

# Project Zeepbelbomen

Rien Maertens  
2<sup>de</sup> Bachelor Informatica

29 november 2015

## Deel I

# Theoretische Vragen

**Opgave 1.** *Geen top in een bladzeepbel heeft een kind in een andere zeepbel.*

*Bewijs.* Stel we hebben een bladzeepbel  $\alpha$  met een top  $T$  die een kind  $K$  heeft in zeepbel  $\beta$ . Neem  $n$  het aantal zeepbellen dat het pad van zeepbel  $\alpha$  naar de wortel bevat. We kunnen de volgende twee gevallen onderscheiden:

**Geval 1** Zeepbel  $\beta$  bevat een top zonder kinderen en is dus een bladzeepbel.

Het pad van de wortel van de boom naar  $\beta$  gaat door  $\alpha$ . Dit pad bevat dus  $n + 1$  zeepbellen. Dit is tegenstrijdig met het gegeven dat het pad van de wortel naar alle bladzeepbellen evenveel zeepbellen bevat. Het gestelde is dus vals. Het is onmogelijk voor een bladzeepbel om een kind te hebben in een andere zeepbel.

**Geval 2** Zeepbel  $\beta$  is geen bladzeepbel, maar bevat wel een pad naar een top zonder kinderen. Deze top is onderdeel van bladzeepbel  $\gamma$ . Het pad van  $\gamma$  naar de wortel van de zeepbelboom bevat  $n + k$  zeepbellen, met  $k$  het aantal zeepbellen in het pad tussen  $\beta$  en  $\gamma$ . Dit is opnieuw tegenstrijdigi met het gegeven dat het pad van de wortel naar alle bladzeepbellen hetzelfde aantal zeepbellen bevat. Ook in dit geval kan een bladzeepbel geen kinderen in andere zeepbellen hebben.

□

**Opgave 2.** *Wat is de maximale diepte van een  $k$ -zeepbelboom met  $n$  sleutels?*

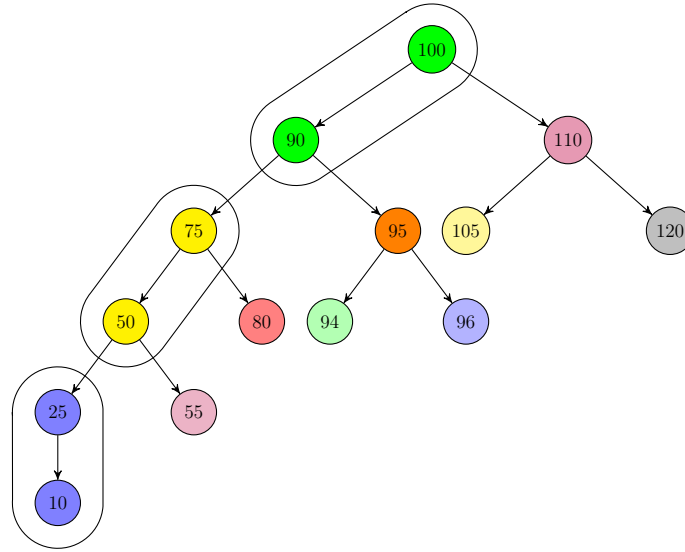
*a. Gemeten in zeepbellen.*

Stel  $d_z$  de diepte in zeepbellen van een zeepbelboom.  $d_z$  bereikt een maximum wanneer iedere zeepbel maar één top omvat, m.a.w. wanneer het pad van de wortel naar een blad evenveel zeepbellen als toppen bevat. De maximale diepte  $d_z$  in zeepbellen van een  $k$ -zeepbelboom met  $n$  sleutels is dus gelijk aan de maximale diepte in toppen van een complete binaire boom met  $n$  toppen. Dit is dus:

$$d_z \leq \lfloor \log_2 n \rfloor \tag{1}$$

b. *Gemeten in toppen van de ten gronde liggende binaire boom.*

Stel  $d_t$  de diepte in toppen van een zeepbelboom.  $d_t$  bereikt een maximum wanneer we één lang pad van zeepbellen maken waar elke top van een zeepbel maximum één kind in dezelfde zeepbel bevat. De andere zeepbellen die kinderen zijn van dit lange pad zijn terug zeepbellen met grootte één. Dit ziet er dan als volgt uit:



Figuur 1: Toppen met een zwarte lijn rond zitten in dezelfde zeepbel. Toppen zonder lijn rond zitten als enige top in de zeepbel.

Het valt op dat wanneer we een niveau zeepbellen toevoegen, we het dubbele aantal toppen moeten toevoegen van de vorige zeepbelboom met een niveau lager. Voor  $n$  het aantal toppen en  $z$  het aantal niveaus in zeepbellen geldt:

$$n \geq \sum_{i=1}^z k2^{z-i} = k(2^z - 1)$$

We lossen dit op naar  $z$ :

$$z \leq \log_2 \left( \frac{n+k}{k} \right)$$

De maximale diepte  $d_t$  van deze zeepbelboom in toppen is gelijk aan het aantal niveaus zeepbellen vermenigvuldigd met het maximale aantal toppen per zeepbel. Dus krijgen we dat  $d_t = k \cdot z$ . Wanneer voegen dit alles samen tot één formule:

$$d_t \leq k \cdot \log_2 \left( \frac{n+k}{k} \right) \quad (2)$$

Waar  $d_t$  gelijk is aan de maximale diepte in toppen.

**Opgave 3.** *Analyse van de verschillende gebalanceerde zeepbelbomen en hun complexiteit.*

a. *Voor toevoegen.*

Het toevoegen van een item aan een gebalanceerde zeepbelboom gebeurt als volgt:

Eerst wordt het item opgezocht in de zeepbelboom. Wanneer het item gevonden werd dan zit het item al in de zeepbelboom en moet er niets gebeuren. In dit geval heeft deze bewerking dezelfde tijds- en geheugencomplexiteit als het opzoeken, respectievelijk  $O(\log n)$  en  $\Theta(1)$  (dit bespreek ik in punt b).

Als de opzoeking uitkomt op `null` dan zit het toe te voegen item niet in de boom. In dit geval wordt er een nieuwe top aangemaakt en bevestigd aan de top die het laatst werd bekeken bij de zoekopdracht. De nieuwe top wordt ook toegevoegd aan de zeepbel waarin de ouder zit. Indien de zeepbel door deze toevoeging teveel toppen bevat wordt deze gesplitst. De manier waarop de zeepbel gesplitst wordt verschilt per implementatie van zeepbelboom:

- **Zeepbelboom1** zal zijn zeepbellen splitsen door eerst te kijken of beide kinderen van de wortel in dezelfde zeepbel zitten. Is dit niet het geval dan wordt een rotatie met de bovenste drie zeepbellen uitgevoerd zodat dit wel zo is. Vervolgens wordt de zeepbel gesplitst zodat elk kind van de oude wortel de wortel wordt van een nieuwe zeepbel. De oude wortel wordt daarna toegevoegd aan de zeepbel van zijn ouder, die eventueel opnieuw wordt gesplitst indien deze nu ook overvol zit. Als de oude wortel ook de wortel van de zeepelboom is dan wordt een nieuwe zeepbel aangemaakt.

Er wordt enkel binnen dezelfde zeepbel gekeken en omdat de grootte van een zeepbel constant is, gebeurt het splitsen van een zeepbel in constante tijd. In het slechtste geval wordt iedere zeepbel vanaf de toegevoegde top tot aan de wortel gesplitst. In opgave 2a (2) hebben we bewezen dat het aantal zeepbellen van de wortel naar een blad maximaal  $\lfloor \log_2 n \rfloor$  is.

We kunnen dus concluderen dat voor Zeepbelboom1 met  $n$  toppen de toevoegbewerking kost  $O(\log n)$  heeft.

- **Zeepbelboom2** splitst door eerst de zeepbelboom zo goed mogelijk te balanceren. Hiervoor worden eerst alle toppen in gesorteerde volgorde in een lijst gestoken en vervolgens opgebouwd tot een zo compleet mogelijke binaire boom. Daarna wordt net zoals Zeepbelboom1 de wortel toegevoegd aan de zeepbel van zijn ouder en vormen de kinderen van deze wortel de wortels van de nieuwe zeepbellen.

Net zoals Zeepbelboom1 wordt er enkel binnen dezelfde zeepbel gekeken, echter zal deze zeepbelboom er net iets langer over doen aangezien altijd alle toppen binnen de zeepbel wordt bekeken in plaats van enkel de bovenste drie. We besluiten dus dat de kost om een een top toe te voegen aan Zeepbelboom2  $O(\log n)$  is voor een boom van  $n$  toppen.

- **Zeepbelboom3** balanceert net zoals Zeepbelboom2 eerst de toppen in de te grote zeepbel, maar in plaats van enkel de wortel toe te voegen aan de bovenliggende zeepbel worden er meerdere toppen toegevoegd. Hoeveel toppen er toegevoegd worden kan worden aangepast. Als deze zeepbelboom werd gemaakt met de normale constructor dan zullen er zoveel mogelijk toppen worden toegevoegd aan de bovenliggende zeepbel tot deze zijn maximum capaciteit heeft bereikt (maar niet te vol zit). Met een tweede constructor kan er ingesteld worden dat er een maximum staat op het aantal omhoog te duwen toppen en kan dit maximum ook meegegeven worden. Er kan ook ingesteld worden dat er, indien mogelijk, net zoveel toppen aan de bovenliggende zeepbel worden toegevoegd tot deze moet splitsen.

Hier geldt hetzelfde als de vorige twee zeepbelbomen: er wordt enkel in dezelfde zeepbel gekeken, wat in constante tijd gebeurt omdat de grootte van de zeepbel constant is. Een splitsing kan een kettingreactie veroorzaken die bovenliggende zeepbellen ook doet

splitsen. Maar er zullen maximum  $O(\log n)$  van deze splitsingen gebeuren. De toevoegbewerking van Zeepbelboom3 zal dus ook kost  $O(\log n)$  hebben voor een boom waar reeds  $n$  elementen in zaten opgeslagen.

b. Voor opzoeken.

De drie gebalanceerde zeepbelbomen gebruiken alledrie dezelfde zoekmethode, dus de complexiteit van deze bewerking is dezelfde. Deze methode werkt zoals een normale binaire zoekboom:

De wortel wordt vergeleken met het gezochte item. Is het item in de wortel kleiner dan dalen we af naar het rechterkind van de wortel. Is het item in de wortel groter dan dalen we af naar het linkerkind van de wortel. Komt de waarde van de wortel overeen met dat van het gezochte item dan hebben we het item gevonden en kunnen we terugkeren uit de functie. We blijven op deze manier zoeken tot we het gezochte item gevonden hebben of tot we niet meer verder kunnen afdalen in de zoekboom (als we null tegenkomen). In dit geval zit het gezochte item niet in de zeepbelboom.

In het slechtste geval zit het item dat wordt gezocht niet in de zeepbelboom en moeten we helemaal tot een blad van de boom afdalen. In (2) heb ik de maximale diepte in toppen berekend. In die formule is  $k$  is een constante dus valt die in asymptotische notatie weg. We kunnen dus besluiten dat het opzoeken van een element in een gebalanceerde zeepbelboom die  $n$  items bevat een kost van  $O(\log n)$  heeft.

**Opgave 4.** *Bepaal voor semi-splay zeepbelbomen de slechtste-geval complexiteit en de geamortiseerde complexiteit.*

We introduceren eerst een paar stellingen die we zullen gebruiken in de antwoorden van deze opgave:

**Stelling 1.** *Het langste pad van de wortel naar een blad in een semi-splay zeepbelboom is naar boven begrenst door  $O(\log(n))$*

*Bewijs.* Het slechtste geval doet zich voor wanneer alle zeepbellen op één lijn staan. De lengte van het pad van de wortel naar de top die zich het verst van de wortel bevindt is dan gelijk aan het aantal zeepbellen vermenigvuldigd met de maximale diepte van een zeepbel.

Het aantal zeepbellen  $z$  in een zeepbelboom met  $n$  toppen en maximaal  $k$  toppen per zeepbel is gelijk aan  $\lfloor n/k \rfloor$ . Aangezien een complete zeepbel altijd gebalanceerd is heeft deze diepte  $\log_2(k)$ . Het langst mogelijke pad van de wortel naar een blad is  $\lfloor n/k \rfloor \cdot \log_2(k) = O(n)$  toppen lang. In asymptotische notatie wordt dit  $O(\log(n))$ . Dit is precies wat we zochten.  $\square$

**Stelling 2.** *Een semi-splay bewerking gebeurt in constante tijd.*

*Bewijs.* Wanneer een splaybewerking wordt uitgevoerd worden alle toppen uit de drie geselecteerde zeepbellen in inorde overlopen en in een lijst gestopt. Bij het inorde itereren worden er  $3 \cdot k$  toppen overlopen, met  $k$  de maximale grootte van een zeepbel. Vervolgens wordt deze lijst opgedeeld in drie lijsten van lengte  $k$  en worden van deze lijsten drie complete gebalanceerde zeepbellen gemaakt. Deze drie zeepbellen worden daarna aan elkaar bevestigd en terug in de zeepbelboom geplaatst.

Omdat er altijd binnen de gegeven drie zeepbelle wordt gekeken zal een semisplay bewerking altijd  $\Theta(k)$  operaties op toppen nodig hebben. Maar omdat de  $k$  in een zeepbelboom constant is kunnen we zeggen dat een splaybewerking in constante tijd of  $\Theta(1)$  tijd verloopt.  $\square$

## Slechtste geval

- **Voor toevoegen**

Het slechtste geval doet zich voor wanneer het toe te voegen item moet toegevoegd worden als kind van de top die zich het verst van de wortel bevindt. Wanneer dit geval is moeten alle toppen over dit pad overlopen worden om de plaats van de toe te voegen top te vinden. Nadat de plaats werd gevonden wordt het nieuwe item toegevoegd, wat een bewerking van constante tijd is. Als de zeepbel waaraan de nieuwe top is toegevoegd vol zit moet deze gebalanceerd worden. Dit gebeurt ook in constante tijd. Vervolgens wordt er vanuit deze zeepbel tot aan de wortel gesplayed.

Uit stelling 1 volgt dat dit langste pad naar boven begrensd wordt door  $O(n)$  en uit stelling 2 volgt dat een splaybewerking in constante tijd gebeurt. We kunnen dus besluiten dat een toevoegbewerking op een boom waar reeds  $n$  toppen in zitten in het slechtste geval  $O(n)$  kost.

- **Voor opzoeken**

Een opzoekbewerking is erg gelijkend op een toevoegbewerking. Het enige verschil is dat bij een zoekbewerking geen top wordt toegevoegd. In het slechtste geval verloopt een opzoekbewerking op een boom van  $n$  toppen in  $O(n)$  tijd.

## Geamortiseerde complexiteit

**Stelling 3.** *Een reeks van  $n > 1$  bewerkingen op een initieel lege semi-splay boom heeft een geamortiseerde complexiteit van  $O(n \log n)$ , dus is de gemiddelde kost  $O(\log n)$  per bewerking.*

*Bewijs.* Ik bewijs dit met behulp van de potentiaalmethode. Voor een lege boom definiëren wij  $\Phi(T) = 0$ . Stel dat  $T$  niet leeg is. Voor een zeepbel  $z$  in een semi-splay boom  $T$  met een willekeurige  $k$ -waarde definiëren we  $A_T(z)$  als het aantal zeepbellen in de deelboom bestaande uit  $z$  en al zijn nakomelingen in  $T$ . Bovendien defineer  $L_T(z) = \log(A_T(z))$  en  $Z(T)$  als het totale aantal zeepbellen in  $T$ . De potentiaal van een semi-splay zeepbelboom definiëren wij als:

$$\Phi(T) = \log(k) \cdot \sum_{z \in T} L_T(z) \quad (3)$$

Met  $k$  de maximale grootte van een zeepbel.

De potentiaal wordt gewijzigd in twee gevallen:

1. Als er een zeepbel wordt toegevoegd (nog zonder de boom te herbalanceren)
2. Als tijdens het semi-splayen als gevolg van een toevoeg- of opzoekbewerking er deelbomen vervangen worden

**Deelresultaat 1.** *Als een nieuw item aan de zeepbelboom  $T$  wordt toegevoegd en het resultaat is de boom  $T'$  dan geldt:*

$$\Phi(T') - \Phi(T) \leq \log(k) \cdot \log(|T'|)$$

Ik heb het hier enkel over het toevoegen zelf en niet de splaybewerking die nog moet volgen.

Als  $T$  leeg was dan klopt dit duidelijk. Stel nu dat  $T$  niet leeg is.



Stel dat de zeepbellen op het pad van de wortel naar de nieuwe top  $z_1, \dots, z_k$  zijn, waarbij  $z_k$  de zeepbel is die de nieuwe top bevat. We hebben hier twee gevallen:

- De nieuwe top werd toegevoegd aan een nieuwe zeepbel:

$$\begin{aligned}\Phi(T') &= \Phi(T) - \log(k) \cdot \left( \sum_{i=1}^{k-1} L_T(z_i) + \sum_{i=1}^k L_{T'}(z_i) \right) \\ &= \Phi(T) + \log(k) \cdot L_{T'}(z_1) + \log(k) \cdot \left( \sum_{i=1}^{k-1} (L_{T'}(z_{i+1}) - L_T(z_i)) \right)\end{aligned}$$

Voor  $i \geq 1$  geldt  $A_T(z_i) \geq A_T(z_{i+1}) + 1$  en  $A_T(z_{i+1}) + 1 = A_{T'}(z_{i+1})$ . Dus  $L_T(z_i) \geq L_{T'}(z_{i+1})$ . Hieruit volgt dat  $L_{T'}(z_{i+1}) - L_T(z_i) \leq 0$ . Uiteindelijk bekomen we:

$$\begin{aligned}\Phi(T') &\leq \Phi(T) + \log(k) \cdot L_{T'}(z_1) \\ &\leq \Phi(T) + \log(k) \cdot \log(Z(T'))\end{aligned}$$

Want de wortelzeepbel heeft namelijk evenveel opvolgers (zichzelf meegeteld) als er zeepbellen in de boom zitten. Een zeepbelboom bevat ook minstens evenveel toppen als zeepbellen dus  $B(T) \leq |T|$  en ook  $\log(B(T)) \leq \log(|T|)$ . We concluderen:

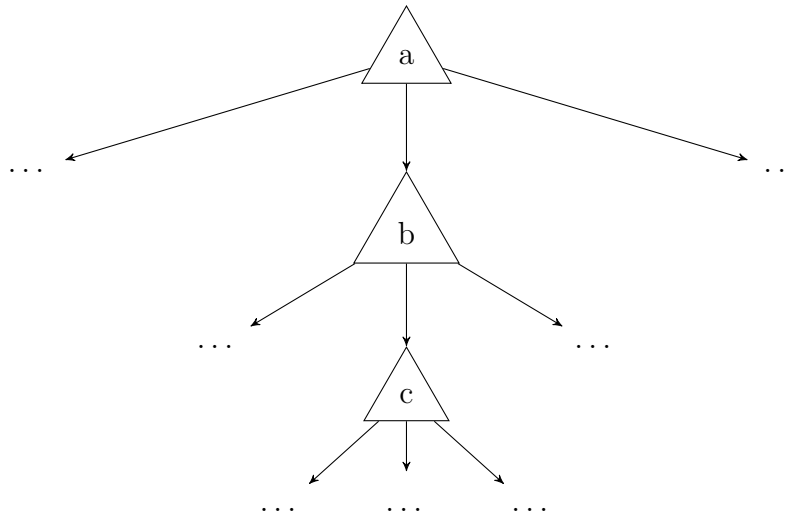
$$\begin{aligned}\Phi(T') &\leq \Phi(T) + \log(k) \cdot \log(|T'|) \\ \Phi(T') - \Phi(T) &\leq \log(k) \cdot \log(|T'|)\end{aligned}$$

- De nieuwe top werd toegevoegd aan een reeds bestaande zeepbel:

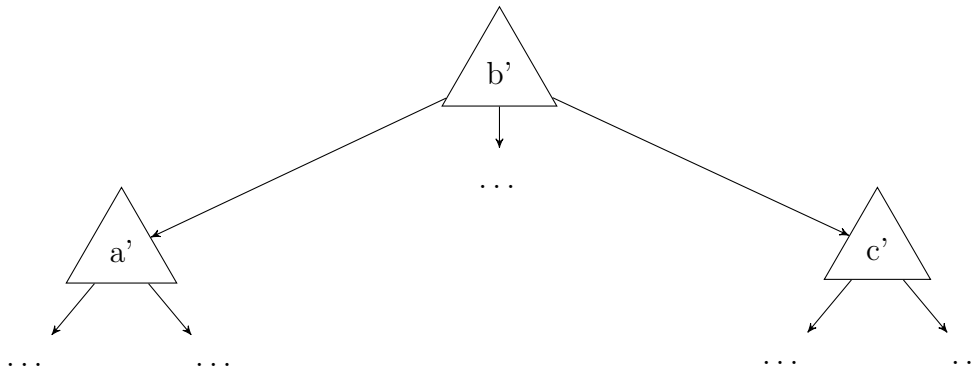
In dit geval is  $\Phi(T) = \Phi(T')$  en dus uiteraard  $\Phi(T) - \Phi(T') = 0 \leq \log(k) \cdot \log(|T'|)$ . Want op zeepbelniveau veranderd er namelijk niets.

**Deelresultaat 2.** *Stel dat een deelboom in een boom  $T$  op een semi-splay manier wordt vervangen en het resultaat is  $T'$ . We noemen de laagste zeepbel in de deelboom van  $T$  die wordt vervangen  $z_l$  en de nieuwe wortel van de nieuwe deelboom van  $T'$  wordt  $z_w$ . Dan geldt:*

$$\Phi(T') - \Phi(T) \leq \log(k) \cdot (2L_{T'}(z_w) - 2L_T(z_l) - 2)$$



Figuur 2: Positie van de zeepbellen voor de splay. De eventuele aanwezigheid van (meerdere) kindzeepbellen wordt aangeduid met “...”



Figuur 3: Positie van de zeepbellen na de splay.

Bij het semi-splayen beschouwen we drie zeepbellen: de zeepbel waarop de splay actie wordt uitgevoerd, zijn ouder en zijn grootouder. We noemen deze respectievelijk  $a$ ,  $b$  en  $c$ . Na het semi-splayen krijgen we drie nieuwe zeepbellen. Deze komen echter niet meer overeen de

zeepbellen van voor de splay want de interne toppen worden tussen deze drie zeepbellen uitgewiseld. We noemen deze drie nieuwe zeepbellen  $a'$ ,  $b'$  en  $c'$ .

We zien op figuur 2 en figuur 3 dat  $c = z_l$  en  $b' = z_w$ . Op zeepbelniveau is een zeepbelboom ook geen binaire boom: een complete zeepbel van grote  $k$  heeft  $k + 1$  kindzeepbellen. Mijn semi-splay algoritme zal wel  $a'$  en  $c'$  bevestigen als meest linkse en meest rechtste kindzeepbel van  $b'$ . Wij hebben dus:

$$\Phi(T') = \Phi(T) + \log(k) \cdot ((L_{T'}(a') + L_{T'}(b') + L_{T'}(c')) - (L_T(a) + L_T(b) + L_T(c)))$$

We zien onmiddellijk dat  $A_{T'}(b') = A_T(a)$  en dus ook  $L_{T'}(b') = L_T(a)$ .

We krijgen vervolgens:

$$\Phi(T') = \Phi(T) + \log(k) \cdot ((L_{T'}(a') + L_{T'}(c')) - (L_T(b) + L_T(c)))$$

Omdat  $c$  een kind is van  $b$  geldt  $A_T(c) < A_T(b)$  en  $L_T(c) < L_T(b)$ .

Dus:

$$\Phi(T') \leq \Phi(T) + \log(k) \cdot ((L_{T'}(a') + L_{T'}(c')) - 2L_T(c))$$

Omdat  $a'$  en  $c'$  kinderen zijn van  $b'$  kunnen we ook afleiden dat  $A_{T'}(a') + A_{T'}(c') \leq A_{T'}(b')$ . Uit de rekenkunde<sup>1</sup> volgt dan dat  $\log(A_{T'}(a')) + \log(A_{T'}(c')) \leq 2\log(A_{T'}(b')) - 2$  dus  $L_{T'}(a') + L_{T'}(c') \leq 2L_{T'}(b') - 2$ .

Waaruit volgt dat:

$$\begin{aligned} \Phi(T') &\leq \Phi(T) + \log(k) \cdot (2L_{T'}(b') - 2L_T(c) - 2) \\ &\leq \Phi(T) + \log(k) \cdot (2L_{T'}(z_w) - 2L_T(z_l) - 2) \\ \Phi(T') - \Phi(T) &\leq \log(k) \cdot (2L_{T'}(z_w) - 2L_T(z_l) - 2) \end{aligned}$$

en dit is precies wat we zochten.

**Deelresultaat 3.** *Als tijdens een toevoegbewerking of een opzoekbewerking in een boom  $T$  langs een pad met  $2s+1$  zeepbellen deelvomen vervangen worden en het resultaat is  $T'$  dan geldt:*

$$\Phi(T') - \Phi(T) \leq \log(k) \cdot (2\log|T'| - 2s)$$

---

<sup>1</sup>Cursusnota's AD2 pagina 49 opmerking 9.

We vervangen  $s$  keer een deelboom. Stel dat de boom na het  $i$ -de vervangen de boom  $T_i$  is, dus  $T_0 = T$  en  $T_s = T'$ . Bovendien zij  $z_{l,i}$  de laagste top van de deelboom die in stap  $i$  word vervangen en  $z_{w,i}$  de wortel van de nieuwe deelboom die in stap  $i$  ontstaat. Dan is

$$\begin{aligned}\Phi(T') &= \Phi(T) + \log(k) \cdot \sum_{i=1}^s (\Phi(T_i) - \Phi(T_{i-1})) \\ &\leq \Phi(T) + \log(k) \cdot \sum_{i=1}^s (2L_{T_i}(z_{w,i}) - 2L_{T_{i-1}}(z_{l,i}) - 2)\end{aligned}$$

De zeepbel die de nieuwe wortel van een net geplaatste deelboom is, is de laagste zeepbel in de volgende stap. Dus  $z_{w,i} = z_{l,i+1}$ . Substitueren we dit dan krijgen we:

$$\begin{aligned}\Phi(T') &\leq \Phi(T) + \log(k) \cdot (2L_{T_s}(z_{w,s}) + \sum_{i=1}^{s-1} (2L_{T_i}(z_{w,i}) - 2L_{T_i}(z_{l,i+1})) - 2s) \\ &\leq \Phi(T) + \log(k) \cdot (2L_{T_s}(z_{w,s}) - 2s) \\ &\leq \Phi(T) + \log(k) \cdot (2\log(|T'|) - 2s)\end{aligned}$$

We hebben nu voor alle deelbewerkingen die de potentiaal wijzigen een bovengrens aangetoond voor het verschil van de potentiaal vóór en na de deelbewerking. Als een top wordt toegevoegd wordt de potentiaal eventueel gewijzigd door het toevoegen van een zeepbel en door het splayen. Bij het opzoeken wordt de potentiaal enkel gewijzigd door het splayen. We kunnen nu onze tabel opstellen met  $T'$  de boom na de bewerking en  $T$  de boom vóór de bewerking. De toevoegbewerking splitsen we nog eens op: we noteren  $\tilde{T}$  als de boom die we verkrijgen na de toevoeging nog vóór de splaybewerking.

Als het aantal zeepbellen van het pad dat wordt doorlopen  $l$  is, dan worden ofwel langs een pad met  $2s + 1 = l$  zeepbellen ofwel langs een pad met  $2s + 1 = l - 1$  zeepbellen deelbomen vervangen. Hierbij is  $2s$  zoals in deelresultaat 3. Er geldt in elk geval  $l \leq 2s + 2$ .

We kunnen de echte kost echter niet beschrijven als het aantal bezochte zeepbellen. We weten wel dat  $t \leq l \cdot \log(k)$ . Hier is  $l$  het aantal zeepbellen op een pad,  $k$  de maximale grootte van een zeepbel en  $t$  het aantal bezochte toppen. (Het is nu ook duidelijk waarom  $\log(k)$  in de potentiaal zit)

	echte kost	gewijzigde kost
top opzoeken	aantal bezochte toppen: $t \leq l \cdot \log(k)$	$= t + \Phi(T') - \Phi(T)$ $= l \cdot \log(k) + \Phi(T') - \Phi(T)$ Deelresultaat 3: $\leq l \cdot \log(k) + \log(k) \cdot (2 \log  T'  - 2s)$ Want $l \leq 2s + 2$ : $\leq 2 \log(k) \log( T' ) + 2$
top toevoegen	aantal bezochte toppen: $t \leq l \cdot \log(k)$	$= t + \Phi(T') - \Phi(T)$ $= t + \Phi(T') - \Phi(\tilde{T}) + \Phi(\tilde{T}) - \Phi(T)$ Deelresultaat 1: $\leq t + \Phi(T') - \Phi(\tilde{T}) + \log(k) \cdot \log( \tilde{T} )$ Deelresultaat 3: $\leq t + \log(k) \cdot (2 \log  T'  - 2s) + \log(k) \cdot \log( \tilde{T} )$ Want $t \leq l \cdot \log(k)$ en $l \leq 2s + 2$ : $\leq 2 \cdot \log(k) \log( T' ) + 2 + \log(k) \cdot \log( \tilde{T} )$ Want $ T  =  T' $ : $\leq 3 \log(k) \log( T' ) + 2$

Omdat voor een initieel lege boom na  $n$  stappen altijd geldt dat  $|T| \leq n$ , volgt hieruit het gestelde. De gewijzigde kost per bewerking is namelijk altijd  $O(\log n)$  en de potentiaal is op het einde altijd ten minste even groot als in het begin.

We hebben zojuist dus bewezen dat een reeks van  $n > 1$  bewerkingen op een initieel lege semi-splay zeepbelboom een geamortiseerde tijdscomplexiteit van  $O(n \log n)$  heeft. De gemiddelde kost per bewerking is dus  $O(\log n)$ .

□

## Deel II

# Implementatie

Hieronder volgt er een korte uitleg over de interne structuur en algoritmes die ik heb gebruikt voor de implementaties van de zeepbelbomen.

## 1 Binaire boom

### 1.1 Node

Mijn binaire boom bestaat intern uit toppen die hun ouder, linker en rechterkind bijhouden. Om problemen met bidirectionele associaties te vermijden kan de verwijzing naar de ouder van een top enkel ingesteld worden wanneer deze top als kind wordt aangeduid van een andere top met de methodes `setRightChild()` en `setLeftChild()`. Omdat de wortel van de binaire boom natuurlijk geen ouder heeft is er ook een methode die deze verwijzing op `null` zet. Een top bevat ook een booleaanse waarde `removed` die functioneert als grafsteen. Daarnaast bevat iedere top ook een verwijzing naar de zeepbel waartoe die top behoort. De volgende functies hebben wat extra uitleg nodig:

`traverseInorder()` traverseer in orde over de kinderen van deze top.

Als argumenten wordt er een `consumer` en een `predicate` meegegeven.

Zodra een top bij het traverseren van de stack komt wordt deze doorgegeven aan de `consumer`. De `predicate` bepaald hoe ver deze iteratie gaat (bijvoorbeeld: enkel in dezelfde zeepbel).

`traverseAndAdd()` maakt gebruik van bovenstaande methode om de kinderen van de top toe te voegen aan de collection `nodes` wanneer ze voldoen aan de `predicate`. Wanneer een top voldoet maar één van zijn kinderen niet meer, dan wordt dit kind toegevoegd aan de collection `children`.

### 1.2 Zeepbel

Om de toppen in groepjes in te delen bestaat de klasse `Zeepbel`. Deze houdt enkel zijn wortel bij en het aantal toppen dat deze bezit. Een zeepbel heeft ook de methodes `topAdded()` en `topsRemoved()` die de zeepbel waarschuwen

wanneer er toppen bijgekomen of weggevalen zijn. Als de methode `balanceBubble` wordt uitgevoerd maakt gebruik van de `TreeBuilder` om zijn toppen intern te balanceren.

### 1.3 Zeepbelboom

In de abstracte klasse `Zeepbelboom` zit de gemeenschappelijke functionaliteit die alle zeepbelboom-implementaties delen. Ik geef wat uitleg over de belangrijkste functies:

`find()` vind een item in de zeepbelboom. Meer uitleg kunt u terugvinden in de javadoc-header. Deze methode wordt door alle basisbewerkingen gebruikt.

`add()` voegt een item toe aan de zeepbelboom. Wanneer een item nog niet in de boom zat wordt deze door de abstracte methode `addToParent` toegevoegd aan de boom. Die methode is abstract `ShrinkingBubbleTree` en `Zeepbelboom4` anders omgaan met volle zeepbellen.

`contains()` geeft `true` terug wanneer een item in de zeepbelboom zit. Het resultaat is `false` wanneer het gezochte item zich niet in de boom bevindt of verwijderd is doormiddel van een grafsteen.

`iterator()` geeft een iterator terug over alle items in de zeepbelboom (die niet verwijderd zijn). Deze maakt gebruik van de `traverseInorder`-methode uit `Node`.

`zeepbelIterator()` iterator over de zeepbellen. Deze methode werkt zoals de `iterator()` methode, maar enkel de wortels van zeepbellen worden bekeken en vervolgens worden deze zeepbellen toegevoegd aan de iterator.

### 1.4 ShrinkingBubbleTree

De abstracte superklasse voor de gebalanceerde zeepbelbomen. Deze heeft de abstracte methode `shrinkBubble` die per implementatie verschilt. Deze methode wordt opgeroepen wanneer er zeepbel te vol zit door een toevoegbewerking en moet gesplitst worden. De verschillende implementaties maken gebruik van enkele hulpmethodes die in deze klasse gedefinieerd zijn de uitleg van

deze hulpmethodes kunt u in de javadoc-headers vinden. Verder zijn ook nog deze functies aanwezig:

**addToParent()** voegt een item toe aan een gegeven top en de zeepbel van deze top. Wanneer deze zeepbel overvol zit wordt de zeepbel gesplitst m.b.v. **shrinkBubble()**.

**remove()** Helaas heb ik een beetje moeten valsspelen en maak ik gebruik van grafstenen om een top te verwijderen uit de boom. Wanneer de boom voor de helft bestaat uit verwijderde toppen wordt de boom opnieuw opnieuw opgebouwd door **rebuildTree**. Omdat een top niet echt verwijderd wordt is het mogelijk dat er zeepbellen helemaal gevuld zijn, maar dat hun iterator leeg is.  
Ik heb geprobeerd een echte implementatie te maken, maar ik kwam vast te zitten bij het debuggen. Deze code kun je nog altijd terugvinden in commentaar onderaan in deze klasse.

**removeAll()** In plaats van achter iedere verwijdermethode te controleren of de boom opnieuw opgebouwd moet worden gaat deze bulkoperatie eerst de gevraagde elementen verwijderen en pas achteraf kijken of de boom moet opnieuw opgebouwd worden. Dit gebeurt omdat deze opbouwoperatie tamelijk duur is.

**rebuildTree()** bouwt de zeepbelboom opnieuw op door eerst alle toppen die niet verwijderd zijn in een lijst te stoppen. Ze worden op een *breadth-first* manier in deze lijst getoken zodat wanneer ze opnieuw toegevoegd worden de boom min of meer dezelfde balancerings heeft. Moest ze gesorteerd in een lijst zitten dan zou de boom zwaar naar rechts overhellen wanneer opnieuw in die volgorde zouden worden toegevoegd.

## 1.5 TreeBuilder

Hulpklasse die wordt gebruikt om van een gesorteerde lijst van toppen een zo goed mogelijk gebalanceerde binaire boom te maken. Volgende functies zijn belangrijk:

**listToTree()** methode die een gebalanceerde binaire boom maakt van een gesorteerde lijst van toppen. Dit gebeurt op een recursieve manier in  $\Theta(n)$  tijd (met  $n$  de lengte van de lijst).



`listToTreeIterative()` iteratieve versie van bovenstaande methode.

Deze maakt gebruik van een `while`-loop en een stack in plaats van een recursieve oproep. Uit testen bleek dat deze methode trager was dan de recursieve methode dus wordt deze functie niet gebruikt.

`attachChildren()` voegt een lijst van  $n + 1$  toppen toe als kinderen van de bladeren van de opgebouwde boom.

## 2 Implementaties gebalanceerde zeepbelbomen

Nu volgt een beschrijving van `splitBubble`-methode van de zeepbelbomen die `ShrinkingBubbleTree` implementeren.

### 2.1 Zeepbelboom1

Wanneer de wortel van de zeepbel twee kinderen heeft die in dezelfde zeepbel zitten wordt er niets veranderd aan de interne structuur. Is dit niet het geval dan wordt er een rotatie uitgevoerd op de bovenste drie toppen tot dit wel zo is. Vervolgens worden de twee kinderen van de wortel genomen als nieuwe wortels voor twee nieuwe zeepbellen. De oude wortel wordt daarna toegevoegd aan de bovenliggende zeepbel. Als deze top de wortel van de zeepbelboom was, en dus geen bovenliggende zeepbel had, dan wordt er een nieuwe zeepbel aangemaakt en ingesteld als wortelzeepbel. Wanneer deze top wel aan een bestaande zeepbel kon toegevoegd worden wordt deze indien nodig opnieuw gesplitst.

### 2.2 Zeepbelboom2

Voor de zeepbel gesplitst wordt wordt deze eerst gebalanceerd met `balanceBubble`. Daarna wordt op dezelfde manier gesplitst als `Zeepbelboom1`.

### 2.3 Zeepbelboom3

Deze zeepbel werkt op dezelfde manier als `Zeepbelboom2` echter worden er hier zoveel mogelijk toppen aan de bovenliggende zeepbel toegevoegd. Deze zeepbelboom kan geconfigureerd worden: er kan gekozen worden om een maximum te zetten op het aantal omhoog te duwen toppen en er kan

ingesteld worden dat er geprobeerd wordt om net genoeg toppen toe te voegen aan de bovenliggende zeepbel dat die verplicht wordt om te splitsen.

## 3 Semi-splay zeepbelboom

### 3.1 Zeepbelboom4

Deze zeepbelboom is een implementatie van de semi-splay zeepbelboom. Bij elke toevoeg- en opzoekbewerking wordt er gesplayed op zeepbelniveau vanaf de zeepbel van de top die het werd toegevoegd op opgezocht tot aan de wortelzeepbel. Ik geef wat meer uitleg bij de volgende methodes:

`addToParent()` voegt een nieuwe top toe aan een top die reeds in de boom zit. De nieuwe top wordt toegevoegd aan de zeepbel van de ouder als hier nog plaats in is. Als deze zeepbel door deze toevoeging volledig gevuld is wordt ze gebalanceerd. Is de zeepbel van de ouder al volledig gevuld dan wordt een nieuwe zeepbel aangemaakt met de nieuwe top als wortel.

`find()` doet hetzelfde als de gelijknamige methode van de superklasse, met dat verschil dat na iedere opzoeking en toevoeging er gesplayt wordt.

`semiSplay()` splayt de zeepbellen vanaf de zeepbel van de meegegeven top. Dit splayen gebeurt als volgt:

1. Eerst worden de drie zeepbellen geselecteerd die moeten worden gesplayed. Eerst wordt de zeepbel geselecteerd waartoe de meegegeven top behoort. Is deze zeepbel niet compleet dan wordt zijn ouderzeepbel in de plaats genomen. Vervolgens worden ook de ouderzeepbel van de laatst geselecteerde zeepbel en zijn ouderzeepbel geselecteerd.
2. Vervolgens worden de toppen van deze zeepbellen in een lijst gestoken door ze in inorde te overlopen. Ook worden de kindertoppen van deze geselecteerde toppen bijgehouden zodat ze later terug kunnen toegevoegd worden.
3. Deze lijst wordt vervolgens in drie gedeeld en van deze deellijsten worden er nieuwe zeepbellen gemaakt met de **TreeBuilder**.
4. Daarna worden deze drie bomen en de kindertoppen uit stap 1 aan elkaar bevestigd.

5. Als laatste stap wordt de bovenste van deze nieuwe drie zeepbellen geslecteerd om te splayen zodat de volgende iteratie kan beginnen.

Deze bewerking herhaalt zich tot de wortelzeepbel of een kindzeepbel van de wortelzeepbel wordt bereikt.

## Deel III

# Experimenten

Nu volgen de resultaten van enkele testen die ik heb uitgevoerd op de zeepbelbomen.

## Opzet

Mijn testresultaten heb ik verkregen door gebruik te maken van de volgende klassen:

**TestResult** : klasse die de tijden en parameters van de verschillende testen voor een zeepbelboom bijhoudt. In deze klasse zijn ook een aantal comparators gedefiniëerd zodat kan gesorteerd worden op naam, tijd om iets toe te voegen ...

**PerformanceTest** : hier gebeuren de echte testen. Deze klasse heeft twee methodes die verschillende benchmarks uitvoeren:

- **testPerK** gaat alle zeepbelboomimplementaties testen met een verschillende  $k$ -waarde en een inputgrootte die gelijk blijft.
- **testPerN** test de zeepbelbomen met een vaste  $k$ -waarde, maar met een stijgende inputgrootte.

Algemeen lopen deze testen als volgt: Eerst wordt de input aangemaakt door een array van integers te maken en door elkaar te shuffelen. Vervolgens wordt iedere bewerking—toevoegen, opzoeken en verwijderen indien geïmplementeerd— tien keer na elkaar uitgevoerd op deze lijst van getallen. Om de resultaten zo stabiel mogelijk te houden wordt iedere reeks voorafgegaan door eenzelfde bewerking die niet wordt gemeten. Ook wordt iedere keer er een nieuwe boom wordt gebruikt de garbagecollection manueel opgeropen.

**PerformanceWriter** : deze klasse bevat de **main()** methode voor de experimenten. Zowel **testPerK** als **testPerN** worden uitgevoerd en de resultaten hiervan worden uitgeschreven naar bestanden.

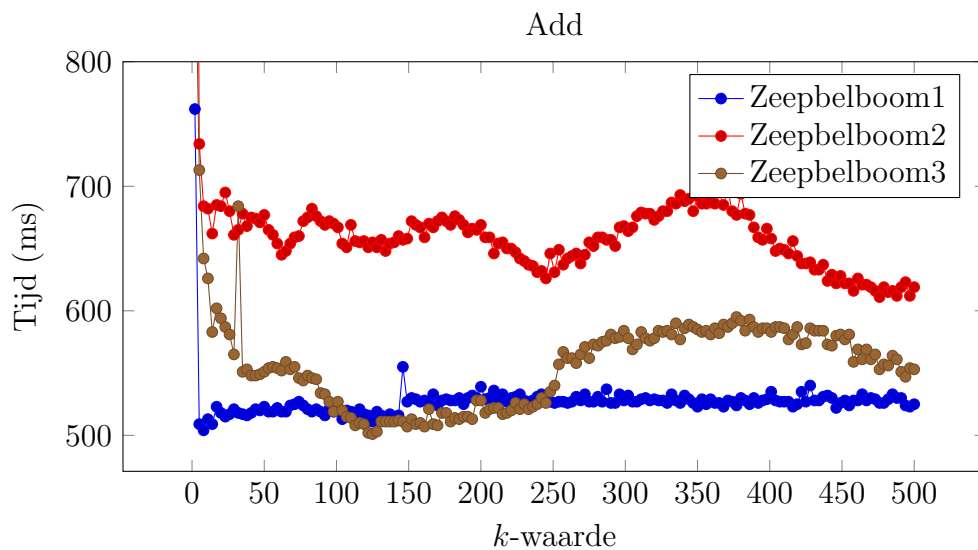
Om de meetresultaten zo stabiel mogelijk te houden heb ik het gehele project gecompileerd in een jar-bestand. Dit gecompileerde bestand heb ik vervolgens

een paar uur laten draaien op een linux-machine. Doormiddel van de kernel-parameter `isolcpus` heb ik één processorkern gereserveerd zodat ik met `taskset` de experimenten heb kunnen uitvoeren op enkel deze processorkern zonder dat andere processen konden interferen.

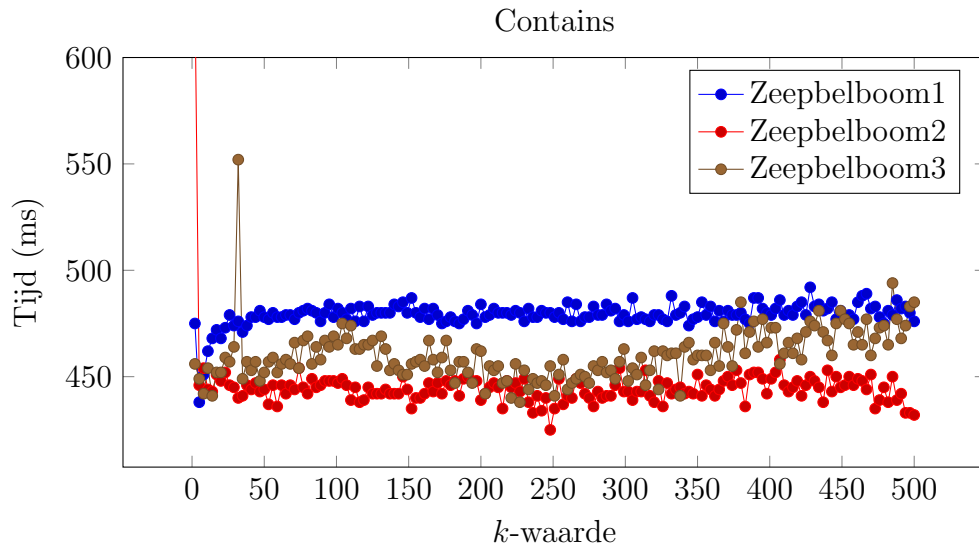
## Resultaten

De resultaten van de meetresultaten door de voorgaande tests kun je in de volgende grafiekjes zien:

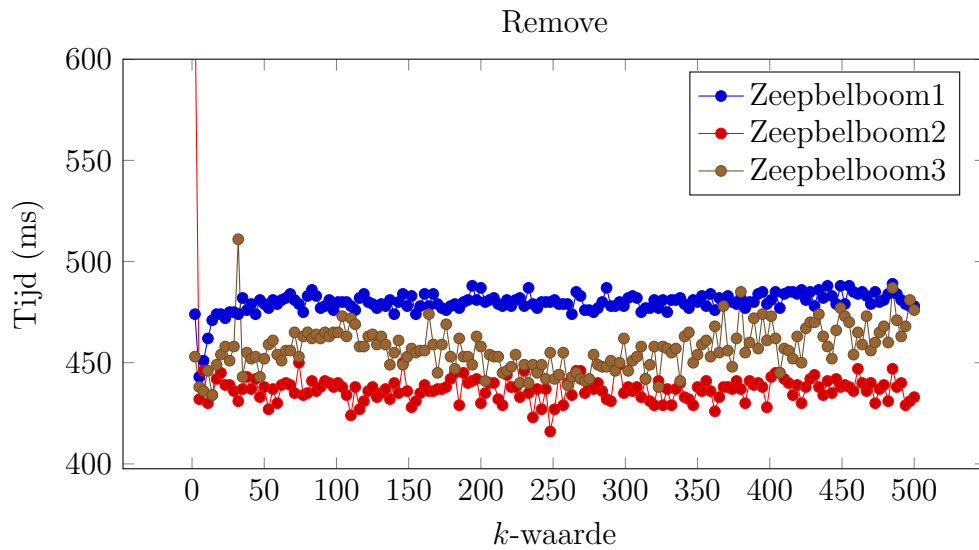
### Gebalanceerde zeepbelbomen



Figuur 4: Toevoegen van 100.000 items aan een gebalanceerde zeepbelboom.



Figuur 5: Opzoekactie in een gebalanceerde zeepbelboom met 100.000 items.



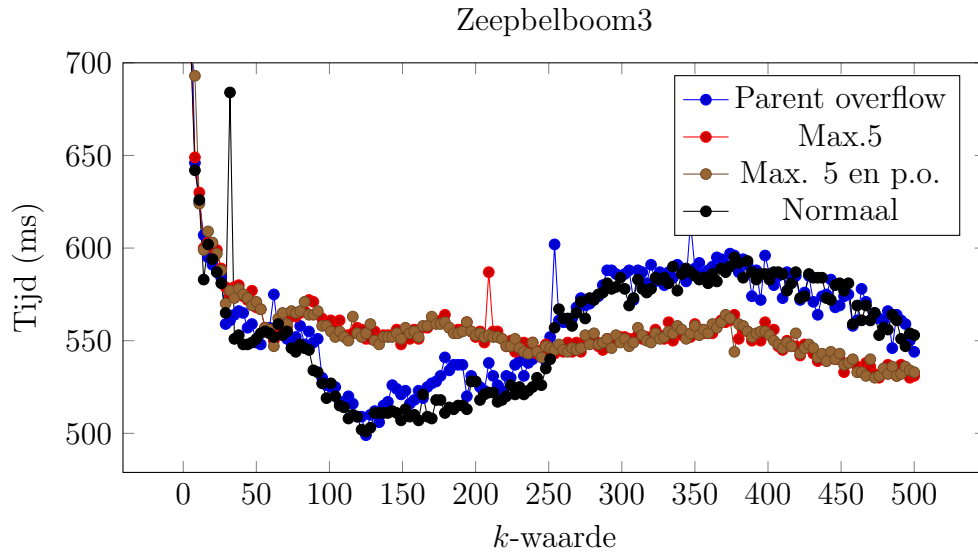
Figuur 6: 100.000 elementen verwijderen uit een gebalanceerde zeepbelboom.

**Toevoegen** Voor de toevoegbewerking merken we in figuur 4 dat de Zeepbelboom1 de beste tijden heeft. Dit komt doordat er veel minder toppen van plaats worden veranderd in de ondeliggende boom. De tweede en derde implementatie duren meestal langer omdat bij iedere splitsing de gehele zeepbel wordt gebalanceerd. Zeepbelboom3 presteert dan weer beter dan de tweede omdat door meerdere toppen toe te voegen aan de ouderzeepbel, er meer plaats is in de nieuwe zeepbellen. Hierdoor zullen er minder splitsingsacties moeten gebeuren en is de toevoegbewerking dus sneller.

**Opzoeken** Op figuur 5 zien we dat Zeepbelboom1 nu de slechtste tijden heeft. Omdat er hier zo weinig mogelijk wijzigingen gedaan worden is deze boom intern minder gebalanceerd dan de andere twee implementaties. Het kan dus langer duren om in Zeepbelboom1 een opzoeking te doen dan bij Zeepbelboom2 en Zeepbelboom3. Zeepbelboom2 presteert beter dan zeepbelboom3 want deze zeepbelboom zal minder zeepbellen bevatten, waardoor de boom in het algemeen ook gebalanceerder is.

**Verwijderen** De verwijderactie in mijn gebalanceerde zeepbelbomen komt eigenlijk neer op een simpele opzoekactie gevolgd door eventueel een booleanse waarde die moet veranderd worden. Daarom kunnen we op figuur 6 zien dat de tijden van de verwijderbewerking overeen komen met die van de opzoeking.

## Zeepbelboom3



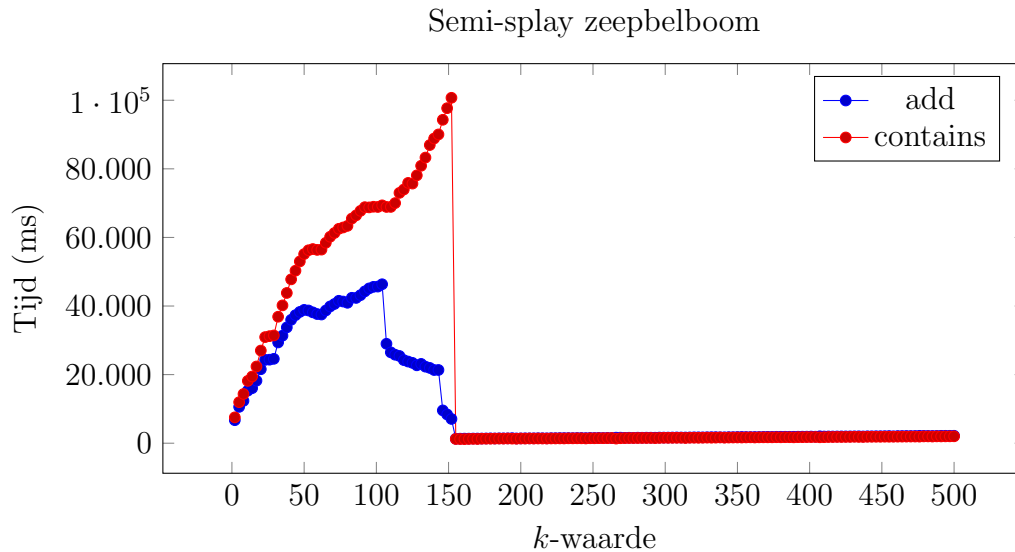
Figuur 7: De verschillende configuraties van Zeepbelboom3.

op figuur 7 kunnen we zien dat het verschil tussen een boom met parent overflow<sup>1</sup> (afgekort p.o.) aan en uit niet significant is. We kunnen wel een verschil zien tussen de zeepbelboom waar een maximum staat op het aantal toppen die toegevoegd mag worden aan de ouderzeepbel. Bij een lage  $k$ -waarde presteert deze iets slechter omdat er door één splitsingsactie terug minder nieuwe zeepbellen bijkomen waardoor er vlugger gesplitst zal moeten worden. Wanneer deze  $k$  groter wordt is dit maximum beter. Er is namelijk een punt waarop er teveel nieuwe zeepbellen worden aangemaakt en deze splitsingsactie trager zal verlopen.

<sup>1</sup>Indien mogelijk worden er zodanig veel toppen toegevoegd aan de ouderzeepbel dat deze verplicht is zich ook te splitsen.



## Semi-splay zeepbelbomen



Figuur 8: Semi-splay zeepbelboom

Ik was enorm verbaasd toen ik figuur 8 zag. Zoals verwacht is een grotere  $k$ -waarde trager voor het toevoegen en opzoeken, er moeten telkens meer bewerkingen op toppen uitgevoerd worden per splaybewerking. Waardoor iedere splaybewerking meer tijd zal innemen hoe groter de  $k$ -waarde is. Maar bij een  $k$  van 150 (en eigenlijk al vanaf 100) doet er zich iets vreemd voor: de toevoeg- en opzoekbewerking verloopt enorm veel vlugger en stijgt ook minder snel in functie van  $k$ . Bij het bespreken van deze merkwaardige resultaten met medeleerlingen hebben ze mij verteld dat hun Zeepbelboom4 ongeveer hetzelfde gedrag vertoont, alleen is het bij mij wel erg extreem. Ik heb helaas geen tijd meer om dit fenomeen te onderzoeken, maar waarschijnlijk zit er ergens een fout in mijn code.

## Besluit

Zeepbelbomen zijn een manier om binaire bomen min of meer gebalanceerd te houden. Zeepbelboom1 presteert het beste wanneer er veel toevoegingen gebeuren. Als er meer opzoekingen gebeuren is het beter om Zeepbelboom2

te gebruiken. Zeepbelboom3 is een middenweg, maar ik kan aanbevelen om een maximum mee te geven wanneer een grote  $k$ -waarde wordt gebruikt. Ten slotte kan ik u afraden om de semi-splay zeepbelboom te gebruiken. Het is namelijk enorm traag, zelfs voor een kleine  $k$ . Daarnaast vertrouw ik mijn eigen implementatie ook niet genoeg.