



 HELSINGIN YLIOPISTO  
 HELSINKI UNIVERSITY  
 UNIVERSITY OF HELSINKI

**Tiedon louhinnan menetelmät**  
**582448 (6 op, 3 ov), syksy 2005**

Hannu Toivonen  
 30.8.2005

Tietojenkäsittelytieteiden laitos



**Kurssin perustiedot**

- Tiedon louhinnan menetelmät, 582448
- 6 op, 3 ov
- Vanhoissa tulkintovaatimuksissa informaatiojärjestelmien linjan valinnainen laudatur-kurssi
- Uusissa tulkintovaatimuksissa tiedonhallinnan linjalla (ent. informaatiojärjestelmät) pakollinen syventävä kurssi
- Esitietovaatimukset: mielellään cum laude, vähintään kuitenkin tietorakenteet + harjoitustyö
- Opettaja: prof. Hannu Toivonen (@cs.helsinki.fi)
- Opetus: 5.9.-28.11. MA 9-12 B222, TO 9-12 BK106
  - ei kuitenkaan periodien taitteissa 17.10.-30.10.
- Kirja: Han ja Kamber: Data mining: Concepts and techniques, Morgan Kaufmann Publishers, August 2000


Tiedon louhinnan menetelmät  
Alustusluento  
2



**Tiedon louhinta**

- <Mitä ja miksi>


Tiedon louhinnan menetelmät  
Alustusluento  
3



**Oppimistavoitteet**

- Kurssin käytyään opiskelija
  - tunnistaa tiedon louhintaongelmasta, soveltuuko siihen luokittelijan oppiminen tai yleisemmin ennustaminen
  - ymmärtää yhden luokittelualgoritmin niin hyvin, että pystyy soveltamaan sitä mielekkäästi ja toteuttamaan sen
  - osaa arvioida luokittelijan tarkkuuden luotettavasti kokeellisesti
  - tunnistaa tiedon louhintaongelmasta, soveltuuko siihen klusterointi
  - ymmärtää yhden klusterointialgoritmin niin hyvin, että pystyy soveltamaan sitä mielekkäästi ja toteuttamaan sen


Tiedon louhinnan menetelmät  
Alustusluento  
4



**Oppimistavoitteet (jatkuu)**

- osaa arvioida kriittisesti klusteroinnin laatua ja ymmärtää käytettyjen muuttujien vaikutuksen saatuun tulokseen
- tunnistaa tiedon louhintaongelmasta, soveltuuko siihen toistuvien hahmojen etsintä
- ymmärtää Apriori-algoritmin niin hyvin, että pystyy soveltamaan sitä mielekkäästi toistuvien hahmojen etsintään ja tarvittaessa toteuttamaan sen
- ymmärtää assosiaatiosäännöt ja niiden tuottamisen Apriori-algoritmillä
- tunnistaa pääpiirteissään joitain muita tiedon louhinnan ongelmatyyppejä
- osaa soveltaa tiedon louhintaprosessia todellisiin louhintaongelmiin

Tiedon louhinnan menetelmät  
Alustusluento  
5



**Kurssin toimintamuoto: ongelma-keskeinen oppiminen**

- Ongelma-keskeinen oppiminen, tutkiva oppiminen
  - laajojen ongelmien tutkiminen ja käsittely pienryhmissä
  - tavoitteellinen, kokonaisvaltainen oppiminen
- Opiskelijoilla aktiivinen rooli
  - aktiivinen oma tiedon hankinta, käsitteiden ja käsitteiden muodostaminen, ratkaisumallien muodostaminen ja arvioiminen sekä kokeilu
- Opettaja on "valmentaja"
  - luennoilla käsitellään vain keskeisimmät asiat sekä ongelma-kohtia
  - harjoitustöiden ohjaus

Tiedon louhinnan menetelmät  
Alustusluento  
6



## Ongelmakeskeisen oppimisen etuja

- Ongelmakeskeisen oppimisen etuja
  - oppiminen on motivoitumpaa ja hauskeempaa
  - kokonaisuuksien omaksuminen ja soveltaminen
  - yhdessä oppimisen hyödyllisyys
  - ryhmätöiden kehittyminen
  - kyky selittää ja kuunnella
  - omien vahvuuksien ja heikkouksien ymmärtäminen ja hyödyntäminen
- Tiedon louhinta on oikeasti ongelma-keskeistä
  - oikeisiin ongelmiin ei ole valmiita oppikirjaratkaisuja
  - ratkaiseminen vaatii kokonaiskuvaa, soveltamiskykyä, ryhmätöitä, ...

Tiedon louhinnan  
menetelmä  
Alustusluento  
7



## Ongelmakeskeisen oppimisen muita puolia

- Työmuoto edellyttää opiskelijoilta omatoimisuutta
- Ryhmätö vaatii yhteistyökykyä
  - konkreettisesti yhteistä aikaa ryhmätöiden tekoon
- Ongelmakeskeinen oppiminen on tehokas tapa harjaantua näissä taidoissa, kun aidosti haluaa sitä
- Jos epäroit, harkitse vaihtoehtoa B: kurssin suorittamista tenttimällä
  - voit seurata kurssin luentoja, harjoitustehtävät ovat julkisia
  - et saa ohjausta etkä pisteitä harjoituksista

Tiedon louhinnan  
menetelmä  
Alustusluento  
8



## Alustava toimintamalli

- Opettaja antaa maanantaina laajan tiedonlouhintaongelman
  - ei yhtä oikeaa vastausta
  - osoitetaan lähdemateriaalia (lähinnä Han et al)
- Kukin opiskelija lukee aiheesta ja pohtii ongelmaa
- Ryhmät keskustelevat ongelmasta, pohtivat ratkaisumalleja käsitteellisellä tasolla, kehittävät niitä yhdessä
- Torstaina keskustelua jatketaan joko ryhmittäin tai kaikkien ryhmien kesken (opiskelijavetoisesti)
- Tarpeen mukaan opettaja täsmentää tai korjaa ongelmaa tai antaa muita lisäohjeita

Tiedon louhinnan  
menetelmä  
Alustusluento  
9



## Alustava toimintamalli (jatkuu)

- Ryhmät jatkavat ratkaisumallin työstämistä
- Seuraavan maanantain luennolla käydään läpi keskeisiä käsitteitä ja mahdollisia ongelma-kohtia
- Ryhmät jatkavat työtä, myös käytännön kokeita
- Seuraavana torstaina ohjausta + mahdollisesti ex tempore -laskareita osaamistason tarkistamiseksi
- Ryhmät jättävät raportin työstään perjantaina klo 15 mennessä
- Ryhmät arvioivat ristiin toistensa raportteja
- Lisäksi posterisessio ryhmäkohtaisista aiheista
  - yhden aiheen esittely muille ryhmille julisteen avulla

Tiedon louhinnan  
menetelmä  
Alustusluento  
10



## Ongelmat ja käsiteltävät aineistot

- Kurssilla käsitellään 3-4 ongelmaa, jotka yhdessä kattavat kurssin oppimistavoitteet
- Ongelma muodostuu tyypillisesti aineistosta, jota halutaan oppia ymmärtämään paremmin tiedon louhinnan avulla
- Esimerkkiaineistot ovat mahdollisuuksien mukaan helposti ymmärrettäviä ja kohtuullisen pieniä
  - havainnollistavia ja helposti käsiteltäviä
  - eivät välttämättä tyypillisiä louhittavia aineistoja
- Omiakin aineistoja saa ehdottaa!

Tiedon louhinnan  
menetelmä  
Alustusluento  
11



## Tuotettavat raportit

- Kustakin ongelmasta:
- Tekninen osa (ryhmä yhdessä): miten ongelmaa lähestyttiin ja käsiteltiin, ehdotetut ratkaisut perusteluineen, suoritettavat kokeet ja niiden tulokset, johtopäätökset sekä tulosten ja ratkaisun (kriittistä) pohdintaa
    - "kohdeyleisönä" muut kurssin kävijät tai muut vastaavassa opintojen vaiheessa olevat (eikä opettaja)
    - arviointi ristikkäin ryhmässä
  - Itsearviointi (henkilökohtaisesti): miten oppimistavoitteet saavutettiin; tehtävän, oppimisen ja opetuksen arviointi
    - palautus vain opettajalle

Tiedon louhinnan  
menetelmä  
Alustusluento  
12



### Opintopiirin järjestäytyminen

- Ryhmien koko n. 4 henkeä
- Ryhmäjako tehdään torstaina klo 9.15
- Kukin ryhmä asettaa omat tavoitteensa
- Sopimus tehdään kirjallisesti
- Ryhmään osallistuminen on sitovaa – epäröivät tapaukset jääköön ulkopuolelle (vaihtoehto B)
- Ryhmissä kiertävät puheenjohtajan ja sihteerin roolit
- Ryhmä suunnittelee itse työskentelytapansa; kommunikointi vapaasti ja monipuolisesti (tapaamiset, sähköposti, puhelin, www, wiki, ...)
- Työtä arvioidaan pääasiassa ryhmittäin

Tiedon louhinnan  
menetelmä  
Aloitustuento  
13



### Kurssin suorittaminen

- Vaihtoehto A
  - ongelma-keskeinen oppiminen
  - sitoutuminen ryhmätyöhön kurssin ajaksi
  - ryhmätöistä yht. maks. 20 pistettä
  - tentti maks. 60 pistettä
    - kurssin läpikäyty edellyttää väh. 30 p kokeesta
  - arvosana ryhmätyön ja tentin perusteella
    - ryhmätyön 10 pinnaa ylittävä osa lisätään koepisteisiin
- Vaihtoehto B
  - tentti maks. 60 pinnaa
  - arvosana pelkän tentin perusteella
  - käytettävissä kurssin materiaalia, ei ohjausta

Tiedon louhinnan  
menetelmä  
Aloitustuento  
14



### Ryhmätyötaitoja

- Auta muita menestymään
- Ota vastuu itsesi lisäksi ryhmästä
- Osoita luottamusta toisia kohtaan olemalla avoin
- Kuuntele, mitä muut sanovat
- Kysy muiden mielipiteitä
- Pyydä toisilta apua
- Ole aktiivinen, osallistu – aktiivisena on mukavampaa
- Vaihda tietoja ja materiaaleja ja kerro työtavoistasi
- Kannusta toisia
- Ole kriittinen asioille, ei ihmisille
- Tee oma osuutesi yhdessä sovitulla tavalla ja yhdessä sovitussa aikataulussa

Tiedon louhinnan  
menetelmä  
Aloitustuento  
15