한국언어정보학회
Korean Society for Language and Information

소식지 제 102 호 (2011년 5월)

발행인: 남승호
편집인: 이해윤
발행처: 한국언어정보학회
발행일: 2011년 5월 16일
URL: http://society.kisti.re.kr/~ksli/

449-791
경기도 용인시 처인구 모현면
한국외국어대학교 인문대학 언어인지과학과 이해윤 교수 연구실내
전화: (031) 330-4284
e-mail: haeyun@hufs.ac.kr

1. 월례 논문발표회

이번 5월 모임에는 이기용 (고려대) 교수님의 특강과, 전홍우 (KISTI) 선생님의 기회논문 그리고 권민재 (한국외대), 김동성 (고려대) 선생님의 일반논문 발표가 있을 예정입니다. 월례회에 대한 문의사항이 있으시면 송민영 연구이사(songmy@dongguk.ac.kr)나 이해윤 총무이사(haeyun@snu.ac.kr)에게 연락 주시기 바랍니다.

2. [언어와 정보] 논문 모집

► [언어와 정보]에 실을 원고를 수시로 모집하고 있습니다. 심사용 논문은 자유로운 스타일로 아래아 한글이나 MS Word로 작성하셔도 됩니다. 어느 경우이든 꼭 논문 요약을 포함시켜 주시기 바라고, 그 외 논문의 스타일은 학회 홈페이지(http://society.kisti.re.kr/~ksli/)에 자세히 소개되어 있으니 참조하시기 바랍니다.

► [언어와 정보]는 일반 논문(general paper) 외에도 소고(squib), 서평 또는 논평(review), 또 이에 대한 답론(reply) 등 여러 형태의 투고를 받습니다. 논문 투고 시 투고 형태를 알려 주시기 바랍니다.

► 논문 심사료와 게재료는 다음과 같습니다. 논문 출판 시 라텍 편집비는 따로 받지 않습니다.

심사료: 30,000원

논문 게재료:

1) 연구비 수혜 논문: 270,000원 (심사료 포함 총 300,000원)
2) 연구비 수혜를 받지 않은 일반논문:
   (i) 저자 중 한 사람 이상이 전임인 경우 170,000원 (심사료 포함 총 200,000원)
   (ii) 모든 저자가 전임이 아닌 경우 20,000원 (심사료 포함 총 50,000원)

► 논문 게재에 관하여 기타 문의사항이 있으시거나, 원고 제출에 관해 의논하시려면 편집위원장인 위혜경 교수님(hkwee@dankook.ac.kr)께 연락하시면 됩니다.

3. 회원 가입 및 회비 납부 안내

► 우리 학회는 회원 여러분들께서 납부하신 회비로 운영됩니다. 원활한 학회 운영을 위해 가급적 빠른 시일 내에 2011년도 연회비를 납부해 주시기 바랍니다. (계좌번호: 우리은행 1002-839-410880 예금주: 이영주)

연회비: 정회원 40,000원, 준회원 20,000원, 기관회원 100,000원
평생회비: 400,000원

► 회원주소록은 연중 관리하고 있습니다. 주소 및 이메일, 전화번호가 바뀌신 분은 변경된 내용을 총무이사 (haeyun@hufs.ac.kr)에게 알려주시면 많은 도움이 되겠습니다. 아울러 주변의 교수님들과 대학원생들이 신입회원으로 많이 가입할 수 있도록 적극적으로 홍보해 주시기를 부탁 드립니다.

4. 학술대회 안내

► Workshop: Semantic Annotation for Computational Linguistic Resources
in conjunction with The 5th IEEE International Conference on Semantic Computing
- 초록 마감: 2011-06-03
- 대회날짜: 2011-09-18 ~ 21
- 대회장소: Stanford University
- 사이트: http://www.sfu.ca/linguistics/wecol2011/

► The 2011 Western Conference on Linguistics (WECOL 2011)
- 초록 마감: 2011-07-01
- 대회날짜: 2011-11-18 ~ 20
- 대회장소: Simon Fraser University (Canada)
- 사이트: http://www.sfu.ca/linguistics/wecol2011/

► The 25th Pacific Asia Conference on Language, Information and Computation (PACLIC 25)
- 논문 마감: 2011-08-08
- 대회날짜: 2011-12-16 ~ 18
- 대회장소: NTU & COLIPS (Singapore)
- 사이트: http://portal.cohass.ntu.edu.sg/PACLIC25/

한국언어정보학회 2011학년도 1학기 월례회 일정표

장소: 대우재단 7층 1실　시간: 9:30-13:00

날짜	시간	구분	발표자	논문제목	사회
3/19 (토)	09:30 10:30	기획1	유현조 (서울대)	한국어 TimeML과 TimeBank	최인철 (경북대)
	Coffee Break (15분)				
	10:45 11:45	특강	전영철 (서울대)	Contrastive expressions in Korean and Hungarian	
	Coffee Break (15분)				
	12:00 13:00	기획2	이은주, 류병래 (충남대)	Broca 실어증 환자의 수동문 이해 양상과 언어학적 쟁점	
4/16 (토)	09:30 10:30	기획1	김지은 (울산대)	The generation of implicit propositions in "alleged" Korean topics	송민영 (동국대)
	Coffee Break (15분)				
	10:45 11:45	특강	송명현, 이정민 (서울대)	Dynamic Prototipicality in English CF-Reduplication	
	Coffee Break (15분)				
	12:00 13:00	기획2	김미숙 (상지대)	Priming Effects in Children's Production of Passive Sentences	
5/21 (토)	09:30 10:30	기획1	전홍우 (KISTI)	**대규모 디지털 학술자원 기반 Virtual Science Brain 활용을 위한 검색 시스템 연구**	김경애 (한경대)
	Coffee Break (15분)				
	10:45 11:45	특강	이기용 (고려대)	**The Semantic Annotation of Motion, Paths and Time in Language**	
	Coffee Break (15분)				
	12:00 12:30	일반	권민재 (한국외대)	**온톨로지 컨셉과 한국어 센스의 매핑 알고리즘의 개발 및 평가**	
	12:30 13:00	일반	김동성 (고려대)	**연관 규칙 마이닝을 활용한 영작문 형태·통사 오류 자동 탐지 및 CFG 오류 규칙 생성**	
6/11 (토)	09:30 10:30	기획1	이우승 (고려대)	TBA	조세연 (강원대)
	Coffee Break (15분)				
	10:45 11:45	특강	최승권 (ETRI)	영어 회화 교육용 대화 시스템에 관한 연구 패턴 기반 자동번역	
	Coffee Break (15분)				
	12:00 12:30	일반	양정석 (연세대)	TBA	
	12:30 13:00	일반	임성출 (경북대)	TBA	

한국언어정보학회 2011 년 5 월 월례회 발표논문 초록

<기획 논문>

대규모 디지털 학술자원 기반 Virtual Science Brain 활용을 위한 검색 시스템 연구

전 홍 우 (KISTI)

연구자들의 연구 주제 선정 및 연구동향 분석하기 위해서 이미 발간된 수많은 논문, 특허, 보고서에서 정보를 수집 및 분석해야 하는 어려움이 있다. 이를 극복하기 위해서 정보추출 연구와 더불어 추출된 정보의 다양한 검색 방법 연구가 필요하다. 그러나 정보추출연구는 활발히 이루어지는 반면 정보의 검색 방법 연구는 상대적으로 주목 받지 못하고 있다. 본 연구에서는 고급 정보들을 다양한 방법을 통해 분석된 결과를 제시하여 사용자들이 새로운 연구 성과를 도출할 수 있는 획기적인 서비스를 소개한다.

<특강>

The Semantic Annotation of Motion, Paths and Time in Language

이 기 용 (고려대)

Travelogues or other types of texts on traveling contain information on events and paths that relate to spatiotemporal locations and durations. Such information, however, requires contextual anchoring and orientations for the correct interpretation of its content that may induce a sequence of appropriate actions. This paper aims at showing how some of the ISO standards on language resource management, particularly ISO-TimeML and ISO-Space, apply to the annotation of semantic content in such texts. The paper proposes some notational modifications to make various IOS standards interoperable for the sustainable management of language resources, while satisfying some requirements on the annotation and representation schemes such as transparency in producing semantic representations with an incremental approach.

<일반논문 1>

온톨로지 컨셉과 한국어 센스의 매핑 알고리즘의 개발 및 평가

권 민 재 (한국외대)

온톨로지는 “an explicit specification of a conceptualization”(Gruber 2003)으로 정의된다. 즉, 이 세상에 존재하는 컨셉들과 그들의 속성 및 관계에 대하여 명세화해 놓은 것을 말한다. 온톨로지는 지식기반 기계번역 시스템 및 검색엔진 등의 분야에서 의미를 처리하는데 있어서 유용하게 사용될 수 있다. 그러나 온톨로지상의 컨셉들은 언어중립적인 매우 추상적인 것이므로 언어정보를 처리하는데 직접적으로 이용할 수 없다. 온톨로지를 언어정보처리에 활용하기 위해서는 온톨로지의 개별 컨셉들에 상응하는 어휘항목들이 매핑되어 있어야 한다. 그러나 온톨로지와 어휘부 사이의 매핑작업은 작업자의 언어적 직관과 주관적 판단에 의존해야 하기 때문에 현실적으로 정확성과 일관성을 유지하기가 어려

울 뿐만 아니라 많은 시간과 비용을 필요로 하는 어려운 과제이다.

본 연구에서는 이러한 문제점을 극복하기 위하여 온톨로지와 어휘부를 반자동적으로 매핑할 수 있는 알고리즘을 설계하였다. 구체적으로, 상위레벨 온톨로지 SUMO의 개별 컨셉에 한국어 명사부의 센스들을 반자동으로 매핑하기 위하여 온라인 한영사전(Google, Yahoo, Naver), Princeton WordNet, SUMO-WordNet-Mappings를 비롯한 다양한 언어자원을 활용하였다. 매핑은 개별 한국어 센스에 상응하는 영어 대응어목록을 한영사전으로부터 추출하는 작업에서부터 시작된다. 매핑알고리즘의 기본 아이디어는 개별 한국어 센스와 대응되는 영어 어휘들을 대상으로 이들 어휘들 간에 공통되는 신셋을 조사하여 그와 매핑되는 SUMO 컨셉을 추적하는 것이다. 이 과정에서 문제가 되는 한 가지 경우로서, 동일한 대응어 집합에 속하는 어휘들이 공유하는 센스가 없는 경우에는 WSD기법에 기반한 예문분석을 통해 개별 어휘의 센스를 확정한 후 이들 센스에 대한 least common subsumer에 해당하는 센스를 WordNet에서 구한 후 그에 대응되는 SUMO 컨셉을 추적하는 방식을 취하였다. 이러한 연구결과 가용한 다양한 언어자원을 활용한 간접 매핑방식을 취했을 때 일관성있고 정확성있는 매핑결과를 얻을 수 있음을 확인할 수 있었다.

Reference

Gruber, Thomas R. (1993): A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition,* 5(2). 199-220.

<일반논문 2>

연관 규칙 마이닝을 활용한 영작문 형태·통사 오류 자동 탐지 및 CFG 오류 규칙 생성

김 동 성 (고려대)

본 연구에서는 일련의 연구에서 수집된 영작문 오류 유형의 정제된 자료를 토대로 연관 규칙을 생성하고, 학습을 통해서 효용성이 검증된 연관 규칙을 활용해서 영작문 데이터의 형태·통사 오류를 자동으로 탐지한다. 영작문 데이터에서 형태·통사 오류를 찾아내는 작업은 많은 시간과 자원이 소요되는 작업이므로 자동화가 필수적이다. 기존의 연구들이 통계적 모델을 활용한 어휘적 오류에 치중하거나 언어 이론적 틀에 근거한 통사 처리에 집중하는 반면에, 본 연구는 데이터 마이닝을 통해서 정제된 데이터에서 연관 규칙을 생성하고 이를 검증한 후 형태·통사 오류를 감지한다. 이전 연구들에서는 이론적 틀에 맞추어진 규칙 생성이나 언어 모델 생성을 위한 대량의 코퍼스 데이터와 같은 다량의 지식베이스 생성이 필수적인데, 본 연구는 적은 양의 정제된 데이터를 활용한다. 영작문 오류 유형의 형태·통사 연관 규칙을 생성하기 위해서 Apriori 알고리즘을 활용하였다. 알고리즘을 통해서 생성된 연관 규칙 중 잘못된 규칙이 생성될 가능성이 있으므로, 상관성 검정, 코사인 유사도와 같은 규칙 효용성의 통계적 검증을 활용해서 타당한 규칙만을 학습하고 축적된 연관 규칙들을 영작문 오류를 자동으로 탐지하는 실험에 활용하였다. 연구 결과로 형태·통사적 문법 오류를 정확하게 탐지함을 알 수 있다. 또한 오류 규칙은 규칙의 패턴이지 정형화된 규칙이 아니다. 정형화된 규칙 생성을 위해서 ATIS-CFG 규칙에 활용해서 CFG 유형의 오류 규칙을 생성하였다.