

Tentamen Kennissystemen

29 maart 2006

15:15 – 18:00 uur

Dit tentamen bestaat uit 2 opgaven op 2 pagina's.

Puntenverdeling:

1a	1b	1c	1d	1e	1f	1g	2a	2b	2c	2d	2e	2f	2g	2h
5	5	10	5	5	5	5	8	5	5	12	5	5	5	5

Totaal aantal punten = 90

Lees de vragen nauwkeurig en let goed op dat je alle onderdelen van de vraag beantwoordt (vergeet bijvoorbeeld niet om je antwoord uit te leggen of een voorbeeld te geven wanneer daarom gevraagd wordt). Succes!

Opgave 1: Reisplanner 9292ov.nl

De website 9292ov.nl is een erg nuttig hulpmiddel bij het plannen van een reis met het openbaar vervoer. De website vertelt je precies hoe je met de trein, bus en/of metro van A naar B kunt komen.

- a) In het college zijn een aantal mogelijke voordelen van het toepassen van kennissystemen genoemd. Noem twee voordelen die voor de OV reisplanner **niet relevant** zijn.

“Retains expertise” is niet relevant omdat er geen sprake is van expert-kennis die bij slechts één expert aanwezig is. Ook is “can solve problems with uncertain or incomplete knowledge” niet heel relevant omdat er weinig onzekerheid is bij het uitvoeren van de taak. Dit zijn de twee meest voor de hand liggende antwoorden. Andere antwoorden met een goede argumentatie zijn ook goed gerekend.

Bij een analyse van een kennissysteem op kennisniveau (*knowledge level*) beschrijf je onder andere de taak / taken die het systeem uitvoert.

- b) Geef de naam van de standaard taak die je in de reisplanner kunt herkennen.

Planning of configuratie.

- c) Noem twee andere aspecten van een kennissysteem die beschreven worden op het kennisniveau. Geef een beschrijving van ieder aspect voor het voorbeeld van de reisplanner.

Een beschrijving van (de afbakening van) het domein (bijvoorbeeld: reisinformatie over openbaar vervoer met trein, bus, metro en tram binnen Nederland).

Een beschrijving van de type kennis die relevant zijn (bijvoorbeeld: kennis over dienstregeling, kennis over verbindingen, kennis over geografie, etc).

Bij de bestudering van een systeem op het symboolniveau (*symbol level*) kijk je naar de manier waarop de kennis in een computer gerepresenteerd wordt. In de OV reisplanner wordt gebruik gemaakt van redundante representaties van sommige kennis in het systeem.

- d) Leg uit wat het voordeel is van het gebruik van redundante representaties.

Je kunt voor iedere taak die je moet uitvoeren gebruik maken van de meest efficiënte representatie.

- e) Beargumenteer aan de hand van de criteria zoals die gegeven zijn in het boek en tijdens het college of de representatie van kennis in de reisplanner expliciet (*explicit* of *declarative*) is.

De kennis is expliciet omdat de kennis autonoom is (dienstregelingen staan op zichzelf), het een duidelijke betekenis heeft (dienstregeling zijn aankomst- en vertrektijden) en er een causaal verband is tussen de representatie en het systeem (een andere dienstregeling geeft een ander resultaat). Dit staat in slide 21 van college 6.

Omdat heel veel mensen deze vraag beantwoord hebben alsof het gaat om de vraag of de kennis “primitief” is, is van deze vraag een bonusvraag gemaakt.

Een potentieel probleem bij de taak die reisplanner uitvoert is het zogenaamde drempeleffect (*threshold effect*).

- f) Leg uit wat dit probleem inhoudt en geef een voorbeeld van dit probleem voor de reisplanner.

Een kleine verandering van de componenten kan een grote verandering voor het resultaat tot gevolg hebben. Een voorbeeld voor de reisplanner is dat een kleine verlenging van de reistijd met de bus zou kunnen veroorzaken dat je de trein mist en daardoor direct een half uur later aankomt.

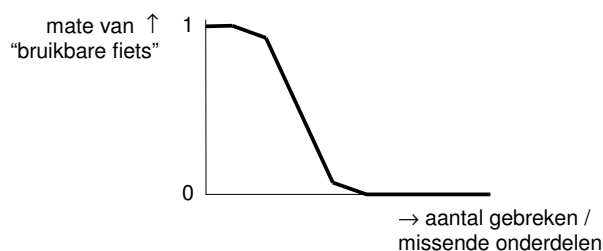
- g) Beschrijf hoe dit probleem over het algemeen wordt opgelost. Leg uit hoe deze oplossing voor de reisplanner zou werken.

Door de componenten een beetje te overschatten. Bijvoorbeeld, de reistijd of overstaptijd iets langer nemen.

Opgave 2: Fietsendiefstal

In Amsterdam worden ieder jaar heel veel fietsen gestolen. De Amsterdamse politie probeert het aantal diefstallen terug te dringen en zo veel mogelijk gevonden fietsen bij de eigenaar terug te bezorgen. Hiervoor zouden ze een kennissysteem kunnen gebruiken. Dit systeem zou allereerst, aan de hand van het aantal gebreken of missende onderdelen, moeten bepalen of de fiets nog een bruikbare fiets is. Een “bruikbare fiets” is een vaag begrip dat met behulp van *fuzzy sets* gerepresenteerd zou kunnen worden.

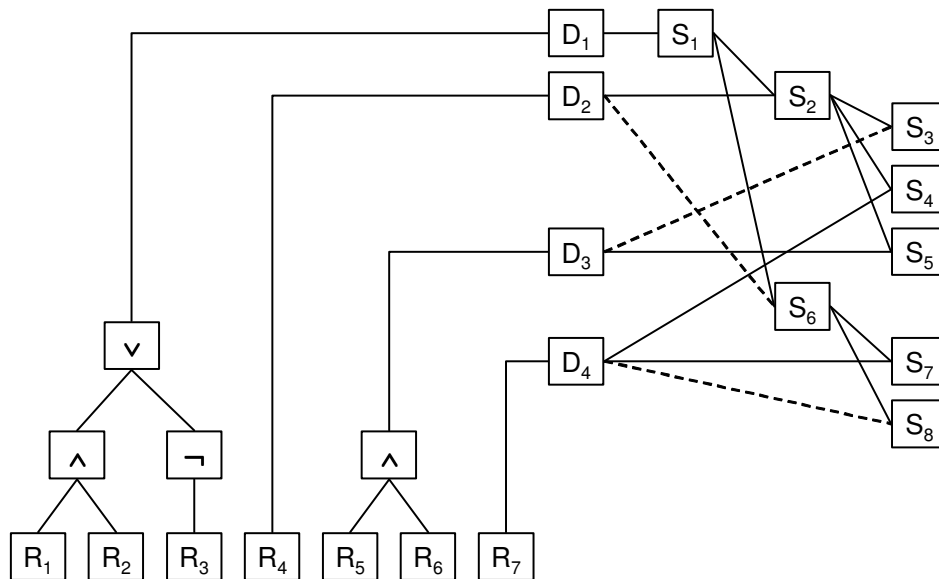
- a) Schets de grafiek voor het vage concept “bruikbare fiets” en leg uit wat er op de X-as en op de Y-as van de grafiek staat.



- b) Een “fietswrak” is het tegenovergestelde van een bruikbare fiets. Leg uit in woorden hoe je de *membership function* (ook wel *characteristic function* genoemd) voor “fietswrak” kunt berekenen vanuit de *membership function* voor “bruikbare fiets”.

Door de inverse van de *characteristic function* te nemen krijg je de *characteristic function* voor “fietswrak”. Oftewel, $f_{wrak}(X) = 1 - f_{bruikbaar}(x)$

Een volgende taak voor het systeem is het classificeren van de fietsen in verschillende typen, zoals “racefiets”, “damesfiets”, “mountainbike” etc. Stel dat we de volgende schematische representatie van de classificatie taak hebben. In dit schema wordt ruwe data (R’s) via logische operatoren ($\wedge$ en, $\vee$ of, $\neg$ niet) gecombineerd tot abstracte gegevens (D’s) en vervolgens geassocieerd in oplossingsclasses (S’s). De oplossingsclasses met een gemeenschappelijke ‘ouder’ zijn onderling onverenigbaar (*mutual exclusive*).



- c) Geef een voorbeeld voor R_5 , R_6 en D_3 voor de classificatie van fietsen.

Als R_5 staat voor “bandbreedte > 3 cm”, R_6 voor “profiel > 5 mm” en staat D_3 voor “terreinbanden”

- d) Geef voor ieder van de volgende vectoren aan welke oplossingsclasses *consistent* zijn, welke *matchen* en welke *inconsistent* zijn.

1. $R = (? \ ? \ 0 \ 1 \ 1 \ ? \ ?)$

$D = (1 \ 1 \ ? \ ?)$

S_1 en S_2 matchen (en zijn consistent)

$S_3 - S_5$ zijn consistent

$S_6 - S_7$ zijn inconsistent

N.B. Als je S_2 consistent noemt omdat de subclasses inconsistent zijn is het ook goed.

2. $R = (1 \ 1 \ 0 \ 0 \ 1 \ 1 \ 0)$

$D = (1 \ 0 \ 1 \ 0)$

S_1 , S_6 en S_8 matchen (en zijn consistent)

$S_2 - S_5$ en S_7 zijn inconsistent.

N.B. Als je S_1 en S_6 consistent noemt omdat de subclasses inconsistent zijn is het ook goed.

- e) Leg uit welke standaard classificatiemethode het meest voor de hand ligt om voor dit model te gebruiken. Beschrijf hoe deze methode werkt.

MC4 is ligt het meest voor de hand omdat we zowel een hiërarchie in de data hebben om data abstractie te doen en een hiërarchie in de oplossingen om een hiërarchische oplossingsmethode te gebruiken.

Een volgende taak van het systeem is het zoeken naar geregistreerde fietsdiefstallen in de buurt van de plaats waar een fiets gevonden is.

- f) Beschrijf het *nearest first* algoritme om in een ruimte efficiënt te kunnen zoeken naar punten in de buurt van een ander punt.

Het nearest-first algoritme deelt een ruimte op in uniforme subruimten en zoekt allereerst in de ruimte waarin het punt ligt vanwaar je vertrekt en vervolgens spiraalsgewijs in de omliggende ruimtes.

Een probleem bij het toepassen van dit algoritme is dat het aantal fietsdiefstallen in het centrum veel groter is dan in stadsdelen verder weg van het centrum.

- g) Leg uit waarom dit een probleem is het efficiënt zoeken.

Omdat je in het centrum zo veel punten binnen een ruimte heb dat je alsnog bijna alle punten moet bekijken. Wanneer je dat op zou lossen door kleinere subruimtes te maken krijg je te veel lege ruimtes waardoor het zoeken ook inefficiënt wordt.

- h) Geef de naam van een representatie van ruimte die dit probleem oplost en beschrijf wat deze representatie inhoudt.

Quad-tree representatie. Bij deze representatie wordt de ruimte in vier kwadranten verdeeld en vervolgens wordt ieder kwadrant dat meer punten bevat dan een vooraf vastgesteld aantal ook weer opgedeeld in vier kwadranten, net zo lang tot ieder kwadrant minder punten bevat dan het vooraf vastgestelde aantal.

Einde tentamen.