



국립국어원 오픈소스 어휘 정보 활용

류창우 <cwryu@debian.org>, KOSSLab 4기



소개

- 프로젝트 중 사용된 외부 오픈소스 데이터에 대한 이야기
 - 공공 저작물에 대한 이야기
-
- KOSSLab **한국어 맞춤법/문법 검사** 프로젝트 진행 중입니다.
 - @debian.org 메일 주소로 알 수 있듯 Debian 개발자
 - 그래서 리눅스 배포판에 포함될 수 있는 오픈소스 라이선스에 신경을 많이 씁니다.



저작권법 제24조의2 공공저작물 관련

- <http://www.law.go.kr/법령/저작권법/제24조의2> (2013.12.30 신설, 2014.7.1 시행)
- 국가/지방자치단체..
 - ..에서 만든 저작물이거나 (즉 공무원이 만든)
 - ..가 계약에 따라 저작권을 100% 보유한 저작물은
- “허락 없이 이용할 수 있다” -> ??
- 예외
 - 국가안전보장
 - 사생활이나 사업상 비밀
 - 다른 법률에서 제한
 - 저작권 위원회에 등록된 국유재산/공유재산

"허락 없이 이용할 수 있다" -> 공공누리 (KOGL)

- <http://www.kogl.or.kr/>
- 출처 표기 의무.
- 비상업적 이용(NC), 2차 저작물 작성 금지(ND) 옵션 가능
- Creative Commons를 노골적으로 베낀 티가 납니다 :-)
- 공공저작물은 NC/ND 제한 없는 제1유형을 선택해야 함
- 큰 논란 없이 오픈소스일 것 같지만 국제적으로 인정 필요 (OSI 등)





주위의 공공 저작물

- 정부나 지자체에서 만든 문서, 사진, 영상 등의 저작물
- 글꼴 (저작권을 소유했다면)
- SW (저작권을 소유했다면)



저작권 정책

1. 저작권법 제24조의2(공공저작물의 자유이용)에 따라 청와대 홈페이지에서 저작재산권의 전부를 보유한 저작물의 경우에는 별도의 이용허락 없이 무료로 자유이용이 가능합니다.

단, 자유이용이 가능한 저작물은 "공공저작물 자유이용허락 표시 기준(공공누리, KOGL) 제1유형 : 출처표시"를 부착하여 개방하고 있으므로 공공누리 표시가 부착된 저작물인지를 확인한 이후에 자유롭게 이용하시기 바랍니다. 자유이용의 경우에는 반드시 저작물의 출처를 구체적으로 표시하여야 합니다.

제 1유형 : 출처표시

- 출처표시
- 상업적, 비상업적 이용 가능
- 변형 등 2차적 저작물 작성 가능



2. 청와대 홈페이지에는 청와대가 저작권 전부를 갖고 있지 아니한 자료도 제공되고 있습니다. 즉, 공공누리가 부착되지 않은 자료들을 사용하고자 할 경우에는 청와대 홈페이지 담당자와 사전에 협의한 이후에 이용하여 주시기 바랍니다.

알림마당

공지사항

보도자료

행사알림

국회매거진

국회정보서비스

정책참고자료실

🔍 국회의원 검색

검색

공지사항

메인 > 알림마당 > 공지사항



대한민국 국회홈페이지에서 알려드리는 소식입니다.

주간식단표(2017.11.20 ~11.26)

번호 | 3990 부서명 | 운영지원과 작성일자 | 2017-11-17 조회수 | 123

주간식단표(2017.11.20 ~11.26)



공공누리 공공저작물 자유이용허락

해당 저작물은 '공공누리' 출처표시-상업적 이용가능-변형가능 조건에 따라 이용할 수 있습니다.

첨부파일 | 📎 ◆주간식단표(2017.11.20 ~11.26).hwp

📎 ◆주간식단표(2017.11.20 ~11.26).pdf

↑ TOP

주 간 식 단 표

2017.11.20 ~ 11.26

연	본관				외관				도서관	현정기념관 (미가)
	본관3식당(5)연대그라운드		본관2식당		외관3식당(관악)		외관2식당(관악)			
	중식1	중식2	본관2식당	본관1식당	중식1	중식2	점심	점심	외관1식당	
20 (월)	콩비지찌개 갈치레구이 도토리묵산채무침 이색/물/부름김치 <11,000>-828kcal	해물볶음우동 창개밥/미소장국 고려대떡/당면발효된 생채/연두부/부름김치 <11,000>-880kcal	갈치죽백기 현미밥/각두기 미도떡/물/부름김치 생채/연두부/부름김치 깨끗이수두부무침/굴 <6,600>-867kcal	대파육개장 창개밥/미소장국 고려대떡/당면발효된 생채/연두부/부름김치 깨끗이수두부무침/굴 816kcal	외관3식당(관악) 대파육개장 창개밥/미소장국 고려대떡/당면발효된 생채/연두부/부름김치 깨끗이수두부무침/굴 <10,000>-954kcal	외관3식당(관악) 대파육개장 창개밥/미소장국 고려대떡/당면발효된 생채/연두부/부름김치 깨끗이수두부무침/굴 <10,000>-954kcal	점심 대파육개장 창개밥/미소장국 고려대떡/당면발효된 생채/연두부/부름김치 깨끗이수두부무침/굴 <10,000>-954kcal	점심 대파육개장 창개밥/미소장국 고려대떡/당면발효된 생채/연두부/부름김치 깨끗이수두부무침/굴 <10,000>-954kcal	외관1식당 대파육개장 창개밥/미소장국 고려대떡/당면발효된 생채/연두부/부름김치 깨끗이수두부무침/굴 <10,000>-954kcal	현정기념관 (미가) 대파육개장 창개밥/미소장국 고려대떡/당면발효된 생채/연두부/부름김치 깨끗이수두부무침/굴 <10,000>-954kcal
21 (화)	마늘보쌈 창개밥/미소장국 고려대떡/당면발효된 생채/연두부/부름김치 깨끗이수두부무침/굴 <11,000>-884kcal	불고기/미소장국 창개밥/미소장국 고려대떡/당면발효된 생채/연두부/부름김치 깨끗이수두부무침/굴 <11,000>-891kcal	죽백기 현미밥/각두기 미도떡/물/부름김치 생채/연두부/부름김치 깨끗이수두부무침/굴 <6,600>-834kcal	동등어김치찌개 창개밥/미소장국 고려대떡/당면발효된 생채/연두부/부름김치 깨끗이수두부무침/굴 854kcal	외관3식당(관악) 죽백기 현미밥/각두기 미도떡/물/부름김치 생채/연두부/부름김치 깨끗이수두부무침/굴 <10,000>-952kcal	외관3식당(관악) 죽백기 현미밥/각두기 미도떡/물/부름김치 생채/연두부/부름김치 깨끗이수두부무침/굴 <10,000>-952kcal	점심 죽백기 현미밥/각두기 미도떡/물/부름김치 생채/연두부/부름김치 깨끗이수두부무침/굴 <10,000>-952kcal	점심 죽백기 현미밥/각두기 미도떡/물/부름김치 생채/연두부/부름김치 깨끗이수두부무침/굴 <10,000>-952kcal	외관1식당 죽백기 현미밥/각두기 미도떡/물/부름김치 생채/연두부/부름김치 깨끗이수두부무침/굴 <10,000>-952kcal	현정기념관 (미가) 죽백기 현미밥/각두기 미도떡/물/부름김치 생채/연두부/부름김치 깨끗이수두부무침/굴 <10,000>-952kcal
22 (수)	우거지해장국 모듬버섯/창개밥 고려대떡/당면발효된 생채/연두부/부름김치 깨끗이수두부무침/굴 <11,000>-856kcal	배달할국수 창개밥/미소장국 고려대떡/당면발효된 생채/연두부/부름김치 깨끗이수두부무침/굴 <11,000>-847kcal	◆아침식단 치킨치즈까스 파프리카볶음밥 생크림양파소프 양송이/양파/양송이 가치피자/피클 <6,600>-894kcal	수상갈비양식 창개밥/미소장국 고려대떡/당면발효된 생채/연두부/부름김치 깨끗이수두부무침/굴 837kcal	◆아침식단 치킨치즈까스 파프리카볶음밥 생크림양파소프 양송이/양파/양송이 가치피자/피클 <10,000>-961kcal	외관3식당(관악) 수상갈비양식 창개밥/미소장국 고려대떡/당면발효된 생채/연두부/부름김치 깨끗이수두부무침/굴 <10,000>-961kcal	외관3식당(관악) 수상갈비양식 창개밥/미소장국 고려대떡/당면발효된 생채/연두부/부름김치 깨끗이수두부무침/굴 <10,000>-961kcal	점심 수상갈비양식 창개밥/미소장국 고려대떡/당면발효된 생채/연두부/부름김치 깨끗이수두부무침/굴 <10,000>-961kcal	외관1식당 수상갈비양식 창개밥/미소장국 고려대떡/당면발효된 생채/연두부/부름김치 깨끗이수두부무침/굴 <10,000>-961kcal	현정기념관 (미가) 수상갈비양식 창개밥/미소장국 고려대떡/당면발효된 생채/연두부/부름김치 깨끗이수두부무침/굴 <10,000>-961kcal
23 (목)	닭개장 가자미구이/우동 창개밥/미소장국 고려대떡/당면발효된 생채/연두부/부름김치 깨끗이수두부무침/굴 <11,000>-862kcal	날치알술비빔밥 창개밥/미소장국 고려대떡/당면발효된 생채/연두부/부름김치 깨끗이수두부무침/굴 <11,000>-830kcal	구절판 현미밥/우동 창개밥/미소장국 고려대떡/당면발효된 생채/연두부/부름김치 깨끗이수두부무침/굴 <6,600>-812kcal	수제함박스테이크 창개밥/미소장국 고려대떡/당면발효된 생채/연두부/부름김치 깨끗이수두부무침/굴 862kcal	외관3식당(관악) 닭개장 가자미구이/우동 창개밥/미소장국 고려대떡/당면발효된 생채/연두부/부름김치 깨끗이수두부무침/굴 <10,000>-955kcal	외관3식당(관악) 닭개장 가자미구이/우동 창개밥/미소장국 고려대떡/당면발효된 생채/연두부/부름김치 깨끗이수두부무침/굴 <10,000>-955kcal	점심 닭개장 가자미구이/우동 창개밥/미소장국 고려대떡/당면발효된 생채/연두부/부름김치 깨끗이수두부무침/굴 <10,000>-955kcal	점심 닭개장 가자미구이/우동 창개밥/미소장국 고려대떡/당면발효된 생채/연두부/부름김치 깨끗이수두부무침/굴 <10,000>-955kcal	외관1식당 닭개장 가자미구이/우동 창개밥/미소장국 고려대떡/당면발효된 생채/연두부/부름김치 깨끗이수두부무침/굴 <10,000>-955kcal	현정기념관 (미가) 닭개장 가자미구이/우동 창개밥/미소장국 고려대떡/당면발효된 생채/연두부/부름김치 깨끗이수두부무침/굴 <10,000>-955kcal
24 (금)	돌솥국 이육향국/아욱차 수제갈비/부름김치 이육향국/아욱차 수제갈비/부름김치 이육향국/아욱차 수제갈비/부름김치 <11,000>-869kcal	창개밥/미소장국 고려대떡/당면발효된 생채/연두부/부름김치 깨끗이수두부무침/굴 <11,000>-869kcal	혼제오리볶음 우동 창개밥/미소장국 고려대떡/당면발효된 생채/연두부/부름김치 깨끗이수두부무침/굴 <6,600>-851kcal	해산물잡채 창개밥/미소장국 고려대떡/당면발효된 생채/연두부/부름김치 깨끗이수두부무침/굴 823kcal	외관3식당(관악) 돌솥국 이육향국/아욱차 수제갈비/부름김치 이육향국/아욱차 수제갈비/부름김치 이육향국/아욱차 수제갈비/부름김치 <10,000>-957kcal	외관3식당(관악) 돌솥국 이육향국/아욱차 수제갈비/부름김치 이육향국/아욱차 수제갈비/부름김치 이육향국/아욱차 수제갈비/부름김치 <10,000>-957kcal	점심 돌솥국 이육향국/아욱차 수제갈비/부름김치 이육향국/아욱차 수제갈비/부름김치 이육향국/아욱차 수제갈비/부름김치 <10,000>-957kcal	점심 돌솥국 이육향국/아욱차 수제갈비/부름김치 이육향국/아욱차 수제갈비/부름김치 이육향국/아욱차 수제갈비/부름김치 <10,000>-957kcal	외관1식당 돌솥국 이육향국/아욱차 수제갈비/부름김치 이육향국/아욱차 수제갈비/부름김치 이육향국/아욱차 수제갈비/부름김치 <10,000>-957kcal	현정기념관 (미가) 돌솥국 이육향국/아욱차 수제갈비/부름김치 이육향국/아욱차 수제갈비/부름김치 이육향국/아욱차 수제갈비/부름김치 <10,000>-957kcal
25 (토)	주5일근무제 실시로 영업을 하지 않습니다.				주5일근무제 실시로 영업을 하지 않습니다.				임시휴무	
26 (일)	주5일근무제 실시로 영업을 하지 않습니다.				주5일근무제 실시로 영업을 하지 않습니다.				임시휴무	

본관1식당, 외관1식당, 도서관식당, 현정기념관식당 메뉴는 3,300원(단, 외부 방문객은 일반기준으로 4,500원)입니다.
식당별 주5일근무제 실시로 영업을 하지 않습니다. 본관3식당, 본관1식당, 외관3식당, 외관1식당, 도서관식당, 현정기념관식당 메뉴는 3,300원(단, 외부 방문객은 일반기준으로 4,500원)입니다.
외관3식당, 외관1식당, 도서관식당, 현정기념관식당 메뉴는 3,300원(단, 외부 방문객은 일반기준으로 4,500원)입니다.
외관3식당, 외관1식당, 도서관식당, 현정기념관식당 메뉴는 3,300원(단, 외부 방문객은 일반기준으로 4,500원)입니다.

* 식재료 구입량에 따라 원산지 변경이 당월에도 발생할 수 있습니다.

홈페이지 도우미

개선 안내

마일리지

마일리지 소개

마일리지 공지

뷰어프로그램다운로드

이메일무단수집거부

개인정보 처리방침

개인정보 처리방침

영상정보처리기기운영관리방침

저작권 정책

🖨 인쇄

· 저작권법 제24조의2(공공저작물의 자유이용)에 따라 서울특별시 홈페이지에서 제공하는 자료 중에서 서울특별시가 저작권사권의 전부를 보유한 저작물의 경우에는 별도의 이용허락 없이 자유이용이 가능합니다.

· 단, 자유이용이 가능한 자료는 **크리에이티브커먼즈라이선스(CCL)**를 부착하여 개방하고 있으므로 CCL 표시가 부착된 저작물인지를 확인한 이후에 자유롭게 이용하시기 바랍니다.



· 자유이용의 경우, 이용자께서는 저작물의 출처를 구체적으로 표시하셔야 하며, 서울특별시 홈페이지 제공 자료의 경우 상업적 이용이 금지되어 있으므로 비영리 목적으로만 이용가능하고, 저작물 변경 없이 그대로 이용해야 합니다.

· CCL 이 부착되지 않은 자료들을 사용하고자 할 경우에는 해당 저작물 담당부서와 사전에 협의한 이후에 이용하여 주시기 바랍니다.

저작권사용조건과 의미

표시

의미

tbs - Mozilla Firefox

파일(F) 편집(E) 보기(V) 이동(S) 북마크(B) 도구(T) 도움말(H)

tbs

tbs.seoul.kr/help/policy.do

서울특별시
가족소통과 아동발달을 위한 신림차유 프로그램운영
서울소식
응답소
정보공개
로그인
회원가입
자유게시판
음원심의 게시판
편성표

시민이 주인이되는 소통의 중심. 서울 수도권 대표 공익방송

tbs
tbs소개
건의신청

FM 95.1Mhz
eFM 101.3Mhz
TV
뉴스
교통정보

웹사이트 도우미
개인정보처리방침
공적자비리신고센터
저작권 정책
자유게시판
음원심의 게시판
공지 게시판
음부즈만 게시판

● 저작권 정책

!
tbs교통방송에서 제공하는 모든 자료는 저작권법에 의하여 보호받는 저작물로서 모든 권리는 tbs교통방송에 있습니다. 이용자는 아래의 저작권 보호정책을 준수하여야 합니다.

OPEN
공공누리
공공저작물 자유이용허락

1. 저작권법 제24조의2(공공저작물의 자유이용)에 따라 tbs교통방송 홈페이지에서 저작재산권의 전부를 보유한 저작물의 경우에는 별도의 이용허락 없이 무료로 자유이용이 가능합니다.
- 단, 자유이용이 가능한 저작물은 "공공저작물 자유이용허락 표시 기준(공공누리, KOGL) 제1유형 : 출처표시"를 부착하여 개방하고 있으며 공공누리 표시가 부착된 저작물인지를 확인한 이후에 자유롭게 이용하시기 바랍니다. 자유이용의 경우에는 반드시 저작물의 출처를 구체적으로 표시하여야 합니다.



국립국어원 자료는?

- 국가의 한국어/한글 정책 기관이니 방대한 자료를 제공
- 사전 - 표준국어대사전
- 말뭉치 등 - 21세기 세종계획
- 기타 보고서 등 자료

[말뭉치](#)[전자사전](#)[용어/문자찾기](#)[통합자료실](#)[회원교류](#)

이용안내

[통합검색](#)[말뭉치](#)[전자사전](#)[용어/문자 찾기](#)[통합자료실](#)[회원교류](#)

> 이용안내



저작권정책 | 이용안내

▶ 저작권 정책

본 누리집에서 별도로 명시되지 않은 모든 자료는 '크리에이티브 코먼스 저작자표시-비영리-변경금지 4.0 국제 라이선스'에 따라 배포·이용될 수 있습니다.

'크리에이티브 코먼스 저작자표시-비영리-변경금지 4.0 국제 라이선스'가 적용되는 본 누리집의 자료는

다만, 저작물을 이용하기 위해서는 다음의 조건을 지켜야 합니다.

- 저작자 표시 : 자료를 사용할 때 저작자를 필수로 표시해야 합니다.
- 영리 목적 사용금지 : 비영리적인 목적으로만 사용·배포해야 합니다.
- 저작물 변경금지 : 자료를 변경하지 않고 사용·배포해야 합니다.



하지만 대부분 자료는 공공누리¹도 오픈소스도 아님

- <https://ithub.korean.go.kr/>
- CC BY-NC-ND 또는 KOGL NC-ND
- 많이 사용하는 데이터임에도 불구하고 비영리-변경금지 제한.
- **넌센스** - "말뭉치" 데이터를 다른 용도로 가공하지 못하면 (ND) 이걸로 뭘 하라는 얘기?
- 공정 사용이 가능한 연구/교육 목적으로만 가능.
- 아마도 인터넷 서비스 용도로만 쓰면 배포 안 하니까 OK?
- NC는? 쓰시는 분들은 쉬쉬하면서 쓰고 계시다는 강한 의심이 듭니다. :-)
- 21세기 세종계획은 이미 10년 전 (2007) 종료 - 오류를 수정해도 배포할 수 없다?



ND 덕분에 오류 수정 사항 배포도 불편

제29회 한글 및 한국어 정보처리 학술대회 논문집 (2017년)

공개와 협업을 통한 세종 형태 분석 말뭉치 오류 개선 방법

한경은⁰, 백슬예, 임재수

(주)카카오

grace.han@kakaocorp.com, cecil.rosa@kakaocorp.com, jamie.lim@kakaocorp.com

Open Sourced and Collaborative Method to Fix Errors of Sejong

Morphologically Annotated Corpora

Gyeong-Eun Han⁰, Seul-Ye Baek, Jae-Soo Lim
KAKAO Corp



민원 결과 답변은: 저작권재산권 전부 소유가 아님

- 아마도 데이터 구축에 참여했던 학교, 연구자의(일부) 소유
- 국민 세금으로 구축한 데이터들도 활용이 제한되는 상황. 단순 보고서 조차도 NC ND 제한
- 자료 구축의 많은 부분을 외주에 의존하는 현실, 입찰 과정에서 저작권 문제를 명시해야 하는데...
- 21세기 세종계획 데이터는 지금부터라도 저작권재산권 양도 받아 라이선스 바뀌어야
- 문화체육관광부고시. 공공저작물 저작권 관리 및 이용 지침. 제6조
 - ② 공공기관 등은 국민의 자유이용이 바람직하다고 판단되는 공공저작물인 경우에는 제1항에 따른 계약 체결 시 2차적저작물작성권을 포함한 저작권재산권의 전부를 취득하도록 노력하여야 한다.
 - ③ 공공기관 등은 권리관계가 명확하지 않은 공공저작물에 대하여는 사후적인 권리처리를 통해 자유이용될 수 있도록 노력하여야 한다.



표준국어대사전은 저작권 등록물

- 포털 사이트 등에 라이선싱



건질 정보는 우리말샘 / 한국어기초사전 뿐

- 2016년 10월 오픈한 사전
- 표준국어대사전 이관 후 사용자 참여
- CC BY-SA 2.0 KR
- 우리말샘 - 사용자 참여 사전 <https://opendict.korean.go.kr/>
- 한국어기초사전 - 자매품, 5만여 단어 <https://krdict.korean.go.kr/>
- 외국어학습사전 - 외국인용 한국어기초사전 (영어, 일본어, 러시아어, 몽골어, 인도네시아어, 베트남어, 스페인어, 아랍어, 타이어, 프랑스어)



다운로드하기

- 우리말샘 (525MB TXT)
 - 450MB XML, XML에 버그로 정보가 누락, 다른 포맷도 버그
- 한국어기초사전(150MB XML)
- 외국어학습사전은다운로드기능 아직 없음



한국어 기초사전

세계인이 누리는 한국어 학습사전

[찾기](#)[자세히 찾기](#)

오늘의 한국어

이틀 명사 [이틀] ★★★

두 날.

[자세히 보기](#)

단어로 찾기

자소로 찾기

주제 및 상황 범주별 찾기

의미 범주별 찾기

구분

☒ 전체 ☒ 단어 ☒ 구 ☒ 문법 · 표현☒ 전체 ☒ 고유어 ☒ 한자어 ☒ 외래어 ☒ 혼종어

품사

☒ 전체☒ 명사 ☒ 대명사 ☒ 수사 ☒ 조사 ☒ 동사 ☒ 형용사☒ 관형사 ☒ 부사 ☒ 감탄사 ☒ 접사 ☒ 의존 명사 ☒ 보조 동사☒ 보조 형용사 ☒ 어미 ☒ 품사 없음 (어근, 줄어든 말 등)

등급별 어휘

☒ 전체 ☒ 초급 ☒ 중급 ☒ 고급 ☒ 없음

다중 매체 정보

☒ 전체 ☒ 사진 ☒ 삽화 ☒ 동영상 ☒ 애니메이션 ☒ 소리 ☒ 없음

의미 범주

전체 전체

주제 및 상황 범주

주제 및 상황 범주

음절수

 음절 ~ 음절 전체 1음절 2음절 3음절 4음절 5음절

찾을 대상

어휘 포함 문자

찾기

내가 살펴본 단어

한국어 발음 듣기 서비스의 일부는 네이버문화재단에서 후원하였습니다.

NAVER 문화재단

찾기

자세히 찾기 결과(총 51,948개)

📄 사전 내려받기

가 가

10개씩 보기 ▾

적용

가¹ 「명사」 [가: ㅈㅅ] ★

1. 어떤 장소나 물건의 둘레나 끝부분.
2. ‘주변’의 뜻을 나타내는 말.

가-¹

(가고, 가는데, 가, 가서, 가니, 가면, 간, 가는, 갈, 갑니다, 갔다, 가라) → 가다¹, 가다²


가² 「조사」

1. 어떤 상태나 상황에 놓인 대상이나 동작의 주체를 나타내는 조사.
2. 바뀌게 되는 대상이나 부정하는 대상임을 나타내는 조사.
3. 앞의 말을 강조하는 뜻을 나타내는 조사.

가-²

(가는데, 가니, 간, 가는, 갈, 갑니다) → 갈다¹, 갈다², 갈다³

자세히 찾기 결과(중)

가¹ 「명사」 [가: 

1. 어떤 장소나 물건

2. '주변'의 뜻을 나타

가-¹

(가고, 가는데, 가, 가

가² 「조사」

1. 어떤 상태나 상황

2. 바뀌게 되는 대상

3. 앞의 말을 강조하는

가-²

(가는데, 가니, 간, 가

가³ (可) 「명사」 [가

1. 회의 등에서 어떤

2. 성적이나 등급을 나

3. 어떤 행동이 가능하거나 허용됨.

☐ 학습 자료 용도

☐ 강의 자료 용도

☒ 연구 자료 용도

☐ 기타

내려받을 파일 형식 및 사전 항목을 선택해 주세요.

파일 형식	<input type="radio"/> MS 엑셀 <input type="radio"/> HWP 한글 <input type="radio"/> MS 워드 <input type="radio"/> 텍스트 <input checked="" type="radio"/> XML
필수 항목	어휘, 구분, 품사, 뜻풀이, 용례
선택 항목	<div><input checked="" type="checkbox"/> 전체</div> <div><input checked="" type="checkbox"/> 로마자 표기 <input checked="" type="checkbox"/> 원어 <input checked="" type="checkbox"/> 발음 <input checked="" type="checkbox"/> 활용</div> <div><input checked="" type="checkbox"/> 파생어 <input checked="" type="checkbox"/> 가봐라 <input checked="" type="checkbox"/> 어휘 등급 <input checked="" type="checkbox"/> 의미 범주</div> <div><input checked="" type="checkbox"/> 주제 및 상황 범주 <input checked="" type="checkbox"/> 전체 참고 <input checked="" type="checkbox"/> 검색용 이형태 <input checked="" type="checkbox"/> 관련어</div> <div><input checked="" type="checkbox"/> 의미 참고 <input checked="" type="checkbox"/> 다중 매체 정보 <input checked="" type="checkbox"/> 문형 <input checked="" type="checkbox"/> 문형 참고</div> <div><input checked="" type="checkbox"/> 부표제어 <input checked="" type="checkbox"/> 부표제어 전체 참고 <input checked="" type="checkbox"/> 부표제어 관련어 <input checked="" type="checkbox"/> 부표제어 의미 참고</div> <div><input checked="" type="checkbox"/> 부표제어 문형 <input checked="" type="checkbox"/> 부표제어 문형 참고</div>

* 내려받기 버튼을 누르면 내려받기 창이 뜹니다.

* 파일 내려받기 창이 뜨면 저장 버튼을 눌러 저장하십시오.

내려 받기

취소



회원 정보

내 단어장

회원 탈퇴

사용자 의견

사전 내려받기

쪽지 관리

사전 내려받기

• 사전 내려받기 파일은 완료 후 30일 동안 유지됩니다. 30일 이내에 파일을 내려 받으시기 바랍니다.

모두 8건의 게시글이 있습니다.

10개씩 보기

번호	요청 일시	어휘 수	진행률	구분	진행 상태	파일	내려받기
8	2017-11-06 17:31:32	51,948개	0% (0 / 51,948)	검색 결과	대기중		<button>내려받기취소</button>
7	2017-11-03 19:46:43	51,948개	100% (51,948 / 51,948)	검색 결과	완료		<button>내려받기</button>
6	2017-10-30 12:58:48	51,948개	100% (51,948 / 51,948)	검색 결과	완료		<button>내려받기</button>
5	2017-10-30 02:17:29	1개	100% (1 / 1)	검색 결과	완료		<button>내려받기</button>
4	2017-10-19 03:12:48	51,949개	100% (51,949 / 51,949)	검색 결과	완료		<button>내려받기</button>



XML은 ISO 24613:2008 LMF 포맷

- [https://en.wikipedia.org/wiki/Lexical Markup Framework](https://en.wikipedia.org/wiki/Lexical_Markup_Framework)
- pylmflib



헤더

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE LexicalResource SYSTEM "DTD_LMF_REV_16.dtd">
<LexicalResource dtdVersion="16">
  <GlobalInformation>
    <feat att="label" val="한국어기초사전 - 국립국어원 제공" />
    <feat att="creationDate" val="2017/11/03 08:02:33" />
    <feat att="languageCoding" val="ISO 639-3" />
  </GlobalInformation>
  <Lexicon>
    <feat att="language" val="kor" />
    <LexicalEntry att="id" val="27733">
      ...
```



단어

```
<LexicalEntry att="id" val="61329">
  <feat att="homonym_number" val="2" />
  <feat att="lexicalUnit" val="단어" />
  <feat att="partOfSpeech" val="명사" />
  <Lemma>
    <feat att="writtenForm" val="사전" />
  </Lemma>
  <feat att="origin" val="辭典" />
  ...
```



단어 설명

<Sense att="id" val="1">

...

<feat att="definition" val="낱말을 모아 일정한 차례에 따라 싣고, 그 발음, 뜻, 어원, 용법 등을 설명한 책." />

<SenseExample>

<feat att="type" val="구" />

<feat att="example" val="외국어 사전." />

</SenseExample>

...



관련 말 레퍼런스도 있고

<SenseRelation>

<feat att="type" val="반대말" />

<feat att="id" val="22109" />

<feat att="lemma" val="올리다" />

<feat att="homonymNumber" val="0" />

</SenseRelation>

<feat att="syntacticPattern" val="1이 2에서 3을 내리다" />

<feat att="definition" val="위에 있는 물건을 아래로 옮기다." />

발음, 사진, 동영상도 있고 (단 오픈소스 아님)

```
<WordForm>
  <feat att="type" val="발음" />
  <feat att="pronunciation" val="사전" />
  <feat att="sound" val="http://dicmedia.korean.go.kr:8899/multimedia/.../sajeon.wav"
/>
</WordForm>
```

```
<Multimedia>
  <feat att="type" val="사진" />
  <feat att="label" val="사전" />
  <feat att="url"
val="http://dicmedia.korean.go.kr:8899/multimedia/.../PIC000195981_700X466.jpg" />
</Multimedia>
```



버그도 있습니다

- "큰따옴표" 항목에서 XML invalid
 <feat att="annotation" val="" ""로 쓴다." />
- 없는 단어 레퍼런스
- 정보 없는 관련 용어 - 우리말샘에서 기초 단어를 선정하면서 남은 흔적?
 <SenseRelation>
 <feat att="type" val="유의어" />
 <feat att="lemma" />
 </SenseRelation>



워밍업 : “사전”으로 이용

- XDXF 오프라인 사전 - GoldenDict 등에서 사용 가능
- https://github.com/soshial/xdxf_makedict/blob/master/format_standard/xdxf_description.md
- XSLT 이용해서 한국어기초사전XML을 변환
- <https://github.com/changwoo/krdict-converted>

사전 - GoldenDict

파일(F) 보기(V) 편집(E) 검색 검색기록(I) Favorites 도움말(H)

사전

사전

From 한국어기초사전

사전

1. 단어 명사

【사: 전】 일이 일어나기 전. 또는 일을 시작하기 전.

사전 검토.
사전 예방.
사전 작업.
사전 점검.
사전 준비.
독감이 유행할 때에는 예방 주사를 맞아 사전에 예방하는 것이 좋다.
사전에 아무런 예고도 없이 선생님은 갑자기 쪽지 시험을 보겠다고 하셨다.
여행 준비는 잘돼 가니? 사전에 꼼꼼히 살펴보는 게 좋을 거야.

사전

1. 단어 명사

【사전】 낱말을 모아 일정한 차례에 따라 실고, 그 발음, 뜻, 어원, 용법 등을 설명한 책.

외국어 사전.
두꺼운 사전.
사전 한 권.
사전을 뒤적이다.
사전을 만든다.
사전을 찾다.
사전을 편찬하다.
사전에 실다.
사전에 실리다.
사전에 올리다.
사전에서 찾아보다.
외국인을 위한 한국어 사전을 만들기 위해서 외국인이 알아야 할 단어들을 정리했다.

Now indexing for full-text search: WordNet 3.0 (En-En)

검색된 사전:

한국어기초사전

Favorites:

검색기록: 1/500

사전

맞춤법 검사 프로젝트에 이용하기

- 2008년부터 시작, hunspell용 한국어 “사전” 데이터
- <https://github.com/spellcheck-ko/hunspell-dict-ko>
- 다수의 단어 데이터가 필요
- 오픈소스 배포가 가능한 단어만 포함
 - 대부분 리눅스 배포판에 포함 (hunspell-ko)





필요한 정보

- 표제어 및 그 단어의 각종 형태론(morphology) 정보
 - 품사
 - 활용 정보 (불규칙 활용 용언 여부, 제한적 활용 등)
 - 기타 정보 - 합성 용언
 - (맞춤법 규정) 제47항 보조 용언은 띄어 쓰기를 원칙으로 하되, 경우에 따라 붙여 쓰기도 허용한다.
 - 다만, 앞말에 조사가 붙거나 **앞말이 합성 용언인 경우**, 그리고 중간에 조사가 들어갈 적에는 그 뒤에 오는 보조 용언은 띄어 쓴다.
 - ...
- 대부분은 자동적인 처리로 정보를 알아낼 수 있으나.. 여전히 수동 작업 필요



예) 걷다 (구름이, 빨래를) vs 걷다 (걸음을)

<feat att="definition" val="구름이나 안개 등이
흩어져 없어지다." />

<WordForm>

<feat att="type" val="활용" />

<feat att="writtenForm" val="걷어" />

규칙 활용

<feat att="definition" val="바닥에서 발을 번갈아
떼어 옮기면서 움직여 위치를 옮기다." />

<WordForm>

<feat att="type" val="활용" />

<feat att="writtenForm" val="걸어" />

ㄷ불규칙 활용



예) 걷다 vs 걷다 - 판별

```
nfd = unicodedata.normalize('NFD', word[:-1])
...
elif nfd[-1] == T_TIKEUT:
    if unicodedata.normalize('NFC', nfd[:-1] + T_RIEUL) + '어' in inflections:
        result = 'ㄷ 불규칙'
    elif unicodedata.normalize('NFC', nfd[:-1] + T_RIEUL) + '아' in inflections:
        result = 'ㄷ 불규칙'
    elif word[:-1] + '어' in inflections:
        result = '규칙'
    elif word[:-1] + '아' in inflections:
        result = '규칙'
    else:
        result = 'ㄷ 불규칙 미확정' # 수동 작업
```

어휘 정보를 모으는 “갈퀴”

- GALKWI - Django 기반 단어 정보 입력/편집 사이트
 - 최초에 Google App Engine 사용
 - 요금 체계 변경 후 중지, 독립 사이트로
 - 바닥부터 약 32000 단어까지
- 문제점
 - 유지 관리 - 정책이 바뀌고 필드가 바뀔 때마다 사이트 개발





Galkwi + Wiki = Galkwiki

- 유사점
 - 다수 기여자들에 의한 정보 생산 및 유지관리
 - 편집 히스토리
 - 사용자 관리
- 차이점
 - 문서가 아닌 정형화된 데이터 - Semantic MediaWiki 사용
 - 편집이 텍스트 편집이 아닌 form 입력 - PageForms (예정)
 - 물관리 - 투표/리뷰 기능 - FlaggedRev (예정)
 - 외부 사용을 위해 데이터 추출 및 export 필요



갈퀴 위키

- XML을 시맨틱 미디어위키 속성으로 import
- 하나의 페이지가 하나의 사전 항목
- import된 데이터 속에서 의미있는 정보 찾아내기
- 미디어위키 API를 통해 가능한 자동화



대문
최근 바뀐
임의 문서로
도움말

도구

여기를 가리키는 문서
가리키는 글의 바뀐
특수 문서 목록
인쇄용 판
고유 링크
문서 정보
이 문서 인용하기
속성 찾아보기

ChangwooRyu 토론 환경 설정 주시문서 목록 기여 로그아웃

문서

토론

읽기

편집

역사 보기



더 보기 ▾

갈퀴 검색



견다 (002) (한국어기초사전 29667)

분류: 사전 항목 | 맞춤법 검사 사전 항목

이 판을 다시 검토하기

정확성: **지점 확인함**

길이: **기초적임**

가독성: **괜찮음**

의견:

편집 승인

"견다 (002) (한국어기초사전 29667) ①"에 대한 사실

RDF 피드 ↗

<i>Has subobject</i>	견다 (002) (한국어기초사전 29667)#1 + ①, 견다 (002) (한국어기초사전 29667)#2 + ①, 견다 (002) (한국어기초사전 29667)#3 + ①
맞춤법 검사:불규칙 활용	ㄷ 불규칙 + ㄹ
맞춤법 검사:표제어	견다 + ㄹ
맞춤법 검사:품사	동사 + ㄹ
사전:동일 단어	견다 (갈퀴 Django 1223) + ㄹ
사전:불규칙 활용	ㄷ 불규칙 + ㄹ
사전:원본	한국어기초사전 + ㄹ
사전:원본 라이선스	CC BY-SA 2.0 KR + ㄹ
한국어기초사전:가져온 시각	15:10:55, 12 10월 2017 + ㄹ
한국어기초사전:구분	단어 + ㄹ
한국어기초사전:동형어 번호	2 + ㄹ
한국어기초사전:발음	견 : 따 (http://dicmedia.korean.go.kr:8899/multimedia/naver/2016/30000/22000/29667 geod-tta.wav) + ㄹ
한국어기초사전:어휘 등급	초급 + ㄹ
한국어기초사전:주제 및 상황 범주	인간 > 신체 행위 + ㄹ
한국어기초사전:표제어	견다 + ㄹ



대문
최근 바뀐
임의 문서로
도움말

도구

여기를 가리키는 문서
가리키는 글의 바뀐
특수 문서 목록
문서 정보

문서 토론

원기 편집

역사 보기



더 보기 ▾

갈퀴 검색



걷다 (002) (한국어기초사전 29667) 편집하기

This page supports semantic in-text annotations (e.g. "[[Is specified as::World Heritage Site]]") to build structured and queryable content provided by Semantic MediaWiki. For a comprehensive description on how to use annotations or the #ask parser function, please have a look at the [getting started](#), [in-text annotation](#), or [inline queries](#) help pages.

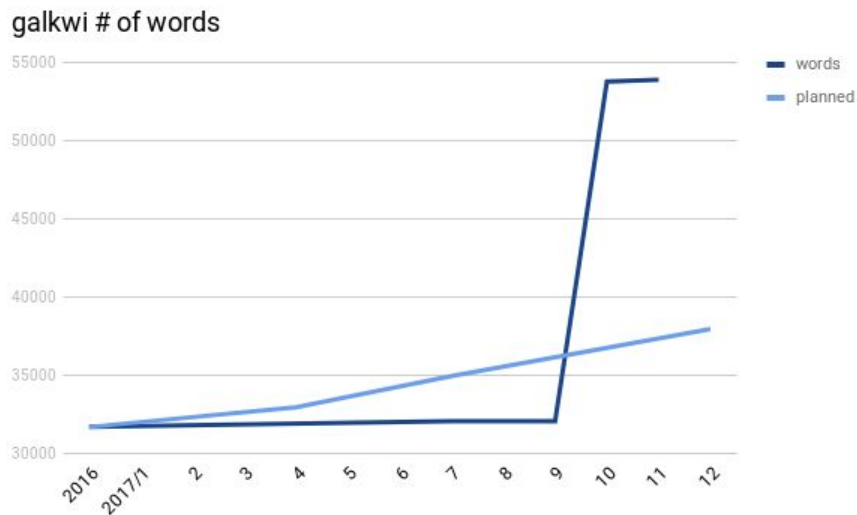


```
<!-- END IMPORT 한국어기초사전 -->
<!-- BEGIN 불규칙활용 자동 검색 -->
{{#set: 사전:불규칙 활용=ㄷ불규칙}}
<!-- END 불규칙활용 자동 검색 -->
<!-- BEGIN 맞춤법 검사 사전 항목 자동 선택 -->
{{맞춤법 검사 사전 항목}}
{{#set:
|맞춤법 검사:표제어=걷다
|맞춤법 검사:품사=동사
}}
<!-- END 맞춤법 검사 사전 항목 자동 선택 -->
<!-- BEGIN 맞춤법 검사 불규칙 활용 자동 검색 -->
{{#set:맞춤법 검사:불규칙 활용=ㄷ불규칙}}
<!-- END 맞춤법 검사 불규칙 활용 자동 검색 -->
```

요약:

☐ 사소한 편집입니다 ☐ 이 문서 주시하기

어휘 대폭 증가: 31979 -> 53940





사전 데이터 버그

- 품사 잘못 - 특히 동사/형용사
 - 기존 데이터와 다를 경우 검사
 - 기존 데이터가 틀렸거나 국립국어원이 틀렸거나 (아직까진 후자가 좀 더 많음 ㅠㅠ)
- 기본적으로 국립국어원 쪽에서 개선해야



회원 정보

내 단어장

회원 탈퇴

사용자 의견

사전 내려받기

쪽지 관리

사용자 의견

종류



전체



사전 내용 오류



장애킨고



서비스 개선 의견



기타 의견

찾을 대상

어휘



Q 찾기

모두 18건의 게시글이 있습니다.

10개씩 보기

구분	어휘	제목	붙임 자료	올린 날짜
사전 내용 오류	횡행하다	형용사가 아니라 동사		2017-11-03 00:30:09
사전 내용 오류	지지부진하다	형용사가 아니라 동사		2017-11-03 00:29:32
사전 내용 오류	되바라지다	형용사가 아니라 동사		2017-11-03 00:28:59
사전 내용 오류	허망하다	동사가 아니라 형용사		2017-11-03 00:22:05
사전 내용 오류	절묘하다	동사가 아니라 형용사		2017-11-03 00:19:44
사전 내용 오류	일정하다	동사가 아니라 형용사		2017-11-03 00:19:07
사전 내용 오류	쓰리다	동사가 아니라 형용사		2017-11-03 00:18:22
사전 내용 오류	성급하다	동사가 아니라 형용사		2017-11-03 00:17:46
사전 내용 오류	비열하다	동사가 아니라 형용사		2017-11-03 00:17:02
사전 내용 오류	부진하다	동사가 아니라 형용사		2017-11-03 00:15:59



우리말샘 추가하면 예상 >10만

- 110만 항목 중...
 - <https://opendict.korean.go.kr/service/dicStat>
 - 속담, 관용구, 전문 용어, 방언, 옛말, 북한어 제외하면 실제 추가할 만한 현대 서울말은 약 20만
- 오히려 너무 많아서 문제? 신어 및 사용자 참여로 추가된 단어들
 - 최대 기여자는 음식 이름 전문
 - 오픈 이후 추가된 각종 유행어들이 구분 없이 들어 있음 (뇌섹남, 거터파크, 성덕, ...)
 - 언어는 변화하는 것이니 사전 등재 자체는 반대하지 않으나, 사용 빈도나 신어 여부 정보가 필요.
 - 박일환 저, 미친 국어사전, 국립국어원의 표준국어대사전 비판 (2015)
- 보류 상태
 - 단어 수로 SMW 처리 부담 우려, 다운로드 파일의 버그, 유행어 등재로 품질 저하 우려



집필 참여하기

사전 통계

어휘 지도

작은 창 사전

들어가기

회원 가입

시험 운영 중

우리말샘

옛 한글

찾기

자세히 찾기

내용 보기

편집 이력

토론하기

☒ 전문가 감수 정보 ?☐ 참여자 제안 정보 ?

단어장에 추가



삼계-탕 (蔘계탕)

편집하기

편집 금지 요청

발음 [삼계탕]

품사 「명사」

「001」 계의 내장을 빼고 인삼, 대추, 찹쌀 따위를 넣어서 끓이는 음식.

▶ 삼계탕은 대게, 찹쌀, 인삼, 대추, 당귀, 닭 가슴살, 밤 등을 넣고 푹 곤다. <일간스포츠 2003년 2월>

▶ 삼계탕, 게살만둣국, 대게매운탕 등 한국인의 입맛에 맞춰 개발한 게요리도 많다. <경향신문 2003년 2월>

다의어(1)

자세히 보기

「001」 계의 내장을 빼고 인삼, 대추, 찹쌀 따위를 넣어서 끓이는 ...

1/1 < >

내가 살펴본 단어

내 단어장

삼계-탕 「001」 x



다른 용도로 이용하는 방법은 없을까요?

- 예문 정보 활용 -> 맞춤법 검사 검증
- 관련어 정보 이용 -> 유의어 사전?
- 외국어학습사전(다운로드 열리면) -> 기계 번역?



정리

- 국가/지자체가 만든 공공저작물을 오픈소스로 활용할 수 있음
- 국립국어원 언어 정보의 대부분은 오픈소스 아님
 - 국가 예산 연구에 대해 라이선스 정책 바뀌어야 함
- 우리말샘/한국어기초사전은 오픈소스 활용 가능
 - 맞춤법 사전에 큰 도움
- 사전 사이트와 데이터 자체의 문제는 아쉬움
 - 개선되었으면



끝 / 질문과 답변

- <https://github.com/spellcheck-ko/hunspell-dict-ko>
- <https://github.com/spellcheck-ko/galkwiki>
- <https://github.com/spellcheck-ko/galkwiki-bots>