

정보원천을 정량적으로 평가하기 위한 몇가지 지표 연구

김 지 현

위대한 령도자 김정일동지께서는 다음과 같이 교시하시였다.

《선진과학기술을 적극 받아들이자면 과학기술정보사업을 활발히 벌려야 합니다.》

(《김정일선집》 증보판 제11권 145페이지)

정보수집은 사용자의 특정한 수요에 따라 관련된 정보원천들에 대하여 여러가지 방법으로 과학적인 검색, 조사, 획득, 정리, 분석을 진행함으로써 최종적으로 필요하고 쓸모있는 정보를 얻는 과정이다.

모든 정보는 일정한 정보원천으로부터 얻어진다.

정보원천은 사람들이 과학연구활동, 생산경영활동과 기타 모든 활동에서 이룩한 성과와 각종 원천기록들을 가공정리하여 얻은 결과물이라고 볼수 있다.

정보원천은 전통적인 인쇄문헌자료는 물론 전자도서자료도 포함하고있다.

과학기술의 급속한 발전은 정보원천량의 방대한 증가를 가져왔다.

통계에 의하면 2010년말까지 존재하고있는 기록된 웹페이지만 해도 600억개에 달한다고 한다.

정보원천의 량이 급속히 증가됨으로써 사람들이 유용한 정보를 얻는데 훨씬 더 많은 시간과 비용을 들이고있다.

정보원천을 과학적인 방법론에 기초하여 정량적으로 평가하고 선별하는것은 적은 시간과 비용으로 더욱 유용한 정보들을 더 많이 수집하기 위한 선차적공정이다.

정보원천평가는 평가자들의 견해와 관점에 따라 결과에서 일련의 차이가 있을수 있다. 평가에 참가하는 성원들이 어떤 지표를 가지고 평가를 진행하는가, 매 지표의 채점값범위를 어떻게 설정하는가, 매 지표의 무게는 어떻게 할당하는가 등이 평가자들의 견해와 관점에 따라 달라질수 있다. 그러므로 정보원천평가결과는 상대적이며 평가자들의 수가 많을수록 그리고 평가자들의 주관적인 영향을 적게 받는 평가지표들에 기초한 평가결과일수록 보다 정확하다고 볼수 있다.

정보수집을 위한 정보원천평가에는 직접적인 평가방법과 간접적인 평가방법이 있다.

직접평가법은 서로 다른 각도에서 정보원천평가지표와 점수기준을 설계하고 평가분석 및 종합하며 종합된 점수값에 근거하여 정보원천의 가치를 평가하는 방법이다.

이렇게 정보원천을 평가하는 방법은 다음과 같다.

첫째: 정보원천평가지표와 점수기준을 설계한다.

둘째: 점수평가기준에 근거하여 각이한 정보원천의 매 지표에 대한 평가를 진행한다.

셋째: 정보원천의 총 점수를 계산한다.

넷째: 총점수의 크기에 따라 정보원천의 가치를 평가한다.

정보원천의 가치평가를 보다 과학적으로 진행하기 위해 중요성정도에 따라 각 지표들에

무게값을 부여하고 매 항목지표가 얻은 점수를 각각 대응되는 무게값으로 곱한 다음 다시 계산하여 합계를 얻는다. 이렇게 정보원천의 가치에 대한 보다 정확한 점수평가를 진행한다.

정보원천평가지표통계표

평 가지 표	정 보 원 천 류 형					
	F_1	F_2	...	F_j	...	F_n
지 표 ₁	f_{11}	f_{21}	...	f_{j1}	...	f_{n1}
지 표 ₂	f_{12}	f_{22}	...	f_{j2}	...	f_{n2}
...
지 표 _i	f_{1i}	f_{2i}	...	f_{ji}	...	f_{ni}
...
지 표 _m	f_{1m}	f_{2m}	...	f_{jm}	...	f_{nm}
총 점	$\sum_{i=1}^m f_{1i}$	$\sum_{i=1}^m f_{2i}$...	$\sum_{i=1}^m f_{ji}$...	$\sum_{i=1}^m f_{ni}$

표에서 f_{ji} 는 정보원천 F_i 의 j 번째 지표에 대한 점수평가자의 평가점수이다. 정보의 가치가 비교적 높은 정보원천은 그 평가점수가 비교적 높다. 만약 가치가 제일 높은 정보원천을 10점으로 평가한다면 가치가 제일 낮은 정보원천은 1점으로 평가한다.

정보원천에 대한 직접적인 평가방법에서는 다음의 지표들을 리용하여 정보원천의 가치를 평가할수 있다.

첫째: 믿음성—믿음성은 정보원천이 제공하는 정보가 진실하고 신뢰할수 있는 정도를 말하는데 이것은 정보원천의 질을 평가하는 제일 중요한 지표이다.

둘째: 시기성—이것은 정보원천이 형성된 때로부터 그로부터 필요한 정보가 리용자들에게 접수되기까지 얼마만한 시간이 걸리는가를 나타낸다.

셋째: 련관성—이것은 정보원천을 통하여 제공되는 정보내용이 사용자의 수요에 부합되는 련관정도를 가리킨다.

넷째: 신규성—이것은 정보원천이 새로운 내용 (견해, 리론, 기술, 가설, 설제, 기술공정 등)을 포함하고있는 정도와 그 갱신빈도를 가리킨다.

다섯째: 전면성—정보원천에 포함된 정보의 폭과 깊이를 가리킨다.

여섯째: 체계성—정보원천에 수집된 정보의 완전성정도를 가리킨다.

일곱째: 사용편리성—정보원천을 획득하고 사용할수 있는 편리성정도를 가리킨다. 레하면 특수한 설비 또는 조작절차를 필요로 하는가 등을 나타낸다.

여덟째: 경제성—정보원천을 획득할 때 지불하는 원가의 크기를 가리킨다. 여기에는 시간원가, 인력원가, 비용원가 등을 포괄한다.

아홉째: 수익성—정보원천이 제공하는 정보의 가치와 그것이 가져오는 사회경제적리익을 가리킨다.

열째: 정보량—포함된 정보량의 크기와 사용자에게 제공하는 유용한 정보량의 크기

를 가리킨다.

이 평가법은 정보원천을 간단히 평가할수 있고 여러가지 유형의 정보원천에 대하여 동시에 채점을 진행할수 있는 우점이 있다. 이 방법은 대체로 정보원천의 중요성을 반영하며 시행이 편리하고 조작이 간단하다. 그러나 이 방법에서는 각 항목지표의 채점에 대하여 채점자의 주관적인 판단에 완전히 의거하며 비록 무게값평가가 첨가되었지만 무게값의 판정 역시 주관적으로 진행하기때문에 채점자의 주관적영향을 비교적 크게 받는다.

우의 정보원천평가법을 리용하여 정보원천을 평가해보자.

정보수집성원들이 다년간의 정보수집사업경험에 기초하여 그들이 속해있는 단위의 각 이한 정보원천에 대한 점수채점을 진행하여 시기성, 종합성, 경제성, 정확성, 획득편리성 등 5가지 지표에 대하여 각종 정보원천들에 대한 평가를 진행하였는데 표와 같다고 하자.

정보원천평가지표점수채점표

평가지표	정보원천유형												
	I	II	III	IV	V	VI	VII	VIII	IX	X	X I	X II	X III
시기성	10	5	1	8	8	10	3	4	8	2	4	8	10
종합성	4	6	10	8	6	9	8	8	9	9	7	6	4
경제성	8	7	4	9	10	9	6	10	6	5	9	8	10
정확성	5	8	8	10	3	10	10	9	10	8	7	8	5
획득편리성	8	7	4	7	10	8	8	6	6	6	4	6	9
총점	35	33	27	42	37	46	35	37	39	30	31	36	38

그중에서 I는 대중전파수단(텔레비존, 방송, 영화 등)이다. II는 정기간행물이고 III는 비정기출판물(종합편집, 설명서, 지도첩 등), IV는 과학기술보고, 기술총화 등, V는 제품자료(건본, 사용설명, 광고 등), VI는 기술문건, VII는 표준문헌, VIII는 특허설명서, IX는 생산계획과 총화, 통계보고서 등, X는 정부출판물, XI는 학술논문, XII는 회의기록, 학술보고, 전람회자료, XIII는 구두정보원천이다. 비용이 제한되어 13개의 정보원천중 10개의 정보원천만 획득하려 한다고 하자.

먼저 정보원천의 가치를 분석한다. 총점에 의하여 I부터 XIII까지의 평가점수를 35, 33, 27, 42, 37, 46, 35, 37, 39, 30, 31, 36, 38로 구분한다. 정보원천들을 그 가치순서대로 배열하면 VI, IV, IX, XIII, V, VIII, XII, I, VII, II, XI, X, III이다.

이러한 평가에 기초하여 정보원천들을 선별할수 있다. 평점에 근거하여 XI(학술논문), X(정부출판물), III(비정기출판물) 등 3가지 정보원천을 버리고 VI, IV, IX, XIII, V, VIII, XII, I, VII, II의 10가지 정보원천을 우선적으로 선택해야 한다.

직접적인 평가방법에 기초하여 정보원천의 가치를 평가하는것은 평가자의 주관적인 견해가 일정하게 반영되는 결함이 있다.

간접평가법은 정보사용자의 정보원천에 대한 수요와 리용정형을 조사하여 정보원천을 평가하는 분석방법이다.

간접평가법에서는 조사표를 리용하여 정보원천평가를 진행한 후 조사표의 자료에 대한 분석대비를 진행하며 최종적으로 정보원천의 가치를 평가결정한다.

이 방법은 정보원천에 대한 직접평가법과는 달리 조사와 점수채점을 호상 결합하여 많은 사용자들에 대한 점수채점통계분석을 진행함으로써 주관적인 점수채점의 부족점을 극복하고 평가결과는 비교적 객관적이다.

그러나 이 방법에서는 비교적 강한 조사표통계분석능력을 지닐것을 요구한다.

정보원천에 대한 간접평가방법은 다음과 같은 절차를 거친다.

첫째로, 조사항목을 설정하고 정보원천평가조사표를 설계한다.

둘째로, 점수채점기준을 설정한다.

셋째로, 조사표를 내준다.

넷째로, 조사표를 회수정리한다.

다섯째로, 자료들을 통계분석한다.

여섯째로, 평가결론을 얻는다.

이 과정에 정보원천평가정형을 과학적으로 분석하기 위하여 반드시 회수된 조사표에 대하여 효과적인 분석을 진행해야 하며 채점이 지나치게 일치하거나 빠지거나 보태진 항목이 비교적 많은 조사표들은 없애고 나머지조사표들에 기록된 사용자의 평가점수를 정보원천간접평가채점표에 종합한다.

정보원천간접평가점수표는 그림과 같다. 이 표는 사용자 m 명으로부터 종합된 n 개의 정보원천에 대한 평가자료이다.

정보원천간접평가점수표

피조사 리용자	정보원천부호					
	F_1	F_2	...	F_j	...	F_n
A_1	a_{11}	a_{12}	...	a_{1j}	...	a_{1n}
A_2	a_{21}	a_{22}	...	a_{2j}	...	a_{2n}
\vdots	\vdots	\vdots	...	\vdots	...	\vdots
A_i	a_{i1}	a_{i2}	...	a_{ij}	...	a_{in}
\vdots	\vdots	\vdots	...	\vdots	...	\vdots
A_m	a_{m1}	a_{m2}	...	a_{mj}	...	a_{mn}
총계	R_1	R_2	...	R_j	...	R_n

표에서 a_{ij} 는 조사된 사용자 A_i 가 정보원천 F_j 에 대하여 진행한 평가점수이다. 비교적 정보값이 있는 정보원천은 그 평가점수가 비교적 높다. 가치가 제일 높은 정보원천은 10점, 가치가 제일 낮은 정보원천은 1점으로 평가한다.

다음의 4가지 지표를 리용하여 정보원천에 대한 비교분석을 진행할수 있다.

① 점수평균값

조사에 동원된 사용자들이 정보원천에 대하여 평가한 점수값들을 리용하여 이 정보원천에 대한 평균점수값을 알수 있는데 그 계산공식은 다음과 같다.

$$R_j = \frac{\sum_{i=1}^m a_{ij}}{m}$$

식에서 R_j 는 정보원천 F_j 에 대한 평균점수이고 m 은 조사에 동원된 사용자수이다. a_{ij} 는 조사된 사용자 A_i 의 정보원천 F_j 에 대한 평가점수이다.

이 평가지표는 정보원천들의 가치를 비교평가하기 위한 가장 일반적인 지표이다. 이 지표에는 해당 정보원천에 대한 평균적인 평가점수는 반영되어있지만 다른 정보원천들과의 호상관계는 직접적으로 반영되어있지 않다.

② 점수비중

모든 조사된 사용자들의 어떤 정보원천에 대한 점수의 합이 모든 조사된 사용자들의 모든 정보원천에 대한 점수의 합에서 차지하는 비중은 곧 해당 정보원천의 점수비중이며 그 계산공식은 다음과 같다.

$$R_j = \frac{\sum_{i=1}^m a_{ij}}{\sum_{j=1}^m \sum_{i=1}^m a_{ij}}$$

식에서 R_j 표식은 정보원천 F_j 의 점수비중이고 n 은 정보원천수량이며 m 은 조사된 사용자수이다. a_{ij} 는 조사된 사용자 A_i 의 정보원천 F_j 에 대한 평가점수이다.

이 평가지표에는 해당 정보원천이 전체 정보원천들과의 관계속에서 얼마만한 비중으로 평가되었는가가 반영되어있다. 실례로 5개의 정보원천들가운데서 어떤 정보원천의 점수비중이 0.4라면 그 원천은 점수비중이 다른 원천들에 비하여 대단히 큰것으로 볼수 있다.

③ 최고점수빈도

어떤 정보원천에 대한 총적인 평가에서 최고점수를 준 사용자수가 전체 조사된 사용자수에서 차지하는 비율은 해당 정보원천의 최고점수빈도이다. 그 계산공식은 다음과 같다.

$$P_{\max} = \frac{m'}{m}$$

식에서 P_{\max} 는 최고점수빈도이고 m' 는 최고점수를 준 사용자수이며 m 은 조사된 전체 사용자수이다.

이 평가지표는 해당 정보원천이 제일 가치있다고 평가하는 사용자들의 비율을 나타내는 량으로서 이 지표가 큰 정보원천일수록 그것이 제일 가치있다고 평가하는 사용자들의 수가 제일 많다는것을 나타낸다.

④ 평균순위지표

평균순위를 계산할 때 우선 사용자들이 매개 정보원천을 평가한 점수들에 대하여 순위를 매기고 사용자의 각종 정보원천에 대한 점수순위를 렬거하여 각종 정보원천의 사용

자평가점수순위표를 작성한다.

그다음 매 사용자의 정보원천점수순위자료를 리용하여 개별적인 정보원천들의 사용자 전체에 대한 정보원천평균순위를 얻는다. 평균순위점수가 제일 낮은것은 정보원천의 가치가 제일 높은것이다.

평균순위를 계산하는 공식은 다음과 같다.

$$M_j = \frac{\sum_{i=1}^m c_{ij}}{m}$$

식에서 M_j 는 정보원천 F_j 의 평균순위이고 c_{ij} 는 조사된 사용자 A_i 의 정보원천 F_j 에 대한 점수순위이며 m 은 조사된 사용자수이다.

간접평가법의 4가지 지표는 단독적으로 리용할수도 있고 동시에 리용할수도 있다. 사용과정에 조사표자료입력의 정확성, 수값계산의 정확성과 자료전환의 정확성에 주의를 돌려 자료의 분석평가에서 과학성을 보장하고 최종적으로 정보원천의 가치를 객관적으로 판단해야 한다.

간접평가법을 리용하여 각이한 정보원천의 가치를 평가해보자.

5명의 정보사용자들이 초록(F_1), 속보(F_2), 정기간행물(F_3), 도서(F_4) 4가지 정보원천에 대하여 평가한 결과가 다음의 표와 같다고 하자.

정보원천평가점수표

정 보 리 용 자	정 보 원 천			
	초록(F_1)	속보(F_2)	정기간행물(F_3)	도서(F_4)
A_1	1	2	3	5
A_2	1	2	5	3
A_3	1	5	3	2
A_4	5	3	2	1
A_5	1	5	2	3
총계	9	17	15	14

먼저 점수평균값을 계산하자.

공식에 의하여 각 정보원천의 점수를 평균한 값은 다음과 같다.

$$R_1 = \frac{9}{5} = 1.8, R_2 = \frac{17}{5} = 3.4, R_3 = \frac{15}{5} = 3, R_4 = \frac{14}{5} = 2.8$$

점수평균값의 크기에 따라 각 정보원천의 중요성을 순서대로 배열하면 F_2, F_3, F_4, F_1 이다.

다음으로 점수비중을 계산하자.

모든 정보원천의 점수총합은 $9+17+15+14=55$ 이다. 공식에 의거하여 각 정보원천의 점수비중을 보면

$$P_1 = \frac{9}{55} \approx 0.16, P_2 = \frac{17}{55} \approx 0.31, P_3 = \frac{15}{55} \approx 0.27, P_4 = \frac{14}{55} \approx 0.26$$

점수비중의 크기에 따라 각 정보원천의 중요성을 순서대로 배열하면 F_2, F_3, F_4, F_1 이다.

다음으로 정보원천들의 최고점수빈도를 계산하자. 공식에 의하여 각 정보원천들의 최고점수빈도를 계산하면 다음과 같다.

$$P_{\max 1} = \frac{1}{5} = 0.2, P_{\max 2} = \frac{2}{5} = 0.4, P_{\max 3} = \frac{1}{5} = 0.2, P_{\max 4} = \frac{1}{5} = 0.2$$

최고점수빈도값에 따라 각 정보원천의 중요성을 순서대로 배열하면 $F_2, F_3=F_4=F_1$ 이다.

마지막으로 정보원천들의 평균순위를 계산하자. 우선 위에서 언급된 표를 다음과 같이 점수순위표로 전환한다.

공식에 의하여 각 정보원천의 평균순위를 계산하면 다음과 같다.

$$M_1 = \frac{4+4+4+1+4}{5} = 3.4, M_2 = \frac{3+3+1+2+1}{5} = 3.4$$

$$M_3 = \frac{2+1+2+3+3}{5} = 2.2, M_4 = \frac{1+2+3+4+2}{5} = 2.4$$

평균순위배열에 따라 각 정보원천의 중요성을 배열하면 F_3, F_4, F_1, F_2 이다.

평가점수순위표

정보원천	피조사리용자				
	A_1	A_2	A_3	A_4	A_5
초록(F_1)	4	4	4	1	4
속보(F_2)	3	3	1	2	1
정기간행물(F_3)	2	1	2	3	3
도서(F_4)	1	2	3	4	2

우에서 서술한 각이한 평가지표에 의하여 정보원천들의 가치를 판정할수 있다.

이 4가지 평가지표를 모두 리용하여 정보원천들을 평가하는데서 여러가지 문제점들이 제기된다.

이 4개의 평가지표를 모두 같은 무게를 주어 정보원천을 평가할것인가 아니면 서로 다른 무게를 줄것인가 하는데 따라 정보원천들에 대한 종합적인 평가가 달라진다.

만일 4개의 평가지표를 모두 같은 무게를 주어 정보원천을 평가한다면 위의 실례에서 5명의 조사된 사람들이 내린 평가결과는 속보(F_2)가 제일 중요하고 그다음은 정기간행물(F_3), 다음은 도서(F_4), 다음은 초록(F_1)이라고 볼수 있다.

만일 4개의 평가지표에서 정보원천의 평균순위지표에 보다 큰 무게를 준다면 속보(F_2)보다 정기간행물(F_3)이 더 가치있는것으로 평가결과가 달라질수 있다.

이 4가지 평가지표를 리용하는데서 기본은 매개 평가지표에 과학적인 무게를 주는것이다.

우리는 정보원천평가방법을 더욱 과학화하고 정보수집에서 나서는 과학리론적문제들을 원만히 풀어나감으로써 사회주의강국건설에 적극 이바지하여야 할것이다.