

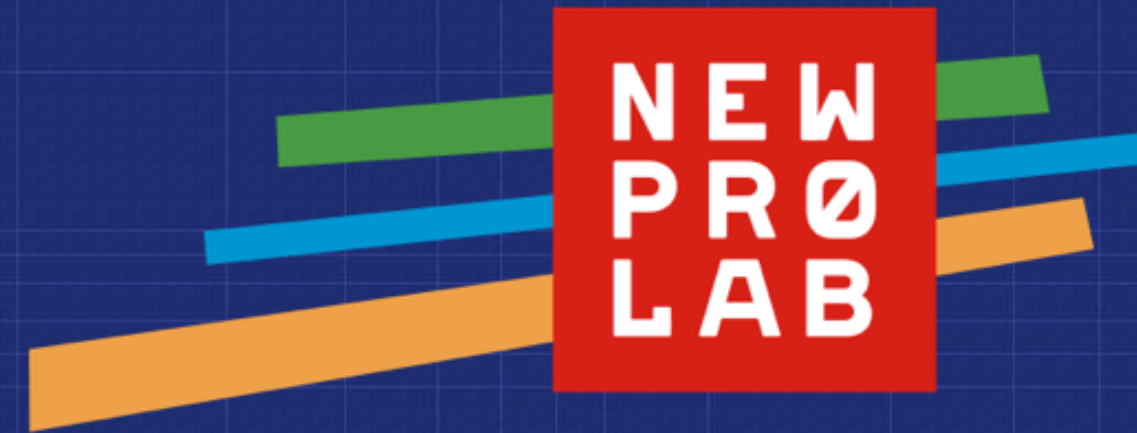


NEW
PRO
LAB

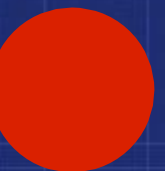
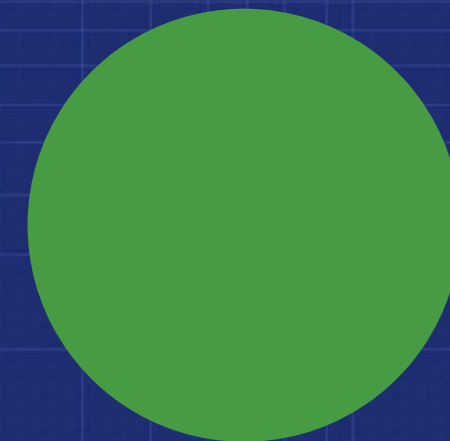
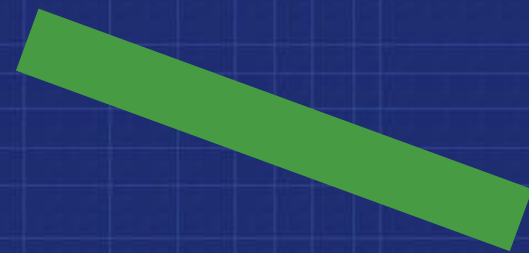
Рекомендательные системы

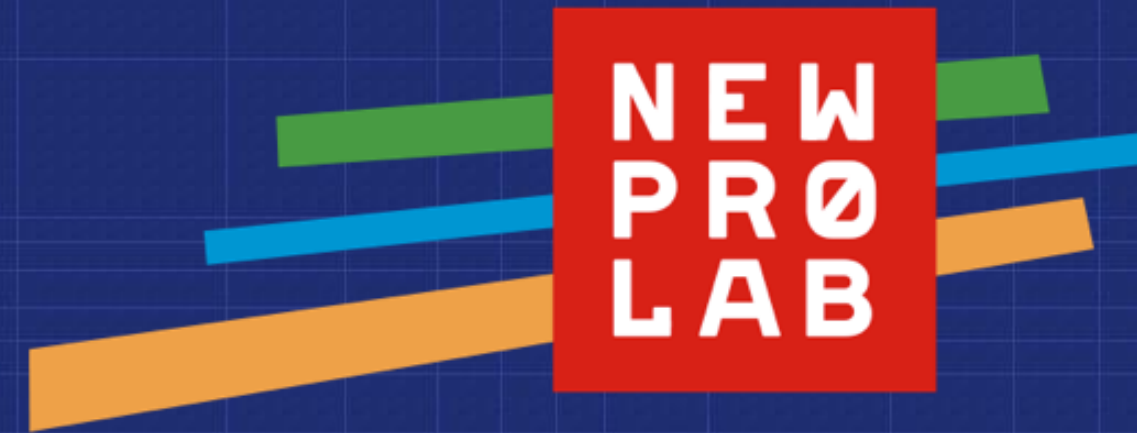
Гриша Сапунов

NEWPROLAB.COM

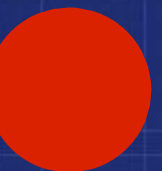
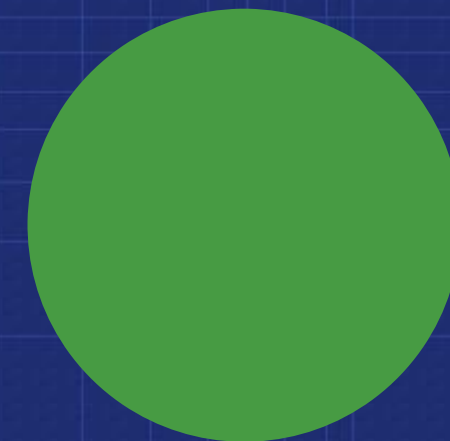
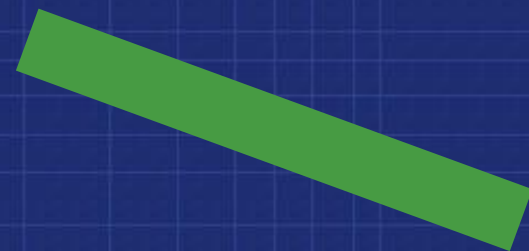


Гриша Сапунов: grigory.sapunov@gmail.com
Дмитрий Игнатов: dmitrii.ignatov@gmail.com



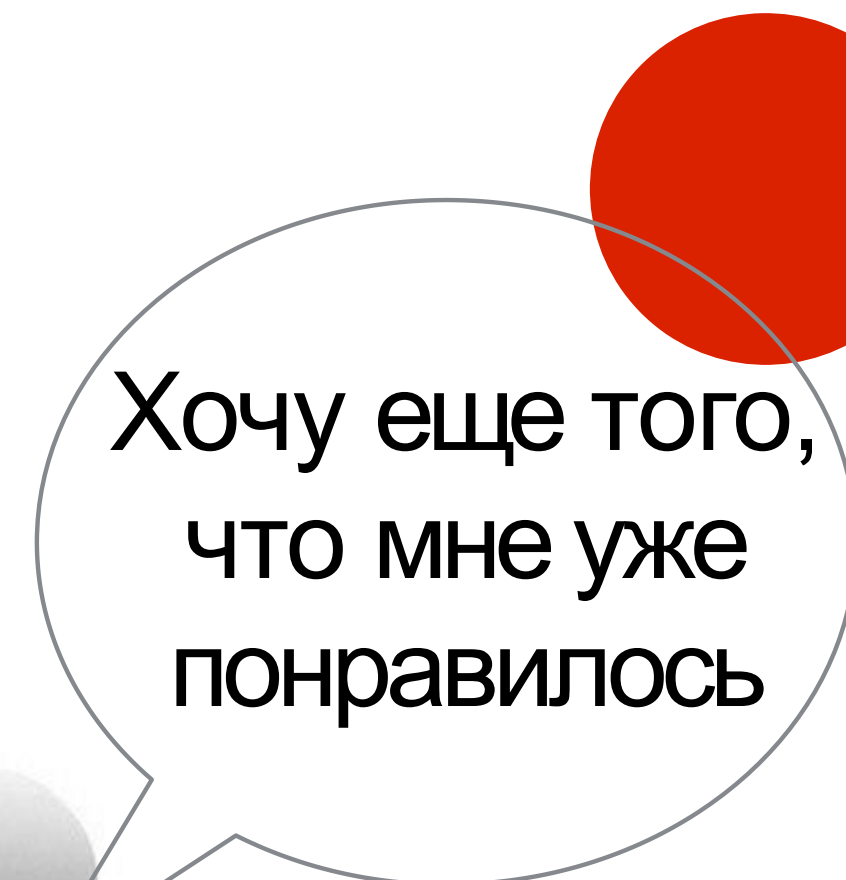


Content-based recommenders





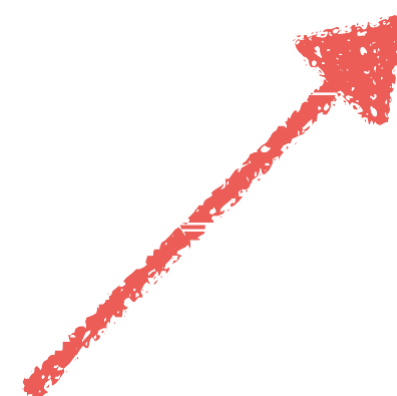
СВ
RecSys



контент продукта



Название	
Режиссер	
Год	
Актер	
Жанр	
...	



Профиль пользователя



КОНТЕНТ продукта

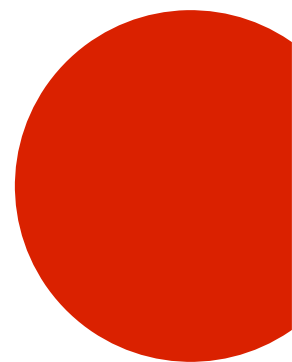


Название	
Режиссер	
Год	
Актер	
Жанр	

- явный (структурированный)
- текстовый контент
- сигнал (аудио, картинки)



Структурированные данные



ЯВНЫЙ КОНТЕНТ



- режиссер
- актеры
- год
- жанр
- бюджет



- бренд
- цвет
- фасон
- страна
- сезон
- цена



- кухня
- чек
- место
- вайфай
- время работы

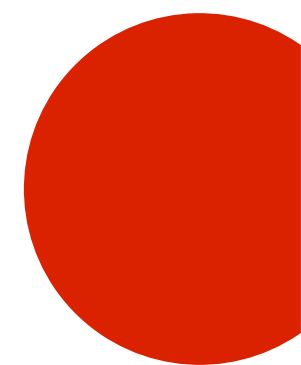




Структурированные данные



бренд	Desigual, Concept Club, ...
цвет	желтое, зеленое, красное, ...
фасон	вечернее, повседневное, ...
страна	Россия, Италия, Франция, ...
сезон	демисезон, лето, зима, ...





Структурированные данные



Desigual	0
Concept Club	1
желтое	1
зеленое	0
красное	0
вечернее	0
повседневное	1
Россия	1
Италия	0
Франция	0
демисезон	1
лето	0
зима	0







NEW
PRO
LAB



Surfingbird

Все ваши интересы в одной ленте!





Новости



Фото



Видео



Арт



Дизайн



Путешествия



Мода



Природа



Животные



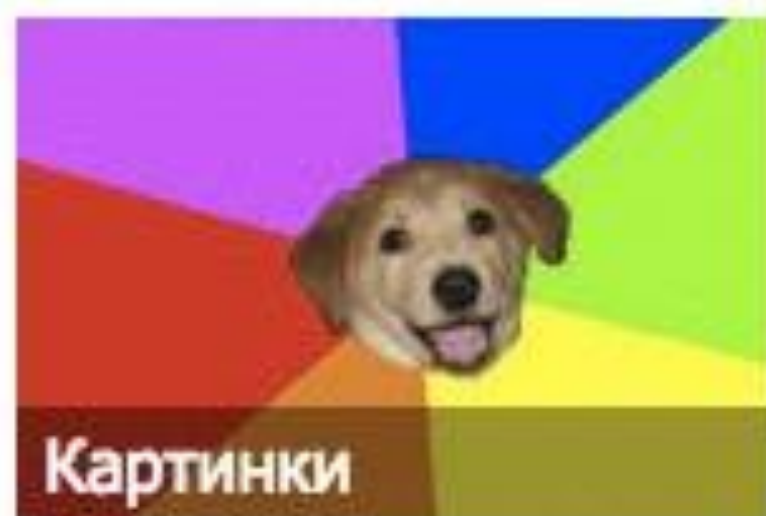
Интерьер



Еда и рецепты



Шоппинг



Картинки



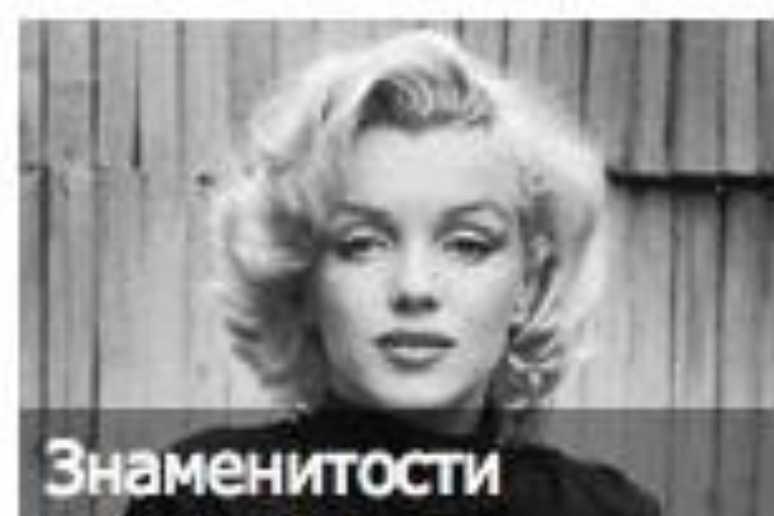
Гифки



Общество



Интернет



Знаменитости



Архитектура



Музыка



Кино



Запросы, комментарии,
описание видео, название
видео

80%



YouTube

RU

machine learning



0:07 / 1:08:39

Lecture 1 | Machine Learning (Stanford)



Stanford

Подписаться

438 260

737 499

+

Добавить в

Поделиться

...

Ещё

 2 797

 56

Дата загрузки: 22 июля 2008 г.

Lecture by Professor Andrew Ng for Machine Learning (CS 229) in the Stanford Computer Science department. Professor Ng provides an overview of the course in this introductory meeting.

Картинки



fastr — a flickr game

метаданные:
название файла

ВИЗУАЛЬНЫЙ КОНТЕНТ:

- цвет
- текстур
- а форма

tags:

- hand-made
- penguin
- toy

NEW
PRO
LAB

Картинки

Google Image Labeler

time left
00:28


score
420

passes
0

seriously dude its jim morrison

label pass

Your partner wants to pass.



zoom out

off-limits

hair
man

my labels

doors
jim morrison
black and white
rock
musician
artist

PANDORA® business model



80,000,000+
registered users in the U.S.



800,000+
songs in the PANDORA® collection



8,000,000,000+
thumbs



Music Genome Project®

- Comprehensive database of music analysis
- Up to 480 attributes per song



Collective Feedback

- 8 billion+ thumbs and growing
- 1.4 billion+ stations created



Individual Engagement

- Real-time adaptation of playlist
- Avg 10+ listening hours per month per listener



National & Local Advertising

- Display, audio and video solutions
- Multi-platform: computers, smartphones, automotive, CE devices

NEW
PRO
LAB

Хочу еще того,
что мне уже
понравилось



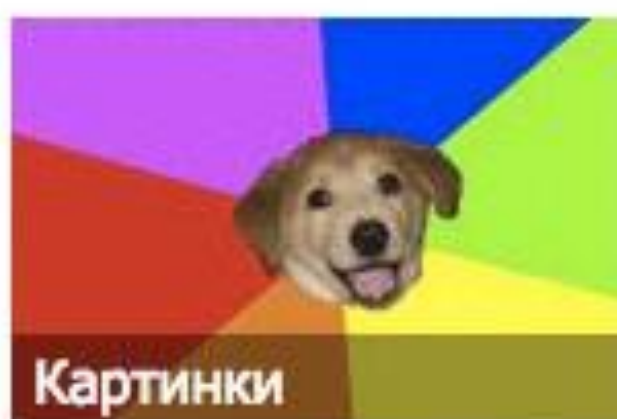
Профиль пользователя





Выбери интересы, чтобы мы могли понять, что тебе рекомендовать

Далее



To get started, tell MovieLens about your preferences by distributing 3 points among your favorite groups of movies below.

Remaining points: 3

+

chick flick, feel-good, touching



Titanic



Dead
Poets
Society



Slumdog
Millionaire

+

computer animation, good versus evil, mythology



Toy Story



The Lord
of the
Rings: The
Fellowship
of the
Ring



Harry
Potter and
the
Philosopher
Stone

+

action, fun movie, special effects



True Lies



The Mask



Men in
Black II

+

blood, dark humor, social commentary



Pulp
Fiction



Kill Bill:
Vol. 1



American
History X

+

classic, masterpiece, quotable



The
Godfather



Psycho



The
Graduate

+

dramatic, good acting, intense



Forrest
Gump



Million
Dollar
Baby



The Social
Network

pick a recommender

⚠ Once you rate 15 movies, you may select a fully personalized recommender ("the warrior", or "the wizard").

- ☐ "the peasant"
non-personalized
- ☒ "the bard"
based on movie group point allocation ([configure](#))
- ☐ "the warrior"
based on ratings
- ☐ "the wizard"
based on ratings

your movies

[ratings \(about your ratings\)](#)

Life Is Beautiful

1997 PG-13 116 min 📅



Forrest Gump

1994 PG-13 142 min 📅



Sav

1996



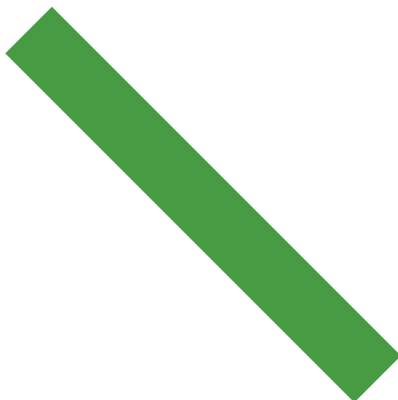
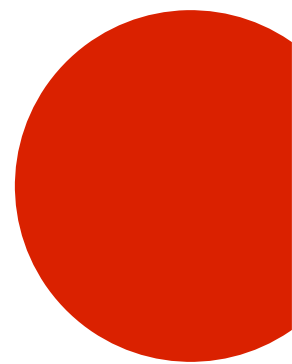
ence of t

118 min 📅





NEW
PRO
LAB



The Professional

1981 108 min 準



Guest from the Future

1984 334 min 準



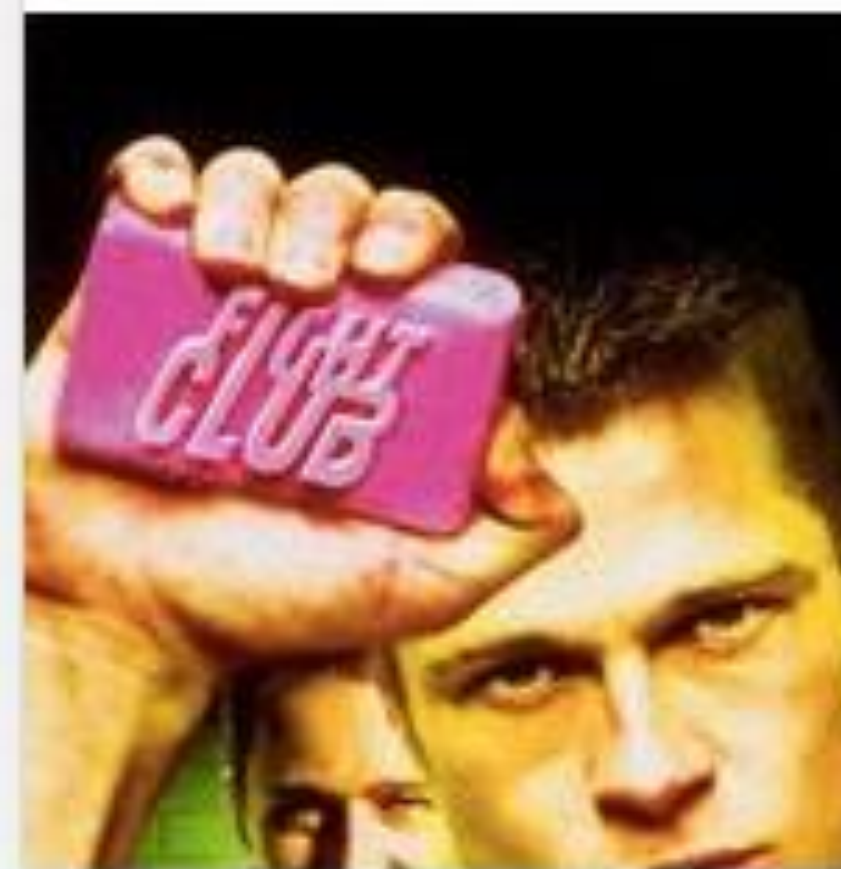
Star Wars: Episode I

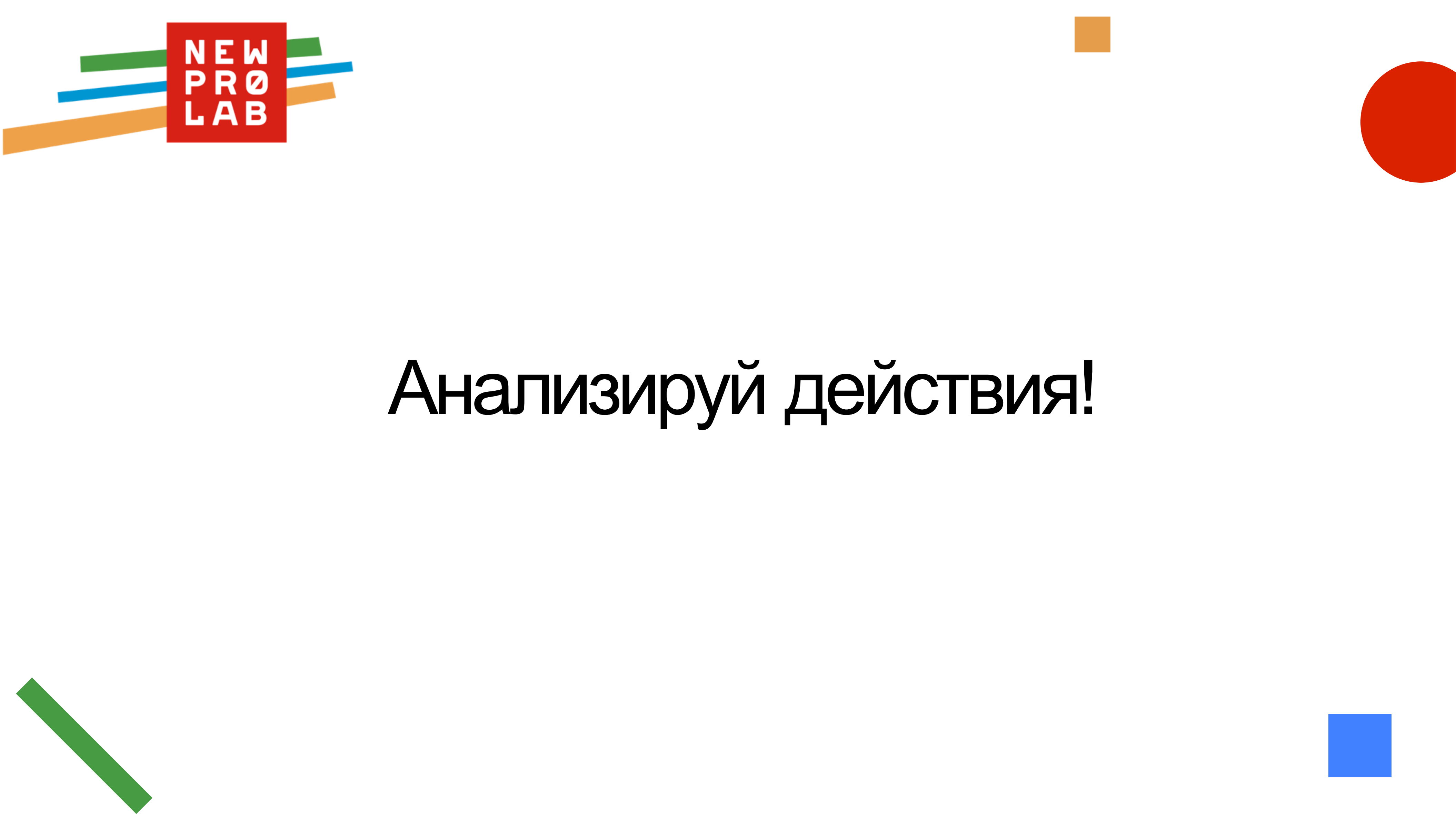
1977 PG 121 min 準



Fight Club

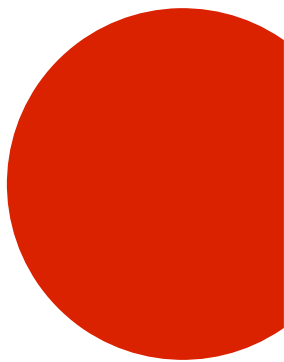
1999 R 139 min 準



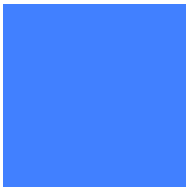
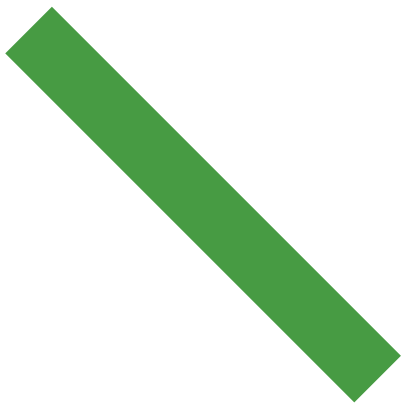


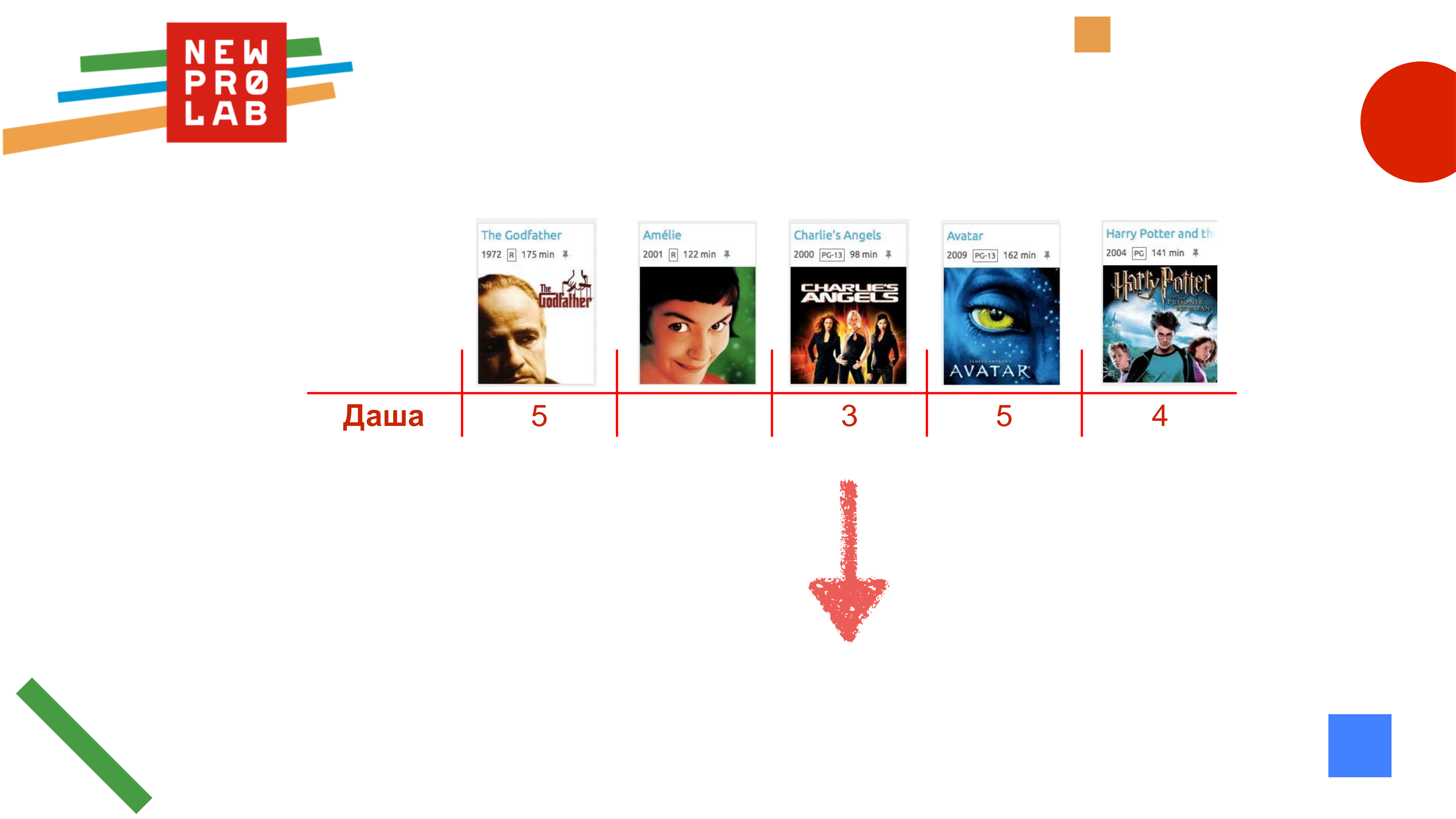
NEW
PRO
LAB

Анализируй действия!



Ваня	3		3	2	3
Маша	5	5	2	5	5
Даша	5		3	5	4
Оля	4		1	4	3
Катя	5	5		5	
Петя	5		2	5	2
Саша	5		1	4	4
Илья	1				5
Андрей	3			3	3
Аня	5		4		4





NEW
PRO
LAB



Даша

5





3

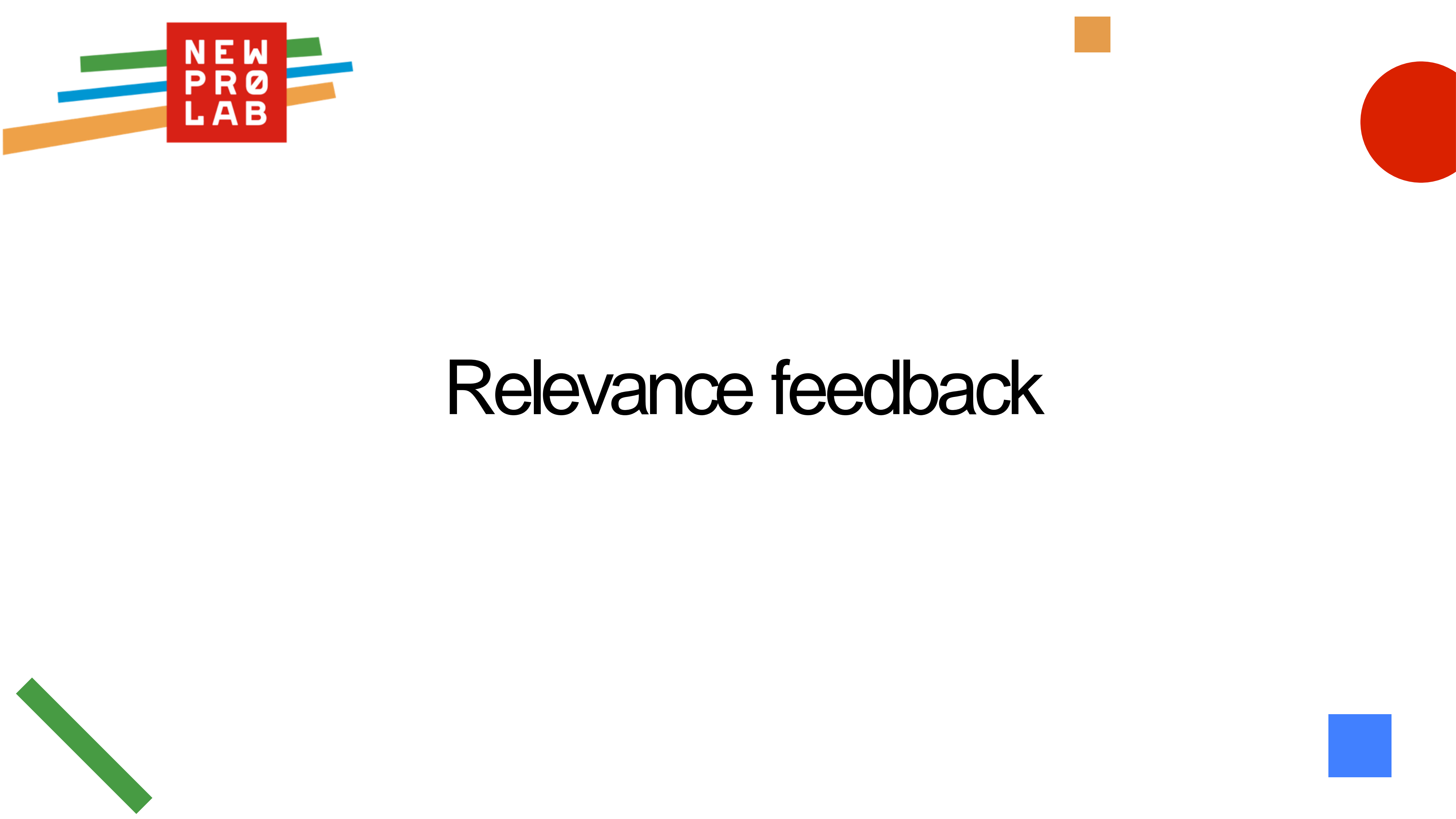
5

4



	Режиссер	Год	Актер	Жанр	Даша
The Godfather 1972 R 175 min					5
Charlie's Angels 2000					3
Avatar 2009	James Cameron	2010	Sam Worthington	Action, Fantasy, Science, Fiction	5
Harry Potter 2004 PG 141					4

	Режиссер	Год	Актер	Жанр	Даша
					5 - 4.25
					3 - 4.25
	James Cameron	2010	Sam Worthington	Action, Fantasy, Science, Fiction	5 - 4.25
					4 - 4.25



NEW
PRO
LAB

Relevance feedback

профиль: new & old

feedback

$$P' = \alpha P + \beta D$$

relevance

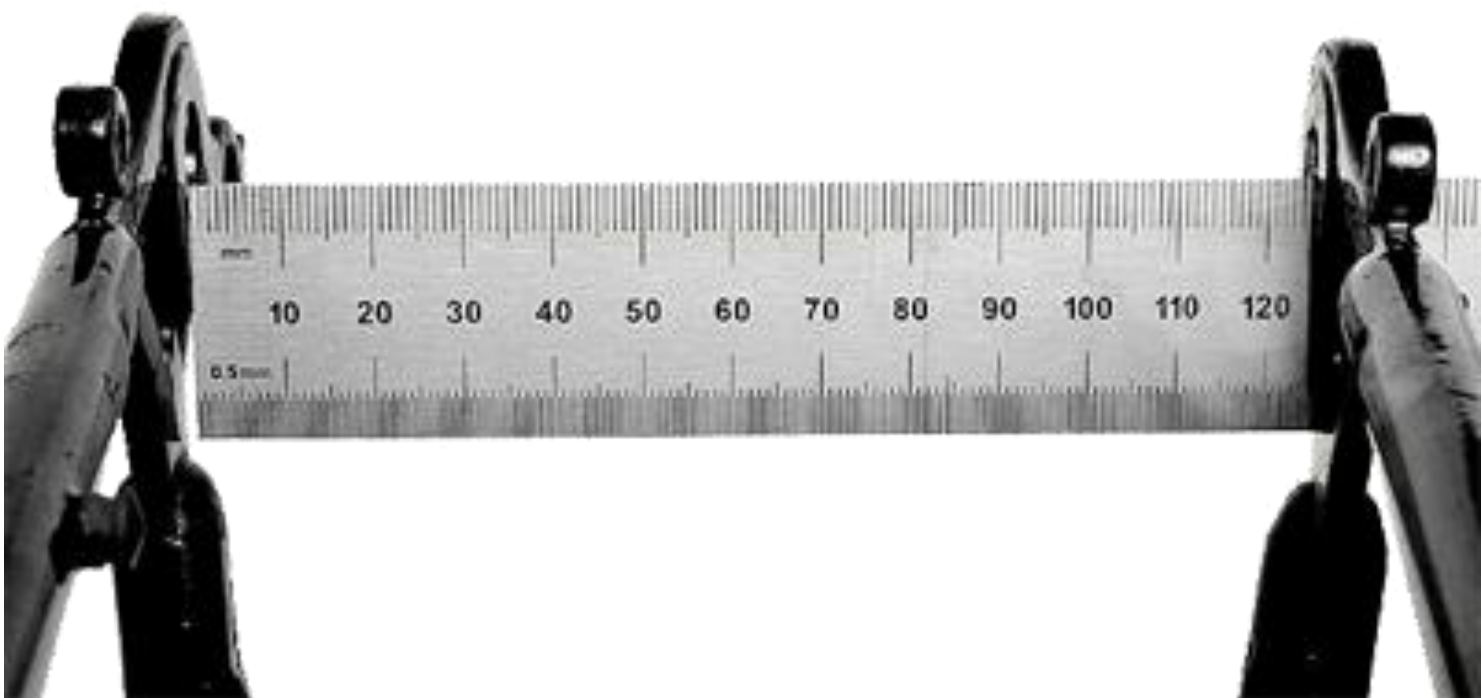
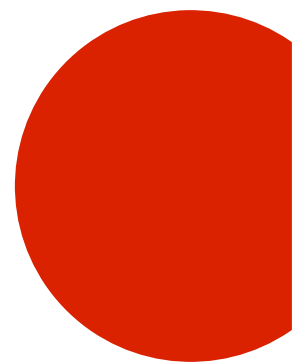
пример

$$P' = 0.5P + 0.5D_+ - D_-$$

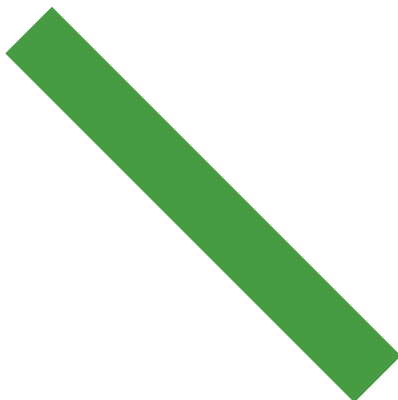
$$P' = \alpha P + \beta D$$

relevance

- интерактивно учим профиль
- α позволяет забывать short-term интересы



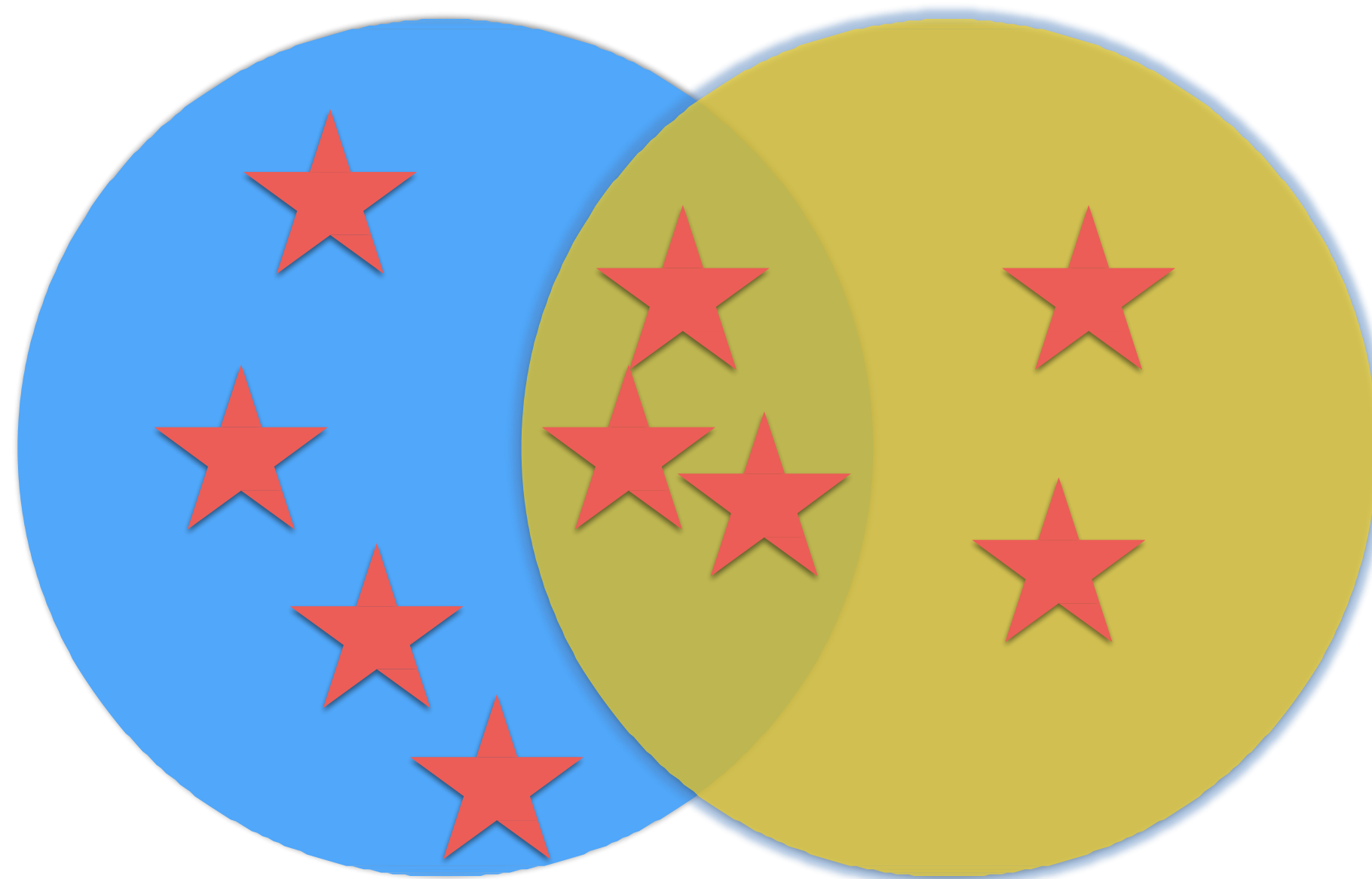
Мера близости



Коэффициент Жаккарда

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

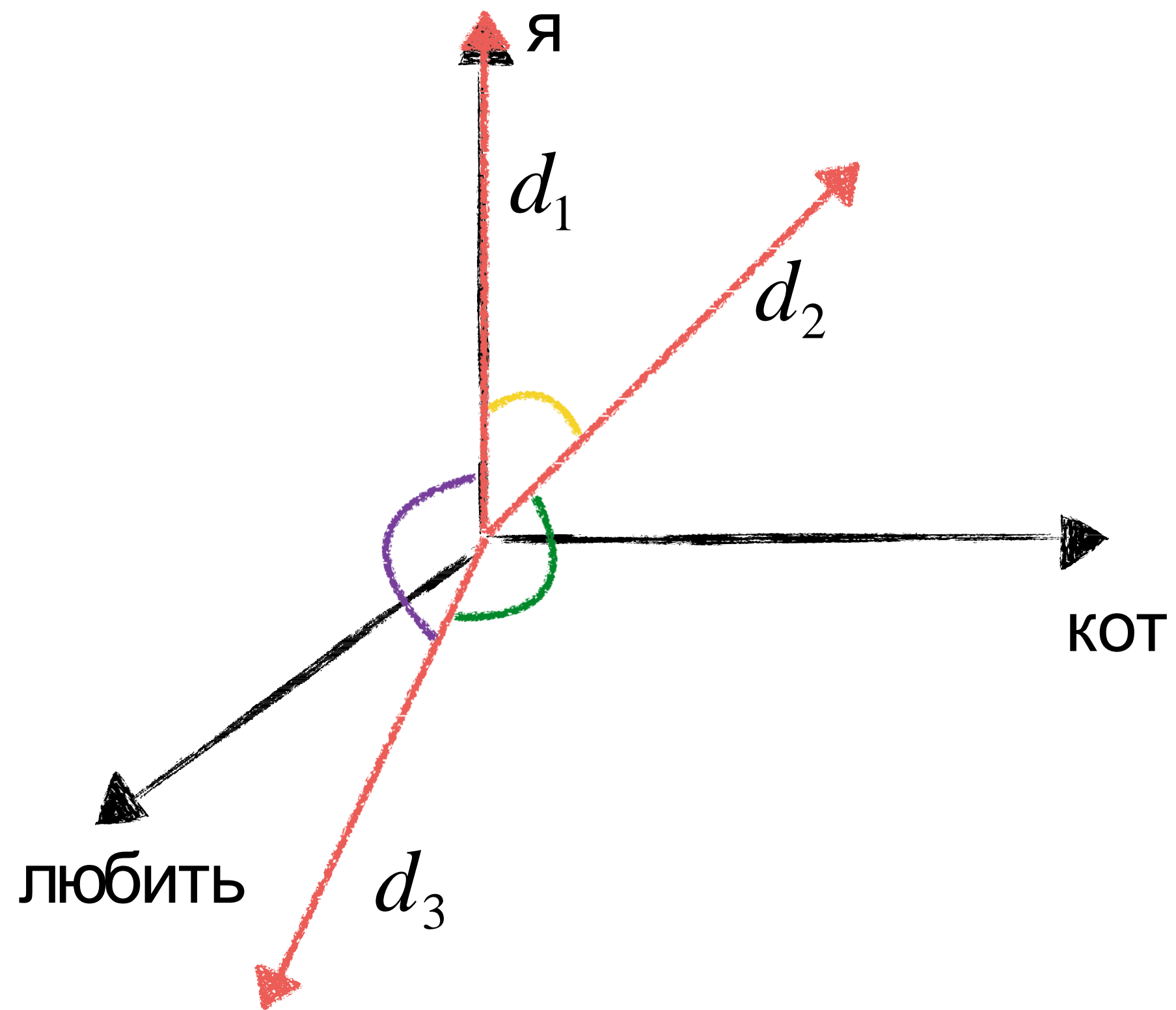
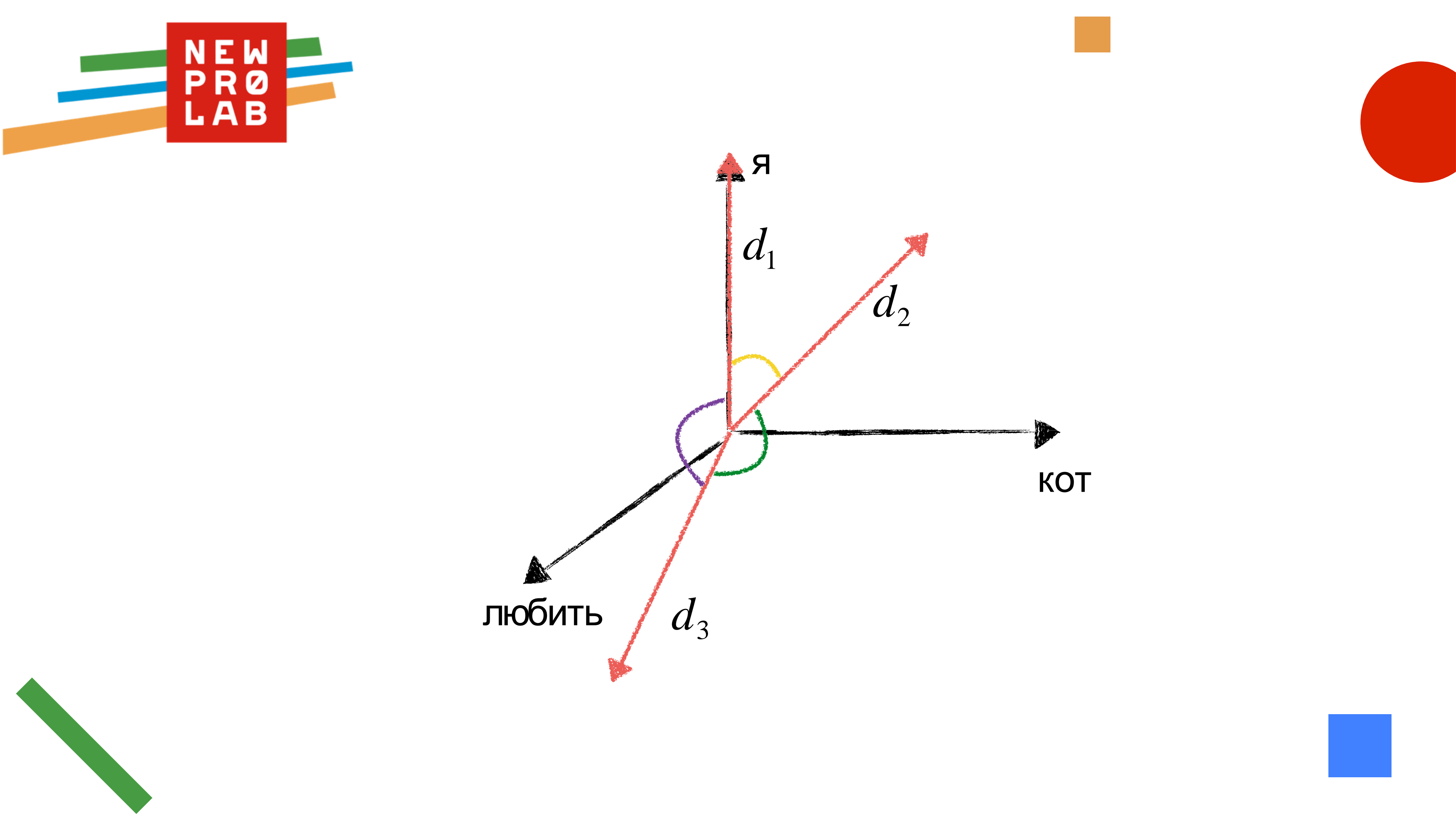
Коэффициент Жаккарда





NEW
PRO
LAB

Косинус




$$a \cdot b = ||a|| \cdot ||b|| \cos \theta$$



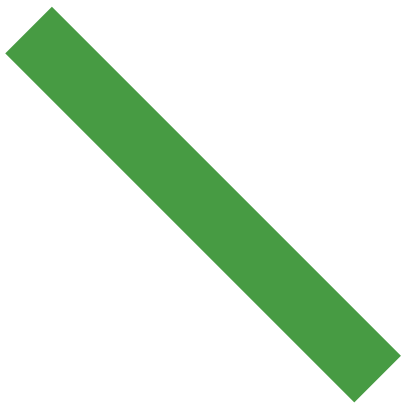
$$\cos \theta = \frac{a \cdot b}{||a|| \cdot ||b||}$$

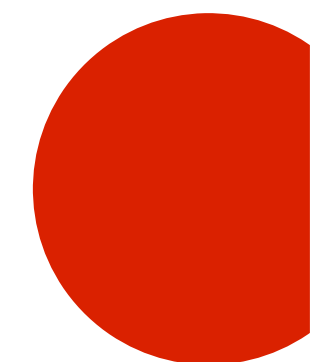
Свойства меры $\cos \theta$

- угол между векторами
- игнорирует элементы вне пересечения
- длина вектора

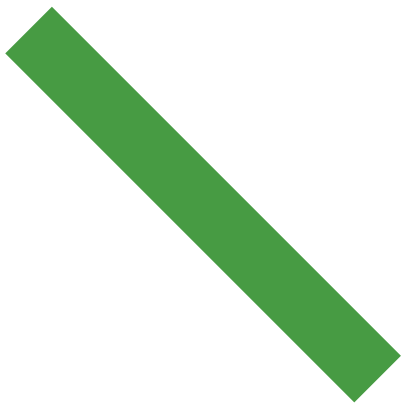


$$a \cdot b = ?$$





	Document	User	
политика	0	1	0
экономика	1	1	1
шоппинг	1	-1	-1
IT	0	1	0
футбол	0	1	0
романтика	1	-1	-1
семья	1	-1	-1
природа	0	0	0
юмор	0	0	0
Россия	1	1	1


$$\cos \theta = -0.16$$

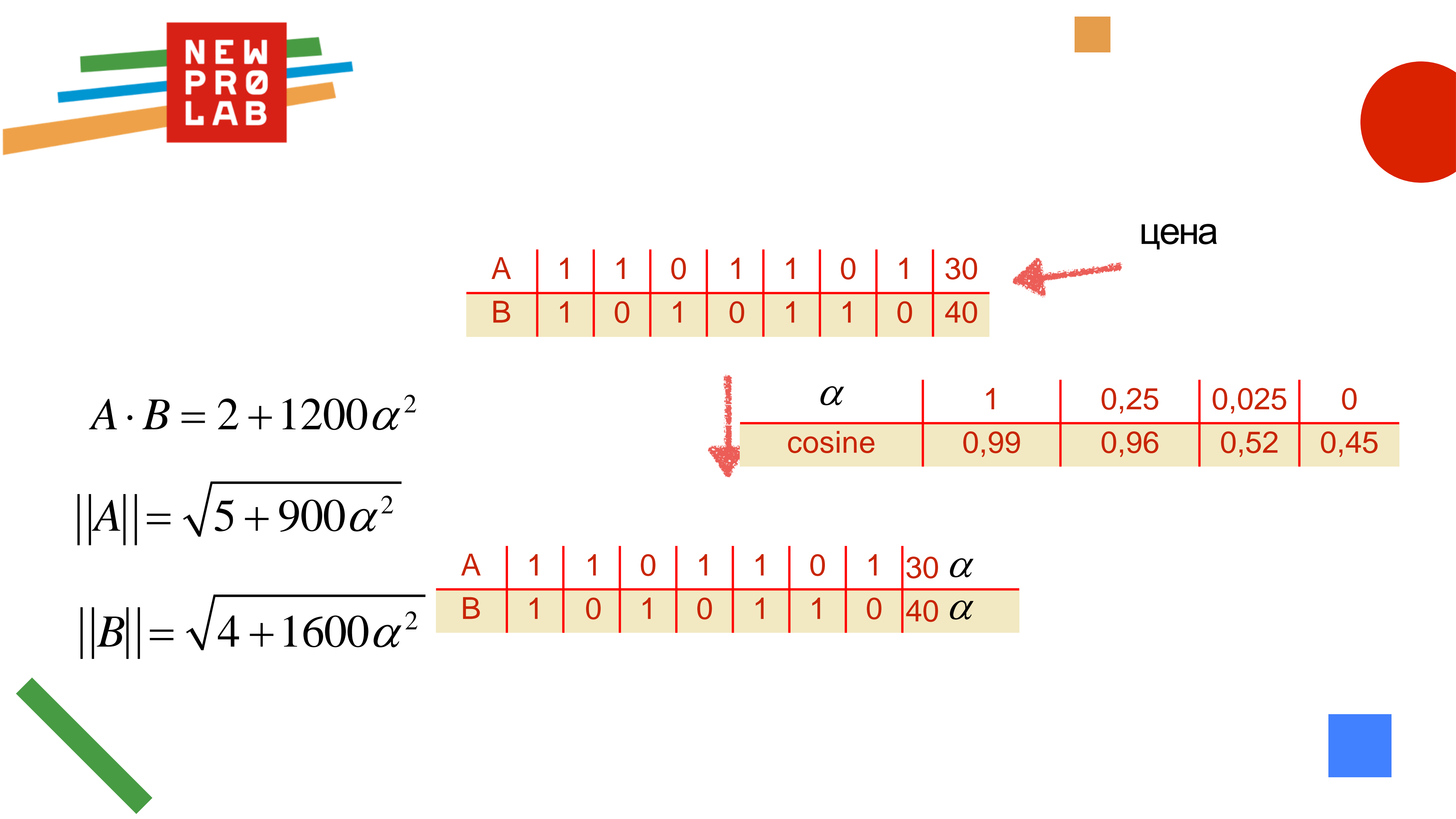
$$||b|| = 5 \quad ||a|| = 8$$

$$a \cdot b = -1$$





	A	B	C
скорость процессора	3.06	2.68	2.92
объем накопителя	500	320	640
размер оперативной памяти	6	4	6



NEW
PRO
LAB

цена

A	1	1	0	1	1	0	1	30
B	1	0	1	0	1	1	0	40



$$A \cdot B = 2 + 1200\alpha^2$$

$$||A|| = \sqrt{5 + 900\alpha^2}$$

$$||B|| = \sqrt{4 + 1600\alpha^2}$$



α	1	0,25	0,025	0
cosine	0,99	0,96	0,52	0,45

A	1	1	0	1	1	0	1	30	α
B	1	0	1	0	1	1	0	40	α

Поправочный коэффициент

- разделить на среднее
- вычесть среднее