

C2C에 기반으로 해외직구 불법거래에 관한 모니터링 시스템 설계 및 분석

신용훈¹, 김정호^{2*}

¹한밭대학교 컴퓨터공학과 박사과정, ²한밭대학교 컴퓨터공학과 교수

Design and analysis of monitoring system for illegal overseas direct purchase based on C2C

Yong-Hun Shin¹, Jeong-Ho Kim^{2*}

¹PH.D. Student, Dept. of Computer Engineering, Han-bat National University

²Professor, Dept. of Computer Engineering, Han-bat National University

요 약 본 논문은, 개인 간 거래 C2C에 기반으로 해외직구 불법 거래에 관한 모니터링 시스템을 제안한다. 관세법에서는 해외직구 물품이 일정 금액 (미화 150불, 단 미국은 미화 200불)이하 또는 자가 사용 물품으로 인정되는 경우에만 제세를 면제토록 규정하고 있다. 과세를 면제받아 구매한 해외직구 물품을 온라인 등에서 되파는 행위는 무신고 밀수입 죄에 해당한다. 그런데도 온라인 중고 사이트에는 이에 대한 되팔이가 증가하여 지속적인 관세법 위반이라는 사회적 이슈로 논란이 되고 있다. 따라서 본 연구에서는 해외직구 관련 불특정 거래 내용을 수집하고, 정보를 빅데이터 방식으로 데이터를 정제하여, 자연어 처리 등을 통해 모니터링 시스템으로 설계하여 판매자와 유사한 형태를 보이는 키워드 분석, 거래방식 분석, 동일성 판별 등을 분석하였다. 해외직구 물품의 불법 거래 단속에 활용이 가능할 것이다.

주제어 : 빅데이터, 융합, 해외직구, 데이터 분석, 데이터마이닝, 크롤링

Abstract In this paper, we propose a monitoring system for illegal overseas direct purchase based on C2C transaction between individuals. The Customs Act stipulates that direct purchases from overseas are exempted from taxation only if they are less than a certain amount (US\$150, but US\$200 in the US) or are recognized as self-used goods. The act of reselling overseas direct purchase items purchased with exemption from taxation online, etc., is a crime of smuggling without a report. Nevertheless, the number of re-sells on online second-hand websites is increasing, and it is becoming a controversial social issue of continuous violation of the Customs Act. Therefore, this study collects unspecified transaction details related to overseas direct purchase, refines the data in a big data method, and designs it as a monitoring system through natural language processing, etc. analyzed. It will be possible to use it to crack down on illegal transactions of overseas direct purchase goods.

Key Words : Big Data, Convergence, Overseas direct Purchase, Data analyze, Data mining, Crawling

*Corresponding Author : Jeong-Ho Kim(jhkim@hanbat.ac.kr)

Received March 21, 2022

Accepted May 20, 2022

Revised May 9, 2022

Published May 28, 2022

1. 서론

4차 산업 혁명의 변화에 따라 모바일 컴퓨팅의 발달, 소셜 미디어의 정보의 폭발적 증가로 인해 거대한 융합 데이터의 세상 속에 살고 있다. 특히 코로나19의 영향 등으로 비대면 온라인 쇼핑이 더욱 활발해지면서 온라인 쇼핑물의 플랫폼 중심으로 경제가 급성장하고 해외직구도 편하게 쇼핑하는 사례가 증가하고 있고, 해외 물품을 선호하는 사회적 분위기가 이용자들을 중심으로 형성되었다. 또한 그 사회적 분위기를 악용한 리셀러(되팔이)들이 등장하여 세계를 면제하여서 구입한 해외직구 물품을 온라인 오픈 마켓 또는 온라인 중고 사이트에서 개인 간 되파는 불법 거래가 성행하고 있다. 관세법에서는 해외직구 물품이 미화 150불, 단 미국은 미화 200불 이하이고, 자가 사용 물품으로 인정되는 때에만 관세 등 세제를 면제하도록 규정하고 있다. 과세를 면제받아 구입한 해외직구 물품을 온라인 등에서 되파는 행위는 관세법 제 269조 밀수입죄, 제270조 부정수입죄, 제270조 부정 감면죄로 무신고 및수입 죄에 해당하게 되어 같은 법 위반으로 처벌받을 수 있다.

해외직구 물품이 자주 거래되는 중고 사이트를 중심으로 전문적인 온라인 되팔이가 등장하여 2018년 이후 서울본부세관에서 '해외직구 물품 되팔기' 근절을 위해 재판매 거래가 활발한 인터넷 커뮤니티 등을 지속적인 모니터링을 시행해 단속 활동을 강화하고 있다[1]. 그런데도 면세혜택을 받고 들여온 제품을 국내에서 온라인 중고 사이트와 모바일 오픈 마켓을 통해 되파는 리셀러들이 '21년 9월에도 적발되었으며 이중 상습 전문 판매자 6명은 관세법 위반으로 입건되는 등 사회적으로 매우 물의를 일으키고 있다[2][3].

본 연구는 우선 온라인 시장 플랫폼의 변화에 따라 해외직구 물품이 자주 거래되는 온라인 중고 사이트를 중심으로 개인 간 거래되는 C2C (Commerce to Commerce) 해외직구 물품 정보 등을 크롤링을 수행하여 데이터를 수집한다.

임시 저장소에 수집된 데이터를 모아 유의미한 데이터와 무의미한 데이터를 정제 후 키워드 분석, 판매패턴 분석, 동일성 분석 등을 할 수 있는 해외직구 물품 불법 거래 모니터링 시스템을 설계하고 관계성을 해석하였다. 따라서 해외직구 물품의 리셀러(되팔이)의 불법거래 단속과 사이버 불법 무역 방지 등에 활용이 가능할 것이다.

2. 관련 연구

2.1 해외직구 관련

관세청에서 제공하는 통계 지표에 의한 2018년에서 2020년까지 3년간 전자상거래물품 해외직구 수입통관 현황에 대한 증감률을 Table 1에 나타내었다[4].

Table 1. E-commerce import customs clearance

Year	number	rate of changeA	price	rate of changeA
2018	3,226	36.8million	2,755%	30.6 million dollars
2019	4,299	33.3million	3,143%	14.1 million dollars
2020	6,357	47.9million	3,753%	19.4 million dollars

해외직구의 수입 현황이 연간 지속해서, 대폭 증가한 이유는 달러 및 엔화 약세로 환율 하락, 해외직구의 저변 확대, 광군절, 블랙 프라이데이 등 글로벌 할인 행사 등과 함께 SNS의 활성화, 물류배송 안정화와 전자결제 편의성 등으로 나타나고 있다[5]. Fig. 1은 2016년 ~ 2020년의 5년 동안, 전자상거래 물품 수입 동향을 나타내고 있으며 해마다 건강식품, 가전제품, 의류 순으로 증가하고 있다[6].

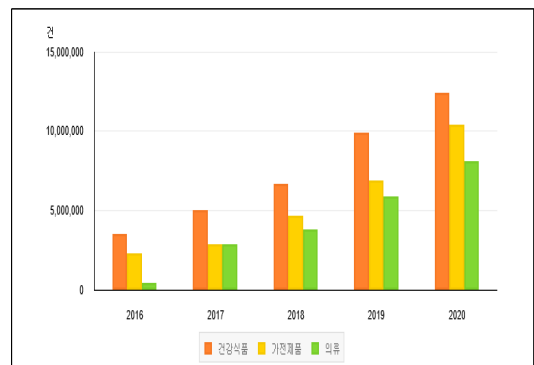


Fig. 1. E-commerce Product Trends

2.2 해외직구 통관제도

해외직구 물품 통관 방법으로는 목록통관과 일반 통관이 있다. 목록통관은 물품 가격, 미화 150불 이하이며, 개인이 자가 사용 목적으로 수입하는 경우에만 특별수송업체의 통관목록 제출만으로 수입신고가 생략, 관세 및 부가세가 부과되지 않는 제도이다. 일반 통관은 목록통관이 되지 않은 수입신고 건을 대상으로 관세사를 통해

세관장에게 수입신고 후 통관되는 제도로 관세 및 부가세를 부과한다[7].

해외직구에 의한 물품 주문, 대금결제 또는 일부가 전자적 방식 또는 온라인을 이용하여 사이버 공간에서 이루어지는 거래로 세 가지 유형으로 구분한다.

첫째는 소비자가 해외 온라인 쇼핑몰에서 제품을 직접 거래한 후 직접 배송을 받는 직접 구매 형태가 있고, 둘째는 해외 판매자가 국내까지 직접 배송하지 않는 물품, 소비자가 배송량이 많아 국내 소비자가 온라인 쇼핑몰에서 직접 구매한 물품을 해외 배송 대행사로 보낸 후 배송 대행업체를 통해 국내로 배송을 받는 형태가 있다.

마지막으로 소비자가 해외 온라인 쇼핑몰에의 모든 행위를 구매 대행업체에 의뢰하여 물품 가격, 물류비, 수수료 등을 지불하고 모든 절차를 위임하는 구매대행이 있다.

2.3 해외직구 중고 거래의 증가

온라인 중고 거래의 증가에 따라 사용자는 실시간으로 해당 플랫폼에 대하여 검색하여 중고 거래 사이트를 이용하는 사람들이 꾸준히 증가하고 있다[8]. 중고 거래의 사례로 중고나라는 네이버 카페에서 2003년 온라인 서비스를 시작하여 2021년 현재 1,800만 명이 넘는 회원 규모를 가지고 중고 거래 플랫폼으로 자리 잡고 있다[9]. 또한, 모바일 기기, SNS, 사물인터넷 등 기술의 발달과 확산으로 인하여 정보의 양과 그 종류가 매우 많아져 빅데이터로 발전하고 있다. 빅데이터는 신뢰성 있는 데이터를 제공하여 해당 기반으로 만들어진 시스템은 더 정확하고 신뢰성 있는 정보를 제공하며 빠르게 선택할 수 있도록 하고 있다[10, 11].

2.4 데이터마이닝

대용량으로 저장된 데이터를 체계적이고 자동으로 통계적 규칙 또는 패턴을 분석하는 방식이다. 데이터로부터 분석에 필요한 데이터를 추출하고 저장한 빅데이터를 정제하여 필요한 데이터 변경한 후 저장한 빅데이터를 자동으로 통계적 규칙 또는 수치 패턴을 분석하여 가치 있는 정보를 Fig. 2와 같이 추출하게 된다. 목표 데이터로부터 데이터 추출되면 전처리를 거쳐 변환하여 패턴을 만들어 데이터 마이닝 분석으로 수행한다[12, 13].

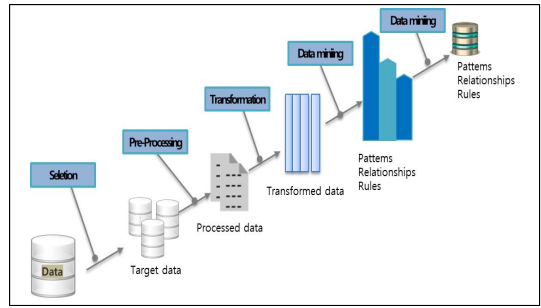


Fig. 2. Data mining

2.5 텍스트 마이닝

텍스트 정보 데이터를 대상으로 하여 필요한 정보를 추출하는 과정으로 정의한다. 텍스트 형태로 만들어진 정형화 되지 않은 데이터를 자연어 처리방식을 이용하여 특정 주제어(검색어)나 문맥을 기반으로 Fig. 3과 같이 데이터를 추출한다[14].

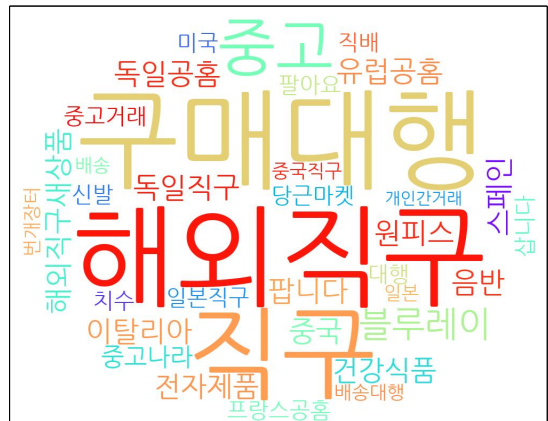


Fig. 3. Text mining

3. 시스템 설계

3.1 시스템 설계

본 논문에서 설계하고자 하는 해외직구 불법 거래 시스템 전체 구조는 Fig. 4에 나타내었다. 해당 시스템은 중고 사이트에서 개인 간 중고 거래의 전후의 다양한 데이터들을 실시간으로 수집하여 해외직구와 관련한 데이터를 정제한 후 자연어 처리 등을 진행하여 해외직구 물품 불법거래 관련 모니터링 시스템의 검색분석, 시각화하는 부분으로 나뉜다.



Fig. 4. System Configuration

시스템을 구현하기 위해 윈도우10 컴퓨터를 사용하고 파이썬을 이용 크롤러를 만들고, 데이터를 수집하고, 수집된 데이터는 SAS로 전처리를 하여 관련 웹서비스를 구현하였다.

3.2 데이터 수집

해외직구 물품이 자주 거래되는 온라인 중고 사이트 중에서 개인 간 거래가 활발한 곳을 선정하였다. 해당 사이트에서 해외직구 불법 거래 의심 판매 글 등 거래되고 있는 불특정 거래 내용을 크롤링으로 수집하였다[15]. 수집된 해당 데이터의 칼럼은 Table 2에 나타내었다.

Table 2. Data Collection Column

Online Used Sites e-commerce)	
· Nickname	· price
· seller phone	· trading area
· product name	· Date
· title	· URL
· e-mail	· payment method

수집 대상은 사전 지정해 놓은 키워드(직구, 해외, 공홈, 대행, 배송 대행지, 지역, 물품명, 국가명)로 필요한 검색 조건을 활용하여 온라인 중고 사이트에서 판매 중이거나 판매 완료된 해외직구 불법 거래 의심 판매 게시글을 수집하였다. 판매 완료된 데이터를 수집하는 것은 판매자가 같은 방법을 이용하여 판매하거나 빈도를 확인할 수 있는 분석하는 근거가 된다.

본 논문에서는 53만 건을 수집하였으며, 데이터의 해당 수집 주기는 해당 웹 사이트에서 부하를 주지 않는 목시적 대기를 위해 20초 주기로 지속하였다. Fig. 5는 일일 수집 건수를 보여주고 있다.

```

File Name: /opt/nextstep/HAV0804P/H1_HAV0804P_20211210.csv
Thu 16 Dec 2021 07:15:01 AM KST
/opt/nextstep/nextstep.py:50: FutureWarning: The default value of regex will change from True to False in a future version. In addition, single character regular expressions will no longer be treated as literal strings when not escaped.
  tgdata["MAKE_DT"] = tgdata["MAKE_DT"].str.replace(".", "")
20211210
srcData Row Count: 1260
srcData Row Count: 1239
srcData Row Count: 1661
srcData Row Count: 96
srcData Row Count: 1159
tgData Row CNT: 2712
Merge Data Row Count: 4730
File Name: /opt/nextstep/HAV0804P/H1_HAV0804P_20211210.csv
Fri 17 Dec 2021 07:15:01 AM KST
/opt/nextstep/nextstep.py:50: FutureWarning: The default value of regex will change from True to False in a future version. In addition, single character regular expressions will no longer be treated as literal strings when not escaped.
  tgdata["MAKE_DT"] = tgdata["MAKE_DT"].str.replace(".", "")
20211210
srcData Row Count: 1111
srcData Row Count: 1227
srcData Row Count: 1731
srcData Row Count: 63
srcData Row Count: 988
tgData Row CNT: 2769
Merge Data Row Count: 4781
File Name: /opt/nextstep/HAV0804P/H1_HAV0804P_20211210.csv
Sat 18 Dec 2021 07:15:01 AM KST
/opt/nextstep/nextstep.py:50: FutureWarning: The default value of regex will change from True to False in a future version. In addition, single character regular expressions will no longer be treated as literal strings when not escaped.
  tgdata["MAKE_DT"] = tgdata["MAKE_DT"].str.replace(".", "")
20211210
srcData Row Count: 1353
srcData Row Count: 1278
srcData Row Count: 1804
srcData Row Count: 63
srcData Row Count: 1119
tgData Row CNT: 2856
    
```

Fig. 5. System Configuration

3.3 비정형 데이터의 정제

Fig. 6은 데이터가 저장되는 과정을 나타내고 있다. 수집된 비정형 데이터는 유의미한 데이터와 무의미한 데이터를 분류하여 데이터 저장소에 구조화하고 각 분류를 만들어 분석하여 데이터베이스에 적용되도록 하였다.

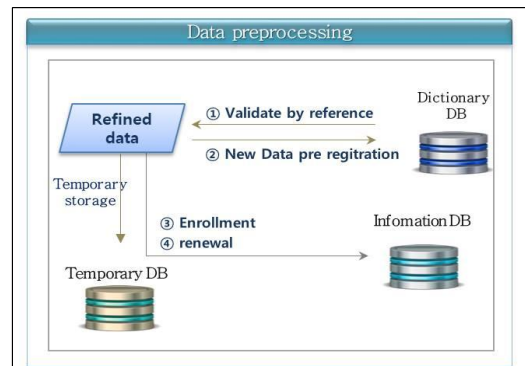


Fig. 6. Data Flow Diagram

비정형 데이터는 판매자의 닉네임, 제목, 휴대 전화번호, 제품명, 가격, 거래지역, 결제 방법, 작성일자 등은 방대한 텍스트로부터 의미 있는 정보를 추출하고 활용하기 위해 자연어 처리를 거쳐 데이터를 구조화하였다. 수집된 외부 데이터에서 판매자 전화번호를 추출하여 문자 및 특수문자를 제거하고 숫자 형 변환을 통해 내부 데이터와 전화번호 형태로 정제하였으며, 판매금액 정제 시에도 수집된 외부 데이터에서 판매금액을 추출하여 문자 및 특수문자 제거와 숫자 형 변환을 통해 금액 형식의 값으로 정제하였다. 거래지역 정제 때 수집된 외부 데이터에서 거래지역을 추출하여 특수문자, 중복단어 제거를 통해 지역명 형식으로 정제하였다.

게시글 주소 URL(uniform resource locator)는 판

매자가 본인이 작성한 글을 삭제해도 모니터링 분석 시스템에서 해석할 수 있도록 수집하였다.

3.4 시스템 탑재

온라인 중고 거래 사이트를 선정하여 일정 시간 간격으로 크롤링 한 데이터를 데이터 저장소에 수집한다. Fig. 7은 자동으로 매일 저장되는 CSV (Comma Separated Values) 파일을 나타내었다.

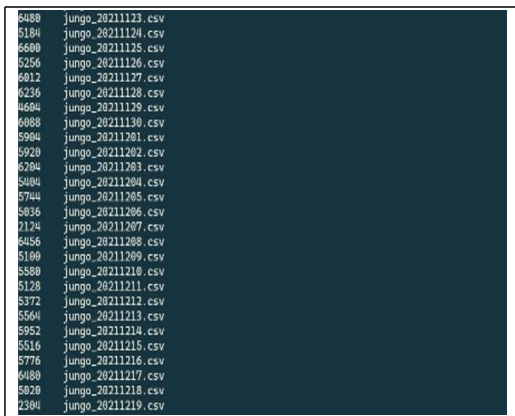


Fig. 7. CSV saved daily

Fig. 8과 같이 수집된 데이터를 임시 저장소에 저장 후 유의미한 데이터 여부를 판별한 뒤 무의미한 데이터와 구분하여, 데이터 전처리 후 분석을 할 수 있도록 데이터베이스에 기록된다.

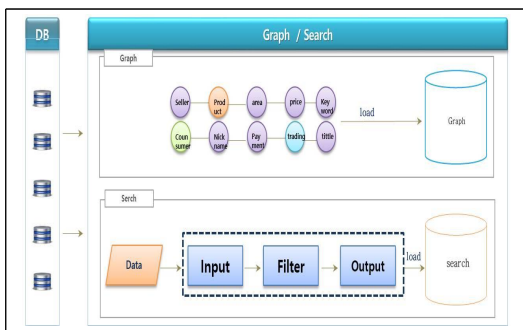


Fig. 8. System Flow Diagram

4. 시스템 분석

4.1 데이터의 시각화

수집된 데이터를 정제하여 사용자가 검색하고 그 결과를 확인하고, 수집현황 등을 실시간 관찰할 수 있도록 대시보드로 Fig. 9에 나타내었다. 대시보드를 통하여 수집현황과 판매 키워드 등 사용자 요구에 대응한 유연성과 편리성을 시각화 표현을 보유하고 있다.



Fig. 9. System Dashboard

4.2 수집데이터의 유사도

Fig. 10는 외부 수집데이터의 기준으로 검색기능을 수행한 결과이다. 조회 기간, 닉네임, 휴대 전화번호, 게시물 제목, 제품명, 거래지역을 세부 조건으로 검색할 수 있다. 이에 따라 해외직구 불법 거래 의심 판매 게시글을 수집하여 판매자 정보와 제품 판매 유사도를 확인하고 자주 거래하는 제품, 유사함 제품군, 게시물 등록 빈도, 거래지역 등으로 관계 예측을 위한 지표로 활용할 수 있다.



Fig. 10. Search and Result

Fig. 11은 수집된 자료를 이용해서 주로 거래하는 품목, 구매 지역, 등록 빈도수와 거래지역 등에 대하여 순위를 구현하였다.

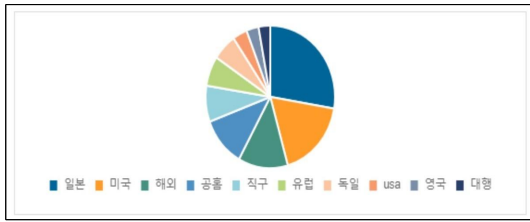


Fig. 11. Ranking of Status

4.3 키워드 정합성과 예측

Fig. 12는 일일 수집되는 데이터의 정합성을 확인하기 위해 외부 데이터의 수집과 함께 저장소에서 수집현황을 구현하였다.

전주 대비 증감을, 금주 비율, 전주 비율에 대해 전처리한 과거의 데이터를 활용한 예측 가능한 추세를 전망할 수 있음을 알 수 있다.

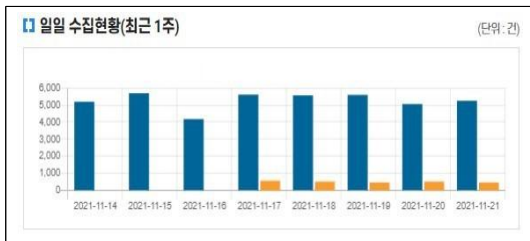


Fig. 12. Collection of Status

Fig. 13은 자주 판매되는 품목에 따라 여러 명의 판매자가 나타나도록 하였다. 이를 통하여 판매 분산 반입의 심 대상자의 판매 내용을 생성하여, 품목별 선호도에 따라 다수의 판매자와 품목별 연결을 나타낼 수 있다.



Fig. 13. Sales Correlation Analysis

5. 결론

온라인 쇼핑의 극대화에 따라 직구 물품을 선호하는

구매자와 사회적 분위기를 악용한 해외직구 재판매자에 대한 사회적 이슈가 제기되고 있다. 본 논문은 온라인 중고 사이트의 개인 간 거래를 중심으로 해외직구 불법 불법 거래에 관한 분석에 대한 모니터링 시스템을 설계하고 분석하였다.

해당 시스템은 닉네임, 휴대 전화번호, 제품명(숫자, 문자, 특수문자) 등을 변환형식에 맞게 최적화된 후 구조화하였다. 거래하는 제품이나 게시글 등록, 해당 빈도수, 거래지역 등을 판매자 정보와 제품명의 유사도를 분석하고, 닉네임 또는 전화번호를 주기적으로 변경할 때도 동일성을 판별, 유사한 형태를 보이는 거래방식을 기준으로 관련 패턴을 도출할 수 있다. 따라서 판매자와 유사한 형태를 보이는 거래방식, 동일성 판별 등 정합성으로 관계 예측을 분석할 수 있었다. 데이터를 시각화하여 해외직구 물품의 리셀러의 불법 거래 단속에 활용할 수 있음을 알 수 있다.

앞으로의 연구는 더 다양한 데이터를 수집·확대하여 해외직구 불법 거래 연구가 필요하고 사이버 불법 무역 방지 등 사회적 문제점들을 해결할 수 있는 계기가 될 수 있으며 발전시킬 필요가 있다.

REFERENCES

- [1] Seoul Customs Hosts "Online Illegal Transaction Prevention Meeting" with e-commerce companies (20.5.29)
- [2] Seoul Customs announces results of special crackdown on overseas direct sales in the first half of the year(21.9.14).
- [3] Seoul Customs Detects 15 Offenders of Direct Buyback Sales during Year-End and New Year Special Enforcement (21.2.4)
- [4] Korea Consumer Agency (2021). Analysis of trends in consumer counseling for international transactions in 2020. Research Report 1-29.
- [5] Korea Consumer Agency (2018)Overseas direct purchase use and damage situation investigation 2018.05, 1-77.
- [6] Korea Customs Service Special Customs Section, E-Country Indicators E-commerce Goods Import Trend <https://www.index.go.kr/potal/main/EachDtlPageD>
- [7] Korea Customs Service, Special Customs Section, Guide to Customs Clearance for Overseas Direct Purchase Useful to Know (2020)
- [8] Kim Seong-ik, Park Gu-rak, Kim Jae-woong, Lee

Yun-yeol, Kim Dong-hyun.(2021). Detailed filter model of C2C e-commerce platform.Proceedings of the Korean Society for Computer and Information Studies ,29(2),85-86

- [9] Seongpil Kim A Study on the User Structure Analysis of Second Hand Market Apps -Focusing on Joongnara, Bungejangter, Danggeunmarket *Journal of the Korean Society for Industrial-Academic Technology* 2021,vol.22,no.7,pp.449-458
DOI : 10.5762/KAIS.2021.22.7.449
- [10] Song, Eugene. The influence of heuristic-systematic information processing leads on perceived risk and transaction intentions in online used goods transactions *Journal of consumer policy studies* 51(1), 2020.4, 51-77(27 pages)
DOI : 0.15723/jcps.51.1.202004.51
- [11] Jaesung Kim, Youngjin Choi, Jaedong Hwang and Wanseop Cho. 2019, "Fisheries science big data platform establishment and meta data management plan", *Journal of the Korean Big Data Society*, vol.4, no.2 pp.93-103.
DOI : 10.36498/kbigdt.2019.4.2.93
- [12] Park Joo-seok.2018 "Comparative Study of Big Data, Open Data, and My Data", *Journal of the Korean Big Data Society*, vol.3, no.1 pp.41-46.
DOI : 10.36498/kbigdt.2018.3.1.41
- [13] Museong Kim, Namgyu Kim Text Augmentation Using Hierarchy-based Word Replacement *Journal of the Korea Society of Computer and Information* 26(1), 2021.1, 57-67 (11 pages)
DOI : 10.9708/jksoci.2021.26.01.057
- [14] Su Ryeon Lee , Choi Eun-jung Comparison of responses to issues in SNS and Traditional Media using Text Mining -Focusing on the Termination of Korea-Japan General Security of Military Information Agreement(GSOMIA)-*Journal of Digital Convergencr* vol.18, no.2 pp.277-284.
DOI : 10.14400/JDC.2020.18.2.27
- [15] Young-Geun Yu, Ki-Bok Nam, &Koo-Rack Park *Implementation of web server monitoring system using crawling technology Journal of the Korea Society of Computer and Information* 24(4), 2019.4, 123-128
DOI : 10.9708/jksoci.2019.24.04.12

신 용 훈(Yong-Hun Shin)

[충원]



- 2007년 2월 : 한밭대학교 컴퓨터공학과(공학석사)
- 2021년 3월 : 한밭대학교 컴퓨터공학과 박사과정
- 관심분야 : 빅데이터, 인공지능, 정보보호
- E-Mail : eten@naver.com

김 정 호(Jeong-Ho Kim)

[충원]



- 1983년 2월 : 경북대학교 전자공학과(공학석사)
- 1994년 2월 : 단국대학교 컴퓨터공학과(공학박사)
- 1983년 3월 ~ 1996년 2월 : 한국전자통신연구원(실장,책임연구원)
- 1997년 2월 ~ 현재 : 한밭대학교 컴퓨터공학과(교수)
- 1997년 2월 ~ 현재 : 한밭대학교 정보기술대학장, 대학원장
- 관심분야 : 빅데이터, 프로토콜공학, 인공지능, 정보보호
- E-Mail : jhkim@hanbat.ac.kr