

Grundlagen der Sequenzanalyse

Winter 2013/14

– Selbstevaluation –

Prof. Dr. Jens Stoye

Technische Fakultät

28. Januar 2014

Sequenzanalyse WS 2013/14

Vorbemerkung

Die folgenden Fragen erheben weder einen Anspruch auf Vollständigkeit, noch ist ein Auftauchen der Fragen ein Indikator für deren Auftretenswahrscheinlichkeit in der Klausur.

– Im positiven, wie im negativen Sinne.

Viel Spaß!

Sequenzanalyse WS 2013/14

Frage 1: Oh-Notation

Welche der folgenden Aussagen über Komplexitätsklassen ist nicht korrekt?

- A $\mathcal{O}(n \log n)$ ist polynomiell
- B $\mathcal{O}(n^2) \subset \mathcal{O}(2^n)$
- C $\mathcal{O}(n) \subset \mathcal{O}(\log n)$
- D $\mathcal{O}(\sqrt{n}) \subset \mathcal{O}(n \log n)$

Sequenzanalyse WS 2013/14

Frage 1: Oh-Notation

Welche der folgenden Aussagen über Komplexitätsklassen ist nicht korrekt?

- A $\mathcal{O}(n \log n)$ ist polynomiell
- B $\mathcal{O}(n^2) \subset \mathcal{O}(2^n)$
- C $\mathcal{O}(n) \subset \mathcal{O}(\log n)$
- D $\mathcal{O}(\sqrt{n}) \subset \mathcal{O}(n \log n)$

Sequenzanalyse WS 2013/14

Frage 2: Sequenzen

Die drei Worte $x = \text{ABA}$ und $y = \text{BABA}$ und $z = \text{BAA}$ sind gegeben. Welche der folgenden Aussagen ist korrekt?

- A y ist Teilwort von x
- B z ist Teilsequenz von y
- C x ist Präfix von y
- D z ist Teilsequenz von x

Sequenzanalyse WS 2013/14

Frage 2: Sequenzen

Die drei Worte $x = \text{ABA}$ und $y = \text{BABA}$ und $z = \text{BAA}$ sind gegeben. Welche der folgenden Aussagen ist korrekt?

- A y ist Teilwort von x
- B z ist Teilsequenz von y
- C x ist Präfix von y
- D z ist Teilsequenz von x

Sequenzanalyse WS 2013/14

Frage 3: Alphabete und Sequenzen

Welche Menge enthält genau alle nicht-leeren Sequenzen über dem Alphabet Σ ?

- A Σ^*
- B Σ^+
- C Σ^1
- D Σ^n

Sequenzanalyse WS 2013/14

Frage 3: Alphabete und Sequenzen

Welche Menge enthält genau alle nicht-leeren Sequenzen über dem Alphabet Σ ?

- A Σ^*
- B Σ^+
- C Σ^1
- D Σ^n

Sequenzanalyse WS 2013/14

Frage 4: Metrik

Welche Variante der Maximal-Matches Distanz ist eine Metrik?

A $d(x, y) := \delta(x||y) + \delta(y||x)$

B $d(x, y) := \log(\delta(x||y))$

C $d(x, y) := \delta(x||y) + \frac{1}{\delta(y||x)}$

D $d(x, y) := \log(\delta(x||y) + 1) + \log(\delta(y||x) + 1)$

Sequenzanalyse WS 2013/14

Frage 4: Metrik

Welche Variante der Maximal-Matches Distanz ist eine Metrik?

A $d(x, y) := \delta(x||y) + \delta(y||x)$

B $d(x, y) := \log(\delta(x||y))$

C $d(x, y) := \delta(x||y) + \frac{1}{\delta(y||x)}$

D $d(x, y) := \log(\delta(x||y) + 1) + \log(\delta(y||x) + 1)$

Sequenzanalyse WS 2013/14

Frage 5: Sequenzvergleich

Wie viele Teilworte der Länge q enthält ein Wort der Länge n ?

A n/q

B $n - q + 1$

C $\lfloor n/q \rfloor - (q + 1)$

D $n + q - 1$

Sequenzanalyse WS 2013/14

Frage 5: Sequenzvergleich

Wie viele Teilworte der Länge q enthält ein Wort der Länge n ?

A n/q

B $n - q + 1$

C $\lfloor n/q \rfloor - (q + 1)$

D $n + q - 1$

Sequenzanalyse WS 2013/14

Frage 6: Sequenzvergleich

Wofür verwendet man die q -Gramm-Distanz und die Maximal-Matches-Distanz in der Praxis?

- A um den Compute-Cluster zu beschäftigen
- B um die Biologen zu beschäftigen
- C um nicht immer die Edit-Distanz berechnen zu müssen
- D um die Edit-Distanz exakt zu berechnen

Sequenzanalyse WS 2013/14

Frage 6: Sequenzvergleich

Wofür verwendet man die q -Gramm-Distanz und die Maximal-Matches-Distanz in der Praxis?

- A um den Compute-Cluster zu beschäftigen
- B um die Biologen zu beschäftigen
- C um nicht immer die Edit-Distanz berechnen zu müssen
- D um die Edit-Distanz exakt zu berechnen

Sequenzanalyse WS 2013/14

Frage 7: Edit-Operationen

Welche Edit-Operationen sind zur Berechnung der *longest common subsequence* mit Hilfe der Edit-Distanz erlaubt?

- A $C, \mathcal{I}_C, \mathcal{D}$
- B $C, \mathcal{S}_C, \mathcal{F}$
- C $\mathcal{S}_C, \mathcal{D}, \mathcal{F}$
- D $C, \mathcal{S}_C, \mathcal{I}_C$

Sequenzanalyse WS 2013/14

Frage 7: Edit-Operationen

Welche Edit-Operationen sind zur Berechnung der *longest common subsequence* mit Hilfe der Edit-Distanz erlaubt?

- A C, I_C, D
- B C, S_C, F
- C S_C, D, F
- D C, S_C, I_C

Sequenzanalyse WS 2013/14

Frage 8: Edit-Distanz

Was gehört an die fehlende Stelle in der Rekursion zur Berechnung der Edit-Distanz von x und y :

$$D(i, j) = \min(\text{_____}, D(i - 1, j) + 1, D(i, j - 1) + 1)$$

- A $D(j - 1, i - 1) + w(x[j], y[i])$
- B $D(i - 1, i - 1) - w(x[i], y[j])$
- C $D(i - 1, j - 1) + w(x[i], y[j])$
- D $D(i - 1, j - 1) + w(x[i - 1], y[j - 1])$

Sequenzanalyse WS 2013/14

Frage 8: Edit-Distanz

Was gehört an die fehlende Stelle in der Rekursion zur Berechnung der Edit-Distanz von x und y :

$$D(i, j) = \min(\text{_____}, D(i-1, j) + 1, D(i, j-1) + 1)$$

- A $D(j-1, i-1) + w(x[j], y[i])$
- B $D(i-1, i-1) - w(x[i], y[j])$
- C $D(i-1, j-1) + w(x[i], y[j])$
- D $D(i-1, j-1) + w(x[i-1], y[j-1])$

Sequenzanalyse WS 2013/14

Frage 9: Scores und Kosten

Was ist der Vorteil von Scores gegenüber Kosten?

- A Scores sind schneller zu berechnen
- B übereinstimmende Teilworte am Anfang und am Ende werden leichter gefunden
- C lokale Alignments sind für Kosten nicht definiert
- D es gibt keinen Vorteil

Sequenzanalyse WS 2013/14

Frage 9: Scores und Kosten

Was ist der Vorteil von Scores gegenüber Kosten?

- A Scores sind schneller zu berechnen
- B übereinstimmende Teilworte am Anfang und am Ende werden leichter gefunden
- C lokale Alignments sind für Kosten nicht definiert
- D es gibt keinen Vorteil

Sequenzanalyse WS 2013/14

Frage 10: Alignment-Definition

Welches Alignment ist ungütig?

A $\begin{pmatrix} - & A & A \\ B & C & - \end{pmatrix}$

B $\begin{pmatrix} - & A & A \\ B & - & C \end{pmatrix}$

C $\begin{pmatrix} - & A & - & A \\ B & - & C & - \end{pmatrix}$

D $\begin{pmatrix} - & A & - & A \\ B & C & - & - \end{pmatrix}$

Sequenzanalyse WS 2013/14

Frage 10: Alignment-Definition

Welches Alignment ist ungütig?

A $\begin{pmatrix} - & A & A \\ B & C & - \end{pmatrix}$

B $\begin{pmatrix} - & A & A \\ B & - & C \end{pmatrix}$

C $\begin{pmatrix} - & A & - & A \\ B & - & C & - \end{pmatrix}$

D $\begin{pmatrix} - & A & - & A \\ B & C & - & - \end{pmatrix}$

Sequenzanalyse WS 2013/14

Frage 11: Alignment

Mit welcher Technik kann man das Alignmentproblem effizient und exakt lösen?

- A** Dynamisches Annealing
- B** Dynamische Programmierung
- C** Dynamische Interpolation
- D** Dynamisches Verbessern

Sequenzanalyse WS 2013/14

Frage 11: Alignment

Mit welcher Technik kann man das Alignmentproblem effizient und exakt lösen?

- A Dynamisches Annealing
- B Dynamische Programmierung**
- C Dynamische Interpolation
- D Dynamisches Verbessern

Sequenzanalyse WS 2013/14

Frage 12: Affine Gapkosten

Welche der folgenden Funktionen beschreibt affine Gapkosten für ein Gap der Länge ℓ ?

A $g(\ell) = d + \ell$

B $g(\ell) = d \cdot (\ell - 1)$

C $g(\ell) = d + (\ell - 1) \cdot e$

D $g(\ell) = d + g(\ell - 1) \cdot e$

Sequenzanalyse WS 2013/14

Frage 12: Affine Gapkosten

Welche der folgenden Funktionen beschreibt affine Gapkosten für ein Gap der Länge ℓ ?

A $g(\ell) = d + \ell$

B $g(\ell) = d \cdot (\ell - 1)$

C $g(\ell) = d + (\ell - 1) \cdot e$

D $g(\ell) = d + g(\ell - 1) \cdot e$

Sequenzanalyse WS 2013/14

Frage 13: Suboptimale Alignments

Mit welchem Algorithmus kann man suboptimale, nichtüberlappende Alignments finden?

- A** mit dem Smith-Kline-Beecham Algorithmus
- B** mit dem Knuth-Morris-Pratt Algorithmus
- C** mit dem Ford-Fulkerson Algorithmus
- D** mit dem Waterman-Eggert Algorithmus

Sequenzanalyse WS 2013/14

Frage 13: Suboptimale Alignments

Mit welchem Algorithmus kann man suboptimale, nichtüberlappende Alignments finden?

- A mit dem Smith-Kline-Beecham Algorithmus
- B mit dem Knuth-Morris-Pratt Algorithmus
- C mit dem Ford-Fulkerson Algorithmus
- D mit dem Waterman-Eggert Algorithmus**

Sequenzanalyse WS 2013/14

Frage 14: Alignmentgraph

Was für ein Graph ist der Alignmentgraph?

- A** gewichtet, gerichtet und unvollständig verbunden
- B** gewichtet, gerichtet und vollständig verbunden
- C** ungewichtet, gerichtet und vollständig verbunden
- D** ungewichtet, ungerichtet und unvollständig verbunden

Sequenzanalyse WS 2013/14

Frage 14: Alignmentgraph

Was für ein Graph ist der Alignmentgraph?

- A** gewichtet, gerichtet und unvollständig verbunden
- B** gewichtet, gerichtet und vollständig verbunden
- C** ungewichtet, gerichtet und vollständig verbunden
- D** ungewichtet, ungerichtet und unvollständig verbunden

Sequenzanalyse WS 2013/14

Frage 15: Alignmentgraph

Welche Aussage trifft auf den Alignmentgraph für *approximatives* Stringmatching zu?

- A** alle Knoten können vom Startknoten aus direkt erreicht werden
- B** alle Knoten der letzten Zeile können vom Startknoten aus direkt erreicht werden
- C** der Endknoten kann direkt vom Startknoten aus erreicht werden
- D** nur die Knoten der ersten Zeile können vom Startknoten aus direkt erreicht werden

Sequenzanalyse WS 2013/14

Frage 15: Alignmentgraph

Welche Aussage trifft auf den Alignmentgraph für *approximatives* Stringmatching zu?

- A alle Knoten können vom Startknoten aus direkt erreicht werden
- B alle Knoten der letzten Zeile können vom Startknoten aus direkt erreicht werden
- C der Endknoten kann direkt vom Startknoten aus erreicht werden
- D nur die Knoten der ersten Zeile können vom Startknoten aus direkt erreicht werden

Sequenzanalyse WS 2013/14

Frage 16: Alignmentgraph

Welche Aussage trifft auf den Alignmentgraph für *lokales* Alignment zu?

- A** alle Knoten können vom Startknoten aus direkt erreicht werden
- B** alle Knoten können vom Endknoten aus direkt erreicht werden
- C** der Endknoten kann direkt vom Startknoten aus erreicht werden
- D** nur die Knoten der ersten Zeile können vom Startknoten aus direkt erreicht werden

Sequenzanalyse WS 2013/14

Frage 16: Alignmentgraph

Welche Aussage trifft auf den Alignmentgraph für *lokales* Alignment zu?

- A** alle Knoten können vom Startknoten aus direkt erreicht werden
- B** alle Knoten können vom Endknoten aus direkt erreicht werden
- C** der Endknoten kann direkt vom Startknoten aus erreicht werden
- D** nur die Knoten der ersten Zeile können vom Startknoten aus direkt erreicht werden

Sequenzanalyse WS 2013/14

Frage 17: Paarweises Sequenzalignment

Wozu verwendet man Backtracking beim Sequenzalignment?

- A** Finden einer optimalen Editsequenz
- B** Finden von suboptimalen, verdeckten Editsequenzen
- C** Finden einer echt verschiedenen, suboptimalen Editsequenz
- D** Finden aller optimalen Editsequenzen

Sequenzanalyse WS 2013/14

Frage 17: Paarweises Sequenzalignment

Wozu verwendet man Backtracking beim Sequenzalignment?

- A Finden einer optimalen Editsequenz
- B Finden von suboptimalen, verdeckten Editsequenzen
- C Finden einer echt verschiedenen, suboptimalen Editsequenz
- D Finden aller optimalen Editsequenzen**

Sequenzanalyse WS 2013/14

Frage 18: Paarweises Sequenzalignment

In welcher Komplexitätsklasse liegt die Laufzeit des Smith-Waterman-Algorithmus für lokales Sequenzalignment (zwei Sequenzen der Länge n , homogene Gapkosten)?

- A $\mathcal{O}(n^3)$
- B $\mathcal{O}(2^n)$
- C $\mathcal{O}(n^2)$
- D $\mathcal{O}(n \log n)$

Sequenzanalyse WS 2013/14

Frage 18: Paarweises Sequenzalignment

In welcher Komplexitätsklasse liegt die Laufzeit des Smith-Waterman-Algorithmus für lokales Sequenzalignment (zwei Sequenzen der Länge n , homogene Gapkosten)?

- A $\mathcal{O}(n^3)$
- B $\mathcal{O}(2^n)$
- C $\mathcal{O}(n^2)$
- D $\mathcal{O}(n \log n)$

Sequenzanalyse WS 2013/14

Frage 19: Paarweises Sequenzalignment

In welcher Komplexitätsklasse liegt die Laufzeit des Smith-Waterman-Algorithmus für lokales Sequenzalignment (zwei Sequenzen der Länge n , generelle Gapkosten)?

- A $\mathcal{O}(n^3)$
- B $\mathcal{O}(2^n)$
- C $\mathcal{O}(n^2)$
- D $\mathcal{O}(n \log n)$

Sequenzanalyse WS 2013/14

Frage 19: Paarweises Sequenzalignment

In welcher Komplexitätsklasse liegt die Laufzeit des Smith-Waterman-Algorithmus für lokales Sequenzalignment (zwei Sequenzen der Länge n , generelle Gapkosten)?

- A $\mathcal{O}(n^3)$
- B $\mathcal{O}(2^n)$
- C $\mathcal{O}(n^2)$
- D $\mathcal{O}(n \log n)$

Sequenzanalyse WS 2013/14

Frage 20: Paarweises Sequenzalignment

In welcher Komplexitätsklasse liegt die Laufzeit des Gotoh-Algorithmus für globales Sequenzalignment (zwei Sequenzen der Länge n , affine Gapkosten)?

- A $\mathcal{O}(n^3)$
- B $\mathcal{O}(3^n)$
- C $\mathcal{O}(n \log n)$
- D $\mathcal{O}(n^2)$

Sequenzanalyse WS 2013/14

Frage 20: Paarweises Sequenzalignment

In welcher Komplexitätsklasse liegt die Laufzeit des Gotoh-Algorithmus für globales Sequenzalignment (zwei Sequenzen der Länge n , affine Gapkosten)?

- A $\mathcal{O}(n^3)$
- B $\mathcal{O}(3^n)$
- C $\mathcal{O}(n \log n)$
- D $\mathcal{O}(n^2)$

Sequenzanalyse WS 2013/14

Frage 21: Paarweises Sequenzalignment

Betrachtet man den Alignmentgraphen des Smith-Waterman-Alignments, so hat dieser an jedem inneren Knoten ...

- A** ... drei eingehende und drei ausgehende Kanten
- B** ... vier eingehende und vier ausgehende Kanten
- C** ... drei eingehende und vier ausgehende Kanten
- D** ... vier eingehende und drei ausgehende Kanten

Sequenzanalyse WS 2013/14

Frage 21: Paarweises Sequenzalignment

Betrachtet man den Alignmentgraphen des Smith-Waterman-Alignments, so hat dieser an jedem inneren Knoten ...

- A ... drei eingehende und drei ausgehende Kanten
- B ... vier eingehende und vier ausgehende Kanten
- C ... drei eingehende und vier ausgehende Kanten
- D ... vier eingehende und drei ausgehende Kanten

Sequenzanalyse WS 2013/14

Frage 22: Paarweises Sequenzalignment

Betrachtet man den Alignmentgraphen des Needleman-Wunsch-Alignments, so hat dieser an jedem inneren Knoten ...

- A** ... drei eingehende und drei ausgehende Kanten
- B** ... vier eingehende und vier ausgehende Kanten
- C** ... drei eingehende und vier ausgehende Kanten
- D** ... vier eingehende und drei ausgehende Kanten

Sequenzanalyse WS 2013/14

Frage 22: Paarweises Sequenzalignment

Betrachtet man den Alignmentgraphen des Needleman-Wunsch-Alignments, so hat dieser an jedem inneren Knoten ...

- A** ... drei eingehende und drei ausgehende Kanten
- B** ... vier eingehende und vier ausgehende Kanten
- C** ... drei eingehende und vier ausgehende Kanten
- D** ... vier eingehende und drei ausgehende Kanten

Sequenzanalyse WS 2013/14

Frage 23: Dotplots

Ein Dotplot wird häufig verwendet, um ...

- A** ... Alignments zu visualisieren
- B** ... die Edit-Distanz zu berechnen
- C** ... lokale Ähnlichkeiten zwischen Sequenzen zu visualisieren
- D** ... den Maximum-Likelihood-Score darzustellen

Sequenzanalyse WS 2013/14

Frage 23: Dotplots

Ein Dotplot wird häufig verwendet, um ...

- A ... Alignments zu visualisieren
- B ... die Edit-Distanz zu berechnen
- C ... lokale Ähnlichkeiten zwischen Sequenzen zu visualisieren
- D ... den Maximum-Likelihood-Score darzustellen

Sequenzanalyse WS 2013/14

Frage 24: q -Gramm Lemma

Wie viele gemeinsame q -Gramme der Länge 3 kommen in einem lokalen Alignment der Länge 9 *mindestens* vor, wenn dort bis zu 2 Fehler erlaubt sind?

- A 1
- B 2
- C 3
- D 42

Sequenzanalyse WS 2013/14

Frage 24: q -Gramm Lemma

Wie viele gemeinsame q -Gramme der Länge 3 kommen in einem lokalen Alignment der Länge 9 *mindestens* vor, wenn dort bis zu 2 Fehler erlaubt sind?

- A 1
- B 2
- C 3
- D 42

Sequenzanalyse WS 2013/14

Frage 25: Schnelle Datenbanksuche

Was ist der Unterschied zwischen *on-line*- und indexbasierter Datenbanksuche?

- A indexbasiert ist bei wenigen Anfragen schneller als *on-line*
- B indexbasiert ist bei vielen Anfragen schneller als *on-line*
- C indexbasiert benötigt weniger Vorverarbeitung als *on-line*
- D weiß nicht

Sequenzanalyse WS 2013/14

Frage 25: Schnelle Datenbanksuche

Was ist der Unterschied zwischen *on-line*- und indexbasierter Datenbanksuche?

- A indexbasiert ist bei wenigen Anfragen schneller als *on-line*
- B indexbasiert ist bei vielen Anfragen schneller als *on-line*
- C indexbasiert benötigt weniger Vorverarbeitung als *on-line*
- D weiß nicht

Sequenzanalyse WS 2013/14

Frage 26: Schnelle Datenbanksuche

Welche Zeitkomplexität hat die *effiziente* Erstellung eines q -Gramm-Index für ein Wort der Länge m ?

- A $\mathcal{O}(|\Sigma|^q + m)$
- B $\mathcal{O}(|\Sigma|^q + qm)$
- C $\mathcal{O}(|\Sigma|^q + m^q)$
- D $\mathcal{O}(|\Sigma|^q + q^m)$

Sequenzanalyse WS 2013/14

Frage 26: Schnelle Datenbanksuche

Welche Zeitkomplexität hat die *effiziente* Erstellung eines q -Gramm-Index für ein Wort der Länge m ?

- A $\mathcal{O}(|\Sigma|^q + m)$
- B $\mathcal{O}(|\Sigma|^q + qm)$
- C $\mathcal{O}(|\Sigma|^q + m^q)$
- D $\mathcal{O}(|\Sigma|^q + q^m)$

Sequenzanalyse WS 2013/14

Frage 27: Wort-Wahrscheinlichkeiten

Was ist die Wahrscheinlichkeit, das Wort `GAGA` in einer DNA-Sequenz x der Länge 18 an Position 5 zu finden? (Die Zeichen von x seien über dem DNA-Alphabet uniform i.i.d.-verteilt.)

- A 0.000390625
- B $1/256$
- C 0.0625
- D $1/4$

Sequenzanalyse WS 2013/14

Frage 27: Wort-Wahrscheinlichkeiten

Was ist die Wahrscheinlichkeit, das Wort `GAGA` in einer DNA-Sequenz x der Länge 18 an Position 5 zu finden? (Die Zeichen von x seien über dem DNA-Alphabet uniform i.i.d.-verteilt.)

- A 0.000390625
- B 1/256**
- C 0.0625
- D 1/4

Sequenzanalyse WS 2013/14

Frage 28: Index-Datenstrukturen

Welche der folgenden ist keine Index-Datenstruktur für die exakte Textsuche:

- A Suffixbaum
- B Burrows-Wheeler-Transformierte
- C Suffix-Array
- D Bipartiter Graph

Sequenzanalyse WS 2013/14

Frage 28: Index-Datenstrukturen

Welche der folgenden ist keine Index-Datenstruktur für die exakte Textsuche:

- A Suffixbaum
- B Burrows-Wheeler-Transformierte
- C Suffix-Array
- D **Bipartiter Graph**

Sequenzanalyse WS 2013/14

Frage 29: Suffixbaum-Definition

Was ist ein *suffix trie*?

- A das selbe wie ein Suffixbaum
- B ein reverser Präfixbaum
- C ein Σ^+ -Baum
- D ein kompakter Σ^+ -Baum

Sequenzanalyse WS 2013/14

Frage 29: Suffixbaum-Definition

Was ist ein *suffix trie*?

- A das selbe wie ein Suffixbaum
- B ein reverser Präfixbaum
- C ein Σ^+ -Baum
- D ein kompakter Σ^+ -Baum

Sequenzanalyse WS 2013/14

Frage 30: Suffixbaum-Konstruktion

Welche Laufzeit hat der Ukkonen-Algorithmus zur *online*-Konstruktion von Suffixbäumen?

- A $\mathcal{O}(n^2)$
- B $\mathcal{O}(n)$
- C $\mathcal{O}(n \log n)$
- D $\mathcal{O}(2^n)$

Sequenzanalyse WS 2013/14

Frage 30: Suffixbaum-Konstruktion

Welche Laufzeit hat der Ukkonen-Algorithmus zur *online*-Konstruktion von Suffixbäumen?

- A $\mathcal{O}(n^2)$
- B $\mathcal{O}(n)$
- C $\mathcal{O}(n \log n)$
- D $\mathcal{O}(2^n)$

Sequenzanalyse WS 2013/14

Frage 31: WOTD

Wofür steht die Abkürzung WOTD?

- A walking out the door
- B write-only top-down
- C word of the day
- D word and tree datastructure

Sequenzanalyse WS 2013/14

Frage 31: WOTD

Wofür steht die Abkürzung WOTD?

- A walking out the door
- B write-only top-down
- C word of the day
- D word and tree datastructure

Sequenzanalyse WS 2013/14

Frage 32: Suffixbaum

Welches Problem kann mit einem Suffixbaum nicht so einfach gelöst werden?

- A** Finden von MUMs und MEMs
- B** Berechnen von multiplen Alignment
- C** Exakte Textsuche
- D** Finden des kürzesten gemeinsamen Teilworts

Sequenzanalyse WS 2013/14

Frage 32: Suffixbaum

Welches Problem kann mit einem Suffixbaum nicht so einfach gelöst werden?

- A Finden von MUMs und MEMs
- B Berechnen von multiplen Alignment**
- C Exakte Textsuche
- D Finden des kürzesten gemeinsamen Teilworts

Sequenzanalyse WS 2013/14

Frage 33: Suffixarrays

Was war die Hauptmotivation für die Einführung von Suffix-Arrays?

- A** sie benötigen weniger Speicher als Suffixbäume
- B** sie sind flexibler als Suffixbäume
- C** sie sind hübscher als Suffixbäume
- D** sie erlauben schnellere Textsuche als Suffixbäume

Sequenzanalyse WS 2013/14

Frage 33: Suffixarrays

Was war die Hauptmotivation für die Einführung von Suffix-Arrays?

- A** sie benötigen weniger Speicher als Suffixbäume
- B** sie sind flexibler als Suffixbäume
- C** sie sind hübscher als Suffixbäume
- D** sie erlauben schnellere Textsuche als Suffixbäume

Sequenzanalyse WS 2013/14

Frage 34: BWT

Welche Kompression eignet sich besonders gut im Anschluss an die Burrows-Wheeler-Transformation?

- A Lempel-Ziv
- B Run-length Encoding
- C JPEG 2000
- D Huffman Encoding

Sequenzanalyse WS 2013/14

Frage 34: BWT

Welche Kompression eignet sich besonders gut im Anschluss an die Burrows-Wheeler-Transformation?

- A Lempel-Ziv
- B Run-length Encoding**
- C JPEG 2000
- D Huffman Encoding

Sequenzanalyse WS 2013/14

Frage 35: Hirschberg-Algorithmus

Welchem Schema folgt der Hirschberg-Algorithmus?

- A Divide and Conquer
- B Greedy
- C Linear Programming
- D Branch and Bound

Sequenzanalyse WS 2013/14

Frage 35: Hirschberg-Algorithmus

Welchem Schema folgt der Hirschberg-Algorithmus?

- A Divide and Conquer
- B Greedy
- C Linear Programming
- D Branch and Bound

Sequenzanalyse WS 2013/14

Frage 36: Multiples Alignment

Über wie viele Vorgänger muss bei der Berechnung einer Zelle der 3-dimensionalen Edit-Matrix beim multiplen Alignment von drei Sequenzen minimiert werden?

- A 3
- B 7
- C 8
- D 15

Sequenzanalyse WS 2013/14

Frage 36: Multiples Alignment

Über wie viele Vorgänger muss bei der Berechnung einer Zelle der 3-dimensionalen Edit-Matrix beim multiplen Alignment von drei Sequenzen minimiert werden?

- A 3
- B 7
- C 8
- D 15

Sequenzanalyse WS 2013/14

Frage 37: NP-Vollständigkeit

Wann ist ein Problem in der Klasse NP?

- A wenn es schwerer zu lösen ist als SAT
- B wenn es in polynomieller Zeit gelöst werden kann
- C wenn ein Zertifikat in polynomieller Zeit berechnet werden kann
- D wenn ein Zertifikat in polynomieller Zeit verifiziert werden kann

Sequenzanalyse WS 2013/14

Frage 37: NP-Vollständigkeit

Wann ist ein Problem in der Klasse NP?

- A wenn es schwerer zu lösen ist als SAT
- B wenn es in polynomieller Zeit gelöst werden kann
- C wenn ein Zertifikat in polynomieller Zeit berechnet werden kann
- D wenn ein Zertifikat in polynomieller Zeit verifiziert werden kann

Sequenzanalyse WS 2013/14

Frage 38: NP-Vollständigkeit

Was sollte man nicht tun, wenn man ein NP-hartes Problem lösen will?

- A einen Approximationsalgorithmus suchen
- B das Problem mit *branch and bound* lösen
- C den Kopf in den Sand stecken
- D eine gute Korrektheitsheuristik entwickeln

Sequenzanalyse WS 2013/14

Frage 38: NP-Vollständigkeit

Was sollte man nicht tun, wenn man ein NP-hartes Problem lösen will?

- A einen Approximationsalgorithmus suchen
- B das Problem mit *branch and bound* lösen
- C **den Kopf in den Sand stecken**
- D eine gute Korrektheitsheuristik entwickeln

Sequenzanalyse WS 2013/14

Frage 39: Carrillo-Lipman

Was für eine Methode ist der Carrillo-Lipman-Algorithmus?

- A Approximationsverfahren
- B Laufzeitheuristik
- C FPT-Algorithmus
- D Korrektheitsheuristik

Sequenzanalyse WS 2013/14

Frage 39: Carrillo-Lipman

Was für eine Methode ist der Carrillo-Lipman-Algorithmus?

- A Approximationsverfahren
- B Laufzeitheuristik**
- C FPT-Algorithmus
- D Korrektheitsheuristik

Sequenzanalyse WS 2013/14

Frage 40: Center-Star

Welchen Approximationsfaktor hat die Center-Star-Methode?

- A 1
- B $1 + \epsilon$
- C 1.5
- D 2

Sequenzanalyse WS 2013/14

Frage 40: Center-Star

Welchen Approximationsfaktor hat die Center-Star-Methode?

- A 1
- B $1 + \epsilon$
- C 1.5
- D 2

Sequenzanalyse WS 2013/14

Frage 41: Divide-and-Conquer Alignment

Wie sind C -optimale Schnitte definiert?

- A** sie minimieren die *multiple additional cost*
- B** sie zwingen das multiple Alignment auf einen optimalen Pfad
- C** sie schneiden die Sequenzen an ihrem linken Ende
- D** sie resultieren in einem optimalen multiplen Alignment

Sequenzanalyse WS 2013/14

Frage 41: Divide-and-Conquer Alignment

Wie sind C -optimale Schnitte definiert?

- A** sie minimieren die *multiple additional cost*
- B** sie zwingen das multiple Alignment auf einen optimalen Pfad
- C** sie schneiden die Sequenzen an ihrem linken Ende
- D** sie resultieren in einem optimalen multiplen Alignment

Sequenzanalyse WS 2013/14

Frage 42: Baumalignment

Welcher Teilschritt des Sankoff-Algorithmus lässt sich effizient lösen?

- A Aufzählen aller Vorgänger eines Eintrags der Edit-Matrix
- B *Minimum Mutation Problem* (Fitch-Algorithmus)
- C Berechnen des optimalen Baumes
- D Berechnen der optimalen Alignmentkosten

Sequenzanalyse WS 2013/14

Frage 42: Baumalignment

Welcher Teilschritt des Sankoff-Algorithmus lässt sich effizient lösen?

- A Aufzählen aller Vorgänger eines Eintrags der Edit-Matrix
- B *Minimum Mutation Problem (Fitch-Algorithmus)*
- C Berechnen des optimalen Baumes
- D Berechnen der optimalen Alignmentkosten

Sequenzanalyse WS 2013/14

Frage 43: Genomalignment

Welche zwei Schritte werden beim Genomalignment durchgeführt?

- A *branch* und *bound*
- B *run* und *jump*
- C MUMs finden und MEMs finden
- D *filtration* und *chaining*

Sequenzanalyse WS 2013/14

Frage 43: Genomalignment

Welche zwei Schritte werden beim Genomalignment durchgeführt?

- A *branch* und *bound*
- B *run* und *jump*
- C MUMs finden und MEMs finden
- D *filtration* und *chaining*

Sequenzanalyse WS 2013/14

ENDE