

가중치를 이용한 채용 정보 검색의 정확성 평가

김형래*, 정경희**, 전도홍**

*한국고용정보원 정보화기획팀

**관동대학교 전자계산공학과

e-mail : unikhee@gmail.com

Accuracy Evaluation of Weighted Recruiting Information Search Result

Hyung-Rae Kim*, Kyoung-Hee Jeong**, Do-Hong Jeon**

*Associate Research Fellow Korea Employment Information Service, Information Strategy Team

**Dept of Computer Science & Engineering, Kwandong University

요 약

취업 정보 시스템에서의 인재정보 및 채용정보의 검색결과에서 원하는 일자리와 인재를 빠른 시간 내에 검색하는 문제는 사용자 만족도에 밀접한 영향을 미치는 중요한 요소 중의 하나라고 할 것이다. 본 연구에서는 키워드를 이용하여 채용정보를 검색할 경우 채용정보 항목에 가중치 값을 적용했을 때의 정확도를 판단하기 위한 정량적 평가의 효과성을 검토한다. 채용정보 항목에 가중치 값을 적용했을 경우와 적용하지 않았을 경우의 채용검색 결과를 취업정보 시스템 관련 담당자를 대상으로 관련성 점수를 5점 척도로 평가한다. 실험결과, 채용정보 항목에 가중치를 적용하였을 때 관련 정확도 평가값이 더 높았으며, 상위 20개의 검색결과 평가값을 순위에 따라 정확성 추이를 살펴본 결과 순위가 높을수록 더 높은 정확성을 보였다.

키워드 : 검색가중치, 취업사이트, 가중치

Key Words : Search Weight, Job Site, Weight

1. 서론

웹 사이트에서 사용자가 원하는 정보를 얻고자 할 경우 방대한 정보의 양 때문에 얼마나 빠른 시간 내에 정보를 얻을 수 있겠는가 하는 문제와 더불어 어떤 검색어를 사용해야 하겠는가 하는 중요한 문제이다.

특히, 취업 정보 시스템에서의 인재정보 및 채용정보의 검색결과에서 원하는 일자리와 인재를 찾아내는 문제는 사용자 만족도에 밀접한 영향을 미치는 중요한 요소 중의 하나라고 할 것이다. 하지만, 사용자가 입력한 검색어에 대한 검색 결과 리스트 중에서 만족도와 관심도 파악은 매우 어려운 문제라고 할 수 있다.

최근에는 쇼핑몰, 포털사이트뿐만 아니라 취업정보 사이트의 검색 방식에 있어서도 검색엔진을 사용하여 키워드 정보 검색방법이 보편화되고 있다.

이에 따라 키워드를 통해 코드화된 입력 값과 코드화되지 않은 입력 값 모두 통합하여 검색하는 기능이 점차 강화되고 있다.

검색 값은 코드화된 검색 값과 코드화되지 않은 검색 값으로 크게 구분 지을 수 있다. 코드화된 검색 값은 항목의 입력 값의 범위가 미리 정하여져 있어 사용자가 선택할 수 있거나 수치 값처럼 크기를 비교할 수 있는 경우를 말하며, 코드화되지 않은 검색 값은 회사정보나 전형방법 등

과 같이 사용자로부터 임의의 값을 입력하는 경우를 말한다. 직종, 학력, 경력, 임금 등의 항목은 코드화 할 수 있는 검색 항목으로써, 검색 항목 별로 코드가 정해진 경우는 정확한 코드가 조건으로 입력된다면 원하는 채용정보를 정확하게 추출할 수 있다. 하지만, 직업이 다양해지고 세분화됨에 따라 특정 톨을 사용할 수 있는 인재를 구하거나, 뛰어난 하나의 기술을 가지고 취업하려는 구직자가 늘어나는 등 취업정보 검색 요구 기능 또한 다양해지고 있다. 따라서 회사정보나 전형방법과 같이 다양한 입력항목 값을 모두 코드화하기 어렵고, 숙련된 사용자에게는 코드화된 검색항목 값을 입력하기 위한 반복된 클릭 행위는 불편을 초래할 수 있다. 키워드 검색방법은 각 항목마다 조건을 선택할 때 발생하는 마우스 클릭 횟수를 줄여주어 사용자들의 정보 검색을 보다 편리하게 지원한다. 하지만 키워드 검색 시 어떤 항목에 더 많은 가중치를 주어 검색 결과의 정확성을 높일 수 있겠는가 하는 문제가 남아있다.

본 논문에서는 키워드 검색시 채용정보 검색의 정확성을 높일 수 있도록 다양한 채용정보 항목에 각기 다른 가중치 값을 적용하였으며, 가중치 적용 값의 효과성과 적절성을 판단하고자 한다.

본 논문의 기여도를 정리하면,

1. 채용정보의 항목에 부여된 가중치 값의 효과성을 검토 하였으며,
 2. 항목 가중치 값이 검색결과 정확성에 미치는 영향을 정량적 평가방법으로 평가하였으며,
 3. 채용정보 항목에 가중치 값을 적용하였을 때의 결과 정확성 평가값이 더 높게 나타났음을 발견하였다.
- 본 논문은 2장에서는 관련연구 기술, 3장에서는 연구방법 및 설문조사에 대하여 설명하였으며, 4장에서는 결과분석, 5장에서는 결론으로 이어진다.

2. 관련연구

김민영(2002)은 검색엔진은 색인과 시소러스의 사용이 취약하며, 검색과 관련한 전문적 지식이 없는 일반인을 대상으로 서비스 되므로 시각과 기술을 이용해 인지적 인터페이스를 구현해야 한다고 하였다. 이에 따라 검색시스템에서 이용자 인터페이스의 효율성을 높이기 위해 사용된 정보각화 기술에 대하여 기술하였으며, 정보검색을 지원할 수 있는 정보시각화 인터페이스 평가요소를 도출하였다[1].

강운이(2000)는 기존 연구에서 개발된 정보 검색시스템의 평가항목들을 바탕으로 정보검색 엔진의 기능성 평가항목을 선정하였다. 설문조사 결과에서 나온 검색엔진 중 상위 5개의 검색엔진을 선정하여 질문지법을 이용한 면접 조사를 하였다. 사용자들이 편리한 기능이나 중요하다고 생각되는 항목을 선택하여 정확률 표시에 대한 항목과 다른 검색엔진의 결과 제시 등의 사용자들의 요구기능을 반영하여 기능성 평가항목을 선정하고 평가하였다. 평가항목 개발과 개선이 요구된 평가방법을 다루었으나 성능적 평가 방법과 평가항목들에 대한 가중치 기준은 향후 문제로 남겨놓았다[2].

김병철 외 2(1999)는 어떤 키워드를 사용해야 할지 고민하는 사용자들의 문제점과 더불어 메타 검색엔진의 낮은 속도 문제점을 해결하고자 하였다. 일반적으로 검색 엔진을 통해 얻는 정보는 단순한 기업 홈페이지이거나 홈페이지 내의 주 정보 정도이며, and, or를 통한 특정 정보를 찾고자하므로 유명 검색엔진들에서는 몇 건 정도를 찾아 내는지 알아보고, 단순 검색의 경우는 홈페이지 자체 DB를 통해 찾아 갈 수 있는 시스템을 설계 구현하고자 하였다[3].

이종욱 외 1(1998)은 현재 세계적으로 많은 검색 엔진들이 인터넷 정보 검색에 사용되어 있으나 정보획득에 소요되는 시간과 비용 및 노력을 줄이기 위해 선택되어질 검색 엔진의 기준을 문제점으로 다루었다. 이에 따라 검색엔진 분류 특성에 따른 검색엔진에 대하여 검색된 문헌의 양, 적합 문헌의 양, 재현율, 정확율, 중복율 등 검색엔진 효율성을 상호 비교, 평가해 봄으로써 이용자들이 인터넷을 통한 정보 검색 시 적절한 검색엔진을 선택하게 하는 기준을 제공하여 주었다[4].

3. 실험개요 및 방법

항목 가중치는 사용자가 검색 결과 후 확인하는 채용정보의 리스트에 나타나는 항목들 중 직종명> 사업체명> 지역> 학력의 순으로 주었으며, 그 외 항목들은 가장 낮은 가중치 값을 적용하였다.

실험 방법은 취업 정보 시스템에서 채용정보 검색 시 사용하는 검색 키워드를 사용자가 임의로 선정하여 검색 결과의 상위 20개를 5점 척도(관련없음 ← 보통 → 관련 높음)로 정확도를 평가하였다.

실험 참여자는 1단어와 2단어로 된 키워드를 임의 선택하였으며, 선정된 각 키워드는 채용정보 항목에 가중치 값을 적용하지 않았을 때의 정확도 값과 가중치 값을 적용했을 때의 정확도 값을 평가한다. 임의로 입력한 키워드는 검색된 결과 리스트 중 제일 먼저 검색된 리스트가 사용자가 원하는 검색 결과와 관련이 높으면 5점, 보통이면 3점, 관련이 없으면 1점으로 체크되어 검색 결과 20개의 평균값을 비교 분석하였다.

참가인원은 모두 17명으로 취업 정보 시스템의 관련 업무를 담당자로 구성하였으며, 설문 방법 및 방식의 일관성을 유지하기 위하여 검색엔진 엔지니어 및 감독자 4명의 감독하여 약 2시간동안 실험이 이루어졌다.

실험자에 의해 평가된 키워드에 대한 평가 평균값은 비가중치 평균값과 가중치 평균값의 차이가 T-검정(쌍체 비교)결과 통계적으로 유의(p값이 0.05보다 작아야 95% 내에서 유의함을 시사)하게 나타나야 한다.

또한, 검색결과는 상위 20위 내에서 순위가 높을수록 높은 정확성을 갖는 것이 좋으며[5], 순위에 따른 정확성 추이를 분석하기 위해서 상위 1위부터 20위까지의 정확도 평가 평균값을 계산하여 지수 평활법(exponential smoothing)으로 평가된 상위 1위부터 20위까지의 추이를 살펴본다.

지수 평활법은 가장 최근 데이터에 가장 큰 가중치가 주어지고 시간이 지남에 따라 가중치가 기하학적으로 감소되는 가중치 이동 평균 예측 기법의 하나로써, 데이터들이 시간의 지수 함수에 따라 가중치를 가지므로 지수 평활법이라고 한다. 이 기법은 가장 최근의 예측 데이터와 주요 판매 데이터 간의 차이에 적합한 평활 상수를 사용함으로써 과거의 데이터를 유지할 필요성을 갖지 않는다. 이러한 접근 방법은 어떤 추세를 갖지 않거나 계절적인 패턴을 나타내는 데이터 또는 추세와 계절성을 모두 갖는 데이터에 사용될 수 있다. 지수 평활법에서는 평활상수(α)를 이용하여 상위에서 하위로 갈수록 더 적은 비중을 주는 방법을 채택하며 기본식은 식 1과 같다.

$$F_{t-1} = \alpha Y_t + (1-\alpha)F_t \quad \dots\dots\dots(\text{식 } 1)$$

(F_{t+1} : 기간 $t+1$ 에서의 예측값,

α : 평활상수($0 \leq \alpha \leq 1$),

Y_t : 기간 t 에서의 실측치, F_t : 기간 t 에서의 예측치)

평활상수 α 값은 값의 특성과 관리자가 좋은 반응율을 나타낸다고 생각하는 값에 의해 결정되며, 여러 개의 α 값에 대해 예측치를 구한다음, 그 중에서 예측 오차를 최소화하는 α 값을 선택한다. 본 논문에서는 추이를 가장 명확하게 보여주는 0.8을 선택하였다.

4. 결과 분석

채용정보 검색결과와 정확성 평가 평균값은 표 1과 같다. 1단어를 키워드로 검색하였을 경우 비가중치일 때 검색결과 관련성 평균값은 3.83점이었고, 가중치 일 때에는 4.26점으로 향상된 것을 알 수 있다.

또, 2단어일 경우는 비가중치 관련성 평균값은 4.26점이며, 가중치일 때 평균값은 4.39점으로 0.13점이 향상되었다.

<표 1> 5점 척도로 평가한 정확성 결과값

구분		사용 키워드	건수	평균(점)
채용정보	비가중치	1단어	9	3.83
		2단어	14	4.26
	가중치	1단어	9	4.26
		2단어	14	4.39

5점 척도로 받은 평균값은 식 2와 같은 방법으로 백점 환산값으로 변환할 수 있으며, 환산한 정확도 평균값은 표 3과 같다.

$$y = (x-1) * 25$$

(단, x=설문조사평균값, y=백점환산값)(식 2)

표 2와 같이 1단어일 경우 채용정보 항목에 가중치 값을 적용하였을 때 검색결과 정확성 향상율은 15.2%이다. 2단어일 경우는 4.0% 향상되었고, 전체를 살펴보면 9.2% 향상되었음을 알 수 있다. 이와 같은 결과는 채용정보 항목에 각각 다른 가중치 값을 적용했을 때 검색결과와 정확도가 더 높음을 알 수 있다.

<표 2>백점으로 환산한 값의 결과값

구분	비가중치(점)	가중치(점)	향상율
1단어	70.75	81.50	15.2%
2단어	81.50	84.75	4.0%
전체	76.13	83.13	9.2%

표 3에서 표 5와 같이 키워드로 입력한 검색결과 값의 비가중치 평균값과 가중치 평균값의 차이가 T-검정결과(쌍체 비교) P=0.0022 t=2.0595로 통계적으로 p값이 0.05보다 작아 95%내에서 유의함이 나타났다.

<표 3> 1단어 t-검정 (쌍체비교) 결과

전체	가중치	비가중치
평균	4.2611	3.8333

분산	1.9624	1.6919
관측수	9	9
피어슨상관계수	0.9357	
가설 평균차	0	
자유도	8	
t통계량	2.5957	
P(T<=t) 단측검정	0.0159	
t 기각치 단측검정	1.8595	
P(T<=t)양측검정	0.0318	
t 기각치 양측검정	2.3060	

<표 4> 2단어 t-검정 (쌍체비교) 결과

전체	가중치	비가중치
평균	4.3929	4.2607
분산	0.9138	0.9239
관측수	14	14
피어슨상관계수	0.9737	
가설 평균차	0	
자유도	13	
t통계량	2.2495	
P(T<=t) 단측검정	0.0212	
t 기각치 단측검정	1.7709	
P(T<=t)양측검정	0.0424	
t 기각치 양측검정	2.1604	

<표 5> 전체 t-검정 (쌍체비교) 결과

전체	가중치	비가중치
평균	4.3413	4.0935
분산	1.2579	1.2067
관측수	23	23
피어슨상관계수	0.9437	
가설 평균차	0	
자유도	22	
t통계량	3.1856	
P(T<=t) 단측검정	0.0021	
t 기각치 단측검정	1.7171	
P(T<=t)양측검정	0.0043	
t 기각치 양측검정	2.0739	

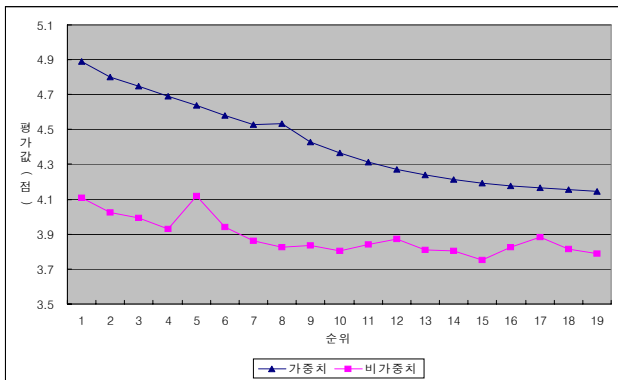
표 6과 같이 키워드 검색결과와 정확도 평가 평균값을 1위와 20위를 비교하여 살펴보면, 가중치 적용결과 값과 비가중치 적용 살펴보면 1단어일 경우 가중치 검색결과 감소율은 평균 12%로 비가중치 검색결과 평가값보다 7% 더 뚜렷한 감소율을 나타냈다.

<표 6> 검색결과 1위와 20위 감소율 비교

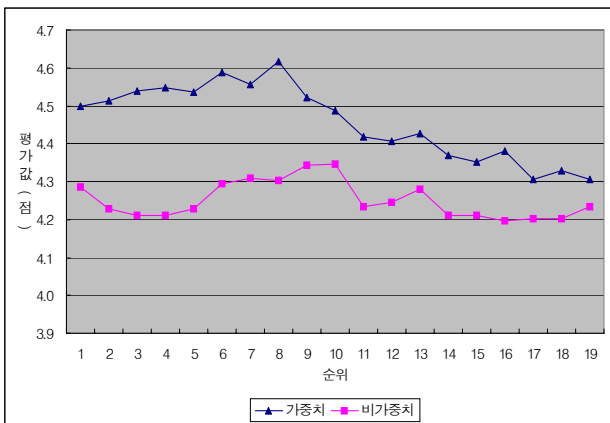
구분	검색1위	검색20위	감소율	
1단어	가중치	4.9	4.1	20%
	비가중치	4.1	3.8	8%
2단어	가중치	4.5	4.3	5%
	비가중치	4.3	4.2	2%
전체	가중치	4.7	4.2	12%
	비가중치	4.2	4.0	5%

검색 결과의 평가값을 상위 1위부터 20위까지의 추이를 지수 평활법으로 나타낸 결과는 그림 1에서 그림 3과 같다.

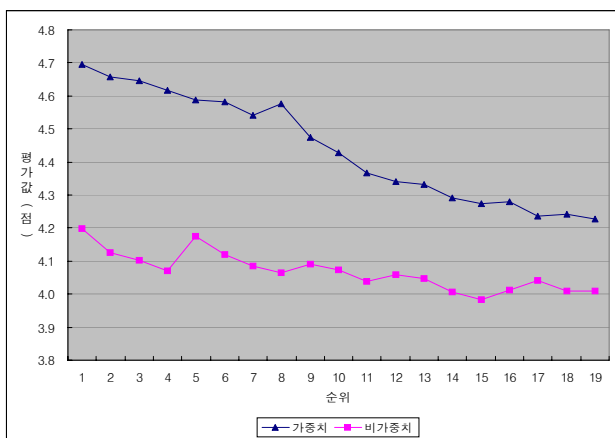
그림에서와 같이 채용정보 항목에 가중치를 적용했을 때 정확성 평가값이 모두 높게 나왔으며, 가중치 값을 적용했을 때 정확도 평가값은 상위 1위가 가장 높았고, 상위 20위까지 1단어, 2단어 모두 하향선을 나타내고 있는 것을 알 수 있다.



(그림 1) 1단어 가중치, 비가중치 정확도 평가 결과



(그림 2) 2단어 가중치, 비가중치 정확도 평가 결과



(그림 3) 전체 가중치, 비가중치 정확도 평가 결과

5. 결론

본 연구에서는 취업 정보 시스템에서 채용정보 검색시 사용자가 입력한 검색어에 대해 항목 가중치 값을 적용하였을 때의 적절성 및 효과성을 도출하고자 하였다.

실험방법은 취업 정보 시스템에서 채용정보 검색 시 사용하는 검색 키워드를 사용자가 임의로 선정하여 상위 20개의 검색결과를 5점 척도(관련없음 ← 보통 → 관련높음)로 평가하였다. 실험참여자는 취업 정보 시스템의 관련 업무를 담당자로 구성하였으며, 1단어와 2단어로 된 키워드를 임의의 선택하여, 채용 정보 항목에 가중치를 적용하였을 때와 가중치 값을 적용하지 않았을 때 나타나는 검색 결과 값의 관련성을 평가하였다.

실험결과, 검색을 통하여 나타난 상위 20개의 가중치 적용 평균값과 비가중치 적용 결과는 T-검정(쌍체비교) 값이 $p=0.0022$ $t= 2.0595$ 로 유의하게 나타났다. 또한 채용정보 항목에 가중치를 적용하지 않았을 때의 검색결과 정확성 평가값은 76.13점이고 가중치 적용 시 정확성 평가 결과 값은 83.13점으로 향상되었다. 마지막으로 상위 20개의 검색결과 평가값을 순위에 따라 정확성 추이 조사 시 순위가 높을수록 높은 정확성을 보였다.

향후 본 논문에서 배제하였던 임금, 지역, 학력 등 검색 조건이 연계된 정확성 제고 방안과 인재정보의 검색 평가 방법에 대한 연구가 필요할 것이다.

참고문헌

- [1] 김민영, “웹 검색 엔진의 정보시각화 인터페이스 평가 요소에 관한 연구”, 중앙대학교 대학원, 2002
- [2] 강운이, “정보검색엔진의 기능성 평가항목의 선정 및 적용”, 전북대학교 교육대학원, 2000
- [3] 김병철, 이명선, 전병태, “검색엔진을 이용한 검색 서비스 설계 및 구현”, Journal of computer science & engineering technology, vol 2, 1999
- [4] 이종욱, 김종택, “인터넷 검색엔진의 효율성 평가에 관한 연구”, 동국대학교 경영대학원 경영논총, vol 22, 1998
- [5] Learning Implicit User Interest Hierarchy for Context in Personalization, Kim and Chan, Applied Intelligence, Springer, 2007.
- [6] 변상석, 이주복, 최진식 공저, “엑셀을 이용한 통계학”, 정익사