

## Unsupervised Learning with Self-Organizing Maps for Named Entity Recognition in the CONLL2003 Dataset

**Authors :** Assel Jaxylykova, Alexnder Pak

**Abstract :** This study utilized a Self-Organizing Map (SOM) for unsupervised learning on the CONLL-2003 dataset for Named Entity Recognition (NER). The process involved encoding words into 300-dimensional vectors using FastText. These vectors were input into a SOM grid, where training adjusted node weights to minimize distances. The SOM provided a topological representation for identifying and clustering named entities, demonstrating its efficacy without labeled examples. Results showed an F1-measure of 0.86, highlighting SOM's viability. Although some methods achieve higher F1 measures, SOM eliminates the need for labeled data, offering a scalable and efficient alternative. The SOM's ability to uncover hidden patterns provides insights that could enhance existing supervised methods. Further investigation into potential limitations and optimization strategies is suggested to maximize benefits.

**Keywords :** named entity recognition natural, language processing, self-organizing map, CONLL-2003, semantics

**Conference Title :** ICMLC 2024 : International Conference on Machine Learning and Cybernetics

**Conference Location :** Bangkok, Thailand

**Conference Dates :** August 15-16, 2024