

Notizen zu Algorithmen II

Jens Ochsenmeier

5. Februar 2018

Inhaltsverzeichnis

- 1 Stringology • 5
 - 1.1 Strings sortieren • 5
 - 1.2 Pattern Matching • 6
 - 1.3 Datenkompression • 13
- 2 Range Minimum Queries • 15
 - 2.1 Lösung 1 — $\langle O(n), O(\log n) \rangle$ • 16
 - 2.2 Lösung 2 — $\langle O(n \log n), O(1) \rangle$ • 16
 - 2.3 Lösung 3 — $\langle O(n \log \log n), O(1) \rangle$ • 17
 - 2.4 Lösung 4 — $\langle O(n), O(1) \rangle$ • 17
 - 2.5 Lowest Common Ancestor • 18

1

Stringology

Inhalt dieses Kapitels:

- Strings sortieren
- Patterns suchen
- Datenkompression

1.1 Strings sortieren

Naive Sortierv Verfahren, wie sie aus der Vorlesung “Algorithmen 1” bekannt sind, sind beim Sortieren von Strings ineffizient, deswegen gibt es für das Sortieren von Strings andere Algorithmen. Ein solcher ist der **Multikey Quicksort**-Algorithmus:

```
mkqsort (S: String Seq, l: N): String Seq
assert  $\forall e, e' \in S: e[1 \dots l - 1] = e'[1 \dots l - 1]$ 
if  $|S| \leq 1$  then return S
pick  $p \in S$  randomly
return concatenation of
  mkqsort ( $\{e \in S : e[l] < p[l]\}$ , l),
  mkqsort ( $\{e \in S : e[l] = p[l]\}$ , l + 1),
  mkqsort ( $\{e \in S : e[l] > p[l]\}$ , l)
```

Abbildung 1.1. Pseudocode-Implementierung des Multikey-Quicksort-Algorithmus.

Dieser Algorithmus sortiert eine String-Sequenz und nimmt an, dass die ersten $l - 1$ Buchstaben bereits sortiert wurden.

Zuerst wird ein zufälliges Pivotelement gewählt. Danach wird die übergebene Sequenz an Strings in drei Teilsequenzen geteilt:

1. Sequenz an Strings, deren l -ter Buchstabe kleiner ist als der l -te Buchstabe des Pivotelements.
2. Sequenz an Strings, deren l -ter Buchstabe derselbe ist wie der l -te Buchstabe des Pivotelements.
3. Sequenz an Strings, deren l -ter Buchstabe größer ist als der l -te Buchstabe des Pivotelements.

Auf die erste und dritte Teilsequenz wird der Algorithmus nun rekursiv mit dem selben Parameter l ausgeführt, da die Buchstaben an der l -ten Position nicht übereinstimmen (müssen) — auf die zweite Teilsequenz wird der Algorithmus rekursiv mit dem Parameter $l + 1$ ausgeführt, weil hier die l -ten Buchstaben aller Wörter in der Sequenz gleich sind.

Die Laufzeit des Algorithmus ist in $O(|S| \log |S| + d)$, wobei d die Summe der eindeutigen Präfixe der Strings in S ist.

1.2 Pattern Matching

Hinweis: In diesem Abschnitt sind Arrays 1-basiert.

In diesem Abschnitt wird es darum gehen, alle oder zumindest ein Vorkommen eines **Patterns** $P = p_1 \dots p_m$ in einem gegebenen **Text** $T = t_1 \dots t_n$ zu finden. Im Allgemeinen ist $n \gg m$, also der Text wesentlich länger als das Pattern, das wir in ihm suchen.

Naives Pattern Matching

Das naive Vorgehen ist, an jeder Position von T zu schauen, ob an dieser das gesuchte Pattern vorkommt. Offensichtlich ist dieser Algorithmus in $O(nm)$, da im schlimmsten Fall für jede Position des Textes das gesamte Pattern durchlaufen werden muss. Dieser Algorithmus kann folgendermaßen implementiert werden:

```

NAIVEPATTERNMATCH (P, T)
i, j := 1
while i ≤ n - m + 1
  while j ≤ m ∧ ti+j-1 = pj do j++
  if j > m then return "P occurs at pos i in T"
  i++
  j := 1

```

Abbildung 1.2. Pseudocode-Implementierung des naiven Pattern-Matching-Algorithmus.

Knuth-Morris-Pratt

Ein anderer Algorithmus zum Finden von Patterns in einem gegebenen Text ist der **Knuth-Morris-Pratt-Algorithmus**. Dieser hat sogar optimale Laufzeit, nämlich $O(n + m)$.

Idee dieses Algorithmus ist es, das Pattern eleganter nach vorne zu verschieben, wenn es einen Mismatch zwischen Text und Pattern gibt. Hierfür brauchen wir ein Hilfswerkzeug:

Für einen String S mit Länge k sei $\alpha(S)$ die Länge des Längsten Präfixes von $S_{1\dots k-1}$, das auch Suffix von $S_{2\dots k}$ ist.¹

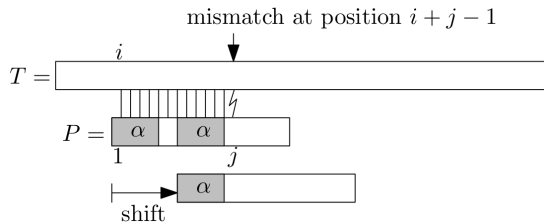


Abbildung 1.3. Idee beim Verschieben des Patterns: α wurde bereits gematcht. Früher als mit dem bereits gematchten Suffix kann das nächste Vorkommen von P nicht auftauchen, also kann man P direkt um $j - 1 - \alpha$ verschieben.

Der Algorithmus besteht aus zwei Teilen:

1. **Border-Array berechnen** ($O(m)$). Damit die oben erläuterten Verschiebungen nachher effizient durchgeführt werden können, berechnen wir für das leere Wort

¹ Wir lassen absichtlich bei Betrachtung des Präfixes den letzten und bei Betrachtung des Suffixes den ersten Buchstaben weg, damit $\alpha(S) = 0$ ist, wenn $|S| = k = 1$ ist.

und jeden Buchstaben in P einen α -Wert. Diese Werte ergeben das **Border-Array**:

$$\text{border}[j] = \begin{cases} -1, & \text{falls } j = 1 \\ \alpha(P_{1 \dots j-1}), & \text{sonst} \end{cases}.$$

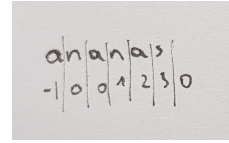


Abbildung 1.4. Beispiel für das Border-Array eines Patterns.

2. **Pattern matchen** ($O(n)$). Nun verwenden wir das erstellte Border-Array, um Vorkommnisse von P in T zu finden. Wir starten sowohl im Text als auch im Pattern an Position 1 und fangen an zu matchen. Kommt es an Position $1 \leq j \leq m$ des Patterns zu einem Mismatch, so können wir P direkt um $j - \text{border}[j] - 1$ verschieben. In Pseudocode sieht das so aus:

```
KMPMATCH (P, T)
i, j := 1
while i ≤ n - m + 1
  while j ≤ m ∧ ti+j-1 = pj do j++
  if j > m then return "P occurs at pos i in T"
  i += j - border[j] + 1
  j := max{1, border[j] + 1}
```

Eine Ausführung des Algorithmus kann also folgendermaßen aussehen:

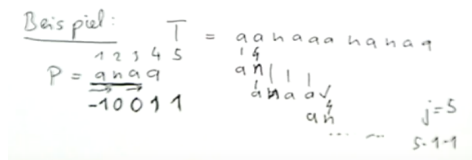


Abbildung 1.5. Beispiel für das Verwenden des Knuth-Morris-Pratt-Algorithmus.

Suffix-Arrays

Im Folgenden werden Arrays wieder mit Position 0 beginnen. Wir verwenden desweiteren folgende Festlegungen:

- Ein **String** ist ein Array von Buchstaben,

$$S[0 \dots n) := S[0 \dots n - 1] := [S[0], \dots, S[n - 1]].$$

- Das **Suffix** S_i sei der Substring $S[i \dots n)$ von S .

- Wir setzen an das Ende jedes Strings ausreichend viele **Endmarkierungen**:
 $S[n] := S[n+1] := \dots := 0$. 0 sei per Definition kleiner als alle anderen vorkommenden Zeichen.

Das **Suffix-Array** eines Strings lässt sich nun folgendermaßen konstruieren:

- Bilde die Menge aller Suffixe S_i ($i = 0, \dots, n-1$) des Strings.
- Sortiere die Menge aller Suffixe des Strings (z.B. mit **Multikey Quicksort**).

0	banana	5	a
1	anana	3	ana
2	nana	1	anana
3	ana	0	banana
4	na	4	na
5	a	2	nana

Abbildung 1.6. Beispiel für die Konstruktion des Suffix-Arrays des Strings "banana".

Mithilfe dieses Suffix-Arrays lassen sich später viele Suchprobleme in Linearzeit lösen. Beispielsweise ist die Suche nach dem längsten Substring, der (eventuell mit Überschneidung) zweimal im Text vorkommt, linear — dafür muss nach Berechnung des Suffix-Arrays der längste String gefunden werden, der Präfix von zwei Strings im Suffix-Array ist (im Beispiel oben wäre das "ana").

Berechnung des Suffix-Arrays in Linearzeit

Das Suffix-Array eines Strings lässt sich in Linearzeit berechnen.² Hier soll lediglich das Prinzip erläutert werden, genauere Angaben gibt es im Paper.

Wir betrachten den String

$$T[0, n) = \underset{\substack{0 \\ x}}{\underset{1}{a}} \underset{2}{b} \underset{3}{b} \underset{4}{a} \underset{5}{d} \underset{6}{a} \underset{7}{b} \underset{8}{b} \underset{9}{a} \underset{10}{d} \underset{11}{o}.$$

Unser Ziel ist das Suffix-Array

$$SA = (12, 1, 6, 4, 9, 3, 8, 2, 7, 5, 10, 11, 0).$$

Wir gehen wie folgt vor:

0. **Suffixe wählen.** Sei

$$B_k = \{i \in [0, n] : i \bmod 3 = k\}$$

und $C = B_1 \cup B_2$ sowie S_C die Menge der entsprechenden Suffixe. C ist also die Menge aller Positionen in T , an denen Suffixe mit einer nicht durch 3 teilbaren Länge beginnen. Hier ist $C = \{1, 4, 7, 10, 2, 5, 8, 11\}$.

² Kärkkäinen, Sanders, Burkhardt: Linear Work Suffix Array Construction

1. **Gewählte Suffixe sortieren.** Wir fügen am Ende von T beliebig viele \emptyset hinzu und bilden zuerst für $k = 1, 2$ die Strings

$$R_k = [t_k t_{k+1} t_{k+2}] [t_{k+3} t_{k+4} t_{k+5}] \cdots [t_{\max B_k} t_{\max B_k+1} t_{\max B_k+2}].$$

Der Charaktere von R_k sind also Tripel. Das letzte Tripel ist immer eindeutig, weil $t_{\max B_k+2} = 0$. Sei $R = R_1 \odot R_2$.

Hier ist

$$R = [\text{abb}][\text{ada}][\text{bba}][\text{do}\emptyset][\text{bba}][\text{dab}][\text{bad}][\text{o}\emptyset\emptyset].$$

Die Ordnung der Suffixe von R stimmt mit der Ordnung der Suffixe S_i überein, deswegen genügt es, die Suffixe von R zu sortieren.

Wir sortieren R nun, indem wir die einzelnen Charaktere von R sortieren und durch ihren Rang in R ersetzen:

$$\text{SA}_R = (8, 0, 1, 6, 4, 2, 5, 3, 7).$$

Nun weisen wir jedem Suffix einen Rang zu. Dazu sei $\text{rank}(S_i)$ der Rang von S_i in C . Für $i \in B_0$ sei $\text{rank}(S_i)$ nicht definiert.

Hier ist $\text{rank}(S_i) = \perp \ 1 \ 4 \ \perp \ 2 \ 6 \ \perp \ 5 \ 3 \ \perp \ 7 \ 8 \ \perp \ 0 \ 0$.

2. **Restliche Suffixe sortieren.** Jeder Suffix $S_i \in S_{B_0}$ sei dargestellt durch $(t_i, \text{rank}(S_{i+1}))$. Da wir alle anderen Suffixe oben schon sortiert haben ist $\text{rank}(S_{i+1})$ hier stets definiert.

Offensichtlich ist

$$S_i \leq S_j \Leftrightarrow (t_i, \text{rank}(S_{i+1})) \leq (t_j, \text{rank}(S_{j+1})),$$

also lassen sich die Paare Radix-sortieren.

Hier ist

$$S_{12} < S_6 < S_9 < S_3 < S_0, \quad \text{weil} \quad (\emptyset, 0) < (a, 5) < (a, 7) < (b, 2) < (x, 1).$$

3. **Zusammenführen.** Das Zusammenführen erfolgt vergleichsbasiert. Beim Vergleichen on $S_i \in S_C$ mit $S_j \in S_{B_0}$ unterscheiden wir zwei Fälle:

$$i \in B_1 : S_i \leq S_j \Leftrightarrow (t_i, \text{rank}(S_{i+1})) \leq (t_j, \text{rank}(S_{j+1}))$$

$$i \in B_2 : S_i \leq S_j \Leftrightarrow (t_i, t_{i+1}, \text{rank}(S_{i+2})) \leq (t_j, t_{j+1}, \text{rank}(S_{j+2}))$$

Hier ist z.B. $S_1 < S_6$ weil $(a, 4) < (a, 5)$ und $S_3 < S_8$ weil $(b, a, 6) < (b, a, 7)$.

Suchen in Suffix-Arrays

Um ein Pattern in einem String zu finden, zu dem man das Suffix-Array konstruiert hat, muss man lediglich ein Suffix finden, dass das gesuchte Pattern als Präfix hat. Man kann so beispielsweise mit binärer Suche in $O(m \log n)$ ein Vorkommen von P in T finden.

Nutzen wir eine zusätzliche Struktur, das **LCP-Array** — dieses speichert in $LCP[i]$ die Länge des längsten gemeinsamen Präfixes von $SA[i]$ und $SA[i - 1]$ — so können wir die Suchzeit auf $O(m + \log n)$ reduzieren.

0	banana	SA =	5	a	LCP =	0	a
1	anana		3	ana		1	a na
2	nana		1	anana		3	an ana
3	ana		0	banana		0	banana
4	na		4	na		0	na
5	a		2	nana		2	na na

Abbildung 1.7. Suffixe, Suffix-Array und LCP-Array des Strings "banana".

Um das LCP-Array berechnen zu können brauchen wir das **invertierte Suffix-Array**. Dieses gibt Aufschluss darüber, wo im Suffix-Array ein bestimmter Suffix steht. Offensichtlich ist $SA^{-1}[SA[i]] = i$.

Der Algorithmus sieht folgendermaßen aus ($O(n)$):

```

CALCULATELCPARRAY ( $SA^{-1}$ , SA)
 $h := 0$ ,  $LCP[1] := 0$ 
for  $i = 1, \dots, n$  do
  if  $SA^{-1}[i] \neq 1$  then
    while  $t_{i+h} = t_{SA[SA^{-1}[i]-1]+h}$  do  $h++$ 
     $LCP[SA^{-1}[i]] := h$ 
     $h := \max(0, h - 1)$ 

```

Suffix-Bäume

Noch anschaulicher, allerdings wesentlich platzverbrauchender, sind **Suffix-Bäume** von Strings. Sie sind formal der *komprimierte Trie der Suffixe* und lassen sich (wenn auch sehr kompliziert) in $O(n)$ berechnen.

Bevor wir den Suffix-Baum eines Strings bilden hängen wir hinten an den String noch einen Charakter dran, der nicht im Alphabet des Strings vorkommt. Das hat den Vorteil, dass anschließend alle Suffixe in einem Blatt des Baums enden.

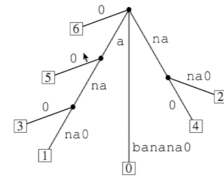


Abbildung 1.8. Beispiel für den Suffix-Baum des Strings "banana".

“Naiv” ist die Erstellung des Suffixbaums in $O(n^2)$. Man kann ihn aber auch aus Suffix-Array und LCP-Array in Linearzeit konstruieren. Dazu hängt man die Suffixe sukzessive in der Tiefe ein, die ihr LCP-Wert angibt:

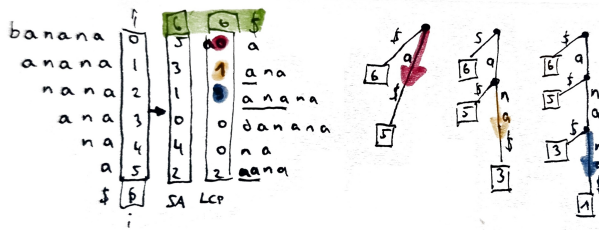


Abbildung 1.9. Sukzessive Konstruktion des Suffix-Baums aus Suffix- und LCP-Array. Zuerst hängt man \$ und das erste Suffix (dessen LCP-Wert immer 0 ist) an die Wurzel. Anschließend nutzt man den LCP-Wert des darauffolgenden Suffixes (hier 1), um festzulegen, wo der Suffix zum Baum hinzugefügt werden muss (durch Pfeile gekennzeichnet).

Die Suche in einem Suffix-Baum ist relativ simpel — man muss lediglich den entsprechenden Kanten entlanglaufen, alle Vorkommen des Patterns liegen im entsprechenden Teilbaum.

Zur Angabe der Komplexitäten sind zwei Fälle zu unterscheiden:

1. Die ausgehenden Kanten sind als Arrays der Größe $|\Sigma|$ gespeichert. Dann ist die Suchzeit in $O(m)$ und der Gesamtplatzbedarf in $O(n|\Sigma|)$.
2. Die ausgehenden Kanten sind als Arrays gespeichert, deren Größe proportional zur Anzahl der Kinderknoten ist. Dann ist die Suchzeit in $O(m \log |\Sigma|)$ und der Gesamtplatzbedarf in $O(n)$.

1.3 Datenkompression

Eine Anwendung der Suffix-Arrays und -Trees ist die **Datenkompression**. Inhalt dieser Vorlesung wird ausschließlich die *verlustfreie Textkompression* sein.

Wörterbuchbasierte Textkompression

Für besonders große Datenbestände bietet sich eine **wörterbuchbasierte Textkompression** an. Grundidee ist, $\Sigma' \subseteq \Sigma^*$ zu wählen und $S \in \Sigma^*$ durch $S' = \langle s'_1, \dots, s'_k \rangle \in \Sigma'^*$ zu ersetzen, sodass $S = s'_1 \cdot \dots \cdot s'_k$ ist. Problem ist der hohe zusätzliche Platzbedarf für das Wörterbuch.

Lempel-Ziv-Kompression

Die **Lempel-Ziv-Kompression** baut das Wörterbuch *on the fly* bei Codierung und Decodierung, sodass dieses nicht explizit gespeichert werden muss.

```

NAIVELZCOMPRESS( $\langle s_1, \dots, s_n \rangle, \Sigma$ )
 $D := \Sigma$  // init dictionary
 $p := s_1$  // current string
for  $i := 2$  to  $n$  do
  if  $p \cdot s_i \in D$  then  $p := p \cdot s_i$ 
  else
    output code for  $p$ 
     $D := D \cup p \cdot s_i$ 
     $p := s_i$ 
output code for  $p$ 

```

```

NAIVELZDECODE( $\langle c_1, \dots, c_k \rangle$ )
 $D := \Sigma$ 
output decode( $c_1$ )
for  $i := 2$  to  $k$  do
  if  $c_i \in D$  then
     $D := D \cup \text{decode}(c_{i-1}) \cdot \text{decode}(c_i)[1]$ 
  else
     $D := D \cup \text{decode}(c_{i-1}) \cdot \text{decode}(c_{i-1})[1]$ 
  output decode( $c_i$ )

```

Abbildung 1.10. Kompressions- und Dekodierungs-Algorithmus für die Lempel-Ziv-Kompression.

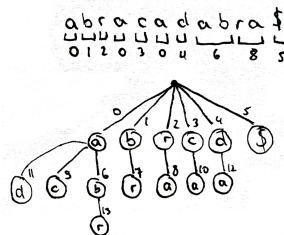


Abbildung 1.11. Beispiel für das Bilden der Lempel-Ziv-Kompression des Wortes "abracadabra". Der Baum wird *on the fly* berechnet und nicht übergeben, sondern nur die komprimierte Information und das Alphabet.

2

Range Minimum Queries

Eine **range minimum Query** gibt für ein array A ($|A| = n$) die Position des kleinsten Elements zwischen zwei Begrenzern $1 \leq l < r \leq n$ zurück:

$$\text{rmq}_A(l, r) = (\arg) \min_{l \leq k \leq r} A[k]$$

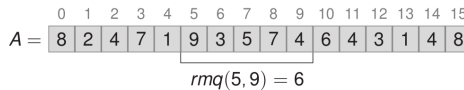


Abbildung 2.1. Beispiel einer range minimum query.

Ziel dieses Kapitels wird sein, einen Algorithmus anzugeben, mit dem eine RMQ-Abfrage in konstanter Zeit beantwortet werden kann, nachdem eine $2n + o(n)$ große Datenstruktur in Linearzeit vorbereitet wurde.

Ein naiver Ansatz, um eine range minimum query auszuführen, ist, einfach das Array zu durchlaufen und das Minimum zu speichern (und wenn nötig zu aktualisieren). Dafür ist keine Vorbereitungsarbeit nötig (also $O(1)$) und die Abfrage ist in $O(n)$. Wir notieren

$$\langle O(1), O(n) \rangle.$$

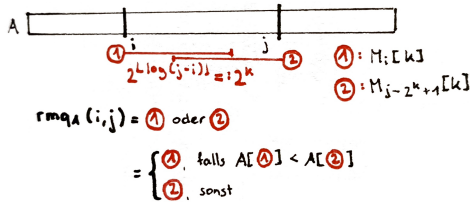


Abbildung 2.3. Funktionsweise des zweiten rmq-Algorithmus.

2.3 Lösung 3 — $\langle O(n \log \log n), O(1) \rangle$

Wir wenden folgende Prozedur an:

1. A in $t = \frac{n}{\log n}$ Blöcke B_0, \dots, B_{t-1} der Größe $\log n$ unterteilen.
2. Array $S[0, t-1]$ mit $S[i] = \min \{x \in B_i\}$ erstellen, rmq-Struktur nach Lösung 2 für S berechnen.
3. Für jeden Block B_i rmq-Struktur nach Lösung 2 berechnen.

Diese Prozedur liegt in $O(n \log \log n)$.

Soll nun $\text{rmq}_A(l, r)$ bestimmt werden, so geht das folgendermaßen:

1. Bestimme die Blöcke $l \in B_{l'}$ und $r \in B_{r'}$.
2. Berechne $m = \text{rmq}_S(l' + 1, r' - 1)$. Wir nutzen also die Struktur über S , um die rmq-Werte der Blöcke zwischen den beiden Grenzblöcken zu berechnen.
3. Es seien k_0, k_1, k_2 die rmq_A -Resultate in den Blöcken l', r' und m .
4. $\text{rmq}_A(l, r) = \arg \min \{A[k_0], A[k_1], A[k_2]\}$.

2.4 Lösung 4 — $\langle O(n), O(1) \rangle$

Zur Konstruktion des Algorithmus mit linearem Platzverbrauch benötigen wir kartesische Bäume.

Kartesischer Baum

Der **kartesische Baum** eines Arrays A ist folgendermaßen definiert:

- Wurzel des Baums ist das (linkste) kleinste Element m des Arrays, mit Position als Label.

- Die Wurzel hat zwei Kindknoten — die Wurzel des linken und die Wurzel des rechten Subarrays von A (jeweils von m aus).
- Rekursion.

Implementierung

Die Implementierung funktioniert so:

1. Partitioniere das Array in Blöcke der Größe s — jeder Block entspricht einem kartesischen Baum.
2. Berechne alle s^2 Möglichkeiten aller $\frac{1}{s+1} \binom{2s}{s}$ möglichen kartesischen Bäume der Größe s in einer Tabelle P . Diese benötigt $O(2^{2s} s^2)$ Speicher, also braucht P für $s = \frac{\log n}{4}$ nur $o(n)$ Speicher.
3. Berechne nach Lösung 2 die Struktur für das Array A' , welche aus den blockweisen Minima von A besteht ($O(n)$ Zeit, $O(n)$ Platz).

2.5 Lowest Common Ancestor

Gegeben sei ein Baum T mit Wurzel, $|T| = n$. Es sei

$$\text{LCA}_T(v, w)$$

der **lowest common ancestor** der Knoten v, w in T , also der Knoten $a \in T$, der

- Vorgänger von v und w ist und
- maximalen Abstand zur Wurzel hat.

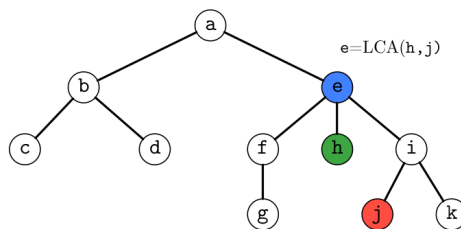


Abbildung 2.4. lowest common ancestor der Knoten h und j ist e .

Von RMQ zu LCA

Lemma 2.5.1. Gibt es eine $\langle f(n), g(n) \rangle$ -Lösung für RMQ, so gibt es eine Lösung in

$$\langle f(2n-1) + O(n), g(2n-1) + O(1) \rangle$$

für LCA.

Beweis. Sei

- T der kartesische Baum des Arrays A ,
- $E[0, \dots, 2n-2]$ das Knoten-Array, die in einer DFS-Euler-Tour von T besucht wurden,
- $L[0, \dots, 2n-2]$ die entsprechende Tiefe der Knoten in E ,
- $R[0, \dots, n-1]$ ein Array mit $R[i] = \min \{j : E[j] = i\}$ für jeden Knoten $i \in T$, also wo in der Euler-Tour der Knoten i das erste Mal auftaucht.

Dann ist

$$\text{LCA}_T(v, w) = E[\text{RMQ}_L(\min \{R[v], R[w]\}, \max \{R[v], R[w]\})].$$

□

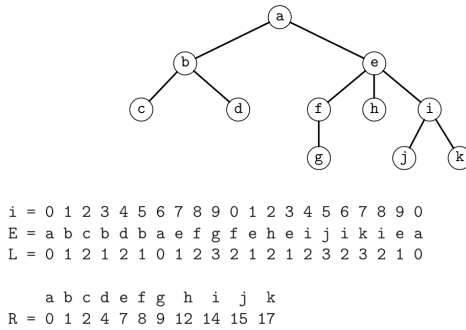


Abbildung 2.5. Konstruktion der Hilfsarrays für LCA.

Achtung: Die Tiefe zweier Knoten, die nacheinander bei der Euler-Tour besucht wurden, kann sich höchstens um 1 unterscheiden, also

$$(L[i] - L[i+1]) \in \{-1, 1\}.$$

Wir müssen also nur RMQs über Arrays lösen, bei denen sich zwei nacheinanderfolgende Elemente um 1 unterscheiden. Das ist das sogenannte **± 1 -RMQ**.

Wir werden nun im Folgenden ausnutzen, dass wir zum Lösen des LCA-Problems nur ± 1 -RMQs betrachten müssen.

LCA in $\langle O(n), O(1) \rangle$ auf $4n + o(n)$ Bits

Wir verwenden folgende Konstruktion:

Wir können nun den RMQ-Wert folgendermaßen berechnen:

```

RMQ (A, l, r)
  lpos := select(l + 1, (), BPext)
  rpos := select(r + 1, (), BPext)
  return rank(rmqexcess±1(lpos, rpos + 1), (), BPext)

```

LCA in $\langle O(n), O(1) \rangle$ auf $2n + o(n)$ Bits

Die Anzahl an benötigten Bits lässt sich im Vergleich zum vorhergehenden Ansatz noch weiter verkleinern. Wir transformieren dazu den kartesischen Baum — der ja ein Binärbaum ist — in einen allgemeinen Baum. Dazu gehen wir folgendermaßen vor:

1. Füge einen Elternknoten zur Wurzel des Baums hinzu (der hinzugefügte Knoten ist also die neue Wurzel).
2. Nehme von jedem Knoten v — von der neuen Wurzel ausgehend — den rechten Kindknoten w , und füge alle Knoten auf dem linken Pfad von w ausgehend zu den Kindknoten von v hinzu.

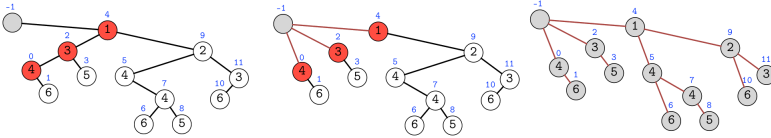


Abbildung 2.7. Kartesischer Baum mit neuer Wurzel. Die roten Knoten in der ersten Grafik sind die Knoten, die zu v — hier die neue Wurzel — hinzugefügt werden. Durch Hinzufügen dieser Knoten zu den Kindknoten von v hat der Baum die Struktur in Grafik 2. Nun wird dieser Prozess rekursiv auf die drei roten Knoten ausgeführt. Ist die Rekursion vollständig abgeschlossen sieht der Baum wie in der dritten Grafik aus.

Diese Transformation lässt sich bei Bedarf auch rückgängig machen; man kann also, wenn nötig, wieder den Binärbaum konstruieren.

Nun lässt sich wieder die Klammernreihe BP von oben bauen. Mit dieser können wir nun RMQ-Abfragen lösen:

```

RMQ (A, l, r)
  lpos := select(l + 2, (), BP)
  rpos := select(r + 2, (), BP)
  return rank(rmqexcess±1(lpos - 1, rpos), (), BP) - 1

```