

Интеллектуальная поисковая система «Exactus»: Опыт участия в семинаре РОМИП'2005

Российский семинар по
Оценке Методов
Информационного Поиска
РОМИП'2005

Институт Системного Анализа РАН
г. Москва

Отличительные особенности Exactus

- использования лингвистических средств обработки текстов, включая морфологический, синтаксический и семантический анализ;
- работа с естественно-языковыми запросами;
- расширение механизмов стандартного поиска по ключевым словам;
- предварительная обработка и расширение запросов пользователя;
- получение данных из источников различного типа и единый механизм работы с этими данными;
- автоматическое подключение новых информационных источников;
- использование специализированных поисковых профилей;
- использование параллельных вычислений.

Преимущества Exactus

- ключевое слово в запросе осознается не как слово-лексема, а как синтаксема, которая выделяется на основании взаимозависимых морфологических, семантических и функциональных признаков;
- семантический и синтаксический анализ не разделяются на два отдельных этапа, а представляют собой единую процедуру;
- используемый в системе коммуникативно-грамматический подход к языку позволяет извлекать из «доступных» машине данных (морфологическая форма, категориальный класс, позиция в предложении) гораздо больше информации, необходимой для точного поиска, чем при традиционных подходах.

Общая схема обработки поисковых запросов



Предобработка поисковых запросов

Запросы подвергаются специальной процедуре предобработки с использованием словаря:

- синонимов и синонимичных именных групп;
- синонимичных предикатов;
- конверсивов;
- стоп-слов.

Анализ текста

Анализ текста проводится в три этапа:

- морфологический анализ;
 - синтаксический анализ;
 - семантический анализ.
- Особенностью является то, что синтаксический и семантический анализ совмещены в единой процедуре.

Выделение информации из текста

Используется три уровня дополнительной информации:

- синтаксическая зависимость отношений между словами в пределах именных групп;
 - значения синтаксем;
 - семантические связи.
- Отличительной особенностью является применение теоретических положений коммуниктивно-грамматической школы к задачам информационного поиска

Подсчет релевантности

Релевантность вычисляется по:

- значениям синтаксиса;
- семантическим связям;
- именованным группам;
- ключевым словам.

Поисковые профили

Поисковый профиль задается пользователем перед отправкой запроса на поисковую машину:

- поиск объектов
 - поиск точных фактов
 - поиск близких ситуаций
 - поиск только по ключевым словам
- Профили это специализированные настройки лингвистических алгоритмов для достижения максимальной эффективности поиска.

О методике тестирования РОМИП'2005: предложения

Учитывать такие важные вещи:

- Расширение запроса близкими по смыслу словами
- Ситуативный поиск
- Вопрос-ответ
- Осмысленность запроса

Слова благодарности организаторам РОМИП'2005

Команда разработчиков Exactus
выражает свою искреннюю
признательность организаторам
РОМИП'2005!

Спасибо за внимание!

Все вопросы и предложения
принимаются по адресу:

matandra@isa.ru