

Исследование и разработка методов реализации вопросно-ответных систем.

Агаев Нурлан, гр.428

Цель работы:

исследование методов реализации современных
вопросно-ответных систем и разработка

построение простой вопросно-ответной системы
для русского языка.

Задача

- Исследование предметной области вопросно-ответного поиска.
- Рассмотрение существующих решений
- Разработка модель простой вопросно-ответной системы для русского языка.
- Реализация разработанной модели.
- Частичная оценка системы.

Актуальность

- Большие массивы информации
- Проблема поиска информации
- Представление пользователю краткого ответа на запрос.
- Отличия от систем информационного поиска.

Пример:

Запрос: Где родился Пушкин?

Результат: Москва

Вопросно-ответный поиск

Выявлена типичная архитектура вопросно-ответной системы:

- 1) Модуль анализа вопроса
- 2) Модуль информационного поиска
- 3) Модуль обработки ответа

Анализ вопроса

- 1) определение типа вопроса;
- 2) определение типа ожидаемого ответа;
- 3) определение фокуса вопроса;
- 4) определение ключевых слов для формирования поискового запроса.

Информационный поиск

IR-система выдаёт список документов.

Задача:

- Извлечь фрагменты.
- Выбрать фрагменты, потенциально содержащие ответ.
- Отранжировать фрагменты.

Обработка ответа

Получение ответа из фрагмента

Использование информации о типе ожидаемого ответа

Создание списка кандидатов, оценка, выбор

Алгоритмы:

– Основанные на шаблонах

Шаблон	Вопрос	Ответ
<AP> such as <QP>	What is autism?	“, <u>developmental disorders</u> such as autism”
<QP>, a <AP>	What is a caldera?	“the Long Valley caldera, <u>a volcanic crater</u> 19 miles long”

– Сбор ответа из N-грамм

Вопросно-ответная система

Главная задача разработки состояла в поисках решения для русского языка в следующих модулях:

- 1)Модуль анализа вопроса
- 3)Модуль обработки ответа

Разработанная модель системы

1. Анализ вопроса

Типы вопросов:

- поиск определения
- библиографического характера
- места
- времени

Определение типа по вопросительным словам и шаблонам.

Информационный поиск

Поисковая системы Google.
Фрагменты документов — это сниппеты.

The image shows a screenshot of a Google search interface. At the top, the Google logo is on the left, and a search bar contains the text "что такое бампер?". To the right of the search bar is a keyboard icon and a blue search button with a magnifying glass. Below the search bar, the word "Поиск" is displayed in red, followed by the text "Результатов: примерно 880 000 (0,22 сек.)".

The main content area is divided into two columns. The left column contains filters: "Все результаты", "Картинки", "Карты", "Видео", "Новости", "Ещё", "Москва", "Изменить место", "Весь Интернет", "Только на русском", "Перевод результатов", and "Показать настройки".

The right column displays search results. The first result is titled "Что означает бампер:" and includes a snippet: "Определения в Бампер — энергопоглощающее устройство автомобиля (на случай Интернета: лёгкого...". Below this snippet is a green link to "ru.wikipedia.org/wiki/Бампер" and the text "Подробнее... »" and "Перейти на сайт".

The second result is titled "Бампер — Википедия" and includes a snippet: "Бампер — энергопоглощающее устройство автомобиля (на случай лёгкого удара) в виде бруса, расположенного спереди (часто и сзади). Бампер ...". Below this snippet is a blue link to "Стандарты - История - Стандарты в Северной Америке - См. также" and the text "Вы посещали эту страницу 18.05.12.".

The third result is titled "Что такое бампер?" and includes a snippet: "В энциклопедиях бампер (английское «bump» означает «удар, столкновение») - описывается, ... Бампер представляет собой разновидность буфера.".

The fourth result is titled "Что такое бампер" and includes a snippet: "В условиях небольших скоростей бампер выполняет роль ударопоглощающего буфера, который гасит энергию удара и предотвращает деформацию ...".

Извлечение ответов

- Выдается 5 кандидатов
- В вопросах библиографического характера и поиска определений с шаблонами «Кто такой <>?», «Кем был <>?», «Что такое <>?» используется прием извлечения ответа из Википедии.
- Для вопросов общего типа — сравнение деревьев разбора:
 1. Если узлы слова в узлах идентичны, то к оценке начисляется 1 балл
 2. Если совпадают только начальные формы слов в узлах — начисляется 0.5 балла
 3. Если слово одного узла, является частью слова другого узла — начисляется 0.5 балла.

Оценка системы

Для частичной оценки системы создан небольшой корпус вопросов.

Оценивалась работа системы над вопросами классов «Person» и «Definition» разработанной типологии

	Person	Definition
Верный ответ	28	10
Неверный ответ	11	8
Не найден ответ	26	22
Всего	75	40

Результаты

- Исследована предметная область вопросно-ответного поиска
- Рассмотрены основные подходы к решению задачи поиска ответа на вопрос
- Разработана модель простой вопросно-ответной системы
- Реализована модель системы
- Оценена работа системы