

사물인터넷에서 참여 기기를 고려한 개인화 정보 검색 기법*

권 준 희**

Personalized Information Retrieval Method considering Participating Device in Internet of Things

Kwon, Joon Hee

〈Abstract〉

Internet of Things is growing rapidly. As it evolves, the amount of data is increasing significantly. It requires a new personalized information retrieval method. Internet of Things is defined as uniquely identifiable interoperable connected object. The first definition of Internet of Things was from Things oriented perspective. However, previous studies about personalized information retrieval method do not consider Things. To meet user's individual needs, previous studies concentrate on only human, not Things. In this paper, we propose a personalized information retrieval method considering participating device in Internet of Things. It provides personalized information using data type preference for each device. Moreover, it provides personalized results by integrating data type preference for set of devices. This paper describes a new personalized retrieval method and algorithm. It consists of five steps. Then, it presents four scenarios using proposed method. The scenarios show our work is more effective and efficient than existing one.

Key Words : Personalization, Information Retrieval, Internet of Things, Device

I. 서론

새롭게 도래하고 있는 4차 산업혁명은 사물인터넷(IoT : Internet of Things) 기술의 사물과 사물, 사람과 사람, 사람과 사물이 서로 연결되는 초연결성이 기반이 된다. 이러한 초연결성은 사회·산업·경제

분야의 생태계 전반을 흔들게 될 것으로 예측되고 있다[1]. 현재 인터넷에 연결되어 있는 기기는 소수에 불과하지만, 사물인터넷 환경이 발전하면서 갈수록 인터넷에 연결되는 사물 기기의 수는 기하급수적으로 증가할 것이다[2].

사물 인터넷의 발전으로 정보의 양은 크게 증가하고 있다. 이에 따라 이러한 정보를 보다 효과적으로 찾아내고자 하는 필요성이 더욱 커지면서 정보 검색 연구에 대한 관심이 증가하고 있다[3]. 정보 검색

* 이 논문은 2017학년도 경기대학교 연구년 수혜로 연구되었음.

** 경기대학교 컴퓨터공학부 교수

(information retrieval)은 대규모의 정보들로부터 요구에 맞는 자료를 찾는 기술을 의미한다[4]. 정보 검색에 관한 많은 연구 중 특히, 수많은 정보로부터 특정 사용자에게 적합한 중요 정보를 찾아내고자 하는 개인화 정보 검색(personalized information retrieval) 기법에 대한 필요성은 매우 크다[5].

개인화 정보 검색 기법은 각 사용자의 선호도와 같은 개인의 특성을 반영하는 요소를 기반으로 개인에 적합한 정보를 제공한다. 그러나, 기존의 전통적인 개인화 정보 검색 기법에서는 사물인터넷 환경을 고려하지 않는다는 문제점을 가진다. 즉, 기존의 연구들은 사람에만 집중하여, 사물 인터넷에서 중요한 요소인 사물을 고려하지 않는다. 사물 인터넷 환경에서는 기존의 환경과는 다르게 정보 검색에 참여하는 다양한 여러 개의 기기가 있다는 특성을 가진다. 이 때 참여 기기는 각 기기 특성에 따라 사용자별 선호 정보가 서로 다르기 때문에 이를 반영한 새로운 개인화 정보 검색 기법이 요구된다.

본 논문에서는 사물인터넷에서 각 사용자의 기기별 데이터 타입 선호도를 기반으로 한 새로운 정보 검색 기법을 제안한다. 제안 기법에서는, 검색에 참여하는 각 기기에 대한 각 사용자의 데이터 타입 선호도에 따라 제공되는 정보가 개인화된다. 또한, 여러 개의 서로 다른 기기들이 검색에 함께 참여하는 경우, 여러 기기들의 데이터 타입 선호도를 융합함으로써 기기 집합의 특성을 반영하는 정보를 검색한다.

본 논문의 구성은 다음과 같다. 2장에서는 제안 기법과 관련된 연구들을 살펴본다. 3장에서는 사물인터넷에서 참여 기기를 고려한 개인화 정보 검색 기법의 처리 과정을 살펴보고, 4장에서는 이를 시나리오에 적용하여 설명한다. 마지막으로 5장에서 결론을 맺는다.

II. 관련연구

사물인터넷이라는 용어는 인터넷과 사물이라는 2개의 중요한 요소로 구성된다[6]. 사물인터넷 시대에서는 모든 사물이 인터넷으로 서로 연결된다. 인터넷에 연결된 사물의 개수는 사람의 수보다 더 많을 수 있으며, 미래 인터넷 시대에서 사물은 매우 중요한 역할을 하게 된다[7]. 사물인터넷은 현재 가장 유망한 기술 중 하나이다. 사물인터넷이 발전하면서 이전과는 비교할 수 없을 정도로 많은 정보들을 다루게 된다[8]. 이에 따라 정보 검색 기법에 대한 필요성은 더욱 커지고 있다.

정보 검색 중 개인화 정보 검색 기법에 대한 연구는 최근까지 활발하게 진행되고 있다. 개인화 정보 검색 기법은 동일한 질의어에 대해서도 사용자의 특성에 따라 서로 다른 검색 결과를 제공하기 때문에 사용자의 요구를 만족시켜 검색 품질이 향상된다는 장점을 가진다[5, 9]. 개인화 정보 검색에 대한 그동안의 연구는 다음과 같다.

사용자 모델링(user modeling) 정보를 이용하여 개인화된 정보 검색에 활용하는 연구들이 있다[10-12]. 사용자 모델은 장기 모델링과 단기 모델링으로 나뉜다. 장기 모델링은 사용자의 변하지 않는 일반적인 관심 사항과 선호도를 의미하는 것으로 사용자 프로파일 정보를 들 수 있다. 단기 모델링은 검색시의 사용자의 행동과 현재 세션에서의 상황들로부터 이끌어낸 사용자 정보를 의미한다.

질의어 확장(query expansion)에 의한 개인화 정보 검색에 대한 연구들이 있다[10, 13, 14]. 질의어 확장은 사용자의 검색 이력 정보 등을 이용하여 사용자의 관심 사항을 바탕으로, 입력된 질의어에 부가적인 질의어를 추가하는 방법 등을 이용하여 사용자가 보다 적합한 검색 결과를 얻게 하는 방법이다.

검색 결과 재정렬 기법(result re-ranking)을 이용한 연구들도 많이 이루어져왔다[10, 15, 16]. 검색 결과

재정렬 기법은 사용자와의 관련성 정도에 따라 검색 결과를 재정렬하여 제공하는 방법이다. 해당 기법에는 사용자 반응이나 사용자의 선호도를 활용하는 다양한 연구들이 진행되어 왔다.

그 외에도 시맨틱 정보를 이용한 검색 기법도 있다. 이러한 연구로는 온톨로지를 이용한 연구 등을 들 수 있다[17]. 최근에는 기계 학습(machine learning)을 이용한 연구도 활발히 진행되고 있다. 이러한 연구로는 딥러닝(deep learning) 기법을 이용하여 사용자 선호도와 특성을 추출하는 연구 등을 들 수 있다[9].

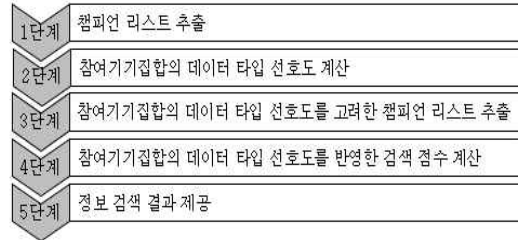
전통적인 환경과는 다르게 사물인터넷 환경에서는 사용자 이외에도 사물이 매우 중요한 요소이다. 그러나, 기존의 개인화 정보 검색에 대한 연구들은 사용자를 중심으로 한 연구에 집중되어 있다. 따라서, 사물인터넷 환경에서 중요한 사물의 특성을 고려한 개인화된 검색 기법에 대한 새로운 연구가 필요하다.

III. 참여 기기를 고려한 개인화 정보 검색 기법

본 장에서는 사물인터넷 환경에서 참여 기기를 고려한 개인화된 top-k 정보 검색 기법을 설명한다. 본 논문에서는 기존의 개인화 정보 검색 기법과는 다르게 각 사용자의 기기별 데이터 타입 선호도를 기반으로 한 정보 검색 기법을 제안한다. 제안하는 검색 기법은 5단계로 구성되며, <그림 1>과 같다.

1단계에서는 챔피언 리스트를 추출한다. 본 논문에서는 전통적으로 top-k 정보 검색에 효율적인 기법으로 알려져 있는 챔피언 리스트(champion list)를 사용한다[4]. 후보 정보 집합에서, 검색 점수에 영향을 주는 항목 중 챔피언 리스트의 요소 정보를 결정하는 기준 항목값이 임계치값 a 이상인 정보만을 추출하여 챔피언 리스트를 구성한다. 본 논문에서는 해당 기준

항목값을 편의상 w 라고 부르기로 한다.



<그림 1> 프로세스 흐름도

2단계에서는 참여 기기 집합의 데이터 타입 선호도를 계산한다. 1단계에서 추출된 챔피언 리스트는 참여 기기 집합의 특성을 반영하지 않는다는 문제점을 가진다. 이러한 문제점을 해결하기 위해 본 논문에서는, 정보 검색에 참여하는 참여 기기 집합 D 에서 정보의 데이터 타입 t 에 대한 사용자 u 의 선호도 $p(D,t,u)$ 를 제안한다. 선호도 $p(D,t,u)$ 는 식 (1)로 구성된다. 여기서, $preference(d,t,u)$ 는 정보 검색에 참여하는 기기 집합 내 기기 d 에서 정보의 데이터 타입 t 에 대한 사용자 u 의 선호도이다. 선호도 $preference(d,t,u)$ 는 사용자 u 의 입력에 의해 주어질 수도 있고, 기기 d 에 대한 사용자 u 의 로깅 정보 등을 기반으로 계산될 수도 있다. 식(1)에서, N 은 정보 검색에 참여하는 기기의 전체 개수이며, D_i 는 기기 집합 D 에 속한 i 번째 기기를 의미한다.

$$p(D,t,u) = \sum_{i=1}^N preference(D_i,t,u) / N \quad (1)$$

3단계에서는 참여 기기 집합의 데이터 타입 선호도를 고려한 챔피언 리스트를 추출한다. 1단계를 통해 추출된 챔피언 리스트의 정보 요소 중 선호도 p 를 고려했을 때 중요하지 않은 정보는 챔피언 리스트의 요소에서 제외한다. 첫째, 챔피언 리스트에 포함된 각

정보에 대해, 참여 기기 집합 D 에서 해당 정보의 데이터 타입 t 에 대한 사용자 u 의 선호도 $p(D,t,u)$ 값이 0인 정보는 챔피언 리스트의 요소에서 제외한다. 둘째, 챔피언 리스트에 포함된 각 정보의 기준 항목값 w 가 임계치값 δ 를 넘지 않는 정보는 챔피언 리스트의 요소에서 제외한다. 선호도 p 를 고려한 챔피언 리스트 요소 정보를 결정하는 기준 항목값 pw 는 다음 식 (2)에 의해 계산된다.

$$pw = w \times p(D,t,u) \quad (2)$$

기준 항목값 pw 가 임계치값 β 를 넘지 않는 정보는 중요하지 않은 정보로 판단하여, 1단계로부터 추출된 챔피언 리스트의 요소에서 제외한다. 이 때, 챔피언 리스트의 각 요소별로 계산한 pw 값을 통해 w 에 대한 임계치값을 계산하는 대신, 식 (3)를 이용하여 임계치값을 한번만 계산함으로써 보다 효율적인 처리가 가능하다.

$$\delta = \beta / p(D,t,u) \quad (3)$$

4단계는 참여 기기 집합의 데이터 타입 선호도를 반영한 검색 점수를 계산한다. 4단계에서는 3단계를 통해 추출된 챔피언 리스트의 각 요소 정보들에 대해서만 검색 점수를 계산한다. 식 (4)는 참여 기기 집합의 데이터 타입 선호도를 반영한 검색 점수 $pscore$ 에 대한 계산식을 보인다. 여기서, $score$ 는 챔피언 리스트의 각 요소 정보들에 대한, 참여 기기 집합의 데이터 타입 선호도를 반영하지 않은 기존의 검색 점수이다.

$$pscore = score \times p(D,t,u) \quad (4)$$

5단계는 4단계를 통해 얻어진 검색 점수를 사용하여 상위 k 개의 정보를 최종 사용자에게 제공한다.

<표 1>은 참여 기기 집합의 데이터 타입 선호도를

고려한 정보 검색 알고리즘을 보인다. 여기서 $w(c)$ 와 $score(c)$ 는 각각 챔피언 리스트의 정보 요소에서 기준 항목값 w 와 검색 점수 $score$ 를 얻어내는 함수이다.

<표 1> 검색 알고리즘

```

1. Algorithm.
2. Begin
3. Input
4.   CandidateSet : 후보 정보 집합
5.   a : 챔피언 리스트 임계치 값
6.   β : 선호도를 고려한 챔피언 리스트 임계치 값
7.   D : 검색에 참여하는 기기 집합
8.   T : 데이터 타입 집합
9.   preferenceSet(D,T,u) :
       { preference(d,t,u) | d ∈ D, t ∈ T }
10. Output
11.   ResultSet : 검색 결과 집합
12. Method
13.   championList = getChampion(CandidateSet,a)
14.   for ( t ∈ T )
15.     compute p(D,t,u)
16.   for ( c ∈ championList )
17.     if ( w(c) < β / p(D,datatype(c),u) )
18.       discard c in championList
19.   for ( c ∈ championList )
20.     ResultSet = ResultSet ∪
       { c, score(c) * p(D,datatype(c),u) }
21.   return ResultSet
22. End.
    
```

1단계와 3단계에서는 기존 연구인 챔피언 리스트를 적용하였다. 해당 연구에서 임계치값 a 와 β 는 검색 속도와 결과에 영향을 미칠 수 있기 때문에 주의하여 값을 선정하는 것이 필요하다. 즉, 임계치값이 커지면 후보 정보 개수가 작아지기 때문에 검색 시간이 줄어드는 장점이 있지만, 작은 개수의 후보 정보만을 추출함으로써 k 개의 결과 정보를 제공하지 못할

수 있다. 그러나, 임계치값이 작아지는 경우에는 k 개의 결과 정보는 모두 제공할 수 있지만, 검색 시간이 오래 소요된다[4].

IV. 실험 및 결과

본 장에서는 시나리오를 통해 제안 기법을 실험하고 결과를 설명한다. 정보를 검색하려고 하는 '나'는 제안 기법을 이용하여 상위 7개의 검색 결과를 요청한다. <표 2>는 후보 정보 전체 집합을 보인다. 여기서 nw 는 검색 점수를 구성하는 항목 중 w 를 제외한 항목값을 의미한다.

<표 2> 후보 정보 전체 집합

정보	데이터 타입	w	nw
c01	이미지	0.45	0.62
c02	음성	0.02	0.17
c03	동영상	0.17	0.8
c04	동영상	0.65	0.34
c05	음성	0.27	0.14
c06	음성	0.14	0.39
c07	음성	0.35	0.61
c08	음성	0.58	0.3
c09	이미지	0.39	0.84
c10	동영상	0.33	0.47
c11	텍스트	0.04	0.06
c12	이미지	0.37	0.23
c13	이미지	0.19	0.6
c14	이미지	0.72	0.8
c15	동영상	0.43	0.17
c16	텍스트	0.41	0.72
c17	텍스트	0.85	0.26
c18	텍스트	0.63	0.88
c19	텍스트	0.29	0.38
c20	동영상	0.25	0.01

단계로 나누어 살펴보면 다음과 같다. 1단계는 챔피언 리스트를 추출한다. 본 시나리오에서는 검색 점수에 영향을 주는 항목값 w 에 대한 임계치값 α 를 0.2로 가정한다. 이에 따른 결과 챔피언 리스트는 <표 3>과 같다. <표 2>에서 나타난 20개의 후보 정보 중 15개의 후보 정보가 추출된다.

<표 3> 챔피언 리스트

정보	데이터 타입	w	nw
c17	텍스트	0.85	0.26
c14	이미지	0.72	0.8
c04	동영상	0.65	0.34
c18	텍스트	0.63	0.88
c08	음성	0.58	0.3
c01	이미지	0.45	0.62
c15	동영상	0.43	0.17
c16	텍스트	0.41	0.72
c09	이미지	0.39	0.84
c12	이미지	0.37	0.23
c07	음성	0.35	0.61
c10	동영상	0.33	0.47
c19	텍스트	0.29	0.38
c05	음성	0.27	0.14
c20	동영상	0.25	0.01

<표 4> 나의 기기별 데이터타입 선호도

기기	데이터 타입	선호도
AI스피커	음성	1
AI스피커	텍스트	0.7
AI스피커	이미지, 동영상	0
스마트TV	동영상	1
스마트TV	텍스트, 음성, 이미지	0.1
스마트폰	텍스트, 음성, 이미지, 동영상	1

이제 4개의 서로 다른 참여 기기 집합에 따른 처리 과정을 각각 살펴본다. 본 시나리오에서는 사물 인터넷 제안 기법의 단계별 프로세스 흐름도에 따라 총 5

넷 기기로 스마트폰, AI스피커, 스마트TV를 고려하는 것으로 한다. 사용자인 '나'의 기기별 데이터 타입 선호도 $preference$ 는 <표 4>와 같다. <표 4>에서, AI스피커는 텍스트 정보가 주어지는 경우 해당 텍스트를 음성으로 변환하여 제공한다고 가정한다. 또한, 참여 기기 집합의 데이터 타입 선호도를 고려한 챔피언 리스트를 추출하기 위한 임계치값 β 는 0.2 이며, 검색 점수 $score$ 는 $0.5xw + 0.5xnw$ 로 계산된다고 한다.

<표 5> 데이터 타입 선호도를 고려한 챔피언 리스트와 검색 점수 (스마트폰)

정보	데이터 타입	w	nw	pScore
c17	텍스트	0.85	0.26	0.555
c14	이미지	0.72	0.8	0.76
c04	동영상	0.65	0.34	0.495
c18	텍스트	0.63	0.88	0.755
c08	음성	0.58	0.3	0.44
c01	이미지	0.45	0.62	0.535
c15	동영상	0.43	0.17	0.3
c16	텍스트	0.41	0.72	0.565
c09	이미지	0.39	0.84	0.615
c12	이미지	0.37	0.23	0.3
c07	음성	0.35	0.61	0.48
c10	동영상	0.33	0.47	0.4
c19	텍스트	0.29	0.38	0.335
c05	음성	0.27	0.14	0.205
c20	동영상	0.25	0.01	0.13

첫째, 참여 기기가 스마트폰인 경우를 살펴보면 다음과 같다. 2단계로 참여 기기 집합의 데이터 타입 선호도를 계산한다. 식 (1)에 의해 스마트폰의 텍스트, 음성, 이미지, 동영상 데이터 타입에 대한 나의 선호도값 p 는 모두 1 이다. 3단계로 스마트폰의 데이터 타입 선호도를 고려한 챔피언 리스트를 추출한다. 식(3)에 의해 모든 데이터 타입에 대한 δ 값은 0.2로, <표 3>에서 w 값이 0.2 미만인 정보는 제외한 챔피언 리

스트는 <표 3>과 동일하다. 4단계는 스마트폰의 데이터 타입 선호도를 반영한 검색 점수를 계산하며, <표 5>와 같다. p 값이 1이므로 $score$ 값의 결과와 $pScore$ 값의 결과는 동일하다. 마지막으로 상위 7개의 결과를 검색 결과로 제공하면 <표 6>과 같다.

<표 6> 최종 검색 결과 (스마트폰)

정보	데이터 타입	pScore
c14	이미지	0.76
c18	텍스트	0.755
c09	이미지	0.615
c16	텍스트	0.565
c17	텍스트	0.555
c01	이미지	0.535
c04	동영상	0.495

<표 7> 데이터 타입 선호도를 고려한 챔피언 리스트와 검색 점수 (스마트폰, AI스피커)

정보	데이터 타입	w	nw	pScore
c17	텍스트	0.85	0.26	0.472
c14	이미지	0.72	0.8	0.38
c04	동영상	0.65	0.34	0.248
c18	텍스트	0.63	0.88	0.642
c08	음성	0.58	0.3	0.44
c01	이미지	0.45	0.62	0.268
c15	동영상	0.43	0.17	0.15
c16	텍스트	0.41	0.72	0.48
c07	음성	0.35	0.61	0.48
c19	텍스트	0.29	0.38	0.285
c05	음성	0.27	0.14	0.205

둘째, 참여 기기 집합이 스마트폰과 AI스피커일 때의 검색 과정은 다음과 같다. 2단계에서는 참여 기기 집합의 데이터 타입 선호도를 계산하며, 이에 대한 결과는 텍스트, 동영상, 음성, 이미지에 대해 각각 0.85, 0.5, 1, 0.5 이다. 3단계로 참여 기기 집합의 데이터 타입 선호도를 고려한 챔피언 리스트는 <표 7>과

같다. 이는 <표 3>의 w 값이 텍스트, 동영상, 음성, 이미지에 대한 각각의 δ 값 0.24, 0.4, 0.2, 0.4미만인 정보를 <표 3>으로부터 제외한 결과이다. 4단계와 5단계를 통해 $pScore$ 값을 계산한 후의 최종 상위 7개의 검색 결과는 <표 8>과 같다.

<표 8> 최종 검색 결과
(스마트폰, AI스피커)

정보	데이터 타입	pScore
c18	텍스트	0.642
c16	텍스트	0.48
c07	음성	0.48
c17	텍스트	0.472
c08	음성	0.44
c14	이미지	0.38
c19	텍스트	0.285

<표 9> 데이터 타입 선호도를 고려한 챔피언 리스트와 검색 점수
(스마트폰, 스마트TV, AI스피커)

정보	데이터 타입	w	nw	pScore
c17	텍스트	0.85	0.26	0.333
c14	이미지	0.72	0.8	0.38
c04	동영상	0.65	0.34	0.33
c18	텍스트	0.63	0.88	0.453
c08	음성	0.58	0.3	0.308
c01	이미지	0.45	0.62	0.268
c15	동영상	0.43	0.17	0.2
c16	텍스트	0.41	0.72	0.339
c07	음성	0.35	0.61	0.336
c10	동영상	0.33	0.47	0.267

세째, 참여 기기 집합이 스마트폰, 스마트TV, AI스피커 3개의 기기로 구성될 때의 검색은 다음과 같이 처리된다. 2단계에서는 참여 기기 집합의 데이터 타입 선호도를 계산하며, 이에 대한 결과는 텍스트, 동영상, 음성, 이미지에 대해 각각 0.6, 0.67, 0.7, 0.5 이다. 3단계로 스마트폰의 데이터 타입 선호도를 고려한 챔피언 리스트는 <표 9>와 같다. 텍스트, 동영상,

음성, 이미지에 대한 δ 값은 각각 0.33, 0.3, 0.29, 0.4 이다. 4단계와 5단계를 통한 최종 검색 결과는 <표 10>과 같다.

<표 10> 최종 검색 결과
(스마트폰, 스마트TV, AI스피커)

정보	데이터 타입	pScore
c18	텍스트	0.453
c14	이미지	0.38
c16	텍스트	0.339
c07	음성	0.336
c17	텍스트	0.333
c04	동영상	0.33
c08	음성	0.308

<표 11> 데이터 타입 선호도를 고려한 챔피언 리스트와 검색 점수
(AI스피커)

정보	데이터 타입	w	nw	pScore
c17	텍스트	0.85	0.26	0.389
c18	텍스트	0.63	0.88	0.529
c08	음성	0.58	0.3	0.44
c16	텍스트	0.41	0.72	0.396
c07	음성	0.35	0.61	0.48
c19	텍스트	0.29	0.38	0.235
c05	음성	0.27	0.14	0.205

네째, 참여 기기 집합이 AI스피커 1개의 기기만으로 구성될 때의 처리 과정을 설명한다. 2단계를 통한 참여 기기 집합의 데이터 타입 선호도는 텍스트, 동영상, 음성, 이미지에 대해 각각 0.7, 0, 1, 0 이다. 3단계로 AI스피커의 데이터 타입 선호도를 고려한 챔피언 리스트는 <표 11>과 같다. 데이터 타입 선호도가 0인 동영상과 이미지 정보를 먼저 제외한 후, 텍스트와 음성에 대해서는 w 값이 각각 0.29, 0.2 미만인 정보를 최종 결과에서 제외한다. 4단계와 5단계를 통한 최종 검색 결과는 <표 12>와 같다.

<표 12> 최종 검색 결과(AI스피커)

정보	데이터 타입	pScore
c18	텍스트	0.529
c07	음성	0.48
c08	음성	0.44
c16	텍스트	0.396
c17	텍스트	0.389
c19	텍스트	0.235
c05	음성	0.205

참여 기기 집합의 데이터 타입 선호도를 고려하지 않은 기존 기법의 검색 결과는 <표 13>과 같다. 기존 기법은 제안 기법의 1단계 과정을 통해 얻은 챔피언 리스트로부터 검색 점수 *score* 값에 따라 추출된 상위 7개의 검색 결과이다.

<표 13> 기존 기법을 이용한 최종 검색 결과

정보	데이터 타입	pScore
c14	이미지	0.76
c18	텍스트	0.755
c09	이미지	0.615
c16	텍스트	0.565
c17	텍스트	0.555
c01	이미지	0.535
c04	동영상	0.495

제안 기법과 기존 기법을 결과 데이터의 정확성 측면에서 비교하면 다음과 같다. 첫째, 참여 기기가 스마트폰으로만 구성된 환경에서는 <표 6>, <표 13>과 같이 제안 기법과 기존 기법의 검색 결과가 동일함을 볼 수 있다. 이는 스마트폰의 경우에는 모든 데이터 타입에 대한 선호도값이 1이기 때문에 데이터 타입 선호도를 고려하지 않는 기존 기법과 동일한 결과가 나타나기 때문이다.

둘째, 참여 기기 집합이 스마트폰과 AI스피커일 때의 검색 결과 <표 8>과 <표 13>의 특성은 다음과 같다. 기존 기법에 비교할 때 제안 기법에서 음성 정보

는 2배 이상 증가, 텍스트 정보는 1.3배 증가, 이미지 정보는 0.3배 감소, 동영상 정보는 제공되지 않는 검색 결과를 보인다. 이는 음성 정보에 대한 선호도가 모두 높은 2개의 기기가 함께 검색에 참여함으로써 음성 정보에 대한 선호도가 더욱 높아졌기 때문이다. 이미지와 동영상 데이터 타입에 대해서는 AI스피커에서의 선호도가 0으로, 2개의 기기가 함께 참여함으로써 해당 정보에 대한 선호도가 낮아지게 된다.

세째, 참여 기기 집합이 스마트폰, 스마트TV, AI스피커 3개의 기기로 구성될 때의 검색 결과 <표 10>과 <표 13>을 비교하면 음성 정보는 2배 이상 증가한 검색 결과를 보인다. 이는 스마트폰과 AI스피커 모두에서 음성 데이터 타입에 대한 선호도값이 모두 1로 가장 높고 스마트TV에서도 선호도값이 0 이상이기 때문이다.

네째, 참여 기기 집합이 AI스피커 1개의 기기만으로 구성될 때의 검색 결과 <표 12>와 <표 13>를 비교한다. 제안 기법에서의 검색 결과는 음성 정보와 텍스트 정보로만 구성되며, 음성 정보의 개수가 3으로 제안 기법의 검색 결과와 비교할 때 가장 크게 증가한 데이터 타입이다. 이에 비해 참여 기기 특성을 고려하지 않는 기존 기법에서는, AI스피커 기기에서는 선호도가 없는 이미지와 동영상 정보를 사용자에게 제공하고 있어 정확도가 떨어지는 것을 알 수 있다.

다음으로, 제안 기법과 기존 기법을 효율성 측면에서 비교한다. 이를 위해, <표 3>의 챔피언 리스트로부터 검색 점수를 계산한 후 상위 7개의 결과를 제공하기 위해 액세스한 정보의 개수 측면에서 제안 기법과 기존 기법을 비교한다. 본 시나리오에서 액세스한 정보의 개수는, 제안 기법은 <표 3>, 참여 기기 집합이 스마트폰과 AI스피커일 때는 <표 7>, 스마트폰, AI스피커, 스마트TV일 때는 <표 9>, 그리고 AI스피커일 때는 <표 11>에 나타난 정보의 개수를 의미한다. 여기서, 참여 기기 집합이 스마트폰으로만 구성되는 경우는 제안 기법과 동일한 결과가 나타나므로 비교하

지 않는다. 15개의 정보를 액세스하는 기존 기법에 비해, <표 7>은 11개로 약 0.73배, <표 9>는 10개로 약 0.67배, <표 11>은 7개로 약 0.47배로 액세스 개수가 줄어들어 기존 기법보다 효율적임을 알 수 있다.

4개의 참여 기기 집합 시나리오를 이용한 제안 기법과 기존 기법의 비교를 통해 다음을 확인할 수 있다. 즉, 제안 기법이 참여 기기의 데이터 타입에 대한 선호도를 반영함으로써 검색 결과 데이터의 정확성이 높으면서도, 사용자에게 필요한 정보만을 추출하여 처리함에 따라 검색 효율성도 높음을 알 수 있다.

본 장에서는 임계치값 α 와 β 를 각각 0.2로 설정한 시나리오를 통해 설명을 하였다. 해당 임계치값이 달라지는 경우에는 3장에서 기술한 것과 같이 챔피언 리스트 특성에 따른 결과가 예상된다. 즉 임계치값이 커질수록 검색 속도가 빨라져 효율성이 높아지지만, k 개의 결과 정보를 제공하지 못할 수 있다. 임계치값이 작아지면 k 개의 결과 정보는 제공하지만, 검색 속도가 느려진다.

V. 결론

사물인터넷 기술을 중심으로 한 4차 산업 혁명 시대가 도래하면서, 사물인터넷 환경 특성을 반영하는 개인화 정보 검색 기법에 대한 요구가 증가하고 있다. 그러나, 기존의 개인화 정보 검색에서는 사용자만을 고려하고 있어 사물인터넷 환경에서 중요한 사물을 고려하지 못한다는 문제점을 가진다.

본 논문에서는 사물인터넷에서 참여 기기를 고려한 개인화 정보 검색 기법을 제안한다. 제안 기법에서는 정보 검색에 참여하는 기기에서 정보의 데이터 타입에 대한 각 사용자의 선호도를 기반으로 개인화된 정보를 각 사용자에게 제공한다. 또한, 여러 개의 서로 다른 기기들이 검색에 함께 참여하는 경우, 여

러 기기들의 데이터 타입 선호도를 융합함으로써 기기 집합의 특성을 반영하는 검색 결과를 제공한다.

본 논문에서는 5단계를 통해 처리되는 검색 기법과 알고리즘을 제안하였다. 다음으로, 제안한 검색 기법을 여러 개의 서로 다른 참여 기기 집합에 따른 시나리오에 적용하였으며, 이를 통해 제안 기법의 정확성과 효율성을 설명하였다.

향후 연구 과제로는, 제안 기법을 구현하고 다양한 실험 조건과 데이터 집합을 이용한 실험 평가가 필요하다. 본 논문에서는 임계치를 특정값으로만 한정하였는데 해당 값을 보다 다양화하여 실험을 하는 것이 필요하다. 그 외에도 후보 정보의 개수, 참여 기기 집합 등 보다 다양한 데이터 집합을 고려하는 것이 필요하다.

참고문헌

- [1] 정민의 · 유성진, “4차 산업혁명 시대의 사물인터넷 산업 발전전략에 관한 연구: 기업측면의 비즈니스 모델혁신 방향을 중심으로,” 한국지능정보시스템학회지능정보연구지능정보연구, 제25권, 제2호, 2019, pp.57-75.
- [2] 서두옥 · 이동호, “사물인터넷 환경에서 CoAP 기반의 저전력, 신뢰성 향상을 위한 경량 프로토콜,” 디지털산업정보학회 논문지, 제15권, 제1호, 2019, pp.21-28.
- [3] 김성림 · 권준희, “사물인터넷에서 소셜 네트워크 사용자 친밀도를 이용한 점진적 검색 기법,” 디지털산업정보학회 논문지, 제14권, 제3호, 2018, pp.1-10.
- [4] Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press, 2008.

- [5] Himani Arya, Jaytrilok Choudhary and Deepak Singh Tomar, "A Survey on Techniques for Personalization of Web Search," *International Journal of Computer Applications*, Vol.94, No.18, 2014, pp.36-41.
- [6] Luigi Atzoria, Antonio Ierab and Giacomo Morabitoc, "The Internet of Things: A survey," *Computer Networks*, Vol.54, No.15, 2010, pp.2787-2805.
- [7] Lu Tan and Neng Wang, "Future internet: The Internet of Things," *Proceedings of the 3rd International Conference on Advanced Computer Theory and Engineering*, 2010, pp.V5-376-V5-380. <http://dx.doi.org/10.1109/ICACTE.2010.5579543>.
- [8] Mouzhi Ge, Hind Bangui and Barbora Buhnova, "Big Data for Internet of Things: A Survey," *Future Generation Computer Systems*, Vol.87, 2018, pp.601-614.
- [9] Yujia Zhou, Zhicheng Dou and Ji-Rong Wen, "Enhancing Re-finding Behavior with External Memories for Personalized Search," *Proceedings of the 13th International Conference on Web Search and Data Mining*, 2020, pp.789-797.
- [10] Liu, J., Liu, C. and Belkin, N. J. "Personalization in text information retrieval: A survey," *Journal of the Association for Information Science and Technology*, Vol.5, No.3, 2019. <https://doi.org/10.1002/asi.24234>.
- [11] Bennett, P., White, R. W., Chu, W., Dumais, S.T., Bailey, P., Borisyuk, F. and Cui, X. "Modeling the impact of short- and long-term behavior on search personalization," *Proceedings of 35th Annual ACM International Conference on Research and Development in Information Retrieval*, 2012, pp.185-194.
- [12] Eickhoff, C., Collins-Thompson, K., Bennett, P.N. and Dumais, S. "Personalizing atypical web search sessions," *Proceedings of the 6th International Conference on Web Search and Data Mining*, 2013, pp.285-294.
- [13] Bai, J., Nie, J.-Y., Bouchard, H. and Cao, G. "Using query contexts in information retrieval," *Proceedings of the 30th Annual ACM International Conference on Research and Development in Information Retrieval*, 2017, pp.15-22.
- [14] Cai, F. and de Rijke, M. "Selectively personalizing query auto-completion," *Proceedings of 39th Annual ACM International Conference on Research and Development in Information Retrieval*, 2016, pp.993-996.
- [15] Shen, X., Tan, B. and Zhai, C., "Context-sensitive information retrieval using implicit feedback," *Proceedings of 28th Annual ACM International Conference on Research and Development in Information Retrieval*, 2005, pp.43-50.
- [16] Lv, Y., Sun, L., Zhang, J., Nie, J.-Y., Chen, W. and Zhang, W., "An interactive implicit feedback approach to personalized search," *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL*, 2006, pp.585-592.
- [17] Ahu Sieg, Bamshad Mobasher and Robin Burke, "Web search personalization with ontological user profiles," *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*, 2007, pp.525-534.

■ 저자소개 ■



권 준 희
Kwon, Joon Hee

2003년 3월-현재
경기대학교 컴퓨터공학부 교수
2002년
숙명여자대학교
컴퓨터과학과(이학박사)
1994년
숙명여자대학교 전산학과(이학석사)
1992년
숙명여자대학교 전산학과(학사)
관심분야 : 사물 인터넷, 정보 검색, 소셜
네트워크 서비스, 공간
데이터베이스
E-mail : kwonjh@kyonggi.ac.kr

논문접수일 : 2020년 2월 17일
수 정 일 : 2020년 2월 27일
게재확정일 : 2020년 3월 2일