

# Sistemas de recomendación

Presente y futuro de la web



Dpt. de Ingeniería del Software e Inteligencia Artificial  
Universidad Complutense de Madrid



- ❑ Almudena Ruiz Iniesta
  - ❑ Estudiante de doctorado
  - ❑ Tesis doctoral en sistemas de recomendación aplicados a entornos e-learning
  - ❑ [almudenari@fdi.ucm.es](mailto:almudenari@fdi.ucm.es)
  - ❑ Desp. 411



## ☐ Aula/Laboratorio

- ☐ 16 y 18 de abril, introducción teórica en aula 6 y 7
- ☐ 20 y 25, sesiones prácticas en laboratorio 6
- ☐ 27, aula 11 o laboratorio 6 (depende)

## ☐ Horario

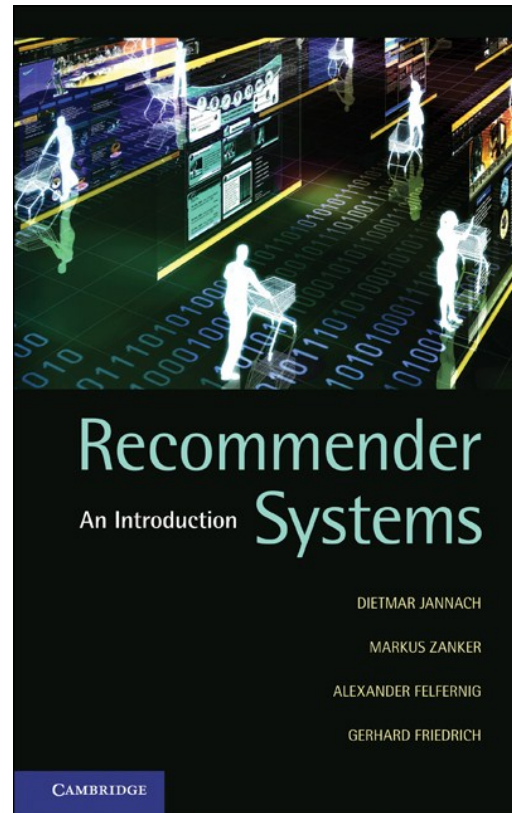
- ☐ de 14.00 a 17.00

- ☐ Conocer el mundo de los sistemas recomendadores
- ☐ Entender cuáles son las principales técnicas de diseño
- ☐ Ser capaz de diseñar mi propio recomendador

- ❑ Introducción a los sistemas de recomendación
- ❑ Técnicas estándar de recomendación
  - ❑ Filtrado colaborativo
  - ❑ Basadas en contenido
  - ❑ Basadas en conocimiento
  - ❑ Híbridas
- ❑ Sistemas de recomendación en la industria
- ❑ Sistemas de recomendación en e-learning
- ❑ Evaluación de los sistemas de recomendación
- ❑ Tendencias en sistemas de recomendación
  - ❑ Generación de explicaciones
  - ❑ Sistemas de recomendación sensibles al contexto
  - ❑ Recomendación y sistemas de social-tagging
- ❑ Desarrollo de prototipos

- ❑ Las notas que seguiremos están basadas en el libro y en el material proporcionado por los autores del mismo en

<http://www.recommenderbook.net/>





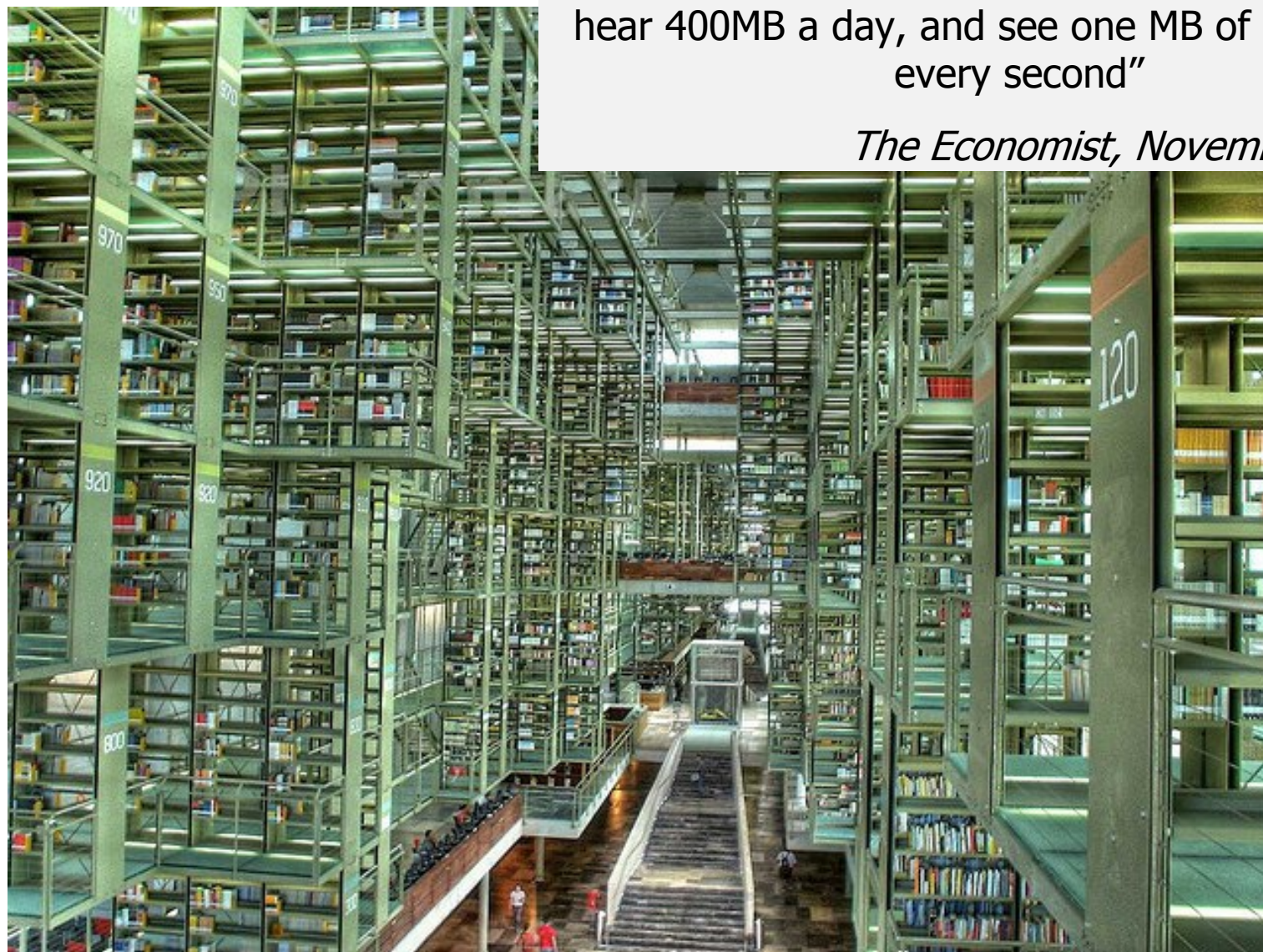
- ☐ Qué es un sistema de recomendación
- ☐ Campo de investigación
- ☐ Técnicas estándar de recomendación
  - ☐ Recomendación colaborativa
  - ☐ Recomendación basada en contenido
  - ☐ Recomendación basada en conocimiento
  - ☐ Recomendación híbrida



- ❑ Qué es un sistema de recomendación
- ❑ Campo de investigación
- ❑ Técnicas estándar de recomendación
  - ❑ Recomendación colaborativa
  - ❑ Recomendación basada en contenido
  - ❑ Recomendación basada en conocimiento
  - ❑ Recomendación híbrida

“People read around ten MB worth of material day,  
hear 400MB a day, and see one MB of information  
every second”

*The Economist, November 30, 2006*





**¿Qué película veo?**

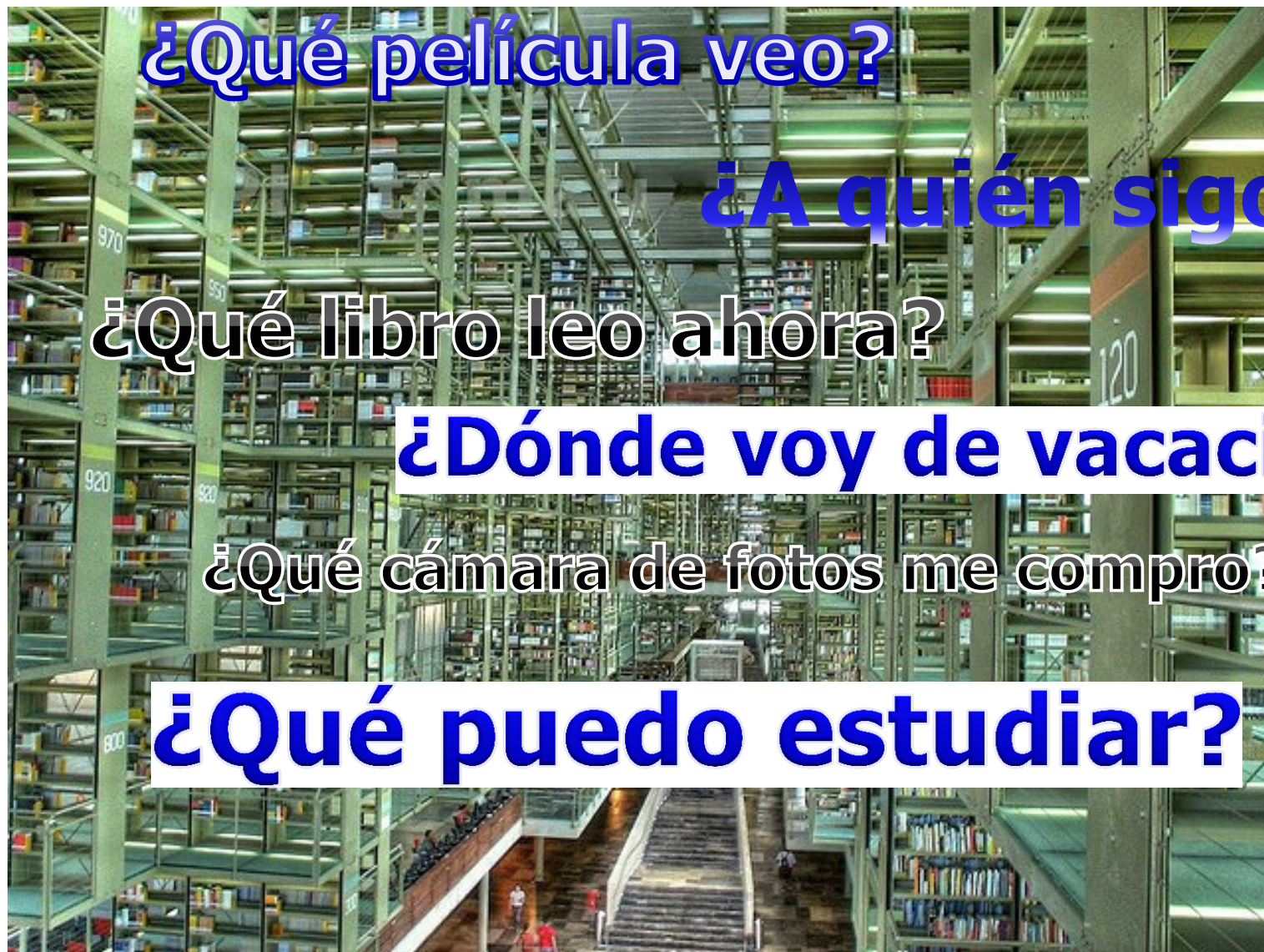
**¿A quién sigo?**

**¿Qué libro leo ahora?**

**¿Dónde voy de vacaciones?**

**¿Qué cámara de fotos me compro?**

**¿Qué puedo estudiar?**



- ☐ Conversaciones con amigos nos daban la respuesta
- ☐ Consejos con un experto
- ☐ En la era de la información recurrimos a la web

- ☐ ¿Nos pueden dar los buscadores la solución? ¿Google tiene la solución?
  - ☐ Hay que saber formular bien la pregunta...
  - ☐ No nos conoce, ¿cuáles son mis necesidades, mis gustos?

- ❑ Los recomendadores (RS) son sistemas que ayudan a emparejar usuarios con productos

*RS are software agents that elicit the interests and preferences of individual consumers [...]*

*and make recommendations accordingly.*

*They have the potential to support and improve the quality of the decisions consumers make while searching for and selecting products online.*

» (Xiao & Benbasat 2007)

- ❑ Cada usuario obtiene una lista de recomendación diferente
  - ❑ Los más valorados/consumidos/vistos no vale para todos los usuarios
- ❑ Requiere que
  - ❑ Los recomendadores tengan un modelo de usuario
    - ❑ Historial de visitas, likes,...

- ❑ Existen diferentes paradigmas de diseño
  - ❑ Basados en los datos que se van a utilizar
  - ❑ En la información proporcionada por el usuario
  - ❑ Características del dominio



- ☐ Qué es un sistema de recomendación
- ☐ Campo de investigación
- ☐ Técnicas estándar de recomendación
  - ☐ Recomendación colaborativa
  - ☐ Recomendación basada en contenido
  - ☐ Recomendación basada en conocimiento
  - ☐ Recomendación híbrida

- ☐ Multidisciplinario
  - ☐ gestionar conocimiento, creación de perfiles, interfaces de usuario, aspectos sociológicos y psicológicos, etc
- ☐ Relativamente joven
- ☐ Con muchas influencias (IR, ML, DM, etc.)
- ☐ Muy aplicado
- ☐ Todavía resulta fácil innovar



ACM Recommender Systems

Barcelona September 2010

$? + ? + ? = !$



20th international  
world wide web conference  
28th October - 1st November 2011  
Hyderabad, India

**SIGIR**  
Special Interest Group  
on Information Retrieval



## ☐ Workshops

- ☐ APRESW: Adaptation, Personalization and Recommendation in the Social-semantic Web (ESWC 2010)
- ☐ SRS: Social Recommender Systems (CSCW 2011)
- ☐ CARS: Context-Aware Recommender Systems (RecSys 2010)
- ☐ WPRRS: Web Personalization, Reputation and Recommender Systems (WI-IAT 2011)
- ☐ RSMEETDB: Recommender Systems meet Databases (DEXA 2011)
- ☐ SMUC: Search and Mining User-generated Contents (CIKM 2010)
- ☐ TEL: Recommender Systems for Technology Enhanced Learning (RecSys 2010)
- ☐ SWM: Social Web Mining (IJCAI 2011)
- ☐ ....

Inicio








Conecta

Descubre

#recsys

Tweets

Personas

Tendencias en Madrid · Cambiar

Di Natale

#PajotesDEP

#CuartoMilenio

#ChapatorraDEP

Morosini

Carmen Porter

Iker Jiménez

Cesc

Coachella

Enrique de Vicente

twitter

© 2012 Twitter

Sobre nosotros

Ayuda

Condiciones

Privacidad

Blog

Estado

Aplicaciones

Recursos

Trabajos

Publicidad

Negocios

Media

Desarrolladores

Resultados de #recsys

Tweets

Más relevantes / Todos

Kieran Hannon

@kieranhannon

8h

Discovery is also necessary [#recsys](#) RT [@TechCrunch](#): Riding The Third Wave of TV Transformation [tcm.ch/UaxWG](#) by [@ericella](#)

Daniel Gallego

@thanos\_malkav

14 de abr

Paper submitted for the 2012 Frontiers In Education Conference! [fie2012.org](#) [#phd](#) [#research](#) [#recsys](#)

Eric Wilson

@peripateric

13 de abr

Why Netflix Never Implemented The Algorithm That Won The Netflix \$1 Million Challenge [zite.to/HCusc1](#) [#recsys](#)

David Elsweller

@delsweil

13 de abr

Our event [#recsys](#) is v. accurate at predicting events people would like to visit, but is still very rarely used. Any literature on this?

toto

@totopampin

13 de abr

[tecnologia elpais.com/tecnologia/201...](#) aerolíneas y recomendaciones sociales [#recsys](#)

Alan Said

@alansaid

11 de abr

and finally, crc of [#umap2012](#) paper titled "Users and Noise: Estimating the Magic Barrier of Recommendation Systems" submitted [#recsys](#)

danielequercia

@danielequercia

11 de abr

Yes! :)RT [@zenogantner](#): RT [@usabart](#): yet another user-centric [#recsys2012](#) paper. [#recsys](#) will become a psychology conference! [#deaddonkey](#)

Simon Doods

@sidooms

10 de abr

Downloaded and installed [@mymedialite](#) 3.0 [ismll.uni-hildesheim.de/mymedialite/do...](#) [#recsys](#)

Bart Knijnenburg

@usabart

10 de abr

Presubmission proofreading yet another user-centric [#recsys2012](#) paper. At this rate [#recsys](#) will become a psychology conference! [#deaddonkey](#)

## ☐ Datasets

- ☐ Netflix (<http://narod.ru/disk/7133213001/netflix.7z.html>)
- ☐ MovieLens (<http://www.grouplens.org/>)
- ☐ Jester (<http://www.ieor.berkeley.edu/~goldberg/jester-data/>)
- ☐ Book-crossing (<http://www.informatik.uni-freiburg.de/~ctiegle/BX/>)
- ☐ KDD Cup (Yahoo!, <http://kddcup.yahoo.com/>)

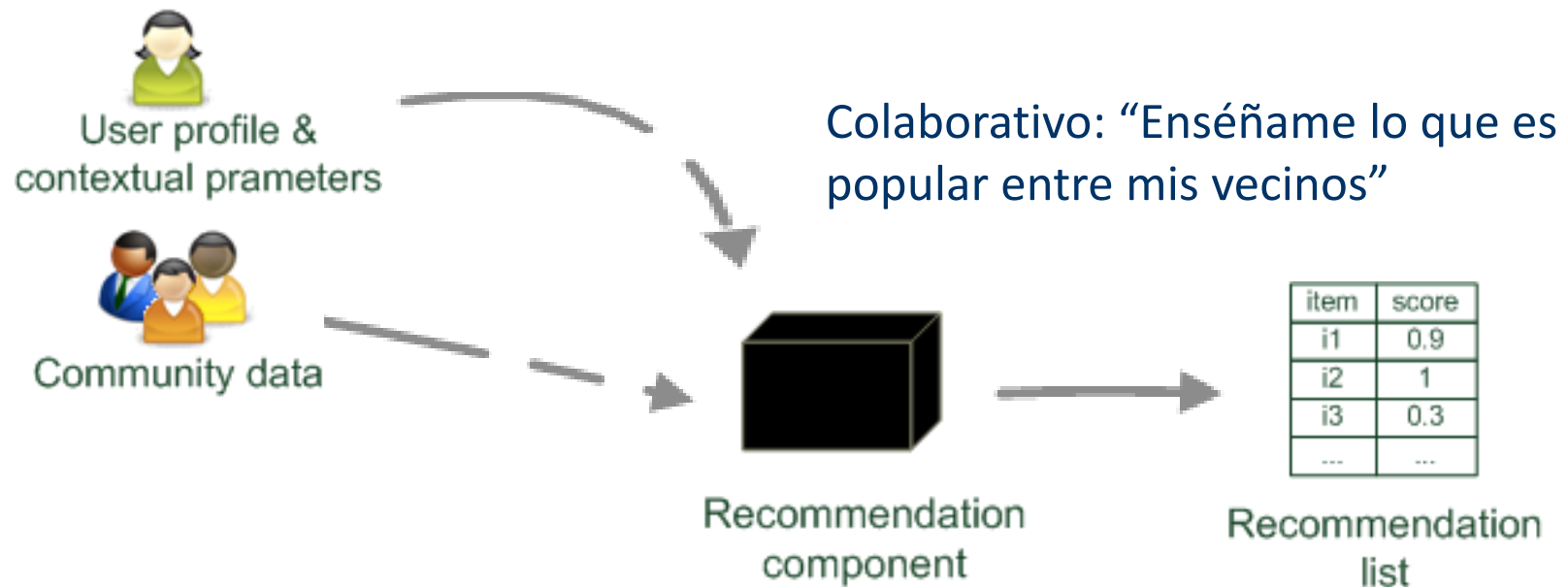
## ☐ Se puede usar información de multitud de sitios sociales

- ☐ Facebook (Like)
- ☐ Twitter (favoritos, RTs)
- ☐ Delicious (recomendación de etiquetas)
- ☐ Quora (recomendar preguntas de interés)

- ❑ Qué es un sistema de recomendación
- ❑ Campo de investigación
- ❑ **Técnicas estándar de recomendación**
  - ❑ Recomendación colaborativa
  - ❑ Recomendación basada en contenido
  - ❑ Recomendación basada en conocimiento
  - ❑ Recomendación híbrida

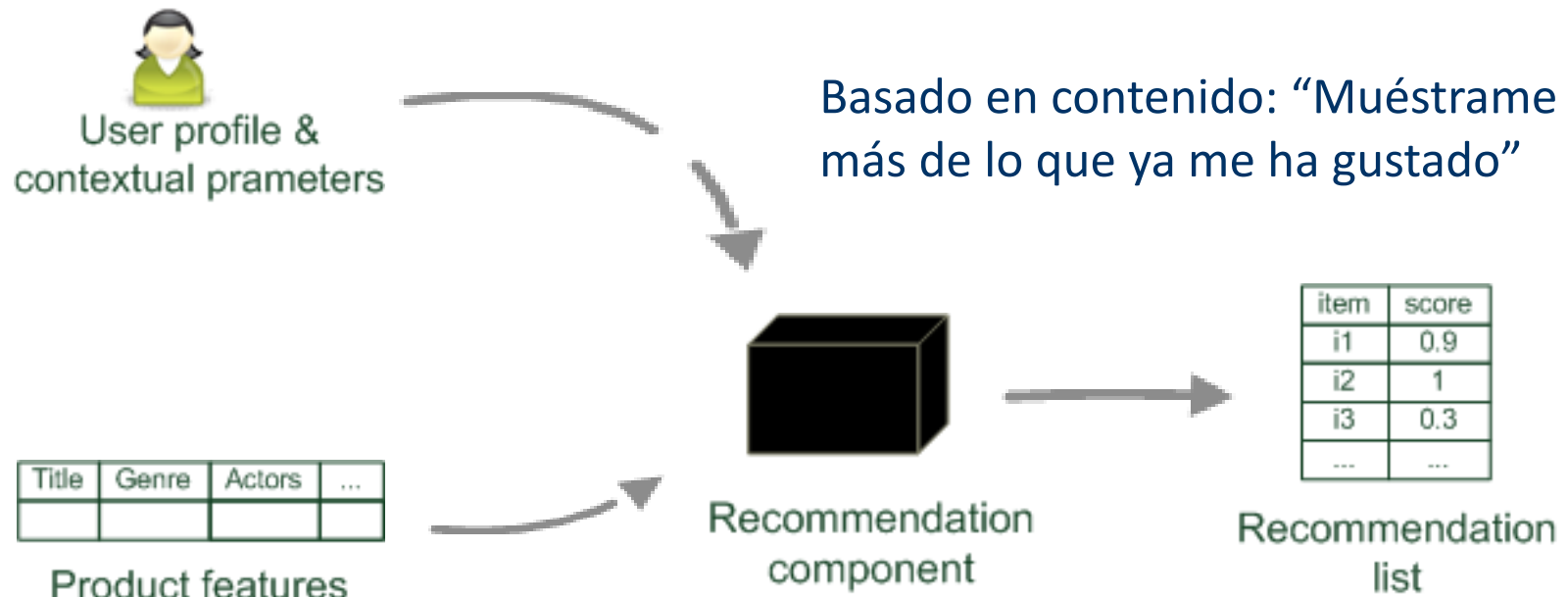
## ❑ Recomendación colaborativa

- ❑ Si los usuarios han compartido algunos de sus intereses en el pasado, tendrán gustos similares en el futuro
  - ❑ A y B tienen una historia parecida. A compra un libro que B no ha visto. Lo más razonable es proponer ese libro también a B

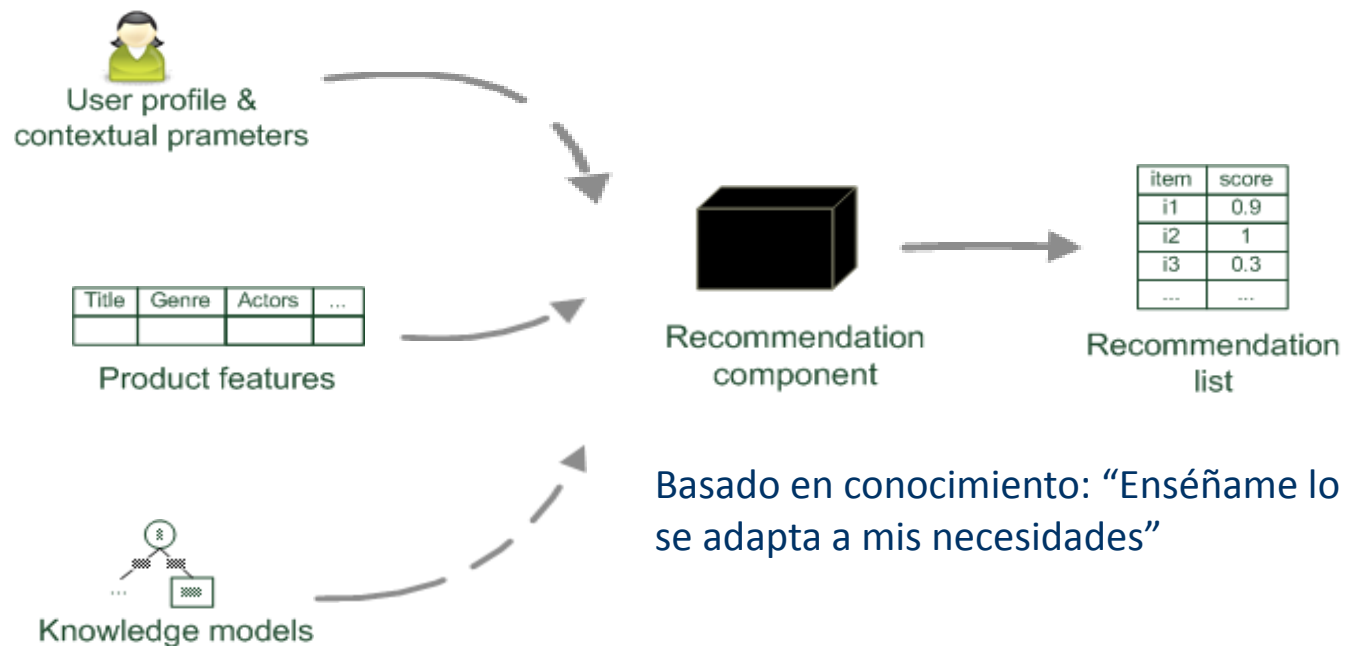




- ❑ Recomendación basada en contenido
  - ❑ Partiendo de una base de elementos con descripción (características) y un perfil con asignaciones de importancia a estas características.  
Emparejar aquellos elementos que mejor cumplan las preferencias del usuario

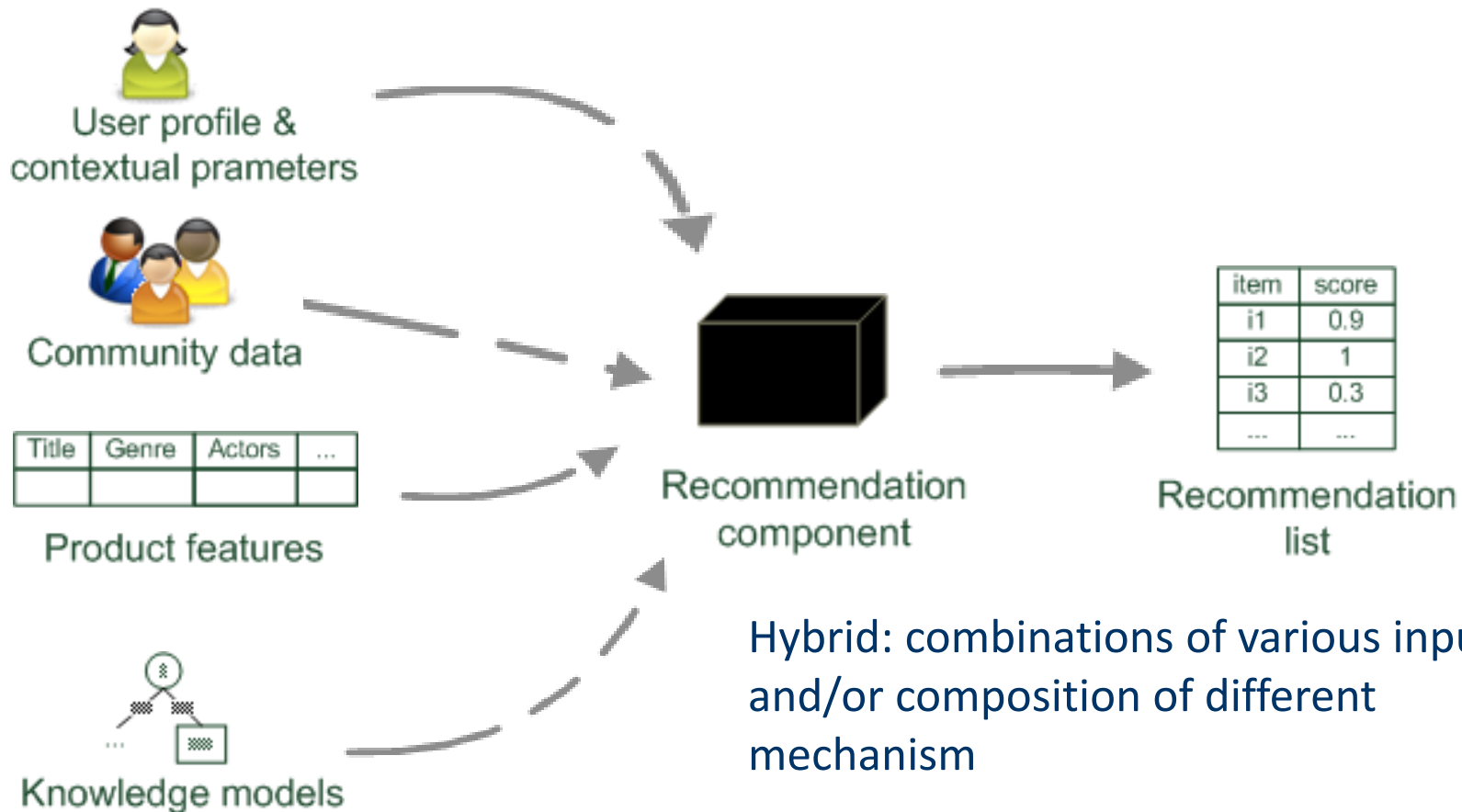


- ❑ Recomendación basada en conocimiento
  - ❑ En muchas ocasiones nos encontraremos con usuarios novatos. No tenemos historial ni preferencias
  - ❑ Recomendadores que están más centrados en el dominio de aplicación
  - ❑ Necesitan información adicional para poder realizar las recomendaciones
    - ❑ Información sobre el usuario y/o sobre los elementos



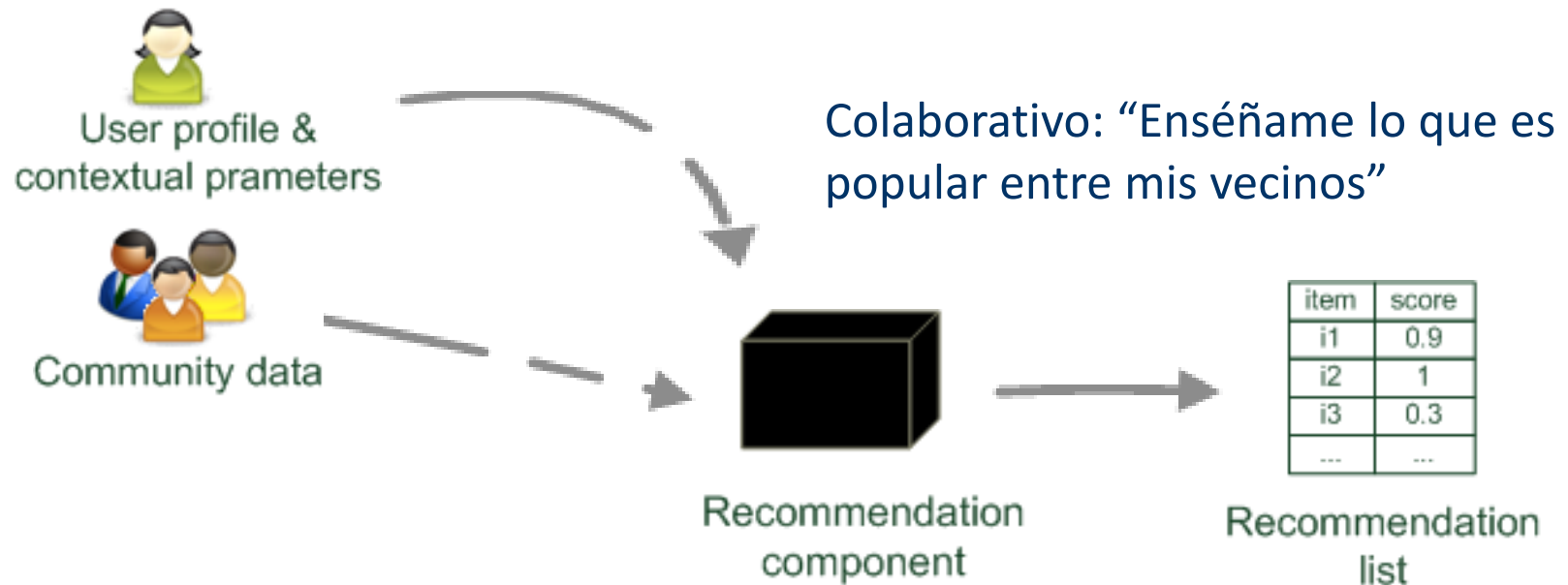
## ❑ Recomendación híbrida

- ❑ Combinación de varias de las técnicas anteriores
- ❑ Obtener “mejores” recomendaciones



- ❑ Qué es un sistema de recomendación
- ❑ Campo de investigación
- ❑ Técnicas estándar de recomendación
  - ❑ Recomendación colaborativa
  - ❑ Recomendación basada en contenido
  - ❑ Recomendación basada en conocimiento
  - ❑ Recomendación híbrida

- ❑ También llamada Filtrado Colaborativo (CF)
- ❑ Es la técnica más empleada
  - ❑ Sus ventajas, rendimiento y limitaciones son bien conocidas
- ❑ Idea
  - ❑ Utilizar lo que dice la gente para recomendar elementos



- ❑ Suposición fundamental y motivación
  - ❑ Los usuarios asignan puntuaciones a los elementos disponibles (implícitas o explícitas)
  - ❑ Usuarios que han tenido gustos similares en el pasado, tendrán gustos similares en el futuro
- ❑ Entrada
  - ❑ Matriz de usuarios-elementos
- ❑ Salida
  - ❑ Una predicción numérica sobre el grado de interés de cada elemento
  - ❑ Una lista de  $n$  elementos recomendados
    - ❑ Que por supuesto no contiene aquellos ya vistos por el usuario

## ☐ La técnica básica

☐ Dado un “usuario activo” (María) y un elemento  $i$  no visto por el usuario

☐ Encontrar el grupo de usuarios (vecinos cercanos) a los que les han gustado los mismos elementos que a María en el pasado y que han valorado el elemento  $i$

☐ Utilizar las valoraciones para predecir si a María le gustará  $i$

☐ Repetir esto para todos los elementos no vistos por María y recomendar los mejores

## ☐ Suposición fundamental e idea

☐ Si los usuarios han tenido gustos similares en el pasado, tendrás gustos similares en el futuro

☐ Las preferencias/gustos de un usuario permanecen estables y consistentes en el tiempo

## □ Ejemplo

- Una base de datos de valoraciones para el usuario actual, María, y las valoraciones de otros usuarios:

	Item1	Item2	Item3	Item4	Item5
María	5	3	4	4	?
User1	3	1	2	3	3
User2	4	3	4	3	5
User3	3	3	1	5	4
User4	1	5	5	2	1

- Determinar si a María le gustará o no el Item5, el cual todavía no ha visto o no ha valorado



- ❑ Algunas preguntas
  - ❑ ¿Cómo medimos la similitud?
  - ❑ ¿Cuántos vecinos tenemos en cuenta?
  - ❑ ¿Cómo generamos una predicción a partir de las valoraciones de los vecinos?

	Item1	Item2	Item3	Item4	Item5
María	5	3	4	4	?
User1	3	1	2	3	3
User2	4	3	4	3	5
User3	3	3	1	5	4
User4	1	5	5	2	1

- Una medida de similitud popular utilizada en CF basado en usuario es la Correlación de Pearson
  - $a, b$  : usuarios
  - $r_{a,p}$  : valoración del usuario  $a$  para el elemento  $p$
  - $P$  : conjunto de elementos que han sido valorados por  $a$  y  $b$
  - Los valores posibles de similitud van desde  $-1$  hasta  $1$

$$sim(a, b) = \frac{\sum_{p \in P} (r_{a,p} - \bar{r}_a)(r_{b,p} - \bar{r}_b)}{\sqrt{\sum_{p \in P} (r_{a,p} - \bar{r}_a)^2} \sqrt{\sum_{p \in P} (r_{b,p} - \bar{r}_b)^2}}$$

- ❑ Una medida de similitud popular utilizada en CF basado en usuario es la Correlación de Pearson
  - ❑  $a, b$  : usuarios
  - ❑  $r_{a,p}$  : valoración del usuario  $a$  para el elemento  $p$
  - ❑  $P$  : conjunto de elementos que han sido valorados por  $a$  y  $b$
  - ❑ Los valores posibles de similitud van desde  $-1$  hasta  $1$

	Item1	Item2	Item3	Item4	Item5	
María	5	3	4	4	?	
User1	3	1	2	3	3	sim = 0,85
User2	4	3	4	3	5	sim = 0,00
User3	3	3	1	5	4	sim = 0,70
User4	1	5	5	2	1	sim = -0,79

- ❑ Una función de predicción común

$$\mathit{pred}(a, p) = \overline{r_a} + \frac{\sum_{b \in N} \mathit{sim}(a, b) * (r_{b,p} - \overline{r_b})}{\sum_{b \in N} \mathit{sim}(a, b)}$$

- ❑ Calcular, si las calificaciones de los vecinos para el elemento no visto  $i$  son más altos o más bajos que la media
- ❑ Combinar las diferencias de valoración - utilizar la similitud con  $a$  como un peso

- ❑ No todas las valoraciones de los vecinos podrían ser igual de "valiosas"
  - ❑ Estar de acuerdo sobre elementos que gustan a todo el mundo no es tan relevante como estar de acuerdo en elementos menos comunes
  - ❑ Posible solución: Dar más peso a los elementos que tienen una mayor varianza
- ❑ Valorar el número de elementos co-valorados
  - ❑ Utilizar "ponderación significativa", por ejemplo, reduciendo el peso de forma lineal cuando el número de elementos co-valorados es bajo
- ❑ Caso de amplificación
  - ❑ Intuición: Dar más peso a los vecinos "muy similares", es decir, donde el valor de similitud es cercano a 1
- ❑ Selección del vecindario
  - ❑ Utilizar umbral de similitud o un número limitado de vecinos

## ❑ Idea general

- ❑ Utilizar la similitud entre elementos (y no entre usuarios) para hacer las predicciones

## ❑ Ejemplo

- ❑ Buscar elementos que son similares al Item5
- ❑ Tomar las valoraciones de María para estos elementos y así predecir la valoración del Item5

	Item1	Item2	Item3	Item4	Item5
María	5	3	4	4	?
User1	3	1	2	3	3
User2	4	3	4	3	5
User3	3	3	1	5	4
User4	1	5	5	2	1

- ❑ Produce buenos resultados en las aproximaciones basadas en elemento
- ❑ Las valoraciones son vistas como un vector n-dimensional
- ❑ La similitud se calcula en base al ángulo entre los dos vectores

$$sim(\vec{a}, \vec{b}) = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| * |\vec{b}|}$$

- ❑ Similitud del coseno ajustada
  - ❑ Tener en cuenta la media de las valoraciones otorgadas por un usuario, transformar las valoraciones originales
  - ❑  $U$ : conjunto de usuarios que han valorados  $a$  y  $b$

$$sim(\vec{a}, \vec{b}) = \frac{\sum_{u \in U} (r_{u,a} - \bar{r}_u)(r_{u,b} - \bar{r}_u)}{\sqrt{\sum_{u \in U} (r_{u,a} - \bar{r}_u)^2} \sqrt{\sum_{u \in U} (r_{u,b} - \bar{r}_u)^2}}$$

- ❑ La función de predicción común

$$\mathit{pred}(u, p) = \frac{\sum_{i \in \mathit{ratedItem}(u)} \mathit{sim}(i, p) * r_{u,i}}{\sum_{i \in \mathit{ratedItem}(u)} \mathit{sim}(i, p)}$$



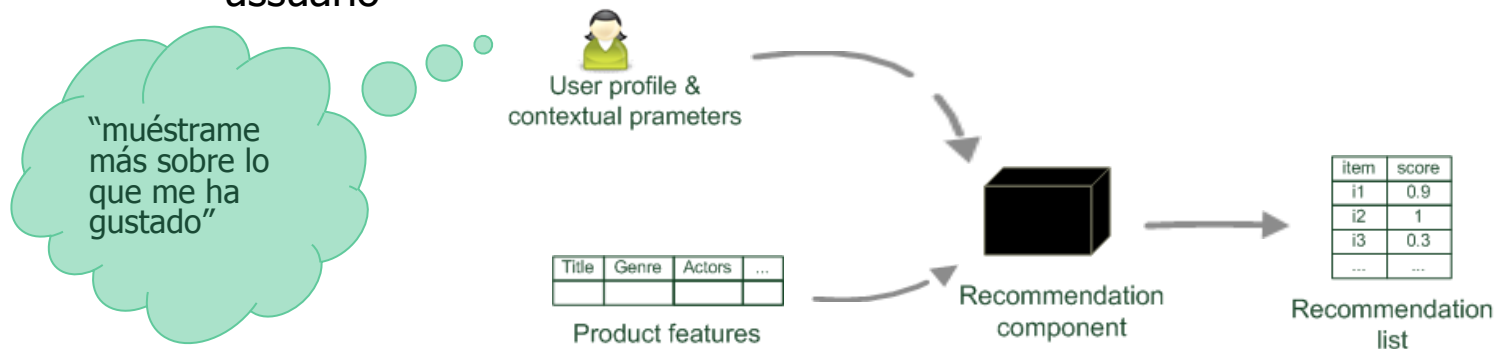
- ❑ Probablemente las valoraciones más precisas
- ❑ Habitualmente se utiliza escalas Likert con valores que van de 1 a 5, ó de 1 a 7
  - ❑ Tema de investigación
- ❑ Principales problemas
  - ❑ Los usuarios nos suelen valorar muchos elementos
    - ❑ El número de las valoraciones disponibles es pequeño → matrices de valoraciones poco densas → recomendaciones de baja calidad
  - ❑ ¿Cómo estimular que un usuario valore muchos elementos?

- ☐ Por lo general recogidos por la tienda web o aplicación en la que se incluye el sistema de recomendación
- ☐ Cuando un usuario compra/visualiza un elemento, muchos recomendadores interpretan este comportamiento con una valoración positiva
- ☐ Clicks, páginas vistas, tiempo empleado en algunas páginas, descargas ...
- ☐ Las valoraciones implícitas pueden recogerse en cualquier momento y no requieren un esfuerzo adicional por parte del usuario
- ☐ Principal problema
  - ☐ Podemos no estar seguros de si el comportamiento se ha interpretado correctamente
  - ☐ Por ejemplo, a un usuario puede no gustarle un libro que ha comprado ya que quizá era un regalo...
- ☐ Las valoraciones implícitas pueden utilizarse como complemento de las explícitas

- ❑ Problema del arranque-en-frío
  - ❑ ¿Cómo recomendar nuevos elementos? ¿Qué recomendar a los nuevos usuarios?
- ❑ Enfoques directos
  - ❑ Preguntar/forzar a los usuarios a valorar un conjunto de elementos
  - ❑ Utilice otro método (por ejemplo, basada en el contenido, demográficas o, simplemente, no personalizado) en la fase inicial
  - ❑ Valoración por defecto: asignar valores por defecto a los elementos que sólo uno de los dos usuarios ha valorado
- ❑ Alternativas
  - ❑ Utilizar mejores algoritmos (más allá de las aproximaciones de los vecinos cercanos)
  - ❑ Ejemplo
    - ❑ En los enfoques del vecino más cercano, el conjunto de los vecinos lo suficientemente similares podría ser demasiado pequeño para hacer buenas predicciones
    - ❑ Asumir “transitividad” entre vecinos

- ❑ Qué es un sistema de recomendación
- ❑ Campo de investigación
- ❑ Técnicas estándar de recomendación
  - ❑ Recomendación colaborativa
  - ❑ Recomendación basada en contenido
  - ❑ Recomendación basada en conocimiento
  - ❑ Recomendación híbrida

- ❑ Aunque los métodos de CF no requerían ninguna información sobre los elementos,
  - ❑ Debería ser razonable utilizar esta información
  - ❑ Recomendar novelas fantásticas a aquellas personas que en el pasado les gustaron las novelas fantásticas
- ❑ Lo que necesitamos
  - ❑ Información sobre los elementos disponibles (contenido), p.ej: género
  - ❑ Algun tipo de perfil de usuario que describa cuáles son sus gustos (preferencias)
- ❑ La tarea
  - ❑ Aprender las preferencias del usuario
  - ❑ Localizar/Recomendar aquellos elementos que son similares a los gustos del usuario



- ❑ La mayoría de los CB han sido aplicadas en la recomendación de textos
  - ❑ Como páginas web o mensajes de un grupo de noticias
- ❑ El contenido de los elementos puede ser representado como un documento de texto
  - ❑ Con descripciones textuales de sus características básicas
  - ❑ Estructurado: Cada elemento es descrito por el mismo conjunto de atributos



Title	Genre	Author	Type	Price	Keywords
The Night of the Gun	Memoir	David Carr	Paperback	29.90	Press and journalism, drug addiction, personal memoirs, New York
The Lace Reader	Fiction, Mystery	Brunonia Barry	Hardcover	49.90	American contemporary fiction, detective, historical
Into the Fire	Romance, Suspense	Suzanne Brockmann	Hardcover	45.90	American fiction, murder, neo-Nazism

- ❑ Desestructurado: descripciones de texto libre

## ■ Representación de los elementos

Title	Genre	Author	Type	Price	Keywords
The Night of the Gun	Memoir	David Carr	Paperback	29.90	Press and journalism, drug addiction, personal memoirs, New York
The Lace Reader	Fiction, Mystery	Brunonia Barry	Hardcover	49.90	American contemporary fiction, detective, historical
Into the Fire	Romance, Suspense	Suzanne Brockmann	Hardcover	45.90	American fiction, murder, neo-Nazism

## ■ Perfil de usuario

Title	Genre	Author	Type	Price	Keywords
...	Fiction	Brunonia, Barry, Ken Follett	Paperback	25.65	Detective, murder, New York

$keywords(b_j)$   
describe el libro  $b_j$   
con un conjunto de palabras clave

## □ Aproximación básica

- Calcular la similitud de los elementos no vistos en el perfil de usuario a partir del solapamiento de palabras clave (keyword) (p.ej, utilizando el coeficiente de Dice)

- O utilizar y combinar distintas métricas

$$\rightarrow \frac{2 \times |keywords(b_i) \cap keywords(b_j)|}{|keywords(b_i)| + |keywords(b_j)|}$$



- ❑ Extracción del contenido
- ❑ La representación simple tiene sus problemas
  - ❑ En particular cuando se extraen automáticamente como
    - ❑ No todas las palabras tienen la misma importancia en el texto
    - ❑ Los documentos más largos tienen una mayor probabilidad de tener mayor superposición con el perfil del usuario
- ❑ Métrica estándar: TF-IDF
  - ❑ Representa los documentos en un espacio multi-dimensional Euclídeo
    - ❑ Vector de términos ponderado
  - ❑ TF: Mide cómo de frecuente aparece un término (densidad en un documento)
    - ❑ Asumiendo que términos más importantes aparecen más frecuentemente
    - ❑ Se debe normalizar para tener en cuenta la longitud del documento
  - ❑ IDF: Intenta reducir el peso otorgado a los términos que aparecen en todos los documentos (palabras sin relevancia)



- ❑ Dada una palabra clave  $i$  y un documento  $j$
- ❑  $TF(i, j)$ 
  - ❑ Frecuencia de la palabra clave  $i$  en el documento  $j$
- ❑  $IDF(i)$ 
  - ❑ Frecuencia inversa calculada como  $IDF(i) = \log \frac{N}{n(i)}$ 
    - ❑  $N$  : número de todos los documentos que pueden ser recomendados
    - ❑  $n(i)$  : número de documentos en  $N$  en los que la palabra clave  $i$  aparece
- ❑  $TF - IDF$ 
  - ❑  $TF-IDF(i, j) = TF(i, j) * IDF(i)$

## □ Frecuencia del término

□ Cada documento es un vector en  $\mathbb{N}^{|v|}$

	Antony and Cleopatra	Julius Caesar	The Tempest	Hamlet	Othello	Macbeth
Antony	157	73	0	0	0	0
Brutus	4	157	0	1	0	0
Caesar	232	227	0	2	1	1
Calpurnia	0	10	0	0	0	0
Cleopatra	57	0	0	0	0	0
mercy	1.51	0	3	5	5	1
worser	1.37	0	1	1	1	0

Vector  $v$  con dimensión  $|v| = 7$

## TF-IDF

- Cada documento es ahora representado con un valor real del vector de pesos  $TF-IDF \in \mathbb{R}^{|v|}$

		Antony and Cleopatra	Julius Caesar	The Tempest	Hamlet	Othello	Macbeth
Antony	157	73	0	0	0	0	
Brutus	4						
Caesar	232						
Calpurnia	0						
Cleopatra	57						
mercy	1.5						
worser	1.3						

	Antony and Cleopatra	Julius Caesar	The Tempest	Hamlet	Othello	Macbeth
Antony	5.25	3.18	0	0	0	0.35
Brutus	1.21	6.1	0	1	0	0
Caesar	8.59	2.54	0	1.51	0.25	0
Calpurnia	0	1.54	0	0	0	0
Cleopatra	2.85	0	0	0	0	0
mercy	1.51	0	1.9	0.12	5.25	0.88
worser	1.37	0	0.11	4.15	0.25	1.95

Ejemplo tomado de <http://informationretrieval.org>

- ❑ Los vectores normalmente son largos y dispersos
- ❑ Eliminar las palabras de parada (stop-words)
  - ❑ Aparecen en todos los documentos
  - ❑ "a", "the", "on", ...
- ❑ Reducir a la raíz
- ❑ Limitar el tamaño
  - ❑ Utilizar sólo las n palabras más representativas para así eliminar el "ruido"
  - ❑ Por ejemplo, utilizar 100 palabras

- ❑ Utilizar conocimiento léxico para utilizar métodos más elaborados de selección
  - ❑ Eliminar palabras que no son relevantes en el dominio
- ❑ Detección de frases y términos
  - ❑ Las frases son más descriptivas que las palabras simples
  - ❑ Por ejemplo "United Nations"
- ❑ Limitaciones
  - ❑ El significado semántico se pierde
  - ❑ Ejemplo: utilizar una palabra en un contexto negativo
    - ❑ "there is nothing on the menu that a vegetarian would like.."
    - ❑ La palabra "vegetarian" recibirá un peso alto
    - ❑ ➡ una coincidencia no deseada para un usuario interesado en restaurantes vegetarianos

- ❑ La similitud se calcula en base al ángulo entre los dos vectores

- ❑  $sim(\vec{a}, \vec{b}) = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| * |\vec{b}|}$

- ❑ Similitud del coseno ajustada

- ❑ Tener en cuenta la media de las valoraciones otorgadas por un usuario, transformar las valoraciones originales

- ❑  $U$ : conjunto de usuarios que han valorados  $a$  y  $b$

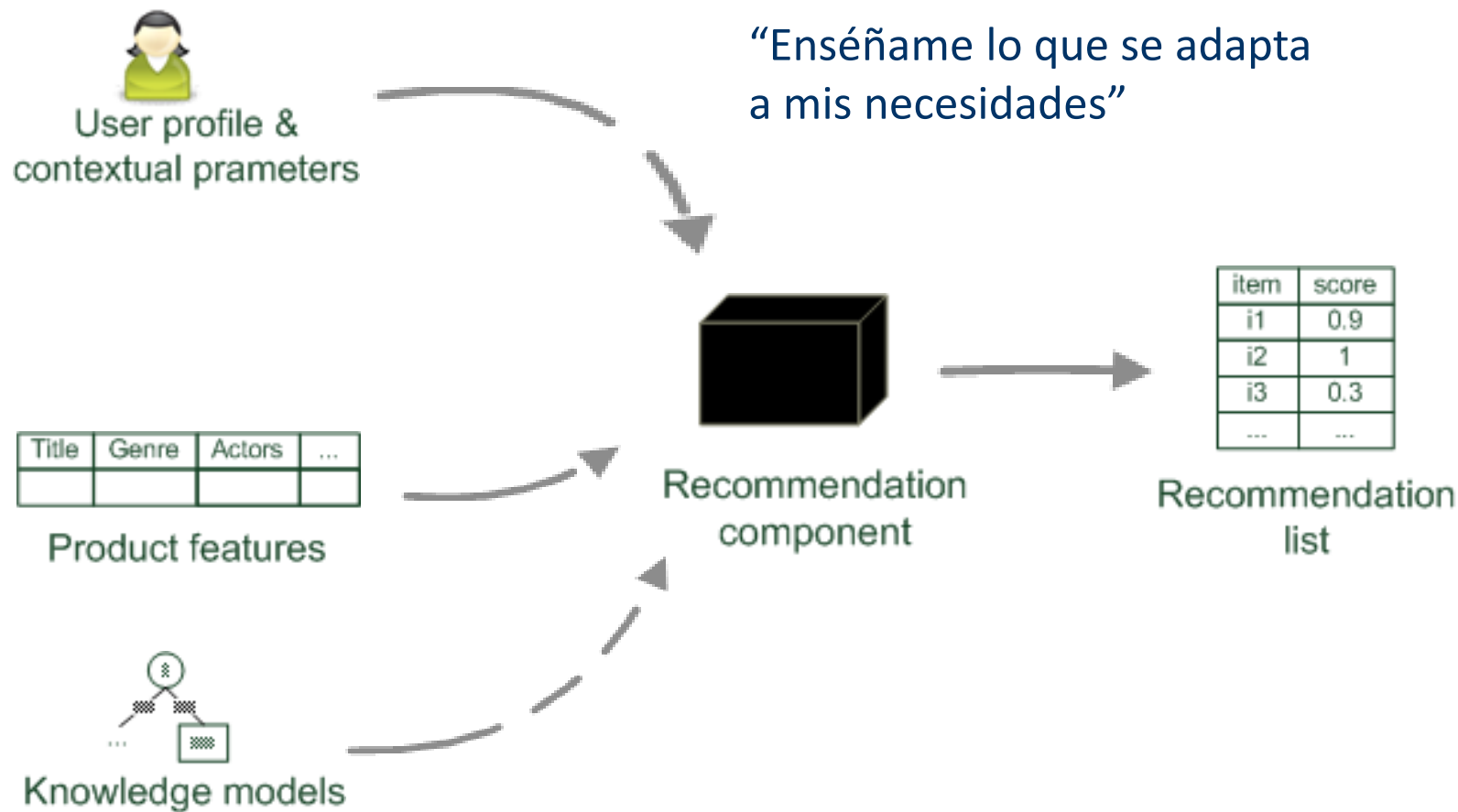
- ❑  $sim(\vec{a}, \vec{b}) = \frac{\sum_{u \in U} (r_{u,a} - \bar{r}_u)(r_{u,b} - \bar{r}_u)}{\sqrt{\sum_{u \in U} (r_{u,a} - \bar{r}_u)^2} \sqrt{\sum_{u \in U} (r_{u,b} - \bar{r}_u)^2}}$

- ❑ Método más sencillo: vecinos más cercanos
  - ❑ Dado un conjunto de documentos  $D$  valorados por el usuario (me gusta/no me gusta)
    - ❑ Explícitamente o Implícitamente
  - ❑ Encontrar los  $n$  vecinos más cercanos de un elemento no visto  $i$  en  $D$ 
    - ❑ Utilizar métricas de similitud para calcular la similitud de dos documentos
  - ❑ Utilizar estos vecinos para predecir la valoración de  $i$ 
    - ❑ Ejemplo.  $k = 5$  elementos más similares a  $i$   
4 de los  $k$  elementos le gustaron al usuario → el elemento  $i$  le gustará

- ❑ Existen otros métodos
  - ❑ Recuperación basada en consulta: el método de Rocchio
  - ❑ Métodos probabilísticos
  - ❑ Métodos de clasificación lineal y de aprendizaje máquina
  - ❑ Métodos de decisión
  - ❑ Selección de características



- ☐ Qué es un sistema de recomendación
- ☐ Campo de investigación
- ☐ Técnicas estándar de recomendación
  - ☐ Recomendación colaborativa
  - ☐ Recomendación basada en contenido
  - ☐ Recomendación basada en conocimiento
  - ☐ Recomendación híbrida



- ❑ Elementos con un bajo número de valoraciones



- ❑ El tiempo transcurrido juega un papel importante
  - ❑ Para determinados productos las valoraciones tienen vida limitada
  - ❑ El estilo de vida y las situaciones del usuario cambian
- ❑ Los usuarios/clientes quieren definir sus requisitos explícitamente
  - ❑ “El coche debe ser de color negro”

- ❑ Basado en restricciones (Constraint-based)
  - ❑ A partir de un conjunto de reglas de recomendación explícitamente definidas
  - ❑ Satisfacen todas las reglas de recomendación
- ❑ Basado en casos (Case-based)
  - ❑ Basado en diferentes métricas de similitud
  - ❑ Recuperar elementos que son similares a los requisitos especificados
- ❑ Ambas aproximaciones son similares en el proceso de recomendación: conversacional
  - ❑ El usuario especifica sus requisitos
  - ❑ El sistema trata de identificar soluciones
  - ❑ Si no se encuentra una solución, el usuario debe “cambiar” sus requisitos

- ❑ Base de conocimiento
  - ❑ Por lo general media entre el modelo de usuario y las características de los elementos
  - ❑ Variables
    - ❑ Características de usuario (requisitos), Características del elemento (Catálogo)
  - ❑ Conjunto de restricciones
- ❑ Conseguir un conjunto de elementos recomendables
  - ❑ Cumplimentar un conjunto de restricciones (aplicables)
  - ❑ La posibilidad de aplicar las restricciones depende del modelo de usuario actual
  - ❑ Las explicaciones otorgan transparencia a la recomendación

- ❑ Encontrar un conjunto de requisitos de usuario de tal manera que un subconjunto de elementos cumpla todas las restricciones
  - ❑ Preguntar al usuario qué requisitos pueden estar relajados/modificados de tal manera que existen algunos elementos que no incumplan ninguna restricción
- ❑ Encontrar un subconjunto de elementos que satisfagan el máximo número posible de restricciones
- ❑ Ordenar los elementos de acuerdo a los pesos otorgados a las restricciones

- ❑ Seleccionar elementos del siguiente catálogo que cumplan los requisitos del usuario

id	price(€)	mpix	opt-zoom	LCD-size	movies	sound	waterproof
P <sub>1</sub>	148	8.0	4×	2.5	no	no	yes
P <sub>2</sub>	182	8.0	5×	2.7	yes	yes	no
P <sub>3</sub>	189	8.0	10×	2.5	yes	yes	no
P <sub>4</sub>	196	10.0	12×	2.7	yes	no	yes
P <sub>5</sub>	151	7.1	3×	3.0	yes	yes	no
P <sub>6</sub>	199	9.0	3×	3.0	yes	yes	no
P <sub>7</sub>	259	10.0	3×	3.0	yes	yes	no
P <sub>8</sub>	278	9.1	10×	3.0	yes	yes	yes

- ❑ Los requisitos del usuario pueden ser
  - ❑ "el precio debe ser menor que 300 €"
  - ❑ "la cámara debe ser idónea para fotografía deportiva"

- Representación del problema como una tupla

$$CSP(X_I \cup X_U, SRS \cup KB \cup I)$$

- Def.

- $X_I, X_U$ : Variables que describen el modelo de usuario y de producto
- KB: Base de conocimiento con restricciones propias del dominio (p.e. **si** propósito=para viajes/ **entonces** longitud focal mínima < 28mm)
- SRS: Requisitos de usuario (p.e. propósito = para viajes)
- I: Catálogo de productos

- Solución: La asignación para la tupla  $\forall x \in X_I (x = v) \in \theta \wedge v \in dom(x)$

*s.t.*  $SRS \cup KB \cup I \cup \theta$  es posible



- ❑ El usuario especifica sus preferencias iniciales
  - ❑ Todas de una vez o de manera incremental con un asistente
- ❑ Al usuario se le presentan un conjunto de elementos que cumplan sus preferencias
  - ❑ Con una explicación sobre por qué es recomendado
- ❑ El usuario debe revisar sus requisitos
  - ❑ Mirar soluciones alternativas
  - ❑ Reducir el número de elementos que coinciden con sus preferencias

- ☐ Ayudar al usuario a elegir una alternativa razonable
  - ☐ No está seguro de qué opción seleccionar
  - ☐ Simplemente no tenemos conocimiento suficiente del dominio para elegir
- ☐ Tipos de valores por defecto
  - ☐ Estáticos
  - ☐ Dependientes
  - ☐ Derivados
- ☐ Selección de la siguiente pregunta
  - ☐ La mayoría de los usuarios no están interesados en especificar valores para todas las propiedades
  - ☐ El sistema identifica qué propiedades pueden ser interesantes para el usuario y que asigne valores sólo a estas

- ☐ "No se ha encontrado solución"
- ☐ Relajación de las restricciones
  - ☐ El objetivo es identificar qué restricciones del conjunto inicial pueden ser relajadas
  - ☐ Hasta que se encuentre una solución
- ☐ El usuario puede modificar sus propuestas

- ❑ Teoría multi-atributo
  - ❑ Cada elemento es evaluado de acuerdo a un conjunto predefinido de utilidades
- ❑ Ejemplo: calidad y económico son dimensiones en el campo de las cámaras digitales

id	value	quality	economy
price	≤250	5	10
	>250	10	5
mpix	≤8	4	10
	>8	10	6
opt-zoom	≤9	6	9
	>9	10	6
LCD-size	≤2.7	6	10
	>2.7	9	5
movies	Yes	10	7
	no	3	10
sound	Yes	10	8
	no	7	10
waterproof	Yes	10	6
	no	8	10

- ❑ Los elementos son recuperados utilizando métricas de similitud

$$\text{similarity}(p, REQ) = \frac{\sum_{r \in REQ} w_r * \text{sim}(p, r)}{\sum_{r \in REQ} w_r}$$

- ❑ Def.

- ❑  $\text{sim}(p, r)$  representa para cada atributo de un elemento  $\phi(p)$  la distancia que existe con el requisito de usuario  $r \in REQ$ .

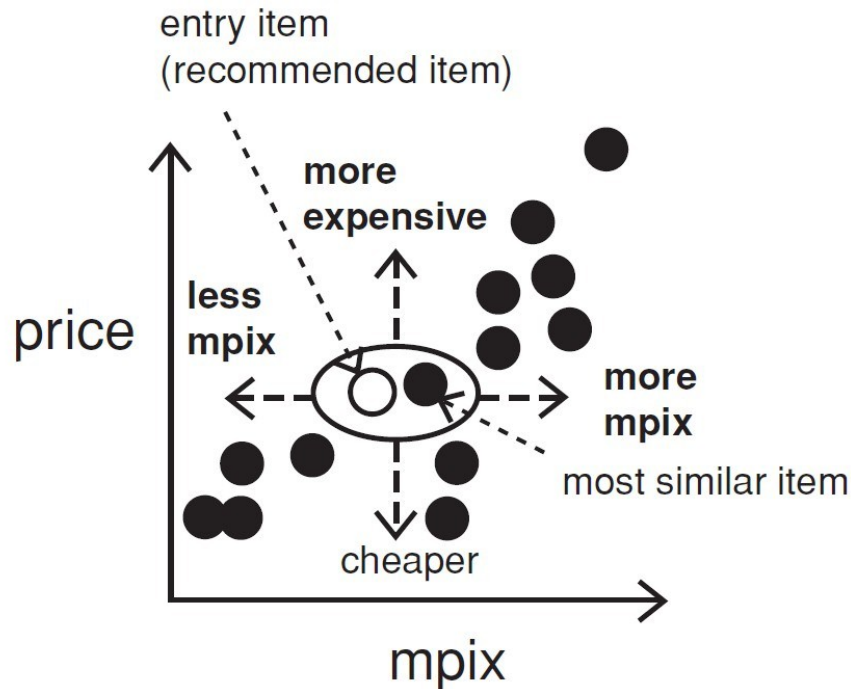
- ❑  $w_r$  es el peso otorgado al requisito  $r$

- ❑ En un caso real, a los usuarios les gustará

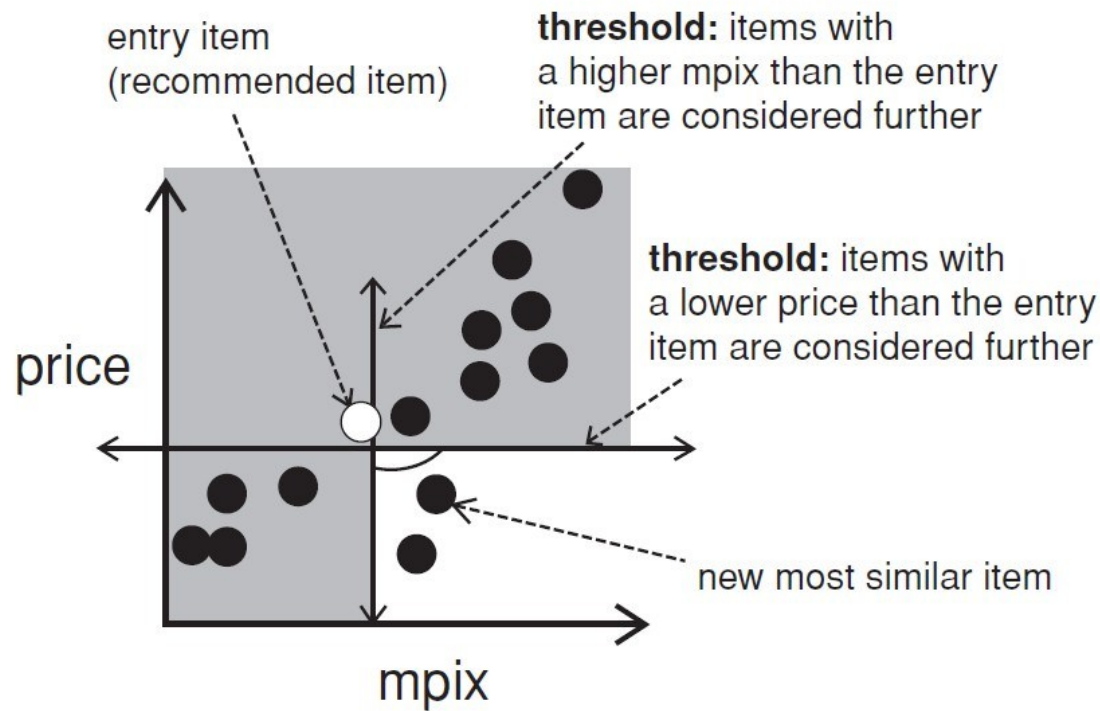
- ❑ Maximizar algunas propiedades, p.e. resolución de la cámara, "more is better"(MIB)

- ❑ Minimizar algunas propiedades, p.e. el precio de la cámara, "less is better"(LIB)

- ❑ Los usuarios pueden no saber qué es lo que realmente quieren
- ❑ Efectuar críticas sobre productos como soporte a la navegación es un método efectivo de obtener productos



- ❑ Criticar sobre múltiples atributos (críticas compuestas) puede mejorar la calidad de la navegación y reducir el tiempo



- ☐ Recomendador de restaurantes en chicago
- ☐ Case-based con crítica unitaria
- ☐ Similitud

*Find your  
Favourite restaurant*



In Vienna you chose:

+43 1 123 123 123

Mariahilferstrasse 123,  
1010 Wien

**Biergasthof**

30€-50€

Local cuisine

local food, central in the city, weekend brunch, room with a view,  
famous for beer, seasonal dishes, group bookings, open all day

For Graz we recommend:

+43 316 45 45 45

Brauhofstrasse 45,  
8023 Graz

**Brauhof**

30€-50€

Local cuisine

local food, own beer, weekend lunch, open all day, private function room,  
famous for beer, seasonal dishes, group bookings, good transport connection

Less \$\$

Nicer

Cuisine

More Quiet

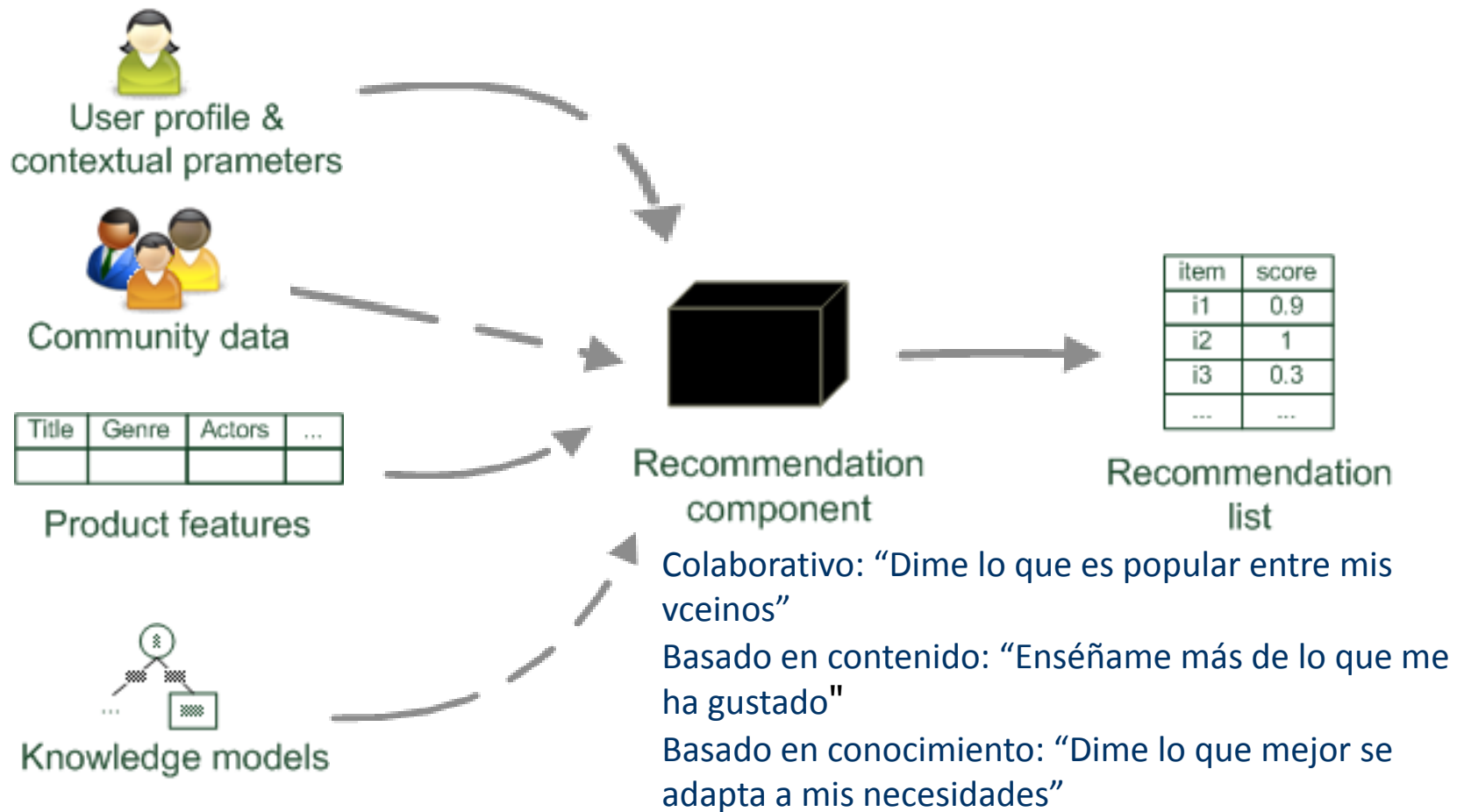
Traditional

Creative

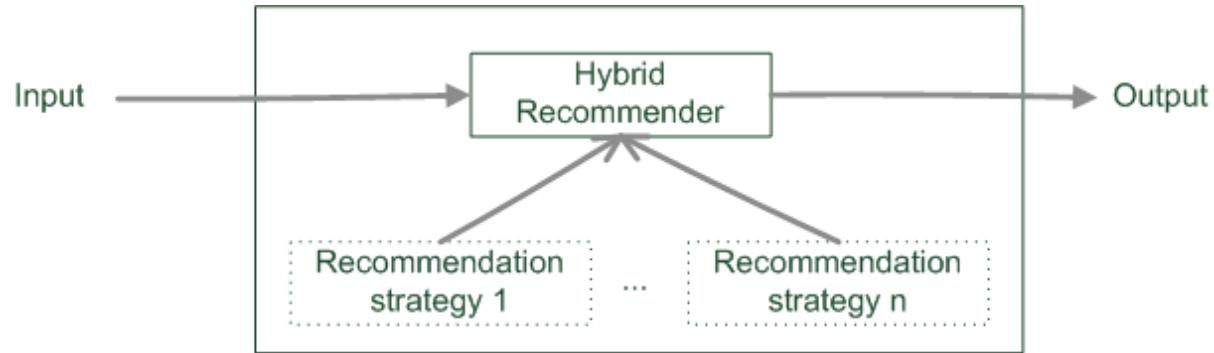
Livelier



- ❑ Qué es un sistema de recomendación
- ❑ Campo de investigación
- ❑ Técnicas estándar de recomendación
  - ❑ Recomendación colaborativa
  - ❑ Recomendación basada en contenido
  - ❑ Recomendación basada en conocimiento
  - ❑ Recomendación híbrida



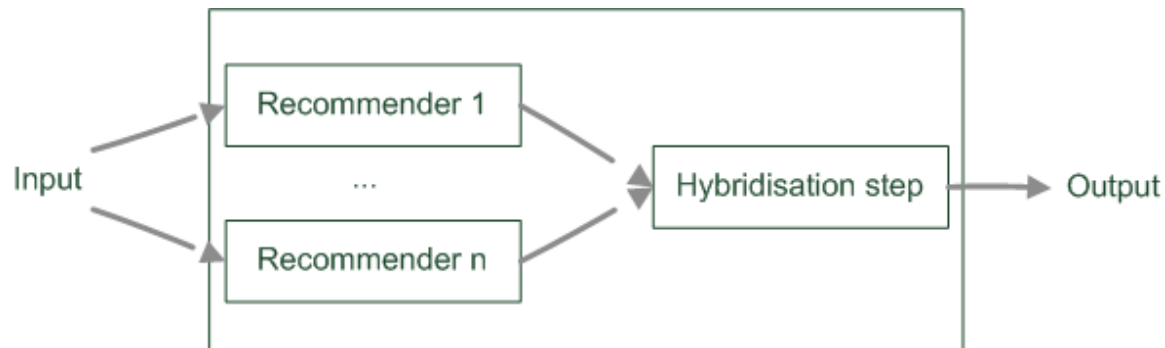
- ❑ Las tres técnicas explicadas anteriormente funcionan todas de una manera correcta, pero ya hemos visto que tienen algunas carencias
  - ❑ Problema del arranque en frío, necesidad de conocimiento, usuarios...
- ❑ Ideal: combinar dos (o más) tipos
  - ❑ hybrida [lat.]: Se dice de todo lo que es producto de elementos de distinta naturaleza
  - ❑ Evitar algunas de las carencias
- ❑ Diferentes estilos de hibridación
  - ❑ Monolítico
  - ❑ Paralelo
  - ❑ Distribuido (Pipelined)



- ❑ Es un híbrido "virtual"
- ❑ Características/Concimiento de distintas fuentes son combinadas

- ❑ Combinación de distintas fuentes de conocimiento
  - ❑ Por ejemplo: Valoraciones y datos de usuario demográficos o requisitos de usuario utilizados para medir la similitud
- ❑ Características híbridas
  - ❑ Características sociales: Películas que le han gustado a un usuario
  - ❑ Características del contenido: Tipo de película (género)
  - ❑ Características híbridas: Al usuario le han gustado muchas películas que son comedia...

- ❑ La salida se obtiene a partir de la combinación de distintas implementaciones
- ❑ Cómo combinar
  - ❑ Pesos
  - ❑ Votaciones
  - ❑ Caso extremo: elección

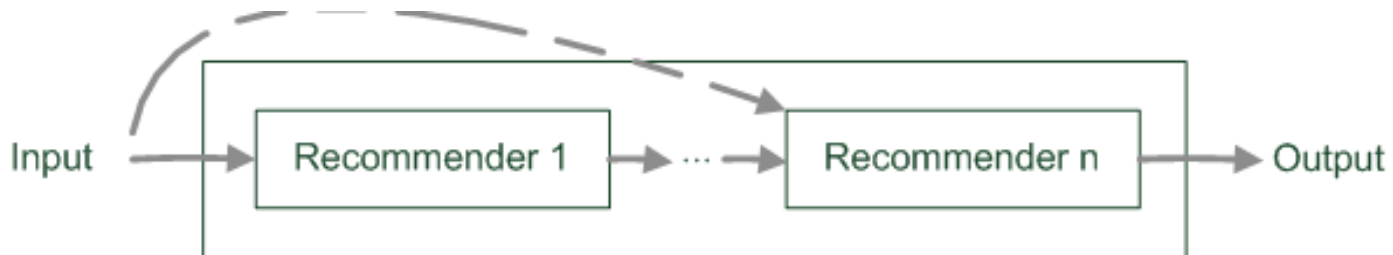


Recommender 1		
Item1	0.5	1
Item2	0	
Item3	0.3	2
Item4	0.1	3
Item5	0	

Recommender 2		
Item1	0.8	2
Item2	0.9	1
Item3	0.4	3
Item4	0	
Item5	0	

Recommender weighted(0.5:0.5)		
Item1	0.65	1
Item2	0.45	2
Item3	0.35	3
Item4	0.05	4
Item5	0.00	

- ❑ Un recomendador pre-procesa alguna entrada para la siguiente etapa
- ❑ Las listas de recomendación se van refinando
  - ❑ El primer recomendador excluye elementos
  - ❑ El segundo recomendador asigna valoraciones







¿Qué recomendadores conozco?

¿Qué páginas que visito frecuentemente tiene recomendadores?

