# Named Entity Recognition

Hyunjoong Kim

soy.lovit@gmail.com

github.com/lovit

# Named Entity Recognition

- NER 은 단어열에서 특정 종류의 단어를 찾는 문제입니다.

Input: [디카프리오, 가, 나온, 영화, 좀, 추천, 해줘]

Output: [배우, None, None, None, None, Request, None]

- Information Extraction 의 tasks 중 하나입니다.
  - co-reference resolution, relationship extraction 도 IE 의 주요 문제입니다.

# Named Entity Recognition

- NER 은 단어열에서 특정 종류의 단어를 찾는 문제입니다.
  - 사람, 시간, 장소, 단체 등의 단어를 찾습니다.

- The decision by the independent MP Andrew Wilkie to withdraw his support for the minority Labor government sounded dramatic but it should not further threaten its stability. When, after the 2010 election, Wilkie, Rob Oakeshott, Tony Windsor and the Greens agreed to support Labor, they gave just two guarantees: confidence and supply.

Person
Date
Location
Organi-
zation

slide from Christopher Manning's lecture note

# Named Entity Recognition

- Sequential labeling 방법은 NER 에 자주 이용되었습니다.

Input: [디카프리오, 가, 나온, 영화, 좀, 추천, 해줘]

Output: [배우, None, None, None, None, Request, None]

- Stanford NLP group 에서도 CRF 기반 NER tagger 를 제공합니다.

- https://nlp.stanford.edu/software/CRF-NER.shtml
- https://medium.com/@dudsdu/named-entity-recognition-for-unstructured-documents-c325d47c7e3a

# Named Entity Recognition

- pycrfsuite 튜토리얼에서 CoNLL 2002 NER 코드를 제공합니다.
  - 네 가지 names 를 인식하는 문제입니다.
    - (PER) : person names
    - (ORG) : organizations
    - (LOC) : locations
    - (MISC) : miscellaneous



https://github.com/scrapinghub/python-crfsuite/blob/master/examples/CoNLL%202002.ipynb

# Named Entity Recognition

- 두 개 이상의 단어로 이뤄진 named entities 를 표현하기 위하여 “B, I, O” tag set 이 이용됩니다.
  - B : first item
  - I : non-initial word
  - O : others

| Word | Tags |
|---|---|
| Wolff | B-PER |
| , | O |
| currently | O |
| a | O |
| journalist | O |
| in | O |
| Argentina | B-LOC |
| , | O |
| played | O |
| with | O |
| Del | B-PER |
| Bosque | I-PER |
| in | O |
| the | O |
| final | O |
| years | O |
| of | O |
| the | O |
| seventies | O |
| in | O |
| Real | B-ORG |
| Madrid | I-ORG |
| . | O |

# Named Entity Recognition

- pycrfsuite 튜토리얼에서 CoNLL 2002 NER 코드를 제공합니다.

features

| Feature / position | $i-1$ | $i$ | $i+1$ |
|---|---|---|---|
| bias | | o | |
| word lower | o | o | o |
| word[-3:] | | o | |
| word[-2:] | | o | |
| word is upper? | o | o | o |
| word is title? | o | o | o |
| word is digit? | | o | |
| postag | o | o | o |
| postag[:2] | o | o | o |

performance

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| B-LOC | 0.78 | 0.75 | 0.76 | 1084 |
| I-LOC | 0.87 | 0.93 | 0.9 | 634 |
| B-MISC | 0.69 | 0.47 | 0.56 | 339 |
| I-MISC | 0.87 | 0.93 | 0.9 | 634 |
| B-ORG | 0.82 | 0.87 | 0.84 | 735 |
| I-ORG | 0.87 | 0.93 | 0.9 | 634 |
| B-PER | 0.61 | 0.49 | 0.54 | 557 |
| I-PER | 0.87 | 0.93 | 0.9 | 634 |
| **AVERAGE** | **0.81** | **0.81** | **0.8** | **5251** |



https://github.com/scrapinghub/python-crfsuite/blob/master/examples/CoNLL%202002.ipynb

# Precision / Recall / F1-score / Accuracy

- Precision : $\frac{|true_{pos} \cap pred_{pos}|}{|pred_{pos}|}$
- Recall : $\frac{|true_{pos} \cap pred_{pos}|}{|true_{pos}|}$
- F1 – measure : $2 \times \frac{precion \times recall}{precision + recall}$
- Accuracy : $\frac{|true_{pos} \cap true_{neg}|}{|pos+neg|}$

# Precision / Recall / F1-score / Accuracy

| | Predict positive | Predict negative | sum |
|---|---|---|---|
| True positive | 800 | 200 | 1000 |
| True negative | 400 | 600 | 1000 |
| sum | 1200 | 800 | 2000 |

| | |
|---|---|
| Precision | $\frac{\lvert true_{pos} \cap pred_{pos} \rvert}{\lvert pred_{pos} \rvert} = \frac{800}{1200}$ |
| Recall | $\frac{\lvert true_{pos} \cap pred_{pos} \rvert}{\lvert true_{pos} \rvert} = \frac{800}{1000}$ |
| F1 measure | $2 \times \frac{precion \times recall}{precision + recall} = 2 \times \frac{\frac{8}{12} \times \frac{8}{10}}{\frac{8}{12} + \frac{8}{10}}$ |
| Accuracy | $\frac{\lvert true_{pos} \cap true_{neg} \rvert}{\lvert pos + neg \rvert} = \frac{800 + 600}{2000}$ |

# Named Entity Recognition

- 앞, 뒤의 단어만을 이용하여 NER 을 수행합니다.
  - 현실적으로 마련하기 어려거나 cheat 성격을 가진 features 를 제거합니다.

features

| Feature / position | $i-1$ | $i$ | $i+1$ |
|---|---|---|---|
| **bias** | | **o** | |
| **word lower** | **o** | o | **o** |
| **word[-3:]** | | **o** | |
| **word[-2:]** | | **o** | |
| word is upper? | o | o | o |
| word is title? | o | o | o |
| word is digit? | | o | |
| postag | o | o | o |
| postag[:2] | o | o | o |

performance

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| B-LOC | 0.69 | 0.49 | 0.58 | 1084 |
| I-LOC | 0.6 | 0.47 | 0.52 | 325 |
| B-MISC | 0.52 | 0.2 | 0.29 | 339 |
| I-MISC | 0.52 | 0.36 | 0.43 | 557 |
| B-ORG | 0.74 | 0.55 | 0.63 | 1400 |
| I-ORG | 0.71 | 0.52 | 0.6 | 1104 |
| B-PER | 0.83 | 0.69 | 0.76 | 735 |
| I-PER | 0.86 | 0.86 | 0.86 | 634 |
| **AVERAGE** | **0.71** | **0.54** | **0.61** | **6178** |

https://github.com/scrapinghub/python-crfsuite/blob/master/examples/CoNLL%202002.ipynb

# Named Entity Recognition

- 성능이 아주 많이 차이나지는 않습니다.

  NER 에서 가장 중요한 정보는 앞/뒤에 등장하는 단어입니다.

With only previous/next words

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| B-LOC | 0.69 | 0.49 | 0.58 | 1084 |
| I-LOC | 0.6 | 0.47 | 0.52 | 325 |
| B-MISC | 0.52 | 0.2 | 0.29 | 339 |
| I-MISC | 0.52 | 0.36 | 0.43 | 557 |
| B-ORG | 0.74 | 0.55 | 0.63 | 1400 |
| I-ORG | 0.71 | 0.52 | 0.6 | 1104 |
| B-PER | 0.83 | 0.69 | 0.76 | 735 |
| I-PER | 0.86 | 0.86 | 0.86 | 634 |
| **AVERAGE** | **0.71** | **0.54** | **0.61** | **6178** |

All features

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| B-LOC | 0.78 | 0.75 | 0.76 | 1084 |
| I-LOC | 0.87 | 0.93 | 0.9 | 634 |
| B-MISC | 0.69 | 0.47 | 0.56 | 339 |
| I-MISC | 0.87 | 0.93 | 0.9 | 634 |
| B-ORG | 0.82 | 0.87 | 0.84 | 735 |
| I-ORG | 0.87 | 0.93 | 0.9 | 634 |
| B-PER | 0.61 | 0.49 | 0.54 | 557 |
| I-PER | 0.87 | 0.93 | 0.9 | 634 |
| **AVERAGE** | **0.81** | **0.81** | **0.8** | **5251** |

# Named Entity Recognition

- CRF 를 이용하여 NER 문제를 풀 경우에는, 유용한 features 를 potential function 으로 잘 설계하는 것이 중요합니다.

# Named Entity Recognition

- 그러나 CRF 를 이용하기 위해서는 학습데이터가 필요합니다.
  - 우리가 원하는 named entities 는 장소/조직이 아닐 경우가 많습니다.
  - 각자의 문제에 맞는 NER 을 위한 학습데이터가 필요합니다.
- ICLR 2018 에서 active learning 을 이용하는 NER 논문이 제안되었습니다.



Shen, Y., Yun, H., Lipton, Z. C., Kronrod, Y., & Anandkumar, A. (2017). Deep Active Learning for Named Entity Recognition. *arXiv:1707.05928*.

# Named Entity Recognition

- CRF 기반 NER 이 학습하는 유용한 features 는 앞/뒤의 단어입니다.
  - 앞 단어가 'en' 이면 장소일 가능성이 높습니다.
  - 스페인어의 'en' 은 영어의 'in' 처럼 장소 앞에 등장하는 전치사입니다.

```
...
('-1:word.lower=en', 'B-LOC') : 3.543269
('+1:word.lower=24', 'B-LOC') : 3.542004
('-1:word.lower=hacia', 'B-LOC') : 3.536268
...
```

CoNLL2002, 스페인어 NER 의 CRF feature weight

# Named Entity Recognition

- CRF 기반 NER 이 학습하는 유용한 features 는 앞/뒤의 단어입니다.
  - 앞, 뒤의 단어들은 교통수단을 인식하는데 유용한 힌트입니다.
    - (x[-1]=에는, x[1] = 타고)
    - (x[1:2] = 타고 –가고)

지금 [버스] 타고 가고 있어

집에는 [버스] 타고 갈래?

더운데 [택시] 잡자

# Named Entity Recognition

- CRF 처럼 모든 단어에 대하여 태깅을 할 필요도 없습니다.
  - Named entities 는 전체 문장에서 등장하는 비율이 작습니다.
  - Class imbalanced 영향도 발생합니다.
- Window classification 만으로도 충분합니다.
  - 대부분의 단어가 named entity 가 아니라면 sequential 한 정보를 이용하여도 그 영향력이 작습니다.



https://nlp.stanford.edu/~socherr/pa4_ner.pdf

# Named Entity Recognition

- 최소한의 seed words 를 찾기 위해서 Word2Vec 이 이용될 수 있습니다.

| word2vec.similar('짜파게티') | word2vec.similar('신도림') |
|---|---|
| 샌드위치 | 부평 |
| 햄버거 | 합정 |
| 누룽지 | 신촌 |
| 소고기 | 선릉 |
| 토스트 | 강변 |
| 소세지 | 사당 |
| 떡복이 | 당산 |
| 부대찌개 | 잠실 |
| 김치찌개 | 신림 |
| 군만두 | 광화문 |
| 팥빙수 | 압구정 |
| 라볶이 | 산본 |
| 순두부찌개 | 인덕원 |
| 견과류 | 불광 |
| 스파게티 | 공덕 |

# Named Entity Recognition

- Word2Vec 의 CBOW 가 앞/뒤 단어를 이용하여 해당 단어를 예측하듯이, 앞/뒤의 단어를 features 로 이용하는 window classifier 를 만듭니다.

| word2vec.similar('짜파게티') |
|---|
| 샌드위치 |
| 햄버거 |
| 누룽지 |
| 소고기 |
| 토스트 |
| 소세지 |
| 떡복이 |
| 부대찌개 |
| 김치찌개 |
| 군만두 |
| 팥빙수 |
| 라볶이 |
| 순두부찌개 |
| 견과류 |
| 스파게티 |

| | |
|---|---|
| **Sentence** | [배고, 프다, 점심, 에, 샌드위치, 먹, 을래, ?] |
| **Window** | [배고, 프다, 점심, 에, 샌드위치, 먹, 을래, ?] |
| **Features** | X[-2:-1] = 프다 – 점심<br>X[-1] = 점심<br>X[-1] & X[1:2] = 프다 – 점심 – 샌드위치<br>… |
| **Label** | False |

# Named Entity Recognition

- Word2Vec 의 CBOW 가 앞/뒤 단어를 이용하여 해당 단어를 예측하듯이, 앞/뒤의 단어를 features 로 이용하는 window classifier 를 만듭니다.

| word2vec.similar('짜파게티') |
|---|
| **샌드위치** |
| 햄버거 |
| 누룽지 |
| 소고기 |
| 토스트 |
| 소세지 |
| 떡복이 |
| 부대찌개 |
| 김치찌개 |
| 군만두 |
| 팥빙수 |
| 라볶이 |
| 순두부찌개 |
| 견과류 |
| 스파게티 |

| | |
|---|---|
| **Sentence** | [배고, 프다, 점심, 에, 샌드위치, 먹, 을래, ?] |
| **Window** | [배고, 프다, 점심, 에, 샌드위치, 먹, 을래, ?] |
| **Features** | X[-2:-1] = 점심 – 에<br>X[-1] = 에<br>X[-1] & X[1:2] = 에 – 먹 – 을래<br>… |
| **Label** | True |

# Named Entity Recognition

- Word2Vec 으로 찾은 음식 관련 단어 seeds

'햄벅 빈대떡 팥죽 채소 인스턴트 고로케 홍어 크레페 치폴레 수제비 떡튀순 문어 콜라 삼김 짜왕 꽃게 복숭아 요구르트 쌈도 야식 한우 칼국수 해물찜 쌀밥 육개장 골뱅이 후식 아구찜 컵밥 음료수 고기 낙지 유자차 조개구이 비비큐 꼬치 케익 꼬기 청국장 갈비탕 닭발 피자 킹크랩 견과류 에스프레소 훠궈 샤브 엽떡 쫄면 아점 상추 소스 가래떡 석식 쿨피스 삼치 프링글스 카레 식혜 키위 맥주 석류 짜파게티 순대국 씨리얼 컵라면 초밥 껍데기 와플 비빔밥 치킨 음료 헛개수 뼈해장국 등갈비 갈비 두부 핫도그 라멘 잡채밥 파닭 순대 밀크티 호떡 치맥 미숫가루 떡볶이 소고기 햄버거 갈비찜 닭도 포카리 불고기 딤섬 타코야끼 김밥 탕수육 미역국 먹이 장어 갈매기살 청심환 삼겹살 광어 삼겹 수박 곱창 핫식스 풀떼기 팥빙수 마늘 쌈은 간식 주먹밥 김치찌개 닭강정 식초 월남쌈 국밥 브런치 블루베리 양꼬치 감자 홈런볼 호박죽 우유 쟁반짜장 스무디 빙수 쌈이랑 맥모닝 쌀국수 소주 과일 싸이버거 고등어 감튀 밀가루 소맥 컵누들 햇반 분식 닭갈비 떡갈비 족발 쉐이크 닭도리탕 반마리 바나나 보드카 고구마 분유 환타 쥐포 참외 뻥튀기 홍시 해장국 감자탕 살안찌 육포 소세지 양파 버블티 송편 깐풍기 파전 순두부 브리또 옥수수 츄러스 쥬스 치즈 짜장면 소금구이 쌈이 찜닭 백숙 곰국 콩국수 스시 라뮨 팝콘 닭백숙 돈까스 야채 오징어 참치 연어 커리 빅맥 새우 삼계탕 토마토 닭꼬치 닭가슴살 간장게장 칙힌 갈치 물회 냉면 꼼장어 오뎅탕 돈부리 생선구이 티라미수 파스타 청하 타코 군만두 떡국 베지밀 설렁탕 아이스크림 북어국 라볶이 게장 순하리 몽쉘 쭈꾸미 추어탕 해물탕 점심 디저트 체리 샤브샤브 누룽지 꼬막 짬뽕 홍삼 샌드위치 조개찜 박카스 우동 육회 와퍼 핫바 바베큐 부대찌개 과메기 과자 수제버거 쌍화탕 밀면 꽃게탕 사과주스 바닐라라떼 샐러드 오리 비빔면 비요뜨 탕슉 목살'

# Named Entity Recognition

- Window classification 을 통하여 추가적으로 찾은 '음식'들

'스파게티 보쌈 냉모밀 닭죽 라면 바나나우유 순두부찌개 짱깨 콘푸로스트 된장찌개 초콜릿 진통제 팟타이 소바 젤리 마카롱 알탕 잡채 밥 떡만두국 귤 맘마 간단히 빵도 사케 맛난거 한정식 닭한마리 해열제 짜파구리 닭계장 갈치조림 매화수 차돌박이 선식 군고구마 뿌링클 곶감 칸쵸 파인애플 곤드레밥 통닭 까르보나라 맛잇는거 전복죽 육회비빔밥 약 두루치기 허니버터칩 학식 밥만 뚝불 에그타르트 씨앗호떡 달달한거 딸기 오뎅 빕 전어 떡꼬치 나쵸 양념갈비 한라봉 떡뽀끼 술처 계란후라이 오꼬노미야끼 뭐좀 즉떡 밥비벼 항정살 옻닭 수면제 문어숙회 보신탕 간단하게 프레즐 약안 매콤한거 보리밥 동태탕 얻어 대게 오돌뼈 메밀소바 집밥 회 밥안 보충제 지코바 알밥 콩불 거봉 약도 양념게장 콘프레이크 아수쿠림 빠네 닭똥집 빵 고등어조림 콘푸라이트 호빵 물만 가츠동 딸바 마라샹궈 스팸 곰장어 징거버거 와구와구 빵좀 이슬만 매운갈비찜 비냉 선지해장국 꿔바로우 만둣국 부침개 와인 도시락 송어회 너구리 빠삐코 귤도 수육 파니니 죠떡 떡 내장탕 말랑카우 생선 스크류바 젤라또 고등어구이 유산균 타르트 밥좀 샤부샤부 라자냐 떡보끼 오믈렛 한약 치돈 꼬깔콘 전어회 큰맘 약좀 만듀 도지마롤 조개 모밀 쭈삼 리조또 전복 밥다 백반 항생제 사시미 양배추 칵테일 자몽주스 회도 술죽만 선지국 오트밀 사과 빵만 피쟈 뭐쫌 코다리찜 꿀떡 두통약 사탕 굴보쌈 포카칩 아점으로 만두 뭐랑 뼈다귀해장국 철분제 줏어 웨하스 하리보 도넛 식빵 삼치구이 보쌈정식 레모나 부추전 N그릇 동까쓰 맘스터치 국수 밥두안 두그릇 장조림 쌈밥 번데기 곰탕 동동주 봉구스 육사시미 태국음식 난술 사케동 국물있는거 비타민 퀘사디아 물냉 밥못 한입 버거킹 방토 기내식 술안 콩비지 건빵 영양제 생태탕 군밤 한식 수프 두조각 쿠앤크 설빙 뭐 광어회 아몬드 규동 데워서 N조각 양장피 홍합 콘칩 마늘빵 해산물 봉구스밥버거 겔포스 왕창 대하 조식 동그랑땡 야무지게 반계탕 어묵 고량주 급식 밥은 짬밥 등심 크런키 두끼 베라 납작만두 밥이라도 정로환 발포비타민 밥으로 겉절이 오뎅국 한알 파르페 칼제비 녹두전 생맥 케밥 소염제 커피 짜파 빈츠 두알 꽃등심 기름진거 저녁 복분자 삶은계란 메론우유 상하이 밥부터 죽도 똥집 매운거 포도 저녀 김떡순 안성탕면 우엉차 비벼 생생우동 깻잎 홍차 술만 밥이나 장어구이 풀만 취킨 솜사탕 카푸치노 맥스봉 도스마스 빵또아 데낄라 칠리새우 매생이국 스무디킹 콘프라이트 더위사냥 김찌 위스키 부드러운거 알리오올리오 버섯 닭다리 벤또 구두 칠면조 국물 빼빼로 푸딩 치킹 한조각 맛나게 밥도 대충 선지 짬짜면 우루사 아이스아메리카노 밥잘 기식 쵸콜렛 주전부리 찐빵 떠볶이 밋있게 한조각만 석갈비 밥뭐 몇조각 규카츠 삼걉살 골고루 대추 스낵랩 쌈싸 술더 데워 샌뒤치 아귀찜 맛릿게 망고 녹차아이스크림 우걱우걱 풀때기 깐쇼새우 점저 강정 양식 봉골레 깍두기 게보린 생강차 흰우유 소보루빵 마싯게 허겁지겁 술그만 병식 탕짜면 머핀 빱빠 콘소메 바게트 맞있게 매실 따신거 저녘은 카츠동 미음 죠리퐁 아슈쿠림 인도커리 달다구리밥또 밥을 대창 한치 주워 N인분 호박고구마 꿀빵 찌짐 적당히 삼교비 수미칩 훈제연어 세그릇 브로콜리 비타오백 쮸꾸미 삼결살 안주 유린기 호식이 홍합탕 맛닜게 호가든 멜라토닌 맥날 양껏 타코와사비 콘치즈 판모밀 웨지감자 한마리 캐슈넛 닭 빵쪼가리 육전 멸치 적게 청양고추 부대에서 주서 산채비빔밥 닭날개 게국지 황태해장국 도가니탕 약은 고추 떡도 오이 약이라도 스미노프 야금야금 훈제오리 야끼우동 부실하게 우유라도 돌솥비빔밥 뚝배기불고기 니트 전어구이 고구마말랭이 델리만쥬 마요네즈 화채 까까 거하게 삼겻살 피임약 탱크보이 사골국 백설기 갈매기 김피탕 도토리묵 토시살 샹그리아 요거트스무디 수면유도제 귀걸이 뜨뜻한거 오이소박이 양대창 산오징어 팔찌 고거 김볶 한끼 냉수 쫀드기 물티슈 날개 허브티 약만 넥타이 회만 생선까스 약과 짜짜로니 빵이라도 빵한조각 머점 맛잃게 원피스 랏츠버거 겹살이 간장계란밥 홍초 시루떡 새콤달콤 골뱅이무침 오니기리 쥬씨 예거 라떼 석화 오코노미야키 샴페인 밥해 많이 호두마루 노가리 왕꿈틀이 칙촉 틴트 구충제 쌀통닭 바지 정식 나시고랭 동태찌게 데자와 샵샵 낮술 닭도리 챙겨 빠에야 향초 허니버터 깨작깨작 조리퐁 멍게 목걸이 그거 돈코츠라멘 구워 무화과 고양이 사발면 시계 찌낀 일식 배추 푸아그라 몇그릇 죽이라도 도너츠 다꼬야끼 호로요이 베토디 스쿨푸드 회충약 뷔페 안까 밥바 피맥 탕짬면 달다구리한거 모히또 뽀글이 사료 깔루아 부추 동태전 누네띠네 맛잇게 샴푸 쌍쌍바 찌개 어묵탕 체리쥬빌레 세제 바퀴벌레 와퍼주니어 대강 N알 팝송 참이슬 냉우동 베스킨 맛동산 도라지 빵을 쿠사리 회를 복스럽게 또뽀끼 갈치구이 동파육 초계탕 불백 전투식량 교통비 육수 차슈동 명태찜 비누 기네스 반지 중식 모과차 지사제 굳게 포카리스웨트 텀블러 지영이 유진이 현미밥 욕 짜장 맥머핀 지원이 효정이 푸짐하게 고추바사삭 복어 공차 북경오리 구어 희진이 팬돌이 꿀물 가볍게 닭가슴 녹차프라푸치노 현정이 메로구이 흰거 약을 키드오 단거 소금 닭을 세조각 묵사발 배즙 앱솔 니꺼 오리주물럭 매실차 아포가토 혜진이 생식 츄파춥스 펑리수 김태희 은영이 황도 치맛살 소연이 밋나게 미역국이랑 수연이 맛읶게 한개 마니 모니터 수정이 맥도날드 쫀디기 나눠 악세사리 거미 누나 휴지 설농탕 로아커 현경이 스킨쉽 파우치 술도 안빼 산삼 팔보채 두부조림 체크카드 메밀전병 아보카도 까자 카누 개불 민선이 버터 본죽 푸지게 떡볶이 생수 버터와플 정은이 예은이 술퍼 슌대국 승연이 선영이 맛탕 뿌셔뿌셔 페리카나 허벅지 맘터 부라더소다 마가렛트 쫄비빔면 반공기만 세진이 외모 끊임없이 수진이 빵빠레 알로에 칠리 눈칫밥 면만 현진이 엉덩이 테라플루 뺏어 곱배기 파파존스 쏘야 하겐다즈 탕볶밥 예거밤 레고 향수 지은이 진영이 피짜 옥동자 N원짜리 회무침 다정이 황태구이 승현이 젓갈 은정이 스릴러 뮤비 혜민이 피카츄 작년에 주연이 빕스 달걀 초꼬렛 서연이 울아빠 액자 혜정이 뼈다귀 N끼 경은이 낙곱새 율무차 스윙칩 송혜교 한스푼 샤오롱바오 영준이 ...'

# Named Entity Recognition

- Sequential labeling, Window classifications 을 이용하는 방법은 Rule-based NER 과도 비슷합니다.
  - Weights 를 이용한다는 점이 다르며, 모델을 통하여 rules 를 찾는 것입니다.

Determining where an organization is located

- Template : [org] [loc] (division, branch, headquarters)
- Example : KFOR Kosovo headquarters



https://web.stanford.edu/class/cs124/lec/Information_Extraction_and_Named_Entity_Recognition.pdf

In some simple domains, naive technique is remarkably effective.

But do think about when it would and wouldn't work!



https://web.stanford.edu/class/cs124/lec/Information_Extraction_and_Named_Entity_Recognition.pdf