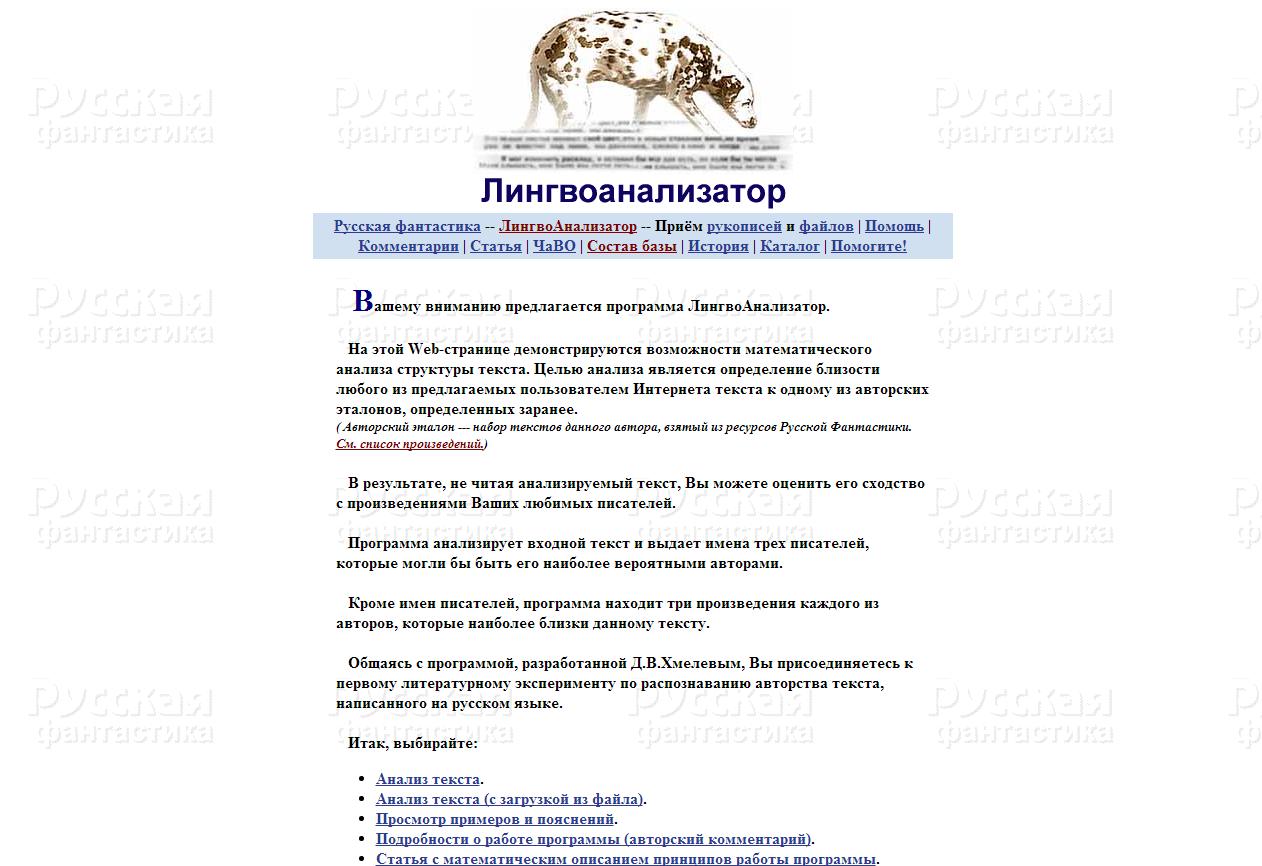
Каталог лингвистических программ и ресурсов в Cети

# Программы анализа и лингвистической обработки текстов

1. Лингвоанализатор (Д. В. Хмелев)

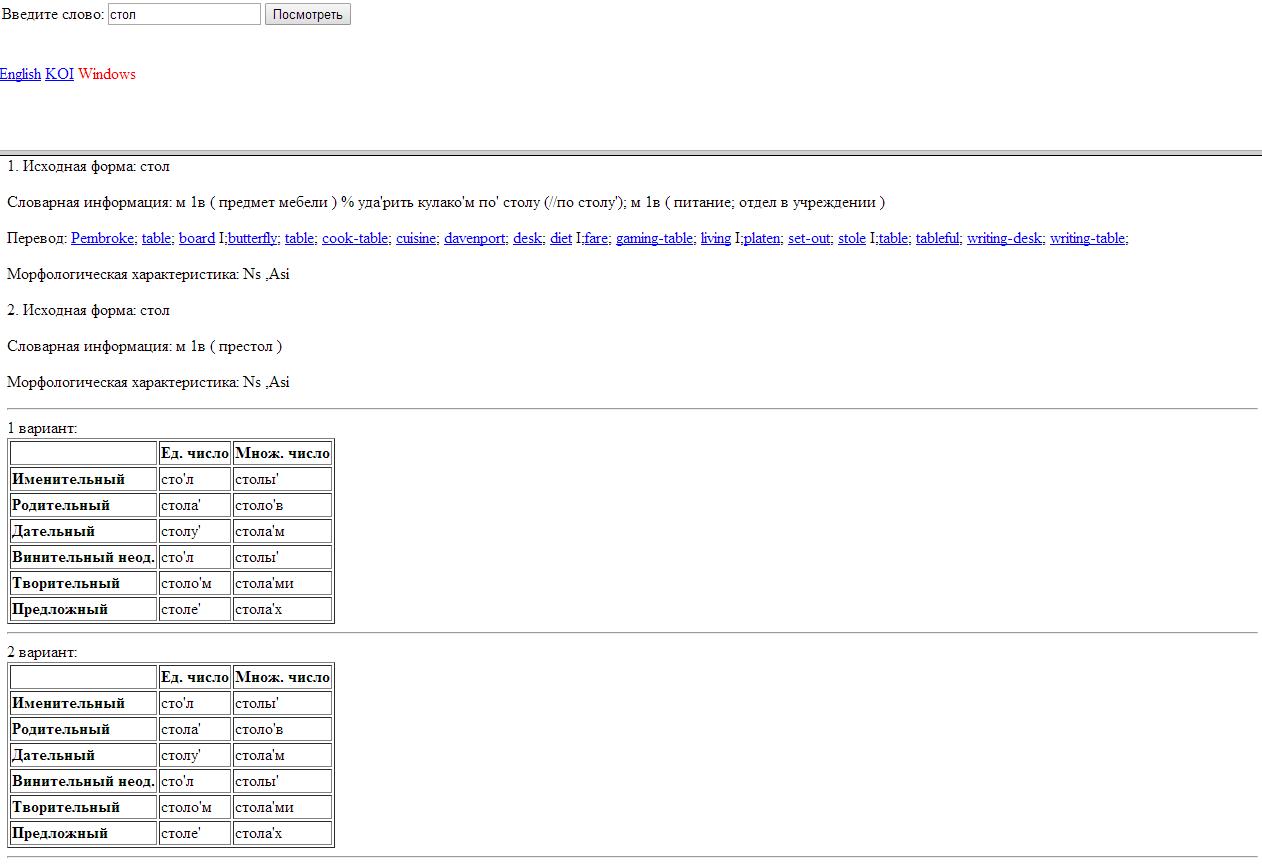
<http://www.rusf.ru/books/analysis/>



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.

1. Морфологический анализатор (С. А. Старостин)

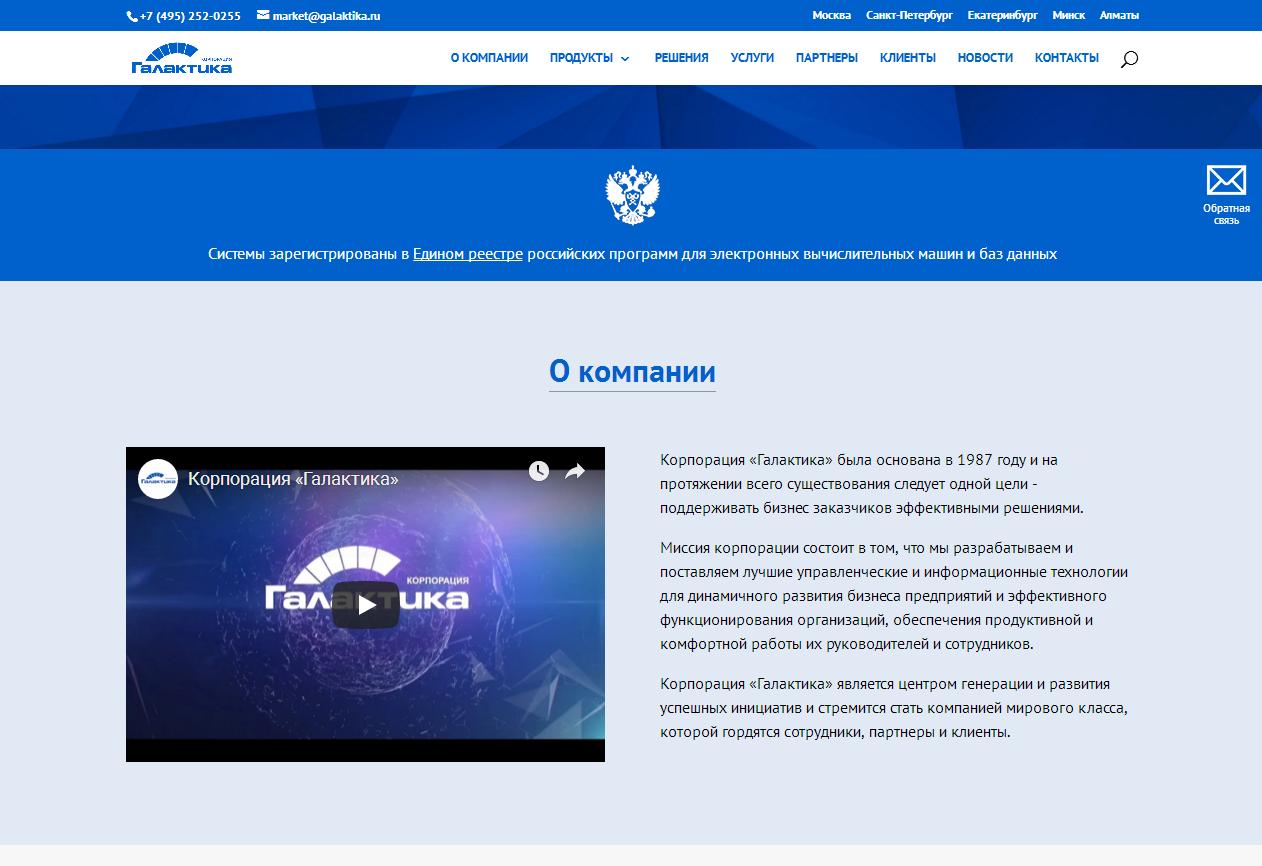
<http://starling.rinet.ru/morph.htm>



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. Galaktika-ZOOM

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



Автоматизированная система поиска и аналитической обработки информации. Мощный инструмент анализа и обработки текста (Text Mining), позволяющий извлекать необходимые сведения из огромного объема данных. Это коммерческая система, имеющая клиентов в рекламе, органах управления и средствах массовой информации.

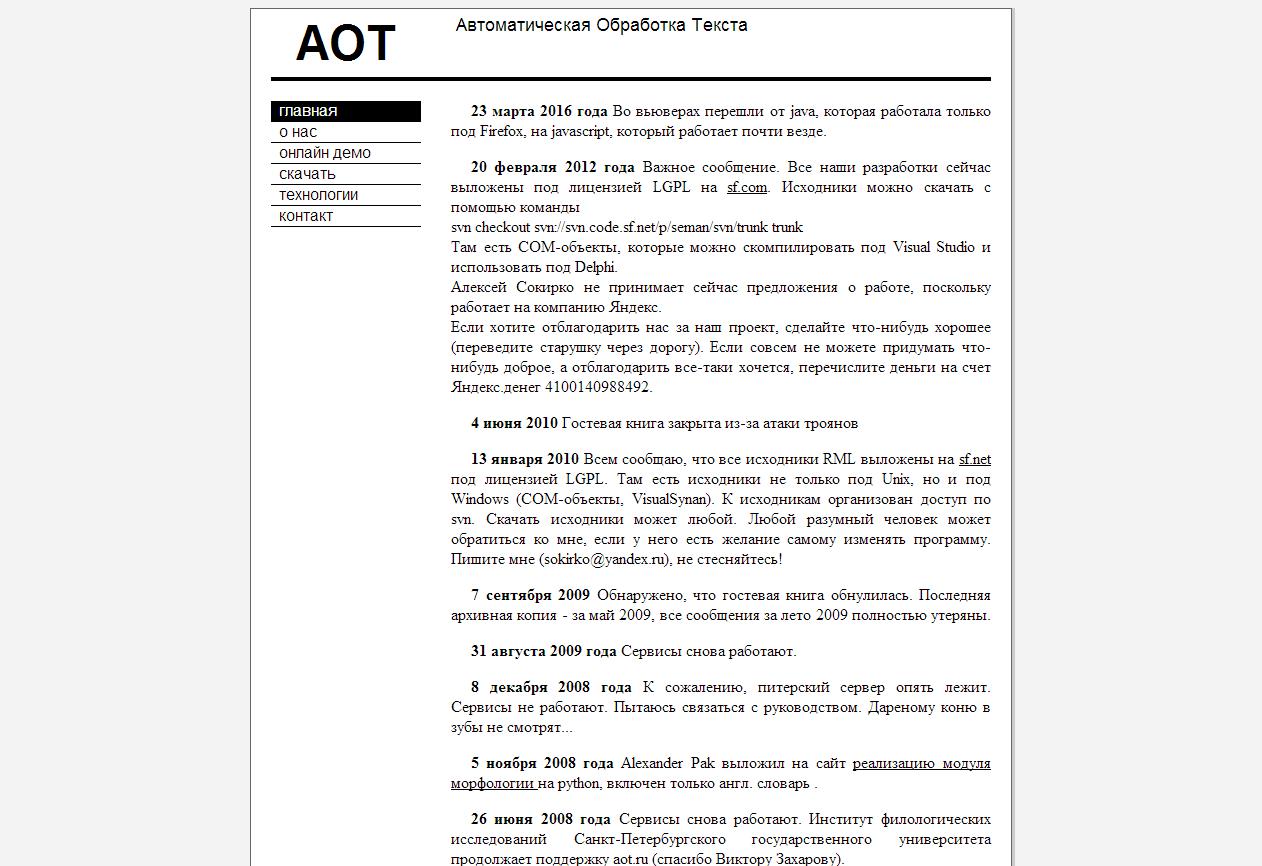
1. [Свежий взгляд/Fresh Eye](http://www.kirsanov.com/fresheye/) (версия 1.21, 1995) (Дмитрий Кирсанов)

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

  
DOS-утилита, реализующая стилистическую проверку русскоязычных текстов. Программа отыскивает в тексте места, где фонетически и морфологически схожие слова расположены в непосредственной близости, что порождает так называемую *паронимию* или "*нечаянную тавтологию*". Программа распространяется без каких-либо ограничений вместе с исходным текстом на C. Есть версия для OS/2.

1. АОТ (автоматическая обработка текста) (Алексей Сокирко)

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

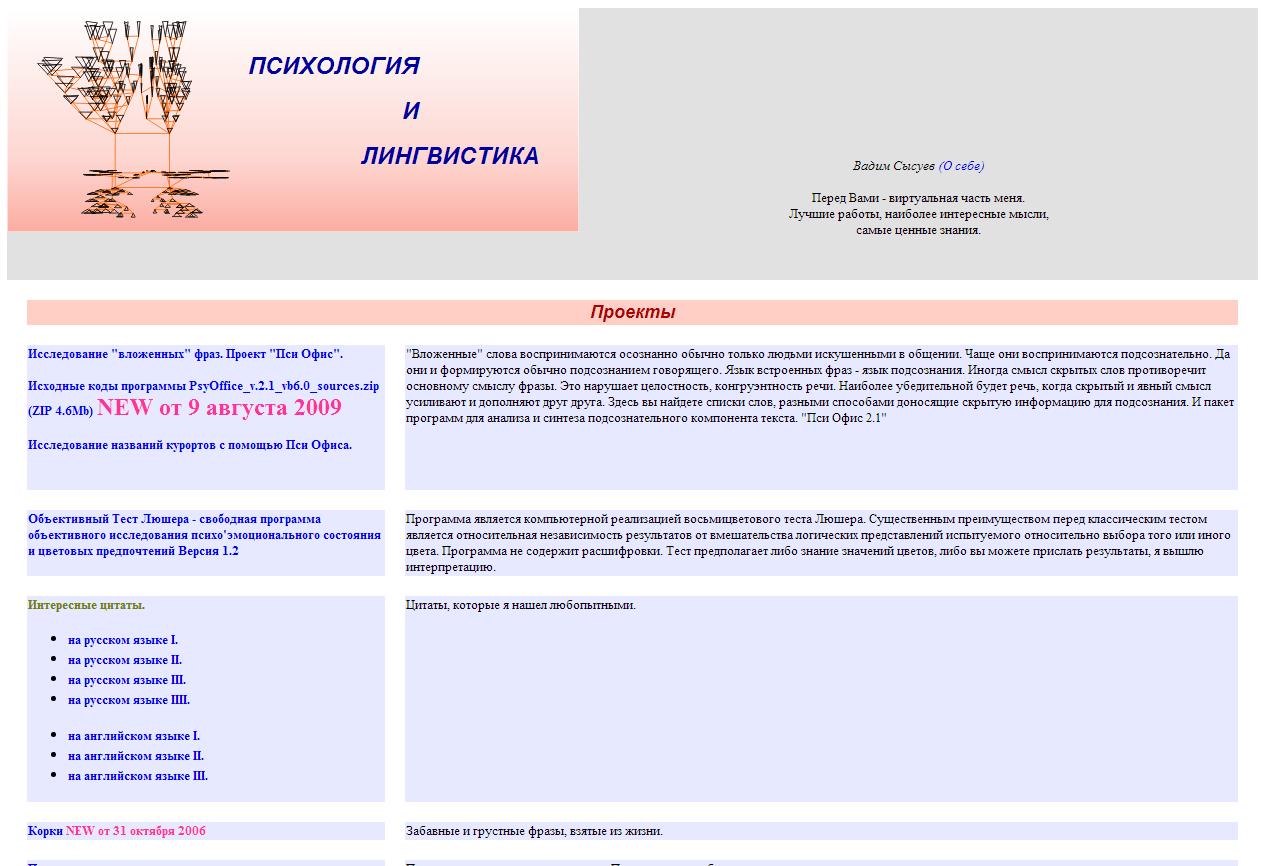


Сайт, на котором представлены разработки бывших сотрудников компании[Диалинг](https://rvb.ru/soft/catalogue/catalogue.html#_dialing), прекратившей свое существование в мае 2001г. Среди предлагаемых продуктов:  
- модуль графематического анализа текста;  
- компоненты морфологического анализа для русск. 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

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

<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.

1. ВААЛ-2000 (Белянин В. П., доктор филологических наук, МГУ; Шалак В. И., кандидат философских наук, Институт философии РАН)

<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. Успешно используется в ГосДуме, правительстве, крупных банках и PR-компаниях. Для бесплатной загрузки доступна старая DOS-версия системы образца 1995 года.

1. Проект «Продвинутая психология»

<http://psyberia.ru/>

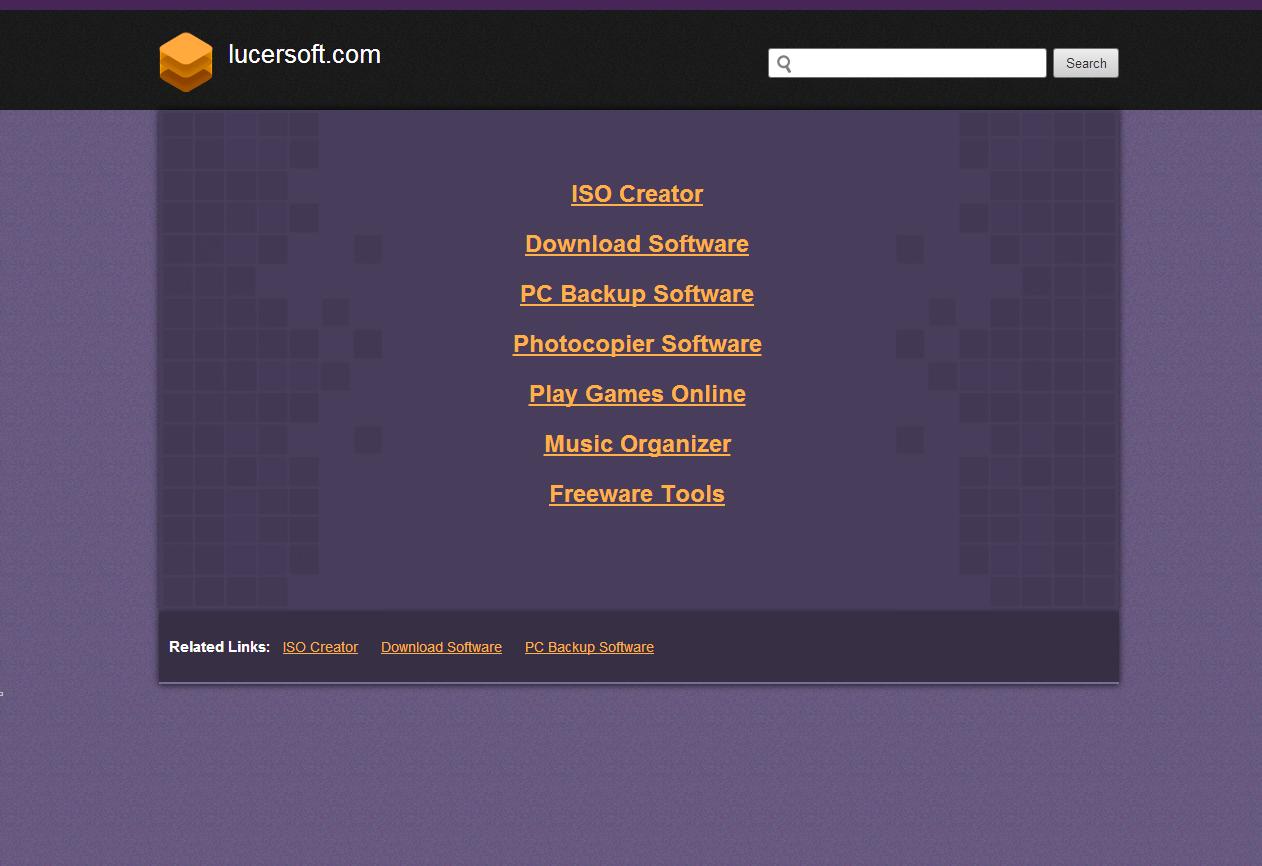


Каталог компьютерных программ на сайте проекта "Продвинутая Психология". Каталог включает следующие разделы: психологические тесты, психологические тренинги и игры, программы для наблюдения биоритмов, психолингвистические программы, психологические справочники и базы данных.

Программы преобразования текстов

1. [HTML Batch Editor](http://www.lucersoft.com/hbeinfo.htm) ([© 2001 Lucersoft](http://www.lucersoft.com/))

<http://ww38.lucersoft.com/hbeinfo.htm>



Компактная утилита для Windows, позволяющая осуществлять пакетные преобразования массивов текстовых или html-файлов с помощью набора заданных шаблонов. Доступна для загрузки в двух вариантах - облегченная freeware версия [HTML Batch Editor v1.1](http://www.lucersoft.com/files/hbe1.zip) и полная shareware ($25) версия [HTML Batch Editor v2.2](http://www.lucersoft.com/files/hbe22.zip).

1. [ClearText Reader's Edition](http://www.gribuser.ru/freeware/cleartxt/index.shtml) (Дмитрий Грибов)

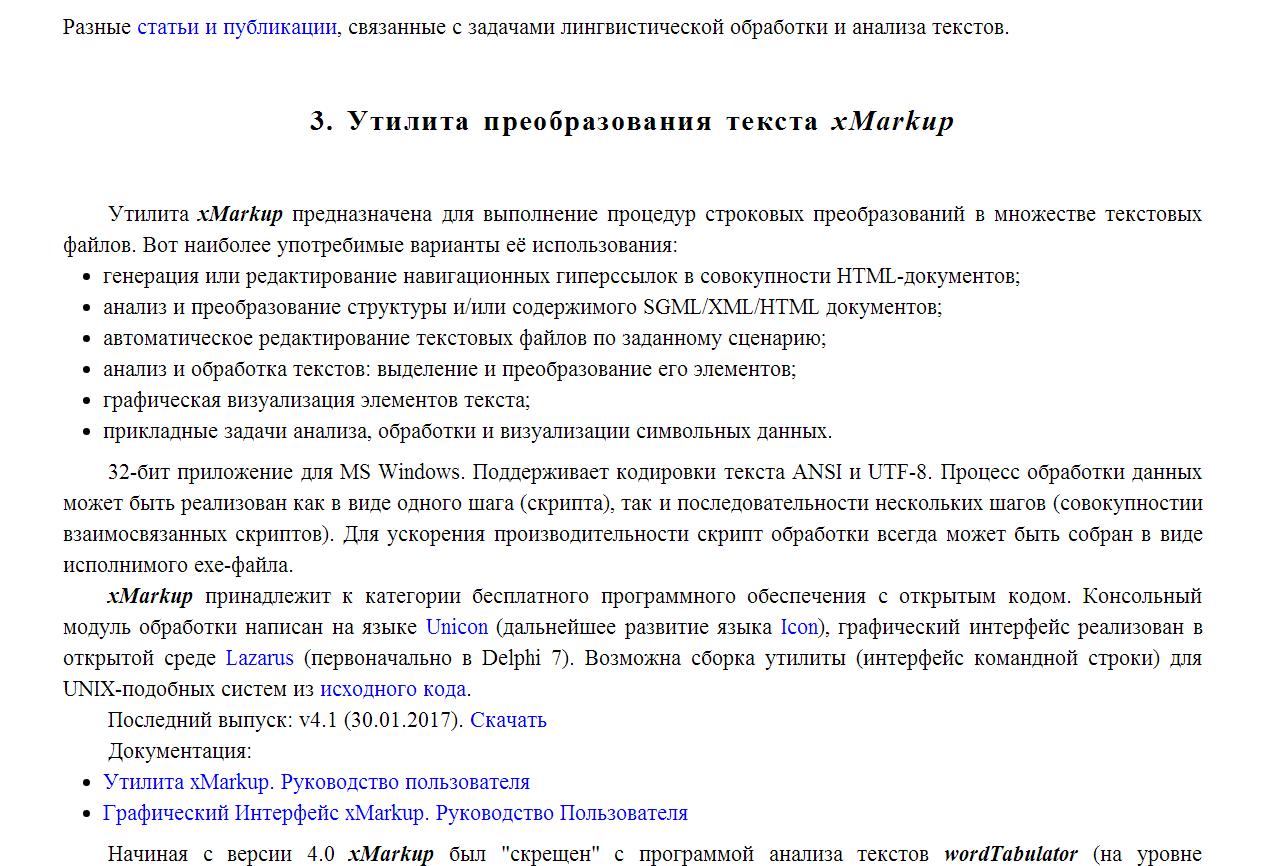
<http://www.gribuser.ru/freeware/cleartxt/index.shtml>



Freeware программа для переформатирования текстовых или html-файлов (Windows). Позволяет осуществлять как вычистку разметки, так и ее добавление. Понятный графический интерфейс, автораспознавание кодировки.

1. [xMarkup v1.6.8](https://rvb.ru/soft/index.html) (© [Логичев С.В.](mailto:slogichev@yandex.ru), 1999-2002)

<https://rvb.ru/soft/index.html>

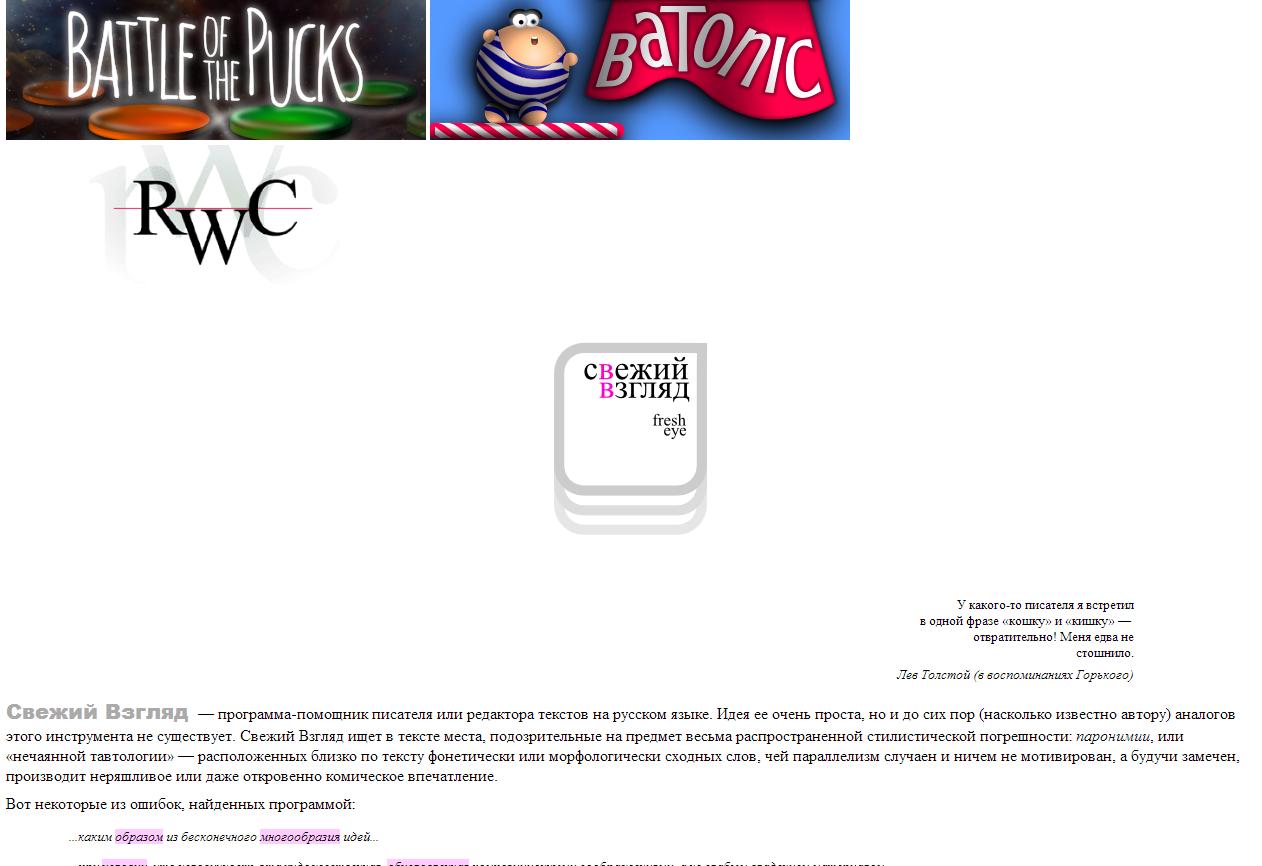


Приведено описание экспериментальной программы, предназначенной для выполнения сколь угодно сложных преобразований множества текстов. Реализована как консольная утилита для Windows 9x/NT/2000/XP. Позволяет изменять, добавлять или удалять заданные элементы обрабатываемых текстов. Встроенный язык обработки. Программу можно получить по e-mail после письма к автору.

Генераторы текстов и «говорящие программы»

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

<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.

2. [The Dada Engine](http://dev.null.org/dadaengine/) (Andrew C. Bulha)

<http://dev.null.org/dadaengine/>



Генератор случайных текстов на основе заданной грамматики (англ.яз.). Работает на UNIX-подобных системах, можно загрузить исходный код. На основе The Dada Engine реализована on-line версия [Генератора Постмодернистских Текстов](http://www.elsewhere.org/cgi-bin/postmodern/).

3. [Chatterbot Central](http://www.simonlaven.com/) (Simon Laven)

<http://www.simonlaven.com/>

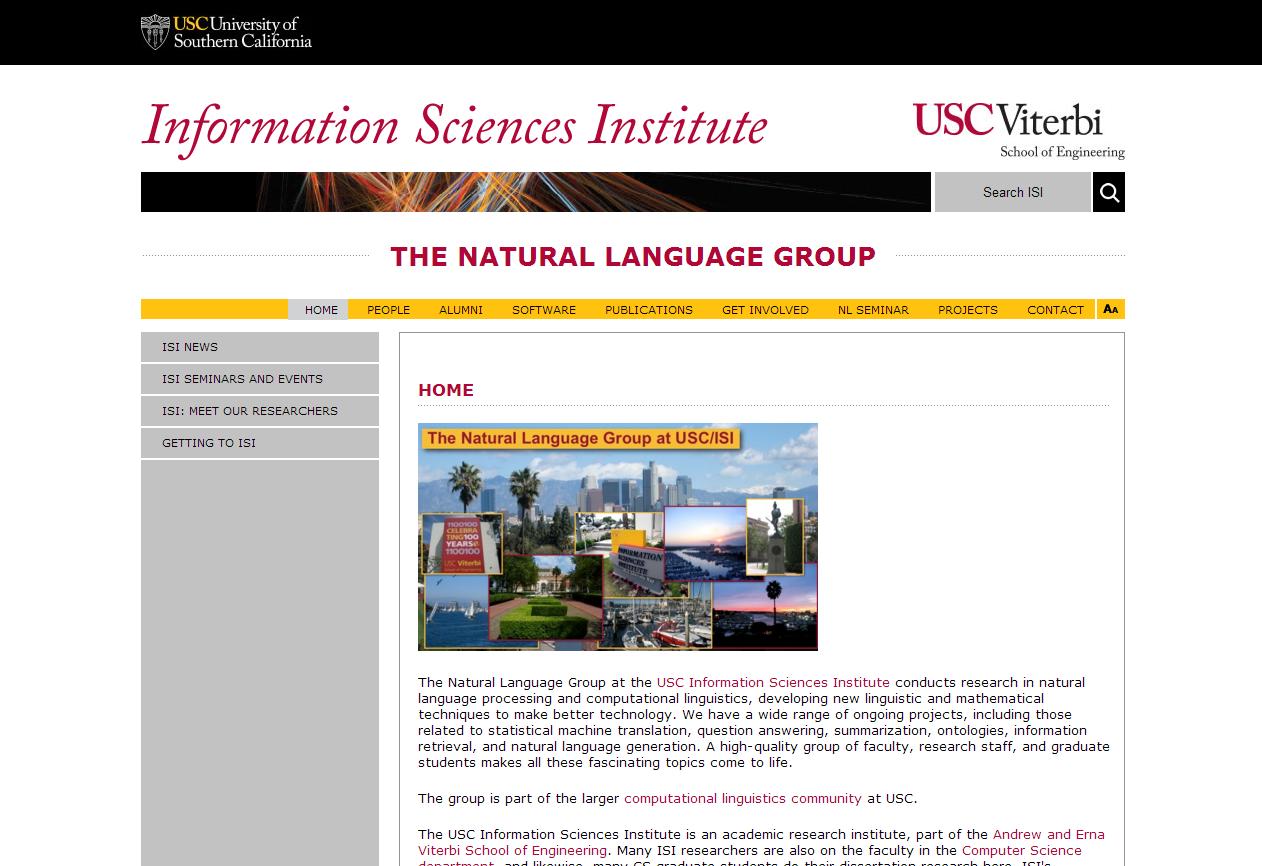


Зоопарк ботов всевозможных мастей и видов. Здесь, похоже, собрана вообще вся информация о говорящих программах, существующих сегодня в сети. Начиная c классики (например, всемирно известной Элизы доктора Фейгенбаума) и кончая самыми последними разработками с использованием языка искусственно-интеллектуальной разметки (AIML).

### Системы обработки естественного языка (NL-Processing)

1. [Natural Language Projects at ISI](http://www.isi.edu/natural-language/nlp-at-isi.html) © [Univ.of Southern California/Information Science Inst.](http://www.isi.edu/), 1999-2001

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



Описание группы проектов по вычислительной лингвистике и машинному переводу, выполняемых в Университете Южной Калифорнии.  
[Webclopedia](http://www.isi.edu/natural-language/projects/webclopedia) - проект, связанный с созданием информационных агентов, способных генерировать ответы пользователю на основе анализа разнородных коллекций ресурсов, доступных в сети.  
[ONTOSAURUS](http://mozart.isi.edu:8003/sensus/sensus_frame.html) - это онтологический тезаурус, включающий 90,000 концептуальных сущностей (словарная база взята из проекта [WordNet](http://www.cogsci.princeton.edu/~wn/)).  
[SUMMARIST](http://www.isi.edu/natural-language/projects/SUMMARIST.html) - мультиязычная система реферирования текстовой информации.  
[ReWrite](http://www.isi.edu/natural-language/projects/rewrite/) - система двуязычного перевода, основанная на статистическом анализе пар текстовых корпусов.

1. Проекты НИИ ИИ: Российский НИИ Искусственного Интеллекта

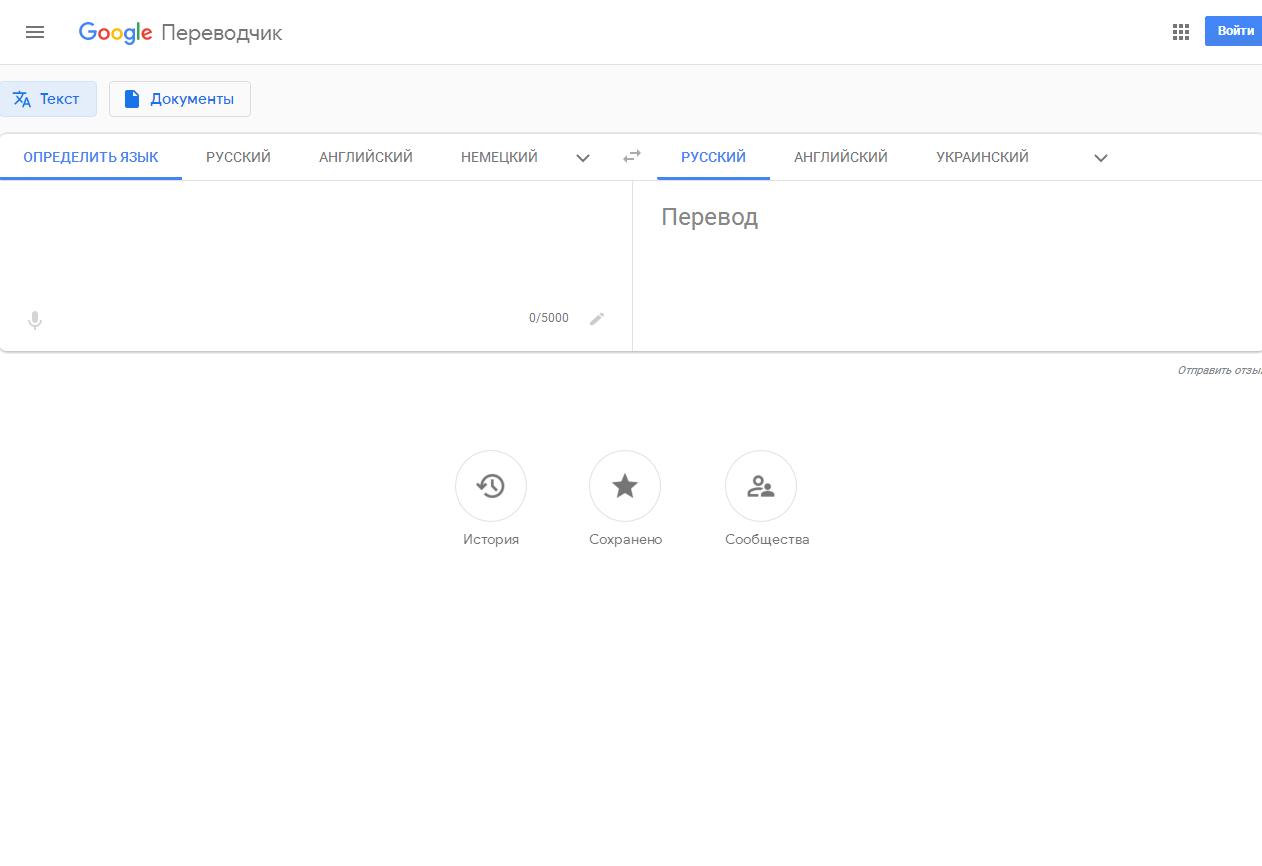
<http://www.artint.ru/>



Описан ряд проектов (Alex, AURA, InBASE, InDoc), реализующих технологии обработки естественного языка в задачах извлечения знаний, поиска, автоматического реферирования для массивов текстов.

1. Google Переводчик © 2008 Google

<https://translate.google.com/>

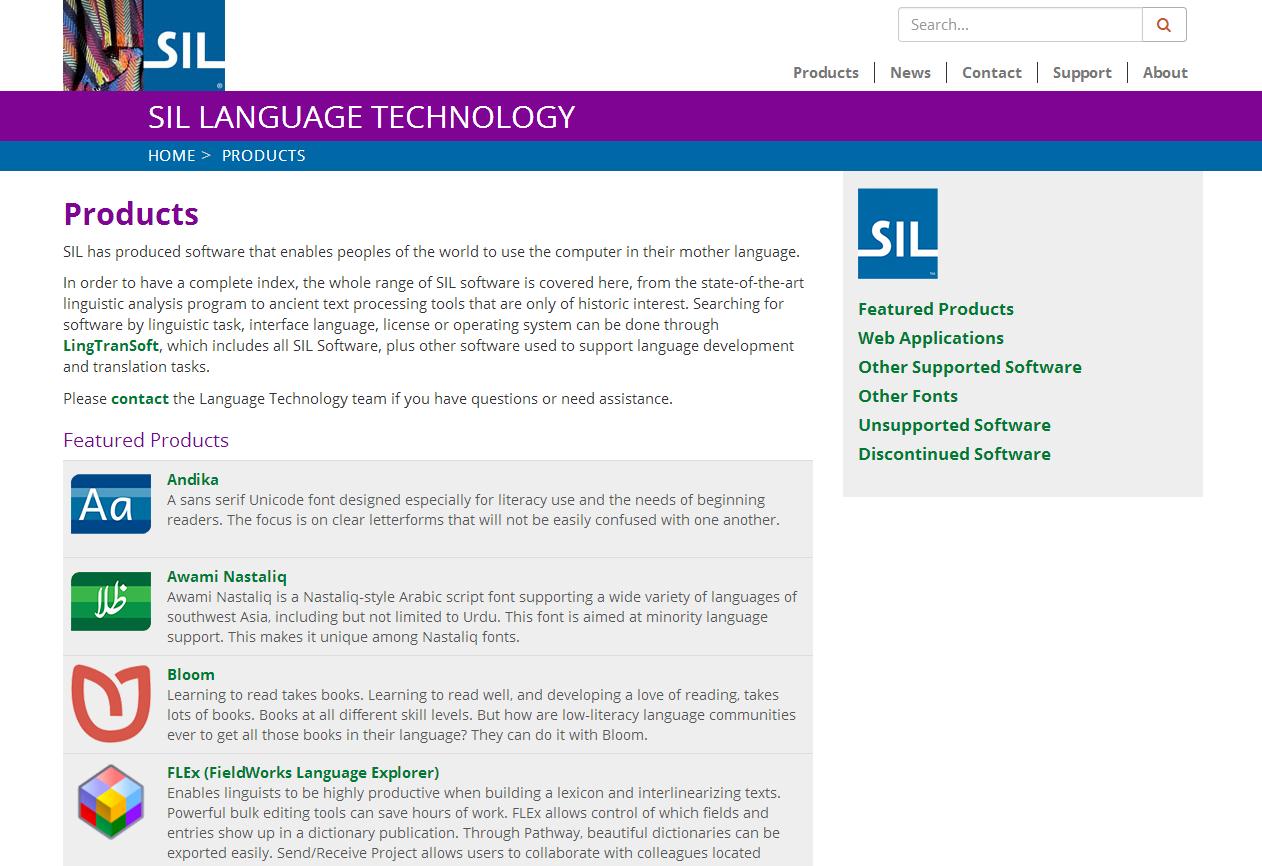


Сервис Google, предлагающий on-line перевод текста или Web-страничек. Поддерживает попарный перевод с 23 различных языков, включая все основные европейские языки и русский.

Коллекции ресурсов

1. [Каталог программ по вычислительной лингвистике](http://www.sil.org/computing/) - 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.

1. [Каталог ресурсов по анализу текстов](http://www.intext.de/textanae.htm) (Harald Klein)

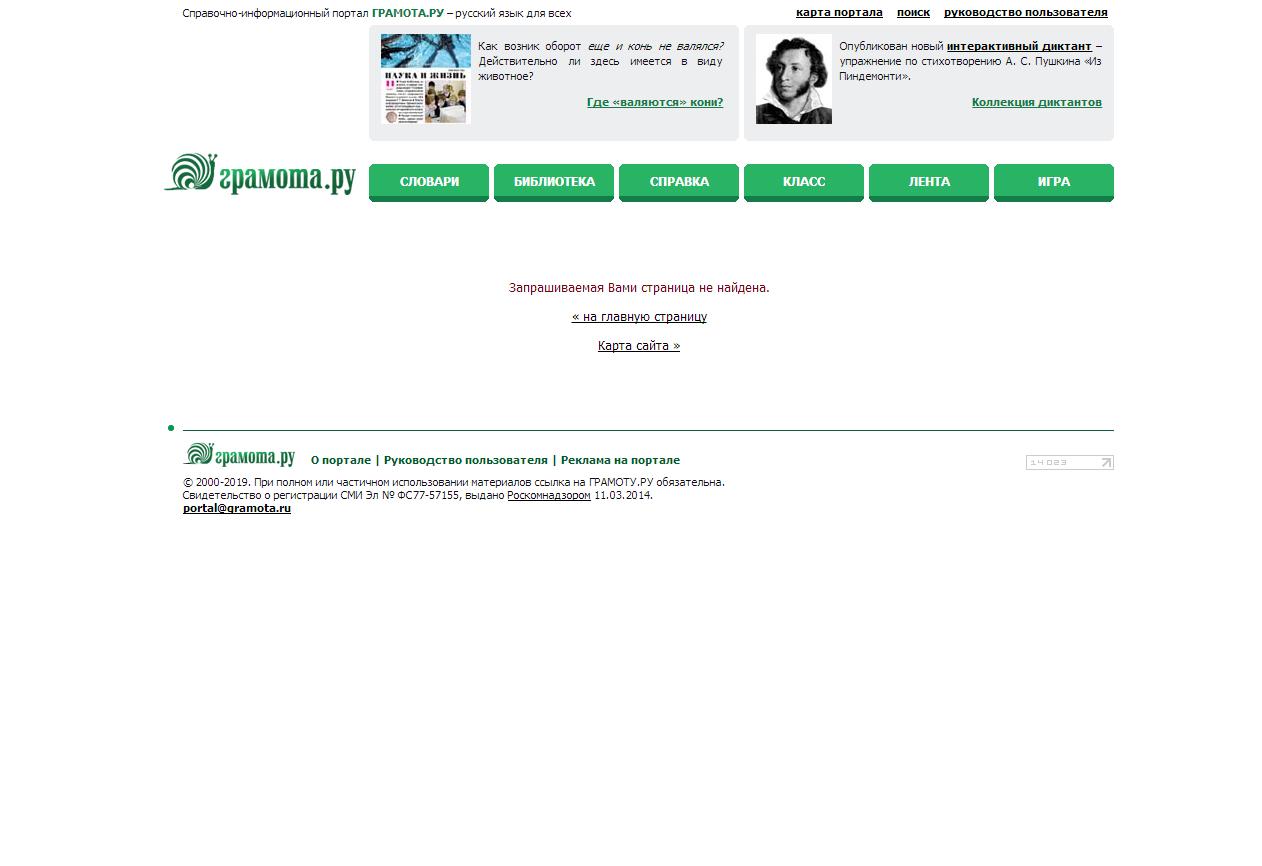
<http://www.intext.de/>



Обширная компиляция ссылок по проблеме анализа текстов (на англ.языке). Большой раздел посвящен тематическим компьютерным программам.

1. [Справочно-информационный портал "Русский язык"](http://www.gramota.ru/index.html) (© [МО "ЭЛЕКС"](mailto:info@gramota.ru), 2000)

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



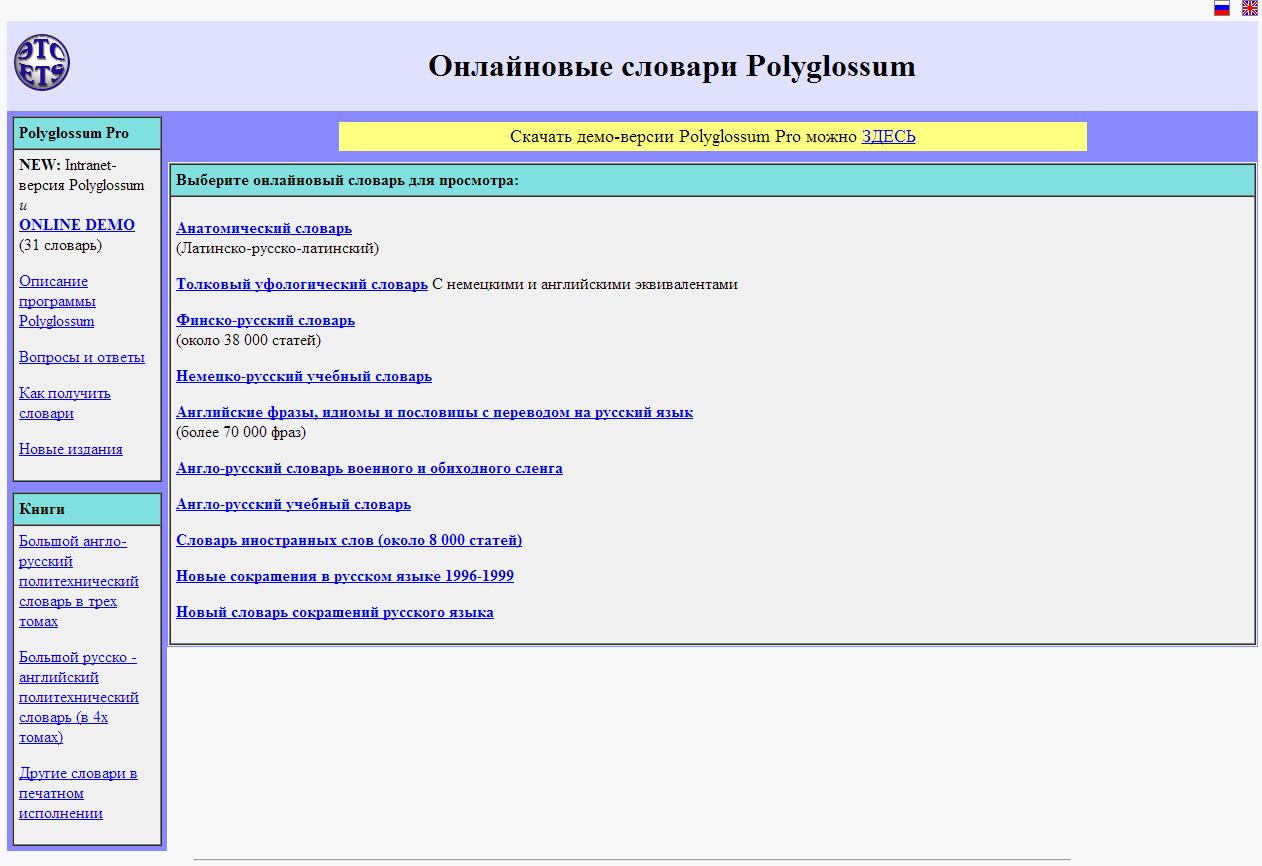
Замечательный ресурс для истинных любителей русской словесности, содержащий массу полезной иноформации. Также включает он-лайновые словари русского языка:  
 - Полный электронный орфографический словарь русского языка под ред. В.В. Лопатина.  
 - Словарь трудностей произношения и ударения в современном 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.

Словари и тезаурусы

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

© 1997-2000 [ETS Publishing House](http://www.ets.ru/udict-r.htm)

<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.

1. Словари Ожегова и Зализняка (С. А. Старостин)

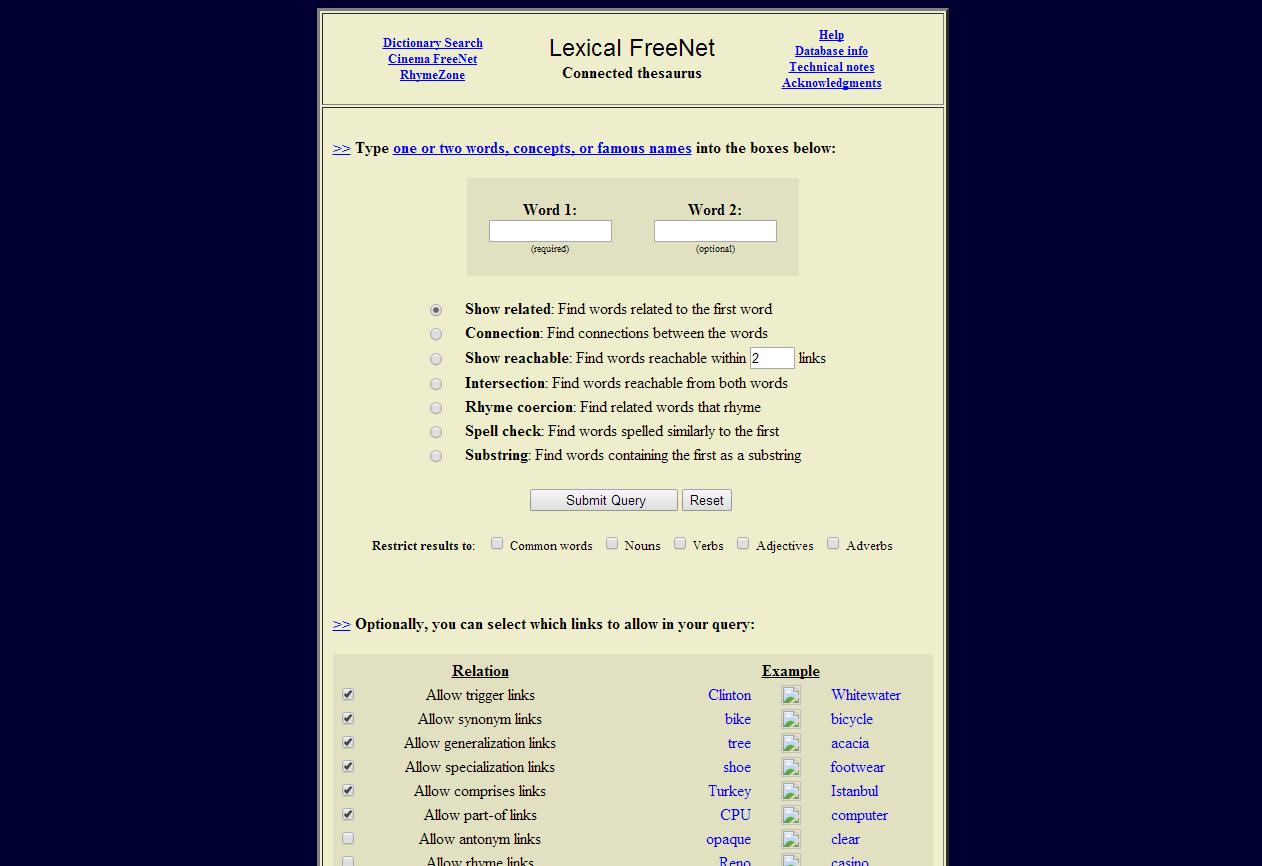
<http://starling.rinet.ru/cgi-bin/main.cgi?flags=wygnmnl>



On-line версии словарей Ожегова и Зализняка.

1. Lexical FreeNet (Datamuse Corporation)

<http://www.lexfn.com/>



Продвинутый on-line тезаурус английского языка. Может использоваться для любых мыслимых целей - решения кроссвордов, поиска синонимов, связанных слов, подбора рифм, орфографической проверки и т.д.