Ideana on tehdä pienehkö mygrep-niminen ohjelma, jonka esikuva on Linux:sta löytyvä grep-työkalu. Grep:ssä ominaisuuksia riittää mutta meidän ohjelma keskittyy ihan sen ytimeen. Grep-ohjelman käyttöohje löytyy tämän dokumentin perästä. Saman löydät Linux:sta komennolla ”man grep” tai vaikka samoilla termeillä google-hakuna. Grep-työkalulla voit etsiä tiedostosta rivit jotka sisältävät haluamasi merkkijonon.

Työ tehdään inkrementteinä joista ensimmäisen tekeminen takaa yhden pisteen projektista; toisen kolme pistettä, kolmannen neljä pistettä ja neljän inkrementin tekeminen antaa kaikki viisi pistettä. Projekti arvioidaan siis pistemäärällä 0 – 5 pistettä ja sen vaikutus kokonaisarvosanaan on kerrottu toisaalla. Inkrementit tulee tehdä järjestyksessä alusta alkaen ja kaikkien inkrementtien tulee toimia siihen inkrementtiin asti jota vastaavaa pistemäärää tavoittelet. Esim. viiden pisteen ohjelmasta tulee löytyä yhden pisteen ohjelman toiminnot. Yhdestä pisteestä hypätään suoraan kolmeen sen vuoksi että työmäärässä tapahtuu myös vähän isompi nousu tuossa kohtaa.

## Inkrementti (takaa yhden pisteen)

Ohjelma käynnistetään komentoriviltä alla listatulla tavalla. Ajoesimerkki on otettu Mac-tietokoneelta; voit ajaa ohjelmaa samalla tavalla myös Windows 10:n Cmd-työkalussa. Tässä ajettavan ohjelman (binääri) nimi on mygrep ja ./mygrep tarkoittaa sitä että se käynnistetään nykyisestä työhakemistosta:

./mygrep

Give a string from which to search for: Erkki Hietalahti

Give search string: rkki

"rkki" found in "Erkki Hietalahti" in position 1

Alla toinen ajo:

./mygrep

Give a string from which to search for: Erkki Hietalahti

Give search string: rkkk

"rkkk" NOT found in "Erkki Hietalahti"

Kuten esimerkeistä näkyy ohjelma kysyy käyttäjältä ”ison” merkkijonon josta etsitään annettua pienempää merkkijonoa. Jälkimmäinen voi löytyä isosta merkkijonosta mistä kohtaa tahansa. Jos etsittävä merkkijono löytyy isosta merkkijonosta sen esiintymiskohta kerrotaan positiosta 0 alkaen. Kaikissa tapauksissa tulostetaan ilmoitus löytymisestä.

Periaatteessa voit ajaa tätä ohjelmaa myös käyttämäsi IDE-välineen (vaikkapa Microsoft Visual Studio) valikosta tai pikanäppäinyhdistelmän avulla mutta jatkossa ohjelmaa kannattaa ajaa komentoriviltä koska sen käyttäminen perustuu oleellisesti komentoriviargumenttien antamiseen ja niitä on IDE-välineellä tyypillisesti kömpelö asettaa.

## Inkrementti (tämä ja 1. inkrementti takaavat kolme pistettä jne.)

Kuten aiemmin sanottiin tämän inkrementin tulee sisältää myös edellisen inkrementin toiminnallisuuden eli jos edelleen ohjelma käynnistetään pelkällä ohjelmatiedoston (binääri) nimellä se toimii kuten edellisessä luvussa on kerrottu.

Mutta jos ohjelma käynnistetään alla kerrotulla tavalla komentoriviltä sen käyttö alkaa muistuttamaan perus-grep -käyttöä komentoriviargumentteineen. Testaa siis ohjelmaasi tämän jälkeen suoraan komentoriviltä.

Alla ajoesimerkki:

./mygrep following man\_grep\_plain\_ASCII.txt

The following options are available:

The grep utility exits with one of the following values:

Ohjelma käy lävitse sille viimeisenä annetun ASCII-muotoisen tiedoston sisällön rivi riviltä ja jos kulloisenkin rivin sisältä löytyy jostain kohtaa haettu merkkijono (2. komentoriviargumentti) ko. rivi tulostetaan; muutoin ei. Toiminta on täsmälleen saman kaltaista kuin grep:in toiminta sen yksinkertaisimmassa käyttötilanteessa.

Annan teidän käyttöönne tekstitiedoston man\_grep\_plain\_ASCII.txt joka sisältää nimensä mukaisesti grep:in manuaalisivun raakana ASCII-tekstinä. Voitte käyttää sitä ohjelmanne testaamisessa; teidän ohjelman tulee löytää samalla tavoin nuo esimerkin kaksi tekstiriviä mitkä näkyvät yo. tulostuksessa jos annatte ohjelmallenne samat komentoriviargumentit ja käytätte samaa man\_grep\_plain\_ASCII.txt -tiedostoa.

## Inkrementti (kaikki tähänastiset inkrementit takaavat neljä pistettä)

Tähän inkrementtiin on poimittu mallina muutama grep:in tekemä yksinkertainen tilastollinen analyysi. Eli sinun mygrep:iä pitää pystyä ajamaan niin että:

* Haluttaessa se tulostaa löydettyjen rivien eteen niiden rivinumerot isossa tiedostossa
* Haluttaessa ajon lopussa mygrep tulostaa tiedon siitä montako riviä tiedostossa oli joista löytyi hakemasi merkkijonon esiintymä.

Näitä molempia ominaisuuksia säädetään komentoriviargumenttien avulla: jos rivinumerointi halutaan siitä ilmoitetaan komentoriviargumentilla; sama pätee rivien lukumäärän laskentaan.

Alla esimerkki ajamisesta:

./mygrep -olo following man\_grep\_plain\_ASCII.txt

32: The following options are available:

245: The grep utility exits with one of the following values:

Occurrences of lines containing "following": 2

Edellisessä tulostuksessa tulostettujen rivien numerot näkyvät rivien edessä; man\_grep\_plain\_ASCII.txt -tiedoston rivillä 32 on täsmälleen tulostuksessa näkyvä sisältö. Samoin löytyneitä rivejä on kaksi kappaletta.

Kuten näet ajoesimerkistä nämä ylimääräiset valinnaiset optiot tarjotaan muodossa -olo. Voit tulkita tämän tarkoittavan sitä että ”-o” kertoo että sen perässä luetellaan haluttuja ohjelman ajoon vaikuttavia optioita (engl. options) ja l = rivinumerointi (engl. line numbering) ja o = löytyneiden rivien lukumääräoptio (engl. occurrences). Nyt komentoriviargumentteja on neljä kolmen sijaan.

Jos optiomääritys on -ol vain rivinumerot tulostetaan ja jos se on -oo vain esiintyneiden rivien lukumäärä tulostetaan. Käyttäjä voi antaa siten näitä valikoiden.

Huom: kun tarjoat tätä inkrementtia opettajalle tarkastukseen myös edellisten inkrementtien pitää toimia!

## Inkrementti (kaikki tähänastiset inkrementit takaavat viisi pistettä)

Tähän inkrementtiin on upotettu seuraavat lisävaatimukset edellisten inkrementtien vaatimusten lisäksi:

* Ohjelmassa pitää soveltaa jossain kohtaa poikkeuskäsittelyä. Luontainen kohta sen soveltamiseen on tiedostokäsittely: onko mainittu tiedosto olemassa, onko siihen lukuoikeus jne. ?
* Haluttaessa merkkijonohaku riviltä tehdään niin että isot ja pienet kirjaimet ovat saman arvoisia. Silloin esim. etsittävät merkkijonot ”erkki”, ”Erkki”, ”eRkKi” ja ”ERKKI” jne. ovat samoja; erkki-merkkijonoa etsitään oli se kirjoitettu millä tavalla tahansa (isoja ja pieniä kirjaimia sekaisin).
* Haluttaessa mygrep ohjelmasi etsiikin ne rivit joilla **ei ole** antamaasi merkkijonoa.

Alta löydät ajoesimerkkejä kustakin vaatimuksesta; niistä näet myös vaaditut optioasetukset komentorivillä.

./mygrep -or following man\_grep\_plain\_ASCII.txt

GREP(1) BSD General Commands Manual GREP(1)

NAME

grep, egrep, fgrep, zgrep, zegrep, zfgrep -- file pattern searcher

SYNOPSIS

grep [-abcdDEFGHhIiJLlmnOopqRSsUVvwxZ] [-A num] [-B num] [-C[num]]

[-e pattern] [-f file] [--binary-files=value] [--color[=when]]

[--colour[=when]] [--context[=num]] [--label] [--line-buffered]

[--null] [pattern] [file ...]

DESCRIPTION

The grep utility searches any given input files, selecting lines that

match one or more patterns. By default, a pattern matches an input line

<… JA PALJON PALJON LISÄÄ RIVEJÄ; ITSE ASIASSA KAIKKI RIVIT POISLUKIEN NE

KAKSI JOISTA LÖYTYY SANA following …>

Optiomäärittelyssä r tarkoittaa käänteistä etsimistä eli etsitään rivejä joilla **ei** ole haettua merkkijonoa (r = engl. reverse search).

Toinen ajoesimerkki:

./mygrep -oi folloWING man\_grep\_plain\_ASCII.txt

The following options are available:

The grep utility exits with one of the following values:

Optio i tarkoittaa että isot ja pienet kirjaimet ovat samanarvoiset (i = engl. ignore case). Siten folloWING löytyy kyllä riviltä jonka osana on teksti following.

Kolmas ajoesimerkki jossa on kaikki optiot päällä:

./mygrep -olori folloWING man\_grep\_plain\_ASCII.txt

1:

2:GREP(1) BSD General Commands Manual GREP(1)

3:

4:NAME

5: grep, egrep, fgrep, zgrep, zegrep, zfgrep -- file pattern searcher

6:

7:SYNOPSIS

8: grep [-abcdDEFGHhIiJLlmnOopqRSsUVvwxZ] [-A num] [-B num] [-C[num]]

9: [-e pattern] [-f file] [--binary-files=value] [--color[=when]]

10: [--colour[=when]] [--context[=num]] [--label] [--line-buffered]

11: [--null] [pattern] [file ...]

12:

13:DESCRIPTION

14: The grep utility searches any given input files, selecting lines that

15: match one or more patterns. By default, a pattern matches an input line

<… TÄSSÄ VÄLISSÄ ON PALJON TULOSTUVIA RIVEJÄ RIVINUMEROINEEN …>

Occurrences of lines NOT containing "folloWING": 299

-olori tarkoittaa: tulostetaan rivinumerot (l), tulostetaan niiden rivien lukumäärä joista ei löytynyt (o), tehdään käänteinen etsiminen eli etsitään rivit joilta merkkijonoa ei löydy (r), ei tehdä eroa isoille ja pienille kirjaimille (i).

Neljäs ajoesimerkki jossa näkyy poikkeuskäsittely:

mv man\_grep\_plain\_ASCII.txt man\_grep\_plain\_ASCII.txt2

./mygrep -olori folloWING man\_grep\_plain\_ASCII.txt

An exception occurred. Exception Nr. -1

Could not find out the size of file "man\_grep\_plain\_ASCII.txt"

Ohjelmassa on siis varauduttu siihen että sille annetaan parametrina olematon tiedosto. Ensimmäinen kohta missä ohjelma käsittelee ko. tiedostoa on sen koon selvittäminen tavuina; juuri tämä käsittely epäonnistuu ensiksi tiedoston puuttumisen takia ja siihen on varauduttu poikkeuskäsittelyllä.