АСОТ (Автоматическая системы обработка текста)

- Процессор, на входе и на выходе которого присутствует текстовая информация на естественном языке
- Моделирование различных языковых процессоров (диалоговое взаимодействие, сжатие информации, реферирование текста, логическая обработка содержания, перевод на другой естественный язык и т.д.)
- «Оптимизация общения человека и машины»

Стратегии

Модульный подход

последовательный анализ по уровням (морфологический, синтаксический, семантический, прагматический)

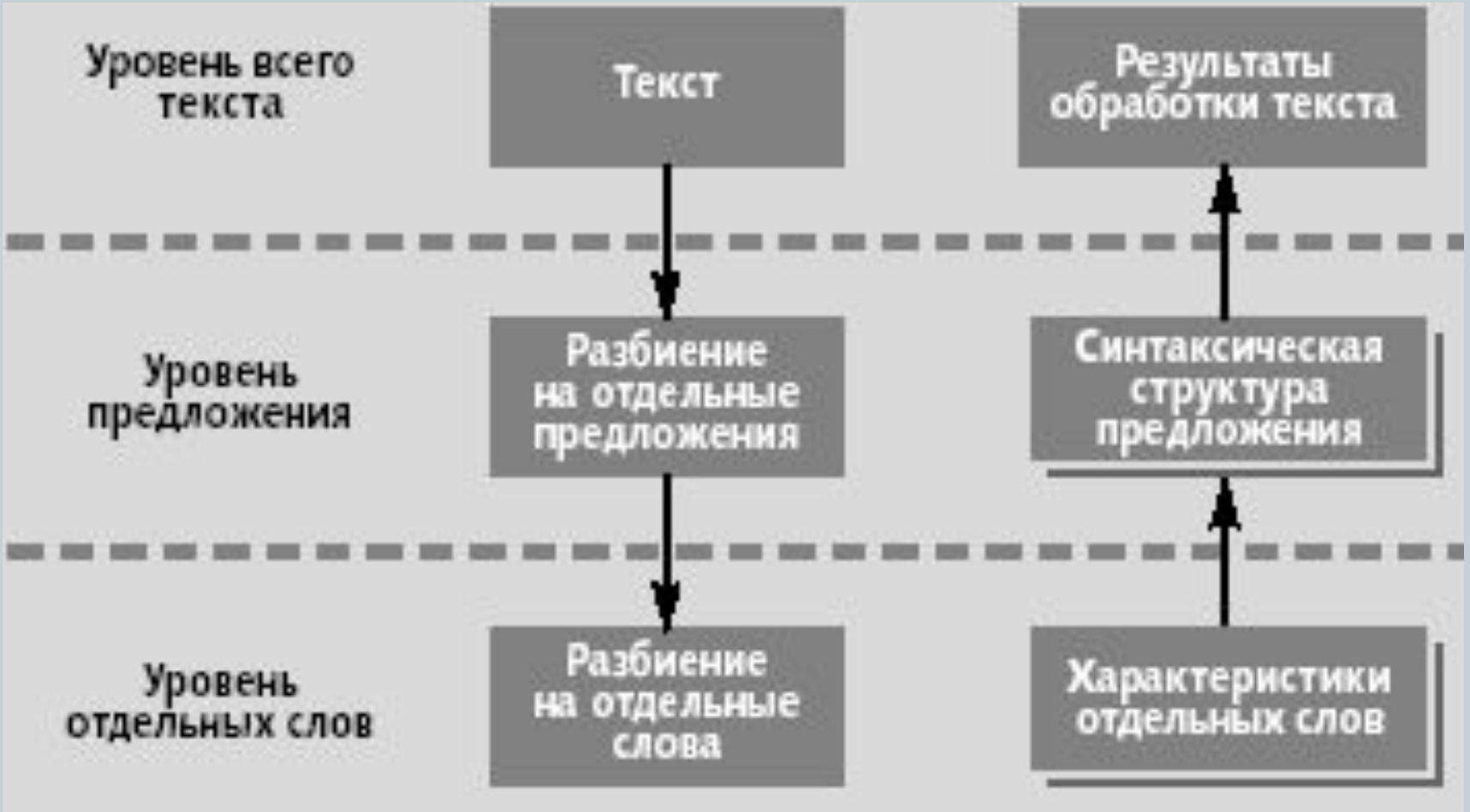
Интегральный подход

Концептуальный анализ

Модульный подход

Модуль морфологического анализа	• Построение морфологической интерпретации слов входного текста
Модуль синтаксического анализа	• Построение дерева зависимостей всего предложения
Модуль семантического анализа	• Построение семантического графа текста

Общая схема обработки текста



Морфологический анализ

- Распознающая роль на входе системы.
- Входной параметр: текстовое представление исходного слова
- Цель и результат: определение морфологических характеристик слова и его основная словоформа.

Рис. 2. Морфологический анализ на основе словаря Зализняка

Синтаксический анализ

- Переход от цепочки лексико-грамматических характеристик, представляющих фразу, к её синтаксической структуре
- Определение взаимосвязи между отдельными словами и частями предложения
- Результат: граф, узлами которого выступают слова предложения

Семантический анализ

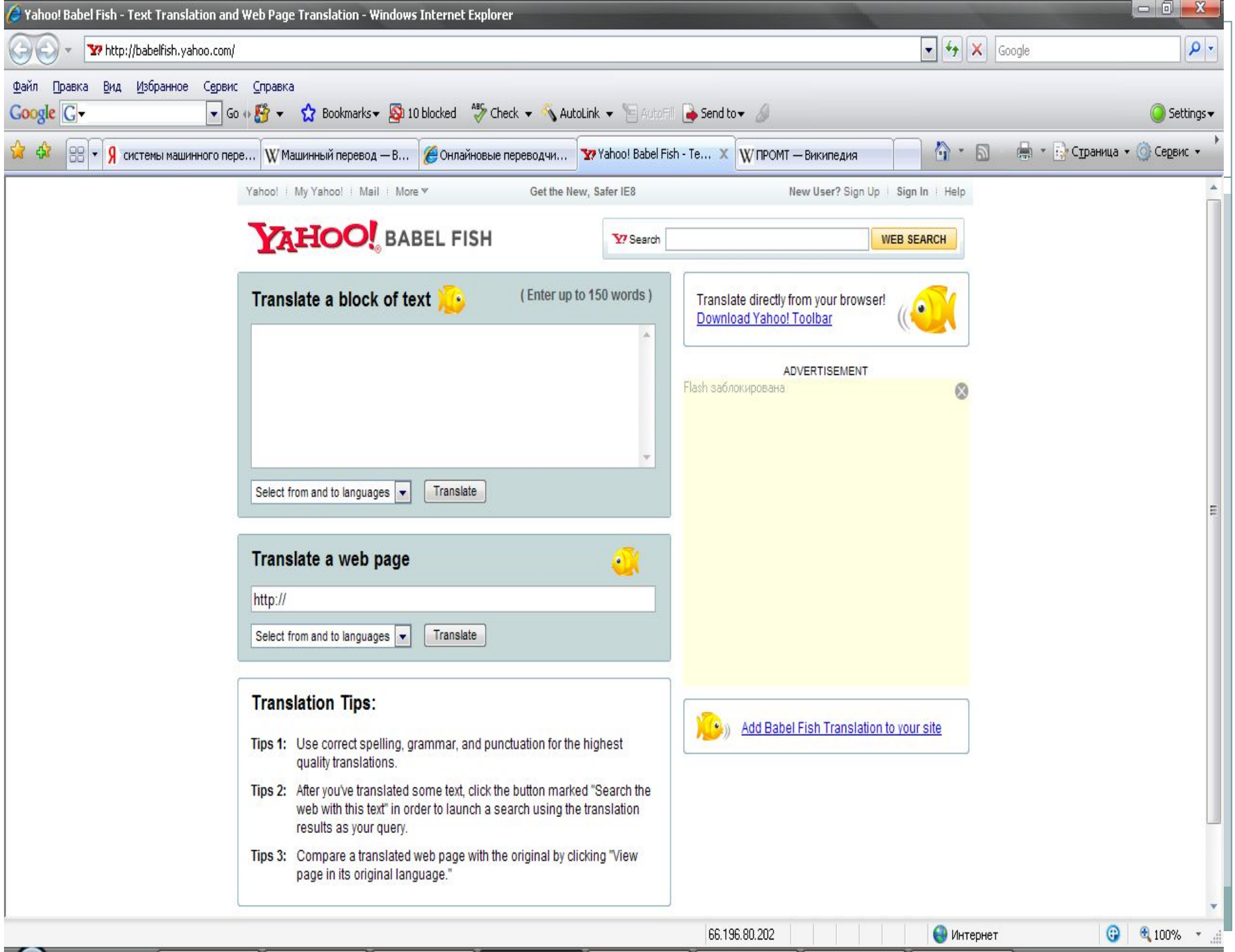
- Поиск фрагментов, формализация, реферирование и т.д.
- Переход от синтаксически проанализированной фразы к её смысловой записи
- Входной параметр: набор деревьев, отражающих синтаксическую структуру каждого предложения
- Основа – тезаурус

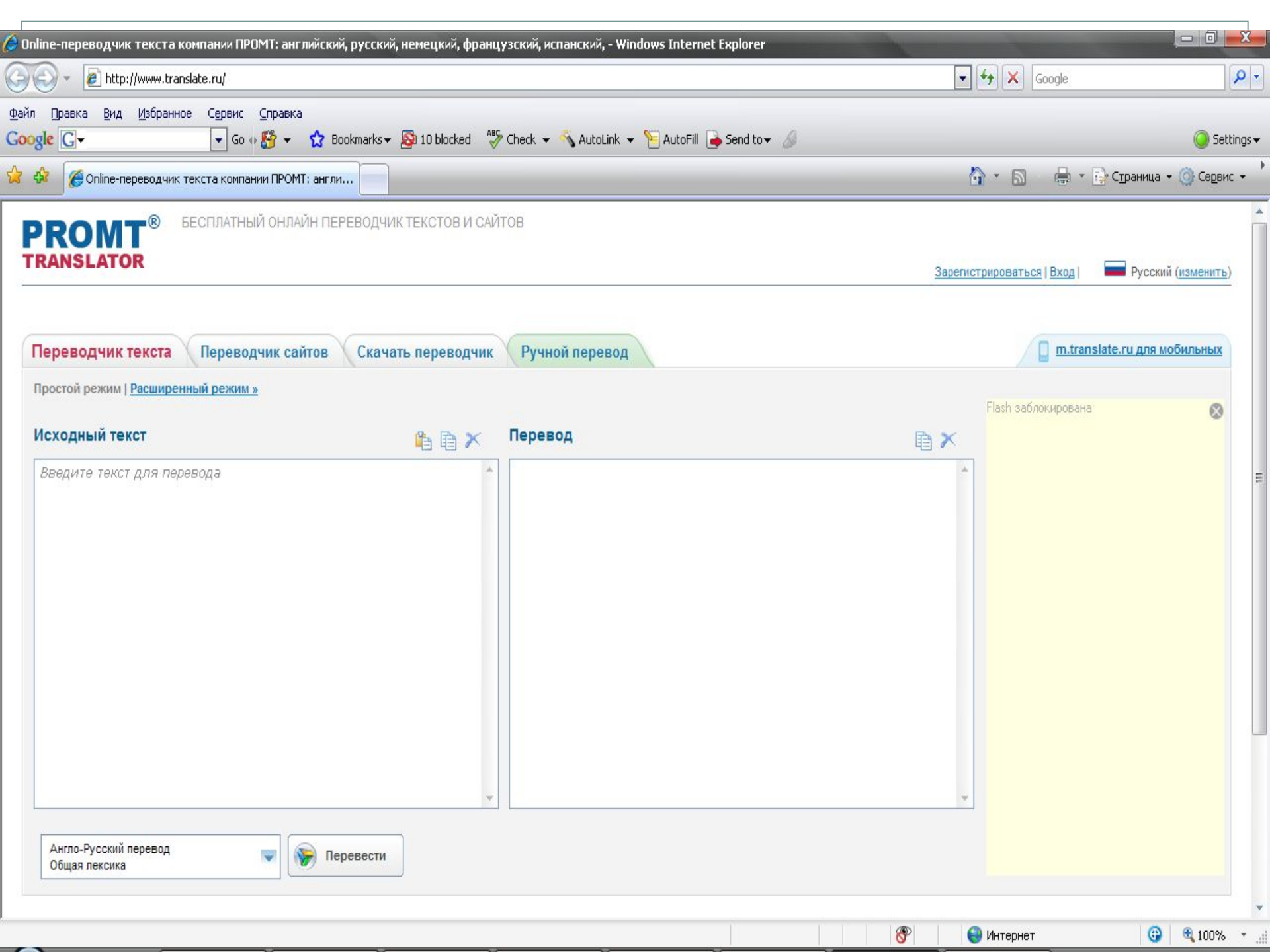
Область реализации

- Системы машинного перевода
  - автоматизированный перевод текста
  - единицы перевода : слова или словосочетания
  - Полнофункциональные коммерческие системы
- Информационно-поисковые системы
  - поиск информации релевантной информационным потребностям пользователя

Системы машинного перевода

- Компания ПРОМТ(www.prompt.ru)
  Текст 500/2000 знаков, web.
- Babel Fish Translation (www.babelfish.altavista.com)
  Текст 150 слов, web. Англ.
- Google Переводчик
- Systran (www.systran.com)
  Текст ~800 знаков, web. Англ.
- PROMT Online Translator [rus/eng]
  (*http://www.translate.ru/*)
- AltaVista [eng]
  (*http://www.world.altavista.com/*)
- TransExp [eng]
  (*http://www.tranexp.com/*)
- Socrat [rus]
  (*http://socrat.ars.ru/cgi-bin/SSISAPI4.0/Socrat.htm*)
- Rustran [rus/eng]
  (*http://www.rustran.com/*)
- ABBY lingvo (http://www.abbyyonline.ru/)


Yahoo! Babel Fish - Text Translation and Web Page Translation - Windows Internet Explorer
http://babelfish.yahoo.com/
Google
Файл Правка Вид Избранное Сервис Справка
Bookmarks 10 blocked Check AutoLink AutoFill Send to Settings
системы машинного пере... Машинный перевод — В... Онлайновые переводчи... Yahoo! Babel Fish - Te... ПРОМТ — Википедия
Страница Сервис
Yahoo! My Yahoo! Mail More
Get the New, Safer IE8
New User? Sign Up Sign In Help
YAHOO! BABEL FISH
Search
WEB SEARCH
Translate a block of text
(Enter up to 150 words)
Select from and to languages
Translate
Translate a web page
http://
Select from and to languages
Translate
Translation Tips:
Tips 1: Use correct spelling, grammar, and punctuation for the highest quality translations.
Tips 2: After you've translated some text, click the button marked "Search the web with this text" in order to launch a search using the translation results as your query.
Tips 3: Compare a translated web page with the original by clicking "View page in its original language."
Translate directly from your browser!
Download Yahoo! Toolbar
ADVERTISEMENT
Flash заблокирована
Add Babel Fish Translation to your site
66.196.80.202
Интернет
100%


Online-переводчик текста компании ПРОМТ: английский, русский, немецкий, французский, испанский, - Windows Internet Explorer
http://www.translate.ru/
Google
Файл Правка Вид Избранное Сервис Справка
Bookmarks 10 blocked Check AutoLink AutoFill Send to Settings
Online-переводчик текста компании ПРОМТ: англи...
Страница Сервис
PROMT® TRANSLATOR
БЕСПЛАТНЫЙ ОНЛАЙН ПЕРЕВОДЧИК ТЕКСТОВ И САЙТОВ
Зарегистрироваться | Вход | Русский (изменить)
Переводчик текста
Переводчик сайтов
Скачать переводчик
Ручной перевод
m.translate.ru для мобильных
Простой режим | Расширенный режим »
Flash заблокирована
Исходный текст
Введите текст для перевода
Перевод
Англо-Русский перевод
Общая лексика
Перевести
Интернет
100%

Информационно-поисковые системы

- Системы, обеспечивающие поиск и отбор необходимых данных в специальной базе с описаниями источников информации (**индексе**) на основе информационно-поискового языка и соответствующих правил поиска.
- Главная задача - поиск информации релевантной информационным потребностям пользователя.
  - Каталоги
  - Поисковые машины
  - Метапоисковые машины

Каталоги

Адреса популярных каталогов:

Зарубежные каталоги:

- Yahoo - **www.yahoo.com**
  Magellan - **www.mckinley.com**

Российские каталоги:

- @Rus - **www.aport.ru**
  Weblist - **www.weblist.ru**
  Улитка - **www.ulitka.ru**

Поисковые машины

Наиболее популярные поисковые машины за рубежом и в России.

Зарубежные поисковые машины:

- Google - www.google.com
  Altavista - www.altavista.com
  Excite - www.excite.com
  HotBot - www.hotbot.com
  Nothern Light - www.northernlight.com
  Go (Infoseek) - www.go.com (infoseek.com)
  Fast - www.alltheweb.com

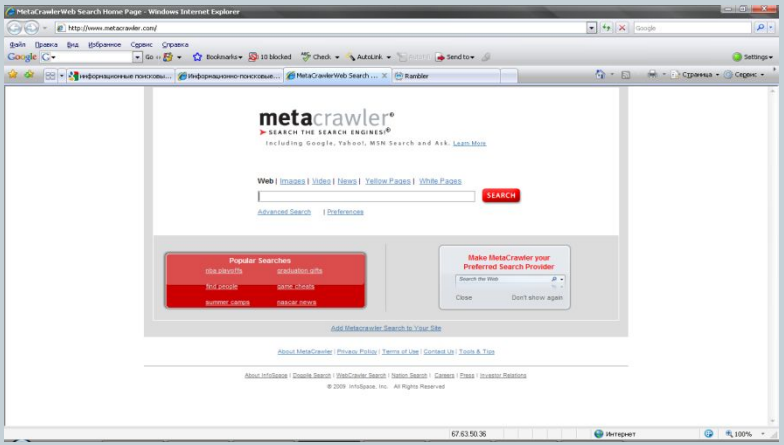
Российские поисковые машины:

- Яndex - www.yandex.ru (или www.ya.ru)
  Рэмблер - www.rambler.ru
  Апорт - www.aport.ru

Метапоисковые системы

Адреса известных метапоисковых систем:

- MetaCrawler - www.metacrawler.com
- SavvySearch - www.savvysearch.com



alltheweb
find it all

Системы интегрального типа

«БОЛЕЕ СОВРЕМЕННЫЙ И БОЛЕЕ АДЕКВАТНЫЙ» Р. ШЕНК

В европейских странах идея интегральной модели появилась в 60-х годах XX в. в связи с созданием систем автоматического перевода.

фрагментарные концептуальные представления:

- морф.анализ
- синт.анализ
- семант. анализ
- сценарии, фреймы, планы.

Концепция Р.Шенка (R.Schank)

- Задача вычислительной семантики – определение процедуры, шаг за шагом сопоставляющей входные предложениям с их смыслом, а также порождающей осмысленные идеи с их воплощением в предложения.

- Основной вопрос – создание представления смысла.

Важны следующие положения:

- 1. Представление смысла не зависит от конкретного языка: «машинным программам, которые могли бы «думать», необходимо оперировать со структурами языка мыслей. Мы надеялись, что такими структурами могли бы представляться передаваемые языком значения».
- 2. Формулируемые процедуры в максимальной степени соответствуют человеческому поведению.
- Эти положения реализованы Р.Шенком и его сотрудников в рамках концепции скриптов.

Система:

- Ищет в тексте диагностические слова
- заполняет пустые слоты в сценарии
- делает ряд концептуальных выводов о смысле текста (в результате чего способна отвечать на поставленные вопросы по содержанию)
- на определенных этапах подключает процедуры
- нельзя получить уровневое представление
- тексты узко ограниченной тематики

Пример: интегральная система анализа Шенка:

1. MARGE (Memory Response Generation in English) - обработка концептуальной информации.

В основе лежит теория концептуальных зависимостей - комплексная теория человеческого мышления.

Работает в двух режимах:

- перефразирование (перевод входной фразы на ЯКЗ)
- концептуальный вывод

2. Модель SAM (Script Applying Mechanism) - компьютерная программа, позволяющая понимать связность текста за счет применения сценариев:

- POLITICS (ведет диалог, моделирует политическую идеологию)
- PAM -> TALE-SPIN - порождение сказок
- FRUMP - машинное реферирование сообщений на нескольких языках, чтение , опирающееся на понятие интереса (Integral Partial Parser)