Kati Kyllönen Opiskelijanumero: 11539913

# Tira harjoitustyö – Sanaindeksointi – Määrittelydokumentti

## Sisällysluettelo

# 1 Tira Harjoitustyö

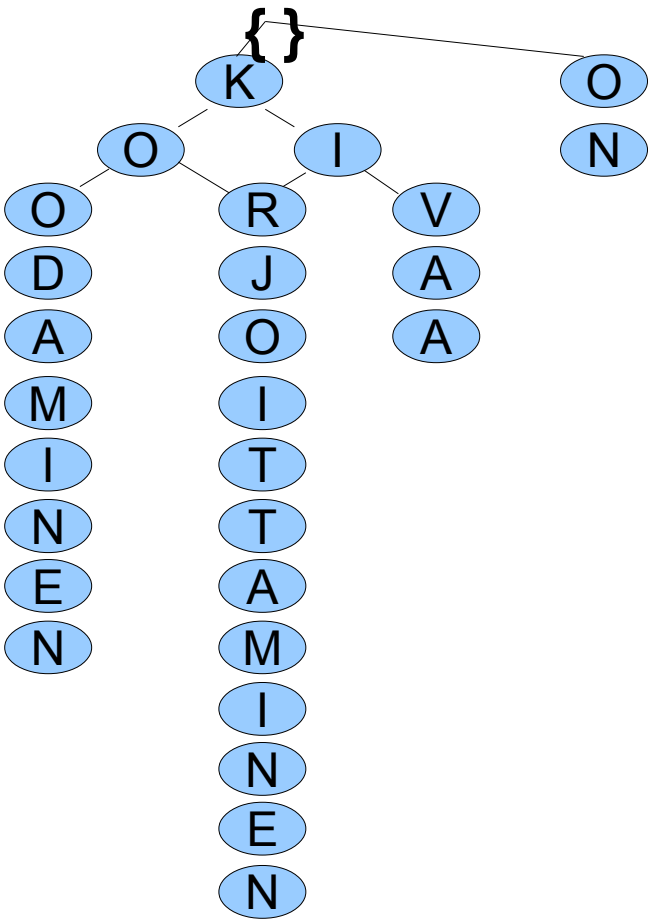
## 1.1 Aihe: Sanaindeksointi

Sanaindeksointiohjelma etsii syötteenä saamistaan tekstitiedostoista haettavan merkkijonon esiintymistiheyden sekä esiintymiskohdat kussakin tiedostossa. Haun suoritettuaan ohjelma tulostaa kaikki tiedostojen ne rivit, joilla haettavat merkkijonot esiintyvät.

Harjoitustyön tarkoituksena on haun toteuttamisen lisäksi tarkkailla ja dokumentoida Trie-hakupuun toteutusratkaisujen välisiä suorituseroja.

## 1.2 Käytettävät algoritmit

Ohjelma muodostaa syötteenä saamistaan tiedostoista Trie-hakupuun, jonka avulla käyttäjän antaman merkkijonon esiintyminen havaitaan. Trie rakenteeseen merkkijonot talletetaan siten, että kukin solmu sisältää yhden kirjaimen. Merkkijonon sisältö saadaan selville lukemalla solmujen arvot juuresta lehteen asti.




*Piirros 1: Esimerkki merkkijonojen tallentamisesta Trie hakupuuhun*

## 1.3 Käytettävät tietorakenteet

Trie rakenteen lapset talletetaan järjestettyyn linkitettyyn listaan sekä dynaamiseen järjestettyyn taulukkoon, jonka kokoa lisätään tarpeen vaatiessa. Harjoitustyössä tarkastellaan ärjestämisajankohdan sekä tiheyden vaikutusta ohjelman suorituskykyyn. Työn edetessä on mahdollista sisällyttää muitakin tietorakenteita, kuten hajautustaulu työn laajuuden niin vaatiessa.

## 1.4 Ohjelman syötteet

Ohjelma kysyy käyttäjältä tiedostoja jotka ladataan hakupuurakenteeseen. Ohjelma kysyy käyttäjältä haettavia merkkijonoja.

## 1.5 Tavoiteltavat aika- ja tilavaativuudet

Ohjelmiston tavoiteltavat aika- ja tilavaativuudet tarkentuvat työn edetessä. Alla lähdemateriaalin (Viite 1: Trie-rakenteen esittely)määrittelemiä aika- ja tilavaativuuksia Trie-rakenteelle. Merkinnällä n tarkoitetaan seuraavassa solmun kaikkien lasten lukumäärää.

1. Dynaamisen järjestetyn taulukkoon

- lisääminen aikavaativuudeltaan O(n)
- hakeminen aikavaativuudeltaan O (log2 n)
- tilavaativuus O(n)

2. Linkitettyyn listaan

    - lisääminen aikavaativuudeltaan O(n)
    - hakeminen aikavaativuudeltaan O(n)
    - tilavaativuudeltaan O(n)

# 2 Lähteet

Tietorakenteet kurssin 58131 kevät 2012 kurssimateriaali: Patrik Floréen

http://www.cs.helsinki.fi/u/floreen/tira2012/tira.pdf

Trie rakenteen esittely: Esa Junttila

http://www.cs.helsinki.fi/u/ejunttil/opetus/tiraharjoitus/trie.html

Algoritmit 2, Luento 6: Timo Männikkö

http://users.jyu.fi/~mannikko/algoritmit2/luennot/luento6.pdf