

Berufsakademie Stuttgart - Staatliche Studienakademie -		Ausbildungsbereich : Technik Fachrichtung: Informationstechnik / Angewandte Informatik
Studienfach: Wissensbasierte Systeme Hilfsmittel: keine		Studienjahrgang/Kurs: - Studienhalbjahr: 5/6 Dozent: Dr. Reichardt 120 Min

LÖSUNGEN

Seite 2	NAME:
---------	-------

Teil 1: Einführung

Aufgabe 1 (6 Punkte)

Alan Turing

- a) Beschreiben Sie kurz und in eigenen Worten den Aufbau und Ablauf des Turing Tests. Wie wird das Ergebnis interpretiert? (3 Punkte)

Teilnehmer sind ein Frager (Interrogator) und zwei Befragte (Person A und B). Die Befragten befinden sich in Raum X und Y und dem Fragenden werden Antworten jeweils nur aus dem Raum übermittelt, ohne dass er die Personen hört oder sieht. Aufgabe des Fragenden ist es, durch geschickte Fragestellung herauszufinden wer in Raum X und wer in Raum Y ist. Aufgabe von Person A ist nun, den Fragenden zu unterstützen, die Aufgabe von Person B ist es, ihn zu einer falschen Entscheidung zu bringen. Person B wird nun durch einen Computer ersetzt. Ist die Erfolgsquote des Fragenden nun weiterhin vergleichbar (d.h. die Aufgabe von Person B wird ähnlich gut erfüllt), so wird der Maschine Denkfähigkeit unterstellt.

- b) Geben Sie ein (fiktives) Beispiel eines kognitiven Systems an, welches die drei Forderungen an kognitive Systeme von Alan Turing erfüllt. Beschreiben Sie jeweils wodurch diese erfüllt werden (Stichworte). (3 Punkte)

Die drei Forderungen sind :

- das System ist in symbolischer Logik programmiert
- verfügt über eine Lernkomponente und
- über eine menschliche Schnittstelle um sich anzupassen und um Instruktionen entgegen nehmen zu können

Typische Beispiele könnten z.B. aus dem Robotik-Bereich gewählt werden, es sollte genannt werden, was dort gelernt wird und wie die Schnittstelle beschaffen ist.

Aufgabe 2 (6 Punkte)

Expertensysteme

- S. Unterlagen

Aufgabe 3 (6 Punkte)

Agenten

In einem unternehmensinternen Helpdesksystem soll ein "intelligenter" Softwareagent eingesetzt werden, der die Helpdeskmitarbeiter bei eingehenden Anfragen bzw. Fehlerfällen unterstützen soll. Seine Aufgabe ist es, einfache Email-Anfragen direkt zu beantworten, sofern dies möglich ist und komplexere Anfragen die der Agent nicht verstehen bzw. bearbeiten kann, an die entsprechenden Helpdeskmitarbeiter weiterzuleiten.

- a) Geben Sie eine geeignete Klassifikation eines Agenten mit obigen Eigenschaften und Aufgaben in der PEAS - Struktur an. (4 Punkte)

P = Performance Measure : Anzahl der zufrieden stellenden Antworten kann rückgekoppelt werden, Bewertungsmaßstab bzgl. signifikanter Stichworte im Mail-Text incl. der Kategorien (Themenfelder).

E = Environment : Helpdesk, Email-Anfragen von Nutzern, Helpdesk-Mitarbeiter

A = Actuators: Weiterleitung einer Email-Anfrage an Helpdesk-Mitarbeiter, Senden einer Antwort-Email (dazu: Textgenerierung mit Templates etc.)

S = Sensors: Empfang einer Email (Text), Zugriff auf Antwortdatenbank

- b) Allgemein: Geben Sie in eigenen Worten an, wie sich "Goal-Based-Agents" von "Utility-Based-Agents" unterscheiden. Machen Sie dabei insbesondere den Einsatzbereich von Utility-Based Agents deutlich und zeigen Sie den Vorteil zu Goal-Based Agents auf. (2 Punkte)

Goal-Based Agents: suchen einen Weg zu einem definierten Ziel (ggf. zu Zwischenzielen). Utility-Based Agents bewerten Zustände des Suchraums mit einer Nützlichkeitsfunktion als Basis für eine Handlungsentscheidung. Dies ist dann wichtig, wenn die Suche eines Ziels zu lange dauern würde (z.B. Schachspiel).

Aufgabe 4 (2 Punkte)

Wissensmanagement

Eine wesentliche Komponente wissensbasierter Systeme und des Knowledge-Managements sind Ontologien. Was versteht man unter einer Ontologie?

Eine Ontologie beschreibt ein gemeinsames Verständnis einer Domäne, welches zwischen Menschen und Applikationen ausgetauscht werden kann. Auch: Die Beschreibung einer Domäne in Form von Objekten und Relationen, sogenannten „Sinnzusammenhängen“.

Teil 2: KI-Sprachen

Aufgabe 1 (7 Punkte)

LISP Grundlagen

Die nachfolgend aufgeführten LISP-Ausdrücke werden in der angegebenen Reihenfolge ausgewertet. Geben Sie nach jedem Schritt an, welchen Wert LISP zurückliefert.

```
> (CAR (CDR '(LIST A (LIST B C) D)))
A
> (SETQ PAAR (APPEND (LIST 'KATJA) (LIST 'PETER)))
(KATJA PETER)
> (SETQ PETER (LIST 'PETER 'MUELLER '27 'KAUFMANN))
(PETER MUELLER 27 KAUFMANN)
> (EVAL (CAR (EVAL (CAR (CDR PAAR)))))
(PETER MUELLER 27 KAUFMANN)
> (DEFUN F(PAAR) (COND ((EQUAL (CAR PAAR) 'PETER) PETER) (T "?")))
F
> (F '(PETER PAN))
(PETER MUELLER 27 KAUFMANN)
> (MEMBER '3 '(1 2 3 4 5))
(3 4 5)
```

Aufgabe 2 (7 Punkte)

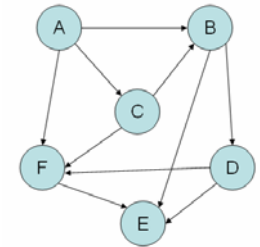
LISP Anwendung

Ein Straßennetz sei durch einen gerichteten azyklischen Graphen repräsentiert. Die Lisp-Funktion CONNECTED(X) liefert zu einem Knoten (Ort) X die Liste der von diesem aus mit einem Schritt erreichbaren Knoten (d.h. die Nachbarknoten bzw. Orte). Die Knoten sind durch Ortsnamen repräsentiert. Ziel ist es, einen Weg von Ort A nach Ort B zu finden.

- a) Geben Sie eine LISP-Funktion **isNeighbour(X,Y)** an, welche bei Eingabe eines Knotens X entscheidet, ob X direkt (in einem Schritt) von Y aus erreichbar ist. Die Ausgabe soll T oder NIL sein. (2 Punkte)
- b) Geben Sie eine LISP-Funktion **isReachable(X,Y)** an, welche überprüft, ob Knoten X von Y aus in beliebig vielen Schritten erreicht werden kann. Die Ausgabe soll T oder NIL sein. (5 Punkte)

Beispiel: (für nebenstehenden Graphen)

```
> (isNeighbour 'C 'E)
NIL
> (isNeighbour 'F 'D)
T
> (isNeighbour 'C 'B)
NIL
> (CONNECTED 'D)
(F E)
> (isReachable 'F 'B)
T
```



```
(defun isNeighbour (X Y)
  (COND ((MEMBER X (CONNECTED Y)) T)))
```

oder:

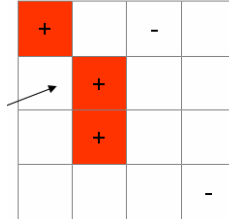
```
(defun isNeighbour (X Y)
  (isIn X (CONNECTED Y)))
(defun isIn (X L)
  (cond ((null L) nil)
        ((equal X (car L)) t)
        (t (isIn X (cdr L)))))

(defun isReachable (X Y)
  (cond ((isNeighbour X Y) t)
        (T (isReachableList X (CONNECTED Y)))))

(defun isReachableList (X L)
  (cond ((null L) nil)
        ((isReachable X (car L)) t)
        (t (isReachableList X (cdr L)))))
```

Aufgabe 3 (12 Punkte)**Prolog**

In einem 4x4 Feld wird Edelmetall gesucht. Ein Detektor kann verwendet werden, dessen Einsatz jedoch recht teuer ist, so dass die Häufigkeit des Einschaltens gering gehalten werden soll. Ihre Aufgabe ist es, ein Prolog Programm zu schreiben, welches anhand der bisher gemachten Untersuchungen ausgeben kann, welches Feld mit Sicherheit solches Edelmetall enthält. Der Detektor kann auf einem Feld aufgestellt werden und gibt dann an, ob auf einem der 8 angrenzenden Felder Edelmetall liegt oder nicht. In nebenstehender Grafik sind 3 positive und 2 negative Untersuchungen durchgeführt worden. Aus diesen lässt sich schließen, dass Feld 1,2 (s. Pfeil) Edelmetall enthalten muß!



Für eine PROLOG-Realisierung soll das Feld zeilenweise als Liste gespeichert sein. Ein p steht dabei für eine positive, ein n für eine negative Messung und k für keine durchgeführte Messung.

- a) Geben Sie ein Prolog-Programm an, welches die Anzahl der Messungen in einem als Parameter übergebenen 4x4 Spielfeld bestimmt. (5 Punkte)

```
?- anzahl([p,k,n,k],[k,p,k,k],[k,p,k,k],[k,k,k,n]), ZAHL).
```

```
ZAHL = 5
```

```
anzahlFeld([p], 1).
anzahlFeld([n], 1).
anzahlFeld([k], 0).
anzahlFeld([p|Rest], Z):-anzahlFeld(Rest, Z1), Z is Z1+1.
anzahlFeld([n|Rest], Z):-anzahlFeld(Rest, Z1), Z is Z1+1.
anzahlFeld([k|Rest], Z):-anzahlFeld(Rest, Z).
anzahl([X], Z) :- anzahlFeld(X,Z).
anzahl([X|Rest], Z) :- anzahlFeld(X,Z1), anzahl(Rest,Z2), Z is Z1 + Z2.
```

Geben Sie ein Prolog-Programm an, welches alle ausgeschlossenen Einzelfelder in einer Zeile mit a markiert (außer p und n Feldern) und die so entstehende Zeile ausgibt. (3 Punkte)

```
?- markiere([p,k,n,k], ZEILE).
```

```
ZEILE = [p,a,n,a]
```

- b) Geben Sie ein Prolog-Programm an, welches aus einem gegebenen Feld eine Spalte extrahiert und ausgibt. Die Spalte wird mit einer Zahl zwischen 1 und 4 angegeben. (3 Punkte)

```
?- spalte([p,k,n,k],[k,p,k,k],[k,p,k,k],[k,k,k,n]), 3, SPALTE).
```

```
SPALTE = [n,k,k,k]
```

```
spalte([[_,_,_],[_,_,_],[_,_,_],[_,_,_],4,[A,B,C,D]).
spalte([[_,_,_],[_,_,_],[_,_,_],[_,_,_],3,[A,B,C,D]).
spalte([[_,_,_],[_,_,_],[_,_,_],[_,_,_],2,[A,B,C,D]).
spalte([[_,_,_],[_,_,_],[_,_,_],[_,_,_],1,[A,B,C,D]).
```

- c) Geben Sie ein Prolog-Programm an, welches in einem 4x4-Feld zeilen- und spaltenweise alle Felder mit a markiert, welche ausgeschlossen werden können (noch nicht die Diagonalen!). (4 Punkte)

Gegeben sei dazu ein Prolog-Programm, welches alle ausgeschlossenen Einzelfelder in einer Zeile mit a markiert (außer p und n Feldern) und die so entstehende Zeile ausgibt.

```
?- markiere([p,k,n,k], ZEILE).
```

```
ZEILE = [p,a,n,a]
```

Beispiel:

```
?- markiereAlle([p,k,n,k],[k,p,k,k],[k,p,k,k],[k,k,k,n]), FELD).
```

```
FELD = [[p,a,n,a],[k,p,a,k],[k,p,k,a],[k,k,a,n]]
```

```
markiereZeilen([], []).
markiereZeilen([X|Rest], [Y|MRest]):-markiere(X,Y),markiereZeilen(Rest,MRest).
markiereAlle([A,B,C,D],[E,F,G,H]):- markiereZeilen([A,B,C,D],[Z1,Z2,Z3,Z4]),
    spalte([Z1,Z2,Z3,Z4],1,S1),
    spalte([Z1,Z2,Z3,Z4],2,S2),
    spalte([Z1,Z2,Z3,Z4],3,S3),
    spalte([Z1,Z2,Z3,Z4],4,S4),
    markiereZeilen([S1,S2,S3,S4],[E1,E2,E3,E4]),
    spalte([E1,E2,E3,E4],1,E),
    spalte([E1,E2,E3,E4],2,F),
    spalte([E1,E2,E3,E4],3,G),
    spalte([E1,E2,E3,E4],4,H).
```

Teil 3: Repräsentation und Methoden**Aufgabe 1 (9 Punkte)****Prädikatenlogik + Resolution**

Eine Aussage wurde in eine Prädikatenlogische Formel umgesetzt. Beweisen Sie formal, daß diese Formel allgemeingültig ist. Dabei seien x und y Variablensymbole, *franz* ein Konstantensymbol und *stirbt*, *liebt* und *traurig* Prädikatssymbole.

$$[(\forall x \forall y ((\text{stirbt}(x) \wedge \text{liebt}(y,x)) \rightarrow \text{traurig}(y))) \wedge (\forall x \exists y \text{liebt}(y,x))] \rightarrow (\text{stirbt}(\text{franz}) \rightarrow (\exists x \text{traurig}(x)))$$

Eine Umformung in eine zu widerlegende skolemisierte, quantorenfreie Form hat bereits stattgefunden. Es ergibt sich:

$$(\neg \text{stirbt}(x) \vee \neg \text{liebt}(y,x) \vee \text{traurig}(y)) \wedge \text{liebt}(f(x_1),x_1) \wedge \text{stirbt}(\text{franz}) \wedge \neg \text{traurig}(y_1)$$

- a) Führen Sie einen Resolutionsbeweis durch und geben Sie zu jedem Schritt die nötigen Unifikatoren an. (7 Punkte)

1:	$\neg \text{stirbt}(x) \vee \neg \text{liebt}(y,x) \vee \text{traurig}(y)$		
2:	$\text{liebt}(f(x_1),x_1)$		
3:	$\text{stirbt}(\text{franz})$		
4:	$\neg \text{traurig}(y_1)$		
R1,3	$5: \neg \text{liebt}(y,\text{franz}) \vee \text{traurig}(y)$	$\sigma_1 = \{ \}$	$\sigma_{mgu} = \{x/\text{franz}\}$
R5,3	$6: \neg \text{liebt}(y_1,\text{franz})$	$\sigma_5 = \{ \}$	$\sigma_{mgu} = \{y/y_1\}$
R2,6	$7: \square$	$\sigma_2 = \{ \}$	$\sigma_{mgu} = \{y_1/f(\text{franz}) \ x_1/\text{franz}\}$

- b) Wie ist die Skolem-Funktion f zu interpretieren? Welche Semantik steckt hinter dieser in dem gegebenen Kontext? (2 Punkte)

Die Funktion f ersetzt den Existenzquantor und repräsentiert somit das „explizite“ Auffinden des erfüllenden Elements. In diesem Fall ist dies als ein Freund oder Verwandter (etc.) von Franz zu interpretieren (und könnte als eine solche Relation in einer Datenbank abgelegt sein).

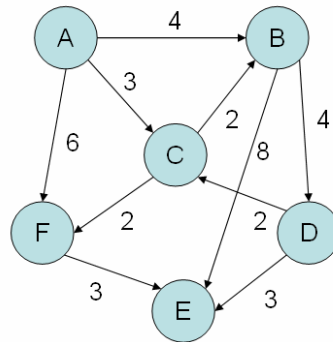
Aufgabe 2 (11 Punkte)

Suchverfahren – A*

Es sei ein A*-Algorithmus für die Graphensuche gegeben. Von Knoten A aus soll der beste Weg zu Knoten E gesucht werden.

- a) Führen Sie den Algorithmus durch und geben Sie nach jedem Schritt an, welche Knoten auf der OPEN und CLOSED Liste stehen, welchen Wert die Funktion $f(x)$ und $g(x)$ hat, welche Alternativen bestehen und welche ausgewählt wird. Geben Sie am Schluß an welchen Pfad mit welchen Kosten der Algorithmus ausgehen wird. Die Funktion $h(x)$ sei ferner gegeben durch folgende Tabelle: (8 Punkte)

H(A)	8
H(B)	3
H(C)	3
H(D)	2
H(E)	0
H(F)	3



Lösungsweg:

Schritt	OPEN	CLOSED
Start	(A)	()
Auswahl von A	(C B F)	(A)
Erreichbar (B,C,F)		
$f(B) = g(B) + h(B)$ $= 4 + 3 = 7$		
$f(C) = 3 + 3 = 6$		
$f(F) = 6 + 3 = 9$		
Auswahl von C	(B F)	(A C)
$g(B_{\text{über } C}) = 5 > g(B) = 4$		
$g(F_{\text{über } C}) = 5 < g(F) = 6$		
$f(F) = 5 + 3 = 8$		
$f(B) = 7$		
Auswahl von B	(F D E)	(A C B)
$f(E) = 12 + 0 = 12$		
$f(D) = 8 + 2 = 10$		
Auswahl von F	(E D)	(A C B F)
$g(E_{\text{über } F}) = 8 < g(E) = 12$		
$f(E) = 8 + 0 = 8$		

Ziel erreicht – Pfad = (A C F E) mit Länge 8

- b) Der A* Algorithmus terminiert unter einer Voraussetzung immer mit dem kürzesten Weg. Welche Voraussetzung ist dies? Begründen Sie warum dadurch immer der kürzeste Weg ausgegeben wird. (3 Punkte)

Die Schätzfunktion h muß optimistisch gewählt sein, d.h. $h(X) \leq \text{kuerzester_weg}(X)$. Begründung: Wird das Ziel auf die CLOSED-Liste gelegt (Ende), so bedeutet dies: $f(\text{Ziel}) = g(\text{Ziel}) + h(\text{Ziel})$ optimal unter allen Möglichkeiten. $H(\text{Ziel})$ ist 0, d.h. $g(\text{Ziel})$ entspricht dem realen Weg(X). Für alle anderen Wege liegt jedoch eine Schätzung vor, die bereits länger ist als dieser Weg. Ist die Schätzung optimistisch, so ist auch jeder reale Weg länger.

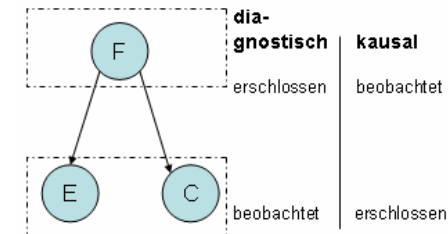
Aufgabe 3 (8 Punkte)

In der realen Welt sind viele Sachverhalte nicht mit klaren Zuordnungen als "wahr" oder "falsch" einzuordnen. Eine Methode, die diesem Umstand Rechnung trägt ist als "Probabilistic Networks" bekannt.

- a) Erläutern Sie in wenigen, eigenen Worten den Unterschied zwischen kausaler und diagnostischer Inferenz und geben Sie je ein Beispiel (incl. Zeichnung eines entsprechenden Netzes) an. (2 Punkte)

Kausale Inferenz : Gegeben sind die Voraussetzungen, geschlossen wird auf die Folgen (kausal davon abhängigen Ereignisse) entlang der gerichteten Kanten eines probabilistischen Netzes.

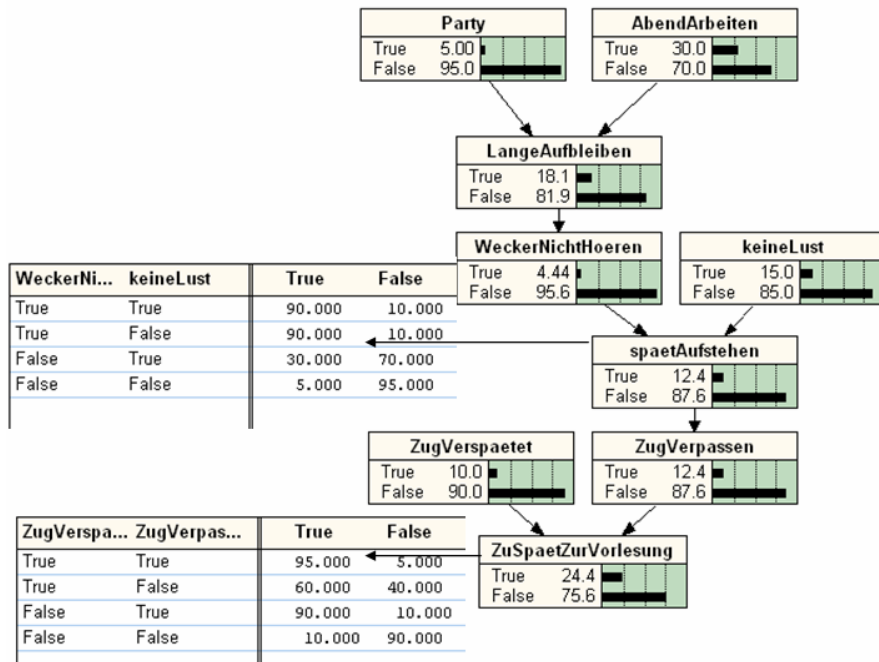
Diagnostische Inferenz: Gegeben sind Beobachtungen, geschlossen wird auf deren Ursachen bzw. auf die Wahrscheinlichkeit der möglichen ursächlichen Ereignisse.



- b) Gegeben seien die folgenden Wahrscheinlichkeitsereignisse und Zusammenhänge in verbaler Form. Erstellen Sie für diese ein Probabilistisches Netz und geben Sie für die Knoten „ZuSpätZurVorlesung“ und „spätAufstehen“ plausible, zu diesen Aussagen passende CPTs für an. (6 Punkte)

Situation: Wie wahrscheinlich ist es, dass **Klaus** rechtzeitig zur Vorlesung erscheint? Gehen Sie der Einfachheit halber davon aus, dass nur diese Abhängigkeiten gelten.

- Wenn am Vorabend eine Party stattfindet, bleibt er sehr wahrscheinlich lange auf.
- Wenn er am Vorabend noch viel gearbeitet hat, so geht das meist nicht so lange, es könnte jedoch mal vorkommen.
- Wenn er lange aufgeblieben ist, so erhöht sich die Wahrscheinlichkeit, dass er den Wecker nicht hört.
- Wenn er noch keine große Lust hat und lange aufgeblieben ist, steht er etwas später auf.
- Wenn er den Wecker nicht hört, so steht er wahrscheinlich erst später auf.
- Wenn er später aufsteht kann es sein, dass er den Zug verpasst.
- Wenn der Zug verspätet ist, so kommt er wahrscheinlich etwas zu spät zur Vorlesung.
- Wenn er den Zug verpasst, so kommt er ziemlich sicher zu spät zur Vorlesung.

**Aufgabe 4 (10 Punkte)**

Ein weiterer Modellierungsansatz für Unschärfe wird durch die Dempster-Shafer Regel beschrieben. Folgender "Kriminalfall" sei mit Hilfe dieser Technik zu klären. Dazu sind die möglichen Täter und deren Attribute, sowie mehrere Beobachtungen gegeben.

Verdächtiger	Abkürzung	Geschlecht	Haarfarbe	Größe
Peter	P	M	Schwarz	1,80
Frank	F	M	Schwarz	1,81
Monika	M	W	Schwarz	1,64
Karl	K	M	Blond	1,79
Eveline	E	W	Braun	1,55
Albert	A	M	Grau	1,72
Helga	H	W	Blond	1,65
Johannes	J	M	Grau	1,76

Dazu seien folgende Beobachtungen gegeben:

- 1.) Zeuge 1 berichtet, daß er eine dunkelhaarige Person beobachtet hat (d.h. Schwarz oder Braun) und ist sich ziemlich sicher, so dass dies mit 0.8 bewertet wird.

- 2.) Zeuge 2 konnte zwar nicht die Haarfarbe erkennen, aber die Größe schätzen. Da er zwei Personen gesehen hat von denen er nicht sagen kann, welche es dann schließlich war, ergibt sich eine Evidenz für die Größe 1,80-1,90 von 0.4 und eine Evidenz für die Größe von 1,70-1,79 von 0.4.

Aufgabe:

- a) Geben Sie die Maße m_1 und m_2 an, die durch die obigen Punkte definiert werden. Geben Sie dabei jeweils an, für welche Personen-Untermenge welche Evidenz vorliegt, sowie die jeweilige Evidenz für die Gesamtmenge Ω . (3 Punkte)

$$m_1(\{P, F, M, E\}) = 0.8 \quad m_1(\Omega) = 0.2$$

$$m_2(\{P, F\}) = 0.4 \quad m_2(\{K, A, J\}) = 0.4 \quad m_2(\Omega) = 0.2$$

- b) Bestimmen Sie die Plausibilität für die Täterschaft von Peter nach Vorlage der beiden (!) Zeugenaussagen. (7 Punkte)

Akkumulation der beiden Maße

$$\text{Konfliktbestimmung } k = m_1(\{P, F, M, E\}) * m_2(\{K, A, J\}) = 0.32 \quad \text{Korrekturfaktor } K = 1/(1-0.32)$$

$$m_{12}(\{P, F\}) = (0.8 * 0.4) + (0.2 * 0.4) * K = 0.4 * K$$

$$m_{12}(\{P, F, M, E\}) = 0.8 * 0.2 * K = 0.16 * K$$

$$m_{12}(\{K, A, J\}) = 0.4 * 0.2 * K = 0.08 * K$$

$$m_{12}(\Omega) = 0.2 * 0.2 * K = 0.04 * K$$

$$\text{Plausibilität}(\{P\}) = m_{12}(\{P, F\}) + m_{12}(\{P, F, M, E\}) + m_{12}(\Omega) = (0.4 + 0.16 + 0.04) * K = 0.6 * K$$

Aufgabe 5 (8 Punkte)

Es wurde das Lernverfahren "Versionenraummethode" (Version Space) eingeführt. Dieses soll nun auf folgende Situation angewandt werden. Ein Fremder aus einer entfernten Galaxie möchte erlernen was ein Hund ist. Dies erfolgt anhand positiver und negativer Beispiele.

- Das erste Wesen ist haarig, hat Schlappohren, läuft auf vier Pfoten und bellt.
- Das zweite Wesen ist haarig, hat spitze Ohren, läuft auf vier Pfoten und miaut.
- Das dritte Wesen ist haarig, hat spitze Ohren, läuft auf vier Pfoten und bellt.

Der Lehrer gibt an, dass nur das erste und dritte Wesen Hunde seien. Aus weiteren Beobachtungen weiß der Fremde, dass

- Wesen immer entweder haarig, nackt oder gepanzert sind
- die Ohren runde, oder spitze Form haben oder Schlappohren sind
- Wesen auf 2 oder 4 Pfoten (resp. Füßen, Hufen) laufen
- Wesen bellen, miauen, brüllen, fauchen, zischen, zwitschern oder sprechen.

- a) Wählen Sie die Attribute und geben Sie die initialen Mengen S und G an. (1 Punkt)

Attribute z.B. ("Haut", "Ohren", "Fortbewegung", "Sprache")

$$S = \{(_, _, _, _)\} \quad G = \{(*, *, *, *)\}$$

- b) Geben Sie die Menge S und G nach jedem vorgelegten Beispiel an. (5 Punkte)

- Das erste Wesen ist haarig, hat Schlappohren, läuft auf vier Pfoten und bellt. (haarig, Schlappohren, vierPfoten, bellt)

$$S = \{(\text{haarig}, \text{Schlappohren}, \text{vierPfoten}, \text{bellt})\} \quad G = \{(*, *, *, *)\}$$

- Das zweite Wesen ist haarig, hat spitze Ohren, läuft auf vier Pfoten und miaut.
(haarig, Spitzohren, vierPfoten, miaut)

$S = \{(haarig, Schlappohren, vierPfoten, bellt)\}$ $G = \{(*, *, *, bellt), (*, Schlappohren, *, *)\}$

- Das dritte Wesen ist haarig, hat spitze Ohren, läuft auf vier Pfoten und bellt.
(haarig, Spitzohren, vierPfoten, bellt)

$S = \{(haarig, *, vierPfoten, bellt)\}$ $G = \{(*, *, *, bellt)\}$

c) Erläutern Sie kurz die Bedeutung des Ergebnisses. (2 Punkte)

Das Klassifikationssystem wird alle Wesen akzeptieren, für die gilt:

$haarig(x) \wedge vierPfoten(x) \wedge bellt(x)$

Es wird alle ablehnen, für die gilt:

$\neg bellt(x)$

Für alle weiteren ist keine Klassifikation möglich, das System müsste noch weiter lernen.

Aufgabe 6 (8 Punkte)

Ein Immobilienmakler will seine Webseite mit einer adaptiven Komponente ausstatten und einem Kunden die Gebrauchtimmobiliën in der Reihenfolge anzeigen, die ihrer Ähnlichkeit zu einem gegebenen Beispiel entsprechen. Dazu wird festgelegt, dass die Beschreibungsmerkmale (Attribute) die folgenden sind: Alter (in Jahren), Ausstattung (gehoben, zeitgemäß, dem Alter entsprechend), Heizungsart (Gaszentral, Öl, Erdwärme), Kaufpreis, Wohnfläche, Grundstücksgröße. Er nimmt an, dass Wohnfläche und Grundstücksgröße den Kunden wichtiger sind als die anderen Attribute. Geben Sie ein sinnvolles und zu den Angaben passendes Ähnlichkeitsmaß an.

Attribute und Skalierungen

Alter (kardinal), Ausstattung (nominal), Heizungsart (nominal), Kaufpreis (kardinal), Wohnfläche (kardinal), Grundstücksgröße (kardinal)

Über kardinal-skalierte Attribute ist eine Distanzfunktion möglich aus der eine Ähnlichkeit abgeleitet werden kann:

$$\begin{aligned} d_{\text{Alter}}(x, y) &= |x - y| & \text{sim}_{\text{Alter}}(x, y) &= 1 - (d_{\text{Alter}}(x, y) / (1 + d_{\text{Alter}}(x, y))) \\ d_{\text{Kaufpreis}}(x, y) &= |x - y| & \text{sim}_{\text{Kaufpreis}}(x, y) &= 1 - (d_{\text{Kaufpreis}}(x, y) / (1 + d_{\text{Kaufpreis}}(x, y))) \\ d_{\text{Wohnfläche}}(x, y) &= |x - y| & \text{sim}_{\text{Wohnfläche}}(x, y) &= 1 - (d_{\text{Wohnfläche}}(x, y) / (1 + d_{\text{Wohnfläche}}(x, y))) \\ d_{\text{Grundstücksgröße}}(x, y) &= |x - y| & \text{sim}_{\text{Grundstücksgröße}}(x, y) &= 1 - (d_{\text{Grundstücksgröße}}(x, y) / (1 + d_{\text{Grundstücksgröße}}(x, y))) \end{aligned}$$

Die nominal-skalierten Attribute können nur direkt verglichen werden, d.h. es genügt folgende

$$\text{Ähnlichkeitsfunktion: } \text{sim}_N(x, y) = \begin{cases} 1 & x = y \\ 0 & \text{sonst} \end{cases}$$

Wichtigkeit (Gewichtungsfaktoren): $w_{\text{Wohnfläche}} = 2$, $w_{\text{Grundstücksgröße}} = 2$, andere sind 1

Es ergibt sich insgesamt:

$$\begin{aligned} \text{sim}(X, Y) &= (w_{\text{Wohnfläche}} \cdot \text{sim}_{\text{Wohnfläche}}(X_{\text{Wohnfläche}}, Y_{\text{Wohnfläche}}) + \\ &w_{\text{Alter}} \cdot \text{sim}_{\text{Alter}}(X_{\text{Alter}}, Y_{\text{Alter}}) + \\ &w_{\text{Grundstücksfläche}} \cdot \text{sim}_{\text{Grundstücksfläche}}(X_{\text{Grundstücksfläche}}, Y_{\text{Grundstücksfläche}}) + \\ &w_{\text{Kaufpreis}} \cdot \text{sim}_{\text{Kaufpreis}}(X_{\text{Kaufpreis}}, Y_{\text{Kaufpreis}}) + \\ &w_{\text{Heizungsart}} \cdot \text{sim}_{\text{Heizungsart}}(X_{\text{Heizungsart}}, Y_{\text{Heizungsart}}) + \\ &w_{\text{Ausstattung}} \cdot \text{sim}_{\text{Ausstattung}}(X_{\text{Ausstattung}}, Y_{\text{Ausstattung}})) / \\ &(w_{\text{Wohnfläche}} + w_{\text{Alter}} + w_{\text{Grundstücksfläche}} + w_{\text{Kaufpreis}} + w_{\text{Heizungsart}} + w_{\text{Ausstattung}}) \end{aligned}$$