

Методика преподавания информационных дисциплин для студентов гуманитарных направлений в КРСУ.

Евтушенко Анна Ивановна, Фейгин Ян Давыдович

В статье рассмотрены основные проблемы преподавания компьютерных дисциплин для студентов гуманитарных направлений (прежде всего лингвистов и филологов) в условиях КРСУ. Задачи ИТ-технологий в гуманитарной области сегодня – это TextMining – интеллектуальный анализ текстов при помощи компьютера. В статье авторы приводят набор тем, которые доступны для изучения студентами-гуманитариями на ГФ и ФМО КРСУ с использованием только свободно распространяемого ПО. В этот перечень вошли: ментальные карты, сканирование и распознавание документов, работа с pdf-форматом, автоматический перевод и реферирование, фоносемантика и контент-анализ. Авторы описывают методику преподавания этих тем и указывают конкретный набор программ, позволяющих эту методику реализовать.

Ключевые слова: компьютерный анализ текстов, ИТ для гуманитариев.

Methods of computer sciences teaching for KRSU students of humanities.

Evtushenko, Anna Ivanovna, Feigin, Yan Davydovich

The article describes the main challenges of teaching computer science for students of humanities (mainly linguists and philologists) in conditions of KRSU. The article discusses the main problems of teaching computer disciplines for students of humanities (primarily linguists and philologists) in KRSU. The tasks of IT technologies in the humanitarian field today are TextMining - intellectual analysis of texts using a computer. In the article, the authors cite a set of topics that are available for study by humanities students at the GD and IRD KRSU using only free software. This list includes: mental maps, scanning and

recognition of documents, working with pdf-format, automatic translation and referencing, phonosemantic and content analysis. The authors describe the methods of teaching these topics and indicate a specific set of programs that allow this method to be implemented.

Key words: Text mining, IT for humanities

КРСУ студенттер гуманитардык үчүн маалымат сабактарды окутуунун методикасы.

Евтүшенко Анна Ивановна, Фейгин Ян Давыдовыч

Статьяда компьютердик окуу боюнча сабак алыпаруу көйгөйлүү маселелери каралган. Гуманитардык багыттагы студенттер учун. (Өзгөчө лингвистика жана филология багытында билим алгандар учун) КОСУ шартында. IT технологиянын бүгүнкү күндөгү маселеси - бул TextMining - текстерди интеллектуалдык талдоо компьютердин жардамы менен.. Ошентип статьяда авторлор КОСУнун ЭАМ факультетинде гуманитарий сабактарды окуган студенттерге көп ыкмаларды сунуштайт. Мисалы: сын-карталар, документтерди сканер кылмай жана аны аныктамай, pdf-формат менен ишти жүргүзүү, бат котормо жана контент-анализ. Авторлор ушул окуу ыкмаларды көрсөтүп негизги программаларды сунуш берип жатышат.

Keywords TextMining, гуманитардык үчүн IT

Сегодня информационные технологии входят в общекультурные компетенции студентов любых специальностей и направлений, в т.ч. и неспециалистов в области ИТ. Помимо общеобразовательного предмета Информатика, где студенты изучают такие базовые для информационных технологий разделы как основы ПК и ОС Windows, работа с электронными документами в текстовом и табличном процессорах Word и Excel, создание презентаций в Power-Point, работа с базами данных в Access, по многим специальностям выделяются значительные кредитные часы на изучение дополнительных разделов современных компьютерных технологий, непосредственно связанных с будущей профессией выпускников.

Для специалистов в областях естественно-технической и экономической на сегодняшний день разработано много программ, позволяющих решать задачи соответствующей профессиональной направленности. И в первую очередь, это такой мощный инструмент, как табличный процессор MS Excel, который намного выделяется по функциональности и интеллектуальности от своих собратьев по пакету Microsoft Office и на сегодняшний день является одним из наиболее популярных приложений в мире. Изучению только его возможностей (вместе с подключаемыми модулями) можно посвятить несколько семестров. Существует целая линейка программ статистической обработки данных, таких как SPSS, Statistica, программ для реализации компьютерной математики – MathCad, MatLab, для бухгалтерского учета – 1С Предприятие, 1С Бухгалтерия и т.д. Для решения дизайнерских и инженерных задач разработаны такие мощные графические процессоры как PhotoShop, Cdraw, AutoCad, Macromedia Flash и т.д. Но для профессий гуманитарной направленности, и прежде всего филологии, лингвистики, политологии положение дел в области профессиональных компьютерных технологий несколько иное...

Одна из основных задач ИТ технологий в гуманитарной сфере - это автоматическая обработка (распознавание и анализ) текстов и речи. В профессиональной среде даже существует соответствующий термин - **Text Mining** - интеллектуальный анализ текста при помощи компьютера.

Text Mining (одна из разновидностей Data Mining) позволяет обнаружить в «сырых» данных ранее неизвестные связи и практически полезные знания, которые можно интерпретировать и использовать для принятия решений в самых разных сферах деятельности. Сегодня анализ текстов (Text Mining) – это набор лингвистических и статистических техник, а также техник машинного самообучения, которые способны моделировать и структурировать информационный контент и текстовые источники в целях бизнес-аналитики, анализа данных, исследований

различного типа. Возможности Text Mining могут применяться для выявления шаблонов и плагиата в тексте, для автоматического размещения информации по интересующим пользователей профилям, для создания обзоров текстов и документов.

К основным элементам Text Mining относятся:

- классификация
- кластеризация
- построение семантических сетей
- извлечение фактов, понятий
- реферирование и аннотирование
- ответы на запросы
- тематическое индексирование
- поиск по ключевым словам

Также в некоторых случаях набор дополняют средства поддержки и создания таксономии и тезаурусов [1].

Определенные успехи в решении этих задач уже есть, но они заметно скромнее, чем в областях инженерно-технических и экономических. На сегодняшний день не существует единой многофункциональной программы, аналогичной, например, Excel, решающей все (или большинство) задач обработки и анализа текста. Есть определённый набор программ, специализированных по 1-3 пунктам из приведённого выше списка, причем наиболее серьезные разработки относятся к сфере коммерческого ПО, т.е. недоступны (по крайней мере, сегодня) для использования в нашем учебном процессе.

В классических учебниках по информационным технологиям для гуманитариев, например, [2] и [3] есть много ценного теоретического материала, но, к сожалению, практически нет указаний, какими программными средствами это можно реализовать, или же предлагаются ссылки на морально устаревшее программное обеспечение времён MS DOS.

Все вышеперечисленное создает определенные проблемы для преподавания таких дисциплин, как «Информационные технологии в лингвистике», «Информационные технологии в гуманитарном образовании», «Компьютерные методы в религиоведении», «Информационно-эмпирические подходы в политической науке» и др. В результате определённых усилий в течение нескольких лет авторам этой статьи удалось составить некоторый «джентельменский набор» программ, который позволяет ознакомить студентов гуманитарных направлений, и прежде всего, лингвистов и филологов, с возможностями современных компьютерных технологий (СКТ) в сфере их профессиональных интересов, используя только свободно распространяемое ПО.

В перечень тем по вышеуказанным дисциплинам вошли:

1. Ментальные карты
2. Сканирование и распознавание текста
3. Работа с pdf-форматом
4. Автоматический перевод текстов.
5. Создание своего электронного словаря в программе EasyWord
6. Дополнительные возможности MS Word
7. Автоматическое реферирование и аннотирование
8. Фоносемантика
9. Контент-анализ

Ментальные карты и методика их преподавания были подробно рассмотрены в нашей статье [4], где желающие могут ознакомиться с преимуществами ментальных карт для представления информации и их использованием в организации учебного процесса на ФМО КРСУ.

В теме *Сканирование и распознавание текста* студенты знакомятся с приемами сканирования документов на планшетном сканере и с распознаванием текста с помощью программы ABBY FineReader. Рассматриваются режимы автоматического разбиения документа на

области (текст, картинка, таблица) и, при необходимости, ручная корректировка этих областей.

Pdf-формат сегодня один из основных форматов электронного документооборота. В теме *Работа с pdf-форматом* мы используем две программы: PDF-XChange.Viewer.v2.5.319 и FoxitPhantom2.2. Студенты учатся основным приемам редактирования и работе со страницами в pdf – документе, распознавания и добавления текстового слоя и возможностям взаимодействия с документами doc-формата.

В теме *Автоматический перевод текстов* рассматриваются возможности и качество перевода off-line переводчика Prompt, и далее студентам предлагается составить сравнительную таблицу для 5-7 on-line переводчиков, работающих с китайским/английским/немецким языком, отсортировать её по критериям, которые студент считает главными и определить top лучших on-line переводчиков.

Название	Количество языков	Качество перевода (по 5 балльной шкале)	Поставьте свой критерий	Оформление сайта, доп_сервисы

В следующей теме учащиеся работают с программой EasyWord, создавая свой персональный электронный словарь (включая транскрипцию слов) по выбранной из предложенного списка тематике. Однако, создание словаря не является самоцелью. Программа является высокоэффективным средством для изучения иностранных языков т.к. в программе EasyWords существует шесть видов упражнений для заучивания слов:

1. “Карточка” – первое знакомство со словом.
2. “Мозаика” – игровой метод, сопоставление слов с переводом.
3. “Выбор перевода” – вырабатывает ассоциативную связь “слово-перевод”.
4. “Выбор слова” – вырабатывает более сложную ассоциацию “перевод-слово”.
5. “Составление слова” – игровой метод, составление слова из букв, позволяет визуально запомнить написание слова.

6. “Написание” – обеспечивает запоминание правильного написания слова.

Эти режимы работы программы предлагаются студентам к самостоятельному изучению.

В *Дополнительных возможностях MS Word* рассматриваются приемы, позволяющие улучшать внешний вид текста, такие как, например, вставка специальных знаков (неразрывный дефис, неразрывный пробел), возможности вставки фонетических значков (для передачи транскрипции), параметры автозамены, возможности автотекста и тезауруса, технологии статистического анализа, в т.ч. и проверка удобочитаемости текста.

Автоматическое реферирование и аннотирование. Студенты работают с двумя программами: 1) ContetAnalyzer v0.52 (DynCat) - portable-версия, проводит частотный анализ и реферировает html-документы и 2) TextAnalys 2.0 – позволяет строить семантические сети и реферировать документы txt-формата. Обе программы позволяют настраивать объем реферата (в % от исходного текста). Студенты готовят текст на 3-5 страниц в двух форматах, проводят реферирование, оценивают качество рефератов по 5-ти балльной шкале, составляют отчет и делают соответствующие выводы о качестве реферирования по обеим программам.

Для изучения темы *Фоносемантический анализ* используются сайты в Интернете, например, www.analizfamili.ru (студенты проводят анализ своей фамилии и имени), и portable-версия программы Vaal-mini (проект ВААЛ), позволяющая получать фоносемантические значения как отдельных слов, так и текстовых массивов на русском или украинском языке.

Студенты готовят 3-4 документа текстового формата – это могут быть речи, выступления какого-либо политического лидера или общественного деятеля за разные годы. Применяв программу Vaal-mini, определяют эмоциональную составляющую всех текстов (1 этап). Далее, в программе EXCEL (этап 2) студенты вводят в табличном виде данные, полученные на 1-м этапе (см.рис1), строят графики по приведенному образцу (этап 3) и делают выводы по полученным результатам

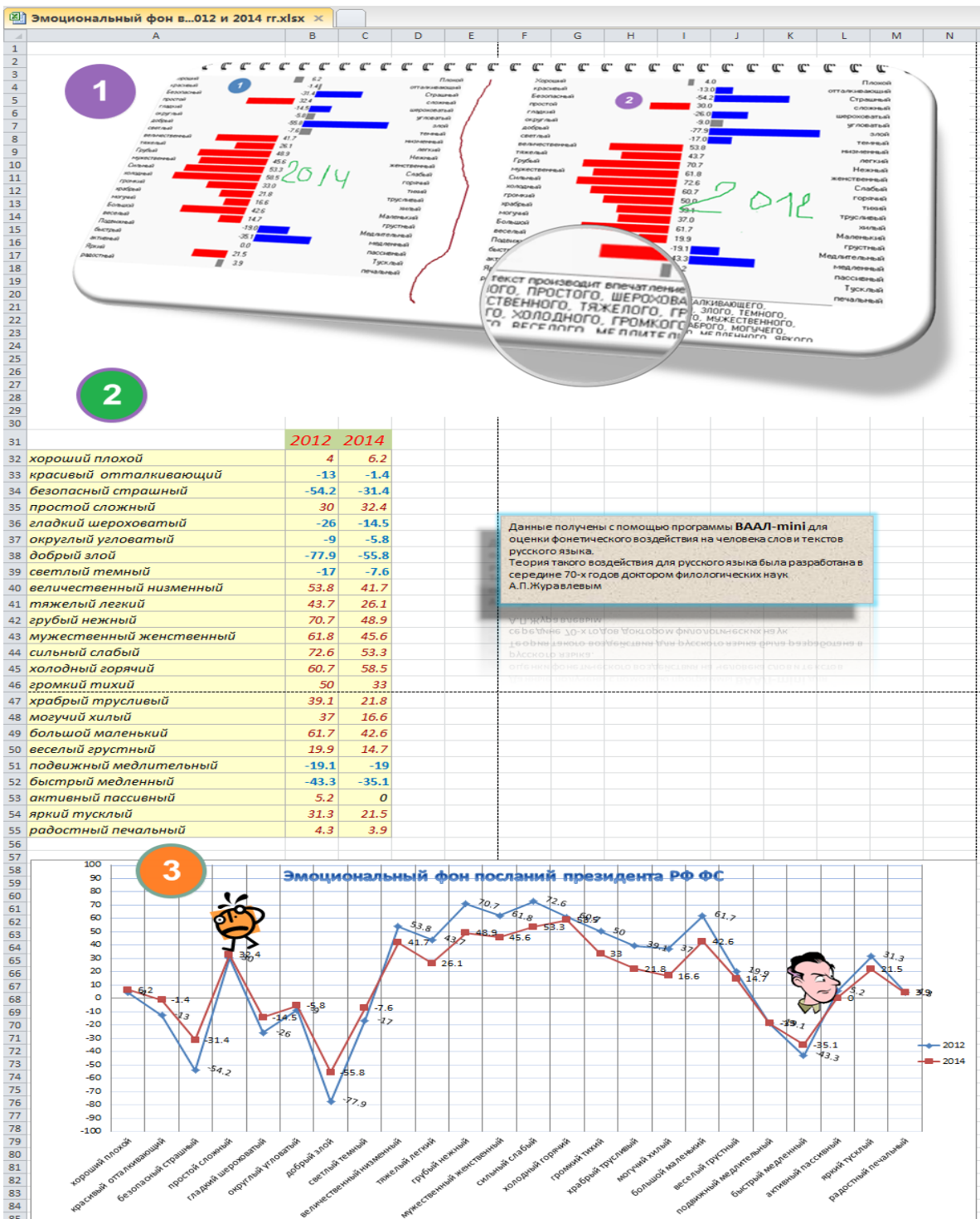


Рис. 1 Этапы проведения фоносемантического анализа

Фоносемантика является частью более общей темы – *Контент-анализа*, где студенты знакомятся с возможностями (в их программной реализации) количественного анализа текстов с целью последующей содержательной интерпретации выявленных числовых закономерностей. Используются уже упомянутые ранее программы ContetAnalyzer v0.52, TextAnalys 2.0 и portable-версия QDA MINER Lite v1.2.

Программа ContetAnalyzer для html-документов в реальном времени позволяет получить различные частотные характеристики слов, словооснов, словосочетаний, на их основе рассчитать вес предложений и составить реферат(см.рис.2)

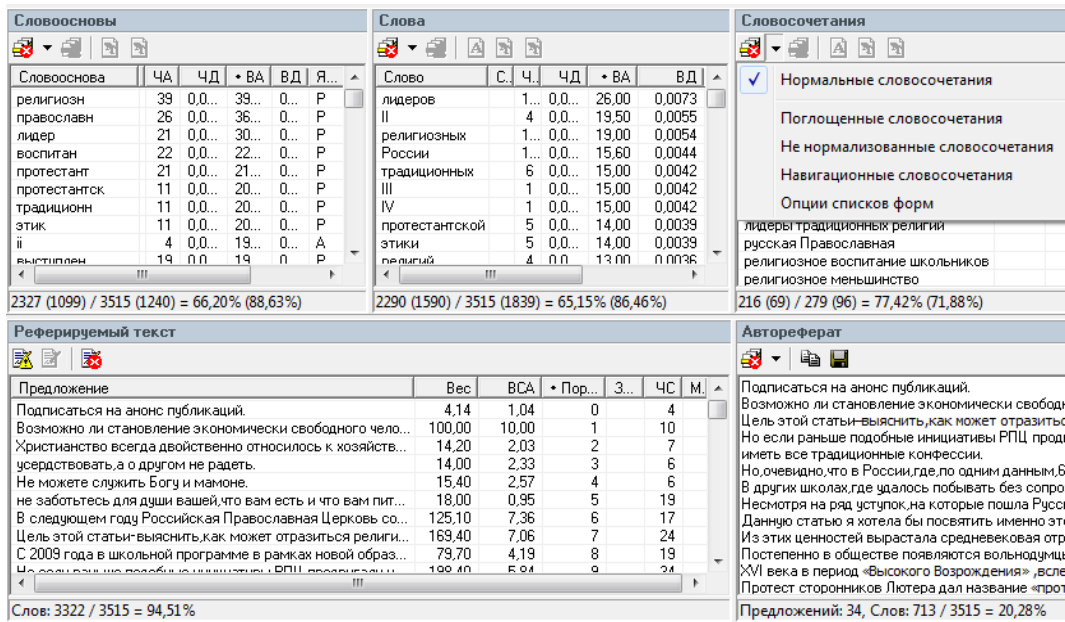


Рис.2 Контент-анализ с помощью программы ContetAnalyzer

Программа TextAnalys 2.0 умеет строить семантическую и тематическую сети для документов в текстовом формате, подсчитывать частотные характеристики слов, веса предложений, строить реферат и проводить ассоциативный поиск (см. рис.3).

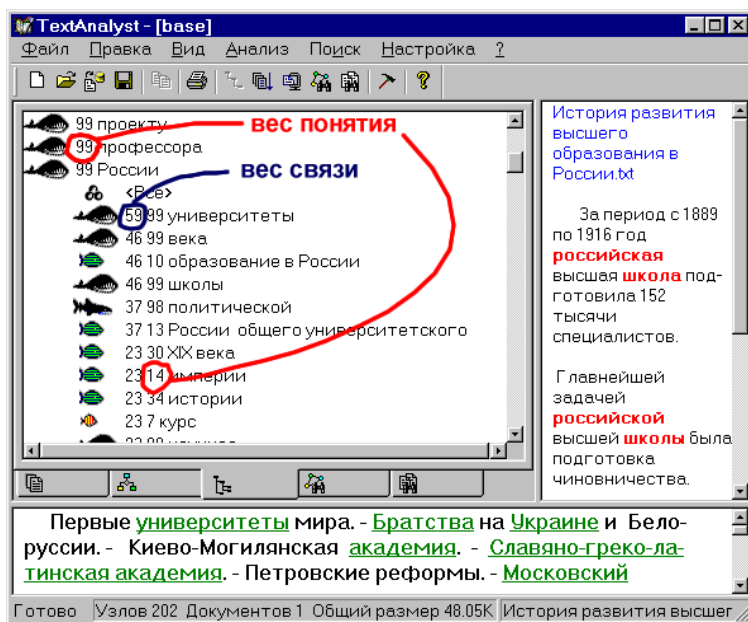


Рис.3. Контент-анализ с помощью программы TextAnalys 2.0

Программа QDA-Lite – программа с ограниченным функционалом , является усечённой версией профессиональной программы QDA, требует ручного кодирования текста (построения тематической иерархии), однако бесплатная и вполне пригодна для изучения основных технологических этапов, используемых в контент-анализе.

Основные возможности:

- Создание категориального аппарата, набора кодов по каждой из категорий
- Количественная оценка каждой из единиц анализа по определенным кодам и категориям
- Отображение результатов квантификации (количественное выражение качественных признаков) в виде различных диаграмм с множеством настроек в EXCEL-стиле.
- Экспорт выборки по категориям и параметрам в файлы (разных форматов) и просмотр отчетов в самой программе

Эта версия программы предназначена для качественного контент-анализа с элементами количественного. Чаще всего, это анализ небольших по объёму массивов материала.

В заключение статьи стоит сказать о том, что авторы выражают уверенность, что рассмотренные выше темы и методики их преподавания могут в достаточной степени познакомить студентов-гуманитариев с основами такой перспективной и активно развиваемой сегодня ИТ отраслью как TextMining и послужить базой для их дальнейших профессиональных достижений.

Литература:

1. Кутукова Е.С. - Технология Text mining, <https://docplayer.ru/37900513-Tehnologiya-text-mining-gosudarstvennyu-universitet-upravleniya.html>.
2. Зубов А.В., Зубова И.И. - Информационные технологии в лингвистике М.: Издательский центр «Академия», 2004. - 208 с.

3. Белоногов Г.Г. Компьютерная лингвистика и перспективные информационные технологии, Издательство: М.: Русский мир, 2004, - 248 с.
4. Евтушенко А.И, Фейгин Я.Д. Использование ментальных карт в учебном процессе на ФМО КРСУ, «Вестник КРСУ», 2015, т15, №3, стр. 97-100

Евтушенко А.И., канд. техн. наук, доцент кафедры Информатика
Кыргызско-Российский Славянский Университет, г.Бишкек
anna_evt@mail.ru, 434388, 0555442264

Фейгин Я.Д. ст. преподаватель кафедры Информатика, КРСУ
feigin_y@yahoo.com 434388, 0772841578