

Was misst der h -Index (nicht)?

Kritische Überlegungen zu einer populären Kennzahl für Forschungsleistungen

Das allgegenwärtige Leistungs- und Effizienzstreben drängt auch Wissenschaftler dazu, Forschungsleistungen zu messen und zu vergleichen. Eine verbreitete Zitationskennzahl ist der h -Index, der der Maximalanzahl Publikationen h eines Forschers entspricht, die je mindestens h -mal zitiert sind. Der Beitrag skizziert die Grundgedanken des h -Index und nimmt basierend auf empirischen Vergleichen kritisch Stellung zu seiner Aussagekraft und Beeinflussbarkeit.



Dr. Magnus Richter
ist Akademischer Rat am Fachgebiet Nachhaltige Produktionswirtschaft und Logistik der TU Ilmenau

Stichwörter: Forschungsleistungsmessung, h -Index, Zitationskennzahlen, Szientometrie

1. Grundgedanken und Eigenschaften des h -Index

Aufgrund der stetigen Verknappung personeller, materieller und finanzieller Ressourcen sowie des allgegenwärtigen Wettbewerbsstrebens sind auch Hochschulen sowie die an ihnen tätigen Wissenschaftler angehalten, leistungsstark zu forschen. In diesem Zuge rücken Vergleiche mit anderen Forschern bzw. **Forschungsleistungsmessung** zunehmend in den hochschulpolitischen Fokus. Ein populäres Maß für Forschungsleistungen ist der h -Index, der durch *Hirsch*, einen argentinischen Physikprofessor, bekannt gemacht wurde (vgl. *Hirsch*, 2005). Der **h -Index** eines Forschers entspricht der **Maximalanzahl** seiner **Publikationen h** , die **jeweils mindestens h Zitationen** aufweisen. Ein h -Index von z. B. 12 impliziert somit wenigstens 12 Publikationen mit je 12 oder mehr Zitationen sowie eine Gesamtzitationsanzahl von mindestens 144. Der h -Index ist aufgrund seines intuitiven Aufbaus auch in den Wirtschaftswissenschaften verbreitet, d. h. viele Ökonomen kennen ihren eigenen

h -Index und geben ihn z. T. auch auf Homepages und in Bewerbungen an.

Abb. 1 zeigt anhand eines sog. *Ferrer*-Graphen (bzw. *Young*-Diagramms) die Struktur des h -Index. Als Beispiel dient das überschaubare Werk des Verfassers. (Die zur Berechnung des h -Index erforderlichen Daten können mittels *Google Scholar* oder der Software *Publish or Perish* (vgl. *Clermont*, 2010) erhoben werden.) Zur Berechnung von h sind die Publikationen $p = 1, \dots, P$ des betreffenden Forschers zunächst **absteigend** nach **ihrer Zitationsanzahl c_p** zu sortieren. Die Publikationen gleichen dann nebeneinander angeordneten Türmen, deren Höhe die jeweilige Zitationsanzahl anzeigt. Im hier betrachteten Beispiel weist die am häufigsten referenzierte Publikation eine Zitationsanzahl von $c_1 = 12$ auf. Dabei zeigt der Index 1 (von c_1) an, dass es sich hierbei um die meistzitierte Publikation (Rangplatz 1) handelt. Auf Rangplatz 2 folgt eine Publikation mit 11 Zitationen ($c_2 = 11$), die sich analog in der Blockhöhe des zweiten Turms (von links) offenbaren. Gemäß dieser Darstellungslogik werden auch alle übrigen Publikationen, d. h. die auf den Rangplätzen 3–7, angeordnet. Bislang noch nicht zitierte Publikationen werden der Übersichtlichkeit halber nicht mit angezeigt.

Der h -Index lässt sich nun ganz einfach am Diagramm ablesen: Hierzu muss lediglich ein im Ursprung fixiertes **Quadrat** solange nach **oben rechts vergrößert** werden, bis es den Rand der „Publikationstürme-Menge“ berührt. Das Aufspannen dieses Quadrats ist in *Abb. 1* in Form des Pfeils dargestellt, der in nord-östliche Richtung zeigt und im vorliegenden Beispiel den **Randpunkt (4;4)** ergibt. Dieser repräsentiert den h -Index des Verfassers in Höhe von 4.

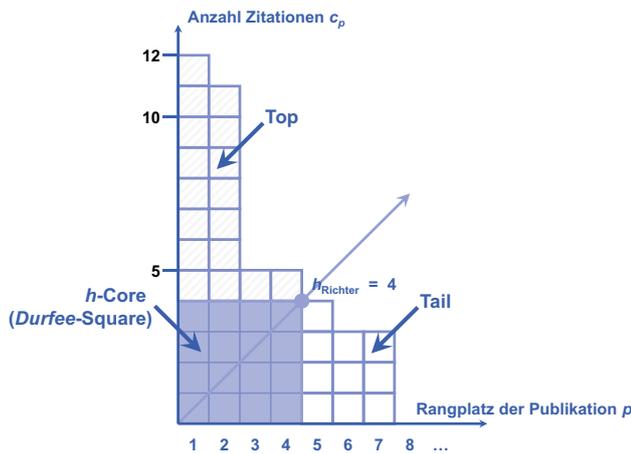


Abb. 1: Exemplarische Zitationskurve des Verfassers

Geometrisch entspricht der h -Index der Seitenlänge des größten Quadrats, das sich zwischen Ursprung des Koordinatensystems und dem oberen rechten, stufenförmig verlaufenden Rand der Zitationskurve einfügen lässt. Dieses in Abb. 1 blau markierte Quadrat wird als **h -Core** oder *Durfee-Square* bezeichnet (vgl. Andrews, 1979, S. 736; Anderson/Hankin/Killworth, 2008, S. 579). Der h -Core ist das zentrale Element bei der Berechnung des h -Index, schließlich sind in ihm die für den h -Index maßgeblichen Publikationen samt der auf sie entfallenden, effektiven Zitationen vereint. In Abschn. 2 werden ergänzend zum h -Core die Flächenbereiche **Top** (schraffierte Kästen über dem h -Core) und **Tail** (leere Kästen rechts vom h -Core) definiert, formalisiert und bezüglich ihrer Relevanz für die Beurteilung von Forschungsleistungen (auf Basis des h -Index) diskutiert. Dabei wird anhand realer Beispiele verdeutlicht, dass sich auch Zitationskurven von Forschern mit gleichem oder zumindest ähnlichem h -Index stark voneinander unterscheiden können. In Abschn. 3 werden Aspekte der **Effizienz** des **Zitiertwerdens** bzw. der **Beeinflussbarkeit** des h -Index diskutiert, wie etwa die Frage, ob die Zitationshäufigkeit vom Publikationstyp abhängt. Dieses Wissen kann nützlich sein, wenn Forscher einen hohen h -Index anstreben. Abschn. 4 fasst die wesentlichen Erkenntnisse des Beitrags zusammen.

2. Die Struktur von Zitationskurven

2.1. Top und Tail als „Nebenschauplätze“ des h -Index

Wie in Abb. 1 erkennbar, bestehen Zitationskurven – neben dem h -Core – aus zwei weiteren Flächen, nämlich dem über dem h -Core befindlichen **Top** sowie dem rechts an den h -Core angegliederten **Tail**. Auf Top und Tail entfällt meist ein beträchtlicher Teil der Zitationen, die das Werk eines Forschers im Laufe der Zeit erhält. So zeigt das Beispiel aus Abb. 1, dass die Publikationen p ($p = 1, \dots, 7$) bis dato ins-

gesamt $12 + 11 + 5 + 5 + 4 + 3 + 3 = 43$ Zitationen erzielen, der h -Core jedoch aus nur $h^2 = 16$ Zitationen besteht. Für den h -Index 4 des Verfassers hätte rein rechnerisch somit ein Bruchteil an Zitationen genügt, nämlich $16 : 43 \approx 37\%$. Die 27 übrigen, nicht im h -Core enthaltenen Zitationen liegen somit in Top (17 bzw. ca. 40%) und Tail (10 bzw. ca. 23%).

Obwohl auch Top- und Tail-Zitationen eines Forschers Ausdruck seiner Wirksamkeit in einer Forschungsgemeinschaft sein können, sind sie für den **aktuellen h -Index irrelevant**. Zwar mag es einen Forscher erfreuen, wenn seine h meistzitierten Publikationen öfter als h -mal zitiert sind oder sein Werk zusätzlich zu den h h -Core-Publikationen weitere ($P-h$) Werke beinhaltet; jedoch: h bleibt h , egal wie groß Top und Tail sein mögen.

Hirsch sind diese „Unterschlagungen“ bewusst. Er postuliert, „that two individuals with similar h s are comparable in terms of their overall scientific impact, even if their total number of papers or ... citations is very different.“ (Hirsch, 2005, S. 16569) Dies hat Kritik am h -Index aufkeimen lassen (vgl. van Raan, 2006, S. 500 f.; Vinkler, 2007, S. 490 f.): So wurden u. a. der verzerrende Einfluss des Erhebungstools für Zitationsdaten sowie mögliche Abhängigkeiten des h -Index von der disziplinären Zugehörigkeit diskutiert (vgl. Bar-Ilan, 2008, S. 269). Auch **Selbstzitationen** wurden in diesem Kontext thematisiert (vgl. Glänzel/Thijs/Schlemmer, 2004). Es ist somit kaum verwunderlich, dass zahlreiche **Modifikationen** und **Alternativen** entwickelt wurden, die die Unzulänglichkeiten des h -Index beheben sollen (vgl. etwa zum g -Index Egghe, 2006, und zum e -Index Zhang, 2009, sowie für einen Überblick über die Entwicklung von Zitationsmaßen Norris/Oppenheim, 2010).

2.2. Empirischer Vergleich von Zitationskurven

Um zu illustrieren, wie stark der h -Index die Verschiedenartigkeit von Forschungsleistungen bzw. Teilaspekte individueller Forschungsprofile verdecken kann, werden im Folgenden zwei äußerst forschungsstarke Professoren anonym miteinander verglichen, die mit h -Indices von 85 bzw. 89 annähernd gleich gut sind.

Sie forschen in den Bereichen Dienstleistungsmarketing bzw. Marketingmanagement und eignen sich aufgrund ihrer ähnlichen Ausrichtung auf internationale und hochrangige Publikationen für einen tragfähigen Leistungsvergleich. Abb. 2 zeigt die Zitationskurven der im Folgenden als X bzw. Y bezeichneten Forscher gemäß der aus Abb. 1 bekannten Darstellungsweise (vgl. für eine Anwendung des h -Index auf 100 Ökonomen Tol, 2009).

Auf der Abszisse sind die nach Zitationshäufigkeit absteigend sortierten Publikationen beider Forscher notiert; die Ordinate gibt die jeweiligen Zitationshäufigkeiten an. Die

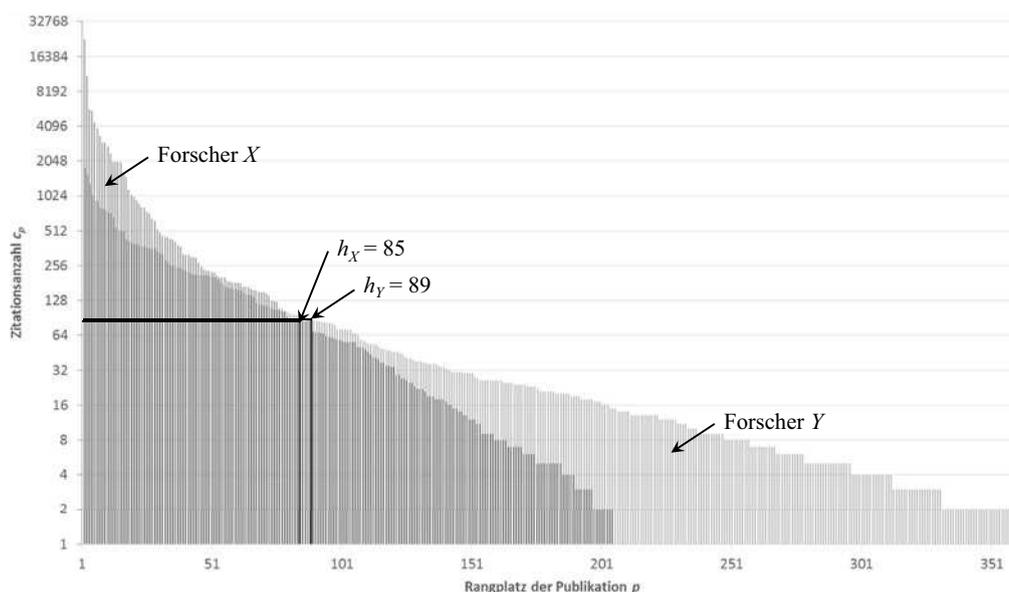


Abb. 2: Empirische Zitationskurven von X und Y (Stand: Januar 2017)

logarithmische Skalierung der Ordinate deutet bereits an, dass sich der Vergleich zwischen X und Y schwierig gestaltet, da ihre meistzitierten Publikationen große Differenzen bezüglich der Zitationsanzahl aufweisen. Eine lineare Skalierung der Zitationsanzahl c_p auf der Ordinate hätte wesentliche Strukturelemente der Zitationskurven, wie z. B. die h -Cores, nicht erkennen lassen, da Xs meistzitierte Publikation über 25.000 Mal zitiert wurde und Ys Zitationskurve somit viel zu klein skaliert worden wäre.

Um den Vergleich zu vereinfachen, sind in Tab. 1 einige wesentliche bibliometrische Daten von X und Y aufgeführt, die neben der Größe der Flächenbereiche von h -Core, Top und Tail auch die Zitationshäufigkeiten der zehn jeweils meistzitierten Publikationen von X und Y beinhalten. Interessant sind hierbei die – in Kontrast zum ähnlichen h -Index – erheblichen Unterschiede beim Abgleich der übrigen Kennzahlen: So führt *google scholar* für Forscher Y insgesamt 658 Veröffentlichungen an, wohingegen Forscher X „nur“ 288 Veröffentlichungen aufweist, d. h. lediglich ca. 44 %.

Mit seiner vergleichsweise „geringen“ Publikationsanzahl generiert X allerdings deutlich mehr Gesamtzitationen, nämlich annähernd 130.000, also 256 % mehr als Y. Dass X trotz seines Vorsprungs bei den Gesamtzitationen letztlich einen geringeren h -Index realisiert als Y, äußert sich logischerweise auch am Flächenanteil seines h -Cores an der Zitationskurve, der mit 5,58 % viel geringer ist als bei Y (21,73 %). Bei Y sind die Gesamtzitationen also „vorteilhafter“ auf den *Durfee*-Square konzentriert.

Dies zeigt sich *vice versa* auch in dem auffallend großen Top der Zitationskurve von X, auf den mehr als 92 % seiner Gesamtzitationen entfallen. Auch die Zitationen ihrer 10 meistzitierten Beiträge verdeutlichen, dass die **relative h -Ineffizienz** von X gegenüber Y vor allem in Xs sehr stark zitierten Top-10-Publikationen begründet liegt. Sie erzielen in Summe 88.164 Zitationen, womit sie (rechnerisch)

einer h -Index-bezogenen „Verschwendung“ von $88.164 - (10 \cdot 85) = 87.314$ Zitationen entsprechen – denn X würde auch dann einen h -Index von 85 erzielen, wenn seine 10 meistzitierten Veröffentlichungen jeweils nur 85 Zitationen besäßen. Lediglich deren Rangplätze würden sich bei je 85 Zitationen ändern, da dann die aktuell auf den Rangplätzen 11–20 befindlichen Veröffentlichungen von X auf die Positionen 1–10 aufrücken würden.

3. Kritische Überlegungen zur Steuerbarkeit des h -Index

Hinsichtlich der Effizienz des Zitiertwerdens stellt sich für „ h -gierige“ Forscher die Frage, wie der h -Index gesteigert werden kann. Bildlich gesprochen ist somit zu klären, welche Publikationen am besten geeignet sind, schnell in den h -Core zu gelangen. Studien, deren Ergebnisse auch von einer explorativen Vorstudie des Verfassers bestätigt wurden, zeigen, dass **Lehrbücher**, **State-of-the-Art**-Aufsätze und **Typisierungen** besonders oft zitiert werden. Die Erhöhung des h -Index kann somit durch Fokussierung auf **breitenwirksame Publikationstypen** begünstigt werden.

Aus Effizienzgesichtspunkten könnte dabei sinnvoll erscheinen, Publikationen so auszugestalten, dass sie möglichst schnell zum h -Index beitragen, dann aber **keine** weiteren Zitationen mehr erhalten – schließlich würde dadurch nur das Top vergrößert. Eine Option, ein solches Versiegen von Zitationen zu steuern, könnte darin bestehen, **Trends** bzw. **Modethemen** zu behandeln, die nach kurzer bis mittlerer Zeitspanne an Relevanz verlieren und dann nicht weiter beachtet werden. Hiergegen sprechen allerdings zwei Aspekte:

Zum einen verändern sich Zitationskurven und h -Indices im Zeitverlauf, sodass Publikationen, die zu einem bestimmten Zeitpunkt den h -Core bilden, ihn später auch

Vergleichskriterium	Term/Berechnung	X	Y	$\Delta (Y-X)$
<i>h</i> -Index	h	85	89	+4
Anzahl Publikationen	P	288	658	+370
Anzahl Zitationen	$\sum_{p=1}^P c_p$	129.470	36.451	-93.019
Flächeninhalt <i>h</i> -Core	h^2	7.225	7.921	+696
Flächenanteil <i>h</i> -Core	$h^2 / \sum_{p=1}^P c_p$	5,58 %	21,73 %	+16,15 %
Flächeninhalt Top	$\sum_{p=1}^h c_p - h^2$	119.358	23.199	-96.159
Flächenanteil Top	$\left(\sum_{p=1}^h c_p - h^2 \right) / \sum_{p=1}^P c_p$	92,19 %	63,64 %	-28,55 %
Flächeninhalt Tail	$\sum_{p=h+1}^P c_p$	2.887	5.331	+2.444
Flächenanteil Tail	$\sum_{p=h+1}^P c_p / \sum_{p=1}^P c_p$	2,22 %	14,63 %	+12,41 %
10 meistzitierte Publikationen		Zitationshäufigkeit c_p		
		X	Y	$\Delta (Y-X)$
$p = 1$		25.018	2.182	-22.836
$p = 2$		23.135	1.784	-21.351
$p = 3$		11.053	1.569	-9.484
$p = 4$		5.749	1.288	-4.461
$p = 5$		5.557	1.037	-4.520
$p = 6$		4.450	932	-3.518
$p = 7$		3.946	917	-3.029
$p = 8$		3.372	806	-2.566
$p = 9$		2.957	794	-2.163
$p = 10$		2.927	774	-2.153

Tab. 1: Bibliometrische Details der Forscher X und Y

wieder verlassen können. Würde X also z. B. versuchen, eine neue Publikation so auszugestalten, dass sie zu gegebener Zeit eine Punktlandung auf 86 Zitationen schafft und dann bei 86 verharrt, um seinen *h*-Index effizient (!) zu steigern, würde er damit einen weiteren Anstieg seines *h*-Index auf 87 durch eben diese Publikation (zumindest zeitweise) blockieren.

Zum anderen erscheint zweifelhaft, dass Forscher rein aus Effizienzgesichtspunkten Interesse am Eindämmen des Tops haben sollten: Eine effiziente Konzentration von Zitationen auf den *h*-Core mag nützlich sein, z. B. wenn Nachwuchsforscher trotz geringer Gesamtzitationen einen ansehnlichen *h*-Index erzielen wollen. Überzählige Zitationen in Top und Tail sind allerdings für das **Wachstum** des *h*-Index von zentralem Stellenwert, zumal die Top-Zitationen c_1-c_h (zusammen mit der Gesamtanzahl P) den *h*-Index nach oben begrenzen und Tail-Publikationen samt Zitationen das „Nachrückpotenzial“ abbilden. Top und Tail als Verschwendung deklarieren und eindämmen zu wollen, ist also kontraproduktiv.

Ein (oft als illegitim betrachtetes) Mittel zur Steigerung des *h*-Index sind **Selbstzitationen**. Sie bezeichnen das Zitieren einer Publikation durch den/die Verfasser selbst. Manchen Kritikern erschienen die durch sie bewirkten Verzerrungen so gravierend, dass sie in einem offenen Brief an die Herausgeber der Zeitschrift *Scientometrics* die Störanfälligkeit des *h*-Index beklagten. V. a. bei Abhängigkeit von Forschern von anderen Wissenschaftlern (z. B. Herausgebern und Gutachtern), so die These, könne **opportunistisches Zitieren** gefördert werden, das mächtigere Forscher letztlich besser dar stehen ließe, als es tatsächlich der Fall ist. Von Selbstzitationen solle der *h*-Index somit abstrahieren (vgl. *Zhivotovsky/Krutovsky, 2008, S. 374 f.*).

An der Zitationskurve in *Abb. 1* wird die Problematik der Selbstzitationen deutlich: So ließe sich der *h*-Index leicht von 4 auf 5 erhöhen, indem der Verfasser seine Publikation auf Rangplatz 5 zitiert. Da diese kurz davor steht, in den *h*-Core zu gelangen, würde hierzu bereits eine weitere Zitation genügen. Irreführend wäre der so erzielte *h*-Index 5 dahingehend, dass er nicht mehr nur die Wahrnehmung der

Arbeiten durch **andere Wissenschaftler** abbildet. Die Legitimität von Selbstzitation hängt somit maßgeblich davon ab, ob sie sachlich wohl begründet ist oder lediglich der Selbstdarstellung dient. Nach Ansicht des Verfassers sollte Selbstzitation **sparsam eingesetzt** werden. Legitim erscheint sie hingegen im Rahmen von **Qualifikationsverfahren**, wie Promotionen und Habilitationen, bei denen es wichtig ist anzuzeigen, welche Vorleistungen erbracht wurden und inwiefern Ko-Autoren (z. B. Betreuer) beteiligt waren. Ähnliches gilt bei der Beitragseinreichung an Zeitschriften, bei der Selbstzitation helfen kann, die **Originalität** und **Exklusivität** eines Manuskripts zu beurteilen.

Weitere Einflussfaktoren auf den h -Index sind die **Sprache** und das **Publikationsorgan**. So ist es naheliegend, dass z. B. in Englisch publizierte Beiträge mehr Leser und potenzielle Zitierende ansprechen als Beiträge, die in Deutsch verfasst sind. Analog gilt für Publikationsorgane (z. B. Fachzeitschriften), dass internationales Erscheinen in einer verbreiteten Sprache mehr Zitationen bedingt als ein Erscheinen in seltenen Sprachen und auf nationaler Ebene.

Dass diese Überlegungen keine akademischen Glasperlenspiele sind, zeigt ein erneuter Blick auf den Vergleich zwischen X und Y (vgl. *Tab. 1*). So mag sich der Leser womöglich gefragt haben, worin die Gründe für den großen Unterschied bei den Gesamtzitationen von X und Y bestehen könnten. Hierzu sei gesagt, dass Y einige seiner frühen Werke auf Deutsch verfasst hat und dieser Zitationsnachteil offenbar nicht kompensierbar ist. X hingegen publiziert von Beginn auf Englisch, d. h. in einer besonders „ h -gefälligen“ Sprache, und in entsprechend verbreiteten Organen. Zudem mag X (Y) der Umstand begünstigt (benachteiligt) haben, dass die weltweit angesehensten Journale in Englisch erscheinen bzw. selbst die renommiertesten deutschsprachigen Zeitschriften lediglich die dritthöchste Kategorie B des sog. *Jourqual 3*-Rankings erreichen.

4. Resümee

Ziel des Beitrags waren die Darstellung und Würdigung des h -Index, einem populären Maß für Forschungsleistung. Der h -Index gleicht der Maximalanzahl h an Publikationen eines Forschers, die je $\geq h$ Zitationen aufweisen. Graphisch entspricht der h -Index der Seitenlänge des größten Qua-

drats, das sich zwischen Ursprung und Rand der Zitationskurve aufspannen lässt. Der h -Index grenzt jedoch Top und Tail von Zitationskurven aus, worin eine seiner wesentlichen Limitation besteht. Anhand eines empirischen Vergleichs von Forschern mit ähnlichem h -Index wurde illustriert, welche deutliche Unterschiede von Zitationskurven zutage treten können, wenn die vom h -Index ausgegrenzten Aspekte doch beachtet werden. Es zeigte sich, dass eine faire Würdigung von Forschungsleistungen oft „mehrerer Blicke“ und Dimensionen bedarf, die der h -Index **nicht** erlaubt. Zudem wurden Ansätze zur Erhöhung des h -Index vorgestellt. Hierzu zählt v. a. die Fokussierung auf internationale Publikationsorgane. Aus moralischen Gründen sollte auf Selbstzitationen hingegen verzichtet werden, sofern sie nicht unbedingt erforderlich sind bzw. Offenlegungspflichten dienen.

Literatur

- Anderson, T. R., Hankin, R. K. S., Killworth, P. D., Beyond the Durfee Square: Enhancing the h -Index to score total Publication Output, in: *Scientometrics*, 76. Jg. (2008), Nr. 3, S. 577–588.
- Andrews, G. E., Partitions and Durfee Dissection, in: *American Journal of Mathematics*, 101. Jg. (1979), Nr. 3, S. 735–742.
- Bar-Ilan, J., Which h -Index? – A Comparison of WoS, Scopus and Google Scholar, in: *Scientometrics*, 74. Jg. (2008), Nr. 2, S. 257–271.
- Clermont, M., Publish or Perish als Auswertungstool für Google Scholar, in: *WiSt – Wirtschaftswissenschaftliches Studium*, 39. Jg. (2010), Nr. 9, S. 418–425.
- Egghe, L., Theory and Practice of the g -Index, in: *Scientometrics*, 69. Jg. (2006), Nr. 1, S. 131–152.
- Glänzel, W., Thijs, B., Schlemmer, B., A bibliometric Approach to the Role of Author Self-Citations in scientific Communication, in: *Scientometrics*, 59. Jg. (2004), Nr. 1, S. 63–77.
- Hirsch, J. E., An Index to quantify an Individual's scientific Research Output, in: *Proceedings of the National Academy of Sciences of the United States of America*, 102. Jg. (2005), Nr. 46, S. 16569–16572.
- Norris, M., Oppenheim, C., The h -Index: a broad Review of a new bibliometric Indicator, in: *Journal of Documentation*, 66. Jg. (2010), Nr. 5, S. 681–705.
- van Raan, A. F. J., Comparison of the Hirsch-Index with standard bibliometric Indicators and with Peer Judgement for 147 chemistry Research Groups, in: *Scientometrics*, 67. Jg. (2006), Nr. 3, S. 491–502.
- Vinkler, P., Eminence of Scientists in the Light of the h -Index and other scientometric Indicators, in: *Journal of Information Science*, 33. Jg. (2007), Nr. 4, S. 481–491.
- Zhang, C.-T., The e -Index, complementing the h -Index for excess Citations, in *PLoS ONE*, 4. Jg. (2009), Nr. 5, S. 1–4.