Clinical Text Normalization

대소문자 통일

CHF → chf

Diabetes → diabetes

약어 확장

BP → blood pressure

Dx → diagnosis

철자 교정

diabetis → diabetes

hypertention → hypertension

날짜/시간 표준화

3/15/23 → 2023-03-15

5pm → 17:00

단위 변환

98.6°F → 37°C

150 lbs → 68 kg

특수문자 처리

HTN-1 → HTN 1

pt./patient → patient

---

1 토큰화
Tokenization

→

2 정규화
Normalization

→

3 표준화
Standardization

→

4 검증
Validation

→

정규화된 텍스트