

Информационный поиск

ИТМО

Октябрь – декабрь 2014

Павел Браславский

при поддержке фонда «Династия»

About myself

2000: PhD thesis on genres and search

2001 – present: Ural Federal University

2005-2012: Yandex

Feb 2012: Kontur labs



ROMIP, Internet Math, RuSSIR, IIR book, YARN, Velpas

IR&NLP: query log mining, automatic summarization & snippet generation, evaluation, CQA, thesauri

<http://kansas.ru/pb>

Состав курса

- Лекции – 4 блока * 4 лекции

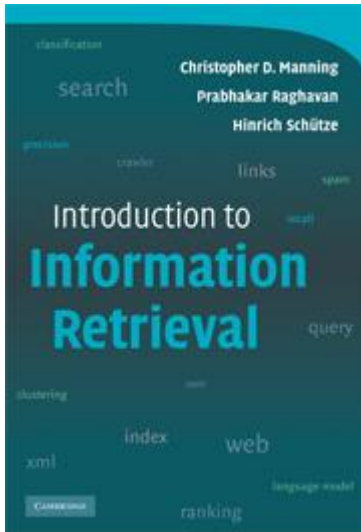
2,3 окт; 30, 31 окт; 20-21 ноя; 18-19 дек (или 25-26 дек)

- Практические задания (!)
- Семинары
- Консультации
- Экзамен

Информация

- Страница курса http://kansas.ru/ir_spb
- Рассылки <http://bit.ly/iir2014> (iir_usu),
year2010@rain.ifmo.ru
- Моя почта pbras@yandex.ru, subj: [ir_spb]

Литература



Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press. 2008.

<http://informationretrieval.org/>



Кристофер Д. Маннинг, Прабхакар Рагхаван, Хайнрих Шютце
Введение в информационный поиск

«Миссия курса»

- Теория + практические задачи
- Основы + примеры конкретных задач

План на 2 и 3 октября

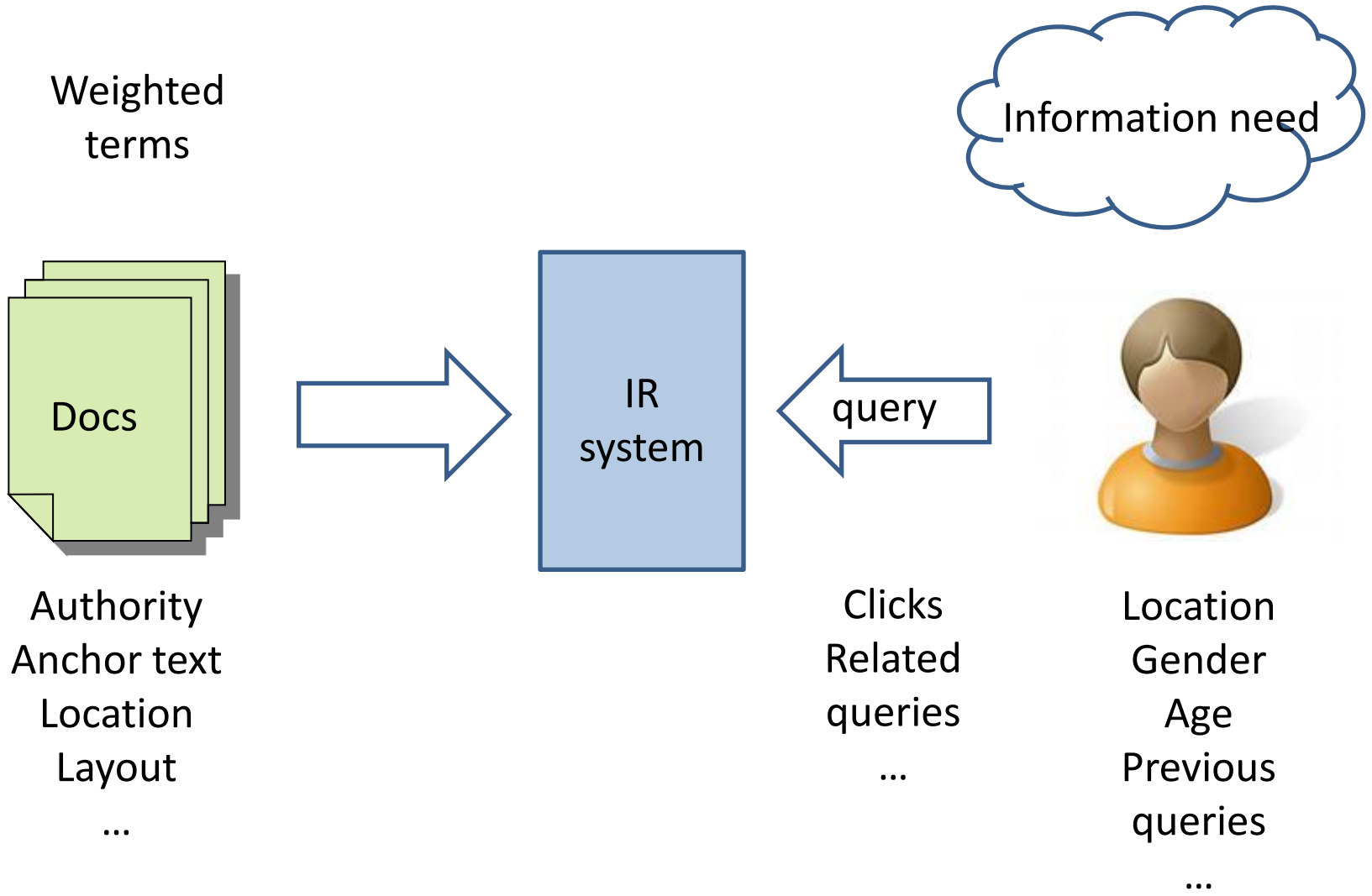
- Что такое информационный поиск?
- Инвертированный индекс. Булев поиск.
- Построение индекса. Распределенное построение индекса. Токенизация.
- Лингвистическая обработка. Русская морфология для информационного поиска.
- Распределение слов в коллекциях и документах.

Информационный поиск

- Уникальное сочетание теоретических и эмпирических подходов
- По-настоящему междисциплинарная область
- Впечатляющий технический прогресс и экономическая эффективность

Не только инвертированные индексы

- Мультимедиа
- Социальные взаимодействия
- Лингвистика
- Машинный перевод
- Психология
- Экономика
- Картографические сервисы



Not only query-response model (!)