Typet lambdakalkyle og semantikk for naturlige språk

INF5820, H2008

Poenget med disse sidene er å gi noen definisjoner som forankrer kapittel 19 i Jurafsky og Martin litt bedre formelt.

Definisjon 1 ***Typer***

- *e og t er basistyper*
- *Hvis $\alpha_1, \alpha_2, \ldots, \alpha_n, \beta$ er typer, så er $\langle \alpha_1, \alpha_2, \ldots, \alpha_n \rangle \rightarrow \beta$ en type.*

Definisjon 2 ***Vokabular***

- *Vokabularet eller leksikonet for et gitt språk består av en mengde konstanter. Hver konstant har en tilhørende type, og det er bare endelig mange typer som har konstanter.*
- *For hver type τ finnes det tellbart mange variable av type τ: $x_\tau^1, x_\tau^2, \ldots$. Som konvensjon bruker vi små bokstaver for variable (og konstanter) av type e, store bokstaver for de av type $e \rightarrow t$, osv.*

Definisjon 3 ***Termer***

- *Variable og konstanter av type τ er termer av type τ. Noen ganger vil vi skrive $t : \tau$ for å si at t er av type τ.*
- *$s(t_1, \ldots, t_n) : \beta$ hvis $t_1 : \alpha_1, t_2 : \alpha_2, \ldots, t_n : \alpha_n, s : \langle \alpha_1, \alpha_2, \ldots, \alpha_n \rangle \rightarrow \beta$.*
- *Hvis x er en variabel, $x : \alpha$, og t en term, $t : \beta$, så $\lambda x[t] : \langle \alpha, \beta \rangle$.*

Dette er det vi trenger for å definere syntaksen. La oss også ta med litt semantikk.

Definisjon 4 ***Domener***

- *Til enhver type τ svarer det et domene D_τ.*
- *D_e er en ikke-tom mengde.*
- *D_t er sannhetsverdiene $\{sann, usann\}$, ofte skrevet $\{1, 0\}$.*
- $D_{\langle \alpha_1, \alpha_2, \ldots, \alpha_n \rangle \rightarrow \beta} = D_\beta^{D_{\alpha_1} \times D_{\alpha_2} \times \cdots \times D_{\alpha_n}}$

Definisjon 5 ***Struktur*** *En struktur (modell) består av et grunndomene D_e og en denotasjonsfunksjon V som tilordner et element i D_τ til enhver konstant av type τ.*

Definisjon 6 ***Variabeltilordning*** *En variabeltilordning g tilordner et element i D_τ til hver variabel av type τ.*

Definisjon 7 ***Tolkningen $t_{V,g}$ av en term t i en struktur $\langle D_e, V\rangle$ relativt til en variabeltilordning g:***

- *Hvis t er en konstant, så er $t_{V,g} = V(t)$*
- *Hvis t er en variabel, så er $t_{V,g} = g(t)$*
- $s(t_1, \ldots, t_n)_{V,g} = s_{V,g}(t_{1,V,g}, t_{2,V,g} \ldots t_{n,V,g})$
- *Hvis $(\lambda x[t]) : \langle \alpha, \beta\rangle$, så er $(\lambda x[t])_{V,g}$ den funksjonen som til hvert element $a \in \alpha$ gir $(t[\underline{a}/x])_{V,g}$ der $\underline{a}$ er et navn på a og $(t[\underline{a}/x])$ er resultatet av å sette inn $\underline{a}$ for alle fri forekomster av x i t.*

Dette er tilstrekkelig til å definere den varianten av typet lambdakalkyle vi trenger. Ofte vil en se at typen $\langle \alpha_1, \alpha_2, \ldots, \alpha_n\rangle \to \beta$ begrenses til at $n = 1$ eller mao. et alle funksjonelle typer er unære og på formen $\langle\alpha\rangle \to \beta$, som da kan skrives $\alpha \to \beta$. Grunnen er at f.eks. en binær funksjon $f : \langle \alpha_1, \alpha_2\rangle \to \beta$ kan erstattes av en unær funksjon $f' : \alpha_1 \to (\alpha_2 \to \beta)$, der $f'(t)(s) = f(t, s)$.

Logikk i dette

Vi kan uttrykke (høyere-ordens) logikk innefor denne rammen. Tanken er at termer av type t skal være logiske formler. Det unære konnektivet $\neg$ kan da innføres som en konstant av type $t \to t$. Tilsvarende kan vi innføre binære konnektiver som termer av type $\langle t, t\rangle \to t$, altså $\wedge : (\langle t, t\rangle \to t)$ og $\vee : (\langle t, t\rangle \to t)$. Vi må legge spesielle krav på tolkningen av disse konnektivene slik vi gjør i vanlig logikk. For eksempel vil $V(\neg)$ være funksjonen som sender 1 til 0 og 0 til 1.

Vikan innføre kvantifikasjon som en kombinasjon av en konstant og et lambdauttrykk. Vi kan for hver type α innføre en term $\forall : ((\alpha \to t) \to t)$ med tolkningen at $\forall_{(\alpha\to t)\to t),\ V,g}(X) = 1$ hvis og bare hvis $X = D_{\alpha\to t}$.