

Руководство по использованию демо-версии компьютерной морфологии для анализа вариантов текста.

1. Демо-версия корпуса.

Для тестирования разработанных лингвистических ресурсов используйте корпус, составленный из трех редакций «Сказания о Мамаевом побоище»: списка Основной редакции РНБ О.IV.22, списка Летописной редакции СПбОИИ №251 и списка Киприановской редакции БАН 32.14.8 (файл “[skazanie_corpus_demo.noc](#)”). Тексты этих списков представлены в соответствии с правилами проекта «СКАТ» и проведено объединение графических вариантов, дополнительно (чтобы избежать неправильной сегментации на словоформы в NooJ) «ять» переобозначен как «Э», йотованное «е» как «Е». Корпус уже преобразован в формат корпуса NooJ.

2. Установка NooJ.

Загрузите и установите NooJ с сайта <http://www.nooj-association.org/> (http://www.nooj-association.org/index.php?option=com_k2&view=item&layout=item&id=4&Itemid=615). Инструкции по установке описаны в руководстве NooJ (https://kitwiki.csc.fi/twiki/pub/NooJ/WebHome/NooJ_Manual.pdf), стр. 18-22.

3. Загрузка корпуса и файлов морфологии.

Запустите NooJ. Для загрузки корпуса выберите меню File > Open > Corpus > filepath (путь к скачанному корпусу на Вашем компьютере). Скопируйте содержимое папки [Lexical Resources](#) в папку ...\\NooJ\\ru\\Lexical Analysis. Укажите, какие морфологические грамматики и словарь Вы будете использовать для анализа корпуса (можно загрузить все файлы из папки [Lexical Resources](#), кроме файла _properties.def), во вкладке меню Info > Preferences > Lexical Analysis > Add resources > Apply.

Перейдите к окну с текстом корпуса в NooJ, в меню появится вкладка CORPUS (см. рисунок 1). Выберите CORPUS > Linguistic Analysis для аннотирования корпуса. Сохраните результаты аннотирования (Ctrl+S или File > Save).

4. Поиск в корпусе.

Возможности NooJ подробно описаны в руководстве NooJ (https://kitwiki.csc.fi/twiki/pub/NooJ/WebHome/NooJ_Manual.pdf), стр. 31-43. Выберите в меню CORPUS > Locate и введите поисковый запрос <V+imperf_c>, нажмите Enter, , установив предварительно значения

- A NooJ regular expression
- All matches
- All occurrences (см. рисунок 2).

Запросы с использованием регулярных выражений пишутся в угловых скобках “< >”. Металингвистическая информация, которую можно использовать в запросах, перечислена в файле _properties.def: в этом файле перечислены части речи и грамматические категории, описываемые в морфологических грамматиках (.nom, .nof). Семантические свойства можно приписывать в словарях (.nod) после указания словоизменительного типа через “+” (в демо-версии указано только свойство “rpor” – собственные имена).

5. Примеры запросов с использованием регулярных выражений NooJ:

- а) Найти все стяженные формы имперфекта: <V+imperf_c> (рисунок 1).

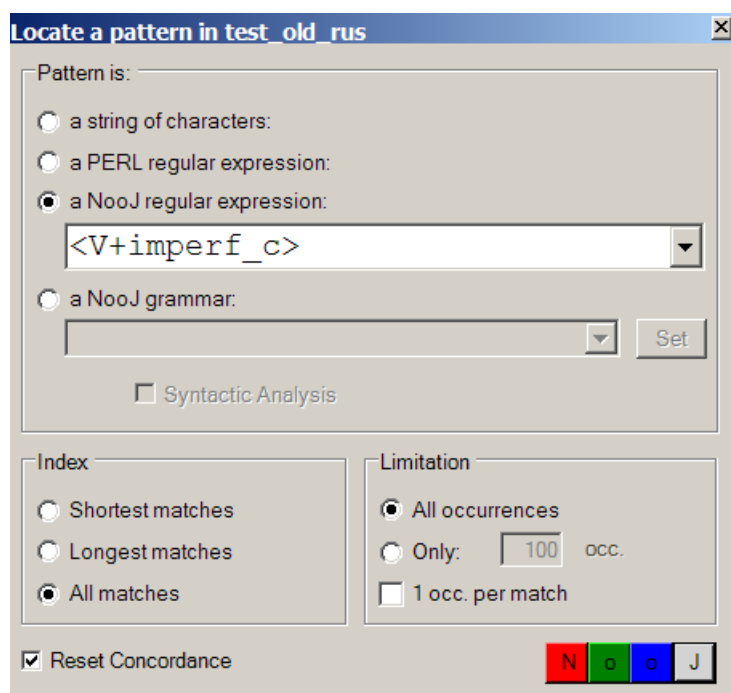


Рисунок 1.

б) Найти все словоупотребления под титулом (в демо-версии обработка титула только для глагола «глаголати»): <titlo>;

в) Найти все именные группы из прилагательного и существительного в именительном падеж: <ADJ+nom><N+nom> (рисунок 2)

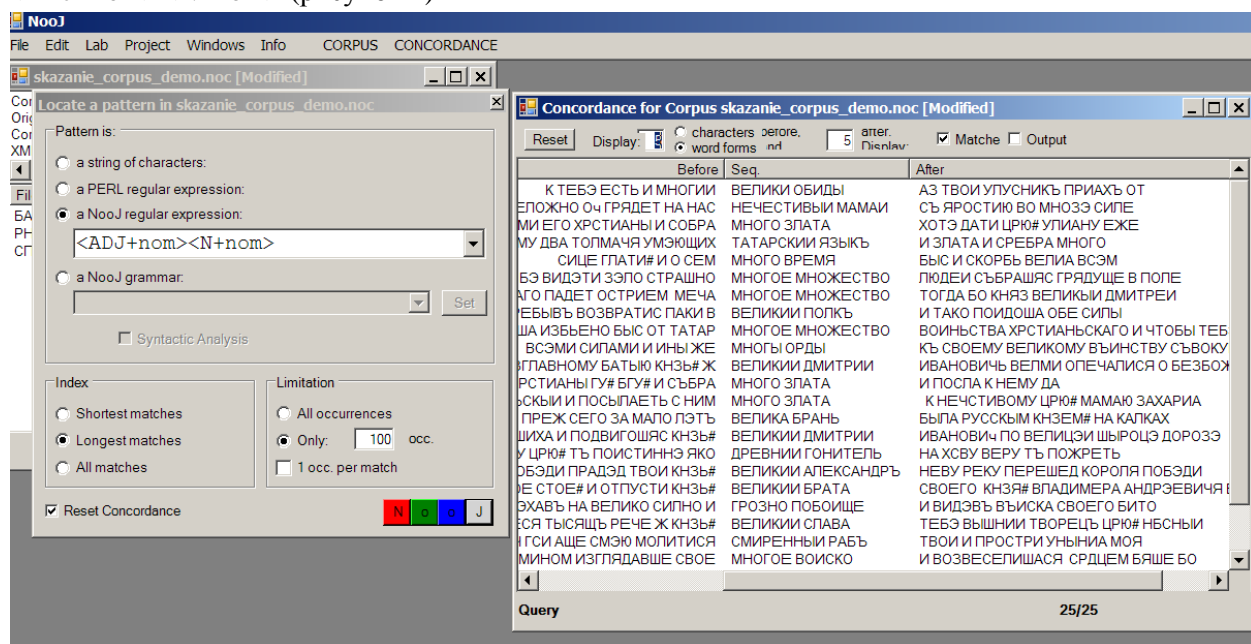


Рисунок 2. Запрос «в» к корпусу и результаты поиска.

г) Пример запросов для поиска по суперлексеме (суперлексема всегда в словаре записывается строчными буквами): а) <ли>, б) <победа>, с) <рязанский>, ниже – запись суперлексеми и ее вариантов в словаре dic_demo.nod.

ли,PART

ЛИ,ли,PART

Л,ли,PART

победа,N+FLX=A

ПОБЕДА,победа,N+FLX=A

ПОБЭДА,победа,N+FLX=A

Locate a pattern in skazanie_corpus_demo.noc

Pattern is:

☐ a string of characters:

☐ a PERL regular expression:

☒ a NooJ regular expression:

☐ a NooJ grammar:

<рязанский>

Set

☐ Syntactic Analysis

Index

☐ Shortest matches

☐ Longest matches

☒ All matches

Limitation

☒ All occurrences

☐ Only: 100 occ.

☐ 1 occ. per match

☒ Reset Concordance

N o J

Рисунок 3. Поиск в корпусе по суперлексе «рязанский».

NooJ - [Concordance for Corpus skazanie_corpus_demo.noc [Modified]]

File Edit Lab Project Windows Info CORPUS CONCORDANCE

Reset Display: 5 characters before, 5 after. Matches Output

Text	Before	Seq.	After
БАН 32.14.8.not	СОТВОРИП ЕСИ НАДО КНЗЕМ# ОЛГОМ	РЯЗАНСКИМ	ВСЯ ГРАДЫ ЕГО И ВПАСТИ
БАН 32.14.8.not	МЭРЫ ТОГДА Ж ОЛЕГЪ КНЯЗ	РЯЗАНЬСКИИ	УСПЫША ЯК МАМАИ КОЧЮЕТ НА
БАН 32.14.8.not	НА ВОРОНОЖЕ В ПРЕДЭЛЕХ ЕГО	РЯЗАНЬСКИХ	И ХОЩЕТЬ ИТИ НА ВЕЛИКОГО
БАН 32.14.8.not	МОСКОВЬСКОГО И ПОСЛА ОЛЕГЪ КНЯЗ	РЯЗАНЬСКИИ	ПОСЛА СВОЕГО КО ЦРЮ# МАМАЮ
БАН 32.14.8.not	ПОСАЖЕНИК И ПРИСЯЖЕНИКЪ ОЛЕГЪ КНЯЗ	РЯЗАНЬСКИИ	МНОГО ТЯ МОЛИТ СПЫШАХ УБО
БАН 32.14.8.not	ПОСЛА ТОИ Ж ОЛЕГЪ КНЯЗ	РЯЗАНЬСКИИ	К ВЕЛИКОМУ КНЗЮ# ЯГАИПУ ОЛГЕРДОВИЧЮ
БАН 32.14.8.not	УБО ПОМЫСЛИША В СЕБЗ ОЛЕГ	РЯЗАНЬСКИИ	И ЯГАИПО ЛИТОВСКИИ ЕГДА УСПЫШИТ
БАН 32.14.8.not	И СПЫШЕВ ЯКО ОЛЕГЪ КНЯЗ	РЯЗАНЬСКИИ	И ЯГАИПО КНЯЗ ЛИТОВЬСКИИ ПРИПОЖИШАСЯ
БАН 32.14.8.not	КНЯЗ ЛИТОВСКИИ И ОЛЕГЪ КНЯЗ	РЯЗАНЬСКИИ	СОВОКУПИШАС С МАМАЕМЪ НА НАС
БАН 32.14.8.not	НЕУТЭШНО СПЫШЕВЪ Ж КНЯЗ ОЛЕГЪ	РЯЗАНСКИИ	ЯКО КНЯЗ ВЕЛИКИИ ДМИТРЕИ ИВАНОВИЧ
БАН 32.14.8.not	ПРОТИВУ МАМАА ОЛЕГ ЖЕ КНЯЗ	РЯЗАНЬСКИИ	УСПЫШАВЪ СИА УСТРАШИСЯ И ВОЗТРЕПЕТА
БАН 32.14.8.not	И УСПЫША ЯКО ОЛЕГЪ КНЯЗ	РЯЗАНЬСКИИ	УСТРАШИС И ВОЗТРЕПАТА ЗЭЛО И
БАН 32.14.8.not	СПЫШЕВ ЖЕ ТО КНЯЗ ОЛЕГЪ	РЯЗАНСКИИ	ЯКО ГРЯДЕТ КНЗЬ# ВЕЛИКИИ ПОБЕДИВЪ
СП6ОИИ №251.not	ГЛЗБЫ СПЫШАВ ТО ОЛЕГЪ КНЯЗ	РЯЗАНСКИИ	ЯКО ЦРЬ# МОМАИ КОЧЮЕТ НА
СП6ОИИ №251.not	ПОСАЖЕНИК И ПРИСЯЖНИК ОЛЕГЪ КНЯЗ	РЯЗАНСКИИ	МНОГО ТЯ МОЛИТ СПЫШАХ ГСНЕ
СП6ОИИ №251.not	КНЗЮ# ЯГАИПУ ЛИТОВСКОМУ ОЛЕГЪ КНЯЗ	РЯЗАНСКИИ	РАДОВАТИСЯ ПИШУ ВЭМ БО ЯКО
СП6ОИИ №251.not	ЧТО ЯГАИПО ЛИТОВСКИИ И ОЛЕГ	РЯЗАНСКИИ	ПРИПОЖИЛИСЯ К МОМАЮ И ПОСЛА
СП6ОИИ №251.not	И ПОВЕДА ЕМУ КАКО ОЛЕГ	РЯЗАНСКИИ	И ЯГАИПО ЛИТОВСКИИ ПРИПОЖИЛИС И
СП6ОИИ №251.not	И ЯГАИПО ЛИТОВСКИИ И ОЛЕГЪ	РЕЗАНСКИИ	НО НЕ СПЕШИТ ЦРЬ# ТОГО
СП6ОИИ №251.not	СПЫШАВ ЖЕ ТО КНЯЗ ОЛЕГ	РЯЗАНСКИИ	ЯКО КНЯЗ ВЕЛИКИИ ДМИТРЕИ ИВАНОВИЧ
СП6ОИИ №251.not	ЯГАИПО ПОРАБОТАЕТ ЕМУ НО ОЛЕГ	РЯЗАНСКИИ	ПРИВОДИТ ИХ НАМ ЖЕ ПОДОБАЕТ
СП6ОИИ №251.not	ДМИТРЕЮ ИВАНОВИЧЮ ЧТО КНЯЗ ОЛЕГЪ	РЯЗАНСКИИ	ПОСЫПАЛ МОМАЮ НА ПОМОЩЬ СВОЮ

Рисунок 4. Результаты поиска по суперлексе «рязанский».