

TAUSTATIIETOTESTI

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.

Valitse seuraaviin kysymyksiin ja väittämiin yksi vastausvaihtoehto.

1. Todennäköisyys sille, että kahta noppaa heitettäessä saadaan niiden summaksi 11 on
a) $1/6$ b) $1/18$ c) $1/9$ d) $1/36$
2. Todennäköisyysjakauma kertoo
a) mikä on tietyn satunnaismuuttujan yhden arvon todennäköisyys
b) miten todennäköisyysmassa jakautuu satunnaismuuttujan eri arvojen kesken
c) miksi todennäköiset tapaukset esiintyvät useammin kuin vähemmän todennäköiset
d) miten todennäköisyysmassa jakautuu satunnaismuuttujan suhteen eri ajanhetkillä
3. Normaalijakauman synonyymi on
a) gaussinen b) binomijakauma c) bayesilainen
4. MDL on lyhenne sanoista
a) Maximum Derivation Length
b) Mean Derivation Limit
c) Minimum Description Length
d) Mean Description Limit
5. Diskurssi viittaa
a) uutiseen b) keskusteluun c) merkitykseen
6. Lauseenjäsennyksellä pyritään
a) löytämään lauseen tärkeimmät sanat
b) muodostamaan lauseen rakenteen kuvaus
c) poistamaan lauseesta monitulkintaisuudet
d) kääntämään lause toiselle kielelle
7. Disambigoinnilla voidaan pyrkiä
a) löytämään lauseen tärkeimmät sanat
b) muodostamaan lauseen rakenteen kuvaus
c) poistamaan lauseesta monitulkintaisuudet
d) kääntämään lause toiselle kielelle
8. SGML on hyvä väline
a) tekstiaineistojen esikäsittelyyn
b) tekstiaineistojen osien ominaisuuksien ja rakenteen kuvaamiseen
c) lauseenjäsennyssääntöjen kuvaamiseen
d) kielenkääntämiseen

9. Perl on hyvä ohjelmointikieli
- tekstiaineistojen esikäsittelyyn
 - tekstiaineistojen osien ominaisuuksien ja rakenteen kuvaamiseen
 - lauseenjäsennyssääntöjen kuvaamiseen
 - kielenkääntämiseen
10. Prolog on hyvä ohjelmointikieli
- tekstiaineistojen esikäsittelyyn
 - tekstiaineistojen osien ominaisuuksien ja rakenteen kuvaamiseen
 - lauseenjäsennyssääntöjen kuvaamiseen
 - kielenkääntämiseen
11. Luonnollisessa kielissä esiintyvillä sanoilla on luonteenomaista
- yksikäsitteisyys
 - monitulkintaisuus
 - tasajakautuneisuus
12. Englanninkielisellä sanalla 'spring' on useampia merkityksiä; oikean tulkinnan löytämiseksi on hyödyllistä tietää sanan
- konteksti
 - Kullback-Leibler -divergenssi
 - kulmakerroin
 - morfologinen analyysi
13. Klusteroinnilla pyritään löytämään
- lauseen oikea tulkinta
 - approksimoimaan jotakin datajoukkoa niin, että muodostetaan lineaarinen malli, joka kulkee datapistejoukon 'läpi' mahdollisimman optimaalisesti
 - löytämään datapisteitä, joiden suhteen annettu datapiste on lähimpänä
 - löytämään datasta joukko ryppäitä niin, että yhden ryppään sisällä olevat alkiot ovat samankaltaisempia kuin jos ne olisi poimittu kahdesta eri ryppästä
14. Kun kieltä tarkastellaan tilastollisena ilmiönä, kiinnitetään huomiota
- kielen ilmiöiden variaatioon
 - kielen ilmiöiden ehdottomaan säännönmukaisuuteen
 - kielen ilmiöiden riippumattomuuteen ajasta, paikasta tai puhujasta