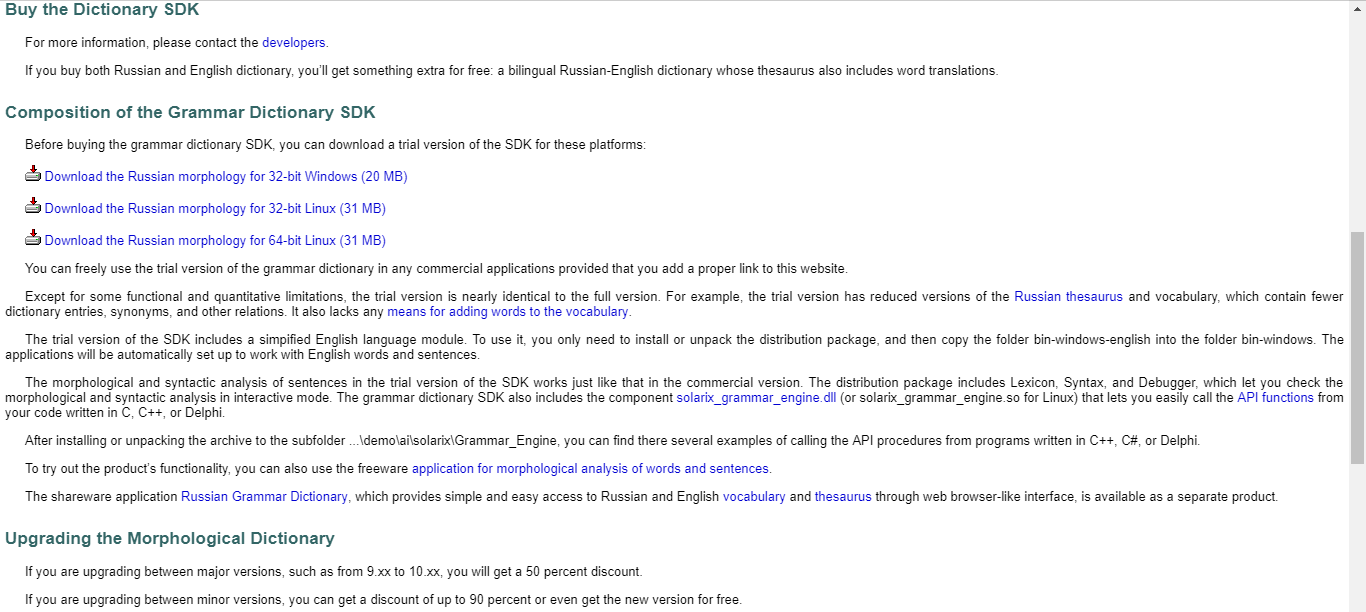
Создание каталога электронных лингвистических ресурсов

1. Компьютерная обработка и анализ текста.

* Russian Morphological Dictionary

<http://www.solarix.ru/grammatical-dictionary-api-en.shtml>

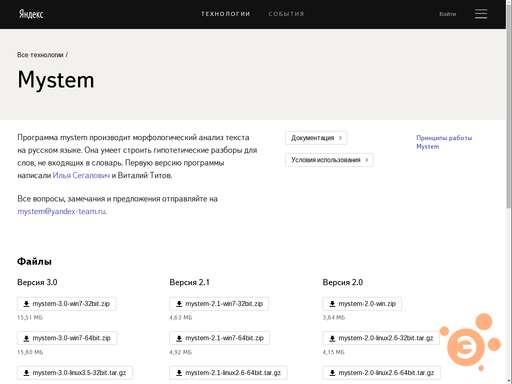


Russian Morphological Dictionary - программа для синтаксического и морфологического анализа русскоязычных текстов.

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

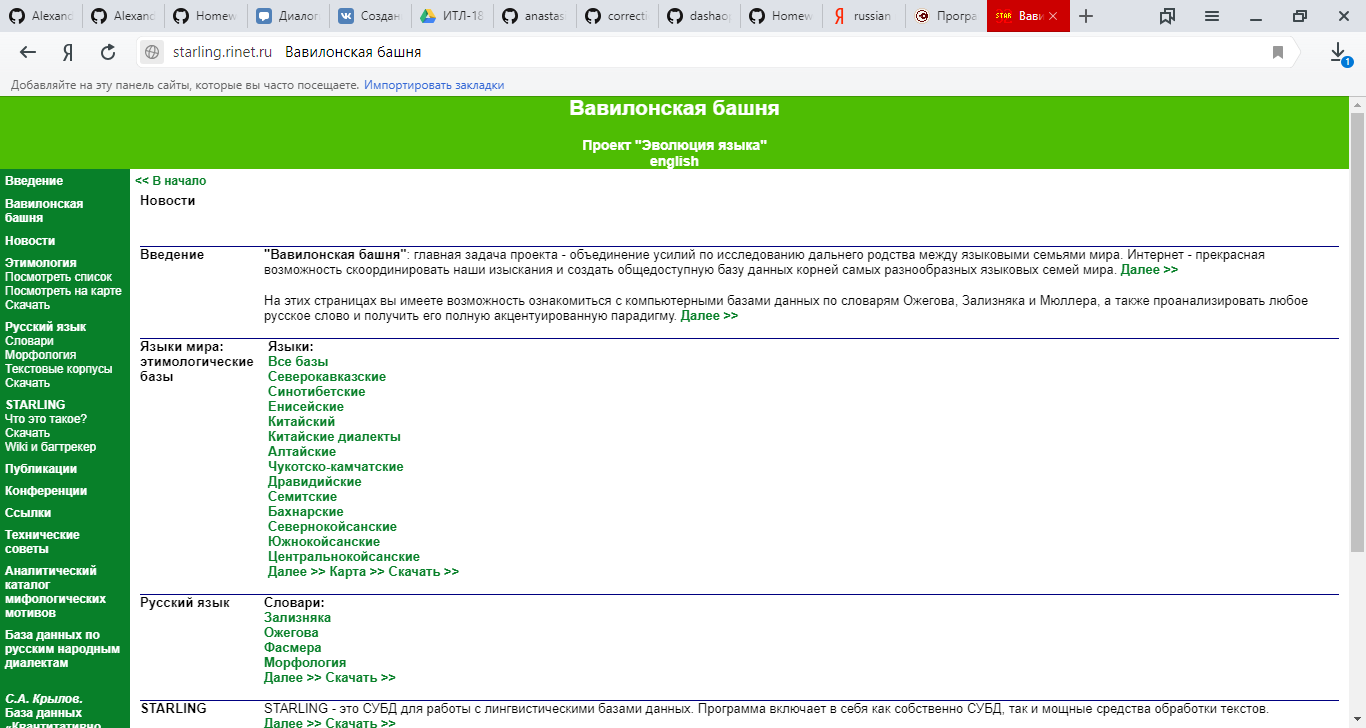
это компактный, очень быстрый и бесплатный морфологический парсер русскоязычных текстов, реализованный также на основе словаря А.Зализняка.

<https://nlpub.ru/Mystem>



* Система StarLing

<http://starling.rinet.ru/index2.php?lan=ru>



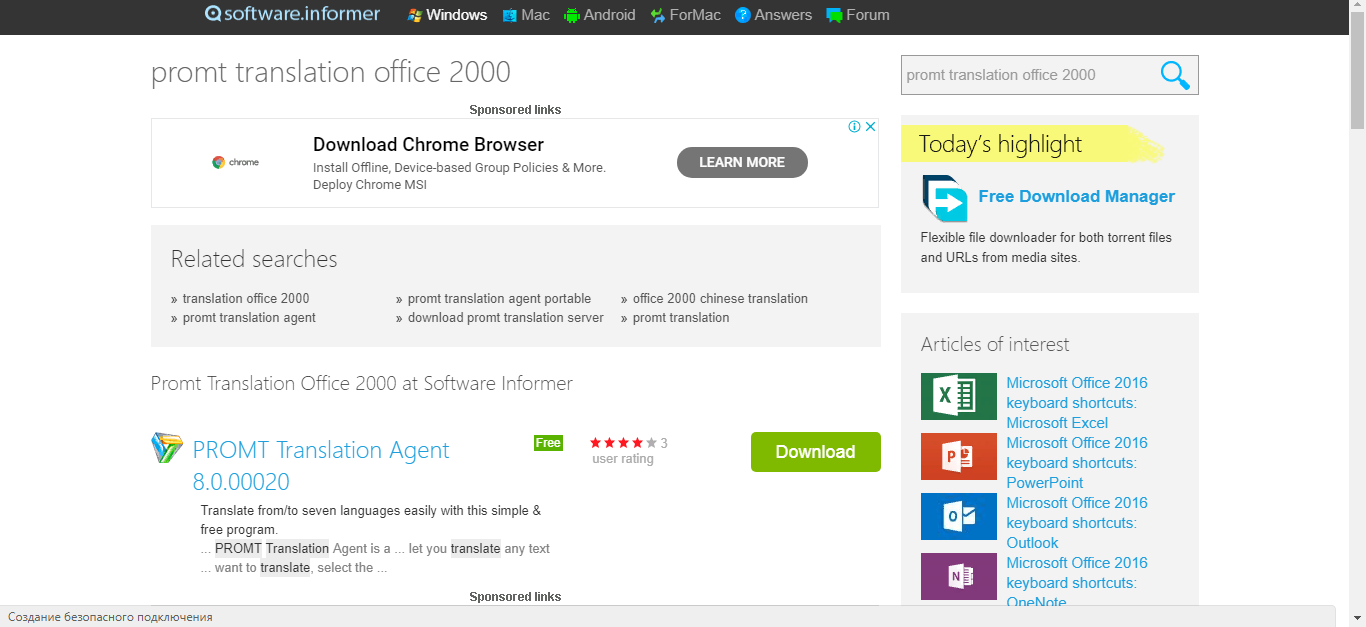
Морфологический анализатор С.А.Старостина для слов русского/английского языков на основе словарей Зализняка (рус.яз.) и Мюллера (англ.яз.). Позволяет получить для вводимого слова базовую форму и морфологическую информацию.

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

Русские системы машинного перевода

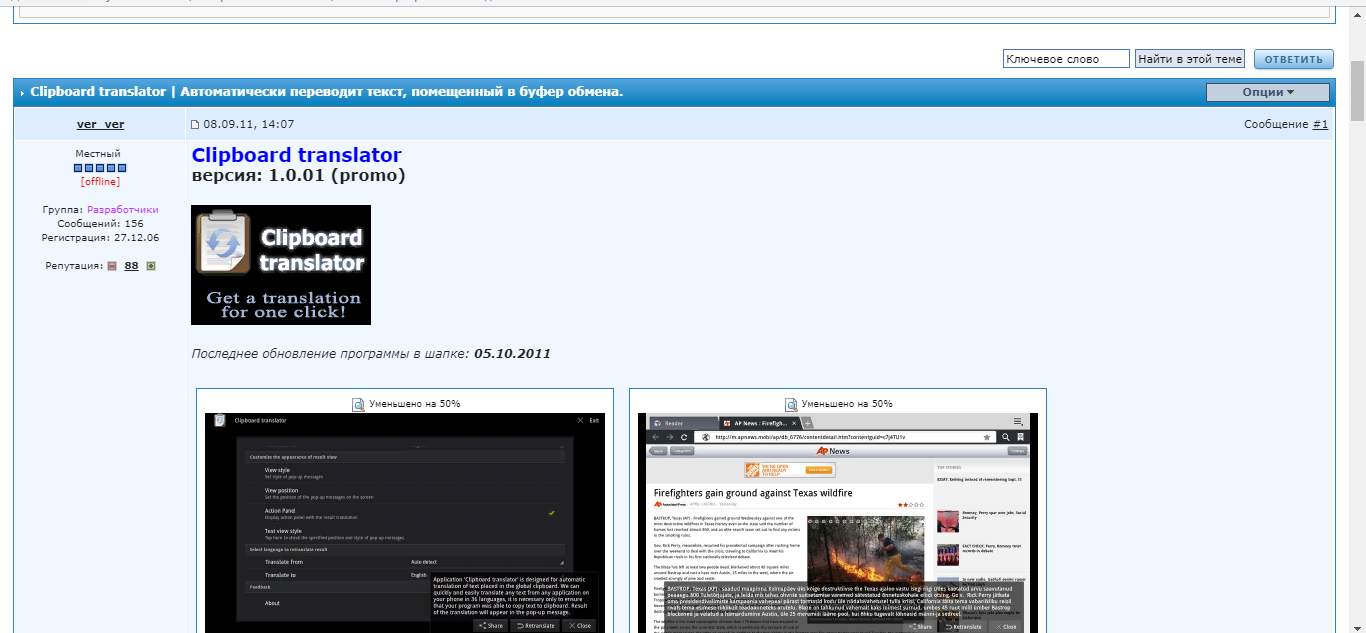
* PROMT Translation Office 2000 (далее — PROMT) ценой 300 долл. — это набор профессиональных инструментов, обеспечивающий перевод с основных европейских языков на русский и обратно. С его помощью можно не только переводить, но и редактировать перевод и работать со словарями всех языковых направлений одновременно.

<http://softwaretopic.informer.com/promt-translation-office-2000/>



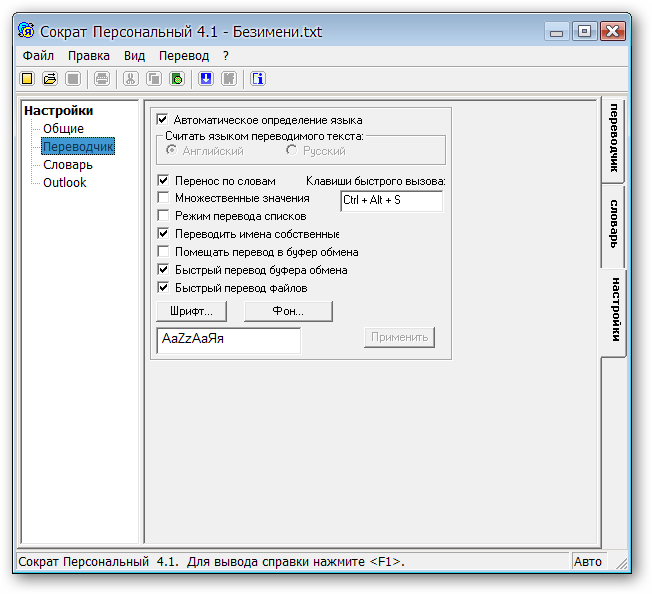
* Clipboard Translator — программа, предназначенная для быстрого перевода текста, предварительно скопированного в буфер обмена. Текст может быть скопирован из любого Windows-приложения (Help, Notepad, Word, Word Perfect, PageMaker и др.);

<http://4pda.ru/forum/index.php?showtopic=265824>



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

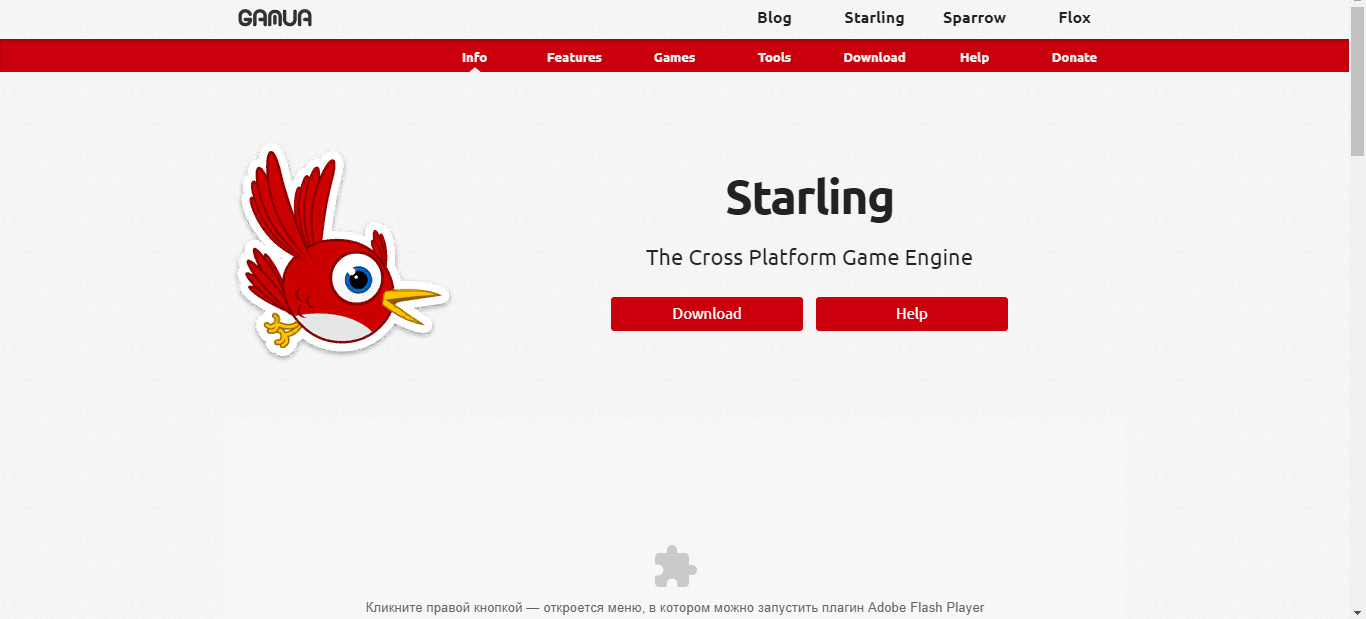
<https://soft.sibnet.ru/soft/5676-socrat-personal-v4-1>



3. Программы анализа и лингвистической обработки и преобразования текстов

* Система StarLing

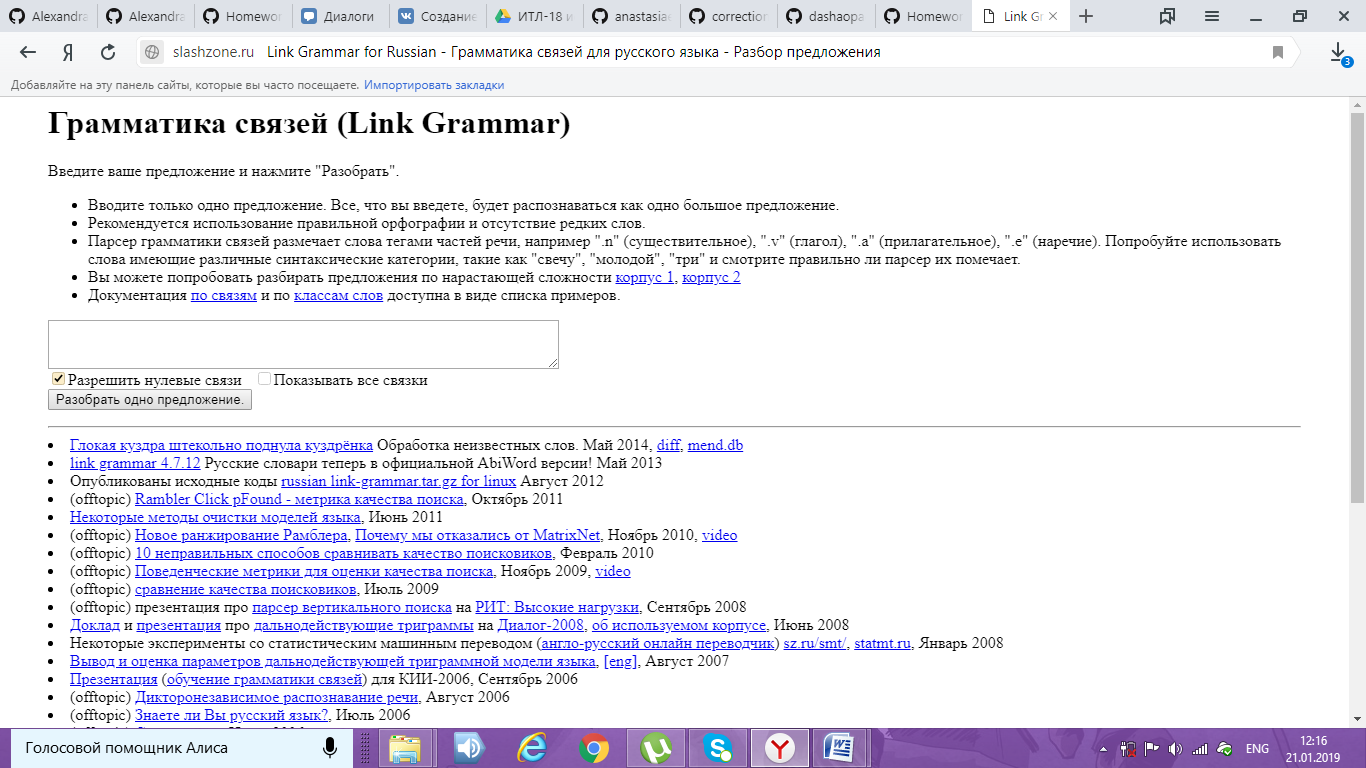
<https://gamua.com/starling/>



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).

* Link Grammar Parser

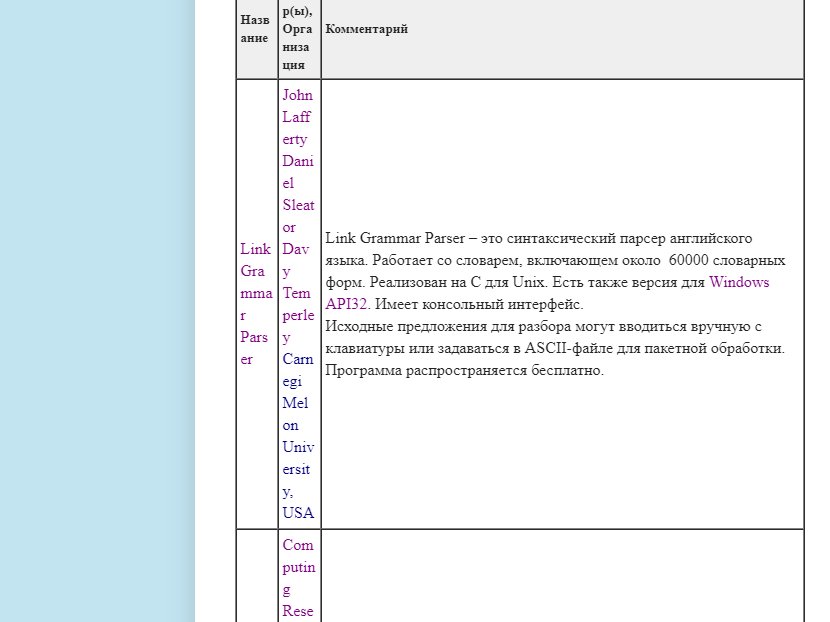
<http://slashzone.ru/parser/>



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.

* Проекты Cíbola/Oleada

<http://redirect.viglink.com/?format=go&jsonp=vglnk_154805873873810&key=172579b97fa4d5e8c1a3c2918a03e499&libId=jr6276mj01012xfz000DA9y8hv6cp&loc=http%3A%2F%2Ftwinpeppers.blogspot.com%2F2012%2F08%2Fc-httpwwwrvbrusoftcatalogueindexhtml.html&v=1&out=http%3A%2F%2Fwww.rvb.ru%2Fsoft%2Fcatalogue%2Findex.html&ref=https%3A%2F%2Fyandex.ru%2F&title=Twinpeppers%3A%20Каталог%20лингвистических%20программ%20и%20ресурсов%20в%20Cети&txt=http%3A%2F%2Fwww.rvb.ru%2Fsoft%2Fcatalogue%2Findex.html>

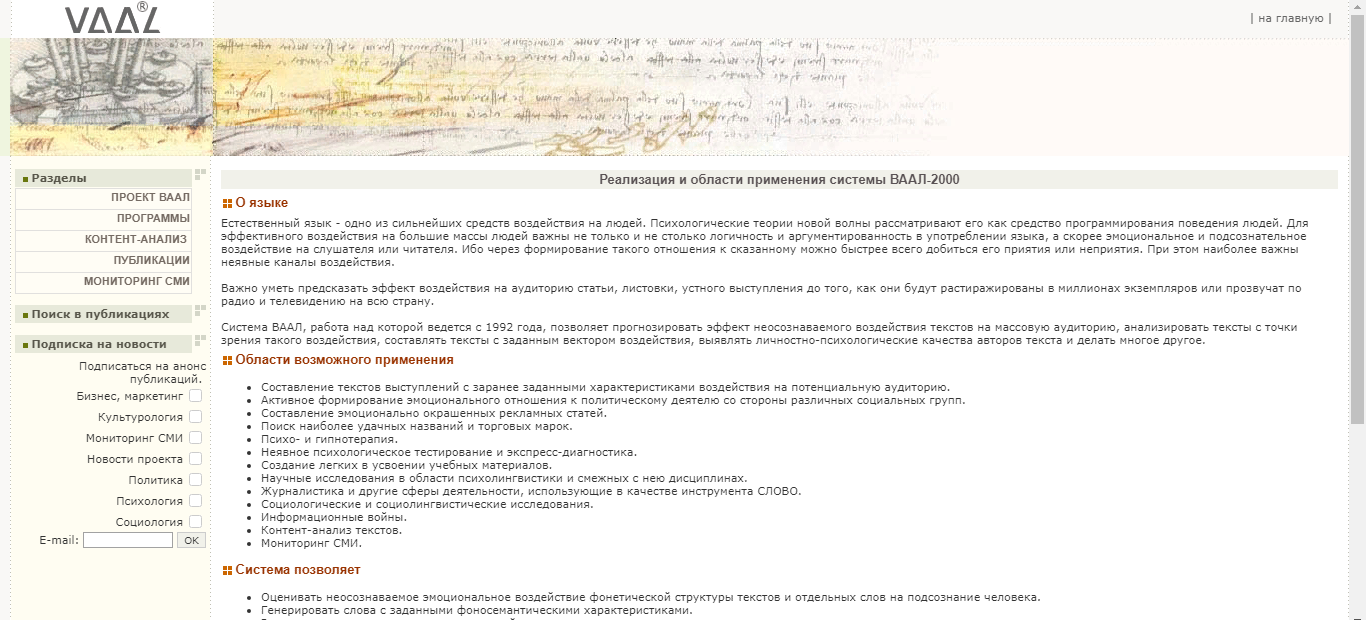


Проекты Cíbola/Oleada реализуют обширные компьютерные системы лингвистического анализа текстов, представленных в Unicode. Компоненты системы включают средства работы с мультиязыковыми текстами (MUTT), построения конкорданса (XConcord) для текстов на более чем 16 языках, статистического анализа, автоматического перевода, различные словари и тезаурусы.

4. Психолингвистические программы

* 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. Для бесплатной загрузки доступна облегченная версия системы ВААЛ-мини.

* ПСИ-Офис версия 2.1

<http://psyberia.ru/soft/psyoffice>



Система психолингвистического анализа текстов. Поддержка русского (словарь 690 тыс.слов) и английского (словарь 430 тыс.слов) языков. Возможность подключения дополнительных словарей.

* ПРИЕМЫ ЖУРНАЛИСТИКИ & PUBLIC RELATIONS, версия 2.03, 2000

<http://kazus.ru/programs/download/5698/>

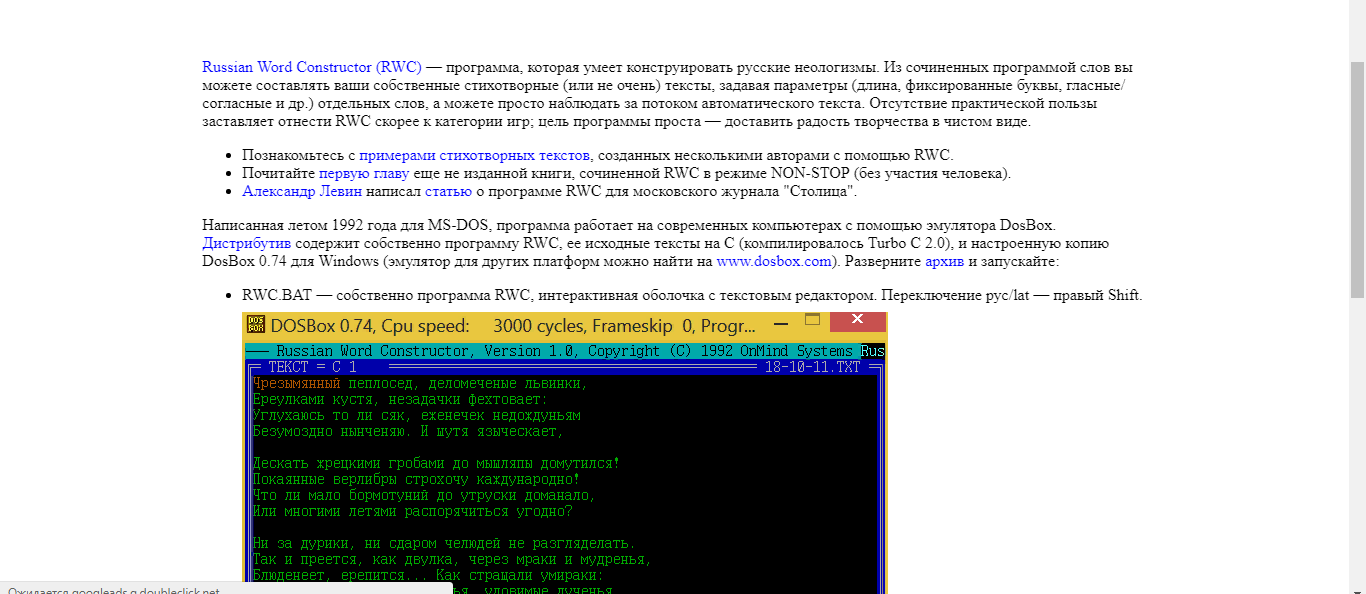


Коммерческая программа, представляющая собой консультанта, помогающего Пользователю при написании различных текстов. Пользователь выбирает ряд параметров, характеризующих желаемый результат, а программа выдает ему рекомендации по написанию текста и иллюстрирует их примерами.

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.

* 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.

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

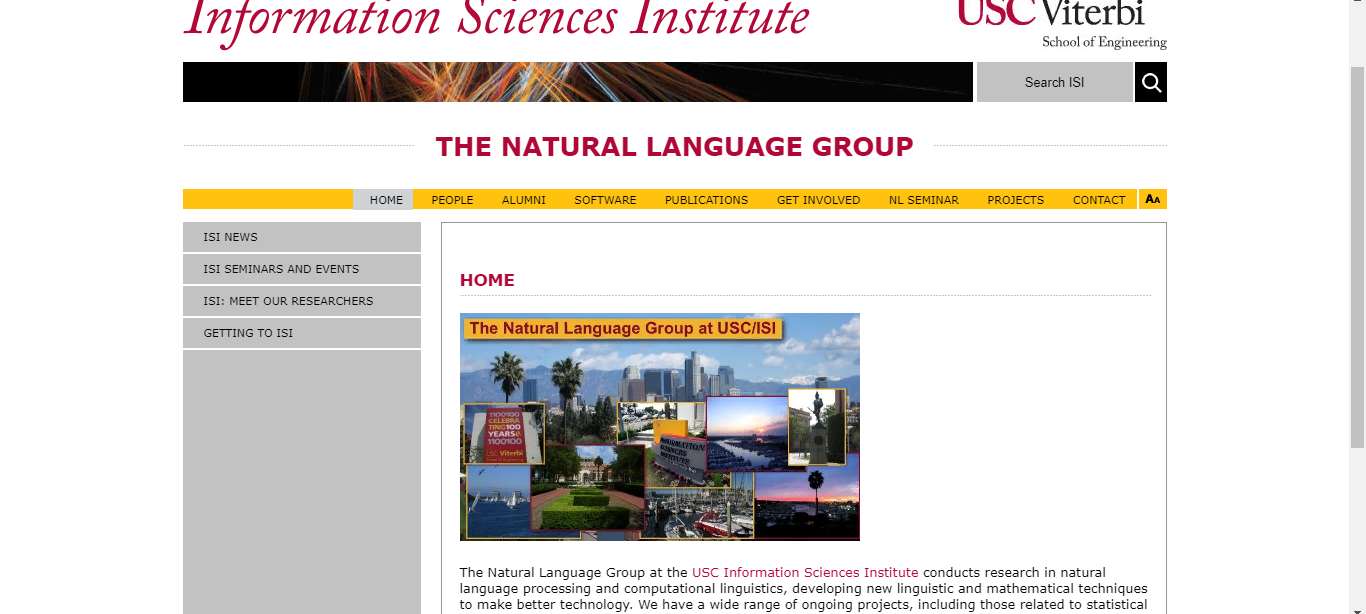


Всемирно известный генератор письменных жалоб Скотта Пейкина. Представляет собой on-line версию программы генерация текста жалобы на заданную персону или организацию. Использует для работы словарь лексической информации, который на сегодняшний день включает более 7000 элементов. Язык реализации - английский.

6. Системы обработки естественного языка

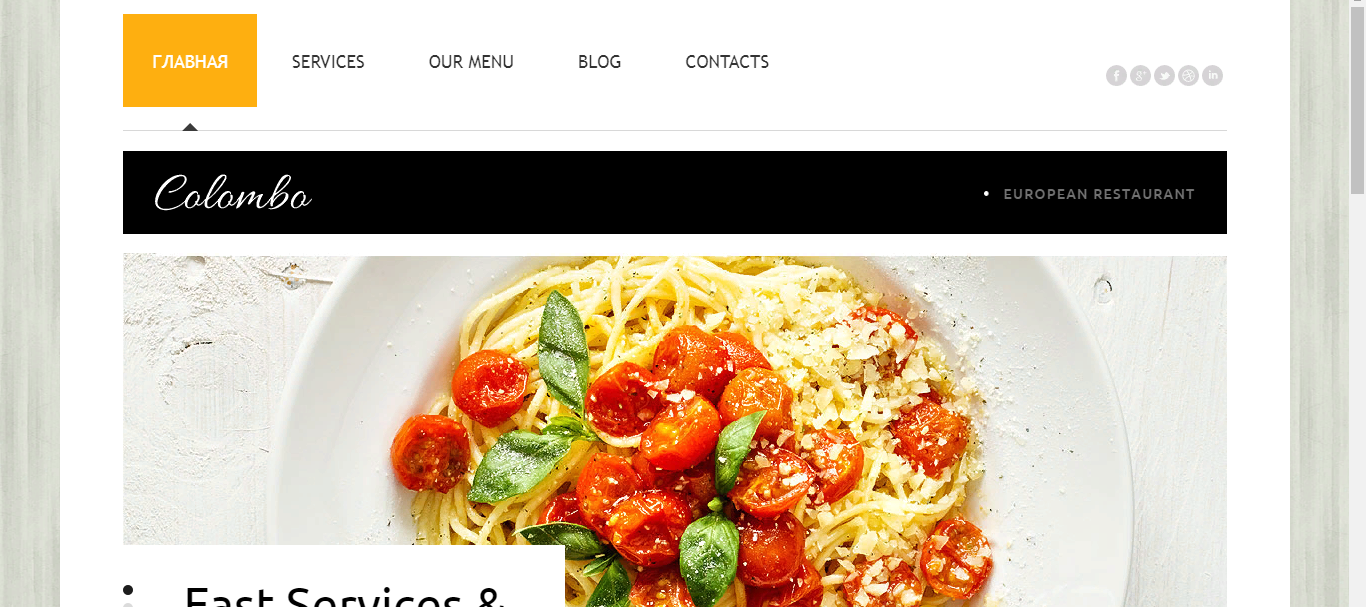
* Natural Language Projects at ISI

<https://www.isi.edu/research_groups/nlg/home>



* 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.

* 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.

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.

* 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.

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.

* 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.

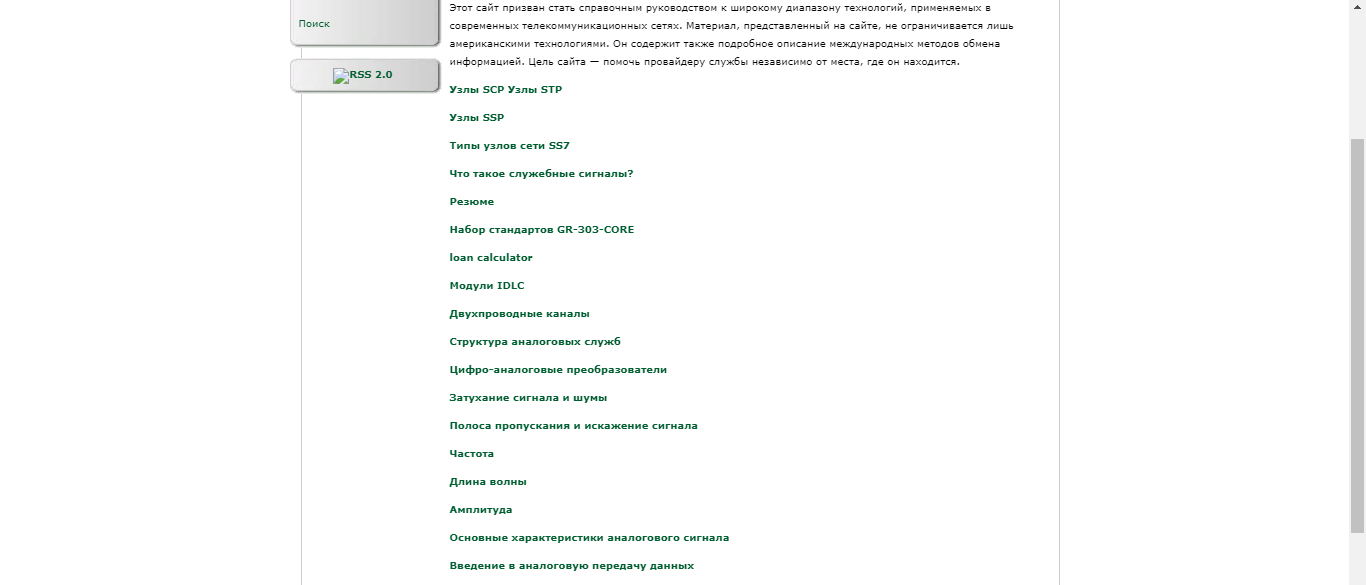
* Корпусный менеджер XAIRA Xaira (XML Aware Indexing and Retrieval Architecture) (<http://www.natcorp.ox.aс.uk>)

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

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.

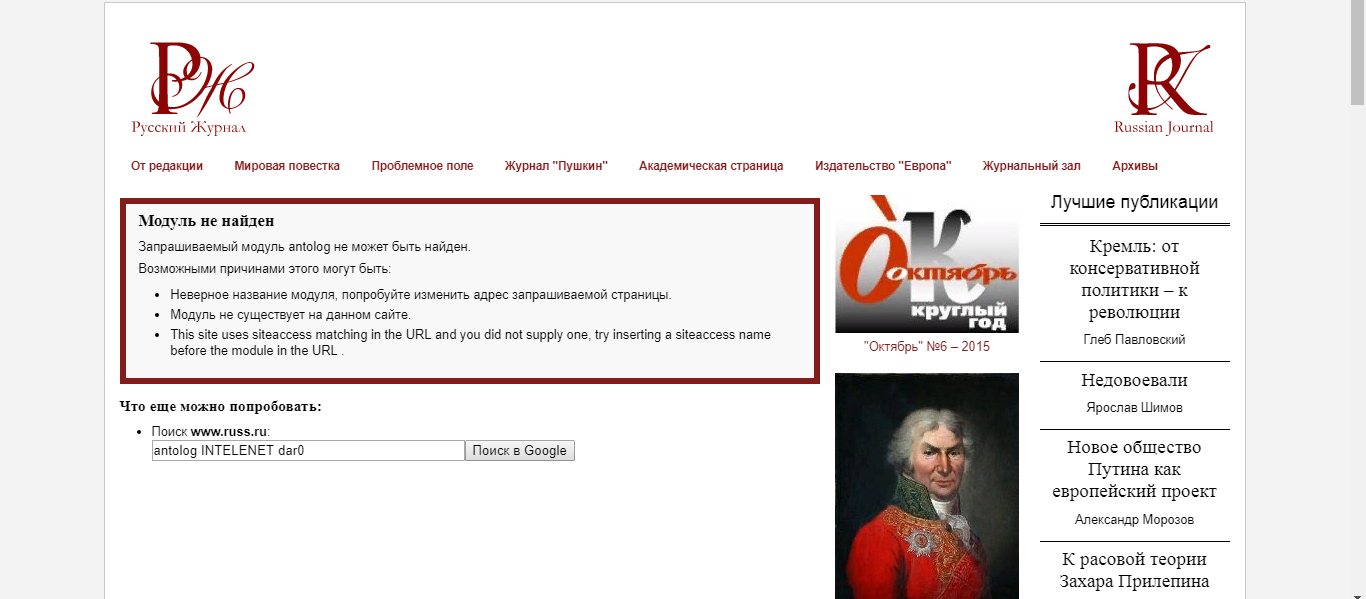
* Словарь В.Даля

<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>

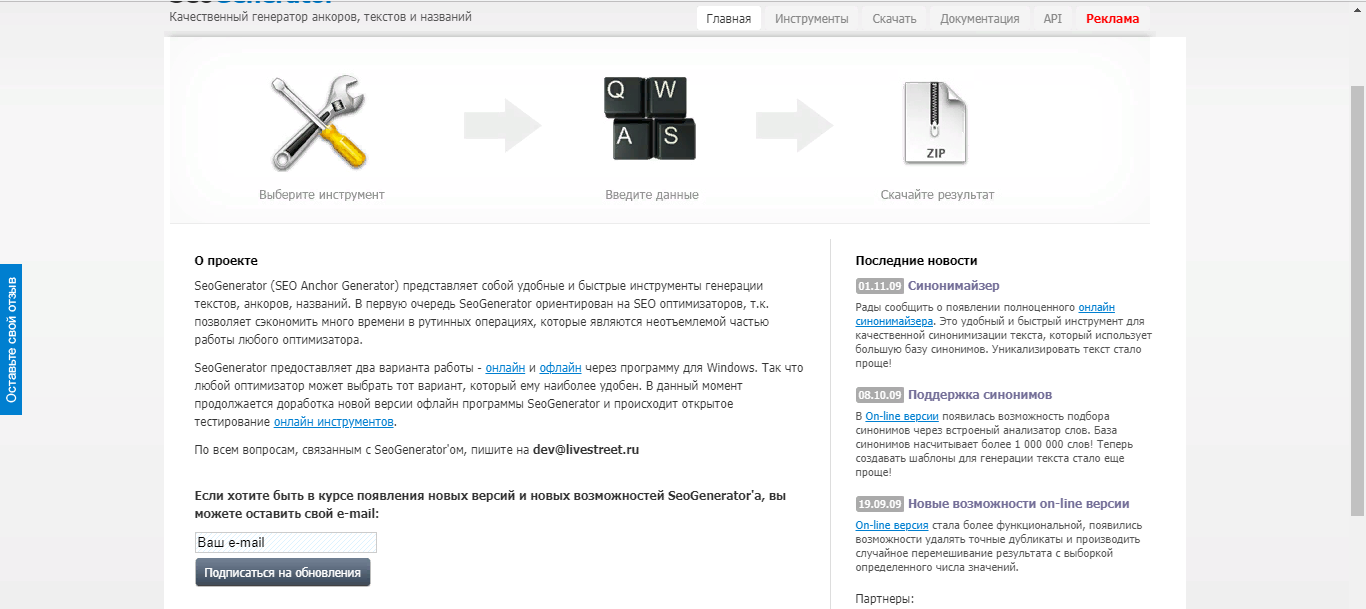


Сетевой проект известного филолога, философа и культуролога Михаила Эпштейна, стартовавший в апреле 2000. Проект связан с изучением феномена однословия, т.е. слова как самоценной сущности. Представляет собой сетевой еженедельник, публикующий эссе по словобразованиям и неологизмам русского языка.

10. Генераторы текстов.

* Сервис SeoGenerator.ru

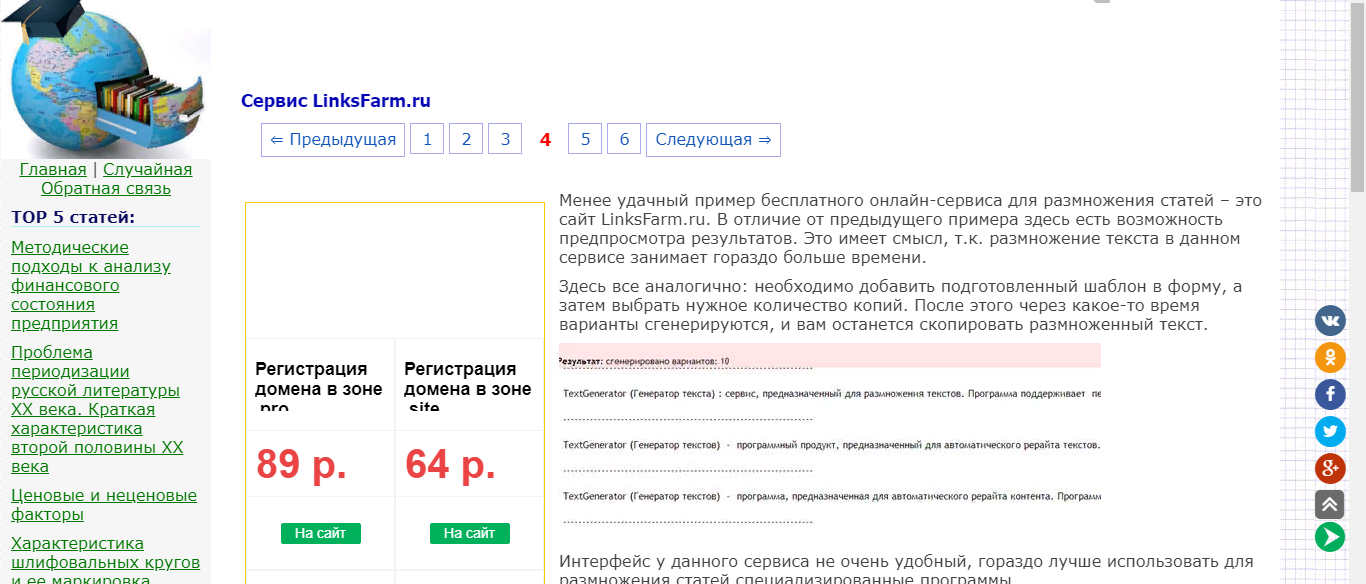
<http://seogenerator.ru/>



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

* Сервис LinksFarm.ru

<https://vikidalka.ru/1-77488.html>

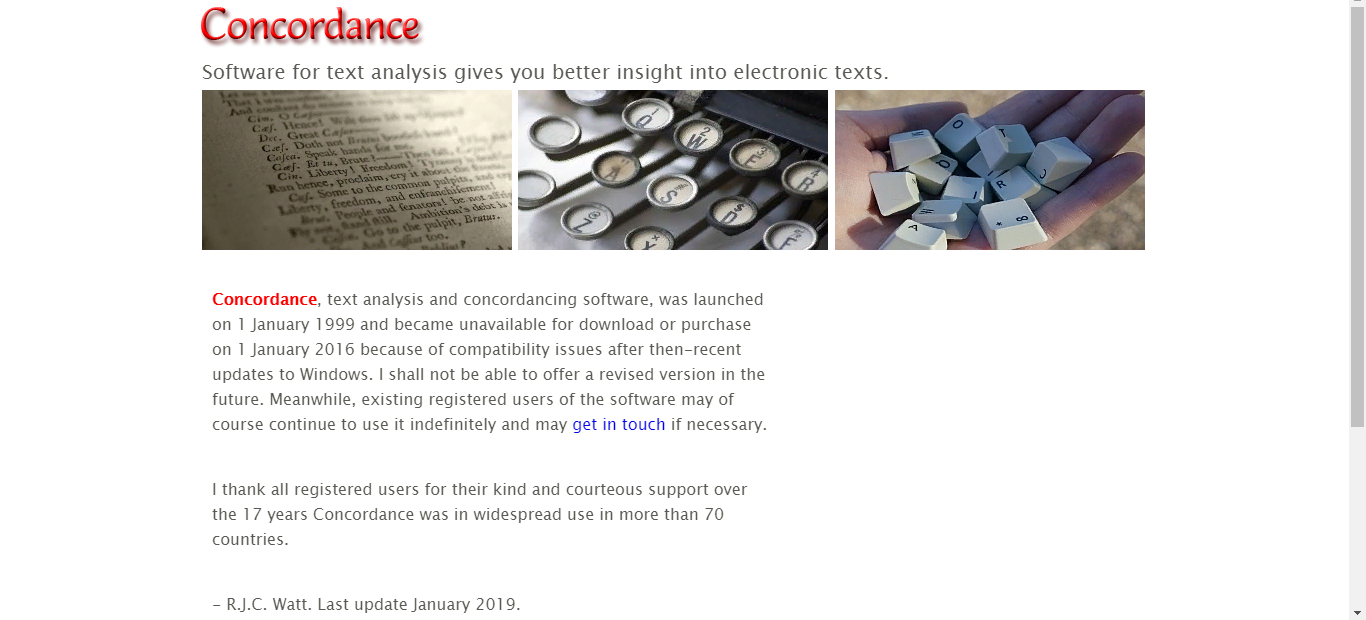


Менее удачный пример бесплатного онлайн-сервиса для размножения статей – это сайт LinksFarm.ru. В отличие от предыдущего примера здесь есть возможность предпросмотра результатов. Это имеет смысл, т.к. размножение текста в данном сервисе занимает гораздо больше времени.

11. Программа для построения конкордансов Concordance 2.0.0

* Concordance

<http://www.concordancesoftware.co.uk>



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

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.

12. Программа снятия омонимии.

* Синтаксический анализатор Диалинг

<http://www.dialog-21.ru/digest/2000/articles/sokirko/>



Система «Диалинг» основана на комбинировании различных вариантов анализа фразы. Здесь также морфологическая омонимия крайне негативно отражается на скорости работы.

* Морфологический анализатор английского языка ENGTWOL.

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.

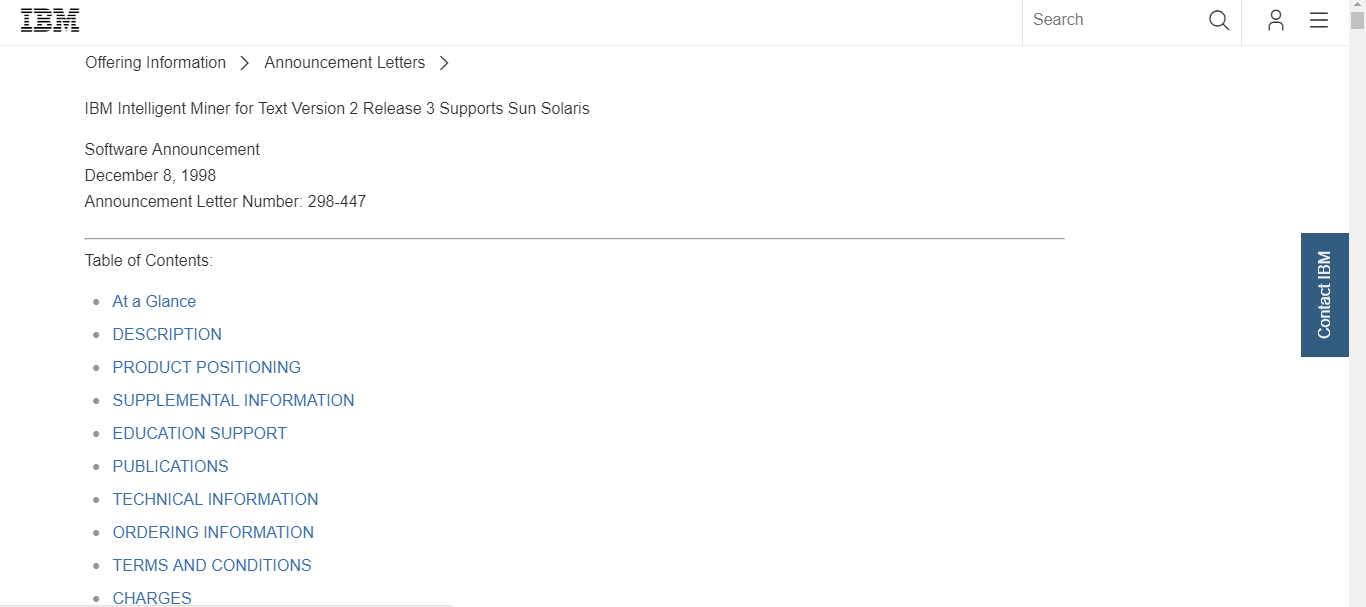
* Hidden Markov Model (HMM) tagging

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.

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

* IBM Intelligent Text Miner

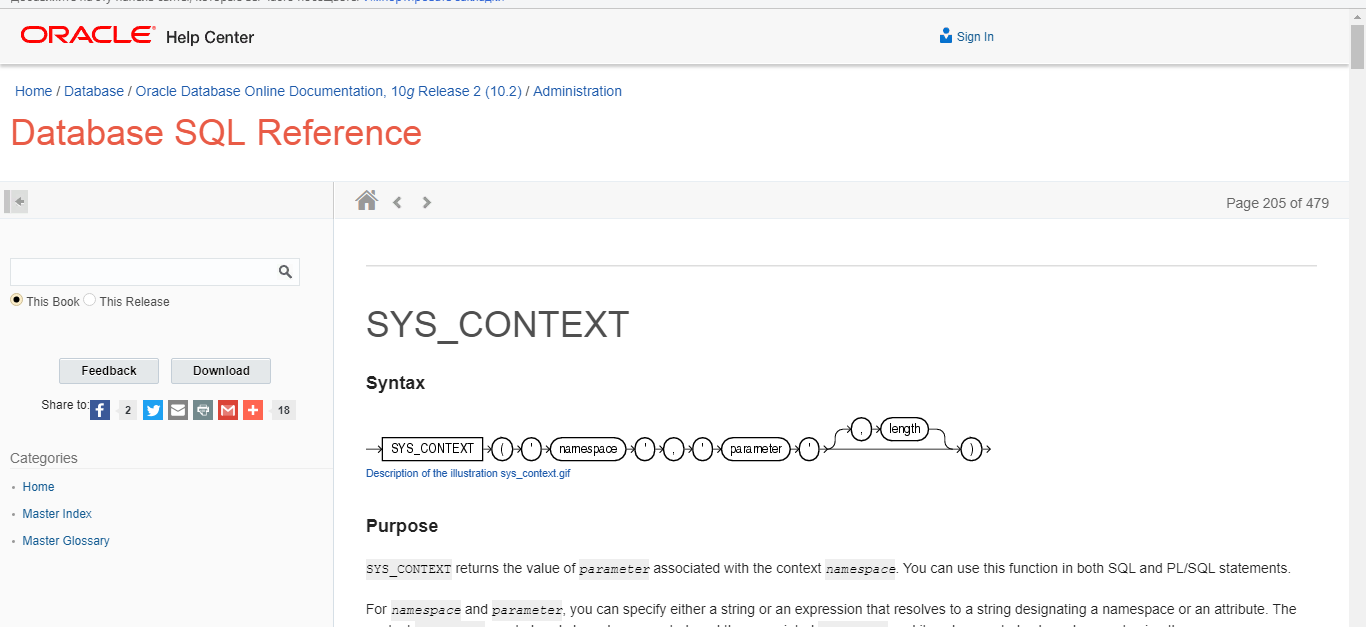
<http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?appname=iSource&infotype=an&letternum=ENUS298-447&subtype=ca&supplier=897>



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.

* Oracle Context

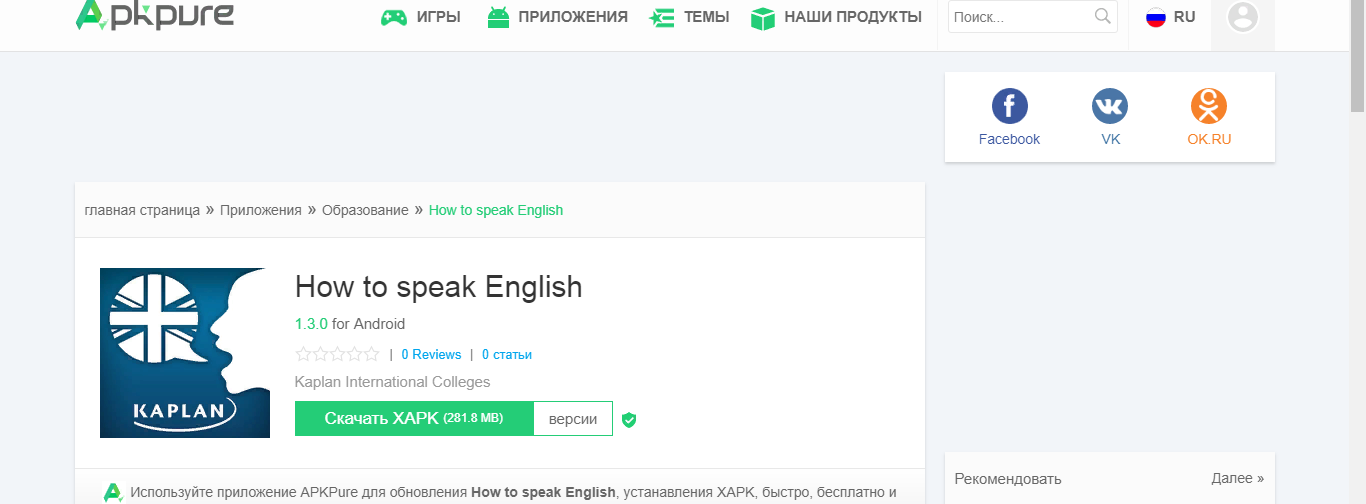
<https://docs.oracle.com/cd/B19306_01/server.102/b14200/functions165.htm>



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.

14. Сетевые ресурсы в обучении языкам.

* How to speak English



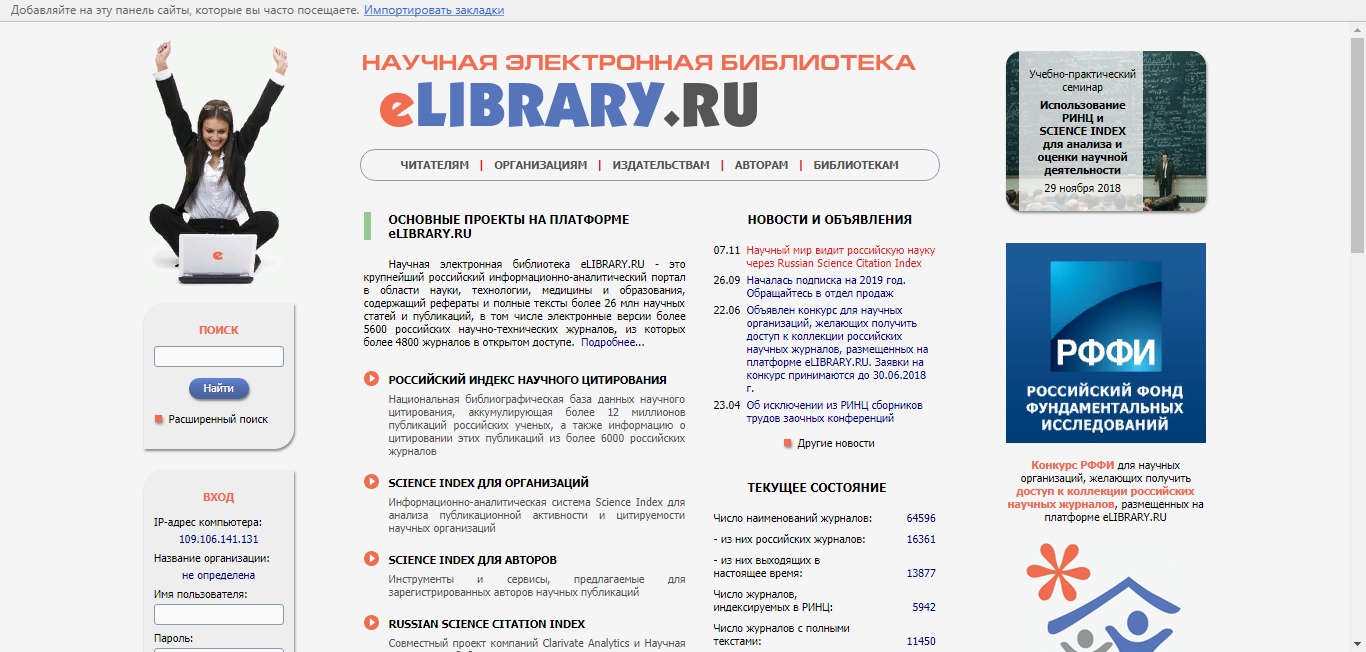
С помощью него вы сможете избавиться от акцента. Это возможно благодаря трем видам упражнения: секции Challenge, разделов Your Sounds и All Sounds. Первая секция построена по принципу «вопрос-ответ», которая позволяет определить уровень владения фонетикой. Your sounds позволит сравнить ваше произношение 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

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> 

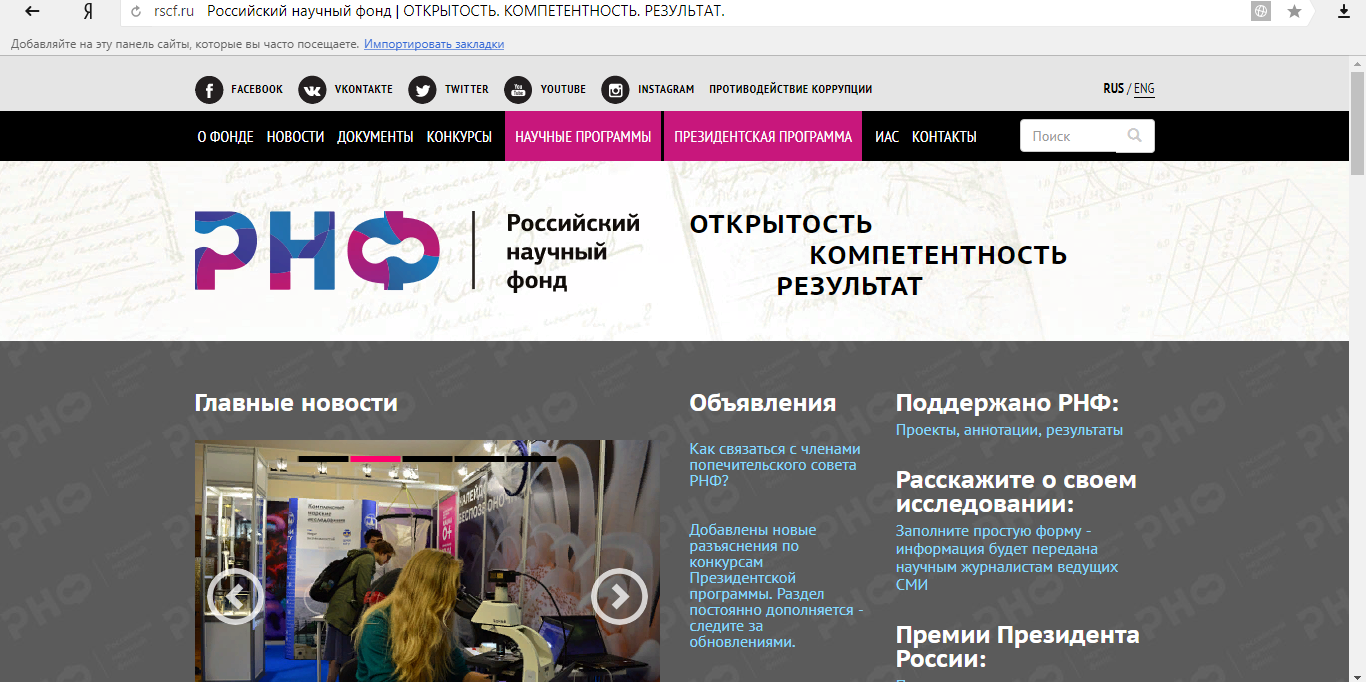
eLIBRARY.RU - крупнейшая в России электронная библиотека научных публикаций, обладающая богатыми возможностями поиска и анализа научной информации. Библиотека интегрирована с Российским индексом научного цитирования (РИНЦ) - созданным по заказу Минобрнауки РФ бесплатным общедоступным инструментом измерения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

* Российский научный фонд.

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

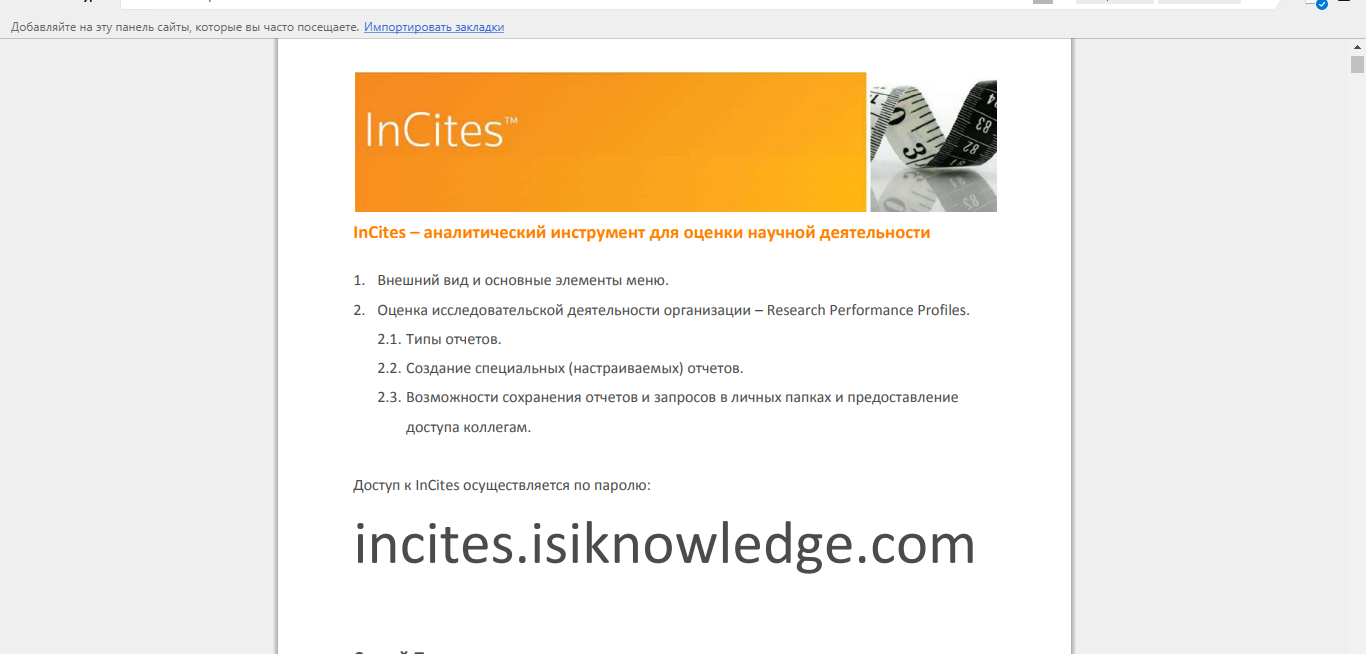
Целью деятельности РНФ является финансовая и организационная поддержка фундаментальных научных исследований и поисковых научных исследований, подготовки научных кадров, развития научных коллективов, занимающих лидирующие позиции в определенной области науки.

* Индексы научного цитирования.

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

Индекс цитирования 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.