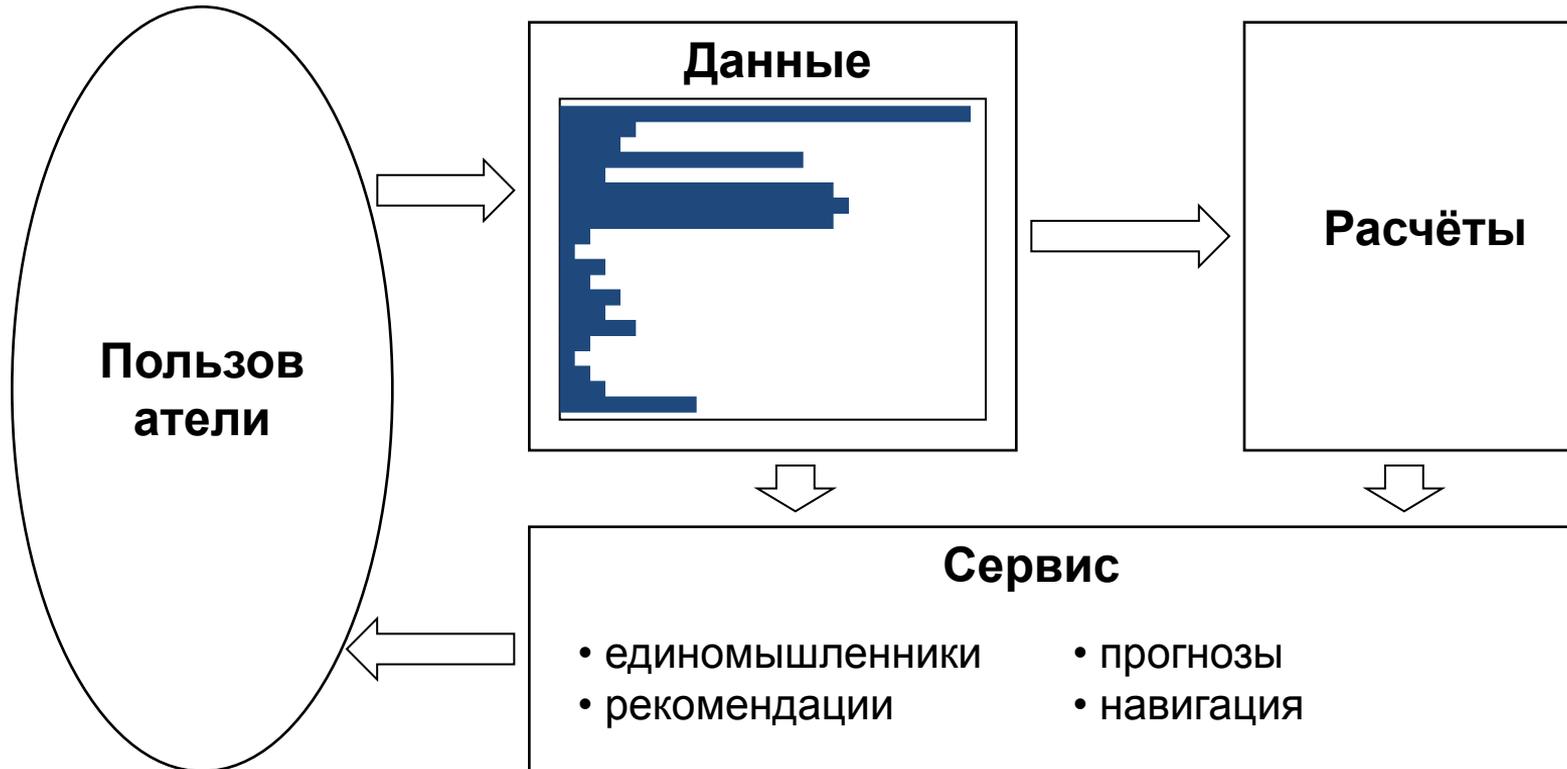


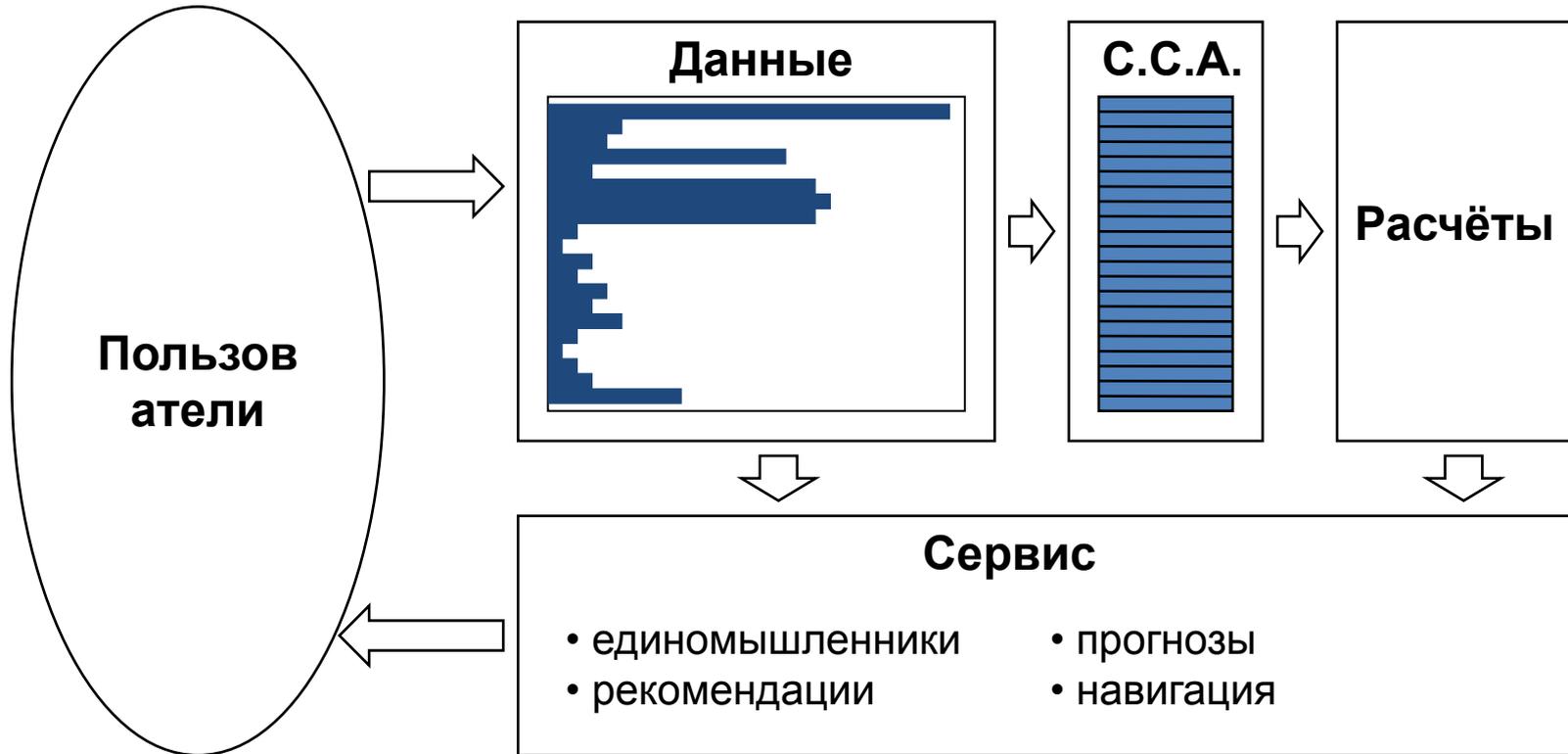


Российские  
интернет-технологии

2011

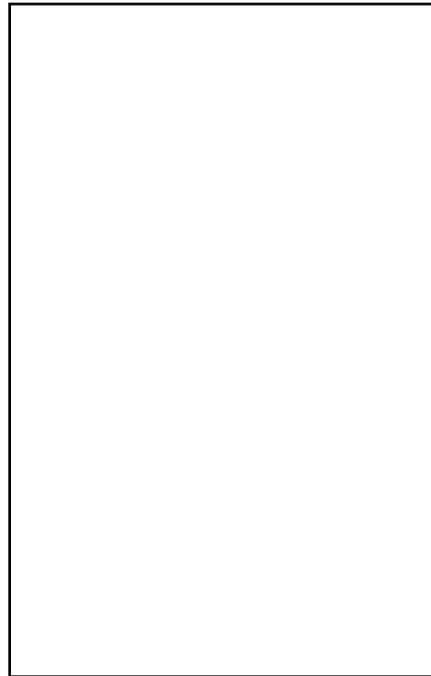
# Факторный анализ в рекомендательной системе imhonet



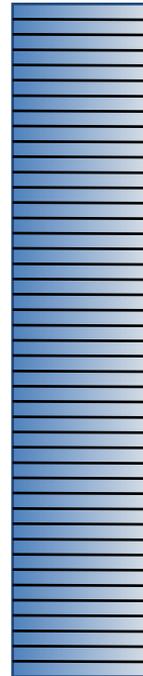


Большая разреженная матрица оценок  $m \times n$   
( $m$  пользователей,  $n$  объектов)

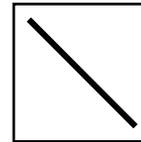
Факторы пользователей  
 $m \times k$



$\approx$



Сингулярные числа  
 $k \times k$

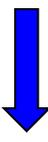
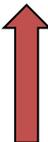
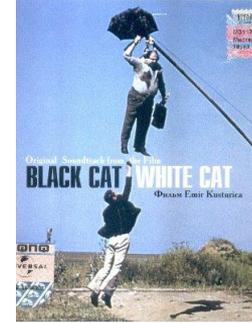
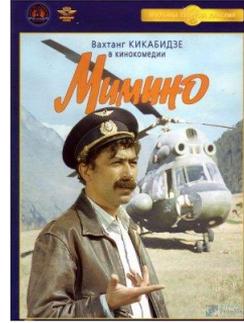
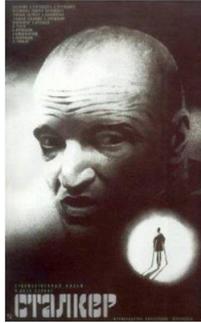


Факторы объектов  
 $k \times n$



$$\epsilon_1 \geq \epsilon_2 \geq \epsilon_3 \geq \epsilon_4 \geq \dots$$

# 1-й фактор



## 2-й фактор

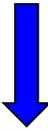
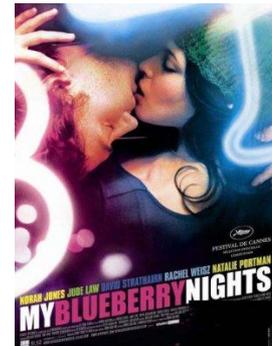
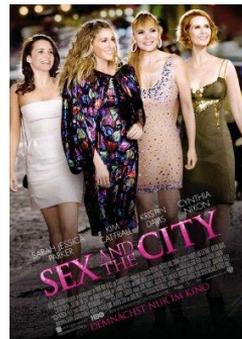
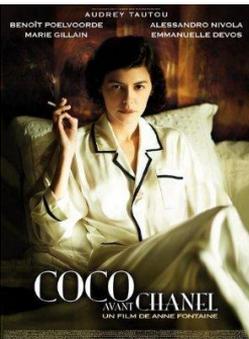
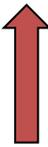
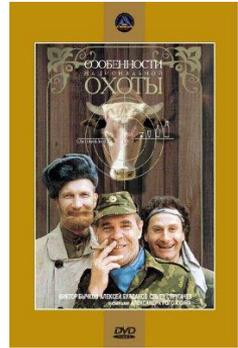
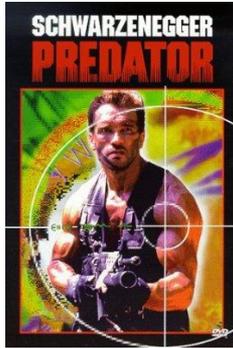




Российские интернет-технологии

2011

### 3-й фактор

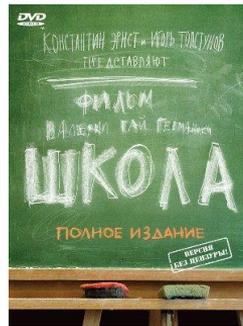
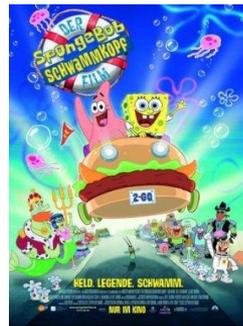
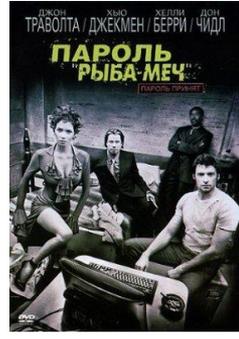
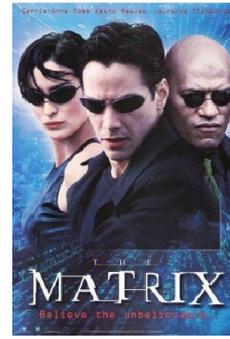
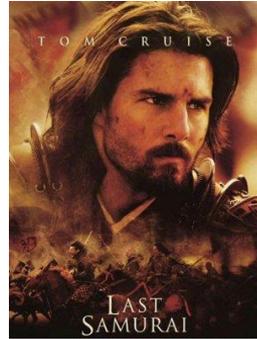
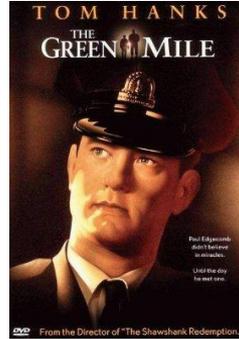
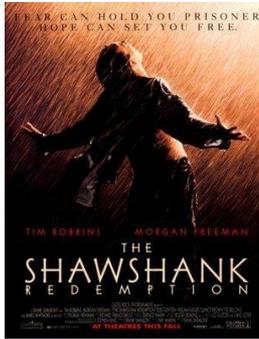




Российские  
интернет-технологии

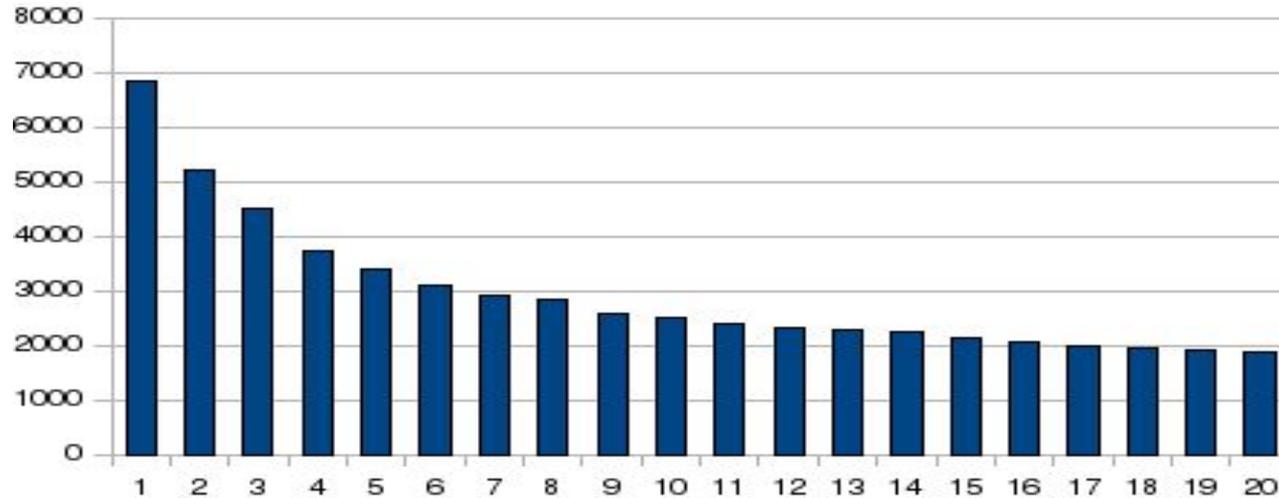
2011

## 4-й фактор





## Значения сингулярных чисел.

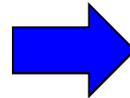
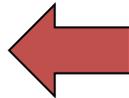


1. ориентация на качество **vs** ориентация на новизну [установка на качество]
2. доброе-простое **vs** трэш-сатира [эмоциональный знак]
3. мужской **vs** женский [пол]
4. высокобюджетный компактный **vs** дешевый сериал [дегги/время]



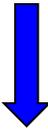
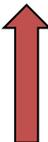
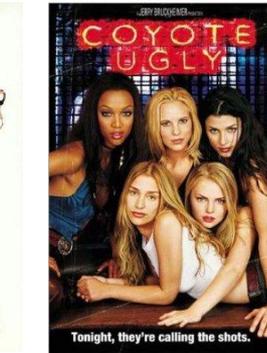
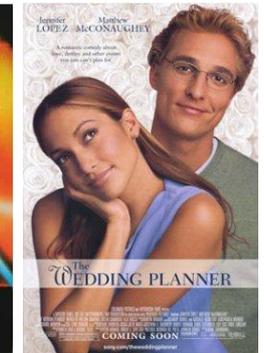
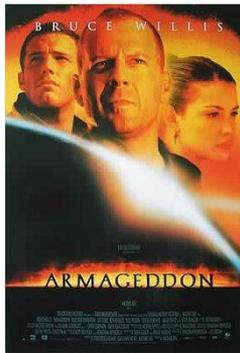
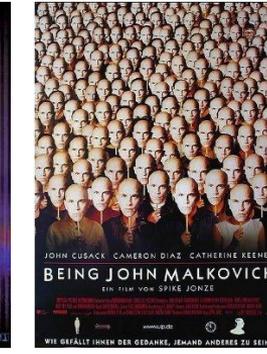
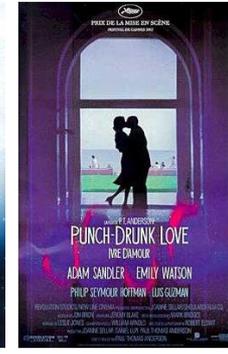
## 1-й фактор для книг по оценкам imhonet.

- Мертвые души
- Преступление и наказание
- Ревизор
- Евгений Онегин
- Горе от ума
- Идиот
- Капитанская дочка
- Шинель
- Гамлет, принц датский
- Братья Карамазовы
- Дон Кихот
- Война и мир
- Вишневый сад
- Илиада
- Приключения Алисы в стране чудес

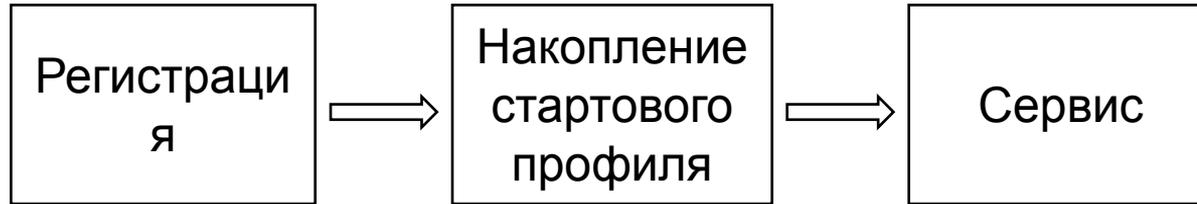


- Код да Винчи (Д.Браун)
- Последний дозор (С.Лукияненко)
- Ангелы и демоны (Д.Браун)
- Дневной дозор (С.Лукияненко)
- Точка обмана (Д.Браун)
- Цифровая крепость (Д.Браун)
- Ночной дозор (С.Лукияненко)
- Вечеринка в стиле “вамп” (А.Кош)
- Волкодав (М.Семенова)
- Чистовик (С.Лукияненко)
- Сумерки (С.Майер)
- Огненный патруль (А.Кош)
- Черновик (С.Лукияненко)
- Профессия: ведьма (О.Громыко)
- Таня Гроттер и Исчезающий Этаж

# 1-й фактор по оценкам Netflix



## Новый пользователь.

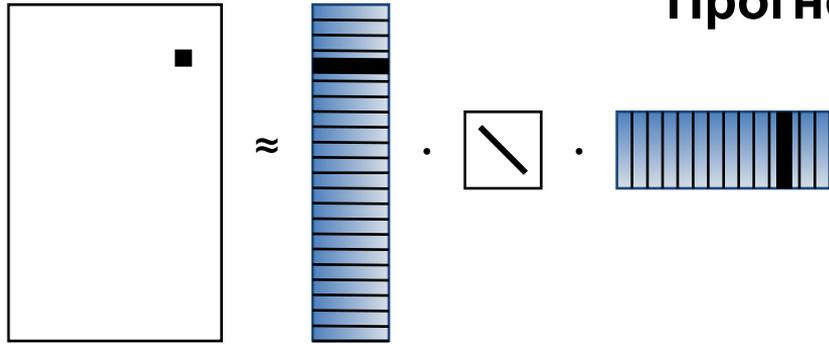


## Пример сбора стартового профиля.





## Прогнозы



$$r_{\text{прогноз}}^{so} = (\bar{p}, E\bar{q}) = \sum_{t=1..k} p_t \cdot \varepsilon_t \cdot q_t$$

## Объяснения

$$r_{\text{прогноз}}^{so} = (\bar{p}, \hat{E}\bar{q}) = \hat{P}_s \cdot \bar{r}_o = \sum_i a_i \cdot r_i^o = \hat{P}_o \cdot \bar{r}_s = \sum_j b_j \cdot r_j^s$$

через оценки этого объекта  
другими пользователями

предыдущими оценками  
пользователя



### Пять сантиметров в секунду

Аниме / Япония, 2007 год

Рекомендовано, так как вам понравилось: Девушка и кот, Семь самураев

Прогноз 8,6 56 рекомендателей

Средняя 8,2 4966 оценок в т.ч. знакомые

Моя оценка



### Пристрели их

Боевики, Триллеры, Приключения, Комедии / США, 2007 год

Рекомендовано, так как вам понравилось: Быть Джоном Малковичем, Типа крутые легавые, Семь самураев

Прогноз 7,3 92 рекомендателя

Средняя 7,1 4970 оценок в т.ч. знакомые

Моя оценка



### Волчок

Драмы, Артхаус, Российские фильмы / Россия, 2009 год

[Смотреть онлайн](#)

Рекомендовано, так как вам понравилось: Иди и смотри, Воображариум доктора Парнаса

Прогноз 7,8 20 рекомендателей

Средняя 7,2 811 оценок

Моя оценка



### Мост в Терабитию

Фантастика, Приключения, Фильмы для детей / США, 2007 год

Рекомендовано, так как вам понравилось: Донни Дарко, Лабиринт Фавна, Коралина в Стране Кошмаров

[Есть на толкучке](#)

Прогноз 8,1 98 рекомендателей

Средняя 7,3 10758 оценок в т.ч. знакомые

Моя оценка

[vbobrikov@imhonet.ru](mailto:vbobrikov@imhonet.ru)

[blog.imhonet.ru/community/1702/](http://blog.imhonet.ru/community/1702/)

[pcaimhonet@googlegroups.com](mailto:pcaimhonet@googlegroups.com)



Российские  
интернет-технологии

2011