

# 웹 크롤링을 이용한 최근 한국교회 설교의 성경 본문 및 제목 분석

강경우<sup>1)</sup> 백석대 교수

권오병<sup>2)</sup> 경희대 교수

## ABSTRACT

As information sharing through the Internet becomes more active, technologies for detecting social phenomena and changes based on information disclosed on the Internet are increasing. In particular, we are detecting the direction of society by using web crawling technology that collects information on the web. One way to understand the direction of Christian society is to use sermon information from church services. In this study, information on Sunday sermons from Korean churches recently released on YouTube was obtained through web crawling, and word cloud analysis was performed. We investigated which book of the Bible the sermon is based on and what keywords appear most often in the sermon title.

Key Words: Web crawling, Sunday sermon, Bible, Sermon title

## I. 서론

인터넷이 보편화 되면서 많은 교회가 홈페이지를 운영하고 이를 통해 성도들이 온라인상에서 소통을 하고 있다. 이와 같은 현상은 2020년 코로나19의 확산과 함께 더욱 가속화 되었다. 코로나 확산 방지를 위해 모든 예배가 성도들 간 비대면 상황에서 운영되면서 각 교회는 홈페이지를 통해 실시간으로 인터넷을 통해 예배 실황을 송출하거나 미리 녹화된 예배 영상을 인터넷상에 공개하였다. 홈페이지를 운영하지 않았거나 존재하더라도 기술적인 문제로 영상 송출이 어려운 교회들은 영상공유 플랫폼인 유튜브 사이트를 통해 예배 영상을 공개하였다. 유튜브를 통해

1) 백석대학교(Baekseok University) 교수, 충청남도 천안시 동남구 백석대학로 1, kwkang@bu.ac.kr

2) 경희대학교(Kyunghee University) 교수, 서울특별시 동대문구 경희대로 26, obkwon@khu.ac.kr

설교영상이 공개되면서 사람들은 출석하는 교회의 설교뿐만 아니라 타교회의 설교까지 볼 수 있는 기회가 생겼다. 이런 현상이 지속되면서 지금은 대부분 교회가 유튜브 상에 채널을 운영하고 유튜브 상에 설교영상을 정기적으로 게시하고 있다. 설교영상은 교회 성도들의 신앙성장에도 도움이 될 뿐 아니라 타교회 교인과 기독교신앙에 관심이 있는 일반인들에게도 도움이 되고 있다.(윤성민, 손동식, 2020)

인터넷에 수많은 콘텐츠가 올라오면서 사람들은 콘텐츠에 대한 정보를 가공하여 원하는 정보를 얻고 있다. 인터넷상에 웹 콘텐츠를 분석하여 정보를 추출하는 기술을 웹 크롤링이라고 부른다. 웹 크롤링은 인터넷을 탐색하고 공개된 웹 페이지를 다운로드 하여 의미 있는 정보를 얻는 기술이다. 이러한 기술을 이용하여 정보를 추출하면 사회의 방향성을 탐지할 수 있다. 최근 들어 서로 다른 동기를 가지고 유튜브나 페이스북 같은 SNS에 공개되는 수많은 콘텐츠를 대상으로 웹 크롤링 기술을 적용하는 사례가 많아지고 있다. 특히 유튜브를 통해 공개되는 수많은 콘텐츠는 웹 크롤링을 통해 정보를 모을 수 있는 중요한 자원이 되고 있다. 유튜브 웹 크롤링의 주요 목표는 관심 있는 동영상에서 특정 데이터 요소를 추출하는 것이다. 이 데이터에는 동영상 제목, 설명, URL, 조회 수, 좋아요, 싫어요, 댓글, 사용자 프로필, 채널 정보 등이 포함될 수 있다. 유튜브는 개발자가 구조화되고 승인된 방식으로 특정 데이터에 액세스할 수 있도록 API(응용 프로그램 프로그래밍 인터페이스)를 제공하는데 다운로드한 HTML을 구문 분석하고 웹 페이지 구조를 처리하는 기존의 웹 스크래핑 대신 API를 사용하면 유튜브에서 데이터를 검색하는 보다 표준화되고 안정적인 방법을 제공하고 있다.(이종화, 2018)

본 연구에서는 웹 크롤링 기술을 이용하여 유튜브에 최근 공개된 한국교회 설교에 대한 정보를 얻어오는 연구를 수행하였다.(이경배, 2019) 유튜브에 공개되는 설교영상에는 교회마다 다르지만 설교 날짜, 설교제목, 성경 본문, 설교자 등의 정보가 함께 게시되는 것이 일반적이다. 하지만 이와 같은 정보를 기술하는 표준이 존재하지 않기 때문에 교회마다 기술하는 방식이 서로 다르다. 본 연구에서는 API를 이용하여 유튜브에 공개된 설교영상 정보를 얻어온 후 날짜정보, 설교제목, 설교자, 성경 본문, 조회 수 정보를 추출하였다. 또한 이 정보를 이용하여 최근 한국교회 주일 설교에서 사용된 성경 본문이 어떻게 분포되는지 조사하고 설교제목의 키워드 분석을 수행하였다. 본 연구에서 포함시키지는 않았지만 이와 같은 분석 결과는 한국 교회 설교 경향을 연구하는데 사용될 수 있을 것으로 기대된다.

## II. 관련 연구

### 1. 유튜브 크롤링 기술

유튜브 크롤링 기술을 활용하면 유튜브 플랫폼에서 공개되어 있는 동영상, 채널, 사용자 상호

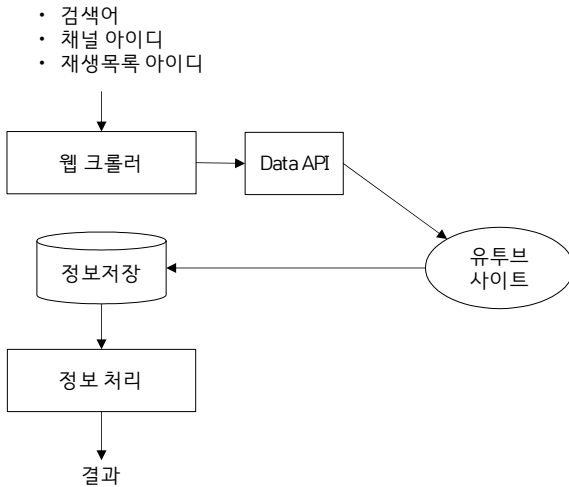
작용 및 기타 관련 정보와 관련된 다양한 데이터 수집이 가능하다. 유튜브 크롤링을 통해 얻을 수 있는 정보를 살펴보면 다음과 같다.

- ◆ 동영상 정보: 개별 동영상에 대한 자세한 정보를 수집할 수 있다. 여기에는 비디오 제목, 설명, 게시 날짜, 기간, 조회 수, 좋아요 수, 싫어요 수, 댓글 및 비디오 태그와 같은 데이터가 포함된다.
- ◆ 채널 세부 정보: 유튜브 크롤링은 유튜브 채널에 대한 데이터를 얻을 수 있다. 여기에는 채널 이름, 채널에 대한 설명, 구독자 수, 총 동영상 수 및 기타 관련 메타데이터가 포함된다.
- ◆ 사용자 상호 작용: 만일 유튜브 댓글을 크롤링하면 비디오에 대한 사용자 상호 작용을 수집할 수 있다. 여기에는 시청자가 게시한 댓글, 댓글에 대한 좋아요 및 싫어요, 댓글에 대한 답변이 포함된다.
- ◆ 비디오 추천: 크롤러는 유튜브의 추천 알고리즘을 분석하여 주어진 비디오 또는 채널에 대한 비디오 추천을 검색할 수 있다. 이 데이터는 사용자 행동 및 콘텐츠 검색 가능성을 이해하는 데 사용할 수 있다.
- ◆ 인기 동영상: 크롤링 유튜브의 인기 섹션을 사용하면 다양한 지역 및 카테고리에서 가장 인기 있고 유행하는 동영상에 대한 데이터를 추출할 수 있다.
- ◆ 비디오 메타데이터 및 통계: 크롤링 유튜브 API는 비디오 해상도, 프레임 속도, 캡션 가용성 및 비디오 처리 상태를 포함하여 비디오 메타데이터 및 통계에 대한 데이터를 제공할 수 있다.
- ◆ 사용자 활동: 유튜브 크롤링을 사용하면 좋아하는 동영상, 재생 목록 및 구독과 같은 사용자 활동에 대한 데이터를 수집할 수 있다. 이 정보는 사용자 기본 설정 및 동작을 이해하는 데 유용할 수 있다.
- ◆ 동영상 카테고리 및 태그: 크롤러는 특정 동영상과 관련된 동영상 카테고리 및 태그에 대한 데이터를 추출하여 콘텐츠를 분류하고 범주화하는 데 도움을 줄 수 있다.
- ◆ 비디오 대본: 크롤링을 통해 비디오 대본을 추출하면 콘텐츠 작성자가 감정 분석 및 키워드 추출과 같은 텍스트 기반 분석을 수행할 수 있다.
- ◆ 콘텐츠 분석 및 트렌드: 유튜브 크롤링을 통해 수집된 데이터를 분석하면 콘텐츠 트렌드, 사용자 참여 패턴, 인기 있는 주제에 대한 통찰력을 얻을 수 있다.
- ◆ 감정 분석: 특정 동영상이나 채널에 대한 시청자의 전반적인 인식과 피드백을 이해하기 위해 감정에 대해 유튜브 댓글을 분석할 수 있다.
- ◆ 콘텐츠 추천 시스템: 사용자 상호 작용 및 비디오 추천에 대한 크롤링 데이터를 사용하여 개인화된 사용자 경험을 위한 콘텐츠 추천 시스템을 개선할 수 있다.

## 2. 유튜브 API를 활용한 웹 크롤링

[그림1]은 유튜브 API를 활용하여 크롤링을 하는 일반적인 방법이다. 그림에 나와 있지 않지만 웹 크롤러가 Data API를 이용하기 위해서는 먼저 구글에서 인증키를 받아야 한다. 유튜브에는 데이터 크롤링 및 API 사용에 관한 정책 및 지침이 있고 개발자는 유튜브 크롤링을 수행하는 동안 이러한 규칙을 지켜야 한다. 웹 크롤러에게 정보를 제공하면 API를 통해 유튜브 사이트에 데이터를 요청할 수 있다. 요청하는 방법은 다양하다.

- 검색어 기반: 검색어를 이용하여 관련 비디오 정보를 얻어 올 수 있다. 제시된 검색어를 기반으로 제공하기 때문에 제공된 비디오 정보가 사용자가 찾는 것일 수도 있고 아닐 수도 있다.
- 채널 아이디 기반: 채널의 아이디를 제공하면 채널에서 제공하는 비디오와 재생 목록의 정보를 얻어 올 수 있다. 재생 목록은 비슷한 종류의 비디오를 모아놓은 저장소로 채널관리자가 만들고 관리한다.
- 재생 목록 아이디 기반: 재생 목록 아이디는 재생 목록을 구분하기 위한 구분자로 이것을 이용하면 재생 목록 내에 속해있는 비디오 정보를 얻어 올 수 있다.



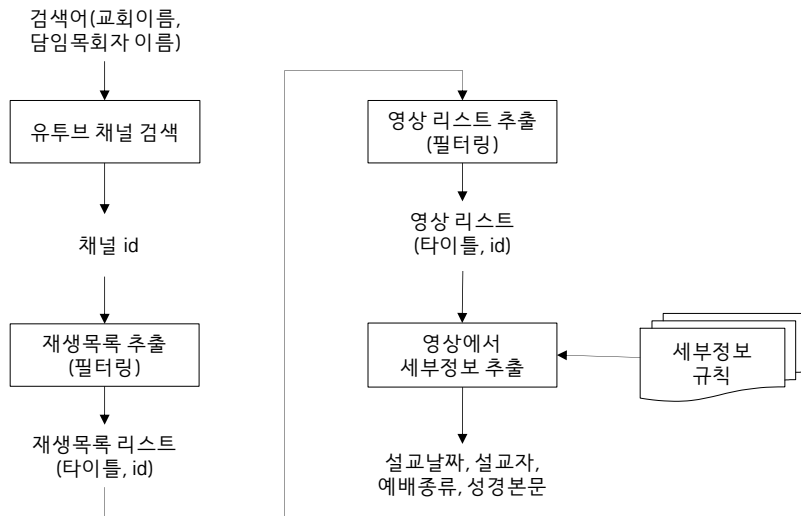
[그림1] 유튜브 웹 크롤링 방법

웹 크롤러가 얻어온 정보는 json형태이고 라이브러리를 이용하여 리스트 형태로 변환 가능하다. 비디오에 대한 정보는 동영상 제목, 설명, URL, 조회 수, 좋아요, 싫어요, 댓글, 사용자 프로필, 채널 정보 등이 있다. 동영상 제목과 설명이 해당 동영상에 대한 자세한 서술형 정보 형태인데 작성 방법은 동영상 게시자에 따라 매우 다양하다. 따라서 원하는 정보를 얻기 위해서는 정보처리 파트를 두어 저장된 정보로부터 원하는 데이터를 얻어낼 수 있어야 한다.

### III. 한국교회 설교 성경구절 크롤링

인터넷 속도가 증가하면서 많은 교회들이 자체적으로 또는 유튜브 채널을 통해 매주 예배 영상을 업로드하고 있다. 이를 통해 교인들은 예배에 직접 참석하지 않아도 온라인을 통해 예배 영상을 시청함으로써 예배에 참여하고 있다. 특히 코로나 바이러스 팬데믹 기간 동안 이런 현상은 더욱 두드러졌다. 한국교회의 예배 순서에는 크게 찬양, 기도, 설교로 이루어지지만 성도들이 느끼는 설교의 비중은 매우 크다. 이 때문에 대부분의 교회는 전체예배 영상 보다는 설교부분만 따로 업로드 하는 것이 일반적이다. 본 연구에서는 웹 크롤링을 통해 유튜브에 최근에 올라간 한국의 주요교회 설교영상에 관한 정보를 분석하는 연구를 진행하였다.

한국교회 목회자들은 설교를 준비함에 있어 성경 구절과 설교 주제를 정하고 이를 기반으로 하여 설교 내용을 준비한다. 이처럼 성경을 기초로 하여 설교를 준비하는 것은 성경이 기독교의 중요한 교리적 기초이며, 대부분 성도들이 성경을 중요한 권위 있는 말씀으로 인식하고 있기 때문이다. 따라서 한국 교회의 설교에서는 주로 성경 구절을 인용하고 성경의 내용을 설명하며 성경의 말씀에 따라 일상생활에 적용하는 방식으로 진행된다. 이러한 방식으로 성경을 사용하고 강조하는 것은 한국교회의 핵심 가치 중 하나이기도 하다. 본 연구에서는 유튜브 API를 이용하여 최근 한국교회의 주일 설교 성경구절과 설교주제를 추출해 보았다.(반재훈, 하중수, 김동현, 2020)



[그림2] 설교정보 추출 방법

각 교회에서 운영하는 유튜브 채널을 교회이름만 가지고 검색하는 데는 한계가 존재한다. 왜냐하면 동일한 이름의 교회가 전국에 많이 존재하기 때문이다. 따라서 정확한 검색을 위해 사용자는 검색어로 교회의 이름과 담임목회자 이름을 이용한다. 그러면 찾고자 하는 교회의

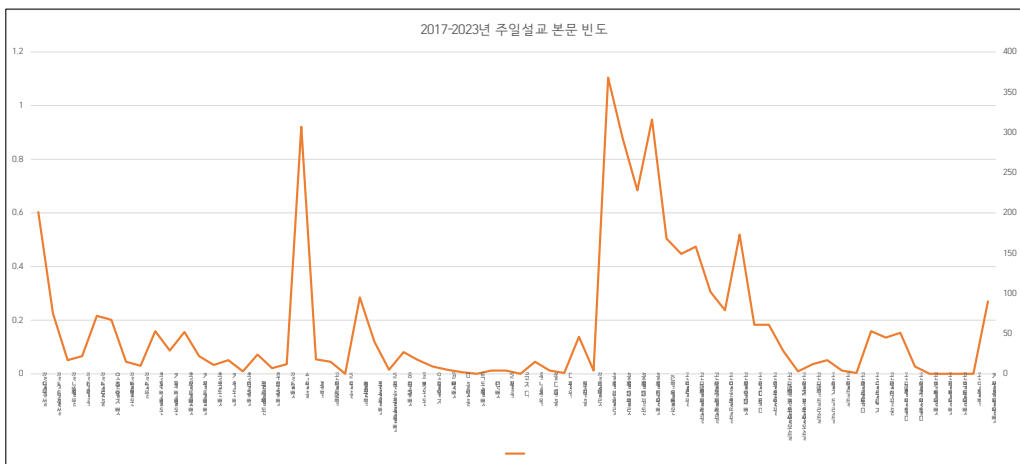
채널을 검색할 수 있다. 채널을 찾은 이후 교회에서 제공하는 재생 목록을 얻어온다. 재생 목록을 얻어올 때 목록 타이틀의 내용을 분석해야 한다. 여기서는 설교영상만 얻어올 것이기 때문에 재생목록 타이틀명에서 ‘설교’ 또는 ‘주일예배’와 같은 단어를 찾는다. 이와 같이 설교영상을 모아놓은 재생 목록을 추출하고 그 다음으로 영상리스트를 추출한다.

영상리스트 중에서 실제 설교영상인지 확인하고 맞는다면 세부정보를 불러온다. 이와 같이 얻어온 정보에서 설교날짜, 설교자, 예배종류, 성경 본문, 설교제목 정보를 추출한다. 이와 같은 정보가 영상 타이틀에 있을 수도 있고 세부정보에 포함되어 있을 수도 있다. 또한, 정보를 표현하는 방법에 표준이 있는 것이 아니기 때문에 각 교회마다 표기하는 방법이 다르다. 따라서 정보를 추출하기 위해서는 추출하고자 하는 정보의 규칙을 정해 주어야 한다. 본 연구에서는 14개 대형교회에서 게시한 유튜브 설교영상 정보를 바탕으로 규칙을 정의하였다.

#### IV. 실험 결과

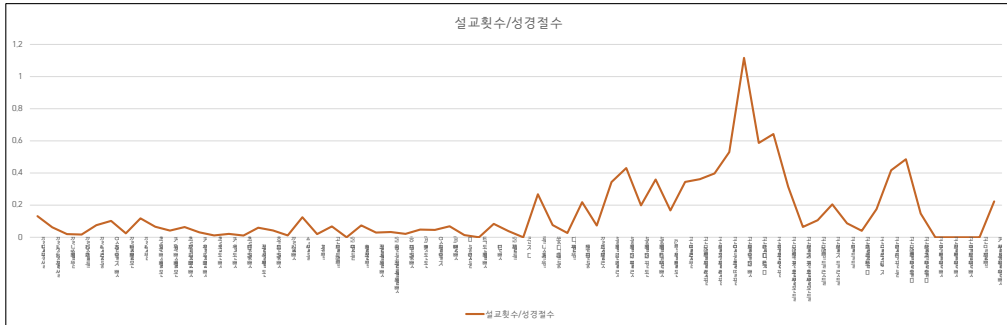
본 연구에서는 유튜브에 올라온 국내 14개 대형교회 주일 설교 정보를 크롤링 하였다. 각 교회가 유튜브에 설교를 정기적으로 올린 기간이 있고 그렇지 않은 기간이 있으며 영상만 올리고 설교의 성경 본문 정보를 올리지 않는 경우도 있기에 크롤링한 정보가 14개 교회의 모든 주일 설교를 대상으로 하지는 않는다. 또한 2017년 이전에는 일부교회만 설교영상을 제공하고 있기 때문에 크롤링 기간을 2017년 이후로 제한하였다.

[그림3]은 주일설교에서 사용한 성경 본문에 대해서 조사한 그래프이다. 그래프는 각 성경 권별 빈도수를 의미한다. 그림에서 보는 바와 같이 설교본문으로 복음서, 시편, 창세기를 많이 사용하고 있음을 알 수 있고 오바다, 나훔, 요한서신서, 유다서 등을 본문으로 하는 주일설교는 찾기가 어려웠다.



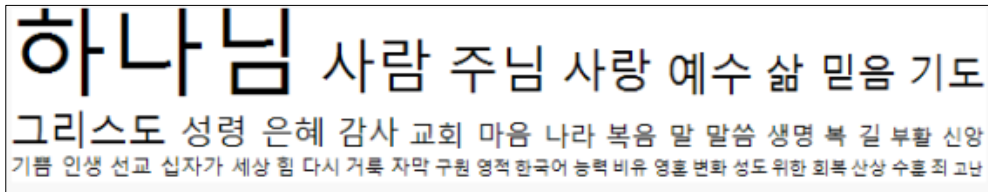
[그림3] 2017-2023년 주일설교에서 사용한 성경 권별 절대 빈도수

[그림4]에서는 성경의 권별 크기에 비례하여 상대 빈도수를 나타내 보았다. 이것은 성경 각 권의 크기가 서로 다르기 때문에 절대 빈도수보다 상대 빈도수가 경향성을 정확히 표현할 것이라고 판단했기 때문이다. 성경의 권별 크기는 성경구절수를 기준으로 하였다. 권별 크기를 감안할 때 에베소서를 본문으로 한 주일 설교 횟수가 가장 많고 신약성경을 본문으로 하는 설교 횟수가 구약성경을 본문으로 했을 때보다 많음을 알 수 있었다.



[그림4] 2017-2023년 주일설교 본문말씀 상대 빈도수

[그림5]에서는 2017년 이후 설교제목에 나타난 주요 키워드를 워드클라우드로 나타낸 결과이다. 그림에서 글자의 크기에 따라 설교제목에 더 많이 나타났다는 뜻이다.



[그림5] 2017-2023년 주일설교제목에 나타난 주요 키워드

[그림6]과 [그림7]에서는 2017년부터 현재까지 기간을 코로나 팬데믹 이전과 이후로 나누어 설교제목 키워드에 대한 분석한 결과를 나타낸다. 코로나 팬데믹 시작 시기는 우리나라를 기준으로 한 것으로 2020년 2월1일을 잡았다.



[그림6] 2017-2020년 1월까지 주일설교제목에 나타난 주요 키워드

**하나님** 사람 주님 예수님 삶 믿음  
 사랑 그리스도 성령 은혜 교회 기도 마음 감사 복음 선교 길 말씀  
 복 인생 부활 비유 생명 신앙 나라 산상 수훈 말 기쁨 세상 십자가 거룩 능력 회복 구원 변화 영적  
 충만 힘 생각 강해 축복 소망 영혼 온전 돌 빛 시작 이유 자녀 가정 고난 성령님 죄 지혜

[그림7] 2020년 2월-2023년 8월까지 주일설교제목에 나타난 주요 키워드

마지막으로 유튜브 상에서 조회 수가 높은 설교를 대상으로 설교제목에 대한 워드클라우드 분석을 수행하였다. 본 연구에서 대상으로 한 설교는 분당우리교회의 주일설교와 날기새(날마다 기막힌 새벽)의 매일설교, 선한목자교회 유기성목사의 설교 그리고 새에덴교회 소강석목사의 설교이다. 각 설교의 제목 키워드에 대한 분석 결과를 순서대로 나타내면 다음과 같다.

**하나님** 예수 기도 사람 믿음 사랑  
 그리스도 삶 성령 주님 교회 복음 생명 은혜 마음 감사 나라 비유  
 인생 말씀 산상 수훈 길 복 회복 거룩 기쁨 영적 선교 힘 능력 7 성도 말 신앙 다시 영혼 변화 고난  
 부활 온전 이유 죄 구원 세상 예배 회개 가지 꿈 생각 충만 돌 십자가 빛

[그림8] 분당우리교회 주일설교제목에 나타난 주요 키워드

**하나님** 사람 말 믿음 사랑 세상 지혜 삶 예수 말씀 힘 은혜  
 죄 마음 기도 고난 심판 이유 사탄 다윗

[그림9] 날기새 매일설교의 제목에 나타난 주요 키워드

**하나님** 주님 은혜 사랑 예수님  
 마음 사람 믿음 감사 기도 기쁨 기적 삶 갈망 구원 동행 생명 순종  
 십자가 예수 그리스도 말 복 부활 빛 세상 자녀 죄 노래 두려움 말씀 열매 영광 지금 찬양

[그림10] 선한목자교회 유기성목사 주일설교제목에 나타난 주요 키워드



**감사 거장 마음 인생 거인 그대**  
**길 사랑 꽃 영혼 코로나 극복 주님 지금 회복 희망**  
 기억 기적 바람 이야기 필요 훌리 again 가슴 겨울 광야 기대 떡 마지막 말씀  
 메세지 미라클 부활 분노 사람 성탄절 송이 암 어게인 어깨 에델바이스 열 위로 은혜 이생  
 일생 절망 처방전 극복 크리스천 행복 향기

[그림11] 새에덴교회 소강석목사 주일설교제목에 나타난 주요 키워드

## V. 결론

본 연구에서는 유튜브 상에 공개된 한국의 대형교회 주일설교 본문 말씀과 제목에 대한 웹 크롤링을 수행하였다. 웹 크롤링은 유튜브에서 제공하는 API를 이용하였고 이를 통해 주일 설교 정보를 얻어온 이후 처리하는 방식을 사용하였다. 설교영상 정보를 작성하는 방식이 교회마다 다르기 때문에 통일된 방법으로 처리하는 것은 불가능하였다. 그럼에도 정보 제공 방식이 비슷한 교회별로 분류하고 그 다음으로는 크롤링한 문서에서 날짜, 예배종류, 설교자, 성경 본문, 설교제목, 조회 수 정보를 추출하였다. 다만, 성경 본문이 둘 이상인 설교인 경우 제일 먼저 나오는 성경 본문만 추출하여 분석에 사용하였다.

웹 크롤링을 통해 주일설교에서 사용한 성경 본문을 살펴보았을 때 한국의 대형교회들은 복음서와 시편, 창세기에 집중되고 있음을 알 수 있었다. 하지만 성경은 권별로 크기가 서로 다기 때문에 빈도수를 절대화 하여 비교하는 것보다 권별 크기에 비례하여 상대 빈도수를 비교하는 것이 더 정확할 것이다. 이를 기반으로 비교하였을 때는 에베소서와 로마서 등 신약성경이 구약성경에 비해 많음을 볼 수 있었다.

두 번째로 설교제목에 나타난 키워드 분석을 수행하였다. 2017년1월1일 이후 2023년 8월까지 게시된 설교의 제목을 분석해 본 결과 하나님, 사람, 주님, 사랑, 예수, 삶, 믿음, 기도와 같은 키워드가 많이 등장하였다. 이것을 코로나가 우리나라에 퍼지기 시작한 2020년 2월을 기준으로 나눠보면 약간 변동이 있음을 알 수 있었다. 코로나 확산 이전에는 하나님, 예수, 사람, 기도, 사랑, 주님 순으로 많이 등장하였는데 코로나 확산 이후에는 하나님, 사람, 주님, 예수님, 삶, 믿음, 사랑 순으로 많이 등장하였다.

연구를 단순히 성경 본문과 설교제목만을 가지고 수행했기 때문에 설교에서 강조하는 내용까지 반영하지는 못했다. 향후 설교영상에 텍스트화 기술을 적용한다면 설교에서 사용되는 용어까지 연구를 수행할 수 있을 것으로 기대된다.

“이 논문은 다른 학술지 또는 간행물에 게재되었거나 게재 신청되지 않았음을 확인함.”

## 참고문헌

- 이종화 (2018). Python을 이용한 SNS 크롤링 시스템 구축, 23(5), 61-76.
- 이경배 (2019). 빅데이터 분석 방법에 의한 사복음서 텍스트의 비교 분석, 제주대학교 석사학위논문.
- 윤성민 (2020). 코로나19 사태에서의 미디어 영상예배를 위한 실천신학적 방법론, 신학과 실천, 69, 67-89.
- 손동식 (2020). 코로나19 시대와 효과적인 온라인 설교에 관한 연구, 대학과 선교, 45, 33-60.
- 안치영, 문상용, 신은혜, 김홍준(2021). 효율적인 웹 크롤링, 데이터 분석 및 시각화 서비스 연구, 디지털문화아카이브지, 4(2), 113-122.
- 반재훈, 하종수, 김동현 (2020). 빅데이터 분석도구 R을 이용한 성경 데이터의 빈도와 소셜 네트워크 분석, 24(2), 166-171.

## 웹 크롤링을 이용한 최근 한국교회 설교 성경 및 제목 분석

### The analysis of recent Korean church sermon Bibles and titles using web crawling

#### 논문초록

인터넷을 통한 정보 공유가 활발해지면서 인터넷에 공개된 정보를 바탕으로 사회 현상과 변화를 탐지하는 기술이 늘고 있다. 특히 웹상에서 정보를 수집하는 웹 크롤링 기술을 이용하여 사회의 방향성을 감지하고 있다. 기독교 사회의 방향성을 알기 위한 방법 중 하나로 교회 예배의 설교 정보를 이용할 수 있다. 본 연구에서는 유튜브 상에 최근에 공개된 한국 대형교회 주일설교에 대한 정보를 웹 크롤링으로 얻어오고 사용되는 성경 본문의 경향과 설교제목에 대한 워드 클라우드 분석을 수행하였다. 성경의 어떤 부분을 본문으로 설교가 이루어지는지, 설교제목에 주로 나타나는 키워드는 어떤 것이 있는지 조사하여 보았다.

주제어: 한국교회 주일설교, 웹 크롤링, 설교 성경, 설교 제목