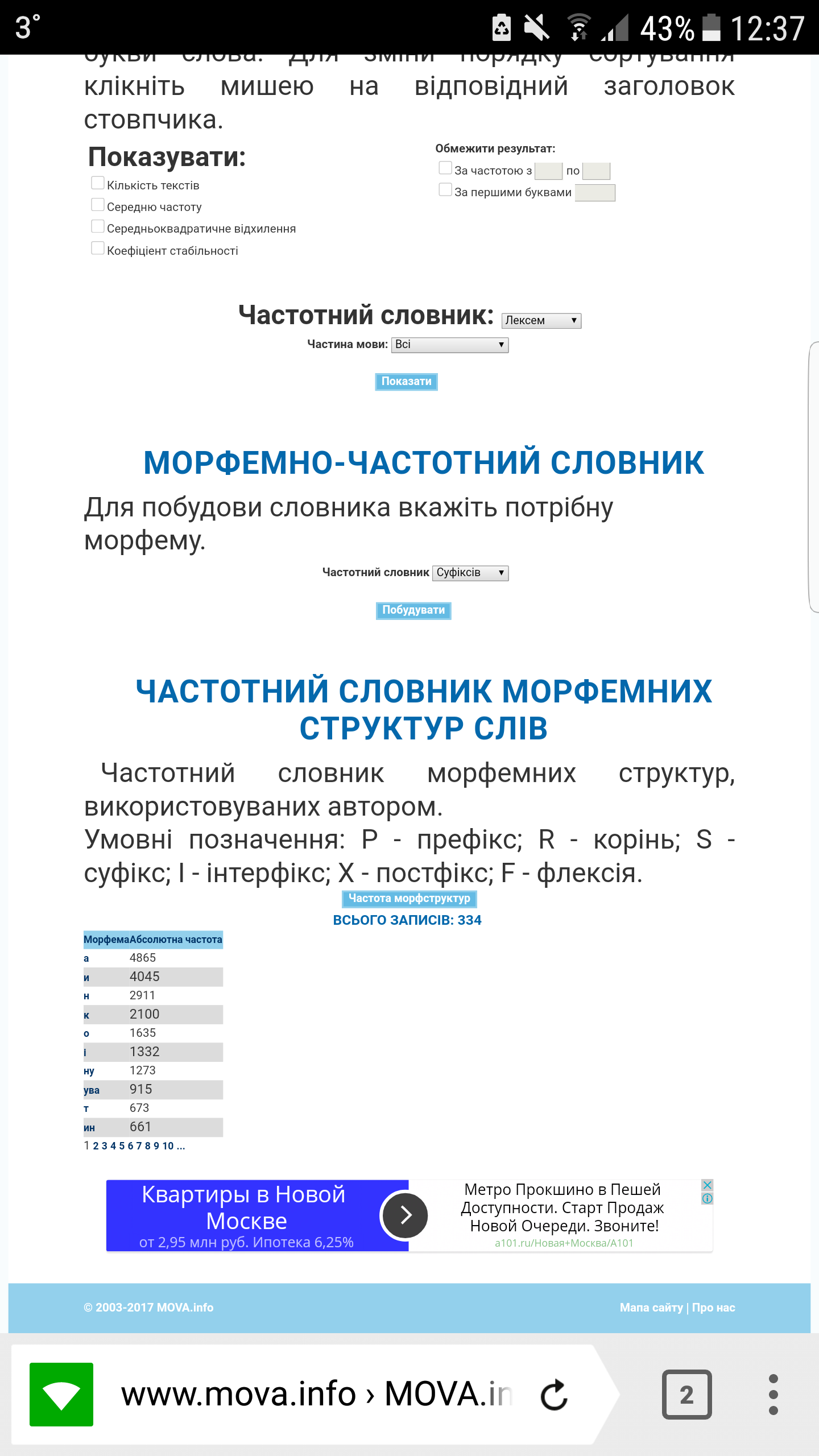
Мы, группа студентов образовательной программы «Фундаментальная и компьютерная лингвистика» Роман Тарасов (группа №2) и Ольга Кустовская (группа №1), провели экспертизу Корпуса текстов украинского языка, созданного сотрудниками КНУ им. Тараса Шевченко на базе образовательного портала mova.info и выяснили следующее:

1. Найти ресурс довольно просто, списки корпусов славянских языков присутствуют на многих сайтах Среди украинских ресурсов ссылки на корпус и сам портал mova.info присутствуют на официальном сайте украинского языка ukrainskamova.com и на странице лаборатории компьютерной лингвистики КНУ. Однако на зарубежных (западных) сайтах ссылки отсутствуют.
2. Что касается дизайна, цветовая гамма подобрана уместно, цвет- жёлто-голубой (стилизация под флаг Украины). Расположение кнопок достаточно удобное. Поиск дополнительных функций типа частотных словарей, синтаксического поиска, N-грамм затруднён. Однако большей части пользователей данные функции, скорее всего, не требуются, поэтому данный недостаток незначителен. Есть и более критичный минус- форма обратной связи расположена неудобно, спрятана в раздел «Про нас».
3. Инструкции для пользователей есть, однако они не охватывают весь функционал Корпуса. «Видеохелпы» отсутствуют, примеры запросов не предусмотрены. Существенным недостатком является одноязычность: так, корпуса многих языков имеют интерфейс на английском языке. В данном случае разработчиками предоставлен лишь украиноязычный интерфейс. Для пользователей Google Chrome существует возможность автоматического встроенного перевода на ряд языков, включая русский, но данный перевод часто не соответствует оригиналу. Так, в начале этой недели в русском переводе присутствовала графа: «Пол автора: Человек/Женщина». Очевидно, что имелось в виду «чоловiк»(укр. мужчина). К концу недели эта неточность перевода была исправлена, но перевод все еще нуждается в доработке. Одна, следует заметить, что украинский текст в целом не представляет сложности для носителя русского языка, т.к. языки являются родственными.
4. Функционал Корпуса позволяет выполнять ряд сложных запросов. Например:
   1. Поиск N-грамм: от 2 до 5 слов
   2. Поиск по грамматическим признакам: часть речи, падеж, время и.т.д.
   3. Поиск по подкорпусам: художественному, публицистическому, поэтическому, фольклорному, законодательному, научному.
   4. Частотные словари по этим подкорпусам с возможностью разбиения на отрасли (например, из научных текстов можно взять конкретно лингвистические) и сферы интересов. Эта функция работает сейчас в тестовом режиме.
   5. Частотные словари произведений украинских писателей и поэтов. Часть этих словарей доступны лишь по предварительной заявке с указанием организации и цели исследовании. В наличии- частотные словари лексем, словоформ. Пример поиска:

Мы хотим посмотреть какие суффиксы прежде всего встречаются в романе украинского писателя Василя Шкляра «Чёрный ворон».

Откроем частотный словарь этого сборника, выберем функцию «Морфемно-частотный словарь (Морфемно-частотний словник)», в поле выбора морфемы напишем: Суффиксов(Суфiксiв). Получили, что самые распространённые- а(4865), и(4045), н(2911). Скриншот:



Также есть указатель частотности морфструктур. Выясним, например слова какой морфструктуры чаще встречались в стихотворениях Леси Украинки. На странице её частотного словаря выберем функцию «Частота морфструктур». Видим, что самые частые структуры- RF, RSF, R. Самые редкие-RXX, RPPRS и другие. Скриншот:

