
저자 (Authors)	박대민
출처 (Source)	한국언론학회 학술대회 발표논문집 , 2015.05, 138-138(1 pages)
발행처 (Publisher)	한국언론학회 Korean Society For Journalism And Communication Studies
URL	http://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE06294508
APA Style	박대민 (2015). 뉴스 자연어처리. 한국언론학회 학술대회 발표논문집, 138-138
이용정보 (Accessed)	고려대학교 163.152.3.*** 2020/07/29 09:31 (KST)

저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

뉴스 자연어처리: <뉴스소스 베타>를 중심으로¹

박대민

(한국언론진흥재단 선임연구위원)

초록

언론학에서 컴퓨터 이용 내용분석, 컴퓨터 보조 질적 자료 분석 소프트웨어, 의미연결망분석 등의 활용이 늘어나면서 자연어처리를 이용하는 경우도 증가하고 있다. 하지만 일반적으로 자연어처리는 하나의 블랙박스로서 방법론적 절차에 대한 엄밀한 검토가 부족하다. 즉 기존 연구에서 이용한 자연어처리 프로그램이 텍스트의 요소를 얼마나 정확하게 재현하는가 하는 자연어처리 성능 문제, 언어학에서 담론분석이라고 일컫는 언론학의 내용분석과 비판적 담론분석을 수행하기 위해 적절한 분석 단위가 무엇인가에 대한 분석 수준 문제 등의 문제가 제기된다. 이 연구에서는 특히 뉴스에 담겨 있는 높은 수준의 논증을 분석하기 위해서는 단어 중심의 구문분석에 초점을 둔 형식 언어학적 접근이 아니라, 문장이나 기사 수준에서 관계를 부여하고 가중치를 계산하는데 필요한 자연어처리가 필요하다고 주장한다. 이 연구는 뉴스에 대한 담론분석을 위하여 개체명, 특히 정보원과 문장 수준의 분석에 초점을 둔 <뉴스소스 베타>를 기획했다. <뉴스소스 베타> 자연어처리의 특징은 정보원 중심의 개체명 인식과 이에 따른 문장 유형 분류, 저널리즘 관행에 따른 최소한의 부분 구문분석을 바탕으로 하는 의미중의성 해소와 대용어 해소, 클러스터링을 통한 중복 기사와 중복 문장의 제거, 저널리즘 가치에 따라 정의된 뉴스 정보원 연결망분석 알고리즘을 통한 가중치 부여를 특징으로 한다. 여기서는 <뉴스소스 베타>의 자연어처리 알고리즘을 설명하고 이와 같은 방법론을 활용하여 수행된 연구 사례를 통해 컴퓨터 보조 담론분석의 가능성을 살펴본 뒤, 추후 뉴스 기사의 자연어처리를 개선하기 위해 필요한 제언을 제시한다.

핵심어: 뉴스 기사의 자연어처리, 담론분석, 뉴스소스 베타, 부분 구문분석, 의미 중의성 해소, 대용어 해소, 클러스터링, 뉴스 정보원 연결망 분석

¹ 이 연구는 서울대학교 ICT사회정책연구센터(CTRC 사업)와 서울대학교 언론정보연구소의 연구기금 지원을 받았습니다.