

해외직구 물품 불법 거래에 관한 모니터링 서비스 설계와 해석

신용훈*, 김정호**, 조진표***

*한밭대학교 정보통신전문대학원 컴퓨터공학과

**한밭대학교 정보통신전문대학원 컴퓨터공학과

***한밭대학교 정보통신전문대학원 컴퓨터공학과

boriwool@korea.kr, jhkim@hanbat.ac.kr, ljich@hanmail.net

Regarding the illegal transaction of overseas direct purchase Monitoring service design and analysis

Yong-Hun Shin*, Jeong-Ho Kim**, Jin-Pyo Jo***

*Dept. of Computer Engineering, Han-bat University

**Dept. of Computer Engineering, Han-bat University

***Dept. of Computer Engineering, Han-bat University

요 약

관세법에서는 해외직구물품이 일정금액(미화 150불, 단 미국은 미화 200불)이하 또는 자기사용 물품으로 인정되는 경우에 제세를 면제토록 규정하고 있으며 관련규정을 어길시 관세법상 무신고 밀수입죄에 해당된다. 본 논문은 해외직구 리셀러(되팔이)가 증가하고 해당 사항이 사회적 이슈로 대두되기에 해외직구 물품 불법거래에 관한 모니터링시스템을 설계하고 해석하였다. 온라인 중고 사이트(e-commerce)에서 거래되고 있는 거래 내용을 크롤링을 통하여 데이터를 수집·전처리를 통해 구조화하고 데이터 정제, 텍스트 범주화, 텍스트 마이닝 등 관계 예측을 해석하였다.

1. 서론

최근 코로나19의 영향으로 비대면 온라인 쇼핑이 활발해지고 해외직구 물품을 선호하는 사회적 분위기를 악용한 '리셀', 즉 해외직구 리셀러(되팔이)가 증가하고 있다. 관세법에서는 해외직구 물품이 첫째, 일정 금액(미화 150불, 단 미국은 미화 200불) 이하이고 둘째, 자기 사용 물품으로 인정되는 때에만 관세 등 제세를 면제하도록 규정하고 있다. 관련 해외직구 물품을 온라인 등에서 되팔이하는 행위는 관세법상 무신고 밀수입죄에 해당하게 된다.

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 오픈마켓을 통해 되팔이한 리셀러들이 최근 '21년 9월 적발되었으며 이중 상습·전문 판매자 6명은 관세법 위반으로 입건되는 등 사회적으로 매우 물의를 일으키고 있다[2].

본 논문은 해외직구 물품이 자주 거래되는 온라인 중고 사이트(e-commerce)를 중심으로 유통되는 해외직구 물품 정보를 크롤링을 이용하여 설계를 진행하였다. 필요한 데이터를 수집하고 정제하여 해외직구 물품 불법 거래에 대한 모니터링 시스템을 구현하여 데이터 정제, 텍스트 범주화, 유사도 분석, 텍스트 마이닝 등 관계 예측을 해석하였다.

제2장에서는 기반연구를 소개하며, 제3장은 수집된 데이터를 정제하고 해외직구 물품 불법 거래에 관한 모니터링 시스템 설계와 해석을 수행하였다. 제4장에서 본 연구의 결론과 향후 연구 방향을 제시한다.

2. 관련 연구

2.1 온라인 중고 사이트와 빅데이터

온라인을 기반으로 하는 중고거래는 1990년대 후반부터 2000년대 초반 사이에 처음 등장한 이후 꾸준히 그리고 상당한 속도로 성장하고 있다[3]. 스마트 기기, SNS, IOT 기술의 발달과 확산으로 인하여 정보의 양과 그 종류가 매우 많아져 빅데이터로 발전하였다. 빅데이터는 신뢰성 있는 데이터를 제공하여

해당 기반으로 만들어진 시스템은 더욱 정확하고 신뢰성 있는 정보를 제공하며 빠르게 선택을 할 수 있도록 할 수 있다. <표 1>은 빅데이터 구조의 핵심 기술을 정리한 것이다[4].

<표1> 빅데이터 단계별 핵심기술

수집	· 내·외부 데이터 연동 · 내·외부 데이터 통합	· Crawling, Open API,FTP · RSS,Log Aggeration · DB Aggeration, Streaming
적재	· 대용량·실시간 데이터 처리 · 분산파일시스템 저장	· Memory Cached · Message Queue · Distributed File, No-SQL
처리	· 데이터 선택, 변환,통합,축소 · 데이터 워크플로·자동화	· Structured Processing · Unstructured Processing · Workflow, Schduler
탐색	· 대화형 데이터 질의 · 탐색적Ad-Hoc 분석	· SQL Like · Distributed Programing · Exploration Visualization
분석	· 빅데이터 마트 구성 · 통계 분석, 고급분석	· Data Mining · Analysis Visualization · Machin Learning
응용	· 보고서·시각화 · 분석 정보	· Data Export/Import · Reporting · Business Visualization

2.2 크롤링

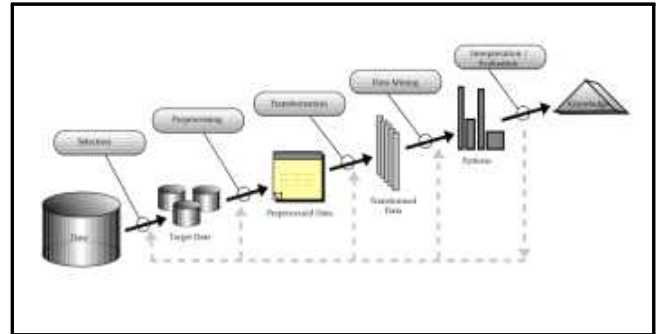
크롤링의 의미는 웹크롤러 또는 웹 스크레이퍼에서 출발한다. 이론적으로 웹페이지에서 데이터(정보)를 가져와서 필요한 데이터만을 수집하는 일련의 작업 전체를 말하며 가장 널리 사용되는 방법은 프로그램을 만들어 해당 웹 서버에 쿼리를 보내어 데이터(HTML이나 기타 웹페이지를 구성하는 파일)를 요청하고 수집하는 역할을 담당하고 이용할 수 있게 하는 원초적인 기술이다[5]. 4차산업 시대에서 웹에서 생산되는 거대한 데이터 중 맞춤형 데이터를 이용하려는 사람들이 늘어나고 있으며 수집하는 기술이 지속해서 발달하고 있다

크롤링의 대상은 온라인상의 웹 환경에서의 모든 정보를 자동 수집할 수 있으며 소셜미디어(트위터, 페이스북 등), 뉴스 등 다양한 웹 공간 미디어들의 콘텐츠 증가로 크롤러의 필요성과 중요성이 더욱 주목받고 있다 간혹 수집을 할 수 있는 API를 제공하는 곳도 존재하지만, 대부분은 필요한 사이트를 수집해야 할 때 제공해 주지 않고 그 사용도 대단히 제한적이라 존재하지 않을 수 있어 지금도 계속 크

롤러는 필요성에 의해 만들어지고 사용되고 있다

2.3 데이터마ining

많은 데이터를 수집·저장한 빅데이터를 자동으로 통계적 규칙 또는 수치 패턴을 분석하여 가치 있는 정보를 추출하는 것으로 (그림1)과 같이 목표 데이터로부터 데이터 추출되면 전처리를 거쳐 변환하여 패턴을 만들어 데이터 마이닝 분석으로 이루어져 있다[6].



(그림1)데이터마이닝

2.4 텍스트마이닝

텍스트마이닝은 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 데이터를 추출한다. 데이터 마이닝과의 차이는 텍스트 정보 데이터와 수치 정보 데이터를 다룬다는 점에서 구분되고 텍스트마이닝은 구조화시키는 과정이 필요하다[7]. 해당 절차를 도식화하면 (그림2)와 같다[8].

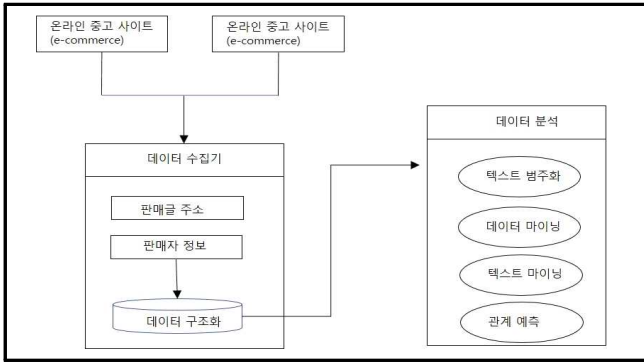


(그림2) 텍스트마이닝 흐름도

3. 제안 시스템 설계

3-1 시스템 구성

해외직구 물품이 자주 거래되는 온라인 중고 사이트(e-commerce)를 두 곳을 선정하여 크롤링 시스템을 설계하였다. 데이터 전처리 및 분석을 위한 시스템을 (그림3)과 같이 구성요소를 설계 구성하였다.



(그림3) 시스템 구성도

3-2 시스템 동작 절차

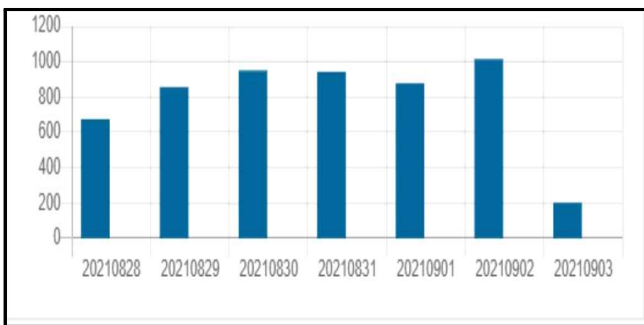
크롤링 대상은 “전체검색” 기능을 활용하여 사전에 정해놓은 키워드(직구, 해외, 공홈, 국가명 등)를 검색 조건으로 활용하여 판매 완료 및 판매 중인 데이터를 전체 수집을 하고 있으며 컬럼은 <표2>를 확인한다.

<표2> 데이터 수집 컬럼

온라인 중고 사이트(e-commerce)	
· 닉네임	· 가격
· 휴대전화번호	· 거래지역
· 제품명	· 작성일자
· 제목	· 게시글 주소(URL)

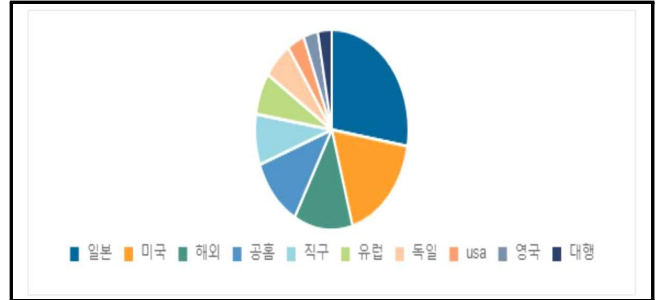
수집 컬럼의 사용은 게시글 주소(URL)는 체증 확보를 위해 글쓴이가 삭제해도 본 시스템에서 사용할 수 있도록 수집했으며 판매자의 닉네임, 제목, 휴대전화번호, 제품명, 가격, 거래지역, 작성일자를 수집하여 관련 데이터 분석을 하기 위함이다.

해당 수집 주기는 10초에서 20초 주기로 지속해서 하였으며 3,000개에서 4,000개 수집하고 있다. (그림4) 일일수집현황은 ‘21.8.28부터 9.3일까지 1주간 수집된 데이터의 양을 보여준다.



(그림4) 일일수집현황

(그림5)는 수집된 자료를 이용해서 주로 거래하는 품목이나 게시글 등록 빈도수, 거래지역 등을 학습하여 키워드 순위를 보여준다.



(그림5) 키워드 랭킹(최근1주)

판매자 정보와 제품의 유사도 분석을 하여 주로 거래하는 제품이나 게시글 등록 빈도수, 거래 지역 등 관계 예측을 위한 지표로 사용토록 하였다. (그림6)에 외부 수집데이터 기준으로 닉네임, 휴대전화번호, 게시글, 제품명, 거래지역을 세부 조건으로 만들어 검색하여 모니터링을 수행함을 알 수 있었다.

The screenshot shows a search interface with filters for 'Search Period' (2021.07.01 to 2021.07.30), 'Site Type', 'Nickname', and 'Phone Number'. Below the search bar is a table of search results with columns: No, 제목 (Title), 제품명 (Product Name), 거래지역명 (Trading Area Name), 작성일시 (Posting Time), and URL. The table lists 10 results, including items like '8인칭 06-507 유닛리터 세트 14.5x34' and '9인칭 60U 상위 (해비메이트) SR'.

(그림6) 검색 및 결과 화면

4. 결론

본 논문에서 해외직구 물품을 선호하는 사회적 분위기를 악용한 ‘리셀’, 즉 해외직구 리셀러(되팔이)에 대한 사회적 동향을 설명하고 이에 대한 해결책으로 웹 크롤링으로 수집된 데이터를 이용하여 설계하고 모니터링을 수행하였다.

수집된 데이터를 자연어처리, 형태소 분석을 통하여 휴대전화번호, 모델·규격명, 품명(숫자, 문자, 특수문자)등을 최적화 하였다. 판매자 정보와 제품명의 유사도 분석을 하고 주로 거래하는 제품이나 게시글 등록, 해당 빈도수, 거래 지역 등을 학습하여 닉네임 또는 전화번호를 주기적으로 변경하는 경우에 동일성을 판별하여 유사한 패턴을 가지는 거래방

식을 기준으로 관련 패턴을 도출 할 수 있는 분석 기법을 적용해 보았다 이를 통해 우범 패턴 분석과 유사한 패턴을 가지는 거래방식 및 동일성 판별 등 관계 예측을 해석하였다.

앞으로는 다양한 데이터 수집 대상을 선정하여 해외직구 불법 거래에 대하여 영향을 미치는 요인들에 대한 개선이 필요하다. 또한 이를 토대로 정보의 근원지와 관계 속에서 데이터를 좀 더 활성화하고 시각화하여 해당 플랫폼을 구축하고 실제 도움을 줄 수 있는 시스템과의 연계를 통하여 개선된 서비스가 가능하다. 따라서 이러한 사이버 불법 무역 방지 등 사회적 문제점들을 해결할 수 있는 계기가 될 수 있으며 점진적으로 발전시킬 필요가 있다.

참고문헌

- [1]서울세관, 해외직구 물품 되팔기 집중단속 실시보도자료 (20.5.29)
- [2]서울세관 상반기 해외직구 되팔이 특별단속 결과 발표 보도자료(21.9.14)
- [3]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)
- [4]김강원 저, 실무로 배우는 빅데이터 기술, 위키북스, 2020년 06월 02일
- [5]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
- [6]J. A. Lara, D. Lizcano, M. A. Martínez, and J. Pazos, "Data preparation for KDD through automatic reasoning based on description logic," Information Systems, vol. 44, pp. 54 - 72, Aug. 2014.
- [7]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)
- [8]Lee, Won-Bael, Kim, Min-Jung, An, Min-Ki, Kim, Yong-Su "Risk Analysis Framework for Overseas Construction Projects by using Text-Minnin" kicem .(2019): 71-72.