RuSSIR
Russian Summer School
in Information Retrieval

# RuSSIR 2008

## Russian Summer School in Information Retrieval

1-5 сентября 2008. Таганрог

Как это было…

# Немного истории

RuSSIR
Russian Summer School in Information Retrieval

- RuSSIR 2007 был первым и проходил в Екатеринбурге 5-12 сентября 2007

- Идея проведения RuSSIR'ов принадлежит Павлу Браславскому
- Школы проводятся совместно с РОМИП
- Школы поддерживаются и спонсируются разными компаниями

# Немного статистики

RuSSIR
Russian Summer School
in Information Retrieval

- Таганрогский технологический институт Южного федерального университета
- 4 полных курса, 1 краткий и 1 лекция
- Конференция молодых ученых (7 докладов)
- 9 иностранных слушателей
- Все курсы по-английски
- 10 часов занятий в день (с перерывами)
- Всего 106 человек

# Курс

RuSSIR
Russian Summer School
in Information Retrieval

- **Text Mining, Information and Fact Extraction**

- Marie-Francine Moens (Katholieke Universiteit Leuven), Belgium

# Text Mining, Information and Fact Extraction

RuSSIR
Russian Summer School in Information Retrieval

В курсе делается широкий обзор методов машинного обучения и их применения к задаче информации из текстовых данных.

- Цель: извлечение конкретных фактов из текста на естественном языке
- Лексическая обработка: извлечение языковых свойств, токенизация, стемминг, POS-разметка, парсинг предложений

# Text Mining, Information and Fact Extraction

RuSSIR
Russian Summer School in Information Retrieval

- Методы классификации: SVM, Байес, принцип максимальной энтропии
- Контекстно-зависимая классификация: Hidden Markov Models, Conditional Random Field, Probabilistic Latent Semantic Analisys, Latent Dirichlet Allocation
- Приложения

# Курс

RuSSIR
Russian Summer School
in Information Retrieval

- **Поиск изображений по содержанию**

- Наталья Васильева
  (HP Labs)
  Санкт-Петербург, Россия

# Поиск изображений по содержанию

RuSSIR
Russian Summer School in Information Retrieval

1) Задачи Image Retrieval: поиск изображений, похожих на заданный пример, поиск по заданной цветовой гамме, примерной форме и т.д.

2) Проблемы Image Retrieval:

- большой разрыв между представлением и семантикой(а интересно именно семантическое содержание)
- субъективность восприятия изображений
- трудность визуализации

# Поиск изображений по содержанию

RuSSIR
Russian Summer School in Information Retrieval

3) Уровни свойств изображения

- Цвет(цветовые пространства, гистограммы)
- Текстура(статистические свойства, фильтры, вэйвлеты)
- Форма(методы выделения границы, кодирование формы)
- Семантический(применение fusion-методов)

# Поиск изображений по содержанию

RuSSIR
Russian Summer School in Information Retrieval

4) Сегментация

5) Многомерное индексирование

- Деревья(R, Quad, VP и тд)
- Locality Sensitive Hashing

6) Приложения

- IBM QBICK - search by image content
- Virage
- Visual SEEk

# Курс

RuSSIR
Russian Summer School
in Information Retrieval

- **Data Structures in IR**

- **Максим Губин (Ask.com, США)**

# Data Structures in IR

RuSSIR
Russian Summer School
in Information Retrieval

Последовательно излагается процесс проектирования поисковой системы. На примерах возникающих проблем показывается применение классических структур данных и алгоритмов: Huffman and LZ coding, Map-Reduce, Bloom filter

- Индексация, структуры хранения данных, методы сжатия, распараллеливание
- Поиск, слияние списков, отсечение, кэширование, построение кластера
- Языковая модель, PageRank

# Курс

RuSSIR
Russian Summer School
in Information Retrieval

- Hands-on Natural Language Processing for Information Access Applications

- **Horacio Saggion (University of Sheffield)**

# Natural Language Processing for Information Access

RuSSIR
Russian Summer School in Information Retrieval

- Извлечение информации из текста
  - выявление именованных сущностей
  - поиск связей между сущностями
  - определение значения сущностей

- Обзор технологий и средств анализа естественного языка на примере системы GATE

# Natural Language Processing for Information Access

RuSSIR
Russian Summer School in Information Retrieval

- Question Answering
  - получение ответа на запрос, сформулированный на естественном языке
  - 3 типа вопросов: факт, перечисление, определение
  - индексация, анализ запроса, получение ответа
- Text Summarization
  - выжимка из текста смысловой сути
  - sentence summarization, article summarization, corpus summarization
  - методы оценки результатов

# Короткий курс

RuSSIR
Russian Summer School in Information Retrieval

- **IR in Social Media**
- Alexey Maykov, Microsoft LiveLabs

- Чем отличается Social Media от обычных СМИ?
- Обзор, применение и архитектура SM
- Сбор данных в SM, различные методы
- Обработка и анализ этих данных

# Лекция. **Темы дня в блогах: Как это работает**

RuSSIR
Russian Summer School in Information Retrieval


- **Антон Волнухин**
- **Андрей Мищенко**

Яндекс


- Что такое «темы дня» в яндекс.блогах?
- Как формируются «темы дня»?
- Особенности формирования

# Конференция молодых ученых

RuSSIR
Russian Summer School in Information Retrieval

**Константин Артемьев**
Метод вероятностного морфологического анализа для задач полнотекстового индексированного поиска

**Александр Сибиряков**
Извлечение мнений о товарах из форумов и блогов с учетом тональности

**Евгений Рабчевский**
Применение лексико-синтаксических шаблонов для автоматизации процесса построения онтологий

**Ольга Пустыльникова**
Автоматическая классификация текстов на основе их структурных признаков. Какую информацию о тексте отражает структура?

**Алексей Владыкин**
Автоматический метод оценки тематической содержательности документов

**Мстислав Масленников**
Самозагрузка правил для извлечения информации из текстов на естественном языке.

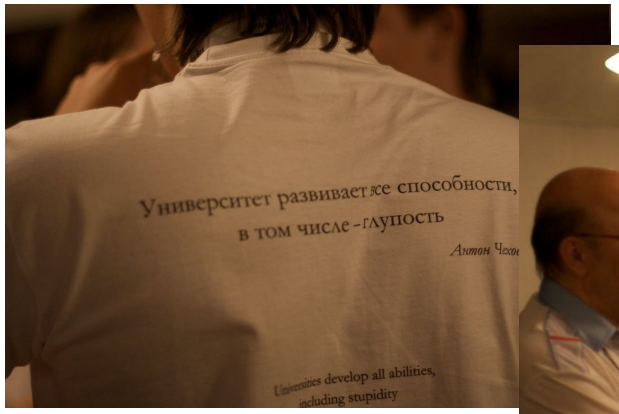
**Ольга Шамина**
Автоматический поиск научных статей в сети Интернет

# Конференция молодых ученых

RuSSIR
Russian Summer School
in Information Retrieval

# Неформальное Welcome Party

RuSSIR
Russian Summer School in Information Retrieval

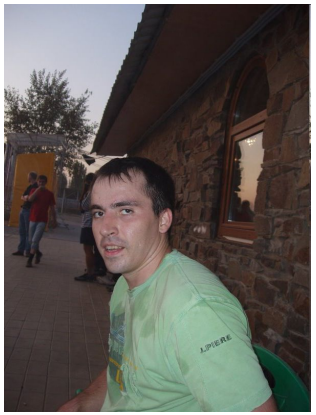



# Неформальное Футбольный матч

RuSSIR
Russian Summer School in Information Retrieval

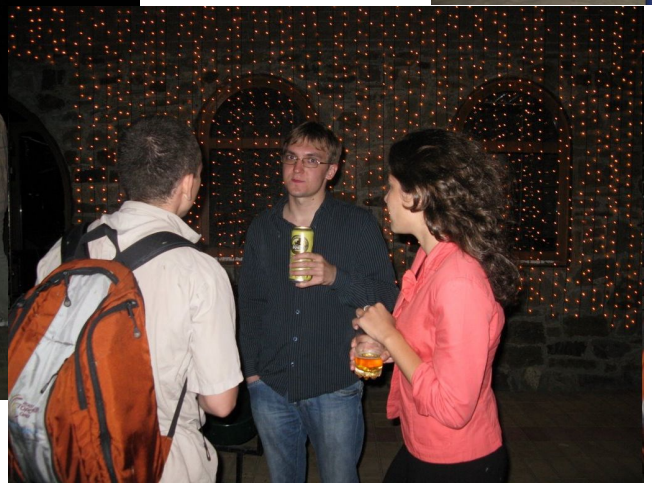
**12 – 0**

**в пользу студентов!**




# Неформальное RuSSIR Party

RuSSIR
Russian Summer School in Information Retrieval

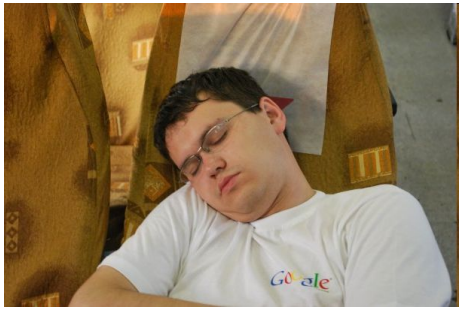



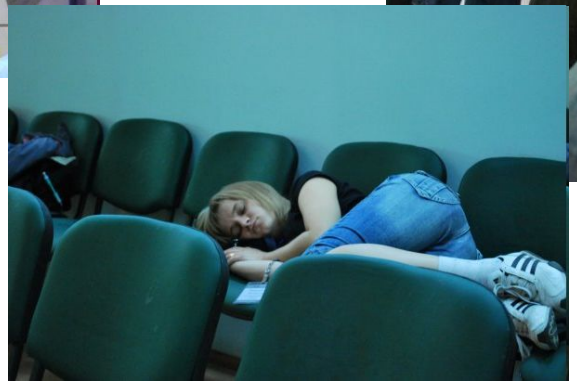
# Неформальное
# Мафия по ночам

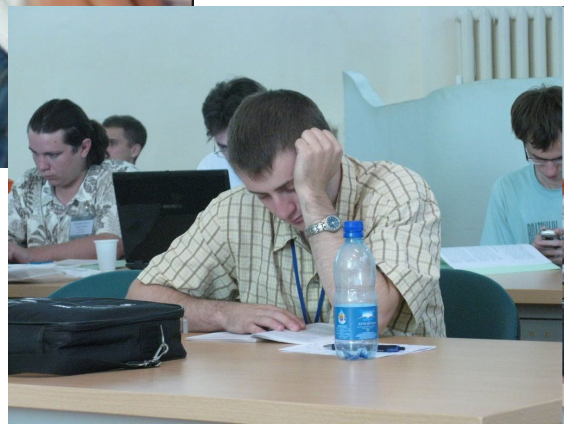
RuSSIR
Russian Summer School
in Information Retrieval

# Неформальное Последствия мафии

RuSSIR
Russian Summer School in Information Retrieval





# А на самом деле:

RuSSIR
Russian Summer School
in Information Retrieval


# Материалы

RuSSIR
Russian Summer School
in Information Retrieval

- Видеозаписи лекций скоро появятся на сайте http://videolectures.net/
- Презентации вы можете почитать уже сейчас http://romip.ru/russir2008/program.html