# 深度學習基礎概論

0408

# 目錄

- Named Entity Recognition(命名實體識別)
- Code
- 中文標註
- https://kknews.cc/zh-tw/code/pvgga9e.html

# Named Entity Recognition(命名實體識別 NER)

- 識別文字中具有特定意義的詞(實體)，主要有人名、地名、轉有名詞等

- 希望將需要識別出來的詞在一段文字中標註出來

- E.g : 地名識別

  台北市是一個交通便利的城市

我們希望 model 把台北市這三個字標記出來

# 命名實體識別 NER

- 命名實體一般包含 3 大類

1. 實體類
2. 時間類
3. 數字類

和 7 小類：人名、地名、組織機構名、時間、日期、貨幣、百分比

# IOB 標註法

- B : beginning，實體的開頭
- I : inside，實體的中間或結尾
- O : outside，不屬於實體

- 台北市是一個交通便利的城市
- B I I O O O O O O O O O O

# BIOES 標註法

- 在 IOB 方法上擴展一個更複雜，但更完備的標註方法
- B : beginning，實體的開頭
- I : inside，實體的中間
- O : outside，不屬於實體
- E : end ，實體的結尾
- S : single ，代表單一個字本身是一個實體

- 台北市是一個交通便利的城市
- B I E O O O O O O O O O O

# Datasets : NCBI

- NCBI 疾病語料庫 是包含 793 個 PubMed 摘要
- Identification of APC2, a homologue of the adenomatous polyposis coli tumour suppressor . The adenomatous polyposis coli ( APC ) tumour-suppressor protein controls the Wnt signalling pathway by forming a complex with glycogen synthase kinase 3beta ( GSK-3beta ) , axin / conductin and betacatenin . Complex formation induces the rapid degradation of betacatenin . In colon carcinoma cells , loss of APC leads to the accumulation of betacatenin in the nucleus , where it binds to and activates the Tcf-4 transcription factor ( reviewed in [ 1 ] [ 2 ] ) . Here , we report the identification and genomic structure of APC homologues . Mammalian APC2 , which closely resembles APC in overall domain structure , was functionally analyzed and shown to contain two SAMP domains , both of which are required for binding to conductin . Like APC , APC2 regulates the formation of active betacatenin-Tcf complexes , as demonstrated using transient transcriptional activation assays in APC - / - colon carcinoma cells . Human APC2 maps to chromosome 19p13 . 3 . APC and APC2 may therefore have comparable functions in development and cancer .

# Datasets : NCBI

- 目的 : 在醫療文章摘要中尋找疾病名稱的 NER
- Training sets : 683
- Validation sets : 100
- Testing sets : 10
- Identification of APC2 , a homologue of the adenomatous polyposis coli tumour suppressor.

O O O O O O O O O B I I I O O

Admatmatous : 腺瘤性，被標標記 B (開頭)

polyposis coli tumour : 結腸瘜肉腫瘤，被標記成 I (內部)

# Code