Goal: Design a data-structure that supports insert, delete, search, and find-by-rank in time $O(\log n)$.

- 1. We choose a red-black tree as the underlying data-structure.
- 2. We store in each node v the size of the sub-tree rooted at v.
- 3. We need to be able to update the size-field in each node without asymptotically affecting the running time of insert, delete, and search. We come back to this step later...

Goal: Design a data-structure that supports insert, delete, search, and find-by-rank in time $\mathcal{O}(\log n)$.

4. How does find-by-rank work? Find-by-rank(k) = Select(root,k) with

```
Algorithm 11 Select(x, i)
```

```
1: if x = \text{null} then return error
```

2: **if** $left[x] \neq null$ **then** $r \leftarrow left[x]$. size + 1 **else** $r \leftarrow 1$

3: **if** i = r **then return** x

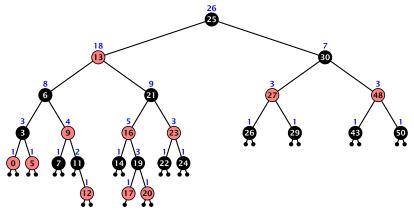
4: if i < r then

5: **return** Select(left[x], i)

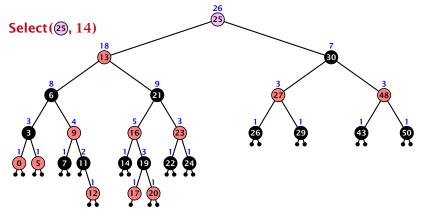
6: **else**

7: **return** Select(right[x], i - r)



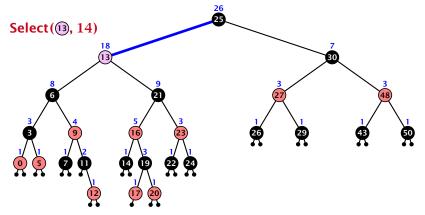


- decide whether you have to proceed into the left or right sub-tree
- adjust the rank that you are searching for if you go right



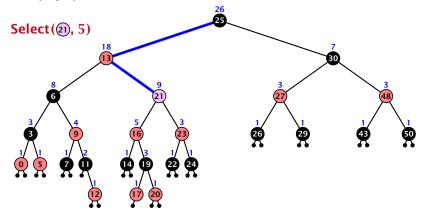
- decide whether you have to proceed into the left or right sub-tree
- adjust the rank that you are searching for if you go right





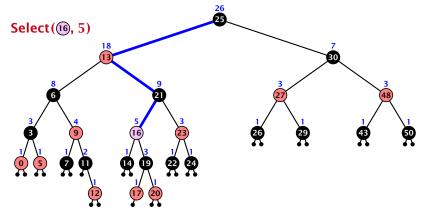
- decide whether you have to proceed into the left or right sub-tree
- adjust the rank that you are searching for if you go right





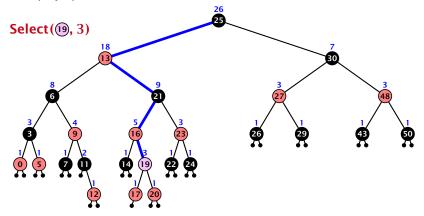
- decide whether you have to proceed into the left or right sub-tree
- adjust the rank that you are searching for if you go right





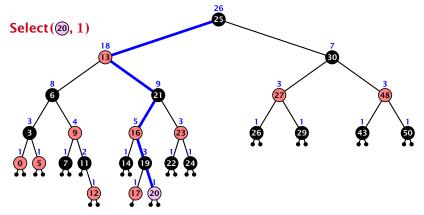
- decide whether you have to proceed into the left or right sub-tree
- adjust the rank that you are searching for if you go right





- decide whether you have to proceed into the left or right sub-tree
- adjust the rank that you are searching for if you go right





- decide whether you have to proceed into the left or right sub-tree
- adjust the rank that you are searching for if you go right



Goal: Design a data-structure that supports insert, delete, search, and find-by-rank in time $O(\log n)$.

3. How do we maintain information?

Search(k): Nothing to do.

Insert(x): When going down the search path increase the size field for each visited node. Maintain the size field during rotations.

Goal: Design a data-structure that supports insert, delete, search, and find-by-rank in time $O(\log n)$.

3. How do we maintain information?

Search(k): Nothing to do.

Insert(x): When going down the search path increase the size field for each visited node. Maintain the size field during rotations.

Goal: Design a data-structure that supports insert, delete, search, and find-by-rank in time $O(\log n)$.

3. How do we maintain information?

Search(k): Nothing to do.

Insert(x): When going down the search path increase the size field for each visited node. Maintain the size field during rotations.

Goal: Design a data-structure that supports insert, delete, search, and find-by-rank in time $\mathcal{O}(\log n)$. \bigcirc

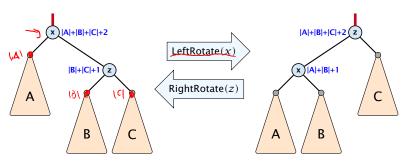
3. How do we maintain information?

Search(k): Nothing to do.

Insert(x): When going down the search path increase the size field for each visited node. Maintain the size field during rotations.

Rotations

The only operation during the fix-up procedure that alters the tree and requires an update of the size-field:



The nodes x and z are the only nodes changing their size-fields.

The new size-fields can be computed locally from the size-fields of the children.

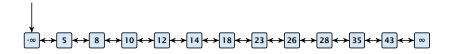
- ightharpoonup time for search $\Theta(n)$
- ightharpoonup time for insert $\Theta(n)$ (dominated by searching the item)
- ▶ time for delete $\Theta(1)$ if we are given a handle to the object, otw. $\Theta(n)$



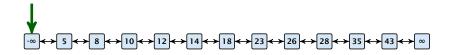
- ightharpoonup time for search $\Theta(n)$
- time for insert $\Theta(n)$ (dominated by searching the item)
- time for delete $\Theta(1)$ if we are given a handle to the object, otw. $\Theta(n)$



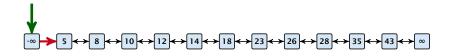
- ightharpoonup time for search $\Theta(n)$
- time for insert $\Theta(n)$ (dominated by searching the item)
- time for delete $\Theta(1)$ if we are given a handle to the object, otw. $\Theta(n)$



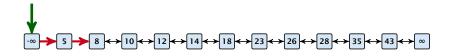
- \blacktriangleright time for search $\Theta(n)$
- time for insert $\Theta(n)$ (dominated by searching the item)
- time for delete $\Theta(1)$ if we are given a handle to the object, otw. $\Theta(n)$



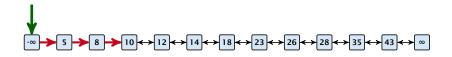
- \blacktriangleright time for search $\Theta(n)$
- time for insert $\Theta(n)$ (dominated by searching the item)
- time for delete $\Theta(1)$ if we are given a handle to the object, otw. $\Theta(n)$



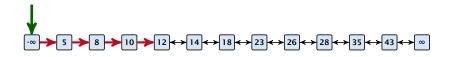
- \blacktriangleright time for search $\Theta(n)$
- time for insert $\Theta(n)$ (dominated by searching the item)
- time for delete $\Theta(1)$ if we are given a handle to the object, otw. $\Theta(n)$



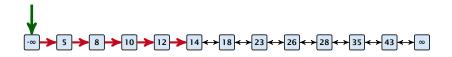
- \blacktriangleright time for search $\Theta(n)$
- time for insert $\Theta(n)$ (dominated by searching the item)
- time for delete $\Theta(1)$ if we are given a handle to the object, otw. $\Theta(n)$



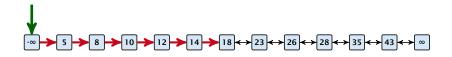
- \blacktriangleright time for search $\Theta(n)$
- time for insert $\Theta(n)$ (dominated by searching the item)
- time for delete $\Theta(1)$ if we are given a handle to the object, otw. $\Theta(n)$



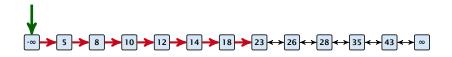
- \blacktriangleright time for search $\Theta(n)$
- time for insert $\Theta(n)$ (dominated by searching the item)
- time for delete $\Theta(1)$ if we are given a handle to the object, otw. $\Theta(n)$



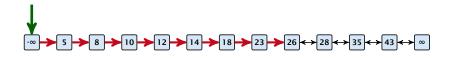
- \blacktriangleright time for search $\Theta(n)$
- time for insert $\Theta(n)$ (dominated by searching the item)
- time for delete $\Theta(1)$ if we are given a handle to the object, otw. $\Theta(n)$



- \blacktriangleright time for search $\Theta(n)$
- time for insert $\Theta(n)$ (dominated by searching the item)
- time for delete $\Theta(1)$ if we are given a handle to the object, otw. $\Theta(n)$



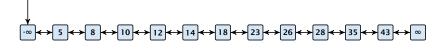
- \blacktriangleright time for search $\Theta(n)$
- time for insert $\Theta(n)$ (dominated by searching the item)
- time for delete $\Theta(1)$ if we are given a handle to the object, otw. $\Theta(n)$



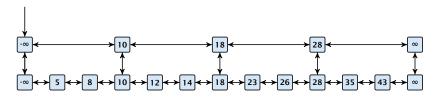
How can we improve the search-operation?

How can we improve the search-operation?

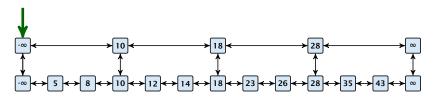
How can we improve the search-operation?



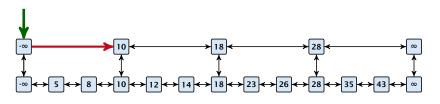
How can we improve the search-operation?



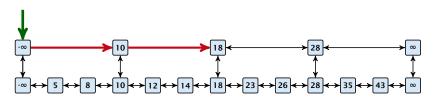
How can we improve the search-operation?



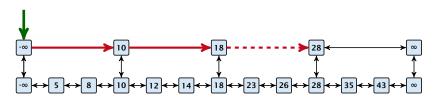
How can we improve the search-operation?



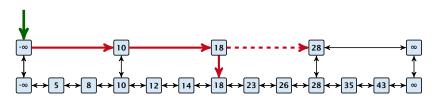
How can we improve the search-operation?



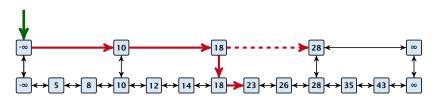
How can we improve the search-operation?



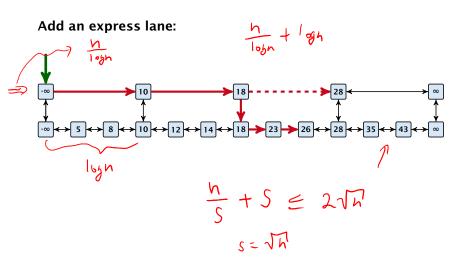
How can we improve the search-operation?



How can we improve the search-operation?

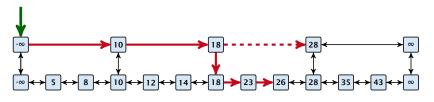


How can we improve the search-operation?



How can we improve the search-operation?

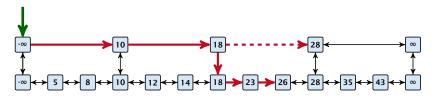
Add an express lane:



Let $|L_1|$ denote the number of elements in the "express lane", and $|L_0|=n$ the number of all elements (ignoring dummy elements).

How can we improve the search-operation?

Add an express lane:

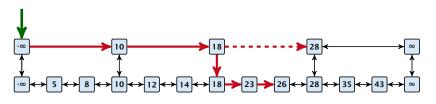


Let $|L_1|$ denote the number of elements in the "express lane", and $|L_0|=n$ the number of all elements (ignoring dummy elements).

Worst case search time: $|L_1| + \frac{|L_0|}{|L_1|}$ (ignoring additive constants)

How can we improve the search-operation?

Add an express lane:



Let $|L_1|$ denote the number of elements in the "express lane", and $|L_0|=n$ the number of all elements (ignoring dummy elements).

Worst case search time: $|L_1| + \frac{|L_0|}{|L_1|}$ (ignoring additive constants)

Choose $|L_1| = \sqrt{n}$. Then search time $\Theta(\sqrt{n})$.

Add more express lanes. Lane L_i contains roughly every $\frac{L_{i-1}}{L_i}$ -th item from list L_{i-1} .

Add more express lanes. Lane L_i contains roughly every $\frac{L_{i-1}}{L_i}$ -th item from list L_{i-1} .

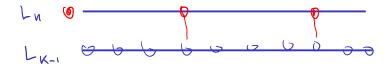
Add more express lanes. Lane L_i contains roughly every $\frac{L_{i-1}}{L_i}$ -th item from list L_{i-1} .

Search(x)
$$(k + 1 \text{ lists } L_0, \ldots, L_k)$$

Find the largest item in list L_k that is smaller than x. At most $|L_k| + 2$ steps.

Add more express lanes. Lane L_i contains roughly every $\frac{L_{i-1}}{L_i}$ -th item from list L_{i-1} .

- Find the largest item in list L_k that is smaller than x. At most $|L_k| + 2$ steps.
- Find the largest item in list L_{k-1} that is smaller than x. At most $\lceil \frac{|L_{k-1}|}{|L_k|+1} \rceil + 2$ steps.



Add more express lanes. Lane L_i contains roughly every $\frac{L_{i-1}}{L_i}$ -th item from list L_{i-1} .

- Find the largest item in list L_k that is smaller than x. At most $|L_k| + 2$ steps.
- Find the largest item in list L_{k-1} that is smaller than x. At most $\lceil \frac{|L_{k-1}|}{|L_k|+1} \rceil + 2$ steps.
- Find the largest item in list L_{k-2} that is smaller than x. At most $\left\lceil \frac{|L_{k-2}|}{|L_{k-1}|+1} \right\rceil + 2$ steps.

Add more express lanes. Lane L_i contains roughly every $\frac{L_{i-1}}{L_i}$ -th item from list L_{i-1} .

- Find the largest item in list L_k that is smaller than x. At most $|L_k| + 2$ steps.
- Find the largest item in list L_{k-1} that is smaller than x. At most $\lceil \frac{|L_{k-1}|}{|L_k|+1} \rceil + 2$ steps.
- Find the largest item in list L_{k-2} that is smaller than x. At most $\left\lceil \frac{|L_{k-2}|}{|L_{k-1}|+1} \right\rceil + 2$ steps.
- **.**..

Add more express lanes. Lane L_i contains roughly every $\frac{L_{i-1}}{L_i}$ -th item from list L_{i-1} .

Search(x)
$$(k + 1 \text{ lists } L_0, \ldots, L_k)$$

- Find the largest item in list L_k that is smaller than x. At most $|L_k| + 2$ steps.
- Find the largest item in list L_{k-1} that is smaller than x. At most $\lceil \frac{|L_{k-1}|}{|L_k|+1} \rceil + 2$ steps. $\leq \frac{|L_k|}{|L_k|} + 3$ Find the largest item in list L_{k-2} that is smaller than x. At
- Find the largest item in list L_{k-2} that is smaller than x. At most $\left\lceil \frac{|L_{k-2}|}{|L_{k-1}|+1} \right\rceil + 2$ steps.
- **.**..
- ► At most $|L_k| + \sum_{i=1}^k \frac{L_{i-1}}{L_i} + 3(k+1)$ steps.

Choose ratios between list-lengths evenly, i.e., $\frac{|L_{i-1}|}{|L_i|} = r$, and, hence, $L_k \approx r^{-k}n$.

Choose ratios between list-lengths evenly, i.e., $\frac{|L_{i-1}|}{|L_i|}=r$, and, hence, $L_k\approx r^{-k}n$.

Worst case running time is: $\mathcal{O}(r^{-k}n + kr)$.

Choose ratios between list-lengths evenly, i.e., $\frac{|L_{i-1}|}{|L_i|}=r$, and, hence, $L_k\approx r^{-k}n$.

Worst case running time is: $\mathcal{O}(r^{-k}n+kr)$. Choose $r=n^{\frac{1}{k+1}}$. Then

$$r^{-k}n + kr$$

Choose ratios between list-lengths evenly, i.e., $\frac{|L_{i-1}|}{|L_i|} = r$, and, hence, $L_k \approx r^{-k}n$.

Worst case running time is: $\mathcal{O}(r^{-k}n + kr)$. Choose $r = n^{\frac{1}{k+1}}$. Then

$$r^{-k}n + kr = \left(n^{\frac{1}{k+1}}\right)^{-k}n + kn^{\frac{1}{k+1}}$$

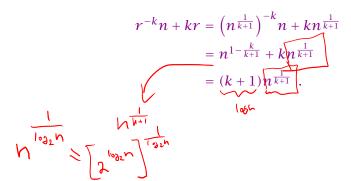
Choose ratios between list-lengths evenly, i.e., $\frac{|L_{i-1}|}{|L_i|} = r$, and, hence, $L_k \approx r^{-k}n$.

Worst case running time is: $\mathcal{O}(r^{-k}n+kr)$. Choose $r=n^{\frac{1}{k+1}}$. Then

$$r^{-k}n + kr = \left(n^{\frac{1}{k+1}}\right)^{-k}n + kn^{\frac{1}{k+1}}$$
$$= n^{1-\frac{k}{k+1}} + kn^{\frac{1}{k+1}}$$

Choose ratios between list-lengths evenly, i.e., $\frac{|L_{i-1}|}{|L_i|} = r$, and, hence, $L_k \approx r^{-k}n$.

Worst case running time is: $\mathcal{O}(r^{-k}n + kr)$. Choose $r = n^{\frac{1}{k+1}}$. Then





Choose ratios between list-lengths evenly, i.e., $\frac{|L_{i-1}|}{|L_i|}=r$, and, hence, $L_k\approx r^{-k}n$.

Worst case running time is: $\mathcal{O}(r^{-k}n+kr)$. Choose $r=n^{\frac{1}{k+1}}$. Then

$$r^{-k}n + kr = \left(n^{\frac{1}{k+1}}\right)^{-k}n + kn^{\frac{1}{k+1}}$$
$$= n^{1-\frac{k}{k+1}} + kn^{\frac{1}{k+1}}$$
$$= (k+1)n^{\frac{1}{k+1}}.$$

Choosing $k = \Theta(\log n)$ gives a logarithmic running time.

How to do insert and delete?

Use randomization instead!

How to do insert and delete?

If we want that in L_i we always skip over roughly the same number of elements in L_{i-1} an insert or delete may require a lot of re-organisation.

Use randomization instead!

How to do insert and delete?

If we want that in L_i we always skip over roughly the same number of elements in L_{i-1} an insert or delete may require a lot of re-organisation.

Use randomization instead!

Insert:

- A search operation gives you the insert position for element x in every list.
- Flip a coin until it shows head, and record the number $t \in \{1, 2, ...\}$ of trials needed.
- lnsert x into lists L_0, \ldots, L_{t-1} .

Delete

- You get all predecessors via backward pointers.
- Delete x in all lists it actually appears in.

Insert:

- A search operation gives you the insert position for element x in every list.
- Flip a coin until it shows head, and record the number $t \in \{1, 2, ...\}$ of trials needed.
- lnsert x into lists L_0, \ldots, L_{t-1} .

Delete

```
You get all predecessors via backward pointers.

Delete - in all lists it actually appears in.
```

Insert:

- A search operation gives you the insert position for element x in every list.
- ▶ Flip a coin until it shows head, and record the number $t \in \{1, 2, ...\}$ of trials needed.
- lnsert x into lists L_0, \ldots, L_{t-1} .

Delete

You get all predecessors via backward pointers.

Insert:

- A search operation gives you the insert position for element x in every list.
- Flip a coin until it shows head, and record the number $t \in \{1, 2, ...\}$ of trials needed.
- ▶ Insert x into lists L_0, \ldots, L_{t-1} .

Delete

Insert:

- A search operation gives you the insert position for element x in every list.
- ▶ Flip a coin until it shows head, and record the number $t \in \{1, 2, ...\}$ of trials needed.
- ▶ Insert x into lists L_0, \ldots, L_{t-1} .

Delete:

- You get all predecessors via backward pointers.
- Delete x in all lists it actually appears in.

Insert:

- A search operation gives you the insert position for element x in every list.
- Flip a coin until it shows head, and record the number $t \in \{1, 2, ...\}$ of trials needed.
- Insert x into lists L_0, \ldots, L_{t-1} .

Delete:

- You get all predecessors via backward pointers.
- ightharpoonup Delete x in all lists it actually appears in.

The time for both operations is dominated by the search

199/209

Insert:

- A search operation gives you the insert position for element x in every list.
- ▶ Flip a coin until it shows head, and record the number $t \in \{1, 2, ...\}$ of trials needed.
- ▶ Insert x into lists L_0, \ldots, L_{t-1} .

Delete:

- You get all predecessors via backward pointers.
- ▶ Delete x in all lists it actually appears in.

The time for both operations is dominated by the search time.

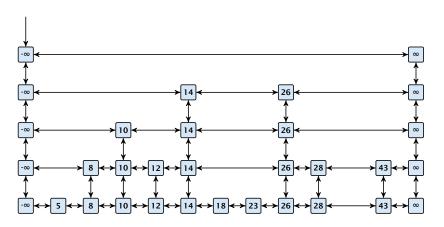
199/209

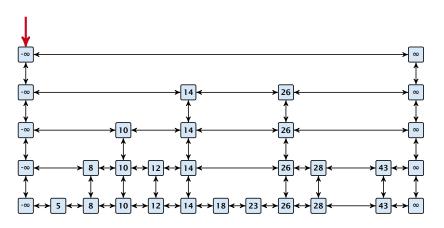
Insert:

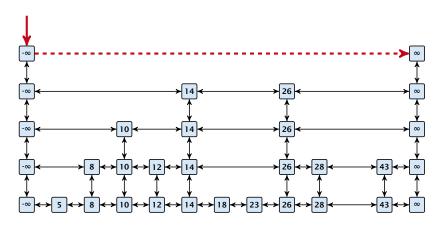
- A search operation gives you the insert position for element x in every list.
- ▶ Flip a coin until it shows head, and record the number $t \in \{1, 2, ...\}$ of trials needed.
- ▶ Insert x into lists L_0, \ldots, L_{t-1} .

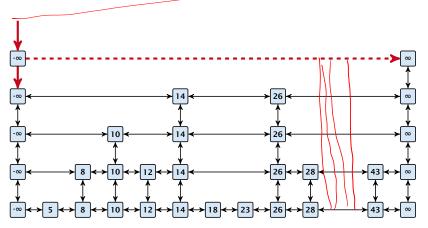
Delete:

- You get all predecessors via backward pointers.
- Delete x in all lists it actually appears in.

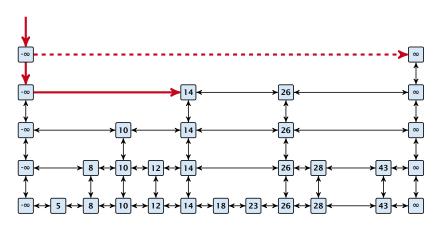


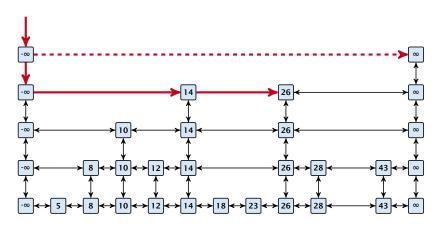


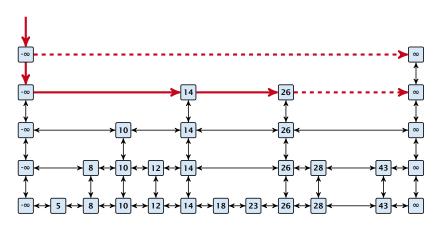


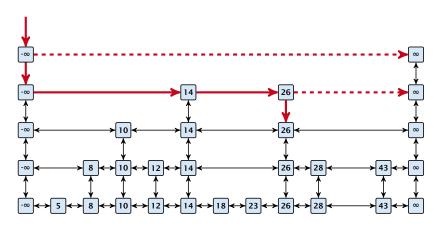


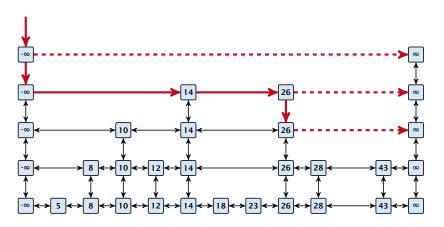




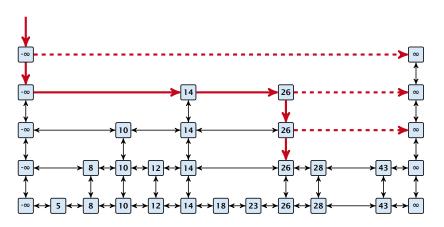




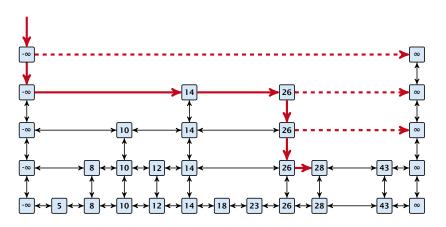


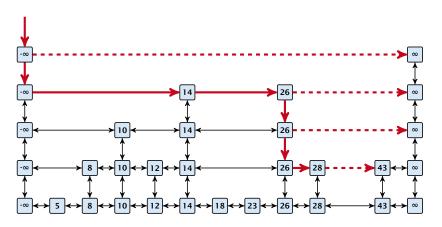


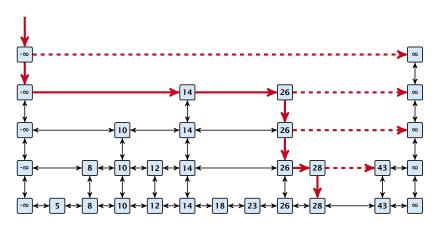
Insert (35):

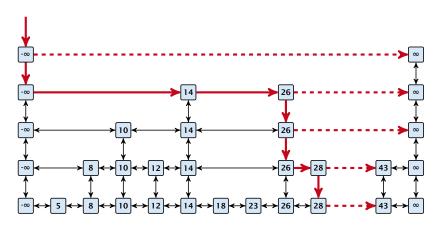


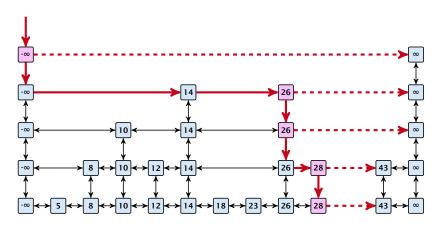
7.5 Skip Lists

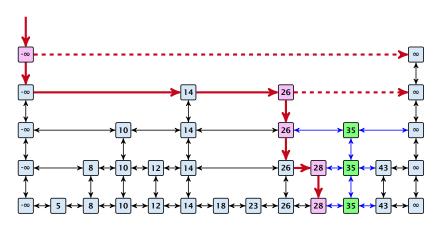












Definition 18 (High Probability)

We say a **randomized** algorithm has running time $\mathcal{O}(\log n)$ with high probability if for any constant α the running time is at most $\mathcal{O}(\log n)$ with probability at least $1 - \frac{1}{n^{\alpha}}$.

Here the O-notation hides a constant that may depend on α .



Definition 18 (High Probability)

We say a **randomized** algorithm has running time $\mathcal{O}(\log n)$ with high probability if for any constant α the running time is at most $\mathcal{O}(\log n)$ with probability at least $1 - \frac{1}{n^{\alpha}}$.

Here the \mathcal{O} -notation hides a constant that may depend on α .



Suppose there are polynomially many events E_1, E_2, \ldots, E_ℓ , $\ell = n^c$ each holding with high probability (e.g. E_i may be the event that the i-th search in a skip list takes time at most $\mathcal{O}(\log n)$).

Suppose there are polynomially many events $E_1, E_2, ..., E_\ell$, $\ell = n^c$ each holding with high probability (e.g. E_i may be the event that the i-th search in a skip list takes time at most $\mathcal{O}(\log n)$).

$$\Pr[E_1 \wedge \cdots \wedge E_{\ell}]$$

Suppose there are polynomially many events $E_1, E_2, ..., E_\ell$, $\ell = n^c$ each holding with high probability (e.g. E_i may be the event that the i-th search in a skip list takes time at most $\mathcal{O}(\log n)$).

$$\Pr[E_1 \wedge \cdots \wedge E_{\ell}] = 1 - \Pr[\bar{E}_1 \vee \cdots \vee \bar{E}_{\ell}]$$

Suppose there are polynomially many events E_1, E_2, \ldots, E_ℓ , $\ell = n^c$ each holding with high probability (e.g. E_i may be the event that the i-th search in a skip list takes time at most $\mathcal{O}(\log n)$).

$$\Pr[E_1 \wedge \cdots \wedge E_{\ell}] = 1 - \Pr[\bar{E}_1 \vee \cdots \vee \bar{E}_{\ell}]$$

$$\geq 1 - n^c \cdot n^{-\alpha}$$

Suppose there are polynomially many events $E_1, E_2, ..., E_\ell$, $\ell = n^c$ each holding with high probability (e.g. E_i may be the event that the i-th search in a skip list takes time at most $\mathcal{O}(\log n)$).

$$\Pr[E_1 \wedge \cdots \wedge E_{\ell}] = 1 - \Pr[\bar{E}_1 \vee \cdots \vee \bar{E}_{\ell}]$$

$$\geq 1 - n^c \cdot n^{-\alpha}$$

$$= 1 - n^{c - \alpha}.$$

Suppose there are polynomially many events $E_1, E_2, ..., E_\ell$, $\ell = n^c$ each holding with high probability (e.g. E_i may be the event that the i-th search in a skip list takes time at most $\mathcal{O}(\log n)$).

Then the probability that all E_i hold is at least

$$\Pr[E_1 \wedge \cdots \wedge E_{\ell}] = 1 - \Pr[\bar{E}_1 \vee \cdots \vee \bar{E}_{\ell}]$$

$$\geq 1 - n^c \cdot n^{-\alpha}$$

$$= 1 - n^{c - \alpha}.$$

This means $\Pr[E_1 \wedge \cdots \wedge E_{\ell}]$ holds with high probability.