

대한항공, 아시아나항공 합병에 대한 매체별 인식 차이도 분석



과목명 : 전산통계 1

담당 교수 : 정원일 교수님

조 : 3

학과: 정보통계수리보험학과

이름 : 이원재, 김가영, 전형우

제출일 : 2020. 12. 12

목 차

I. 서론 / 1

- I. 주제 설명 및 선택 이유 / 1
- II. 예측 결과 / 1

II. 본론 / 2

- I. 분석 방식 1 - 설명 / 2
- II. 신문사 데이터 / 4
 - 1) 중앙일보 / 4
 - 2) 한겨레 / 5
 - 3) 경향신문 / 6
- III. 웹사이트 데이터 / 6
 - 1) 다음 / 7
 - 2) zum / 8
- IV. 방송사 데이터 / 9
 - 1) SBS / 9
- V. 분석방식 2 - 설명 / 9
- VI. 방송사 데이터 / 11
 - 1) kbs, sbs 차이 분석 / 11

III. 결론 / 12

- I. 실제 분석결과와 예상결과 차이 비교 / 12
- II. 매체 별 차이 / 12
- III. 추가 분석 방향 / 12

IV. 부록 / 13

- I. 참고 문헌 / 13

I. 서론

I. 주제 설명 및 선택 이유

본 보고서는 지난 11월 중순 화제가 되었던 대한항공/아시아나 항공의 합병 이슈에 대해 웹에서 포털사이트, 신문사, 방송사 별로 어떤 주요 단어의 출현 빈도가 잦았는 지에 대해 비교하고 분석하는 보고서이다.

대한항공과 아시아나항공의 합병 소식이 처음 언론에 보도된 시점은 2020년 11월 12일이다. 따라서 이 보고서가 작성되고 있는 2020년 12월 12일과 정확히 한 달의 차이가 있기 때문에 11월 12일부터 1달간 웹 사이트에 올라왔던 기사들을 분석해, 합병 소식이 퍼지기 시작한 이후 어떤 단어의 출현 빈도가 잦았는지 분석을 하기에 시기가 알맞다는 생각이 들어 이 주제를 선정하게 되었다.

또한 대한항공과 아시아나항공의 합병에 대해 합병 이후 대한항공이 국제선 여객노선과 주요 화물노선의 점유율의 합을 70% 이상으로 가져가게 되면서, 사실상 국내 항공시장의 지배적 사업자를 넘어 독점 사업자에 가까운 힘을 형성할 것으로 보아 긍정적인 여론도 있었지만, 부채비율이 굉장히 높은 두 회사간의 합병에 대해 빛이 더 커지는 대한항공에 대해 우려하는 부정적인 여론도 있었기 때문에 두 키워드에 등장하는 단어의 비율로 이를 파악하기 쉬울 것이라는 생각이 들었다.

II. 예측 결과

각 웹사이트 별로 각 '대한항공' 키워드와 '아시아나항공' 키워드를 검색했을 경우, 두 회사의 인수합병 소식이 가장 많이 나왔을 것이기에 '대한항공'을 검색했을 때 '아시아나항공'의 출현 빈도가 가장 잦을 것이고 '인수', '합병' 등의 단어도 출현 빈도가 잦을 것이라고 생각되고, 그 반대의 경우도 그럴 것이라고 예상된다.

또한 두 회사의 합병 이후 항공시장의 최대 독점 기업이 될 것이므로, 이와 관련해 에어서울, 제주항공, 진에어 등 저가항공사의 입장과 상황에 대한 기사 역시 많아질 것으로 예측이 되어 '에어서울', '제주항공' 등의 단어도 데이터에 잡힐 것으로 생각된다.

그리고 12월 이후로 코로나 확진자의 증가 추세가 엄청났는데, 이에 따라 타격을 입은 항공시장에 대한 기사도 많을 것으로 보아 '코로나' 키워드나 '관광' 등의 기존 코로나와 관련한 항공시장의 이슈에 대한 키워드도 종종 보일 것으로 예상된다.

II. 본문

I. 분석 방식 1 – 설명

먼저 교수께서 수업 시간에 웹 크롤링에 대해서 알려주신 방법에 관해서 설명해보도록 하겠다. 이용 코드의 예시는 본 팀의 웹 크롤링 예시 중 중앙일보에서 '대한항공'이란 키워드를 이용해 추출할 때 이용한 코드를 이용했다.

```
library(KONLP)
library(rvest)
library(httr)
library(XML)
warning()
```

우선 진행을 위해서 필요한 library들을 인코딩 해주었다.

```
url1 = 'https://news.join.com/Search/JoongangNews?page='
url2 = '&keyword=%EB%8C%80%ED%95%9C%ED%95%AD%EA%B3%B5&PeriodType=OneMonth&SortType=New&SearchCategoryType=JoongangNews'
news_page = list()

for(page in 1:24){
  url_joongang = paste0(url1,page,url2)
  web = GET(url_joongang)
  html = htmlTreeParse(web, useInternalNodes = T, trim = T, encoding = 'utf-8')
  rootNode = xmlRoot(html)
  title = c()
  for(i in 1:10){
    xpath1 = '//*[@id="content"]/div[2]/div[2]/ul/li['
    xpath2 = ']/div/h2/a'
    xpath = paste0(xpath1, i, xpath2)
    news = xpathSApply(rootNode, xpath, xmlValue)
    title[i] = news
  }
  news_page[[page]] = title
}
news_page
news_page1<-unlist(news_page)
news_page1
```

크롤링 하려는 중앙일보의 주소는 다음과 같았다.

'https://news.join.com/Search/JoongangNews?page=11&keyword=%EB%8C%80%ED%95%9C%ED%95%AD%EA%B3%B5&PeriodType=OneMonth&SortType=New&SearchCategoryType=JoongangNews'

위 주소 속에서 총 페이지 속에서 페이지가 하나씩 넘어갈 때마다 빨간색 부분만 바뀌기 때문에 그 부분을 제외하고 앞부분을 url1, 뒷부분을 url2로 나누고 for 함수에 넣어줘서 page를 계속해서 증가시켜줬고, paste0함수를 이용해 url1, page, url2를 묶어 페이지에 해당하는 url 주소로 만들어 주었다. 이후 이 주소를 GET() 함수로 바꿔 web이란 변수에 저장시켜 이를 htmlTreeParse() 함수를 이용해 html 변수에 넣어줬고, title이란 빈 벡터를 지정해줘 각 페이지에서 원하는 내용을 넣어주려고 했다.

이후 다시 한번 for()문이 등장한다. 뉴스 페이지 속에서 얻고자 했던 기사 제목 각각의 주소가 '//*[@id="content"]/div[2]/div[2]/ul/li[1]/div/h2/a' 다음과 같았고, 이 역시 page가 변하는 것처럼 몇 번째 줄에 있는 기사인 지에 따라 값이 1~10으로 바뀌기 때문에 이를 나눠주고 이후

paste0()으로 다시 묶어 주는 방안을 택했다. 이후 앞서 있던 rootNode와 만들어낸 주소의 위치를 xpathSApply() 함수를 통해 기사 제목을 얻어낼 수 있었고, 이를 title의 i번째 자리에 입력시켜 줬다. 이중 for문 중 밖에 있는 for문에서 앞서 지정해준 빈 리스트 값인 news_page에 크롤링해온 값들을 저장해주었다. 이와 같은 방법을 총 24번 반복했고, 반복문을 통해 나온 값을 unlist() 해 list 화를 풀어줬다.

```
#전처리
news_pre<-gsub("[\r\n\t]",'',news_page1)
news_pre
news_pre1<-gsub("[[:punct:]]","",news_pre)
news_data = news_pre1
str(news_data)
news_data.df<-as.data.frame(news_data)
str(news_data.df)
news_data.df
```

그 데이터를 gsub() 함수를 이용해 각각 전처리해주었고, 전처리 된 데이터를 as.data.frame을 이용해 dataframe 형태로 바꿔 새로운 변수 news_data.df에 저장시켜줬다.

```
#파일 저장과 읽기
setwd("C:/Users/user/Desktop/2020 - 2/전산통계1/기말프로젝트")
write.csv(news_data,"news_data_JA.csv",quote = F)
news_data<-read.csv("news_data_JA.csv",header = T,stringsAsFactors = F, encoding = 'utf-8')
str(news_data)
names(news_data)<-c("no","news_text")
head(news_data)
```

이후 수집한 자료를 반영구적으로 이용하기 위해서 write.csv()함수를 이용해 csv파일로 저장해 줬고, 이를 read.csv()로 다시 읽어내 data.frame으로 열어줬다. 이후 column이름을 'no',와 'news_text'로 지정해주었다.

```
#말뭉치 생성
newsCorpus<-Corpus(VectorSource(news_data))
newsCorpus
inspect(newsCorpus)
TDM<-TermDocumentMatrix(newsCorpus,control=list(wordLengths=c(4,16)))
Encoding(TDM$dimnames$Terms) = 'utf-8'
TDM
news.df<-as.data.frame(as.matrix(TDM))
str(news.df)
```

VectorSource() 함수와 Corpus() 함수를 이용해 말뭉치 객체를 생성했다. 이후 inspect()을 이용해 어떻게 생성됐는지를 확인해주었다. 선별한 말뭉치를 TermDocumentMatrix() 함수를 이용해 길이가 2~8개 사이의 단어만 선별해서 행렬을 만들어줬다. 'utf-8'을 인코딩해 한글 또한 가능하게 만들어 줬고, 선별한 TDM 변수를 행렬로 만들어준 뒤, 다시 dataframe화 시켜주었다.

```
#단어 출현 빈도수 구하기
wordResult<-sort(rowSums(news.df),decreasing = T)
wordResult[1:10]
#단어구름 생성
myNames<-names(wordResult)
myNames
```

rowSums() 함수를 이용해 중복되는 것들을 모두 합쳐주었고, 이를 sort() 시켜줘 가장 많은 빈도로 나온 것부터 나열할 수 있게 되었다. 이를 wordResult 변수에 입력시켜주고, 상위 10개의 단어를 확인해보았다. 이후 names()를 이용해 단어 이름을 추출했다.

```
#단어 빈도수 구하기
df<-data.frame(word=myNames,freq=wordResult)
head(df)
df
#단어구름 생성
pal<-brewer.pal(12,"Paired")
wordcloud(df$word,df$freq,min.freq = 2,
          random.order =F,scale=c(4,0.7),
          rot.per = .1,colors = pal,family="malgun")
```

추출한 전체 단어와 단어 출현 빈도수를 이용해 dataframe 생성해주었다. 마지막으로 brewer.pal()을 이용해 전체 12가지 색으로 먼저 나눠준 뒤, wordcloud() 함수를 이용해 검색 단어와 연관되어 나오는 단어들을 구름화 시켜 시각화해 주었다.

II. 신문사 데이터 분석

1) 중앙일보



- '대한항공'을 검색했을 때 -



- '아시아나항공'을 검색했을 때 -

다음은 중앙일보에서 '대한항공'과 '아시아나항공' 키워드를 검색했을 때 나오는 결과이다. 검사한 세 가지 신문사 중 가장 많은 정보가 있었기 때문에 단어 구름을 통해 시각화해 봤을 때 뽁뽁하게 많은 단어가 구름을 채우고 있는 모습을 볼 수 있다. 각각 가장 많이 나온 내용은 물론 자기 자신이었고, 다음으로 상대 단어가 나오는 것을 봐 지난 한 달간 두 기업의 합병 소식이 눈에 띄는 내용이었음을 알 수 있다. 공통으로 '구조조정', '합병', '항공빅딜' 등의 단어가 나타나는 것이 이를 더욱 힘써주고 있다.

주목할 점은 둘의 차이점이다. '대한항공'을 검색했을 때만 '백신'이란 키워드가 나오는 것을 확인할 수 있다. 이는 대한항공이 대한민국 국적기로서 국가에서 필요한 백신을 이송하는 역할을 한다는 것을 봐 한국 내 지위가 더 높음을 알 수 있다. 이 같은 추측은 '김해신공항'이란

키워드를 통해 또 알 수 있다. 최근 김해 인근 신공항의 백지화와 관련해 많은 여론이 오가고 있다. 항공업에 매우 밀접한 공항 관련 키워드가 '대한항공' 검색 시에만 나오는 것을 봐 두 회사의 규모를 모르는 사람도 단어들을 보고 대한항공이 국가와 관련된 더욱 큰 기업이란 것을 알 수 있다. 그리고 '아시아나항공'을 검색했을 땐 '대한항공', '인수합병'과 관련된 내용이 대다수인 거에 비해 '대한항공'에선 위 단어와 관련되지 않는 단어도 상대적으로 많다는 점에서도 확인할 수 있다.

재밌는 점은 코로나 상황이 이전에 등장하지 않았을 법한 단어가 등장했다는 것이다. 공통으로 '국제관광비행'이란 단어가 등장했다. 이는 최근 등장한 비행기를 타고 한반도 한 바퀴를 돌고 오는 비행을 의미한다. 코로나 이전엔 없었던 관광 형태가 현재 키워드 검색에서 나타나는 것을 봐, 현재 비행 관광 형태 및 상황을 알아볼 수 있다.

2) 한겨레



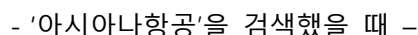
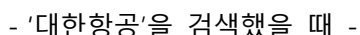
- '대한항공'을 검색했을 때 -



- '아시아나항공'을 검색했을 때 -

다음은 한겨레신문에서 두 단어를 검색했을 때이다. 우선 앞선 중앙일보와의 단어 구름을 비교해봤을 때, 전체적인 구름의 밀도, 단어의 개수에서 큰 차이를 보인다는 것을 알 수 있다. 이에 상대적으로 두 신문사 간의 영향력, 규모, 그리고 해당 문제에 관한 관심도 차이를 알 수 있다. 공통 단어는 앞선 분석과 비슷하게 나타난다. 양사의 단어, '구조조정', '통합', '인수', '점유율', 'kcgi' 등의 단어가 눈에 띈다. 양사 합병 소식이 나오기 전엔 등장하지 않았을 법한 단어들이 현재 많이 등장하는 것으로 봐 사회적 상황에 대해 알 수 있다. 재밌는 점은 '아시아나항공' 검색 시 나오는 '조원태'라는 단어이다. 한진그룹의 회장인 조원태 회장이 본사를 검색했을 때보다 '아시아나항공'을 검색 시에 나타나고 있다. 이는 인수합병에서 아시아나항공이

3) 경향신문



III. 웹사이트 데이터

앞서 국내의 메인 신문사에서 추출한 데이터의 분석에 대해 보았다. 이번에는 국내의 포털사이트인 다음과 Zum 사이트에서 뉴스에 대한항공을 검색했을 경우에 나오는 기사 타이틀들에서 나오는 단어의 빈도들에 대해 확인하고, 단어 구름을 형성했다.

1) 다음(daum)



- '대한항공'을 검색했을 때 -

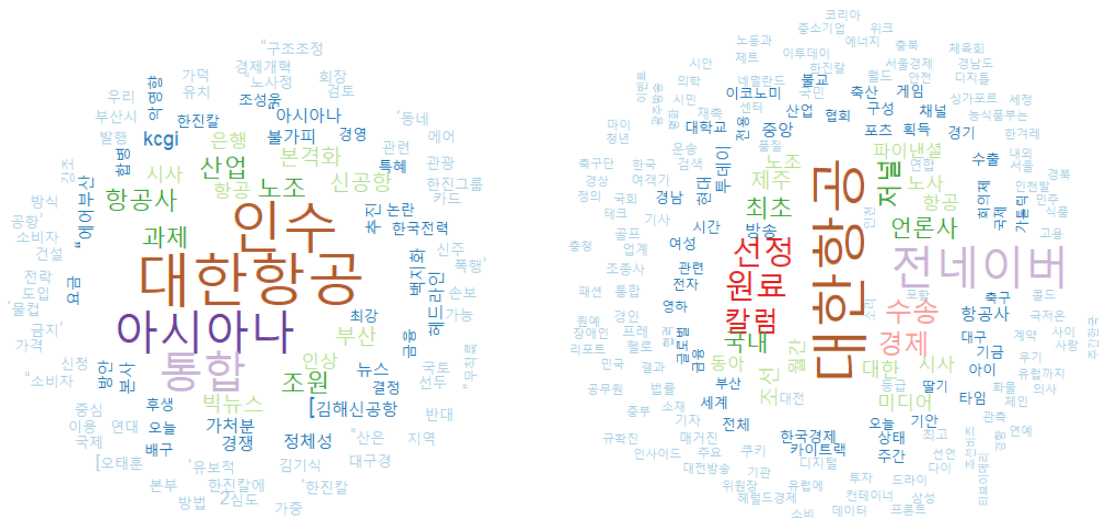


- '아시아나항공'을 검색했을 때

포털사이트 다음에서 '대한항공'과 '아시아나항공'을 검색했을 때 나오는 결과들이다. 이 경우에서 대한항공을 검색했을 때에는 검색어가 대한항공이었음에도 불구하고 아시아나의 빈도가 더 크게 단어 구름에 있을 정도로 대한항공과 아시아나 항공의 높은 연관성을 확인할 수 있었다. 그리고 '구조조정' '경영권' 등의 단어도 눈에 띄는데, 이는 대한항공의 사업구조가 아시아나를 인수함에 따라 변경이 될 것이라는 여론의 기대감이 반영된 것과 아시아나 인수에 따른 경영권의 변화 등이 사회의 큰 관심거리라는 것을 뜻한다. 그리고 아시아나항공을 검색했을 때의 단어 그림에서 가장 크게 보이는 단어는 기간산업안정 기금을 뜻하는 '기안기금' 이었다. 아시아나 항공이 기안기금의 지원 대상이었기 때문에 그런 것으로 보인다. 또한, '제주항공'이 아시아나에 이어 기안기금의 2호 지원 대상이 되었기 때문에 제주항공의 단어 등장이 많은 것으로 보인다. 이것은 최근 들어서 기안기금이 아시아나 항공과 제주항공의 큰 이슈였다는 것을 뜻하고, '321억' 은 기안기금의 액수로 보인다. 또한 다른 저가 항공회사인 '에어서울' 그리고 아시아나 항공과 이슈가 있던 '대한항공' 역시 보인다.

주목할 점은 대한항공을 검색했을 때의 결과에서 아시아나 항공과의 합병 이슈가 더 많이 눈에 띈다는 것이다. 대한항공보다 '아시아나', '인수' 등의 단어가 더 크게 표시될 정도로 대한항공 검색기록에선 아시아나 항공과의 합병 이슈가 더 큰 관심거리였지만, 아시아나 항공의 검색기록에서는 '기안기금' '제주항공'등 현재 아시아나 항공의 상황인 기안기금이 더 주된 관심거리였다. 이는 아시아나 항공은 이미 인수가 되기로 확정된 회사이기 때문에 더 신경 쓸 요소가 없고, 오히려 아시아나를 인수하게 되면서 대한항공이 사업구조와 경영권 등 많은 문제를 해결해야 하기에 대한항공 쪽에 아직 인수와 관련된 이슈가 많이 나오고 있다고 생각된다.

1) SBS



- '대한항공'으로 검색했을 때-

- '아시아나'로 검색했을 때

방송사 SBS에서 각각의 키워드로 검색해보았을 때 제일 많이 언급된 키워드는 아무래도 대한항공임을 알 수 있다. 따라서 대한항공이 아시아나와 인수합병한 것이 제일 주된 이슈임을 두 그림을 통해서 파악할 수 있다. 여기서 주목해볼야 할 점은 '대한항공'을 검색했을 때 '아시아나', '통합', '인수'등과 더불어, '노조', '항공사', '가처분'등 많이 나온 반면에, '아시아나'로 검색했을 땐 '대한항공', '수송'등 이외에는 언급이 적음을 알 수 있다. 대한항공이 인수합병기업이고, 아시아나가 피인수합병 기업이기 때문에, 이러한 내용이 '대한항공' 측에서 더 많이 언급되었다고 짐작해볼 수 있다. 또한, 대한항공이 아시아나와 인수합병을 한다면 주가에도 영향을 미칠 것이기 때문에, 주식시장이 활발한 지금, '김해신공항', '인수', '한진 칼', 'kcgi' 등과 같은 연관 키워드인 나온 것이라고 볼 수 있다.

V. 분석방식 2 - RSelenium을 이용

웹 크롤링엔 여러 가지 방법이 있다. 앞서 언급한 방식처럼 탐색할 페이지의 url과 타이틀의 xpath를 찾아 조건문을 이용하여 수동으로 데이터를 모을 수 있다. 하지만 이 모든 과정이 수동으로 이루어지며 데이터들의 정보를 직접 찾아서 코드를 만들어야 하기 때문에 많은 양의 데이터가 필요하거나, 경로가 복잡해질 경우엔 어려움이 발생할 수 있다. 또한 우리가 보는 대부분의 웹 페이지는 동적 웹 페이지로, 서버에 있는 데이터들이 스크립트에 의해 가공 처리 된 후 생성되어 웹 페이지에 전달되어 웹 페이지가 계속해서 바뀌게 된다. 따라서 일반적인 방법으로는 크롤링이 어려울 수 있다. 이러한 경우에 사용하기 용이한 것이 RSelenium이다. 여기서 Selenium이란 웹 애플리케이션을 테스트하기 위한 포터블 프레임워크를 뜻하는데, webdriver라는 API를 이용하여 운영체제에 설치된 브라우저를 제어한다. 브라우저를 직접 동작시키는 것은 이러한 동적

웹페이지들의 데이터 또한 가져올 수 있으며, 코드를 구성해놓으면 크롤링이 자동으로 진행된다는 장점이 있다.

RSelenium을 이용하여 웹 크롤링을 진행한 방식은 다음과 같다. 일반적으로 클릭하여 페이지를 이동하고 값을 찾는 방식이 아닌 함수만을 이용하여 사용자의 제어 없이 크롤링을 진행하게 한다.

```
#RSelenium을 사용한 웹크롤링
library(RSelenium)
library(seleniumPipes)
library(rvest)
library(httr)
|
#파싱
remDr = remoteDriver(
  port = 4445L,
  browserName="chrome"
)
remDr$open()
remDr$navigate("https://news.kbs.co.kr/search/search.do#1")
webElem <- remDr$findElement(using = 'xpath', value='//*[@id="input-search"]')
search_word= "대한항공"
webElem$sendKeysToElement(list(search_word))
search1<- remDr$findElement(using="xpath", value = '//*[@id="content"]/div/div[1]/div[1]/fieldset/button')
search1$clickElement()
```

우선 명령 프롬프트를 연 다음에 selenium이 실행될 수 있는 환경을 갖춘 후 필요한 패키지를 설치하여 준다. 이후 다음의 함수들을 사용하여 진행한다.

1) RSelenium을 이용하여 해당 브라우저를 실행시키기

- remDr\$open(): 원하는 브라우저의 창을 열어준다.
- remDr\$navigate("url"): 지정된 url에 들어간다.
- remDr\$findElement(using=,value=): 접속된 페이지 안에 해당 xpath 값을 찾는다.
- webElem\$sendKeysToElement(): 원하는 위치에 찾고자 하는 값을 넣는다.
- Search1\$clickElement(): 클릭하고자 하는 버튼의 값을 지정 후 클릭하도록 한다.

```
head_list = list()
j = 2;
n = 0;
while(n < 13){
  head_page<-c()
  for (i in 1:10){
    print(i)
    temp = NA;
    try({temp<- remDr$findElement(using="xpath", value = paste0('//*[@id="searchList"]/li[' ,as.character(i),']/a/span[2]/em'))
    temp<- unlist(temp$getElementText())
    })
    head_page[i] <- temp
    Sys.sleep(1)
  }
  next1<- remDr$findElement(using="xpath", value = paste0('//*[@id="pagingList"]/li[' ,as.character(j),']/a'))
  next1$clickElement()
  n = n+1
  head_list[[n]]<-head_page
  Sys.sleep(10)
  if (j <6){
    j = j + 1
  }
}
```

2) 해당 페이지에 접속까지 마친 후, 페이지 별로 그 안에 있는 타이틀을 수집한다.

보통 url과 타이틀의 xpath 값은 해당 구역에 맞춰진 하나의 숫자 값을 제외하고는 모두 공통된다. 따라서 데이터를 모으고자 하는 범위를 설정하여, 페이지의 쪽수와 타이틀의 개수를 선택한다.

- while 문과 for 문을 이용하여 반복적으로 추출한다. (페이지의 경우 13쪽, 타이틀 10개)

두 방송사 모두 대한항공과 더불어 '아시아나', '인수', '통합'이라는 키워드가 제일 많이 추출되었다. 이외에 공통적으로 언급된 키워드로는 '가격 인상', '저가항공사', '구조조정', '악영향', '소비자' 등이 보이는 것으로 보아 인수합병에 따른 여파와 독과점에 대한 부정적 인식이 많이 있음을 알 수 있었다. 두 방송사 모두 비슷한 키워드가 도출되었다. 위에서 분석한 SBS가 주식시장에 더욱 초점이 맞춰져 있는 것과 다르게 MBC와 KBS는 인수합병에 따른 소비자의 여론을 더욱 반영한 것으로 보인다.

III. 결론

I. 실제 분석결과와 예상결과 차이 비교

상대적으로 예상결과와 실제 분석결과가 유사하게 나왔다. 대한항공/아시아나항공의 이슈가 11월 중순에 매우 화제였던 만큼, 대한항공을 검색했을 경우 아시아나항공이, 아시아나항공을 검색했을 경우 대한항공이 서로 가장 많은 빈도로 출현할 것이라고 생각했다. 실제 이와 같은 결과를 보여주었다. 또한, 에어서울, 제주항공 등 다른 항공사의 단어 출현 빈도 역시 많을 것으로 예측했는데, 역시 아시아나 항공 합병 이후 기간산업안정기금의 수혜자가 될 제주항공이나, 다른 항공사인 에어서울의 출현 빈도로 이를 증명해주었다. 그리고 코로나로 인해 새로 등장한 상품 형태인 '관광비행' 등의 키워드가 등장하며 해외에 착륙하지 않고 공중에서 비행 후 다시 돌아오는 '미착륙' 등의 키워드 역시 등장했다. 또한 코로나와 관련해서 백신을 운반하는 대한항공의 역할이 강조되며 '백신' 등의 키워드 역시 눈에 띄었다.

II. 매체 별 차이

동일 매체 별 차이는 매우 흥미롭게 보여 졌다. 특히 신문매체의 차이가 두드러졌다. 보수언론인 중앙일보에선 합병에 대해 다소 긍정적인 단어들을 찾을 수 있었다. 이에 반해, 진보언론인 경향신문에선 다시 부정적인 단어가 계속해서 등장이 두드러졌다. 이를 통해 각 언론사들이 위 상황에 대해서, 평소 정치사회적 가치관이 어떻게 유지하고 있는 지를 확인할 수 있었다.

타 매체 별 보여주는 단어 차이가 보였다. '기안기금'이란 단어가 신문매체에선 등장하지 않았지만, 다른 두 매체에선 높은 등수로 등장하는 것을 찾을 수 있었다. 이에 매체 별 차이도 있긴 하지만, 이는 정보의 양의 차이로 인식할 수 있다. 상대적으로 적은 데이터를 가진 신문매체와 달리 온라인 매체는 더욱 방대한 데이터들을 얻을 수 있기 때문에 여러 방안의 단어를 찾을 수 있었다. 이에 의거해, 정확한 정보 분석을 위해선 표본의 수가 중요하다는 것을 다시 한 번 볼 수 있게 된다.

III. 추가 분석 방향

11월 12일부터 약 1달간의 데이터를 통해 대한항공과 아시아나 항공의 합병 소식이 등장한 직후부터 지금까지의 기사에서는 양 사에게 가장 주된 뉴스였음을 알 수 있다. 이 점을 보

고 난 후 인수합병 이슈 전의 검색 결과에 따른 단어 구름을 보며 시점에 따른 검색어들의 차이점을 판단할 수 있을 것이다.

올해 초 발병한 코로나 바이러스와 관련된 분석 또한 가능할 것이다. 코로나 바이러스에 의해 가장 직접적인 영향을 받은 산업이 항공업이기 때문에 2월부터의 데이터를 수집 후 코로나 바이러스 상황의 격상 기간에 따라 나눠 분석을 진행할 수 있다. 기간 별 코로나 확산율과 관련해서 어떤 키워드 등이 중심인 지를 파악하는 것 또한 재밌을 것이다.

타 산업과 관련 지어 분석을 해보는 것 또한 가능하다. 검색어 속에서도 다수 등장한 ‘국제관광비행’, ‘무착륙’ 같은 여행은 코로나 바이러스로 인한 인간들의 여행에 대한 심리 분석을 이용해 등장한 신상품이라고 여겨진다. 위 상품 이외에도 타 산업, 상황과 함께 분석할 경우 관련 상품을 만들어 현 시점의 최대한의 방안을 택할 수 있을 것이다.

IV. 부록

I. 참고 문헌

“[R.아르] 웹페이지에서 원하는 요소 가져오기(크롤링)”, 가벼운 블로그, 2018.03.01, <https://lightblog.tistory.com/161>.

“RSelenium(셀레니움)을 이용한 동적 웹페이지 크롤링하기”, Henry’s Quantopia, 2019.09.17, <http://henryquant.blogspot.com/2019/09/rselenium.html>.

“[withR]Rselenium을 이용한 웹스크래핑 - 웹드라이버 실행하기”, 신나는연구소, 2019.12.26, <https://medium.com/excitinglab/with-rselenium%EC%9D%84-%EC%9D%B4%EC%9A%A9%ED%95%9C-%EC%9B%B9%EC%8A%A4%ED%81%AC%EB%9E%98%ED%95%91-%EC%9B%B9%EB%93%9C%EB%9D%BC%EC%9D%B4%EB%B2%84-%EC%8B%A4%ED%96%89%ED%95%98%EA%B8%B0-b53bbf99645>.