

ToNK MANUAL

Text Of NetworK

KAIST

연락처

문의사항 및 시스템 오류에 대한 질의사항이 있을 경우 아래의 연락처로 문의 주시기 바랍니다.

문일철

icmoon@smslab.kaist.ac.kr

한국과학기술원

산업 및 시스템 공학과 조교수

오혜연

alice.oh@kaist.edu

한국과학기술원

전산학과 조교수

Alice Haeyun Oh

Copyright Notification and Disclaimer

COPYRIGHT (c) 2009-2011 Ill Chul Moon, Alice Heayun Oh - Korea Advanced Institute of Science and Technology (KAIST), 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Republic of Korea - ALL RIGHTS RESERVED.

By downloading the files provided herein you indicate your agreement with this copyright notice and disclaimer. Permission to use, copy and modify this version of the software or any parts of it and its documentation is hereby granted for RESEARCH ONLY purposes and provided that the above copyright notice and this permission notice appear intact in all copies of the software, that you do not sell the software, nor include the software in a commercial package.

The release of this software into the public domain does not imply any obligation on the part of the authors to release future versions into the public domain. The authors are free to make upgraded or improved versions of the software available for a fee or commercially only. Commercial licensing of the software is available by contacting Dr. Ill Chul Moon (icmoon@smslab.kaist.ac.kr) or Dr. Alice Heayun Oh (alice.oh@kaist.edu). The software is provided "as is" and without warranty of any kind, express, implied or otherwise, including without limitation, any warranty of merchantability or fitness for a particular purpose. The software is experimental only and has not been designed for, tested or approved for hazardous use. The use of the software is at the user's own risk. Any conclusions you may draw based on the software or its use are your own. We expressly disclaim any responsibility or liability for any and all adverse effects, including personal, bodily, property or business injury, and for damages or loss of any kind whatsoever, resulting directly or indirectly, whether from negligence, intent or otherwise, from the use or disuse of the software, from errors in the software, or from misunderstandings arising from the software itself, its us or its documentation.

분석 과정 미리 보기

TONK는 크게 2 가지 분석이 있습니다. 첫 번째는 문서의 내용 및 메타 정보를 분석하는 것이고 두 번째는 문서의 개념어 간 연결 관계를 분석하는 것입니다. 이 분석들은 일정 단계까지는 동일한 과정을 거치지만 그 이후 분석에서는 다른 분석 알고리즘을 이용하게 됩니다. 2 가지 분석에 대한 자세한 내용은 다음과 같습니다.

I. 문서 내용 및 메타 정보 분석

문서의 내용을 분석하는 것은 분석 대상인 문서들이 어떤 화제를 가장 비중 있게 다루었는지 알아보는 것입니다. 메타 정보를 분석하는 것은 문서의 내용을 분석한 결과를 토대로 하여 화제(이슈), 저자의 영향력, 문서 간 관계 등을 파악하는 것입니다. 특히 메타 정보를 확인하는 것은 ORA라는 프로그램을 별도로 이용합니다. 다음은 문서 내용 및 메타 정보 분석에 대한 과정입니다.

A. 프로젝트 생성

일반적으로 워드나 한글을 사용하여 처음 문서작업을 시작할 때 ‘새 문서’라는 것을 들어보셨을 겁니다. TONK에서는 ‘새 분석 프로젝트’를 하나의 분석 단위로 사용하고 있는데요. 이 분석 프로젝트는 다시 4 개의 하위폴더로 구성되어 있습니다. 분석 대상을 모아놓은 문서, 분석 결과로 얻은 저자 정보를 모아놓은 저자 그리고 저자와 완료되면 사용자가 원하는 결과를 획득 할 수 있게 됩니다.

이슈의 관계를 나타내는 네트워크, 사전을 모아놓은 사전입니다. 하나의 프로젝트를 완성하기 위해서는 B. 본문 정제, C. 자연어 처리, D. 사전 생성, F. 저자-이슈 분석, H. 저자-이슈-성향 분석 J. 포스트 참조관계망을 순서대로 실행해야 합니다. 이들은 분석 프로젝트에 수반된 기능들로써 실행이 모두 완료되면 사용자가 원하는 결과를 획득 할 수 있게 됩니다.

B. 본문 정제

본문 정제는 바로 이후에 이어질 ‘자연어 처리’의 실행을 원활하게 하기 위해 행해지는 기능으로 프로그램이 문서를 인식하기 쉽게 정제합니다. 빈 칸이 너무 많을 경우를 대비하여 여유 공란을 삭제한다거나 분석에 필요하지 않은 특수 기호나 숫자, 문장 기호를 삭제할 수 있습니다. 또한 영어 소문자를 대문자로, 영어 대문자를 소문자로도 바꿀 수 있습니다.

"해외원정 도박으로 사회에 물의를 일으켰지만 다시 일어날 수 있도록 기회를 준 국민 여러분께 조금이나마 보답하고 어려운 이웃에게 봉사하고 싶어 행사를 시작했다"고 말했다. 황기순은 작년까지 모금한 성금으로 전동 휠체어 27대, 일반 휠체어 847대를 구입해 장애인들에게 전달했다. engine@yna.co.kr(끝)저작권자(c)연합뉴스 무단전재·재배포금지. >

문장 기호 삭제를 실행시키기 전 화면

amp quot 해외원정 도박으로 사회에 물의를 일으켰지만 다시 일어날 수 있도록 기회를 준 국민 여러분께 조금이나마 보답하고 어려운 이웃에게 봉사하고 싶어 행사를 시작했다 amp quot 고 말했다 황기순은 작년까지 모금한 성금으로 전동 휠체어 27대 일반 휠체어 847대를 구입해 장애인들에게 전달했다 engine yna co kr 끝 저작권자 c 연합뉴스 무단전재 재배포금지 amp gt

문장 기호 삭제를 실행시킨 후 화면

C. 자연어 처리

자연어 처리는 형태소, 품사, 구문분석, NER 정보를 생성하는 것입니다. 이는 바로 이후에 이어질 사전을 생성하는 것을 돋기 위해 행해지는 것으로 프로그램이 단어를 개개로 인지할 수 있도록 문서를 세분화 시키는 작업입니다. 음식을 먹는다고 할 때 통째로 삼켜 먹는 것보단 이로 씹어서 소화시키는 것이 편리하겠죠! 자연어 처리를 음식을 씹어먹는 과정이라 생각하시면 됩니다.

무려(nc) 되(xsv) 었(ep) 던(etm) '(s) 패닉(nc) '(s) 상황(nc) 까지(jx) 는(jx) 안가(nc) ((s) 뉴욕(nc) =(s) UAE(nc))(s) 두바이(nc) 정부(nc) 소유(nc) 최대(nc) 지주(nc) 회사(nc) 01(co) ↳ (etm) 두바이(nc) (xsv) 었(ep) 다고(ec) ㅎㅎ(px) 는(etm) 사실(nc) 상(xsn) 의(jm) '(s) 모라토리엄(nc) '(s) 선언(nc) 소식(nc) 되(xsv) 는(etm) 모습(nc) 를(jc) 보이(pv) 고(ec) 있(px) 다(eft) .(s)

자연어 처리 실행 후 생성된 형태소의 예

D. 사전 생성

사전은 특정 단어들을 모아놓은 것으로 바로 이후에 이어질 ‘저자-이슈 분석’ 혹은 ‘저자-이슈 성향 분석’을 실행시키기 위해 필수적으로 행해져야 하는 기능입니다. TONK에서는 ‘사전’의 개념이 가장 큰 특징이라 해도 과언이 아닌데요. 이러한 사전은 크게 4 가지 종류가 있으며 각 사전들을 문서에 적용시키는 방식으로 ‘저자-이슈 분석’과 ‘저자-이슈 성향 분석’을 위한 가장 최적화된 단어들을 수집합니다.

E. 저자-이슈 분석

‘저자-이슈 분석’은 문서의 내용을 파악하는데 핵심적인 역할을 하는 토픽을 추출해내는 역할을 합니다. 토픽은 바로 이전 과정에서 생성하였던 사전을 기반으로 문서에서 가장 비중 있게 다루어지고 서로 관련 있는 단어들을 한대 묶어 나타낸 것입니다.

토픽
노무현_행사_공천_사진_운동_서거_추모_영장_모금_후원_기부_세상_김대중_민주_묘역_전시_재단_민주주의_구속_조성 방송_대사_시청_드라마_직원_현장_대원_사무_인기_파견_확인_도미니카_대사관_여성_추노_예능_외교관_촬영_유재광_출연 수정_해임_여론_제출_의견_투표_만심_원안_방법_국정_당론_분열_여야_사과_논란_사퇴_현법_이정현_최고위원_지도 PD수첩_소송_방송_광우병_청구_문화_가정_출신_검사_민사_마련_순례_배상_변호사_증거_국방부_입영_고통_자료_기각

생성된 토픽

F. 저자-이슈-성향 분석

‘저자-이슈-성향 분석’은 문서의 내용을 파악하는데 핵심적인 역할을 하는 토픽과 그의 성향을 추출해내는 역할을 합니다. 토픽은 바로 이전 과정에서 생성하였던 사전을 기반으로 문서에서 가장 비중 있게 다루어지고 서로 관련 있는 단어들을 한데 묶어 나타낸 것입니다.

토픽ID	성향	비중값	토픽
20	+	0.056048	두바이_지수_글락_달러_위기_화물_투자_페닉_금융_약재_자금_반동_선언_금리_아부다비_대안_확인_중국_국제_신중
11	+	0.054248	도박_구속_카지노_마카오_언예인_기소_적발_자금_가수_인기_상습_비카_유명_대학교수_호텔_지도_입간_선고_인천_밸런트
17	+	0.051025	판매_휴대폰_애플_삼성전자_모토로라_이익_영업_매출_점유_인기_삼성_개발_SK텔레콤_비다_환경_부문_경제_터치_풀이_도전
10	+	0.050700	의원_공천_영장_구속_개발_인증_뱅킹_전남도_기각_청구_꼴인_지방선거_승천_현직_보안_현금_자금_증거_인멸_도주
16	+	0.049886	선언_신청_공사_업무_소송_선고_재정_집행_조기_지도_기소_경찰_후원_환경_대강_보수_행정_정지_공작_파업
12	+	0.049796	의원_수정_해임_제출_중국_분열_의견_원안_확인_선언_시퇴_여론_이전_논란_민주_합의_축구_민주주의_최고위원_재단

G. 사용자 해석

‘저자-이슈 분석’ 혹은 ‘저자-이슈-성향 분석’을 통해 토픽이 생성되면 사용자는 이를 구성하는 단어들을 토대로 하여 수집한 정보에서 어떤 화제들이 주로 다뤄졌는지 추측할 수 있습니다. ‘저자-이슈-성향 분석’과 같은 경우는 토픽 이외에 토픽의 성향까지 알 수 있습니다.

토픽ID	성향	비중값	토픽
11	+	0.052499	남북_남북자_언론_합의_청와대_원칙_대변인_협의_국군_회견_지지_수정_자료_사설_선거_국정_발굴_이전_남측_조기

‘저자-이슈-성향 분석’의 성향값이 나타난 예

토픽ID	비중값	토픽
1	0.071175	노무현_행사_공천_사진_운동_서거_추모_영장_모금_후원_기부_세상_김대중_민주_묘역_전시_재단_민주주의_구속_조성

‘저자-이슈 분석’의 결과가 나타난 예

예를 들어 ‘**노무현_행사_공천_사진_운동_서거...**’를 보면 단어들의 조합이 전 노무현 대통령과 김대중 대통령의 서거와 관련된 화제라는 것을 알 수 있습니다. 토픽을 산출해 내는 과정까지는 프로그램의 역할이지만 이를 보고 화제를 파악하는 일은 사용자의 몫인데요. **토픽을 구성하는 모든 단어들이 서로 간의 연관성을 100% 보여주지 못하는 경우도 있으므로 사용자의 해석이 필수적이게 됩니다.**

예문을 다시 살펴보면 ‘**노무현_행사_공천_사진_운동_서거_추모_영장_모금...민주주의_구속_조성**’에서 ‘사진’, ‘전시’, ‘조성’은 해당 토픽에서 말하고자 하는 화제와는 관련이 멀다는 것을 알 수 있습니다. 따라서 사용자는 이를 배제하고 주요 화제를 파악할 수 있어야 하죠. 이렇게 사용자의 판단까지 이루어지면 문서의 내용을 분석하는 것은 종료됩니다.

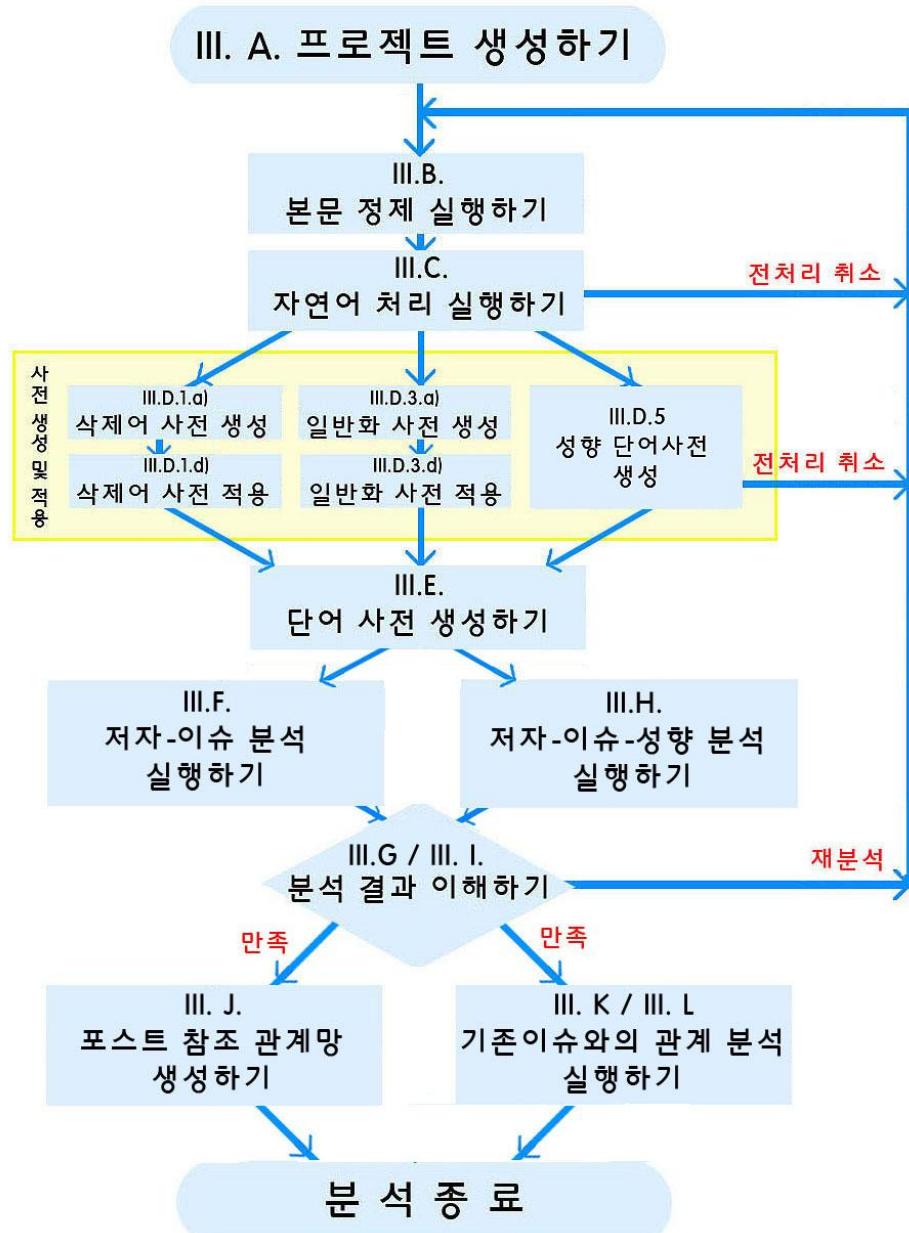
H. 포스트 참조관계망

생성된 토픽과 사용자의 판단으로 문서의 내용이 파악되면 이를 토대로 문서의 메타 정보를 분석할 수 있습니다. 포스트 참조관계망은 저자가 작성한 문서가 전체 문서 내에서 얼마만큼의 영향력을 나타냈는지 산출해냅니다. 문서의 조회 수, 찬성 수, 반대 수 등을 기준으로 문서의 영향력을 산출한 후 ORA를 이용하여 결과를 열람합니다. 이로써 TONK의 첫 번째 분석이 종료됩니다.

I. 기존 이슈와의 관계 분석

‘포스트 참조관계망’의 과정에서와 같이 생성된 토픽과 사용자의 판단으로 문서의 내용이 파악되면 이를 토대로 문서의 메타 정보를 분석할 수 있습니다. ‘기존 이슈와의 관계 분석’은 ‘저자-이슈 분석’과 ‘저자-이슈-성향 분석’의 결과를 중심으로 프로젝트에서 다루어진 토픽을 다른 프로젝트의 토픽과 비교하여 시간에 흐름에 따라 토픽의 비중이 어떻게 변화하였는지 측정합니다. 이로써 TONK의 첫 번째 분석이 종료됩니다.

● 첫번째 분석 - 문서 내용 및 메타 정보 분석



- ① 프로젝트 생성하기(III.A. page 53) : 새 분석 프로젝트를 생성합니다.
- ② 본문 정제 실행하기 (III.A. , page 63) : 특정 기호나 숫자, 영어 소문자 등을 본문에서 정제합니다.
- ③ 자연어 처리 실행하기(III.C. , page 70) : 형태소, 품사, 구문분석, NER 정보를 생성합니다.
- ④ 사전 생성 및 적용
 - 삭제어 사전 생성(III.D.1.a), page 78) : 문서에서 필요하지 않은 단어들을 수집합니다.
 - 일반화 사전 생성(III.D.3.a), p age 99) : 문서에서 동일한 의미의 단어를 수집합니다.
 - 성향 단어사전 생성(III.D.5.a), p age 116): 문서에서 특정 성향을 띤 단어를 수집합니다.
 - 삭제어 사전 적용(III.D.1.d), page 88) : 문서에서 필요하지 않은 단어들이 삭제됩니다.
 - 일반화 사전 적용(III.D.3.d), page 108) : 동일한 의미의 단어가 한 단어로 통일이 됩니다.
- ⑤ 단어 사전 생성하기(III.E. , p age 125) : 문서에서 필요한 단어들을 수집합니다.
- ⑥ 저자-이슈 분석 실행하기(III.F. , page 135) : 토픽을 생성합니다.
- ⑦ 저자-이슈-성향 분석 실행하기(III.H. , page 170) : 토픽과 토픽의 성향을 산출합니다.
- ⑧ 분석 결과 이해하기(III.G., page 160, III.I., page 170) : 사용자가 산출된 토픽을 기반으로 하여 정보를 파악합니다.
- ⑨ 포스트 참조관계망 실행하기 (III.J. , page 199): 문서의 메타 정보를 분석합니다.
- ⑩ 기존 이슈와의 관계 분석 실행하기(III.K., page 226, III.L., page 238): 문서의 메타 정보를 분석합니다.
- ⑪ 분석 종료: 분석이 완료됩니다.

II. 문서의 개념어 간 연결관계

문서의 개념어 간 연결관계를 분석하는 것은 TONK에서 지정한 특정 개념을 가진 개념어들이 서로 관계가 있는지의 여부를 파악하는 것입니다. 다시 말하면, TONK는 개념어들을 추출하여 이들 간의 관계가 존재하는지 알아봅니다. 특히 결과를 확인하는 과정에서는 ORA라는 프로그램을 별도로 이용합니다. 다음은 문서의 개념어 간 연결관계에 대한 과정입니다.

A. 프로젝트 생성

일반적으로 워드나 한글을 사용하여 처음 문서작업을 시작할 때 ‘새 문서’라는 것을 들어보셨을 겁니다. TONK에서는 ‘새 분석 프로젝트’를 하나의 분석 단위로 사용하고 있는데요. 이 분석 프로젝트는 다시 4개의 하위폴더로 구성되어 있습니다. 분석 대상을 모아놓은 문서, 분석 결과로 얻은 저자 정보를 모아놓은 저자 그리고 저자와 이슈의 관계를 나타내는 네트워크, 사전을 모아놓은 사전입니다. 하나의 프로젝트를 완성하기 위해서는 B. 본문 정제, C. 자연어 처리, D. 사전 생성, E. 문서내 관계망을 순서대로 실행해야 합니다. 이들은 분석 프로젝트에 수반된 기능들로써 실행이 모두 완료되면 사용자가 원하는 결과를 획득 할 수 있게 됩니다.

B. 본문 정제

본문 정제는 바로 이후에 이어질 ‘자연어 처리’의 실행을 원활하게 하기 위해 행해지는 기능으로 프로그램이 문서를 인식하기 쉽게 정제합니다. 빈 칸이 너무 많을 경우를 대비하여 여유 공란을 삭제한다거나 분석에 필요하지 않은 특수 기호나 숫자, 문장 기호를 삭제할 수 있습니다. 또한 영어 소문자를 대문자로, 영어 대문자를 소문자로도 바꿀 수 있습니다.

"해외원정 도박으로 사회에 물의를 일으켰지만 다시 일어날 수 있도록 기회를 준 국민 여러분께 조금이나마 보답하고 어려운 이웃에게 봉사하고 싶어 행사를 시작했다"고 말했다. 황기순은 작년까지 모금한 성금으로 전동 휠체어 27대, 일반 휠체어 847대를 구입해 장애인들에게 전달했다. engine@yna.co.kr(끝)저작권자(c)연합뉴스 무단전재·재배포금지. >

문장 기호 삭제를 실행시키기 전 화면

amp quot 해외원정 도박으로 사회에 물의를 일으켰지만 다시 일어날 수 있도록 기회를 준 국민 여러분께 조금이나마 보답하고 어려운 이웃에게 봉사하고 싶어 행사를 시작했다 amp quot 고 말했다 황기순은 작년까지 모금한 성금으로 전동 휠체어 27대 일반 휠체어 847대를 구입해 장애인들에게 전달했다 engine yna co kr 끝 저작권자 c 연합뉴스 무단전재 재배포금지 amp gt;

문장 기호 삭제를 실행시킨 후 화면

C. 자연어 처리

자연어 처리는 형태소, 품사, 구문분석, NER 정보를 생성하는 것입니다. 이는 바로 이후에 이어질 사전을 생성하는 것을 돋기 위해 행해지는 것으로 프로그램이 단어를 개개로 인지할 수 있도록 문서를 세분화 시키는 작업입니다. 음식을 먹는다고 할 때 통째로 삼켜 먹는 것보단 이로 씹어서 소화시키는 것이 편리하겠죠! 자연어 처리를 음식을 씹어먹는 과정이라 생각하시면 됩니다.

우려(nc) 되(xsv) 었(ep) 던(etm) '(s) 패닉(nc) '(s) 상황(nc) 까지(jx) 는(jx) 안가(nc) ((s) 뉴욕(nc) =(s) UAE(nc))(s) 두바이(nc) 정부(nc) 소유(nc) 최대(nc) 지주(nc) 회사(nc) 이(co) ㄴ (etm) 두바이(nc) (xsv) 었(ep) 다고(ec) 하(px) 는(etm) 사실(nc) 상(xsn) 의(jm) '(s) 모라토리엄(nc) '(s) 선언(nc) 소식(nc) 되(xsv) 는(etm) 모습(nc) 들(jc) 보이(pv) 고(ec) 있(px) 다(eft) .(s)

자연어 처리 실행 후 생성된 형태소의 예

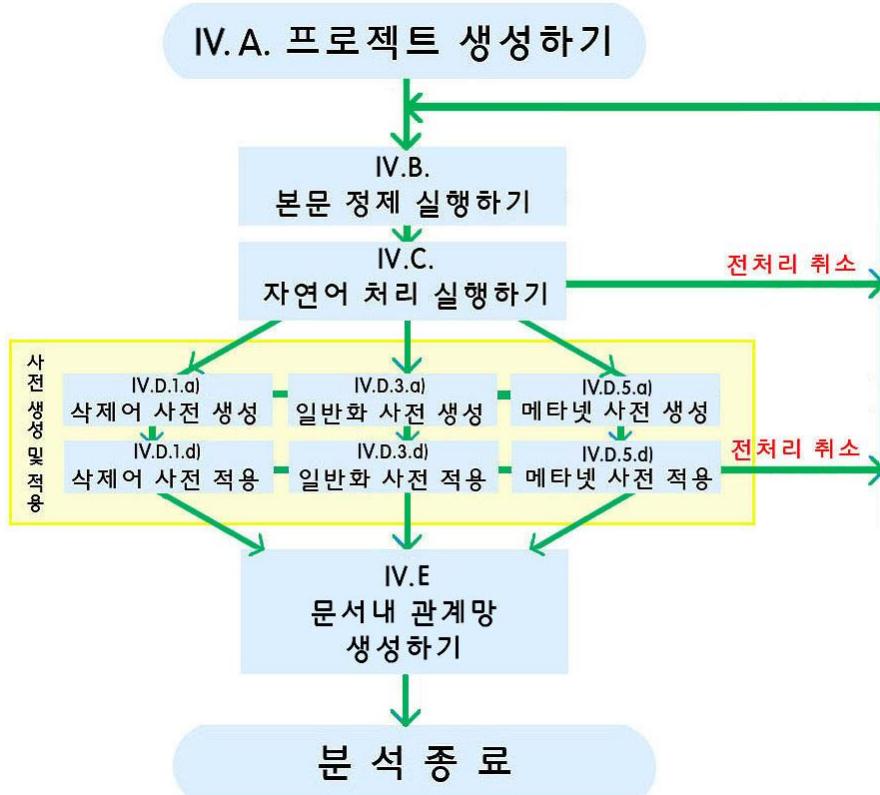
D. 사전 생성

사전은 특정 단어들을 모아놓은 것으로 바로 이후에 이어질 ‘문서내 관계망’을 실행시키기 위해 필수적으로 행해져야 하는 기능입니다. TONK에서는 ‘사전’의 개념이 가장 큰 특징이라 해도 과언이 아닌데요. 이러한 사전은 총 3 가지가 있으며 각 사전들을 문서에 적용시키는 방식으로 문서내 관계망을 위한 가장 최적화된 단어들을 수집합니다.

E. 문서내 관계망

3 개의 사전이 모두 생성되면 이를 기반으로 단어와 단어 사이에 관계가 존재하는지 알아봅니다. 여기서 단어는 문서 내에 있는 모든 단어가 아니라 TONK 가 지정한 특정 단어들을 말합니다. 이 단어를 개념이라 일컫는데요. 개념이란 기관, 행동, 자원, 장소 등 특정한 개념을 가지고 있는 단어를 말합니다. 특정한 개념은 총 8 가지가 있으며 온톨로지라 부릅니다. 사용자가 온톨로지를 직접 선택하면 TONK 가 온톨로지에 해당하는 개념어를 찾아 이들의 관계를 산출해냅니다. 추출된 개념어를 열람하고 이들간의 관계를 확인하는 과정은 ORA 라는 프로그램을 통해 이루어집니다. 이로써 두번째 분석인 문서의 개념어 간 연결관계가 종료됩니다.

● 두번째 분석 - 문서의 개념어 간 연결관계



- ① 프로젝트 생성하기(IV.A. page 249) : 새 분석 프로젝트를 생성합니다.
 - ② 본문 정제 실행하기 (IV.A. , page 259) : 특정 기호나 숫자, 영어 소문자 등을 본문에서 정제합니다.
 - ③ 자연어 처리 실행하기(IV.C. , page 266) : 형태소, 품사, 구문분석, NER 정보를 생성합니다.
 - ④ 사전 생성 및 적용
 - 삭제어 사전 생성(IV.D.1.a), page 273) : 문서에서 필요하지 않은 단어들을 수집합니다.
 - 일반화 사전 생성(IV.D.3. a), p age 294) : 문서에서 동일한 의미의 단어를 수집합니다.
 - 메타넷 사전 생성(IV.D.5. a), p age 312) : 메타넷 사전을 생성합니다.
 - 삭제어 사전 적용(IV.D.1.d), page 283) : 문서에서 필요하지 않은 단어들이 삭제됩니다.
 - 일반화 사전 적용(IV.D.3.d), page 303) : 동일한 의미의 단어가 한 단어로 통일이 됩니다.
 - 메타넷 사전 적용(IV.D.5. a), p age 318) : 메타넷 사전을 문서에 적용합니다.

- ⑤ 문서내 관계망 생성하기(IV.E., page 326) : 문서에서 단어들 간의 관계가 존재하는지 분석합니다.
- ⑥ 분석 종료: 분석이 완료됩니다.

목차

연락처	1
분석 과정 미리 보기	2
CHAPTER I TONK 와 친해지기	17
A. 실행 환경 및 구성.....	17
1. 소프트웨어 환경	17
2. 플랫폼 환경	17
B. 설치하기	18
C. TONK 실행하고 끝내기	23
1. TONK 실행하기	23
2. TONK 끝내기.....	24
CHAPTER II TONK 화면 살펴보기	25
A. 제목 표시줄	26
B. 메뉴 막대	27
1. 파일	27
2. 보기	31
3. 전처리.....	32
5. 분석	37
6. 도움말.....	43
C. 표준 도구	44
1. 첫번째 텍스트 파일	44
2. 이전 텍스트 파일	44
3. 다음 텍스트 파일	44

4. 마지막 텍스트 파일		44
5. 글꼴		44
6. 글꼴 크기		44
7. 전처리 되돌리기		44
8. 여유 공란 삭제		44
9. 문장 기호 삭제		45
10. 심볼 삭제		45
11. 숫자 삭제		45
12. 영어 대문자화		45
13. 영어 소문자화		45
14. 바꾸기		45
15. 형태소/품사/구문분석/NER 생성		45
17. 일반화 사전 적용		45
19. 도움말		46
D. 컴포넌트 트리		47
E. 작업창		48
F. 메시지 판넬		49
1. 콘솔		49
2. 속성		50
G. 상태표시줄		51
H. 기타 일러두기		52
CHAPTER III 첫번째 분석 - 문서의 내용 및 메타 정보		53
A. 프로젝트 생성하기		53

1.	새 분석 프로젝트 생성하기.....	53
2.	새 분석 프로젝트 생성 처리 결과 보기	56
3.	분석 프로젝트 내보내기	59
4.	분석 프로젝트 가져오기	61
B.	본문 정제 실행하기	63
1.	본문 정제	63
2.	본문정제 처리결과 보기.....	66
C.	자연어 처리 실행하기.....	70
1.	자연어 처리.....	70
2.	자연어 처리 처리결과보기	73
D.	사전 생성하기.....	75
1.	삭제어 사전	78
2.	삭제어 사전 처리결과 보기	91
3.	일반화 사전	99
4.	일반화 사전 처리 결과 보기	110
5.	성향 단어 사전	116
E.	단어 사전 생성하기	125
1.	단어 사전	125
2.	단어 사전 처리결과 보기	132
F.	저자-이슈 분석 실행하기.....	135
1.	저자-이슈 분석.....	135
2.	저자-이슈 분석 처리 결과 보기 – 이슈와 문서 중심으로 보기	139
3.	저자-이슈 분석 처리 결과 보기 – 저자 중심으로 보기	157
G.	저자-이슈 분석 결과 이해하기.....	160
1.	전체 토픽.....	160
2.	토픽	162
3.	저자	163
4.	문서	165
5.	토픽간 관계	167

H. 저자-이슈-성향 분석 실행하기	170
1. 저자-이슈-성향 분석	170
2. 저자-이슈-성향 분석 처리결과 보기.....	174
I. 저자-이슈-성향 분석 결과 이해하기	191
1. 전체 토픽.....	191
2. 토픽	194
3. 저자	195
4. 문서	197
J. 포스트 참조관계망 생성하기.....	199
1. 포스트 참조관계망 생성.....	199
2. 포스트 참조관계망 처리 결과 보기	202
K. 기존 이슈와의 관계 분석 실행하기 – 저자-이슈 분석.....	226
1. 기존 이슈와의 관계 분석 – 저자-이슈 분석 중심으로	226
2. 기존 이슈와의 관계 분석 처리 결과 보기	233
L. 기존 이슈와의 관계 분석 실행하기 – 저자-이슈-성향 분석	238
1. 기존 이슈와의 관계 분석 – 저자-이슈-성향 분석 중심으로.....	238
2. 기존 이슈와의 관계 분석 처리 결과 보기	244
CHAPTER IV 두번째 분석 - 문서의 개념어 간 연결관계.....	249
A. 프로젝트 생성하기	249
1. 새 분석 프로젝트 생성하기.....	249
2. 새 분석 프로젝트 생성 처리 결과 보기	252
3. 분석 프로젝트 내보내기	255
4. 분석 프로젝트 가져오기	257
B. 본문 정제 실행하기	259
1. 본문 정제	259
2. 본문정제 처리결과 보기.....	262
C. 자연어 처리 실행하기.....	266

1. 자연어 처리.....	266
2. 자연어 처리 처리결과보기	269
D. 사전 생성하기.....	271
1. 삭제어 사전	273
2. 삭제어 사전 처리결과 보기	286
3. 일반화 사전	294
4. 일반화 사전 처리 결과 보기	305
5. 메타넷 사전	312
6. 메타넷 사전 처리 결과 보기	320
E. 문서내 관계망 생성하기.....	326
1. 문서내 관계망 생성	326
2. 문서내 관계망 처리결과 보기	330
부록	361
품사 리스트	361

Chapter I TONK 와 친해지기

A. 실행 환경 및 구성

1. 소프트웨어 환경

Windows 7 64bit/32bit

Windows Vista 64bit/32bit

Windows XP 64bit/32bit

Java Virtual Machine Sun JRE 1.6 이상

각 운영체제의 bit 체계에 맞는 TONK Installer 를 이용해야 합니다.

2. 플랫폼 환경

a) 최소사양

CPU Quad-Core

RAM 4 GB 이상

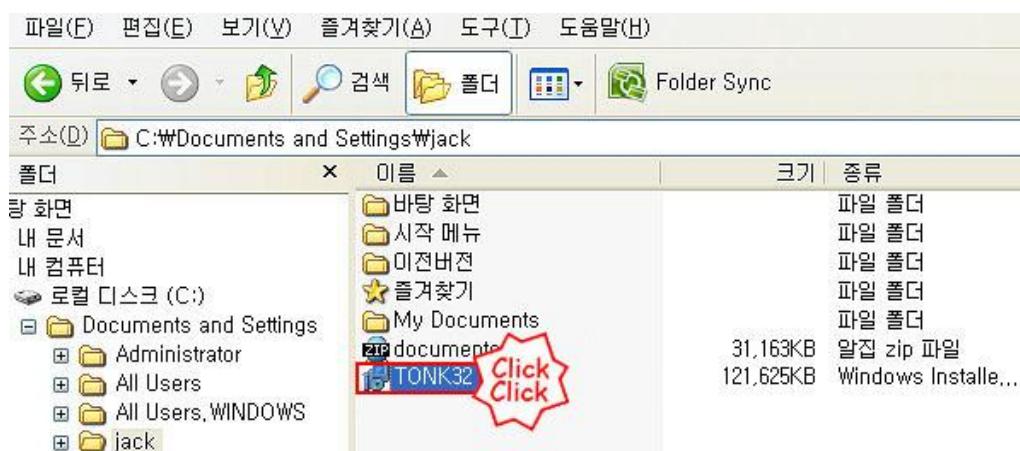
b) 권장사양

CPU Dual Quad-Core

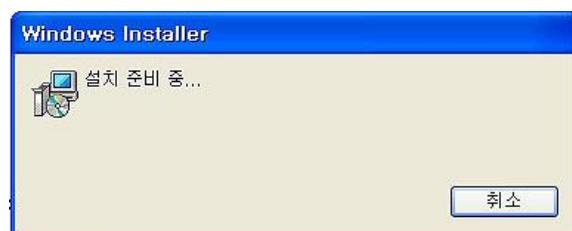
RAM 6GB 이상

B. 설치하기

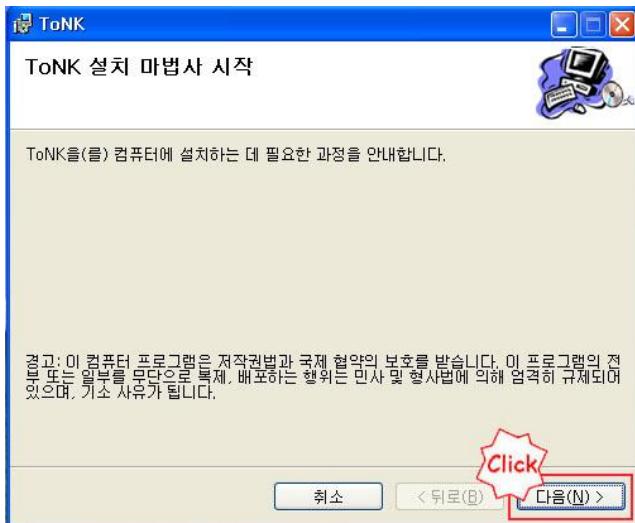
- TONK32.exe 를 더블 클릭합니다.



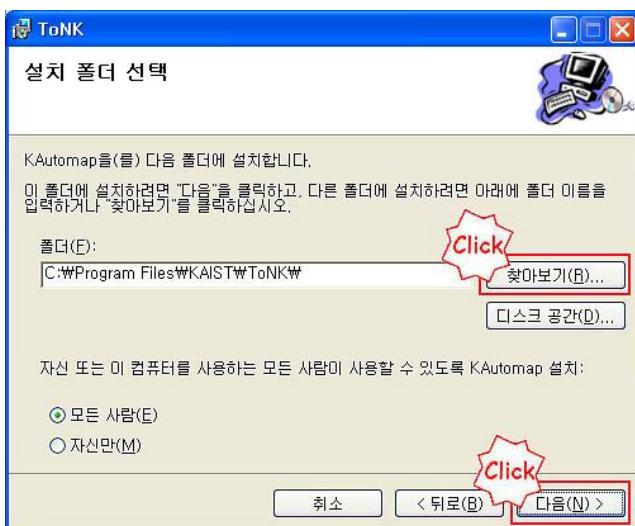
- 'Windows Installer' 대화상자가 나타나면 다음 화면이 나타날 때까지 기다립니다. 이때 <취소>를 누르면 설치가 중단됩니다.



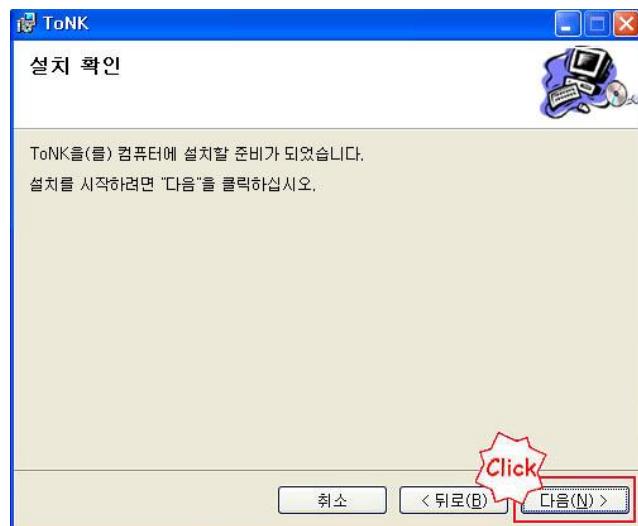
3. ‘TONK 설치 마법사 시작’ 대화 상자가 나타나면 <다음> 버튼을 클릭합니다. 이 때 <취소>를 누르면 설치가 중단됩니다.



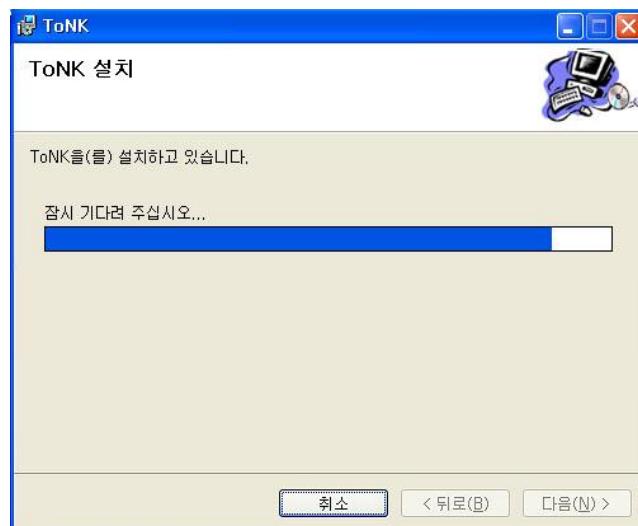
4. ‘설치 폴더 선택’ 대화상자가 나타나면 <찾아보기> 버튼을 클릭하여 설치할 장소를 선택한 후 <다음>을 클릭합니다. 이 때 <뒤로>를 클릭하면 이전 대화상자로 이동하며 <취소>를 클릭하면 설치가 중단됩니다.



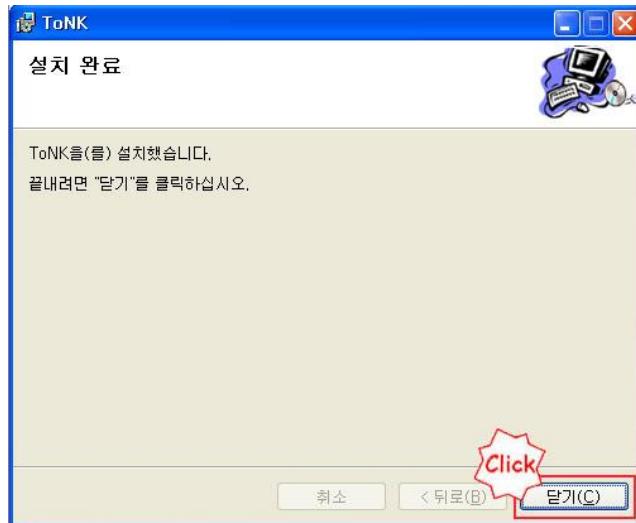
5. 설치 확인 대화상자가 나타나면 <다음>을 클릭합니다. 이 때 <뒤로>를 클릭하면 이전 대화상자로 이동하며 <취소>를 클릭하면 설치가 중단됩니다.



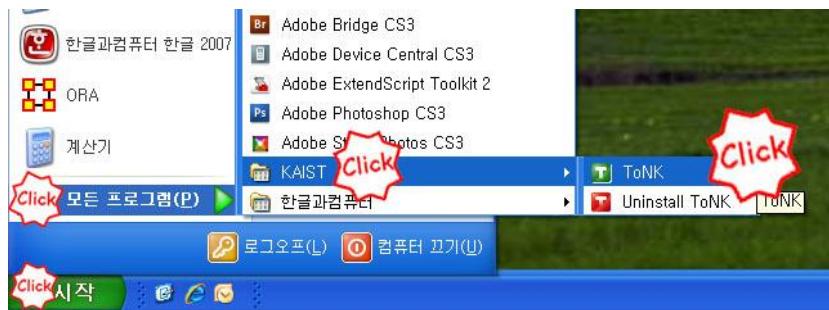
6. 'TONK 설치' 대화상자가 나타나면 막대가 채워질 때까지 기다립니다. 이 때 <취소>를 클릭하면 설치가 중단됩니다.



7. ‘설치 완료’ 대화상자가 나타나면 <닫기> 를 누릅니다.



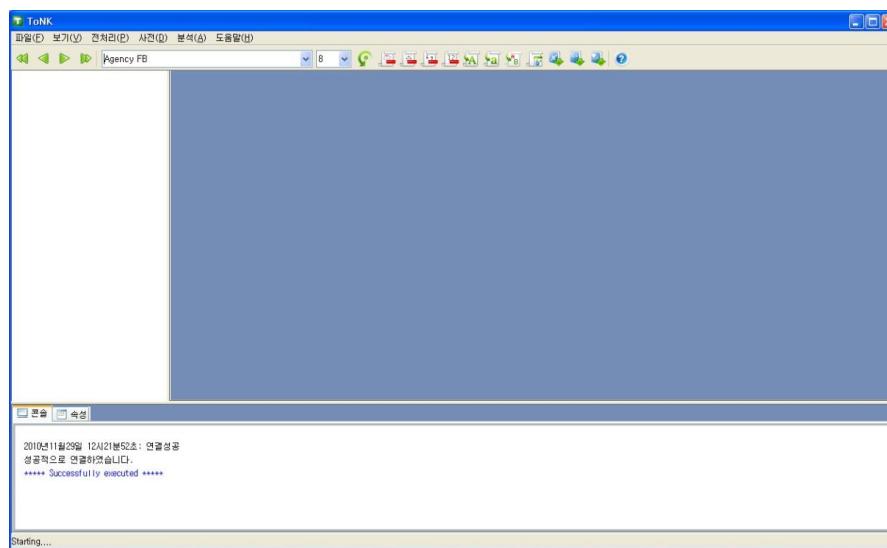
8. 설치된 TONK 프로그램을 확인하려면 **시작—모든프로그램—KAIST — TONK** 를 차례대로 클릭합니다.



9. 프로그램이 실행되고 있는 모습입니다. 좌측에 나타나는 검정색 대화상자는 사용자가 직접 활용하는 것은 아니며 프로그램을 사용하는 동안 나타나게 됩니다. 이 대화상자는 프로그램을 사용하는 동안 에러가 발생하였을 때 원인을 찾아내는 역할을 합니다. 만약 프로그램에 에러가 발생하게 되면 당시 화면을 캡처하여 KAIST로 보내주시면(연락처 참조, page 1) 문제를 해결해드립니다.



프로그램이 시작하는 화면

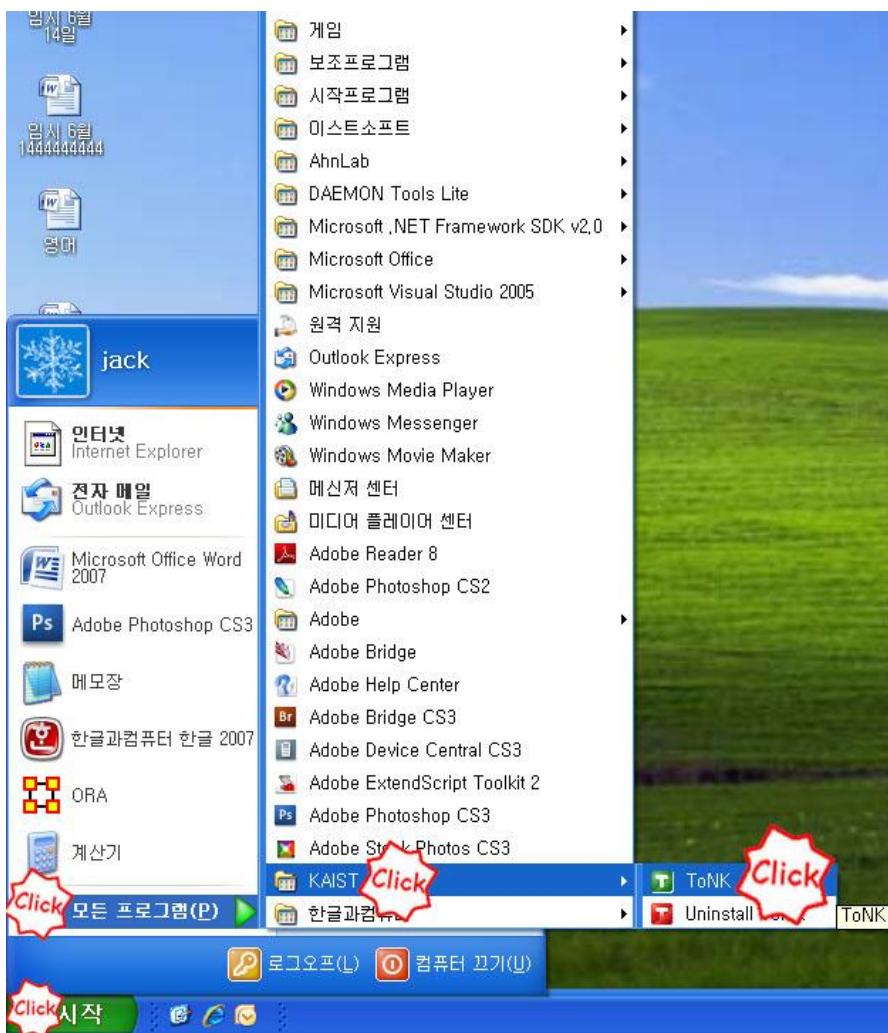


프로그램을 시작하였을 때 초기화면

C. TONK 실행하고 끝내기

1. TONK 실행하기

TONK를 실행하려면 윈도우의 <시작> 도구를 클릭한 후 모든 프로그램 — KAIST — TONK를 차례로 클릭합니다.



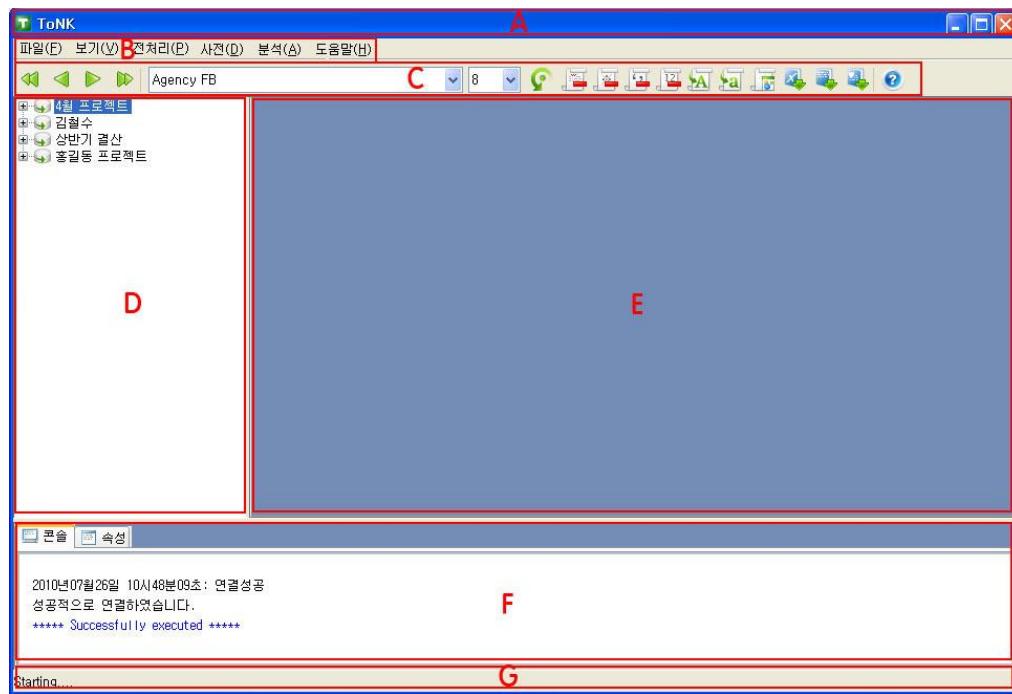
2. TONK 끝내기

열려있는 모든 분석 탭과 TONK 를 한꺼번에 종료하려면 <파일> 메뉴 막대에 <나가기(X)> 를 클릭합니다.



Chapter II TONK 화면 살펴보기

TONK를 실행하였을 때 나타나는 전체화면은 다음과 같습니다.



A. 제목 표시줄

제목표시줄 오른쪽에는 프로그램의 창의 크기를 조절하거나 조절할 수 있는 ‘최소화’, ‘최대화’ 및 ‘이전 크기로 복원’, ‘닫기’ 단추가 존재합니다.



B. 메뉴 막대

TONK의 모든 기능을 주제별로 모아 놓은 곳으로 각 메뉴를 클릭하면 세부항목이 펼쳐집니다.

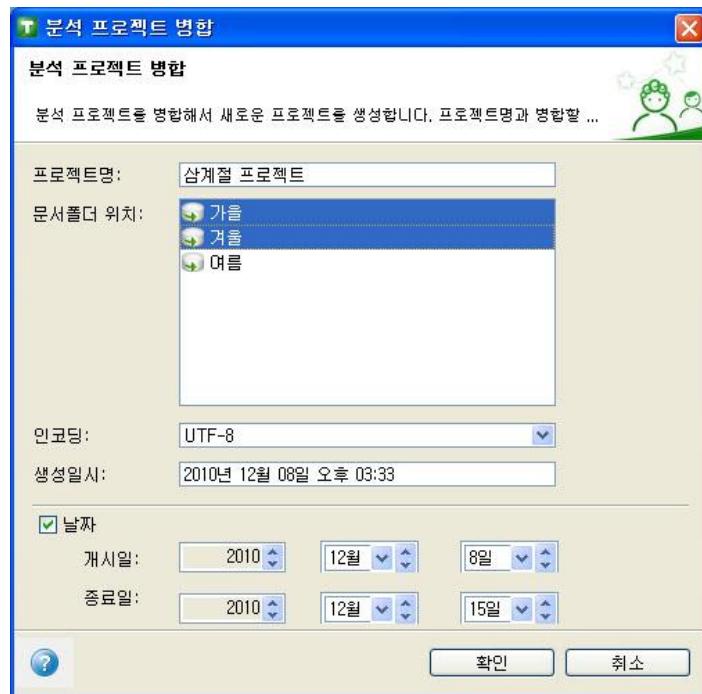
파일(F) 보기(V) 전처리(P) 사전(D) 분석(A) 도움말(H)

1. 파일

파일을 관리를 하는 기능을 모아두었습니다.



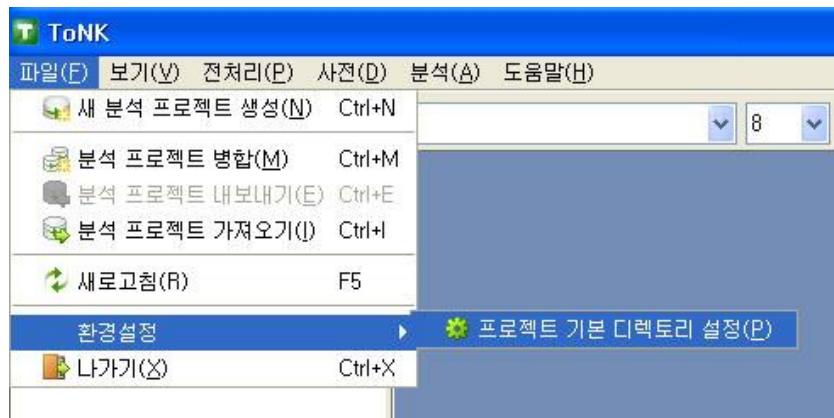
- a) 새 분석 프로젝트 생성 : 새로운 프로젝트를 생성합니다.
- b) 분석 프로젝트 병합 : 분석프로젝트를 병합해서 새로운 프로젝트를 생성합니다.



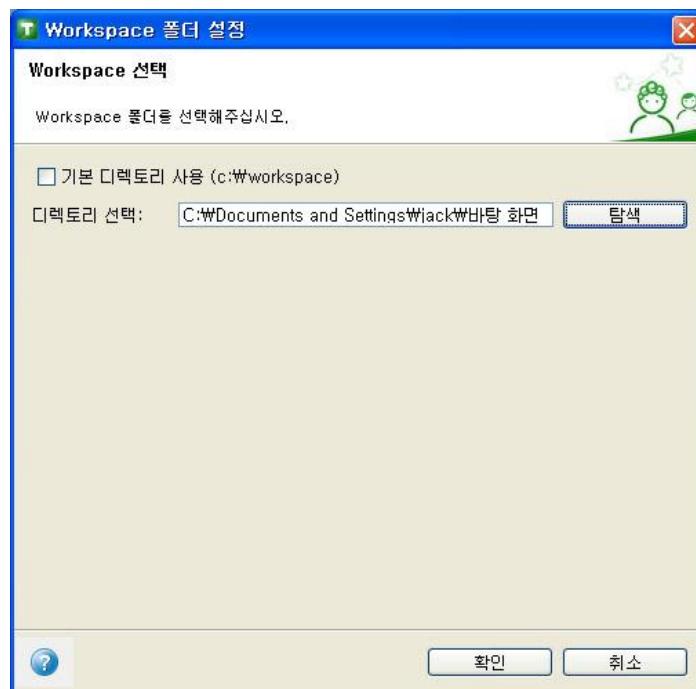
‘프로젝트명’에는 생성하고자 하는 프로젝트의 이름을 기입하고 ‘문서폴더 위치’에는 병합하고자 하는 프로젝트를 동시 클릭합니다. 하단의 ‘날짜’는 병합하는 문서들의 문서 생성일을 별도로 지정해주는 것으로 개시일부터 종료일 사이에 생성된 문서만 병합이 됩니다.

- c) 분석 프로젝트 내보내기 : 분석한 프로젝트를 저장합니다.
- d) 분석 프로젝트 가져오기 : 분석된 프로젝트를 불러옵니다.
- e) 새로고침 : 작업공간에 있는 모든 프로젝트를 다시 읽습니다.

(e-1) 파일 – 환경설정 – 프로젝트 기본 디렉토리 설정을 클릭합니다.



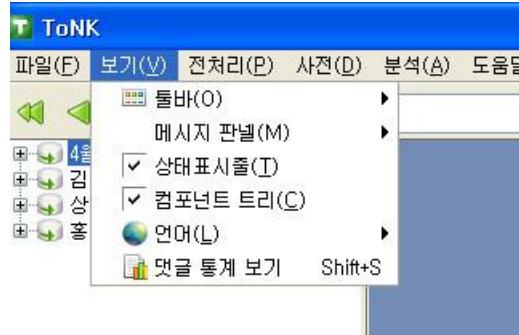
(e-2) 기본 디렉토리(C://Workspace) 이외에 사용자가 임의로 저장하고 싶은 곳이 있다면 '기본 디렉토리 사용'의 체크마크를 해지합니다. 이후 <탐색>을 클릭하여 저장할 곳을 지정한 후 확인 버튼을 클릭합니다.



- f) 환경 설정 : 분석 프로젝트의 기본 디렉토리를 설정합니다.
- g) 나가기 : TONK 를 종료합니다.

2. 보기

화면 구성을 전환할 수 있는 기능을 모아두었습니다.



a) 툴바

하부메뉴로 표준 도구가 있으며 체크마크가 있으면 화면에 해당 기능이 나타나며 체크마크가 없으면 화면에 해당 기능이 사라집니다.

b) 메시지 패널

하부메뉴로 콘솔 패널과 속성 패널이 있으며 체크마크가 있으면 화면에 해당 기능이 나타나며 체크마크가 없으면 화면에 해당 기능이 사라집니다.

c) 상태 표시줄

체크마크가 있으면 하단 화면에 상태표시줄이 나타나며 체크마크가 없으면 화면 상태표시줄이 사라집니다. 체크마크의 유무는 보기의 상태표시줄을 클릭하면 됩니다.

d) 컴포넌트 트리

체크마크가 있으면 화면에 컴포넌트 트리가 나타나며 체크마크가 없으면 화면에 컴포넌트 트리가 사라집니다. 체크마크의 유무는 보기의 컴포넌트 트리를 클릭하면 됩니다.

e) 언어

한국어와 영어가 있으며 메뉴막대의 언어를 변경할 수 있습니다.

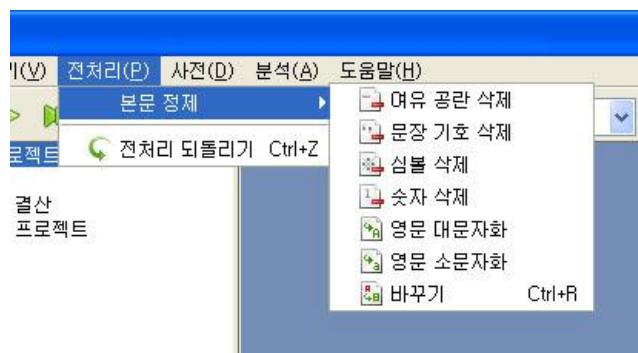
3. 전처리

분석을 시작하기 전에 실행하는 기능을 모아두었습니다.



a) 본문 정제

본문 정제에 관한 총 7 가지 하부메뉴가 있습니다.



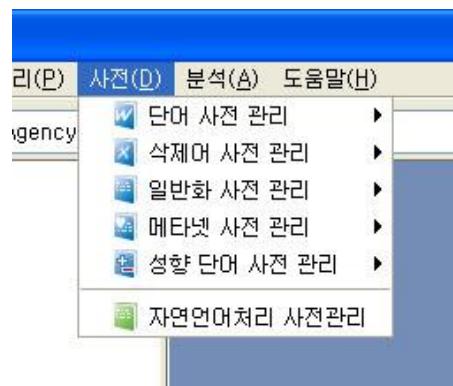
- (a-1) 여유 공란 삭제 : 2 칸 이상으로 띄어진 공란을 원문에서 삭제합니다.
- (a-2) 문장 기호 삭제 : 따옴표나 느낌표 등 문장 기호를 원문에서 삭제합니다.
- (a-3) 심볼 삭제 : 목욕탕이나 당구장 표시 등 심볼을 원문에서 삭제합니다.
- (a-4) 숫자 삭제 : 조사와 같이 쓰이지 않고 단독으로 쓰인 숫자를 원문에서 삭제합니다.
- (a-5) 영어 대문자화 : 원문에 있는 알파벳을 모두 대문자화 시킵니다.
- (a-6) 영어 소문자화 : 원문에 있는 알파벳을 모두 소문자화 시킵니다.
- (a-7) 바꾸기 : 원문에 있는 특정 단어를 다른 단어로 변환시킵니다.

b) 전처리 되돌리기

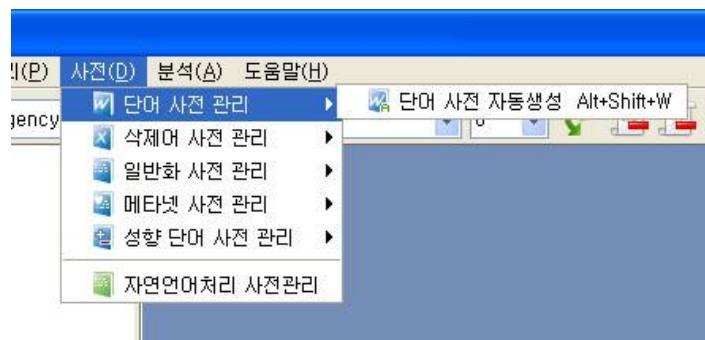
문서에 대한 기존 전처리 명령들을 취소하고 프로젝트 생성시의 상태로 되돌립니다.

4. 사전

분석에 필요한 사전을 모아두었습니다.



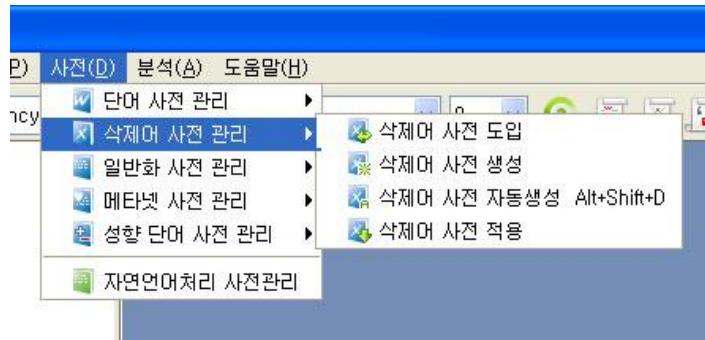
a) 단어 사전 관리



(a-1) 단어 사전 자동생성 : 단어사전을 프로그램을 통해 생성합니다.

b) 삭제어 사전 관리

삭제어 사전에 관한 총 4 가지 하부메뉴가 있습니다.



(b-1) 삭제어 사전 도입 : 기존에 생성되었던 삭제어 사전의 파일을 컴퓨터에서 불러옵니다.

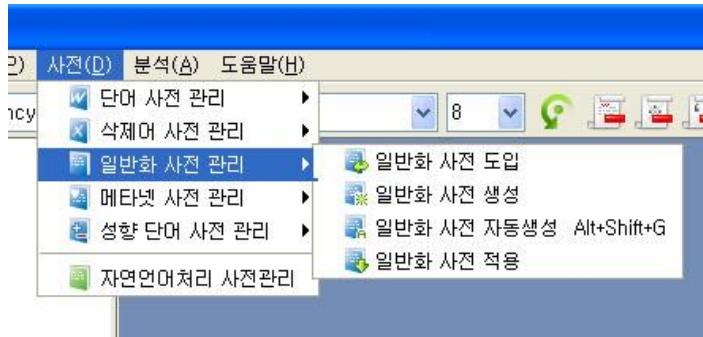
(b-2) 삭제어 사전 생성 : 사용자가 직접 단어를 추가/삭제하기 위하여 초기화된 삭제어 사전을 생성합니다.

(b-3) 삭제어 사전 자동생성 : 삭제어 사전을 프로그램을 통해 생성합니다.

(b-4) 삭제어 사전 적용 : 생성한 삭제어 사전을 프로젝트에 적용합니다.

c) 일반화 사전 관리

일반화 사전에 관한 총 4 가지 하부메뉴가 있습니다.



(c-1) 일반화 사전 도입 : 기존에 생성되었던 일반화 사전의 파일을 컴퓨터에서 불러옵니다.

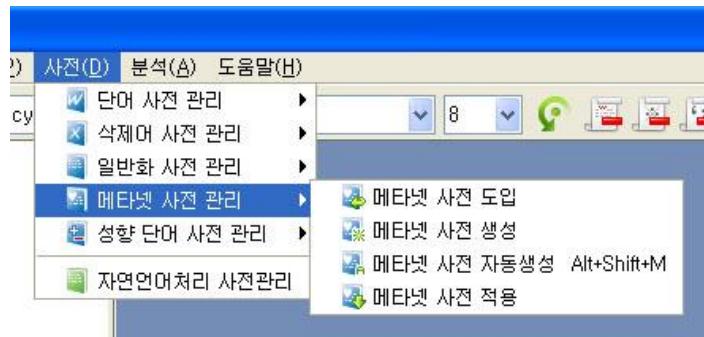
(c-2) 일반화 사전 생성 : 사용자가 직접 단어를 추가/삭제하기 위하여 초기화된 일반화 사전을 생성합니다.

(c-3) 일반화 사전 자동생성 : 일반화 사전을 프로그램을 통해 생성합니다.

(c-4) 일반화 사전 적용 : 생성한 일반화 사전을 프로젝트에 적용합니다.

d) 메타넷 사전 관리

메타넷 사전에 관한 총 4 가지 하부메뉴가 있습니다.



(d-1) 메타넷 사전 도입 : 기존에 생성되었던 메타넷 사전의 파일을 컴퓨터에서 불러옵니다.

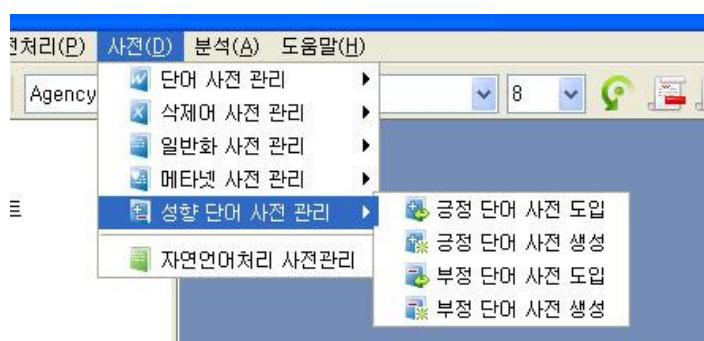
(d-2) 메타넷 사전 생성: 사용자가 직접 단어를 추가/삭제하기 위하여 초기화된 메타넷 사전을 생성합니다.

(d-3) 메타넷 사전 자동생성 : 메타넷 사전을 프로그램을 통해 생성합니다.

(d-4) 메타넷 사전 적용 : 생성한 메타넷 사전을 프로젝트에 적용합니다.

e) 성향 단어 사전 관리

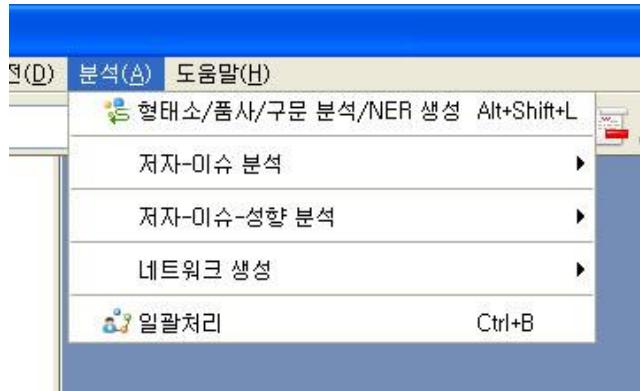
성향 단어사전에 관한 총 4 가지 하부메뉴가 있습니다.



- (e-1) 긍정 단어 사전 도입 : 기존에 생성되었던 긍정 단어사전의 파일을 컴퓨터에서 불러옵니다.
- (e-1) 긍정 단어 사전 생성 : 사용자가 직접 단어를 추가/삭제하기 위하여 초기화된 긍정 단어사전을 생성합니다.
- (e-1) 부정 단어 사전 도입 : 기존에 생성되었던 부정 단어사전의 파일을 컴퓨터에서 불러옵니다.
- (e-1) 부정 단어 사전 생성 : 사용자가 직접 단어를 추가/삭제하기 위하여 초기화된 부정 단어사전을 생성합니다.

5. 분석

본격적으로 분석을 실행하는 기능을 모아두었습니다.



a) 형태소/품사/구문분석/NER 생성

분석 문서의 형태소, 품사, 구문분석, NER 정보를 생성합니다.

b) 저자-이슈 분석

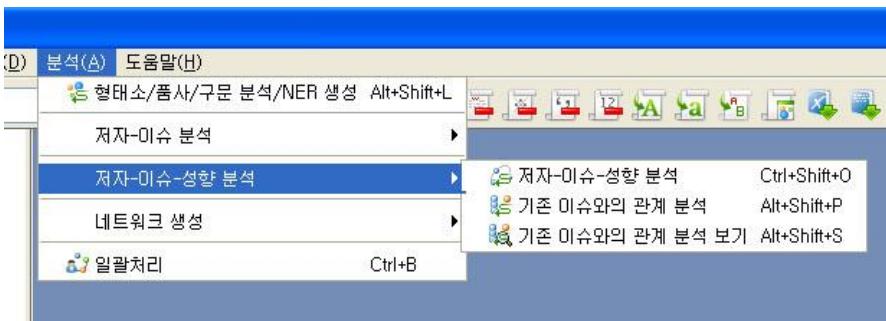


(b-1) 저자-이슈 분석 : 토픽을 추출해냅니다.

(b-2) 기존 이슈와의 관계 분석 : ‘저자-이슈 분석’을 글자로 한 해당 프로젝트의 ‘저자-이슈 분석’ 결과와 다른 분석 프로젝트의 ‘저자-이슈 분석’과의 연관관계를 찾아냅니다.

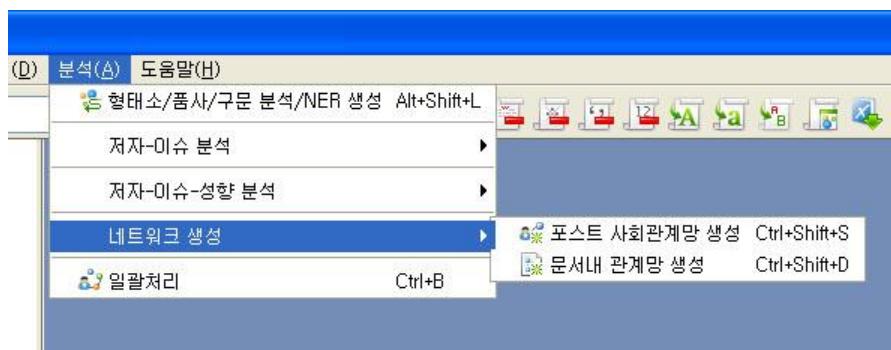
(b-3) 기존 이슈와의 관계 분석 보기 : ‘저자-이슈와의 관계 분석’ 결과를 확인합니다.

c) 저자-이슈-성향 분석



- (c-1) 저자-이슈-성향 분석 : 토픽과 토픽의 성향을 추출해냅니다.
- (c-2) 기존 이슈와의 관계 분석 : '저자-이슈-성향 분석'을 골자로 한 해당 프로젝트의 '저자-이슈-성향 분석' 결과와 다른 분석 프로젝트의 '저자-이슈-성향 분석'과의 연관관계를 찾아냅니다.
- (c-3) 기존 이슈와의 관계 분석 보기 : '기존 이슈와의 관계 분석' 결과를 확인합니다.

d) 네트워크 생성

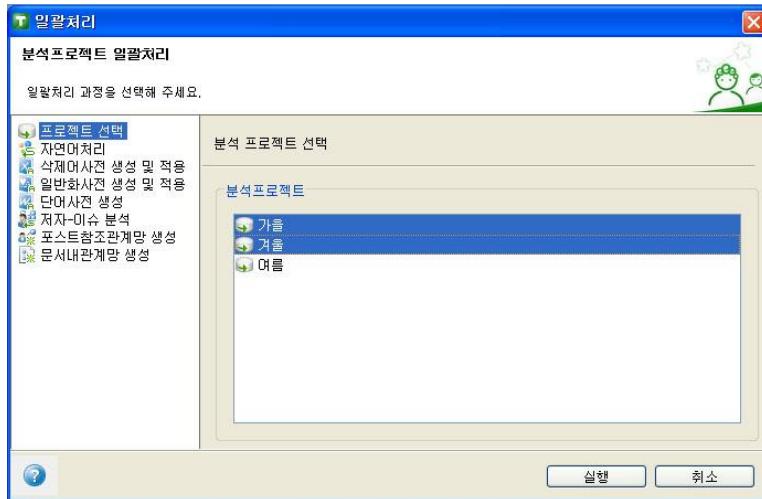


- (d-1) 포스트 사회 관계망 생성 : 분석 문서와 저자 이슈 간의 상관관계를 네트워크로 구성합니다.
- (d-2) 문서내 관계망 생성 : 분석 문서 내부의 핵심 단어 간 상관관계를 네트워크로 구성합니다.

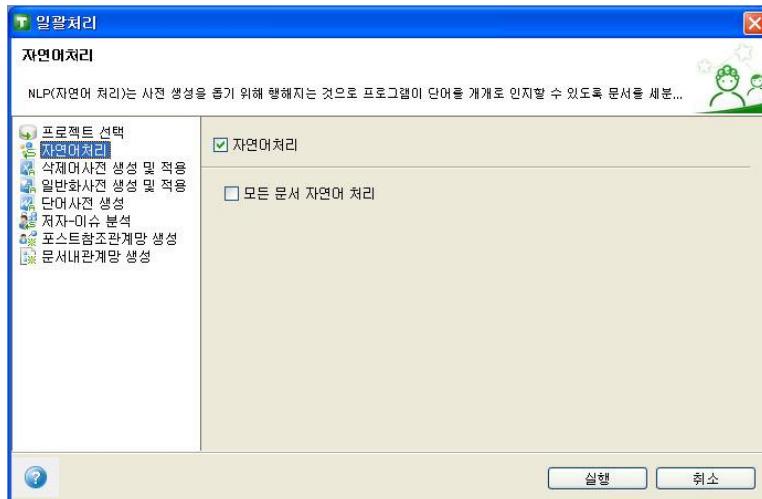
e) 일괄처리

여러가지 프로젝트에서 동일한 분석을 실행하고자 할 시 개별적으로 실행해야 하는 부담을 덜고 한꺼번에 처리해주는 기능입니다. 보다 자세한 설명은 해당 챕터를 참고하시기 바랍니다.

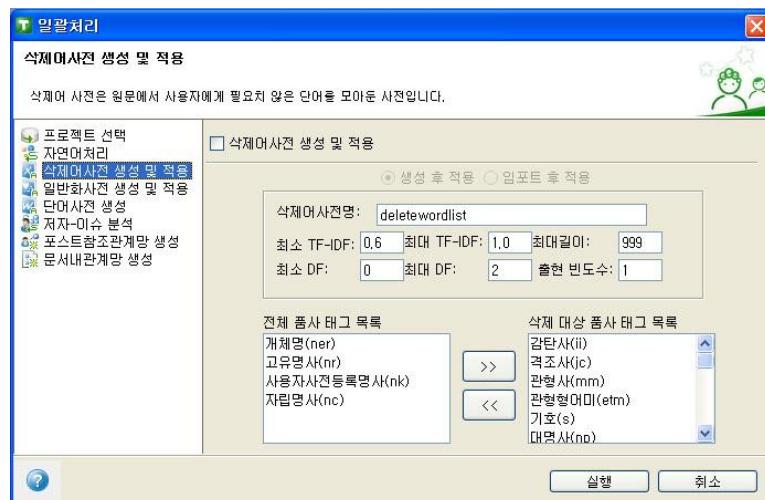
(e-1) 프로젝트 선택 : 일괄 처리할 프로젝트를 동시에 클릭합니다.



(e-2) 자연어 처리 : '자연어 처리'에 클릭을 하면 자연어 처리를 적용합니다. '모든 문서 자연어 처리'에 체크를 하면 이미 자연어 처리가 되어있는 문서일지라도 모든 문서를 대상으로 자연어 처리를 실행합니다. (참고- III. C. 1. 자연어처리, page)+

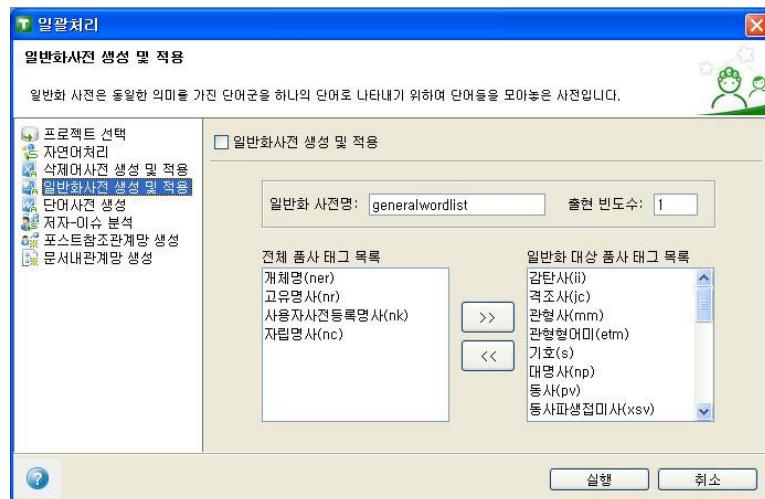


(e-3) 삭제어 사전 생성 및 적용 : '삭제어 사전 생성 및 적용'에 체크를 하면 직접 대화상자 내에서 생성하여 적용할 수 있으며 또한 기존에 생성된 삭제어 사전을 불러오기 하여 적용할 수도 있습니다. 아래의 설정을 사용자가 조절할 수 있습니다. (참고 – III. D.1. 삭제어 사전, page)

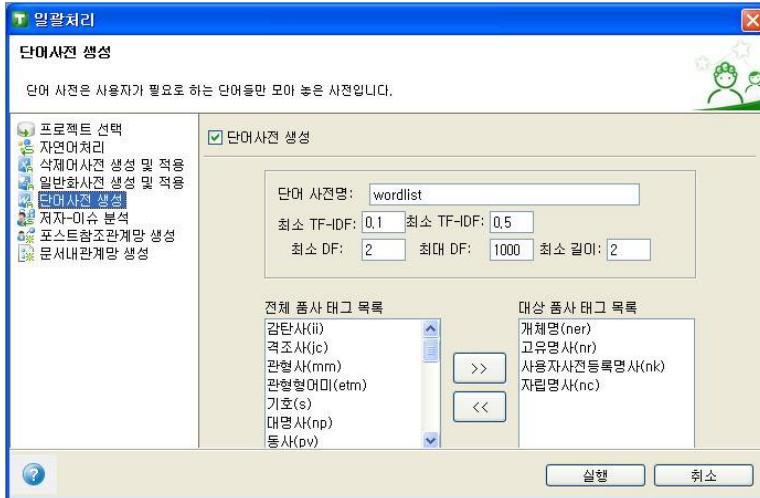


(e-4) 일반화사전 생성 및 적용 : ‘일반화 사전 생성 및 적용’에 체크를 하면 일반화 사전이 아래의 설정대로 생성하여 적용이 됩니다. 설정은 사용자가 자유로 조절할 수 있습니다.

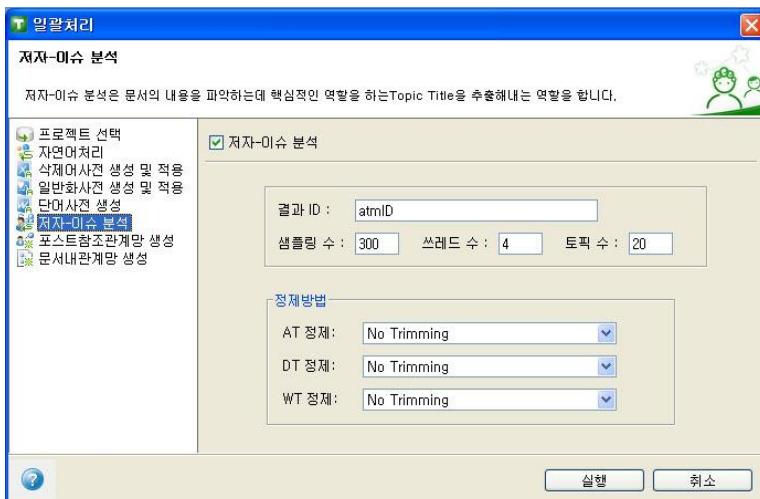
(참고- III. D.3. 일반화 사전, page 99)



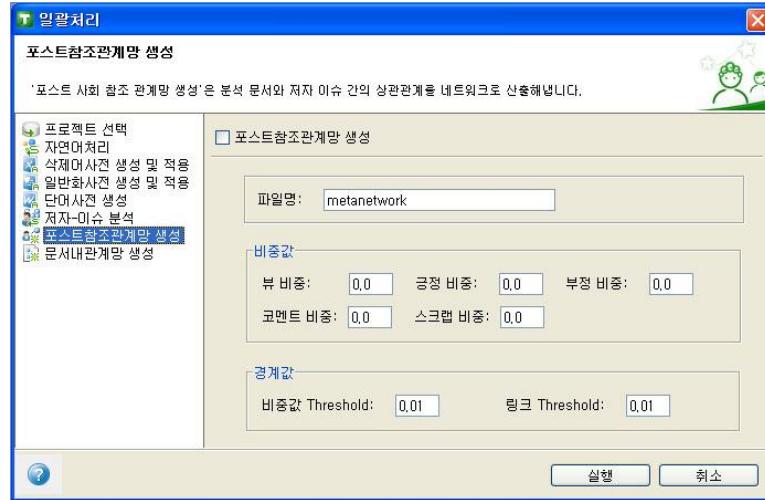
(e-5) 단어사전 생성 : ‘단어사전 생성’에 체크를 하면 단어 사전이 아래의 설정대로 생성됩니다. 설정은 사용자가 자유로 조절할 수 있습니다. (참고- III.E.1. 단어사전, page 125)



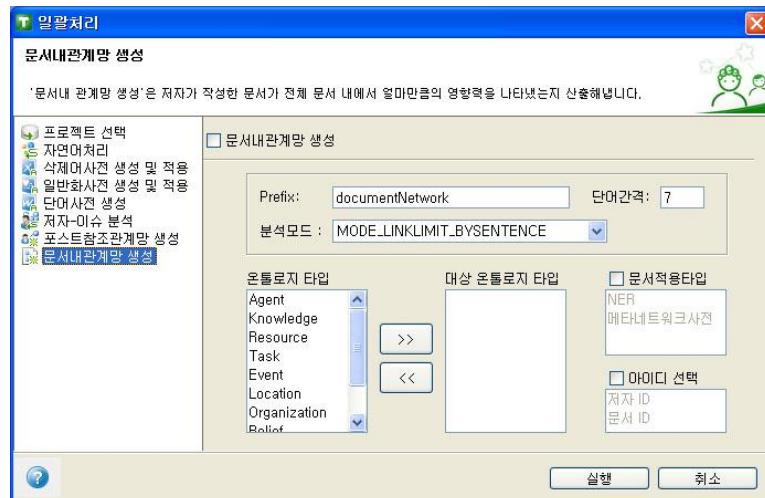
(e-6) 저자-이슈 분석 : '저자-이슈 분석'에 체크를 하면 아래의 설정대로 '저자-이슈 분석'이 실행됩니다. (III. E. 1. 저자-이슈 분석, page 135)



(e-7) 포스트 참조 관계망 생성 : '포스트참조관계망 생성'에 체크를 하면 아래의 설정대로 '저자-이슈 분석'이 실행됩니다. (III.J.1. 포스트 참조관계망 생성, page 199)

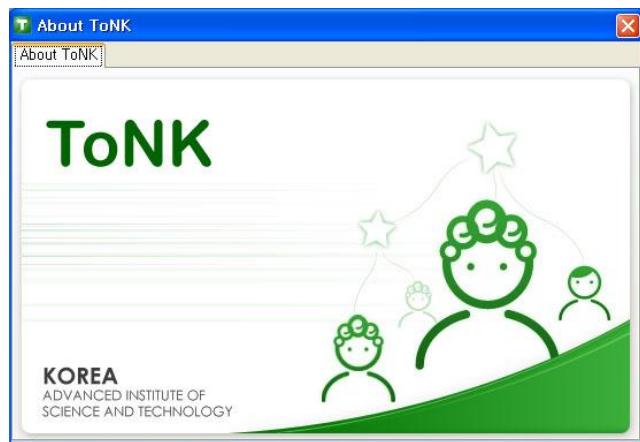


(e-8) 문서내 관계망 생성 : ‘문서내 관계망 생성’에 체크를 하면 아래의 설정대로 ‘저자-이슈 분석’이 실행됩니다. (IV. E. 1. 문서내 관계망 생성, page 326)



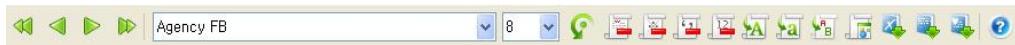
6. 도움말

TONK 의 관련 정보가 나타납니다.



C. 표준 도구

자주 쓰이는 주요 기능을 빠른 아이콘 형태로 표시하여 쉽게 사용할 수 있습니다.



1. 첫번째 텍스트 파일



분석 문서에서 목록의 맨 처음에 나타난 문서를 새 탭으로 생성합니다.



2. 이전 텍스트 파일

분석 문서에서 기존에 열려있는 문서탭의 바로 이전의 문서를 새 탭으로 생성합니다.



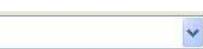
3. 다음 텍스트 파일

분석 문서에서 기존에 열려있는 문서탭의 바로 다음의 문서를 새 탭으로 생성합니다.



4. 마지막 텍스트 파일

분석 문서에서 목록의 맨 마지막에 나타난 문서를 새 탭으로 생성합니다.



5. 글꼴

글꼴을 변경할 수 있습니다.



글꼴 크기를 변경할 수 있습니다.



7. 전처리 되돌리기

문서에 대한 기존 전처리 명령들을 취소하고 프로젝트 생성시의 상태로 되돌립니다



8. 여유 공란 삭제

2칸 이상으로 띄어진 공란을 원문에서 삭제합니다.

9. 문장 기호 삭제



따옴표나 느낌표 등 문장 기호를 원문에서 삭제합니다.

10. 심볼 삭제



목욕탕이나 당구장 표시 등 심볼을 원문에서 삭제합니다.

11. 숫자 삭제



조사와 같이 쓰이지 않고 단독으로 쓰인 숫자를 원문에서 삭제합니다.

12. 영어 대문자화



원문에 있는 알파벳을 모두 대문자로 변환합니다.

13. 영어 소문자화



원문에 있는 알파벳을 모두 소문자로 변환합니다.

14. 바꾸기



원문에 있는 특정 단어를 다른 단어로 변환합니다.

15. 형태소/품사/구문분석/NER 생성



분석 문서의 형태소, 품사, 구문분석, NER 을 생성합니다.

16. 삭제어 사전 적용



생성한 삭제어 사전을 프로젝트에 적용합니다.

17. 일반화 사전 적용



생성한 일반화 사전을 프로젝트에 적용합니다.

18. 메타넷 사전 적용 

생성한 메타넷 사전을 프로젝트에 적용합니다.

19. 도움말 

TONK 의 관련 정보가 나타납니다.

D. 컴포넌트 트리

사용자가 생성하고 분석을 진행/종료한 프로젝트들이 보여집니다.

프로그램을 처음 실행하였을 때

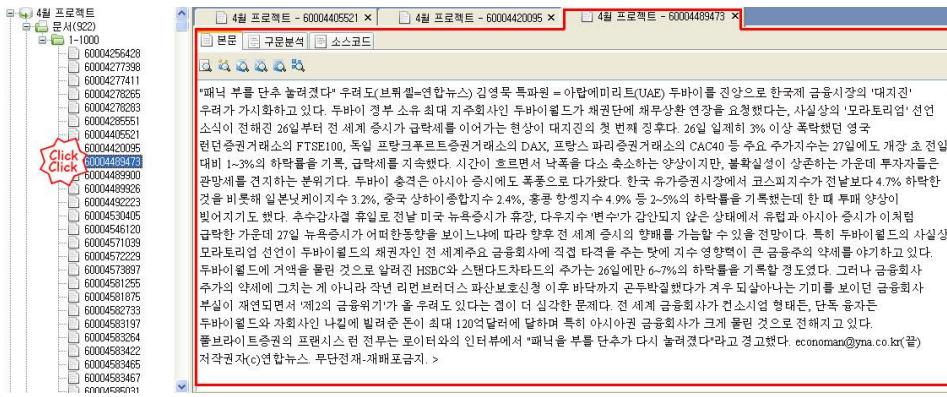
기존에 생성하였던 프로젝트가 있을 때

TONK를 실행시키면 컴포넌트 트리의 초기화면에는 빈 공란이 나타납니다. 하지만 기존에 프로젝트를 생성하였다면 TONK의 시작과 함께 컴포넌트 트리에 프로젝트가 보여집니다. 분석이 완료된 임의의 프로젝트를 더블 클릭하면 이를 구체적으로 분석하기 위한 4개의 주요 요소(문서, 저자, 네트워크, 사전)와 결과(저자-이슈-성향 결과 혹은 저자-이슈 분석결과)가 하위 폴더의 형태로 존재합니다. 이 외에 새로운 프로젝트를 불러오거나(방법: **메뉴막대 — 파일 — 분석 프로젝트 가져오기**), 기존 문서를 삭제할 수 있습니다(방법: **마우스 오른쪽 버튼 — 삭제**).

생성한 모든 프로젝트는 사용자가 지정한 작업폴더(최초 설정- C://Workplace, page 29, (d-2) 참조)에 저장되며 프로그램 종료 후 재부팅 시 컴포넌트 트리에 자동으로 보여집니다. 문서를 삭제하는 경우에는 컴포넌트 트리의 목록에서만 사라지는 것이 아니라 사용자가 지정한 작업폴더에서도 함께 사라지게 됩니다.

E. 작업창

TONK의 주요 활동 무대라고 할 수 있는 작업창은 전체 화면의 우측에 넓게 자리잡고 있으며 컴포넌트 트리 및 메뉴막대에서 실행한 명령을 수행하는 창입니다. 컴포넌트 트리에서 프로젝트를 개괄적으로 관리했다면 작업창은 컴포넌트 트리에서 나열되었던 문서를 실제로 열람하고 세부 분석도 할 수 있습니다. 컴포넌트 트리에서 사용자가 열람하고자 하는 문서를 더블 클릭 하면 작업창에는 해당 탭이 생성되고 여러 문서를 열람 할 경우 탭들은 우측으로 나열됩니다. 탭의 상단에는 해당 문서의 프로젝트 명과 아이디가 표시됩니다.

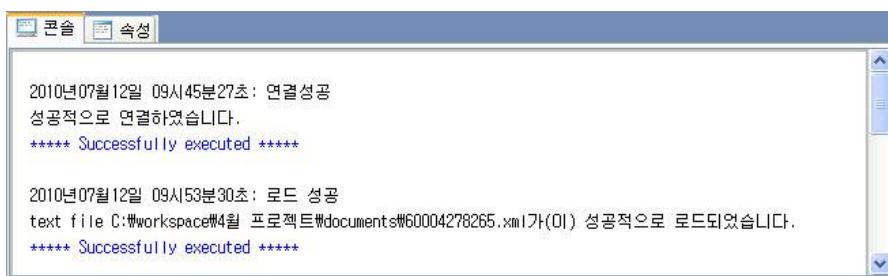


F. 메시지 판넬

1. 콘솔

TONK에서 실행되고 있는 명령의 시작과 완료를 보고하거나 명령 과정의 성공여부가 표시됩니다. 첫째 줄에는 명령 시간이 나타나며 둘째 줄에는 명령의 종류가, 셋째 줄에는 명령의 성공여부가 나타납니다.

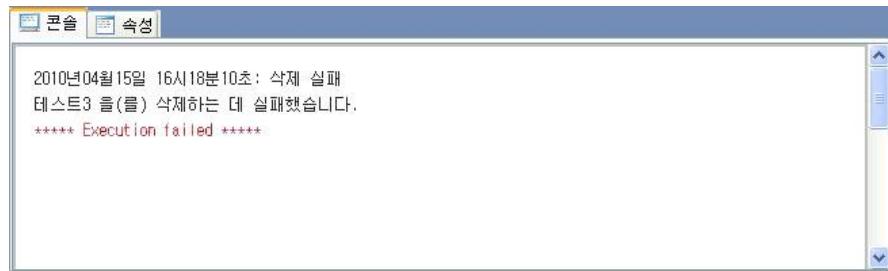
- a) 새 프로젝트를 생성하는 명령을 이상 없이 완료하였을 때



The screenshot shows the 'Console' tab of the TONK interface. The output window displays the following text:

```
2010년07월12일 09시45분27초: 연결성공  
성공적으로 연결하였습니다.  
***** Successfully executed *****  
  
2010년07월12일 09시53분30초: 로드 성공  
text file C:\workspace\4월 프로젝트\documents\60004278265.xml(이) 성공적으로 로드되었습니다.  
***** Successfully executed *****
```

- b) '테스트 3.txt'를 삭제하는 명령을 실행하지 못했을 때

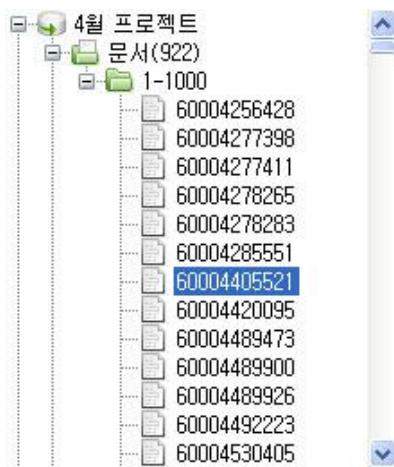


The screenshot shows the 'Console' tab of the TONK interface. The output window displays the following text:

```
2010년04월15일 16시18분10초: 삭제 실패  
테스트3 을(를) 삭제하는 데 실패했습니다.  
***** Execution failed *****
```

2. 속성

컴포넌트 트리의 임의의 폴더나 문서를 클릭하면 해당 항목의 문서명, 위치, 생성일, 크기 등과 같은 구체적인 정보가 나타납니다.



임의의 문서(6000448900)를 클릭했을 때

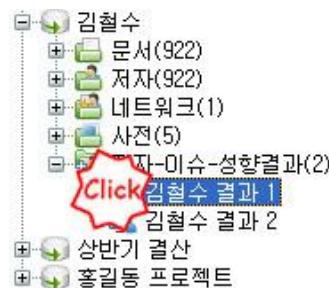
Info	
문서명	60004405521
위치	C:\workspace\4월 프로젝트\documents\60004405521.xml
작성자	연합뉴스
생성일	2009-10-15 16:46:21
크기	31,321 bytes
댓글 수	0
본문 문장 수	0

속성에서 나타나는 해당 문서(6000448900)의 관련 정보

사용자가 선택한 항목에 따라 보여지는 관련 문서 항목은 각각 다르며, 특히 <III.A.2.b>저자, page 57>에서 다뤄질 동일한 저자명의 정보를 구분 하는 데에도 유용하게 쓰입니다.

G. 상태표시줄

컴포넌트 트리에서 임의의 항목을 클릭하면 상태표시줄에 해당 주소가 나타납니다. 문서의 저장된 위치를 쉽게 알 수 있습니다.



‘저자-이슈-성향 결과’의 ‘김철수 결과 1’을 클릭

C:\workspace\김철수\Atoms\김철수 결과 1\김철수 결과 1.xml

‘김철수 결과 1’이 저장된 주소

H. 기타 일러두기

컴포넌트 트리에서 등장하는 용어를 정리한 표입니다.

-ID	프로그램이 해당 정보에 임의로 부여하는 숫자로써 사용자는 해당 정보의 이름을 가지고 인식할 수 있지만 프로그램은 아이디를 가지고 문서를 인식합니다. 즉, 프로그램만의 언어입니다. 예를 들면 저자에는 저자 아이디가, 네트워크에는 네트워크 아이디가 할당됩니다.
-명	해당 정보의 이름을 의미하며, 문서에는 문서명이, 저자에는 저자명이, 결과에는 결과명이, 네트워크에는 네트워크 명이 있습니다.
문서명	문서의 이름입니다.
저자명	문서를 작성한 사람이나 기관의 이름입니다. 주목할 점은 문서명이나 문서아이디, 저자 아이디 등은 오로지 하나씩 존재하지만, 저자명은 문서에서 추출한 저자의 이름이기 때문에 저자 본인은 각자 다른 사람이더라도 이름은 동일할 수 있다는 것입니다. 따라서 저자명은 중복이 가능합니다.
저자 ID	프로그램이 문서에 대한 각각의 저자에 임의로 부여한 숫자입니다.
결과명	저자-이슈 분석이나 저자-이슈 성향 분석을 실행한 후 생성되는 결과문서의 이름입니다.
결과 ID	프로그램이 임의로 부여한 숫자가 아니라 사용자가 저자-이슈 분석이나 저자-이슈-성향 분석을 실행 하는 과정에서 지정한 아이디를 말합니다. 결론적으로 결과명과 결과 아이디는 같습니다.
네트워크명	포스트 참조관계망이나 문서내 관계망을 실행한 후 생성되는 네트워크의 이름입니다.
네트워크 ID	프로그램이 임의로 부여한 숫자가 아니라 사용자가 포스트 참조관계망이나 문서내 관계망을 실행하는 과정에서 지정한 아이디를 말합니다. 결론적으로 네트워크명과 네트워크 아이디는 같습니다.

Chapter III 첫번째 분석 - 문서의 내용 및 메타 정보

문서의 내용을 분석하는 것은 분석 대상인 문서들이 어떤 화제를 가장 비중 있게 다루었는지 알아보는 것입니다. 메타 정보를 분석하는 것은 문서의 내용을 분석한 결과를 토대로 하여 화제(이슈), 저자의 영향력, 문서 간 관계 등을 파악하는 것입니다. 특히 메타 정보를 확인하는 것은 ORA라는 프로그램을 별도로 이용합니다. 다음은 문서 내용 및 메타 정보 분석에 대한 과정입니다.

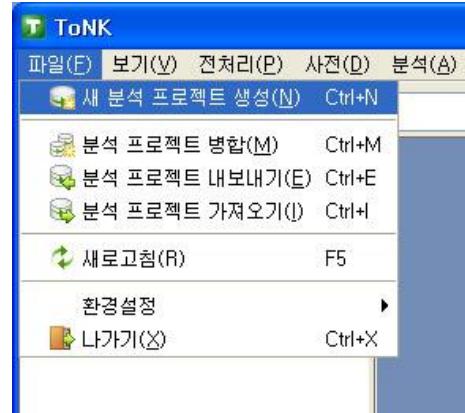
A. 프로젝트 생성하기

새로운 분석을 시작하기 위해 ‘새 분석 프로젝트’를 생성합니다. 워드나 한글 프로그램에서 새 문서를 생성하듯이 TONK에서는 새 분석 프로젝트를 생성합니다. 프로젝트는 총 4 개의 하위 폴더로 구성되었으며, 분석 대상인 문서가 모여 있는 ‘문서’, 문서의 저자들이 모여 있는 ‘저자’, 문서, 저자, 이슈의 관계를 나타내는 네트워크, 사전을 모아놓은 ‘사전’, 저자-이슈 분석과 저자-이슈-성향 분석의 결과가 모여 있는 ‘저자-이슈 결과’와 ‘저자-이슈-성향 결과’가 있습니다. 생성된 프로젝트는 파일로 자동으로 저장이 되며 이를 임의로 삭제하지 않는 이상 추후 TONK를 실행하였을 때 컴포넌트 트리에 자동으로 나타납니다.

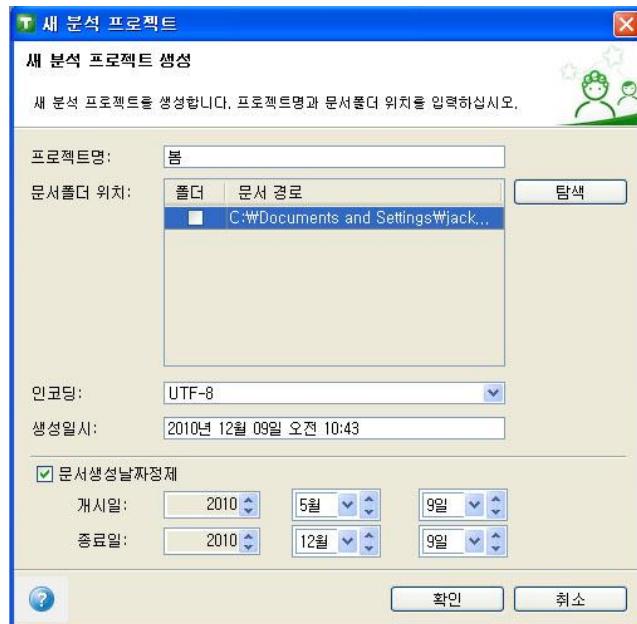
Comment [j1]: 모르겠음

1. 새 분석 프로젝트 생성하기

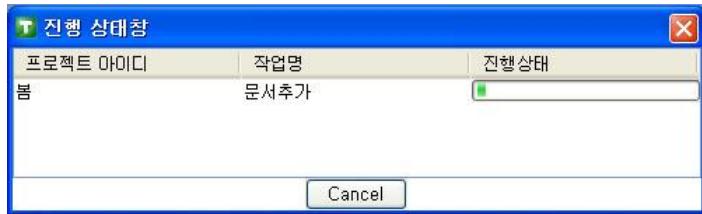
- TONK 프로그램을 실행하면 컴포넌트 트리와 작업창에 빈 화면이 나타납니다. 새 분석 프로젝트를 실행하려면 <파일>에서 ‘새 분석 프로젝트 생성’을 누릅니다.



- b) ‘새 분석 프로젝트’ 대화상자가 나타나면 ‘프로젝트 명’에는 본인이 원하는 제목을 기입합니다. ‘문서 폴더 위치’에는 우측의 <탐색>을 클릭하여 분석하고자 하는 문서가 위치해 있는 폴더를 지정한 후 <확인>을 클릭합니다. ‘문서생성날짜정제’에 체크하여 개시일과 종료일을 설정하면 선택된 문서 폴더 안에서 지정한 기간 안에 생성된 문서가 수집됩니다.



- c) ‘진행 상태창’ 대화상자가 나타나면 연두색 막대가 채워질 때까지 기다립니다. 이때 <Cancel>을 클릭하면 ‘자연어 처리’가 중단됩니다. 이 과정에서는 분석 프로젝트의 하위폴더를 생성하고 분석 대상을 추출합니다. 이때 <취소>를 클릭하면 새 분석 프로젝트 생성이 중단됩니다.



- d) 작업이 완료되면 컴포넌트 트리에 프로젝트가 생성됩니다. 생성된 프로젝트를 더블 클릭하면 이를 구성하고 있는 4 가지 (문서, 저자, 네트워크, 사전) 하위 폴더가 나타납니다.



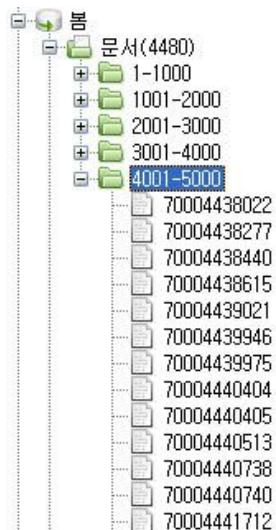
- e) 생성된 프로젝트는 파일의 형태로 사용자가 지정한 작업폴더(최초 설정- C://Workplace, page 29, (d-2) 참조)에 저장되며 하단의 속성에서 저장된 위치 등 상세정보를 확인할 수 있습니다.
- f) 명령이 성공적으로 수행되었다면 속성에는 ***** Successfully executed ***** 가 표기됩니다.

2. 새 분석 프로젝트 생성 처리 결과 보기

생성된 분석 프로젝트는 컴포넌트 트리와 작업창에서 확인합니다. 프로젝트를 더블 클릭하면 총 4 개의 하위폴더가 보여지며 새 분석 프로젝트를 바로 생성한 상태에서는 2 개의 하위 폴더(문서, 저자)만 열람이 가능합니다.

a) 문서

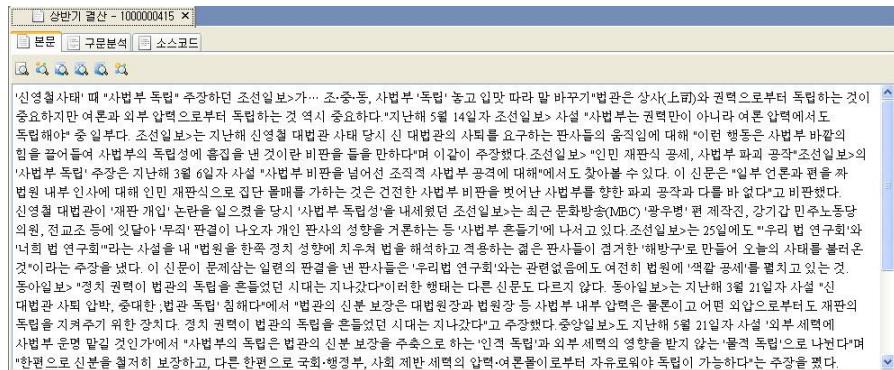
‘문서’는 사용자가 새 프로젝트를 생성할 시 분석 대상으로 선정했던 문서를 모아놓은 폴더입니다. 새 분석 프로젝트를 생성한 후 컴포넌트 트리 및 작업창에서 바로 열람이 가능합니다.



‘문서’를 더블클릭 하면 문서의 개수가 많은 경우를 감안하여 1000 단위¹로 나누어 생성된 하위 폴더가 차례로 나타납니다. 나열된 문서들은 문서명으로 보여지게 되며 괄호 안의 숫자는 문서의 개수를 의미합니다.

이 중 하나의 문서를 클릭하면 메시지 패널의 속성에서 관련 정보(문서명, 위치, 작성자, 생성일 등)가 표시되고 상태표시줄에는 문서가 저장된 위치가 보여집니다. 문서를 더블 클릭하면 작업창에 문서의 내용이 보여집니다. 수집된 된 문서를 삭제하고 싶을 경우에는 해당 사전에 마우스 오른쪽 버튼을 눌러 삭제를 클릭합니다.

문서에 있는 임의의 문서를 더블 클릭했을 때 작업창은 다음과 같습니다.



수집된 문서들의 원문을 새로 생성된 작업창에서 확인할 수 있습니다.

b) 저자

‘저자’는 ‘문서’로 수집된 문서의 저자를 모아놓은 폴더입니다. 새 분석 프로젝트를 생성한 후 컴포넌트 트리에서만 열람이 가능합니다. 이후 <III.F 저자-이슈 분석 실행하기, page 135>에서 다뤄질 ‘저자-이슈 분석’을 실행해야만 ‘저자’의 작업창을 열람할 수 있습니다. 괄호 안의 숫자는 총 저자의 개수를 의미합니다.

¹ 문서의 개수가 10,000이라면 1-1000, 1001-2000, 2001-3000 등과 같이 총 10 개의 하위폴더가 생성됩니다.



‘저자’를 더블 클릭하면 목록에서는 저자명¹이 나열됩니다. 만일 동일한 저자명이 목록에 나타날 경우, 이는 오류가 난 것이 아니라 단지 문서에 대한 추출된 저자명이 같을 뿐² 실제 저자는 다른 사람입니다. 바꾸어 말하면, 저자들은 각자 다른 사람이더라도 이름은 동일할 수 있기 때문에 저자명은 중복 되어서 나타날 수 있습니다.

‘저자 ID’란 프로그램이 각 저자에 임의로 부여한 숫자입니다. 만약 저자명이 같아서 내용파악이 어려울 경우 속성의 ‘저자 아이디’ 란을 참고하면 쉽게 이해할 수 있습니다.

목록 중 임의의 저자명 클릭하면 속성에 관련 정보(저자명, ID, 작성문서 수)가 표시되고 상태표시줄에는 문서가 저장된 위치가 나타납니다. 다시 임의의 저자명을 더블 클릭하면 작업창에는 저자에 대한 분석내용이 보여 지는데 이는 ‘저자-이슈 분석’을 실행한 후에만 나타납니다.

¹ 저자명은 문서를 작성한 사람이나 기관의 이름입니다.

² ‘저자’에서 다루는 정보는 동일한 저자명을 한 대 모아 분석하는 것이 아니라 각 문서에 대한 저자를 개별적으로 분석하는 것입니다.

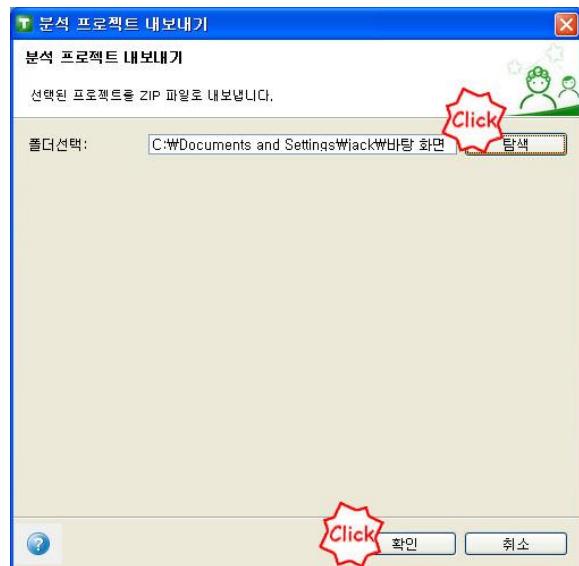
3. 분석 프로젝트 내보내기

TONK는 새 프로젝트를 만들과 동시에 사용자가 지정한 작업폴더(최초 설정- C://Workplace, page 29, (d-2) 참조)에 저장을 합니다. 하지만 사용자가 분석내용을 파일로 저장하고자 한다면 ‘분석 프로젝트 내보내기’를 사용합니다.

- a) <파일>에서 ‘분석 프로젝트 내보내기’를 클릭합니다.



- b) ‘분석 프로젝트 내보내기’ 대화상자가 나타나면 <탐색>을 클릭하여 저장할 장소를 선택한 후 <확인>을 누릅니다.



- c) 작업이 완료되면 사용자가 지정한 곳에 선택한 프로젝트가 생성된 것을 확인할 수 있습니다.

파일을 저장한 위치 및 파일 이름은 정확히 기억하는 것이 좋으므로 찾기 쉬운 곳에 저장해 둡니다. 또한 <새 폴더>를 클릭하여 관련된 문서를 한 폴더에 저장해 둔다면 문서 관리하기가 수월해 집니다. 작업이 완료되면 지정한 위치에 선택한 프로젝트가 생성된 것을 확인할 수 있습니다.

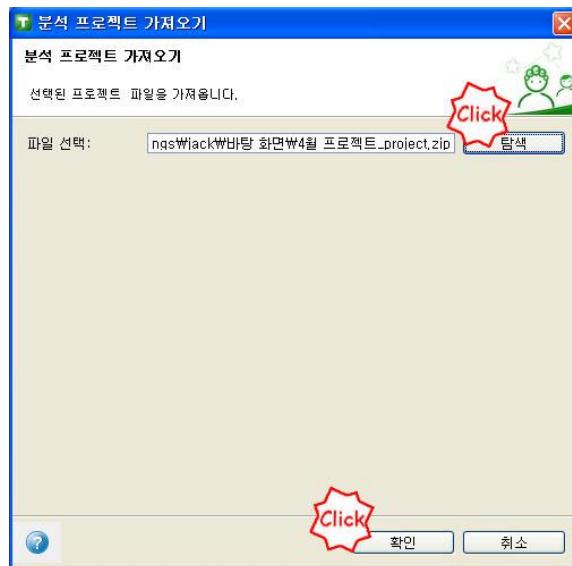
4. 분석 프로젝트 가져오기

저장되어 있는 TONK 파일을 열기 위해 ‘분석 프로젝트 가져오기’를 사용합니다.

- a) <파일>에서 ‘분석 프로젝트 가져오기’를 클릭합니다.



- b) ‘분석 프로젝트 가져오기’ 대화상자에서 <탐색>을 클릭하여 TONK으로 불러올 파일을 찾아 선택한 후 <확인>을 클릭합니다. 불러온 파일은 Zip 형태이어야 합니다.



- c) 작업이 완료되면 컴포넌트 트리에 선택한 프로젝트가 생성된 것을 확인할 수 있습니다.

B. 본문 정제 실행하기

1. 본문 정제

새 프로젝트를 생성한 후 문서의 본문을 정렬해서 보고자 할 때 사용합니다. 다음과 같이 본문 정제에는 총 7 가지 종류가 있습니다.

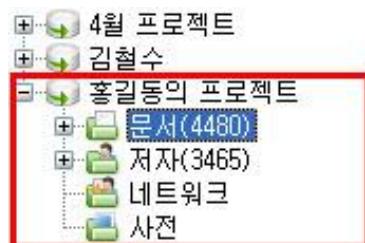
종 류	기 능
여유 공란 삭제	2칸 이상으로 띄어진 공란을 원문에서 삭제
문장 기호 삭제	따옴표나 느낌표 등 문장 기호를 원문에서 삭제
심볼 삭제	목욕탕이나 당구장 표시 등 심볼을 원문에서 삭제
숫자 삭제	조사와 같이 쓰이지 않고 단독으로 쓰인 숫자를 원문에서 삭제
영어 대문자화	원문에 있는 알파벳을 모두 대문자로 변환
영어 소문자화	원문에 있는 알파벳을 모두 소문자로 변환
바꾸기	원문에 있는 특정 단어를 다른 단어로 변환

본문 정제는 분석하는데 있어서 필수적인 과정은 아니나 사용자의 필요에 따라 원문을 정제해서 보고자 할 때 유용합니다. 본문 정제는 컴포넌트 트리의 ‘문서’에 있는 모든 문서¹에 적용됩니다. 다음은 본문 정제를 실행하는 과정입니다.

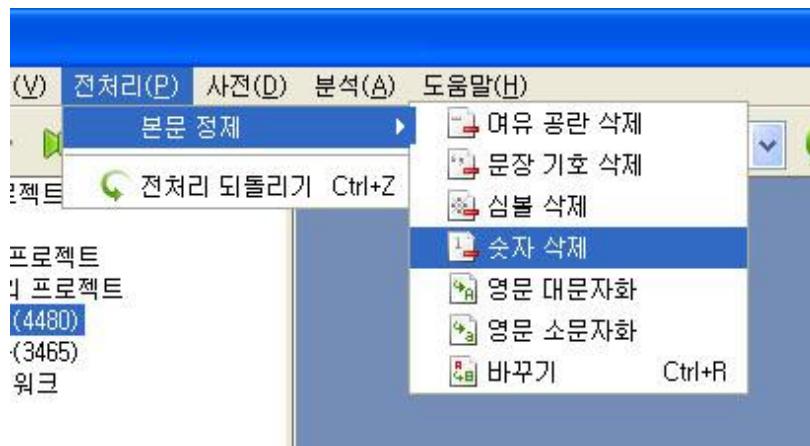
- a) 컴포넌트 트리에서 커서를 해당 프로젝트의 범위 안²에 클릭합니다.

¹ 주의할 점은 각각의 문서에 따로 적용되는 것이 아니라 ‘문서’에 있는 모든 문서에 적용된다는 것입니다.

² 프로젝트의 하위폴더(문서, 저자, 네트워크, 사전)에 커서를 두어도 됩니다. 하지만 해당 프로젝트가 아닌 다른 프로젝트에 커서를 두면 실행이 되지 않습니다.



- b) 메뉴막대의 <전처리>에서 ‘본문 정제’에 커서를 가져간 후 원하는 항목을 클릭합니다.



- c) 작업이 완료되면 결과를 확인하기 위해 ‘문서’에서 임의의 문서를 더블 클릭합니다. 작업창에 탭이 생성되면 결과를 확인할 수 있습니다.

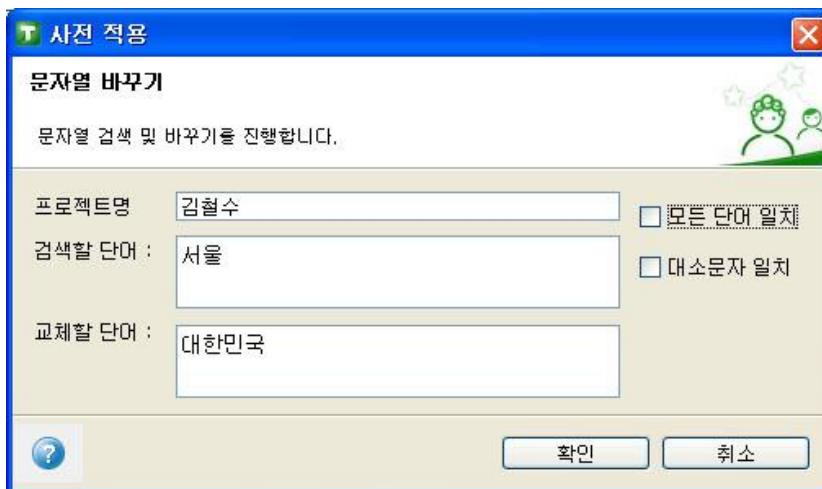
아시아 시장이 개장할 때 금가격 상승세가 두드러집니다. 머지않아 금투기로 인한 피해를 가장 많이 보는 사람은 그런 불행을 피할 수 있었으면 좋겠네요. p.s 2 어제 연방준비은행과 유럽중앙은행 모두 출구전략 가능성에 대해 달려에 대한 불안감도 점차 설 자리를 잃고 있죠

숫자 삭제를 실행시키기 전

아시아 시장이 개장할 때 금가격 상승세가 두드러집니다. 머지않아 금투기로 인한 피해를 가장 많이 보는 사람은 그런 불행을 피할 수 있었으면 좋겠네요 [p.s] 어제 연방준비은행과 유럽중앙은행 모두 출구전략 가능성에 대해 내 달러에 대한 불안감도 점차 설 자리를 잊고 있죠

숫자 삭제를 실행시킨 후

- d) 명령이 성공적으로 수행되었다면 속성에는 ***** Successfully executed ***** 가 표기됩니다.
- 바꾸기를 실행하고자 할 경우 ‘검색할 단어’에 변환할 단어를, ‘교체할 단어’에 변환 되어질 단어를 기입하고 확인을 클릭합니다.



2. 본문정제 처리결과 보기

본문 정제를 실행하였을 때의 결과는 ‘문서’에 있는 임의의 문서를 더블 클릭 한 후 작업창에 생성되는 새 탭에서 확인합니다.

a) 여유 공란 삭제

금융시장을 강타하고 있습니다. 닉터둠 교수는 유로권과 IMF에서 재정지원 할것이기 때문에, 디폴트는 잘사는 나라들인 영국, 프랑스, 독일 모두 재정적자가 늘어나고 있고, 다른 나라를 지원할 정도로 넉넉한 서 벗어난다고 하더라도 포르투갈, 아일랜드, [REDACTED] 이탈리아 같은 나라가 또 기다리고 있습니다.

여유 공란 삭제를 실행시키기 전

금융시장을 강타하고 있습니다. 닉터둠 교수는 유로권과 IMF에서 재정지원 할것이기 때문에, 디폴트는 잘사는 나라들인 영국, 프랑스, 독일 모두 재정적자가 늘어나고 있고, 다른 나라를 지원할 정도로 넉넉한 서 벗어난다고 하더라도 포르투갈, 아일랜드, 이탈리아 같은 나라가 또 기다리고 있습니다.

여유 공란 삭제를 실행시킨 후

2칸 이상의 여유 공란이 사라졌습니다.

b) 문장 기호 삭제

" 해외원정 도박으로 사회에 물의를 일으켰지만 다시 일어날 수 있도록 기회를 준 국민 여러분께 조금이나마 보답하고 어려운 이웃에게 봉사하고 싶어 행사를 시작했다"고 말했다. 황기순은 작년까지 모금한 성금으로 전동휠체어 27대, 일반휠체어 847대를 구입해 장애인들에게 전달했다. yna.co.kr(끝) 저작권자 © 연합뉴스 무단전재·재배포금지 >

문장 기호 삭제를 실행시키기 전

amp quot 해외원정 도박으로 사회에 물의를 일으켰지만 다시 일어날 수 있도록 기회를 준 국민 여러분께 조금이나마 보답하고 어려운 이웃에게 봉사하고 싶어 행사를 시작했다 amp quot 고 말했다 황기순은 작년까지 모금한 성금으로 전동휠체어 27대 일반휠체어 847대를 구입해 장애인들에게 전달했다 engine yna co kr 끝 저작권자 © 연합뉴스 무단전재·재배포금지 amp gt;

문장 기호 삭제를 실행시킨 후

‘@’ 와 ‘.’ ‘(‘ ’) 등이 생략되었습니다.

c) 심볼 삭제

우리는 봉인가?후플러스>는 현대차의 수출용과 내수용 차량의 가격 비교를 시도했다. 현대차는 수 나타났다. 우리나라의 경제 성장과 함께 걸어온 현대차에 대해 남다른 애정을 갖고 있는 우리 국민들. 드는 이유는 무엇일까? 어떤 외압에도 굴복하지 말고, 소신껏 있는 그대로를 밝혀주시길 ★ 기대 만땅...

심볼 삭제를 실행시키기 전

우리는 봉인가?후플러스>는 현대차의 수출용과 내수용 차량의 가격 비교를 시도했다. 현대차는 수 나타났다. 우리나라의 경제 성장과 함께 걸어온 현대차에 대해 남다른 애정을 갖고 있는 우리 국민들. 드는 이유는 무엇일까? 어떤 외압에도 굴복하지 말고, 소신껏 있는 그대로를 밝혀주시길 기대 만땅...

심볼 삭제를 실행시킨 후

★ 가 생략되었습니다.

d) 숫자 삭제

아시아 시장이 개장할 때 금가격 상승세가 두드러집니다. 머지않아 금투기로 인한 피해를 가장 많이 보는 사람은 그런 불행을 피할 수 있었으면 좋겠네요. p.s 2 어제 연방준비은행과 유럽중앙은행 모두 출구전략 가능성에 대해 내 달러에 대한 불안감도 점차 설 자리를 잊고 있죠

숫자 삭제를 실행시키기 전

아시아 시장이 개장할 때 금가격 상승세가 두드러집니다. 머지않아 금투기로 인한 피해를 가장 많이 보는 사람은 그런 불행을 피할 수 있었으면 좋겠네요. p.s 어제 연방준비은행과 유럽중앙은행 모두 출구전략 가능성에 대해 내 달러에 대한 불안감도 점차 설 자리를 잊고 있죠

숫자 삭제를 실행시킨 후

단독으로 쓰인 숫자 2 가 생략되었습니다.

e) 영문 대문자화

대주교 사택 밑에서 구조됐으며, 포르토프랭스 쇼핑센터에서는 25살 여성 호텔라인 로사나가 구조됐다. 아남았던 15살 소녀를 숨지게 했다고 목격자들이 전했다. 목격자들은 경찰이 경고사격을 하다 오발사 중국 일간 데일리메일은 전했다. newglass@yna.co.kr(끝)저작권자(c)연합뉴스. 무단전재-재배포금지. >

영문 대문자를 실행시키기 전

대주교 사택 밑에서 구조됐으며, 포르토프랭스 쇼핑센터에서는 25살 여성 호텔라인 로사나가 구조됐다. 아남았던 15살 소녀를 숨지게 했다고 목격자들이 전했다. 목격자들은 경찰이 경고사격을 하다 오발사 중국 일간 데일리메일은 전했다. NEWGLASS@YNA.CO.KR(끝)저작권자(c)연합뉴스. 무단전재-재배포금지. >

영문 대문자를 실행시킨 후

영어 소문자가 대문자로 변환되었습니다.

f) 영문 소문자

사소송 1심, PD수첩 무죄' 편(서울=연합뉴스) 임은진 기자 = MBC PD 수첩'은 지난 20일 법원이 PD 수선 아래사 빈손의 사인과 관련한 어머니와의 인터뷰 자료를 내보낸다. 이 자료에서 빈손의 어머니는 "1 손의 어머니가 딸의 수술을 집도한 병원을 상대로 한 소장도 가지고 있다고 밝혔다. 이 소장에는 의사가

영문 소문자를 실행시키기 전

사소송 1심, pd수첩 무죄' 편(서울=연합뉴스) 임은진 기자 = mbc pd 수첩'은 지난 20일 법원이 pd 수선 아래사 빈손의 사인과 관련한 어머니와의 인터뷰 자료를 내보낸다. 이 자료에서 빈손의 어머니는 "1 손의 어머니가 딸의 수술을 집도한 병원을 상대로 한 소장도 가지고 있다고 밝혔다. 이 소장에는 의사가

영문 소문자화를 실행시킨 후

영어 대문자가 소문자로 변환되었습니다.

g) 바꾸기

서울 연합뉴스 임은진 기자 개그맨 황기순이 다음달 동료 연예인들과 함께 사이클을 타고 전국 각지를 돌며 장애인 돋기 성금을 모금하는 사랑 더하기 사이클대행진 을 벌인다 2000년부터 시작해 올해로 9회째를 맞는 이번 행사에서 황기순은 다음달 13 23일 서울을 출발해 수원 천안 대전 대구 부산 등 7개 도시를 돌며

바꾸기를 실행시키기 전

대한민국 연합뉴스 임은진 기자 개그맨 황기순이 다음달 동료 연예인들과 함께 사이클을 타고 전국 각지를 돌며 장애인 돋기 성금을 모금하는 사랑 더하기 사이클대행진 을 벌인다 2000년부터 시작해 올해로 9회째를 맞는 이번 행사에서 황기순은 다음달 13 23일 대한민국을 출발해 수원 천안 대전 대구 부산 등 7개 도시를

바꾸기를 실행시킨 후

‘서울’ 이 ‘대한민국’으로 변환되었습니다.

C. 자연어 처리 실행하기

1. 자연어 처리

새 프로젝트를 생성한 후 각 문장에 대한 형태소, 품사, 구문분석, NER 정보를 생성합니다. 이는 사전¹을 만들기 위해 실행하는 것으로, 프로그램이 단어를 개개로 인지할 수 있도록 문서를 세분화시킵니다. 다음은 형태소, 품사, 구문분석, NER 정보에 관한 표입니다.

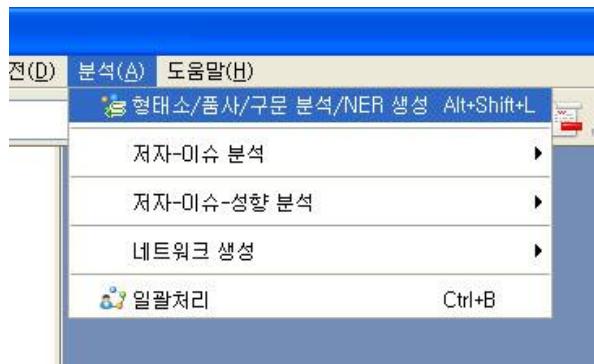
종 류	의 미	예
형태소	뜻을 가진 가장 작은 말의 단위	어미, 선어말어미 등
품사	단어를 기능, 형태, 의미에 따라 나눈 갈래	명사, 대명사, 수사, 조사, 동사, 형용사, 관형사, 부사, 감탄사
구문분석	문장의 구/절을 구분하는 것	명사구, 서술어구, 부사어구 등
NER	고유명사	인명, 지명, 국명, 책명 등

사용자의 필요에 따라 실행되는 본문 정제와는 달리, 자연어 처리는 사전을 생성하기 위한 또 하나의 절차로써 필수적으로 실행되어야만 합니다. 즉, 자연어 처리를 실행해야만 사전을 생성할 수 있으며 본문 정제는 사전을 생성하는 것과는 관련이 없습니다. 주의할 점은 자연어 처리를 실행한 후 이전 단계인 본문 정제를 실행하면 새 프로젝트를 생성한 후 적용되었던 모든 명령이 취소²된다는 것입니다. 다음은 자연어 처리를 실행하는 과정입니다.

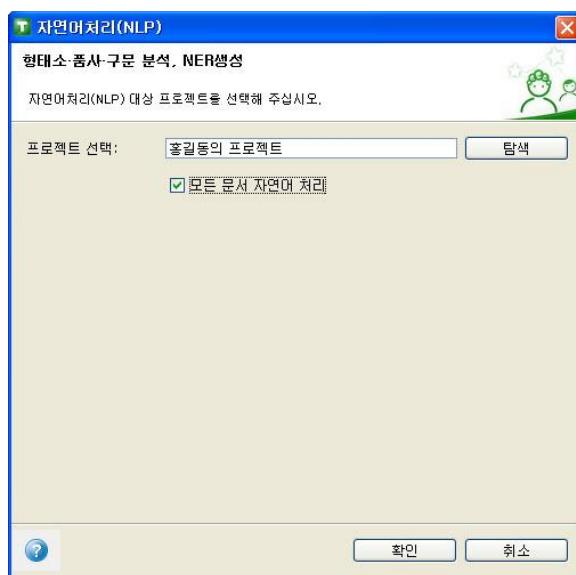
¹ 사전은 문서의 토픽을 추출하기 위한 일종의 도구로 쓰이며 총 5 가지 종류가 있습니다.

² 본문 정제를 실행시킨 상태에서 사전 생성으로 넘어가고 싶다면 다시 자연어 처리를 실행합니다.

- a) 자연어 처리를 실행하려면 **메뉴막대 – 분석 – 형태소/품사/구문분석/NER 생성**을 누릅니다.

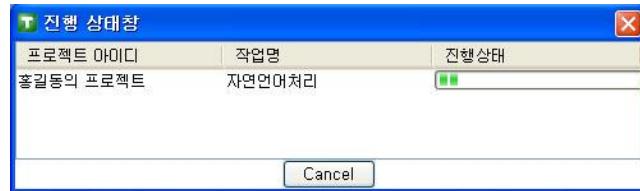


- b) ‘자연어 처리’ 대화상자가 나타나면 오른쪽의 <탐색>을 눌러 적용시키고자 하는 프로젝트를 선택한 후 <확인>을 클릭합니다. ‘모든 문서 자연어 처리’에 체크를 하지 않으면 기존에 자연어 처리가 되어있는 문서는 제외한 나머지 문서에 자연어 처리가 실행됩니다. 그러나 ‘모든 문서 자연어 처리’에 체크를 하면 기존에 자연어 처리의 여부와 상관 없이 모든 문서를 대상으로 자연어 처리가 실행됩니다.



*컴포넌트 트리에서 적용시키고자 하는 프로젝트에 이미 클릭이 되어 있을 경우, ‘프로젝트 선택’란에 별도의 탐색을 하지 않아도 자동으로 해당 프로젝트명이 나타납니다.

- c) ‘진행 상태창’ 대화상자가 나타나면 연두색 막대가 채워질 때까지 기다립니다. 이때 <Cancel>을 클릭하면 ‘자연어 처리’가 중단됩니다.



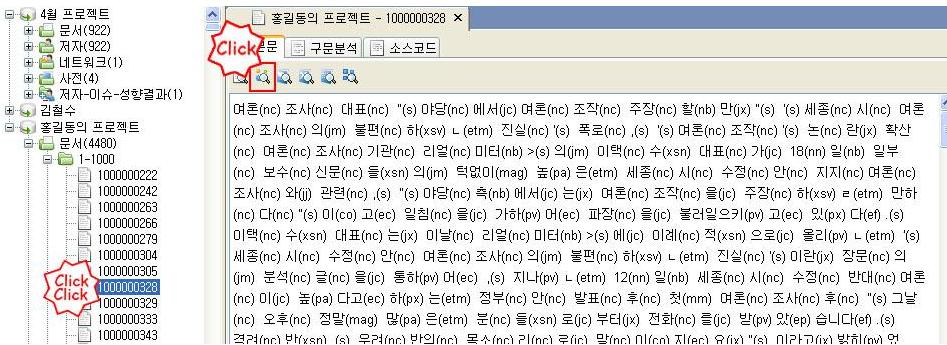
- d) 명령이 성공적으로 수행되었다면 속성에는 ***** Successfully executed ***** 가 표기됩니다.

2. 자연어 처리 처리결과보기

자연어 처리를 실행하였을 때의 결과는 ‘문서’에서 임의의 문서¹를 더블 클릭 한 후 작업창에 생성되는 새 탭에서 확인합니다.

a) 형태소와 품사가 생성된 화면

컴포넌트 트리에서 임의의 문서를 더블 클릭한 뒤 새 탭이 생성되면 하위 탭인 HTML 탭에서 형태소 분석 결과 보기 아이콘을 클릭합니다. 형태소에 대한 자세한 정보는 <부록-품사리스트, page 361>에서 확인할 수 있습니다.



b) NER 이 생성된 화면

컴포넌트 트리에서 임의의 문서를 더블 클릭한 뒤 새 탭이 생성되면 HTML 탭에서 NER 보기 아이콘을 클릭합니다.

¹ 문서탭 자체에 새로고침 기능이 없기 때문에 이미 문서탭을 열어놓았던 상태이더라도 제거한 뒤 해당 문서를 클릭하여 다시 생성하여야 합니다.



D. 사전 생성하기

TONK에서 사전이란 사용자가 문서에서 특정 단어들을 수집해 놓은 것을 말합니다. 자연어 처리를 실행한 후 문서가 단어 단위로 세분화됨에 따라 사용자는 사용자가 원하는 특정 단어만 추출할 수 있게 됩니다. 이를 기반으로 사전이 만들어지면 문서에 적용시켜 사용자가 원하는 분석을 실행합니다.

TONK에서 제공하는 사전은 총 5 가지¹로, 원문에서 사용자에게 필요치 않은 단어들을 모아놓은 **삭제어 사전**, 동일한 의미를 가진 단어들을 한 단어로 통합해 놓은 **일반화 사전**, 개개의 단어에 온톨로지가 부여된 단어들을 모아놓은 **메타넷 사전**, 사용자의 주관에 따라 긍정과 부정의 의미를 가진 단어를 모아놓은 **성향 단어사전**, 사용자가 필요로 하는 단어들만 모아 놓은 **단어 사전**이 있습니다.

사전 생성 이후 실행할 수 있는 분석은 총 3²가지로 각 분석마다 요구되는 사전이 다르므로 사용자는 분석 목적에 맞게 사전을 생성하도록 합니다. 특히 이번 장<I. 문서 내용 및 메타 정보 분석>에서는 ‘저자-이슈 분석’과 ‘저자-이슈-성향 분석’을 실행하기 때문에 필요한 사전은 삭제어 사전, 일반화 사전, 성향 단어사전, 단어사전입니다.

사전을 만드는 방법에는 사용자가 직접 단어를 추가하거나 삭제하기 위해 만드는 ‘생성’과 원문에서 단어를 추가하는 ‘사전 원문에서 생성’이 있습니다. 또한 프로그램을 통해 자동으로 생성하는 ‘자동생성’과 사용자의 필요에 따라 기존에 생성되었던 사전의 파일을 컴퓨터에서 ‘도입’하는 방법도 있습니다.

이렇게 만들어진 사전을 문서에 ‘적용’을 실행합니다. ‘적용’을 실행하는 과정에서는 사용자의 분석 목적에 따라 특별히 정해진 순서가 있습니다.

분석명	순서
저자-이슈 분석	삭제어 사전 적용 – 일반화 사전 적용 – 단어사전 생성 – 저자-이슈 분석
저자-이슈 성향 분석	삭제어 사전 적용 – 일반화 사전 적용 – 성향 단어사전 생성 – 단어사전 생성 – 저자-이슈-성향 분석

¹ 사전의 종류는 4 가지이지만, 같은 종류의 사전을 여러 개 만들 수도 있습니다.

² 저자-이슈 분석, 저자-이슈-성향분석, 문서내 관계망 분석입니다.

‘저자-이슈 분석’과 ‘저자-이슈-성향 분석’은 최종적으로 사전을 적용 시킨 후에 문서에 남아 있는 단어들과, 단어 사전으로 수집된 단어들의 교집합을 대상으로 서로 관련 있는 단어들을 모아놓은 토픽을 추출하여 비중에 따라 내림차순으로 정렬하는 것입니다.

일반적으로 고려해보면 여타 사전들처럼 단어사전도 적용을 실행시켜야 하지 않을까라는 추측을 할 수 있으나 특이하게도 단어사전은 ‘자동생성’ 기능만 존재하고 ‘적용’ 기능은 존재하지 않습니다. 대신에 ‘자동 생성’을 통해 수집된 단어들이 ‘저자-이슈 분석’과 ‘저자-이슈-성향 분석’을 실행하는 과정에서 쓰이게 됩니다.

사전들은 토픽을 추출하고 단어간 관계를 알아내는 목적 이외에 다른 용도로 쓰일 수 있는 기능이 아직까지는 없기 때문에 단어사전에 따로 ‘적용’이라는 기능을 만들어 놓고 다시 ‘저자-이슈 분석’과 ‘저자-이슈-성향 분석’을 실행하는 것이 아니라, 조금이라도 절차를 덜기 위해 바로 ‘저자-이슈 분석’과 ‘저자-이슈-성향 분석’으로 넘어갑니다. 즉, 단어사전의 ‘적용’이란 개념을 단어사전에서는 분석 목적에 따라 ‘저자-이슈 분석’과 ‘저자-이슈-성향 분석’으로 간주할 수 있습니다.

단, ‘저자-이슈-성향 분석’은 저자들이 특정 성향을 가지고 있으며 이들이 사용하는 단어의 정보가 존재할 경우에만 분석이 가능하므로 성향 단어사전을 추가로 생성해야 합니다. 이 성향 단어사전은 단어사전과 마찬가지로 실제 분석과정에 이용되므로 적용단계가 존재하지 않습니다.

주의할 사항은 사전을 만들어 놓기만 하고 ‘적용’을 실행시키지 않는다면 ‘저자-이슈 분석’의 결과에 반영이 되지 않기 때문에 사용자에게 필요한 사전은 꼭 ‘적용’을 실행시켜야 한다는 것입니다. 또한 4¹가지 종류의 사전을 모두 생성하는 것이 필수적인 것은 아니나, 여러 가지 특성이 다른 사전을 문서에 적용시킴으로써 보다 더 정확성 높은 토픽을 산출할 수 있다는 것도 명심해야 합니다. 다음은 사전 정보를 요약해 놓은 표입니다.

종 류		기 능	결 과
1	삭제어 사전	원문에서 사용자에게 필요치 않은 단어를 모아둔 것	삭제어 단어가 원문에서 삭제됨
2	일반화 사전	원문에서 동일한 의미를 가진 단어를 통일된 단어로 모아둔 것	일반화 단어가 원문에서 통일됨

¹ 삭제어 사전, 일반화 사전, 메타넷 사전, 성향 단어사전입니다. 단어사전은 ‘저자-이슈 분석’과 ‘저자-이슈-성향 분석’으로 넘어가기 때문에 필수적으로 생성되어야 합니다.

4	성향 단어 사전	사용자의 주관에 따라 긍정과 부정의 의미를 가진 단어를 모아둔 것	문서에서 비중 있게 다루어지는 토픽의 성향을 파악함-
5	단어 사전	원문에서 사용자가 원하는 특정 단어를 모아둔 것	토픽이 생성됨

사전이 생성되면 컴포넌트 트리의 ‘사전’에 목록이 추가 되고 속성에는 사전명, 생성일, 위치, 크기와 같은 상세정보가 나타납니다. 사전을 적용시킨 결과는 **문서—임의의 문서 – 문서 탭**에서 해당 사전의 아이콘을 눌러 확인할 수 있습니다.

사전을 만드는 것에서부터 적용시키는 과정까지 하나씩 살펴봅시다.

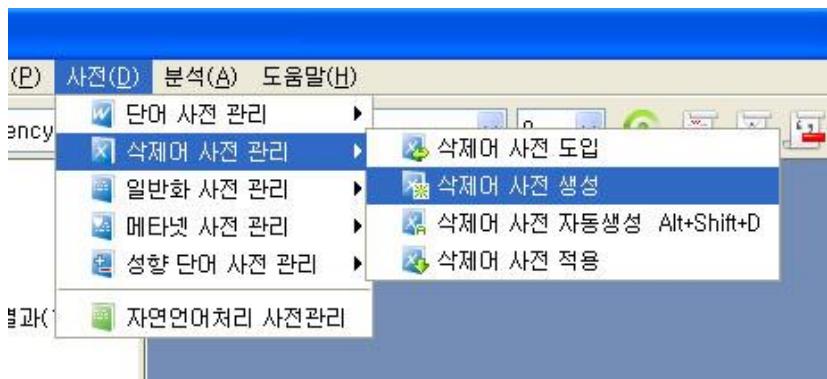
1. 삭제어 사전

삭제어 사전은 원문에서 사용자에게 필요치 않은 단어를 모아둔 사전입니다. 삭제어 사전을 만드는 방법에는 (1) 사용자가 직접 단어를 추가하거나 삭제하기 위해 초기화된 사전을 만드는 ‘삭제어 사전 생성’과 (2) 원문에서 삭제어를 추가하는 ‘삭제어 사전 원문에서 생성’ (3) 프로그램을 통해 자동으로 생성하는 ‘삭제어 자동생성’ (4) 사용자의 필요에 따라 기존에 생성되었던 사전의 파일을 컴퓨터에서 불러오는 ‘삭제어 사전 도입’이 있습니다. 삭제어 사전을 생성한 후 사용자가 이를 문서에 적용하고자 한다면 ‘삭제어 사전 적용’을 실행합니다.

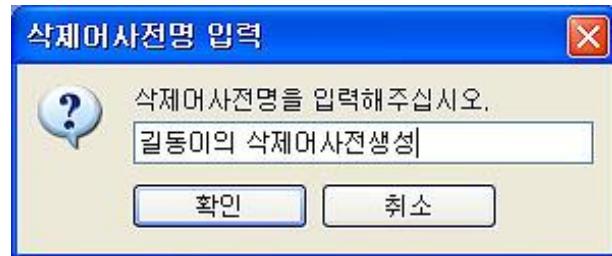
a) 삭제어 사전 생성

‘삭제어 사전 생성’은 사용자가 직접 단어를 추가/삭제하기 위하여 초기화된 삭제어 사전을 생성합니다. 다음은 ‘삭제어 사전 생성’을 실행하는 과정입니다.

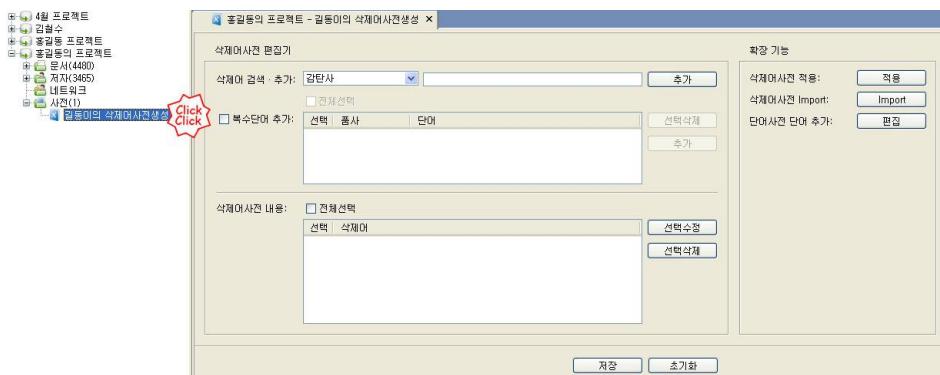
(a-1) (방법 1) ‘삭제어 사전 생성’을 실행하려면 **메뉴막대 – 사전 – 삭제어 사전 관리 – 삭제어 사전 생성**을 누릅니다. 또는 (방법 2) 컴포넌트 트리의 ‘사전’에 오른쪽 마우스를 클릭한 후 ‘삭제어 사전 생성’을 누릅니다.



(a-2) ‘삭제어 사전명 입력’ 대화상자가 나타나면 기입란에 본인이 원하는 이름을 쓰고 <확인>을 누릅니다.



(a-3) 이후 컴포넌트 트리의 ‘사전’에 삭제어 사전이 생성됩니다. 생성된 삭제어 사전을 더블 클릭하면 작업창에 삭제어 사전 편집탭이 나타나며 원하는 단어를 추가하거나 삭제할 수 있습니다.



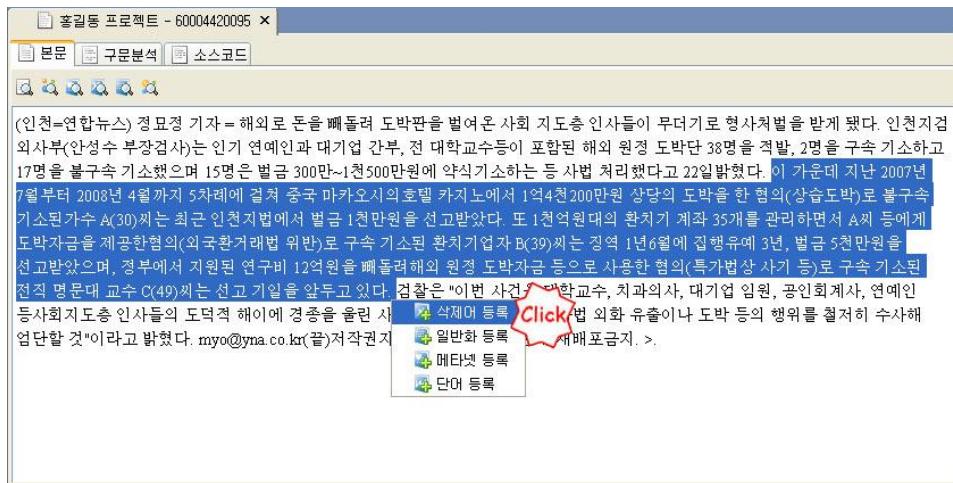
(a-4) 생성된 삭제어 사전은 파일의 형태로 사용자가 지정한 작업폴더(최초 설정- C://Workplace, page 29, (d-2) 참조)에 자동 저장되며 메시지 패널의 속성에서 저장된 위치 등 상세정보를 확인할 수 있습니다.

(a-5) 명령이 성공적으로 수행되었다면 속성에는 ******* Successfully executed ******* 가 표기됩니다.

b) 삭제어 사전 원문에서 생성

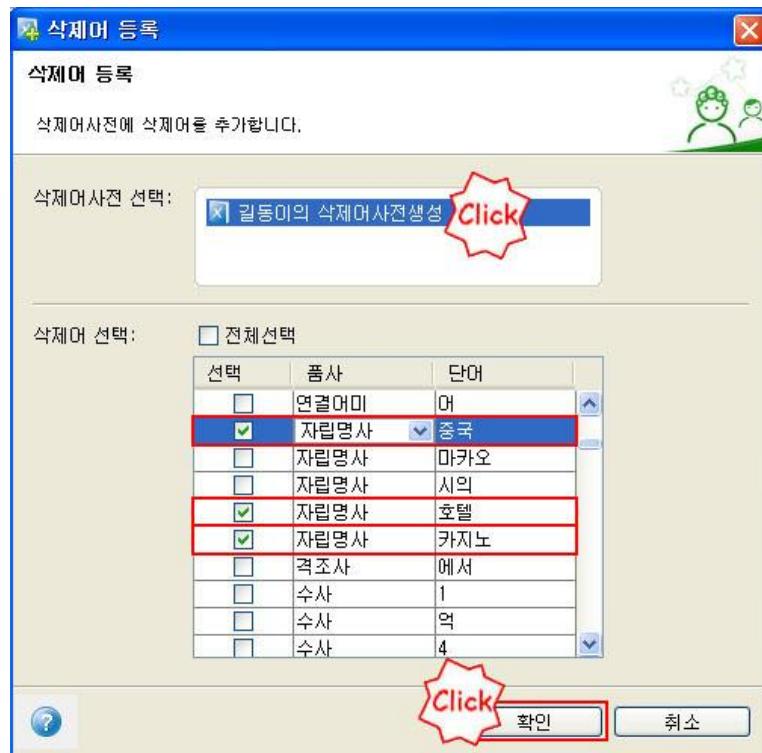
'삭제어 사전 원문에서 생성'은 원문에 드래그를 하여 단어를 추가하는 방식으로 사용자가 원문에서 직접 삭제어를 등록하여 기존에 생성된 삭제어 사전에 단어를 추가합니다. 다음은 '삭제어 사전 원문에서 생성'을 실행하는 과정입니다.

(b-1) 컴포넌트 트리 '문서'에서 추가하고자 하는 삭제어가 포함되어 있는 문서를 클릭하여 작업창에 새 탭을 생성합니다. 삭제어로 추가하고자 하는 단어나 문장에 마우스로 드래그합니다. 이후 마우스 오른쪽버튼을 클릭하여 '삭제어 등록'을 누릅니다.



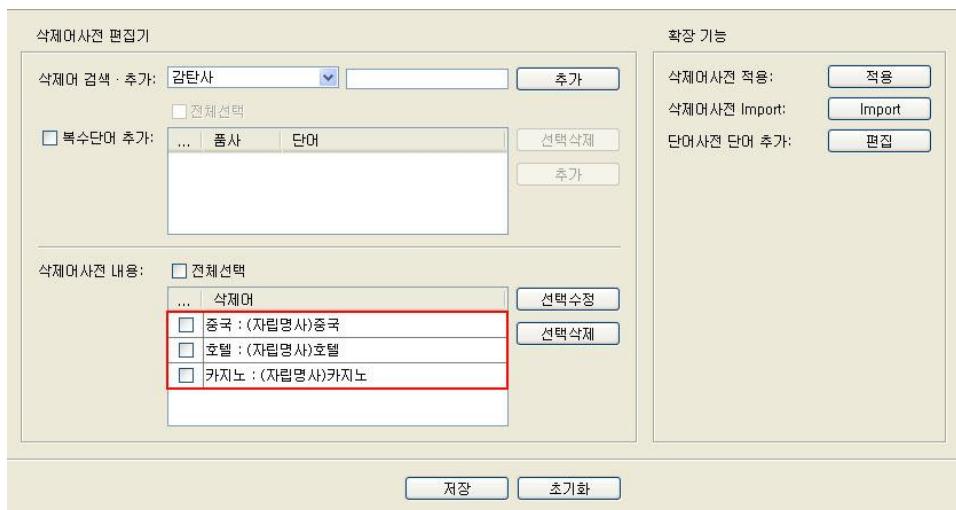
(b-2)

‘삭제어 등록’ 대화상자가 나타나면 ‘사전 선택’에서 추가하려는 삭제어 사전에 클릭을 합니다.
 ‘삭제어 선택’란은 사용자가 드래그를 한 단어/문장의 품사가 자동으로 구분이 되어 나타납니다. 추가하고자 하는 삭제어에 체크를 한 후 <확인>을 누릅니다.



자립명사인 ‘참여’를 삭제어로 선택

(b-3) 컴포넌트 트리의 ‘사전’에서 해당 삭제어 사전을 더블 클릭하여 작업창에 새 탭을 생성합니다.
 삭제어로 선택한 단어가 삭제어 사전에 등록된 것을 확인합니다.



기존의 삭제어 사전에 추가한 경우

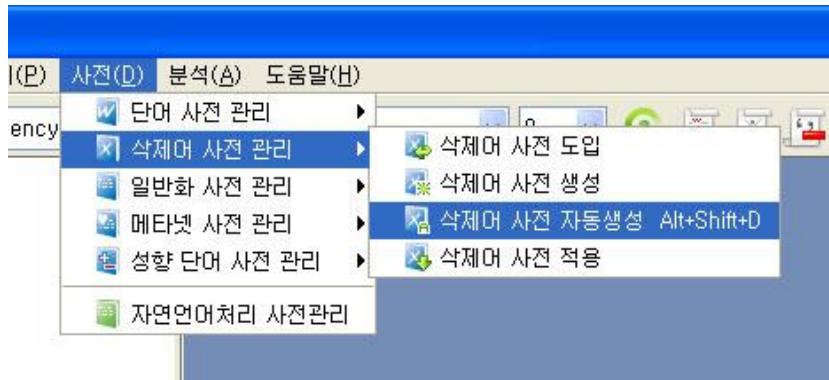
(b-4) 생성된 삭제어 사전은 파일의 형태로 사용자가 지정한 작업폴더(최초 설정- C://Workplace, page 29, (d-2) 참조)에 자동 저장되며 메시지 판넬의 속성에서 저장된 위치 등 상세정보를 확인할 수 있습니다.

(b-5) 명령이 성공적으로 수행되었다면 속성에는 ***** Successfully executed ***** 가 표기됩니다.

c) 삭제어 사전 자동생성

삭제어 사전을 프로그램을 통해 생성합니다. 다음은 ‘삭제어 사전 자동생성’을 실행하는 과정입니다.

(c-1) ‘삭제어 사전 자동 생성’을 실행하려면 메뉴막대 – 전처리 – 삭제어 사전 관리 – 삭제어 사전 자동생성을 누릅니다.

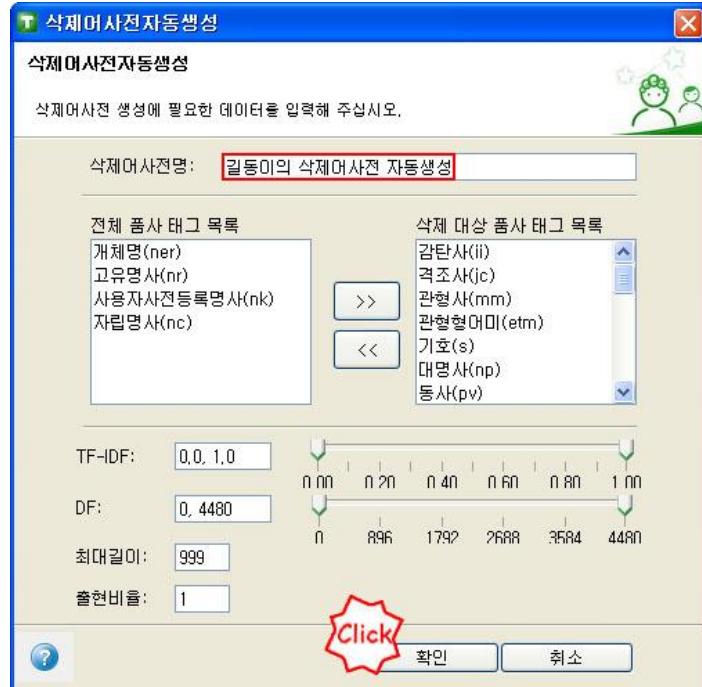


(c-2)

‘삭제어 사전 자동생성’ 대화상자가 나타나면 ‘삭제어 사전명’에는 본인이 원하는 제목을 씁니다.

‘전체 품사 태그 목록’에서 분석하고자 하는 품사를 클릭한 후 추가버튼을 눌러 ‘대상 품사 태그 목록’으로 이동시킵니다. 만약 이동 시킨 품사를 삭제하고자 하는 경우에는 삭제버튼을 누릅니다. 하지만 사용자가 분석하기 용이하도록 이상적인 수치와 품사¹를 기본적으로 나타내주기 때문에 특정한 목적 이외에는 그대로 사용합니다.

¹ 일반적으로 명사를 중심으로 한 토픽을 얻는 것이 전체 주제를 파악하는데 용이하므로 명사관련 품사를 제외한 모든 품사가 주로 선택됩니다. 품사에 대한 보다 자세한 정보는 페이지 360쪽을 참고합니다.



- i. TF-IDF 는 한 문서 내에서 단어가 쓰인 횟수와 단어가 전체 문서 집합에서 등장하는 횟수의 역수를 곱한 것으로 수치가 높을수록 해당 단어가 특정 문서에서 자주 쓰였다는 것을 나타냅니다. 즉, TF-IDF 값이 높을수록 그 단어는 특정 이슈에서 자주 쓰였을 가능성이 높습니다. 주의할 사항은 삭제어 사전에서의 TF-IDF 값을 적용하는 방식은 사용자가 지정한 범위를 제외한 나머지 부분에서 적용된다는 것입니다. 다시 말하면, 지정한 수치의 이외의 값을 가진 단어들이 삭제됩니다. 아래의 붉은 막대로 표현된 부분이 사용자가 지정한 값의 이외의 부분으로 삭제어 사전에 추가됩니다. 만약 범위를 0.0과 1.0을 지정한다면 삭제어는 존재하지 않습니다.



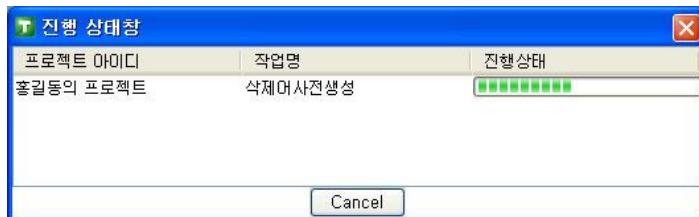
- ii. 한 단어가 등장하는 문서의 수를 나타냅니다. DF 값이 높으면 아주 일반적인 단어를 의미하는 것이고 DF 값이 낮으면 매우 특수한 경우의 단어를 의미합니다. 일반적으로 우리가 관심 있어하는 단어는 중간 정도의 값을 지닌 DF 입니다. DF 값이 매우 높은

경우는 -은, -는, -이, -가와 같이 문서의 성향과 주제에 상관없이 자주 들어가는 단어를 의미하게 되고, DF 값이 매우 낮은 경우에는 오·탈자가 섞인 단어, 혹은 매우 특수한 경우에만 쓰이는 단어가 선별되게 됩니다. 그러므로 삭제어 사전에서는 중간 정도의 범위를 벗어난 DF 값을 가진 단어들로 생성되는 것이 중요합니다. 다시 말하면, DF 값을 적용하는 방식은 사용자가 지정한 범위를 제외한 나머지 부분에서 적용됩니다. 아래의 붉은 막대로 표현된 부분이 사용자가 지정한 값의 이외의 부분으로 삭제어 사전에 추가됩니다.

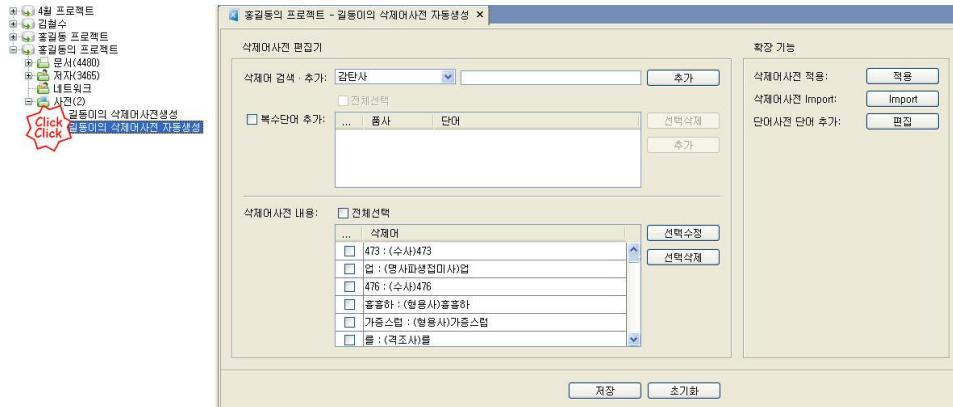


- iii. 최대길이는 단어의 최대 길이입니다.
- iv. 출현빈도수는 전체 문서 집합에서 출현한 횟수입니다.

(c-3) ‘진행 상태창’ 대화상자가 나타나면 연두색 막대가 채워질 때까지 기다립니다. 이때 <Cancel>을 클릭하면 ‘삭제어 사전 자동생성’이 중단됩니다.



(c-4) 작업이 완료되면 컴포넌트 트리의 ‘사전’에 삭제어 사전이 생성됩니다. 생성된 삭제어 사전을 더블 클릭하면 작업창에 삭제어 사전 편집 탭이 나타나며 프로그램에서 추출한 형태소 이외에 사용자가 원하는 단어를 추가하거나 삭제할 수 있습니다.



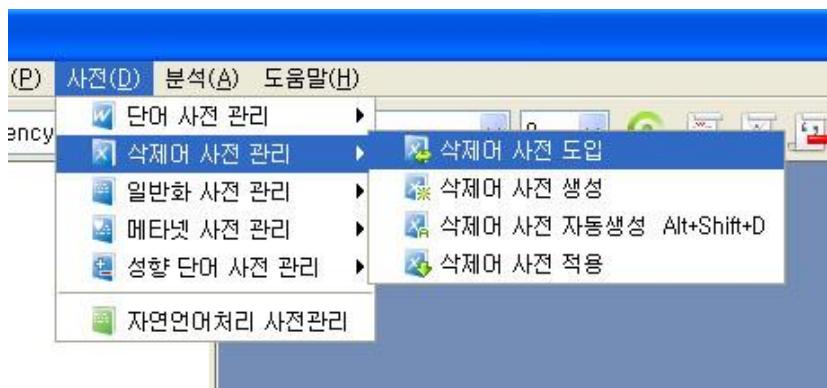
(c-5) 생성된 삭제여 사전은 파일의 형태로 사용자가 지정한 작업폴더 (최초 설정- C://Workplace, page 29, (d-2) 참조)에 자동 저장되며 메시지 판넬의 속성에서 저장된 위치 등 상세정보를 확인할 수 있습니다.

(c-6) 명령이 성공적으로 수행되었다면 속성에는 ***** Successfully executed ***** 가 표기됩니다.

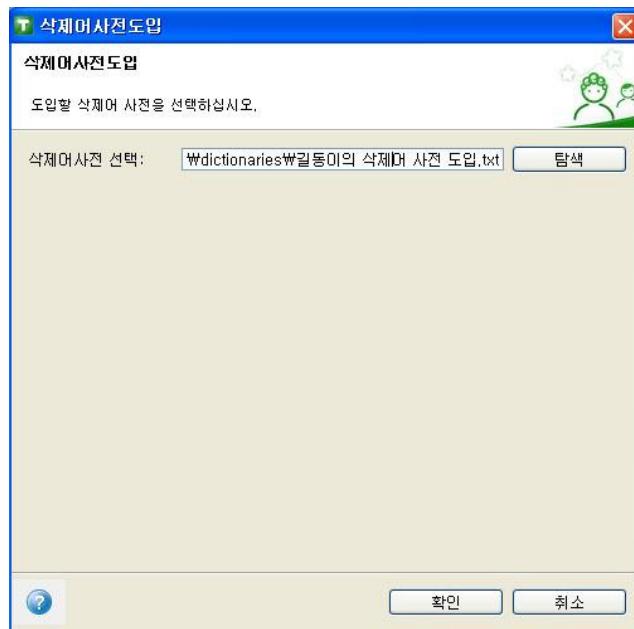
d) 삭제여 사전 도입

기존에 생성되었던 삭제여 사전의 파일을 컴퓨터에서 불러옵니다. 다음은 ‘삭제여 사전 도입’을 실행하는 과정입니다.

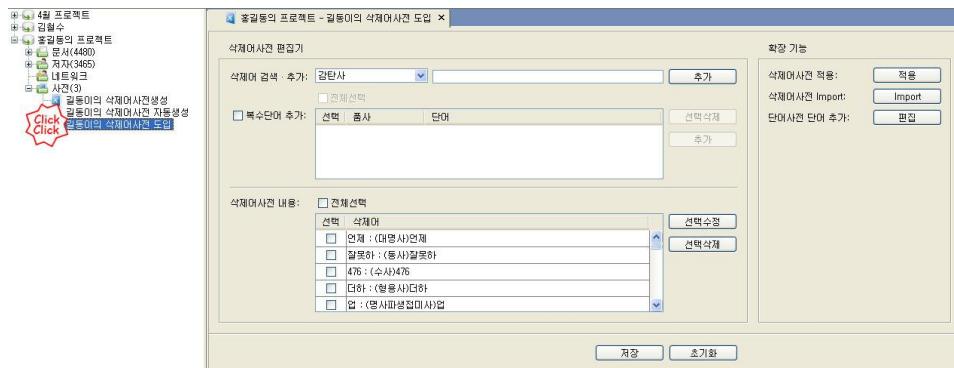
(d-1) ‘삭제여 사전을 도입’을 실행하려면 **메뉴막대 – 사전 – 삭제여 사전 관리 – 삭제여 사전 도입**을 누릅니다.



(d-2) ‘삭제어 사전 도입’ 대화상자가 나타나면 <탐색>을 눌러 ‘사전’로 불러올 파일을 선택한 후 <확인>을 클릭합니다.



(d-3) 이후 컴포넌트 트리의 ‘사전’에 삭제어 사전이 생성됩니다. 불러온 삭제어 사전을 더블 클릭하면 작업창에 삭제어 사전 편집 탭이 나타나며 사용자가 원하는 단어를 추가하거나 삭제할 수 있습니다.

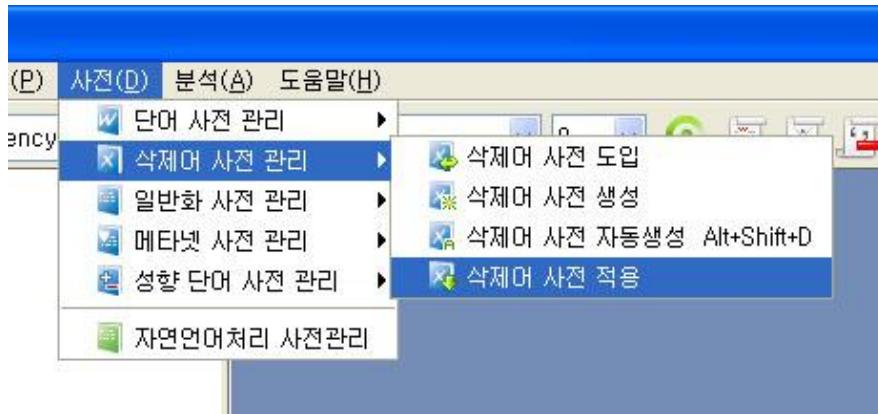


(d-4) 명령이 성공적으로 수행되었다면 속성에는 ***** Successfully executed ***** 가 표기됩니다.

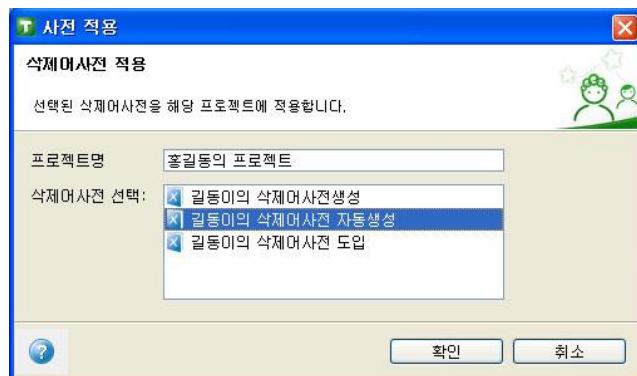
e) 삭제어 사전 적용

생성한 삭제어 사전을 문서에 적용합니다. 다음은 ‘삭제어 사전 적용’을 실행시키는 과정입니다.

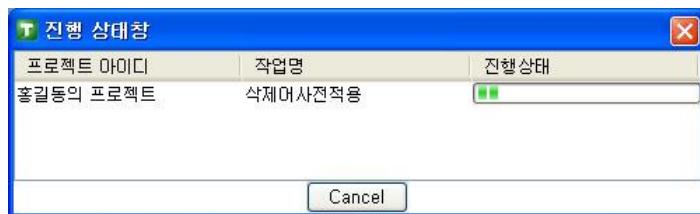
(e-1) ‘삭제어 사전 적용’을 실행하려면 메뉴 막대 – 사전 – 삭제어 사전 관리 – 삭제어 사전 적용을 누릅니다.



(e-2) ‘사전 적용’ 대화상자가 나타나면 프로젝트에 적용시킬 삭제어 사전을 선택한 후 <확인>을 클릭합니다.



(e-3) ‘진행 상태창’ 대화상자가 나타나면 연두색 막대가 채워질 때까지 기다립니다. 이때 <Cancel>을 클릭하면 ‘삭제어 사전 적용’이 중단됩니다.



(e-4) 명령이 성공적으로 수행되었다면 속성에는 `***** Successfully executed *****` 가 표기됩니다.

2. 삭제어 사전 처리결과 보기

삭제어 사전을 적용시킨 결과를 확인하려면 **컴포넌트 트리—문서**에서 임의의 문서¹를 더블 클릭합니다. 좌측에 새 탭이 생성되면 **HTML 탭 — 삭제어 보기 아이콘**을 클릭하여 삭제어가 빨간색으로 표시되어 있고 취소선이 그어져 있는 것을 확인합니다.

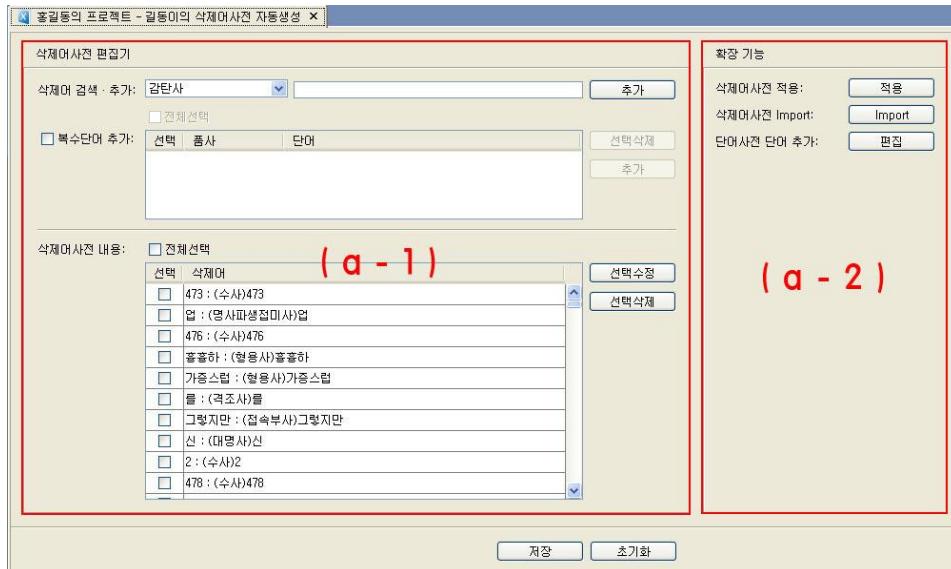


‘사전’에서 삭제어 사전을 더블 클릭하면 작업창에 해당 사전을 편집할 수 있는 탭이 나타나며 프로그램에서 추출한 단어 이외에 사용자가 원하는 단어를 추가하거나 삭제할 수 있습니다.

a) 삭제어사전

컴포넌트 트리의 ‘사전’에 있는 임의의 삭제어 사전을 클릭했을 때 작업창은 다음과 같습니다.

¹ 사전의 ‘적용’은 ‘문서’에 있는 모든 문서에 적용이 되기 때문에 임의의 문서에서 열람할 수 있습니다.



탭의 상단에는 프로젝트 명과 사전종류가 표기되며 하위 메뉴에는 ‘삭제어사전 편집기’과 ‘확장 기능’ 총 2 가지가 있습니다.

- 저장 저장 : 편집된 내용을 저장합니다.
- 초기화 초기화 : 명령이 실행되지 않은 초기의 상태로 돌아갑니다.

(a-1) 삭제어 사전 편집기

i. 삭제어 검색 추가

삭제어를 검색하고자 한다면 선택란에 해당하는 품사를 선택한 후 우측란에 찾고자 하는 삭제어를 기입하면 하단에 조건에 일치하는 검색어가 나타납니다.

삭제어를 추가하고자 한다면 선택란에서 해당하는 품사를 선택한 후 우측란에 추가하고자 하는 삭제어를 기입합니다. 이후 <추가>를 누르면 하단에 조건에 일치하는 삭제어가 생성됩니다.

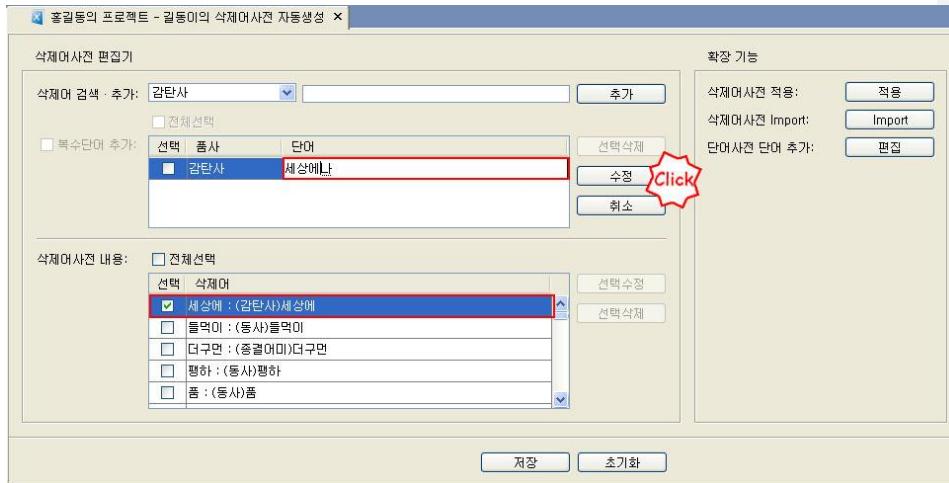


고유명사를 선택하고 ‘홍길동’을 추가 했을 때

복수 단어를 삭제어로 추가하고자 한다면 복수단어 추가란에 체크를 한 후 <추가>를 누릅니다.

ii. 삭제어사전 내용

삭제어 사전에 등록되어 있는 삭제어들이 나타납니다. 임의의 삭제어의 선택란에 체크를 하고 <선택삭제>를 클릭하면 삭제어 사전에서 삭제됩니다. 임의의 삭제어의 선택란에 체크를 하고 <선택수정>을 클릭하면 삭제어 검색 추가로 이동을 하게 되며 해당 단어에 더블클릭을 하여 단어를 수정합니다.

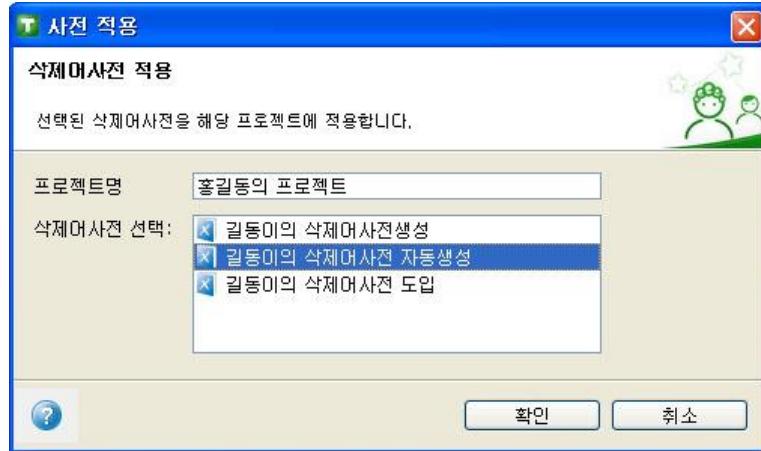


(a-2) 확장 기능



i. 삭제어 사전 적용

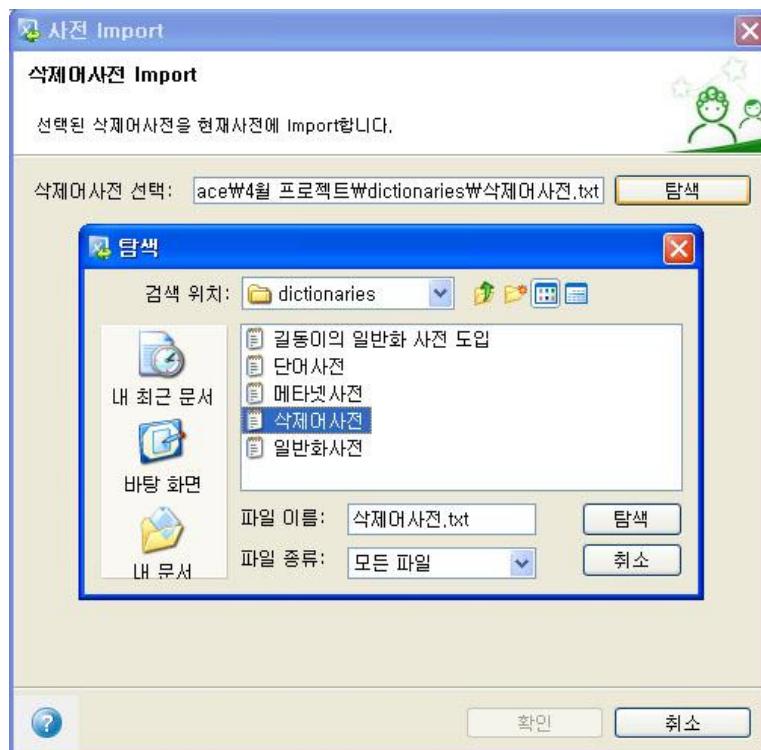
삭제어 사전을 프로젝트에 적용합니다. <적용>을 누르면 아래와 같은 대화상자가 나타납니다. 삭제어 사전 선택란에서 적용하고자 하는 삭제어 사전을 선택한 후 <확인>을 클릭합니다.



‘길동이의 삭제어사전 자동생성’을 적용하고자 할 때

ii. 삭제어 사전 Import

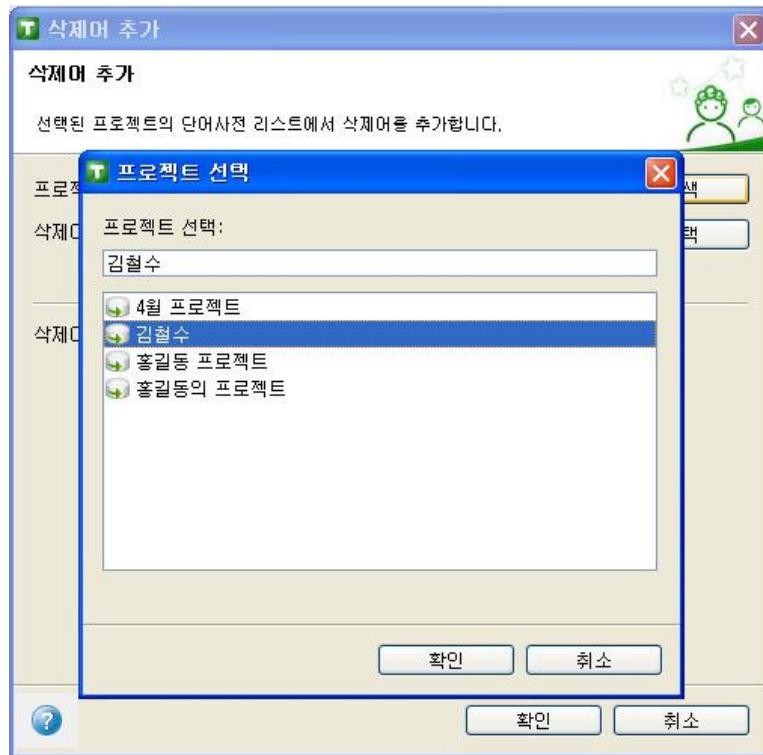
기존에 생성된 삭제어 사전을 불러옵니다.



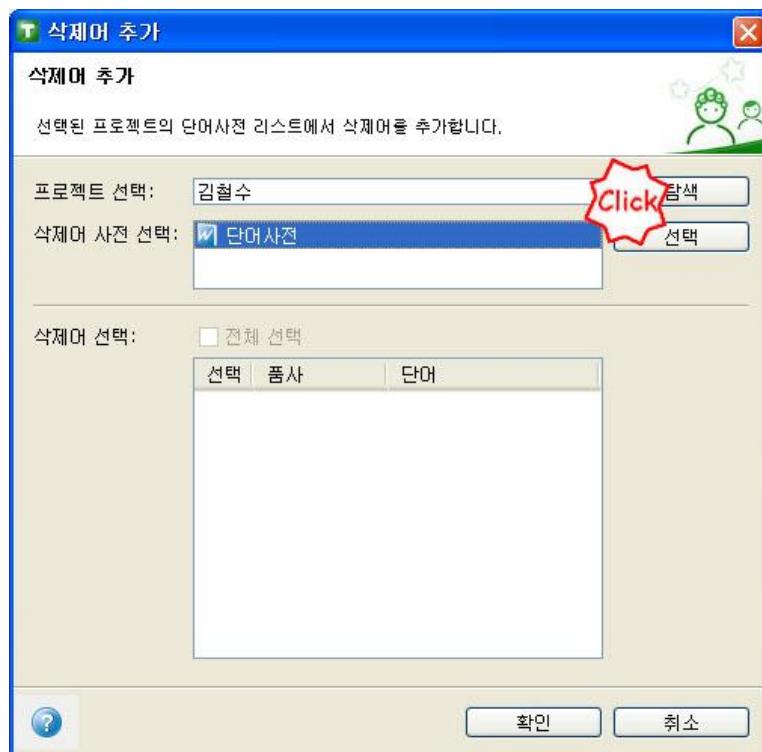
iii. 단어 사전 단어 추가

다양한 프로젝트의 단어 사전으로부터 사용자가 원하는 단어를 선택해 삭제어 사전에 추가합니다. 주의할 점은 선택한 프로젝트에 생성된 단어사전이 반드시 있어야 합니다. 다음은 단어 사전의 단어를 삭제어 사전에 추가하는 과정입니다.

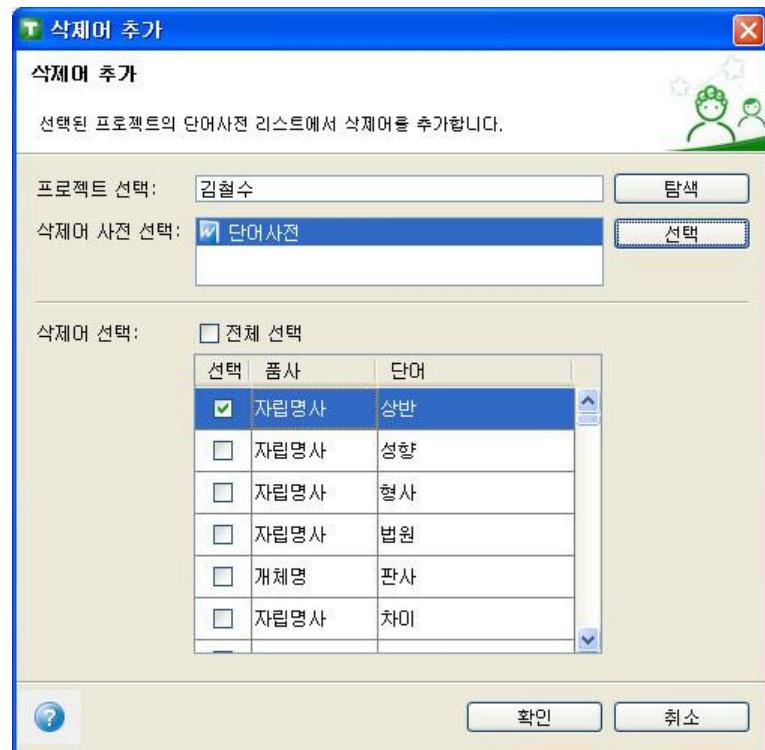
(iii-1) <편집>을 클릭합니다. ‘삭제어 추가’ 대화상자가 나타나면 프로젝트 선택란에 불러오고자 하는 단어 사전이 속해있는 프로젝트를 <탐색>을 눌러 선택한 후 <확인>을 누릅니다.



(iii-2) 사전 선택란에는 불러온 프로젝트가 가지고 있는 단어 사전 목록이 나타납니다. 원하는 단어 사전을 클릭하고 <선택>을 누릅니다.



(iii-3) 삭제어 선택란에는 선택한 단어 사전의 단어들이 나열됩니다. 사용자가 추가하고자 하는 단어에 체크표시를 하고 <확인>을 누릅니다.



자립명사인 ‘상반’을 삭제어 사전에 추가

(iii-4) 속성의 ‘단어수’에서 단어가 추가된 것을 확인할 수 있습니다.

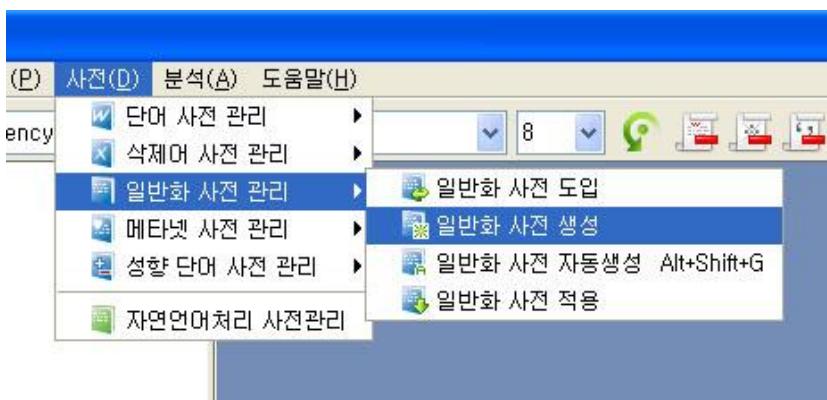
3. 일반화 사전

일반화 사전은 동일한 의미를 가진 단어군을 하나의 단어로 나타내기 위하여 사용자가 지정한 단어와 동일 의미의 단어군의 쌍을 모아놓은 사전입니다. 일반화 사전을 만드는 방법에는 (1) 사용자가 직접 단어를 추가하거나 삭제하기 위해 초기화된 사전을 만드는 ‘일반화 사전 생성’과 (2) 원문에서 일반 어를 추가하는 ‘일반화 사전 원문에서 생성’ (3) 프로그램을 통해 자동으로 생성하는 ‘일반화 사전 자동생성’ (4) 사용자의 필요에 따라 기존에 생성되었던 사전의 파일을 컴퓨터에서 불러오는 ‘일반화 사전 도입’이 있습니다. 일반화 사전을 생성한 후 사용자가 이를 문서에 적용시키고자 한다면 ‘일반화 사전 적용’을 실행시킵니다.

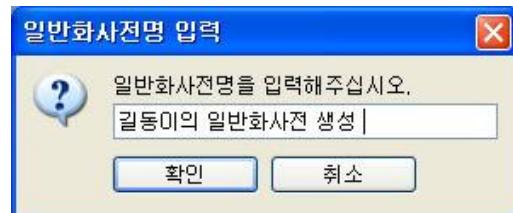
a) 일반화 사전 생성

‘일반화 사전 생성’은 사용자가 직접 단어를 추가/삭제하기 위하여 초기화된 일반화 사전을 생성합니다. 다음은 ‘일반화 사전 생성’을 실행하는 과정입니다.

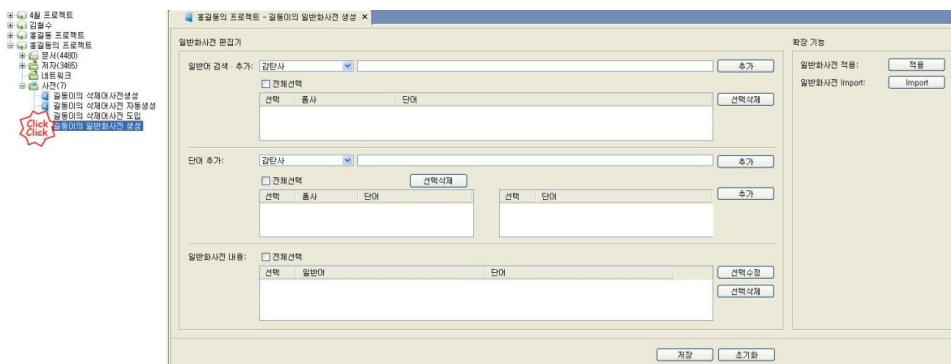
(a-1) ‘일반화 사전 생성’을 실행하려면 (방법 1) **메뉴 막대 – 사전 – 일반화 사전 관리 – 일반화 사전 생성**을 누릅니다. 또는 (방법 2) 컴포넌트 트리의 ‘사전’에 오른쪽 마우스를 클릭한 후 ‘일반화 사전 생성’을 누릅니다.



(a-2) ‘일반화 사전명 입력’ 대화상자가 나타나면 기입란에 본인이 원하는 이름을 쓰고 <확인>을 누릅니다.



(a-3) 이후 컴포넌트 트리의 ‘사전’에 일반화 사전이 생성됩니다. 생성된 일반화 사전을 더블 클릭하면 작업창에 일반화 사전 편집 탭이 나타나며 원하는 단어를 추가하거나 삭제할 수 있습니다.



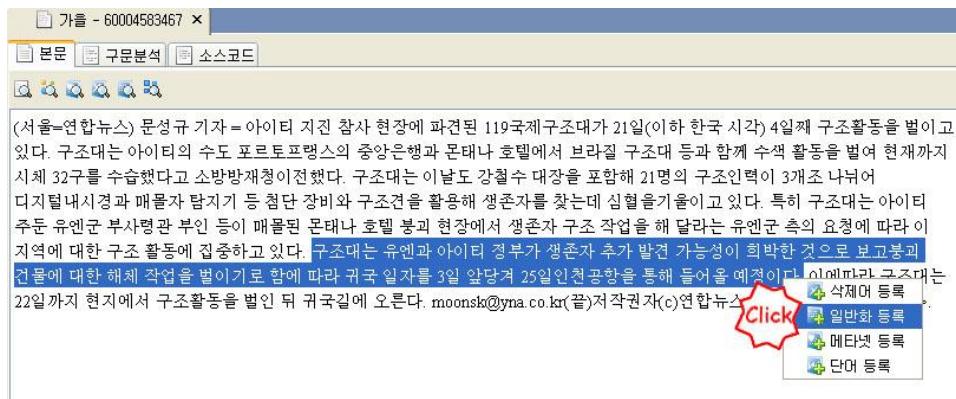
(a-4) 생성된 일반화 사전은 파일의 형태로 사용자가 지정한 작업폴더(최초 설정- C://Workplace, page 29, (d-2) 참조)에 자동 저장되며 메시지 패널의 속성에서 저장된 위치 등 상세정보를 확인할 수 있습니다.

(a-5) 명령이 성공적으로 수행되었다면 속성에는 ***** Successfully executed ***** 가 표기됩니다.

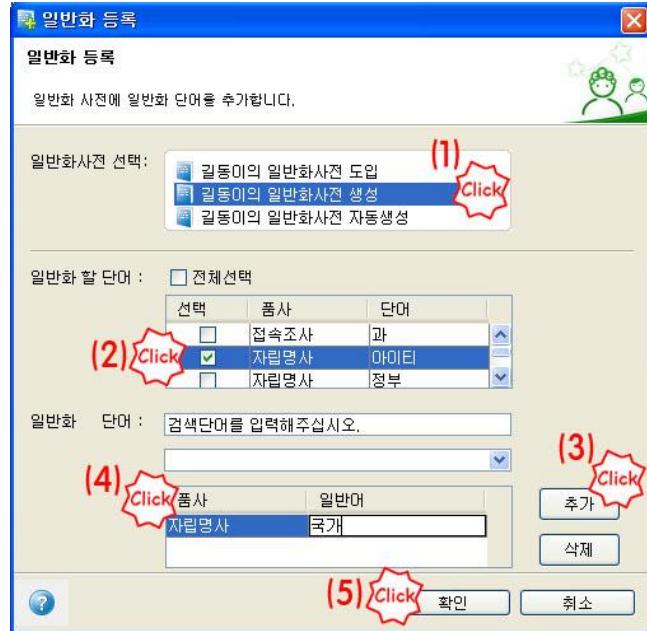
b) 일반화 사전 원문에서 생성

‘일반화 사전 원문에서 생성’은 원문에 드래그를 하여 단어를 추가하는 방식으로 사용자가 원문에서 직접 일반어를 등록하여 사전을 생성하거나 기존에 생성된 일반화 사전에 일반어와 단어의 쌍을 추가합니다. 다음은 ‘일반화 사전 원문에서 생성’을 실행하는 과정입니다.

(b-1) 컴포넌트 트리 ‘문서’에서 추가하고자 하는 일반어가 포함되어 있는 문서를 작업창에 새 탭을 생성합니다. 일반어로 추가하고자 하는 단어나 문장에 마우스로 드래그합니다. 이후 마우스 오른쪽버튼을 클릭하여 ‘일반어 등록’을 누릅니다.



(b-2) ‘일반화 등록’ 대화상자 가 나타나면 ‘사전 선택’에서 추가하려는 일반화 사전을 클릭합니다. ‘일반화 할 단어’란은 사용자가 드래그를 한 단어/문장의 품사가 자동으로 구분이 되어 나타납니다. ‘일반화 할 단어’ 란에 추가하고자 하는 단어를 체크한 후 오른쪽 하단의 <추가>버튼을 누릅니다.



자립명사인 ‘아이티’를 일반어로 선택

‘일반화 단어’의 ‘품사’란에 일반어의 해당 품사를 선택한 후 ‘일반어’란에 일반어를 기입한 후 Enter 를 누른 후 <확인>을 클릭합니다.

(b-3) 컴포넌트 트리의 ‘사전’에서 해당 일반화 사전을 더블 클릭하여 작업창에 새 탭을 생성합니다. 일반어로 선택한 단어가 일반화 사전에 등록된 것을 확인합니다.

일반화사전 내용:		
선택	일반어	단어
<input type="checkbox"/>	국제기구 : (자립명사)국제기구	유엔 : (자립명사)유엔
<input checked="" type="checkbox"/>	국가 : (자립명사)국가	아이티 : (자립명사)아이티

기존의 삭제어 사전에 추가한 경우

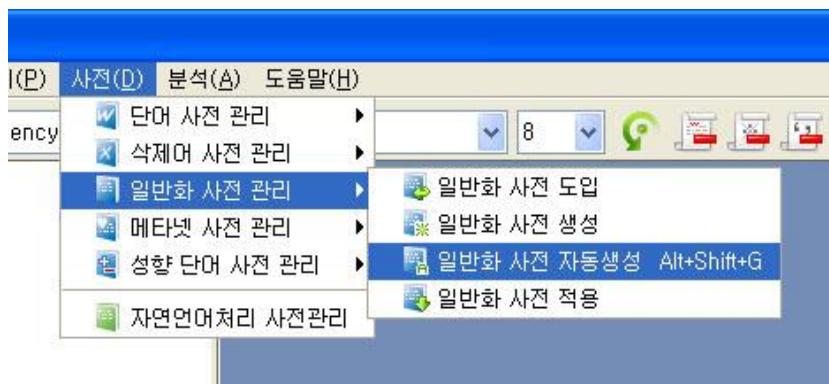
(b-4) 생성된 일반화 사전은 파일의 형태로 사용자가 지정한 작업폴더(최초 설정- C://Workplace, page 29, (d-2) 참조)에 자동 저장되며 메시지 판넬의 속성에서 저장된 위치 등 상세정보를 확인할 수 있습니다.

(b-5) 명령이 성공적으로 수행되었다면 속성에는 ******* Successfully executed ******* 가 표기됩니다.

c) 일반화 사전 자동생성

일반화 사전을 프로그램을 통해 생성합니다. 다음은 ‘일반화 사전 자동생성’을 실행하는 과정입니다.

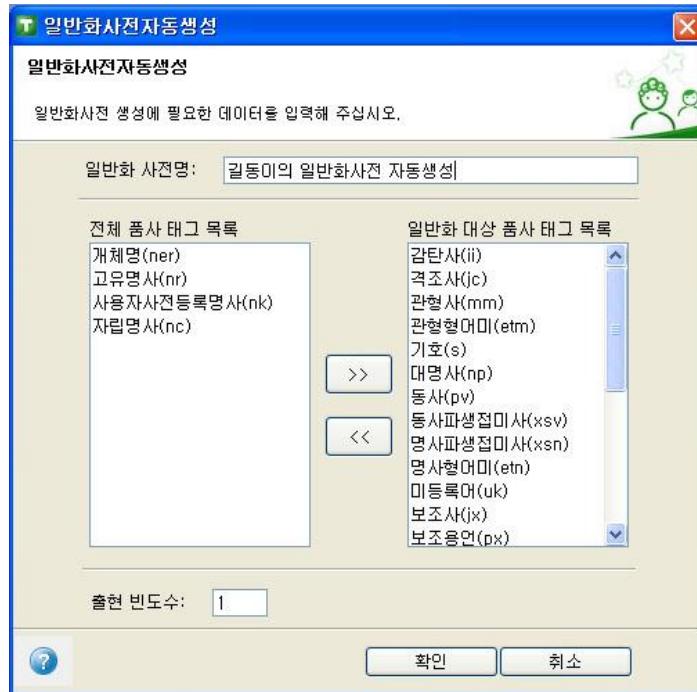
(c-1) ‘일반화 사전 생성’을 실행하려면 **메뉴막대 – 사전 – 일반화 사전 관리 – 일반화 사전 자동생성을** 누릅니다.



(c-2)

- ‘일반화 자동생성’ 대화상자가 나타나면 ‘일반화 사전명’에는 본인이 원하는 제목을 씁니다.
- ‘전체 품사 태그 목록’에서 분석하고자 하는 품사를 클릭한 후 추가버튼을 눌러 ‘대상 품사 태그 목록’으로 이동시킵니다. 만약 이동 시킨 품사를 삭제하고자 하는

경우에는  삭제버튼을 누릅니다. 하지만 사용자가 분석하기 용이하도록 이상적인 수치와 품사를 기본적으로 나타내주기 때문에 특정한 목적 이외에는 그대로 사용합니다.

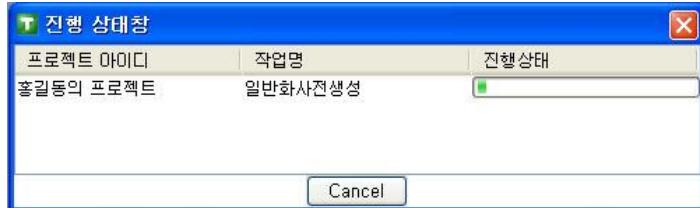


iii. 출현 빈도수는 문서 집합에서 출현한 횟수입니다.

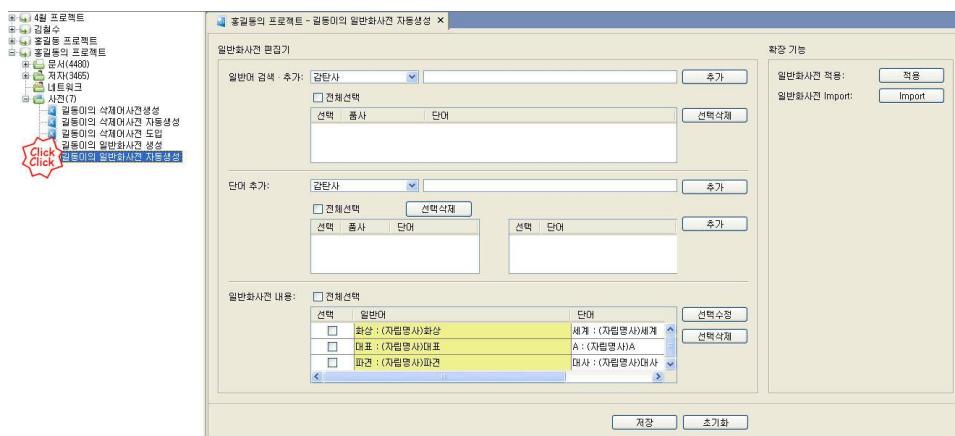
동시 출현 빈도수가 높을수록 두 단어 사이의 연관관계가 강해지므로 한 단어가 다른 단어의 동일단어일 가능성도 높아집니다. 설정이 끝나면 확인 버튼을 클릭합니다.

(c-3) ‘진행 상태창’ 대화상자가 나타나면 연두색 막대가 채워질 때까지 기다립니다. 이때 <Cancel>을 클릭하면 ‘일반화 사전 생성’이 중단됩니다.

¹ 일반적으로 명사를 중심으로 한 토픽을 얻는 것이 전체 주제를 파악하는데 용이하므로 명사관련 품사를 제외한 모든 품사가 주로 선택됩니다. 품사에 대한 보다 자세한 정보는 페이지 360 쪽을 참고합니다.



(c-4) 작업이 완료되면 컴포넌트 트리의 ‘사전’에 일반화 사전이 생성됩니다. 생성된 일반화 사전을 더블 클릭하면 작업창에 일반화 사전 편집 탭이 나타나며 프로그램에서 추출한 단어 이외에 사용자가 원하는 단어를 추가하거나 삭제할 수 있습니다.



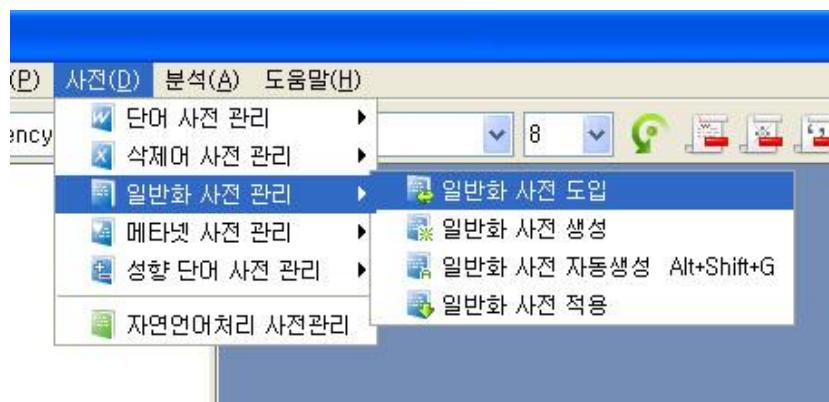
(c-5) 생성된 일반화 사전은 파일의 형태로 사용자가 지정한 작업폴더(최초 설정-C://Workplace, page 29, (d-2) 참조)에 자동 저장되며 메시지 판넬의 속성에서 저장된 위치 등 상세정보를 확인할 수 있습니다.

(c-6) 명령이 성공적으로 수행되었다면 속성에는 ***** Successfully executed ***** 가 표기됩니다.

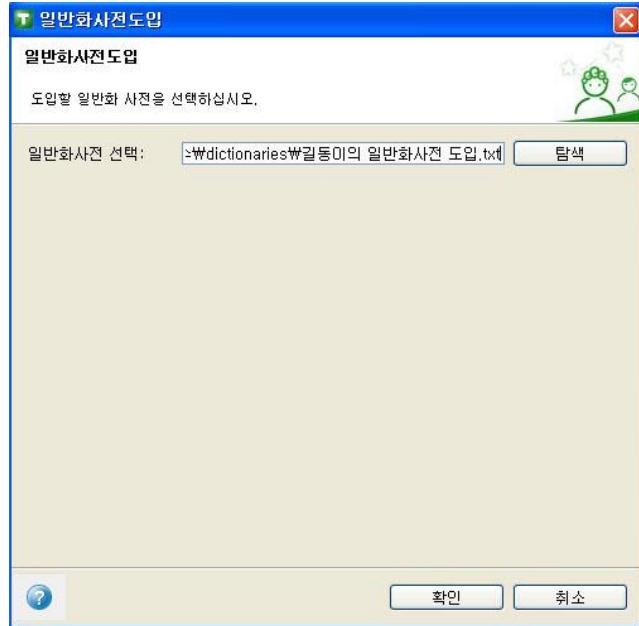
d) 일반화 사전 도입

기존에 생성되었던 일반화 사전의 파일을 컴퓨터에서 불러옵니다. 다음은 ‘일반화 사전 도입’을 실행하는 과정입니다.

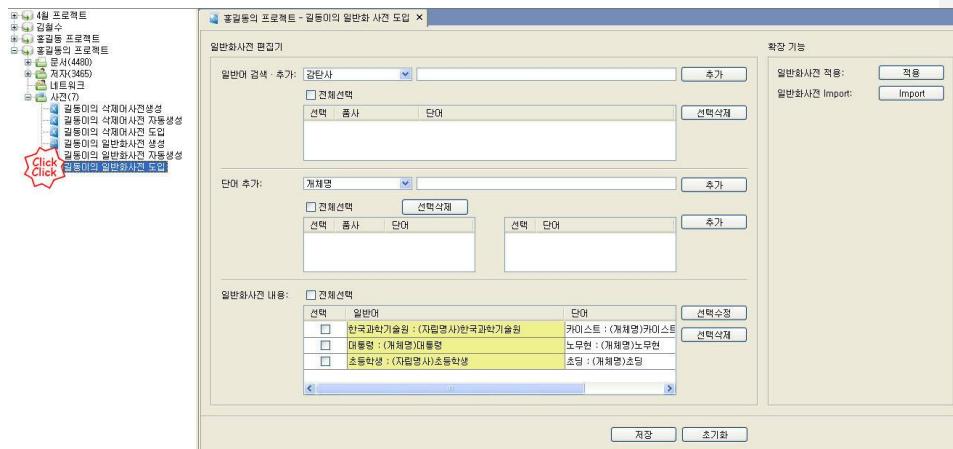
(d-1) ‘일반화 사전 도입’을 실행하려면 **메뉴막대 – 사전 – 일반화 사전 관리 – 일반화 사전 도입**을 누릅니다.



(d-2) ‘일반화 사전 도입’ 대화상자가 나타나면 <탐색>을 눌러 ‘사전’로 불러올 파일을 선택한 후 <확인>을 클릭합니다.



(d-3) 이후 컴포넌트 트리의 ‘사전’에 일반화 사전이 생성됩니다. 불러온 일반화 사전을 더블 클릭하면 작업창에 일반화 사전 편집 탭이 나타나며 사용자가 원하는 단어를 추가하거나 삭제할 수 있습니다.

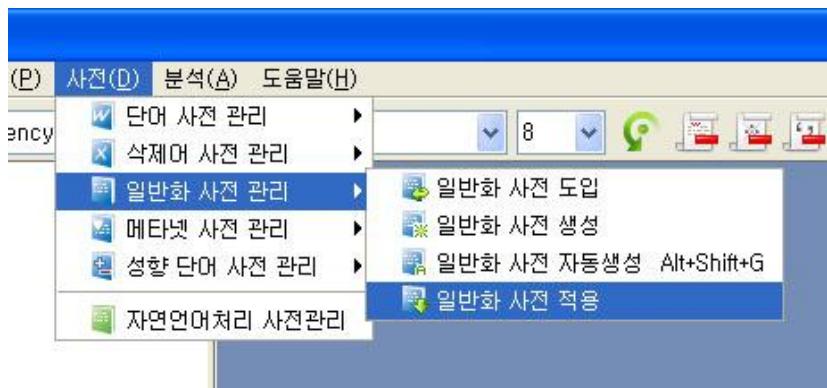


(d-4) 명령이 성공적으로 수행되었다면 속성에는 ***** Successfully executed ***** 가 표기됩니다.

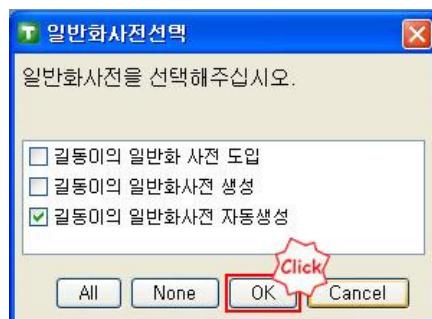
e) 일반화 사전 적용

생성한 일반화 사전을 프로젝트에 적용합니다. 다음은 ‘일반화 사전 적용’을 실행하는 과정입니다.

(e-1) ‘일반화 사전 적용’을 실행하려면 메뉴 막대 – 사전 – 삭제어 사전 관리 – 삭제어 사전 적용을 누릅니다.

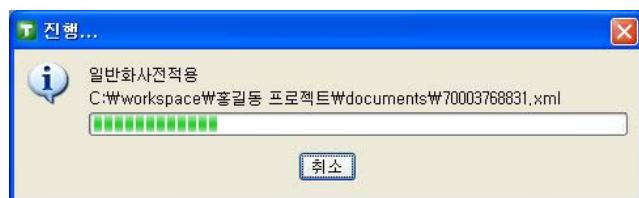


(e-2) ‘일반화 사전 선택’ 대화상자가 나타나면 프로젝트에 적용시킬 일반화 사전을 선택한 후 <OK> 를 클릭합니다.



- i. All : 목록에 있는 사전을 모두 선택합니다.
- ii. None : 체크 표시가 되어 있는 항목을 해제합니다.
- iii. OK : 선택을 완료하고 다음 단계로 넘어갑니다.
- iv. Cancel : ‘일반화 사전 선택’ 대화상자를 취소합니다.

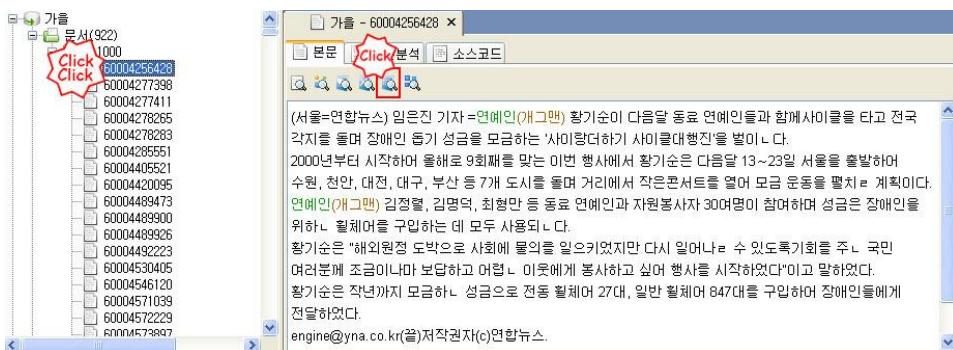
(e-3) ‘진행’ 대화상자가 나타나면 막대가 모두 100%에 도달할 때까지 기다립니다. 이때 <Cancel> 을 클릭하면 ‘일반화 사전 적용’이 중단됩니다.



(e-4) 명령이 성공적으로 수행되었다면 속성에는 ***** Successfully executed ***** 가 표기됩니다.

4. 일반화 사전 처리 결과 보기

일반화 사전을 적용시킨 결과를 확인하려면 **컴포넌트 트리—문서**에서 임의의 문서¹를 더블 클릭합니다. 좌측에 새 탭이 생성되면 **HTML 탭 — 일반화 보기** 아이콘을 클릭하여 단어는 갈색으로, 일반어는 연두색으로 변환된 것을 알 수 있습니다.

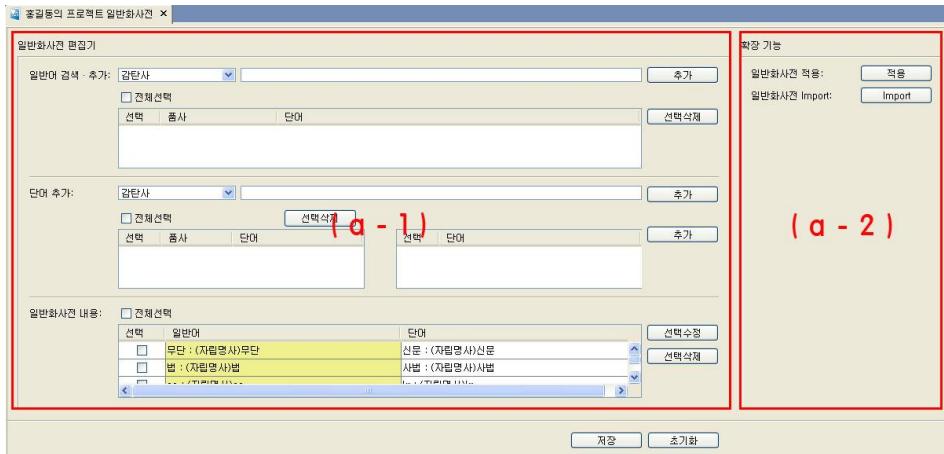


‘사전’에서 일반화 사전을 더블 클릭하면 작업창에 해당 사전을 편집할 수 있는 탭이 나타나며 프로그램에서 추출한 단어 이외에 사용자가 원하는 단어를 추가하거나 삭제할 수 있습니다.

a) 일반화 사전

컴포넌트 트리의 ‘사전’에 있는 임의의 일반화 사전을 클릭했을 때 작업창은 다음과 같습니다.

¹ 사전의 ‘적용’은 ‘문서’에 있는 모든 문서에 적용이 되기 때문에 임의의 문서에서 열람할 수 있습니다.



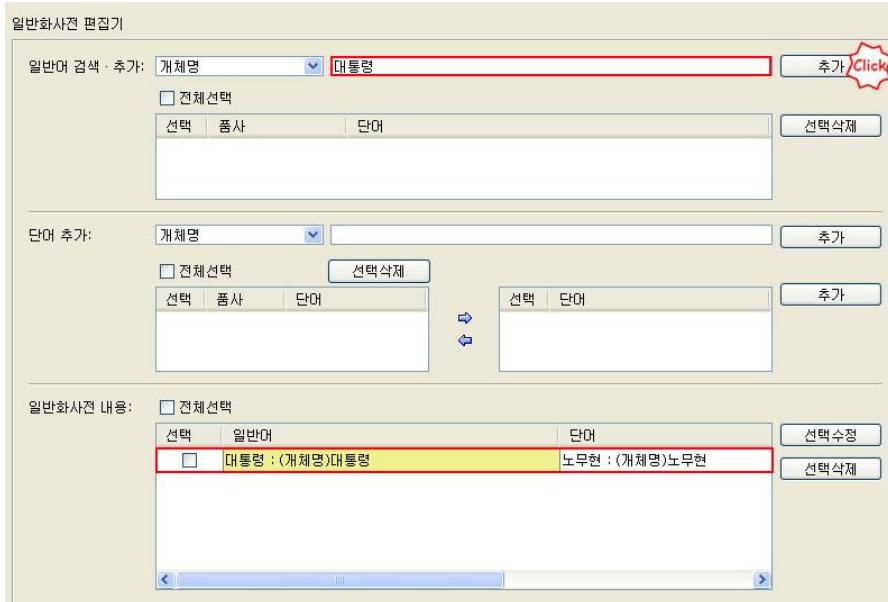
탭의 상단에는 프로젝트 명과 사전종류가 표기되며 하위 메뉴에는 ‘일반화사전 편집’과 ‘확장 기능’ 총 2 가지가 있습니다.

- 일반어는 동일한 의미를 가진 단어(들)이 통일되려는 단어입니다.
- 단어는 일반어로 통일하려는 단어(들)입니다.

(a-1) 일반화사전 편집기

i. 일반어 검색

일반어를 검색하고자 한다면 선택란에 해당하는 품사를 선택한 후 선택란 아래에 있는 일반어란에 찾고자 하는 일반어를 기입하면 하단에 조건에 일치하는 검색어가 나타납니다.



‘대통령’을 검색했을 때

ii. 일반어 추가

새로운 일반어를 생성하고자 한다면 일반어와 단어가 동시에 생성되어야 합니다. 일반화 사전은 일반어와 단어를 쌍을 이루어 모아놓은 사전이기 때문에 일반어에 해당하는 단어 또한 같이 추가가 되어야 합니다. 다음은 일반어를 생성하는 과정입니다.

(ii-1) 선택란에서 해당하는 품사를 선택한 후 선택란 아래에 있는 일반어란에 추가하고자 하는 일반어를 기입합니다. 이후 <확인>을 누르면 하단에 조건에 일치하는 일반어가 생성됩니다. 일반어가 생성되면 이에 해당하는 단어를 생성하기 위해 하단에 ‘단어 추가’로 이동합니다.



개체명인 ‘국가’를 일반어로 추가했을 때

*생성된 일반어를 삭제하고자 하는 경우 선택란에 체크를 한 후 **선택삭제**를 누릅니다.

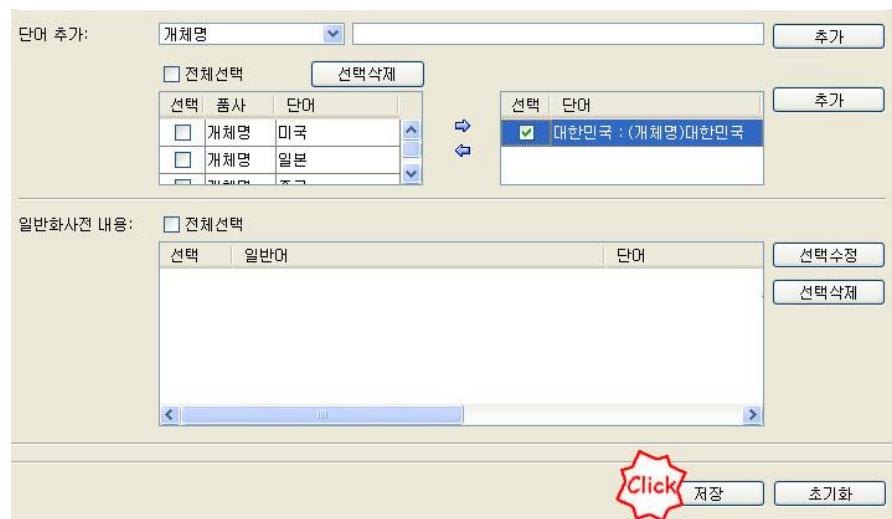
(ii-2) 선택란에 해당하는 품사를 선택한 후 선택란 좌측에 있는 단어란에 생성하고자 하는 단어를 기입하면 하단에 조건에 일치하는 단어가 생성됩니다.



‘대한한국’을 일반어인 ‘대한민국’에 추가하려 할 때

*선택란에 체크를 한 후 **선택삭제**를 누르면 생성된 단어가 삭제됩니다.

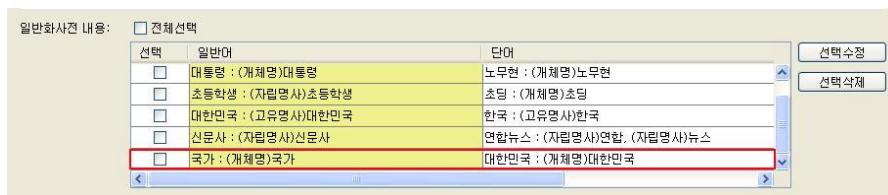
(ii-3) 생성된 단어에 체크를 한 후 **→** 을 클릭하면 단어가 우측으로 이동합니다. 우측란으로 이동된 단어만이 일반어와 쌍을 이루어 사전에 추가될 수 있습니다.



‘대한민국’을 우측으로 이동시켰을 때

단어를 생성했지만 일반화 사전에 추가하는 것을 취소하려면 우측으로 이동한 단어에 체크를 한 후 을 누르면 좌측으로 단어가 이동합니다.

(ii-4) 이후 <확인>을 누르면 하단에 조건에 일치하는 일반어가 생성됩니다. 저장을 클릭하면 생성된 일반어가 일반화 사전에 저장이 됩니다.



iii. 일반어의 단어 추가

이미 생성된 일반어와 단어의 쌍에 여러 개의 단어를 추가할 수 있습니다. (방법 1) (a-1) i.에서 언급한 방식대로 일반어를 검색합니다. 이후 ‘단어 추가’로 이동하여 ii.(ii-2)에서 언급한 방식대로 단어를 추가합니다. 혹은 (방법 2) 하단에서 일반어를 직접 찾아 체크를 한 후 을 클릭합니다. ‘일반화 검색·추가’란에 일반어가 나타나면 ‘단어 선택’으로 이동하여 ii.(ii-2)에서 언급한 방식대로 단어를 추가합니다.

일반어를 직접 찾아 <선택수정>을 클릭한 화면

- i. 선택수정 : 선택란에 체크가 되어 있는 일반어를 수정합니다.
- ii. 선택삭제 : 선택란에 체크가 되어 있는 일반어를 삭제합니다.
- iii. 저장 : 편집된 내용을 저장합니다.
- iv. 초기화 : 명령이 실행되지 않은 초기의 상태로 돌아갑니다.

(a-2) 확장기능



i. 일반화사전 적용

<적용>을 누르면 생성한 일반화 사전을 프로젝트에 적용합니다.

ii. 일반화사전 Import

<Import>를 누르면 생성한 기존에 생성한 일반화 사전을 현재 프로젝트에 가져옵니다.

5. 성향 단어 사전

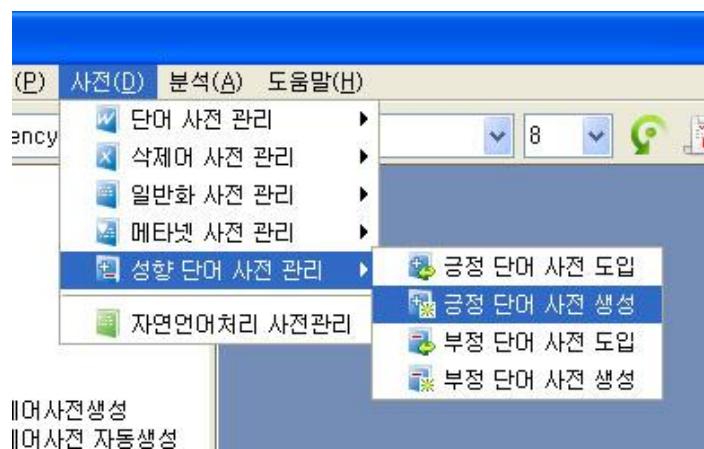
성향 단어 사전은 사용자의 주관에 따라 원문에서 긍정의미를 가진 단어와 부정의미를 가진 단어를 모아둔 사전입니다. 사전의 종류는 2 가지로 긍정 단어사전과 부정 단어사전이 있습니다. **중요한 점은 긍정 및 부정의 의미가 사용자의 입장에서 부정과 긍정의 의미를 갖는다는 것입니다.** 이는 일반적인 긍정 및 부정의 의미가 아니라 만약 부정적인 글이더라도 그 부정적인 의미가 사용자에게 긍정적으로 다가온다면 사용자에게는 긍정단어라는 이라는 뜻입니다.

성향 단어 사전을 만드는 방법에는 (1) 사용자가 직접 단어를 추가하거나 삭제하기 위해 초기화된 사전을 만드는 ‘사전 생성’과 (2) 사용자의 필요에 따라 기존에 생성되었던 사전의 파일을 컴퓨터에서 불러오는 ‘사전 도입’이 있습니다. 성향 단어 사전을 생성한 후 사용자가 이를 문서에 적용하고자 한다면 ‘저자-이슈 성향 분석’을 실행합니다.

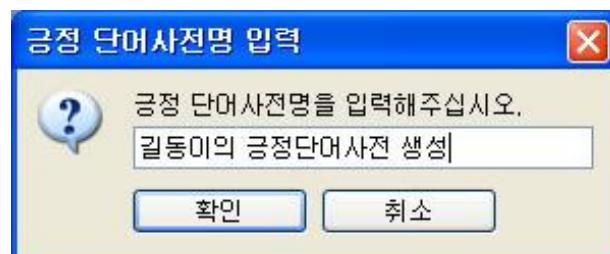
a) 긍정 단어 사전 생성

‘긍정 단어 사전 생성’은 사용자가 직접 단어를 추가/삭제하기 위하여 초기화된 긍정 단어 사전을 생성합니다. 다음은 ‘긍정 단어 사전 생성’을 실행하는 과정입니다.

(a-1) (방법 1) ‘긍정 단어 사전 생성’을 실행하려면 **메뉴막대 – 사전 – 성향 단어 사전 관리 – 긍정 단어 사전 생성**을 누릅니다. 또는 (방법 2) 컴포넌트 트리의 ‘사전’에 오른쪽 마우스를 클릭한 후 ‘긍정 단어사전 생성’을 누릅니다.



(a-2) ‘긍정 단어 사전명 입력’ 대화상자가 나타나면 기입란에 본인이 원하는 이름을 쓰고 <확인>을 누릅니다.



(a-3) 이후 컴포넌트 트리의 ‘사전’에 긍정 단어 사전이 생성됩니다. 생성된 긍정 단어 사전을 더블 클릭하면 작업창에 긍정 단어 사전 편집탭이 나타나며 원하는 단어를 추가하거나 삭제할 수 있습니다.



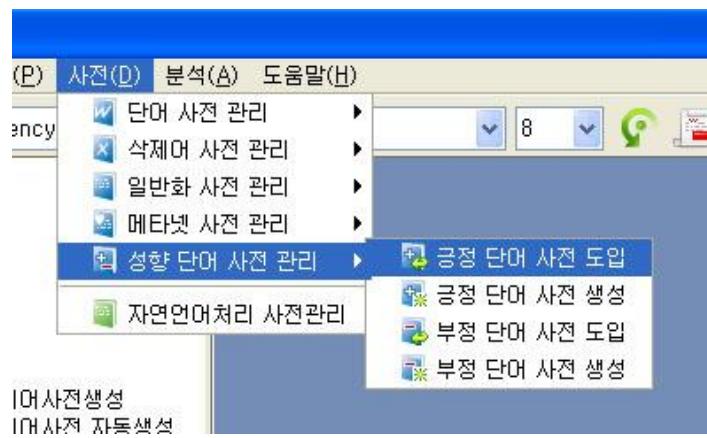
(a-4) 생성된 긍정 단어 사전은 파일의 형태로 사용자가 지정한 작업폴더(최초 설정-C:/Workplace, page 29, (d-2) 참조)에 자동 저장되며 메시지 판넬의 속성에서 저장된 위치 등 상세정보를 확인할 수 있습니다.

(a-5) 명령이 성공적으로 수행되었다면 속성에는 ***** Successfully executed ***** 가 표기됩니다.

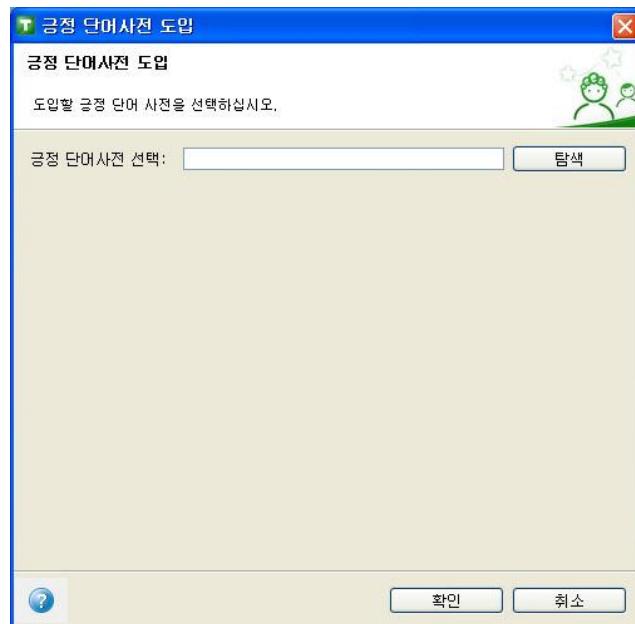
b) 긍정 단어 사전 도입

기존에 생성되었던 긍정 단어 사전의 파일을 컴퓨터에서 불러옵니다. 다음은 ‘긍정 단어 사전 도입’을 실행하는 과정입니다.

(b-1) ‘긍정 단어 사전을 도입’을 실행하려면 **메뉴막대 – 사전 – 성향 단어 사전 관리 – 긍정 단어 사전 도입**을 누릅니다.



(d-2) '긍정 단어사전 도입' 대화상자가 나타나면 <탐색>을 눌러 '사전'로 불러올 파일을 선택한 후 <확인>을 클릭합니다.



(d-3) 이후 컴포넌트 트리의 ‘사전’에 긍정 단어 사전이 생성됩니다. 불러온 긍정 단어 사전을 더블 클릭하면 작업창에 긍정 단어 사전 편집 탭이 나타나며 사용자가 원하는 단어를 추가하거나 삭제할 수 있습니다.

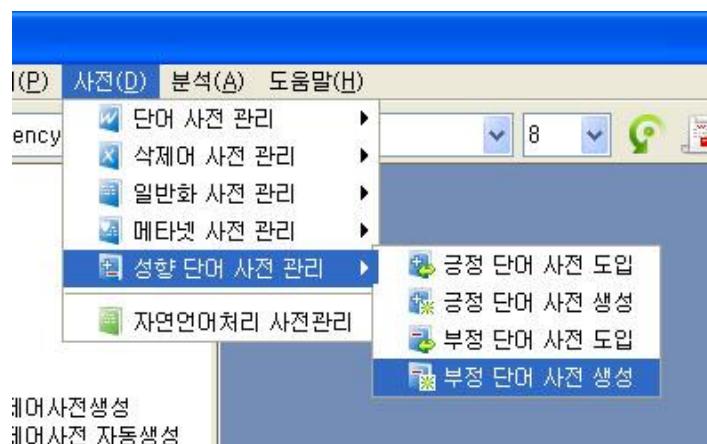


(d-4) 명령이 성공적으로 수행되었다면 속성에는 ***** Successfully executed ***** 가 표기됩니다.

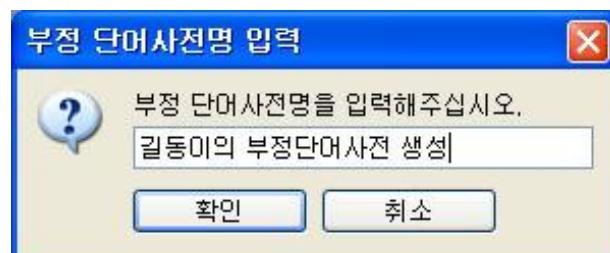
c) 부정 단어 사전 생성

‘부정 단어 사전 생성’은 사용자가 직접 단어를 추가/삭제하기 위하여 초기화된 긍정 단어 사전을 생성합니다. 다음은 ‘부정 단어 사전 생성’을 실행하는 과정입니다.

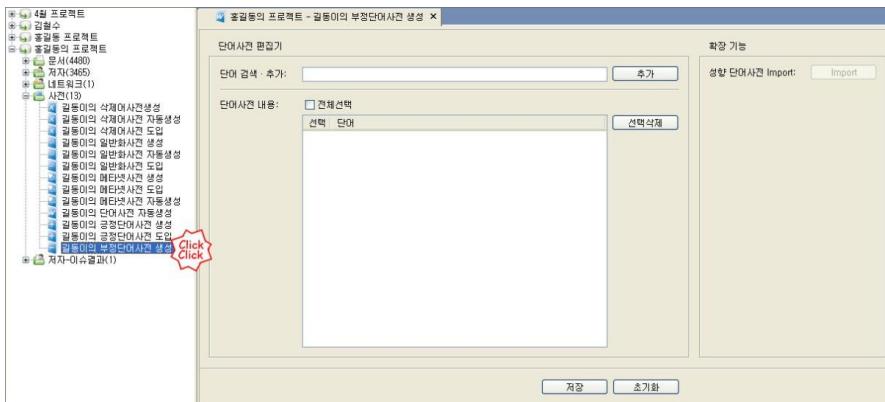
(c-1) (방법 1) ‘부정 단어 사전 생성’을 실행하려면 **메뉴막대 – 사전 – 성향 단어 사전 관리 – 부정 단어 사전 생성**을 누릅니다. 또는 (방법 2) 컴포넌트 트리의 ‘사전’에 오른쪽 마우스를 클릭한 후 ‘부정 단어사전 생성’을 누릅니다.



(c-2) ‘부정 단어 사전명 입력’ 대화상자가 나타나면 기입란에 본인이 원하는 이름을 쓰고 <확인>을 누릅니다.



(c-3) 이후 컴포넌트 트리의 ‘사전’에 부정 단어 사전이 생성됩니다. 생성된 부정 단어 사전을 더블 클릭하면 작업창에 부정 단어 사전 편집탭이 나타나며 원하는 단어를 추가하거나 삭제할 수 있습니다.



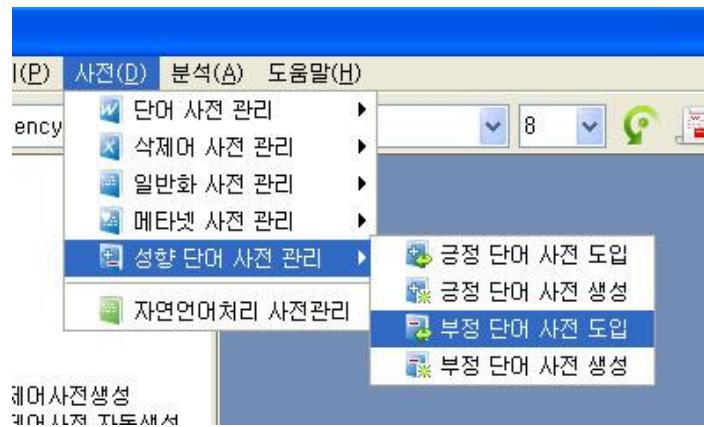
(c-4) 생성된 부정 단어 사전은 파일의 형태로 사용자가 지정한 작업폴더(최초 설정-C://Workplace, page 29, (d-2) 참조)에 자동 저장되며 메시지 패널의 속성에서 저장된 위치 등 상세정보를 확인할 수 있습니다.

(c-5) 명령이 성공적으로 수행되었다면 속성에는 ***** Successfully executed ***** 가 표기됩니다.

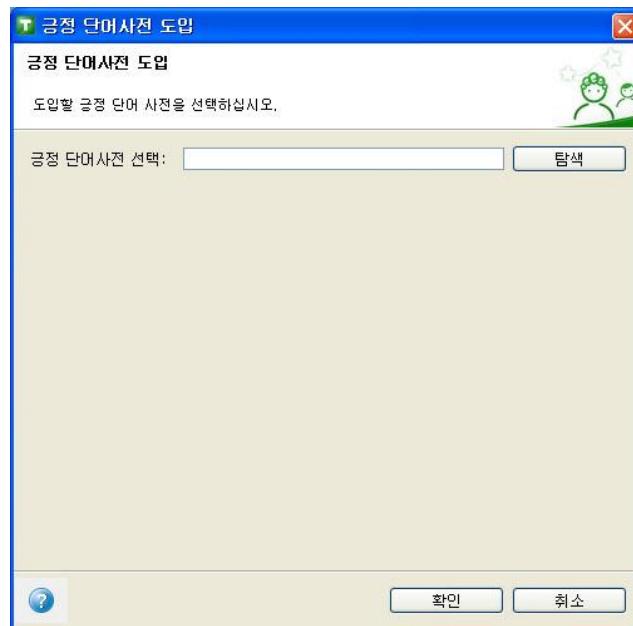
d) 부정 단어 사전 도입

기존에 생성되었던 긍정 단어 사전의 파일을 컴퓨터에서 불러옵니다. 다음은 ‘부정 단어 사전 도입’을 실행하는 과정입니다.

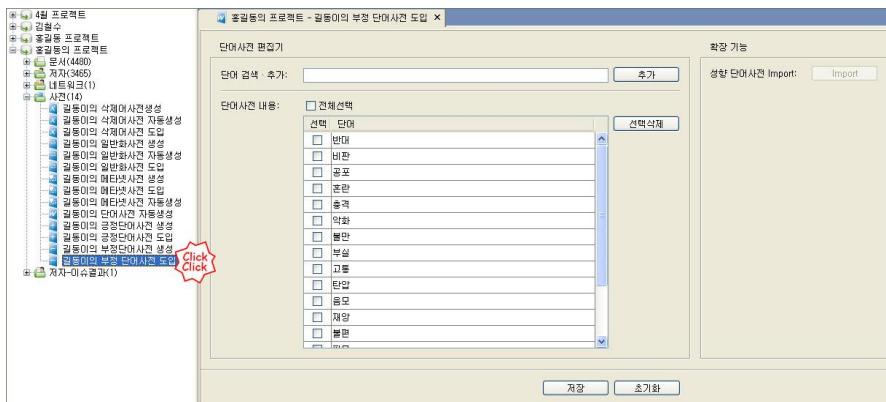
(d-1) ‘부정 단어 사전을 도입’을 실행하려면 **메뉴막대 – 사전 – 성향 단어 사전 관리 – 부정 단어 사전 도입**을 누릅니다.



(d-2) ‘부정 단어사전 도입’ 대화상자가 나타나면 <탐색>을 눌러 ‘사전’로 불러올 파일을 선택한 후 <확인>을 클릭합니다.



(d-3) 이후 컴포넌트 트리의 ‘사전’에 부정 단어 사전이 생성됩니다. 불러온 부정 단어 사전을 더블 클릭하면 작업창에 부정 단어 사전 편집 탭이 나타나며 사용자가 원하는 단어를 추가하거나 삭제할 수 있습니다.



(d-4) 명령이 성공적으로 수행되었다면 속성에는 ***** Successfully executed ***** 가 표기됩니다.

E. 단어 사전 생성하기

1. 단어 사전

단어 사전은 사용자가 필요로 하는 단어들만 모아 놓은 사전입니다. 여타 사전들과는 달리 단어 사전은 ‘적용’의 기능이 없고 바로 ‘저자-이슈 분석’이나 ‘저자-이슈-성향 분석’으로 이어지기 때문에 맨 마지막에 적용되어야 하며 또한 필수적으로 만들어져야 합니다. 단어 사전을 만드는 방법에는 프로그램을 통해 자동으로 생성하는 ‘단어사전 자동생성¹’을 사용합니다. 단어 사전을 생성한 후 사용자가 이를 문서에 적용시키는 내용은 별도로 <III.F. 저자-이슈 분석 실행하기, page 135>, <III. 저자-이슈-성향 분석 실행하기, page 170>에서 다루겠습니다.

일반적으로 고려해보면 여타 사전들처럼 단어 사전도 적용을 실행시켜야 하지 않을까라는 추측도 할 수 있으나 특이하게도 단어사전은 ‘자동생성’만 존재하고 ‘적용’ 기능은 존재하지 않습니다. 대신 ‘자동생성’을 통해 수집된 단어들이 ‘저자-이슈 분석’과 ‘저자-이슈-성향 분석’ 과정에서 쓰이게 됩니다.

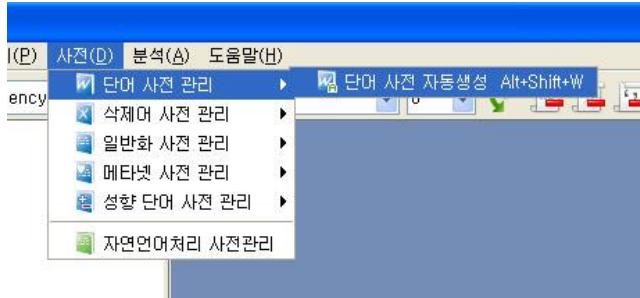
사전들은 토픽을 추출하는 목적 이외에 다른 용도로 쓰일 수 있는 기능이 아직까지는 없기 때문에 단어사전에 따로 ‘적용’이라는 기능을 만들어 놓고 다시 ‘저자-이슈 분석’과 ‘저자-이슈-성향 분석’을 실행하는 것이 아니라, 조금이라도 절차를 덜기 위해 바로 ‘저자-이슈 분석’과 ‘저자-이슈-성향 분석’으로 넘어갑니다. 즉, 사전의 ‘적용’이란 개념을 단어사전에서는 분석 목적에 따라 ‘저자-이슈 분석’이나 ‘저자-이슈-성향 분석’으로 간주할 수 있습니다.

a) 단어 사전 자동생성

단어사전을 프로그램을 통해 생성합니다.

(a-1) ‘단어사전 자동생성’을 실행하려면 **메뉴막대 – 사전 – 단어 사전 관리 – 단어 사전 자동생성을 누릅니다.**

¹ 단어사전은 ‘생성’과 ‘도입’ 기능이 존재하지 않습니다.



(a-2)

- i. ‘단어 사전 자동생성’ 대화상자가 나타나면 ‘단어 사전명’에는 본인이 원하는 이름을 줍니다.
- ii. ‘전체 품사 태그 목록’에서 분석하고자 하는 품사를 클릭한 후 추가버튼을 눌러 ‘대상 품사 태그 목록’으로 이동시킵니다. 만약 이동 시킨 품사를 삭제하고자 하는 경우에는 삭제버튼을 누릅니다. 하지만 사용자가 분석하기 용이하도록 이상적인 수치와 품사¹를 기본적으로 나타내주기 때문에 특정한 목적 이외에는 그대로 사용합니다.

¹ 일반적으로 형용사나 동사보다는 명사를 추출 하는 것이 전체 주제를 파악하는데 용이하므로 고유명사, 대명사, 사용자 사전 등록 명사, 의존명사, 자립명사가 주로 사용됩니다. 품사에 대한 보다 자세한 정보는 페이지 360 쪽을 참고합니다.



- i. TF-IDF는 한 문서 내에서 단어가 쓰인 횟수와 단어가 전체 문서 집합에서 등장하는 횟수의 역수를 곱한 것으로 수치가 높을수록 해당 단어가 특정 문서에서 자주 쓰였다는 것을 나타냅니다. 즉, TF-IDF 값이 높을수록 그 단어는 특정 이슈에서 자주 쓰였을 가능성이 높습니다. 주의할 사항은 단어 사전에서의 TF-IDF 값을 적용하는 방식은 삭제어 사전 자동생성과는 달리 사용자가 지정한 범위 안에서 적용된다는 것입니다. 다시 말하면, 사용자가 지정한 범위에 속한 값을 가진 단어들이 포함됩니다. 아래의 붉은 막대로 표현된 부분이 사용자가 지정한 구간으로 단어 사전에 추가됩니다. 만약 범위를 0.0 과 1.0 을 지정한다면 모든 단어가 포함됩니다.



- ii. DF : 한 단어가 등장하는 문서의 수를 나타냅니다. DF 값이 높으면 아주 일반적인 단어를 의미하는 것이고 DF 값이 낮으면 매우 특수한 경우의 단어를 의미합니다. 일반적으로 우리가 관심 있어하는 단어는 중간 정도의 값을 지닌 DF입니다. DF 값이 매우 높은 경우는 -은, -는,

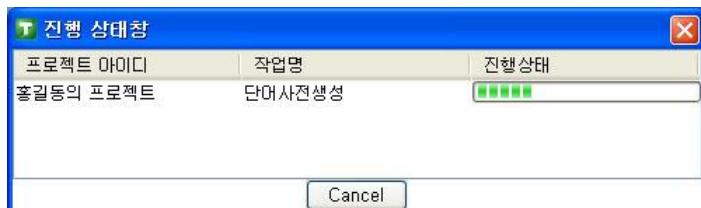
-이, -가와 같이 문서의 성향과 주제에 상관없이 자주 들어가는 단어를 의미하게 되고, DF 값이 매우 낮은 경우에는 오·탈자가 섞인 단어, 혹은 매우 특수한 경우에만 쓰이는 단어가 선별되게 됩니다. **그러므로 중간 정도의 범위를 지닌 DF 값으로 단어사전을 생성하는 것이 중요합니다.** 아래의 붉은 막대로 표현된 부분이 사용자가 지정한 부분으로 단어 사전에 추가됩니다.



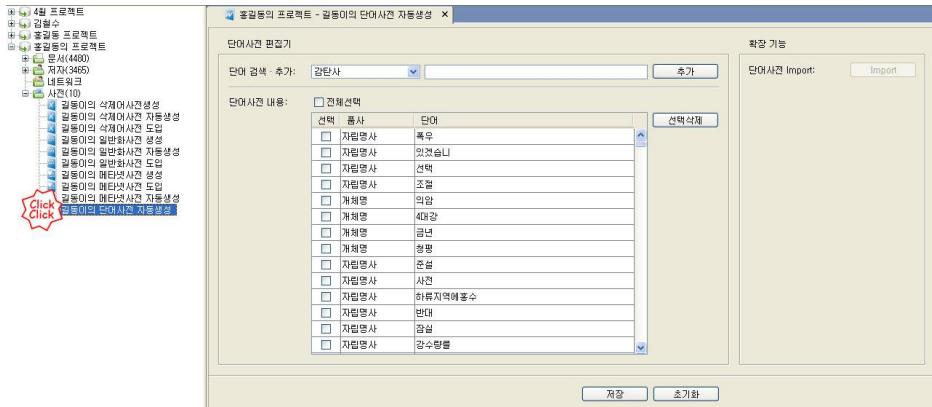
iii. 최소길이는 단어의 최소 길이입니다.

설정이 끝나면 <확인>을 클릭합니다.

(a-3) ‘진행 상태창’ 대화상자가 나타나면 연두색 막대가 채워질 때까지 기다립니다. 이때 <Cancel>을 클릭하면 ‘자연어 처리’가 중단됩니다.



(a-4) 작업이 완료되면 컴포넌트 트리의 ‘사전’에 단어 사전이 생성됩니다. 생성된 단어 사전을 더블 클릭하면 작업창에 단어사전 편집 탭이 나타나며 프로그램에서 추출한 단어 이외에 사용자가 원하는 단어를 추가하거나 삭제할 수 있습니다.

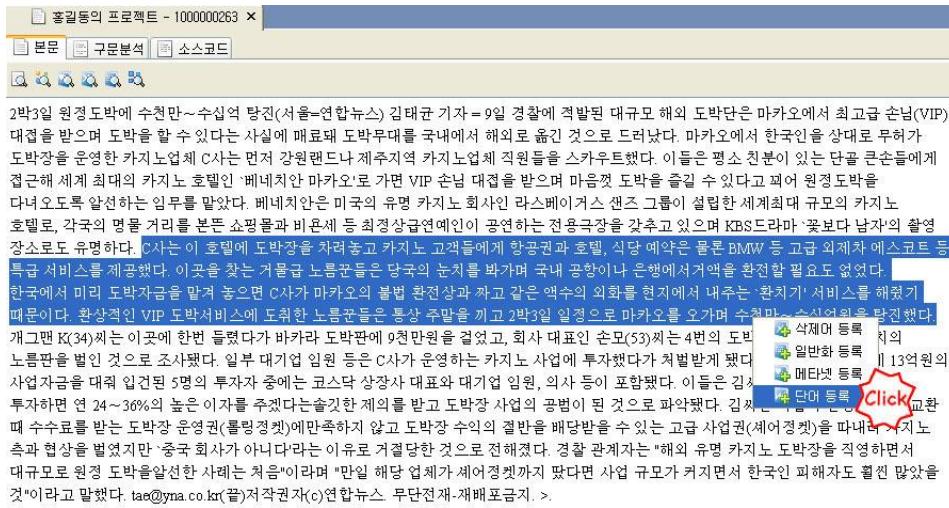


(a-5) 명령이 성공적으로 수행되었다면 속성에는 ***** Successfully executed ***** 가 표기됩니다.

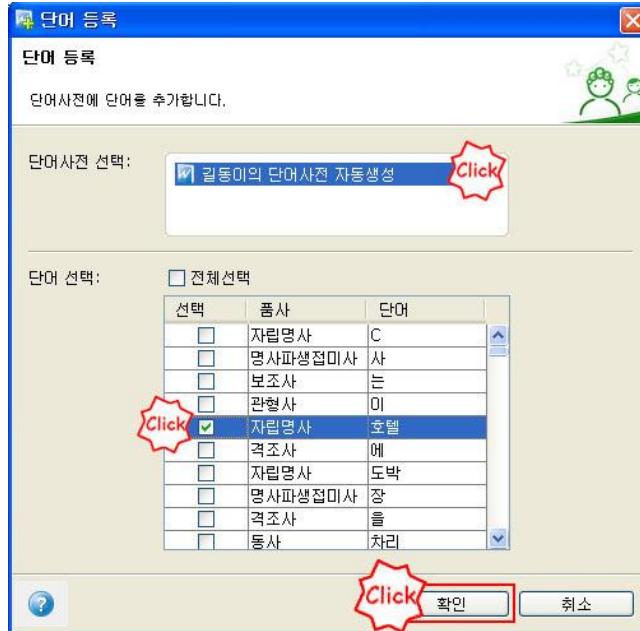
b) 단어 사전 원문에서 생성

'단어 사전 원문에서 생성'은 원문에 드래그를 하여 단어를 추가하는 방식으로 사용자가 원문에서 직접 단어를 등록하여 사전을 생성하거나 기존에 생성된 단어 사전에 단어를 추가합니다. 다음은 '단어 사전 원문에서 생성'을 실행하는 과정입니다.

(b-1) 컴포넌트 트리 '문서'에서 임의의 문서를 클릭하여 작업창에 새 탭을 생성합니다. 단어로 추가하고자 하는 단어나 문장에 마우스로 드래그합니다. 이후 마우스 오른쪽 버튼을 클릭하여 '단어 등록'을 누릅니다.

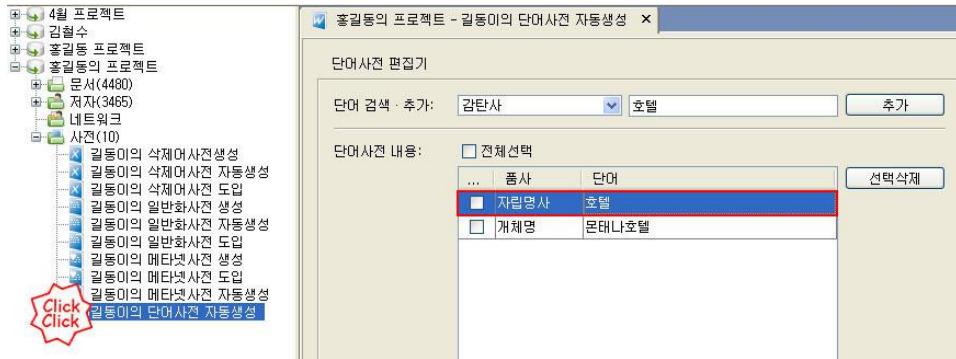


(b-2) ‘단어 등록’ 대화상자가 나타나면 ‘사전 선택’에서 추가하려는 단어 사전을 클릭합니다. ‘단어 선택’란은 사용자가 드래그를 한 단어/문장의 품사가 자동으로 구분이 되어 나타납니다. 추가하고자 하는 단어에 체크를 한 후 <확인>을 누릅니다.



자립명사인 ‘호텔’을 단어로 선택

(b-3) 컴포넌트 트리의 ‘사전’에서 해당 단어 사전을 더블 클릭하여 작업창에 새 탭을 생성합니다.
선택한 단어가 단어 사전에 등록된 것을 확인합니다.



기존의 단어사전에 추가된 ‘호텔’

(b-4) 생성된 단어 사전은 파일의 형태로 사용자가 지정한 작업폴더(최초 설정- C://Workplace, page 29, (d-2) 참조)에 자동 저장되며 메시지 패널의 속성에서 저장된 위치 등 상세정보를 확인할 수 있습니다.

(b-5) 명령이 성공적으로 수행되었다면 속성에는 ***** Successfully executed ***** 가 표기됩니다.

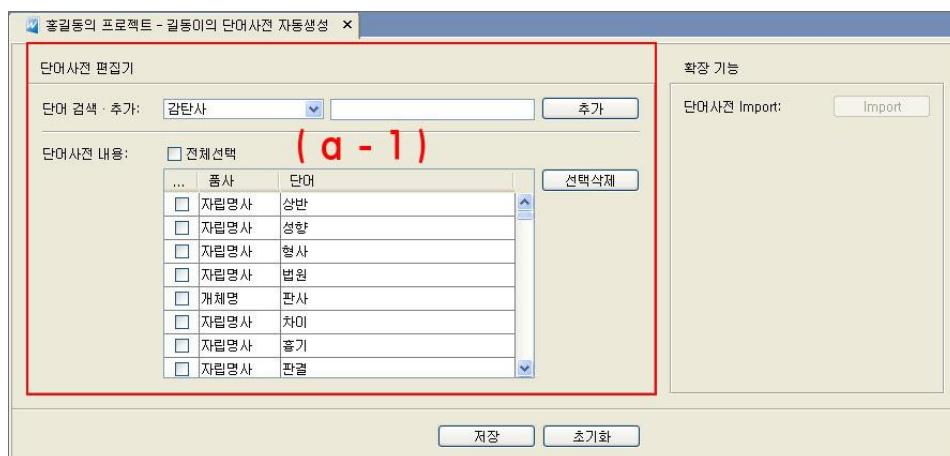
2. 단어 사전 처리결과 보기

단어 사전의 ‘적용’이란 개념은 ‘저자-이슈 분석’이나 ‘저자-이슈-성향 분석’과 마찬가지이기 때문에¹ 이들의 결과가 나타나는 **컴포넌트 트리—저자-이슈 분석 결과** 혹은 **컴포넌트 트리 — 저자-이슈-성향 분석 결과**에서 확인합니다. 이에 대한 자세한 내용은 <III.F. 2. 저자-이슈 분석 처리결과 보기—이슈와 문서 중심으로 보기, page 139> 과 <III.H. 2. 저자-이슈-성향 분석 처리결과 보기, page 174>에서 다루겠습니다.

컴포넌트 트리—사전에서 생성된 단어 사전을 더블 클릭하면 작업창에 해당 사전을 편집할 수 있는 탭이 나타나며 프로그램에서 추출한 단어 이외에 사용자가 원하는 단어를 추가하거나 삭제할 수 있습니다.

a) 단어 사전 편집

컴포넌트 트리의 ‘사전’에 있는 임의의 단어 사전을 클릭했을 때 작업창은 다음과 같습니다.



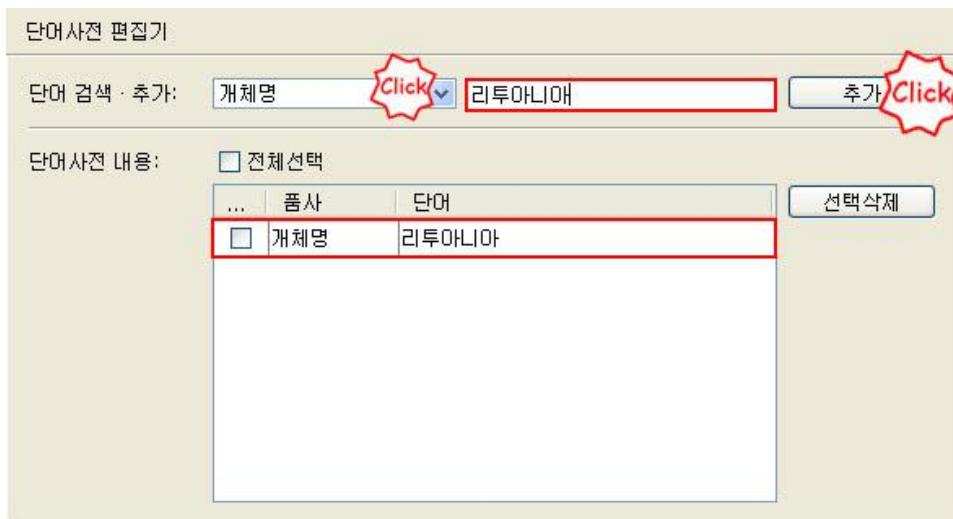
¹ 삭제어 사전과 일반화 사전이 적용된 결과는 **컴포넌트 트리—문서**에서 확인합니다.

탭의 상단에는 프로젝트 명과 사전종류가 표기되며 하위 메뉴에는 ‘단어 사전 편집’과 ‘확장 기능’ 총 2 가지가 있습니다. 기존에 등록된 단어를 찾을 수 있고 새로운 단어를 추가할 수도 있습니다.

(a-1) 단어 사전 편집기

i. 단어 검색

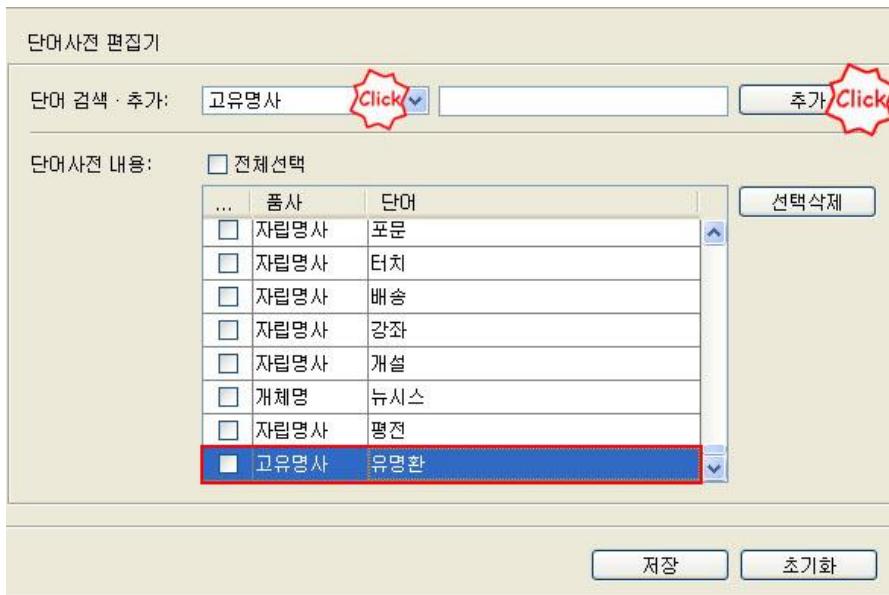
단어를 검색하고자 한다면, 선택란에 해당하는 품사를 클릭한 후 단어란에 찾고자 하는 단어를 기입하면 하단에 조건에 일치하는 검색어가 나타납니다.



개체명을 선택하고 ‘리투아니아’를 검색했을 때

ii. 단어 추가

선택란에서 해당하는 품사를 클릭한 후 단어란에 추가하고자 하는 단어를 기입합니다. 이후 우측에 <추가>를 누르면 하단에 조건에 일치하는 단어가 생성됩니다.



개체명을 선택하고 ‘호랑나비’를 추가했을 때

생성된 단어를 삭제하고자 한다면 선택란에 체크를 한 후 **선택삭제**를 클릭하면 선택된 단어가 삭제됩니다.

(ii-1) 저장 **저장** : 편집된 내용을 저장합니다.

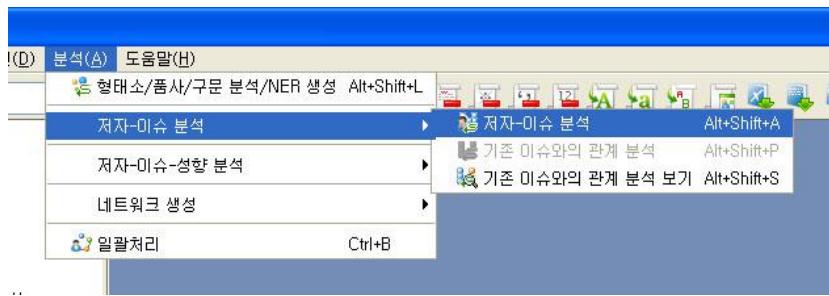
(ii-2) 초기화 **초기화** : 명령이 실행되지 않은 초기의 상태로 돌아갑니다.

F. 저자-이슈 분석 실행하기

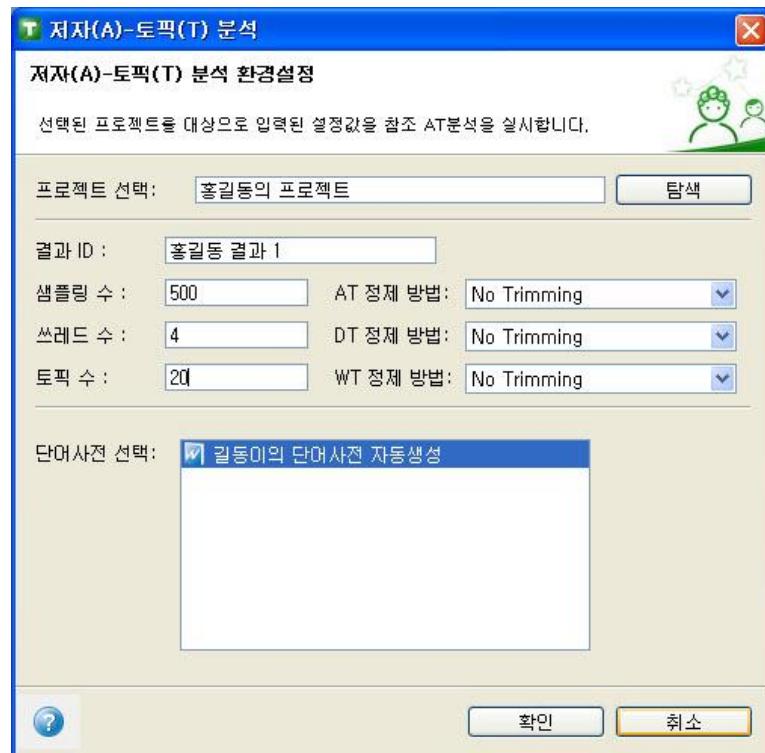
1. 저자-이슈 분석

단어사전을 생성한 후 ‘저자-이슈 분석’을 실행합니다. 저자-이슈 분석은 단어 사전에 수집된 단어들과 다른 사전(들)을 적용 시킨 후에 문서에 남아 있는 단어들의 교집합을 대상으로 합니다. 무엇보다도 ‘저자-이슈 분석’에서는 분석 과정보다 이를 실행하여 얻어지는 결과 메뉴들에 대한 충분한 이해력과 활용력이 요구되며 지금부터 저자-이슈 분석 과정을 살펴보도록 하겠습니다.

- a) ‘저자-이슈 분석’을 실행하려면 **메뉴막대 – 분석 – 저자-이슈 분석**을 누릅니다.



- b) ‘저자(A)-토픽(T) 분석’ 대화상자가 나타납니다.



(b-1) 프로젝트 선택에는 적용 대상이 되는 프로젝트입니다. 우측에 <탐색>을 클릭하여 해당 프로젝트를 선택한 후 <확인>을 클릭합니다.

(b-2) 결과 ID에는 사용자가 원하는 결과명을 기입합니다.

(b-3) 샘플링 수는 ‘저자-이슈 분석’을 더욱 정밀하게 하기 위하여 반복적으로 수행하는 통계처리 횟수입니다.

(b-4) 쓰레드 수는 샘플링을 병렬적으로 수행하기 위하여 지정하는 동시 작업 횟수입니다.

(b-5) 토픽 수에는 몇 개의 토픽을 추출할 것인지 기입합니다.

(b-6) 단어사전 선택에는 적용하고자 하는 단어사전을 체크합니다.

(b-7) AT 정제 방법 : 산출되는 저자의 비중값을 정제합니다.

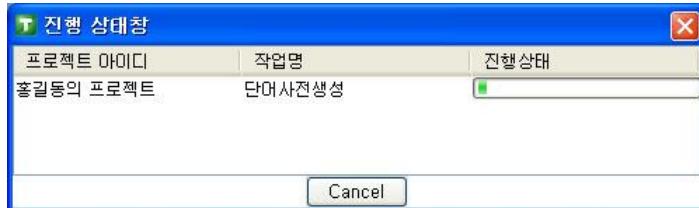
(b-8) DT 정제 방법 : 산출되는 문서 비중값을 정제합니다.

(b-9) WT 정제 방법 : 산출되는 단어의 비중값을 정제합니다.

- No Trimming : 처리를 실행하지 않음
- Probability Drop Trimming : 가장 큰 비중값으로 떨어지는 지점까지만 결과값을 표시
- Log Probability Drop Trimming : 비중값을 Log 로 치환한 후 가장 큰 값으로 떨어지는 지점까지 표시
- Top Topic Matching Trimming : 검색된 단어, 저자, 문서로 해당 토픽을 가장 높은 비중값으로 들고 있을 확률

‘저자(A)–토픽(T) 분석’의 권장수치는 샘플링 수에는 1000 개 문서 시 500 회, 쓰레드 수는 CPU 의 Core 개수, 토픽 수는 1000 개 기준 20 개입니다. 설정이 끝나면 <확인>을 클릭합니다.

- c) ‘진행 상태창’ 대화상자가 나타나면 연두색 막대가 채워질 때까지 기다립니다. 이때 <Cancel>을 클릭하면 ‘자연어 처리’가 중단됩니다.



- d) 작업이 완료되면 컴포넌트 트리의 ‘저자-이슈 결과’에 ‘저자-이슈 분석’의 결과¹가 생성됩니다. 생성된 결과를 더블 클릭하면 작업창에 새 탭이 나타나며 프로그램에서 추출한 ‘전체토픽, 저자, 토픽, 문서, 토픽간 관계’의 값을 확인할 수 있습니다.

¹ ‘저자-이슈 결과’에서 결과명과 결과 아이디는 동일하게 사용합니다. ‘문서’에서도 문서명과 문서 아이디는 동일하게 사용합니다.



- e) 생성된 분석 결과는 파일의 형태로 사용자가 지정한 작업폴더(최초 설정- C://Workplace, page 29, (d-2) 참조)에 자동 저장되며 메시지 팬넬의 속성에서 저장된 위치 등 상세정보를 확인할 수 있습니다.
- f) 명령이 성공적으로 수행되었다면 속성에는 [***** Successfully executed *****](#) 가 표기됩니다.

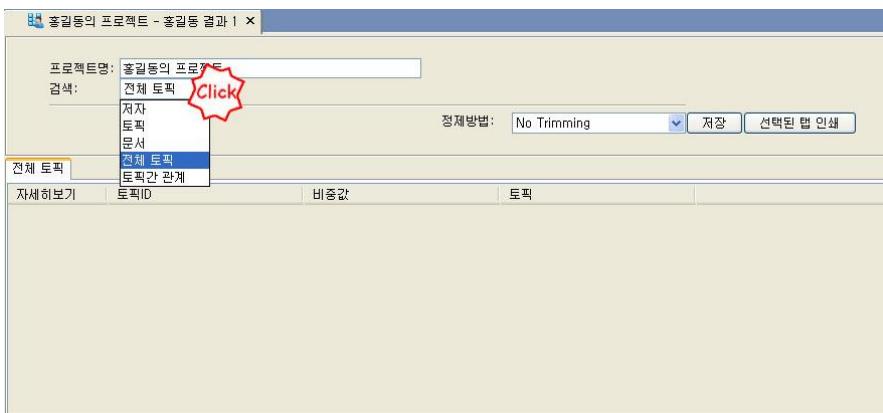
2. 저자-이슈 분석 처리 결과 보기 – 이슈와 문서 중심으로 보기

컴포넌트 트리의 ‘저자-이슈결과’는 ‘저자-이슈 분석’을 실행한 결과가 나타나는 곳입니다. 나열된 결과 문서들은 결과명으로 보여지게 되며 괄호 안의 숫자는 결과문서의 개수를 의미합니다.



‘저자-이슈결과’에서는 결과명과 결과 아이디를 동일하게 사용합니다. 여기에서 특이사항은 원칙적으로 프로그램이 결과에 대한 아이디를 부여 해야 하나, 사용자가 ‘저자(A)-토픽(T) 분석 환경설정 대화상자—결과 ID’에 결과 아이디를 직접 기입함으로써 아이디를 부여했다는 점입니다. 생성된 결과를 삭제하고 싶을 경우 해당 사전에 마우스 오른쪽 버튼을 눌러 삭제를 클릭합니다.

컴포넌트 트리의 ‘저자-이슈 결과’에 있는 결과 문서를 클릭했을 때 초기의 작업창의 모습은 다음과 같습니다.



탭 상단에는 프로젝트 명과 결과명이 표기됩니다. 결과 탭의 하위 선택란에는 전체 토픽, 저자, 토픽, 문서, 토픽간 관계 총 5 가지 메뉴가 있습니다.

정제방법은 <검색>을 눌렀을 때 표시되는 결과의 개수를 제한하는 역할을 합니다. 주의할 점은 앞서 ‘저자-이슈 분석’을 실행할 시에 설정하였던 AT/DT/WT 정제 방법과 기능이 같으며 중복해서 적용이 가능합니다.



- No Trimming : 처리를 실행하지 않음
- Probability Drop Trimming : 가장 큰 비중값으로 떨어지는 지점까지만 결과값을 표시
- Log Probability Drop Trimming : 비중값을 Log 로 치환한 후 가장 큰 값으로 떨어지는 지점까지 표시
- Top Topic Matching Trimming : 검색된 단어, 저자, 문서로 해당 토픽을 가장 높은 비중값으로 들고 있을 확률
- 저장 : 현재 보이는 결과를 엑셀에서 열 수 있는 파일 형태로 저장합니다.
- 선택된 탭 인쇄 : 현재 보이는 결과를 인쇄한다.

a) 전체 토픽

전체 토픽은 ‘저자-이슈 분석’을 실행한 후 생성되는 결과메뉴 중 하나로, 추출한 토픽을 한눈에 확인할 수 있습니다. 가장 비중 있게 다뤄진 순서대로 토픽이 나열됩니다.

(방법 1) 선택란에서 전체 토픽을 클릭한 후 <검색>을 누릅니다. 혹은 (방법 2) 컴포넌트 트리의 ‘저자-이슈 분석’에서 해당 결과문서에 마우스 오른쪽을 클릭한 후 ‘전체 토픽’을 누릅니다.

자세히보기	토픽ID	비중값	토픽
1	0.077963	신증플루_환자_감염_백신_예방_병원_사망_학교_확산_치료_발생_검사_대책_바이러스_보건_학생_정부_접종_의료...	
2	0.070794	생각_사람_부동산_가격_국민_아파트_나라_부동_시작_정도_이상_지금_인간_현재_정책_국가_개발_경제_세계_현실...	
3	0.068264	세종시_의원_수정_국민_한나라당_문재_정치_국회_대표_투표_원안_수도_총리_정부_여론_행정_민주당_반대_토론...	
4	0.064513	금지_기능_전제_기자_아이폰_결과_관계_배포_동안_가운데_서울_설명_중심_무단_기술_전행_방문_종합_문화_상당...	
5	0.059949	사업_북한_정보_4대강_문제_대통령_회담_이명박_남북_관계_남북정상회담_정상_죽진_예산_관련_가능_원칙_대화...	
6	0.052410	문제_조사_관련_칠레_배포_금지_기자_무단_프리우스_도박_사과_내용_에폭_간접_관계_경우_주장_박찬호_설명...	
7	0.051027	연합뉴스_그리스_내포_금지_전재_저작권자_무단_지원_정부_이날_재정_회의_계획_방안_추가_독일_티니_필요_국...	
8	0.050429	시장_경제_미국_위기_한국_금융_중국_국가_세계_가동_상황_투자_유럽_우려_두바이_분석_사태_문제_운항_전망...	
9	0.050158	보도_방송_MBC_언론_판결_사설_검찰_PD수첩_이번_영화_새해_클래식_마바타_법원_경제동_논란_판사_발언_내...	
10	0.048329	선언_시국_국민_정치_교사_전교조_교육_사회_경제_행위_자유_민주주의_행동_공무원_표현_참여_불법_학생_시민...	
11	0.047202	도요타_리콜_미국_일본_자동차_차량_감독_사태_결합_세계_연합뉴스_판매_이번_대표_품질_생산_문제_기술_사고...	
12	0.045309	아이티_지진_구조_발생_피해_활동_현지_지원_구호_대_주민_문화_건물_연합뉴스_규모_사망_상황_현장_기자_국제...	
13	0.044627	김연_선수_한국_올림픽_금메달_연합뉴스_연기_훈련_경기_시간_최고_밴쿠버_프로그램_여자_김연아_배포_대회...	
14	0.041163	스마트폰_국내_시장_사용_업체_삼성전자_판매_제품_세계_개발_사업_기술_출시_흘해_애플_인터넷_연구_기능_서...	

(a-1) 자세히 보기 : 해당 토픽¹으로 새 탭이 생성되어 이동합니다.

(a-2) 토픽 ID : 프로그램이 비중값의 수치²를 내림차순으로 각 토픽에 부여한 숫자입니다.

(a-3) 비중값 : 각 토픽이 모든 문서³에서 얼마나 비중 있게 다루어졌는지 수치로 나타낸 것입니다. 수치의 범위는 0~1 까지이며 1에 가까울수록 비중이 커집니다.

(a-4) 토픽 : 단어사전에 수집된 단어들을 대상으로 프로그램이 서로 관련 있는 단어들을 그룹화 시켜 놓은 것입니다. 토픽의 총 개수는 ‘저자이슈 분석 대화상자-토픽 수’에서 사용자가 지정하였던 개수입니다⁴. 단어들은 하이픈으로 연결되어 있으며 20 개씩⁵ 모아져 있습니다.

b) 저자

¹ page 95, c)Topic 을 참조합니다.

² 아이디(ID)란 프로그램이 해당 정보에 ‘임의로’ 부여하는 숫자입니다. 예외적으로 토픽은 비중값의 수치를 기준으로 아이디가 부여됩니다.

³ 문서에 수집된 모든 문서를 말합니다.

⁴ 예를 들어 ‘저자이슈 분석 대화상자-토픽 수’에서 10을 기제하였다면 총 토픽의 개수는 10개가 됩니다.

⁵ 토픽에서 보여지는 단어들의 순서는 해당 토픽 안에서 가장 비중이 높은 상위의 단어 20개를 수집한 것입니다.

저자는 ‘저자-이슈 분석’을 실행한 후 생성되는 결과메뉴 중 하나로, 문서를 작성한 개개의 저자가 추출된 토픽 중에서 어떤 것을 가장 비중 있게 다루었는지 그리고 저자가 작성한 문서는 어떤 것들이 있는지를 열람할 수 있습니다. 가장 비중 있게 다뤄진 순서대로 토픽이 나열됩니다.

(방법 1) 선택란에서 저자를 클릭한 후 하단에 저자 아이디란이 생기면 사용자가 열람하고자 하는 저자 아이디를 선택한 후 <검색>을 누릅니다. 혹은 (방법 2) 컴포넌트 트리의 ‘저자-이슈 분석’에서 해당 결과문서에 마우스 오른쪽을 클릭한 후 ‘저자’를 누릅니다.

The screenshot shows a search interface for topic analysis. The search parameters are set to '저자' (Author) and 'All' for both topic and document counts. The search method is 'No Trimming'. The results table lists various topics with their term scores. A red starburst highlights the 'Click' button next to the author selection field.

관련문서보기	토픽ID	토픽	비중값	Term Score
	1	3세중시_의원_수정_국민_한보나_교당	0.272000	0.651116
	2	상각_시감_부동산_가격_국민	0.192000	0.392737
	3	9보도_방송_MBC_언론_판결	0.104000	0.148969
	4	18노무현_대통령_서민_사람_상각	0.080000	0.093603
	5	4급지_기능_건재_기자_아이폰	0.072000	0.076657
	6	19오프간_대통령_경찰_사람_자신	0.049000	0.031642
	7	6문제_조사_관련_질례_배포	0.040000	0.019075
	8	14스마트폰_국내_시장_사용_업체	0.040000	0.019075
	9	16기업_개성공단_북한_회장_시간	0.032000	0.006120
	10	1신종플루_환자_감염_백신_예방	0.024000	-0.000815
	11연합뉴스_그리스_배포_급지_건재	0.016000	-0.007030	
	12도요타_카풀_미국_일본_자동차	0.016000	-0.007030	

좌측에 선택란은 결과의 개수를 제한하는 역할을 합니다. 사용자에 필요에 따라 실행되므로 필수적으로 행해져야 되는 것은 아닙니다. 원하는 수치를 적거나 항목을 클릭한 후 <확인>을 클릭하면 결과의 수가 기입한 수만큼 제한된 것을 확인할 수 있습니다.

The screenshot shows the 'Trimming' dropdown menu open. The options listed are 'No Trimming', 'Probability Drop Trimming', 'Log Probability Trimming', and 'Top Topic Probability'. A red starburst highlights the 'Click' button next to the 'Trimming' dropdown.

- 저자명 : 해당 저자 아이디에 대한 저자명을 나타냄
 - 토픽수 : 결과 값으로 나타나는 토픽의 개수를 제한
 - 문서수 : 결과 값으로 나타나는 문서의 개수를 제한
 - No Trimming : 처리를 실행하지 않음
 - Probability Drop Trimming : 가장 큰 비중값으로 떨어지는 지점까지만 결과값을 표시
 - Log Probability Trimming : 비중값을 Log 로 치환한 후 가장 큰 값으로 떨어지는 지점까지 표시
 - Top Topic Matching Trimming : 검색된 단어, 저자, 문서로 해당 토픽을 가장 높은 비중값으로 들고 있을 확률
 - 저장 : 현재 보이는 결과를 엑셀에서 열 수 있는 파일 형태로 저장합니다.
 - 문서 내보내기 : 현재 보여지는 문서들을 원래 데이터 형태대로 지정된 폴더에 저장한다.
 - 선택된 탭 인쇄 : 현재 보이는 결과를 인쇄한다.

(b-1) 토픽

토Pic 문서	관련문서보기	토Pic ID 토픽	비중값	Term Score
		3세 증시_의원_수정_국민_한나라당	0.272000	0.651116
		2생각사람_부동산_기획_국민	0.192000	0.392717
		외보도_방송_MBC_언론_파급	0.104000	0.148969
		18노무현_미통령_사법_법률_집권_성작	0.080000	0.095603
		1금융_기금_증권_제재_기획_부동산	0.072000	0.076657
		19정부_기획_통화_법률_자산	0.068000	0.061452
		6문화_조사_언론_설립_베풀포	0.040000	0.019075
		14스마트폰_국민_시장_사용_업체	0.040000	0.019075
		16기금_개설공정_북한_회장_시간	0.032000	0.008120
		1신종플루_한국_감염_백신_예방	0.024000	-0.000815
		7연예_뉴스_스타_배포_금지_언제	0.016000	-0.007030
		11도로주차_리croft_기술_모포_자동차	0.016000	-0.007030

문서보기	문서ID	문서 작성일	토Pic	비중값	Term Score	문서 제목
	1000000000000000000	2010-01-17 00:00:00	3세 증시_의원_수정_국민_한나라당	0.272000	0.651116	노무현은 예언자? 국토영합발전 주장 등명상 화제만..

- i. 관련 문서 보기 : 별도의 이동 없이 하단에 해당 토픽을 비중 있게 다룬 문서들을 나열합니다.
 - (i-1) 문서 보기 : 우측으로 문서탭이 열리면서 해당 문서를 열람할 수 있습니다.
 - (i-2) 문서 ID : 상단에서 사용자가 선택한 토픽을 비중 있게 다룬 문서의 아이디입니다.
 - (i-3) 문서 작성일 : 상단에서 사용자가 선택한 토픽을 비중 있게 다룬 문서의 작성일입니다.
 - (i-4) 토픽 : 상단에서 사용자가 선택한 토픽을 비중 있게 다룬 문서의 토픽입니다.

(i-5) 비중값 : 상단에서 사용자가 선택한 토픽을 비중 있게 다른 문서에서 해당 저자가 이 토픽을 얼마나 비중 있게 다루었는지 수치로 나타냅니다. 따라서 하단의 비중값 또한 상단의 비중값과 비슷해야 합니다.

(i-6) Term Score : 비중값을 재가공하여 만든 지표로써 해당 단어, 문서, 사람 등이 토픽에 어느 정도 연관되어 있는지 나타냅니다. 이 지표는 비중값과 다르게 해당 토픽의 높은 순위에 위치한 단어, 저자, 문서를 다른 토픽에서는 어디에 위치해 있는지 판별하여 특징적으로 나타난 토픽의 사람, 문서, 단어 등을 지표상에서 높게 산출합니다. 즉, 비중값은 해당 토픽만을 염두해 두고 그 중요도를 수치화하는데 반해, Term Score 는 해당 토픽에서의 순위만이 아니라 다른 토픽에서의 순위까지 생각하여 해당 토픽에서의 중요도를 판별합니다. 다시 말하면, 비중값이 1 차 결과물이라면 Term Score 가 2 차 결과물이라 할 수 있습니다.

(i-7) 문서 제목: 상단에서 사용자가 선택한 토픽을 비중 있게 다른 문서의 제목입니다.

- ii. 토픽 ID : 프로그램이 토픽에 부여한 숫자입니다.
- iii. 토픽 : 총 20 개의 단어로 구성되었던 토픽들 중에서 비중이 큰 상위의 단어 4 개를 모아둔 것입니다.
- iv. 비중값 : 해당 저자가 각 토픽을 얼마나 비중 있게 다루었는지 수치로 나타낸 것입니다. 수치의 범위는 0~1 까지이며 1 에 가까울수록 비중이 커집니다.

(b-2) 문서

토릭 문서			
토릭보기	문서ID	문서 작성일	문서 제목
★	1000000304	2010-01-17 00:00:00	노무현은 예언자? 국토균형발전 주장 동영상 화제만발..
토픽ID	토픽	비중값	Term Score
3	세종시_의원_수정_국민_한나라당	0.272000	0.651116
2	생각_사람_부동산_가격_국민	0.192000	0.392737
9	보도_방송_MBC_언론_판결	0.104000	0.148969
18	노무현_대통령_서민_사람_생각	0.080000	0.093603
4	금지_가능_전재_기자_마이폰	0.072000	0.076657
19	아프간_대통령_정권_사람_자신	0.048000	0.031642
6	문제_조사_관련_질례_내포	0.040000	0.019075
14	스마트폰_국내_시장_사용_업체	0.040000	0.019075
16	기업_개성공단_북한_회장_시간	0.032000	0.008120
1	신증플루_환자_감염_백신_예방	0.024000	-0.000815
7	연합뉴스_그리스_배포_금지_전재	0.016000	-0.007030
11	도요타_리콜_미국_일본_자동차	0.016000	-0.007030

- i. 토픽보기 : 해당 문서에서 비중 있게 다른 토픽들을 나열합니다.

(i-1) 토픽 ID : 상단의 문서에서 비중 있게 다루어진 토픽의 아이디입니다.

(i-2) 토픽 : 상단의 문서에서 비중 있게 다루어진 토픽입니다.

(i-3) 비중값 : 상단의 문서에 대하여 토픽이 얼마나 비중 있게 다루어졌는지 수치로 나타냅니다.

(i-iv) Term Score : 비중값을 재가공하여 만든 지표로써 해당 단어, 문서, 사람 등이 토픽에 어느 정도 연관되어 있는지 나타냅니다. 이 지표는 비중값과 다르게 해당 토픽의 높은 순위에 위치한 단어, 저자, 문서를 다른 토픽에서는 어디에 위치해있는지 판별하여 특징적으로 나타난 토픽의 사람, 문서, 단어 등을 지표상에서 높게 산출합니다. 즉, 비중값은 해당 토픽만을 염두해 두고 그 중요도를 수치화하는데 반해, Term Score 는 해당 토픽에서의 순위만이 아니라 다른 토픽에서의 순위까지 생각하여 해당 토픽에서의 중요도를 판별합니다. 다시 말하면, 비중값이 1 차 결과물이라면 Term Score 가 2 차 결과물이라 할 수 있습니다.

- ii. 문서 ID : 프로그램이 해당 문서에 임의로 부여한 숫자입니다.
- iii. 문서 작성일: 문서가 생성된 일시입니다.
- iv. 문서 제목: 문서의 제목입니다.

c) 토픽

토픽은 ‘저자-이슈 분석’을 실행한 후 생성되는 결과메뉴 중 하나로 해당 토픽이 각각의 단어, 저자, 문서를 얼마나 비중 있게 다루었는지 열람할 수 있습니다. 단어, 저자, 문서는 가장 비중 있게 다뤄진 순서대로 나열됩니다.

(방법 1) 선택란에서 토픽을 클릭한 후 하단에 토픽 아이디란이 생기면 사용자가 열람하고자 하는 토픽 아이디를 선택한 후 <검색>을 누릅니다. 혹은 (방법 2) 컴포넌트 트리의 ‘저자-이슈 분석’에서 해당 결과문서에 마우스 오른쪽을 클릭한 후 ‘토픽’를 누릅니다.



우측의 선택란은 결과의 개수를 제한하는 역할을 합니다. 사용자에 필요에 따라 실행되므로 필수적으로 행해져야 되는 것은 아닙니다. 원하는 수치를 적거나 항목을 클릭한 후 확인 버튼을 클릭하면 결과의 수가 제한된 것을 확인할 수 있습니다.



- 토픽명 : 해당 토픽 ID에 대한 토픽명을 나타냄
- 토픽 수 : 결과 값으로 나타나는 토픽의 개수를 제한
- 문서 수 : 결과 값으로 나타나는 저자의 개수를 제한
- No Trimming : 처리를 실행하지 않음
- Probability Drop Trimming : 가장 큰 비중값으로 떨어지는 지점까지만 결과값을 표시
- Log Probability Drop Trimming : 비중값을 Log로 치환한 후 가장 큰 값으로 떨어지는 지점까지 표시
- Top Topic Matching Trimming : 검색된 단어, 저자, 문서로 해당 토픽을 가장 높은 비중값으로 들고 있을 확률
- 저장 : 현재 보이는 결과를 엑셀에서 열 수 있는 파일 형태로 저장합니다.
- 문서 내보내기 : 현재 보여지는 문서들을 원래 데이터 형태대로 지정된 폴더에 저장한다.

- 선택된 템 인쇄 : 현재 보이는 결과를 인쇄한다.

(c-1) 단어

단어ID	단어	비중값	Term Score
740	신종플루	0,080826	0,970521
5371	환자	0,025676	0,280340
3364	감염	0,024423	0,265490
745	백신	0,020396	0,218231
2209	예방	0,019955	0,207801
3392	병원	0,016882	0,170597
747	사망	0,016405	0,157855
1318	학교	0,016176	0,153974
8	확산	0,014781	0,136518
2210	치료	0,014587	0,146756
657	발생	0,012538	0,096368
1497	검사	0,012273	0,112412
1400	대책	0,011231	0,096622
3369	바이러스	0,010242	0,102886
510	보건	0,009907	0,096123
1072	학생	0,009412	0,082689
4	정부	0,009254	0,043050
14076	접종	0,009236	0,091868
6756	의료	0,009130	0,086653
9763	확진	0,008671	0,085727
447	경우	0,008371	0,047514

- 단어 ID: 프로그램이 단어에 임의로 부여한 숫자입니다.
- 단어: 단어사전에 수집된 단어입니다.
- 비중값: 단어사전에 수집된 단어들이 해당 토픽을 얼마나 비중 있게 다루었는지 수치로 나타낸 것을 말합니다. 수치의 범위는 0 ~ 1 까지이며 1 에 가까울수록 비중이 커집니다.
- Term Score : 비중값을 재가공하여 만든 지표로써 해당 단어, 문서, 사람 등이 토픽에 어느 정도 연관되어 있는지 나타냅니다. 이 지표는 비중값과 다르게 해당 토픽의 높은 순위에 위치한 단어, 저자, 문서를 다른 토픽에서는 어디에 위치해있는지 판별하여 특징적으로 나타난 토픽의 사람, 문서, 단어 등을 지표상에서 높게 산출합니다. 즉, 비중값은 해당 토픽만을 염두해 두고 그 중요도를 수치화하는데 반해, Term Score 는 해당 토픽에서의 순위만이 아니라 다른 토픽에서의 순위까지 생각하여 해당 토픽에서의 중요도를 판별합니다. 다시 말하면, 비중값이 1 차 결과물이라면 Term Score 가 2 차 결과물이라 할 수 있습니다.

(c-2) 저자

생성문서보기	저자ID	저자명	비중값	Term Score
[파일]	10599815	강원일보	0,731707	2,876687
[파일]	10601759	새전북신문	0,689655	2,737580
[파일]	10599078	제민일보	0,682759	2,571145
[파일]	10614099	국제신문	0,661692	2,497503
[파일]	10600561	전북일보	0,643836	2,476592
[파일]	10601297	대전일보	0,637931	2,462097
[파일]	10600010	전북도민일보	0,624413	2,370064
[파일]	10670630	augustx	0,622642	2,354323
[파일]	10582084	충북일보	0,619910	2,421747
[파일]	10588243	브레이크뉴스	0,612245	2,193206
[파일]	10677396	정다운미웃	0,603175	2,039749
[파일]	10671913	대마왕	0,600000	2,039480
[파일]	10614630	전북도민일보	0,600000	2,220126
[파일]	10600433	경기일보	0,599222	2,204435

- i. 생성문서보기 : 해당 저자가 쓴 문서(들)의 관련 정보들을 열람할 수 있습니다. 별도의 화면 이동 없이 작업창 하단에 실행됩니다. 이는 본 토픽과는 관련이 없습니다.

문서보기	문서ID	문서 작성일	문서 제목	토픽
[파일]	10599815	2009-09-01 00:00:00	신종플루 거점병원 의료진 잇단 감염	신종플루_환자_감염_백신_예방

- (i-1)문서보기 : 우측으로 새 탭이 열리면서 해당 저자가 쓴 문서를 열람할 수 있습니다.
 (i-2)문서 ID : 프로그램이 문서에 임의로 부여한 숫자입니다.
 (i-3)문서 작성일 : 문서가 생성된 일시입니다.
 (i-4)문서 제목: 문서의 제목입니다.
 (i-5)토픽 : 해당 문서에서 가장 비중 있게 다뤄지는 토픽이 나타납니다.

- ii. 저자 ID: 프로그램이 해당 저자에 임의로 부여한 숫자입니다.
 iii. 저자: 저자명입니다.
 v. 비중값: 저자들이 해당 토픽을 얼마나 비중 있게 다루었는지 수치로 나타낸 것을 말합니다. 수치의 범위는 0~1 까지이며 1에 가까울수록 비중이 커집니다.

vi. Term Score: 비중값을 재가공하여 만든 지표로써 해당 단어, 문서, 사람 등이 토픽에 어느 정도 연관되어 있는지 나타냅니다. 이 지표는 비중값과 다르게 해당 토픽의 높은 순위에 위치한 단어, 저자, 문서를 다른 토픽에서는 어디에 위치해있는지 판별하여 특징적으로 나타난 토픽의 사람, 문서, 단어 들을 지표상에서 높게 산출합니다. 즉, 비중값은 해당 토픽만을 염두해 두고 그 중요도를 수치화하는데 반해, Term Score 는 해당 토픽에서의 순위만이 아니라 다른 토픽에서의 순위까지 생각하여 해당 토픽에서의 중요도를 판별합니다. 다시 말하면, 비중값이 1 차 결과물이라면 Term Score 가 2 차 결과물이라 할 수 있습니다.

(c-3) 문서

문서 보기	저자 문서 보기	문서 ID	문서 작성일	비중값	Term Score	저서명	문서 제목
■	■	10599815	2009-09-01 00:00:00 0_731707	2.876687	강원일보	신종플루 가정방역 의료진 잇단 감염	
■	■	10601753	2009-09-01 00:00:00 0_689655	2.737580	세계신문	신종플루 확산 아디카서	
■	■	10598238	2009-09-01 00:00:00 0_682758	2.731145	제주일보	신종플루 의심환자 10명 조치 학계	
■	■	10514699	2009-09-01 00:00:00 0_643936	2.497930	제주신문	기상청 "최악" 물류단속 플루 대책 /유경길	
■	■	10600561	2009-09-01 00:00:00 0_643936	2.476592	전북일보	[신종플루 확산] 물류단속 플루 대책 /유경길	
■	■	10601297	2009-09-02 00:00:00 0_657381	2.462097	대전일보	신종플루 확산 대처 방지	
■	■	10600010	2009-09-01 00:00:00 0_624413	2.370084	전북도민일보	대한가 계간 신종플루 예방 네 살	
■	■	1067630	2009-09-25 00:00:00 0_625642	2.354323	augustus	신종플루 가정방침을 전부에 뿐나다	
■	■	10582084	2009-09-03 00:00:00 0_619910	2.421747	충북일보	신종플루에 노동 경력을 떨으 엄마"	
■	■	10588243	2009-09-03 00:00:00 0_612245	2.193206	브레이크뉴스	신종플루 사망자 4명으로 늘어, '무' 줄태	
■	■	10677396	2009-09-31 00:00:00 0_603175	2.039749	정다운미술	신종플루 예방 위해 대형병원의 사용 광고 활동 필요하다	
■	■	10514630	2009-09-01 00:00:00 0_600000	2.220126	전북도민일보	<미경총> 전북도 청정복지사·신종플루 예방과 개인위생	
■	■	10671913	2009-09-24 00:00:00 0_600000	2.039480	대마왕	신종플루 가볍게 지나온으면...	
■	■	10600433	2009-09-01 00:00:00 0_599222	2.204495	경기일보	신종플루 대유행" 예고 가을철 철저한 대비를	

- i. 문서 보기 : 우측으로 새 탭이 열리면서 해당 문서를 열람할 수 있습니다.
- ii. 저자 문서 보기 : 해당 문서 이외에 작성한 저자의 다른 문서들을 열람할 수 있습니다. 별도의 화면 이동 없이 작업창 하단에 실행됩니다.

문서 보기	문서 ID	문서 작성일	문서 제목	토픽
■	10585274	2009-09-02 00:00:00	[신종플루 확산 '비상'] 노인 무방비 노출 대책 마련 시급	신종플루_환자_감염_백신_예방
■	10585275	2009-09-02 00:00:00	[신종플루 확산 '비상'] 현혈자 감소 혈액 수급 차질	지역_행사_경기_월드컵_도시
■	10585276	2009-09-02 00:00:00	[신종플루 확산 '비상'] 道 허霆동안 학생 7명 확진 판정	신종플루_환자_감염_백신_예방
■	10585298	2009-09-02 00:00:00	[평창] 신종플루 예방 29억천만원 지원	신종플루_환자_감염_백신_예방
■	10585353	2009-09-02 00:00:00	[평창] 30만 인파 몰린다! 신종플루 대책 고심	신종플루_환자_감염_백신_예방
■	10585360	2009-09-02 00:00:00	[평성] 신종플루 방역 활동 강화	신종플루_환자_감염_백신_예방
■	10599231	2009-09-01 00:00:00	[평창] 보건교사 없는데 신종 플루 대응?	신종플루_환자_감염_백신_예방
■	10599792	2009-09-01 00:00:00	[정선] 정선 36개교 신종플루 24시간 비상방역체계 유지	신종플루_환자_감염_백신_예방
■	10599815	2009-09-01 00:00:00	신종플루 확산방지 공직기강 확립 지시	신종플루_환자_감염_백신_예방
■	70004127300	2009-11-27 21:49:23	두바이 쇼케에 금융시장 곤두박질	시장_경제_미국_위기_환경
■	70004242575	2009-12-21 21:38:42	노 전대통령 묘역 조성...민주당 도당 일일첫집	생각_사람_부동산_가격_국민

- (ii-1) 문서 보기 : 우측으로 새 탭이 열리면서 해당 문서를 열람할 수 있습니다.
- (ii-2) 문서 ID : 프로그램이 문서에 임의로 부여한 숫자입니다.
- (ii-3) 문서 작성일 : 문서가 생성된 일시입니다.
- (ii-4) 문서 제목 : 문서의 제목입니다.

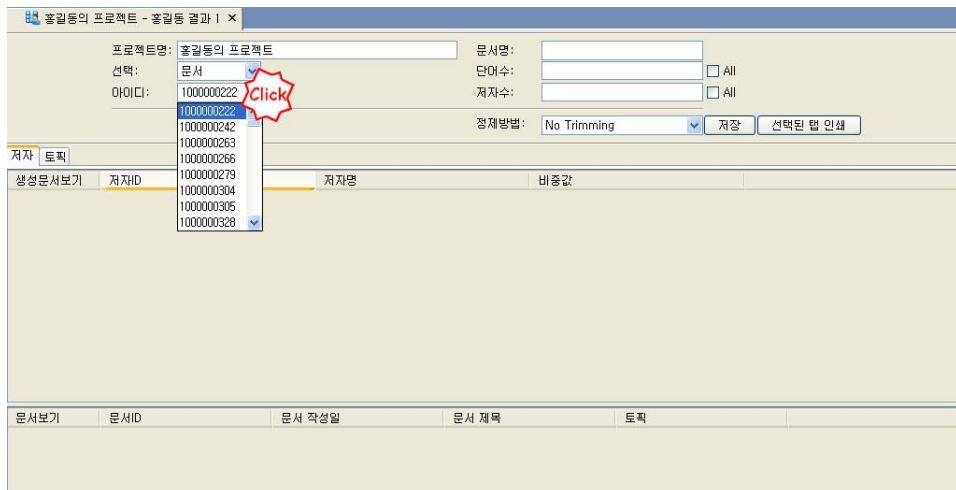
(ii-5) 토픽 : 해당 문서에서 가장 비중 있게 다뤄진 토픽이 나타납니다.

- iii. 문서 ID : 프로그램이 해당 문서에 임의로 부여한 숫자입니다.
- iv. 문서 작성일 : 문서가 생성된 일시입니다.
- v. 비중값 : 문서들이 해당토픽을 얼마나 비중 있게 다루었는지 수치로 나타낸 것을 말합니다. 수치의 범위는 0~1 까지이며 1에 가까울수록 비중이 커집니다.
- vi. Term Score : 비중값을 재가공하여 만든 지표로써 해당 단어, 문서, 사람 등이 토픽에 어느 정도 연관되어 있는지 나타냅니다. 이 지표는 비중값과 다르게 해당 토픽의 높은 순위에 위치한 단어, 저자, 문서를 다른 토픽에서는 어디에 위치해있는지 판별하여 특징적으로 나타난 토픽의 사람, 문서, 단어 등을 지표상에서 높게 산출합니다. 즉, 비중값은 해당 토픽만을 염두해 두고 그 중요도를 수치화하는데 반해, Term Score 는 해당 토픽에서의 순위만이 아니라 다른 토픽에서의 순위까지 생각하여 해당토픽에서의 중요도를 판별합니다. 다시 말하면, 비중값 1 차 결과물이라면 Term Score 가 2 차 결과물이라 할 수 있습니다.
- vii. 저자명 : 저자의 이름입니다.
- viii. 문서 제목 : 문서의 제목입니다.

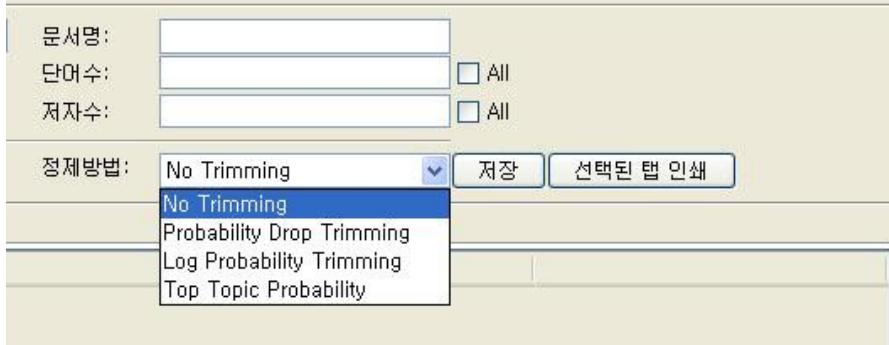
d) 문서

문서는 ‘저자-이슈 분석’을 실행한 후 생성되는 결과메뉴 중 하나로, 각 문서의 저자 정보를 확인할 수 있고, 가장 비중 있게 다뤄진 토픽을 열람할 수도 있습니다. 가장 비중 있게 다뤄진 순서대로 나열됩니다.

(방법 1) 선택란에서 문서를 클릭한 후 하단에 문서 아이디란이 생기면 사용자가 열람하고자 하는 문서 아이디를 선택한 후 <검색>을 누릅니다. 혹은 (방법 2) 컴포넌트 트리의 ‘저자-이슈 분석’에서 해당 결과문서에 마우스 오른쪽을 클릭한 후 ‘도큐먼트’를 누릅니다.



좌측에 선택란은 결과의 개수를 제한하는 역할을 합니다. 사용자에 필요에 따라 실행되므로 필수적으로 행해야 되는 것은 아닙니다. 원하는 수치를 적거나 항목을 클릭한 후 <확인>을 클릭하면 결과의 수가 제한된 것을 확인할 수 있습니다.



- 문서명 : 해당 문서 아이디에 대한 문서명을 나타냄
- 단어 수 : 결과 값으로 나타나는 단어의 개수를 제한
- 저자 수 : 결과 값으로 나타나는 저자의 개수를 제한
- No Trimming : 처리를 실행하지 않음
- Probability Drop Trimming : 가장 큰 비중값으로 떨어지는 지점까지만 결과값을 표시
- Log Probability Drop Trimming : 비중값을 Log 로 치환한 후 가장 큰 값으로 떨어지는 지점까지만 표시

- Top Topic Matching Trimming: 검색된 단어, 저자, 문서로 해당 토픽을 가장 높은 비중값으로 들고 있을 확률
- 저장 : 현재 보이는 결과를 엑셀에서 열 수 있는 파일 형태로 저장합니다.
- 선택된 탭 인쇄 : 현재 보이는 결과를 인쇄한다.

(d-1) 저자

저자	토픽		
생성문서보기	저자ID	저자명	비중값
	60004657467	연합뉴스	1,000000

- i. 생성문서보기 : 해당 문서의 저자가 쓴 다른 문서들을 별도의 화면 이동 없이 하단에 나타납니다.

문서보기	문서ID	문서 작성일	문서 제목	토픽
	60004657467	2010-03-03 17:16:09	도요타, 美서 오토론 5년간 무미자	도요타_리콜_미국_일본_자동차

- (i-1) 문서 보기: 우측으로 문서탭이 열리면서 해당 저자가 쓴 문서를 열람할 수 있습니다.
 (i-2) 문서 ID : 프로그램이 문서에 임의로 부여한 숫자입니다.
 (i-3) 문서 작성일 : 문서가 생성된 일시입니다.
 (i-4) 문서 제목: 문서의 제목입니다.
 (i-5) 토픽 : 해당 문서에서 가장 비중 있게 다뤄지는 토픽이 나타납니다.

- ii. 저자 ID: 프로그램이 해당 저자에 임의로 부여한 숫자입니다.
 iii. 저자명 : 해당 문서의 저자명입니다.
 iv. 비중값: 저자가 해당 문서에서 얼마나 비중 있게 다루어졌는지 수치로 나타낸 것을 말합니다. 수치의 범위는 0~1 까지이며 1에 가까울수록 비중이 커집니다. 문서에 대한 저자는 단 1 명이기 때문에 값은 무조건 1이 나오게 됩니다.

(d-2) 토픽

저자	토픽	비중값	Term Score
	11 도요타_리콜_미국_일본_자동차	0.335106	0.932668
	6 문제_조사_관련_칠레_배포	0.191489	0.425793
	4 금지_가능_전재_기자_아이폰	0.122340	0.217223
	7 연합뉴스_그리스_배포_금지_전재	0.085106	0.120226
	2 생각_사람_부동산_가격_국민	0.053191	0.050141
	14 스마트폰_국내_시장_사용_업체	0.047872	0.040083
	17 선거_지역_지방_후보_단체	0.026596	0.006636
	16 기업_개성공단_북한_회장_시간	0.021277	0.000561
	10 선언_시국_국민_정치_교사	0.015957	-0.004170
	13 김연_선수_한국_올림픽_금메달	0.015957	-0.004170
	15 지역_행사_경기_월드컵_도시	0.015957	-0.004170
	3 세종시_의원_수정_국민_한나라당	0.010638	-0.007093
	5 사업_북한_정부_4대강_문제	0.010638	-0.007093
	18 노무현_대통령_서민_사람_생각	0.010638	-0.007093

- i. 관련 문서 보기 : 해당 토픽을 비중 있게 다른 문서들을 별도의 화면 이동 없이 작업창 하단에 나열합니다.

문서보기	문서ID	문서 작성일	토픽	비중값	Term Score	문서 제목
	20023375897	2009-11-19 00:00:00	생각_사람_부동산_가격_국민	0.031250	-0.000147	내불부동산이 아주 빅딜선 되는구나.투기꾼...

- (i-1) 문서보기 : 우측으로 새 탭이 열리면서 해당 저자가 쓴 문서를 열람할 수 있습니다.
- (i-2) 문서 ID : 프로그램이 문서에 임의로 부여한 숫자입니다.
- (i-3) 문서 생성일: 문서가 생성된 일시입니다.
- (i-4) 토픽 : 해당 문서에서 가장 비중 있게 다뤄지는 토픽이 나타납니다.
- (i-5) 비중값 : 모든 문서가 해당 토픽을 얼마나 비중 있게 다루었는지 수치로 나타낸 것을 말합니다. 수치의 범위는 0~1 까지이며 1에 가까울수록 비중이 커집니다.
- (i-6) 문서 제목 : 문서의 제목입니다.

- ii. 토픽 ID : 프로그램이 해당 토픽에 임의로 부여한 숫자입니다.
- iii. 토픽 : 해당 문서에서 가장 비중 있게 다뤄지는 토픽이 나타납니다.
- iv. 비중값 : 해당 문서가 각 토픽을 얼마나 비중 있게 다루어졌는지 수치로 나타낸 것을 말합니다. 수치의 범위는 0~1 까지이며 1에 가까울수록 비중이 커집니다.
- v. Term Score : 비중값을 재가공하여 만든 지표로써 해당 단어, 문서, 사람 등이 토픽에 어느 정도 연관되어 있는지 나타냅니다. 이 지표는 비중값과 다르게 해당

토픽의 높은 순위에 위치한 단어, 저자, 문서를 다른 토픽에서는 어디에 위치해있는지 판별하여 특징적으로 나타난 토픽의 사람, 문서, 단어 들을 지표상에서 높게 산출합니다. 즉, 비중값은 해당 토픽만을 염두해 두고 그 중요도를 수치화하는데 반해, Term Score 는 해당 토픽에서의 순위만이 아니라 다른 토픽에서의 순위까지 생각하여 해당 토픽에서의 중요도를 판별합니다. 다시 말하면, 비중값이 1 차 결과물이라면 Term Score 가 2 차 결과물이라 할 수 있습니다.

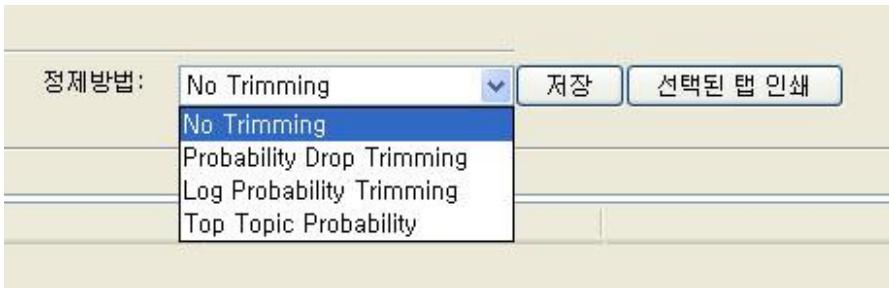
e) 토픽간 관계

토픽간 관계는 ‘저자-이슈 분석’을 실행한 후 생성되는 결과메뉴 중 하나로, 해당 토픽이 각각의 토픽과 얼마만큼 유사한지 열람할 수 있습니다. 토픽은 해당 토픽과 가장 유사하게 다뤄진 순서대로 나열됩니다.

(방법 1) 선택란에서 토픽간 관계를 클릭한 후 좌측에 토픽 아이디란이 생기면 사용자가 열람하고자 하는 토픽 아이디를 선택한 후 <검색>을 누릅니다. 혹은 (방법 2) 컴포넌트 트리의 ‘저자-이슈 결과’에서 해당 결과문서에 마우스 오른쪽을 클릭한 후 ‘토픽간 관계’을 누릅니다.



우측에 선택란은 결과의 개수를 제한하는 역할을 합니다. 사용자에 필요에 따라 실행되므로 필수적으로 행해져야 되는 것은 아닙니다. 원하는 수치를 적거나 항목을 클릭한 후 <확인>을 클릭하면 결과의 수가 제한된 것을 확인할 수 있습니다.



- No Trimming : 처리를 실행하지 않음
- Probability Drop Trimming : 가장 큰 비중값으로 떨어지는 지점까지만 결과값을 표시
- Log Probability Drop Trimming : 비중값을 Log로 치환한 후 가장 큰 값으로 떨어지는 지점까지만 표시
- Top Topic Matching Trimming: 검색된 단어, 저자, 문서로 해당 토픽을 가장 높은 비중값으로 들고 있음을 확률
- 저장 : 현재 보이는 결과를 엑셀에서 열 수 있는 파일 형태로 저장합니다.
- 선택된 탭 인쇄 : 현재 보이는 결과를 인쇄합니다.

(e-1) 토픽간 관계

토픽간 관계	
토픽ID	유사도
2	0.070249 생각_사람_부동산_가격_국민_아파트_나라_부동_시작_정도_이상_지금_인간_현재_정책_국가_개발_경제_세계_현실
12	0.067752 아이티_지진_구조_발생_피해_활동_현지_지원_구호_대_주민_평화_건물_연합뉴스_규모_서망_상황_기자_국제
4	0.057643 금지_가능_전체_기자_아이폰_결과_관계_배포_동안_가운데_서울_설명_중심_무단_기업_진행_방문_증합_문화_상당
8	0.051057 시장_경제_미국_위기_한국_금융_중국_국가_세계_가능_상황_투자_유럽_우려_두바이_분석_사태_문제_은행_전망
10	0.049303 선언_시국_국민_정치_교사_건강_교육_사회_경계_행위_자유_민주주의_행동_공무원_표현_참여_불법_학생_시민_대한민국
6	0.048602 문제_조사_관련_칠레_배포_금지_기자_무단_포리우스_도박_사과_내용_왜곡_진행_관계_경우_주장_박찬호_설명_확인
5	0.044507 사업_북한_정부_4대강_문제_대통령_회담_이명박_남북_관계_남북정상회담_정상_죽전_예산_관련_가능_원천_대화_개최_입장
19	0.042634 아프간_대통령_정권_사람_자신_세력_작전_파병_미국_생각_남한_언론_시작_노무현_접속_지금_통일_미군_협상_출신
18	0.039954 노무현_대통령_서민_사람_생각_로스쿨_대학_이명박_정치_도입_출신_시험_사전_제도_지도_본인_아래_시절_자식_세력
7	0.038321 연합뉴스_그리스_배포_금지_전재_저작권자_무단_지원_정부_미남_재정_회의_계획_방안_추가_독일_EU_필요_국가_기자
14	0.034232 소마트폰_국내_시장_사건_단체_암살_직원_국정원_북한_경찰_사법_이스라엘_갈등_혁명_정보_자금_인사_사설_두바이_하마스_공작
20	0.030804 김대중_장악_사건_단체_암살_직원_국정원_북한_경찰_사법_이스라엘_갈등_혁명_정보_자금_인사_사설_두바이_하마스_공작
3	0.028800 세종시_의원_수정_국민_한나라당_문제_정치_국회_대표_투표_원안_수도_총리_정부_여론_행정_민주당_반대_토론_박근혜
16	0.025667 기업_개성공단_북한_회장_시간_의상_관광_임주_정부_근로자_기대_자금_수준_기본_제한_통행_재개_경제_기간_변화
17	0.014939 선거_지역_지방_후보_단체_기자_참여_공사_예정_출마_시작_지방선거_투입_평가_공천_기초_실시_대표_사퇴_교과
11	0.010487 도요타_리콜_미국_일본_자동차_차량_구조_제품_세계_연합뉴스_판매_이번_대표_품질_생산_문제_가족_사고_안전
15	0.010413 지역_행사_경기_월드컵_도시_이번_시설_취소_기간_복지_대회_여행_개최_예상_본선_축구_남아공_부분_예정_평가

- i. 토픽 ID : 프로그램이 해당 토픽에 임의로 부여한 숫자입니다.
- ii. 유사도: 해당 토픽이 각 토픽과 얼마나 유사하게 다루어졌는지 수치로 나타냅니다.
수치의 범위는 0 ~ 1 까지이며 1에 가까울수록 유사도가 커집니다.

- iii. 토픽: 단어 사전에 수집된 단어들을 대상으로 프로그램이 서로 관련 있는 단어들을 모아두었습니다. 토픽에서 보여지는 단어들의 순서는 해당 토픽 안에서 가장 비중이 높은 상위의 단어 20 개를 수집한 것입니다.

3. 저자-이슈 분석 처리 결과 보기 – 저자 중심으로 보기

‘저자-이슈 분석’을 실행하고 나면 토픽을 산출하는 것 뿐만 아니라 <III. A. 3. 새 분석 프로젝트 생성 처리결과 보기, page 56> 언급하였듯이 저자 정보도 열람할 수 있게 됩니다. 다시 말하면, 새 분석 프로젝트를 생성하고 난 후 ‘저자’의 작업창은 열람할 수 없으나 ‘저자- 이슈 분석’을 실행하고 나면 이를 열람할 수 있습니다. 컴포넌트 트리의 ‘저자’에 있는 임의의 문서를 클릭했을 때 작업창은 다음과 같습니다.

The screenshot shows a software window titled '홍길동의 프로젝트 - 불광동휘발류 - x'. It has three main tabs:

- a)** 저자명: -불광동휘발류- 글 쓴 횟수: 1회
홍길동 결과 1
ATOM결과보기 [저장] 선택된 템 인쇄
- b)** 토픽

저자-이슈 탐색 결과보기	토픽ID	토픽	비중값	Term Score
	18	노무현_대통령_서민_사람_생각	0.275000	0.556667
	17	선거_지역_지방_후보_단체	0.125000	0.154473
	13	김연_연수_한국_풀림픽_금메달	0.100000	0.101264
	5	사업_북한_정부_4대강_문제	0.050000	0.015975
	8	시장_경제_미국_위기_한국	0.050000	0.015975
	10	선언_시국_국민_정치_교사	0.050000	0.015975
	1	김종률루_한자_감염_백신_예방	0.025000	-0.009341
	3	성격_사람_부동산_가격_국민	0.025000	-0.009341
	3	세종시_의원_수정_국민_한나라당	0.025000	-0.009341
	4	금지_가능_전재_기자_마이플	0.025000	-0.009341
	6	문제_조사_관련_칠레_베포	0.025000	-0.009341
	7	문화뉴스_그리스_해포_금지_전재	0.025000	-0.009341
	9	보도_방송_MBC_언론_판결	0.025000	-0.009341
	11	도오터_리클_미국_인보_자동차	0.025000	-0.009341
- c)** 문서보기

문서ID	문서 작성일	문서 제목	토픽
1000002287	2010-02-02 00:00:00	6월2일 지방선거 이런 공약 내세우는 자 1순위 당선...	노무현_대통령_서민_사람_생각

저자 탭 상단에는 프로젝트 명과 저자가 표기되며 하위 프레임으로는 저자 정보, 관련 토픽, 저자 문서 총 3 가지가 있습니다.

a) 저자 정보

The screenshot shows the 'Author Information' tab with the following fields:

- 저자명: -불광동휘발류-
- 글 쓴 횟수: 1회
- 홍길동 결과 1
- ATOM결과보기 [저장] 선택된 템 인쇄

(a-1) 저자명 : 저자의 이름입니다.

(a-2) 글 쓴 횟수 : 해당 저자가 모든 문서 중에서 글을 쓴 총 횟수입니다.

(a-3) 저자-이슈 분석 결과 : 사용자가 열람하고자 하는 결과 문서를 선택란에서 고를 수 있습니다.

(a-4) ATOM 결과보기 : 해당 결과문서에 대한 ‘저자-이슈 분석’의 결과를 새 탭으로 확인할 수 있습니다.

(a-5) 저장 : 현재 보이는 결과를 엑셀에서 열 수 있는 파일 형태로 저장합니다.

(a-6) 선택된 탭 인쇄 : 현재 보이는 결과를 인쇄한다.

b) 관련 토픽

토픽		토픽ID	토픽	비중값	Term Score
		18	노무현_대통령_서민_사람_생각	0,275000	0,556667
		17	선거_지역_지방_후보_단체	0,125000	0,154473
		13	김연_선수_한국_올림픽_금메달	0,100000	0,101264
		5	사업_북한_정부_4대강_문제	0,050000	0,015975
		8	시장_경제_미국_위기_한국	0,050000	0,015975
		10	선언_시국_국민_정치_교사	0,050000	0,015975
		1	신풍풀루_환자_감염_백신_예방	0,025000	-0,009341
		2	생각_사람_부동산_가격_국민	0,025000	-0,009341
		3	세종시_의원_수정_국민_한나라당	0,025000	-0,009341
		4	금지_가능_전재_기자_마이폰	0,025000	-0,009341
		6	문제_조사_관련_칠레_배포	0,025000	-0,009341
		7	연합뉴스_그리스_배포_금지_전재	0,025000	-0,009341
		9	보도_방송_MBC_언론_판결	0,025000	-0,009341
		11	도요타_리콜_미국_일본_자동차	0,025000	-0,009341

(b-1) 저자-이슈 탐색 결과보기 : 해당 토픽에 관한 저자-이슈 분석 결과를 새 탭으로 확인할 수 있습니다.

(b-2) 토픽 ID : 프로그램이 토픽에 부여한 숫자입니다.

(b-3) 토픽 : 단어 사전에 수집된 단어들을 대상으로 프로그램이 서로 관련 있는 단어들을 모아두었습니다. 토픽에서 보여지는 단어들의 순서는 해당 토픽 안에서 가장 비중이 높은 상위의 단어 20 개를 수집한 것입니다.

(b-4) 비중값: 해당 저자가 각 토픽을 얼마나 비중 있게 다루었는지 수치로 나타낸 것입니다. 수치의 범위는 0~1 까지이며 1에 가까울수록 비중이 커집니다.

c) 저자 문서

문서보기	문서ID	문서 작성일	문서 제목	토픽
	1000002287	2010-02-02 00:00:00	6월2일 지방선거 이런 꼴약 내세우는 자 1순위 당선...	노무현_대통령_서민_사람_생각

- (c-1) 문서보기 : 해당 문서가 새 탭으로 나타납니다.
- (c-2) 문서 ID : 프로그램이 해당 문서에 임의로 부여한 숫자입니다.
- (c-3) 문서 작성일 : 문서가 생성된 일시입니다.
- (c-4) 문서 제목: 문서의 제목입니다.
- (c-5) 토픽 : 해당 문서에서 가장 비중 있게 다뤄지는 토픽이 나타납니다.

G. 저자-이슈 분석 결과 이해하기

정해진 순서대로 저자-이슈 분석을 실시하여 토픽을 얻었습니다. 이제부터 분석 결과를 해석하는 것은 사용자의 판단을 요구합니다. 이번 장에서는 생성된 토픽을 사용자가 어떻게 활용할 수 있는지 알아보겠습니다.

1. 전체 토픽

새 프로젝트를 생성할 시 사용자가 분석대상으로 지정했던 문서들이 어떤 화제에 관해서 가장 비중 있게 다루어졌는지 순위별로 확인할 수 있습니다. 생성된 토픽에 대한 토픽 ID는 화제의 순위 별로 매겨지기 때문에 토픽 ID 1 번에 해당하는 토픽이 가장 비중 있게 다루어졌다고 말할 수 있습니다. 1 순위이기 때문에 당연히 비중값도 제일 높습니다.

전체 토픽			
자세히보기	토픽ID	비중값	토픽
1	1	0.062044	스마트폰_시장_삼성전자_세계_환경_국내_삼성_점유_기자_지수_생산_판매_시작_상승_주도_계획_직원_가격_금융_사무
2	2	0.060365	기자_경찰_후원_낙동강_계획_노무현_현장_하천_장관_축진_출신_환경_여부_구역_활영_시작_예정_협의_지정_외...
3	3	0.058481	판결_재판_언론_두바이_비판_PDF수첩_법_개혁_소송_광우병_사법_국민_단독_청구_보수_무죄_기소_환율_법관_급락
4	4	0.057297	구조_아이티_활동_지진_구호_포르토프랑스_대사_건물_호텔_현장_의료_유엔_대월_잔해_기자_기연_시작_마련_도...
5	5	0.056568	의원_국민_총리_세종시_질문_제출_추진_수정_논란_장관_대변인_공구_대화_시설_논란_기자_배상_기수_국정_기적
6	6	0.055270	회담_남북정상회담_남북_추진_합의_해결_촉구_필요_준비_시기_체제_청와대_투표_원칙_비판_의견_학습_전환_평...
7	7	0.051813	도박_자금_구속_연예인_기소_카지노_공사_미카오_실험_기수_설시_인기_모형_적발_설계_유명_입건_투자_검증...
8	8	0.051256	국내_애물_시장_개발_아이_판매_서비스_휴대폰_페드_전자_국적_효과_모토로이_투자_SK텔레콤_드로_아드_신청...
9	9	0.050933	도요타_리콜_차량_결합_프리우스_자동차_대통_관세_수리_발표_사고_운전_폐발_간사장_제동장치_실시_언론_소...
10	10	0.050385	선거_행사_행정_예정_계획_지도_공천_수입_발표_국토_의원_영장_동원_검토_의장_여성_출재_고통_통합_확실
11	11	0.049825	세종시_수장_해임_여론_도시_간의_대표_원안_지지_대학_한국_회의_이전_미상_발전_현신_사퇴_연구_당론_분열
12	12	0.048996	선수_풀립픽_훈령_한국_밴쿠버_캐나다_세계_밴쿠버_동계올림픽_축구_쇼트트랙_생태_예약_인간_분선_가입_남자...
13	13	0.047761	영화_아바다_감독_판례_총행_세계_개봉_상영_구자철_국내_매출_하임크_나이지리아_이승렬_작품_공격_동원_러...
14	14	0.046648	한국_이동국_홍콩_허정무_감독_이를_칼국_중국_개막_연속_공격수_분부_김보경_헤딩_시기_급복_영상감독_분선...
15	15	0.044872	박찬호_거의_투수_선발_점검_종사_불법_보상_시즌_도로_운동_투입_구원_슬리기_구단_플라멩코_사전_생산_이...
16	16	0.043706	위기_사태_금융_기준_브레이크_세계_품질_건설_기술_재정_총액_모델_고객_사이_도로터자동차_신형_이상_대표...
17	17	0.043310	지방_필요_국제_시작_선거_기관_확인_안정화_문화_방문_역사_업무_국방부_시망_한국_검사_종합_도률_통계_도시
18	18	0.042369	회장_위원_복귀_기자_동계올림픽_유치_미간회_자격_민심_여권_예정_피초비그_한국_활동_전시_시청_애완도글...
19	19	0.039325	중국_교전_경력_훈련_의견_마지막_취재_사과_액시코_브리길_언급_발표_공동_하리_해결_발표_갈래_이규현_대만...
20	20	0.038858	방송_기자_국민_관광_영상_시정_공개_현행_타이타닉_재개_관람_논란_개정_출연_상승_수출_약재_복수국적_부문...

생성된 토픽

스마트폰_시장_삼성전자_세계_환경_국내_삼성_점유_기자_지수_생산_판매_시작_상승_주도_계획_직원_가격_금융_사무

1 번 토픽

1 번 토픽의 단어들의 조합이 ‘스마트폰’, ‘시장’, ‘삼성전자’, ‘세계’가 나온 것으로 보아 ‘삼성전자가 새로운 스마트폰을 세계에 출시하지 않았을까?’라고 추측할 수 있습니다. 사용자가 예상한 내용이 맞는지 확인하기 위해 **자세히 보기**–토픽탭 생성 – 문서로 이동합니다. 이 메뉴는 1 번 토픽에 관하여 가장 비중 있게 다룬 문서들을 보여줍니다. 상위 10 위 안에 랭크되어 있는 문서들의 제목이 모두 삼성전자의 스마트폰 출시에 관련한 것으로 보아 위에서 추측하였던 내용이 맞았다는 것을 알 수 있습니다. 실제 문서를 확인하려면 <문서보기>를 클릭합니다.

토픽

- 삼성, 스마트폰 시장 대반격 나섰다
- (스마트폰혁명)④선수 빼긴 국내사
- [fn사설] 삼성스마트폰, 세계 명품 되려면
- 삼성전자 스마트폰의 ‘반격’ 애플과 점유율 격차 줄인다
- 삼성전자, 스마트폰 주도권 잡겠다
- 삼성 스마트폰의 반격
- 삼성 “스마트폰 3배 이상 팔겠다”
- 스마트폰이 휴대폰업체 실적 ‘좌지우지’
- “올 판매되는 휴대폰 6대 중 1대는 스마트폰”
- 삼성 “스마트폰 3배 더 팔겠다”
- <‘한국폰’ 애플 공세에 수익성 악화>
- 안드로이드폰, 마이폰 마성 넘본다
- 삼성 “스마트폰 3배 더 팔겠다”

1 번 토픽을 많이 다룬 문서의 제목

다시 화제를 바꾸어, 1 번 토픽의 단어들을 살펴보면 ‘기자’, ‘금융’, ‘사무’와 같은 단어들은 방금 확인한 삼성전자의 스마트폰 출시와 100%의 연관성을 보여주지 못합니다. 따라서 사용자가 여타 세부 메뉴를 클릭하여 화제를 정확히 파악할 수 있어야 합니다.

2. 토픽

전체 토픽에서 다루었던 내용을 토대로 4 번 토픽이 아이티 지진 참사 구조활동을 화제로 하였다는 것을 알아냈습니다. 이제부터는 4 번 화제를 비중 있게 다른 저자들을 알아보고자 합니다. 이는 해당 화제에 관하여 어떤 기관이나 단체가 가장 비중 있게 다루고 있는지 살펴볼 때 유용합니다.

문서보기	저자문서보기	문서ID	문서 작성일	비중값	Term Score	저자명	문서 제목
		60004587...	2010-01-24...	0.675676	2,876687	연합뉴스	아이티 지진 11월만에 생존자 구조(종합)
		60004587...	2010-01-24...	0.644737	2,737580	연합뉴스	아이티 지진 11월만에 생존자 구조(종합2보)
		60004583...	2010-01-21...	0.634921	2,571145	연합뉴스	119구조대 아이티서 시체 32구 수습
		70004402...	2010-01-24...	0.576087	2,497503	경향신문	구조 작업 종료... 구호활동에 주력
		60004587...	2010-01-23...	0.564103	2,476592	연합뉴스	아이티 사망자 11만명 확인... 29% 또 구조
		70004404...	2010-01-25...	0.561905	2,462097	세계일보	"수색·구조 활동 종료... 시신 15만구 수습"
		60004592...	2010-01-27...	0.548872	2,370064	연합뉴스	<아이티 한국 구조대... 불무기 방역> 화제
		60004583...	2010-01-21...	0.538462	2,364323	연합뉴스	<아이티... 기적의 생존자 있다라>
		70004393...	2010-01-21...	0.529412	2,421747	경향신문	세계가 팔걸은 구호... 항구도 기능 회복
		60004585...	2010-01-22...	0.528302	2,193206	연합뉴스	아이티 파견 119구조대 25일 귀국
		70004393...	2010-01-21...	0.515625	2,039749	강원일보	아이티 지진 기적의 생존자
		60004590...	2010-01-25...	0.513889	2,220126	연합뉴스	<긴급구호대원이 된 아이티 향상>
		60004590...	2010-01-25...	0.511364	2,039480	연합뉴스	아이티 파견 1차 긴급구호대 귀국
		70004398...	2010-01-22...	0.508475	2,204435	파이낸...	아이티 파견 119구조대 25일 귀국

4 번 토픽을 클릭하였을 때의 결과

토픽— 아이디 — 4 —저자를 클릭합니다. 이 메뉴는 4 번 토픽을 비중 있게 다른 저자들을 보여줍니다.

1 순위부터 5 순위까지 랭크되어 있는 저자들은 연합뉴스, 연합뉴스, 연합뉴스, 경향신문, 연합뉴스 신문 순입니다. 이를 보아 연합뉴스가 가장 외신을 많이 다룬 것을 알 수 있습니다.

생성문서보기	저자ID	저자명	비중값	Term Score
[*]	60004587505	연합뉴스	0.675676	2,876687
[*]	60004587844	연합뉴스	0.644737	2,737580
[*]	60004583467	연합뉴스	0.634921	2,571145
[*]	70004402430	경향신문	0.576087	2,497503
[*]	60004587100	연합뉴스	0.564103	2,476592

4 번 토픽에 대한 저자 정보

3. 저자

토픽에서 다루었던 내용을 토대로 연합뉴스가 외신 관련 기사를 많이 다룬다는 것을 알아냈습니다. 이제부터는 연합뉴스의 저자들 중 임의의 저자를 선택하여 그가 어떤 토픽을 가장 비중 있게 다루며 그가 쓴 문서는 어떤 것이 있는지 살펴보도록 하겠습니다. 이는 특정 인물이 어떤 화제를 주로 다루며 어떤 글을 많이 썼는지 살펴볼 때 유용합니다.

저자 — 아이디 — 600004587505 — 토픽 을 클릭합니다. 이 메뉴는 해당 저자가 어떤 토픽을 가장 비중 있게 다루었는지 보여줍니다. 1 순위와 2 순위의 비중값의 차이가 크게 나는 것을 감안할 때 1순위인 4 번 토픽에 관해 가장 비중 있게 다룬 것을 알 수 있습니다.

관련문서보기	토픽ID	토픽	비중값	Term Score
4	4	구조_마이티_활동_지진_구호	0.675676	0.901223
8	8	국내_애플_시장_개발_아이	0.027027	0.497248
12	12	선수_올림픽_훈련_한국_밴쿠버	0.027027	0.125627
14	14	한국_미동국_홍콩_허정무_감독	0.027027	0.055432
17	17	지방_필요_국제_시작_선거	0.027027	0.055432
19	19	중국_파견_경력_훈련_의견	0.027027	0.031909
1	1	스마트폰_시장_삼성전자_세계_환경	0.013514	0.021459
2	2	기자_경찰_후원_낙동강_계획	0.013514	0.021459
3	3	판결_재판_언론_두바이_비판	0.013514	0.003976
5	5	의원_국민_총리_세종시_질문	0.013514	-0.002515
6	6	회담_남북장상회담_남북_추진_합의	0.013514	-0.006842
7	7	도박_자금_구속_면예인_기소	0.013514	-0.006842
9	9	도요타_리콜_차량_결함_프리우스	0.013514	-0.006842
10	10	선거_행사_행정_예정_계획	0.013514	-0.007836

이를 쉽게 확인하려면 좌측 선택란에서 ‘Probability Drop Trimming’을 클릭하면 가장 큰 값으로 다루어진 토픽을 한번에 알 수 있습니다.

프로젝트명:	홍길동의 프로젝트	저자명:	연합뉴스										
선택:	저자	토픽수:	<input type="checkbox"/> All										
아이디:	60004587505	문서수:	<input type="checkbox"/> All										
<input type="button" value="검색"/>		정제방법:	Probability Drop Trimming										
<input type="radio"/> 토픽 <input type="radio"/> 문서													
<table border="1"> <thead> <tr> <th>관련문서보기</th> <th>토픽ID</th> <th>토픽</th> <th>비중값</th> <th>Term Score</th> </tr> </thead> <tbody> <tr> <td></td> <td>4</td> <td>구조_아이티_활동_지진_구호</td> <td>0,675676</td> <td>0,901223</td> </tr> </tbody> </table>				관련문서보기	토픽ID	토픽	비중값	Term Score		4	구조_아이티_활동_지진_구호	0,675676	0,901223
관련문서보기	토픽ID	토픽	비중값	Term Score									
	4	구조_아이티_활동_지진_구호	0,675676	0,901223									

저자 – 아이디 – 60004587505 – 문서를 클릭합니다. 이 메뉴는 해당 저자가 어떤 문서들을 가장 비중 있게 다루었는지 보여줍니다. 문서가 1개 나온 것을 감안할 때 이 저자가 쓴 글은 단 1개 뿐이며 그 문서가 4 번 토픽과 관련 있는 글이라는 것을 다시 한번 확인할 수 있습니다.

프로젝트명:	홍길동의 프로젝트	저자명:	연합뉴스																																								
선택:	저자	토픽수:	<input type="checkbox"/> All																																								
아이디:	60004587505	문서수:	<input type="checkbox"/> All																																								
<input type="button" value="검색"/>		정제방법:	No Trimming																																								
<input type="radio"/> 토픽 <input type="radio"/> 문서																																											
<table border="1"> <thead> <tr> <th>토픽보기</th> <th>문서ID</th> <th>문서 작성일</th> <th>문서 제목</th> </tr> </thead> <tbody> <tr> <td>*</td> <td>60004587505</td> <td>2010-01-24 07:54:27</td> <td>아이티 지진 11일만에 생존자 구조(종합)</td> </tr> </tbody> </table>				토픽보기	문서ID	문서 작성일	문서 제목	*	60004587505	2010-01-24 07:54:27	아이티 지진 11일만에 생존자 구조(종합)																																
토픽보기	문서ID	문서 작성일	문서 제목																																								
*	60004587505	2010-01-24 07:54:27	아이티 지진 11일만에 생존자 구조(종합)																																								
<table border="1"> <thead> <tr> <th>토픽ID</th> <th>토픽</th> <th>비중값</th> <th>Term Score</th> </tr> </thead> <tbody> <tr> <td>4</td> <td>구조_아이티_활동_지진_구호</td> <td>0,675676</td> <td>0,901223</td> </tr> <tr> <td>8</td> <td>국내_애플_시장_개발_아이</td> <td>0,027027</td> <td>0,497248</td> </tr> <tr> <td>12</td> <td>선수_풀립피_훈련_한국_밴쿠버</td> <td>0,027027</td> <td>0,125627</td> </tr> <tr> <td>14</td> <td>한국_미동국_홍콩_허정무_감독</td> <td>0,027027</td> <td>0,055432</td> </tr> <tr> <td>17</td> <td>지방_필요_국제_시작_선거</td> <td>0,027027</td> <td>0,055432</td> </tr> <tr> <td>19</td> <td>중국_파견_경력_훈련_의견</td> <td>0,027027</td> <td>0,031909</td> </tr> <tr> <td>1</td> <td>스마트폰_시장_삼성전자_세계_환경</td> <td>0,013514</td> <td>0,021459</td> </tr> <tr> <td>2</td> <td>기자_경찰_후원_낙동강_계획</td> <td>0,013514</td> <td>0,021459</td> </tr> <tr> <td>3</td> <td>파견_재판_언론_두바이_비판</td> <td>0,013514</td> <td>0,003976</td> </tr> </tbody> </table>				토픽ID	토픽	비중값	Term Score	4	구조_아이티_활동_지진_구호	0,675676	0,901223	8	국내_애플_시장_개발_아이	0,027027	0,497248	12	선수_풀립피_훈련_한국_밴쿠버	0,027027	0,125627	14	한국_미동국_홍콩_허정무_감독	0,027027	0,055432	17	지방_필요_국제_시작_선거	0,027027	0,055432	19	중국_파견_경력_훈련_의견	0,027027	0,031909	1	스마트폰_시장_삼성전자_세계_환경	0,013514	0,021459	2	기자_경찰_후원_낙동강_계획	0,013514	0,021459	3	파견_재판_언론_두바이_비판	0,013514	0,003976
토픽ID	토픽	비중값	Term Score																																								
4	구조_아이티_활동_지진_구호	0,675676	0,901223																																								
8	국내_애플_시장_개발_아이	0,027027	0,497248																																								
12	선수_풀립피_훈련_한국_밴쿠버	0,027027	0,125627																																								
14	한국_미동국_홍콩_허정무_감독	0,027027	0,055432																																								
17	지방_필요_국제_시작_선거	0,027027	0,055432																																								
19	중국_파견_경력_훈련_의견	0,027027	0,031909																																								
1	스마트폰_시장_삼성전자_세계_환경	0,013514	0,021459																																								
2	기자_경찰_후원_낙동강_계획	0,013514	0,021459																																								
3	파견_재판_언론_두바이_비판	0,013514	0,003976																																								

4. 문서

임의의 문서에 대한 저자와 토픽에 관해 알아보겠습니다. 이는 해당 문서를 중심으로 이 글을 쓴 저자가 누구인지 그리고 이 글의 화제는 무엇인지 알아 보고자 할 때 유용합니다.

문서 — 아이디 — 60004587505 — 저자를 클릭합니다. 이 메뉴는 해당 문서의 저자를 보여줍니다. 일반적으로 문서에 대한 저자는 단 1명이기 때문에 목록에는 1명의 저자가 뜨게 되며 비중값도 항상 1이 나오게 됩니다.

생성문서보기	저자ID	저자명	비중값
[file icon]	60004256428	연합뉴스	1,000000

문서 — 아이디 — 60004587505 — 토픽 을 클릭합니다. 이 메뉴는 해당 문서에 관하여 가장 비중 있게 다뤄진 토픽을 보여줍니다. 1순위와 2순위의 비중값의 차이가 크게 나는 것으로 보아 1순위인 7번 토픽에 관해 가장 비중 있게 다룬 것을 알 수 있습니다. 이를 쉽게 확인하려면 우측 선택란에서 'Probability Drop Trimming'을 클릭하면 가장 큰 값으로 다루어진 토픽을 한번에 알 수 있습니다.

프로젝트명:	홍길동의 프로젝트	문서명:	황기순, 전국																																																												
선택:	문서	단어수:																																																													
아이디:	60004256428 	저자수:																																																													
검색		정제방법: No Trimmir																																																													
<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th>저자</th> <th>토픽</th> <th>비중값</th> <th>Term Score</th> </tr> </thead> <tbody> <tr><td></td><td>기도박_자금_구속_연예인_기소</td><td>0,238095</td><td>0,855056</td></tr> <tr><td></td><td>2기?문서 :원_낙동강_계획</td><td>0,111111</td><td>0,190387</td></tr> <tr><td></td><td>10선거_할사_행정_예정_계획</td><td>0,111111</td><td>0,111417</td></tr> <tr><td></td><td>1 스마트폰_시장_삼성전자_세계_환경</td><td>0,079365</td><td>0,091809</td></tr> <tr><td></td><td>15박찬호_계약_투수_선발_점검</td><td>0,063492</td><td>0,050143</td></tr> <tr><td></td><td>19 중국_파견_경력_훈련_의견</td><td>0,063492</td><td>0,029489</td></tr> <tr><td></td><td>8 국내_애플_시장_개발_마이</td><td>0,047619</td><td>0,024764</td></tr> <tr><td></td><td>17 지방_필요_국제_시작_선거</td><td>0,047619</td><td>0,015926</td></tr> <tr><td></td><td>5 의원_국민_총리_세종시_질문</td><td>0,031746</td><td>0,011842</td></tr> <tr><td></td><td>13 영화_마비타_감독_관객_흥행</td><td>0,031746</td><td>0,001176</td></tr> <tr><td></td><td>18 회장_위원회_복귀_기자_동계올림픽</td><td>0,031746</td><td>0,001176</td></tr> <tr><td></td><td>3 판결_재판_언론_두바이_비판</td><td>0,015873</td><td>-0,004321</td></tr> <tr><td></td><td>4 구조_마이티_활동_지진_구호</td><td>0,015873</td><td>-0,008094</td></tr> <tr><td></td><td>6 회담_남북정상회담_남북_추진_합의</td><td>0,015873</td><td>-0,008094</td></tr> </tbody> </table>				저자	토픽	비중값	Term Score		기도박_자금_구속_연예인_기소	0,238095	0,855056		2기?문서 :원_낙동강_계획	0,111111	0,190387		10선거_할사_행정_예정_계획	0,111111	0,111417		1 스마트폰_시장_삼성전자_세계_환경	0,079365	0,091809		15박찬호_계약_투수_선발_점검	0,063492	0,050143		19 중국_파견_경력_훈련_의견	0,063492	0,029489		8 국내_애플_시장_개발_마이	0,047619	0,024764		17 지방_필요_국제_시작_선거	0,047619	0,015926		5 의원_국민_총리_세종시_질문	0,031746	0,011842		13 영화_마비타_감독_관객_흥행	0,031746	0,001176		18 회장_위원회_복귀_기자_동계올림픽	0,031746	0,001176		3 판결_재판_언론_두바이_비판	0,015873	-0,004321		4 구조_마이티_활동_지진_구호	0,015873	-0,008094		6 회담_남북정상회담_남북_추진_합의	0,015873	-0,008094
저자	토픽	비중값	Term Score																																																												
	기도박_자금_구속_연예인_기소	0,238095	0,855056																																																												
	2기?문서 :원_낙동강_계획	0,111111	0,190387																																																												
	10선거_할사_행정_예정_계획	0,111111	0,111417																																																												
	1 스마트폰_시장_삼성전자_세계_환경	0,079365	0,091809																																																												
	15박찬호_계약_투수_선발_점검	0,063492	0,050143																																																												
	19 중국_파견_경력_훈련_의견	0,063492	0,029489																																																												
	8 국내_애플_시장_개발_마이	0,047619	0,024764																																																												
	17 지방_필요_국제_시작_선거	0,047619	0,015926																																																												
	5 의원_국민_총리_세종시_질문	0,031746	0,011842																																																												
	13 영화_마비타_감독_관객_흥행	0,031746	0,001176																																																												
	18 회장_위원회_복귀_기자_동계올림픽	0,031746	0,001176																																																												
	3 판결_재판_언론_두바이_비판	0,015873	-0,004321																																																												
	4 구조_마이티_활동_지진_구호	0,015873	-0,008094																																																												
	6 회담_남북정상회담_남북_추진_합의	0,015873	-0,008094																																																												

여기에서 7 번 토픽에 관한 <연관문서보기>를 클릭하면 ‘문서 아이디 60004256428’에 대한 문서들이 나열되는 것이 아니라 7 번 토픽을 비중 있게 다른 문서들이 하단에 나열됩니다. 이것은 **토픽—아이디—7—문서** 을 선택해도 같은 결과가 나타납니다.

문서보기	문서ID	문서 작성일	토픽	비중값	Term Score	문서 제목
	70003657189	2009-08-10 0...	도박_자금_구속_연예...	0,615385	2,016455	인기 개그맨도 마카오 원정도박
	70003659610	2009-08-10 1...	도박_자금_구속_연예...	0,582192	2,006824	연예인 마카오 원정도박
	60004278265	2009-08-09 0...	도박_자금_구속_연예...	0,573171	1,925368	[천900억원대 마카오 원정 도박단 적발]
	70003961415	2009-10-22 2...	도박_자금_구속_연예...	0,567901	1,835917	마카오 원정도박단 '철퇴'
	70003654483	2009-08-08 1...	도박_자금_구속_연예...	0,567164	1,758044	연예인 도박 또 적발! 인기가수 A씨 등 5명 포함
	70003962388	2009-10-22 2...	도박_자금_구속_연예...	0,567164	1,754483	마카오원정 도박단 무더기 적발
	70003655625	2009-08-01 1...	도박_자금_구속_연예...	0,554622	1,600823	활개치는 해외원정도박, 사람위는 뭐하나?
	60004277411	2009-08-07 2...	도박_자금_구속_연예...	0,544118	1,601860	해외원정 도박 연예인 등 무더기 적발!(종합)
	70003682522	2009-08-16 1...	도박_자금_구속_연예...	0,539823	1,411845	[주간이슈] 또 연예인 끝 원정도박 외
	70003685373	2009-08-17 0...	도박_자금_구속_연예...	0,533784	1,467066	또 블러드전 연예계 도박 망령…도박에 빠지는 연예인 왜'
	70003654379	2009-08-08 1...	도박_자금_구속_연예...	0,532258	1,415368	인기 연예인 끝 해외원정 도박
	60004277398	2009-08-07 2...	도박_자금_구속_연예...	0,531250	1,265750	해외 원정도박 유명 연예인 겹찰 적발

5. 토픽간 관계

토픽간 관계에서는 해당 토픽이 각각의 토픽과 얼마나 유사한지 열람할 수 있습니다. 예를 들어 분석 대상인 문서들이 가장 비중 있게 다룬 화제가 4 개라고 가정합시다. 하지만 ‘저자-이슈 분석’을 실행할 당시 사용자가 토픽의 개수를 임의로 지정하였기 때문에 사용자가 정확히 4 개를 지정하지 않는 이상 본래의 화제의 개수와 일치하지 않습니다. 만약 사용자가 토픽의 개수로 20 을 지정하였다면 본래의 4 개인 토픽이 20 개로 불어나, 결국 사용자는 4 개의 화제를 20 개로 불려서 판단하게 됩니다. 이와 같은 상황을 감안하여 결과값의 정확성을 더욱 높이기 위해 토픽간 관계에서는 토픽간의 유사도를 측정하여 생성된 토픽이 실제로는 동일한 화제를 다루고 있는지 확인할 수 있습니다.

11 번 토픽을 살펴봅시다.

의원_국민_총리_세종시_질문_제출_추진_수정_논란_장관_대변인_공구_대화_사설_훈란_기각_배상_기수_국정_기적

이 토픽의 내용은 위에서 분석하였던 과정을 토대로 세종시 원안 분쟁에 관한 내용이란 것을 알아냈습니다. 그렇다면 18 번 토픽과 가장 유사한 토픽을 열람해 보겠습니다. **토픽간 관계 –아이디 – 5 – 검색**을 클릭합니다. 검색 결과는 5 번 토픽을 제외한 모든 토픽이 나타납니다. 토픽은 해당 토픽과 가장 유사하게 다뤄진 순서대로 나열됩니다.

프로젝트명:	홈길동의 프로젝트	토평명:	의원_국민_총리_세종시_집문
선택:	도파간 판례	제작:	
아이디:	5 Click	제작일:	2018-08-28
		제작자:	김현수
		제작단위:	경상남도
		제작방법:	No Trimming
		저장:	저장
		선택된 템 인쇄	
		검색	

1 순위는 11 번 토픽으로 나타났습니다. 이는 5 번 토픽이 11 번 토픽과 가장 유사성이 높다는 것을 뜻합니다.

세종시_수정_해임_여론_도시_건의_대표_원안_지지_대학_찬성_회의_미전_미상_발전_혁신_사회_연구_당론_분열

11 번 토픽

그렇다면 11 번 토픽은 어떤 문서들에서 주로 나타났는지 확인하기 위해 토픽— 아이디— 11 — 문서 를 클릭합니다.



프로젝트명:	홍길동의 프로젝트	토픽명:	세종시_수정_해임_여론_도시	선택:	선택: 토픽	선택수:	<input type="checkbox"/> 셀
아이디:	11 Click	문서수:	<input type="checkbox"/> All	검색:		정제방법:	No Trimming
저장 문서내보내기 선택							
문서보기	저자문서보기	문서ID	문서 작성일	Term S...	저자명	문서 제목	
[1]		70004476607	2010-02-08 0...	2,876687	마전일보	세종시 원안 지지율, 수정안 앞서	
[2]		70004477996	2010-02-08 0...	2,737580	충부매일	“세종시 원안 지지” 수정안보다 앞서	
[3]		7000447107	2010-02-07 2...	2,571145	한겨레	‘원안’ 여론 높아지고 ‘천이’ 고민 깊어가고	
[4]		70004478772	2010-02-08 0...	2,497503	서울신문	[사설] 청 출리 해임안 세종시 해별 본질 아니다	
[5]		70004476976	2010-02-07 2...	2,476592	전남일보	‘나주혁신도시 정부지원 확대해야’	
[6]		70004478452	2010-02-08 0...	2,462097	출혈투데이	【박건옥 칼럼】‘세종시’ 오래 끌어봐야...	
[7]		70004476781	2010-02-07 2...	2,370064	충도일보	천미 천복 세종시 수정안 충력전 “설미 최대 분수령”	
[8]		70004474986	2010-02-07 1...	2,354323	파이낸셜...	정 윤찬 정치권 ‘꼼꼼의 적’, 해임안 가결 촉각	
[9]		70004476793	2010-02-07 2...	2,421747	충도일보	“총리해임안” 세종시 정국 광점 부상	
[10]		70004478493	2010-02-08 0...	2,193206	마전일보	여야 결연후 세종시 충보 출력전	
[11]		70003654228	2009-08-08 0...	2,039749	부산일보	“자체혁 과신하면 유혹에 쉽게 빠져”	
[12]		70004475345	2010-02-07 1...	2,220126	파이낸셜...	세종시 정면다급로 치닫는 정치권	
[13]		70004480861	2010-02-08 1...	2,039480	광주일보	‘親朴도’ 정총리 해임안 통조론	
[14]		70004479230	2010-02-08 0...	2,204435	영남일보	천미-천박 세종시 싸움 금주 최고조	

토픽 ID 11 번을 클릭한 화면

문서 제목

세종시 원안 지지율, 수정안 앞서
 "세종시 원안 지지" 수정안보다 앞서
 '원안' 여론 높아지고 '친미' 고민 깊어가고
 [사설] 정 총리 해임안 세종시 해법 본질 아니다
 "나주혁신도시 정부지원 확대해야"
 [박건옥 칼럼] '세종시' 오래 끌어봐야…
 친미·친박 세종시 수정안 출력전 "설이 최대 분수령"
 정운찬, 정치권 '공공의 적'..해임안 가결 촉각
 '총리해임안' 세종시 정국 쟁점 부상
 여야, 설연휴 세종시 홍보 출력전
 "자제력 과신하면 유혹에 쉽게 빠져"
 세종시 정면대결로 치닫는 정치권
 親朴도 '정총리 해임안' 동조론
 친미-친박 세종시 싸움 금주 최고조

문서 제목을 확대한 화면

문서들의 문서 제목을 살펴보면 11 번 토픽의 내용도 5 번 토픽과 같이 세종시 원안에 관한 분쟁을 주로 다루었다는 것을 파악할 수 있습니다. 이는 11 번 토픽과 5 번 토픽이 실제로는 동일한 화제이나 사용자가 토픽 수를 지정하는 부분에서 나뉘어졌다는 것도 알 수 있습니다.

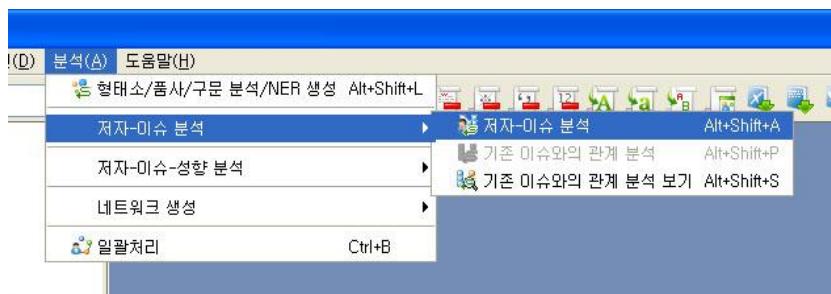
사용자가 분석 문서들만 가지고 실제 몇 개의 화제가 다루어졌는지 파악하는 것은 어렵기 때문에 토픽간 관계 을 통하여 토픽의 수를 활용하면 정확하게 파악합니다.

H. 저자-이슈-성향 분석 실행하기

1. 저자-이슈-성향 분석

저자-이슈 성향분석에서는 문서들을 분석 대상으로 사용자의 주관적인 기준을 설정하여 어떤 화제가 어떤 성향을 가지고 다루어졌는지 파악합니다. 앞에서 다루었던 저자-이슈 분석과는 다른 개념으로 문서에서 다루어지는 주된 토픽뿐만 아니라 토픽의 성향까지 파악할 수 있습니다. 무엇보다도 ‘저자-이슈-성향 분석’에서는 저자-이슈 분석과 같이 분석 과정보다 이를 실행하여 얻어지는 결과 메뉴들에 대한 충분한 이해력과 활용력이 요구되며 지금부터 저자-이슈 성향 분석 과정을 살펴보도록 하겠습니다.

- a) ‘저자-이슈-성향 분석’을 실행하려면 **뉴막대 – 분석 – 저자-이슈- 성향 분석 - 저자-이슈-성향 분석**을 누릅니다.



- b) ‘저자(A)-토픽(T)-의견(O) 분석’ 대화상자가 나타납니다.



(b-1) 프로젝트 선택에는 적용 대상이 되는 프로젝트입니다. 우측에 <탐색>을 클릭하여 해당 프로젝트를 선택한 후 <확인>을 클릭합니다.

(b-2) 결과 ID에는 사용자가 원하는 결과명을 기입합니다.

(b-3) 샘플링 수는 ‘저자-이슈-성향 분석’을 더욱 정밀하게 하기 위하여 반복적으로 수행하는 통계처리 횟수입니다.

(b-4) 쓰레드 수는 샘플링을 병렬적으로 수행하기 위하여 지정하는 동시 작업횟수입니다.

(b-5) 토픽 수에는 몇 개의 토픽을 추출할 것인지 기입합니다.

(b-6) 단어사전 선택에는 적용하고자 하는 단어사전을 체크합니다.

(b-7) AT 정제 : 산출되는 저자의 비중값을 정제합니다.

(b-8) AO 정제 : 산출되는 저자의 성향값을 정제합니다.

(b-9) DT 정제 : 산출되는 문서의 비중값을 정제합니다.

(b-10) DO 정제 : 산출되는 문서의 성향값을 정제합니다.

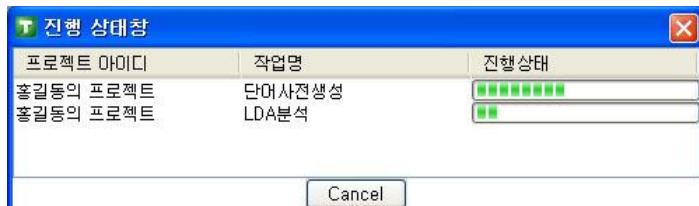
(b-11) WT 정제 : 산출되는 단어의 비중값을 정제합니다.

(b-12) WO 정제 : 산출되는 단어의 성향값을 정제합니다.

- No Trimming : 처리를 실행하지 않음
- Probability Drop Trimming : 가장 큰 비중값으로 떨어지는 지점까지만 결과값을 표시
- Log Probability Drop Trimming : 비중값을 Log 로 치환한 후 가장 큰 값으로 떨어지는 지점까지 표시
- Top Topic Matching Trimming : 검색된 단어, 저자, 문서로 해당 토픽을 가장 높은 비중값으로 들고 있을 확률

‘저자(A)－토픽(T)－의견(O) 분석’의 권장수치는 샘플링 수에는 1000 개 문서 시 500 회, 쓰레드 수는 CPU 의 Core 개수, 토픽 수는 1000 개 기준 20 개입니다. 단 ‘저자-이슈-성향분석’에서는 긍정 성향의 토픽 20 개와 부정성향의 토픽 20 개가 생성되어 주어진 토픽의 2 배만큼의 토픽이 생성됩니다. 설정이 끝나면 <확인>을 클릭합니다.

- c) ‘진행 상태창’ 대화상자가 나타나면 연두색 막대가 채워질 때까지 기다립니다. 이때 <Cancel>을 클릭하면 ‘자연어 처리’가 중단됩니다.



- d) 작업이 완료되면 컴포넌트 트리의 ‘저자-이슈-성향 결과’에 분석의 결과¹가 생성됩니다. 생성된 결과를 더블 클릭하면 작업창에 새 탭이 나타나며 프로그램에서 추출한 ‘전체토픽, 저자, 토픽, 문서’의 값을 확인할 수 있습니다.

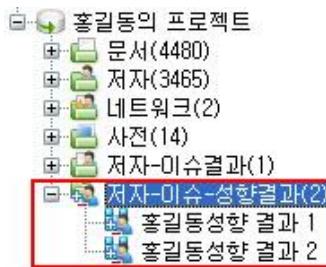
¹ ‘저자-이슈-성향 결과’에서 결과명과 결과 아이디는 동일하게 사용합니다. ‘문서’에서도 문서명과 문서 아이디는 동일하게 사용합니다.



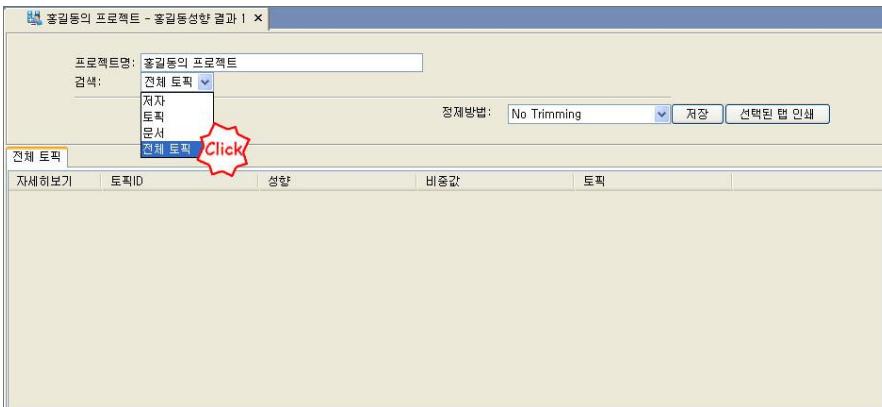
- e) 생성된 분석 결과는 파일의 형태로 사용자가 지정한 작업폴더(최초 설정- C://Workplace, page 29, (d-2) 참조)에 자동 저장되며 메시지 판넬의 속성에서 저장된 위치 등 상세정보를 확인할 수 있습니다.

2. 저자-이슈-성향 분석 처리결과 보기

컴포넌트 트리의 ‘저자-이슈-성향 결과’는 ‘저자-이슈-성향 분석’을 실행한 결과가 나타나는 곳입니다. 나열된 결과 문서들은 결과명으로 보여지게 되며 괄호 안의 숫자는 결과문서의 개수를 의미합니다.



‘저자-이슈-성향 결과’에서는 결과명과 결과 아이디를 동일하게 사용합니다. 여기에서 **특이사항은 원칙적으로 프로그램이 결과에 대한 아이디를 부여 해야 하나, 사용자가 ‘저자-이슈-성향 분석 대화상자—결과 ID’에 결과 아이디를 직접 기입함으로써 아이디를 부여했다는 점입니다.** 생성된 결과를 삭제하고 싶을 경우 해당 사전에 마우스 오른쪽 버튼을 눌러 삭제를 클릭합니다.



탭 상단에는 프로젝트명과 결과명이 표기됩니다. 결과 탭의 하위 선택란에는 전체 토픽, 저자, 토픽, 문서를 포함하여 총 4 가지 메뉴가 있습니다.

정제방법은 <검색>을 눌렀을 때 표시되는 결과의 개수를 제한하는 역할을 합니다.

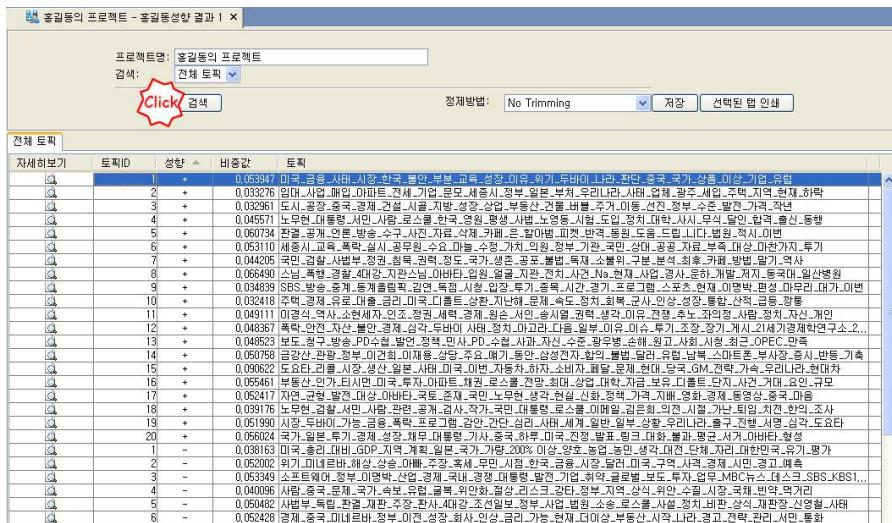


- No Trimming : 처리를 실행하지 않음
- 비중값 Drop Trimming : 가장 큰 비중값으로 떨어지는 지점까지만 결과값을 표시
- Log 비중값 Drop Trimming : 비중값을 Log로 치환한 후 가장 큰 값으로 떨어지는 지점까지 표시
- Top 토픽 Matching Trimming : 검색된 단어, 저자, 문서로 해당 토픽을 가장 높은 비중값으로 들고 있을 확률
- 저장 : 현재 보이는 결과를 엑셀에서 열 수 있는 파일 형태로 저장합니다.
- 선택된 템 인쇄 : 현재 보이는 결과를 인쇄합니다.

a) 전체 토픽

전체 토픽은 ‘저자-이슈-성향 분석’을 실행한 후 생성되는 결과메뉴 중 하나로, 추출한 토픽과 성향을 한눈에 확인할 수 있습니다. 초기 화면에는 성향별로 토픽이 나열되며 원하는 항목을 클릭하여 그 항목을 기준으로 토픽을 나열할 수 있습니다. 예를 들어 비중값 항목을 클릭하면 비중값을 기준으로 토픽이 나열됩니다.

(방법 1)선택란에서 전체 토픽을 클릭한 후 <검색>을 누릅니다. 혹은 (방법 2) 컴포넌트 트리의 ‘저자-이슈-성향 결과’에서 해당 결과문서에 마우스 오른쪽을 클릭한 후 ‘전체 토픽’을 누릅니다.



- (a-1) 자세히 보기 : 해당 토픽¹으로 새 탭이 생성되어 이동합니다.

(a-2) 토픽 ID : 프로그램이 비중값의 수치²를 내림차순으로 각 토픽에 부여한 숫자입니다.

(a-3) 성향 : 사용자의 주관으로 형성한 성향 단어사전을 사용하여 토픽의 성향을 구분합니다. +는 긍정 성향이며 -는 부정 성향입니다.

(a-4) 비중값 : 각 토픽이 모든 문서³에서 얼마나 비중 있게 다루어졌는지 수치로 나타낸 것입니다. 수치의 범위는 0~1 까지이며 1에 가까울수록 비중이 커집니다.

(a-4) 토픽: 단어사전에 수집된 단어들을 대상으로 프로그램이 서로 관련 있는 단어들을 그룹화 시켜 놓은 것입니다. 토픽의 총 개수는 ‘저자(A)-토픽(T)-의견(O) 분석 대화상자-토픽 수’에서 사용자가 지정하였던 개수입니다⁴. 단어들은 하이픈으로 연결되어 있으며 20개씩⁵ 모아져 있습니다.

¹ page 95, c)Topic 을 참조합니다.

² 아이디(ID)란 프로그램이 해당 정보에 ‘임의로’ 부여하는 숫자입니다. 예외적으로 Topic-Title은 Probability의 수치를 기준으로 아이디가 부여됩니다.

³ 문서에 수집된 모든 문서를 말합니다.

⁴ 예를 들어 ‘저자(A)-토픽(T) 분석 대화상자-토픽 수’에서 10을 기제하였다면 총 토픽의 개수는 10개가 됩니다.

5 토픽에서 보여지는 단어들의 순서는 해당 토픽 안에서 가장 비중이 높은 상위의 단어 20 개를 수집한 것입니다.

b) 저자

저자는 ‘저자-이슈-성향 분석’을 실행한 후 생성되는 결과메뉴 중 하나로, 문서를 작성한 개개의 저자가 추출된 토픽들 중에서 어떤 토픽을 어떤 성향으로 가장 비중 있게 다루었는지 그리고 저자가 작성한 문서는 어떤 것들이 있는지를 열람할 수 있습니다. 초기 화면에는 성향별로 토픽이 나열되며 원하는 항목을 클릭하여 그 항목을 기준으로 토픽을 나열할 수 있습니다. 예를 들어 비중값 항목을 클릭하면 비중값을 기준으로 토픽이 나열됩니다.

(방법 1)선택란에서 저자를 클릭한 후 하단에 저자 아이디란이 생기면 사용자가 열람하고자 하는 저자 아이디를 선택한 후 <검색>을 누릅니다. 혹은 (방법 2)컴포넌트 트리의 ‘저자-이슈-성향 결과’에서 해당 문서에 마우스 오른쪽을 클릭한 후 ‘저자’를 누릅니다.

관련문서보기	토픽ID	토픽	비중값	Term Score
[redacted]	1	3세증시_의원_수정_국민_한나라당	0.272000	0.651116
[redacted]	2	상자_사람_부동산_가격_국민	0.192000	0.392737
[redacted]	9	보도_방송_MBC_언론_판결	0.104000	0.148969
[redacted]	18	노무현_대통령_서민_사법_생각	0.080000	0.093603
[redacted]	4	글자_기능_전재_기자_아이	0.072000	0.076657
[redacted]	19	오프간_대통령_경찰_사법_자신	0.049000	0.031642
[redacted]	6	문제_조사_관련_질례_배포	0.040000	0.019075
[redacted]	14	스마트폰_국내_시장_사용_업체	0.040000	0.019075
[redacted]	16	기업_개성공단_북한_회장_시간	0.032000	0.006120
[redacted]	1	신종플루_환자_감염_백신_예방	0.024000	-0.000815
[redacted]	7	연합뉴스_그리스_배포_글자_전재	0.016000	-0.007030
[redacted]	11	도요타_리콜_미국_일본_자동차	0.016000	-0.007030

문서보기	문서ID	문서작성일	토픽	비중값	Term Score	문서제목
[redacted]						

좌측에 선택란은 결과의 개수를 제한하는 역할을 합니다. 사용자에 필요에 따라 실행되므로 필수적으로 행해져야 되는 것은 아닙니다. 원하는 수치를 적거나 항목을 클릭한 후 <확인>을 클릭하면 결과의 수가 기입한 수만큼 제한된 것을 확인할 수 있습니다.



- 저자명 : 해당 저자 아이디에 대한 저자명을 나타냄
- 토픽수 : 결과 값으로 나타나는 토픽의 개수를 제한
- 문서수 : 결과 값으로 나타나는 문서의 개수를 제한
- No Trimming : 처리를 실행하지 않음
- Probability Drop Trimming : 가장 큰 비중값으로 떨어지는 지점까지만 결과값을 표시
- Log Probability Trimming : 비중값을 Log 로 치환한 후 가장 큰 값으로 떨어지는 지점까지 표시
- Top Topic Matching Trimming : 검색된 단어, 저자, 문서로 해당 토픽을 가장 높은 비중값으로 들고 있을 확률
- 저장 : 현재 보이는 결과를 엑셀에서 열 수 있는 파일 형태로 저장합니다.
- 문서 내보내기: 현재 보여지는 문서들을 원래 데이터 형태대로 지정된 폴더에 저장한다.
- 선택된 템 인쇄 : 현재 보이는 결과를 인쇄합니다.

(b-1) 토픽

토직	문서				
관련문서보기					
5	+				
7	+				
19	+				
18	+				
6	+				
20	+				
1	+				
2	+				
3	+				
4	+				
8	+				
9	+				
10	+				
11	+				
문서보기					
문서ID	문서 작성일	토픽	비중값	Term Score	문서 제목
1000000242	2010-01-17 00:00:00	판결_공개_언론_방송_수구	0.253125	0.836104	노무현대통...

i. 관련 문서 보기 : 별도의 이동 없이 하단에 해당 토픽을 비중 있게 다른 문서들을 나열합니다.

(i-1) 문서 보기 : 우측으로 문서탭이 열리면서 해당 문서를 열람할 수 있습니다.

(i-2) 문서 ID : 상단에서 사용자가 선택한 토픽을 비중 있게 다른 문서의 아이디입니다.

(i-3) 문서 작성일 : 상단에서 사용자가 선택한 토픽을 비중 있게 다른 문서의 작성일입니다.

(i-4) 토픽 : 상단에서 사용자가 선택한 토픽을 비중 있게 다른 문서의 토픽입니다. 따라서 하단의 토픽 또한 상단의 토픽과 비슷해야 합니다.

(i-5) 비중값 : 상단에서 사용자가 선택한 토픽을 비중 있게 다른 문서에서, 해당 저자가 이 토픽을 얼마나 비중 있게 다루었는지 수치로 나타냅니다.

(i-6) Term Score : 비중값을 재가공하여 만든 지표로써 해당 단어, 문서, 사람 등이 토픽에 어느 정도 연관되어 있는지 나타냅니다. 이 지표는 비중값과 다르게 해당 토픽의 높은 순위에 위치한 단어, 저자, 문서를 다른 토픽에서는 어디에 위치해 있는지 판별하여 특징적으로 나타난 토픽의 사람, 문서, 단어 등을 지표상에서 높게 산출합니다. 즉, 비중값은 해당 토픽만을 염두해 두고 그 중요도를 수치화하는데 반해, Term Score는 해당 토픽에서의 순위만이 아니라 다른 토픽에서의 순위까지 생각하여 해당 토픽에서의 중요도를 판별합니다. 다시 말하면, 비중값이 1차 결과물이라면 Term Score 가 2차 결과물이라 할 수 있습니다.

(i-7) 문서 제목: 상단에서 사용자가 선택한 토픽을 비중 있게 다른 문서의 제목입니다.

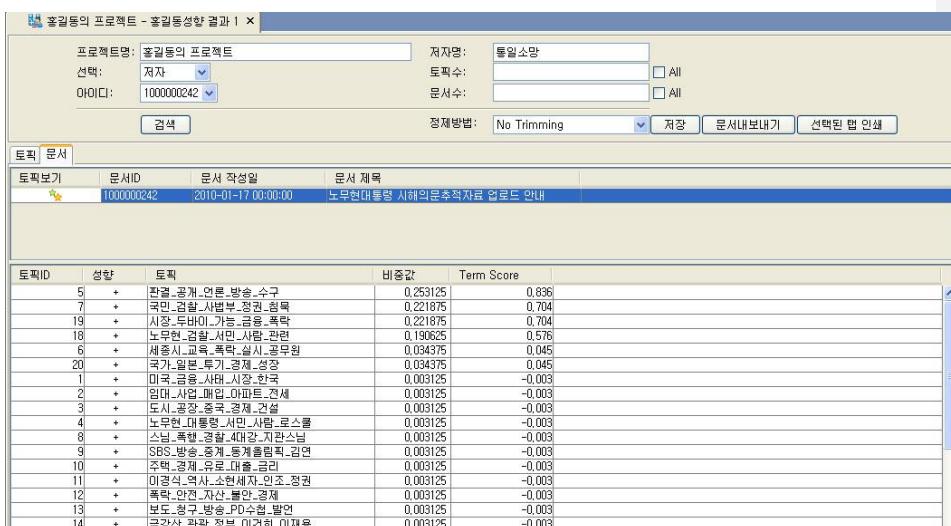
ii. 토픽 ID : 프로그램이 토픽에 부여한 숫자입니다.

iii. 토픽 : 총 20 개의 단어로 구성되었던 토픽들 중에서 비중이 큰 상위의 단어 4 개를 모아둔 것입니다.

iv. 비중값 : 해당 저자가 각 토픽을 얼마나 비중 있게 다루었는지 수치로 나타낸 것입니다. 수치의 범위는 0~1 까지이며 1 에 가까울수록 비중이 커집니다.

- v. Term Score : 비중값을 재가공하여 만든 지표로써 해당 단어, 문서, 사람 등이 토픽에 어느 정도 연관되어 있는지 나타냅니다. 이 지표는 비중값과 다르게 해당 토픽의 높은 순위에 위치한 단어, 저자, 문서를 다른 토픽에서는 어디에 위치해 있는지 판별하여 특징적으로 나타난 토픽의 사람, 문서, 단어 들을 지표상에서 높게 산출합니다. 즉, 비중값은 해당 토픽만을 염두해 두고 그 중요도를 수치화하는데 반해, Term Score는 해당 토픽에서의 순위만이 아니라 다른 토픽에서의 순위까지 생각하여 해당 토픽에서의 중요도를 판별합니다. 다시 말하면, 비중값이 1 차 결과물이라면 Term Score 가 2 차 결과물이라 할 수 있습니다.

(b-2) 문서



The screenshot shows a software interface for document analysis. At the top, there are search filters for 'Project Name: Hong Gil Dong's Project', 'Author: 투일소망', 'Topic ID: 1000000242', and 'Document ID: 1000000242'. Below these are buttons for '검색' (Search), '저장' (Save), '문서내보내기' (Export Document), and '선택된 템 인쇄' (Print Selected Item). The main window has tabs for 'Topic' and 'Document'. Under 'Topic', there is a table with columns: Topic ID, Category, Document ID, Document Creation Date, Document Title, and a large number of rows showing topic details. One row is highlighted in blue: Topic ID 1000000242, Category '+', Document ID 1000000242, Creation Date 2010-01-17 00:00:00, Title '노무현 대통령 시대의 문화 충무원'. Under 'Document', there is another table with columns: Document ID, Category, Topic ID, Topic Name, Term Score, and a large number of rows showing document details. One row is highlighted in blue: Document ID 1000000242, Category '+', Topic ID 5, Topic Name '판결_공개_언론_방송_수구', Term Score 0.936.

Topic ID	Category	Topic Name	Term Score
5	+	판결_공개_언론_방송_수구	0.936
7	+	국민_검찰_사법부_정치_침몰	0.221875
19	+	사장_누리마이_가는_금융_폭락	0.221875
18	+	노무현_검찰_시민_사람_권력	0.190625
6	+	세종시_교육_폭력_급식_충무원	0.094375
20	+	국가_일본_투기_경제_경쟁	0.094375
1	+	미국_금융_서비스_시장_한국	0.003125
2	+	임대_사업_매입_아파트_건설	0.003125
3	+	도시_공장_중국_경제_건설	0.003125
4	+	노무현_교통_서민_사람_로스쿨	0.003125
8	+	스님_폭행_경찰_대마강_지관스님	0.003125
9	+	SBS_방송_경제_동아일보_김연	0.003125
10	+	주택_경제_유로_대출_금리	0.003125
11	+	미공식_역사_소현세자_인조_정권	0.003125
12	+	폭락_안전_자산_불안_경제	0.003125
13	+	보도_경구_방송_PD수첩_발언	0.003125
14	+	금강산_관광_정보_이건희_이재용	0.003125

- i. 토픽보기 : 해당 문서에서 비중 있게 다룬 토픽들을 나열합니다.

- (i-1) 토픽 ID : 상단의 문서에서 비중 있게 다루어진 토픽의 아이디입니다.
- (i-2) 성향 : 사용자의 주관으로 형성한 성향 단어사전을 사용하여 토픽의 성향을 구분합니다. + 는 긍정 성향이며 -는 부정 성향입니다.
- (i-3) 토픽 : 상단의 문서에서 비중 있게 다루어진 토픽입니다.

(i-4) 비중값 : 상단의 문서에 대하여 토픽이 얼마나 비중 있게 다루어졌는지 수치로 나타냅니다.

(i-5) Term Score : 비중값을 재가공하여 만든 지표로써 해당 단어, 문서, 사람 등이 토픽에 어느 정도 연관되어 있는지 나타냅니다. 이 지표는 비중값과 다르게 해당 토픽의 높은 순위에 위치한 단어, 저자, 문서를 다른 토픽에서는 어디에 위치해 있는지 판별하여 특징적으로 나타난 토픽의 사람, 문서, 단어 등을 지표상에서 높게 산출합니다. 즉, 비중값은 해당 토픽만을 염두해 두고 그 중요도를 수치화하는데 반해, Term Score 는 해당 토픽에서의 순위만이 아니라 다른 토픽에서의 순위까지 생각하여 해당 토픽에서의 중요도를 판별합니다. 다시 말하면, 비중값이 1 차 결과물이라면 Term Score 가 2 차 결과물이라 할 수 있습니다.

ii. 문서 ID : 프로그램이 해당 문서에 임의로 부여한 숫자입니다.

iii. 문서 작성일: 문서가 생성된 일시입니다.

iv. 문서 제목: 문서의 제목입니다.

c) 토픽

저자는 ‘저자-이슈-성향 분석’을 실행한 후 생성되는 결과메뉴 중 하나로, 문서를 작성한 개개의 저자가 추출된 토픽들 중에서 어떤 토픽을 어떤 성향으로 가장 비중 있게 다루었는지 그리고 저자가 작성한 문서는 어떤 것들이 있는지를 열람할 수 있습니다. 초기 화면에는 비중값로 토픽이 나열되며 원하는 항목을 클릭하여 그 항목을 기준으로 토픽을 나열할 수 있습니다.

(방법 1) 선택란에서 토픽을 클릭한 후 하단에 토픽 아이디란이 생기면 사용자가 열람하고자 하는 토픽 아이디를 선택한 후 <검색>을 누릅니다. 혹은 (방법 2) 컴포넌트 트리의 ‘저자-이슈-성향 결과’에서 해당 결과문서에 마우스 오른쪽을 클릭한 후 ‘토픽’을 누릅니다.



우측의 선택란은 결과의 개수를 제한하는 역할을 합니다. 사용자에 필요에 따라 실행되므로 필수적으로 행해져야 되는 것은 아닙니다. 원하는 수치를 적거나 항목을 클릭한 후 확인 버튼을 클릭하면 결과의 수가 제한된 것을 확인할 수 있습니다.



- 토픽 명 : 해당 토픽 ID 에 대한 토픽명을 나타냄
- 토픽 수 : 결과 값으로 나타나는 토픽의 개수를 제한
- 문서 수 : 결과 값으로 나타나는 저자의 개수를 제한
- No Trimming : 처리를 실행하지 않음
- Probability Drop Trimming : 가장 큰 비중값으로 떨어지는 지점까지만 결과값을 표시
- Log Probability Drop Trimming : 비중값을 Log 로 치환한 후 가장 큰 값으로 떨어지는 지점까지만 표시
- Top Topic Matching Trimming : 검색된 단어, 저자, 문서로 해당 토픽을 가장 높은 비중값으로 들고 있을 확률
- 저장 : 현재 보이는 결과를 엑셀에서 열 수 있는 파일 형태로 저장합니다.
- 문서 내보내기: 현재 보여지는 문서들을 원래 데이터 형태대로 지정된 폴더에 저장한다.
- 선택된 템 인쇄 : 현재 보이는 결과를 인쇄합니다.

(c-1) 단어

단어	저자	문서	단어ID	단어	비중값	Term Score	긍정	부정
1350	미국		0,043096		0,279118	0,538545	0,461455	
1351	금융		0,029306		0,213077	0,662422	0,337578	
1255	사태		0,024135		0,160748	0,470114	0,529886	
1435	시장		0,022411		0,146568	0,541156	0,458844	
1085	한국		0,020687		0,137015	0,572681	0,427319	
742	불안		0,017240		0,119311	0,473269	0,526731	
262	부분		0,015516		0,088971	0,550610	0,449390	
1320	교육		0,015516		0,122377	0,917629	0,082371	
2038	성장		0,015516		0,109193	0,788939	0,211061	
483	미유		0,013792		0,081271	0,758714	0,241286	
1352	위기		0,013792		0,081903	0,393053	0,606947	
2126	두바이		0,013792		0,102474	0,460141	0,539859	
159	나라		0,012068		0,084495	0,514248	0,485752	
236	판단		0,012068		0,093385	0,522099	0,477901	
368	중국		0,012068		0,076981	0,364913	0,635087	
595	국가		0,012068		0,073932	0,498895	0,501105	

- i. 단어 ID: 프로그램이 단어에 임의로 부여한 숫자입니다.
- ii. 단어: 단어사전에 수집된 단어입니다.
- iii. 비중값: 단어사전에 수집된 단어들이 해당 토픽을 얼마나 비중 있게 다루었는지 수치로 나타낸 것을 말합니다. 수치의 범위는 0~1 까지이며 1에 가까울수록 비중이 커집니다.
- iv. Term Score : 비중값을 재가공하여 만든 지표로써 해당 단어, 문서, 사람 등이 토픽에 어느 정도 연관되어 있는지 나타냅니다. 이 지표는 비중값과 다르게 해당 토픽의 높은 순위에 위치한 단어, 저자, 문서를 다른 토픽에서는 어디에 위치해있는지 판별하여 특징적으로 나타난 토픽의 사람, 문서, 단어 등을 지표상에서 높게 산출합니다. 즉, 비중값은 해당 토픽만을 염두해 두고 그 중요도를 수치화하는데 반해, Term Score는 해당 토픽에서의 순위만이 아니라 다른 토픽에서의 순위까지 생각하여 해당 토픽에서의 중요도를 판별합니다. 다시 말하면, 비중값이 1 차 결과물이라면 Term Score 가 2 차 결과물이라 할 수 있습니다.
- v. 긍정 : 해당 단어가 문서 내에서 얼마만큼 긍정 성향을 가지고 표현되었는지 수치로 나타냅니다.
- vi. 부정 : 해당 단어가 문서 내에서 얼마만큼 부정 성향을 가지고 표현되었는지 수치로 나타냅니다.

(c-2) 저자

단어	저자	문서				
생성문서보기	저자ID	저자명	비중값	Term Score	긍정	부정
[*]	1003381	디에쉘	0.934483	4.973329	0.875000	0.125000
[*]	1002068	이-쁜한세상	0.420000	1.164409	0.571429	0.428571
[*]	1014353	출바른사람들	0.315385	0.934856	0.352941	0.647059
[*]	1002293	TwilightZone	0.261290	0.903352	0.625000	0.375000
[*]	1014115	호두까기	0.220000	0.467670	0.078431	0.921569
[*]	1009113	마리내1000년	0.183333	0.367745	0.294118	0.705882
[*]	1000266	슬픈한국	0.109639	0.391094	0.700855	0.299145
[*]	1003803	숨겨진질서	0.081579	0.105993	0.474359	0.525641
[*]	1000914	펄스꿈스	0.055000	0.087419	0.500000	0.500000
[*]	1000343	진실승리	0.054676	0.021001	0.457190	0.542810
[*]	1000016	무서운할배	0.050000	0.000000	0.041667	0.958333
[*]	1000000594	삼돌이	0.050000	0.000000	0.090909	0.909091
[*]	1000000791	진실승리	0.050000	0.000000	0.050000	0.950000
[*]	1005380	시멘트위민들레	0.050000	0.000000	0.050000	0.950000

- i. 생성 문서 보기 : 해당 저자가 쓴 문서(들)의 관련 정보들을 열람할 수 있습니다. 별도의 화면 이동 없이 작업창 하단에 실행됩니다. 이는 본 토픽과는 관련이 없습니다.

문서보기	문서ID	문서 작성일	문서 제목	토픽
[*]	1000000593	2010-01-29 00:00:00	マイペースとアメリカの危機 対決...	미국_금융_사회_시장_한국

- (i-1) 문서보기 : 우측으로 새 탭이 열리면서 해당 저자가 쓴 문서를 열람할 수 있습니다.
 (i-2) 문서 ID : 프로그램이 문서에 임의로 부여한 숫자입니다.
 (i-3) 문서 작성일 : 문서가 생성된 일시입니다.
 (i-4) 문서 제목: 문서의 제목입니다.
 (i-5) 토픽 : 해당 문서에서 가장 비중 있게 다뤄지는 토픽이 나타납니다.

- ii. 저자 ID: 프로그램이 해당 저자에 임의로 부여한 숫자입니다.
 iii. 저자: 저자명입니다.
 iv. 비중값: 저자들이 해당 토픽을 얼마나 비중 있게 다루었는지 수치로 나타낸 것을 말합니다. 수치의 범위는 0~1 까지이며 1에 가까울수록 비중이 커집니다.
 v. Term Score: 비중값을 재가공하여 만든 지표로써 해당 단어, 문서, 사람 등이 토픽에 어느 정도 연관되어 있는지 나타냅니다. 이 지표는 비중값과 다르게 해당 토픽의 높은 순위에 위치한 단어, 저자, 문서를 다른 토픽에서는 어디에 위치해있는지 판별하여 특징적으로 나타난 토픽의 사람, 문서, 단어 등을 지표상에서 높게 산출합니다. 즉, 비중값은 해당 토픽만을 염두해 두고 그 중요도를 수치화하는데 반해, Term Score 는 해당 토픽에서의 순위만이 아니라 다른 토픽에서의 순위까지 생각하여 해당 토픽에서의 중요도를 판별합니다. 다시 말하면, 비중값이 1 차 결과물이라면 Term Score 가 2 차 결과물이라 할 수 있습니다.

- vi. 긍정 : 해당 저자가 문서 내에서 얼마만큼 긍정 성향을 가지고 표현하였는지 수치로 나타냅니다.
 - vii. 부정 : 해당 저자가 문서 내에서 얼마만큼 부정 성향을 가지고 표현하였는지 수치로 나타냅니다.

(c-3) 문서

단어	제자	문서					
문서보기	자자문서보기	문서ID					
		문서작성일					
		비중값					
		Term Score					
		긍정					
		부정					
		저자명					
		문서제목					
1000000593	2010-01-29 00:00:00..	0.934483	4.973329	0.875000	0.125000	0.142000	아이패드와 미국의 위치 탐색...
1000000598	2010-01-29 00:00:00..	0.420000	1.164049	0.571429	0.428571	0.142000	이쁜 세상에 ■◆ 미국 달러의 부정... 다른 나라를 찾는길이 죄...
1000000440	2010-01-24 00:00:00..	0.315305	0.934866	0.642934	0.647059	0.125000	미네르바 vs 승수와 강남수
1000001045	2010-02-05 00:00:00..	0.261290	0.903352	0.625000	0.375000	0.142000	파닉, 금융시장 평균이 되는가 ■■■■■
1000000622	2010-02-05 00:00:00..	0.220000	0.467670	0.078431	0.921569	0.142000	유럽인은 자신을 물먹이다... ■◆ 유럽, 뉴욕 즐지 대학등 등 ... 주간 미수 경제 미...
1000001012	2010-02-05 00:00:00..	0.202500	0.641264	0.619048	0.380952	0.142000	산하이미...
1000000303	2010-02-05 00:00:00..	0.183333	0.367745	0.294118	0.705882	0.142000	월별에 따라가는 사람들의 계...
1000001059	2010-02-05 00:00:00..	0.109639	0.391094	0.700895	0.299145	0.142000	월별 경기 평판의 퍼마 가증-1/...
1000000995	2010-02-05 00:00:00..	0.083784	0.111310	0.461538	0.534563	0.142000	달리거나 끌거나 폭력적 폭도나...
1000001081	2010-02-05 00:00:00..	0.083784	0.107814	0.455698	0.544303	0.142000	달리거나 끌거나 폭력적 폭도나...
1000000586	2010-02-05 00:00:00..	0.081579	0.105993	0.474393	0.523641	0.142000	달리거나 끌거나 폭력적 폭도나...
1000000865	2010-02-05 00:00:00..	0.081579	0.106364	0.474393	0.526541	0.142000	달리거나 끌거나 폭력적 폭도나...
1000000501	2010-01-19 00:00:00..	0.050000	0.087419	0.500000	0.500000	0.142000	【풀무나】시즌 원활한 판매는...
1000000323	2010-01-19 00:00:00..	0.050000	0.000000	0.041667	0.958333	0.142000	경악!! 여행 즈음 사고 대처가...

- i. 문서 보기 : 우측으로 새 탭이 열리면서 해당 문서를 열람할 수 있습니다.
 - ii. 저자 문서 보기 : 해당 문서 이외에 작성한 저자의 다른 문서들을 열람할 수 있습니다.
별도의 화면 이동 없이 작업창 하단에 실행됩니다.

문서보기	문서ID	문서작성일	문서제목	토픽
	10585274	2009-09-02 00:00:00	[신증플루 확산 '비상'] 노인 무병비 노출 대책 마련 시급	신증플루...환자·감염·백신·예방
	10585275	2009-09-02 00:00:00	[신증플루 확산 '비상'] 환경위생 감소 혈액 수급 차질	지역...행사·경기·월드컵·도시
	10585276	2009-09-02 00:00:00	[신증플루 확산 '비상'] 1회 허동등 학생 7명 확진 판정	신증플루...환자·감염·백신·예방
	10585298	2009-09-02 00:00:00	[강릉] 신증플루 예방 2역 8천만원 지원	신증플루...환자·감염·백신·예방
	10585353	2009-09-02 00:00:00	[평창] 30만 인구 몰리다! 신증플루 대책 고심	신증플루...환자·감염·백신·예방
	10585360	2009-09-02 00:00:00	[횡성] 신증플루 방역 활동 강화	신증플루...환자·감염·백신·예방
	10599230	2009-09-01 00:00:00	[명성] 보건교사도 없는데 신증플루 대응?	신증플루...환자·감염·백신·예방
	10599231	2009-09-01 00:00:00	[정선] 정선 36개교 신증플루 24시간 비상방역체계 유지	신증플루...환자·감염·백신·예방
	10599792	2009-09-01 00:00:00	신증플루 확산방지 공직기강 확진자 7명	신증플루...환자·감염·백신·예방
	10599815	2009-09-01 00:00:00	신증플루 거점병원 의료진 잇단 감염	신증플루...환자·감염·백신·예방
	70004127300	2009-11-27 21:49:23	두바이→ 코크스 공장→ 전국 노동자→ 두바이	시장...경제·미국·위기·한국
	7000424525	2009-12-21 23:39:42	노조(자본)→ 민족·주조...미주조...도단...의원...차진	생각...원...봉독...사건...국민

- (ii-1) 문서 보기 : 우측으로 새 탭이 열리면서 해당 문서를 열람할 수 있습니다.
 - (ii-2) 문서 ID : 프로그램이 문서에 임의로 부여한 숫자입니다.
 - (ii-3) 문서 작성일 : 문서가 생성된 일시입니다.
 - (ii-4) 문서 제목 : 문서의 제목입니다.
 - (ii-5) 토픽 : 해당 문서에서 가장 비중 있게 다뤄진 토픽이 나타납니다.

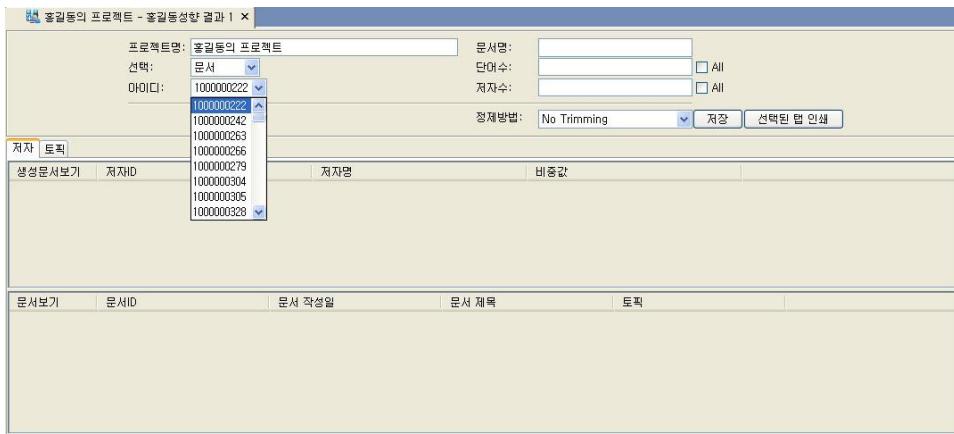
- iii. 문서 ID : 프로그램이 해당 문서에 임의로 부여한 숫자입니다.

- iv. 문서 작성일 : 문서가 생성된 일시입니다.
- v. 비중값 : 문서들이 해당토픽을 얼마나 비중 있게 다루었는지 수치로 나타낸 것을 말합니다. 수치의 범위는 0~1 까지이며 1에 가까울수록 비중이 커집니다.
- vi. Term Score : 비중값을 재가공하여 만든 지표로써 해당 단어, 문서, 사람 등이 토픽에 어느 정도 연관되어 있는지 나타냅니다. 이 지표는 비중값과 다르게 해당 토픽의 높은 순위에 위치한 단어, 저자, 문서를 다른 토픽에서는 어디에 위치해있는지 판별하여 특징적으로 나타난 토픽의 사람, 문서, 단어 등을 지표상에서 높게 산출합니다. 즉, 비중값은 해당 토픽만을 염두해 두고 그 중요도를 수치화하는데 반해, Term Score 는 해당 토픽에서의 순위만이 아니라 다른 토픽에서의 순위까지 생각하여 해당토픽에서의 중요도를 판별합니다. 다시 말하면, 비중값이 1 차 결과물이라면 Term Score 가 2 차 결과물이라 할 수 있습니다.
- vii. 긍정 : 해당 문서가 얼마만큼 긍정 성향을 가지고 표현하였는지 수치로 나타냅니다.
- viii. 부정 : 해당 저자가 문서 내에서 얼마만큼 긍정 성향을 가지고 표현하였는지 수치로 나타냅니다.
- ix. 저자명 : 저자의 이름입니다.
- x. 문서 제목 : 문서의 제목입니다.

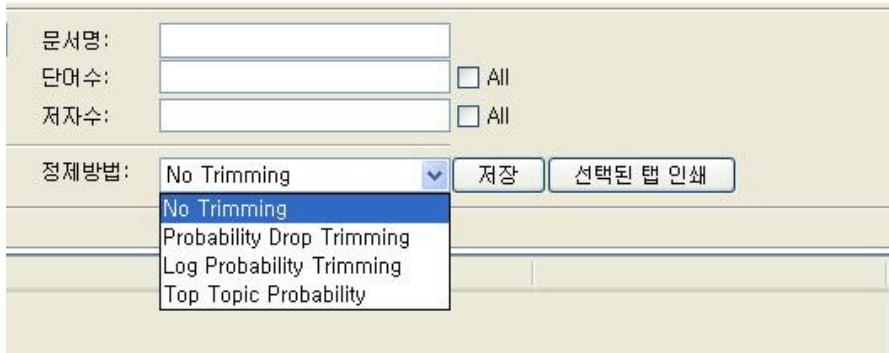
d) 문서

문서는 ‘저자-이슈-성향 분석’을 실행한 후 생성되는 결과메뉴 중 하나로, 각 문서의 저자 정보를 확인할 수 있고, 가장 비중 있게 다뤄진 토픽과 토픽의 성향을 열람할 수도 있습니다.

(방법 1) 선택란에서 문서를 클릭한 후 하단에 문서 아이디란이 생기면 사용자가 열람하고자 하는 문서 아이디를 선택한 후 <검색>을 누릅니다. 혹은 (방법 2) 컴포넌트 트리의 ‘저자-이슈-성향 결과’에서 해당 결과문서에 마우스 오른쪽을 클릭한 후 ‘도큐먼트’를 누릅니다.



좌측에 선택란은 결과의 개수를 제한하는 역할을 합니다. 사용자에 필요에 따라 실행되므로 필수적으로 행해져야 되는 것은 아닙니다. 원하는 수치를 적거나 항목을 클릭한 후 <확인>을 클릭하면 결과의 수가 제한된 것을 확인할 수 있습니다.



- 문서명 : 해당 문서 아이디에 대한 문서명을 나타냄
- 단어 수 : 결과 값으로 나타나는 단어의 개수를 제한
- 저자 수 : 결과 값으로 나타나는 저자의 개수를 제한
- No Trimming : 처리를 실행하지 않음
- Probability Drop Trimming : 가장 큰 비중값으로 떨어지는 지점까지만 결과값을 표시
- Log Probability Drop Trimming : 비중값을 Log 로 치환한 후 가장 큰 값으로 떨어지는 지점까지만 표시

- Top Topic Matching Trimming: 검색된 단어, 저자, 문서로 해당 토픽을 가장 높은 비중값으로 들고 있을 확률
- 저장 : 현재 보이는 결과를 엑셀에서 열 수 있는 파일 형태로 저장합니다.
- 선택된 템 인쇄 : 현재 보이는 결과를 인쇄합니다.

(d-1) 저자

저자	토픽
생성문서보기	저자ID 저자명 비중값 1000343 진실승리 1,000000

- i. 생성 문서 보기 : 해당 문서의 저자가 쓴 다른 문서들을 별도의 화면 이동 없이 하단에 나타납니다.

문서보기	문서ID	문서 작성일	문서 제목	토픽
	1000000222	2010-01-17 00:00:00	[특보]새총시수정안 반대여론 급상승!--"지방선거 영향 클것"	폭력_안전_자산_불안_경제

(i-1) 문서 보기: 우측으로 문서탭이 열리면서 해당 저자가 쓴 문서를 열람할 수 있습니다.

(i-2) 문서 ID : 프로그램이 문서에 임의로 부여한 숫자입니다.

(i-3) 문서 작성일 : 문서가 생성된 일시입니다.

(i-4) 문서 제목: 문서의 제목입니다.

(i-5) 토픽 : 해당 문서에서 가장 비중 있게 다뤄지는 토픽이 나타납니다.

ii. 저자 ID: 프로그램이 해당 저자에 임의로 부여한 숫자입니다.

iii. 저자명 : 해당 문서의 저자명입니다.

iv. 비중값: 저자가 해당 문서에서 얼마나 비중 있게 다루어졌는지 수치로 나타낸 것을 말합니다. 수치의 범위는 0~ 1 까지이며 1에 가까울수록 비중이 커집니다. 문서에 대한 저자는 단 1 명이기 때문에 값은 무조건 1이 나오게 됩니다.

(d-2) 토픽

저자	토픽	비중값	Term Score
관련문서보기	토픽ID	성향	토픽
11	+	이경식_역사_소현세자_인조_정권	0.275000 0.593479
12	+	폭락_안전_자산_불안_경제	0.275000 0.593479
1	+	미국_금융_사태_시장_한국	0.025000 -0.005995
2	+	임대_사업_매입_아파트_전세	0.025000 -0.005995
3	+	도시_공장_중국_경제_건설	0.025000 -0.005995
4	+	노무현_대통령_시민_사람_로스쿨	0.025000 -0.005995
5	+	판결_공개_언론_방송_수구	0.025000 -0.005995
6	+	세종시_교육_폭락_실시_공무원	0.025000 -0.005995
7	+	국민_검찰_사법부_정권_침묵	0.025000 -0.005995
8	+	스님_폭행_경찰_내대강_지관스님	0.025000 -0.005995
9	+	SBS_방송_증계_동계올림픽_김연	0.025000 -0.005995
10	+	주택_경제_유로_대출_금리	0.025000 -0.005995
13	+	보도_청구_방송_PD수첩_발언	0.025000 -0.005995
14	+	금강산_관광_정부_미건희_미재용	0.025000 -0.005995

- i. 관련 문서 보기 : 해당 토픽을 비중 있게 다룬 문서들을 별도의 화면 이동 없이 작업창 하단에 나열합니다.

문서보기	문서ID	문서 작성일	토픽	비중값	Term Score	문서 제목
	1000001041	2010-02-05 00:00:00	이경식_역사_소현세자_인조_정권	0.624194	4.572056	내가 밖온 정치의 더러움을 확인시켜주는 추노
	1000000650	2010-01-27 00:00:00	이경식_역사_소현세자_인조_정권	0.742105	3.286933	'추노'가 명품드라마 인 이유..
	1000000903	2010-02-05 00:00:00	이경식_역사_소현세자_인조_정권	0.350000	0.928377	통일에 만족하는 사람들에게...
	1000000222	2010-01-17 00:00:00	폭락_안전_자산_불안_경제	0.275000	0.593479	[특보] 새증시 수장인 반대여론 '상승승!!'—'지방선거 영향 끝'◆
	1000000998	2010-02-05 00:00:00	미국_금융_시태_시장_한국	0.220000	0.457670	◆ 미국 달러의 부활--- 다른 나라를 앞질리며 최선의 길 ◆
	1000000407	2010-01-21 00:00:00	국민_검찰_사법부_정권_침묵	0.110000	0.200222	전국에서 불난 국민들이 남기 목줄을 쳐기 시작!!!
	1000000328	2010-01-18 00:00:00	국가_일본_투기_경제_성장	0.050000	0.000000	경악! 일본 조사기 한 대포가 여론조작 현장 실토토!!!
	1000000350	2010-01-20 00:00:00	국가_일본_투기_경제_성장	0.050000	0.000000	★총국 속보★세종시 수장안의 전신!! 그리고 경운천의 자살률~~
	1000000415	2010-01-25 00:00:00	국가_일본_투기_경제_성장	0.050000	0.000000	「산영필사」# 사법부 특집「조장하던 [조선일보]가...
	1000000579	2010-02-05 00:00:00	국가_일본_투기_경제_성장	0.050000	0.000000	통일증집에서 평화증집으로 남북관계 폐리(廢立)를 전환해야
	1000000594	2010-01-29 00:00:00	국가_일본_투기_경제_성장	0.050000	0.000000	★ MBC 낭복점(浪福點)이 빛난 논란, 같은해 대법의 시의 표명
	1000000791	2010-02-04 00:00:00	국가_일본_투기_경제_성장	0.050000	0.000000	##[특보]기방선거 정권집권론, 계상도 빼고 전국적으로 절대우세!

(i-1) 문서보기 : 우측으로 새 탭이 열리면서 해당 저자가 쓴 문서를 열람할 수 있습니다.

(i-2) 문서 ID : 프로그램이 문서에 임의로 부여한 숫자입니다.

(i-3) 문서 생성일: 문서가 생성된 일시입니다.

(i-4) 토픽 : 해당 문서에서 가장 비중 있게 다뤄지는 토픽이 나타납니다.

(i-5) 비중값 : 모든 문서가 해당 토픽을 얼마나 비중 있게 다루었는지 수치로 나타낸 것을 말합니다. 수치의 범위는 0~1 까지이며 1에 가까울수록 비중이 커집니다.

(i-6) Term Score : 비중값을 재가공하여 만든 지표로써 해당 단어, 문서, 사람 등이 토픽에 어느 정도 연관되어 있는지 나타냅니다. 이 지표는 비중값과 다르게 해당 토픽의 높은 순위에 위치한 단어, 저자, 문서를 다른 토픽에서는 어디에 위치해있는지 판별하여 특징적으로 나타난 토픽의 사람, 문서, 단어 등을 지표상에서 높게 산출합니다. 즉, 비중값은 해당 토픽만을 염두해 두고 그 중요도를 수치화하는데 반해, Term Score 는 해당 토픽에서의 순위만이 아니라 다른 토픽에서의 순위까지 생각하여 해당 토픽에서의 중요도를 판별합니다. 다시 말하면, 비중값이 1 차 결과물이라면 Term Score 가 2 차 결과물이라 할 수 있습니다.

(i-7) 문서 제목 : 문서의 제목입니다.

- ii. 토픽 ID : 프로그램이 해당 토픽에 임의로 부여한 숫자입니다.
- iii. 토픽 : 해당 문서에서 가장 비중 있게 다뤄지는 토픽이 나타납니다.
- iv. 비중값 : 해당 문서가 각 토픽을 얼마나 비중 있게 다루어졌는지 수치로 나타낸 것을 말합니다. 수치의 범위는 0~1 까지이며 1에 가까울수록 비중이 커집니다.
- v. Term Score : 비중값을 재가공하여 만든 지표로써 해당 단어, 문서, 사람 등이 토픽에 어느 정도 연관되어 있는지 나타냅니다. 이 지표는 비중값과 다르게 해당 토픽의 높은 순위에 위치한 단어, 저자, 문서를 다른 토픽에서는 어디에 위치해있는지 판별하여 특징적으로 나타난 토픽의 사람, 문서, 단어 등을 지표상에서 높게 산출합니다. 즉, 비중값은 해당 토픽만을 염두해 두고 그 중요도를 수치화하는데 반해, Term Score 는 해당 토픽에서의 순위만이 아니라 다른 토픽에서의 순위까지 생각하여 해당 토픽에서의 중요도를 판별합니다. 다시 말하면, 비중값이 1 차 결과물이라면 Term Score 가 2 차 결과물이라 할 수 있습니다.

I. 저자-이슈-성향 분석 결과 이해하기

정해진 순서대로 저자-이슈-성향 분석을 실시하여 토픽과 토픽의 성향을 알아냈습니다. 이제부터 분석 결과를 해석하는 것은 사용자의 판단을 요구합니다. 이번 장에서는 생성된 토픽과 성향을 사용자가 어떻게 활용할 수 있는지 알아보겠습니다.

1. 전체 토픽

새 프로젝트를 생성할 시 사용자가 분석대상으로 지정했던 문서들이 어떤 화제에 관해서 그리고 어떤 성향으로 가장 비중 있게 다루어졌는지 순위별로 확인할 수 있습니다. 토픽의 수는 긍정 성향과 부정 성향이 동시에 나타나므로 ‘저자-이슈-성향 분석’에서 사용자가 지정하였던 토픽의 수의 2 배로 산출됩니다.

생성된 토픽에 대한 토픽 ID는 ‘저자-이슈 분석’과는 같이 문서에서 비중 있게 다루어진 화제의 순위별로 매겨졌기 때문에 1 번 토픽이 가장 비중값이 높다고 말할 수 있습니다. 또한 주의할 점은 연결된 20 개의 토픽들의 나열순서가 성향값을 기준으로 나열한 것이 아니라 성향값과 관계 없이 전체 문서 안에서 비중 있게 다루어진 단어 순서대로 나타난다는 것입니다.

전체 토픽을 클릭하였을 때 최초의 설정은 성향별로 비중값이 높은 순서대로 나타나며 사용자의 기호에 따라 해당항목을 클릭하여 나열 순서를 정렬할 수 있습니다.

생성된 토픽

4대강_사업_낙동강_문제_공사_토적_시작_불질_지역_국민_중단_조사_사진_필요_오염_준설_정부_전국_상황_모습

-14 번 토팩

부정성향을 가진 -14 번 토픽의 단어들의 조합이 ‘4 대강’, ‘사업’, ‘낙동강’, ‘문제’가 나온 것으로 보아 ‘4 대강 사업이 국민 여론과 자연환경에 끼치는 영향’이라고 추측할 수 있습니다. 사용자가 예상한 내용이 맞는지 확인하기 위해 **자세히 보기**-토팁 생성 – 문서로 이동합니다. 이 메뉴는 -3 번 토픽에 관하여 가장 비중 있게 다룬 문서들을 보여줍니다. 상위 10 위 안에 랭크되어 있는 문서들의 제목이 모두 4 대강 사업에 대한 국민 여론의 반발에 관련한 것으로 보아 위에서 추측하였던 내용이 맞았다는 것을 알 수 있습니다. 실제 문서를 확인하려면 <자세히보기>를 클릭합니다.

문서 제목
이명박이 잔인하게 박살낸...대한민국의 모습(사진)//-----
4대강 사업으로 사라지고 있는 우리나라 바경 13선 모음 자료
서울대교수 4대강 진실' 외면하는 죽중동 질타 파문!!!
전국에서 뿔난 국민들이 삽바기 목줄을 죄기 시작!!!
노무현은 예언자? 국토균형발전 주장 동영상 화제만발..
잠시후 11시 MBC PD수첩 필 시청바랍니다.
"생명의 땅, 꼭 자전거 길로 내줘야 하나" 편
★충격 속보★세종시 수정안의 진상!! 그리고 정운찬의 자살골~~
이명박))이러면 안되지..이건 **** 반칙 *** 입니다~~~~~!!
수구꼴통 언론들 아주 발악을 하는구만...

부정 3 번 토픽을 많이 다룬 문서의 제목

다시 화제를 바꾸어, -14 번 토픽의 단어들을 살펴보면 ‘필요’, ‘사진’ 같은 단어들은 방금 확인한 4 대강 사업에 대한 국민 여론의 반발과는 100% 의 연관성을 보여주지 못합니다. 따라서 사용자가 여타 세부 메뉴를 클릭하여 화제를 정확히 파악할 수 있어야 합니다.

2. 토픽

전체 토픽에서 다루었던 내용을 토대로 -14 번 토픽이 4대강 사업에 대한 국민 여론의 반발을 화제로 하였다는 것을 알아냈습니다. 이제부터는 -14 번 화제를 비중 있게 다른 저자들을 알아보고자 합니다. 이는 해당 화제에 관하여 어떤 기관이나 사람이 어떤 성향으로 다루고 있는지 살펴볼 때 유용합니다.

홍길동의 프로젝트 - 홍길동성향 결과 1 ×							
프로젝트명:	홍길동의 프로젝트	토픽명:	4대강_사업_낙동강_문제_공사				
선택:	선택: 토픽	선택수:	<input type="checkbox"/> All				
아이디:	-14	문서수:	<input type="checkbox"/> All				
검색		정제방법:	No Trimming	<input type="checkbox"/> 저장	<input type="checkbox"/> 문서내보내기	<input type="checkbox"/> 선택된 항목만	
단어 저자 문서							
문서보기	저자문서보기	문서ID	문서 작성일	비중값	Term Score	긍정	부정
		1000001068	2010-02-05 00:00:00	0.753571	3,445596	0.068966	0.931034
		1000000902	2010-01-19 00:00:00	0.379167	1,291106	0.206897	0.793103
		1000000951	2010-02-05 00:00:00	0.374074	1,502667	0.071429	0.928571
		1000000407	2010-01-21 00:00:00	0.368182	1,292972	0.300000	0.700000
		1000001042	2010-02-05 00:00:00	0.208930	0.796905	0.062500	0.937500
		1000000350	2010-01-20 00:00:00	0.122222	0.224178	0.111111	0.888889
		1000000604	2010-01-26 00:00:00	0.116667	0.261658	0.964103	0.435097
							FreeDark
							"성향의 땅, 즉 자연가 절로 내줘야 하나" 품

-14 번 토픽을 클릭하였을 때의 결과

토픽—아이디 -14— 저자를 클릭합니다. 이 메뉴는 -14 번 토픽을 비중 있게 다른 저자들을 보여줍니다.
 1 순위부터 6 순위까지 랭크되어 있는 저자들은 파란푸른, 돌아온단합, 주먹이운다, 노루귀, 산음, 육스라이브 순입니다. 아래의 그림에서 볼 수 있듯이 긍정 성향값과 부정 성향값을 비교하면 부정 성향값이 더 큽니다. 이를 통해 상기 저자들은 -14 번 토픽에 관하여 부정적인 성향을 띠고 있다는 것을 알 수 있습니다.

단어	저자	문서	생성문서보기	저자ID	저자명	비중값	Term Score	긍정	부정
				1014301	파란푸른	0,753571	3,445596	0,068966	0,931034
				1000000333	돌아온단합	0,433333	1,608749	0,428571	0,571429
				1003144	주먹이운다	0,379167	1,291106	0,206897	0,793103
				1000169	노루귀	0,368182	1,292972	0,300000	0,700000
				1004020	산음	0,209890	0,796905	0,062500	0,937500
				1000000913	육스라이브	0,183333	0,351691	0,285714	0,714286

-14 번 토픽에 대한 저자 정보

3. 저자

토익에서 다루었던 내용을 토대로 파란푸른, 돌아온단합, 주먹이운다, 노루귀, 산음, 육스라이브 등의 저자가 4대강 관련 토픽에 대해 부정적인 성향을 나타낸다는 것을 알아냈습니다. 이제부터는 전체 문서의 저자들 중 임의의 저자를 선택하여 그가 어떤 토픽을 어떤 성향을 가지고 가장 비중 있게 다루었으며 그가 쓴 문서는 어떤 것이 있는지 살펴보도록 하겠습니다. 이는 특정 인물이 어떤 화제를 주로 다루며 어떤 성향의 글을 많이 썼는지 살펴볼 때 유용합니다.

저자 — 아이디 — 1014364 — 토픽 을 클릭합니다. 이 메뉴는 해당 저자가 어떤 토픽을 어떤 성향으로 가장 비중 있게 다루었는지 보여줍니다. 긍정 성향 토픽에서 1순위와 2순위의 비중값의 차이가 크게 나는 것을 감안할 때 1순위인 +15 번 토픽에 관해 비중 있게 다룬 것을 알 수 있습니다. 부정 성향 토픽에서는 1순위와 2순위의 비중값의 차이가 크게 나지 않으므로 -1 번과 -9 번 토픽에 관해 비중 있게 다룬 것을 알 수 있습니다.

홍길동의 프로젝트 - 홍길동성향 결과 1 ×									
프로젝트명: 홍길동의 프로젝트			문서명:	이명박이 잔인하게 박살난					
선택:	문서	단어수:							
아이디:	1000001068	저자수:							
<input type="button" value="검색"/>						정제방법: No Trimming			
저자		토픽							
관련문서보기	토픽ID	성향	토픽	비중값	Term Score				
<input type="checkbox"/>	17	+	자연_균형_발전_대상_마바타	0.366667	0.835267				
<input type="checkbox"/>	1	+	미국_금융_사태_시장_한국	0.033333	-0.003996				
<input type="checkbox"/>	2	+	임대_사업_매입_아파트_전세	0.033333	-0.003996				
<input type="checkbox"/>	3	+	도시_공장_중국_경제_건설	0.033333	-0.003996				
<input type="checkbox"/>	4	+	노무현_대통령_서민_사람_로스쿨	0.033333	-0.003996				
<input type="checkbox"/>	5	+	판결_공개_언론_방송_수구	0.033333	-0.003996				
<input type="checkbox"/>	6	+	세종시_교육_폭락_실시_공무원	0.033333	-0.003996				
<input type="checkbox"/>	7	+	국민_검찰_사법부_정권_첨목	0.033333	-0.003996				
<input type="checkbox"/>	8	+	스님_폭행_경찰_4대강_지관스님	0.033333	-0.003996				
<input type="checkbox"/>	9	+	SBS_방송_중계_동계올림픽_김연	0.033333	-0.003996				
<input type="checkbox"/>	10	+	주택_경제_유로_대출_금리	0.033333	-0.003996				
<input type="checkbox"/>	11	+	미경식_역사_소현세자_민조_정권	0.033333	-0.003996				
<input type="checkbox"/>	12	+	폭락_안전_자산_불안_경제	0.033333	-0.003996				
<input type="checkbox"/>	13	+	보도_청구_방송_PD수첩_발언	0.033333	-0.003996				
<input type="checkbox"/>	14	+	금강산_관광_정부_이건희_이재용	0.033333	-0.003996				
<input type="checkbox"/>	15	+	도요타_리콜_시장_생산_일본	0.033333	-0.003996				
<input type="checkbox"/>	16	+	부동산_인가_티시먼_미국_투자	0.033333	-0.003996				
<input type="checkbox"/>	18	+	노무현_검찰_서민_사람_관련	0.033333	-0.003996				
<input type="checkbox"/>	19	+	시장_두바이_가능_금융_폭락	0.033333	-0.003996				
<input type="checkbox"/>	20	+	국가_일본_투기_경제_성장	0.033333	-0.003996				
<input type="checkbox"/>	14	-	4대강_사업_낙동강_문제_공사	0.753571	3.445596				
<input type="checkbox"/>	19	-	4대강_사업_스님_정부_예산	0.075000	0.169876				
<input type="checkbox"/>	3	-	소프트웨어_정부_이명박_산업_경제	0.039286	0.063580				
<input type="checkbox"/>	13	-	좌파_그리스_노조_경제_재정	0.039286	0.063580				

이를 쉽게 확인하려면 좌측 선택란에서 ‘Probability Drop Trimming’을 클릭하면 가장 큰 값으로 다루어진 토픽을 한번에 알 수 있습니다.

The screenshot shows the search interface for the Hong Gil-dong Project. The search parameters are set to '작자' (Author) and '1014364'. The '정제방법' (Filter Method) dropdown is set to 'Probability Drop Trimming'. A red circle highlights the 'Click' button next to this dropdown. Below the search bar is a table titled '토픽' (Topic) showing three topics with their IDs, polarities (+ or -), terms, and term scores. The first topic, ID 15, has a high term score of 5.167137.

관련문서보기	토픽ID	성향	토픽	비중값	Term Score
	15	+	도요타_리콜_시장_생산_일본	0.930952	5.167137
	1	-	미국_총리_대비_GDP_지역	0.420000	1.164409
	9	-	스님_시장_시간_환율_의혹	0.220000	0.467670

저자 – 아이디 – 100343 – 문서를 클릭합니다. 이 메뉴는 해당 저자가 어떤 문서들을 가장 비중 있게 다루었는지 보여줍니다. 문서가 1개 나온 것을 감안할 때 이 저자가 쓴 글은 단 1개뿐이며 그 문서가 +11 번 토픽과 관련 있는 글이라는 것을 확인할 수 있습니다.

The screenshot shows the search interface for the Hong Gil-dong Project. The search parameters are set to '저자' (Author) and '100343'. A red circle highlights the 'Click' button next to the author ID. Below the search bar is a table titled '문서' (Document) showing one document entry. The document ID is 1000000222, it was written on 2010-01-17 at 00:00:00, and its subject is '[특보] 새 종시수정안 반대여론 급상승!!--"지방선거 영향 클것"'.

문서보기	문서ID	문서 작성일	문서 제목
★	1000000222	2010-01-17 00:00:00	[특보] 새 종시수정안 반대여론 급상승!!--"지방선거 영향 클것"

Below the document table is another table titled '토픽' (Topic) showing three topics related to the document. The first topic, ID 11, has a high term score of 0.593.

토픽ID	성향	토픽	비중값	Term Score
11	+	미경식_역사_소현세자_인조_정권	0.275000	0.593
12	+	폭락_안전_자산_불안_경제	0.275000	0.593
15	-	여론_조사_세종시_수정_반대	0.709756	3.394

4. 문서

임의의 문서에 대한 저자와 토픽에 관해 알아보겠습니다. 이는 해당 문서를 중심으로 이 글을 쓴 저자가 누구인지 그리고 이 글의 화제는 무엇인지 알아 보고자 할 때 유용합니다.

문서 — 아이디 — 1000001068 — 저자를 클릭합니다. 이 메뉴는 해당 문서의 저자를 보여줍니다. 일반적으로 문서에 대한 저자는 단 1명이기 때문에 목록에는 1명의 저자가 뜨게 되며 비중값도 항상 1이 나오게 됩니다.



문서 — 아이디 — 1000001068 — 토픽 을 클릭합니다. 이 메뉴는 해당 문서에 관하여 가장 비중 있게 다뤄진 토픽과 그것의 성향을 보여줍니다. 긍정 성향에서 1순위와 2순위의 비중값의 차이가 크게 나는 것으로 보아 1순위인 17 번 토픽에 관해 가장 비중 있게 다룬 것을 알 수 있습니다. 부정 성향에서도 1순위와 2순위의 비중값의 차이가 크게 나는 것으로 보아 1순위인 -14 번 토픽에 관해 가장 비중 있게 다룬 것을 알 수 있습니다. 이를 쉽게 확인하려면 우측 선택란에서 'Probability Drop Trimming'을 클릭하면 가장 큰 값으로 다루어진 토픽을 한번에 알 수 있습니다.

총길동의 프로젝트 - 총길동성향 결과 1 ×						
프로젝트명:	총길동의 프로젝트	문서명:	이명박이 잔인하게 박살난			
선택:	문서	단어수:				
아이디:	1000001068	저자수:				
<input type="button" value="검색"/>						정제방법: No Trimming
저자 토픽						
관련문서보기	토픽ID	성향	토픽	비중값	Term Score	
<input type="checkbox"/>	17	+	자연_균형_발전_대상_아바타	0.366667	0.835267	
<input type="checkbox"/>	1	+	미국_금융_사태_시장_한국	0.033333	-0.003996	
<input type="checkbox"/>	2	+	임대_사업_매입_아파트_전세	0.033333	-0.003996	
<input type="checkbox"/>	3	+	도시_공장_중국_경제_건설	0.033333	-0.003996	
<input type="checkbox"/>	4	+	노무현_대통령_서민_사람_로스쿨	0.033333	-0.003996	
<input type="checkbox"/>	5	+	판결_공개_언론_방송_수구	0.033333	-0.003996	
<input type="checkbox"/>	6	+	세종시_교육_폭락_실시_공무원	0.033333	-0.003996	
<input type="checkbox"/>	7	+	국민_경찰_사법부_정권_침묵	0.033333	-0.003996	
<input type="checkbox"/>	8	+	스님_폭행_경찰_대강_지관스님	0.033333	-0.003996	
<input type="checkbox"/>	9	+	SBS_방송_중계_동계올림픽_김연	0.033333	-0.003996	
<input type="checkbox"/>	10	+	주택_경제_유로_대출_금리	0.033333	-0.003996	
<input type="checkbox"/>	11	+	이경식_역사_소현세자_인조_정권	0.033333	-0.003996	
<input type="checkbox"/>	12	+	폭락_안전_자산_불안_경제	0.033333	-0.003996	
<input type="checkbox"/>	13	+	보도_청구_방송_PD수첩_발언	0.033333	-0.003996	
<input type="checkbox"/>	14	+	금강산_관광_정부_이건희_이재용	0.033333	-0.003996	
<input type="checkbox"/>	15	+	도요타_리콜_시장_생산_일본	0.033333	-0.003996	
<input type="checkbox"/>	16	+	부동산_인가_티시먼_미국_투자	0.033333	-0.003996	
<input type="checkbox"/>	18	+	노무현_경찰_서민_사람_관련	0.033333	-0.003996	
<input type="checkbox"/>	19	+	시장_두바이_가능_금융_폭락	0.033333	-0.003996	
<input type="checkbox"/>	20	+	국가_일본_투기_경제_성장	0.033333	-0.003996	
<input type="checkbox"/>	14	-	4대강_사업_낙동강_문제_공사	0.753571	3.445596	
<input type="checkbox"/>	19	-	4대강_사업_스님_정부_예산	0.075000	0.169876	
<input type="checkbox"/>	3	-	소프트웨어_정부_이명박_산업_경제	0.039286	0.063580	
<input type="checkbox"/>	13	-	좌파_그리스_노조_경제_재정	0.039286	0.063580	

여기에서 +17 번 토픽에 관한 <관련문서보기>를 클릭하면 ‘문서 아이디 1000001068’에 대한 문서들이 나열되는 것이 아니라 +17 번 토픽을 비중 있게 다른 문서들이 하단에 나열됩니다. 이것은 **토픽—아이디—17—문서**를 선택해도 같은 결과가 나타납니다.

문서보기	문서ID	문서 작성일	토픽	비중값	Term Score	문서 제목
<input type="checkbox"/>	1000000304	2010-01-17 00:00:00	자연_균형_발전_대상_아바타	0.623077	2.431624	노무현은 예언자? 국토균형발전 주장 동영상 화제만발..
<input type="checkbox"/>	1000000683	2010-01-19 00:00:00	자연_균형_발전_대상_아바타	0.611688	3.105742	제임스 카메론 감독의 <아바타>가 MB에게 전하는 메시지~!
<input type="checkbox"/>	1000001068	2010-02-05 00:00:00	자연_균형_발전_대상_아바타	0.366667	0.835267	이명박이 잔인하게 박살낸.. 대한민국의 모습(사진)//-----
<input type="checkbox"/>	1000000263	2010-01-17 00:00:00	자연_균형_발전_대상_아바타	0.262500	0.679309	노무현생각 - 국토균형발전 ⑧

J. 포스트 참조관계망 생성하기

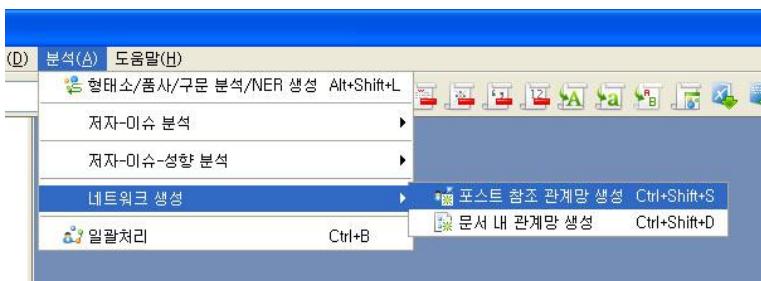
생성된 토픽에 대해 사용자의 판단까지 이루어지게 되면 문서의 메타 정보를 파악할 수 있습니다. 메타 정보는 저자, 주제(이슈), 저자의 영향력, 문서가 참조된 관계(링크된 문서)입니다. 이전 과정에서 실행되었던 분석들이 문서 내에서 주로 다루어진 내용을 파악하는 일이었다면 이번 장부터는 생성된 결과를 기반으로 한 메타 정보를 파악합니다. 다음은 ‘포스트 참조관계망 생성’의 자세한 내용입니다.

1. 포스트 참조관계망 생성

‘포스트 참조관계망 생성’은 저자가 작성한 문서가 전체 문서 내에서 얼마만큼의 영향력을 나타냈는지 산출해냅니다. 문서의 조회 수, 찬성 수, 반대 수, 스크랩 수, 댓글 수를 기준으로 사용자가 별도의 비중 값을 설정하여 문서의 영향력을 측정합니다. 메타 네트워크가 생성되면 ORA 를 이용하여 결과를 확인합니다.

‘포스트 참조관계망 생성’은 토픽을 산출한 결과를 대상으로 하기 때문에 반드시 <III.G. 저자-이슈 분석 결과 이해하기, page 160> 과 <III.H. 저자-이슈-성향 분석 결과 이해하기, page 191>까지 마친 상태에서 실행되어야 하며 특정 저자의 영향력을 측정할 때 유용합니다. 다음은 ‘포스트 참조관계망 생성’을 실행하는 과정입니다.

- 포스트 사회관계망 생성을 실행하려면 **메뉴막대 – 분석 – 네트워크 생성 – 포스트 참조 관계망 생성**을 누릅니다.



- b) ‘포스트참조관계망생성기’ 대화상자가 나타나면 ‘메타넷 파일명’에 원하는 이름을 기재한 후 ‘저자-이슈 분석’ 혹은 ‘저자-이슈-성향 분석’ 결과에서 적용대상이 되는 결과 문서를 클릭합니다.



사용자가 각 항목에 가중치¹의 수치를 지정합니다.

상기의 예시에는 뷰 비중, 긍정 비중, 스크랩 비중의 수치에 가중치 0.5 를, 반대에는 -0.5 를, 코멘트 비중에는 1 을 두었습니다. 이는 긍정적으로 반응한 요소에 가중치를 크게 두어 긍정적인 영향력을 가진 저자들을 순서대로 나열하고자 할 때 활용할 수 있습니다. 반대로 부정 비중에 1 을 두고 나머지 요소에マイ너스 수치를 두면 부정적으로 반응한 요소에 가중치를 두어 부정적인 영향력을 가진 저자들이 순서대로 산출됩니다.

(b-1) 뷰 비중 : 조회 수

(b-2) 긍정 비중: 찬성 수

(b-3) 부정 비중: 반대 수

¹ 가중치란 평균치를 산출할 때 개별치에 부여되는 중요도입니다. ‘포스트 참고 관계망 생성기’에서 사용자가 별도의 가중치를 들 수 있습니다.

(b-4)스크랩 비중: 스크랩 수

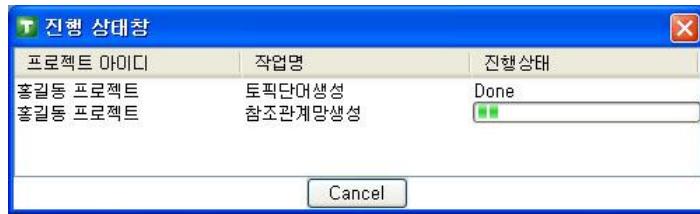
(b-5) 코멘트 비중: 댓글 수

(b-6) 비중값 Threshold : 네트워크를 생성할 때 특정 비중값 이하의 저자 토픽 관계는 생략하도록 한다.

(b-7) 링크 Threshold : 네트워크를 생성할 때 특정 링크값 이하의 저자와 저자와의 관계는 생략한다.

‘포스트참조관계망생성기’의 권장수치는 뷰 비중, 긍정 비중, 스크랩 비중은 0.5, 부정 비중은 -0.5, 코멘트 비중값은 1 입니다. 설정이 끝나면 <확인>을 클릭합니다.

c) ‘진행 상태창’ 대화상자가 나타나면 연두색 막대가 채워질 때까지 기다립니다. 이때 <Cancel>을 클릭하면 ‘포스트 사회관계망 생성’이 중단됩니다.



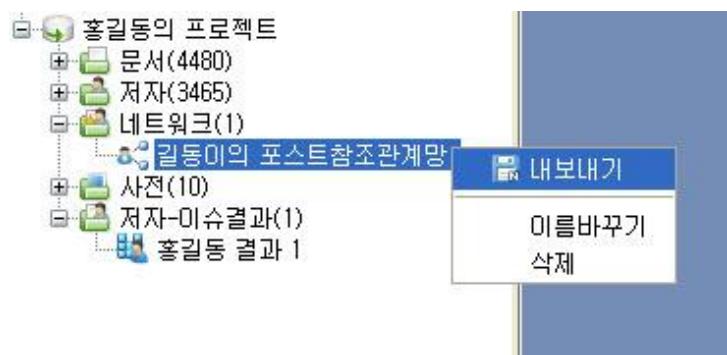
d) 명령이 성공적으로 수행되었다면 속성에는 ***** Successfully executed ***** 가 표기됩니다.

2. 포스트 참조관계망 처리 결과 보기

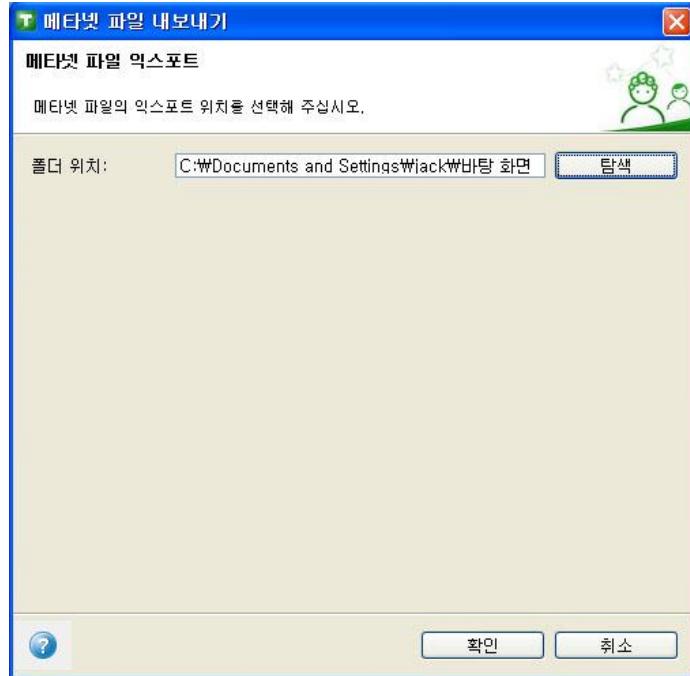
'포스트 참조관계망 생성'을 실행하였을 때의 결과는 ORA에서 확인합니다. 다음은 '포스트 참조관계망 생성'의 결과를 확인하는 과정입니다.

a) TONK에서 파일 내보내기

(a-1) 컴포넌트 트리의 '네트워크'에서 생성된 포스트 참조관계망 문서에 마우스 오른쪽을 누른 후 '내보내기'를 클릭합니다.



(a-2) '메타넷 파일 내보내기' 대화상자가 나타나면 <탐색>을 클릭하여 저장할 장소를 지정한 후 <확인>을 누릅니다.



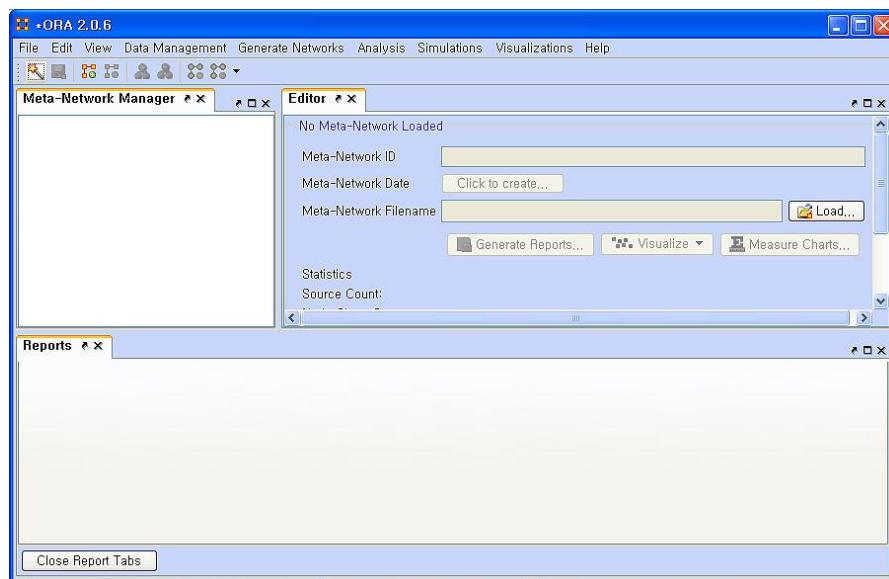
지정한 위치로 포스트 참고관계망 문서(메타 문서)가 생성됩니다.

b) ORA에서 파일 불러오기

(b-1) 바탕화면에 있는 ORA 아이콘을 더블 클릭하여 ORA를 실행합니다.

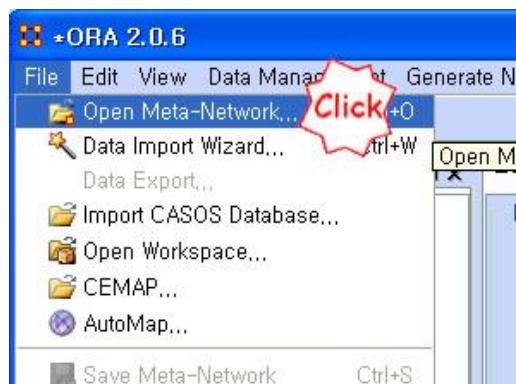


ORA 아이콘을 더블 클릭

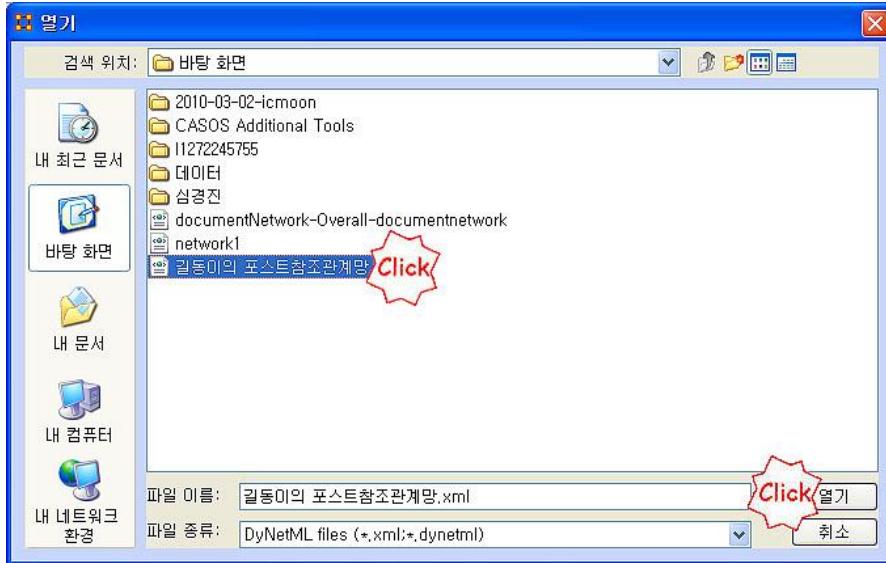


ORA 를 실행하였을 때 초기화면

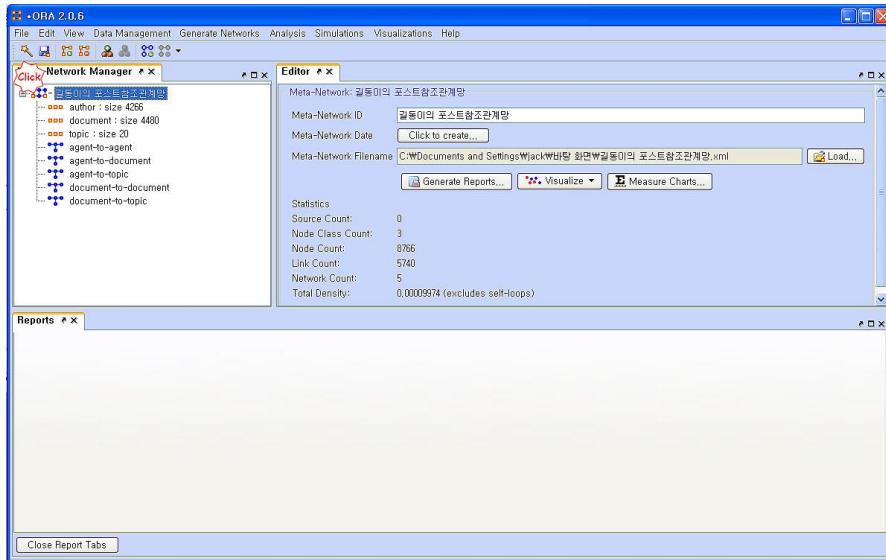
(b-2) ORA 가 실행되면 **File — Open Meta Network** 을 클릭합니다.



(b-3) ‘열기’ 대화상자가 나타나면 익스포트를 실행시킨 문서를 클릭한 후 <확인>을 누릅니다.



(b-4) Mata-Network Manager 에 TONK 에서 생성했던 저자-이슈-문서 네트워크가 나타납니다. + 를 클릭하면 분석 문서의 하위메뉴인 저자, 문서, 토픽 등이 나타납니다.

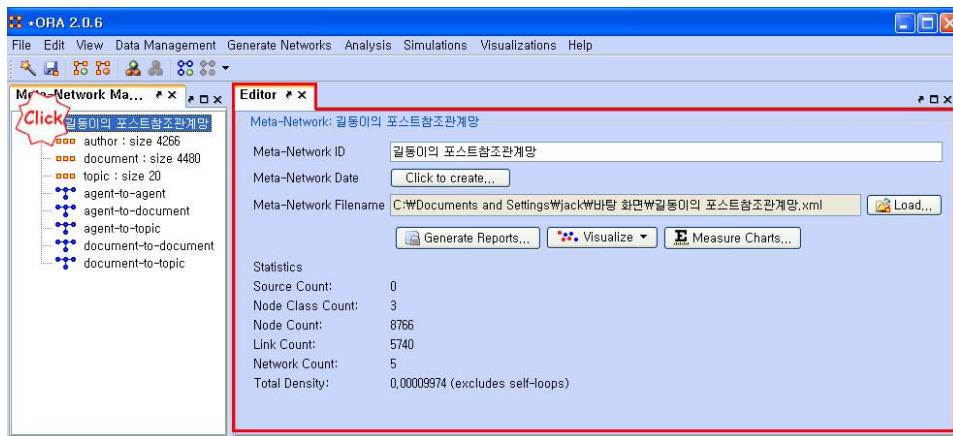


c) ORA 에서 결과 열람하기

Mata-Network Manager 에 나타나는 저자-이슈-문서 네트워크의 하위 메뉴는 총 8 가지로, **분석 대상인 개별 개체 항목을 일컫는 Node Set 과 이들간의 연결관계를 나타내는 Network**로 이루어져 있습니다. Node Set 에는 author, document, topic 이 있으며 Network 에는 agent-to-agent, agent-to-document, agent-to-topic, document-to-document, document-to-topic 이 있습니다. 저자-이슈-문서 네트워크의 자세한 내용은 다음과 같습니다.

(c-1) 저자-이슈-문서 네트워크

TONK 의 **분석—포스트 참조 관계망 생성 — 포스트 참조 관계망 생성기-메타넷 파일명**에서 지정했던 이름과 같습니다. Mata-Network Manager 에 생성된 메타 네트워크를 클릭하면 우측 Editor 란에 관련정보가 나타납니다.

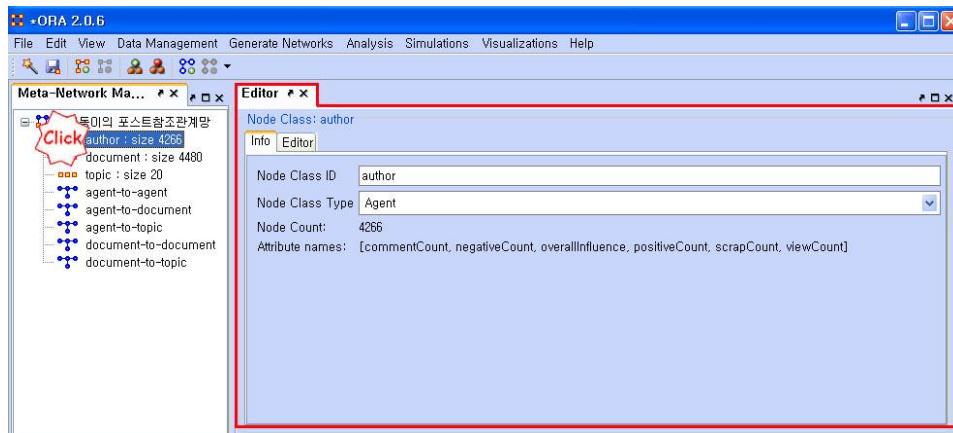


- i. Meta-Network ID : 메타 네트워크의 이름을 나타내며 이름을 변경할 수 있습니다.
- ii. Meta-Network Date : TONK 에서 메타 네트워크를 만들었던 날짜입니다.
- iii. Meta-Network Filename : 메타 네트워크가 저장된 장소입니다.
- iv. Source Count : 문서 출처의 개수입니다.
- v. Node Class Count : 메타 네트워크에 존재하는 총 Node 의 수입니다.
- vi. Link Count : 메타 네트워크에 존재하는 총 연결 관계의 수입니다.
- vii. Network Count : 메타 네트워크에 존재하는 총 네트워크의 수입니다.
- ii. Total Density : 네트워크의 조밀도입니다.

(c-2) Author

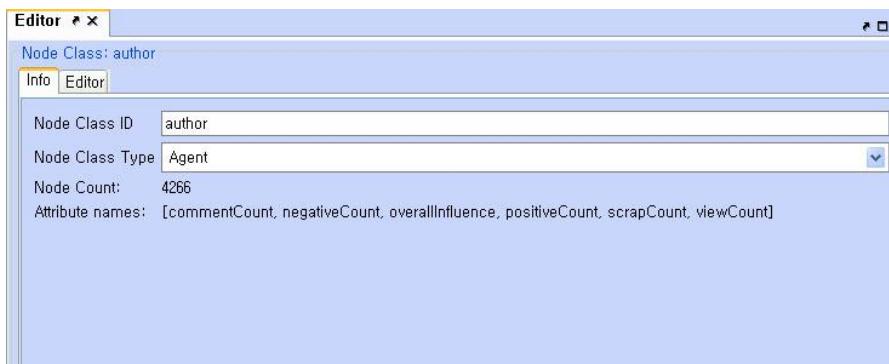
저자-이슈-문서 네트워크의 하위 메뉴로 나타나는 저자는 TONK에서 메타 네트워크가 산출되어진 프로젝트의 저자입니다. size는 저자의 수를 나타내며 문서에 댓글을 남긴 저자의 수도 포함합니다. TONK는 문서를 작성한 저자의 수만 인식하기 때문에 ORA에서 보여지는 수치보다 더 작을 수 있습니다.

Meta-Network Manager의 저자를 클릭하면 우측 Editor란에 관련정보가 나타나며 Info와 Editor 탭으로 구성되어 있습니다.



i. Info

선택된 Node의 기본 정보를 나타냅니다.



(i-1) Node Class ID : 선택된 Node 의 아이디를 나타내며 아이디를 변경할 수 있습니다.

(i-2) Node Class Type : 저자가 가지고 있는 온톨로지입니다.

(i-3) Node Count : 저자의 수입니다.

(i-4) Attributes names : 저자의 특성치 목록입니다.

ii. Editor

생성된 메타 네트워크의 결과를 수치로 확인합니다.

The screenshot shows a software interface titled "Node Class: author". It has tabs for "Info" and "Editor", with "Editor" selected. A search bar labeled "Find:" is present. The main area is a table with columns: Node ID, Node Title, commentCount, negativeCount, overallInfluence, positiveCount, scrapCount, and viewCount. The "overallInfluence" column is highlighted with a red border. On the right side, there's a sidebar with buttons for "Nodes" (Create, Move, Delete, Merge, Clean), "Attributes" (Create, Delete, Import), and other controls. At the bottom, there are buttons for "Select All" and "Clear All", and a status message: "0 item(s) selected, 4266 visible, 4266 total."

Node ID	Node Title	commentCount	negativeCount	overallInfluence	positiveCount	scrapCount	viewCount
0	염병이	0.0	0.0	0.0	0.0	0.0	11.0
760	Kramer	2.0	0.0	2.0	0.0	0.0	2.0
783	시선유도	14.0	0.0	14.0	0.0	0.0	14.0
799	하와유	9.0	0.0	9.0	0.0	0.0	9.0
800	o777을할기0o	16.0	0.0	16.0	0.0	0.0	0.0
801	나나그리고...	7.0	0.0	7.0	0.0	0.0	0.0
802	통일소망	12.0	0.0	12.0	0.0	0.0	12.0
810	민중불교2ccr	11.0	0.0	11.0	0.0	0.0	11.0
813	Choi_Archi	0.0	0.0	0.0	0.0	0.0	352.0
821	꽃병기술자	16.0	0.0	16.0	0.0	0.0	16.0
827	o777을할기0o	16.0	0.0	16.0	0.0	0.0	0.0
829	주니	14.0	0.0	14.0	0.0	0.0	14.0

(ii-1) Find : 사용자가 원하는 특정 Node Id 및 Node Title 을 검색할 수 있습니다.

(ii-2) Node ID : 문서 수집 시 저자에 부여한 숫자입니다.

(ii-3) Node Title : 저자의 이름입니다.

(ii-4) comment Count : 저자가 쓴 글의 댓글 수에 가중치를 적용한 값입니다.

(ii-5) negative Count : 저자가 쓴 글의 반대 수에 가중치를 적용한 값입니다.

(ii-6) **overallinfluence** : 저자의 영향력을 수치로 나타냅니다. 실제 사용자가 확인해야 할 항목입니다.

(ii-7) positive Count : 저자가 쓴 글의 찬성 수에 가중치를 적용한 값입니다.

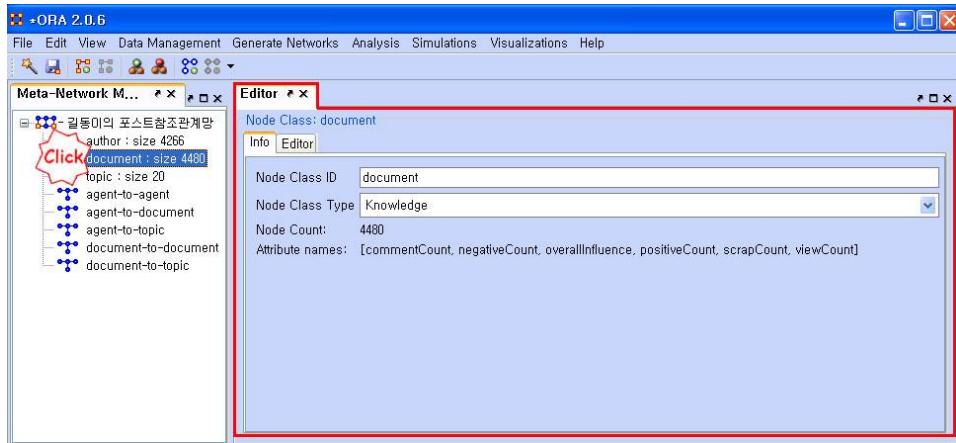
(ii-8) scrap Count : 저자가 쓴 글의 스트랩 수에 가중치를 적용한 값입니다.

(ii-9) view Count : 저자가 쓴 글의 조회 수에 가중치를 적용한 값입니다.

(c-3) 문서

저자-이슈-문서 네트워크의 하위 메뉴로 나타나는 문서는 TONK에서 메타 네트워크가 산출되어진 프로젝트의 문서입니다. size는 문서의 수를 나타내며 TONK에서 해당 프로젝트의 문서의 수와 같습니다.

Meta-Network Manager의 문서를 클릭하면 우측 Editor란에 관련정보가 나타나며 Info와 Editor 탭으로 구성되어 있습니다.



i. Info

선택된 Node의 기본 정보를 나타냅니다.



(i-1) Node Class ID : 선택된 Node의 아이디를 나타내며 아이디를 변경할 수 있습니다.

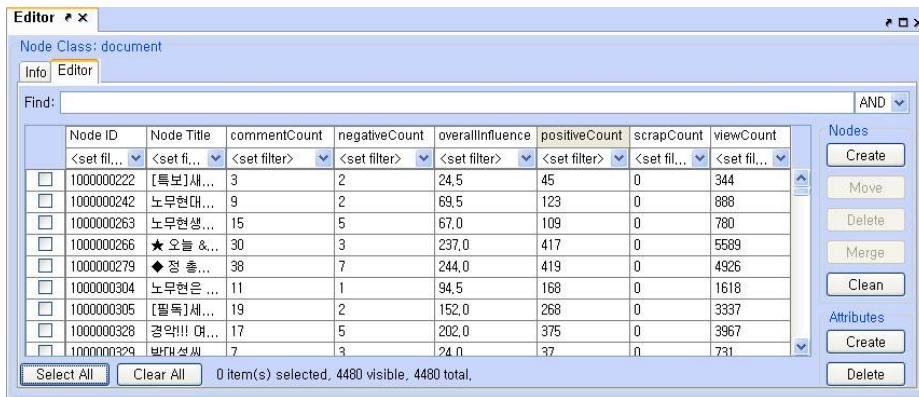
(i-2) Node Class Type : 저자가 가지고 있는 온톨로지입니다.

(i-3) Node Count : 저자의 수입니다.

(i-4) Attributes names : 저자의 특성치 목록입니다.

ii. Editor

생성된 메타 네트워크의 결과를 수치로 확인합니다.



The screenshot shows a software window titled "Editor" with a toolbar at the top. Below the toolbar is a header bar with tabs: "Info" (which is selected) and "Editor". A search bar labeled "Find:" is followed by a dropdown menu set to "AND". The main area is a table with columns: Node ID, Node Title, commentCount, negativeCount, overallInfluence, positiveCount, scrapCount, and viewCount. There are filter dropdowns for each column. The table contains several rows of data. On the right side of the window, there is a sidebar with buttons for "Nodes" (Create, Move, Delete, Merge, Clean) and "Attributes" (Create, Delete).

	Node ID	Node Title	commentCount	negativeCount	overallInfluence	positiveCount	scrapCount	viewCount
<input type="checkbox"/>	1000000222	[특보]새...	3	2	24,5	45	0	344
<input type="checkbox"/>	1000000242	노무현대...	9	2	69,5	123	0	888
<input type="checkbox"/>	1000000263	노무현생...	15	5	67,0	109	0	780
<input type="checkbox"/>	1000000266	★ 오늘 &...	30	3	237,0	417	0	5589
<input type="checkbox"/>	1000000279	◆ 정 총...	38	7	244,0	419	0	4926
<input type="checkbox"/>	1000000304	노무현은 ...	11	1	94,5	168	0	1618
<input type="checkbox"/>	1000000305	[활복]세...	19	2	152,0	268	0	3337
<input type="checkbox"/>	1000000328	경악!!! 여...	17	5	202,0	375	0	3967
<input type="checkbox"/>	1000000329	반대설씨	7	3	24,0	37	0	731

Buttons at the bottom left: "Select All" and "Clear All". Status bar: "0 item(s) selected, 4480 visible, 4480 total."

(ii-1) Find : 사용자가 원하는 특정 Node Id 및 Node Title 을 검색할 수 있습니다.

(ii-2) Node ID : 문서에 부여한 숫자입니다.

(ii-3) Node Title : 문서의 제목입니다.

(ii-4) comment Count : 문서의 댓글 수에 가중치를 적용한 값입니다.

(ii-5) negative Count : 문서의 반대 수에 가중치를 적용한 값입니다.

(ii-6) overallinfluence : 문서의 영향력을 수치로 나타냅니다.

(ii-7) positive Count : 문서의 찬성 수에 가중치를 적용한 값입니다.

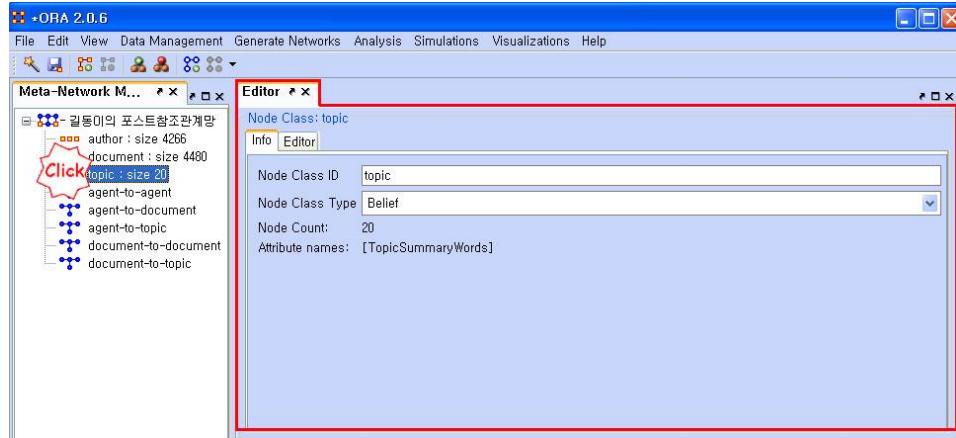
(ii-8) scrap Count : 문서의 스크랩 수에 가중치를 적용한 값입니다.

(ii-9) view Count : 문서의 조회 수에 가중치를 적용한 값입니다.

(c-4) Topic

저자-이슈-문서 네트워크의 하위 메뉴로 나타나는 토픽은 TONK에서 메타 네트워크가 산출되어진 프로젝트의 토픽입니다. size 는 토픽의 수를 나타내며 이는 TONK에서 해당 프로젝트의 토픽의 수와 같습니다.

Meta-Network Manager 의 저자를 클릭하면 우측 Editor 란에 관련정보가 나타나며 Info 와 Editor 탭으로 구성되어 있습니다.



i. Info

선택된 Node 의 기본 정보를 나타냅니다.



(i-1) Node Class ID : 선택된 Node 의 아이디를 나타내며 아이디를 변경할 수 있습니다.

(i-2) Node Class Type : 저자가 가지고 있는 온톨로지입니다.

(i-3) Node Count : 저자의 수입니다.

(i-4) Attributes names : 저자의 특성치 목록입니다.

ii. Editor

생성된 메타 네트워크의 결과를 확인합니다.

The screenshot shows the 'Editor' window for the 'topic' class. The left pane displays a table with columns: Node ID, Node Title, and TopicSummaryWords. The right pane contains buttons for managing nodes and attributes. The table data is as follows:

Node ID	Node Title	TopicSummaryWords
1	백신_하루...	백신_하루_가족
2	노무현_대...	노무현_대통령_후보
3	한국_일본...	한국_일본_시법부
4	기자_대통령...	기자_대통령_파이낸셜뉴스
5	학생_김연...	학생_김연_기자
6	직원_기자...	직원_기자_부모
7	북한_김대...	북한_김대중_남북
8	미국_중국...	미국_중국_한번
9	세종시_의...	세종시_의원_서울

Buttons on the right include: Create, Move, Delete, Merge, Clean, Attributes, Create, and Delete.

(ii-1) Find : 사용자가 원하는 특정 Node Id 및 Node Title 을 검색할 수 있습니다.

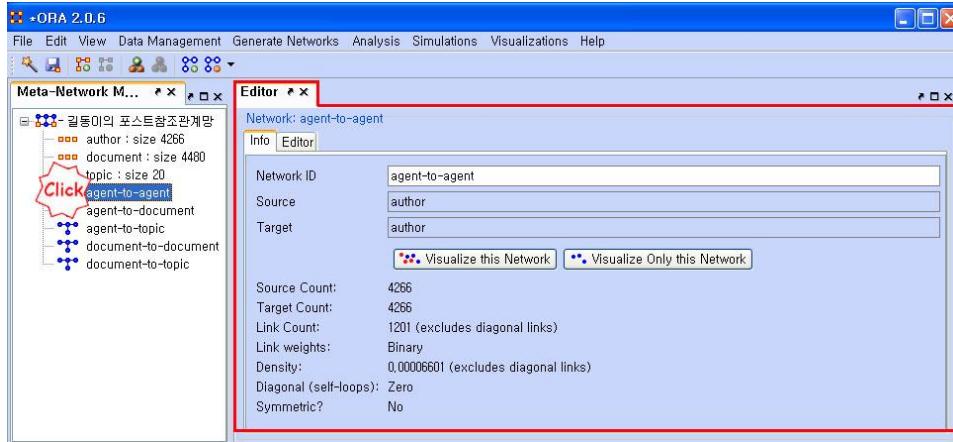
(ii-2) Node ID : 토픽에 부여한 숫자입니다.

(ii-3) Node Title : 토픽 입니다.

(ii-4) Topic Summary Words : 토픽 입니다.

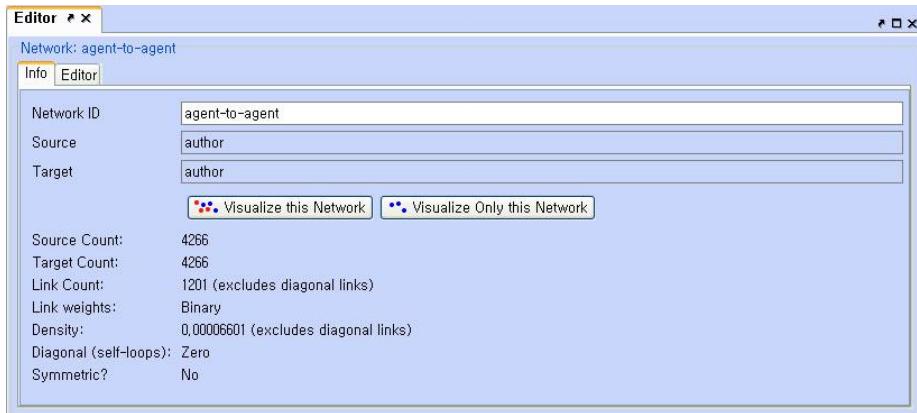
(c-5) agent-to-agent

Agent에서 Agent로의 네트워크, 즉 **agent** 와 **agent** 와의 관련 여부를 수치나 선으로 나타냅니다. Mata-Network Manager 의 agent-to-agent 를 클릭하면 우측 Editor 란에 관련정보가 나타나며 Info 와 Editor 탭으로 구성되어 있습니다.



i. Info

선택된 Network 의 기본 정보를 나타냅니다.



(i-1) Network ID : 네트워크의 이름을 나타내며 이름을 변경할 수 있습니다.

(i-2) Source : 연관관계의 시작점입니다.

(i-3) Target : 연관관계의 도착점입니다.

- 메타 네트워크의 모든 연관 관계를 보여줍니다.

선택된 Network 의 연관 관계를 보여줍니다.

(i-4) Source Count : 연관관계의 시작점의 개수입니다.

(i-5) Target Count : 연관관계의 도착점의 개수입니다.

(i-6) Link Count : agent-to-agent 네트워크에 존재하는 총 연관관계의 수입니다.

(i-7) Link weights : 연관관계의 중요도입니다.

(i-8) Density : 네트워크의 조밀도입니다.

ii. Editor

(ii-1) Binary view : 가중치가 적용되지 않은 두 Node 간의 관계를 보고자 할 때 사용되며 Yes 혹은 No 를 나타내는 체크박스의 형태로 나타냅니다. Source 가 행이며 Target 이 열입니다.

	진실승리	통일소망	일등미인	궁금해여-	조치	포시타노	25시	톨키
진실승리	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
통일소망	<input type="checkbox"/>	<input checked="" type="checkbox"/>						
일등미인	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
궁금해여-	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
조치	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
포시타노	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
25시	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
톨키	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
휘록	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
율라쁨따이	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

(ii-2) Numeric view : 가중치가 적용된 두 Node 간의 관계를 보고자 할 때 사용되며 수치로 나타냅니다. 행과 열 모두 저자입니다.

Editor

Network: agent-to-agent

Info Editor

Binary View Numeric View Row count = 4266 Column count = 4266

	진실승리	통일소망	일등미인	궁금해여-	조치	포시타노	25시	통키
진실승리	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
통일소망	0.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0
일등미인	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
궁금해여-	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
조치	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
포시타노	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
25시	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
통키	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
휘록	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
클라쁨따이	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Convert Links Remove Links Symmetrize Remove Self Loops

(c-6) agent-to-document

Agent에서 document로의 네트워크, 즉 Agent인 저자와 문서와의 관련 여부를 수치나 선으로 나타냅니다. Meta-Network Manager의 agent-to-document를 클릭하면 우측 Editor란에 관련정보가 나타나며 Info와 Editor탭으로 구성되어 있습니다.

***ORA 2.0.6**

File Edit View Data Management Generate Networks Analysis Simulations Visualizations Help

Meta-Network M... Editor

Network: agent-to-document

Info Editor

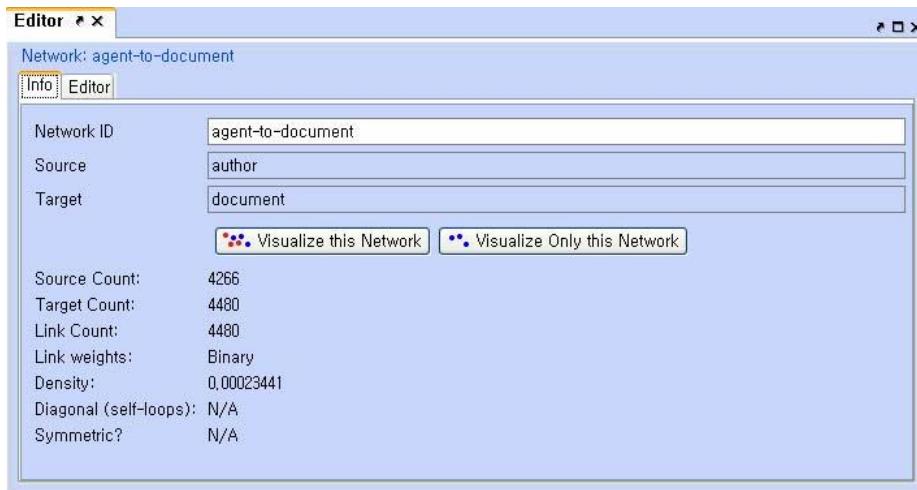
Network ID: agent-to-document
Source: author
Target: document

Visualize this Network Visualize Only this Network

Source Count: 4266
Target Count: 4480
Link Count: 4480
Link weights: Binary
Density: 0.00023441
Diagonal (self-loops): N/A
Symmetric?: N/A

i. Info

선택된 Network의 기본 정보를 나타냅니다.



(i-1) Network ID : 네트워크의 이름을 나타내며 이름을 변경할 수 있습니다.

(i-2) Source : 연관관계의 시작점입니다.

(i-3) Target : 연관관계의 도착점입니다.

[Visualize this Network] - 메타 네트워크의 모든 연관 관계를 보여줍니다.

[Visualize Only this Network] - 선택된 Network 의 연관 관계를 보여줍니다.

(i-4) Source Count : 연관관계의 시작점의 개수입니다.

(i-5) Target Count : 연관관계의 도착점의 개수입니다.

(i-6) Link Count : agent-to-agent 네트워크에 존재하는 총 연관관계의 수 입니다.

(i-7) Link weights : 연관관계의 중요도입니다.

(i-8) Density : 네트워크의 조밀도입니다.

ii. Editor

(ii-1) Binary view : 가중치가 적용되지 않은 두 Node 간의 관계를 보고자 할 때 사용되며 Yes 혹은 No 를 나타내는 체크박스의 형태로 나타냅니다. Source 가 행이며 Target 이 열입니다.

Editor

Network: agent-to-document

Info Editor

Binary View (selected) Numeric View Row count = 4266 Column count = 4480

	[특보]세종...	노무현대통...	노무현생각...	★ 오늘 &q...	◆ 정 총리...	노무현은 ...	[필독]세종...	경악!!! 여...
전실승리	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
통일소망	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>					
일등미인	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
궁금해여-	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
조치	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
포시타노	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
25시	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
통카	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
휘록	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
클라쁨따이	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Convert Links Remove Links

(ii-2) Numeric view : 가중치가 적용된 두 Node 간의 관계를 보고자 할 때 사용되며 수치로 나타냅니다. Source 가 행이며 Target 이 열입니다.

Editor

Network: agent-to-document

Info Editor

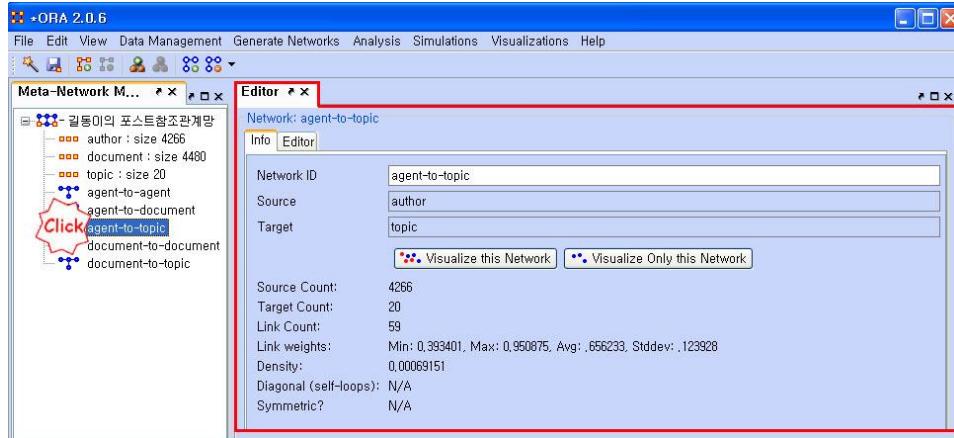
Numeric View (selected) Binary View Row count = 4266 Column count = 4480

	[특보]세종...	노무현대통...	노무현생각...	★ 오늘 &q...	◆ 정 총리...	노무현은 ...	[필독]세종...	경악!!! 여...
전실승리	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
통일소망	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
일등미인	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
궁금해여-	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
조치	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
포시타노	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
25시	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
통카	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
휘록	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
클라쁨따이	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Convert Links Remove Links

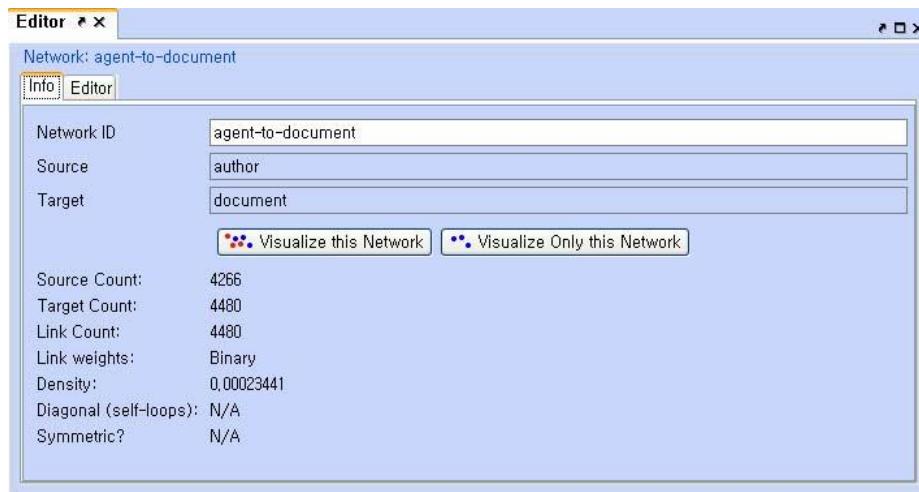
(c-7) agent-to-topic

Agent에서 Topic으로의 네트워크, 즉 Agent인 저자와 토픽과의 관련 여부를 수치나 선으로 나타냅니다. Meta-Network Manager의 agent-to-topic를 클릭하면 우측 Editor란에 관련정보가 나타나며 Info와 Editor탭으로 구성되어 있습니다.



i. Info

선택된 Network의 기본 정보를 나타냅니다.



(i-1) Network ID : 네트워크의 이름을 나타내며 이름을 변경할 수 있습니다.

(i-2) Source : 연관 관계의 시작점입니다.

(i-3) Target : 연관 관계의 도착점입니다.

Visualize this Network - 메타 네트워크의 모든 연관 관계를 보여줍니다.

Visualize Only this Network 선택된 Network의 연관 관계를 보여줍니다.

(i-4) Source Count : 연관 관계의 시작점의 개수입니다.

(i-5) Target Count : 연관 관계의 도착점의 개수입니다.

(i-6) Link Count : agent-to-agent 네트워크에 존재하는 총 연관 관계의 수입니다.

(i-7) Link weights : 연관 관계의 중요도입니다.

(i-8) Density : 네트워크의 조밀도입니다.

ii. Editor

(ii-1) Binary view : 가중치가 적용되지 않은 두 Node 간의 관계를 보고자 할 때 사용되며 Yes 혹은 No 를 나타내는 체크박스의 형태로 나타냅니다. Source 가 행이며 Target 이 열입니다.

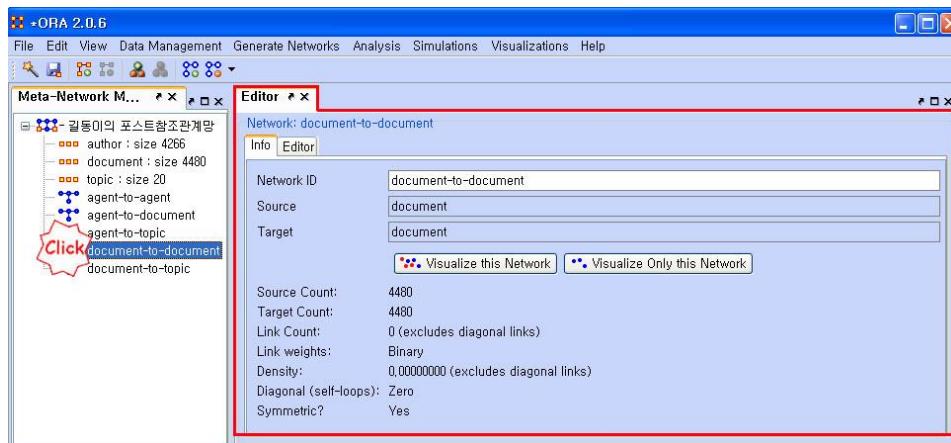
	[특보]세종...	노무현대통...	노무현생각...	★ 오늘 &q...	◆ 정 총리...	노무현은 ...	[필독]세종...	경악!! 여...
전설승리	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
통일소망	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>					
일등미인	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
궁금해여-	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
조치	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
포시타노	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
25시	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
통기	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
휘를	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
플라멩파이	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

(ii-2) Numeric view : 가중치가 적용된 두 Node 간의 관계를 보고자 할 때 사용되며 수치로 나타냅니다. Source 가 행이며 Target 이 열입니다.



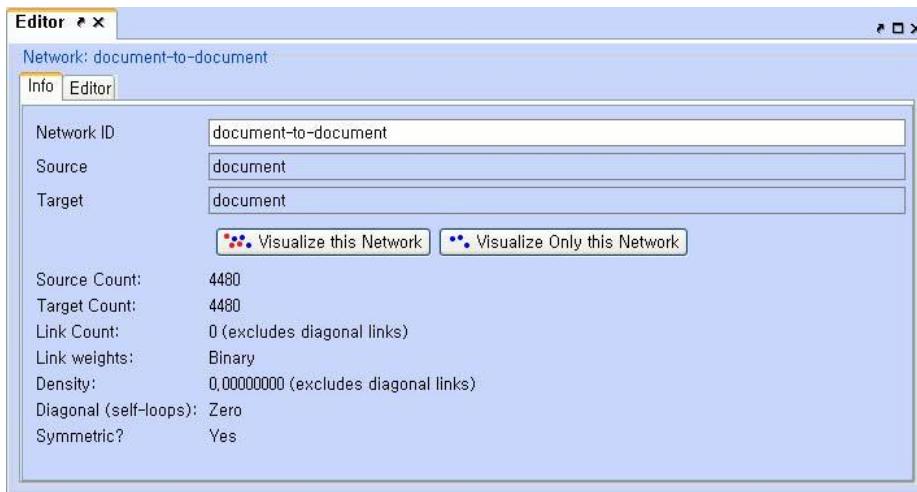
(c-8) document-to-document

문서에서 문서로의 네트워크, 즉 문서와 문서의 관련 여부를 수치나 선으로 나타냅니다. Meta-Network Manager의 document-to-document를 클릭하면 우측 Editor란에 관련정보가 나타나며 Info와 Editor 탭으로 구성되어 있습니다.



i. Info

선택된 Network의 기본 정보를 나타냅니다.



(i-1) Network ID : 네트워크의 이름을 나타내며 이름을 변경할 수 있습니다.

(i-2) Source : 연관 관계의 시작점입니다.

(i-3) Target : 연관 관계의 도착점입니다.

[Visualize this Network] - 메타 네트워크의 모든 연관 관계를 보여줍니다.

[Visualize Only this Network] - 선택된 Network 의 연관 관계를 보여줍니다.

(i-4) Source Count : 연관 관계의 시작점의 개수입니다.

(i-5) Target Count : 연관 관계의 도착점의 개수입니다.

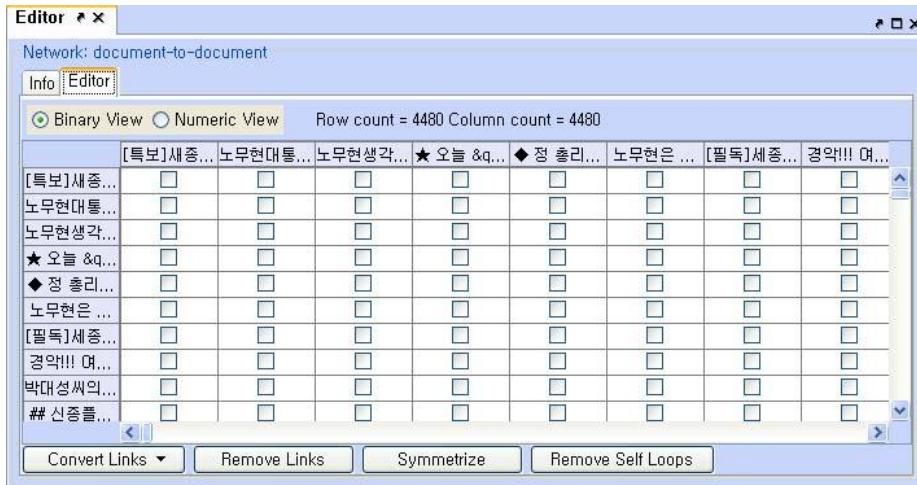
(i-6) Link Count : agent-to-agent 네트워크에 존재하는 총 연관 관계의 수입니다.

(i-7) Link weights : 연관 관계의 중요도입니다.

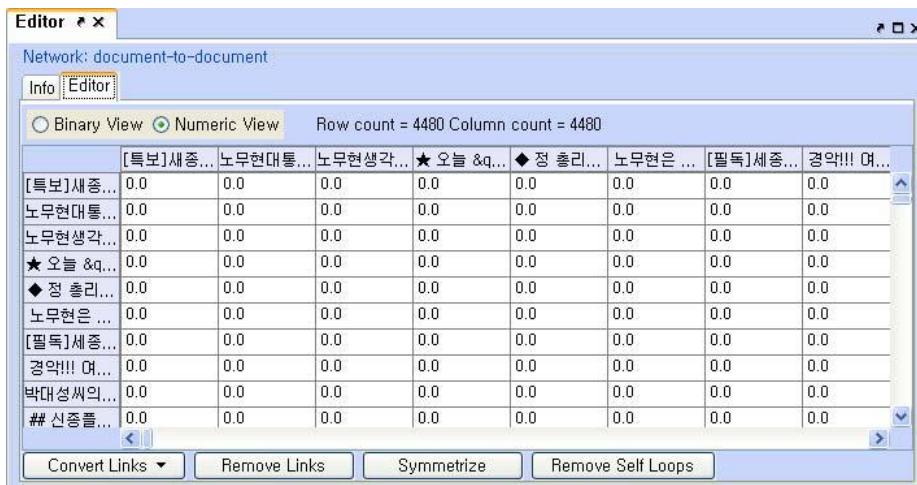
(i-8) Density : 네트워크의 조밀도입니다.

ii. Editor

(ii-1) Binary view : 가중치가 적용되지 않은 두 Node 간의 관계를 보고자 할 때 사용되며 Yes 혹은 No 를 나타내는 체크박스의 형태로 나타냅니다. Source 가 행이며 Target 이 열입니다.

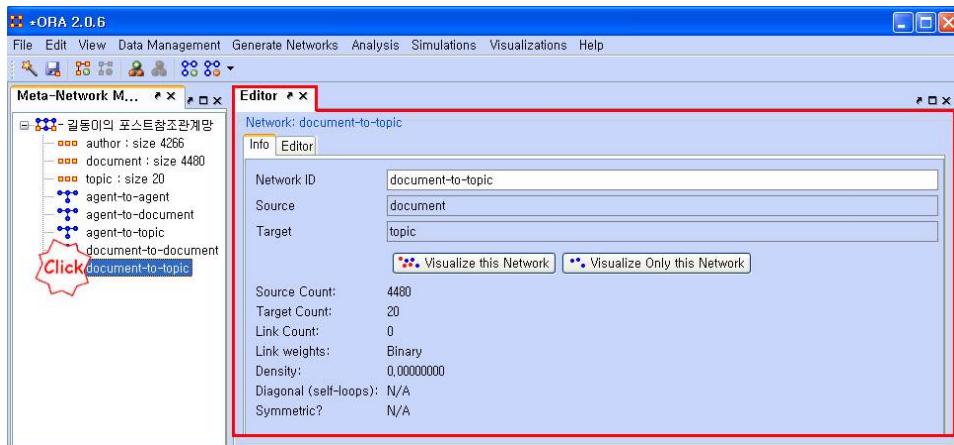


(ii-2) Numeric view : 가중치가 적용된 두 Node 간의 관계를 보고자 할 때 사용되며 수치로 나타냅니다. Source 가 행이며 Target 이 열입니다.



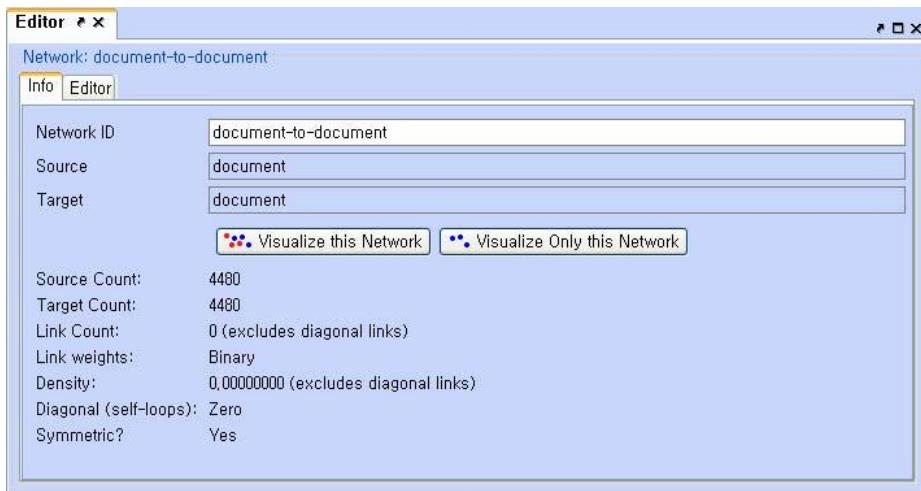
(c-9) document-to-topic

문서에서 토픽으로의 네트워크, 즉 문서와 토픽과의 관련 여부를 수치나 선으로 나타냅니다. Meta-Network Manager의 document-to-topic을 클릭하면 우측 Editor 란에 관련정보가 나타나며 Info와 Editor 탭으로 구성되어 있습니다.



i. Info

선택된 Network의 기본 정보를 나타냅니다.



(i-1) Network ID : 네트워크의 이름을 나타내며 이름을 변경할 수 있습니다.

(i-2) Source : 연관관계의 시작점입니다.

(i-3) Target : 연관 관계의 도착점입니다.

- Visualize this Network - 메타 네트워크의 모든 연관 관계를 보여줍니다.
- Visualize Only this Network 선택된 Network의 연관 관계를 보여줍니다.

(i-4) Source Count : 연관 관계의 시작점의 개수입니다.

(i-5) Target Count : 연관 관계의 도착점의 개수입니다.

(i-6) Link Count : agent-to-agent 네트워크에 존재하는 총 연관관계의 수 입니다.

(i-7) Link weights : 연관 관계의 중요도입니다.

(i-8) Density : 네트워크의 조밀도입니다.

ii. Editor

(ii-1) Binary view : 가중치가 적용되지 않은 두 Node 간의 관계를 보고자 할 때 사용되며 Yes 혹은 No 를 나타내는 체크박스의 형태로 나타냅니다. Source 가 행이며 Target 이 열입니다.

(ii-2) Numeric view : 가중치가 적용된 두 Node 간의 관계를 보고자 할 때 사용되며 수치로 나타냅니다. Source 가 행이며 Target 이 열입니다.

Editor * x Network: document-to-document

Info Editor

Binary View Numeric View Row count = 4480 Column count = 4480

	[특보]새증...	노무현대통...	노무현생각...	★ 오늘 &q...	◆ 정 총리...	노무현은 ...	[필독]새증...	경악!!! 며...
[특보]새증...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
노무현대통...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
노무현생각...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
★ 오늘 &q...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
◆ 정 총리...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
노무현은 ...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
[필독]새증...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
경악!!! 며...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
박대성씨의...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
# 신증플...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Convert Links ▾ Remove Links Symmetrize Remove Self Loops

K. 기존 이슈와의 관계 분석 실행하기 — 저자-이슈 분석

이번 장에서는 문서의 메타 정보를 파악할 수 있습니다. 메타 정보는 저자, 주제(이슈), 저자의 영향력, 문서가 참조된 관계(링크된 문서)입니다. 이전 과정에서 실행되었던 분석들이 문서 내에서 주로 다루어진 내용을 파악하는 일이었다면 이번 장부터는 생성된 결과를 기반으로 한 메타 정보를 파악합니다. 다음은 ‘기존 이슈와의 관계 분석’의 자세한 내용입니다.

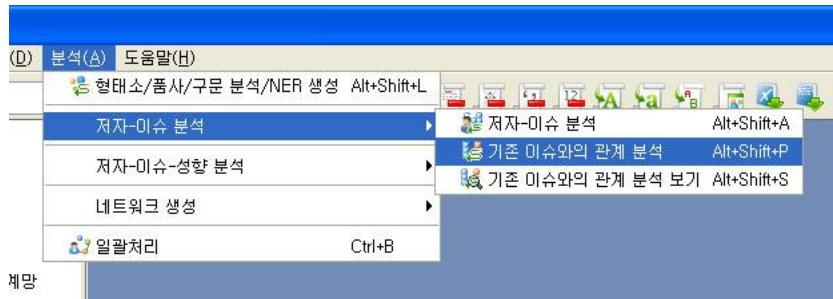
1. 기존 이슈와의 관계 분석 – 저자-이슈 분석 중심으로

‘저자-이슈 분석’을 골자로 한 ‘기존 이슈와의 관계 분석’은 프로젝트에서 주로 다른 토픽들과 다른 프로젝트에서 주로 다른 토픽들 간의 유사도를 측정합니다. 이는 한 프로젝트 내에서 토픽들간의 유사도를 측정하는 토픽간 관계와는 다른 개념으로 **사용자가 직접 비교 대상이 되는 프로젝트들을 지정하여 이들간의 토픽들을 분석합니다.**

‘저자-이슈 분석’을 골자로 한 ‘기존 이슈와의 관계 분석’은 각 프로젝트의 토픽을 산출한 결과를 대상으로 하기 때문에 반드시 <**III.G. 저자-이슈 분석 결과 이해하기, page 160**>와 <**III.H. 저자-이슈-성향 분석 결과 이해하기, page 191**>까지 마친 상태에서 실행되어야 합니다.

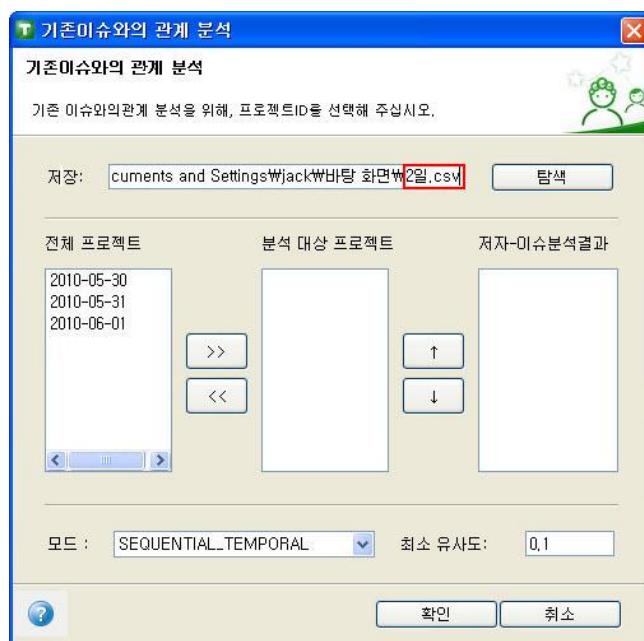
‘기존 이슈와의 관계 분석’의 결과는 2 가지 종류가 있습니다. 첫 번째는 한 프로젝트에서 다뤄진 토픽이 그 **이후 프로젝트**에서는 얼마만큼의 비중을 가지고 다루어졌는지 측정합니다. 두 번째 분석은 한 프로젝트에서 다뤄진 토픽이 그 **이전 프로젝트**에서는 얼마만큼의 비중을 가지고 다루어졌는지 측정합니다. 여기에서 말하는 ‘이전 프로젝트’와 ‘이후 프로젝트’는 우리가 쉽게 생각할 수 있는 시간적 개념뿐만 아니라 선택된 프로젝트 간의 전후 관계도 의미합니다. 다음은 ‘저자-이슈-문서 네트워크 생성’을 실행하는 과정입니다.

- a) 기존 이슈와의 관계 분석을 실행하려면 **메뉴막대 – 분석 – 저자-이슈 분석 – 기존 이슈와의 관계 분석**을 누릅니다.



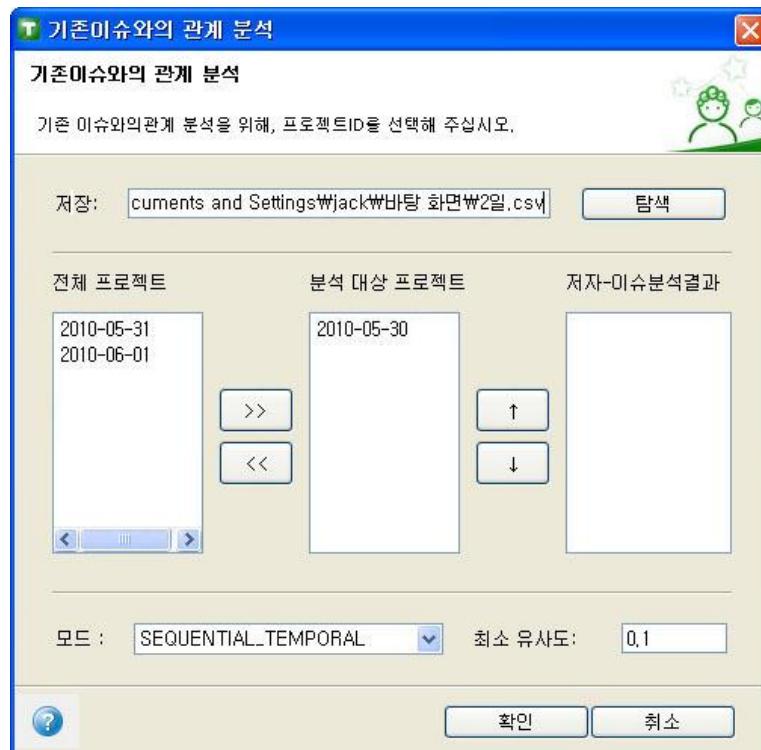
b)

(b-1) '기준이슈와의 관계 분석' 대화상자가 나타나면 <탐색>을 클릭한 후 기준 이슈와의 관계 분석의 결과가 저장되는 장소를 선택합니다. '저장'에 주소가 나타나면 원하는 결과의 이름과 확장자인.csv 를 입력합니다.



바탕화면에 '2 일'이라고 저장한 예

(b-2) 이후 ‘전체 프로젝트’에서 분석하고자 하는 프로젝트를 클릭한 후  추가 버튼을 눌러 ‘분석 대상 프로젝트’로 이동시킵니다. 만약 이동 시킨 프로젝트를 취소 하고자 하는 경우에는  삭제 버튼을 누릅니다.



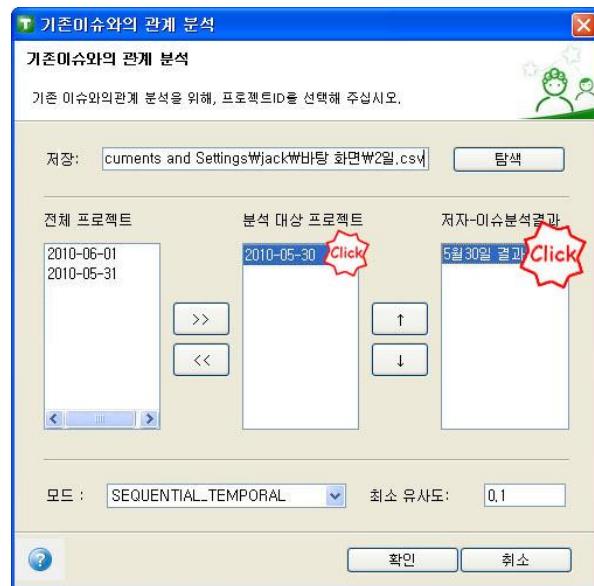
‘분석 대상 프로젝트’로 이동한 2010-05-30 프로젝트

만약 컴포넌트 트리에 나타난 프로젝트가 4개 이더라도 프로젝트 내에 저자-이슈 분석 결과가 존재하지 않는다면 대화상자의 ‘전체 프로젝트’에서 목록이 나타나지 않습니다. 따라서 **프로젝트 뿐만 아니라 프로젝트의 저자-이슈 분석결과도 반드시 존재해야 합니다.**



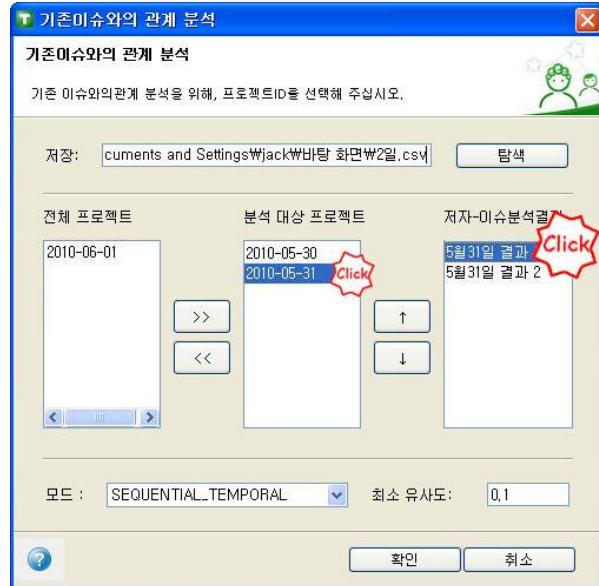
‘저자-이슈 분석’의 결과를 산출한 프로젝트

(b-3) 선택된 프로젝트가 ‘분석 대상 프로젝트’로 이동하면 다시 프로젝트를 클릭합니다. ‘저자-이슈 분석결과’에 해당 프로젝트의 결과명이 나타납니다. 분석하고자 하는 결과문서를 클릭합니다.



'2010-05-30' 프로젝트의 '5월 30일 결과 1'을 분석대상으로 선택

(b-4) 다시 '전체 프로젝트'로 돌아옵니다. 앞에서 선택한 프로젝트와 비교하고자 하는 프로젝트를 클릭하여 '분석 대상 프로젝트'로 이동시킨 후, '저자-이슈 분석결과'에서 결과문서를 누릅니다. 비교 분석하고자 하는 프로젝트가 여러 개일 경우 위와 같은 과정을 반복합니다.



'2010-05-31' 프로젝트의 '5 월 31 일 결과 1'을 분석대상으로 선택

(b-5) 모드

- i. SEQUENTIAL_TEMPORAL : 분석 대상이 시간대별로 3 가지 프로젝트가 존재한다고 가정할 경우, 시간에 흐름에 맞춰 유사도를 측정합니다. 예를 들어 Monday, Tuesday, Wednesday라는 이름을 가진 프로젝트에서는 Monday와 Tuesday의 유사도를 계산하고, Tuesday와 Wednesday간의 유사도를 계산합니다.
- ii. END_TIME_BASED : 분석 대상이 시간대별로 3 가지 프로젝트가 존재한다고 가정할 경우, 시간의 흐름보다는 마지막 시간대의 분석 결과가 어느 시간대의 분석 결과와 유사한지 측정합니다. 예를 들어 Monday, Tuesday, Wednesday라는 이름을 가진 프로젝트에서는 Wednesday와 Monday의 유사도를 계산하고, Wednesday와 Tuesday간의 유사도를 계산합니다.

(b-6) 최소 유사도

사용자가 분석 결과를 유사하다고 판정하기 위한 최소한의 유사도 값입니다. 범위는 0에서 1 사이이며 주로 0.3¹을 사용합니다.

(b-7) 버튼: 위 아래 버튼을 클릭하여 결과가 나타나고자 하는 프로젝트를 가장 상위에 위치시킵니다. 상기의 예시에서는 ‘2010-05-30’ 프로젝트가 맨 위에 있으므로 기존 프로젝트에 ‘2010-05-30’ 프로젝트 결과가 나타나게 됩니다.

설정이 끝나면 <확인>을 누릅니다.

c) 명령이 성공적으로 수행되었다면 속성에는 `***** Successfully executed *****` 가 표기됩니다.

¹ 일반 블로그와 게시판을 분석한 결과는 0.3 이상의 유사도를 지닌 관계가 유의미하다고 판단합니다.

2. 기존 이슈와의 관계 분석 처리 결과 보기

- a) ‘기존 이슈와의 관계 분석’의 결과를 확인하려면 **메뉴—분석—저자-이슈 분석 — 기존 이슈와의 관계 분석 보기**를 클릭합니다.



- b) ‘탐색’ 대화상자가 나타나면 ‘기존 이슈와의 관계 분석’ 대화상자의 ‘저장’에서 생성한 파일을 찾아 클릭을 합니다.



- c) 작업창에 나타나는 초기화면은 다음과 같습니다.



(c-1) 분석 방법: ‘기준 이슈와의 관계 분석’ 대화상자의 ‘모드’에서 설정하였던 분석 방법입니다.

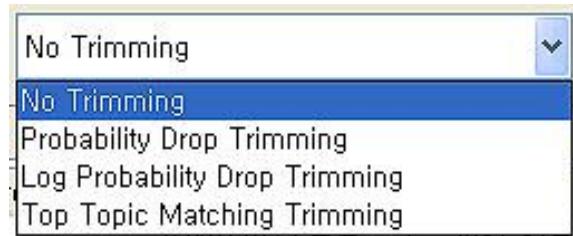
(c-2) 기준 프로젝트: 분석 대상으로 지정하였던 프로젝트들 중에서 사용자가 특정 프로젝트를 선택하여 이를 기준으로 나머지 프로젝트들과의 유사성을 판단합니다. 이 때 특정 프로젝트를 기준 프로젝트라고 정의합니다.

(c-3) 기준 토픽 : 기준 프로젝트를 선택한 후 세부적으로 이의토픽을 선택합니다. 주의할 점은 ‘프로젝트 선택’에서만 그치는 것이 아니라 이에 해당하는 토픽 또한 반드시 지정해주어야 한다는 것입니다. 지정하지 않을 시에는 기본적으로 토픽 1에 대한 유사도가 산출됩니다.

(c-4) 토픽명 : 기준 토픽의 숫자에 해당하는 토픽입니다.

(c-5) 대상 프로젝트 : 앞에서 설정한 기준 프로젝트 이외에 비교 대상이 되는 프로젝트입니다. 메뉴에는 기준 프로젝트를 제외한 나머지 프로젝트가 나타납니다

(c-6) 정제방법 : <검색>을 눌렀을 때 표시되는 결과의 개수를 제한합니다.



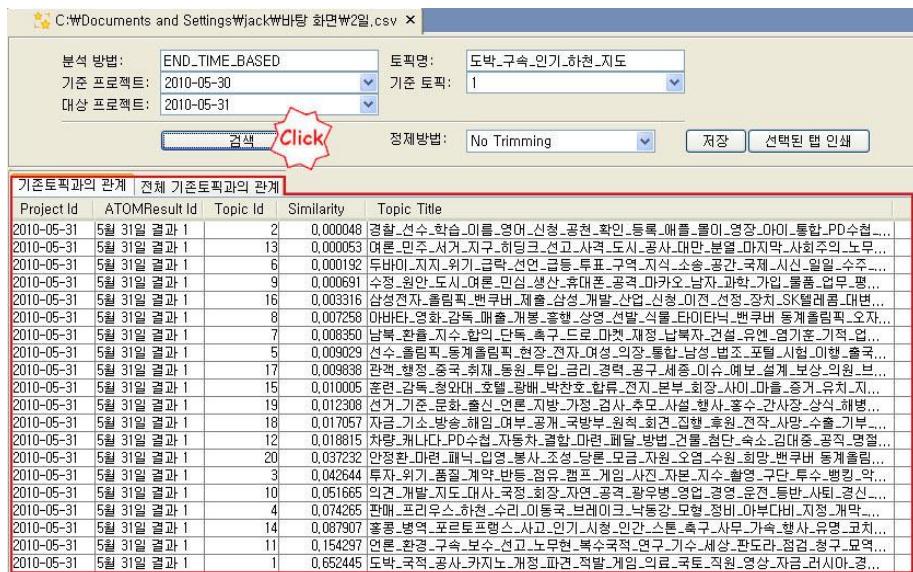
- i. No Trimming : 처리를 실행하지 않음
- ii. Probability Drop Trimming : 가장 큰 비중값으로 떨어지는 지점까지만 결과값을 표시
- iii. Log Probability Drop Trimming : 비중값을 Log 로 치환한 후 가장 큰 값으로 떨어지는 지점까지 표시
- iv. Top Topic Matching Trimming : 검색된 단어, 저자, 문서로 해당 토픽을 가장 높은 비중값으로 들고 있을 확률

(c-7) 검색 : 조건 설정이 마무리 되면 결과를 확인합니다.

d) 분석 결과

기준이 되는 프로젝트를 선택하고 이에 해당하는 토픽을 클릭한 후 <검색>을 누르면 ‘기준 이슈와의 분석’의 결과가 나타납니다. 하위 메뉴는 2 가지로 토픽들 간의 유사성을 사용자가 선택하여 열람할 수 있는 기준토픽과의 관계와 사용자가 이를 선택할 필요 없이 화면에 전부 나타내주는 전체 기준토픽과의 관계가 있습니다.

(d-1) 기준토픽과의 관계



The screenshot shows a search interface for '기준 토픽과의 관계'. The search bar contains 'No Trimming'. A red starburst surrounds the 'Click!' button next to the search bar. The results table has columns: Project Id, ATOMResult Id, Topic Id, Similarity, and Topic Title. The table lists 20 rows of data, each showing a project ID, topic ID, similarity score (e.g., 0.000048, 0.000053), and a topic title (e.g., 경찰_선수_학습_이름_영어_신청_공천_확인_등록_애플_물미_영장_마이_통합_PD수첩...).

Project Id	ATOMResult Id	Topic Id	Similarity	Topic Title
2010-05-31	5월 31일 결과 1	2	0.000048	경찰_선수_학습_이름_영어_신청_공천_확인_등록_애플_물미_영장_마이_통합_PD수첩...
2010-05-31	5월 31일 결과 1	13	0.000053	여론_민주_서거_지구_해밀크_선고_사격_도시_공사_대만_분열_마지막_사회주의_노무...
2010-05-31	5월 31일 결과 1	6	0.000194	두바이_지지_위기_급락_선언_급등_투표_구역_지식_소송_공간_국제_시신_일일_수주...
2010-05-31	5월 31일 결과 1	9	0.000691	수정_원안_도시_여론_민심_생산_휴대폰_공격_마카오_남자_과학_가입_물품_업무_평...
2010-05-31	5월 31일 결과 1	16	0.003316	삼성전자_홀리파크_백화점_제출_삼성_개발_산업_신청_미전_설정_장치_SK텔레콤_대변...
2010-05-31	5월 31일 결과 1	8	0.007258	마비타_영화_감독_매출_개봉_홀리파크_상영_선발_식사를_타이타닉_백쿠버_동계올림픽_오자...
2010-05-31	5월 31일 결과 1	7	0.008351	남북_한글_지수_합의_단독_축구_드로_마켓_재정_남북자_건설_유엔_임기_훈_기적_업...
2010-05-31	5월 31일 결과 1	5	0.009029	선수_홀리파크_동계올림픽_현장_전자_아트_의장_틀립_남성_별조_포털_시월_미행_풀국...
2010-05-31	5월 31일 결과 1	17	0.009838	관객_행정_중국_취재_동원_투입_금리_경력_공구_세종_미수_예보_설계_보상_의원_브...
2010-05-31	5월 31일 결과 1	15	0.010005	훈련_감독_청와대_호텔_광배_박찬호_합류_전지_본부_회장_사이_마들_증거_유치_지...
2010-05-31	5월 31일 결과 1	19	0.012308	선거_기준_문화_출신_언론_지방_가정_검사_주모_사설_행사_홍수_간사장_상식_해병...
2010-05-31	5월 31일 결과 1	18	0.017057	자금_기소_방송_해임_여부_공개_국방부_원희_회견_집행_후원_전작_서망_수출_기부...
2010-05-31	5월 31일 결과 1	12	0.018815	차량_캐나다_PD수첩_자동차_콜한_마련_폐달_방법_건물_첨단_속소_김대중_공작_명절...
2010-05-31	5월 31일 결과 1	20	0.037232	안정한_마련_페닉_입영_봉사_조성_당문_도모_자원_오염_수원_회망_백쿠버_동계올림...
2010-05-31	5월 31일 결과 1	3	0.042641	투자_위기_통길_계약_반동_점유_캠프_개입_사전_자본_지수_촬영_구단_투수_랭킹_의...
2010-05-31	5월 31일 결과 1	10	0.051665	의견_개발_지도_대사_국정_회장_자연_환경_광우병_영업_경영_운전_돌반_시론_경신...
2010-05-31	5월 31일 결과 1	4	0.074265	판매_프리우스_한국_수리_미동국_브레이크_낙동강_모현_정비_마부단비_지정_개막...
2010-05-31	5월 31일 결과 1	14	0.087907	홍콩_병역_포로_포령스_사고_인기_시청_인간_스론_축구_사무_기술_행사_유명_코치...
2010-05-31	5월 31일 결과 1	11	0.154297	언론_환경_구속_보수_선고_노무현_복수국적_연구_기수_세상_판도라_점검_청구_묘역...
2010-05-31	5월 31일 결과 1	1	0.652445	도박_국적_공사_카지노_개정_파견_적발_개입_의료_국토_직원_영상_자금_러시아_경...

- i. Project Id : 기준이 되는 프로젝트와 비교·분석되는 프로젝트의 Id 입니다.
- ii. ATOMResult Id : 기준이 되는 프로젝트와 비교·분석되는 프로젝트의 저자-이슈 분석의 결과명 입니다.
- iii. Topic Id : 기준 토픽의 아이디입니다.
- iv. Similarity: 기준이 되는 토픽이 다른 프로젝트의 토픽들과 얼마나 유사하게 다루어졌는지 수치로 나타냅니다. 수치의 범위는 0~1 까지이며 1에 가까울수록 유사도가 커집니다.
- v. 토픽: 단어 사전에 수집된 단어들을 대상으로 프로그램이 서로 관련 있는 단어들을 모아둔 것입니다.

상기의 예시에서는 기준 프로젝트를 **2010-05-30** 으로 지정하고 이에 대한 기준 토픽으로 1 번을 설정했습니다. 대상 프로젝트는 **2010-05-30** 프로젝트와 비교 대상이 되는 **2010-05-31** 프로젝트가 나타납니다.

Project By 토픽 메뉴의 토픽 Id 에서는 **2010-05-30** 프로젝트의 1 번 토픽과 가장 유사하게 다루어진 토픽의 Id 가 나타나게 됩니다. 2 번 토픽 이 Similarity 의 수치¹가 가장 높기 때문에 **2010-05-30** 프로젝트의 1 번 토픽과 가장 유사성이 높다고 해석할 수 있습니다. 마찬가지로 기준이 되는 프로젝트를 **2010-05-31** 로 설정하고 이의 토픽을 2 번을 선택하면 **2010-05-30** 프로젝트의 1 번 토픽 이 가장 유사도가 높게 나타나는 것도 확인할 수 있습니다.

이 결과는 **2010-05-30** 프로젝트에서 가장 비중 있게 다루어진 1 번 토픽 이 **2010-05-31** 프로젝트에서는 2 번째로 비중 있게 다루어진 2 번 토픽과 대응되면서 시간이 지남에 따라 ‘도박-구속-인기-하천-지도’를 화두로 한 내용의 비중이 약간 떨어진 것을 알 수 있습니다.

(d-2) 전체 기준토픽과의 관계

¹ 최대 유사도 값을 0 으로 지정하였기 때문에 0 이상의 값이 산출됩니다.

Click

Topics	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0.652	0.007	0.027	0.002	0.003	0.007	0.072	0.067	0.038	0.006	0.01	0.006	0.05	0.028	0.045	0.008	0.067	0.016	0.086	0.014
2	0	0.001	0.122	0.139	0.057	0.01	0.004	0.23	0.118	0	0.003	0.001	0.043	0.099	0.004	0.043	0.016	0	0.04	0.021
3	0.043	0.01	0.147	0.03	0.002	0.001	0.016	0.028	0.045	0.07	0.264	0	0.002	0.015	0.007	0	0.01	0.005	0.002	0.056
4	0.074	0.001	0.213	0.012	0.006	0.026	0.06	0.037	0	0.223	0.007	0.005	0.035	0.002	0	0	0.002	0.097	0.02	0.006
5	0.009	0.019	0.013	0.233	0.011	0.004	0.006	0.04	0.014	0.014	0.006	0.006	0.012	0.421	0.052	0.002	0.015	0.001	0.014	0.007
6	0	0.013	0.002	0.005	0.002	0.007	0.003	0.044	0.001	0	0.399	0.033	0.004	0	0	0	0.015	0.008	0.001	0.001
7	0.008	0.015	0.052	0.021	0.169	0.159	0.015	0.07	0.015	0.001	0.073	0.022	0.047	0.018	0.002	0.006	0.006	0.01	0.005	0.004
8	0.007	0.003	0.006	0.318	0.002	0	0.006	0.017	0	0	0	0.001	0.005	0.007	0.001	0.004	0.001	0.626	0	0.026
9	0.001	0	0.014	0.002	0.016	0.005	0.001	0.017	0.019	0.012	0.008	0.504	0.021	0	0.003	0	0.004	0	0.038	0.003
10	0.052	0.108	0.02	0.035	0.04	0.005	0.04	0.019	0.103	0.019	0.003	0.003	0.037	0	0.007	0.027	0.11	0.005	0.026	0.047
11	0.154	0.004	0.02	0.026	0.073	0.002	0.003	0.032	0.049	0.021	0.002	0.001	0.102	0.099	0.013	0.218	0.011	0.019	0.008	0.066
12	0.019	0.001	0.01	0.003	0.029	0.006	0.043	0.035	0.001	0.232	0.024	0.003	0.008	0.104	0.025	0.111	0	0.008	0.007	0.003
13	0	0.002	0.052	0.022	0.087	0.008	0.028	0.046	0.001	0.001	0.002	0.155	0.111	0.014	0.033	0	0.002	0.003	0.005	0.085
14	0.088	0.031	0.045	0.011	0.005	0.003	0.06	0.056	0.013	0.041	0	0.006	0.009	0.054	0.048	0.026	0.017	0.001	0.205	0.068
15	0.01	0.049	0.04	0.562	0.002	0.01	0	0.028	0	0.001	0.001	0.001	0.002	0.016	0.005	0.012	0.013	0.016	0.05	0.282
16	0.003	0.006	0.054	0.016	0.023	0.009	0.01	0.014	0.196	0	0.005	0.014	0.076	0.469	0.003	0.007	0.002	0.023	0.002	0.013
17	0.01	0.059	0.061	0.002	0.017	0.004	0.014	0.062	0	0.006	0	0.07	0.018	0	0.041	0.031	0.022	0.222	0.001	0.002
18	0.017	0.001	0.019	0.046	0.162	0.022	0.456	0.056	0.002	0.009	0.01	0.001	0.009	0.045	0.007	0.045	0.155	0	0.006	0.042
19	0.012	0.184	0.064	0.002	0.021	0	0.122	0.159	0.002	0.006	0.004	0.029	0.253	0.012	0	0.14	0.006	0.004	0.003	0.011
20	0.037	0.001	0.053	0.006	0.01	0.015	0.074	0.133	0.009	0.001	0.036	0.015	0.042	0.06	0.054	0.003	0.008	0.013	0.044	0.015

행은 대상 프로젝트의 토픽 ID이고 열은 기준 프로젝트의 토픽 ID입니다. 기준 프로젝트의 토픽 ID 6 번과 대상 프로젝트의 토픽 ID 7 번의 유사도는 표에서도 알 수 있듯이 0.159입니다. 이처럼 사용자가 토픽을 직접 설정하지 않고도 유사성을 한눈에 확인할 수 있습니다. 유사도가 0 으로 표기된 곳은 유사성이 거의 없다는 뜻입니다.

L. 기존 이슈와의 관계 분석 실행하기 — 저자-이슈-성향 분석

이번 장에서는 문서의 메타 정보를 파악할 수 있습니다. 메타 정보는 저자, 주제(이슈), 저자의 영향력, 문서가 참조된 관계(링크된 문서)입니다. 이전 과정에서 실행되었던 분석들이 문서 내에서 주로 다루어진 내용을 파악하는 일이었다면 이번 장부터는 생성된 결과를 기반으로 한 메타 정보를 파악합니다. 다음은 ‘기존 이슈와의 관계 분석’의 자세한 내용입니다.

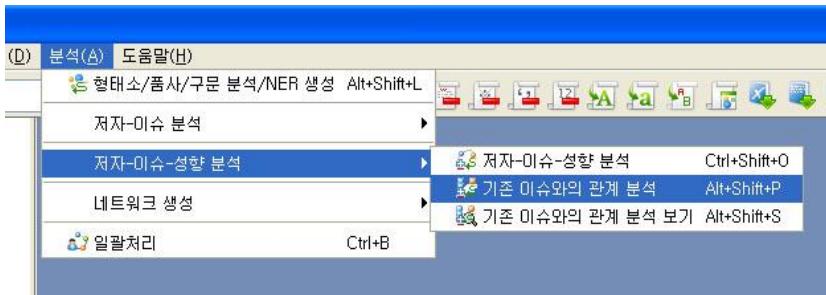
1. 기존 이슈와의 관계 분석 – 저자-이슈-성향 분석 중심으로

‘저자-이슈-성향 분석’을 골자로 한 ‘기존 이슈와의 관계 분석’은 프로젝트에서 주로 다른 토픽들과 다른 프로젝트에서 주로 다른 토픽들 간의 유사도를 측정합니다. 이는 한 프로젝트 내에서 토픽들간의 유사도를 측정하는 토픽간 관계와는 다른 개념으로 사용자가 직접 비교 대상이 되는 프로젝트들을 지정하여 이들간의 토픽들을 분석합니다.

‘저자-이슈-성향 분석’을 골자로 한 ‘기존 이슈와의 관계 분석’은 각 프로젝트의 토픽을 산출한 결과를 대상으로 하기 때문에 반드시 <III.G. 저자-이슈 분석 결과 이해하기, page 160>와 <III.H. 저자-이슈-성향 분석 결과 이해하기, page 191>까지 마친 상태에서 실행되어야 합니다.

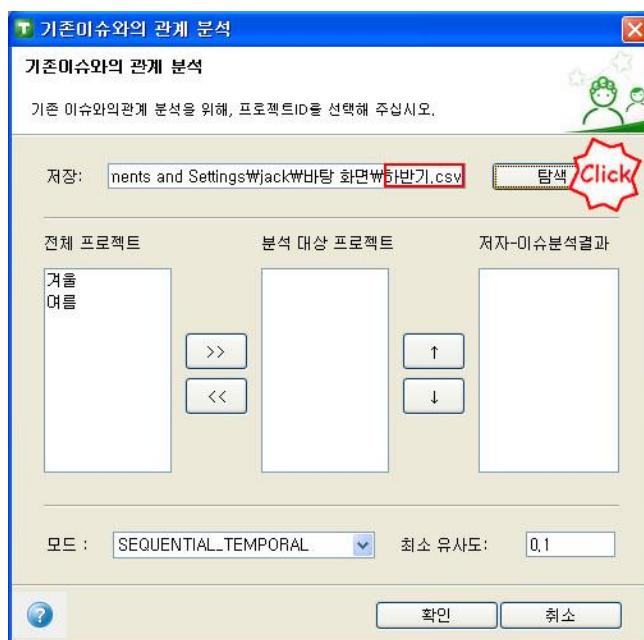
‘기존 이슈와의 관계 분석’의 결과는 2 가지 종류가 있습니다. 첫 번째는 한 프로젝트에서 다뤄진 토픽이 그 이후 프로젝트에서는 얼마만큼의 비중을 가지고 다루어졌는지 측정합니다. 두 번째 분석은 한 프로젝트에서 다뤄진 토픽이 그 이전 프로젝트에서는 얼마만큼의 비중을 가지고 다루어졌는지 측정합니다. 여기에서 말하는 ‘이전 프로젝트’와 ‘이후 프로젝트’는 우리가 쉽게 생각할 수 있는 시간적 개념뿐만 아니라 선택된 프로젝트 간의 전후 관계도 의미합니다. 다음은 ‘저자-이슈-문서 네트워크 생성’을 실행하는 과정입니다.

- a) 기존 이슈와의 관계 분석을 실행하려면 **메뉴막대 – 분석 – 저자-이슈-성향 분석 – 기존 이슈와의 관계 분석**을 누릅니다.



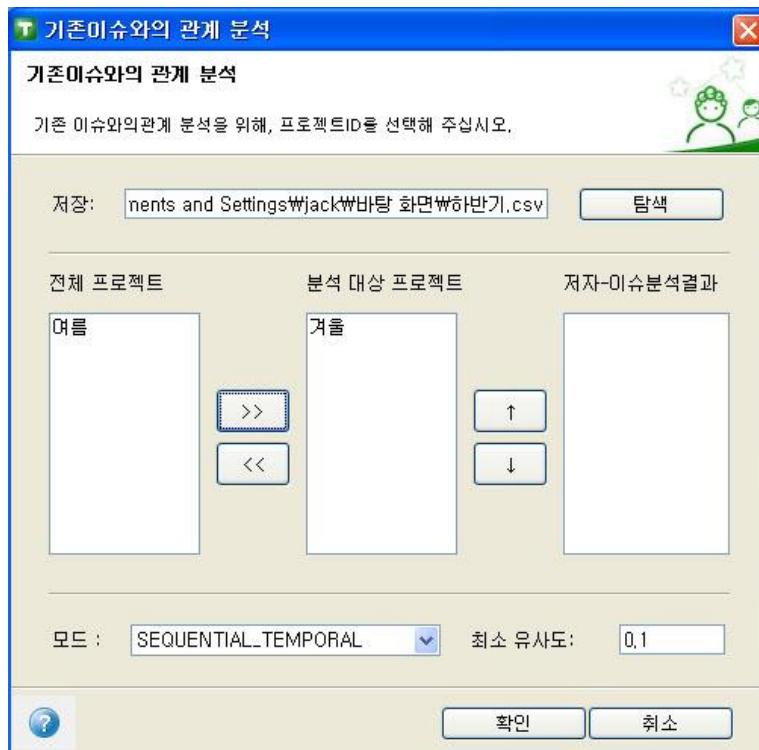
b)

(b-1) ‘기존이슈와의 관계 분석’ 대화상자가 나타나면 <탐색>을 클릭한 후 기존 이슈와의 관계 분석의 결과가 저장되는 장소를 선택합니다. ‘저장’에 주소가 나타나면 원하는 결과의 이름과 확장자인.csv 를 입력합니다.



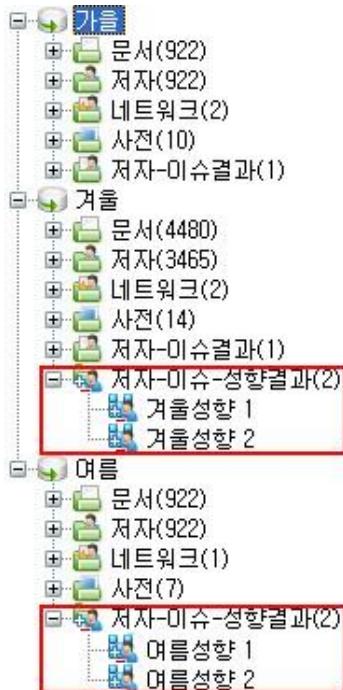
바탕화면에 ‘하반기’라고 저장한 예

(b-2) 이후 ‘전체 프로젝트’에서 분석하고자 하는 프로젝트를 클릭한 후  추가 버튼을 눌러 ‘분석 대상 프로젝트’로 이동시킵니다. 만약 이동 시킨 프로젝트를 취소 하고자 하는 경우에는  삭제 버튼을 누릅니다.



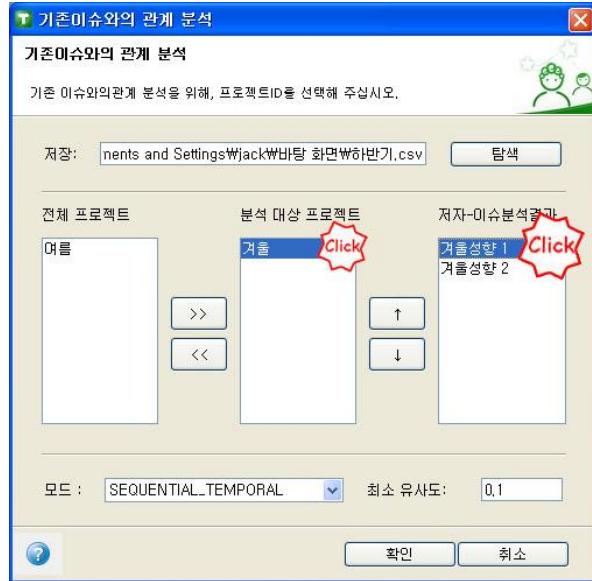
‘분석 대상 프로젝트’로 이동한 ‘겨울’ 프로젝트

만약 컴포넌트 트리에 나타난 프로젝트가 3 개 이더라도 프로젝트 내에 저자-이슈-성향 분석 결과가 존재하지 않는다면 대화상자의 ‘전체 프로젝트’에서 목록이 나타나지 않습니다. 따라서 프로젝트 뿐만 아니라 프로젝트의 저자-이슈-성향 분석결과도 반드시 존재해야 합니다.



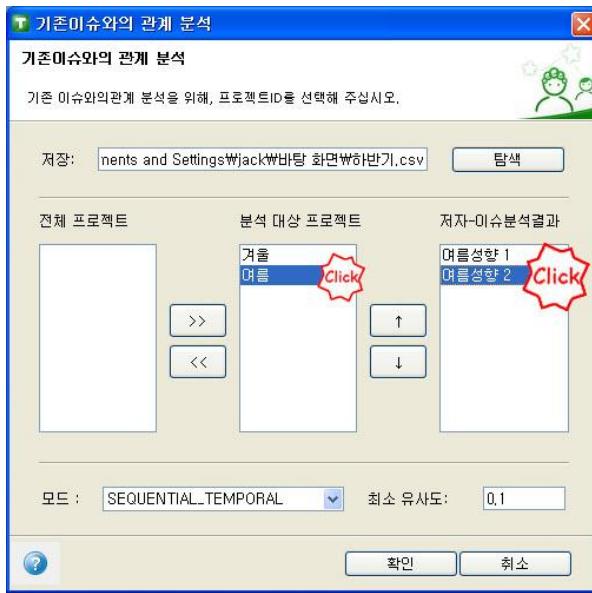
'저자-이슈-성향 분석'의 결과를 산출한 프로젝트

(b-3) 선택된 프로젝트가 '분석 대상 프로젝트'로 이동하면 다시 '겨울' 프로젝트를 클릭합니다.
'저자-이슈-성향 분석결과'에 해당 프로젝트의 결과명이 나타납니다. 분석하고자 하는 결과문서를
클릭합니다.



‘겨울’ 프로젝트의 ‘겨울성향 1’을 분석대상으로 선택

(b-4) 다시 ‘전체 프로젝트’로 돌아옵니다. 앞에서 선택한 프로젝트와 비교하고자 하는 프로젝트를 클릭하여 ‘분석 대상 프로젝트’로 이동시킨 후, ‘저자-이슈-성향분석결과’에서 결과문서를 누릅니다. 비교 분석하고자 하는 프로젝트가 여러 개일 경우 위와 같은 과정을 반복합니다.



‘여름’ 프로젝트의 ‘여름성향 1’을 분석대상으로 선택

(b-5) 모드

- i. SEQUENTIAL_TEMPORAL : 분석 대상이 시간대별로 3 가지 프로젝트가 존재한다고 가정할 경우, **시간에 흐름에 맞춰** 유사도를 측정합니다. 예를 들어 Monday, Tuesday, Wednesday 라는 이름을 가진 프로젝트에서는 Monday 와 Tuesday 의 유사도를 계산하고, Tuesday 와 Wednesday 간의 유사도를 계산합니다.
- ii. END_TIME_BASED : 분석 대상이 시간대별로 3 가지 프로젝트가 존재한다고 가정할 경우, 시간의 흐름보다는 **마지막 시간대의 분석 결과가 어느 시간대의 분석 결과와 유사한지** 측정합니다. 예를 들어 Monday, Tuesday, Wednesday 라는 이름을 가진 프로젝트에서는 Wednesday 와 Monday 의 유사도를 계산하고, Wednesday 와 Tuesday 간의 유사도를 계산합니다.

(b-6) 최소 유사도

사용자가 분석 결과를 유사하다고 판정하기 위한 최소한의 유사도 값입니다. 범위는 0에서 1 사이이며 주로 0.3¹을 사용합니다.

(b-7) 버튼: 위 아래 버튼을 클릭하여 결과가 나타나고자 하는 프로젝트를 가장 상위에 위치시킵니다. 상기의 예시에서는 ‘겨울’ 프로젝트가 맨 위에 있으므로 기존 프로젝트에 ‘겨울’ 프로젝트가 나타나게 됩니다.

설정이 끝나면 <확인>을 누릅니다.

c) 명령이 성공적으로 수행되었다면 속성에는 ******* Successfully executed ******* 가 표기됩니다.

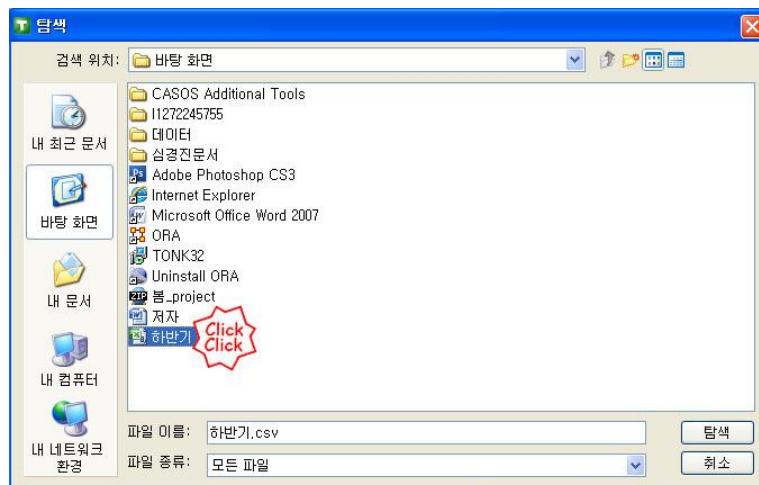
¹ 일반 블로그와 게시판을 분석한 결과는 0.3 이상의 유사도를 지닌 관계가 유의미하다고 판단합니다.

2. 기존 이슈와의 관계 분석 처리 결과 보기

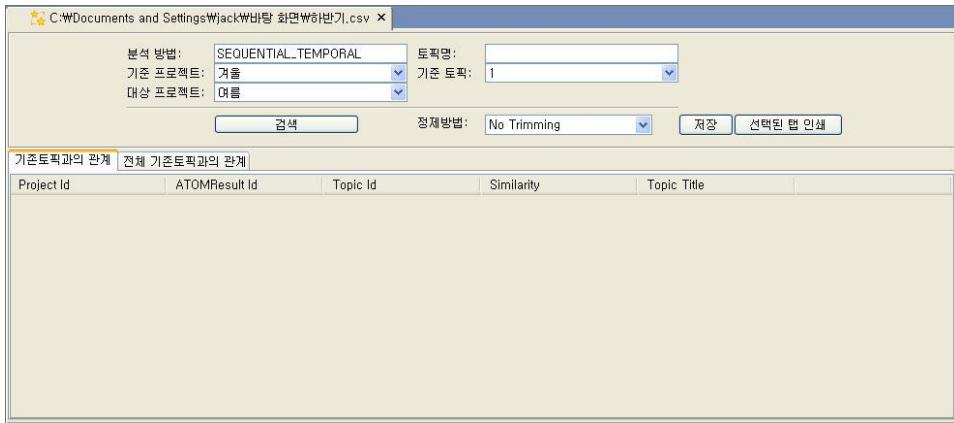
- a) ‘기존 이슈와의 관계 분석’의 결과를 확인하려면 **메뉴 – 분석 – 저자-이슈-성향 분석 – 기존 이슈와의 관계 분석 보기**를 클릭합니다.



- b) ‘탐색’ 대화상자가 나타나면 ‘기존 이슈와의 관계 분석’ 대화상자의 ‘저장’에서 생성한 파일을 찾아 클릭을 합니다.



- c) 작업창에 나타나는 초기화면은 다음과 같습니다.



(c-1) 분석 방법: ‘기준 이슈와의 관계 분석’ 대화상자의 ‘모드’에서 설정하였던 분석 방법입니다.

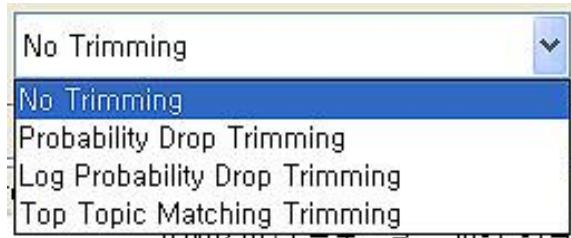
(c-2) 기준 프로젝트: 분석 대상으로 지정하였던 프로젝트들 중에서 사용자가 특정 프로젝트를 선택하여 이를 기준으로 나머지 프로젝트들과의 유사성을 판단합니다. 이 때 특정 프로젝트를 기준 프로젝트라고 정의합니다.

(c-3) 기준 토픽 : 기준 프로젝트를 선택한 후 세부적으로 이의토픽을 선택합니다. 주의할 점은 ‘프로젝트 선택’에서만 그치는 것이 아니라 이에 해당하는 토픽 또한 반드시 지정해주어야 한다는 것입니다. 지정하지 않을 시에는 기본적으로 토픽 1에 대한 유사도가 산출됩니다.

(c-4) 토픽명 : 기준 토픽의 숫자에 해당하는 토픽입니다.

(c-5) 대상 프로젝트 : 앞에서 설정한 기준 프로젝트 이외에 비교 대상이 되는 프로젝트입니다. 메뉴에는 기준 프로젝트를 제외한 나머지 프로젝트가 나타납니다

(c-6) 정제방법 : <검색>을 눌렀을 때 표시되는 결과의 개수를 제한합니다.



i. No Trimming : 처리를 실행하지 않음

- ii. Probability Drop Trimming : 가장 큰 비중값으로 떨어지는 지점까지만 결과값을 표시
- iii. Log Probability Drop Trimming : 비중값을 Log로 치환한 후 가장 큰 값으로 떨어지는 지점까지 표시
- iv. Top Topic Matching Trimming : 검색된 단어, 저자, 문서로 해당 토픽을 가장 높은 비중값으로 들고 있을 확률

(c-7) 검색 : 조건 설정이 마무리 되면 결과를 확인합니다.

d) 분석 결과

기준이 되는 프로젝트를 선택하고 이에 해당하는 토픽을 클릭한 후 <검색>을 누르면 ‘기준 이슈와의 분석’의 결과가 나타납니다. 하위 메뉴는 2 가지로 토픽들 간의 유사성을 사용자가 선택하여 열람할 수 있는 기준토픽과의 관계와 사용자가 이를 선택할 필요 없이 화면에 전부 나타내주는 전체 기준토픽과의 관계가 있습니다.

(d-1) 기준토픽과의 관계

Project Id	ATOMResult Id	Topic Id	Similarity	Topic Title
여름	여름상향 1	12	0.169931	의원_수정_해임_제출_공격_분열_의견_원안_확인_선언_사이트_여론_미경_호란_민주주의_회고위원회_재단
여름	여름상향 1	15	0.149681	지방_선거_노무현_김해_서거_죽모_의원_분열_민주_행사_마련_봉하마을_출신_김대중_백지_한창_민심_청당_민주주의_작동
여름	여름상향 1	-3	0.361690	여론_수정_수출_풀이_국정_통제_공직_원안_방송_사설_기념_여론_풀이_공작_비속_실장_수주_율단_권태신_사회주의
여름	여름상향 1	-4	0.115454	의원_해임_수정_당론_제출_기독_국정_조속_신청_의견_특포_민본_민주_미정현_호란_조기_결의_사고_통합
여름	여름상향 1	-8	0.138443	남북_언론_정부_대변인_협의_최근_납북자_지지_수정_자료_사설_원칙_협의_미경_선거_국정_조기_방송_의원_지방

- i. Project Id : 기준이 되는 프로젝트와 비교·분석되는 프로젝트의 Id입니다.
- ii. ATOMResult Id : 기준이 되는 프로젝트와 비교·분석되는 프로젝트의 저자-이슈 분석의 결과명입니다.
- iii. Topic Id : 기준 토픽의 아이디입니다.
- iv. Similarity: 기준이 되는 토픽이 다른 프로젝트의 토픽들과 얼마나 유사하게 다루어졌는지 수치로 나타냅니다. 수치의 범위는 0~1 까지이며 1에 가까울수록 유사도가 커집니다.
- v. Topic Title: 단어 사전에 수집된 단어들을 대상으로 프로그램이 서로 관련 있는 단어들을 모아둔 것입니다.

상기의 예시에서는 기준 프로젝트를 **겨울**로 지정하고 이에 대한 기준 토픽으로 -15 번을 설정했습니다. 대상 프로젝트는 **겨울** 프로젝트와 비교 대상이 되는 **여름** 프로젝트가 나타납니다.

기준토픽과의 관계 메뉴의 토픽 Id 에서는 **겨울** 프로젝트의 -15 번 토픽 과 가장 유사하게 다른 토픽의 Id 가 나타나게 됩니다. 12 번 토픽 이 Similarity 의 수치¹가 가장 높기 때문에 **겨울** 프로젝트의 -15 번 토픽과 가장 유사성이 높다고 해석할 수 있습니다. 마찬가지로 기준이 되는 프로젝트를 **여름**으로 설정하고 이의 토픽을 12 번을 선택하면 **겨울** 프로젝트의 -15 번 토픽 이 가장 유사도가 높게 나타나는 것도 확인할 수 있습니다.

(d-2) 전체 기준토픽과의 관계

		기준토픽과의 관계																				전체 기준토픽과의 관계																																	
To...		1	3	4	5	6	8	9	14	15	16	17	18	19	-2	-4	-6	-7	-8	-9	-11	-14	-15	-16	-17	-18	-19	-20																											
2		0	0	0	0	0	0	0.195	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																		
3		0	0	0	0	0	0	0	0	0	0	0	0	0	0.148	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0																		
4		0	0	0	0	0	0	0	0	0	0	0	0	0	0.191	0	0	0	0	0.172	0	0	0	0.115	0	0	0	0.101	0	0	0	0	0	0	0	0																			
6		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.12	0	0	0	0	0	0	0	0	0	0	0	0																	
9		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.103	0	0																	
12		0	0	0	0	0	0.181	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.11	0	0	0	0	0	0	0	0	0	0	0	0.172	0	0																
14		0	0.163	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.173	0.215	0	0.144	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0												
15		0	0	0.31	0	0	0	0	0	0	0	0	0	0	0.197	0	0	0	0	0	0	0	0	0	0	0.15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0										
17		0	0	0	0	0	0	0	0	0.114	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0										
19		0	0	0	0	0	0	0	0	0	0	0.105	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0									
20		0.161	0	0	0	0	0	0	0	0	0	0	0	0	0.275	0.142	0	0	0.233	0.109	0	0.108	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0									
-1		0	0	0	0	0	0	0	0.163	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0								
-2		0	0	0	0	0.181	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.159	0							
-3		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.362	0	0	0	0	0	0	0	0	0	0	0	0.167	0	0	0	0	0	0	0	0	0								
-4		0	0	0	0	0.144	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.115	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.104	0								
-5		0	0	0	0	0	0	0	0	0	0	0	0	0	0.128	0.184	0	0	0.108	0	0.117	0.113	0	0	0	0	0.114	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0								
-6		0	0	0	0	0	0	0	0	0.312	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.158	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0								
-8		0	0	0	0.153	0	0	0	0.115	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.138	0	0	0	0.249	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0					
-9		0	0	0	0.104	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
-10		0	0	0	0	0	0	0.136	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
-11		0	0	0	0	0	0	0	0.124	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.201	0	0	0	0	0	0		
-17		0	0	0	0	0	0	0	0	0.435	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
-19		0	0	0	0.101	0	0	0	0	0	0	0	0.103	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

¹ 최대 유사도 값을 0 으로 지정하였기 때문에 0 이상의 값이 산출됩니다.

행은 대상 프로젝트의 토픽 ID이고 열은 기준 프로젝트의 토픽 ID입니다. 기준 프로젝트의 토픽 ID -15 번과 대상 프로젝트의 토픽 ID 12 번의 유사도는 표에서도 알 수 있듯이 0.17입니다. 이처럼 사용자가 토픽을 직접 설정하지 않고도 유사성을 한눈에 확인할 수 있습니다. 유사도가 0으로 표기된 곳은 유사성이 거의 없다는 뜻입니다.

Chapter IV 두번째 분석 - 문서의 개념어 간 연결관계

문서의 개념어 간 연결관계를 분석하는 것은 TONK에서 지정한 특정 개념을 가진 개념어들이 서로 관계가 있는지의 여부를 파악하는 것입니다. 다시 말하면, TONK는 개념어들을 추출하여 이들 간의 관계가 존재하는지 알아봅니다. 특히 결과를 확인하는 과정에서는 ORA라는 프로그램을 별도로 이용합니다. 다음은 문서의 개념어 간 연결관계 분석에 대한 과정입니다.

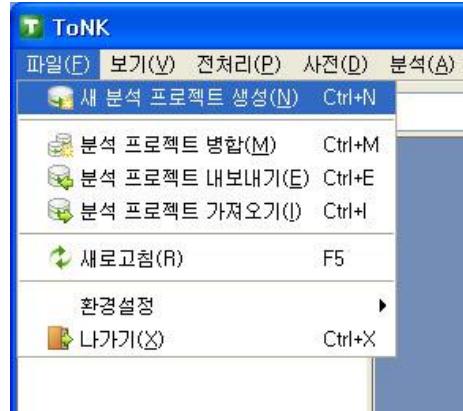
A. 프로젝트 생성하기

새로운 분석을 시작하기 위해 ‘새 분석 프로젝트’를 생성합니다. 워드나 한글 프로그램에서 새 문서를 생성하듯이 TONK에서는 새 분석 프로젝트를 생성합니다. 프로젝트는 총 4개의 하위 폴더로 구성되었으며, 분석 대상인 문서가 모여 있는 ‘문서’, 문서의 저자들이 모여 있는 ‘저자’, 문서, 저자, 이슈의 관계를 나타내는 네트워크, 사전을 모아놓은 ‘사전’, 저자-이슈 분석과 저자-이슈-성향 분석의 결과가 모여 있는 ‘저자-이슈 결과’와 ‘저자-이슈-성향 결과’가 있습니다. 생성된 프로젝트는 파일로 자동으로 저장이 되며 이를 임의로 삭제하지 않는 이상 추후 TONK를 실행하였을 때 컴포넌트 트리에 자동으로 나타납니다.

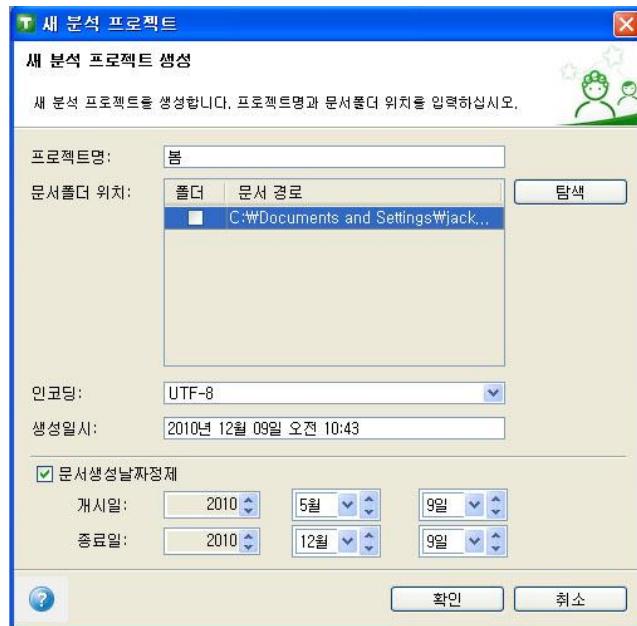
Comment [j2]: 모르겠음

1. 새 분석 프로젝트 생성하기

- TONK 프로그램을 실행하면 컴포넌트 트리와 작업창에 빈 화면이 나타납니다. 새 분석 프로젝트를 실행하려면 <파일>에서 ‘새 분석 프로젝트 생성’을 누릅니다.

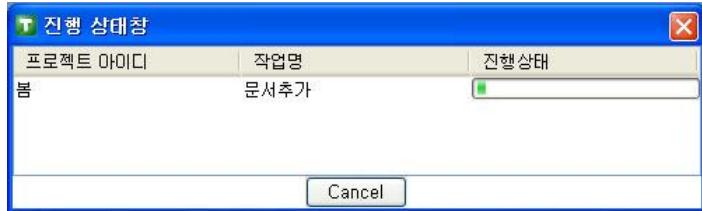


- b) ‘새 분석 프로젝트’ 대화상자가 나타나면 ‘프로젝트 명’에는 본인이 원하는 제목을 기입합니다. ‘문서 폴더 위치’에는 우측의 <탐색>을 클릭하여 분석하고자 하는 문서가 위치해 있는 폴더를 지정한 후 <확인>을 클릭합니다. ‘문서생성날짜정제’에 체크하여 개시일과 종료일을 설정하면 선택된 문서 폴더 안에서 지정한 기간 안에 생성된 문서가 수집됩니다.



- c) ‘진행 상태창’ 대화상자가 나타나면 연두색 막대가 채워질 때까지 기다립니다. 이때 <Cancel>을 클릭하면 ‘자연어 처리’가 중단됩니다. 이 과정에서는 분석 프로젝트의

하위폴더를 생성하고 분석 대상을 추출합니다. 이때 <취소>를 클릭하면 새 분석 프로젝트 생성이 중단됩니다.



- d) 작업이 완료되면 컴포넌트 트리에 프로젝트가 생성됩니다. 생성된 프로젝트를 더블 클릭하면 이를 구성하고 있는 4 가지 (문서, 저자, 네트워크, 사전) 하위 폴더가 나타납니다.



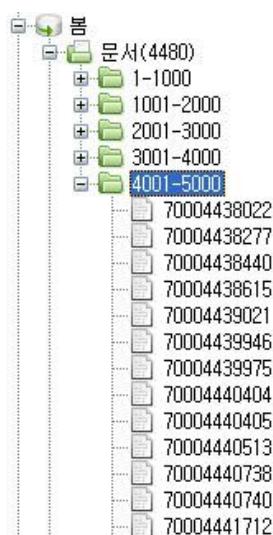
- e) 생성된 프로젝트는 파일의 형태로 사용자가 지정한 작업폴더(최초 설정- C://Workplace, page 29, (d-2) 참조)에 저장되며 하단의 속성에서 저장된 위치 등 상세정보를 확인할 수 있습니다.
- f) 명령이 성공적으로 수행되었다면 속성에는 ***** Successfully executed ***** 가 표기됩니다.

2. 새 분석 프로젝트 생성 처리 결과 보기

생성된 분석 프로젝트는 컴포넌트 트리와 작업창에서 확인합니다. 프로젝트를 더블 클릭하면 총 4 개의 하위폴더가 보여지며 새 분석 프로젝트를 바로 생성한 상태에서는 2 개의 하위 폴더(문서, 저자)만 열람이 가능합니다.

a) 문서

‘문서’는 사용자가 새 프로젝트를 생성할 시 분석 대상으로 선정했던 문서를 모아놓은 폴더입니다. 새 분석 프로젝트를 생성한 후 컴포넌트 트리 및 작업창에서 바로 열람이 가능합니다.



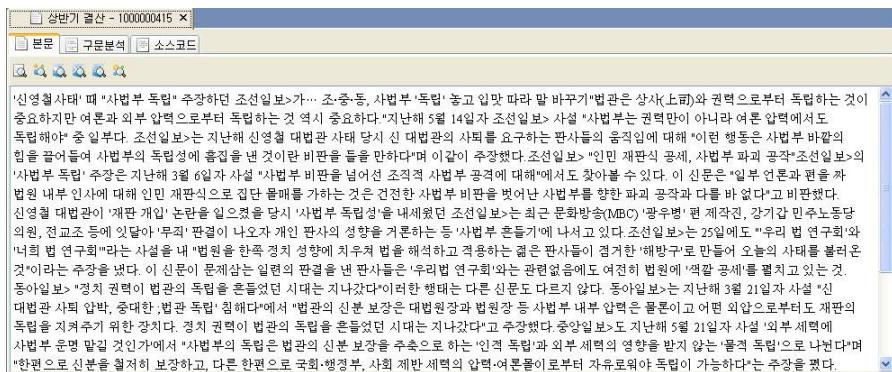
‘문서’를 더블클릭 하면 문서의 개수가 많은 경우를 감안하여 1000 단위¹로 나누어 생성된 하위 폴더가 차례로 나타납니다. 나열된 문서들은 문서명으로 보여지게 되며 괄호 안의 숫자는 문서의

¹ 문서의 개수가 10,000 이라면 1-1000, 1001-2000, 2001-3000 등과 같이 총 10 개의 하위폴더가 생성됩니다.

개수를 의미합니다. ‘문서’에서는 예외적으로 프로그램이 문서에 대한 문서 아이디¹를 별개로 지정하지 않고 기존의 문서명과 동일하게 사용합니다. 즉, ‘문서’에서 문서명과 문서 아이디는 같습니다.

이 중 하나의 문서를 클릭하면 메시지 패널의 속성에서 관련 정보(문서명, 위치, 작성자, 생성일 등)가 표시되고 상태표시줄에는 문서가 저장된 위치가 보여집니다. 문서를 더블 클릭하면 작업창에 문서의 내용이 보여집니다. 수집된 된 문서를 삭제하고 싶을 경우에는 해당 사전에 마우스 오른쪽 버튼을 눌러 삭제를 클릭합니다.

문서에 있는 임의의 문서를 더블 클릭했을 때 작업창은 다음과 같습니다.



수집된 문서들의 원문을 새로 생성된 작업창에서 확인할 수 있습니다.

b) 저자

‘저자’는 ‘문서’로 수집된 문서의 저자를 모아놓은 폴더입니다. 새 분석 프로젝트를 생성한 후 컴포넌트 트리에서만 열람이 가능합니다. 이후 <III.F 저자-이슈 분석 실행하기, page 135>에서 다뤄질 ‘저자-이슈 분석’을 실행해야만 ‘저자’의 작업창을 열람할 수 있습니다. 괄호 안의 숫자는 총 저자의 개수를 의미합니다.

¹ 문서 아이디는 문서명과는 다른 개념으로 프로그램에서 임의로 지정한 숫자형태의 아이디입니다.



‘저자’를 더블 클릭하면 목록에서는 저자명¹이 나열됩니다. 만일 동일한 저자명이 목록에 나타날 경우, 이는 오류가 난 것이 아니라 단지 문서에 대한 추출된 저자명이 같을 뿐² 실제 저자는 다른 사람입니다. 바꾸어 말하면, 저자들은 각자 다른 사람이더라도 이름은 동일할 수 있기 때문에 저자명은 중복 되어서 나타날 수 있습니다.

‘저자 ID’란 프로그램이 각 저자에 임의로 부여한 숫자입니다. 만약 저자명이 같아서 내용파악이 어려울 경우 속성의 ‘저자 아이디’ 란을 참고하면 쉽게 이해할 수 있습니다.

목록 중 임의의 저자명 클릭하면 속성에 관련 정보(저자명, ID, 작성문서 수)가 표시되고 상태표시줄에는 문서가 저장된 위치가 나타납니다. 다시 임의의 저자명을 더블 클릭하면 작업창에는 저자에 대한 분석내용이 보여 지는데 이는 ‘저자-이슈 분석’을 실행한 후에만 나타납니다.

¹ 저자명은 문서를 작성한 사람이나 기관의 이름입니다.

² ‘저자’에서 다루는 정보는 동일한 저자명을 한 대 모아 분석하는 것이 아니라 각 문서에 대한 저자를 개별적으로 분석하는 것입니다.

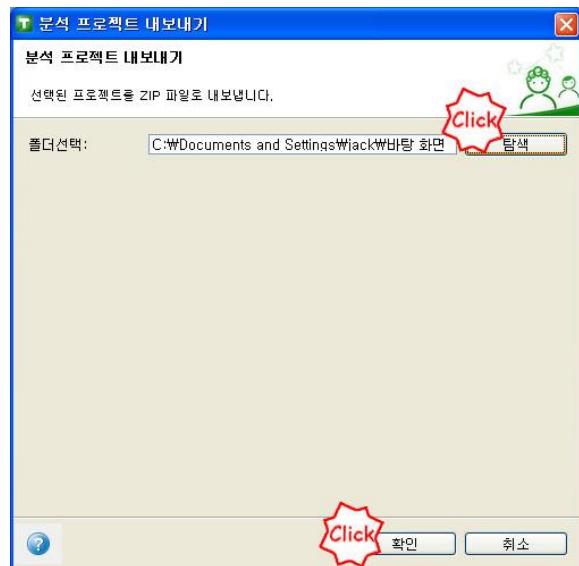
3. 분석 프로젝트 내보내기

TONK는 새 프로젝트를 만들과 동시에 사용자가 지정한 작업폴더(최초 설정- C://Workplace, page 29, (d-2) 참조)에 저장을 합니다. 하지만 사용자가 분석내용을 파일로 저장하고자 한다면 ‘분석 프로젝트 내보내기’를 사용합니다.

- <파일>에서 ‘분석 프로젝트 내보내기’를 클릭합니다.



- ‘분석 프로젝트 내보내기’ 대화상자가 나타나면 <탐색>을 클릭하여 저장할 장소를 선택한 후 <확인>을 누릅니다.



- c) 작업이 완료되면 사용자가 지정한 곳에 선택한 프로젝트가 생성된 것을 확인할 수 있습니다.

파일을 저장한 위치 및 파일 이름은 정확히 기억하는 것이 좋으므로 찾기 쉬운 곳에 저장해 둡니다. 또한 <새 폴더>를 클릭하여 관련된 문서를 한 폴더에 저장해 둔다면 문서 관리하기가 수월해 집니다. 작업이 완료되면 지정한 위치에 선택한 프로젝트가 생성된 것을 확인할 수 있습니다.

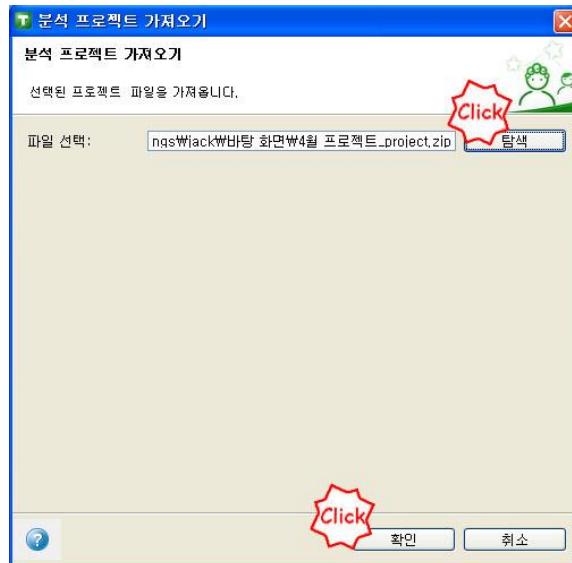
4. 분석 프로젝트 가져오기

저장되어 있는 TONK 파일을 열기 위해 ‘분석 프로젝트 가져오기’를 사용합니다.

- a) ‘파일’에서 ‘분석 프로젝트 가져오기’를 클릭합니다.



- b) ‘분석 프로젝트 가져오기’ 대화상자에서 <탐색>을 클릭하여 TONK으로 불러올 파일을 찾아 선택한 후 <확인>을 클릭합니다. 불러온 파일은 Zip 형태이어야 합니다.



- c) 작업이 완료되면 컴포넌트 트리에 선택한 프로젝트가 생성된 것을 확인할 수 있습니다.

B. 본문 정제 실행하기

1. 본문 정제

새 프로젝트를 생성한 후 문서의 본문을 정렬해서 보고자 할 때 사용합니다. 다음과 같이 본문 정제에는 총 7 가지 종류가 있습니다.

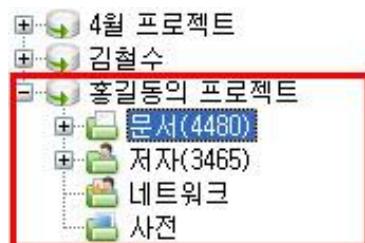
종 류	기 능
여유 공란 삭제	2칸 이상으로 띄어진 공란을 원문에서 삭제
문장 기호 삭제	따옴표나 느낌표 등 문장 기호를 원문에서 삭제
심볼 삭제	목욕탕이나 당구장 표시 등 심볼을 원문에서 삭제
숫자 삭제	조사와 같이 쓰이지 않고 단독으로 쓰인 숫자를 원문에서 삭제
영어 대문자화	원문에 있는 알파벳을 모두 대문자로 변환
영어 소문자화	원문에 있는 알파벳을 모두 소문자로 변환
바꾸기	원문에 있는 특정 단어를 다른 단어로 변환

본문 정제는 분석하는데 있어서 필수적인 과정은 아니나 사용자의 필요에 따라 원문을 정제해서 보고자 할 때 유용합니다. 본문 정제는 컴포넌트 트리의 ‘문서’에 있는 모든 문서¹에 적용됩니다. 다음은 본문 정제를 실행하는 과정입니다.

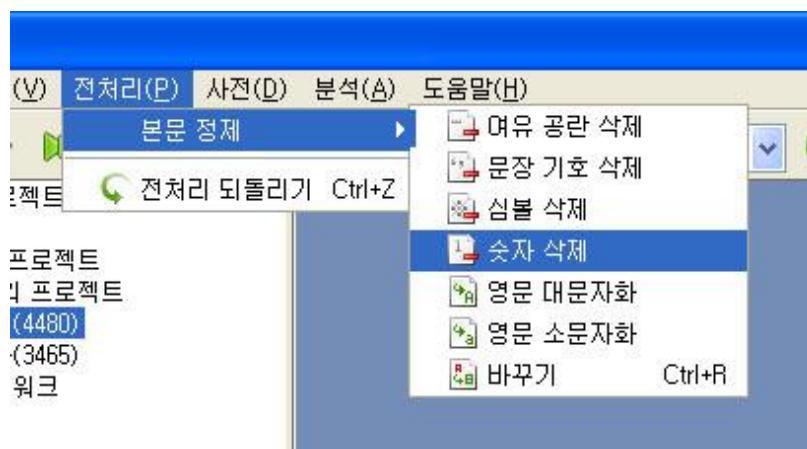
- a) 컴포넌트 트리에서 커서를 해당 프로젝트의 범위 안²에 클릭합니다.

¹ 주의할 점은 각각의 문서에 따로 적용되는 것이 아니라 ‘문서’에 있는 모든 문서에 적용된다는 것입니다.

² 프로젝트의 하위폴더(문서, 저자, 네트워크, 사전)에 커서를 두어도 됩니다. 하지만 해당 프로젝트가 아닌 다른 프로젝트에 커서를 두면 실행이 되지 않습니다.



- b) 메뉴막대의 <전처리>에서 ‘본문 정제’에 커서를 가져간 후 원하는 항목을 클릭합니다.



- c) 작업이 완료되면 결과를 확인하기 위해 ‘문서’에서 임의의 문서를 더블 클릭합니다. 작업창에 탭이 생성되면 결과를 확인할 수 있습니다.

아시아 시장이 개장할 때 금가격 상승세가 두드러집니다. 머지않아 금투기로 인한 피해를 가장 많이 보는 사람은 그런 불행을 피할 수 있었으면 좋겠네요. p.s 2 어제 연방준비은행과 유럽중앙은행 모두 출구전략 가능성에 대해 달려에 대한 불안감도 점차 설 자리를 잃고 있죠

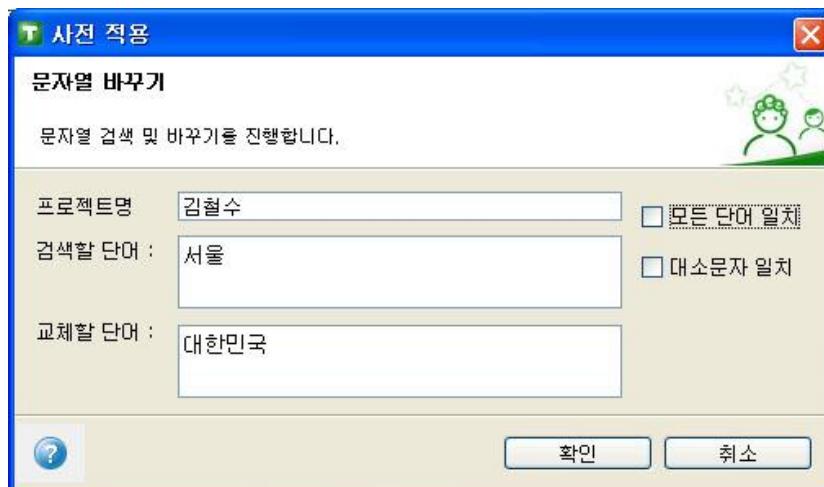
숫자 삭제를 실행시키기 전

아시아 시장이 개장할 때 금가격 상승세가 두드러집니다. 머지않아 금투기로 인한 피해를 가장 많이 보는 사람은 그런 불행을 피할 수 있었으면 좋겠네요 [p.s] 어제 연방준비은행과 유럽중앙은행 모두 출구전략 가능성에 대해 내 달러에 대한 불안감도 점차 설 자리를 잊고 있죠

숫자 삭제를 실행시킨 후

d) 명령이 성공적으로 수행되었다면 속성에는 ***** Successfully executed ***** 가 표기됩니다.

1. 바꾸기를 실행하고자 할 경우 ‘검색할 단어’에 변환할 단어를, ‘교체할 단어’에 변환 되어질 단어를 기입하고 확인을 클릭합니다.



2. 본문정제 처리결과 보기

본문 정제를 실행하였을 때의 결과는 ‘문서’에 있는 임의의 문서를 더블 클릭 한 후 작업창에 생성되는 새 탭에서 확인합니다.

a) 여유 공란 삭제

금융시장을 강타하고 있습니다. 닉터둠 교수는 유로권과 IMF에서 재정지원 할것이기 때문에, 디폴트는 잘사는 나라들인 영국, 프랑스, 독일 모두 재정적자가 늘어나고 있고, 다른 나라를 지원할 정도로 넉넉한 서 벗어난다고 하더라도 포르투갈, 아일랜드, [REDACTED] 이탈리아 같은 나라가 또 기다리고 있습니다.

여유 공란 삭제를 실행시키기 전

금융시장을 강타하고 있습니다. 닉터둠 교수는 유로권과 IMF에서 재정지원 할것이기 때문에, 디폴트는 잘사는 나라들인 영국, 프랑스, 독일 모두 재정적자가 늘어나고 있고, 다른 나라를 지원할 정도로 넉넉한 서 벗어난다고 하더라도 포르투갈, 아일랜드, 이탈리아 같은 나라가 또 기다리고 있습니다.

여유 공란 삭제를 실행시킨 후

2칸 이상의 여유 공란이 사라졌습니다.

b) 문장 기호 삭제

" 해외원정 도박으로 사회에 물의를 일으켰지만 다시 일어날 수 있도록 기회를 준 국민 여러분께 조금이나마 보답하고 어려운 이웃에게 봉사하고 싶어 행사를 시작했다" 고 말했다. 황기순은 작년까지 모금한 성금으로 전동휠체어 27대, 일반휠체어 847대를 구입해 장애인들에게 전달했다. yna.co.kr(끝) 저작권자 © 연합뉴스 무단전재·재배포금지 >

문장 기호 삭제를 실행시키기 전

amp quot 해외원정 도박으로 사회에 물의를 일으켰지만 다시 일어날 수 있도록 기회를 준 국민 여러분께 조금이나마 보답하고 어려운 이웃에게 봉사하고 싶어 행사를 시작했다 amp quot 고 말했다 황기순은 작년까지 모금한 성금으로 전동휠체어 27대 일반휠체어 847대를 구입해 장애인들에게 전달했다 engine yna co kr 끝 저작권자 © 연합뉴스 무단전재·재배포금지 amp gt;

문장 기호 삭제를 실행시킨 후

‘@’ 와 ‘.’ ‘(‘ ’) 등이 생략되었습니다.

c) 심볼 삭제

우리는 봉인가?후플러스>는 현대차의 수출용과 내수용 차량의 가격 비교를 시도했다. 현대차는 수 나타났다. 우리나라의 경제 성장과 함께 걸어온 현대차에 대해 남다른 애정을 갖고 있는 우리 국민들. 드는 이유는 무엇일까? 어떤 외압에도 굴복하지 말고, 소신껏 있는 그대로를 밝혀주시길 ★ 기대 만땅...

심볼 삭제를 실행시키기 전

우리는 봉인가?후플러스>는 현대차의 수출용과 내수용 차량의 가격 비교를 시도했다. 현대차는 수 나타났다. 우리나라의 경제 성장과 함께 걸어온 현대차에 대해 남다른 애정을 갖고 있는 우리 국민들. 드는 이유는 무엇일까? 어떤 외압에도 굴복하지 말고, 소신껏 있는 그대로를 밝혀주시길 기대 만땅...

심볼 삭제를 실행시킨 후

★ 가 생략되었습니다.

d) 숫자 삭제

아시아 시장이 개장할 때 금가격 상승세가 두드러집니다. 머지않아 금투기로 인한 피해를 가장 많이 보는 사람은 그런 불행을 피할 수 있었으면 좋겠네요. p.s 2 어제 연방준비은행과 유럽중앙은행 모두 출구전략 가능성에 대해 내 달러에 대한 불안감도 점차 설 자리를 잊고 있죠

숫자 삭제를 실행시키기 전

아시아 시장이 개장할 때 금가격 상승세가 두드러집니다. 머지않아 금투기로 인한 피해를 가장 많이 보는 사람은 그런 불행을 피할 수 있었으면 좋겠네요. p.s 어제 연방준비은행과 유럽중앙은행 모두 출구전략 가능성에 대해 내 달러에 대한 불안감도 점차 설 자리를 잊고 있죠

숫자 삭제를 실행시킨 후

단독으로 쓰인 숫자 2 가 생략되었습니다.

e) 영문 대문자화

대주교 사택 밑에서 구조됐으며, 포르토프랭스 쇼핑센터에서는 25살 여성 호텔라인 로사나가 구조됐다. 아남았던 15살 소녀를 숨지게 했다고 목격자들이 전했다. 목격자들은 경찰이 경고사격을 하다 오발사 중국 일간 데일리메일은 전했다. newglass@yna.co.kr(끝) 저작권자(c)연합뉴스 무단전재-재배포금지. >

영문 대문자를 실행시키기 전

대주교 사택 밑에서 구조됐으며, 포르토프랭스 쇼핑센터에서는 25살 여성 호텔라인 로사나가 구조됐다. 아남았던 15살 소녀를 숨지게 했다고 목격자들이 전했다. 목격자들은 경찰이 경고사격을 하다 오발사 중국 일간 데일리메일은 전했다. NEWGLASS@YNA.CO.KR(끝) 저작권자(c)연합뉴스 무단전재-재배포금지. >

영문 대문자를 실행시킨 후

영어 소문자가 대문자로 변환되었습니다.

f) 영문 소문자

사소송 1심, PD수첩 무죄' 편(서울=연합뉴스) 임은진 기자 = MBC PD 수첩'은 지난 20일 법원이 PD 수선 아래사 빈손의 사인과 관련한 어머니와의 인터뷰 자료를 내보낸다. 이 자료에서 빈손의 어머니는 "1 손의 어머니가 딸의 수술을 집도한 병원을 상대로 한 소장도 가지고 있다고 밝혔다. 이 소장에는 의사가

영문 소문자를 실행시키기 전

사소송 1심, pd수첩 무죄' 편(서울=연합뉴스) 임은진 기자 = mbc pd 수첩'은 지난 20일 법원이 pd 수선 아래사 빈손의 사인과 관련한 어머니와의 인터뷰 자료를 내보낸다. 이 자료에서 빈손의 어머니는 "1 손의 어머니가 딸의 수술을 집도한 병원을 상대로 한 소장도 가지고 있다고 밝혔다. 이 소장에는 의사가

영문 소문자화를 실행시킨 후

영어 대문자가 소문자로 변환되었습니다.

g) 바꾸기

서울 연합뉴스 임은진 기자 개그맨 황기순이 다음달 동료 연예인들과 함께 사이클을 타고 전국 각지를 돌며 장애인 돋기 성금을 모금하는 사랑 더하기 사이클대행진 을 벌인다 2000년부터 시작해 올해로 9회째를 맞는 이번 행사에서 황기순은 다음달 13 23일 서울을 출발해 수원 천안 대전 대구 부산 등 7개 도시를 돌며

바꾸기를 실행시키기 전

대한민국 연합뉴스 임은진 기자 개그맨 황기순이 다음달 동료 연예인들과 함께 사이클을 타고 전국 각지를 돌며 장애인 돋기 성금을 모금하는 사랑 더하기 사이클대행진 을 벌인다 2000년부터 시작해 올해로 9회째를 맞는 이번 행사에서 황기순은 다음달 13 23일 대한민국을 출발해 수원 천안 대전 대구 부산 등 7개 도시를

바꾸기를 실행시킨 후

‘서울’ 이 ‘대한민국’으로 변환되었습니다.

C. 자연어 처리 실행하기

1. 자연어 처리

새 프로젝트를 생성한 후 각 문장에 대한 형태소, 품사, 구문분석, NER 정보를 생성합니다. 이는 사전¹을 만들기 위해 실행하는 것으로, 프로그램이 단어를 개개로 인지할 수 있도록 문서를 세분화시킵니다. 다음은 형태소, 품사, 구문분석, NER 정보에 관한 표입니다.

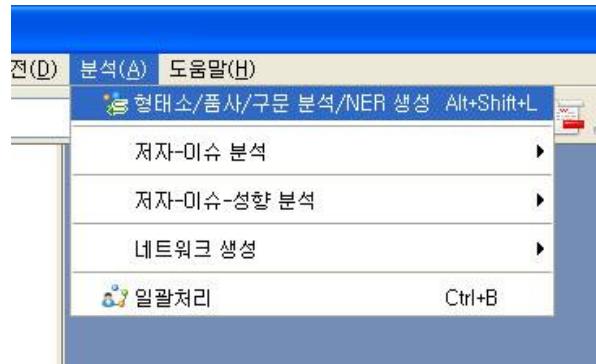
종 류	의 미	예
형태소	뜻을 가진 가장 작은 말의 단위	어미, 선어말어미 등
품사	단어를 기능, 형태, 의미에 따라 나눈 갈래	명사, 대명사, 수사, 조사, 동사, 형용사, 관형사, 부사, 감탄사
구문분석	문장의 구/절을 구분하는 것	명사구, 서술어구, 부사어구 등
NER	고유명사	인명, 지명, 국명, 책명 등

사용자의 필요에 따라 실행되는 본문 정제와는 달리, 자연어 처리는 사전을 생성하기 위한 또 하나의 절차로써 필수적으로 실행되어야만 합니다. 즉, 자연어 처리를 실행해야만 사전을 생성할 수 있으며 본문 정제는 사전을 생성하는 것과는 관련이 없습니다. 주의할 점은 자연어 처리를 실행한 후 이전 단계인 본문 정제를 실행하면 새 프로젝트를 생성한 후 적용되었던 모든 명령이 취소²된다는 것입니다. 다음은 자연어 처리를 실행하는 과정입니다.

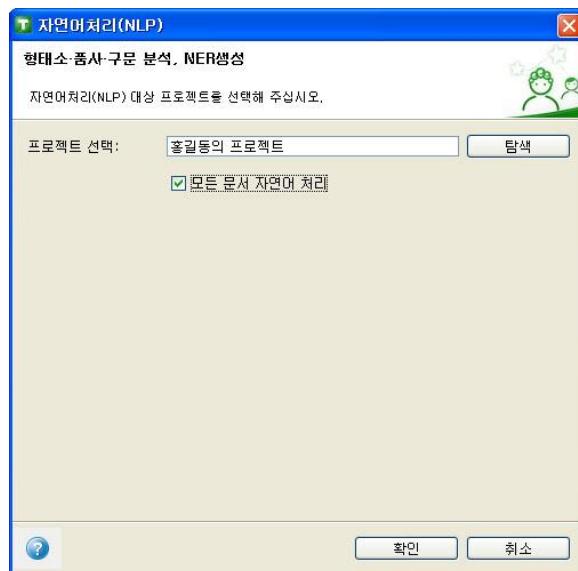
- a) 자연어 처리를 실행하려면 메뉴막대 – 분석 – 형태소/품사/구문분석/NER 생성을 누릅니다.

¹ 사전은 문서의 토픽을 추출하기 위한 일종의 도구로 쓰이며 총 3 가지 종류가 있습니다.

² 본문 정제를 실행시킨 상태에서 사전 생성으로 넘어가고 싶다면 다시 자연어 처리를 실행합니다.

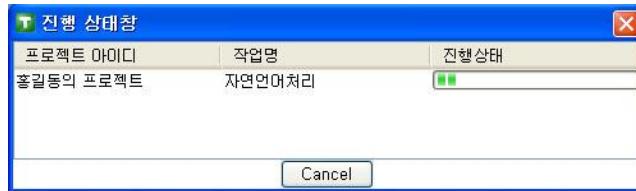


- b) ‘자연어 처리’ 대화상자가 나타나면 오른쪽의 <탐색>을 눌러 적용시키고자 하는 프로젝트를 선택한 후 <확인>을 클릭합니다. ‘모든 문서 자연어 처리’에 체크를 하지 않으면 기존에 자연어 처리가 되어있는 문서는 제외한 나머지 문서에 자연어 처리가 실행됩니다. 그러나 ‘모든 문서 자연어 처리’에 체크를 하면 기존에 자연어 처리의 여부와 상관 없이 모든 문서를 대상으로 자연어 처리가 실행됩니다.



*컴포넌트 트리에서 적용시키고자 하는 프로젝트에 이미 클릭이 되어 있을 경우, ‘프로젝트 선택’란에 별도의 탐색을 하지 않아도 자동으로 해당 프로젝트명이 나타납니다.

- c) ‘진행 상태창’ 대화상자가 나타나면 연두색 막대가 채워질 때까지 기다립니다. 이때 <Cancel>을 클릭하면 ‘자연어 처리’가 중단됩니다.



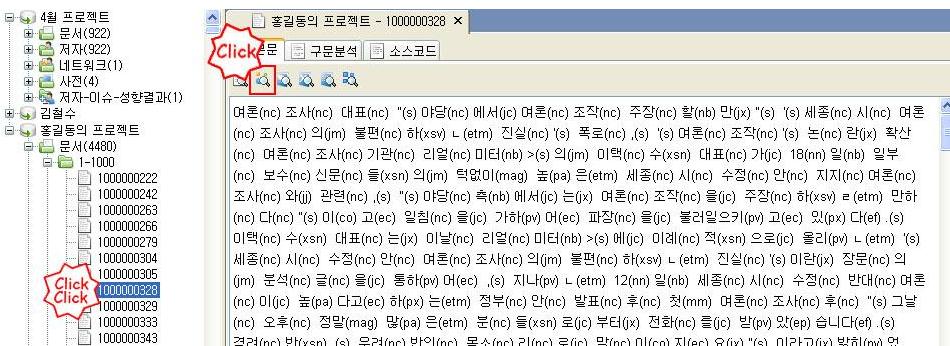
- d) 명령이 성공적으로 수행되었다면 속성에는 ***** Successfully executed ***** 가 표기됩니다.

2. 자연어 처리 처리결과보기

자연어 처리를 실행하였을 때의 결과는 ‘문서’에서 임의의 문서¹를 더블 클릭 한 후 작업창에 생성되는 새 탭에서 확인합니다.

a) 형태소와 품사가 생성된 화면

컴포넌트 트리에서 임의의 문서를 더블 클릭한 뒤 새 탭이 생성되면 하위 탭인 HTML 탭에서 형태소 분석 결과 보기 아이콘을 클릭합니다. 형태소에 대한 자세한 정보는 <부록-품사리스트, page119>에서 확인할 수 있습니다.



b) NER 이 생성된 화면

컴포넌트 트리에서 임의의 문서를 더블 클릭한 뒤 새 탭이 생성되면 HTML 탭에서 NER 보기 아이콘을 클릭합니다.

¹ 문서탭 자체에 새로고침 기능이 없기 때문에 이미 문서탭을 열어놓았던 상태이더라도 제거한 뒤 해당 문서를 클릭하여 다시 생성하여야 합니다.



D. 사전 생성하기

TONK에서 사전이란 사용자가 문서에서 특정 단어들을 수집해 놓은 것을 말합니다. 자연어 처리를 실행한 후 문서가 단어 단위로 세분화됨에 따라 사용자는 사용자가 원하는 특정 단어만 추출할 수 있게 됩니다. 이를 기반으로 사전이 만들어지면 문서에 적용시켜 사용자가 원하는 분석을 실행합니다.

TONK에서 제공하는 사전은 총 5 가지¹로, 원문에서 사용자에게 필요치 않은 단어들을 모아놓은 **삭제어 사전**, 동일한 의미를 가진 단어들을 한 단어로 통합해 놓은 **일반화 사전**, 개개의 단어에 온톨로지가 부여된 단어들을 모아놓은 **메타넷 사전**, 사용자의 주관에 따라 긍정과 부정의 의미를 가진 단어를 모아놓은 **성향 단어사전**, 사용자가 필요로 하는 단어들만 모아 놓은 **단어 사전**이 있습니다.

사전 생성 이후 실행할 수 있는 분석은 총 3²가지로 각 분석마다 요구되는 사전이 다르므로 사용자는 분석 목적에 맞게 사전을 생성하도록 합니다. 특히 이번 장<IV. 문서의 개념어간 연결 관계>에서는 ‘문서내 관계망’을 실행하기 때문에 필요한 사전은 삭제어 사전, 일반화 사전, 메타넷 사전입니다.

사전을 만드는 방법에는 사용자가 직접 단어를 추가하거나 삭제하기 위해 만드는 ‘생성’과 원문에서 단어를 추가하는 ‘사전 원문에서 생성’이 있습니다. 또한 프로그램을 통해 자동으로 생성하는 ‘자동생성’과 사용자의 필요에 따라 기존에 생성되었던 사전의 파일을 컴퓨터에서 ‘도입’하는 방법도 있습니다.

이렇게 만들어진 사전을 문서에 ‘적용’을 실행합니다. ‘적용’을 실행하는 과정에서는 사용자의 분석 목적에 따라 특별히 정해진 순서가 있습니다.

분석명	순서
문서내 관계망	삭제어 사전 적용 – 일반화 사전 적용 – 메타넷 사전 적용

주의할 사항은 사전을 만들어 놓기만 하고 ‘적용’을 실행시키지 않는다면 ‘문서내 관계망’의 결과에 반영이 되지 않기 때문에 사용자에게 필요한 사전은 꼭 ‘적용’을 실행시켜야 한다는 것입니다. 또한 3³가지 종류의 사전을 모두 생성하는 것이 필수적인 것은 아니나, 여러 가지 특성이 다른 사전을

¹ 사전의 종류는 5 가지이지만, 같은 종류의 사전을 여러 개 만들 수도 있습니다.

² 저자-이슈 분석, 저자-이슈-성향분석, 문서내 관계망 분석입니다.

³ 삭제어 사전, 일반화 사전, 메타넷 사전입니다.

문서에 적용시킴으로써 보다 더 정확성 높은 단어간의 관계를 산출할 수 있다는 것도 명심해야 합니다.
다음은 사전 정보를 요약해 놓은 표입니다.

종 류		기 능	결 과
1	삭제어 사전	원문에서 사용자에게 필요치 않은 단어를 모아둔 것	삭제어 단어가 원문에서 삭제됨
2	일반화 사전	원문에서 동일한 의미를 가진 단어를 통일된 단어로 모아둔 것	일반화 단어가 원문에서 통일됨
3	메타넷 사전	원문에서 온톨로지가 부여된 단어를 모아둔 것	온톨로지가 부가된 개념어 생성

사전이 생성되면 컴포넌트 트리의 ‘사전’에 목록이 추가 되고 속성에는 사전명, 생성일, 위치, 크기와 같은 상세정보가 나타납니다. 사전을 적용시킨 결과는 **문서—임의의 문서 – 문서 탭**에서 해당 사전의 아이콘을 눌러 확인할 수 있습니다.

사전을 만드는 것에서부터 적용시키는 과정까지 하나씩 살펴봅시다.

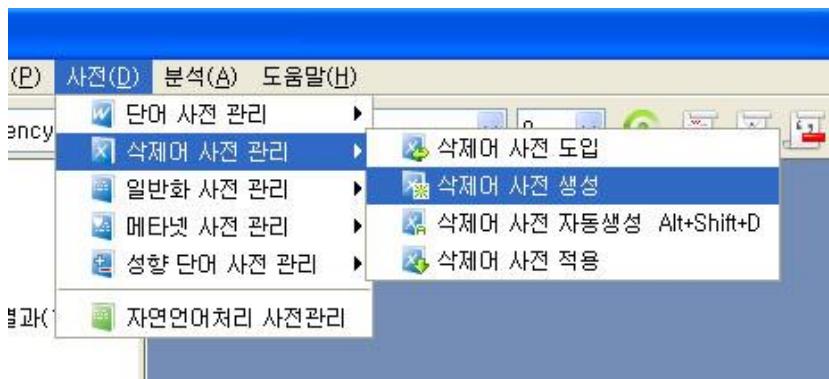
1. 삭제어 사전

삭제어 사전은 원문에서 사용자에게 필요치 않은 단어를 모아둔 사전입니다. 삭제어 사전을 만드는 방법에는 (1) 사용자가 직접 단어를 추가하거나 삭제하기 위해 초기화된 사전을 만드는 ‘삭제어 사전 생성’과 (2) 원문에서 삭제어를 추가하는 ‘삭제어 사전 원문에서 생성’ (3) 프로그램을 통해 자동으로 생성하는 ‘삭제어 자동생성’ (4) 사용자의 필요에 따라 기존에 생성되었던 사전의 파일을 컴퓨터에서 불러오는 ‘삭제어 사전 도입’이 있습니다. 삭제어 사전을 생성한 후 사용자가 이를 문서에 적용하고자 한다면 ‘삭제어 사전 적용’을 실행합니다.

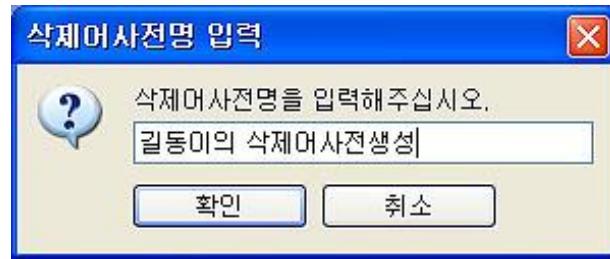
c) 삭제어 사전 생성

‘삭제어 사전 생성’은 사용자가 직접 단어를 추가/삭제하기 위하여 초기화된 삭제어 사전을 생성합니다. 다음은 ‘삭제어 사전 생성’을 실행하는 과정입니다.

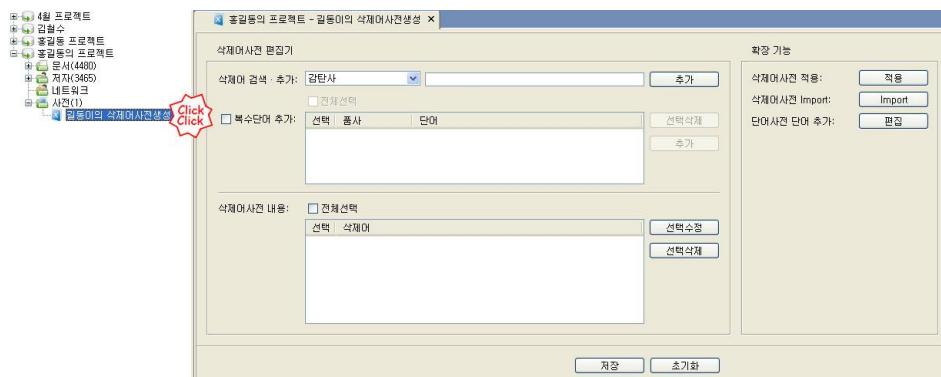
(a-1) (방법 1) ‘삭제어 사전 생성’을 실행하려면 **메뉴막대 – 사전 – 삭제어 사전 관리 – 삭제어 사전 생성**을 누릅니다. 또는 (방법 2) 컴포넌트 트리의 ‘사전’에 오른쪽 마우스를 클릭한 후 ‘삭제어 사전 생성’을 누릅니다.



(a-2) ‘삭제어 사전명 입력’ 대화상자가 나타나면 기입란에 본인이 원하는 이름을 쓰고 <확인>을 누릅니다.



(a-3) 이후 컴포넌트 트리의 ‘사전’에 삭제어 사전이 생성됩니다. 생성된 삭제어 사전을 더블 클릭하면 작업창에 삭제어 사전 편집탭이 나타나며 원하는 단어를 추가하거나 삭제할 수 있습니다.



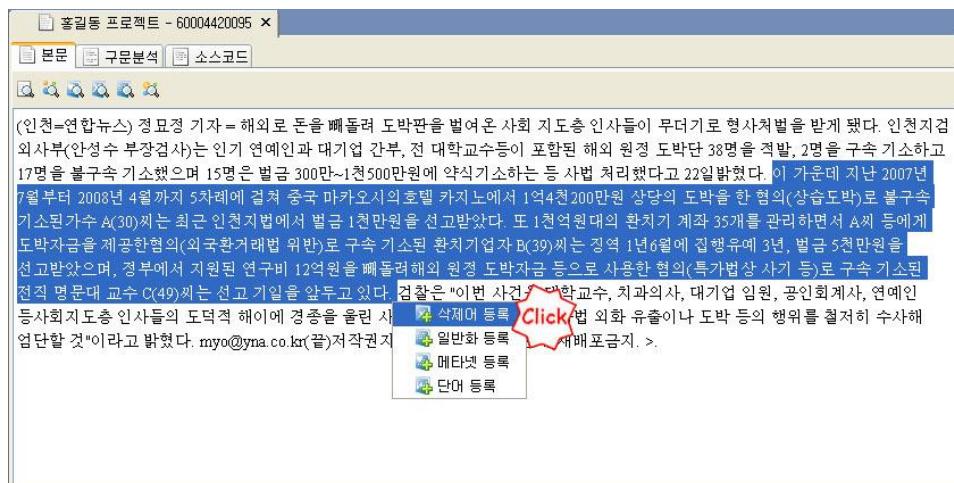
(a-4) 생성된 삭제어 사전은 파일의 형태로 사용자가 지정한 작업폴더(최초 설정- C://Workplace, page 29, (d-2) 참조)에 자동 저장되며 메시지 패널의 속성에서 저장된 위치 등 상세정보를 확인할 수 있습니다.

(a-5) 명령이 성공적으로 수행되었다면 속성에는 ******* Successfully executed ******* 가 표기됩니다.

d) 삭제어 사전 원문에서 생성

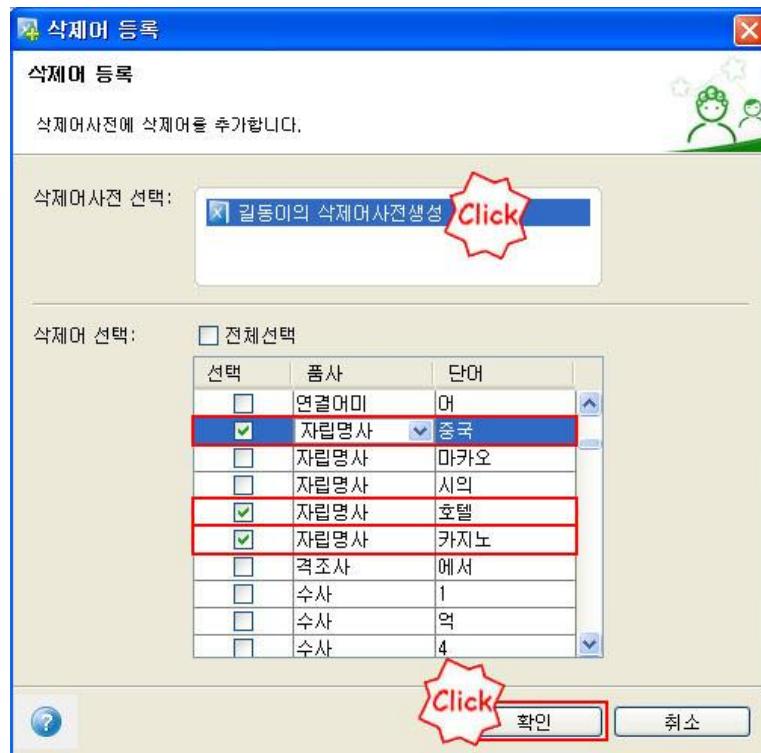
'삭제어 사전 원문에서 생성'은 원문에 드래그를 하여 단어를 추가하는 방식으로 사용자가 원문에서 직접 삭제어를 등록하여 기존에 생성된 삭제어 사전에 단어를 추가합니다. 다음은 '삭제어 사전 원문에서 생성'을 실행하는 과정입니다.

(b-1) 컴포넌트 트리 '문서'에서 추가하고자 하는 삭제어가 포함되어 있는 문서를 클릭하여 작업창에 새 탭을 생성합니다. 삭제어로 추가하고자 하는 단어나 문장에 마우스로 드래그합니다. 이후 마우스 오른쪽버튼을 클릭하여 '삭제어 등록'을 누릅니다.



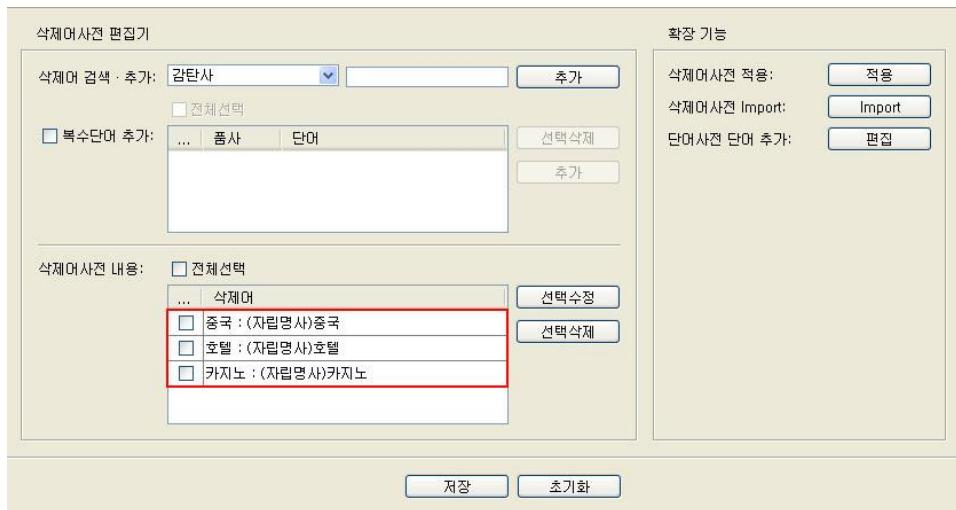
(b-2)

‘삭제어 등록’ 대화상자가 나타나면 ‘사전 선택’에서 추가하려는 삭제어 사전에 클릭을 합니다.
 ‘삭제어 선택’란은 사용자가 드래그를 한 단어/문장의 품사가 자동으로 구분이 되어 나타납니다. 추가하고자 하는 삭제어에 체크를 한 후 <확인>을 누릅니다.



자립명사인 ‘참여’를 삭제어로 선택

(b-3) 컴포넌트 트리의 ‘사전’에서 해당 삭제어 사전을 더블 클릭하여 작업창에 새 탭을 생성합니다.
 삭제어로 선택한 단어가 삭제어 사전에 등록된 것을 확인합니다.



기존의 삭제어 사전에 추가한 경우

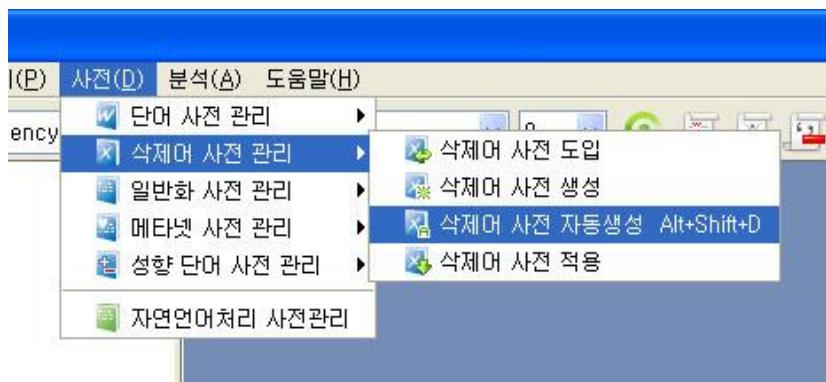
(b-4) 생성된 삭제어 사전은 파일의 형태로 사용자가 지정한 작업폴더(최초 설정- C://Workplace, page 29, (d-2) 참조)에 자동 저장되며 메시지 판넬의 속성에서 저장된 위치 등 상세정보를 확인할 수 있습니다.

(b-5) 명령이 성공적으로 수행되었다면 속성에는 ***** Successfully executed ***** 가 표기됩니다.

e) 삭제어 사전 자동생성

삭제어 사전을 프로그램을 통해 생성합니다. 다음은 ‘삭제어 사전 자동생성’을 실행하는 과정입니다.

(c-1) ‘삭제어 사전 자동 생성’을 실행하려면 메뉴막대 – 전처리 – 삭제어 사전 관리 – 삭제어 사전 자동생성을 누릅니다.



(c-2)

'삭제어 사전 자동생성' 대화상자가 나타나면 '삭제어 사전명'에는 본인이 원하는 제목을 씁니다.

'전체 품사 태그 목록'에서 분석하고자 하는 품사를 클릭한 후 추가버튼을 눌러 '대상 품사 태그 목록'으로 이동시킵니다. 만약 이동 시킨 품사를 삭제하고자 하는 경우에는 삭제버튼을 누릅니다. 하지만 사용자가 분석하기 용이하도록 이상적인 수치와 품사¹를 기본적으로 나타내주기 때문에 특정한 목적 이외에는 그대로 사용합니다.

¹ 일반적으로 명사를 중심으로 한 토픽을 얻는 것이 전체 주제를 파악하는데 용이하므로 명사관련 품사를 제외한 모든 품사가 주로 선택됩니다. 품사에 대한 보다 자세한 정보는 페이지 360쪽을 참고합니다.



- i. TF-IDF 는 한 문서 내에서 단어가 쓰인 횟수와 단어가 전체 문서 집합에서 등장하는 횟수의 역수를 곱한 것으로 수치가 높을수록 해당 단어가 특정 문서에서 자주 쓰였다는 것을 나타냅니다. 즉, TF-IDF 값이 높을수록 그 단어는 특정 이슈에서 자주 쓰였을 가능성이 높습니다. 주의할 사항은 삭제어 사전에서의 TF-IDF 값을 적용하는 방식은 사용자가 지정한 범위를 제외한 나머지 부분에서 적용된다는 것입니다. 다시 말하면, 지정한 수치의 이외의 값을 가진 단어들이 삭제됩니다. 아래의 붉은 막대로 표현된 부분이 사용자가 지정한 값의 이외의 부분으로 삭제어 사전에 추가됩니다. 만약 범위를 0.0 과 1.0 을 지정한다면 삭제어는 존재하지 않습니다.



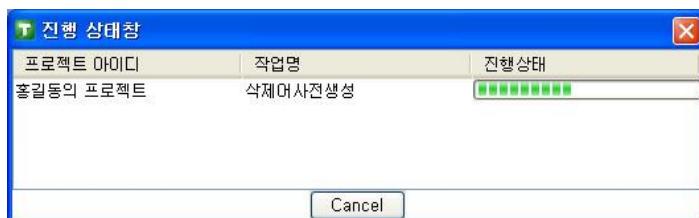
- ii. 한 단어가 등장하는 문서의 수를 나타냅니다. DF 값이 높으면 아주 일반적인 단어를 의미하는 것이고 DF 값이 낮으면 매우 특수한 경우의 단어를 의미합니다. 일반적으로

우리가 관심 있어하는 단어는 중간 정도의 값을 지닌 DF입니다. DF 값이 매우 높은 경우는 -은, -는, -이, -가와 같이 문서의 성향과 주제에 상관없이 자주 들어가는 단어를 의미하게 되고, DF 값이 매우 낮은 경우에는 오·탈자가 섞인 단어, 혹은 매우 특수한 경우에만 쓰이는 단어가 선별되게 됩니다. **그러므로 삭제어 사전에서는 중간 정도의 범위를 벗어난 DF 값을 가진 단어들로 생성되는 것이 중요합니다.** 다시 말하면, DF 값을 적용하는 방식은 사용자가 지정한 범위를 제외한 나머지 부분에서 적용됩니다. 아래의 붉은 막대로 표현된 부분이 사용자가 지정한 값의 이외의 부분으로 삭제어 사전에 추가됩니다.



- iii. 최대길이는 단어의 최대 길이입니다.
- iv. 출현빈도수는 전체 문서 집합에서 출현한 횟수입니다.

(c-3) ‘진행 상태창’ 대화상자가 나타나면 연두색 막대가 채워질 때까지 기다립니다. 이때 <Cancel>을 클릭하면 ‘삭제어 사전 자동생성’이 중단됩니다.



(c-4) 작업이 완료되면 컴포넌트 트리의 ‘사전’에 삭제어 사전이 생성됩니다. 생성된 삭제어 사전을 더블 클릭하면 작업창에 삭제어 사전 편집 탭이 나타나며 프로그램에서 추출한 형태소 이외에 사용자가 원하는 단어를 추가하거나 삭제할 수 있습니다.



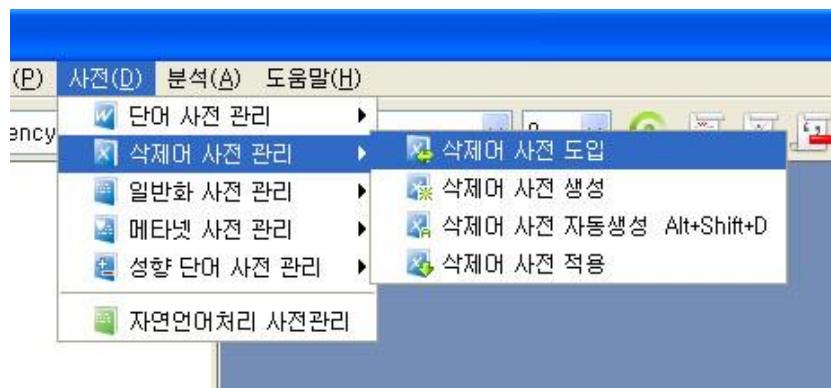
(c-5) 생성된 삭제여 사전은 파일의 형태로 사용자가 지정한 작업폴더 (최초 설정- C://Workplace, page 29, (d-2) 참조)에 자동 저장되며 메시지 판넬의 속성에서 저장된 위치 등 상세정보를 확인할 수 있습니다.

(c-6) 명령이 성공적으로 수행되었다면 속성에는 ***** Successfully executed ***** 가 표기됩니다.

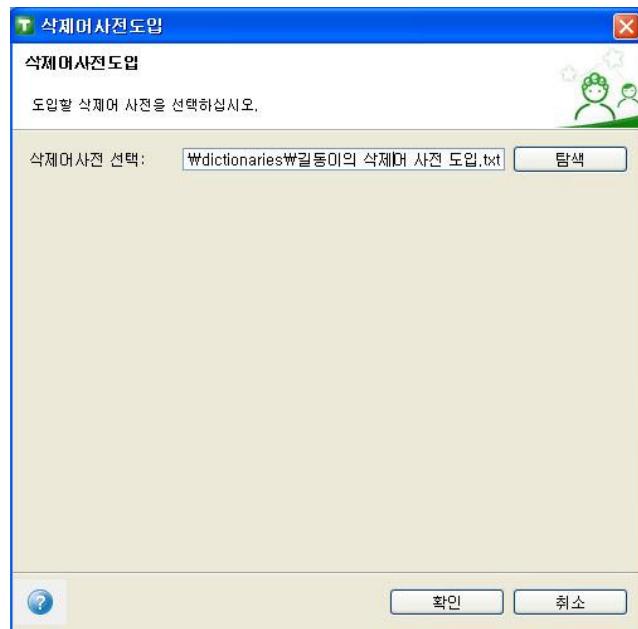
f) 삭제여 사전 도입

기존에 생성되었던 삭제여 사전의 파일을 컴퓨터에서 불러옵니다. 다음은 ‘삭제여 사전 도입’을 실행하는 과정입니다.

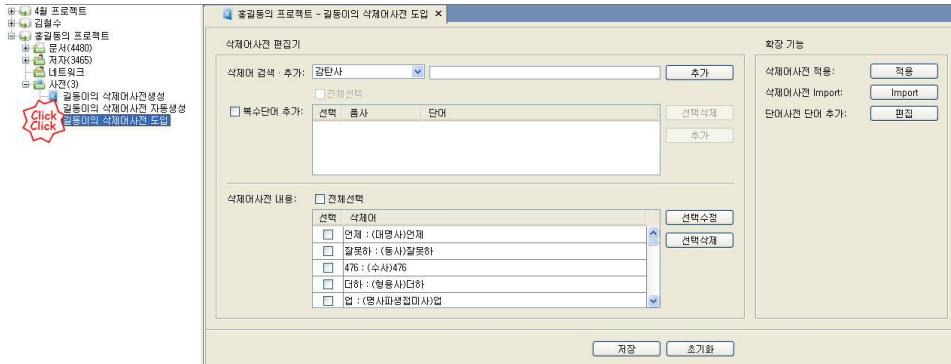
(d-1) ‘삭제여 사전을 도입’을 실행하려면 **메뉴막대 – 사전 – 삭제여 사전 관리 – 삭제여 사전 도입**을 누릅니다.



(d-2) ‘삭제어 사전 도입’ 대화상자가 나타나면 <탐색>을 눌러 ‘사전’로 불러올 파일을 선택한 후 <확인>을 클릭합니다.



(d-3) 이후 컴포넌트 트리의 ‘사전’에 삭제어 사전이 생성됩니다. 불러온 삭제어 사전을 더블 클릭하면 작업창에 삭제어 사전 편집 탭이 나타나며 사용자가 원하는 단어를 추가하거나 삭제할 수 있습니다.

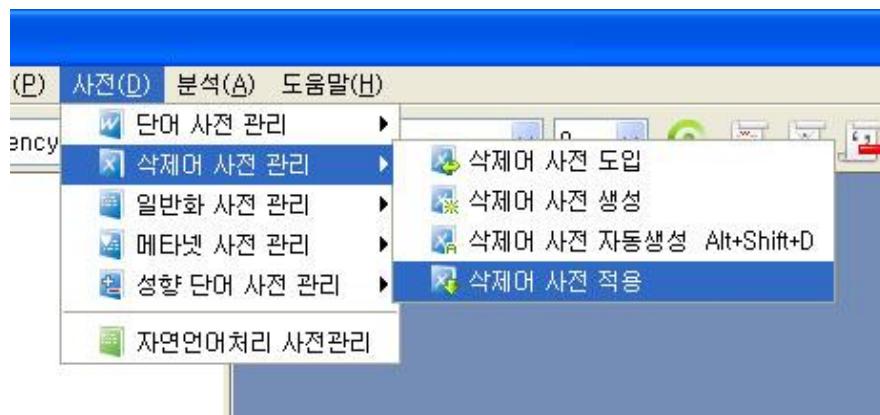


(d-4) 명령이 성공적으로 수행되었다면 속성에는 ***** Successfully executed ***** 가 표기됩니다.

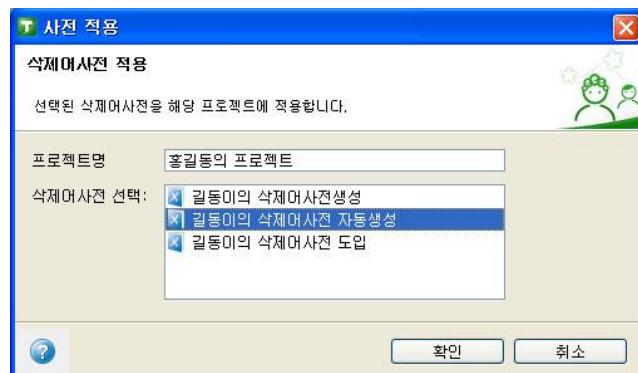
e) 삭제어 사전 적용

생성한 삭제어 사전을 문서에 적용합니다. 다음은 ‘삭제어 사전 적용’을 실행시키는 과정입니다.

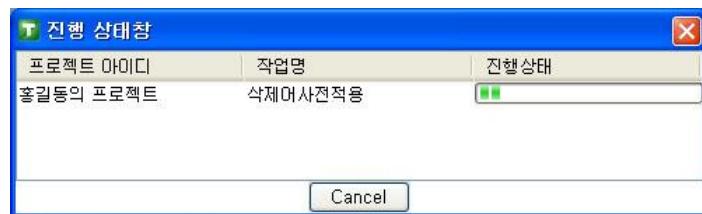
(e-1) ‘삭제어 사전 적용’을 실행하려면 메뉴 막대 – 사전 – 삭제어 사전 관리 – 삭제어 사전 적용을 누릅니다.



(e-2) ‘사전 적용’ 대화상자가 나타나면 프로젝트에 적용시킬 삭제어 사전을 선택한 후 <확인>을 클릭합니다.



(e-3) ‘진행 상태창’ 대화상자가 나타나면 연두색 막대가 채워질 때까지 기다립니다. 이때 <Cancel>을 클릭하면 ‘삭제어 사전 적용’이 중단됩니다.



(e-4) 명령이 성공적으로 수행되었다면 속성에는 ******* Successfully executed ******* 가 표기됩니다.

2. 삭제어 사전 처리결과 보기

삭제어 사전을 적용시킨 결과를 확인하려면 **컴포넌트 트리—문서**에서 임의의 문서¹를 더블 클릭합니다. 좌측에 새 탭이 생성되면 **HTML 탭 — 삭제어 보기 아이콘**을 클릭하여 삭제어가 빨간색으로 표시되어 있고 취소선이 그어져 있는 것을 확인합니다.



‘사전’에서 삭제어 사전을 더블 클릭하면 작업창에 해당 사전을 편집할 수 있는 탭이 나타나며 프로그램에서 추출한 단어 이외에 사용자가 원하는 단어를 추가하거나 삭제할 수 있습니다.

a) 삭제어사전

컴포넌트 트리의 ‘사전’에 있는 임의의 삭제어 사전을 클릭했을 때 작업창은 다음과 같습니다.

¹ 사전의 ‘적용’은 ‘문서’에 있는 모든 문서에 적용이 되기 때문에 임의의 문서에서 열람할 수 있습니다.



탭의 상단에는 프로젝트 명과 사전종류가 표기되며 하위 메뉴에는 ‘삭제어사전 편집기’과 ‘확장 기능’ 총 2 가지가 있습니다.

- 저장 저장: 편집된 내용을 저장합니다.
- 초기화 초기화: 명령이 실행되지 않은 초기의 상태로 돌아갑니다.

(a-1) 삭제어 사전 편집기

i. 삭제어 검색 추가

삭제어를 검색하고자 한다면 선택란에 해당하는 품사를 선택한 후 우측란에 찾고자 하는 삭제어를 기입하면 하단에 조건에 일치하는 검색어가 나타납니다.

삭제어를 추가하고자 한다면 선택란에서 해당하는 품사를 선택한 후 우측란에 추가하고자 하는 삭제어를 기입합니다. 이후 <추가>를 누르면 하단에 조건에 일치하는 삭제어가 생성됩니다.

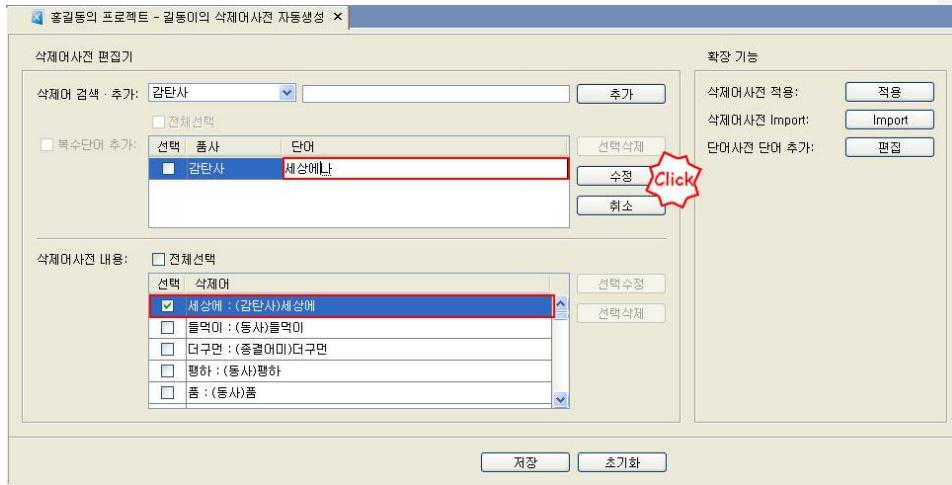


고유명사를 선택하고 ‘홍길동’을 추가 했을 때

복수 단어를 삭제어로 추가하고자 한다면 복수단어 추가란에 체크를 한 후 <추가>를 누릅니다.

ii. 삭제어사전 내용

삭제어 사전에 등록되어 있는 삭제어들이 나타납니다. 임의의 삭제어의 선택란에 체크를 하고 <선택삭제>를 클릭하면 삭제어 사전에서 삭제됩니다. 임의의 삭제어의 선택란에 체크를 하고 <선택수정>을 클릭하면 삭제어 검색 추가로 이동을 하게 되며 해당 단어에 더블클릭을 하여 단어를 수정합니다.



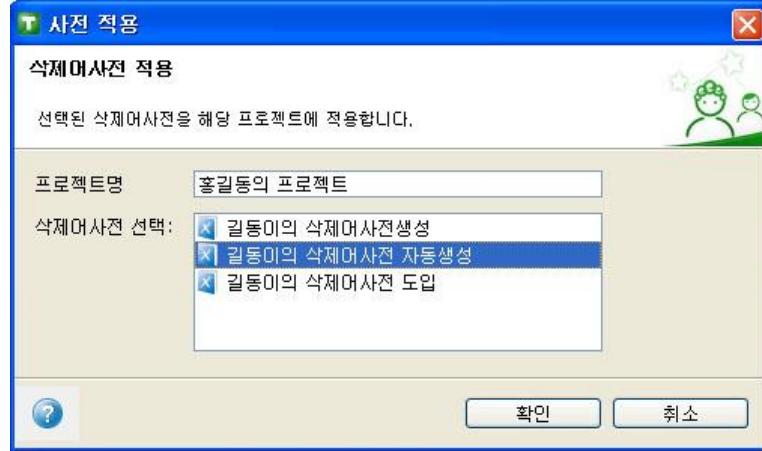
(a-2) 확장 기능



i. 삭제어 사전 적용

삭제어 사전을 프로젝트에 적용합니다. <적용>을 누르면 아래와 같은 대화상자가 나타납니다.

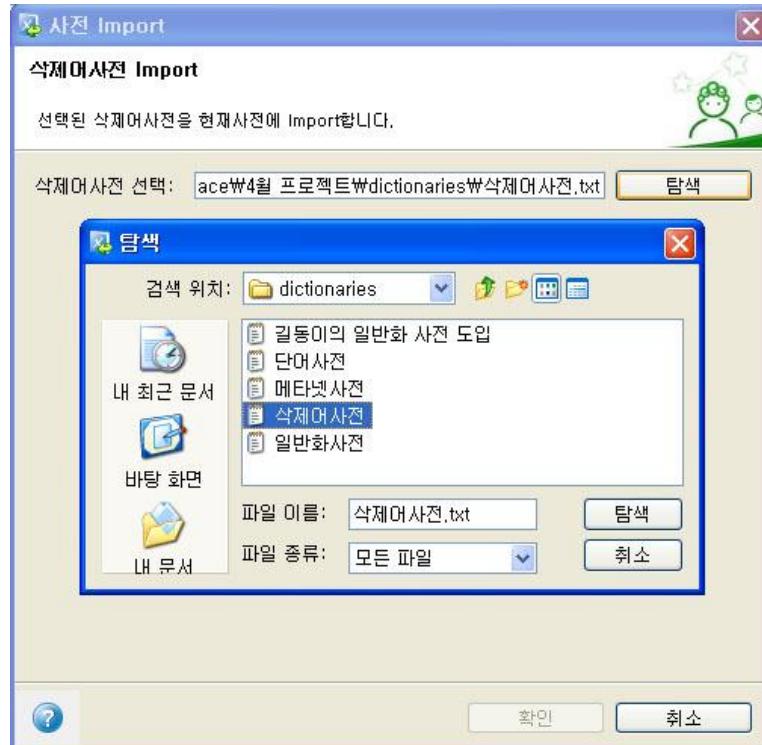
삭제어 사전 선택란에서 적용하고자 하는 삭제어 사전을 선택한 후 <확인>을 클릭합니다.



'길동이의 삭제어사전 자동생성'을 적용하고자 할 때

ii. 삭제어 사전 Import

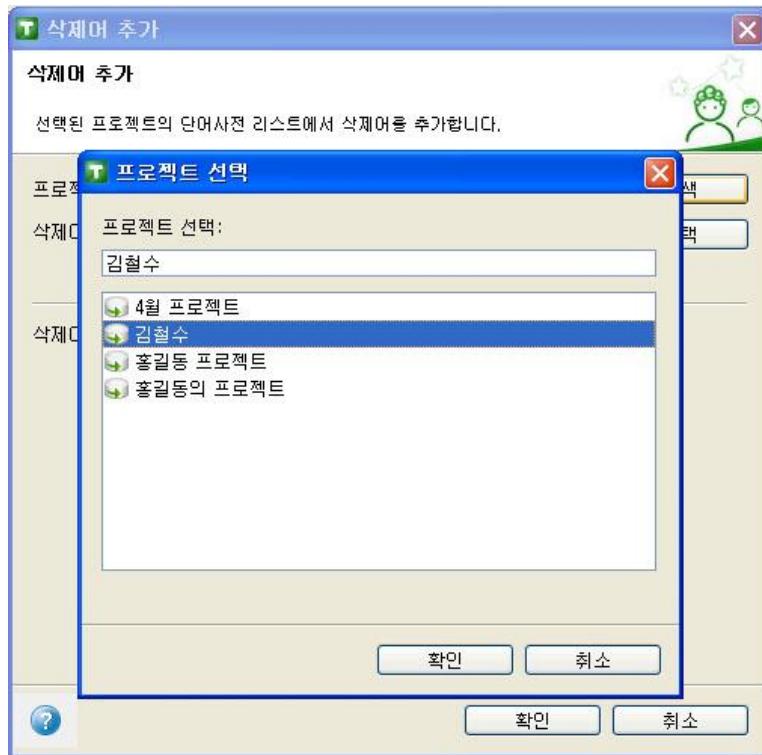
기존에 생성된 삭제어 사전을 불러옵니다.



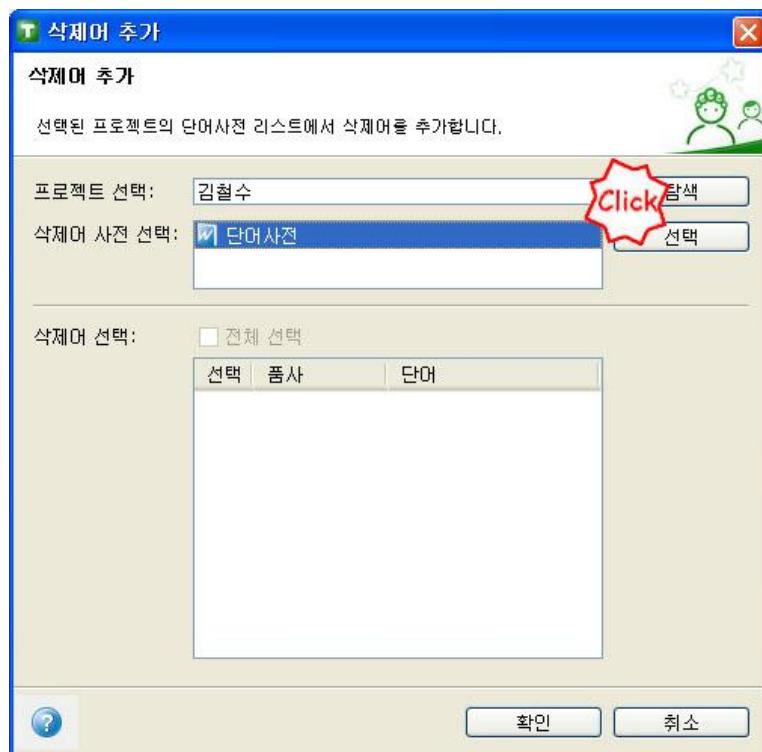
iii. 단어 사전 단어 추가

다양한 프로젝트의 단어 사전으로부터 사용자가 원하는 단어를 선택해 삭제어 사전에 추가합니다. 주의할 점은 선택한 프로젝트에 생성된 단어사전이 반드시 있어야 합니다. 다음은 단어 사전의 단어를 삭제어 사전에 추가하는 과정입니다.

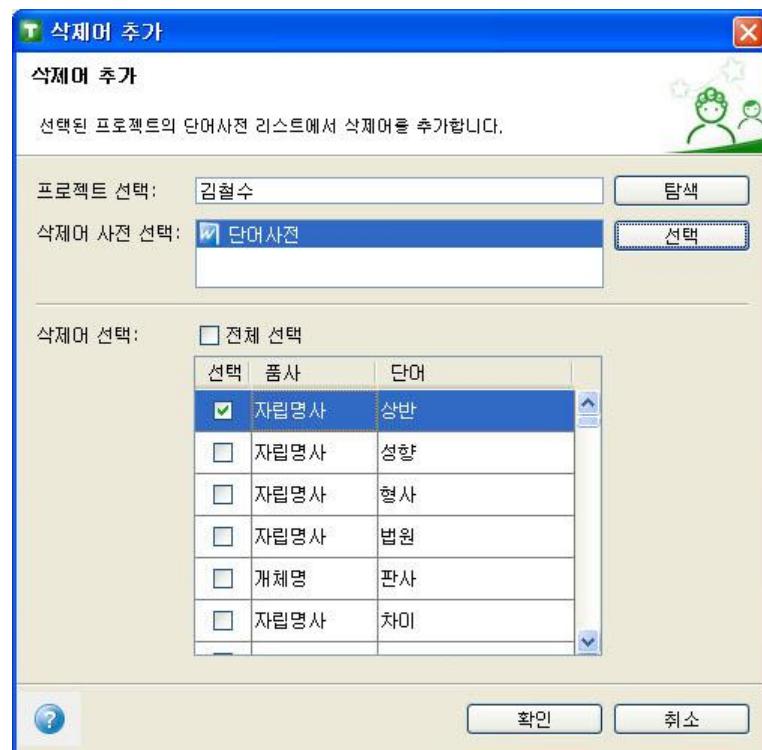
(iii-1) <편집>을 클릭합니다. ‘삭제어 추가’ 대화상자가 나타나면 프로젝트 선택란에 불러오고자 하는 단어 사전이 속해있는 프로젝트를 <탐색>을 눌러 선택한 후 <확인>을 누릅니다.



(iii-2) 사전 선택란에는 불러온 프로젝트가 가지고 있는 단어 사전 목록이 나타납니다. 원하는 단어 사전을 클릭하고 <선택>을 누릅니다.



(iii-3) 삭제어 선택란에는 선택한 단어 사전의 단어들이 나열됩니다. 사용자가 추가하고자 하는 단어에 체크표시를 하고 <확인>을 누릅니다.



자립명사인 ‘상반’을 삭제어 사전에 추가

(iv-4) 속성의 ‘단어수’에서 단어가 추가된 것을 확인할 수 있습니다.

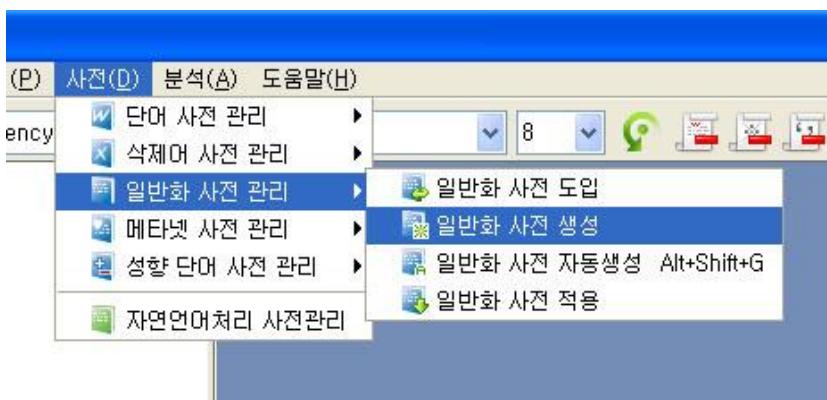
3. 일반화 사전

일반화 사전은 동일한 의미를 가진 단어군을 하나의 단어로 나타내기 위하여 사용자가 지정한 단어와 동일 의미의 단어군의 쌍을 모아놓은 사전입니다. 일반화 사전을 만드는 방법에는 (1) 사용자가 직접 단어를 추가하거나 삭제하기 위해 초기화된 사전을 만드는 ‘일반화 사전 생성’과 (2) 원문에서 일반 어를 추가하는 ‘일반화 사전 원문에서 생성’(3) 프로그램을 통해 자동으로 생성하는 ‘일반화 사전 자동생성’ (4) 사용자의 필요에 따라 기존에 생성되었던 사전의 파일을 컴퓨터에서 불러오는 ‘일반화 사전 도입’이 있습니다. 일반화 사전을 생성한 후 사용자가 이를 문서에 적용시키고자 한다면 ‘일반화 사전 적용’을 실행시킵니다.

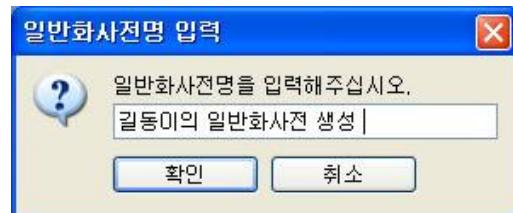
a) 일반화 사전 생성

‘일반화 사전 생성’은 사용자가 직접 단어를 추가/삭제하기 위하여 초기화된 일반화 사전을 생성합니다. 다음은 ‘일반화 사전 생성’을 실행하는 과정입니다.

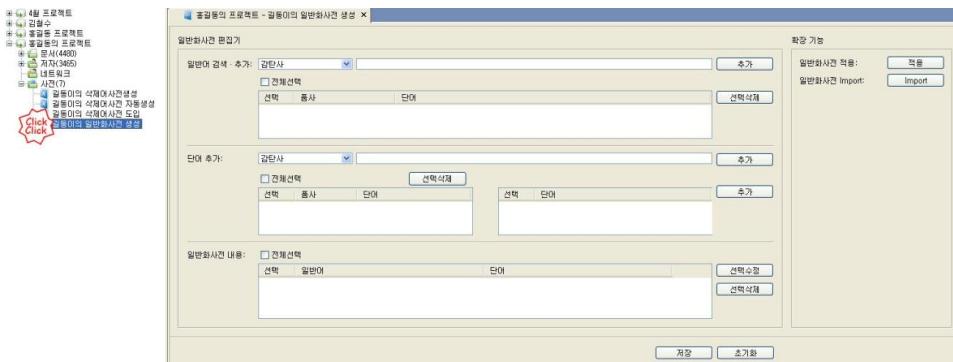
(a-1) ‘일반화 사전 생성’을 실행하려면 (방법 1) **메뉴 막대 – 사전 – 일반화 사전 관리 – 일반화 사전 생성**을 누릅니다. 또는 (방법 2) 컴포넌트 트리의 ‘사전’에 오른쪽 마우스를 클릭한 후 ‘일반화 사전 생성’을 누릅니다.



(a-2) ‘일반화 사전명 입력’ 대화상자가 나타나면 기입란에 본인이 원하는 이름을 쓰고 <확인>을 누릅니다.



(a-3) 이후 컴포넌트 트리의 ‘사전’에 일반화 사전이 생성됩니다. 생성된 일반화 사전을 더블 클릭하면 작업창에 일반화 사전 편집 탭이 나타나며 원하는 단어를 추가하거나 삭제할 수 있습니다.



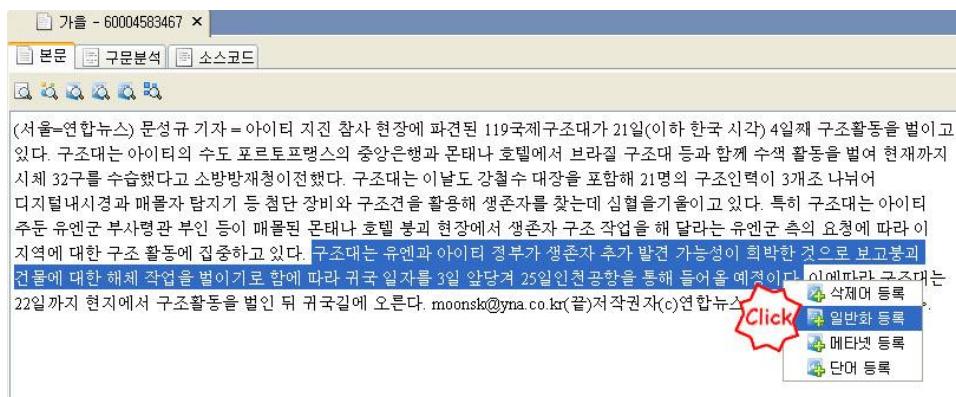
(a-4) 생성된 일반화 사전은 파일의 형태로 사용자가 지정한 작업폴더(최초 설정- C://Workplace, page 29, (d-2) 참조)에 자동 저장되며 메시지 패널의 속성에서 저장된 위치 등 상세정보를 확인할 수 있습니다.

(a-5) 명령이 성공적으로 수행되었다면 속성에는 ***** Successfully executed ***** 가 표기됩니다.

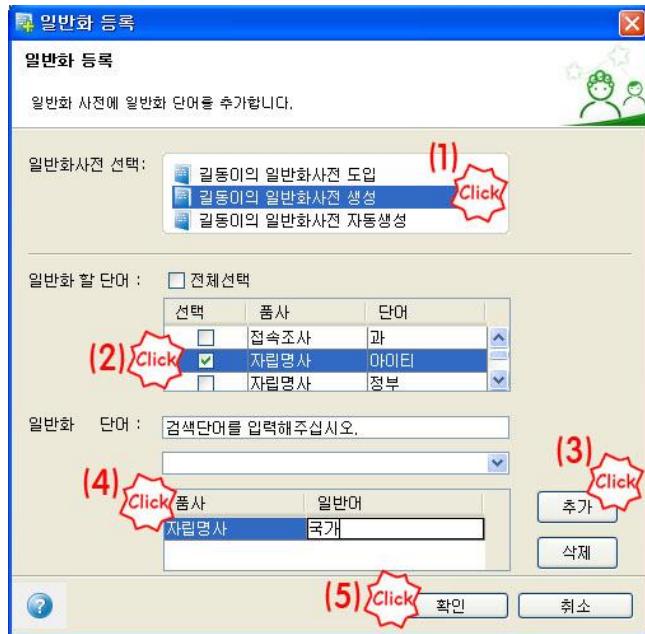
b) 일반화 사전 원문에서 생성

'일반화 사전 원문에서 생성'은 원문에 드래그를 하여 단어를 추가하는 방식으로 사용자가 원문에서 직접 일반어를 등록하여 사전을 생성하거나 기존에 생성된 일반화 사전에 일반어와 단어의 쌍을 추가합니다. 다음은 '일반화 사전 원문에서 생성'을 실행하는 과정입니다.

(b-1) 컴포넌트 트리 '문서'에서 추가하고자 하는 일반어가 포함되어 있는 문서를 작업창에 새 탭을 생성합니다. 일반어로 추가하고자 하는 단어나 문장에 마우스로 드래그합니다. 이후 마우스 오른쪽버튼을 클릭하여 '일반어 등록'을 누릅니다.



(b-2) '일반화 등록' 대화상자 가 나타나면 '사전 선택'에서 추가하려는 일반화 사전을 클릭합니다. '일반화 할 단어'란은 사용자가 드래그를 한 단어/문장의 품사가 자동으로 구분이 되어 나타납니다. '일반화 할 단어'란에 추가하고자 하는 단어를 체크한 후 오른쪽 하단의 <추가>버튼을 누릅니다.



자립명사인 ‘아이티’를 일반어로 선택

‘일반화 단어’의 ‘품사’란에 일반어의 해당 품사를 선택한 후 ‘일반어’란에 일반어를 기입한 후 Enter 를 누른 후 <확인>을 클릭합니다.

(b-3) 컴포넌트 트리의 ‘사전’에서 해당 일반화 사전을 더블 클릭하여 작업창에 새 탭을 생성합니다. 일반어로 선택한 단어가 일반화 사전에 등록된 것을 확인합니다.

일반화사전 내용:		
선택	일반어	단어
<input type="checkbox"/>	국제기구 : (자립명사)국제기구	유엔 : (자립명사)유엔
<input checked="" type="checkbox"/>	국가 : (자립명사)국가	아이티 : (자립명사)아이티

기존의 삭제어 사전에 추가한 경우

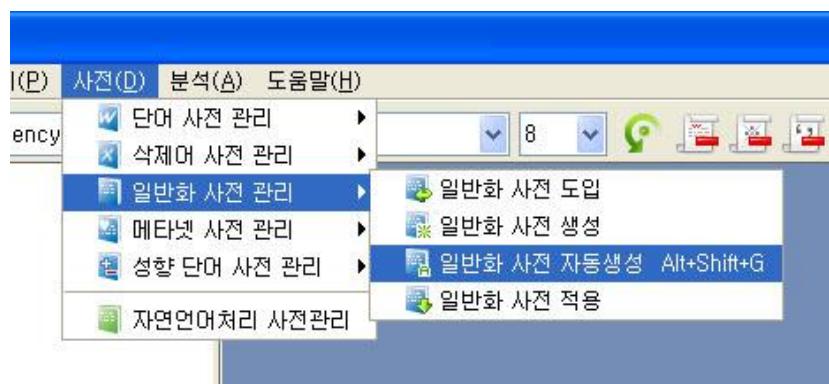
(b-4) 생성된 일반화 사전은 파일의 형태로 사용자가 지정한 작업폴더(최초 설정- C://Workplace, page 29, (d-2) 참조)에 자동 저장되며 메시지 판넬의 속성에서 저장된 위치 등 상세정보를 확인할 수 있습니다.

(b-5) 명령이 성공적으로 수행되었다면 속성에는 ******* Successfully executed ******* 가 표기됩니다.

c) 일반화 사전 자동생성

일반화 사전을 프로그램을 통해 생성합니다. 다음은 ‘일반화 사전 자동생성’을 실행하는 과정입니다.

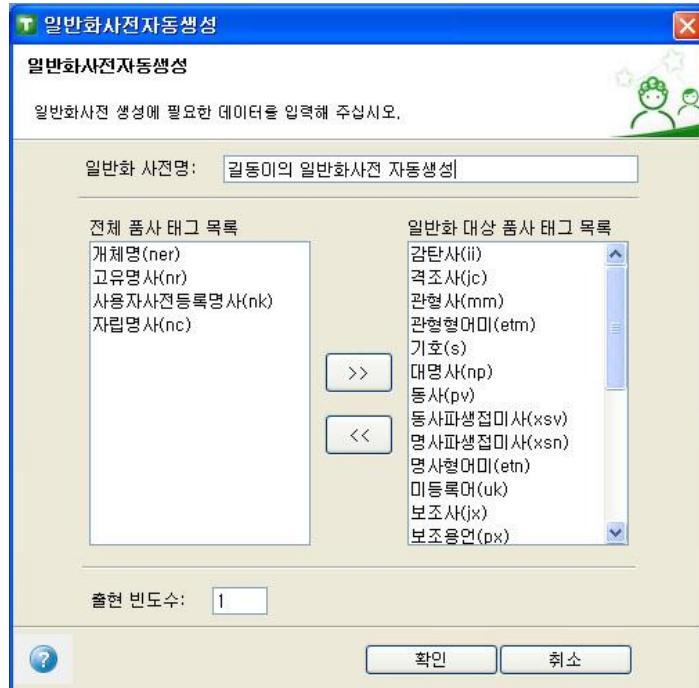
(c-1) ‘일반화 사전 생성’을 실행하려면 **메뉴막대 – 사전 – 일반화 사전 관리 – 일반화 사전 자동생성을** 누릅니다.



(c-2)

- ‘일반화 자동생성’ 대화상자가 나타나면 ‘일반화 사전명’에는 본인이 원하는 제목을 씁니다.
- ‘전체 품사 태그 목록’에서 분석하고자 하는 품사를 클릭한 후 추가버튼을 눌러 ‘대상 품사 태그 목록’으로 이동시킵니다. 만약 이동 시킨 품사를 삭제하고자 하는

경우에는  삭제버튼을 누릅니다. 하지만 사용자가 분석하기 용이하도록 이상적인 수치와 품사를 기본적으로 나타내주기 때문에 특정한 목적 이외에는 그대로 사용합니다.

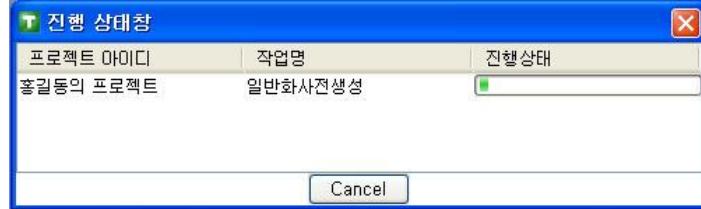


iii. 출현 빈도수는 문서 집합에서 출현한 횟수입니다.

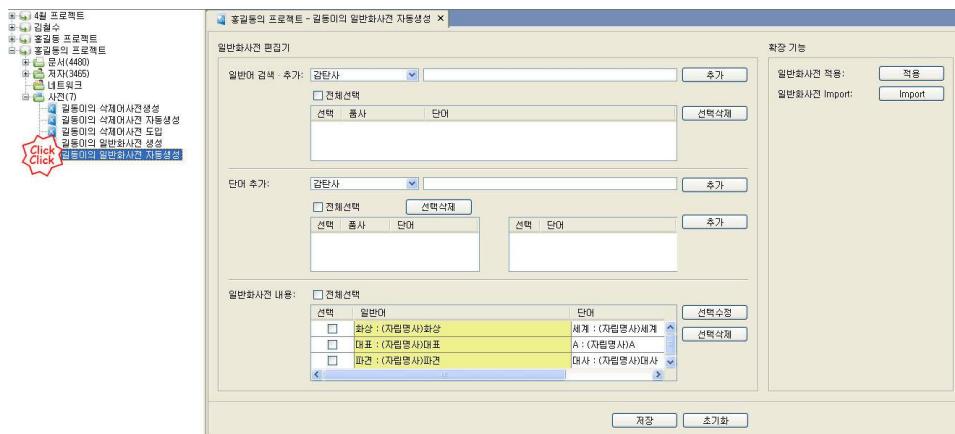
동시 출현 빈도수가 높을수록 두 단어 사이의 연관관계가 강해지므로 한 단어가 다른 단어의 동일단어일 가능성도 높아집니다. 설정이 끝나면 확인 버튼을 클릭합니다.

(c-3) ‘진행 상태창’ 대화상자가 나타나면 연두색 막대가 채워질 때까지 기다립니다. 이때 <Cancel>을 클릭하면 ‘일반화 사전 생성’이 중단됩니다.

¹ 일반적으로 명사를 중심으로 한 토픽을 얻는 것이 전체 주제를 파악하는데 용이하므로 명사관련 품사를 제외한 모든 품사가 주로 선택됩니다. 품사에 대한 보다 자세한 정보는 페이지 360 쪽을 참고합니다.



(c-4) 작업이 완료되면 컴포넌트 트리의 ‘사전’에 일반화 사전이 생성됩니다. 생성된 일반화 사전을 더블 클릭하면 작업창에 일반화 사전 편집 탭이 나타나며 프로그램에서 추출한 단어 이외에 사용자가 원하는 단어를 추가하거나 삭제할 수 있습니다.



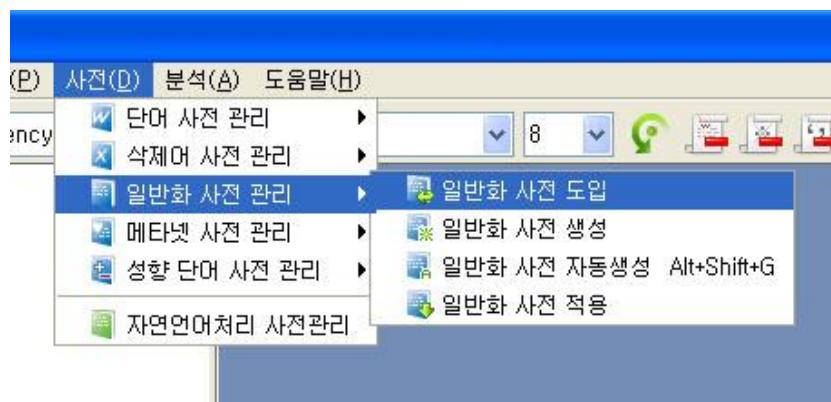
(c-5) 생성된 일반화 사전은 파일의 형태로 사용자가 지정한 작업폴더(최초 설정-C://Workplace, page 29, (d-2) 참조)에 자동 저장되며 메시지 판넬의 속성에서 저장된 위치 등 상세정보를 확인할 수 있습니다.

(c-6) 명령이 성공적으로 수행되었다면 속성에는 ***** Successfully executed ***** 가 표기됩니다.

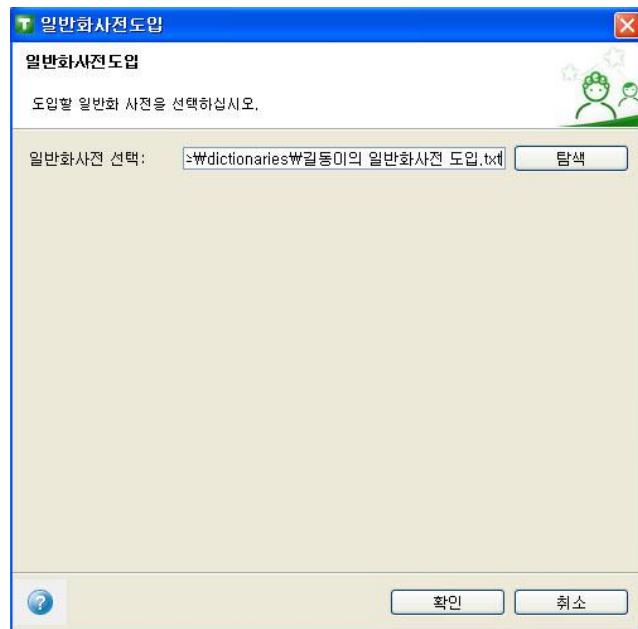
d) 일반화 사전 도입

기존에 생성되었던 일반화 사전의 파일을 컴퓨터에서 불러옵니다. 다음은 ‘일반화 사전 도입’을 실행하는 과정입니다.

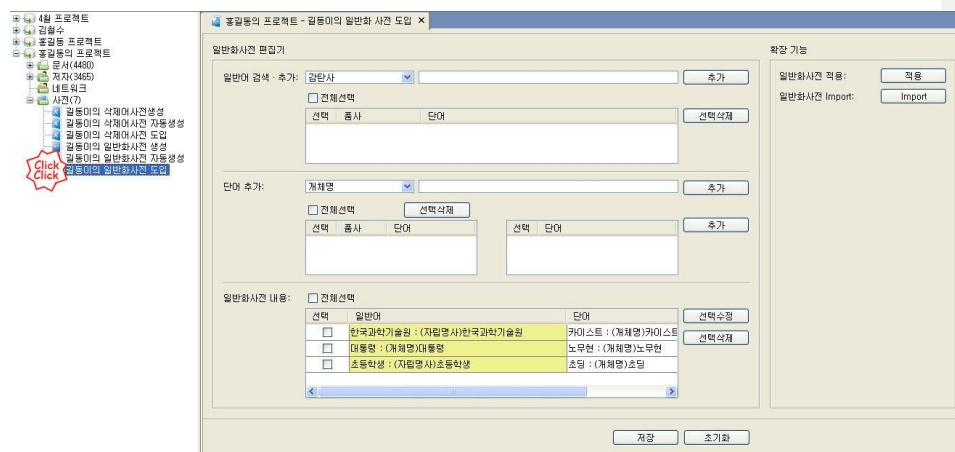
(d-1) ‘일반화 사전 도입’을 실행하려면 **메뉴막대 – 사전 – 일반화 사전 관리 – 일반화 사전 도입**을 누릅니다.



(d-2) ‘일반화 사전 도입’ 대화상자가 나타나면 <탐색>을 눌러 ‘사전’로 불러올 파일을 선택한 후 <확인>을 클릭합니다.



(d-3) 이후 컴포넌트 트리의 ‘사전’에 일반화 사전이 생성됩니다. 불러온 일반화 사전을 더블 클릭하면 작업창에 일반화 사전 편집 탭이 나타나며 사용자가 원하는 단어를 추가하거나 삭제할 수 있습니다.

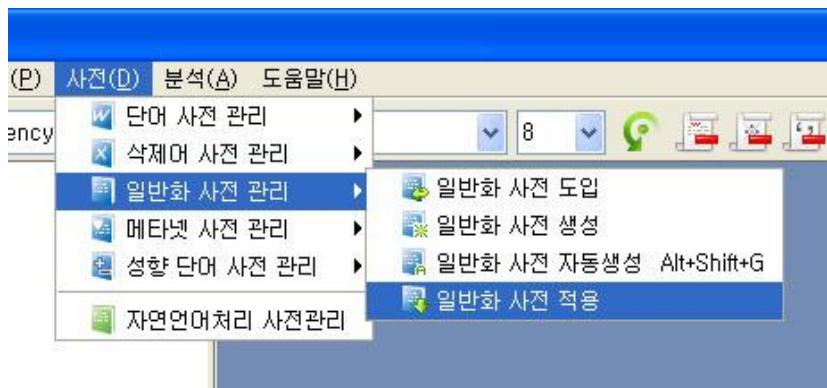


(d-4) 명령이 성공적으로 수행되었다면 속성에는 ***** Successfully executed ***** 가 표기됩니다.

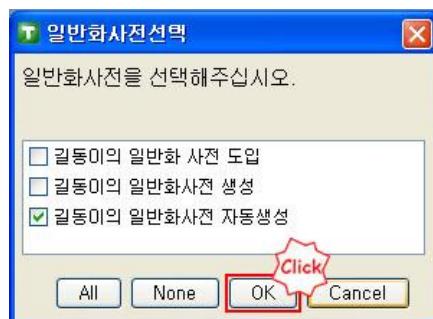
e) 일반화 사전 적용

생성한 일반화 사전을 프로젝트에 적용합니다. 다음은 ‘일반화 사전 적용’을 실행하는 과정입니다.

(e-1) ‘일반화 사전 적용’을 실행하려면 메뉴 막대 – 사전 – 삭제어 사전 관리 – 삭제어 사전 적용을 누릅니다.

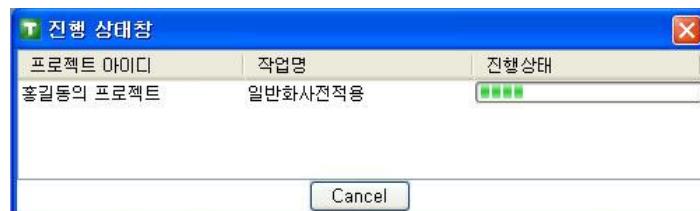


(e-2) ‘일반화 사전 선택’ 대화상자가 나타나면 프로젝트에 적용시킬 일반화 사전을 선택한 후 <OK> 를 클릭합니다.



- i. All : 목록에 있는 사전을 모두 선택합니다.
- ii. None : 체크 표시가 되어 있는 항목을 해제합니다.
- iii. OK : 선택을 완료하고 다음 단계로 넘어갑니다.
- iv. Cancel : ‘일반화 사전 선택’ 대화상자를 취소합니다.

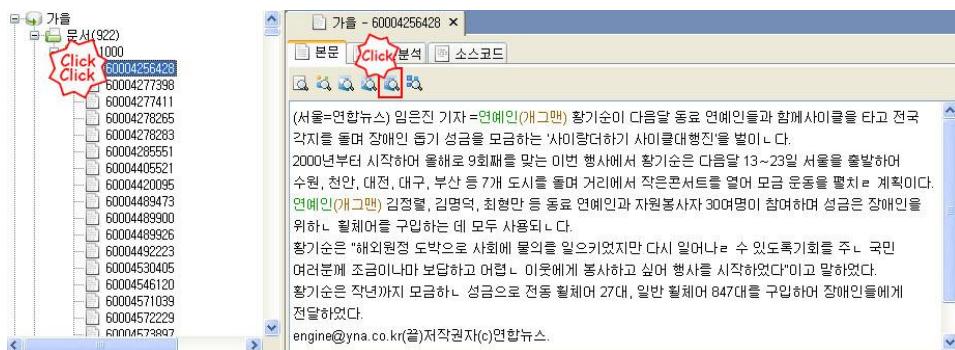
(e-3) ‘진행’ 대화상자가 나타나면 막대가 모두 100%에 도달할 때까지 기다립니다. 이때 <Cancel> 을 클릭하면 ‘일반화 사전 적용’이 중단됩니다.



(e-4) 명령이 성공적으로 수행되었다면 속성에는 ***** Successfully executed ***** 가 표기됩니다.

4. 일반화 사전 처리 결과 보기

일반화 사전을 적용시킨 결과를 확인하려면 **컴포넌트 트리—문서**에서 임의의 문서¹를 더블 클릭합니다. 좌측에 새 탭이 생성되면 **HTML 탭 — 일반화 보기** 아이콘을 클릭하여 단어는 갈색으로, 일반어는 연두색으로 변환된 것을 알 수 있습니다.

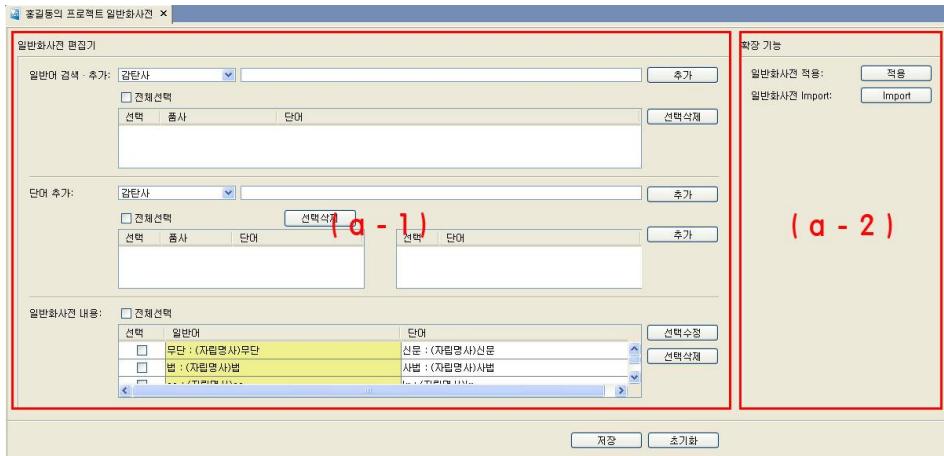


‘사전’에서 일반화 사전을 더블 클릭하면 작업창에 해당 사전을 편집할 수 있는 탭이 나타나며 프로그램에서 추출한 단어 이외에 사용자가 원하는 단어를 추가하거나 삭제할 수 있습니다.

a) 일반화 사전

컴포넌트 트리의 ‘사전’에 있는 임의의 일반화 사전을 클릭했을 때 작업창은 다음과 같습니다.

¹ 사전의 ‘적용’은 ‘문서’에 있는 모든 문서에 적용이 되기 때문에 임의의 문서에서 열람할 수 있습니다.



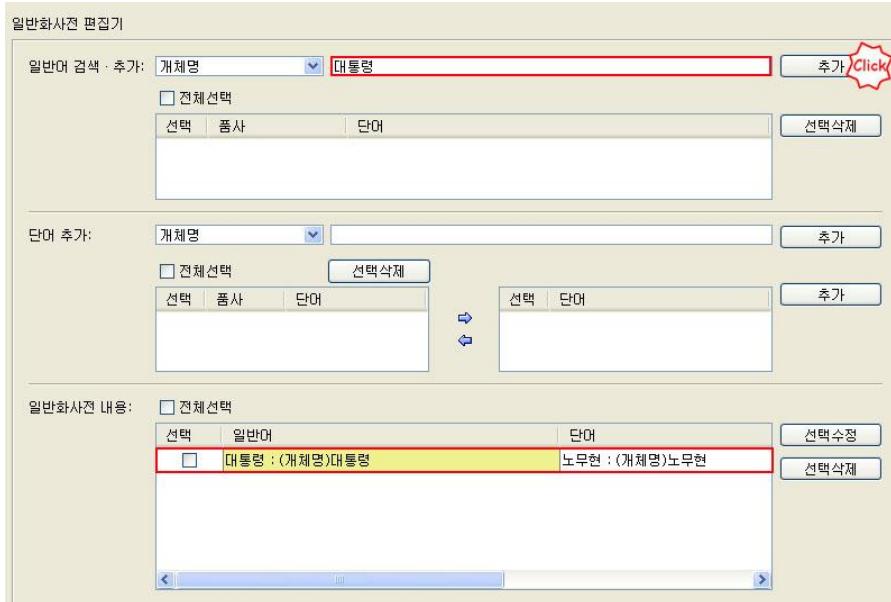
탭의 상단에는 프로젝트 명과 사전종류가 표기되며 하위 메뉴에는 ‘일반화사전 편집’과 ‘확장 기능’ 총 2 가지가 있습니다.

- 일반어는 동일한 의미를 가진 단어(들)이 통일되려는 단어입니다.
- 단어는 일반어로 통일하려는 단어(들)입니다.

(a-1) 일반화사전 편집기

i. 일반어 검색

일반어를 검색하고자 한다면 선택란에 해당하는 품사를 선택한 후 선택란 아래에 있는 일반어란에 찾고자 하는 일반어를 기입하면 하단에 조건에 일치하는 검색어가 나타납니다.

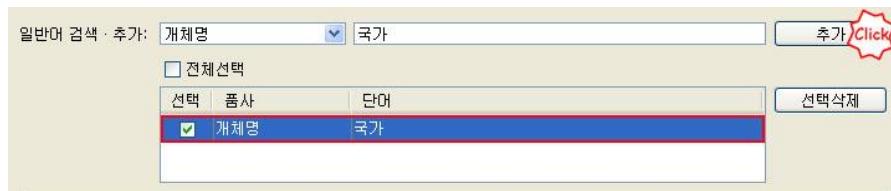


‘대통령’을 검색했을 때

ii. 일반어 생성

새로운 일반어를 생성하고자 한다면 일반어와 단어가 동시에 생성되어야 합니다.
일반화 사전은 일반어와 단어를 쌍을 이루어 모아놓은 사전이기 때문에 **일반어에 해당하는 단어 또한 같이 추가가 되어야 합니다.** 다음은 일반어를 생성하는 과정입니다.

(ii-1) 선택란에서 해당하는 품사를 선택한 후 선택란 아래에 있는 일반어란에 추가하고자 하는 일반어를 기입합니다. 이후 <확인>을 누르면 하단에 조건에 일치하는 일반어가 생성됩니다. 일반어가 생성되면 이에 해당하는 단어를 생성하기 위해 하단에 ‘단어 추가’로 이동합니다.



개체명인 ‘국가’를 일반어로 추가했을 때

*생성된 일반어를 삭제하고자 하는 경우 선택란에 체크를 한 후 **선택삭제**를 누릅니다.

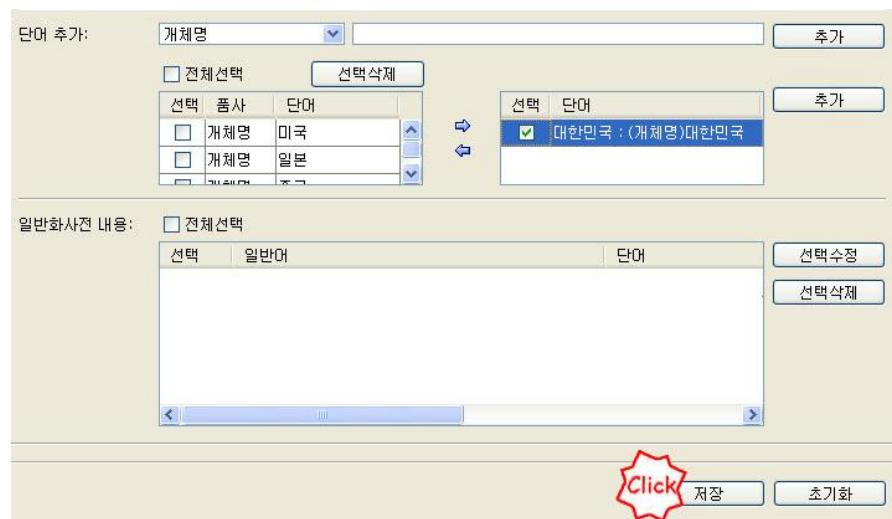
(ii-2) 선택란에 해당하는 품사를 선택한 후 선택란 좌측에 있는 단어란에 생성하고자 하는 단어를 기입하면 하단에 조건에 일치하는 단어가 생성됩니다.



'대한민국'을 일반어인 '국가'에 추가하려 할 때

*선택란에 체크를 한 후 **선택삭제**를 누르면 생성된 단어가 삭제됩니다.

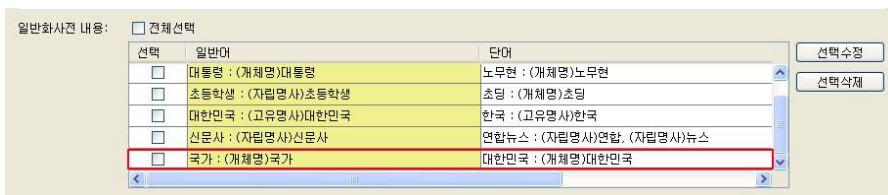
(ii-3) 생성된 단어에 체크를 한 후 **▶**을 클릭하면 단어가 우측으로 이동합니다. 우측란으로 이동된 단어만이 일반어와 쌍을 이루어 사전에 추가될 수 있습니다.



'대한민국'을 우측으로 이동시켰을 때

단어를 생성했지만 일반화 사전에 추가하는 것을 취소하려면 우측으로 이동한 단어에 체크를 한 후  을 누르면 좌측으로 단어가 이동합니다.

(ii-4) 이후 <확인>을 누르면 하단에 조건에 일치하는 일반어가 생성됩니다. 저장을 클릭하면 생성된 일반어가 일반화 사전에 저장이 됩니다.



iii. 일반어의 단어 추가

이미 생성된 일반어와 단어의 쌍에 여러 개의 단어를 추가할 수 있습니다. (방법 1) (a-1) i.에서 언급한 방식대로 일반어를 검색합니다. 이후 ‘단어 추가’로 이동하여 ii.(ii-2)에서 언급한 방식대로 단어를 추가합니다. 혹은 (방법 2) 하단에서 일반어를 직접 찾아 체크를 한 후  을 클릭합니다. ‘일반화 검색·선택’란에 일반어가 나타나면 ‘단어 추가’로 이동하여 ii.(ii-2)에서 언급한 방식대로 단어를 추가합니다.



일반어를 직접 찾아 <선택수정>을 클릭한 화면

- 선택수정 : 선택란에 체크가 되어 있는 일반어를 수정합니다.
- 선택삭제 : 선택란에 체크가 되어 있는 일반어를 삭제합니다.
- 저장 : 편집된 내용을 저장합니다.
- 초기화 : 명령이 실행되지 않은 초기의 상태로 돌아갑니다.

(a-2) 확장기능



- 일반화사전 적용

<적용>을 누르면 생성한 일반화 사전을 프로젝트에 적용합니다.

ii. 일반화사전 Import

<Import>를 누르면 생성한 기준에 생성한 일반화 사전을 현재 프로젝트에 가져옵니다.

5. 메타넷 사전

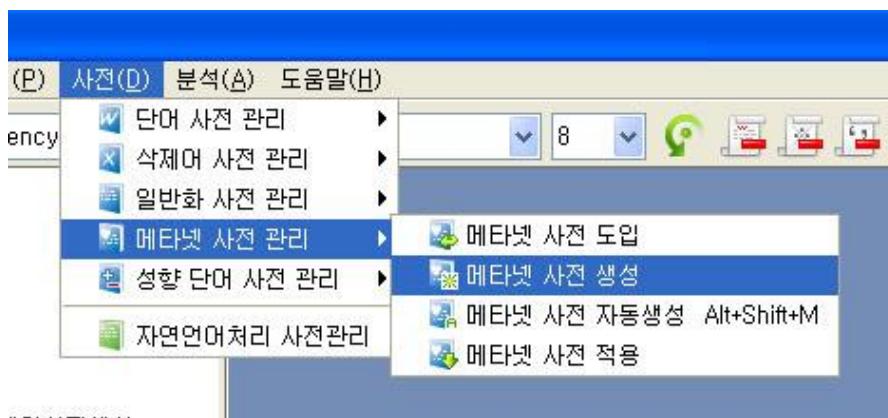
메타넷 사전은 단어의 성격을 명확하게 하기 위하여 온톨로지를 단어에 부여한 사전입니다. 온톨로지란 단어의 성격을 명확하게 할 수 있는 특정 개념으로 장소, 기관, 행동 등이 있습니다. 사용자가 단어에 온톨로지를 부여하여 단어와 온톨로지가 쌍으로 존재하게 됩니다.

메타넷 사전을 만드는 방법에는 (1) 사용자가 직접 단어를 추가하거나 삭제하기 위해 초기화된 사전을 만드는 ‘메타넷 사전 생성’과 (2) 원문에서 메타넷어를 추가하는 ‘메타넷 사전 원문에서 생성’ (3) 프로그램을 통해 자동으로 생성하는 ‘메타넷 사전 자동생성’ (4) 사용자의 필요에 따라 기존에 생성되었던 사전의 파일을 컴퓨터에서 불러오는 ‘메타넷 사전 도입’이 있습니다. 메타넷 사전을 생성한 후 사용자가 이를 문서에 적용시키고자 한다면 ‘메타넷 사전 적용’을 실행시킵니다.

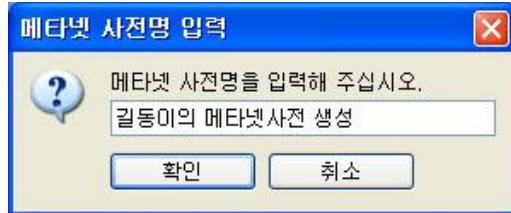
a) 메타넷 사전 생성

‘메타넷 사전 생성’은 사용자가 직접 단어를 추가/삭제하기 위하여 초기화된 일반화 사전을 생성합니다. 다음은 ‘일반화 사전 생성’을 실행하는 과정입니다.

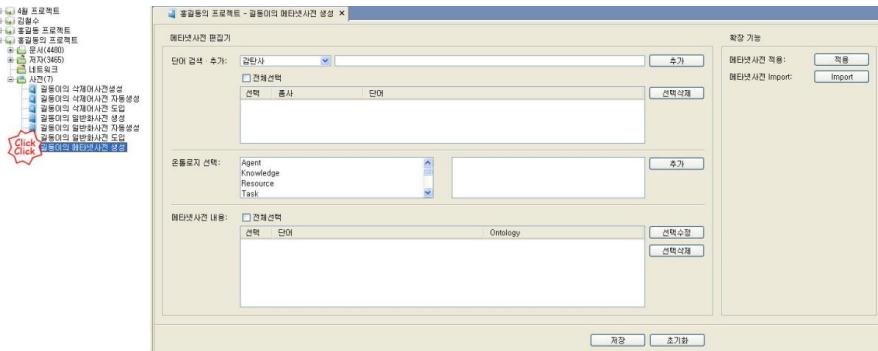
(a-1) ‘메타넷 사전 생성’을 실행하려면 (방법 1) **메뉴 막대 – 사전 – 메타넷 사전 관리 – 메타넷 사전 생성**을 누릅니다. 또는 (방법 2) 컴포넌트 트리의 ‘사전’에 오른쪽 마우스를 클릭한 후 ‘메타넷 사전 생성’을 누릅니다.



(a-2) ‘메타넷 사전명 입력’ 대화상자가 나타나면 기입란에 본인이 원하는 이름을 쓰고 <확인>을 누릅니다.



(a-3) 이후 컴포넌트 트리의 ‘사전’에 메타넷 사전이 생성됩니다. 생성된 메타넷 사전을 더블 클릭하면 작업창에 메타넷 사전 편집 탭이 나타나며 원하는 단어를 추가하거나 삭제할 수 있습니다.



(a-4) 생성된 메타넷 사전은 파일의 형태로 사용자가 지정한 작업폴더(최초 설정 C://Workplace, page 29, (d-2) 참조)에 자동 저장되며 메시지 판넬의 속성에서 저장된 위치 등 상세정보를 확인할 수 있습니다.

(a-5) 명령이 성공적으로 수행되었다면 속성에는 ***** Successfully executed ***** 가 표기됩니다.

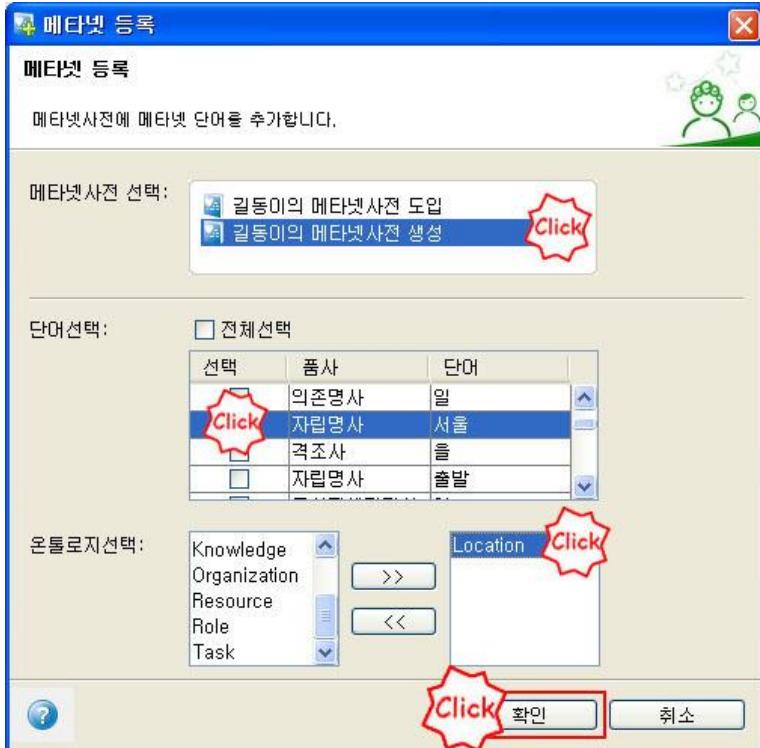
b) 메타넷 사전 원문에서 생성

'메타넷 사전 원문에서 생성'은 원문에 드래그를 하여 단어를 추가하는 방식을 사용하며 사용자가 원문에서 직접 메타넷어를 등록하여 사전을 생성하거나 기존에 생성된 메타넷 사전에 단어를 추가합니다. 다음은 '메타넷 사전 원문에서 생성'을 실행하는 과정입니다.

(b-1) 메타넷 사전 원문에서 생성 컴포넌트 트리 '문서'에서 추가하고자 하는 단어가 포함되어 있는 문서를 클릭하여 작업창에 새 탭을 생성합니다. 메타넷 사전으로 추가하고자 하는 단어나 문장에 마우스로 드래그합니다. 이후 마우스 오른쪽 버튼을 클릭하여 '메타넷 등록'을 누릅니다.



(b-2) '메타넷 등록' 대화상자가 나타나면 '사전 선택'에서 메타넷어가 추가되려는 메타넷 사전을 클릭합니다. '단어 선택'란은 사용자가 드래그를 한 단어/문장의 품사가 자동으로 구분이 되어 나타납니다. 추가하고자 하는 단어에 체크를 한 후 '온톨로지 선택'으로 이동하여 해당 온톨로지를 추가합니다. 설정이 완료되면 <확인>을 클릭합니다.



'서울'에 'Location'을 부여

(b-3) 컴포넌트 트리의 '사전'에서 해당 메타넷 사전을 더블 클릭하여 작업창에 새 탭을 생성합니다.
선택한 단어가 메타넷 사전에 등록된 것을 확인합니다.



- i. 선택수정 [선택수정] : 선택란에 체크가 되어 있는 메타넷어를 수정합니다.
- ii. 선택삭제 [선택삭제] : 선택란에 체크가 되어 있는 메타넷어를 삭제합니다.

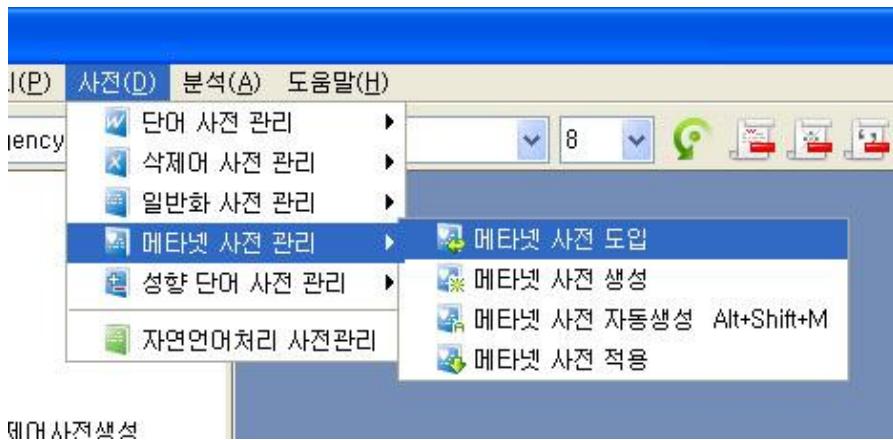
(b-4) 생성된 삭제어 사전은 파일의 형태로 사용자가 지정한 작업폴더(최초 설정- C://Workplace, page 29, (d-2) 참조)에 자동 저장되며 메시지 판넬의 속성에서 저장된 위치 등 상세정보를 확인할 수 있습니다.

(b-5) 명령이 성공적으로 수행되었다면 속성에는 ******* Successfully executed ******* 가 표기됩니다.

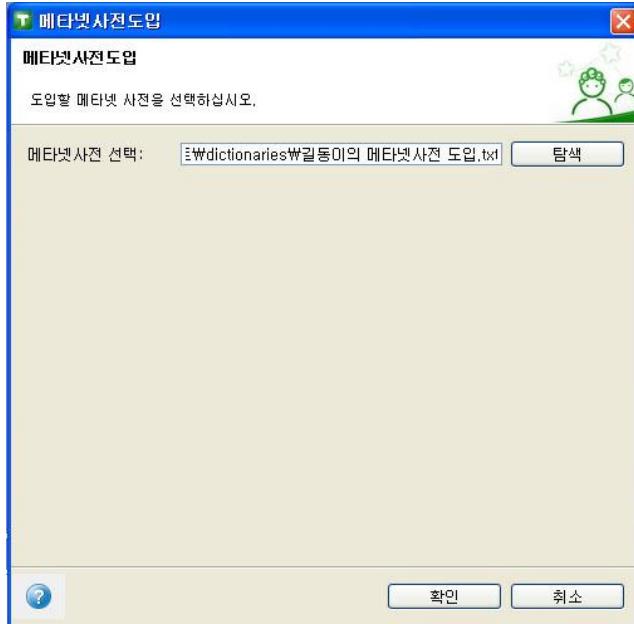
c) 메타넷 사전 도입

기존에 생성되었던 메타넷 사전의 파일을 컴퓨터에서 불러옵니다. 다음은 ‘메타넷 사전 도입’을 실행하는 과정입니다.

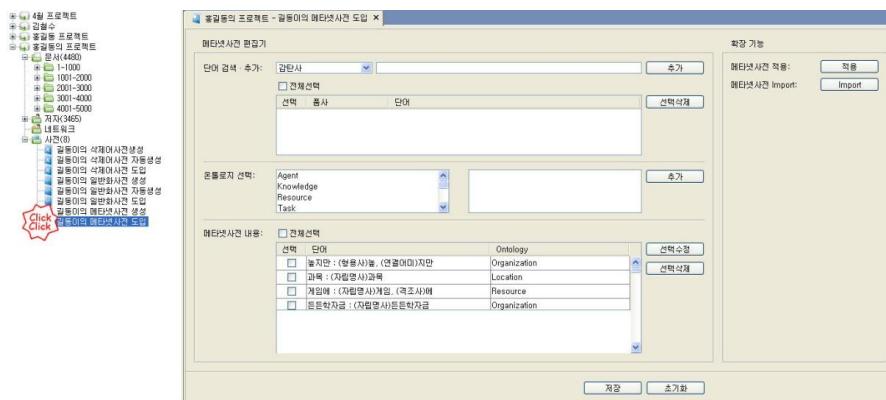
(c-1) ‘메타넷 사전 도입’을 실행하려면 **메뉴막대 – 분석 – 메타넷 사전 관리 – 메타넷 사전 도입**을 누릅니다.



(c-2) ‘메타넷 사전 도입’ 대화상자가 나타나면 <탐색>을 눌러 ‘사전’로 불러올 파일을 선택한 후 <확인>을 클릭합니다.



(c-3) 이후 컴포넌트 트리의 ‘사전’에 메타넷 사전이 생성됩니다. 불러온 메타넷 사전을 더블 클릭하면 작업창에 일반화 사전 편집 탭이 나타나며 사용자가 원하는 단어를 추가하거나 삭제할 수 있습니다.

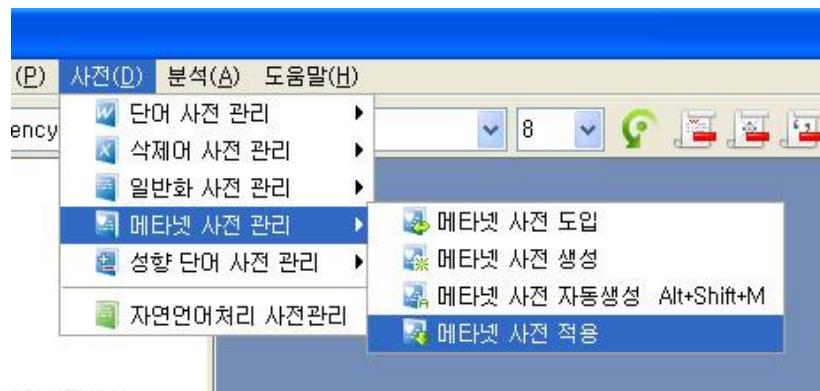


(c-4) 명령이 성공적으로 수행되었다면 속성에는 ******* Successfully executed ******* 가 표기됩니다.

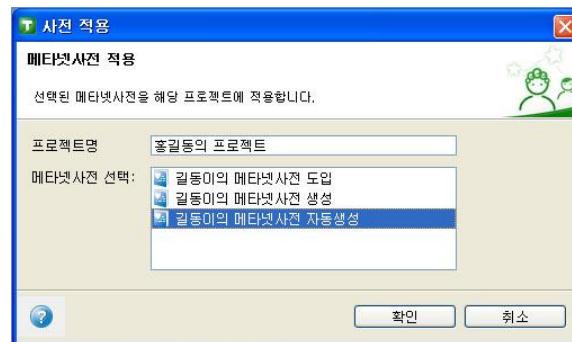
d) 메타넷 사전 적용

생성한 메타넷 사전을 프로젝트에 적용합니다. 다음은 ‘메타넷 사전 적용’을 실행하는 과정입니다.

(d-1) ‘일반화 사전 적용’을 실행하려면 **메뉴막대 – 사전 – 메타넷 사전 관리 – 메타넷 사전 적용**을 누릅니다.

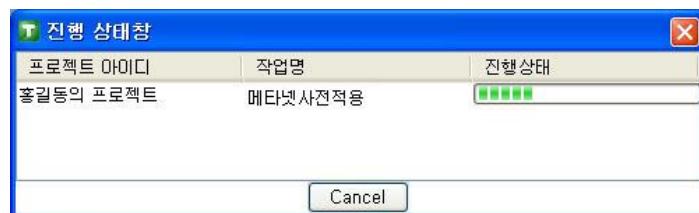


(d-2) ‘메타넷 사전 선택’ 대화상자가 나타나면 프로젝트에 적용시킬 메타넷 사전을 선택한 후 <OK> 를 클릭합니다.



- i. All : 목록에 있는 사전을 모두 선택합니다.
- ii. None : 체크 표시가 되어 있는 항목을 해제합니다.
- iii. OK : 선택을 완료하고 다음 단계로 넘어갑니다.
- iv. Cancel : ‘일반화 사전 선택’ 대화상자를 취소합니다.

(d-3) ‘진행’ 대화상자가 나타나면 막대가 모두 100%에 도달할 때까지 기다립니다. 이때 <Cancel> 을 클릭하면 ‘일반화 사전 적용’이 중단됩니다.



(d-4) 명령이 성공적으로 수행되었다면 속성에는 ***** Successfully executed ***** 가 표기됩니다.

6. 메타넷 사전 처리 결과 보기

메타넷 사전을 적용시킨 결과를 확인하려면 **컴포넌트 트리—문서**에서 임의의 문서¹를 더블 클릭합니다. 좌측에 새 탭이 생성되면 **HTML 탭 — 메타넷 보기** 아이콘을 클릭하여 단어가 빨간색으로 온톨로지가 하늘색으로 변한 것을 알 수 있습니다.

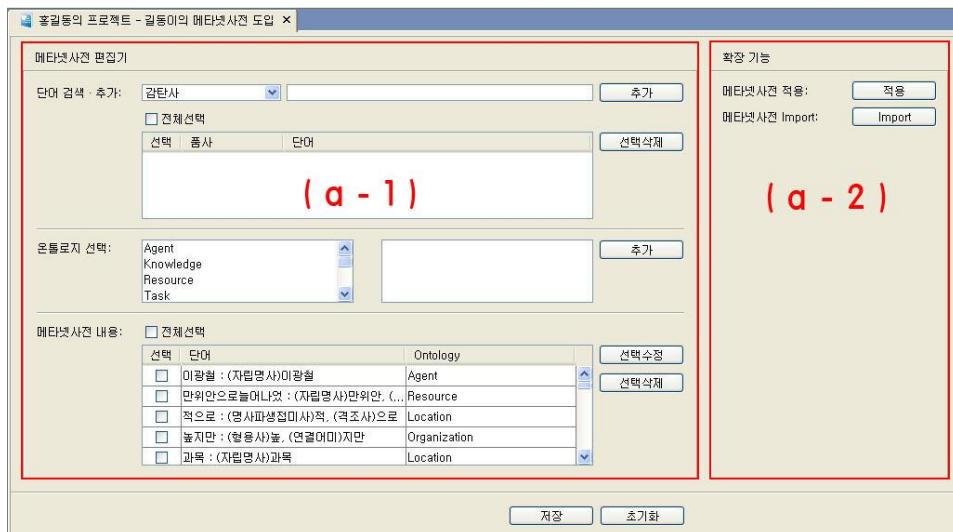


‘사전’에서 메타넷 사전을 더블 클릭하면 작업창에 해당 사전을 편집할 수 있는 탭이 나타나며 프로그램에서 추출한 단어 이외에 사용자가 원하는 단어를 추가하거나 삭제할 수 있습니다.

a) 메타넷 사전

컴포넌트 트리의 ‘사전’에 있는 임의의 메타넷 사전을 클릭했을 때 작업창은 다음과 같습니다.

¹ 사전의 ‘적용’은 ‘문서’에 있는 모든 문서에 적용이 되기 때문에 임의의 문서에서 열람할 수 있습니다.



탭의 상단에는 프로젝트 명과 사전종류가 표기되며 하위 메뉴에는 ‘메타넷사전 편집기’과 ‘확장 기능’ 총 2 가지가 있습니다.

- 메타넷어는 개념어와 온톨로지를 합한 말입니다.
- 온톨로지는 단어의 성격을 나타낼 수 있는 특정 개념입니다. 온톨로지가 단어에 부여되면 그 단어를 개념어라 부릅니다.

(a-1) 메타넷 사전 편집기

i. 메타넷 검색 추가

메타넷어를 검색하고자 한다면 선택란에 해당하는 품사를 선택한 후 선택란 왼쪽에 있는 메타넷어란에 찾고자 하는 메타넷어를 기입하면 좌측에 조건에 일치하는 검색어가 나타납니다.

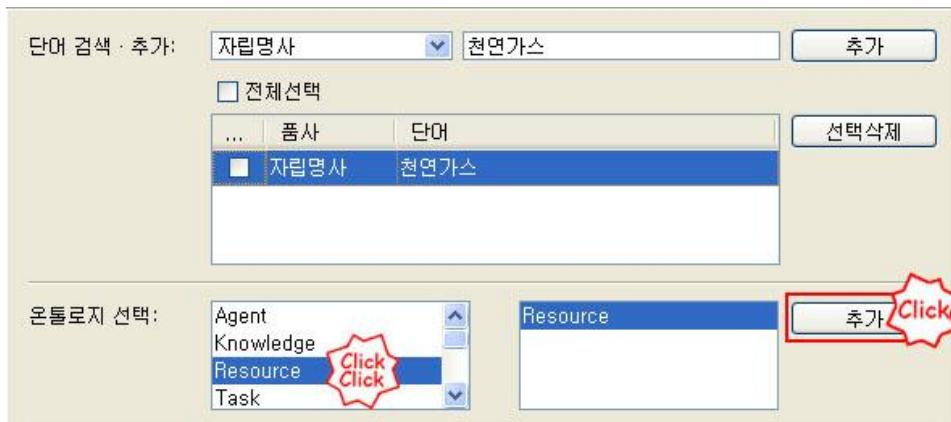
새로운 메타넷어를 생성하고자 한다면 단어에 온톨로지를 반드시 부여해야 합니다.
메타넷 사전은 단어에 온톨로지를 부여한 메타넷어를 모아둔 사전이기 때문에 단어에 해당하는 온톨로지 또한 같이 추가가 되어야 합니다. 다음은 메타넷어를 생성하는 과정입니다.

(i-1) 선택란에 해당하는 품사를 선택한 후 선택란 좌측에 있는 단어란에 생성하고자 하는 단어를 기입하면 하단에 조건에 일치하는 단어가 생성됩니다.



*생성된 단어를 삭제하고자 하는 경우 선택란에 체크를 한 후 **선택삭제**를 누릅니다.

(i-2) 단어가 생성되면 이에 해당하는 온톨로지를 생성하기 위해 하단에 ‘온톨로지 선택’으로 이동합니다. ‘온톨로지 선택’에서 생성된 단어에 해당하는 온톨로지를 선택한 후 더블클릭하면 우측으로 이동합니다. 여러 개의 온톨로지를 이동시켜 한 단어에 추가할 수도 있습니다. **우측란으로 이동된 온톨로지만이 단어에 부여되어 사전에 추가될 수 있습니다.**



*우측으로 이동한 온톨로지를 취소하려면 우측으로 이동한 해당 온톨로지를 더블클릭합니다.

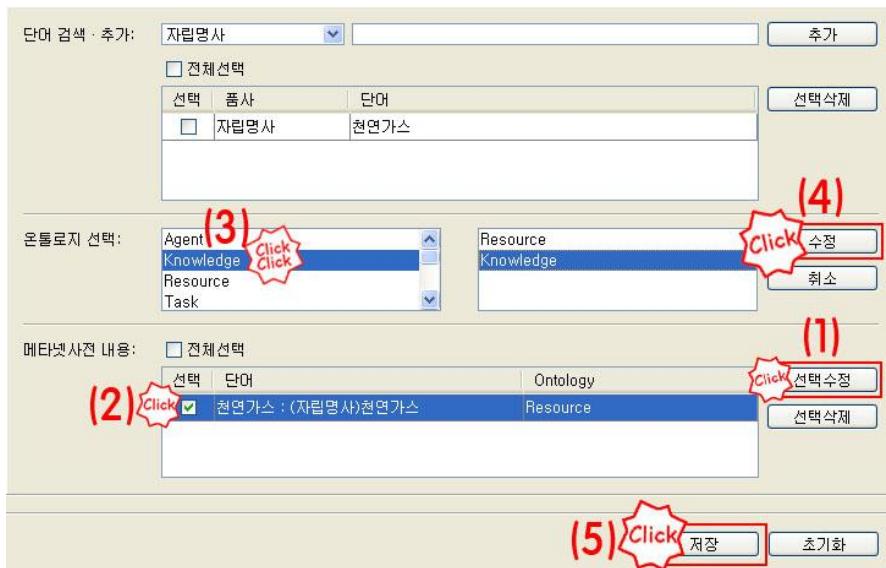
(i-3) 이후 <확인>을 누르면 좌측에 조건에 일치하는 메타넷어가 생성됩니다. 저장을 클릭하면 생성된 메타넷어가 메타넷 사전에 저장이 됩니다.



선택란에 체크를 한 후 **선택삭제**를 누르면 생성된 메타넷어가 삭제됩니다.

b) 메타넷어의 온톨로지 추가

이미 생성된 단어와 온톨로지의 쌍에 사용자의 필요에 따라 여러 개의 온톨로지를 추가로 부여할 수 있습니다. (방법 1) <(a-1) i. , page 321>에서 언급한 방식대로 단어를 검색합니다. 이후 ‘온톨로지 선택’으로 이동하여 <i.(i-2), page 322>에서 언급한 방식대로 온톨로지를 추가합니다. 혹은 (방법 2) 하단에서 메타넷어를 직접 찾아 선택란에 체크를 한 후 **선택수정**을 클릭합니다. ‘온톨로지 선택’란에 온톨로지가 나타나면 <i.(i-2), page 322>에서 언급한 방식대로 온톨로지를 추가합니다.



메타넷어를 직접 찾아 <선택수정>을 클릭한 화면

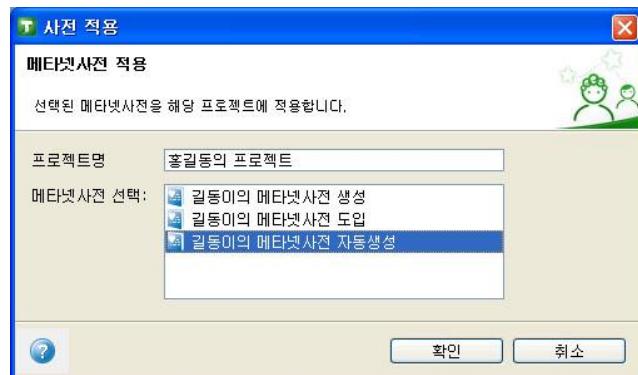
- (ii -1) 저장 : 편집된 내용을 저장합니다.
(ii -2) 초기화 : 명령이 실행되지 않은 초기의 상태로 돌아갑니다.

(a-2) 확장 기능



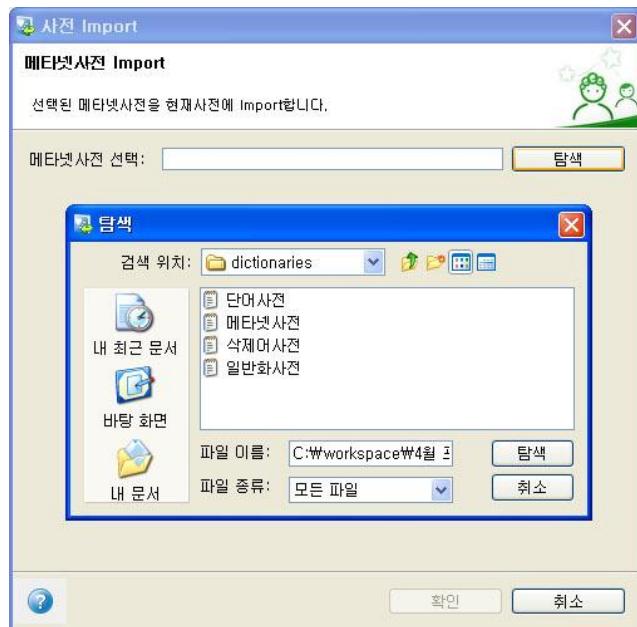
i. 메타넷사전 적용

메타넷 사전을 프로젝트에 적용합니다. <적용>을 누르면 아래와 같은 대화상자가 나타납니다.
메타넷 사전 선택란에서 적용하고자 하는 메타넷 사전을 선택한 후 <확인>을 클릭합니다.



ii. 메타넷사전 Import

<Import>를 누르면 생성한 기준에 생성한 일반화 사전을 현재 프로젝트에 가져옵니다.



E. 문서내 관계망 생성하기

3 개의 사전이 모두 생성되면 이를 기반으로 단어와 단어 사이에 관계가 존재하는지 알아봅니다. 여기서 단어는 문서 내에 있는 모든 단어가 아니라 TONK 가 지정한 특정 단어들을 말합니다. 이 단어를 개념어라 부르며 기관, 행동, 자원, 장소 등 특정한 개념을 가지고 있습니다. 여기서 특정한 개념을 온톨로지라 부르며 총 8 가지가 있습니다. 사용자가 온톨로지를 직접 선택하면 TONK 가 온톨로지에 해당하는 개념어를 찾아 이들의 관계를 산출해 냅니다. 추출된 개념어를 열람하고 이들간의 관계를 확인하는 과정은 ORA 라는 프로그램을 통해 이루어집니다. 다음은 ‘문서내 관계망 생성’의 자세한 내용입니다.

1. 문서내 관계망 생성

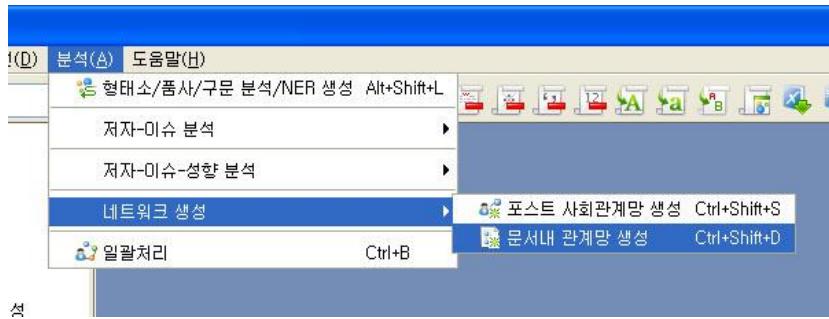
‘문서내 관계망 생성’은 문서 내에서 특정 개념에 해당하는 단어들을 추출하여 이 단어들과의 관계를 선으로 연결하거나 수치로 나타냅니다. 특정 개념은 총 10 가지로 사람, 장소, 정보, 자원, 임무, 사건, 장소, 기관, 신념, 역할, 행동이 있습니다. TONK 에서는 이러한 특정 개념을 ‘온톨로지’라고 정의하며 온톨로지를 가지고 있는 단어를 개념어라 부릅니다. 예를 들어¹ ‘이명박’은 ‘사람’이라는 온톨로지를 가지고 있는 개념이며 ‘서울’은 ‘장소’라는 온톨로지를 가지고 있는 개념입니다. 개념어를 파악하는 것 이외에도 개념어와 개념어 사이의 관계를 파악할 수 있습니다. 이는 문서 내에 존재하는 개념어 간의 거리를 사용자가 수치로 지정하여 결정하며 단어간격이라는 설정을 통하여 조절할 수 있습니다.

사용자는 원하는 온톨로지를 직접 선택하여 관계망을 생성한 후 ORA 를 이용하여 결과를 확인합니다. ‘문서내 관계망 생성’은 반드시 <IV.D 사전 생성하기, page 271>까지 마친 상태에서 실행되어야 하며 특정 단어들의 연관 관계를 파악하는 데에 유용합니다.

다음은 ‘문서내 관계망 생성’을 실행하는 과정입니다.

- a) 문서내 관계망 생성을 실행하려면 **메뉴막대 – 분석 – 네트워크 생성 – 문서내 관계망 생성**을 누릅니다.

¹ 온톨로지에 여러 개의 개념어가 속할 수 있습니다. 사람이라는 온톨로지에는 이명박, 반기문, 노무현, 홍길동 등이 해당됩니다.



b)



(b-1) ‘문서내 관계망 분석’ 대화상자가 나타나면 ‘Prefix’에는 본인이 원하는 관계망명을 씁니다.

관계망이 생성되면 본인이 ‘Prefix’에 지정한 이름에 Overall-document network 이라는 태그가 붙습니다. 생성된 관계망은 각 문서 별로 여러 개가 저장되기 때문에 하나의 파일로 나타내기 위하여 개별 문서들을 종합한 결과물을 따로 추출한 후 ‘Overall-document network’이라는 접미사를 붙인 것입니다. 위와 반대로 개별 문서 별로 만들어진 문서내 관계망은 그 폴더에서 접미사가 문서 아이디로 지정되어 있습니다.

(b-2) ‘온톨로지¹ 타입’에서 분석하고자 하는 온톨로지를 클릭한 후 추가버튼을 눌러 ‘대상 온톨로지 타입’으로 이동시킵니다. 만약 이동 시킨 온톨로지를 삭제하고자 하는 경우에는 삭제버튼을 누릅니다.

(b-3) 단어간격은 개념어와 개념어 간의 관계가 존재하는지의 여부를 결정하는 기능으로써 사용자가 관계가 존재할 수 있는 범위를 문서 내에 있는 단어들 간의 거리 수치(word distance)를 기준으로 지정합니다. 예를 들어 단어간격을 7로 지정하면, 두 개념어 사이에 7개의 단어보다 적은 숫자의 개념어가 존재했을 때 그 두 개념어 사이의 관계가 존재한다고 인식합니다. 영어의 경우에는 대부분 단어간격 7을 기본으로 하고 있으며 한국어는 아직까지 표준이 정해지지 않았습니다.

(b-4) 분석 모드에서 **MODE_LINKLIMIT_BYSENTENCE** 는 개념어를 추출하는데 있어서 문장을 단위로 단어간격을 적용하며 **MODE_LINKLIMIT_BYDOCUMENT** 는 문서를 단위로 단어간격을 적용합니다.

(b-5) 문서적용타입 : 문서적용타입에 체크표시를 하면 사용자의 기호에 따라 특정 사전 혹은 NER 을 적용시킬 수 있습니다. 만약 사용자가 NER 만을 선택할 경우, 자연어 처리도구에서 선택된 온톨로지만을 이용하여 문서내 관계망을 구성하고 사용자가 생성한 메타넷 사전을 선택할 경우, NER 을 무시하고 사용자의 기호대로 문서내 관계망을 구성합니다. 체크표시를 하지 않으면 NER 및 사전 모두 적용됩니다.

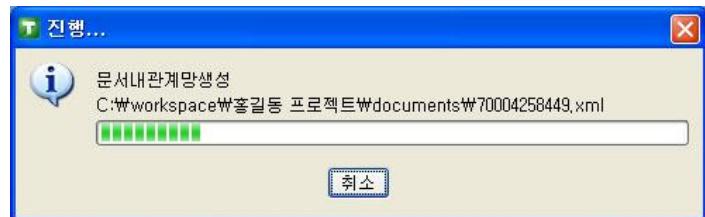
(b-6) 아이디 선택 : 원 자료에서 제공되는 저자 ID 및 문서 ID 를 이용하여 저자 및 문서 노드를 생성하고 이를 네트워크에 추가합니다. 선택하지 않았을 경우에는 저자 ID 및 문서 ID 정보를 이용하지 않습니다.

첫번째 방법은 문장이 끝났을 경우 단어들 간의 거리(Word distance)를 세는 기록을 초기화시키고 두번째 방법은 문서가 끝났을 경우 단어들 간의 거리(Word distance)를 초기화 시킵니다. 일반적으로 첫 번째 방법을 이용하여 생성되는 문서내 관계망이 두 번째 방법을 통한 것보다 조밀도가 떨어집니다.

설정이 끝나면 <확인> 을 클릭합니다.

¹ 온톨로지는 TONK 에서 지정한 특정 개념입니다. 사람, 장소, 정보, 자원, 임무, 사건, 장소, 기관, 신념, 역할, 행동이 있습니다.

- c) ‘진행 상태창’ 대화상자가 나타나면 연두색 막대가 채워질 때까지 기다립니다. 이때 <Cancel>을 클릭하면 ‘문서내 관계망 생성’가 중단됩니다.



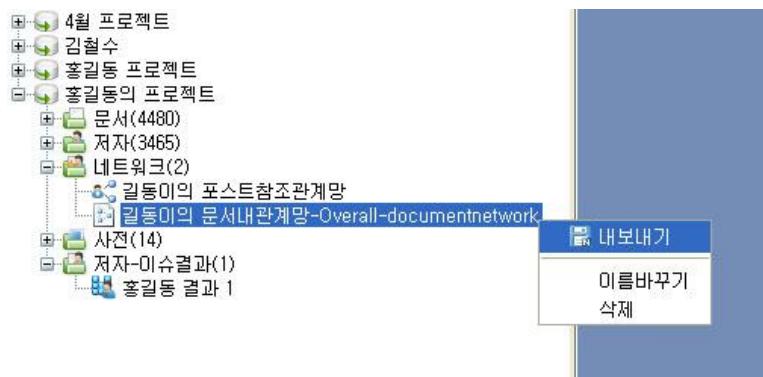
- d) 명령이 성공적으로 수행되었다면 속성에는 ***** Successfully executed ***** 가 표기됩니다.

2. 문서내 관계망 처리결과 보기

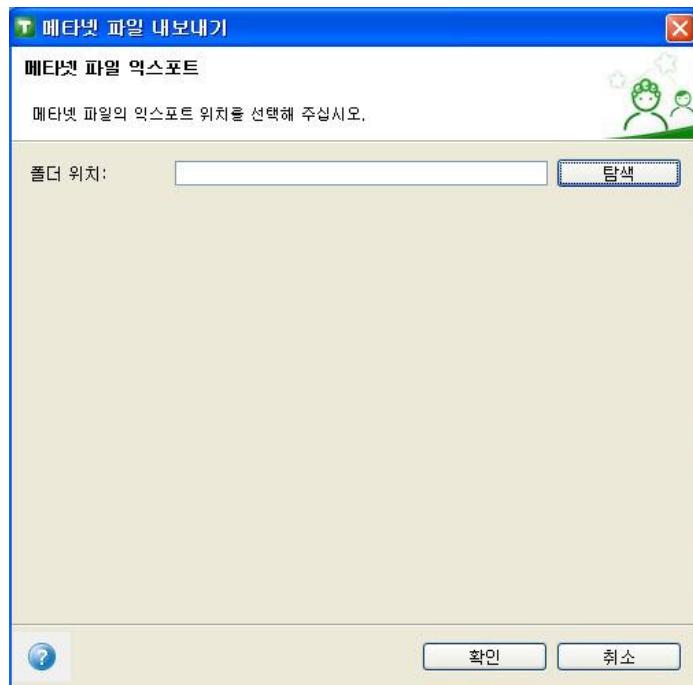
문서내 관계망 생성을 실행하였을 때의 결과는 ORA에서 확인합니다. 다음은 ‘문서내 관계망 생성’의 결과를 확인하는 과정입니다.

a) TONK에서 파일 내보내기

(a-1) 컴포넌트 트리의 ‘네트워크’에서 생성된 문서내 관계망 문서에 마우스 오른쪽을 누른 후 ‘내보내기’를 클릭합니다.



(a-2) ‘메타넷 파일 내보내기’ 대화상자가 나타나면 <탐색>을 클릭하여 저장할 장소를 지정한 후 <확인>을 누릅니다.

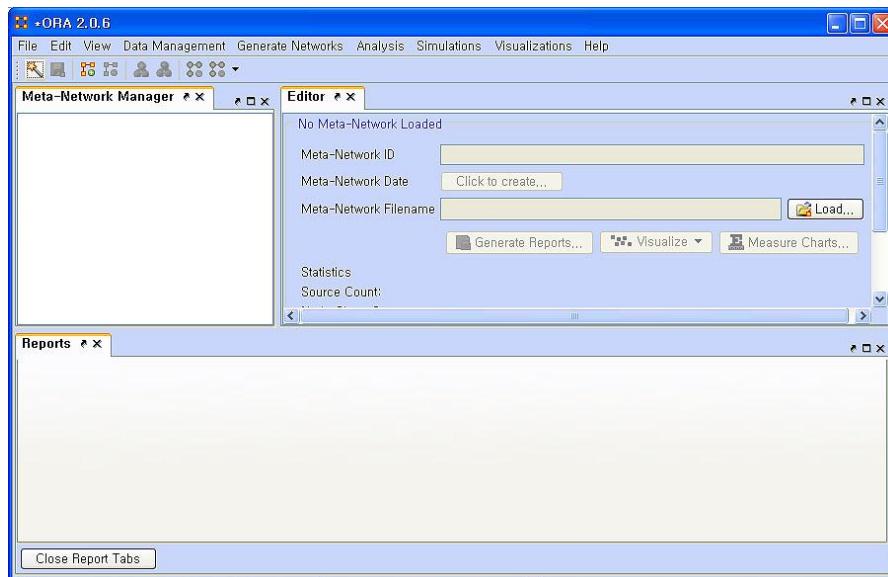


b) ORA에서 파일 불러오기

(b-1) 바탕화면에 있는 ORA 아이콘을 더블 클릭하여 ORA를 실행합니다.



ORA 아이콘을 더블 클릭

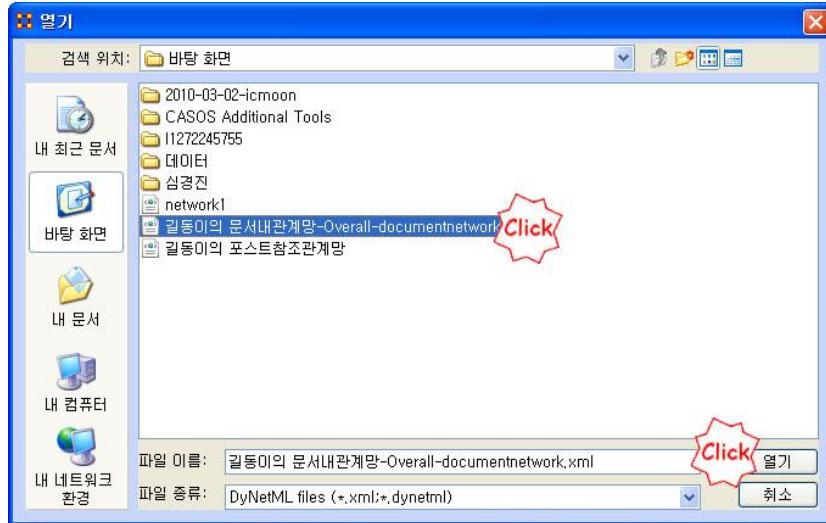


ORA 를 실행하였을 때 초기화면

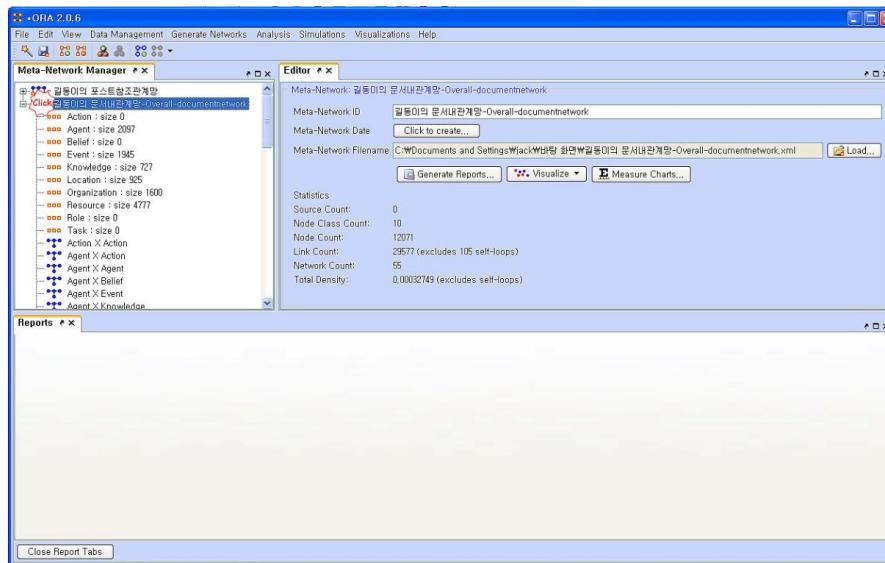
(b-2) ORA 가 실행되면 **File — Open Meta Network** 을 클릭합니다.



(b-3) ‘열기’ 대화상자가 나타나면 익스포트를 실행시킨 문서를 클릭한 후 <확인>을 누릅니다.



(b-4) Mata Network Manager 에 TONK 에서 생성했던 문서내 관계망 문서가 나타납니다. + 를 클릭하면 관계망 문서의 하위 메뉴인 저자, 문서, 토픽 등이 나타납니다.

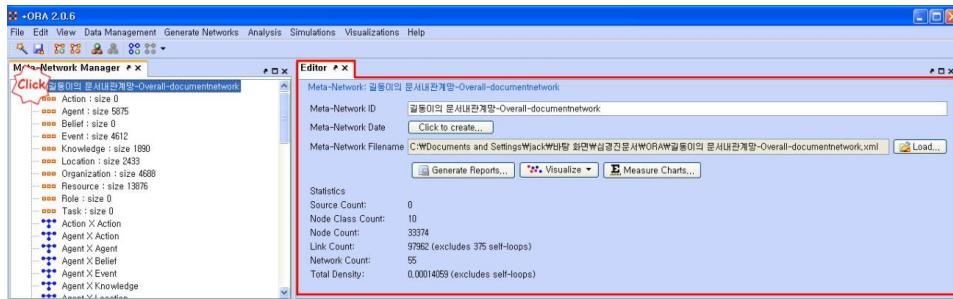


c) ORA에서 결과 열람하기

Mata-Network Manager에 나타나는 저자-이슈-문서 네트워크의 하위 메뉴는 총 8 가지로, 분석 대상인 개별 항목을 일컫는 **Node Set**과 이들간의 연결관계를 나타내는 **Network**로 이루어져 있습니다. Node Set에는 Action, Agent, Belief, Event, Knowledge, Location, Organization, Resource, Role, Task가 있으며 Network에는 총 55개로 페이지 357을 참조합니다. 문서내 관계망에 대한 자세한 내용은 다음과 같습니다.

(c-1) 문서내 관계망

Mata-Network Manager에 나타난 결과는 TONK에서 생성한 문서내 관계망입니다. TONK의 **분석 — 문서내 관계망 생성 — 문서내 관계망 분석-Prefix**에서 지정했던 이름¹과 같습니다. 생성된 문서내 관계망을 클릭하면 Editor란에 관련정보가 나타납니다.



36

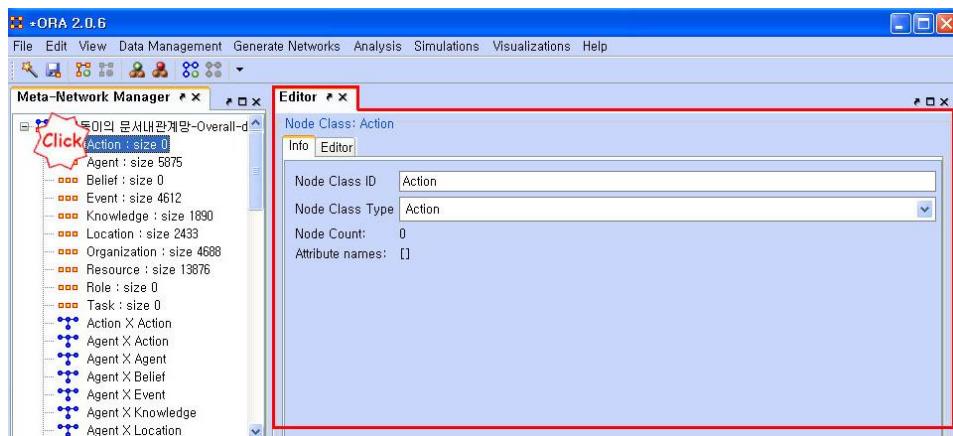
- i. Meta-Network ID : 문서내 관계망의 아이디를 나타내며 변경할 수 있습니다.
- ii. Meta-Network Date : TONK에서 문서내 관계망을 만들었던 날짜입니다.
- iii. Meta-Network Filename : 문서내 관계망이 저장된 장소입니다.
- iv. Source Count : 문서 출처의 개수입니다.
- v. Node Class Count : 문서내 관계망에 존재하는 총 Node의 수입니다.
- vi. Link Count : 문서내 관계망에 존재하는 총 연결관계의 수입니다.
- vii. Network Count : 문서내 관계망에 존재하는 총 네트워크의 수입니다.
- viii. Total Density : 네트워크의 조밀도입니다.

¹ Prefix에서 지정한 이름에 Overall-documentnetwork 태그가 붙습니다.

(c-2) Action

문서내 관계망의 하위 메뉴로 나타나는 Action 은 TONK 에서 선택한 온톨로지 입니다. size 는 Action 에 해당하는 개념어의 수 입니다. Action 은 TONK 의 **분석—문서내 관계망 생성 — 문서내 관계망 분석- 온톨로지 탑입**에서 사용자가 Action 을 선택하였기 때문에 Action 에 해당하는 개념어의 정보가 나타납니다. 만약 Action 을 선택하지 않았을 경우 하위메뉴로 생성이 되지 않습니다. 주의할 점은 size 가 0 이 나왔을 경우, 사용자가 TONK 에서 온톨로지 탑입을 선택하지 않은 것이 아니라, 온톨로지에 해당하는 개념어가 없다는 것입니다.

Action 를 클릭하면 우측 Editor 란에 관련정보가 나타나며 Info 와 Editor 탭으로 구성되어 있습니다.



i. Info

선택된 Node 의 기본 정보를 나타냅니다.



(i-1) Node Class ID : 선택된 Node 의 아이디를 나타내며 아이디를 변경할 수 있습니다.

(i-2) Node Class Type : 선택된 Node 에 해당되는 온톨로지입니다.

(i-3) Node Count : 해당 온톨로지에 존재하는 총 개념어 수 입니다.

(i-4) Attributes names : 저자의 특성치 목록입니다.

ii. Editor



(ii-1) Find : 사용자가 원하는 특정 Node Id 및 Node Title 을 검색할 수 있습니다.

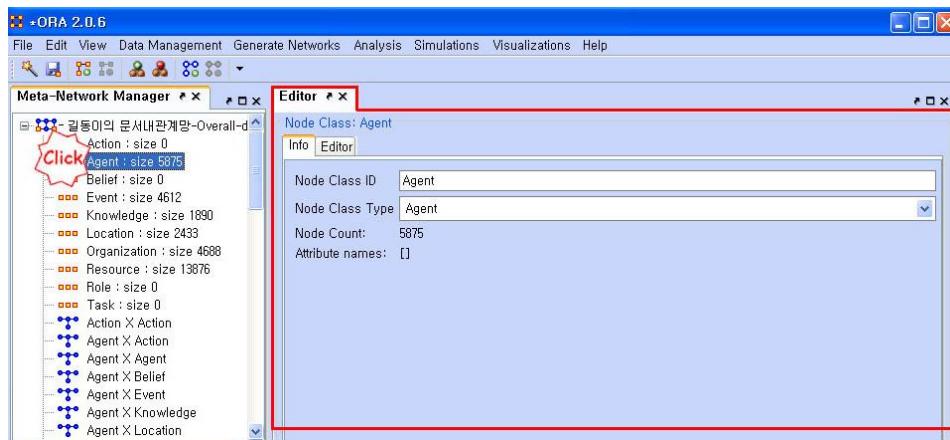
(ii-2) Node ID : Action 에 부여한 숫자입니다.

(ii-3) Node Title : Action 에 해당하는 개념어입니다.

(c-3) Agent

문서내 관계망의 하위 메뉴로 나타나는 Agent 는 TONK에서 선택한 온톨로지 입니다. size 는 Agent에 해당하는 개념어의 수입니다. Agent는 TONK의 **분석—문서내 관계망 생성 — 문서내 관계망 분석-온톨로지** 탑입에서 사용자가 Agent를 선택하였기 때문에 Agent에 해당하는 개념어의 정보가 나타납니다. 만약 Agent를 선택하지 않았을 경우 하위 메뉴로 생성이 되지 않습니다. 주의할 점은 size 가 0 이 나왔을 경우, 사용자가 TONK에서 온톨로지 탑입을 선택하지 않은 것이 아니라, 온톨로지에 해당하는 개념어가 없다는 뜻입니다.

Agent를 클릭하면 우측 Editor란에 관련정보가 나타나며 Info와 Editor탭으로 구성되어 있습니다.



i. Info

선택된 Node의 기본 정보를 나타냅니다.



- (i-1) Node Class ID : 선택된 Node 의 아이디를 나타내며 아이디를 변경할 수 있습니다.
- (i-2) Node Class Type : 선택된 Node 에 해당되는 온톨로지입니다.
- (i-3) Node Count : 해당 온톨로지에 존재하는 총 개념이 수입니다.
- (i-4) Attributes names : 저자의 특성치 목록입니다.

ii. Editor



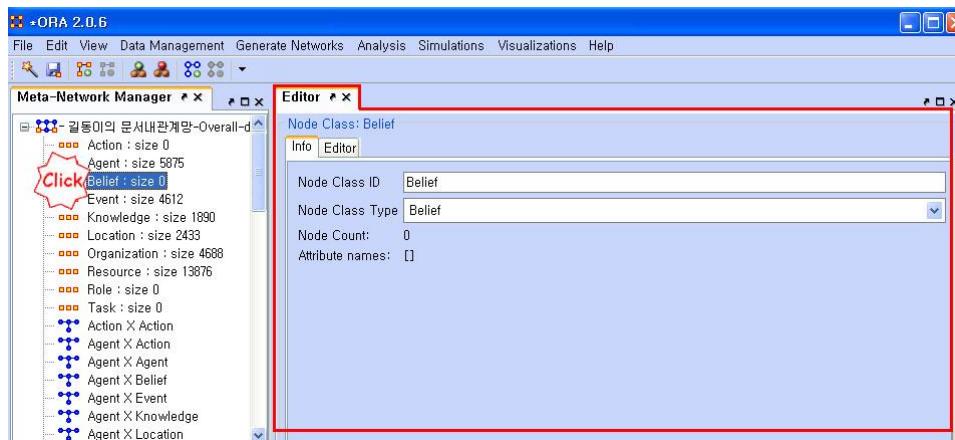
- (ii-1) Find : 사용자가 원하는 특정 Node Id 및 Node Title 을 검색할 수 있습니다.
- (ii-2) Node ID : Agent 에 부여한 숫자입니다.

(ii-3) Node Title : Agent 에 해당하는 개념어입니다.

(c-4) Belief

문서내 관계망의 하위 메뉴로 나타나는 Belief 는 TONK에서 선택한 온톨로지 입니다. size 는 Belief 에 해당하는 개념어의 수 입니다. Belief 는 TONK 의 **분석—문서내 관계망 생성 — 문서내 관계망 분석-온톨로지** 타입에서 사용자가 Belief 를 선택하였기 때문에 Belief 에 해당하는 개념어의 정보가 나타납니다. 만약 Belief 를 선택하지 않았을 경우 하위 메뉴로 생성이 되지 않습니다. 주의할 점은 size 가 0 이 나왔을 경우, 사용자가 TONK에서 온톨로지 타입을 선택하지 않은 것이 아니라, 온톨로지에 해당하는 개념어가 없다는 뜻입니다.

Belief 를 클릭하면 우측 Editor 란에 관련정보가 나타나며 Info 와 Editor 탭으로 구성되어 있습니다.



i. Info

선택된 Node 의 기본 정보를 나타냅니다.



(i-1) Node Class ID : 선택된 Node 의 아이디를 나타내며 아이디를 변경할 수 있습니다.

(i-2) Node Class Type : 선택된 Node 에 해당되는 온톨로지입니다.

(i-3) Node Count : 해당 온톨로지에 존재하는 총 개념이 수 입니다.

(i-4) Attributes names : 저자의 특성치 목록입니다.

ii. Editor



(ii-1) Find : 사용자가 원하는 특정 Node Id 및 Node Title 을 검색할 수 있습니다.

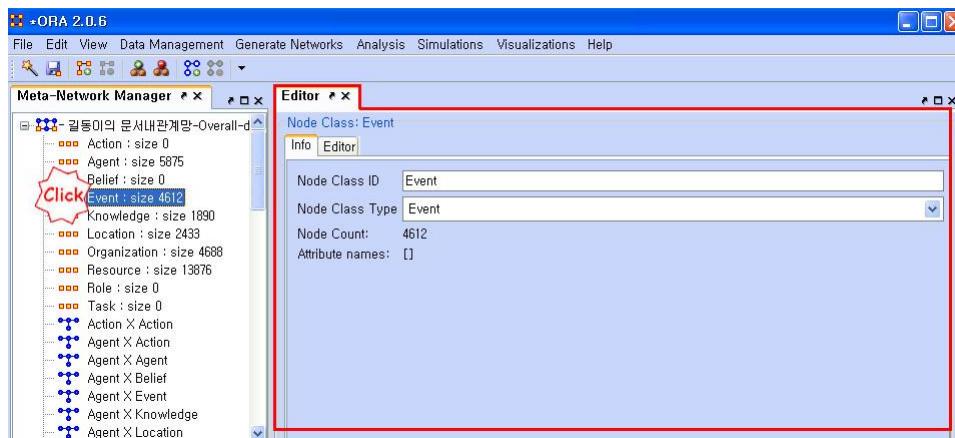
(ii-2) Node ID : Belief 에 부여한 숫자입니다.

(ii-3) Node Title : Belief에 해당하는 개념어입니다.

(c-5) Event

문서내 관계망의 하위 메뉴로 나타나는 Event는 TONK에서 선택한 온톨로지 입니다. size는 Event에 해당하는 개념어의 수입니다. Event는 TONK의 **분석—문서내 관계망 생성 — 문서내 관계망 분석-온톨로지** 타입에서 사용자가 Event를 선택하였기 때문에 Event에 해당하는 개념어의 정보가 나타납니다. 만약 Event를 선택하지 않았을 경우 하위 메뉴로 생성이 되지 않습니다. 주의할 점은 size가 0이 나왔을 경우, 사용자가 TONK에서 온톨로지 타입을 선택하지 않은 것이 아니라, 온톨로지에 해당하는 개념어가 없다는 뜻입니다.

Event를 클릭하면 우측 Editor란에 관련정보가 나타나며 Info와 Editor탭으로 구성되어 있습니다.



i. Info

선택된 Node의 기본 정보를 나타냅니다.



(i-1) Node Class ID : 선택된 Node 의 아이디를 나타내며 아이디를 변경할 수 있습니다.

(i-2) Node Class Type : 선택된 Node 에 해당되는 온톨로지입니다.

(i-3) Node Count : 해당 온톨로지에 존재하는 총 개념어 수 입니다.

(i-4) Attributes names : 저자의 특성치 목록입니다.

ii. Editor

The screenshot shows the 'Editor' window for the 'Event' node class. The 'Editor' tab is selected. The window includes a search bar ('Find') and a filter dropdown ('AND'). A table lists nodes with columns for 'Node ID' and 'Node Title'. The table contains the following data:

Node ID	Node Title
<set filter>	<set filter>
2010-01-17	2010-01-17
16:29:30	16:29:30
17일	17일
지난 15일	지난 15일
지방선거	지방선거
1090	1090
6월2일	6월2일
12시 10분경부터	12시 10분경부터
3.5초	3.5초

At the bottom, there are buttons for 'Select All' and 'Clear All', and a status message: '0 item(s) selected, 4612 visible, 4612 total.' To the right, there are two vertical toolbars: 'Nodes' (Create, Move, Delete, Merge, Clean) and 'Attributes' (Create, Delete).

(ii-1) Find : 사용자가 원하는 특정 Node Id 및 Node Title 을 검색할 수 있습니다.

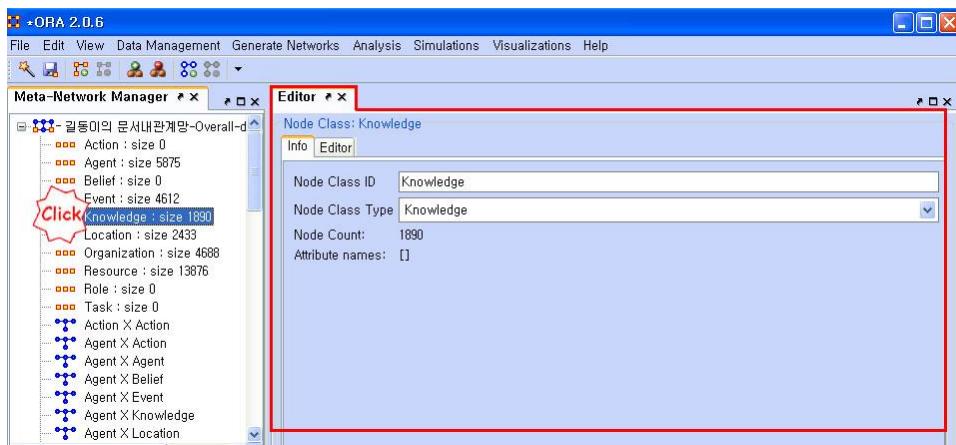
(ii-2) Node ID : 프로그램이 Event 에 부여한 숫자입니다.

(ii-3) Node Title : Event 에 해당하는 개념어입니다.

(c-6) Knowledge

문서내 관계망의 하위 메뉴로 나타나는 Knowledge 는 TONK 에서 선택한 온톨로지 입니다. size 는 Knowledge 에 해당하는 개념어의 수 입니다. Knowledge 는 TONK 의 **분석—문서내 관계망 생성 — 문서내 관계망 분석-온톨로지 탑입**에서 사용자가 Knowledge 를 선택하였기 때문에 Knowledge 에 해당하는 개념어의 정보가 나타납니다. 만약 Knowledge 를 선택하지 않았을 경우 하위 메뉴로 생성이 되지 않습니다. 주의할 점은 size 가 0 이 나왔을 경우, 사용자가 TONK 에서 온톨로지 탑입을 선택하지 않은 것이 아니라, 온톨로지에 해당하는 개념어가 없다는 뜻입니다.

Knowledge 를 클릭하면 우측 Editor 란에 관련정보가 나타나며 Info 와 Editor 탭으로 구성되어 있습니다.



i. Info

선택된 Node 의 기본 정보를 나타냅니다.



(i-1) Node Class ID : 선택된 Node 의 아이디를 나타내며 아이디를 변경할 수 있습니다.

(i-2) Node Class Type : 선택된 Node 에 해당되는 온톨로지입니다.

(i-3) Node Count : 해당 온톨로지에 존재하는 총 개념어 수 입니다.

(i-4) Attributes names : 저자의 특성치 목록입니다.

ii. Editor

Find:		AND
Node ID	Node Title	
<set filter>	<set filter>	
<input type="checkbox"/> 경제방에포털	경제방에포털	
<input type="checkbox"/> 지만아고라	지만아고라	
<input type="checkbox"/> K151	K151	
<input type="checkbox"/> 신동마	신동마	
<input type="checkbox"/> 에너지	에너지	
<input type="checkbox"/> 정부청사	정부청사	
<input type="checkbox"/> 인터뷰	인터뷰	
<input type="checkbox"/> 한국일보	한국일보	
<input type="checkbox"/> 동아일보	동아일보	

Nodes

Attributes

Buttons: Create, Move, Delete, Merge, Clean, Create, Delete.

Select All | Clear All | 0 item(s) selected, 1890 visible, 1890 total.

(ii-1) Find : 사용자가 원하는 특정 Node Id 및 Node Title 을 검색할 수 있습니다.

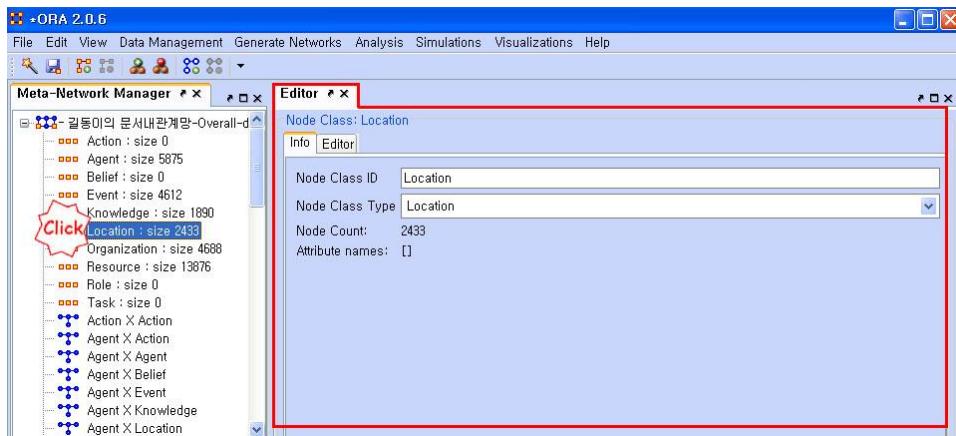
(ii-2) Node ID : Knowledge 에 부여한 숫자입니다.

(ii-3) Node Title : Knowledge 에 해당하는 개념어입니다.

(c-7) Location

문서내 관계망의 하위 메뉴로 나타나는 Location 은 TONK 에서 선택한 온톨로지 입니다. size 는 Location 에 해당하는 개념어의 수 입니다. Location 은 TONK 의 **분석—문서내 관계망 생성 — 문서내 관계망 분석- 온톨로지** 타입에서 사용자가 Location 을 선택하였기 때문에 Location 에 해당하는 개념어의 정보가 나타납니다. 만약 Location 을 선택하지 않았을 경우 하위메뉴로 생성이 되지 않습니다. 주의할 점은 size 가 0 이 나왔을 경우, 사용자가 TONK 에서 온톨로지 타입을 선택하지 않은 것이 아니라, 온톨로지에 해당하는 개념어가 없다는 뜻입니다.

Location 을 클릭하면 우측 Editor 란에 관련정보가 나타나며 Info 와 Editor 탭으로 구성되어 있습니다.



i. Info

선택된 Node 의 기본 정보를 나타냅니다.



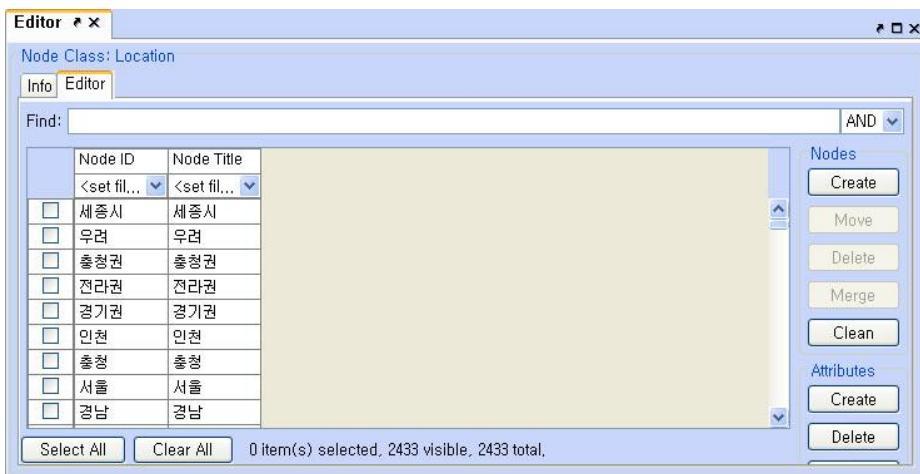
(i-1) Node Class ID : 선택된 Node 의 아이디를 나타내며 아이디를 변경할 수 있습니다.

(i-2) Node Class Type : 선택된 Node 에 해당되는 온톨로지입니다.

(i-3) Node Count : 해당 온톨로지에 존재하는 총 개념이 수입니다.

(i-4) Attributes names : 저자의 특성치 목록입니다.

ii. Editor



(ii-1) Find : 사용자가 원하는 특정 Node Id 및 Node Title 을 검색할 수 있습니다.

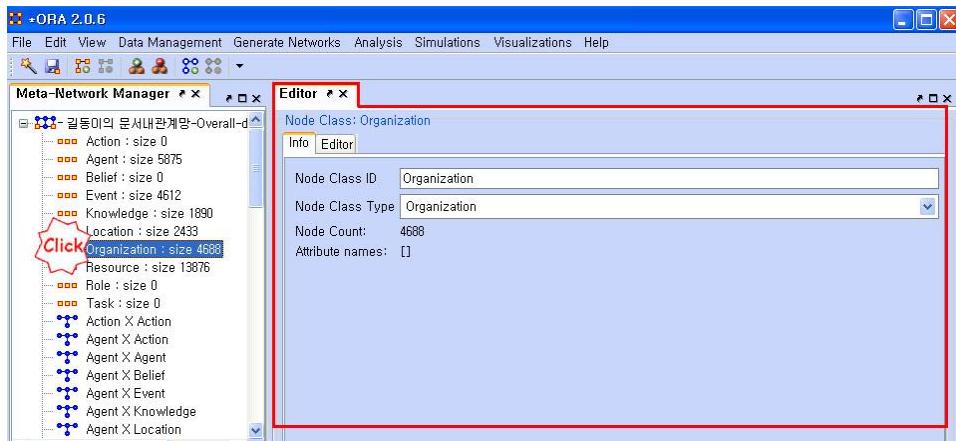
(ii-2) Node ID : Location 에 부여한 숫자입니다.

(ii-3) Node Title : Location 에 해당하는 개념어입니다.

(c-8) Organization

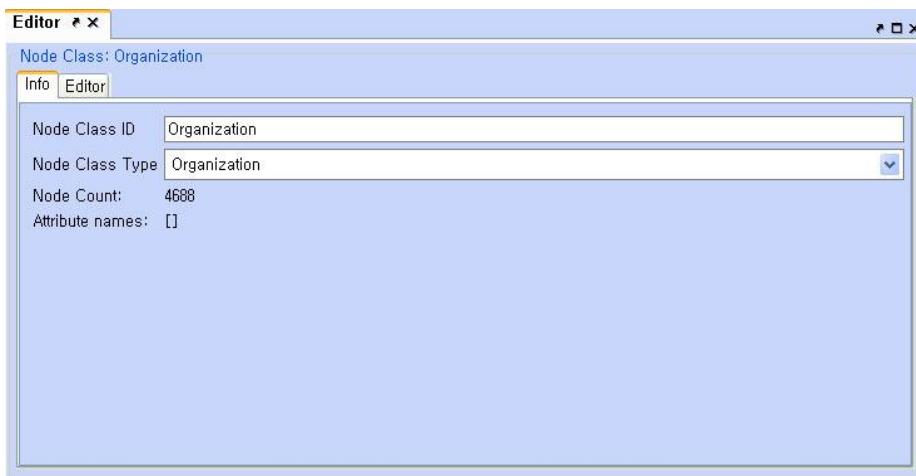
문서내 관계망의 하위 메뉴로 나타나는 Organization 은 TONK에서 선택한 온톨로지 입니다. size 는 Action 에 해당하는 개념어의 수 입니다. Organization 은 TONK 의 **분석—문서내 관계망 생성 — 문서내 관계망 분석-온톨로지 탑**에서 사용자가 Organization 을 선택하였기 때문에 Organization 에 해당하는 개념어의 정보가 나타납니다. 만약 Organization 을 선택하지 않았을 경우 하위메뉴로 생성이 되지 않습니다. 주의할 점은 size 가 0 이 나왔을 경우, 사용자가 TONK 에서 온톨로지 탑을 선택하지 않은 것이 아니라, 온톨로지에 해당하는 개념어가 없다는 뜻입니다.

Organization 을 클릭하면 우측 Editor 란에 관련정보가 나타나며 Info 와 Editor 탭으로 구성되어 있습니다.



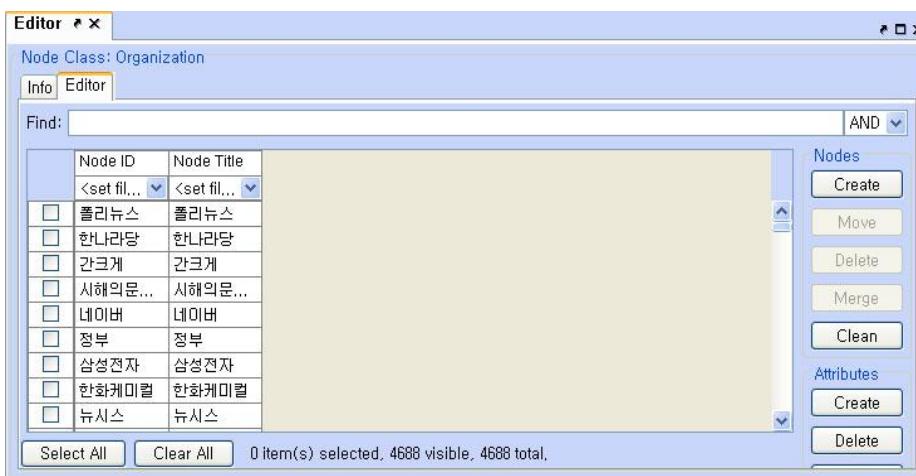
i. Info

선택된 Node 의 기본 정보를 나타냅니다.



- (i-1) Node Class ID : 선택된 Node 의 아이디를 나타내며 아이디를 변경할 수 있습니다.
- (i-2) Node Class Type : 선택된 Node 에 해당되는 온톨로지입니다.
- (i-3) Node Count : 해당 온톨로지에 존재하는 총 개념이 수입니다.
- (i-4) Attributes names : 저자의 특성치 목록입니다.

ii. Editor



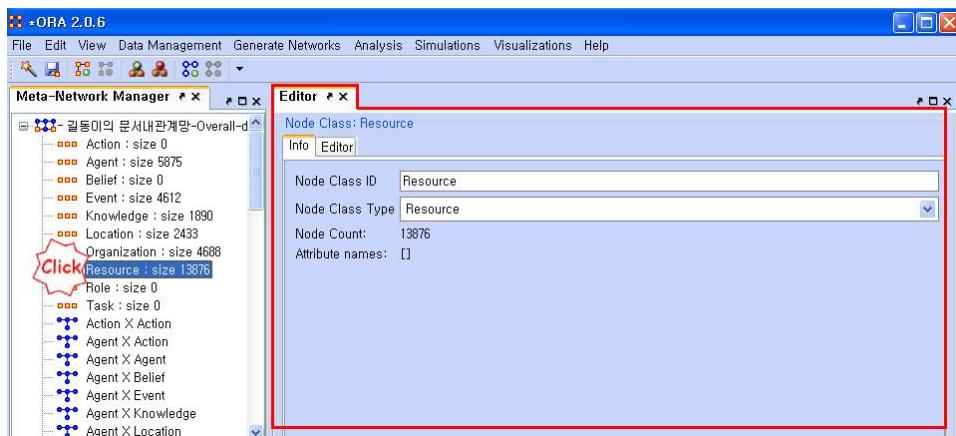
- (ii-1) Find : 사용자가 원하는 특정 Node Id 및 Node Title 을 검색할 수 있습니다.
- (ii-2) Node ID : Organization 에 부여한 숫자입니다.

(ii-3) Node Title : Organization 에 해당하는 개념어입니다.

(c-9) Resource

문서내 관계망의 하위 메뉴로 나타나는 Resource 는 TONK 에서 선택한 온톨로지 입니다. size 는 Resource 에 해당하는 개념어의 수 입니다. Resource 는 TONK 의 **분석—문서내 관계망 생성 — 문서내 관계망 분석-온톨로지** 타입에서 사용자가 Resource 를 선택하였기 때문에 Resource 에 해당하는 개념어의 정보가 나타납니다. 만약 Resource 를 선택하지 않았을 경우 하위 메뉴로 생성이 되지 않습니다. 주의할 점은 size 가 0 이 나왔을 경우, 사용자가 TONK 에서 온톨로지 타입을 선택하지 않은 것이 아니라, 온톨로지에 해당하는 개념어가 없다는 뜻입니다.

Resource 를 클릭하면 우측 Editor 란에 관련정보가 나타나며 Info 와 Editor 탭으로 구성되어 있습니다.



i. Info

선택된 Node 의 기본 정보를 나타냅니다.



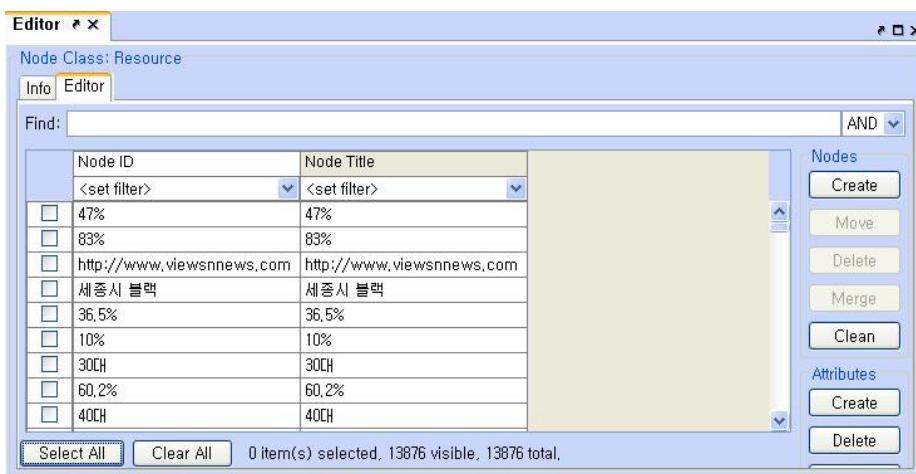
(i-1) Node Class ID : 선택된 Node 의 아이디를 나타내며 아이디를 변경할 수 있습니다.

(i-2) Node Class Type : 선택된 Node 에 해당되는 온톨로지입니다.

(i-3) Node Count : 해당 온톨로지에 존재하는 총 개념어 수 입니다.

(i-4) Attributes names : 저자의 특성치 목록입니다.

ii. Editor



(ii-1) Find : 사용자가 원하는 특정 Node Id 및 Node Title 을 검색할 수 있습니다.

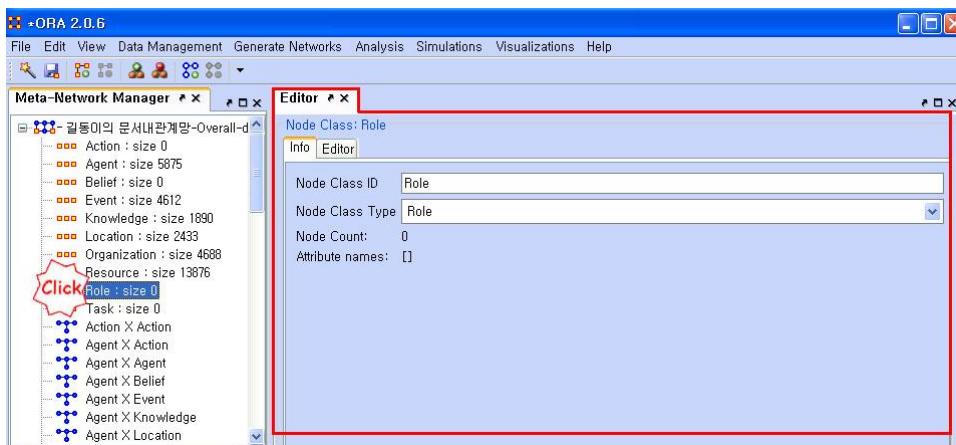
(ii-2) Node ID : Resource 에 부여한 숫자입니다.

(ii-3) Node Title : Resource에 해당하는 개념어입니다.

(c-10) Role

문서내 관계망의 하위 메뉴로 나타나는 Role은 TONK에서 선택한 온톨로지 입니다. size는 Role에 해당하는 개념어의 수입니다. Role은 TONK의 **분석—문서내 관계망 생성 — 문서내 관계망 분석-온톨로지** 탑입에서 사용자가 Role을 선택하였기 때문에 Role에 해당하는 개념어의 정보가 나타납니다. 만약 Role을 선택하지 않았을 경우 하위 메뉴로 생성이 되지 않습니다. 주의할 점은 size가 0이 나왔을 경우, 사용자가 TONK에서 온톨로지 탑입을 선택하지 않은 것이 아니라, 온톨로지에 해당하는 개념어가 없다는 뜻입니다.

Role을 클릭하면 우측 Editor란에 관련정보가 나타나며 Info와 Editor탭으로 구성되어 있습니다.



i. Info

선택된 Node의 기본 정보를 나타냅니다.



- (i-1) Node Class ID : 선택된 Node 의 아이디를 나타내며 아이디를 변경할 수 있습니다.
- (i-2) Node Class Type : 선택된 Node 에 해당되는 온톨로지입니다.
- (i-3) Node Count : 해당 온톨로지에 존재하는 총 개념이 수입니다.
- (i-4) Attributes names : 저자의 특성치 목록입니다.

ii. Editor

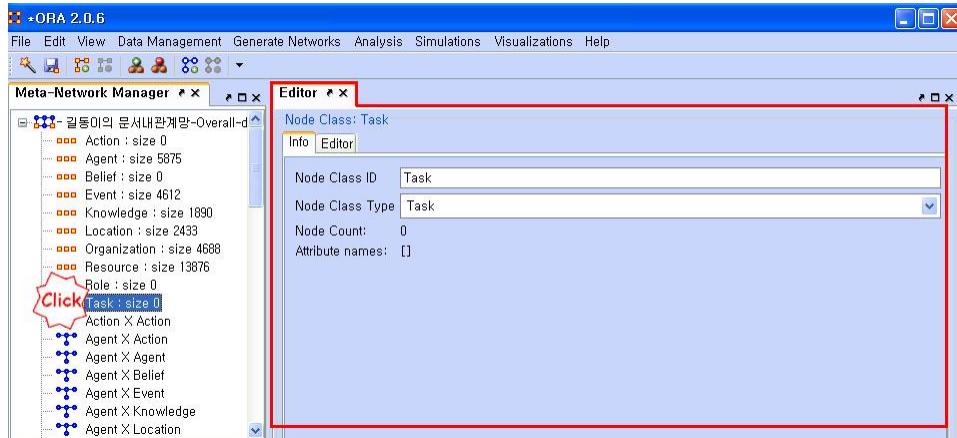


- (ii-1) Find : 사용자가 원하는 특정 Node Id 및 Node Title 을 검색할 수 있습니다.
- (ii-2) Node ID : Role 에 부여한 숫자입니다.
- (ii-3) Node Title : Role 에 해당하는 개념이 입니다.

(c-11)Task

문서내 관계망의 하위 메뉴로 나타나는 Task는 TONK에서 선택한 온톨로지 입니다. size는 Task에 해당하는 개념어의 수입니다. Task는 TONK의 **분석—문서내 관계망 생성 — 문서내 관계망 분석- 온톨로지 탑**에서 사용자가 Task를 선택하였기 때문에 Task에 해당하는 개념어의 정보가 나타납니다. 만약 Task를 선택하지 않았을 경우 하위 메뉴로 생성이 되지 않습니다. 주의할 점은 size가 0이 나왔을 경우, 사용자가 TONK에서 온톨로지 탑을 선택하지 않은 것이 아니라, 온톨로지에 해당하는 개념어가 없다는 뜻입니다.

Task를 클릭하면 우측 Editor란에 관련정보가 나타나며 Info와 Editor탭으로 구성되어 있습니다.



i. Info

선택된 Node의 기본 정보를 나타냅니다.



(i-1) Node Class ID : 선택된 Node 의 아이디를 나타내며 아이디를 변경할 수 있습니다.

(i-2) Node Class Type : 선택된 Node 에 해당되는 온톨로지입니다.

(i-3) Node Count : 해당 온톨로지에 존재하는 총 개념어 수 입니다.

(i-4) Attributes names : 저자의 특성치 목록입니다.

ii. Editor



(ii-1) Find : 사용자가 원하는 특정 Node Id 및 Node Title 을 검색할 수 있습니다.

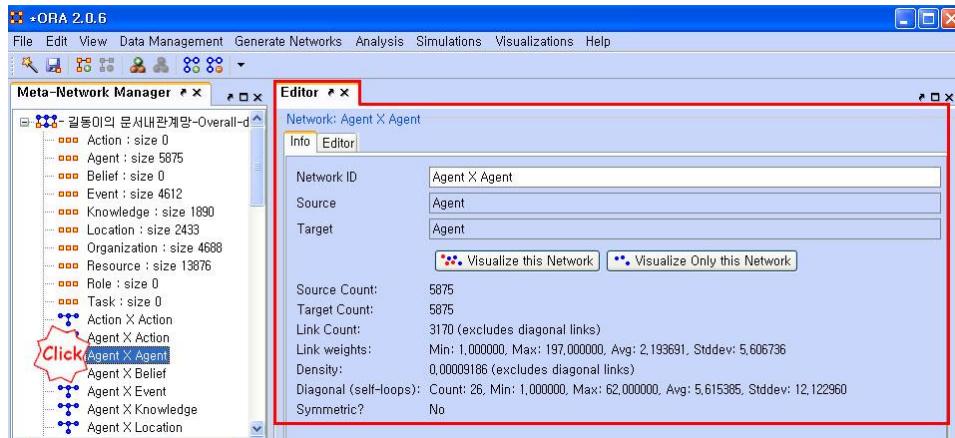
(ii-2) Node ID : Task 에 부여한 숫자입니다.

(ii-3) Node Title : Task 에 해당하는 개념이입니다.

(c-12) Network

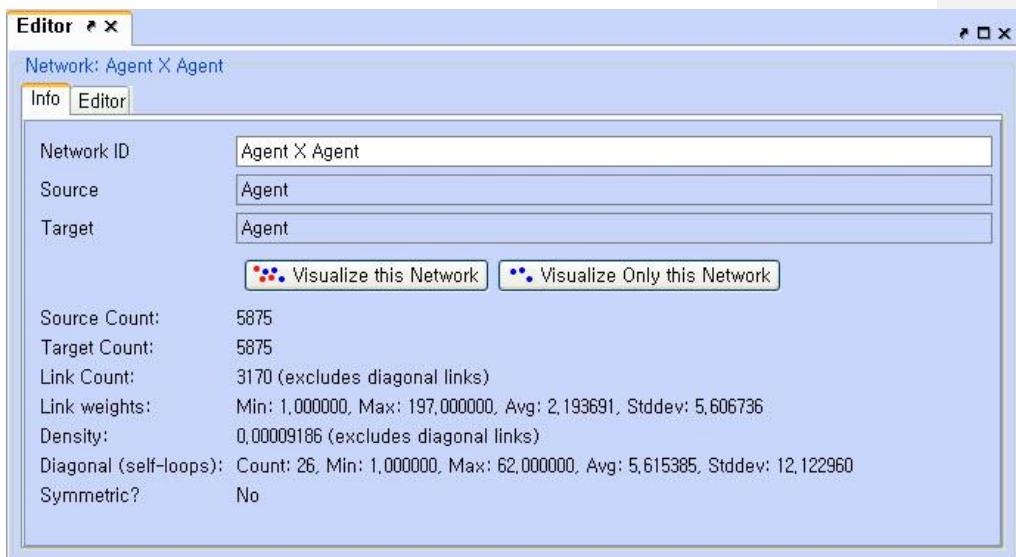
문서내 관계망의 하위 메뉴로 나타나는 Network 에는 총 55 가지가 있으며 페이지 357 을 참조합니다. 이 중 대표적으로 Agent 와 Agent 와의 관계를 알아보겠습니다.

Agent X Agent 를 클릭하면 우측 Editor 란에 관련정보가 나타나며 Info 와 Editor 탭으로 구성되어 있습니다.



i. Info

선택된 Network 의 기본 정보를 나타냅니다.



(i-1) Network ID : 문서내 관계망의 아이디를 나타내며 변경할 수 있습니다.

(i-2) Source : 연관 관계의 시작점입니다.

(i-3) Target : 연관 관계의 도착점입니다.

- 문서내 관계망의 모든 연관 관계를 보여줍니다.

- 선택된 Network 의 연관 관계를 보여줍니다.

(i-4) Source Count : 연관관계의 시작점의 개수입니다.

(i-5) Target Count : 연관관계의 도착점의 개수입니다.

(i-6) Link Count : agent-to-agent 네트워크에 존재하는 총 연관관계의 수 입니다.

(i-7) Link weights : 연관관계의 중요도입니다.

(i-8) Density : 네트워크의 조밀도입니다.

ii. Editor

(ii-1) Binary view : 가중치가 적용되지 않은 두 Node 간의 관계를 보고자 할 때 사용되며 Yes 혹은 No 를 나타내는 체크박스의 형태로 나타냅니다. Source 가 행이며 Target 이 열입니다.

Editor

Network: Agent X Agent

Info Editor

Binary View Numeric View Row count = 5875 Column count = 5875

	박근혜	김혜영	노무현	최신호	미성기	SCHOTT	Misubishi	있었던적
박근혜	✓							
김혜영		✓						
노무현	✓		✓					
최신호					✓			
미성기								
SCHOTT							✓	
Misubishi								
있었던적								
윤석열								
정								
노			✓					

Convert Links ▾ Remove Links Symmetrize Remove Self Loops

여기서 말하는 가중치는 ‘저자-이슈-문서 네트워크’의 가중치와는 다른 것으로 **두 개념어가 동시에 등장한 횟수입니다.** 저자-이슈-문서 네트워크’에서의 가중치는 저자가 쓴 글의 댓글 수, 조회 수, 찬성 수 등에 부여 합니다. 하지만 문서내 관계망에서의 가중치는 두 개념어 사이의 연결 관계에 가중치를 부여합니다.

(ii-2) Numeric view : 가중치가 적용된 두 Node 간의 관계를 보고자 할 때 사용되며 수치로 나타냅니다. Source 가 행이며 Target 이 열입니다.

Editor

Network: Agent X Agent

Info Editor

Binary View Numeric View Row count = 5875 Column count = 5875

	박근혜	김혜영	노무현	최신호	미성기	SCHOTT	Misubishi	있었던적
박근혜	3.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
김혜영	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
노무현	8.0	0.0	62.0	0.0	0.0	0.0	0.0	0.0
최신호	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
미성기	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
SCHOTT	0.0	0.0	0.0	0.0	0.0	0.0	2.0	0.0
Misubishi	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
있었던적	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
윤석열	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
정	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
노	0.0	0.0	3.0	0.0	0.0	0.0	0.0	0.0

Convert Links ▾ Remove Links Symmetrize Remove Self Loops

● 기타 네트워크

	Network 종류	상세 정보
1	Action X Action	Action 에 해당하는 개념어와 Action 에 해당하는 개념어 사이의 연결 관계
2	Agent X Action	Agent 에 해당하는 개념어와 Action 에 해당하는 개념어 사이의 연결 관계
3	Agent X Agent	Agent 에 해당하는 개념어와 Agent 에 해당하는 개념어 사이의 연결 관계
4	Agent X Belief	Agent 에 해당하는 개념어와 Belief 에 해당하는 개념어 사이의 연결 관계
5	Agent X Event	Agent 에 해당하는 개념어와 Event 에 해당하는 개념어 사이의 연결 관계
6	Agent X Knowledge	Agent 에 해당하는 개념어와 Knowledge 에 해당하는 개념어 사이의 연결 관계
7	Agent X Location	Agent 에 해당하는 개념어와 Location 에 해당하는 개념어 사이의 연결 관계
8	Agent X Organization	Agent 에 해당하는 개념어와 Organization 에 해당하는 개념어 사이의 연결 관계
9	Agent X Resource	Agent 에 해당하는 개념어와 Resource 에 해당하는 개념어 사이의 연결 관계
10	Agent X Role	Agent 에 해당하는 개념어와 Role 에 해당하는 개념어 사이의 연결 관계
11	Agent X Task	Agent 에 해당하는 개념어와 Task 에 해당하는 개념어 사이의 연결 관계
12	Belief X Action	Belief 에 해당하는 개념어와 Action 에 해당하는 개념어 사이의 연결 관계
13	Belief X Belief	Belief 에 해당하는 개념어와 Belief 에 해당하는 개념어 사이의 연결 관계
14	Belief X Role	Belief 에 해당하는 개념어와 Role 에 해당하는 개념어 사이의 연결 관계
15	Event X Action	Event 에 해당하는 개념어와 Action 에 해당하는 개념어 사이의 연결 관계
16	Event X Belief	Event 에 해당하는 개념어와 Belief 에 해당하는 개념어 사이의 연결 관계
17	Task X Role	Task 에 해당하는 개념어와 Role 에 해당하는 개념어 사이의 연결 관계
18	Task X Task	Task 에 해당하는 개념어와 Task 에 해당하는 개념어 사이의 연결 관계
19	Task X Event	Task 에 해당하는 개념어와 Event 에 해당하는 개념어 사이의 연결 관계
20	Task X Location	Task 에 해당하는 개념어와 Location 에 해당하는 개념어 사이의 연결 관계
21	Task X Organization	Task 에 해당하는 개념어와 Organization 에 해당하는 개념어 사이의 연결 관계
22	Event X Event	Event 에 해당하는 개념어와 Event 에 해당하는 개념어 사이의 연결 관계
23	Event X Location	Event 에 해당하는 개념어와 Location 에 해당하는 개념어 사이의 연결 관계
24	Event X Organization	Event 에 해당하는 개념어와 Organization 에 해당하는 개념어 사이의 연결 관계
25	Event X Role	Event 에 해당하는 개념어와 Role 에 해당하는 개념어 사이의 연결 관계
26	Knowledge X Action	Knowledge 에 해당하는 개념어와 Action 에 해당하는 개념어 사이의 연결 관계
27	Knowledge X Belief	Knowledge 에 해당하는 개념어와 Belief 에 해당하는 개념어 사이의 연결 관계
28	Knowledge X Event	Knowledge 에 해당하는 개념어와 Event 에 해당하는 개념어 사이의 연결 관계

29	Knowledge X Knowledge	Knowledge 에 해당하는 개념어와 Knowledge 에 해당하는 개념어 사이의 연결 관계
30	Knowledge X Location	Knowledge 에 해당하는 개념어와 Location 에 해당하는 개념어 사이의 연결 관계
31	Knowledge X Organization	Knowledge 에 해당하는 개념어와 Organization 에 해당하는 개념어 사이의 연결 관계
32	Knowledge X Resource	Knowledge 에 해당하는 개념어와 Resource 에 해당하는 개념어 사이의 연결 관계
33	Knowledge X Role	Knowledge 에 해당하는 개념어와 Role 에 해당하는 개념어 사이의 연결 관계
34	Knowledge X Task	Knowledge 에 해당하는 개념어와 Task 에 해당하는 개념어 사이의 연결 관계
35	Location X Action	Location 에 해당하는 개념어와 Action 에 해당하는 개념어 사이의 연결 관계
36	Location X Belief	Location 에 해당하는 개념어와 Belief 에 해당하는 개념어 사이의 연결 관계
37	Location X Location	Location 에 해당하는 개념어와 Location 에 해당하는 개념어 사이의 연결 관계
38	Location X Organization	Location 에 해당하는 개념어와 Organization 에 해당하는 개념어 사이의 연결 관계
39	Location X Role	Location 에 해당하는 개념어와 Role 에 해당하는 개념어 사이의 연결 관계
40	Organization X Action	Organization 에 해당하는 개념어와 Action 에 해당하는 개념어 사이의 연결 관계
41	Organization X Belief	Organization 에 해당하는 개념어와 Belief 에 해당하는 개념어 사이의 연결 관계
42	Organization X Organization	Organization 에 해당하는 개념어와 Organization 에 해당하는 개념어 사이의 연결 관계
43	Organization X Role	Organization 에 해당하는 개념어와 Action 에 해당하는 개념어 사이의 연결 관계
44	Resource X Action	Resource 에 해당하는 개념어와 Action 에 해당하는 개념어 사이의 연결 관계
45	Resource X Belief	Resource 에 해당하는 개념어와 Belief 에 해당하는 개념어 사이의 연결 관계
46	Resource X Event	Resource 에 해당하는 개념어와 Event 에 해당하는 개념어 사이의 연결 관계
47	Resource X Location	Resource 에 해당하는 개념어와 Location 에 해당하는 개념어 사이의 연결 관계
48	Resource X Organization	Resource 에 해당하는 개념어와 Organization 에 해당하는 개념어 사이의 연결 관계
49	Resource X Resource	Resource 에 해당하는 개념어와 Resource 에 해당하는 개념어 사이의 연결 관계
50	Resource X Role	Resource 에 해당하는 개념어와 Role 에 해당하는 개념어 사이의 연결 관계
51	Resource X Task	Resource 에 해당하는 개념어와 Task 에 해당하는 개념어 사이의 연결 관계
52	Role X Action	Role 에 해당하는 개념어와 Action 에 해당하는 개념어 사이의 연결 관계

53	Role X Role	Role 에 해당하는 개념어와 Role 에 해당하는 개념어 사이의 연결 관계
54	Task X Action	Task 에 해당하는 개념어와 Action 에 해당하는 개념어 사이의 연결 관계
55	Task X Belief	Task 에 해당하는 개념어와 Belief 에 해당하는 개념어 사이의 연결 관계

부록

품사 리스트

	품 사 명	정 의
1	감탄사(ii)	말하는 이의 놀람·느낌·응답 등을 간단히 나타내는 말.
2	격조사(jc)	체언이나 용언의 명사형 아래에 붙어서, 그 체언이나 용언의 명사형이 문장 안에서 다른 말에 대하여 가지는 자리를 나타내는 조사.
3	고유명사(nr)	특정한 사물 및 사람의 고유 기호나 명칭.
4	관형사(mm)	체언(體言) 또는 체언형(形)을 수식하는 역할을 하는 품사(品詞).
5	관형형어미(etm)	관형사를 만들어주는 어미.
6	기호(s)	인간의 지식·의지·감정을 어떤 물리현상을 통하여 나타내는 하나의 표현형식.
7	대명사(np)	사람이나 사물, 장소나 방향을 직접 가리키는 기능을 하는 품사.
8	동사(pv)	사람이나 사물의 움직임 또는 작용을 나타내는 말로 문장의 주체가 되는 말의 서술어 기능을 하는 품사.
9	동사파생접미사(xsv)	동사파생어를 만들어주는 접미사.
10	명사파생접미사(xsn)	명사파생어를 만들어주는 접미사.
11	명사형어미 (etn)	문장에서 용언의 어간에 붙어 명사와 같은 기능을 수행하게 하는 어미. '-음', '-기' 따위가 있다.
12	미등록어(uk)	자연어 처리로 분석되지 않는 단어.
13	보조사(jx)	체언, 부사, 활용 어미 따위에 붙어서 어떤 특별한 의미를 더해 주는 조사. '은', '는', '도', '만', '까지', '마저', '조차', '부터' 따위가 있다.
14	보조용언(px)	본용언과 연결되어 그것의 뜻을 보충하는 역할을 하는 용언. '가지고 싶다'의 '싶다', '가게 되다'의 '되다' 따위이다.
15	사용자사전 등록 명사 (nk)	사용자가 임의로 등록한 명사.
16	선어말어미 (ep)	어말 어미 앞에 나타나는 어미. '-시-', '-옵-' 따위와 같이 높임법에 관한 것과 '-았-', '-는-', '-더-', '-겠-' 따위와 같이 시상(時相)에 관한 것이 있다.
17	속격조사 (jm)	문장 안에서, 앞에 오는 체언이 뒤에 오는 체언의 관형어임을 보이는 조사.
18	수사(nn)	사물의 수량이나 순서를 나타내는 단어의 한 부류.
19	연결어미(ec)	어간에 붙어 다음 말에 연결하는 구실을 하는 어미. '-게', '-고', '-(으)며', '-(으)면', '-(으)니', '-아/어', '-지' 따위가 있다.

20	의존명사(nb)	의미가 형식적이어서 다른 말 아래에 기대어 쓰이는 명사. '것', '따름', '뿐', '데' 따위가 있다.
21	일반부사 (mag)	용언 또는 다른 말 앞에 놓여 그 뜻을 분명하게 하는 품사. 활용하지 못하며 성분 부사와 문장 부사로 나뉜다. '매우', '가장', '과연', '그리고' 따위가 있다.
22	자립명사(nc)	다른 말의 도움을 받지 아니하고 단독으로 쓰일 수 있는 명사.
23	접두사(xp)	파생어를 만드는 접사로, 어떤 단어의 앞에 붙어 새로운 단어가 되게 하는 말. '맨손'의 '맨-', '들볶다'의 '들-', '시퍼렇다'의 '시-' 따위가 있다.
24	접속부사 (maj)	앞의 체언이나 문장의 뜻을 뒤의 체언이나 문장에 이어 주면서 뒤의 말을 꾸며 주는 부사. '그러나', '즉', '또는', '및' 따위가 있다
25	접속조사 (jj)	두 단어를 같은 자격으로 이어 주는 구실을 하는 조사. '와', '과', '하고', '(이)나', '(이)랑' 따위가 있다.
26	종결어미 (ef)	한 문장을 종결되게 하는 어말 어미
27	지정사 (co)	무엇이 무엇이라고 지정하는 단어로, 긍정의 '이다'와 부정의 '아니다'가 있다.
28	형용사 (pa)	사물의 성질이나 상태를 나타내는 품사
29	형용사 파생 접미사(xsm)	사랑스러운, 그런 단어. - 스러운 접미사로 형용사를 만들어주는 것