

# A quienes les gustó esta charla también les gustó ...

Cómo los Sistemas de Recomendación revolucionaron la  
industria del comercio electrónico

Lic. Ernesto M. Mislej

Charla de Borrachos Depto. de Computación - FCEyN - UBA

`emislej@gmail.com`

27 de Junio de 2013



Afro-Celt Sound System

# Definición

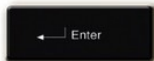
Los Sistemas de Recomendación (RS), rec. systems, engines, frameworks, platforms es una disciplina derivada del Data Mining que se centra en el diseño de filtros sobre colecciones de items que son del gusto o del interés del usuario.

Para tal fin se utilizan la historia de calificaciones realizada por el usuario y por la comunidad de usuarios.

Se trabajan sobre dominios ligados al gusto como son las películas, programas de TV, video por demanda, música, libros, espectáculos, restaurantes, noticias, entre otros.

*How Endless Choice Is Creating Unlimited Demand*

# The Long Tail



Why the Future of Business  
Is Selling Less of More

The Long Tail

CHRIS ANDERSON

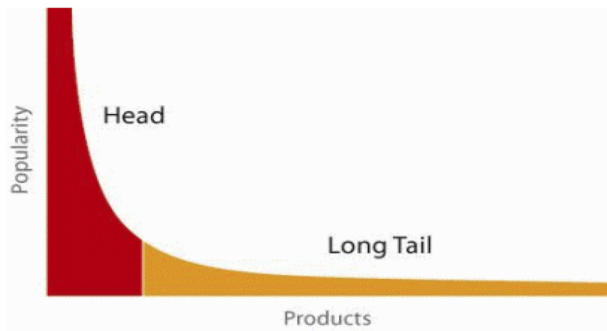
*"Anderson's insights influence Google's strategic thinking in a profound way.*

*READ THIS BRILLIANT AND TIMELY BOOK."*

*—ERIC SCHMIDT, CEO, GOOGLE*

# The Long Tail

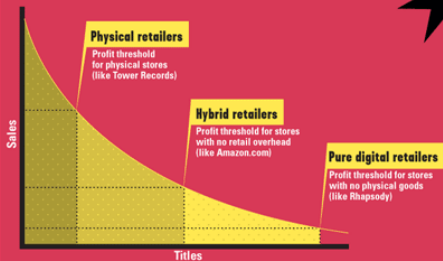
- *Recomendaciones de items muy poco populares y aún desconocidos por el usuario, pero que serán de su agrado!*
- 20% de los items acumulan el 80% de los votos
- **Serendipia** (a.k.a **chiripa**) ocurre cuando encontrás algo *supuestamente* inesperado.



# THE BIT PLAYER ADVANTAGE

Beyond bricks and mortar there are two main retail models – one that gets halfway down the Long Tail and another that goes all the way. The first is the familiar hybrid model of Amazon and Netflix, companies that sell physical goods online. Digital catalogs allow them to offer unlimited selection along with search, reviews, and recommendations, while the cost savings of massive warehouses and no walk-in customers greatly expands the number of products they can sell profitably.

Pushing this even further are pure digital services, such as iTunes, which offer the additional savings of delivering their digital goods online at virtually no marginal cost. Since an extra database entry and a few megabytes of storage on a server cost effectively nothing, these retailers have no economic reason not to carry *everything* available.



## "IF YOU LIKE BRITNEY, YOU'LL LOVE ..."

Just as lower prices can entice consumers down the Long Tail, recommendation engines drive them to obscure content they might not find otherwise.



Source: Amazon.com

# NETFLIX

The Netflix Prize

# The Netflix Prize

- **Training data:** 100 millones de calificaciones, 48.000 usuarios, 17.770 películas
- 6 años de información: 2000-2005
- **Test data:** Algunas últimas calificaciones de cada usuarios (2.8 millones)
- Criterio de evaluation: root mean squared error (RMSE)
- *Netflix Cinematch* RMSE: 0.9514
- **Concurso** 2700+ equipos
- \$1M USD para aquél quien supere en un 10 % a *Netflix Cinematch*
- **Grand Prize to team *BellKor's Pragmatic Chaos*** 21 de septiembre de 2009.



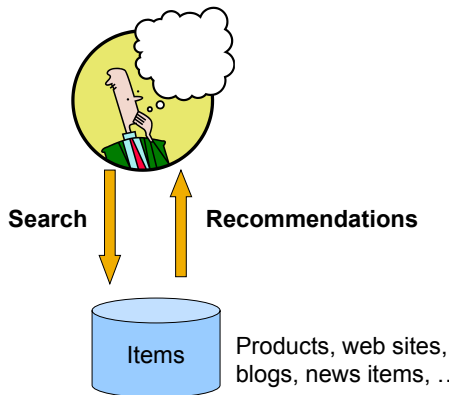
A close-up portrait of a man with dark hair and a mustache, looking slightly to the left. He is wearing a dark, possibly leather, vest over a light-colored shirt. The background is blurred, showing what appears to be an interior setting with a window or light source on the left.

Juan Carlos

A close-up portrait of a young man with long, dark, wavy hair and a mustache. He is smiling and looking slightly to the left of the camera. He is wearing a light blue tank top. The background is a plain, light-colored wall with a window or doorway visible behind him.

René

# Sistemas de Recomendación



# Ejemplos de Sistemas de Recomendación

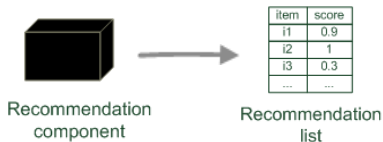
- 1 Ofrecer artículos periodísticos en sitios de noticias online basados en los intereses del lector.
- 2 Ofrecer sugerencias de productos en una tienda online basados en el historial de compra.
- 3 Encontrar cuál es el grupo de fans con más ganas para viajar a ver un recital.
- 4 Ofrecer cursos de capacitación a personas conociendo sus capacidades y su historia curricular.
- 5 Sugerir destinos de viaje en relación a las búsquedas realizadas en una agencia de viajes online.

# Paradigmas de los sistemas de recomendación

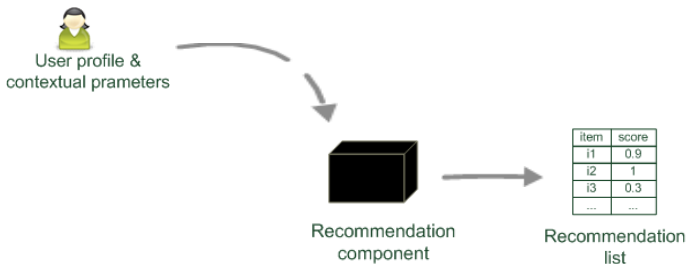
Los *recsys* utilizan diferentes tecnologías, podemos clasificarlas dentro de este grupo de 2 (+1) categorías.

- **Basados en Contenido** el sistema encuentra características de los ítems pertenecientes al catálogo y encuentra ítems aún no vistos. Por ejemplo, si el usuario vio muchas películas *cowboys*, el sistema recomendará otras películas de ese género.
- **Filtros Colaborativos** el sistema encuentra otros usuarios similares respecto de sus historiales de compra y recomendará ítems aún no vistos.
- **Basados en Conocimiento:** el sistema emula el comportamiento de un vendedor realizando preguntas para conocer cuáles son nuestros intereses

# Paradigmas de los sistemas de recomendación

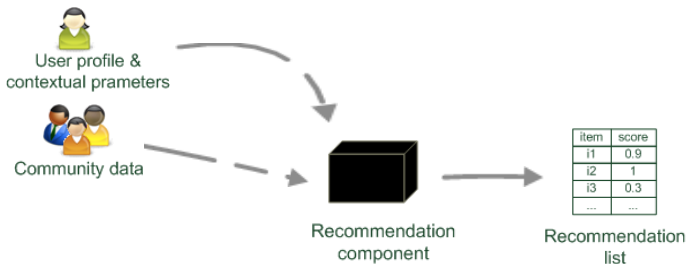


# Paradigmas de los sistemas de recomendación



**Filtros personalizados:** *Sólo mostrame estás categorías*

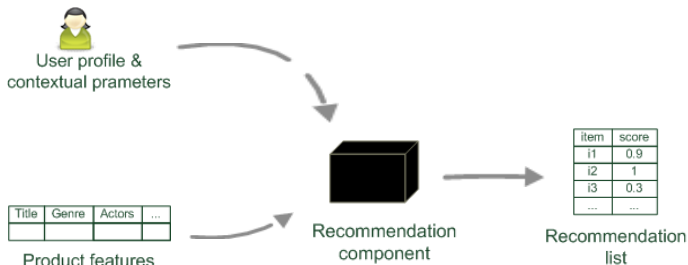
# Paradigmas de los sistemas de recomendación



**Collaborative:** *Contame qué le gustó a mis pares*

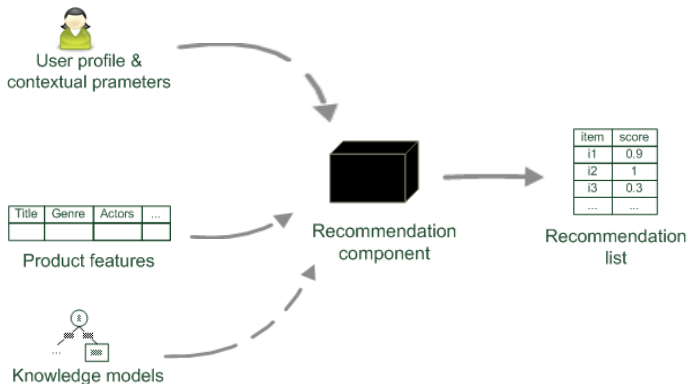


# Paradigmas de los sistemas de recomendación



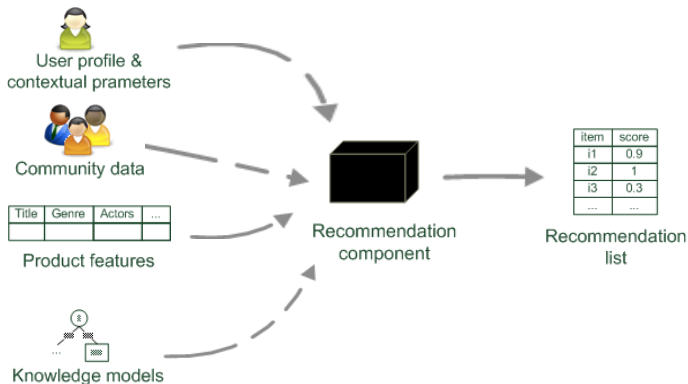
**Content-based:** *Dame más de lo que me gusta*

# Paradigmas de los sistemas de recomendación



**Knowledge-based:** *¿Qué estás buscando?*

# Paradigmas de los sistemas de recomendación



**Híbridos:** combinando los diferentes mecanismos

# Content-based recommendation

**Idea:** Recomendar ítems al usuario  $x$  similares a los ítems que previamente el usuario  $x$  calificó positivamente. Recomendar *fantasy novels* a la gente que le gustó *fantasy novels* en el pasado.

Qué necesitamos:

- características del ítem como el género. A ello lo llamamos (*content*).
- construir un perfil del usuario que describen sus gustos (*preferences*).

El objetivo:

- inducir preferencias
- reconocer ítems *similares* a los preferidos del usuario en el pasado.



Videos de gatos  
¿Dónde está el género?



Duplicados

# Collaborative Filtering

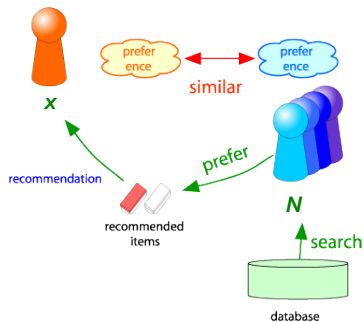
- $C$  = conjunto de Usuarios
- $S$  = conjunto de Items
- Función de utilidad  $u : C \times S \rightarrow R$ 
  - $R$  = conjunto de calificaciones

## Matriz de utilidad:

|       | Item 1 | Item 2 | Item 3 | Item 4 | Item 5 |
|-------|--------|--------|--------|--------|--------|
| Alice | 5      | 3      | 4      | 4      | ?      |
| Berta | 3      | 1      | 2      | 3      | 3      |
| Clara | 4      | 3      | 4      | 3      | 5      |
| Diana | 5      | 1      | 4      | 3      | 4      |
| Erika | 1      | 5      | 5      | 2      | 1      |

# Collaborative Filtering

- Consideremos al usuario  $x$
- Identificar al grupo  $N$  de otros usuarios tal que sus calificaciones son *similares* a las calificaciones de  $x$
- Luego estimar las calificaciones de los ítems aún-no-vistos de  $x$  utilizando las calificaciones del grupo  $N$





# Dimensionality Reduction

Un enfoque sería intentar estimar las celdas vacías de la matriz de utilidad como el producto de 2 matrices (*thin matrices*) rectangulares.

## UV-Decomposition

$$\begin{bmatrix} 5 & 2 & 4 & 4 & 3 \\ 3 & 1 & 2 & 4 & 1 \\ 2 & & 3 & 1 & 4 \\ 2 & 5 & 4 & 3 & 5 \\ 4 & 4 & 5 & 4 & \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \\ u_{31} & u_{32} \\ u_{41} & u_{42} \\ u_{51} & u_{52} \end{bmatrix} \times \begin{bmatrix} v_{11} & v_{12} & v_{13} & v_{14} & v_{15} \\ v_{21} & v_{22} & v_{23} & v_{24} & v_{25} \end{bmatrix}$$



Cold Start

FRANK & ERNEST® by Bob Thaves



© 1990 BY NEA INC THAVES 6-13

Explanation



Feedback



CARS

Context-Aware

Recommender Systems



WE CARRY OUR OWN CONTRACTS

\$99 DOWN DELIVERS TAX AND LICENSE WEEKLY

FINE USED

599

100% FINANCING  
ON APPROVED CREDIT

KB Recommendation

# PROs and CONs

## Content-based:

**PROs:** No hace falta comunidad, solo func. de distancia entre ítems

**CONs:** Content-descriptions, cold start para usuarios nuevos, no hay sorpresa.

## Collaborative Filtering (CF):

**PROs:** Serendipia de resultados, reconoce segmentos del mercado

**CONs:** Requiere feedback, cold start para usuarios nuevos e ítems.

## Knowledge-based:

**PROs:** Recomendaciones determinísticas, calidad controlada, no hay cold-start, remite al diálogo del vendedor.

**CONs:** Conocer el dominio del catálogo, no admite entorno dinámicos.



Sistema de Recomendación  
de Vinos



# Referencias

- **The Netflix prize.** <http://www.netflixprize.com/>
- *The Long Tail: Why the Future of Business is Selling Less of More*, by Chris Anderson.
- *Recommender Systems: An Introduction*, by Markus Zanker, Alexander Felfernig, Gerhard Friedrich.
- *Recommender Systems Handbook*, by Lior Rokach, Bracha Shapira.
- **The ACM Conference Series on Recommender Systems.** <http://recsys.acm.org/>
- *Recommendation Systems - Mining of Massive Datasets, Chap. 9*, by Anand Rajaraman, Jeff Ullman y Jure Leskovec.