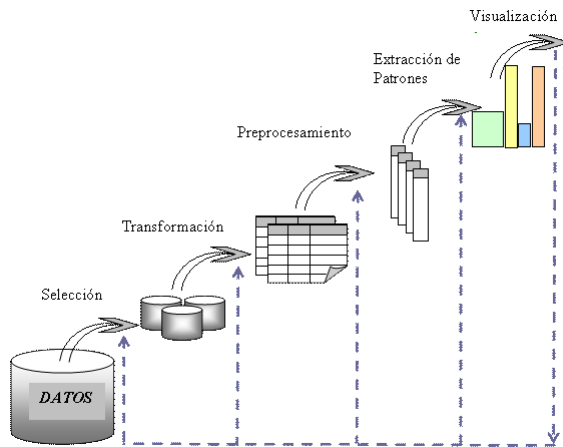


## PREDICA: Sistema de Recomendación



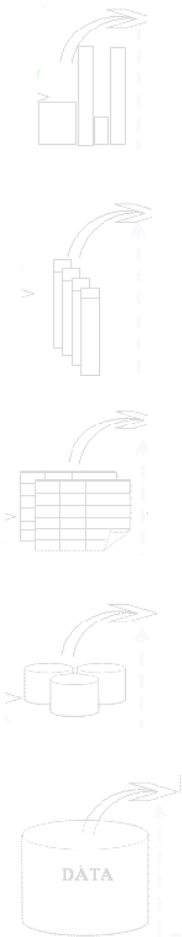
Marta Millán  
Patricia Trujillo  
Oscar Bedoya  
Edward Gallego  
Edward Ortiz

# Sistemas de Recomendación

---

## □ Formulación Común

- Estimación ratings items no vistos por usuario
- Estimación basada en ratings dados por usuario sobre otros items e información adicional
- Recomendación de items con más alto rating



# Formulación Problema




$C = \{c_1, c_2, \dots, c_n\}$  usuarios

$I = \{i_1, i_2, \dots, i_k\}$  Items



$u$ : función de utilidad del item  $i$  para usuario  $c$






$$u : C \times I \rightarrow R$$



donde  $R$  es un conjunto totalmente ordenado  
(enteros no negativos o reales en un rango)


$$\forall c \in C \quad i'_c = \arg \max_{i \in I} u(c, i)$$

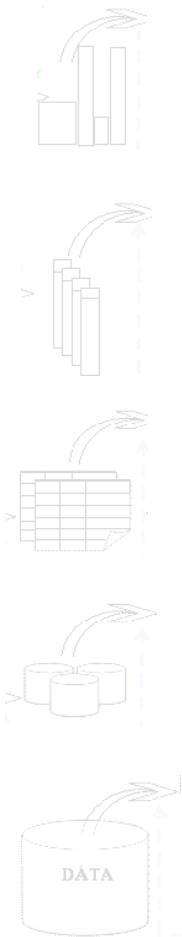

# Formulación Problema

- 
- ❑  $c$  puede definir un perfil
  - ❑ Cada item tiene asociadas características
  - ❑ Problema de recomendación
    - $U$  no definida para  $C \times I$  completo
    - Necesidad de extrapolar  $u$
    - Construcción matrices ratings (usuario x item)
    - Estimación ratings items no evaluados
    - Recomendación rating estimados Top N
    - Intrusive
- 
- 
- 
- 

# Sistemas de Recomendación

---

- ❑ Software predicción items (**movies, music, books, news, web pages**) de interés usuario
- ❑ información perfil del usuario
- ❑ Técnicas más usadas (cómo se hace la recomendación)
  - Content-based recommendations
  - Collaborative recommendations
  - Hybrid approaches



# Recomendación: Clasificación

## ❑ Basados en Contenido

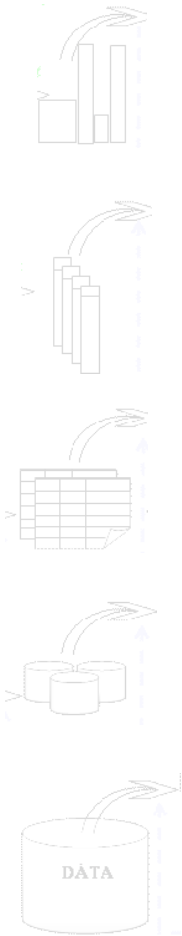
- ❑ Definición de items a partir de características
- ❑ Aprende perfil usuario a partir características objetos accedidos
- ❑ Item-item
- ❑ Recomiendan items similares

## ❑ Filtrado Colaborativo

- ❑ Comunidades usuarios (social)
- ❑ Pueden sorprender

## ❑ Híbridos

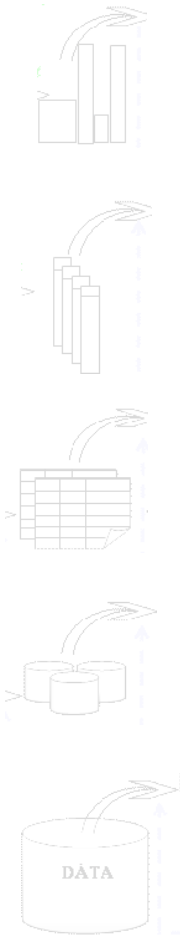
Combinación modelos anteriores



# Enfoque basado en contenido

---

- Information retrieval
- Items representados características
- Utilidad item  $i$  para usuario  $c$  basado en utilidades asignadas por usuario  $c$  a items similares a  $i$

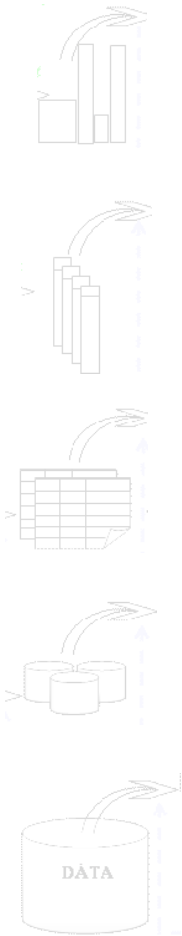


# Enfoque basado en contenido

---

- Limitaciones

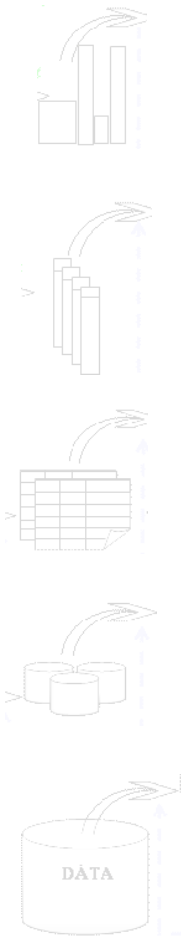
- Consumen tiempo
- No adecuados entornos dinámicos
- Entornos muy grandes con items ingresando con mucha frecuencia
- Problema nuevo usuario





# Filtrado Colaborativo

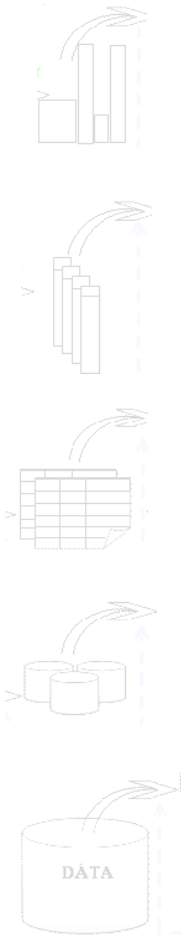
- Basado en opiniones usuarios
- Recomendaciones : opiniones de usuarios similares a  $u_a$
- Items no representados en términos de características
- Utilidad  $u(c,i)$  basada utilidades  $u(c_j,i)$   $c_j \in C$  similares a  $c$



# Filtrado Colaborativo

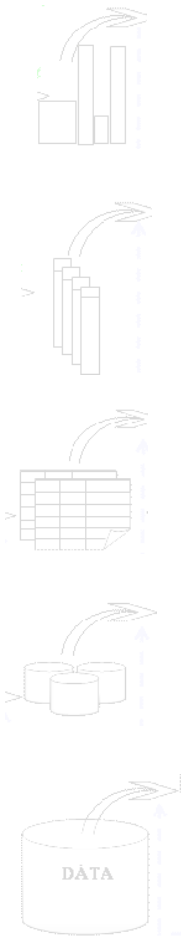
---

- ❑ Basados en Memoria o en Modelo
- ❑ Aplicables multiples dominios
- ❑ Independientes items a recomendar
- ❑ Centradas en usuarios
- ❑ Usan diferentes tipos de datos: preferencias, demográficas, ratings



# Filtrado Colaborativo

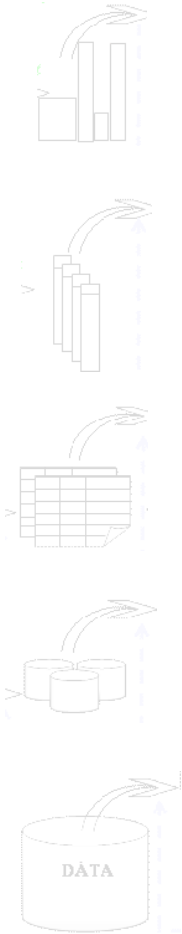
- Utilidad de item  $i$  para usuario  $c$  estimado a partir de utilidades asignadas al item  $i$  por usuarios similares a  $i \in I$
- Basados en juicios de comunidad o vecindario de  $u_a$ 
  - Comunidad: concepto importante
  - Usuarios contribuyen desempeño sistema



# Filtrado Colaborativo

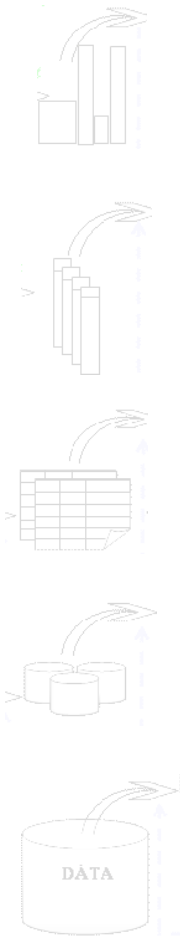
## □ Basados en Memoria

- Heurísticas predicción ratings
- Basados en colección completa de items previamente evaluados por los usuarios
- Valor desconocido rating  $r_{c,i}$  agregado de ratings de otros usuarios (usualmente los N más similares) para el mismo item



# Filtrado Colaborativo

- ❑ Basados en Memoria
  - Varios enfoques similitud usuarios
  - Similitud entre usuarios
    - ratings items evaluados por ambos
    - basado en correlación
    - basado en coseno

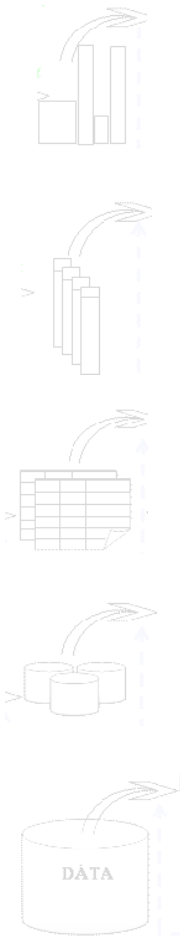


# Filtrado Colaborativo

---

## ❑ Basados en Modelo

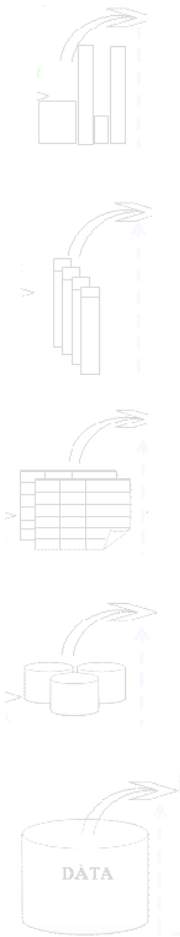
- Usan conjunto de ratings para aprender un modelo
- Modelo usado para predecir ratings
  - Enfoques probabilísticos para estimar rating basados en modelos:
    - » Cluster models
    - » Bayesian Networks



# Filtrado Colaborativo

---

- ❑ Basados en Modelo
  - Limitación: usuario por cluster
  - Nuevo usuario
  - Item nuevo
  - Dispersión



# Filtrado Colaborativo

## Recomendaciones para usuario C

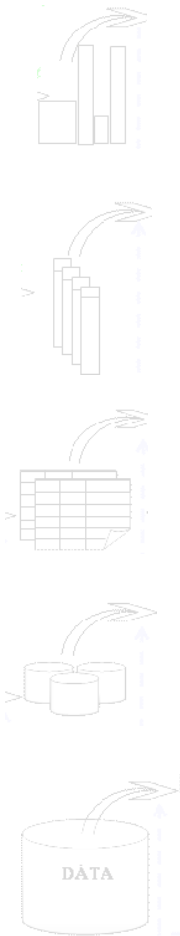
	Book 1	Book 2	Book 3	Book 4	Book 5	Book 6
Customer A	X			X		
Customer B		X	X		X	
Customer C		X	X			
Customer D		X				X
Customer E	X				X	

- B es muy similar a C (ha comprado los libros que C ha comprado). Book 5 altamente recomendado
- D se parece algo. Book 6 se recomienda
- A y E no se parecen a C

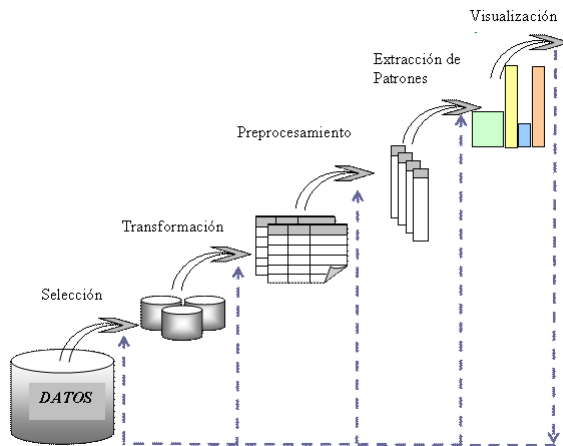


# Hibridos

- ❑ Combinación contenido y colaborativo
  - Implementar separados y combinar predicciones
  - Incorporar características de contenido en enfoque colaborativo
  - incorporar características colaborativas en contenido
  - construir modelo unificado



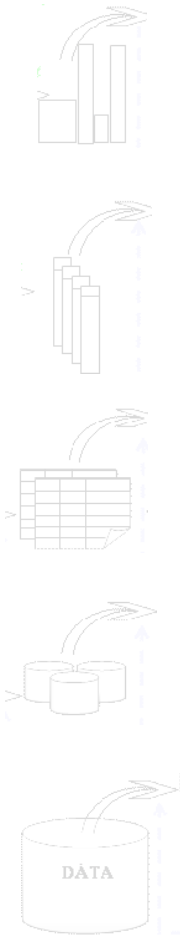
## Sistema Recomendación PREDICA



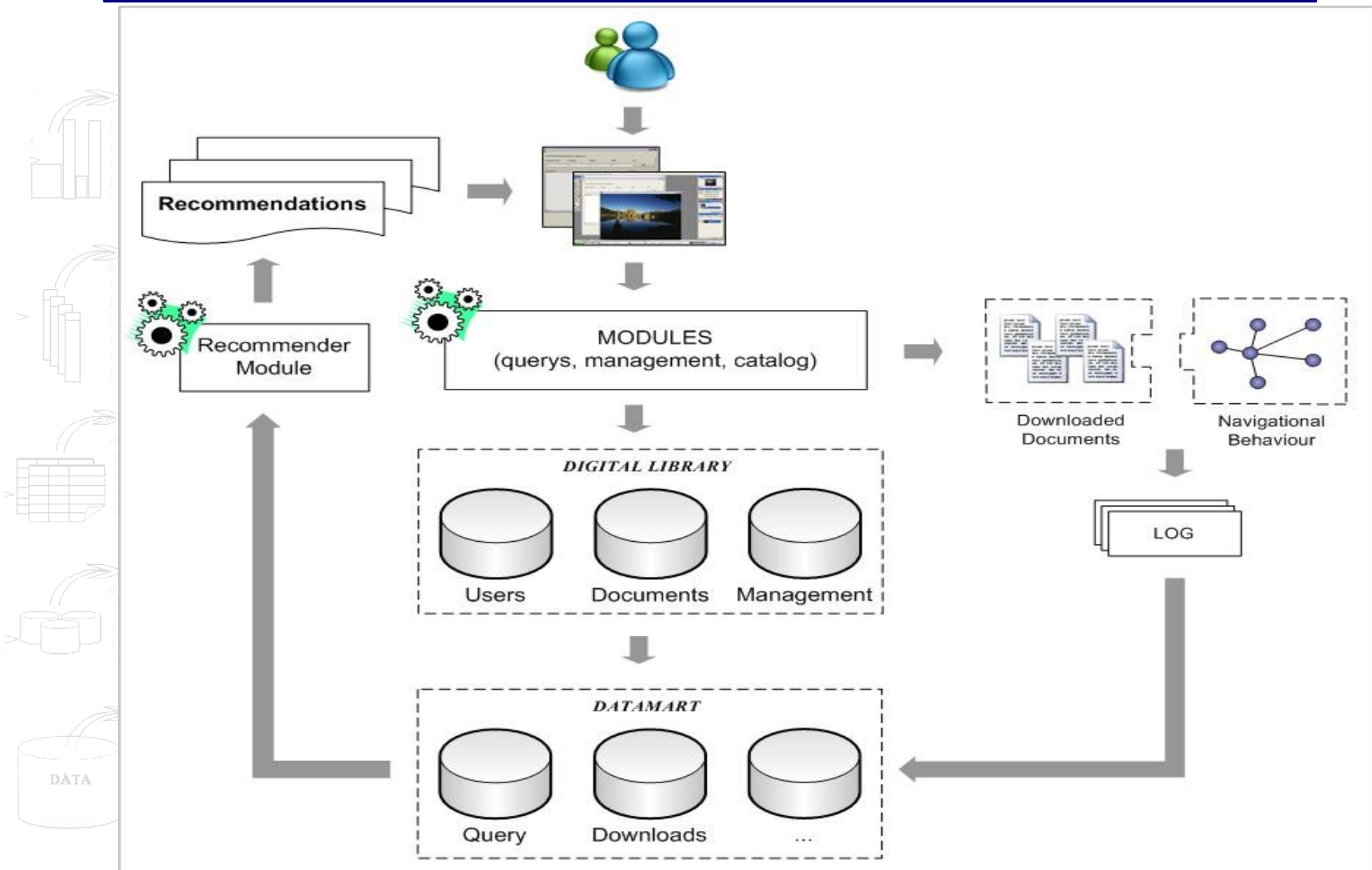
# Modelo Recomendación PREDICA

---

- ❑ Antecedentes
  - No intrusivo
  - Híbrido
  - Dominio de Aplicación
  - Ratings implícitos
  - Medidas de similitud
  - Énfasis en conducta usuario
  - Limitaciones usuario nuevo
  - Dispersión

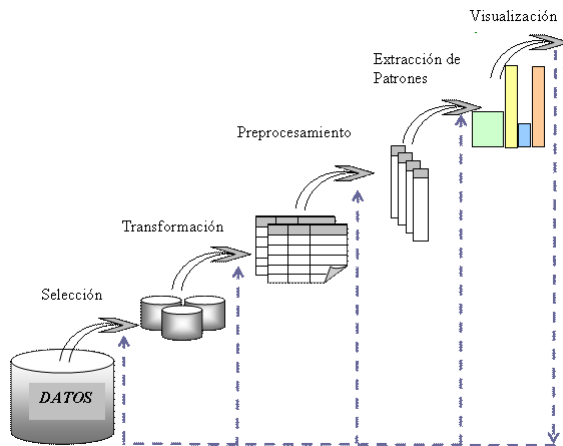


# Arquitectura SR PREDICA



# KDD

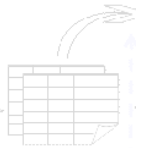
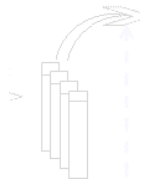
## Enfoques



# Predica

---

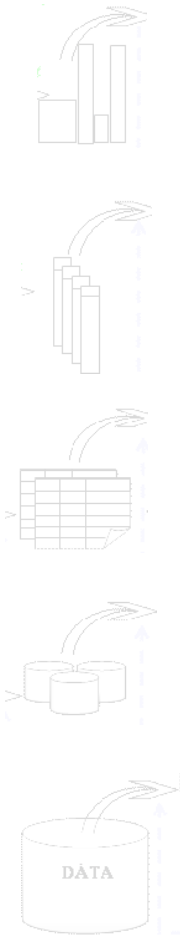
- ❑ Basado en Filtrado colaborativo
- ❑ Contrucción Clusters de usuarios
- ❑ Similitud usuarios: datos demográficos, psicográficos y de navegación
- ❑ Problema de usuario nuevo
- ❑ Posibilidad de ponderación factores



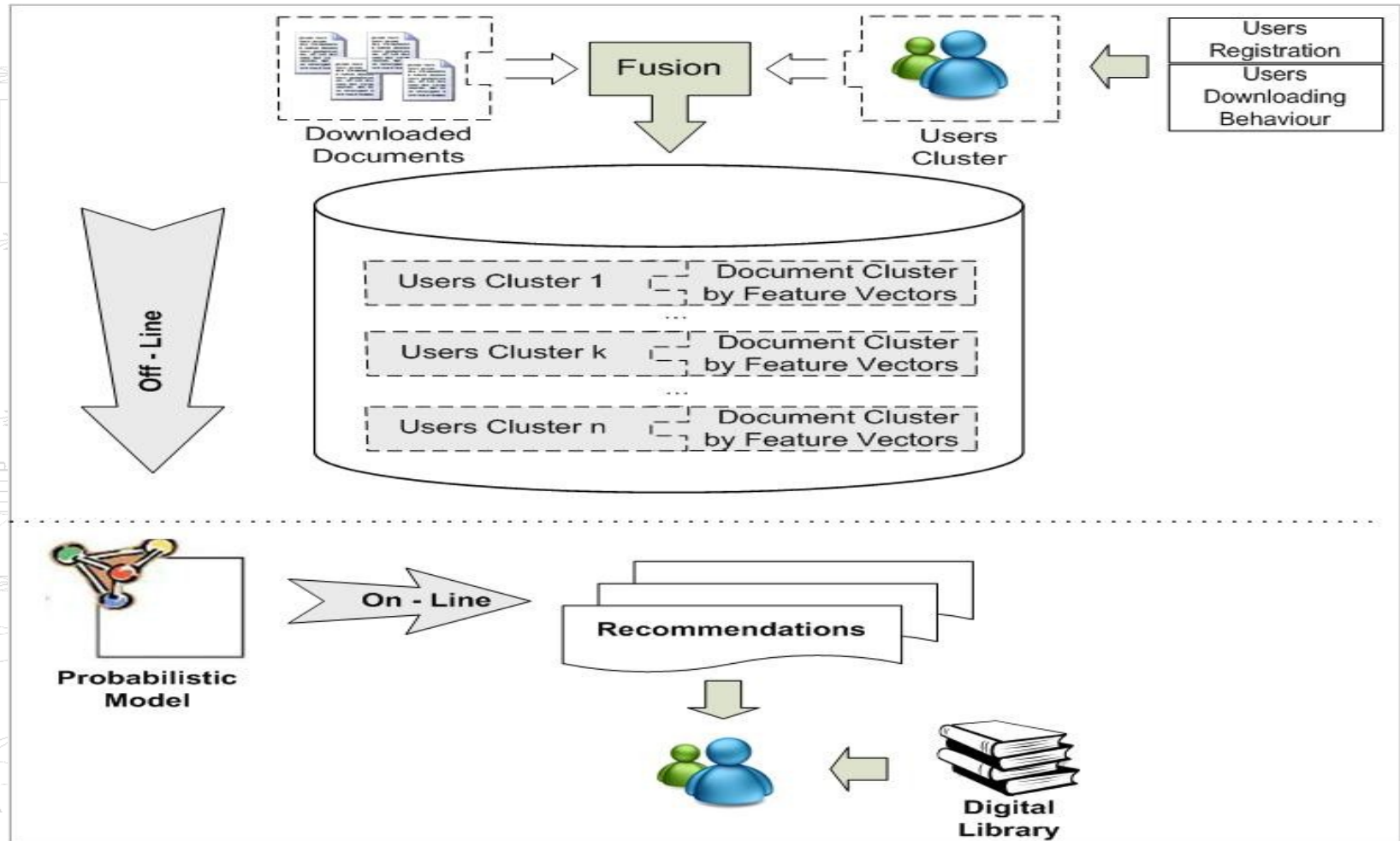
# Predica

---

- ❑ Meta-clusters usuarios-documentos
- ❑ Metadatos de documentos
- ❑ Modelo probabilidad para recomendar
- ❑ Recalculo de clustering: administrador



# Basado en Modelo-Memoria (I)



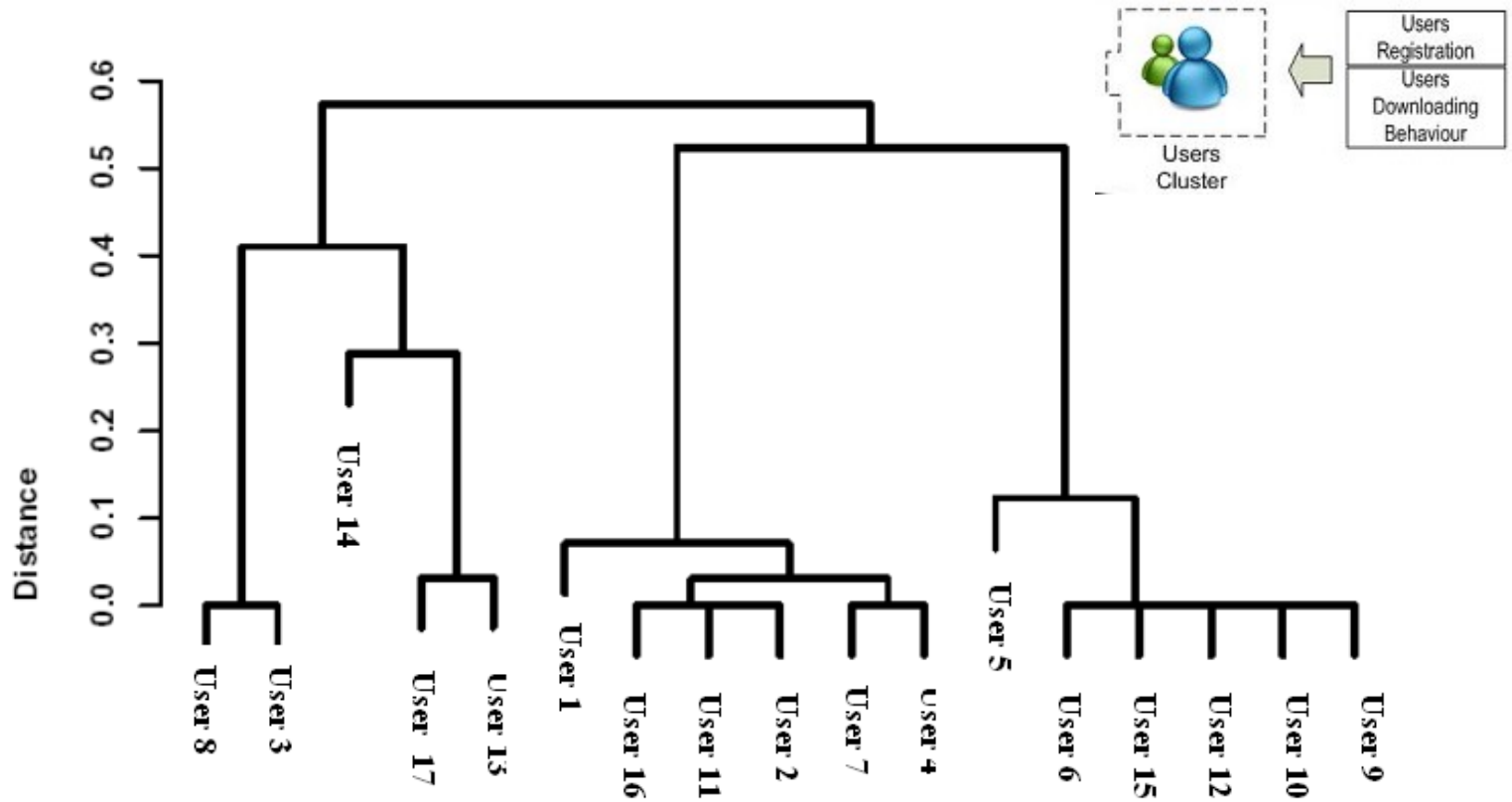


# Características

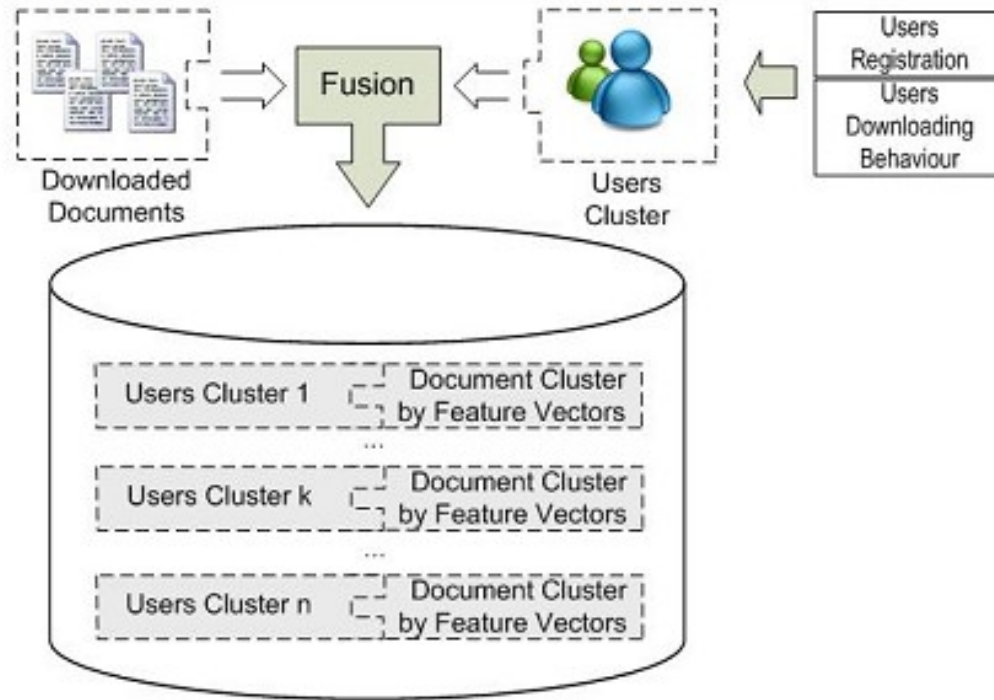
Static			Dynamic																																
Demographic Characteristics			Psychographic Characteristics																																
Age Range	University Relation	Academic Degree	Interest Areas			Downloaded Documents																													
			Area 1	...	Area m	Document 1	...	Document n																											
<div><div>User A</div><table><tr><td>User B</td><td>1</td><td>0</td></tr><tr><td>1</td><td>a</td><td>b</td></tr><tr><td>0</td><td>c</td><td>d</td></tr></table><div><math>\lambda_1</math></div></div>			User B	1	0	1	a	b	0	c	d	<div><div>User A</div><table><tr><td>User B</td><td>1</td><td>0</td></tr><tr><td>1</td><td>a</td><td>b</td></tr><tr><td>0</td><td>c</td><td>d</td></tr></table><div><math>\lambda_2</math></div></div>			User B	1	0	1	a	b	0	c	d	<div><div>User A</div><table><tr><td>User B</td><td>1</td><td>0</td></tr><tr><td>1</td><td>a</td><td>b</td></tr><tr><td>0</td><td>c</td><td>d</td></tr></table><div><math>\lambda_3</math></div></div>			User B	1	0	1	a	b	0	c	d
User B	1	0																																	
1	a	b																																	
0	c	d																																	
User B	1	0																																	
1	a	b																																	
0	c	d																																	
User B	1	0																																	
1	a	b																																	
0	c	d																																	

$$S(A, B) = \sum_{i=1}^3 \lambda_i S_i(A, B),$$

# Clusters de usuarios

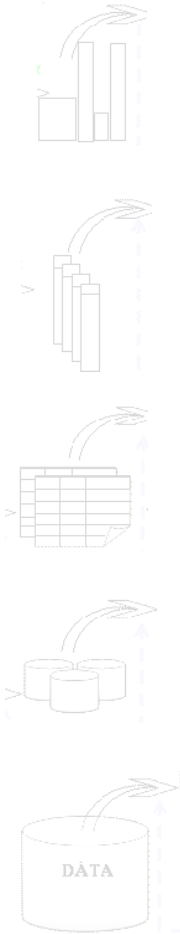


# Meta-Clusters



# Función de Predicción

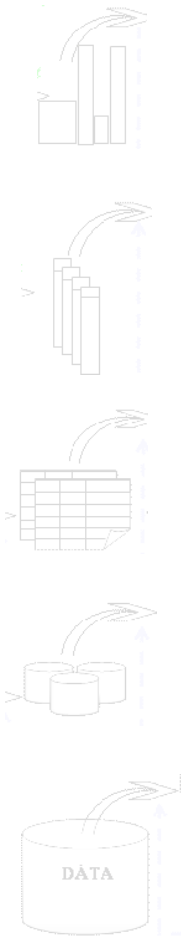
$$P_s(D_j|U_a) \approx P(SD(f_j)|U_a, C_k) + P(D_j|SD(f_j), C_k)$$



# Predica: Recomendación (II)

---

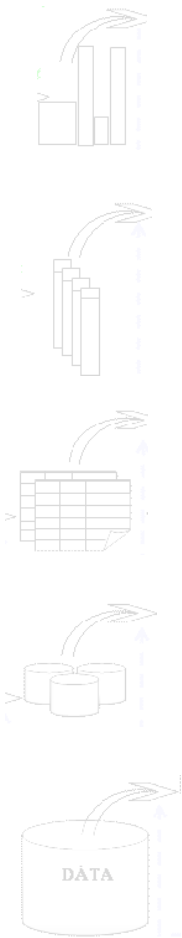
- ❑ Matrices de co-ocurrencia de descargas
- ❑ Basada en Reglas Hebb
  - Activación neuronas relación entre conceptos mas fuerte
  - Falta de activación pierde vigencia
  - Co-activacion conceptos\_ relación
  - fortaleza de la asociación entre dos conceptos  $i$  y  $j$  es proporcional al grado de activación y tiempo



# Predica: Recomendación (II)

---

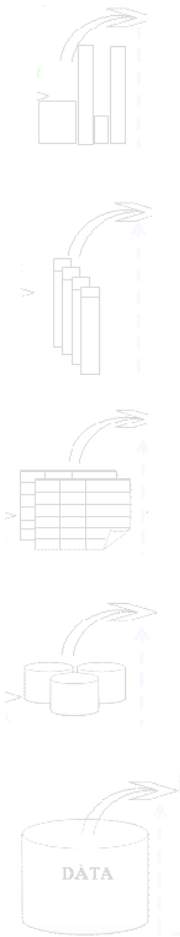
- ❑ Activaciones en Predica - acciones consultar, descargar un documento o leer el resumen.
- ❑ Acción asociada con importancia representada en el grado de activación  $A(i)$ .
- ❑  $A(i)$  es un valor numérico entre 0.0 y 1.0



# Cálculo activaciones

---

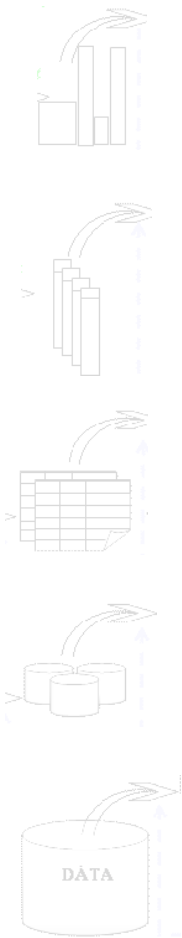
- grado de activación documento basado en dos estrategias
  - explícita usuario califica documento
  - implícita determinados por el experto (importancia de acciones en documentos)



# Requerimientos

---

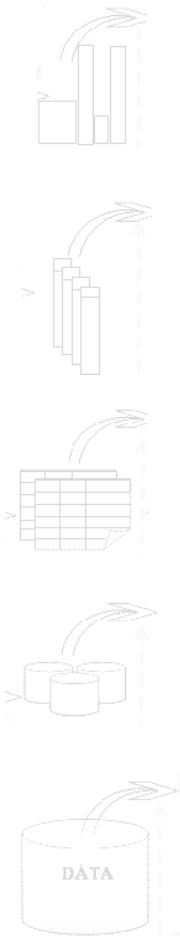
- ❑ mantener valores activaciones en un medio persistente
- ❑ cargar al sistema la información de las activaciones realizadas durante el día
- ❑ Posibilidad de fortalecer una nueva relación entre documentos que genere nueva recomendación.





# Uso de activaciones

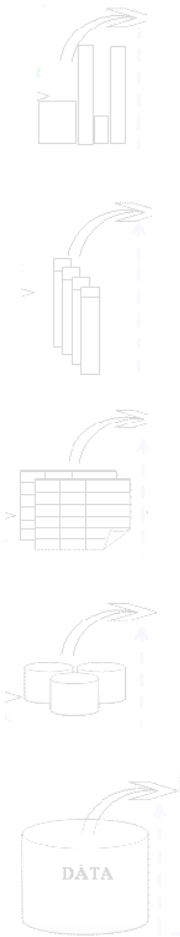
- ❑ recomendaciones hechas cuando el usuario activa documento
- ❑ Si no existe ningún documento no se puede recomendar. De lo contrario, se listan los documentos que tienen los cinco mayores grados de activación.



# Modelo III Colaborativo

---

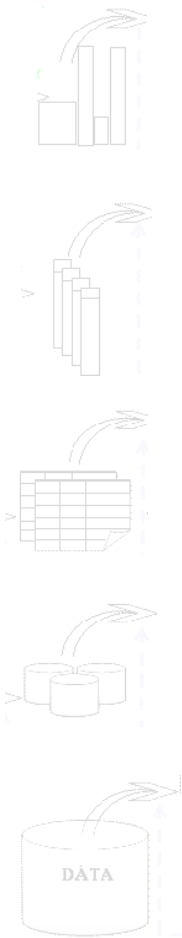
- ❑ No simetría similitud usuarios
- ❑ Importancia diferenciar historias de conducta (tiempo)
- ❑ Propuesta medida asimétrica simple
- ❑ Evaluación de resultados: satisfactoria
- ❑ Mejoras en precisión de recomendaciones



# Enfoque basado en Memoria

---

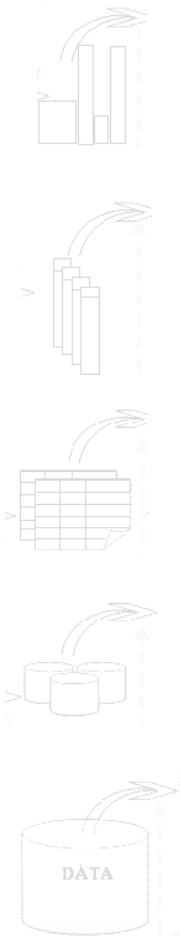
- ❑ Colaborativo
- ❑ Medida de similitud solo navegacional
- ❑ similitud usuarios dinámica
- ❑ cálculo en línea
- ❑ Precálculo de matriz de similitud
- ❑ Resultados experimentales promisorios



# Enfoque basado en Memoria

---

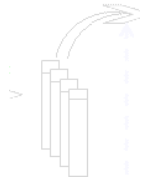
- ❑ Uso función similitud asimétrica
- ❑ Cálculo de vecindarios para  $u_a$
- ❑ Recomendaciones basadas en items vecindario mas frecuentes



# Temas abiertos en SR

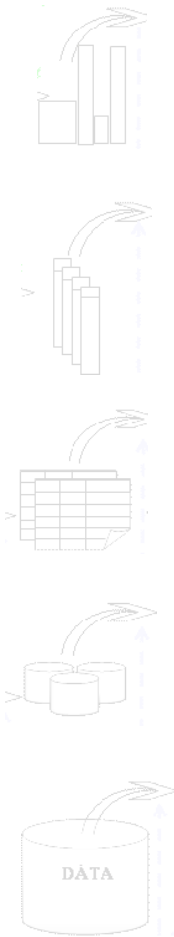
---

- ❑ Métodos de recomendación más efectivos
- ❑ Métodos rango amplio de aplicaciones reales
- ❑ Mejores metodos para representar conducta del usuario
- ❑ Incorporación de información contextual



# Temas abiertos en SR

---

- 
- ❑ Utilización de rating multicriterio
  - ❑ Desarrollo de métodos menos “intrusive” y más flexibles
  - ❑ Medidas más efectivas para evaluar desempeño