Математика в филологии

- Криптография и расшифровка древних текстов
- Обоснование авторства текстов
- Перевод текстов с одного языка на другой
- Лексический анализ текстов
- Другие задачи…

Статистические методы анализа лексики

Лексика представляет собой *статистически* организованную структуру:

- Вероятностные характеристики слова проявляются в неодинаковой частотности их в речи, в многообразных видах лексических связей
  - Установлено, например, что самые частотные слова в естественном языке, как правило, являются наиболее краткими, наиболее древними, наиболее простыми по морфологической структуре, наиболее многозначными
- Статистические методы используются для изучения характера семантических связей между словами.
  - Так, например, установлено, что слова, часто встречающиеся вместе в определенном отрезке текста, теснее связаны между собой по смыслу, чем слова, реже появляющиеся рядом в этом же отрезке текста.

Математическая лингвистика

- **Математическая лингвистика -** математическая дисциплина, разрабатывающая формальный аппарат для описания строения естественных и некоторых искусственных языков. Возникла в 50-х годах 20 века.

- Базируется на методах алгебры, теории алгоритмов и теории автоматов.

- Направления математической лингвистики:
  - Изучение способов математического описания правильных текстов (в первую очередь предложений)
    - Для описания строения (синтаксической структуры) предложения можно либо выделить в нём "составляющие" — группы слов, функционирующие как цельные синтаксические единицы, либо указать для каждого слова те слова, которые от него непосредственно зависят (если такие есть). Математические объекты, возникающие при таком описании структуры предложения, называются деревом составляющих (1-й способ) и деревом синтаксического подчинения (2-й способ).
  - Теория формальных грамматик (Н.Хомский)
    - Изучает способы описания закономерностей, которые характеризуют уже не отдельный текст, а всю совокупность правильных текстов того или иного языка. Эти закономерности описываются путём построения "формальной грамматики" — абстрактного "механизма", позволяющего с помощью единообразной процедуры получать правильные тексты данного языка вместе с описаниями их структуры.
    - Используется в основном для при разработке и описании искусственных языков (например, языков программирования)
  - Построение аналитических моделей языка, в которых на основе тех или иных данных о речи, считающихся известными (например, множества правильных предложений), производятся формальные построения, дающие некоторые сведения о структуре языка.

Лингвистика и алгебра

Лев Владимирович Щерба, 1930 г.

«Глокая куздра штеко будланула бокра и кудрячит бокрёнка»

$$y = x + a$$

Математический язык

- Математика: мышление, чувствование и язык.
- Язык – это система условных знаков, принятых в некотором сообществе и обеспечивающая коммуникацию его членов.
- Язык математики как и любой другой язык состоит из совокупности высказываний (предложений). Математические высказывания это математические символы, объединенные формулой.

Математика – язык символов и формул.

Математический язык (продолжение)

- Язык в широком смысле – это словарь, грамматика, рассказы, повести, пьесы и романы, написанные на этом языке.
- В математическом языке:
  - словарь и грамматика – *математическая операционная система*
  - рассказы, повести и прочее – *математические модели*

Математический язык (продолжение)

МАТЕМАТИЧЕСКАЯ ОПЕРАЦИОННАЯ СИСТЕМА

Идеализированные объекты

Числа,

Фигуры,

Векторы,

Функции,

Матрицы,

…

Абстрактные структуры

(например, множества)

Действия (операции):

Сложение,

Вычитание,

Умножение,

Деление,

Дифференцирование,

Интегрирование,

…

Аксиомы, теоремы, определения

Элементы теории множеств

- Множество – первичное понятие современной математики, это понятие не определяется через другие понятия а только поясняется.
- **Множество** –
  - «объединение в одно целое объектов, хорошо различимых нашей интуицией или нашей мыслью» (Георг Кантор, 1845-1918, немецкий математик, основатель теории множеств);
  - совокупность каких-либо объектов
- Объекты, входящие в множество – ***элементы множества.*** Например: числа, буквы, люди и т.п.

Элементы теории множеств (продолжение)

- Множества, состоящие из конечного числа элементов – *конечные множества*
- Множества, состоящие из бесконечного числа элементов – *бесконечные множества*
- Обозначения:
  - Множества – A, B, X
  - Элементы множества – a, b, x

Элементы теории множеств (продолжение)

- Обозначения:

$x \in X$	Объект *x* есть элемент множества *X*
$x \notin X$	Объект *x* не принадлежит множеству *X*
$A \subset B$	Множество А содержится в множестве В (входит в множество В)
$\varnothing$	Пустое множество

Множества А и В называются ***равными*** (А = В), если они состоят из одинаковых элементов.

Элементы теории множеств (продолжение)

- **Числовые множества**
  - Множество натуральных чисел $N = \{1,\ 2,\ 3,\ \ldots\}$
  - Множество целых чисел $Z = \{\ldots, -3,\ -2,\ -1,\ 0,\ 1,\ 2,\ 3,\ \ldots\}$
  - Множество рациональных чисел *Q*
  - Множество действительных чисел *R*

$$N \subset Z \subset Q \subset R$$

Элементы теории множеств (продолжение)

Упражнения:

1. Какие из следующих множеств геометрических фигур на плоскости равны между собой:
   - A – множество всех квадратов;
   - B – множество всех прямоугольников;
   - C – множество всех четырехугольников с прямыми углами;
   - D – множество всех прямоугольников с равными сторонами;
   - F – множество всех ромбов с прямыми углами

2. Для каждого из слов: «сосна», «осколок», «насос», «колос» составьте множество его различных букв. Имеются ли среди них равные?

Алгебраические операции над множествами

- **Объединением множеств А и В** называется новое множество, которое обозначается $A \cup B$ и состоит из всех элементов, принадлежащих хотя бы одному из множеств А или В, т.е

$$A \cup B = \{ x \in A \text{ или } x \in B \}$$

Например: $\{1,2,3\} \cup \{2,3,4\}=\{1,2,3,4\}$

- **Пересечением множеств А и В** называется новое множество, которое обозначается $A \cap B$ и состоит из всех элементов, принадлежащих одновременно множествам А и В, т.е.

$$A \cap B = \{ x \in A \text{ и } x \in B \}$$

Например: $\{1,2,3\} \cap \{2,3,4\}=\{2,3\}$

Алгебраические операции над множествами

- **Разностью множеств А и В** называется новое множество, которое обозначается А\В и состоит из всех элементов множества А, не принадлежащих множеству В, т.е

$$A \setminus B = \{ x \in A \text{ и } x \notin B \}$$

Например: {1,2,3} \ {2,3,4}={1}

- **Симметрическая разность АΔВ** есть множество всех элементов, принадлежащих или А, или В (но не обоим вместе)

$$A \Delta B = (A \setminus B) \cup (B \setminus A)$$

Например: {1,2,3} Δ {2,3,4}={1,4}

Алгебраические операции над множествами

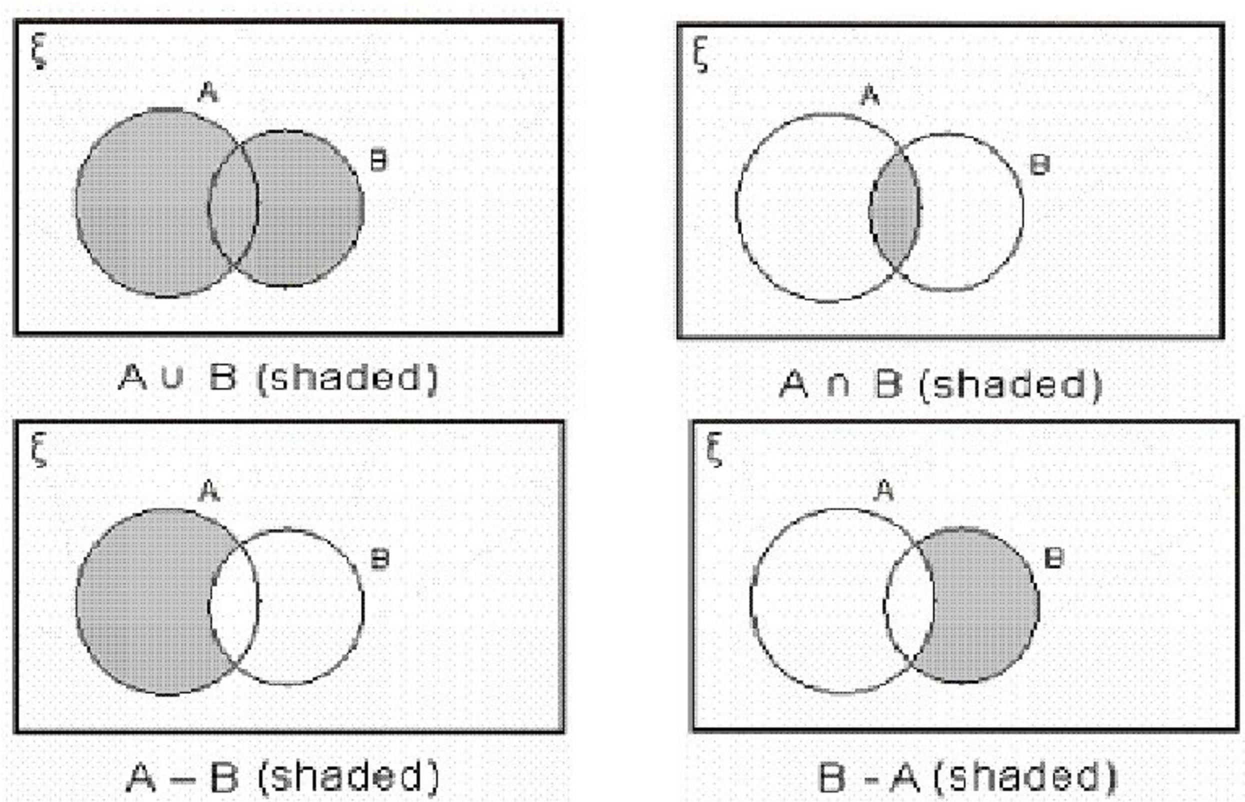
- **Декартовым произведением множеств A и B** называется новое множество, обозначаемое $A \times B$, элементами которого являются всевозможные пары (a, b), где $a \in A$, $b \in B$, то есть $A \times B = \{(a, b) \mid a \in A, b \in B\}$.

  Например, если $A=\{1, 2, 3\}$, $B=\{1, 3, 4\}$,

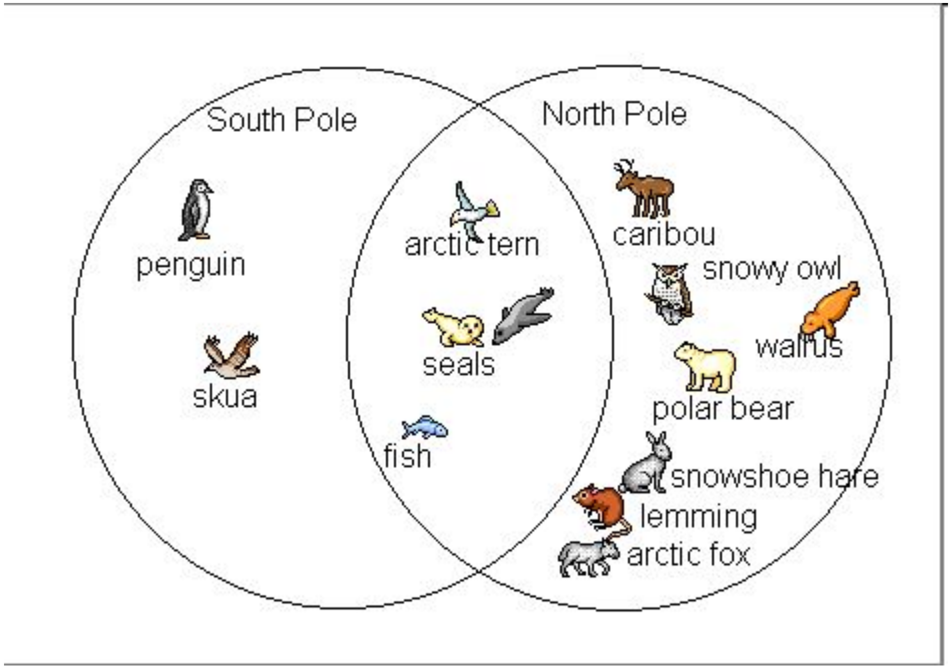
  то $A \times B=\{(1, 1), (1, 3), (1, 4), (2, 1), (2, 3), (2, 4), (3, 1), (3, 3), (3, 4)\}$.

  Отметим, что с декартовым произведением связано понятие координатной плоскости. Множество координат точек координатной плоскости является декартовым произведением $R \times R$, где R – множество действительных чисел – координаты точек по оси x и оси y, соответственно.

Алгебраические операции над множествами. Круги Эйлера или диаграммы Венна.



Алгебраические операции над множествами. Круги Эйлера или диаграммы Венна.



Алгебраические операции над множествами

Упражнения

1. Выпишите все подмножества множества B = {1, 2, 3}

2. Запишите множество A перечислением его элементов, если A = {$x \in N$, $2<x<8$}

3. Даны два множества: A = {1, 2 ,3 ,4} и B = {3, 4, 5, 6}. Записать множества, представляющие:
а) объединение $A \cup B$;
б) пересечение $A \cap B$;
в) разность $A \setminus B$, $B \setminus A$;
г) симметрическую разность $A \Delta B$, $B \Delta A$;
д) декартово произведение $A \times B$.

4. Определить пересечением или объединением множеств A={5,7,8} и B={1,5,6} является множество C = {1,5,6,7,8}?

5. Проверить выполняется ли переместительный закон умножения для декартова произведения двух множеств, т.е. верно ли, что $A \times B = B \times A$? В качестве множеств A и B возьмите множества: A={2}; B={1,3}.

6. Проверьте на примере множеств A={2}; B={1,3} и C={4,5} выполняется ли сочетательный закон для декартова произведения, т.е. верно ли, что $(A \times B) \times C = A \times (B \times C)$.

Численность множества

Пусть ***A*** и ***B*** – конечные множества. Число элементов множества ***A*** условимся обозначать символом ***m(A)*** и называть ***численностью*** множества ***A.***

Число элементов объединения и разности двух конечных множеств:

Определим численность объединения множеств *A* и *B*.

Если множества ***A*** и ***B*** не пересекаются, то $m(A \cup B) = m(A) + m(B)$.

Таким образом, численность объединения конечных непересекающихся множеств равна сумме численностей этих множеств.

Если множества ***A*** и ***B*** пересекаются, то в сумме $m(A) + m(B)$ число элементов пересечения $A \cap B$ содержится дважды: один раз в $m(A)$, а другой – в $m(B)$. Поэтому, чтобы найти численность объединения $m(A \cup B)$, нужно из указанной суммы вычесть $m(A \cap B)$.

Таким образом: $m(A \cup B) = m(A) + m(B) - m(A \cap B)$

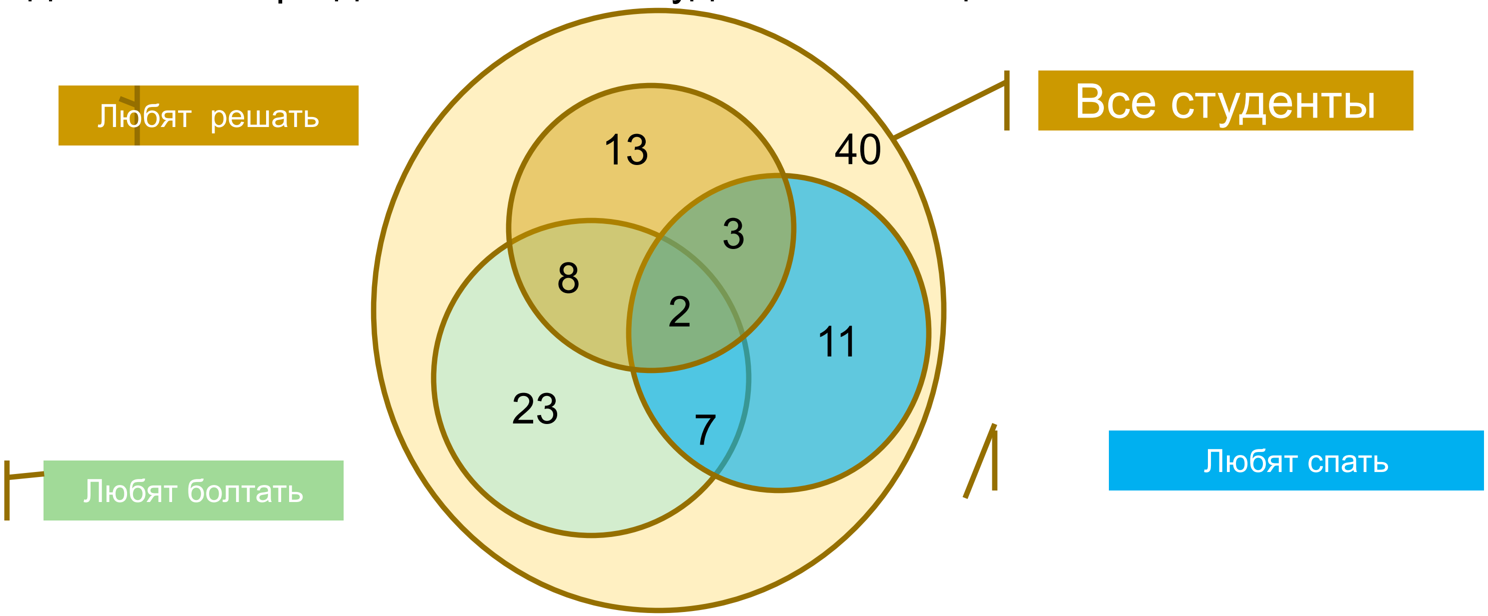
Определим теперь численность разности множеств *A* и *B.*

Если множества ***A*** и ***B*** не пересекаются, то $A \setminus B = A$, и поэтому $m(A \setminus B) = m(A)$.
Если множества ***A*** и ***B*** пересекаются, то $m(A \setminus B) = m(A) - m(A \cap B)$.
Если $B \subset A$, то $A \cap B = B$, и, следовательно, $m(A \setminus B) = m(A) - m(B)$.

Использование теории множеств для решения задач

Задача 1

В группе 40 студентов. Из них 23 любят болтать на занятиях, 13 — решать задачи, 11 любят на занятиях спать. Среди тех, кто болтает на занятиях, постоянно засыпают — 7, а среди тех, кто решает задачи, засыпают только 3. Болтать и решать задачи умеют 8 человек; а 2 человека успевают на одной паре делать все три дела. Сколько студентов вообще ничего не любят?



Использование теории множеств для решения задач

Задача 2

В группе из 100 туристов 70 человек знают английский язык, 45 знают французский язык и 23 человека знают оба языка. Сколько туристов в группе не знают ни английского, ни французского языка?

Решение задачи:

Обозначим:

U – универсальное множество, т.е. множество всех туристов,
A – множество туристов, знающих английский язык,
B – множество туристов, знающих французский язык.

Необходимо найти количество туристов, не знающих ни одного языка, т.е. количество элементов множества $\mathbf{D = U \setminus (A \cup B)}$.

Дано (по условию): $m(U) = 100$ (чел.)
$m(A) = 70$ (чел.)
$m(B) = 45$ (чел.)
$m(A \cap B) = 23$ (чел.)

Найти: $m(D) = m(U) – m(A \cup B)$ - ?

Решение: Используя формулу, находим количество туристов, знающих хотя бы один язык:

$\mathbf{m(A \cup B) = m(A) + m(B) – m(A \cap B)} = 70 + 45 - 23 = 92,$ $\Rightarrow$
количество туристов, не знающих ни одного языка:
$\mathbf{m(D) = m(U) - m(A \cup B)} = 100 – 92 = 8$ (чел.)

Ответ: 8 чел.

Использование теории множеств для решения задач

Задача 3
20 мальчиков поехали на пикник. При этом 5 из них обгорели, 8 были сильно покусаны комарами, а 10 остались всем довольны. Сколько обгоревших мальчиков не было покусано комарами? Сколько покусанных комарами мальчиков также и обгорели?

Задача 4
Из 40 предложений 30 содержат предлог «в», 27 предлог «на», в пяти предложениях нет ни того, ни другого. Сколько предложений содержат оба предлога?