

TEHTÄVÄMÄÄRITTELY

Aihe: Sanaindeksoija

Sovellus lukee käyttäjän antamista tekstitiedostoista sanat hakupuurakenteeseen, jonka jälkeen käyttäjän on mahdollista kysellä sovellukselta sanojen tai niiden yhdistelmien (haluamiensa merkkijonojen) esiintymisestä kyseisissä tekstitiedostoissa.

Sovelluksen toiminnallisuus:

- tekstitiedoston sisäänluku
 1. sisäänluettavien tiedostojen maksimilukumäärän kysyminen
 - a) jos lukumäärä on 0 tai pienempi, ohjelman päättäminen
 - b) jos lukumäärä on suurempi kuin 0
 2. tiedoston nimen kysyminen
 3. tiedoston nimen tarkastus
 - a) jos nimi on tyhjä ja sisäänlukuja on tehty, siirtyminen sanahakuun
 - b) jos nimi on tyhjä, mutta sisäänlukuja ei ole tehty, siirtyminen tilaston tulostamiseen
 - c) jos nimi ei ole tyhjä
 4. tiedoston löytymisen tarkistus
 - a) jos tiedostoa ei löydy, ilmoitus käyttäjälle ja paluu kohtaan 1.
 - b) jos tiedosto löytyy
 5. tiedoston läpiluku ja tallennus hakupuuhun
 6. paluu kohtaan 2.
- sanahaku
 1. haettavan merkkijonon pyytäminen
 2. haettavan merkkijonon tarkastus
 - a) jos merkkijono on tyhjä, siirtyminen tilaston tulostamiseen
 - b) jos merkkijono ei ole tyhjä
 3. annetun merkkijonon haku hakupuurakenteesta
 4. osumien tulostaminen käyttäjälle
 5. osumien lukumäärän tulostaminen käyttäjälle
 6. paluu kohtaan 1.
- tilaston tulostaminen
 1. sisäänluettujen tiedostojen, haettujen merkkijonojen ja osumien lukumäärien tulostaminen käyttäjälle
 2. käytön päättäminen

Toteutettavat algoritmit ja tietorakenteet:

- String-taulukko, johon tallennetaan tiedostojen nimet ja viittaukset sisältöolioon
- kullekin tiedostolle luodaan sisältöolio, jonka String-taulukkoon tallennetaan tiedostosta luetut rivit
- Trie-puu, johon tallennetaan maksimissaan yhden rivin pituisia merkkijonoja
 - rajaukset

- tallennettavat merkkijonot aloitetaan sanojen alusta, joten haku ei löydä merkkijonoja sanan keskeltä
- tekstitiedostoa käsitellään rivi kerrallaan, joten haettavan merkkijonon tulee sijaita yhdellä rivillä, jotta se löytyy
- rakenteet
 - merkkijonojen 1. kirjaimien tallennusta varten maksimitaulukko, jonka alkio toimii aloitussolmuna linkitetylle listalle
 - linkitetyt listat, joihin tallennetaan aloitussolmusta lähtevät kirjainjärjestykset
 - kukin linkitetyn listan solmu toimii aloitussolmuna seuraavalle linkitetylle listalle
- int-tila, johon tallennetaan kunkin tekstitiedoston osuuden lukumäärät
- tehtyjä valintoja muutetaan tarvittaessa tehokkaampiin ja toimivampiin toteutuksen edistyessä

Syötteet:

- ascii-muotoisia tekstitiedostoja
- käyttäjä antaa sovellukselle tiedoston nimen
- tiedoston on sijaittava samassa hakemistossa sovelluksen kanssa, jotta sovellus sen löytää

Tavoiteltava aika- ja tilavaativuus:

- tekstitiedoston sisäänluku
- sanahaku
- tulosten tulostaminen

Lähteet:

- kurssisivulla annetut materiaalit
- Tira-kurssin (kevät 2012) materiaalit
- Java 7 API Specification