

Основы лингвистического обеспечения АСОИУ



Лекция №1

Компьютерная лингвистика

К.т.н., доцент Филиппович Анна Юрьевна

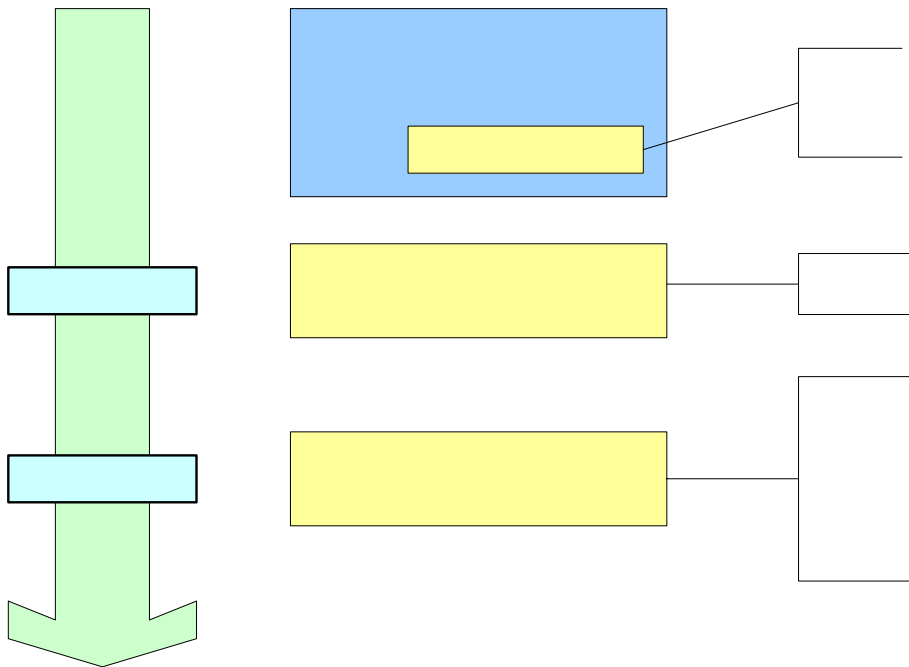
Лингвистика

- **Лингвистика** (от лат. lingua - язык) — это наука о языке.

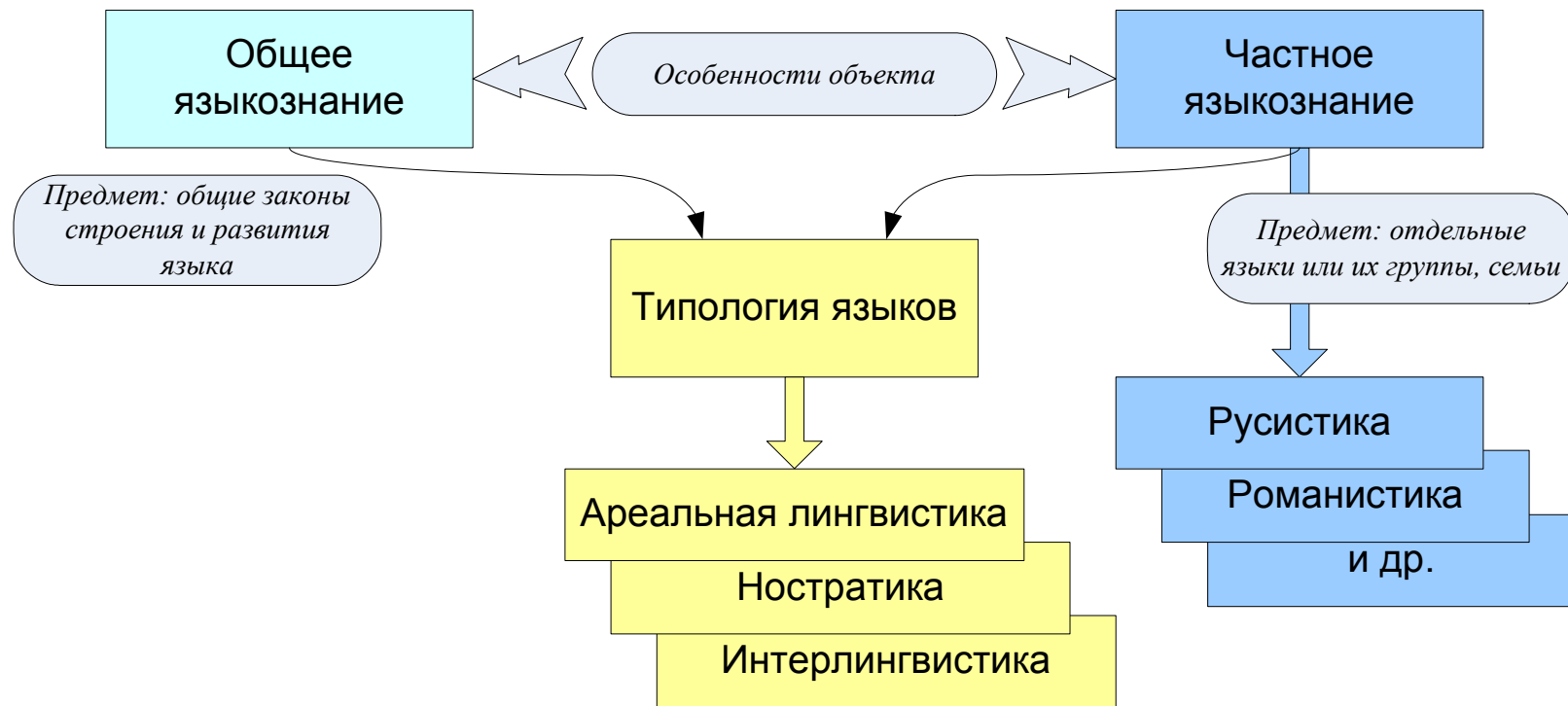
Лингвистика = языкознание = языковедение

- Предметом **языкознания** является язык во всем объеме его свойств и функций, строение, функционирование и историческое развитие языка.
 - **Язык** — это естественно возникшая в человеческом обществе и развивающаяся система облеченных в звуковую форму знаковых единиц, способная выразить всю совокупность понятий и мыслей человека и предназначенная прежде всего для целей коммуникации.
-

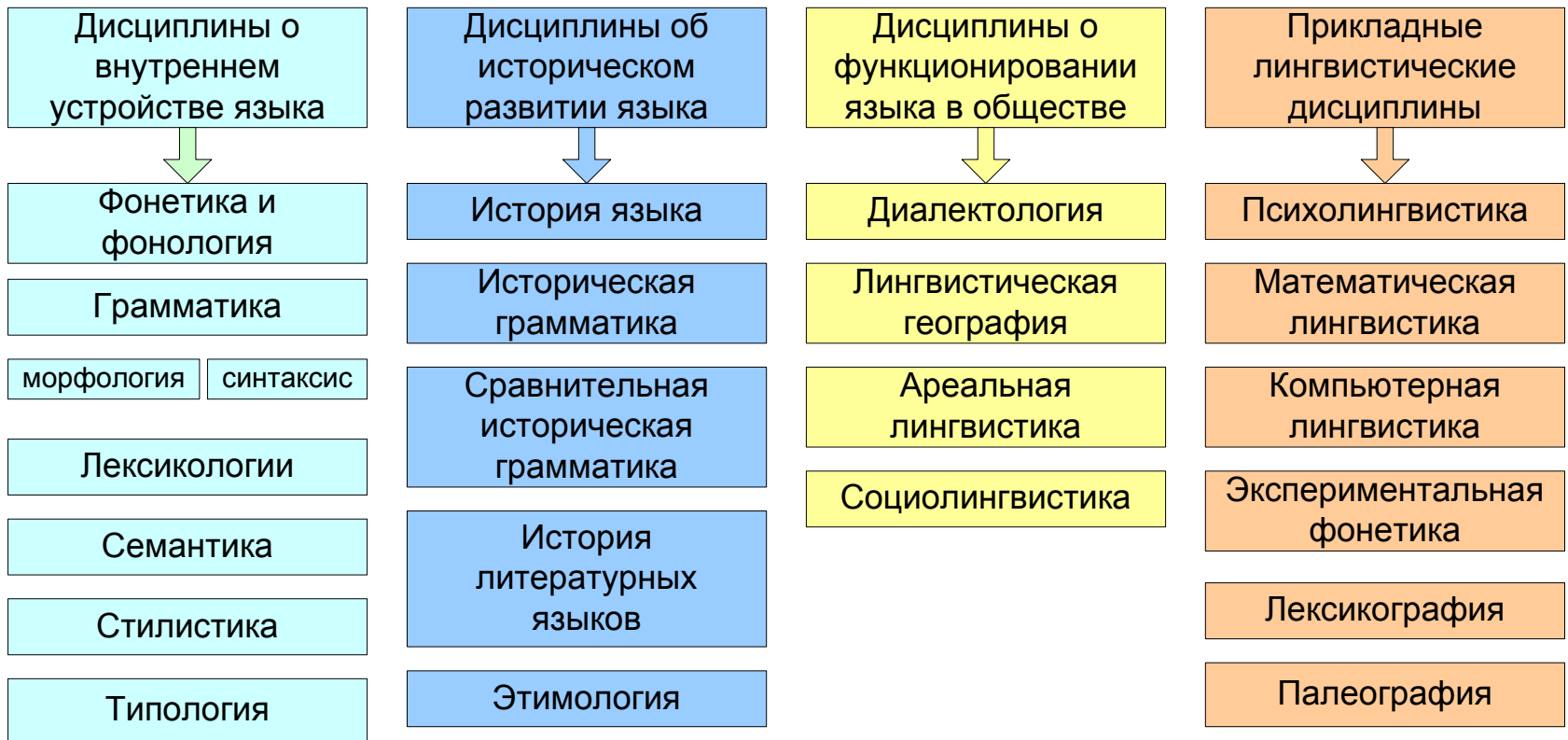
Предмет языкознания



Система языковедческих дисциплин



Классификация языковедческих дисциплин



...

Прикладная лингвистика

- **Прикладная лингвистика** — направление в языкознании, занимающееся разработкой методов решения практических задач, связанных с использованием языка.
-

Прикладные задачи языкознания

- создание и совершенствование письменностей;
 - создание систем: транскрипции, транслитерации иноязычных текстов, стенографии, письма для слепых;
 - повышение эффективности редакционно-издательских процессов типографского набора на основе рационального использования свойств языка;
 - упорядочение и стандартизация научно-технической терминологии;
 - изучение процессов и создание правил образования названий новых изделий, товаров, химических веществ;
-

Прикладные задачи языкознания

- разработка методов: смыслового отождествления текстов, аннотирования, реферирования, адекватного преобразования текстов в иноязычную форму;
 - отбор языкового материала в целях совершенствования преподавания языков;
 - создание специальных лингвистических справочников, создание и изучение искусственных языков.
-

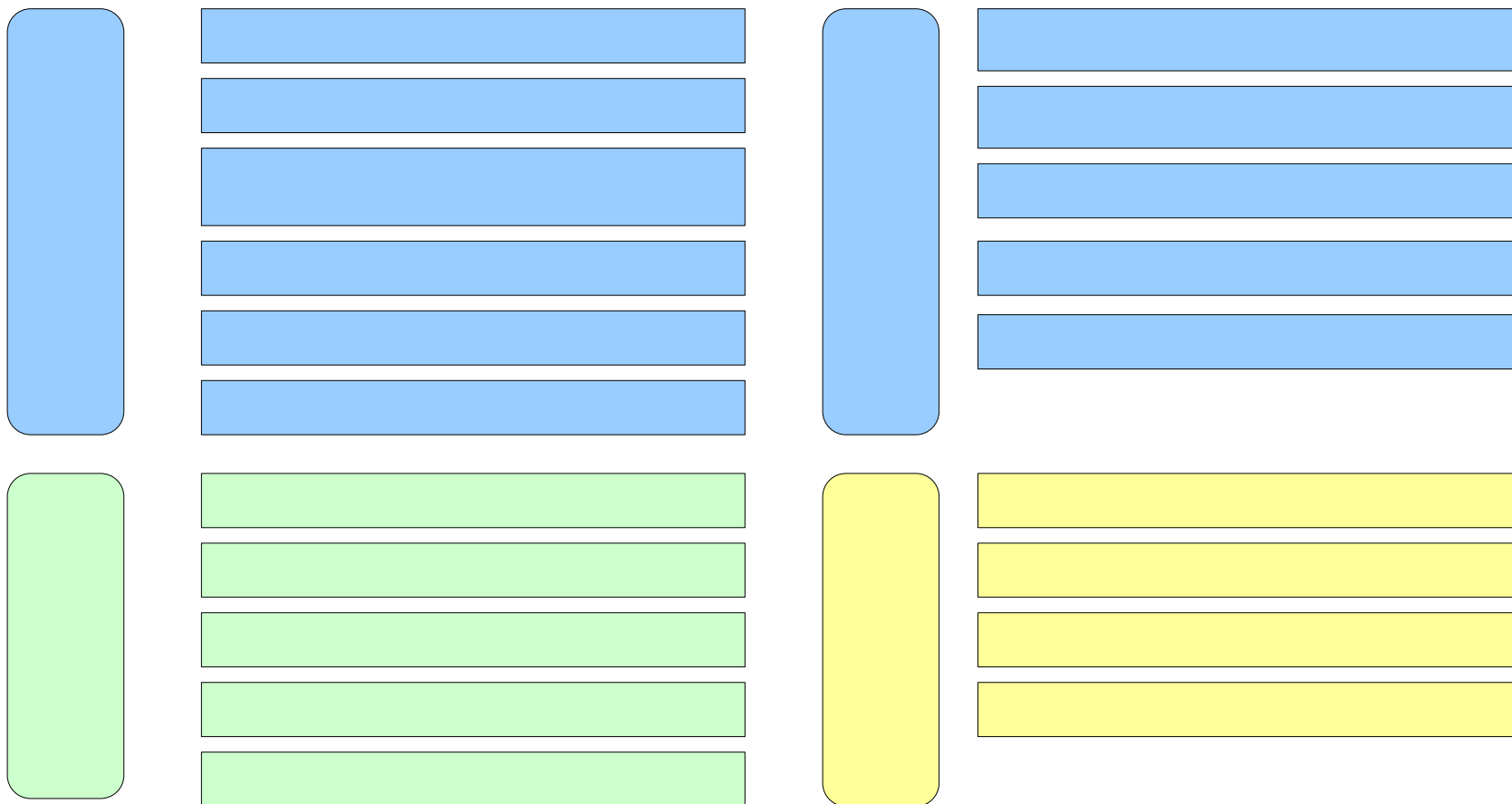
Термин «Прикладная лингвистика»



Прикладная лингвистика и функции языка

- С функциональной точки зрения **прикладная лингвистика** может быть определена как научная дисциплина, в которой изучаются и разрабатываются способы оптимизации функционирования языка.
-

Прикладная лингвистика и функции языка



Методы прикладной лингвистики

- ❑ Описание (классификации);
 - ❑ Моделирование;
 - ❑ Специальные методы и методики других наук.
-

Компьютерная лингвистика

- **Компьютерная лингвистика** - направление в прикладной лингвистике, ориентированное на использование компьютерных инструментов – программ, компьютерных технологий организации и обработки данных – для моделирования функционирования языка в тех или иных условиях, ситуациях, проблемных сферах и т.д., а также вся сфера применения компьютерных моделей языка в лингвистике и смежных дисциплинах.
-

Примеры технических специальностей, связанных с компьютерной лингвистикой

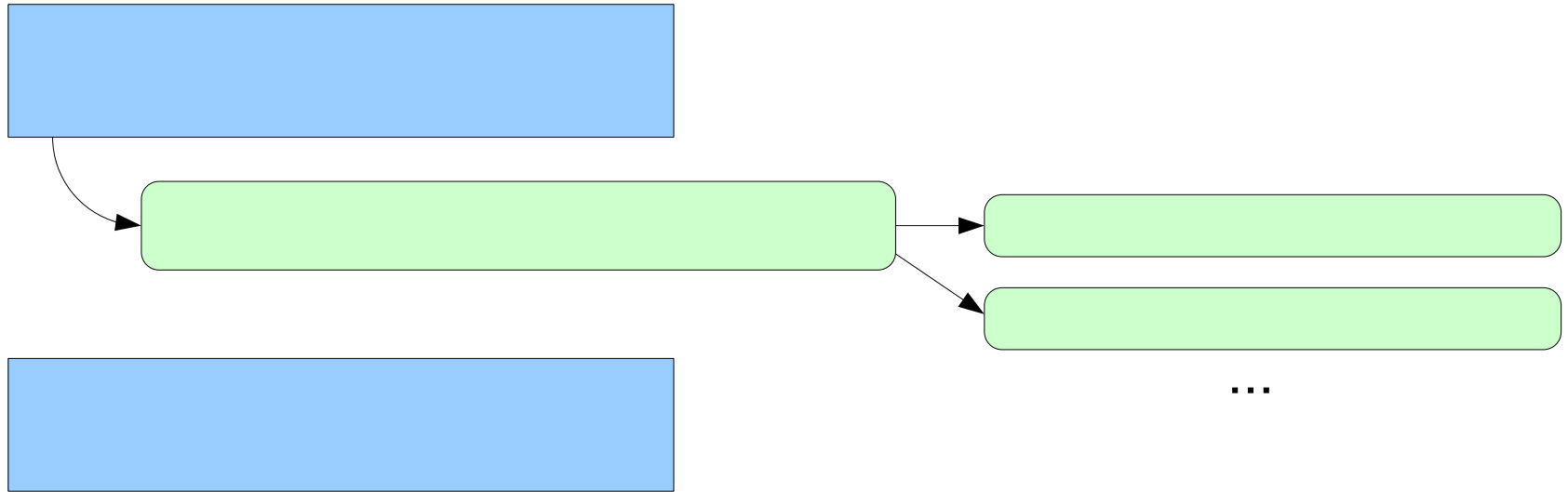
05.13.17 — Теоретические основы информатики:

- Область 5. Разработка и исследование моделей и алгоритмов анализа данных, обнаружения закономерностей в данных и их извлечения, разработка и исследование методов и алгоритмов анализа текста, устной речи и изображений.
- Область 6. Разработка методов, языков и моделей человеко-машинного общения; разработка методов и моделей распознавания, понимания и синтеза речи, принципов и методов извлечения данных из текстов на естественном языке.
- Область 9. Разработка новых интернет-технологий, включая средства поиска, анализа и фильтрации информации, средства приобретения знаний и создание онтологии, средств интеллектуализации бизнес-процессов.
- Область 10. Разработка основ математической теории языков и грамматик,

05.13.06 — Автоматизация технологических процессов и производств (по отраслям):

- — ... разработка лингвистического обеспечения автоматизированных систем различного типа АСУП, АСУТП и др.
-

Инструментарий компьютерной лингвистики



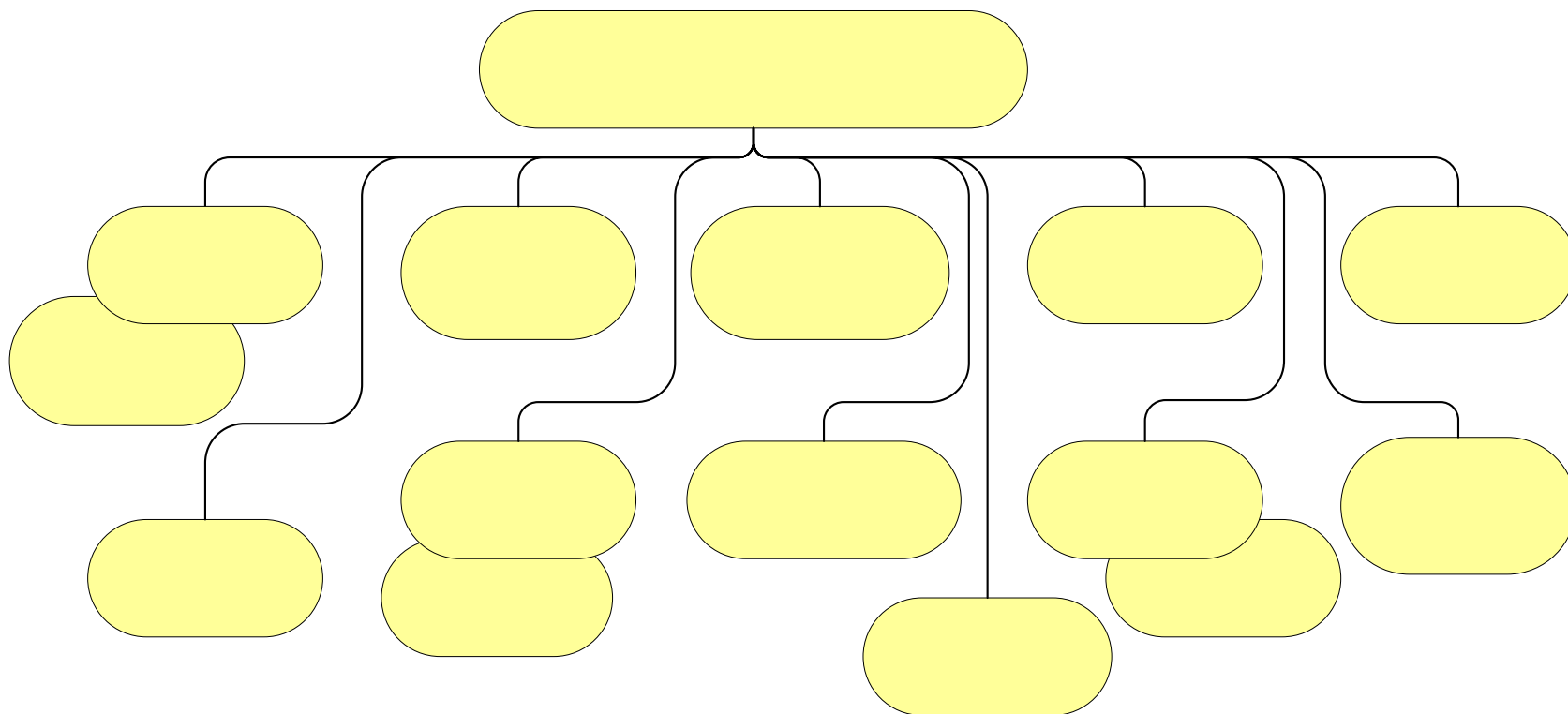
Направления исследований и разработок КЛ

- ❑ Лингвистическая семантика.
 - ❑ Язык и представление знаний.
 - ❑ Коммуникация, диалог и речевой акт.
 - ❑ Язык и Интернет.
 - ❑ Корпусная лингвистика.
 - ❑ Формальные модели языка.
 - ❑ Лингвистика текста.
 - ❑ Компьютерная лексикография.
-

Направления исследований и разработок КЛ

- ❑ Компьютерная лексикография.
 - ❑ Тезаурусы и онтологии.
 - ❑ Речевые технологии.
 - ❑ Системы поиска и классификации.
 - ❑ Компьютерный анализ документов.
 - ❑ Машинный перевод.
 - ❑ Вопросно-ответные системы.
-

Информационные и программные средства лингвистического обеспечения АСОИУ.



Текстовые редакторы, процессоры и настольные издательские системы

- **Текстовый процессор** — вид прикладной компьютерной программы, предназначенной для производства любого вида печатной информации.
- **Настольная издательская система** — комплекс компьютерных аппаратных и программных средств, служащих для предпечатной подготовки оригинал макетов печатной продукции.

Примеры программ:

- Crimson Editor, EditPlus, EmEditor, Notepad, SoloPad, Write, WordPad, Блокнот;
 - Adobe InCopy, Lotus WordPro, Microsoft Word, Apple iWork Pages, WordPerfect;
 - Adobe PageMaker, Adobe InDesign, QuarkXPress, Scribus, Microsoft Publisher, Apple Pages.
-

Программы-конструкторы шрифтов

- **Программы-конструкторы шрифтов** - программы предназначенные, для создания шрифтов.

Примеры:



FontLab



Fontographer



Softy



FontCreator



FontForge
(PfaEdit)

и др.

Системы оптического распознавания (OCR-системы)

- **OCR-системы**
(optical character recognition)
предназначены для ввода печатного текста для печатных и электронных изданий.

Примеры:

- *OCR CuneiForm;*
 - *ABBYY FineReader;*
 - *Recognita Plus;*
 - *TypeReader;*
 - *CharacterEyes;*
 - *IRIS OCR;*
 - *Easy Reader;*
 - *OmniPage Professional*
и др.
-

Системы машинного перевода

- **Системы машинного перевода** – системы, осуществляющие автоматический перевод текстов на одном естественном языке в эквивалентный по содержанию текст на другом языке.

Примеры:

- PROMT
 - Pragma
 - TRADOS
 - DEJA VU
- и др.
-

Электронные словари

□ Лингвистические словари

- описывают языковые единицы: слова, устойчивые словосочетания и др.

□ Энциклопедические словари

- научные или научно-популярные справочные издания, представляющие собой систематизированный свод знаний в каких-либо областях.

Примеры:

- Проект «Русские словари»
<http://www.slovari.ru/>
 - Яндекс-словари
<http://slovari.yandex.ru/>
 - «Кругосвет» – универсальная энциклопедия
<http://www.krugosvet.ru/>
 - Википедия – свободная энциклопедия
<http://ru.wikipedia.org/>
 - Система электронных переводных словарей Lingvo
<http://www.lingvo.ru/>
-

ИПС

- **ИПС (информационно-поисковая система)** - это система, обеспечивающая поиск и отбор необходимых данных в специальной базе с описаниями источников информации (индексе) на основе информационно-поискового языка и соответствующих правил поиска.

Примеры:

- www.mail.ru
 - www.yandex.ru
 - www.rambler.ru
 - www.aport.ru
 - www.google.com
 - www.altavista.com
 - www.excite.com
 - www.yahoo.com
-