

Информационные технологии в обработке текстов

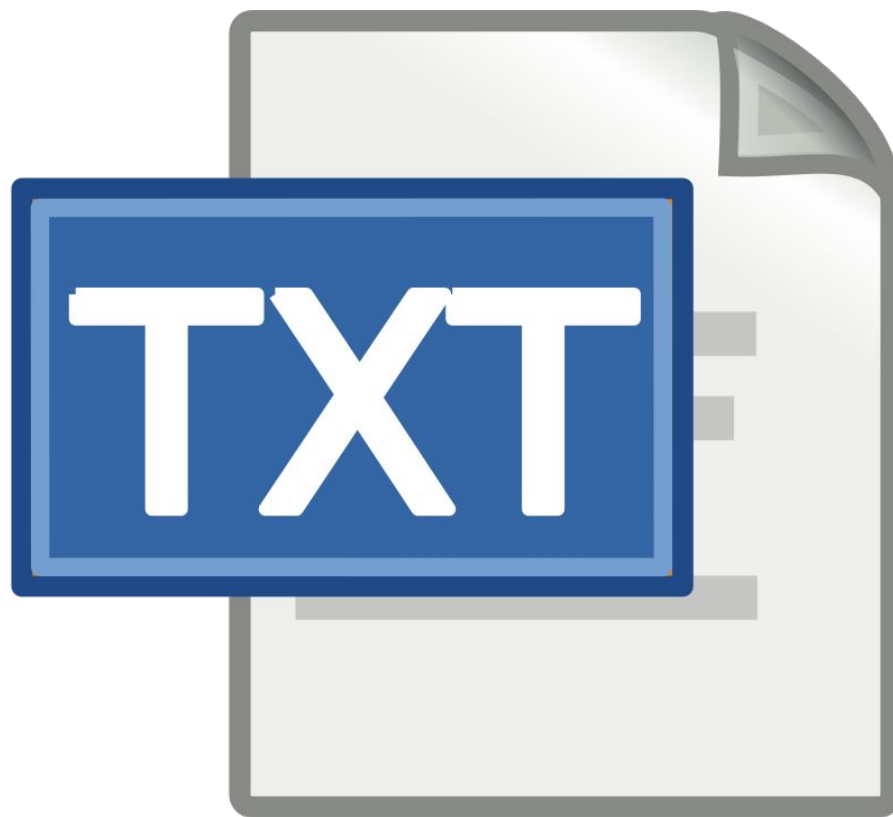
Подготовила:
Студентка 1 курса
факультета РГФ
2 группы
Воробьева
Екатерина

Под информационными технологиями понимается вся совокупность программного и аппаратного обеспечения, методов и средств организации автоматизированной обработки информации, представленной в цифровой форме.

Представление текста

“Текст - это упорядоченный набор слов, предназначенный для того, чтобы выразить некий смысл. В лингвистике термин используется в широком значении, включая в себя и устную речь”.

Текстовые файлы



Правила машинного набора текста

1. Все слова разделяются только пробелом.
2. Знаки препинания примыкают к предыдущему слову.
3. Скобки и кавычки всех видов примыкают к первому и последнему слову заключенного в них текста.
4. Текст разрывается только в конце абзаца.
5. Большие форматированные пробелы делаются вставкой символа табуляции.

Шрифт

Шрифт - это графический рисунок букв, цифр и символов, обладающий общими для всех символов стилистическими особенностями изображения.

Параметры шрифта

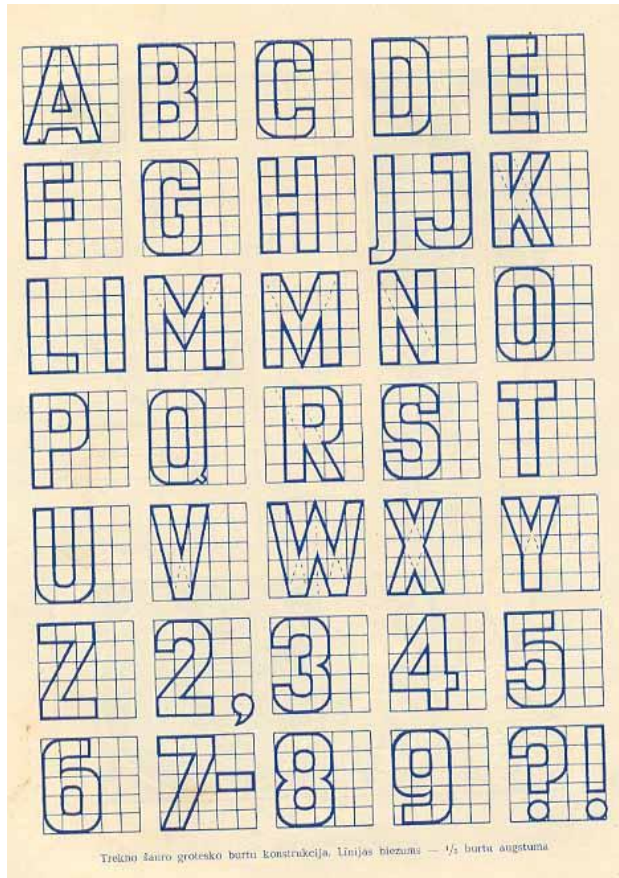
1. Рисунок шрифта
2. Кегль(размер шрифта) - чаще всего задается в типографских пунктах.
3. Начертание
4. Подчеркивание или зачеркивание шрифта

Antiqua

Антиквенный шрифт

*The most Captivating Type,
unquestionably excelling all others,
said Johan Verboet concerning
his Types, in the Year*

Акцидентный
шрифт



Рубленный шрифт

Структурирование текста

Параметры:

1. Выравнивание.
2. Отступы.
3. Абзацный отступ (красная строка).
4. Интервалы.
5. Буквица - *крупная выступающая первая буква абзаца.*

Поиск

Задача: поиск фрагментов, содержащих некоторые понятия, в достаточно большом массиве.

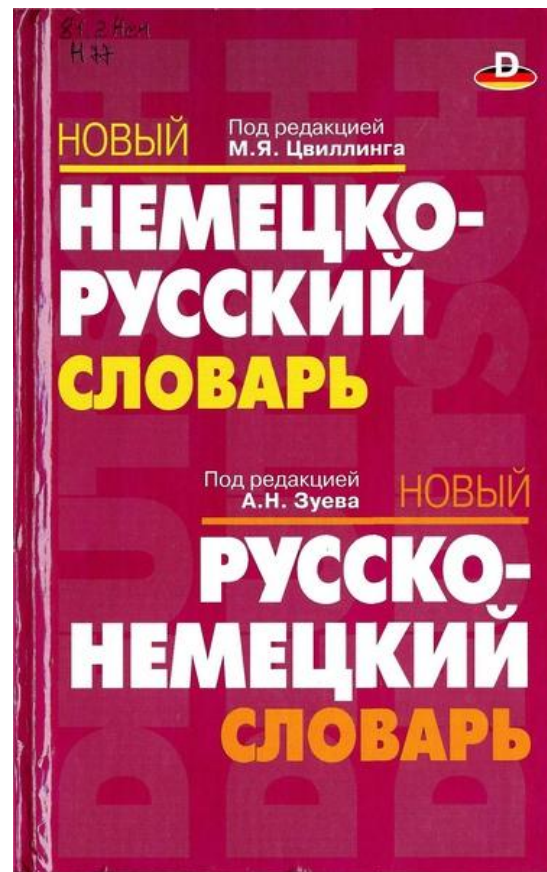
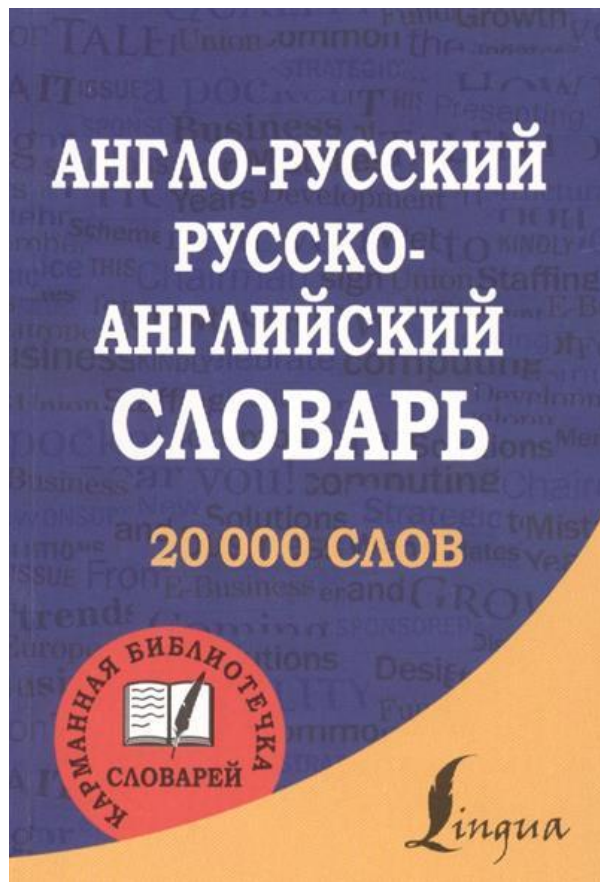
Первый подход

По сути, это *описание фрагмента текста, удовлетворяющего некоторым условиям, по тем частям, которые в нем содержатся, и их порядку.*

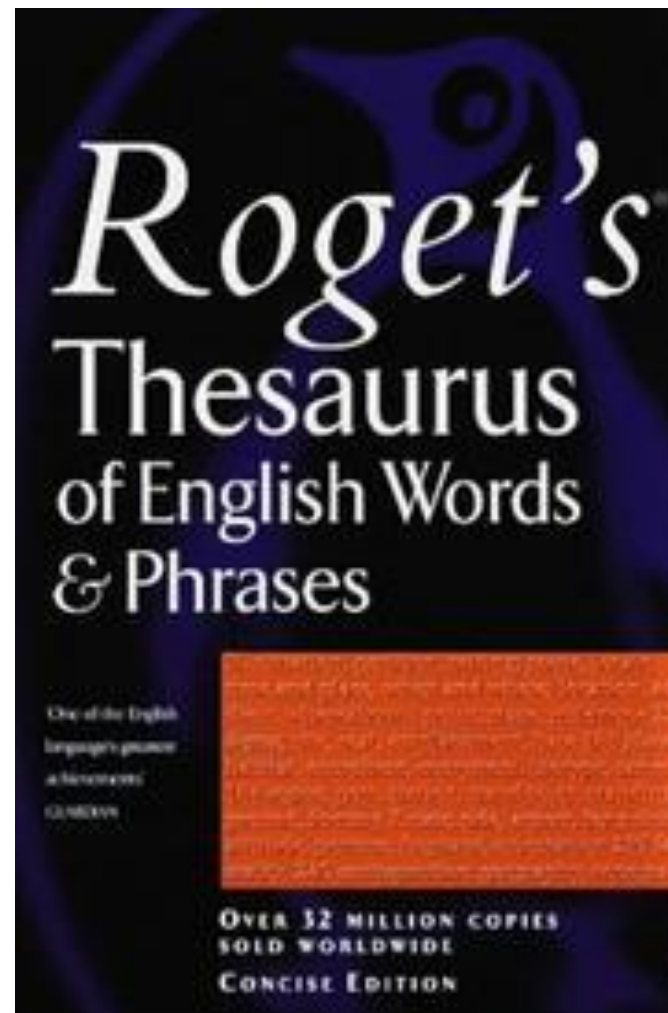
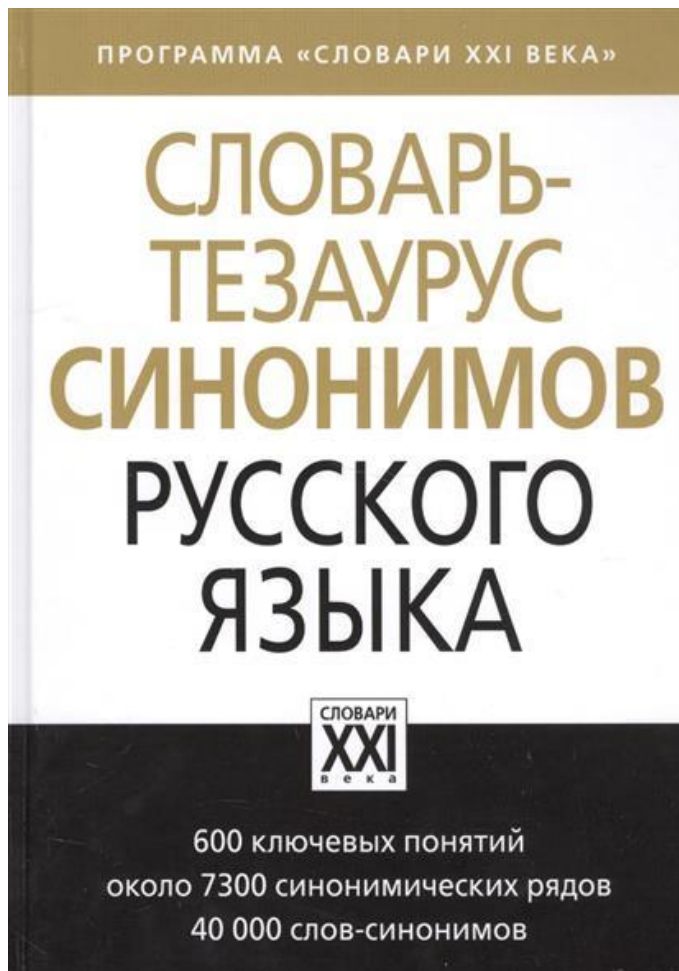
Второй подход

Он предусматривает предварительное создание *специального вида базы для ускорения поиска - индекса.*

Словари



Тезаурусы



- Современные системы обработки позволяют создавать краткие обзоры текстов (рефераты) или готовить перевод с одного естественного языка на другой.
- С применением специализированных по областям знания словарей современные системы автоматизированного перевода создают подстрочник, который может дать представление о смысле текста и в дальнейшем помочь переводчику в переводе документа.

Спасибо за внимание!