

# Online Learning for Solving Data Availability Problem in Natural Language Processing

Buddhi Wathsala  
*b.wathsala.bw@gmail.com*

Prof. Albus Dumbledore  
*ad@ucsc.cmb.ac.lk*

Prof. Severus Snape  
*ss@ucsc.cmb.ac.lk*

*University of Colombo School of Computing*

## Abstract

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. [1], [2], [3]

## 1 Intro

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

## 2 Related Works

### 2.1 AA

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. [4], [5]

## References

- [1] D. Stewart. (Oct. 25, 2018). Thomas kuhn, [Online]. Available: [www.famousscientists.org/thomas-kuhn/](http://www.famousscientists.org/thomas-kuhn/).
- [2] T. E. of Encyclopaedia Britannica. (Oct. 25, 2018). Thomas s. kuhn american philosopher and historian, [Online]. Available: <https://www.britannica.com/biography/Thomas-S-Kuhn>.
- [3] T. Nickles, “Thomas kuhn”, in *British journal for the philosophy of science*, vol. 53, Oct. 2002, pp. 1–312, ISBN: 0521796482.

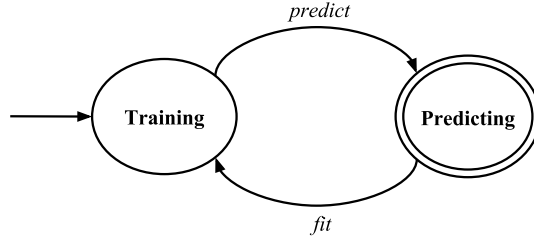


Figure 1: Research Design

Mini Batch	Precision	Recall	F-Measure
1	20.2481	30.8333	24.444
2	31.9227	44.5519	37.1945
3	30.5363	49.4973	37.7708
4	29.5513	52.2244	37.7446

Table 1: Precision, Recall, and F-Measure value variation for 4 mini batches in the CRF model

---

**Algorithm 1** Data Preprocessing

---

```

1: procedure PREPROCESS(file)
2:   text  $\leftarrow$  read(file)
3:   vocabulary  $\leftarrow$  CREATE_VOCABULARY(text)
4:   data  $\leftarrow$  []
5:   tokens  $\leftarrow$  []
6:   for sentence in text do
7:     s_words  $\leftarrow$  VOCABULARY_INDEX(sentence)
8:     s_tokens  $\leftarrow$  TOKENS(sentence)
9:     data.add(s_words)
10:    tokens.add(s_tokens)
11:  end for
12:  return (data, tokens)
13: end procedure

```

---

- [4] A. Bird. (Oct. 25, 2018). Stanford encyclopedia of philosophy - thomas kuhn, [Online]. Available: <https://plato.stanford.edu/entries/thomas-kuhn/#1>.
- [5] T. F. Orman, "Paradigm as a central concept in thomas kuhn's thought", vol. 6, International Journal of Humanities and Social Science, Oct. 2016.