

Report 1.3

Lynx: System for Financial News Analysis Using NLP and Graphs

Andrei Zhdanov Mikhail Sofin

November 4, 2025

I. Team

Team Akira:

Andrei Zhdanov (an.zhdanov@innopolis.university)

Mikhail Sofin (m.sofin@innopolis.university)

II. Project topic

Lynx: System for Financial News Analysis Using NLP and Graphs.

III. Links

Link to Github repository with all source code and notebooks: [Lynx](#)

Link to Miro: [Miro](#)

IV. What has been done so far

In the previous report, we explored the NER + RE task for the financial and political domains in English. Over the past three weeks, our focus has been on two main areas: the graph frontend and investigating the NER + RE domain task in Russian.

A. Russian NER task

As part of the solution to the problem of extracting named entities (NER) from texts in Russian, a comprehensive analysis of the available tools was carried out, as a result of which the most optimal solution was chosen — the [Natasha](#) library, which is a specialized tool for natural language processing with a focus on Russian-language content. This library offers a full-featured set of features for extracting named entities, including recognition of key types of objects: names of people (PER), geographical names (LOC) and organizations (ORG). In addition to the basic NER, Natasha provides advanced functionality for extracting structured facts such as dates, sums of money, and addresses, and integrates tools for linguistic text processing, including lemmatization, morphological analysis, tokenization, and segmentation. This model was also trained on news, which suits us perfectly.

```

    v from natasha import (
        Segmenter,
        MorphVocab,
        NewsEmbedding,
        NewsMorphTagger,
        NewsSyntaxParser,
        NewsNERTagger,
        NamesExtractor,
    )

    segmenter = Segmenter()
    morph_vocab = MorphVocab()
    emb = NewsEmbedding()
    morph_tagger = NewsMorphTagger(emb)
    syntax_parser = NewsSyntaxParser(emb)
    ner_tagger = NewsNERTagger(emb)
    names_extractor = NamesExtractor(morph_vocab)

```

Fig. 1. Natasha model

To begin with, we developed a pipeline using Natasha and Rule Based RE.

```

Запрос:
{
    "text": "Путин немного подумав, как сообщил он, купил так давно желаемую Америку."
}

Ответ:
{
    "results": [
        {
            "sentence": "Владимир Путин немного подумав, как сообщил он, купил так давно желаемую Америку",
            "entities": [
                {"text": "Америку", "type": "LOC", "start": 9, "stop": 10},
                {"text": "Владимир Путин", "type": "PER", "start": 1, "stop": 2}
            ],
            "relations": [
                {
                    "subject": "Владимир Путин",
                    "subject_type": "PER",
                    "relation": "купил",
                    "relation_type": "FINANCIAL_TRANSACTION",
                    "object": "Америку",
                    "object_type": "LOC",
                    "sentence": "Путин немного подумав, как сообщил он, купил так давно желаемую Америку"
                }
            ]
        }
    ...
}

```

Fig. 2. endpoint doc

B. Russian RE task

For the RE task, we decided to make our model from scratch for more practice and see how much more effective it would be than the rule based RE. To do this, we first developed our dataset. We've retrieve about 1,000 sentence from RBK-type news feeds.

```

1 text
2 Орбан заявил о нежелании вводить евро в Венгрии из-за «развала» Евросоюза
3 "Использование евро может теснее связать Венгрию с Евросоюзом, заявил Орбан, отметив свое нежелание запускать этот процесс. По его словам, вв
4 Сентябрь в Москве второй раз с начала века оказался солнечнее августа
5 На Tesla подали в суд из-за аварии со сгоревшими в Cybertruck студентами
6 "В США родители двух студентов колледжа, погибших в аварии с Tesla Cybertruck, утверждают, что трагедия случилась из-за технической неисправн
7 Семин назвал неожиданным уход гендиректора «Спартака»
8 "Бывший тренер сборной России положительно оценил работу Олега Мальцева в «Спартаке», отметил, что специалист большое внимание уделял развити
9 Уличных музыкантов «парализовали» работу музыкальной школы в Петербурге
10 Администрация центрального района Санкт-Петербурга подготовила проект, в случае принятия которого уличным музыкантам запретят играть на двух
11 Ранкультуры предложили критерии привлечения гражданского центра к работе в сфере культуры
12 Юристы предложили для всех государственных учреждений контракт с онлайн-кинотеатрами, видеосервисами и на телеканалах. Вопросы будут сконцентриров
13 В Ростове засно проговорили к похищенному организатору покушения из СБУ
14 Южный окружной военный суд в Ростове на Дону засно проговорил к похищенному лишению свободы подполковника Службы безопасности Украины (СБУ)
15 "Кубъ узнала, что Сарсик расформировал бывшую группировку "кортица"
16 Группировка «Днепр» (бывшая «Кортица») отвела за распределение снарядов между бригадами. Руководивший ее командующий объединенных сил ВСУ в
17 В Москве наградили лауреатов ИР-премии «В карде»
18 Сменная гражданско Касаткина досрочно завершила сезон
19 Теннисистка в этом сезоне впервые с 2019 года больше матчи проиграла, чем выиграла. Она показалась на плотный график и эмоциональные проб
20 Умерла автор цикла книг «Хроники Ратири» Джилли Купер
21 Британская писательница Джилли Купер умерла в возрасте 88 лет. Она стала знаменитой благодаря серии любовных романов «Хроники Ратири»
22 Гладков сообщил о погибшем в результате ракетного удара по Белгороду
23 "Один человек погиб и еще один пострадал в результате ракетного удара по Белгороду, сообщил Гладков. По его словам, погибший помогал спасать
24 В Долгопрудном эвакуировали 11 из-за задымления
25 "Посетители торгового центра «ГородБелгород» в Долгопрудном эвакуированы из-за задымления, сообщили РБК в Главном управлении МЧС по Моск
26

```

Fig. 3. raw dataset

Then use LLM to create RE dataset from raw sentence.

```

{
  "sentence": "Орбан заявил о нежелании вводить евро в Венгрии из-за «развала» Евросоюза.",
  "entities": [
    {"id": "T1", "text": "Орбан", "type": "PER", "start": 0, "end": 5},
    {"id": "T2", "text": "Венгрия", "type": "LOC", "start": 37, "end": 44},
    {"id": "T3", "text": "Евросоюз", "type": "ORG", "start": 56, "end": 65}
  ],
  "relations": [
    {
      "head": "T1",
      "tail": "T3",
      "relation": "POLITICAL_STATEMENT",
      "trigger": "заявил о нежелании вводить евро"
    },
    {
      "head": "T1",
      "tail": "T2",
      "relation": "POLITICAL_SOCIAL_CHANGE",
      "trigger": "отказ от введения евро в стране"
    }
  ],
  {
    "sentence": "Сентябрь в Москве второй раз с начала века оказался солнечнее августа.",
    "entities": [
      {"id": "T1", "text": "Москва", "type": "LOC", "start": 11, "end": 17}
    ],
    "relations": []
  },
  {
    "sentence": "На Tesla подали в суд из-за аварии со сгоревшими в Cybertruck студентами.",
    "entities": [
      {"id": "T1", "text": "Tesla", "type": "ORG", "start": 3, "end": 8},
      {"id": "T2", "text": "США", "type": "LOC", "start": 0, "end": 0},
      {"id": "T3", "text": "Cybertruck", "type": "ORG", "start": 47, "end": 57}
    ]
  }
}

```

Fig. 4. RE dataset

For our dataset BiLSTM architecture model with an Attention mechanism was chosen. This decision was made for the following key reasons. BiLSTM is optimally suited for text processing, as it captures bidirectional contextual dependencies, which is crucial for defining relationships between entities. Adding an attention mechanism allows the model to automatically focus on the most relevant words in a sentence, ignoring noise, which is especially useful when working with long sequences. Architecture is a direct application of the material learned in the course(LSTM, Attention). The model is compact and efficient for training on limited computing resources, which is ideal for the purposes of this project. Such hybrid architectures (BiLSTM + Attention) are widely and successfully used in RE tasks[1], [2], including in the biomedical field, showing high efficiency in extracting complex relationships without the need for manual feature construction.

```

class BiLSTMAttentionRE(nn.Module):
    def __init__(self, vocab_size, embed_dim=32, hidden_dim=64, num_classes=num_classes):
        super().__init__()
        self.embedding = nn.Embedding(vocab_size, embed_dim, padding_idx=0)
        self.bilstm = nn.LSTM(embed_dim, hidden_dim // 2, bidirectional=True, batch_first=True)
        self.attention = nn.Linear(hidden_dim, 1)
        self.fc = nn.Linear(hidden_dim * 4, num_classes)

    def forward(self, input_ids, positions):
        embeds = self.embedding(input_ids)
        lstm_out, _ = self.bilstm(embeds)

        attn_scores = self.attention(lstm_out).squeeze(-1)
        attn_weights = torch.softmax(attn_scores, dim=-1)
        context = torch.sum(lstm_out * attn_weights.unsqueeze(-1), dim=-1)

        combined = torch.cat([context, positions], dim=-1)
        out = self.fc(combined)
        return out

model = BiLSTMAttentionRE(vocab_size)
criterion = nn.CrossEntropyLoss()
optimizer = optim.Adam(model.parameters(), lr=0.001)

```

Fig. 5. model architecture

C. Layout UI/UX implementation

Modern theme-supporting components and cards were created by using React + Vite + React Router tech stack. All elements are adaptive to screen change and made in warm colors.



Fig. 6. Layout Presentation

D. force-graph with link labeling

All relations from model are now parses and renders as RelationalCard objects. They include object and subject via nodes on a graph, and relation as labeled link.



Fig. 7. Relation Rendering

V. Work Distribution

- Andrei Zhdanov: Creating dataset for RE task, discover models for russian NER, develop model for russian RE task
- Mikhail Sofin: Frontend design implemented via layout and context cards with force graphs used, also connected the backend side with model.

VI. Plan for the Next Weeks

In the next three weeks, we plan to finish integrating the models, and, if possible, improve and finalize the models for the RE task. Test our app on a golden set and improve user interface.

References

- [1] G. Alfattni, N. Peek, and G. Nenadic, “Attention-based bidirectional long short-term memory networks for extracting temporal relationships from clinical discharge summaries,” *Journal of Biomedical Informatics*, vol. 123, p. 103915, Nov. 2021, ISSN: 1532-0464. DOI: [10.1016/j.jbi.2021.103915](https://doi.org/10.1016/j.jbi.2021.103915). [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1532046421002446>.
- [2] T.-H. Nguyen, T.-D. Nguyen, H.-Q. Nguyen, T.-T. Nguyen, and T.-B. Ho, “Relation extraction in biomedical texts based on multi-head attention model with syntactic dependency feature: Modeling study,” *JMIR Medical Informatics*, vol. 10, no. 10, e41136, Oct. 2022, ISSN: 2291-9694. DOI: [10.2196/41136](https://doi.org/10.2196/41136). [Online]. Available: <https://medinform.jmir.org/2022/10/e41136>.