

학습기능을 가진 효율적인 품사 부착 도구 설계 및 구현

안유미^o 오진영 차정원
창원대학교 컴퓨터공학과
sha237@changwon.ac.kr, psyche.ojy@gmail.com, jcha@changwon.ac.kr

Design and implementation of an efficient part-of-speech annotation tool that has the study facility

Yu-Mi Ahn^o Jin-Young Oh Jung-Won Cha
Dept. of Computer Engineering, Changwon National University

요 약

본 논문에서는 자바 기반의 품사부착 코퍼스 작성 도구를 제안 및 구현한다. 본 시스템에서는 각 사용자가 독립적으로 실행하지만 주요 데이터베이스는 서버에서 관리함으로써 지식을 공유할 수 있고, 품사부착 작업에 있어 사전에 만들어진 어절 후보로부터의 선택 및 사용자 입력이 가능하도록 한다. 고빈도 오류 어절의 자동 표시 기능, 용례 검색을 통한 도움말 기능, 코멘트를 기반으로 구성된 집단 지식을 이용한 도움말 확장 기능 및 사전 검색 기능을 구현한다. 또한, 일관성 검사를 통해 품사부착 결과에 대한 신뢰도 증가 및 작업의 편의성을 증대시킬 수 있도록 설계한다.

주제어 품사태깅, 품사부착도구

1. 서론

인터넷 문서가 증가하면서 품사태깅에 대한 요구는 증가되었다. 품사 부착 코퍼스 작성은 이런 품사태깅 시스템의 성능 향상을 위해 반드시 필요한 과정이다. 그러나 코퍼스 작성은 시간과 노력이 많이 들어가는 과정이다. 따라서 우수한 작성 도구의 개발이 필요하다.

현재 대부분의 품사부착은 전용 도구가 아닌 공용 에디터를 사용하여 사용자가 직접 품사를 수정한다. 그러나 현실적으로 모든 단어에 대한 적절한 형태소와 문맥적 의미를 잘 알고 있는 사람이라 할지라도 방대한 문장에 대해 완벽하게 품사부착 작업을 하는 것은 무리가 있다. 특히, 한글의 경우 단어마다의 의미가 매우 다양하여 이러한 작업에 더욱 어려움이 따른다. 일반 에디터를 이용하여 품사부착 코퍼스를 작성할 경우, 지식공유가 어려우며, 일관성 검사와 여러 가지 도움말 등의 어려움이 존재한다.

이처럼 수작업을 통한 코퍼스 작성의 어려움에 따라 최근 몇 년간 미국을 비롯한 국내외 여러 분야에서 관련 분야에 대한 다양한 언어분석을 위한 기본 말뭉치 가공 도구들이 개발 되었다.

미국을 비롯한 외국의 경우 이러한 도구의 개발은 국내에 비해 활성화 되어있다. 현재 개발된 국외 도구에는 영국의 'MATE workbench', 타이완의 'Opinion Annotation Tool(OAT)', 미국의 'Annotation Graph AP', 핀란드의 'DepAnn' 등이 있다.

MATE workbench[1]는 음성 데이터나 문서 데이터를 가공하여 분석결과를 부착하고 표현하며 질의를 던질 수 있는 도구이다. 이 도구는 사람이 쉽게 가공된 문서를 생성할 수 있도록 설계되었으며 다른 사람들과 쉽게 공유하여 작업을 진행할 수 있고 다양한 부착 방법을 사용할 수 있다. 다양한 부착 방법을 사용하여 생성된 말뭉

치도 XML로 변환이 가능하다. 또한 자바로 개발되어 어떤 환경에서도 실행할 수 있다.

Opinion Annotation Tool(OAT)[2]는 의견(opinion)을 부착할 수 있는 도구이다. 이 도구는 GUI형식으로 개발된 도구로서 각 레벨의 의견을 부착할 수 있다. 즉 단어, 부분문장, 전체문장 레벨의 의견을 부착할 수 있다. 또한 문서의 주제에 따라서 적절한 의견을 부착할 수 있는 기능도 제공한다. 이 도구는 중국어, 영어, 일본어에 대해서 문서를 작성할 수 있다.

Annotation Graph AP[3]는 Time-Series Data에 대해서 언어정보를 부착할 수 있는 도구이다. 이 도구는 사용자가 특별한 분야의 가공도구를 신속히 개발할 수 있도록 그래프를 이용하는 기반구조를 제공한다. 따라서 응용프로그램 API, 입출력 라이브러리, 그래픽 인터페이스 등을 제공한다.

DepAnn[4]는 Dependency treebanks를 위한 언어정보 부착도구로, 그래픽과 텍스트 기반 인터페이스 모두를 제공한다. 도구의 입력 형식은 TIGER-XML을 사용하고 있다. 이 도구는 사용자가 문장 구조에 문제가 있는 부분에 대해 사용자 의견을 추가할 수 있도록 하며, 트리 구조를 손쉽게 이동하거나 수정할 수 있다. 또한, 자동적인 일관성 검사를 통해 그 문장 구조와 주석과 encoding 검증을 수행한다.

국내의 경우 포항공대의 'POSTAG Workbench(PWB)', 솔트룩스의 '의미 태깅 시스템(Semano)' 등이 있다.

포항공대 isoft 연구실에서 개발된 품사태거인 POSTAG는 자동으로 품사를 할당하는 프로그램이다. 이 프로그램의 학습문서를 제작하는 도구인 PWB[5]는 인터넷 브라우저에서 실행되는 자바 애플릿으로 구현되어 있다. 이 프로그램은 형태소 분석된 결과를 입력으로 받아서 그 결과를 그래프 형태의 그림으로 표현한다. 사용자는 이 그래프에서 원하는 경로를 결정하면 프로그램에서

자동으로 저장하도록 되어 있다. 이 프로그램은 포항공대에서 개발된 품사체계에 맞도록 설계되어 있고 애플릿으로 작성되어 있어 느리고 다중사용자 실행이 되지 않는 단점이 있다.

솔트룩스의 '의미 태깅 시스템(Semano)[6]은 자연어 처리와 텍스트 마이닝 기술에 기반 하여, 사람이름, 기업명, 주소, 기술명, 제품명, 프로젝트 등의 개체명과 사건, 원인, 결과, 상황(상태) 등의 단위지식 정보를 추출하고, 온톨로지 인스턴스 생성 및 의미 테마데이터 생성 기능을 제공하는 시스템이다. 이 Semano에서는 각 도메인에 맞는 다양한 개체명 사전을 로딩하고 자체개발한 XRE 문법과 규칙에 기반 하여 의미 정보를 추출할 수 있다.

그러나 이러한 도구들은 표준화된 언어정보를 부착하는 것이 아닌 각자 기준에 근거하여 작성되고 있어 자원을 공유하는데 한계를 가지고 있다. 본 논문에서는 이러한 도구의 필요에 따라, 세종21 프로젝트 품사태깅 기준에 따라 코퍼스를 작성하고, 자동 학습 기능을 가진 품사부착 도구를 제안한다. 여기서 제안한 도구에서는 사용자가 각각의 어절에 대한 후보를 입력하지 않고, 사전에 생성된 후보를 단순 선택하는 방식을 택한다.

본 논문의 구성은 다음과 같다. 2장에서는 구현된 도구에 대한 구조 및 주요 기능을 설명한다. 마지막으로 3장에서 본 개발 도구에 대한 결론을 내린다.

2. 품사 부착 도구

지금까지 효율적인 언어분석을 위한 기본 말뭉치 가공 도구들이 개발되었다. 도구들이 편리한 인터페이스를 제공한다면 약간의 훈련으로도 충분히 도구를 다룰 수 있어 코퍼스 작성에 효율적일 것이다. 그러나 도구 사용에 능숙한 사용자라도 관련 분야의 지식이 부족하다면 작업에 많은 제약이 따르게 된다[7].

특히, 품사 부착의 경우 각각의 품사에 대한 표기를 알아야 정확도가 높은 품사 부착 작업이 가능해진다. 정확도가 높은 품사 부착 작업이 이루어지기 위해서는 품사부착과 동시에 사전 검색을 통한 품사 확인, 도움말 기능을 통한 사전 작업 확인, 이전 완료 문장에 대한 확인 및 수정 등 많은 기능을 동시에 수행해야 한다. 많은 작업을 동시에 수행한다면 혼란이 발생할 수 있고, 품사 부착 과정에서 일관된 작업과정이 없으면 작업에 있어 추가 시간 및 혼란을 발생시킨다.

이러한 혼란과 번거로움을 최소화하기 위한 품사 부착 도구를 개발하였다. 제안된 도구는 품사 부착 시 필요한 다양한 기능을 한 작업 공간 내에서 최대한 이루어질 수 있도록 구성하고, 완료된 문서에 대해 빠른 확인 및 수정이 가능하다. 또한, 다른 사람의 작업 결과를 확인하고 의견을 달거나 다른 사람의 의견을 확인할 수 있다.

그림 1은 사용자의 작업 구조를 보여주고, 그림 2는 품사 부착 작업의 전체적인 모습이다.

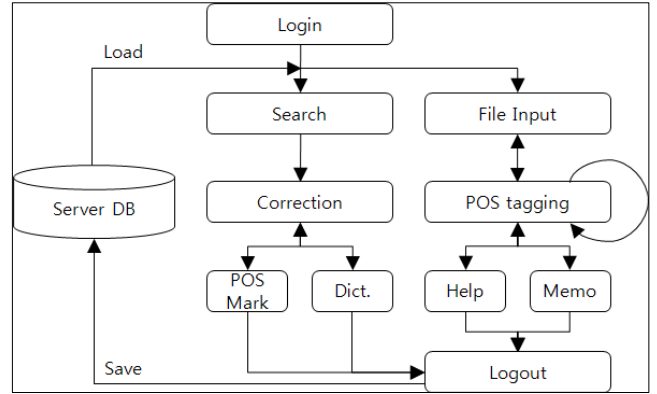


그림 1. 사용자의 작업 과정

그림 1에서 사용자가 로그인을 하여 작업하고자 하는 파일을 선택하면 그림 2와 같이 각 문장에 대한 후보가 나타난다. 여기서 용례검색, 사전검색, 사용자 의견 등을 참조하여 품사부착을 한다. 그림 2에서는 본 논문에서 제안한 도구의 작업공간을 보인다. 그림의 왼쪽은 사용자가 후보 중에서 정답을 선택하는 공간이고, 왼쪽 아래에는 현재까지 작업한 문장의 리스트를 볼 수 있다. 그리고 오른쪽은 용례를 검색하고 사용자 의견을 작성하고, 사전을 검색하는 공간이다.

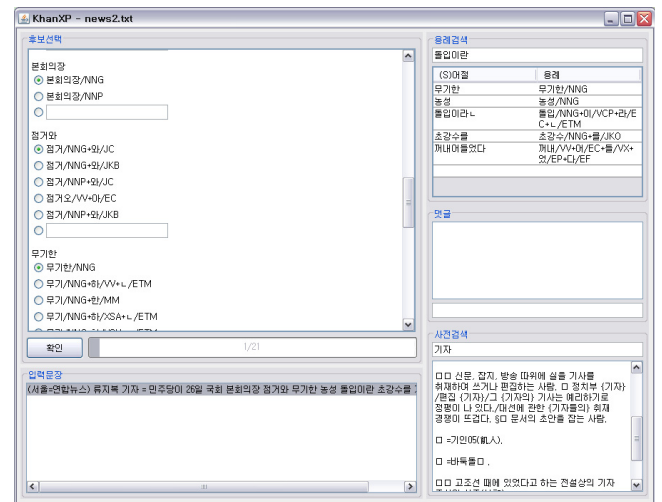


그림 2. 품사 부착 작업 공간

본 논문에서 제안한 도구는 다음과 같은 큰 특징을 가지고 있다.

2.1 데이터베이스 공유

대량의 문서에 대하여 품사 부착 작업을 할 경우, 많은 사람이 함께 작업하는 것이 대부분이다. 많은 사람이 저마다 어절에 대한 품사 선택에 있어 자신만의 생각에 맞추어 작업을 하게 될 경우, 만약 사용자가 품사에 대해 잘못 알고 있는 상태에서 작업을 한다면 잘못된 품사 부착에 대한 오류의 인식 및 확인이 힘들어진다. 많은 사용자가 비슷한 어절에 대해 작업을 한다면 다양한 결과 및 의견 확인을 통하여 더 정확도 높은 작업을 할 수

있을 것이다.

폼사 부착 작업에서 발생할 수 있는 오류 문제나 능력 저하의 문제 등을 해결하기 위하여 본 도구는 프로그램을 실행할 때 중앙서버에서 관리하는 데이터베이스를 다운로드 받아 사용자 작업공간에 저장하게 되고 이 정보를 바탕으로 지식을 공유한다. 또한 프로그램을 종료할 때 작업한 내용을 다시 중앙서버에 전달함으로써 폼사 부착 작업 시 서로의 의견과 결과를 확인 할 수 있으며, 사용자들이 함께 작업하는 것과 같은 효과를 준다.

2.2 출력 된 후보로부터의 선택

폼사 부착 작업에서 발생하는 번거로움 중 하나는 각 어절에 대한 폼사 정보를 사용자가 직접 작성해야 한다는 것이다. 아무런 정보가 없는 상황에서의 폼사를 일일이 작성해야 할 경우와 어절에 대한 폼사 후보가 제공되는 상황에서 알맞은 후보를 선택하여 작성하는 경우, 후보가 제공되는 상황에서 알맞은 후보를 단순히 선택하기만 하면 되는 경우의 작업 속도는 많은 차이가 있을 것이다.

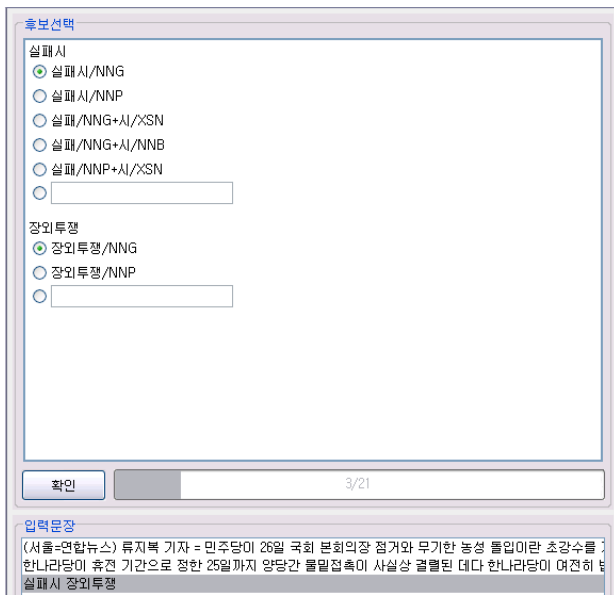


그림 3. 각 어절에서 나타나는 폼사 후보들 중 문맥에서 선택 될 가능성이 가장 높은 답이 자동으로 선택되어 출력 된 모습

이러한 번거로움을 해결하고 작업 속도를 향상시키기 위해 어절에 대해 생성 된 후보를 보여주고 선택만 하면 작업이 완료 된다. 만약 알맞은 후보가 존재하지 않을 경우 새로운 태그를 직접 입력할 수 있도록 하였다. 각 어절에 대해 생성된 후보는 문맥과 상관없이 그 어절에서 우선순위가 가장 높은 폼사부터 낮은 폼사 순으로 출력 된다. 또, 어절 마다 문맥상 사용될 가능성이 높은 후보가 미리 선택되어 출력되도록 하였다. 이러한 기능들을 통해 간단한 선택만으로 폼사 부착이 가능하며, 오타에 의한 폼사 오류 확률을 낮추고, 모든 어절을 수정하지 않아도 되기 때문에 빠른 폼사 부착 작업이 가능하다.

그림 3은 어절에 대한 각각의 후보들과 문맥상에서 사용 될 가능성이 가장 높은 폼사 후보가 선택되어 작업 화면에 출력 되는 모습을 보여준다. 아래의 '입력문장'은 현재까지 작업한 문장의 리스트를 보여준다. 이 리스트에서 문장을 선택하면 문장의 입력 오류로 어절분리가 잘못된 경우 또는 오타가 있는 경우에 대해서 사용자가 문장을 수정할 수 있도록 하였으며, 이는 바로 작업장에 적용된다. 뿐만 아니라 문장을 선택하면 문서 내에서 작업 완료된 문장들에 대한 선택 결과를 확인 가능하다.

2.3 학습을 통한 고빈도 오류 어절 자동 표시

본 도구는 정답 가능성이 높은 후보를 미리 선택하여 출력해 준다. 만약 이 후보가 틀렸을 경우 사용자는 틀린 어절 폼사를 수정하게 된다. 많은 사용자가 작업을 할 경우, 유사 문장의 어절에 대해 다른 결과를 선택할 경우가 생긴다. 각 어절에 대해 어절 폼사를 시스템이 추천한 후보가 아닌 다른 폼사로 한번 이상 수정 한 경우, 특정 표시로 사용자에게 알려 준다면 폼사 부착 작업의 효과를 증가시킬 수 있을 것이다. 특정 표시가 있는 부분에 대해서는 사용자가 더 관심을 가지고 작업을 할 수 있을 뿐 아니라, 추천 후보에 대한 반복적 오류 수정을 통해 작업의 질을 향상 시킬 수 있기 때문이다.

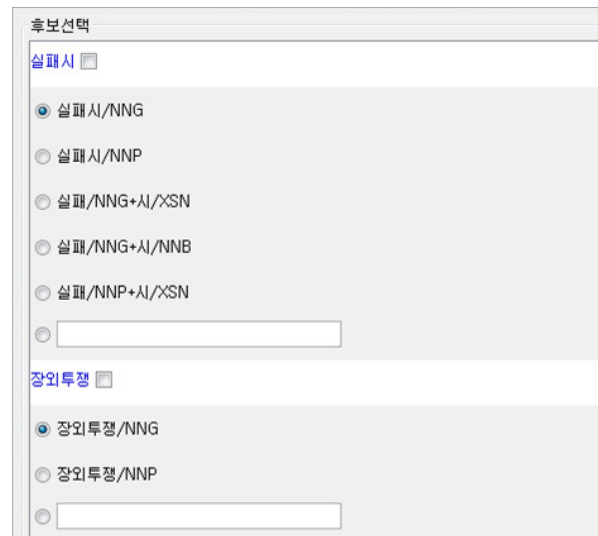


그림 4. 같은 문맥에서 다른 어절을 선택 한 경우, 오류 어절이 자동으로 표시 된 모습. 오류어절은 파란색으로 표시, 고빈도 오류어절은 다시 빨간색으로 변경

어절에 대해 신규 작업의 경우 시스템 추천 후보와 현재 사용자 수정 결과를 비교하여 다를 경우 횡수를 증가시킨다. 수정이나 책갈피 기능을 사용하는 작업에 대해서는 시스템 추천 후보, 이전 수정작업 완료된 결과, 현재 작업한 결과 비교를 통해 자동적이고 능동적으로 특정 표시에 대한 변화가 가능하도록 하였다. 폼사 부착이 완료 된 특정 어절수와 특정 어절 중 수정 된 수를 이용하여 식(1)과 수식을 이용하여 오류발생 정도를 표시하였다.

$$socre = p_i - 1.65 \sqrt{\frac{p_i(1-p_i)}{n_i}} \quad (1)$$

여기서 n_i 는 특정 어절 i 가 출현한 수이고 p_i 는 수정 횟수를 n_i 로 나눈 값이다. 만약 한 번도 다르지 않다면 '검정색, 한 번 이상 수정하면 '파란색', 식(1)의 결과가 0.6 이상이면 '빨간색'으로 색깔로 효과를 달리 주었다. 그림 4를 통해 이와 같은 결과를 확인할 수 있다.

2.4 용례 검색을 통한 도움말 기능

많은 어절에 대해 품사 부착 작업을 하다보면 문장에서 어떤 품사로 사용 되었는지 모르는 어절이 생길 수 있다. 특히 한글의 경우 단어마다의 의미가 매우 다양하여 문장에서 쓰이는 품사에 혼돈이 있을 수 있다. 또한 여러 문장을 시간 간격을 두고 작업 할 경우 이전 작업에 대한 결과를 기억하기 어려울 경우가 있다. 이전 작업에서 현재 작업과 유사 문맥에서 나타난 어절의 품사와 다른 품사가 선택되었다면 그 문서의 질과 신뢰도는 낮아지게 된다. 유사 문맥에 대해 동일한 결과의 품사가 부착되어야 하는 품사 부착 작업에서는 큰 문제가 된다. 만약 이전에 작업한 어절에 대한 정보를 검색하거나, 다른 사용자가 작업한 결과를 손쉽게 확인할 수 있다면 작업의 효율성은 높아질 수 있을 것이다.

이러한 효과에 착안하여 용례검색을 통한 도움말 기능을 구현하였다. 이 기능을 사용하면 확실하지 않은 어절에 대한 검색을 통하여 작업할 수 있어 전문 지식이 부족한 사용자라 할지라도 조금 더 정확성 높은 품사 부착 작업을 할 수 있다.

품사 부착 작업을 하고 있을 경우, 특정 어절에 대한 검색은 직접 어절을 입력하면 전후 두 어절이 함께 출력되므로 문맥상에서의 어절에 대한 품사를 확인할 수 있다. 또, 작업 중인 어절에 대해서는 직접 입력 또는 작업 중인 어절을 클릭하면 자동으로 검색하여 결과를 출력해 준다.

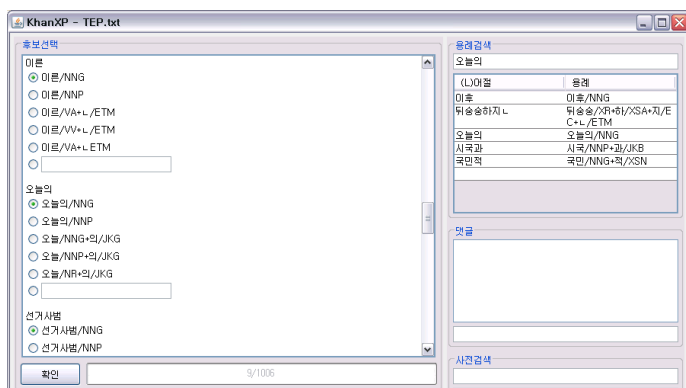


그림 5.작업 중인 어절을 클릭하여 용례 확인

그림 5에서 보는 것과 같이 작업을 진행 중일 때에도 어절에 대해 실시간으로 용례를 확인할 수 있다.

용례 검색에 대한 다른 기능은 그림 6에서 보는 것과 같이 문서 수정을 위한 검색이 가능하다.

본 도구는 중앙서버에서 다운로드한 데이터베이스(중앙DB)와 현재 사용자의 컴퓨터에서 작업한 문서들의 결과 및 정보가 저장되는 데이터베이스(사용자DB) 두 개로 나누어진다. 용례를 검색하면 (L)로 표시된 사용자DB와 (S)로 표시된 중앙DB의 결과를 출력한다. 수정 가능한 것은 (L)로 표시된 용례결과로 용례를 선택하면 해당 결과들을 왼쪽 작업창에서 확인 할 수 있다. 어절에 대한 사용자의 이전 작업 결과 확인 및 잘못된 부분에 대한 수정을 위해서는 문서 수정 작업창을 통해 쉽게 이루어질 수 있다. 체크박스를 이용하여 한 번에 여러 개의 용례를 수정할 수 있다.

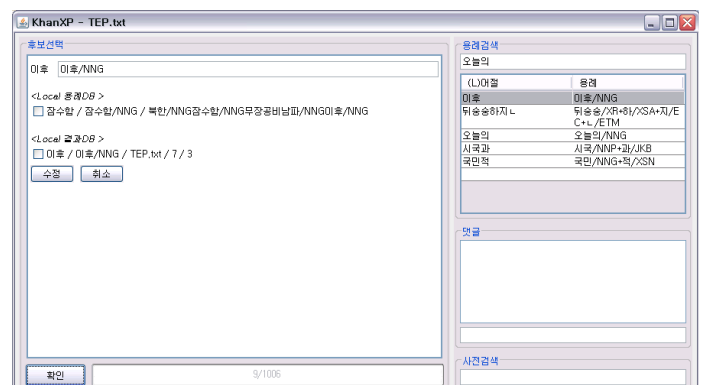


그림 6. 문서 수정을 위한 검색에서의 용례 확인

2.5 집단 지식을 이용한 도움말 확장

집단 지식을 이용한 도움말의 확장은 도움말(용례검색)의 확장이라고 말할 수 있다. 용례검색 시 결과와 함께 출력되는 코멘트 형식의 집단 지식은 품사 부착 작업의 효율을 높일 수 있다. 작업자는 특정 어절에 대한 고려사항을 기술하여 공동작업자가 참조하게 할 수 있다.

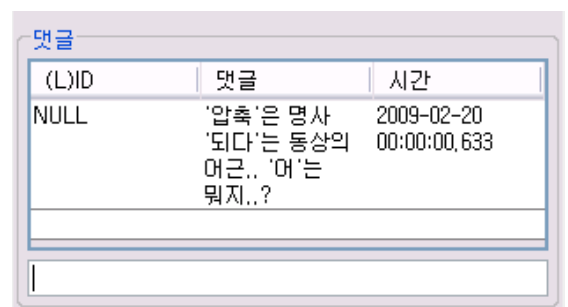


그림 7. '압축되어' 어절에 대한 코멘트 정보

하지만 이 코멘트 정보는 자신만 확인 할 수 있는 정보로도 활용할 수 있으며, 많은 공동작업자가 참조할 수 있게 중앙데이터베이스에 전달하는 방법으로 나누어진다. 중앙데이터베이스에 전달하면 특정 어절에 대해 자신이 선택한 후보에 대한 코멘트를 남겨 참고하는 사람들에게 어절에 대한 분석을 전해 줄 수도 있다. 등록한 사용자는 코멘트를 삭제 할 수 있으며, 데이터베이스

스에는 같은 어절에 대하여 등록 된 모든 코멘트와 날짜 정보의 조합을 함께 가지고 있으므로 같은 내용을 등록 하여도 모두 등록이 가능하다.

그림 7은 사용자가 용례검색을 한 뒤에 특정 어절에 대한 자신의 의견을 저장한 것이다. 그림 7의 코멘트는 모든 사용자가 확인 할 수 없는 상태이지만, 로그인을 통해 중앙데이터베이스에 전달하면, 모든 사용자는 ‘압축 되어’ 라는 어절 용례검색으로, 용례 결과와 이미 저장된 코멘트 내용을 확인 할 수 있는 것이다.

2.6 사전 및 품사 표기법 검색

단어마다 다양한 의미가 존재 할 수 있다. 품사 부착 작업을 하면서 용례검색을 통한 도움말을 이용할 수 있지만, 사전적인 의미가 필요한 경우가 생긴다. 이를 위해서 국어사전을 검색하여 그 의미와 품사를 참조할 수 있는 기능을 추가하였다. 사전적 의미를 통해서 문맥상의 어절 의미를 파악하여 작업의 편리함을 도와준다. 작업을 하며 독립적으로 바로 검색이 가능하다.

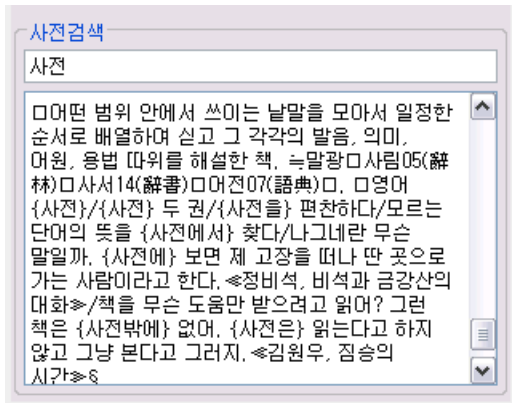


그림 8. 문장 내의 어절 의미 파악에 도움을 주기 위한 사전 검색 기능

2.7 자동 일관성 검사

일관성 검사란 특정 어절에 대한 입력 자료가 미리 정해진 기준과 일치 하는지를 확인하는 것이다. 일관성 검사는 같은 문맥에 대한 동일한 어절의 결과는 같아야 한다는 가정으로 설계된 기능이다.

앞에서 기술된 용례 검색 시에 이미 저장된 동일 어절에 대한 결과 문맥을 파악하여 현재 저장하려는 문맥과의 유사도를 계산한다. 그리고 같은 문맥임에도 불구하고 저장 된 작업과 다른 품사가 선택 된 경우에는 사용자에게 어절에 대한 선택 결과가 사전에 완료된 작업과 다르다는 것을 알려주는 형식으로 구성한다. 일관성이 틀릴 경우 나타나는 경고를 확인하여 사용자는 현재 작업 중인 결과나 이전 결과를 모두 수정할 수 있다.

저장 된 어절의 결과의 전후 어절을 포함하여 문맥의 품사를 확인하고 현재 선택한 어절에 대해 문맥의 품사를 확인하여 비교 후 어느 정도의 유사성을 가지는가에 따라 일관성 유지 여부를 알려준다.

일관성 검사의 사용은 사용자 입장에서는 품사 부착

작업의 일관성을 유지하고, 작업의 저장과 동시에 알려주므로 빠른 수정이 가능하다. 관리자 입장에서는 완료 문서 및 사용자에게 대한 신뢰도를 증가시킬 수 있고, 사후 결과 문서의 사용에 있어 효율이 극대화 될 것이다.

3. 결론

국내의 경우 언어처리를 위한 말뭉치 가공 도구가 일반적이지 않다. 하지만 증가하는 많은 양의 데이터를 활용하고자하는 요구는 증가하며, 품사 부착 작업의 처리 효율성을 높이기 위해 편리한 도구의 필요성 또한 높아지고 있다. 하지만 개발된 도구들은 동시에 많은 사용자들을 관리하기 어려우며, 느리고, 사용자들 간의 정보 공유가 어렵다는 것 등의 단점을 가지고 있다.

본 도구는 설계 시 이러한 단점들에 대한 기능들에 중점을 두었다. 앞에서 기술된 것처럼 작업 시에 사전적 의미의 검색을 위한 사전을 부착하였고, 용례검색으로 어절의 문맥을 확인할 수 있으며, 집단 지식을 이용한 도움말 확장으로 서로의 의견을 공유할 수 있다. 다시 말해 위의 기능은, 기능 구성에 대한 효율을 강조했다. 품사 부착 작업과 동시에 사전을 검색하며, 용례검색을 하고 부착적으로 집단 지식을 보여주며 최대한의 페이지 내에서의 동선을 줄이도록 하였다. 또한 다음 기능들을 통해서 핵심적으로 본 논문에서 강조하고자 하는 것은 학습기능을 가진 것이라 할 수 있겠다. 일관성 검사를 통한 문맥상의 어절 품사 부착 결과의 오류를 찾는다. 즉, 이전에 저장된 문맥 흐름과 어절결과의 수집 및 학습으로 문맥의 오류를 찾고 사용자에게 알려준다. 또, 고빈도 오류 어절 자동 표시를 통하여 품사 부착 작업의 각 어절에 대한 위험도를 색으로 표시하였다. 이러한 기능들로 본 도구는 품사 부착 작업의 효율을 높였으며, 도구 필요성을 증명했다고 할 수 있겠다.

마지막으로 일관성의 기능의 방안으로 기존의 동일 파일들에 대하여 다른 사용자에게 할당 한 후 품사 부착 작업에 대한 결과를 서로 비교함으로써, 문맥에 대한 흐름 측정 설계로 더욱 성능을 높일 수 있을 것이다. 또한, 평가 방법으로 관리자가 일부 문장에 대한 품사 부착 작업을 수행 한 후, 그 문서의 결과를 바탕으로 기존 각 사용자가 수행한 품사 부착의 결과와 비교하여 추가 평가를 한다면 사용자 평가와 동시에 각 문서의 정답 문장 (또는 어절) 확률로서의 문서 신뢰도를 평가할 수 있을 것이다. 이와 같은 방법들을 이용하여 평가방법을 보완한다면 품사 부착 작업에의 효율은 더욱 증가할 것이다.

참고 문헌

- [1] <http://mate.nis.sdu.dk/about/>
- [2] Lun-Wei Ku, Yu-Ting Liang and Hsin-Hsi Chen, "Opinion Extraction, Summarization and Tracking in News and Blog Corpora", 2006, AAI
- [3] <http://agtk.sourceforge.net/>
- [4] Tuomo Kakkonen, "DepAnn - An Annotation Tool for Dependency Treebanks",
- [5] 차정원, 일반화된 한국어 미등록어 추정, 석사학위 논문, 1999

[6] Saltlux Site, <http://in2.saltlux.com/>

[7]이인근, 황도삼, 권순학, “온톨로지 구축 시스템(An Ontology Construction System)”, 2006, 한글 및 한국어 정보처리 학술대회