

국가과학기술정보센터(NDSL)의 논문 콘텐츠 가치평가에 관한 연구 A Study on Value Evaluation for Article Contents on NDSL

이 선 희, 김 혜 선
한국과학기술정보연구원

Seon-Hee Lee, Hye-Sun Kim
Korea Institute of Science and Technology Information

요약

국내 과학기술 연구자의 요구를 만족시키는 고품질 콘텐츠를 제공하기 위한 기초작업으로 국가과학기술정보센터(NDSL) 콘텐츠의 가치평가 기준을 수립할 필요가 있다. 본 연구는 NDSL의 대표적인 콘텐츠인 논문의 가치에 영향을 미치는 요인들을 도출하고 각 요인별 측정 방안을 모색하였다. 제시된 콘텐츠 가치평가 기준에 의거하여 NDSL 콘텐츠 표본을 평가하고, 콘텐츠 가치평가 방안 기준이 적절한지를 분석하였다.

I. 서론

한국과학기술정보연구원(KISTI)은 1962년 이후 과학기술분야 콘텐츠를 국내 연구자들에게 제공하여 왔다. KISTI가 운영하는 전자도서관인 국가과학기술정보센터(National Digital Science Library 이하 NDSL)는 1억 건이 넘는 과학기술분야 콘텐츠를 제공하고 있으며 이용자들이 의하여 연간 1억 뷰와 550만 건의 다운로드가 이루어지고 있다. 그러나 망라성에 의하여 구축된 콘텐츠 중에서 품질이 우수하지 못하거나 이용률이 낮은 콘텐츠는 서비스 속도를 내기 못하거나 재현율을 낮추는 요인이 될 수 있다. 따라서 이용자의 요구를 만족시키는 고품질 콘텐츠를 제공하기 위한 기초 작업으로 전통적으로 공공재로 인식되어온 콘텐츠에 대한 가치평가 기준을 수립할 필요가 있다. NDSL 콘텐츠 중 국내의 학술지나 학회에서 발표된 논문은 이용의 70% 이상을 차지하는 핵심 콘텐츠이다. 본 연구는 NDSL 논문 콘텐츠 가치에 영향을 미치는 요인들을 추출하고 요인별 측정 방안을 제시하였다. NDSL에서 검색어로 추출한 표본 논문을 제시된 콘텐츠 가치평가 기준에 의거하여 평가하였으며, 콘텐츠 가치평가 방안 기준의 장점과 한계점을 분석하고 보완안을 제안하였다.

II. 콘텐츠 가치평가 기준 및 평가 방안

본 연구는 유승훈(2009)의 지식정보 가치평가 모형을 참조하여, 학술논문에 대한 콘텐츠 가치평가지표에 이용적 가치와 내재적 가치를 적용하였다. (표 1)에서 보는 바와 같이 이용성 가치 세부평가지표는 전문(full-text)접근성, 이용빈도(전문 이용빈도, 페이지뷰 이용빈도), 언어가독성이다. 내재성 가치 세부평가지표는 권위, NDSL저널이용지표, 완전성, 최신성, 정확성이다. 정확성을 제외한 세부평가지표는 NDSL 논문의 메타데이터, 통계데이터, 전문 존재 여부 등에 의해 객관적으로 분석할 수 있다. 정확성은 정성적 평가로 찾고자 하는 주제 혹은 검색어와 검색결과로 나온 논문 내용의 연관성을 평가한다. 정확성은 객관적인 세부평가지표의 결과가 정확한지를 평가하기 위하여 도입하였다.

표 1. 논문 콘텐츠 가치별 세부평가지표

가치	세부평가지표		평가대상	판단기준	
				객관적	주관적
1. 이용성	1.2 이용빈도	1.1 전문(full-text)접근성	논문	o	
		1.2.1 전문 이용빈도		o	
		1.2.2 페이지뷰 이용빈도		o	
	1.3 언어 가독성	o			
2. 내재성	2.1 권위		수목학술지	o	
	2.2 NDSL 이용지표		논문	o	
	2.3 완전성			o	
	2.4 최신성			o	
	2.5 정확성				o

콘텐츠 가치 세부평가지표별 점수기준은 (표 2)와 같다. 논문 가치평가 점수는 세부평가지표별 점수를 합한 (24점 = 8개 * 3점) 후에 NDSL저널이용지표를 합한 것이다.

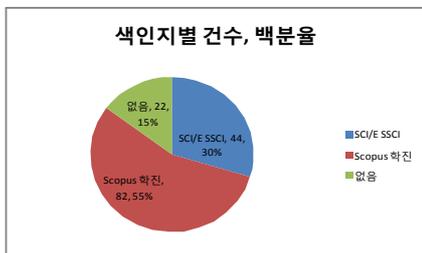
표 2. 논문 콘텐츠 가치 세부평가지표별 점수 기준

가치	세부평가지표	점수	점수기준	
1. 이용성	1.1 전문접근성	3	3점: OA 1점: KESLI 구독시 접근 가능 0점: 전문접근 불가, KESLI 비구독	
	1.2 이용빈도	1.2.1 전문 이용빈도	3	3점: 5건 이상, 2점: 3건 이상 1점: 1건 이상, 0점: 없음 *전문다운로드 건수와 페이지뷰 건수는 평가기준은 동일, 값은 별도 처리
		1.2.2 페이지뷰 이용빈도	3	
	1.3 언어가독성	3	3점: 한글 2점: 영어 1점: 스페인어, 중국어, 일어, 불어, 독어 0점: 기타 언어	

2. 내재성	2.1 권위	3	3점: SCI, SCIE, SSCI, AHCI 등재 2점: SCOPUS, 학진등재/후보 1점: 기타, 0점: 미등재
	2.2 NDSL저널 이용지표	이용지표	NDSL저널이용지표(3년간 해당 저널 원문보기 건수를 국내저널 평균원문보기 건수로 나눔)
	2.3 완전성	3	3점: 전문, 초록, 저자, 출판사, 발행년도, 학술지명, 권호정보 있음. 2점: 초록, 저자, 출판사, 발행년도, 학술지명, 권호정보 있음. 전문 없음. 1점: 최소한 서지정보. 저자, 출판사, 학술지명, 발행년도 있음. 전문, 초록 없음. 0점: 기타
	2.4 최신성	3	3점: 2010년 이후, 2점: 2005-2009년 1점: 2000-2004년, 0점: 2000년 이전
	2.5 정확성	3	3점: 내용이 정확 2점: 내용이 다소 정확 1점: 내용이 다소 관련, 0점: 내용 무관

III. 콘텐츠 가치평가 평가 결과

NDSL에서 검색어 ‘(기관리포지토리 or 기관레포지토리) or (institutional and (repository or repositories)) or 기관저장소’로 실행하여 얻은 결과 867건 중에서 표본 148건(전체의 17%)을 추출하였다. 해외논문은 결과 중 상위 100건을 추출하고 국내논문은 48건을 추출하여 세부평가지표 점수기준에 의거하여 점수를 환산하였다. 점수가 높을수록 가치가 높은 것으로 평가된다. 이용빈도는 2013년 7월부터 12월까지 6개월간 NDSL의 논문 전문 다운로드 건수와 페이지뷰 건수를 활용하였다. 전체 148건 중 최고점수는 25.65점으로 노영희 교수의 논문으로 ‘dCollection의 링크드데이터 구축에 관한 연구’였다. 상위 30위는 모두 국내논문이었다. 상위 50위 안에 해외논문은 단 8건이었다. 국내논문이 높은 평가를 받은 것은 NDSL 이용자들이 국내문헌을 더 많이 활용하고 있는 현실을 반영하고 있다. NDSL저널이용지표가 있는 경우는 41건으로 모두 국내 학진 등재저널이었다. <그림 1>과 같이 세계적으로 권위 있는 색인지인 SCI/E, SSCI 수록 학술지의 논문은 전체 148건 중 44건(30%)이었고 모두 해외논문이었다. SCOPUS와 학진등재지에 수록된 논문까지 합하면 전체의 85%가 국내외 색인지에 수록된 우수 논문이었다.



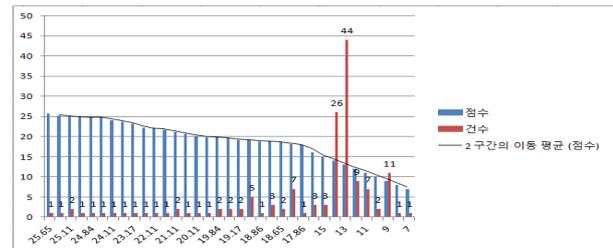
▶▶ 그림 1. 논문 수록 학술지의 색인지 등재비율

국내논문의 경우, 전문접근성, 언어가독성, 서지완전성, NDSL저널이용지표에서 높은 점수를 얻었고 해외논문은 동일한 세부평가기준에서 상대적으로 점수를 적게 얻었다. 정확도 분석결과 ‘매우 관련 있음(3점)’인 논문은 148건 중 134건(90.5%)이었고 ‘대체로 관련 있음(2점)’은

8건(5.4%)이었고 ‘일부 관련 있음(1점)’은 1건이었고 ‘전혀 관련 없음(0점)’이 5건(3.3%)이었다. 검색어인 ‘리포지토리’가 문헌정보학이나 IT분야뿐만 아니라 핵 관련 분야에서도 활용되는 동음이의어이어서, 용어는 동일하나 활용하는 내용이 전혀 다른 원자력 분야 논문이 나타나기도 했다. 정확도를 제외한 객관적 세부평가기준들은 검색결과와 기준 없이도 메타데이터나 전문 제공 여부 등 콘텐츠 자체만으로도 평가할 수 있었다.

IV. 결론 : 콘텐츠 가치평가 방안에 대한 분석

콘텐츠 가치평가의 장점은 NDSL 논문 가치평가 모델로 활용이 가능하며 자동화하여 평가할 수 있고 가치를 수치화할 수 있고 인포그래픽으로 표현 가능하다는 것이다. 한계점은 NDSL에서 제공되는 논문은 약 7천만 건으로 평가하려면 노력과 시간이 소요된다는 것이다. 또한 세부평가지표에 대한 검토와 조정이 필요하다. NDSL저널이용지표는 국내학술지에만 해당하고 해외학술지는 해당되지 않아, 해외학술지 보다 국내학술지에 수록된 논문의 가치가 더 높은 것으로 계산되었다. <그림 2>와 같이 동일한 평가점수가 많았다. 예를 들면, 148건 중 13점이 44건으로 29.7%를 차지하였다. 14점(26건)과 12점(9건)을 합하면 53.3%이다. 가치평가에 동일한 점수에 해당하는 건수가 많아 논문 전체인 7천만 건을 대상으로 평가한다면 많은 논문이 동일한 점수에 해당하므로 결과를 세분화되는 것이 쉽지 않다. 만일 개인화 서비스에 가치평가 방안을 적용할 경우, 동일한 가치에 속하는 논문이 많아 서비스에 혼란을 초래할 수 있다. 정확성의 경우는 평가자의 판단이 필요하며 기계적 분석보다 이용자의 판단이 중요하다. 원자력 분야 리포지토리가 결과에 나온 것처럼 관련 없는 주제를 평가하는 기준도 있어야 한다는 문제점이 있다.



▶▶ 그림 2. 콘텐츠 가치평가 점수별 해당 건수

향후 보완방안을 살펴보면 다음과 같다. 이용자 관점에서 만족도조사 등이 실시하여 평가하여야 한다. 세부평가 지표 중 이용자에 따라 상대적인 특성을 지닌 정확성, 언어가독성, 최신성 등은 이용자만족도조사 시 활용되는 것도 바람직해 보인다. 내재성의 권위나 NDSL저널이용지표는 논문이 수록된 학술지가 평가대상으로 향후 논문 위주로 가치를 평가할 수 있는 방안이 모색되어야 한다.

참고 문헌

- [1] 김혜선, 대학도서관 외국학술지 구독 영향요인 분석, 이화여자대학교대학원 박사학위논문, 2012.
- [2] 유승훈, 허재용, 안윤기, “지식정보 가치평가 모형의실증적 개발”, 정보관리연구, 제40권, 제1호, pp.113-132, 2009년.
- [3] 한국문헌정보학회, 최신 문헌정보학의 이해, pp. 43-45, 한국문헌정보학회, 2013.
- [4] NDSL(<http://www.ndsl.kr>)