

# 논문 영향력 예측을 위한 저자별 논문의 피인용 패턴 분석

## An Analysis on the Pattern of Citation in Articles for the Prediction of Article Impact

이혜진, 한국과학기술정보연구원, hyejin@kisti.re.kr

이춘실, 숙명여자대학교 문헌정보학과, cslee@sookmyung.ac.kr

Hyejin Lee, Korea Institute of Science and Technology Information

Choon-shil, Lee, Sookmyung Women's University

본 연구는 최근에 발표된 단위 논문별 영향력을 예측하기 위해 해당 저자가 이전에 발표했던 논문들의 년도별 피인용 추이를 분석하였다. 최근 IF가 가장 높은 "Cell & Tissue Engineering" 분야와 그 분야에서 IF가 가장 높은 저널 "Stem Cell"을 선정하여 최근에 실린 논문의 저자정보를 파악하고, 해당 저자의 이전에 발표한 논문들을 추출하였다. 분석 결과, 저자가 발표한 논문들이 최근에 인용이 많이 일어나면 이후 발표한 논문의 인용빈도가 높아지는 것으로 나타났으며, 특히 동일한 h-index를 가진 저자들 간의 비교에서 더욱 두드러지게 나타났다.

### 1. 서론

최근까지 학술적 연구성과의 질을 평가하기 위해 성과에 대한 영향력을 측정하는 다양한 시도가 이루어지고 있다. 전통적으로 영향력은 출판된 논문의 피인용 빈도를 기반으로 하고 있으며, 대표적인 영향력 지수로 JIF(Journal Impact Factor), JII(Journal Immediacy Index), h-index, g-index 등이 있다. 그 중 가장 보편적인 영향력 측정 지수인 JIF는 학술지 단위의 인용정보임에도 불구하고 각 학술지에 게재된 단일 논문과 해당 논문의 저자 영향력 측정에 확장하여 적용함에 따라 많은 문제점들이 지적되고 있다(김관준, 이재윤, 2010). 따라서 개별 논문과 연구자의 연구성과 측정을 위하여 h-index를 활용하는 사례가 점차 증가하고 있다. 특히 저자의 영향력 측정은 해당 저자가 발표한 논문의 우수성을 판단할 수 있는 근거로써 활용도가 크다.

h-index는 2005년 Hirsch가 제안한 지수로 기존의 영향력 지수의 한계점을 극복하기 위해 개발되었으며, 이 후 Web of Science(이하 WOS)에서 새로운 인용지수로 채택됨으로서 연구성과 측정을 위한 보편적인 지표로 받아들여지고 있다. h-index는 지속적으로 상당수의 논문을 발표하고, 그 중 다수의 논문들이 인용이 많이 되어야 높은 지수를 받을 수 있다는 점에서 질과 양적인 측면을 모두 고려하여 평가한다는 장점을 가지고 있다. 하지만 서로 다른 분야를 비교하는 용도로 사용될 수 없고, 한 가지 수치로 개별 연구자를 대략적으로 추정할 수밖에 없다는 단점이 있다. 그럼에도 불구하고, 지수가 매우 직관적이고, 기관, 국가의 연구성과 등 다양한 성과 측정에 활용 가능하기 때문에 인정을 받고 있는 지수이다.

최근에는 h-index의 단점을 보완하기 위하여 인용네트워크 분석, 년도별 분석, 공저자수를 반영한 분석 등을 활용하여 h-index를

변형하거나 응용한 지표들이 꾸준히 늘어나고 있다.

본 연구에서는 최근에 발표된 논문의 영향력을 예측하기 위하여 h-index를 보완할 수 있도록 해당 저자가 이전에 발표한 논문의 피인용 빈도에 대한 시간적 추이를 분석하였다. 동일한 h-index를 가진 저자라고 하여도 시간에 따라 논문의 인용도가 증가하거나 감소하는 특성을 보이게 되므로 시간적 추이를 기반으로 논문의 영향력을 예측하는 것이 가능할 것으로 판단된다.

## 2. 연구대상 및 방법

본 연구는 대상 주제분야와 저널 선정, 저자추출, 해당 저자의 이전 발표논문 년도별 인용빈도 추출, 저자의 h-index 추출, 년도별 피인용 패턴 분석의 과정으로 진행하였다.

먼저, JCR(Journal Citation Reports)에서 최근 3년(2010년부터 2012년) 평균 Impact Factor(이하 IF)가 높으면서 증가 추세를 보이는 주제분야를 선정하였다. 결과, "Cell & Tissue Engineering" 분야이며 IF 값은 2010년 3.38, 2011년 3.72, 2012년 3.87이다. 대상 저널은 해당 주제 중 IF가 두 번째로 높은 저널인 "Stem Cell"로 하였다. IF가 가장 높은 저널은 "Cell Stem Cell"이었으나, 2013년 논문이 아직 많이 발표되지 않았기 때문에 본 연구의 대상에 포함시키지 않았다. WOS에서 "Stem Cell" 저널에 2013년에 발표된 논문은 2013년 6월말 기준 총 76건으로, 76명의 1저자가 2006년부터 2012년까지 7년 동안 발표한 논문을 추출하였다. 2006년을 시작년도로 설정한 이유는 이 분야에서 유도만능줄기세포가 세계 최초로 만들어지면서 전세계가 줄기세포 연구에 박차를 가한 시기이기 때문이다. 분석결과의 오류를 방지하기 위하여 발표 논문이 5건 이하인 저자는 분석에서 제외하였

다. 최근 논문의 1저자가 발표한 논문은 총 1,552건이었으며, 논문별 평균 총 인용빈도는 15.6이고, h-index의 평균은 6이다.

본 연구는 76명의 저자가 "Stem Cell"에 발표한 논문의 영향력을 예측하는 것이므로 분석 대상은 76명이 발표한 논문이며, 분석 범위는 논문들의 년도별 피인용 빈도이다. 년도별 피인용 패턴을 살펴보기 위해 본 연구에서는 저자의 년도별 피인용빈도와 빈도수 랭킹을 측정하였다. 년도별 평균 피인용빈도는 해당저자의 년도별 피인용 빈도를 발표논문 누적 건수로 나누었으며, 빈도수 랭킹은 년도별 평균 피인용빈도수를 기준으로 오름차순으로 저자별로 순위화하였다.

## 3. 분석 결과

### 3.1 저자별 발표 논문 현황 분석

저자별 발표논문 분석 결과, 평균 발표논문 수는 20.4건이고, 2006년 이후로 가장 많은 논문을 발표한 저자는 총 102건을 발표했으며, h-index는 18로 나타났다. 가장 높은 h-index를 가진 저자는 20으로 발표논문은 총 43건으로 발표 논문빈도 순위는 4위이다.

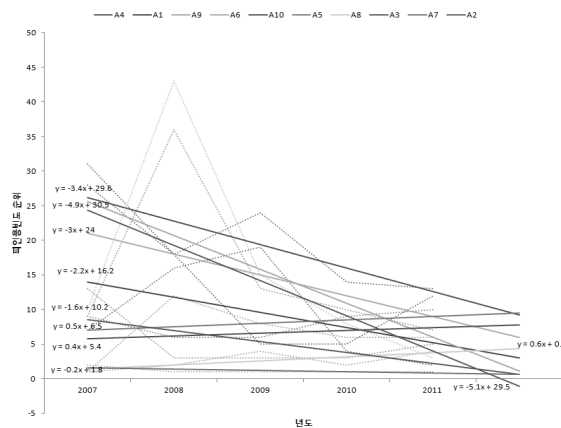
년도별 발표논문 분석 결과, 2011년에 가장 많은 논문발표가 이루어졌으며, 저자별 평균 발표 논문수도 가장 높게 나타났다.

<표 1> 년도별 발표논문 현황

	'06	'07	'08	'09	'10	'11	'12
발표 논문수	103	141	177	211	226	262	251
발표 저자수	38	45	60	59	67	64	62
평균 발표 논문수	2.7	3.1	3.0	3.6	3.4	4.1	4.0

### 3.2 년도별 피인용 빈도 분석

년도별 피인용 빈도 분석은 2006년부터 2010년까지의 피인용빈도 추이를 통해 2011년, 2012년 피인용빈도와 순위를 추정하여 실제 피인용빈도와 순위를 비교하였다. 순위 추정은 Microsoft Excell 2007의 선형 추세를 활용하여 2011년도와 2012년도의 피인용빈도를 산출하였다. 피인용빈도수는 데이터 값의 제한이 없기 때문에 정규화를 위해 저자별 피인용빈도를 순위화하였다.



<그림 1> 년도별 피인용 빈도 순위 추세선

분석 결과, 2011년, 2012년 추정값과 실제 값 사이에는 각각 0.92, 0.9의 강한 양의 상관관계가 나타나 최근 논문이 많이 인용될수록 이후 논문은 많이 인용되는 것으로 나타났다. 2012년 피인용빈도 순위 상위 10위 내의 저자의 선형 추세를 살펴보면 <그림 1>과 같다. 순위에 대한 추정이기 때문에 기울기가 음수이면 순위가 점점 높아지는 것을 의미한다. 10명의 저자 중 음의 기울기가 큰 저자는 A2(-4.9), A4(-3.4), A6(-3)이며, 양의 기울기가 발생한 저자는 A8(+0.6), A5(+0.5), A10(+0.4)이다. A4와 A6의 경우 2012년도의 피인용빈도 순위가 7위이지만 최근 들어 상승세를 보이고 있는 저자이며, h-index는

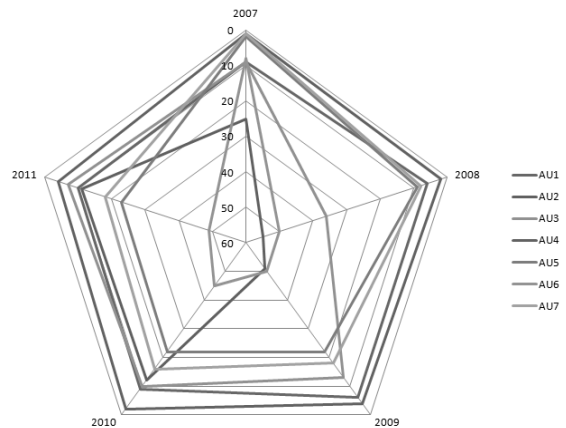
각각 5, 11이다. A5와 A10의 경우는 각각 9위와 10위로 최근 들어 논문의 피인용빈도가 낮아지는 저자였으며, h-index는 11, 6이다.

<표 2> 년도별 피인용 빈도 순위

	2007	2008	2009	2010	2011	2012
A1	7	16	19	4	2	3
A2	31	18	5	5	12	5
A3	13	3	3	3	5	2
A4	28	18	24	14	13	7
A5	9	6	6	9	10	9
A6	9	36	13	10	7	7
A7	2	1	1	1	1	1
A8	1	2	4	2	4	4
A9	10	43	15	8	3	5
A10	1	12	8	6	6	10

### 3.3 동일 h-index 저자별 피인용 패턴 비교

다음으로는 동일한 h-index를 갖는 저자들 간에 시간에 따라 인용이 어떻게 다르게 나타나고, 그것이 차후 인용에 어떤 영향을 줄 수 있을지를 분석하였다.

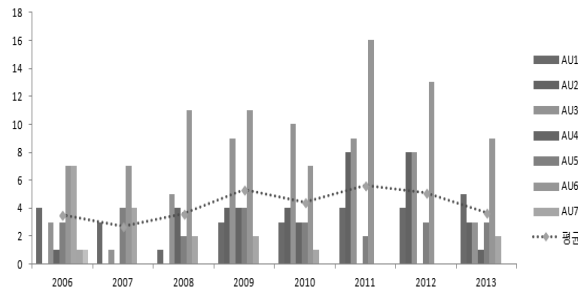


<그림 2> 년도별 피인용 패턴(동일 h-index를 가진 저자)

h-index가 11인 저자들의 년도별 피인용 패턴을 분석한 결과는 <그림 2>와 같다. h-index

11인 저자들을 선정한 이유는 본 연구 대상에서 h-index 11인 값을 가장 많은 저자들이 가지고 있었기 때문이다.

분석 결과, AU6의 논문들은 최근보다 과거에 많이 인용되었으며, AU2의 논문들은 최근에 인용이 많이 되는 것으로 나타났다. <그림 3>은 저자별로 논문발표 건수를 평균과 비교하여 나타낸 표로 AU6은 매년 많은 논문을 발표하고 있지만 다른 저자들의 논문들에 비해 인용이 많이 일어나지 않은 것으로 파악할 수 있다. AU2는 최근에 들어 평균 이상의 논문을 발표하는 것으로 나타났다. AU5는 년도별로 논문을 고르게 발표하고 있지만 논문들의 인용빈도에 있어서는 현재보다 과거에 논문인용이 많이 일어났다. AU3은 다른 저자들보다 비교적 많이 발표하면서 해당 논문들이 최근에 많이 인용되는 것으로 나타났다.



<그림 3> 년도별 논문 발표건수

2006년부터 2010년까지의 피인용 빈도 순위 패턴과 2012년 피인용빈도 순위를 비교하여 살펴본 결과, 최근 인용과 발표논문이 많은 AU2와 AU3은 76명 저자 중 각각 11위, 7위를 차지하였으며, 과거에 인용이 많이 된 AU5와 AU6은 각각 34위, 55위를 차지하는 것으로 나타났다. 이 중 특이한 점은 대부분 저자의 경우 논문 생산량이 높으면 피인용 빈도가 높게 나타났으나, AU6의 경우는 최근 2년 동안 논문 생산량이 매우 높은 저자이지만 최근 피인용 빈도가 매우 낮은 것으로 나타났다.

#### 4. 결론 및 고찰

연구 결과, 저자가 발표한 논문이 최근에 인용이 많이 일어나고 있다면 이후 발표한 논문은 인용이 많이 될 수 있음을 추정할 수 있었다. 이러한 결과는 동일한 h-index를 가진 저자들 간의 비교에서 더욱 두드러지게 나타났다. 하지만 h-index가 매우 높거나 피인용 빈도 순위가 꾸준히 높은 저자의 경우 추세선이 완만하므로 예외로 나타났다.

본 연구는 특정 주제분야의 저널을 대상으로 분석하였기 때문에 연구 결과에 대한 신뢰성 확보와 구체적인 논문별 인용예측을 위해 대상 데이터를 확대하고, 분석 방법에 있어서 순위별, h-index별 저자를 그룹화하여 특성을 분석하거나 예측 공식을 정교화하는 추가적인 연구가 필요하다.

#### 참고문헌

김관준, 이재윤. 2010. 학술지 영향력 측정을 위한 h-지수의 응용에 관한 연구. 『정보관리학회지』, 27(1): 269-287.

박지연 외. 2010. 기관단위 연구성과 평가방법에 관한 연구: h-지수 및 변형지수를 중심으로. 『정보관리학회지』, 27(1): 249-267.

Egghe, L. 2010. "The Hirsh-index and related impact measures." In *Blaise Cronin(ed.), Annual Review of Information Science and Technology*, 44: 65-114.

Hirsch, J. E. 2005. "An index to quantify an individual's scientific research output." *Proceedings of the National Academy of Science*. 102: 16569-16572.