

TAUSTATIIETOTESTI JA YHTEENVETO VASTAUKSISTA

Oheisen kysymysjoukon avulla on tarkoitus saada tietoa siitä, minkälaiset taustatiedot osallistujilla on kurssin tematiikkaan ja menetelmistöön liittyen. Vastauksia käytetään vain opetuksen ohjaamisen apuna, ei opiskelijoiden arviointiin esimerkiksi kurssin suorittamisen osana.

Vastauksen antoivat ensimmäiselle luennolle osallistuneet 17 opiskelijaa. Kävi ilmi, että kysymysjoukon puitteissa tarkasteltuna lähtötaso on hyvä eli suurin osa vastasi suurimpaan osaan kysymyksistä oikein. Kysymyksessä 10 on kaksi vastausvaihtoehtoa, joita voidaan pitää oikeina.

Sulkeissa on esitetty, kuinka moni valitsi k.o. vastausvaihtoehdon.

Valitse seuraaviin kysymyksiin ja väittämiin yksi vastausvaihtoehto.

1. Todennäköisyys sille, että kahta noppaa heitettäessä saadaan niiden summaksi 11 on
a) 1/6 **b) 1/18 (16)** c) 1/9 d) 1/36 (1)
2. Todennäköisyysjakauma kertoo
a) mikä on tietyn satunnaismuuttujan yhden arvon todennäköisyys
b) miten todennäköisyysmassa jakautuu satunnaismuuttujan eri arvojen kesken (17)
c) miksi todennäköiset tapaukset esiintyvät useammin kuin vähemmän todennäköiset
d) miten todennäköisyysmassa jakautuu satunnaismuuttujan suhteen eri ajanhetkillä
3. Normaalijakauman synonyymi on
a) gaussinen (17) b) binomijakauma c) bayesilainen
4. MDL on lyhenne sanoista
a) Maximum Derivation Length (2)
b) Mean Derivation Limit (3)
c) Minimum Description Length (7)
d) Mean Description Limit (3)

Yksi vastaajista oli kiinnittänyt huomionsa siihen, että alunperin sanasta Description puuttui yksi i-kirjain. Ihmisen kielikyvyllä on sinänsä ominaista pystyä ymmärtämään ilmaisuja, vaikka mukana olisi melko runsaastikin kohinaa (esimerkiksi satunnaisia kirjoitusvirheitä).

5. Diskurssi viittaa
a) uutiseen **b) keskusteluun (11)** c) merkitykseen (2)
6. Lauseenjäsennyksellä pyritään
a) löytämään lauseen tärkeimmät sanat (1)
b) muodostamaan lauseen rakenteen kuvaus (16)
c) poistamaan lauseesta monitulkintaisuudet
d) kääntämään lause toiselle kielelle

Lauseenjäsennyksen määritelmän mukainen kuvaus on vaihtoehto B. Lauseenjäsennyksellä voidaan sinänsä ajatella edistettävän myös muiden vaihtoehtojen mukaisia tehtäviä.

7. Disambigoinnilla voidaan pyrkiä
a) löytämään lauseen tärkeimmät sanat

- b) muodostamaan lauseen rakenteen kuvaus
- c) poistamaan lauseesta monitulkintaisuudet (17)**
- d) kääntämään lause toiselle kielelle

Ehkäpä termi disambiguointi sellaisenaan ei ollut kaikille etukäteen tuttu, mutta vaihtoehto C osattiin valita viisaasti hyödyntämällä tietoa englannin kielen sanan 'ambiguous' merkityksestä.

8. SGML on hyvä väline
- a) tekstiaineistojen esikäsittelyyn
 - b) tekstiaineistojen osien ominaisuuksien ja rakenteen kuvaamiseen (14)**
 - c) lauseenjäsennyssääntöjen kuvaamiseen (2)
 - d) kielenkääntämiseen

SGML on tekstiaineistojen merkkäuskieli, joka on XML:n taustalla. Kumpikaan niistä ei ole kurssin sisältöosuutta, mutta kysymyksellä testattiin lähiterminologian tuntemusta.

9. Perl on hyvä ohjelmointikieli
- a) tekstiaineistojen esikäsittelyyn (12)**
 - b) tekstiaineistojen osien ominaisuuksien ja rakenteen kuvaamiseen (1)
 - c) lauseenjäsennyssääntöjen kuvaamiseen (3)
 - d) kielenkääntämiseen (1)

Perl on hyvä kieli tekstiaineistojen esikäsittelyyn ja sitä voidaan hyödyntää kurssin harjoitustyössä. Myös muut välineet tulevat kyseeseen eli harjoitustyössä käytettäviä välineitä voi valita varsin vapaasti hyödyntäen tarkoituksenmukaisesti omia aiemmin hankittuja taitoja.

10. Prolog on hyvä ohjelmointikieli
- a) tekstiaineistojen esikäsittelyyn (3)
 - b) tekstiaineistojen osien ominaisuuksien ja rakenteen kuvaamiseen (1)
 - c) lauseenjäsennyssääntöjen kuvaamiseen (10)**
 - d) kielenkääntämiseen (myös mahdollinen järkevä vastaus) (1)**

11. Luonnollisessa kielissä esiintyvillä sanoilla on luonteenomaista
- a) yksikäsitteisyys
 - b) monitulkintaisuus (17)**
 - c) tasajakautuneisuus

12. Englanninkielisellä sanalla 'spring' on useampia merkityksiä; oikean tulkinnan löytämiseksi on hyödyllistä tietää sanan
- a) konteksti (17)**
 - b) Kullback-Leibler -divergenssi
 - c) kulmakerroin
 - d) morfologinen analyysi

13. Klusteroinnilla pyritään löytämään
- a) lauseen oikea tulkinta
 - b) approksimoimaan jotakin datajoukkoa niin, että muodostetaan lineaarinen malli, joka kulkee datapistejoukon 'läpi' mahdollisimman optimaalisesti (1)
 - c) löytämään datapisteitä, joiden suhteen annettu datapiste on lähimpänä
 - d) löytämään datasta joukko ryppäitä niin, että yhden ryppään sisällä olevat alkio ovat samankaltaisempia kuin jos ne olisi poimittu kahdesta eri ryppästä (16)**

14. Kun kieltä tarkastellaan tilastollisena ilmiönä, kiinnitetään huomiota
- a) kielen ilmiöiden variaatioon (15)**
 - b) kielen ilmiöiden ehdottomaan säännönmukaisuuteen
 - c) kielen ilmiöiden riippumattomuuteen ajasta, paikasta tai puhujasta (2)