1. Formaalsed keeled ja seosed

Teoreetiline informaatika on täppisteadus, mis kasutab oluliselt diskreetse matemaatika mõisteid ja tulemusi.

Järgnevas esitame kursuse jooksul enim kasutatavate formaalsete keelte alaste mõistete definitsioonid:

sümbol – mittedefineeritav (nagu punkt ja sirge geomeetrias). Sümbolid esitatakse graafiliste kujunditena; käesolevas konspektis kasutatavateks sümboliteks on enamasti ladina või kreeka tähed (nii suur- kui väiketähed, nii indeksiga kui ilma) ning numbrid;
tähestik või **alfabeet**– lõplik sümbolite hulk;
sõna või **string** – lõplik sümbolite jada (ilma nendevaheliste tühikuteta). Tavaliselt on tähestikku moodustavad tähed lineaarselt järjestatud. Selle järjestuse abil saab ka sõnade vahel defineerida nn. **leksikograafilise järjestuse** $<<$: $X_1 \ldots X_n << Y_1 \ldots Y_m$ parajasti siis, kui võrreldavates sõnades esimese vastavatel kohtadel oleva teineteisest erineva tähepaari (X_i, Y_i) korral $X_i < Y_i$, kus $<$ on tähestikus defineeritud järjestus.
sõna w pikkus $|w|$ – sümbolite arv sõnas w; ϵ – tühi sõna ($|\epsilon| = 0$);
sõna **prefiks** – sõna mistahes algussümbolitest koosnev osasõna. Näiteks sõna $tigu$ prefiksiteks on ϵ, t, ti, tig ja $tigu$.
sõna **sufiks** – sõna mistahes lõpusümbolitest koosnev osasõna;
formaalne keel – mingisse tähestikku kuuluvatest sümbolitest moodustatud sõnade hulga osahulk (näit. tähestikus Σ kõikide sõnade hulk Σ^*);
w^R - sõna, mis on saadud sõnast w tähtede järjekorra vastupidiseks muutmise teel, näiteks $abbcihh^R = hhicbba$;
palindroom – sõna w, mille korral $w^R = w$ (s.t. sõna, mis ühtib oma tähtede vastupidises järjekorras kirjutamisel saadud sõnaga);
sõnade **korrutis e. konkatenatsioon** – sõna, mis on saadud antud sõnade järjest kirjutamisel. Kui L_1 ja L_2 on mingid keeled tähestikus Σ, siis defineerime nende **korrutise** $L_1L_2 = \{xy|x \in L_1, y \in L_2\}$, aga samuti $L^0 = \{\epsilon\}$ ning $L^{i+1} = LL^i$, $i > 0$;
Hulka $L^* = \cup_{i=0}^{\infty} L^i$ nimetatakse keele L **Kleene'i katteks**. Defineerime ka $L^+ = \cup_{i=1}^{\infty} L^i$.

Näide. Olgu $\Sigma = \{0, 1\}$ ning $L_1 = \{01, 11\}$ ja $L_2 = \{1, 10\}$. Siis $L_1L_2 = \{011, 0110, 111, 1110\}$ ning $L_2^* = \{\epsilon, 1, 10, 11, 110, 1010, \ldots\}$.

1.1. Ülesanne. Olgu antud tähestik $\Sigma = \{\emptyset, 2, x, \alpha\}$, kusjuures sümbolite vahel on defineeritud järgmine järjestusseos: $\emptyset < 2 < x < \alpha$. Leida
1) sõnade $x\alpha\alpha$, $\alpha2\emptyset\emptyset$ ja $x\alpha2$ vaheline leksikograafiline järjestus;
2) väikseim ja suuruselt eelviimane kolmesümboliline sõna;
3) sõna $\alpha2\emptyset\emptyset22$ kõikide prefiksite hulk ja kõikide sufiksite hulk;
4) mingi kuuesõnaline keel tähestikus Σ;
5) mingi lõpmatu keel tähestikus Σ.

1.2. Ülesanne. Leida lühim sõna w tähestikus $\{a, j, k, r, s\}$, mille korral $w^R = w$ ja mis saadakse järgmistest sõnadest neile tähtede vahelekirjutamise teel: a) $kajakas$; b) $rajakas$. Formuleeri ülesande lahendamisel kasutatud algoritm.

1.3. Ülesanne (iseseisvalt). Leida minimaalne tähtede arv, millega saab tähestikus {a, b, i, k, s} kirjutatud sõna "kassisaba" täiendada palindroomini. Millise palindroomi saate?

1.4. Ülesanne. Olgu L_1={suur, väike, kodu} ja L_2={ahv, rätik}. Leida L_1L_2.

1.5. Ülesanne. Kirjutada välja kõik keele $L = \{a, b\}$ Kleene'i kattesse kuuluvad sõnad, mille pikkus on väiksem kui 3.

Hulgateoreetiliste mõistete osas kasutame järgmisi tähiseid:

$\mathcal{P}(X)$ või 2^X – hulga X osahulkade hulk;
$X - Y = \{x \in X | x \notin Y\}$ - hulkade X ja Y hulgateoreetiline vahe;
$\mathcal{N} = \{1, 2, \ldots\}$ - naturaalarvude hulk ($\mathcal{N}^0 = \{0, 1, 2, \ldots\}$).

Olgu $X_1, X_2, \ldots, X_n$ mingid hulgad. Hulka $X_1 \times \ldots \times X_n = \{(x_1, \ldots, x_n) | \, x_1 \in X_1, \ldots, x_n \in X_n\}$ nimetatakse hulkade $X_1, \ldots, X_n$ *ristkorrutiseks* või *otsekorrutiseks*. Juhul, kui $X_i = X, i = 1, \ldots, n$, siis tähistatakse nende otsekorrutist sümboliga X^n. Hulga X^n mistahes osahulka nimetatakse n-aarseks seoseks hulgal X. Juhul, kui $n = 2$, räägitakse **binaarsest seosest**. Seda, et elemendid x ja y on seoses R, tähistatakse kas xRy või $(x, y) \in R$.

Binaarset seost R hulgal X nimetatakse:
1) *refleksiivseks*, kui xRx iga $x \in X$ korral,
2) *sümmeetriliseks*, kui mistahes $x, y \in X$ korral xRy parajasti siis, kui yRx,
3) *transitiivseks*, kui sellest, et xRy ja yRz järeldub, et ka xRz.
Refleksiivset, sümmeetrilist ja transitiivset seost nimetatakse *ekvivalentsiseoseks*. Kui hulgal X on defineeritud mingi ekvivalentsiseos R, siis jaguneb hulk X paarikaupa lõikumatuteks ekvivalentsiklassideks, kus ühte klassi kuuluvad parajasti paarikaupa omavahel seoses R olevad elemendid. Iga ekvivalentsiklass on määratud oma mistahes elemendi poolt, mistõttu ekvivalentsiklassi kuuluva mistahes elemendi võib valida selle klassi esindajaks.

Näide. Olgu hulgaks X naturaalarvude hulk. Defineerime hulgal X seose $/_7$ järgmiselt: loeme, et $x/_7y$ parajasti siis, kui $x - y$ jagub arvuga 7.

1.6. Ülesanne. Tähistagu hulk A Tallinna linnaliinide bussipeatuste hulka. Defineerime peatustevahelise seose ρ järgmiselt: $x\rho y$ parajasti siis, kui leidub bussiliin, mille bussid sõidavad peatusest x peatusse y. Kas seos ρ on refleksiivne, sümmeetriline, transitiivne?

1.7. Ülesanne. Olgu $H = \{a, b, c\}$ ja $X = \{x, y\}$. Leida:
1) ristkorrutis $H \times X$;
2) millised omadustest (refleksiivsus, sümmeetria, transitiivsus) on seosel
$\rho = \{(b, b), (b, a), (c, a), (a, a), (a, c), (a, b)\}$;
3) mingi (binaarne) sümmeetriline seos hulgal H;
4) mingi (binaarne) transitiivne seos hulgal H;
5) ülesandes 2) antud binaarse seose ρ ja seose $\tau = \{(b, x), (c, y)\}$ korrutis $\rho\tau$;
6) ρ^2, ρ^3 ja ρ^4, kus ρ on ülesandes 2) antud binaarne seos.

Olgu P mingi binaarsete seoste omaduste hulk. Seose R **P-katteks** nim. minimaalset seost $R' \supseteq R$, mis rahuldab kõiki omadusi hulgast P.
Seoste $P \subseteq X_1 \times X_2$ ja $Q \subseteq X_2 \times X_3$ korrutiseks nimetatakse seost $P \circ Q = \{(a, b) \in X_1 \times X_3 |$ leidub element $x \in X_2$ nii, et $(a, x) \in P; (x, b) \in Q\}$.

Kui P=$\{transitiivsus\}$, siis binaarse seose R **transitiivne** kate R^+ on leitav järgmise algoritmi abil:
1) $(a, b) \in R \Rightarrow (a, b) \in R^+$;
2) $(a, b) \in R^+, (b, c) \in R \Rightarrow (a, c) \in R^+$;
3) R^+ sisaldab vaid aksioomide 1) ja 2) abil saadud elemente.
Seose R transitiivse katte võib kirjutada ka kujul $R^+ = \cup_i \underbrace{R \circ R \circ \ldots \circ R}_{i}$, kus $i > 0$.

1.8. Ülesanne. Olgu $H = \{a; b; c; d\}$ ja olgu $Q = \{(a, b); (d, a); (b, d); (c, a)\}$. Leida seose Q P-kate, kui 1) P=$\{refleksiivsus\}$ ja 2) P=$\{transitiivsus\}$.

1.9. Ülesanne. Olgu naturaalarvude hulgal defineeritud seos $Q = \{(5, 6); (6, 7); (7, 8)\}$ ning olgu P=$\{$*refleksiivsus, sümmeetria, transitiivsus*$\}$. Millised on seose Q P-katte ekvivalentsiklassid?