

Tiedonhakumenetelmät, kl. 2001

Luennot: ke 10-12 A414

Harjoitukset: ke 8-10 B453

Suoritustapa (alustavasti):

koe pe 18.5. klo 10-14 Auditorio

n. 35 pistettä

harjoitukset 10 pistettä

harjoitustyö n. 15 pistettä

Kokeen voi suorittaa uusintakokeessa; muut komponentit mukana. Kokeesta on saatava noin puolet pisteistä, samoin yhteispisteistä.

Harjoituksia noin 10 kertaa, ei muodollisesti pakollisia - suositellaan!

- sisältö kuuluu kurssiin
- lisäpisteet

Kurssin sisällöstä

Johdanto (motivointi, sisällön kuvailu, materiaali)

Dokumentin perusmallit

Kyselyt: Boolean malli, vektorimalli, muut

Dokumenttien indeksointi

Dokumenttien ryhmittely

Menetelmiä: merkkijonohaku, nimikirjoitustiedostot

Informaation suodatus (filtering)

Hyperteksti ja tiedonhaku

Digitaaliset kirjastot (digital libraries)

Dokumentin metatiedot

Multimedia ja tiedonhaku

Hakukoneet

(kaikkea ei varmaan ehditä ...)

Kurssin / aihepiirin materiaali

Baeza-Yates, R. & Ribeiro-Neto, B., Modern Information Retrieval. Addison-Wesley, 1999. (BYRN)

Salton, G., Automatic Text Processing. Addison-Wesley, 1989 (Part 3, ss. 227-373 pääosin). (S)

Korfhage, R., Information Storage and Retrieval. Wiley, 1997. (K)

Witten, I.H., Moffat, A. & Bell, T.C., Managing Gigabytes. Van Nostrand Reinhold, 1994. (osia) (WMB)

Sparck Jones, K. & Willett, P., Readings in Information Retrieval. Morgan Kaufmann, 1997. (klassisten IR-artikkelien kokoelma) (SJW)

Järvelin, K., Tekstietiedonhaku tietokannoista. Suomen ATK-Kustannus Oy, 1995. (J) (osia, yleistä taustaa)

- alkuperäisartikkeleita
- lehtien teemanumeroita
 - Comm. ACM 38,4 (1995): digital libraries
 - Comm. ACM 35,12 (1992): information filtering
 - Interacting with Computers 10, 2-3 (1998): HCI and information retrieval
 - IEEE Computer 32, 2 (1999): digital libraries

- lehtiä:
 - Information Retrieval (Kluwer; 1999-)
 - Information Processing and Management
 - JASIS

- konferensseja:
 - ACM SIGIR
 - Digital Libraries

Aihepiiri CR-luokituksessa:

- H.3 Information Storage and Retrieval
- H.3.3 Information Search and Retrieval
- H.3.7 Digital Libraries
- H.3.1 Content Analysis and Indexing

- yhteyksiä:
 - tietokannat
 - käyttöliittymät
 - informaatiotutkimus (kirjastotiede)

Kurssimateriaali

- kurssin kotisivulla (pdf)
- kurssikansiossa (A412)

- luentokalvojen kopiot
(ei välttämättä täydellisenä)
- harjoitustehtävät
- harjoituksiin liittyviä artikkeleita

Kalvokopiot osoittavat käsiteltyjen asioiden pääkohdat, mutta eivät sovellu itsenäiseen opiskeluun.

Kotisivulla lisäksi tietoja ym.

- kannattaa seurata!

Harjoitustyö tehdään 2 hengen ryhmissä

- aloitus 22.1.
aiheet, tavoite, työnjako, arvostelu
- kurssin suoritukseen pakollinen

ACM:n CR-luokitus:

H.3 INFORMATION STORAGE AND RETRIEVAL**H.3.0 General****H.3.1 Content Analysis and Indexing**

Abstracting methods
Dictionaries
Indexing methods
Linguistic processing
Thesauruses

H.3.2 Information Storage

File organization
Record classification [**]

H.3.3 Information Search and Retrieval

Clustering
Information filtering
Query formulation
Relevance feedback
Retrieval models
Search process
Selection process

H.3.4 Systems and Software

Current awareness systems
(selective dissemination of
information--SDI) [**]

Distributed systems

Information networks

Performance evaluation

(efficiency and effectiveness)

Question-answering (fact retrieval)
systems [**]

User profiles and alert services

H.3.5 Online Information Services

Commercial services

Data sharing

Web-based services

H.3.6 Library Automation

Large text archives

H.3.7 Digital Libraries

Collection

Dissemination

Standards

Systems issues

User issues

H.3.m Miscellaneous