

뇌파, 시선추적 및 인공지능 기술에 기반한 디지털 도서관 인터페이스 연구: 암묵적 적합성 피드백 활용을 중심으로*

Digital Library Interface Research Based on EEG, Eye-Tracking, and Artificial Intelligence Technologies: Focusing on the Utilization of Implicit Relevance Feedback

김현희 (Hyun-Hee Kim)**

김용호 (Yong-Ho Kim)***

초 록

본 연구는 디지털 도서관의 콘텐츠를 탐색하는 동안 이용자의 암묵적 적합성 피드백을 활용하여 적합성을 판단하기 위해 뇌파 기반 및 시선추적 기반 방법들을 제안하고 평가해 보았다. 이를 위해서 32명을 대상으로 하여 동영상, 이미지, 텍스트 데이터를 활용하여 뇌파/시선추적 실험들을 수행하였다. 제안된 방법들의 유용성을 평가하기 위해서, 딥러닝 기반의 인공지능 방법들을 경쟁 기준으로 사용하였다. 평가 결과, 주제에 적합한 동영상과 이미지(얼굴/감정)를 선택하는 데에는 뇌파 컴포넌트 기반 방법들(av_P600, f_P3b)이 높은 분류 정확도를 나타냈고, 이미지(객체)와 텍스트(신문 기사)를 선택하는 데에는 인공지능 기반 방법 즉, 객체 인식 기반 방법과 자연언어 처리 방법이 각각 높은 분류 정확도를 나타냈다. 끝으로, 뇌파, 시선추적 및 인공지능 기술에 기반한 디지털 도서관 인터페이스를 구현하기 위한 지침 즉, 암묵적 적합성 피드백에 기반한 시스템 모형을 제안하고, 분류 정확도를 향상시키기 위해서 미디어별로 적합한 뇌파 기반, 시선추적 기반 및 인공지능 기반 방법들을 제시하였다.

ABSTRACT

This study proposed and evaluated electroencephalography (EEG)-based and eye-tracking-based methods to determine relevance by utilizing users' implicit relevance feedback while navigating content in a digital library. For this, EEG/eye-tracking experiments were conducted on 32 participants using video, image, and text data. To assess the usefulness of the proposed methods, deep learning-based artificial intelligence (AI) techniques were used as a competitive benchmark. The evaluation results showed that EEG component-based methods (av_P600 and f_P3b components) demonstrated high classification accuracy in selecting relevant videos and images (faces/emotions). In contrast, AI-based methods, specifically object recognition and natural language processing, showed high classification accuracy for selecting images (objects) and texts (newspaper articles). Finally, guidelines for implementing a digital library interface based on EEG, eye-tracking, and artificial intelligence technologies have been proposed. Specifically, a system model based on implicit relevance feedback has been presented. Moreover, to enhance classification accuracy, methods suitable for each media type have been suggested, including EEG-based, eye-tracking-based, and AI-based approaches.

키워드: 뇌파, 시선추적, 인공지능, 암묵적 적합성 피드백, 디지털 도서관 인터페이스

EEG, eye tracking, artificial intelligence, implicit relevance feedback, digital library interfaces

* 이 논문은 2020년 대한민국 교육부와 한국연구재단의 인문사회분야 중견연구자지원사업의 지원을 받아 수행된 연구임(NRF-2020S1A5A2A01040945).

** 명지대학교 문헌정보학과 명예 교수(kimhh@mju.ac.kr) (제1저자, 교신저자)

*** 부경대학교 미디어커뮤니케이션학부 명예 교수(yh1228kim@gmail.com) (공동저자)

■ 논문접수일자: 2024년 2월 15일 ■ 최초심사일자: 2024년 2월 29일 ■ 게재확정일자: 2024년 3월 4일

■ 정보관리학회지, 41(1), 261-282, 2024. <http://dx.doi.org/10.3743/KOSIM.2024.41.1.261>

© Copyright © 2024 Korean Society for Information Management

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

1. 서론

디지털 도서관의 데이터베이스 내에 텍스트, 이미지(비주얼), 오디오 및 동영상(오디오-비주얼)이 생산되어 축적되고 있다. 많은 연구들이 뇌파(electroencephalography, EEG), 시선 추적(eye tracking, ET) 같은 비용 효율적인 신경과학 연구방법들을 사용하여 이용자의 특성이나 다양한 종류의 미디어를 수용할 수 있는 정보 검색 환경에서 맞춤형 정보 서비스를 제공할 수 방안들을 제안하고 있다(Davis, Spapé, & Ruotsalo, 2022; Jacucci et al., 2019). 뇌파 정보는 이용자의 인지 과정과 집중 상태를 드러낸다. 시선 정보는 이용자가 관심 있는 대상을 식별할 수 있게 해주고, 시선 경로 분석을 통해 이러한 대상에 대한 반응을 측정할 수 있게 한다. 따라서, 이 두 종류의 정보는 미디어 콘텐츠 분석을 위해 상호 보완적으로 사용될 수 있다.

디지털 도서관의 검색 환경에서, 이용자의 주관적인 선호도(또는 적합성)를 추론하기 위해서 이용자가 자신의 선호도 정보를 직접 입력하게 하거나, 도서관 시스템과의 상호 작용에 의해 수집된 데이터(예: 클릭 수, 체류 시간)를 사용할 수 있다(Davis, Spapé, & Ruotsalo, 2022). 이용자에게 입력을 요청하는 것은 항상 실용적이지 않으며, 클릭 수나 체류 시간과 같은 데이터를 적합성의 대리 지표로 사용할 때 신뢰성을 확보하기 어렵다. 뇌파 및 시선 추적을 기반으로 한 방법들은 암묵적인 적합성 피드백을 활용하여 이용자의 선호도와 관심사를 더 정확하게 이해할 수 있게 해 준다.

미디어 콘텐츠를 검색할 때, 콘텐츠 자체(내

부 정보)나 이용자 데이터, 맥락 정보와 같은 외부 정보를 활용할 수 있다. 뇌파 및 시선 추적 기반 방법들은 콘텐츠 자체를 다루는 것이 아니라 이용자 반응(외부 정보)만을 다룬다는 한계를 가지고 있다. 최근에는 딥러닝을 이용하여 미디어 콘텐츠의 의미를 분석하는 연구들이 진행되고 있다. 딥러닝 기반 객체 검출 알고리즘의 비약적인 발전에도 불구하고 대체로 높은 분류 정확도를 나타내는 텍스트, 이미지 분석과는 다르게 동영상(비디오) 분석에서는 여전히 어려움을 겪고 있다. 이는 비디오를 요약하는 경우를 살펴보면, 시청자가 전체 비디오를 시청하면서 추론하는 비디오의 실제 의미적 주제와 관련된 객체(또는 인물)가 포함된 비디오 쇼트(shot)로 구성된 비디오 요약에서 추론되는 주제 사이의 의미적 차이 때문으로 생각된다. 뇌파 및 시선추적 기반 방법은 오디오-비주얼 시맨틱스에 기반하여 콘텐츠 의미 분석을 수행하기 때문에 이러한 의미적 격차 문제를 어느 정도 완화시킬 수 있을 것으로 기대한다.

의미적 격차 문제, 이용자 반응만을 이용하는 문제를 보완하기 위해서 인공지능 기술에 의한 미디어 콘텐츠의 내용 분석(예, 객체, 얼굴 및 행동 분석) 결과를 이용자 반응을 기반으로 한 뇌파/시선추적 기반 방법들에 의한 미디어 콘텐츠 분석 결과와 서로 결합한다면 좀 더 정확한 미디어 콘텐츠 분석을 수행할 수 있을 것으로 생각된다(Udurume et al., 2023).

본 연구의 목표는 뇌파와 시선추적 기술에 의해서 수집된 암묵적 적합성 피드백을 활용하고, 인공지능 기술에 의하여 미디어 콘텐츠 내용을 분석할 수 있는 디지털 도서관 인터페이

스를 설계하기 위한 지침을 제안하여 개인형 맞춤형 정보 서비스를 제공하는 데에 있다. 이를 위해서 첫째, 이용자의 암묵적 적합성 피드백을 활용하여 오디오-비주얼, 비주얼 및 텍스트 자료의 의미 분석을 수행할 뇌파 모형 및 시각 모형 기반 방법들을 제안하였다. 둘째, 제안된 방법들의 유용성을 평가하기 위해서 6개의 연구 문제들을 설정하고 이들을 검증하기 위해서 두번의 뇌파/시선추적 실험들을 수행하였다. 상기 평가에서 인공지능 기법들이 경쟁 기준으로 사용되었다. 셋째, 연구 문제들의 검증 결과와 이론적 근거에 기반하여, 뇌파, 시선추적 및 인공지능 기술에 기반한 디지털 도서관 인터페이스를 구현하기 위한 지침 즉, 암묵적 적합성 피드백에 기반한 디지털 도서관 시스템 모형을 제안하고, 적합성 판정의 분류 정확도를 향상시키기 위해서 미디어별로 적합한 뇌파 기반, 시선추적 기반 및 인공지능 기반 방법들을 제시하였다.

2. 선행 연구

적합성 판정, 감정 인식, 미디어 콘텐츠 분석과 요약, 시스템 인터페이스에 뇌파, 시선추적 및 인공지능 중 하나의 방법을 적용한 단일 모달(modal) 방식, 두 개 이상의 단일 모달 방식들을 결합하여 적용한 멀티 모달 방식에 대한 선행 연구들을 기술한다.

2.1 단일 모달

첫째, 뇌파 기반 방법을 적용한 연구들은 살

펴보면, Eugster et al. (2016)은 뇌파 실험에서 피험자들에게 위키피디아 문서를 읽도록 한 후 이때 나오는 뇌파 신호를 기반으로 명시적인 피드백 없이 암묵적 피드백으로 새로운 정보를 추천할 수 있는 뇌-관련성 패러다임을 제안하였다. 저자들은 뇌파 신호의 분석을 통해서 주제에 적합한 단어들이 부적합한 단어들에 비해 양전위 최대값이 더 높아지는 P600 효과가 있음을 확인했다. Schindler와 Bublatzky (2020)는 공포스런 얼굴은 무표정한 얼굴보다 EPN(early posterior negativity, 자극 제시 이후 200~350ms 구간에서 음전위 최저 뇌파값이 나타나는 컴포넌트)의 음전위 최소값이 더 낮아진다고 보고하였다.

둘째, 시선추적 정보를 적합성과 관련시키는 연구들을 살펴보면, Gwizdka와 Zhang(2015)은 피험자들이 관심을 갖는 웹페이지를 볼 때 관심을 갖지 않는 웹페이지를 볼 때보다 동공 크기(pupil size)가 확대된다는 것을 발견하고, 피험자들의 인지적 노력 차이를 분석하는 데에 동공 크기가 가장 중요한 예측 지표(metric)라고 주장하였다. Wang et al. (2019)은 시선추적을 이용하여 사람의 시각적 주의와 비디오의 주요 객체 판단 간에 강한 정적 상관관계가 있다고 주장하였다.

Bhattacharya et al. (2020)은 피험자들이 주제에 적합한 뉴스 기사를 판독한 경우와 부적합한 뉴스 기사를 판독한 경우를 비교할 경우 이들 간에 서로 다른 시선 경로 패턴을 보일 것이라고 가정하였다. 저자들은 피험자들이 뉴스 기사를 읽고 질문에 제시된 주제에 적합 또는 부적합이라고 평가하는 시선추적 실험을 수행하였다. 피험자의 시선 경로를 이미지로 인

코딩하여 이 이미지를 사용하여 합성곱 신경망(convolutional neural network, CNN) 분류기를 훈련하였다. 훈련된 분류기는 최대 80% 정확도로 뉴스 기사의 주제 관련성을 예측하였다. 윤정원과 신수연(2021)은 시선추적과 회상·인식 테스트를 이용하여, 페이스북의 건강 정보 게시물 형식이 대학생들의 페이스북 게시물 시선 주시 패턴과 인지 테스트 결과에 미치는 영향을 분석하였다. 저자들은 이용자들이 정보를 포함하고 있는 영역에 주의를 집중하며, 정보를 포함하지 않는 사진보다는 본문의 내용에 먼저 시선을 집중한다고 보고하였다.

셋째, Ul Haq et al. (2022)은 딥러닝을 활용하여 관심있는 객체(object of interest)를 탐지하는 비디오 요약 프레임워크를 제안했다. 제안된 프레임워크의 성능을 평가하기 위해 세 가지 유형의 데이터셋을 이용하여 실험한 결과, 모두 100%에 가까운 정확도를 얻었다. Borgalli와 Surve(2022)는 이미지(얼굴)에 나타나는 얼굴 가지의 감정을 인식하기 위해 맞춤형 CNN 아키텍처를 사용하였다. K겹 교차 검증(k-fold cross-validation) 방법에 의해서 세 개의 데이터셋을 이용하여 분석한 결과, 91.58%의 정확도를 얻었다.

2.2 멀티 모달

Gwizdka et al. (2017)은 뇌파 및 시선추적 기술을 사용하여 신문 기사를 읽고 적합성 판단을 하는 과정을 분석하였다. 저자들은 다양한 적합성 수준의 신문 기사를 평가하는 데 사용되는 인지 과정에 차이가 있으며, 이러한 차이는 뇌파와 시선추적을 통해 감지될 수 있다

고 주장했다. Golenia et al. (2018)은 뇌파와 시선추적의 조합을 통해 이미지 검색에서 암묵적 적합성 피드백의 활용 가능성을 조사하였다. 분석 결과, 뇌파와 시선추적 정보를 결합한 멀티 모달 방식의 정확도(85.9%)가 단일 모달 방식의 정확도(시선추적: 81.0%, 뇌파: 76.9%)보다 높게 나타났다.

Zheng et al. (2018)은 시선추적과 뇌파 정보가 상호 보완적으로 이용되어 감정 인지의 정확도를 높인다는 것을 증명하였다. 저자들은 피험자들에게 짧은 동영상들을 시청하게 한 후 다양한 시선추적 지표들과 정량화 뇌파(Quantitative EEG, QEEG)를 이용하여 네 가지 감정을 구분할 수 있는지 조사하였다. 분석 결과, 멀티 모달 방식을 사용한 경우의 분류 정확도(85.11%)가 단일 모달 방식을 사용한 경우의 분류 정확도(시선추적: 67.82%, 뇌파: 70.33%)보다 높게 나타났다. 더 나아가 뇌파 정보는 행복한 감정을 잘 구분해 주는 반면 시선추적 정보는 공포 감정을 잘 가려내 준다고 보고하였다.

Jacucci et al. (2019)은 시스템이 검색 결과에 대한 이용자의 신경생리학적 반응을 해석함으로써 암묵적 적합성 판단의 평가를 개선할 수 있는지 여부를 조사했다. 저자들은 뇌파 신호와 눈동자 움직임에서 실시간으로 적합성 피드백을 사용하였고, 그들의 접근 방식이 신경적응형 정보 검색 시스템의 설계를 위한 하나의 지침이 될 수 있다고 보고하였다. Udurume et al. (2023)은 얼굴에 나타난 감정을 식별하기 위해서 얼굴 인식과 음성 인식 그리고 QEEG를 사용했다. 분석 결과, 멀티 모달 방식의 정확도(80.8%)가 각 단일 모달 방식의 정확도(얼굴 인식: 78%, 음성 인식: 74.5%, 뇌파: 69.5%)

보다 높게 나타났다.

본 연구는 상기 선행 연구들을 분석한 후 뇌파 및 시선추적 기반 방법들을 제안하고, 이들 방법들과 인공지능 기반 방법들을 비교분석한 후 암묵적 적합성 피드백에 기반한 디지털 도서관 인터페이스를 구현하기 위한 지침을 제시하였다.

3. 미디어 콘텐츠의 의미 분석 방법

동영상 내용에 대한 의미추론과정은 시각 채널을 통해 유입되는 시각 정보와 청각 채널을 통해 유입되는 청각 정보의 통합과정이며 이는 작업기억 내에서 이루어진다(Mayer, 2005). 이 통합과정에서 새로운 감각정보를 의미적 추론의 대상으로 삼기 위해서는 장기기억에 저장된 지식의 도움이 필요하다. 오디오-비주얼 시맨틱스는 시각 정보와 청각 정보로 구성된 오디오-비주얼 콘텐츠의 의미와 중요성을 파악하는 분야이다(Foley & Kwan, 2015). 한편 비주얼 시맨틱스는 이미지의 비주얼 요소(예, 객체, 인물)와 캡션의 의미와 중요성을 다루고, 텍스트 시맨틱스는 텍스트의 의미와 중요성을 다룬다. 본 연구는 이러한 오디오-비주얼, 비주얼 및 텍스트의 의미 분석을 위해서 뇌파 및 시각 모형 기반 방법들을 제안하였다. 뇌파(EEG)는 외부 자극에 의해서 나타나는 전위 변화를 다룬 사건관련유발전위(event related potentials, ERP)와 뇌파를 정량적인 수치로 표현한 정량화 뇌파(QEEG)로 구분된다. 본 연구는 실험 데이터로 이미지, 텍스트와 함께 객체를 주제로 한 다큐멘

터리 영상물을 다루기 때문에 ERP가 QEEG보다 더 적합하다고 판단되어 뇌파 및 시각 모형을 ERP의 P600, P3b 컴포넌트들을 이용하여 구성하였다.

3.1 뇌파 및 시각 모형 기반 방법 제안

이미지 정보를 인지할 때 어떤 ERP 컴포넌트를 사용하는지에 따라서 활성화되는 두뇌 영역에 달라진다고 보고되고 있다. 예를 들어, P3a 컴포넌트(300ms 근처에서 양전위 최고값을 나타내는 컴포넌트)를 사용할 때는 전두엽, P3b 컴포넌트(400ms 근처에서 양전위 최고값을 나타내는 컴포넌트)를 사용할 때는 두정엽과 측두엽, 그리고 P600 컴포넌트(600ms 근처에서 양전위 최고값을 나타내는 컴포넌트)를 사용할 때는 중앙엽이 활성화된다고 알려졌다(Allegretti et al., 2015; Davis, Spapé, & Ruotsalo, 2022). 본 연구는 이 중에서 두 개의 두뇌 영역이 활성화되는 P3b 컴포넌트를 이미지(객체)의 적합성 측정에 사용하였고, 이를 o_P3b(object P3b)로 지칭하였다.

한편 텍스트의 적합성 측정에 P600 컴포넌트를 적용하는 연구들(Eugster et al., 2016; Ye et al., 2022)과 P3b 컴포넌트를 적용하는 연구(Eugster et al., 2014)가 있다. 실험 데이터를 사전에 분석한 결과, P3b 효과가 P600 효과보다 더 크게 나타나 P3b를 채택하고, 이를 t_P3b(text P3b)로 지칭하였다. 이와 같이 P3b 효과가 더 크다는 것은 실험 데이터가 구조화 텍스트이기 때문일 수 있다. 즉, 신문 제목의 두서너 개의 단어들(0.5초 동안, 대략 1.65~2.5 개 단어를 관독할 수 있음)에서 주제를 파악할

수 있을 것으로 생각된다.

주제에 적합한 오디오-비주얼 정보를 포함하고 있는 동영상 시청할 때 반응 시점이 이미지보다 늦어질 것으로 가정하여 P600 컴포넌트를 사용하였고(Kim & Kim, 2019a; 2019b), 이를 av_P600(audio-visual P600)로 지칭하였다. 본 연구는 감정 식별에 대한 기존 뇌파 연구들의 결과가 다양하여 일률적인 패턴을 보이지 않아서 이미지(얼굴/감정)의 적합성 판정에 P3b 컴포넌트를 이용하였고, 이를 f_P3b(face P3b)로 지칭하였다.

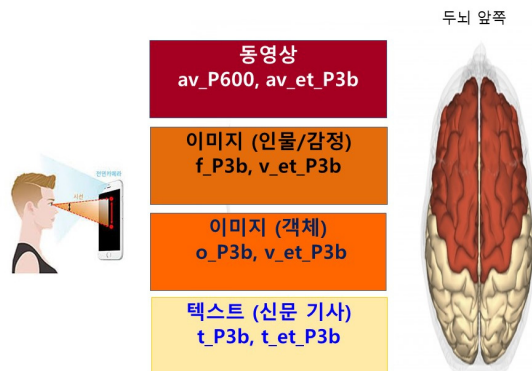
시선추적 분석에 사용되는 지표에는 고정(fixation), 도약(saccade), 동공 크기 등이 있고 이 중에서 동공 크기는 적합성 판정에 가장 중요한 지표로 알려져 있다. 따라서, 동영상, 이미지 및 텍스트 자극에 대한 시선추적 기반 방법을 구현하기 위해서 동공 크기를 이용하였다. 구체적으로, 뇌파에 사용된 ERP의 P3b 컴포넌트를 시선추적에 적용하여 et_P3b(eye tracking P3b) 컴포넌트로 지칭하고 이를 400ms 근처에서 양전위 최고 동공 크기값을 나타내는 컴포넌트로 사용하였다. 이 컴포넌트를 동영상, 이

미지, 텍스트에 적용한 경우 각각 av_et_P3b, v_et_P3b(visual et_P3b), t_et_P3b로 지칭하였다.

미디어 종류에 따라서 두뇌 활성화 영역이 달라질 수 있다. 일반적으로 멀티미디어 즉, 동영상, 이미지에 의한 자극은 두뇌의 앞쪽(예, 전두엽)이 활성화되는 반면, 언어(텍스트)에 의한 자극은 두뇌의 뒤쪽(예, 두정엽)이 활성화된다고 알려져 있다(DeLong, Quante, & Kutas, 2014; Kim & Kim, 2019a; 2019b)(〈그림 1〉참조).

3.1.1 뇌파 모형 기반 방법

상기에서 언급한 컴포넌트들에서 중요한 채널들을 선정하기 위해서 본 연구에서 수행한 뇌파/시선추적 실험들의 데이터를 이용하였고, 이때 사용된 유의 수준은 0.01이다. 이후 SPSS 23의 인공 신경망(Artificial Neural Network, ANN)을 활용하여 컴포넌트 분석에 의해서 선정된 채널들의 가중치를 구한 후 뇌파 모형을 구성하였다.



〈그림 1〉 미디어별로 활성화되는 두뇌 영역과 사용된 ERP 컴포넌트들

1) av_P600 컴포넌트를 활용한 모형

(1) 채널 선정: 동영상의 경우, 주제 적합 양전위 최고점의 평균값과 주제 부적합 양전위 최고점의 평균값 간에 유의미한 차이가 나는 채널은 총 6개(전두엽: Fz, F3, 전두-중앙엽: FCz, FC3, FC4; 중앙엽: Cz)로 나타났다. 이에 따라서, 동영상에 의한 자극은 두뇌의 앞쪽이 활성화된다는 이론은 입증되었다.

(2) av_P600_uV 모형: 실험 데이터를 인공 신경망(SPSS 사용)으로 분석하여 상기 6개 채널의 가중치를 구한 후 공식을 구성하였다. 인공 신경망에 의한 여러 분석 결과들 중에서 학습 및 검정 데이터의 평균 분류 정확도가 높은 경우 즉, 77.0%(학습 데이터), 90%(검정 데이터)를 선정하였다. 이때 총 345 사례(주제 적합: 196, 주제무관: 149) 중 학습으로 311개(90%)가 사용되었고, 34개(10%)가 검정에 사용되었다. 이후, 아래와 같이 상기 학습 및 검정 데이터의 분류 결과를 산출할 때 사용된 6개 채널의 가중치에 기초하여 공식 1을 구성하였다.

$$av_P600_uV = \sum_{i=1}^n A(CH_i) \times W(CH_i) \quad (1)$$

여기서, A(CH_i)는 채널 CH_i의 뇌파값(ERP 진폭값)이고, W(CH_i)는 채널 CH_i의 가중치이다. av_P600_uV(공식 1)는 여섯 개 채널 각각의 A값과 W값을 곱한 후 이를 합산하여 계

산된다(<표 1> 참조).

2) P3b 컴포넌트를 활용한 모형

(1) 채널 선정: 이미지(얼굴)의 경우, 주제 적합 양전위 최고점의 평균값과 주제 부적합 양전위 최고점의 평균값 간에 유의미한 차이가 나는 채널은 총 4개(전두엽: f8; 두정엽: p8; 측두엽: t8; 측두-두정엽: tp8)로 나타났다. 이미지(객체)의 경우, 적합 양전위 최고점의 평균값과 부적합 양전위 최고점의 평균값 간에 유의미한 차이가 나는 채널은 총 7개(전전두엽: FP1; 전두엽: F3; 중앙엽: Cz, C3; 두정엽: P3; 전두-중앙엽: FC3; 중앙-두정엽: CP3)로 나타났다. 한편 텍스트의 경우, 적합 양전위 최고점의 평균값과 부적합 양전위 최고점의 평균값 간에 유의미한 차이가 나는 채널은 총 7개(두정엽: Pz, P7; 중앙엽: C3; 측두엽: T7, 전두-측두엽: FT7; 중앙-두정엽: CP3; 측두-두정엽: TP7)로 나타났다. 이미지는 두뇌의 앞쪽을 활성화시킨다는 예측과 달리 두뇌 전체가 활성화되는 경향을 보였고, 텍스트는 예측한 대로 두뇌의 뒤쪽을 활성화시키고 있었다.

(2) f_P3b_uV 모형, o_P3b_uV 모형 및 t_P3b_uV 모형: 세 개의 모형들을 구성하기 위해서 상기 P600_uV 모형과 동일한 방식을 사용하였다. 구체적으로, 세 개의 컴포넌트들(f_P3b, o_P3b, t_P3b)에서 중요한 채널들을 각각 선정하기 위해서 실험 데이터를 이용하였고, 인공

<표 1> 채널별 가중치(av_P600_uV)

av_P600_uV(동영상)(공식 1)											
채널	가중치	채널	가중치	채널	가중치	채널	가중치	채널	가중치	채널	가중치
Fz	0.20	F3	0.13	FCz	0.11	FC3	0.21	FC4	0.19	Cz	0.15

신경망을 활용하여 각 컴포넌트 분석에 의해서 선정된 채널들의 가중치를 구한 후 뇌파 모형들을 다음과 같이 구성하였다(공식 2-1~2-3).

$$f_{P3b_uV} = o_{P3b_uV} = t_{P3b_uV} = \sum_{i=1}^n A(CH_i) \times W(CH_i) \quad (2)$$

여기서, f_{P3b_uV} 는 네 개 채널 각각의 A값과 W값을 이용하여 계산되고, o_{P3b_uV} 와 t_{P3b_uV} 는 일곱 개 채널 각각의 A값과 W값을 이용하여 계산된다(〈표 2〉 참조).

3.1.2 시각 모형 기반 방법

동영상, 이미지 및 텍스트 자극에 대한 시각 모형 기반 방법을 구현하기 위해서 동공 크기를 이용하였다. 구체적으로, 주제 적합 자극과 주제 비적합 자극 간의 동공 크기 차이를 분석하기 위해서 et_{P3b} 컴포넌트를 사용하였다(4.1.3장 참조).

3.2 인공지능 기반 방법

본 연구는 인공지능 기술을 구현하기 위해서

Amazon Rekognition을 이용하였고, 자연언어 분석을 위해서 Amazon Comprehend를 이용하였다. OpenAI의 ChatGPT-4는 Amazon 소프트웨어의 분석 결과와 비교하기 위해서 이용하였다. 인공지능 기술을 활용하여 동영상, 이미지 및 텍스트를 분석하는 방법을 기술하면 다음과 같다.

첫째, 각 동영상을 분석하기 위해서 객체-텍스트 기반 방법을 채택하였다. Amazon Rekognition을 이용하여 동영상에서 주제와 관련된 객체가 있는지 파악한다. 만약 동영상이 객체를 포함하고 있다면 객체가 포함된 동영상의 키프레임(이미지)(들)을 추출하여 해당 키프레임의 객체에 대한 신뢰도 점수를 계산하였다. 한편 Amazon Comprehend를 이용하여 검색 질문(〈표 3〉)과 제시되는 동영상의 텍스트로 변환된 내레이션에서 각각 핵심 구절들을 추출한 후, 이들 핵심 구절들이 서로 얼마나 일치되는지를 측정하여 각 내레이션의 적합도 점수를 계산하였다. 이후 신뢰도 점수와 적합도 점수를 합하여 각 동영상의 종합 점수를 구하고 이 종합 점수들을 내림차순으로 정렬한 후, 원하는 수만큼의

〈표 2〉 채널별 가중치(P3b_uV)

f_P3b_uV(이미지/얼굴)(공식 2-1)							
채널	가중치	채널	가중치	채널	가중치	채널	가중치
F8	0.33	P8	0.26	T8	0.25	TP8	0.16
o_P3b_uV(이미지/객체)(공식 2-2)							
채널	가중치	채널	가중치	채널	가중치	채널	가중치
FP1	0.38	F3	0.09	C3	0.08	Cz	0.07
P3	0.04	FC3	0.30	CP3	0.03		
t_P3b_uV(텍스트)(공식 2-3)							
채널	가중치	채널	가중치	채널	가중치	채널	가중치
Pz	0.10	P7	0.16	C3	0.19	T7	0.16
FT7	0.20	CP3	0.10	TP7	0.09		

동영상을 선택하였다.

둘째, 이미지의 객체 인식을 위해 Amazon Rekognition을 이용하여 각 이미지(객체)에 대한 신뢰도 점수를 계산하고, 이미지의 얼굴/감정 분석을 위해 Amazon Rekognition(Facial Analysis)을 이용하여 각 이미지에 대한 성별, 감정 등에 대한 신뢰도 점수를 측정한다. 이후 원하는 수만큼의 이미지를 각각 선택하였다.

셋째, 신문 기사의 자연언어 분석을 위해서 Amazon Comprehend를 이용하여 상기에서 기술한 동영상의 내레이션과 같은 방식으로 분석하였다.

4. 주제 적합성 판정을 위한 분류 정확도 측정과 평가

4.1 뇌파와 시선추적 실험

4.1.1 피험자, 실험 데이터 및 시스템

뇌파는 성별, 나이, 교육 수준에 따라서 차이가 있다고 알려졌다(Evans, Cui, & Starr, 1995; Pöld et al., 2018). 두 번에 걸쳐서 뇌파/시선추

적 실험들을 진행하였다(IRB 승인번호: IRB 승인번호: MJU-2020-06-006-03). 1차 실험에서 사용한 실험 데이터는 <표 3>의 1~2번(동영상)이고 20~30대의 12명의 피험자들(학부생: 10명, 대학원생: 2명)을 대상으로 하였다. 2차 실험에서 사용한 실험 데이터는 1~6번(동영상, 이미지 및 텍스트)이며 20~40대의 20명의 피험자들(학부생: 16명, 대학원생: 2명, 연구원: 2명)을 대상으로 하였다. 연구 문제들을 검증하기 위해서 이 두 실험 결과의 데이터는 통합해서 분석하였다.

실험 데이터로 동영상 16개(재생시간: 5.1초~10.5초), 이미지(얼굴) 8개, 이미지(객체) 4개, 신문 기사(텍스트) 4개를 이용하여 총 여섯 개의 질문에 적합한 자료(들)를 선택하도록 하였다(<표 3> 참조). 예를 들어, 텍스트의 경우, 4개의 신문 기사의 제목과 기사 본문(요약)들을 피험자들에게 보여준 후 질문 주제에 적합한 신문 기사(들)를 선택하도록 하였다.

실험 장비는 뉴로 스캔의 뇌파기기(Neuroscan Synamp Amplifier)와 토비테크놀로지의 시선추적 장치(Tobii Pro Nano)를 이용하였다. 뇌파 및 시선추적 데이터는 Curry 7.0, Tobii Pro Lab 및 SPSS 23을 이용하여 분석하였다.

<표 3> 실험 데이터

No.	데이터 유형/전체 자료수/적합 자료수	질문
1	동영상 1/8개/3개	신라의 상징인 금관이나 범(절과 탐으로 표현)을 다룬 동영상(들)을 검색한다.
2	동영상 2/8개/3개	서로 뒤바뀐 김구와 윤봉길의 회중 시계를 다룬 동영상(들)을 검색한다.
3	이미지(남녀 얼굴) 1/4개/2개	놀란 표정의 얼굴 이미지(들)를 검색한다.
4	이미지(여자 얼굴) 2/4개/2개	기쁜 표정의 여자 얼굴 이미지(들)를 검색한다.
5	이미지(객체)/4개/1개	금관 이미지(들)를 검색한다.
6	텍스트(신문 기사)/4개/1개	미국과 중국의 인공지능 연구의 강점을 비교하는 신문 기사(들)를 검색한다.

4.1.2 실험 절차

뇌파/시선추적 실험 절차를 설명하면 다음과 같다(〈그림 2〉 참조). 피험자에게 실험 내용을 설명하였다. 이후 피험자의 눈 위치를 교정하는 캘리브레이션(calibration) 과정을 거친 후 사전 테스트를 수행하여 피험자에게 실험 절차를 숙지시켰다. 1차 실험에서는 피험자에게 동영상, 2차 실험에서는 동영상, 이미지 및 텍스트를 제시한 후 각 자극에 대해서 질문에서 요구하는 주제와 적합한지 적합성 판정(주제 비적합: 1, 주제 부분적합: 2, 주제 적합: 3)을 하도록 하였다.

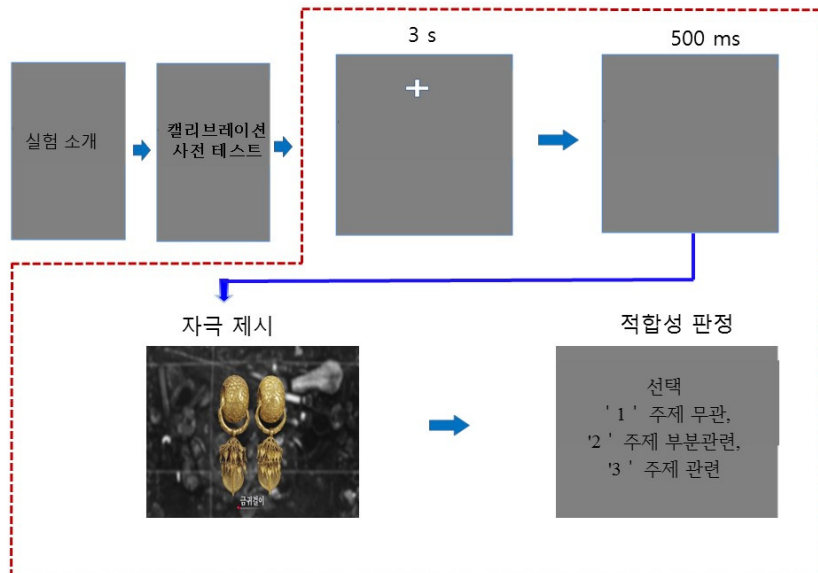
4.1.3 뇌파 및 시각 반응 측정과 분석

기저값 설정은 자극의 시작점을 찾아서 -200ms에서 시작점까지의 평균 뇌파값을 뇌파 기저값으로 하였다. 한편 자극 제시 이전에 보여주는

회색 스크린을 보는 200ms 동안의 평균 동공 크기를 동공 크기 기저값으로 사용하였다. 시작점으로부터 1,000ms에 대한 뇌파 및 동공 크기 데이터를 각 자극에 대한 뇌파 및 시각 반응으로 보고 다음과 같이 측정된 후 분석하였다.

1) 동영상

동영상(뇌파)의 경우, 총 에폭 512개(32명 X 16개, 피험자당 16개 에폭들)는 주제 적합 196개, 주제 부분적합 167개, 주제 비적합 149개로 구분되었다. 적합성 판정을 위해서 주제 부분적합 에폭들은 애매성 때문에 제외시키고 345개의 에폭들만 이용하였다. 이러한 345개 에폭 데이터를 입력 데이터로 하여 CURRY 7.0 프로그램과 SPSS 23을 이용하여 각 피험자의 동영상별로 양전위 최고 뇌파값(av_P600)을 계산하였다. 동공 크기의 경우, 데이터는 Tobii



각 자극별 빨간 점선으로 표시된 박스 반복

〈그림 2〉 실험 절차

Pro Lab 프로그램에 의해서 수집되었고, 수집된 데이터를 이용하여 각 동영상의 동공 크기(et_P3b)를 측정하였다. 구체적으로 피험자가 각 동영상을 보는 처음 0.2초 ~ 0.5초 동안의 동공 크기를 네 개의 단위(0.2 ~ 0.5)로 각각 분석하였다. 각 단위의 동공 크기는 동공 크기 기저값과의 차이로 계산한 표준값을 이용하였다. 이 표준값을 이용하여 각 피험자의 동영상 별로 av_et_P3b값을 계산하였다. 예를 들어, 피험자의 기저값이 2.92mm이고 0.2초 단위의 동공 크기의 평균값이 3.4mm이면 0.2초 단위의 표준값은 0.48mm(=3.4-2.92)이 된다. 이와 동일한 방식으로 나머지 단위들(0.3초 ~ 0.5초)의 표준값들(0.51mm, 0.53mm, 0.52mm)을 계산한다. 이후 이 네 가지 표준값들 중에서 가장 값이 큰 0.53mm가 av_et_P3b값이 된다.

2) 이미지와 텍스트

이미지와 텍스트의 뇌파 분석은 상기에 기술한 방법과 동일하게 분석하였다. 단 피험자들이 동영상과 달리 적합성 판정에서 주제 부분적합을 선택하지 않았다. 따라서, 이미지(객체)인 경우, 총 예폭 80개(피험자당 4개 예폭들)는 주제 적합 20개와 주제 비적합 60개로 구분되었다. 이미지(얼굴/감정)인 경우, 총 예폭 160개(피험자당 8개 예폭들)는 주제 적합 60개와 주제 비적합 100개로 구분되었다. 텍스트의 경우, 총 예폭 80개는 주제 적합 20개와 주제 비적합 60개로 구분되었다. 이러한 예폭 데이터를 입력 데이터로 하여 P3b 컴포넌트를 이용하여 각 피험자의 양전위 최고 뇌파값을 계산하였다. 동공 크기의 경우도 상기에 기술한 방법과 동일하게 동공 크기(et_P3b)를 측정하여 분석하였다.

신문 기사에서는 적합성 판정을 위해서 동공 크기와 함께 시선 경로 패턴도 사용하였다. 구체적으로, 피험자들이 신문 제목을 보고 적합한 자료로 판단되면 기사 본문을 자세히 살펴볼 것으로 예상되어 피험자 간에 유사한 시선 경로 패턴이 나타날 것으로 가정하였다. 한편 비적합한 자료를 본 경우, 기사 본문을 읽지 않거나 단지 훑어볼 것으로 예상되며, 이로 인해 피험자 간에 서로 다른 시선 경로 패턴이 나타날 것으로 가정하였다. 시선의 움직임을 시각화하기 위해서 시선의 고정과 도약이 이루어지는 위치와 도약 경로의 순서를 공간적으로 보여주는 시선 플롯(gaze plot)을 활용하였다. 개인 간의 시선 플롯 이미지의 유사도는 소프트웨어(Duplicate Photo Cleaner)로 측정하여 시선 경로 패턴을 분석하였다.

4.2 연구 문제 설정과 검증

논문의 후반부에 기술할 디지털 도서관 인터페이스 지침에서 적합성 판정의 분류 정확도를 향상시키기 위해 미디어별로 적합한 미디어 콘텐츠 의미 분석 방법(예, 뇌파, 시선추적 및 인공지능 기반 방법)을 제안하고자 한다. 이를 위해, 다음에 기술한 여섯 가지 연구 문제들을 설정하고 검증하였다. 연구 문제들을 검증하기 위해, 제안된 각 방법의 분류 정확도와 방법들 간의 상관 관계를 분석하였다. 분류 정확도를 계산하는 과정을 설명하면 다음과 같다. 자극의 주제 관련성을 분석하기 위해, 자극에 의해 유발된 피험자들(동영상: 32명, 이미지/텍스트: 20명)의 뇌파 및 시선추적 데이터를 분석하였다. 뇌파 기반 방법은 네 개의 모형들(av_P600_uV,

f_P3b_uV, o_p3b_uV, t_p3b_uV)을 활용한 뇌파 모형(EEG model) 기반 방법과 30개 채널의 뇌파값들(av_P600, f_P3b, o_p3b, t_p3b)을 모두 활용한 뇌파 컴포넌트(EEG component) 기반 방법으로 구분하여 분석하였다. 또한 시선추적 기반 방법은 시각 모형(et_P3b)을 활용한 시각 모형 기반 방법과 시선 경로 패턴 기반 방법이 사용되었다(〈그림 3〉 참조).



〈그림 3〉 뇌파 및 시선추적 기반 방법

다음은 SPSS의 인공 신경망(artificial neural network, ANN)을 이용하여 실험에서 얻은 뇌파 및 시선추적 데이터를 10겹 교차 검증으로 분석하여 분류 정확도를 계산하였다. 구체적으로 동영상의 경우 모든 데이터(345 케이스)를 무작위로 10개의 동일한 크기의 데이터셋으로 나누고, 9개의 데이터셋에서 학습하고, 1개의 데이터셋에서 테스트를 진행하였다. 이 절차를 분류 정확도의 평균을 얻기 위해서 10번 반복하였다. 동영상과 유사한 방법으로 이미지와 텍스트의 분류 정확도를 측정하였다. 한편 인공지능 기반 방법에 의한 분류 정확도도 측정하였다(3.2장 참조).

4.2.1 연구 문제 1: 동영상 자료를 대상으로 한 경우 신경과학 연구방법(뇌파와 시각 모형)과 딥러닝을 활용한 객체-텍스트 인식 기반 방법 간에 어떤 차이가 있는가?

첫째, 뇌파 기반, 시각 모형, 객체-텍스트 인식 기반 방법들의 적합성 관정에 대한 분류 정확도를 비교해 보았다. 모든 채널들을 사용한 뇌파 컴포넌트 기반 방법(av_P600[all CHs])과 시각 모형 기반 방법(av_et_P3b)을 결합한 멀티모달 방식의 분류 정확도(78.12%)가 가장 높게 나타났다. 이어서 뇌파 컴포넌트 기반 방법의 정확도(75.14%), 시각 모형 기반 방법의 정확도(69.69%), 뇌파 모형 기반 방법(av_P600_uV)의 정확도(67.63%), 객체-텍스트 인식 기반 방법의 정확도(66.7%)순으로 나타났다(〈표 4〉 참조). 객체-텍스트 인식 방법의 정확도가 상대적으로 낮게 나타난 이유는 〈표 3〉의 질문 1의 경우, 주제에 적합한 동영상 1이 핵심적인 주제인 신라의 개혁, 즉 법을 내레이션에서 다루고 있음에도 불구하고, 주제와 관련된 객체(예, 절, 탑)의 이미지를 포함하지 않아서 객체-텍스트 인식 방법에 의해서 선정되지 않았기 때문이다(3.2장 참조).

둘째, 신경과학 연구방법과 객체-텍스트 인식 기반 방법 간의 상관 관계를 살펴보면 다음과 같다. 시각 모형 기반 방법이 주제와 관련된 객체를 포함하는 동영상에 높은 가중치를 주는 경향이 있음을 발견하고 시각 모형 및 객체-텍스트 인식 기반 방법이 객체를 통해 상호 연관될 수 있다고 가정하였다. 이들 간의 관계를 좀 더 자세히 살펴보기 위해서 시각 모형 기반 방법, 뇌파 기반 방법, 동영상에 포함된 세 가지 시각 요소(객체, 자막, 배경) 간의 상관관계를

조사했다. 예를 들어, 각 동영상을 네 가지 태그 (객체, 자막, 배경, 없음)를 사용하여 분류하였다. 만약 동영상이 객체와 자막을 포함하고 있다면, 해당 동영상에는 두 가지 태그 즉, 객체와 자막이 할당된다.

상관관계 분석 결과, 시각 모형 기반 방법과 객체 간에 정적 상관관계를 발견했다($r=0.217$). 이는 동영상에 주제 관련 객체가 포함되어 있으면 해당 동영상이 높은 av_et_P3b값을 가질 가능성이 높다는 것을 의미한다. 이외에 객체와 자막 간에도 정적 상관관계($r=0.362$)가 있었다. 반면에, 시각 모형 기반 방법과 배경 간에는 부적 상관관계가 있었다($r=-0.223$). 이는 동영상이 배경 정보를 포함하고 있다면 해당 동영상이 낮은 av_et_P3b값을 가질 가능성이 높다는 것을 의미한다. 한편 뇌파 기반 방법과 상기에 서 기술한 세 가지 시각 요소 간에 유의미한 정적 또는 부적 상관관계는 없는 것으로 나타났다.

4.2.2 연구 문제 2: 동영상 자료를 대상으로 한 경우 뇌파 기반 방법의 분류 정확도와 시각 모형 기반 방법의 분류 정확도 간에 어떤 차이가 있는가?

뇌파 기반 방법과 시각 모형 기반 방법의 적합성 판정에 대한 분류 정확도를 비교해 본 결과, 시각 모형 기반 방법의 분류 정확도(69.69%)가

뇌파 컴포넌트 기반 방법의 정확도(75.14%)보다 낮게 나타났지만, 뇌파 모형 기반 방법의 정확도(67.63%)보다 다소 높게 나타났다(〈표 4〉 참조). 이와 같이 뇌파 컴포넌트 기반 방법이 높은 분류 정확도를 나타낸 이유는 30개 채널의 뇌파값들을 활용하여 더 많은 정보를 분석했기 때문으로 생각된다.

4.2.3 연구 문제 3: 이미지(얼굴/감정) 자료를 대상으로 한 경우 신경과학 연구방법(뇌파와 시각 모형)의 분류 정확도와 딥러닝 기반 얼굴 인식 기반 방법의 분류 정확도 간에 어떤 차이가 있는가?

뇌파 컴포넌트/시각 모형 기반 방법의 분류 정확도(90.09%)가 가장 높게 나타났다. 이어서 뇌파 컴포넌트 기반 방법의 정확도(85.02%), 얼굴 인식 기반 방법의 정확도(75%), 뇌파 모형 기반 방법의 정확도(71.48%), 시각 모형 기반 방법의 정확도(60.97%)순으로 나타났다(〈표 5〉 참조). 얼굴 인식 기반 방법은 기쁨, 무표정은 대체로 정확하게 인식하는 반면 놀람과 공포는 뚜렷하게 구분하지 못하는 경우가 있었다. 〈표 3〉의 질문 2(놀란 표정의 얼굴)의 경우 Amazon Rekognition은 두 개의 놀란 표정의 얼굴들 중 하나는 놀람으로 다른 하나는 공포로 인지하였으나, ChatGPT-4에서는 이 두 이미지들을 모

〈표 4〉 분류 정확도(동영상)

방법	분류 정확도(%)
뇌파 컴포넌트(av_P600)(all CHs)	75.14
뇌파 모형(av_P600_uV)	67.63
시각 모형(av_et_P3b)	69.69
뇌파 컴포넌트/시각 모형	78.12
객체-텍스트 인식	66.67

〈표 5〉 분류 정확도(이미지/[얼굴/감정])

방법	분류 정확도(%)
뇌파 컴포넌트(f_P3b)(all CHs)	85.02
뇌파 모형(f_P3b_uV)	71.48
시각 모형(v_et_P3b)	60.97
뇌파 컴포넌트/시각 모형	90.09
얼굴 인식	75

두 놀람으로 분류하였다.

4.2.4 연구 문제 4: 이미지(객체) 자료를 대상으로 한 경우 신경과학 연구방법(뇌파와 시각 모형)의 분류 정확도와 딥러닝 기반 객체 인식 기반 방법의 분류 정확도 간에 어떤 차이가 있는가?

객체 인식 기반 방법의 분류 정확도(100%)가 가장 높게 나타났고, 이어서 뇌파 컴포넌트/시각 모형 기반 방법의 정확도(89.15%), 시각 모형 기반 방법의 정확도(87.20%), 뇌파 컴포

넌트 기반 방법의 정확도(86.33%), 뇌파 모형 기반 방법의 정확도(74.50%)순으로 나타났다(〈표 6〉 참조).

4.2.5 연구 문제 5: 텍스트 자료(신문 기사)를 대상으로 한 경우 신경과학 연구방법(뇌파와 시선추적)의 분류 정확도와 딥러닝 기반 자연언어 처리 방법의 분류 정확도 간에 어떤 차이가 있는가?

자연언어 처리 방법의 분류 정확도(100%)가 가장 높게 나타났다(〈표 7〉 참조). 뇌파 기

〈표 6〉 분류 정확도(이미지/객체)

방법	분류 정확도(%)
뇌파 컴포넌트(o_P3b)(all CHs)	86.33
뇌파 모형(o_P3b_uV)	74.50
시각 모형(v_et_P3b)	87.20
뇌파 컴포넌트/시각 모형	89.15
객체 인식	100

〈표 7〉 분류 정확도(텍스트)

방법	분류 정확도(%)
뇌파 컴포넌트(t_P3b)(all CHs)	86.67
뇌파 모형(t_P3b_uV)	85.86
시각 모형(t_et_P3b)	61.28
시선 경로 패턴	76.57
뇌파 컴포넌트/시각 모형	92.32
뇌파 컴포넌트/시선 경로 패턴	89.68
자연언어 처리	100

〈표 8〉 분류 정확도(단일 모달 대 멀티 모달)

미디어	분류 정확도(%) (단일 모달 대 멀티 모달)
동영상	70.82 대 78.12
이미지(얼굴/감정)	72.49 대 90.09
이미지(객체)	82.68 대 89.15
텍스트(신문 기사)	77.6 대 91

반 방법의 평균 분류 정확도(86.27%)가 시선추적 기반 방법의 평균 분류 정확도(68.93%)보다 높게 나타났다. 시선 경로 패턴 기반 방법의 분류 정확도(76.57%)가 시각 모형 기반 방법의 정확도(61.28%)보다 높게 나타났으나, 뇌파 컴포넌트 기반 방법과 결합시에는 뇌파 컴포넌트/시각 모형 기반 방법의 정확도(92.32%)가 뇌파 컴포넌트/시선 경로 패턴 기반 방법의 정확도(89.68%)보다 높게 나타났다.

4.2.6 연구 문제 6: 신경과학 연구방법(뇌파와 시선추적)에서 미디어별 단일 모달 방식과 멀티 모달 방식 간의 분류 정확도에는 어떤 차이가 있는가?

동영상과 이미지에서 멀티 모달 방식(뇌파 컴포넌트/시각 모형)의 분류 정확도가 단일 모달 방식들(뇌파 기반 및 시각모형)의 평균 분류 정확도보다 높게 나타났다. 텍스트에서도 멀티 모달 방식들(뇌파 컴포넌트/시선추적)의 평균 분류 정확도가 단일 모달 방식들(뇌파 기반 및 시선추적)의 평균 분류 정확도보다 높게 나타났다(〈표 8〉 참조). 특히 이미지(얼굴/감정)와 텍스트 자료에서 멀티 모달과 단일 모달 간에 분류 정확도 차이가 크게 나타났다(이미지[얼굴/감정]: 17.6%, 텍스트: 13.4%).

5. 뇌파, 시선추적 및 인공지능을 활용한 디지털 도서관의 인터페이스 설계를 위한 지침

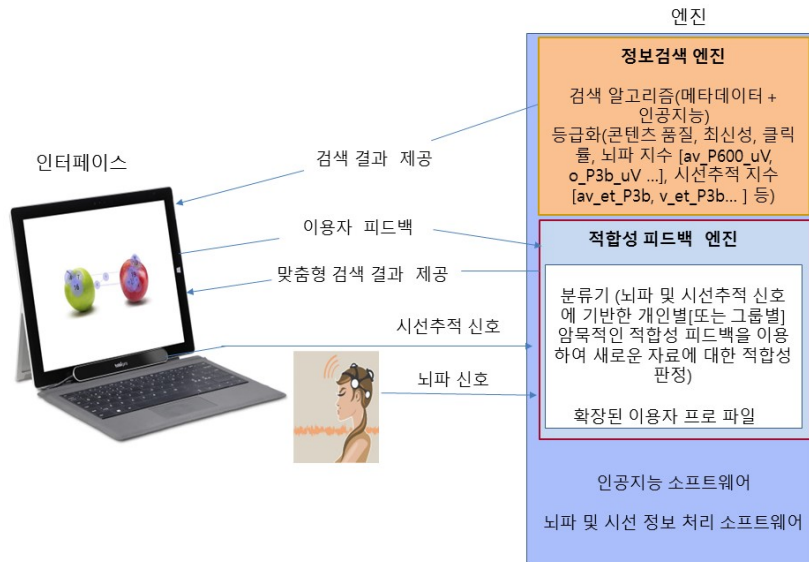
연구 문제들의 검증 결과와 이론적 근거를 기반으로 개인별 또는 그룹별 맞춤형 정보 검색 및 서비스를 가능하게 하는 디지털 도서관 인터페이스를 구현하기 위한 지침을 제안한다. 구체적으로, 암묵적 적합성 피드백에 기반한 디지털 도서관 시스템 모형을 제안하고, 적합성 판정의 분류 정확도를 향상시키는 위해서 미디어별로 적합한 뇌파 기반, 시선추적 기반 및 인공지능 기반 방법들을 제시한다.

5.1 디지털 도서관 시스템 모형

디지털 도서관 시스템 모형은 인터페이스, 뇌파 장치, 시선추적 장치, 정보검색 엔진 및 적합성 피드백 엔진으로 구성된다(〈그림 4〉 참조). 뇌파 장치는 웨어러블 헤드셋 또는 귀 착용형 뇌파 기기에서 뇌파를 용이하게 측정할 수 있도록 하고, 시선추적 장치는 거치형을 활용한다.

5.1.1 정보검색 엔진

디지털 도서관은 이용자의 검색 질문에 대해



〈그림 4〉 시스템 모형

여 정보검색 엔진을 이용하여 검색한 후 검색 결과를 이용자에게 제공한다(〈그림 4〉 참조). 이를 위해서 시스템에 있는 메타 데이터와 인공지능에 기반한 검색 알고리즘을 활용한다. 예를 들어, 특정 객체에 대한 이미지 자료를 검색하는 경우 데이터 베이스의 메타 데이터(예, 주제, 년도 등)와 인공지능(예, 객체 인식) 기반 방법을 사용하여 검색한 후 그 결과를 적합도순으로 정렬한 후 제시한다. 적합도를 계산할 때 사용하는 기준에는 내부 정보(콘텐츠의 품질, 최신성 등)와 외부 정보(클릭률, 뇌파 지수, 시선추적 지수 등)가 있다.

이때 사용하는 뇌파 및 시선추적 지수는 일종의 클릭률과 같은 것으로 클릭률은 대리 지표인 반면 뇌파 및 시선추적은 직접 지표로 사용된다. 뇌파 지수로 이용자들이 특정 이미지(객체) 자료에 대해서 반응하여 나타내는 뇌파 모형값(o_P3b_Uv)의 평균을 사용한다. 여기

서 이미지(객체)의 뇌파 모형값은 상기에서 기술한 공식 2-2를 이용하여 구한다(〈표 9〉 참조)(3.1.1장 참조). 한편 시선추적 지수로 시각 모형값(v_et_P3b)의 평균이 사용된다(4.1.3장 참조). 개인별 차이가 있긴 하지만 일반적으로 관심있는 자료를 볼 때 관심이 없는 자료보다 높은 뇌파값 또는 높은 동공 크기값을 나타낸다.

5.1.2 적합성 피드백 엔진

이용자가 위에서 예시한 검색 결과를 브라우징할 때 실시간으로 얻은 이미지에 반응하는 뇌파 및 시선추적 신호를 암묵적 피드백으로 활용하기 위해서 이 신호를 적합성 피드백 엔진에 있는 분류기를 이용하여 분류한다. 만약 분류 결과, 관심을 갖는 이미지로 판단되면 그 이미지와 주제가 유사한 다른 이미지들을 검색하여 맞춤형 검색 결과를 보여 주거나 해당 이미지에 대한 자세한 정보를 보여주는 절차를

〈표 9〉 이용자 프로파일의 뇌파/시선추적 데이터

이용자 ID	그룹 ID	ERP 유형	적합성 판정	미디어 ID	뇌파 모형값 (o_P3b_uV)	시각 모형값 (v_et_P3b)			
1005	6	P3b	관련	2-10-1501	12.86	0.53			
채널별 뇌파값(o_P3b)									
채널	뇌파값	채널	뇌파값	채널	뇌파값	채널	뇌파값	채널	뇌파값
O2	8.347	P8	7.164	CP3	9.913	C3	12.518	FT7	7.935
O1	8.641	C4	10.418	CPz	11.014	FCz	14.828	FC3	14.183
Oz	9.862	TP8	9.17	Cz	13.788	Fz	15.537	F3	18.334
Pz	10.558	T8	13.174	FC4	15.561	F4	19.161	FP2	34.658
P4	11.305	P7	8.105	FT8	15.996	F8	25.308	F7	6.137
CP4	9.904	P3	9.375	TP7	4.64	T7	12.123	FP1	11.367

진행할 수 있다(〈그림 4〉 참조). 또는 이용자에게 검색 결과 중 적합한 자료(들)를 선택하게 하여 명시적인 적합성 피드백을 받을 수 있는 경우에는 이를 암묵적 적합성 피드백과 함께 결합하여 상기 작업을 진행할 수 있다.

상기 작업을 하기 위해서는 이용자 프로 파일에 뇌파/시선추적 정보를 포함시켜 이용하는 방안을 고려해 볼 수 있다. 즉, 사전에 뇌파/시선추적 측정을 수행하여 이용자의 주제 적합, 비적합에 반응하는 뇌파/시선추적 정보를 이용자 프로 파일에 저장한다(〈표 9〉 참조). 사전 테스트에서는 피험자에게 검색 질문을 제시한 후 자극(예, 동영상, 이미지, 텍스트)을 보여주고 검색 질문과의 관련도(관련, 부분관련, 무관)를 평가하도록 한다. 하나의 예로 〈표 9〉에서 보여준 것처럼 특정 이미지(객체)에 대한 각 채널의 뇌파값(o_P3b), 뇌파 모형값(o_P3b_uV), 시각 모형값(v_et_P3b), ERP 유형, 적합성 판정 결과 등을 저장하여 학습 데이터로 저장한다. 이 학습 데이터를 이용하여 새로운 자극에 의해 실시간으로 나타나는 뇌파 및 시선추적 신호(검정 데이터)를 분류기를 이용하여 분류

하는 과정을 거치게 된다. 만약 이런 개인적인 정보가 없다면 이용자 프로파일에 저장되어 있는 그룹 ID(예, 나이, 성별, 선호도 등으로 구분된 그룹)를 활용하여 해당 이용자에게 가장 적합한 그룹을 체크하고 그 그룹에 속해 있는 이용자들의 뇌파 및 시선 정보를 활용할 수 있으나 개인별 정보를 활용하는 경우와 비교하여 분류 정확도가 낮아질 수 있다.

5.2 미디어별로 적합한 뇌파 기반, 시선추적 기반 및 인공지능 기반 방법

동영상과 이미지에서 멀티 모달 방식의 분류 정확도가 단일 모달 방식들의 평균 분류 정확도보다 높게 나타났다(〈표 8〉 참조). 텍스트에서도 멀티 모달 방식들의 평균 분류 정확도가 단일 모달 방식들의 평균 분류 정확도보다 높게 나타났다. 그러나 어느 경우에도 모든 방법들을 사용하는 것은 비효율적이다. 따라서 정보 검색시 타겟이 되는 미디어의 유형에 따라 효율적인 방법(들)을 제안하면 다음과 같다(〈그림 5〉 참조).

동영상 유형 1: 뇌파 컴포넌트(av_P600) + 객체-텍스트 인식 (기준점: 자극이 시작되는 지점)
동영상 유형 2: 뇌파 컴포넌트(av_P600) + 시선 경로 패턴
이미지 (얼굴/감정): 뇌파 컴포넌트(f_P3b), 얼굴 인식, 뇌파 컴포넌트(f_P3b + EPN) + 시각 모형(v_et_P3b)
이미지 (객체): 객체 인식, 시각 모형(v_et_P3b), 객체 인식 + 시각 모형
구조화 텍스트: 자연언어 처리, 뇌파 컴포넌트(t_P3b), 시선 경로 패턴
비구조화 텍스트: 자연언어 처리, 뇌파 컴포넌트(t_P3b, t_P600)

〈그림 5〉 미디어별로 적합한 뇌파 기반, 시선추적 기반 및 인공지능 기반 방법

1) 동영상을 두 가지로 구분해 보면 객체, 인물, 동물 등을 다룬 다큐멘터리 동영상을 유형 1로 구분하고, 복잡하고 다양하며 개념적인 주제에 관한 동영상(예, 심리 영화)을 유형 2로 구분한다. 유형 1인 경우에는 뇌파 컴포넌트(av_P600) 기반 방법이 효율적으로 보인다. 이때 높은 분류 정확도를 얻기 위해서 객체-텍스트 기반 방법을 뇌파 컴포넌트 기반 방법과 함께 사용하는 방안을 고려해 볼 수 있다. 본 연구에서는 동영상이 시작되는 지점을 기준으로 하여 뇌파 반응을 분석하였으나 효율성을 높이기 위해서 기준점을 달리할 수 있다. 예를 들어, 먼저 객체-텍스트 기반 방법을 사용하여 동영상 내 관심 자극(예, 객체)을 인식한 후, 해당 자극이 나타난 지점을 기준으로 하여 일정 기간 동안 뇌파 신호의 변화를 분석하는 방안을 활용할 수 있다. 한편 유형 2인 경우에는 개념적인 주제 분석을 위해서 동영상의 많은 부분을 분석해야 할 필요가 있다. 따라서, 동영상 전체 (또는 일부)를 시청하는 동안 정보 요구에 적합한 동영상 자료인 경우 이용자들 간에

유사한 시선 경로 패턴을 보이고, 비적합한 동영상 자료를 본 경우 무작위 시선 경로 패턴을 보일 것이라고 가정한 Shi et al. (2017)의 연구 결과를 활용할 수 있다. 즉, 뇌파 컴포넌트 및 시선 경로 패턴 기반 방법을 결합한 멀티 모달 방식을 채택하여 시간적 제한을 갖는 뇌파 컴포넌트 기반 방법을 보완할 수 있다.

2) 이미지(얼굴)에 나타나는 감정을 인식하는 데에는 뇌파 컴포넌트 기반 방법(f_P3b)이나 얼굴 인식 기반 방법을 사용할 수 있다. 분류 정확도를 높이기 위해서는 뇌파 컴포넌트들(f_P3b, EPN)과 시각 모형을 함께 사용하는 멀티 모달 방식을 채택할 수 있다.

3) 이미지(객체)를 인식하는 데에는 객체 인식 기반 방법, 시각 모형 기반 방법(v_et_P3b)을 이용할 수 있다. 본 연구에서는 인공지능 기술에 의해서 인식이 비교적 쉬운 금관을 사용하여 객체 인식 기반 방법의 분류 정확도가 높게 나타났다. 그러나, 특정한 형태나 구조가 없는 객체 즉, 구름, 추상적인 예술 작품 또는 불분명하거나 중첩된 객체는 객체 인식 기반 방법에 의해서 식별되기 어려울 수 있다. 이런 경우에는 객체 인식 및 시각 모형 기반 방법을 결합한 멀티 모달 방식을 고려해 볼 수 있다.

4) 텍스트는 구조화 텍스트와 비구조화 텍스트로 구분할 수 있다. 제목이 먼저 나오고 그 다음에 기사 본문이 나오는 신문 기사와 같은 구조화 텍스트는 자연언어 처리 방법, 뇌파 컴포넌트 기반 방법(t_P3b), 시선 경로 패턴 기반 방법을 이용할 수 있다. 한편 소셜 미디어 게시물과 같은 비구조화 텍스트는 자연언어 처

리 방법, 뇌파 컴포넌트 기반 방법이 효율적으로 이용될 수 있을 것이다. 이때 사용할 수 있는 컴포넌트에는 t_P3b 또는 t_P600(text P600)을 적용할 수 있다. t_P600 컴포넌트를 제안한 이유는 비구조화 텍스트는 구조화 텍스트에 비해 추론적 처리(inferential processing)에 더 많은 시간이 소요될 수 있기 때문이다(Ye et al., 2022).

6. 결론과 제언

본 연구는 적합성 판정을 위해서 뇌파의 av_P600과 P3b, 시선추적의 et_P3b와 시선 경로 패턴을 사용하여 뇌파 기반 및 시선추적 기반 방법들을 제안하였다. 제안된 방법들의 유용성을 평가하기 위해서 딥러닝 기술을 사용한 인공지능 기반 방법들을 경쟁 기준으로 사용하였다. 평가 결과, 주제에 적합한 동영상과 이미지(얼굴/감정)를 선택하는 데에는 뇌파 컴포넌트 기반 방법이 높은 분류 정확도를 나타냈고, 주제에 적합한 이미지(객체)와 신문 기사를 선택하는 데에는 인공지능 기반 방법이 높은 분류 정확도를 보여 주었다. 끝으로, 연구 문제들의 검증 결과와 이론적 근거를 바탕으로, 암묵적 적합성 피드백에 기반한 디지털 도서관의 인터페이스 구현을 위한 지침을 제안하였다. 본 연구의 의의는 명시적 적합성 피드백 수집

이 용이하지 않은 정보 검색 환경에서 암묵적 적합성 피드백을 활용하여 이용자의 검색 결과에 대한 만족도를 높인다는 점에 있다.

본 연구는 뇌파의 ERP만 다루고 QEEG를 다루지 않았다. 특히 동영상 유형 2의 경우 QEEG 기반 방법이 적합할 것으로 예측된다. 따라서, 동영상에 QEEG 기반 방법을 적용하고 이를 본 연구 결과와 비교하는 후속 연구가 필요하다. 본 연구는 분류 정확도를 계산하기 위해서 뇌파 및 시각 데이터에서 특징들을 추출하여 이들을 분석하는 전통적인 기계 학습 방법을 채택하였다. 앞으로 많은 데이터가 수집되면 적합성 판정에 대한 분류 정확도를 높일 수 있을 것으로 예측되는 딥러닝 방식을 채택한 후속 연구도 필요해 보인다. 구체적으로 시계열 데이터(뇌파 및 시각 원시 데이터) 또는 이미지(뇌파 신호 및 동공 크기 변동 그래프)를 장단기 메모리(long shortterm memory, LSTM)나 CNN을 이용하여 분석하는 연구이다.

뇌파/시선추적 실험의 피험자들이 20~40대 남성들로 구성되어 있어서 연구 결과를 다른 연령, 성별, 교육 수준에 일반화하는데 제한이 있을 수 있다. 또한 본 연구의 실험 데이터로 사용된 재생시간이 30초 이내의 짧은 다큐멘테이션 동영상, 이미지(얼굴/감정), 이미지(객체) 및 신문 기사(구조화 텍스트)와 다른 유형의 데이터 또는 대규모 데이터를 뇌파/시선추적 실험에 사용한다면 그 결과가 달라질 수 있다.

참 고 문 헌

- 윤정원, 신수연 (2021). 페이스북 건강정보 게시물 형식이 시각적 주의와 인지결과에 미치는 영향. 한국 문헌정보학회지, 55(3), 219-237. <http://doi.org/10.4275/KSLIS.2021.55.3.219>
- Allegretti, M., Moshfeghi, Y., Hadjigeorgieva, M., Pollick, F. E., Jose, J. M., & Pasi, G. (2015). When relevance judgement is happening? An EEG-based study. Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, 719-722. <https://doi.org/10.1145/2766462.2767811>
- Bhattacharya, N., Rakshit, S., Gwizdka, J., & Kogut, P. (2020). Relevance prediction from eye-movements using semi-interpretable convolutional neural networks. Proceedings of the 2020 Conference on Human Information Interaction and Retrieval, 223-233. <https://doi.org/10.1145/3343413.3377960>
- Borgalli, M. R. A. & Surve, S. (2022). Deep learning for facial emotion recognition using custom CNN architecture. Journal of Physics: Conference Series, 2236(1), 012004. <https://doi.org/10.1088/1742-6596/2236/1/012004>
- Davis, K. M., Spapé, M., & Ruotsalo, T. (2022). Contradicted by the brain: Predicting individual and group preferences via brain-computer interfacing. IEEE Transactions on Affective Computing, 14(4), 3094-3105. <https://doi.org/10.1109/TAFFC.2022.3225885>
- DeLong, K. A., Quante, L., & Kutas, M. (2014). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. Neuropsychologia, 61, 150-162. <https://doi.org/10.1016/j.neuropsychologia.2014.06.016>
- Eugster, M. J., Ruotsalo, T., Spapé, M. M., Barral, O., Ravaja, N., Jacucci, G., & Kaski, S. (2016). Natural brain-information interfaces: Recommending information by relevance inferred from human brain signals. Scientific Reports, 6(1), 38580. <https://doi.org/10.1038/srep38580>
- Eugster, M. J., Ruotsalo, T., Spapé, M. M., Kosunen, I., Barral, O., Ravaja, N., Jacucci, G., & Kaski, S. (2014). Predicting term-relevance from brain signals. Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval, 425-434. <https://doi.org/10.1145/2600428.2609594>
- Evans, W. J., Cui, L., & Starr, A. (1995). Olfactory event-related potentials in normal human subjects: Effects of age and gender. Electroencephalography and Clinical Neurophysiology, 95(4), 293-301. [https://doi.org/10.1016/0013-4694\(95\)00055-4](https://doi.org/10.1016/0013-4694(95)00055-4)

- Foley, J. J. & Kwan, P. (2015). Feature extraction in content-based image retrieval. In Khosrow-Pour, D. B. A. ed. *Encyclopedia of Information Science and Technology*, Third Edition. Pennsylvania: IGI Global.
- Golenia, J. E., Wenzel, M. A., Bogojeski, M., & Blankertz, B. (2018). Implicit relevance feedback from electroencephalography and eye tracking in image search. *Journal of Neural Engineering*, 15(2), 026002. <http://dx.doi.org/10.1088/1741-2552/aa9999>
- Gwizdka, J. & Zhang, Y. (2015). Differences in eye-tracking measures between visits and revisits to relevant and irrelevant web pages. *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 811-814. <https://doi.org/10.1145/2766462.2767795>
- Gwizdka, J., Hosseini, R., Cole, M., & Wang, S. (2017). Temporal dynamics of eye-tracking and EEG during reading and relevance decisions. *Journal of the Association for Information Science and Technology*, 68(10), 2299-2312. <https://doi.org/10.1002/asi.23904>
- Jacucci, G., Barral, O., Dae, P., Wenzel, M., Serim, B., Ruotsalo, T., Pluchino, P., Freeman J., Gamberini, L., Kaski, S., & Blankertz, B. (2019). Integrating neurophysiologic relevance feedback in intent modeling for information retrieval. *Journal of the Association for Information Science and Technology*, 70(9), 917-930. <https://doi.org/10.1002/asi.24161>
- Kim, H. H. & Kim, Y. H. (2019a). Video summarization using event-related potential responses to shot boundaries in real-time video watching. *Journal of the Association for Information Science and Technology*, 70(2), 164-175. <https://doi.org/10.1002/asi.24103>
- Kim, H. H. & Kim, Y. H. (2019b). ERP/MMR algorithm for classifying topic-relevant and topic-irrelevant visual shots of documentary videos. *Journal of the Association for Information Science and Technology*, 70(9), 931-941. <https://doi.org/10.1002/asi.24179>
- Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer ed. *The Cambridge Handbook of Multimedia Learning*. New York: Cambridge University Press, 134-146.
- Pöld, T., Bachmann, M., Päske, L., Kalev, K., Lass, J., & Hinrikus, H. (2018). EEG spectral asymmetry is dependent on education level of men. *World Congress on Medical Physics and Biomedical Engineering 2018*, 2, 405-408. https://doi.org/10.1007/978-981-10-9038-7_76
- Schindler, S. & Bublatzky, F. (2020). Attention and emotion: An integrative review of emotional face processing as a function of attention. *Cortex*, 130, 362-382. <https://doi.org/10.1016/j.cortex.2020.06.010>
- Shi, Z. F., Zhou, C., Zheng, W. L., & Lu, B. L. (2017). Attention evaluation with eye tracking

- glasses for EEG-based emotion recognition. 8th International IEEE/EMBS Conference on Neural Engineering (NER), 86-89. <https://doi.org/10.1109/NER.2017.8008298>
- Udurume, M., Valverde, E. C., Caliwag, A., Kim, S., & Lim, W. (2023). Real-time multimodal emotion recognition based on multithreaded weighted average fusion. *Journal of the Ergonomics Society of Korea*, 42(5), 417-433. <http://doi.org/10.5143/JESK.2023.42.5.417>
- Ul Haq, H. B., Asif, M., Ahmad, M. B., Ashraf, R., & Mahmood, T. (2022). An effective video summarization framework based on the object of interest using deep learning. *Mathematical Problems in Engineering*, 2022, 7453744. <https://doi.org/10.1155/2022/7453744>
- Wang, W., Song, H., Zhao, S., Shen, J., Zhao, S., Hoi, S. C., & Ling, H. (2019). Learning unsupervised video object segmentation through visual attention. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3064-3074.
- Ye, Z., Xie, X., Liu, Y., Wang, Z., Chen, X., Zhang, M., & Ma, S. (2022). Towards a better understanding of human reading comprehension with brain signals. *Proceedings of the ACM Web Conference*, 380-391. <https://doi.org/10.1145/3485447.3511966>
- Zheng, W. L., Liu, W., Lu, Y., Lu, B. L., & Cichocki, A. (2018). Emotionmeter: a multimodal framework for recognizing human emotions. *IEEE Transactions on Cybernetics*, 49(3), 1110-1122. <http://dx.doi.org/10.1109/TCYB.2018.2797176>

<p>• 국문 참고문헌에 대한 영문 표기 (English translation of references written in Korean)</p>
--

- Yoon, Jung Won & Syn, Sue Yeon (2021). How do formats of health related Facebook posts effect on eye movements and cognitive outcomes?. *Journal of the Korean Society for Library and Information Science*, 55(3), 219-237.
<http://doi.org/10.4275/KSLIS.2021.55.3.219>