

Решения задачи распознавания именованных сущностей совершенствуются уже несколько десятилетий. На протяжении этого периода происходит постепенный переход от моделей, основанных на скрупулёзно конструируемых «руками» правилах, к самостоятельно обучающимся системам, конструируемым независимо от языка. Одними из первых самообучающихся моделей были модели основанные на вероятностных моделях таких как HMM и CRF. С теми или иными модификациями в начале 2000-х эти методы пытались применять в различных областях, например, в медицине, где, впрочем, даже наиболее успешные работы показывали достаточно низкие результаты около 0.7 F1 [4]. Тем не менее, как оказалась, даже с помощью чистого CRF при использовании оптимизационных алгоритмов, учитывающих удалённый контекст, можно получить неплохие результаты [2]. Прогресс в технологии обучения искусственных нейронных сетей позволил успешно применить нейросетевые архитектуры к задаче NER, что существенно улучшило качество распознавания именованных сущностей современными алгоритмами. При этом стоит отметить, что независимо от используемого подхода, лучший результат дают модели, адаптированные для входных данных, особенно при высокой специфичности обрабатываемых текстов, например, в твитах [3]. Одной из сильных сторон нейросетевого подхода является возможность посимвольной обработки слов [5] для получения признаков, которые могут использоваться другими алгоритмами - элементами сложной модели. Так, самые эффективные современные модели используют комбинации нейросетевых и вероятностных моделей [1]

Список литературы

- [1] L. T. Anh, M. Y. Arkhipov, and M. S. Burtsev. Application of a hybrid bi-lstm-crf model to the task of russian named entity recognition, 2017.
- [2] Flavio Massimiliano Cecchini and Elisabetta Fersini. Named entity recognition using conditional random fields with non-local relational constraints, 2013.
- [3] Alan Ritter, Sam Clark, Oren Etzioni, et al. Named entity recognition in tweets: an experimental study. In *Proceedings of the conference on empirical methods in natural language processing*, pages 1524–1534. Association for Computational Linguistics, 2011.
- [4] Burr Settles. Biomedical named entity recognition using conditional random fields and rich feature sets. In *Proceedings of the International Joint Workshop on Natural Language Processing in Biomedicine and Its Applications*, JNLPBA '04, pages 104–107, Stroudsburg, PA, USA, 2004. Association for Computational Linguistics.
- [5] Emily Sheng and Prem Natarajan. A byte-sized approach to named entity recognition, 2018.