sie werden durch die Informationen des Löser bestimmt.

1. Probleme mit Einfach Zuständen

Für den Problemlöser ist klar, in welchem Zustand er sich befindet und was seine möglichen Aktionen bewirken werden. Wie bereits im Roomba-Beispiel (Kap. 2.3.2) erwähnt, kann dies als endlicher Zustandsautomat modelliert werden.

2. Probleme mit Mehrfach-Zuständen

Der genaue Zustand, in dem sich der Problemlöser befindet, ist nicht bekannt, und der Problemlöser weiß nicht, was seine Aktionen bewirken werden.

Anhand des Roomba-Beispiels: In dem Extremfall dass der Roomba keine Sensoren hat, kann er als mehrfaches Problem modelliert werden. In einem solchen Fall kann der Startzustand einer der Zustände S1 bis S8 sein (siehe Abb. 2.3).

Eine mögliche Lösung besteht darin, die Menge der aktuell möglichen Zustände zu verwenden, um die Aktionen des Löser zu bestimmen:

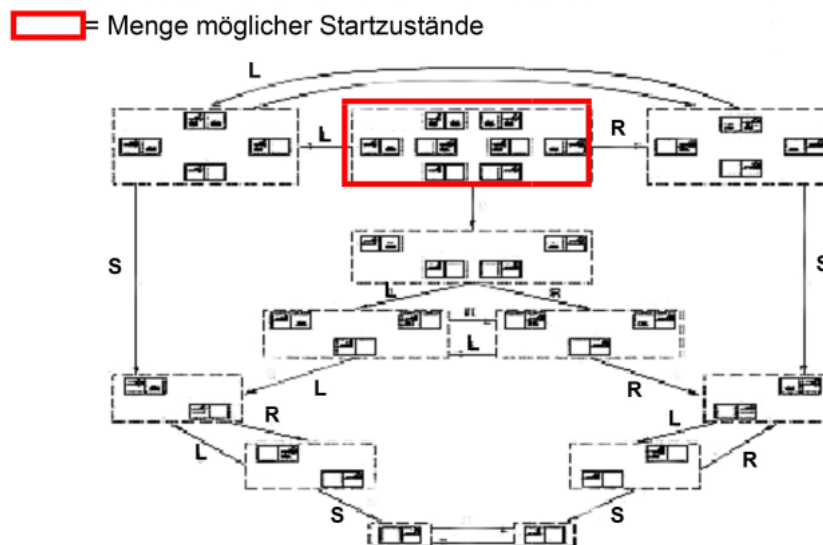


Abbildung 2.5: Staubsaugerwelt als Mehrfach-Problem

3. Zufall-Probleme

Der L?ser hat keine vollst?ndige Kenntnis einer sich st?ndig ver?ndernden Welt. Er kann nur die lokale Umgebung wahrnehmen.

Auch hier wieder das Roomba-Beispiel: Der Roboter befindet sich im Zustand S1 oder S3 und kann die Aktionen: Saugen, nach rechts fahren, Saugen ausf?hren.

Der Roboter saugt und bewegt sich dann nach rechts. Wenn er aber sich vor der Bewegung im Zustand S3 befand, geht er zum Zustand S8 ?ber. Er versucht dann zu saugen aber die Aktion schlagt fehl, da in dem Zustand kein Staub vorhanden ist.

Um dieses Problem zu l?sen, ben?tigt der Roboter einen Sensor, der das Vorhandensein von Staub erkennt.



Abbildung 2.6: Roomba Zufallproblem

4. Explorations-Probleme

Der L?ser hat keine Kenntnis von der Welt und muss seine Umgebung erkunden, um die m?glichen Zust?nde und die Auswirkungen seiner Aktionen zu erfahren.

Ein gutes Beispiel daf?r ist der Marsrover. Der Rover muss zun?chst Daten sammeln, um seine Umgebung kennenzulernen und eine Karte zu erstellen. Mit dieser Karte kann es dann erfolgreich die Pfadfindung durchf?hren.

2.5 Rationaler autonomer Agent als Probleml?ser

Ein rationaler autonomer Agent ist ein Wesen, das seine Welt permanent wahrnehmen und unabh?ngig reagieren kann.

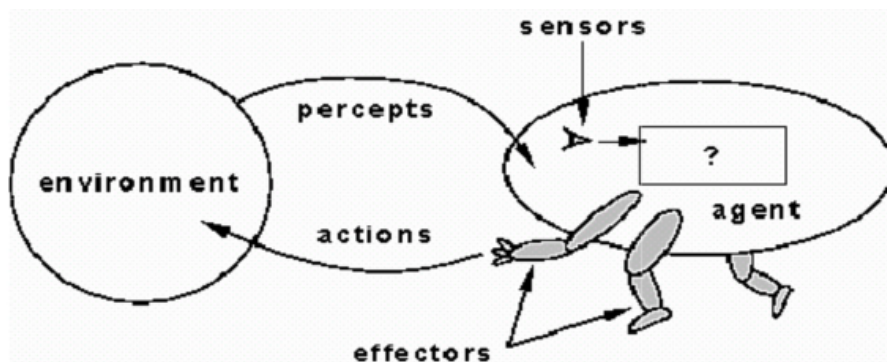


Abbildung 2.7: Rationaler autonomer Agent

Die Abbildung 2.7 veranschaulicht die Interaktion zwischen einem Agenten und seiner Umgebung unter Verwendung seiner Perzeptoren (PAGE = Percepts, Actions, Goals, Environment).

2.5.1 Einfaches Beispiel für einen rationalen autonomen Agenten

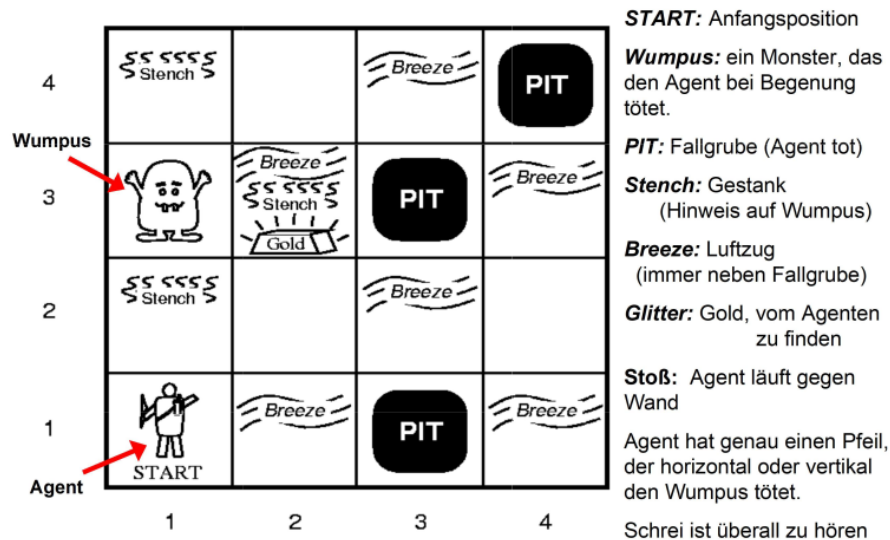


Abbildung 2.8: Wumpus World

Wumpus World ist eine Höhle mit 4/4 Räumen, die durch Gängen verbunden sind. Es sind also insgesamt 16 Räume miteinander verbunden. Es gibt einen wissensbasierten Agenten, der durch diese Welt gehen wird. Die Höhle hat einen Raum mit einem Monster namens Wumpus, das jeden frisst, der den Raum betritt. Wumpus kann vom Agenten erschossen werden, aber der Agent hat einen einzigen Pfeil. In der Welt von Wumpus gibt es mehrere endlose Lochräume und wenn ein Agent in ein Loch fällt, wird er dort für immer gefangen sein. In einem der Räume dieser Höhle befindet sich ein Haufen Gold. Das Ziel des Agenten ist es also, das Gold zu finden und aus der Höhle zu kommen, ohne in das Loch zu fallen oder von Wumpus gefressen zu werden. Der Agent wird belohnt, wenn er mit Gold herauskommt, und er bekommt eine Strafe, wenn er von Wumpus gefressen wird oder in ein Loch fällt.

Umgebung von Wumpus World

- Ein 4x4-Raster von Räumen.
- Der Agent beginnt im Feld [1,1] und zeigt nach rechts.
- Standort von Wumpus und Gold werden zufällig ausgewählt, außer Feld [1,1].
- Jedes Quadrat der Höhle kann mit Wahrscheinlichkeit 0,2 eine Grube sein, außer dem ersten Quadrat

Eigenschaften von Wumpus World

- Die Welt ist teilweise beobachtbar, da der Agent nur die nahe Umgebung wahrnehmen kann, in der er sich befindet.
- Die Welt ist deterministisch, da die Auswirkungen aller Handlungen bekannt sind.
- Die Reihenfolge ist wichtig, also ist die Welt sequentiell.
- Die Welt ist statisch wie Wumpus und die Gruben bewegen sich nicht.

- Die Umgebung ist diskret.
- Einzelagentenumgebung (Wumpus wird nicht als Agent betrachtet, da er statisch ist).

2.5.2 Arten von rationalen Agenten

Reflex Agenten

Ein Reflexagent ist eine Entit?t, die reaktiv ist und mit Sensoren arbeitet. Dieser Agent setzt sich keine Ziele und k?mmert sich nicht um die Auswirkungen seiner Handlungen.

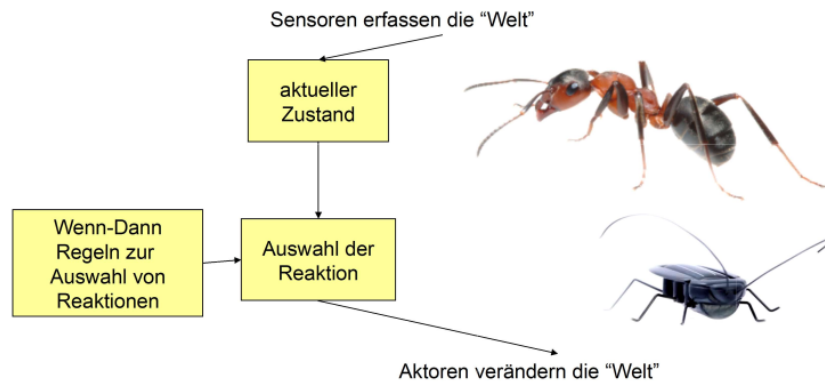


Abbildung 2.9: Reflex-Agenten

Ziel orientierter Agent

Zielorientierte Agenten planen ihre Handlungen und antizipieren die Auswirkungen dieser Handlungen, um ihre Ziele zu erreichen.

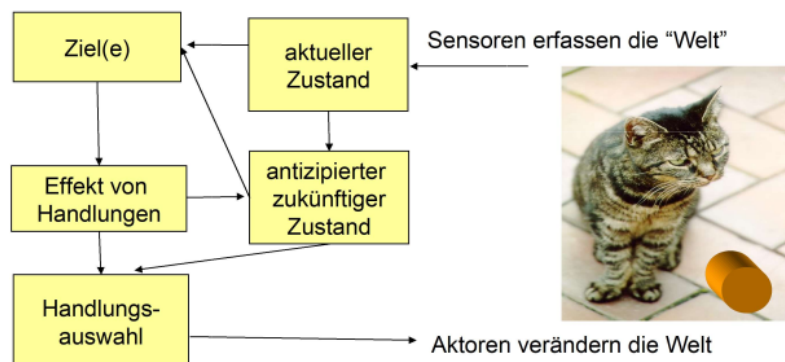


Abbildung 2.10: Ziel-orientierter Agenten

Nutzen-orientierter Agent

Ein nutzen-orientierter Agent wägt die Kosten und Gewinne von Handlungen ab, um die Effektivität seiner Handlungen beim Erreichen eines Ziels zu maximieren.

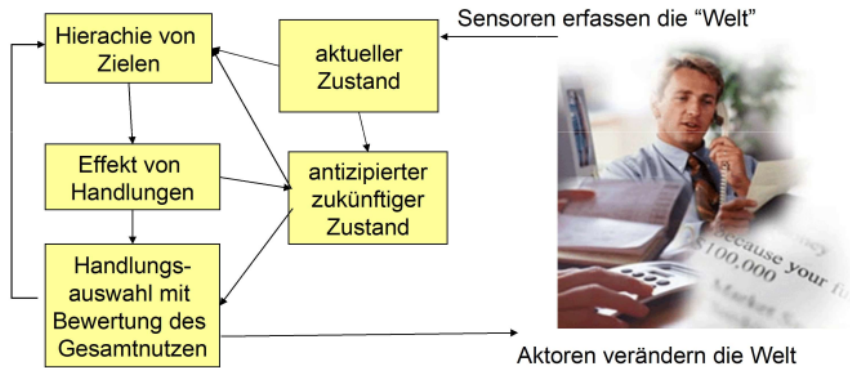


Abbildung 2.11: Nutzen-orientierter Agenten

Lern-fähiger Agent

Diese Art von Agent ist in der Lage, die Effekte seiner Handlungen zu lernen, um seine Ziele besser zu erreichen, indem er seine Handlungen effektiver auswählt.

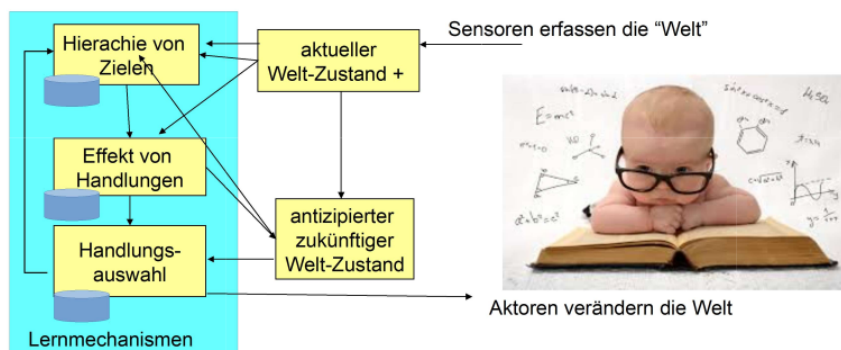


Abbildung 2.12: Lern-fähiger Agenten

Emotionaler Agent

Dieser Agent verwendet neben rationalem Denken eine emotionale Bewertung, um reale menschliche Entscheidungsprozesse besser zu simulieren.

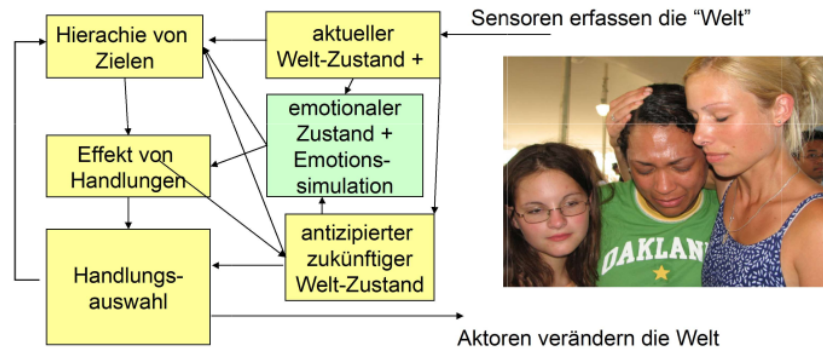


Abbildung 2.13: Emotionaler Agent

Kreativer Agent

Dieser Agent hat die Fähigkeit, neue Aktionen zu entwickeln, um neue Wege zur Lösung von Problemen zu schaffen, die nicht vorprogrammiert waren.

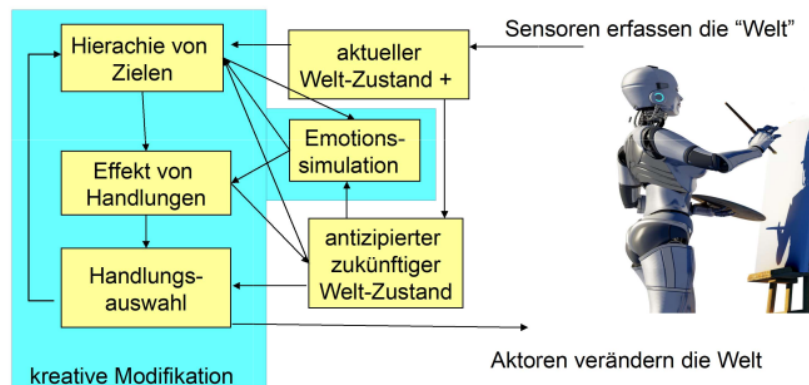


Abbildung 2.14: Kreativer Agent

Kooperative Agenten

Eine Situation, in der mehrere Agenten zusammenarbeiten, um komplexe Aufgaben zu lösen. Dies erfordert die Modellierung kooperativer Verhaltensstrategien und das Verständnis der Absichten anderer Agenten.



Abbildung 2.15: Robocup competition: Cooperative agents

3. Problemlösung als Suchaufgabe

3.1 Wegsuche ohne Karte

“Wegsuche ohne Karte” bedeutet Wegfindung in einer unbekannten Umgebung ohne eine Karte, die den Agenten leitet.

Beispiel: **Roboter R** befindet sich in einem unbekannten Gebiet und muss sich zu einem Zielobjekt bewegen. Dies kann durch Anwendung eines **Bug-Algorithmus** gelöst werden, der voraussetzt, dass der Roboter mit Sensoren ausgestattet ist, um Hindernisse und das **Zielobjekt S** zu erkennen.

3.1.1 Bug Algorithmen Beispiel

Ein Beispiel für einen Bug-Algorithmus ist wie folgt:

1. Wenn das **Zielobjekt S** in Sichtweite ist, fährt **Roboter R** direkt darauf zu
2. Wenn **S** nicht in Sicht ist, aber stattdessen ein Hindernis vorhanden ist, bewegt sich **R** gemäß einer bestimmten Regel um das Hindernis herum (z. B. im Uhrzeigersinn).
3. **R** scannt erneut nach dem Objekt **S** und wiederholt die Schritte 1 und 2, bis das Ziel erreicht ist.

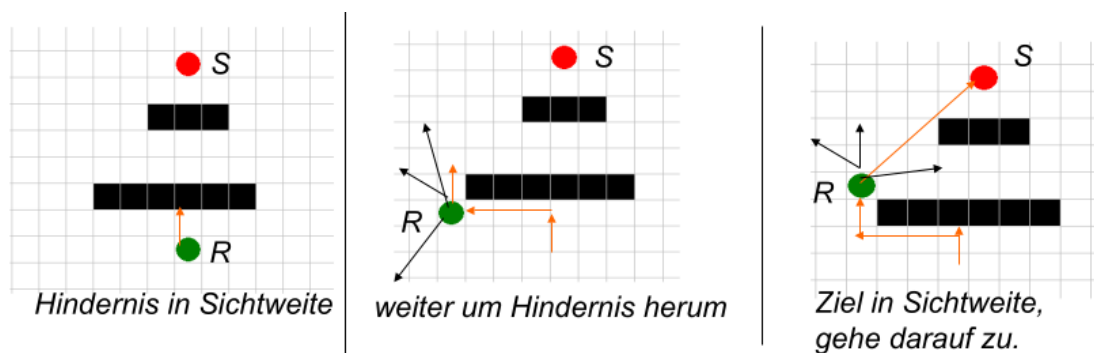


Abbildung 3.1: Beispiel von Bug-Algorithmus Verfahren

3.1.2 Problem mit dem Bug-Algorithmus

In bestimmten Situationen (z.B. siehe Abb. 3.2) ist der Roboter mit dem Ansatz in Abschnitt 3.1.1 nicht in der Lage, das Ziel zu finden.

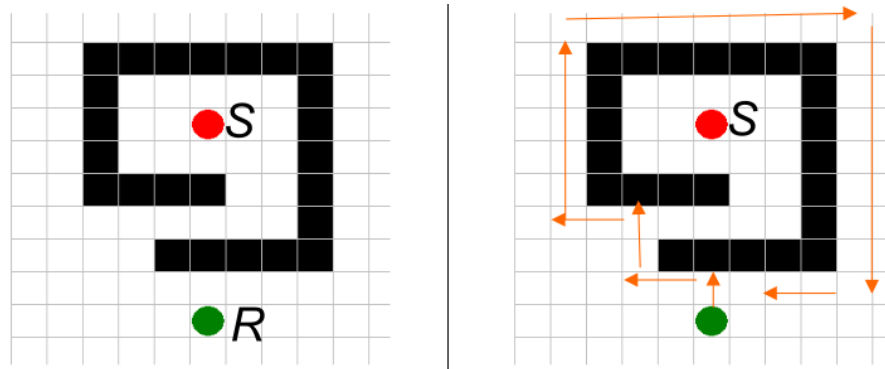


Abbildung 3.2: Der Roboter kann das Ziel nicht sehen

Der Algorithmus muss verbessert werden, z. B. durch Bewegen gegen den Uhrzeigersinn, um eine Bewegung in einem kontinuierlichen Kreis zu vermeiden.

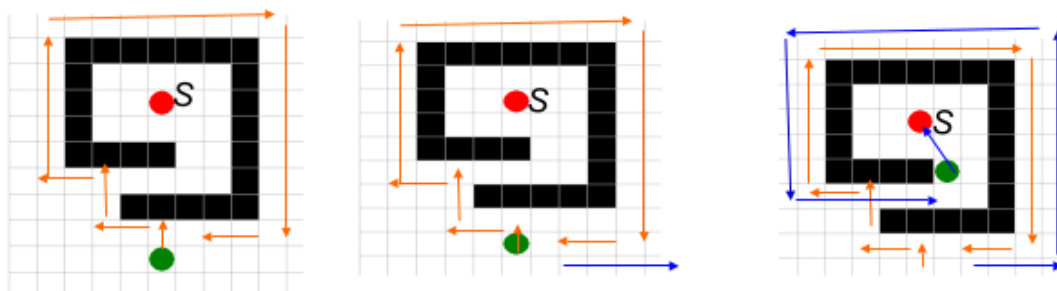


Abbildung 3.3: Gegen den Uhrzeigersinn bewegen, um das Ziel zu erreichen

3.2 Repräsentation von Suchräumen

3.2.1 Suchraum als Karte

Ein Suchraum wird normalerweise als grafische Karte dargestellt. Diese Karten können als Wegenetz oder als Gitter mit benachbarten Zellen dargestellt werden, wie in Abbildung 3.4 gezeigt.

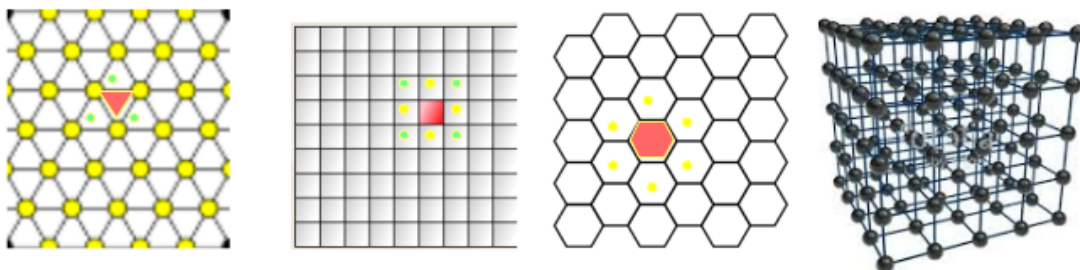


Abbildung 3.4: Beispiele von Gittermustern

Je nach verwendetem Gittermuster werden unterschiedliche Suchgraphen basierend auf der Anzahl der Nachbarn jeder Zelle im Gitter gebildet. Zum Beispiel: Eine dreieckige Zelle hat sechs Nachbarn, eine quadratische Zelle hat 4 (oder 8, wenn Diagonalen erlaubt sind) und eine sechseckige Zelle hat 6.

3.2.2 Charakterisierung von Suchproblemen in Graphen

Um ein Suchproblem in einem Graphen darzustellen, ist es wichtig, einen **Startzustand S**, einen gewünschten **Endzustand Z** und die **möglichen Aktionen zu definieren, um von einem Zustand zum anderen zu gelangen**.

Um diese Suchproblemlogik anzuwenden, in einem gerichteten Graphen G :-

- Zustände: Knoten im Graph G
- Anfangs- und Zielzustand: ausgezeichnete Knoten in G
- Zustandsübergänge: Entspricht dem Passieren von Kanten in G

3.3 Wegsuche als systemstisches Ablaufen von Graphen

Bei der Wegfindung mit Hilfe eines Graphen müssen ein **Startknoten** und eine **Funktion zum Testen**, ob der Zielknoten erreicht wurde, definiert werden. Mit Hilfe des Startknotens und dieser Funktion kann eine Folge von Knoten gefunden werden, die die Testfunktion erfüllen können. Wichtig ist, dass die für die Sequenz ausgewählten Knoten **benachbarte Knoten sind, die durch Kanten verbunden** sind.

3.3.1 Expliziter Graphen und sukzessive Expansion

Es ist wichtig, zwischen einem expliziten Graphen und einem “gedachten” Graphen als Repräsentation eines Suchraums für Lösungen zu unterscheiden.

Expliziter Graphen

Eine explizite Darstellung der Knoten und Kanten ist verfügbar (z.B Adjazentmatrix). Dies ist nur für kleinere Knotenmengen praktikabel und die meisten Probleme erfordern keine vollständige Repräsentation des Suchgraphen.

Sukzessive Expansion

Diagramm ist verfügbar, aber nicht vollständig explizit. Der Suchprozess erfolgt durch Knotenerweiterung, indem Knoten für Knoten vorgegangen wird.

3.3.2 Generelle Wegfindungsstrategie

Zyklen und Mehrfachbesuche sind zu vermeiden, da sie den Suchbaum exponentiell wachsen lassen können. Generell sollte man keinen bereits benutzten Weg nochmal gehen, keine Wege mit Zyklen kreieren, einen bereits besuchten oder ausgebauten Zustand nicht nochmal besuchen oder erzeugen.

3.3.3 Generelle Bewertung von Suchverfahren

Bewertungskriterien eines Pfadfindungsprozesses sind wie folgt:

Korrektheit

Es ist wichtig, dass die Wegfindungslösung tatsächlich eine Lösung des Problems ist.

Vollständigkeit

Existiert eine Lösung, terminiert der Algorithmus nach endlicher Zeit und generiert eine Lösung.

Optimalität

Die optimalste Lösung wird gefunden, wenn mehrere möglich sind.

Zeitkomplexität

Die Zeit, die im worst-case/average-case benötigt wird, um eine optimale Lösung zu finden.

Speicherkomplexität

Wie viel Speicher, die im worst-case/average-case benötigt wird.

3.4 Arten von Suchverfahren

Es gibt viele Möglichkeiten, eine Wegfindungssuche durchzuführen. Diese nächsten Abschnitte werden darauf eingehen.

3.4.1 Breitensuche in expliziten Graphen

Breitensuche ist auf Englisch “breadth first search”. Für diesen Algorithmus werden für jeden Knoten zusätzliche Daten benötigt: **noch nicht gesucht** (weiß dargestellt), **entdeckt, aber noch nicht verarbeitet** (grau dargestellt), **verarbeitet** (schwarz dargestellt).

Überblick über den Ablauf:

1. Wähle einen Startknoten aus dem Graphen und legen Sie diesen in die Warteschlange, Q. Alle Knoten in Q sind grau markiert.
2. Solange es Elemente in Q gibt, markiere alle Nachbarn der Elemente grau und füge diese der Warteschlange Q hinzu.
3. Nachdem alle Nachfolger eines Knotens grau markiert wurden, markiere den Knoten schwarz und entferne ihn aus der Warteschlange. Prüfen Sie, während Sie die Nachfolgerknoten grau markieren, ob der zu markierende Knoten der Zielknoten ist.
4. Wiederhole Schritt 2 und Schritt 3, bis die Warteschlange Q leer ist.

Komplexitätsbetrachtung

Speicherbedarf und Zeitbedarf sind exponentiell. Dies führt dazu, dass große Graphen unlösbar sind und sogar relativ kleine Graphen zu lange brauchen, um praktikabel zu sein (z.B Graphen Tiefe 12 brauch 35 Jahre Rechenzeit).

3.4.2 Uniforme Kostensuche

Bei diesem Suchalgorithmus wird die Breitensuche so verändert, dass auch die **Kosten der Nachbarknoten** berücksichtigt werden. Da die Kosten zweier benachbarter Knoten normalerweise bereits bekannt sind, kann **die Warteschlange in einen Heap umstrukturiert werden**, der in **aufsteigender Reihenfolge der Pfadkosten** sortiert ist. Auf diese Weise wird gehofft, dass der kürzeste Weg vom Startknoten zum Zielknoten gefunden wird.

3.4.3 Modifizierte Uniforme Kostensuche / Dijkstra

Um eine optimale Lösung für ein Wegsucheproblem zu finden, ist es notwendig, alternative Pfade parallel zu konstruieren. Wenn sich diese Pfade am selben Punkt treffen, kann der **kürzeste Pfad beibehalten werden**, während der Rest aus dem Suchbaum entfernt wird, um die kürzeste Gesamtroute zu finden.

Wenn ein Pfad zum Zielknoten gefunden wird, werden alle **anderen parallelen Pfade fortgesetzt**, es sei denn, ihre Kosten übersteigen den bereits gefundenen Pfad. Dies führt zu einer besseren Effizienz bei der Suche nach dem kürzesten Weg.

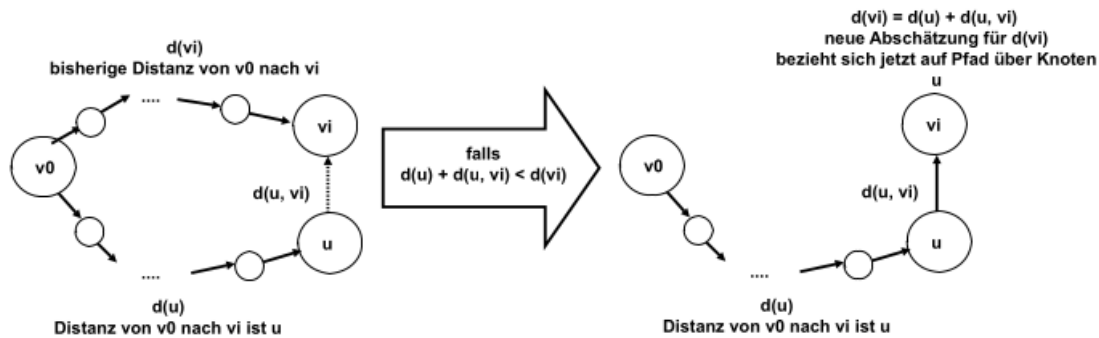


Abbildung 3.5: Dijkstra-Algorithmus Pfadsuche

Dieser Algorithmus ist in Abbildung 3.5 demonstriert. Anhand der Abbildung kann man sehen, wie die beiden Routen von v_0 nach vi verglichen werden und die Kosten verglichen werden, um nur die kürzere Route beizubehalten.

3.4.4 Ablauf eines Graphen mit Tiefensuche

Im Gegensatz zur Traversierung mit Breitensuche verläuft die Traversierung mit Tiefensuche auf möglichst langen Wegen und führt die Weiß-Grau-Schwarz-Markierung (siehe Abschnitt 3.4.1) durch. Dieser Algorithmus wird normalerweise rekursiv implementiert.

Ablauf von Tiefensuche

1. Wähle einen Startknoten v_s . Markiere den Knoten als grau ein.
2. Für jeden Nachbarknoten v_k von v_s , der weiß markiert ist, führe folgendes durch:
 - (a) Füge v_k zum Stack hinzu
 - (b) Markiere v_k als grau
 - (c) Falls v_k Nachbarn hat, führe Schritt (2) rekursiv aus
 - (d) Wenn v_k keine Nachbarn hat oder seine Nachbarn schwarz markiert sind, markiere v_k schwarz und entferne v_k vom Stack
3. Der Algorithmus endet, sobald der Stack leer ist.

Komplexitätsbetrachtung

Nur benachbarte Knoten des aktuellen Suchpfads werden auf dem Stack gespeichert. Dies bedeutet viel weniger Knotenerweiterungen im Vergleich zur Breitensuche, was zu weniger Speicherverbrauch führt. Die Zeitkomplexität ist jedoch ebenso wie die Breitensuche exponentiell.

3.4.5 Tiefensuche und Backtracking-Algorithmen

In realen Anwendungen beinhaltet eine Lösung die Verwendung vieler verschiedener Komponenten. Die endgültige Lösung verwendet daher normalerweise viele Teillösungen, die möglicherweise nicht richtig erweitert werden können (z.B: erreicht die Teillösung eine Sackgasse). In diesen Fällen muss die Teillösung durch eine weniger komplexe Lösung ersetzt werden.

Genereller Ablauf von einem Backtracking-Algorithmus

Gegeben sind ein Array “SolutionComponents”, das alle Werte der endgültigen Lösung enthält, und die Funktionen “FirstTrialValue()”, “NextTrialValue()” und “CheckValid()”.

1. Der ersten Komponente wird der erste Versuchswert gegeben:

```
SolutionComponents[0] = FirstTrialValue(0)
```

2. Die Gültigkeit der Lösung wird überprüft:

```
CheckValid(SolutionComponents)
```

3. Wenn die bisherige Lösung gültig ist, fahren Sie mit dem nächsten Wert fort (i = 1, 2, 3...) und überprüfe den Wert erneut

```
SolutionComponents[i] = FirstTrialValue(i)  
CheckValid(SolutionComponents)
```

4. Wenn der CheckValid-Funktion an einem bestimmten Punkt false zurückgibt, versuche es mit den nächsten Werten. Wenn alle Werte ebenfalls falsch zurückgeben, führe ein Backtracking durch, indem Sie i um i reduzieren und den nächsten Wert für dieses i versuchen. Erhöhen Sie dann i wieder um 1 und probieren Sie alle Werte aus.
5. Schritt 4 wird so oft wie nötig wiederholt. Für den Fall, dass i null erreicht und es keine Versuchswerte mehr für i = 0 gibt, gibt es keine mögliche Lösung.

3.4.6 Limitierte Tiefensuche

Es ist möglich, die Nachteile der Tiefensuche zu vermeiden, indem man eine maximale Tiefe des Pfades einstellt. Das macht die Suche vollständig aber nicht immer optimal. Um diese Idee zu erweitern, kann eine iterative Tiefensuche versucht werden.

3.4.7 Iterative Tiefensuche

Bei dieser Methode wird die Tiefe mit jeder Iteration erhöht, um sicherzugehen, dass eine Lösung gefunden wird.

3.4.8 Bidirektionale Suche

Anstatt nur vom Startknoten aus zu beginnen, führen Sie eine Suche vom Start- und vom Zielknoten aus durch. Wenn sich die beiden Verfahren in der Mitte treffen, ist eine Lösung gefunden.

3.5 KI-Suchverfahren

Es gibt zwei Klassen von Suchverfahren: **blinde Suchverfahren** und **KI-Suchverfahren**. Blinde Suchverfahren sind auf einem bestimmten Schema basiert, das unabhängig von dem jeweiligen Problem ist. Einige Beispiele hierfür sind die in den vorangegangenen Kapiteln behandelten Verfahren wie Breitensuche, Tiefensuche, Bidirektionale Suche usw.

KI-Suchverfahren hingegen nutzen problemspezifisches Vorwissen zur Eingrenzung des Suchraums. Es handelt sich um informierte heuristische Suchverfahren.

Blinde Suchverfahren erfordern, dass eine Lösung durch systematische und erschöpfende Suche in einem Suchgraphen gefunden wird, was ineffizient und kein problemspezifisches Wissen nutzt.

Informierte Suchprozesse hingegen nutzen problemspezifische Eigenschaften, um die Effizienz der Knotenexpansion zu verbessern.

3.5.1 Greedy Search

Die Greedy-Suche ist eine modifizierte Breitensuche (siehe Abschnitt 3.4.1), bei der nur die Knoten mit den geringsten Kosten in die Warteschlange aufgenommen werden.

Ablauf von Greedy Search

Wenn die Kosten des aktuellen Knotens zum Zielknoten unbekannt sind, **müssen diese Kosten geschätzt werden**, und dann wird der Nachbar mit den geringsten Kosten ausgewählt. Die Funktion, die diese Kosten schätzt, wird **heuristische Funktion** genannt.

Der Unterschied zwischen Greedy Search und Uniform Cost Search (siehe Abschnitt 3.4.1) besteht darin, dass bei der Greedy Search **die Kosten von einem Knoten zum Zielknoten** berechnet werden und nicht von einem Knoten zum anderen.

Eigenschaften von Greedy Search

Greedy Search bietet tendenziell schnelle Lösungen, die oft, aber nicht immer, der optimale Weg sind.

Greedy Search ist ähnlich wie die Tiefensuche mit Backtracking, nicht vollständig, und erfordert eine gute Heuristik für eine bessere Güte des Verfahrens.

3.5.2 Der A* Algorithmus

Der A*-Algorithmus ist ein neuer Ansatz, der auf Greedy Search (Abschnitt 3.5.1) und dem Dijkstra-Algorithmus (Abschnitt 3.4.3) aufbaut. A* arbeitet basierend auf der Funktion:

$$f(n) = g(n) + h(n)$$

Wobei $f(n)$ die geschätzten Kosten der billigsten Lösung ist, $g(n)$ die Kosten für die Bewegung von der Ausgangszelle zur aktuellen Zelle, und $h(n)$ die geschätzten Kosten für die Bewegung von der aktuellen Zelle zur Zielzelle. Mit der Funktion $f(n)$ werden die Kosten berechnet, und der Rest des Algorithmus läuft wie bei einer Breitensuche ab. Um eine optimale Lösung zu erzielen, sollte die heuristische Funktion $h(n)$ die Entfernung zum Ziel nicht überschätzen.

Es ist auch möglich, die Gewichtung der Komponenten der A*-Kostenfunktion zu ändern:-

$$f(n) = \alpha * g(n) + \beta * h(n)$$

Wenn β viel größer als α ist, dann nähert sich die Suche einer Greedy Search. Andersrum ist es ähnlich wie bei einer Breitensuche.

3.5.3 Pfadplanung in Computerspielen

Pathing-Algorithmen werden in Spielen häufig zur Navigation von Figuren in der virtuellen Welt verwendet. In frühen 3D-Spielen wurde dies durch Wegpunkte erreicht. NPCs können sich nur durch Wegpunkte bewegen und der Pfad wird mit dem A*-Algorithmus (Abschnitt 3.5.2) berechnet (Abb. 3.6).

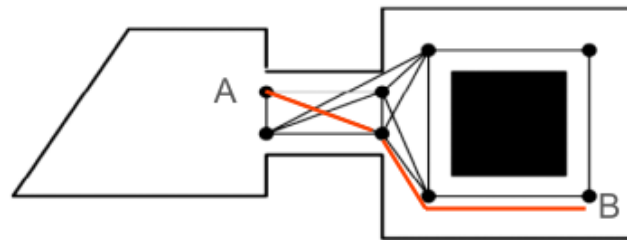


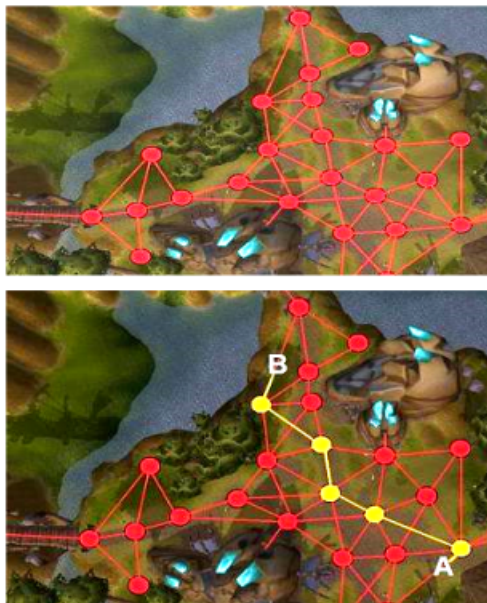
Abbildung 3.6: Spielnavigation mit Wegpunkten

Dieser Ansatz mit Wegpunkten schränkt jedoch die Bewegung der Spielfiguren stark ein. Dieser Ansatz mit Wegpunkten schränkt jedoch die Bewegung der Spielfiguren stark ein. Ein besserer Ansatz ist die Verwendung von Polygon-Netze.

Bereiche, in denen sich die Figuren frei bewegen können, werden als Polygone dargestellt, und diese Bereiche sind durch Knoten miteinander verbunden. Bei diesem Ansatz können sich die Figuren freier bewegen, und der Ansatz kann weiterhin mit bekannten Algorithmen wie A* berechnet werden.

● Wegpunkte

– wenig Bewegungsfreiheit



● Polygon-Meshes

– bessere Bewegungsfreiheit

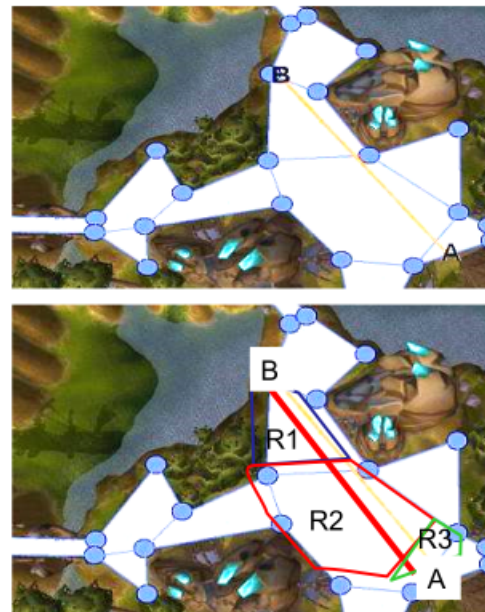


Abbildung 3.7: Spielnavigation mit Polygon-Meshes vs Wegpunkte

4. Suchverfahren für Strategiespiele

Es gibt viele Arten von Spielen, die eine KI erfordern, um gegen sie zu spielen, z. B. rundenbasierte Strategie- oder Kartenspiele, Strategiespiele mit einem Zufallselement (z. B. Würfel) und Spiele, die eine strategische Positionierung von Einheiten erfordern.

Im Gegensatz zu normalen Suchproblemen sind Spiele insofern einzigartig, als der Gegner unberechenbar ist und die Rechenleistung normalerweise begrenzt ist.

4.1 MiniMax-Algorithmus

MiniMax ist ein Algorithmus für rundenbasierte Nullsummenspiele. Bei diesem Algorithmus **maximiert** Spieler A seine Gewinnchancen, wobei er davon ausgeht, dass Spieler B versuchen wird, die Gewinnchancen von Spieler A zu **minimieren**.

4.1.1 Ablauf des MiniMax-Algorithmus

Der Algorithmus läuft wie folgt ab:

1. Erzeuge einen Suchbaum, wobei die Wurzel die Startposition des Spiels ist und das Ende des Baums die möglichen Endzustände des Spiels darstellt.
2. Berechne die Gewinnchancen für Spieler A für jeden Zweig von den Endzuständen zurück zum Ursprung berechnet.
3. Für jede Ebene im Suchbaum berechne der Wert für die Gewinnchancen von A eines Knotens aus den Werten seiner Nachfolgeknoten. Falls Spieler A am Zug ist, dann ist der Knotenwert der Maximum der Nachfolger, und falls Spieler B am Zug ist, dann ist der Knotenwert der Minimum der Nachfolger.
4. Nachdem alle Knoten berechnet wurden, wählt Spieler A den Weg, der ihm am meisten nützt.

Dies lässt sich am besten anhand eines einfachen Spiels namens “Nimm” demonstrieren (siehe Abb. 4.1). In diesem Spiel gibt es n Objekte. Die Spieler entfernen abwechselnd 1, 2 oder 3 der Objekte. Der Spieler, der das letzte Objekt nimmt, verliert.

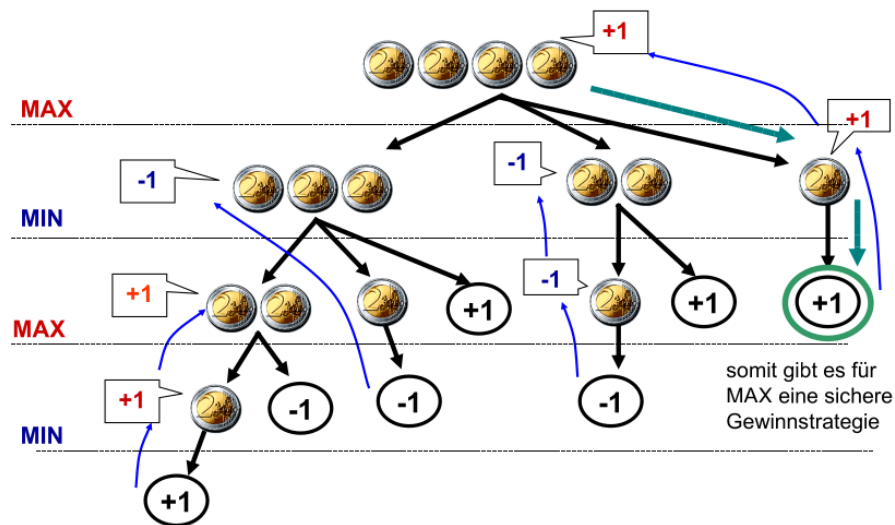


Abbildung 4.1: Bottom-up Bewertung der Knoten aus der Sicht von MAX

4.1.2 Eigenschaften des MiniMax-Algorithmus

Der Algorithmus ist vollständig, wenn der Suchbaum endlich ist, und ist optimal, wenn gegen einen optimalen Spieler gespielt wird. Der Algorithmus hat ein Zeitbedarf von $O(b^m)$ mit Suchtiefe m und Verzweigungsfaktor b und hat ein Platzbedarf von $O(b * m)$.

Dieser Algorithmus kann modifiziert werden, indem man die Tiefe des Baumes begrenzt (nur ein paar Runden vorausschaut) und den Nutzen einer Position mit einer guten Bewertungsfunktion abschätzt.

4.1.3 Verbesserung des MiniMax-Algorithmus durch Pruning

Pruning, auch bekannt als $\alpha\beta$ -Search oder $\alpha\beta$ -Pruning, ist eine Idee, um die Effizienz des Minimax-Algorithmus zu verbessern, indem frühzeitig erkannt wird, welchem Zweig des Suchbaums nicht gefolgt werden muss.

Unter der Annahme, dass MIN und MAX jeweils den für sie optimalen Weg wählen, ist es möglich frühzeitig zu erkennen, welche Verzweigungen des Suchbaums das Ergebnis nicht beeinflussen und daher nicht berechnet werden müssen. Unter Verwendung des Nimm-Beispiels können Pfade geschnitten werden, die ein schlechteres Ergebnis liefern als bereits erkundete Pfade.

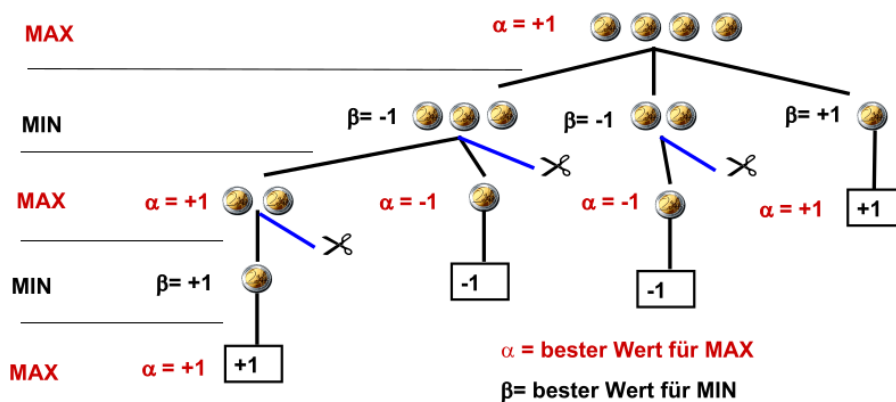


Abbildung 4.2: Beschneiden des MinMax-Nimm-Beispiels

4.1.4 Komplexität von MiniMax-Pruning

Im schlimmsten Fall, wenn die Knoten in einer ungünstigen Reihenfolge sind, dann ist es dasselbe wie bei einem normalen MiniMax. Im besten Fall könnte man die Mindestanzahl an Positionen berechnen lassen. Dies kann erreicht werden, indem die Reihenfolge der Bewegungen basierend auf der Effizienz der Bewegung neu geordnet wird. Zum Beispiel ist beim Schach das Schlagen einer Figur mit einem Turm oft vorteilhaft und sollte zuerst sequenziert werden.

4.1.5 MiniMax in Spielen mit Zufälligkeit

Die Zufälligkeit wird unter Verwendung zusätzlicher Knoten im Suchbaum gehandhabt. Für die Bewertung erhält jeder Zweig der Zufallsknoten den Durchschnitt aller möglichen Zufallsergebnisse.

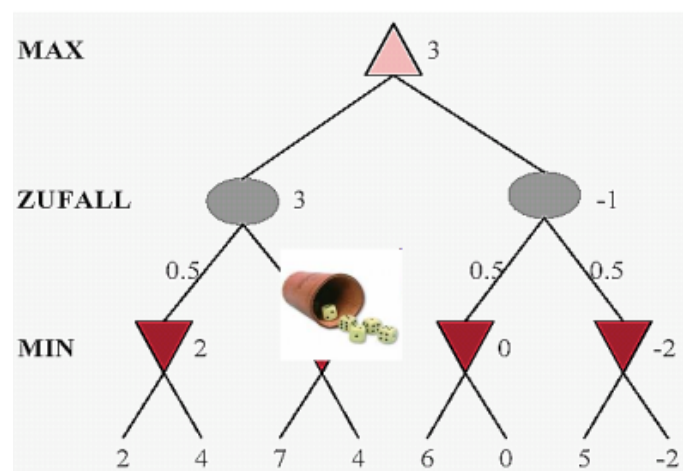


Abbildung 4.3: Zusätzlicher Zufallsknoten mit Mittelwert-Bewertung

4.2 Monte Carlo Tree-Search

In großen Suchräumen wie Schach oder Go ist der MiniMax-Algorithmus weniger geeignet, da zu viele Knoten erweitert werden müssten. Die Idee von Monte Carlo Tree Search (MCTS) ist, dass ein Suchbaum wie in MiniMax ebenfalls generiert wird, aber anstatt alle möglichen Positionen zu erweitern, werden die Züge durch zufällige Züge über (sehr) viele Spiele erweitert.

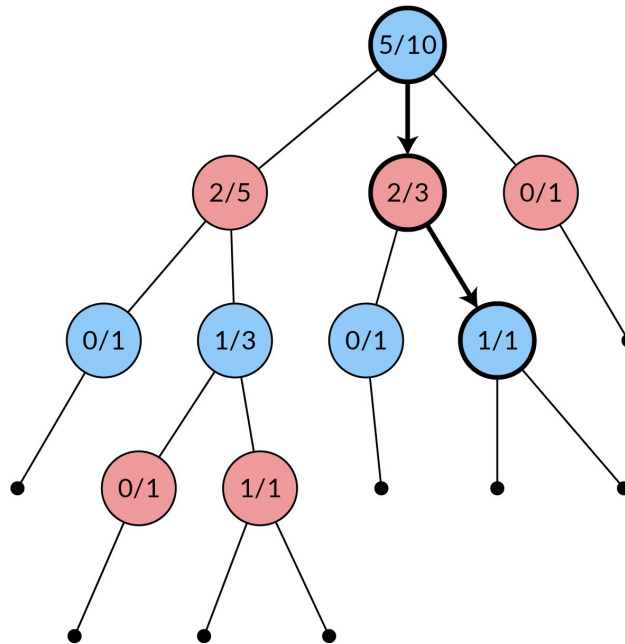
4.2.1 Ablauf des MCTS-Algorithmus

Die folgenden Schritte werden wiederholt durchgeführt:-

Die folgenden Schritte werden so lange wiederholt, bis der Algorithmus zum Abbrechen aufgefordert wird:

1. **Selection.** Max beginnt bei der Wurzel, um einen günstigen Zug zu wählen. Wenn noch keine Informationen vorliegen, wird ein zufälliger Zug gewählt.
2. **Expansion.** Wenn ein ausgewählter untergeordneter Knoten nicht der Endzustand ist, werden alle möglichen Nachfolger hinzugefügt und einer ausgewählt.
3. **Simulation.** Ab dem ausgewählten Knotenpunkt wird das Spiel gespielt, bis es mit einem Sieg, einer Niederlage oder einem Remis endet.

- Der daraus resultierende Suchbaum sieht in etwa so aus wie in der Abbildung 4.4.



Jeder Knoten hat einen Wert von *Siege/Spiele*. Mit diesem Wert kann der beste Knoten basierend auf einem hohen Verhältnis von Gewinnen zu gespielten Spielen ausgewählt werden.

5. Constraints

Bei vielen Problemen gibt es Abhängigkeiten in den Lösungskomponenten. Es gibt zum Beispiel Fälle, in denen wenn X den Wert A hat dann hat Y den Wert B. Oder wie im Spiel Sudoku kann dieselbe Zahl nicht zweimal in derselben Reihe oder Diagonale vorkommen. Probleme, bei denen Constraints eine Rolle spielen, und wie sie gelöst werden können, werden in diesem Kapitel behandelt.

5.1 Lösen von Constrain-Problemen

Allgemeines Constraint Solving Problem

Bei einer gegebenen Menge von Variablen innerhalb eines bestimmten Wertebereichs und einer Menge von Beschränkungen erlaubter Kombinationen der Variablenwerte wird nach einer konkreten Anordnung der Variablen gesucht, die alle Beschränkungen erfüllen kann.

Constraint Solving Problem als Suche

Solche Probleme können als Suchproblem dargestellt werden, wobei:

- **Suchraum:** Menge von Variablen und deren Domänen
- **Anfangszustand:** Alle Variablen sind noch unbelegt
- **Zielzustand:** alle Variablen belegt und Constraints erfüllt
- **Zieltest:** Überprüfe, ob alle Variablen die Constraints erfüllen

5.1.1 Ansätze zur Lösung von Constraint Solving Problemen

Ansatz 1: Generiere und Teste

Generiere Kombinationen von Werten für die Variablen basierend auf ihren Domänen und teste, ob die Bedingungen erfüllt sind. Wiederhole so oft wie nötig.

x_1	x_2	x_3	Test
1	1	1	nein
1	1	2	nein
1	2	1	nein
1	2	2	nein
2	1	1	nein
2	1	2	nein
2	2	1	ja

Tabelle 5.1: Ergebnis des Generierens und Testens

Ansatz 2: Tiefensuche mit Backtracking

Die Werte für die Variablen werden systematisch generiert, um nicht alle möglichen Kombinationen generieren zu müssen.

x_1	x_2	x_3	Test
1	1	1	nein
1	1	2	nein
1	2	2	nein
1	2	1	nein
2	2	1	ja

Tabelle 5.2: Ergebnis von Tiefensuche und Backtracking

Durch den Einsatz von Heuristiken und anwendungsspezifischen Merkmalen kann die Lösung von Constraint-Problemen effizienter gestaltet werden. Unter Verwendung von Constraints kann der Zustandsraum umstrukturiert werden, um die Constraints zum leichteren Lösen des Problems auszunutzen.

5.1.2 Optimierung der Lösung von Constraint Problemen

Die folgenden Optimierungen lassen sich anhand eines konkreten Beispiels besser erklären. Ein Beispiel für ein Problem mit Constraints ist die Einfärbung von Staaten in Australien mit drei Farben, so dass benachbarte Staaten niemals dieselbe Farbe haben.



Abbildung 5.1: Karte der Staaten von Australien

Anwendung von Constraint Netzen

Ein Netz kann erstellt werden, wenn Probleme binäre Beschränkungen haben. Die Variablen sind die Knoten und die Kanten verbinden die Knoten mit Nebenbedingungen.

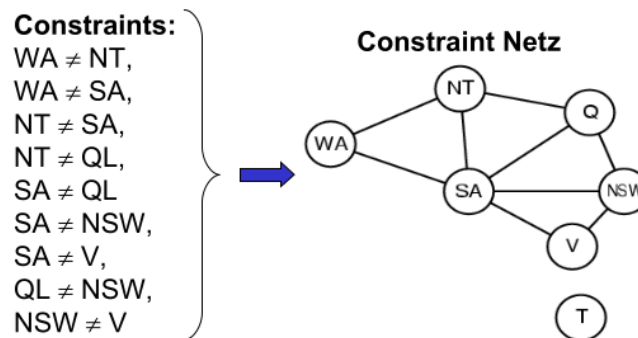


Abbildung 5.2: Constraints-Netz

Das Constraints Netz kann in Heuristiken verwendet werden, um die Nutzung der Tiefensuche mit Backtracking zu verbessern.

Abalauf von Tiefensuche mit Backtracking zur Lösung von Constraint Problemen

Mit Hilfe einer **Degree Heuristik**, bei der mit der Variable begonnen wird, die die meisten Constraints aufweist, kann die Suche optimiert werden. Based on the Constraint Netz in Abb. 5.2, können wir sehen, dass SA die meisten Einschränkungen hat und die Suche damit beginnen sollte. Danach wird die Suche wie eine typische Tiefensuche (siehe Abschnitt 3.4.5) fortgesetzt:-

1. In jedem Schritt wird einer Variablen ein Wert zugeordnet.
2. Wenn die Variable nicht mehr so erweitert werden kann, dass sie den Beschränkungen entspricht, wird ein Backtracking durchgeführt.

Eine weitere Heuristik, die verwendet werden kann, ist die **Minimum-Remaining-Value-Heuristik**. Bei dieser Heuristik wird die Variable ausgewählt, die die kleinste Anzahl möglicher Werte hat.

Abgesehen von der Auswahl der Variablen, der ein Wert zugewiesen werden soll, ist es auch möglich, das Beschränkungsnetz zu verwenden, um den Wert für eine Variable auszuwählen, der andere Variablen am wenigsten einschränken würde. Dies ist als **Least-Constraining-Value Heuristik** bekannt. Dazu schaut man sich mit dem Constraint Mesh an, mit welchen anderen Variablen eine Variable verbunden ist und welche möglichen Werte es für diese Variablen noch gibt.

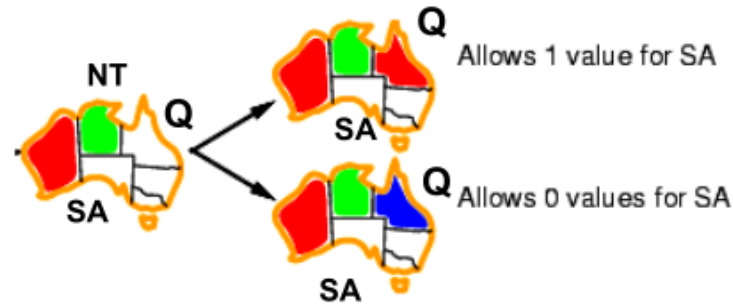


Abbildung 5.3: Least-Constraining-Value Heuristik

Angenommen, der Algorithmus versucht anhand des Australien-Färbebeispiels, eine Farbe für Q auszuwählen. Da NT in diesem Fall bereits grün ist, kann Q nur rot oder blau sein, aber da die Wahl von Rot es für SA unmöglich macht (weil WA auch schon rot ist), wird die Farbe Blau für Q gewählt (siehe Abb. 5.3).

5.1.3 Constraint Propagation

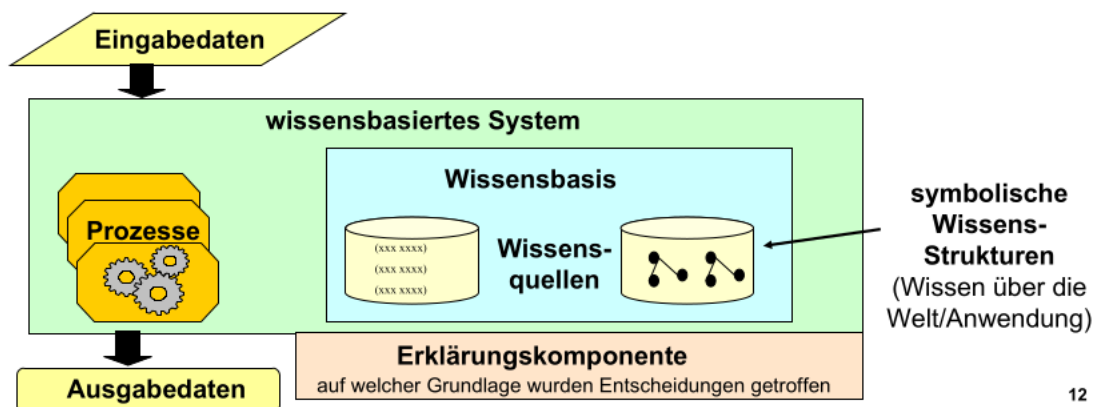
Die Idee der Constraint-Fortpflanzung besteht darin, dass, bevor einer Variablen ein Wert zugewiesen wird, zuerst die Berechnung der Auswirkung dieses Werts auf alle anderen Constraints ausgeführt werden muss.

Der einfachste Weg, dies zu erreichen, ist die Anwendung der Idee der Kantenkonsistenz. Kante A bis Kante B ist genau dann konsistent, wenn für jeden Wert von A mindestens ein zulässiger Wert für B vorhanden ist. Wenn Werte auf diese Weise eliminiert werden, können inkonsistente Werte aus den Domänen der Knoten entfernt werden.

6. Wissensbasierte Systeme

Wissensbasierte Systeme verwenden wissensbasierte Informationsverarbeitung, um Lösungen für Probleme zu berechnen. Diese Art der Informationsverarbeitung ist ein Paradigma, das eine strikte Trennung von Wissen und den zur Lösung des Problems erforderlichen Verarbeitungsmechanismen erfordert.

In einem klassischen System besteht ein Programm aus Programmen und Algorithmen. Ein wissensbasiertes System hingegen besteht aus symbolisch dargestelltem Wissen und Verarbeitungsmechanismen, um dieses Wissen zur Problemlösung zu nutzen.



12

Abbildung 6.1: Überblick über ein wissensbasiertes System

6.1 Wissensdarstellung in wissensbasierten Systemen

Wissensbasierte Informationsverarbeitung basiert auf der Vorstellung, dass viele Probleme nur lösbar sind, wenn man Erfahrung mit dem Problem und den Lösungsmöglichkeiten dafür hat. Dies wirft die Frage auf: Wie stellen wir Wissen so dar, dass es von Computern verstanden und zur Problemlösung verwendet werden kann?

Wissen wird dabei als Klassen, Kategorien und Konzepte mit Hierarchien, Zusammensetzungen und Klassifikationen dargestellt. Dies ist eine ontologische Erfassung einer Anwen-

dungsdomäne bekannt. Diese Art der Datenmodellierung lässt sich am besten mit einer Mindmap vergleichen (siehe Abb. 6.2).

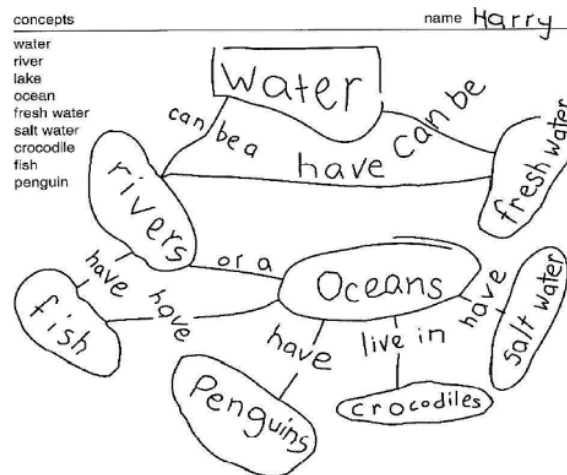


Abbildung 6.2: Beispiel einer Mindmap von einem Grundschüler

Diese Klassen, Kategorien und Konzepte werden letztlich in der Implementierung durch programmtypische Datenstrukturen wie Objekt-Arrays und Pointer abstrahiert (Abb. 6.3).

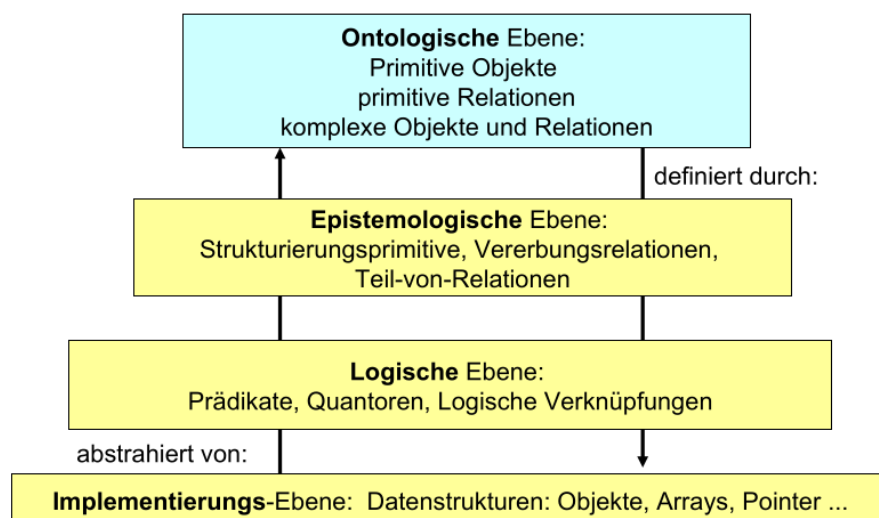


Abbildung 6.3: Ebenen der Wissensrepräsentation

Es gibt viele Möglichkeiten, Wissen explizit darzustellen. Eine Anwendung kann logikbasierte Ansätze wie Fakten und Regeln verwenden. Andere können Rahmen und Fälle oder strukturierte Vererbungsnetze ähnlich objektorientierter Modelle verwenden.

6.2 Schnittstelle zur Interaktion mit einem Wissensbasierten System

Damit ein Benutzer mit einem wissensbasierten System interagieren kann, muss eine Schnittstelle vorhanden sein, die mindestens die folgenden Funktionen ausführen kann:

- **Tell(WBS, knowledge-item):** Füge dem System neues Wissen hinzu. z.B: "Die Welt ist Rund"
- **Ask(WBS, question):** Anfrage stellen. z.B: ist die Welt eine Kugel?

Allein durch die Verwendung dieser beiden Funktionen kann ein wissensbasiertes System verwendet werden, um Probleme innerhalb seiner Wissensbasis zu lösen.

6.3 Ansätze zur Wissensverarbeitung

Eine wichtige Komponente eines wissensbasierten Systems ist die Inferenzmaschine, die verwendet wird, um unsere Antworten auf Fragen zu finden. Es gibt viele Ansätze für die Implementierung einer solchen Maschine zur Verarbeitung von vorhandenem Wissen.

Prinzip der Wissensabstraktion

Diese Methode wird als induktive Inferenz bezeichnet: Ableitung allgemeiner Aussagen aus Spezialfällen. z.B:

- Beobachtung: Studenten Hans und Fritz haben ein Laptop
- Schlussfolgerung: Alle Studenten haben Laptops

Prinzip von Wissen über Wissensintegrität

Diese Methode wird als deduktive Inferenz bezeichnet: Durch Erkennung von Regeln und Invarianzen aus formalen Beschreibungen wird explizites Wissen explizit gemacht. z.B:

- Beobachtung: Aussage A ist wahr und Aussage B ist wahr.
- Schlussfolgerung: Auch die Aussage A oder B ist wahr.

Prinzip von Wissen über Kausalzusammenhänge

Diese Methode wird als abduktive Inferenz bezeichnet: Die Ursachen werden aus Beobachtungen abgeleitet: A ist der Grund dafür, dass B passiert ist, wenn also B beobachtet wird, ist A passiert. z.B:

- Beobachtung: Licht ist an.
- Schlussfolgerung: Lichtschalter ist auf Position "Ein".

Wissen über Analogien

Diese Methode wird als analoge Inferenz bezeichnet: Neue Erkenntnisse werden aus bereits bekannten abgeleitet. z.B:

- Beobachtung: durch ein dickes Rohr kann viel Wasser fließen.
- Schlussfolgerung: durch ein dickes Kabel kann viel Strom fließen.

Prinzip der Verwendung von empirischem Wissen

Diese Methode wird als probabilistische Inferenz bezeichnet: Schlussfolgerung, dass eine Wirkung eintritt, basierend auf der Wahrscheinlichkeit, dass sie aufgrund der Situation eintritt. z.B:

- Wenn 100 Personen im Raum sind, kann man davon ausgehen, dass mindestens zwei von ihnen denselben Geburtstag haben.

Prinzip der Verwendung von unscharfem Wissen

Diese Methode wird als Fuzzy-Inferenz bezeichnet: Schlussfolgerungen auf der Grundlage der Wahrscheinlichkeit, dass ein Objekt zu einer Menge gehört. z.B:

- Hans ist vierzig Jahre alt. Alte Menschen sind weise.
- Zugehörigkeit Hans zur Menge der jungen Menschen: 0.6
- Zugehörigkeit Hans zur Menge der alten Menschen: 0.4
- Ask: Ist Hans Weise? Antwort: Na ja, geht so

6.4 Aufgaben beim Entwurf eines Wissenbasierten Systems

Es ist es wichtig, von Anfang an herauszufinden, welche Arten von Wissen es zu erwerben gilt und wie diese repräsentiert und gespeichert werden. Daneben muss auch entschieden werden, welche Inferenzmechanismen verwendet werden sollen.

Dann stellt sich die Frage, wie die Wissensbasis gefüllt werden soll. Dazu gibt es einfache Methoden, z. B. indem man einen Experten dazu bringt, sein Wissen mit einem Texteditor, einem Formular oder einem Mikrofon einzugeben. Es gibt aber auch komplexere Ansätze wie die Entwicklung einer Schnittstelle zur Umwandlung von Datenbankinhalten in ein Format, das von einem wissensbasierten System verarbeitet werden kann, oder die Extraktion neuen Wissens mit Hilfe künstlicher neuronaler Netze.

7. Regelbasierte Wissensverarbeitung und Expertensysteme

In vielen Fällen reichen typische numerische Herangehensweisen an Probleme nicht aus und es ist menschliche Expertise erforderlich. Dienstleistungen von Experten sind jedoch nicht immer verfügbar und teuer. Mit der Anwendung symbolischer Problemlösung unter Verwendung wissensbasierter Systeme (siehe Kapitel 6) kann menschliches Fachwissen mithilfe von Computern automatisiert und skaliert werden. Diese Systeme sind als Expertensysteme bekannt.

Einige Anwendungsbeispiele für Expertensysteme wären im Dienst der Medizin, wo viele seit rund 30 Jahren im Einsatz sind, in der Wirtschaft, wo es zur Kreditwürdigkeitsprüfung und zum Abwägen von Investitionsrisiken verwendet wird, und in Kundendienst-Chatbots.

7.1 Aufbau eines regelbasierten Wissensbasierten Systems

Regelbasierte Systeme basieren auf dem Prinzip eines Produktionssystems zur Termersetzung (Emil Post 1943). Eine Übersicht über ein komplettes Produktionssystem ist in Abbildung 7.1 dargestellt.

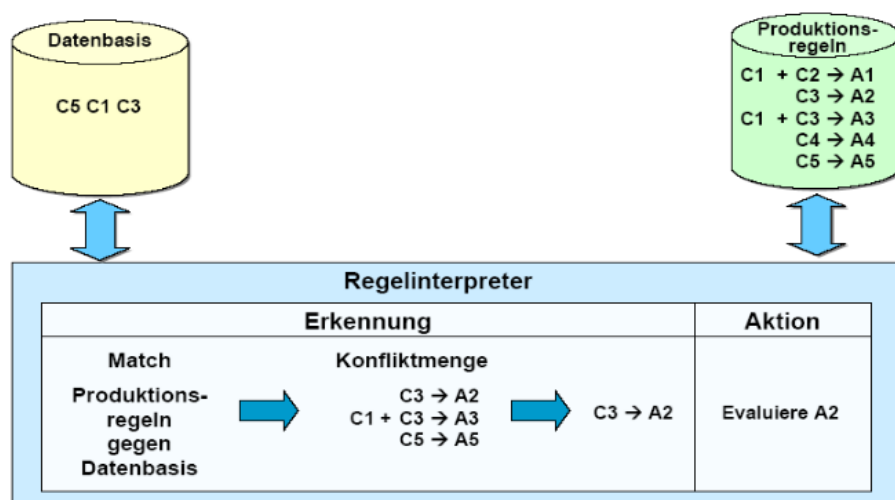


Abbildung 7.1: Produktionssystem zur Termersetzung

Eine Produktion, auch als “Regel” bekannt, besteht aus zwei Teilen (Abb. 7.2).

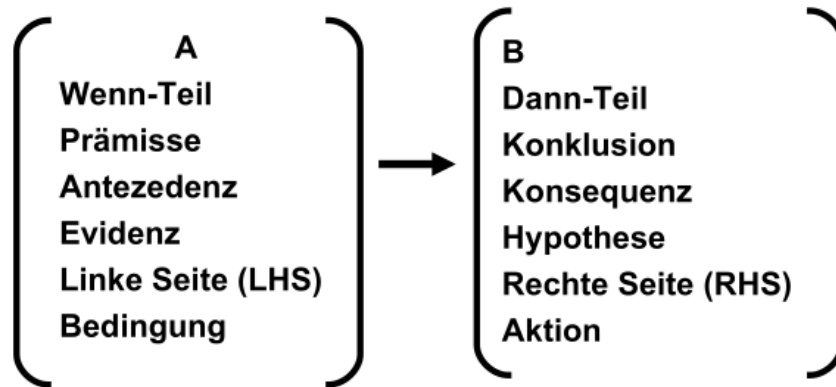


Abbildung 7.2: Regelstruktur eines regelbasierten Expertensystems

Grundsätzlich basieren die Regeln auf einer Kombination der Teile von A und Teilen von B. Zum Beispiel: Wenn A1 und A2 und An, **dann** B1 und B2 und Bn und Bm und so weiter.

Anhand dieser Regeln kann Wissen von Experten gebildet werden. Tatsächlich arbeiten viele Expertensysteme mit einer einfachen Form von IF <Bedingung> THEN <Konsequenz>.

7.2 Zustandsraumdarstellung eines Wissensverarbeitungsproblems

Die vier wichtigsten Komponenten für so ein System, die man kennen muss, sind Fakten, Regeln und Anfragen.

Fakten sind Einträge in der Fakten- oder Datenbasis. Für jedes Problem beginnt das System mit einer Beschreibung des gesamten bekannten Wissens über eine Problemstellung.

Regeln werden verwendet, um einen Lösungsschritt zu erzeugen. Mit Hilfe von Regeln werden neue Fakten berücksichtigt und einige frühere verworfen. Das bedeutet, dass die Faktenbasis durch die Anwendung von Regeln verändert wird.

Anfragen, die als Prädikate konstruiert sind, beschreiben die Bedingungen, die ein Problem umgeben. Wenn das Prädikat in der Datenbasis gefunden wird (nachdem Fakten einige Regeln durchlaufen haben), ist das Problem gelöst.

Eine Übersicht über die Beziehungen zwischen diesen drei Komponenten ist in Abbildung 7.3 dargestellt.



Abbildung 7.3: Beziehung zwischen Fakten, Regeln und Abfragen

Dabei ist es wichtig zu beachten, dass der Zustandsraum nicht definiert, wie ein Problem mit den Regeln gelöst werden kann. Hierfür ist eine Suchfunktion erforderlich.

7.3 Kompilierung von Regelbasen

Ein naiver Ansatz für eine Regelinterpreter-Implementierung wäre, in jeder Schleife alle Elemente in einer Datenbasis zu vergleichen. Dies wäre langsam und ineffizient. Dieser Prozess kann durch zwei Methoden effizienter gestaltet werden: Begrenzung der zu verarbeitenden Einträge und Begrenzung der zu verarbeitenden Bedingungen.

Um die Anzahl der Einträge pro Schleife zu begrenzen, ist es möglich, die Ergebnisse eines Vergleichs zu speichern und nur die Änderungen in der Datenbasis auf neue oder entfernte Regeln zu prüfen. Beim zweiten Ansatz können ähnliche Tests nur einmal verarbeitet werden, da die Bedingungen für Regeln oft nicht disjunkt sind.

Ansatz eines Rete-Algorithmus zur Regelbasenkompilierung

Regeln werden in ein Netz aus Codestücken umgewandelt, die Bedingungen darstellen. Ähnliche Codestücke werden kombiniert, um sicherzustellen, dass sie höchstens einmal verarbeitet werden. Ein Beispiel für ein Regelnetz ist in Abbildung 7.4 dargestellt, und die Kombination dieses Netzes ist in Abbildung 7.5 zu sehen.

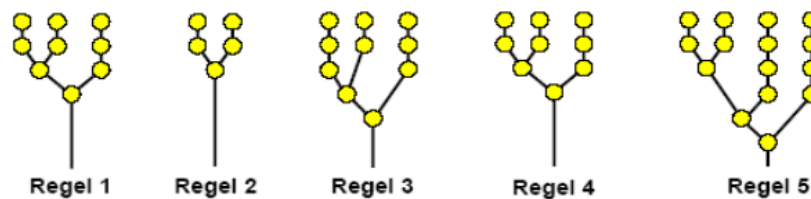


Abbildung 7.4: Beispiel eines Regelnetzes

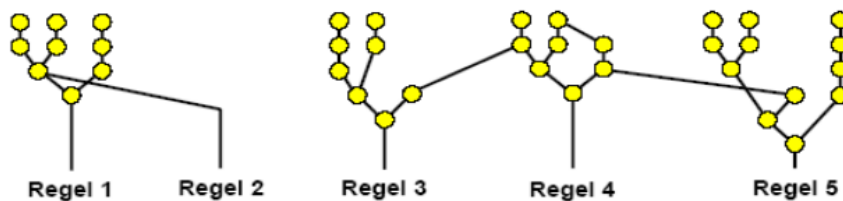


Abbildung 7.5: Kombinierte Regelnetze

7.4 Konfliktauflösung von Regelsystemen

In vielen Fällen sind mehrere Regeln anwendbar, so dass sich die Frage stellt, welche Regeln man anwenden sollte. Die Antwort auf diese Frage hängt von der Reihenfolge der Regeln in der Regelbasis ab. In kommutativen Systemen kann eine schlechte Wahl der Regeln zu langen Lösungen führen. In nicht-kommutativen Systemen kann die Schwierigkeit der Suche aufgrund der Rückverfolgung exponentiell sein.

Ansatz 1: Heuristische Regelauswahl

Wie bei den meisten Suchproblemen können Heuristiken den Prozess effizienter machen. Ein gängiger Ansatz besteht darin, sich an der Häufigkeit der verwendeten Regel zu orientieren, d. h. jede Regel darf nur einmal verwendet werden. Ansonsten kann es von Vorteil sein, Regeln zu bevorzugen, die auf den neuesten generierten Daten beruhen. Schließlich ist es auch möglich, eine Implementierung zu erstellen, die auf spezifischen Merkmalen des Problems basiert. Das bedeutet, dass Regeln, die mehr mit dem aktuellen Problem zu tun haben, bevorzugt werden.

Ansatz 2: Meta-Regeln

Metaregeln sind Regeln über Regeln. In diesem Fall wäre ein Ansatz, Regeln über die Auswahl von Regeln in Form von Regeln zu bilden. Einige Meta-Regeln wären zum Beispiel: Regeln mit billigeren Verfahren gegenüber teuren Verfahren bevorzugen, Regeln mit weniger gefährlichen Ansätzen gegenüber gefährlicheren Ansätzen bevorzugen, Regeln von Experten gegenüber Anfängern bevorzugen.

7.5 Aufgaben bei der Entwicklung eines Expertensystems

Um ein Expertensystem zu entwickeln, müssen bestimmte Schritte unternommen werden. An erster Stelle steht die Durchführung einer Machbarkeitsanalyse. Expertensysteme sind nur in bestimmten Problemfällen anwendbar, daher ist es wichtig, sich zu vergewissern, dass ein Expertensystem der richtige Methode ist.

Danach ist es wichtig zu entscheiden, wie mit dem System interagiert werden soll. Das bedeutet, dass festgelegt werden muss, welche Aufgaben unterstützt werden sollen und welche Art von Dialog zwischen Benutzern und Experten geführt werden muss.

Sobald dies erreicht ist, ist der nächste Schritt der Wissenserwerb. Es müssen Experten gefunden und ihr Wissen gesammelt werden. Hier ist es wichtig, die richtigen Fragen zu stellen, um das gegebene Problem zu lösen.

Als Nächstes folgt die Umsetzung des Wissens in Regeln, die Implementierung der Regeln in ein wissensbasiertes System und dann das Testen.

8. Formaler Logik

8.1 Wissensrepräsentation und wissensbasierte Verarbeitung auf der Grundlage formaler Logik

Es gibt Fälle, in denen Schlussfolgerungen aus Fakten, Gegenständen oder Zusammenhängen gezogen werden müssen. Zum Beispiel manchmal Fragen wie: Hat Objekt A Merkmal B? Oder ist A mit B verwandt? müssen beantwortet werden. Dies kann mit einem System geschehen, das formale Logik verwendet.

Die Idee, die hinter der Funktionsweise eines Logiksystems steht, ist die Definition der anwendungsspezifischen Fakten (durch eine Logiksprache) und die anschließende Einspeisung dieser Definition in einen allgemeinen anwendungsunabhängigen Verarbeitungsmechanismus. Der Verarbeitungsmechanismus generiert dann eine Lösung auf der Grundlage der definierten Fakten. Dieser Verarbeitungsmechanismus erfolgt durch logische Schlussfolgerungen und wird auch als Inferenzmechanismus bezeichnet.

8.1.1 Syntax und Semantik einer Logiksprache

Um Fakten und Zusammenhänge zu definieren, ist eine formale Sprache notwendig. Die Syntax einer Logiksprache definiert, welche Zeichenketten korrekt ausgedrückt werden (auch bekannt als Terme, Sätze, Ausdrücken oder Formeln).

Bei der Semantik geht es darum, wie die Sprache definiert, unter welchen Umständen ein Ausdruck wahr ist. Semantik legt fest, wie die Sprache definiert, unter welchen Umständen ein Ausdruck in der gegebenen "Welt" wahr ist.

Die Aussagenlogik beschreibt Beziehungen zwischen Ausdrücken, die wahr oder falsch sein können, und abstrahiert von der natürlichen Sprachstruktur der Aussage.

Syntax der Aussagenlogik

Die in der formalen Sprache ausgedrückte Aussagenlogik hat eine lexikalische und eine strukturelle Komponente. Die lexikalische Komponente besteht aus grundlegenden Symbolen wie Konstanten (TRUE, FALSE), Satzsymbolen (A, B, C), logischen Verknüpfungssymbolen ($\wedge \vee \neg \Leftrightarrow \Rightarrow$) und Klammern. Die Strukturkomponente besteht aus Regeln, die syntaktisch korrekte Sätze (Aussagen) bilden. Ein Beispiel für einen Satz ist: $A \vee \neg B \vee C \wedge D$ was bedeutet: "A oder nicht B oder C und D".

Semantik aussagenlogischer Ausdrücke

Die Semantik (d.h. die Bedeutung) eines aussagenlogischen Ausdrucks wird als eine Interpretation bezeichnet. Eine Interpretation ist eine Abbildung I , die Sätze auf ihre Wahrheitswerte abbildet.

8.1.2 Begriffe der Aussagenlogik-Syntax

Literale

Eine aussagenlogische Variable ist als positives Literal bekannt (z.B: A), und negierte sind als negative Literale bekannt (z.B: $\neg A$).

Klausel

Eine Formel, die aus einer Disjunktion von n Literalen besteht, ist eine Klausel (z.B: $A \vee \neg B \vee C \wedge D$).

Konjunktive Normalform

Eine Formel, die aus einer Konjunktion von Klauseln besteht, hat die konjunktive Normalform (z.B: $(A \vee \neg B \vee C \wedge D) \wedge (A \vee C \wedge E) \wedge \neg F$).

8.1.3 Belegung, Modell und Erfüllbarkeit

Gegeben sei ϕ ein Satz, σ eine Belegung für die Aussagevariablen in ϕ und I eine Interpretationsfunktion mit $I(\phi) = TRUE$. Dann ist:

- ϕ gültig unter der Belegung σ
- σ ist ein Modell von ϕ .

Notation dafür wäre $\sigma \models \phi$.

Gegeben sei ϕ ein Satz dann ist ϕ erfüllbar genau dann wenn ϕ mindestens ein Modell hat. Eine Menge von F Formeln ist erfüllbar genau dann wenn eine Belegung σ existiert, unter der alle Formeln ϕ_i aus F gültig sind.

8.2 Lösen von SAT-Problemen mit SAT-Solver

SAT-Probleme bedeuten boolesche Erfüllbarkeitsprobleme. Solche Probleme können mit einem SAT-Solver gelöst werden. Dies kann erfolgen, indem einfach boolesche Formeln ϕ an einen SAT-Solver übergeben werden, der dann nach $I(\phi) = TRUE$ auflöst.

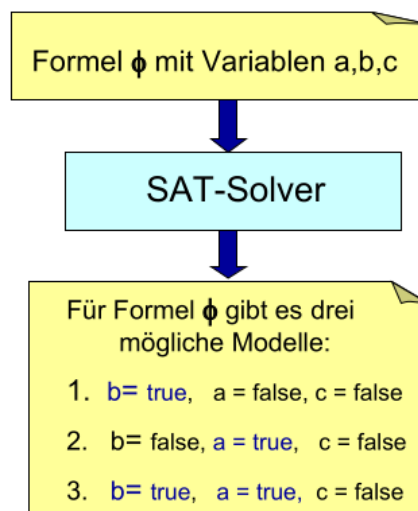


Abbildung 8.1: Verwendung eines SAT-Solvers

Abbildung 8.1 zeigt die Ausgabe der Variablen a,b,c mit der folgenden Formel:

$$\phi = (a \vee b \vee c) \wedge (a \Rightarrow \neg(a \wedge c)) \wedge (c \Rightarrow (a \wedge c \vee b \wedge c)) \wedge (\neg a \Rightarrow (\neg c \wedge b \vee \neg b \wedge c))$$

Was die folgende Wahrheitstabelle ergibt:

b	a	c	$I(\phi)$
false	false	false	false
true	false	false	true
false	true	false	true
true	true	false	true
false	false	true	false
true	false	true	false
false	true	true	false
true	true	true	false

Abbildung 8.2: Wahrheitstabelle des SAT-Problems

Ein Beispiel für eine Anwendung eines SAT Solvers ist das Spiel Sudoku. Man kann die Einschränkungen der Sudoku-Felder mit einer Logiksprache definieren und dann einen SAT-Solver verwenden, um ein Modell zu finden, das das Problem lösen kann.

8.3 Ausdrucksstärke Logiksprachen

Die Aussagenlogik funktioniert in einigen Fällen, ist aber auf die Modellierung von Fakten beschränkt. Dies kann zu kompliziert werden, wenn alle strukturellen Beziehungen explizit als Fakten dargestellt werden müssen. Mit Aussagenlogik gibt es keine Quantifizierung, keine Relationen und keine Funktionen. Logiksprache muss also erweitert werden, um aussagekräftiger zu sein.

8.3.1 Prädikatenlogik 1. Stufe

Die logische Sprache wird um Prädikate (boole'schen Funktionen), Variablen und Quantoren (ermöglicht es, über eine Menge von Objekten mit Hilfe einer Variablen zu sprechen) erweitert.

Prädikatenlogik 1. Stufe (First Order Logic) beschreibt Objekte, ihre Eigenschaften und ihre Beziehungen zueinander. Es gibt bei Prädikatenlogik auch eine Unterscheidung zwischen Syntax und Semantik.

Syntax: Lexikalischer Teil

- Konstanten-Symbole: TRUE, FALSE, A, B, Hans, Paul...
- Variablen-Symbole: x, y, ...
- Funktions-Symbole: plus, minus, mul, Vater_von...
- Prädikaten-Symbole: student, hat_Computer, P, Q...
- Logische Verknüpfungen: $\vee \wedge \neg \Leftrightarrow \Rightarrow$
- Quantoren: \exists, \forall
- Gleichheit: =

Syntax: Struktureller Teil

- **Terme:** Konstanten-Symbole, $f(t_1, t_2, t_3, \dots, t_n)$ wenn f eine Funktionsymbol und t_1 bis t_n Terme sind.
- **Atomare Formel:** $P(t_1, t_2, t_3, \dots, t_n)$ wenn Prädikatsymbol P und t_1 bis t_n Terme sind.
- **Komplexe Formel:** jede atomare Formel ist eine Formel. Wenn R und S Formeln sind und x ein Variablen-Symbol dann sind folgende Formeln: R , $\neg R$, $R \wedge S$, $R \vee S$, $R \Rightarrow S$, $\exists x : R$, $\forall x : R$

Beispiele von Sachverhalten-Formulierung in Prädikatenlogik

- Alle Kinder lieben Eiscreme: $\forall x : Kind(x) \Rightarrow liebt_Eiscreme(x)$
- Es gibt einen Baum der Nadeln hat: $\exists x : Baum(x) \wedge hat_Nadeln(x)$
- Die Mutter einer Person ist dessen weibliches Elternteil:

$$\forall x \forall y : Mutter(x, y) \Leftrightarrow (weiblich(x) \wedge Elternteil(x, y))$$

8.3.2 Semantik der Prädikatenlogik

Es wird eine Interpretationsfunktion verwendet, die der Aussagelogik ähnelt, aber komplexer ist. Der Wahrheitsgehalt einer Formel wird durch die Struktur, $S = (U, I)$ bestimmt, wobei U das "Universum" (beobachteter Gegenstandsbereich) und I die Interpretationsabbildung (die Bestandteile einer Formel, die die Objekte und Relationen im Gegenstandsbereich abbildet) sind.