

# Digital Libraries WS 2018/2019

## Übungsblatt 1

Aaron Winziers - 1176638; Michael Wolz - 1195270

5. November 2018

### Aufgabe 1

### Aufgabe 2

(a)

Über die Anzahl der Publikationen und Zitate sagt der h-Index nur sehr wenig aus. Lediglich dass der Autor mindestens 10 Veröffentlichungen hat die alle mindestens zehn mal zitiert wurden sind.

Die folgende Tabelle listet zwei Beispiele auf die beide einen h-Index von 10 hätten:

Publikation #	Zitate	Zitate
1	10	100
2	10	254
3	10	156
4	10	620
5	10	148
6	10	1024
7	10	415
8	10	184
9	10	220
10	10	784
11	0	-
12	0	-

Hier sind die Zahlen der Zitate zwar sehr extrem und unwahrscheinlich aber trotzdem wird illustriert dass der h-Index eines Autors alleine nicht stets viel aussagend ist.

**(b)**

Die verschiedenen Fachgebiete könnten problematisch sein, da es Unterschiede in der Größe des Forschungsgebiets gibt. Wenn es weniger Forscher in einem Gebiet gibt, kann es sein, dass deren Forschung sehr spezialisiert ist und daher weniger zitiert wird.

Das Alter kann auch Einfluss auf die Wertlegung auf den h-Index haben. Erstens kann der h-Index eines jüngeren Autors nicht so hoch sein wie ein älterer, da er noch nicht so viele Publikationen hat, und je länger eine Publikation veröffentlicht ist, desto wahrscheinlicher ist es, dass sie eine hohe Anzahl an Zitaten hat. Hierdurch hat der jüngere Kandidat einen starken Nachteil, weil der h-Index nur unmöglich „besser“ sein kann als sein älterer Kollege.

**(c)**

1. Wo sie publiziert haben

Wichtig ist, nicht nur wie viel, sondern auch bei welchen Organisationen ein Kandidat publiziert hat, bzw. an welchen Konferenzen er/sie teilgenommen hat. Der Ruf einer Organisation ist zwar meist in deren Gebieten bekannt, jedoch ist dies nicht quantifizierbar.

2. Frequenz der Publikationen

Die Häufigkeit, mit der ein Autor oder Kandidat publiziert, kann als Kriterium verwendet werden und ist quantifizierbar.

3. Relevanz der Forschung zur Ausschreibung

Bei einer Berufung muss es natürlich erst eine Ausschreibung geben, die detailliert angibt, welcher Bereich die Professur abdecken soll und ob ein Kandidat auf diese Beschreibung passt, kann durch seine Publikationen eingesehen werden. Dieses Kriterium ist nicht quantifizierbar.

## **Aufgabe 3**

**a)**

- Information systems  $\Rightarrow$  Information systems applications  $\Rightarrow$  Digital libraries and archives
- Information systems  $\Rightarrow$  Information retrieval
- Information systems  $\Rightarrow$  Information retrieval  $\Rightarrow$  Retrieval models and ranking
- Applied computing  $\Rightarrow$  Computers in other domains

- Networks  $\Rightarrow$  Network architectures

**b)**

**Vorteil:** Durch die freie Wahl von Schlagwörtern können diese deutlich genauer auf die entsprechende Arbeit angepasst werden und dem Nutzer bereits vor dem Lesen mehr Informationen über den Inhalt der Arbeit liefern.

**Nachteil:** Das Einordnen von freien Schlagwörtern ist deutlich komplizierter. Nutzer haben keine klare Übersicht über die vorhandenen Schlagwörter. Wenn ein Nutzer sich für eine bestimmte Thematik interessiert, ist es deutlich einfacher für ihn sich an einem Klassifikationssystem zu orientieren.

**c)**

Klassifikation freie Schlagwörter Titel Volltext

Die meisten Treffer werden bei der Volltext-Suche erwartet, da es dort aufgrund der Menge an Wörtern/Sätzen deutlich wahrscheinlicher ist eine entsprechende Suchanfrage zu beantworten. Dafür ist diese Form der Suche relativ ungenau, da wahrscheinlich viele Publikationen gefunden werden, welche nicht unbedingt mit der gesuchten Thematik übereinstimmen. Bei geeignet gewählten freien Schlagwörtern sind relativ gute Ergebnisse zu erwarten, falls die Suchanfrage darauf ausgerichtet ist. Leider könnten so aber Publikationen wegfallen, welche inhaltlich eventuell sogar besser passen würden, der Autor aber die Schlagwörter nicht vergeben hat. Dies trifft auch auf die Klassifikation zu. Hier muss der Autor seine Arbeit ebenfalls korrekt klassifizieren, um es dem Nutzer zu ermöglichen seine Suchanfrage bestmöglich zu beantworten. Es werden mehr Ergebnisse erwarten als bei den freien Schlagworten. Bei der Suche auf Basis des Titels sind ebenfalls genauere Treffer zu erwarten als bei der Volltext-Suche, jedoch ist durch eine fehlende Klassifizierung nicht zu garantieren, dass es sich dabei um die richtige Domäne handelt (Filter etc. ausgenommen). So könnte die Suchanfrage nach *Latex* sich sowohl auf das Textsatzsystem, als auch auf das Material beziehen.

## Aufgabe 4

Siehe Anhang: Aufgabe4.java