

2013

국민대학교

자연어처리 연구실

장은서, 권도형, 박소영

지도교수: 강승식

[CopyCheck: 한글문서 표절검사 시스템 version 1.0 사용자 매뉴얼]

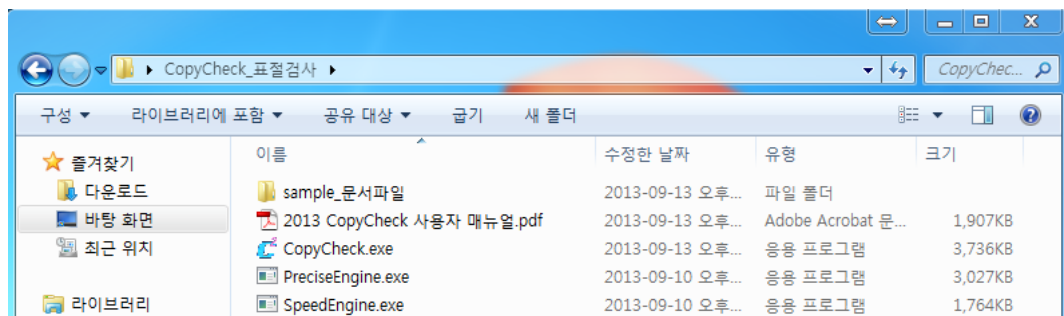
본 문서는 표절검사 프로그램 CopyCheck의 사용자 매뉴얼로써 레포트, 학위 논문 등 여러 문서 파일에 대해 표절 검사 프로그램을 구동시키는 방법을 설명한다.

목차

1 프로그램 설치 및 실행 파일	2
2 표절검사 프로그램 실행 - 윈도우 기반	3
2.1 표절검사 문서 파일 선택	6
2.1.1 '파일' 단위로 선택하기	6
2.1.2 '폴더' 단위로 선택하기	7
2.2 표절검사 모드	8
2.3 표절검사 결과보기	9
2.3.1 검사 결과 보기	9
2.3.2 표절구간 시각화	12
2.4 표절검사 결과 저장하기	13
2.5 표절검사 결과 파일 불러오기	13
3. 프로그램의 기능 제약	14
3.1 프로그램 사용 환경	15
3.2 프로그램 실행	15
3.3 문서 포맷	15
3.4 프로그램 문제 해결	15
4. 기타	16

1 프로그램 설치 및 실행 파일

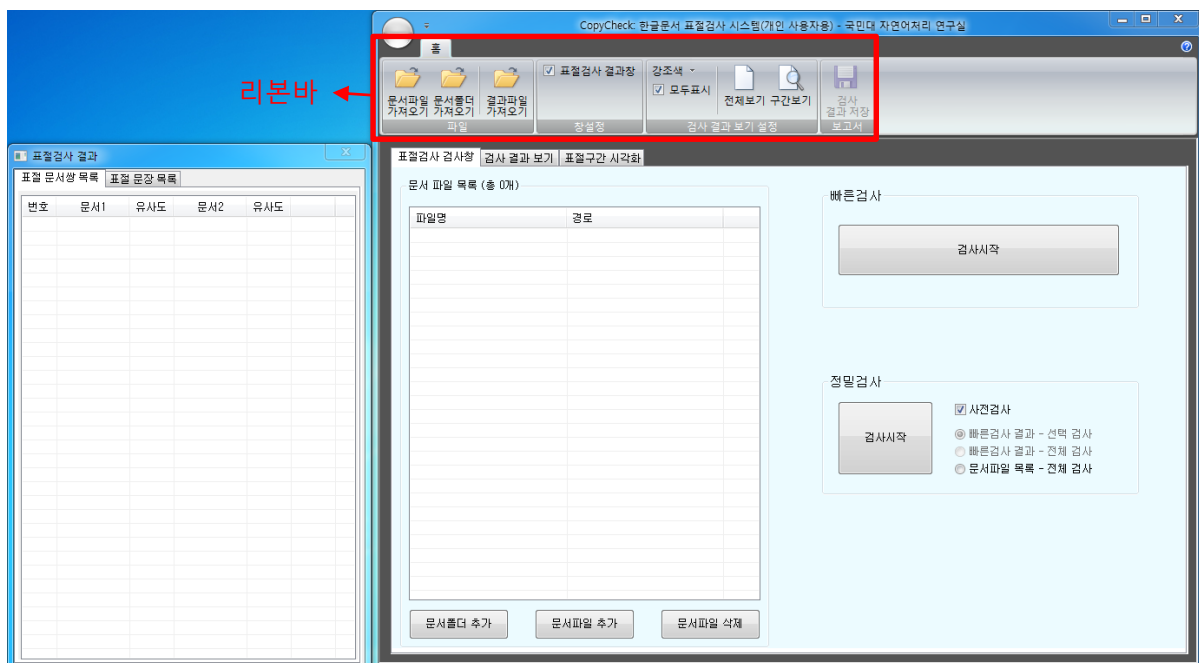
본 문서는 CopyCheck 사용자 매뉴얼로써 표절검사를 하기 위한 각 단계별 프로그램 사용 방법을 설명한다. 표절 검사 프로그램 폴더를 열어보면 [그림 1]과 같이 'sample_문서파일'폴더와 'CopyCheck.exe', 'PreciseEngine.exe', 'SpeedEngine.exe', '2013 CopyCheck 사용자 매뉴얼'이 존재함을 볼 수 있다. 이 폴더 안에 존재하는 '.exe' 파일들은 표절검사 프로그램 수행에 있어 꼭 필요한 파일이며 같은 폴더(경로)에 존재하여야 표절검사 프로그램이 올바르게 동작한다. 그리고 'sample_문서파일'폴더는 표절검사 프로그램 실행을 위한 예제 파일들이 존재한다.



[그림 1] 표절검사 프로그램 파일 구성물

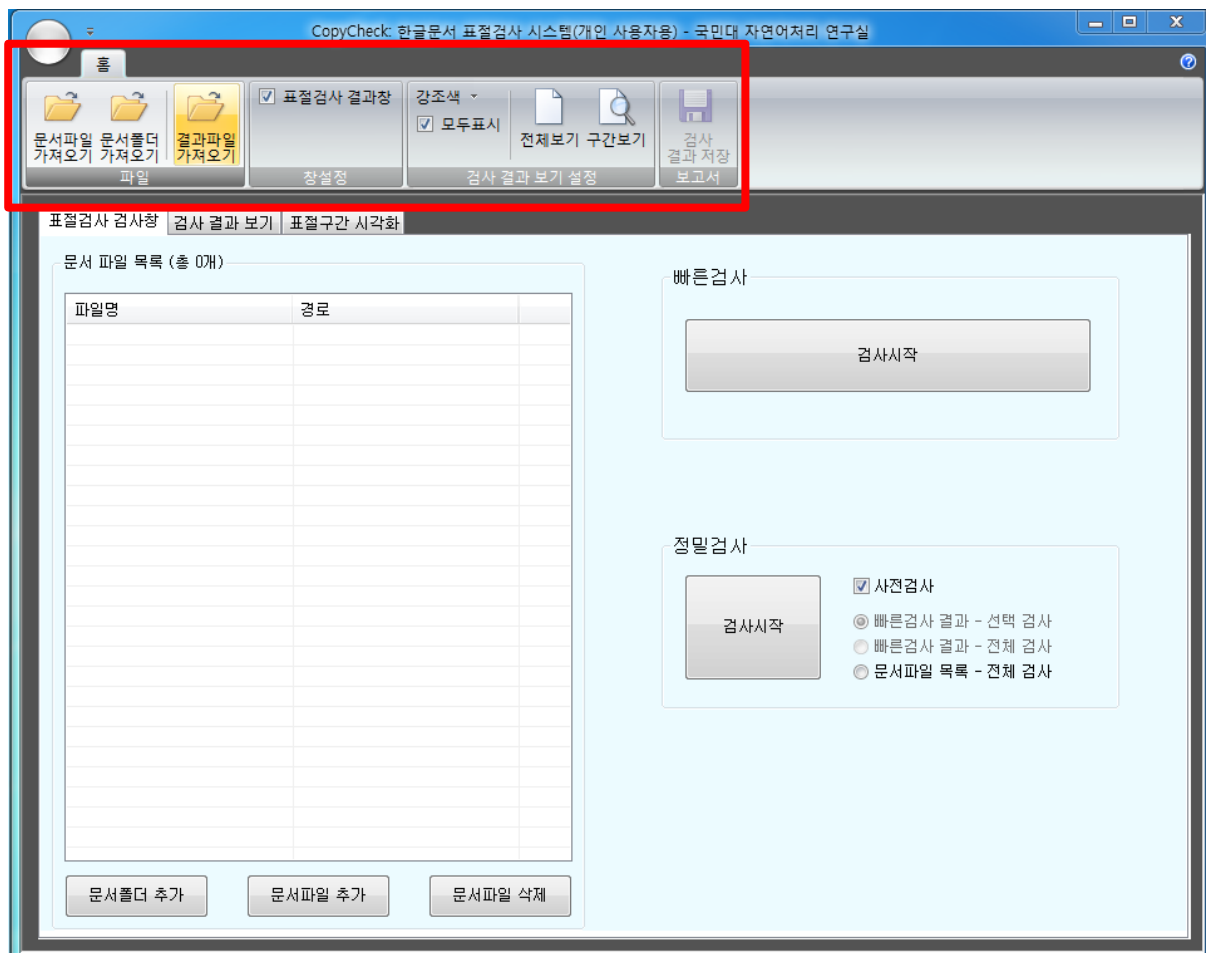
2 표절검사 프로그램 실행 - 윈도우 기반

CopyCheck.exe 파일을 실행하면 [그림 2]과 같이 CopyCheck 초기 화면이 나타난다. 초기 화면은 [그림 2]과 같이 표절검사를 진행할 main화면과 main화면 좌측에 위치한 '표절검사 결과'창으로 구성된다. main 화면의 상단은 '홈' 1개의 탭으로 구성되어 있으며, 리본바 하단은 '표절검사 검사창', '검사 결과 보기', '표절구간 시각화' 3개의 탭으로 구성되어 있다. 초기 프로그램을 실행하면 '표절검사 검사창' 탭에서 시작하게 된다.



[그림 2] CopyCheck.exe 프로그램 실행 화면

'홈' 탭의 메뉴는 [그림 3]와 같이 '파일', '창설정', '검사 결과 보기 설정', '보고서' 4개의 카테고리가 있다. '파일' 카테고리에는 표절검사를 위한 문서파일 가져오기, 문서폴더 가져오기, 결과파일 가져오기를 선택하기 위한 버튼들이 있으며 '창설정' 카테고리에는 main화면의 좌측에 위치한 '표절검사 결과'창 활성화를 설정할 수 있는 체크박스가 있다. '검사 결과 보기 설정' 카테고리에는 표절검사 결과를 보기 위한 강조색, 모두표시, 전체보기, 구간보기를 선택하기 위한 버튼들이 있다. '보고서' 카테고리에는 표절검사 수행 후, 결과 파일을 저장하는 '검사 결과 저장' 버튼이 있다.



[그림 3] 표절검사 main 화면 - 리본바

표절검사 과정은 크게 3단계로 구성되어 있다.

- 첫째, 표절 검사할 파일/폴더 선택
- 둘째, 표절검사 모드 선택
- 셋째, 표절검사 결과보기 모드 선택

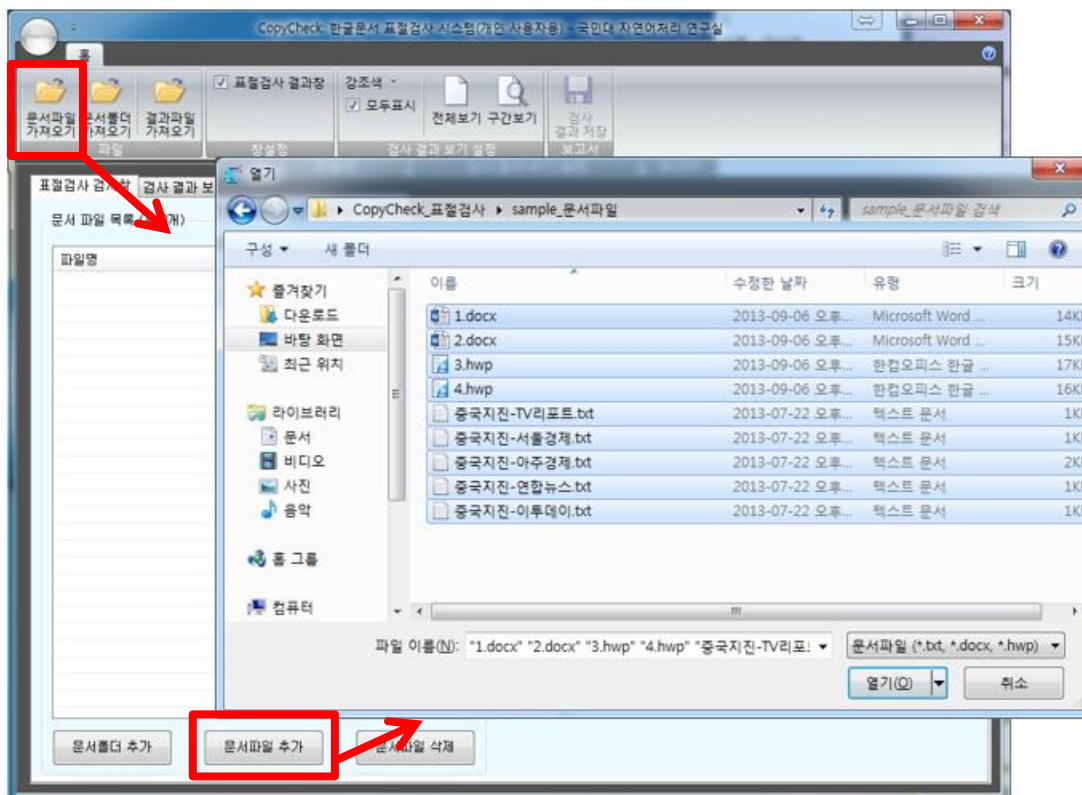
'CopyCheck: 한글문서 표절검사 시스템 version 1.0'은
현재 한글(.hwp), 워드(.docx), 텍스트(.txt)의 문서 포맷
만을 지원합니다.

2.1 표절검사 문서 파일 선택

표절 검사할 문서 파일을 선택하는 방법은 '파일' 단위로 선택하는 방법과 '폴더' 단위로 선택하는 방법 2가지가 있다. 또한 파일을 선택하는 버튼이 상단과 하단에 2가지 있으며, 상단에 위치한 버튼은 기존의 파일 목록을 없애고 새롭게 목록을 생성하는 기능이며, 하단에 위치한 버튼은 기존의 파일 목록에 추가되는 기능이다.

2.1.1 '파일' 단위로 선택하기

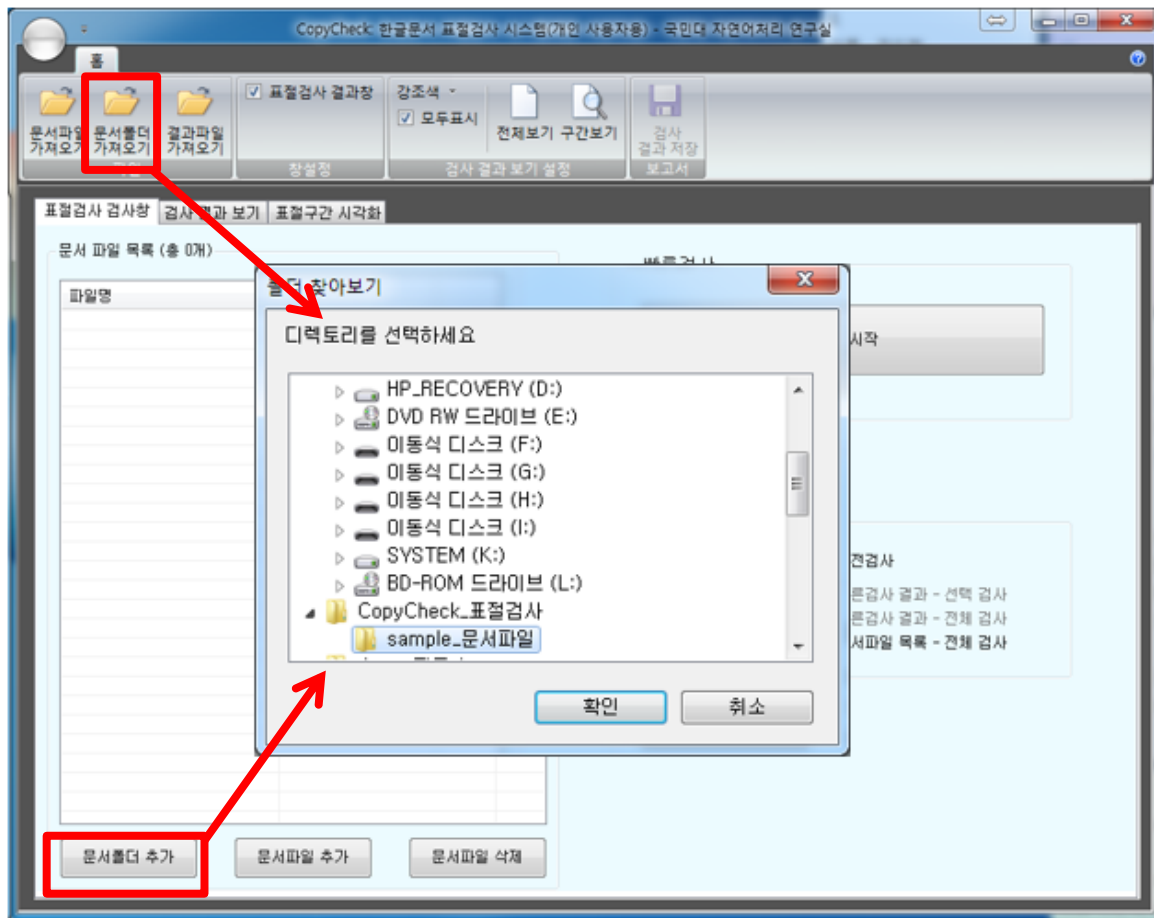
표절 검사할 파일을 '파일' 단위로 선택하는 방법은 상단의 '문서파일 가져오기' 또는 하단의 '문서파일 추가' 버튼을 클릭하는 것이다. 클릭 시 파일을 선택할 수 있도록 탐색기 창이 나타나며, 표절 검사할 문서 파일들을 선택하고 열기 버튼을 누르면 된다. [그림 4]는 'sample_문서파일' 폴더 아래의 파일들을 선택한 예이다.



[그림 4] 표절검사 과정 - '파일' 단위로 선택한 화면

2.1.2 '폴더' 단위로 선택하기

표절 검사할 파일을 '폴더' 단위로 선택하는 방법은 상단의 '문서폴더 가져오기' 또는 하단의 '문서폴더 추가' 버튼을 클릭하는 것이다. 클릭 시 폴더를 선택할 수 있도록 탐색기 창이 나타나며, 표절 검사할 문서 폴더를 선택하고 확인 버튼을 누르면 된다. [그림 5]은 'sample_문서파일' 폴더를 선택한 예이다.

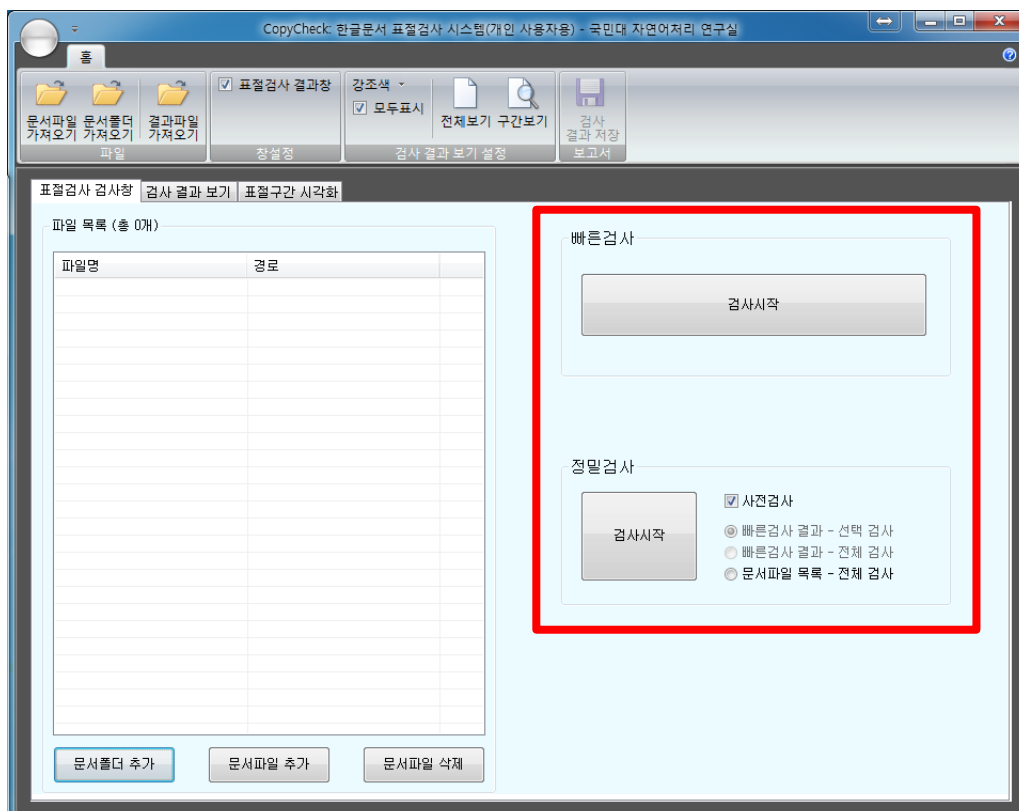


[그림 5] 표절검사 과정 - '폴더' 단위로 선택한 화면

2.2 표절검사 모드

표절검사 모드는 '빠른검사'와 '정밀검사'가 있다. '빠른검사'는 검사 속도는 빠르지만 '정밀검사'에 비해 표절 의심 구간을 파악하는 정확도가 떨어질 수 있다. 반면 '정밀검사'의 경우는 '빠른검사'에 비해 검사 속도가 느린 편이지만 표절 의심 구간을 파악하는 정확도가 '빠른검사'에 비해 높다.

'빠른검사'는 검사 후 '빠른검사 결과 - 선택 검사'와 '빠른검사 결과 - 전체 검사' 2가지 모드가 있다. 먼저 '빠른검사 결과 - 선택 검사'는 빠른검사를 통해 표절이 의심되는 문서 파일들을 추려내고, 사용자가 원하는 문서 파일들을 선택하여 정밀 검사하는 기능이다. 다음으로 '빠른검사 결과 - 전체 검사'는 빠른 검사를 통해 추려낸 문서 파일들에 대하여 전체적으로 정밀 검사하는 기능이다. '문서파일 목록 - 전체 검사'의 버튼을 체크한 후, 정밀검사의 검사시작 버튼을 누르면 사용자가 업로드한 문서 파일 전체에 대해 정밀검사를 진행한다.

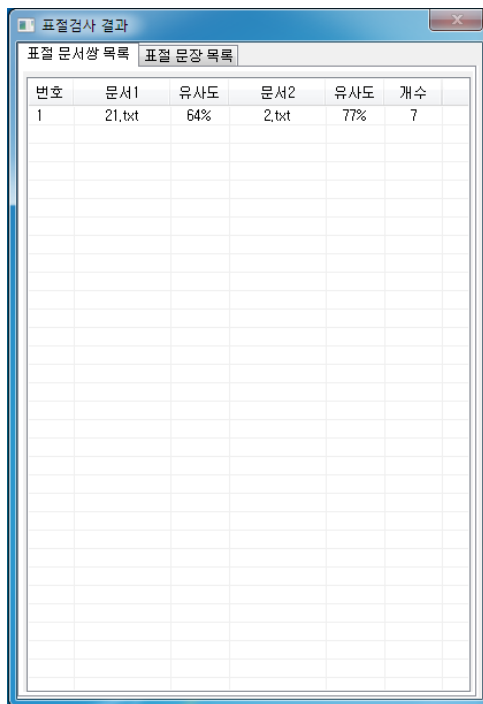


[그림 6] 표절검사 main 화면 - 검사 모드

2.3 표절검사 결과보기

표절 검사를 진행한 후, 표절 의심 문서 파일들이 존재하면 [그림 7]과 같이 '표절검사 결과' 창의 '표절 문서쌍 목록' 탭에 한 행마다 한 쌍의 의심되는 파일들의 '이름'과 '유사도', 의심되는 '문장 수'가 출력된다. 상세하게 보고 싶은 한 행을 더블 클릭하면 자동으로 '표절검사 결과'창은 [그림 8]와 같이 '표절 문서쌍 목록' 탭에서 '표절 문장 목록' 탭으로 이동하게 되고, main화면은 '표절검사 검사창' 탭에서 '검사 결과 보기' 탭으로 이동한다.

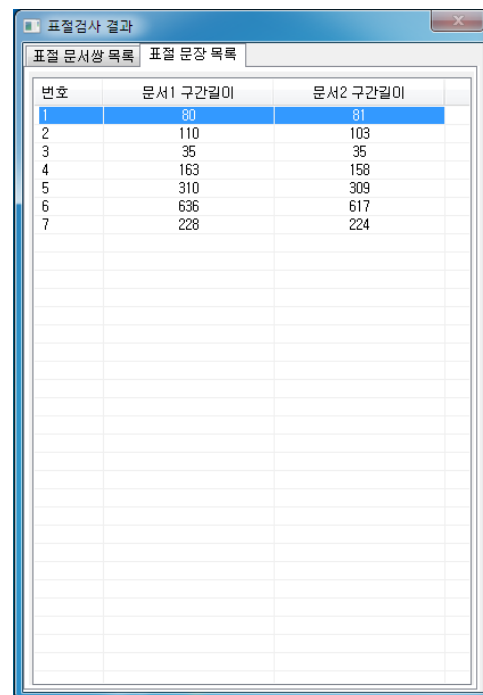
표절검사 결과는 사용자가 업로드한 문서 파일들 중에서 표절 의심 문서 파일들에 대해 1:1 비교 방식으로 나타내며, 결과를 보는 방법은 '검사 결과 보기'와 '표절구간 시각화' 2가지가 있다. '검사 결과 보기'는 표절 의심 문서 파일들 중 각각의 의심되는 문장을 보여준다. 또한 '표절구간 시각화'는 '검사 결과 보기'와는 달리 표절 의심 문서 파일들의 내용이 보여지지 않고 표절 의심되는 구간의 분포도를 보여준다. 자세한 내용은 각각의 카테고리에서 설명하도록 한다.



번호	문서1	유사도	문서2	유사도	개수
1	21.txt	64%	2.txt	77%	7

[그림 7] 표절검사 결과창

- 표절 문서쌍 목록



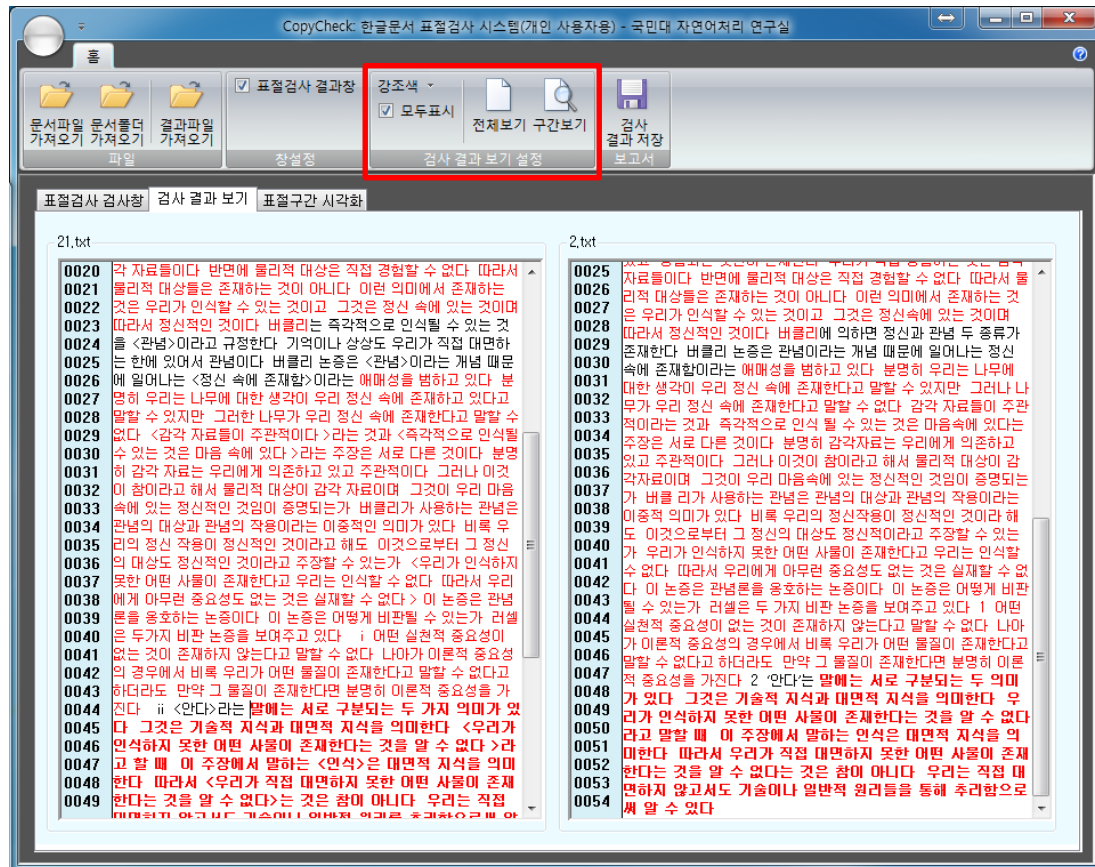
번호	문서1 구간길이	문서2 구간길이
1	80	81
2	110	103
3	35	35
4	163	158
5	310	309
6	636	617
7	228	224

[그림 8] 표절검사 결과창

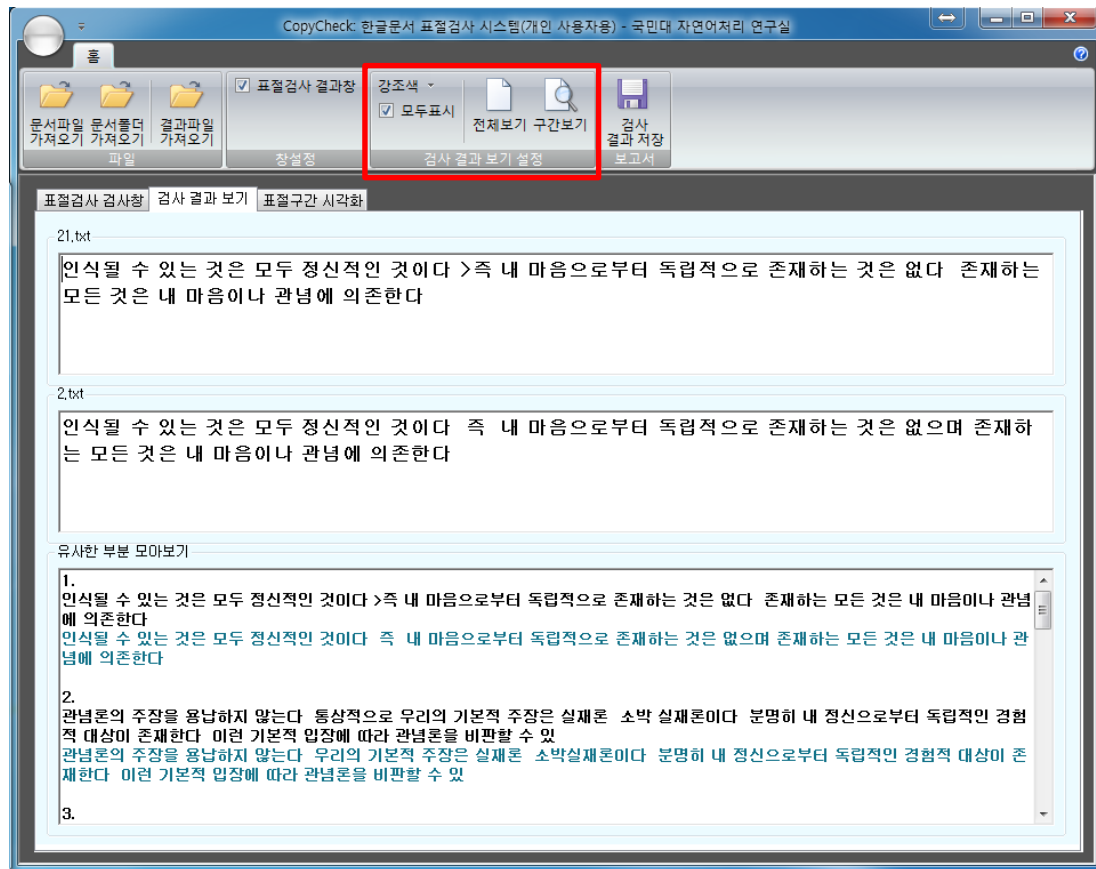
- 표절 문장 목록

2.3.1 검사 결과 보기

‘검사 결과 보기’는 ‘전체보기’와 ‘구간보기’ 2가지 방법이 있다. 이는 리본바에 위치한 ‘검사 결과 보기 설정’ 카테고리에서 선택할 수 있다. ‘전체보기’와 ‘구간보기’는 [그림 8]와 같은 ‘표절검사 결과’창에서 각 행을 클릭하면 해당 행의 표절 의심 구간에 대해 비교하여 볼 수 있다. ‘전체보기’는 [그림 9]와 같이 표절 의심 문서 파일의 내용 전체를 보여주며, 그 중 의심되는 문장을 ‘강조색’을 통해 보여준다. 반면 ‘구간보기’는 [그림 10]과 같이 표절 의심 문서 파일 중 의심되는 문장 별로 비교하여 보여준다.



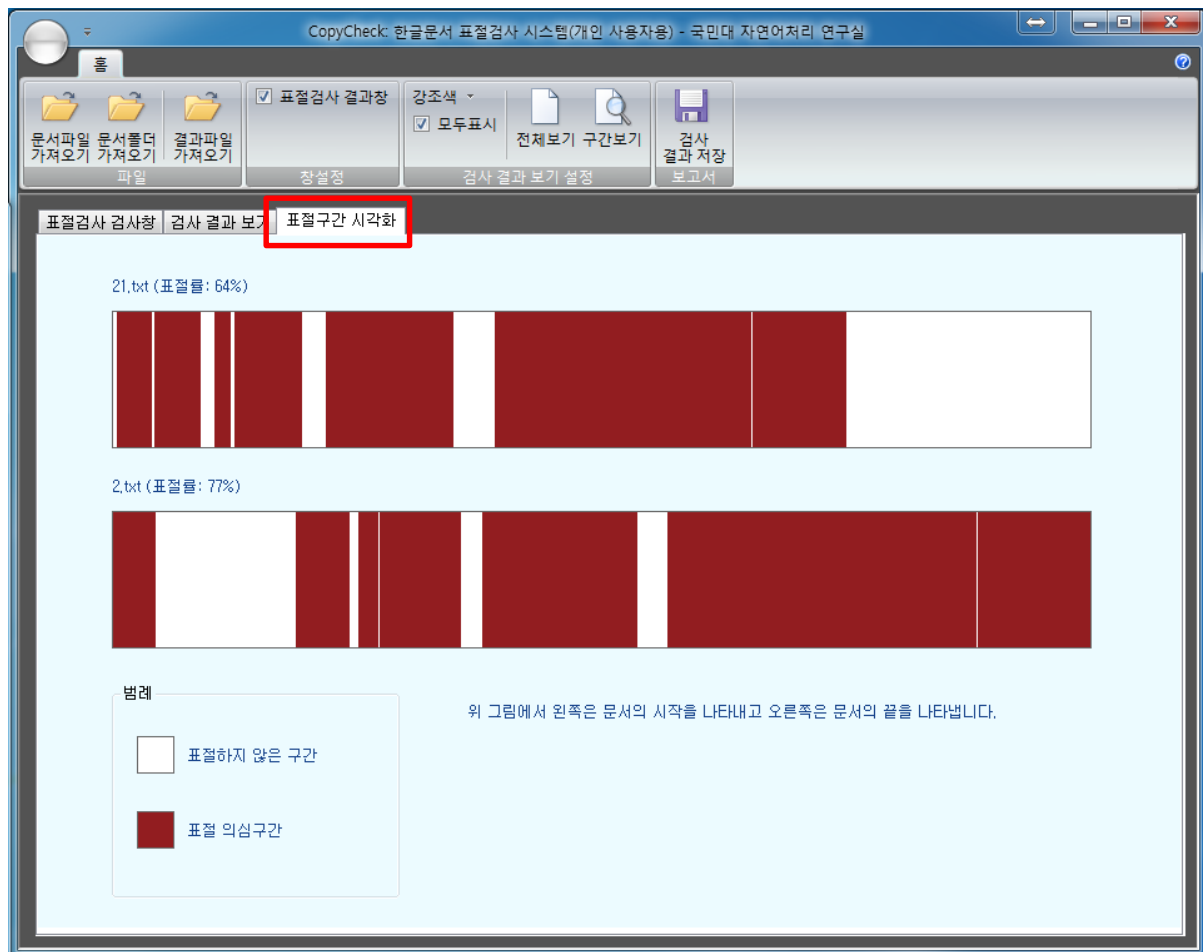
[그림 9] 검사 결과 보기 - 전체보기



[그림 10] 검사 결과 보기 - 구간보기

2.3.2 표절구간 시각화

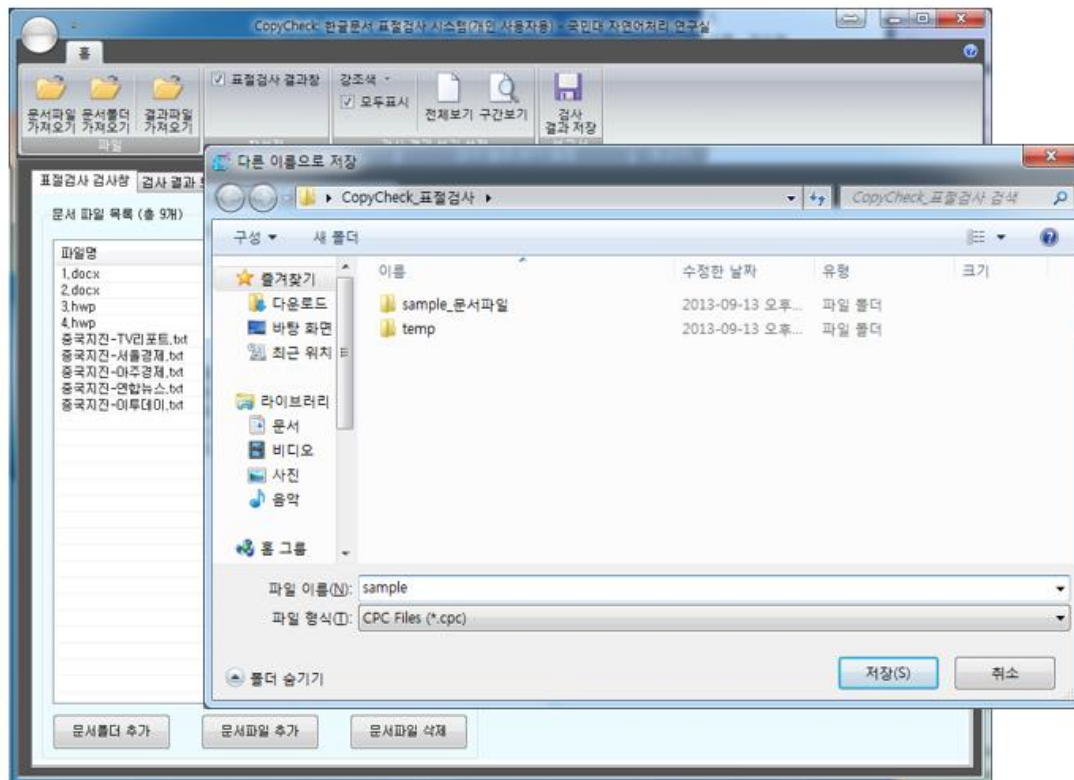
‘표절구간 시각화’는 앞서 설명한 ‘검사 결과 보기’와는 달리 실제 의심되는 문서 파일의 내용이 보이지 않는다. 다만, [그림 11]와 같이 표절 의심 문서 파일에 대해서 표절 의심 구간인 곳들을 표시하여 사용자가 표절 의심구간의 분포를 한 눈에 확인할 수 있다는 이점이 있다.



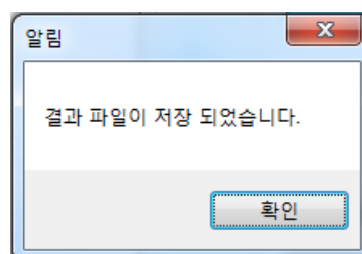
[그림 11] 표절구간 시각화

2.4 표절검사 결과 저장하기

표절검사 결과를 저장하고 싶다면 '보고서' 카테고리의 '검사 결과 저장' 버튼을 클릭한다. 이 때 [그림 12]과 같이 탐색기 창이 뜨면 원하는 경로를 선택한 다음 원하는 파일 이름을 입력하고 저장버튼을 누른다. 이와 같이 진행하면 [그림 13]와 같이 '알림'창이 뜬다.



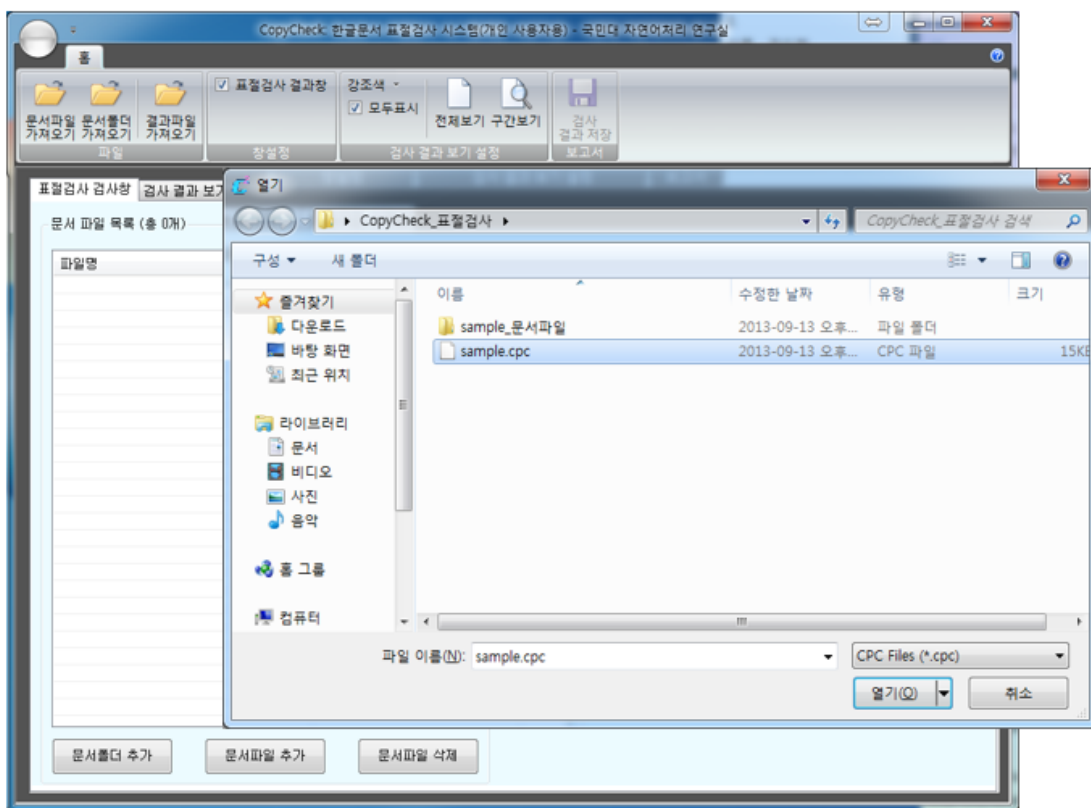
[그림 12] 표절검사 main화면 - 검사 결과 저장



[그림 13] 결과 저장 '알림'창

2.5 표절검사 결과 파일 불러오기

이전에 표절검사를 진행한 결과를 다시 보고 싶다면, [그림 14]와 같이 '파일' 카테고리의 '결과파일 가져오기' 버튼을 클릭한다. 이때 주의할 점은 이전에 검사를 진행하고 나서 검사 결과 파일(*.cpc)을 저장하였을 때 가능하다.



[그림 14] 표절검사 main화면 - 결과 파일 가져오기

3. 프로그램의 기능 제약

3.1 프로그램 사용 환경

- Windows 7, 8 환경에서 테스트되었으며, 다른 버전의 Windows에서 정상 작동을 보장할 수 없음
- Linux, OS X 등 기타 운영체제는 지원하지 않음

3.2 프로그램 실행

- CopyCheck는 한번에 하나의 프로그램만 실행 가능

3.3 문서 포맷

- 한글(.hwp), MS워드(.docx), 텍스트(.txt)파일 이외의 문서 포맷은 지원하지 않음
 - txt파일 중 character set이 UTF-8인 경우는 읽지 못할 수 있음

3.4 프로그램 문제 해결

- 프로그램이 비정상적으로 종료된 경우: 임시 파일이 저장되어 있는 폴더가 삭제되지 않은 경우, 이후에 진행되는 검사에 영향을 미칠 수 있으므로 프로그램이 위치한 폴더 아래의 temp폴더를 수동으로 삭제해야 함

4. 기타

관련 문의 사항은 아래의 메일 주소로 보내주십시오.

- nlplab.kmu@gmail.com