Лабораторная работа 1.

1. Слов – 67

Символов (без пробелов) – 539

Символов (с пробелами) – 605

Символов в латинской графике – 93

Чисел – 3 (цифр 6)

Средняя длина слов – 8.04477612

Средняя длина слов русского языка – 9.8

Средняя длина слов в латинской графике – 5.79569892

Русскому

а) Ø – норвежский; ö ß – немецкий; ρ ω – греческий; ё – русский; š – чешский

б) th – английский; sch – немецкий; šč – чешский

в) et – французский; the – английский; der – немецкий; och – шведский; için – турецкий

4)

| **Критерий** | **Английский** | **Немецкий** |
| --- | --- | --- |
| Типичные артикли | a/an, the | die, der, des, das, dem, |
| Указательные местоимения | this, that, those, these | der, dieser, jener, solcher, es, selbst |
| Местоимения 3-го лица | he, she, it, they | er, sie, es, sie, Sie |
| Отдельные формы вспомогательных глаголов | do/does, is/am/are, have/has; did, was/were, had; will, will be, will have | haben, sein, warden; hatte, war, wurde; gehabt, gewesen, geworden |
| Основные предлоги и союзы | in, on, of, after, for, until, with, if, because, at, so, as, that, by, since, | mit, nah, aus, zu, von, an, seit, bis, in, ohne, bei, über, für, unter |
| Другие частотные слова | have, see, know, say, can, get, make, time, take, think | das Leben, lesen, machen, denken, reisen, spielen, warten, werden, wohnen, das Alter, wollen, die Eltern |

5)

| **Критерий** | **Русский** |
| --- | --- |
| Типичные артикли | - |
| Указательные местоимения | тот, этот, такой, таков, столько |
| Местоимения 3-го лица | он, она, оно, они |
| Отдельные формы вспомогательных глаголов | начать, уметь, любить, рад, иметь желание, стать, успеть, нравиться, должен, готов, обязан, достоин |
| Основные предлоги и союзы | в, к, с, от, над, под, через, про, у, на, около, по, и, а, но, да, если, как, так, тоже, то, чтобы, потому что |
| Другие частотные слова | я, быть, он, она, они, мочь, год, тот, человек, сказать, быть, себя, один, ещё, уже, говорить, мой, знать, жизнь, два, другой, очень, первый, раз, где |

6) Прилагательные на эсперанто имеют одно и то же окончание “…a”, а существительные “…o”

1. Лингвистическая защита информации – нахождение закономерностей и частотности, шифровка, дешифровка текстов

2. Прикладную и компьютерную лингвистику нельзя считать синонимами, так как компьютерная лингвистика входит в прикладную. Также компьютерная лингвистика решает конкретную прикладную задачу, к примеру, автоматизация, в отличии от прикладной, которая решает какие угодно прикладные задачи с использованием знания о языке. Помимо этого, в компьютерной лингвистике требуется знание в машинном обучении, следовательно её от прикладной лингвистики и отличает эта компьютерная составляющая.

3. Основные направления компьютерной лингвистики:

* Информационный поиск
* Машинный перевод
* Чат-боты и диалоговые системы
* Анализ текста (Text Mining)
* Создание лингвистических ресурсов
* Синтез, распознавание речи
* Анализ тональностей

Создание лингвистических ресурсов.

Лингвистические ресурсы подразделяются на лексические и текстовые.

**Лексические ресурсы** – словари (термин не имеет однозначности), тезаурусы (семантический словарь, в котором представлены разные смысловые связи (отношение род-вид, часть-целое, синонимы, ассоциации; к примеру, стол – разновидность мебели)) и онтологии – набор понятий определённой области знаний, который ориентирован на многократное использование для различных задач, отличие от тезауруса в узконаправленности, также онтологии часто могут быть представлены в виде графа.

**Текстовые ресурсы** – корпуса (массив текстов, собранный по определённому принципу и обладающий лингвистической разметкой (присвоение тегов).

Свойства корпуса – репрезентативность (разнообразие явлений в текстах (обработка текстов в Твиттере и Вк)), сбалансированность.

Разметка может быть экстралингвистическая (частое размечание про текст и о тексте (об авторе, жанр, дата написания), структурная (разметка глав, абзацев, предложений), собственно-лингвистическая (морфологическая, разметка по части речи; синтаксическая, связь между словами; семантическая, значения тех или иных слов, где цвета, где запахи; анафорическая, к какому слову относится местоимение, прописываем, где возникает анафора).

1. С точки зрения компьютерной грамматики “Ромашка” будет лишним, так как начинается на Р; С точки зрения компьютерной семантики слово “Липа” будет лишним, так как это дерево, а остальное цветы.
2. An = прилагательное; Nn = существительное в именительном падеже ; Ng = существительное в единственном числе родительного падежа; N(A)n = существительное, образованное от прилагательного; N(V)n = существительное, образованное от глагола; V = глагол; D = наречие.
3. N1nN2g ↔ A(N2)nN1n, AnN1 ↔ N(A)nN1g, V из Ng ↔ N(V)n из Ng
4. а) Яблоко из магазина ↔ Магазинное яблоко

б) Трудолюбивый студент переписал аккуратно плохую курсовую