

## **КОМПЬЮТЕРНАЯ ОБРАБОТКА ТЕКСТА. ПРОГРАММА ПОДСЧЕТА СЛОВ ДЛЯ СОПОСТАВИТЕЛЬНОГО АНАЛИЗА ТЕКСТА ПО ЗАДАННОМУ КРИТЕРИЮ**

*Бул макалада текстке салыштырма талдоо жүргүзүү үчүн атайын түзүлгөн программа (DELPHI) жөнүндө айтылат.*

*Данная статья – об осуществлении сопоставительного анализа текста с помощью программы подсчета слов (на DELPHI) по заданному критерию.*

*Cet article est de realisation de l'analyse comparee du texte a l'aide du programme composee (au DELPHI)*

В настоящее время существует множество программ для работы с текстом. Среди них специализированные текстовые процессоры, предназначенные для обработки больших объемов данных, программы, а также программные средства разработки, в основе которых лежит написание программного кода. Такие программные комплексы часто обладают определенными особенностями, которые облегчают программисту процесс создания программы. Например, выделяются ключевые слова конкретного языка программирования, “подсказчик кода” или code completion, позволяющие не сделать ошибку в выборе конкретного оператора, кроме того, пользователь всегда может настроить редактор текста под себя: увеличить или уменьшить размер шрифта, изменить цвет текста и так далее. Но каждый такой программный комплекс ориентирован в основном на конкретный язык программирования или тип документа, реже позволяет работать с несколькими типами файлов. Для программиста часто возникает необходимость работы с множеством текстовых файлов, содержащих программный код. Для этого используют специальные текстовые редакторы, которые поддерживают множество типов файлов и обладают возможностью гибкой настройки. Поэтому хотелось создать программу, которая позволит быстро и удобно обрабатывать программный код, компилировать его при необходимости для обработки текстовых данных.

В настоящее время такие программы создаются при помощи систем программирования, таких как Microsoft Visual Studio, Borland Delphi, Borland C++ и других. Перед тем, как начать работать над программой, необходимо было выбрать оптимальную среду разработки и наиболее удобный язык программирования. Object Pascal, с которым работает Delphi, является самым удобным средством разработки на данный момент, и система Delphi обладает развитыми средствами отладки. Как известно, в Delphi есть возможность использовать компоненты, то есть дополнительные части кода, созданные сторонними разработчиками. Таким образом, например,

удалось организовать плавающие настраиваемые меню и быструю обработку текста, проверку орфографии и использование словарей. Так как программа представляет собой простой текстовый редактор и немногим более функциональна, чем стандартный Блокнот. В данном случае рассматривается следующая задача: даны тексты одного содержания на двух разных языках: кыргызский язык и французский язык. Нужно осуществить сопоставительный анализ текстов на двух языках по статистической обработке данных с подсчетом количества использования частей речи (существительных и прилагательных) на текстах сопоставляемых языков. В качестве исходных текстов берем тексты на примере произведении «Жамиля» Ч.Айтматова на кыргызском и французском языках. Для решения поставленной задачи используем следующий алгоритм: создаем исходный электронный текст на двух сопоставляемых языках в виде перевода соответствующего текста (на примере повести Ч.Айтматова «Жамиля»); снабжаем дополнительно специальным символами-показателями части речи перед соответствующим словом в тексте. Например, слова существительные снабжаем символом (S), слова прилагательные символом (A). Таким образом создаем новый модифицированный текст для дальнейшей обработки; создание программного счетчика подсчета входного слова – части речи по данному критерию в модифицированном тексте; выводим количество вхождений обработки текста требуемых слов на сопоставляемых языках как результат обработки текстов. Указанный алгоритм реализован в системе программирования Delphi. Программа состоит из двух модулей: Unit1, Unit2. В модуле Unit1 находится список искомых слов. С помощью Unit1 создается графический пользовательский интерфейс, поэтому форма, соответствующая модулю, содержит такие компоненты, как Button1, 2, 3, 4, Memo1,2. В Memo1 представлен список искомых слов, Memo 2 содержит обрабатываемый текст. Имеется возможность сохранять как результаты поиска, которые содержатся в Memo1, файл под определенным именем. Предварительно надо провести подготовительную работу по преобразованию в текстовый формат с расширением txt. В модуле Unit 2 находятся функции, применяемые в Unit1. События регулируются четырьмя кнопками Button 1, 2, 3, 4. Это преобразовывается следующим образом: открыть шаблон поиска, сохранить шаблон поиска; открыть текст; считать. Для открытия и сохранения файлов используются компоненты TOpenDialog, TSaveDialog. Разделителем является компонент TSplitter. Процедурой Open1 Click загружается в Memo1 выбранный файл. Процедурой Save Click сохраняется выбранный файл в Memo1. Процедурой Open2 Click в Memo2 загружается выбранный файл.Процедурой Button1 Click делается подсчет слов. Используются две переменные для строки и пять – для целых чисел. Процедурой Button1 Click исполняется функция nfindRecur, которая находится в модуле Unit2. Эта функция делает рекурсивный поиск искомой подстроки по всему тексту. С помощью функции Pos находится место первого вхождения искомого слова в обрабатываемом тексте. С помощью Delete удаляется текст, находящийся до первого искомого слова в тексте.

**Чынгыз Айтматов «Жамила»**

Ар дайым бир жакка жол жүрөрдө, мен ушул алкагы жөнөкөй жыгачтан жасалган сүрөттүн алдына келип турам. Мына эртең да айылга жөнөймүн. Сүрөттү карап, мен андан жолума ак тилек бата алып жаткан өңдүү, аны көпкө көз айрыбай тиктейм.

Ушул күнгү чейин бул сүрөттү эч бир көргөзмөлөргө да берген жокмун, ал тургай айылдан туугандарым келгенде, көздөн далдалап бекитип коём. Анча эле жашыргандай эмнеси бар, уяттуубу деп, кокус оюңарга кетип жүрбөсүн, - жок бул сүрөттүн эч кандай ыксыз жайы деле жок, же болбосо, ага «көз тийип» жетет дегендей ал бир ашкан укмуш да эмес. Биринчи көрүшкө жөнөкөй эле кадыресе сүрөт. Бетине тартылган жер кандай жөнөкөй болсо, сүрөт өзү дагы ошондой жөнөкөй

Кийинчерээк берки үйдүн ээси дүйнөдөн кайтып, артында аялы менен тестиер эки уулу калат. Илгертен калган адат боюнча агайын-туугандар жесирдин башын байлап коёлу деп, арбак, кудайга тууралап менин атама никелештирип коюшат. Ал үй жонунан өзүнчө түтүн болуп саналып, мал оокаты, чарбагы бөлөк болгону менен, чынында биз барыбыз бир үй-бүлө болчубуз.

### *Traduction de Louis Aragon*

#### **Djamilia**

Et me revoila devant ce petit tableau dans son cadre modeste. Demain, des le matin, il me vaut aller a l'ail! et je regarde le tableau longuement et attenlivement comme si il allait me dire bon voyage. Ce tableau-la, je ne l'ai jamais encore envoye aux expositions. De plus, quand il vient chez moi, de l'ail, des gens de ma parente, je m'e fforce ,de le dissimuler. Non qu'il y ait en lui quelque raison de honte, mais c'est loin d'etre un exemple d'art. il est simple, comme simple est la terre qui y est representee. Dans la profondeur du tableau, il y a la ligne d'un ciel fane d'automne, ie vent, par-dessus unr lointaine rangee de montagnes, chasse de rapides petits nuages pie. Au premier plan, c` est un rouge brun, la steppe des absinthes. Et le chemin noir qui n'a guere eu le temps de secher, apres les pluies recentes, Sur les bas-cotes, secs, se serrent des arbrisseaux brises. Le long de l'orniere detrempee, viennent s'aligner les traces de deux voyageurs; plus elles s' enfoncent loin, plus elles sont faibles sur le chemin, et, quant aux voyageurs, on dirait qu'ils n'ont plus qu'un pas a faire pour surtir du cadre.

Form 1

Открыть шаблоны поиска    Сохранить шаблоны поиска    Открыть текст

стр  
по

Djamilia

Et me revoila devant ce (A) petit (S) tableau dans son cadre (A) modeste.  
Demain, des le matin, (S) il me vaut aller a l'ail et (S) je regarde le tableau longuement et attenlivement comme si (S) il allait me dire (A) bon voyage.

Ce (S) tableau-la, (S) je ne l'ai jamais encore envoye aux expositions. De plus, quand (S) il vient chez moi, de l'ail, des gens de ma parente, (S) je m'e fforce ,de le dissimuler. Non qu'il y ait en lui quelque (S) raison de (S) honte, mais c'est loin d'etre un (S) exemple d'art. il est simple, comme simple est la (S) terre qui y est representee.

Dans la (S) profondeur du (S) tableau, il y a la (S) ligne d'un ciel fane d'automne, ie vent, par-dessus unr lointaine (A) rangee de (S) montagnes, chasse de rapides (A) petits (S) nuages pie. Au (S) premier (S) plan, c` est un (A) rouge brun, la (S) steppe des absinthes. Et le (S) chemin (A) noir qui n'a guere eu le (S) temps de secher, apres les (S) pluies recentes, Sur les bas-cotes, secs, se serrent des (S) arbrisseaux (A) brises. Le (S) long de l'orniere detrempee, viennent s'aligner les traces de deux voyageurs; plus elles s' enfoncent loin, plus elles sont faibles sur le chemin, et, quant aux voyageurs, on dirait qu'ils n'ont plus qu'un pas a faire pour surtir du cadre. L'un d'eux ... mais j'anticipe un peu.

C'etait au temps de ma prime jeunesse.

(S) J'n etait dans la (A) troisieme annee de la guerre. Sur les fronts (A) lointains, quelque part devant Koursk et Orel, se battaient nos (S) peres et nos (S) freres, et nous, qui n'avions pas alors nos quinze ans, (S) nous travaillions au kolkhoze. (S) Un travail d'homme.

(A) Jour, quotidien, reposait sur nos epaules mal assurees, (S) Le travail, etait particulierement dur , aux jours de la maison. Des semaines entieres, (S) nous n'etions pas meme passes a la maison et nous avions donne le jour et la nuit aux champs, sur l'aire ou en chemin vers la gare ou se menait ou (S) le grain.

Par l'un de ces jours (A) torides, quand (S) les fa1 cilles, eut-on dit, etaient portes au rouge a force de moissonner, moi, comme je m'en retourmais de la gare sur la bitchka a vide, (S) je me decidai a tourner bride vers la maison.

Tout a cote du gue, sur la colline, la ou se termine la rue, il y a deux (S) cours (A) entourees d'une (A) solide cloture de bois. Autour de la propriete s'elevent des (S) peupliers. Ce sont nos (S) maison. Nos deux (S) familles vivent en voisinage depuis de (A) lointaines epoques. Pour

Считать

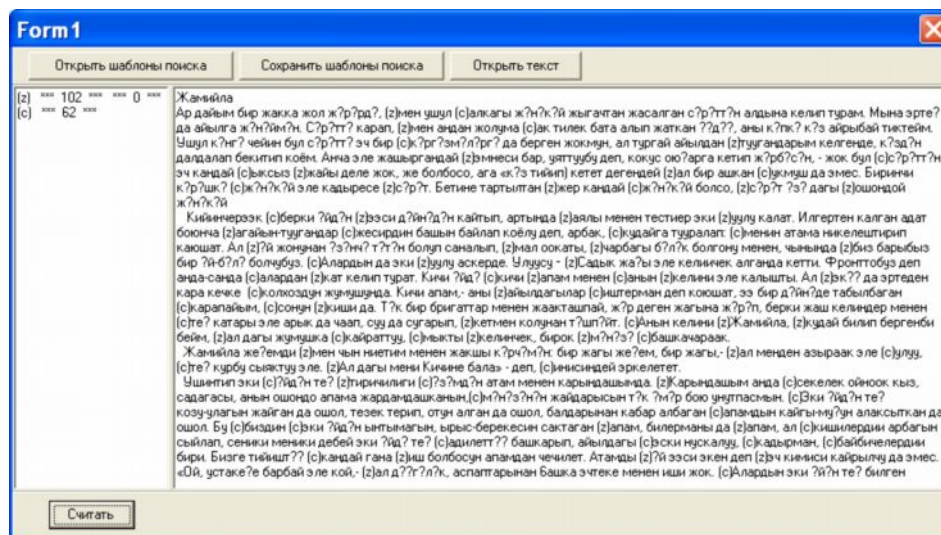


Таблица частоты использования существительных и прилагательных на кыргызском и французском языках на примере повести Ч.Айтматова «Жамияла»

Язык	Объем книжной страницы	Части речи	
		сущ.	прил.
Кыргызский	4,5 стр.	102	62
Французский	10 стр.	161	66

Из результатов обработки текстов можно сделать следующий вывод.

Использование существительных в кыргызском языке меньше, чем во французском языке, примерно на 60 %. Использование прилагательных в кыргызском и французском языках почти совпадает. Задача компьютерной обработки текста с целью организации лингвистического анализа текста является актуальным и достаточно сложным вопросом исследования. Это обстоятельство объясняется новизной направления науки – математической лингвистики и компьютерной обработки естественного языка как новой ветви в лингвистике, которая недостаточно развита не только в нашей стране, но и во многих странах мира. В связи с этим отсутствуют соответствующие кадры, кроме того, компьютерная обработка текста еще не получила широкого применения в стране, отсутствуют кадры, умело использующие инструменты информатики в лингвистических задачах.

### СПИСОК ЛИТЕРАТУРЫ

1. Баячорова Б.Ж. Основы информатики: Учебник для студентов вузов. – Бишкек, 2001.
2. Милош Е., Джорупбеков С. Парадигмы программирования: Учебное пособие. – Б., 2009.
4. Макьюин К. Дискурсивные стратегии для синтеза текста на естественном языке. – Новое в зарубежной лингвистике. Вып. XXIV, Компьютерная лингвистика. – М., 1989.

5. Попов Э.В., Преображенский А.Б. Особенности реализации ЕЯ-систем. – Искусственный интеллект. Кн. 1. Системы общения и экспертные системы. – М., 1990.
6. Преображенский А.Б. Состояние развития современных ЕЯ-систем. – Искусственный интеллект
7. Баранов А.Н. Категории искусственного интеллекта в лингвистической семантике. – М., 1987.
8. Советов Б.Я. Информационная технология. – М.: Высшая школа, 1992.
9. Тимофеев А.В. Информатика и компьютерный интеллект. – М.: Педагогика, 1991. – 128 с.
10. Айтматов Ч. «Ш томдон турган чыгармалар», I-том. – Фрунзе: Кыргызстан, 1982 ж.
11. Lois Aragon “ Djamilia” d`après Tch. Aitmatov. – Paris, 2001.