

In [40]:

```

# 워드클라우드 import
from wordcloud import WordCloud, STOPWORDS

# 이미지 저장 및 크기 변경, 필터적용
from PIL import Image

# # 한글 자연어 처리 패키지
from konlpy.tag import Okt
okt = Okt()

# 자연어 처리 (영어권 언어 처리하는 패키지)
# => 시드폴더 열어서 korean stopwords 직접 만들어 사용.
import nltk

# 빈도수 분석
from nltk import FreqDist
# 자연어 처리에서 토큰화를 수행하는데 사용
nltk.download('punkt')

# 자연어 처리에서 불용어를 제거하는데 사용(리소스 다운)
nltk.download('stopwords')

# stop words import
from nltk.corpus import stopwords

# 단어 토큰화
from nltk.tokenize import word_tokenize

# 그래프 차트 그리기 위한 패키지
import matplotlib.pyplot as plt

# 넘파이 import
import numpy as np

# 판다스 import
import pandas as pd

import os

```

```

[nltk_data] Downloading package punkt to
[nltk_data] C:\Users\TECH2_25\AppData\Roaming\nltk_data...
[nltk_data] Package punkt is already up-to-date!
[nltk_data] Downloading package stopwords to

```

```
[nltk_data]      C:\Users\TECH2_25\AppData\Roaming\nltk_data...
[nltk_data]      Unzipping corpora\stopwords.zip.
```

In [35]:

```
# 불용어 정의
print(stopwords.words('korean'))
```

['첫', '감소', '만에', 'OTT', '실적', '주가', '가입자', '콘텐츠', '증시', '에', '등', '3
5', '의', '것', '↓', '↑', '←', '→', '↖', '↗', '↘', '↙', '넷플릭스', '하여', '"', '""', '"""', '...', '...', '...',
'...', '...', '...', '...', '...', '[', ']', '(', ')', '{', '}', '만한', '의', '좀', '그만', '했으
면', '1위', '2위', '3위', '4위', '5위', '6위', '7위', '8위', '9위', '10위', '첫번째',
'두번째', '세번째', '네번째', '다섯번째', '여섯번째', '일곱번째', '여덟번째', '아홉번째',
'열번째', '은', '이제', '됐다', '수', '.', '...', '...', '!', '?', '이어', '외신도',
'또', '다시', '굴욕', '!', '@', '#', '\$', '%', '^', '&', '*', '은', '는', '이', '가',
'아', '휴', '아이구', '아이쿠', '아이고', '어', '나', '저희', '따라', '의해', '을',
'를', '에', '의', '가', '으로', '로', '에게', '뿐이다', '의거하여', '근거하여',
'입각하여', '기준으로', '예하면', '예를', '들면', '예를 들자면', '저', '소인',
'소생', '저희', '지말고', '하지만', '하지마라', '다른', '물론', '또한', '그리
고', '비길수', '없다', '해서는', '안된다', '뿐만', '아니라', '만이', '아니다', '만
은', '아니다', '막론하고', '관계없이', '그치지', '않다', '그러나', '그런데', '하지만',
'든간에', '논하지', '않다', '따지지', '않다', '설사', '비록', '더라도', '아니
면', '만', '못하다', '하는', '편이', '낫다', '불문하고', '향하여', '향해서',
'향하다', '쪽으로', '틈타', '이용하여', '타다', '오르다', '제외하고', '이', '외에',
'이', '밖에', '하여야', '비로소', '한다면', '몰라도', '외에도', '이곳', '여기',
'부터', '기점으로', '따라서', '할', '생각이다', '하려고하다', '이리하여', '그리하여',
'그렇게', '함으로써', '하지만', '일때', '할때', '앞에서', '중에서', '보는데서',
'으로써', '로써', '까지', '해야한다', '일것이다', '반드시', '할줄알다', '할수있
다', '할수있어', '임에', '틀림없다', '한다면', '등', '등등', '제', '겨우', '단
지', '다만', '할뿐', '딩동', '댕그', '대해서', '대하여', '대하면', '훨씬', '얼마나',
'얼마만큼', '얼마큼', '남짓', '여', '얼마간', '약간', '다소', '좀', '조금',
'다수', '몇', '얼마', '지만', '하물며', '또한', '그러나', '그렇지만', '하지만',
'이외에도', '대해', '말하자면', '뿐이다', '다음에', '반대로', '말하자면', '이와',
'반대로', '바꾸어서', '말하면', '바꾸어서', '한다면', '만약', '그렇지 않으면', '까악',
'툭', '딱', '빠격거리다', '보드득', '비격거리다', '파당', '응당', '해야한다',
'에', '가서', '각', '각각', '여러분', '각종', '각자', '제각기', '하도록하다',
'와', '과', '그러므로', '그래서', '고로', '한', '까닭에', '하기', '때문에', '거니와',
'이지만', '대하여', '관하여', '관한', '과연', '실로', '아니나다를가', '생각한대
로', '진짜로', '한적이있다', '하곤하였다', '하', '하하', '허허', '아하', '거바',
'와', '오', '왜', '째서', '무엇때문에', '어찌', '하겠는가', '무슨', '어디',
'어느곳', '더군다나', '하물며', '더욱이는', '어느때', '언제', '야', '이봐', '어이',
'여보시오', '흐흐', '흥', '휴', '헉헉', '헐떡헐떡', '영차', '여차', '어기여차',
'공공', '아야', '앗', '아야', '칼칼', '졸졸', '작작', '뚝뚝', '주룩주룩', '쇼',
'우르르', '그래도', '또', '그리고', '바꾸어말하면', '바꾸어말하자면', '혹은', '혹
시', '답다', '밋', '그에', '따르는', '때가', '되어', '즉', '지든지', '설령',
'가령', '하더라도', '할지라도', '일지라도', '지든지', '몇', '거의', '하마터면',
'인젠', '이젠', '된바에야', '된이상', '만큼', '어찌됐든', '그위에', '게다가', '점에
서', '보아', '비추어', '보아', '고려하면', '하게될것이다', '일것이다', '비교적',
'좀', '보다더', '비하면', '시키다', '하게하다', '할만하다', '의해서', '연이서',
'이어서', '잇따라', '뒤따라', '뒤이어', '결국', '의지하여', '기대여', '통하여', '자
마자', '더욱더', '불구하고', '얼마든지', '마음대로', '주저하지', '않고', '곧',
'즉시', '바로', '당장', '하자마자', '밖에', '안된다', '하면된다', '그래', '그렇
지', '요컨대', '다시', '말하자면', '바꿔', '말하면', '즉', '구체적으로', '말하자면',
'시작하여', '시초에', '이상', '허', '헉', '허걱', '바와같이', '해도좋다', '해
도된다', '게다가', '더구나', '하물며', '와르르', '팍', '펑', '펼렁', '동안', '이
래', '하고있었다', '이었다', '에서', '로부터', '까지', '예하면', '했어요', '해요',
'함께', '같이', '더불어', '마저', '마저도', '양자', '모두', '습니다', '가까스로',
'하려고하다', '즈음하여', '다른', '다른', '방면으로', '해보요', '습니까', '했어요',
'말할것도', '없고', '무릎쓰고', '개의치않고', '하는것만', '못하다', '하는것이',
'낫다', '매', '매번', '들', '모', '어느것', '어느', '로써', '갖고말하자면', '어
디', '어느쪽', '어느것', '어느해', '어느', '년도', '라', '해도', '언젠가', '어떤
것', '어느것', '저기', '저쪽', '저것', '그때', '그럼', '그러면', '요만한걸', '그

래 ', '그때 ', '저것만큼 ', '그저 ', '이르기까지 ', '할 ', '줄 ', '안다', '할', '힘이',
 '있다', '너 ', '너희 ', '당신 ', '어찌', '설마 ', '차라리 ', '할지언정 ', '할지라도 ',
 '할망정 ', '할지언정 ', '구토하다 ', '계우다 ', '토하다 ', '메쓰겁다 ', '옆사람 ', '튀
 쳇', '의거하여', '근거하여 ', '의해', '따라 ', '힘입어 ', '그 ', '다음', '버금', '두번째
 로 ', '기타 ', '첫번째로 ', '나머지는 ', '그중에서 ', '견지에서 ', '형식으로 ', '쓰여',
 '입장에서 ', '위해서 ', '단지 ', '의해되다 ', '하도록시키다 ', '뿐만아니라 ', '반대로 ',
 '전후 ', '전자', '앞의것 ', '잠시 ', '잠깐 ', '하면서 ', '그렇지만 ', '다음에 ', '그러한
 즉 ', '그런즉 ', '남들 ', '아무거나', '어찌하든지 ', '같다 ', '비슷하다 ', '예컨대 ', '이
 렇 정도로 ', '어떻게 ', '만약 ', '만일 ', '위에서 ', '서술한바와같이 ', '인 ', '듯하다 ',
 '하지 ', '않는다면 ', '만약에 ', '무엇 ', '무슨 ', '어느 ', '어떤 ', '아래윗', '조차',
 '한데 ', '그럼에도 ', '불구하고 ', '여전히 ', '심지어 ', '까지도 ', '조차도 ', '하지 ',
 '않도록 ', '않기', '위하여', '때 ', '시각 ', '무렵 ', '시간 ', '동안', '어때 ', '어떠한
 ', '하여금 ', '네 ', '에 ', '우선 ', '누구 ', '누가 ', '알겠는가 ', '아무도 ', '줄은모른
 다 ', '줄은 ', '몰랐다 ', '하는 ', '김에 ', '검사겸사', '하는바 ', '그런 ', '까닭에',
 '한 ', '이유는', '그러니 ', '그러니까', '때문에', '그', '너희 ', '그들 ', '너희들', '타
 인', '것 ', '것들 ', '너', '위하여 ', '공동으로 ', '동시에 ', '하기 ', '위하여', '어찌하
 여 ', '무엇때문에 ', '붕붕', '윙윙', '나', '영영 ', '휘익 ', '윙윙 ', '오호 ', '아하 ',
 '어쨌든 ', '만 ', '못하다 ', '하기보다는 ', '차라리 ', '하는 ', '편이 ', '낫다 ', '흐흐
 ', '놀라다 ', '상대적으로 ', '말하자면 ', '마치 ', '아니라면 ', '쉴 ', '그렇지 ', '않으면
 ', '그렇지 ', '않다면 ', '안 ', '그러면 ', '아니었다면 ', '하든지 ', '아니면 ', '이라면
 ', '좋아 ', '알았어 ', '하는것도 ', '그만이다 ', '어쩔수 ', '없다 ', '하나 ', '일', '일
 반적으로 ', '일단 ', '한편으로는 ', '오자마자 ', '이렇게되면 ', '이와같다면 ', '전부 ',
 '한마디', '한항목 ', '근거로 ', '하기에 ', '아울러 ', '하지 ', '않도록 ', '않기 ', '위해
 서 ', '이르기까지 ', '이 ', '되다 ', '로 ', '인하여 ', '까닭으로 ', '이유만으로 ', '이로
 ', '인하여 ', '그래서 ', '이 ', '때문에 ', '그러므로 ', '그런 ', '까닭에 ', '알 수 있다
 ', '결론을 ', '낼 수 있다 ', '으로 ', '인하여 ', '있다 ', '어떤것 ', '관계가 있다', '관련
 이 있다', '연관되다 ', '어떤것들 ', '에 ', '대해 ', '이리하여 ', '그리하여 ', '여부 ',
 '하기보다는 ', '하느니 ', '하면', '할수록 ', '운운 ', '이러이러하다 ', '하구나 ', '하도다
 ', '다시말하면 ', '다음으로 ', '에 ', '있다 ', '에 ', '달려 ', '있다 ', '우리들 ', '오히
 려 ', '하기는한데 ', '어떻게 ', '어떻해 ', '어찌됐어 ', '어때 ', '어째서 ', '본대로 ',
 '자 ', '이', '이쪽 ', '여기 ', '이것 ', '이번 ', '이렇게말하자면 ', '이런', '이러한 ',
 '이와 ', '같은 ', '요만큼 ', '요만한 ', '것 ', '얼마 ', '안 ', '되는 ', '것 ', '이만큼
 ', '이 ', '정도의', '이렇게 ', '많은 ', '것 ', '이와 ', '같다 ', '이때 ', '이렇구나 ',
 '것과 ', '같이 ', '끼익 ', '빠걱', '따위', '와', '같은', '사람들', '부류의 ', '사람들 ',
 '왜냐하면 ', '중의하나 ', '오직 ', '오로지 ', '에 ', '한하다 ', '하기만 ', '하면 ', '도착
 하다', '까지', '미치다 ', '도달하다 ', '정도에 ', '이르다 ', '할 ', '지경이다 ', '결과에
 ', '이르다 ', '관해서는 ', '여러분 ', '하고 ', '있다 ', '한 ', '후', '혼자 ', '자기 ',
 '자기집', '자신 ', '우에 ', '종합한것과같이', '총적으로 ', '보면 ', '총적으로 ', '말하면
 ', '총적으로 ', '대로 ', '하다 ', '으로서 ', '참 ', '그만이다 ', '할 ', '따름이다 ', '쿵
 ', '탕탕', '광광 ', '둥둥 ', '봐', '봐라', '아이야 ', '아니 ', '와아 ', '응 ', '아이 ',
 '참나 ', '년 ', '월 ', '일 ', '령 ', '영 ', '일 ', '이 ', '삼 ', '사 ', '오 ', '육 ',
 '륙 ', '칠 ', '팔 ', '구 ', '이천육 ', '이천칠 ', '이천팔 ', '이천구 ', '하나 ', '둘 ',
 '셋', '넷', '다섯 ', '여섯 ', '일곱 ', '여덟 ', '아홉 ', '령 ', '영 ', '이 ', '있 ',
 '하 ', '것 ', '들 ', '그 ', '되 ', '수 ', '이 ', '보 ', '않 ', '없 ', '나 ', '사람',
 '주', '아니 ', '등 ', '갈 ', '때 ', '년 ', '가 ', '한 ', '지 ', '대하 ', '오 ', '말 ',
 '일 ', '그렇 ', '위하 ', '때문', '그것 ', '두 ', '말하 ', '알 ', '그러나 ', '반 ', '못하
 ', '일 ', '그런 ', '또 ', '문제 ', '더 ', '사회 ', '많 ', '그리고 ', '좋 ', '크', '따르
 ', '중 ', '나오 ', '가지 ', '씨 ', '시키 ', '만들 ', '지금 ', '생각하 ', '그러 ', '속 ',
 '하나 ', '집 ', '살 ', '모르 ', '적 ', '월 ', '데 ', '자신', '안 ', '어떤 ', '내 ', '내
 ', '경우 ', '명 ', '생각', '시간 ', '그녀', '다시 ', '이런 ', '앞 ', '보이 ', '번 ', '나
 ', '다른 ', '어떻 ', '여자 ', '개 ', '전 ', '들 ', '사실 ', '이렇 ', '점 ', '싶 ', '말
 ', '정도 ', '좀 ', '원 ', '잘 ', '통하 ', '놓']

In [36]:

데이터 로드

nflx_news = pd.read_csv('./nflx0415_0430.csv')

In [37]:

데이터 전처리

```
nflx_wc = nflx_news.drop(['뉴스 식별자', '일자', '언론사', '기고자',
'통합 분류1', '통합 분류2', '통합 분류3', '사건/사고 분류1', '사건/사
고 분류2', '사건/사고 분류3', '인물', '위치', '기관', '특성추출(가중치
순 상위 50개)', '본문', 'URL', '분석제외 여부'], axis = 1)
```

In [38]:

nflx_wc

Out[38]:

	제목	키워드
0	지난 백 번의 특별사면은 정말 국민을 통합했을 까[그때 그 뉴스]	특별사면,국민,통합,국회,동의,일반,사면,특별사면,정 권,임기,시국,사범,특사,집중,...
1	과거 현재 미래 인문학백화점 '민속이란 삶이 다'展	과거,미래,인문학,백화점,민속,서울,삼청동,동쪽,경북 궁,국립민속박물관,민속,특별전,...
2	[OTT다방] 회의하다 모유 나온 男, 임신 고백하고 맘카페 만든 사연	회의,모유,임신,고백,맘카페,사연,리뷰,켄타로,넷플릭 스,히야마,임신,추천,작품,세상...
3	여자들의 뜨거운 비명을 들어라 [김봉석 윤이나 의 정기구독]	여자들,비명,김봉석,윤이나,정기구독,애플,TV,플러스, 로어,세상,함성,소설,앤솔러지...
4	[임병선의 메멘토 모리] 마릴린 먼로의 죽음 보도 된 것과 달랐다	마릴린,먼로,죽음,보도,미국,여배우,세상,명성,미모,명 예,주검,발견,5일,현지시간,...
...
894	삼성전자만? 아마존 폐북도 '장투의 배신'	삼성전자,아마존,폐북도,장투,배신,1~2년,서학개미 들,수익,저조,한숨,삼성전자,네이...
895	"차기작은 상어 멸치 김 등 수산물 시리즈"	차기작,상어,멸치,수산물,시리즈,저주토끼,부커,최종, 후보작,저주,토끼,작가,정보,번...
896	배우 전여빈 넷플릭스 시리즈 '너의 시간 속으로' 캐스팅	배우,전여,넷플릭스,시리즈,캐스팅,강릉,출신,배우,전 여빈,사진,넷플릭스,시리즈,캐스...
897	박찬욱 고레에다 연출 한국영화 2편 칸 경쟁부문 진출...이정재도 감독으로 초청	박찬욱,고레,2편,연출,한국,영화,경쟁,진출,이정재,감 독,초청,박찬욱,감독,일본,고...
898	넷플릭스, 2021년 한국 매출 6317억원 2020년보 다 52% 증가	넷플릭스,한국,매출,2020년,52%,증가,넷플릭스,한국 법,넷플릭스서비스스코리아,감...

899 rows × 2 columns

In [39]:

제목 불용어처리 후 컬럼 추가

```
# nflx_wc의 df에서 '제목' 열의 모든 데이터를 공백문자를 이용해 하나의  
문자열로 만듦
```

```
sentence = ' '.join(nflx_wc['제목'])
```

불용어 처리할 리스트 호출

```
stop_words = set(stopwords.words('korean'))
```

이어붙인 문자열 토큰화

```
words = word_tokenize(sentence)

# 분리된 단어들 중 불용어에 해당하는 단어를 제외하고 필터링
title_filter_words = [word for word in words if not word in
stop_words]

# nflx_kw 컬럼에 추가하기
nflx_wc['title filter'] = ' '.join(title_filter_words)
```

In [21]:

nflx_wc

Out[21]:

	제목	키워드	title_filter
0	지난 백 번의 특별사면은 정말 국민을 통합했을까[그때 그 뉴스]	특별사면,국민,통합,국회,동의,일반,사면,특별사면,정권,임기,시국,사범,특사,집중,...	지난 백 번의 특별사면은 정말 국민을 통합했을까 그때 뉴스 과거 현재 미래 인문학백...
1	과거 현재 미래 인문학백화점 '민속이란 삶이다'展	과거,미래,인문학,백화점,민속,서울,삼청동,동쪽,경복궁,국립민속박물관,민속,특별전,...	지난 백 번의 특별사면은 정말 국민을 통합했을까 그때 뉴스 과거 현재 미래 인문학백...
2	[OTT다방] 회의하다 모유 나온 男, 임신 고백하고 맘카페 만든 사연	회의,모유,임신,고백,맘카페,사연,리뷰,켄타로,넷플릭스,히야마,임신,추천,작품,세상...	지난 백 번의 특별사면은 정말 국민을 통합했을까 그때 뉴스 과거 현재 미래 인문학백...
3	여자들의 뜨거운 비명을 들어라 [김봉석 윤이나의 정기구독]	여자들,비명,김봉석,윤이나,정기구독,애플,TV,플러스,로어,세상,함성,소설,앤솔러지...	지난 백 번의 특별사면은 정말 국민을 통합했을까 그때 뉴스 과거 현재 미래 인문학백...
4	[임병선의 메멘토 모리] 마릴린 먼로의 죽음 보도된 것과 달랐다	마릴린,먼로,죽음,보도,미국,여배우,세상,명성,미모,명예,주검,발견,5일,현지시간,...	지난 백 번의 특별사면은 정말 국민을 통합했을까 그때 뉴스 과거 현재 미래 인문학백...
...
894	삼성전자만? 아마존 폐북도 '장투의 배신'	삼성전자,아마존,폐북도,장투,배신,1~2년,서학개미들,수익,저조,한숨,삼성전자,네이...	지난 백 번의 특별사면은 정말 국민을 통합했을까 그때 뉴스 과거 현재 미래 인문학백...
895	"차기작은 상어 멸치 김 등 수산물 시리즈"	차기작,상어,멸치,수산물,시리즈,저주토끼,부커,최종,후보작,저주,토끼,작가,정보,번...	지난 백 번의 특별사면은 정말 국민을 통합했을까 그때 뉴스 과거 현재 미래 인문학백...
896	배우 전여빈 넷플릭스 시리즈 '너의 시간 속으로' 캐스팅	배우,전여,넷플릭스,시리즈,캐스팅,강릉,출신,배우,전여빈,사진,넷플릭스,시리즈,캐스...	지난 백 번의 특별사면은 정말 국민을 통합했을까 그때 뉴스 과거 현재 미래 인문학백...
897	박찬욱 고레에다 연출 한국영화 2편 칸 경쟁부문 진출...이정재도 감독으로 초청	박찬욱,고레,2편,연출,한국,영화,경쟁,진출,이정재,감독,초청,박찬욱,감독,일본,고...	지난 백 번의 특별사면은 정말 국민을 통합했을까 그때 뉴스 과거 현재 미래 인문학백...
898	넷플릭스, 2021년 한국 매출 6317억원 2020년보다 52% 증가	넷플릭스,한국,매출,2020년,52%,증가,넷플릭스,한국법,넷플릭스서비시스코리아,감...	지난 백 번의 특별사면은 정말 국민을 통합했을까 그때 뉴스 과거 현재 미래 인문학백...

899 rows × 3 columns

In [48]:

print(os.listdir())

```
['.ipynb_checkpoints', '09_crawling_google_news.ipynb', 'chromedriver.exe', 'CSV파일', 'Kaggle_Stock_Data.py', 'mask_image.png', 'netflix_news.csv', 'news_crawling.p
```

In [50]:

In [32]:

localhost:8888/nbconvert/html/TIL/시각화프로젝트/NFLX0415~0430 WordCloud.ipynb?download=false

```
# freq_dist = FreqDist(all_titles.split())

# # 단어 추출
# word = freq_dist.most_common()

# # 워드클라우드 생성
# font_path='C:/Windows/Fonts/malgun.ttf'
# wordcloud =
WordCloud(background_color='white',font_path=font_path).generate_fro

# # 그래프 출력
# plt.imshow(wordcloud, interpolation='bilinear')
# plt.axis('off')
# plt.show()
```

In [33]:

```
# icon = Image.open('./mask_image-removebg.png')
# plt.imshow(icon)

# mask = Image.new("RGB", icon.size, (255, 255, 255))
# mask.paste(icon,icon)
# mask = np.array(mask)
```