

# Using Nature Language Processing to Improve Optical Character Recognition

**Lining Xu**

Tsinghua University

**Yongxu Wu**

Beijing Information Science and Technology University

**Abstract:** OCR (Optical Character Recognition) has developed over 100 years. However, if the document or picture is stained, it could not work well. Considering that words in text can be connected by logical relationship, with the help of the idea that reducing the size of word stock which references from license plate recognition, this paper established N-GRAM model, used the results of Google search engine to improve its text sparsity. The use of residual features of the original stained characters can improve the recognition rate and accuracy with the help of a smaller size of the word stock successfully.