

Luento 6 Tiedon esitysmuodot

Lukujärjestelmät
Kokonaisluvut
Liukuluvut
Merkit, merkkijonot
Totuusarvot
Kuvat, äänet, hajut(?)

3.4.2004 Copyright Teemu Kerola 2003 1

Tiedon tyypit ⁽³⁾

- **Kommunikointi ihmisen kanssa**
 - kuva, ääni, merkit, ...
- **Laitteiston sisäinen talletus**
 - kuvaformaatit, ääniformaatit, pakkausstandardit, ...
 - kokonaisluvut, liukuluvut, merkit, merkkistöt
 - ohjelmat
- **Suorittimen omia lajinaan ymmärtämät tyypit**
 - on olemassa konekäskyjä tälle tietotyypille
 - kokonaisluvut
 - liukuluvut (useimmat suorittimet nykyään)
 - totuusarvot (jotkut suorittimet)
 - merkit (jotkut suorittimet)
 - konekäskyt

3.4.2004 Copyright Teemu Kerola 2003 2

Tiedon esitys laitteistossa ⁽³⁾

- **Kaikki tieto** koneessa on binääribitteinä (0 tai 1)
 - binäärijärjestelmän numerot: 0, 1
 - helppo toteuttaa piireillä
 - helppo suunnitella logiikkaa Boolean algebran avulla
- **Muisti jaettu tasapituisiin sanoihin** (word)
 - sana = word = 32 bittiä (16 bittiä, 64 bittiä, ...)
- **Yleensä sana on jaettu tasapituisiin 8-bittisiin tavuihin** (byte)

sana

1000 1101	1010 1101	1011 1100	1111 0001
tavu	tavu	tavu	tavu

3.4.2004 Copyright Teemu Kerola 2003 3

Tiedon esitys laitteistossa ⁽²⁾

- Tietoa siirretään muistiväylää pitkin sanoina
 - joskus useampi kuin yksi sana kerrallaan (lohko)
- Suorittimen rekisterit ovat yleensä yhden tai kahden sanan mittaisia
 - 1 sana: kokonaisluku, pieni liukuluku
 - 1 sana: 1 merkki tai 4 merkkiä
 - 2 sanaa: pitkä kokonaisluku, iso liukuluku

3.4.2004 Copyright Teemu Kerola 2003 4

Tiedon esitys ⁽³⁾

- Kysymys: miten esittää erityyppisiä tietoja?
- Vastaus: koodataan ne biteiksi
 - kaikki tieto on koneessa biteinä
- Kaikelle käsitellylle tiedolle on omat koodausmenetelmänsä
 - kaikkia koodausmenetelmiä ei ole standardoitu
 - samalla tietotyypillä voi olla useita koodausmenetelmiä
 - kokonaisluvut, liukuluvut, merkit, merkkijonot, kuvat, ...
 - ongelma: ymmärtävätkö koneet toisiaan?
 - tiedon esitysmuotoa voidaan joutua muuttamaan, kun tietoa siirretään koneelta toiselle

3.4.2004 Copyright Teemu Kerola 2003 5

Suorittimen ymmärtämä tieto ⁽⁵⁾

- Kaikki tieto koneessa on koodattuna biteiksi
- **Muistissa** voidaan esittää kaikki tieto millä tahansa sovitulla esitystavalla (koodauksella)
- **Suoritin** osaa tehdä operaatioita joillakin esitystavalla koodatuille tiedoille
 - kokonaisluvut ja liukuluvut (aina) TTK-91: kokonaisluvut
 - totuusarvot, merkit ja merkkijonot (joskus)
 - kuvat ja äänet (ei yleensä ellei erikoistunut suoritin)
 - hajut (ei vielä)
- Muiden tietojen käsittely tapahtuu ohjelmallisesti
 - esim. merkkejä (merkkien esitysmuotoa) voidaan käsitellä kokonaislukuoperaatioilla ja aliohjelmilla

3.4.2004 Copyright Teemu Kerola 2003 6

Binäärijärjestelmä (2)

- Kantaluku 2, numerot 0 ja 1
 - numeroiden painoarvot oikealta vasemmalle:
 $1=2^0, 2=2^1, 4=2^2, 8=2^3, 16=2^4, 32=2^5, \dots$
 - kymmenjärjestelmässä painoarvot ovat
 $1=10^0, 10=10^1, 100=10^2, 1000=10^3, \dots$

3.4.2004 Copyright Teemu Kerola 2003 7

Binäärilukuesimerkkejä (9)

3.4.2004 Copyright Teemu Kerola 2003 8

Binäärilukujen laskutoimitukset (3)

+	0	1
0	0	1
1	1	10

*	0	1
0	0	0
1	0	1

$$\begin{array}{r} 11 \\ 101101 \\ +1100 \\ \hline 111001 \end{array} \quad \begin{array}{l} 45 \\ +12 \\ =57 \end{array}$$

$$\begin{array}{r} 101 \\ *101 \\ \hline 101 \\ +101 \\ \hline 11001 \end{array} \quad \begin{array}{l} 5 \\ *5 \\ =25 \end{array}$$

3.4.2004 Copyright Teemu Kerola 2003 9

Binääripiste (3)

- Binääriluvuilla voi olla myös binääriosia (vrt. desimaaliosa)

3.4.2004 Copyright Teemu Kerola 2003 10

Binääripiste-esimerkkejä (10)

3.4.2004 Copyright Teemu Kerola 2003 11

Muunnokset lukujärjestelmien välillä (5)

- 2-järjestelmä \Rightarrow 10-järjestelmä
 - esitettiin jo edellä
- 10-järjestelmä \Rightarrow 2-järjestelmä
 - kokonaisosa ja desimaaliosa erikseen
 - kokonaisosa:
 - jaa toistuvasti 2:lla, kunnes 0 jäljellä
 - ota jakojäännökset käännetyssä järjestyksessä

3.4.2004 Copyright Teemu Kerola 2003 12

Big vs. Little Endian (3)

- Miten monitavuiset arvot talletetaan?

Sanan osoite: 0x1200: [] [] [] []
 talleta 0x11223344 ?? tavuosoitteet: 0x1200 0x1201 0x1202 0x1203

Big-Endian: eniten merkitsevä tavu pienimpään osoitteeseen → [0x11] [0x22] [0x33] [0x44]
 Little-Endian: vähiten merkitsevä tavu pienimpään osoitteeseen → [0x44] [0x33] [0x22] [0x11]

3.4.2004 Copyright Teemu Kerola 2003 19

Big vs. Little Endian (5)

- Monitavuisen tiedon (sana-) osoite on sama molemmissa tapauksissa
- Tavujen järjestys on erilainen
- Suorittimen suunnittelija päättää
 - Matematiikkapiiriin tulee tietää miten luvut esitetty
 - Täytyy ottaa huomioon siirrettäessä tietoa verkon yli
- Power-PC: bi-endian - molemmat moodit käytössä
 - voidaan valita ohjelmakohtaisesti
 - etuoikeutetussa tilassa voidaan vielä valita erikseen
 - suoritin osaa laskea kummallakin tavalla talletetuilla luvuilla

TTK-91: big-endian

3.4.2004 Copyright Teemu Kerola 2003 20

Negatiiviset luvut (4)

- Etumerkkibitti erikseen
- Yhden komplementtiesitys
- Kahden komplementtiesitys
- Vakiolisäys
 - Esim lisää 127 ($=2^8 - 1$)
 - yleensä: 2^{bittikm} - 1
 - Talleta etumerkittömänä

arvo talletus
 +57 = 0011 1001
 sign bit = MSB = most significant bit
 luku -57 = 1011 1001 talletusmuoto
 -57 = 1100 0110
 "sign" bit (+1)
 -57 = 1100 0111
 "sign" bit (+1)
 -57 = 0100 0110
 -57 + 127 = 70
 +57 = 0111 1000
 +57 + 127 = 184

3.4.2004 Copyright Teemu Kerola 2003 21

Kahden komplementti (7)

- Useimmiten käytössä
- Etu: vain yksi nolla
 - yhden komplementtissa kaksi nollaa
 - +0 = 0000 0000 -0 = 1111 1111
- Helpot muunnokset: arvo ↔ esitysmuoto
 - miten arvo -57 esitetään?
 • 1100 0110 +1 = 1100 0111
 - mitä arvoa esitysmuoto 1100 0111 tarkoittaa?
 • - (0011 1000 +1) = -0011 1001 = -57

3.4.2004 Copyright Teemu Kerola 2003 22

Liukuluvut (3)

- Tietokoneessa ei ole reaalilukuja tai rationaalilukuja (matemaattiset käsitteet)
- Aina rajallinen esityksen tarkkuus
 - lukuja π (pii), $\sqrt{2}$, tai $1/3$ ei voi esittää tarkasti
 - esim. luvut 1.000000000 ja luvut 1.000000001 ovat yhtäsuuria (joissakin esityksissä)
- Yleinen reaalilukuja vastaava esitysmuoto on liukukuesitysmuoto float, double, real
 - 32 bittiä, noin 7-8 desimaalinumeron tarkkuus
 - 64 bittiä, noin 16-17 desimaalinumeron tarkkuus

3.4.2004 Copyright Teemu Kerola 2003 23

Liukulukujen esitys (5)

“+” “14” “1.23”
 sign exponent mantissa or significand
 (exponentti) (mantissa)

3.4.2004 Copyright Teemu Kerola 2003 24

IEEE 32-bit Floating Point Standard ⁽¹⁾

IEEE Standard 754

“+”	“14”	“1.1875” = “1.0011”
sign	exponent	mantissa or significand
	(2:n potenssi)	

- Etumerkki
 - 1 bitti, '+' koodattu 0:ksi, '-' koodattu 1:ksi
 - etumerkkibitti $S \Rightarrow$ etumerkin arvo = $(-1)^S$
 - +n eli 0:n arvo on +1
 - n eli 1:n arvo on -1

3.4.2004 Copyright Teemu Kerola 2003 25

IEEE 32-bit FP Standard ⁽²⁾

“+”	“15”	“1.1875” = “1.0011”
sign	exponent	mantissa or significand

- 8 bittiä eksponentille, lisättyä 127:llä (biased form)
 - todellinen arvo exponent = 5 store \rightarrow talletettu arvo $5+127 = 132 = 1000\ 0100$
 - exponent = -1 store \rightarrow $-1+127 = 126 = 0111\ 1110$
 - exponent = 0 store \rightarrow $0+127 = 127 = 0111\ 1111$
- esitys muodot 0 ja 255 erikoistapauksia
 - laajennettu arvoalue: hyvin pienet luvut, NaN, $\pm\infty$
 - talletettu arvoalue: **1 - 254** \Rightarrow tod. arvoalue: **-126 - 127** (arvoalue)

3.4.2004 Copyright Teemu Kerola 2003 26

IEEE 32-bit FP Standard ⁽⁷⁾

“+”	“15”	“0.1875” = “0.0011”
sign	exponent	mantissa or significand

1/8 = 0.1250
1/16 = 0.0625
0.1875

- 23 bittiä mantissalle siten, että ...
 - Binääripiste (.) on heti ensimmäisen bitin jälkeen
 - Mantissa on normalisoitu: vasemmanpuolimmainen bitti on 1
 - Vasemmanpuolimmaista (eniten merkitsevä) bittiä (1) ei talleteta (implied bit, piilobitti) \rightarrow 24 bitin mantissa!

mantissa eksponentti
0.0011 “15”
1.1000 “12”
1000 “12”

3.4.2004 Copyright Teemu Kerola 2003 27

IEEE 32-bit FP Values ⁽⁹⁾

23.0 = $+10111.0 * 2^0 = +1.0111 * 2^4 = ?$
 $4+127=131$

0	1000 0011	011 1000 0000 0000 0000 0000
sign	exponent	mantissa or significand
1 bit	8 bits	23 bits

1.0 = $+1.0000 * 2^0 = ?$
 $0+127 = 127$ 0x3F800000

0	0111 1111	000 0000 0000 0000 0000 0000
sign	exponent	mantissa or significand
1 bit	8 bits	23 bits

3.4.2004 Copyright Teemu Kerola 2003 28

IEEE 32-bit FP Values ⁽⁶⁾

0x40740000

0	1000 0000	111 1000 0000 0000 0000 0000
sign	exponent	mantissa or significand
1 bit	8 bits	23 bits

$X = ?$ $X = (-1)^0 * 1.1111 * 2^{(128-127)}$
 $= 1.1111_2 * 2$
 $= (1 + 1/2 + 1/4 + 1/8 + 1/16) * 2$
 $= (1 + 0.5 + 0.25 + 0.125 + 0.0625) * 2$
 $= 1.9375 * 2 = 3.875$

3.4.2004 Copyright Teemu Kerola 2003 29

Merkit ⁽⁵⁾

- Yleensä 1 tavu per merkki
- ASCII, 7 bittiä/merkki (+ tark. bitti?)
 'A' = 0x41, 'a' = 0x61, LF = 0x0A
- EBCDIC, 8 bittiä/merkki
- ISO/IEC 8859-15 ('Latin-9'),
 - 8-bittiä/merkki, 256 eri merkkiä käytössä
 - mukana myös ä, ö, š, €

Suomalaiset merkit eri merkistöissä:
<http://www.hut.fi/~kmustone/skandit.html>

3.4.2004 Copyright Teemu Kerola 2003 30

UCS ja Unicode ⁽⁵⁾

- UCS - Universal Character Set
- Samat merkit, eri standardit
- 2 tavua eli 16 bittiä per merkki
 - 65536 merkkiä koko maailmassa käytössä oleville n. 200000 symbolille
- Kontrollimerkit
 - 0x0000-001F and 0x0080-009F
 - 0x007F = DELETE, 0x0020 = SPACE
- UCS:ssä myös 8-bittiset koodi "rivit"
 - eri alueille tai tarkoituksiin (zone) omat 8-bittiset koodinsa

3.4.2004

Copyright Teemu Kerola 2003

31

UCS koodistoja

- Merkit välillä 0x0000-00FF (16 bittiä) samassa järjestyksessä kuin Latin-9 merkitössä (8 bittiä)
 - 16-bittisen UCS:n "rivi 00" = 8-bittinen Latin-9
- Myös muut aakkoset:
 - I-zone = Kanji (0x4E00-9FFF, 20992 merkkiä)

	4E0	4E1	4E2	4E3	4E4	4E5	4E6	...
0	一	弓	北	丰	习	乐	习	...
	4E00	4E10	4E20	4E30	4E40	4E50	4E60	...
1	丁	丑	兩	卯	彳	乘	乡	...
	4E01	4E11	4E21	4E31	4E41	4E51	4E61	...

- Ei om

3.4.2004

Copyright Teemu Kerola 2003

32

Muita UCS koodistoja

- Heprea: 0x0590-05FF

	059	05A	05B	05C	05D	05E	05F
0	א	ב	ג	ד	ה	ו	ז
	0590	0591	0592	0593	0594	0595	0596
1	ח	ט	י	כ	ל	מ	נ
	0597	0598	0599	059A	059B	059C	059D

- Arabia: 0x0600-06FF

	060	061	062	063	064	065	066	067
0	ا	ب	ت	ث	ج	ح	خ	د
	0600	0601	0602	0603	0604	0605	0606	0607
1	ذ	ر	ز	س	ش	ص	ض	ط
	0608	0609	060A	060B	060C	060D	060E	060F

- Gujarati: 0x0A80-0AFF

	0A8	0A9	0AA	0AB	0AC	0AD
0	અ	બ	ક	દ	ધ	ણ
	0A80	0A81	0A82	0A83	0A84	0A85
1	ત	થ	જ	ઝ	ઞ	ટ
	0A86	0A87	0A88	0A89	0A8A	0A8B

<http://www.unicode.org/charts/>

3.4.2004

Copyright Teemu Kerola 2003

33

Merkkijonot ⁽⁵⁾

- Yleensä peräkkäin talletettu joukko tavuja
- Lisäksi tarvitsee jollain tavalla koodata merkkijonon pituus
 - laitetaan loppuun erikoismerkki
 - C-kieli: '\0' = 0x00
 - tai toteutetaan tietueena

20 "Ei yleensä nyt enää!"

pituus merkkijono

- ei omia konekäskyjä, käsittely aliohjelmilla

- kokonaisluku- ja bittimanipulointikäskyt
- joissakin koneissa "stropy" ja "strem" konekäskyt

3.4.2004

Copyright Teemu Kerola 2003

34

Totuusarvot ⁽⁵⁾

- Boolean TRUE ja FALSE
- Yleensä koodattu TRUE=1, FALSE=0
 - muttei aina!
 - Totuusarvolauseke **A and B** = kokon.lukulauseke **A*B**
- Usein Boolean arvo per sana
 - loput 31 bittiä nollia
 - ohjelmointikielten Boolean muuttujat
- Joskus pakatussa muodossa 32 arvoa per sana
- Ei omia konekäskyjä, manipulointi aliohjelmilla
 - kokonaisluku- ja bittimanipulointikäskyt
 - haluttu konekäsky "JTRUE ..." voidaan toteuttaa konekäskynä "JPOS ..." (jos TRUE = 1)

3.4.2004

Copyright Teemu Kerola 2003

35

Kuvat

- Monta kuvastandardia
 - yleisyys, siirrettävyys, pakkaustiheys
 - näyttöä varten tarvittavan laskennan määrä
- Kuvatiedoston alussa otsake kertoo talletusformaatin
- Viiva- ja vektorikuvat
 - kuva koodattuna objekteina
 - ympyrä, monikulmio, käyrä, alueen väri
- Rasterikuvat
 - kuva koodattuna pisteinä
 - kunkin pisteen väri koodattu esim. 24 bitillä

3.4.2004

Copyright Teemu Kerola 2003

36

Kuvat

- Kuvat on yleensä pakattu mahdollisimman vähän tilaa vievää muotoon
 - optimoitu tilan, ei laskennan mukaan
 - purkaminen voi vaatia paljon laskentaa
- GIF, JPEG, TIFF, BMP,
- Ei omia konekäskyjä, manipulointi aliohjelmilla

3.4.2004

Copyright Teemu Kerola 2003

37

Videokuva

- Vie hyvin paljon muistitilaa
- Talletus kuva kerrallaan, esim. 25 kuvaa/sek
 - 1 sekunti hyvälaatuista videokuva pakkaamattomassa muodossa 20 MB
- Talletus ”incrementaalisesti”
 - kun seuraava kuva poikkeaa edellisestä vain vähän ...
 - talleta vain muutokset edelliseen

3.4.2004

Copyright Teemu Kerola 2003

38

Videostandardit

- MPEG (Moving Pictures Expert Group)
- AVI (Audio Visual Interleave)
- MOV, INDEO, FLI, GL, DVD, ...
- Ei omia konekäskyjä, manipulointi aliohjelmilla tai ...
- Erikoisprosessoreilla (GPU), joiden käskykanta suunniteltu (jonkin standardin mukaisten kuvien) kuvankäsittelyyn
 - grafiikkakorteilla

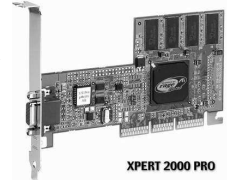
3.4.2004

Copyright Teemu Kerola 2003

39

Grafiikkakortit

- Esim. 4-256 MB VRAM (dual-port) muistia ...
 - 2 lukua/kirjoitusta samanaikaisesti
- ... tai ”tavallista”, mutta hyvin nopeaa RAMia
- Nopea väylä (ennen PCI, nyt AGP) suorittimelle
- Näytönohjaus monitoristandardien (VGA, XGA, RGB, ...) mukaisesti
- Oma suoritin (GPU)
 - lukee videodataa ja generoi näytettävän kuvan näyttöpuskuriin, josta monitori sen näyttää
- Voi olla integroitu emolevyn kanssa



3.4.2004

Copyright Teemu Kerola 2003

40

Äänet

- Täydellinen äänidata
 - 44100 näytettä/sek, 16 b/näyte, 88KB /sek
- Syntetisoitu ääni
 - MIDI-käskyjä
 - Music Instrument Digital Interface
 - ”Soita nuotti N voimakkuudella V”
- Ei omia konekäskyjä, manipulointi aliohjelmilla tai ...
- Erikoisprosessoreilla, joiden käskykanta suunniteltu äänen käsittelyyn
 - äänikortit

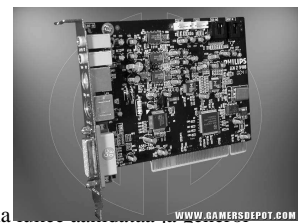
3.4.2004

Copyright Teemu Kerola 2003

41

Äänikortit

- Esim. 4-64 MB VRAM tai RAM muistia
- Nopea väylä (esim. PCI) suorittimelle
- Oma suoritin, joka
 - äänet kaiuttimille tai vahvistimeen
 - kaiuttimet tai vahvistin kiinni äänikortilla
- Voi olla integroitu emolevyn (tai grafiikkakortin) kanssa



3.4.2004

Copyright Teemu Kerola 2003

42

Maku, haju, tunto ja muu data ⁽³⁾

- Tähtien kirkkaus, hajut, veneen tyyppi, tunteen palo,
- Toteutus sovelluskohtaisesti, ei vielä yleisiä standardeja
 - kokonaisluvut (diskreetti data)
 - liukuluvut (jatkuva data)
- Ei omia konekäskyjä, manipulointi omilla aliohjelmilla

3.4.2004

Copyright Teemu Kerola 2003

43

Haju

ks. HS artikkeli 5.5.2000

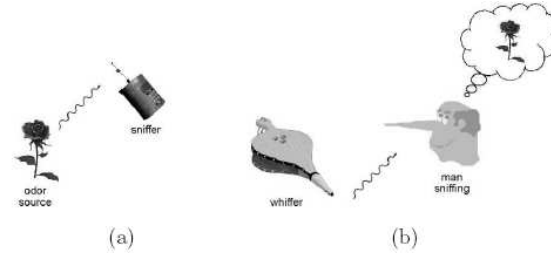


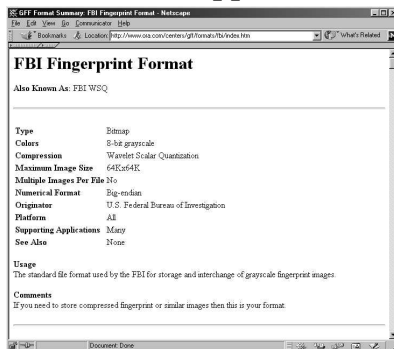
Fig. 1. The building blocks of an odor communication system.

3.4.2004

Copyright Teemu Kerola 2003

44

-- Luennon 6 loppu --



3.4.2004

Copyright Teemu Kerola 2003

45