

Рекомендательные системы на основе вложенных тегов

Исполнил:

Чеботаев А. П.

Научный руководитель:

Бухановский А. В., д.т.н

Кафедра Компьютерных Технологий, ФИТиП

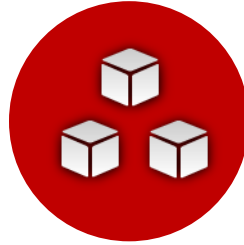
СПб 2011

Задачи рекомендательных систем

пользователи



объекты



функция «полезности»

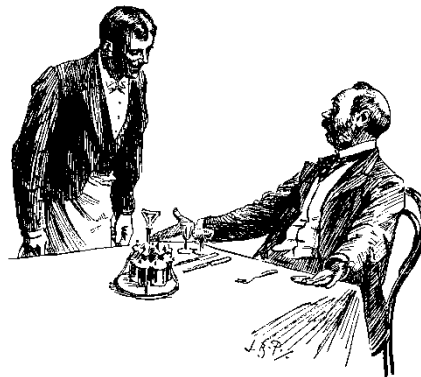
$$u(\text{green circle with people icon}, \text{red circle with cubes icon})$$

Современные РС

- Упорядочивание

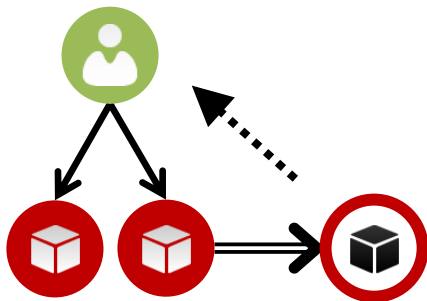
Направления развития

- Обоснование
- Фильтрация



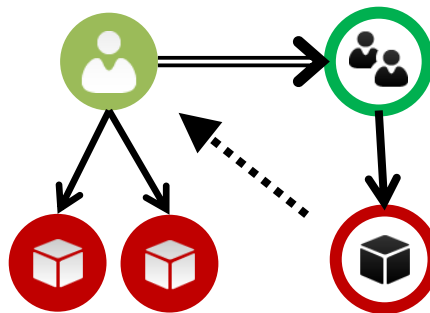
Существующие методы

Контентные методы



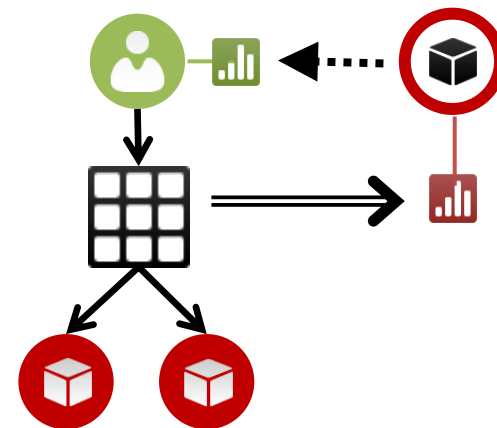
- ✗ Узкие рекомендации
- ✗ Однородность объектов

Совместная фильтрация



- ✗ Ресурсоемкость
- ✗ Похожие пользователи не являются обоснованием

Скрытые факторы

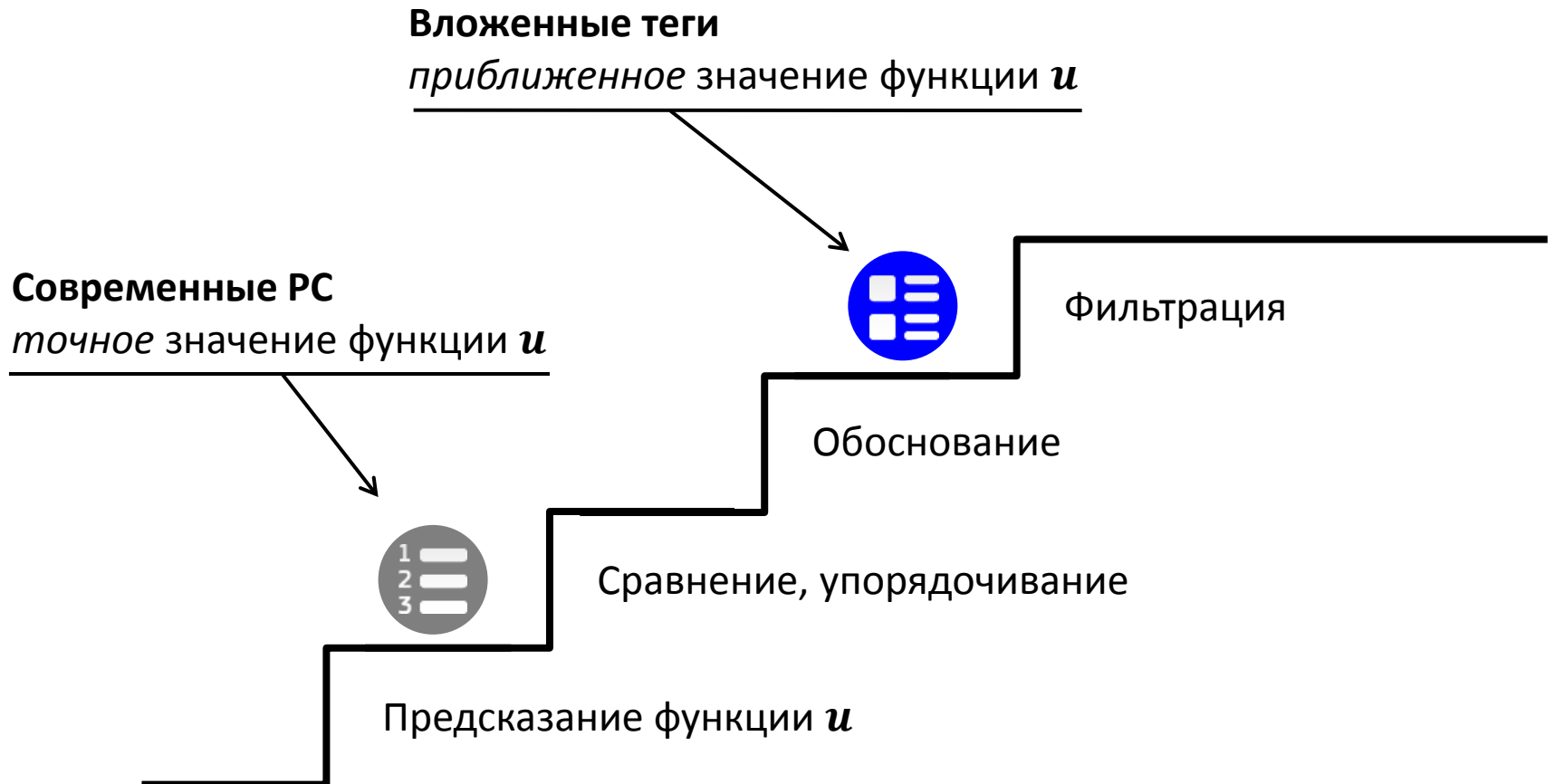


- ✗ Ресурсоемкость
- ✗ Факторы являются скрытыми

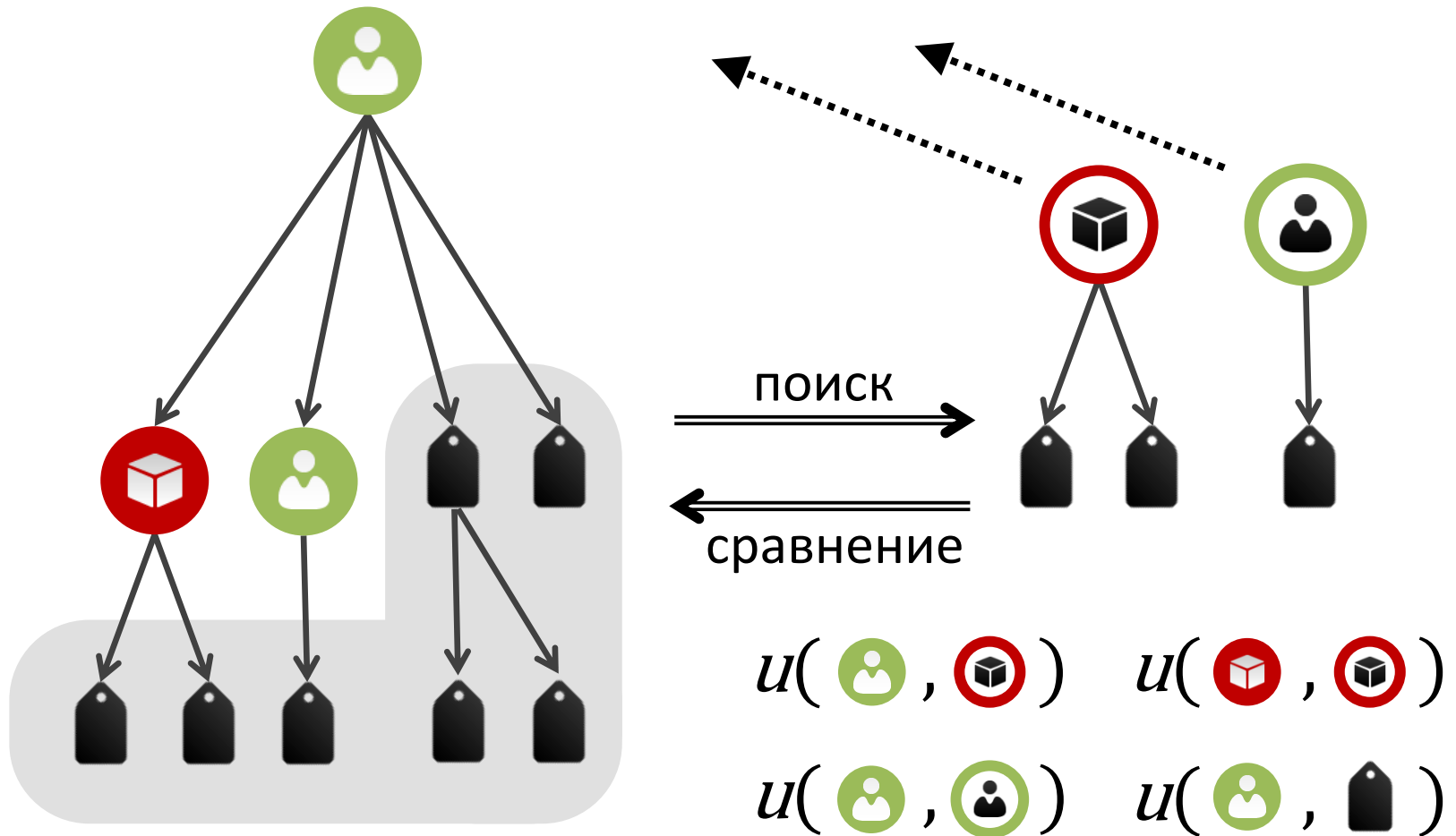
⇒ поиск

.....→ рекомендация

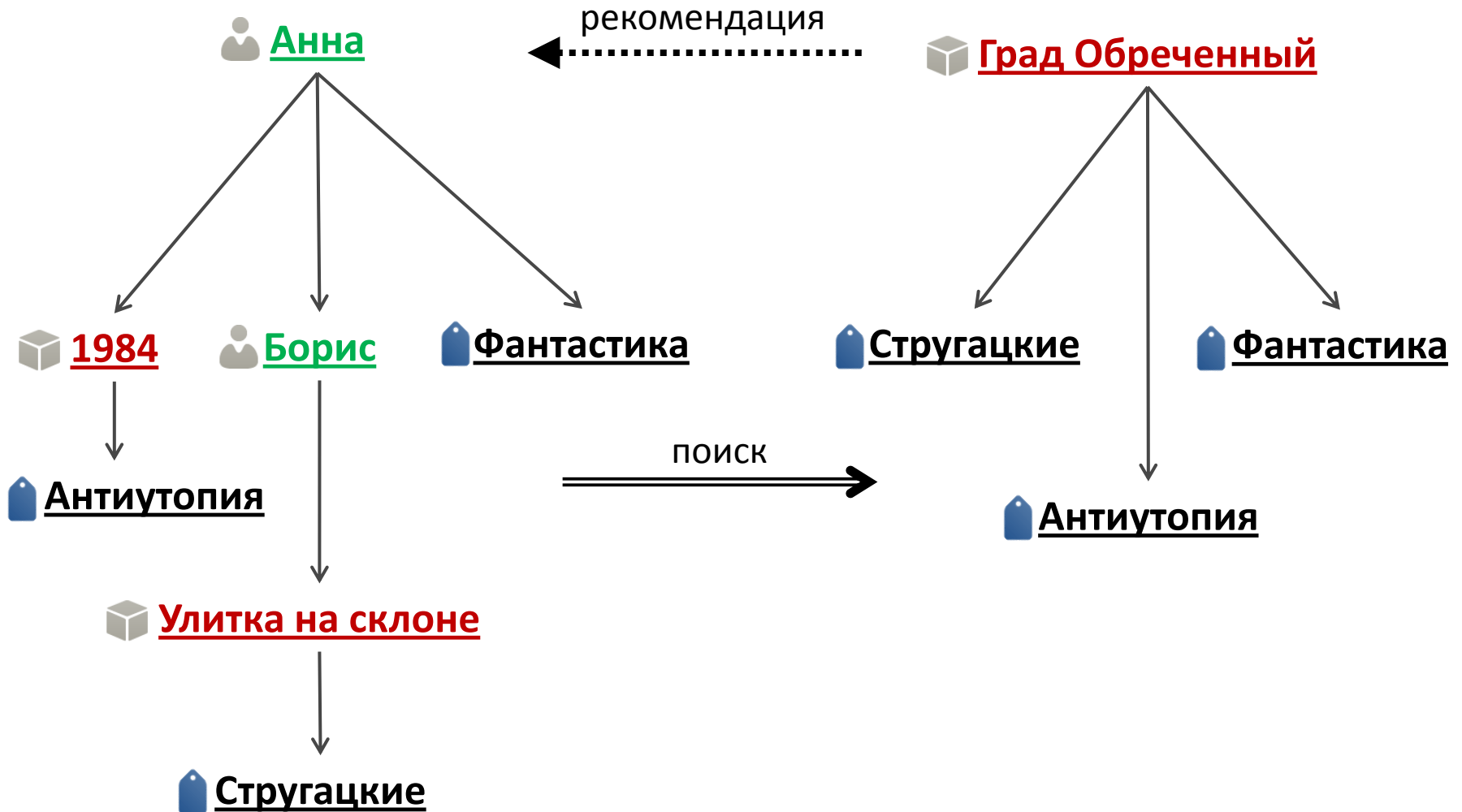
Рекомендательные системы



Вложенные теги

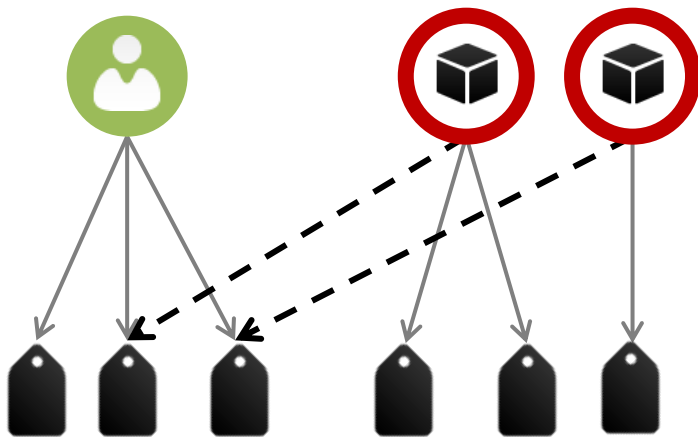


Пример



Поиск объектов

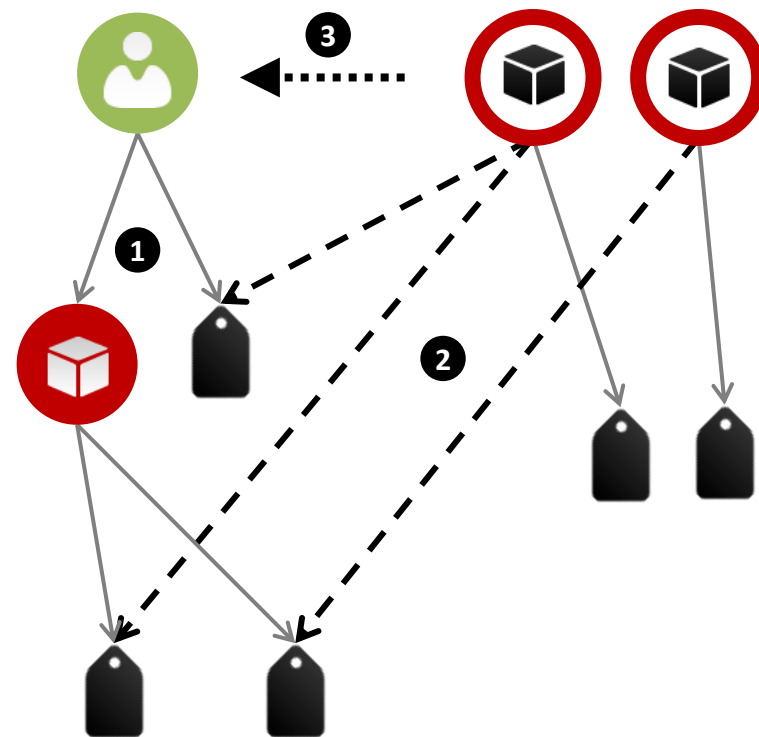
1. Поиск компонент
2. Поиск их родителей
3. Сравнение, фильтрация



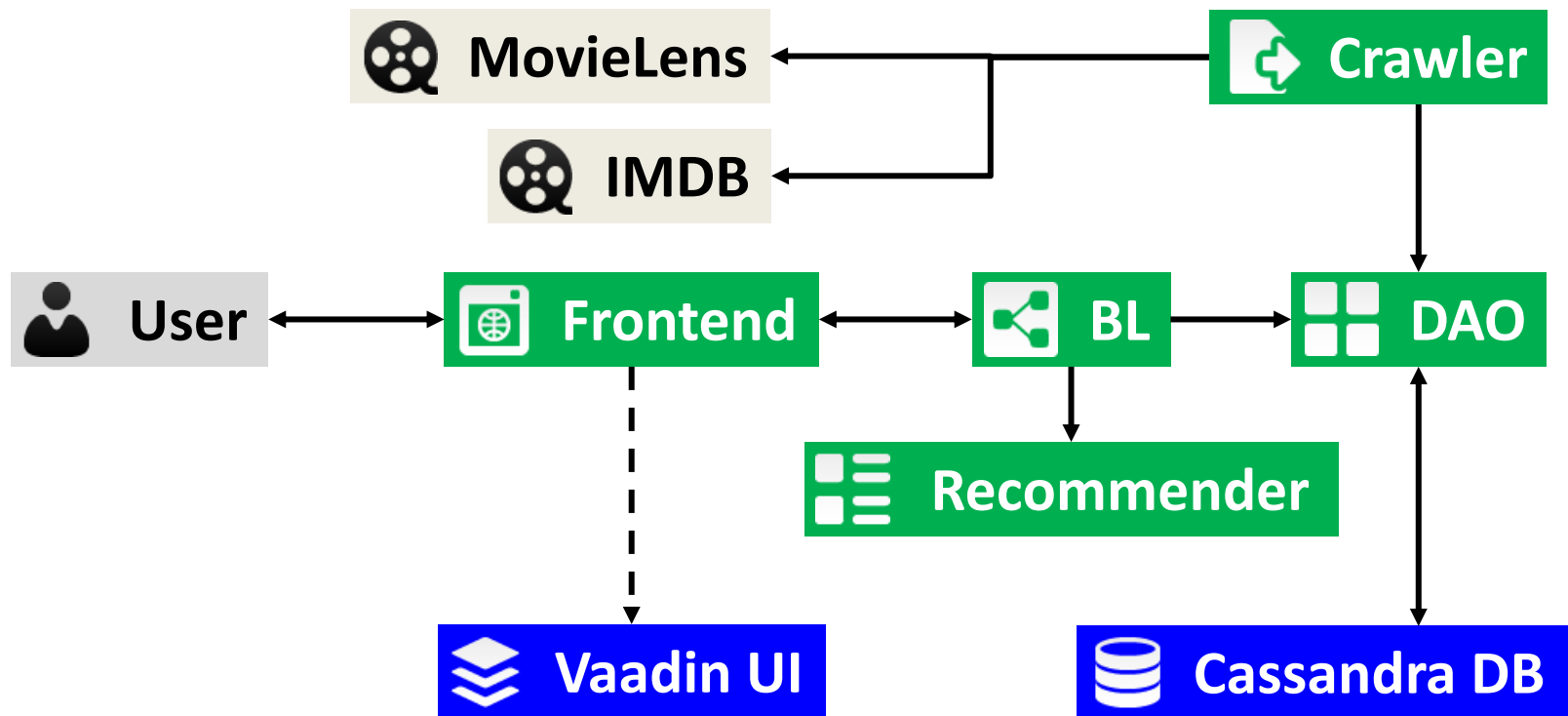
Поиск — $O(n^2)$

Сравнение — $O(n)$

n — среднее количество компонент



Экспериментальный стенд



10 000 фильмов

350 000 тегов (актеры, режиссеры, метки пользователей)

Обоснование рекомендаций

Interface showing a user profile and a list of recommended items with their associated tags and scores.

User Profile: Anton Chebotayev (Антон Чеботаев)

Recommended Items and Tags:

- jedi-knight** 10,0 | **star wars** 10,0 | **Robert (I) Rodriguez** 10,0
- George (I) Lucas** 10,0 | **Samuel L. Jackson** 10,0 | **nonlinear** 10,0
- death-star** 10,0 | **vampire-hunter** 10,0 | **Quentin Tarantino** 10,0
- blood** 10,0 | **stop-action** 10,0 | **gangster** 10,0 | **Four Rooms** 10,0
- Bruce Willis** 10,0 | **coffee** 10,0 | **neo-noir** 10,0
- slow-motion-action-scene** 10,0 | **lightsaber** 10,0 | **vampire-slayer** 10,0
- violence** 10,0 | **space-battle** 10,0 | **George Clooney** 10,0

Item Details:

- Илья Пименов** 10,7
Bruce Willis Jason Statham Pulp Fiction Gary Oldman Johnny Depp
Evil Dead, The Brad Pitt
- Snatch** 5,2
slow-motion-action-scene stop-action Guy Ritchie Tim Fal Day
Jason Statham neo-noir blood narrated-by-character
nonlinear-timeline gangster violence rhyming-slang
stuffed-in-a-trunk yardie gangster-comedy hasidic-jew
man-eating-pig jewish-mafia head-brace attacked-by-a-d
- Star Wars: Episode II - Attack of the Clones** 5,0
death-star jedi-knight George Lucas lightsaber space battle
Samuel L. Jackson star wars
murdered-before-giving-protagonist-information prequel-cult-film
mistaking-reality-for-dream god becoming-evil galactic-war
character-feels-around-for-missed-head foreshadow
mechanical-hand closing-the-eyes-of-a-dead-person
fate-of-the-universe changeling decapitation ends-with-a-wedding
- From Dusk Till Dawn** 4,5
Quentin Tarantino Robert (I) Rodriguez vampire-hunter
George Clooney vampire-slayer blood violence mexican-and
pencil-stabbing holy-water-gun role-in-hand pre-columbian
bloody-hand knife-in-hand boss constrictor head-chopped off
obscene-gesture held-hostage border-guard bloody-spr

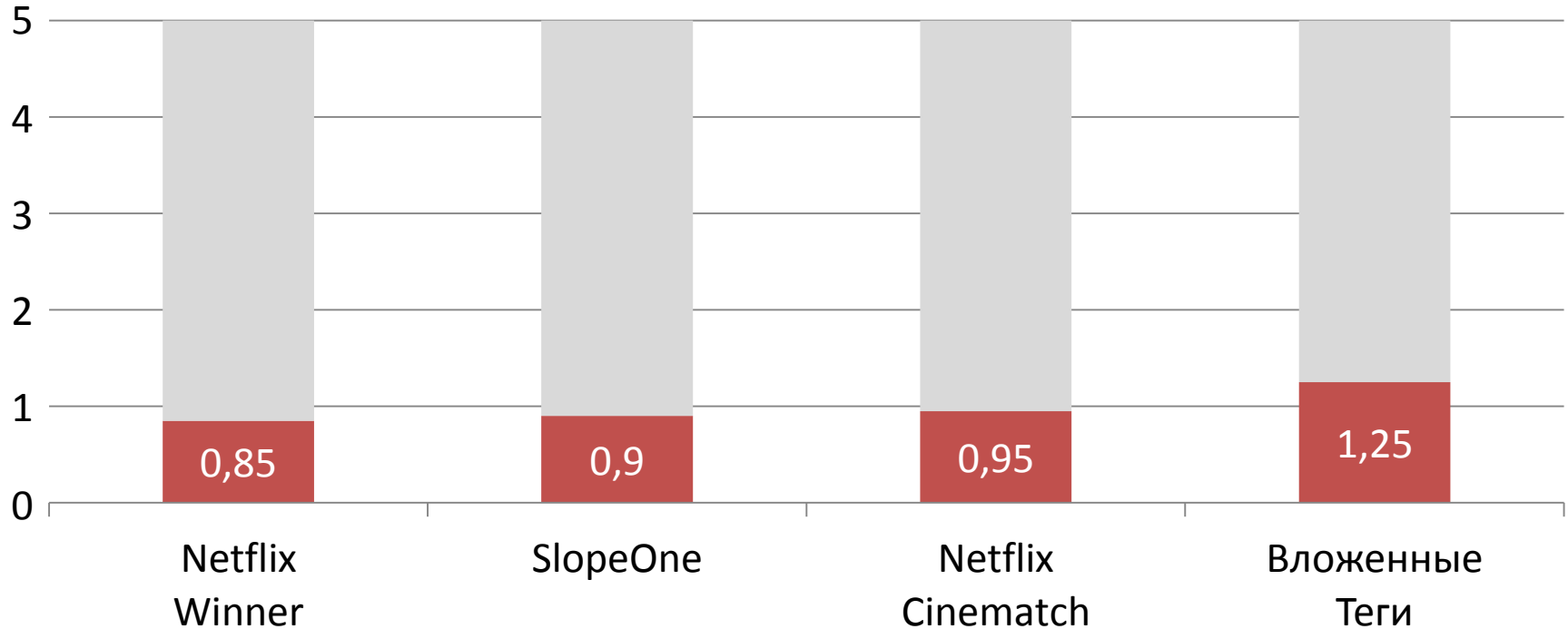
Navigation Icons: User profile, Tag, Tag, Item (highlighted in red), Tag, Item (highlighted in red), User profile.

Фильтрация рекомендаций

The diagram illustrates the filtering of recommendations. At the top, four icons represent different types of input: a green circle with a person icon (User), a red circle with a cube icon (Item), and two black tag icons. Blue arrows point from these icons to a search bar. The search bar contains the text: "Антон Чебограев", "Pulp Fiction", "Guy Ritchie", "gangster", and "Add Search Keywords". Below the search bar is a green button labeled "Get Something Like That". Below the button, three movie recommendations are listed, each with a red circle containing a cube icon and a blue arrow pointing to the recommendation title. The recommendations are:

- Kill Bill: Vol. 2** (8,7)
 - Tags: Quentin Tarantino, Samuel L. Jackson, neo-noir, blood, nonlinear-timeline, violence, barstow-california, reference-to-annie-oakley, death-by-snakebite, reference-to-spiderman, perth-australia, tragic-heroine, roll-call, wedding-singer, sheath, shaolin-temple, master-apprentice-relationship, hitlist, one-woman-army, intentional-goof, wedding-chapel, swordsmanship, head-in-toilet, helplessness, shot-back-to-back, costumer, eye-ripped-out, duology, mother-child-reunion, holding-head-under-water
- Snatch** (8,0)
 - Tags: slow-motion-action-scene, stop-action, Guy Ritchie, Tim Faraday, Jason Statham, neo-noir, blood, narrated-by-character, nonlinear-timeline, gangster, violence, rhyming-slang, stuffed-in-a-trunk, yardie, gangster-comedy, hasidic-jew, man-eating-pig, jewish-mafia, head-brace, attacked-by-a-dog, running-out-of-ammo, concorde, unintelligible, bare-knuckle-fighting, fake-gun, dog-fighting, shot-through-a-wall, Eli Cross, Guluzar Ritchie, Taha Ali Reza
- From Dusk Till Dawn** (7,8)
 - Tags: Quentin Tarantino, Robert (I) Rodriguez, vampire-hunter, George Clooney, vampire-slayer, blood, violence, mexican-band, pencil-stabbing, holy-water-gun, hole-in-hand, pre-columbian, bloody-hand, knife-in-hand, boa-constrictor, head-chopped-off, obscene-gesture, held-hostage, border-guard, bloody-spray, body-horror, stab-wound, melting-face, Gianni Nunnari, Elizabeth Avellan, Lawrence Bender, Meir Teper, Paul Helleman, Nellie Kurtzman, John (I) Esposito

Неточные оценки



1 000 000 рейтингов
6 000 пользователей
3 900 фильмов
100 000 ключевых слов



1. Обучение на 800 000 рейтингов
2. Проверка на 200 000 рейтингов
3. Измерение среднеквадратичной ошибки

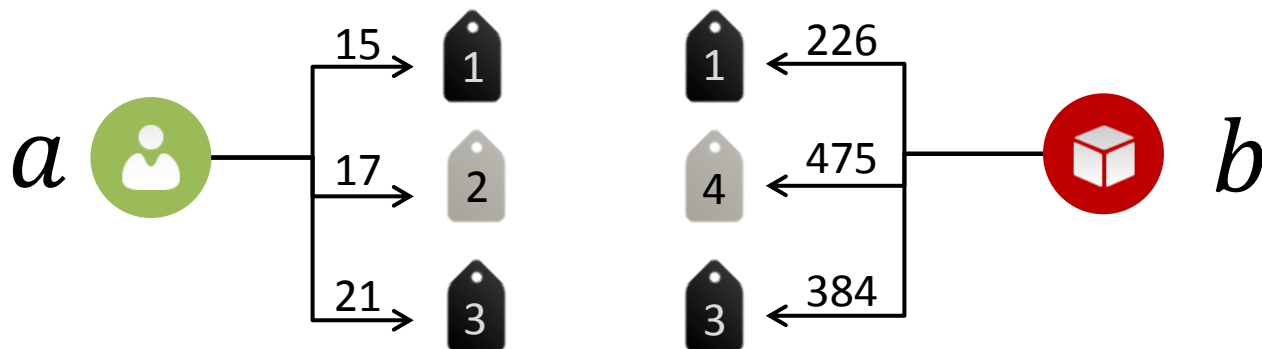
Результаты

1. Выявлены проблемы существующих рекомендательных систем
2. Предложен метод, решающий проблемы обоснования и фильтрации рекомендаций
3. Построен экспериментальный стенд рекомендательной системы по предложенному методу
4. Произведено сравнение результатов предложенного метода с существующими

Выводы

- Отсутствие явных факторов не позволяет фильтровать рекомендуемые объекты
- Возможно построение *обосновывающих* рекомендательных систем с функциональностью *фильтрации* рекомендаций при меньших затратах ресурсов
- Допустимо использование предложенного метода в промышленных системах

Сравнение объектов



- Объект a связан с $S_a = \{o_{a_1}, o_{a_2}, \dots, o_{a_n}\}$
Вес связи $o_1 \rightarrow o_2$ обозначим $w(o_1, o_2)$
- $pw(a, o) = \frac{w(a, o_i)}{\sum_i^{o_i \in S_a} w(a, o_i)}$ — взвешенный вес
- $u(a, b) = \sum_i^{o_i \in S_b \cup S_a} \min(pw(o_i, a), pw(o_i, b))$

Развитие метода

- Функции сравнения
- Алгоритмы поиска
- Качество описания объектов
 - Анализ рецензий фильмов
 - Фильтрация пользовательских данных