

TagBench: 언어 정보 태깅 도구

서형원⁺·김재훈¹·최명길²·남유림³·권홍석⁴

TagBench: a tagging tool for linguistic information

Hyeong-Won Seo⁺, Jae-Hoon Kim¹ · Myung-Gil Kim² · Yoo-Rim Nam³ · Hong-Suk Kwon⁴

초록

일반적으로 한글 문서에 포함된 각종 언어 정보(형태소, 구품음, 기반구의 품사나 의미 정보 등)를 일일이 수정하는 것에는 많은 어려움이 따른다. 또한 지금까지 개발된 언어 정보 수정 도구들은 이런 언어 정보를 쉽게 수정할 수 없다. 따라서 본 논문에서는 한글 문서의 각종 언어 정보를 쉽게 태깅하고 수정할 수 있도록 도와주는 TagBench를 제안한다.

본문

본 논문에서 제안하는 TagBench의 기능은 다음과 같다.

- 1) 통합 - 이전 문장과 통합하고 문장번호는 현재 문장번호로 한다.
- 2) 분리 - 현재 문장을 두 문장으로 분리한다.
- 3) 주석 - 문장에 필요한 정보를 입력한다.
- 4) 분석 - 품사 부착기를 구동하여 새로운 분석 결과를 보여준다.
- 5) 저장 - 수정된 문장을 저장하고 자동으로 품사 태깅을 수행하여 DB정보를 수정한다.
- 6) 임시저장 - 수정한 문장을 임시 저장하여 다음 태깅할 때에 다시 복귀하도록 한다.
- 7) 복원 - 현재 문장에 대한 정보를 이전 문장으로 되돌린다.
- 8) 편집 - Focusing 된 곳에서 편집 가능하면 편집할 수 있도록 한다.
- 9) 출력 - 수정이 완료된 파일에 한해 텍스트 혹은 XML 파일 형식으로 출력한다.

이 외에도 각 품사의 형태소와 구품음 그리고 기반구의 의미 태그를 수정하거나 오류를 수정할 수 있고, 수정된 데이터를 하나의 서버에서 DB로 관리하기 때문에 빠르게 이전 데이터나 다른 사람이 수정한 결과를 가져와서 검토해보는 것이 가능하다. 그리고 관리자는 오류를 수정하는 과정에서 생긴 미해결 문장을 통합하여 해결하고 그것을 각 사용자에게 재편성해줄 수 있다. 그림1은 TagBench의 메인 화면이다.

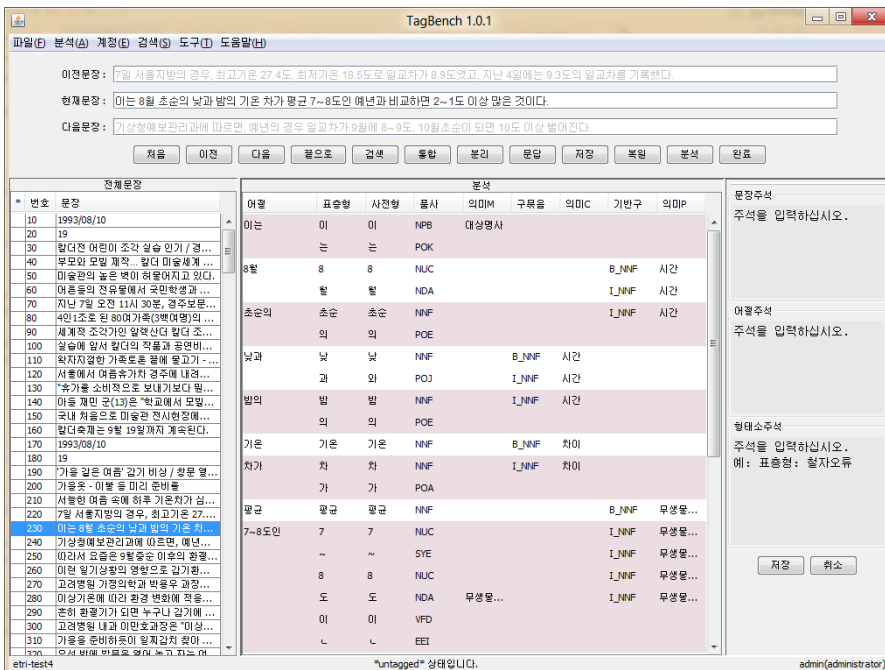


그림 1 TagBench의 메인화면. 화면 하단 왼쪽은 전체문장창, 가운데는 분석창, 오른쪽에는 주석창이 위치하고 있다.

+ 서형원 (한국해양대학교 컴퓨터공학과), E-mail: wonn24@gmail.com, Tel: 051)410-4896
 1 김재훈 (한국해양대학교 컴퓨터공학과), E-mail: jhoon@hhu.ac.kr, Tel: 051)410-4574
 2, 3, 4 최명길, 남유림, 권홍석 (한국해양대학교 컴퓨터공학과), Tel: 051)410-4896