

Übungen zu Ideen der Informatik

<https://www.mpi-inf.mpg.de/departments/algorithms-complexity/teaching/winter21/ideen/>

Blatt 5: Websuche

Abgabeschluss: 29.11.2021

Aufgabe 1 Vorkommenslisten (10 Punkte)

Betrachten Sie die folgenden Textdokumente:

- T_1 : Der Bundestag hat das neue Infektionsschutzgesetz gebilligt.
- T_2 : Der neue Bundestag hat mehr Abgeordnete als je zuvor.
- T_3 : Was soll das neue Infektionsschutzgesetz ändern?

- a) Erstellen Sie die Vorkommenslisten für die Wörter *Bundestag*, *Infektionsschutzgesetz*, *Abgeordnete*, *neue* und *ändern*. (5 Punkte)
- b) Für einen Text T und ein Wort w schreiben wir $w \in T$, falls das Wort w in T vorkommt, anderenfalls $w \notin T$, was gleichbedeutend ist mit $\neg(w \in T)$. Wir können solche Aussagen mit logischen Operatoren verknüpfen. Für Wörter $w = \text{Bundestag}$ und $v = \text{Infektionsschutzgesetz}$ sowie Texte $T = T_1$ und $S = T_2$ gilt beispielsweise:

$$(w \in T) \wedge (v \notin S).$$

Hinweis: Haben Sie die Symbole gerade nur überflogen? Lesen Sie den Einführungstext zu dieser Teilaufgabe nochmal und stellen Sie sicher, dass Sie die Notation verstanden haben.

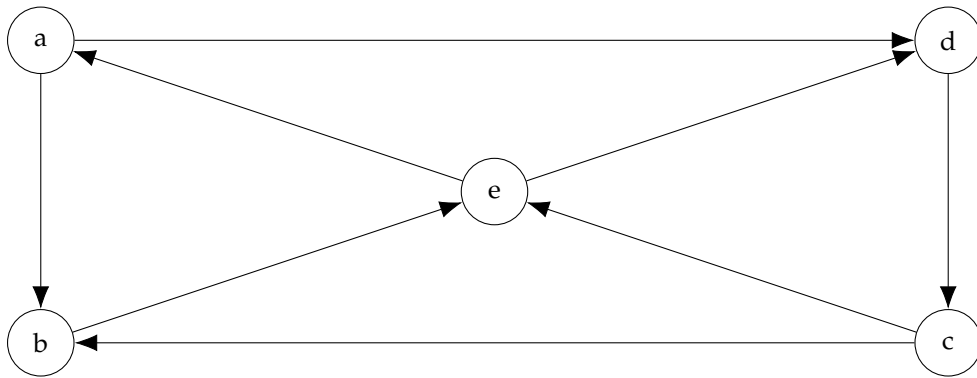
Beantworten Sie die folgenden Fragen (jeweils mit Begründung):

- (1) Für welche Texte $T \in \{T_1, T_2, T_3\}$ gilt $\text{Bundestag} \in T$? (1 Punkt)
 - (2) Für welche Wörter w der deutschen Sprache gilt $(w \in T_1) \wedge (w \in T_2)$? (1 Punkt)
 - (3) Für welche Texte $T \in \{T_1, T_2, T_3\}$ gilt $(\text{Abgeordnete} \in T) \vee (\text{ändern} \in T)$? (1 Punkt)
- c) In der Vorlesung wurde dargestellt, wie Sie durch Mischen von Vorkommenslisten Ergebnisse zu einer Suchanfrage finden können, die zwei Wörter kombiniert (z.B. *Barack* und *Obama*). Stellen Sie dar, wie Sie nach einem ähnlichen Prinzip Ergebnisse zu einer Suchanfrage finden können, die ein Wort w_1 enthält, ein anderes Wort w_2 aber explizit *nicht*. (2 Punkte)

Aufgabe 2 Pagerank (10 Punkte)

Betrachten Sie den folgenden Graphen. Die Knoten stehen für Webseiten und die Kanten für Verweise zwischen diesen Webseiten. Wir möchten für alle Knoten im Graphen ihre Relevanz bestimmen.

- a) Welche Probleme träten bei der Bestimmung der Relevanz nach dem Simulationsprinzip auf, wenn Sie den Pfeil (d,c) umdrehten (mit Begründung)? (1 Punkt)
- b) Stellen Sie das Relevanzgleichungssystem nach dem Prinzip aus der Veranstaltung auf und lösen Sie es. (6 Punkte)



- c) Sie betreiben neben dem Studium einen Online-Shop für bunte Socken und ärgern sich darüber, dass dieser in den Suchergebnissen einer prominenten Suchmaschine nur weit unten auftaucht. Nehmen Sie an, dass die Suchmaschine das Ranking von Suchergebnissen für bunte Socken allein über die in der Vorlesung eingeführte Variante des Pagerank bestimmt.
- (1) Würde es der Sichtbarkeit Ihres Online-Shops helfen, wenn die Universität des Saarlandes auf Ihrer Webseite einen Link zu Ihrem Online-Shop einrichtet (mit Begründung)? (1 Punkt)
 - (2) Ein findiger Kommilitone schlägt vor, Sie könnten in großem Stil Webseiten kaufen und mit diesen Webseiten auf Ihren Online-Shop verlinken. Kann der Vorschlag des Kommilitonen funktionieren (mit Begründung)? (1 Punkt)
 - (3) Würde sich Ihre Antwort zur vorigen Teilaufgabe ändern, wenn der Kommilitone vorschläge, die gekauften Webseiten auch untereinander zu verlinken (mit Begründung)? (1 Punkt)

Aufgabe 3 Soziale Netzwerke (10 Punkte)

In der Vorlesung haben wir besprochen, wie man das WWW als Graphen modellieren kann: mit einzelnen Webseiten als Knoten und Links zwischen den Webseiten als Kanten. Ähnlich kann man auch soziale Netzwerke (z.B. LinkedIn) modellieren: Hier sind die Personen die Knoten und eine Kante signalisiert, dass zwei Personen miteinander bekannt sind. Im Unterschied zum WWW haben die Kanten hier keine (öffentlich sichtbare) Richtung: Man sieht nur, dass zwei Personen sich kennen, aber nicht, wer wem eine Kontaktanfrage geschickt hat. Daher zeichnen wir die Kanten ohne Pfeile, etwa so (am Beispiel der Beziehung $\{Ada, Bob\}$):



Hinweis: Wenn eine Kante zwischen den Knoten u und v eine Richtung hat, notieren wir sie als (u, v) , wenn sie keine Richtung hat, als $\{u, v\}$. Das erste Konstrukt nennt sich Tupel, hier spielt die Reihenfolge der Elemente eine Rolle, also $(u, v) \neq (v, u)$. Das zweite Konstrukt nennt sich Menge, hier spielt die Reihenfolge der Elemente keine Rolle, d.h. $\{u, v\} = \{v, u\}$.

- a) Stellen Sie die folgende Liste von Kontakten aus einem sozialen Netzwerk als Graph dar (Sie können die Namen der Personen mit ihren Anfangsbuchstaben abkürzen). (2 Punkte)

Kontakte = $[\{Ana, Bob\}, \{Ana, David\}, \{Ana, Ela\}, \{Bob, Carl\}, \{Bob, Gina\},$
 $\{Carl, Fred\}, \{Carl, Gina\}, \{David, Ela\}, \{Ela, Fred\}, \{Ela, Gina\}]$

- b) Welcher Kontakt ist von *David* am weitesten entfernt (mit Begründung)? (1 Punkt)
- c) Welchen neuen Kontakt würden Sie *Fred* vorschlagen, wenn Sie ihm jemanden nennen wollen, den er voraussichtlich kennt (mit Begründung)? (1 Punkt)

- d) Jede Person im Netzwerk kann die Posts und Re-Posts der anderen Personen sehen, mit denen sie verbunden ist.
- (1) *David* hat etwas gepostet und hofft darauf, dass sein Post von allen Personen im Netzwerk gesehen wird. Wie häufig muss der Post mindestens geteilt werden, damit jede Person im Netzwerk die Chance hat, den Post zu sehen – und welche Person(en) müssten in diesem Fall den Post teilen? (2 Punkte)
 - (2) Wie häufig müsste ein Post von *Gina* mindestens geteilt werden, damit jede Person im Netzwerk die Chance hat, den Post zu sehen – und welche Person(en) müssten in diesem Fall den Post teilen? (1 Punkt)
 - (3) Nun abstrakt formuliert: Wenn in *irgendeinem* sozialen Netzwerk *irgendwer* postet. . .
 - (i) . . . welche Personen haben eine Chance, den Post zu sehen (d.h. in wessen Newsfeeds taucht der Post auf)? (1 Punkt)
 - (ii) . . . welche Personen haben *keine* Chance, den Post zu sehen (d.h. in wessen Newsfeeds taucht der Post *nicht* auf)? (1 Punkt)
 - (iii) . . . was muss demnach gegeben sein, damit alle Personen im Netzwerk eine Chance haben, den Post zu sehen? (1 Punkt)

Ich habe für die Videos, die Nachbereitung und das Übungsblatt etwa Stunden gebraucht.
(Ann-Sophie fertigt aus diesen Zahlen eine Statistik an. Kurt und Corinna sehen nur diese Statistik. Wir möchten wissen, ob der Schwierigkeitsgrad in etwa richtig ist.)

Websuche war spannend okay langweilig
schwierig okay einfach