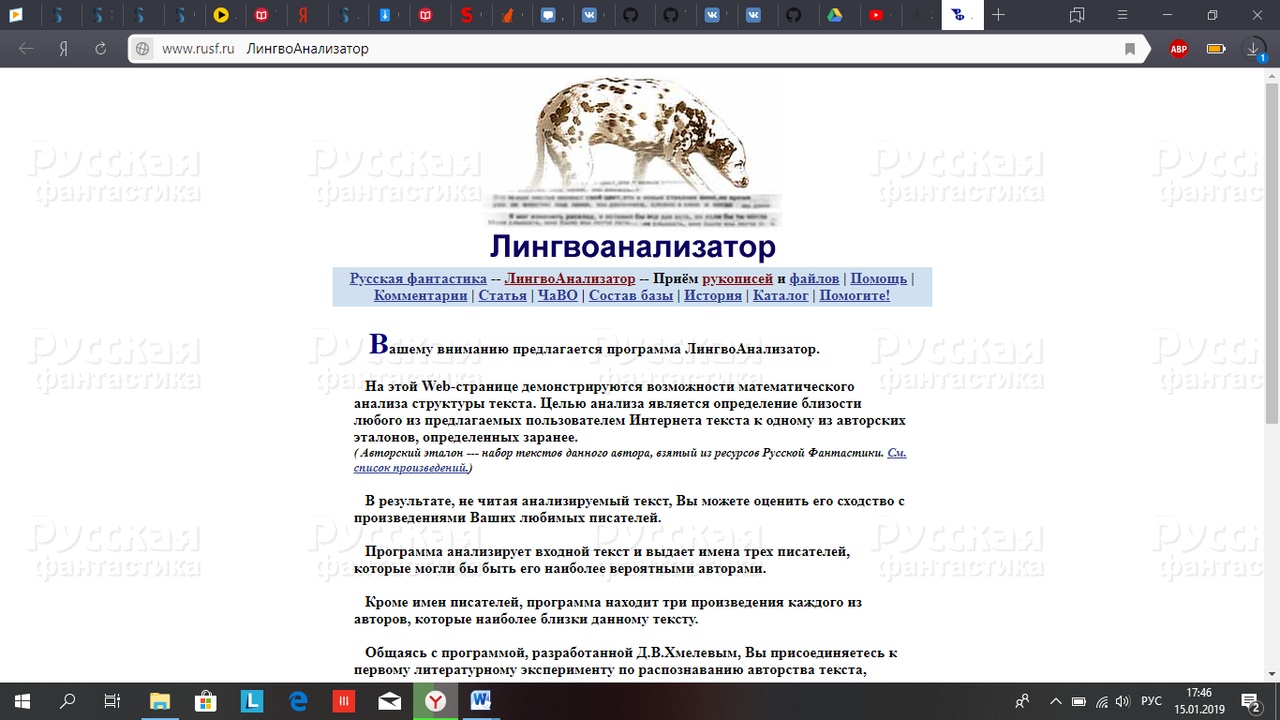
Лингвистические программы и ресурсы

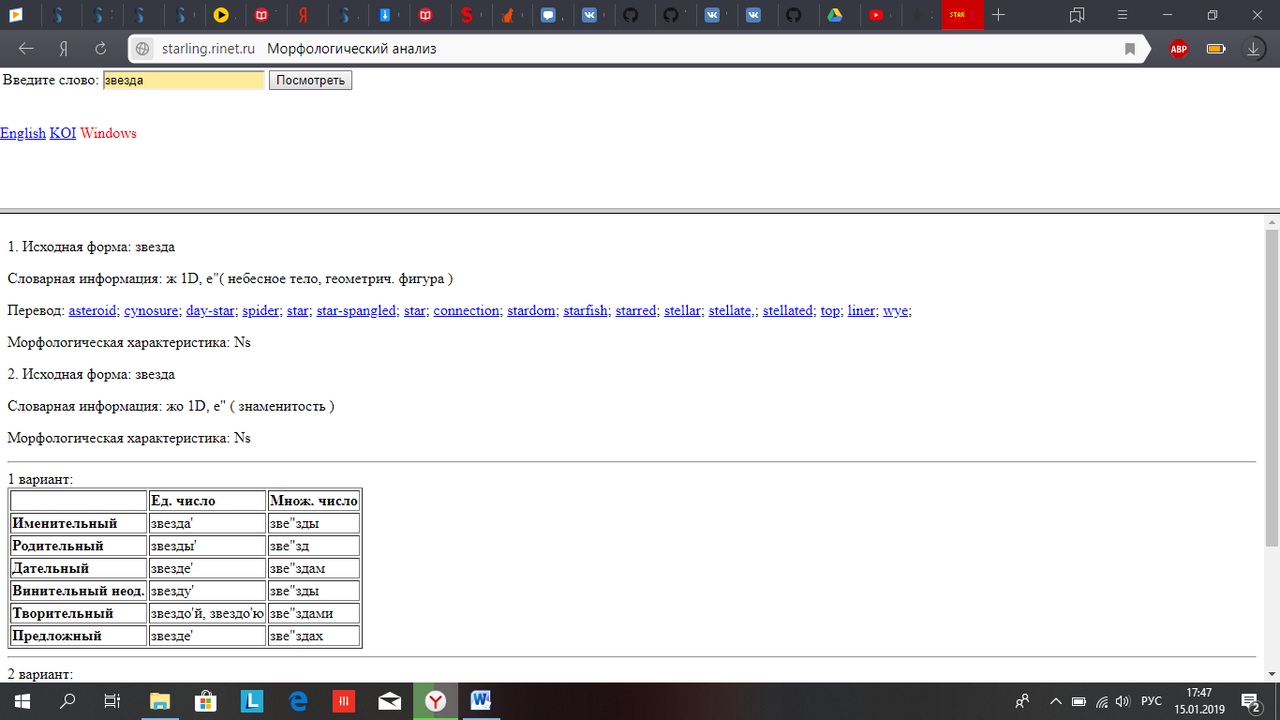
# Программы анализа и лингвистической обработки текстов

1. ЛингвоАнализатор <http://www.rusf.ru/books/analysis/>



программы математического анализа структуры текста. Целью анализа является определение близости любого из предлагаемых пользователем текстов к одному из авторских эталонов, определенных заранее. Программа анализирует входной текст и выдает имена трех писателей, которые могли бы быть его наиболее вероятными авторами. Кроме этого, программа находит три произведения каждого из авторов, которые наиболее близки данному тексту.

1. Морфологический анализ <http://starling.rinet.ru/morph.htm>



On-line версия программы морофлогического анализа слов русского/английского языков. Позволяет получить для вводимого слова базовую форму и морфологическую информацию. Программа реализована на основе словарей Зализняка (рус.яз.) и Мюллера (англ.яз.).

1. АОТ Автоматическая Обработка Текста <http://www.aot.ru/cgi-bin/morph.py>

Сайт, на котором представлены разработки бывших сотрудников компании Диалинг, прекратившей свое существование в мае 2001г. Среди предлагаемых продуктов:

- модуль графематического анализа текста;

- компоненты морфологического анализа для русск. и англ.яз.;

- модуль автоматического уничтожения омонимии;

- модуль семантического анализа текста;

- различные тезаурусы.

Также опубликована диссертация А.Сокирко "Семантические словари в автоматической обработке текста" по теме машинного перевода.

# Психолингвистические программы

1. Пси-Офис <http://psy-two.narod.ru/>

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).

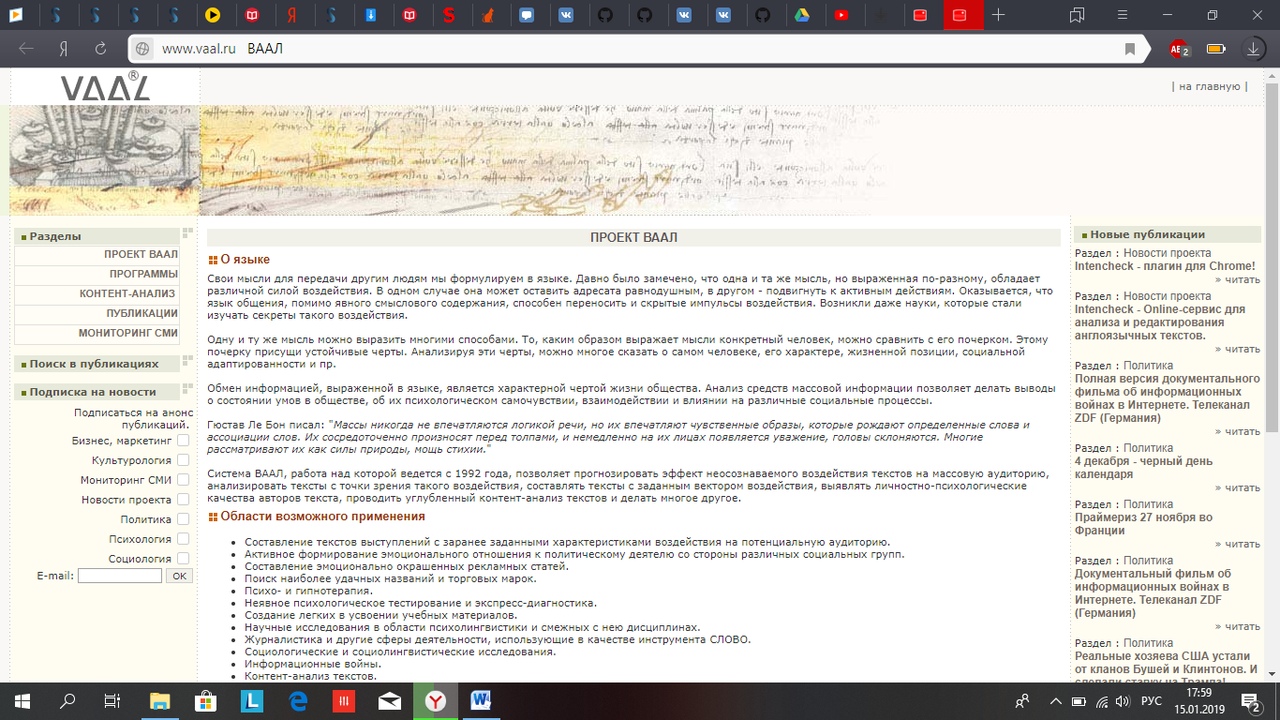
- программа синтеза подсознательного компонента текста.

1. Лаборатория ПСИТОН <http://psyton.ru/index.html>



Позволяет повысить валидность результатов значительно сократив время тестирования. Результаты предоставляются в виде: показателей по шкалам, психографического профиля и итоговой интерпретации. Данные психодиагностики можно распечатать и сохранить. Комплект включает базисные стандартизированные психодиагностические методики для проведения многоаспектной и специально-ориентированной психодиагностики.

1. ВААЛ-2000 <http://www.vaal.ru/>

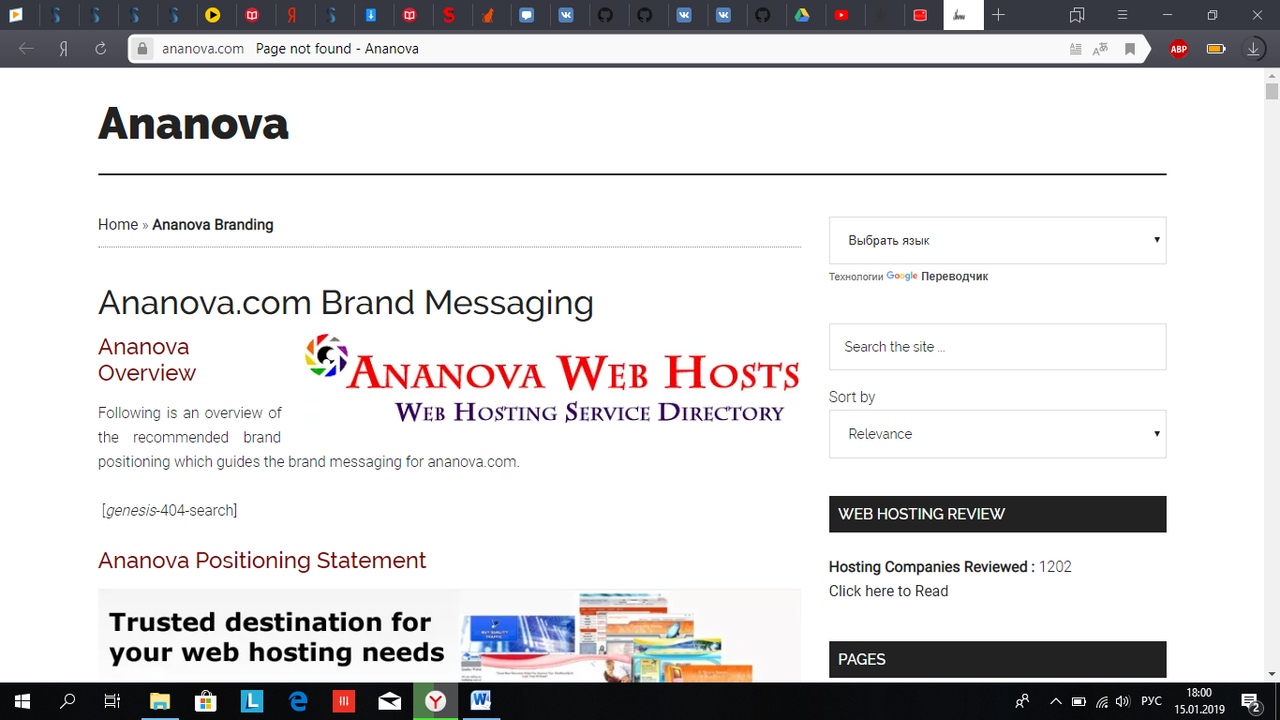


Психолингвистическая экспертная система лексического и контент-анализа текстов. Основные задачи - прогноз эффекта неосознаваемого воздействия текста на массовую аудиторию, анализ текстов с точки зрения такого воздействия, генерация текста с заданным вектором воздействия, выявление личностно-психологических качеств автора текста.

Система реализована в виде набора DLL-библиотек, которые подключаются к наиболее популярному текстовому процессору Word for Windows. Просто в главном меню появляется новый пункт. Такой способ реализации позволяет сохранить для пользователя привычную удобную среду создания документов и максимально облегчает освоение системы ВААЛ.

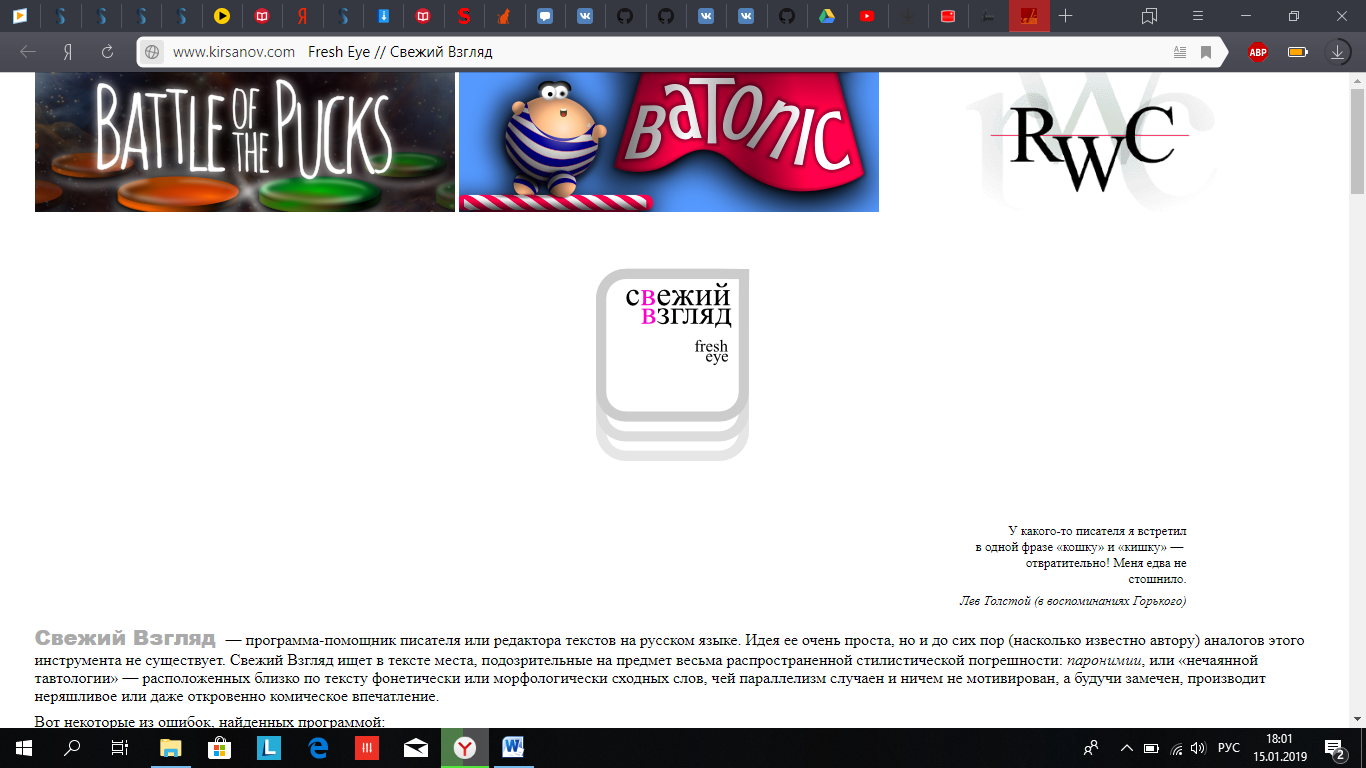
# Генераторы текстов и "говорящие" программы

1. ANANOVA <https://ananova.com/video/lo_video?category=summary>



Новостное агенство в Великобритании, использующее на своем сайте виртуального анимированного диктора. Т.е. симпатичная виртуальная девушка, сверкая глазками и шевеля губками, синтезированным голосом рассказывает вам о последних новостях (естественно, на английском языке).

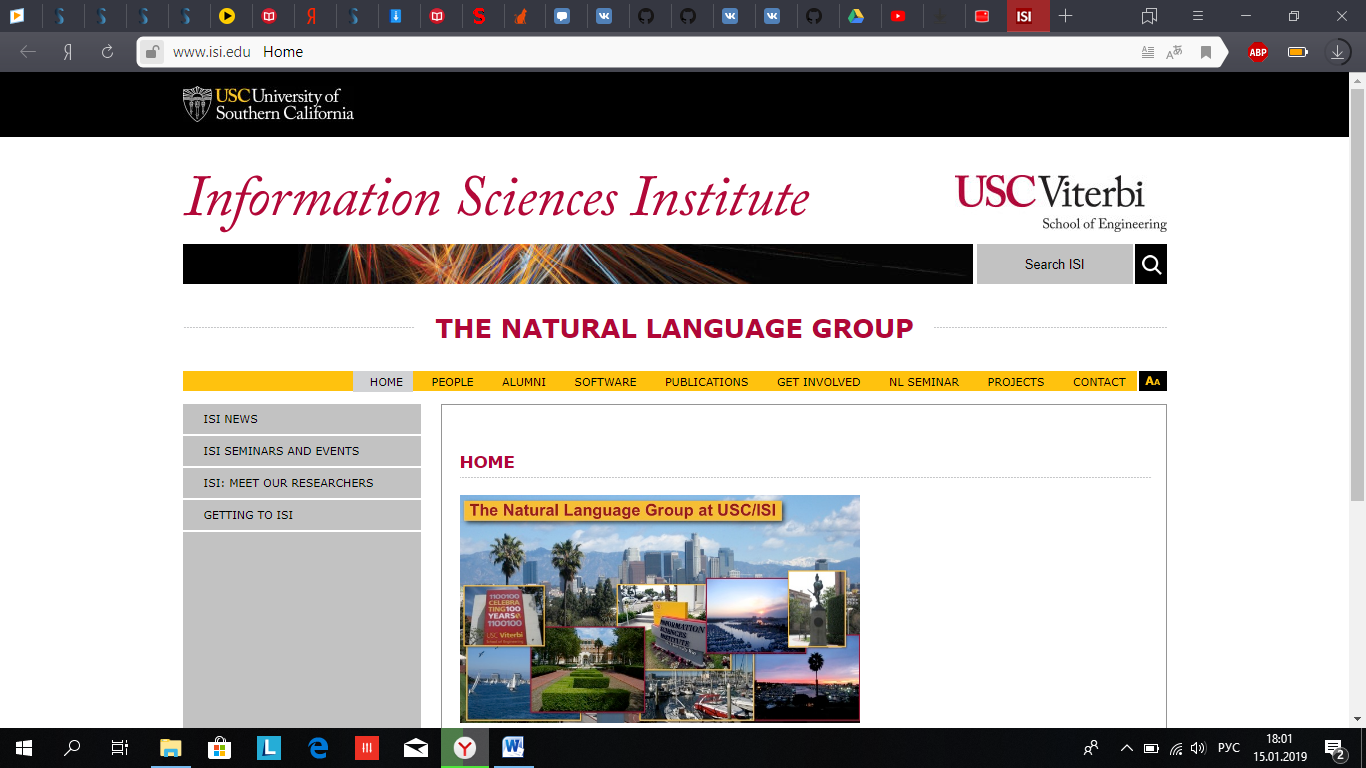
1. Russian Word Constructor <http://www.kirsanov.com/fresheye/>



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.

Системы обработки естественного языка

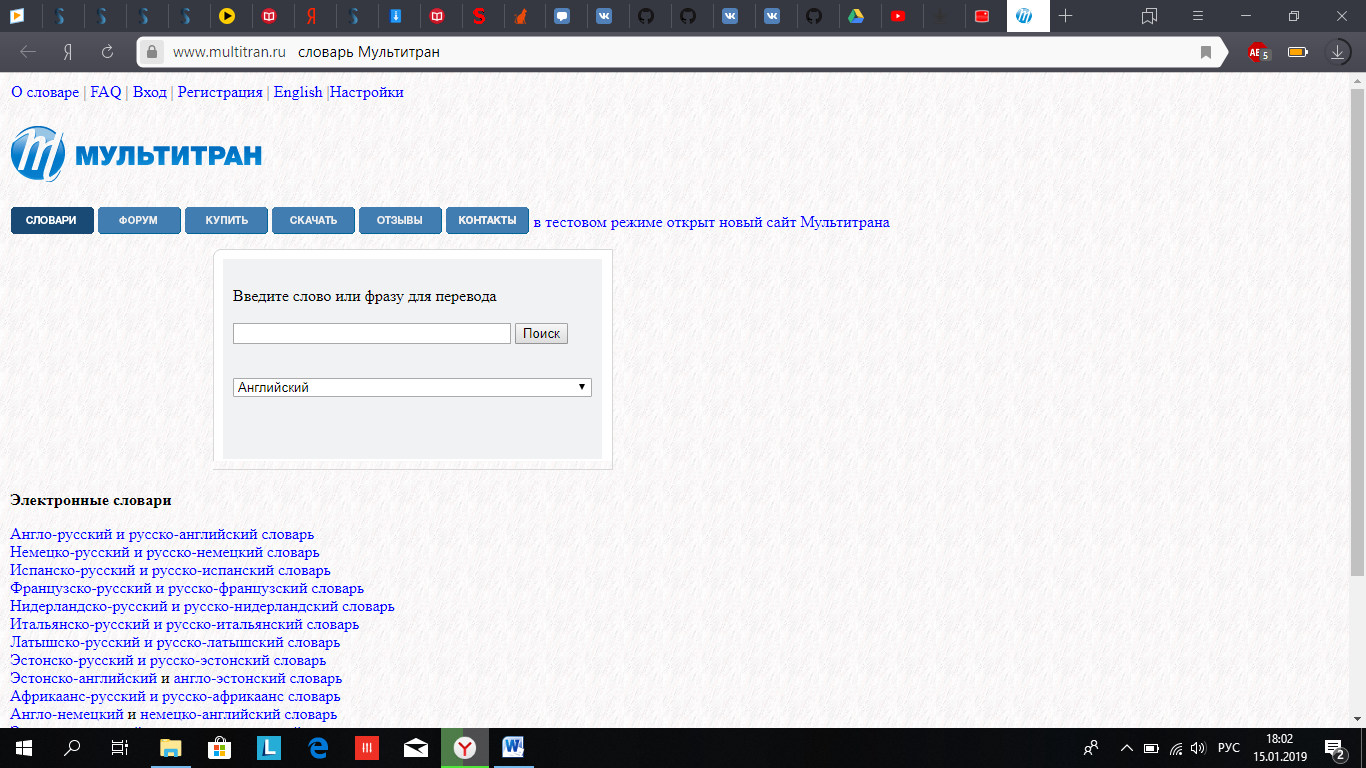
1. Natural Language Projects at ISI <https://www.isi.edu/research_groups/nlg/home>



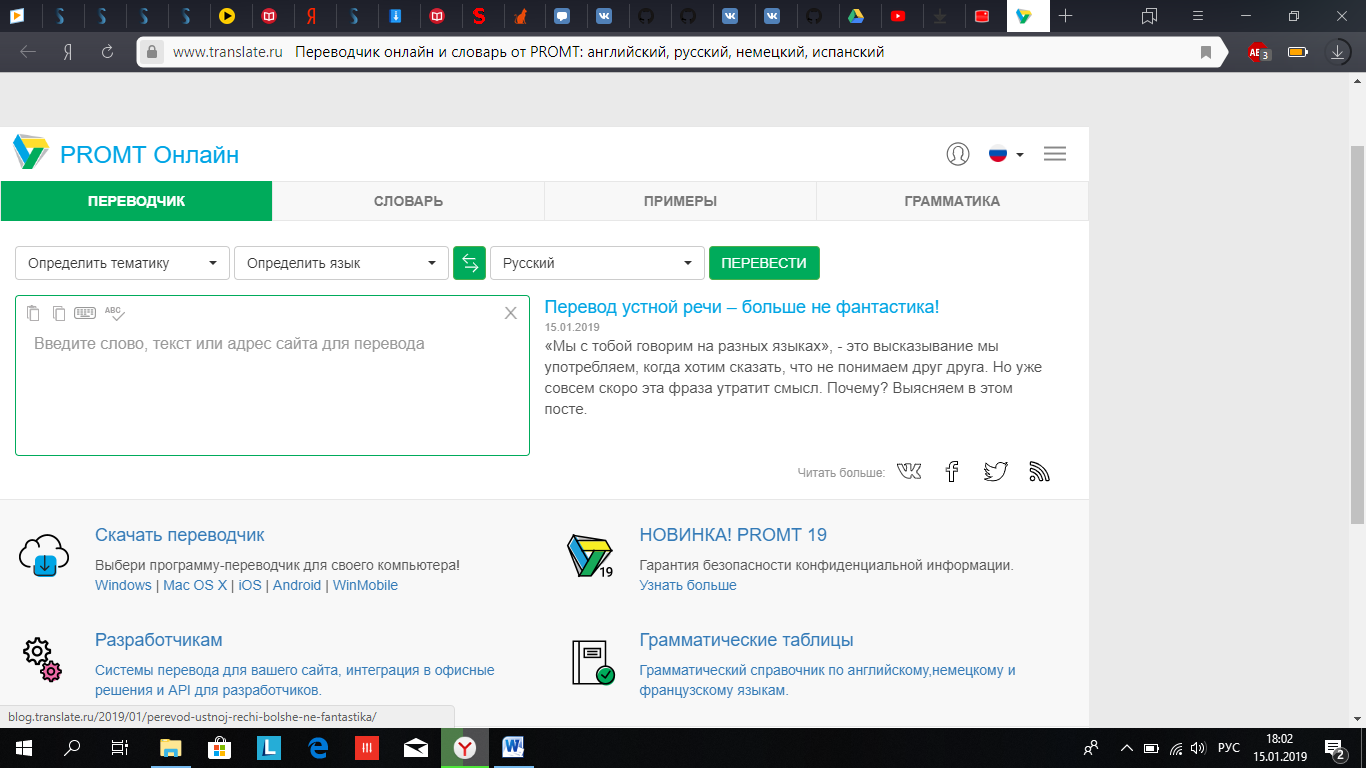
Описание группы проектов по вычислительной лингвистике и машинному переводу, выполняемых в Университете Южной Калифорнии.

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

1. Мультитран <https://www.multitran.ru/>



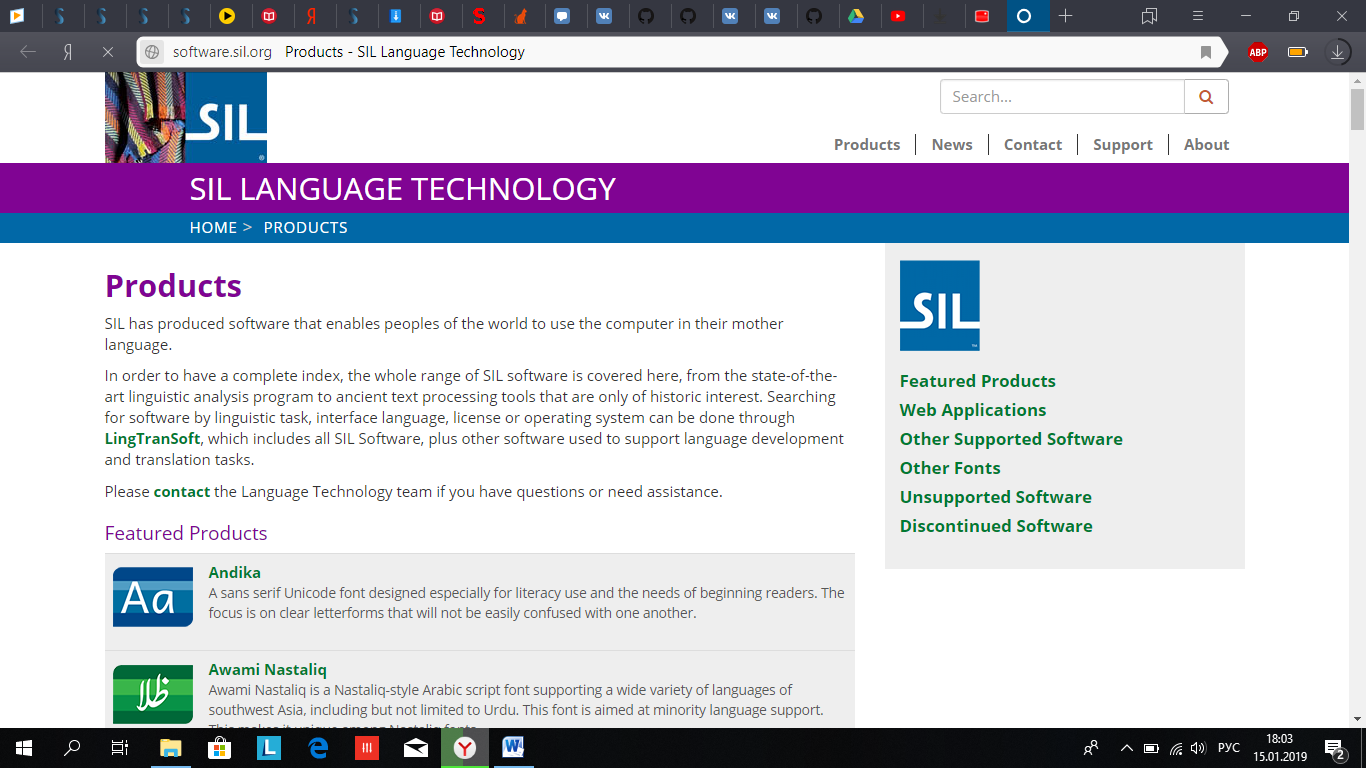
Система автоматического перевода, поддерживающая русский, английский, немецкий, французский и испанский языки. Содержит более пяти миллионов терминов и предоставляет возможности алфавитного, морфологического и фразового поиска.

1. Translate.Ru <http://www.translate.ru/>
2. 

Портал on-line перевода компании ПРОМТ. Поддерживает русский, английский, немецкий, французский, итальянский и испанский языки.

# Коллекции ресурсов

1. Каталог программ по вычислительной лингвистике <http://software.sil.org/products/>

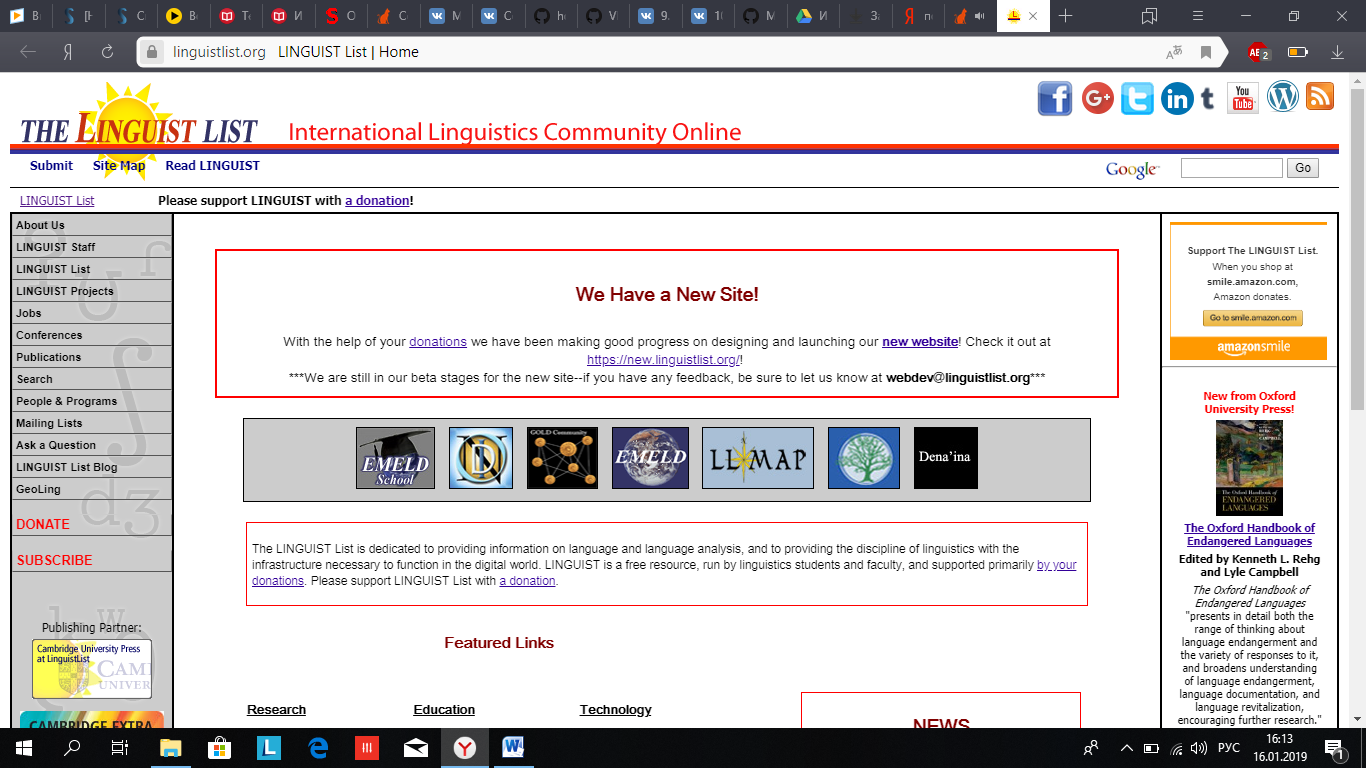


Обширный каталог программ по вычислительной лингвистике, разработанных в рамках проекта SIL. Представленные в каталоге программы носят явно академический характер и требуют значительных усилий для изучения и применения. Не рекомендуются неискушенным в академической лингвистике пользователям. Все программы,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

1. Text Analysis Info <http://www.textanalysis.info/>
2. 

Text Analysis Info - это бесплатный информационный портал, посвященный анализу контента межчеловеческого общения (в основном текстов). Также представлены различные программы, предназначенные для обработки таких источников как аудио-, видео- или речевых данных.

1. The Linguist List <https://linguistlist.org/>

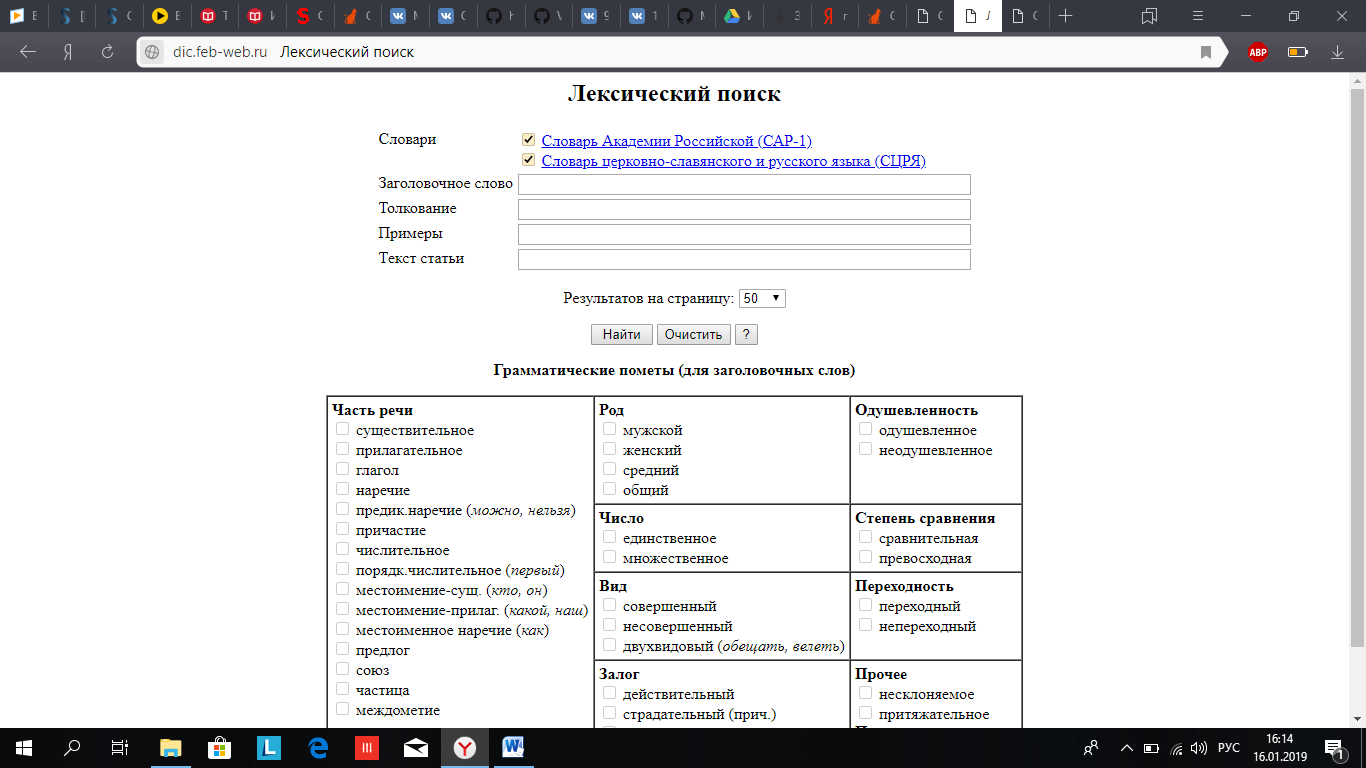


Большой каталог программного обеспечения для различных областей компьютерной обработки текстов и лингвистики.

# электронные словари и тезаурусы

1. Сводный исторический словарь русского языка XVIII–XX веков

<http://dic.feb-web.ru/rusdict/index.htm>

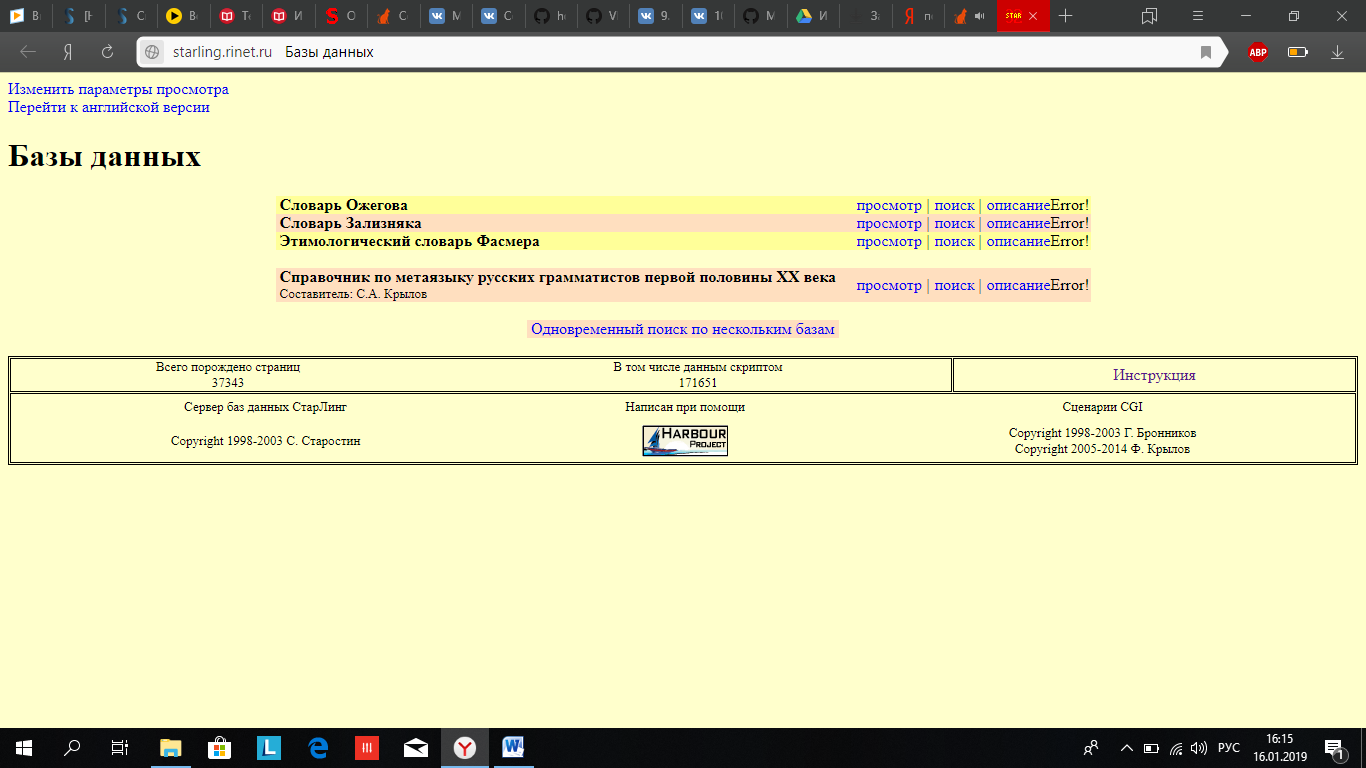


Проект создания интегрированной словарной базы данных русского языка XVIII–XX веков, включающей глубокую разметку словарных статей и возможность поиска по зонам. Текущая реализация включает поиск по словарям:

Словарь Академии Российской, по азбучному порядку расположенный: В 6 т. (1806–1822).

Словарь церковнославянского и русского языка: В 4 т. (1847).

1. Словари Ожегова и Зализняка <http://starling.rinet.ru/cgi-bin/main.cgi?flags=wygnmnl>

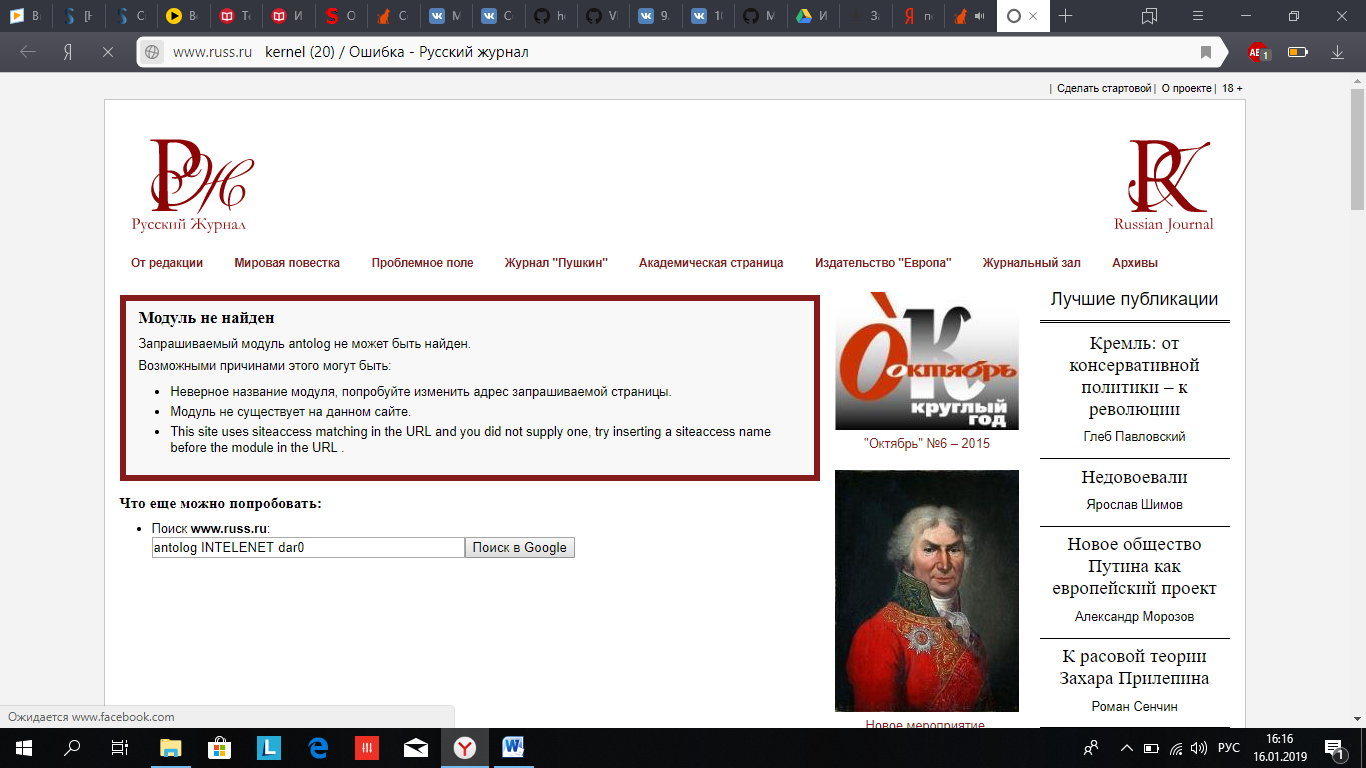


On-line версии словарей Ожегова и Зализняка.

Первый экран интерфейса к этимологической базе данных предлагает выбор из имеющихся баз, а также настройку параметров интерфейса.

Главное поле выбора позволяет указать, к какой базе данных Вы хотите сделать запрос.

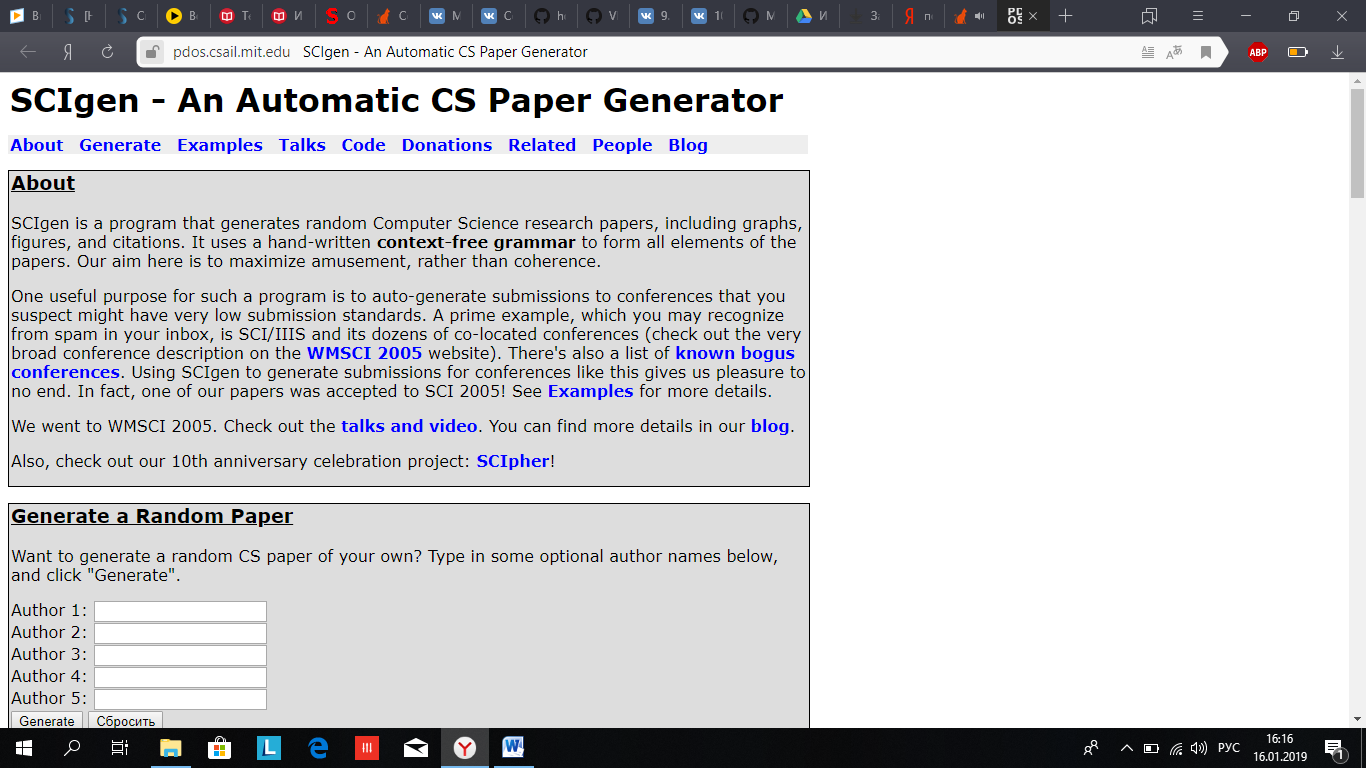
1. Лексикон "Дар слова" <http://www.russ.ru/antolog/INTELNET/dar0.html>



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.

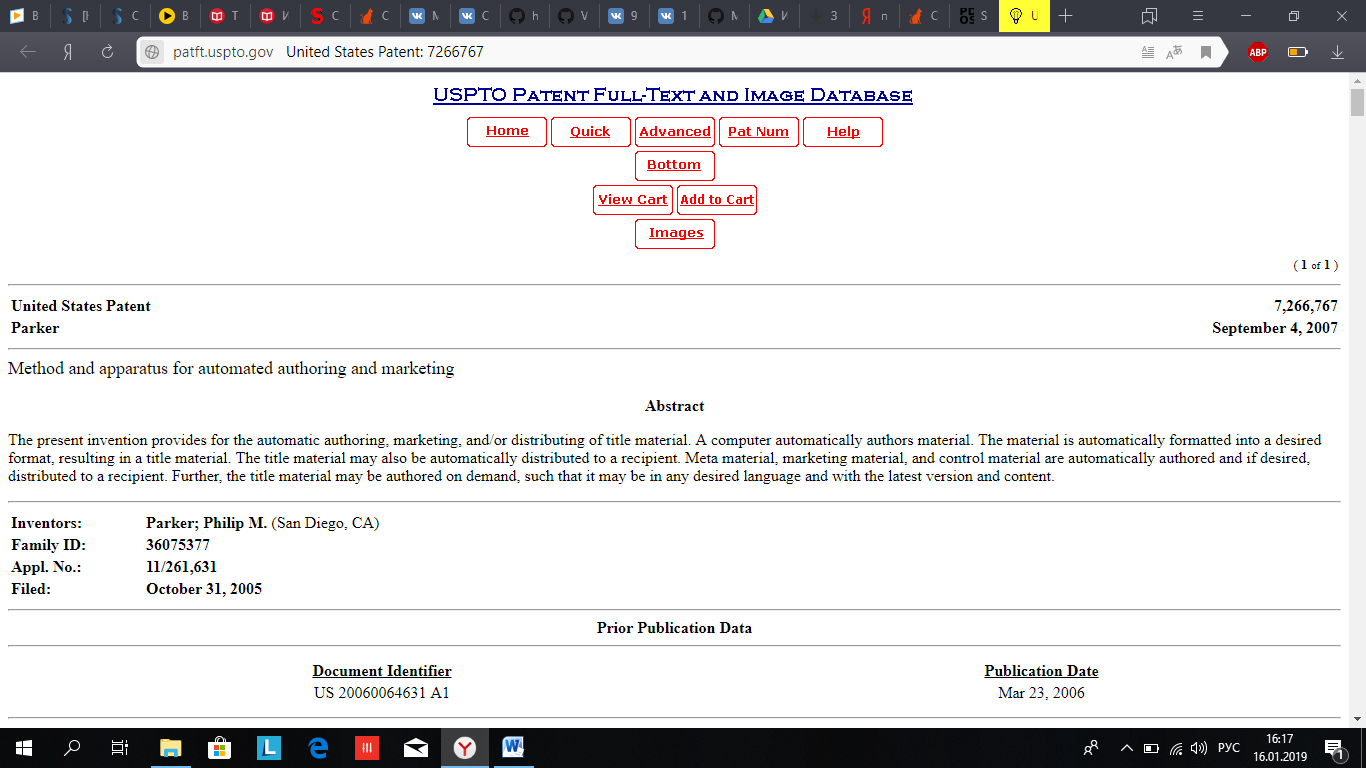
# генераторы текстов (on-line версия программы морфологического анализа слов русского/английского языков С.А. Старостина)

1. SCIgen <https://pdos.csail.mit.edu/archive/scigen/>



SCIgen - это программа-генератор случайных псевдо научных текстов по вычислительной технике, включающих графики, иллюстрации и цитирования других работ. Для формирования элементов статьи используется универсальная грамматика, не зависящая от контекста. Одним из вариантов применения такой программы авторы называют автогенерацию статей для конференций с сомнительно низкими стандартами для принимаемых публикаций

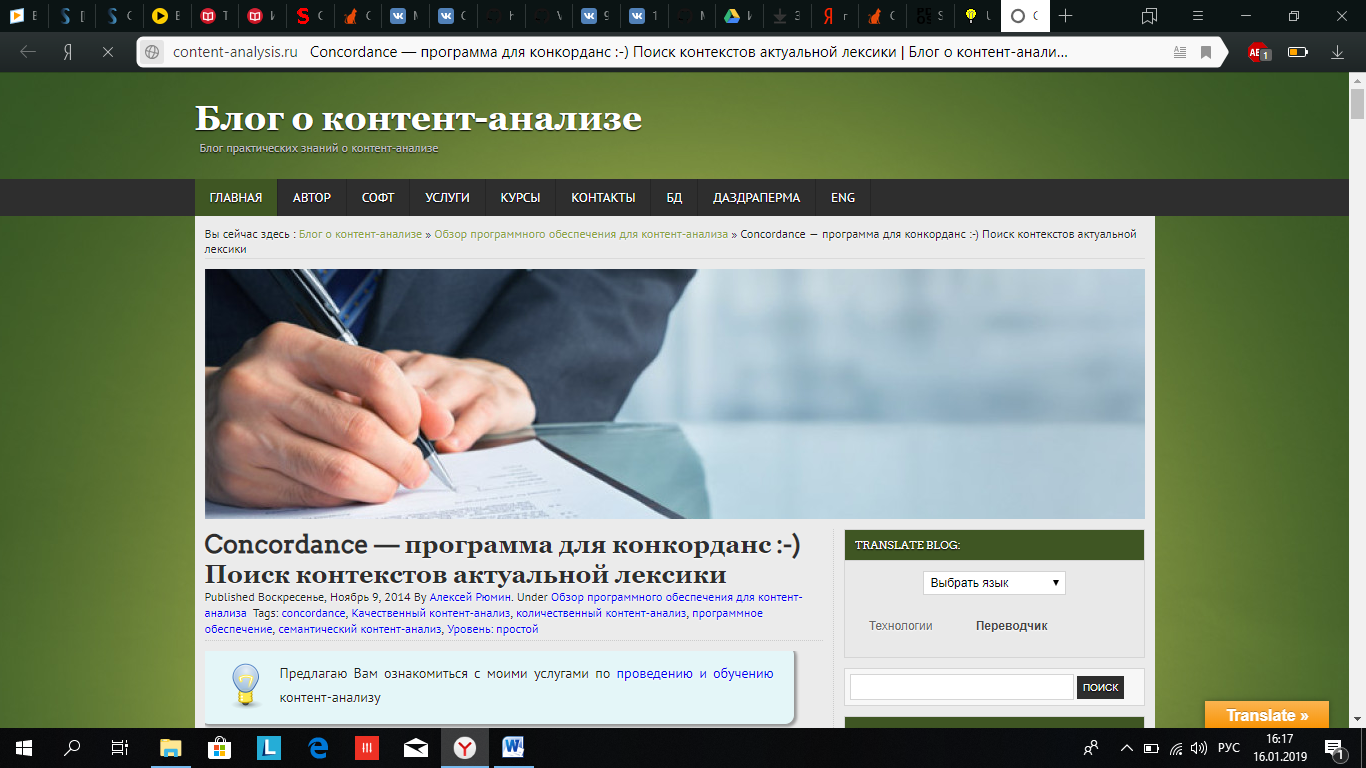
1. Method and apparatus for automated authoring and marketing <http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PALL&p=1&u=%2Fnetahtml%2FPTO%2Fsrchnum.htm&r=1&f=G&l=50&s1=7,266,767.PN.&OS=PN/7,266,767&RS=PN/7,266,767>



Патент на изобретение (англ.), описывающий методику автоматической генерации специализированных отчетов и справочников на основе публично доступной в Интернет информации. Данные по заданным предметным областям предварительно извлекаются из интернет-источников и накапливаются в базе данных. Пользователь задает предметную область, жанр и набор опций (язык публикации, шаблоны оформления страниц, графиков и таблиц) и система автоматически генерирует контент, который экспортируется в MS Word документ. С помощью набора макросов документ за считанные минуты автоматически форматируется и проверяется на орфографию, генерируются заголовки разделов, номера страниц, таблица содержания и библиографические метаданные.

# • программа для построения конкордансов Concordance 2.0.0

<http://content-analysis.ru/index.php/soft/concordance-programma-dlya-konkordans-poisk-kontekstov-aktualnoj-leksiki/>



Коммерческая программа для построения конкордансов и частотных списков для Win9x/NT/2000 (регистрация 80$). Обработка текстов в кодировках, поддерживаемых Windows, настройка на заданный алфавит. Сохранение результатов в виде HTML-файлов. Богатые средства анализа текстов. Возможность бесплатной загрузки для пробного ознакомления.