

## ITIA4 Tiedonhaun perusteet

Tentti 11.11.2016

*Kaikki tehtävät arvostellaan pistein 0 – 5.*

1. Alla on kaksi väittämää, joissa oletetaan, että karsinta tehdään indeksiavaimille ja kyselyjen avaimille. Pitääkö toinen niistä, molemmat vai ei kumpikaan paikkaansa? Perustele miksi.
  - a) Stemmaus eli karsinta ei koskaan alenna tarkkuutta täystäsmäyttävässä, Boolean mallia noudattavassa hakujärjestelmässä.
  - b) Stemmaus eli karsinta ei koskaan alenna saantia täystäsmäyttävässä, Boolean mallia noudattavassa hakujärjestelmässä.
  
2. Määrittele lyhyesti seuraavat käsitteet:
  - a) keskitarkkuuksien keskiarvo eli mean average precision (MAP)
  - b) AND-operaatio
  - c) sanamuotojen normalisointi
  - d) aihelevanssi
  - e) tf\*idf-painotus
  
3. Mitkä ovat hakukoneen keskeiset komponentit ja toiminnot? [ohje: kaavakuvin tuettu essee, pituus avoin].
  
4. Mitä ongelmia luonnollisen kielen morfologiset, syntaksiset ja semanttiset ilmiöt aiheuttavat tiedonhauulle? [ohje: kaavakuvin tuettu essee, pituus avoin].