**Лабораторная работа 1**

**Уровень 1**

**Лингвистическая защита информации** - это область информационной безопасности, которая использует знания из лингвистики для защиты информации от несанкционированного доступа. Лингвистические знания используются, например, для разработки сложных паролей, аутентификации по голосу или тексту, анализа текстов на предмет возможных угроз безопасности и т.д.

**Нельзя считать синонимами прикладную и компьютерную лингвистику**. Прикладная лингвистика занимается применением лингвистических знаний в практических областях, как например, машинный перевод, обработка естественного языка и т.д. В то время как компьютерная лингвистика изучает применение компьютеров и информационных технологий в анализе и обработке языка.

**Основные направления компьютерной лингвистики включают в себя машинный перевод, обработку естественного языка, извлечение** информации, генерацию текста, распознавание речи и другие. Одним из направлений является машинный перевод, который изучает автоматическое переведение текста с одного языка на другой с помощью компьютерных программ и алгоритмов. Это направление активно развивается с развитием искусственного интеллекта и глубокого обучения.

1. Раздел "[Лингвистика. Язык](https://edu.vsu.ru/mod/assign/view.php?id=918508)", Упражнения 1-6

1. Определите статистические показатели приведенного ниже текста смешанного языкового типа.

[Проект](https://edu.vsu.ru/mod/resource/view.php?id=919336)ы Cíbola/Oleada реализуют обширные компьютерные системы лингвистического анализа текстов, представленных в Unicode. Компоненты системы включают средства работы с мультиязыковыми текстами (MUTT), построения конкорданса (XConcord) для текстов на более чем 16 языках, статистического анализа, автоматического [перевод](https://edu.vsu.ru/mod/assign/view.php?id=919076)а, различные словари и тезаурусы. Некоторые версии этих компонентов доступны для бесплатной загрузки после процедуры формальной регистрации. Все компоненты реализованы в среде X11 Window System для SunOs и Solaris (источник: Проекты Cíbola/Oleada <http://rvb.ru/soft/catalogue/c01.html>).

|  |  |
| --- | --- |
| Слов | 35 |
| Символов (без пробелов) | 186 |
| Символов (с пробелами) | 256 |
| Символов в латинской графике | 74 |
| Чисел | 2 |
| Средняя длина слов | 6.77 |

2. Средняя длина слов тексте - 6,77 слов.

Средняя длина слов в латинской графике составляет 2,86 слов.

У русского языка такая же средняя длина слов

3. Определите, каким языкам соответствуют следующие специфические буквы, буквосочетания и слова:

а) Ø ö ß ρ ω ё š, - **Норвежский, Шведский, Датский**

б) th sch šč, - **Немецкий**

в) et, the, der, och, için. – **Немецкий, Английский, Турецкий**

4. Создайте диагностический словарь для определения языка на материале текстов на двух разных языках (на ваш выбор). Для этого заполните следующую таблицу.

Английский: Learning English is an exciting journey that opens up a world of opportunities. It's a language spoken by millions of people around the globe, making it essential for communication in today's interconnected world. By practicing English regularly, whether through reading, writing, speaking, or listening, you can improve your skills and become more confident in using the language.

Французский: Apprendre l'anglais est un voyage passionnant qui ouvre un monde d'opportunités. C'est une langue parlée par des millions de personnes à travers le monde, ce qui la rend essentielle pour la communication dans le monde interconnecté d'aujourd'hui. En pratiquant régulièrement l'anglais, que ce soit en lisant, en écrivant, en parlant ou en écoutant, vous pouvez améliorer vos compétences et devenir plus confiant dans l'utilisation de la langue.

|  |  |  |
| --- | --- | --- |
|  | Англ | Франц |
| Типичные артикли | “the”, “a” и “an” | “le”, “la”, “les” |
| Указательные местоимения | That, this | - |
| Местоимения 3-го лица | You, it | “je”, “tu”, “il”, “elle” |
| Отдельные формы вспомогательных глаголов | - | “avoir”, “faire”. |
| Основные предлоги и союзы | “in”, “on”, of, “and”, “but”, | de”, “à”, “en”, “pour” |
| Другие частотные слова |  |  |

6. Прочитайте несколько фраз на эсперанто. Назовите морфологические диагностические показатели этого языка, учитывая, что существительные и прилагательные на эсперанто всегда имеют одни и те же окончания.

Рус.                                                                                         Эсперанто

зеленое дерево                                                                  verda arbo

старый человек                                                                   maljuna viro

хороший друг                                                                      bela amiko

Сущ и прил на эсперанто имеют одинаковые окончания, что упрощает их распознавание.

Эсперанто имеет простую систему склонений существительных и прилагательных, основанную на числе и падеже.

Глаголы на эсперанто можно определить по характерным окончаниям, которые изменяются в зависимости от времени, лица и числа.

Наречия и местоимения также имеют свои уникальные окончания, которые помогают их идентифицировать.

Числительные на эсперанто также имеют характерные окончания, позволяющие их отличить от других частей речи.

Эсперанто не имеет артиклей и множественного числа (кроме некоторых исключений), что также является диагностическим показателем.

В эсперанто используются приставки и суффиксы для формирования новых слов и изменения их значения.

Информация, информационные технологии

1. Найдите лишнее в приведенном ниже списке. Решите данную задачу с точки зрения компьютерной семантики и компьютерной грамматики.

Ландыш, левкой, лаватера, лютик, люпин, ромашка, липа.

Если программа предназначена для работы с данными о растениях, то словом “липа” в компьютерной семантике может называться не только растение, но и процесс подделки данных

Если рассматривать список как текст, записанный на языке программирования, то лишними могут быть знаки препинания или пробелы, поскольку они не являются значимыми символами в большинстве языков программирования.

2.

а) AnNn ↔ N(A)nNg - соотношение между прилагательным и существительным. ("великое море" (Апресян, 1966)).

b) VNa ↔ N(V)nNg - соотношение между глаголом и существительным. ("прыгать высоко" (Апресян, 1966)).

c) N1nN2g ↔A(N2) nN1n - соотношение между двумя существительными, где одно из них является прилагательным для другого. ("городская жизнь" (Апресян, 1966)).

d) VвNa ↔ N(V)nвNa - соотношение между глаголом с предлогом и существительным.("думать о счастье" (Апресян, 1966)).

e) DV ↔ A(D)nN(V)n - соотношение между деепричастием и прилагательным с существительным. ("стремиться к успеху" (Апресян, 1966)).

3.

а) NвNa ↔ V(N)nвNg (врачебный визит ↔ визит врача)

б) NвNa ↔ V(N)nNa (аромат сада ↔ ароматный сад)

в) NвNa ↔ V(N)nNa (выход из дома ↔ выхожу из дома)

4.

а) теннисистка из России ↔ российская теннисистка (N1n из N2g ↔ A(N2)nN1n)

б) Мужчина среднего возраста покупал дом. (AnNnVDAaNa)