Tietojenkäsittelyteorian perusteet T/Y
T-79.1001/1002

Pekka Orponen

Tietojenkäsittelyteorian laboratorio, TKK

Syksy 2005

T–79.1001/1002 Tietojenkäsittelyteorian perusteet T/Y
Introduction to Theoretical Computer Science T/Y

What can one do with a computer?

This course divides roughly into three parts. We examine three models of a computer, their power and limitations, i.e., the classes of languages that they can decide.

It turns out that each class of *automata* corresponds to a class of *grammars*. The classes of automata and grammars are:

1. Finite state automata & regular expressions
2. Pushdown automata & context-free grammars
3. Turing machines & unrestricted grammars

The first teaching period of the course (=Y version) covers finite state automata, regular expressions and context-free grammars. The second period covers the rest.

Course versions

The two-period (T) version corresponds to old course T-79.148. It is compulsory in the computer science curriculum, and in many specialisations of the pre-2005 telecommunications engineering curriculum.

The one-period (Y) version covers the most broadly useful tools in the area. It is a compulsory basic (P) course in the post-2005 telecommunications engineering program.

Practical arrangements

Registration: obligatory, by TOPI

Lectures: Thu 14–16 T1, in Finnish by Pekka Orponen

Tutorials: Tue 12–13, 13–14, 16–17 (in English), Tue 17–18, Wed 12–13, 13–14, 14–15, 15–16 — choose one and register by TOPI

Computer exercises: obligatory, using WWW system "Regis"

Course home page:
`http://www.tcs.hut.fi/Studies/T-79.1001/`

Course newsgroup: `opinnot.tik.tkt` on `news.tky.hut.fi`

- Please ask (and answer!) your questions there!

To pass the course

1. Pass the computer exercises *before taking the exam*.
2. *After passing the computer exercises* pass the exam by scoring at least 30/60 (exams typically in May, Aug, Oct, Dec, Mar)
   - You may earn up to 6 points on the exam by participating in the exercises.
   - 3 homework problems a week, 1 exam point for every 5 homework problems prepared for exercise session
3. For the Y version of the course, exam is graded to 30 points, and max number of exercise points is 3.

Material

Lecture notes (in Finnish) and solutions of *demonstration* exercises (in Finnish) are distributed through Edita.

Recommended textbook Michael Sipser, Introduction to the Theory of Computation, PWS Publishing 1997. (supplementary for Finnish-speaking students; likely necessary for non-Finnish-speaking students)

Some additional material on the course WWW page.

TIETOJENKÄSITTELYTEORIA

- Matemaattinen oppi siitä, mitä tietokoneella on mahdollista tehdä ja kuinka tehokkaasti.
- Tarjoaa matemaattisia käsitteitä ja menetelmiä tietojenkäsittelyjärjestelmien mallintamiseen ja analysointiin sekä selkeiden ja tehokkaiden ratkaisujen laatimiseen.

Tietojenkäsittelyteorian osa-aloja

Laskettavuusteoria
: Mitä tietokoneella voi tehdä periaatteessa?
- Turing, Gödel, Church, Post (1930-luku); Kleene, Markov (1950-luku).

Laskennan vaativuusteoria
: Mitä tietokoneella voi tehdä käytännössä?
- Hartmanis, Stearns (1960-luku); Cook, Levin, Karp (1970-luku); Papadimitriou, Sipser, Håstad, Razborov ym. (1980-).

Automaatti- ja kielioppiteoria
: Tietojenkäsittelyjärjestelmien perustyyppien ominaisuudet ja kuvausformalismit.
- Chomsky (1950-luku); Ginsburg, Greibach, Rabin, Salomaa, Schützenberger ym. (1960-luku)

Ohjelmien oikeellisuus
: Tietojenkäsittelyjärjestelmien matemaattisesti eksakti määrittely ja oikean toiminnan verifiointi.
- Dijkstra, Hoare (1960-luku); Manna, Pnueli, Scott ym. (1970-).

Muuta
- algoritmien suunnittelu ja analyysi (Knuth, Hopcroft, Tarjan ym.)
- kryptologia (Rivest, Shamir, Adleman ym.)
- rinnakkaisten ja hajautettujen järjestelmien teoria (Lamport, Lynch, Milner, Valiant ym.)
- koneoppimisteoria (Valiant ym.)
- jne.

Tällä kurssilla käsitellään lähinnä automaatteja ja kielioppeja sekä hieman laskettavuusteorian alkeita. Muita aiheita käsitellään Tietojenkäsittelyteorian laboratorion muilla kursseilla.

1. Matemaattisia peruskäsitteitä

1.1 Joukot

Joukko (engl. set) on kokoelma alkioita. Alkiot voidaan ilmoittaa joko luettelemalla, esim.

$$S = \{2, 3, 5, 7, 11, 13, 17, 19\}$$

tai jonkin säännön avulla, esim.

$$S = \{p \mid p \text{ on alkuluku}, 2 \leq p \leq 20\}.$$

Jos alkio a kuuluu joukkoon A, merkitään $a \in A$, päinvastaisessa tapauksessa $a \notin A$. (Esim. $3 \in S$, $8 \notin S$.)

Tärkeä erikoistapaus on *tyhjä joukko* (engl. empty set) $\emptyset$, johon ei kuulu yhtään alkiota.

Jos joukon A kaikki alkiot kuuluvat myös joukkoon B, sanotaan että A on B:n *osajoukko* (engl. subset) ja merkitään $A \subseteq B$. [Kirjallisuudessa esiintyy myös merkintä $A \subset B$.] Jos A *ei* ole B:n osajoukko merkitään $A \not\subseteq B$. Siis esim.

$$\{2, 3\} \subseteq S, \qquad \{1, 2, 3\} \not\subseteq S.$$

Triviaalisti on voimassa $\emptyset \subseteq A$ kaikilla A.

Joukot A ja B ovat samat, jos niissä on samat alkiot, so. jos on $A \subseteq B$ ja $B \subseteq A$. Jos on $A \subseteq B$, mutta $A \neq B$, sanotaan että A on B:n *aito osajoukko* (engl. proper subset) ja merkitään $A \subsetneq B$. Edellä olisi siis voitu myös kirjoittaa $\{2, 3\} \subsetneq S$ ja $\emptyset \subsetneq A$ jos $A \neq \emptyset$.

Joukon alkioina voi olla myös toisia joukkoja (tällöin puhutaan usein “joukkoperheestä”), esim.

$$X = \{\emptyset, \{1\}, \{2\}, \{1,2\}\}.$$

Jonkin perusjoukon A kaikkien osajoukkojen muodostamaa joukkoperhettä sanotaan A:n *potenssijoukoksi* (engl. powerset) ja merkitään $\mathcal{P}(A)$:lla; esim. edellä on $X = \mathcal{P}(\{1,2\})$. [Koska n-alkioisen perusjoukon A potenssijoukossa on 2^n alkiota (HT), käytetään kirjallisuudessa potenssijoukolle myös merkintää 2^A.]

Huomaa, että $A \subseteq B$ jos ja vain jos $A \in \mathcal{P}(B)$.

Joukkoja voidaan kombinoida *joukko-operaatioilla*, joista tärkeimmät ovat:

yhdiste (engl. union)

$$A \cup B = \{x \mid x \in A \text{ tai } x \in B\},$$

esim. $\{1,2,3\} \cup \{1,4\} = \{1,2,3,4\}$.

leikkaus (engl. intersection)

$$A \cap B = \{x \mid x \in A \text{ ja } x \in B\},$$

esim. $\{1,2,3\} \cap \{1,4\} = \{1\}$.

erotus (engl. difference)

$$A - B = \{x \mid x \in A \text{ ja } x \notin B\},$$

esim. $\{1,2,3\} - \{1,4\} = \{2,3\}$.
[Erotukselle käytetään myös merkintää $A \setminus B$.]

Joukko-operaatioita koskevat tietyt laskulait, joista tärkeimmät ovat yhdisteen ja leikkauksen *liitännäisyys*:

$$A \cup (B \cup C) = (A \cup B) \cup C, \quad A \cap (B \cap C) = (A \cap B) \cap C$$

ja *vaihdannaisuus*:

$$A \cup B = B \cup A, \quad A \cap B = B \cap A$$

sekä näiden *osittelulait*:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C),$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

Jos kaikki tarkasteltavat joukot ovat jonkin yhteisen “universaalijoukon” U osajoukkoja, sanotaan erotusta $U - A$ joukon A *komplementiksi* (U:n suhteen) ja merkitään $\bar{A}$:lla.

Yhdiste-, leikkaus- ja komplementointioperaatioita yhdistävät tärkeät ns. *de Morganin kaavat*:

$$\overline{A \cup B} = \bar{A} \cap \bar{B},$$

$$\overline{A \cap B} = \bar{A} \cup \bar{B}.$$

Lisäksi joukkojen erotus voidaan esittää leikkauksen ja komplementoinnin avulla seuraavasti:

$$A - B = A \cap \bar{B}.$$

Jos joukkoperheen $\mathcal{A}$ jäsenet on *indeksoitu*, esim.

$$\mathcal{A} = \{A_1, A_2, A_3, \dots\},$$

niin yhdisteelle ja leikkaukselle voidaan käyttää lyhennemerkintöjä

$$\bigcup_{i \geq 1} A_i = A_1 \cup A_2 \cup A_3 \cdots \qquad \text{ja} \qquad \bigcap_{i \geq 1} A_i = A_1 \cap A_2 \cap A_3 \cdots$$

Indeksien ei tarvitse olla edes luonnollisia lukuja, vaan *indeksijoukkona* voi olla mikä tahansa joukko I. Tällöin käytetään merkintöjä

$$\mathcal{A} = \{A_i \mid i \in I\}$$

ja

$$\bigcup_{i \in I} A_i, \qquad \bigcap_{i \in I} A_i.$$

1.2 Relaatiot ja funktiot

Olkoot A ja B joukkoja. Alkioiden $a \in A$ ja $b \in B$ *järjestettyä paria* (engl. ordered pair) merkitään (a, b). Huomaa, että joukkoina on aina $\{a, b\} = \{b, a\}$, mutta jos $a \neq b$, niin järjestettyinä pareina on $(a, b) \neq (b, a)$.

Joukkojen A ja B *karteesinen tulo* (engl. Cartesian product) määritellään

$$A \times B = \{(a, b) \mid a \in A \text{ ja } b \in B\},$$

esim.

$$\begin{aligned} & \{1,2,3\} \times \{1,4\} \\ = \ & \{(1,1),(1,4),(2,1),(2,4),(3,1),(3,4)\}. \end{aligned}$$

Relaatio R joukolta A joukolle B on karteesisen tulon $A \times B$ osajoukko:

$$R \subseteq A \times B.$$

Jos $(a, b) \in R$, niin merkitään myös aRb ja sanotaan että alkio a on *relaatiossa (suhteessa)* R alkioon b. Tätä *infix*-merkintää käytetään varsinkin silloin, kun relaation nimenä on jokin erikoismerkki, esim. $\leq$, $\prec$, $\equiv$, $\sim$.

Jos relaation R *lähtöjoukko* (engl. domain) A ja *maalijoukko* (engl. range) B ovat samat, so. $R \subseteq A \times A$, sanotaan että R on relaatio *joukossa* A.

Relaation $R \subseteq A \times B$ *käänteisrelaatio* (engl. inverse relation) on relaatio $R^{-1} \subseteq B \times A$,

$$R^{-1} = \{(b, a) \mid (a, b) \in R\}.$$

Jos $R \subseteq A \times B$ ja $S \subseteq B \times C$ ovat relaatioita, niin niiden *yhdistetty relaatio* (engl. composite relation) $R \circ S \subseteq A \times C$ määritellään:

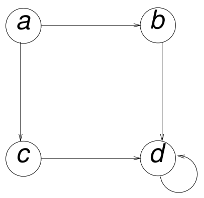
$$R \circ S = \{(a, c) \mid \exists b \in B \text{ s.e. } (a, b) \in R, (b, c) \in S\}.$$

Varsinkin jos joukot A ja B ovat äärellisiä, relaatiota $R \subseteq A \times B$ voi olla havainnollista tarkastella *suunnattuna verkkona* t. *graafina*, jonka *solmuina* ovat joukkojen A ja B alkiot ja solmusta $a \in A$ on *kaari* ("nuoli") solmuun $b \in B$, jos ja vain jos $(a,b) \in R$.

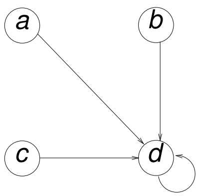
Olkoon esimerkiksi joukossa $A = \{a,b,c,d\}$ määritelty relaatio $R \subseteq A \times A$,

$$R = \{(a,b),(a,c),(b,d),(c,d),(d,d)\}.$$

Relaation R graafiesitys on



Relaation $R \circ R$ graafiesitys puolestaan on



Relaatio $f \subseteq A \times B$ on *funktio*, jos kukin $a \in A$ on relaatiossa f täsmälleen yhden $b \in B$ kanssa. Tällöin käytetään yleisten relaatiomerkintöjen sijaan tavallisesti merkintöjä $f : A \to B$ ja $f(a) = b$.

Funktioita koskee kaikki mitä edellä yleisesti on todettu relaatioista, mutta historiallisista syistä funktioiden yhdistäminen merkitään toisin päin kuin yleisten relaatioiden: jos $f : A \to B$ ja $g : B \to C$ ovat funktioita, niin niiden yhdistetty funktio määritellään kaavalla $(g \circ f)(a) = g(f(a))$, so. relaatioina

$$g \circ f = \{(a,c) \mid \exists b \in B \text{ s.e. } f(a) = b, g(b) = c\}.$$

Olkoon $f : A \to B$ funktio.

- f on *surjektio* (engl. onto map), jos jokainen $b \in B$ on jonkin $a \in A$ kuva, so. jos $f(A) = B$.
- f on *injektio* (engl. one-to-one map), jos kaikki $a \in A$ kuvautuvat eri alkioille, so. jos $a \neq a' \Rightarrow f(a) \neq f(a')$.
- f on *bijektio*, jos se on sekä injektio että surjektio, so. jos jokainen $b \in B$ on yhden ja vain yhden $a \in A$ kuva.

1.3 Ekvivalenssirelaatiot

Ekvivalenssirelaatiot ovat matemaattisesti täsmällinen muotoilu sille yleiselle idealle, että oliot ovat keskenään *samankaltaisia* jonkin kiinnostavan ominaisuuden X suhteen. Ominaisuuteen X perustuva ekvivalenssirelaatio osittaa tarkasteltavien olioiden joukon *ekvivalenssiluokkiin*, jotka vastaavat ominaisuuden X eri arvoja. (Kääntäen mielivaltainen olioiden joukon ositus Π määrää tietyn abstraktin samankaltaisuusominaisuuden, nim. sen että oliot ovat samankaltaisia jos ne sijoittuvat samaan osituksen Π luokkaan.)

Osoittautuu, että yleinen “samankaltaisuusrelaation” idea voidaan kiteyttää seuraaviin kolmeen ominaisuuteen.

Määritelmä 1.1 Relaatio $R \subseteq A \times A$ on

1. *refleksiivinen*, jos aRa $\forall a \in A$;
2. *symmetrinen*, jos $aRb \Rightarrow bRa$ $\forall a, b \in A$;
3. *transitiivinen*, jos $aRb, bRc \Rightarrow aRc$ $\forall a, b, c \in A$.

Määritelmä 1.2 Relaatio $R \subseteq A \times A$, joka toteuttaa edelliset ehdot 1–3 on *ekvivalenssirelaatio*. Alkion $a \in A$ *ekvivalenssiluokka* (relaation R suhteen) on

$$R[a] = \{x \in A \mid aRx\}.$$

Ekvivalenssirelaatioita merkitään usein R:n sijaan alkioiden samankaltaisuutta korostavilla symboleilla $\sim$, $\equiv$, $\simeq$ tms.

Esim. Olkoon

$$A = \{\text{kaikki 1900-luvulla syntyneet ihmiset}\}$$

ja aRb voimassa, jos henkilöillä a ja b on sama syntymävuosi. Tällöin R on selvästi ekvivalenssi, jonka ekvivalenssiluokat koostuvat keskenään samana vuonna syntyneistä henkilöistä. Luokkia on 100 kappaletta, ja “abstraktisti” ne vastaavat 1900-luvun vuosia 1900, …,1999.

Lemma 1.3 Olkoon $R \subseteq A \times A$ ekvivalenssi. Tällöin on kaikilla $a, b \in A$ voimassa:

$$R[a] = R[b] \quad \text{joss} \quad aRb.$$

Tod. Helppo; sivuutetaan. □

Lemma 1.4 Olkoon $R \subseteq A \times A$ ekvivalenssi. Tällöin R:n ekvivalenssiluokat muodostavat A:n *osituksen* erillisiin epätyhjiin osajoukkoihin, so.:

- $R[a] \neq \emptyset$ kaikilla $a \in A$;
- $A = \bigcup_{a \in A} R[a]$;
- jos $R[a] \neq R[b]$, niin $R[a] \cap R[b] = \emptyset$, kaikilla $a, b \in A$.

Tod. Helppo; sivuutetaan. □

Kääntäen jokainen perusjoukon A ositus erillisiin epätyhjiin luokkiin A_i, $i \in I$, määrää vastaavan ekvivalenssirelaation:

$$a \sim b \quad \Leftrightarrow \quad a \text{ ja } b \text{ kuuluvat samaan luokkaan } A_i.$$

1.5 Induktiopäättely

Olkoon $P(k)$ jokin luonnollisten lukujen ominaisuus. Jos on voimassa:

1. $P(0)$ ja
2. kaikilla $k \geq 0$:

$$P(k) \Rightarrow P(k+1),$$

niin $P(n)$ on tosi kaikilla $n \in \mathbb{N}$. □

Esimerkki.
Väite. Kaikilla $n \in \mathbb{N}$ on voimassa kaava

$$P(n): \quad (1+2+\cdots+n)^2 = 1^3+2^3+\cdots+n^3.$$

Todistus.

1. Perustapaus: $P(0): \quad 0^2 = 0.$

(jatkuu)

T–79.1001/1002 Tietojenkäsittelyteorian perusteet

2. Induktioaskel: Oletetaan, että annetulla $k \geq 0$ kaava

$$P(k): \quad (1+2+\cdots+k)^2 = 1^3+2^3+\cdots+k^3$$

on voimassa. Tällöin on myös:

$$\begin{aligned}
&(1+2+\cdots+k+(k+1))^2 \\
&= (1+\cdots+k)^2 + 2(1+\cdots k)(k+1) + (k+1)^2 \\
&= 1^3+\cdots+k^3 + 2\cdot\frac{k(k+1)}{2}\cdot(k+1) + (k+1)^2 \\
&= 1^3+\cdots+k^3 + k(k+1)^2 + (k+1)^2 \\
&= 1^3+\cdots+k^3 + (k+1)^3.
\end{aligned}$$

On siis todettu, että kaavan $P(k)$ totuudesta seuraa kaavan $P(k+1)$ totuus, so. että $P(k) \Rightarrow P(k+1)$, kaikilla $k \geq 0$. Luonnollisten lukujen induktioperiaatteen 1.9 nojalla voidaan nyt päätellä, että kaava $P(n)$ on voimassa kaikilla $n \in \mathbb{N}$. $\square$

Pekka Orponen syksy 2005