감성과학
제22권 2호, 2019
<연구논문>

pISSN 1226-8593
eISSN 2383-613X
Sci. Emot. Sensib.,
Vol.22, No.2, pp.49-58, 2019
https://doi.org/10.14695/KJSOS.2018.22.2.49

인공지능 검색 서비스 활용에 따른 서비스 사용성 평가: 네이버 앱을 중심으로

Usability Evaluation of Artificial Intelligence Search Services Using the Naver App

황신희[1], 주다영[2†]

Shin Hee Hwang[1], Da Young Ju[2†]

Abstract

In the era of the 4th Industrial Revolution, artificial intelligence (AI) has become one of the core technologies in terms of the business strategy among information technology companies. Both international and domestic major portal companies are launching AI search services. These AI search services utilize voice, images, and other unstructured data to provide different experiences from existing text-based search services. An unfamiliar experience is a factor that can hinder the usability of the service. Therefore, the usability testing of the AI search services is necessary. This study examines the usability of the AI search service on the Naver App 8.9.3 beta version by comparing it with the search services of the current Naver App and targets 30 people in their 20s and 30s, who have experience using Naver apps. The usability of Smart Lens, Smart Voice, Smart Around, and AiRS, which are the Naver App beta versions of their artificial intelligence search service, is evaluated and statistically significant usability changes are revealed. Smart Lens, Smart Voice, and Smart Around exhibited positive changes, whereas AiRS exhibited negative changes in terms of usability. This study evaluates the change in usability according to the application of the artificial intelligence search services and investigates the correlation between the evaluation factors. The obtained data are expected to be useful for the usability evaluation of services that use AI.

Key words: Artificial Intelligence, Mobile Application, Search Service, Usability Evaluation, User Experience

요 약

4차 산업 혁명 시대에 인공지능은 IT 기업을 중심으로 기업들의 핵심 사업 전략이 되고 있다. 그리고 국내외 주요 포탈 기업들 또한, 인공지능 기반의 검색 서비스를 출시하고 있다. 인공지능 검색 서비스는 이미지·음성과 같은 비정형 데이터를 활용하며 검색 패러다임을 확장시켰다. 하지만 기존의 텍스트 기반의 검색 서비스와 다른 인터페이스를 제공한다. 익숙하지 않은 인터페이스는 서비스의 사용성을 저해할 수 있는 요소로, 인공지능 검색 서비스를 이용에 따른 사용성에 변화를 알아볼 필요가 있다. 본 연구는 네이버앱 8.9.3 베타버전을 사례로 인공지능 검색 서비스를 실험한다. 실험은 네이버앱 사용 경험이 있는 20대와 30대 30명을 대상으로, 네이버앱의 인공지능 검색 서비스인 스마트 렌즈, 스마트 보이스, 스마트 어라운드, AiRS 추천 콘텐츠의 사용성을 기존의 네이버앱 검색과 비교하여 평가한다. 실험분석 결과, 기존의 네이버앱 검색과 비교하여 통계적으로 유의미한 사용성 변화가 있는 것으로 나타났다.

※ 이 논문은 2019년도 정부(문화체육관광부)의 재원으로 한국콘텐츠진흥원의 지원을 받아 수행된 연구임(No. R2017030009, 차세대 이동공간 인포테인먼트 콘텐츠 및 인터랙션 개발).

※ 이 논문은 산업통산자원부 R&D사업 '창조혁신형 디자인고급인력양성사업(N0001436)'의 지원으로 진행되었음.

[1] 황신희: 연세대학교 디자인인텔리전스 석사과정

[2] † (교신저자) 주다영: 연세대학교 글로벌융합기술원 교수 / E-mail : dyju@yonsei.ac.kr / TEL : 032-749-5829

스마트 렌즈, 스마트 보이스, 스마트 어라운드는 양(+)의 상관관계가, AiRS 추천 콘텐츠는 음(-)의 상관관계가 있었다. 본 연구는 인공지능 검색 서비스를 적용에 따른 사용성 변화를 평가하고 분석한 것으로, 추후 인공지능을 활용한 서비스의 사용성 평가 연구에 유용한 자료가 될 것으로 기대한다.

주제어: 검색서비스, 모바일 어플리케이션, 인공지능, 사용자 경험, 사용성 평가

1. 서론

4차 산업 혁명 시대에 인공지능은 IT 기업을 중심으로 기업들의 핵심 사업 전략이 되고 있다(Jang, 2016). 인공지능은 빅데이터를 통한 딥러닝 기술로 끊임없이 학습 및 발전을 거듭하고, 자율주행 자동차, 무인항공기, 사물인터넷, 헬스 케어, 지식서비스 등 기술 기반의 산업에서 적극적으로 활용되고 있다. 시장 조사기관 가트너는 2018년 디지털 트렌드 중 하나로 지능형 앱 및 분석에 대해 언급 하였으며, 향후 몇 년 내에 비지니스 인텔리전스(BI)의 90%에 인공지능이 적용될 것을 예상하였다(Cearley, Burke, Searle, & Walker, 2017). 인공지능 분야는 2020년까지 전 세계적으로 주목받는 분야일 것으로 예측 및 발표되었고(Elliot & Andrews, 2017), 주요 IT기업을 중심으로 인공지능기반 산업에 적극적인 투자가 이루어지고 있다(Makridakis, 2017).

검색 분야에서 인공지능은 검색 패러다임을 변화 확장시키는 역할을 한다. 스튜어트 러셀 교수는 검색 엔진의 인공지능을 도입으로 검색 산업이 현재의 1조 달러에서 10조 달러 규모로 성장할 것이라고 전망하였다(Russell & Norving, 2016). 인공지능 검색 서비스는 사용자에게 빅데이터를 통해 더 많은 정보를 제공하고, 머신러닝/딥러닝과 같은 기술로 더 맞춤화된 정보를 제공한다. 그리고 이미지·음성과 같은 비정형 데이터를 검색에 활용 가능하게 한다. 하지만 비정형 데이터를 활용한 검색 서비스는 기존의 텍스트 기반의 검색과는 다른 경험을 제공한다는 점에서 서비스의 사용성 변화를 연구할 필요가 있다.

이에 본 연구에서는 새로운 검색 기술인 인공지능 검색 서비스를 활용함에 따라 사용성이 어떻게 변화하는지를 확인하고자 하며, 다음 두 가설을 검증한다. 첫 번째 가설은 인공지능 검색 서비스를 이용하면 사용성에 변화가 있을 것이라는 가설이다. 두 번째 가설은 인공지능 검색 방법에 따라 사용성 평가 요소별 변화의 차이가 있을 것이라는 가설이다. 연구는 Fig. 1과 같은 프로세스로 진행된다. 첫 번째, 문헌 조사를 통해 본 연구의 주요 키워드인 인공지능 검색 서비스, 사용성 평가의 개념과 연구 및 사례 대해 알아본다. 두 번째, 문헌 조사를 바탕으로 실험 가설 검증에 적합한 실험 대상, 실험 대상자를 선정하고, 사용성 평가 질문과 실험 프로세스를 설계한다. 마지막으로, 실험 후 SPSS 25.0ver으로 *t*검정을 하여 통계적 유의미함을 알아본다.

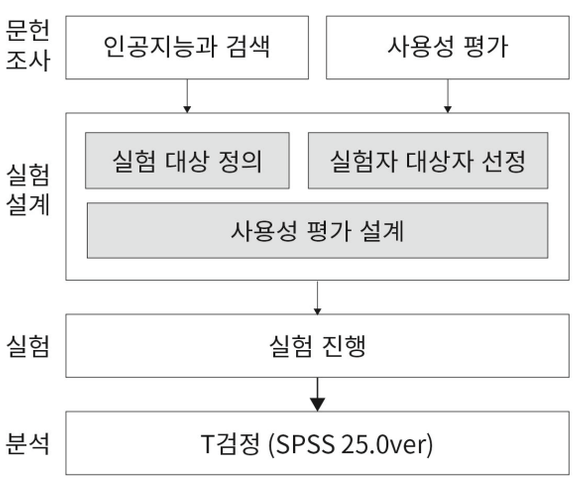

Fig. 1. 연구 프로세스

2. 이론적 배경

2.1. 인공지능과 검색

인공지능은 인간의 사고 능력을 컴퓨터로 구현시킨 기술이다. 인공지능의 기원은 앨런 튜링의 '컴퓨터와 인간이 말하는 것을 구분할 수 있는가?'를 실험한 '튜링 테스트'로(Turing, 1995), 이후에 1956년 다트머스 국제 학회에서 '지능을 가진 기계'를 '인공지능'이라고 명명하며 개념이 정립되었다(Mccarthy et al., 2006).

검색 서비스는 인공지능의 기술력을 바탕으로 사용

자의 성향과 의도에 맞춤화된 검색 결과를 제공하는 것을 목표한다(Choi & Kim, 2016). 그리고 인공지능 기술 기반의 검색 서비스는 글로벌 IT기업을 중심으로 연구 및 개발되고 있다. 구글은 딥러닝 기반의 랭크브레인(RankBrain) 알고리즘을 통해 동일한 키워드를 검색하여도 사용자의 선호 콘텐츠나 사이트에 따라 다른 검색 결과를 제공한다(Sullivan, 2016). 페이스북은 딥페이스(Deep Face)을 적용하여 친구의 얼굴을 구별한다(Taigman et al., 2014). 이외에도 애플의 시리(Siri), 삼성의 빅스비(Bixby), IBM의 왓슨(Watson)과 같은 서비스는 음성을 기반으로 한 인공지능 검색 서비스 중 하나이다.

2.2. 사용성 평가

사용성은 '명시된 조건 하에서 사용될 때 소프트웨어 제품이 사용자에게 이해되고, 학습되고, 사용되고, 매력적일 수 있는 능력', '특정 사용 환경에서 효율성, 효율성 및 만족도를 가진 지정된 목표를 달성하기 위해 특정 사용자가 제품을 사용할 수 있는 범위'라고 정의되었다(Abran et al., 2003).

닐슨(1994)은 사용성 평가를 제품 경쟁력을 높이기 위해 사용되는 방법 중 하나로 적용되는 대상에 따라 그 방법과 지표가 상이하다고 정의하였으며, 피터모빌(2004)은 IT서비스의 사용자 경험을 평가하는 방법으로 허니콤 모델을 제시하였고, 사용성 평가 요소로 유용성, 용이성, 매력성, 검색성, 접근성, 신뢰성, 가치성을 Fig. 2와 같이 제시하였다. 그리고 사용성 평가 요소에 대하여 다음과 같이 정의하였다. 유용성은 사용자가 시스템을 통해 자신이 원하는 정보를 얻는 것을 의미한다. 용이성은 사용자가 제품을 사용하기 쉬운 것을 의미한다. 매력성은 사용자가 제품에 대해 감성적인 만족을 느끼는 것을 의미한다. 검색성은 제품이 제공하는 정보가 사용자가 찾지 않아도 사용자에게 인지되는 것을 의미한다. 접근성은 제품에 접근 가능한 사용자의 범위가 넓은 것을 의미한다. 신뢰성은 제품이 제공하는 정보 혹은 서비스에 대하여 사용자가 믿을 수 있다고 느끼는 것을 의미한다. 가치성은 제품이 사용자와 제공자 모두에게 가치가 있고, 모두에게 만족을 주는 것을 의미한다(Morville, 2004).

Fig. 2. 사용자 경험 허니콤 모델

비정형 데이터를 활용한 검색 서비스를 대상으로 사용성 평가를 진행한 선행 연구에서는 평가하는 대상에 따라 평가 요소와 평가 기준을 다르게 하였으며, 평가 대상에 따라 사용성에 차이가 있었다. 김종애(2009)는 학술데이터베이스에서 텍스트 검색과 이미지 검색을 검색 소요시간, 처리동작횟수, 사용용이성, 검색메뉴 이해도, 검색결과 제시방식의 이해용이성, 만족도를 기준으로 사용성 평가하였다. 그리고 검색 소요시간, 처리동작횟수의 측면에서 텍스트 기반 검색이 더 효율적이라고 평가하였다. 최영선(2014)은 의약품 정보검색시스템에서 텍스트 기반과 이미지 기반 검색을 검색시간, 검색과정, 검색결과에서 피험자의 주관적인 만족도를 측정하였다. 피험자의 만족도는 검색 대상에 따라 달랐으며, 피험자는 의약품의 이미지 정보를 검색한 경우에 이미지 기반 검색을 유용하다고 생각하였다. 조유정 외(2017)는 GOMS 모델을 변형하여 음성 인터페이스에 최적화된 새로운 평가 모델을 제안하였다. 그리고 IPTV의 사용성을 해당 모델을 활용하여 평가하였다. 실험 결과 일반적으로 리모컨을 활용하여 사용하는 기능에서 사용성의 변화가 없었지만, 텍스트를 입력하여 검색할 경우 음성을 통한 검색의 사용성이 향상된 것을 확인하였다. 복경수 외(2014)는 시간과 위치를 기반의 소셜 검색 기법을 제안하고, 불필요한 연산량을 감소시켜 검색 속도를 향상시킨다는 것을 확인하였다. 고광수 외(2011)는 특허정보 검색 분야에서 텍스트 마이닝을 활용하여 검색 정확도를 높이고, 검색 시간을 단축시킨다는 것을 확인하였다.

3. 실험 대상 및 방법

3.1. 실험 대상 정의

본 연구는 네이버앱 8.9.3 베타버전을 사례로 실험을 진행한다. 네이버는 최근 1년간 국내 검색 포털 사이트의 약 70%의 시장 점유율을 차지하고 있는 기업이다(Internettrend, 2018). 네이버가 최근에 공개한 네이버앱 베타버전은 네이버앱이 기존에 제공하고 있던 인공지능 검색 서비스인 스마트 렌즈, 스마트 보이스, 스마트 어라운드를 표면적으로 강조하고 있고, 인공지능 연관 콘텐츠를 추천 서비스인 AiRS 추천 콘텐츠가 추가 되어있다(Shim, 2018).

스마트 렌즈 스마트렌즈는 이미지 인식기술 기반의 검색 서비스이다. 스마트렌즈는 휴대폰의 카메라 기능을 통해 사진을 찍거나, 휴대폰에 저장되어있는 이미지를 불러와 검색 정보를 입력한다. 활용 가능한 기능으로는 동일한 이미지 찾기, QR코드 링크 이동하기, 사진 속의 텍스트 번역 및 검색하기 등이 있으며, 쇼핑 및 일반 검색 영역에서 주로 활용된다.

스마트 보이스 스마트 보이스는 음성 인식기술 기반의 검색 서비스이다. 네이버의 챗봇 서비스인 네이버i가 네이버앱에 적용된 형태로, 소리를 통해 검색 내용을 입력한다. 입력 내용에 대한 답변은 청각 정보와 시각정보를 동시에 제공된다. 음악 제목 찾기 기능과 더불어, 기존의 텍스트 입력으로 검색하던 대부분의 검색이 가능하기 때문에 기능이 활용되는 분야는 다양하다.

스마트 어라운드 스마트 어라운드는 인공지능 장소 검색 서비스로, 네이버의 인공지능 장소 추천 시스템인 에어스페이스(AiRSPACE, AiRS + Place and Context Embedded)가 적용되어있다. 이 서비스는 사용자의 위치 정보를 활용하여 사용자 주변의 유명 음식점, 가볼만한 곳, 문화 강좌를 추천한다.

AiRS 추천 콘텐츠 AiRS는 인공지능기반의 추천 서비스로, CF(Collaborative Filtering, 협력 필터)기술과 RNN(Recurrent Neural Network, 인공신경망)기술이 적용되어있다. AiRS 추천 콘텐츠는 사용자가 보고 있는 블로그 혹은 뉴스 콘텐츠의 텍스트 분석결과와 사용자의 콘텐츠 소비 패턴을 활용하여 연관 콘텐츠를 추천한다.

3.2. 실험 대상자 선정

피험자는 네이버앱의 사용 경험이 있는 20대와 30대이다. 모바일 환경에서 기존의 텍스트 중심의 검색 서비스와 인공지능을 활용한 검색 서비스의 사용 경험을 동일한 환경에서 비교하기 위하여 기존에 네이버앱 경험이 있는 사용자를 대상으로 하고, 피험자가 인공지능 검색 서비스의 개념을 단시간에 이해 가능해야 한다는 점에서 비교적 모바일 활용도가 높은 연령을 대상으로 한다.

제이콥 닐슨은 5명을 대상으로 사용성 평가를 진행하면 사용성에 대한 모든 문제를 파악할 수 있다고 하였다(Nielsen, 2000). 그리고 선행 연구에 따르면 30명의 표집크기면 자료 분석 결과가 유의미하다고 간주된다(Hong et al., 1996). 결과적으로, 선행연구에 따라 본 연구에서는 총 32명의 피험자를 선정하였다.

3.3. 사용성 평가 설계

본 연구에서는 피터모빌의 허니콤 모델을 바탕으로 사용성 평가 요소를 유용성, 용이성, 검색성, 매력성, 접근성으로 정하였다(Morville, 2004). 피터모빌의 허니콤 모델 중 신뢰성은 네이버 자체가 제공하는 정보의 질과 연관된 요소이고, 가치성은 사용자뿐만 아니라 서비스 제공자인 네이버의 입장도 고려된 평가요소라고 판단하여 본 실험의 사용성 평가 요소에서는 제외하였다. 사용성 평가 요소의 조작적 정의는 Table 1과 같으며, 이를 기준으로 설문 문항을 구성하였다. 그리고 각 설문 문항에 대한 답변을 리커트 5점 척도(매우 그렇다, 그렇다, 보통이다, 그렇지 않다, 매우 그렇지 않다)로 받았다.

Table 1. 사용성 평가 요소의 조작적 정의

사용성 평가 요소	조작적 정의
유용성	사용자가 서비스를 통하여 원하는 정보를 얻는 것
용이성	사용자가 서비스를 사용하기 쉬운 것
검색성	서비스가 제공하는 정보가 사용자에게 노출되어 인지되는 것
접근성	서비스에 접근 가능한 사용자의 범위가 넓은 것
매력성	사용자가 서비스에 대해 감성적인 만족을 느끼는 것

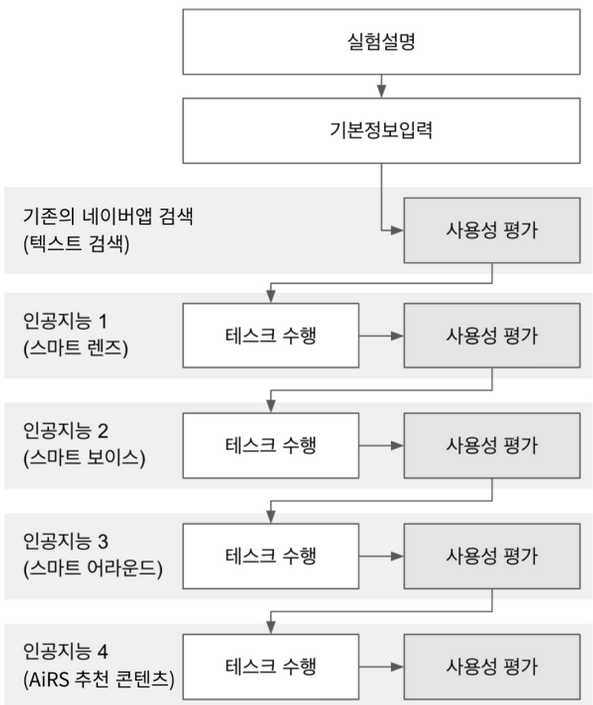

Fig. 3. 실험 프로세스

3.4. 실험 프로세스 설계

실험은 진행에 필요한 네이버앱 베타버전은 정식 출시되지 않은 서비스로 설치과정이 번거롭고, 실험하는 기능이 피험자에게 익숙하지 않은 기능일 수 있다는 점에서 대면 설문조사를 통해 실험을 진행한다.

실험 프로세스를 Fig. 3과 같다. 첫 번째, 실험 참여자에게 실험 개요를 설명한다. 두 번째, 실험 참여자의 인구 통계학적 정보를 수집한다. 세 번째, 인공지능 검색을 활용하기 전 네이버앱의 사용성을 기존 네이버앱의 텍스트 기반 검색의 사용경험을 바탕으로 평가한다. 마지막으로, 네이버 앱의 인공지능 검색 서비스인 스마트 렌즈, 스마트 보이스, Table 2 인공지능 검색서비스 활용을 위한 검색 테스크 스마트 어라운드, AiRS 추천 콘텐츠 각각에 대하여 테스크를 수행하며 서비스를 이용하고, 사용성을 평가하는 프로세스를 반복한다. 검색 테스크는 한정된 시간 내에 인공지능 검색 서비스의 특징에 대한 이해를 돕기 위하여 각 검색 서비스의 활용도가 높은 예시를 중심으로 Table 2와 같이 구성하였다.

결과 분석 후, 음(-)의 상관관계가 있었던 AiRS 서비스에 한하여 사용성이 더 나빠졌다고 대답한 피험자

Table 2. 인공지능 검색 서비스 활용을 위한 검색

서비스	검색 테스크
스마트 렌즈	스마트 렌즈 기능을 사용하여 제시된 동일한 옷을 검색해주세요.
	스마트 렌즈 기능을 사용하여 약통 이름을 번역해주세요.
스마트 보이스	스마트 보이스 기능을 두 번 사용해 주세요. 예시 : 날씨, 최근 개봉 영화, 롱 패딩 등
스마트 어라운드	스마트 어라운드를 활용하여 내 주변 맛집을 검색해주세요.
AiRS 추천 콘텐츠	다음 순서에 따라 검색을 진행해 주세요. 1. 검색창에서 '영화 신비한 동물사전'을 검색해 주세요. 2. 눈길을 끄는 내용의 블로그를 클릭해주세요. 3. AiRS 추천 콘텐츠를 사용하여 추천되는 콘텐츠를 확인하고, 눈길을 끄는 콘텐츠를 선택해주세요.

21명 중 10명을 대상으로 그렇게 생각한 이유에 대하여 회상적 발성사고법으로 심층인터뷰를 진행하였다.

3.5. 결과 분석

실험을 통해 도출된 결과는 설문지의 데이터 코딩을 거쳐 SPSS 25.0ver을 사용하였으며, 유의수준은 $p<0.05$로 설정하여 통계적 자료를 처리하였다. 본 실험에 적용되는 대립 가설은 총 두 가지이다. 첫 번째, 검색 서비스를 이용하면 사용성에 변화가 있을 것이다. 두 번째, 인공지능 검색 방법에 따라 사용성 평가 요소별 변화의 차이가 있을 것이다. 본 실험에서는 한 집단의 사전과 사후 점수 차이가 있다는 대립가설을 통계적으로 검증하는데 적합한 분석 방법인 대응표본 t검정을 실시하였다.

4. 실험 분석 결과

4.1. 실험 대상자의 인구통계학적 특성

실험은 2018년 11월 12부터 11월 18일까지 7일간 진행되었으며, 해당 기간 동안 20대와 30대 32명을 샘플링 하였다. 그 중 네이버앱 사용 경험이 없어 실험

에 적합하지 않은 2명을 제외하였고, 총 30명의 실험 결과를 분석하였다. 분석 대상자의 인구 통계학적 사항에 대한 빈도 분석 결과는 Table 3과 같다. 성별은 남성이 15명(50%), 여성이 15명(50%)로 동일한 비율을, 연령대는 20대 20명(66.7%), 30대 10명(33.3%)의 비율 구성을 보였다.

Table 3. 분석 대상자의 인구통계학적 특성

조사내용		빈도	%
성별	남	15	50
	여	15	50
	소계	30	100
연령별	20대	20	66.7
	30대	10	33.3
	소계	30	100

4.2. 인공지능 검색 서비스 이용에 따른 사용성 변화

기존의 네이버앱의 사용성과 네이버앱 iOS 8.9.2의 인공지능 검색 서비스인 스마트 렌즈, 스마트 보이스, 스마트 어라운드, AiRS 추천 콘텐츠의 사용성 평가를 분석한 결과, 스마트 렌즈, 스마트 보이스, 스마트 어라운드에서는 양(+)의 상관관계가, AiRS에서는 음(-)의 상관관계가 있었다. 결과적으로, 첫 번째 가설이었던 '검색 서비스를 이용하면 사용성에 변화가 있을 것이다.'라는 가설은 채택되었다.

4.3. 사용성 평가 요소별 변화

사용성 평가 요소는 유용성, 용이성, 검색성, 접근성, 매력성 총 다섯 가지로, 설문은 요소별 해당하는 질문으로 구성되었다. 각 질문에 대한 답변은 리커트 5점 척도를 활용하여 수집 되었고, 평가 결과는 최저점을 1점으로, 최고점을 5점으로 점수를 환산하였다. 동일한 평가 요소에 해당하는 문항에 대해서는 각 점수의 평균을 최종 값으로 사용하였다. 기존의 네이버앱 검색 서비스였던 텍스트 검색의 사용성 평가 요소의 평균은 Table 4와 같이 유용성이 3.68, 용이성이 3.28, 검색성이 3.20, 접근성이 3.98, 매력성이 3.70으로 나타났다. 그리고 인공지능 검색 서비스 사용에 따른 사용성 변화 정도와 변화의 통계적 유의미함을 검증하기 위하여 대응표본 *t*검정을 실시한 결과는 Table 5와 같다.

Table 4. 텍스트 검색 사용성 평가 기술통계

검색 서비스	평가요소	평균	표준편차
기존 네이버앱	유용성	3.68	0.91
	용이성	3.28	0.83
	검색성	3.20	0.85
	접근성	3.98	0.71
	매력성	3.70	0.95

Table 5. 인공지능 검색 이용에 따른 사용성 변화 대응표본 t검정

검색 서비스	평가 요소	평균		표준편차		t값	p값
		기존 검색	인공 지능	기존 검색	인공 지능		
스마트 렌즈	유용성	3.68	3.83	0.91	1.07	-0.77	0.07
	용이성	3.28	3.90	0.83	1.07	-3.15	0.44
	검색성	3.20	3.32	0.85	0.92	-0.70	0.01
	접근성	3.98	3.75	0.71	1.02	1.25	0.66
	매력성	3.70	3.70	0.95	1.41	0.00	0.85
스마트 보이스	유용성	3.68	3.93	0.91	0.85	-1.61	0.00
	용이성	3.28	3.97	0.83	1.15	-3.26	0.57
	검색성	3.20	3.62	0.85	0.83	-2.75	0.00
	접근성	3.98	4.12	0.71	0.78	-0.94	0.04
	매력성	3.70	3.95	0.95	0.90	-1.53	0.00
스마트 어라 운드	유용성	3.68	3.75	0.91	0.84	-0.44	0.00
	용이성	3.28	3.82	0.83	0.93	-3.15	0.07
	검색성	3.20	3.83	0.85	0.82	-4.23	0.01
	접근성	3.98	3.85	0.71	1.04	0.70	0.61
	매력성	3.70	3.70	0.95	1.16	0.00	0.26
AiRS 추천 콘텐츠	유용성	3.68	2.98	0.91	0.84	4.58	0.00
	용이성	3.28	2.75	0.83	1.07	2.74	0.24
	검색성	3.20	2.65	0.85	1.11	2.72	0.08
	접근성	3.98	2.93	0.71	1.22	4.71	0.89
	매력성	3.70	2.52	0.95	1.26	5.16	0.31

텍스트 검색을 사용하였을 때와 비교하여 스마트 렌즈를 사용하였을 때 사용성 변화는 Fig. 4와 같이 검색성이 통계적으로 유의미한 변화가 있었고, 양(+)의 상관관계가 있었다. 검색성은 평균이 3.32 (t=-0.70, p= 0.01)로 0.12 증가하였다.

스마트 보이스를 사용하였을 때 사용성 변화는 Fig. 5와 같이 유용성, 검색성, 접근성, 매력성이 통계적으로 유의미한 변화가 있었고, 모두 양(+)의 상관관계가 있었다. 유용성은 평균이 3.93 (t=-1.61, p= 0.00)로 0.25 증가하였고, 검색성은 평균이 3.62 (t=2.75, p=0.00)로

Fig. 4. 텍스트 검색과 스마트 렌즈 평균 비교

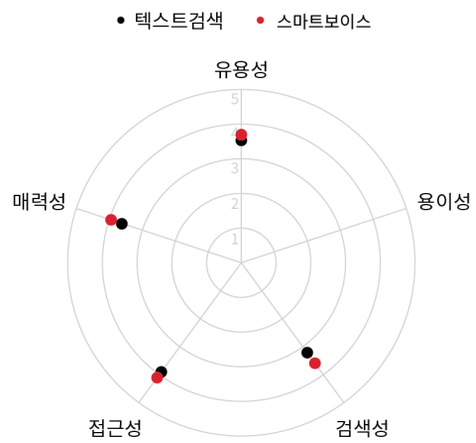

Fig. 5. 텍스트 검색과 스마트 보이스 평균 비교

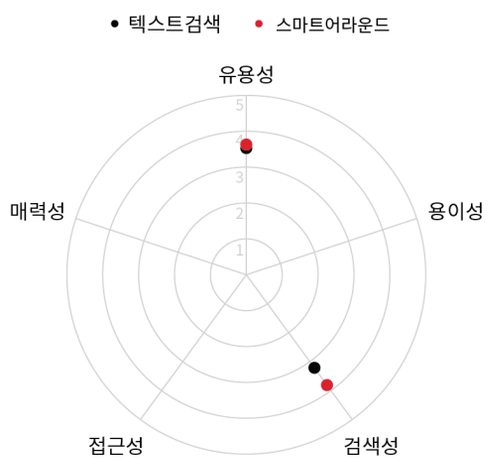

Fig. 6. 텍스트 검색과 스마트 어라운드 평균 비교

0.42 증가하였고, 접근성은 평균이 4.12 (t=-0.94, p=0.04)로 0.13 증가하였고, 매력성은 평균이 3.95(t =-1.53, p=0.00)로 0.25 증가하였다.

Fig. 7. 텍스트 검색과 AiRS 추천 콘텐츠 평균 비교

스마트 어라운드를 사용하였을 때 사용성 변화는 Fig. 6과 같이 유용성, 검색성이 통계적으로 유의미한 변화가 있었고, 모두 양(+)의 상관관계가 있었다. 유용성은 평균이 3.75 (t=-0.44, p=0.00)로 0.07 증가 하였고, 검색성은 평균이 3.83 (t=-4.23, p=0.01)로 0.63 증가하였다.

AiRS 추천 콘텐츠를 사용하였을 때 사용성 변화는 Fig. 7과 같이 유용성이 통계적으로 유의미한 변화가 있었고, 음(-)의 상관관계가 있었다. 유용성은 평균 2.98 (t=4.58, p=0.00)로 0.70 감소하였다.

사용성 평가 분석 후, 사용성의 변화가 음(-)의 상관관계가 있었던 AiRS에 대하여 사용성이 더 나빠졌다고 대답한 피험자 21명 중 10명을 대상으로 그렇게 생각한 이유에 대하여 회상적 발성사고법으로 심층 인터뷰를 진행하였다. 피험자들은 서비스가 추천해주는 키워드와 콘텐츠가 보고 있던 콘텐츠와 연관성이 떨어지고, 실질적으로 원하는 콘텐츠를 추천해주지 않는다고 느낀다고 말하였다. 그 결과 기존의 텍스트 검색과 비교하여 차별성을 느끼지 못하여 유용성이 낮게 평가한 것으로 나타났다. 통계적으로 유의미한 요소는 아니었지만 AiRS 추천 콘텐츠의 접근성에 관한 의견으로 콘텐츠를 보고 있는 화면에 한하여서만 AiRS 추천 콘텐츠를 사용할 수 있다는 점에서 서비스의 접근성이 좋지 않다고 느낀다고 대답하였다.

결과적으로, 두 번째 가설이었던 '인공지능 검색 방법에 따라 사용성 평가 요소별 변화의 차이가 있을 것이다.'라는 가설은 채택되었다.

5. 결론 및 제언

본 연구는 인공지능 검색 서비스를 사용함에 따른 사용성 변화를 확인하는 것이며, 다음 연구 가설을 검증하였다. ‘검색 서비스를 이용하면 사용성에 변화가 있을 것이다.’, ‘인공지능 검색 방법에 따라 사용성 평가 요소별 변화의 차이가 있을 것이다.’. 본 연구는 네이버앱 iOS 8.9.2 베타버전의 인공지능 검색 기능인 스마트 렌즈, 스마트 보이스, 스마트 어라운드, AiRS 추천 콘텐츠를 대상으로 실험을 설계하였으며, 사용성 평가를 서비스의 유용성, 용이성, 검색성, 접근성, 매력성 측면에서 진행하였다.

실험은 네이버앱 경험이 있는 20대와 30대 32명을 샘플링하여 직접대면 설문조사를 실시하였고, 그 중 실험에 적합한 30명의 데이터를 분석하였다. 그러므로 본 실험의 결과는 20대와 30대 만의 한정된 결과이다. 실험은 피험자에게 인공지능 검색 서비스를 경험하기 이전에 기존의 네이버앱의 텍스트 검색에 대한 사용성 평가 설문을 한 후, 네이버앱 베타버전의 인공지능 검색 서비스를 사용하고 사용성 평가를 진행하였다. 그리고 사용성 평가 분석 후, 사용성이 음(-)의 상관관계가 있었던 AiRS 추천 콘텐츠에 한하여 부정적으로 답변을 했던 사용자 21명 중 10명을 대상으로 회상적 발성사고법으로 심층 인터뷰를 진행하였다.

실험 분석 결과 인공지능 검색을 사용하였을 때 기존의 텍스트 기반 검색과 비교하여 통계적으로 유의미한 사용성 변화가 있음을 확인하였다. 각 검색에서 요소별 변화는 스마트 렌즈에서 검색성이, 스마트 보이스에서 유용성, 검색성, 접근성, 매력성이, 스마트 어라운드에서 유용성과 검색성이 양(+)의 상관관계가 있었으며, AiRS 추천 콘텐츠에서 유용성이 음(-)의 상관관계가 있었다. 그리고 AiRS 추천 콘텐츠의 사용성에 대하여 부정적 답변한 피험자 10명 중 9명은 해당 서비스를 통해 추천된 키워드와 콘텐츠가 만족스럽지 않고, 기존의 텍스트 기반의 검색과 차이를 느끼지 못하였다고 대답하였다.

기존의 연구에서는 검색 대상과 검색 방법이 적절하게 적용된 경우 검색 서비스의 사용성이 향상 되는 것을 확인하였다. 텍스트 기반의 정보검색과 이미지 기반의 정보 검색을 비교한 선행 연구에서는 검색되는 콘텐츠에 따라 사용성이 다르게 나타났다(Kim, 2009; Choi, 2014). 그리고 본 연구에서도 이미지 기반의 인공지능 검색을 사용하였을 때 사용성이 향상되었다. 음성검색을 IPTV에서 활용한 선행 연구와 위치와 시간을 소셜 검색에서 활용한 선행연구에서는 검색 속도가 향상되는 것으로 나타났다(Cho et al., 2017; Bok et al., 2014). 본 연구에서도 음성기반, 위치기반의 인공지능 검색을 사용하였을 때 텍스트 검색과 비교하여 사용성이 향상되었다. 텍스트 마이닝을 특허정보 컨텐츠 검색에서 활용한 선행연구에서는 검색 정확도가 높아지고, 검색 시간이 단축되었다(Go et al., 2011), 하지만 본 연구의 텍스트 마이닝을 활용한 검색 콘텐츠 추천에서는 사용성이 낮아지는 것을 확인하였다.

기존의 텍스트 중심 검색의 사용성은 평가 요소별 평균 점수가 보통을 의미하는 3점 이상의 점수를 받고 있었으며 보통 이상의 사용성을 가지고 있다고 판단된다. 그리고 본 연구에서는 사용성 변화에 양(+)의 상관관계가 있던 네이버앱의 스마트 렌즈, 스마트 보이스, 스마트 어라운드 사례를 통해 선행연구의 결과와 동일하게 검색되는 콘텐츠가 인공지능 검색 서비스가 제공하는 인터페이스와 부합한 경우, 사용성이 증가하는 것을 확인하였다. 특히 사용성 평가 요소 중 검색성의 점수가 검색 서비스 사례 모두에서 유의미한 변화가 있었는데, 이러한 결과는 인공지능 검색 서비스는 검색에 있어서 적절한 인터페이스를 제공 하고, 제공된 인터페이스가 사용자가 다양한 정보에 정보에 접근할 수 있게 하기 때문이라고 판단한다. 하지만 음(-)의 상관관계가 있던 AiRS추천의 사용성 평가 점수가 모두 3점 이하의 점수로 특히 유용성의 변화가 유의했던 점에서 인공지능의 검색 서비스가 적절한 콘텐츠를 추천하지 못하여 유용하지 못하다고 판단되었을 경우 보통 이하의 사용성으로 이어질 가능성이 있다.

본 연구는 인공지능 검색 서비스를 실제 서비스 중인 서비스를 대상으로 하여 피험자가 직접 서비스를 이용해보고 연구했다는 점에서 의의가 있으며, 추후의 인공지능을 활용한 서비스의 사용성 평가와 더불어 인공지능 검색 서비스 개발 및 연구 분야에 유용한 자료가 될 것으로 기대한다. 하지만 본 연구에는 다음

과 같은 한계가 있다. 첫 번째, 본 연구 대상인 네이버 앱 8.9.3 베타버전의 검색 서비스는 모두 딥러닝 기술이 적용되어 있다. 검색 서비스에서 딥러닝 기술은 사용자의 피드백을 통해 검색 알고리즘을 다시 학습시키며 검색 결과의 정확도를 향상시키는 역할을 한다. 그러므로 본 실험 이후 지속적으로 검색 기능이 향상되며 이것이 이후의 사용성에 영향을 줄 가능성이 있다. 두 번째, 실험에 사용된 네이버앱 베타버전은 상용화되지 않은 버전이다. 기존 버전의 네이버앱과 비교하여 인터페이스의 차이가 크게 있으며, 피험자가 해당 앱의 인터페이스가 익숙하지 않다는 점이 평가 결과에 영향을 주었을 가능성이 있다. 세 번째, 피험자가 각각의 인공지능 검색 서비스를 체험하는 과정에서 임의의 검색 키워드를 제시하였는데, 이 키워드는 인공지능 검색 서비스에서 비교적 잘 검색되는 키워드였다는 점에서 실제로 검색 환경과 차이가 있을 수 있다. 마지막으로, 샘플 사이즈가 작고 대상을 20대와 30대로 한정하여 실험 결과에 오차가 있을 수 있다. 추후 인공지능 검색 서비스를 대상으로 사용성 평가를 설계할 경우, 충분한 수의 피험자를 모집하고, 대상자가 서비스에 적응할 수 있는 시간과 환경을 제공할 것을 권장한다.

REFERENCES

Abran, A., Khelifi, A., Suryn, W., & Seffah, A. (2003). Usability meanings and interpretations in ISO standards. *Software Quality Journal, 11*(4), 325-338. DOI: 10.1023/A:1025869312943

Bok, K. S., Ahn, M. J., Im. J. T., & Yoo, J. S. (2014). Efficient location based social search considering time property in mobile environments. *KIISE Transactions on Computing Practices, 20*(4), 243-247.

Cearley, D. W., Burke, B., Searle, S., & Walker. M. J. (2011). The top 10 strategic technology trends for 2018. *Gartner*. Retrieved from https://www.gartner.com/smarterwithgartner/gartner-top-10-strategic-technology-trends-for-2018/

Cho, Y. J., Han, S. H., Lee, K.T., Jung, G. S., & Cha, J. H. (2017). Study on the usability testing for voice-controlled user interface : Focused on IPTV usage. *Proceedings of the Korean Information Science Society* (pp. 1334-1336).

Choi, Y. S. (2014). *Comparative study on usability evaluation of text vs. graphic-based drugs retrieval system*. Master's Thesis. University of Sungkyunkwan, Seoul.

Choi, Y. R., & Kim, K. H. (2016). Artificial intelligence overview and application examples. *Industrial Engineering Magazine, 23*(2), 23-29.

Elliot, B. & Andrews, W. (2017). *A Framework for Applying AI in the Enterprise*. USA: Gartner. ID: G00336031.

Go, G. S., Jung, W. K., Shin, Y. G., Park, S. S., &Jang, D. S. (2011). A Study on Development of Patent Information Retrieval Using Textmining. *Journal of the Korea Academia-Industrial Cooperation Society*. 12(8), 3677-3688.

Hong, J. S., Park, O. H., & Choi, C. H. (1996). *Analysis of Survey Methods and Statistical Data*. Seoul: Packyongsa.

Internettrend. (2018). Search engine trends. *Internettrend*. Retrieved from http://www.internettrend.co.kr/

Jang, P. S. (2016). [EU] 2016 Davos forum: What are our strategies for the forthcoming Fourth Industrial Revolution?. *Science & Technology Policy, 26*(2), 12-15.

Kim, J. A. (2009). Usability evaluation of text-based search and visual search of a multidisciplinary library database. *Journal of the Korean Society for Information Management, 26*(3), 111-129. DOI: 10.3743/KOSIM.2009.26.3.111

Makridakis, S. (2017). The forthcoming artificial intelligence (AI) revolution: Its impact on society and firms. *Futures, 90*, 46-60. DOI: 10.1016/j.futures.2017.03.006

Mccarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (2006). A proposal for the dartmouth summer research project on artificial intelligence, august 31, 1955. *AI Magazine, 27*(4), 12. DOI: 10.1609/aimag.v27i4.1904

Morville, P. (2004). User experience design. *Semantic Studios*. Retrieved from https://semanticstudios.

com/user_experience_design/

Nielsen, J. (1994). Usability inspection methods. In *Conference companion on Human factors in computing systems* (pp. 413-414). New York: ACM.

Nielsen, J. (2000). Why you only need to test with 5 users. *Nielsen Norman Group*. Retrived from https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/

Russell, S. & Norvig, P. (1995). *Artificial Intelligence: A Modern Approach.* New Jersey: Prentice-Hall.

Shim, J. S. (2018). Naver has a strong AI passion. More than you think. *Dongascience*. Retrieved from http://dongascience.donga.com/news.php?idx=21978

Sullivan, D. (2016). FAQ: All about the google rankbrain algorithm. *Search Engine Land*. Retrieved from http://searchengineland.com/faq-all-about-the-new-google-rankbrain-algorithm-234440

Taigman, Y., Yang, M., Ranzato, M. A., & Wolf, L. (2014). Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1701-1708).

Turing, A. M. (1995). Computing machinery and intelligence. *Brian Physiology and Psychology, 213*. DOI: 10.1007/978-1-4020-6710-5_3

Unger, R., & Chandler, C. (2009). *A project guide to UX design*. India: Pearson.

원고접수: 2019.01.23
수정접수: 2019.05.26
게재확정: 2019.06.15

ⓒ 2019 (by) the authors. This open access article is distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided that the original work is properly cited.