

Московский Физико-Технический Институт

ДИПЛОМНАЯ РАБОТА:
**«Методы выявления взаимосогласованных
структур сходства в системах
взаимодействующих объектов»**

Выполнил студент 175 группы:

Лексин В.А.

Научные руководители:

член-корреспондент РАН, д.ф.-м.н. Рудаков К.В.

н.с., к.ф.-м.н. Воронцов К.В.



Постановка задачи

- Разработать и реализовать специальный вариант технологии АКС, приспособленный для решения задач персонализации услуг и поиска ресурсов в Интернет
- Выявить и визуализировать кластерные структуры ресурсов и пользователей.



Исходные данные

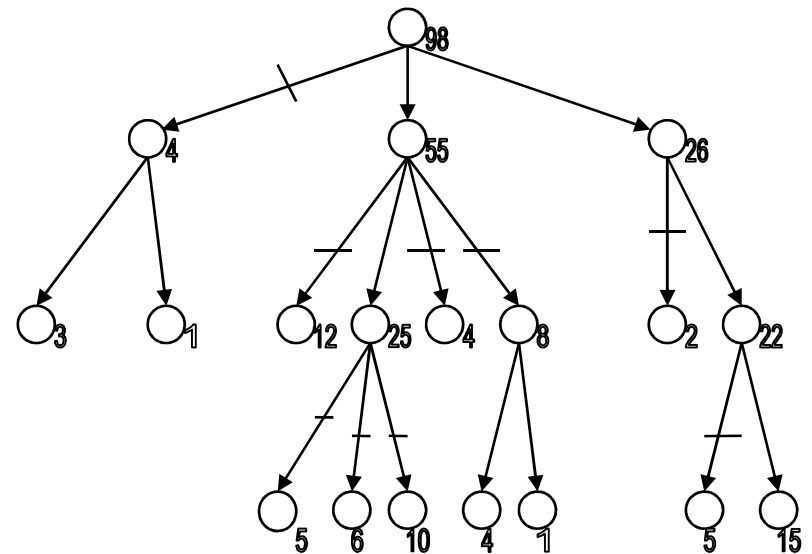
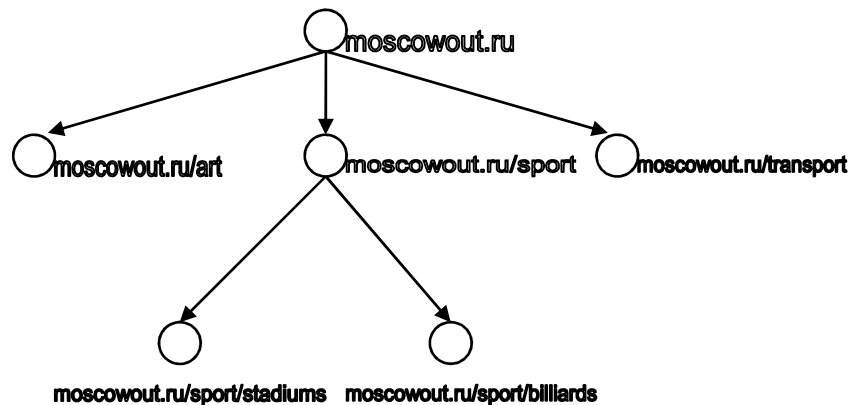
- Объем лога 4Гб
- 14 606 пользователей
- 1 972 636 ресурсов (из них 129 600 были выбраны)
- Интервал времени: 1 неделя работы поисковой системы Яндекс



Общая схема обработки данных

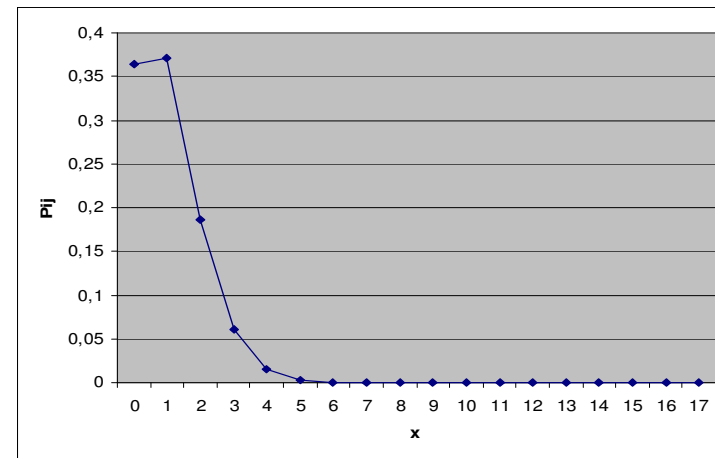
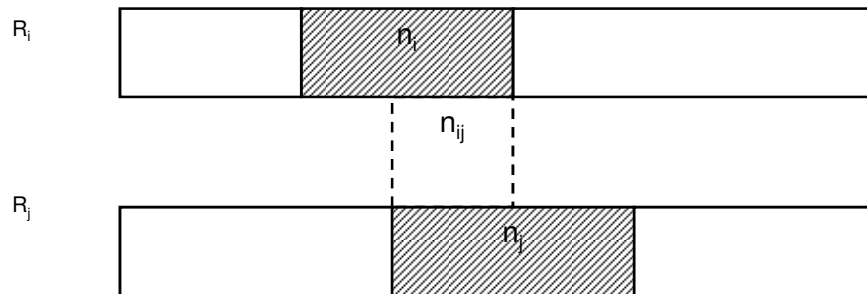
1. Формирование полных словарей пользователей и ресурсов
2. Редукция словарей
3. Формирование матрицы посещений
4. Формирование разреженной матрицы сходства ресурсов
5. Построение карты сходства всех ресурсов (карта Интернета)
6. Построение локальной карты сходства в окрестности заданного ресурса.

Формирование и редукция словарей пользователей и ресурсов



Пример построения и редукции словаря ресурсов с порогом 13 посещений

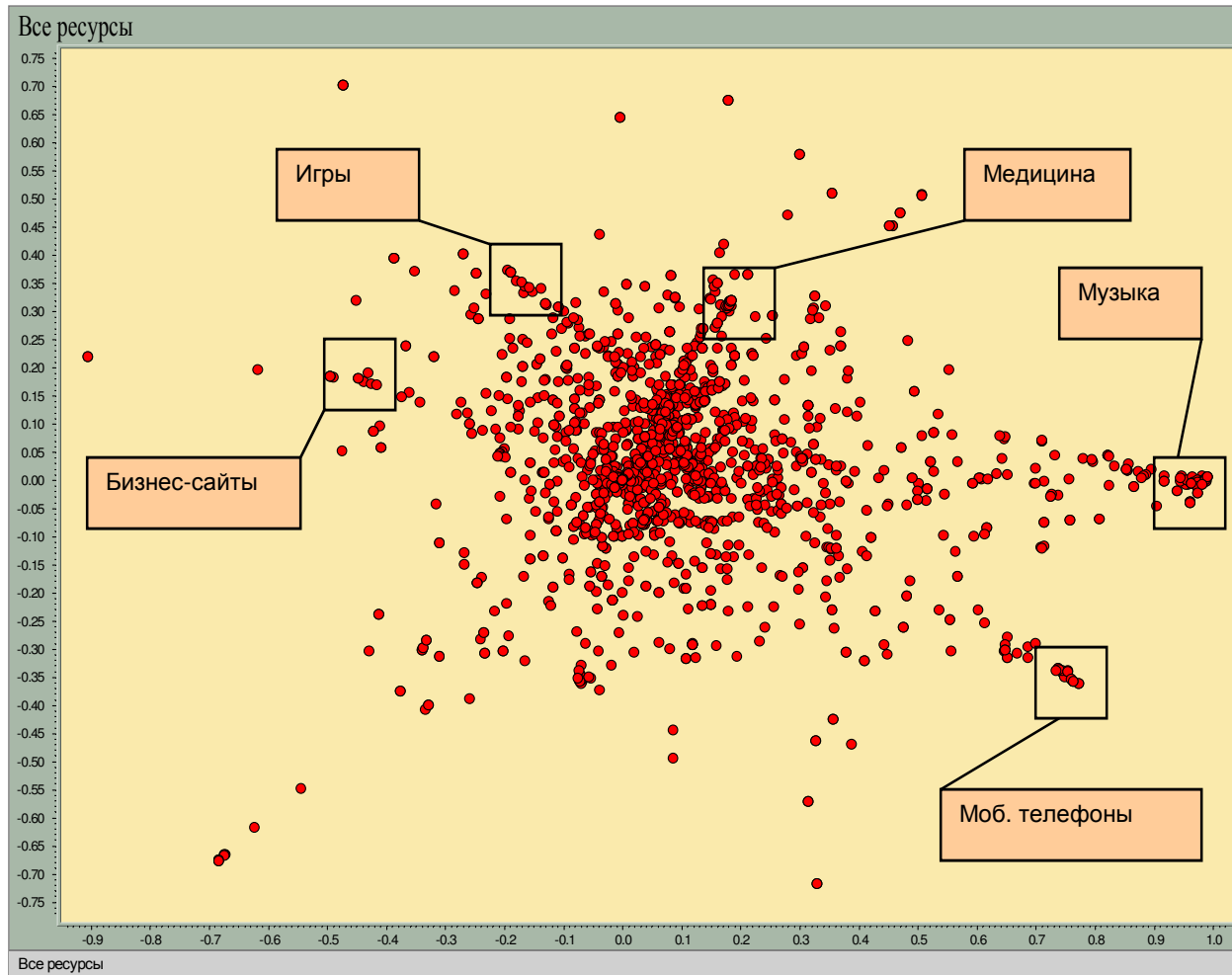
Вычисление оценок сходства ресурсов



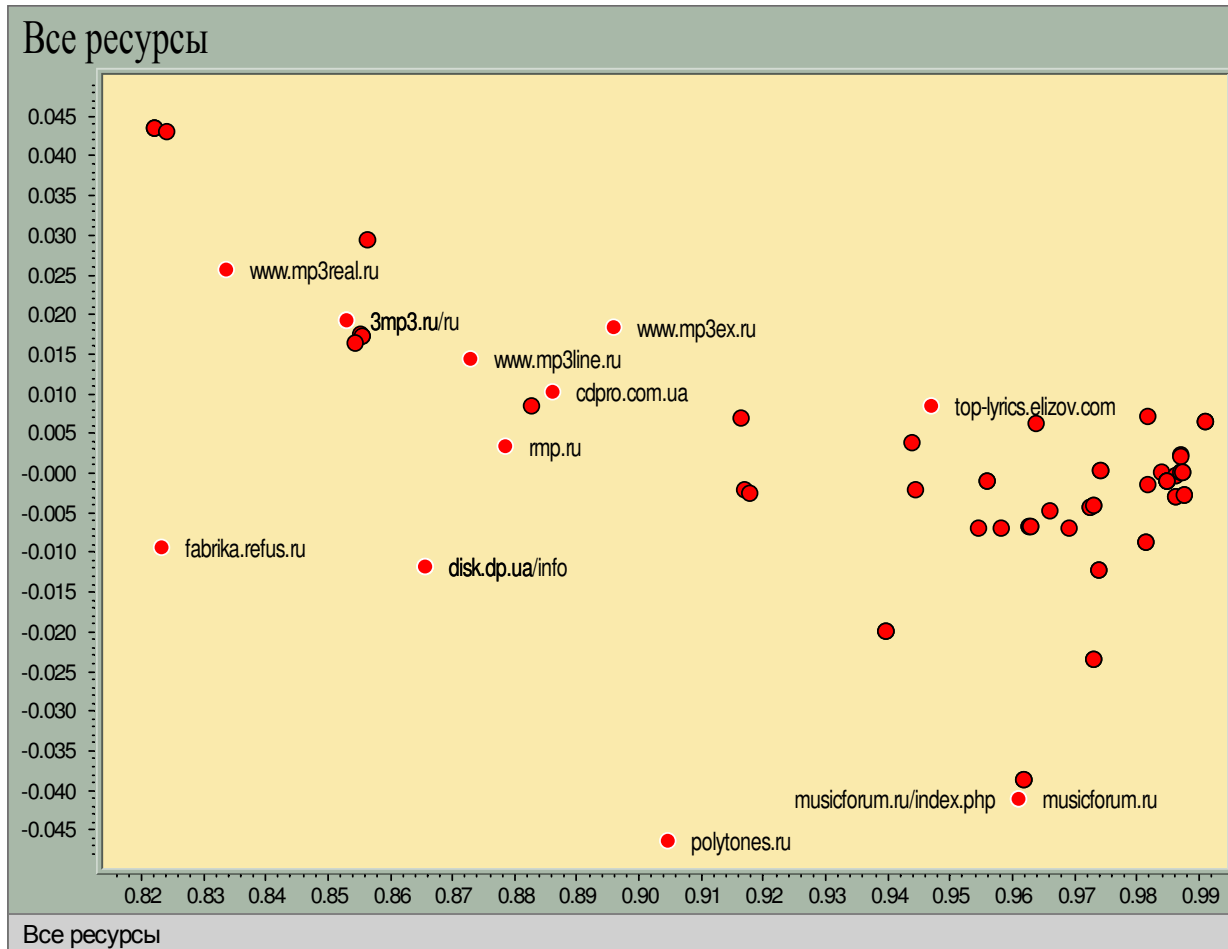
$$P_{ij} = P(n_{ij} = x) = \frac{C_{n_i}^x C_{U-n_i}^{n_j-x}}{C_U^{n_j}} \quad \rho(i, j) = \left(\frac{|\ln \alpha|}{|\ln P_{ij}|} \right)^3$$



Карта сходства ресурсов



Фрагмент карты сходства



- my3blka.narod.ru
- www.mp3zzz.ru
- mp3load.ru
- mp3-unique.liveinternet.ru
- mp3forum.ru
- www.mp3collection.ru
- mp3rank.ru
- www.mp3real.ru
- top-lyrics.elizov.com
- rmp.ru
- www.mp3ex.ru
- mp3.allabout.ru
- www.mp3line.ru
- www.filesale.ru
- www.oxid.ru
- www.agharta.ru
- 3mp3.ru
- moremusic.com.ru
- www.miditext.ru
- ru.mp3s-download.com
- www.rusmuz.ru
- meloman.tehnofil.ru
- www.mp3search.ru
- www.mp3tones.ru
- thesearch.ru
- www.muзzone.com
- www.mp3.myfind.ru
- www.mp3zone.ru
- www.russianmusic.de
- www.zaycev.net
- www.mp3host.ru
- disk.dp.ua
- www.mp3up.ru
- www.uliss.ru
- mp3.volpi.ru
- musicforum.ru
- cdpro.com.ua



Основные результаты

- Разработаны и реализованы эффективные алгоритмы анализа логов поисковой машины
- Построена карта сходства ресурсов Интернет
- Предложена методика построения локальных карт сходства