

Самарский государственный архитектурно-
строительный университет

Компьютерный анализ классификационных систем
при автоматическом индексировании текстов.

Выполнил: Байских Геннадий
студент 1-го курса

Научные руководители: проф. д.т.н. Пиявский С.А.
ассистент Будаев Д.С.

Краткие исторические сведения по УДК.

- В основу УДК была положена "Десятичная классификация" М. Дьюи.
- В 1876 г. вышло первое, очень краткое, издание таблиц ДКД.
- В 1905 г. в Брюсселе на французском языке вышло первое сводное издание таблиц новой десятичной классификации.
- В 1933 г. на французском языке закончилось второе полное издание таблиц десятичной классификации, теперь оно называлось "Универсальная десятичная классификация".
- В 1962 г. в стране в качестве единой обязательной классификации принята Универсальная десятичная классификация (УДК), а с 1963 г. в научно-технических издательствах, редакциях научно-технических журналов, в органах научно-технической информации, научно-технических библиотеках введено обязательное индексирование всех публикаций по Универсальной десятичной классификации, т. е. все информационные материалы в области естественных и технических наук издаются с индексами Универсальной десятичной классификации (УДК).

Краткие исторические сведения по ББК

- Сегодня мы отмечаем 30-летие ББК.
- Процесс внедрения ББК начался с выходом в свет первого выпуска таблиц.
- Распространение ББК сдерживалось крупным недостатком системы обозначений.
- В 1968 г. вышло в свет последнее издание таблиц классификации для массовых библиотек под редакцией З.Н.Амбарцумяна, содержательно переработанное в соответствии с ББК.
- Таблицы ББК для массовых библиотек (в одном томе) вышли в свет лишь в 1977 г. и переизданы в 1988 г.. В 1978, 1986 и 1988 гг. публиковался вариант для детских и школьных библиотек.
- В 1989 г. вышел в свет специализированный вариант ББК для краеведческих каталогов.
- Сегодня ББК - это система вариантов и изданий.
- Библиотекам нужны живые, постоянно развивающиеся КС, оперативно реагирующие на все изменения в общественной и научной жизни.

Цель работы:

Целью моей работы было автоматизировать процесс индексирования текстов при их вводе в интернет и добиться наиболее высокой скорости работы программы.

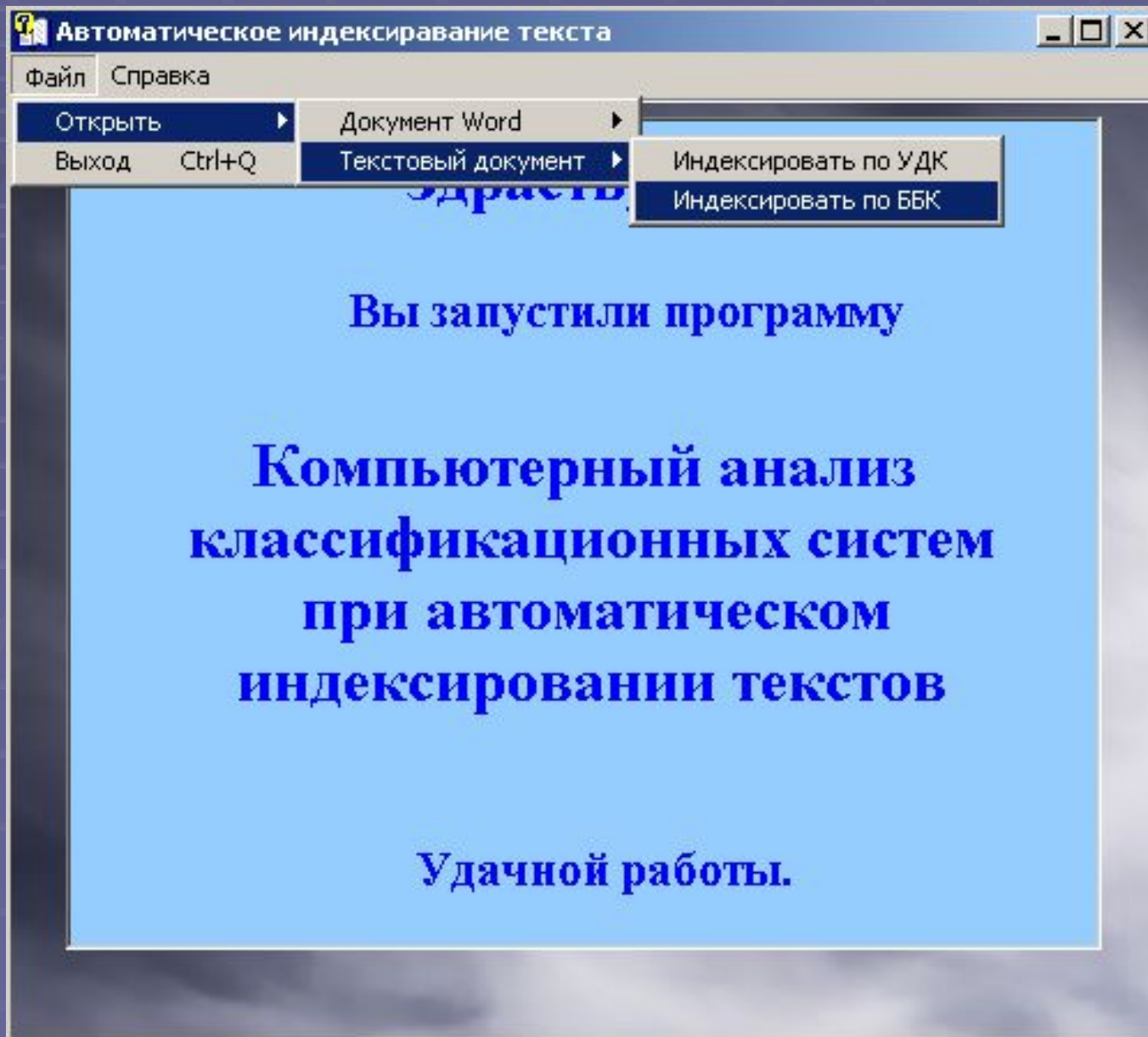
Фрагмент УДК обрабатываемый программой

- 681.3.01 - принцип обработки данных
- 681.3.02 - машины и их составные части по конструкции и компоновке
- 681.3.04 - представление данных: числами, алфавитом, кодами
- 681.3.06 - математическое обеспечение (программирование). Программы ЦВМ.
- 681.3.07 - аспекты запоминающих устройств (ЗУ) и ввода- вывода.
- 681.3.08 - принципы измерения, измерительные приборы, погрешности измерения и их коррекция.
- 681.31 - вычислительная техника в целом.
- 681.32 - цифровые или преимущественно цифровые системы, машины и устройства.
- 681.321 - настольные калькуляторы (простые, управляемые) вручную машины, на которых каждая операция выполняется шаг за шагом.
- 681.322 - автоматические цифровые вычислительные машины (ЦВМ). Комплектные машины с хранимой модифицируемой программой, которая может включать операции принятия решений и условные операции.
- 681.323 - специализированные цифровые системы или комплектные машины.
- 681.324 - вычислительные системы. Многомашинные комплексы. Сети ЦВМ.
- 681.325 - устройства, блоки, элементы для преобразования данных сортировки и счета

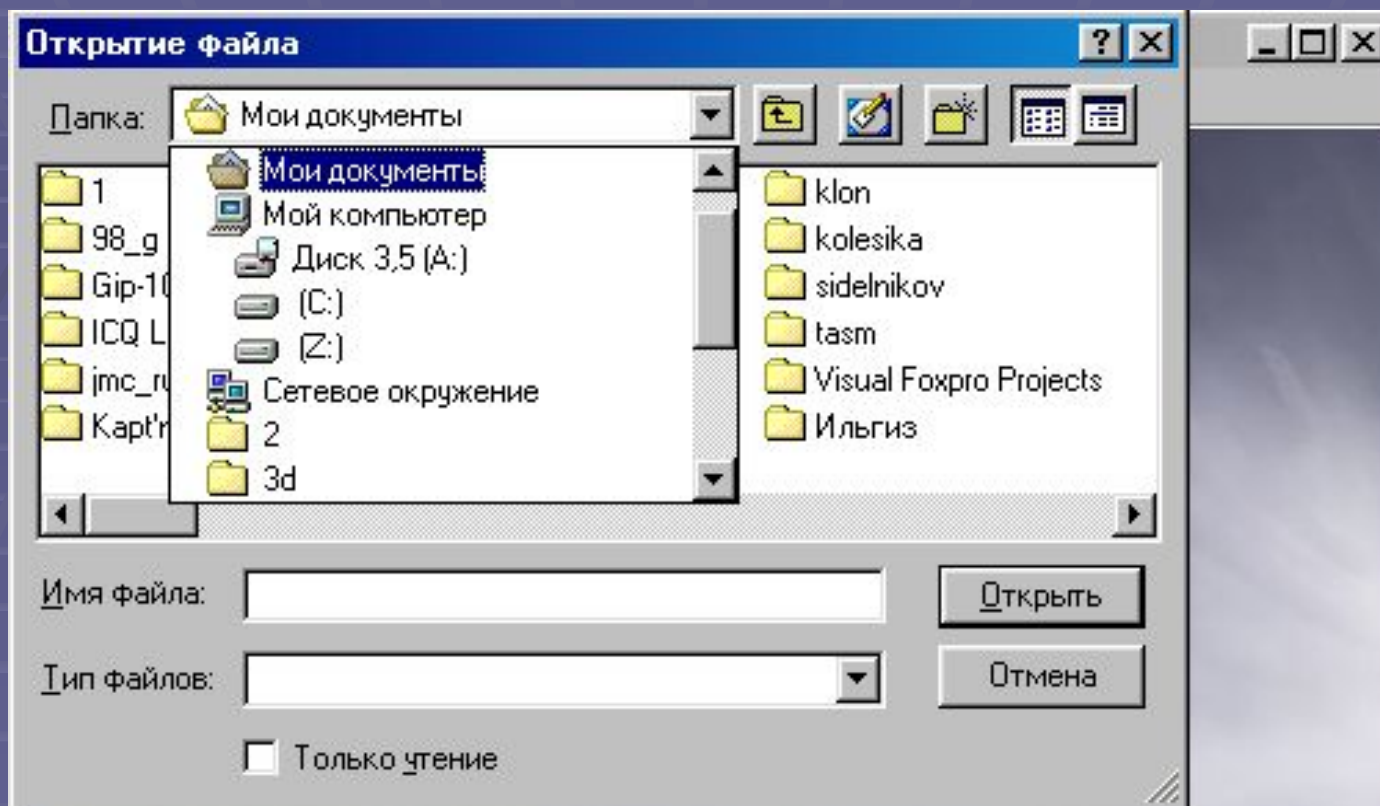
Фрагмент ББК обрабатываемый программой

- 32.97- Вычислительная техника
- 32.973- Компьютеры (электронно-вычислительные машины)
- 32.973.1- Аналоговые компьютеры (электронно-вычислительные машины непрерывного действия)
- 32.973.2- Цифровые компьютеры (электронно-вычислительные машины дискретного действия)
- 32.973.202- Компьютерные системы и сети
- 32.973.202-018.2- Программное обеспечение компьютерных сетей
- 32.973.202-04- Коммуникационное оборудование
- 32.973.26- Микро-ЭВМ. Персональные компьютеры
- 32.973.26-018- Программирование для микро-ЭВМ
- 32.973.26-018.1- Языки программирования
- 32.973.26-018.2- Программное обеспечение
- 32.973.26-02- Проектирование. Архитектура микро-ЭВМ
- 32.973.26-04- Элементы и узлы микро-ЭВМ

Главное меню



Открытие текста



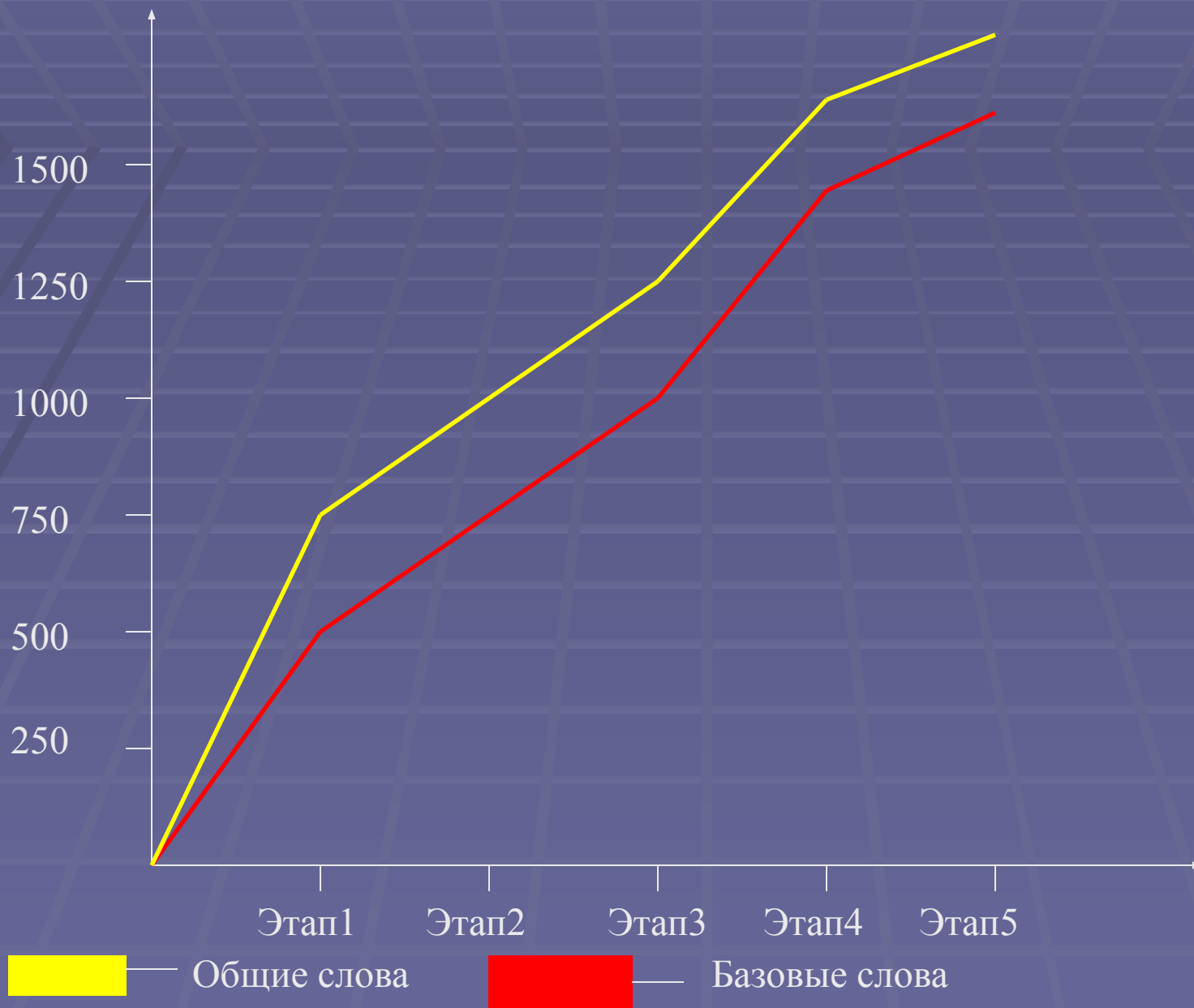
Результат

The screenshot shows a window titled "Автоматическое индексирование текста" (Automatic text indexing). The window has a menu bar with "Файл" (File) and "Справка" (Help). The main area is divided into sections. At the top, there is a label "Индексируемый текст" (Text to be indexed) above a text box containing the snippet: "Проектирование вычислительных систем – создание комплекта конструкторской и программной документации, необходимой для производства и эксплуатации....". Below this, there are two sections for search results. The first section is labeled "Рекомендуемый индекс" (Recommended index) and "Совпадений" (Matches). It shows an index of "681.3.07", a title "аспекты запоминающих устройств (ЗУ) и ввода- вывода.", and 30 matches. The second section is labeled "Ближайший индекс" (Closest index) and "Совпадений" (Matches). It shows an index of "681.322", a title "автоматические цифровые вычислительные машины (ЦВМ).", and 8 matches.

Рекомендуемый индекс	Название	Совпадений
681.3.07	аспекты запоминающих устройств (ЗУ) и ввода- вывода.	30

Ближайший индекс	Название	Совпадений
681.322	автоматические цифровые вычислительные машины (ЦВМ).	8

Диаграмма прироста слов



Корреляционная связь

	32.97	32.973	32.973.1	32.973.2	32.973.2 02	32.973.2 02-018. 2	32.973.2 02-04	32.973.2 6	32.973.2 6-018	32.973.2 6-018.1	32.973.2 6-018.2	32.973.2 6-02	32.973.2 6-04
681.3.01	1	2	1	0	9	8	7	3	1	1	4	3	1
681.3.02	5	3	4	6	7	13	6	2	4	1	4	1	5
681.3.04	2	3	4	17	15	11	10	11	7	6	4	1	2
681.3.06	5	2	3	9	6	7	6	16	8	1	2	3	2
681.3.07	2	4	4	3	9	15	18	10	6	1	3	4	2
681.3.08	1	3	4	9	6	14	9	6	8	1	4	3	2
681.31	0	1	2	8	6	3	8	4	7	3	2	1	4

Названия УДК и ББК при которых заметны сгущения.

- **681.3.04** - представление данных: числами, алфавитом, кодами
32.973.2- Цифровые компьютеры (электронно-вычислительные машины дискретного действия) **(17)**
- **681.3.04** - представление данных: числами, алфавитом, кодами
32.973.202- Компьютерные системы и сети **(15)**
- **681.3.06** - математическое обеспечение (программирование).
Программы ЦВМ.
32.973.26- Микро-ЭВМ. Персональные компьютеры **(16)**
- **681.3.07** - аспекты запоминающих устройств (ЗУ) и ввода-вывода.
32.973.202-04- Коммуникационное оборудование **(18)**

Дальнейшее развитие программы

Дальнейшее развитие будет направлено на то, что бы программа обрабатывала все разделы УДК и ББК. Затем она будет загружена в один из серверов, и поисковые системы будут пользоваться как дескрипторным так и классификационным способом, что поможет пользователю быстрее найти интересующий его текст.