WordNet
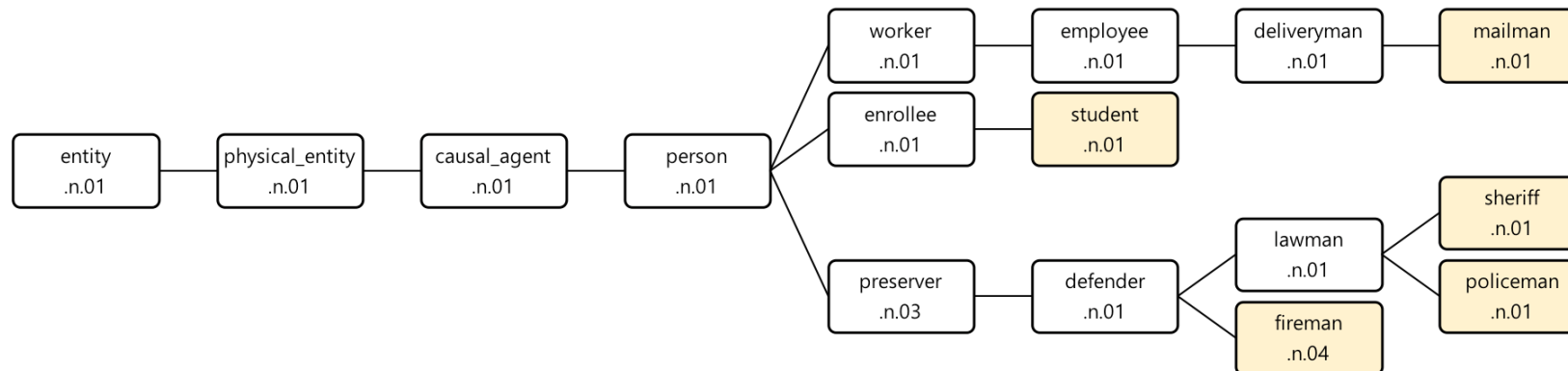
Ki Hyun Kim

nlp.with.deep.learning@gmail.com

WordNet

- Thesaurus(어휘분류사전, 시소러스)

- 심리학 교수인 George Armitage Miller 교수가 지도하에 프린스턴 대학에서 1985년부터 만든 프로그램

- 동의어 집합(Synset) 또는 상위어(Hypernym)나 하위어(Hyponym)에 대한 정보가 특히 잘 구축되어 있는 것이 장점
  - Directed Acyclic Graph(유방향 비순환 그래프)

WordNet

- NLTK or Web
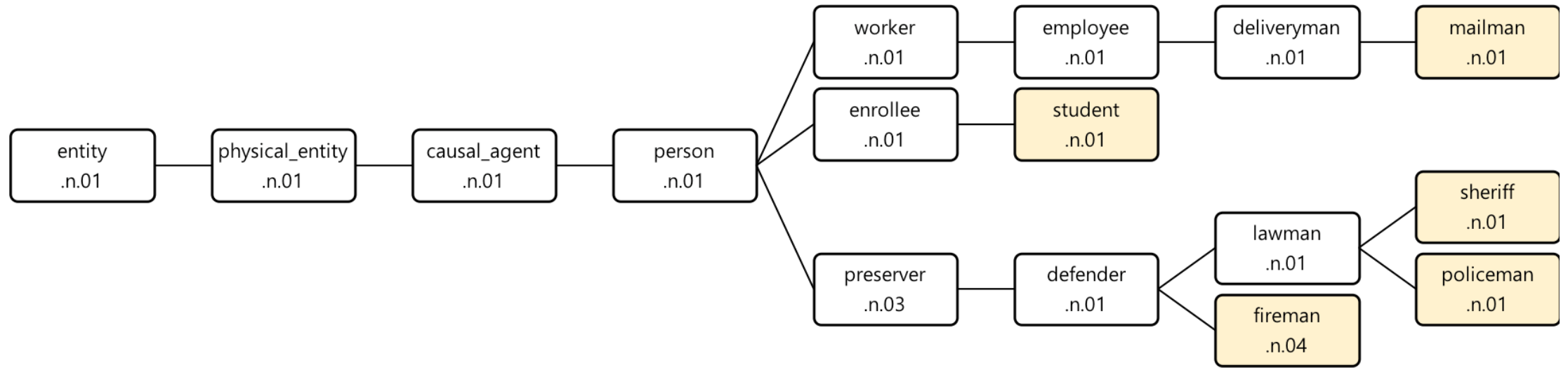 - http://wordnetweb.princeton.edu/perl/webwn

WordNet

- Search Result: bank

Noun

- S: (n) **bank#1** (sloping land (especially the slope beside a body of water)) *"they pulled the canoe up on the bank"; "he sat on the bank of the river and watched the currents"*
- S: (n) depository financial institution#1, **bank#2**, banking concern#1, banking company#1 (a financial institution that accepts deposits and channels the money into lending activities) *"he cashed a check at the bank"; "that bank holds the mortgage on my home"*
- S: (n) **bank#3** (a long ridge or pile) *"a huge bank of earth"*
- S: (n) **bank#4** (an arrangement of similar objects in a row or in tiers) *"he operated a bank of switches"*
- S: (n) **bank#5** (a supply or stock held in reserve for future use (especially in emergencies))
- S: (n) **bank#6** (the funds held by a gambling house or the dealer in some gambling games) *"he tried to break the bank at Monte Carlo"*
- S: (n) **bank#7**, cant#2, camber#2 (a slope in the turn of a road or track; the outside is higher than the inside in order to reduce the effects of centrifugal force)
- S: (n) savings bank#2, coin bank#1, money box#1, **bank#8** (a container (usually with a slot in the top) for keeping money at home) *"the coin bank was empty"*
- S: (n) **bank#9**, bank building#1 (a building in which the business of banking transacted) *"the bank is on the corner of Nassau and Witherspoon"*
- S: (n) **bank#10** (a flight maneuver; aircraft tips laterally about its longitudinal axis (especially in turning)) *"the plane went into a steep bank"*
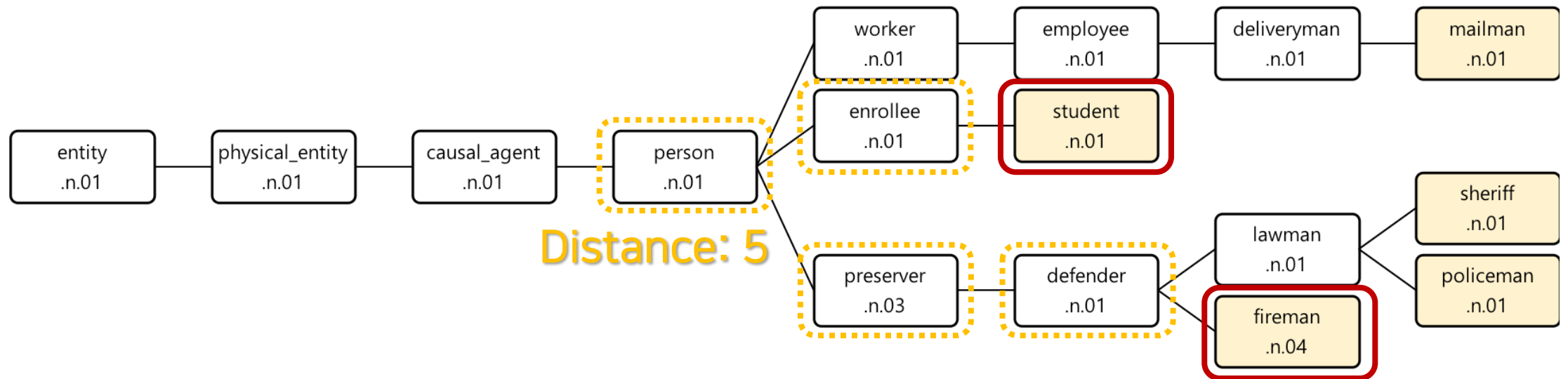
Korean WordNet

이름	기관	웹사이트
KorLex	부산대학교	http://korlex.pusan.ac.kr/
Korean WordNet(KWN)	KAIST	http://wordnet.kaist.ac.kr/

Hierarchy in WordNet

Distance between Words

$$\text{similarity}(w, w') = -\log \text{distance}(w, w')$$

Summary

- WordNet을 통해
  - 단어의 계층적 구조를 파악할 수 있음
  - 동의어 집합(synset)을 구할 수 있음
  - 단어 사이의 유사도를 계산할 수 있음

이 모든 것이 코퍼스 없이 가능!

- 하지만 다음의 경우,
  - 어떤 단어에 대한 자료를 만들던 날, Miller 교수님이 부인과 다투었음
  - 특정 도메인(또는 수집된 데이터)에 특화된 수치를 계산하고 싶음
  - 신조어나 사전에 등록되지 않은 단어

Data-driven 방식의 필요성