

# ODPORÚČANIE **V PRAXI**

PETER GAŠPAR  
6.3.2019

# Odporúčanie

## **Motivácia:**

odstrániť preťaženie množstvom informácií

## **Úloha:**

*personalizovaný* výber relevantných položiek pre používateľa

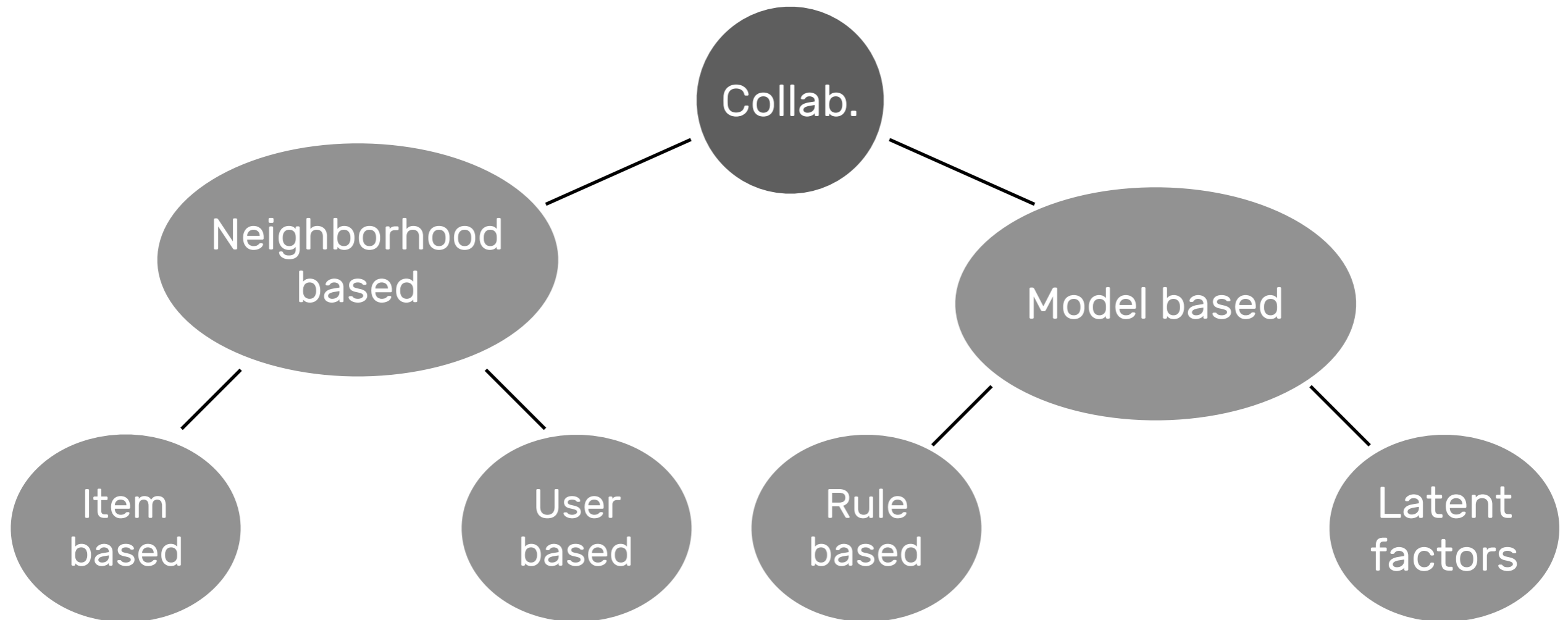
## **Ciele:**

zníženie zaťaženia používateľa,

zlepšenie používateľského zážitku,

zvýšenie predaja, ...

# Základné prístupy



**Matrix Factorization**  
Neural Networks

# História

**GroupLens** - výskumná skupina (est. 1992)

**Amazon** - odporúčanie v e-commerce

- vynález item-based collaborative filtering (1998)<sup>1,2,3</sup>

**Netflix** - Netflix Prize (2006)

- 100 miliónový dataset hodnotenia filmov,
- potreba zlepšiť RMSE o 10%, výhra 1 milión US\$,
- pokračovanie sa nekonalo.

1) <https://dl.acm.org/citation.cfm?doid=371920.372071>

2) <https://ieeexplore.ieee.org/document/1167344>

3) <https://www.google.com/patents/US6266649>

	Item 1	Item 2	Item 3	Item 4	Item 5
User 1	0	3	0	3	0
User 2	4	0	0	2	0
User 3	0	0	3	0	0
User 4	3	0	4	0	3
User 5	4	3	0	4	0

# Odporúčanie stretneme najmä v e-shopoch

## Mohlo by sa vám páčiť



8,9 / 10 NOVÉ

**LAST MINUTE: Jarná dovolenka  
s wellness a polpenziou**

📍 Ružomberok

**109 €**

> Zobrazíť viac



8,8 / 10 Platí do 20.12.2019

**Výhľad na Tatry, wellness a 2 deti  
zdarma v hotel Tulipán\*\*\***

📍 Vysoké Tatry - Tatranská Lomnica

**od 139 €**

> Zobrazíť viac



**Kurz na klavír, organ alebo kurz spevu**

📍 Trnava a Bratislava

**15 €**

> Zobrazíť viac



# Ako v praxi?

Odporúčanie poznáme najmä z online obchodov.

Dnes si ale ukážeme zaujímavé prípady použitia aj v iných doménach:

- šport,
- reality,
- ubytovanie,
- zdravotníctvo.

# Šport - stávkovanie

Odporúčanie pri tipovaní futbalových zápasov.

- Na ktorý zápas vsadiť?
- Koľko tipovať? (z pohľadu financií)
- Na ktorý tím stavať? (predpoveď výhry)

**Výber črt:** OneR, InfoGain, GainRatio a korelácia

**Významné črty:** nedávne výsledky zápasov, rozdiely v počte gólov, H2H hry



# Šport - stávkovanie

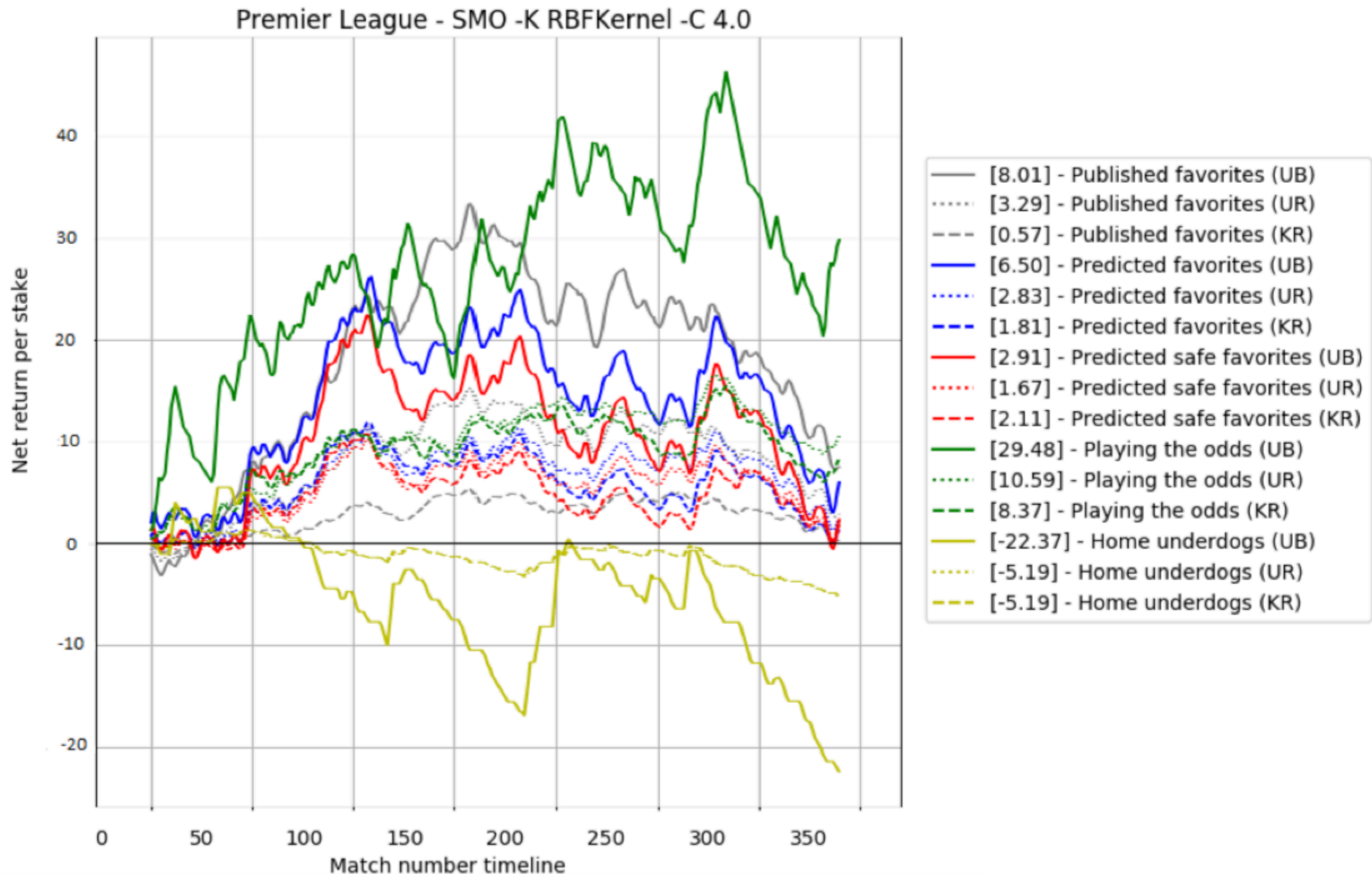
## **Predikcia tímu, ktorý zvíťazí:**

- najlepšie RandomForest (54,37% úspešnosť),
- za predpokladu, že vždy vykonáme stávku na všetko.

## **Odporúčanie:**

- výber *vhodných* stávok,
- využíva viaceré stratégie tipovania (závisí od používateľa),
- overenie simuláciou.

# Šport - stávkovanie



# Šport - stávkovanie

League	Accuracy	Kernel	C	Reduction technique
LaLiga	55.30%	linear	0.5	PCA
Premier League	60.09%	radial	1	GainRatio/Corr.
Serie A	57.83%	linear	0.125	None
Bundesliga	51.87%	sigmoid	4.0	None
Pro League	49.52%	radial	2.0	GainRatio

**Table 3: The LibSVM parameters with the highest accuracy.**

## Problémy:

- nemožnosť použiť krížovú validáciu pre overenie,
- každý chce tipnúť správne.

# Odporúčanie domov

Personalizované odporúčanie domov je náročné.

Každý produkt je predaný len **raz**.

Priemerný Američan vyberá dom **4.3 mesiaca**.

42%

Ľudí prvýkrát  
kupuje dom

78%

absolvuje súkromnú  
prehliadku domu

79%

kupuje dom  
online

43%

navštívi "otvorený"  
dom

# Odporúčanie domov

## Zillow.com (USA)

- 110 miliónov domov
- 1.2 mil. aktívnych ponúk

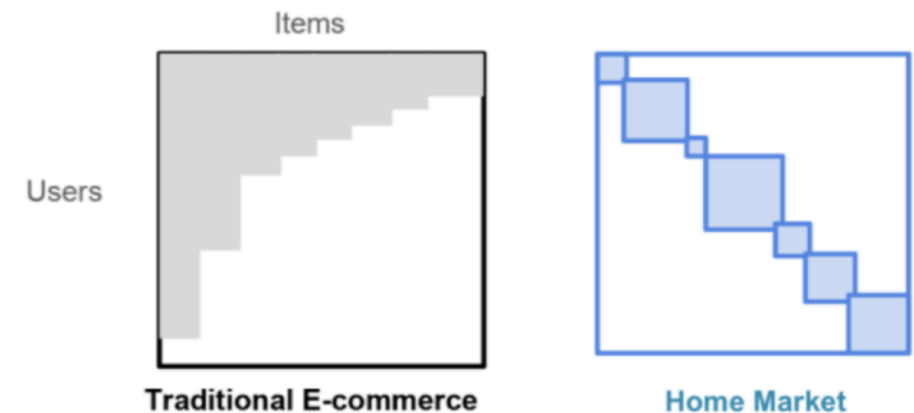


Figure 3. Illustrative depiction of the User-Item matrix for traditional e-commerce vs. home shopping.

## Hybridný prístup:

- kolaboratívny prístup (pre dlhšie zverejnené domy),
- obsahový prístup (pre nové domy),
- záver: kombinácia výsledkov z oboch odporúčačov.

# Odporúčanie domov

## Kolaboratívny prístup:

- Implicit Matrix Factorization for collaborative filtering,
- aplikujú pritom na každý región zvlášť,
- signál implicitnej spätnej väzby: čas prítomnosti na stránke s detailom domu.

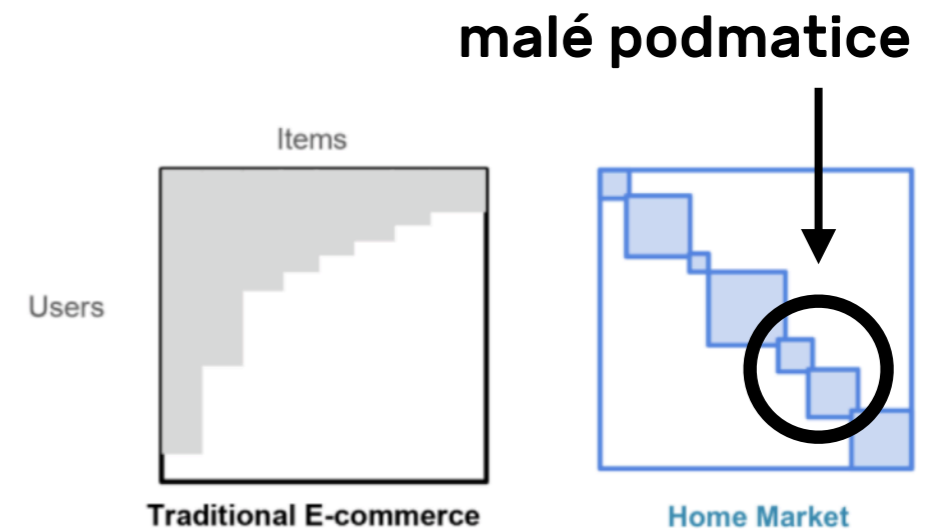


Figure 3. Illustrative depiction of the User-Item matrix for traditional e-commerce vs. home shopping.

# Odporúčanie domov

## Obsahový prístup:

- využíva profil používateľa,
- na základe neho získa najvhodnejšie domy,
- pre obsahové odporúčanie sa využíva Gradient Boosted Decision Tree (GBDT).

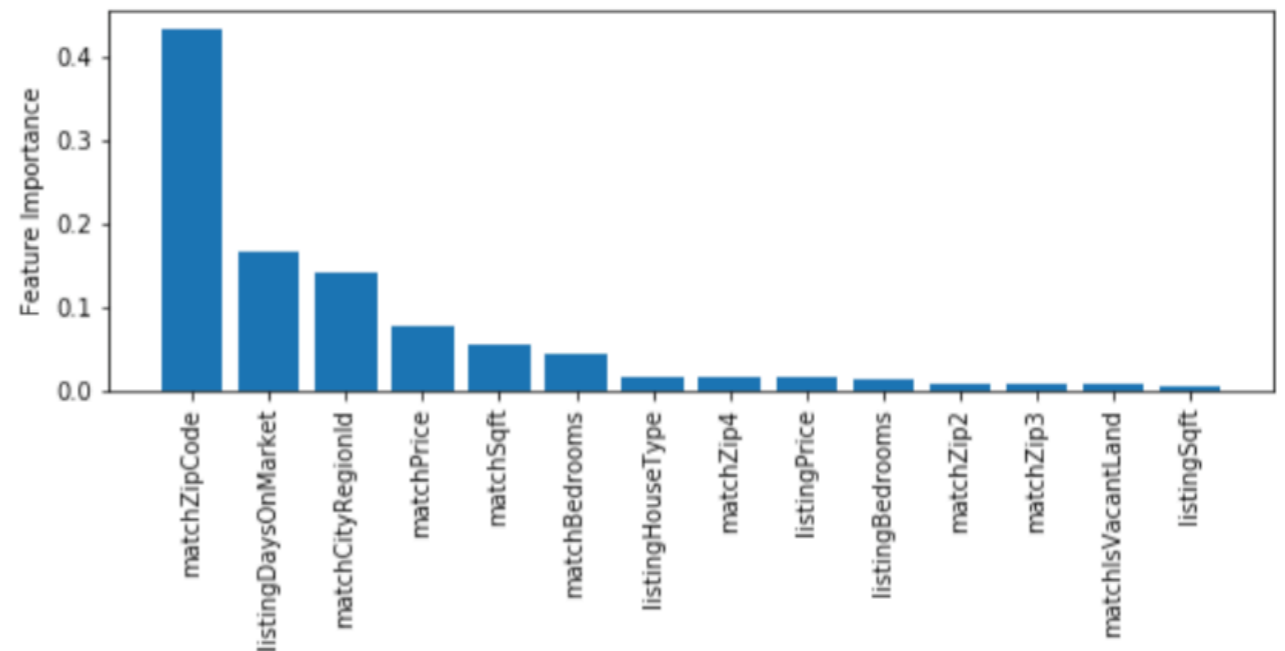


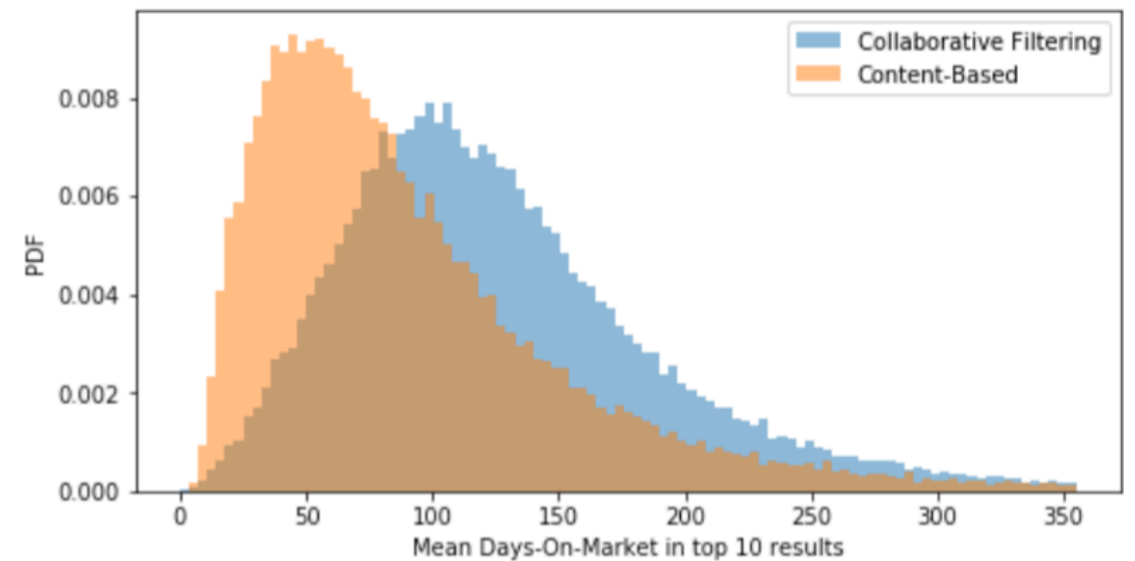
Figure 6. GBDT content model feature importance

Najdôležitejšie črty: PSČ, počet dní zverejnenia, región, cena.

# Odporúčanie domov

Metric	Relative Uplift
Home click-through rate	+14.30%
Home agent contact rate	+17.20%
Home save rate	+18.60%
Home share rate	+13.70%

**Table 1. Impact of home recommendations on website metrics.**



**Figure 7. Mean days on the market for content-based and collaborative filtering models.**



# Personalizácia recenzií

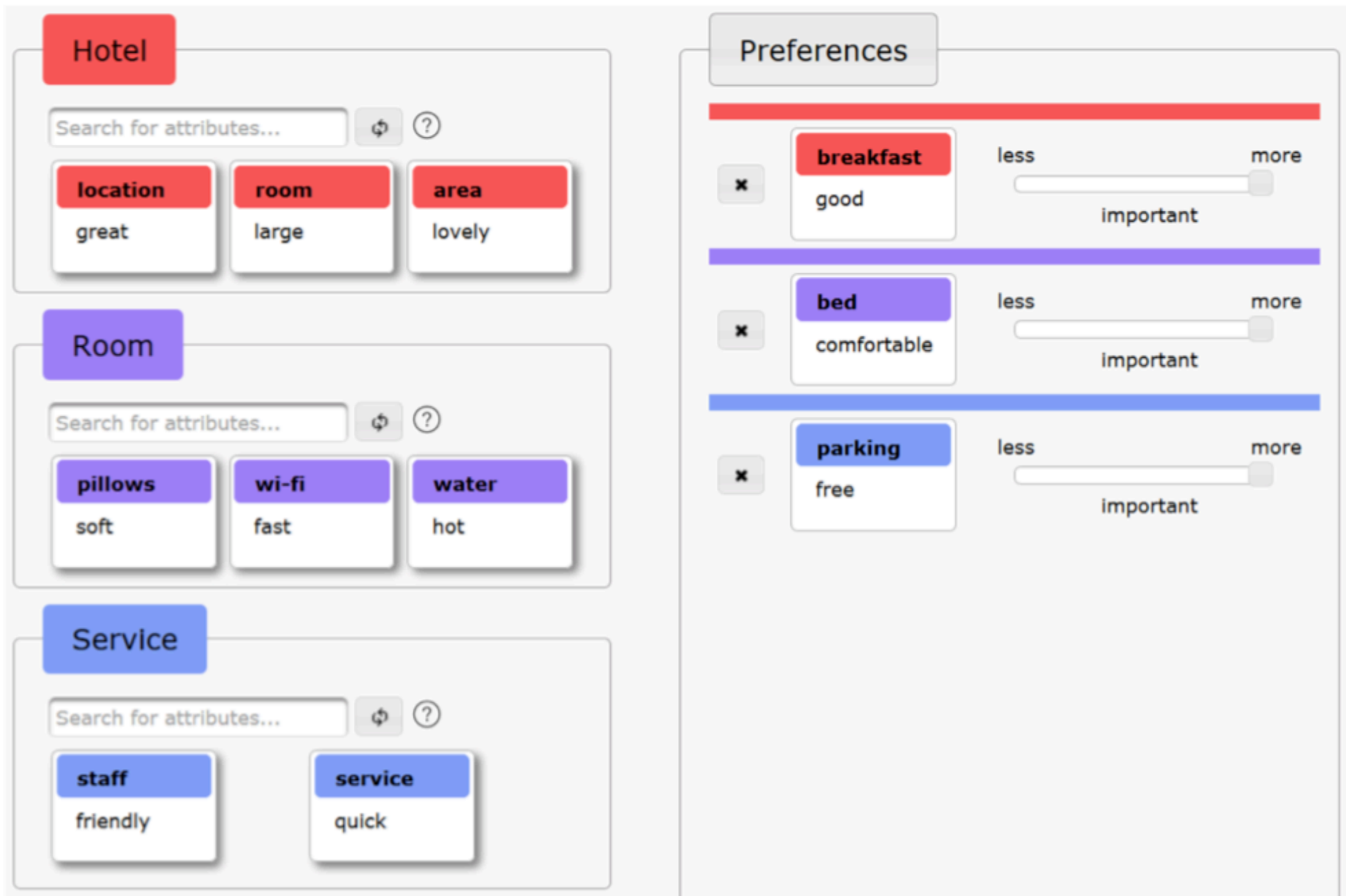
## Čím sa zaoberali:

- Personalizované odporúčanie recenzií hotelov.
- Vizualizácia recenzií, pre zlepšenie rozhodovania používateľa.
- Preskúmanie mechanizmov, ktoré umožňujú používateľom kontrolovať vizualizáciu.

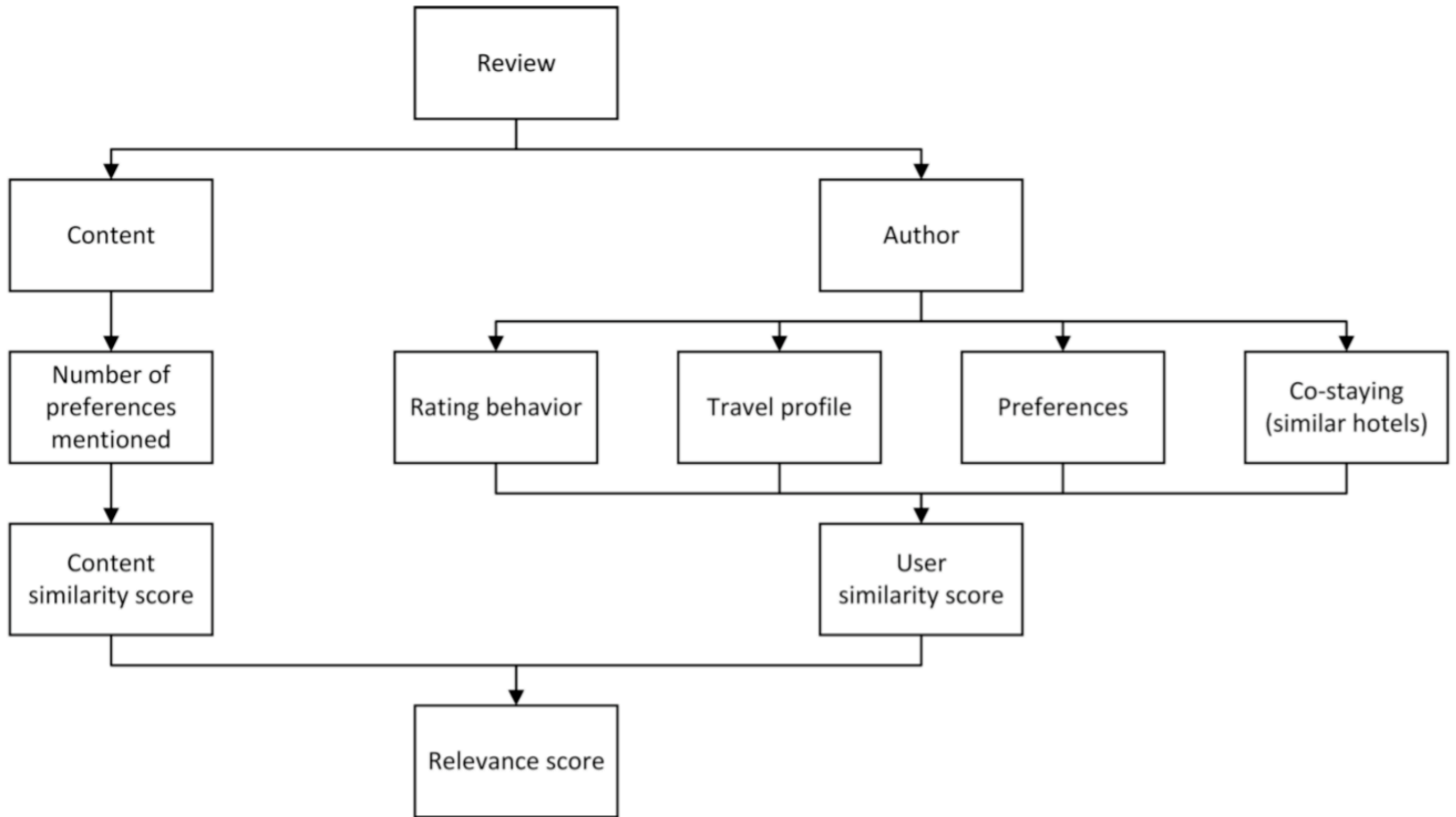
# Personalizácia recenzií

## Personalizované odporúčanie recenzií hotelov

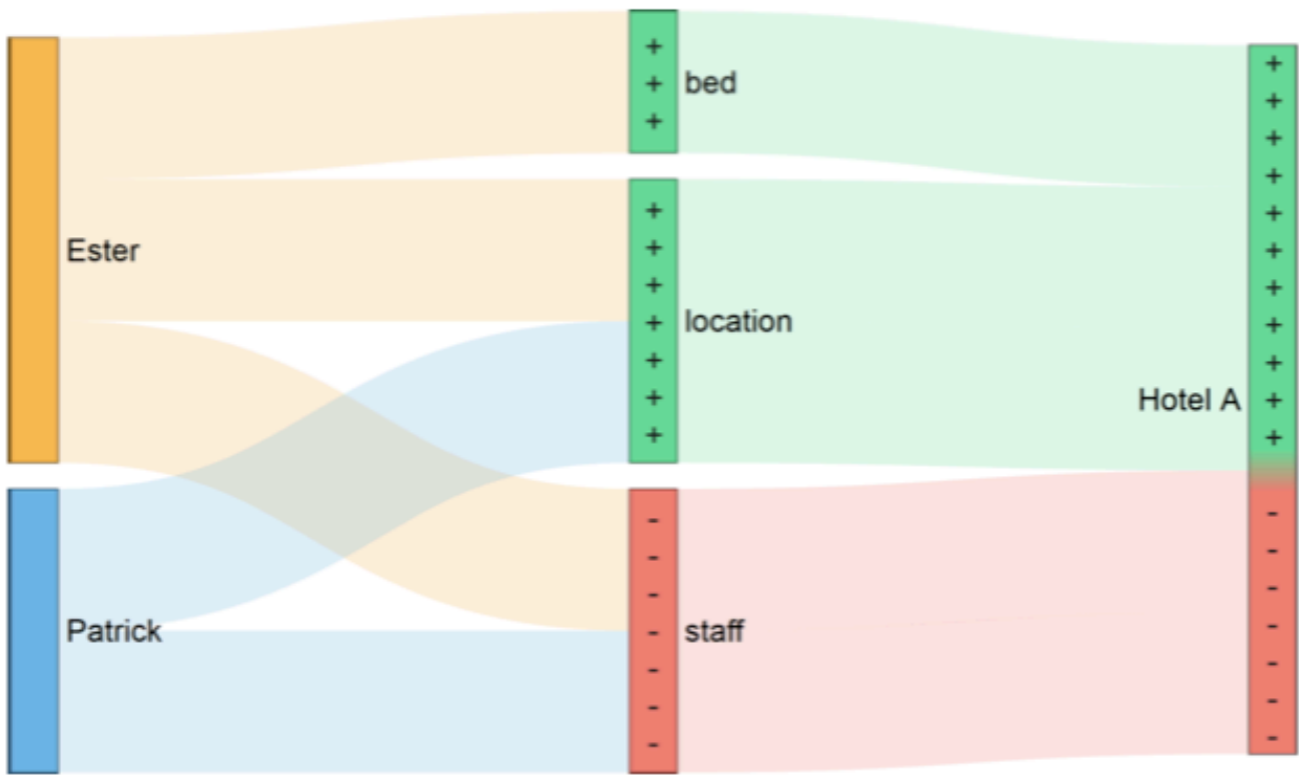
- prioritizujeme recenzie, ktoré sa zaoberajú tým, čo **používateľa zaujíma** (napr. mäkkosť matracov),
- prioritizujeme recenzie od **ľudí, ktorí sa podobajú** na používateľa (napr. biznismen).
  - *Vzniká implicitná sociálna sieť*



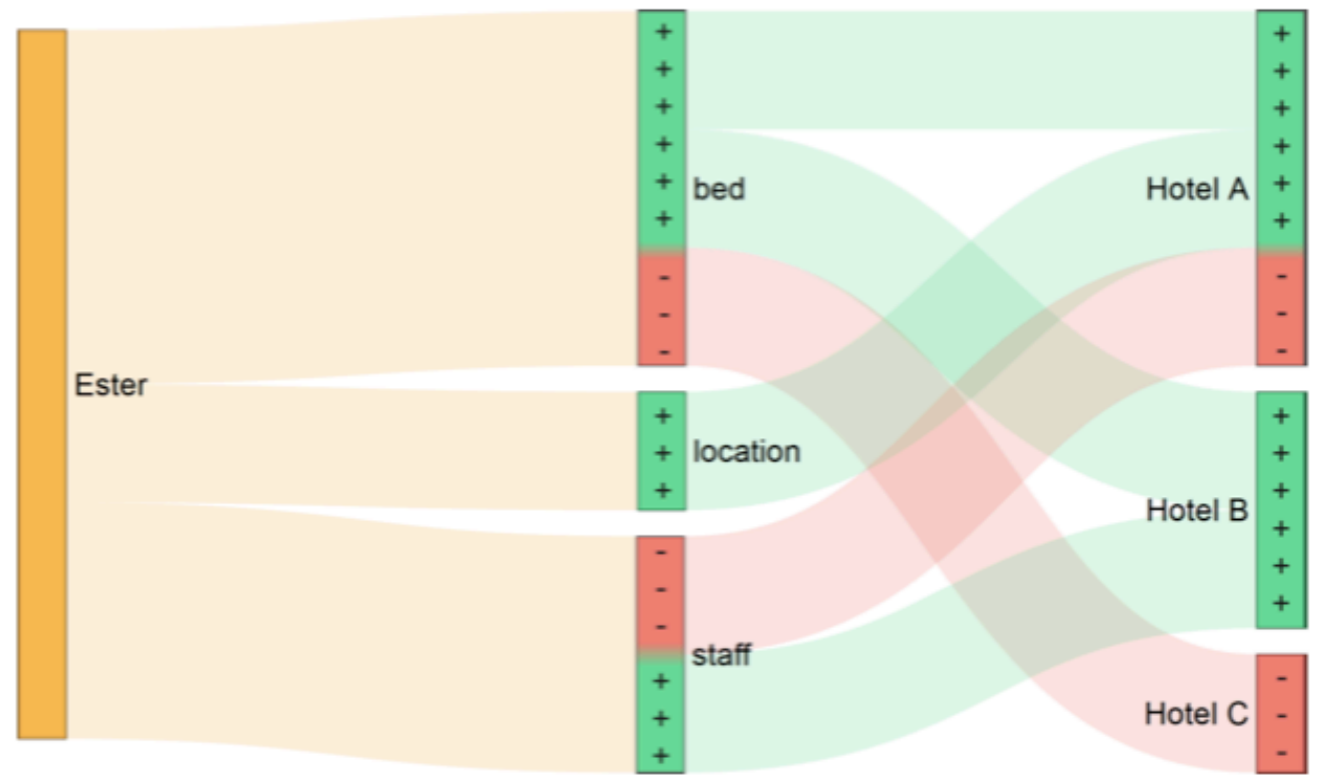
**Získavanie preferencií používateľa.** Používatelia môžu chytiť-a-pustiť relevantné témy z kategórií naľavo do preferencií napravo. Posuvníky sa využívajú na prispôsobenie dôležitosti každého atribútu.



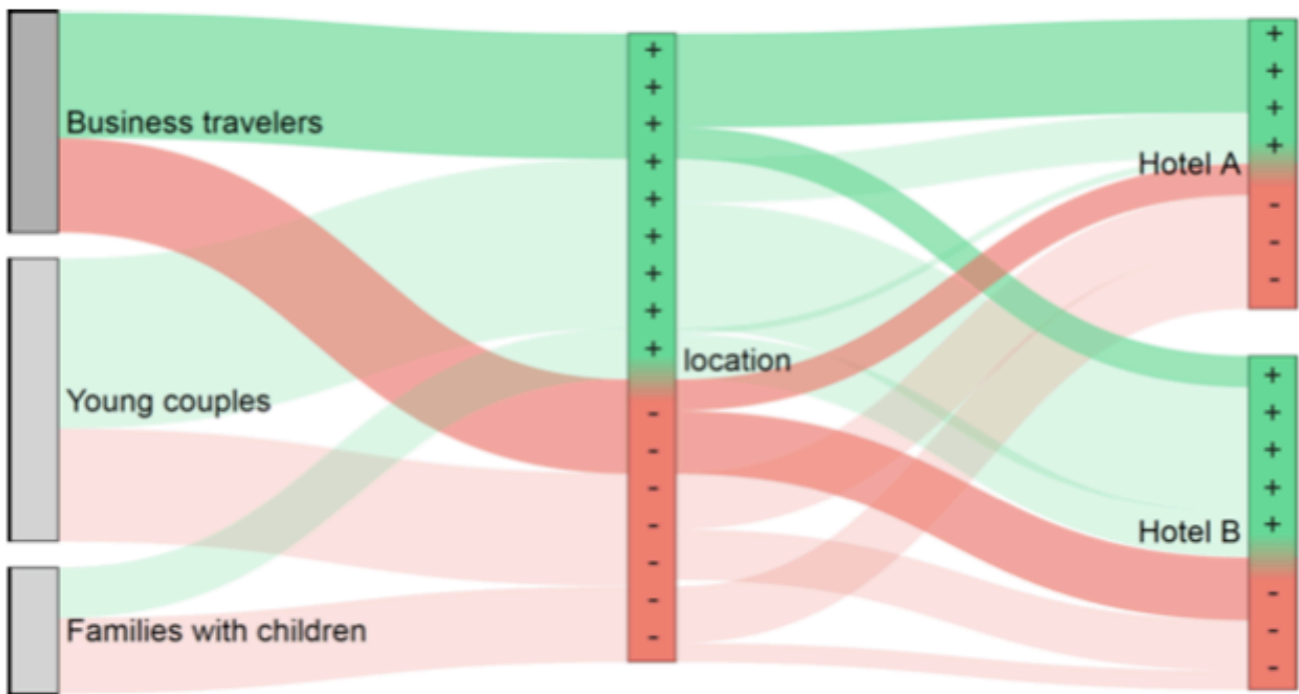
**Výpočet skóre relevancie pre recenziu.** Recenzia sa zobrazuje pri odporúčanom hoteli.



(a)



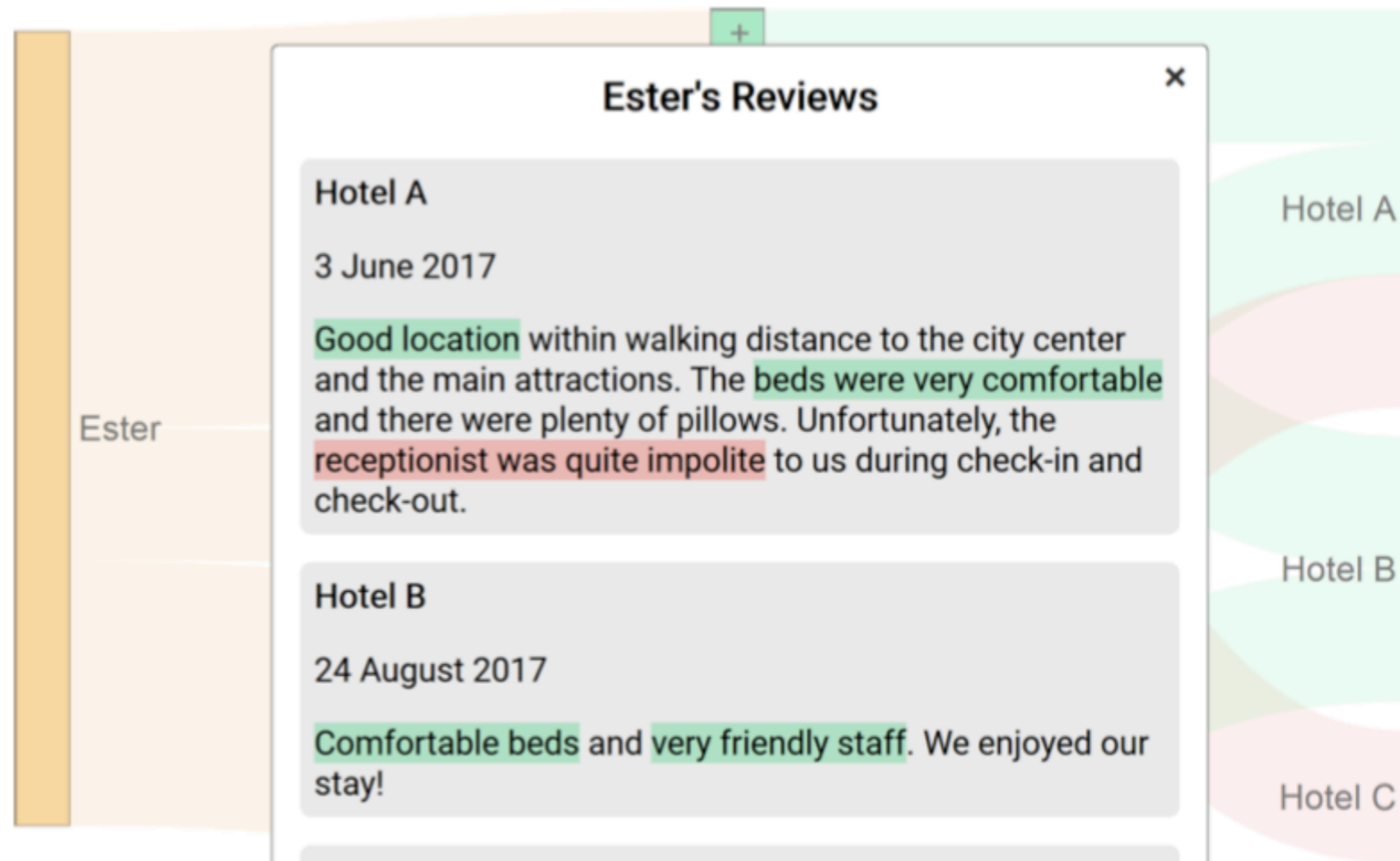
(b)



(c)



(d)



# Personalizovaná kontrola symptómov

Vyhľadávanie sa využíva na odhalenie ochorení pred tým, než navštívite odborníka.

1% dopytov v Googli smerujú na symptómy (Pinchin. 2016)

**Hlavný problém:** dokáže nahradiť lekára?

**Možné riešenie:** Kontrolovanie symptónov cez špecializované systémy

Pinchin: Personalized symptom checker using medical claims (2016)  
<https://blog.google/products/search/im-feeling-yucky-searching-for-symptoms/>

Kafle et al.: Personalized symptom checker using medical claims (2018)  
[http://ceur-ws.org/Vol-2216/healthRecSys18\\_paper\\_1.pdf](http://ceur-ws.org/Vol-2216/healthRecSys18_paper_1.pdf)

# Personalizovaná kontrola symptómov

## Problémy súčasných systémov:

- vyžadujú zadanie symptómov od používateľa (bez kontextu),
- niektoré sa s ním snažia interagovať a získavať ďalšie informácie,
- potreba manuálnej anotácie bázy znalostí,
- nezohľadňujú viaceré bázy znalostí.



# Personalizovaná kontrola symptómov

Autori navrhli systém na kontrolovanie príznakov.

Využívajú a kombinujú overené lekárske zdroje<sup>1</sup>.

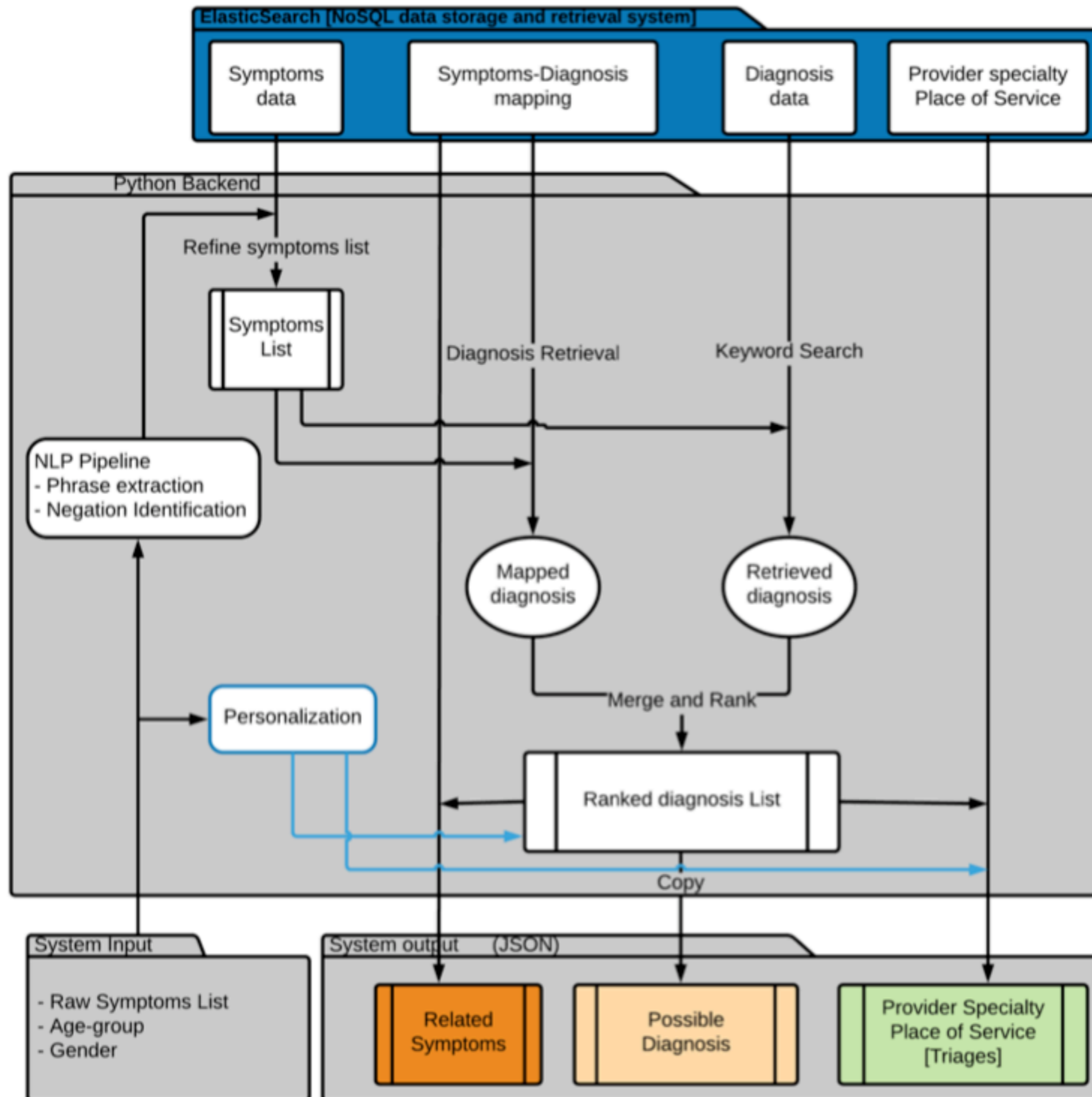
Automatizované vytvorenie BZ vyladené medikmi.

Systém vyžaduje od používateľa zadanie symptómov, ale autori sa snažia aplikovať rôzne NLP prístupy na redukciu šumu textu.

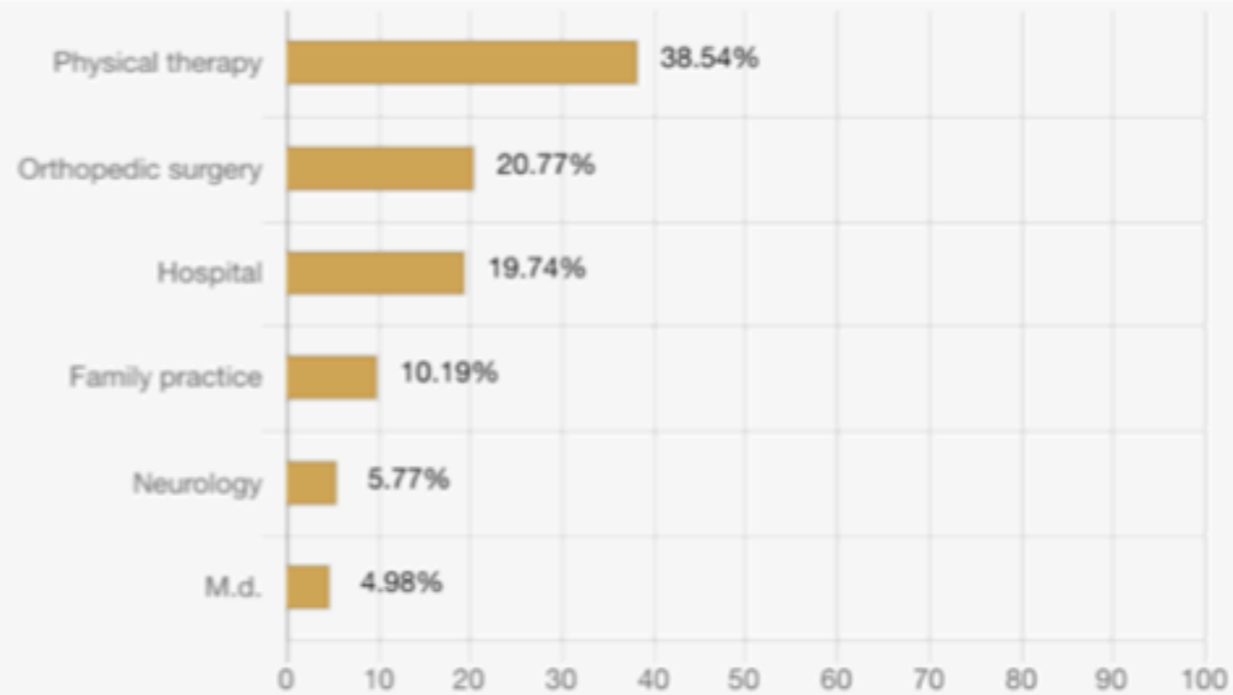
# Personalizovaná kontrola symptómov

Na základe symptómov sa snažia určiť diagnostiku, pričom využívajú historické lekárske záznamy používateľa - **personalizácia**.

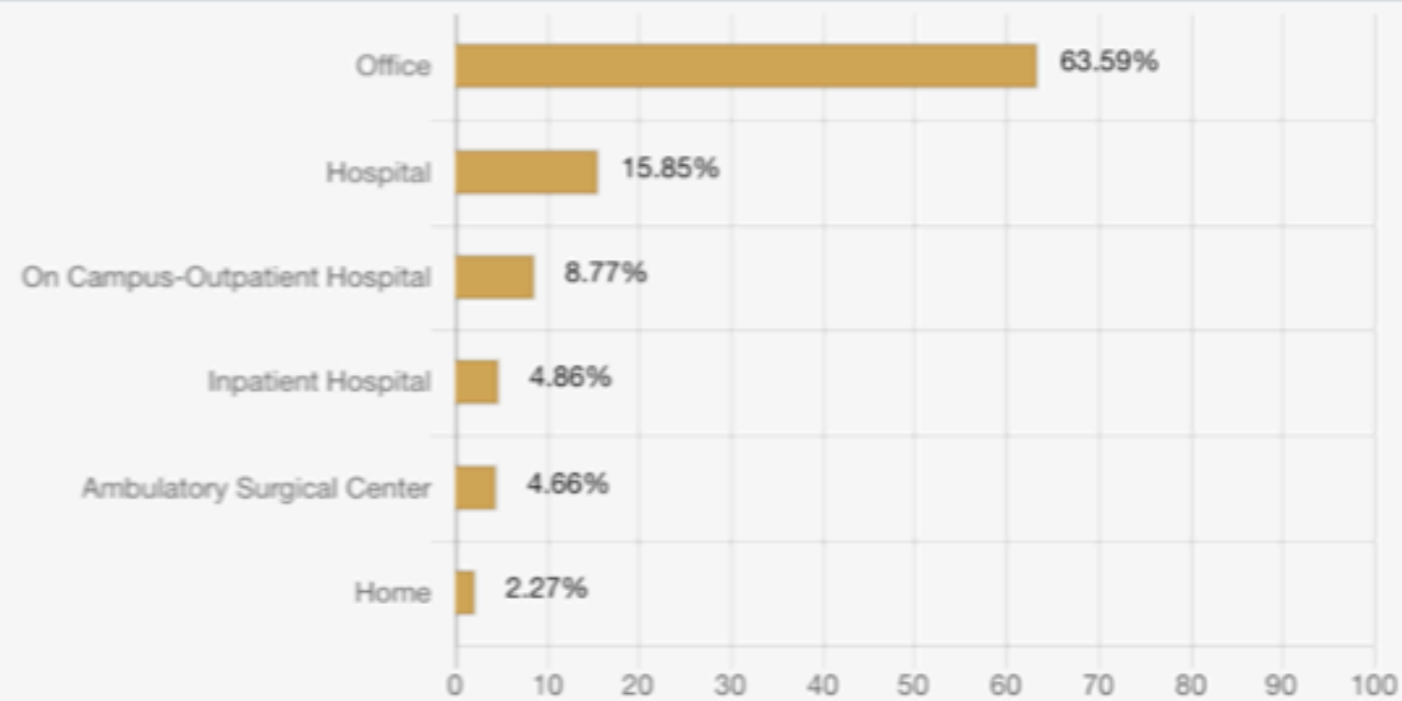
- systém na základe príznakov a demografie určí diagnózu a následne výsledky preusporiada s využitím histórie diagnóz pacienta.



## Most frequent Providers' Specialty



## Most frequent Place of Service



# Personalizovaná kontrola symptómov

45 kariet pacientov

<b>Metric</b>	<b>Emergent care</b>	<b>Non emergent care</b>	<b>Self-care</b>	<b>Overall</b>
Top@1	20.00	20.00	50.00	29.55
Top@3	20.00	33.33	57.14	36.36
Top@5	33.33	53.33	64.29	50.00
Top@10	46.67	60.00	71.43	59.09
Top@20	53.33	66.67	78.57	65.91

**Table 2: Accuracy evaluation (%) of symptom checker across diagnosis requiring emergency, non-emergency and self-care.**

# Ďalšie zaujímavé domény

<https://recsys.acm.org/recsys18/workshops/>

<https://recsys.acm.org/recsys18/program/>

a predchádzajúce ročníky...