

Форум \ «Общие вопросы» \ «Задача»

Ожидаемый результат

- ПО, умеющее отвечать на запросы о спортивных событиях
- Запросы — строка текста
- Ответы — XML

Запросы

- Матчи Зенита в августе 2009
- Матчи Спартака, закончившиеся ничьёй
- Матчи Спартак-ЦСКА в Петербурге
- Победы сине-бело-голубых
- Хоккей в Москве
- ...

Ответы

- XML, UTF-8
- Список событий, удовлетворяющих условию запроса, включая вложенные события
- Один запрос — один XML файл

Иерархия событий

- Событие — чемпионат
 - Событие — матч
 - *Событие — факт в рамках матча*
 - *Игрок забил гол*
 - *Игрок получил красную карточку*
 - *Замена игрока*
 - ...

Характеристики события

- Идентификатор
- Тип
- Название
- Дата / время
- Место
- Команды-участники
- Результат
- Вид спорта
- Идентификатор события-родителя

Формальная спецификация XML

- Будет ...

Исходные данные

- Тексты новостей
 - Тренировочный корпус (сегодня)
 - Тестовый корпус (9 июля)
- Дампы Википедии

Оценка результатов

- Начало - 10 июля вечером
- Окончание - ?
- Тестовый корпус — у всех команд один
- Тестовые запросы — у каждой команды свои
- Количество запросов — 20-50 (?)
- Время на получение результатов — 10 минут
- Способ оценки — вручную

Вопросы?

Как жить?

«В прикладной лингвистике ситуация фрактальная - каждая маленькая проблемка на самом деле размером со все остальные - то есть бесконечна. Это касается выделения слов, определения конца предложения, транслитерации, исправления опечаток - даже таких мелочей. Если делать по-хорошему, уйдёт вся жизнь.»

Ашманов, 2001 г

Выводы

- **Чем проще, тем лучше**
- Разделить на подзадачи
- Определить интерфейсы
- Собрать решение целиком, соединив «пустые» модули
- Составить свои тестовые запросы и ответы
- Постепенно наполнять решение содержанием