

연구성과의 질 제고를 위한 논문평가 모형개발

이혁재* · 여운동** · 이상필***

1. 서론

최근 정부가 “국가연구개발사업등의성과평가및성과관리에관한법률(이하 성과평가법)”을 제정하는 등 연구개발의 성과평가에 대한 관심이 높아지고 있다. 이는 국가연구개발에 대한 예산의 증가와도 무관하지 않다. 조사·분석·평가 대상 국가연구개발사업 투자 총액은 1999년의 3조 7,067억 원에서 2004년의 7조 827억 원까지 13.8%의 연평균 증가율로 꾸준히 증가하였다(국가과학기술위원회, 2005). 이와 같이 정부의 연구개발에 대한 예산이 급격한 증가추세를 보이고 있는 가운데 막대한 공적 자금이 어떠한 목적으로 어떠한 경로를 통하여 쓰이고 있으며, 자금투입의 결과는 어떠한지에 대한 관심이 증대되는 것은 당연한 일이라고 할 수 있다. 특히 최근에는 연구개발의 결과물을 가시화하고 이의 질을 평가하려는 다양한 시도가 이루어지고 있다.

정부의 연구개발 관련 예산의 증가와 더불어 한국이 생산해 낸 논문성과물 역시 성장세를 보이고 있다. 최근 10년간 한국이 발표한 논문 수의 추이를 살펴보면(한국과학기술원, 2005) Thomson Scientific의 NSI(National Science Indicator) 데이터베이스에 수록된 2004년의 논문 중 한국이 발표한 논문의 양은 19,279편이며 이것은 세계 총 논문 수인 787,677 편에 해당하는 수치이다. 이는 2003년 대비 2.62%(492편)가 증가한 수치이다. 또한 최근 10년간 세계 순위는 1995년의 23위에서 2004년에는 14위로 상승하였다. 세계 논문 점유율 역시 1995년의 0.68%에서 2004년에는 1.96%로 꾸준한 증가를 보이고 있다. 이와 같은 논문성과물의 양적인 성장의 상당 부분은 정부의 연구개발 관련 예산의 증가와도 관련이 있다고 볼 수 있다. 최근 정부의 “성과평가법” 제정은 이와 같은 연구개발의 양적인 성장과 함께 질적인 성장을 도모하기 위한 정부의 노력을 잘 나타내는 것이라 할 수 있다.

이러한 정부의 의지를 반영하여 2005년 12월에는 표준성과지표를 발표하였으며(과학기술부, 2005) 이는 2006년 국가연구개발사업 조사·분석·평가 추진 계획에도 반영되었다. 국가연구개발사업 표준성과지표에서 발견할 수 있는 흥미로운 사실은 국가연구개발사업을 바라보는 모든 관점에 대해 논문성과가 포함되어 있다는 점이다. 이는 연구개발이 그 특성상 많은 부분이 과학적 지식의 창조에 그 가치를 부여하며 따라서 다양한 연구개발의 결과물 중 학술지에 게재되는 논문이 가장 일차적이면서 동시에 가장 중요한 결과물로 고려되고 있다는 사실을 의미한다. 연구개발 자체를 목적으로 하며 결과활용 측면에서는 전반적인 지식증진을 목적으로 하는 경우의 연구개발사업은 논문성과가 가장 핵심적인 지표이다. 또한 다른 목적을 가지는 연구개발사업의 유형에 대해서도 핵심지표를 제외한 다른 성과지표들 중에서 논문성과가 대부분 가장 높은 순위를 차지하고 있다. 이러한 사실은 국가연구개발사업 전반에 걸쳐 논문성과의 중요성을 대변해 주는 명백한 근거임을 의미한다.

현재 국내외에서 논문성과물의 질을 평가하기 위하여 특정 논문(또는 논문 집단)의 피인용 수가

* 이혁재, 한국과학기술정보연구원 선임연구원, 02-3299-6059, hlee@kisti.re.kr

** 여운동, 한국과학기술정보연구원 연구원, 02-3299-6017, wdyeo@kisti.re.kr

*** 이상필, 한국과학기술정보연구원 책임연구원, 02-3299-6066, splee@kisti.re.kr

널리 사용되고 있다. 이는 논문게재를 연구개발의 산출물(output)로, 인용(citation)을 타 연구자들에 의한 연구결과의 인식(recognition)으로 보는 철학에 기본을 둔다. 논문의 인용에 대한 정량적 지표로는 단순한 피인용수로부터 Thomson ISI사에서 제공하는 학술지별 영향지수(Impact Factor, IF) 등이 있다. 이들 지표 중 학술지의 영향지수는 기준 연도 이전의 2개년 동안 특정 학술지에 게재된 논문이 기준 연도에 인용된 횟수를 동일 2개년 동안 게재된 논문의 총 수로 나눈 값을 의미한다. 비록 이 지표가 연구성과의 질을 판단하기 위해 사용되고 있기는 하지만 애초에 이 지표는 학술지에 수록된 논문의 “평균적” 인용빈도로부터 학술지의 순위 선정, 평가, 비교를 목적으로 만들어진 것이다(Garfield, 1994). 또한 이 지표는 국가별, 연도별, 분야별로 차별화 된 속성을 보인다(Moed et al, 1984; MacRoberts et al, 1989). 이로 인해 영향 지수로부터 연구성과물의 질을 평가하기 위해서는 이에 대한 사전 지식이 반드시 필요하며, 통상적으로 활자화된 연구성과가 적거나 요구하는 형태로 산출물을 내기 어려운 분야에서는 분야별 조정이 없는 일방적인 평가가 대단히 불리한 제도로 작용할 수 있다는 점을 염두에 두어야 한다(이상필 등, 2005). 그러나 이러한 분야별 조정은 위에서 언급한 국가 최상위의 과학기술관련 평가인 국가연구개발사업의 표준성과지표에서도 반영되지 않고 있다.

2005년 한국기술혁신학회에서 개최한 원탁회의에서 학문적으로도 연구기획과 연구성과 평가에 대한 새로운 패러다임의 필요성이 주장된 바가 있다.¹⁾ 상기 원탁회의에서는 과학기술 연구현장의 연구자들에 대해 연구기획평가의 올바른 패러다임이 정착되기 위해서는 분야별 특수성이 반영되어야 한다는 점을 강조하였으며, 한국기술혁신학회 회장에게 분야별 특수성을 공식화해줄 것을 요청한 바 있다. 한국기술혁신학회는 상기 내용에 대한 공식적인 문서를 과학기술혁신학회에 전달한 바 있다. 실제로 현재 수행되고 있는 대부분의 논문성과물 평가에 있어 이와 같은 분야별 특수성이 전혀 고려되지 못하고 있는 실정이다(한국기술혁신학회, 2006). 따라서 빠른 시일 내에 이러한 분야별 특수성에 대한 연구와 함께 적절한 논문 평가 지표를 개발하여 가이드라인으로 제시할 필요가 있다.

본 연구에서는 위에서 언급한 바와 같은, 논문의 질적 평가에서 발생하는 문제점을 해결하기 위해 분야별 특수성이 반영된 논문성과 평가지표를 제시하고 이를 적용할 때 고려하여야 할 사항에 대하여 언급하고자 한다.

II. 본론

1. 기본방향

본 연구에서는 논문성과의 질적 평가를 위한 일련의 정량적 지표를 제공하고자 하였다. 이를 위해 JRC(Journal Citation Report)의 자료를 기반으로 한 실증분석을 통해 현재 사용되고 있는 영향지수 등의 지표가 분야별로 큰 편차가 있음을 보이고자 하였다. 이들 지표에 분야별 특수성을 고려할 수 있는 요소들을 첨가하여 분야간 비교·평가가 가능한 지표를 산출하였으며, 아울러 논문성과의 국제적 비교·평가가 가능하도록 지표를 설계하였다. 지표 산출에 사용되는 자료는 Thomson Scientific의 데이터베이스를 활용하였다.

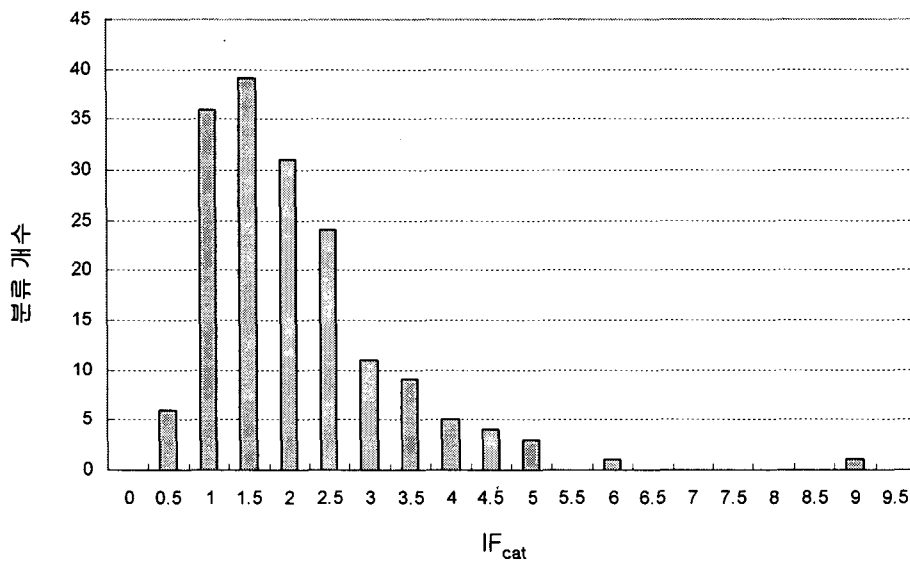
2. 분야별 특수성에 대한 실증분석

1) 한국과학기술단체총연합회, 2005/4/22.

1) JCR 분석

2004년 JCR에는 총 5,968개의 학술지가 포함되어 있으며, 이들은 170개의 세부분류(이하 JCR 분류)로 나뉜다. <그림 1>은 각 학술지를 JCR 분류별로 구분하고, 해당 분야에 속하는 학술지에 포함된 모든 논문들의 피인용수를 분야별 총 논문수로 나누어서 JCR 분류별 IF인 IF_{cat} 를 구한 후 IF_{cat} 별 분류 개수의 분포를 나타낸 그림이다.

<그림 1> JCR 분류로 재구성한 IF_{cat} 별 분류개수



분석 결과 가장 높은 IF_{cat} 값을 나타내는 분류는 Multidisciplinary Science로서 8.99의 IF_{cat} 값을 가지며, 가장 적은 값을 가진 분류는 Engineering-Marin로 0.14의 IF_{cat} 값을 가지므로, IF_{cat} 의 최대 격차가 8.85인 것으로 파악되었다. 또한 그림에서 나타난 학술지의 IF_{cat} 분포곡선은 정규분포가 아니며 낮은 IF_{cat} 에 다수의 분류가 집중되어 있음을 알 수 있다. 예를 들어 IF_{cat} 가 1-1.5인 구간에는 39개의 분류가 존재하는 반면 8.5-9인 구간에는 1개의 분류만이 존재한다. 이와 같이 정규분포가 아닌 IF_{cat} 의 특성은 평가지표로서의 문제점을 무엇보다도 잘 드러내 주고 있다. 따라서 아무런 조정없이 IF가 학제 분야간 평가에서 성과지표로 이용될 때는 높은 IF를 가진 특정 분야가 상대적으로 적은 노력으로 우수한 평가를 받을 수 있게 되는 불평등을 초래할 수 있다.

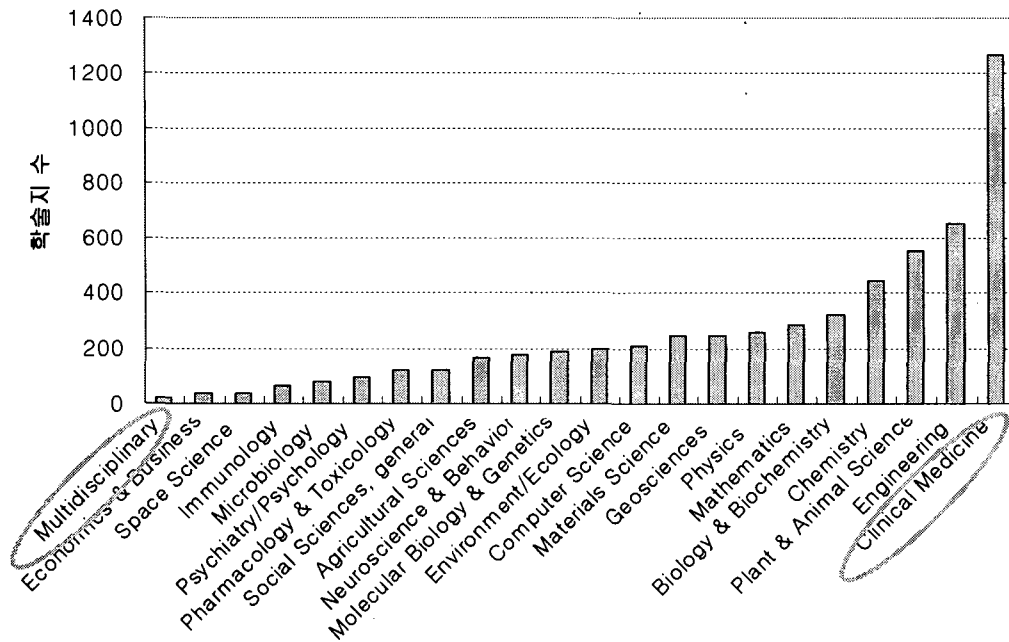
2) JCR-ESI 연계 분석

(1) 분류별 학술지수

서로 다른 연구 분야의 연구자가 자신의 연구 결과를 SCI 등재 학술지에 발표하고자 하는 경우를 가정할 때 학술지가 많은 분야의 경우가 그렇지 못한 경우에 비하여 논문 발표가 더 쉬울 것이라는 사실은 쉽게 예상할 수 있다. 뿐만 아니라 학술지 수가 많게 되면 그만큼 특정 논문이 인용될 가능성도 높아지게 된다. 본 연구에서는 이와 같은 분류별 학술지 수의 편중 현상을 분석하였다. 이를 위해 현재 22개의 대분류를 가지고 있는 ESI(Essential Science Indicators) 데이터베이스와 JCR 데이터베이스를 연동하여 분석하였다. 다시 말해 특정 ESI 분류에 속하는 학술지에 대한 논문 수 및 피인용 수 등의 정보를 JCR 데이터베이스에서 추출하여 활용하였다.

<그림 2>는 ESI 분류별 학술지 수를 분석한 그림이다. 분석 결과 Multidisciplinary, Space Science, Immunology 등은 상대적으로 학술지 수가 적은 것으로 나타났으며, Chemistry, Plants & animal Science, Clinical Medicine 등은 학술지수가 많은 것으로 나타났다. 특히 Clinical Medicine은 전체 학술지 중 22%를 차지하여 학술지 수에 있어 분야 편중이 심각한 것을 알 수 있다. 학술지 수가 가장 많은 Clinical Medicine과 가장 적은 Multidisciplinary의 차이는 1,245 개 (66.5배)로 나타났다.

<그림 2> ESI 분류별 학술지 수



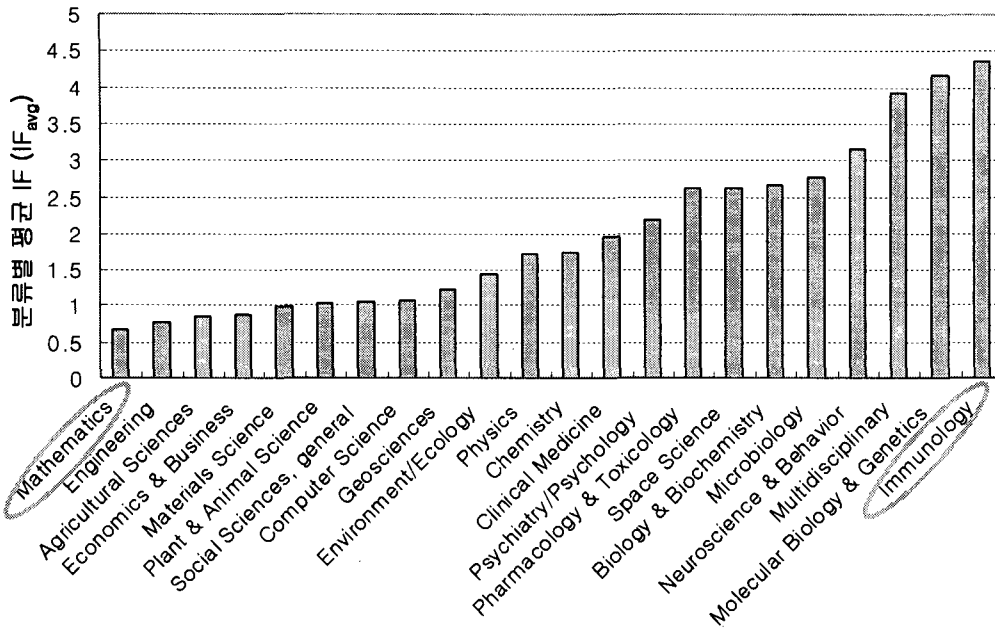
(2) ESI 분류별 평균 IF(IF_{avg})

본 연구에서는 또한 ESI 분류에 속하는 학술지들의 평균 IF를 산출하였다. 이를 위해 ESI 분류에 해당하는 학술지들에 대한 정보를 JCR 데이터베이스에서 추출하였으며 이들 학술지의 평균 IF를 산출하였다.

분류별 IF를 산출하는 데에는 두 가지 방법을 고려하였다. 하나는 ESI 각 분류에 속하는 모든 학술지들의 IF를 합하여 그 학술지들의 수로 나누는 방법이고, 또 다른 방법은 JCR에서 제공하는 학술지별 IF를 구하듯 ESI 각 분류에 속하는 모든 학술지의 논문들이 가진 전체 피인용수를 전체 논문수로 나누는 것이다. 본 연구에서는 전자를 IF_{avg} 로 하고, 후자를 IF_{cat} 로 표현하였다.

<그림 3>은 IF_{avg} 를 ESI 분류별로 나타낸 그림이다. 그림에서 알 수 있듯이 Mathematics, Engineering, Agricultural Science 등은 낮은 IF_{avg} 값을 나타내고 있으며, Immunology, Molecular Biology & Genetics, Multidisciplinary 등이 높은 IF_{avg} 값을 보이고 있다. 주로 응용학문은 IF_{avg} 가 낮으며 바이오관련 학문은 높은 IF_{avg} 를 나타내고 있으나 분야별 특성을 정확히 파악하기 위해서는 보다 면밀한 분석이 필요하다. 주목할 점은 앞서 분석에서 학술지 수가 적은 것으로 나타났던 Immunology가 큰 IF_{avg} 값을 보이는 것이다. 이는 Immunology가 가지는 학문적인 특성과 함께 인용패턴의 특성이 반영된 결과로 추정된다. Multidisciplinary의 경우는 Immunology처럼 IF_{avg} 값과 학술지수의 순위 차이가 많이 나는데, 이것은 Nature나 Science 등 세계적으로 인지도가 높은, 즉 IF가 높은 학술지가 주로 포함되어 있기 때문으로 추정된다. IF_{avg} 가 가장 낮은 Mathematics와 가장 높은 Immunology의 차이는 3.7이며 비율로는 6.6배이다.

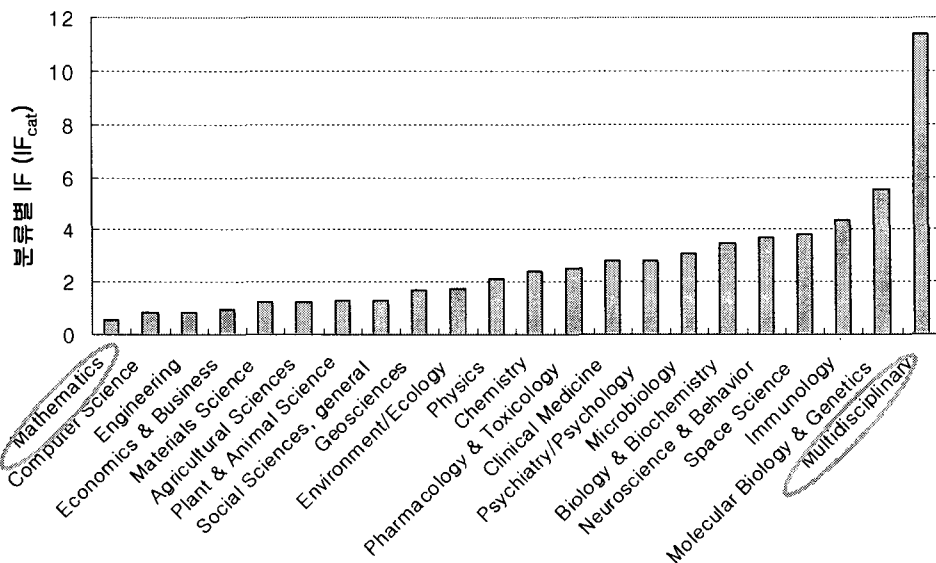
<그림 3> ESI 분류별 평균 IF(IF_{avg})



(3) ESI 분류별 평균 IF(IF_{cat})

<그림 4>는 앞서 설명한 ESI 분류별 IF를 산출하는 두 가지 방법 중 후자의 방법으로 IF_{cat}을 구하여 나타낸 그림이다. 분석 결과 IF_{cat}에 따른 분야별 양상은 IF_{avg} 분석 결과와 유사하다. 특히 Mathematics 분야의 경우는 IF 산출방법에 무관하게 가장 낮은 IF를 보였다. 본 연구에서는 Multidisciplinary가 다른 분류와 비교하여 유독 높은 IF_{cat}을 나타내고 있다는 사실도 주목할 만한 점이다.

<그림 4> ESI 분류별 IF(IF_{cat})



분야별 특수성에 대한 실증분석 결과 논문을 대상으로 하는 연구성과 평가에서 끊임없이 제기되던 IF의 학문분야별 차이를 정량적으로 제시할 수 있었다. 학문분야별 IF 등의 차이는 주로 분야별 인용패턴의 차이에 기인하는 것으로 알려져 있으나, 그 외에도 학술지의 수, 연간 연구자 일인당 논문 수 등 여러 가지 요인들이 복합적으로 작용하고 있다고 보는 것이 타당할 것이다. 이상의 분석 결과가 제시하는 것은 서로 다른 분야의 연구자들의 연구결과를 IF와 같이 획일화된 척도를 이용하여 평가할 경우 특정 분야에 속한 연구자들에 대하여 불합리적인 결과를 초래할 수 있다는 점이다.

3. 지표의 선정

위에서 살펴본 바와 같이 학술지 수(또는 논문 수)나 IF와 같은 피인용 기반의 지표는 분야별 특수성을 내재하고 있어 연구결과의 평가에 직접적으로 사용될 경우 문제를 야기할 소지가 있다. 본 연구에서는 분야별 특수성을 반영한 일련의 지표 및 활용 시 유의사항 등을 제시하여 객관적이며 공정하게 평가하는 데에 활용될 수 있도록 하였다.

1) 논문 수 (Publications; P)

논문 수는 논문성과 분석에 있어 가장 기초적인 자료로 사용된다. 적용 가능 대상은 광범위하여 연구자, 연구팀, 연구센터, 대학 등 각종 기관, 국제 규모의 연구센터, 국가 등이며, 특정 학문 또는 기술 분야에 대하여 세부 분야나 학제·기술간 융합 분야까지 적용이 가능하다. 그러나 논문 수 자체만으로는 큰 의미를 부여할 수 없다. 예를 들어 ‘특정 연도에 SCI 등재 학술지에 한국인 저자가 발표한 노화관련 논문의 수가 몇 편이다’라는 사실 그 자체만으로는 어떠한 의미도 부여하기 어렵다. 그러나 이를 동일한 분야에 대한 다른 집단의 논문 수와 비교하면, 두 집단 사이의 노화관련 연구의 활동도를 비교할 수 있는 자료가 된다. 따라서 논문 수를 평가 지표로 사용할 경우에는 제 2의 대상과 비교가 이루어지도록 하는 것이 바람직하다.

2) 피인용 수 (Citations; C)

논문의 인용정보는 해당 논문의 질, 영향(impact) 그리고 유용성(utility) 등을 판단할 수 있는 근거를 제공한다. 과학기술계는 다양한 연구주체 사이의 의견교환(communication)에 의하여 움직인다고 해도 과언이 아니다. 이러한 의견교환의 매개체 중에서 대단히 중요한 역할을 하고 있는 것이 바로 과학기술 논문이며, 연구주체들은 다른 연구주체의 연구결과를 자신의 논문에서 인용함으로써 그 가치를 부여하게 된다. 논문의 인용은 자신의 새로운 연구결과가 있을 수 있도록 도움을 준 이전의 논문을 명시하거나, 해당 과학기술 분야에서 중요한 발명이나 발견으로 간주할 수 있는 논문을 명시하기 위하여 이루어진다. 어떠한 목적이건 간에 특정 논문이 다른 논문에서 많이 인용되었다는 것은 그 논문이 과학기술적으로 질이 우수하거나 영향력을 가진다는 것을 의미하므로, 논문성과의 평가에 있어 피인용 수를 분석하는 것은 충분한 의미를 가진다. 이와 같은 맥락에서 논문의 피인용 수를 간접적 동료평가라고 하기도 한다.

3) 평균 피인용수 (Average Citation Rate; ACR)

논문의 피인용 수를 논문성과의 평가에 활용할 경우 논문 한 건의 피인용 수를 평가에 이용하는 일은 극히 드물며, 일반적으로 특정 연구주체가 발표한 논문군의 평균 피인용수를 계산하여 사용

하게 된다. 그러나 이 지표 역시 단독으로 사용할 경우에는 유용성이 감소한다. 다시 말해 특정 연구주체가 발표한 논문군의 평균 피인용 수가 몇 회라는 사실은 단순한 숫자 이상의 의미를 갖기 어렵다. 따라서 제 2의 비교대상을 도입하는 것이 필요하다. 본 연구에서는 이 때 도입되는 기준지표로 분야별 평균 피인용률(Field-Average Citation Rate: ACR(F))을 제안한다. 이는 피인용 수가 학문 분야별로 큰 차이를 보이기 때문이므로 분야별 특성을 고려해 주기 위해서이다.

4) 분야별 평균 피인용률 (Field-Average Citation Rate; ACR(F))

논문의 질을 평가하기 위하여 흔히 사용되는 논문 수, 피인용 수 및 영향지수는 분야별로 큰 차이를 보인다는 사실은 이미 분야별 특수성에 대한 실증분석에서 보인 바 있다. 그러나 분야간 비교·평가가 요구되는 경우에는 이와 같은 분야별 특수성을 고려할만한 지표가 존재하지 않아 전문가의 주관에 전적으로 의존해야 하는 상황이다. 본 연구에서는 이와 같은 분야별 속성을 반영할 수 있는 정량적 기준 지표를 제공하고자 한다.

분야별 평균 피인용률은 (식 1)에 의해서 산출할 수 있다. 여기서 ACR(F)는 분야 F의 평균 피인용률을 의미하며, C(F)와 P(F)는 각각 분야 F에 속하는 모든 논문의 피인용수 및 총 논문수를 의미한다. J_i 는 분야 F에 속한 N_j 개의 학술지 중 i 번째 학술지를, $C(J_i, F)$ 및 $P(J_i, F)$ 는 각각 학술지 J_i 에 게재된 논문의 총피인용수와 총논문수를 의미한다.

$$ACR(F) = \frac{C(F)}{P(F)} = \frac{\sum_{i=1}^{N_j} C(J_i, F)}{\sum_{i=1}^{N_j} P(J_i, F)} \quad (\text{식 1})$$

ACR(F)는 특정 분야에 속하는 학술지에 게재된 논문이 인용된 전체 횟수의 전체 논문 수에 대한 비율이므로 특정 분야에 속하는 논문이 평균적으로 인용되는 회수, 즉 해당분야의 평균 수준에 대한 정보를 제공한다. 따라서 평가자는 평가대상 논문의 피인용수 및 이 논문이 속한 분야의 ACR(F)를 참고하여 논문의 질을 평가할 수 있다. 단 이 지표를 사용하기 위해서는 적당한 분류체계를 선정하고 각 분야에 해당하는 학술지 목록을 작성하는 작업이 선행되어야 한다.

5) 분야 내 학술지 순위 (Journal Ranking; JR)

분야간 비교·평가가 요구되는 경우에 그 효용성에 문제가 있는 영향지수(IF)를 보완할 수 있는 지표로서 분야 내 학술지의 순위를 제공한다. 영향지수는 학술지에 수록된 논문의 피인용수에 근간을 두므로 분야별로 큰 차이를 보이고 그 결과 분야간 비교·평가에는 적합하지 않다. 이러한 경우에는 단순한 영향지수가 아닌 분야 내 학술지의 순위가 유효한 것으로 알려져 있다. 따라서 본 연구에서는 이를 논문의 질을 평가하기 위한 기준 지표로 제시한다. 분야별 평균 피인용률의 경우와 마찬가지로 적당한 분류체계가 선정되고 각 분야에 해당하는 학술지 목록이 선정되면 분야 내의 학술지 순위는 각 학술지의 영향지수로부터 부여할 수 있다.

6) 순위보정 영향지수 (Rank-Normalized Impact Factors; rnIF)

분야 내 학술지 순위를 사용하여 영향지수를 보정하면 분야간 비교·평가가 가능해진다(Pudovkin, 2004). 그동안 영향지수가 가지는 분야별 특수성을 보정해 주기 위하여 많은 노력이

있었다(Sen, 1982; Marshakova-Shaikevich, 1996; Sombatsompot et al, 2004; van Leeuwen et al, 2002; Hirst, 1978; Ramirez et al, 2000; Pudovkin, 2004). 그 중 추가적인 자료의 확보 없이 가장 간단하면서도 효과적으로 영향지수를 보정할 수 있는 방법은 (식 2)와 같이 보정된 분야 내의 학술지 순위를 사용하는 방법이다.

$$rnIF = \frac{(N_j - R_j + 1)}{N_j} \quad (\text{식 2})$$

여기서 rnIF는 순위보정 영향지수, N_j 는 해당 분야 내의 학술지 수, R_j 는 분야 내 특정 학술지의 순위이다. rnIF를 산출하기 위해서는 JCR에서 제공하는 정보 이외의 특별한 다른 정보를 필요로 하지 않는다. rnIF는 분야 내에서 특정 학술지가 차지하는 위상을 의미한다. 특정 학술지의 rnIF가 x 라는 값을 가진다면 이는 $(1-x) \times 100\%$ 의 학술지가 이 학술지보다 상위의 영향지수를 가진다는 사실을 의미한다. 예를 들어 2000년 기준으로 학술지 Genetics는 JCR 분류 중 Genetics & Heredity 분야의 114개 학술지 중에서 17위이다. 이 때 rnIF는 $(114-17+1)/117 = 0.860$ 이며 이는 해당 분야에서 $(1-0.860) \times 100\% = 14\%$ 의 학술지가 Genetics보다 상위에 있음을 의미한다. 특정 학술지가 분야 내 최상위인 경우 rnIF는 1이 되며, 최하위인 경우는 0에 가깝게 된다.

7) 분야 내 학술지 인용역치 (Citation Thresholds; CT)

분야 내 학술지의 순위에 근거하여 학술지 사이의 구간을 설정하고 이에 대한 정보를 활용한다. 순위를 정하는 기준은 논문 수, 피인용 수, 영향지수 등이 될 수 있다. 예를 들어 특정 분야에 속하는 학술지들을 영향지수 기준으로 10% 단위별로 10개 구간에 대하여 나누어 배분할 수 있다. 이는 임의의 학술지가 영향지수 기준으로 특정 분야 내에서 차지하는 위상에 대한 정보를 제공한다. ESI 데이터베이스에서 제공하는 바와 같이 분류별 피인용수에 대한 구간 정보 역시 분야 내 인용역치의 좋은 예이다.²⁾

8) 수준평가 지수 (Quality Factor; Q-Factor)

특정 논문집단의 수준을 평가하기 위하여 논문집단의 평균 피인용률과 해당 분야 논문의 평균 피인용률의 비율을 수준평가 지수(이하 Q-Factor) 형태로 제공한다. 이는 평가대상인 논문집단의 평균 피인용률인 ACR(Average Citation Rate)을 해당 분야의 국제 평균인 분야별 평균 피인용률인 ACR(F)와 비교한 값이다. 따라서 Q-Factor의 값이 1이면 평가대상 논문집단의 수준은 세계평균임을 의미한다. Q-Factor는 (식 3)과 같은 방법으로 산출한다. 여기서 C_i 는 N개의 평가대상 논문 중 i번째 논문의 피인용수를 의미한다.

$$Q-Factor = \frac{ACR}{ACR(F)} = \frac{\sum_{i=1}^N C_i / N}{\frac{\sum_{j=1}^{N_j} C(J_j, F)}{N_j} \cdot \frac{\sum_{j=1}^{N_j} P(J_j, F)}{N_j}} \quad (\text{식 3})$$

2) <http://portal.isiknowledge.com/portal.cgi?DestApp=ESI&Func=Frame>

Q-Factor 값이 유의성을 가지려면 평가대상의 크기가 큰 것이 바람직하다. 다시 말해 평가대상이 개별 연구자인 경우보다 연구집단, 연구기관, 연구개발사업, 부처 또는 국가 등과 같이 충분히 많은 개체(논문)수를 제공할 수 있는 경우가 통계학적으로 보다 많은 의미를 부여할 수 있다.

9) 활동도 지수 (Activity Index)

위에서 살펴본 지표는 모두 논문의 인용에 기반을 둔 지표들이다. 그러나 논문 게재 건수 역시 연구 성과를 평가하기 위한 지표로 활용될 수 있다. 단 논문 게재 건수 역시 분야별 특성을 보이므로 이를 고려해 주는 것이 바람직하다. 활동지수는 이러한 목적에 적합한 지표이다. 활동도 지수는 특정 연구주체가 전체 논문 게재 건수를 기준으로 특정 분야에서 차지하는 비중을 나타내는 것으로 논문 게재의 활동도 또는 집중도를 의미한다. 활동도 지수는 다음과 같이 산출할 수 있다 (Frame, 1977).

$$\text{Activity Index} = \frac{\frac{\text{특정연구주체가 특정 분야에 발표한 논문 수}}{\text{특정분야의 전체 논문 수}}}{\frac{\text{특정 연구주체가 발표한 모든 논문 수}}{\text{모든 분야의 전체 논문 수}}}$$

10) 매력도 지수 (Attractivity Index)

논문의 피인용수 역시 활동도 지수와 동일한 개념으로 취급할 수 있다. 이 경우 활동도 지수를 나타내는 식에 논문 게재 건수 대신 논문의 피인용수가 사용되며, 이 지수를 매력도 지수로 명명한다 (Schubert et al., 1986). 매력도 지수는 특정 연구주체가 전체 논문 피인용수를 기준으로 특정 분야에서 차지하는 비중을 나타내는 것으로 논문 피인용 활동도 또는 연구내용의 매력도를 의미한다. 매력도 지수는 다음과 같이 산출할 수 있다.

$$\text{Attractivity Index} = \frac{\frac{\text{특정연구주체가 특정 분야에 발표한 논문의 피인용수}}{\text{특정분야 전체 논문의 피인용수}}}{\frac{\text{특정 연구주체가 발표한 모든 논문의 피인용수}}{\text{모든 분야 전체 논문의 피인용수}}}$$

4. 지표의 적용

위에서 제시한 지표는 이론적으로는 개인으로부터 국가에 이르기까지 모든 연구개발 주체에 대하여 적용이 가능하다. 그러나 대부분의 통계 지표가 그러하듯이 본 연구에서 제시된 지표들 역시 적당한 대상에 대하여 적용할 경우에 그 유의성을 찾을 수 있다. 여기서는 연구개발의 주체, 즉 평가대상을 미시(micro), 중간(meso) 및 거시(macro)의 범위로 구분하고 각각에 대하여 어떤 지표를 적용하는 것이 유용한지에 대하여 언급하도록 하겠다.

미시적 평가는 평가대상이 개인 연구자이거나 국가연구개발사업의 가장 작은 단위인 연구개발과제인 경우를 의미한다. 이와 같은 평가 대상의 특징은 평가 대상이 되는 논문의 수가 그리 많지

않다는 점이다. 따라서 통계학적으로 지표에 영향을 줄 수 있는 인자가 다수 존재하게 된다. 한 예로 개인의 논문성과를 평가할 경우 피인용 회수는 시간에 대해 변하게 된다. 다시 말해 올해에 특정 논문의 피인용수가 1회였던 것이 내년에 2회로 증가하게 되면 이는 피인용 수가 2배 증가했음을 의미한다. 따라서 이를 포함하는 지표 역시 2배 증가한 값을 보이게 된다. 이는 평가 대상의 규모가 작기 때문에 발생하는 현상이다. 따라서 피인용수나 논문당 (평균) 피인용수와 같은 지표를 사용할 때는 분야별 평균 피인용률이나 분야 내 학술지 인용역치 등과 같은 비교자료와 함께 사용해야 한다.

인용정보를 평가에 활용하는 것이 여의치 않을 경우는 해당 논문에 게재된 학술지의 영향지수를 활용하되, 분야 간 비교가 필요할 경우에는 학술지가 속한 분야 내 학술지 순위 또는 순위보정 영향지수와 같은 지표를 기준으로 전문가에 의한 정성적 평가를 내리는 것이 바람직하다(Pudovkin, 2004).

중간적 평가는 평가대상이 연구개발사업, 연구기관, 부처(청) 등인 경우를 의미한다. 이 경우는 평가의 규모가 비교적 크므로 게재 논문 수나 피인용 회수 등의 지표가 미시적 평가의 평가대상에 비해 급격하게 변화하지는 않는다. 따라서 정량적 평가지표를 활용하여도 통계학적 의미를 찾을 수 있다. 따라서 본 연구에서 제시한 Q-factor와 같은 정량적 지표를 평가에 적용해도 큰 무리가 없다. 단 평가목적에 따라 연구팀이나 부서와 같은 비교적 작은 규모의 대상에 대한 평가를 병행할 필요도 있으며, 이때에는 미시적 평가에 사용된 지표 및 전문가 평가를 병행하는 것을 고려하는 것이 바람직하다.

거시적 평가의 경우는 평가 대상이 범부처 연구개발사업 등과 같이 국가차원의 규모를 가지는 연구개발활동에 대한 평가를 의미한다. 국가 간 비교·평가 역시 이에 해당한다. 거시적 평가의 경우 그 규모가 크기 때문에 평가대상 논문의 피인용수 등의 변화가 전체 지표에 미치는 영향이 크지 않다. 따라서 위에서 제시한 Q-factor와 같은 정량적 지표를 평가에 적용하는 것이 가능하다. 실례로 2003-2004년에 국가과학기술위원회에서 주관했던 “BT분야 국가연구개발에 대한 심층 분석 및 평가”의 경우 1994년 이후 국가차원에서 지원된 BT 분야 국가연구개발에 의한 논문성과물의 질적 수준을 평가하기 위해 Q-factor와 유사한 정량적 평가지표(CPP/FCSm)가 사용되기도 하였다(국가과학기술위원회 등, 2004).

본 연구에서 제시한 정량적 지표와 적용 가능한 평가대상의 범위를 <표 2>에 정리하였다. 평가자들은 평가 대상의 규모와 목적에 따라 아래의 지표 중 적당한 것을 선택적으로 적용할 것을 권장한다.

본 연구에서 제시한 지표 중 분야별 평균 피인용률 등 분야별 접근이 필요한 지표를 적용하기 위해서는 학문분류체계를 선정하는 과정이 선행되어야 한다. 이는 표준기술분류, 미래유망신기술 분류, NTRM 분류체계 등 현존하는 분류체계를 면밀히 검토하여 선정되어야 한다. 일단 분류체계가 선정된 이후에는 해당 분류에 속하는 학술지를 선정하는 작업이 이루어져야 한다. 이와 같은 분류체계 및 학술지의 선정작업은 과학기술계 학회 등 각 분야 전문가 집단의 합의를 통하여 이루어져야 지표의 활용에 있어 발생할 수 있는 공정성이나 객관성 문제를 방지할 수 있을 것으로 사료된다.

지표를 적용할 때 염두에 두어야 할 다른 사항은 시차(time-lag)에 대한 문제이다. 문헌의 인용 분석은 해당 문헌이 분석에 필요한 충분한 피인용수를 확보한 이후에 가능해진다. 네덜란드 CWTS의 연구에 의하면, 분야에 따라 다를 수는 있으나 일반적으로 한 문헌이 학술지 게재 이후 최대 피인용수를 확보하는 데에는 2-4년의 시간이 소요된다고 한다.³⁾ 따라서 인용 분석을 활용한 평가대상을 선정할 때 이 기간이 반드시 참고 되어야 한다. 예를 들어 2005년 말에 게재된 논문에 대한 인용 분석을 할 경우라면 적어도 2년(2006-2007년)의 시간이 경과한 이후에, 즉 2008년 초 이후에 분석을 하는 것이 분석 결과에 대한 유의성을 확보할 수 있다는 의미가 된다.

3) CWTS Graduate Course in Science and Technology Studies 2004, Measuring Science: Assessment of Research Performance and Discovery of Patterns of Scientific and Technological Development, Center for Science and Technology Studies, Leiden University, The Netherlands, 2004.

<표 1> 정량적 평가지표의 적용 대상

지표	대상	미시 (micro)	...	거시 (macro)
기본지표	논문 수 (Publications; P)	○		○
	피인용 수 (Citations; C)	○		○
	논문당 (평균) 피인용수 (Average Citation Rate; ACR) ⁴⁾	○		○
기준지표 ⁺	분야별 평균 피인용률 (Field-Average Citation Rate; $ACR(F)$)	○		○
	분야 내 학술지 순위 (Journal Ranking; JR)	○		
	분야 내 학술지 인용역치 (Citation Thresholds; CT)	○		
	순위보정 영향지수 (Rank-Normalized Impact Factors; $rnIF$)	○		
주지표	수준평가 지수 (Quality Factor; Q -Factor)			○
	활동도 지수 (Activity Index; AI)			○
	매력도 지수 (Attractivity Index; AAI)			○

III. 결론

논문성과물에 대한 평가에 있어 분야별 특성이 고려되어야 한다는 지적은 오래 전부터 존재하여 왔다. 그러나 국내에서 이에 대한 실증분석을 통한 문제제기는 이루어지지 않은 것이 현실이다. 본 연구에서는 Thomson Scientific사의 JCR 및 ESI 데이터베이스 분석을 통해 각종 분류별로 학술지 수, 피인용 수 및 영향지수(IF)등의 지표가 분야별 특수성을 보임을 정량적으로 증명하였다.

JCR 분류별로 산출한 IF_{cat} 별 학술지 수의 분포를 분석한 결과 정규분포를 보이지 않으며 특정 IF_{cat} 구간에 학술지가 집중되어 있음을 알 수 있었다. ESI 분류별 학술지 수를 분석한 결과 Clinical Medicine분야가 압도적으로 많은 학술지를 보유하고 있어 분야간 불균형의 가장 대표적인 예를 보여주는 사실을 확인할 수 있었다. 또한 ESI 분류체계별 IF를 두가지 방법으로 구한 결과 정도의 차이는 있으나 두 경우 모두 분야별 편중이 존재한다는 사실을 알 수 있었다. 특히 Mathematics 분야의 경우 계산방법에 무관하게 가장 낮은 IF를 보였으며, 이와 같은 분야는 분야별 특수성을 고려하지 않은 현재의 평가체계에서는 불이익을 당할 수 있음을 의미한다.

분야별 특수성에 대한 실증분석 결과 논문을 평가대상으로 하는 연구성과에 있어 분야별 특수성에 대한 고려는 반드시 필요하다라는 사실을 확인할 수 있었으며, 본 연구에서는 이와 같은 분야별 특수성을 고려하여 논문의 질을 평가하기 위한 일련의 지표를 제시하였다. 본 연구에서 제시하는 지표는 기본지표, 기준지표 및 주지표의 세 종류로 요약될 수 있다. 기본지표의 경우는 논문 수, 피인용 수 및 논문당 평균 피인용 수 등 가장 기본적이며 일차적인 지표들이다. 기준지표는 분야별 평균 피인용률, 분야 내 학술지 순위, 순위 보정 영향지수 등과 같이 기본지표 사용 시 비교 대상이 되거나, 분야간 비교가 필요할 때 활용 가능한 지표들이다. 끝으로 주지표는 수준평가지수, 활동도 지수 및 매력도 지수와 같이 기본지표와 기준지표로부터 얻을 수 있는 지표를 의미한다. 본 연구에서는 또한 이들 지표를 활용할 때 유의해야 할 사항을 제시하였다. 각 지표를 적용함에 있어 적용대상의 크기 및 시차 등을 고려하여야 지표의 유용성을 높일 수 있으며, 분야별 접근이 필요한 분야별 평균 피인용률과 같은 지표를 적용하기 위해서는 분류체계 및 학술지의 선정 등 과학기술계 전문가 집단의 합의가 반드시 선행되어야 할 것으로 사료된다.

4) CWTS의 경우는 CPP(Citations Per Publication)로 표현하기도 한다.

참고문헌

- 국가과학기술위원회, 한국과학기술기획평가원, 『BT 분야 국가연구개발 심층분석 및 평가에 관한 연구』, 2004.9.
- 이상필, 문영호, 이혁재, 여운동, “연구기획평가 지원을 위한 분야별 논문의 피인용 통계 분석”, 한국기술혁신학회, 2005.8.
- 국가과학기술위원회, 『2005년도 국가연구개발사업 조사·분석·평가 결과』, 한국과학기술기획평가원, 2005.8.
- 과학기술부, 『표준성과지표 해설집』, 2005.11.
- 과학기술부, 과학기술혁신본부, 『2006년도 국가연구개발사업 조사·분석·평가 추진계획』, 2005.12.
- 김인묵 외, 『기초과학연구지원사업별 평가전략 정립을 위한 정책조사연구』, 한국과학재단, 2002.
- 한국기술혁신학회, 『연구성과의 질 제고를 위한 논문평가 모형개발 (1)』, 2006.
- Frame, J.D., “Mainstream Research in Latin America and Caribbean”, *Interciencia*, 2, 1977, p.143.
- Garfield, E., “The ISI Impact Factor, Current Contents Print Edition”, 1994; <http://scientific.thomson.com/free/essays/journalcitationreports/impactfactor/>
- Hirst, G., “Discipline Impact Factor: A Method for Determining Core Journal Lists”, *Journal of the American Society for Information Science*, 29, 1978, pp.171-172.
- MacRoberts, M. H., MacRoberts, B.R., “Problems of Citation Analysis: A Critical Review”, *Journal of the American Society for Information Science*, 40(5), 1989, pp. 342-349.
- Marshakova-Shaikovich, I., “Bibliometric Evaluation of Russian Science Journals”, *Vestnik Rossiiskoi Akademi Nauk (The Herald of Russian Academy of Sciences)*, 73(9), 1996, pp.788-796.
- Moed, H. F., Burger, W.J.M., Frankfort, J.G., van Raan, A.F.J., “The Application of Bibliometric Indicators: Important Field- and Time-Department Factors to be Considered”, *Scientometrics*, 8, 1984, pp. 177-203.
- Pudovkin A., Garfield E., “Rank-Normalized Impact Factor: A Way to Compare Journal Performance Across Subject Categories”, *Proceedings of the 67th ASIS&T Annual Meeting*, 41, 2004, pp.504-515.
- Ramirez, A.M., Garcia, E.O., Del Rio, J.A., “Renormalized Impact Factors”, *Scientometrics*, 47, 2000, pp.3-9.
- A. Schubert, T. Braun, “Relative indicators and relational charts for comparative assessment of publication output and citation impact”, *Scientometrics*, 9, 1986.
- Sombatsompop, N., Premkamolnetr, T., Markpin, T., “A Modified Method for Calculating the Impact Factors in Journals in ISI Journal Citation Reports: Polymer Science Category in 1997-2001”, *Scientometrics*, 60(1), 2004, pp.217-235.
- Sen, B.K., “Normalized Impact Factor”, *Journal of Documentation*, 48, 1982, pp.318-329.
- van Leeuwen, T.N., Moed, H.F., “Development and Application of Journal Impact Factor Measures in the Dutch Science System”, *Scientometrics*, 53(2), 2002, pp.249-266.