

인용문헌 분석을 통한 학술 논문의 수명 및 계보에 관한 연구*

The Study on the Genealogy and Impact Factor of Papers by Citation Analysis

정 준 민(Jun-Min Chung)**

목 차

- | | |
|---------------------|---------------------|
| 1. 서론 | 4. 인용문헌 분석 - 이론적 해석 |
| 2. 인용문헌 분석 - 가설적 접근 | 5. 결론 |
| 3. 인용문헌 분석 - 계량적 모형 | |

초 록

인용문헌 분석을 과거 집단적 분석에서 탈피하여 개개 문헌이 갖는 행위를 시간축에 따른 영향력으로 해석하여 단순히 인용 회수가 아닌 인용 체인내에서 문헌의 영향력으로 해석하여 보았다. 영향력 지수는 직접 인용이 되거나 비록 직접 인용은 되지 않았더라도 특정 문헌을 인용한 문헌이 다시 다른 문헌에 의해 인용되어질 경우, 특정 문헌의 영향을 받았다는 체인 구조 내에서의 연결 효과를 설명하고 있다. 실험의 결과는 인용되지 않은 문헌의 grace period를 도입하여 문헌의 수명을 인용 관계 속에서 풀었고 이렇게 살아남아 영향력 지수가 높은 문헌을 하나의 데이터베이스로 서비스 할 경우, 학술지 시장의 활성화와 도서관 제정의 압박을 풀 수 있을 것이라 예측을 내놓았다.

ABSTRACT

The citation analysis is not applied to group measurements but to individual papers finding their impact factors among citation chains which are created by tracing citing-cited relationship between any two papers in data set. The individual impact factor is measured by adding each value derived from citation chain. Each paper's impact factor index is calculated by adding the values of each index by direct citing-cited relationship, and the values of each impact factor made by indirectly influencing to the papers in the citation chain. The research introduces a grace period, in which the system holds the papers not cited by other papers yet, but are expected to be cited within this period. Eventually not cited papers after the grace period would be eliminated by the system. The experiment suggests a reasonable database in which the highly influenced papers are gathered, and could be serviced in stead of buying databases filled with worthless not-cited-papers.

키워드: 인용분석, 계보, 생산성, 계량서지학, 영향력 지수

Citation Analysis, Genealogy, Productivity, Bibliometrics, Impact Factor

* 이 논문은 2005년도 전남대학교 학술연구비 지원에 의하여 연구되었음.

** 전남대학교 문헌정보학과 교수(wizard@chonnam.ac.kr)

논문접수일자: 2010년 4월 26일 최초심사일자: 2010년 4월 29일 게재확정일자: 2010년 5월 11일
한국문헌정보학회지, 44(2): 357-379, 2010. [DOI:10.4275/KSLIS.2010.44.2.357]

1. 서론

과학기술 논문의 폭발적 증가는 그만큼 연구자들의 연구 성과가 많음을 보여주기도 하지만 동시에 연구자들이 논문을 쓰기 전, 검색을 통하여 점검하고 확인해야 할 시간과 비용이 증가한다는 말이기도 하다. 작금의 과학기술은 바로 이러한 논문들을 통해 시간과 공간의 제약을 뛰어넘는 엄청난 발전을 거듭해 왔다. 후자는 앞으로 과연 연구할 만한 주제와 대상이 남아있을까 의심이 들 정도로 많은 양의 연구 성과물이 쏟아져 나왔고 또 앞으로도 그 이상 생산 되리라는 데엔 의심의 여지가 없다. 그렇다면 그렇게 많이 생산된 논문의 가치는 어떨까? 과연 그것이 생산해 낸 양 만큼이나 그 가치가 있을까? 만일 우리가 그 가치를 알 수 있다면 그 척도는 무엇일까?

계량서지학에서 과학기술 문헌의 가치와 그 평가 방법을 알아내기 위하여 무수히 많은 실험과 가설을 동원하였지만 현재까지 가장 객관적이며 또 합리적인 방법으로 연구논문 스스로 만들어낸 참고문헌(인용문헌)에 대한 수량학적 분석을 들 수 있다. 인용문헌이란 동료 연구자에 의해 자신이 쓴 논문의 가치를 공식적으로 인정받기 위해 인용되어지는 문헌으로 그 목적과 방법은 다양하지만 한 가지 분명한 것은 그것이 인위적이기 보다는 차라리 자연적이라는 것이다.

인용문헌은 주로 자신이 논문에서 나타내고자 하는 주제와 방법에 대한 역사적, 이론적 근거와 함께 선임 연구자에 대한 학술적 업적에 대한 예우로 사용되어져 왔다(김갑선 2002; Bonzi & Snyder 1991). 같은 주제의 논문을 쓰면서

도 학계에서 인정받지 못하거나 근거가 빈약한 논문은 거의 인용하지 않았을 뿐 아니라 자신의 논문이 인용되지 않은 같은 부류의 논문에 대해서는 표절 시비까지 일어나는 게 현실이고 보면 인용문헌의 가치는 단순히 목록을 넘어 그 분야 전문가 집단에 의한 자정(自淨) 작업이라고 보아도 과언이 아니다. 그러나 형식 그렇듯이 모든 것이 수량적으로 이해되거나 설명되어질 수는 없다. 그 수량적 표현 이면에는 그것에 힘을 부여할 만한 이론과 기준이 있어야 하는 것이다. 인용문헌 역시 수량적 표현을 위한 보다 이론적이고 기술적인 배경이 있어야 한다. 단순히 다른 논문에 많이 인용되었다고 좋은 논문은 아니라고 본다. 즉, 인용빈도가 함축하고 있는 다른 요인도 같이 검토된다면 보다 명확하고 객관적인 척도가 될 수 있으리라 본다.

인용문헌에 대한 연구는 크게 관계, 유용성, 그리고 가치에 대한 연구가 있다. 또한 이들 연구의 공통점은 인용행위에 대해 통계적 분석을 시도하고 있음이다. 관계를 규명하기 위한 수단인 동시인용(co-citation)와 동일문헌 인용(bibliographic coupling)의 경우, 동시 인용 빈도와 같은 문헌의 공통적 인용 행위를 계량적으로 측정하는 통계적 행위이다. 그 관련성이 깊으면 깊을수록 두 문헌 나아가 두 문헌의 저자 사이의 주제적 관련성을 표시하고 이렇게 구축된 지도를 통해 주제적 성향을 파악할 수 있게 되는 것이다. 이 분야에 대한 연구는 현재에도 꽤 많이 진행되고 있으며 나름 그 성과를 보이고 있다(김희정 2005; 이재운 2008). 이렇듯 저자에 대한 동시인용 또는 동일문헌 인용 같은 경우 클러스터링 기법과 그래프 이론에 입각하여 나름 실험적 법칙을 찾고 있다.

인용문헌을 통한 수명에 대한 연구는 상당히 광범위하게 이루어졌으나 그 방법의 단순함과 결과에 미치는 요인이 너무 많아 그렇게 성공적이지는 않다. 과거 특정 연도에 생산된 문헌의 해를 거듭하는 과정에서의 인용빈도수를 측정하거나, 최신의 특정 연도를 대상으로 그 해에 생산된 문헌이 갖는 인용문헌의 생산된 연도를 비교함으로써 그 주제 영역의 수명이나 유용성을 측정한다. 그러나 한번 생산된 문헌의 이용 가치가 줄 것이라는 것은 누구나 예측가능하지만 실제로 문헌을 분석해 보면 단지 집단적 의미 외에는 개별 문헌에서는 연도와 상관없이 지속적으로 인용된 문헌도 있음을 볼 때, 단순히 통계적으로 수명 또는 유용성을 평가하는 것은 집단적 해석 외에는 큰 의미를 찾을 수 없다. 그러나 미디어 환경변화에 따른 문헌 수명의 변화를 발견하는 것은 나름 의미 있는 분석이라 본다(Shin 2009).

학술저널에 대한 가치 평가로 가장 많이 사용되어지는 지표로 JIF(Journal Impact Factor)가 있다. 1975년부터 ISI(Institute for Scientific Information)에서 Journal Citation Reports를 통해 저널 상호간의 인용패턴을 이용해 저널영향력을 발표하고 있는데(Garfield 2006) 그 평가 방법은 여러 학자에 의해 조금씩 개량되고 보강되고 있지만(Habibzadeh, F. & Yadollahie, M. 2008; Neuhaus, C., Marx, W., & Daniel, H. 2009) 평가지표를 개발하는데 인용문헌이 갖는 의미는 같다고 볼 수 있다. 특정연도 t에 있어서 특정저널의 JIF는 같은 해 t에 특정저널의 t-1년과 t-2년에 실린 논문들의 인용된 수를 t-1년과 t-2년에 특정저널에 실린 논문의 수로 나눈 값을 말한다. 이는 평가연도의 앞 2년에

걸쳐 생산된 논문이 얼마만큼 인용이 되는가를 표시하는 것으로 저널의 입수경로와 매체에 따른 편차가 있음을 상기해 볼 때 앞으로 상당한 변화를 예견할 수 있다. 개별 논문 또는 저자에 대한 인용문헌 평가지표도 이미 소개된 예가 있다. 그렇지만 엄밀히 말하면 개별 논문이나 저자에 대한 평이기 보다 인용관계를 통해 학문 추이를 분석하거나 변화를 예견하고자 하는 시도이며(Garfield, E., Pudovkin, A.I. & Istomin, V.S. 2003) 인용관계와 인용횟수에 의한 인용의 랭크추이를 분석함으로써 개별논문의 가치를 비교하는 논문도 있으나(Sidiropoulos, A., & Manolopoulos, Y. 2005) 이들 논문들은 인용이라는 사실에만 초점을 맞추으로써 인용의 가치를 단정 짓고 있다. 즉, 인용된 사실보다 인용의 결과와 추이가 중요함에도 불구하고 인용된 역사적 사실만을 전제함으로써 인용효과에 대한 결과적 해석을 간과한 것이다.

본 연구는 인용문헌의 특성을 보다 객관적이고 다양한 척도로 분석하여 실제로 과학기술 문헌의 가치 평가의 기준으로 삼아보고자 한다. 집단적 지표의 의미가 아닌, 그리고 인용문헌을 단순히 유용성에 대한 해석이 아닌 그 자체로 이미 하나의 객체이며 시대적으로 상속되어지는 가치를 갖는 생명체로 인식하고자 한다. 즉, 생산된 후 그 가치와 속성을 다음 세대에 심어주기 위한 행위로 인용(상속)을 선택하였고 그 과정에 그 가치가 생존함을 보이는 것이다. 인용되지 않는 문헌은 죽은 문헌이라 볼 수 있지만 최소 한번 이상 인용이 되어 다른 문헌에 영향을 끼치고 그 영향을 받은 문헌이 지속적으로 인용되거나 앞의 과정(상속)을 밟는다면 비록 더 이상 인용이 되지 않더라도 그 생명

력은 존재한다고 볼 수 있는 것이다. 같은 의미에서 비록 여럿에게 인용이 되었지만 그 문헌을 인용한 다른 문헌들이 더 이상 인용되지 않는다면 첫 문헌조차 그 가치가 생존할 수 없게 되므로 죽은 문헌과 마찬가지가 된다.

2. 인용문헌 분석 - 가설적 접근(실험세팅)

본 연구는 크게 문헌의 계보(genealogy)에 관한 연구와 문헌의 수명(lifetime)에 관한 연구로 대별되며 기존의 연구에서 보여준 인용문헌의 집체적 분석이 아닌 개개 인용문헌에 대한 객체적 해석을 전제하여 각 논문들의 가치를 평가할 수 있는 방안을 설계하여 보았다.

먼저 논문이 생산되고 생산된 논문이 인용되는 시점에 대한 분석을 통하여 과거 인용문헌에 대한 수명을 측정하듯이 단순히 특정연도의 논문이 다른 연도에 비해 상대적으로 얼마만큼 인용되는가에 따라 그 주제 문헌의 수명을 계산하였다면 본 연구에서는 인용이 됨을 전제로 얼마만큼 빨리 인용되는가를 측정하여 그 주제의 잠복기(grace period)를 계산하여 보았다. 특정 주제의 논문이 생산되어지고 그것이 인용되는 것을 전제할 때 평균 어느 정도 시기를 기다리면 특정 논문이 인용되거나 그렇지 않음을 판단하기 위함이다. 물론 모든 논문이 그러하진 않지만 대개 일정 기간이 경과할 때까지 인용되지 않으면 그 논문이 인용될 가능성은 거의 0에 가깝다 할 수 있다[가설 1]. 결국 인용되지 않은 논문은 그 논문의 생산과 무관하게 의미론적으로 존재하지 아니함을 알 수 있다.

일단 인용된 논문이 정리되어지고 나면 그 논문이 얼마나 많이 그리고 얼마나 오래 인용되었나를 따져 그 논문의 그 주제에 끼친 영향력을 파악하였다. 이 과정에서 어느 논문은 1) 특정 시기에 집중적으로 인용되거나, 2) 거의 한 두 편의 논문만이 인용하였거나 아니면 3) 매해 지속적으로 인용되고 있음을 알 수 있었다. 그런데 한번 인용된 문헌의 경우, 이 문헌을 인용한 논문이 일정기간이 경과할 때 까지 다른 문헌에 의해 인용되지 않으면 본인이 인용한 문헌 역시 더 이상 아무에게도 영향력을 행사할 수 없게 되므로 숫자상으로는 비록 자신이 한번 인용되었다고 하더라도 실제로는 전혀 인용되지 않음과 같게 되는 것이다. 이 같은 추론을 전제로 가설을 설정하여 보면 개개 문헌의 피인용 평가는 그 문헌을 인용한 문헌이 지속적으로 살아 있을 때 의미를 가지며 만일 그렇지 아니할 경우 피인용 문헌 역시 인용되지 않는 결과를 낳는다.

위의 두 가지 사실을 전제로 보면, 모든 문헌은 단지 인용됨이 중요한 것이 아니라 지속적으로 그 영향력을 직간접적으로 끼칠 때 의미를 갖는다. 뿐만 아니라 그 영향력은 자신이 직접 인용되었을 경우 뿐 아니라 자신을 인용한 문헌이 누군가에 의해 인용되어 그 생명력을 가질 때 본인이 인용되었음 역시 의미가 있으므로 자신을 중심으로 하나의 인용 체인을 형성해 가면서 그 영향력을 갖게 되는 것이다. 그렇다면 여기서 다음과 같은 의문을 갖게 된다.

첫째, 특정 문헌이 직접적으로 인용된 경우가 더 큰 영향력을 갖는가? 아니면 둘째, 비록 간접적으로 인용된 효과를 갖지만 특정 문헌을 인용한 문헌이 연속적

으로 인용 고리를 갖는 것이 보다 더 영향력이 큰가?

첫 번째 경우, 인용한 문헌이 더 이상 인용되지 않을 경우, 피인용 문헌까지도 사라질 위기를 초래할 수 있으며 두 번째의 경우는 적은 피인용 횟수를 가진 오래된 문헌이 그 긴 인용 체인 덕분에 크게 평가되어질 소지가 있게 된다. 이 두 가지 문제는 인용과 피인용 관계에서 피인용 문헌이 갖는 영향력(dif: direct impact factor)을 어떻게 부여하느냐와 인용 체인관계 속에서 간접 영향력(iif: indirect impact factor)을 어떻게 측정 또는 정의하느냐에 따라 차이를 보일 뿐 아니라 적절한 기준을 만들 수 있으리라 본다[가설 2].

Goffman은 정보검색에 있어 특정 질의에 대해 특정문헌을 찾는 행위에서 한걸음 더 나아가 그 특정문헌의 인용문헌, 그리고 그것의 인용문헌도 특정 질의에 적합한 문헌임을 증명함으로써 검색의 효율을 높였으며 그 때 인용문헌 체인에 놓인 문헌의 적합성 측정을 위해 인용/피인용 문헌간의 적합성을 p로 보고 특정 질의에 대한 문헌의 적합성을 체인에 있는 문헌의 거리에 따라 p, p², p³, p⁴, ... 라 놓았으며 그 값이 임계점을 넘으면 관련문헌으로 정의하였다(Goffman 1968). 마찬가지로 인용 체인관계의 제일 끝, 즉 최종 인용문헌이 인용한 문헌의 영향력을 dif라 했을 때, 그 문헌에 의해 인용된 문헌의 영향력은 바로 직전의 문헌에 의한 dif와 그 문헌이 갖는 영향력 dif의 간접인용 효과에 따라 dif*dif(=iif)가 추가되어 dif + dif*dif가 된다. 이런 식으로 N개의 문헌이 연결된 인용체인을 이어가다 보면 가장 근원이 된 문헌의

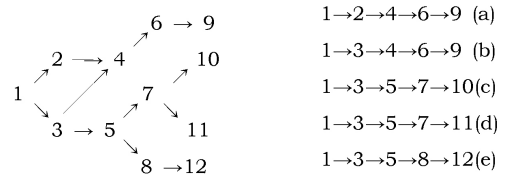
영향력,

$$IF_1 = \frac{dif(1 - dif^{n-1})}{1 - dif}$$

이를 일반화 시키면,

$$IF_k = \frac{dif(1 - dif^{n-k})}{1 - dif} \quad \text{공식 A}$$

인용 체인의 n번째 문헌은 최종 인용문헌이므로 자신이 인용된 적이 없는 문헌이다 그러므로 그 문헌의 영향력은 0이 되는 것이며 그 앞의 문헌 즉 n-1번째 문헌의 영향력은 dif가 됨을 알 수 있다. 그러나 인용 체인이 단일 체인이 아니므로 각 문헌의 영향력은 단일 체인으로 조각내어 그 값을 계산한 후 병렬 체인의 값을 모두 합산하여 주면 된다.



1번 문헌의 영향력 지수는 채널(a), (b), (c), (d) 그리고 (e)의 값을 합산하면 되는데 각 채널의 계산은 공식 A를 따른다. 문헌 간 영향력 계수 dif를 .2로 놓고 계산하여 보면,

$$\frac{.2 * (1 - .2^4)}{1 - .2} * 5 = .2496 \text{ 이 된다.}$$

왜냐하면 1번 문헌의 인용체인 크기는 모두 5이기 때문이다.

여기서 우리는 Goffman이 말했듯이 적합성(영향력)에 대한 논의가 제기된다. 비록 인용 체인의 첫 문헌의 영향력이 체인에 미치더라도 그 미치는 정도가 어느 정도인가를 측정할 필요가 있다. 그 값은 dif에 따라 그것의 몇 제곱까지 의미 있는 값인지를 산출해 내야하며 그 과정은 실증적 분석이 전제됨을 말한다. 본 연구는 그와 같은 가능성을 전제로 실험적 분석을 시도하여 봄으로써 그 가능성과 효용성을 확인하여 보고자 한다.

3. 인용문헌 분석 - 계량적 모형(실험분석)

3.1 데이터 수집

본 연구는 크게 문헌의 계보(genealogy)에 관한 연구와 문헌의 영향지수(impact factor)에 관한 연구로 대별되며 SCOPUS에 등재된 저널에 실린 과학기술 논문과 그 인용문헌을 대상 데이터로 한다. 연구범위는 SCOPUS에서 온라인으로 제공하는 데이터로 하며 검색의 효율과 검색된 데이터의 재현율을 높이기 위해 분석 관련 주제의 가장 포괄적인 키워드를 추출하여 이를 조합 검색을 통하여 추출하였다.

SCOPUS는 2004년 11월 Elsevier에 의해 개발된 현존하는 가장 큰 초록 및 색인을 수록한 상업 데이터베이스로 출발하였다. ISI의 Web of Science를 능가할 인용문헌 관리시스템을 갖고 있으며 WoS에서조차 그 시스템을 따라갈 정도의 성장을 하고 있다(Gorraiz, J. & Schloegl, C. 2008).

본 연구는 SCOPUS를 통해 아래의 주제어를 갖는 문헌을 추출하였으며 추출된 문헌으로부터 인용문헌을 발췌하여 그 인용문헌이 검색된 문헌군에 속한 경우만을 추려 실험 데이터를 구축하였다. 이렇게 실험집단을 구축하지 않으면 인용문헌의 체인을 만들 수 없을 뿐 아니라 인용문헌의 적합성(영향력)을 객관적으로 판정하기가 쉽지 않기 때문이다. 즉, 같은 주제어를 가진 집단 내에서 상호 인용 관계를 파악하고 그것을 체인화 함으로써 실험 집단의 특성을 명확히 하고 체인 내에서의 상호 적합성(영향력)을 확인할 수 있기 때문이다.

실험데이터수집	탐색 주제어	검색 문헌수
2009년 4월 6일	bibliometrics	2,086건

본 실험은 실험을 통하여 주제집단에 대한 사실 데이터를 추출하거나 분석하고자 함이 아니라 인용 체인과 문헌의 영향력 평가 모형을 만들기 위한 행위이므로 실험 데이터가 갖는 특질은 보다 현상을 명확하게 표현하고 드러낼 필요가 있다.

3.2 데이터 기초 통계

실험 데이터로 추출된 문헌 2,086건 중 이들 문헌을 인용한 횟수는 전체 8,954건이었으며 문헌 집합 중, 초기 실험 데이터로 선정된 데이터에 포함되는 데이터에 의해 인용된 횟수는 2,524건, 데이터에 포함되지 않은 외부 데이터에 의해 인용된 횟수는 6,430건이었다. 실험 데이터 내의 문헌을 인용한 문헌은 실험 데이터 외에 8,546건이 더 있었으나 그 주제는 bibliometrics

이기 보다는 인접 영역으로 8,546건에 비해 그들이 인용한 횟수가 6,430건인 것으로 보아 평균 한 번이 조금 넘는 인용횟수로 크게 의미를 부여할 필요는 없다. 반면 실험 데이터 내부를 보면 인용된 문헌 721건이 내부적으로 2,524회 인용되어 평균 세 번이 넘는 인용횟수를 보인다.

다. 그러나 만일 실제로 문헌 하나 하나의 영향력을 따지는 경우라면 모든 데이터를 열린 상태에서 취급하여야 할 것이다.

〈표 1〉은 실험 데이터의 출판년도에 따른 피인용 분석 내용을 보여준다.

〈표 1〉에서 보듯이 753개의 문헌은 실험 데이

〈표 1〉 출판연도에 따른 피인용 통계

문헌의 출판년	A	B	C	Total
1975			1	1
1985	1	1		2
1986		3	1	4
1987	1	2	1	4
1988		2		2
1989	2	3	2	7
1990	3	3	3	9
1991	4	6	4	14
1992	4	1	7	12
1993	2	2	8	12
1994	4	2	4	10
1995	10	10	18	38
1996	6	14	19	39
1997	7	10	30	47
1998	7	12	23	42
1999	16	27	46	89
2000	18	46	51	115
2001	25	53	66	144
2002	24	50	66	140
2003	32	63	69	164
2004	25	54	72	151
2005	43	73	79	195
2006	90	128	79	297
2007	119	117	57	293
2008	169	71	15	255
2009				
Total	612	753	721	2,086

A: 인용되지 않음

B: 실험 데이터 외부에서만 인용됨

C: 실험 데이터 및 외부 데이터로부터 인용됨

터에서 보면 612개의 실제로 전혀 인용되지 않은 문헌과 마찬가지로 전혀 인용되지 않음으로 평가되어지고 있다. 그러나 이 데이터가 인용된다고 평가되기 위해서는 이들 문헌이 인용한 피 인용 문헌이 데이터 내에 편입되어야하는데 결국 모든 데이터를 완전히 분석하기 전까지는 불가능한 일이다. 아마 WoS(Web of Science)의 JIF처럼 SCOPUS가 이들 데이터를 분석해 그 결과를 보여줄 수 있다면 실현 가능하리라 본다.

〈표 2〉는 실험 데이터 2,086건 중 인용된 721건의 첫 피인용 시기를 나타낸 것이다.

〈표 3〉은 실험 데이터가 실제로(인용문헌을 실험 데이터로 국한하지 않을 경우) 얼마나 많이 인용되는가를 연대별로 나열한 것이다. 특히 신규로 정의된 값은 잠복기(약3년: grace period)

가 경과되지 않은 문헌으로 아직 인용되지 않은 문헌의 수를 가리킨다. 이것을 오직 실험 데이터 내로 한정시켜 표현하면 〈표 4〉와 같다.

〈표 4〉에 제시된 내용은 실험 데이터 2,086건 중 내부적으로 인용된 721건의 피 인용 빈도와 연도별 관계를 보여주고 있다. 총 2,524건의 피 인용 회수와 평균 3.5회의 피 인용 회수를 보여주고 있다. 그러나 만일 이것을 단순히 실험 데이터 전체에 적용하면 평균 1.2회의 피 인용 회수를 나타내어 인용 분석의 왜곡이 심화될 여지가 있게 된다.

전혀 인용되지 않은(612건) 또는 외부로부터만 인용된(753건) 문헌의 연도별 분포는 〈표 5〉와 같다. 인용되지 않은 612건 안에는 잠복기에 있는 문헌 378건이 포함되어 있다.

〈표 2〉 출판 후 첫 피인용 때까지 걸린 시간(year)

인용된 해	문헌 수	%	인용된 해	문헌 수	%
출판 원년	90	12.50%	출판 후 16년	0	0.00%
출판 후 01년	190	26.39%	출판 후 17년	0	0.00%
출판 후 02년	150	20.83%	출판 후 18년	2	0.28%
출판 후 03년	98	13.61%	출판 후 19년	0	0.00%
출판 후 04년	55	7.64%	출판 후 20년	0	0.00%
출판 후 05년	46	6.39%	출판 후 21년	0	0.00%
출판 후 06년	29	4.03%	출판 후 22년	0	0.00%
출판 후 07년	23	3.19%	출판 후 23년	0	0.00%
출판 후 08년	16	2.22%	출판 후 24년	0	0.00%
출판 후 09년	9	1.25%	출판 후 25년	0	0.00%
출판 후 10년	2	0.28%	출판 후 26년	0	0.00%
출판 후 11년	5	0.69%	출판 후 27년	0	0.00%
출판 후 12년	4	0.56%	출판 후 28년	0	0.00%
출판 후 13년	1	0.14%	출판 후 29년	0	0.00%
출판 후 14년	0	0.00%	출판 후 30년	0	0.00%
출판 후 15년	0	0.00%	출판 후 31년	1(무시)	0.00%

* 총 721개 문헌이 실험 데이터 내부적으로 인용되어졌으며 1개 논문이 특이하게 출판 후 31년 만에 인용되어져 데이터의 오류를 줄이기 위해 무시하였음. 결과적으로 평균 2.71년이 경과 후 처음으로 인용이 되어졌으며 만일 31년 후 인용된 문헌을 포함할 경우 2.75년의 평균값을 얻을 수 있었다.

〈표 3〉 연도별 출판물의 피인용 횟수

Total	피인용	1975	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996
1	500 이상													
1	400-499													
0	300-399													
0	200-299													
4	100-199													
1	90-99													
4	80-89									1				
2	70-79													
3	60-69													
7	50-59				1					1	1			
8	40-49									1				
23	30-39							1						1
55	20-29								1	1				3
146	10-19			1				1	2	1	1	3	6	5
190	5-9						2	1	1	2	4	1	7	5
61	4												4	
58	3													2
85	2	1									1			
72	1										1		1	3
	외부 인용		1	3	2	2	3	3	6	1	2	2	10	14
	인용 안됨		1		1		2	3	4	4	2	4	10	6
	신규													
2,086		1	2	4	4	2	7	9	14	12	12	10	38	39

Total	피인용	1997	1998	1999	2000	2001	2002	2003	2004	2005	grace period			
											2006	2007	2008	2009
1	500 이상					1								
1	400-499	1												
0	300-399													
0	200-299													
4	100-199		1				1		1	1				
1	90-99					1								
3	80-89					1	1	1						
2	70-79			1		1								
3	60-69	1		1					1					
4	50-59			1	1					1	1			
7	40-49	1			2	1	1	1			1			
21	30-39	2	2	2	2	2		3		3	3	2		
50	20-29	3	3	4	4	7	4	8	10	2	3	2		
126	10-19	6	5	11	7	13	17	25	16	16	8	1	1	

Total	피인용	1997	1998	1999	2000	2001	2002	2003	2004	2005	grace period			
											2006	2007	2008	2009
167	5-9	6	3	12	10	16	16	17	19	23	26	15	4	
57	4		1	3	9	5	8	5	9	5	3	8	1	
56	3	2		2	6	3	9	5	4	7	10	6	2	
83	2	2	4	2	4	8	5	3	8	13	15	15	4	
67	1	6	4	7	6	7	4	1	4	8	9	8	3	
753	외부 인용	10	12	27	46	53	50	63	54	73	128	117	71	
234	인용 안됨	7	7	16	18	25	24	32	25	43	Not Defined Yet			
378	신규										90	119	169	
1,932		47	42	89	115	144	140	164	151	195	297	293	255	

〈표 4〉 실험데이터 출판연별 피인용 횟수

total	1	0	0	0	0	0	1	3	11	28	95	43	95	138	306	721
	100회 이상 인용	90-99 회 인용	80-89 회 인용	70-79 회 인용	60-69 회 인용	50-59 회 인용	40-49 회 인용	30-39 회 인용	20-29 회 인용	10-19 회 인용	5-9 회 인용	4회 인용	3회 인용	2회 인용	1회 인용	
1975														1		1
1985																0
1986																1
1987												1				1
1988																0
1989												1	1			2
1990										1			1		1	3
1991											1	1			2	4
1992							1			2	1	1		1	1	7
1993									1				2	1	4	8
1994														4		4
1995											3	4	5	2	4	18
1996										3	5		4	3	4	19
1997	1								1	1	3	5	3	4	12	30
1998								1	1	2	4	2	2	4	7	23
1999								1	1	2	11	3	8	7	13	46
2000									1	2	7	3	7	12	19	51
2001										5	7	5	8	10	31	66
2002								1	1	2	11	3	4	18	26	66
2003									1	6	13	1	12	11	25	69
2004									1	1	16	4	16	9	25	72
2005										3	9	3	7	17	40	79
grace period	2006								1		4	3	8	20	43	79
	2007											3	6	12	36	57
	2008												1	2	12	15
	2009															

〈표 5〉 미인용 문헌 연도별 분포

Total	피인용	1975	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996
753	외부 인용		1	3	2	2	3	3	6	1	2	2	10	14
234	인용 안됨		1		1		2	3	4	4	2	4	10	6
378	신규													

	피인용	1997	1998	1999	2000	2001	2002	2003	2004	2005	grace period			
											2006	2007	2008	2009
	외부 인용	10	12	27	46	53	50	63	54	73	128	117	71	
	인용 안됨	7	7	16	18	25	24	32	25	43	Not Defined Yet			
	신규										90	119	169	

3.3 모형 설계

결국 인용된(내부적으로) 721건의 논문과 잠복기에 있는 논문 378건 만이 살아있는 논문으로 평가되어졌으며 987건(47%)의 논문은 그 수명을 다했다고 할 수 있다. 그러나 첫 인용 때까지 걸린 시간과 평균시간을 비교해 보면 평균시간을 넘어 인용된 192건의 논문의 예처럼 실험 데이터 내에서 약 9%의 확률로 회생의 가능성은 있다고 본다. 이 또한 확률적 의미일 뿐 실제 인용된다 하더라도 2차 시간 제약은 약 3년으로 70%가 이 안에 속한다고 보면 잠복기 2.7년을 6년으로 잡을 경우 그래도 590여 편(6년 이전 논문 134편 + 잠복기 내의 문헌 중 약 94% = 전체의 약 28%)의 논문은 영구히 제명이 되어도 무방하다.

721편 논문의 인용체인을 만들기 위해 모든 데이터를 피인용 - 인용관계로 묶어 보았다. 이렇게 묶인 데이터에는 인용된 문헌과 인용한 문헌으로 대별되어지며 인용한 문헌 내에는 다른 문헌에 의해 인용된 문헌도 포함되어져 있다. 또 일부 인용한 문헌 집단 내에는 향후 분석

에서 더 이상 인용되지 않는 문헌으로 분류되어 사라질 문헌이 포함되어 있으며 결과적으로 그 문헌으로부터 인용된 문헌 관계 역시 삭제시킬 예정이다. 전체 데이터(실험 데이터 및 실험 데이터를 인용하고 있는 실험 데이터 외부 데이터)를 대상으로 하면 13,563개의 피인용 - 인용관계가 나오고 그것을 실험 데이터 내로 좁히면 2,524개의 관계를 정리하였다.

이들 관계를 체인으로 엮기 위해 인용문헌이 피 인용되는 관계 쌍을 찾아 그 관계가 끝날 때까지 연속적으로 이어나가 인용 체인을 구축하였다. 구축된 관계 쌍은 〈표 6〉과 같은 모습을 보인다.

〈표 6〉의 예시는 하나의 피인용 - 인용 관계를 하나의 체인 고리라 할 때 이런 고리가 11회 반복되고 있음을 보여주며 본 실험 데이터에서 가장 긴 체인으로 이와 같은 연결 고리가 4개 나타났다. 전체 연결 고리에 대한 통계는 〈표 7〉과 같다.






〈표 7〉에서 보듯이 고리 개수의 기대치는 4.8개로 평균적으로 모든 체인은 약 5개의 고리를 갖는다. 5개의 고리를 갖는 체인을 중심으로 체인의 형태와 특성을 알아보았다.

<표 6> 인용 체인의 예시

피인용-인용			피인용-인용			피인용-인용			피인용-인용		
67	153	○	153	258	○	258	582	○	582	668	○
668	949	○	949	1,091	○	1,091	1,489	○	1,489	1,540	○
1,540	1,690	○	1,690	1,952	○	1,952	2,037	×			

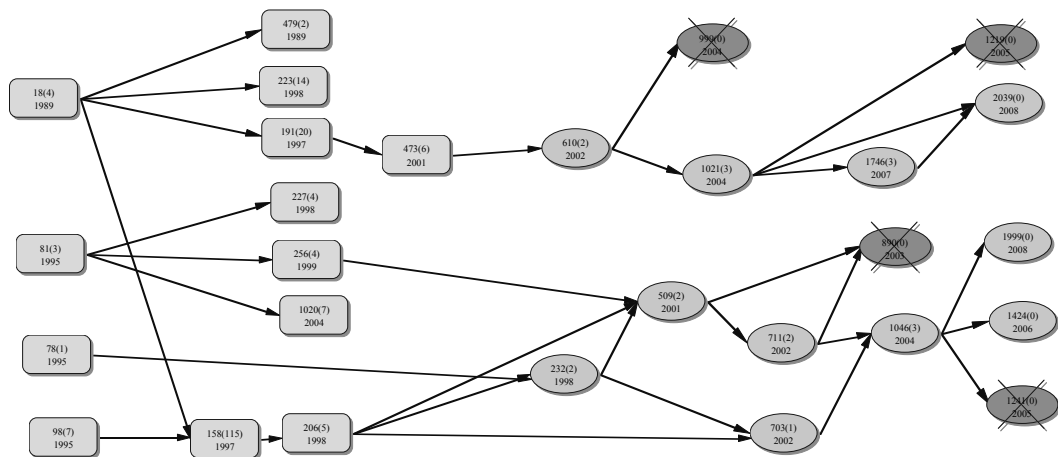
<표 7> 고리 개수와 그에 따른 체인 수

고리 개수	1	2	3	4	5	6	7	8	9	10	11
체인 수	476	896	1,647	2,127	2,206	1,882	1,360	709	249	52	4

<그림 1>에서 ,  (또는 둥근 모형에 X 표시된 노드) 및  (또는 둥근 모형 노드)는 문헌을 표시하며 화살표는 화살표의 왼쪽 문헌이 오른쪽 문헌에 의해 인용되었음을 나타내며 노드 안에 표시된 수는 문헌 고유번호 인용횟수 및 출판연도를 나타낸다. 는 이미 죽은 논문이며 는 사라질 위험성이 높은 문헌 집단이다.

<그림 1>에서도 나타나듯이 18(4), 78(1), 81(3), 98(7)번 문헌은 그 인용횟수가 얼마 되

지 않지만 그 문헌을 인용한 문헌 중에 158(115), 191(20)처럼 피인용 빈도가 높은 문헌이 있는가 하면 초기 문헌을 인용한 문헌 집단이 적절히 피인용 되어 더 이상 인용되지 않더라도 사라질 가능성은 희박하다. 한편 체인의 오른쪽 끝에 존재하는 최근 문헌의 경우, 잠복기(grace period)가 전제되지 않거나 앞으로 1, 2년이 경과할 때까지 더 이상 인용되지 않는다면 지금의 죽은 문헌보다 더 많은 문헌이 연쇄적으로 사라질 수 있게 되는 것이다. 1046(3)의 경우,



<그림 1> 5개의 고리를 가진 체인 샘플

자신을 인용한 문헌 중 하나는 사라져버렸고 나머지 둘도 비슷한 상황에 놓여 있어 1046(3) 문헌이 사라지면 연쇄적으로 711(2), 703(1), 509(2), 232(2) 문헌도 사라질 뿐 아니라 206(5) 문헌도 상당히 불안해질 수밖에 없다. 결국 시간이 흐르면서 인용되었다는 사실만으로 평가 되는 것이 아니라 살아남는 논문이 제대로 평가받을 수 있게 되는 것이다. 2002년 논문 중에는 같은 저널 같은 호에 상호 인용한 예가 발견 되는데 [678(1) ↔ 684(1)] 이 경우는 과학기술의 발전으로 논문이 배포되기 전, 회색논문의 형태에서 서로 참조가 일어나기 때문이라 생각되어지며 경우에 따라서는 자기 논문을 자신이 인용하는 경우로 이 경우 역시 문헌의 품질이 우수하지 않을 경우, 시간이 경과하면 사

라지게 되어있다.

13,563개의 인용체인의 상호 영향력을 측정하기 위하여 피인용 - 인용 관계의 영향력 $p= .2$ 에 서부터 .6까지 .1 단위로 실험을 실시하였다. 실험 은 grace period 3년을 전제하여 3년이 경과할 때까지 인용되지 않은 문헌을 포함한 경우(pif) 와 배제한 경우(Qpif), 이렇게 두 가지 상황을 놓고 하였다. 상위 20위 이내에 드는 문헌은 대 개 1990년대에 출판된 것이며 2000년대 문헌이 3개 나타남을 보인다(표 8 참조). 다만 죽은 문헌으로 인한 순위 변동은 크게 일어나지 않음을 보이고 있다. 다만 하위권 문헌에선 죽은 문헌으로 인해 전혀 영향력을 갖지 못하는 문헌이 다수(59개 문헌) 발생함을 볼 때, 시간의 경과 에 따라 상당한 수의 문헌이 데이터베이스 내에

〈표 8〉 피인용 영향력 평가 상위 20위

		pif				
ID		0.20	0.30	0.40	0.50	0.60
98	1995	382.77652050	652.55465601	1002.45972398	1465.39257813	2089.16989870
47	1992	369.35354804	631.19476579	974.61575238	1437.65283203	2079.01908031
48	1992	356.31143444	607.88473180	935.94230866	1374.19824219	1973.14023199
158	1997	285.11682253	480.80172766	727.38029466	1041.60546875	1446.03313766
189	1997	159.75296307	272.68836890	419.63656806	614.66796875	877.99981978
279	1999	159.04503808	268.41902028	406.34605568	582.20312500	808.59906048
100	1995	144.99520696	247.66741715	381.91947796	561.81738281	808.44210094
326	1999	141.95990784	241.40239119	369.07582464	535.23046875	753.84232704
513	2001	126.48975360	213.29260440	322.24174080	460.03125000	635.30296320
67	1993	119.39620651	202.75362525	310.06001312	451.24365234	641.28088375
138	1996	113.09442048	192.54049109	295.12873984	429.99023438	610.15280640
134	1996	108.42820280	185.51192864	286.93762580	424.19628906	614.93642158
21	1990	107.11674061	181.70565574	276.97990554	400.49609375	562.87196774
151	1996	105.85467392	179.93300198	275.26837248	400.10351563	566.17012224
49	1992	104.64909005	177.99725998	272.66882458	397.18359375	563.72466278
18	1989	103.03775867	175.75647033	270.18765496	395.22851563	563.72348191
418	2000	78.24798720	132.61911030	201.76335360	290.60156250	405.59063040
257	1999	67.14497536	114.12789066	174.49634816	253.25781250	357.34288896
350	2000	66.47071283	112.45449855	170.97462886	246.56054688	345.47601562
258	1999	65.82650470	112.23313375	172.50839757	252.46484375	360.56394547

〈표 8〉 피인용 영향력 평가 상위 20위(계속)

ID		Qpif				
		0.20	0.30	0.40	0.50	0.60
98	1995	365,33535570	622,51884531	955,88003758	1396,88476563	1991,49953710
47	1992	368,67751604	629,82991879	972,11258438	1433,29345703	2071,70599231
48	1992	355,20967444	605,85214180	932,53878866	1368,79199219	1964,86215199
158	1997	268,52411853	452,49458866	684,30611866	979,99609375	1361,28376166
189	1997	159,56736307	272,13576890	418,39496806	612,29296875	873,91021978
279	1999	156,29303808	263,60102028	398,89005568	571,45312500	793,81506048
100	1995	144,94784696	247,47922715	381,40235796	560,66113281	806,17842094
326	1999	141,52790784	240,34939119	367,05982464	531,85546875	748,65832704
513	2001	125,20975360	211,01260440	318,72174080	455,03125000	628,58296320
67	1993	118,33828651	200,94909525	307,33617312	447,39990234	636,08152375
138	1996	110,96130048	188,81326109	289,31369984	421,45898438	598,11824640
134	1996	108,36964280	185,33903864	286,52930580	423,35253906	613,34954158
21	1990	105,22554061	178,37895574	271,76070554	392,80859375	552,00476774
151	1996	103,43547392	175,85930198	269,10517248	391,29101563	554,00692224
49	1992	104,38029005	177,50195998	271,84002458	395,87109375	561,72786278
18	1989	98,20479867	167,45313033	257,37933496	376,54101563	537,35900191
418	2000	78,24798720	132,61911030	201,76335360	290,60156250	405,59063040
257	1999	67,01697536	113,78589066	173,79234816	252,00781250	355,32688896
350	2000	63,14879283	106,82046855	162,47478886	234,52929688	329,11665562
258	1999	65,73050470	111,99913375	172,06039757	251,71484375	359,41194547

피인용 - 인용 관계의 영향력은 어느 값을 주든 큰 변화는 없었다. 이론적으로는 p값이 크면 클 수록 긴 고리를 가진 문헌이 유리하리라 보였지만 폭 넓게 인용되어지는 데에도 같이 작용하므로 그 변화는 미미함을 보였다. 그러나 피인용 횟수가 높은 20개 문헌을 조사해 보면 영향력 20위 순위에 오직 7개 문헌만 포함되어 있으며 특히, 최신 문헌의 경우는 영향력이 높지 않음을 알 수 있다. 〈표 9〉는 피인용 횟수 상위 20위를 보여주고 있다.

〈표 9〉 피인용 횟수 상위 20위

영향력순위	id	year	count
4	158	1997	115
3	48	1992	49
25	219	1998	34
32	296	1999	34
58	642	2002	30
2	47	1992	29
74	971	2004	26
15	49	1992	25
10	67	1993	25
43	612	2002	25
119	1,319	2006	23
6	279	1999	22
19	350	2000	22
86	744	2003	22
37	205	1998	21
45	191	1997	20
73	1,061	2005	19
24	177	1997	17
31	541	2001	17
23	153	1996	16

4. 인용문헌 분석 - 이론적 해석(결과해석)

[가설 1]

특정 주제의 논문이 생산되어지고 그것이 인용되는 것을 전제할 때 평균 어느 정도 시기를 기다리면 특정 논문이 인용되는가? 또 어느 정도 기간이 경과할 때까지 인용되지 않으면 그 논문이 인용될 가능성은 거의 0에 가깝다 할 수 있겠는가?

[가설 2]

인용한 문헌이 더 이상 인용되지 않을 경우, 피인용 문헌까지도 사라질 위기를 초래할 수 있으며 적은 피인용 횟수를 가진 오래된 문헌이 그 긴 인용 체인 덕분에 크게 평가되어 질 소지가 있게 된다. 이 두 가지 문제는 인용과 피인용 관계에서 피인용 문헌이 갖는 영향력(dif: direct impact factor)을 어떻게 부여하느냐와 인용 체인관계 속에서 간접 영향력(iif: indirect impact factor)을 어떻게 측정 또는 정의하느냐에 따라 차이를 보일 뿐 아니라 적절한 기준을 만들 수 있으리라 본다.

4.1 실험 데이터의 적정성 문제

문헌에 대한 행태연구, 특히 인용문헌에 대한 연구를 수행하기 위해선 실험 데이터를 구축하는 것이 가장 힘든 작업이다. 데이터를 수집하고 분석하기 위해선 수집된 데이터가 유효성을 가져야 하는데 원 문헌 및 그것을 인용한 인용문헌을 동시에 구축하기 위해선 데이터베

이스가 갱신되는 시점을 잘 파악하여 데이터의 변동이 없을 때 데이터를 구축해야한다. 그러나 실제로 SCOPUS에 접속하여 데이터를 수집하고자 하면 원 문헌을 구축하고 그 원 문헌의 인용문헌을 하나 하나 재검색을 시도해야 하는데 그 작업이 상당한 시간을 요한다.

본 연구에서는 'bibliometrics'에 관한 문헌 2,086개를 일차적으로 구축한 다음 문헌을 생산 연도의 내림차순으로 정렬하였으며 수집한 데이터 항목에 수집 당시의 인용문헌 개수를 표시케 하여 추후 인용문헌을 수집할 때 수치의 오류를 최소화 하려 하였다. SCOPUS가 무작위로 데이터를 갱신할 뿐 아니라 오래된 문헌을 수시로 편입하여 구축된 문헌 이외에 관련 문헌이 발생할 뿐 아니라 인용문헌 검색을 위해 새로이 접속할 경우, 이미 구축된 실험 데이터와 무관하게 배열된 데이터에서 인용문헌을 찾아 원 문헌과 일치시키는데 상당한 어려움을 겪었다. 본 연구는 일의 수월성을 위하여 Excel을 활용하였으며 인용 고리를 표현하기 위해 Inspiration이라는 mind map 관련 소프트웨어를 활용하였다.

일차적으로 구축된 원 문헌과 이차적으로 수집된 인용문헌을 상호 연결시켜 인용문헌 중 원 문헌에 포함되지 않은 문헌을 골라내어 따로 데이터 처리를 하였으며 원 문헌과 인용문헌을 연결 짓는 과정에서 피인용과 인용의 관계를 1:1로 구축하였다. 즉, 인용된 문헌 721개와 그를 인용한 문헌 중 2,086개의 원 문헌 데이터에 속한 문헌을 일차적으로 묶어내었다. 이 일차적 배열에서 인용문헌 내에 피 인용된 문헌과 단순히 인용만으로 그 역할을 다한 문헌 이렇게 두 종류의 문헌을 구분 지었으며 이와 같은 작업을 연속적으로 수행하여 인용 고

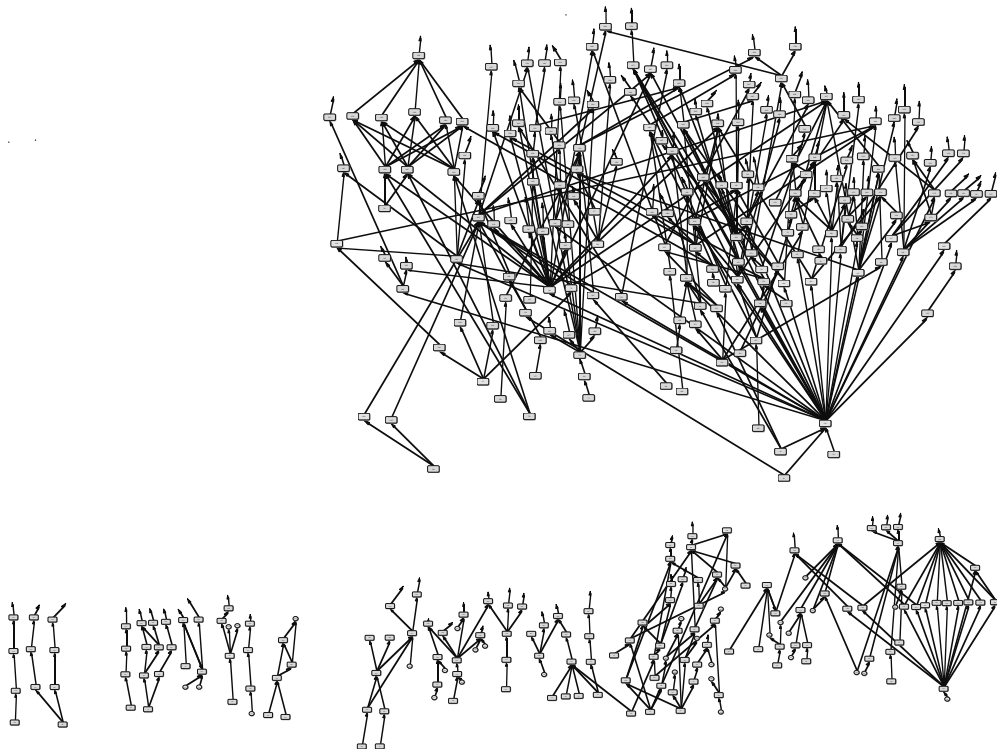
리를 완성하였다. <그림 2>는 완성된 인용 고리를 Inspiration으로 표출한 모습이다.

거의 대부분의 문헌이 상호 얽혀 있으며 <그림 2>의 아래 부분은 일부 문헌을 발췌하여 다양한 구성을 보이려 하였다. 문제는 그림에 표현되지 않은 실험 데이터 외부의 데이터로 실험 데이터가 이들 외부 데이터로부터 인용이 되었음에도 불구하고 경우에 따라서는 전혀 인용되지 않은 문헌으로 평가된다는 사실이다. 앞장에서 언급하였듯이 특정한 주제를 주지 않고 전체 학술정보 데이터베이스를 분석한다면 가능하겠지만 피 인용 - 인용 관계를 해석하기 위해선 피인용 문헌이나 인용 문헌 모두 닫힌 공간

에 존재해야 하기 때문에 주제적 특성을 갖는 데이터의 경우는 어쩔 수 없이 피 인용 문헌군을 도메인으로 설정할 수밖에 없는 것이다. 결국 특정 시간을 전제한 피 인용 문헌군과 그에 상응하는 인용 문헌 집단이 모두 같은 공간에서 그 관계를 정의함으로써 실험 데이터는 적정함을 갖추었다.

4.2 잠복기(grace period) 설정의 문제

피 인용과 인용의 관계를 설정함에 있어 개개 문헌에 있어서는 인용되어지는 행위와 결과가 기록으로 의미를 갖지만 집단의 시각에서 보



<그림 2> Inspiration으로 표현한 인용 고리

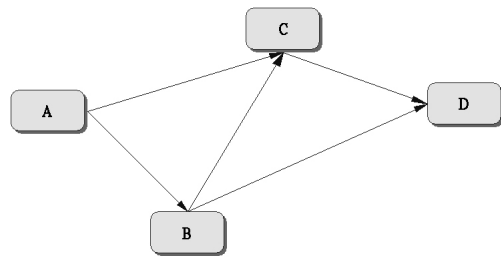
면 시간의 흐름에 따라 새로운 행위와 결과가 생겨나고 지나간 기록이 퇴색되는 시간 축에서 움직인다. 인용 역시 시간의 흐름 속에 생성과 소멸하는 행위이지 각인되어 기억되는 행위는 아니다. 그럼에도 불구하고 한번 인용된 것은 인용이 된 이력이 아닌 인용한 문헌이 갖는 내용적 표출에서 의미가 있다. 즉, 인용은 행위로서가 아닌 교류를 통하여 영향을 끼치는 결과의 의미를 갖는 것이다. 결국 인용은 피 인용 문헌의 기록이 아닌 인용 문헌의 영향력으로 생존해 가는 것이다. 그렇기 때문에 인용한 문헌이 향후 인용되지 않는다면 그 영향력 또한 소멸하게 되는 것이다. 여기서 제기되는 문제는 영향력의 소멸과 승계는 ‘어떻게 표출되는가?’이다. 승계는 인용 문헌이 다른 문헌에 의해 인용되어질 때 발생할 것이며 소멸은 인용 문헌이 아무에게도 인용되지 않을 때라는 것이다. 그러나 인용이 되는 시점은 행위로써 누구나 인지할 수 있지만 인용되지 않음은 알 수가 없다.

그러나 문헌이 인용되지 않음은 예측을 할 수는 있다. 특정 주제 집단 내에서 과거 문헌들이 보인 인용 행태 분석을 통하여 하나의 문헌이 인용이 된다면 그 시기는 모든 문헌이 출판되어 다른 문헌에 의해 인용될 때까지 걸리는 시간의 기대치를 구하면 된다. 평균시간(기대치)를 넘을 때까지 인용되지 않는 문헌은 확률적으로 인용되지 않을 가능성이 높다는 말이다. 그러므로 인용되지 않은 문헌이라 하더라도 출판 후 기간이 기댓값에 못 미칠 경우는 앞으로 인용되어질 가능성이 상당히 높음을 암시하고 있다. 물론 확률과 분포에 대한 가설적 해석이지만 이 역시 수학적모델이라고 본다면 인용되어질 시기에 대한 기댓값은 문헌에 있어서는

생존의 시기이며 그 이후로는 소멸의 해석도 의미 있게 되는 것이다. 여기서 생성과 소멸이라 함은 특정 데이터베이스의 편입과 삭제의 맥락에서 이해되어야 하며 향후 정보 서비스의 품질과 비용에 관한 전제가 되는 것이다.

4.3 인용 고리 설계의 문제

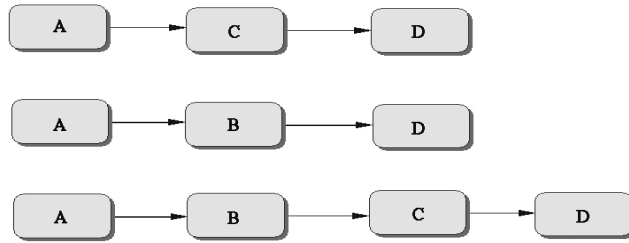
문헌이 인용되어지는 구조적(synthetic) 과정을 영향력 차원에서 보면 직접 인용되어지는 경우와 동시에 타 문헌에 의해 간접적으로 영향력을 끼치는 경우도 발생한다.



〈그림 3〉 문헌의 피 인용 - 인용 관계도

A 문헌은 B와 C 문헌에 의해 인용되어졌지만 그 영향력은 B, C 는 물론 D 문헌에 까지도 끼치고 있으며 C 문헌의 경우는 A 문헌을 직접 인용함으로써 그 영향력을 크게 받고 있지만 동시에 B 문헌을 통해 간접적으로 A 문헌의 영향력을 추가로 받게 되는 것이다. 결국 A 문헌의 경우, 그 영향력 경로는 〈그림 4〉와 같다.

즉, 문헌 A는 3개의 인용 체인을 갖게 되는 것이며 각 체인으로부터 영향력 값을 동시에 부여받게 되는 것이다. 두 문헌간의 영향력 지수를 p 라 하면 문헌 A가 갖는 영향력 값은 3개의 직렬 값을 병렬로 더하면 된다. 즉, 각 인용



〈그림 4〉 영향력 경로 예시

체인의 영향력 값(공식 A 참조)을 병렬 체인 수만큼 각각의 값을 더하면 되는 것이다.

$$p_A = (p + p^2) + (p + p^2) + (p + p^2 + p^3)$$

직렬 체인 직렬 체인 직렬 체인

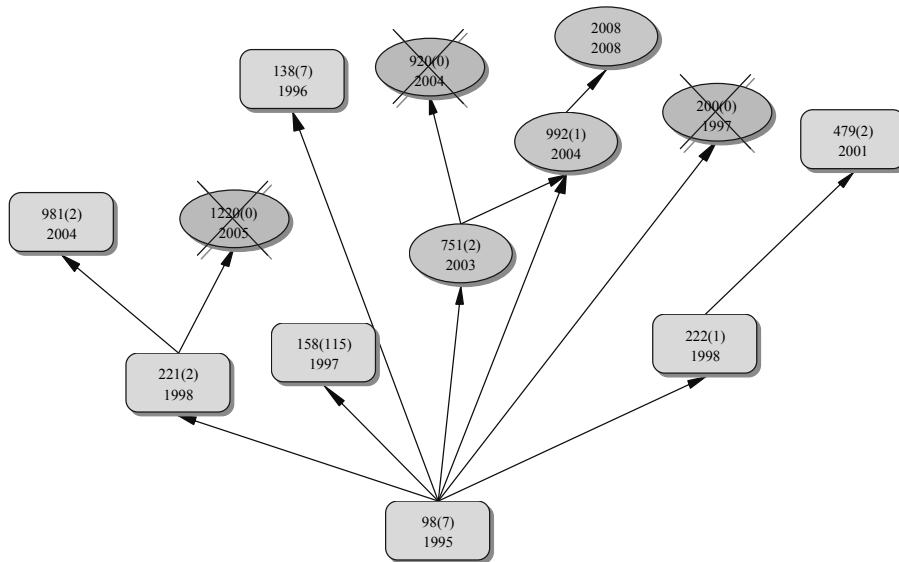
〈그림 5〉 영향력 지수

4.4 영향력 지수의 개발과 의미

인용행위가 시간에 대한 변수라면 결과는 이 미 다른 문헌에 의해 반영되어지고 시간의 흐름 속에 그 의미를 갖게 된다. 그러므로 그 영향력을 측정하거나 보여줄 수 있다면 단지 기록으로만 존재하는 인용 행위를 시간에 대한 결과로 해석할 수 있으며 그 추이를 분석하면 단기 예측 속에 가치 있는 문헌을 추려 제공할 수 있게 되는 것이다. 즉 영향력 지표를 지금이 아닌 미래 시점에서의 문헌의 가치를 나타내는 지수로 설정하면 상대적으로 높은 지수를 갖는 문헌을 모아 데이터베이스를 만들어낼 수 있으며 도서관이나 연구기관이 매년 전자저널의 기간호를 구독하는데 투자하는 비용을 줄여 기초 자원 데이터베이스(steady DB)로 대체할 수 있게 되는 것이다. 영향력 지수 값은 원리만 지켜진다

면 더 개량시킬 수 있다. 피 인용문헌이 갖는 후발 문헌에 대한 직, 간접의 영향력을 수렴하여 측정된 개개 문헌이 갖는 실시간 지수를 영향력 지수라 하며 이 지수는 기록으로서의 의미를 갖는 인용의 행위를 시간에 따른 결과로 해석하기 위한 방편으로 문헌은 한번 생산되어지지만 그 영향력은 인용을 통하여 관련 문헌 집단에 깊게 자리매김하게 되는 것이다.

〈표 8〉과 〈표 9〉에서 보여주듯이 영향력 지수는 피 인용 회수보다 훨씬 합리적이다. 〈표 8〉에서 1위를 차지한 문헌 98번은 오직 7번 인용이 되었음에도 불구하고 1,535개의 인용 체인을 만들어내고 있다(그림 6 참조). 인용 체인이 많다는 것은 98번 문헌을 직, 간접으로 인용하고 있는 문헌들이 상호간에 인용행위가 많이 발생하고 있음을 의미한다. 이는 결코 죽은 문헌에 의해 인용 체인이 무너질 가능성이 적다는 것을 의미하고 결국은 시간 축 속에서 문헌의 미래가치를 표현한다 하겠다. 그러므로 이 문헌을 위시하여 높은 영향력 지수를 갖는 문헌은 인용 회수보다는 인용 체인이 많음을 의미하고 그만큼 안정적으로 활용이 된다는 것을 뜻한다. 그러나 인용 회수가 여전히 적음에 대해서는 논란의 여지가 많다. 98번 문헌을 인용한 문헌은 7개로 그 중에 두 개(138,158) 문헌이 영향력



〈그림 6〉 98번 문헌의 인용 체인

20위 안에 들어 있으며 한 문헌은 죽은 문헌이며 다른 한 문헌 역시 잠복기가 지나면 연쇄적으로 죽을 가능성이 높은 문헌이다.

결국 인용 회수가 많고 적음의 문제가 아니라 탄탄한 문헌에 의해 인용되어지고 아니고에 따라 그 문헌의 영향력 뿐 만아니라 삶과 죽음의 문제도 거론되는 것이다.

5. 결론

계량서지학 영역에서 인용문헌 분석이 차지하는 비중은 상당히 크다. 그러나 관심과 규모에 비해 그 내용은 너무 한정적이다. 인용문헌의 행태가 갖는 집단적 가치에 관심을 갖다보니 개개 인용행위에 대한 해석이 그렇게 많지는 않다. 물론 인용문헌이 갖는 용도와 해석의 폭이 넓다 보니 정형화된 틀을 만들기 어려움

도 있겠지만 시간 축에 대한 행위이기 때문에 갖는 한계가 있다. 그럼에도 불구하고 개별 인용문에 대한 연구는 내용에 대한 계량적 해석에서 크게 벗어나지 않고 있다. 본 연구에서 보여주는 행태분석은 단순히 인용의 내재적 문제가 아닌 패턴에 대한 수량적 해석이며 시간의 흐름에 따른 인용의 결과적 분석이다.

한번 인용된 기록이 기록으로 의미를 갖는 것이 아니라 그것의 가치에 따라 실제로 인용행위가 평가받는 것으로 피인용된 문헌조차도 그것을 인용한 문헌의 가치가 없어지면 인용행위조차 의미를 상실하며 결국 피인용 사실도 의미가 없게 되는 것이다. 이렇듯 인용 행위를 유기적으로 해석함으로써 인용에 대한 새로운 해석을 시도하였으며 그것을 집단이 아닌 개개 문헌의 영향력으로 평가하여 봄으로써 인용문헌 분석의 새로운 시도를 하였다. 결과는 많이 인용된 문헌에 비해 상대적으로 적은 피인용

횃수를 갖는 문헌이 그 인용 체인이 길면 그 가치가 높아짐을 보임으로써 피인용 횃수와 함께 인용 체인에 대한 의미를 찾았다.

본 연구의 결과를 가지고 실제 학술정보 시장에서의 활용방안을 제시하여 보면 다음과 같다.

전자도서관 서비스가 일반화 되면서 종이로 된 문서가 점차 전자문서로 바뀌고 많은 도서관이 전자저널을 구독하는 상황에서 저널의 기간 호들이 도서관에 남아있지 않을 뿐 아니라 구독계약에서도 제외되고 있다. 그렇다면 도서관은 기간 호에 대해 어떻게 대처해야할까? 이용자들이 언제 얼마나 이용할지도 모르는 지난 호에 대해 지속적으로 구독을 하거나 일방적으로 구독을 중지할 수도 없고, 이 모두를 이용자들이 알아서 해결하라 하기에는 엄청난 부대비용이 들어가는 일이다.

앞의 실험 결과에서도 보았듯이 오래된 문헌

의 상당부분은 인용되어지지 않는(81%) 게 현실이라고 보면 인용되어질 가능성이 있는 또는 많이 인용되어지는 논문만을 모아 하나의 패키지로 만든다면 도서관은 도서관 나름대로 출판업자는 출판업자 나름대로 저렴하고 이윤창출이 가능한 상품을 개발할 수 있다고 본다. 즉, 일정기간(grace period)의 저널만을 정기구독하고 그 기간이 경과한 저널의 논문은 하나의 데이터베이스로 만들어 출판업자는 좀 더 비싼 가격에 그러나 전체를 구독할 때 보다는 싸게 공급한다면 도서관은 도서관대로 개인연구자는 연구자대로 많은 시간과 비용을 절감할 수 있으리라 본다. 물론 인용되지 않은 논문에 대한 데이터베이스는 검색은 자유롭게 하지만 그 논문에 대한 복사 서비스는 상당한 비용을 지불하여야 이런 논의가 의미를 갖는다.

참 고 문 헌

- [1] 강대신, 문성빈. 2009. 연구성과의 질적 평가를 위한 계량정보학적 분석에 관한 연구. 『정보관리학회지』, 26(3): 377-394.
- [2] 김갑선. 2002. 참고문헌의 동시공존현상-한국사회과학자들의 인용동기와 참고문헌의 분석. 『한국문헌정보학회지』, 36(4): 21-47.
- [3] 김희정. 2005. 저자 동시인용분석에 의한 국내의 기록관리학 분야의 지적구조비교에 관한 연구. 『한국문헌정보학회지』, 39(3): 207-224.
- [4] 유소영, 이재운. 2008. 학제적 분야의 정보서비스를 위한 학술지 인용 분석에 관한 연구: Y대학교 생명공학과를 중심으로. 『정보관리학회지』, 25(4): 283-308.
- [5] 이재운. 2008. 서지적 저자결합분석 - 연구동향 분석을 위한 새로운 접근 -. 『정보관리학회지』, 25(1): 173-190.
- [6] Bonzi, Susan & Snyder, H. W. 1991. "Motivations for Citation: A Comparison of Self Citation

- and Citation to Others.” *Scientometrics*, 21(2): 245-254.
- [7] Eun-Ja Shin, 2009. “The Dispersion Phenomenon of Journal Citations in a Digital Environment.” *Journal of the Korean Society for Information Management*, 26(2): 211-222.
- [8] Garfield, E. 2006. “The History and Meaning of the Journal Impact Factor.” *Journal of the American Medical Association*, 295(1): 90-93.
- [9] Garfield, E., Pudovkin, A. I. & Istomin, V. S. 2003. “Why do we need Algorithmic Historiography?” *Journal of the American Society for Information Science and Technology*, 54(5): 400-412.
- [10] Goffman, W. 1968. “An Indirect Method of Information Retrieval.” *Information Storage & Retrieval*, 4(4): 361-373.
- [11] Gorraiz, J., & Schloegl, C. 2008. “A bibliometric analysis of pharmacology and pharmacy journals: Scopus versus Web of Science.” *Journal of Information Science*, 34(5): 715-725.
- [12] Haas, Eric, Wilson, Glen Yahola, Cobb, Casey D., Hyle, Adrienne E., Jordan, Kitty, & Kearney, Kerri S. 2007. “Assessing Influence on the Field: An Analysis of Citations to *Educational Administration Quarterly*, 1979-2003.” *Educational Administration Quarterly*, 43(4): 494-513.
- [13] Habibzadeh, F., & Yadollahie, M. 2008. “Journal weighted impact factor: a proposal.” *Journal of Informetrics*, 2: 164-172.
- [14] Neuhaus, C., Marx, W., & Daniel, H. 2009. “The Publication and Citation Impact Profiles of *Angewandte Chemie* and the *Journal of American Chemical Society* Based on the Sections of *Chemical Abstracts*: A Case Study on the Limitations of the Journal Impact Factor.” *Journal of the American Society for Information Science and Technology*, 60(1): 176-183.
- [15] Sidirpoulos, A., & Manolopoulos, Y. 2005. “Generalized Comparison of Graph-Based Ranking Algorithms for Publications and Authors.” *The Journal of Systems and Software*, 79: 1679-1700.
- [16] Wilson, C.S., & Tenopir, C. 2008. “Local Citation Analysis, Publishing and Reading Patterns: Using Multiple Methods to Evaluate Faculty Use of an Academic Library’s Research Collection.” *Journal of the American Society for Information Science and Technology*, 59(9): 1393-1408.

〈부록〉 영향력지수 상위 10위 문헌(Bibliometrics, 2009년 3월 현재)

(Appendix: Impact Factor Top 10 Articles(Bibliometrics, March, 2009))

1. López Piñero, J.M., & Terrada, M.L. 1992. “Bibliometric indicators and the evaluation of

- medical scientific activity.(II). Scientific communication in distinct areas of the medical sciences." *Medicina Clinica*, 98(3): 101-106. [Cited 53 times].
2. Gøtzsche, P.C., Krog, J.W., & Moustgaard, R. 1995. "Bibliometric analysis of Danish medical research 1986-1992." *Ugeskrift for læger*, 157(37): 5075-5081. [Cited 13 times].
 3. López Piñero, J.M., & Terrada, M.L. 1992. "Bibliometric indicators and the evaluation of medical scientific activity.(III). The indicators of information production, circulation and dispersion, consumption and the repercussions." *Medicina Clinica*, 98(4): 142-148. [Cited 81 times].
 4. Seglen, P.O. 1997. "Why the impact factor of journals should not be used for evaluating research." *British Medical Journal*, 314(7079): 498-502. [Cited 437 times].
 5. García Río, F., Dorgham, A., Serrano, S., Terreros Caro, J.G., Álvarez-Sala, R., Prados, C., Álvarez-Sala, J.L., & Villamor, J. 1997. "Twenty-year evolution of bibliometric indicators for authorship and citation in Archivos de Bronconeumología." *Archivos de Bronconeumologia*, 33(1): 20-26. [Cited 7 times].
 6. Mela, G.S., Cimmino, & M.A., Ugolini, D. 1999. "Impact assessment of oncology research in the European Union." *European Journal of Cancer*, 35(8): 1182-1186. [Cited 33 times].
 7. Alvarez-Sala, J.L., Prados, C., & Alvarez-Sala, R. 1995. "The impact of our journals." *Archivos de Bronconeumologia*, 31(8): 403-406. [Cited 10 times].
 8. Garcia Rio, F. 1999. "Estrategias para una busqueda bibliografica eficiente. bibliometria. valoracion critica." *Archivos de Bronconeumologia*, 35(SUPPL. 1): 27-30. [Cited 13 times].
 9. García-Río, F., Serrano, S., Dorgham, A., Alvarez-Sala, R., Ruiz Peña, A., Pino, J.M., Alvarez-Sala, J.L., & Villamor, J. 2001. "A bibliometric evaluation of European Union research of the respiratory system from 1987-1998." *European Respiratory Journal*, 17(6): 1175-1180. [cited 25 times].
 10. Camí, J., Fernández, M.T., & Gómez Caridad, I. 1993. "Spanish scientific production in biomedicine and health. A study via the Science Citation Index(1986-1989)." *Medicina Clinica*, 101(19): 721-731. [Cited 54 times].

• 국문 참고자료의 영어 표기

(English translation / romanization of references originally written in Korean)

- [1] Dae-Shin Kang, & Sung-Been Moon. 2009. "A Study on Informetric Analysis for Measuring the Qualitative Research Performance." *Journal of the Korean Society for Information*

- Management*, 26(3): 377-394.
- [2] Kap-Seon Kim. 2002. "The Co-occurrence Phenomenon of Both Korean and Non-Korean Literatures Within the Korean Reference-An Analysis on the Citation Motivations and References by Social Scientists." *Journal of the Korean Society for Library and Information Science*, 36(4): 21-47.
- [3] Hee-Jung Kim. 2005. "A Study on Comparison of Intellectual Structure in Records Management and Archives Using Author Cocitation Analysis." *Journal of the Korean Society for Library and Information Science*, 39(3): 207-224.
- [4] So-Young Yu, & Jae-Yun Lee. 2008. "Journal Citation Analysis for Library Services on Interdisciplinary Domains: A Case Study of Department of Biotechnology, Y University." *Journal of the Korean Society for Information Management*, 25(4): 283-308.
- [5] Jae-Yun Lee. 2008. "Bibliographic Author Coupling Analysis: A New Methodological Approach for Identifying Research Trends." *Journal of the Korean Society for Information Management*, 25(1): 173-190.

