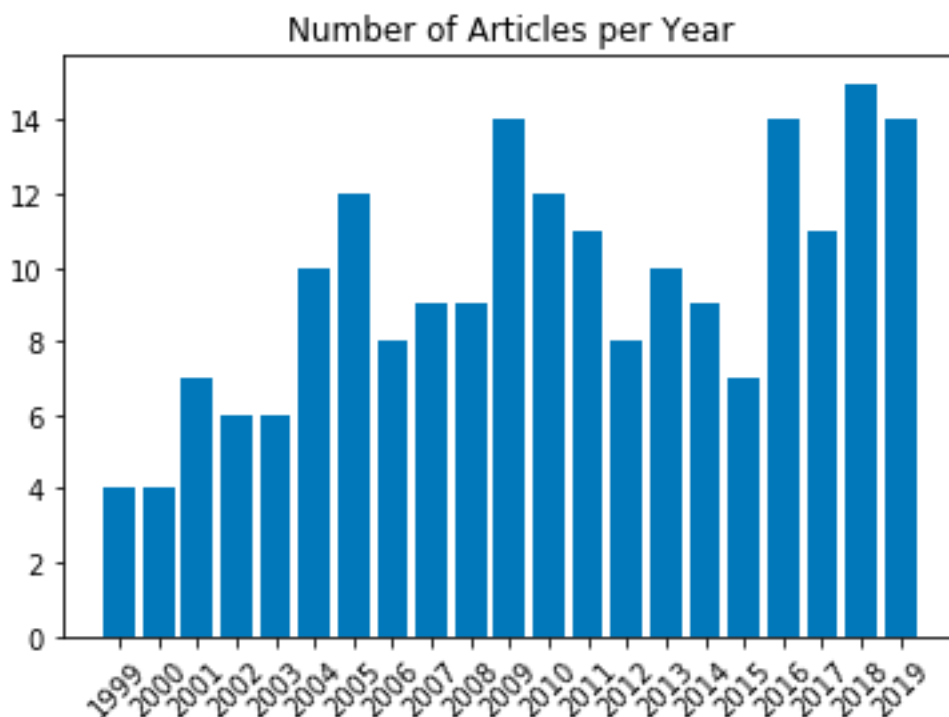


1. 대상 데이터 수집 및 정제

- RISS에서 제공하는 ‘한국음악치료학회’에 게재된 논문을 수집한다.
- 수집된 논문은 1999년부터 2019년까지 ‘한국음악치료학회’에 게재된 논문으로 수집 결과 총 239건을 데이터로 얻을 수 있었다. (data_raw.csv)
- 수집한 정보는 제목, 저자, 학회명, 학회지명, 권호사항, 년도, 언어, 키워드, KDC 및 KCI 정보, 페이지 정보, 링크 정보, 초록이며 이중 제목, 저자, 년도, 권호 사항, 키워드, 초록을 따로 추출해서 별도의 데이터로 만들었다. (data_essential.csv)
- 키워드 및 초록 정보는 비어있는 정보가 있으므로 이를 ‘한국음악치료학회’ 웹 사이트에서 직접 논문을 찾아 수작업으로 채워넣었다. (data_refined.csv)
- 중복된 논문, 영어 논문, 초록이 없는 북 리뷰 및 학회 규정 사항은 분석 대상에서 제외하였으며, 총 200건이 최종 분석 대상이다. 제외된 논문과 그 근거를 별도의 파일로 작성하였다. (reason.csv)
- 1999년부터 2003년까지는 키워드 정보가 없다

2. 대상 데이터 탐색

- 저자 별 논문 빈도 수를 계산할때 저자는 통일되지 않은 문자 양식(황은영, 황은영Hwang, 황은영HwangEunYoung)으로 인해 처리가 어렵다. (수행하지 않음)
- 연도 별 논문 빈도 수를 계산한 결과를 보면 2009년을 기준으로 논문 수가 떨어지는 것을 확인할 수 있다. 이후 2016년에 전년 대비 급증한다. (frequency_year.csv)

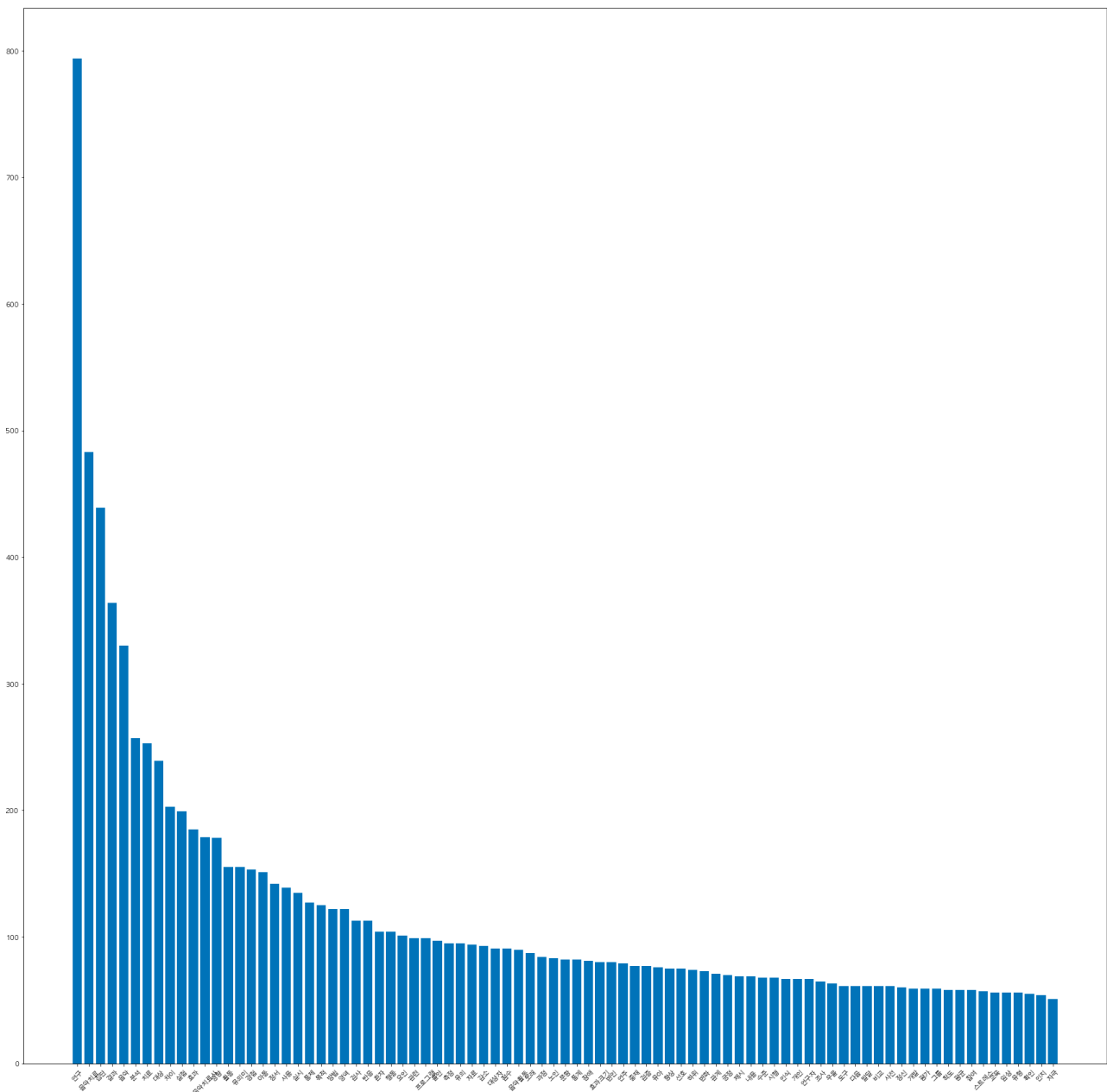


- 연도 별 논문 빈도 수를 계산한 결과를 보면 2009년을 기준으로 논문 수가 떨어지는 것을 확인할 수 있다. 이후 2016년에 전년 대비 급증한다. (frequency_year.csv)

- 사용된 키워드에 대한 빈도 분석을 수행한다. 이때 MV 키워드는 키워드 없음을 나타내는 약속어이다. 총 908 개의 키워드가 존재한다. 전체 결과를 파일로 저장한다. (frequency_keyword.csv)
- 한글 분석이기 때문에 영어로된 키워드를 지우고 MV 키워드를 지운다. 해당 키워드는 형태소 분석 시에 필요한 사전 구축에 사용한다. 빈도수가 1인 키워드까지 포함한 결과를 frequency_keyword_full.csv로 저장하고 빈도수가 1인 키워드를 제외한 결과를 frequency_keyword_partial.csv로 저장한다.
- 연도별 키워드의 변화를 살펴본다. (keyword_year.csv). 1999년부터 2003년까지는 키워드가 존재하지 않기 때문에 분석에 있어서 제외해야한다. 가장 주목할만한 점은 음악치료사에 대한 키워드가 이전에 비해서 증가한 것을 확인할 수 있다. 또한 메타분석 역시 2010년대 이후 꾸준히 발표된다.

3. 대상 데이터 탐색

- 초록의 단어를 형태소 분석하여서 명사만 추출한 결과로 얻은 단어를 통해 논문 주제의 변화를 살펴본다.
- 형태소 분석을 위해서 Komoran을 사용하였으며 이는 사용자 정의 사전을 추가할 수 있기에 용이하다. 사용자 정의 사전으로는 키워드를 하나의 일반명사(NNG)로 인식하도록 설정하였다.
- 초록에 사용된 명사 단어에 대한 빈도를 저장한다. (frequency_abstract_full.csv)



- 단어 분석에 있어서 일반적으로 한 글자 단어는 잘 사용하지 않으며 일정 빈도 이상(임의로 50으로 설정)인 단어들을 추출한다. (frequency_abstract_partial.csv)
- 연도별로 초록에 사용된 단어에 변화를 살펴본다. 여기서 대상이 되는 단어는 단어 길이가 2이상이고 빈도수가 50이상인 단어이다. (abstract_year_word.csv)
- 연도별로 초록에 사용된 단어는 문헌의 개수 또는 문헌의 길이에 따라 달라질 수 있다. 즉, 특정 문헌에 집중적으로 많이 등장한 단어로 인해 전체적인 빈도수가 증가할 수 있으므로 이를 보정하기 위해서 해당 단어가 사용된 논문의 수를 연도별로 같이 파악한다. (abstract_year_article.csv)
- 공기 빈도를 파악함으로써 어떠한 단어가 같은 초록에서 사용되었는지를 파악한다. 여기서 단어의 기준은 길이가 2이상이며 빈도수가 50이상인 명사이다. (coword_matrix.csv)

4. 토픽 모델링

- 수행된 결과에 대한 해석이 중요하다. 특히 사항은 보이지 않는다.