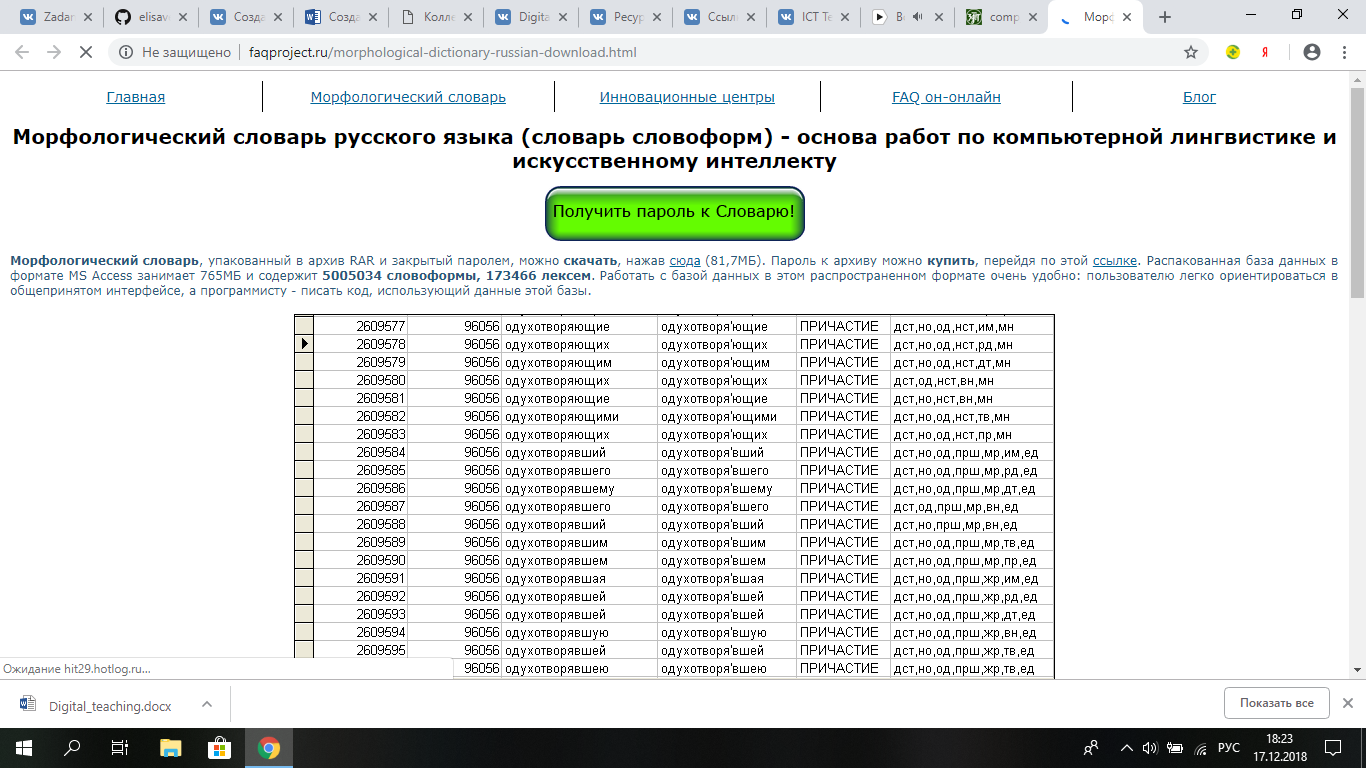
**1.Компьютерная обработка и анализ текста.**

://faqproject.ru/morphological-dictionary-russian-download.html

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. 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.



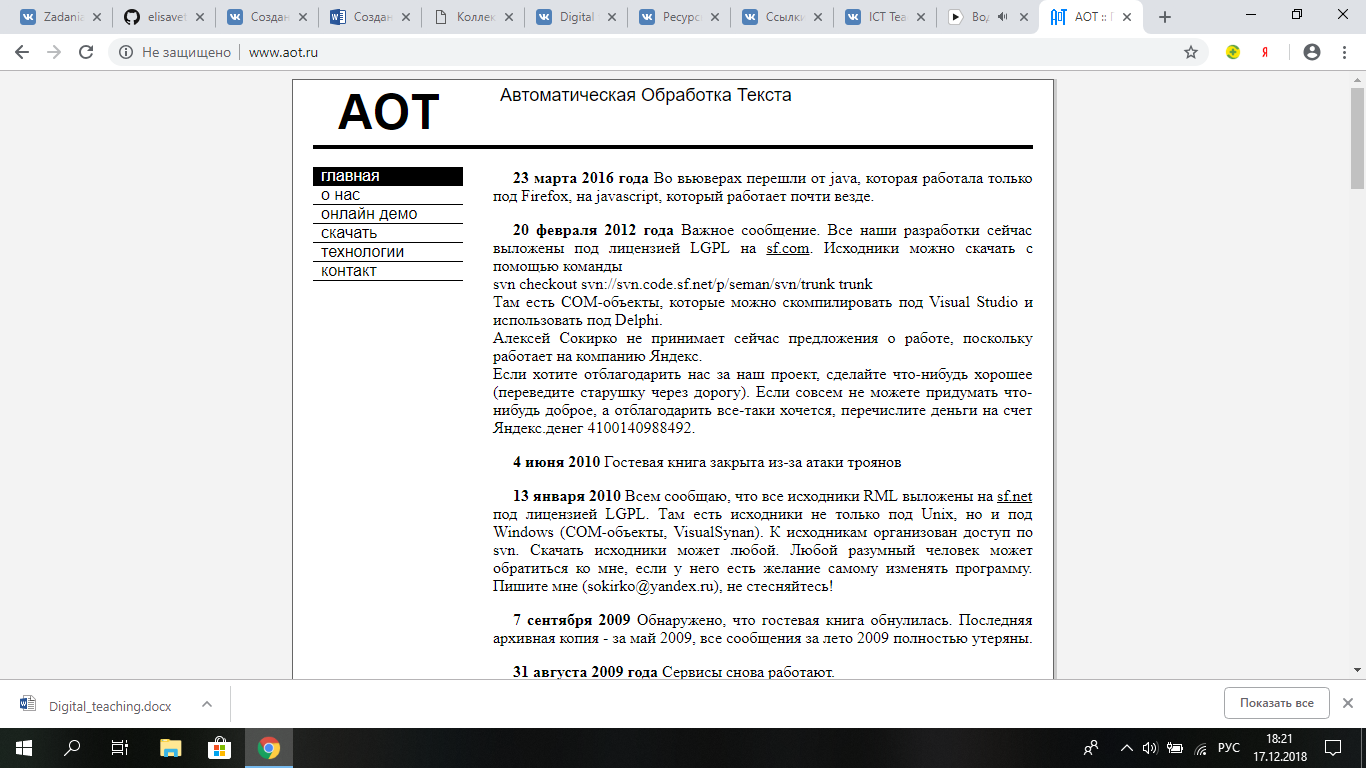
<http://www.aot.ru/>

Программа 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;

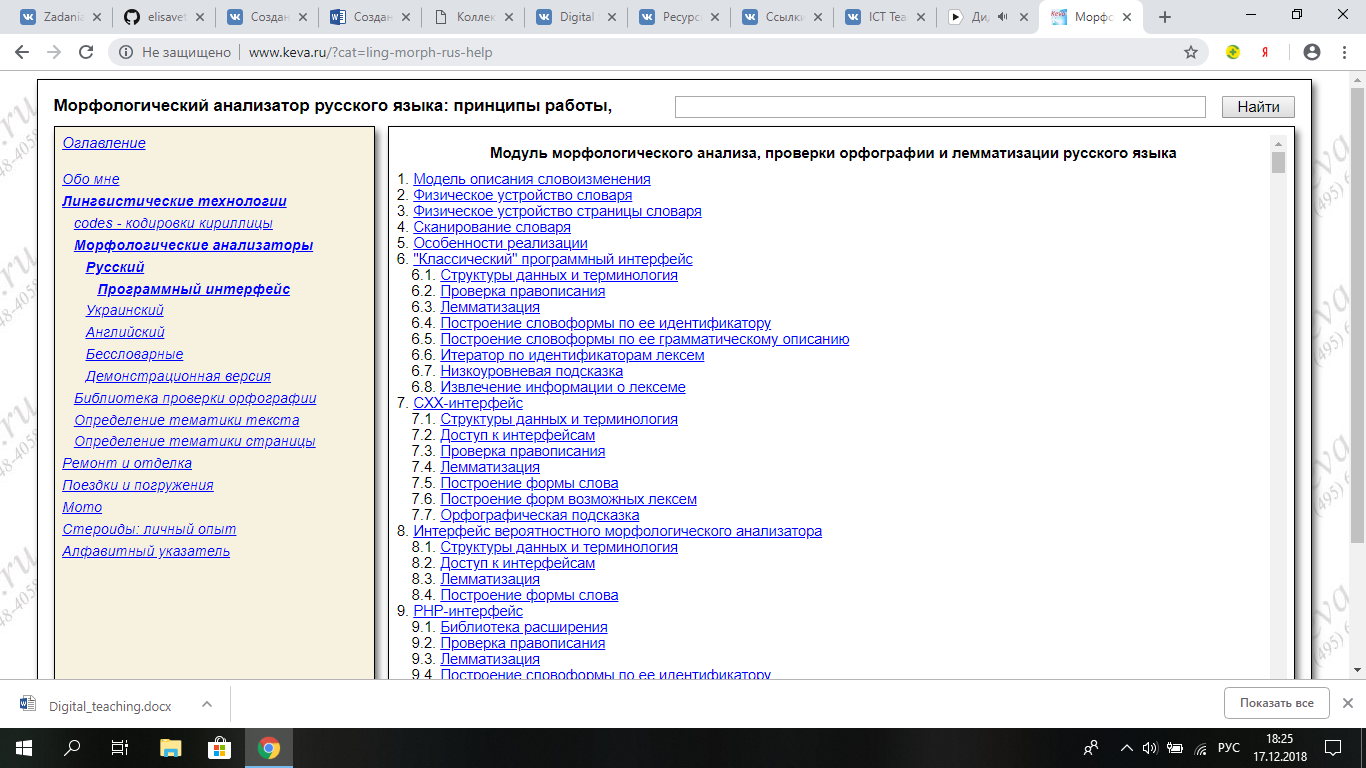
 морфологический анализатор текстов на русском и английском языках;

 построение конкордансов для заданной совокупности текстов



<http://www.keva.ru/ling/rus/help.htm>

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. В случае многозначности введенной формы выводятся все варианты анализа. Именно возможность получить варианты анализа введенной в программу формы представляется привлекательной с точки зрения нашего проекта возможностью, т.к. эти варианты дадут почву для объективного определения места морфа в системе языка.



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

**Программы, связанные с вычислительной лингвистикой:**

**• программы анализа и лингвистической обработки и преобразования текстов;**

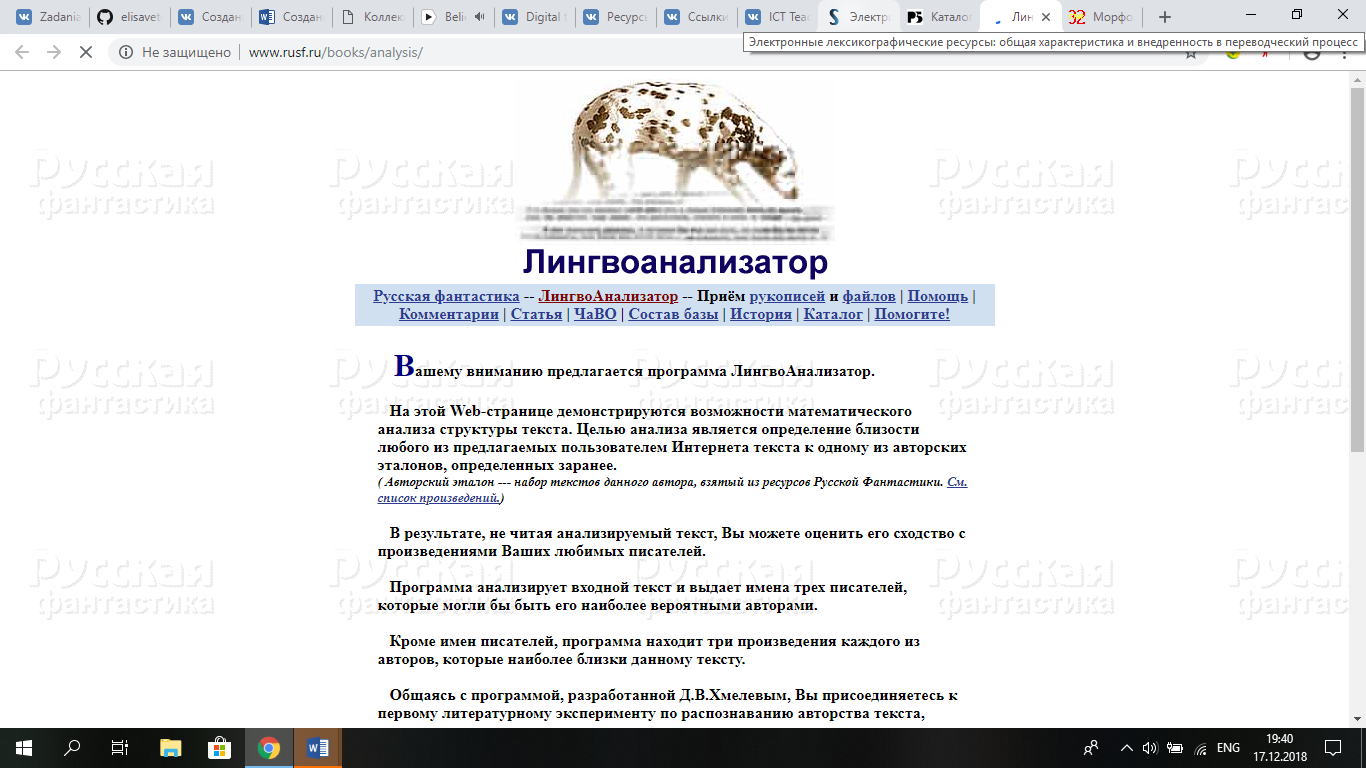
<http://opennlp.apache.org/index.html>

OpenNLP - это организационный центр "open source" проектов, связанных с машинной обработкой естественного языка под эгидой Apache. OpenNLP предлагает обширный Java-инструментарий обработки текстов на основе методов машинного обучения. Он включает средства токенизации, выделения предложений, разметки частей речи, выделения имен собственных, разбора текста и разрешения перекрестных ссылок. Имеется документация на английском языке.



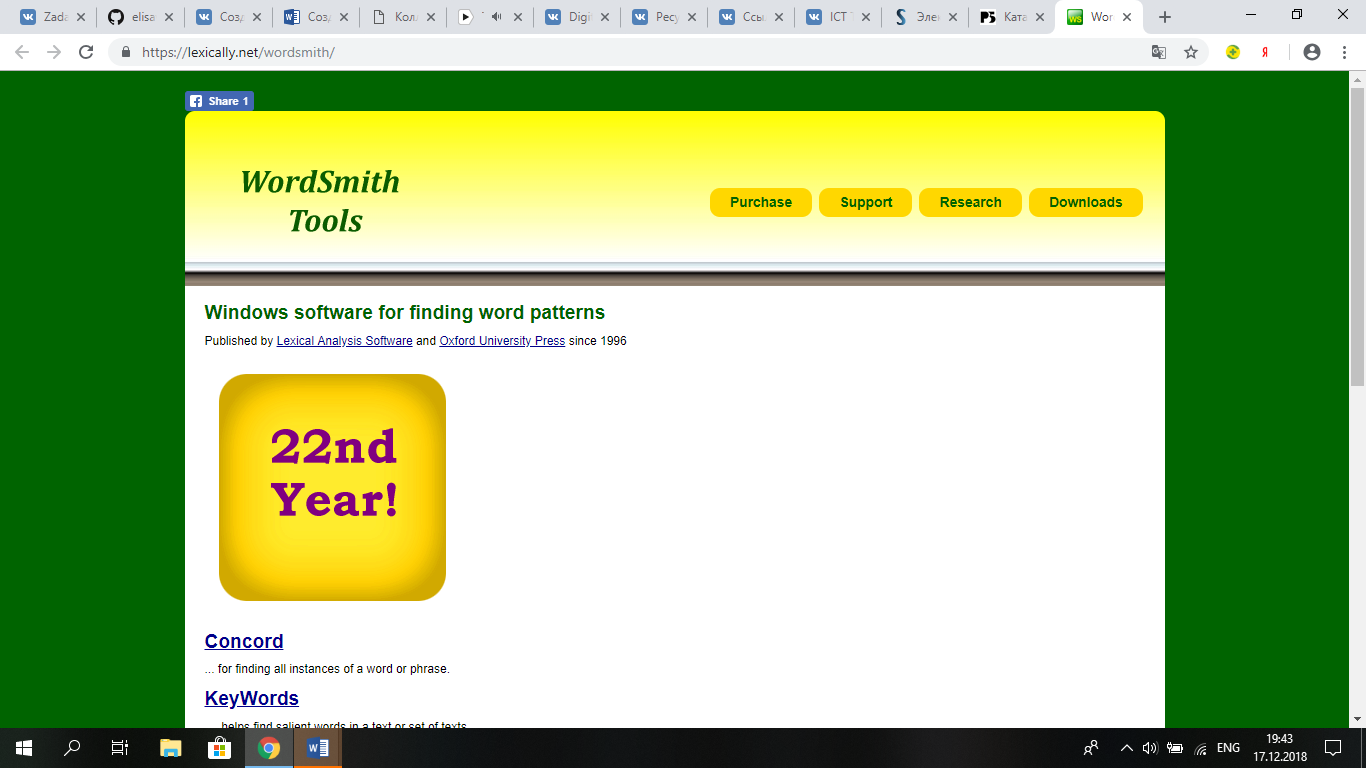
<http://www.rusf.ru/books/analysis/>

On-line версия программы математического анализа структуры текста. Целью анализа является определение близости любого из предлагаемых пользователем текстов к одному из авторских эталонов, определенных заранее. (*Авторский эталон* - это набор текстов данного автора, взятый из [ресурсов Русской Фантастики](http://www.rusf.ru/books/analysis/list.htm)). 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.



<https://lexically.net/wordsmith/>

WordSmith Tools - это многофункциональный программный комплекс для исследования поведения слов в текстах. Компонент WordList позволяет построить список отдельных слов или кластеров слов, упорядоченных по алфавиту или частоте. Компонент Concord используется для построения конкордансов. Компонент KeyWords позволяет построить список ключевых слов. Имеется настройка на тип языка в обрабатываемых текстах.

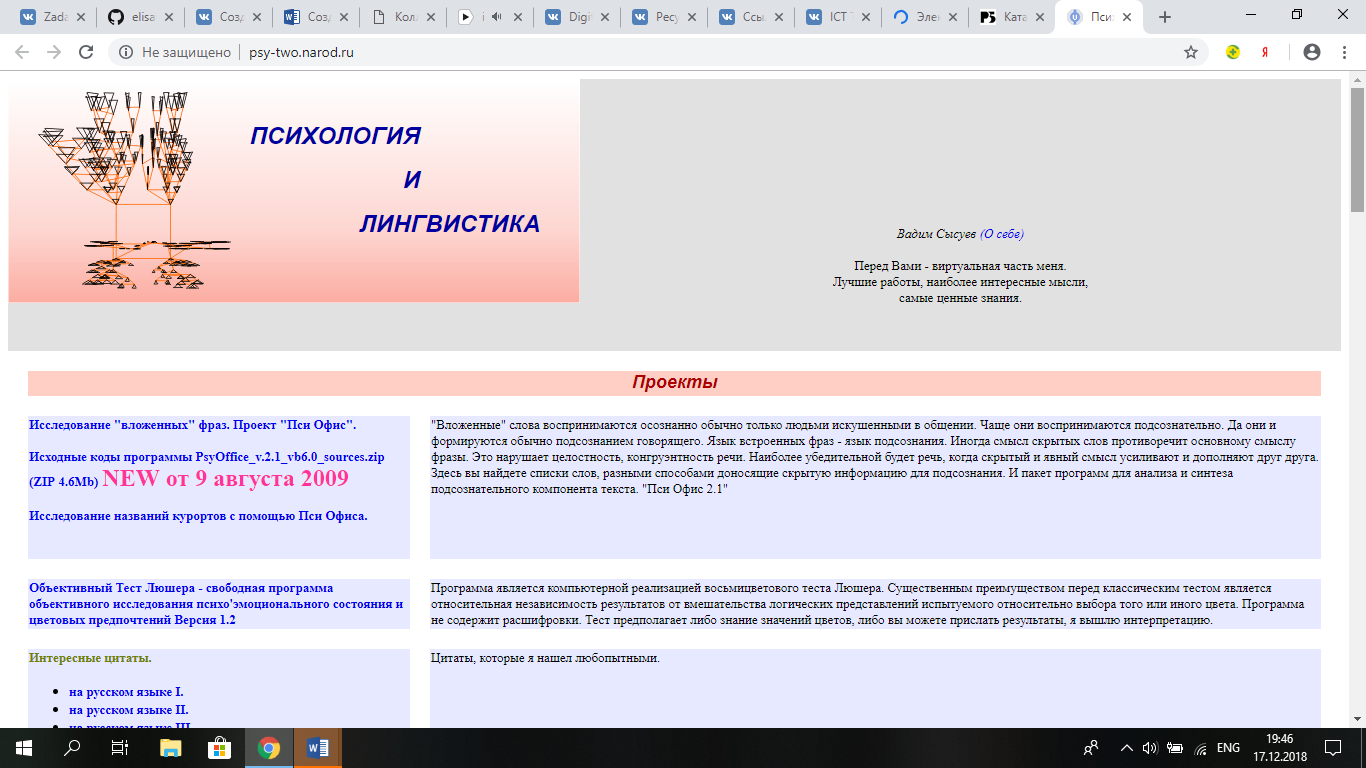


**• психолингвистические программы;**

<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.

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



<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 текстовому процессору Word для Windows.



<http://psyton.ru/index.html>

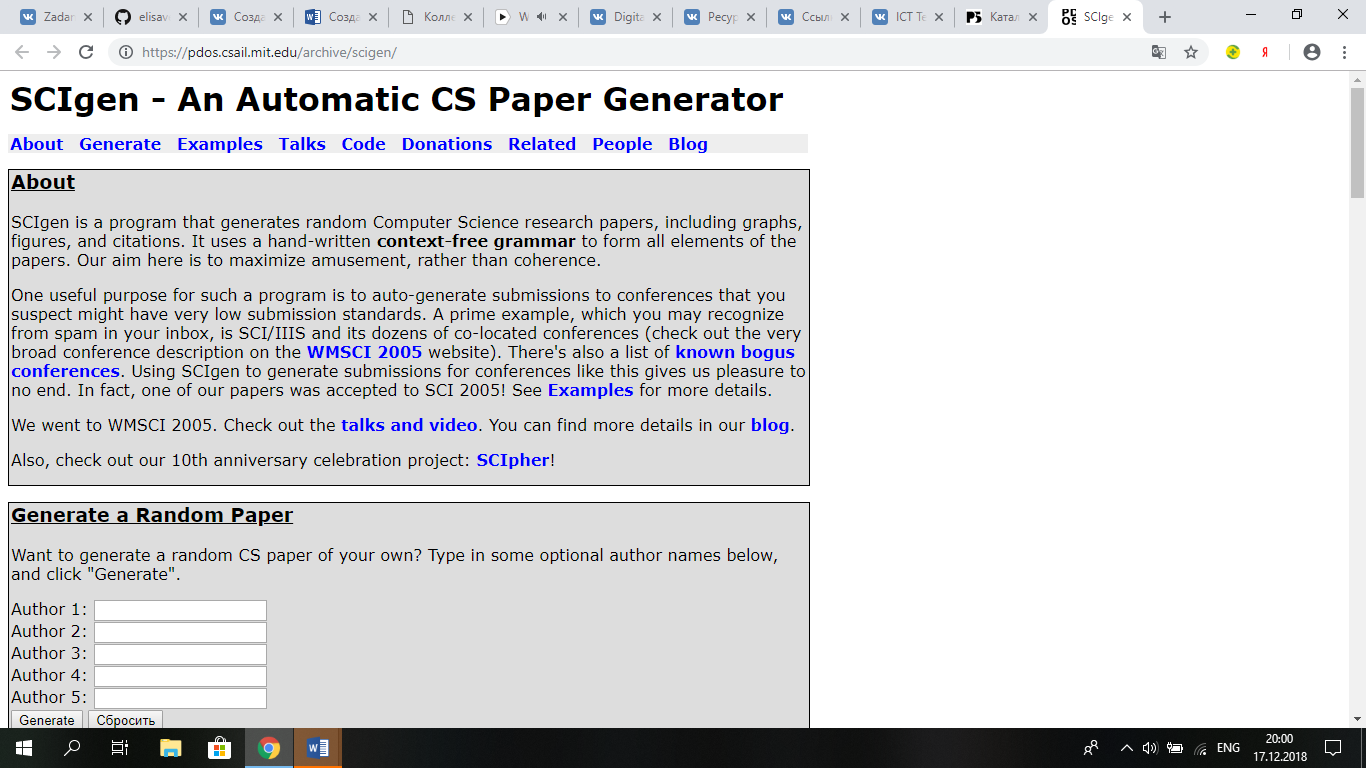
Профессиональные психодиагностические программы



**• генераторы текстов и "говорящие" программы;**

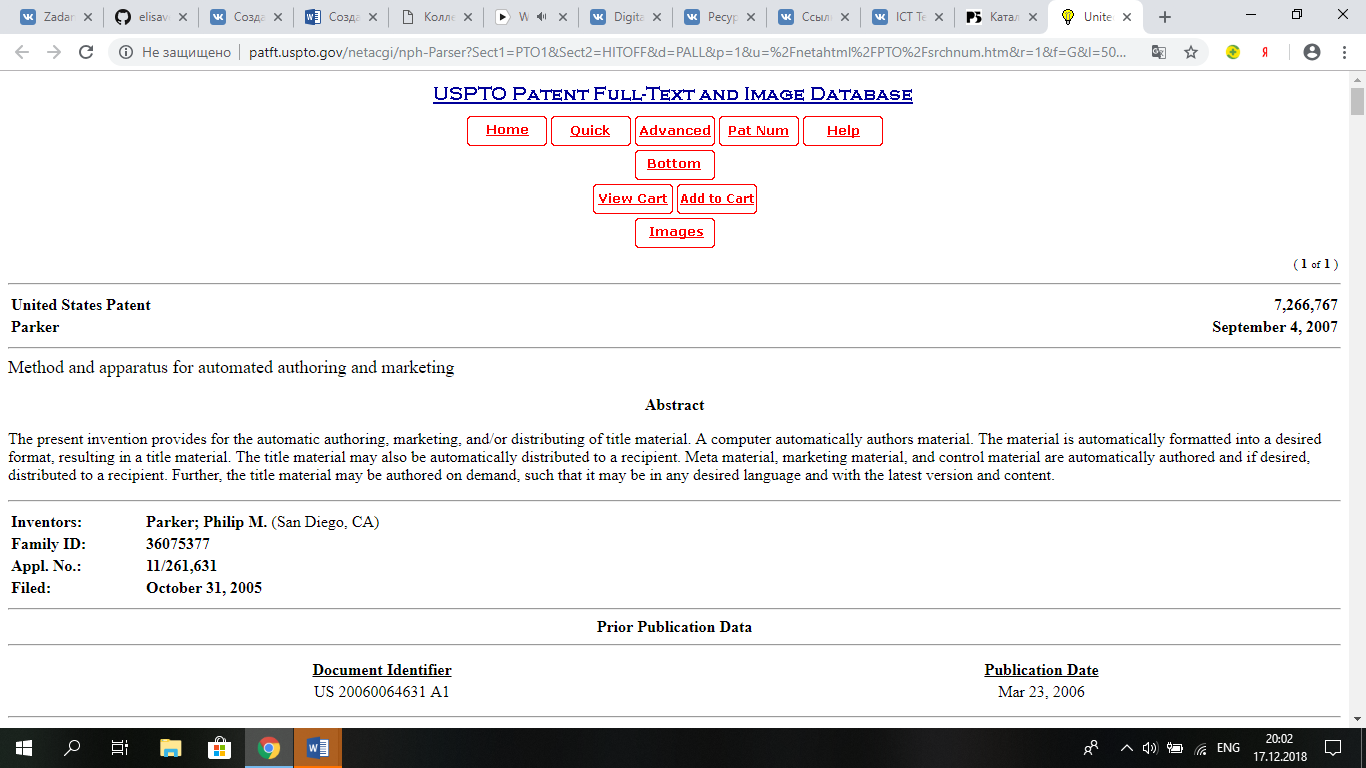
<https://pdos.csail.mit.edu/archive/scigen/>

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. Ученый Cyril Labbe из Joseph Fourier University в Гренобле выявил более 120 искусственно сгенерированных публикаций, включенных в труды 30 международных конференций за период 2008-13.



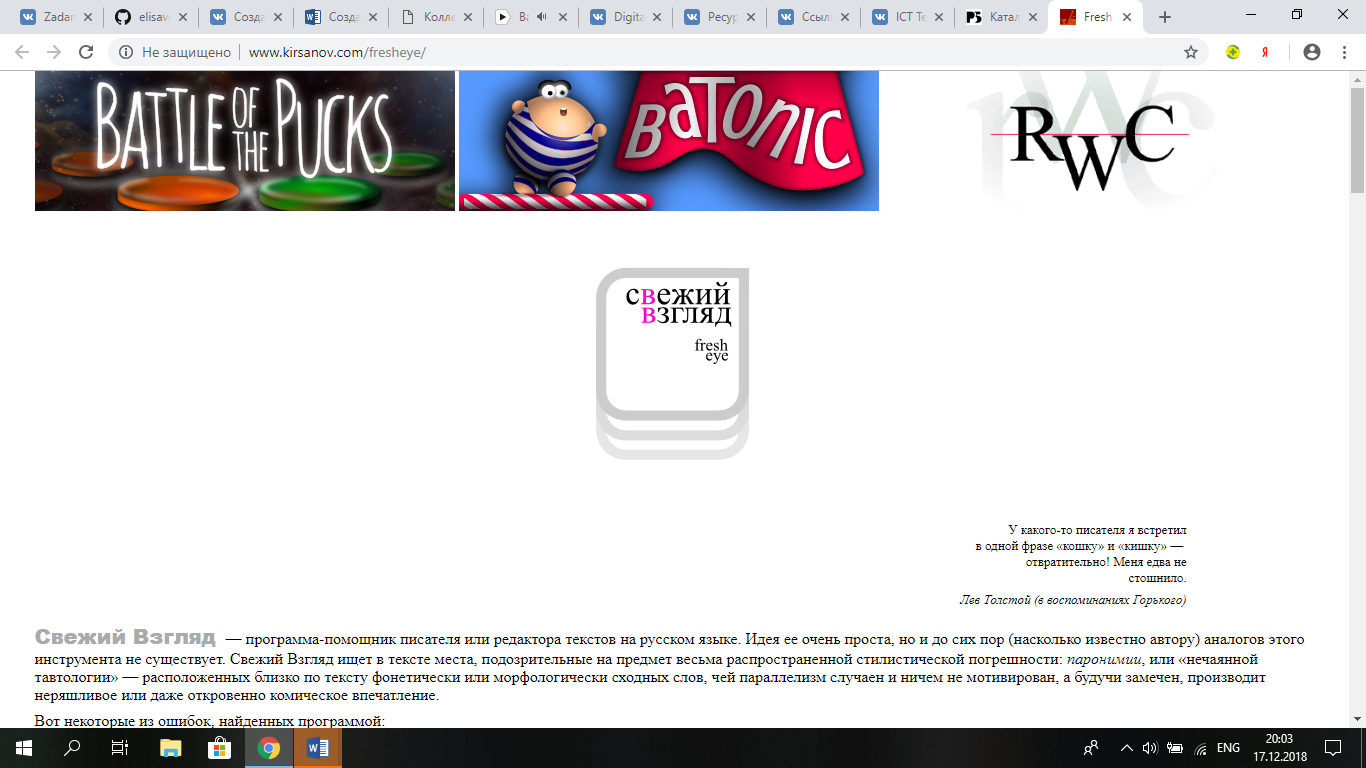
<http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PALL&p=1&u=%2Fnetahtml%2FPTO%2Fsrchnum.htm&r=1&f=G&l=50&s1=7,266,767.PN.&OS=PN/7,266,767&RS=PN/7,266,767>

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.



<http://www.kirsanov.com/fresheye/>

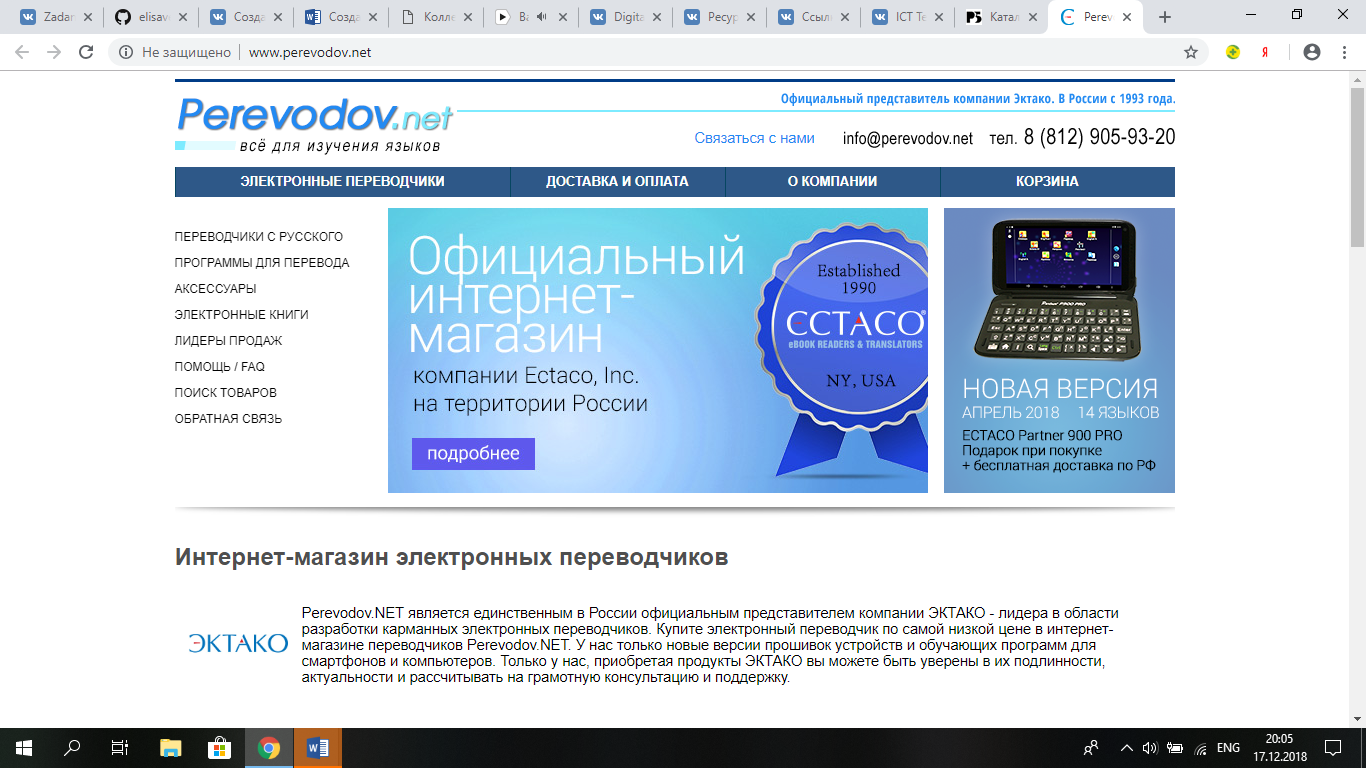
Экспериментальная программа для генерации русскоязычных стихоподобных текстов ("инструмент поэта"). Программа способна конструировать русские неологизмы на основе заданного словаря с лексико-статистической информацией. Несмотря на свою "древность", программа интересная и может позабавить своими словесными изобретениями. Прилагается утилита для создания авторских рабочих словарей.



**• системы обработки естественного языка;**

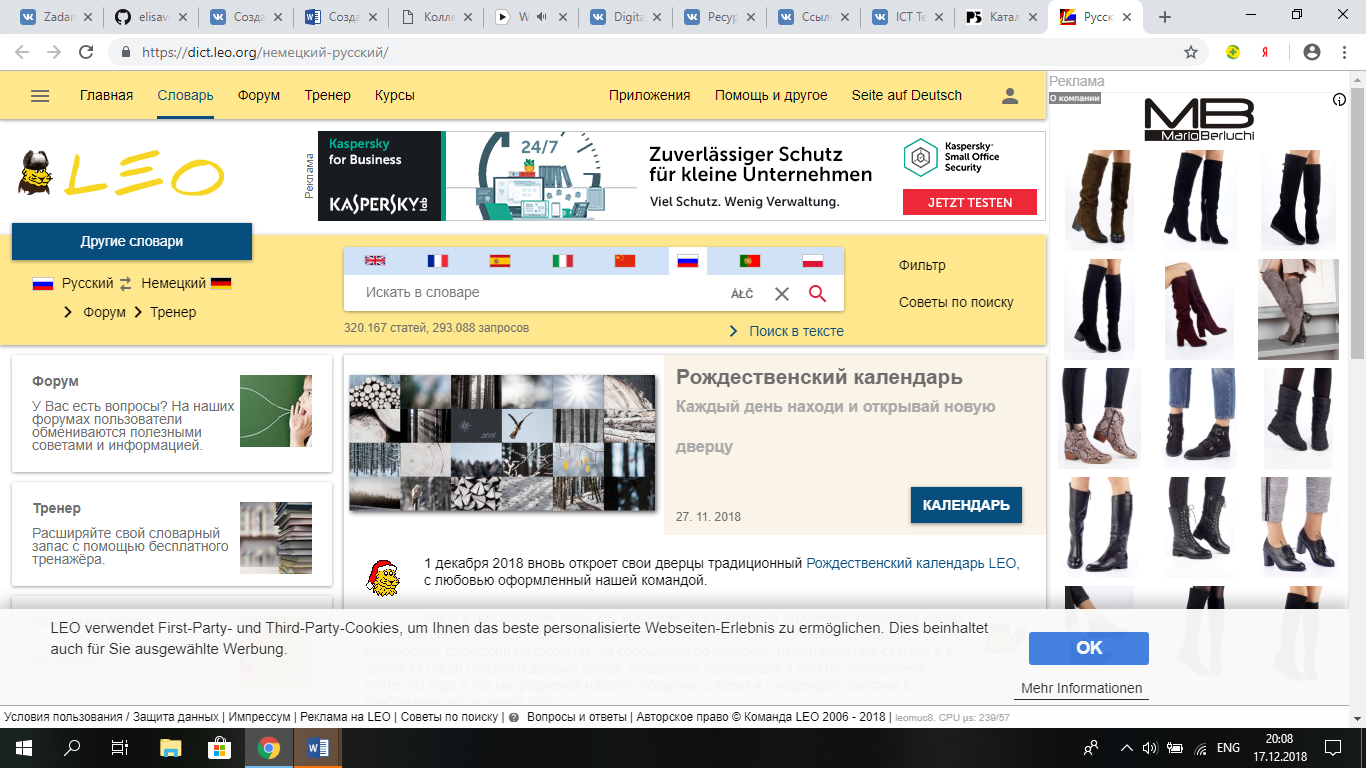
<http://www.perevodov.net/>

Система on-line перевода компании Ectaco, специализирующейся на разработке электронных карманных переводчиков. Поддерживает русский, немецкий и французский языки.



<https://dict.leo.org/%D0%BD%D0%B5%D0%BC%D0%B5%D1%86%D0%BA%D0%B8%D0%B9-%D1%80%D1%83%D1%81%D1%81%D0%BA%D0%B8%D0%B9/>

Система англо-немецкого перевода и словарь, содержащий более 3300000 словарных статей.



<https://www.translate.ru/>

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

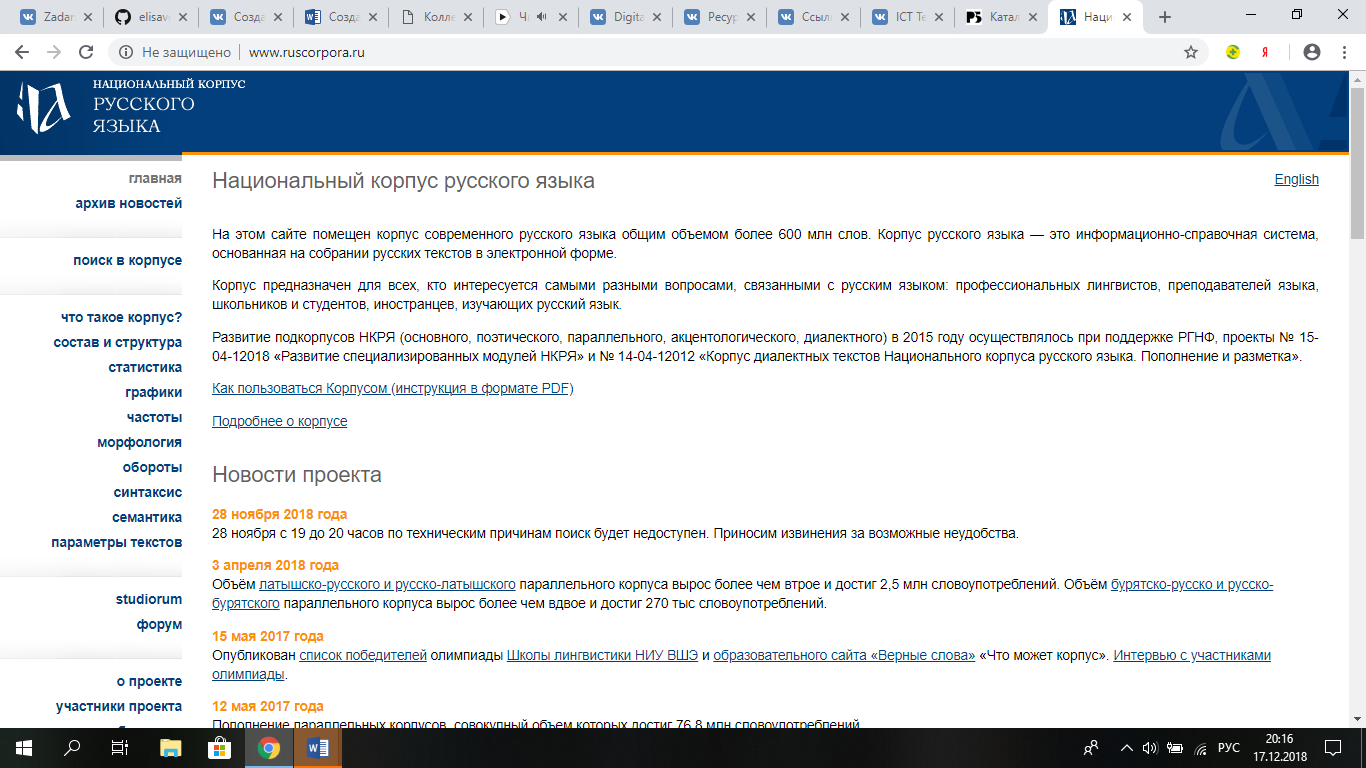


**• создание и использование текстовых корпусов, национальные корпусы языков и коллекции текстов;**

<http://www.ruscorpora.ru/>

На этом сайте помещен корпус современного русского языка общим объемом более 600 млн слов. Корпус русского языка — это информационно-справочная система, основанная на собрании русских текстов в электронной форме.

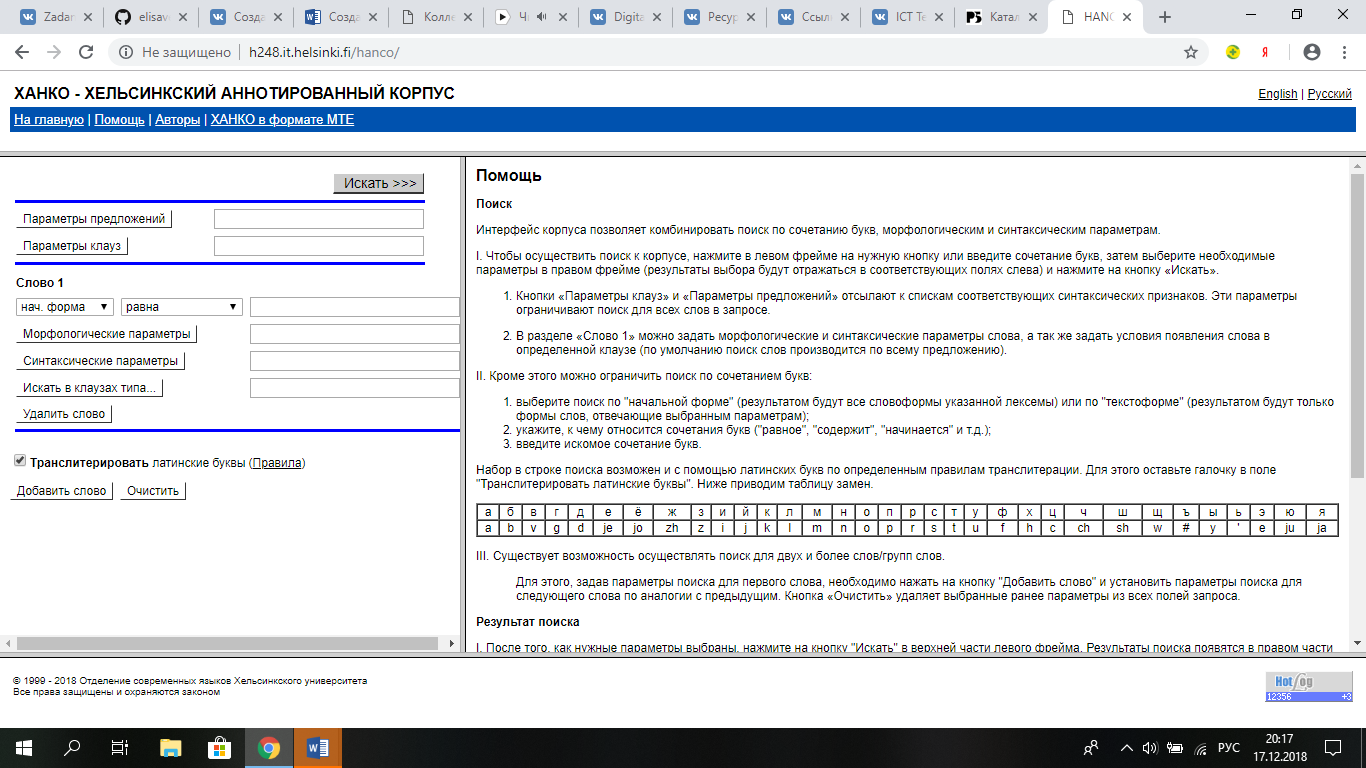
Корпус предназначен для всех, кто интересуется самыми разными вопросами, связанными с русским языком: профессиональных лингвистов, преподавателей языка, школьников и студентов, иностранцев, изучающих русский язык.



<http://h248.it.helsinki.fi/hanco/>

### Типы разметки в ХАНКО

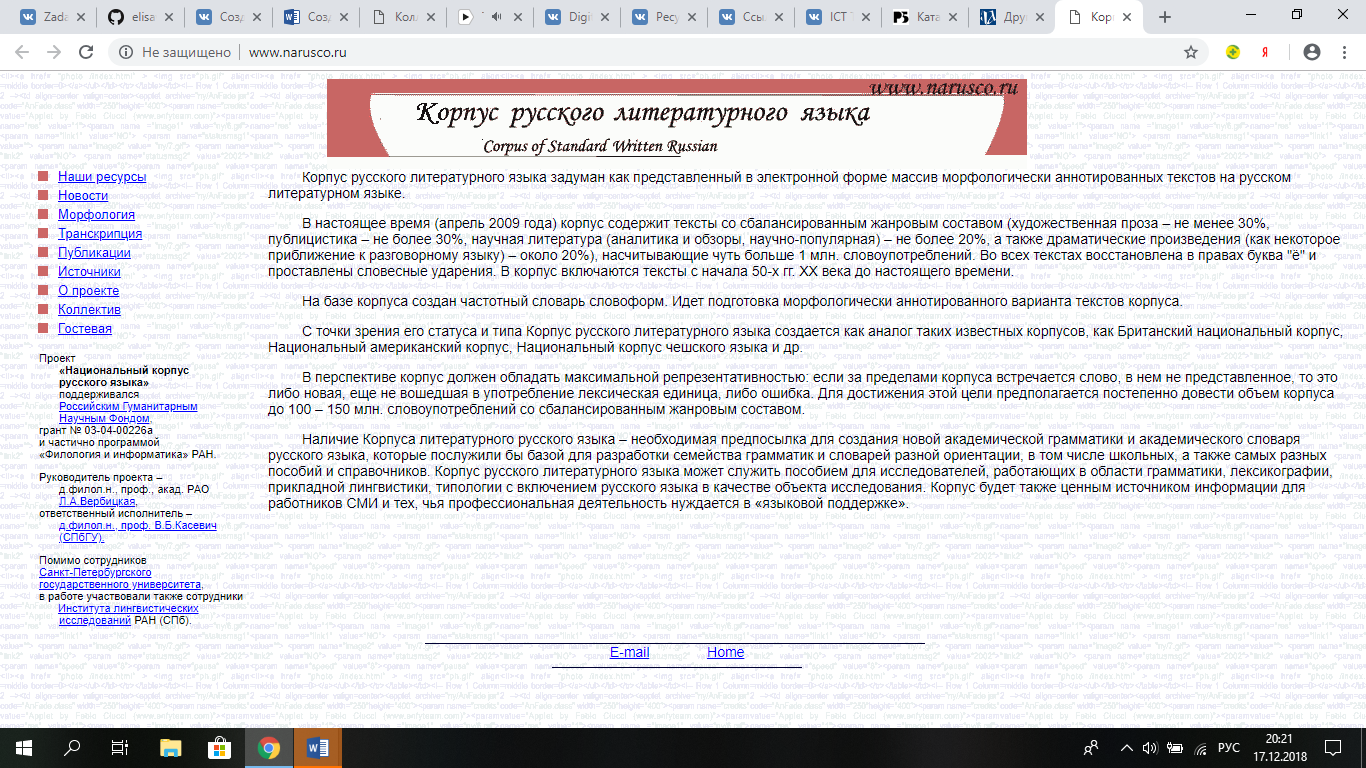
* [**Морфологическая**](https://ru.wikipedia.org/wiki/%D0%9C%D0%BE%D1%80%D1%84%D0%BE%D0%BB%D0%BE%D0%B3%D0%B8%D1%8F_(%D0%BB%D0%B8%D0%BD%D0%B3%D0%B2%D0%B8%D1%81%D1%82%D0%B8%D0%BA%D0%B0)) **разметка.** Полная морфологическая характеристика каждой текстоформы с возможностью указать спорные случаи, имеющие неоднозначную трактовку.
* [**Синтаксическая**](https://ru.wikipedia.org/w/index.php?title=C%D0%B8%D0%BD%D1%82%D0%B0%D0%BA%D1%81%D0%B8%D1%81&action=edit&redlink=1) **разметка.** Выделены два типа единиц: простое предложение (клауза) и сложное предложение. Синтаксическая разметка выполнена в рамках синтаксической теории членов предложения.
* **Метаразметка.** ХАНКО, будучи довольно однородным по внешним текстовым параметрам корпусом, содержит минимальную [метаинформацию](https://ru.wikipedia.org/wiki/%D0%9C%D0%B5%D1%82%D0%B0%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D0%B5): номер журнала, автор текста, тип текста.



<http://www.narusco.ru/>

Национальный корпус – это корпус, который обеспечивает максимально полное отражение лексики и грамматики языка. Настоящий Корпус русского литературного языка отражает употребление слов, словоформ, грамматических конструкций, словосочетаний русского литературного языка, начиная с середины 20 в. и до настоящего времени. Полнота корпуса достигается тогда, когда отсутствие в нем слова, словоформы, грамматической конструкции, более или менее устойчивого словосочетания означает, что эти единицы или явления:

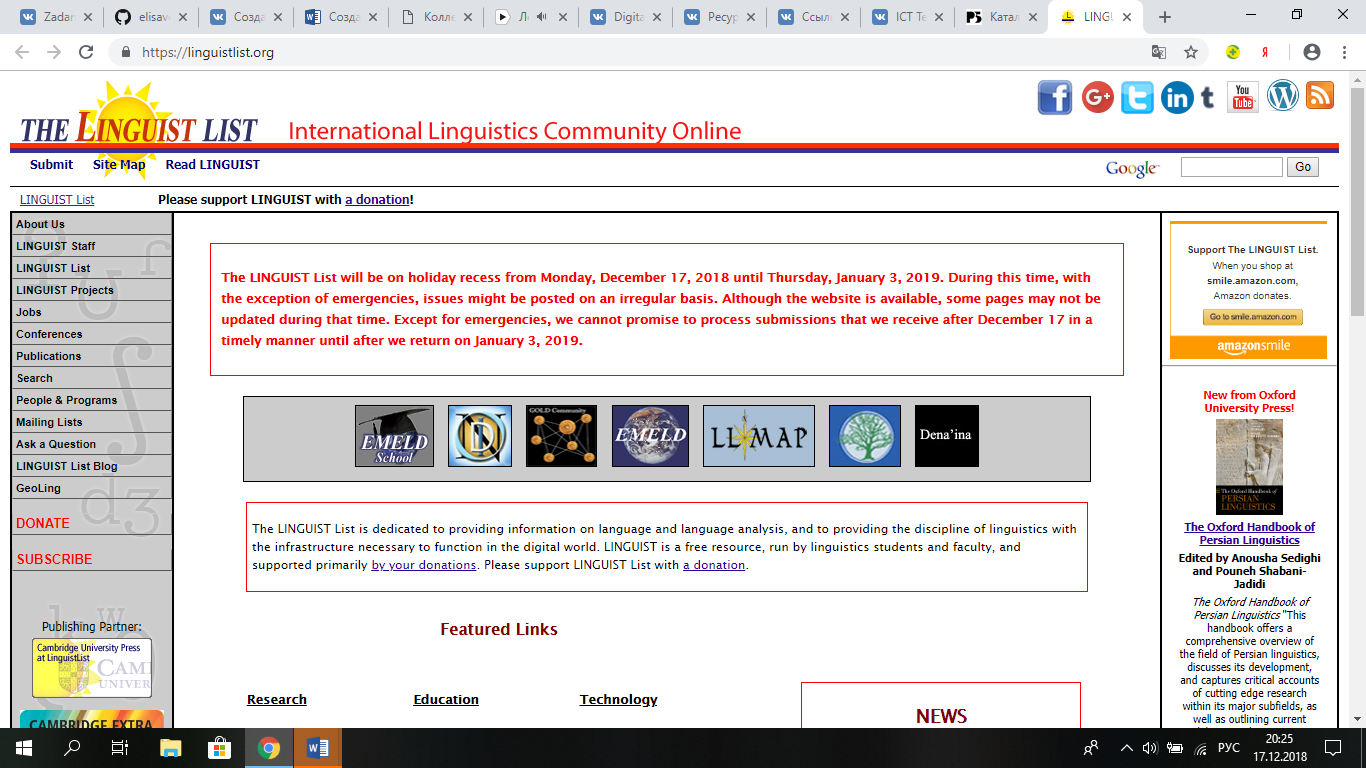
* принадлежат к сугубо индивидуальному (присущему только лишь данному автору) словоупотреблению и не встречаются в текстах других авторов;
* являются ошибкой, ненормативным употреблением;
* являются анахронизмом, явно устаревшим словоупотреблением;
* являются не ассимилированным (т.е. еще не вошедшим в русский язык) заимствованием.



**• коллекции ресурсов;**

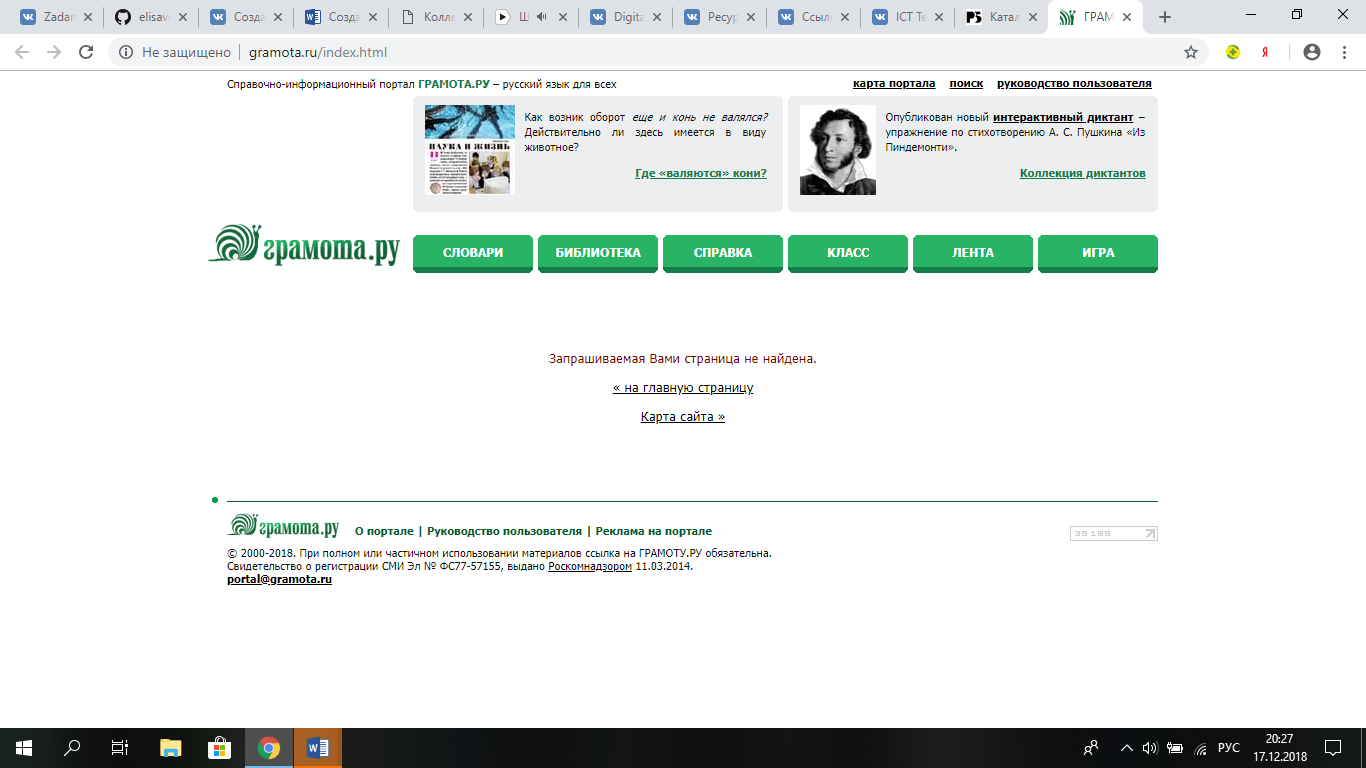
<https://linguistlist.org/>

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



<http://gramota.ru/index.html>

Замечательный ресурс для истинных любителей русской словесности, содержащий массу полезной иноформации. Также включает on-line словари русского языка:  
 - Полный электронный орфографический словарь русского языка под ред. В.В. Лопатина.  
 - Словарь трудностей произношения и ударения в современном русском языке.  
 - Новый словарь русского языка. Толково-словообразовательный.  
 - Русское словесное ударение. Словарь нарицательных имён.  
 - Русское словесное ударение. Словарь имён собственных.  
На сайте опубликован исчерпывающий [указатель ссылок](http://www.gramota.ru/links.html) на интернет-ресурсы, посвященные русскому языку и литературе.



<http://www.textanalysis.info/>

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.



**• электронные словари и тезаурусы,**

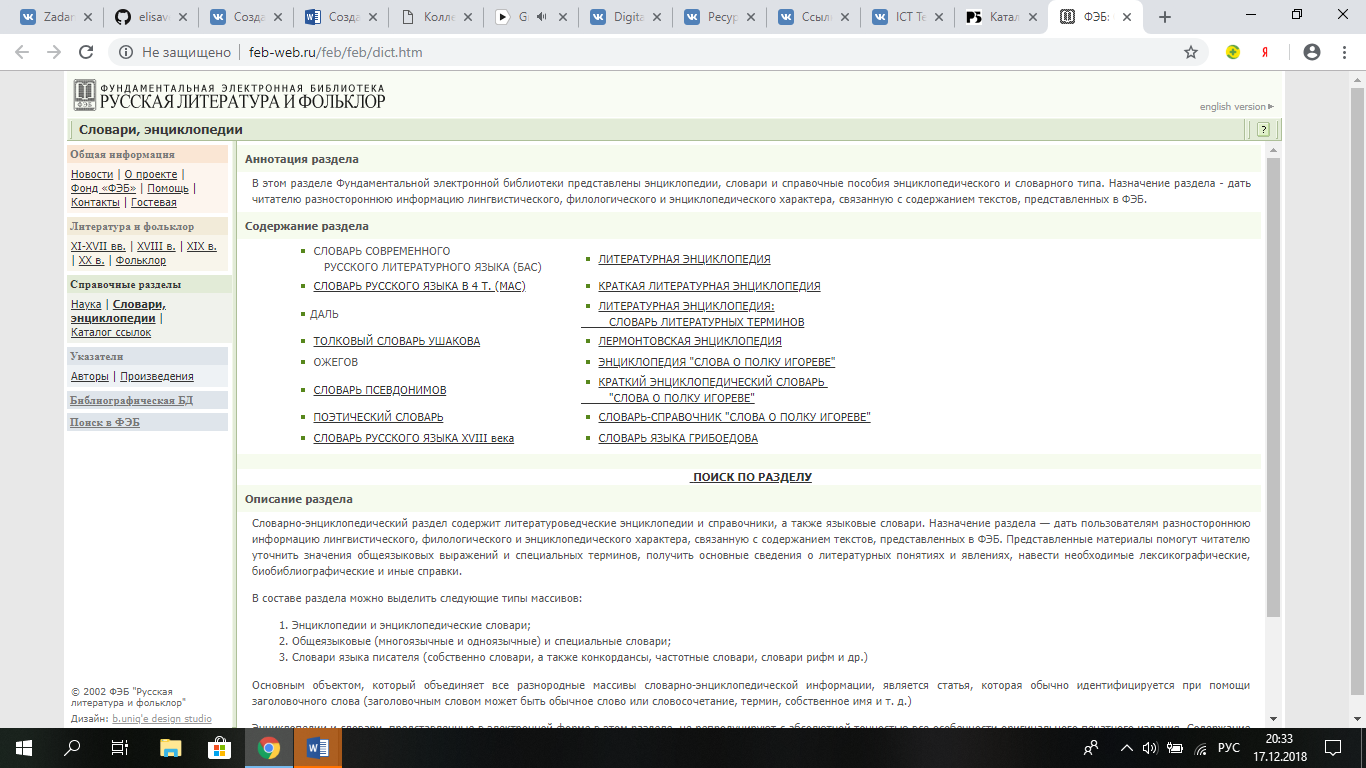
<https://www.babla.ru/>

On-line словарь для 25 языков, включая русский и китайский. Перевод включает различные варианты (синонимы), произношение и примеры употребления. Кроме словаря сайт содержит другие полезные разделы: поиск предложных оборотов, спряжение глаголов, фразы для различных категорий лексики (составление резюме, написание научных работ, деловая переписка, бизнес, путешествия), тесты, упражнения и многое другое.



<http://feb-web.ru/feb/feb/dict.htm>

Энциклопедии, словари и справочные пособия на сайте Фундаментальной Электронной Библиотеки (ФЭБ).



<http://www.ets.ru/udict-r.htm>

Коллекция on-line словарей: анатомический (латынь), уфологический, сокращений, иностранных слов, финско-русский, немецко-русский, англо-русский фразеологический и ряд других. Кроме этого на сайте представлены различные словари Polyglossum и языковые программы.

