

ИНСТРУМЕНТЫ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ ОБРАБОТКИ ТЕКСТА ДЛЯ СОПОСТАВИТЕЛЬНОГО АНАЛИЗА

Макалада ар түрдүү тилдерди талдоо жана тексти лингвистикалык изилдөө маселелери каралган.

В вопросах лингвистического исследования текста важное место занимает сопоставительный статистический анализ текстов на различных языках народов мира.

The article is about the linguistic research of a text.

В современном информационном обществе возможности ученого-гуманитария значительно расширились в связи с появлением новых "продвинутых" информационных технологий (НИТ). Компьютерный анализ текстов представляет собой одну из наиболее перспективных областей применения математических методов и НИТ в гуманитарном исследовании. Именно здесь существует большой спрос на современные программные средства компьютеризированного анализа текстовых структур и их компьютерной визуализации.

В течение двух последних десятилетий компьютер стал объектом исследования многих гуманитарных наук. В настоящее время активно исследуются возникающие социо- и этнокультурные традиции коммуникации. Ученые, разрабатывающие такие новые направления, как медиаэкология и эколингвистика, свидетельствуют о том, что электронные средства массовой коммуникации значительно трансформируют способ восприятия и концептуализации действительности человеком¹.

Активно исследуются возможности использования компьютера в методике и практике преподавания иностранных языков.

Инструменты информационных технологий обработки текста для сопоставительного анализа содержат обращение к вопросам использования языка программирования Delphi и к использованию инструментариев информационных технологий при обработке сопоставляемых материалов на кыргызском и английском языках, а также о парадигме программирования в среде Delphi, модульное

¹ McLuhan 1962, ВебиНе:электронная версия, Postman 1976, Ерасов, 1997, Леонтович, 2000.

программирование и модули. Примеры использования в визуальном программировании часто используемых простых компонентов типа TButton, TLabel, TEdit, алгоритм организации, сопоставительного анализа частоты, использования главных членов предложения в кыргызском и английском языках и его компьютерная реализация.

Для решения задач даны тексты одного содержания на двух языках (кыргызский и английский). Нужно осуществить сопоставительный анализ текстов на двух языках при помощи статистической обработки данных подсчетом количества вхождений главных членов предложения на текстах сопоставляемых языков. Для решения этой задачи используем следующий алгоритм:

- Создаем электронный текст на двух сопоставляемых языках в виде перевода соответствующего текста (Т. Касымбеков. «Сынган кылыч» на кыргызском, «Broken Sword» на английском языках).
- Снабжаем дополнительным специальным символом-показателем главные члены предложения, соответствующие в тексте: S – подлежащее, P – сказуемое. Таким образом, создаем искусственный текст.
- Создание программного счетчика подсчета вхождения слова по данному критерию в тексте.
- Вывод количества вхождений требуемых слов в тексте на сопоставляемых языках.

Данная программа реализована в системе программирования Delphi, которая состоит из двух модулей Unit1. и Unit2. Unit1 заключен графическим интерфейсом. В компоненте T.Memo2 находится обрабатываемый текст. Имеется возможность сохранить как результаты поиска, которые создаются в Memo1, файл под заданным определенным именем. Предварительно надо привести подготовительную работу по преобразованию с расширением .txt, преобразованный из Word документа.

В Unit2 находится функция, применяемая в Module1 в служебной функции. В Module1 события регулируются четырьмя кнопками. Используются события on click «Открыть шаблоны поиска, сохранить, открыть текст, считать и содержаться», а второе Memo поля содержат свои методы, методы save to file и load from file.

Учитывая, что нет точных лингвистических критериев определения членов предложений ни на одном национальном языке, используем искусственный прием с предварительным выявлением членов предложений самим исследователем (человеком). Это является общей проблемой в лингвистике. Используя указанные обозначения электронного текста на кыргызском и английском языках, получаем новый текст, с которым будем работать по выявлению частоты использования главных членов

предложений по указанным главным критериям в виде знаков. Текст представляется в виде файлов с расширением txt, который состоит из элементов в виде строк как последовательность слов. Подсчет вхождения слов (S) и (P) производим с помощью программы Delphi. Результат анализа текстов в первой части в объеме 24 страниц дал следующие результаты

Таблица 1

:Результаты анализа текстов

Языки	Подлежащее	Сказуемое
кыргызский	122	161
английский	48	35

Фрагменты текстов на кыргызском и английском языках даны на примере романа Т.Касымбекова «Сынган кылыч».

Фрагменты преобразованного текста с использованием главных членов предложения текста на кыргызском языке:

БИРИНЧИ БОЛУМ

1865—жыл....

Генерал (S)Черняев боз чаптуу денде жалгыз кездуу дурбу салып турду. Бет алдындагы туз талаа гана эмес, те алда кайда карарган чоочун Ташкен (S) дубалы, мунарыктын алдында жаткан дарактуу (S) кыштактар, сай бойлорундагы кудурейген (S) камыштар, андан бери канат-туу чегирткелердей жайнап, акырын жылып келаткан квп (S) аскер даана (P) жатты. Кылка — кылка атчандар. Арты керунбей ийрилип, чубалган жввлер. Боз чан, ар жердей бирткелеп тутеп, кайра жерге суйрвлуп, суйдан, сарала тулан. чептерге, жалгыз аяк кек жантактарга самсып тегулет. Алыстан (S) ат (P) кишенейт. Адамдын жан-жуну кулакка илешет. Кокондуктар мынчалык камдуу келерин (S) генерал (P) кутпеген эле. Бирок, журвгунун орту оче элек курагы, аксаргыл жузунвн кандайдыр бир кооптонуу керунбоду, женишинин тагдырына санаркоо болбоду. Кайта, эч капарсыз гана, каралжын саргыч мурутунун алдынан текебер жымыйып тиктеп турду. Бу ийри кылыччандарды биринчи жолу (P) керуп отурабы?

It was 1865. General Chernayev stood on a hilltop covered with grass long scorched by the torrid Asian sun and looked through a telescope. He could now see not only the steppe before him but also the dark outlines of Tashkent in the distance, the green villages with high minarets, the reeds lining the rivers, and the host of Khokand warriors crawling like ants through the valley. There were mounted detachments and long lines of infantry. Dust rose high in the air to

settle again on the ground, on the sparse feather-grass turned yellow and the bushes of camel burr.

The neighing of horses and unsteady din of hundreds of human voices could be heard on the hilltop. The general had not expected the Khokand commander to be so shrewd. However, his well-groomed face showed no doubts that victory could be his. The general was carefully watching the picture that was unfolding before him, a satisfied smile played on his lips under the brown brush of moustache. This was not the first time he had seen these warriors with their curved swords. He had 1,500 soldiers armed with repeating rifles and artillery. What was there to worry about? The general had seen nothing to provoke concern so far.

Пример применения пакета

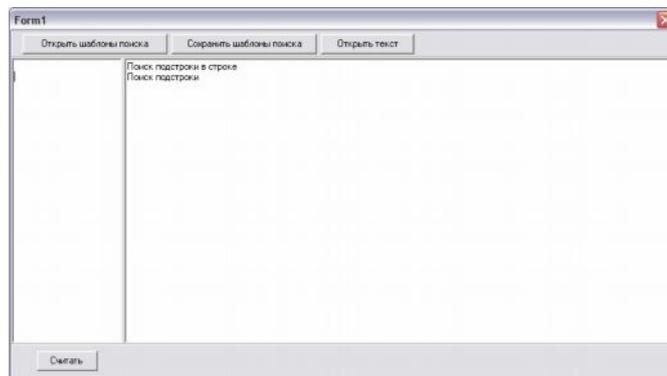


Рис. 1. Запуск приложения

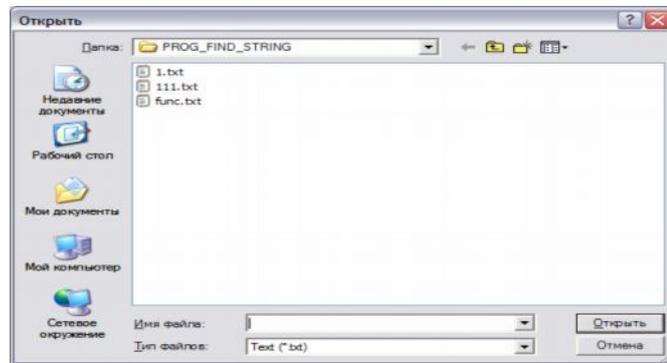


Рис. 2. Загрузка текстового файла для анализа

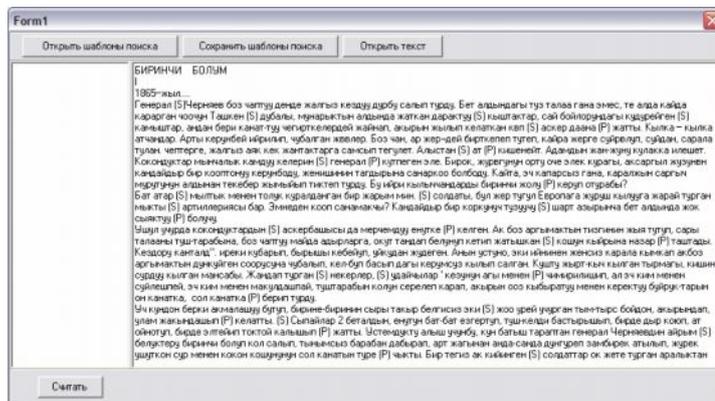


Рис. 3. Текст после загрузки файла

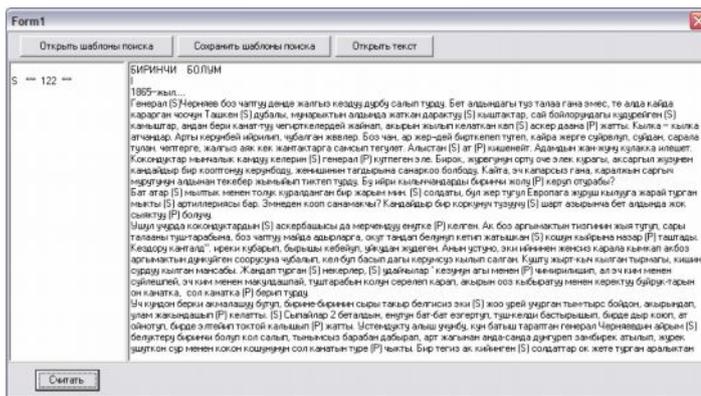


Рис. 4. Подсчет по признакам S

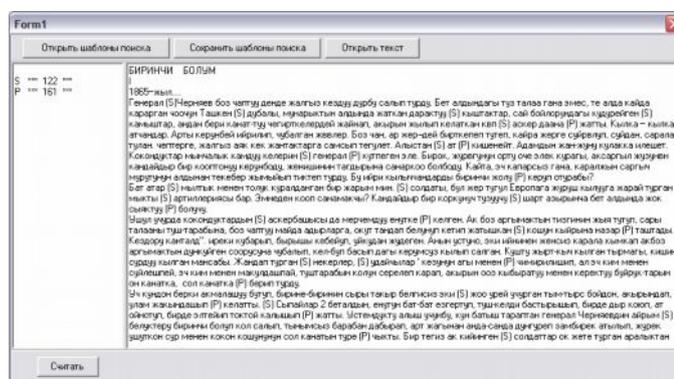


Рис. 5. Подсчет по признакам P

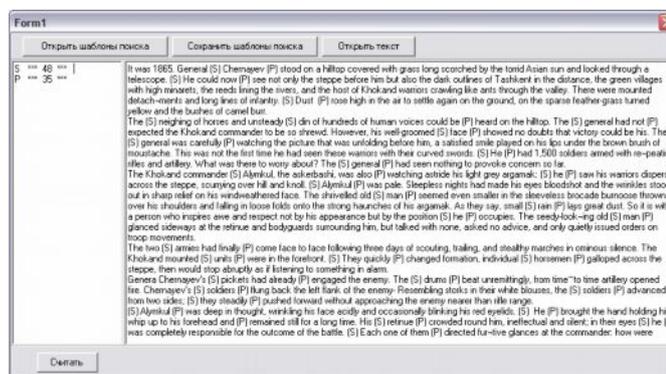


Рис. 6. Текст на английском после анализа

Список литературы

1. Основы информатики: Учебник для студентов вузов /Баячорова Б. Ж. – Бишкек, 2001.. – 258 с.
2. Парадигмы программирования. Учебное пособие /Милош Е., Джорупбеков С. –Бишкек, 2009. – 208 с.

3. Кузьмин Ю.Г. Перевод как мыслительно-речевая деятельность //Тетради переводчика, вып. 12. – М.: Междунар. отношения, 1975.
4. Цвиллинг М.Я., Туровер Г.Я. О критериях оценки перевода //Тетради переводчика, вып. 15. – М.: Междунар. отношения, 1973– С. 5.
5. Словарь русского языка. В 4-х т. Т.Ш. – М.: Гос. изд-во иностр. и нац. словарей, 1959.
6. Гак В.Г. Сопоставительная лексикология. – М.: Междунар. отношения, 1977. – С. 5.
7. Цвиллинг М.Я., Туровер Г.Я. О критериях оценки перевода. – С. 6.
8. Комиссаров В.Н., Туровер Г.Я. Перевод как лингвистический источник //Тетради переводчика, вып. 12. – М.: Междунар.отношения, 1975. – С. 27.
9. Апресян Ю.Д. Об одном правилесложения лексических значений //Проблемы структурной лингвистики, 1971. – М.: Наука, 1972. – С. 442.
10. Кьюо Дж., Джеанини М. Объектно-ориентированное программирование: Учебный курс. – СПб.: Питер, 2005. – 238 с.
11. Омуралиев А., Табышалиев Р. Информатика негиздери. Окуу куралы. – Б., 2002. – 360 б.
12. Баячорова Б.Ж., Омуралиев А.С. Эсептоочу машиналар, информатика жана программалоо. – Фрунзе, 1988. – 107 б.
13. Математика. Терминдердин тушундурмо создугу. – Фрунзе: Кыргыз Совет энциклопедиясынын башкы редакциясы, 1989. – 208 б.
14. Омуралиев А. Маалыматтар технологиясы/ КТМУ. - Бишкек, 2001. – 317 б.