Создание каталога электронных лингвистических ресурсов

1. Компьютерная обработка и анализ текста.

• - Russian Morphological Dictionary

(http://www.solarix.ru/grammatical-dictionary-api-en.shtml)

работает с входным ASCII-текстом. Используется морфологический словарь А.Зализняка, включающий 120.000 слов. Реализована на SWI-Prolog для Windows.

Основан на сайте <http://www.ruscorpora.ru>. Программа быстро и с опорой на авторитет указанного словаря определяет грамматические признаки слов. При обращении к текстам социолектной принадлежности это может обеспечить доказательную атрибуцию морфов, используемых в речи пользователей социальных сетей. С другой стороны, необходимо иметь в виду проблему ограниченности словника словаря А.Зализняка, в котором отсутствуют имена собственные, некоторые неологизмы последнего времени, сравнительные формы вроде постарше, наречия вида по-детски, многие сложные слова, пишущиеся через дефис, многие наречия на -о и -е (последняя задача не снимается введением синкретического класса “наречие/краткая форма прилагательного”). Соответственно, мы прогнозируем затруднения при определении грамматической принадлежности новых для системы литературного языка слов.

• Mystem (авторы - Илья Сегалович, Виталий Титов, компания Yandex) –

это компактный, очень быстрый и бесплатный морфологический парсер русскоязычных текстов, реализованный также на основе словаря А.Зализняка. Доступны для загрузки версии для Windows и Linux. Работает как консольное приложение и имеет различные режимы представления результатов. В общем, программа Mystem производит морфологический анализ литературного нормативного текста на русском языке. Для слов, отсутствующих в словаре, порождаются гипотезы на основании частотности суффиксов. Следовательно, неологизмы и окказионализмы, появление которых легко прогнозировать в социолекте, не получат при использовании этой программы достаточно аргументированного опознания, однако факт наличия гипотезы, сформированной на основе имеющихся в программе сведений о функционировании литературного языка, не может не порадовать. К сожалению, в подавляющем количестве отзывов, которые оставляют пользователи этой программы, отмечается сложность установки программы и введения нужных параметров исследования.

• Рабочее Место Лингвиста

http://www.aot.ru компания Dialing (Москва)),

предлагает анализ текстов для построения систем автоматического перевода с русского на английский язык (и наоборот). Включает ряд автономных компонентов: • синтаксический анализатор текстов на русском языке; • морфологический анализатор текстов на русском и английском языках; • построение конкордансов для заданной совокупности текстов. Система написана на языке C++ и работает в среде Windows 9x/2000/NT. Программа имеет множество позитивных откликов от исследователей различных научных областей, однако создается впечатление, что проект приостановлен, т.к. по указанному адресу программы нет, не предлагается и никакого альтернативного пути знакомства с продуктом. • Морфологический анализатор (http://www.keva.ru/ling/rus/help.html, автор - С.А.Старостин)

– это онлайн-версия программы морофлогического анализа слов русского/английского языков. Позволяет получить для вводимого слова базовую форму и морфологическую информацию. Программа реализована на основе словарей А.А. Зализняка и В.К. Мюллера (английский язык). В Морфологический анализатор может быть введено любое русское или английское слово в произвольной грамматической форме.

Программой анализа выдаются следующие сведения для русского слова: a) исходная слоформа (по Зализняку); b) словарная информация, то есть морфологический индекс русского слова и имеющиеся комментарии из Грамматического Словаря Зализняка; c) перевод, то есть набор словарных статей из словаря Мюллера, в которых содержится соответствующее русское слово, c готовыми отсылками на соответствующие словарные статьи; d) морфологическая характеристика введенного русского слова.

2. Информационный поиск, машинный перевод, компьютерная лексикография, анализ и синтез текста речи, автоматическое извлечение лингвистических данных и знаний.

Русские системы машинного перевода

На российском рынке представлены два производителя, поставляющих системы МП, — компания ПРОМТ и компания «Арсеналъ». Оба производителя поставляют несколько продуктов данного класса, ориентированных на решение разных задач.

• PROMT Translation Office 2000 (далее — PROMT) ценой 300 долл. — это набор профессиональных инструментов, обеспечивающий перевод с основных европейских языков на русский и обратно. С его помощью можно не только переводить, но и редактировать перевод и работать со словарями всех языковых направлений одновременно. http://softwaretopic.informer.com/promt-translation-office-2000/

В PROMT входят следующие коллекции словарей:

«Легкая промышленность» (180 долл.);

«Тяжелая промышленность» (180 долл.);

«Коммерция» (99 долл.);

«Наука» (120 долл.);

«Техника» (199 долл.).

Для обеспечения высокого качества перевода в системе PROMT предусмотрена возможность настройки на перевод конкретного текста — посредством подключения специализированных предметных словарей, поставляемых отдельно, а также создания собственных пользовательских словарей. Удобным средством настройки системы является также возможность выбора тематики документа: какие словари подключать, какие слова оставить без перевода и как обрабатывать специальные конструкции типа электронного адреса, даты и времени.

Система PROMT включает следующие модули:

PROMT— профессиональная среда для перевода;

Dictionary Editor — средство для пополнения и редактирования словарей систем машинного перевода семейства PROMT;

PROMT Electronic Dictionary — электронный словарь, предоставляющий пользователю широкие возможности доступа к лексико-грамматической информации, собранной в специализированных словарях семейства PROMT. Может быть использован при любой работе с текстами (например, для быстрого получения информации о переводных эквивалентах данного слова или словосочетания);

WebView — браузер, позволяющий получить синхронный перевод HTML-страниц при навигации в Интернете. WebView содержит два окна для отображения HTML-страниц: верхнее отображает исходную страницу, полученную из Интернета, нижнее — ее перевод с сохранением ссылок, картинок, вставленных объектов и т.д. Осуществлять переход по ссылкам можно как в верхнем окне, содержащем исходный текст, так и в нижнем, содержащем перевод;

SmarTool — инструмент, реализующий функции перевода в приложениях Microsoft Office 97 (Word, Excel) и Microsoft Office 2000 (Word, Excel, PowerPoint, FrontPage, Outlook). Меню и инструментальная панель перевода встраиваются во все основные приложения Microsoft Office 2000 и Microsoft Office 97, что позволяет получить перевод открытого документа непосредственно в этих приложениях;

QTrans — программа, предназначенная для быстрого перевода неформатированного текста. С ее помощью можно легко и быстро перевести текст, текстовый файл или буфер обмена (Clipboard). Для улучшения качества перевода можно выбрать подходящую тематику, подключить специализированные словари и зарезервировать слова;

Clipboard Translator — программа, предназначенная для быстрого перевода текста, предварительно скопированного в буфер обмена. Текст может быть скопирован из любого Windows-приложения (Help, Notepad, Word, Word Perfect, PageMaker и др.);

«Интегратор» — средство доступа ко всем приложениям пакета.

Сократ Персональный 4. http:/vsofte.biz/451-sokrat-personalnyy.html

«Сократ Персональный 4.0» (25 долл.) — система автоматизированного перевода с английского на русский и наоборот.

Он может переводить обыкновенные документы, фрагменты текстов из буфера обмена Windows, а также очень удобен для перевода сообщений встроенных справочных систем (HELP) англоязычных программ.

«Сократ Персональный» — наиболее мощный продукт, позволяющий использовать дополнительные тематические словари и создавать свои собственные, расширяя словарный запас..

С помощью словаря вы можете получить перевод искомого слова следующими способами:

набрать слово в поле ввода, расположенном в верхнем правом окне словаря. Перемещение по словарной базе осуществляется по мере ввода букв, до тех пор пока не будет получено максимально возможное совпадение;

вставить слово в поле ввода из буфера обмена. В этом случае будет осуществлен быстрый переход к слову, максимально совпадающему с введенным;

выбрать ранее переведенное слово из окна истории поля ввода, после чего будет осуществлен быстрый переход к тому слову, которое имеет максимально возможное совпадение с введенным;

выделить слово в другом приложении и, удерживая клавишу Shift, щелкнуть по выделению правой кнопкой мыши. Во всплывающем окне появится перевод выделенного слова;

использовать сочетание горячих клавиш, предварительно поместив необходимое слово в буфер обмена.

3. Программы анализа и лингвистической обработки и преобразования текстов

• Link Grammar Parser http://slashzone.ru/parser/

Link Grammar Parser – это синтаксический парсер английского языка. Работает со словарем, включающем около 60000 словарных форм. Реализован на C для Unix. Есть также версия для Windows API32. Имеет консольный интерфейс.

Исходные предложения для разбора могут вводиться вручную с клавиатуры или задаваться в ASCII-файле для пакетной обработки. Программа распространяется бесплатно.

• Проекты Cíbola/Oleada

Автор: Computing Research Laboratory (CLR)

New-Mexico State University, USA.

Проекты Cíbola/Oleada реализуют обширные компьютерные системы лингвистического анализа текстов, представленных в Unicode. Компоненты системы включают средства работы с мультиязыковыми текстами (MUTT), построения конкорданса (XConcord) для текстов на более чем 16 языках, статистического анализа, автоматического перевода, различные словари и тезаурусы. Некоторые версии этих компонентов доступны для бесплатной загрузки после процедуры формальной регистрации. Все компоненты реализованы в среде X11 Window System для SunOs и Solaris.

• Система StarLing https://gamua.com/starling/

СУБД StarLing, позволяющая работать с мультиязычными текстами большой длины, с транскрипционными знаками, с удобным поиском, с анализом и синтезом словоформ по словарю Зализняка, с переводом по словарю Мюллера. Есть функции для сравнительно-исторических исследований (глоттохронология). Для загрузки доступны полные DOS и Windows версии системы. Для обеих версий системы требуется предустановка системных фонтов, также доступных для загрузки (DOS, Windows).

Кроме этого можно загрузить словари Ожегова и Зализняка в DBF-формате.

В режиме on-line на сайте доступна этимологическая база для различных языков.

4. Психолингвистические программы

• ПСИ-Офис версия 2.1 http://psyberia.ru/soft/psyoffice

Система психолингвистического анализа текстов, включающая 3 компоненты:

- программа поиска вложенных слов в тексте, т.е. слов, "спрятанных" внутри и на переходах между словами.

- программа поиска повторяющихся фрагментов текста при анализе "автоматического письма" (такие тексты пишутся с целью анализа текущих подсознательных процессов).

- программа синтеза подсознательного компонента текста.

Поддержка русского (словарь 690 тыс.слов) и английского (словарь 430 тыс.слов) языков. Возможность подключения дополнительных словарей. Работает в Windows 9x/2000/NT. Freeware.

• ВААЛ-2000 http://www.vaal.ru/proekt/vaal2000.php

Психолингвистическая экспертная система лексического и контент-анализа текстов. Основные задачи - прогноз эффекта неосознаваемого воздействия текста на массовую аудиторию, анализ текстов с точки зрения такого воздействия, генерация текста с заданным вектором воздействия, выявление личностно-психологических качеств автора текста. Система реализована в виде набора DLL-библиотек, которые подключаются к текстовому процессору Word для Windows. Успешно используется в ГосДуме, правительстве, крупных банках и PR-компаниях. Для бесплатной загрузки доступна облегченная версия системы ВААЛ-мини.

• ПРИЕМЫ ЖУРНАЛИСТИКИ & PUBLIC RELATIONS, версия 2.03, 2000

Коммерческая программа, представляющая собой консультанта, помогающего Пользователю при написании различных текстов. Пользователь выбирает ряд параметров, характеризующих желаемый результат, а программа выдает ему рекомендации по написанию текста и иллюстрирует их примерами. В основе программы - анализ 12500 текстов мастеров русского языка, начиная от писателей-классиков и заканчивая современными журналистами и политиками. К сожалению в программе отсутствуют возможности анализа написанного Пользователем текста.

Для загрузки доступна демо-версия, являющаяся облегченным вариантом профессиональной версии. Демо-версия также включает руководство пользователя и демонстрационный обучающий ролик. Программа требует для работы русские Windows 9x/2000/NT и Microsoft Word.

5. Генераторы текстов или «говорящие программы»

• Russian Word Constructor(RWC) версия 1.0, 1992 http://www.kirsanov.com/rwc.

Экспериментальная программа для генерации русскоязычных стихоподобных текстов ("инструмент поэта"). Программа способна конструировать русские неологизмы на основе заданного словаря с лексико-статистической информацией. Несмотря на свою "древность", программа интересная и может позабавить своими словесными изобретениями. Прилагается утилита для создания авторских рабочих словарей. Распространяется бесплатно. Есть исходные тексты на языке C. Работает как консольное приложение MS-DOS.

• Scott Pakin's automatic complaint-letter generator http://wwwcsag.ucsd.edu/individual/pakin/complaint

Всемирно известный генератор письменных жалоб Скотта Пейкина. Представляет собой on-line версию программы генерация текста жалобы на заданную персону или организацию. Использует для работы словарь лексической информации, который на сегодняшний день включает более 7000 элементов. Язык реализации - английский.

• Cyrano Server http://www.nando.net/toys/cyrano/version2/compose-cyrano.html

On-line версия программы генерации любовных писем. Пользователь задает стиль сочинения, свое имя, обращение к любимой и некоторые атрибуты, ее характеризующие, и программа ловко сочиняет письмо нужной тональности. По-моему, очень полезная вещь для неспособных в литературном плане любовников. Язык реализации - английский.

6. Системы обработки естественного языка

• Natural Language Projects at ISI https://www.isi.edu/research\_groups/nlg/home

Описание группы проектов по вычислительной лингвистике и машинному переводу, выполняемых в Университете Южной Калифорнии.

Webclopedia - проект, связанный с созданием информационных агентов, способных генерировать ответы пользователю на основе анализа разнородных коллекций ресурсов, доступных в сети.

ONTOSAURUS - это онтологический тезаурус, включающий 90,000 концептуальных сущностей (словарная база взята из проекта WordNet).

SUMMARIST - мультиязычная система реферирования текстовой информации.

ReWrite - система двуязычного перевода, основанная на статистическом анализе пар текстовых корпусов.

• ARM Engine 4.0 http://www.abbyy.ru/products/arme/index.htm

ABBYY Retrieval & Morphology (ARM) Engine - это коммерческая инструментальная система, позволяющая встраивать в разрабатываемые приложения функции полнотекстового поиска и морфологического анализа текстов. Поддерживает практически все европейские языки, включая русский. Реализована в виде API с использованием технологии COM-объектов. Демо-версию можно получить бесплатно.

• Следопыт 3.0 http://www.sledopyt.ru/

Интеллектуальная поисковая система, применимая для смыслового и нечеткого поиска документов на русском и/или английском языках. При поиске учитывается морфология и изменение грамматических форм элементов поискового запроса. Понимает все основные форматы документов для Windows. Цена от $9.9 до $1290 в зависимости от комплектации. Демо-версию можно взять здесь.

7. Коллекции ресурсов.

• Каталог программ по вычислительной лингвистике http://www.sil.org/computing/

Обширный каталог программ по вычислительной лингвистике, разработанных в рамках проекта SIL. Представленные в каталоге программы носят явно академический характер и требуют значительных усилий для изучения и применения. Не рекомендуются неискушенным в академической лингвистике пользователям. Все программы, а их более 60, являются бесплатными и доступны для online загрузки. Есть исходные коды.

• Каталог программ по вычислительной лингвистике в сети Internet http://www.sil.org/linguistics/computing.html

Очень обширная коллекция ссылок на программы чрезвычайно широкой лингвистической направленности в сети Internet. Для всех платформ. Но совсем нет программ, посвященных русскому языку. Представлены как бесплатные так и коммерческие программы.

• TELRI Research Archive of Computational Tools and Resources (TRACTOR) http://www.tractor.de/

Архив лингвистических программ и ресурсов в рамках проекта TELRI. TELRI - это ассоциация исследователей, целью которой является выявление и продвижение лингвистических ресурсов и программ, связанных с обработкой текстовых и речевых данных для языков центральной и западной Европы. Пока архив не очень-то большой.

8. Создание и использование тестовых корпусов, национальные корпусы языков и коллекции текстов.

• Корпусный менеджер Manatee (Bonito)

Поисковая система Manatee (корпусный менеджер) представляет собой программное обеспечение для работы с корпусами текстов. Система Manatee состоит из двух частей: сервера (manateesrv) и графического пользовательского интерфейса (GUI - graphical user interface) Bonito, работающего на стороне клиента, созданного Павлом Рыхли (Pavel Rychly) и группой NLPlab (Natural Language Processing Laboratory) на факультете информатики Университета им. Масарика.

Система Manatee легка и удобна в использовании и позволяет проводить поиск отдельных атрибутов (словоформа, лемма, тэг); использовать регулярные выражения и логические операторы; имеет средства задания структуры (границы предложения и др.); быстро обрабатывает сложные запросы;ь имеет возможность использования шаблонов. Программа сохраняет историю запросов пользователя; имеет возможность просмотра морфологических характеристик словоформы и отображение лемм.

• Корпусный менеджер IMS Corpus Workbench (CQP)

IMS Corpus Workbench – это набор программных средств для работы с большими, лингвистически размеченными массивами текстов (корпусами) Одним из этих средств является IMS Corpus Query Processor (CQP) – специальная поисковая программа для лингвистических исследований [6]. В программе реализованы разные способы задания запросов: простой поиск, подобный использованию Google, задан по умолчанию; поиск точной последовательности с использованием упрощенного языка запросов (кавычки); поиск с использованием всех возможностей CQP. Версии программы CQP, работающей в системе Windows, пока не существует.

• Корпусный менеджер XAIRA Xaira (XML Aware Indexing and Retrieval Architecture)

– это программа поиска и вывода лингвистической информации в больших или малых документах в формате XML. Она является развитием программы SARA, изначально разработанной в рамках проекта British National Corpus (http://www.natcorp.ox.aс.uk). Xaira осуществляет поиск основы по основе слова или конкретного слова, показывает частоту слов, дает возможность вводить ограничения по частоте, использовать правила лемматизации, сравнивать частоту данного слова или леммы с частотами других слов, появляющихся в том же контексте.

Программа осуществялет упорядочение результатов по статистическим данным, отражающим силу коллокационных связей, предлагает варианты запросов, ищет по начальным и конечным XML тэгам, также как и по словам, использует комбинации разных запросов.

Данная система состоит из следующих частей: - программа indexer, которая индексирует отдельные XML файлы; - server, программа, которая осуществляет взаимодействие программы клиента с файлами данных; - программа client, которая осуществляет взаимодействие сервера и пользователя

9. Электронные словари и тезаурусы.

• Словарь языка А.С.Грибоедова http://www.inforeg.org.ru/concord/index.htm

Словарь языка А.С. Грибоедова реализует полное лексикографическое описание творчества великого русского писателя. Словарь охватывает все известные тексты А.С. Грибоедова и включает более 12 тыс. лексем и более 120 тыс. словоупотреблений, снабженных подробной лексикографической информацией. По своей структуре словарь представляет собой алфавитно-частотный конкорданс с грамматической информацией. Каждое словоупотребление описывается своей базовой формой; морфологическими характеристиками; названием произведения и его фрагмента; контекстом; гипертекстовой ссылкой на полный текст произведения.

Очень удобный и мощный аппарат поиска.

• Словарь В.Даля http://vidahl.agava.ru/

On-line версия знаменитого словаря В.Даля "Толковый словарь живого великорусского языка". Оригинальный текст © В.И.Даль, 1880-1882. Адаптация к современному языку и издание © Изд. "Цитадель", г. Москва, 1998 г.

• Лексикон "Дар слова" http://www.russ.ru/antolog/INTELNET/dar0.html

Сетевой проект известного филолога, философа и культуролога Михаила Эпштейна, стартовавший в апреле 2000. Проект связан с изучением феномена однословия, т.е. слова как самоценной сущности. Представляет собой сетевой еженедельник, публикующий эссе по словобразованиям и неологизмам русского языка. Каждую неделю подписчикам высылается одно или несколько новых слов, с дефиницией, толкованиями и примерами употребления. Все публикуемые слова являются уникальными словообразованиями, которых не найти ни в одном словаре русского языка.

10. Генераторы текстов.

• Сервис SeoGenerator.ru

Существует достаточно мало сервисов, позволяющих в режиме онлайн размножить текст. Одним из самых удобных вариантов может стать ресурс SeoGenerator.ru.

Заготовленный шаблон необходимо размещать в специальной форме, как показано на скрине ниже.

Здесь же показано, как выставить нужное количество вариантов текстов, и где находится кнопка для копирования результатов. При нажатии на нее результат будет сохранен в буфер обмена, после чего его останется вставить в любой удобный документ (например, в блокнот или Word).

Примечательно то, что данный ресурс содержит еще и синонимайзер. Чтобы воспользоваться им, необходимо перейти по вкладке «СИНОНИМАЙЗЕР», расположенной в верхней части экрана. Затем нужный текст вставляется в специальную форму, а после нажатия на кнопку «Обработать» вы увидите результат. Теперь при наведении на выделенные слова вы сможете выбрать подходящие синонимы для них.

Не очень удобный синонимайзер, но использовать его можно.

• Сервис LinksFarm.ru

Менее удачный пример бесплатного онлайн-сервиса для размножения статей – это сайт LinksFarm.ru. В отличие от предыдущего примера здесь есть возможность предпросмотра результатов. Это имеет смысл, т.к. размножение текста в данном сервисе занимает гораздо больше времени.

Здесь все аналогично: необходимо добавить подготовленный шаблон в форму, а затем выбрать нужное количество копий. После этого через какое-то время варианты сгенерируются, и вам останется скопировать размноженный текст.

Интерфейс у данного сервиса не очень удобный, гораздо лучше использовать для размножения статей специализированные программы.

11. Программа для построения конкордансов Concordance 2.0.0

Concordance – гибкая система для анализа текстов, которая позволяет получить детальную информацию об электронных текстах и производить глубокий и всесторонний анализ текстов на других языках.

Сайт разработчика: http://www.concordancesoftware.co.uk

С помощью Concordance можно подсчитать количество слов и частотность для списка слов. Данный инструмент широко применяется при анализе текстовых данных, в науках, связанных с языками, лингвистике, для извлечения информации, лексикографии, переводе, а также в ряде других коммерческих областей и научных дисциплин. Concordance может быстро построить соответствия с указанием контекста для каждого из слов, выполнить обработку текстов практически любых размеров.

С помощью Concordance можно одновременно просматривать полный список слов, найденные соответствия и исходный текст, а также просматривать оригинальный текст, просто нажав на любое из слов, после чего будут показаны все появления данного слова в контексте. Возможен анализ текстов из любых других программ Windows посредством использования буфера обмена, либо заданием набора входных файлов. Concordance реализует поиск веб-соответствий, конвертируя после этого результат анализа в HTML файлы, связанные между собой. Наличие списка выбора и «черного списка» позволяет задать слова, которые должны быть включены или исключены из вашего анализа.

Concordance работает почти cо всеми языками, поддерживаемыми системой Windows. Concordance позволяет выбрать вариант сортировки слов: поиск фраз, учет расстановки слов при поиске, а также использовать регулярные выражения для поиска. Просмотреть статистику по тексту можно по типу слов, процентному соотношению, количеству символов и предложений. Полностью поддерживается полноэкранный предпросмотр перед печатью и печать с контролем размера страниц, полей, заголовков, нижних колонтитулов и шрифтов. Concordance, пожалуй, наиболее гибкое и мощное аналитическое программное обеспечение для анализа текстов.

12. Программа снятия омонимии.

• Синтаксический анализатор Диалинг .

В первом из них (ЭТАП) используется «фильтровый метод» синтаксического анализа, при котором сначала строится полный набор допустимых гипотетических синтаксических связей между словами анализируемой фразы, а затем из этого набора удаляются (фильтруются) недопустимые связи (синтагмы). Этот процесс продолжается до тех пор, пока оставшиеся связи не образуют дерево, являющееся искомой синтаксической структурой. Система «Диалинг» основана на комбинировании различных вариантов анализа фразы. Здесь также морфологическая омонимия крайне негативно отражается на скорости работы.

• В качестве примера альтернативного «локального синтаксического» подхода, основанного на использовании всевозможных синтаксических правил типа предложного управления или согласований главного и подчиненного слов, можно привести морфологический анализатор английского языка ENGTWOL [3].

Эта система включает словарь основ объемом 56000 единиц и вручную составленную базу данных, содержащую более тысячи правил, являющихся запретами на появление определенных последовательностей грамматических классов в текстах.

• Другим широко известным вероятностным подходом является алгоритм, основанный на использовании скрытой Марковской модели (Hidden Markov Model (HMM) tagging).

Основная идея алгоритма заключается в том, чтобы для каждого слова, входящего в предложение, выбрать грамматический класс (тэг) таким образом, чтобы максимизировать функцию: P(word|tag)\*P(tag|previous n tags), где P(tag|previous n tags) - условная вероятность (вычисленная по размеченному корпусу), появления данного тэга tag, при условии, что предыдущие n тэгов уже определены. P(word|tag) - условная вероятность (также вычисленная по корпусу) появления в данном месте слова word, при условии, что это слово имеет данный грамматический класс tag [3].

13. Программы аннотирования, реферирования, распознавания, синтеза текстов.

• IBM Intelligent Text Miner http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?appname=iSource&infotype=an&letternum=ENUS298-447&subtype=ca&supplier=897

Технология Text Mining представляет собой одну из разновидностей методов Data Mining и подразумевает процессы извлечения знаний и высококачественной информации из текстовых массивов. Это обычно происходит посредством выявления шаблонов и тенденций с помощью средств статистического изучения шаблонов. Такая технология глубинного анализа текстов способна «просеивать» большие объемы неструктурированной информации и выявлять из них только самое значимое, чтобы человеку не приходилось самому тратить время на добычу ценных знаний «вручную».

Text Mining – это алгоритмическое выявление и обнаружение в «сырых» данных доселе неизвестных корреляций и связей, ранее неизвестных практически полезных знаний, которые можно интерпретировать и использовать для принятия решений в самых разных сферах деятельности человека. Результаты Text Mining могут быть использованы для математического прогнозирования, анализа социальной обстановки и анализа рынков.

• Oracle Context https://docs.oracle.com/cd/B19306\_01/server.102/b14200/functions165.htm

Текст, полученный путем соединения отрывочных фрагментов, лишен гладкости, его трудно читать. Кроме того, источники информации вовсе не всегда являются текстами, ведь необходимо подготавливать аннотации и на видеозаписи, к примеру, спортивных соревнований, или формировать сводные данные по биржевым таблицам. Перечисленные инструменты реферирования рассчитаны на обработку только текстовой информации. И, наконец, они не могут работать сразу с несколькими источниками. Так, скажем, многочисленные ленты новостей в Web сообщают об одних и тех же событиях, и на этот случай мог бы оказаться полезен инструмент, способный выделить общие места и новую информацию.

14. Сетевые ресурсы в обучении языкам.

• How to speak English

Приложение было создано в 2014 году компанией Kaplan. С помощью него вы сможете избавиться от акцента. Это возможно благодаря трем видам упражнения: секции Challenge, разделов Your Sounds и All Sounds. Первая секция построена по принципу «вопрос-ответ», которая позволяет определить уровень владения фонетикой. Your sounds позволит сравнить ваше произношение с академическим. Третья же секция хранит в себе не только международный фонетический алфавит, но и научит вас распознавать речевые ошибки и их возникновения.

• British Council

– это мировая организация с учебными центрами в более 100 странах. Многие учебники издаются под редакцией этой организации. Неудивительно, что одноименное приложение пользуется популярностью.

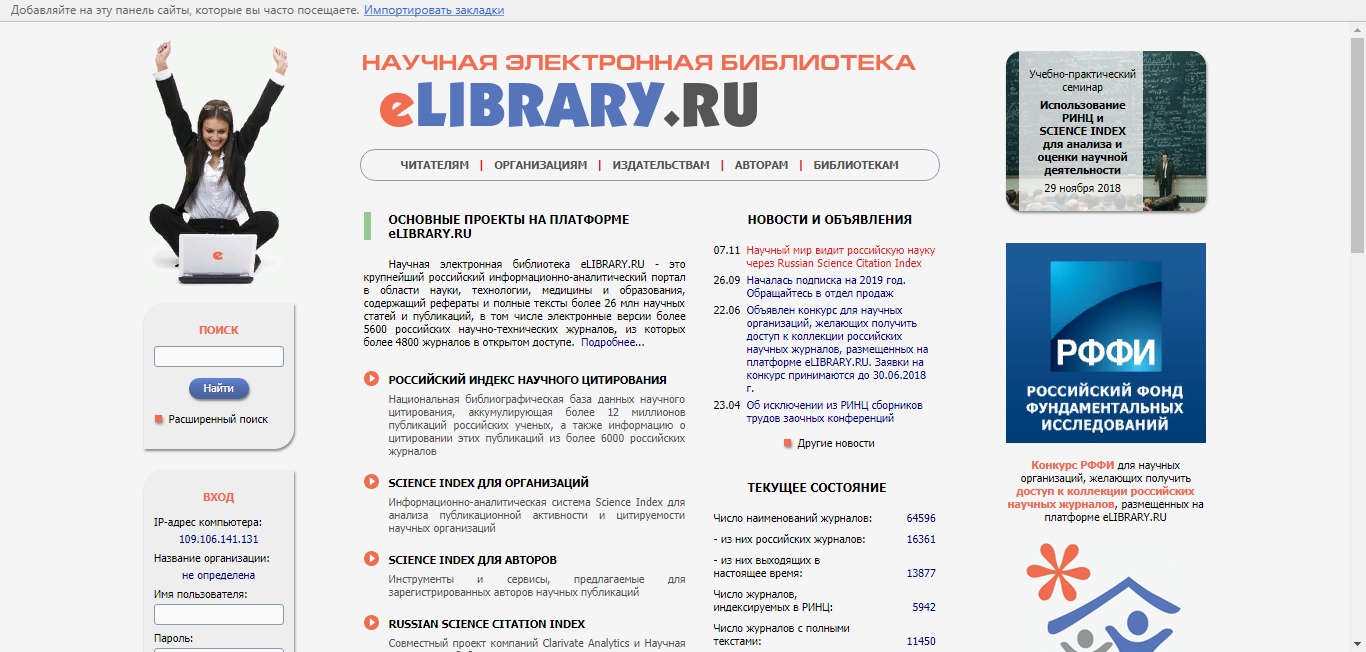
Приложения позволяют изучать английский язык через игры, видео, подкасты. Ученику будут предложены упражнения разных уровней сложностей. С помощью British Council вы узнаете новые слова, изучите грамматику, овладеете правильным произношением.

15. Информационные электронно-образовательные ресурсы.

* Электронная библиотека e-library

https://elibrary.ru/elibrary\_about.asp

eLIBRARY.RU - крупнейшая в России электронная библиотека научных публикаций, обладающая богатыми возможностями поиска и анализа научной информации. Библиотека интегрирована с Российским индексом научного цитирования (РИНЦ) - созданным по заказу Минобрнауки РФ бесплатным общедоступным инструментом измерения публикационной активности ученых и организаций. eLIBRARY.RU и РИНЦ разработаны и поддерживаются компанией "Научная электронная библиотека".



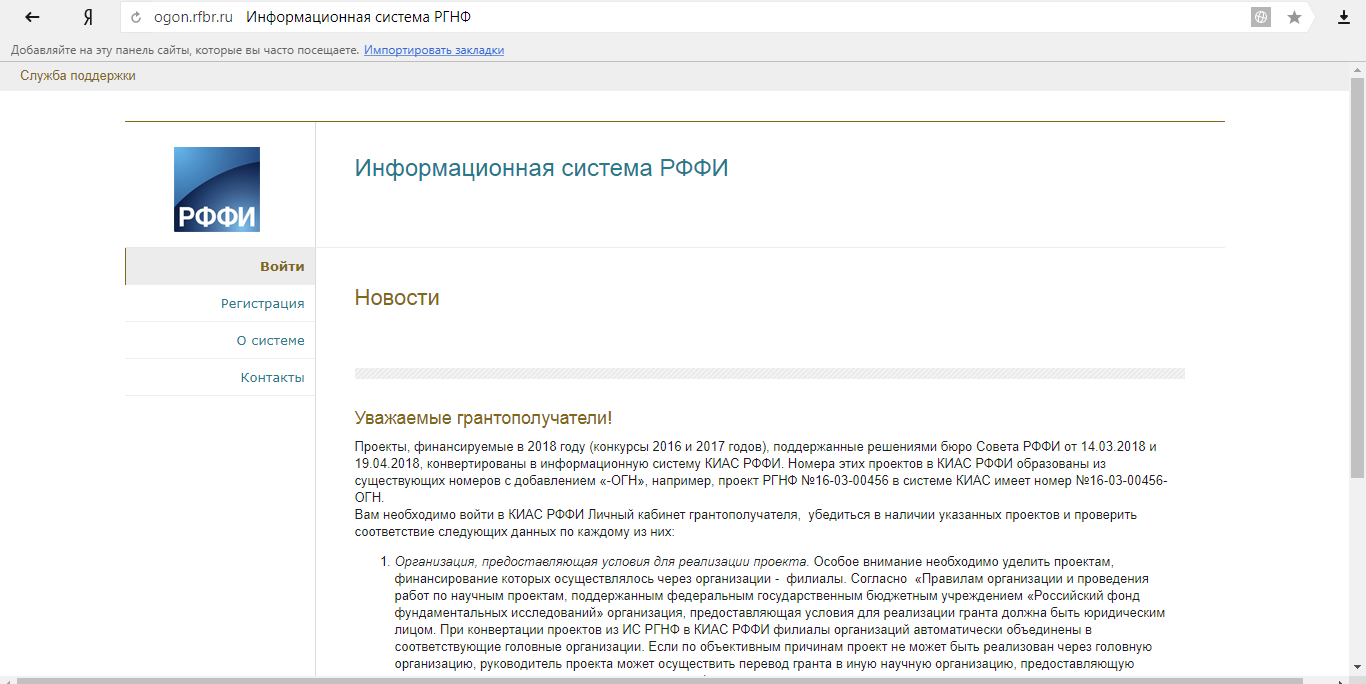
* Информационная система российского гуманитарного научного фонда.

http://ogon.rfbr.ru/

Информационная система Российского гуманитарного научного фонда (ИС РГНФ) представляет собой специальное программное обеспечение, разработанное для потребностей РГНФ.

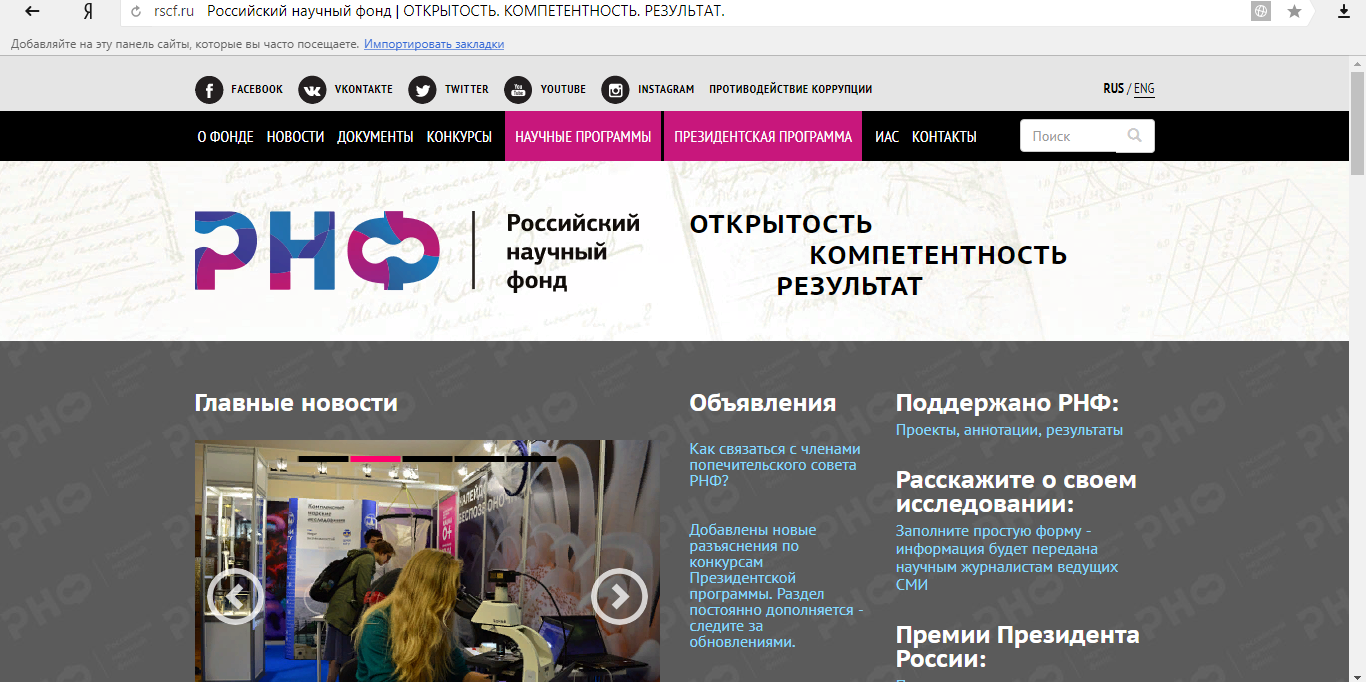
Назначение ИС РГНФ:

* Удаленное оформление заявок на конкурсы РГНФ, подготовка к печати экземпляров зарегистрированных заявок.
* Удаленное оформление отчетов по поддержанным РГНФ проектам, подготовка к печати экземпляров зарегистрированных отчетов.
* Проведение удаленной экспертизы заявок и отчетов, поступивших в РГНФ.
* Подготовка финансовых документов по финансируемым проектам.
* Сбор, автоматизированная обработка и хранение информации о поступающих в РГНФ заявках и отчетах.



* Российский научный фонд.

http://rscf.ru/ru/node/2167

Целью деятельности РНФ является финансовая и организационная поддержка фундаментальных научных исследований и поисковых научных исследований, подготовки научных кадров, развития научных коллективов, занимающих лидирующие позиции в определенной области науки. 

* Индексы научного цитирования.

http://disszakaz.ru/publikatsii/rints-rossiyskiy-indeks-nauchnogo-tsitirovaniya/

Индекс цитирования — реферативная база данных научных публикаций, индексирующая ссылки, указанные в пристатейных списках этих публикаций и предоставляющая количественные показатели этих ссылок (такие как суммарный объем цитирования, индекс Хирша и др.).

Индекс цитирования является одним из самых распространенных наукометрических показателей и применяется (для формальной оценки) в научных и бюрократических кругах многих стран. Альтернативами индексу цитирования являются экспертная оценка и оценка по импакт-фактору научных журналов.



* WoS для оценки научной деятельности.

<https://wos-scopus.com/wp-content/uploads/2015/04/Incites1.pdf>

RSCI является отдельной базой данных, не входящей в основное ядро базы WoS. –Web of Science Core Collection. Однако она полностью интегрирована с поисковой платформой WoS по аналогии с китайским (Chinese Science Citation Database) , латиноамериканским (SciELO Citation Index) и корейским (Korea Citation Index) индексами научного цитирования. Подписчикам WoS уже доступны все материалы RSCI. Полный перечень журналов, включенных в базу RSCI, представлена сайте Научной электронной библиотеки eLIBRARY.RU. 