

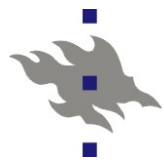
HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Johdatus tietojenkäsittelytieteeseen - silmäys tietojenkäsittelyn ydineknologioihin

**Matemaattis-luonnontieteellinen tiedekunta
Tietojenkäsittelytieteen laitos**

Kurssin sisältö

Lähde: Peter J. Denning: Great Principles of Computing (Communications of the ACM, 46, 11, marraskuu 2003, sivut 15-20).

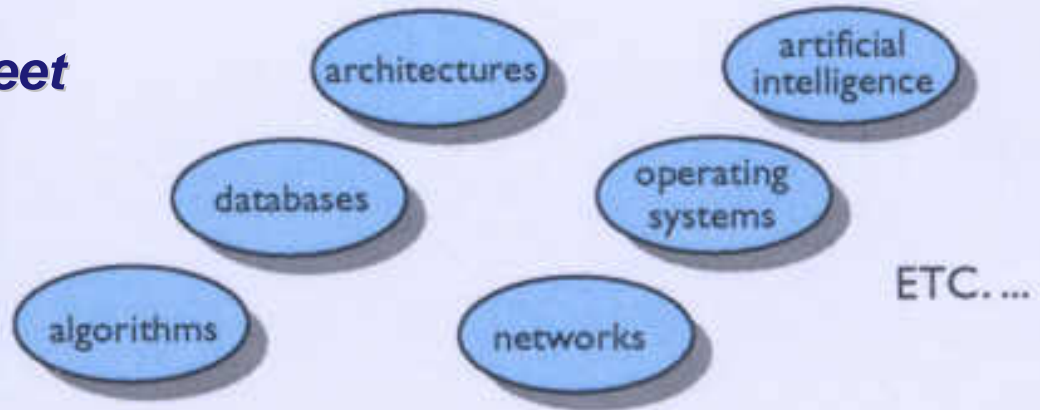


- Luku 1: Historiaa**
- Luku 2: Kokonaiskuva**
- Luku 3: Eettiset perusteet**
- Luku 7: COMPUTING PRACTICES**

programming
engineering systems
modeling
innovating
applying

Luku 4:

CORE TECHNOLOGIES



Luku 6:

GREAT
PRINCIPLES
OF
COMPUTING

DESIGN

simplicity, performance, reliability,
evolvability, security

Luku 5:

MECHANICS

computation, communication, coordination,
automation, recollection



Ydinteknologiat 1950-luvulla

1. algoritmit (*algorithms*)
2. numeeriset menetelmät (*numerical methods*)
3. laskennan mallit (*computation models*)
4. kääntäjät (*compilers*)
5. ohjelmointikielet (*programming languages*)
6. logiikkapiirit (*logic circuits*)



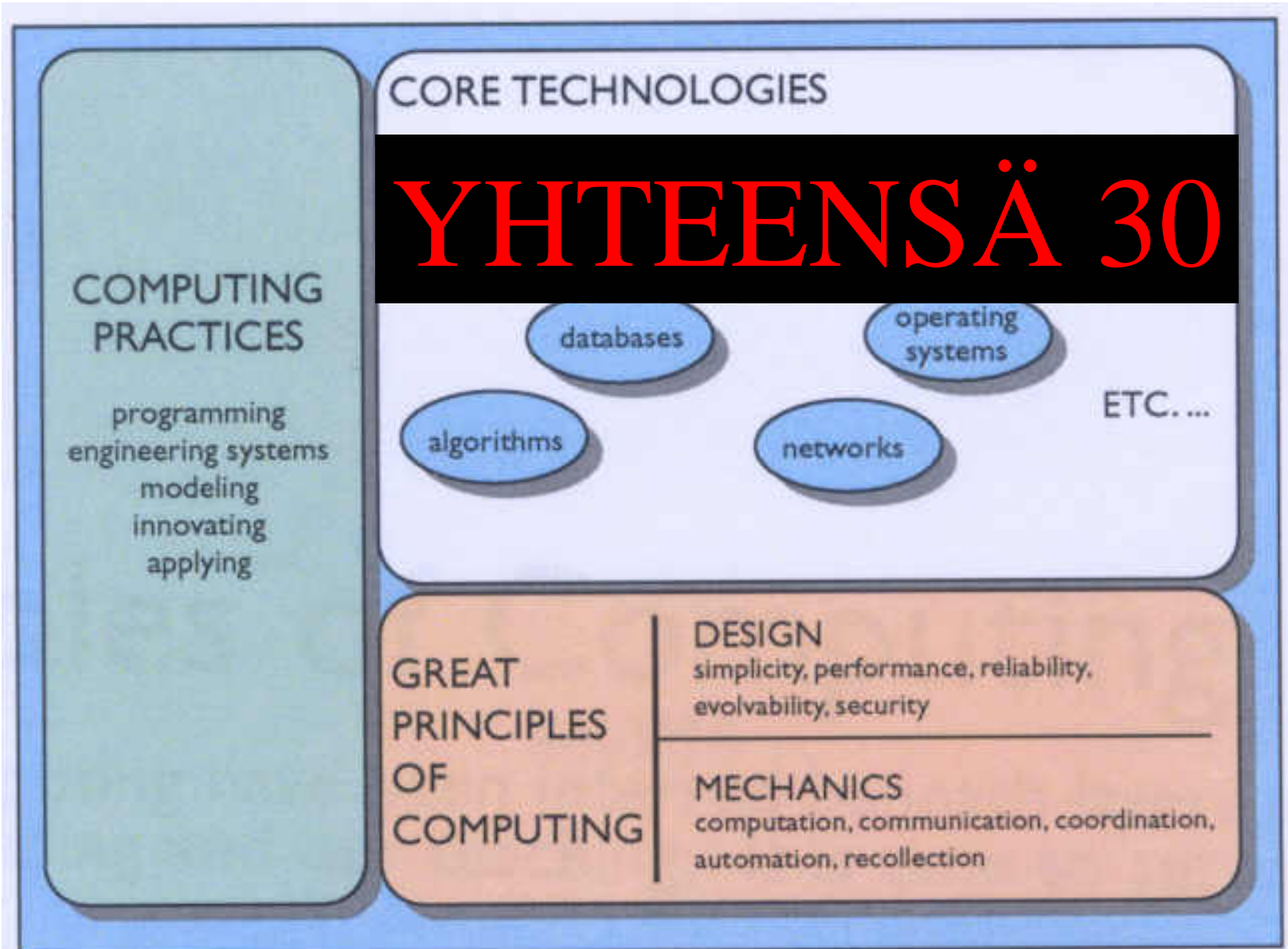
Ydinteknologioita 1990-luvulle tultaessa lisää

7. käyttöjärjestelmät (*operating systems*)
8. tiedonhaku (*information retrieval*)
9. tietokannat (*databases*)
10. tietoverkot (*networks*)
11. tekoäly (*artificial intelligence, AI*)
12. ihmisen ja tietokoneen vuorovaikutus (*human-computer interactions, HCI*)
13. ohjelmistotekniikka (*software engineering*)



Ydinteknologioita 2003

Lähde: Peter J. Denning: Great Principles of Computing (Communications of the ACM, 46, 11, marraskuu 2003, sivut 15-20).





Algoritmit (*algorithms*)

- Persialainen matemaatikko ja tähtitieteilijä Al-Khwarizmi.
- Äärellinen joukko täsmällisiä, suoritettavissa olevia ohjeita, jotka ohjaavat päättyvää tehtävän suorittamista.
- Tietojenkäsittelyssä algoritmit lopulta suoritetaan tietokoneella.
 - Väliin tarvitaan monta ydinteknologiaa.
- Algoritmitutkimuksessa
 - kehitetään algoritmeja
 - analysoidaan niiden ominaisuuksia



Tekoäly (*artificial intelligence, AI*)

- Tietokoneohjelman osoittama äly.
- Vaikeasti määriteltävissä.
- Tekoälytutkimuksessa tarkastellaan järjestelmiä, jotka automatisoivat älykästä toimintaa edellyttäviä tehtäviä:
 - ohjaus (*control*),
 - suunnittelu ja ajoitus (*planning and scheduling*)
- Sovellusalueita: puheen tunnistus, asiantuntijajärjestelmät, videopelit, yms.



Kääntäjät (*compilers*)

- Ohjelmointikielen kääntäjä on tietokoneohjelma, joka muuntaa ohjelmointikielisen lähdekoodin (*source code*) konekieliseen muotoon (*object code*).

- Käännös on usein kaksivaiheinen:
 - Lähdekoodi välimuotoon
 - Välimuotoinen koodi objektikoodiksi
 - selaaaja tunnistaa alkionimet (*token*)
 - jäsennin tunnistaa kielen rakenteet
 - semanttisessa analyysissä tutkitaan mm tietotyyppien käytön oikeellisuus
 - objektikoodin optimointi



Laskennallinen tiede (*computational science*)

- Muiden tieteenalojen tutkimusongelmia kuvaavien mallien ratkaisemista tietokoneen avulla.
- Mallien muodostaminen usein yhteistyötä.
- Mallien ratkaisu- ja analysointimenetelmien tutkiminen ja kehittäminen on tietojenkäsittelytieteen tutkimusalue.
- Laskennallinen
 - biologia
 - lääketiede
 - kemia
 - fysiikka
 - tilastotiede
 - jne



Tieteellinen laskenta (*scientific computing*)

- Eri tieteenaloilla käytettävien matemaattisten mallien numeerisia ratkaisumenetelmiä sekä niiden tietokonetoteutuksia.
- Numeerinen analyysi osa tieteellistä laskentaa.
- Usein ”numeronmurskausta” eli pitkiä laskutoimituksia.
- Tieteellinen laskenta – laskennallinen tiede: kietoutuvat usein yhteen.



Tietokoneen rakenne (*computer architecture*)

- Tietokonearkkitehtuuri on tietokoneiden rakenteen suunnittelun taustalla oleva teoria.
 - Laitteiston suunnittelu siten, että laitteisto käyttäytyy ohjelmoijien olettamalla tavalla.
 - Toteutusteknologioiden (esim. puolijohteiden) käyttäminen siten, että laitteisto on ”mahdollisimman hyvä”.

- Mahdollisimman hyvä riippuu suunnittelun tavoitteista:
 - hinta vs nopeus
 - koko, paino, virrankulutus



Tiedon louhinta (*data mining*)

- Suurista tietomassoista etsitään kaavaimia (*pattern*), kuten assosiaatiosääntöjä.
- Käytetään laskennallisia tekniikoita esimerkiksi
 - tilastotieteellisiä menetelmiä,
 - tiedon haun menetelmiä,
 - koneoppimisen menetelmiä,
 - hahmontunnistuksen menetelmiä.
- *KDD: Knowledge-Discovery in Databases*
- Tavoitteena löytää tietomassasta (data) aiemmin tunnistamatonta ja mahdollisesti hyödyllistä tietoa.



Tietoturva (*data security*)

- Tietoturva (*information/data security*) on tiedon luotettavuudelle asetettuja kriteereitä.
 - saatavuus (*availability*)
 - luottamuksellisuus (*confidentiality*)
 - pääsynhallinta (*access control*)
 - salaus (*encryption*)
 - eheys (*integrity*)
 - tarkistussumma (*checksum*)
 - tarkistuskoodi (*cyclic redundancy check, CRC*)
 - digitaalinen (sähköinen) allekirjoitus (*digital signature*)
 - kiistämättömyys (*non-repudiation*)
 - tunnistus (*identification*)
 - todennus (*authentication*)
- Oikea tieto oikeille ihmisille oikeaan aikaan!



Tietorakenteet (*data structures*)

- Tietorakenteet ovat
 - tapoja, miten tieto talletetaan tietokoneen muistiin ja
 - operaatioita, joiden avulla tietoja päästään käyttämään.
- Tietorakenteiden valinta (suunnittelussa) vaikuttaa olennaisesti tiedonkäsittelyn tehokkuuteen (mm suoritus aika, muistitilan tarve, virhealttius).
- Ohjelmointikielissä on yleensä valmiit ja tehokkaat tietorakenteiden käsittelymahdollisuudet.
- Pino, jono, lista, hajautustaulu, puu, taulukko, ...



Tietokannat (*databases*)

- Tietokokoelma, joka muodostaa hallinnollisen kokonaisuuden.
- Tietokannan tietomalli (*data model*) määrää tiedon rakenteen ja käsittelyn (kyselykielet).
- Tietokannoille on tyypillistä tietoriippumattomuus:
 - tietokannoissa tiedon rakenteen kuvaus on erillään ohjelmista
- Tietokannan hallintajärjestelmällä (*database management system, DBMS*) perustetaan tietokanta ja hallitaan sen tietoja.



Sanan transaktio (*transaction*) merkityksiä

- *Transaction processing* tarkoittaa yleensä suomeksi transaktioiden käsittelyä tietokantojen yhteydessä.
- *Event handling* tarkoittaa yleensä suomeksi tapahtumankäsittelyä tapahtumaohjatuissa järjestelmissä, kuten esim. graafisissa käyttöliittymissä.
 - näppäintä painettu
 - hiirtä liikutettu
 - valittu toimenpide
 - ajastin
- Varovaisuutta: *Transaction processing* joskus tapahtumakäsittelyä yms sotkua! Entä keskeytys (interrupt)?



Transaktio (*transaction*) tietokantojen yhteydessä

- Transaktioiden käsittelyllä hallitaan mm. tietokannan tietojen samanaikaista käyttöä.
 - Toimintaketjuja, joita ei saa keskeyttää.
- Samanaikaisuuden hallinta takaa transaktioiden jälkeen tietojen oikeellisuuden: ACID-säännöt
 - Atomisuus (Atomicity)
 - Oikeellisuus (Consistency)
 - Eristys (Isolation)
 - Pysyvyys (Durability)



Päätöksenteon tukijärjestelmät (*decision support systems, DSS*)

- Ohjelmistoja, jotka tukevat päätöksentekoa organisaatioissa.
 - ei automaattisia päätöksiä
 - vuorovaikutteisia
 - laajentaa käyttäjän kognitiivista päätöksentekokykyä

- *DSS* on käsitteenä laaja
 - Johdon tietojärjestelmät (*management information systems, MIS*)
 - Ylimmän johdon tietojärjestelmät (*executive information systems, EIS*)

- Monitieteistä ja monta tkt:n ydinteknologiaa, esim, tietokannat, käyttöliittymät, tekoäly, visualisointi, ...



Hajautettu tietojenkäsittely (*distributed computation*)

- Fyysisesti eri paikoissa verkossa olevien tietokoneiden yhteistoiminta tehtävän suorittamiseksi.
- Käyttäjät ja tietojenkäsittelykapasiteetti yhdistetään läpinäkyvästi, avoimesti ja skaalautuvasti (kun tarvitaan enemmän, niin saadaan vaivattomasti enemmän)
- Tavoitteena parempi resurssien saatavuus, vikasietoisuus (*fault-tolerance*) ja suoritusteho.



Rinnakkaislaskenta (*parallel computation*)

- Tehtävä jaetaan osatehtäviin, joita suoritetaan rinnakkain (samanaikaisesti) usealla suoritimella.
- Tavoitteena nopeampi tehtävän valmistuminen.
- Tavoitteena parempi resurssien saatavuus, vikasietoisuus (*fault-tolerance*) ja suoritusteho.
- Tutkimuskohteita:
 - Laitteistoarkkitehtuurit, erityisesti prosessorien välinen ja prosessorien ja muistien välinen kytkentä.
 - Rinnakkaislaskentaan soveltuvat algoritmit.
 - Säikeiden välinen kommunikointi.



Sähköinen kaupankäynti (e-commerce)

- Tuotteiden tai palveluiden jakelu, osto, myynti, markkinointi ja tarjonta tietoverkkojen välityksellä.
- Sähköisen kaupankäynnin järjestelmä on monitieteinen.
- Tarvitaan mm
 - toimiva tietoteknologia,
 - sopivia liiketoimintamalleja ja
 - riittävä tietoturvaan perustuva luottamus.
- Usein tarvittavia toimintoja:
 - Sähköinen varainsiirto (*electronics fund transfer*).
 - Tuotantoketjun hallinta (*supply chain management*).
 - Välitön transaktioiden käsittely (*online transaction processing*).
 - Sähköinen tiedonvaihto (*electric data interchange, EDI*).
 - Automatisoidut varastokirjanpitojärjestelmät.
 - Automatisoidut tiedonkeruujärjestelmät.



Tietokonegrafiikka (*computer graphics*)

- Kattaa visuaalisen tietojenkäsittelyn.
 - Kuvien synteettinen tuottaminen
 - Reaalimaailmasta peräisin olevan visuaalisen informaation ja tilatiedon (*spatial information*) muokkaaminen.
- Joitakin osa-alueita:
 - tosiaikainen kolmiulotteisten kuvien esittäminen (*3-D rendering*),
 - animointi,
 - videosignaalin käsittely,
 - visuaalisten tehosteiden luonti ja muokkaus,
 - kuvan (*image*) muokkaaminen ja mallintaminen.
- Sisältää usein matemaattisia malleja ja laskentaa.



Ihmisen ja tietokoneen vuorovaikutus (*human-computer interaction, HCI*)

- Monitieteistä: mm. estetiikka, muotoilu, psykologia, jne
- Tietojenkäsittelytieteessä keskitytään käyttöliittymään (*user interface, UI*).
 - Ohjelmisto.
 - Laitteisto, myös oheislaitteet.
- Tietokoneet ja tietokonejärjestelmät käyttäjäystävällisemmiksi (*user-friendly*) ja helppokäyttöisemmiksi!



Tiedonhaku (*information retrieval*)

- Tiedon – tekstiä, ääntä, kuvaa, dataa – etsimistä (*search*) dokumenteista.
- Dokumenttien etsimistä.
- Dokumentteja kuvaavan tiedon (*metadata*) etsimistä.
- Etsintä tietokannoista ja tietoverkoista.
- Aluksi tieteellisten julkaisuiden sisältämän informaation haun automatisointia.
- Webin hakukoneet nykyisin ehkä yleisimmin käytettyjä sovelluksia.



Luonnollisen kielen käsittely (*natural-language processing*)

- Tekoälyn ja kielitieteen yhteisellä maaperällä.
- Tutkitaan luonnollisen kielen automaattisen tuottamisen ja ymmärtämisen ongelmia.
- Tietokannan tiedoista luonnollista puhetta.
- Puheesta esitysmuoto, jota tietokoneen on helppo käsitellä.



Tietoverkot (*networks*)

- Tietoliikenneyhteyksillä ja tietoliikenneprotokollilla yhteen kytkettyjen tietokoneiden järjestelmä.
- Tutkimusalueita:
 - Tietoliikennelaitteet.
 - Tiedon esitysmuodot.
 - Tietoturva.
 - Tietoliikenneprotokollat.
 - Verkonhallinta (*network management*).
 - Langaton tiedonsiirto (*wireless communication*).
 - Liikkuva tietojenkäsittely (*mobile computing*).



Käyttöjärjestelmät (*operating systems, OS*)

- Ohjelmisto, joka hallinnoi tietokoneen laitteistoa ja ohjelmistoja.

- Käyttöjärjestelmä palvelee muita ohjelmia:
 - Muistin hallinta ja jakaminen (allocation).
 - Käskyjen suorituksen järjestäminen (prioritizing).
 - Oheislaitteiden hallinta.
 - Tietoliikenteen tukeminen.
 - Tiedostojen hallinta.



Käyttöjärjestelmät

■ Huolehtii

- keskeytyksistä (*interrupts*),
- ajastimista (*timers*),
- prosesseista (*processes*) ja säikeistä (*threads*) sekä niiden vuorottamisesta (*scheduling*),
- samanaikaisuuden hallinnasta (*concurrency control*)
- samanaikaisesti suoritettavien ohjelmien eristämisestä ja
- prosessien välisestä kommunikoinnista (*interprocess communication*).



Käyttöjärjestelmät

- Tutkimusalueita ovat mm
 - muistinhallinta (*memory management*),
 - tiedostojärjestelmät (*file systems*),
 - samanaikaisuuden hallinta (*concurrency control*),
 - vikasietoisuus (*fault-tolerance*) ja
 - virrankulutuksen hallinta.



Ohjelmointikielet (*programming languages*)

- Täsmällisesti määritelty tapa antaa tietokoneelle toimintaohjeet.
 - Syntaksi (*syntax*) eli lauseoppi (sanasto ja kielioppisäännöt).
 - Semantiikka (*semantics*) eli merkitysoppi.
- Ohjelmointikielessä määritellään mm. ohjelmoijan käytössä olevat
 - tietotyypit (data types),
 - tietorakenteet (data structures),
 - lauseet
 - jne
- Tutkimusalueita ovat mm. ohjelmointikielten ominaisuudet ja ohjelmointimallit (paradigmat).



Tosiaikajärjestelmät (*real-time systems*)

- Järjestelmiä – laitteisto ja ohjelmisto – joiden on täytettävä aikavaatimus.
- Tosiaikajärjestelmän ei välttämättä tarvitse olla nopea, mutta tulos on oltava valmis aikarajaan (*deadline*).
- Luokitellaan koviin (*hard*) ja pehmeisiin (*soft*) sen mukaan kuinka ehdottomia aikarajat ovat.
 - Kovan tosiaikajärjestelmän tulos on aina virheellinen, jos aikaraja ylittyy.



Robotiikka (*robots*)

- Robotti on elektro-mekaaninen laite, joka tekee tehtäviä autonomisesti tai ennalta ohjelmoidusti.
- Robotiikassa tarvitaan elektroniikan, mekaniikan ja ohjelmistotekniikan hallintaa.
- Tiettyyn tehtävään soveltuvan robotin kehittämiseen tarvitaan mm
 - havaintoja tekeviä tunnistimia (*sensors*),
 - ohjausalgoritmeja ja
 - robotin mekaanista toimintaa ohjaavat säätimet (*actuators*).



Ohjelmistotekniikka (*software engineering*)

- Ohjelmistojen suunnitteluun, toteuttamiseen ja ylläpitoon kuuluvia tekniikoita ja käytäntöjä.
 - Tietojenkäsittelytieteen ydinteknologioita.
 - Projektinhallintaa (*project management*).
 - Insinööritaitoa (*engineering*).
 - Sovellusalueen tietämystä.
- Ohjelmistotekniikassa kustannukset ja luotettavuus ovat yhtä keskeisiä kuin perinteisimmillä insinööritaidon alueilla.



Ohjelmistotekniikka

- IEEE:n standardi 610.12 määrittelee, että ohjelmistotekniikka on
 - systemaattisen, kurinalaisen ja ilmaistavissa olevan menettelytavan käyttämistä ohjelmiston kehittämisessä, käytössä ja ylläpidossa sekä
 - tällaisten menettelytapojen tutkimista.



Supertietokoneet (*supercomputers*)

- Aikansa laskentateholtaan suorituskykyisimpiä tietokoneita.
- Laskentatehon kasvattaminen on yleensä tapahtunut
 - lisäämällä innovatiivisesti rinnakkaisuutta käskyjen käsittelyssä,
 - huolellisella muistihierarkian suunnittelulla ja
 - prosessorin rakenteen yksityiskohtaisella suunnittelulla.
- Yleensä suunniteltu tietyn tyyppiseen tietojenkäsittelyyn – useimmiten numeeriseen laskentaan.



Virtuaalitodellisuus (*virtual reality*)

- Käyttäjä on vuorovaikutuksessa tietokoneella simuloidun ympäristön kanssa.
 - Simuloitu ympäristö voi olla todellisuuden kaltainen (esim. lentäjäkoulutus)
 - todellisuuden kaltainen (esim. lentäjäkoulutus) tai
 - todellisuudelle vieras (esim. monet videopelit).

- Simuloidussa ympäristössä on
 - yleensä visuaalisia kokemuksia
 - tavallisella näyttölaitteella tai
 - erityisellä stereoskooppisella näytöllä.
 - usein myös kuvan kanssa synkronoitua ääntä.



Konenäkö (*vision*)

- Tutkitaan, miten tietokone saadaan ”ymmärtämään” kuvien sisältöä.

- Kuvista etsitään tiettyä tarkoitusta palvelevaa informaatiota:
 - Sovelluksia esim.
 - lääketieteessä,
 - laitteen ohjauksessa,
 - laadunvalvonnassa.

- Kehitettävää riittää...tekoälyä, signaalinkäsittelyä, neurobiologiaa, matematiikkaa, fysiikkaa (valon heijastuminen pinnoista), ...



Visualisointi (*visualization*)

- Menetelmät, joilla luodaan kuvia, kaavioita tai animaatioita.
- Tavoitteena on parantaa tiedon välittymistä.
- Sovelluksia esim.
 - tieteissä,
 - tekniikassa,
 - tuotekehityksessä ja tuotannossa,
 - opetuksessa ja
 - lääketieteessä.
- Tietokonegrafiikka on visualisoinnin tärkein apuväline.



Työnkulku (*workflow*)

- Organisaation työtehtävien tekemisen järjestäminen tietokonejärjestelmiä apuna käyttäen.
 - Miten työtehtävät järjestetään?
 - Kuka suorittaa minkäkin tehtävän?
 - Missä järjestyksessä työtehtävät on suoritettava?
 - Mitkä ovat tehtävän aloittamisen edellytykset?
 - Miten tietovirrat tukevat tehtävän suorittamista?
 - Miten tehtävien etenemistä seurataan?

- Työnkulun tukijärjestelmissä (*workflow systems*) on usein kaksi osaa:
 - Työnkulun mallintaminen (*workflow modeling component*).
 - Työnkulun seuranta (*workflow execution component, workflow run-time system*)