

Übungsblatt 10

21. Januar 2009

Hinweis: Soweit nicht anders angegeben, gibt es für jede korrekt bearbeitete Teilaufgabe einen Punkt. Die Abgabe der Hausübungen ist bis spätestens zum Beginn der nächsten Vorlesung möglich – entweder persönlich direkt vor der Vorlesung oder per Einwurf in den Briefkasten des Instituts (Informatikzentrum, zweiter Stock, vor Raum 238).

Aufgabe 23 (Der Web-Graph)

- a) Eine Webseite hat In-Degree k , wenn sie genau k In-Links besitzt. Man hat festgestellt, daß die Anzahl aller Webseiten mit In-Degree k proportional zu $\frac{1}{k^{2,1}}$ ist.
Wie groß ist die Wahrscheinlichkeit dafür, daß eine zufällig gewählte Webseite einen In-Degree von 1 besitzt? Wie groß ist der In-Degree einer Webseite im Mittel (Erwartungswert)? (2 Punkte)
- b) Nehmen wir an, daß wir durch einen Webcrawl einen Teil des Webs lokal vorliegen haben. Durch Analyse des zugehörigen Web-Graphen stellen wir fest, daß der mittlere In-Degree 9 beträgt. Welche Aussagen können wir nun über den mittleren Out-Degree treffen?

Aufgabe 24 (Größenschätzung)

- a) Seien A und B zwei Websuchmaschinen. Nehmen wir an, daß wir durch geschickte Anfragestellung an die beiden Suchmaschinen in der Lage waren, jeweils eine große (gleichverteilte) Stichprobe aus den zugehörigen Indexen zu ziehen. Bei der Analyse dieser beiden Stichproben stellen wir fest, daß 30% der von A indexierten Seiten auch von B indexiert wurden. Außerdem wurden 50% der von B indexierten Seiten auch von A indexiert. Wie groß ist der Index von Suchmaschine A (relativ gesehen zur Größe von Bs Index)?
- b) Durch Analyse einer repräsentativen Stichprobe des Webs wissen wir, daß Term a in etwa 0,5% aller Dokumente auftritt, Term b in etwa 0,07% und Term c in etwa 2,2%.
Eine Suchmaschine liefert auf die Anfrage „ a “ rund 28 Millionen Treffer, auf die Anfrage „ b “ rund 5 Millionen Treffer und auf die Anfrage „ c “ rund 135 Millionen Treffer. Schätzen Sie, wie groß der Index dieser Suchmaschine ist! (2 Punkte)