



자연언어처리와 지식그래프: 한국어와 글로벌 활동

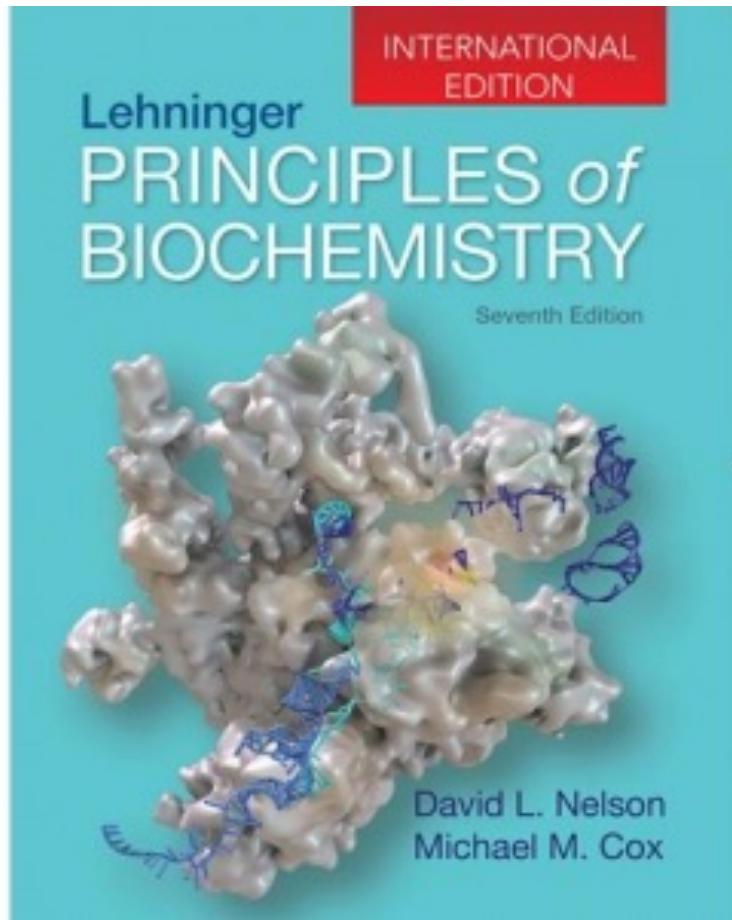
최기선

[지금 이 사람]
국어 정보화와 전문용어
표준화의 선구자- 최기선
한국과학기술원 교수를 만나다

<https://kiss.kstudy.com/public/public2-article.asp?key=50894264>

학부 1학년 때 생각

- 컴퓨터가 언어를 이해할 수 있다면? 사고를 한다면?
- 컴퓨터가 생리학과 같은 조직이 된다면?



전공

수학

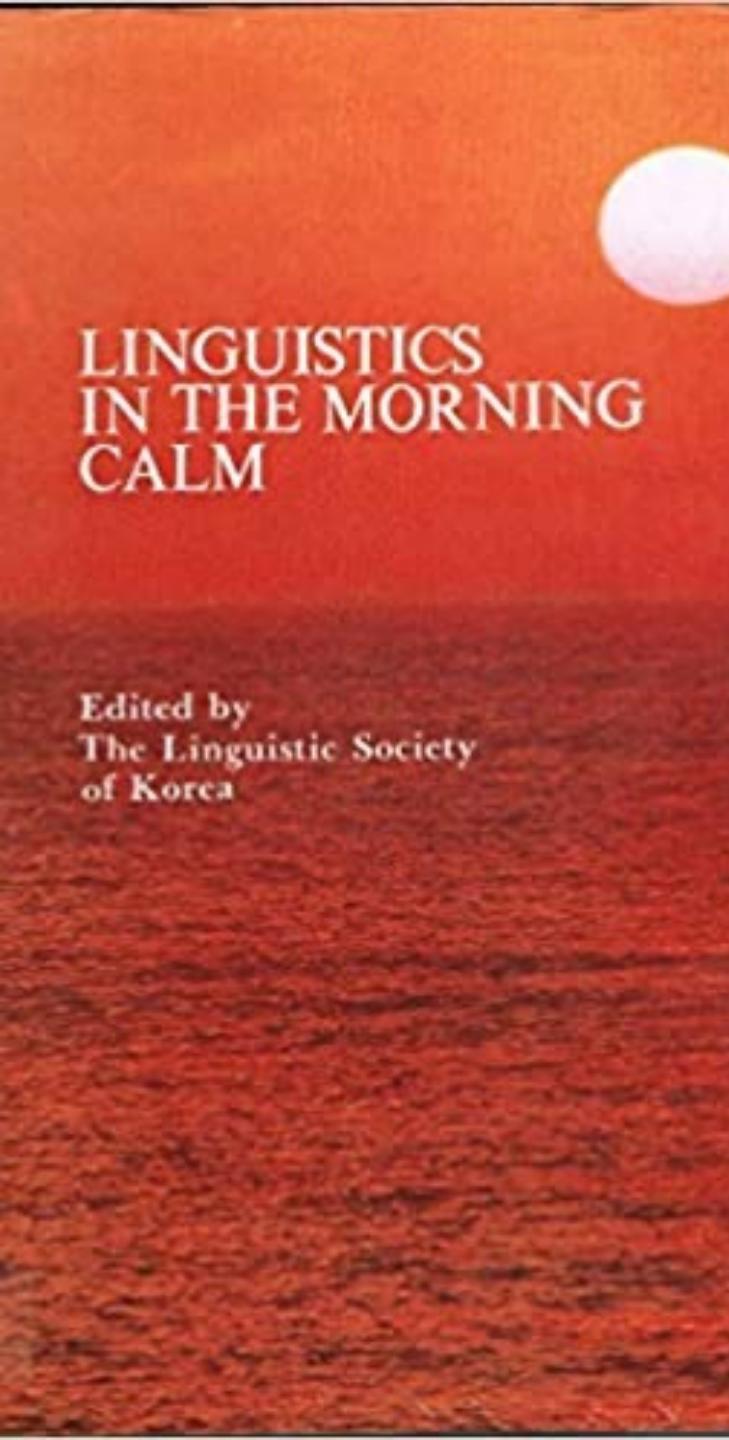
수치해석, 학부

컴퓨터그래픽스 “Hidden surface removal”, 석사

운영체제 (scheduling), 박사 1년

한글 폰트 제작, 코드, 에디터, 박사1년

- 신문사 글자사용빈도 - 코드, 모눈지와 글자꼴 - 폰트
- 한글공학 (1994) 책 편집



LINGUISTICS IN THE MORNING CALM

Edited by
The Linguistic Society
of Korea

조우: 맴돌던 생각의 확인

- SICOL-1981
 - Seoul International Conf. on Linguistics
- 한국외국어대
 - 게시판
- 롯데호텔 장소
- 논문 중 “.... Grammar on infinite memory ...”
- 키워드
 - Noam Chomsky
 - Grammar

허가 받기 전 - 전공 변경 전

SICOL-1981 도중 (박사과정 2년차, 3학기)

- 논문발표 듣고 가능성 확인
- 종로서적센터 (종각 옆, 현재는 없음) 3층 외국서적
 - 책 발견, 3만원 (40년전)
 - 어머니께 부탁, 며칠 후 어머니와 함께 종로서적센터 방문
 - 어머니께서 확인하시고 구입해주심.
- 용어를 모르니, 묻고 또 찾고, 10번 읽으니 이해.
 - 강의를 공식 청강하기로
 - 연세대 국어국문학과, 남기심 교수님 방문
 - 연세대 국어학특강, Truth-conditional semantics 청강 (4학기째)

Model paper 로서

- A Theory of Syntactic Recognition for Natural Language, Mitchell Marcus (1978), MIT PhD Thesis

종로서적에서 구입한 책의 박사학위논문 원판

- 박사학위논문을 책 출간
- 논문 전개 패턴
- 핵심 원칙
- 가벼운 예
- 원칙의 구현 제약
- 실증
- 파서 구조와 학습 가능성
- 효율적 파싱 클래스

A Theory of Syntactic Recognition for Natural Language

by

Mitchell Philip Marcus

A.B., Harvard University (1972)

Submitted in Partial Fulfillment of the Requirements for the Degree of

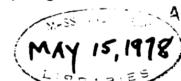
Doctor of Philosophy

at the

Massachusetts Institute of Technology

February, 1978

Signature of Author _____ Signature redacted
Department of Electrical Engineering and Computer Science, October 13, 1977
Certified by _____ Signature redacted Thesis Supervisor _____
Accepted by _____ Chairman, Departmental Committee on Graduate Student _____
Archives



모험 - 자연언어처리 공부 시작, (1981)

인터넷 시대 이전

KAIST 도서관

- 저널 요약 서비스
- 일본 정보처리학회 저널 등
- 논문 신청 - 국내, 일본국회도서관, 영국로얄도서관
- 저자, 편집장에게 편지

오전에 도서관에서 논문 찾기

오후에 도착한 논문 분류, 읽기 - 연구노트 만들기

모두 실험해보기 - 실험결과 노트

Study Group 참석

형식문법연구회

- 매주 월요일 저녁, 중앙대 강형렬 교수님 연구실 등
- 논문에 쓰일 계산가능 모델 발견 (박사과정 5학기)

서울대
어학연구소 여름
특강

- "Script, Plan, Goal and Understanding", Roger Schank
- 강의 준비, 특강

Implementation Tool

- Logic programming, Prolog
 - Edinburgh University, 박사학위논문 결과물, C-Prolog
 - 일본 5세대 컴퓨팅 프로젝트 - AI의 실용화 연구
 - 심볼릭계산 컴퓨터 등장

Life cycle

1. Motivation
2. Motivation의 가능성 확인
3. Self-motivated Learning
4. Presentation
5. Study Group - 인접 분야
6. Implementation - 모델의 실현
7. 국제적 논문 제출
8. Challenge in the unseen context



Determinism
Hypothesis
must be
supported
by the
Structure of
the parser

“Strictly Deterministic” for Parser:

Except “Garden-path sentences”



Human cannot understand
without *difficulty*

*Backtracking if no
enough
“lookahead”*

*Hypothesis-driven
vs. Data-driven*
• *Partially data-driven*

*Hypothesis-driven
= top-down*



Garden path sentence

- E.g.,
- “*The horse raced past the barn fell*”.
- 말이 마굿간을 지나서 달려나갔다 넘어졌다.
- 마굿간을 지나서 달려간 말이 넘어졌다.

Some examples which motivate the structure of the parser

- The parser must
 - Reflect expectations:
 1. I called John to make Sue feel better.
 2. I wanted John to make Sue feel better.
 3. I called [_NP John] [_S to make Sue feel better].
 4. I wanted [_S John to make Sue feel better].
 - Have some sort of Look-ahead:
 - *rf.* Look-ahead parser in program language
 - Be partially data driven

Context-free grammar vs. Efficient Parsability

- Syntactic parser
 - “A theory of syntactic recognition of natural language.” (1978) - Mitchell Marcus.
 - “Towards Minimal Data structures for Deterministic Parsing”, *IJCAI* (1979) - D. Shipman, M. Marcus,
 - “Learning structural descriptions of grammar rules from Examples”, *Linguistics, IJCAI* (1979), R. Berwick,
 - “Context-freeness of the language accepted by Marcus’ parser”, *ACL* (1987) - R. Nozohoor-Farshi,
 - “Parsing Efficiency, computational complexity and the evaluation of grammatical theories”, *Linguistic inquiry* (2008) - Berwick and Weinberg,

국제학술대회 발표

박사학위과정 3년째

일본 Logico-Linguistic Society of Japan

- 연구실 방문 (Tokyo 대, ETL, Tokyo 공대, Kyoto 대 방문)
 - International Colleague 만들어짐. (상호 인정)

미국 Literary and Linguistic Computing
발표

- C-Prolog 저자 (DHD Warren) - SRI 방문
- UC Berkeley, Lofti Zadeh (Fuzzy Theory), Robert Wilensky (AI Plan) 등 방문

직접 해보고 국제적으로
인정받도록 스스로
노력하라.

전문용어의 연구

- 한국어 Key-word in Context (KWIC) 개발
 - Concordance
- TKE (Terminology and Knowledge Engineering), (1990), “KOCP: An Extended OCP for Korean and English”
 - Oxford Concordance Package

The screenshot shows the 'Concordance' tab of the OCP interface. The search term 'jeito' has been entered in the search bar, and the search window size is set to 50. The results list 345 hits, with the first few shown in the table below:

Hit	KWIC	Text Excerpt
33		jeito. Com você não da. Este trabalho me mata.
34		ssim. Com assim. Não come desse jeito. Com assim. Não segue o filhote desse jeito
35		ito com os pais! Não fale desse jeito comigo! Não me magoe deste jeito. É preciso
36		se? Por que você está falando desse jeito comigo? Você parece estar um pouco exaltado.
37		ava convencido. Tom não gostou do jeito como Maria cozinhava. Eu tenho medo dos polic
38		aria de te ouvir cantar. Fascina-me o jeito como você canta. A população anda mais anten
39		iram para acompanhá-los. Eu amo o jeito como você me surpreende. E você chama isso
40		em senso não falaria comigo daquele jeito. Cristianismo e islamismo são duas religiões
41		ar o jeito disso. Tom está pegando o jeito da coisa. Ela está tremendo de frio. Ela
42		cozinhar carne? Para quê? Não tem jeito de abrir esta porta. O idoso morreu de
43		encontrar aqui! Estão buscando um jeito de acelerar o rendimento da máquina sem alte
44		ambicioso. Eu o reconheci pelo seu jeito de andar. A amabilidade não é seu forte.
45		onsigo, mais quero. Eu gosto do seu jeito de andar. Quer beber? Escove os dentes antes
46		enteceu. O Tom encontrou um novo jeito de anunciar as notícias. Em altas altitudes
47		trava. Brincar é sem dúvida o melhor jeito de aprender. Encontrei um jeito de mudar o
48		jeito de baixar vídeos do YouTube? Eu tento, mas
49		ação de teclas. Você conhece algum jeito de caminhar. Amanhã nos encontraremos no res
50		novo. O que tenho de novo é o jeito de cantar me agrada. Eu sei que você
51		gosto do seu jeito de cantar. Seu jeito de cantar. Na manhã seguinte, o boneco de
52		inteligente e atraente. Eu adoro o seu jeito de cantar. Seu jeito de cantar me agrada.
53		a voz. Acho que Tom odeia o meu jeito de cantar. Vocês já sabiam disso? Tu já
54		dizer "eu te amo". Ela deu uma forma de fazer isso. Deve haver um jeito de contornar a situação
55		embarrassante em que se
56		jeito de convencer o Tom a trabalhar com a
57		jeito de cuidar de mim. Eu adoro seu jeito de cuidar de mim. Vocês já sabiam disso? Tu já

Specific- Domain Language

Terminology Science로의 초대

- 한국어 KWIC에서
- 일본전문용어 협회장의 초대 (1990, 가을)

국립국어원

- 이공학 전문용어 표준화 (1997-2006)

전문용어 표준화에 대한 공부

- ISO/TC37 전문용어 표준화 위원회
총무간사 방문
- Christian Galinski (비엔나)



International Recognition Procedure

My History of International Recognition procedure

한국어 KWIC - 데이터 현상: 연구개발 - 실증

국내외 발표

일본, 미국, 비엔나 등: Colleague 만남

국내에서의 인정, 필요성을 평소에 피력

국제적 관련자 방문, 필요한 공식절차

International Service and Recognition

ISO의 요청: 국제적 니즈의 도래

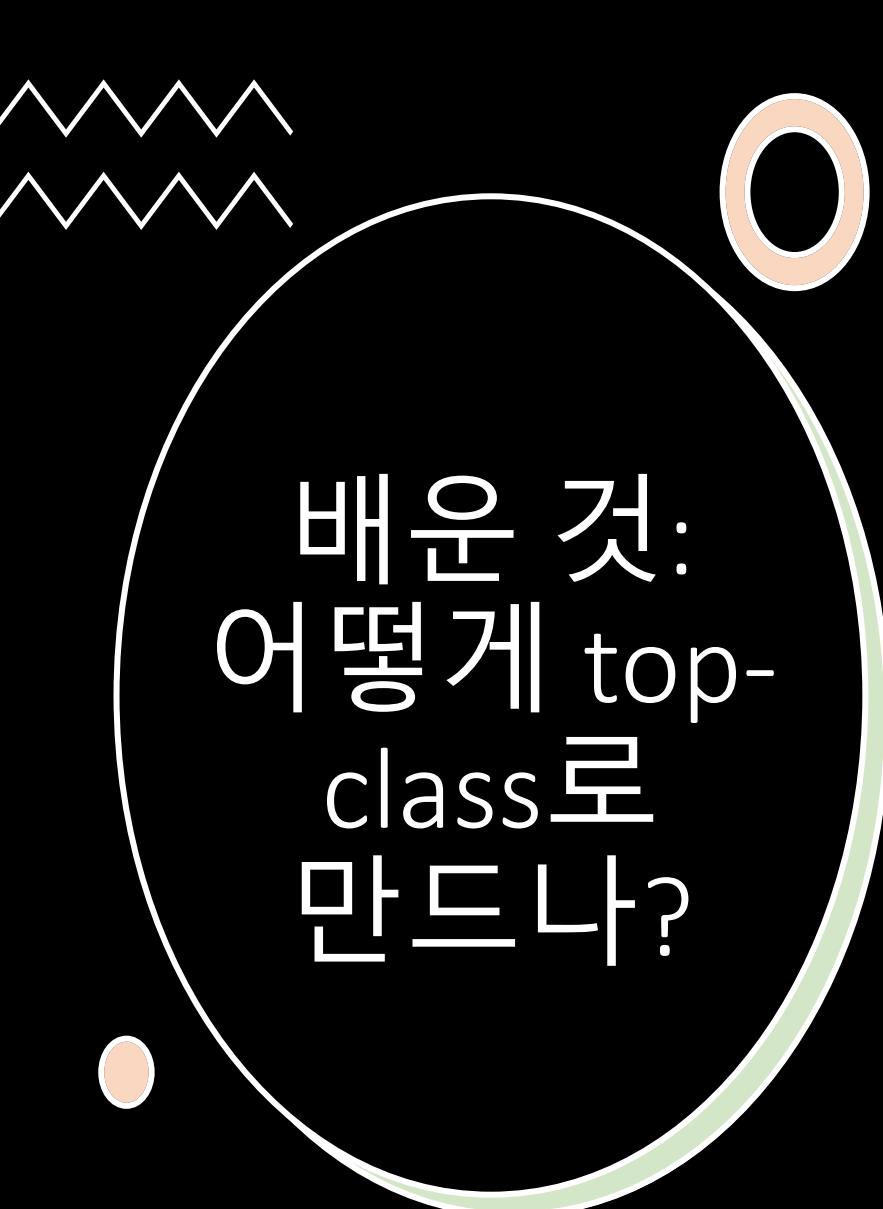
1993년 당시 질문:

왜 아시아에서는 Top-class conference가 개최되지 않는가?

1995, 3rd NLPRS (Natural language processing
Symposium) 한국 개최

어떻게 했나?

- 국제적 top-class invited speaker (3명)
- 주변국가 top-class scholar 초빙 (중국 2, 일본 2, 동남아시아 3)
- 일본 자연언어처리학회, PC 위원장 초빙



배운 것:
어떻게 top-
class로
만드나?

- Procedure in General
 - 1. 중요 인사 초빙
 - 2. 주변 중요 인사 초빙
 - 3. 주변 top-class 인정
 - 4. 국제적 인정 받음

그 다음 효과

아시아 중심체 AFNLP 창설

- Asia Federation of NLP (2005-),
<http://www.afnlp.org/>
- Regional Representative, Board members (종신직)

Asian Initiative

- Not just advanced NLP
- but support on the weak NLP area in Asia



Very very
private
experience: //

- 1998, Korterm 창립
(Korea terminology research center for language engineering)
 - ISO/TC37 Terminology 위원회 임원 초빙 - 강연
 - Eafterm (East-Asia terminology forum) 창립 멤버

2000년 5월 29일(월)
LREC 아테네 학회
중

- Terminology resources and computation workshop 개최 함
 - <http://lrec-conf.org/lrec2000/korterm.kaist.ac.kr/lrec2000/index.html>
- Terminology workshop을 LREC 언어공학 컨퍼런스에서 최초 개최
- Leadership, Christian Galinski (chair), 최기선 (associate chair)
- All traditional leaders all together
 - They would like to know about the organizers.

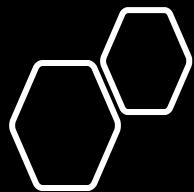
5월 30일(화)

Christian Galinski와 조찬

ISO 표준 만드는 방법 등 전수

저녁 때

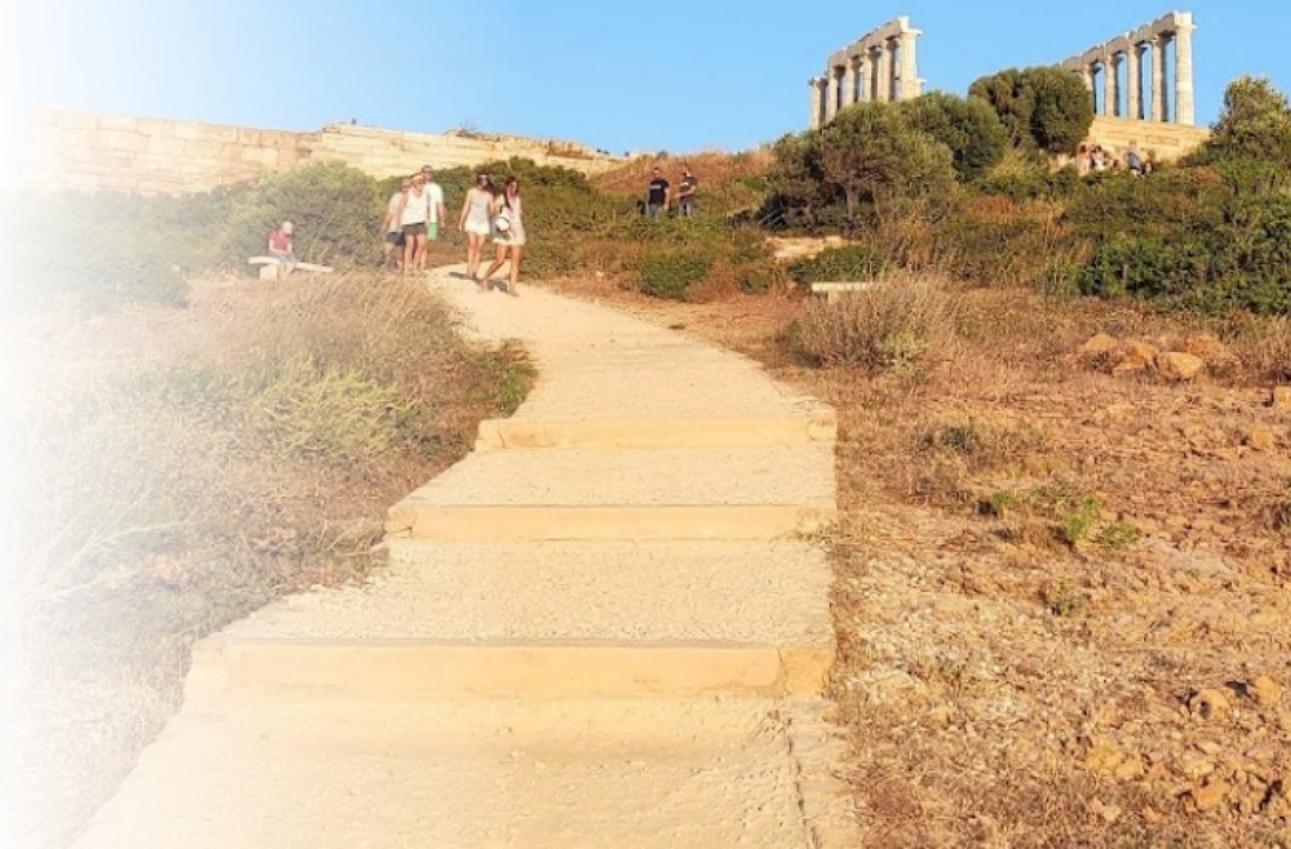
- 수요일날 시외버스를 타고
- 포세이돈 신전에 가서 의논하자고 함.

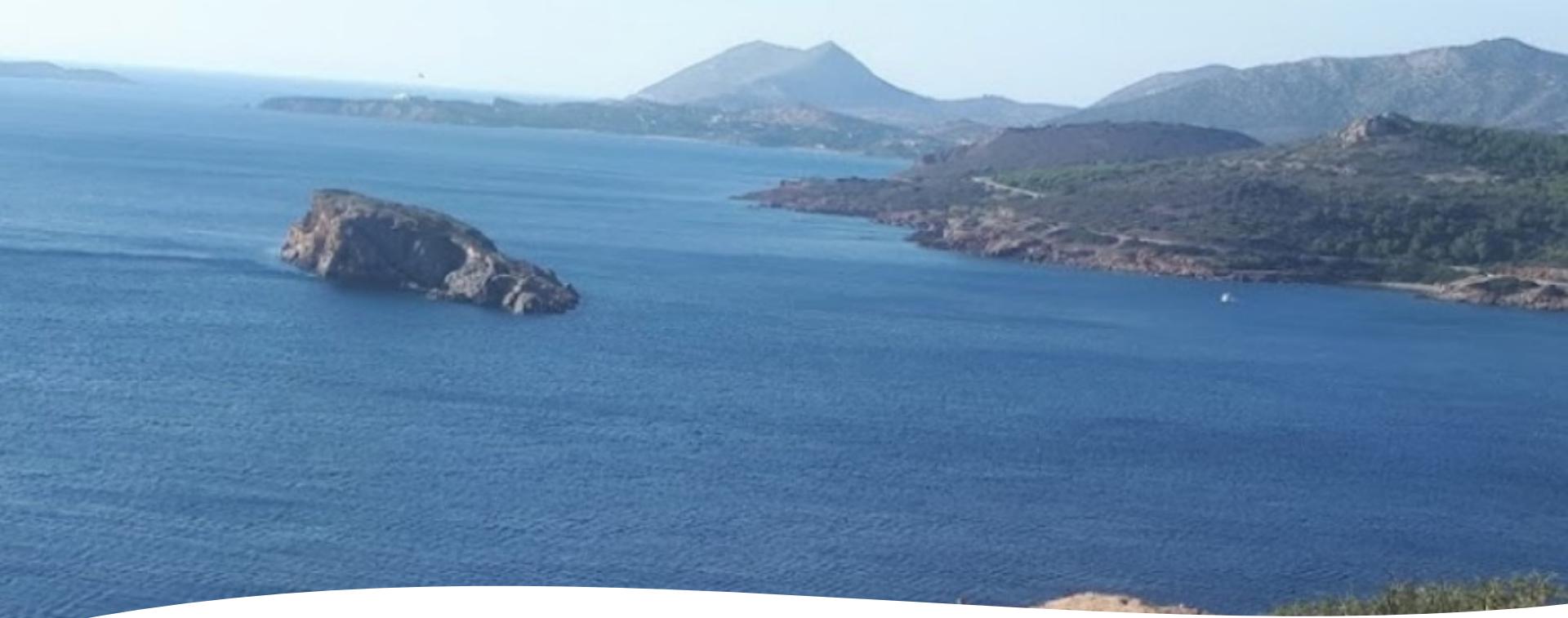


5월31일(수)

- 아침 7시 버스 정거장
- 포세이돈 신전 앞 하차

포세이돈 신전에서





해변에서
바위섬
돌아오기



해변에서 발동력 놀이보트 타고
바다 한 가운데에서 수영하기

6월1일(목) 점심

- ISO/TC37 leadership lunch meeting에 초청 받음
 - 일종의 oral test
 - ISO/TC37/SC4 새로운 커미티 창설 제안서 작성
 - 6월2일(금)에 제안서 검토하기로 함.

2001년 유럽 내 리더 방문

2002년 창설하기
위하여

유럽 내 리더
방문

Global leadership
인정

Language Resource Management Standard Committee 창립

2002

- 언어자원관리
- 데이터 관리
- Annotation Standard

현재까지

- Committee manager, 종신직
- <https://www.iso.org/committee/297592.html>
- 국제 community에서의 leadership
- 국제표준 임원으로서 한국 2번째 (2002년)



ISO/TC 37

ISO/TC 37/SC 4

Language resource management

ABOUT

SECRETARIAT: [KATS](#)

Committee Manager: [Prof Key-Sun Choi](#)

Chairperson (until end 2023): [Ms Nicoletta Calzolari Zamorani](#)

ISO Technical Programme Manager [TPM]: [Mme Mercè Ferrés Hernández](#)

ISO Editorial Manager [EM]: [Ms Nicola Perou](#)

Creation date: 2001

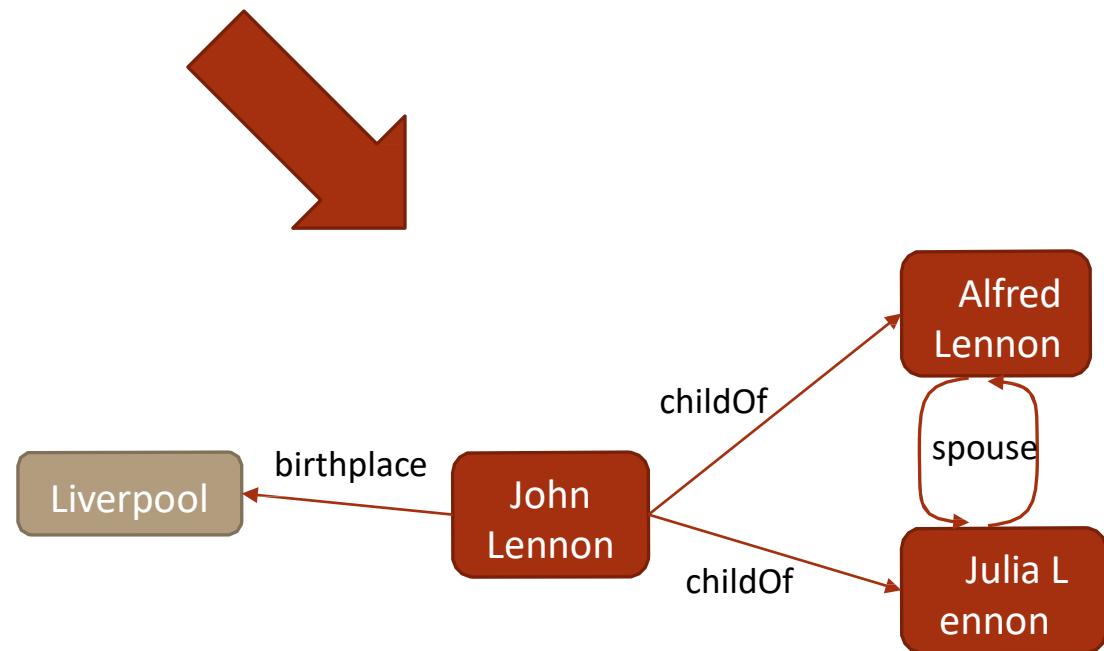
SCOPE

Standardization of the modelling, specification, design, documentation and encoding of digital language resources to enable integration, interchange and replicability.

언어와 지식 그래프

Why do we need NLP?

John was born in Liverpool, to Julia and Alfred Lennon.



NLP로 문장을 그래프

*John was born in Liverpool,
to Julia and Alfred Lennon.*

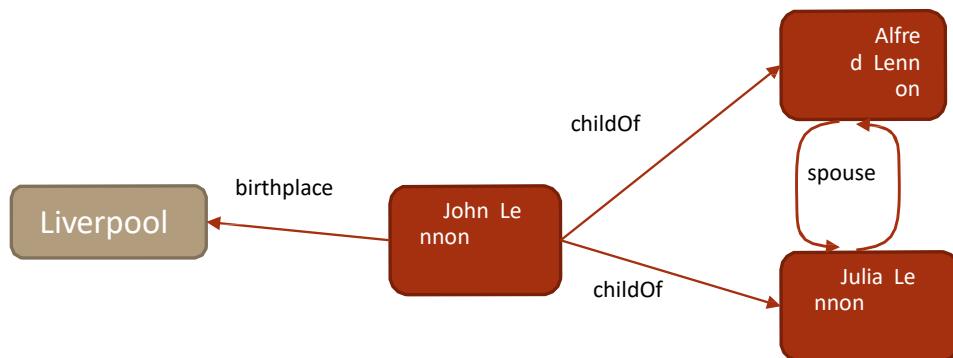
<John Lennon, hasBirthplace,
Liverpool>

<John Lennon, isChildOf, Alfred
Lennon>

<John Lennon, isChildof, Julia
Lennon>

<Alfred Lennon, isSpouseOf,
Juia Lennon>

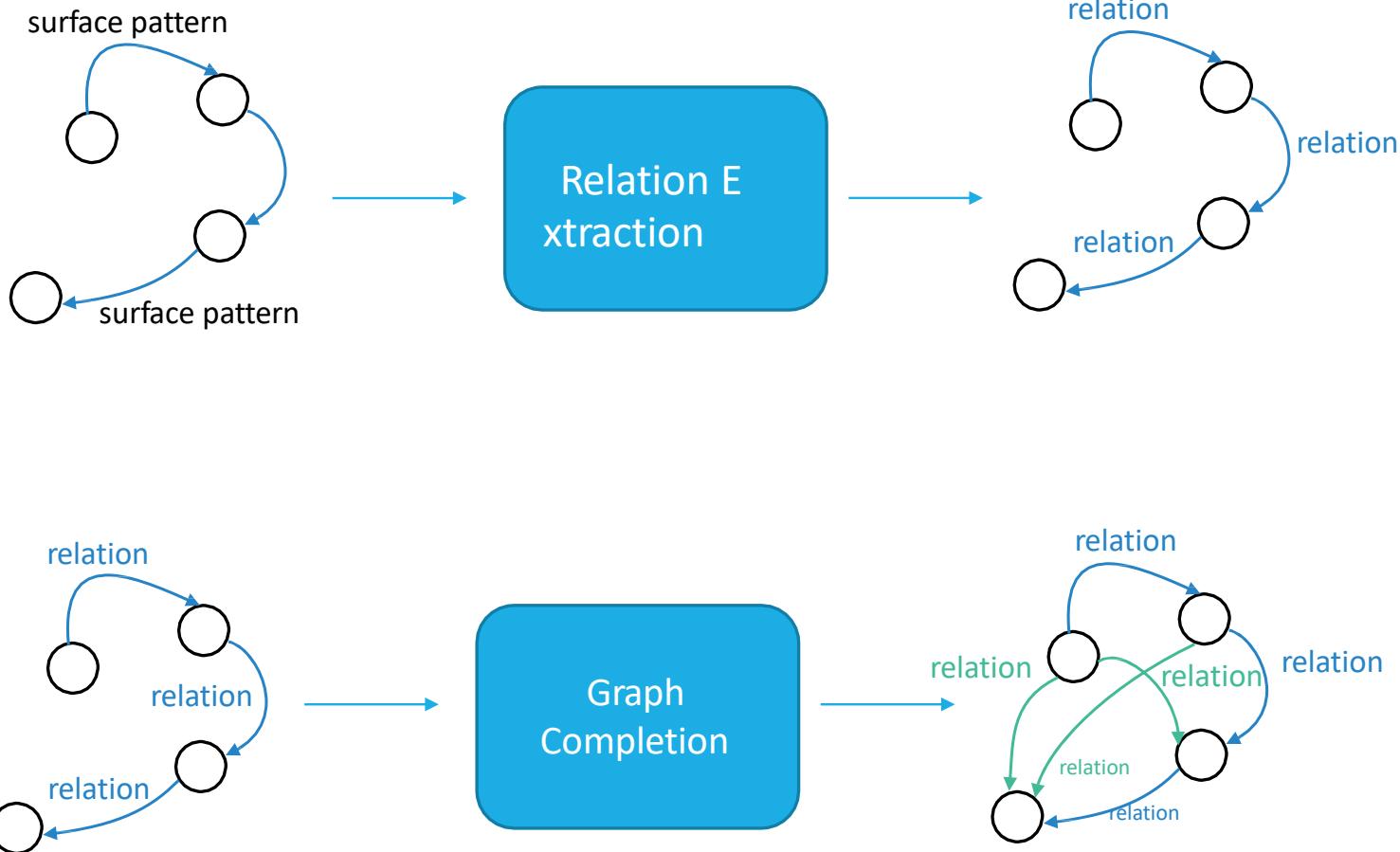
<Julia Lennon, isSpouseOf,
Alfred Lennon>



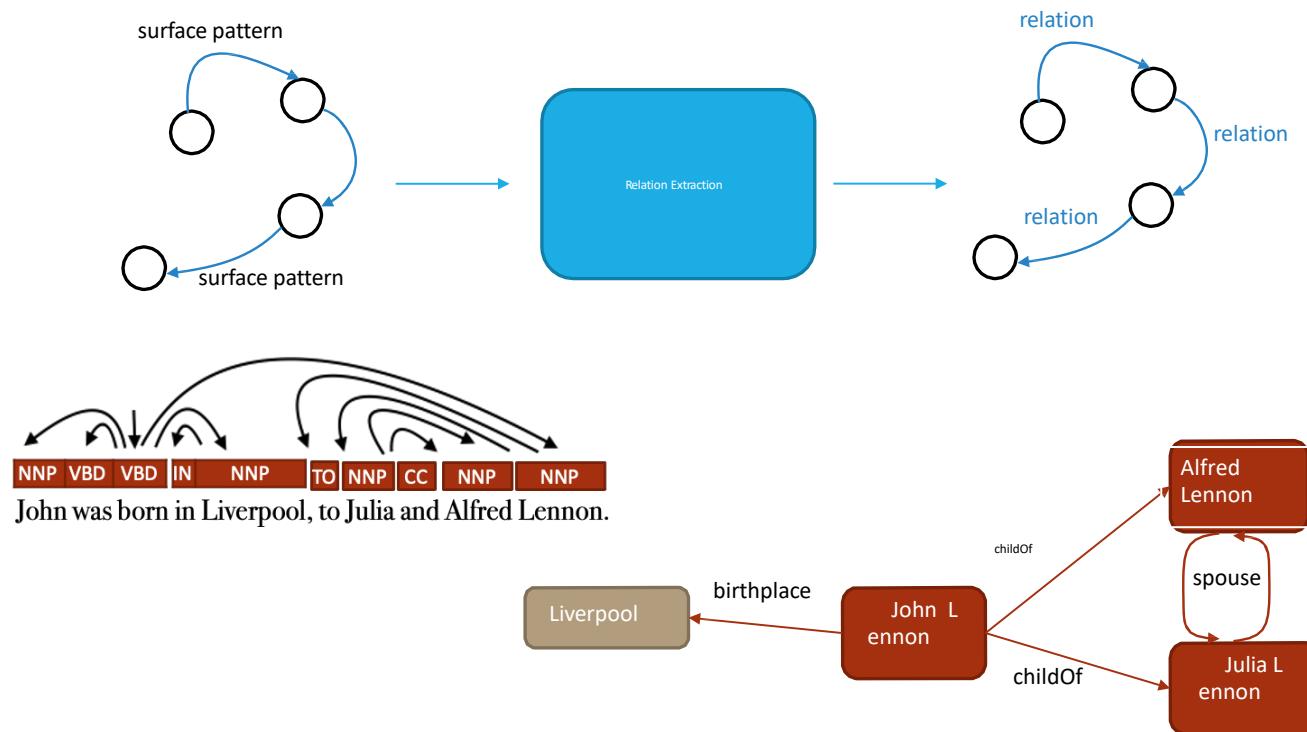
Embedding Techniques

- MATRICES, TENSORS, AND NEURAL NETWORKS

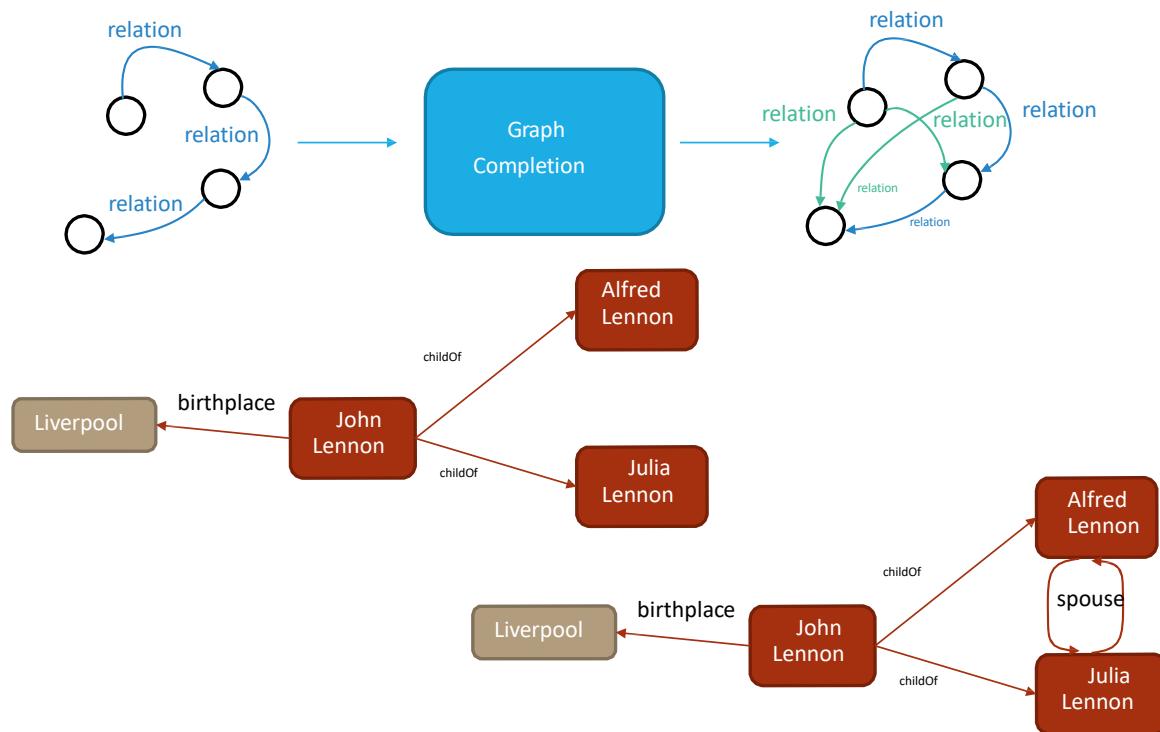
Two related tasks



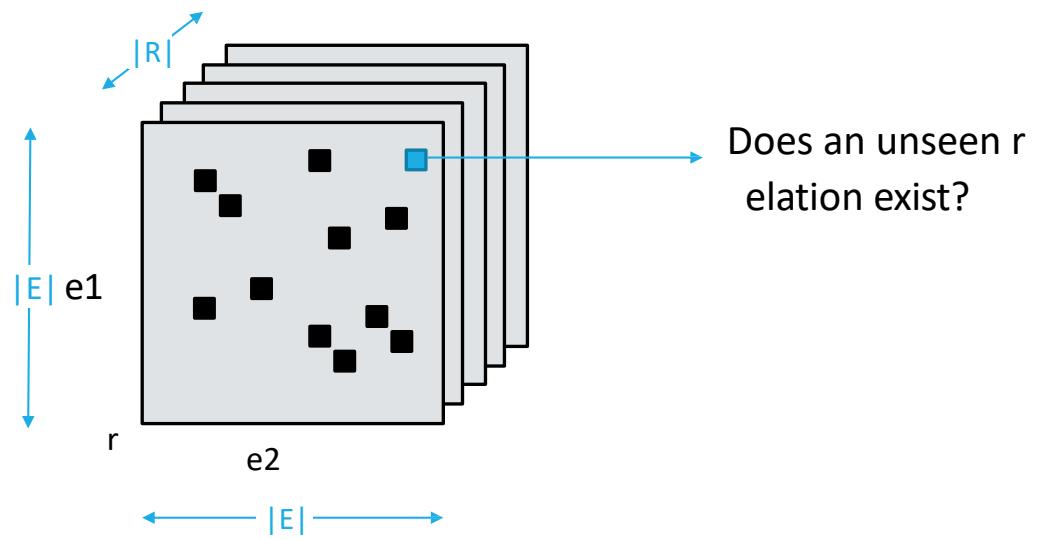
Relation extraction: predicate linking



Graph completion: Entity + relation linking + alpha



Tensor Formulation of KG



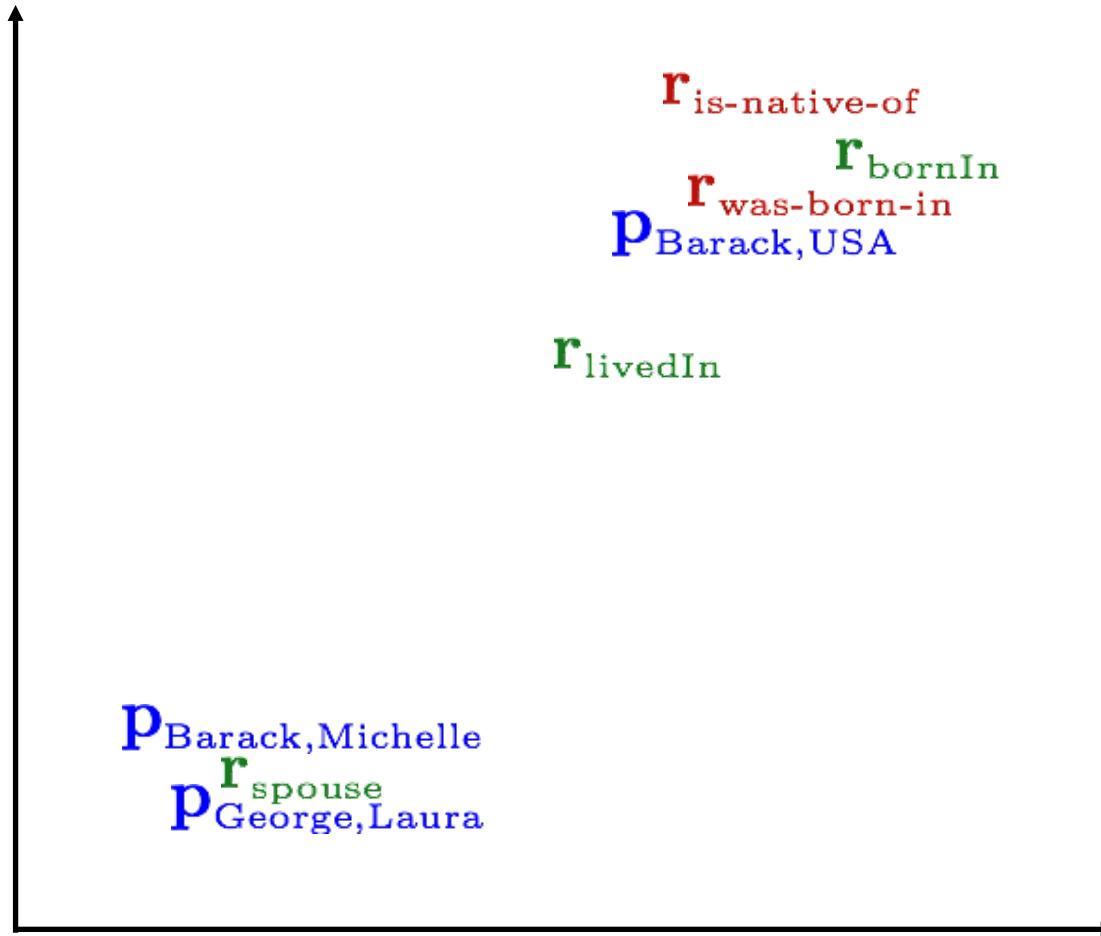
Relation Extraction as a Matrix

Sanders was born in Brooklyn, to Dorothy and Eli Sanders.

Entity Pairs

	<i>was born in</i> <i><nsbjpass:born<-nmod:in-</i>	<i>was born to</i>	<i>and</i>	<i>birthplace(X,Y)</i>	<i>spouse(X,Y)</i>
Bernie Sanders, Brooklyn	1				?
Bernie Sanders, Dorothy Sanders		1			
Bernie Sanders, Eli Sanders			1		
Dorothy Sanders, Eli Sanders				1	?
Barack Obama, Hawaii	1				1
Barack Obama, Michelle Obama			1		1

Relation Embeddings



대표 연구

자연언어처리와 지식그래프 (시맨틱웹)

- 국가 인공지능 프로젝트: ExoBrain (WiseKB) 지식베이스 구축

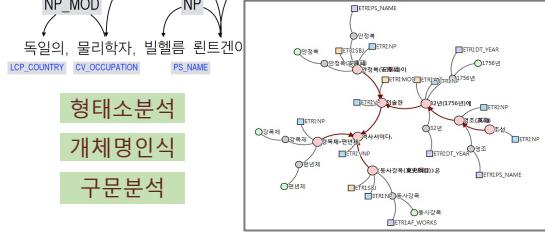
인간처럼 학습, 판단 가능한 미래 AI 기술



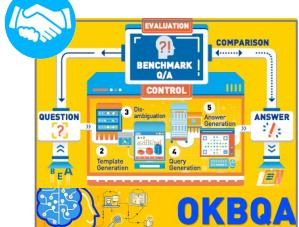
- ## ▶ Mobile 질의응답 (Eurostar2)



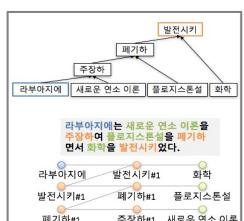
- ▶ 한국어 자연언어처리, WordNet, FrameNet



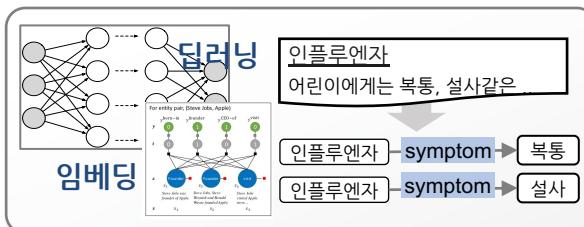
-  www.okbqa.org
Open KB and QA



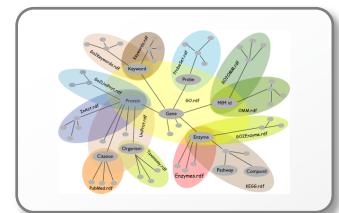
- ## ▶ 자연어맥락 지식 그래프 생성



- ▶ 인공지능 기반 정보 추출 기술



- ## ▶ 대용량 DB 구축 및 표준화



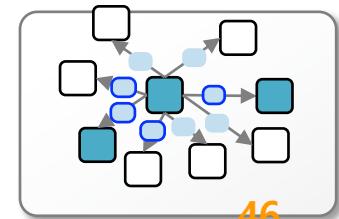
- ## ▶ 개체 연결 기법



- 국가지정 언어자원은행 유



- ## ▶ 그래프(문서/DB) 요약 기술



최기선 (KAIST 전산학부 명예교수, 건양대 의료인공지능학과 석좌교수) kschoi@kaist.ac.kr, kschoi@konyang.ac.kr

국제협력연구과제

- LOD2, EU FP7
 - Semantic Linked Open Data
- QAMEL (<http://qamel.eu/>) , EuroStar2
 - Question-Answering for Mobile Environment

ISO/TC37/SC4 (언어자원관리 표준위원회)

- Committee manager: Key-Sun Choi (창립자, 종신직)

수상

- 한글날 (2015) 유공 옥조근정훈장
- 한글 및 한국어 자연언어처리 효시와 국제화
- Eugen Wüster Prize (2014), 전문용어학의 공학화
 - Infoterm, UNESCO Multi-lingual Chair, Vienna Univ.
- KAIST 연구대상 (1998)



전산학부장 (2006-2011)

AFNLP 회장 (Asia Federation of NLP) (2009-2010), 창립



Infoterm 현재 이사장

COLING 2022

THE 29TH INTERNATIONAL CONFERENCE ON
COMPUTATIONAL LINGUISTICS

October 12 - 17, 2022

Gyeongju, Republic of Korea