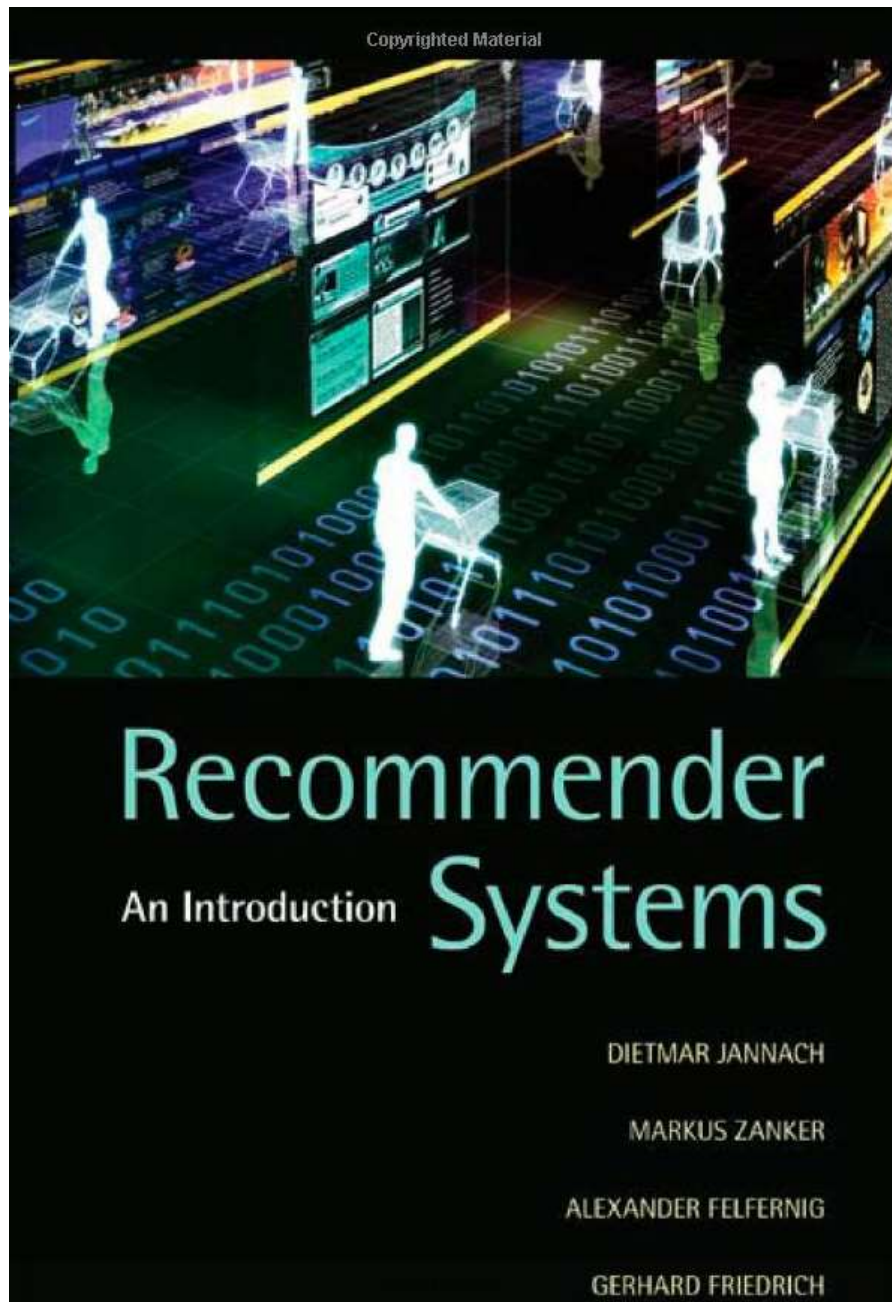


AJÁNLÓ RENDSZEREK



- a szabványos, egyenvásárlóra tervezett **tömegtermékek** ideje lejárt
 - többféle vásárló, többféle igényét kielégítő, **heterogén** (sokféle) **termékek** gyártásának irányába kell elmozdulni
- ↓
- szükség van ajánló rendszerekre

Motiváció

A heterogenitás dilemmája:

- Megvan a sokféle termék, változatos kínálat 😊
- A vásárlónak rengeteg lehetőségből kell választani (információ túlterhelés) ☹️

A kínálatot szűrni/rangsorolni kell!

Ajánló rendszerek:

- A vásárlónak csak a releváns termékeket mutassuk 😊
- Egyénre/testreszabott vásárlás 😊

Gazdasági előnyök az eladók számára

- **Webes vásárlásokat** (vásárlói kosarak létrejönnek) könnyű nyomon követni 😊

Ha a webshop böngészése közben **érdekes termékekkel** találkozik a vásárló, talán vásárolni is fog!

- **Visszatérő vásárlók** (vásárlások száma növekszik) 😊

Ha a konkurenciához képest nálunk **érdekesebb termékekkel** találkozik a vásárló – mert a régebbi ténykedéseiből épített vásárlói profil szerint ajánlunk kedvére valót -, talán emiatt már legközelebb is inkább hozzánk tér be vásárolni.

- **Járulékos eladások** (a kosár átlag mérete növekszik) 😊


Ha vásárlás közben további **érdekes termékekkel** találkozik a vásárló, talán azokból is vásárolni fog.

Ajánló rendszerek

localhost/recommendation/recommendations.php

Alkalmazások Daniel Bognar (@b... Darkfly Video (@da... GDPR Rico x Miss Mood -... Next Level ft. Quer... BLR x Rave & Crave... Dj Free & Purebeat... Kygo & Ellie Gouldi...

Moviee ©


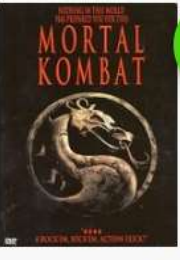








firstuser

- My profile
- Rating history
- Recommendations**
- About
- Logout

Current location: **Veszprém, HU**

Do you want to change? [CHANGE LOCATION](#)

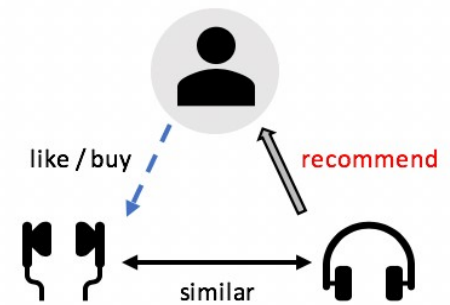
The recommended movies list in order of your predicted taste:

		
Hudson Hawk	Mortal Kombat	The Golden Compass
1991	1995	2007
★ 3.6/10	★ 4.4/10	★ 5.6/10
 	 	 

Definíció

Az ajánlórendszerek speciális **információsűrő rendszerek**, amelyek olyan **tartalmat kínálnak a felhasználónak**, amely nagy valószínűséggel érdekes lesz a számára.

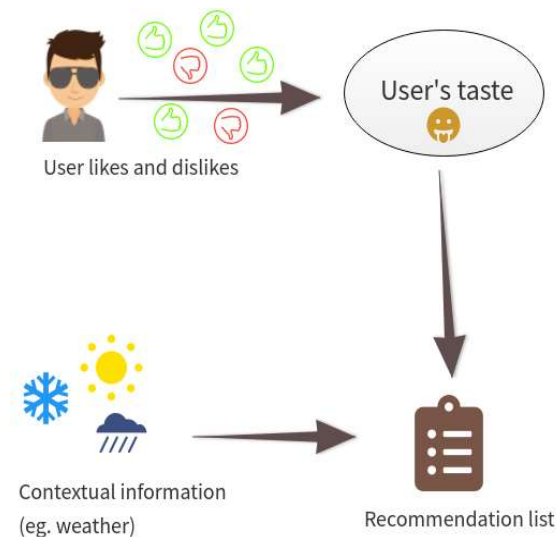
A legelterjedtebb változatok:



Tartalomalapú ajánlás

A		✓	✗	✓	✓
B			✓	✗	✗
C		✓	✓	✗	
D		✗		✓	
E		✓	✓	?	✗

Kollaboratív ajánlás



Kontextus függő ajánlás

Megközelítések

- **Kollaboratív ajánlás:** A felhasználónak olyan termékeket ajánlunk, melyeket más, **hasonló ízlésű** felhasználók magasra értékelték.
- **Tartalomalapú módszerek:** A felhasználónak olyan termékeket ajánlunk, melyek tartalmukban hasonlóak azokhoz, amelyeket a **múltban magasra értékelt**.
- **Kontextus függő:** Figyelembe vesszünk valamilyen jellemzőt, amely csak **egy adott kontextusban érdekes**.
- **Hibrid módszerek:** Az előző módszerek valamilyen szintű ötvözése.

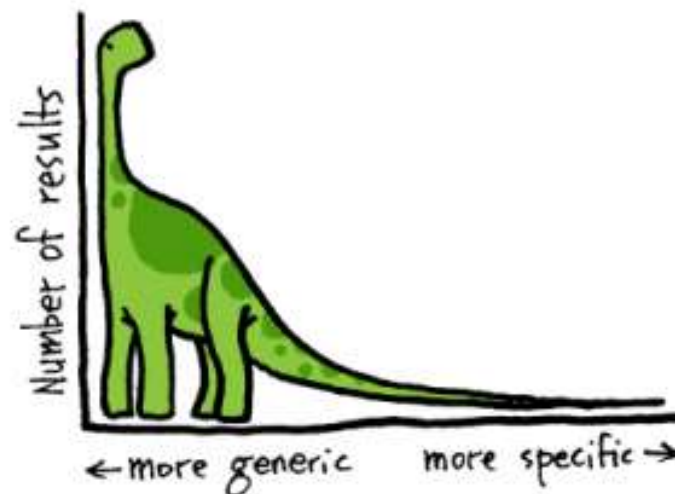
Néhány ajánlórendszer megközelítés összehasonlítása

Ajánlórendszer típusa	Előny	Hátrány
Kollaboratív	Bőséges értékelésekkel és sok felhasználóval pontos ajánlás.	Hideg-indítás probléma az új felhasználók esetében (Hideg indítás problémáról akkor beszélünk, amikor egy új felhasználóról nincs információnk és nem tudja az algoritmus, hogy mit ajánljon neki).
Tartalomalapú	Nincs szükség a többi felhasználó értékeléseire az ajánlás elkészítéséhez.	Nincs váratlan termék az ajánlások között, túlságosan kiszámítható ajánlások.
Tudásalapú	Determinisztikus ajánlások, nincs hidegindítás.	A tudás direkt megszerzése túl kényelmetlen.
Hibrid	A vegyes alkalmazás miatt, elkerüli néhány hátrányát az egyes megközelítéseknek.	Komplex, bonyolult megvalósítási mód.
Demográfiai	Az ajánlások csak az adott régióra érvényesek, így elkerülve az érvénytelen ajánlásokat.	Kevés a kutatás még ebben a típusú megközelítésben.

Kiértékelés

Valós célfüggvények:

- Ügyfél elégedettsége!
 - Ajánlott elemek megvásárlása (likeolása)
 - Többet vásárolnak mintha nem lenne ajánlás?
-
- A Long Tail jelentkezésének klasszikus példája az Amazon.com sikere. A webáruház olyan **közösségi ajánlórendszer**t épített árukatalógusa és megrendelő oldalai köré, amely minden professzionális marketingnél hatékonyabban hívja fel a vásárlók figyelmét a számukra csábító termékekre.
 - A "hosszú farok" szerint **megfelelően hatékony, korszerű logisztikával a pár embert érdeklő termékeket is ugyanolyan hatásfokkal és haszonnal lehet értékesíteni** (tehát támogatni, forgalmazni, piacra dobni is), mint a hagyományosan sikerbiztos termékeket.



Számítanak-e a generációs különbségek?

- Általában véve: **NEM**
- Egyes területeken **IGEN**:
 - eszközhasználat
 - kommunikáció módja
 - megértés
 - ügyfélkezelés (reakciók, interakciók)
 - üzenet a felhasználónak (marketing)



Kívánalmak általában

A VEVŐ

- Olcsóbban (legolcsóbban)
- Gyorsan (gyorsabban)
- Könnyebben
- Megbízható eladótól
- Egyéb szempontok:
 - Fizetési mód
 - Ügyfélkezelés
 - Presztízs
-

A KERESKEDŐ

- Több vevő
- Gyorsabb kiszolgálás
- Nagyobb bevétel (-profit)
-

Ajánlási feladat formális leírása

C - felhasználók

S - termékek

$u : C \times S \rightarrow R$ hasznosságfüggvény

(R halmaz, teljes rendezéssel).

- explicit értékelés: **milyen skálán?**
- implicit értékelés: **kattintások, oldallátogatások, eltöltött idő stb.**

Tanulási feladat:

határozzuk meg azt az

$u' : C \times S \rightarrow R$ leképezést, amely

- a lehető legjobban közelíti u -t és
- teljesen definiált a teljes $C \times S$ téren.

Ajánló rendszer kimenete: $\forall c \in C$ vásárló számára a legérdekesebb/leghasznosabb új termék(ek) megadása, azaz:

$$s_c \in S, \text{ hogy } u(c, s_c) \text{ maximális}$$

1. Kollaboratív ajánlás

Alapgondolat:

- Felhasználók (explicit vagy implicit) értékelik a tartalmakat
- Azok a felhasználók akik a múltban hasonló dolgokat kedveltek a jövőben is hasonlóan fognak viselkedni

					
A		✓	✗	✓	✓
B			✓	✗	✗
C		✓	✓	✗	
D		✗		✓	
E		✓	✓	?	✗

Kollaboratív ajánlórendszerek

					
A		✓	✗	✓	✓
B			✓	✗	✗
C		✓	✓	✗	
D		✗		✓	
E		✓	✓	?	✗

- Egy adott felhasználó számára egy adott (még nem értékelt) termék értékelését hasonló ízlésű felhasználók által a **termékre adott értékelések összesítéséből becsüljük**.
- Lépések:
 1. Felhasználópárok hasonlósági faktorának megállapítása.
 2. Adott felhasználóhoz hasonló ízlésű felhasználók megkeresése.
 3. A hasonló ízlésű felhasználók által adott értékek aggregálásával az ismeretlen értékelés becslése.

Kollaboratív ajánló rendszer

- A továbbiakban egy termék (pl. film) **hasznosságán** mindig egy **adott skálán mozgó értékelést** (rating) értünk, azaz, hogy egy felhasználó milyen magasra pontozta, értékelte (volna) az adott terméket. Ennek megfelelően a következő jelölést vezetjük be:

$$r_{c,s} \equiv u(c,s)$$

- Egy adott felhasználó számára egy adott (még nem értékelt) termék értékelését most **a hasonló ízlésű felhasználók által a termékre adott értékelések összesítéséből becsüljük:**

$$\hat{r}_{c,s} \leftarrow \forall r_{c_i,s} \text{ alapján, ahol: } c \sim c_i$$

Az ajánlás menete

1. Felhasználópárok **hasonlósági faktor**ának megállapítása:

$$\forall x, y \in \mathcal{C} : \text{sim}(x, y) \text{ meghatározása}$$

2. Adott felhasználóhoz **hasonló ízlésű felhasználók** megkeresése:

$$\forall c \in \mathcal{C} : c \rightarrow \hat{\mathcal{C}} = \{c' \mid c' \in \mathcal{C}, \text{sim}(c, c') \text{ nagy}\}$$

3. A hasonló ízlésű felhasználók által adott értékek **aggregálásával** az ismeretlen értékelés becslése:

$$\hat{r}_{c,s} = \text{aggr}_{\forall c' \in \hat{\mathcal{C}}} (r_{c',s})$$

Felhasználó párok hasonlósága

- Általában a felhasználók által közösen értékelt S_{xy} termékhalmaz alapján dolgozunk:

$$S_{xy} = \{s \in \mathcal{S} \mid r_{x,s} \neq \emptyset, r_{y,s} \neq \emptyset\}$$

- A mindkettőjük által értékelt termékek pontszámait egy-egy \vec{x} és \vec{y} vektorként tekintve, definiálhatjuk a hasonlóságot az ezek által bezárt szög koszinuszaként:

$$\text{sim}(x, y) = \cos(\vec{x}, \vec{y}) = \frac{\vec{x} \cdot \vec{y}}{|\vec{x}| \cdot |\vec{y}|} = \frac{\sum_{s \in S_{xy}} r_{x,s} \cdot r_{y,s}}{\sqrt{\sum_{s \in S_{xy}} r_{x,s}^2} \cdot \sqrt{\sum_{s \in S_{xy}} r_{y,s}^2}}$$

Hasonló ízlésű felhasználók megkeresése

- Adott egy c felhasználó, a hozzá leginkább hasonló ízlésű néhány másik felhasználót szeretnénk megtalálni.
- Kiszámíthatjuk minden x, y felhasználó párra $sim(x,y)$ -t, majd rendezzük. **Nagyon lassú lehet nagy adathalmazok esetén, mivel az adatok nem férnek el az operatív memóriában.**
- **Gráfot építhetünk** a felhasználókból, ahol a hasonló ízlésűek valamilyen szempontból „közel” helyezkednek el egymáshoz.

Értékelések aggregálása

- Valamilyen átlagot számolunk:

- a) Egyszerű átlag

$$\hat{r}_{c,s} = \frac{1}{N} \sum_{\forall c' \in \hat{C}} r_{c',s}, \text{ ahol } N = |\hat{C}|$$

- b) Felhasználók hasonlóságával súlyozott átlag

$$\hat{r}_{c,s} = k \cdot \sum_{\forall c' \in \hat{C}} \text{sim}(c, c') \cdot r_{c',s}, \text{ ahol } k = \frac{1}{\sum_{\forall c' \in \hat{C}} |\text{sim}(c, c')|}$$

- c) Súlyozott átlag, normalizált értékelésekkel

$$\hat{r}_{c,s} = \bar{r}_c + k \cdot \sum_{\forall c' \in \hat{C}} \text{sim}(c, c') \cdot (r_{c',s} - \bar{r}_{c'})$$

Modellalapú módszerek – kollaboratív ajánlás esetén

- **Klaszterezésen alapuló módszerek:** a hasonló ízlésű felhasználókat egy-egy osztályba soroljuk, ezen belül naiv bayes-i osztályozót használunk a becslésre.
- **Bayes-hálók:** a csomópontok a termékek, lehetséges állapotaik az egyes értékelések (pl. 1-10-ig)
- További **gépi tanulási módszerek** használata ...

Kollaboratív ajánló rendszerek

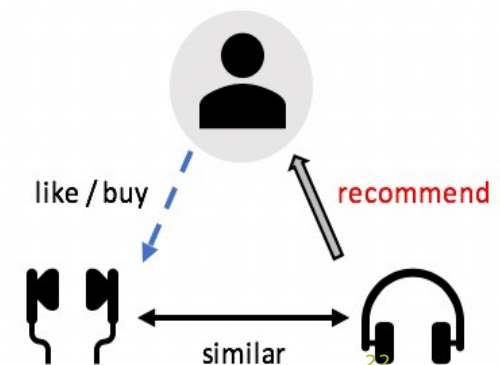
Előnyök:

- Tetszőleges típusú termékek
 - A felhasználók értékrendjük szerint bármilyen típusú terméket képesek pontozni, így mindenféle termékhez rendelhető hasznosság.
 - A termékeknek nem is kell klasszikus értelemben összehasonlíthatónak lenniük.
- Minden megkülönböztethető
 - A felhasználók értékeléskor a teljes tartalmat nézik, így nem lehet olyan probléma, mint pl. a tartalomalapú ajánlás esetében, hogy a kívánt jellemzők megegyeznek, így egy jó és rossz filmet meg tudunk különböztetni.
- Hátrányok:
 - Sok felhasználó, sok értékelés kell (minden felhasználóhoz legyen elég sok hasonló ízlésű, akik ráadásul értékelték is a kérdéses terméket
 - Új felhasználó problémája
 - Új termék problémája
 - Külső információt nehéz bevonni

					
A		✓	✗	✓	✓
B			✓	✗	✗
C		✓	✓	✗	
D		✗		✓	
E		✓	✓	?	✗

2. Tartalom alapú ajánló rendszerek

- Kizárólag a **célfelhasználó** korábbi értékelései alapján **ajánlunk**
- Az **elemeket jellemzőkkel** írjuk le
pl: színészek, rendező, műfaj, főszereplő stb.
- A jellemzőkre egy **osztályozót**/regresszort építhetünk
- **Tanítás** a célfelhasználó korábbi értékelései alapján



Tartalom alapú ajánló rendszerek

Előnyök

- Független a többi felhasználótól
- Egyetlen felhasználó esetén is életképes a rendszer
- Hidegindítás problémája kisebb
- Ritkaság nem okoz gondot
- „Egységiségek”-nél is működik
- Új vagy nem felkapott elemeket is ajánlhat (long tail)
- Értelmezhető modell (pl. legnagyobb súlyú tüzelő jellemzők)

Tartalom alapú ajánlás - Hátrányok

- **Termékjellemzők előállításának problémája**

- Szöveges dokumentumok esetén jól automatizálható, de más termékek (pl. képek, filmek) esetén manuálisan kell megadni (szereplők, műfaj, stb.)
→ sok munka ☹
- Megint más termékeknél esetleg egyáltalán nem tudunk olyan (skalár) jellemzőket megadni, melyen keresztül értelmesen összehasonlíthatnánk azokat.

- **Megkülönböztethetlenség problémája**

- Ha két termék jellemzői azonosak (a két ennes megegyezik), a rendszer számára megkülönböztethetetlenek.

Pl. Egy jól és egy rosszul megírt tudományos cikk ugyanabban a témában.

- **Túl specializáció**

- Elképzelhető, hogy nem mindig a már magasan értékeltékhez leginkább hasonló elem ajánlása az ideális

Pl. Étterem ajánló rendszer: amennyiben a felhasználó még csak kínai és magyar ételeket értékelt, soha nem fogunk neki görög éttermet ajánlani (még ha az a legjobb is a városban), hiszen túlságosan különböző az eddig (jónak) értékelt elemektől

- **Egy új felhasználónak egy ideig nem tudunk értékelést adni (megfelelő méretű tanító adatbázis kell!)**

Tartalomalapú ajánló rendszer

- Egy adott felhasználó számára egy termék hasznosságát a hasonló termékek **ugyanazon felhasználó által megadott hasznosságaiból becsüljük** és a legnagyobb becsltet ajánljuk:

$$\hat{u}(c, s) \leftarrow \forall u(c, s_i) \text{ alapján, ahol: } s \sim s_i$$

- Klasszikusan szöveges dokumentumokra alkalmazzuk.
- A dokumentumokat tartalmilag pl. kulcsszavak gyakoriságával, fontosságával jellemezhetjük.
- Két dokumentum hasonló, ha tartalmuk, pl. jellemző kulcsszavaik hasonlóak.
- Feltételezzük, hogy tartalmában a már kedvelt dokumentumokhoz hasonlóra vágyik a felhasználó.

Az ajánlás menete

1. Az egyes dokumentumokhoz **tartalmi profilt** készítünk:

$$\forall s \in \mathcal{S} : s \rightarrow \text{Content}(s)$$

2. Az értékelt dokumentumok alapján **felhasználói profil** készül:

$$\forall c \in \mathcal{C} : \{(\text{Content}(s), u(c, s)) \mid s \in \mathcal{S}_c\} \rightarrow \text{ContentBasedProf}(c)$$

3. Ezek alapján az ismeretlen **hasznosságokat** becsüljük:

$$\hat{u}(c, s) = \text{score}(\text{Content}(s), \text{ContentBasedProfile}(c))$$

Dokumentum profil készítése (pl. egy lehetséges heurisztikára)

- Legyenek a rendszerben nyilvántartott kulcsszavak:

$$\sigma_1, \sigma_2, \dots, \sigma_k$$

- Az s_i dokumentumban σ_j kulcsszó gyakorisága:

$$f_{i,j}$$

- Egy dokumentum profilja legyen az a vektor, mely az összes kulcsszavak $w_{i,j}$ fontosságait tartalmazza erre a dokumentumra:

$$\text{Content}(s_i) = \vec{w}_{s_i} = (w_{i,1}, w_{i,2}, \dots, w_{i,k})$$

Dokumentum profil készítése (pl. egy lehetséges heurisztikára)

- Egy **kulcsszó fontossága** sokféleképp definiálható, de pl. mondhatjuk, hogy arányos a normalizált előfordulási gyakoriságával:

$$TF_{i,j} = \frac{f_{i,j}}{\max_z f_{i,z}}$$

- A túl sok dokumentumban szereplő kulcsszavak viszont nem túl informatívak, ezek súlyát hivatott csökkenteni az **inverse-document-frequency** tényező:

$$IDF_j = \log \frac{n}{n_j}, \text{ ahol } n_j = \left| \left\{ i \mid f_{i,j} > 0 \right\} \right|$$

- Így a **fontosság**:

$$w_{i,j} = TF_{i,j} \cdot IDF_j$$

Felhasználói profil készítése (pl. egy lehetséges heurisztikára)

- A felhasználó profilja lehet pl. szintén egy vektor, mely a múltban elolvasott dokumentumok alapján az adott felhasználó egyes kulcsszavakra vonatkozó $v_{i,j}$ preferenciáit tartalmazza:

$$\text{ContentBasedProfile}(c_i) = \vec{v}_{c_i} = (v_{i,1}, v_{i,2}, \dots, v_{i,k})$$

A preferenciák meghatározása történhet a már értékelt dokumentumok \vec{w}_{s_i} vektorainak valamilyen átlagolásával.

- Így a \vec{v} preferenciavektor szintén tekinthető egy TF-IDF vektornak.

Hasznosságbecslés (pl. egy lehetséges heurisztika)

- A hasznosságot definiálhatjuk pl. a \vec{v} és \vec{w} vektorok által bezárt szög koszinuszaként:

$$\begin{aligned}\hat{u}(c_i, s_l) &= \text{score}(\text{Content}(s_l), \text{ContBasedProfile}(c_i)) = \\ &= \cos(\vec{v}_{c_i}, \vec{w}_{s_l}) = \frac{\vec{v}_{c_i} \cdot \vec{w}_{s_l}}{|\vec{v}_{c_i}| \cdot |\vec{w}_{s_l}|} = \frac{\sum_{j=1}^k v_{i,j} \cdot w_{l,j}}{\sqrt{\sum_{j=1}^k v_{i,j}^2} \cdot \sqrt{\sum_{j=1}^k w_{l,j}^2}}\end{aligned}$$

Hasznosságbecslés (modell-alapú módszerek)

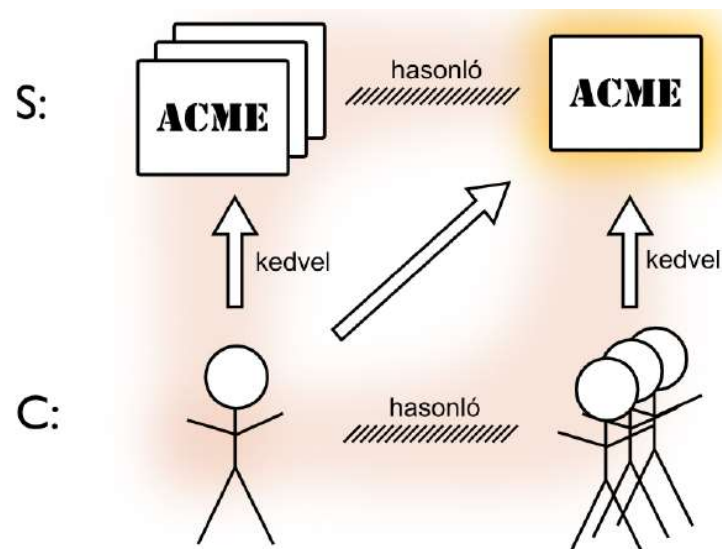
- A bemutatott vektoros becslés a **heurisztikus módszerek** családjába tartozik, és természetesen csak egy a sok lehetséges algoritmus közül.
- Vannak ún. **modell-alapú módszerek**
 - **Content (s)** itt is valamilyen statisztika a dokumentumokról.
 - **ContentBasedProfile(c)** viszont valamilyen bonyolultabb modell.
 - A hasznosságbecslés a modell kimenetén jelenik meg, az adott dokumentum profilját a bemenetére adva:

$$\hat{u}(c_i, s_l) = Model_{c_i}(Content(s_l))$$

Pl. Naiv bayes-i osztályozó felépítése a kulcsszógyakoriságból

Pl. Neurális hálózatok

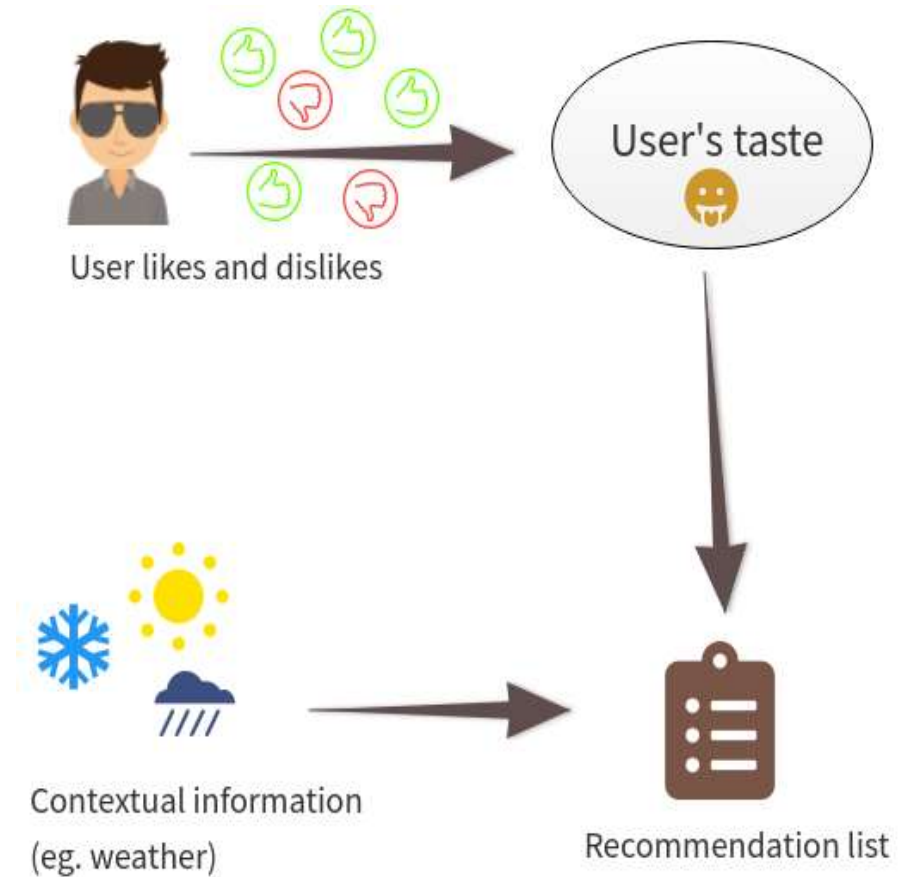
3. Hibrid rendszerek létrehozásának lehetőségei



- **tartalom alapú** → **kollaboratív**
 - „tanulható” felhasználóknál hiányzó értékelésekre becslés tartalom alapú módszerekkel
 - a becsült értéket (kis súllyal) felhasználjuk kollaboratív módszerekben
 - **kollaboratív** → **tartalom alapú**
 - jellemzőket nyerünk ki a többi felhasználó és azok értékelései alapján
- Pl: szakirodalom ajánlásnál hivatkozások alapján

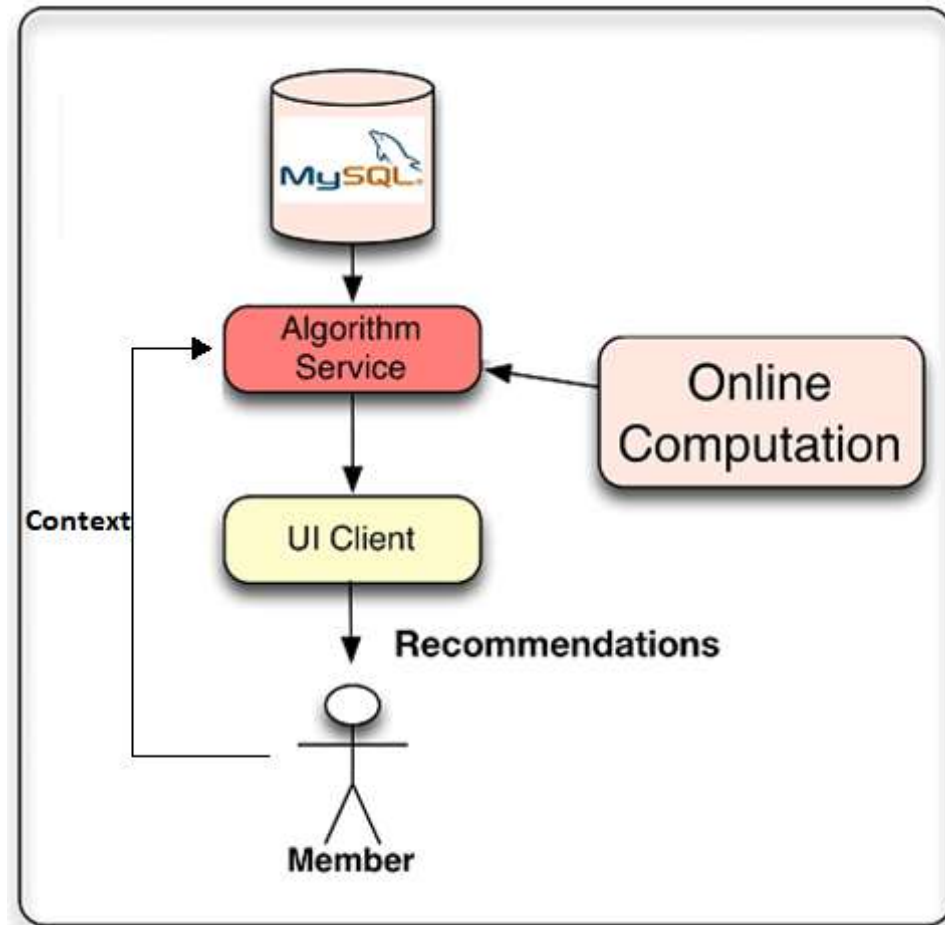
4. Kontextus függő ajánló rendszerek – egy konkrét megvalósítás

- Utazási ajánlatok helyszínekre
- Télen és nyáron különböző
- Kontextus: évszak



Kontextus függő ajánló rendszerek

- Filmek ajánlása a felhasználó ízlése alapján, amely felhasználja a **felhasználó lokációját**
- A felhasználó lokációja körüli filmek ajánlási sorrendbe állítása a cél



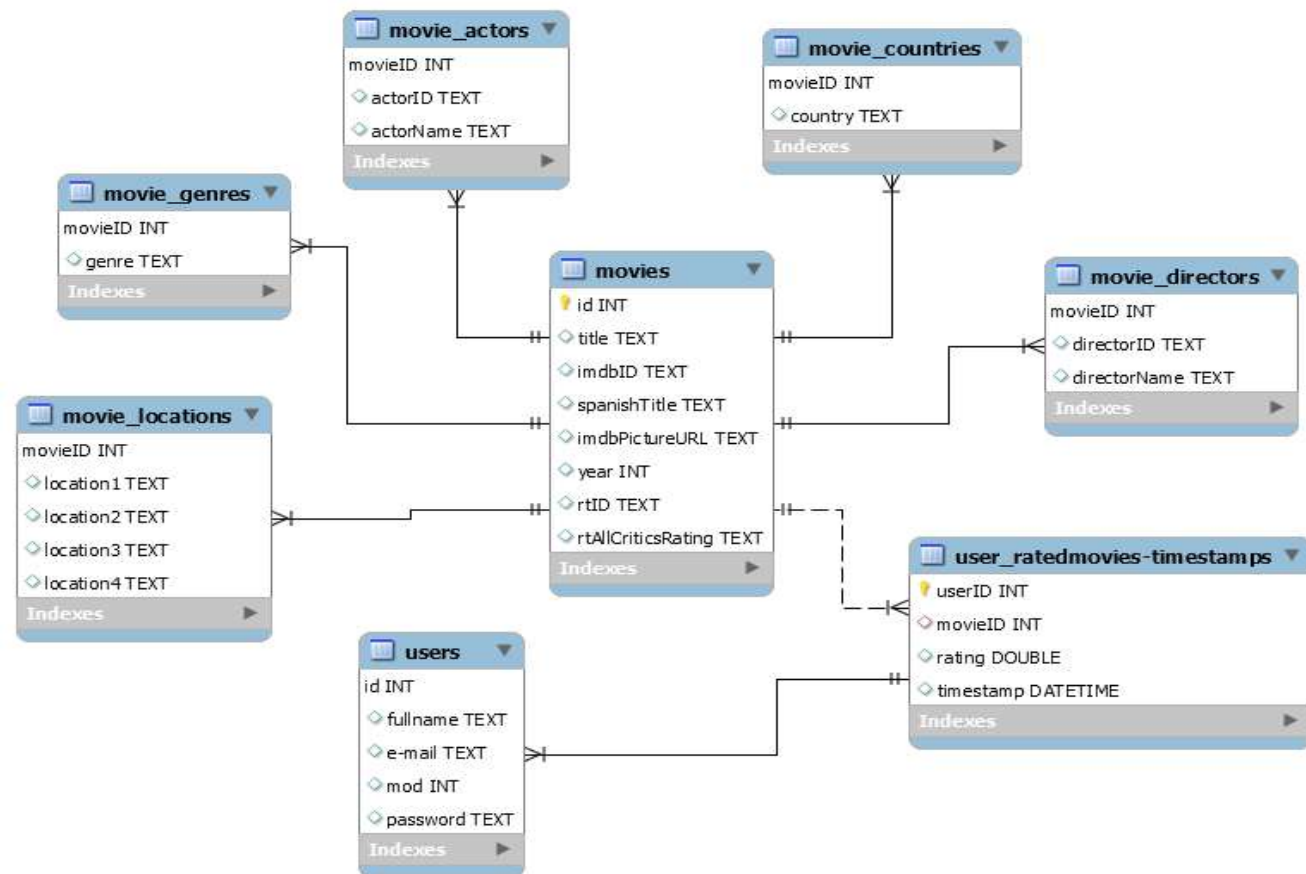
Kontextus függő ajánló rendszerek Adatbázis

MovieLens adatbázis

- 2113 felhasználó
- 10197 film
- 20 műfaj
- 4060 rendező
- 95321 színész
- 72 ország
- 47899 helyszín
- **855598 értékelés**

Filmek értékelése:

0.5 - 5



Kontextus függő ajánló rendszerek Felhasználói interfész

The screenshot shows a web browser at the URL localhost/recommendation/recommendations.php. The browser's address bar and tabs are visible at the top. The application interface is divided into a left sidebar and a main content area.

Left Sidebar (User Profile):

- Moviee © logo
- Profile picture of a man with glasses and a suit, labeled "firstuser".
- Navigation menu:
 - My profile
 - Rating history
 - Recommendations (highlighted)
 - About
 - Logout

Main Content Area:

- Current location: **Veszprém, HU**
- Do you want to change? [CHANGE LOCATION](#)
- The recommended movies list in order of your predicted taste:

Movie Title	Year	Rating	Like Icon	Heart Icon
Hudson Hawk	1991	★ 3.6/10	Green thumbs up	Red heart
Mortal Kombat	1995	★ 4.4/10	Green thumbs up	Red heart
The Golden Compass	2007	★ 5.6/10	None	Red heart

The "The Golden Compass" card also features a small number "37" in the bottom right corner.

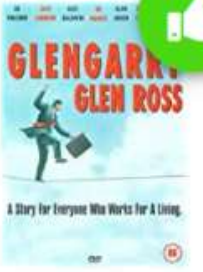





My ratings

localhost/recommendation/myratings.php

Alkalmazások Daniel Bogнар (@bo Darkfy Video (@dar AnswerMiner - Predi GDPR Rico x Miss Mood - Next Level ft. Query BLR x Rave & Crave Dj Free & Purebeat

Movie 🎬

- My details
- My rated movies**
- The recommendations
- About

 Glengarry Glen Ross 1992 ★ 8.5/10	 Strange Days 1995 ★ 6/10	 Stalker 1979 ★ 8.2/10
 Crouching Tiger, Hidden Dragon	 ONE MILLION YEARS B.C.	 REPO MAN