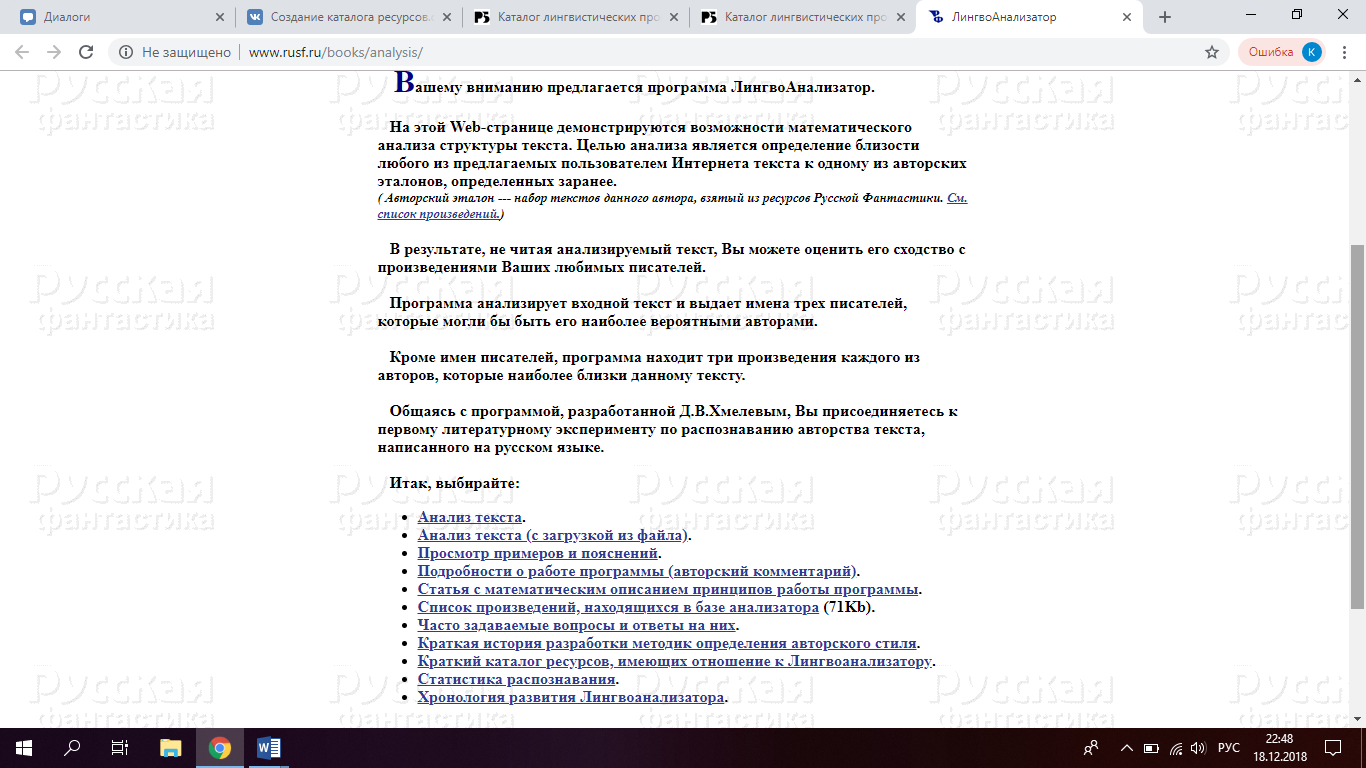
- Компьютерная обработка и анализ текста

|  |  |
| --- | --- |
| [Лингвоанализатор](http://www.rusf.ru/books/analysis/index.htm) Д.В Хмелев  http://www.rusf.ru/books/analysis/ | [Д.В.Хмелев](mailto:dima@vvv.srcc.msu.su) |

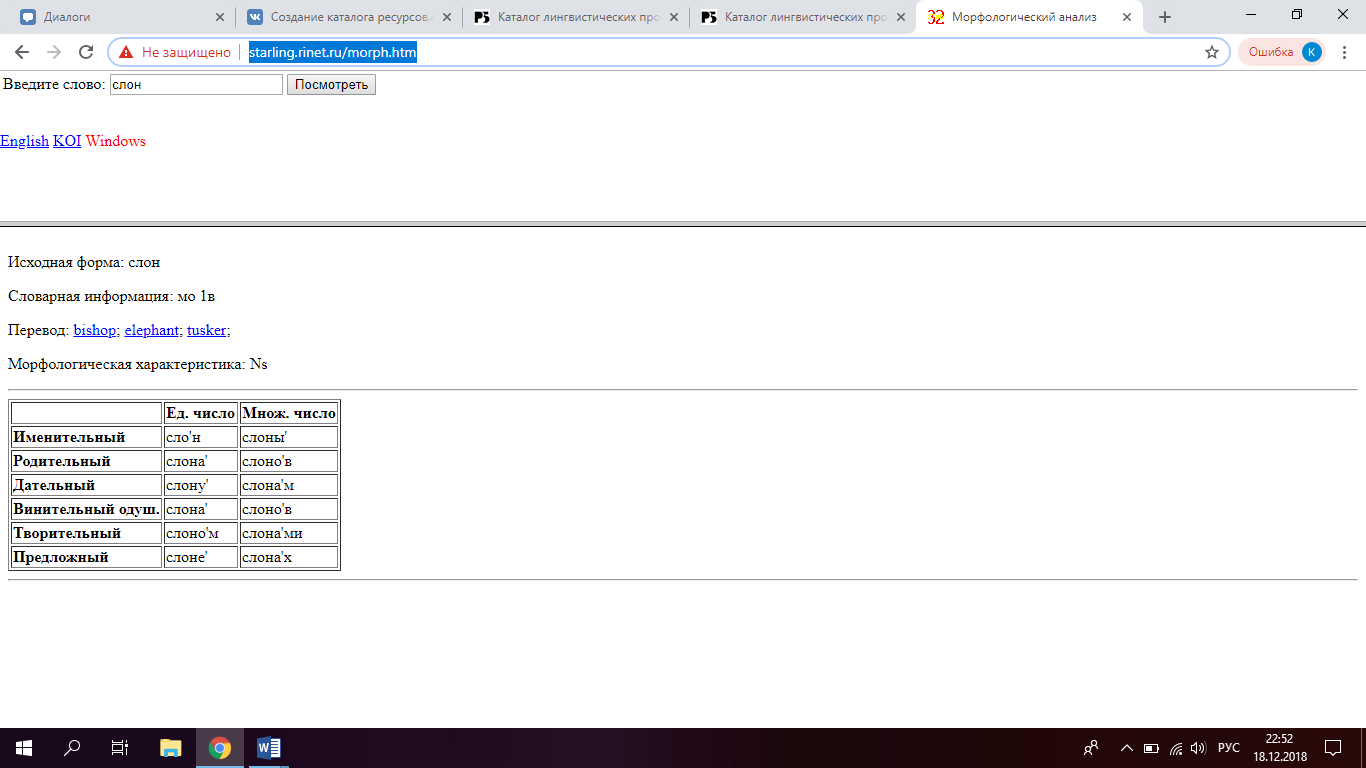
On-line версия программы математического анализа структуры текста. Целью анализа является определение близости любого из предлагаемых пользователем текстов к одному из авторских эталонов, определенных заранее. (*Авторский эталон* - это набор текстов данного автора, взятый из [ресурсов Русской Фантастики](http://www.rusf.ru/books/analysis/list.htm)).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



[Морфологический анализатор](http://starling.rinet.ru/morph.htm)

<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.).



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

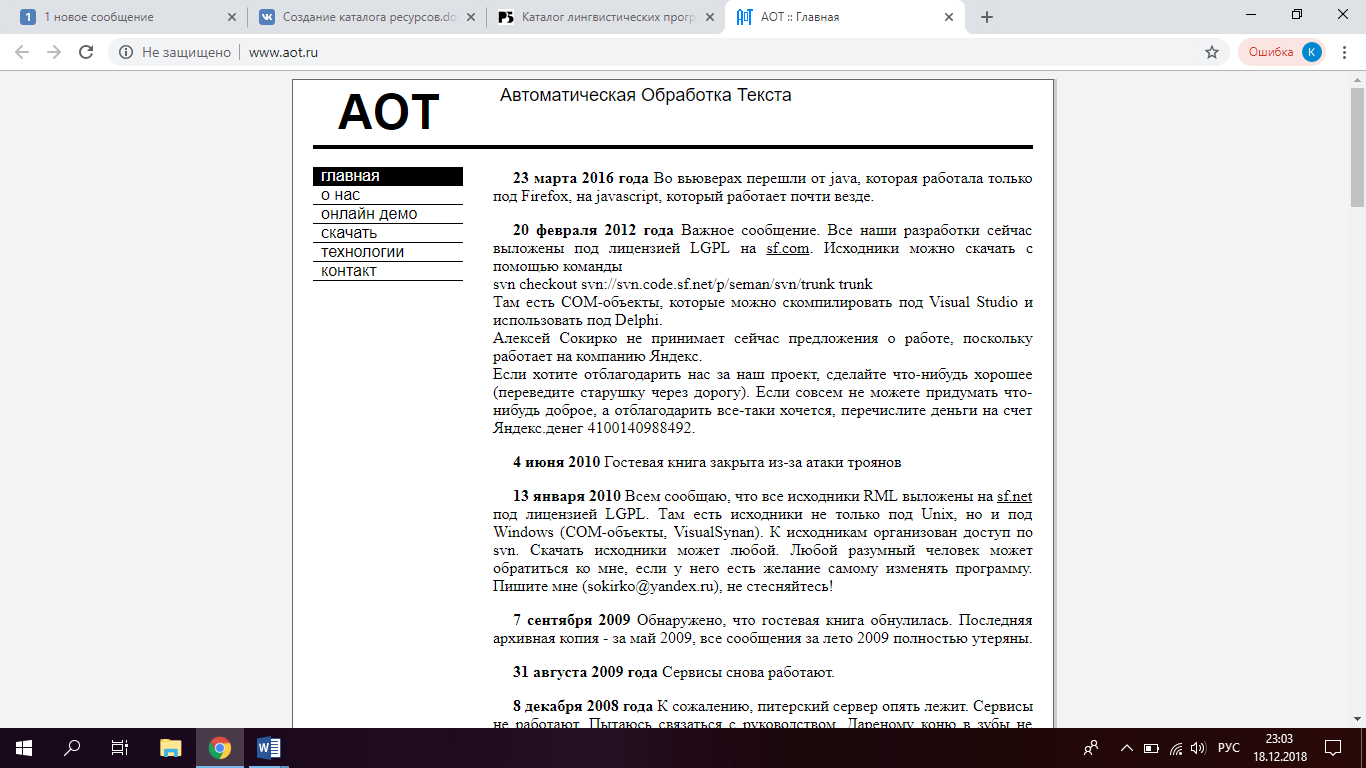
Программы, связанные с вычислительной лингвистикой:

• программы анализа и лингвистической обработки и преобразования текстов;

**АОТ**

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

Сайт, на котором представлены разработки бывших сотрудников компании [Диалинг](https://rvb.ru/soft/catalogue/catalogue.html" \l "_dialing), 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.



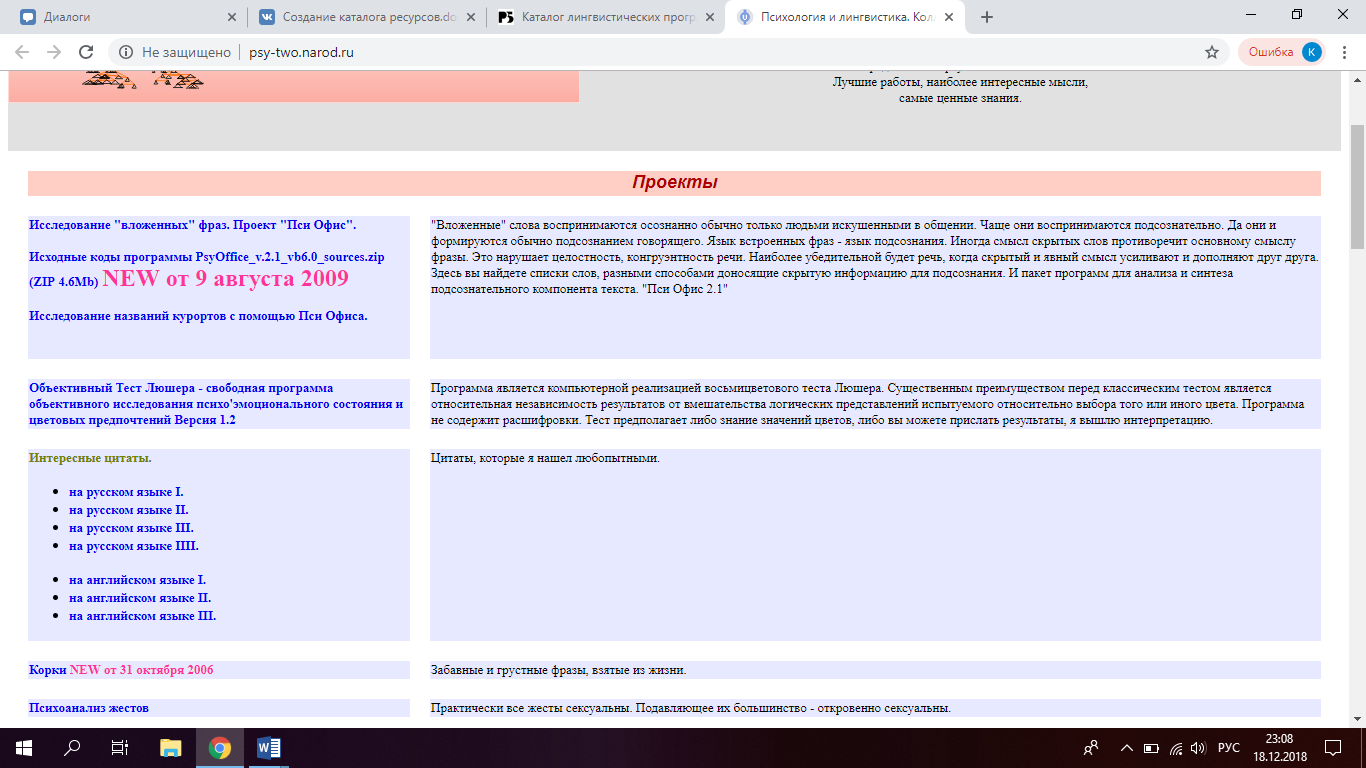
**• психолингвистические программы**

[ПСИ-Офис версия 2.1](http://psy-two.narod.ru/)

**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.

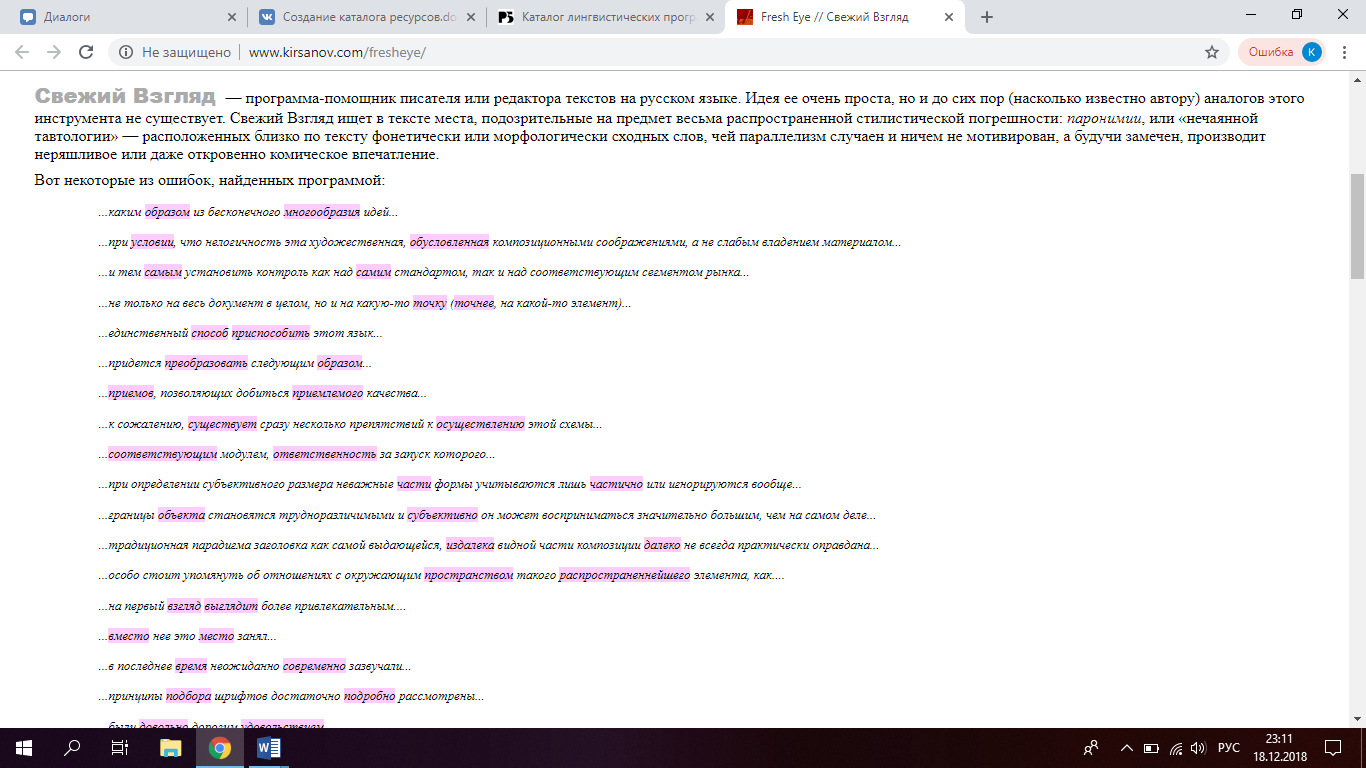
Поддержка русского (словарь 690 тыс.слов) и английского (словарь 430 тыс.слов) языков. Возможность подключения дополнительных словарей. Работает в Windows 9x/2000/NT. Freeware.



**Генераторы текстов и «говорящие» программы**

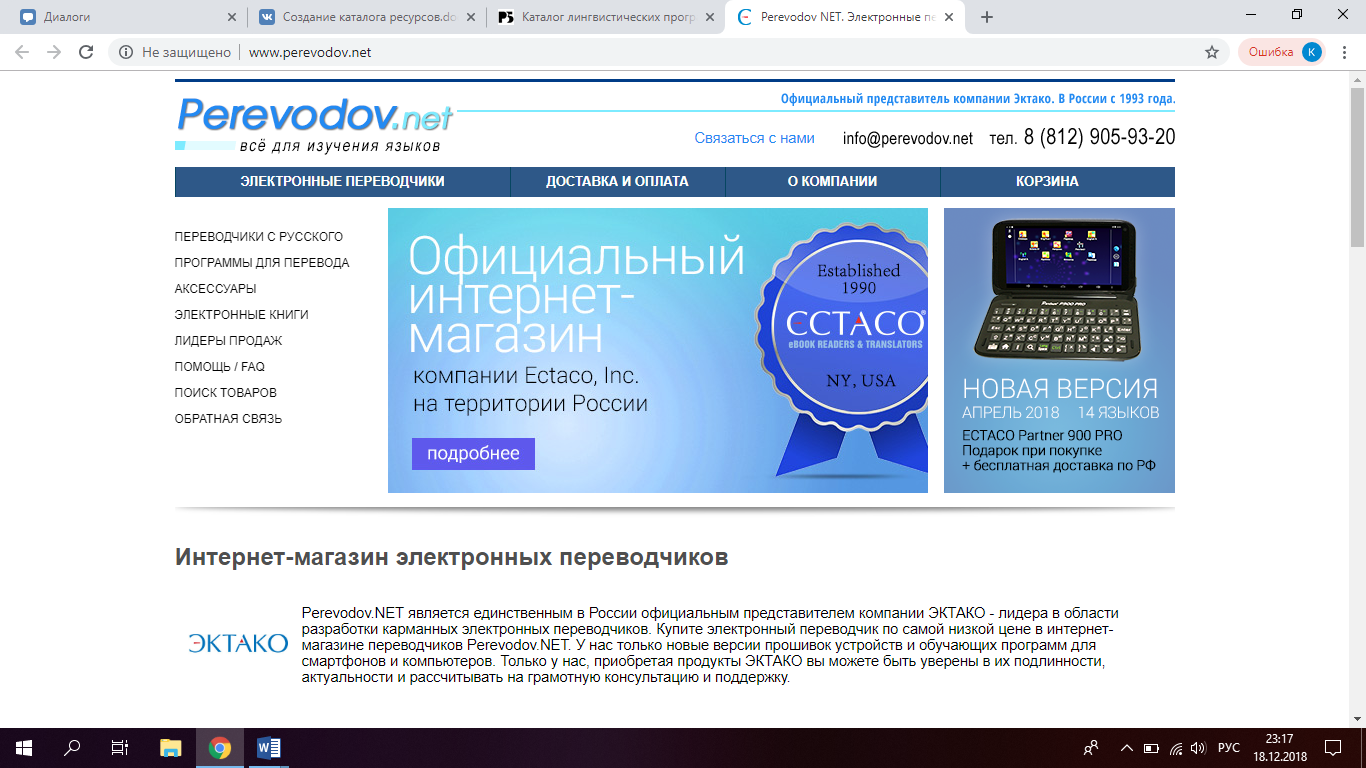
[Russian Word Constructor](http://www.kirsanov.com/fresheye/)(RWC)  
версия 1.0, 1992

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**

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

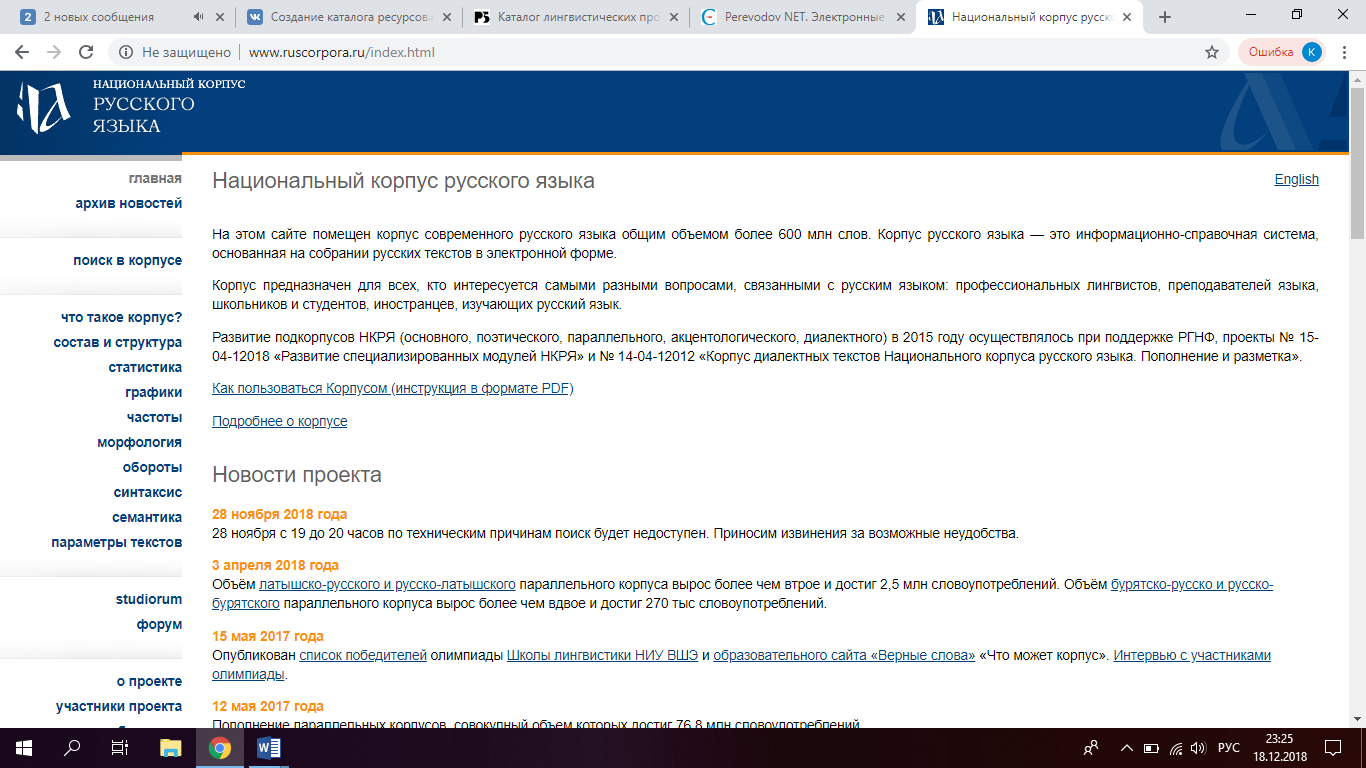


**• создание и использование текстовых корпусов, национальные корпусы языков и коллекции текстов**

<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.

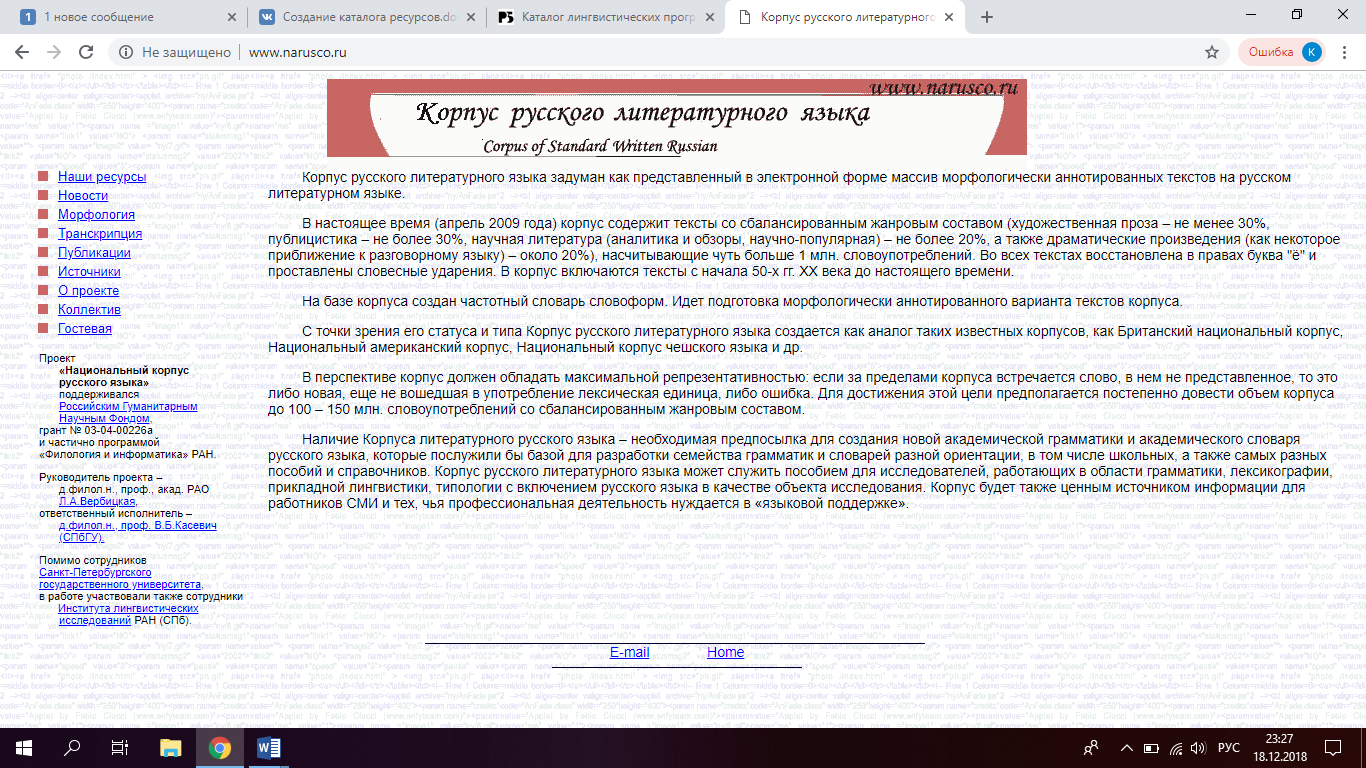
Корпус предназначен для всех, кто интересуется самыми разными вопросами, связанными с русским языком: профессиональных лингвистов, преподавателей языка, школьников и студентов, иностранцев, изучающих русский язык.



<http://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.



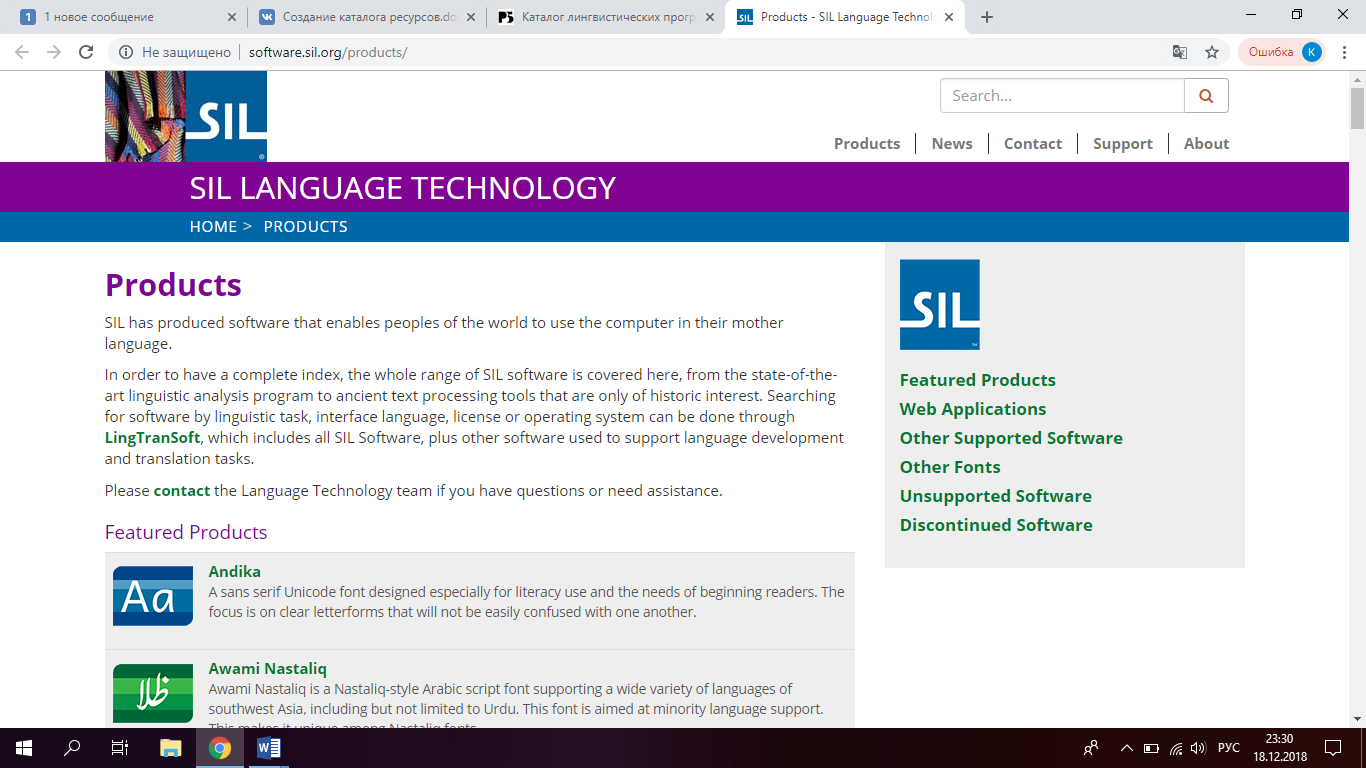
**Коллекции ресурсов**

Каталог программ по вычислительной лингвистике SIL International

SIL International (Summer Institute of Linguistics)

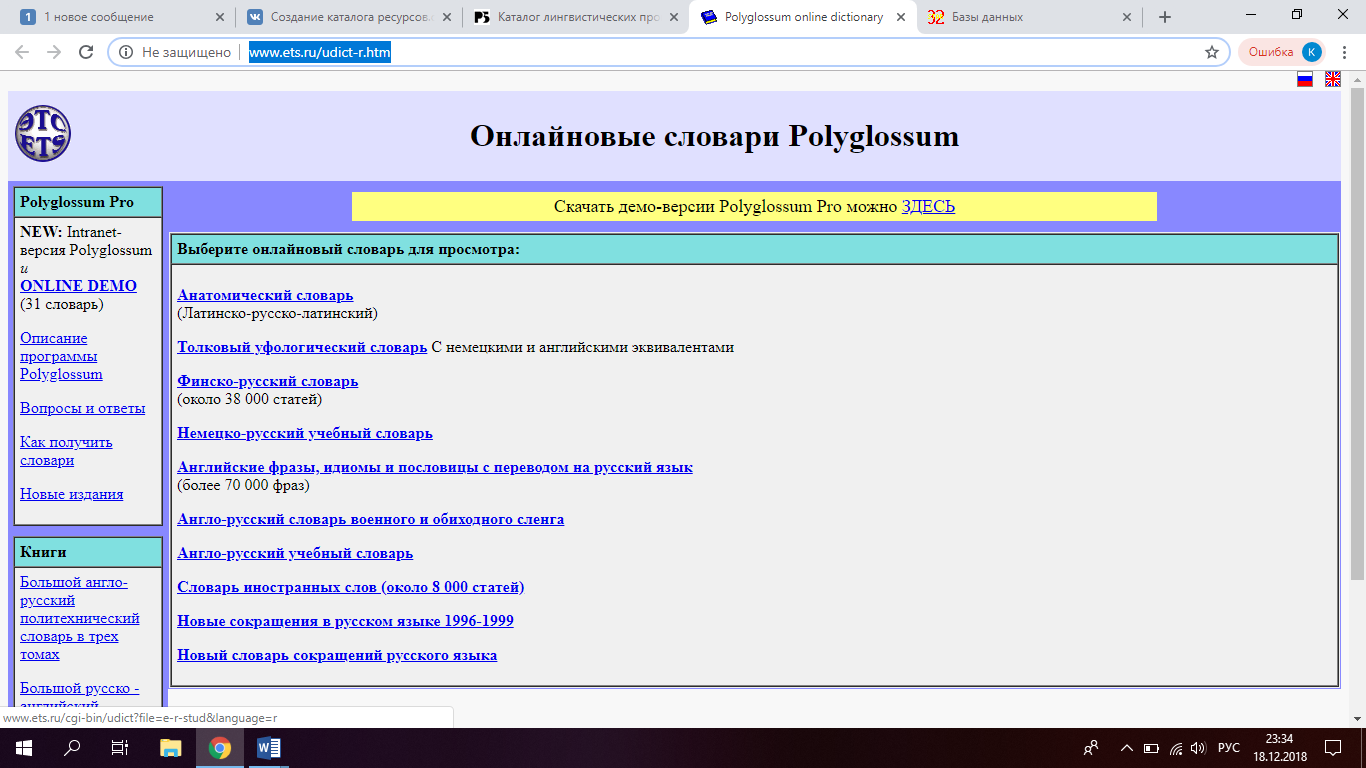
<http://software.sil.org/products/>

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



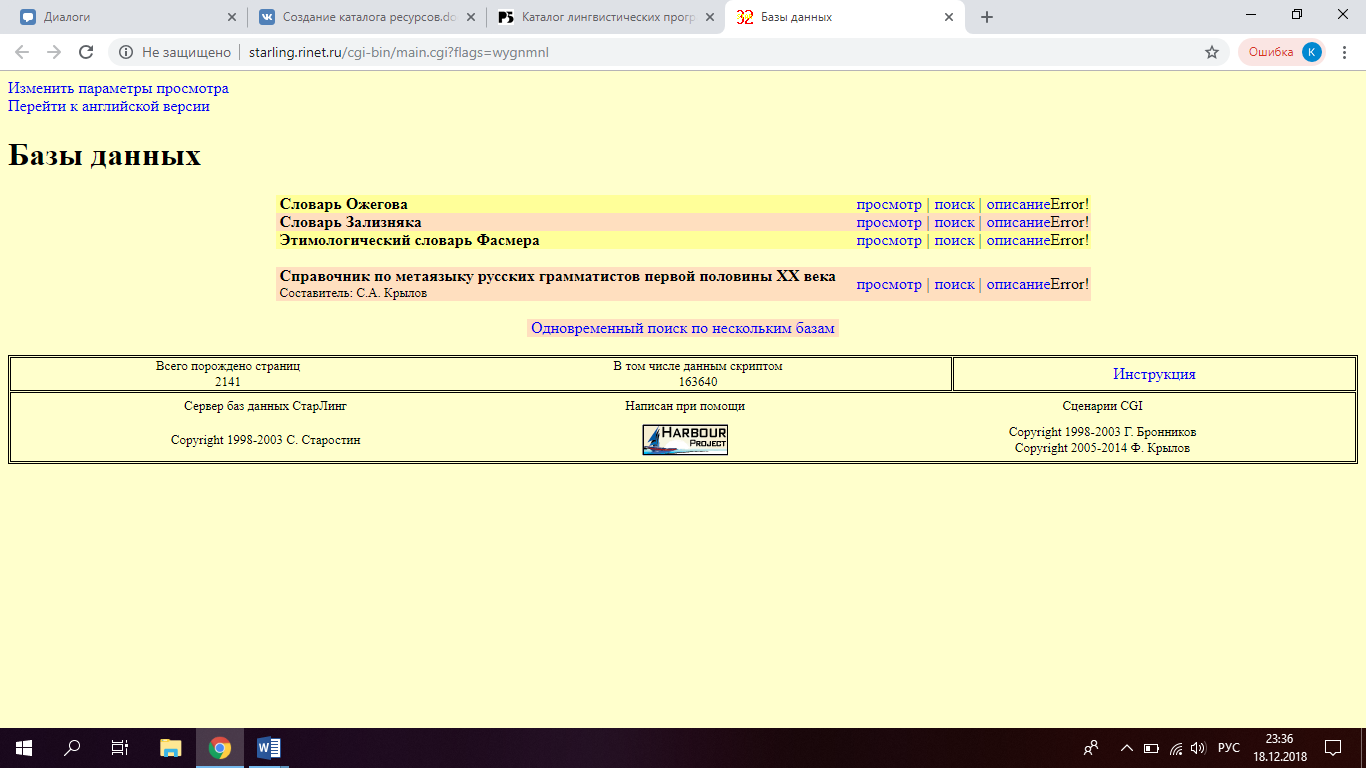
[On-line словари Издательского Дома "ЭТС"](http://www.ets.ru/udict-r.htm)

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



[Словари Ожегова и Зализняка](http://starling.rinet.ru/cgi-bin/main.cgi?flags=wygnmnl)

<http://starling.rinet.ru/cgi-bin/main.cgi?flags=wygnmnl>



[Толковый словарь Merriam-Webster](http://www.m-w.com/)

<https://www.merriam-webster.com/>

Онлайн версия известного словаря

