КОРПУСНАЯ ЛИНГВИСТИКА

ОБЪЕКТ, ПРЕДМЕТ, ЦЕЛЬ, ЗАДАЧИ, ИСТОРИЯ СОЗДАНИЯ, ПРИМЕРЫ

ПОНЯТИЕ КОРПУСНОЙ ЛИНГВИСТИКИ

Корпусная лингвистика - раздел языкознания, занимающийся разработкой, созданием и использованием текстовых корпусов с применением компьютерных технологий. Термин впервые введен в употребление в 60-е годы XX века.

Корпусная лингвистика сформировалась как отдельный раздел науки о языке в первой половине 90-х годов XX века.

ЛИНГВИСТИЧЕСКИЙ КОРПУС ЭТО СОВОКУПНОСТЬ ТЕКСТОВ:

собранных в соответствии с определенными принципами

представленных в электронном виде

унифицированных

размеченных по определенному стандарту

обеспеченных специализированной поисковой системой.

ЦЕЛЕСООБРАЗНОСТЬ СОЗДАНИЯ КОРПУСОВ

представление данных в реальном контексте достаточно большая представитель ность данных

возможность многократного использования единожды созданного корпуса для решения различных задач

ОБЪЕКТ И ПРЕДМЕТ КОРПУСНОЙ ЛИНГВИСТИКИ

Объект корпусной лингвистики – корпус текстов, который, с одной стороны, представляет собой исходный речевой материал для корпусной лингвистики и для других лингвистических дисциплин, с другой стороны, является результатом деятельности корпусной лингвистики.

Предмет корпусной лингвистики - теоретические основы и практические механизмы создания и использования представительных массивов языковых данных, предназначенных для лингвистических исследований в интересах широкого круга пользователей.

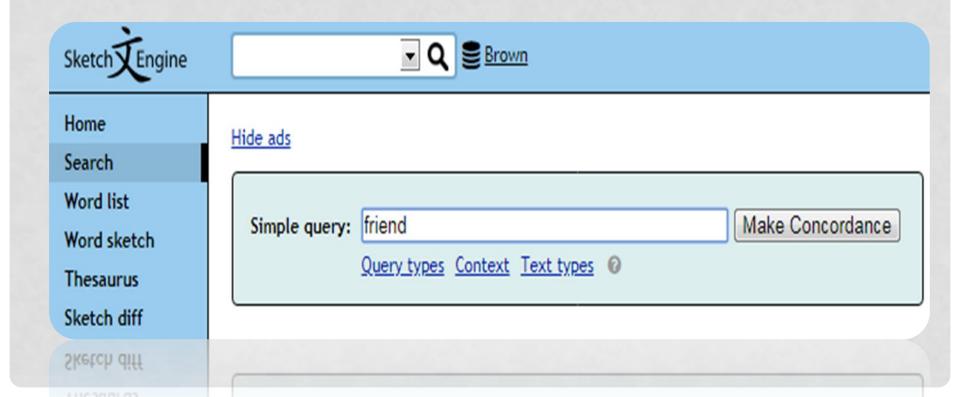
ЦЕЛЬ КОРПУСНОЙ ЛИНГВИСТИКИ - ИССЛЕДОВАНИЕ ЕСТЕСТВЕННОГО ИСПОЛЬЗОВАНИЯ ЯЗЫКА

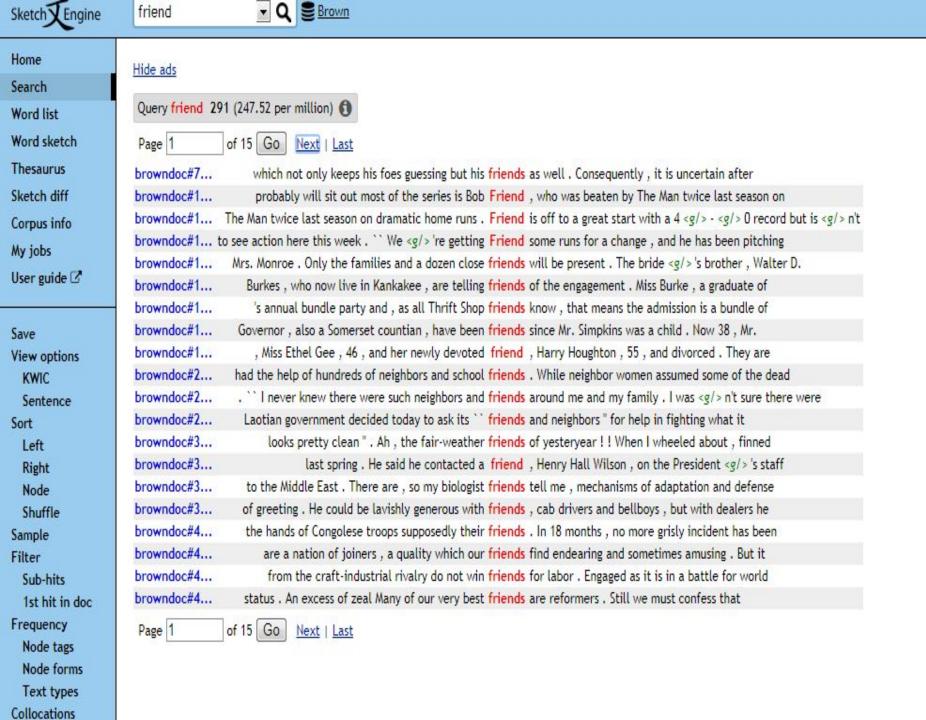
Задачи:

- Сбор текстов с определенной целью
- Их машинная обработка
- Помощь в создании словарей (лексикографическая поддержка)
- Составление конкордансов (список встречающихся в тексте словоформ со словесным окружением, расположенных в алфавитном порядке)
- Составление частотных словарей
- Создание национальных корпусов
- Исследование использования естественного языка в разных регистрах
- Помощь в создании лингводидактических пособий

ИСТОРИЯ СОЗДАНИЯ ЛИНГВИСТИЧЕСКИХ КОРПУСОВ

Brown Corpus - создан в 1960-е годы в
 Университете Брауна, содержал 500
 фрагментов текстов по 2 тысячи слов в каждом.





Search Word list

Word sketch

Thesaurus

Sketch diff

Corpus info

My jobs

User guide 2

Save

Change options

Cluster

Sort by freq

Hide gramrels More data

Less data

Sketch grammar

Translate

- Arabic
 - Bulgarian
 - Czech
 - Dutch
 - French
 - German
 - Italian
 - Polish
 - Russian

 - Spanish

Menu position

friend (noun)
Brown freq = 285 (242.41 per million)

object of		
		30.18
meet	2	6.96
ask	2	6.79
make	7	6.19
become	2	5.85
have	13	5.47
see	2	5.41
be	27	4.43

subject o	of .	
		18.60
continue	2	7.66
hold	2	6.49
make	2	4.40
be	18	3.85
friend	s are	
have	3	3.36

1	modifier		
			39.65
	close	Z	9.45
	fair-weather	2	9.15
	peer	2	8.96
	intimate	2	8.81
	neighborhood	3	8.75
	old	11	8.57
	old friends		
	liberal	2	8.55
	good	14	8.22

2

8.05

9.72

7.94

5.86

Linguisii	_	0.05
boy	4	7.45
few	3	6.86
many	4	6.69
and/or		
		24.91
relative	3	9.74

Fnolish

neighbor

anyone

school

pp of-i		
Arthur and Arthur		10.18
mine	7	10.75
Heard	2	10.44

5

pp obj of-i		
		6.67
Many	2	9.23
lot	3	8.37

pp obj to-i 4.56		
		4.56
write	2	6.69

predicate of		
		1.40
stranger	2	9.69

pp at-i		
		1.40
court	2	7.86

ИСТОРИЯ СОЗДАНИЯ ЛИНГВИСТИЧЕСКИХ КОРПУСОВ

- **1970 годы** частотный словарь русского языка Л. Н. Засориной, 1 миллион слов
- 1980-е годы- русский корпус, созданный в Университете Уппсалы, Швеция, 1-1,5 миллиона слов
- 1980-е годы Bank of English, Бирмингемский Университет, на данный момент Word Banks Online, 500-550 миллионов слов
- 1985 год Машинный фонд русского языка, создававшийся по инициативе А. П. Ершова
- 1991-1994 годы British National Corpus, BNC, 100 миллионов слов

THANKYOU FOR ATTENTION!