

# 미술품 거래 빅데이터를 이용한 작가 분석 시스템 구현

이 승 경\*\*, 임 종 태\*\*\*

## 목 차

요약	3.2 데이터 수집
1. 서론	3.3 한국어 자연어처리
2. 관련 연구	4. 시스템 구현 결과
2.1 크롤링	5. 결론
2.2 자연어처리	참고문헌
3. 작가 분석 시스템	Abstract
3.1 시스템 구성도	

## 요약

국내 미술시장 규모는 2018년 기준으로 최근 5년간 매년 21.9%씩 증가하는 성장세를 유지하고 있다. 미술품 유통 플랫폼은 화랑과 오프라인 경매 방식, 그리고 온라인 경매로 다양해지고 있다. 미술시장은 작품의 제작(창작), 유통(무역), 소비(매수) 등 3개 분야로 구성되어 있으며, 경제적 가치는 물론 예술적 가치에 대한 인식이 확산되면서 투자 수단으로써 관심도가 높아지고 있다. 작품을 재테크 수단으로 생각하는 소비자는 작품의 객관적 정보에 대한 욕구가 높아지지만, 예술시장 유통 분야의 정보 제공이 폐쇄적이고 불균형해 객관적이고 신뢰할 수 있는 통계를 수집·분석하는 데 한계가 있다.

본 연구는 예술시장 유통 분야에 대한 빅데이터 수집과 정형·비구조적 데이터 분석을 통해 객관적이고 신뢰할 수 있는 미술품 유통 현황을 파악한다. 이를 통해 현재 시장에서 저자의 분석을 객관적으로 제공할 수 있는 시스템을 구현하고자 한다. 본 연구에서는 미술품 유통 사이트에서 저자 정보를 수집하고 일간지 매일경제에서 저자의 기사를 수집·분석해 작가별 연관 단어의 빈도를 산출했다. 이를 통해 본 연구에서는 소비자에게 객관적이고 신뢰할 수 있는 정보를 제공하는 것을 목표로 한다.

*표제어: 미술시장, 빅데이터, 크롤링, 한국어 자연어처리, 데이터시각화*

접수일(2021년 02월 09일), 수정일(1차:2021년 03월 16일), 게재확정일 (2021년 03월 16일)

\* 본 연구는 한국콘텐츠진흥원 국가 R&D연구 지원에 의하여 연구되었음.

\*\* 공주대학교 컴퓨터공학부 컴퓨터공학전공 학사과정졸업,

\*\*\* 공주대학교 컴퓨터공학부 교수

### 1. 서론

국내 미술시장 규모는 최근 5년간(2018년 기준) 21.9% 증가한 4,482억 원 규모로 증가하고 있으며, 거래 건수도 3만 9,367점으로 31.6% 증가하는 추세로 5년 연속 성장세를 유지하고 있다. 거래 방식도 화랑, 오프라인의 경매, 그리고 온라인 경매로 미술품 유통 플랫폼이 다양화되고 있다. 미술시장은 작품의 생산(창작), 유통(매매), 소비(구매)의 3가지 영역으로 이루어져 있으며, 미술적 가치뿐만 아니라 경제적 가치의 인식이 확산함에 따라 재테크의 수단으로도 관심이 높아지고 있다(Ministry of Culture, 2019).



Fig. 1-1 The Trend of Art Market Transactions for 11 Years(Ministry of Culture 2019)

미술품 작품을 재테크의 수단으로 생각하는 소비자는 작가와 작품에 대한 객관적인 정보의 필요성이 증가하고 있지만, 화랑, 경매회사, 아트페어를 중심으로 하는 현재 미술시장 유통영역에서 정보 제공이 폐쇄적이고 불균형적이므로, 객관적이고 신뢰성이 있는 통계자료를 수집 및 분석하는 것에 대한 한계

가 있다.

<Fig 1-2>는 뉴시스와 한국미술사가감정협회가 함께 설립한 미술품 유통 가격정보 사이트인 K-Artprice 메인 홈페이지다.

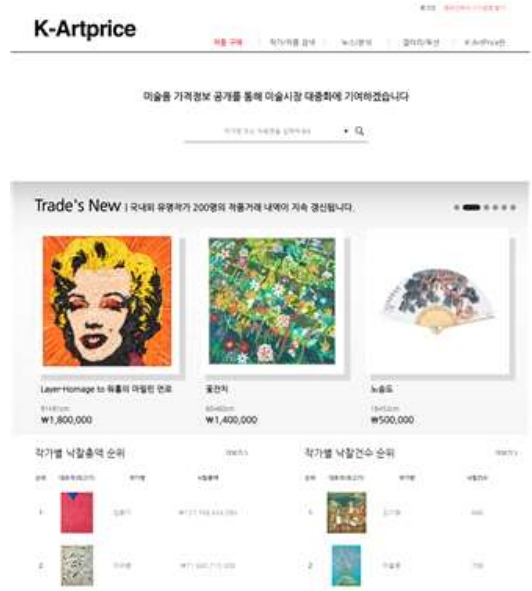


Fig. 1-2 K-Artprice

K-Artprice에서는 국내 경매회사를 통해 거래된 미술품의 가격과 작가 정보를 작품의 크기와 소재별로 정량적으로 제공하고 있으나 작가와 작품에 대한 상세한 정보를 정성적으로 알기에는 한계가 있다.

아래 <Fig. 1-3>은 매일경제에서 미술작가 강요배를 검색한 화면이다. 매일경제에서는 미술 작가명으로 검색하여 문화/공연/전시로 상세내용을 기사로 전달하고 있으며, 국내 미술시장에 관한 기사와 문화 관련 콘텐츠가 풍부하다. 하지만 미술품에 대한 콘텐츠 수는 많지만, 작가와 작품에 대한 정보를 알기는 어렵다.

본 연구는 미술시장 유통 영역에 관한 빅데이터 수집하고, 정형 데이터와 비정형 데이터 분석을 통하여 객관적이고 신뢰성이 있는 미술품 유통 현황

및 실태를 파악한다. 이를 통해 현재 시장에서 작가에 대한 정량적/정성적 분석을 객관적으로 제공할 수 있는 시스템을 구현한다.



Fig. 1-3 MK Search

본 연구에서는 Python을 분석 도구로 활용하여, 데이터의 작가별 단어의 출현 빈도수와 연관 단어의 추출을 통해 소비자들에게 객관적인 작가의 정보를 제공하고자 한다.

## 2. 관련 연구

### 2.1 크롤링

웹에 공개된 정보는 대부분 HTML 형식이며 이를 가져와서 데이터베이스에 저장하려면 데이터 가공이 필요하다. 광고 등의 불필요한 정보를 제거하고, 필요한 정보만 가져오려면 사이트의 구조를 식별하여야 한다(WIKIPEDIA, 2017). 웹에서 공개된 정보를 추출하고 웹 문서의 구조를 분석하는 것을 스크래핑 기술이라 한다. 최근에는 로그인해야 정보에 접근할 수 있는 사이트도 많이 있다. 이 경우 단순히 URL

을 알고 있는 것만으로는 유용한 정보에 접근할 수 없으므로 정확한 스크래핑을 위해 로그인해서 필요한 웹 페이지에 접근한다.

이러한 스크래핑을 지원하는 프로그램이 웹 크롤러(Web Crawler)이다. 웹 크롤러는 웹 문서와 이미지 등을 주기적으로 수집하여 자동으로 데이터베이스화하는 프로그램이다.

웹 크롤러가 하는 작업을 웹 크롤링(Web Crawling)이라 하며, 보통 웹 크롤러를 사용하여 웹 문서의 복사본을 생성하고 검색 엔진은 이렇게 생성된 데이터를 인덱싱하여 빠른 검색을 할 수 있도록 한다. 파이썬에서 이용할 수 있는 웹 크롤러 라이브러리인 BeautifulSoup과 Selenium이 대표적이다.

Beautifulsoup 은 HTML 및 XML 파일에서 데이터를 가져오기 위한 Python 라이브러리이며, 선호하는 파서와 함께 작동하여 구문 분석 트리를 탐색, 검색 및 수정하는 방법을 제공한다(WIKIPEDIA, 2004). Selenium에 비교하면 빠르게 수집할 수 있다.

Selenium은 웹 브라우저 원격 조작에 사용하는 라이브러리이다(WIKIPEDIA, 2020). 자바스크립트를 많이 사용하는 웹 사이트는 스크래핑할 때 웹 브라우저를 사용하지 않으면 제대로 수집할 수 없다. 이런 경우에는 웹 브라우저를 원격으로 조작하는 방식으로 모든 웹 페이지에서 데이터를 스크래핑할 수 있다. Selenium을 이용하면 자동으로 URL을 열고 클릭할 수 있으며, 스크롤 하거나, 문자를 입력하는 등의 다양한 조작을 자동화할 수 있다. 또한, 화면을 캡처해서 이미지로 저장하거나 HTML의 특정 부분을 꺼내는 것도 가능하다. 구글 크롬, 파이어폭스, 인터넷 익스플로러, 오페라 등의 다양한 웹 브라우저를 원격으로 조작할 수 있다.

### 2.2 자연어 처리

자연어 처리는 인간의 언어 현상을 컴퓨터와 같은 기계를 이용해서 모사할 수 있도록 연구하고 이를 구현하는 인공지능의 주요 분야 중 하나이다. 자연어 처리 기술은 정보검색, QA 시스템, 문서 자동 분류, 신문기사 클러스터링, 대화형 Agent 등 다양한 응용에 사용하고 있다(WIKIPEDIA,2010).

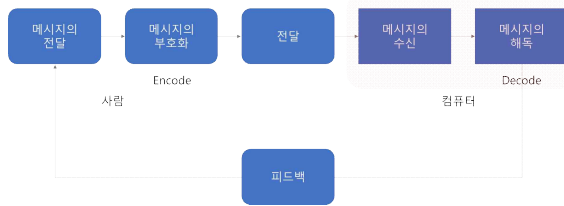


Fig. 2-1 NLP Process

<Fig. 2-1>은 자연어 처리 과정이다. 한국어 문장을 분석할 때 가장 문제가 되는 것은 단어를 구분하는 것이다. 영어의 경우 공백으로 잘라 쉽게 단어를 나눌 수 있지만, 한국어에는 조사가 붙어있으므로 단순히 공백으로 잘라서는 단어를 구분할 수 없다. 따라서 한국어 문장을 최소 단위인 형태소로 나누어 주는 형태소 분석(Morphological Analysis)을 수행한다.

한국어 형태소 분석 라이브러리에서는 파이썬에서 사용할 수 있는 오픈 소스를 이용하여, 직접 사전을 준비하고, 형태소 분석 알고리즘을 구현하지 않아도 형태소 분석을 할 수 있다. 이번 연구에서는 은전한닢, 꼬꼬마, 한나눔, Okt, 코모란 등 5개 오픈 소스 형태소 분석기를 파이썬 환경에서 사용할 수 있도록 인터페이스를 통일한 지도 학습 기반 한국어 자연어처리 패키지를 사용한다(Lee, 2019).

<Tab. 2-1>과 <Fig. 2-1>은 클래스별 성능지표이다.

Tab. 2-1 Comparison between POS tagging classes(KoNLPy)

Class	Loading	Execution
-------	---------	-----------

	time	time
Hannanum	5.6988	35.7163
Komorran	5.4866	25.6008
Hannanum	0.6591	8.8251
Okt(Twitter)	1.4870	2.4714
Mecab	0.0007	0.2838

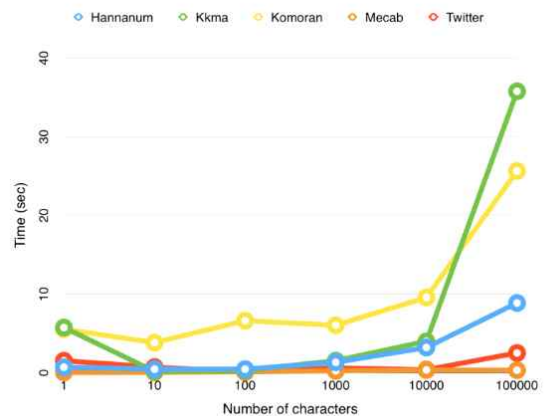


Fig. 2-1 Number of characters(KoNLPy)

자연어 처리에서 속도만큼 중요한 지표는 형태소 분석 품질이다. ‘아버지가방에들어가신다’ 문장을 분석한 결과는 <Tab. 2-2>와 같다.

Tab. 2-2. Comparison of quality of morpheme analysis

Hanna num	Kkma	Komorran	Mecab	Okt(Twitter)
		아버		
아버	아버	지가방에	아버	아버
지가방에	지 /	들어가신	지 /	지 /
들어가/	NNG	다 /	NNG	Noun
N		NNP		
	가방		가 /	가방
이 / J	/ NNG		JKS	/ Noun

시 대 / E	에 / JKM	방 / NNG	에 / Josa
	들어 가 / VV	에 / JKB	들어 가신 / Verb
	시 / EPH	들어 가 / VV	다 / Eomi
	나대/ EFN	신다 / EP+EC	

본 연구는 속도와 품질을 비교하기 위하여 Okt 클래스를 사용한다.

Word2Vec은 2013년 구글 연구팀이 발표한 기법으로 가장 널리 쓰이고 있는 단어 임베딩 모델로, 한글 문장의 단어를 벡터로 변환하는 도구이다(Lee, 2019). Word2Vec은 한글 문장의 단어 사이의 연결을 기반으로 단어와 단어의 연관성을 분석하여 내부 단어들을 벡터로 만들어준다.

Word2Vec은 CBOW와 Skip-gram 두 가지의 모델을 가지고 있다. CBOW는 주변에 있는 문맥 단어들을 가지고 타깃 단어 하나를 맞추는 과정에서 학습이 된다. Skip-gram 모델은 타깃 단어를 가지고 주변 문맥 단어가 무엇일지 예측하는 과정에서 학습된다(Lee, 2019).

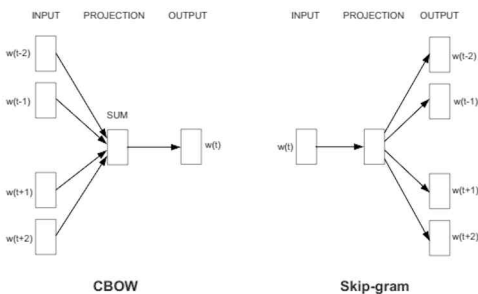


Fig. 2-2 CBOW, Skip-gram model(Mikolov et al., 2013a)

### 3. 작가 분석 시스템

#### 3.1 시스템 구성도

본 연구의 작가 분석 시스템은 국내 미술시장의 객관적인 정보 전달을 위해 K-Artprice의 정형 데이터와 매일경제에서 작가에 관한 기사인 데이터를 융합하여 분석 결과를 시각화한다. <Fig. 3-1>은 작가 분석을 위한 시스템 구성도이다.

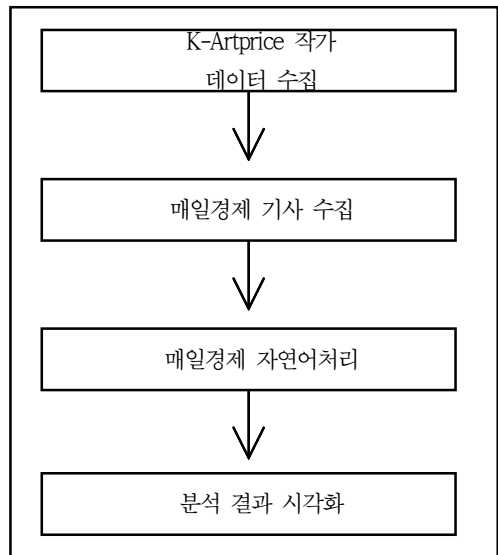


Fig. 3-1 System configuration diagram

K-Artprice에 작가 데이터를 파이썬 웹 크롤링인 BeautifulSoup, Selenium을 사용하여 수집하고, 수집된 작가 대상으로 BeautifulSoup을 이용하여 매일경제 작가별 기사 주소를 수집한다. 수집된 기사 주소에 접속하여 기사 내용을 수집한다. 수집된 매일경제 기사 내용에 관해서 문장에 대한 형태소를 분석하고, Word2Vec을 사용하여 단어 간 유사도 계산을 수행한다. 마지막으로 최종 분석 결과에 대한 시각화를 해주는 시각화 모듈을 구현한다.

#### 3.2 데이터 수집

K-Artprice는 미술품 국내 경매회사를 통해 거래된 미술품의 최신 가격을 투명하게 제공한다. 국내외 주요 작가들과 작품 정보를 손쉽게 볼 수 있으며, 작가의 작품 거래 정보를 수집할 수 있는 점에서 선정했다.

매일경제는 국내 주요 일간지로 문화/공연/전시로 상세내용을 기사로 전달하고 있으며, 국내 미술시장에 관한 기사가 많이 작성되고 있고, 매일경제신문 프리미엄 기사에 문화 관련 콘텐츠가 풍부한 점에서 선정했다.

K-Artprice에 등록되어있는 국내외 유명작가 200명 중 거래가 활발히 이루어지고 최근 5년간 낙찰액이 많은 국내 작가 100명을 선정했다. 100명의 작가의 데이터를 수집하기 위하여 웹 크롤링 기술 중 웹 자동화 프레임워크인 Selenium을 사용하여 작가명 검색 자동화 프로그램을 만들었다. HTML 문서의 구문을 분석하기 위한 Python 패키지인 BeautifulSoup을 사용하여 Selenium에서 전달받은 페이지 정보를 100명 작가의 국적, 생애, 학력, 최근 5년간 낙찰액, 그리고 낙찰 전수를 수집한다.

수집한 데이터는 유지와 관리가 용이한 데이터베이스에 저장하였다. 100명의 작가를 상위, 중위, 하위 등급으로 나누어 주기 위해 최근 5년간 낙찰액 기준으로 정렬했다.

genre	name	country	birth	death	sex	school	price	count	percentile
1 회화	김환기	대한민국	1913	1974	남성	서울대학교 미술학	140395303690	472	1
2 회화	이우환	대한민국	1936		남성	서울대학교 철학	60051852500	484	0.9898989898989899
3 회화	박서보	대한민국	1931		남성	홍익대학교 회화과	330080228050	272	0.9797979797979798
4 회화	정승환	대한민국	1932		남성	서울대학교 회화과	31163046450	192	0.9696969696969697
5 회화	전경자	대한민국	1924	2015	여성	도로 여자미술전문학교 미술학 동회	18527500000	317	0.9595959595959596
6 회화	박수근	대한민국	1914	1965	남성	양구공립보통학교	17534286000	132	0.9494949494949495
7 회화	김창열	대한민국	1929		남성	서울대학교 미술학	16429552600	406	0.9393939393939393
8 회화	이종섭	대한민국	1916	1996	남성	도로 문리대학 미술과	11328340000	36	0.9292929292929293
9 회화	이대원	대한민국	1921	2005	남성	서울대학교 법학	8714790000	422	0.9191919191919191
10 회화	윤형근	대한민국	1928	2007	남성	홍익대학교 서양화	8586740900	162	0.9090909090909090

Fig. 3-2 Top10 Based on the amount of bid in the last 5 years

<Tab. 3-1>은 상위, 중위, 하위 작가 리스트이다.

Tab. 3-1 Artist List

	상위작가	중위작가	하위작가
1	김환기	오수환	문신
2	이우환	임옥상	윤병락
3	박서보	변관식	오승윤
4	정상화	이건용	김동유
5	전경자	김기린	박영선
6	박수근	사석원	김형근
7	김창열	최영림	안중식
8	이중섭	최올가	최영욱
9	이대원	윤명로	도성욱
10	윤형근	곽인식	황규백
11	장옥진	박생광	안창홍
12	유영국	강요배	이수억
13	백남준	이숙자	박정희
14	김종학	황염수	김인승
15	오치균	이인성	변종하
16	도상봉	이동엽	신영복
17	이용노	고영훈	이만익
18	이강소	임직순	김규진
19	하종현	변시지	박득순
20	김정희	김구림	이수동
21	남관	강익중	홍경택
22	김태호	황용엽	마리킴
23	전광영	박고석	박항률
24	이성자	전혁림	한묵
25	서세옥	황영성	이쾌대
26	김기창	민경갑	김병중
27	권옥연	권순철	최종태
28	권영우	오세열	이항성
29	이배	하인두	이희돈
30	오윤	박노수	이두식
31	정창섭	윤중식	권기수
32	김홍수	배병우	류병엽
33	안영일	이동기	이승오
34			김점선

상위, 중위, 하위 등급의 작가를 대상으로 매일경제에서 문화산업, 공연/전시/음악 카테고리 중심으로 기사 수

집을 수행한다. 기사 수집을 하기 위해 Python의 Newspaper 모듈을 사용한다. Newspaper는 사용자가 지정한 URL에서 텍스트만 추출을 해주는 모듈로 간편하게 기사 내용을 수집할 수 있다. Newspaper에 URL을 지정하기 위해 기사 제목에서 뉴스를 불러올 기사 주소를 BeautifulSoup을 사용하여 기사 주소를 추출한다. 추출된 기사 주소로 newspaper 모듈에 직접 URL을 지정함으로써 기사 내용을 수집한다.

<Fig. 3-3>과 <Fig. 3-4>는 Newspaper를 사용한 코드와 수집된 데이터 결과이다.

```

article = newspaper.Article(URL, language='ko')
article.download()
article.parse()
    
```

Fig. 3-3 Newspaper Code

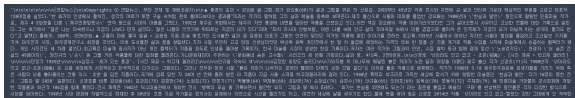


Fig. 3-4 Newspaper Result

### 3.3 한국어 자연어 처리

한국어 자연어 처리를 위해서 다음과 같은 순서를 가진다.

첫 번째, 기사 내용에서 분석에 불필요한 문자열인 이메일과 전화번호 문자를 제거한다.

두 번째, 작가별 단어 빈도수와 키워드를 알아보기 위해 명사 추출을 한다. 이 때에 명사를 추출하기 위해서 형태소 분석이 필요하다. 형태소 분석을 위해서는 기사의 글자나 단어, 어절, 문장처럼 작은 의미 단위인 토큰(token)으로 나누어야 하는데 이를 토큰 분리(tokenization)라고 한다. 토큰 분리 후에는 형태소 사전과 품사 사전을 활용하여 어간 추출(stemming)과 품사부착(Part Of Speech tagging, POS-tagging) 과정을 거친다(Park, 2016).

세 번째, 작가와 작가, 작가와 단어의 의미 유사성을 알아보기 위해 형태소 분석을 한 데이터를 가지고 Word2Vec의 Skip-gram 모델을 사용한다.

#### 3.3.1 문자열 제거

\n\n제임스 터렐, 아틀란티스\n\n\n\n도상봉, 고궁  
 풍경, 37.9×45.5cm, 캔버스에 유채, 1973\n\n\n\n\n지  
 난해 한국국제아트페어\n\n\n[연합뉴스]\n\nCopyrights  
 © 연합뉴스 무단 전재 및 재배포금지\n\n국내 최  
 대 미술품 장터인 한국국제아트페어(KIAF)가 25~  
 29일 강남구 삼성동 코엑스에서 장을 열고 약 1만  
 점의 작품을 거래한다.주최기관인 (사)한국화랑협회  
 는 18일 기자간담회에서 18회제를 맞이한 KIAF 열  
 개를 공개했다.  
 (하략)

Fig. 3-5 Data Example

<Fig. 3-5>과 같이 명사 추출과 문장과 관련 없는 문자열인 이메일, URL, 전화번호, 한글 자음 및 모음, 특수 기호 제거를 위해서 Python의 정규 표현식을 사용하여 데이터 전처리를 한다. 정규 표현식은 검색 패턴에 일치되는 문자열을 원하는 문자열로 바꿔준다. 본 논문은 패턴에 제거하고 싶은 이메일, URL, 전화번호, 한글 자음/모음, 특수 기호 패턴을 만들고, 바꿀 문자열은 공백으로 지정하였다. 데이터 전처리를 한 결과는 <Fig. 3-6>과 같다.







을 주고 작가별 주변 단어를 예측했다. 또한 앞의 3.3.2 절에서 수행한 명사를 추출한 데이터를 불러오고, 중심 단어로부터 좌우 10개 단어, 최소 3회 이상 출현단어로 단어 의미의 유사성을 계산하였다.

<Fig. 3-11>은 상위작가 간 상위 13개 단어 유사도 결과이다. 보면 ‘장욱진’ 단어에는 김환기 작가와 박수근 작가가 0.60 이상의 유사도를 가지고 단어가 중복됨을 알 수 있다.

키워드	김환기	이우환	박서보	정상화	천경자	박수근	김장겸
김필데오	0.72326529						
황금주	0.700553417						
정화	0.656545997						
유영국	0.656276047						
모닝스타	0.629343002						
장욱진	0.629072845					0.645925581	
박수근	0.612311482					0.643274546	
남관	0.601460993						
야상곡	0.600616932						
푸르름	0.59142971						
정상화		0.819935262	0.871290684				0.697657943
박서보		0.816536546		0.871290565			0.636865135

Fig. 3-11 Example of Word Similarity Results Between Top Artists

#### 4. 시스템 구현 결과

본 시스템은 국내 미술시장의 객관적인 정보 전달을 위해 K-Artprice의 정형 데이터와 매일경제 작가 검색 결과인 비정형 데이터를 융합하여 현재 시장에서 작가에 대한 정량적과 정성적 분석을 통해 객관적으로 정보를 제공한다. K-Artprice과 매일경제 데이터 수집에서 웹 크롤링 기술로 Selenium,

BeautifulSoup, 그리고 Newspaper를 사용하여 작가의 정보를 수집하고, 작가를 최근 5년 낙찰액으로 상위/중위/하위 작가로 나누었다. 매일경제사이트에서 [문화산업]과 [공연/전시/음악] 카테고리로 검색한 기사들에 대한 자연어 처리를 위해서 오픈소스 형태소 분석기인 한글 자연어처리 패키지인 KoNLP를 사용하였고, 단어 간 의미의 유사성은 Word2Vec을 사용하여 계산하였다.

Tab. 4-1 Top 10 Results by Top Artist MK articles Word Frequency

단어	빈도수	단어	빈도수
서울	6300	대표	4347
옥션	5504	그림	4256
현대	5123	화랑	3771
김환기	4881	국내	3723
아트	4793	세계	3620

Tab. 4-2 Top 10 Results by Bottom Artist MK articles Word Frequency

단어	빈도수	단어	빈도수
그림	1110	갤러리	793
서울	1007	예술	749
미술관	897	아트	740
문화	839	대표	729
대통령	827	현대	615

Tab. 4-3 Top 10 Results by Middle Artist MK articles Word Frequency

단어	빈도수	단어	빈도수
갤러리	1811	현대	978
아트	1554	화랑	966
미술관	1443	대표	903
서울	1098	예술	873
그림	976	문화	725

<Tab. 4-1>, <Tab. 4-2>, 그리고 <Tab. 4-3>은 매 일경제 기사들을 형태소 분석을 통해 출현 빈도가 많은 상위 단어 10개에 대한 결과이다. 유의미한 분석을 위해서 작품, 작가, 경매, 미술, 전시, 화가의 단어는 제거하였다. 하지만 빈도수만을 가지고 기사

에 있는 주요 단어를 한눈에 파악하기가 어렵다.

아래 그림 <Fig. 4-1>, <Fig. 4-2>, 그리고 <Fig. 4-3>은 출현 빈도가 많은 상위 단어들의 시각화한 결과이다. 각 상위, 중위, 하위 작가에 속하는 특징에 따라 출현 빈도가 다를 수 있다.



Fig. 4-1 Visualize Word Frequency for Top Artist articles



Fig. 4-2 Visualize Word Frequency for Middle Artist articles



Tab. 4-6 Word2Vec by Bottom Artist

단어	중복	작가 수	중복 작가		
문학진	4	김인승	박득순	박영선	변중하
임직순	4	김인승	박득순	박영선	변중하
김상유	3	김인승	박득순	변중하	
손용성	3	박득순	박영선	변중하	
오용	3	김병중	김형근	이두식	
이종우	3	김인승	박영선	변중하	
강형구	2	김동유	홍경택		
권수현	2	류병엽	이만익		
노은	2	김동유	홍경택		
달주	2	박득순	박영선		

그림 <Fig. 4-4>는 K-Artprice에서 최근 5년동안의 낙찰금액으로 작가들을 상위, 중위, 하위 구간으로 나누어 작가의 전체 정보를 시각화한 결과이다. 이 그림은 수집한 기사 정보의 날짜를 기준으로 기간을 설정하고, 작가들의 전체 기사 건수를 제공하는 상위 작가 분석 시스템 화면이다.

이를 통해서 작가의 주 화법과 비슷한 화법을 사용하는 작가들을 파악할 수 있었으며, 정형 데이터와 공신력 있는 시가를 활용한 비정형 데이터를 함께 활용함으로써 단어들의 연관성 분석에 대한 신뢰성을 확보했다.



Fig. 4-4 Artist Analysis System

그림 <Fig. 4-5>는 상위작가 김환기의 연관 단어들을 분석한 시각화 결과이다. 그림에서 알 수 있듯이 김환기를 중심으로 박수근, 유영국, 장욱진 등 상위 작가 간의 연관성이 높은 결과를 보여준다. 또한, 김환기 작가 작품의 특징인 점화, 푸르름과 같은 단어가 나타났다.

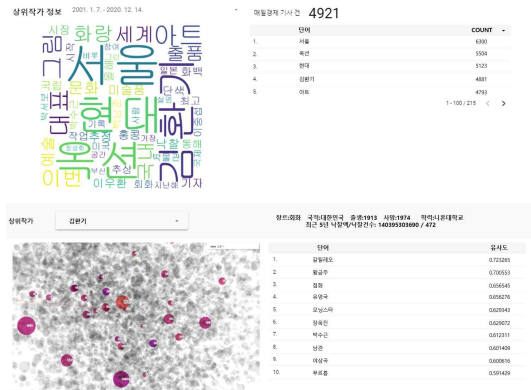


Fig. 4-5 Kim Whan Ki Display

## 5. 결론

세계 미술시장이 성장하면서 미술품은 감상을 위한 소비재일 뿐만 아니라 투자의 대상이 되었다. 미술품의 가치에 관한 연구가 이루어져 왔으며 미술품의 가치를 시장 논리에서 다루는 것에 대한 비판이

거듭 제기되고 있고, 문화와 예술의 대상 이유로 경제적인 논리를 적용하여 분석하는 것이 어려웠다.

최근 미술품을 재테크의 수단으로 인식하여 투자의 대상물로 그 가치를 연구하고 다양한 학문의 영역으로 다각화된 분석이 필요하다(Woo, 2011). 본 연구는 현재 미술시장의 유통 영역에서 정보 제공이 폐쇄적이고 불균형적인 문제를 객관적이고 신뢰성 있는 통계자료를 수집 및 분석을 하고 결과를 제공하는 시스템 구현에 목적을 두었다.

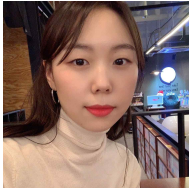
본 연구에서는 정형 데이터인 K-Artprice의 국내 작가 100명을 대상으로 최근 5년 낙찰액 기준으로 작가를 상위, 중위, 하위로 나누었고 작가별 비정형 데이터인 매일경제 기사를 분석하였다. 분석 결과 작가들을 상위, 중위, 하위 구간별로 연관이 높게 측정되고 상위, 중위, 하위 작가 중에서도 상위에 속한 작가들의 특징이 두드러지게 나타났다. 또한, 작가 간 중복으로 언급이 많이 되는 단어가 나타나고 있어서 주로 같이 활동하는 작가가 누구인지 알 수 있었고, 작가 간의 사적인 관계가 있다는 해석도 가능함을 보였다. 또한 원하는 화법, 색상을 주면 가장 유사도가 높은 작가나 그림명을 알 수 있다. 따라서 본 연구는 소비자가 미술품 구매에 폭넓은 정보를 제공하는 도구가 될 수 있다.

하지만, K-Artprice의 경우 등록되어 있는 국내 작가는 최근 5년간 낙찰액이 최소 2억 이상인 유명 작가이며 정보가 충실하게 제공이 되고 있고, 이는 국내 협회 등록 작가 4만 명 중 0.5% 상위작가의 정보만을 제공하고 있다. 이 정보는 충분치 않은 숫자이며 따라서 많은 정보를 제공하기 어려웠다. 그리고 신진 작가들의 정보를 제공하는 사이트의 경우에도, 작가의 기본 정보만을 제공하고 있어 작가의 트렌드, 특징 및 특성에 대해 알 수 없었다. 신진작가의 정보가 충분치 않아 소비자는 구매에 어려움이 있을 수 있음을 파악하였다. 국내 미술시장의 활성화와 성장을 위해서는 국내 원로작가뿐만 아니라 신진작가들의 객관적이고 신뢰성 있는 정보 제공이 이

루어져야 한다.

## Reference

- [1] Ministry of Culture(2019), Survey on the Art Market, Korea Arts Management Service 2018. (문화체육관광부(2019), 미술시장실태조사 2018 기준)
- [2] Lee KiChang(2019), Sentence Embedding Using Korean Corpora, Acom Publishing Co., 95-130. (이기창(2019), 한국어임베딩, 에이콘출판, 95-130)
- [3] Mikolov et al(2013a), Efficient Estimation of Word Representations in Vector Space.
- [4] Mikolov et al(2013b), Distributed Representations of Words and Phrases and their Compositionality.
- [5] Oh MiRa(2015), A Study on How to Boost the Art Market -Focusing on young artists program. (오미라(2015), 미술시장의 활성화 방안 연구 -신진작가 발굴 시스템 연구 중심으로)
- [6] Park Deamin(2016), Natural Language Processing of News Articles : A Case of NewsSource beta, Korean Society For Journalism And Communication Studies, 4-52. (박대민(2016), 뉴스 기사의 자연어처리: <뉴스소스 베타>를 중심으로, 한국언론학회, 4-52)
- [7] Woo SookYi(2011), A Study on the Analysis of Arts Market Status and Potential Investment of an Art Work. (우숙이(2011), 미술시장 현황과 미술품의 투자 가치 분석에 관한 연구)
- [8] WIKIPEDIA(2004), Beautiful Soup. (위키피디아(2004), 뷰티풀 스푼)
- [9] WIKIPEDIA(2010), NLP. (위키피디아(2010), 자연어처리)
- [10] WIKIPEDIA(2017), Scraping (위키피디아(2017), 스크래핑)
- [11] WIKIPEDIA(2020), Selenium. (위키피디아(2020), 셀리니움)



**Lee, Seung Kyung (ketia56@naver.com)**

Lee Seung-kyung received a bachelor's degree from Kongju University. She is currently working at Wiseite. Her research interests include art and machine learning.



**Lim, JongTae (jtlim@kongju.ac.kr)**

Lim JongTae is a professor in Division of Computer Engineering of Kongju National University. He received her M.S. and Ph.D. in KAIST. His current research area is BigData Processing, Deep Learning, and Transaction Processing in Blockchain Mechanism.

# Art transaction using big data Artist analysis system implementation

SeungKyung Lee\*, JongTae Lim\*\*

## ABSTRACT

The size of the domestic art market has increased 21.9% over the past five years as of 2018 to KRW 448.2 billion and the number of transactions has also increased 31.6% to 39,367 points maintaining growth for the fifth consecutive year. Art distribution platforms are diversifying from galleries and auction-style offline to online auctions. The art market consists of three areas: production (creation), distribution (trade), and consumption (buying) of works and as the perception of artistic value as well as economic value spreads interest is also increasing as a means of investment. Consumers who purchase works and think of them as a means of investment technology have an increased need for objective information about their works, but there is a limit to collecting and analyzing objective and reliable statistics because information provision in the art market distribution area is closed and unbalanced.

This paper identifies objective and reliable art distribution status and status through big data collection and structured and unstructured data analysis on art market distribution areas. Through this, we want to implement a system that can objectively provide analysis of authors in the current market.

This study collected author information from art distribution sites and calculated the frequency of associated words by writer by collecting and analyzing the author's articles from Maeil Business, a daily newspaper. It aims to provide consumers with objective and reliable information.

*Keywords: art market, bigdata, crawling, korean NLP, data visualization*

---

\* This study was supported by the National R&D Research Support of the Korea Creative Content Agency.  
\* First Author, Graduated from Kongju National University with a bachelor's degree in computer engineering.  
\*\* Corresponding Author, Professor, Division of Computer Engineering, Kongju National University