



Faglig kontakt under eksamen:
Institutt for datateknikk og informasjonsvitenskap
Heri Ramampiaro, 73593440

EKSAMEN I EMNE IT2801 INFORMASJONGJENFINNING

Onsdag 18. desember 2008.
Tid: kl 09.00 – 13.00 (4 timer)

BOKMÅL

Hjelpemidler: D – Ingen trykte eller håndskrevne tillatt. **Kun** typegodkjent kalkulator er tillatt.

Sensuren faller: 18. januar 2008.

Svar **kort og konsist** på alle spørsmålene. **Stikkord** foretrekkes fremfor lange forklaringer.

Oppgave I (25%)

1. Data vs. Informasjonsgjenfinning
 - a. Hvilke tre viktige karakteristikk som skiller data gjenfinning fra informasjonsgjenfinning? (2%)
 - b. Forklar i stikkordsform hva er hensikten med informasjonsgjenfinningssystem? (1.5%)
 - c. Gi to eksempler hver på datagjenfinnings og informasjonsgjenfinningsapplikasjoner. (1.5%)
2. Forklar hvilke roller ”*index terms*” har i informasjonsgjenfinningssammenheng? (4%)
3. Hvilke 5 viktige tekstoperasjoner bør man kjøre for å forenkle gjenfinning av informasjonen? Forklar. Hvilke hensikter har disse operasjonene? (5%)
4. Forklar kort prinsippene bak invertert indeks (inverted index). Bruk gjerne eksempel til å støtte forklaringa di. (5%)
5. Forklar hvordan *Standard Rocchio*-formelen ser ut og hva den brukes til. Bruk eksempel til å støtte forklaringen din. (6%)

Oppgave II (25%)

1. Tegn et blokkdiagram (med firkanter og piler) over en generell arkitektur til multimedia informasjonsgjenfinningssystem (MIRS). Forklar kort hvilke(n) rolle(r) hver av blokkene har. (6%)
2. Læreboka tegner opp en figur over multimedia datamodell. Hvilke tre lag (layers) denne modellen består av? Gi minst et eksempel på noe som hører til hvert lag. (6%)

3. "Feature" er et sentralt begrep i multimedia gjenfinning. Hva er hensiktene med "features"? Hvilke tre krav ekstrahering av features (feature extraction) bør oppfylle? (4%)
4. Audio kan klassifiseres til tale og musikk.
 - a. Lag en tabell som forklarer forskjellene mellom disse to audiotypene. Du må ha med minst 4 features. (2%)
 - b. Vis hvordan du kan bruke disse forskjellene til å klassifisere audio ved hjelp av steg for steg klassifikasjon (step-by-step classification). Tips: Tegn opp et flytdiagram som viser stegene. (2%)
 - c. Er det mulig å skille tale og musikk basert på frekvensspekteret (frequency spectrum)? Begrunn svaret ditt. (2%)
5. List opp hvilke aktuelle features som kan brukes til gjenfinning av bilder. (3%)

Oppgave III (30%)

1. Sammenlikn sannsynlighetsmodellen (Probabilistic model) og vektorbasert modellen (Vector Space Model). Hvilke ville du foretrekke dersom du skulle lage en tekstgjenfinningssystem? Begrunn svaret ditt. (7%)
2. Hvilke mål kan brukes for å involvere brukeren i evalueringen av et informasjonsgjenfinningssystem (IR-system)? Forklar. (4%)
3. Utvidelse av spørringer.
 - a. Forklar hva er hensiktene med utvidelse av spørringer (queries). (2%)
 - b. Forklar kort prinsippene med User Relevant Feedback. Bruk gjerne figur til å støtte forklaringen din. (3%)
 - c. Hva er hovedforskjellen mellom "automatic local analysis" og "automatic global analysis"? (4%)
4. Forklar prinsippene bak "association clusters". Hva brukes det til? (4%)
5. "Suffix tree" og "suffix trie" brukes ofte til å bygge opp indekser. Konstruer et "suffix trie" og "suffix tre" av følgende tekst: (6%)
 "A tornado can cause sever destructions. Twister is a movie about a tornado"

Oppgave IV (20%)

Svar rett/galt på følgende utsagn. Hvert **riktig** og **begrunnet** svar belønnes med **2** poeng. **Feilsvar** får **-1,5** poeng. Mens **ubegrunnet** eller **ingen svar** er **0** poeng.

1. Fargepiksler kan ikke brukes til å sammenlikne to bilder.
(Rett/Galt)
2. Det er praktisk å sammenlikne avstanden mellom to audio filer ved å summere forskjellene mellom samplene (samples) i de to filene.
(Rett/Galt)
3. R-frame er en betegnelse for gjennomsnittsbildet i en bildesamling.
(Rett/Galt)
4. Thesaurus kan brukes i utvidelse av spørringer.
(Rett/Galt)
5. Crawlere brukes i distribuerte web-søkemotorer.
(Rett/Galt)

6. Bildeinformasjon kan gjenfinnes ved hjelp av gjenfinningsmetoder for andre media.
(Rett/Galt)
7. En SQL-database er godt egnet til multimediegjenfinning.
(Rett/Galt)
8. Audiogjenfinningsystemer kan bruke teknikker kjent fra tekstgjenfinningsystemer. (Rett/Galt)
9. Bevegelsesinformasjon er ikke så nyttig som feature til å gjenfinne video. (Rett/Galt)
10. *Micon* er en viktig feature for bilder og brukes i bildegjenfinning, og kan sidestilles med index terms for tekst gjenfinning.
(Rett/Galt)