DAI-Labor
TU Berlin

Grundlagen der Künstlichen Intelligenz

08.12.2005: Modallogik

According to the intentional stance, an agent is assumed to decide to act and communicate based on its beliefs about its environment and its desires and intentions.
nach D.C. Dennet: The Intentional Stance

Dr.-Ing. Stefan Fricke
stefan.fricke@dai-labor.de

A|O|T
Agententechnologien in betrieblichen Anwendungen und der Telekommunikation

Gliederung

Probleme klassischer Logik

Einführung

Klassische Logik beschreibt Tatsachen und Zusammenhänge, aber..

- ⇨ **... kein Handeln, keine Prozesse**
  - → Der Effekt einer Aktion ist unvorhersagbar in nichtdeterministischen und in dynamischen Umgebungen
- ⇨ **... keine Zeit**
- ⇨ **... keine Überzeugungen, Motivationen, etc**
  - → Ich hoffe, du glaubst, dass ich weiß, dass der Zahnarztbesuch Schmerzen bereitet. Ich beabsichtige nicht, Schmerzen zugefügt zu bekommen. (Aber ich beabsichtige zum Zahnarzt zu gehen)

Modallogik

Einführung

- ⇨ **Idee: Aus logischen Zusammenhängen und Inferenzen der Prädikatenlogik sinnvolle Sachverhalte und Folgerungen durch Einführung so genannter Modaloperatoren ermöglichen.**
- ⇨ **Modaloperatoren werden vor eine Formel geschrieben und geben ihr eine neue Interpretation. Z. B. für p = es_regnet**

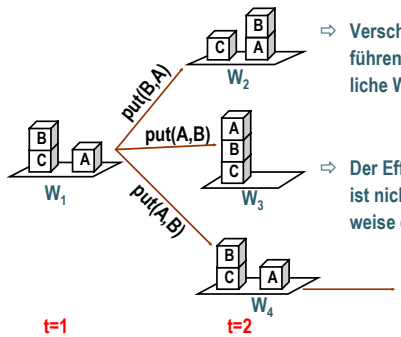
ICH_GLAUBE p vs. ICH_HOFFE ¬p

IRGENDWANN p und ICH_WEISS IRGENDWANN p

ICH_GLAUBE DU_HOFFST p vs. ICH_HOFFE DU_GLAUBST p

Possible Worlds

Einführung



- ⇨ **Verschiedene Aktionen führen in unterschiedliche Welten.**
- ⇨ **Der Effekt einer Aktion ist nicht notwendigerweise deterministisch**

Modallogik

Einführung

- ⇨ **Erreichbarkeitsrelation $R(W_1, W_2)$: W_2 ist von W_1 aus erreichbar.**
- ⇨ **Möglichkeit: ◊P ist folgerbar in einer Welt W genau dann, wenn P wahr ist in *mindestens* einer möglichen Welt**
  - → $\exists w': R(w,w') \wedge w' \models P$
- ⇨ **Notwendigkeit: □P ist folgerbar in einer Welt W genau dann, wenn P wahr ist in *jeder* möglichen Welt**
  - → $\forall w': R(w,w') \Rightarrow w' \models P$

Axiomsschemata für Modallogiken

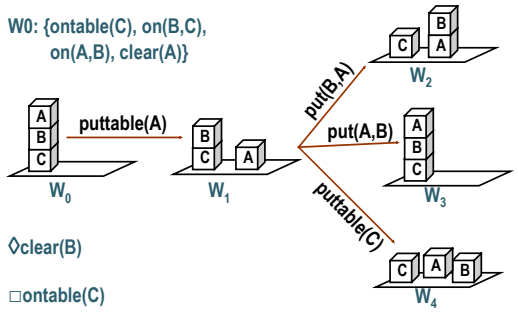
K:	$\Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$	(Distribution)
D:	$\Box A \rightarrow \neg\Box\neg A$	(Konsistenz)
T:	$\Box A \rightarrow A$	(Reflexivität)
B:	$A \rightarrow \Box\Diamond A$	(Symmetrie)
4:	$\Box A \rightarrow \Box\Box A$	(Transitivität)
NEC:	$A \rightarrow \Box A$	(Notwendigkeit)

Possible Worlds Semantik

- ⇨ **Ein Modell ist nicht nur eine, sondern eine Menge von Welten …**
  - → … generiert durch Aktionsauswahl,
  - → … generiert durch eine Sequenz von Aktionen,
  - → … generiert durch mehrere mögliche Ergebnisse einer Aktion,
  - → … generiert durch Wissenslücken.
- ⇨ **Possible Worlds dienen zur Beschreibung dieses Modells…**

Beispiel für Possible Worlds in der Blockswelt

- ⇨ W0: {ontable(C), on(B,C), on(A,B), clear(A)}

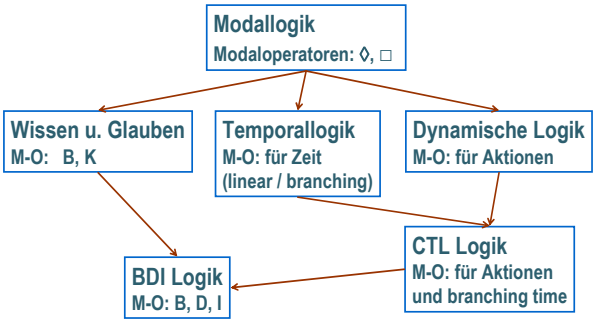


- ⇨ ◊clear(B)
- ⇨ □ontable(C)

Gliederung

- ⇨ **Einführung in Modallogik**
- ⇨ **Modallogiken für Zeit und Handeln**
  - → Temporallogik
  - → Dynamische Logik
- ⇨ **Intentionalität**
- ⇨ **Epistemische Logik mit Belief, Desire, Intention**
- ⇨ **BDI-Agenten**
- ⇨ **Zusammenfassung**

Modallogiken im Zusammenhang



Logik von Glauben und Wissen

- ⇨ $K_a\varphi$ Agent a weiß φ
- ⇨ $Bel_a\varphi$ Agent a glaubt φ
- ⇨ **Ein Agent weiß / glaubt eine Proposition φ genau dann, wenn φ in allen möglichen Welten gilt.**

Eigenschaften von Wissen und von Glauben

(1) $K_a\varphi \wedge K_a(\varphi \rightarrow \psi) \rightarrow K_a\psi$

(2) $K_a\varphi \rightarrow \varphi$ Wissen

(3) $K_a\varphi \rightarrow K_a(K_a\varphi)$ positive Introspektion

(4) $\neg K_a\varphi \rightarrow K_a(\neg K_a\varphi)$ negative Introspektion

⇨ Für $Bel_a\varphi$ analog:
- → gilt (1)
- → gilt nicht (2)
- → gilt (3)
- → kann (4) gelten, ist aber problematisch.

Beispiel

⇨ 2 Personen sitzen einander gegenüber und wissen (1), dass mindestens einer von ihnen eine weiße Stirn hat (2). B sagt, er wisse nicht, ob er eine weiße Stirn habe (3). Daraus folgert A, dass er selbst eine weiße Stirn hat.

1. $K_A(\neg White(A) \rightarrow K_B(\neg White(A))$ (1)
2. $K_A(K_B(White(A) \vee White(B)))$ (2)
3. $K_A(\neg K_B(White(B)))$ (3)

⇨ Kann A daraus schließen, dass er selbst eine weiße Stirn hat?

Beispiel

Inferenzregeln **A1**: $K_A\varphi \wedge K_A(\varphi \rightarrow \psi) \rightarrow K_A\psi$; **A2**: $K_A\varphi \rightarrow \varphi$
R2: $\varphi \rightarrow \psi \wedge K_A\varphi \rightarrow K_A\psi$

1. $K_A(\neg White(A) \rightarrow K_B(\neg White(A))$	
2. $K_A(K_B(White(A) \vee White(B)))$	
3. $K_A(\neg K_B(White(B)))$	
4. $\neg White(A) \rightarrow K_B(\neg White(A))$	1, A2; $X \rightarrow Y \Leftrightarrow \neg X \vee Y$
5. $K_B(\neg White(A) \rightarrow White(B))$	2, A2
6. $K_B(\neg White(A)) \rightarrow K_B(White(B))$	5, A1
7. $\neg White(A) \rightarrow K_B(White(B))$	4, 6
8. $\neg K_B(White(B)) \rightarrow White(A)$	Umkehrschluss von 7
9. $K_A(White(A))$	3, 8, R2

Dynamische Logik

⇨ Aktionssymbole a, b, ...

⇨ a;b Sequenz

⇨ a+b nichtdeterministische Auswahl

⇨ p? deterministische Auswahl (IF-THEN)
Aktion, basierend auf Wahrheitswert von p

⇨ a* 0 oder mehr Wiederholungen von a

⇨ <a>p a macht p möglicherweise wahr (analog zu ◊)

⇨ [a]p a macht p notwendigerweise wahr (analog zu □)

Temporallogik

⇨ Momente T mit einer partiellen Ordnung <, die die Vorher-Beziehung ausdrückt.
- → Jeder Moment entspricht einer Possible World

⇨ p U q p ist wahr bis q wahr wird (UNTIL)

⇨ Xp p ist im nächsten Moment wahr (NEXT)

⇨ Pp p war in einem früheren Moment wahr (PAST)

⇨ Ep p ist irgendwann in der Zukunft wahr (EVENTUALLY)

⇨ Ap p wird immer wahr sein (ALWAYS) $Ap \equiv \neg E \neg p$

Gliederung

Fragestellungen zum Verhalten komplexer Systeme

- Wie lässt sich das Verhalten eines Systems vorhersagen ohne Kenntnis seiner internen Struktur?
- Welche Repräsentationen eignen sich zur Definition
  - des beobachtbaren Verhaltens?
  - des erwarteten Verhaltens?
- Die Intentionalität gibt Antworten auf die Beschreibung des Verhaltens komplexer Systeme...

Zur Theorie des Intentional Stance

- Der Philosoph Daniel Dennett unterscheidet zwischen drei Heransgehensweisen, wenn man Vorhersagen über ein System treffen will:
  - physical stance,
  - design stance,
  - intentional stance...

Physical Stance

- Der physical stance beschreibt Entitäten in physikalischen Begriffen.
- Zum Beispiel
  geometrische Körper in der Mathematik,
  Mechanik im Maschinenbau,
  Statik in der Architektur.

Design Stance

- Der design stance nimmt an, dass ein Objekt zu einem Zweck entwickelt wurde und sich entsprechend verhält.
- Die Erklärung erfolgt anhand funktionaler Begriffe.
- Beispiel: Das Verhalten eines Autos wird über Funktionen wie Gas geben, Kupplung treten, Lenken, Bremsen beschrieben.

Intentional Stance

- Der intentional stance beschreibt das Verhalten einer Entität als das eines rationalen Agenten.
  - Es ist immer auch möglich, dasselbe Verhalten in rein physikalischen oder funktionalen Begriffen zu erklären.
- Aus pragmatischen Gründen ist die intentionale Perspektive jedoch unverzichtbar. Was können das für pragmatische Gründe sein?
- Die Komplexität der Entität.
  - U.U. ist es sinnvoll, einem Thermostaten die „Absicht“ zuzubilligen, die Temperatur konstant zu halten.

Theorie des Intentional Stance

- Die Kernidee ist, das Verhalten von Individuen durch die Zuschreibung von Wünschen und Überzeugungen rational verständlich und damit prognostizierbar zu machen.
- Diese Zuschreibungen beschreiben jedoch keine Tatsachen, aus denen das Verhalten aufgrund kausaler Gesetzmäßigkeiten erschlossen werden kann.
  - Vielmehr stellen intentionale Beschreibungen ein Abstraktionswerkzeug dar, das mit familiären Begriffen operiert.

Schachcomputer als Beispiel für den Intentional Stance

- **Wie setzt man sich adäquat mit einem Schachcomputer auseinander, dessen exakte innere Implementation man nicht kennt?**
  - Er hat *Wissen* über die Figuren auf den Feldern und die Bewegungsmöglichkeiten der Figuren;
  - Er hat den *Wunsch*, das Spiel zu gewinnen.
- **Mit diesen Zuschreibungen kann man gute Vorhersagen oder Erklärungsansätze über sein Verhalten liefern.**

Gliederung

- **Einführung in Modallogik**
- **Modallogiken für Zeit und Handeln**
- **Intentionalität**
- **Epistemische Logik mit Belief, Desire, Intention**
  - Modaloperatoren Belief, Desire, Intention
  - Axiomatisierungen für B, D, I
- **BDI-Agenten**
- **Zusammenfassung**

Belief, Desire, Intention (BDI) bietet dreierlei:

- **ein philosophisches Modell des menschlichen Denkens und Handelns,**
  - [Bratman, 1987]
- **verschiedene Architekturimplementierungen,**
  - (IRMA, PRS, JACK, JIAC)
- **eine abstrakte logische Semantik.**

Epistemische Modallogiken für Agenten

Epistemische Logik nutzen, um

- **mentale Attitüden von Agenten zu modellieren:**
  - Beliefs, Desires, Goals, Know How, Intentions, …,
- **das Verhalten eines Agenten zu begründen,**
- **das Verhalten eines Agenten vorhersehbar zu machen,**
- **sinnvolle Handlungen zu generieren.**

Belief, Desire, Intention

- **$Bel_x p$: Agent x glaubt p.**
- **$Des_x p$: Agent x wünscht Zustand p**
  - Ziele ermöglichen in die Zukunft gerichtetes Handeln.
- **$Int_x p$ Agent x beabsichtigt, den Zustand p zu erreichen**
  - Intentionen sind mit Handlungen verknüpft

Eigenschaften von Zielen (Goals, bzw. Desires)

- **Desires können nicht erfüllbar oder gar inkonsistent sein.**
  - Goals sind Teilmengen von Desires: erreichbar und konsistent.
- **Der Agent soll glauben, dass sein Ziel erreichbar ist.**
  - Das verhindert, dass der Agent Ziele annimmt, von denen er glaubt, dass sie unerreichbar sind.
  - Diese Eigenschaft wird Realismus genannt.
- **Ziele (und auch Intentionen) können überprüft werden.**

Eigenschaften von Intentionen

- Jede Intention soll erfüllbar sein.
- Die Intentionen sollen gegenseitig konsistent sein.
- Sie sollen mit den Beliefs des Agenten vereinbar sein.
  - $Int_x p$ impliziert nicht $Des_x p$
- Ein Agent muss nicht unbedingt an die Erfüllbarkeit einer Intention glauben, ausschließen darf er sie jedoch nicht.

Possible Worlds in der Computation Tree Logic (CTL)

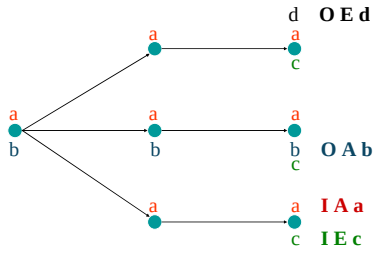
- Eindeutige Vergangenheit, verzweigende Zukunft
- Ein **Pfad** ist eine maximale Menge von Momenten, ausgehend von der Gegenwart bis in alle Momente in der Zukunft, entlang einem Zweig gemäß <
- **Situation** = Welt zu einem Zeitpunkt
- **Zustandsformeln** beziehen sich auf eine Situation
- **Pfadformeln** beziehen sich auf einen Pfad …

Operatoren für Possible Worlds

- Modale Operatoren für **Situationen:**
  - **A** *a* (Always) In jeder möglichen Zukunft gilt *a*
  - **E** *a* (Eventually) In einer possible world gilt *a*
  - z.B. E B p (irgendwann glauben, dass es regnet)
- Modale Operatoren für **Pfade:**
  - **O** *f* (optional) für mindestens einen Pfad ist *f* wahr
  - **I** *f* (inevitable) für alle ausgehenden Pfade ist *f* wahr
  - z.B. O A B p (in einem möglichen Leben immer glauben, dass es regnet)

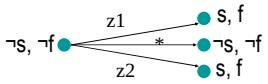
Possible Worlds in der CTL

b: Cindy ist in Berlin
a: Berlin ist Hauptstadt Deutschlands
c: Cindy ist in Celle
d: es ist Herbst



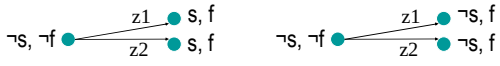
Beispiel für den Einsatz der Possible Worlds Semantik

- Ein Agent glaubt, dass es unvermeidlich ist, dass eine Zahnfüllung (f) durch Schmerz (s) begleitet wird.



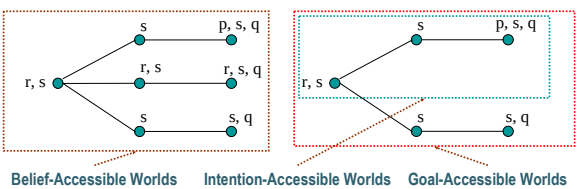
s: Schmerz;
f: Füllung;
z1: Zahnarzt 1;
z2: Zahnarzt 2;
*: andere Aktion

- Der Agent hat das Ziel, eine Zahnfüllung zu bekommen, aber nicht notwendigerweise das Ziel, Schmerz zu erleiden.



Beziehungen zwischen B, D und I

- Goal-Accessible Worlds sind Subwelten der Belief-Accessible Worlds eines Agenten.
- Intention Accessible Worlds sind Subwelten der Goal-Accessible Worlds.



Belief-Accessible Worlds Intention-Accessible Worlds Goal-Accessible Worlds

Gliederung

- Einführung in Modallogik
- Modallogiken für Zeit und Handeln
- Intentionalität
- Epistemische Logik mit Belief, Desire, Intention
  - Modaloperatoren Belief, Desire, Intention
  - Axiomatisierungen für B, D, I
- BDI-Agenten
- Zusammenfassung

KD45: Axiomatisierung für epistemische Logiken

- $Bel\ (p \rightarrow q) \rightarrow (Bel\ p \rightarrow Bel\ q)$ (K)
- $Bel\ p \rightarrow \neg Bel\ \neg p$ (D)
- $Bel\ p \rightarrow Bel\ Bel\ p$ (4)
- $\neg Bel\ p \rightarrow Bel\ \neg Bel\ p$ (5)

Inferenzregeln:

- $p \wedge p \rightarrow q \rightarrow q$ (Modus Ponens)
- $p \rightarrow Bel\ p$ (Notwendigkeit)

Schlussfolgerungen in BDI sind nicht trivial

Bel (clever(John) AND John=partner(Sally))
impliziert nicht:
Bel (clever(Sally))

Des (Zahnarztbesuch) AND Bel (Zahnarztbesuch => Schmerz)
impliziert nicht:
Des (Schmerz)

Axiomatisierungen von Agenten 1

- Wenn der Agent eine Formel beabsichtigt, dann muss er sie als Ziel haben und auch an sie glauben (starker Realismus):
  - $Int_a x \Rightarrow Des_a x$ und $Des_a x \Rightarrow Bel_a x$
    - x steht z.B. für optionally(eventually(Diplom))
  - (d.h., Agenten haben keine beliebigen Absichten)
- Realismus: Agenten glauben an ihre Intentionen und Ziele
  - $Int_a x \Rightarrow Bel_a Int_a x$ und $Des_a x \Rightarrow Bel_a Des_a x$
- schwacher Realismus: Nicht an die Nichterfüllbarkeit glauben
  - $Int_a x \Rightarrow \neg Bel_a Int_a \neg x$

Axiomatisierungen von Agenten 2

- Absichten werden durch Ziele gestützt
  - $Int_a x \Rightarrow Des_a Int_a x$
- Bewusstsein bezüglich der Ereignisse
  - unabhängig vom Ausgang (Gelingen oder Scheitern)
  - $done(e) \Rightarrow Bel_a\ done(e)$
    - done(e): Ereignis e hat stattgefunden (erfolgreich oder nicht)
- Fallen lassen von Absichten nach endlicher Zeit
  - $Int_a x \Rightarrow I\ E\ \neg Int_a x$

Verpflichtung (commitment) gegenüber Intentionen

- Intentionen lösen Handlungen aus:
  - Aktivierung eines mit der Intention verknüpften Handlungsplans.
- Entsprechend ist ein Agent gegenüber seinen Intentionen zu Handlungen verpflichtet (committed).
- Agenten können Intentionen entweder aufrechterhalten oder verwerfen.
  - z.B. in Abhängigkeit davon, ob die Handlung erfolgreich durchgeführt wurde.

Drei Arten des Commitments

- Ein **blindly commited** Agent verfolgt seine Intentionen, bis er glaubt, sie erreicht zu haben.

$$Int_a I\ E\varphi \Rightarrow I\ Int_a I\ E\varphi\ U\ Bel_a\varphi$$

  - Es ist unabdingbar, dass die Intention aufrechterhalten wird, bis („until“) der Agent φ erreicht zu haben glaubt.
  - Was geschieht aber, wenn der Agent BEL(¬φ) in seiner Wissensbasis hat?

Drei Arten des Commitments

- Ein **single-minded** Agent behält seine Absichten, bis er glaubt, dass er sie erreicht hat oder niemals erreichen wird.

$$Int_a I\ E\varphi \Rightarrow I\ Int_a I\ E\varphi\ U\ (\ Bel_a\varphi \vee \neg Bel_a\ O\ E\ \varphi\ )$$

  - Die Intention wird fallen gelassen, wenn der Agent nicht mehr an die Erfüllbarkeit des Zustands glaubt.
  - Was bedeutet das für ein Ziel $Des_a\ E\ \varphi$?

Drei Arten des Commitments

- Ein **open-minded** Agent bleibt seinen Intentionen solange verpflichtet, bis er sie als erreicht ansieht oder er keine **entsprechenden Ziele** mehr hat.

$$Int_a I\ E\varphi \Rightarrow I\ Int_a I\ E\varphi\ U\ (\ Bel_a\varphi \vee \neg Des_a\ O\ E\ \varphi\ )$$

  - Die Intention wird auch dann fallen gelassen, wenn das unterstützende Ziel nicht mehr existiert.

Axiomatisierungen von Agenten 3

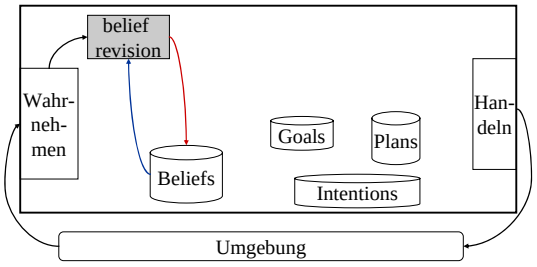
- Ist eine Formel notwendigerweise gültig, dann soll der Agent nicht gezwungen sein, sie als Ziel oder Absicht aufzunehmen.
  - $Bel_a I\ \varphi \not\Rightarrow Des_a I\ A\ \varphi$
- Ist eine Formel $(\varphi \rightarrow \psi)$ notwendigerweise gültig und hat der Agent das Ziel (oder die Absicht) φ, dann soll er nicht gezwungen sein, auch b als Ziel (oder Absicht) zu haben.
  - $Bel_a I\ (\varphi \rightarrow \psi) \wedge Des_a I\ \varphi \not\Rightarrow Des_a I\ \psi$

Gliederung

- Einführung in Modallogik
- Modallogiken für Zeit und Handeln
- Intentionalität
- Epistemische Logik mit Belief, Desire, Intention
- BDI-Agenten
  - Allgemeine BDI-Architektur
  - Agent-0
- Zusammenfassung

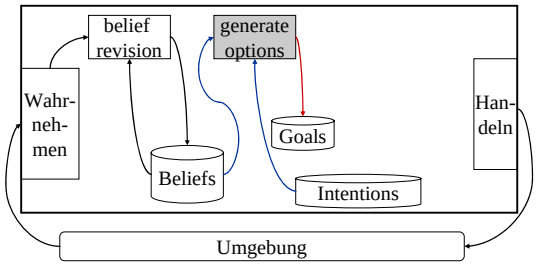
Modell einer BDI-Agentenarchitektur

1. Belief revision function: Bel* x Perception → Bel*



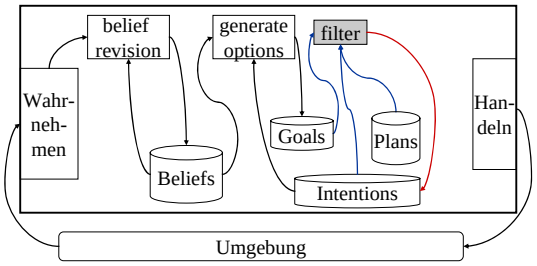
Modell einer BDI-Agentenarchitektur

2. Option generation : Bel* x Int* → Goal*



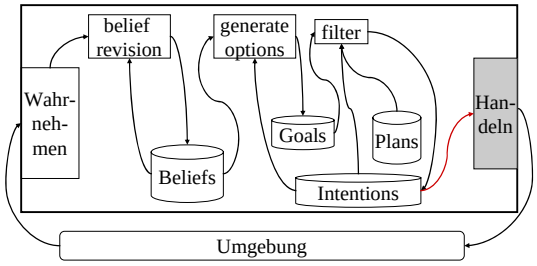
Modell einer BDI-Agentenarchitektur

3. filter : Goal* x Int* x Plan* → Int*



Modell einer BDI-Agentenarchitektur

4. ausführen: Int* → Action



Gliederung

- ⇨ Einführung in Modallogik
- ⇨ Modallogiken für Zeit und Handeln
- ⇨ Intentionalität
- ⇨ Epistemische Logik mit Belief, Desire, Intention
- ⇨ BDI-Agenten
  - → Allgemeine BDI-Architektur
  - → Agent-0
- ⇨ Zusammenfassung

AGENT-0 [Shoham 1993]

Agent-Oriented Programming besteht nach Shoham aus:

- ⇨ logischem System zur Beschreibung mentaler Zustände,
- ⇨ interpretierter Programmiersprache für Agenten sowie
- ⇨ Agentifizierungsprozess: Kette von Übersetzungsschritten.

AGENT-0: Logisches System

AGENT-0 unterstützt folgende Sprachelemente

- ⇨ belief (mental) - BEL
- ⇨ commitment (mental) - CMT
- ⇨ capability (nicht mental) - CAN
- ⇨ Grammatik = modale Prädikatenlogik, erweitert um Zeitpunkte

mentaler Belief-Operator

BEL(<agent>, <timepoint>, <fact>)

BEL(a, t1, BEL(b, t2, like(a, b, t3)))

„Agent *a* glaubt zum Zeitpunkt *t1*, dass Agent *b* zum Zeitpunkt *t2* an *like(a,b,t3)* glaubt“

nicht-mentaler Capability-Operator

CAN(<agent>, <timepoint>, <fact>)

CAN(a, t1, open(door, t2))

Fähigkeit wird durch Zustand ausgedrückt

„Agent *a* kann zum Zeitpunkt *t1* sicherstellen, dass die Tür zum Zeitpunkt *t2* offen ist“

Mentaler Commitment-Operator

CMT(<agent>, <agent>, <timepoint>, <fact>)

CMT(a, b, t1, open(door, t2))

„Agent *a* ist Agent *b* zum Zeitpunkt *t1* zur Türöffnung verpflichtet“

Logisches System: Annahmen

⇨ **Konsistenz von Beliefs und Commitments**

⇨ **Good faith:**
CMT(a,b,t,x) <=> BEL(a, t, CAN(a, t, x))

⇨ **Introspection:**
CMT(a,b,t,x) <=> BEL(a, t, CMT(a, b, t, x))
¬CMT(a,b,t,x) <=> BEL(a ,t, ¬CMT(a, b, t, x))

Die AGENT-0 -Sprache: communicative actions und private actions

⇨ **inform(t, a, fact)**
→ zum Zeitpunkt *t* dem Agenten *a* das *fact* zusenden

⇨ **request(t, a, action)**
→ Agent *a* zu Handlung *action* auffordern
→ unrequest und refrain analog.

⇨ **DO(t, p-action)**
→ zum Zeitpunkt *t* die lokale Methode *p-action* ausführen

Die AGENT-0 -Sprache: Commitment Rules

COMMIT(msgcond, mtlcond, agent, action)

msgcond = (From, Type, Content)

„aufgrund einer empfangenen Nachricht in einem bestimmten Zustand einem anderen Agenten zu einer Aktion verpflichtet sein.“

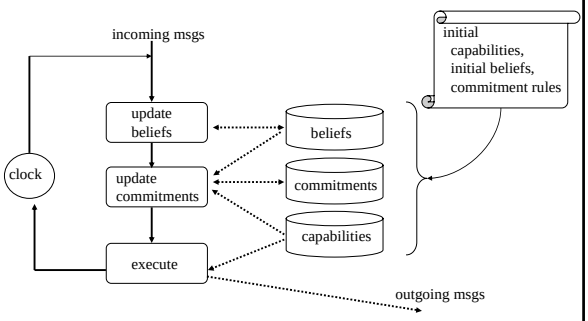
existenzquantifizierte Variable

```
COMMIT(   (?a, REQUEST, ?action),
          (BEL [ !t, myfriend( ?a ) ]),
          ?a,
          ?action )
```

allquantifizierte Variable

„requests von Freunden werden ausgeführt“

Agent-0: Architektur und Interpreterschleife



Agent-0: Interpreterschleife: 1. Belief-Update

betrifft inform-Sprechakte:

inform(t1, a, fact(x, t2))

- Einfügen des Fakts in die Belief-Datenbasis
- Entfernen inkonsistenter Fakten

Agent-0: Interpreterschleife: 2. Commitment-Update

commit(msgcond,mtlcond,a,action)

Die Verpflichtung wird angenommen, wenn

a) msgcond mappt msg,
b) mtlcond ist wahr,
c) action ist aktuelle capability,
d) Agent ist nicht zu REFRAIN(action) verpflichtet.

unrequest(action): Entfernen der Handlung aus Commitment-DB

Agent-0: Interpreterschleife: 3. Execute

3. Ausführung von Commitments in Abhängigkeit der Zeit:

INFORM(t, b, fact)
a) senden
b) assert(BEL(t,a, BEL(t,b,fact)))

REQUEST, UNREQUEST
senden

DO(t, action)
a) action checken,
b) ggfs. ausführen

Gliederung

- ⇨ Einführung in Modallogik
- ⇨ Modallogiken für Zeit und Handeln
- ⇨ Epistemische Logiken
- ⇨ Zusammenfassung

Modallogiken — Zusammenfassung

	Syntax	Inferenz	Semantik
Modal	Möglichkeit ◊ und Notwendigkeit □	Inferenzregeln für ◊ und □	Possible Worlds
Dynamisch	Sequenz, Verzweigung, Test	Inferenzregeln für Ergebnisse	Possible Worlds mit Übergängen
Temporal	Zeitpunkte oder Intervalle, zeitliche Beziehungen	Inferenz für Aussagen mit zeitlicher Ergänzung	Possible Worlds mit vielfältigen Übergängen

Zusammenfassung Intentionalität

- Der intentional stance bietet eine pragmatische und natürliche Methode, die Regeln und Verhaltensmuster unabhängig von der tatsächlichen Implementation erkennen lässt.
- Es geht nicht darum, ob ein Agent wirklich etwas „glaubt“, „wünscht“ oder „beabsichtigt“.
- BDI erweist sich als besonders nützlich, wenn ein Agent über andere Agenten „nachdenken“ soll.

BDI-Logiken: Zusammenfassung

- Der Agent wird als intentionales System beschrieben
- Modallogik wird für nicht-monotones Schließen verwendet
  - Modaloperatoren Bel, Des, Int
  - verschiedene Axiomatisierungen sind möglich
- BDI überbrückt die Lücke zwischen Wissen und Handeln
  - Aktionen werden über Intentionen, Intentionen über Ziele und Ziele über Wissen motiviert.

Vorteile einer *expliziten Repräsentation* von Zielen in Agenten

- höhere Fehlertoleranz:
  - Schlägt eine Handlung fehl, kann der Agent besser wieder aufsetzen.
- bessere Koordination:
  - Widersprüchliche Ziele nicht gleichzeitig verfolgen.
  - Subsumierende Ziele nur einmal verfolgen.
  - Ständiges Überdenken und Umplanen bei veränderten Randbedingungen ist möglich.

BDI-Logiken: Diskussion und Ausblick

- Welche Modaloperatoren verwenden?
  - Wissen, Know How, Kommunikation
  - Commitment, Joint Goal, Joint Commitment
- Die Axiomatisierungen sind problematisch
  - Constraintformalismen drücken keine Aktivitäten aus
  - Wo kommen Ziele her?
- Der Umgang mit Widersprüchen und veralteten Informationen ist unklar.

Weitere Probleme klassischer Logik

Zusammenfassung

Klassische Logik ist als Repräsentationsformalismus ungeeignet für viele Aspekte der wirklichen Welt:

- zur Repräsentation kontinuierlicher Größen,
- zur Repräsentation des Unbekannten,
- zur Repräsentation unsicheren, probabilistischen Wissens,
  - „morgen ist 80%ige Regenwahrscheinlichkeit“
- zur Repräsentation von Relativität und Unschärfe.
  - „Paul ist ziemlich groß“.

DAI-Labor
TU Berlin

Grundlagen der Künstlichen Intelligenz

08.12.2005: Modallogik

Abschluss der ersten Teils GKI „symbolische KI“

ab nächste Woche: statistische Verfahren

Dr.-Ing. Stefan Fricke
stefan.fricke@dai-labor.de

AIOIT
Agententechnologien in betrieblichen Anwendungen und der Telekommunikation

Referenzen

- **Agent-0:**
  - Y. Shoham. Agent-oriented programming. Artificial Intelligence, 60. S. 51 - 92, 1993.
  - www.agentbuilder.com: kommerzielles Agent-0
- **BDI:**
  - Modeling Rational Agents within a BDI – Architecture (Anand. S. Rao, Michael P. Georgeff)
  - Andreas Kerlin, Anatolij Zubow, Daniel Göhring: Glauben und Absichten. Seminararbeit HU Berlin, 2002, http://www.drgoehring.de/uni/papers/Glauben_und_Absichten_062002.pdf

Anhänge

BDI-Programm

```
Beliefs = Beliefs0
Intentions = Intentions0 Goals = Goals0
while true do
        get next perceipt p
        Beliefs = brf(Beliefs,p)
        Intentions = options(Goals,Intentions)
        Intentions = filter(Beliefs, Goals, Intentions)
        π = plan(Beliefs, Intentions)
        while not (empty(π) or succeeded (Intentions, Beliefs) or
                    impossible(Intentions, Beliefs)) do
                α = head(π)
                execute(α)
                π = tail(π)
                get next perceipt p
                Beliefs = brf(Beliefs,p)
                if not sound(π, Intentions, Beliefs) then
                    π = plan(Beliefs, Intentions)
        end while
end while
```

Dropping intentions that are impossible or have succeeded

Reactivity, replan

Klassische Logik und Agenten — Einleitung

Schwierige Anwendbarkeit der Prädikatenlogik für bestimmte Umgebungstypen:

- **Nichtdeterministisch** → **unzuverlässige Aktionen**
- **Partiell beobachtbar** → **unvollständiges Wissen**
- **Kontinuierlich** → **Repräsentationsproblem**
- **Dynamisch** → **unzuverlässiges Wissen**

Agent-0 – Zusammenfassung

- **Der Agent ist im Wesentlichen ein Regelinterpreter.**
- **Ziele werden nicht modelliert.**
- **Zeitpunkte sind problematisch im Zusammenhang mit Wissen.**
- **Wenig Aufmerksamkeit wird dem belief-revision geschenkt.**
- **AgentBuilder ist eine kommerzielle Erweiterung von Agent-0**
  - mit Sprechakten und Interaktionsprotokollen
  - www.agentbuilder.com