

---

---

# Chemical Name Entity Recognition

---

---

The slide features decorative horizontal lines: a thick teal line at the top, a thin teal line below it, and another thick teal line at the bottom. Two short, thick brown dashes are positioned horizontally, one on the left and one on the right, centered vertically between the middle thin teal line and the bottom thick teal line.

# Preprocessing :

1. Sentence and word Tokenization
2. Lemmatization
3. Stop Word Removal
4. Generation of N-grams

# Features Used:

- 1.) Lemma and POS tag of the word.
- 2.) Digit Count
- 3.) Upper and lower case counts
- 4.) Punctuation Characters
- 5.) Case Pattern Features
- 6.) Character N-grams

# Model Used:

CRF++ software was utilized for training CRF models.

We trained the model with 90% data and tested the model with the remaining data.

We got an accuracy of 66.67%.