

h-지수 및 변형 지수에 대한 연구

A Study on the h-Index and Its Variants

이재윤, 경기대학교 {memexlee@kgu.ac.kr}

Lee, Jae-Yun, Kyonggi University

인용을 통한 연구성과 지표인 h-지수는 2005년에 제안된 이후 공식의 간단명료함, 용이한 산출 방법, 지수의 강건성 등이 인정되면서 계량정보학 분야를 비롯한 학술 공동체의 활발한 후속 논의가 이례적으로 신속하게 이어지고 있다. 후속 연구는 주로 h-지수의 타당성 검토, h-지수를 이용한 특정 분야의 개인 성과 측정 시도, h-지수의 타 영역에의 응용, h-지수의 보완 등을 주제로 전개되고 있다. 이 연구에서는 우선 h-지수와 이를 보완한 g-지수를 비롯한 변형 지수에 대해서 현재까지 전개되고 있는 논의를 살펴보았다. 그리고 인용빈도가 높지 않은 경우를 고려하여 개선한 지수를 제안하고 실험적인 측정을 시도하였다.

1. 서론

Hirsch(2005)가 제안한 h-지수는 인용을 통해서 개인의 연구 성과를 측정하려는 시도이다. 캘리포니아 대학 샌디에이고 캠퍼스의 물리학자인 Jorge Hirsch는 발표전 논문을 2005년 여름에 먼저 E-print archive에 올렸는데, 일주일 후 *Nature*지의 소식란에 Ball(2005)이 Hirsch의 논문과 h-지수를 소개한 것이 신속하면서도 폭넓은 관심을 끈 계기가 되었다.

h-지수는 공식의 간단명료함, 손쉬운 산출 방법, 지수의 강건성 등이 인정되면서 발표 이후 계량정보학 분야 뿐만 아니라 여러 학술 공동체의 활발한 후속 논의가 이례적으로 신속하게 이어지고 있다. 후속 연구는 주로 h-지수의 타당성 검토, h-지수를 이용한 특정 분야의 개인 성과 측정 시도, h-지수의 타 영역에의 응용, h-지수의 보완 등에 대해서 전개되고 있다.

이 연구에서는 우선 h-지수와 이를 보완한 g-지수를 비롯한 변형 지수를 살펴본 다음, 개선된 새로운 지수를 몇 가지 제안하였다. 그리고 실제 자료를 대상으로 각 지수의 특성을 검증해보았다.

2. h-지수

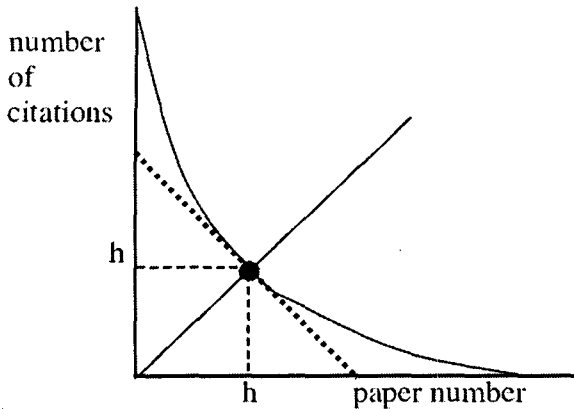
2.1 h-지수의 정의 및 특성

h-지수의 정의는 다음과 같다(Hirsch 2005).

어떤 과학자의 N_p 개 논문 중에서 h 개의 논문은 최소 h 번 이상씩 인용되었고, 나머지 $N_p - h$ 개 논문은 모두 개별 인용빈도가 h 번 이하이면 그의 h-지수는 h 이다.

풀어서 설명하자면, 어떤 연구자의 논문을 인용빈도가 높은 순부터 나열하였을 때 논문의 인용빈도가 논문의 순위보다 크거나 같은 마지막 논문의 순위가 그 연구자의 h-지수가 된다. 예를 들어 h-지수가 20이라면, 최소 20회 이상 인용된 논문을 20편 발표했다는 뜻이다.

인용빈도의 분포가 <그림 1>의 굵은 선과 같다면 총 인용빈도는 굵은 선 아래의 면적이다. 이 그림에서 h-지수는 원점에서 45도 각도로 출발한 직선과 마주치는 지점의 x좌표(순위)가 된다. 실제 인용빈도는 실수가 아닌 정수이므로 이 그림처럼 곡선형태가 아닌 계단식으로 감소한다.



<그림 1> 인용빈도 분포와 h-지수 (Hirsch 2005)

<표 1>에 가상의 자료로 h-지수를 산출한 예를 소개하였다. Web of Science와 같은 인용 색인 DB를 이용하면 이와 같은 인용 순위를 손쉽게 산출할 수 있다.

물론 h-지수 이외에 총 인용빈도 등 연구자를 평가하는 다른 지표도 여러 가지가 있다. Hirsch는 다른 지표에 대해서는 다음과 같이 비판하였다.

- 발표논문 총 수: 생산성을 반영하지만 논문의 중요도나 영향력은 측정하지 못한다.
- 총 인용빈도: 총 영향력을 나타내지만 완벽한 산출이 어려우며 공저자로 참여한 한 두 논문의 인용빈도가 예외적으로 높은 경우에 과대 평가될 위험이 있다.
- 논문당 평균 인용빈도: 연구 경력이 긴 연구자와 짧은 연구자를 비교할 수 있지만, 완벽한 산출이 어려우며 생산성이 높은 연구자에게 불리한 단점이 있다.
- 일정 빈도 이상 인용된 논문의 수: 앞선 지수들의 단점을 대부분 극복할 수 있으나, 인용빈도의 기준값을 정하는 것이 임의적이어서 누군가에게는 유리하게, 또 다른 누군가에게는 불리하게 작용할 수 있다.
- 인용빈도로 일정 순위 이상인 상위 논문의 인용빈도 집합: 앞선 지수들의 단점을 대부분 극

<표 1> 연구자의 논문 인용빈도 가상 사례

인용 순위	연구자 A의 논문	연구자 B의 논문	연구자 C의 논문	연구자 D의 논문
1	10	4	3	20
2	7	4	3	3
3	5	4	1	2
4	4	4	1	1
5	3	3	1	1
6	2	2		
7	2	2		
8	2	2		
9	1	1		
10	1	1		
총인용	37	27	9	27
평균인용	3.7	2.7	1.8	5.4
h	4	4	2	2

복할 수 있다. 그러나 하나가 아닌 여러 개의 값을 살펴야 하며, 일정 순위를 정하는 것이 임의적이어서 역시 누군가에게 유리하거나 불리한 상황이 발생한다.

이에 반해서 h-지수는 인용빈도가 높은 소수 논문에 좌우될 여지를 없애고 연구 생산성과 발표논문의 중요도를 함께 반영하는 척도라고 평가된다(Lehmann et al. 2006).

Hirsch(2005)는 h-지수를 제안한 같은 논문에서 h-지수가 연구 경력이 오래되었을수록 유리한 점을 감안하여 m이라는 척도도 제안하였다. 연구자의 최초 논문 발표 이후 경과한 햇수를 n이라고 하였을 때 m은 다음과 같이 산출된다.

$$m = \frac{h}{n}$$

즉, m은 h-지수로 측정되는 연구자의 인용 영향력이 어느 정도로 빠르게 달성되었는가를 반영한다. 연구자의 영향력을 물체의 이동에 비유해서 말하자면, h-지수는 총 이동 거리에 해당하고 m은 평균 이동 속도에 해당한다고 할 수 있다. m이 크면 클수록 연구자의 인용 영향력이 급속하게 향상되었음을 나타낸다.

2.2 h-지수를 이용한 연구자 성과 측정 사례

h-지수는 이용하는 인용정보원에 따라서 다른 결과가 나타날 수 있다. 많은 경우에 Thomson ISI의 Web of Science 데이터베이스를 이용하여 측정하지만, Google Scholar나 Scopus를 이용하는 방안도 논의되고 있으며, Lehmann(2006)과 같이 특정 분야 원문 데이터베이스에 인용정보가 포함된 것을 이용하는 경우도 있다. 어느 정보원을 이용하더라도 동명이인의 문제가 있을 수 있기 때문에 인용정보를 검색한 후 직접 상위 인용빈도 문헌을 확인하는 절차를 거쳐서 h-지수를 확정하고 있다. h-지수는 인용빈도 상위 문헌만 대상으로 하므로 이처럼 직접 눈으로 검증할 수 있다는 것이 장점이다.

처음 Hirsch(2005)가 h-지수를 제안하면서 시험적으로 측정한 대상은 물리학 분야와 생물학

및 생의학 분야의 유명 연구자였다. 그 이후 다양한 분야에서 h-지수를 측정하여 소개하였고, 개인이 아닌 연구팀 단위 측정이나 학술지, 용어 등에 대해서 적용하는 시도도 발표되었다. <표 2>에 주요 응용 사례를 정리하였다.

3. h-지수에 대한 비판과 보완

3.1 h-지수에 대한 기존 비판

◎ 자기인용에 다소 영향 받음

Hirsch(2005)는 h-지수를 제안하면서 자기인용에 크게 좌우되지 않는다고 주장하였으나, Cronin & Meho(2006)가 미국 정보학자에 대해서 분석한 결과에 따르면 31명 분석대상자 중에서 자기인용을 포함한 결과 순위가 4위 상승한 경우와 7위 하락한 경우가 있었다. 이는 자기인용의 포함 여부가 최소한 일부 연구자의

<표 2> h-지수 응용 사례

연구자	대상 분야	측정단위	인용 출처	비고
Hirsch(2005)	물리학 생물학/생의학 물리학	개인 개인 학과	Web of Science	최고 110 (E. Witten) 최고 191 (S. H. Snyder) UCSD 물리학과 - 118
Glänzel & Persson(2005)	계량정보학 (Price 메달 수상자)	개인	Web of Science	최고 17 (T. Braun)
Bar-Ilan(2006)	계량정보학 (Price 메달 수상자)	개인	Google Scholar Web of Science	최고 23 (L. Leydesdorff) 최고 18 (A.F.J. van Raan / T. Glänzel)
Cronin & Meho(2006)	정보학 (미국)	개인	DIALOG (SCI, SSCI, A&HCI)	최고 20 (N.J. Belkin) - 자기인용 포함 최고 19 (N.J. Belkin) - 자기인용 제외
Bornmann & Daniel(2005)	생의학	개인	SCI	박사후 연구비 지원 승인자 평균 3.84 박사후 연구비 지원 미승인자 평균 2.72
Batista et al. (2005)	물리학 화학 생물학/생의학 수학	개인	Web of Science	최고 37 최고 29 최고 24 최고 14 브라질 연구자가 포함된 논문만 분석함
Lehmann et al. (2006)	고에너지물리학	개인	SPIRES	개인별 인용빈도의 최고, 평균, 중간값과 지수 비교
Kelly & Jennions(2006)	생태학/진화생물학	개인	Web of Science	학술지별, 성별 차이 분석
van Raan (2006)	화학/화공학 (네덜란드)	연구팀	Web of Science	발간년도 포함 3년 이내 인용만 채택
Braun et al. (2005)	전 분야	학술지	Web of Science	2001년 발간 논문에 대한 인용 측정
Rousseau(2006a)	정보학	학술지	Web of Science	JASIS 1991-2000 10년간
Banks(2006)	물리학	용어	Web of Science	h보다 m 위주 분석

h-지수에는 영향을 끼친다는 것을 시사한다.

$$h_f = h / \langle N_a \rangle = h^2 / N_a^{(T)}$$

◎ 분야간 비교가 어려움

Hirsch(2005)도 언급하였지만 h-지수로는 상이한 분야의 연구자를 비교하기가 어렵다. 분야에 따라서 값의 차이가 크기 때문이다 (Batista et al. 2005; Cronin & Meho 2006; Kelly & Jennions 2006). 이는 특히 공저자 수에 영향을 많이 받는 것으로 알려져 있다(Batista et al. 2005).

$\langle N_a \rangle$ 는 h-지수 산정에 포함된 h개 논문의 평균 공저자 수이며 $N_a^{(T)}$ 는 h개 논문의 총 저자수(중복 허용)를 뜻한다. 이 공식은 분야별 차이를 극복할 수 있을 뿐만 아니라, 단독저자로 참여한 경우보다 공저자로 참여한 논문의 인용 영향력을 낮춰주는 효과도 얻을 수 있다. 다만 각 논문의 공저자 수를 모두 파악하지 않으면 산출이 불가능한 것이 단점이다.

◎ 동률값이 흔하여 변별력이 낮음

h-지수는 동률 순위의 연구자가 많이 나타나므로 변별력이 약하다(Batista et al. 2005). 계량 정보학자에 대한 Glänzel & Persson(2005)의 연구에서는 1위인 17을 기록한 연구자가 3명으로 나타났으며, Cronin & Meho(2006)이 조사한 미국 정보학자의 h-지수 순위는 공동 11위가 무려 8명에 달하였다.

◎ g-지수

Egghe(2006a; 2006b)는 일단 h-지수 산정에 포함된 인용빈도 상위 논문이 추가로 인용되더라도 지수값의 향상에 전혀 기여하지 못하는 점을 지적하고 이를 보완하는 g-지수를 제안하였다.

◎ 연구 생산성의 영향을 많이 받음

h-지수는 이론적인 최고값이 발표한 논문의 수이므로 많은 논문을 발표할수록 유리하다. van Raan(2006)은 화학 분야 연구팀에 대한 전문가들의 평판과 h-지수를 비교해본 결과 h-지수는 규모가 작은 연구팀에게 불리하게 나타났다고 보고하였다. 학술지에 대해서 측정할 경우에도 마찬가지로 수록 논문이 많은 학술지가 유리하게 된다.

주어진 논문집합에서 인용빈도 상위 g개 논문의 인용빈도 합이 g의 제곱 이상인 최하위 순위 g가 이 논문집합의 g-지수이다. 이 경우 상위 g+1개 논문의 인용빈도 합은 (g+1)의 제곱보다 작다.

풀어서 나타내면, 어떤 연구자의 논문을 인용빈도가 높은 순부터 나열하였을 때, 논문의 인용빈도 누적합계가 논문 순위의 제곱보다 크거나 같은 마지막 논문의 순위가 그 연구자의 g-지수값이 된다. 위 정의에서 논문집합이라는 표현을 쓴 이유는 h-지수나 g-지수를 저자 단위가 아닌 다양한 대상에 대해서 측정하는 상황을 고려했기 때문이다.

3.2 h-지수의 보완/변형 지수

◎ h_f 지수

Batista et al.(2005)은 h-지수의 수준이 분야마다 다를 수 있다고 지적하고, 그 원인이 상이한 공저 관례 때문이라고 주장하였다. 이들은 공저자 수를 고려하여 보정한 h_f 지수를 제안하였다. 브라질의 4개 기초과학분야에 대해서 측정해본 결과 h-지수와 달리 h_f 지수는 값의 분포가 분야마다 큰 차이 없이 거의 일치하는 것으로 나타났다. h_f 지수 공식은 다음과 같다.

예를 들어 g-지수가 20이라면, 인용빈도 상위 논문의 인용빈도 합이 20×20=400회 이상이면 21×21=441회보다는 작다는 뜻이다. g-지수 산출과정에는 상위권 논문의 갯수만 아니라 개별 인용빈도가 그대로 반영되므로 이를 무시하는 h-지수에 비해서 직관과 더 일치하는 결과를 얻을 수 있다고 Egghe는 주장하였다.

<표 1>의 가상 자료에 대해 적용해보면 연

구자 A, B, C, D의 h-지수는 각각 4, 4, 2, 2이지만, g-지수는 5, 4, 2, 5가 된다. h-지수가 똑같이 4인 연구자 A와 B, 그리고 같은 2가 되는 연구자 C와 D중에서 상위 논문의 인용빈도가 더 높은 연구자 A와 연구자 D가 각각 더 큰 g-지수를 가진다. 따라서 g-지수로 평가한 결과가 직관에 더 가깝다는 Egghe의 주장은 설득력이 있다. 단점이라면, 인용빈도를 그대로 반영하므로 연구자 D와 같이 단 한 편의 논문 때문에 높은 지수값이 산출되는 경우가 발생할 수 있다는 점이다.

◎ h_s-지수와 g_s-지수

h-지수는 일단 h위 이상에 자리잡은 논문의 인용빈도가 이후에 아무리 높아지더라도 지수값에 반영되지 않는 단점이 있다. 반면에 g-지수는 지나치게 최상위 논문의 인용빈도에 좌우된다는 약점이 있다. 이 연구에서는 두 경우의 절충방안으로, h위 이내 논문의 인용빈도의 제곱근을 합하여 산출하는 h_s-지수를 다음 공식과 같이 제안한다.

$$h_s = \sum_{r=1}^h \sqrt{C(r)}$$

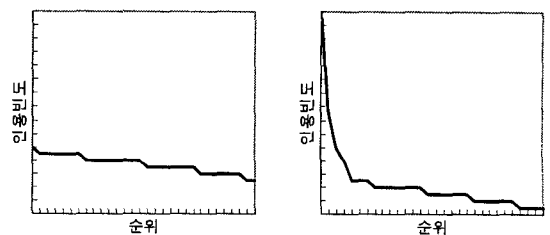
여기서 C(r)은 순위가 r인 논문의 인용빈도이다. g-지수처럼 인용빈도를 그대로 합산하지 않고 제곱근을 취해서 합하는 이유는 네트워크의 부익부 빈익빈 원리(Barabasi 2002) 때문이다. 일단 인용빈도가 높아서 주목을 받은 논문은 그렇지 못한 논문에 비해서 추가 인용되기 쉽다. 따라서 인용빈도가 두 배라는 것이 곧 두 배의 영향력이라고 받아들이기는 어렵다. h회 이상 인용된 논문의 인용빈도에 제곱근을 취하여 합산한 h_s-지수는, h-지수와 달리 최상위 논문의 인용빈도 차이를 지수값에 반영하면서도 g-지수처럼 한 두 논문에 의해 지수값이 좌우되는 상황은 방지하는 장점을 가진다. h_s-지수값은 h-지수값보다 항상 크거나 같으며 h가 1인 경우에만 같다.

<표 1>의 가상 자료에 적용해보면 연구자 A, B, C, D의 h_s-지수는 각각 10.0, 8.0, 3.5, 6.2이다. h-지수로는 동물이던 A와 B, 그리고 C와 D 사이의 우열이 각각 드러나므로 h_s-지수가 변별력이 높음을 알 수 있다. 또한 연구자 D의 경우에는 인용빈도가 최상위 논문에 집중된 덕분에 g-지수로는 연구자 A와 함께 공동 1위였지만, h_s-지수로는 네 명중 세 번째로 나타났다. 이와 같이 h_s-지수는 최상위 논문의 인용빈도를 제한적으로 반영함으로써 h-지수와 g-지수의 상반된 단점을 동시에 극복할 수 있다.

h_s-지수와 동일한 방식으로 g-지수를 개량한 g_s-지수를 측정할 수 있다. g_s-지수는 g-지수로 산출된 g보다 인용빈도 순위가 높거나 같은 논문의 인용빈도에 제곱근을 취하여 합하는 지수이다. h_s-지수보다 더 많은 논문의 인용빈도를 반영할 수 있다는 장점은 있으나, g값 자체가 최상위 논문의 인용빈도에 크게 좌우되므로 g_s-지수도 정도는 덜하지만 마찬가지로 문제점을 가진다.

◎ L-지수

h-지수는 <그림 1>에서와 같이 인용빈도 분포선과 원점에서 출발한 45도 선과의 교차점의 x좌표에 해당한다. 그런데 연구자의 논문이 인용된 실제 빈도 분포는 곡선이 아닌 계단식으로 감소할뿐더러 <그림 1>과 달리 <그림 2>처럼 특이한 인용빈도 분포를 나타내는 경우도 적지 않다.



<그림 2> 다양한 인용빈도 분포 형태

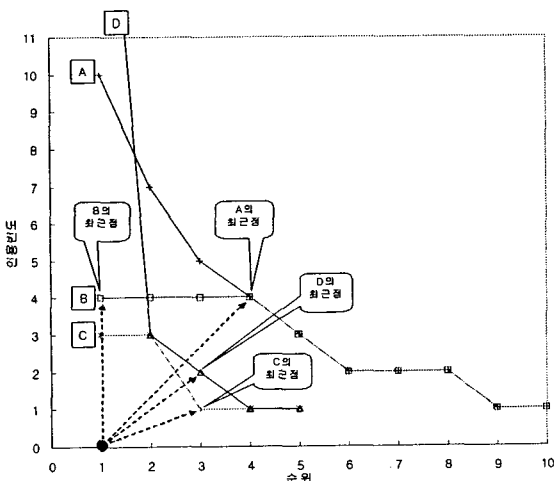
이 연구에서는 h-지수에서 사용된 45도 선을 버리고 각 연구자의 상이한 인용빈도 분포 형태를 반영할 수 있는 새로운 지수로 L-지수를 다음과 같이 제안한다.

특정 논문집합에 속한 논문을 인용빈도가 큰 순서로 정렬하여 순위가 r인 논문의 인용빈도를 C(r)이라고 하면, 이 논문집합의 L-지수는 각 점 (r,C(r))과 점 (1,0) 사이의 거리 L(r) 중에서 최단 거리이다. 거리는 유클리드 거리로서 다음 공식으로 산출한다.

$$L(r) = \sqrt{(r-1)^2 + C(r)^2}$$

풀어서 설명하자면, 각 논문의 인용빈도를 높은 것부터 그래프에 표시하였을 때 각 점과 좌표(1,0)과의 거리를 측정하여 가장 짧은 거리를 해당 논문집합(연구자)의 L-지수로 삼는다. 거리측정의 기준이 되는 원점으로 (0,0)이 아닌 (1,0)을 사용하는 이유는, 순위의 최저값은 1이고 인용빈도의 최저값은 0이기 때문이다.

앞의 가상 사례에 대해서 최근점을 파악하여 L-지수를 산출하는 예를 <그림 3>에 제시하였다. 이 그림에서 연구자 A, B, C, D의 최근점은 각각 (4,4), (1,4), (3,1), (3,2)이고 L-지수는 5.0, 4.0, 2.2, 2.8이 된다. 연구자 B는 연구자 A와 h-지수로는 같은 4였지만, h 지점에서 상위



<그림 3> L-지수 산출 예

로 올라가도 인용빈도가 증가하지 않아서 최근점이 원점 (1,0)과 더 가까워지는 결과가 되었다. 이와 같이 L-지수는 인용빈도의 분포 형태가 상이한 자료를 비교할 경우에 h-지수보다 공정한 산출이 가능하다는 것을 알 수 있다.

◎ h_A-지수

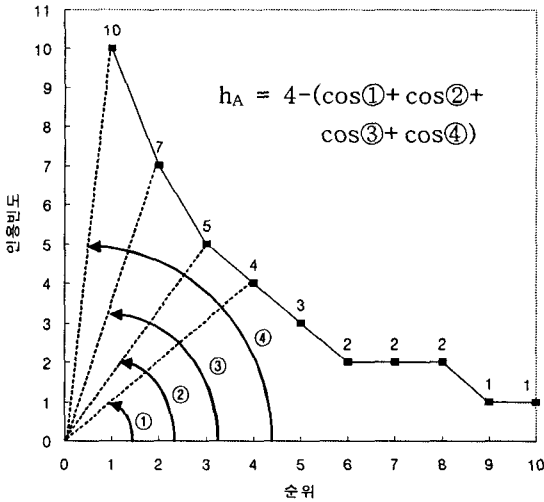
새로운 척도로 좌표나 거리가 아닌 각도와 코사인 계수를 이용한 지수인 h_A-지수를 다음과 같이 제안한다.

특정 논문집합에 속한 논문을 인용빈도가 큰 순서로 정렬하여 순위가 r인 논문의 인용빈도를 C(r)이라고 하면 각 논문을 점 (r,C(r))로 나타낼 수 있다. h-지수로 산출된 순위 h 이상인 각 논문마다 원점 (0,0)으로부터 각 논문 (r,C(r))을 잇는 직선과 x축 사이의 각도에 대한 코사인 값을 구한 후 1에서 뺀 다음 합산한 값 h_A가 이 논문집합의 h_A-지수이다. 이를 공식으로 나타내면 다음과 같다.

$$h_A = \sum_{r=1}^h \left(1 - \frac{r}{\sqrt{r^2 + C(r)^2}}\right) = h - \sum_{r=1}^h \frac{r}{\sqrt{r^2 + C(r)^2}}$$

코사인 값을 1에서 빼주는 이유는 각도가 클수록 코사인 값은 작아지기 때문이다. 1에서 뺀 코사인 값은 논문의 인용빈도가 0일 때 0부터 시작하여 인용빈도가 커질수록 1에 가까워진다. h_A-지수는 1 미만인 값을 h개 더하는 것이므로 h-지수보다 항상 작다.

<그림 4>는 앞의 가상 자료에서 연구자 A의 경우를 나타낸 것이다. 연구자 A의 h-지수는 4이므로, h_A-지수는 4위 이내 논문으로 산출한다. 인용빈도 1위부터 4위까지 논문의 좌표는 각각 (10,1), (7,2), (5,3), (4,4)이고 원점에서 각 점을 이은 선과 x축과의 각도는 <그림 4>의 ①, ②, ③, ④와 같다. 각각에 대해서 코사인을 취한 후 1에서 빼면 0.90, 0.73, 0.49, 0.29가 되고 이를 합한 2.40이 연구자 A의 h_A-



<그림 4> h_A-지수 산출 예

지수이다.

지금까지 살펴본 각 지수를 가상 자료에 포함된 네 연구자에 적용한 결과를 종합하면 <표 3>과 같다. 이 표에 나타났듯이 h_A-지수는 h_S-지수와 마찬가지로 연구자 A와 B 사이의 우열과 연구자 C와 D 사이의 우열을 각각 판별할 수 있으며, 연구자 D와 같이 한 논문 때문에 지나치게 높은 값을 가지는 것을 방지하는 특성이 있다.

4. 실제 자료에 대한 측정 결과

앞에서 살펴본 각 지수를 다양한 실제 자료에 대해 적용해본 다음 지수간 상관관계를 분석해보았다. 조사는 개인, 학과, 대학, 학술지를 단위로 수행하였는데, 지수의 특성 면에서는 차이가 거의 없었으므로 이 글에서는 개인 연구자에 대한 산출 결과만 보고한다.

계량정보학 분야의 최고 권위인 Price 메달 수상자에 대해서는 Glänzel & Persson(2005)이 1986년 이후 인용 자료를 수집해서 분석한 바 있다. Egghe(2006b)는 1972년 이후의 인용 자료를 수집하여 더 완전한 분석을 수행하였는데, 여기에 포함된 14명에 대한 인용빈도 자료

<표 3> 가상 자료에 대한 각 지수 산출 결과

	연구자 A의 논문	연구자 B의 논문	연구자 C의 논문	연구자 D의 논문
인용빈도합계	37	27	9	27
인용빈도평균	3.7	2.7	1.8	5.4
h	4	4	2	2
g	5	4	2	5
L	5.0	4.0	2.2	2.8
h _S	10.0	8.0	3.5	6.2
g _S	11.8	8.0	3.5	9.6
h _A	2.40	2.00	1.13	1.40

를 이용하여 앞에서 다룬 여러 지수를 산출해 보았다.

<표 4>에 14인에 대한 지수 산출 결과를 제시하였다. 각 지수의 특성을 비교하기 위해서 인용빈도 1위 논문의 인용빈도(max)와 연구자별 상위 16위 이내 논문의 인용빈도 합계(sum16)를 제시하였다. 16위 이내 합계를 산출한 이유는 Egghe(2006b)에 수록된 각 저자별 인용빈도 리스트 중에서 가장 짧은 경우가 16위 논문까지였기 때문이다.

<표 4>에서 지수에 따라 순위가 가장 달라지는 저자는 H. White이다. White는 h-지수로는 12로 가장 낮은 값이지만, g-지수, h_S-지수, g_S-지수, h_A-지수로는 Leydesdorff, Egghe, Rousseau의 세 연구자를 제치는 것으로 나타났다. White의 상위 논문 인용빈도 합계(sum16)가 다른 세 연구자의 두 배 내외에 이르는 것을 감안할 때 h-지수는 상위 논문의 개

<표 4> Price 메달 수상자 14인의 지수 산출 결과

연구자	max	sum16	h	g	h _S	g _S	L	h _A
Garfield	625	2364	27	59	270.8	391.4	36.9	21.5
Narin	112	1063	27	40	190.5	243.9	35.6	18.2
Braun	125	913	25	38	165.7	227.3	33.6	16.5
van Raan	108	554	19	27	104.2	136.4	24.2	12.0
Small	305	1393	18	39	146.9	193.8	23.3	13.4
Schubert	124	714	18	30	111.1	158.9	24.8	12.2
Glänzel	124	555	18	27	99.8	136.9	25.5	11.9
Moed	108	614	18	27	104.7	136.7	24.2	12.0
Martin	156	634	16	27	95.5	133.7	22.6	11.1
Ingwerson	120	646	13	26	84.3	114.6	16.4	9.5
Egghe	47	345	13	19	61.5	82.5	17.0	8.4
Leydesdorff	79	351	13	19	61.7	81.6	17.0	8.4
Rousseau	25	238	13	15	51.0	57.9	15.3	7.6
White	128	601	12	25	74.9	111.5	17.0	8.6

<표 5> 개인 자료에 대한 지수간 스피어맨 순위상관

	h	g	h _s	g _s	L	h _A
max	0.333	0.675	0.597	0.628	0.419	0.597
sum16	0.646	0.908	0.881	0.873	0.632	0.868

별 인용빈도를 무시하는 특성 때문에 White에 대해서 과소평가하였다고 판단된다. 정도는 덜 하지만 L-지수도 이와 유사한 성향을 보였다.

<표 5>의 지수간 순위상관 분석 결과에서도 흥미로운 사실이 드러난다. 최댓값과의 상관관을 보면 여섯 가지 지수 중에서 g-지수와 g_s-지수가 0.6대로 높았고, 그중에서도 특히 g-지수는 유일하게 99% 수준에서 최댓값과 유의한 상관관이 있는 것으로 나타났다(유의수준 99.2%). 앞에서 살펴본 바와 같이 g-지수가 인용빈도를 그대로 합산한 결과로 최상위 논문의 인용빈도에 크게 좌우된다는 점이 다시 확인되었다. 또한 g_s-지수도 이런 g-지수의 단점을 어느 정도 공유하고 있음이 드러났다(98.4% 유의수준에서 최댓값과 상관 있음). 반면에 합산을 하지 않는 h-지수와 L-지수는 95% 유의수준에서도 최댓값과 상관관이 없는 것으로 나타났다.

5. 결론

최근 제안된 h-지수와 이를 변형한 g-지수는 짧은 기간에도 불구하고 다양한 응용 사례가 발표되고 있다.

이 연구에서 두 지수를 사례와 함께 검토해 본 결과 h-지수는 상위 논문의 인용빈도가 높은 연구자에 대한 과소평가의 문제가 나타났고, g-지수는 최상위 논문의 인용빈도에 크게 좌우된다는 단점이 드러났다.

이런 단점을 보완하기 위해서 네 가지 새로운 지수를 제안하였으며, 그중에서도 h_s-지수와 h_A-지수는 기존 지수의 단점을 극복할 수 있는 것이라고 기대된다. 향후 다양한 자료에 대해서 실측을 수행하여 제안한 지수의 타당성과 응용가능성을 검증할 계획이다.

참고문헌

- Ball, P. 2005. Index aims for fair ranking of scientists. *Nature* 436, 900. <<http://www.nature.com/nature/journal/v436/n7053/full/436900a.html>>.
- Banks, M. G. 2006. An extension of the Hirsch Index: Indexing scientific topics and compounds. To be published in *Scientometrics*. Also available from <<http://arxiv.org/abs/physics/0604216>>.
- Barabasi, Albert-Laszlo. 2002. *Linked: The New Science of Networks*. Perseus Books Group.
- Bar-Ilan, J. 2006. H-index for Price medalists revisited. *ISSI Newsletter*, 2(1): 3-5.
- Batista, P. D., M. G. Campiteli, O. Kinouchi, and A. S. Martinez. 2005. An index to quantify an individual's scientific research valid across discipline. Submitted to *Scientometrics*. Also available from <<http://arxiv.org/abs/physics/0509048>>.
- Bornmann, L., and H.-D. Daniel. 2005. Does the h-index for ranking of scientists really work? *Scientometrics*, 65(3): 391-392.
- Braun, T., W. Glänzel, and A. Schubert. 2005. A Hirsch-type index for journals. *The Scientist*, 19(22): 8.
- Cronin, Blaise, and Lokman Meho. 2006. Using the h-index to rank influential information scientists. *Journal of the American Society for Information Science and Technology*, 57(9): 1275-1278.
- Egghe, L. 2006a. An improvement of the H-index: the G-index. *ISSI Newsletter*, 2(1), 8-9.
- Egghe, L. 2006b. Theory and practise of the g-index. To be published in *Scientometrics*.
- Glänzel, W., and O. Persson. 2005. H-index for Price medalists. *ISSI Newsletter*, 1(4): 15-18.
- Hirsch, J. E. 2005. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102: 16569-16572.
- Lehmann, S., A. D. Jackson, and B. E. Lautrup. 2006. Measures and Mismeasures of Scientific Quality. [online]. [cited 2006.6.30]. <<http://arxiv.org/abs/physics/0512238>>.
- Rousseau, Ronald. 2006a. A case study : evolution of JASIS' Hirsch index. [online]. [cited 2006.6.28]. <<http://eprints.rclis.org/archive/00005430/>>.