

Information Retrieval – IR

Franco Giustozzi
Bases de Datos Avanzadas

REPASO



- Hacia una definición de Recuperación de Información.
- Sistemas de Recuperación de Información.
- Modelos de Recuperación de Información.
- Evaluación de la Recuperación de Información.
- Lenguajes de Consulta.
- Estrategias de Búsqueda.
- Indexado y Búsqueda.

AGENDA



- Vamos a ver algunas aplicaciones de la IR.
 - Recuperación de Información en la WEB.
 - Búsquedas Inteligentes. Sistemas Recomendadores.
 - Un Sistema Recomendador para la Recuperación Personalizada de objetos de Aprendizaje.



RECUPERACIÓN DE INFORMACIÓN EN LA WEB

RECUPERACIÓN DE INFORMACIÓN EN LA WEB



- Hoy en día, existen muchos sitios WEB que manejan un cuadal de información considerable.
- Estos sitios suelen tener un buscador propio para permitirle al usuario localizar y acceder a la información que necesita en una forma más directa y más rápida, evitando tener que navegar todo el sitio.
- Al igual que en la búsqueda en la WEB, buscar información dentro de un sitio determinado, muchas veces no se obtienen los resultados esperados.
- Suceden problemas de sinonimia y polisemia de términos.

RECUPERACIÓN DE INFORMACIÓN EN LA WEB



Un **buscador** tiene tres componentes principales:

- Un **CRAWLER**: recorre la colección y crea un repositorio local con el contenido de las paginas visitadas.
- Un **INDEXADOR** procesa dicho repositorio y genera un índice invertido del mismo.
- Las consultas ingresadas son enviadas al **MOTOR DE BÚSQUEDA**, el cual consulta al índice y retorna los resultados de las busquedas al usuario.

RECUPERACIÓN DE INFORMACIÓN EN LA WEB



Arquitectura de un buscador.

RECUPERACIÓN DE INFORMACIÓN EN LA WEB



- Un Motor de Búsqueda es Nutch (lucene.apache.org/nutch/).
- Aquí research.google.com/pubs/InformationRetrievalandtheWeb.html se pueden encontrar varios artículos relacionados a la IR y la Web.



BÚSQUEDAS INTELIGENTES

SISTEMAS RECOMENDADORES

BÚSQUEDAS INTELIGENTES

SISTEMAS RECOMENDADORES



Cuando se busca información, en particular cuando se busca información en la WEB, se advierte:

- Una sobrecarga de información, lo que obliga a los usuarios a explorar espacios demasiado densos.
- Por otro lado, muchas veces un material dado no es el adecuado para un usuario, ya que cada uno posee preferencias personales.
 - Sería bueno que sean consideradas en el momento de la búsqueda.

BÚSQUEDAS INTELIGENTES

SISTEMAS RECOMENDADORES

- Según [Resnick et al., 1997], [Terveen et al., 2001] un **SISTEMA RECOMENDADOR** ayuda a resolver este tipo de problemas.
- Pues son capaces de seleccionar, de forma automática y personalizada, el material que mejor se adapte a las preferencias o necesidades de un usuario.
- Estos sistemas utilizan distintas técnicas para razonar sobre las preferencias de los usuarios y sobre las descripciones semánticas del material disponible.
 - ¿Cómo?, modeladas a través de perfiles personales.

BÚSQUEDAS INTELIGENTES

SISTEMAS RECOMENDADORES



- Este tipo de sistemas generalmente se modelan como **sistemas multiagentes**.
- Se puede definir un **agente** como una aplicación informática con capacidad para decidir cómo actuar para alcanzar sus objetivos.

BÚSQUEDAS INTELIGENTES

SISTEMAS RECOMENDADORES



Algunas ventajas de estos **agentes**:

- Pueden funcionar fiablemente en un entorno rápidamente cambiante e impredecible (como es la WEB).
- Pueden configurarse con diferentes perfiles para tomar decisiones de acuerdo a las necesidades del usuario y hacer tareas más específicas y personalizadas.

BÚSQUEDAS INTELIGENTES

SISTEMAS RECOMENDADORES

¿Cómo logramos esa personalización?

- La personalización de los resultados de una búsqueda se sustenta en los **metadatos**, tanto del usuario como de los documentos, dado que permiten evaluar la cercanía de un documento al perfil del usuario.
- Los **metadatos** descriptivos de un documento, en general, siguen un estándar de metadatos.

¿Cómo se representan los **metadatos**?

- Se pueden representar de varias maneras.
 - Una puede ser la forma de vectores donde cada elemento del vector está compuesto por tres partes:
 - nombre del metadato.
 - valor que tiene el metadato.
 - un peso que indique la importancia que posee esa característica para que el recurso sea aprovechado por el usuario.

BÚSQUEDAS INTELIGENTES

SISTEMAS RECOMENDADORES



- La ventaja de un **SISTEMA RECOMENDADOR** es que permite mejorar los resultados de una búsqueda porque presenta los resultados de distinta forma según el usuario que haya realizado la consulta.
- Sin embargo, el mayor problema para el uso de **SISTEMAS RECOMENDADORES** es la falta de metadatos en los documentos.



SISTEMA RECOMENDADOR PARA LA RECUPERACIÓN DE OBJETOS DE APRENDIZAJE

BÚSQUEDAS INTELIGENTES

SISTEMAS RECOMENDADORES

SISTEMA RECOMENDADOR PARA LA RECUPERACIÓN DE OBJETOS DE APRENDIZAJE

- En el dominio de la educación existe gran cantidad y diversidad de material que puede contribuir al proceso enseñanza-aprendizaje.
- Utilizando buscadores se puede obtener todo tipo de material desde la Web.
 - Problemas:
 - falta de estructura en la información.
 - No siempre el resultado es el esperado por el usuario, porque no considera otros factores relevantes (preferencias personales).
- Hay formas de acceder a información más estructurada, utilizando repositorios.
- Tener en cuenta las preferencias de los usuarios en el momento de la búsqueda.

BÚSQUEDAS INTELIGENTES

SISTEMAS RECOMENDADORES

SISTEMA RECOMENDADOR PARA LA RECUPERACIÓN DE OBJETOS DE APRENDIZAJE



Entonces, el objetivo es:

Desarrollar un sistema recomendador que ayude a los usuarios a encontrar recursos educativos de acuerdo a sus características y preferencias, y brinde los resultados ordenados de acuerdo a su perfil.

BÚSQUEDAS INTELIGENTES

SISTEMAS RECOMENDADORES

SISTEMA RECOMENDADOR PARA LA RECUPERACIÓN DE OBJETOS DE APRENDIZAJE

- Según [Wiley], un **Objeto de Aprendizaje (OA)** es *“cualquier recurso digital que puede ser utilizado repetidamente para facilitar el aprendizaje”*.
 - Pueden adquirir diferentes formatos y soportes.
 - Pueden ser reutilizados, actualizados, combinados, separados, referenciados y sistematizados.
- Se necesitan **potentes repositorios**.

BÚSQUEDAS INTELIGENTES

SISTEMAS RECOMENDADORES

SISTEMA RECOMENDADOR PARA LA RECUPERACIÓN DE OBJETOS DE APRENDIZAJE

- Un **Repositorio de Objetos de Aprendizaje** se entiende como una gran colección de OAs:
 - Estructurado como una base de datos.
 - Con **metadatos** asociados.
 - Generalmente se pueden acceder via Web.
- Algunos Repositorios son:
 - ARIADNE (European Association open to the World, for Knowledge Sharing and Reuse)
<http://www.ariadne-eu.org/>
 - OER Commons (Open Educational Resources)
www.oercommons.org
 - FLOR (Federación Latinoamericana de Repositorios)
<http://ariadne.cti.espol.edu.ec/FederatedClient>
 - MERLOT (Multimedia Educational Resource for Learning and Online Teaching)
www.merlot.org

BÚSQUEDAS INTELIGENTES

SISTEMAS RECOMENDADORES

SISTEMA RECOMENDADOR PARA LA RECUPERACIÓN DE OBJETOS DE APRENDIZAJE



- Los **metadatos** son un conjunto de atributos necesarios para describir las principales características de un recurso.
- **LOM (Learning Object Metadata)** es el estándar de metadatos de la **IEEE** para los OAs:
 - Especifica la sintaxis y la semántica de un conjunto mínimo de metadatos necesario para identificar, administrar, localizar y evaluar un OA.

BÚSQUEDAS INTELIGENTES

SISTEMAS RECOMENDADORES

SISTEMA RECOMENDADOR PARA LA RECUPERACIÓN DE OBJETOS DE APRENDIZAJE



- Los **sistemas recomendadores** ayudan a las personas a encontrar lo que necesitan especialmente en la Web.
- Se ha incrementado el diseño de sistemas multiagentes para abordar el desarrollo de estos sistemas.
- Los **sistemas recomendadores** tienen distintas aplicaciones:
 - El dominio de la educación es un buen candidato, pues:
 - Hay mucha diversidad y cantidad de material.
 - Accesible vía repositorios o vía la Web.
 - Puede ser utilizado por un docente y/o alumno.

BÚSQUEDAS INTELIGENTES

SISTEMAS RECOMENDADORES

SISTEMA RECOMENDADOR PARA LA RECUPERACIÓN DE OBJETOS DE APRENDIZAJE



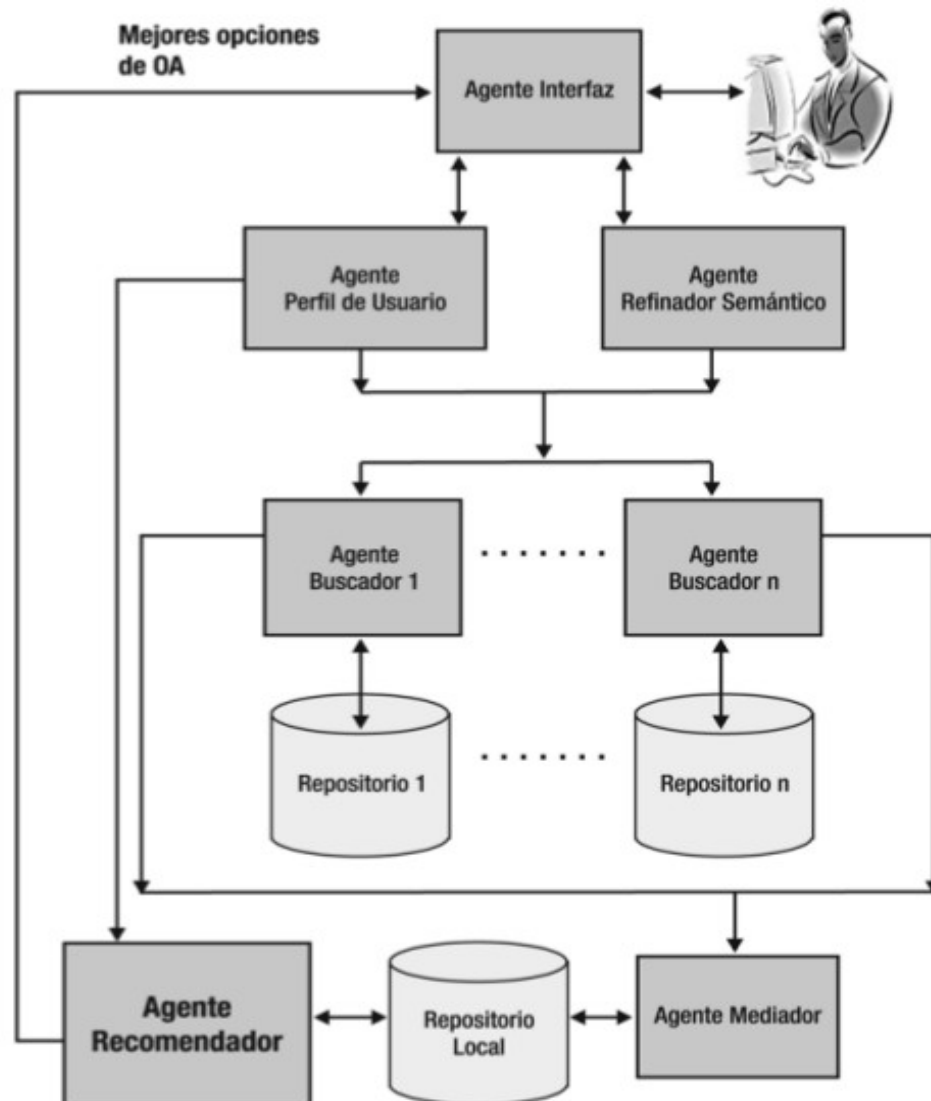
Arquitectura Propuesta del Sistema Recomendador.

- Consta de los siguientes agentes:
 - El Agente Interfaz (Agente-I).
 - El Agente Refinador Semántico (Agente-RS).
 - El Agente Perfil de Usuario (Agente-PU).
 - Los Agentes Buscadores (Agente-Bi).
 - El Agente Mediador (Agente-M).
 - El Agente Recomendador (Agente-R).

BÚSQUEDAS INTELIGENTES

SISTEMAS RECOMENDADORES

SISTEMA RECOMENDADOR PARA LA RECUPERACIÓN DE OBJETOS DE APRENDIZAJE



Agente-I: se encarga de capturar los datos ingresados por el usuario.

Agente-RS: genera la estrategia de búsqueda asociada al interés del usuario.

Agente-PU: construye y mantiene el perfil del usuario.

Agente-Bi: se encargan de encontrar los objetos educativos (OA) que satisfacen la temática y las restricciones de interés del usuario en los distintos repositorios.

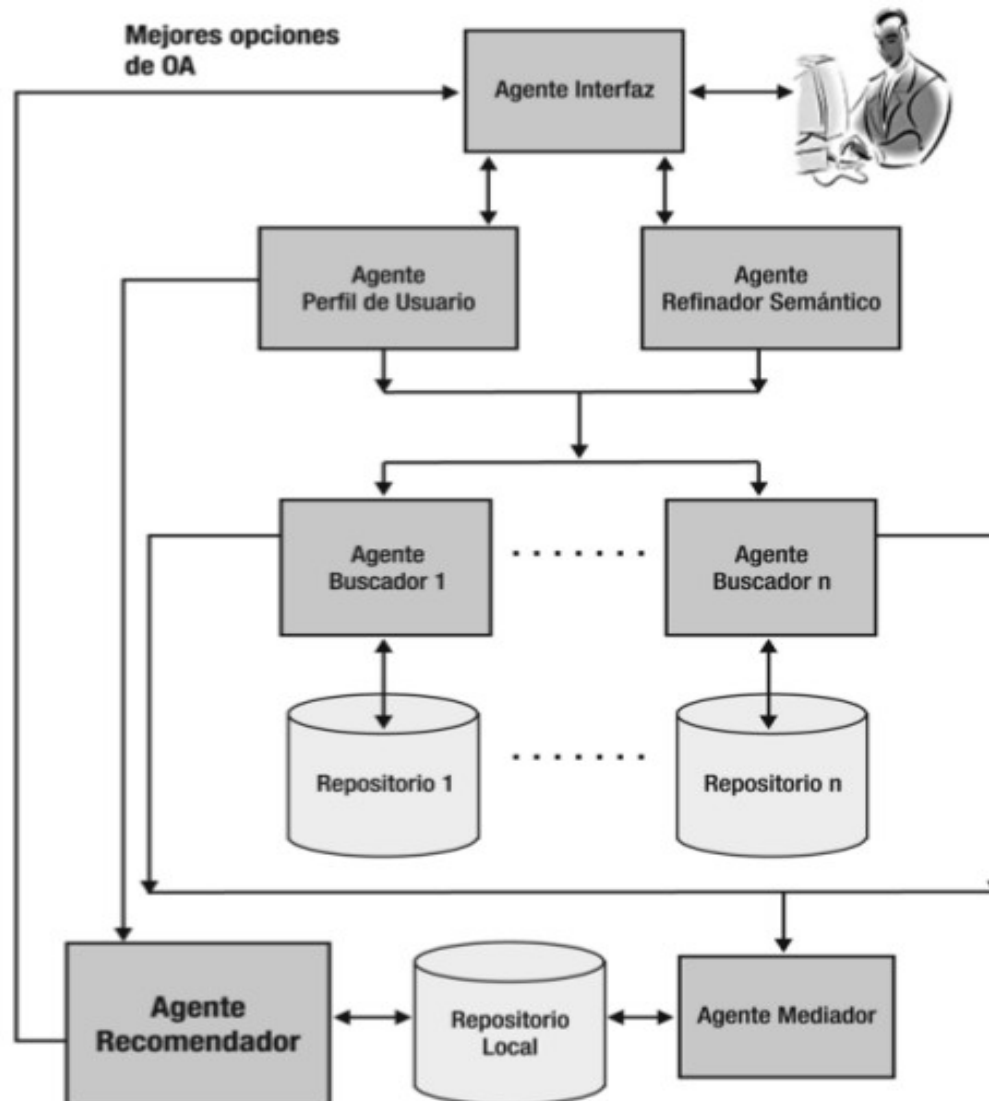
Agente-M: integra lo encontrado por cada Agente-Bi y soluciona conflictos.

Agente-R: selecciona los mejores objetos de acuerdo al perfil del usuario.

BÚSQUEDAS INTELIGENTES

SISTEMAS RECOMENDADORES

SISTEMA RECOMENDADOR PARA LA RECUPERACIÓN DE OBJETOS DE APRENDIZAJE



BÚSQUEDAS INTELIGENTES

SISTEMAS RECOMENDADORES

SISTEMA RECOMENDADOR PARA LA RECUPERACIÓN DE OBJETOS DE APRENDIZAJE

Diseño del Agente Recomendador (Agente-R)

- Su objetivo es encontrar el Objeto de Aprendizaje que satisfaga más al usuario, teniendo en cuenta la temática, las restricciones y las preferencias (su perfil).
 - Las **preferencias** que el usuario tiene respecto a los OA (por ejemplo, el idioma o el contexto académico).
 - Las **restricciones** (por ejemplo, su duración máxima).
 - Tanto las preferencias como las restricciones pueden ser graduadas (valores en el intervalo $[0,1]$)

BÚSQUEDAS INTELIGENTES

SISTEMAS RECOMENDADORES

SISTEMA RECOMENDADOR PARA LA RECUPERACIÓN DE OBJETOS DE APRENDIZAJE



Diseño del Agente Recomendador (Agente-R)

- El grado de intención asociado a cada OA dependerá de:
 - Las restricciones del usuario.
 - El grado de sus preferencias.
 - La satisfacción esperada de las preferencias a través de un OA que cuenta con ciertas características (metadatos).
 - También es posible considerar la confianza en la fuente de la información (grado de confianza).

BÚSQUEDAS INTELIGENTES

SISTEMAS RECOMENDADORES

SISTEMA RECOMENDADOR PARA LA RECUPERACIÓN DE OBJETOS DE APRENDIZAJE



Diseño del Agente Recomendador (Agente-R)

- Tomaron como base al modelo **g-BDI** (Sistema multi-contextos compuesto por contextos y reglas que representan sus relaciones).
- Se tienen contextos para:
 - Representar sus creencias (Belief Context).
 - Representar sus deseos (Desire Context).
 - Representar sus intenciones (Intention Context).

BÚSQUEDAS INTELIGENTES

SISTEMAS RECOMENDADORES

SISTEMA RECOMENDADOR PARA LA RECUPERACIÓN DE OBJETOS DE APRENDIZAJE

Diseño del Agente Recomendador (Agente-R)

Belief Context: modela las características que presentan los OAs. Estas características están definidas a través de metadatos definidos de acuerdo al estandar LOM y escritos en XML. Seleccionaron un subconjunto de metadatos, los que consideraron mas reelevantes para el Recomendador (idioma, contexto academico, etc).

Con estos datos del objeto, el agente aplicará las reglas para inferir su grado de creencia b_{ij} en que un objeto O_i satisfaga las distintas preferencias p_j del usuario, representado por la fórmula: $B(b_{ij}; O_i; p_j)$.

Algunas de los metadatos que usaron son:

- Metadato Interaccion.
- Metadato Rol.
- Metadatos Contexto académico y Nivel de Conocimiento.
- Metadatos Estilo de Aprendizaje.

BÚSQUEDAS INTELIGENTES

SISTEMAS RECOMENDADORES

SISTEMA RECOMENDADOR PARA LA RECUPERACIÓN DE OBJETOS DE APRENDIZAJE

Diseño del Agente Recomendador (Agente-R)

Desire Context: El deseo global del Agente-R es encontrar el OA que más satisfaga al usuario, teniendo en cuenta la temática, las restricciones y las preferencias.

En este contexto se representan las preferencias que el usuario tiene respecto a los OA (idioma, contexto academico) y las restricciones (duración máxima). Estas pueden ser graduadas, expresando lo que el usuario desea o rechaza en diferentes grados.

Intention Context: En este caso, las intenciones serán los objetivos educativos que se intentan alcanzar a través del mejor objeto seleccionado. Las intenciones dependeran de las restricciones del usuario respecto de los OA, sus preferencias y tambien de la satisfacción esperada de las preferencias a través de un recurso educativo.

Por ejemplo, en que medida un OA cuyo tipo de recurso está catalogado en su metadato como [ejercicio, lectura] satisface a la preferencia de que el recurso sea “práctico”.

BÚSQUEDAS INTELIGENTES

SISTEMAS RECOMENDADORES

SISTEMA RECOMENDADOR PARA LA RECUPERACIÓN DE OBJETOS DE APRENDIZAJE



Ejemplo:

Supongamos que José es un estudiante de Ingeniería que está buscando documentos con información sobre Matrices dado que está cursando un primer curso de Álgebra. Por otra parte, tiene mucho conocimiento sobre el idioma inglés, pero no sobre el portugués. José prefiere que el estilo de aprendizaje sea práctico.

Además, desearía que el tiempo que le lleve el desarrollo del material no sea superior a 60 minutos, y que el documento a buscar sea gratuito.

BÚSQUEDAS INTELIGENTES

SISTEMAS RECOMENDADORES

SISTEMA RECOMENDADOR PARA LA RECUPERACIÓN DE OBJETOS DE APRENDIZAJE

Preferencia	Prioridad	Descripción
Idioma="español"	1.0	Idiomas que el usuario desea que presenten los documentos.
Idioma="inglés"	0.8	
Rol="estudiante"	1.0	Si el que realiza la búsqueda lo hace como estudiante o docente.
Interacción="baja"	0.7	Grado de interacción que el usuario desea tener con el objeto.
Contexto Académico="universitario"	0.7	Nivel educativo en que el usuario desea el material.
Nivel de Conocimiento="bajo"	1.0	Nivel de dificultad que el usuario desea sobre el tema.
Estilo de Aprendizaje="práctico"	1.0	
Restricción		Descripción
Duración Máxima="60"	0.7	Tiempo máximo que el usuario espera que le tome trabajar con el objeto.
Costo="no"	1.0	Si restringe la búsqueda a los recursos que no sean pagos o permite los que tienen costo.

Extracto del perfil del usuario correspondiente a Jose

BÚSQUEDAS INTELIGENTES

SISTEMAS RECOMENDADORES

SISTEMA RECOMENDADOR PARA LA RECUPERACIÓN DE OBJETOS DE APRENDIZAJE

Metadatos	O_1	O_2	O_3	O_4
Language	inglés	español	español	inglés
Format	doc	pdf	doc	pdf
Learning Resource Type	lectura	[ejercicio, lectura]	diapositiva	ejercicio
Interactivity Level	bajo	bajo	bajo	alto
Intended End User Role	estudiante	estudiante	estudiante	docente
Context	universitario	universitario	secundario	universitario
Difficulty	media	baja	media	alta
Typical Learning Time	40	50	20	50
Cost	no	no	no	no

Metadatos relevantes de los OAs recuperados

BÚSQUEDAS INTELIGENTES

SISTEMAS RECOMENDADORES

SISTEMA RECOMENDADOR PARA LA RECUPERACIÓN DE OBJETOS DE APRENDIZAJE

- A partir de un conjunto de reglas en el **Belief Context** del **Agente-R** se calcula el grado de satisfacción esperada de cada una de las preferencias del usuario.

Ejemplos de estas reglas son:

- (R1) Si InteractivityLevel(Oi)=bajo
Entonces $B(Oi, \text{interactividad}=\text{baja}, 1)$
- (R2) Si InteractivityLevel(Oi)=medio
Entonces $B(Oi, \text{interactividad}=\text{baja}, 0.6)$
- (R3) Si Context(Oi)=universitario and Difficulty(Oi)=media
Entonces $B(Oi, (\text{contexto}=\text{universitario}) \text{ and } (\text{nivel}=\text{bajo}), 0.7)$
- (R4) Si Context(Oi)=secundario and Difficulty(Oi)=alta
Entonces $B(Oi, (\text{contexto}=\text{universitario}) \text{ and } (\text{nivel}=\text{bajo}), 0.8)$
- (R5) Si LearningResourceType(Oi)=ejercicio
Entonces $B(Oi, \text{estilo}=\text{práctico}, 1)$
- (R6) Si LearningResourceType(Oi)=[ejercicio, lectura]
Entonces $B(Oi, \text{estilo}=\text{práctico}, 0.6)$

BÚSQUEDAS INTELIGENTES

SISTEMAS RECOMENDADORES

SISTEMA RECOMENDADOR PARA LA RECUPERACIÓN DE OBJETOS DE APRENDIZAJE

El **Agente-R** a partir de las restantes restricciones y preferencias va a seleccionar los que sean más adecuados para él (intenciones de alcanzar sus objetivos), considerando:

- La prioridad dada por el usuario a cada preferencia (p_k, d_k).
- El grado de satisfacción b_{ik} de cada preferencia p_k por las característica de un recurso.
- Como el promedio de las satisfacciones esperadas de las distintas preferencias

$$I(O_i, P) = \frac{\sum_{j=1}^n d_j \times b_{ij}}{n} \quad , \text{ donde } P = p_1 \wedge p_2 \wedge \dots \wedge p_n$$

BÚSQUEDAS INTELIGENTES

SISTEMAS RECOMENDADORES

SISTEMA RECOMENDADOR PARA LA RECUPERACIÓN DE OBJETOS DE APRENDIZAJE

- Computando estos valores para los cuatro objetos se tienen los siguientes grados de intención:

$$I(O1, P) = 0.76 \quad I(O2, P) = 0.80 \quad I(O3, P) = 0.71 \quad I(O4, P) = 0.56$$

- Finalmente, el **Agente-R** ordena los objetos según el valor decreciente del grado obtenido: O2, O1, O3, O4.
- Luego el sistema recomienda a José el curso O2 como el más adecuado y además, le da una lista ordenada de las restantes alternativas.

Fin del Ejemplo.

BÚSQUEDAS INTELIGENTES

SISTEMAS RECOMENDADORES

SISTEMA RECOMENDADOR PARA LA RECUPERACIÓN DE OBJETOS DE APRENDIZAJE



- Implementaron un prototipo del **Agente Recomendador (Agente-R)**.
 - Utilizaron los lenguajes SWI-Prolog y Ruby.
- Las pruebas con los casos de uso fueron satisfactorias.
- Durante la experimentación, encontraron un problema.
 - Información incompleta en los metadatos educativos de los OA en los repositorios evaluados.

Bibliografía



- Bender, C. M., Deco, C., Tópicos avanzados de Bases de datos.
- Baeza-Yates, R., Ribeiro-Neto, B- (eds.), Modern Information Retrieval. New York. ACM Press, 1999.
- A Recommender System for Learning Objects Personalized Retrieval.
- Van Rijsbergen, C. J. Information Retrieval. Butterworths, 1979.
- Más las referencias citadas en las slides.