

Hertentamen Kennissystemen

11 juni 2007
18:30 – 21:15 uur

Dit tentamen bestaat uit 4 opgaven op 2 pagina's.

Puntenverdeling:

1a	1b	2a	2b	2c	2d	2e	3a	3b	3c	3d	3e	4a	4b	4c
3	8	8	8	5	3	5	5	3	5	12	5	3	5	5

Totaal aantal punten = 83

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: Symbolen en betekenis

Volgens de *physical symbol hypothesis* zijn symbool-systemen een voldoende voorwaarde voor intelligent gedrag. Kennissystemen werken intern met symbolen.

a) Waarin verschilt een symbool van een willekeurige markering (*marking or pattern*)?

*Een symbool kan worden herkend en geschreven door een “herkenner” (recognizer).
Eventueel: het heeft een betekenis.*

b) Noem twee manieren om betekenis aan symbolen toe te kennen en beschrijf kort wat deze methoden inhouden.

*Reference semantics (denotatie): symbolen toekennen aan dingen in de echte wereld.
Truth semantics: waarheidswaarden toekennen aan logische uitdrukkingen.*

Opgave 2: Vermindering van files

Rijkswaterstaat werkt aan een kennissysteem om de files te reduceren. Het systeem detecteert of het verkeer normaal doorstroomt, of dat er sprake is van een doorstromingsprobleem. Wanneer het verkeer niet goed verloopt, zal het proberen de automobilisten via aanwijsborden naar alternatieve wegen te sturen.

a) Noem vier voorwaarden voor het toepassen van kennissystemen en bespreek of daar in dit geval aan voldaan is.

Vier van de volgende punten moeten genoemd worden (1 punt) en op zo'n manier besproken worden voor dit specifieke systeem dat duidelijk is wat de betekenis is van de voorwaarde.

- *Cost-effective*
- *Human expertise sufficiently available*
- *Well structured knowledge (not: common sense knowledge)*
- *Limited problem domain*
- *Traditional methods do not apply*
- *Expertise is available during design*

- b) Bij het bouwen van een kennissysteem worden vier verschillende fases onderscheiden. Geef de namen van deze fases en leg uit wat er in ieder van deze fases plaatsvindt.

- *Probleem identificatie: vocabulaire en doelen bepalen.*
- *Conceptualisatie: type en rol van kennis bepalen, bijv. m.b.v. hard-op-denk protocollen.*
- *Formalisatie: redeneerstappen en zoektechnieken bepalen.*
- *Implementatie: systeem bouwen m.b.v. software engineering technieken.*

- c) Welke standaard taak kun je in dit systeem herkennen? Leg je antwoord uit.

Diagnose. Er is sprake van een constatering van een disfunctioneren en poging tot reparatie.

Er zijn twee verschillende niveau's waarop je over een kennissysteem kunt nadenken.

- d) Hoe heet het niveau waarop vraag 2c) relevant is?

Kennisniveau.

- e) Geef de naam van het andere niveau en geef twee vragen voor dit systeem die op dat niveau relevant zijn.

Symboolniveau.

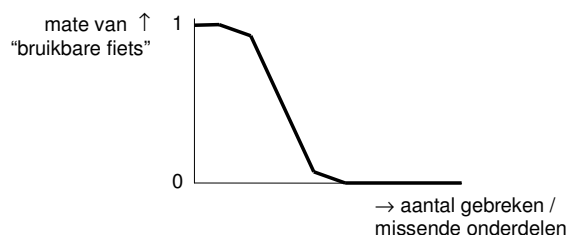
Hoe representeren we kennis.

Welke zoekalgoritmen gebruiken we?

Opgave 3: 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 bepaalt aan de hand van het aantal gebreken of missende onderdelen 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.

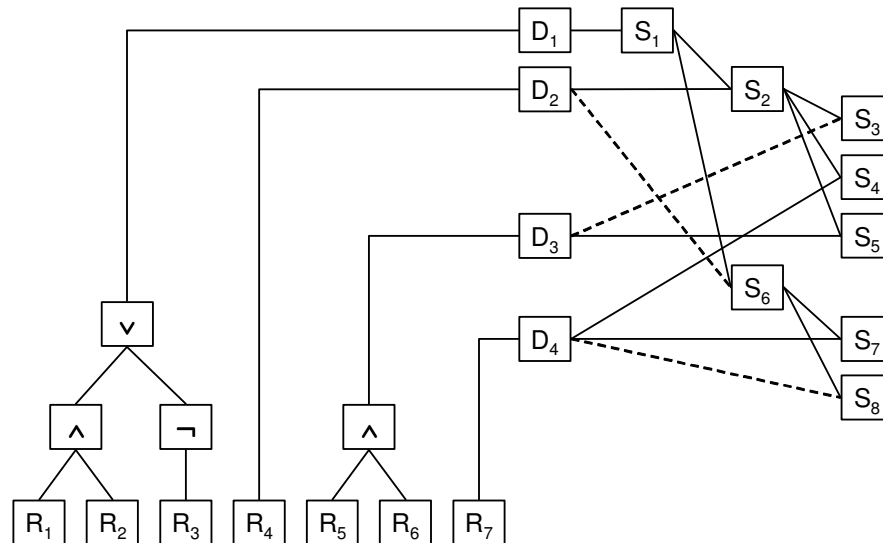
- a) Schets de grafiek voor het vage concept "bruikbare fiets" en leg uit wat er op de X-as en op de Y-as van de grafiek staat.



- 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, $f_{\text{wrak}}(X) = 1 - f_{\text{bruikbaar}}(x)$

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 geclassificeerd in oplossingsclasses (S’s). De oplossingsclasses met een gemeenschappelijke ‘ouder’ zijn onderling onverenigbaar (*mutual exclusive*).



c) Geef een voorbeeld wat R_5 , R_6 en D_3 zouden kunnen zijn 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.

Opgave 4: Onzekerheid

Een van de manieren om met onzekerheid om te gaan in kennissystemen is met behulp van kansrekening. Hierbij wordt vaak gebruik gemaakt van de regel van Bayes.

a) Waarom is de regel van Bayes vaak nuttig?

Omdat we vaak weten hoe vaak symptoom X gegeven Y voorkomt, terwijl we geïnteresseerd zijn in de kans op Y gegeven symptoom X .

Stel dat we weten dat de kans op koorts 1% is. We weten verder dat 90% van de mensen met koorts rode vlekken heeft. We nemen aan dat deze kansen gelijk zijn voor mensen die een tentamen doen en mensen die dat niet doen. We weten ook dat 50% van de mensen die tentamen doen rode vlekken hebben en dat slechts 5% van de mensen die geen tentamen doen rode vlekken hebben.

b) Je kijkt tijdens het tentamen naast je en ziet dat je buurman rode vlekken heeft. Hoe groot is de kans dat 'ie koorts heeft? Laat je berekening zien.

*Gevraagd: kans op koorts gegeven rode vlekken $P(K|R)$. We weten de kans op rode vlekken gegeven koorts $P(R|K) = 0.9$. We weten ook dat kans op koorts $P(K) = 0.01$. Bayes' regel zegt dat $P(K|R) = P(R|K) * P(K) / P(R)$. Tijdens een tentamen is de kans op rode vlekken $P(R) = 0.5$*

*De kans op koorts is dan $P(K|R) = 0.9 * 0.01 / 0.5 = 0.009 * 2 = 0.018 = 1,8 \%$.*

c) Je zit in de bus en ziet iemand met rode vlekken. Hoe groot is de kans dat 'ie koorts heeft? Laat je berekening zien.

*Zelfde argumentatie. Enige verschil is dat kans op rode vlekken zonder tentamen $P(R) = 0.05$. De kans op koorts is dan $P(K|R) = 0.9 * 0.01 / 0.05 = 0.009 * 20 = 0.18 = 18 \%$.*

Einde tentamen.