

빅데이터를 이용한 중고 거래 분석 시스템 연구

안병태
안양대학교 교양대학 교수

A Study for Used Transaction Analysis System using Big Data

Byeongtae Ahn
Professor, Liberal & Arts College, Anyang University

요 약 최근 중고 거래를 지원하는 중고 거래 사이트가 증가함에 따라 사용자는 실시간으로 다양한 정보를 검색하고자 한다. 이러한 새로운 변화는 전자상거래 기반에서 새로운 유형의 C2C(Commerce to Commerce) 거래가 활성화 되었다. 그러나 각각의 중고 거래 사이트는 고유한 특성들이 있어 전체 표준화가 어렵다. 따라서 본 논문에서는 사용자가 사용한 거래 데이터를 실시간으로 제공하고 원하는 정보를 신속하게 제공하는 시스템을 연구하였다. 본 논문에서는 인터넷 전자 상거래를 통한 중고품 통합 거래 시스템 개발에 필요한 크롤러 시스템을 연구하고, 정의된 형태소 분석기를 통해 사용자가 원하는 웹 환경에서 정보 제공이 가능하도록 하였다. 따라서 본 연구에서는 다양한 중고 물품 사이트에 접속하지 않고도 사용자가 원하는 정보를 제공하는 시스템을 설계하였다.

주제어 : 빅데이터, 전자상거래, 데이터 분석, 중고 거래, 개인간거래

Abstract Recently, as the number of used trading sites supporting used trading increases, users want to search for a variety of information in real time. This new change has enabled a new type of C2C (Commerce to Commerce) transaction in the e-commerce base. However, since each used trading site has its own characteristics, it is difficult to standardize the whole. Therefore, in this paper, we studied a system that provides the transaction data used by the user in real time and provides the desired information quickly. In this paper, we researched the crawler system necessary for the development of the integrated trading system for used goods through Internet e-commerce, and made it possible to provide information in the web environment desired by the user through the defined morpheme analyzer. Therefore, in this study, we designed a system that provides information desired by users without accessing various used goods sites.

Key Words : Big Data; e-Commerce; Data Analysis; Used Transaction; C2C

1. 서론

인터넷 사용이 보편화되면서 온라인을 기반으로 하는 전자상거래가 일상화되었다. 전자 상거래는 주로 기업의 주도로 이루어지는 B2C의 형태로 이루어져있다. 하지만, 인터넷의 발전으로 기업이 소비자와 직접 연결될 수 있

도록 할 뿐 아니라 소비자들 간 연결의 기틀을 마련하며 새로운 거래 유형 C2C 창출에 까지 영향을 미쳤다. 이러한 변화로 인하여 온라인상에서 소비자의 역할은 제품을 판매하는 판매자의 역할 또한 수행하며 그 영역이 점차 확대되어 가는 추세가 되어갔다. 그 결과 소비자가 제품을 구매하는 방법이 점점 다변화 되고, 더욱 저렴하게 제

*Corresponding Author : Byeongtae Ahn(ahnbt@anyang.ac.kr)

Received April 6, 2021
Accepted June 20, 2021

Revised April 30, 2021
Published June 28, 2021

품을 구매하기 위해서 온라인 중고 거래 사이트를 이용하는 구매자들이 꾸준히 증가하고 있다. 특히, 현재 저 성장과 소비 침체가 장기화하면서 중고품 경제(Second-hand Economy)가 급성장하고 있다[1]. Fig. 1.은 2017년부터 2018년까지 소비자 심리지수와 중고 거래 플랫폼의 매출 규모를 나타낸 표이다. 꺾은선 그래프는 소비자 심리지수이며 오른쪽 지표를 참고해서 확인한다. 주황색 막대그래프는 중고나라, 초록색 막대그래프는 번개장터의 매출액이며 왼쪽 지표를 참고한다. 매출액의 규모는 100만 달러 기준이며 소비자 심리지수는 크게 감소를 하는 반면 중고거래 매출 규모는 기하급수적으로 커지는 것을 확인할 수 있다.

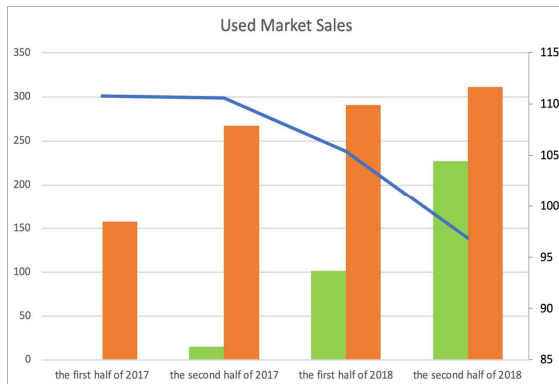


Fig. 1. Consumer sentiment index and used market sales graph

중고 시장은 개인간 거래가 많이 시장 규모를 정확하게 집계하기 어렵다. 하지만 유통업계에서 추산하는 규모는 중고 시장의 대표적인 중고차 시장을 제외하고도 20조원대로 추산하고 있다. 그렇기 때문에 잠재력이 높은 시장으로 평가되며, 모바일 앱 론칭, 상품 검수와 같은 서비스를 강화하면서 중고시장에서 새로운 수익을 창출하려는 움직임이 많아지고 있다. 이러한 시장 추세로 인하여 G마켓, 옥션과 같은 기존의 온라인 전자 상거래를 제공하는 대형 플랫폼에서도 중고거래 시장을 겨냥한 서비스를 개발하여 제공하며, GPS 정보를 이용하여 지역 기반의 중고 상품을 제공하는 당근마켓과 같이 많은 중고 거래 플랫폼이 등장하였다. 본 연구에 앞서 다양한 중고 거래 서비스를 분석한 결과 상품이 서비스 별로 분산되어 존재하는 점을 가장 큰 문제점으로 잡았다. 대표적으로 ‘맥북 프로 2019’를 검색했을 때 극단적으로 상품의 개수가 1개밖에 나오지 않는 경우도 존재하였다. 또한, 가격 분포도도 서비스 별로 상품의 가격이 평균 10만원 정도

차이가 있었고, 최대 30만원이 넘게 차이가 나는 경우가 있었다. 이를 해결하기 위해서 하나의 웹 사이트에 모든 중고거래 상품을 조회할 수 있는 통합 플랫폼을 구축하였다[2].

본 연구의 2장에서는 웹 서비스 구성을 위한 표준과 관련 연구를 알아보고, 3장에서 국내.외 사례를 알아본다. 4장에서는 중고 거래 통합플랫폼 시스템 설계도를 제안하고 5장에서는 결론 및 향후과제를 제시한다.

2. 관련연구

본 연구에서 제시하는 방안은 다른 중고거래 웹 서비스에 상품을 크롤링해와 통합 플랫폼에서 조회 및 검색을 할 수 있는 서비스를 개발하는 것이다. 단순히 보이지만 하루 평균 60만개 이상의 데이터를 처리해야하며, 소비자들에게 실시간으로 상품을 보여줄 수 있는 서비스를 구성하였다[3]. 본 연구에서는 자바스크립트 언어를 이용하여 웹 시스템의 클라이언트와 서버를 구성하였으며, 파이썬 언어를 이용하여 크롤러와 데이터 분석을 구현하였다. 마지막으로 데이터베이스로는 json 타입을 지원하는 mongodb를 이용하였다.

2.1 웹 시스템 구성

클라이언트 기반 웹 시스템의 구성 방법에서는 서버 기반의 구조에서 사용한 톰캣과 같은 웹 어플리케이션 서버와 엔진엑스와 같은 웹 서버의 기능을 대신하기 위해 저 사양의 서버에서도 동작이 가능한 노드제이에서 서버를 사용하였다[5]. 웹 어플리케이션은 사용자 인터페이스를 만들기 위해 사용되는 오픈소스 프로그래시브 자바스크립트 프레임워크의 한 종류인 뷰(Vue.js)를 사용하였다. 뷰는 다른 Angular, Backbone, React 자바스크립트 프레임워크보다 매우 작고 가벼우며 복잡도도 낮다. 코어 라이브러리는 선언형 렌더링과 컴포넌트 구성에 초점을 두며 기존 페이지에 임베드가 가능하다. 라우팅, 상태관리, 빌드 도구화와 같이 복잡한 어플리케이션에 필요한 고급 기능들은 공식적으로 유지 보수되는 지원 라이브러리와 패키지를 통해 제공된다[6]. 뷰(Vue.js)는 일반적인 다른 자바스크립트 프레임워크에서 사용하는 MVC(Model-View-Control) 패턴과 다르게 MVVM(Model-View-ViewModel) 패턴을 사용한다. 이 MVVM 패턴은 MVC 패턴의 변형으로 뷰(View)의

추상화를 만드는 것이 핵심이다. 뷰(View)의 추상화는 재사용할 수 있고 테스트하기 쉽다. 뷰(View)의 추상화를 통해 응용프로그램 구조는 단순해지고, 이상적으로, 시각 디자인과 표현 논리를 독립적으로 구현할 수 있다. 즉, MVC 패턴에서 컨트롤러 역할처럼 데이터를 관리하고 액션을 처리하는 역할을 담당한다. Fig. 2는 뷰모델 인스턴스가 처리하는 작업들이다.

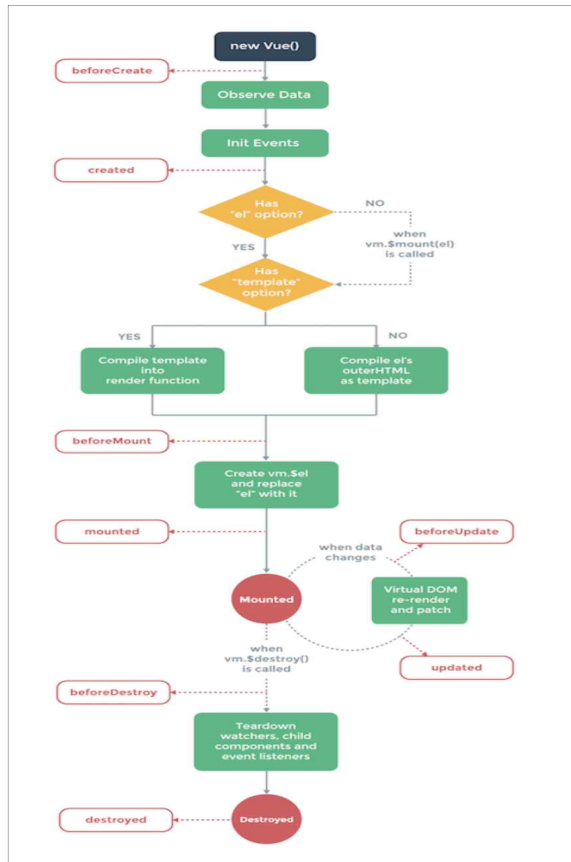


Fig. 2. Flow Chart of View Model Instance

기존의 웹 페이지 개발 방식에서는 HTML 파일을 기본으로 하여 자바스크립트를 HTML 파일에 포함시켜 사용하였으나 본 연구에서는 먼저 자바스크립트 라이브러리를 기본으로 실행시킨다. 이후 웹 페이지 생성을 위한 자바스크립트 파일에서 필요한 라이브러리를 결합시켜 데이터를 공유할 수 있게 하고, 템플릿 파일에 존재하는 HTML 데이터를 로딩하여 사용자에게 제공하는 웹 페이지를 만들어 낸다. 따라서 개별적 라이브러리의 추가 로딩이 필요가 없으며, 데이터의 공유가 가능하고, 템플릿 파일의 변경만으로 웹 페이지를 변경할 수 있다. 이러한 구조는 여러 자바스크립트 라이브러리의 결합을 용이하

게 하고 하나의 틀 안에서 동작할 수 있게 한다[7].

2.2 크롤링 구현

최근 빅데이터 관련 직종의 인기도와 수요도 증가하고 있다. 그 결과 다양한 데이터 관련 직종과 업무가 생겨나는데 대표적으로 데이터 분석과 관리, 데이터 분석 서버 관리, 데이터 분석 기획, 데이터 분석 개발, 리뷰 크롤러 개발, 텍스트 분석, 웹 크롤러 개발, 크롤링 데이터 분석, 데이터 수집 시스템 개발 및 크롤링 개발 등의 업무를 있다[8]. 본 연구에 필요한 중고 상품 데이터를 기업의 DB에 접속하여 쉽게 명령어 한 줄로 원하는 데이터를 가져올 수 있는 계정 정보가 있다면 문제가 없겠지만 대부분이 그렇듯이 기업의 보안과 고객 정보 보호차원에서 기업의 데이터는 보호 받고 있다. 이러한 데이터는 다양한 특성을 갖고 있다[9].

2.3 데이터 분석 구현

중고거래 상품들은 모두 개인 소비자들의 위주의 게시글이라 모두 비정형 텍스트다. 이 텍스트를 전처리 하기 위해서는 자연어 처리 기술이 필요하다. 한글로 작성된 게시글에서 기술 개체 식별을 위해서는 교착어의 특징을 갖는 한국어의 특징을 고려할 필요가 있다. 또한, 실시간 서비스에 알맞은 빠른 처리 속도가 필요했기 때문에 본 연구에서는 중고 상품에 적용 가능한 한국어 형태소 분석기(OKT, MeCab, Komoran, KKma)의 성능을 비교 평가하고 분석하였고 그 결과가 그림 4에 있다. 좌측에 시간은 초단위로 진행을 했다[10]. Fig. 3은 형태소 분석기 성능 비교 그래프를 나타낸 것이다.

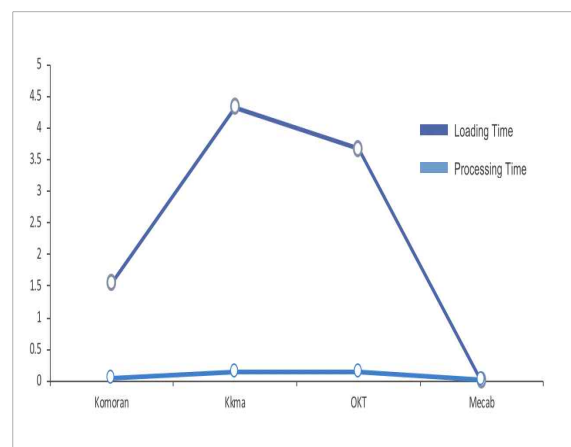


Fig. 3. Morphological Analyzer Performance Comparison Graph

실험은 하나의 중고 상품 게시글에 대하여 평균적인 라이브러리 로딩시간과 처리시간을 표시하였다. X축은 각각의 형태소 분석기 라이브러리이며, y축은 품사 태깅 시간 및 라이브러리 로딩 시간이다[11]. 본 연구에서 개발한 시스템은 실시간으로 데이터를 업데이트하여 사용자에게 보여줘야 하는 시스템으로 처리속도가 가장 중요했다. 평가 분석한 형태소 분석기들의 성능을 비교 평가한 결과 MeCab 분석기가 압도적으로 빠른 처리속도를 보여주었으며 추출된 색인어를 비교한 결과 분석기들 간에 비슷한 결과를 도출함을 확인하였다. 하지만 다른 형태소 분석기와 마찬가지로 최근에 새로 만들어진 신조어 또는 고유명사가 포함된 문장 등은 제대로 분석해 내지 못하는 문제점이 있다. Fig. 4는 Mecab 분석 결과를 나타낸 것이다.



Fig. 4. MeCab Analysis Results

연구에서 원하는 데이터는 ‘맥북에어’ 라는 데이터지만 한국어 문법 특성때문에 ‘-에’ 단어와 ‘-어’ 단어가 각각 부사격 조사와 동사로 분류되고 있다. 대부분의 검색 엔진과 알고리즘에는 Mecab이 적합하지만 본 연구에서 사용하는 중고 상품 데이터 조회 및 분석에는 적합하지 않아 Mecab 형태소 분석기를 보완하였다.

3. 국내.외 사례

3.1 국내 사례

* 중고나라 : 총 누적 2100만명의 회원을 보유한 국내 최대 중고거래 플랫폼으로써 지난 2015년까지 자체 플랫폼을 가지지 않고 네이버라는 국내 최대 포털의 카페 서비스를 활용하여 운영해 왔다. 자체 앱을 개발하여 플랫폼 간의 상품 공유를 통해 사용자를 앱으로 유입시켰다.

* 당근마켓 : ‘당신 근처의 중고 직거래 마켓’이라는 정

체성을 갖고 2015년 출시하였다. 이전까지의 중고거래 플랫폼과는 다르게 사용자의 위치에 기반을 둔 서비스이다. 사용자가 거주하고 있는 지역을 등록하면 지역에서 실시간으로 거래되고 있는 물품을 확인할 수 있다.

* 다나와 : 2000년 디지털 카메라 가격 비교 서비스를 제공하는 서비스로 시작하여 이후 컴퓨터 부품 시세 정보도 함께 제공하기 시작하였다. 현재는 백색가전외에 자동차, 중고 제품 등도 함께 제공한다. 소비자들이 주로 이용하는 오픈 마켓의 시세를 모두 제공해주기 때문에 사용자들이 물품구매전 가격 파악에 용이하다.

3.2 국외사례

* Craigslist : 중고물품 뿐만 아니라 집과 구인광고 등도 올라오는 사이트로써 1995년 미국 샌프란시스코에서 첫 서비스를 시작하였고, 2000년 다른 미국도시들로 확장, 현재는 50개 국가에서 운영되는 서비스이다. 지역을 선택하고 물품을 검색하여 자신이 원하는 물품을 찾는 방식이다.

* 아마존 : 구매자로부터 반품된 like-new, open-box, pre-owned 상품들을 할인하여 재판매하는 아마존의 서비스로써 아마존이라는 큰 기업에서 자체적으로 판매하는 중고 상품이다보니 소비자는 좀 더 신뢰를 갖고 구매할 수 있다.

* 이베이 : 다국적 전자상거래 기업. C2C와 B2C 판매를 증대한다. 경매 방식의 판매와 즉시 구매 형식의 판매 방식을 제공한다. 구매자는 수수료를 내지 않고, 판매자만 일정 개수 이상의 물품을 판매할 때 수수료를 부과한다.

4. 시스템 설계

Fig. 5는 우리 프로젝트 기획의 1단계 시스템 구성도를 나타낸 것이다. 3개 사이트를 대상으로 크롤러 봇을 일정 시간 간격으로 호출하여 데이터를 수집한다. 수집된 데이터는 데이터가 유효한 데이터인지 확인하고 유효하지 않은 데이터를 제외하고 데이터 분석 과정을 거친다. 이 과정에서 미리 정해놓은 카테고리에 해당하게 되면 각 카테고리로 분류를 하고 필요한 정보를 추출하여 데이터베이스에 기록한다.

크롤러 봇은 1분 간격으로 각 사이트를 크롤링 하며, 이전에 기록된 게시글 이후부터 가장 최신 게시글을 가져온다. 하나의 크롤러 봇은 하나의 사이트를 담당하며

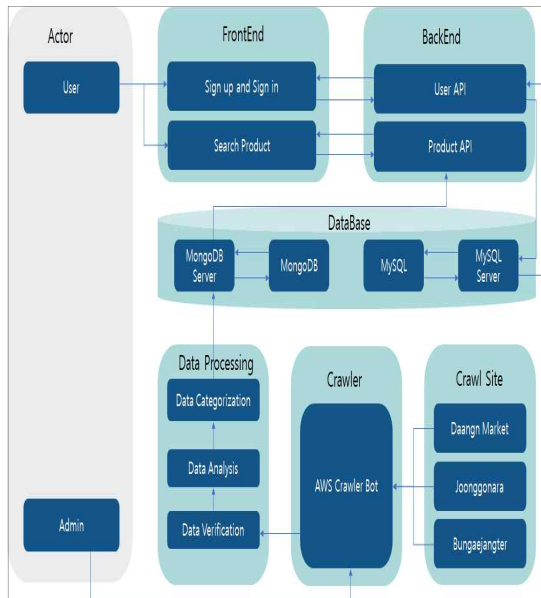


Fig. 5. System Structure Diagram

가장 최신글 인지를 판별하여 이를 기록하고 종료한다. 크롤링된 모든 글을 데이터 처리과정으로 전달하지 않고 1차적으로 유효한 데이터인지를 검증한다.

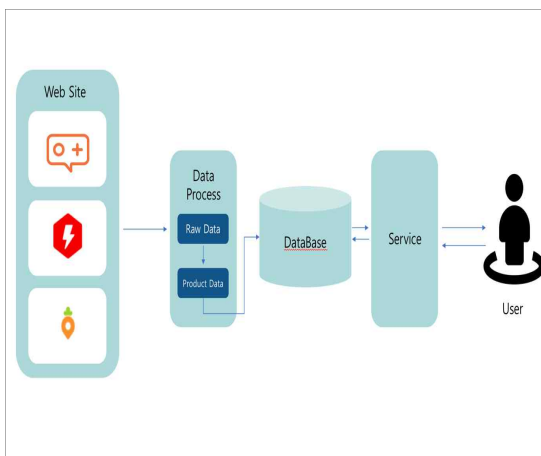


Fig. 6. Portal System Flow Diagram.

Fig. 6은 포털 시스템 흐름도를 나타낸 것이다. 웹에서 크롤링을 통해 데이터를 수집하고 수집된 데이터 중 일부는 생산 데이터로 저장된다. 저장된 데이터는 데이터 분석기를 통해 사용자에게 필요한 정보로 제공된다.

Fig. 7은 데이터 저장 시스템을 나타낸 것이다. 수집된 데이터는 데이터 처리과정을 거쳐 무의미한 데이터와 의미 데이터로 분류되어 데이터 저장소에 저장된다. 1차적으로 필터링된 데이터에 대하여 추가적인 검증을 통해 불필요한 것을 걸러냅니다. 글의 제목, 본문과 카테고리

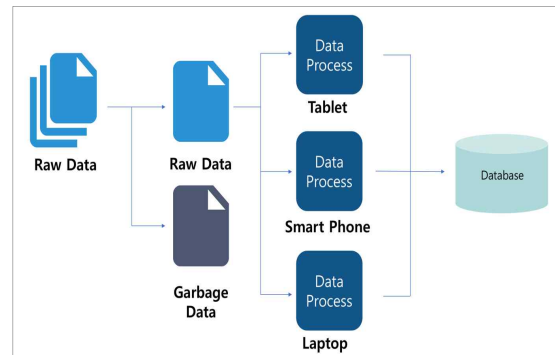


Fig. 7. Data Storage System

분류를 위해 데이터 셋의 일치를 통해 각각에 맞는 카테고리에 분류하게 된다. 각 카테고리마다 필요로 하는 세부 옵션 추출하는 과정을 수행한다[14]. 사용자는 카테고리 뿐만 아니라 자신이 원하는 옵션에 맞는 데이터를 검색할 수 있어야 하므로 주요 옵션을 추출하는 과정을 필요로 한다. 카테고리 분류, 가격과 세부 옵션을 추출한 글은 데이터 처리 과정을 통해 알게된 정보와 원시 데이터를 모두 MongoDB에 저장한다. MongoDB는 RDBMS의 table이 아니라 collection 단위로 JSON형태의 document를 저장한다[15]. 이와 별개로 유저 정보를 관리하는 DB를 두어야 하므로 이는 RDBMS인 MySQL을 사용한다. REST API 서버 작성을 위해 Node.js를 이용하였다[9, 10]. 이 시스템에선 단순히 DB에서 데이터를 가져오는 저비용 행위가 많이 이루어질 것이라는 점에서 해당 프레임워크를 사용한다. 직접적인 판매글 작성은 제공하지 않기 때문에 해당 서버에서 가장 많이 이루어지는 작업은 게시글 목록을 read하는 행위이고, 주로 JSON형식의 데이터를 다룬다. 웹 프론트 부분은 vue.js를 이용하여 개발하였다[16]. 사용자는 자신이 찾고자 하는 상품의 카테고리를 선택한 후 원하는 스펙과 가격대를 선택함으로써 세세한 옵션을 검색할 수 있게 한다.

5. 결론 및 향후 과제

본 연구에서는 인터넷 전자상거래를 통한 통합 중고 거래 시스템 개발에 필요한 크롤러 시스템을 설계하였다. 그리고, 형태소 분석기를 정의하여 웹 환경에서 사용자들이 조회할 수 있는 서비스에 대해서 연구하였다. 이러한 과정에 필요한 중고거래 크롤링 API는 추후 시스템 구축 시 많은 도움이 될 것이다. 향후 과제로는 이러한 시스템 설계를 기반으로 시스템을 구축할 것이다.

REFERENCES

- [1] R. J. Gordon. (2020). Does the new economy measure up to the great inventions of the past? *J. Econ. Perspect.* 14, 49-74.
- [2] B. Carlsson. (2019). The Digital Economy: What Is New and What Is Not? *Struct. Chang. Econ. Dyn.* 15, 245-264.
- [3] R. Armagan, Yeni Ekonomi ve Türkiye. Suleyman Demirel Üniversitesi IIBF Dergisi 2000, 5, 139-153.
- [4] Akyazi, H.; Kalça, A. Yeni Ekonomi ve İktisat Bilimi. *Liberal Düşünce Dergisi* 2019, 29, 221-242.
- [5] BARIŞIK, S.; Yirmibeşçik, O. Türkiye'de Yeni Ekonomi'nin Olusum Surecini Hızlandırmaya Yonelik Uyum Cabaları. *ZKU Sosyal Bilimler Dergisi* 2020, 2, 39-62.
- [6] Viskari, S.; Pekka, S.; Marko, T. Implementation of Open Innovation Paradigm, Cases: Cisco Systems, Dupont, IBM, Intel, Lucent, P&G, Philips and Sun Microsystems; Lappeenranta University of Technology Research Report 189; Lappeenranta University of Technology: Lappeenranta, Finland, 2020.
- [7] Conboy, K.; Mikalef, P.; Dennehy, D.; Krogstie, J. Using business analytics to enhance dynamic capabilities in operations research: A case analysis and research agenda. *Eur. J. Oper. Res.* 2020, 281, 656-672.
- [8] Mikalef, P.; Boura, M.; Lekakos, G.; Krogstie, J. Big data analytics capabilities and innovation: The mediating role of dynamic capabilities and moderating effect of the environment. *Br. J. Manag.* 2019, 30, 272-298.
- [9] Taylor, T. Thinking about a new economy. *Public Interest* 2001, 24, 3-19.
- [10] Addo-Tenkorang, R.; Helo, P.T. Big data applications in operations/supplychain management: A literature review. *Comput. Ind. Eng.* 2020, 101, 528-543.
- [11] Huang, B.; Jin, L.; Lu, Z.; Yan, M.; Wu, J.; Hung, P.C.; Tang, Q. RDMA-driven MongoDB: An approach of RDMA enhanced NoSQL paradigm for large-scale data processing. *Inf. Sci.* 2019, 502, 376-393, doi:10.1016/j.ins.2019.06.048.
- [12] Schäffer, E.; Mayr, A.; Fuchs, J.; Sjarov, M.; Vorndran, J.; Franke, J. Microservice-based architecture for engineering tools enabling a collaborative multi-user configuration of robot-based automation solutions. *Procedia CIRP* 2019, 86, 86-91.
- [13] Fabian, K.; Philipp, B. Return of the JS: Towards a Node.js-Based Software Architecture for Combined CMS/CRM Applications. *Procedia Comput. Sci.* 2020, 141, 454-459.
- [14] Boran, F.E. Genç, S.; Kurt, M.; Akay, D. A multi-criteria intuitionistic fuzzy group decision making for supplier selection with TOPSIS method. *Expert Syst. Appl.* 2021 36, 11363-11368.
- [15] T. J. Barker & Z. B. Zabinsky. (2011). A multicriteria decision making model for reverse logistics using analytical hierarchy process. *Omega*, 39(5), 558-573.
- [16] Zheng Xu. (2017). The analytics and applications on supporting big data framework in wireless surveillance networks, *International Journal of Social and Humanistic Computing, Volans* 2(3), 141-149.

안 병 태(Byeongtae Ahn)

[정회원]



· 1999년 2월 : 국민대학교 컴퓨터과학부(이학사)

· 2006년 8월 : 경상대학교 컴퓨터과학부(공학박사)

· 2012년 3월 ~ 현재 : 안양대학교 교양대학 컴퓨터전공 교수

· 관심분야 : 블록체인, 전자상거래, 암호

화폐, 스마트 컨트랙트

· E-Mail : ahnbt@anyang.ac.kr