



# Preliminary Report of III&CYUT for NTCIR-11 MedNLP-2

Team ID: III&CYUT

Liang-Pu Chen

[eit@iii.org.tw](mailto:eit@iii.org.tw)



# Processing Steps

---

- Step 1: Format Processing
- Step 2: CRF Training & Testing
- Step 3: ICD10 Classifier (normalize)



# Features for CRF

---

- Term
  - Unigram
  - Stick continuous numerical
- Part-of-speech
  - MeCab
  - Kuromoji
- Dictionary
  - MEDIS Byomei Master
  - MEDIS Shojo Shoken Master



# CRF Data Format

Term	POS1	POS1	POS2	POS2	Dictionary	IOB2
2028	名詞	数	名詞	数	Begin	B-t
年	名詞	接尾	名詞	接尾	Y-name	l-t
11	名詞	数	名詞	数	N	l-t
月	名詞	副詞可能	名詞	副詞可能	Y-name	l-t
頃	名詞	副詞可能	名詞	副詞可能	N	l-t
よ	助詞	格助詞	助詞	格助詞	N	l-t
り	助詞	格助詞	助詞	格助詞	Y-name	l-t
、	記号	読点	記号	読点	N	O
胸	名詞	一般	名詞	一般	N	B-c
痛	名詞	一般	名詞	一般	Y-name	l-c
出	名詞	サ変接続	名詞	サ変接続	N	O
現	名詞	サ変接続	名詞	サ変接続	N	O



# CRF Data Format

Term	POS1	POS1	POS2	POS2	Dictionary	IOB2
2028	名詞	数	名詞	数	Begin	B-t
年	名詞	接尾	名詞	接尾	Y-name	l-t
11	名詞	数	名詞	数	N	l-t
月	名詞	副詞可能	名詞	副詞可能	Y-name	l-t
頃	名詞	副詞可能	名詞	副詞可能	N	l-t
よ						l-t
り						l-t
、	記号	読点	記号	読点	N	O
胸	名詞	一般	名詞	一般	N	B-c
痛	名詞	一般	名詞	一般	Y-name	l-c
出	名詞	サ変接続	名詞	サ変接続	N	O
現	名詞	サ変接続	名詞	サ変接続	N	O

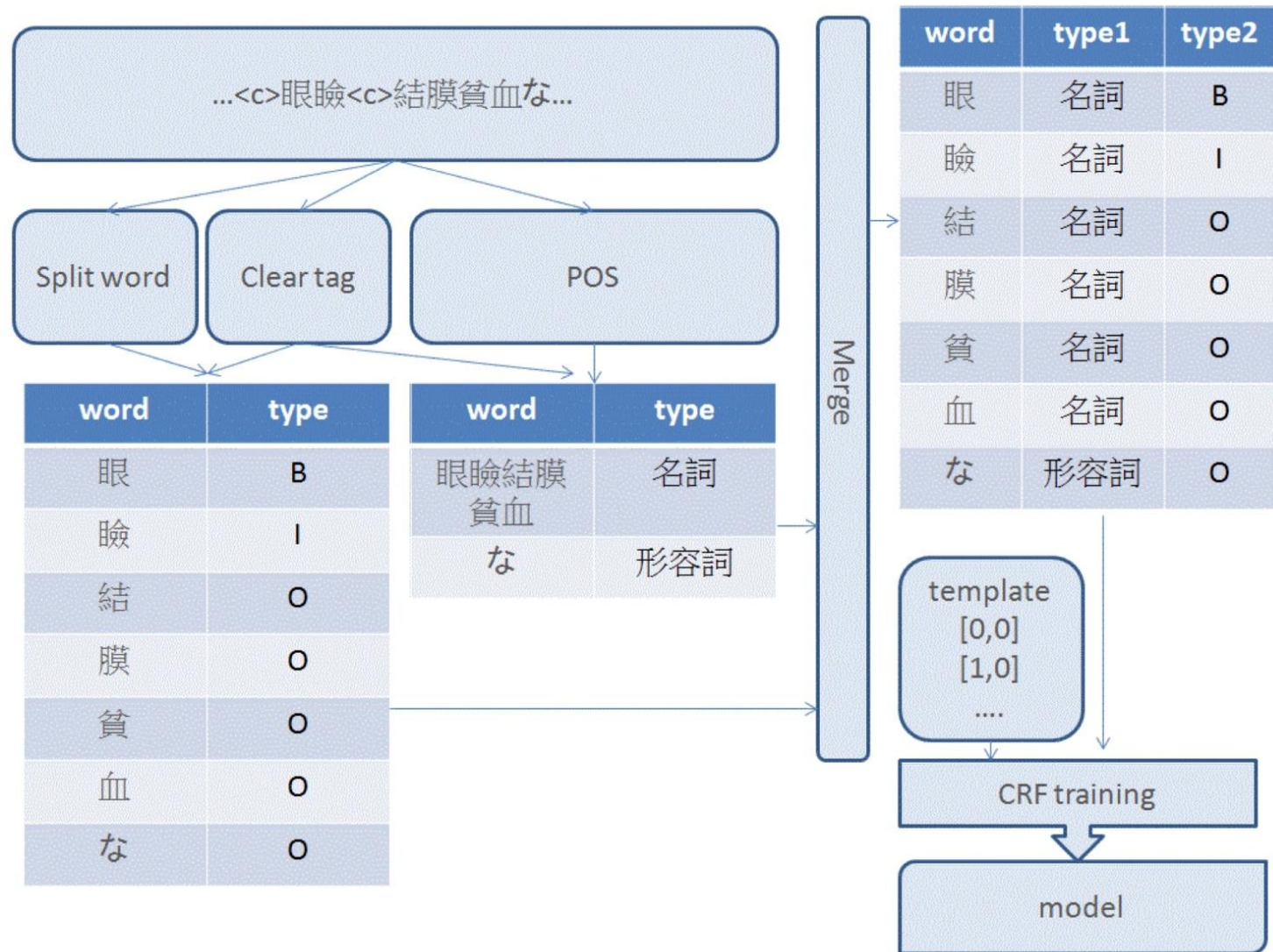
Combine this term and previous one then check it existed in dictionary or not

胸痛

Y-name

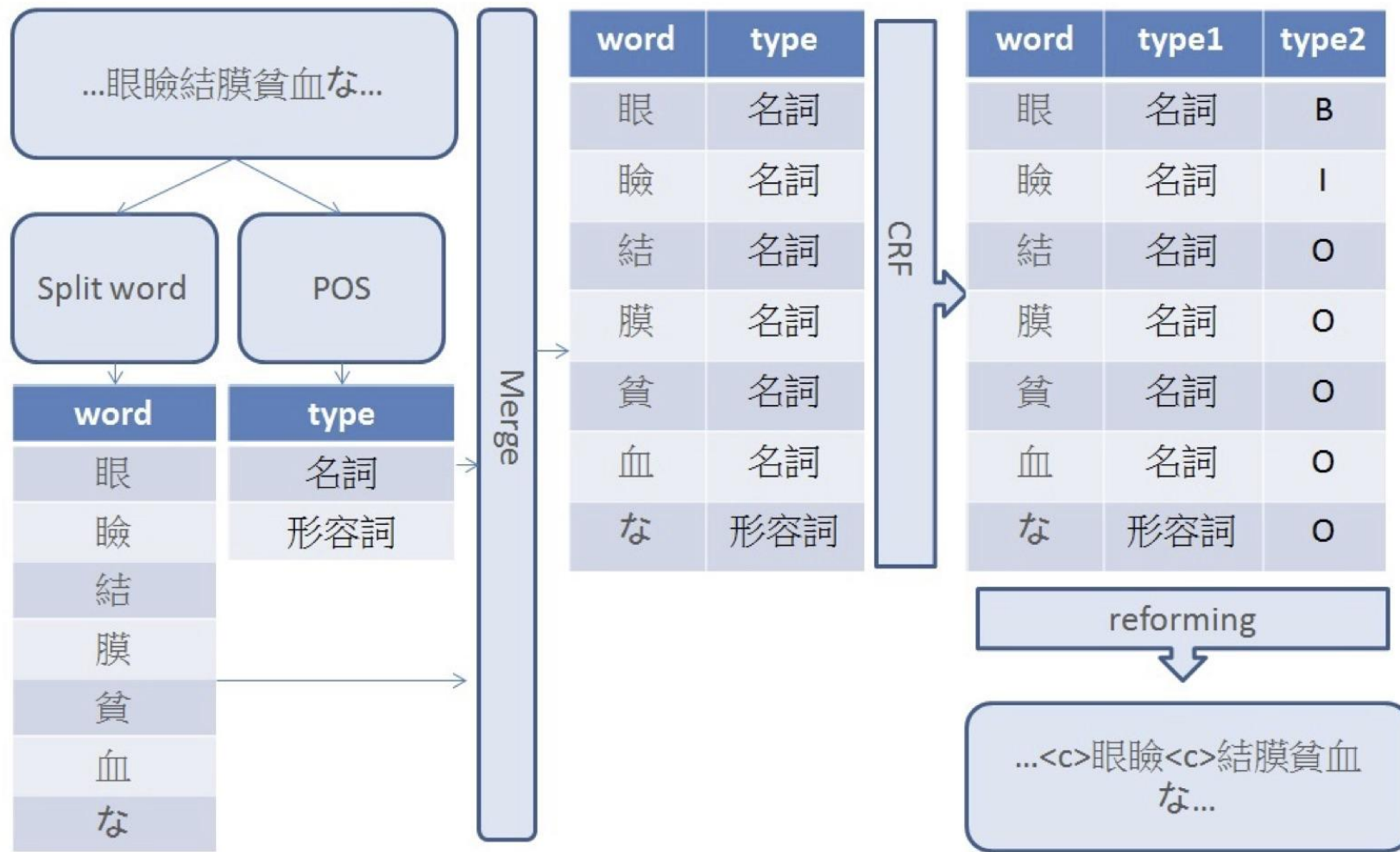


# CRF Training





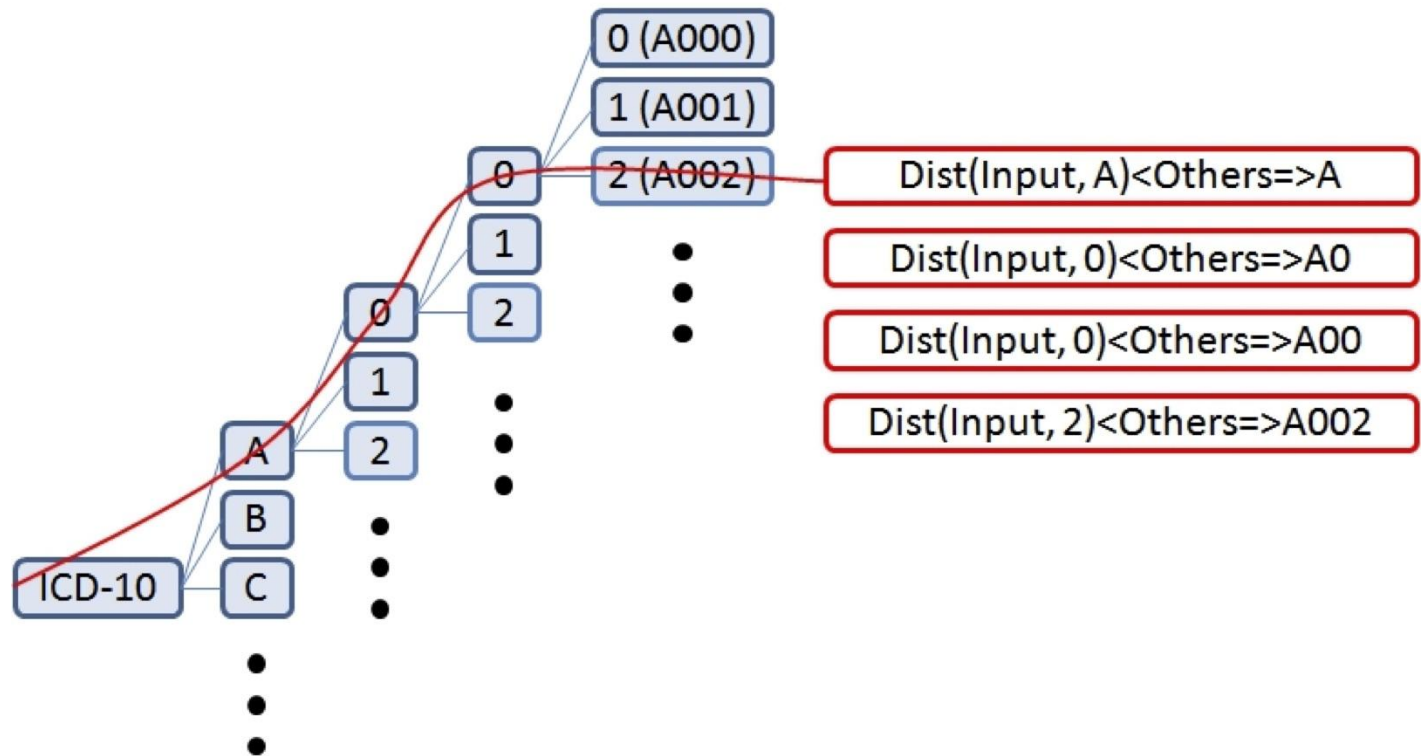
# CRF Testing





# ICD10 Classifier

- ICD10 has hierarchical categories
- Build a tree structure to approach







# Official Result

Tag	Precision	Recall	F1-score
Positive	62.68	48.62	54.76
Family	37.21	76.19	50.00
Negation	51.86	51.57	51.71
Suspicion	6.56	7.27	6.90



**Thank you for your attention**  
**Poster MedNLP09**