

KoMCI 2000 학술지 정보 처리 방법에 대하여 - 학술지 정보의 통일을 제안하며 -

<인용>은 논술이나 주장의 출전으로 다른 저자의 논문이나 논고, 자료 또는 구술을 차용하는 것을 의미한다. 학술지의 인용행태에 따른 인용분석은 인용하는 문헌과 인용된 문헌간의 관계를 밝히는 것으로, 본 KoMCI (Korean Medical Citation Index) 2000 사업은 우리나라 의학분야 연구자들이 연구에 인용한 문헌을 대상으로 인용분석기법을 적용하여 국내외 문헌의존도와 핵심자료를 규명하기 위한 것이다.

즉, 우리나라 의학분야 연구자들이 국내의 특정학술지를 얼마나 자주 인용하고 있는지, 어떤 학술지들이 그 특정학술지를 인용하고 있는지, 각 학술지가 자체 학술지를 얼마나 많이 인용하고 있는지, 그 학술지가 출판된 후 얼마나 빨리, 그리고 오랫동안 인용되고 있는지에 대해 알 수 있도록 한 것이 그 목적인다고 할 수 있다. 이러한 인용분석을 위해서는 원논문(인용한 문헌)과 참고문헌(인용된 문헌)의 서지 기술요소 중에서 인용색인 데이터베이스 구축시에 필요한 데이터요소와 형식을 결정하여 정보처리하여야 한다. 이때 참고문헌의 정확한 서지사항 입력과 함께 기술방법을 표준화하여야 신뢰도 높은 인용분석이 될 수 있다.

KoMCI 2000 학술지에 게재된 논문은 총 7,821편으로 이 논문이 인용한 참고문헌은 170,254개이고, 이중에서 국내문헌은 17,857개(10.5%)이다. 또한 국내문헌 중에서 국내학술지만을 인용한 참고문헌은 14,511개(81.3%)인데 이 수치는 권호, 면수, 연도표기에 있어서 오류를 범하거나 누락한 703개(4.8%)의 문헌을 포함한 것으로 이 중 328개(2.2%)가 권호표기를, 236개(1.6%)가 연도표기를, 109개(0.8%)가 면수표기에 오류를 범하였다. 출판예정인 문헌을 참고문헌으로 인용한 경우는 16개, 비의학 관련 학술지를 인용한 경우로 학술지를 확인할 수 없는 경우는 76개로 나타났다. 또한 KoMCI 2000 학술지가 참고문헌으로 인용한 학술지는 실제로는 416종인데 저자가 이명, 별칭, 영문명 등 여러 형태의 학술지 명칭을 사용함으로써 1,073종으로 나타나 분산된 학술지 정보의 기술방법을 통일시키는 방안의 필요성이 제기되었다.

연구자가 기술한 서지사항은 연구자의 인용행태와 각 학술지의 편집의도에 따라 그 기술방법이 상이할 수밖에 없다. 그러나 이를 있는 그대로 입력하고 저자나 출판사의 오류를 수정치 않거나 참고문헌의 기술방법을 표준화하지 않는다면 동일한 값을 가지는 데이터가 분산 처리될 수밖에 없어 결국 각종인용지표의 결과값을 신뢰할 수 없게 된다.

따라서 KoMCI 2000 학술지의 정보처리를 정확하고 효율적으로 하기 위하여 서지정보 입력에 앞서 KoMCI 2000 학술지 69종을 인용색인 데이터베이스로 구축할 경우 필요한 데이터 요소 및 각 정보 처리시에 필요한 사항을 결정하였다. 그리고 연구자가 기술한 서지사항에 대하여 각종 서지 데이터베이스의 검색 및 실제 학술지의 확인작업을 통해 참고문헌의 정확한 서지정보를 입력하였다. 또한 학술지명이 저자에 의하여 이명이나 별칭 등으로 사용되었을 때 시스템상에서 여러 형태로 표기된 학술지명을 동일한 학술지명으로 인식할 수 있도록 학술지 명칭을 통일하여 각종 인용지표가 분산되지 않도록

록 데이터를 처리하였다. 끝으로 국내 의학관련 학술지의 학술지명칭을 통일하여 KoMCI 2000 학술지 정보처리에 사용한 “(대한의학회 제정) 한국의학학술지명 및 약어명칭 표기방안”을 제안하여 각 학회가 유용하게 사용할 수 있도록 권고하고자 한다.

1. KoMCI 2000 학술지의 정보 처리 방법

KoMCI 2000 학술지 69종의 입력은 원논문과 참고문헌으로 나누어 진행하였다. 원논문과 참고문헌에 고유의 source 논문ID를 부여하여 원논문과 참고문헌을 서로 연결시킴으로써 각 데이터의 인용관계를 확인하는 작업에 활용할 수 있도록 하였다.

원논문의 입력은 KoMCI 2000 사업의 각종 인용정보에 영향을 주지는 않으나 향후 KoMCI 사업이 인용색인정보 시스템으로서 발전해 나갈 때 기초자료가 되어 보다 원활한 시스템을 구축할 수 있을 것으로 기대한다. 원논문의 데이터요소는 학술지명, 한글논문명, 영문논문명, 한글저자명, 영문저자명, 발표년도, 발표월, 권(volume), 호(issue), 시작페이지, 마지막 페이지, 문헌유형이다. 문헌유형의 경우 article, case report, review, editorial, letter만을 대상으로 하되 우리나라 특성상 심포지움 자료 (전체 논문이 실린 것)도 review로 간주하여 입력하였다. 또한 학술대회 자료집의 경우 실제 학술지의 발표년도와 권(volume)이 일치하고 페이지를 승계하고 있을 경우 입력대상에 포함하였다.

참고문헌은 각종 인용지표에 영향을 주는 것으로 저자의 오류 혹은 편집상의 실수로 서지사항이 잘못 표기된 경우 기사명과 첫번째 저자, 발행년도 순으로 각종 서지 데이터베이스 및 실제 학술지를 참조하여 수정하는 것을 원칙으로 하였다. 참고문헌의 데이터요소는 첫번째 저자명, 학술지명, 연도, 권, 시작페이지, 자료의 유형이다.

참고문헌 입력시 문제점은 저자명의 경우 원논문이 한글로 된 학술지임에도 불구하고 각 학술지의 편집의도에 따라 영문저자명을 사용했거나, 저자명이 누락되었거나, 동일한 저자명의 표기형식이 다르거나, 동명이인인 경우로 서지사항의 확인작업이 어렵다는 것이다. 또한 학술지명의 경우 학술지명이 누락되었거나, 학술지명의 기술형태가 이명, 별칭, 약어, 영문명 등으로 다양하게 사용되어 동일한 값을 가지는 데이터가 분산될 수 있다는 것이다. 따라서 학술지명은 일단 초기 입력시 저자가 사용한 학술지명의 표기형태를 인정하여 그대로 입력하였다. 이후 이명, 별칭, 한자, 약어, 영문명 등에 대한 학술지명 전거파일을 작성, 통일된 학술지명으로 시스템이 자동 변환하여 동일한 학술지로 인식, 입력된 레코드와 합산하여 각종지표를 계산할 수 있도록 하였다. 향후 인용색인데이터베이스 구축시에는 저자명과 학술지명에 대한 전거통제(authority control)를 해야만 인용분석과 정보검색이 용이해질 것이다.

발행년도는 연호를 사용한 경우는 연도로 계산하여 입력하였고 발행년 표시가 없거나 오기로 추정되는 경우는 서지사항 확인 후 실제 발행년도를 입력하였다. 단, 한 참고문헌 안에 복수의 발행년도로 표기된 경우는 공란으로 입력, 통계처리시 제외될 수밖에 없었다. 특히 연도표기가 잘못된 경우 각종 인용지표의 분석에 지대한 영향을 미치는 변수가 되므로 정확한 데이터의 입력을 위하여 기사명을 중심으로 각종 서지 데이터베이스를 검색하고 각 학술지의 논문을 실제 대조하여 이를 수정하였다.

자료의 유형상 저자가 단행본인지 학술지인지 구분하지 않고 참고문헌을 기술한 경우 실제로도

단행본 혹은 학술지로 볼 수 있는 자료, 특히 학술대회 자료집 중 그 성격이 학술지의 부록으로 규정할 수 있는지가 모호한 경우는 권호표기가 되어 있고 발간년도가 실제 학술지와 일치하는 경우 학술지의 부록으로 인정하여 각종 인용지표에 포함시켰다.

권, 호가 애매한 신문 등의 자료는 출간된 월, 일을 기입하여 “한겨레신문 0814 (한겨레신문 8월 14일자)”와 같이 처리하였고, 전자문헌을 인용한 경우 저자가 기술한 그대로 입력하였는데 이는 아직 서지 기술규칙이 정립되지 않았기 때문이다.

출판 예정인 미간행자료는 서지사항이 불분명하여 검색이 쉽지 않고 발행여부가 불명확하였는데, 각종 서지 데이터베이스와 실제 학술지를 확인한 후 발행된 자료에 한하여 서지사항을 수정하여 입력하였다.

2. 학술지정보의 통일방안

KoMCI 학술지를 정보처리하는 과정에서 가장 어려웠던 점은 전술한 바와 같이 KoMCI 2000 학술지 69종이 인용한 학술지의 경우 실질적으로 416종인데 반해 저자가 이명, 별칭, 한글약어명, 영문잡지명, 영문약어명, 이전학술지명 등 여러 가지 명칭을 사용함으로써 1,073종으로 분산되어 출현하였다는 것이다. 이에 학술지의 서지정보를 체계적으로 정리할 필요성이 제기되어 학술지명 및 학술지약어명, 학술지의 영문명, 학술지의 영문약어명, ISSN, 창간년도, 간기, 주제분야, 이전서명 및 다른 표제서명 등을 학술지의 표제지, 학회의 home page, 각종 서지 데이터베이스, 연구자의 인용형태를 통해서 분석하였다. 이렇게 분석한 한국의학학술지의 서지정보를 “(대한의학회 제정) 한국의학학술지명 및 약어명칭 표기방안”으로 제안하고자 한다.

의학관련 학술지의 선정은 다음과 같다.

첫번째 군은 대한의학학술지편집인협회의 “국내 의학학술지 평가”에서 2000년까지 KoreaMed 학술지로 선정된 69종의 학술지를 KoMCI 2000 학술지로 규정하고 이를 자모순에 의하여 학술지번호(JID) 1번에서 69번까지 할당하였다.

두번째 군은 대한의학학술지편집인협회의 2001년도 “국내 의학학술지 평가”에서 KoreaMed 학술지로 추가된 8종을 JID 70번에서 77번까지 부여하고, 현재 발간되고 있는 전문적인 의학관련 분야의 학술지이거나 69종의 KoMCI 학술지가 3회이상 인용한 국내 의학학술지를 JID 301번에서 512번까지 부여하였다.

세번째 군은 KoMCI 학술지가 한두번 인용한 간행물이거나 비의학관련 학술지, 폐간된 의학학술지 등을 포함한 학술지군으로 JID 701번에서 874번이다.

이중 첫번째와 두번째 군 289종에 대하여 다음과 같은 원칙으로 “(대한의학회 제정) 한국의학학술지명 및 약어명칭 표기방안”을 작성하였다.

- 한글학술지명: 학술지 표제지의 서명을 학술지명으로 하였다.
- 한글학술지약어명: ‘대한’ → 대, 잡지 → ‘지’, ‘한국’ → ‘한’의 기본원칙 하에 학술지명을 대표할 수 있는 단어나 단어의 약어를 포함하되, 최대 5자를 넘지 않는 범위에서 선정하였다.

- 영문학술지명: 학술지 표제지 및 책등서명, 판권기, 학회홈페이지, 각 서지색인 데이터베이스 등의 정보를 근거로 영문학술지명을 부여하였다.
- 영문학술지약어명: MEDLINE의 학술지명 및 약어명칭 표기법에 준하고 최신 학술지 약어명 등은 PubMed journal browser 및 “Abbreviations, Acronyms & Initials (<http://kamje.or.kr>)”를 참조하였다.
- 영문코드: 영문학술지명에서 각 단어의 첫 자를 취하고, journal_code가 중복될 경우 겹치지 않는 다음 단어에서 첫 자를 취하여 최대 6자를 넘지 않도록 하였다.
- 주제분야: 외국학술지의 주제분야와 통일하기 위하여 SCI가 사용하는 학술지 주제분류 “Subject Category: SC codes” list를 그대로 적용하여 학술지번호(JID) 1번에서 512번의 289종에 대하여 Subject Category 부여작업을 하였다.

학술지정보의 통일을 위하여 활용한 정보는 Korean Index Medicus, KoreaMed, JCR 2000, 한국학술정보 데이터베이스, 의학연구정보센터 데이터베이스, 보건연구정보센터 데이터베이스, 연구개발정보센터 데이터베이스, 한국의학도서관협의회 종합목록 데이터베이스, 국회도서관 정기간행물기사색인 데이터베이스, 국립중앙도서관 한국문헌번호센터 데이터베이스이고, 실물확인이 가능한 학술지의 경우는 이를 우선으로 하였다.

본 사업에서 제외된 사항은 다음과 같다.

첫째, 저자의 소속기관명 입력은 이번 사업에서 제외하였다. 그러나 이는 연구자 개인의 연구행태 및 각급 연구기관의 연구수준을 평가할 수 있는 도구로 활용할 수 있는 근거가 되므로 추후 사업에서는 기술해야 하는 데이터 요소이다.

둘째, 저자명에 대한 통일화 방안은 사업내용에서 제외하였다.

셋째, 주제색인에 사용될 수 있는 주제어(MeSH)의 처리는 제외하였다.

그러나 한 분야에서 인용 및 정보 탐색을 수행할 때 충분한 정보가 없더라도 다양한 분야의 논문을 최대한 이용하기 위해서는 중요한 필수요소이다.

넷째, 69종의 학술지 중 2000년도에 발행이 되지 않은 Journal of Genetic Medicine은 인용문헌정보 입력에서 제외하였다.

다섯째, 학술지 서지정보는 본 사업의 진행기간인 2001년 9월에서 12월에 조사된 것으로 2002년 1월 이후 변경된 내용은 반영치 않았다.

3. 맺는 말

본 사업은 우리나라 최초로 의학관련 학술지를 대상으로 인용색인 데이터베이스를 구축할 때 필요한 각 데이터 요소를 선정하여 입력하고, 의학학술지 정보를 통일화할 수 있는 방안을 모색하여 한국 의학 및 의학관련 학술지의 인용도를 분석하였다는데 큰 의의를 둔다.

인용색인정보의 정보처리 측면에서 볼 때 각 학회의 학술지 출판시에 고려해야 할 문헌기술 관련 사항을 보면 다음과 같다.

첫째, 서지기술의 부정확성은 인쇄상의 잘못도 있지만 주로 저자와 관련된 문제이다. 이를 각 학회의 편집위원회가 책임을 맡아 인용문헌 작성시 매체별 필수 요소와 최소한의 규칙을 모든 저자에게 주지시키고 확인할 필요가 있다.

둘째, 기본적으로 참고문헌의 기술형식을 표준화할 필요가 있다.

참고문헌의 필수 기술요소 중에서 학술지명은 이전서명, 이명, 별칭, 영문명, 약어명 등이 저자 혹은 학술지의 기술원칙에 따라 다양하게 표현되고 있어 전거통제(Authority control)를 해야 하는 요소이다. 또한 저자명 역시 영문, 영문약어, 한문 등이 혼용되어 나타나 이를 통제할 수 있는 준거가 필요하다. 따라서 KoMCI 2000 사업에서는 학술지정보를 통일할 수 있는 준거를 마련하여 “(대한의학회 제정) 한국의학학술지명 및 약어명칭 표기방안”을 제안하였다. 학술문헌의 탐색문제에 인용색인개념을 적용한 것은 1964년 Science Citation Index가 처음이다. 이 SCI는 인용된 문헌과 인용한 문헌간을 체계적으로 편성한 색인으로서 색인작성과정에 색인자의 개입이 불필요하고 문헌간의 상호관계를 파악할 수 있는 색인기법으로 학술지의 평가도구로 인정받고 있다. 국내에서도 의학계가 지속적인 지원과 관심을 갖고 KoMCI 2000 사업을 바탕으로 SCI와 같은 색인을 계속 구축해 간다면 각 주제영역별 연구자의 실적과 학술정보망 및 연구주제의 발전과정을 분석, 파악하는 도구로 유용하게 사용될 수 있을 것이다. KoMCI 사업은 국내 의학관련 학술지의 평가도구로서 국내 의학학술지의 질적 향상을 도모하여 의학발전에 기여할 것으로 본다.

가톨릭대학교 성의교정 도서관 부장

한국의학도서관협의회 고문 김 호 배