# Recommender Systems

## Ing. Tomáš Řehořek

tomas.rehorek@fit.cvut.cz



INTELLIGENCE

ROUP

Computational Intelligence Group (CIG) **Department of Theoretical Computer Science** Faculty of Information Technology (FIT)

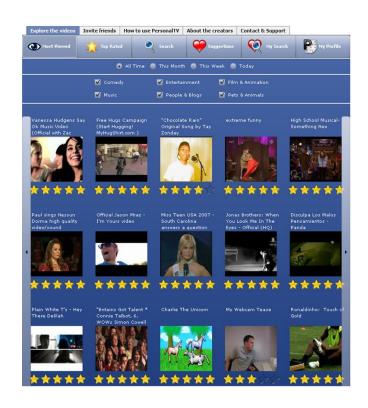
COMPUTATIONAL Czech Technical University (CTU) in Prague



Modgen, s.r.o.

# Recommender Systems: Quick Intro

- Software systems providing suggestions for items to be of use to a user
  - Based on her past behavior (item purchases, ratings with stars, views...)
  - Based on behavior of other users of the system



## **Content-based Recommendation**

- Uses meta-data about users/items in the system
- Attempts to recommend items similar to those liked by the user in the past

## **Collaborative Filtering**

- Most prominent family of algorithms in Recommender Systems
- Analyses purchase/rating history in the whole userbase, recommends items liked by similar users

# Collaborative Filtering

	Movie 1	Movie 2	Movie 3	 Movie <i>n</i>
User 1	2	?	?	?
User 2	?	5	4	?
User 3	4	?	1	3
÷				
User <i>m</i>			2	

## INPUT

- Set of *users*  $U = \{u_1, \dots, u_m\}$ ,
- Set of *items*  $I = \{i_1, ..., i_n\}$ ,
- Set of *ratings*  $R = \{r_1, \dots, r_k\}$ 
  - $r_i \in U \times I \times \mathbb{R}$  (explicit ratings) or
  - $r_i \in U \times I \times \{0,1\}$  (implicit ratings)
- Target user  $u_x \in U$

### OUTPUT

- Top-N recommendations
  - Set  $\{i_1^{rec}, \dots, i_N^{rec}\} \subset I$  of items that the user will most likely to appreciate
- Rating predictions
  - Predict ratings of unknown couples from  $U \times I$

# Our Research

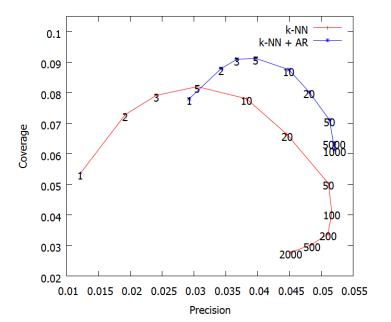
### **II. Matrix-Based Algorithms**

- User-based *k*-Nearest Neighbors
- Item-based *k*-Nearest Neighbors
- Matrix Factorization

UserID	ltem1	Item2	Item3	Item4	Item5	ltem6	Item7	• • •
1	?	3	?	?	4	1	3	
2	3	?	4	?	5	?	1	
3	?	2	5	5	3	1	?	
4	2	?	?	?	2	?	5	
5	3	?	?	5	4	?	2	
6	1	4	?	4	?	3	?	
:								•••

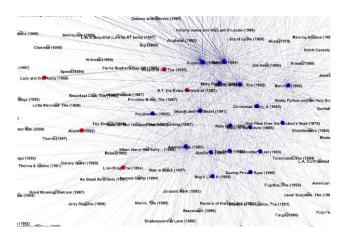
## I. Multi-Objective Optimization

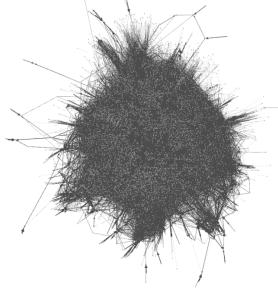
• Trade-offs between the accuracy and the diversity of items being recommender



### **III. Rule-Based and Graph-based Algorithms**

- Association Rules
- Sequential Patterns



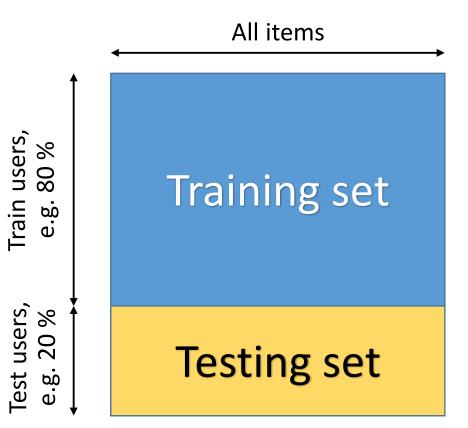


# Split Validation of a Recommender System

• To validate the Recommender, one may divide the **users** into:

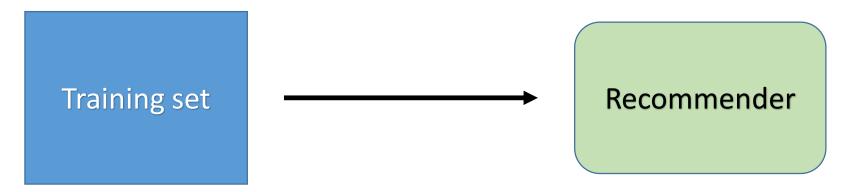
Train users,

- **Training set**, which is fully submitted to the Recommender
- **Testing set**, submitted only partially and used to evaluate the Recommender



# Split Validation of a Recommender System

• From the training set, all the items are submitted to the system



- From the **testing set**, for **each user**, the set of ratings in divided into:
  - Observation subset the ratings/purchases submitted to the system
  - Testing subset the ratings/purchases used to evaluate the system

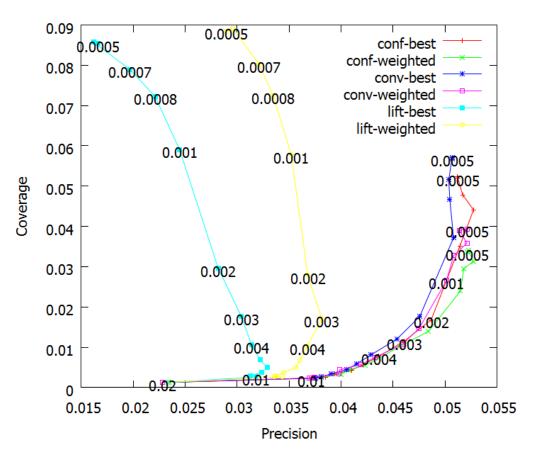


# Our Research: Multi-Objective Optimization

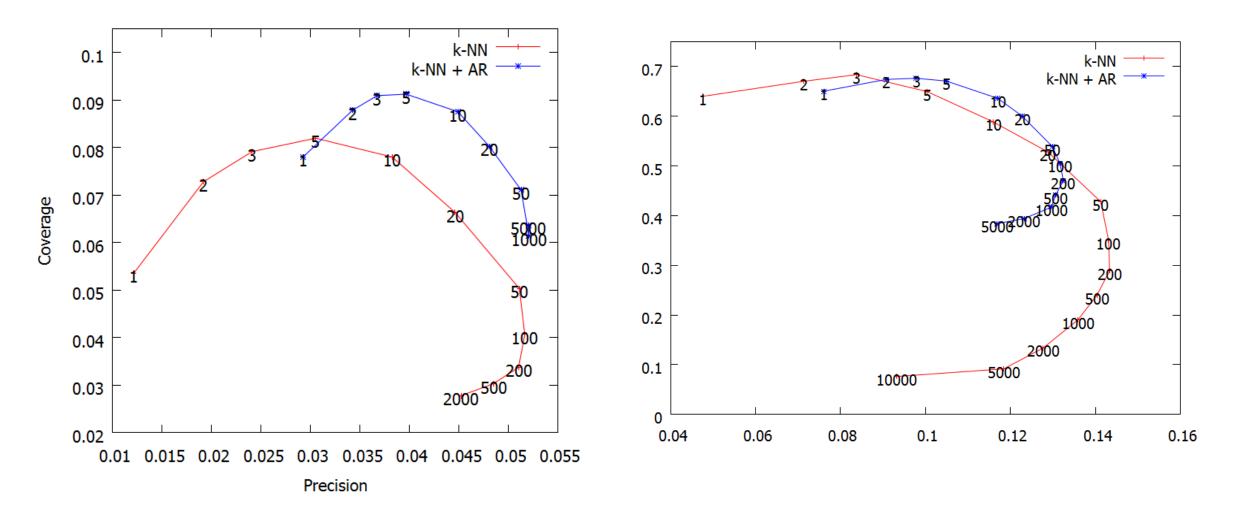
k-Nearest Neighbors

0.7 k-NN 0.6 10 20 0.5 50 Coverage 0.4 100 0.3 200 500 0.2 1000 2000 0.1 5000 10000 0 0.04 0.05 0.06 0.07 0.08 0.09 0.11 0.12 0.13 0.14 0.15 0.1 Precision

### **Association Rules**

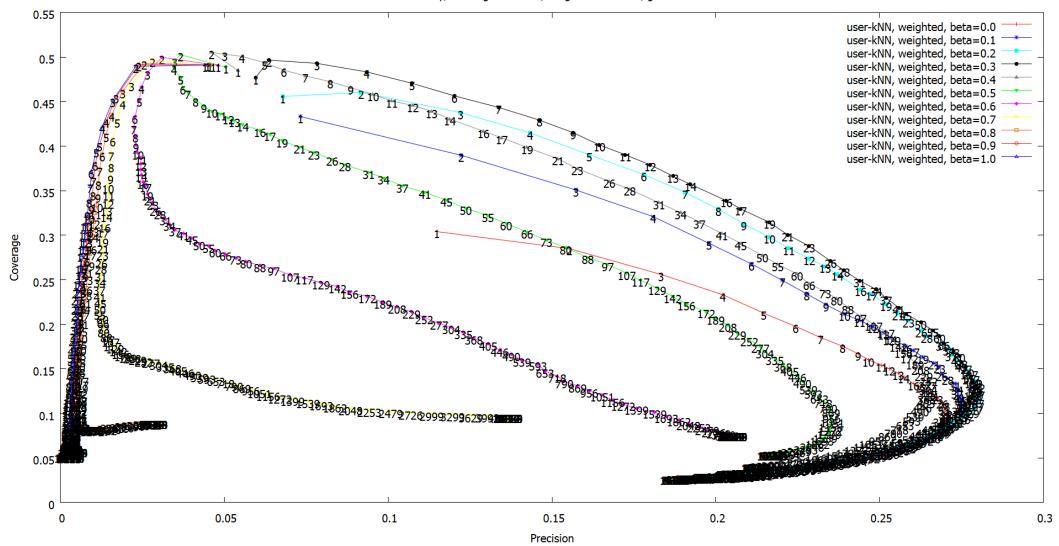


# **Our Research: Model Ensembles**

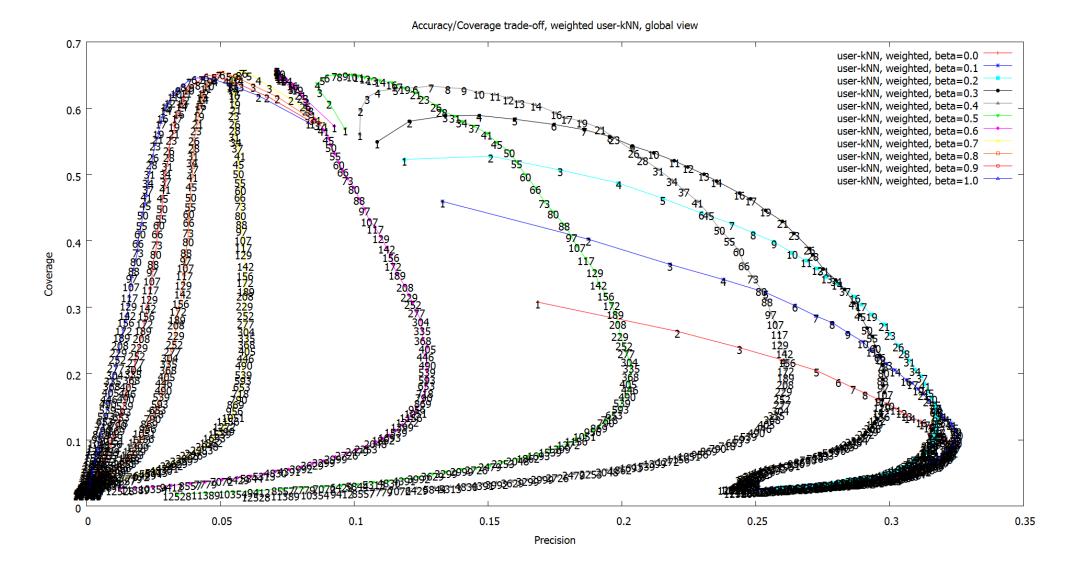


## Our Research: Bestseller Penalization (MovieLens1M)

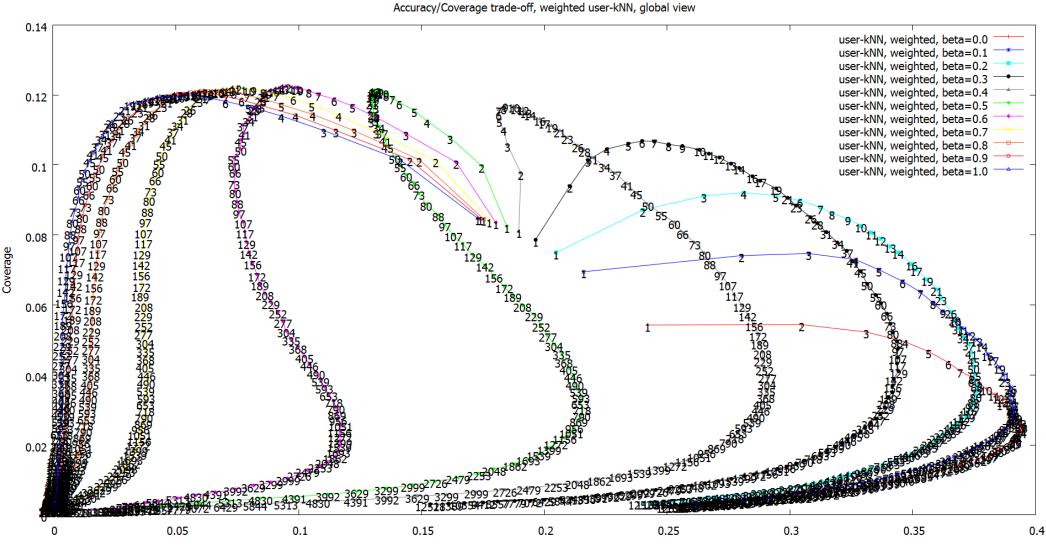
Accuracy/Coverage trade-off, weighted user-kNN, global view



## Our Research: Bestseller Penalization (MovieLens10M)

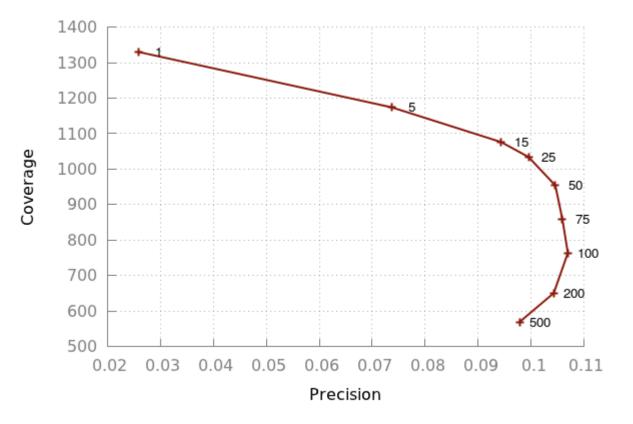


## Our Research: Bestseller Penalization (Libimseti.cz)



Precision

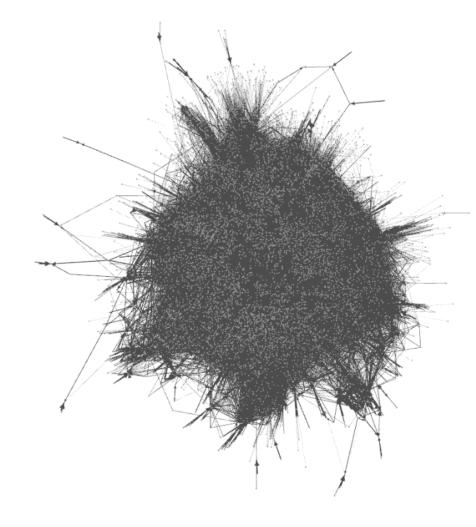
# **Business Impact?**



2. n25	<b>339</b> (8.79%)
3. n100	<b>209</b> (5.42%)
4. n500	<b>207</b> (5.37%)
5. n5	<b>120</b> (3.11%)
6. old	<b>99</b> (2.57%)

isthereanydeal.com

# Live demo: Visualizing Association Rules

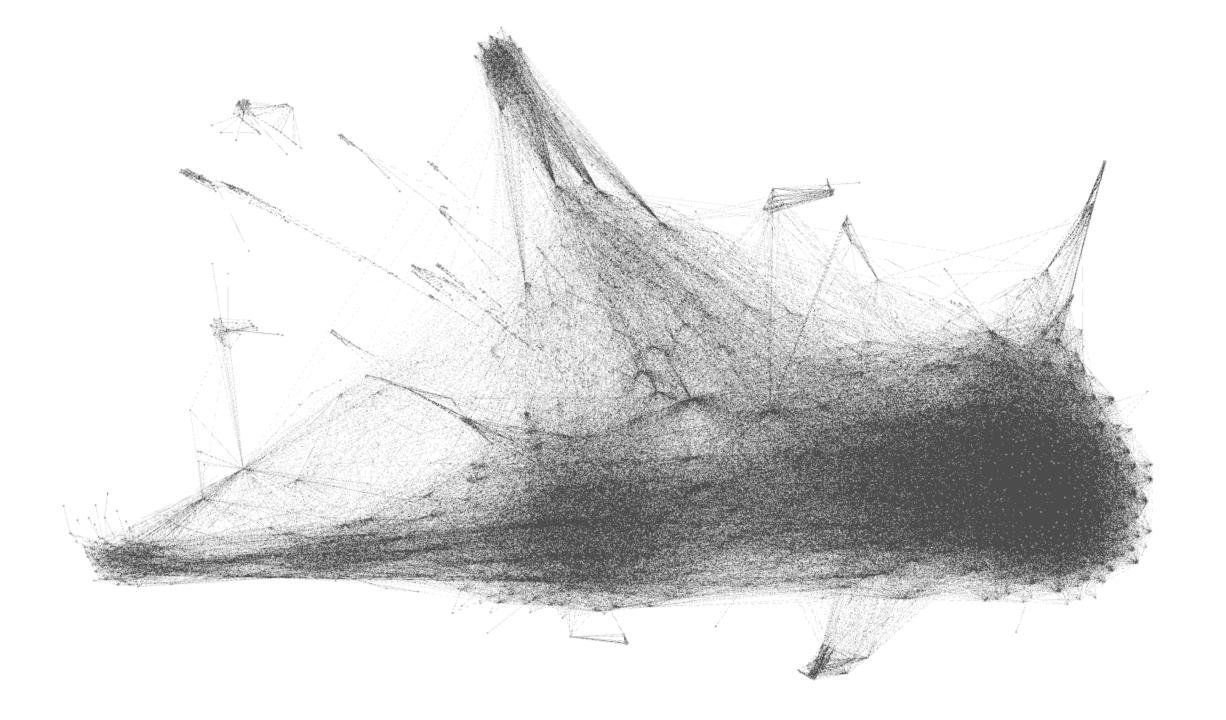


Indiana Jones and the Last Crusade (1989) tville (1998) Beetlejuide (1988) Ene Is Beautiful (La Vita AT bella) (1997) Raising Arizona (198 Stand by Me (1986) Airplanel (1980) Alien (1979) Big (1988 Clueless (1995 **Butch Cassidy** Witness (1985) 1980mbo (1941 Aliens (1986) Die Hard (1988) Ferris Bueller's Day Off (1986) 02 The (1939) (1997)Speed (1994) Lady and the Tramp (1955) E.T. the Extra-Terrestrial (1982) Breakfast Club, The 1985 Veapon (1987) Dogs (1992) Princess Bride, The (1987) Monty Python and the Holy Gra Little Mermaid, The (1989) Christmas Story, A (1983) Godfat Beauty and the Beast (1991) Pocahontas (1995) Toy Stopual 189 Red October Gond (Maddunting (1997) One Flew Over the Cuckoo's Nest (1975) ken Run (2000) Ghostbusters (1984) Aladd (1992) Titanic (1997) Blade (1990) When Harry Met Sally Apollo 13 7(39) (1985) hindler List (1993) Babe (1995 ape (1993) Terminator, The (1984) Thelma & Louise (1991) Galaxy Quest (1999) Lion King the (1994) Men in Black (1997) Saving Private Ryan (1998) Bud's Life A (1998) As Good As It Gets (1597 jest Gump (1994) American Fugitive, The (1993) Good Morning, Vietnam (1987) ow (1994) Jurassic Park (1993)

Dances with Wolves (1990)

w (1994) Usual Suspects, The (199 Jerry Maguire (1996) Matrix, The (1999) Raiders of the Lone Art of 1984)Lambs, The (1991) Braveheart (1995) Fargo (1996) Pulp Fic

≥ (1992)





- Recommendation as a cloud service
- Domain-independent
  - IPTV, VoD, Eshops, Cultural events, Sport facilities...
- Fast and scalable implementation of CF algorithms
  - Parallel Association Rules
  - Parallel k-NN
  - Parallel Matrix Factorization
- Real-time model updates
- Adjusting recommendation to fit business needs
  - Filtering, Boosting

# Live demo: Boosting and Filtering

#### https://modgen.net:8766/goout/

#### 'schedule' > 1386279637 and "komedie" not in 'tag'

#### User 86118115275

#### Items purchased

#### Count: 19

!itemId	Purchase time	!Total purchases	name	schedule	score	tag	type
event-163790	2013-11-21 17:47:29	690	Lucie, větší než malé množství lásky	1395615540.0	None	{'muzikal', 'divadlo', 'show'}	event
event-165498	2013-11-21 17:47:53	42	Strakonický dudák	1400194740.0	None	{'drama', 'divadlo'}	event
event-17737	2013-11-21 17:48:49	323	Vražda v salónním coupé	1386889140.0	None	{'top', 'komedie', 'divadlo'}	event
event-17749	2013-11-21 17:49:03	308	Drobečky z perníku	1392677940.0	None	{'divadlo', 'tragikomedie'}	event
event-17737	2013-11-21 17:50:47	323	Vražda v salónním coupé	1386889140.0	None	{'top', 'komedie', 'divadlo'}	event
event-5810	2013-11-21 17:51:03	131	39 stupňů	1386889140.0	None	{'drama', 'divadlo'}	event
event-212343	2013-11-21 17:51:31	8	Zahradní slavnost	1386889140.0	None	{'drama', 'divadlo'}	event
event-184929	2013-11-21 17:52:10	225	Jedenácté přikázání	1403737140.0	None	{'komedie', 'divadlo'}	event
event-184929	2013-11-21 17:53:47	225	Jedenácté přikázání	1403737140.0	None	{'komedie', 'divadlo'}	event
event-5810	2013-11-21 17:53:50	131	39 stupňů	1386889140.0	None	{'drama', 'divadlo'}	event
event-100659	2013-11-21 17:53:59	560	Kapka medu pro Verunku	1401663540.0	None	{'pro-deti', 'muzikal', 'pohadka', 'divadlo'	event
event-166156	2013-11-21 17:54:05	232	Sejdeme se Na Cibulce	1387321140.0	None	{'debata', 'divadlo', 'show'}	event
event-137165	2013-11-21 17:54:31	185	Růžové brýle	1391900340.0	None	{'drama', 'divadlo'}	event
event-134719	2013-11-21 17:54:47	21	Viny	1390258740.0	None	{'drama', 'divadlo', 'show'}	event
event-8363	2013-11-21 17:54:57	31	Příběhy ze Starého zákona II	1386889140.0	None	{'drama', 'divadlo', 'literatura'}	event
event-111031	2013-11-21 17:55:08	256	Aspects of Alice	1388444340.0	None	{'cerne', 'divadlo'}	event
event-140055	2013-11-21 17:55:35	155	Faust – Muž mezi Bohem a Ďáblem	1393109940.0	None	{'cerne', 'divadlo'}	event
event-209912	2013-11-21 17:56:41	68	Strauss Gala	1386889140.0	None	{'klasicka', 'koncerty'}	event
erformer-15694	2013-11-21 17:56:41	174	Symfonický orchestr Českého rozhlasu	None	0	{'klasicka', 'kapely'}	performe

Booster

Filter

Items recommended

if "drama" in 'tag' then 2 else 1

Submit

#### Result

Count: 10

Time: 155.82 ms

!itemId	!Total purchases	name	schedule	score	tag	type
event-176612	105	Dva	1390863540.0	None	{'drama', 'divadlo'}	event
event-5147	47	César a Drana	1387234740.0	None	{'drama', 'divadlo'}	event
event-5368	30	Šoa	1390517940.0	None	{'drama', 'divadlo'}	event
event-11271	30	Polední úděl	1390863540.0	None	{'drama', 'divadlo'}	event
event-198430	81	Vánoční koncert Strauss gala	1386889140.0	None	{'klasicka', 'koncerty'}	event
event-108968	40	Amerikana III	1400799540.0	None	{'balet', 'tanec', 'divadlo'}	event
event-142810	113	Babička	1390690740.0	None	{'drama', 'divadlo'}	event
event-202294	29	Ateliér Open Space 1	1387493940.0	None	{'workshop', 'divadlo'}	event
event-8444	66	Carmen	1399330740.0	None	{'opera', 'divadlo'}	event
event-30166	42	Česká mše vánoční – Dětská opera Praha	1387666740.0	None	{'klasicka', 'divadlo', 'koncerty'}	event

## Parallel Computing Challanges in Recommender Systems

- Online Recommender Systems must satisfy many tough requirements:
  - Handling click-streams of millions of users in real-time
  - 100 ms to build recommendations based on database of 1.000.000 users?
  - Datasets for online portals are as large as 250 GB of compressed rating data!

## Sample problems

### **Find Nearest Neighbors**

Given vector  $u \in \mathbb{R}^n$  and set  $U = \{u'_1, \dots, u'_m\}$ , find subset of k vectors from U which are most similar to u

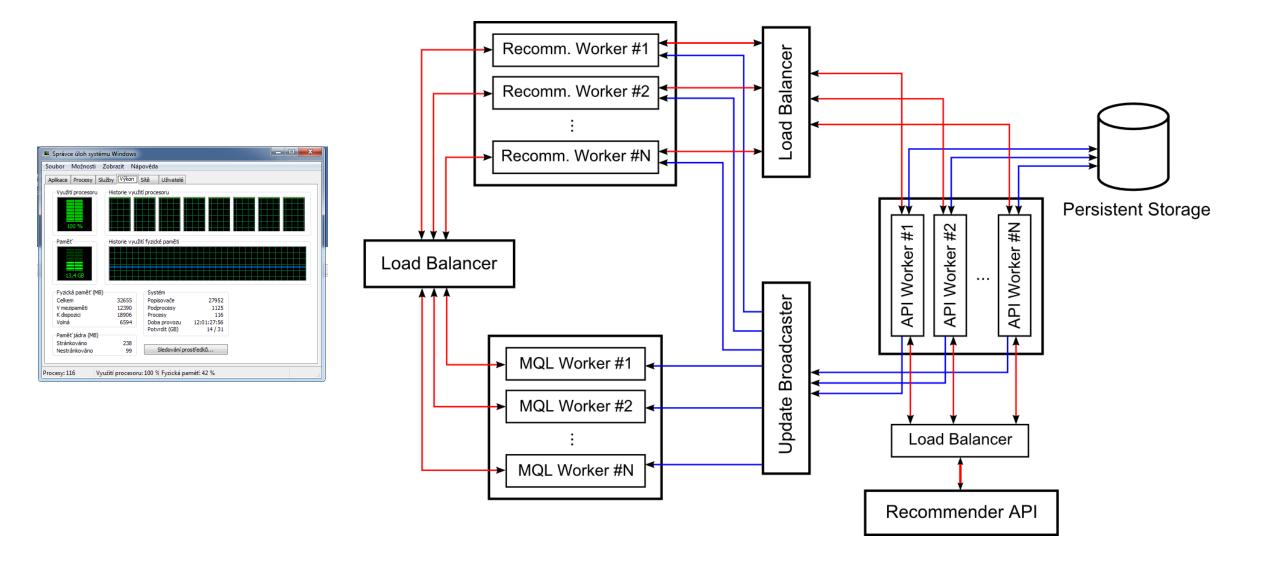
### **Find Frequent Itemsets**

Given sets 
$$I = \{i_1, ..., i_n\}$$
 and  $U = \{u_1, ..., u_m\}$ ,  
 $u_i \subseteq I$ , find set  $X \subseteq 2^I$  such that  $x \in X$  iff  

$$\frac{|\{u \in U | x \subseteq u\}|}{|I|} \ge s_{min}$$
for some  $s_{min} \in [0,1]$ 

- for  $n, m = 10^6$  or even more (Youtube, last.fm)
- within small, fixed amount of time (several 10s of milliseconds?)

# Parallel Computing: Modgen Architecture



# Thank you for your attention!

Ing. Tomáš Řehořek tomas.rehorek@modgen.net