

Research on the Design of a Deep Learning-Based Automatic Web Page Generation System

Jung-Hwan Kim*, Young-beom Ko*, Jihoon Choi*, Hanjin Lee**

*Ph.D, Dept. of Digital Management, Korea University, Sejong, Korea

*Ph.D, Dept. of Digital Management, Korea University, Sejong, Korea

*Ph.D, Dept. of Digital Management, Korea University, Sejong, Korea

**Professor, School of Creative Convergence Education, Handong Global University, Pohang, Korea

[Abstract]

This research aims to design a system capable of generating real web pages based on deep learning and big data, in three stages. First, a classification system was established based on the industry type and functionality of e-commerce websites. Second, the types of components of web pages were systematically categorized. Third, the entire web page auto-generation system, applicable for deep learning, was designed. By re-engineering the deep learning model, which was trained with actual industrial data, to analyze and automatically generate existing websites, a directly usable solution for the field was proposed. This research is expected to contribute technically and policy-wise to the field of generative AI-based complete website creation and industrial sectors.

▶ **Key words:** Deep Learning, Artificial Intelligence, Automatic Website Generation, E-commerce, Website Classification, Web Page Components

[요 약]

본 연구는 폭증하는 디지털 비즈니스의 수요 증가를 감당하기 위하여 AI를 활용한 새로운 제작 방법을 모색하는데 목적이 있다. 이에 딥러닝과 빅데이터를 기반으로 실제 웹페이지 생성 가능 시스템을 구축하고자 하였다. 첫째, 이커머스 웹사이트 기능을 바탕으로 분류체계를 수립하였다. 둘째, 웹페이지 구성요소의 유형을 체계적으로 분류하였다. 셋째, 딥러닝이 적용가능한 웹페이지 자동생성시스템 전체를 설계하였다. 실제 데이터를 학습하여 구현된 딥러닝 모델이 기존 웹사이트를 분석하고 자동생성되도록 재설계 함으로써, 산업에서 바로 사용가능한 방안을 제안했다. 나아가 체계가 부족했던 웹사이트 레이아웃 및 특징에 대한 분류체계를 수립했다는 측면에서 의의가 있다. 이는 향후 생성형 AI 기반의 웹사이트 연구 및 산업 분야에 크게 기여할 수 있을 것이다.

▶ **주제어:** 딥러닝, 인공지능(AI), 웹사이트 자동생성, 이커머스, 웹사이트 분류체계, 웹페이지 구성요소

- First Author: Jung-Hwan Kim, Corresponding Author: Hanjin Lee
- *Jung-Hwan Kim (kjhbond@gmail.com), Dept. of Digital Management, Korea University
- *Young-beom Ko (jedihs7480@gmail.com), Dept. of Digital Management, Korea University
- *Jihoon Choi (kemp32@kolon.com), Dept. of Digital Management, Korea University
- **Hanjin Lee (cus@handong.edu), School of Creative Convergence Education, Handong Global University
- Received: 2024. 01. 09, Revised: 2024. 02. 05, Accepted: 2024. 02. 05.

I. Introduction

COVID-19 팬데믹으로 인해 디지털 전환이 가속화되고, 그 중요성과 시급성이 높아졌고[1], 정보통신기술(ICT)이 발전함에 따라 비즈니스 모든 영역에 디지털 기술이 결합되어 변화를 일으키고 있다[2]. 이에 많은 경영진들이 디지털 전환에 주요 자원을 할당하면서 웹사이트 수요와 실제 수가 크게 증가하였다[3]. 글로벌 통계 리서치업체인 인터넷라이브스탯(internetlivestats)[4]에 따르면, 2010년 대비 2023년 현재의 글로벌 웹사이트 수는 무려 5배가 넘게 폭발적으로 증가한 약 11억 개를 기록했다. 이렇게 지속적으로 증가하는 웹사이트 제작의 수요에 따라 생산성 증가가 필요했고, 이를 충족할 시스템 및 서비스의 발전이 이뤄져 왔다. 초기 웹페이지가 직접 제작하던 방식에서 발전하여 CMS(Content Management System)를 활용한 생산성의 혁신으로 수요를 따라잡았다. 최근에는 해상도에 따라 왜곡되지 않도록 방지하는 반응형 웹(Responsibility Web) 기술을 포함하여[5] 웹 접근성(Web Accessibility)도 준수할 수 있는 시스템이 상용화되었다. 즉, No-Code 방식으로 제공할 수 있는 웹사이트 자동생성 및 관리 시스템을 활용해 생산성 및 사용성이 대폭 개선된 제작환경이 조성되었다.

한편 인공지능 연구는 생각보다 역사가 오래되었지만[6] 최근 ChatGPT가 발표되는 등 세계를 강타한 놀라운 서비스 덕분에 세상에 주목을 받기 시작했다. 이와 함께 관련 연구도 더욱 활발해지며 인공지능에 관한 많은 성과가 이뤄지고 있다. 웹사이트 제작산업도 웹빌더의 등장으로 생산성이 크게 증가했지만, 이제는 인공지능을 융합하여 더욱 생산성을 높이고 품질도 함께 높이는 효율적인 방법을 모색해야만 한다. 일부 발빠른 기업과 선행연구자들은 웹사이트 자동생성시스템과 인공지능을 접목하여 다양한 시도를 하고 있다. 10Web이나 워스와 같이 직접 모델을 학습하고 융합하여 실제 서비스하는 사례도 등장했다[6]. 웹사이트 자동생성시스템과 인공지능의 융합에 관한 선행연구는 UI(User Interface)에 대한 패턴을 바탕으로 시제품(목업)-디자인을 미리 만들어주거나[7] 예제 기반으로 빠른 생성을 도와주기도 한다[8-9]. 아울러 스케치를 기반으로 인공지능이 분석 및 생성을 도와주는 형태[10] 등 크게 세가지 분야로 연구 및 발전이 이뤄지고 있다[11].

하지만, 이렇게 활발함에도 불구하고 아직은 전문가가 아닌 일반인이 웹사이트를 완전히 완성할 수는 없다. 이것은 웹사이트 제작 분야와 관련한 인공지능의 한계와 관련이 깊다. 많은 선행연구들이 지적한 바와 같이 아직은 사

용자의 의도를 매우 잘 분석하거나, 웹사이트의 구조와 기능을 매우 잘 이해하기 어려운 상황이다. 게다가 웹사이트 기반 코드를 보다 효율적으로 생성하여 완성도 높은 결과를 보여주는 인공지능 모델이 아직은 없기에, 보다 정확하고 전문적인 결과를 만들어낼 기술이 필요하다.

본 연구에서는 현재까지 가장 발전된 기술인, 웹사이트 자동생성시스템을 기반으로 딥러닝 모델을 융합해보고자 한다. 이를 통해 전문가들이 만든 것과 같이 높은 품질의 자동생성 결과를 보여주는 시스템을 설계하여 최근의 요구사항을 충족시킬 수 있는 방법을 연구하고자 한다. 이를 위해 대상 웹사이트의 범위를 선정하고, 웹사이트의 분류체계를 마련할 것이다. 나아가 웹페이지를 구성하는 요소에 대한 분류체계도 제안하고자 한다. 최종적으로 분류체계 기반의 딥러닝 모델을 융합하여 웹페이지를 자동으로 생성해주는 시스템 설계를 제안할 것이다.

II. Preliminaries

1. Related works

1.1 CMS (Contents Management System)

2000년대 웹사이트 제작 전문업체가 직접 만들어야만 이용할 수 있는 시대에서, 최근 10년 사이 CMS(Content Management System)로 인해 웹사이트를 제작하는 패러다임이 크게 바뀌었다. 유행을 선도한 CMS는 오픈소스가 주류였는데, 이것을 설치하면 관리자 화면이 빠르게 생성되었다. 간단한 설정만 하면 사용자 화면도 쉽게 생성되어 디자인만 변경하면 빠르게 웹사이트를 완성할 수 있어 제작 생산성이 크게 향상되었다. 당시 '워드프레스 (WordPress)'는 세계에서 가장 인기있는 오픈소스였는데, 다양한 플러그인을 스토어에서 손쉽게 설치하고, 복잡한 웹 사이트를 생성 및 관리할 수 있는 시스템으로 알려져 크게 확장되었으며[12], 현재는 글로벌 전체 웹사이트의 43%가 워드프레스로 제작되었다고 홍보하고 있다.

국내에서도 2010년대 중반, 규모가 큰 공공기관인 '서울시청'과 '경기도청' 웹사이트를 워드프레스 기반으로 제작하라는 사업공고가 발표되었다. 당시 해당 글로벌 웹사이트는 연구진이 직접 참여하여 워드프레스를 기반으로 개발을 주도한 경험이 있다. Aladdin(2018)에 따르면 워드프레스 리뷰에서 PHP라는 개발언어를 몰라도 되는 상황과 워드프레스의 장점에 대해 언급하며, 서비스 이용자 측면으로의 기술변화가 크게 이뤄지고 있었음을 시사한다.

2010년대 초반에 무료 오픈소스 웹사이트 생성 및 콘텐츠

츠 관리 시스템인 워드프레스가 등장하며 제작의 생산성을 크게 높였다[13]. 이에 발맞춰 완성된 웹사이트 템플릿만을 전문으로 판매하는 플랫폼이 생기며 다시 한번 생산성과 품질이 크게 향상되는 계기가 되었다. 테마포레스트는 구매한 템플릿을 워드프레스에 업로드 및 클릭 몇 번만으로 바로 웹사이트가 되는 혁신적인 경험을 제공했으며 현재까지 서비스 중이다. 이런 에반토는 템플릿의 범위를 더 확장하여 웹 소스와 영상 조각까지 판매하며, 다양한 템플릿을 보급했고, 생산성 향상에 크게 일조하였다. 이러한 템플릿 기반의 제작은 생산성만이 아니라, 결과물의 품질과 제작기간의 단축이 가능하므로, 전문가에게 큰 영향을 주고 있다.

1.2 Web-Builder System

웹빌더는 CMS의 단점을 명확히 보완하여 등장한 사용자 친화적인 웹사이트 제작 서비스이다. 웹빌더는 코드를 작성하지 않는 것이 아니라, 아예 할 수 없도록 고안되었다. 파워포인트처럼 디자인 화면을 바탕으로 웹사이트를 편집할 수 있게 해주며[14], 다른 서비스로부터 별도의 테마를 구매해서 설치하는 번거로움 없이, 자체 서비스로부터 몇 번의 클릭으로 선택과 사이트 생성이 빠르게 이뤄진다. 웹빌더 서비스에서는 누구나 무료 또는 적은 비용으로 템플릿을 사용하여 웹사이트를 만들고 온라인 비즈니스를 쉽게 시작할 수 있다[15]. 이 시스템은 전자상거래 플랫폼이나 기초적인 수준의 콘텐츠관리시스템(CMS)도 포함하며, 디자인 또는 기술이 없는 개인이 웹사이트를 만들 수 있도록 지원한다[16].

wix는 현재 글로벌에서 가장 유명하다고 평가받는 웹빌더 서비스이다. 현재는 글로벌 가입자 수가 무려 1억명이 넘으며, 웹사이트 자동생성시스템 서비스 분야의 선두주자이기에 다양한 기업이 워드의 서비스를 표방하여 비슷한 서비스를 출시하였는데, 기본적으로 제공되는 기능은 다음과 같다[17].

- 텍스트, 이미지 및 비디오 추가
- 배경 및 다른 요소 변경
- 사이트 메뉴 편집 및 페이지 전환 선택
- 슬라이드쇼 및 호버 상자 등 상호작용 요소 추가
- 소셜미디어, 연락처 양식 및 블로그 추가
- '회원 전용'을 위한 비공개 페이지 추가
- 제품 갤러리와 같은 '스토어 요소' 추가 등

1.3 Comparison of Web Builders and Developers

다음으로 이러한 '웹사이트 자동생성시스템 기반 개발'과 기존의 제작방식인 웹개발자를 같은 환경에서 비교하면 얼마나 차이가 있을까? 탐색이 필요하다. Forbes(2019)는 위의 비교를 개발속도, 사용성, 제작된 웹사이트의 수, 그리고 제작비용에 대하여 다음의 <Table 1>과 같이 조사하고, 결론을 언급하고 있다[14].

Table 1. Comparison of Web Builders and Developers

	wix	shopify	developer
Speed (hour)	2	2	40
Price (USD) /Month	\$4-30	\$30-300	\$2,000 -10,000
Count (2019)	3,373,287	981,768	170-200 Million

Forbes(2019)는 20명의 전문가를 대상으로 '웹빌더 웹 개발자가 하는 일을 얼마나 효과적으로 복제하는가?'에 대해 전문가 설문을 실시했다. 그 결과 10점 만점 중 4.8점 수준으로 복제가 된다는 결론을 제시한다. 또한 '웹빌더가 전통적인 웹 개발 산업에 미친 영향은 무엇인가?'의 질문에서는 놀랍게도 영향은 매우 미미하며 웹개발자에게 심각한 위협을 주지 않는다는 의견을 피력했다. 그 이유는, 해당 방법이 사용자 친화적인 방법을 제공하지만, 사용자 정의와 복잡성 측면에서 웹 개발 전문가에 의한 맞춤형 웹 개발과 비교했을 때 여전히 제한 사항이 있다는 것이다. 아울러 웹빌더는 개인 및 소규모 비즈니스에 적합하지만, 더 복잡하고 커다란 프로젝트의 경우에는 웹 개발자가 여전히 중요한 역할을 한다고 밝혔다[14].

하지만, 이것은 개발자들이 체감하는 내용에 대한 설문 결과이므로 사실과는 거리가 있을 수 있다. 개발자들은 자신의 도구에 익숙하며, 결코 웹빌더를 사용하지 않는다는 현장의 환경이 고려되지 않았기 때문이다. 즉, 웹빌더가 필요한 사람은 개발자가 아닌 비 개발자이며 이들은 일반 사용자 일수도 있고, IT 전문가 중 개발자가 아닌 기획자, 디자이너 그리고 마케터나 영업자일 수 있다. 이들에게 No-Code 기반의 웹빌더와 같은 서비스는 기술의 장벽을 해결해 주는 유일한 수단이자, 웹사이트를 직접 제작하기 위해 필요한 서비스이므로 대상이 바뀌면 크게 다른 결과를 얻을 것으로 예상된다. 본 연구자들도 10년 이상의 웹 개발 경력을 갖고 있으나, 웹빌더의 사용성과 품질이 지속적으로 향상된다면 Forbes(2019)의 설문결과와 다른 결론을 제시할 것으로 보인다. 오히려 최근의 트렌드를 보면

개발자에 대한 수요가 크게 줄어들 것이라 예측되며, 웹 개발 산업에 미치는 영향이 매우 클 것이라고 판단한다. 이것은 기술이나 서비스가 발전하며 직업이 사라지거나 형태가 변화하는 것과 유사한 맥락으로 보인다.

1.4 Artificial Intelligence-based Web Builder

Kaluarachchi의 연구에서는 47개의 웹사이트 자동생성시스템에 인공지능이 적용된 사례에 대해 조사했다[11]. 이를 기반으로 크게 3가지 분류 즉, 1)예제기반, 2)목업기반 그리고 3)인공지능 기반으로 나눌 수 있다며 <Fig 1>와 같이 정리하였다.

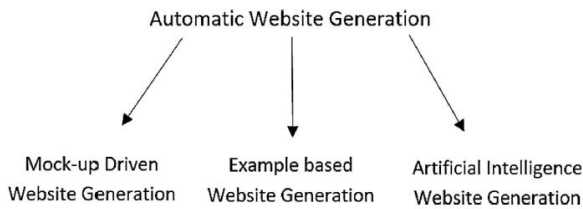


Fig. 1. Automatic Website Generation Methods

첫 번째, 목업기반 웹사이트 자동생성은 답러닝을 적용시킨 웹사이트 자동생성 방법에서 가장 많은 연구가 있었다. 이것은 대략의 설계도를 그리면 그것이 개발물로 자동 생성되는 것처럼 작동하는 방법을 일컫는다. GUI(Graphical User Interfaces)를 설계하고 코드로 자동변환하는 것과 유사한 동작수행을 말한다[18].

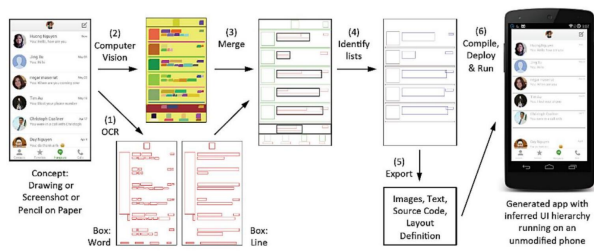


Fig. 2. Illustration of Mockup-based UI Code Generation

<Fig 2>은 목업기반의 접근에 대한 작동 예시이다. 목업기반 작동방식은 참고할 이미지를 바탕으로 UI컴포넌트나 텍스트를 찾는다. 이를 답러닝 모델의 Object Detection 알고리즘을 활용해 그룹 분리와 코드 생성을 하는 방식으로 작동한다.

두 번째, 예제 기반 웹사이트 생성은 사용자가 스케치한 결과를 바탕으로 온라인 상에 공개되어 있는 많은 예제와의 유사도를 측정한다. 그 후 결과를 보여주고 사용자가 선택하도록 안내한다. 여기서 분석된 레이아웃과 컴포넌트

를 바탕으로 비슷한 예제를 추천하기 위해서는, 대량의 데이터 수집이 선행되어야 한다. 이에 구현의 난이도가 상대적으로 더 높다. 하지만 이러한 한계점에도 불구하고, 실제 예제 웹사이트를 기반으로 생성하기 때문에, 산업에서 보다 빠르게 사용할 수 있는 현실적인 방법이 될 수 있다. 선행연구를 조사한 결과, PC 기반의 웹사이트를 다루는 사례에서 점차 Mobile 기반의 어플리케이션 개발의 영역을 다루는 쪽으로 증가하고 있다[18,19]. <Fig 3>은 스케치를 바탕으로 가장 유사한 결과물을 추천하는 Forrest Huang 외(2019)의 연구인 스와이어(Swire)의 개요[19].

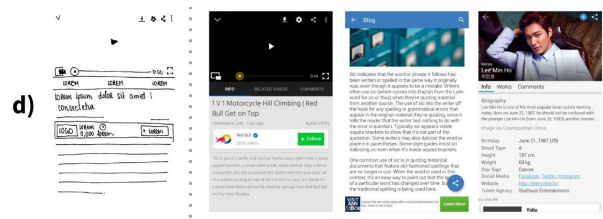


Fig. 3. Illustration of Example-based UI Code Generation

세 번째, 인공지능 기반의 웹사이트 생성은 글로벌 매출 1위의 웹빌더 서비스 업체인 wix사의 Artificial Design Intelligence(ADI) 서비스를 들 수 있다[17]. 해당 서비스는 고객 설문 데이터를 바탕으로 가장 적합한 유형의 템플릿을 선별한다. 그 후 입력된 데이터를 바탕으로 입력 및 최적화 과정을 거쳐 생성이 지원된다. 하지만 이는 광범위한 인공지능의 영역일 뿐, 답러닝 모델 기반의 혁신적인 생성형 결과가 아니므로, 자칫 템플릿의 추천 정도로 여겨질 수 있다. 그럼에도 불구하고, 예제 기반의 제작과 마찬가지로 산업에서 바로 사용할 수 있는 현재의 가장 진보된 방법으로 여겨진다.

선행연구를 바탕으로 목업기반의 답러닝을 바탕으로 컴포넌트나 레이아웃을 탐지하고 GUI 코드를 자동생성해주는 연구가 꽤 오래전부터 진행되어 왔음을 알 수 있다. 이들의 결과물은 상당히 유용하고 발전가능한 연구임에 틀림없지만, Kaluarachchi이 한계로 언급한 것[11]과 유사하게 다음의 한계점을 느낄 수 있었다.

먼저 선행연구의 결과들은 웹사이트의 제작이라는 범위보다 단일 페이지에서의 섹션(Section) 수준을 손쉽게 배치할 수 있는 부분이다. 즉, 전체 웹페이지를 만들어주거나 추천을 통해서라도 완성되는 결과를 볼 수 없으므로 실제 산업에서 활용하기에는 더딘 기초 수준의 분석과 생성이라는 한계점이 있다.

다음으로, 구성요소(컴포넌트)의 탐지와 다양성이 적은 컴포넌트 코드의 재배치 수준으로 작동한다는 한계가 있다.

GUI가 디자이너의 창의적 영역이라는 점을 고려할 때, 이 결과에 대하여 전문가 또는 일반인의 심미성과 만족도에 대한 가설을 세우기 어렵다. 여기서는 색상의 조화나 폰트와 같은 전체적인 디자인 품질에 관한 기대를 할 수 없다. 마지막으로, 탐지의 영역으로는 이해하기 어려운, 시스템 이용자의 제작의도에 대한 분석이 누락된 것이다. 즉, UI의 구조는 분석과 생성이 가능하지만, 이용자가 이미지를 강조하고자 의도하는 것인지, 정보전달이 중요한 것인지는 정확히 반영되기 어렵다.

III. The Proposed Scheme

먼저, 본 연구에서 분석 및 제작 가능한 범위 설정을 통해 웹페이지 자동생성을 위한 딥러닝 모델의 학습 범위와 연구 완성도를 향상시키고자 했다. 이러한 이유로, 분석 및 자동생성 되는 웹사이트의 카테고리를 “전문 전자상거래(이하 이커머스) 쇼핑몰”로 한정했다. 종합몰이나, 포털과 같은 디자인은 일반적인 쇼핑몰과 레이아웃이나 제작의 범위가 크게 다르기 때문에, 분류체계의 수립이나 모델의 학습 등에 어려움이 클 것으로 판단되었다.

본 연구에서 제안하는 시스템은 국내 웹호스팅 플랫폼 서비스를 영위하는 업체들에서 가장 많이 사용되는 프로그래밍 언어인 PHP언어의 버전 7.x와 무료로 배포되는 MariaDB를 활용해 서버사이드를 개발할 수 있다[11, 18]. 나아가 인공지능 모델의 서빙 API 구현을 위한 파이썬 언어를 활용해 전체 서비스를 구현할 수 있다. 사용자 화면은 일반적인 HTML, CSS3.0, Javascript 등 언어의 제약 없이 모두 구현이 가능하다.

이번 분석에서 제안하는 딥러닝에 의한 웹페이지 자동생성기법은 사용자 제작의도를 담고 있는 벤치마킹 웹사이트의 웹페이지를 분석하는 것에서 시작한다. 분석의 작동 방법은 웹페이지 캡처 후 1. 단락을 탐지하고, 2. 단락을 유형별로 분류하며, 3. 요소를 탐지하고, 3. 특징을 탐지하여 종합적인 태그 데이터를 추출하는 것이다. 그리고 이를 미리 정의된 템플릿의 태그 데이터와 비교하여 가장 유사한 템플릿을 가져와 조립하는 형태로 <Fig 4>과 같이 작동할 수 있다.

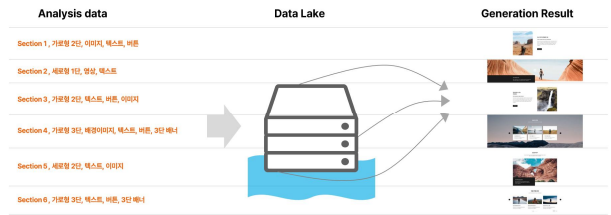


Fig. 4. Deep Learning-based Analysis and Generation

3.1 Section Type Classification

위의 방법으로 웹페이지를 데이터화 하기 위해서는 페이지를 구성하는 각 요소에 대한 정의가 필요하다. 그러나 선행연구에서는 웹페이지 구성요소를 디자인과 요소의 관점에서 분류한 결과를 찾기 어려웠다. 이러한 이유로 본 연구에서는 직접 다양한 웹사이트의 메인 페이지를 수집하여 분류체계를 수립하고자 했다. 그리고 전문 쇼핑몰의 구성 요소에 대한 분석을 위하여 관련 데이터베이스를 많이 축적하여 서비스하고 있는 '디비컷(Dbcut)' 웹서비스를 활용했다[20]. 디비컷에서는 쇼핑몰을 카테고리 기준으로 잘 정리해 두었으며, 여기서 3,000종 이상의 쇼핑몰 메인 페이지를 수기로 수집하여 확인 및 체계를 수립할 수 있었다.

<Fig 5> 과 같이 모든 쇼핑몰은, 정보를 가로 단으로 구분하고 있었고, 이를 섹션(Section)으로 정의하였으며, 세로 단의 경우 1단에서 10단이 넘는 경우까지 매우 다양함을 알 수 있었다. 간혹, 같은 성격의 섹션임에도 2~3층으로 구성된 경우도 있었으며 이러한 섹션이 갖는 특성이 매우 유사함을 확인할 수도 있었다. 예를 들어, 조사된 모든 쇼핑몰은 가장 높은 위치에 있는 섹션을 Header로 취급하고 있으며, 가장 낮은 위치에 있는 섹션을 Footer로 취급하고 있었다.

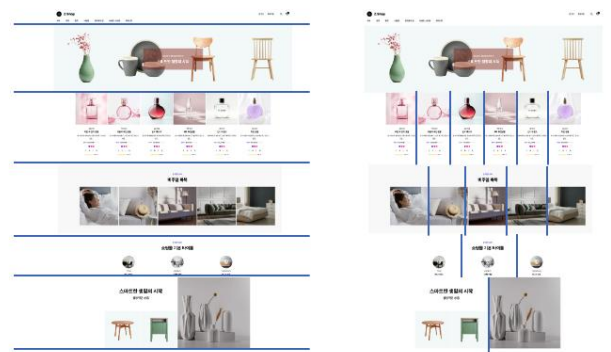


Fig. 5. Horizontal, Vertical Section Classification

나머지 가운데의 여러 개 섹션은 Body로 취급하는데, 별도로 Body라는 섹션은 존재하지 않았다.

그래서 Header와 Footer 및 7가지 단에 따른 섹션과 2~3층으로 구성된 '갤러리' 섹션을 포함했다. 이에 최종적으로 <Table 2>와 같이 쇼핑몰 메인 페이지를 구성하는 레이아웃을 총 10개의 섹션 유형 분류체계로 수립할 수 있었다[20].

Table 2. Web Page Section Type Classification

Type	Description
Header	a paragraph located at the very top or left side of a Webpage, which includes the logo and menu
Footer	a paragraph located at the very bottom of a Webpage, containing site information and links related to privacy protection
Banner	paragraph includes promotional messages and images. The main visual under the Header is also one of the banners
2 Block	This refers to a paragraph where images or text are divided into two areas. Typically, it's a design paragraph where text is split between the left and right sides
3 Block	a paragraph divided into three areas
4 Block	paragraph divided into four areas. In shopping mall Websites, there are sections where product lists are displayed in four items per line
5 Block	a paragraph divided into five areas. Similar to the '4block', it is commonly used in shopping mall Websites
6 Block	a paragraph divided into six areas. It is frequently used in shopping mall sites, especially in the category areas
n Block	This paragraph, divided into more than six areas, is used when there are many categories, similar to the '6block'
Gallery	This refers to a paragraph with images where the division of areas is ambiguous. This also includes unsorted image paragraphs

3.2 Element Type Classification

다음으로 레이아웃을 구성하는 섹션에 대한 분류체계를 수립한 후, 각 섹션을 구성하는 웹 요소(Element)가 무엇인가를 분석하는 것이 필요했다. 본 연구에서 정의하는 섹션을 구성하는 요소(Element)란, Image, Text와 같이 눈에 보이는 가장 작은 단위의 HTML Tag 요소이며, 추가로 정보전달을 위해 Tag 요소가 조합되어 만들어진 기능적 요소(Functional Element)를 포함한다[11, 20].

이에 3.1과 마찬가지로 기존에 수집된 3,000장의 이미지를 활용하여 분류체계를 수립하였다. 여기서는 최초 100장을 기준으로 분류를 수립한 후, 이를 바탕으로 다음 100장을 기준으로 검토하는 작업과 분류 불가요소를 바탕으로 추가 분류기준을 수립하는 등의 과정을 반복적으로 거쳤다. 처음 분류체계는 구조가 존재하는 형태로 제작되었지만, 추후 데이터 전처리과정 중 가공의 용이성과 실제 딥러닝 모델의 학습 등에 크게 영향을 미칠 수 있어 보완적인 분류체계를 수립하였다[18]. 이는 포괄적이면서 태그 기반으로 작동할 수 있도록 설계되었는데, 다음과 같다.

- 마스터(Header, Footer), 메뉴, 그리드, 페이지, 타이틀, 텍스트, 이미지, 영상, 배너, 표, 게시판, 폼, 지도, 버튼, 아이콘, 라인, 파일

3.3 Deep Learning-Based Automatic Web Page Generation System Design

딥러닝이 적용된 웹사이트 자동생성 시스템은 <Fig. 6>과 같이 사용자가 입력한 벤치마크 웹사이트의 URL을 통

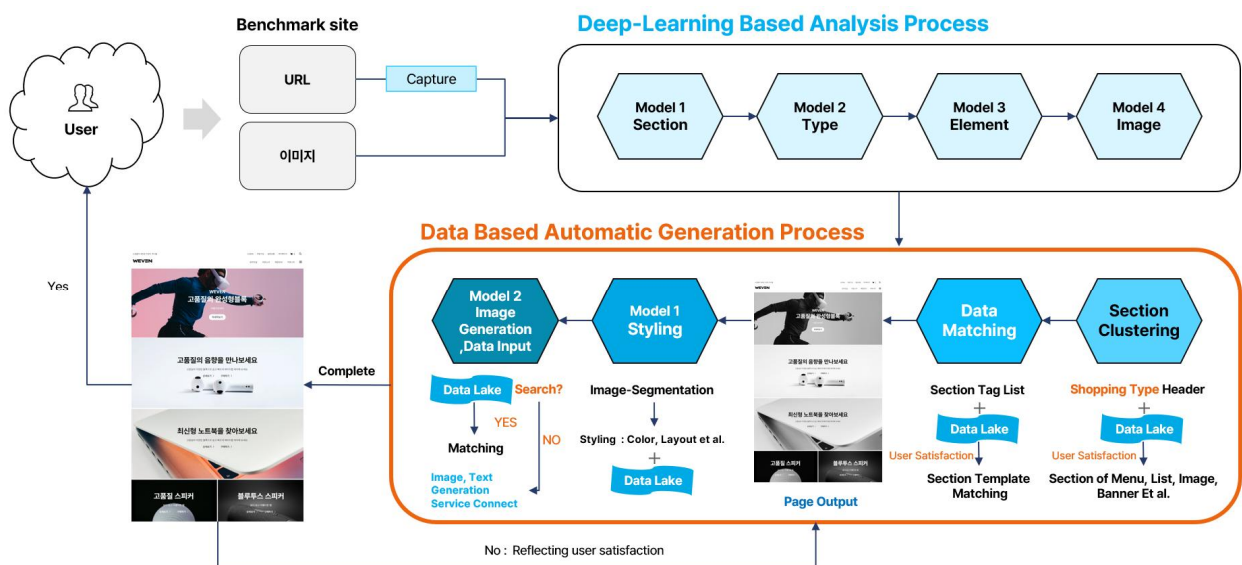


Fig. 6. Design of Deep Learning-Based Automatic Web Page Generation System

한 웹페이지 캡처와 스크래핑 데이터를 바탕으로 딥러닝 모델 4개의 분석을 진행했다. 이를 통해 앞서 정의한 섹션의 유형과 요소의 유형 등 다양한 데이터를 추출하고, 이 데이터를 조합하여 데이터레이크에 탑재된 섹션을 군집화했다. 이후 유사도가 가장 높은 데이터를 바탕으로 섹션 템플릿을 추천하여 섹션별로 분석의 순서대로 조합되었다.

1차 웹페이지를 생성하고 분석된 데이터로부터 대표 컬러 및 웹사이트의 산업 성격 등을 고려하여 이미지나 컬러 및 적합한 콘텐츠를 보정하는 순으로 전체 생성 프로세스를 설계하였다. 최종적으로 소비자에게 전달될 웹페이지가 만족스러운지를 확인하는 절차를 두어, 불만족 시 매칭 및 생성의 절차에 대한 결과값을 입력하고 머신러닝을 통해 비슷한 결과가 재발되지 않도록 장치를 두었다. 그리고 다시 생성 절차를 시작하여 새로운 결과를 제공함으로써, 점진적으로 만족할 수 있도록 시스템을 설계했다[18, 20].

여기서 웹페이지의 분석은 총 4단계의 딥러닝 모델을 기반으로 이루어졌다. 1)섹션, 2)섹션의 유형분류, 3)요소의 탐지, 4)섹션의 특징 탐지로 구성하였으며, 예상되는 결과는 <Table 3>과 같다. 먼저, 섹션탐지 단계에서는 YOLO모델을 기반으로 웹 페이지에서의 섹션을 추출한다. 그리고 탐지된 섹션 이미지를 잘라서 각 섹션의 분류를 분석한다. 분류의 분석은 분류모델로 인공지능 시장에서 단연 선두를 달리는 EfficientNet V2를 사용했다. 그리고 바로 요소 및 섹션특징 분석을 비동기로 동시에 호출한다. 섹션의 특징 탐지로는 그 크기를 추출하는 것과 로고의 위치를 파악하거나 주요 색상추출 등을 수행한다. 나아가 요소 분석을 통해 섹션을 구성하는 요소들을 추출하는 분석 과정을 수행했다. 여기서 추출된 데이터는 각각의 태그로 지정되어 웹 페이지의 템플릿을 매칭할 수 있게 된다.

Table 3. Web page Section type classification

Model	Expected analysis result data
Section Detection	Header, 1, 2, 3, 4, 5...n, Footer
Section Type Classification	Header Logo Left, Right Footer Logo Left, Right
Object Detection	Element Type Element Size Element Color Element Position
Section Specification Detection	Section Width Section Key Color is include Text?

본 연구에서 제안된 설계에서 템플릿은 매우 중요하다. 이 템플릿은 반드시 별도 관리도구에 의해 등록 및 수정과 분류 작업이 가능하도록 고안된 시스템이 필요하다. 여기서는 구현된 분석용 딥러닝 모델을 활용하여 일관된 태그가 입력되도록 했다. 추후 모델의 업데이트가 생기면 자동으로 태그 업데이트가 이뤄져 분석 데이터와 매칭에 버전 오류가 없도록 제공한다. 이번 분석에서는 템플릿 관리 도구가 갖춰져 있다는 전제 하에 제안하는 부분으로써 실제 구현을 위한 제안은 후속연구가 필요할 것이다. 본고에서는 탐색적 차원에서 전체 과정의 설계를 중심기술하였다.

IV. Conclusions

본 연구는 급증한 웹사이트 제작수요에 따라 발전하고 있는 혁신적인 인공지능의 기술을 융합하여 웹사이트 자동생성시스템에 딥러닝을 적용한 새로운 시스템을 설계하는 것에 중점을 두었다. 이 시스템은 전통적인 웹사이트 제작의 절차를 일정한 규칙과 표준을 바탕으로 분석할 수 있도록 틀을 만들어 제공했다. 아울러 딥러닝 모델을 학습되고 또 데이터레이크 구축을 통한 지속적인 매칭이 가능한 방안을 제시했다.

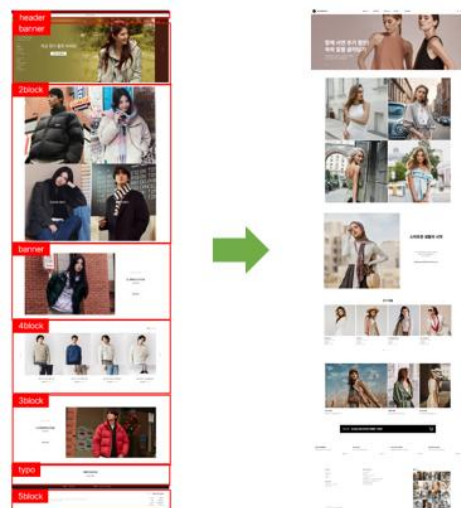


Fig. 7. Analysis And Matching Result

제안된 시스템은 <Fig 7>과 같이 입력된 웹사이트의 단락을 나누고, 단락을 분류하며, 요소와 특징을 추출하여 데이터기반 템플릿을 매칭할 수 있다. 추출된 데이터는 <Table 4>와 같이 태그의 목록으로 변환된다.

Table 4. Analysis of Labeling by Types

Type	Labeling Classification
Master Header	Logo_Left, Center_Menu, White
Banner	Title, Image, Button, Width1800 Height800
2Block	Image, Floor2, Width1200, Photo
Banner	Left_Image, Photo, Right_Title, Button, Width1200
4Block	Photo, Bottom_Title, Shop_List, Width1200, Slider
Banner	Right_Image, Photo, Left_Title, Button, Width1200
Banner	Center_Title, Button
Master Footer	Logo_Left White, Right_Menu

이 시스템이 구현될 경우 종래의 목업 기반, 예제 기반 생성 기법의 한계를 극복할 수 있을 것이다. 특히 생성형 AI가 갖는 일종의 '기계적인(컴퓨터스러운)' 생성의 결과의 한계를 뛰어 넘고, 전문가들이 만들어 둔 템플릿을 기반으로 산업 현장에서 바로 사용할 수 있는 높은 품질의 결과를 기대할 수 있을 것이다[21]. 또 템플릿의 조화에 따라 무수히 다른 결과를 활용할 수 있고, 이것은 최근 급증하는 온라인 비즈니스의 수요를 충족할 수 있는 새로운 패러다임이 될 수 있으리라 여겨진다[22].

처음 시도에 따른 웹페이지 구성요소의 분류와 실제 섹션을 구성하는 단위요소의 분류체계는 딥러닝 모델의 실제 구현을 통해 많은 보완이 필요할 것이다. 우리가 사용하는 디지털 기기와 소프트웨어(SW) 연계 IoT 장비, 키오스크 등도 초기 일정한 규격을 바탕으로 사용성 기반의 산업적 발전과 창의적인 요소가 더해져 왔다[23,24]. 이렇듯 웹페이지 구성요소에 대한 표준은 웹디자인과 제작산업에 큰 영향을 미칠 것으로 기대된다.

게다가 최근 생성형 AI에 대한 연구가 활발히 진행되고 있지만, 단순한 원천 데이터(raw data) 기반의 생성과 단일 단락 수준의 개발에 그쳤던 부분을 어떻게 접근해야 하는지에 대해서도 고찰했다[25]. 다시 말해 실제 현업에서 활용할 수 있을 정도의 전체 완성물으로써 웹페이지를 제작할 수 있는 가능성을 제시했다. 프롬프트를 통해 사용자 의도를 수차례 반복해 입력해야 했던 현재의 방식에서 나아가, 완성물을 한 번에 제공하고 이를 분석하는 모델로 시스템을 설계한 것이다. 이로써 같은 모델로 미리 구성해 둔 데이터와의 매칭 기반 시스템이 운용되는 프레임워크를 적용할 수 있다는 학술적 시사점을 제공하게 되었다.

그럼에도 불구하고, 본 연구는 다음의 몇 가지 한계점을 갖고 있어 밝혀둔다. 가장 먼저, 제안하는 시스템은 전문

가 수준의 대규모 템플릿의 존재와 관리 시스템이 필수로 구현되어야 함을 전제로 한다. 이 템플릿의 품질에 따라 생성된 웹페이지의 품질이 결정되기 때문이다. 이것은 최근 인공지능의 연구와 산업의 발전이 컴퓨터에게 일을 시키고 높은 부가가치를 창출하고자 하는 등 IT산업의 방향 [26]과 부합하지 않는다. 즉, 여전히 사람의 개입이 많이 필요하므로 템플릿을 자동화 할 수 있는 추가적인 방안이 필요하다.

그리고 딥러닝 모델 기반의 벤치마크 웹사이트 분석은 사용자 설계(UI: User Interface) 측면의 디자인만 우선 고려되었다. 연구의 한계 상 기능적인 측면의 분석이나 의도가 분석되지 않아 생성된 페이지에 필요한 기능이 반영되지 못했다. 정적 웹페이지를 기반으로 분석하는 데에는 사용자 행위기반의 동적기능을 분석할 수 없으며, 일정한 패턴이 있더라도 오류율이 많은 부분으로써 후속 연구를 통해 지속적으로 보완해 나가야 할 것이다[27].

또, 딥러닝 모델이 각 단락과 요소의 탐지와 특징 분류 등에 있어 예상한 수준의 정확도를 보일 것이라 가정한 측면이다. 이를 기반으로 전체 시스템을 설계한 만큼, 모델을 학습시키기 위하여 대량의 데이터 수집 및 가공이 필요해 보인다. 실제 인공지능 기술의 활용에서 딥러닝 모델이나 알고리즘은 선행연구로부터 참조하므로 쉽게 접근이 가능하다. 하지만 이를 특정 서비스에 맞게 훈련하기 위해 요구되는 대량의 데이터 수집 및 가공이 큰 비중을 차지하는 현실의 한계와 일맥상통하는 부분이다[28].

마지막으로 웹페이지 단락과 구성요소의 분류체계를 기반으로 설계한 시스템은 전체 시스템의 설계를 제안하였으므로 후속연구를 통해 이를 실제로 입증해야 한다. 그리고 학술적, 산업적 가치가 확인을 위하여 사용성 평가를 통해 효과성 및 효율성을 측정해 볼 필요가 있다[29].

종합적으로 볼 때, 본 연구는 인공지능과 웹사이트 자동 생성 분야에서 절차적인 혁신을 추구하여 기여한 점이 크다. 동시에 산업의 실제 적용이 가능한 새로운 연구방향과 가능성을 제시했지만, 자동화된 웹 제작에 새로운 지평을 열어 가기에 많은 후속연구가 필요하다[30].

REFERENCES

- [1] ZHU, Meng; YANG, Yang; HSEE, Christopher K. The mere urgency effect. *Journal of Consumer Research*, Oct. 2018, 45.3: 673-690, DOI : <https://doi.org/10.1093/jcr/ucy008>
- [2] Park A., Lee, S-B, and Song, J. (2020). "Application of AI based

- Chatbot Technology in the Industry," *Journal of the Korea Society of Computer and Information*, July, 25(7), pp. 17-25, DOI : <https://doi.org/10.9708/jksci.2020.25.07.017>
- [3] Kim, R. Y. (2020). "The impact of COVID-19 on consumers: Preparing for digital sales." *IEEE Engineering Management Review*, 48(3): 212-218, DOI : <https://doi.org/10.1109/emr.2020.2990115>
- [4] Internetlivestats (2023). "Total number of Websites." Retrieved 10, 2023, from <https://www.internetlivestats.com/>
- [5] Parlakkilic, A. (2022). "Evaluating the effects of responsive design on the usability of academic Websites in the pandemic." *Education and Information Technologies* 27(1), pp. 1307-1322, DOI : <https://doi.org/10.1007/s10639-021-10650-9>
- [6] Muthukrishnan, Nikesh, et al. (2020) "Brief history of artificial intelligence." *Neuroimaging Clinics* 30(4), pp. 393-399, DOI : <https://doi.org/10.1016/j.nic.2020.07.004>
- [7] Xu, Y., Bo, L., Sun, X., Li, B., Jiang, J., & Zhou, W. (2021). image2emmet: Automatic code generation from web user interface image. *Journal of Software: Evolution and Process*, 33(8), e2369, DOI : <https://doi.org/10.1002/smr.2369>
- [8] Hashimoto, Y. and T. Igarashi (2005). Retrieving Web Page Layouts using Sketches to Support Example-based Web Design. *SBM*, DOI : <http://dx.doi.org/10.2312/SBM/SBM05/155-164>
- [9] Chang, K. S-P. and Myers B. A. (2012). WebCrystal: understanding and reusing examples in Web authoring. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. DOI : <https://doi.org/10.1145/2207676.2208740>
- [10] Baule, D., von Wangenheim, C. G., von Wangenheim, A., Hauck, J. C., and Junior, E. C. V. (2021). Automatic code generation from sketches of mobile applications in end-user development using Deep Learning. *arXiv preprint arXiv:2103.05704*. DOI : <https://doi.org/10.48550/arXiv.2103.05704>
- [11] Kaluarachchi, T. and M. Wickramasinghe (2023). "A systematic literature review on automatic Website generation." *Journal of Computer Languages*, 75, DOI : <https://doi.org/10.1016/j.col.2023.101202>
- [12] Mohammed, A., C. M. Rahman, and M. S. Abdulkarim. "The Scientific Comparison between Web-Based Site and Web-Builder (Open Source) Project: Functionalities, Usability, Design and Security." *International Journal of Scientific Research and Management (IJSRM)*, 6, DOI : <https://doi.org/10.18535/ijrm/v6i6.ec05>
- [13] WhichCmsToChoose.com (2016). "WordPress - CMS review, advantages and disadvantages." from <http://whichcmstochoose.com/wordpress.html>.
- [14] Forbes, J. (2019). "Web Builders and the Future of the Web Development Industry." DOI : <https://urn.fi/URN:NBN:fi:amk-2019100619610>
- [15] Nguyen, T. (2020). "Web design fundamentals: Web Builder software." *Australian Law Librarian* 28(3), pp. 165-167, DOI : <https://search.informit.org/doi/10.3316/agis.20210907053029>
- [16] Pillai, A., Shinohara, K., and Tigwell, G. W. (2022). Website Builders Still Contribute To Inaccessible Web Design. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility* (pp. 1-4), DOI : <https://doi.org/10.1145/3517428.3550368>
- [17] Ndukwe, W. (2019). "Website Builders: A Tool In Web Design From A Graphic Design Perspective." DOI : <https://urn.fi/URN:NBN:fi:amk-201904266105>
- [18] Newman, M. W. and J. A. Landay (2000). Sitemaps, storyboards, and specifications: A sketch of Web site design practice. *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques*. DOI : <https://doi.org/10.1145/347642.347758>
- [19] Huang, F., Canny, J. F., & Nichols, J. (2019). Swire: Sketch-based user interface retrieval. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-10), DOI : <https://doi.org/10.1145/3290605.3300334>
- [20] DBcut (2023). "shopdesign." from www.dbcut.com/shopdesign.
- [21] Lee, H. & Gu, H. (2022) "Delphi Research on Usability Test Framework of Metaverse Platform - Case of Roblox, Zepeto, and Gathertown". *Journal of the Korea Contents Association*, Oct, 22(9), pp. 179-193, DOI : <https://doi.org/10.5392/JKCA.2022.22.09.179>
- [22] Chai, B-S. Min, D-H, and Lee, H-J. (2023). "Research to Improve the Quality of Non-Contact Education Systems." *The Society of Convergence Knowledge Transactions*, March, 11(1), pp. 89-100, DOI : <https://doi.org/10.22716/sckt.2023.11.1.008>
- [23] Choi, H-S., Ko, Y-B., Min, D-H and Lee, H-J. (2023). "Research on Recognition Differences of PSD Features in KIOSK." *The Society of Convergence Knowledge Transactions*, June, 11(2), pp. 45-55, DOI : <https://doi.org/10.22716/sckt.2023.11.2.015>
- [24] Lee, H-E, Yoon, J-W. Park, Y-H, and Lee, H-J. (2023). "Expanding the Universal Lifelong Learning through Digital Education Engagements." *The Society of Convergence Knowledge Transactions*, December, 11(4), pp. 105-111. DOI : <https://doi.org/10.22716/sckt.2023.11.4.039>
- [25] Kim, H.J, Kim, Y.J, Yoon, D.H, and Lee, H.J. (2024). "Empirical Research on the Interaction between Visual Art Creation and Artificial Intelligence Collaboration." *The Journal of Convergence on Culture Technology*, January, 10(1), pp. 517-524, DOI : <https://doi.org/10.17703/jcct.2024.1.517>
- [26] Park, H. and Kim, K. (2023) "Convolutional Neural Network Model Using Data Augmentation for Emotion AI-based Recommendation Systems," *Journal of the Korea Society of Computer and Information*. December, 28(12), pp. 57-66, DOI: <https://doi.org/10.9708/JKSCI.2023.28.12.057>.
- [27] Lee, H-D, Kim, S-G, Na, S-C, Ham, J-Y, and Kwak, C-H. (2024)

"A Study on Traffic Vulnerable Detection Using Object Detection-Based Ensemble and YOLOv5," Journal of the Korean Society of Computer and Information, January, 29(1), pp. 61-68, DOI : <https://doi.org/10.9708/jksoci.2024.29.01.061>

- [28] Park, J.-S., Kim, S.-Y., Yoon, Y.-C., & Kim, S. K. (2023). "Optimizing CNN Structure to Improve Accuracy of Artwork Artist Classification," Journal of the Korea Society of Computer and Information, September, 28(9), pp. 9-15, DOI : <https://doi.org/10.9708/jksoci.2023.28.09.009>
- [29] Lee, H., Kwon, S. and Min, D. (2021) "The Empirical Research on the User Satisfaction of Mobile Grocery Shopping Customer Journey," Journal of Information Technology Applications and Management, August, 28(4), pp. 59-78, DOI : <https://doi.org/10.21219/jitam.2021.28.4.059>
- [30] Park, S., Lee, Y., Lee, H.-J. (2024), "Research on Enhancing Customer Experience through AI-Supported Review Generation," Transaction of the Korean Institute of Electrical Engineers, 73(2) pp. 334-342, DOI : <https://doi.org/10.5370/KIEE.2024.73.2.13>

Authors



Jung-Hwan Kim received the B.S. degree in Management from Korea University in 2009 and the Ph.D. degrees in Digital Management (DDM) from Korea University, Korea in 2024 respectively.

Dr. Kim opened the cutting edge business in Generative AI powered solution, as named "Weven". He is currently the founder and CEO of the Weven Corp. since 2022. He is interested in Deep Learning, Artificial Intelligence, Automatic Process Engineering Start-ups and Entrepreneurship.



Young-beom Ko is currently Ph.D candidates in the Department of Digital Management (DDM), Korea University from 2019. He founded IT company "Myeongseok" and had been carrying out various SI projects.

His main interests are the ERP, e-Commerce, Big Data, Cryptocurrency and Blockchain.



Jihoon Choi is currently Ph.D candidates in the Department of Digital Management (DDM), Korea University from 2021. He is currently in charge of the Motorad Sales Branch in Kolon Mobility Group.

He is interested in Digital Mobility, Brand Marketing, e-Commerce and ESG Management.



Hanjin Lee received the B.S. in Sociology, M.S. degree in Communication from Yonsei University, Korea in 2006 and 2011. Furthermore he received the Ph.D. degrees in Digital Management, Korea University, 2021.

Dr. Lee's research interests are focused on AI biz innovation, as well as the intersection between user behavior and digitalization. He has gained 16 years experience in the e-Commerce industry (Naver, eBay, and Coupang Corp.)