

Воронежский государственный университет

Проектная работа

на тему:

Понятие информационных технологий в лингвистике.

Автоматизированное рабочее место лингвиста.

Понятие естественного и искусственного языка. Виды
искусственных языков.

Подготовила студентка ВГУ

1 курс

Направление «Лингвистика»

Профиль «ТиМПИЯиК»

Онучина Анна Вениаминовна

Воронеж- 2018

Мотивация

Современное информационное пространство обязывает человека, изначально обладающего определенными знаниями, опытом и компетенциями, уметь обучаться: находить нужную информацию, уметь пользоваться различными источниками. Постоянное приобретение дополнительных знаний помогает решать профессиональные и личные задачи.

Ресурсы, в которых можно найти информацию о необходимости и эффективности применения информационных технологий в процессе лингвистической деятельности:

<https://qwizz.ru/информационные-технологии-лингвис/>

Теория.

1. Понятие информационных технологий в лингвистике.

В настоящее время лингвистическая наука находится в тесной связи с информационными технологиями. С помощью технологий можно успешно решать всевозможные задачи лингвистики. В то же время сама эта наука оказывает значительное воздействие на формирование современных инноваций.

Информационные технологии в лингвистике являются совокупностью законов, средств и методов, позволяющих получать, хранить, передавать, распространять и преобразовывать информацию о языке и принципах его функционирования при помощи современных компьютерных программ.

<https://qwizz.ru/информационные-технологии-лингвис/> Перейдя по этой ссылке, вы сможете более подробно ознакомиться с понятием информационных технологий в лингвистике.



2. Автоматизированное рабочее место лингвиста.

Общие сведения:

Автоматизированное рабочее место (АРМ) — это рабочее место специалиста, оснащенное персональным компьютером, программным обеспечением и совокупностью информационных ресурсов индивидуального или коллективного пользования, которые позволяют ему вести обработку данных с целью получения информации, обеспечивающей поддержку принимаемых им решений при выполнении профессиональных функций.

Что касается лингвистики, рабочее место специалиста данной сферы должно быть оснащено совокупностью аппаратных, программных и лингвистических средств, необходимых для автоматической обработки лингвистических данных.

Программы и ресурсы, необходимые лингвисту:

Электронные словари:

ABBYY Lingvo

Babylon-Pro

АльфаЛекс

Pop-up Dictionary

SKIIN

Albion Explanatory Dictionary

Rdic

EDictionary

VU-Dictionary

Parley Box

Context

MultiLex

Контекстные словари:

Artefact Dictionary

TranslateIt!

TransLite

MyDic

QDictionary

LearnWords

ABBY LINGVO

Контекстный словарь представляет собой словарь каждая запись которого содержит контекст употребления одного определенного значения слова. Контекстный словарь создается с целью дальнейшего, более эффективного изучения слов в смысловом окружении, контексте.

<http://www.contdict.ru> – это словарь, работающий в онлайн-режиме, который позволяет пользователю работать более чем с десятком языков, а так же имеет собственный каталог обучающих видео- <http://www.contdict.ru/videos>

Лингвистические банки данных:

Под лингвистическим банками данных понимаются представленные в электронной форме языковые источники (корпусы текстов) и лингвистические описания.

<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.

<https://corpus.byu.edu/bnc> - это корпус текстов из 100 миллионов слов, содержащий образцы письменного и разговорного британского английского языка из широкого круга источников. Корпус охватывает британский английский конца XX в., представленный широким

разнообразием жанров, и задуман как образец типичного разговорного и письменного британского английского языка того времени.

Lancaster/Oslo-Bergen Corpus - имеет в своем составе два подкорпуса - Leeds-Lancaster Treebank и Lancaster Parsed Corpus, - снабженных синтаксической разметкой.

Также существует отдельный сайт, который содержит каталог лингвистических программ и ресурсов в сети -

<https://rvb.ru/soft/catalogue/catalogue.html#Resources>

3.Понятие естественного и искусственного языка. Виды искусственных языков.

«Естественный» и «искусственный» - это деление языков по происхождению.

Естественный язык - в лингвистике и философии языка язык, используемый для общения людей и не созданный искусственно (в отличие от искусственных языков)

Искусственные языки - специальные языки, которые, в отличие от естественных, сконструированы целенаправленно. Они могут конструироваться с помощью естественного языка или ранее построенного искусственного языка.

Различают следующие виды искусственных языков:

- Языки программирования и компьютерные языки - языки для автоматической обработки информации с помощью ЭВМ.
- Информационные языки - языки, используемые в различных системах обработки информации.
- Формализованные языки науки - языки, предназначенные для символической записи научных фактов и теорий математики, логики, химии и других наук.
- Языки несуществующих народов, созданные в беллетристических или развлекательных целях, например: эльфийский язык, придуманный Дж. Толкином клингонский язык, придуманный Марком Окрандом для фантастического сериала «*Star Trek*», язык На'ви, созданный для фильма «Аватар».

- Международные вспомогательные языки - языки, создаваемые из элементов естественных языков и предлагаемые в качестве вспомогательного средства межнационального общения.

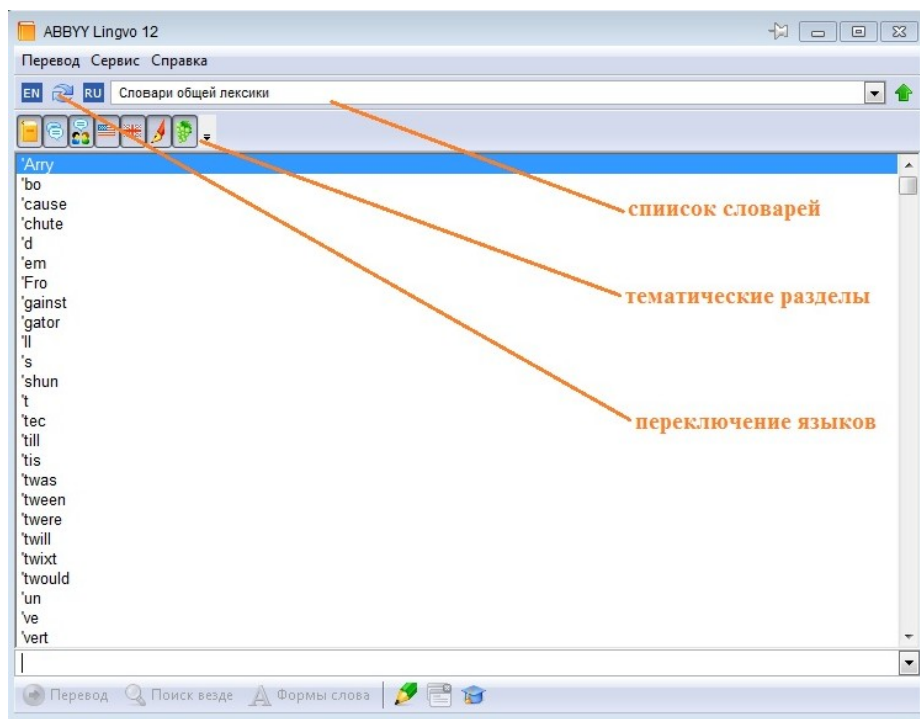
Практика.

Ознакомление с ABBYY Lingvo 12 Multilingual Edition

Это электронный словарь, который содержит современную лексику из разных тематических областей, как универсальную, так и специальную.



Качаем этот словарь по ссылке – <https://soft.sibnet.ru/soft/12491-abbyy-lingvo-12-ml/download> и устанавливаем.



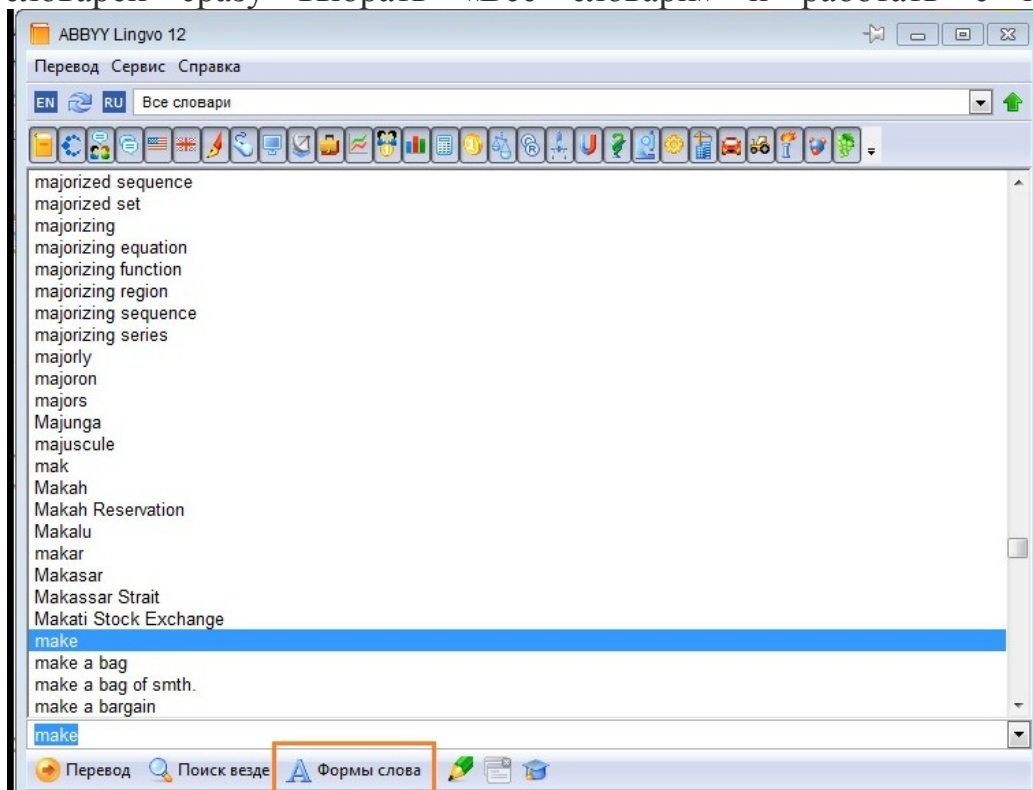
В списке словарей по умолчанию доступно 5 групп:

- словарь общей лексики;
- словари по естествознанию;
- словари по информатике;
- словари по экономике и праву;
- технические словари.

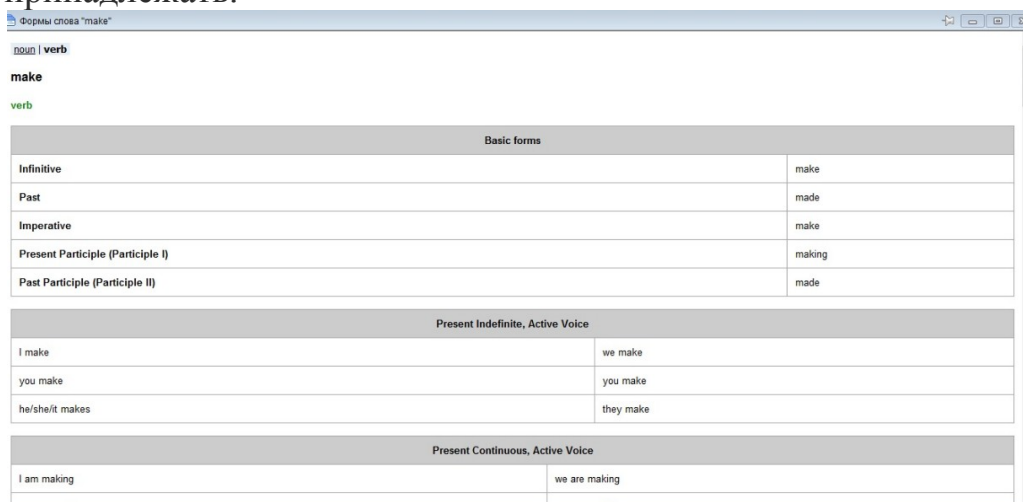
Каждая группа состоит из определенных тематических разделов. Например, выбрав словарь общей лексики, видим, что он состоит из следующих тематических ниш: универсальный (LingvoUniversal), идиомы (Idioms), неформальное общение (Informal), американский английский (American), британский английский (GreatBritain), виноделие (Wine).

Если кнопка тематического раздела нажата то, при переводе, система будет осуществлять поиск слова в данной теме. Для удобства можно в списке

словарей сразу выбрать «Все словари» и работать с ними всегда.



Перед нами откроется окно, в котором можно посмотреть всевозможные формы например слова «make», для каждой части речи, к которой оно может принадлежать.



Следующей крайне полезной функцией программы является перевод слов при помощи их выделения и нажатия комбинации клавиш Ctrl+C+C. Допустим, вы читаете одну из моих статей на английском языке. Попробуйте выделить незнакомое слово и, зажав Ctrl, два раза нажать клавишу C. Вот, что из этого

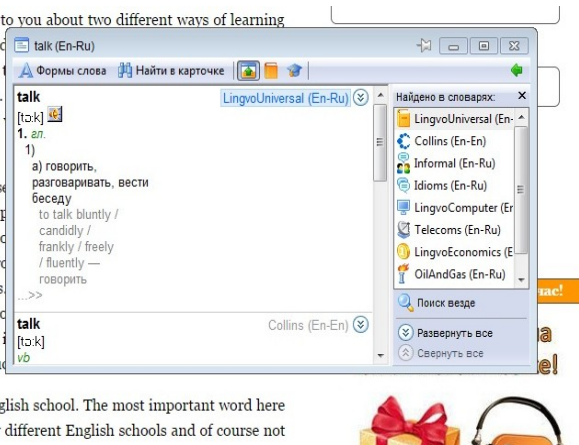
получится:

Hello my friends! Today I'd like to talk to you about two different ways of learning

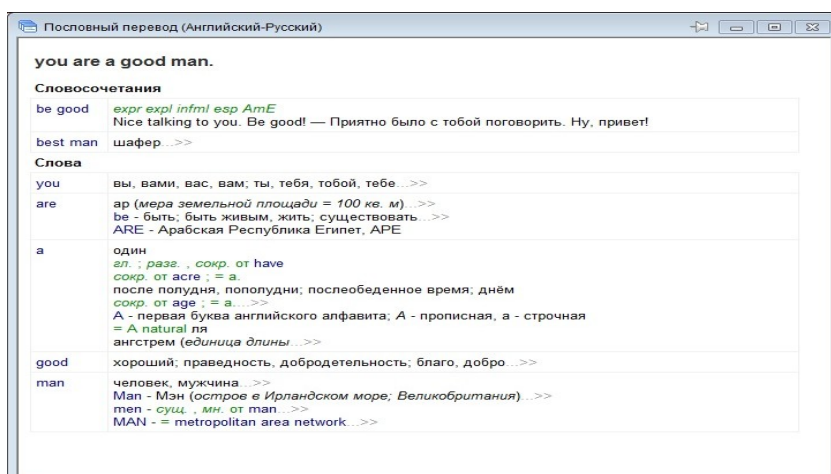
English — common classes and individual lessons. It's an important question for those who want to learn English. Years ago I was thinking about it as well. Now I know two ways of learning from my point of view. You should make the right decision.

What do we get visiting common classes? We learn how to learn. Learning to communicate with other people is the main goal of group lessons. As you well know, one of the problems is getting rid of all the psychological problems. The main problem is common for 90% of learners. It's a problem during a long period. You should start trying to speak with other people. A good learning English school can give such

I want to stress these words — good English school. The most important word here is «good». Nowadays there are so many different English schools and of course not




Однако данный словарь не предназначен для перевода случайных фраз и предложений. Программа выдаст перевод фразы лишь в том случае, если данное выражение является устойчивым, идиомой или фразовым глаголом. Если же вам требуется перевести какое-то случайное предложение, то вам будет предложен перевод каждого слова в отдельности. Например, давайте переведем фразу «You are a good man». Вводим выражение в строку перевода и нажимаем Enter.







Ознакомление с Translate.ru

По этой ссылке мы переходим на сайт словаря <https://www.translate.ru>

Определить язык  Русский Введите слово для перевода... 

Словарь PROMT Онлайн - это самый быстрый и удобный перевод слов и фраз, который включает в себя:




-  **Граматику**
Род у существительных, склонение существительных и прилагательных, степени сравнения прилагательных, спряжение глаголов;
-  **Транскрипцию и произношение слов и их переводов**
Можно прослушать, как произносится искомое слово, а также все его переводы, для английского языка доступна транскрипция;
-  **Примеры переводов**
Для каждого запроса осуществляется поиск устойчивых словосочетаний с искомым словом в словарной базе и поиск примеров употребления в реальных текстах;
-  **Ссылки на другие ресурсы и словарные сервисы,**
такие как Википедия, Dictionary.com и Merriam-Webster и другие.






На данном изображении перечислены основные достоинства и возможности данного словаря.

В поисковую строку мы вводим слово, которое требует перевода, предварительно выбрав исходный язык(однако программа способна сама определить язык по заданному слову)

















PROMT Онлайн 👤 🇷🇺 ☰

ПЕРЕВОДЧИК СЛОВАРЬ ПРИМЕРЫ ГРАММАТИКА


Английский  Русский make  

make [meɪk] , *Глагол*
make / made / made / making / makes

-  **делать** (choice, arrangements)
-  **заставлять**
-  **сделать**
-  **вынуждать** (compel)
-  **зарабатывать**
-  **достигать** (achieve)
-  **составлять** (be, constitute, equal)
-  **заключать** (contract)
-  **ставить** (record)
-  **снимать** (film)
-  **производить** (tools)
-  **произносить** (speech)
-  **осуществлять** (payment)
-  **совершать** (error)
-  **заслуживать** (earn)
-  **выполнять** (Путешествия) (hybrid translation)

Словосочетания (593)

 Поиск...

- [black face make up](#) - грим "блэкфейс"
- [blackface make up](#) - грим "блэкфейс"
- [borat: cultural learnings of america for make benefit glorious nation of kazakhstan](#) - Борат: культурные исследования Америки в пользу славного государства Казахстан
- [eye make up](#) - макияж глаз
- [eye make up remover](#) - средство для снятия макияжа с глаз
- [fine feathers make fine birds](#) - одежда красит человека
- [full make up](#) - полный макияж
- [make a bad break](#) - проговариваться
- [make a balk](#) - совершать промах
- [make a balk of good ground](#) - упускать удобный случай
- [make a baulk](#) - совершать промах
- [make a baulk of good ground](#) - упускать удобный случай
- [make a beeline](#) - мчаться
- [make a big difference](#) - иметь большое значение
- [make a blot](#) - поставить кляксу
- [make a bolt for it](#) - помчаться
- [make a bonfire of](#) - сжигать
- [make a botch](#) - напортачить
- [make a break with](#) - порвать
- [make a buck](#) - зарабатывать

Перевод с английского на русский выполнен PROMT®

Словарь показывает нам всевозможные переводы данного слова, а так же словосочетания с ЭТИМ словом.

Затем в меню сайта мы выбираем графу «Примеры» и программы выдает нам всевозможные примеры использования данного слово в предложениях и текстах с различными вариантами контекстного перевода, а так же предлагает выбрать тематику употребления данного слова.

ПЕРЕВОДЧИК	СЛОВАРЬ	ПРИМЕРЫ	ГРАММАТИКА
Английский	Русский	make	
<input type="text" value="Поиск..."/>			
<p>Переводы: все⁸³³²⁵ Другие переводы⁴⁷¹⁴⁰ сделать³³⁶⁰ делает¹⁷¹⁷ сделал¹⁰⁷³ делать¹⁰⁶⁸ убедитесь¹⁰⁴⁰ осуществления⁸⁶⁹ сделали⁷²⁹ показать все</p> <p>Тематика: все⁸³³²⁵ Политика и экономика⁴⁵⁴⁶⁴ Разговорная речь¹⁴⁹¹³ Общая лексика¹³¹⁴⁸ Software³⁷¹¹ Интернет²⁹¹⁹ Финансы²²³⁴ показать все</p>			
I can hardly make a speech without feeling nervous.		Я едва ли могу произнести речь без того, чтобы нервничать.	
Can you make a photocopy, please?		Вы можете сделать ксерокопию?	
Make a stupid man to pray - he'll hurt his forehead.		Заставь дурака Богу молиться - он себе и лоб расшибёт.	
Part of the problem, of course, is that we make resolutions to do only things that we are not otherwise likely to do.		Часть проблемы заключается в том, что мы принимаем решения делать только те вещи, которые иначе мы, возможно, не сделали бы.	
A hearing will be scheduled within a month for the judge to make a more permanent ruling on the law.		Слушание будет назначено в течение месяца, на котором судья вынесет более постоянное решение по этому закону.	

Также сайт дает возможность ознакомиться с грамматикой данного слова. Для этого нам нужно перейти в раздел «Грамматика» в меню сайта.

ПЕРЕВОДЧИК	СЛОВАРЬ	ПРИМЕРЫ	ГРАММАТИКА
Английский	make		
<input type="text" value="make, Глагол"/>			
<p> make [meɪk] - <i>Глагол</i></p> <p>ВСЕ ФОРМЫ INDEFINITE CONTINUOUS PERFECT PERFECT CONTINUOUS INFINITIVES PARTICIPLES</p>			
<p><i>Present Indefinite</i></p> <p>I make you make he/she/it makes we make you make they make</p>		<p><i>Present Perfect</i></p> <p>I have made you have made he/she/it has made we have made you have made they have made</p>	
<p><i>Past Indefinite</i></p> <p>I made you made he/she/it made we made you made they made</p>		<p><i>Past Continuous</i></p> <p>I was making you were making he/she/it was making we were making you were making they were making</p>	
		<p><i>Present Continuous</i></p> <p>I am making you are making he/she/it is making we are making you are making they are making</p>	
		<p><i>Present Perfect Continuous</i></p> <p>I have been making you have been making he/she/it has been making we have been making you have been making they have been making</p>	
		<p><i>Past Perfect</i></p> <p>I had made you had made he/she/it had made we had made you had made they had made</p>	
		<p><i>Past Perfect Continuous</i></p> <p>I had been making you had been making he/she/it had been making we had been making you had been making they had been making</p>	

Ознакомление с Корпусами

<https://corpus.byu.edu/bnc> - это корпус текстов из 100 миллионов слов, содержащий образцы письменного и разговорного британского английского языка из широкого круга источников. Корпус охватывает британский английский конца XX в., представленный широким разнообразием жанров, и задуман как образец типичного разговорного и письменного британского английского языка того времени.

На главной странице мы видим строку поиска, в которую необходимо вставить интересующее слово. В результате откроется страница с примерами предложений, в которых оно встречается.

Например, проверим, а используется ли глагол «recommend» с инфинитивом, т.е., а можно ли сказать «recommend to do».

Вставляем в строку поиска «recommend to»

Корпус показывает нам, что нельзя.

Вот еще одна база данных <http://corpus.leeds.ac.uk/protected/query.html> , которая состоит из разных корпусов, в том числе и BNC.

Здесь поиск того же «recommend to» дал другие результаты, и выяснилось, что хоть и редко, но «recommend to do» все-таки встречается.

Ознакомление с синтезаторами речи

Программы-синтезаторы речи с каждым годом всё больше входят в нашу жизнь. Они позволяют нам более досконально учить иностранные языки, переводят тексты в удобный аудиоформат, используются в функционале различных служебных программ и многое другое. И когда у некоторых из нас возникает потребность воспроизвести онлайн какой-либо текст в аудиоформате, тогда многие из нас обращаются к различным сервисам и программам по синтезу речи, способным помочь нам в трансформации нужного там текста. Данные программы могут во многом облегчить труд лингвиста.

Чтобы перевести текст в речь онлайн можно воспользоваться сервисом fromtexttospeech <http://www.fromtexttospeech.com> . Он работает по принципу конвертации текста в аудиофайл формата mp3, который затем можно скачать себе на компьютер. Сервис поддерживает конвертацию текста величиной в 50 тыс. символов, что является достаточно значительным объёмом.



Convert text to speech

Simply copy and paste your text in the box below :

Create Audio File

Convert text to speech

Listen to the MP3 file :



Для работы с сервисом fromtexttospeech перейдите на него, в опции «Select Language» выберите «Russian» .

В большом окне введите (вставьте) нужный для озвучки текст, затем нажмите на кнопку «Create Audio File». Текст будет обработан, затем вы сможете послушать полученный результат, а потом и скачать его себе на ПК.

Для этого нажмите правой клавишей мыши на «Download audio file» и выберите в появившемся меню «Сохранить объект как».

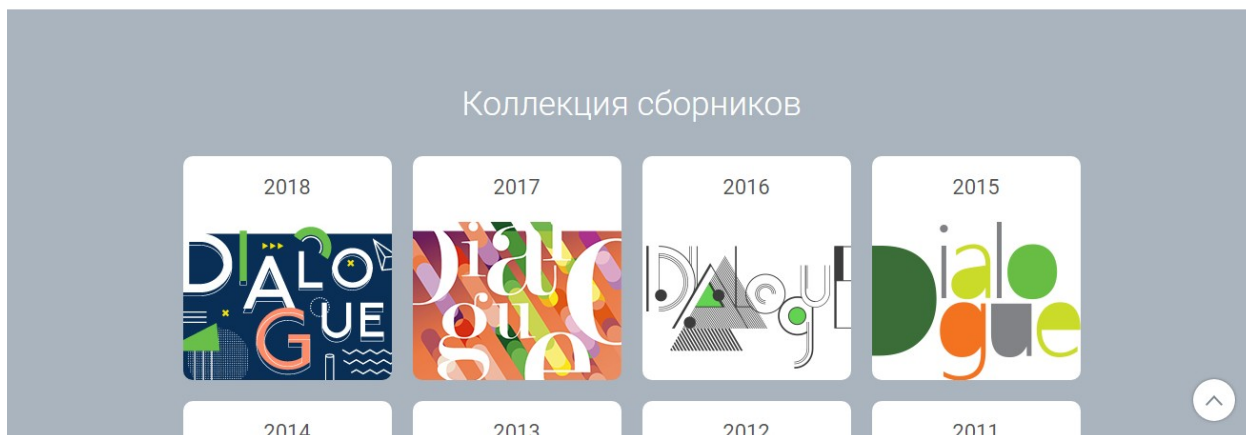
Ознакомление с лингвистическими ресурсами

Изучим возможности лингвистических ресурсов на примере сайта <http://www.dialog-21.ru> – на данном сайте представлены тексты всех докладов международной конференции «Диалог», посвященной проблемам и вопросам лингвистики и языкознания.

Данный сайт содержит коллекцию сборников с докладами, начиная с 2000г.

основным направлениям NLP (информационный поиск, машинный перевод, компьютерная лексикография, анализ и синтез речи, автоматическое извлечение лингвистических данных и др.), но и работы в области лексической семантики, синтаксиса, теории коммуникации.

- статьи, посвященные лингвистическому анализу русского языка, предполагающие знание этого языка у читателя, подаются на русском языке (с обязательной английской аннотацией).



Выбираем любой, например, сборник за 2012 год.

DIALOGUE 2019 Предыдущие конференции 2018 2017 2016 2015 2014

Сборник Dialogue Evaluation О конференции

Сборник 2012 [ВЫБРАТЬ СБОРНИК](#)

СОДЕРЖАНИЕ	ФОРМАТ PDF	ДОПОЛНИТЕЛЬНО
Титульный лист ✓ Статьи Аннотации	Полная версия сборника. Том 1 Полная версия сборника. Том 2	Статьи, публикуемые на сайте

(A) Anisimovich K. V., Druzhkin K. Ju., Minlos F. R., Petrova M. A., Selegey V. P., Zuev K. A.
Syntactic and semantic parser based on ABBYY Compreno linguistic technologies (v. 2: 91)
Antonova A. A., Misyurev A. V.
Russian dependency parser SyntAutom at the DIALOGUE-2012 parser evaluation task (v. 2: 104)
Апресян В. Ю.
"Тут" как показатель темпоральной близости (v. 1: 1)
Arkhipov A. V., Zakharov L. M., Krivnova O. F., Kodzasov S. V., Lebedev A. A.
Russian intonation corpus: a preliminary report (v. 1: 18)

(B) Проект частотного словаря русских идиом (v. 1: 28)
Беликов В. И., Селегей В. П., Шаров С. А.
Прологомены к проекту Генерального интернет-корпуса русского языка (v. 1: 37)
Бочаров В. В., Алексеева С. В., Грановский Д. В., Остапук Н. А., Степанова М. Е., Суриков А. В.
Сегментация текста в проекте «Открытый корпус» (v. 1: 51)
Описание гэппинга в системе автоматического перевода (v. 1: 61)
Богданова Н. В.
О проекте словаря дискурсивных единиц русской речи (на корпусном материале) (v. 1: 71)
Большаков И. А., Большакова Е. И.
Автоматический морфоклассификатор русских именных групп (v. 1: 81)
Борисова Е. Г.
Роль дискурсивных слов в управлении пониманием текста (v. 1: 93)

(C) Chetviorkin I. I., Braslavski P. I., Loukachevitch N. V.
Sentiment analysis track at ROMIP 2011 (v. 2: 1)
Chetviorkin I. I.
Testing the sentiment classification approach in various domains - ROMIP 2011 (v. 2: 15)

Выбрав раздел статьи, мы можем ознакомиться с текстом любого доклада, представленного на конференции за выбранный год.

Задания для закрепления изученного материала:

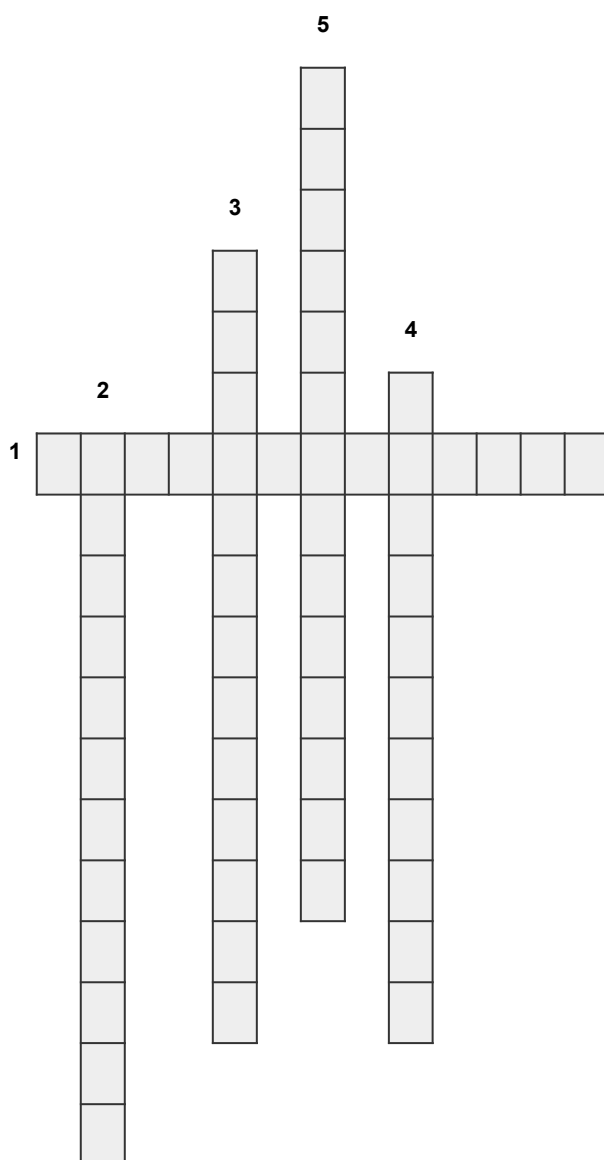
1. Карточки Quizlet

<https://quizlet.com/354914343/Закрепление-материала-по-изученной-теме-flash-cards>

2. Подбор Quizlet

<https://quizlet.com/354914343/match>

3. Кроссворд



1. Как называется язык, создаваемый из элементов естественных языков и предлагаемый в качестве вспомогательного средства межнационального общения?
2. Как называется язык, используемый для общения людей и не созданный искусственно?
3. Как называется специальный язык, который сконструирован целенаправленно?
4. Как называется язык, используемый в различных системах обработки информации?
5. Как называется словарь каждая запись которого содержит контекст употребления одного определенного значения слова?

4. Пройдите тест

<https://docs.google.com/forms/d/1UDTjB5pZrVIuatSpk4MQYFq7MeSnsyama2BiMAtpPrs/edit?usp=sharing>

5. Ответьте на вопросы в развернутом виде

https://docs.google.com/forms/d/1V7HpCLyH34U1mSE-KtWjDyi_dU-5Nq3Jzermefvc6lY/edit?usp=sharing

Ресурсы, которыми я пользовалась для создания урока: [html](http://biouroki.ru/workshop/crossgen.html)

<http://biouroki.ru/workshop/crossgen.html>

<https://quizlet.com>

<http://www.google.md/intl/ru/forms/about>

Ответы на кроссворд

