

MPHTI 27.41.19

DOI: <https://doi.org/10.26577/JMMCS.2020.v107.i3.06>**Ж.М. Кожирбаев*** , **Ж.А. Есенбаев** Частное учреждение "National Laboratory Astana г. Нур-Султан, Казахстан
*e-mail: zhanibekkm@gmail.com

РАСПОЗНАВАНИЕ ИМЕНОВАННЫХ ОБЪЕКТОВ ДЛЯ КАЗАХСКОГО ЯЗЫКА

Распознавание именованных объектов (NER) считается одной из важных задач обработки естественного языка (NLP). Это способ распознавания объектов реального мира, таких как географическое положение, имя человека, организация и т. д., которые встречаются в предложении. Существует несколько подходов, основанных на созданных вручную правилах грамматики и статистических моделях, таких как машинное обучение и гибридные методы, для решения задачи распознавания именованных объектов. Цель данной работы состоит в том, чтобы поэкспериментировать с методами, основанными на статистическом подходе и на машинном обучении, и проверить как они справляются с агглютинативным казахским языком. В данной работе представлено распознавание именованных объектов на основе подхода машинного обучения, называемого условным случайным полем (CRF), как статистический метод. А также мы используем гибридный подход, сочетающий двунаправленную модель нейронной сети с долгой краткосрочной памятью (LSTM) и модель CRF. Это современный подход к распознаванию именованных объектов. Модель с перекрестным проверенным рандомизированным поиском показывает оценку f1 с 0,95. Гибридная модель LSTM-CRF показывает оценку f1 с 0,88. Результаты выглядят довольно хорошо, и это не требует каких-либо особенностей разработки по сравнению с моделью CRF. Для проведения экспериментов, был создан корпус (kazNER) для задачи NER с такими метками, как имя человека, местоположение, организация и другие. Корпус состоит из 29629 предложений, которые содержат хотя бы одно собственное существительное, содержащее только теги части речи.

Ключевые слова: распознавание именованных объектов, модель с условным случайным полем, нейронная сеть с долгой краткосрочной памятью, векторное представление слов

Ж.М. Кожирбаев*, Ж.А. Есенбаев

"National Laboratory Astana" жеке мекемесі, Нұр-Сұлтан қ., Қазақстан

*e-mail: zhanibekkm@gmail.com

Қазақ тіліндегі жалқы есімдерді тану

Жалқы есімдерді тану (NER) табиғи тілді өңдеудің (NLP) маңызды тапсырмаларының бірі болып саналады. Бұл сөйлемде кездесетін географиялық атауларды, адамның есімдерін, ұйымның аттарын және т.б. сияқты нақты жалқы есімдерді танудың тәсілі. Қолмен жасалған грамматикалық ережелер мен статистикалық модельдерге негізделген бірнеше тәсілдер бар, мысалы, жалқы есімдерді тану мәселесін шешуде машиналық үйрету және гибридік әдістер. Бұл жұмыстың мақсаты - статистикалық тәсіл мен машиналық оқытуға негізделген әдістермен тәжірибе жасау және олардың агглютинативті қазақ тілімен қалай жұмыс жасайтындығын тексеру. Бұл жұмыста шартты кездейсоқ өріске (CRF) негізделген

статистикалық тәсілмен қазақ тілінің жалқы есімдерін таңу ұсынылған. Біз сондай-ақ екі бағытты ұзақ қысқа мерзімді жады (LSTM) негізіндегі нейрондық желі және CRF моделімен біріктіретін гибридік әдісті қолданамыз. Бұл жалқы нысандарды таңудың қазіргі заманғы тәсілі. Кросс-расталған рандомизацияланған іздеу моделі 0,95 f1 көрсеткішінде тоқтаса, LSTM-CRF гибридік моделі 0,88 f1 көрсетеді. Нәтижелер өте жақсы көрінеді және CRF моделімен салыстырғанда ешқандай дизайн ерекшеліктерін қажет етпейді. Тәжірибелер үшін NER тапсырмасына адамның есімі, географиялық атаулар, ұйым атаулары және басқалар сияқты белгілері бар корпус (kazNER) құрылды. Корпус 29629 сөйлемнен тұрады, олардың әрқайсысында кем дегенде бір жалқы есім бар.

Түйін сөздер: жалқы есімдерді таңу; шартты кездейсоқ өріс моделі; ұзақ қысқа мерзімді жады; сөздердің векторлық көрінісі

Z.M. Kozhirbayev*, Z.A. Yessenbayev

Private Institution "National Laboratory Astana Nur-Sultan, Kazakhstan

*e-mail: zhanibekkm@gmail.com

Named entity recognition for the Kazakh language

Named Entity Recognition (NER) is considered one of the important tasks of natural language processing (NLP). This is a way of recognizing real world objects, such as geographical location, person's name, organization, etc., that are found in a sentence. There are several approaches based on manually created grammar rules and statistical models, such as machine learning and hybrid methods, to solve the problem of recognizing named entities. The aim of this work is to experiment with methods based on statistical approach and machine learning, and to check how they deal with agglutinative Kazakh language. This paper presents the recognition of named objects based on a machine learning approach called conditional random field (CRF) as a statistical method. We also use a hybrid approach combining a bidirectional neural network model with long-term short-term memory (LSTM) and a CRF model. This is a modern approach to the recognition of named objects. The cross-validated randomized search model shows an f1 score of 0.95. The hybrid LSTM-CRF model shows an f1 score of 0.88. The results look pretty good and it doesn't require any design specifics compared to the CRF model. For the experiments, a corpus (kazNER) was created for the NER task with such marks as a person's name, location, organization and others. The corpus consists of 29,629 sentences that contain at least one proper noun containing only part of speech tags.

Key words: named entity recognition; conditional random field; long-term short-term memory; word embeddings

1 Введение

За последнее десятилетие огромный прогресс был достигнут в области обработки естественного языка с появлением подходов машинного обучения и доступности вычислительных ресурсов для хранения и обработки огромного количества данных. Если большинство неструктурированных текстовых данных, доступных не только из традиционных средств массовой информации, но и из социальных сетей, можно структурировать, это дал бы возможность получить богатые знания из собранных данных. Извлечение именованных объектов составляет основную задачу для предоставления важной информации из полуструктурированных и неструктурированных текстовых источников. В этой работе будет представлено одно из известных решений задачи распознавания именованных объектов для казахского языка с применением условных случайных полей (CRF). А также мы использовали гибридный подход, сочетающий двунаправленную модель LSTM и модель CRF. Гибридный подход применялся на векторном представлении слов. Для проведения экспериментов был построен корпус с местонахождением,

организацией, именами и другие. Такая задача обусловлена целями проекта по разработке инструментов обработки текста на казахском языке, частью которого является настоящая работа. Нет сомнений в том, что в подходе нет новизны, но это не было целью. Цель состоит в том, чтобы поэкспериментировать с методами, основанными на статистическом подходе и на машинном обучении, и проверить как они справляются с агглютинативным казахским языком.

2 Обзор литературы

История задачи NER начинается с Sixth Conference on Message Understanding (MUC-6) в 1996 году [9], где задачи были сосредоточены на извлечении информации. В процессе постановки целей это выглядело как отдельная задача при извлечении объектов из документов. Чтобы определить объект, был введен термин «именованная сущность», и задача была названа как распознавание именованной сущности. Предыдущее исследование по извлечению информации из неструктурированных текстов проводилось с целью определения значимости «единиц информации», таких как имена людей, организаций, местоположений и числовые выражения, такие как время, дата, деньги и процентные выражения. Существует большое количество исследований, проведенных по NER для многих других языков. Для тщательного обзора работ по NER читателю рекомендуется обратиться к недавнему обзору [10]. Этот раздел ограничен кратким обзором исследований по NER для казахского языка [7, 8]. В предыдущей работе авторы утверждают, что их модель на основе CRF и особенности, полученные из результатов подхода морфологического анализа, значительно улучшают производительность системы с 69,91

3 Методология

В этом разделе представлен краткий обзор методов, которые были применены к созданному корпусу в этой задаче. CRF и LSTM были выбраны в качестве основного подхода для определения именованных объектов, представленных в предложении. Тем не менее, другие методы машинного обучения также используются для сравнения результатов и демонстрации влияния выбора признаков.

3.1 Случайный лес (Random Forest)

Поскольку задача NER рассматривается как простая задача классификации, древовидная модель Random Forest (RF) будет представлена с использованием простой карты объектов. Было доказано, что простые древовидные модели демонстрируют значительную производительность в задачах классификации. RF-классификатор, одна из самых точных древовидных моделей, может выучить основные правила, по которым помечаются термины. Выбор правильных признаков играет важную роль в производительности системы [1].

3.2 Наивный байесовский классификатор для полиномиальных моделей

Наивный байесовский (NB) подход всесторонне применяется к задачам NLP. Этот метод основан на принципе максимальной апостериорной вероятности. Для классифициро-

ванного объекта вычисляются функции правдоподобия каждого из названных классов сущностей, и из них вычисляются апостериорные вероятности названных классов сущностей. Объект принадлежит к названному классу, для которого апостериорная вероятность максимальна [5]. Есть две известные модели: многозначные модели и многовариантные модели Бернулли. Чтобы упорядочить соответствующий именованный класс сущностей n^* для нового термина w , он вычисляет:

$$p(c_i|w_i) = \frac{p(c_i)p(w_i|c_i)}{p(w_i)} \quad (1)$$

3.3 Условное случайное поле (CRF)

Условные случайные поля, являющиеся дискриминационной вероятностной моделью, чаще всего используются для решения проблем мечения и сегментации последовательностей [2]. Поскольку CRF является контролируемым алгоритмом машинного обучения, для его обучения требуется обучающая выборка достаточного размера. CRF может учитывать контекст; например, CRF с линейной цепью может предсказывать последовательности меток для последовательностей входных данных, в то время как дискретный классификатор предсказывает метки только для одной выборки. Приведенная ниже формула предназначена для CRF, где y - выходная переменная, а X - входная последовательность:

$$p(y|X, \lambda) = \frac{1}{Z(X)} \exp \sum_{i=1}^n \sum_j \lambda_j f_j(X, i, y_{i-1}, y_i) \quad (2)$$

Последовательность выходных выборок моделируется как нормализованное произведение функции функции.

3.4 Нейронная сеть с долгой краткосрочной памятью (LSTM)

Еще одна важная стратегия построения высокопроизводительного метода глубокого обучения - это понимание того, какой тип нейронной сети лучше всего подходит для решения проблемы NER, учитывая, что текст представляет собой последовательный формат данных. Но не любой тип LSTM справляется с этой задачей, так как использование стандартного LSTM для прогнозирования будет учитывать только «прошлую» информацию в последовательности текста. Нам нужно использовать двунаправленные LSTM для NER, поскольку контекст охватывает последовательные и будущие метки в последовательности. Двунаправленный LSTM представляет собой комбинацию двух LSTM: один движется вперед «справа налево», а другой - назад «слева направо» [11].

3.5 Гибридный подход (LSTM-CRF)

Дана входная последовательность $x = (x_1, \dots, x_m)$, то есть слова предложения и последовательность состояний вывода $s = (s_1, \dots, s_m)$, то есть теги именованного объекта. В

условных случайных полях мы смоделировали условную вероятность того, что выходная последовательность состояний дает входную последовательность:

$$p(s_1, \dots, s_m | x_1, \dots, x_m) \quad (3)$$

Мы сделали это путем определения карты объектов, которая отображает всю входную последовательность x в паре с полной последовательностью состояний s в некоторый вектор пространственных объектов d -измерения:

$$\Phi(x_1, \dots, x_m, s_1, \dots, s_m) \subset \mathbb{R}^d \quad (4)$$

Тогда мы можем моделировать вероятность в виде лог-линейной модели с вектором параметров $w \subset \mathbb{R}^d$ как:

$$p(s|x; w) = \frac{\exp(w \cdot \Phi(x, s))}{\sum_{s'} \exp(w \cdot \Phi(x, s'))}, \quad (5)$$

где, s' охватывает все возможные выходные последовательности. Мы можем рассматривать выражение $w \cdot \Phi(x, s) = score_{crf}(x, s)$ как оценку того, насколько хорошо последовательность состояний соответствует данной входной последовательности. Идея состоит в том, чтобы заменить функцию линейной оценки нелинейной нейронной сетью. Мы определяем *score* как:

$$score_{lstm-crf}(x, s) = \sum_{i=0}^n W_{s_{i-1}, s_i} \cdot LSTM(x)_i + b_{s_{i-1}, s_i}, \quad (6)$$

где W_{s_{i-1}, s_i} и b - весовой вектор и смещение, соответствующие переходу от s_{i-1} к s_i , соответственно. Функции оценки также называются textit потенциальные функции. После построения этой функции оценки мы можем оптимизировать условную вероятность $p(s|x; W, b)$, как в обычном CRF, и распространять ее обратно через сеть [12].

4 Источник данных

Набор данных собран из корпуса kazdet: NLA-NU Казахский банк деревьев зависимости [3], который аннотирован для леммы, части речи, морфологии и отношений зависимости в соответствии с Universal Dependency 2 и хранится в формате UD-native CoNLL-U format. По состоянию на декабрь 2018 года банк содержит 61 тыс. предложений и 934,7 тыс. токенов. Это довольно большой корпус с множеством аннотаций. Однако для задачи распознавания именованных сущностей нет аннотаций. Из kazdet корпуса было извлечено 29629 предложений, которые содержат хотя бы одно собственное существительное, содержащее только теги части речи. После этого был создан корпус (kazNER) для задачи NER. Поскольку теги IOB стали стандартным способом представления структур фрагментов в файлах, корпус kazNER будет в этом формате. Формат тегов IOB содержит теги вида:

- В - *TAG_TYPE* – для слова в начальном отрезке;
- I - *TAG_TYPE* – для слов внутри отрезка;
- O – вне любого отрезка.

Теги IOB далее подразделяются на следующие классы:

- LOC = Местоположение объекта;
- ORG = Организация;
- PER = Имя человека;
- OTH = любая другая именованная сущность, например имя питомца, название книги и т. д.

На рисунке 1 показано распределение слов по тегам с тегом O (вне любого фрагмента) и без него.

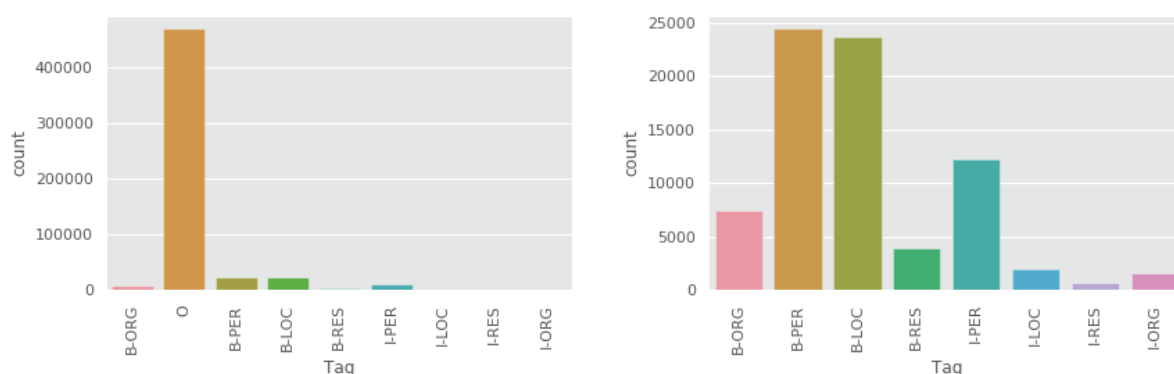


Рисунок 1 – Распределение слов по тегам с (слева) и без (справа) тега O

5 Эксперименты и обсуждения

Для того чтобы использовать методы, упомянутые в разделе “Методология”, необходимо предоставить набор признаков для правильного распознавания тегов. Поскольку классификаторы RF и NB не учитывают контекст, примерный набор признаков, такой как прописные буквы, тип слова (заголовок, строка, цифра), длина слова, символы алфавита, будет представлен для построения этих моделей. Однако метод CRF учитывает контекст. Sklearn-crfsuite [6] позволяет извлекать особенности слова в виде словаря, готового для использования с моделью:

- Текущие слова;
- Предыдущие слова;

- Следующие слова;
- Текущие POS-теги;
- Предыдущий и следующий POS-теги.

Группа Stanford NLP применила эти признаки при использовании CRF для задачи NER [4]. Признаки, такие как прописная буква, нижняя буква, цифра были извлечены для текущего, предыдущего и следующего слова. В дополнение к рекомендованным, в набор были добавлены признаки, такие как *istitle*, *iscamelcase*, *isabbv*, *has_hyphen*. Модель CRF использует алгоритм LGBFS (градиентный спуск с использованием метода L-BFGS) с упорядоченной упругой сеткой ($C1 + C2$). Значения упругой сетки регуляризации можно настроить, чтобы проверить их влияние на производительность. Сначала начальные значения были выбраны как $C1 = 0,1$ и $C2 = 0,1$ для модели CRF (CRF1). Затем параметры настраивались как $C1 = 10$ и $C2 = 0.1$ (CRF2). В третьем эксперименте (CRF3) использовался перекрестный рандомизированный поиск (Randomized CV Search), который представляет собой исчерпывающий поиск по сетке всех комбинаций параметров.

В гибридном подходе, сочетающий двунаправленную модель LSTM и модель CRF, мы сопоставляем предложения с последовательностью чисел, а затем дополняем последовательность. Обратите внимание, что мы увеличили индекс слов на единицу, чтобы использовать ноль в качестве значения заполнения. Это сделано потому, что мы хотим использовать параметр *mask_zero* слоя внедрения, чтобы игнорировать входные данные со значением ноль. Модель обучается с использованием алгоритма обратного распространения. Оптимизация параметров выполняется с помощью *rmsprop*. Гиперпараметры выбираются на основе производительности набора данных разработки.

Метрика f1-показателя будет использоваться для оценки производительности модели, поскольку точность не является хорошей метрикой для несбалансированного набора данных *kazNER*.

Таблица 1 – f1-показатель на тестовом наборе моделей NB, RFC и CRF

Мо- дели	B- PER	I- PER	B- LOC	I- LOC	B- ORG	I- ORG	B- OTH	I- OTH	Общий с “O” меткой	Общий без “O” метки
NB	0.19	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.81	0.06
RFC	0.40	0.00	0.41	0.54	0.45	0.00	0.18	0.02	0.89	0.32
CRF1	0.97	0.97	0.96	0.90	0.80	0.69	0.70	0.60	0.99	0.92
CRF2	0.90	0.92	0.89	0.85	0.72	0.59	0.47	0.34	0.98	0.85
CRF3	0.98	0.98	0.98	0.92	0.87	0.81	0.82	0.76	0.99	0.95

Таблица 1 показывает, что модели NB и RF работали довольно плохо. Значения f1-показателя большинства классов были равны 0. Причиной этого является отсутствие необходимых признаков для принятия соответствующих решений. Кроме того,

Таблица 2 – Показатели на тестовом наборе гибридной моделей LSTM-CRF

Метки	Точность	Полнота	F1-показатель
ORG	0.73	0.73	0.73
PER	0.90	0.93	0.91
LOC	0.92	0.94	0.93
RES	0.58	0.69	0.63
Среднее	0.86	0.89	0.88

обе модели не учитывают контекст и просто запоминают слова и теги, которых недостаточно для точного распознавания. По сравнению с дискретными классификаторами, классификатор CRF показывает отличные результаты. После оптимизации видно, что более низкие значения регуляризации упругой сети ($C1 + C2$) приводят к наилучшей производительности модели - особенно для $C1$. Модель с перекрестным проверенным рандомизированным поиском показывает приличную оценку f1 с 0,95 в целом без тега «O».

На Таблице 2 приведены результаты эксперимента на гибридной модели LSTM-CRF. Результаты выглядят довольно хорошо, и это не требует каких-либо особенностей разработки по сравнению с моделью CRF. Преимущество CRF здесь не очень заметно, но если бы у нас был набор данных с более сложными именованными объектами, это было бы безусловно хорошим результатом.

6 Заключение

В данной работе представлена модель NER на основе статистического подхода и машинного обучения для казахского языка. Несмотря на то, что проведенные эксперименты показывают значительную производительность, модель может развиваться по различным направлениям. В будущем эта работа будет улучшена за счет изучения новых признаков, которые влияют на распознавание объектов. Набор данных kazNER будет дополнен новыми предложениями.

Так как NER работает с большим корпусом, нейронные сети очень эффективны в поиске именованных объектов в данных, чтобы обеспечить превосходную модель NER. Использование современных методов, таких как сочетание нейронной сети LSTM и CRF, прост и часто дает хорошие результаты, есть некоторые потенциальные недостатки [13]. Если мы не видели слово во время предвычисления, мы должны закодировать его как неизвестное и вывести его значение из окружающих его слов. Часто слово postfix или prefix содержит много информации о значении слова. Использование этой информации очень важно, если вы имеете дело с текстами, которые содержат много редких слов, и вы ожидаете много неизвестных слов во время вывода. Для кодирования информации на уровне символов мы будем использовать вложения символов и LSTM для кодирования каждого слова в вектор. Мы можем использовать практически все, что создает один вектор для последовательности символов, представляющих слово.

7 Благодарности

Эта работа проводилась в рамках грантов N^oAP05134272 и N^oAP08053085, финансируемых Министерством образования и науки Республики Казахстан.

Список литературы

- [1] Gislason P.O., Benediktsson J.S. and Johannes R. Random forests for land cover classification // *Pattern Recognition Letters*. - 2006. - Vol. 27. - P. 294–300.
- [2] Lample G., Ballesteros M., Subramanian S., Kawakami K. and Dyer C. Neural architectures for named entity recognition // *arXiv*. - 2016. - Vol. 1603.01360. - P. 1–11.
- [3] Makazhanov A., Sultangazina A., Makhambetov O. and Yessenbayev Z. Syntactic annotation of kazakh: Following the universal dependencies guidelines. a report // *Proceedings of the 3rd International Conference on Turkic Languages Processing (TurkLang 2015)*. - Kazan, Russia, 2015. - P. 338–350.
- [4] Manning C., Surdeanu M., Bauer J., Finkel J., Bethard S. and McClosky D. The Stanford CoreNLP natural language processing toolkit // *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*. - Baltimore, Maryland, USA, 2014. - P. 55-60.
- [5] Murphy K. Naive bayes classifiers // *Journal of the University of British Columbia*. - 2006. -Vol. 18. - P. 1–60.
- [6] Korobov M. sklearn-crfsuite // URL: <https://scikit-learn.org>. Accessed June 5, 2020.
- [7] Tolegen G., Toleu A., Mamyrbayev O. and Mussabayev R. Named Entity Recognition for Kazakh Using Conditional Random Fields // *CICLing2019: Springer Lecture Notes in Computer Science*. - La Rochelle, France, 2019.
- [8] Tolegen G., Toleu A. and Xiaoqing Z. Neural Named Entity Recognition for Kazakh // *Proceedings of the 4-th International Conference on Computer Processing of Turkic Languages*. - Bishkek, 2016. - P. 118–127.
- [9] Vilain M., Burger J., Aberdeen J., Connolly D. and Hirschman L. A model-theoretic coreference scoring scheme // *Proceedings of the 6th conference on Message understanding*. - Columbia, Maryland, USA, 1995. -P. 45–52.
- [10] Yadav V. and Bethard S. A survey on recent advances in named entity recognition from deep learning models // *Proceedings of the 27th International Conference on Computational Linguistics*. - Santa Fe, New Mexico, USA, 2018. -P. 2145–2158.
- [11] Hochreiter S. and Jurgen S. Long short-term memory // *Neural computation* 9. - 1997. - P. 1735–1780.
- [12] Huang Z., Wei X. and Kai Y. Bidirectional LSTM-CRF models for sequence tagging // *arXiv*. - 2015. - Vol. 1508.01991. - P. 1–10.
- [13] Dong C., Jiajun Z., Chengqing Z., Masanori H. and Hui D. Character-based LSTM-CRF with radical-level features for Chinese named entity recognition // *In Natural Language Understanding and Intelligent Applications*. - 2016. -P. 239–250.

References

- [1] Gislason P.O., Benediktsson J.S. and Johannes R. "Random forests for land cover classification", *Pattern Recognition Letters* vol. 27 (2006): 294–300.
- [2] Lample G., Ballesteros M., Subramanian S., Kawakami K. and Dyer C. "Neural architectures for named entity recognition", *arXiv* vol. 1603.01360 (2016): 1–11.
- [3] Makazhanov A., Sultangazina A., Makhambetov O. and Yessenbayev Z. "Syntactic annotation of kazakh: Following the universal dependencies guidelines. a report", *Proceedings of the 3rd International Conference on Turkic Languages Processing (TurkLang 2015)* (Kazan, Russia, 2015): 338–350.
- [4] Manning C., Surdeanu M., Bauer J., Finkel J., Bethard S. and McClosky D. "The Stanford CoreNLP natural language processing toolkit", *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations, Baltimore, Maryland, USA* (2014): 55-60.
- [5] Murphy K. "Naive bayes classifiers", *Journal of the University of British Columbia* vol. 18 (2006): 1–60.

- [6] Korobov M. "sklearn-crfsuite", URL: <https://scikit-learn.org>. Accessed June 5, 2020.
- [7] Tolegen G., Toleu A., Mamyrbayev O. and Mussabayev R. "Named Entity Recognition for Kazakh Using Conditional Random Fields", *CICLing2019: Springer Lecture Notes in Computer Science* (La Rochelle, France, 2019).
- [8] Tolegen G., Toleu A. and Xiaoqing Z. "Neural Named Entity Recognition for Kazakh", *Proceedings of the 4-th International Conference on Computer Processing of Turkic Languages* (Bishkek, 2016): 118–127.
- [9] Vilain M., Burger J., Aberdeen J., Connolly D. and Hirschman L. "A model-theoretic coreference scoring scheme", *Proceedings of the 6th conference on Message understanding* (Columbia, Maryland, USA, 1995): 45–52.
- [10] Yadav V. and Bethard S. "A survey on recent advances in named entity recognition from deep learning models", *Proceedings of the 27th International Conference on Computational Linguistics* (Santa Fe, New Mexico, USA, 2018): 2145–2158.
- [11] Hochreiter S. and Jurgen S. "Long short-term memory", *Neural computation* 9 (1997): 1735–1780.
- [12] Huang Z., Wei X. and Kai Y. "Bidirectional LSTM-CRF models for sequence tagging", *arXiv* vol. 1508.01991 (2015): 1–10.
- [13] Dong C., Jiajun Z., Chengqing Z., Masanori H. and Hui D. "Character-based LSTM-CRF with radical-level features for Chinese named entity recognition", *In Natural Language Understanding and Intelligent Applications* (2016): 239–250.