
Funktionaalisen- ja olioparadigman suorituskykyvertailu Scala-kielessä

LuK-tutkielma
Turun yliopisto
Tulevaisuuden teknologioiden laitos
Labra
2019
Jaakko Paju

Sisältö

1	Johdanto	1
2	Ohjelmointiparadigmat	3
2.1	Olioparadigma	3
2.2	Funktionaalinen paradigma	4
2.3	Paradigmoille tyypillisiä ominaisuuksia	4
3	Scala	6
3.1	Muuttujat	6
3.2	Tietotyytit	7
3.3	Kontrollirakenteet	8
3.3.1	Ehtolauseet	8
3.3.2	Silmukat	8
3.4	Metodit ja funktiot	9
3.4.1	Funktioliteraalit	10
3.4.2	Korkeamman asteen funktiot	10
3.5	Luokat ja oliot	11
3.5.1	Class	11
3.5.2	Case class	11
3.5.3	Singleton-oliot	12
3.5.4	Piirretyypit	13

4	Kokoelmat	14
4.1	Järjestykselliset kokoelmat	15
4.2	Suorituskykyvertailut	17
4.2.1	Iterointi	17
4.2.2	Satunnainen haku	18
4.2.3	Kokoelman luominen	19
4.2.4	Kokoelman loppuun lisääminen	21
4.2.5	Häntäoperaatio	22
5	Yhteenveto	24
	Lähdeluettelo	25

1 Johdanto

Funktionaalinen ohjelmointi kasvattaa suosiotaan jatkuvasti. Useisiin yleiskäyttöisiin ja alunperin imperatiivisiin ohjelmointikieliin on lisätty ominaisuuksia funktionaalisesta ohjelmoinnista. Esimerkiksi suosittu oliokielet Java, Python ja C++ ovat kaikki lainanneet funktionaalisesta ohjelmoinnista anonyymit sekä korkeamman asteen funktiot.

Funktionaalisen paradigman keskeisistä käsitteitä ovat muuttumattomat arvot ja tietorakenteet. Muuttumattomien arvojen käyttäminen usein lisää kopioimisen tarvetta verrattuna ohjelmiin, joissa käytetään muuttuvia arvoja. Kopioimisen seurauksena muistia pitää varata ja vapauttaa useammin kuin muuttuvia arvoja käytettäessä. Tämä herättää kysymyksen funktionaalisten ohjelmien suorituskyvystä verrattuna imperatiivisiin ohjelmiin.

Tutkielmassa perehdytään tarkastelemaan millaisia vaikutuksia funktionaalisella paradigmalla on suorituskykyyn verrattuna olioparadigmaan. Tutkielman suorituskykyvertailut keskittyvät Scala-ohjelmointikieleen, sillä se on suunniteltu tukemaan sekä funktionaalista että olioparadigmaa, jolloin suorituskyvyn vertailu näiden kahden paradigman välillä on mielekästä ja suoraviivaista. Tarkastelu kohdistuu erityisesti muuttumattomiin kokoelmiin. **TODO: Lisää toinen tutkittava aihe** Tutkielma on suoritettu perehtymällä aiheita käsittelevään kirjallisuuteen.

Luvussa 2 esitellään molemmat vertailun kohteena olevat paradigmat. Luvussa 3 esitellään Scala-kielen rakenteet, ja miten ne tukevat kumpaakin paradigmaa. Luvussa 4 tarkastellaan Scalan standardikirjaston muuttumattomien kokoelmien suorituskykyä ja ver-

rataan sitä muuttuvien kokoelmien suorituskyyyn. Viimeisenä luvussa 5 esitellään joh-
topäätökset ja kootaan tutkielman tulokset.

TODO: Lisää esittely toisesta tutkittavasta aiheesta

2 Ohjelmointiparadigmat

Ohjelmointiparadigmat luokittelevat ohjelmointikieliä sen perusteella, miten kieli on suunniteltu mallintamaan ongelmia ja minkälaisia mekanismeja kieli tarjoaa näiden ongelmien ratkaisuun. Yhdessä ohjelmointikielessä voi olla vaikutteita useammasta paradigmasta. Tällaista kieltä kutsutaan *moniparadigmaiseksi* kieleksi. Ohjelmointiparadigmat voi jakaa karkeasti kahteen yläluokkaan: *imperatiivisiin* ja *deklaratiivisiin*. [1, Luku 6]

Imperatiiviset kielet keskittyvät ohjelmointiongelman ratkaisuun määrittelemällä *miten* tietokoneen tulisi ratkaista laskettava ongelma. Ohjelmat siis rakentuvat peräkkäisistä käskyistä, jotka muokkaavat ohjelman tilaa ja näin ratkaisevat ongelman. Imperatiivisista kielistä puhutaan matalan tason kielinä, sillä ongelman ratkaisu mallinnetaan niissä tietokoneen näkökulmasta. [2, Luku 1]

Deklaratiiviset kielet ratkaisevat ongelman vastaamalla kysymykseen *mitä* tietokoneen tulisi tehdä ongelman ratkaisemiseksi. Käytännössä tämä tarkoittaa ongelman ratkaisun kuvailemista ilman toteutuksen yksityiskohtiin uppoutumista. Deklaratiiviset mielletään yleensä korkean tason kielinä, sillä ne mallintavat ongelmanratkaisua ohjelmoijan näkökulmasta. [2, Luku 1]

2.1 Olioparadigma

Olio-ohjelmoinnin perusajatus on kuvata ongelmaa olioiden avulla, jotka kukin kuvaavat jotakin ongelma-alueen käsitettä. Olioiden tarkoitus on kapseloida kuvaamansa käsitteen tieto ja tila sisäänsä, sekä tarjota operaatioita kapseloidun tiedon muokkaamiseen

ja tarkasteluun. olioparadigma on osa imperatiivisia paradigmoja, sillä ongelmanratkaisu tapahtuu peräkkäisillä komennoilla, jotka muuttavat olioiden ja samalla koko ohjelman tilaa kohti ratkaistua ongelmaa. [2, Luku 1] [1, Luku 10]

2.2 Funktionaalinen paradigma

Funktionaaliset ohjelmat rakentuvat yksittäisistä funktiosta, joita yhdistämällä luodaan yhä isompia funktiota. Näillä funktioilla ei ole tilaa, ja ne ovat puhtaita eli eivät aiheuta *sivuvaikutuksia*. Sivuvaikutuksia voivat olla esimerkiksi muuttujan arvon muuttaminen, poikkeuksen nostaminen, tiedoston lukeminen tai kirjoittaminen, pyyntö tietokantaan tai verkon ylitse sekä näytölle piirtäminen. Tilaton ja puhdas funktio palauttaa tilanteesta riippumatta tietyllä syötteellä aina saman arvon. [3, Luku 1]

Ohjelmointikieli ilman sivuvaikutuksia olisi käytännössä hyödytön, joten funktionaaliset ohjelmointikielet tarjoavat erilaisia mekanismeja sivuvaikutusten hallintaan. Funktionaalisen ohjelmoinnin katsotaan yleensä kuuluvan deklaratiiiviseen paradigmaan. [1, Luku 11]

2.3 Paradigmoille tyypillisiä ominaisuuksia

Lauseke on ohjelmointikielen rakenne, jonka suoritus tuottaa arvon. Lauseke voi koostua joko arvosta tai operaattorista, jota on sovellettu yhteen tai useampaan lausekkeeseen. Esimerkiksi $3+2$ on numeerinen lauseke, jossa on yksi operaattori $+$, jota sovelletaan kahteen numeeriseen arvoon 3 ja 2. Lausekkeen arvo voidaan sijoittaa muuttujaan, antaa parametriksi funktioon tai sen arvo voidaan palauttaa funktiosta. Lauseke on minkä tahansa ohjelmointikielen perusrakenne, eli lausekkeita on funktionaalisissa sekä oliokielissä. [1, Luku 6]

Funktionaalisessa ja olioparadigmassa molemmissa on muuttujia, mutta niitä käsitellään eri tavoilla. Funktionaalisissa kielissä muuttujaan sijoitettua arvoa ei voi muuttaa

alustuksen jälkeen, kun taas imperatiivissa kielissä muuttujan arvoa voi yleensä muuttaa. Sama pätee tietorakenteisiin: imperatiivisissa kielissä niiden muuttaminen alustamisen jälkeen on sallittua, funktionaalisissa ei. [3, Luku 3] Funktionaalisissa kielissä funktioita kohdellaan kuin mitä tahansa muitakin arvoja: niitä voi sijoittaa muuttujiin, antaa parametrina toiseen funktioon tai käyttää funktion palautusarvona. Olikielissä tämä ei aina ole mahdollista. [1, Luku 6]

Komento on ohjelmointikielen rakenne, jonka suoritus ei aina tuota arvoa ja saattaa aiheuttaa sivuvaikutuksia. Komennot saattavat esimerkiksi muuttaa olioiden ja muuttujien tilaa, lukea käyttäjän tuottamaa syötettä tai nostaa poikkeuksen. Komentojen suorittamisella saattaa olla sivuvaikutuksia, joten niiden suoritusjärjestys on merkityksellinen. Imperatiivisiin kieliin komennot kuuluvat oleellisesti, mutta funktionaalsiin eivät.

Imperatiivisissa kielissä toisto toteutetaan silmukoilla. Silmukkaa suoritetaan ennalta määriteltä määrä tai kunnes silmukan suoritusta säätelevä ehtolause muuttuu epätodeksi. Ehtolauseen arvo määriytyy muuttujan arvon mukaan, jota muutetaan silmukan suorituksen aikana. Funktionaalisissa kielissä muuttujien arvoa ei voi muuttaa, ja siitä syystä silmukat olisivat hyödyttömiä. Silmukoiden sijaan käytetään rekursiota, eli funktio kutsuu itseään kunnes funktioon määriteltä perustapaus saavutetaan. [1, Luku 6 ja 11]

Kummankin paradigman ominaisuuksissa on omat ongelmakohtansa. Imperatiivisia ohjelmia on tavallisesti vaikea säikeistää, sillä jos useat säikeet käsittelevät samaa oliota samanaikaisesti, saattaa olion tila olla odottamaton. Funktionaalisessa ohjelmassa samaa ongelmaa ei ole, koska luomisen jälkeen olion tila ei muutu, ja sitä voi turvallisesti käyttää useampi säie. [4, Luku 6]

Toiston toteuttaminen rekursion avulla toimii hyvin, kunhan rekursio ei ajaudu liian syväksi. Yleensä tämä tarkoittaa, ettei käsiteltävän datan määrä ole suuri. Jokainen rekursiivinen funktiokutsu kasvattaa kutsupinoa, kunnes tapahtuu ylivuoto ja ohjelma kaatuu. Usein kääntäjä pystyy optimoimaan rekursiiviset kutsut käännösvaiheessa silmukoiksi, jos rekursio on funktion viimeinen operaatio. [4, Luku 8]

3 Scala

Scala on staattisesti tyypitetty moniparadigmainen käännettävä ohjelmointikieli, joka yhdistää funktionaalisen ohjelmoinnin ja olio-ohjelmoinnin ominaisuuksia. Scala on korkean tason kieli ja sen syntaksi on kompaktia ja eleganttia. Scalan kääntäjä ja tyyppijärjestelmä takaavat tyyppiturvallisuuden kääntämisen aikana muutamaa poikkeusta lukuunottamatta. Scala-koodi on tarkoitettu käännettäväksi Javan tavukoodiksi, ja se mahdollistaa Java-kirjastojen käyttämisen suoraan Scalasta. Tavukoodia ajetaan Javan virtuaalikoneessa (JVM). Tässä tutkielmassa tarkastellaan Scalan versiota 2.12.10. [5, Introduction] [4, Luku 2]

3.1 Muuttujat

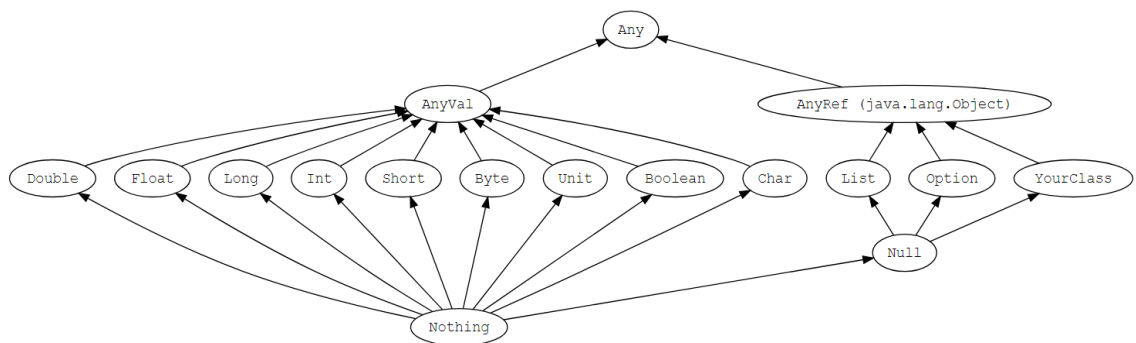
Scalassa on tavallisten muuttujien lisäksi vakioita, jotka ovat muuttujia, joiden arvoa ei alustuksen jälkeen voi muuttaa. Vakion määrittelyyn käytetään avainsanaa **val** ja muuttujan määrittelyyn **var**. Funktionaalisessa Scalassa käytetään pääasiassa **val**-avainsanalla määriteltyjä muuttujia, ja imperatiivisessa tyyliässä voidaan käyttää molempia.

Muuttujan nimen jälkeen määritellään muuttujan tyyppi, joka erotetaan muuttujan nimestä kaksoispisteellä. Esimerkiksi **val x: Int = 1** alustaa kokonaislukuvakion *x*, jonka arvo on 1. Muuttujan tyyppimäärittelyn voi jättää myös kirjoittamatta, sillä Scalan kääntäjä osaa yleensä päätellä muuttujan tyypin asetetun arvon perusteella. Edellisen esimerkin voi siis halutessaan kirjoittaa muodossa **val x = 1**, ja Scala osaa päätellä muuttujan olevan tyypiltään kokonaisluku. [5, Basics]

Muuttujat ovat myös staattisesti tyypitettyjä, joten muuttujan tyyppi ei voi muuttua ajon aikana. Scalassa myös funktiot ovat arvoja, joten niitä voidaan sijoittaa muuttujiin. Esimerkiksi `val add1 = (x: Int) => x + 1` luo *funktioliteraalin* ja sijoittaa sen muuttujaan nimeltä `add1`. Funktioita käsitellään tarkemmin luvussa 3.4. [4, Luku 1]

3.2 Tietotyypit

Scala on puhtaasti oliokieli, eli kaikki arvot ja muuttujat ovat olioita. Jokainen Scalan tietotyyppi perii luokan `Any` ja luokka `Nothing` on jokaisen tietotyypin alaluokka. Scalassa luokka voi olla arvoluokka tai viiteluokka. Arvoluokat perivät luokan `AnyVal` ja niihin ei voi sijoittaa null-arvoa. Viiteluokat perivät luokan `AnyRef`, joka on tyypialias Javan luokalle `Object`. Jokaisen viiteluokan alaluokka on `Null`. Tietotyyppien hierarkia on kuvattu kuvassa 3.1. [4, Luku 5]



Kuva 3.1: Scalan luokkahierarkia

Alkeistietotyyppinä kuvaavat arvoluokat `Byte`, `Short`, `Int`, `Long`, `Char`, `Float`, `Double`, `Boolean` ja `Unit`. Kaikki paitsi `Unit` vastaavat Javan vastaavia kääreluokkia. `Unit` on erityinen tietotyyppi, joka kuvastaa funktion tyhjää paluuarvoa. Sen voi ajatella vastaavan Javan `void`-avainsanaa. Merkkijonoja Scalassa kuvaa tietotyyppi `String`, joka on tyypialias Javan merkkijonoa kuvaavalle tietotyyppille. [4, Luku 5]

Scalan alkeistietotyypit muutetaan käänkösvaiheessa primitiivityypeiksi, esimerkiksi

Scalan `Int` käännetään 32-bittiseksi sanaksi, aivan kuten Javan `int`. Tämä takaa yhteensopivuuden Java-kirjastojen kanssa. Se parantaa myös suorituskkyä, sillä primitiiviarvolle ei tarvitse allokoida muistia ajon aikana. [4, Luku 6]

3.3 Kontrollirakenteet

3.3.1 Ehtolauseet

Ehtolauseita kirjoitetaan Scalassa *if/else*-lausekkeiden avulla. Lauseketta voi käyttää imperatiiviseen tyyliin eli jos ehto arvioituu todeksi, suoritetaan if-lohko, muuten else-lohko.

```
val x = true
if (x) {
  println("True")
} else {
  println("False")
}
// Tulostaa "True"
```

Scalan if-lausekkeella on myös arvo, eli if-lauseke palauttaa arvon. Arvon palauttavaa if-lauseketta voidaan käyttää myös funktionaalisessa ohjelmoinnissa. Esimerkiksi `val y = if (x >= 0) "positive" else "negative"` asettaa muuttujan `y` arvoksi joko merkkijonon "positive"tai "negative", riippuen muuttujan `x` arvosta. Huomattavaa on myös, että if-lauseke voidaan kirjoittaa yhdelle riville ilman aaltosulkeita. [6, Luku 2.1]

3.3.2 Silmukat

Silmukoita Scalassa on kolmenlaisia: **while**, **do** ja **for**. While-silmukka suorittaa sitä seuraavaa lohkoa 0-n kertaa, kunnes annettu ehto arvioituu epätodeksi. Do-silmukka toimii vastaavalla tavalla, mutta sitä suoritetaan 1-n kertaa.

```
val x = false
while (x)
  println("while")
```

```
do
  println("do")
  while (x)
```

Yllä oleva esimerkki ei suorita while-silmukkaa kertaakaan, ja do-silmukan kerran. [6, Luku 2.5]

Scalan **for**-lauseke on monipuolisempi kuin monissa muissa kielissä. Sitä voi käyttää perinteiseen tapaan silmukkana, jolloin for-lausekkeen *generaattori* antaa määriteltystä arvoista yksi kerrallaan muuttujan *i* kautta. Lauseketta voidaan käyttää myös usean generaattorin kanssa ja arvoja voidaan suodattaa lausekkeen sisällä. Kuten if-lauseke, myös for-lauseke voi palauttaa arvon. Seuraavaksi esimerkki kummastakin käyttötavasta.

```
for (i <- 1 to 10)
  println(i)
```

Silmukkana lauseketta voi käyttää yllä olevan esimerkin tapaan tulostamaan kaikki numerot yhdestä kymmeneen.

```
val res = for {
  x <- 1 to 10
  if x % 2 == 0
} yield (x * 2) // Vector(4, 8, 12, 16, 20)
```

Yllä olevassa esimerkissä luodaan kokonaisnumerovektori yhdestä kymmeneen, suodatetaan parilliset arvot, palautetaan jäljelle jääneet arvot kahdella kerrottuna ja asetetaan palautettu vektori muuttujaan *res*. Tällainen **for**-lauseke on yleinen funktionaalisessa Scalassa. [6, Luku 2.6]

3.4 Metodit ja funktiot

Jokainen operaatio Scalassa on metodikutsu. Myös tavanomaiset aritmeettiset operaattorit on toteutettu metodikutsuina. Esimerkiksi kahden kokonaisluvun yhteenlasku $1 + 2$ on lyhyempi versio metodikutsusta $1.(+)(2)$. Scalassa on mahdollista määritellä funktio kahdella tavalla: olion metodiksi tai funktioliteraaliksi. Metodi voidaan määritellä **def**-

avainsanalla. Scalassa funktio palauttaa automaattisesti viimeisen lausekkeensa arvon, joten **return**-avainsanan käyttö on vapaaehtoista ja sitä käytetään vain harvoin. [6, Luku 1.4] [5, Basics]

3.4.1 Funktioliteraalit

Funktioliteraali on funktio, jota käsitellään kuten mitä tahansa muuttujaa. Yleensä funktioliteraali sijoitetaan muuttujaan tai annetaan parametriksi toiselle funktiolle. Kuten muillakin muuttujilla, funktioilla ja metodeilla on Scalassa aina tyyppi. Tyypin voi määritellä eksplisiittisesti, mutta monesti Scala osaa päätellä funktion tyypin. Seuraavassa esimerkissä määritellään kummallakin tavoilla funktio, joka kertoo onko parametrina annettu kokonaisluku parillinen. Funktion tyyppi on siis `Int => Boolean`. [4, Luku 8]

```
def isEven1(x: Int): Boolean = x % 2 == 0
val isEven2: (Int => Boolean) = x => x % 2 == 0
```

3.4.2 Korkeamman asteen funktiot

Funktio voi ottaa parametriksi toisen funktion tai palauttaa funktion. Tällaisia funktioita kutsutaan *korkeamman asteen funktioiksi*. Korkeamman asteen funktiot ovat yksi funktionaalisen ohjelmoinnin kulmakivistä. Seuraavassa esitellään korkeamman asteen funktio, joka ottaa parametrina kokonaisluvun sekä funktion kokonaisluvusta kokonaislukuun, ja palauttaa kutsuu tätä funktiota parametrin kokonaisluvulle.

```
def mapValue(v: Int, m: Int => Int): Int = m(v)

def double(x: Int): Int = x * 2

mapValue(2, double)
mapValue(2, _ * 3)
```

Kuten yllä olevasta esimerkistä huomataan, annetaan `double`-funktio parametrinä ihan kuten mikä tahansa muukin muuttuja. Esimerkin viimeisellä rivillä käytetään funktioliteraalia `_ * 3`, joka on lyhyempi muoto funktioliteraalille `x => x * 3`. [6, Luku 12]

3.5 Luokat ja oliot

Scalassa on useita erilaisia luokka- ja oliotyyppisiä sekä tapoja luoda instansseja luokista. Tässä luvussa esitellään pääpiirteittäin erilaiset luokkatyypit.

3.5.1 Class

Luokka voidaan määritellä Scalassa **class**-avainsanalla. Luokka voi olla tyhjä tai se voi sisältää metodeita ja muuttujia. Luokasta luodaan instanssi **new**-avainsanalla. Luokan sisältämät kentät voidaan määritellä konstruktorissa heti luokan nimen jälkeen su-
luissa. Konstruktorissa määritellyt kentät ovat oletuksena **private val**. Luokasta saa-
daan uusi instanssi **new**-avainsanan avulla. Ihmistä kuvaava luokka ja instanssi voidaan
määritellä seuraavalla tavalla:

```
class Person(val name: String, var age: Int) {  
    def isAdult = age >= 18  
}  
val p = new Person("Matti", 30)
```

Luokassa Person on kaksi kenttää: julkinen muuttumaton kenttä nimi ja julkinen muutet-
tava ikä. Lisäksi luokassa on metodi, joka kertoo onko henkilö täysi-ikäinen. [5, Classes]

3.5.2 Case class

Scalassa on myös erityisiä case-luokkia, joita käytetään yleensä kuvaamaan muuttuma-
tonta dataa. Case-luokka määritellään kuten tavallinen luokka, mutta konstruktorissa mää-
riteltävät kentät ovat oletuksena **public val**. Case-luokkaan voidaan määritellä me-
todeita samoin kuin tavalliseenkin luokkaan. Case-luokan instanssia luotaessa ei tarvitse
käyttää **new**-avainsanaa. Puhelinmallia kuvaava case-luokka ja instanssi voidaan määri-
tellä seuraavalla tavalla:

```
case class Phone(brand: String, model: String)  
val p = Phone("Google", "Pixel4")
```

Scala-kääntäjä lisää case-luokkiin `toString`-, `hashCode`-, `equals`- ja `copy`-metodeita oletustoteutuksilla. Lisäksi kääntäjä lisää *kumppaniolioon* `apply`-tehdasmetodin, jonka avulla uusien instanssien luonti tapahtuu. [4, Luku 15]

Tavallisia luokkia käytetään varsinkin oliotyylisessä ohjelmoinnissa, kun taas case-luokat ovat käytössä erityisesti funktionaalisessa Scalassa, kun halutaan käyttää muuttumattomia olioita. Tavallisten luokkien ja case-luokkien yhtäsuuruusvertailut eroavat toisistaan merkittävästi. Tavallisia luokkia verrataan viittauksen mukaan, eli osoittaako viittaus samaan olioon JVM:ssä. Case-luokkien yhtäsuuruutta verrataan olion kenttien arvon mukaan. [4, Luku 15]

3.5.3 Singleton-oliot

Useat muut oliokielet tarjoavat mahdollisuuden kirjoittaa staattisia luokkia tai staattisia kenttiä luokkiin. Scalassa tällaista staattisuutta ei ole, vaan vastaava toiminnallisuus mahdollistetaan erityisien *singleton*-olioiden kautta. Singleton-oliosta on ajon aikana olemassa vain yksi instanssi, ja Scala luo sen automaattisesti. Singleton-olion on mahdollista periä luokka. Viittaaminen olioon tapahtuu yksinkertaisesti olion nimellä. [5, Singleton objects]

Singleton-oliota voidaan kutsua myös kumppaniolioksi, jos se on määritelty saman nimisen luokan yhteydessä. Kumppanioliot näkevät ilmentymiensä yksityiset kentät, ja ilmentymät näkevät kumppaniolionsa yksityiset kentät. Luvun 3.5.2 Phone-luokalle voidaan luoda kumppaniolio seuraavalla tavalla:

```
object Phone {  
    def iphone11 = new Phone("Apple", "iPhone_11")  
}  
val iphone = Phone.iphone11
```

Tässä tapauksessa kumppanioliossa on vain yksi metodi, jonka avulla voidaan luoda helposti instansseja. [4, Luku 4]

3.5.4 Piirrettyypit

Piirrettyyppejä käytetään Scalassa määrittelemään olioiden tarjoamien palveluiden rajapintoja ja mahdollistamaan ohjelmakoodin uudelleenkäytettävyyttä. Uusi piirrettyyppi luodaan **trait**-avainsanalla. Piirrettyyppien on myös sallittua periä toisia piirrettyyppejä tai luokkia. Myös singleton-olioihin on mahdollista liittää piirrettyyppejä. Piirrettyypin voi ajatella olevan Javan rajapintaluokan ja abstraktin luokan yhdistelmä. [4, Luku 6 ja 12]

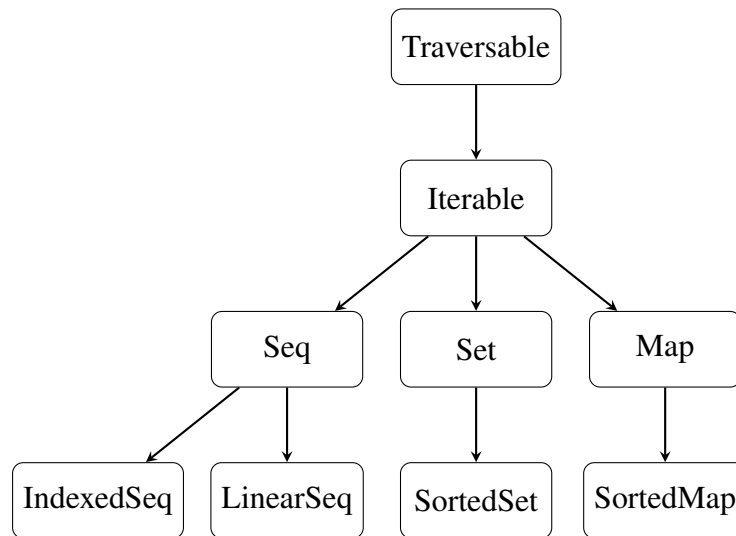
Piirrettyypin jäseniksi voidaan määritellä metodeita tai muuttujia. Jäsenet voivat olla abstrakteja tai konkreetteja. Abstraktille jäsenelle tulee toteuttavan luokan tarjota toteutus, kun taas konkreettelle jäsenille ei. Yhdessä piirreluokassa voi olla sekä abstrakteja että konkreetteja jäseniä. Tietyissä erityistapauksissa metodi voidaan merkitä abstraktiksi, vaikka sille on annettu toteutus piirrettyypissä. [6, Luku 10]

4 Kokoelmat

Tutkimuksen kohteena on Scalan standardikirjaston kokoelmat versiosta 2.8 versioon 2.12. Standardikirjastossa on toteutus useimmille yleisimmille kokoelmaluokille kuten taulukoille, listoille, joukoille(`Set`) ja assosiaatiotauluille(`Map`). Scalan kokoelmaluokat ovat paketissa `scala.collection`. Muuttumattomat kokoelmat ovat alipaketissa `immutable` ja muuttuvat alipaketissa `mutable`. [7]

Muuttumattoman kokoelman sisältämät alkiot eivät voi muuttua alustamisen jälkeen. Tämä tarkoittaa, ettei alkioden lisääminen, poistaminen tai uudelleenjärjestäminen ole mahdollista. Kokoelman muuttamista muistuttavat operaatiot, kuten `map`, `reverse`, `fold` ja kokoelmien yhdistäminen, palauttavat aina uuden kokoelman jättäen alkuperäisen kokoelman ennalleen. Käytännössä kuitenkin aina ei kaikkia muuttumattoman kokoelman alkioita ei tarvitse kopioida, vaan tietorakenteet pyrkivät käyttämään hyväkseen rakenteellista jakamista parantaakseen suorituskkyä ja optimoidakseen muistinkäyttöä. Muuttuvissa kokoelmissa on nimensä mukaisesti mahdollista vaihtaa alkioden järjestystä, lisätä ja poistaa alkioita kokoelmasta luomatta uutta kokoelmaa. [7] [4, Luku 22]

Kokoelmaluokat noudattavat piirreluokkien määrittelemää hierarkiaa, joka on paketissa `scala.collection`. Hierarkia on sama sekä muuttumattomille että muuttuville kokoelmille. Ylimpänä hierarkiassa on `Traversable`-piirreluokka, jolla on yksi välitön aliluokka, `Iterable`. `Iterable`:lla on kolme välitöntä aliluokkaa `Seq`, `Set` ja `Map`, jotka edustavat järjestyksellisiä kokoelmia, joukkoja ja assosiaatiotauluja. Kuvassa 4.1 näytetään lisäksi vielä muutama aliluokka. [7]



Kuva 4.1: Kokoelmien hierarkia

4.1 Järjestykselliset kokoelmat

TODO: Etsi parempi ilmaus järjestykselliselle Suorituskykyvertailuun valittiin muuttumattomista kokoelmista `List`, joka tyypillinen rekursiivinen linkitetty rakenne funktionaaliessa ohjelmoinnissa, ja `Vector`, jossa alkion haku indeksin mukaan on tehokkaampaa. Muuttuvista kokoelmista valittiin `Array`, joka on kiinteän kokoinen taulukko, `ArrayBuffer`, joka on dynaamisesti kasvava taulukko ja `ListBuffer`, joka on dynaamisesti kasvava linkitetty rakenne. Kaikki kyseiset muuttuvat tietorakenteet ovat tyypillisiä olio-ohjelmoinnissa käytettyjä tietorakenteita.

`List` on abstrakti luokka, joka kuvaa yhteen suuntaan linkitettyä rekursiivista muuttumatonta listaa, jolla on kaksi konkreettista toteutusta: `::` ja `Nil`. Epätyhjää listaa kuvaavassa luokassa `::` on kaksi jäsentä: alkio, jota kutsutaan listan *pääksi*, ja *häntä* joka kuvaa listan muita alkioita. Tyhjää listaa ja samalla listan loppumista kuvaa singleton-olio `Nil`. Esimerkiksi luvut 1 ja 2 sisältävä lista voidaan ilmaista näin: `1 :: 2 :: Nil`. Listan `head`-ja `tail`-operaatiot sekä listan alkuun lisääminen tapahtuvat vakioajassa. Aikakompleksisuus alkion hakemiselle indeksin perusteella on lineaarinen. [8] [7]

`Vector` on taulukon ominaisuuksia jäljittelevä muuttumaton tietorakenne, joka on

toteutettu puurakenteena, jonka solmut ovat korkeintaan 32-alkioisia taulukoita. Lehtisolmujen taulukot sisältävät vektorin varsinaisia alkioita, ja muut solmut sisältävät alemman tason taulukoita. Jos esimerkiksi alustetaan uusi 70-alkioinen vektori, on se kolmen taulukon puu, jossa juurisolmuna olevalla taulukolla on kolme lapsitaulukkoa, joista kahdessa on 32 alkioita ja yhdessä 6 alkioita. Tämä mahdollistaa taulukoiden rakenteellisen jakamisen eri vektori-instanssien kesken, mikä vähentää kopioimista ja muistinkäyttöä. Satunnaisten haku-, muokkaus- ja poisto-operaatioiden aikakompleksisuus on $\log(32, N)$, eli operaatioiden voidaan ajatella tapahtuvan lähes vakioajassa. [7] [9, Luku 4]

`Array` on Scalassa erityinen kokoelma, joka mahdollistaa primitiivityyppisten, viit-taustyyppisten ja geneeristen alkioden tallentamisen. Primitiivityyppisiä alkioita käsitel-lään ilman alkioden käärimistä olioiksi. Esimerkiksi kokonaislukuja sisältävä

`Array[Int]` kääntyy Javan `int[]`-taulukoksi. Kuten Javassa, taulukon koko ei voi muuttua alustamisen jälkeen. Verrattuna Javan taulukoihin `Array` tarjoaa kuitenkin huo-mattavasti enemmän operaatioita. Haku- ja muokkausoperaatiot tapahtuvat vakioajassa. Häntäoperaatio vie kuitenkin lineaarisen ajan, sillä jokainen alkio, poislukien ensimmäi-nen, joudutaan kopioimaan uuteen taulukkoon. [7]

`ArrayBuffer` on dynaamisesti kasvava listarakenne, joka käyttää alkioden tallen-tamiseen taulukkoa. Tästä johtuen useimmat operaatiot vastaavat aikakompleksisuudel-taan `Array`:ta. Lisäksi `ArrayBuffer` mahdollistaa alkion lisäämisen taulukon alkuun, loppuun sekä keskelle. Alkuun ja keskelle lisäämisen aikakompleksisuus on lineaarinen. Muut operaatiot tapahtuvat pääasiassa vakioajassa, poislukien täyteen taulukkoon lisää-minen, joka vie lineaarisen ajan. [7]

`ListBuffer` on dynaamisesti kasvava muuttuva kokoelma joka on toteutettu linki-tettynä rakenteena. Poiketen `List`:stä, `ListBuffer` mahdollistaa alkioden lisäämisen rakenteen alkuun ja loppuun vakioajassa. Satunnaisen alkion haku- ja muokkausoperaa-tiot sekä rakenteen keskelle lisääminen ovat aikakompleksisuudeltaan lineaarisia. Myös häntäoperaatio vie lineaarisen ajan, sillä kaikki hännän alkiot täytyy kopioida uuteen tie-

torakenteeseen. [7]

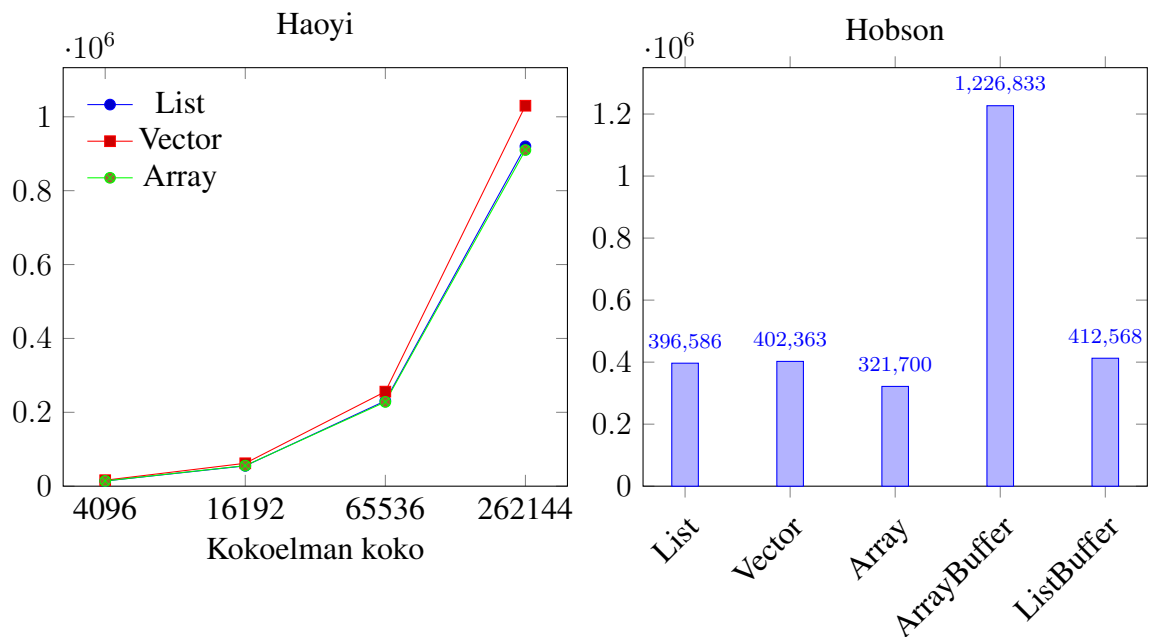
4.2 Suorituskykyvertailut

Tutkielmassa on käytetty kokoelmien suorituskykymittauksia kahdesta eri lähteestä: Li Haoyi:n[10] ja Toby Hobsonin[11] mittauksista. Haoyin mittauksissa mitataan aikaa, jonka operaatioiden suorittaminen vie, eli pienempi arvo tarkoittaa parempaa suorituskykyä. Hobsonin mittauksissa tarkastellaan operaatioiden määrää sekunnissa, jolloin suuri luku takooirtaa parempaa suorituskykyä. Hayoin mittauksissa suorituskykyä on mitattu eri kooisilla kokoelmilla. Kaavioihin on valittu kyseisen operaation kannalta mielekkäät koot kokoelmille.

Järjestyksellisten kokoelmien tapauksessa kiinnostuksen kohteena on yleisimpien operaatioiden suorituskyky: kokoelman luominen, iterointi, alkion satunnainen haku sekä lisääminen ja häntäoperaatio.

4.2.1 Iterointi

Iteroinnissa kokoelman jokainen alkio käydään läpi. Lähteenä käytettävissä suorituskykymittauksissa ei tutkittu rinnakkaisten kokoelmien suorituskykyä, joten iterointi tapahtui järjestyksessä. Haoyin mittauksissa `Array:n` iteroiminen **while**-silmukalla sekä `ArrayBuffer:n` iteroiminen tapahtui mittaamattoman lyhyessä tai jopa negatiivisessa ajassa, joten on kyseisiä mittauksia ei otettu huomioon. Kaaviosta 4.2 voidaan huomata, että suorituskyvyissä ei ollut suuria eroavaisuuksia, paitsi Hobsonin mittauksissa `ArrayBuffer` oli 3-4 kertaa muita kokoelmia nopeampi. Haoyin mittauksissa merkittävää eroa ei ollut, vaikkakin `Array` oli hieman muita nopeampi.



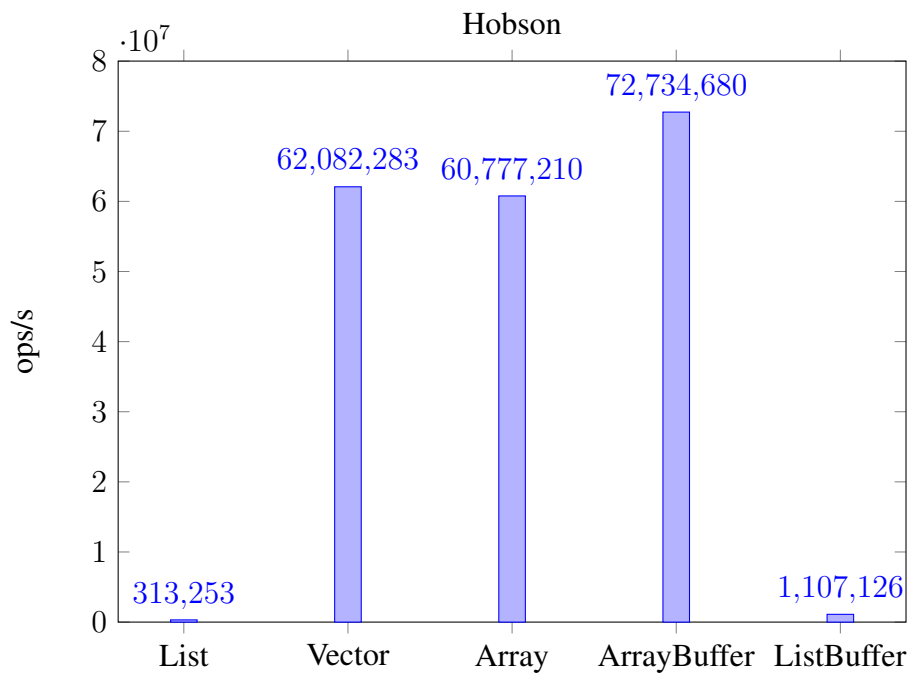
Kuva 4.2: Iterointi

4.2.2 Satunnainen haku

Nimensä mukaan satunnainen haku tarkoittaa alkion hakemista kokoelmasta satunnaisen indeksin perusteella. Linkitetyissä rakenteissa haku vaatii alkioden käymistä läpi peräkkäin kunnes haettu indeksi saavutetaan. Taulukkopohjaisissa tietorakenteissa vastaavaa rajoitetta ei ole, vaan alkion muistiosoite voidaan laskea suoraan indeksin perusteella. Tämän perusteella on perusteltua olettaa taulukkojen olevan huomattavasti linkitettyjä rakenteita suorituskykyisempiä.

Hobsonin mittaukset kaaviossa 4.3 tukevat tätä hypoteesia. `ListBuffer` sekä `List` ovat jopa useita satoja kertoja hitaampia kuin taulukoihin pohjautuvat kokoelmat. Hieman yllättäen `ListBuffer` on kuitenkin noin 3,5 kertaa nopeampi kuin `List`, vaikka `ListBuffer` käyttää sisäisesti alkioden tallentamiseen `List`:iä. Taulukon ominaisuuksia mukaileva `Vector` ja `Array` olivat suorituskyvyltään lähes identtisiä. `ArrayBuffer` oli noin 15% nopeampi kuin muut taulukkopohjaiset kokoelmat, mikä on hieman yllättä-

vää, sillä sisäisesti alkiot tallennetaan käyttäen `Array`:ta.



Kuva 4.3: Satunnainen haku

4.2.3 Kokoelman luominen

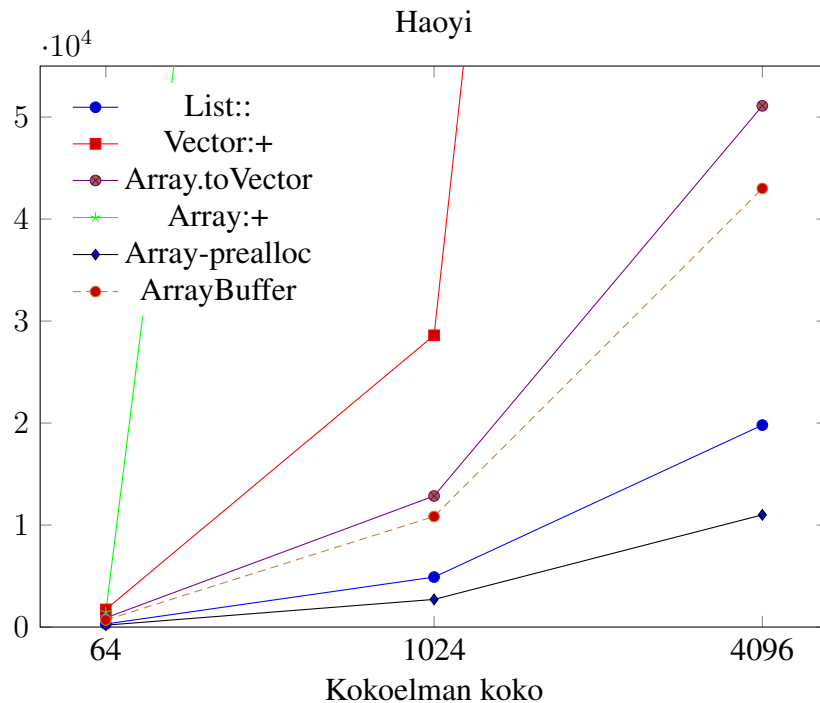
Kokoelmia rakennetaan useasti lisäämällä alkioita yksi kerrallaan kokoelmaan. Joskus kokoelma luodaan toisen kokoelman tai kokoelmatyyppin alkioista. Taulukkoa (`Array`) luotaessa täytyy yleensä päättää miten monen alkion taulukko luodaan. Mikäli täyteen taulukkoon halutaan lisätä alkioita, täytyy luoda uusi suurempi taulukko ja kopioida kaikki alkiot alkuperäisestä taulukosta uuteen taulukkoon ja lisätä uusi alkio uuteen taulukkoon. Taulukon kaikkien alkioden kopiointi uuteen taulukkoon todennäköisesti huonontaa suorituskyyä huomattavasti.

Hayoin vertailuissa (kaavio 4.4) on testattu useita tapoja kokoelmien luomiseen. `Array: +` kuvastaa tilannetta, jossa luodaan kokoelma lisäämällä alkio täyteen taulukkoon. Oletustusti sen suorituskyy on koko vertailun huonoin. `Array-prealloc` kuvaa tilannetta, jossa luodaan valmiiksi oikean kokoinen taulukko, ja sen suorituskyy on vertailun paras.

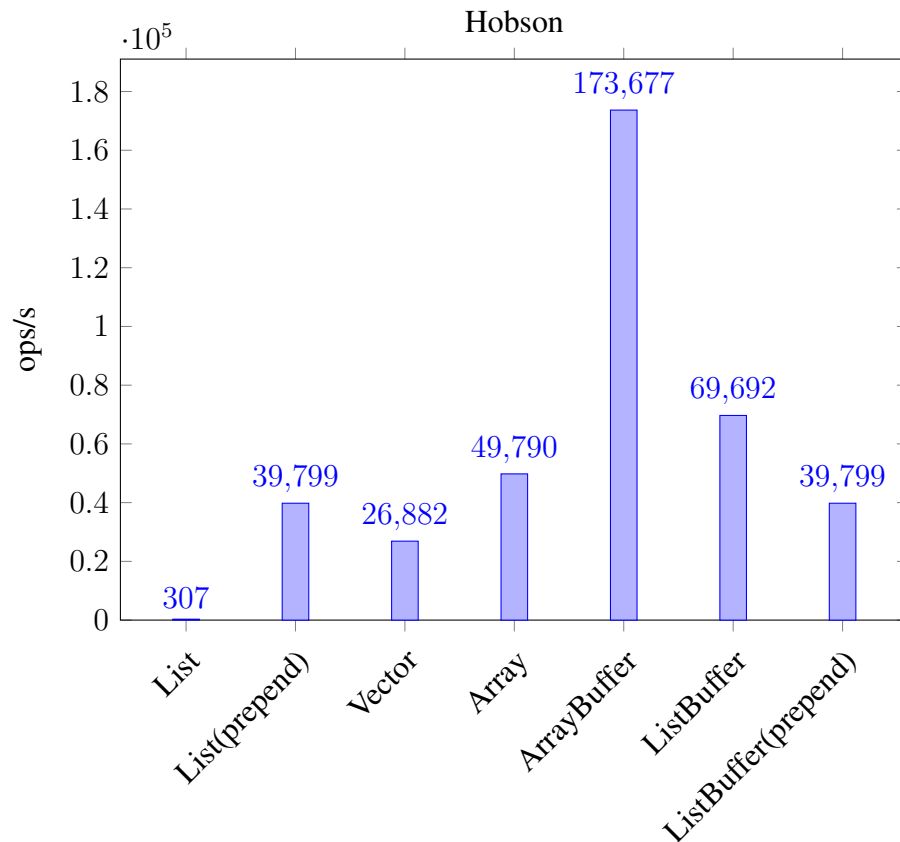
`List::` rakentaa linkitetyn rakenteen lisäämällä alkioita linkitetyn listan alkuun, ja sen on ilmoitettu tapahtuvan vakioajassa. [7] Hayoin mittaukset osoittavat tämän todeksi ja ainoastaan `Array-prealloc` oli `List::`aa suoritussykyisempi.

Kun `ArrayBuffer:n` sisäinen taulukko tulee täyteen, luodaan uusi, kaksi kertaa alkuperäistä suurempi, taulukko ja vanhan taulukon arvot kopioidaan uuteen. Suorituskyky on linkitettyä rakennetta huonompi, mutta hieman `Array.toVector`:ia parempi.

`Vector:+` aikakompleksisuuden pitäisi olla lähes vakio [7], mutta mittausten perusteella suorituskyky on varsin heikko. Erityisesti kun luodaan yli 1024 alkion kokoelmia ja sisäiseen puurakenteeseen tulee yksi taso lisää, suorituskyky heikkenee merkittävästi. Suorituskykyisempi tapa luoda `Vector` on luoda toinen kokoelma ja kutsua sen `toVector`-metodia. Haoyin mittausten `Array.toVector` on tilanne, jossa luodaan `Array` pre-alloc-tyylisesti ja luodaan sen pohjalta `Vector`. Suorituskyky on huomattavasti parempi kuin lisäämällä alkioita yksi kerrallaan.



Kuva 4.4: Kokoelman luominen



Kuva 4.5: Kokoelman loppuun lisääminen

4.2.4 Kokoelman loppuun lisääminen

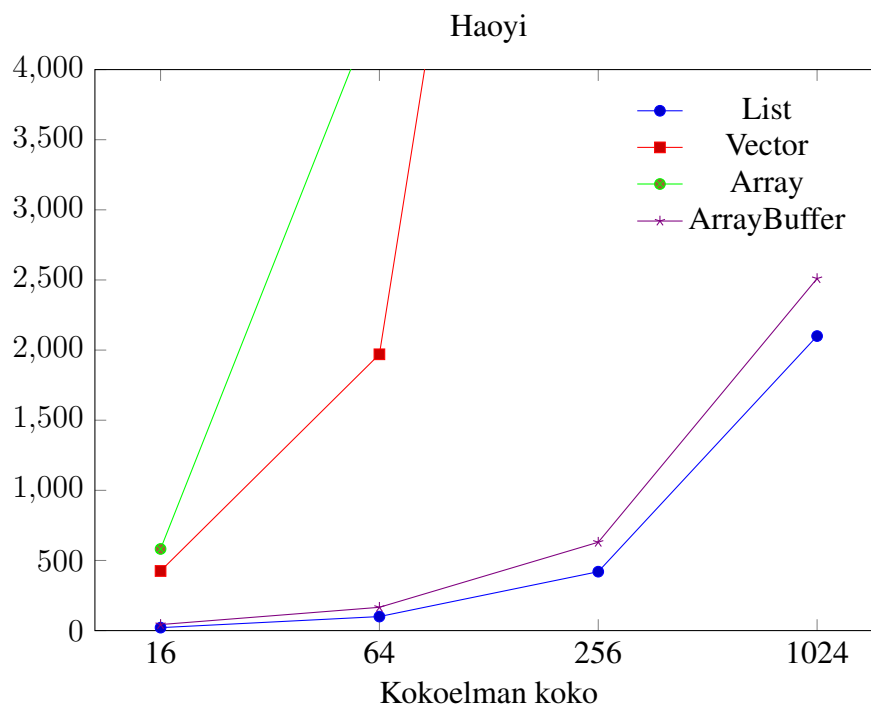
Asetelma Hobsonin mittauksessa kokoelman loppuun lisäämisestä on hieman samankaltainen kuin Haoyin kokoelman luomisen mittauksessa, mutta varsinkin `List` suoriutuu loppuun lisäämisestä huomattavasti heikommin. Loppuun lisättäessä, täytyy jokainen alkio kopioida uuteen kokoelmaan ja uusi alkio lisätä uuden kokoelman loppuun. `List-prepend` tarkoittaa tilannetta jossa alkio lisätään rakenteen alkuun, ja kokoelman järjestys käännetään. Kaaviosta 4.5 voidaan huomata, että tällä tavoin suoritussyky on jopa yli 100 kertaa parempi.

Muuttuvasta linkitetystä rakenteesta, `ListBuffer`:sta on mittaus kummallakin edellisessä kappaleessa mainitulla tavalla. Suoraan kokoelman loppuun lisäämisellä saavutetaan paras suoritussyky, ja se on jopa yli 70% parempi kuin lisäämällä alkuun ja kään-

tämällä kokoelman järjestys. Prepend-tyylillä suorituskky on identtinen `List`-prepend kanssa.

Taulukoista `ArrayBuffer` suoriutuu selvästi parhaiten, ollen yli kolme kertaa parempi kuin seuraavaksi paras taulukkorakenne `Array`. On syytä huomioida, että kyseisessä mittauksessa luodaan valmiiksi oikean kokoinen `Array`. `Vector` suoriutuu tästäkin mittauksesta yllättävän huonosti, ollen noin 50% huonompi kuin `Array` ja 30% huonompi kuin `List`-prepend.

4.2.5 Häntäoperaatio



Kuva 4.6: Häntäoperaatio

Haoyin häntäoperaation suorituskkyä mittaavassa tutkimuksessa aloitetaan n -kokoisesta kokoelmasta ja tehdään häntäoperaatiota jäljelle jäävälle kokoelmalle, kunnes jäljellä on tyhjä kokoelma. Mittauksen tulos on kulunut aika, jolloin pieni arvo tarkoittaa hyvää suorituskkyä. Kaaviossa 4.6 on mittauksen tulokset.

`List` suoriutuu oletetusti parhaiten. Muuttumattoman linkitetyn rakenteen häntään saadaan viittaus yksinkertaisesti palauttamalla ensimmäisen alkion seuraaja, ja se tapahtuu aina vakioajassa. `ArrayBuffer` suoriutui mittauksessa vain hieman `List`:iä huonommin.

Selvästi huonoiten suoriutui `Array`. Tulos on oletettu, sillä jokaisella häntäoperaatiolla kaikki paitsi taulukon ensimmäinen alkio täytyy kopioida uuteen taulukkoon. `Vector` on hieman `Array`:ta parempi, mutta sen suorituskky on 256:n alkion kokoelmalla jo noin 20 kertaa huonompi kuin kahdella suorituskkyisimmällä kokoelmalla.

5 Yhteenveto

Lähdeluettelo

- [1] M. Gabbrielli ja S. Martini, *Programming Languages: Principles and Paradigms*. Springer, 2010.
- [2] M. L. Scott, *Programming Language Pragmatics, 4th Edition*. Morgan Kaufmann, 2015.
- [3] P. Chiusano ja R. Bjarnason, *Functional Programming in Scala*. Manning, 2015.
- [4] M. Odersky, L. Spoon ja B. Venners, *Programming in Scala, 3rd Edition*. Artima, 2016.
- [5] (2019). Tour of Scala, url: <https://docs.scala-lang.org/tour/tour-of-scala.html> (viitattu 03.01.2020).
- [6] C. Horstmann, *Scala for the Impatient, 2nd Edition*. Addison-Wesley Professional, 2017.
- [7] (2017). Scala Collections v2.8-2.12, url: <https://docs.scala-lang.org/overviews/collections/> (viitattu 12.02.2020).
- [8] (2019). Scala API v2.12, url: <https://www.scala-lang.org/files/archive/api/2.12.x> (viitattu 03.01.2020).
- [9] V. Theron ja M. Diamant, *Scala High Performance Programming*. Packt, 2016.
- [10] L. Hayoi. (2016). Benchmarking Scala Collections, url: <http://www.lihaoyi.com/post/BenchmarkingScalaCollections.html> (viitattu 18.02.2020).

- [11] T. Hobson. (2016). Benchmarking Scala Collections, url: <https://www.tobyhobson.com/posts/scala/collections-performance/> (viitattu 18.02.2020).