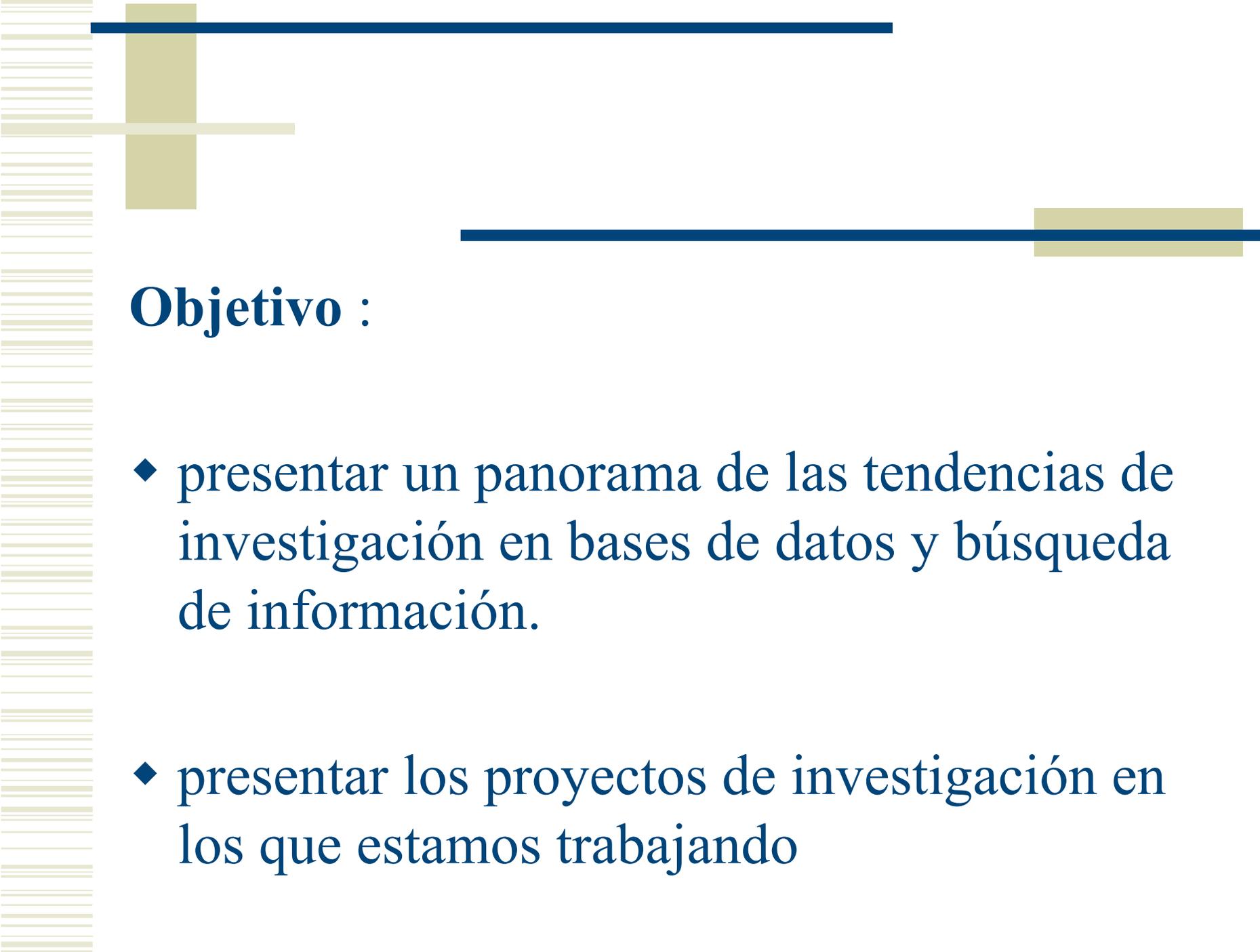




Tendencias actuales de Investigación en Bases de Datos

Claudia Deco - Cristina Bender

Mails: deco@fceia.unr.edu.ar
bender@fceia.unr.edu.ar



Objetivo :

- ◆ presentar un panorama de las tendencias de investigación en bases de datos y búsqueda de información.
- ◆ presentar los proyectos de investigación en los que estamos trabajando



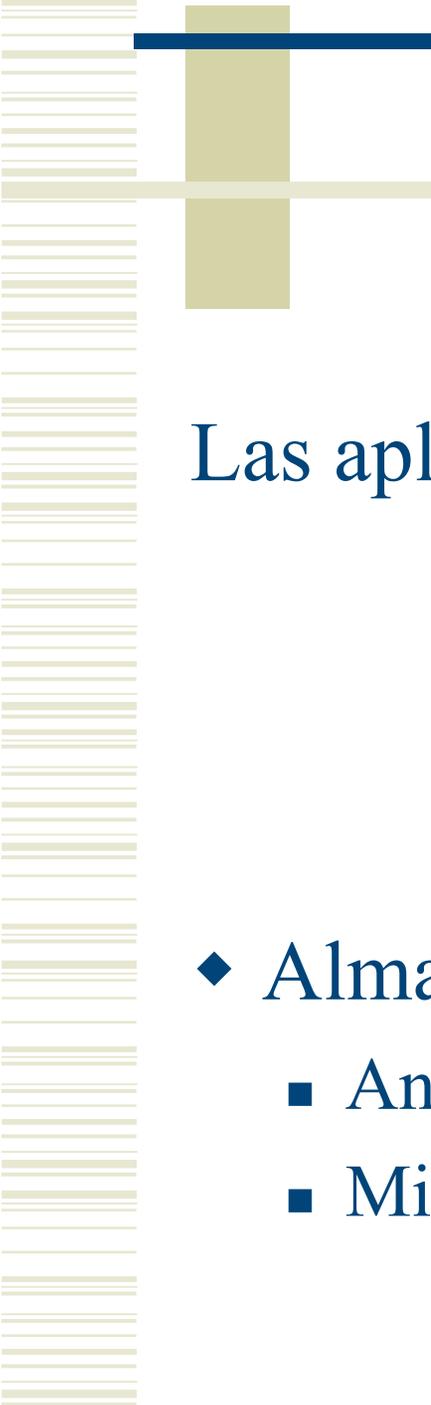
En la actualidad, se trabaja en ...

Datawarehouse

Bases de datos espaciales, temporales

Sistemas de recuperación de información

Datos en la Web



Datawarehouse

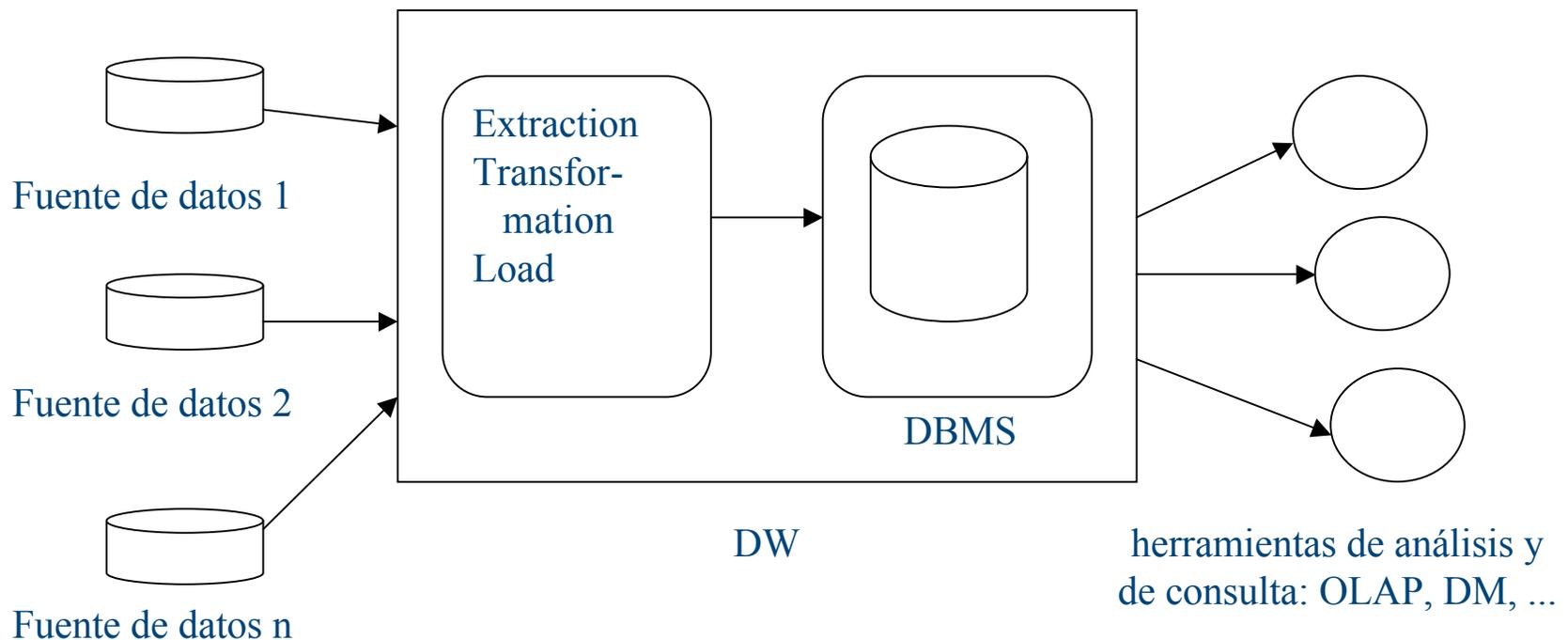
Las aplicaciones de BD se pueden clasificar en:

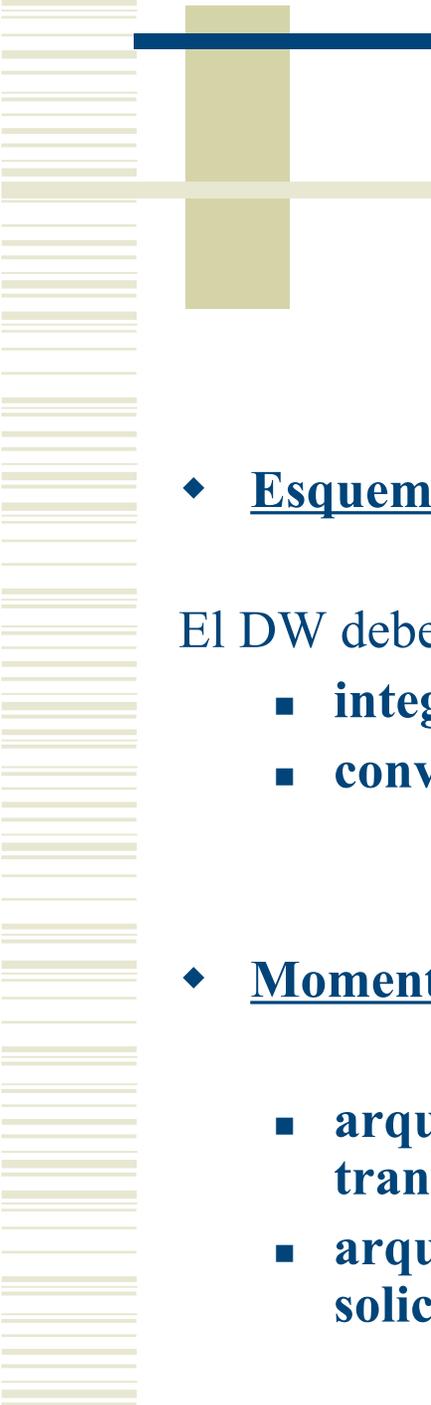
- ◆ procesamiento de transacciones
- ◆ sistemas de ayuda a la decisión

- ◆ Almacenes de datos (Datawarehouse).
 - Análisis de datos (OLAP).
 - Minado de datos (Datamining).

Datawarehouse (DW)

es un depósito de información **integrada** a partir de **varias fuentes** guardada según un **esquema unificado** en un **único lugar**.





Problemas

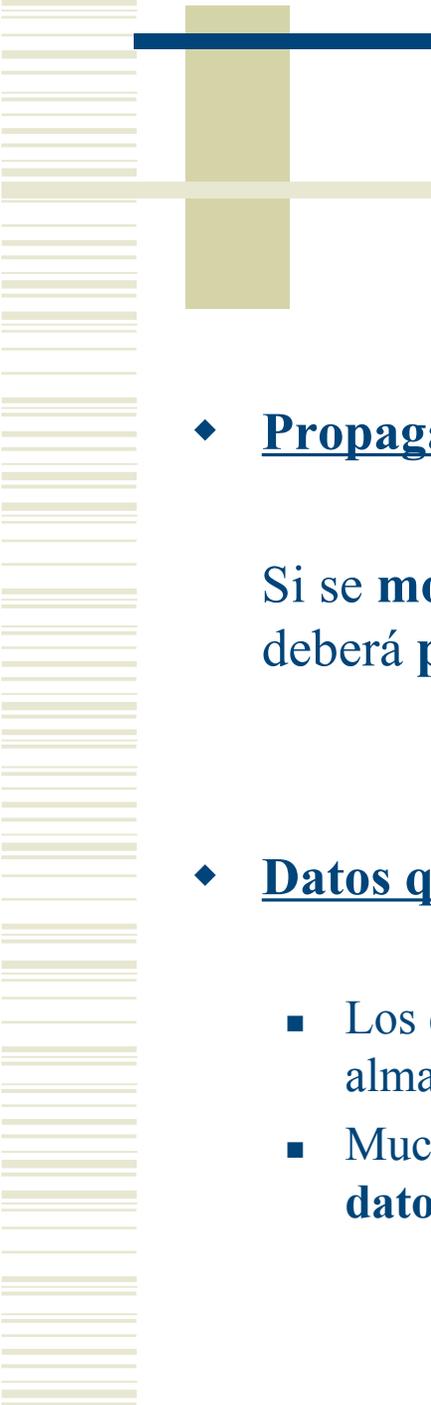
◆ Esquema que debe utilizarse

El DW debe:

- **integrar** los **esquemas** (diseño) de las distintas fuentes de datos y
- **convertir** los **datos** al esquema integrado antes de guardarlos.

◆ Momento y manera de recoger los datos

- **arquitectura orientada a orígenes de datos** (las fuentes de datos transmite la información nueva)
- **arquitectura orientada a destinos de datos** (el almacén de datos solicita los datos a las fuentes).



Problemas

- ◆ Propagación de las actualizaciones

Si se **modifican los esquemas** (diseño) de los orígenes de datos, esto deberá **propagarse al DW**.

- ◆ Datos que se deben resumir

- Los **datos** generados por sistemas **transaccionales** son **grandes** para almacenarlos.
- Muchas consultas para la **toma de decisión** se pueden resolver a partir de **datos resumidos obtenidos por agregación**.

OLAP: Análisis de datos

Para **resumir** los datos se utilizan:

- ◆ **funciones de agregación**

- en SQL son limitadas => se **extiende** SQL

- ◆ **histogramas**

- existen **extensiones** de SQL

```
SELECT percentil, avg(saldo) FROM cuenta
GROUPBY N_tile(saldo, 10) AS percentil
```

- ◆ **agregación sobre varios atributos**

- tablas de **referencias cruzadas**

```
SELECT color talla, sum(numero) FROM ventas
GROUPBY color, talla WITH CUBE.
```

Modelo n-dimensional (n atributos), generará un **n-cubo de 2ⁿ vértices**

Se definen operadores de refinamiento de consultas sobre n-cubos



Data Mining

Es la **búsqueda de información** de importancia ó
“descubrimiento del conocimiento”
en **grandes volúmenes de datos**.

- ◆ La información obtenida puede representarse mediante un conjunto de **reglas**.

Ejemplo:

“las mujeres jóvenes con ingresos mensuales superiores a \$5000 son las personas con mayores probabilidades de comprar coches deportivos de tamaño pequeño”.

Tipos de reglas

Clasificación:

Encontrar reglas que dividan los datos en **grupos disjuntos**.

Intenta **descubrir reglas** que clasifiquen datos nuevos en **base a datos históricos**

Ejemplo: **clasificar a un nuevo cliente** para conceder ó no un crédito.

Asociación:

Encontrar reglas que **asocien** datos de las transacciones.

Ejemplo: \forall transacción T, compra (T, pan) \rightarrow compra (T, leche)

donde: T es una variable, rango(T) = {todas las transacciones}.



Bases de Datos Temporales

Ejemplo:

¿Quién ha trabajado en ventas más tiempo de lo que Juan trabajó con un sueldo de \$20000?

Aplicaciones:

- ◆ Bancarias,
- ◆ Controles de personal,
- ◆ Registros médicos,
- ◆ Inventario,
- ◆ Programación de reservas (avión, tren, hoteles...),
- ◆ Aplicaciones científicas (ej. monitoreo del tiempo).

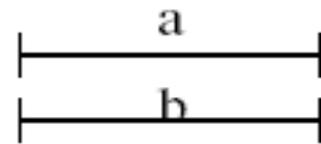
- 
- ◆ En una base de **datos** temporal cada hecho registrado tiene una **marca de tiempo**.
 - ◆ **Tiempo válido:** conj. de intervalos de t durante los que el hecho es **verdadero**.
 - ◆ **Tiempo de transacción:** intervalo de t durante el cual ese hecho es **cierto en el sistema de BD**.
 - ◆ **Relación temporal:**
 - cada tupla tiene un t asociado cuando es verdadera.
 - Se define **Algebra Temporal, SQL Temporal**



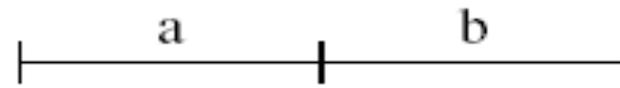
a before b



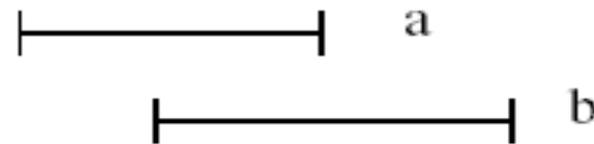
a equals b



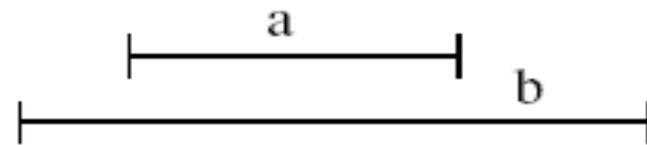
a meets b



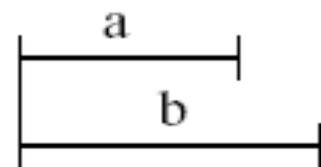
a overlaps b



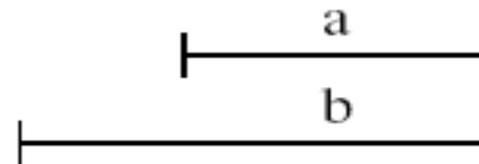
a during b



a starts b



a finishes b





Bases de Datos Espaciales

Registran información en **puntos, líneas y regiones**.

El espacio de interés puede ser:

- Una abstracción bidimensional de la superficie de la tierra.
- Un modelo del cerebro humano.
- Una representación 3D de la disposición de una cadena de moléculas de proteína.
- etc.

Qué necesita ser representado?

point



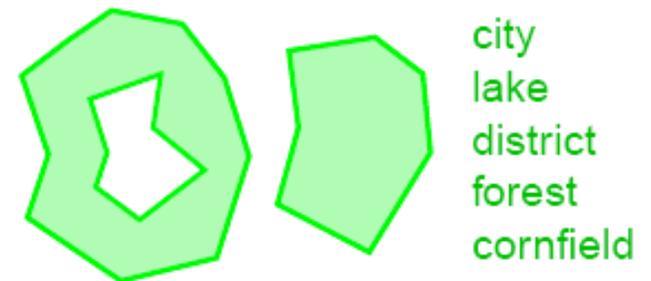
(location of object in space
but not its extent relevant)

line



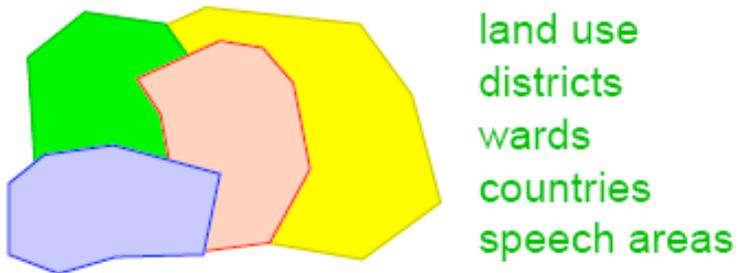
(connections in space,
movement through space)

region

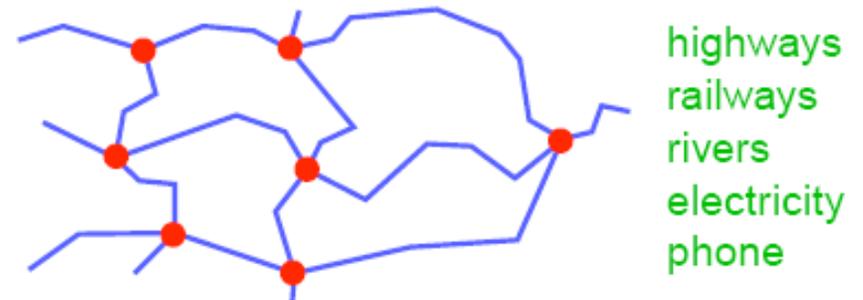


(extent of an object relevant)

partition



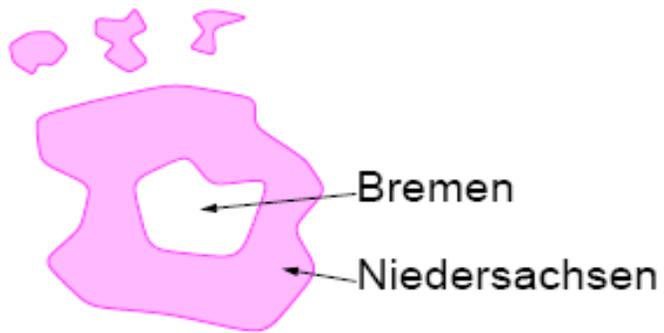
Spatially embedded *network* (graph)



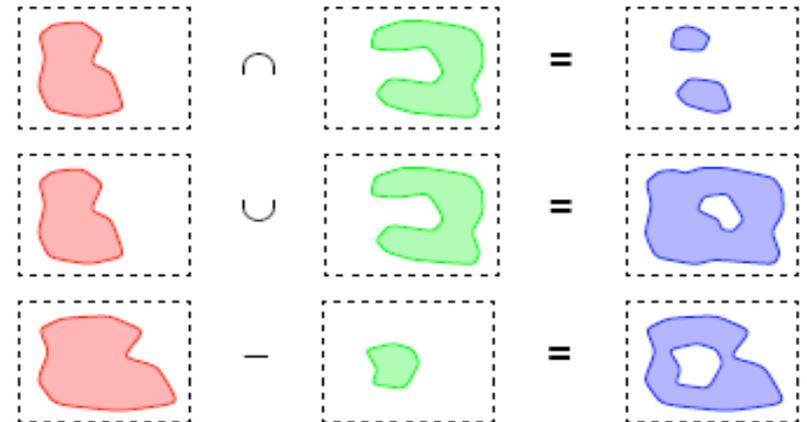
Others: nested partitions, digital terrain models

Algunas operaciones

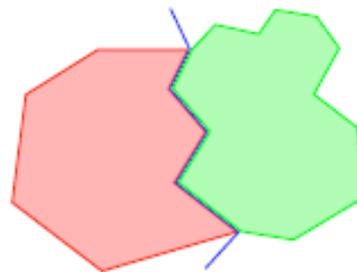
application-driven requirements



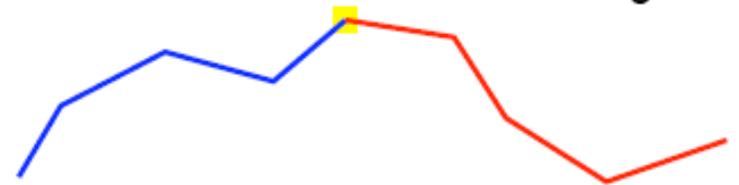
formal requirements



adjacent regions



meeting lines





Operaciones Espaciales

- ◆ **Predicados espaciales que devuelvan valores booleanos.**
 - Relaciones topológicas: igual, disjunto, adyacente, intersección, cubre, contiene, fuera, etc.
 - Orden espacial: detrás, en_frente, debajo, por_sobre, etc.
 - Relaciones direccionales: norte, sur, este, noreste, etc.
- ◆ **Operaciones espaciales que devuelvan valores numéricos.**
 - Área, perímetro, diámetro, distancia, maxdist, mindist, etc.
- ◆ **Operaciones espaciales que devuelva nuevos objetos espaciales.**
 - Operaciones de construcción: unión, intersección, diferencia, centro, borde, etc.
 - Operaciones de transformación: extender, rotar, trasladar, etc.
- ◆ **Operaciones en colecciones de objetos espacialmente relacionados.**
 - Operaciones generales: voronoi, mascercano, componer, descomponer, etc.
 - Operaciones para particiones: fusión, superimposición, cubrir, etc.
 - Operaciones para redes: camino_mas_corto, etc.

Ejemplos

Encontrar los nombres de todos los países que son vecinos de USA.

```
SELECT    C1.Name AS "Vecinos de USA"  
FROM      Country C1, Country C2  
WHERE     Touch(C1.Shape, C2.Shape) = 1 AND C2.Name = 'USA'
```

Encontrar los países que atraviesan todos los ríos listados en la tabla River.

```
SELECT    R.Name, C.Name  
FROM      River R, Country C  
WHERE     Cross(R.Shape, C.Shape) = 1
```



Bases de Datos Espacio - Temporales

Es un sistema de base de datos cuyos objetos tienen una **geometría que cambia a lo largo del tiempo**

-> sistemas que tienen la capacidad de gestionar geometrías en cambio continuo

Por ejemplo:

- ◆ **Sistema de control del tráfico**
- ◆ **Sistema de gestión catastral** (parcelas cambian su forma con el tiempo)

Ejemplo en STSQL

Se agregan nuevos tipos de datos:

“moving point” y “moving region” :

Flights (id:string, from:string, to:string, route: *mpoint*)

Encontrar una ruta entre dos instantes de tiempo:

```
SELECT      trajectory (Route(7:00..9:00))  
FROM        flights  
WHERE       id="UA207"
```



Bases de Datos y Web

Tecnologías de las dos culturas

La **Web** nos provee de:

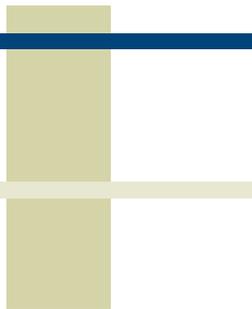
- ◆ Una **infraestructura** global y un conjunto de **estándares** que soportan el **intercambio** de **documentos**.
- ◆ Un formato de **presentación** para **hipertextos**. (HTML)
- ◆ **Interfaces** bien diseñadas para **recuperación de documentos**. (Técnicas de recuperación de información)
- ◆ Es la **base de datos más grande**.



Tecnologías de las dos culturas

Las **bases de datos**, nos ofrecen:

- ◆ Técnicas de **almacenamiento** y **lenguajes de consulta**, que proveen acceso eficiente a grandes cuerpos de datos muy estructurados.
- ◆ **Modelos de datos**, y métodos para **estructurar datos**.
- ◆ Mecanismos para mantener la **integridad y consistencia** de los datos.



Surge la necesidad de un Puente para poder consultar a la Web como a una base de datos

Solución:

- Un **formato** nuevo, XML, para **intercambiar datos con estructura**.
- Un **nuevo modelo de datos semiestructurados**, que relaja la sintaxis de sistemas de base datos muy estructurados.

Datos Semiestructurados

- ◆ Son datos sin esquema o auto-descriptibles
- ◆ La información sobre la estructura está junto con los datos.
- ◆ Representación mediante una lista de etiquetas-valor.

Ejemplo:

```
{ name: {first: "Pablo", last: "Pérez"},  
  age: 44,  
  email: pablo@hotmail.com  
}
```

XML

- Componentes básicos:**
- elemento (texto)
 - etiquetas (definidas por el usuario)

Ejemplo:

```
<people>
  <person>
    <name> Alan </name>
    <age> 42 </age>
    <email> agb@abc.com </email>
  </person>
  <person>
    <name> Patsy </name>
    <age> 36 </age>
    <email> ptn@abc.com </email>
  </person>
</people>
```

DTD

- ◆ **Describen los elementos disponibles en un documento XML.**
- ◆ **Esto introduce el concepto de:**
 - **Documento-bien-formado (los tags se abren y se cierran)**
 - **Documento-válido (tiene un DTD asociado).**

Ejemplo

```
<!DOCTYPE db [  
  <!ELEMENT db (person*)>  
  <!ELEMENT person (name,age,email)>  
  <!ELEMENT name (#PCDATA)>  
  <!ELEMENT age (#PCDATA)>  
  <!ELEMENT email (#PCDATA)>  
] >
```



Lenguajes de marcado

- ◆ **XML** (eXtensible Markup Language), con sus respectivos DTD (Document Type Definition): para **intercambio** de datos
- ◆ **RDF** (Resource Description Framework): recomendado como estándar para los **metadatos**.
- ◆ **OWL** (Ontology Web Language): estándar para realizar **anotaciones de ontologías** en la web
 - es un lenguaje para publicar y compartir ontologías en la web
 - desarrollado por el W3C .

RDF

- ◆ Permite la **descripción** y el **procesamiento** de **metadatos** de cualquier dominio
- ◆ Usa **XML** como lenguaje de base

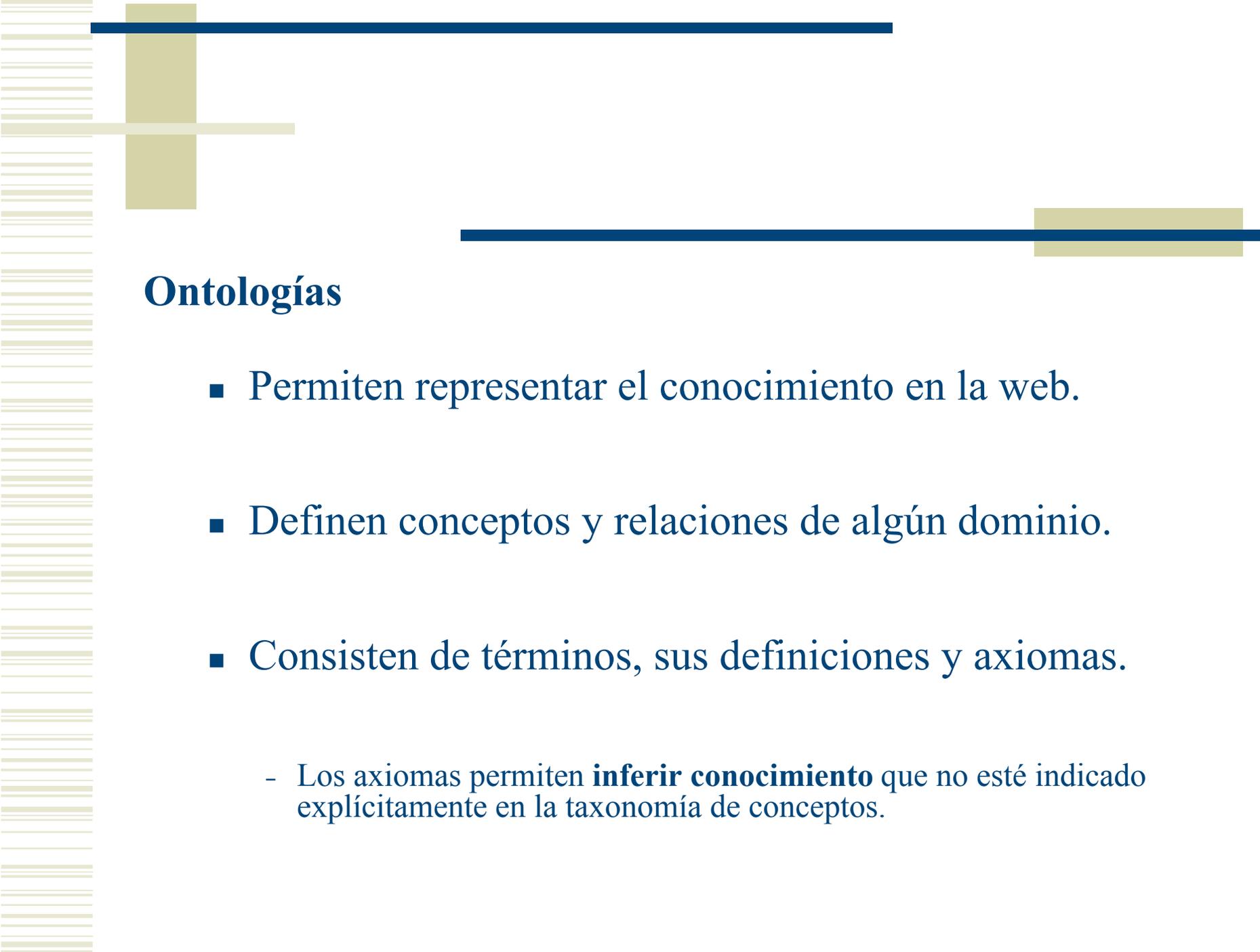
Ejemplo “La página web <http://www.infovis.net> fue creada por Juan”

Sujeto: <http://www.infovis.net> (recurso)

Predicado: creada (propiedad, tiene un creador)

Objeto: Juan (el valor de la propiedad)

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:s="http://description.org/schema/">
  <rdf:Description about="http://www.infovis.net">
    <s:Creador>Juan</s:Creador>
  </rdf:Description>
</rdf:RDF>
```



Ontologías

- Permiten representar el conocimiento en la web.
- Definen conceptos y relaciones de algún dominio.
- Consisten de términos, sus definiciones y axiomas.
 - Los axiomas permiten **inferir conocimiento** que no esté indicado explícitamente en la taxonomía de conceptos.

OWL

- ◆ Usa XML como lenguaje de base

Ejemplo

```
<owl: Class rdf:ID="Female">
```

```
<rdfs:subClassOf rdf:resource:"#Animal"/>
```

```
<owl:disjointWith rdf:resource:"#Male"/>
```

```
<owl: Class>
```

Information Retrieval

Recupera datos con la **mejor coincidencia** con el patrón dado.

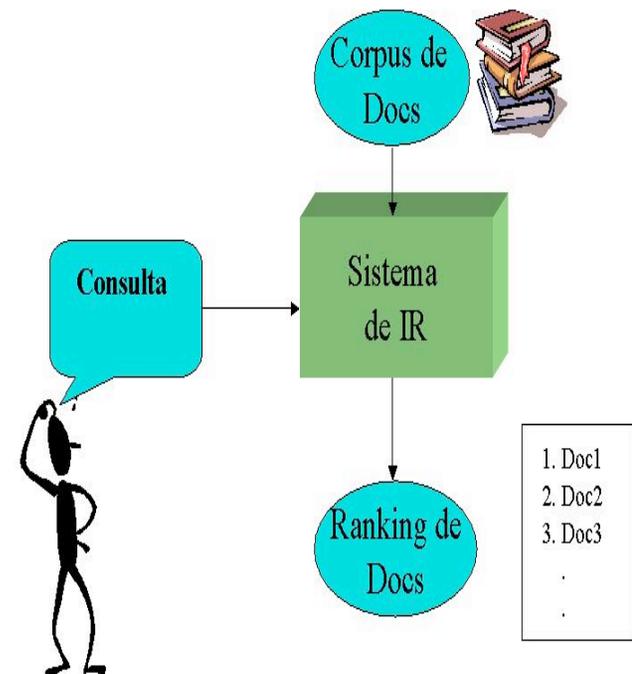
IR versus Data Retrieval

Dados:

- ♦ Un **corpus** de documentos textuales en lenguaje natural.
- ♦ Una **consulta** de usuario en la forma de un string de texto.

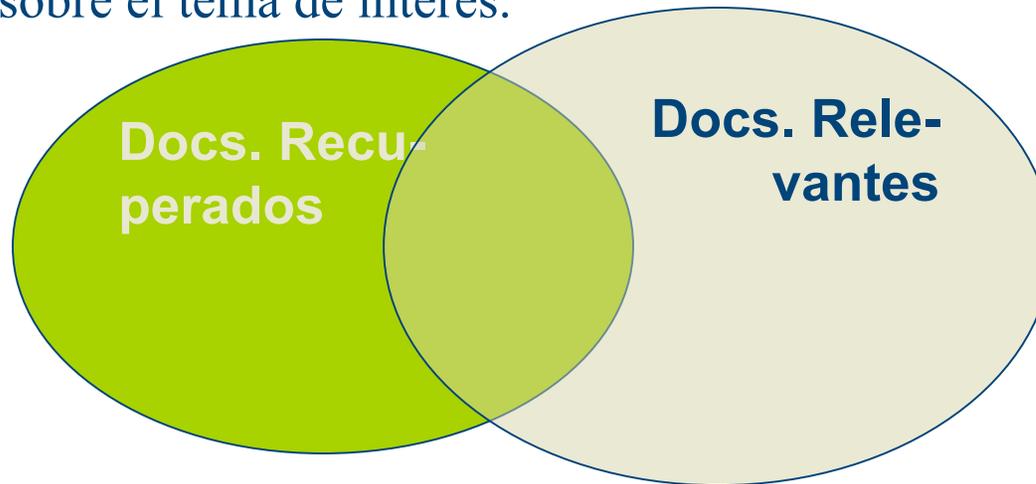
Encontrar:

- ♦ Un **conjunto rankeado** de documentos que son **relevantes** para la consulta



Realizada una búsqueda:

el conj. de docs recuperados **no coincide** totalmente con el conj. de docs relevantes sobre el tema de interés.



Una **búsqueda** será **óptima** cuando estos **dos conjuntos coincidan**

→ todos los docs recuperados sean relevantes
y todos los docs relevantes sean recuperados.



Problemas

Problemas con los datos

- **Datos distribuidos.**
- **Datos volátiles.** (dinámica de Internet)
- **Gran volumen.** (crecimiento exponencial de la Web).
- **Datos no estructurados y redundantes** (30% duplicado).
- **Calidad de los datos.** (no hay proceso ni control editorial)
- **Datos heterogéneos.** (estructural, semántica)

Problemas con los usuarios

- **Cómo especificar la consulta**



Indicadores

Para evaluar el resultado de una búsqueda:

- ♦ **Precisión:** ratio de docs relevantes sobre el número total de docs recuperados

$$\text{Precisión} = \frac{\text{Número de documentos relevantes recuperados}}{\text{Número de documentos recuperados}}$$

- ♦ **Recall:** proporción de docs relevantes que son recuperados.

$$\text{Recall} = \frac{\text{Número de documentos relevantes recuperados}}{\text{Número total de documentos relevantes}}$$



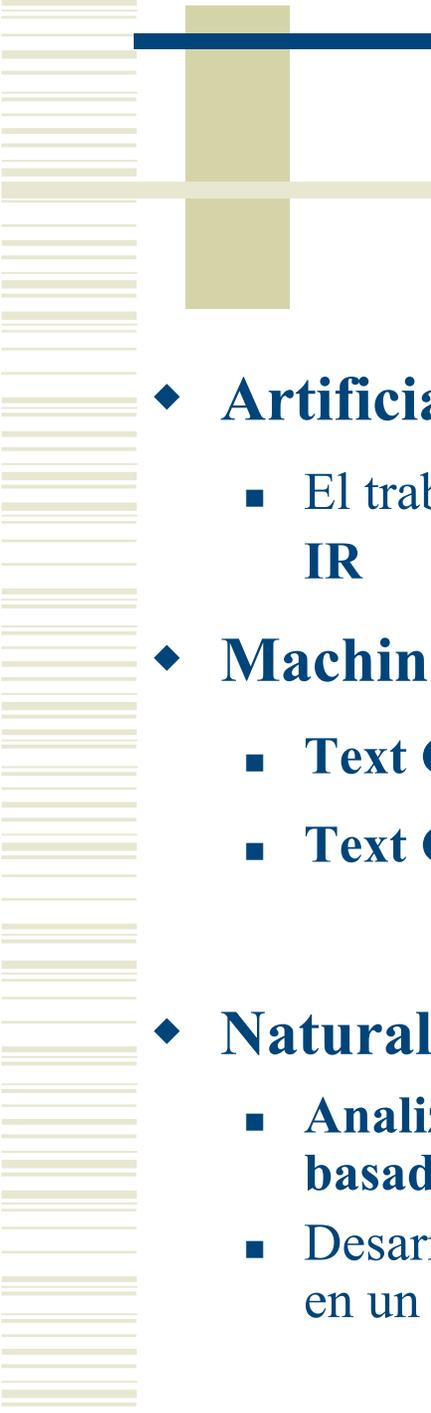
Áreas Relacionadas con IR

- ◆ Database Management
- ◆ Library and Information Science
- ◆ Artificial Intelligence
- ◆ Natural Language Processing
- ◆ Machine Learning



Areas Relacionadas

- ◆ Database Management
 - Recientemente se ha volcado a los *datos semiestructurados* (XML) y esto lo ha llevado más **cerca** de la **IR**
- ◆ Library and Information Science
 - Se focaliza en los aspectos del **usuario humano** de la IR (interacción humano-computadora, interface de usuario, visualización).
 - El trabajo reciente sobre *bibliotecas digitales* la llevó más **cerca** de las Ciencias de la **Computación** y de la **IR**



Areas Relacionadas

◆ Artificial Intelligence

- El trabajo en **ontologías** y **agentes inteligentes** la lleva más **cerca** de la **IR**

◆ Machine Learning

- **Text Categorization** Clasificación automática de **jerarquías** (Yahoo).
- **Text Clustering** Agrupamiento de **resultados** de consulta de la RI.

◆ Natural Language Processing

- **Analizar sintaxis** y **semántica** de textos puede permitir **recuperación basada en *significado*** más que en keywords.
- Desarrolla **métodos** para identificar piezas específicas de información en un documento (***information extraction***)



Estrategia de Búsqueda

Una **estrategia de búsqueda** es una expresión lógica compuesta por distintos conceptos combinados con conectores lógicos AND, OR y NOT.

- ◆ **Mejorar la precisión** depende de la **correcta preparación** de esta estrategia.



Recursos lingüísticos que se utilizan en la IR

- Diccionarios
- Diccionarios multilinguales
- Ontologías
- Tesauros



Query Expansion

Se propone un **refinamiento** semántico
que utiliza **conocimiento lingüístico**,
para la preparación
de una **estrategia** de búsqueda
que **mejore la precisión** de los resultados.



Espacios Métricos

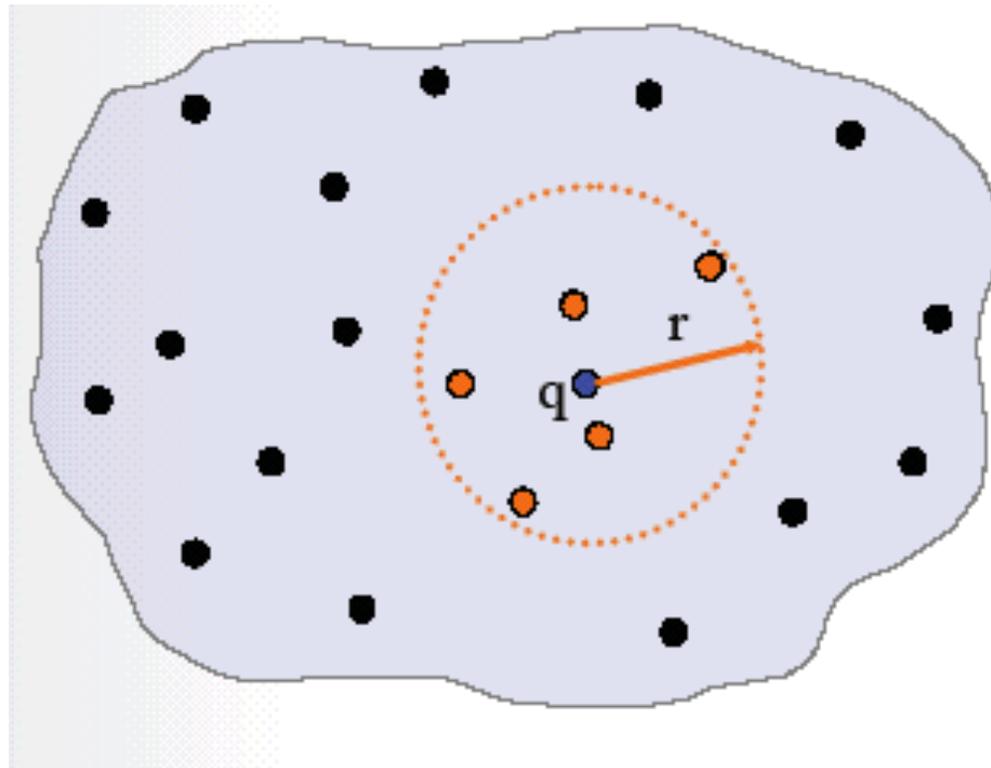
Bases de datos tradicionales.

- Los datos tienen una estructura exacta y bien definida.
- Búsquedas exactas, por igualdad/desigualdad.
- `SELECT * FROM Alumno WHERE Ciudad = 'Rosario';`

Bases de datos *no estructuradas*.

- No se pueden realizar búsquedas exactas.
- La operación más típica es la *búsqueda por similitud*.
- Algunos ejemplos : imágenes, texto, huellas digitales, música, cadenas de ADN, ...

- ◆ Búsqueda por similitud: recuperar los objetos de la base de datos más semejantes a uno dado.





Espacio métrico = Universo de objetos
+ Función de distancia

Ej. Colección de palabras + Distancia de edición

- ◆ **Def.:** Sean O_1 y O_2 dos objetos del universo de objetos posibles. La distancia (disimilaridad) se denota con $d(O_1, O_2)$
- ◆ **Propiedades de la distancia:**
 - $d(A, B) = d(B, A)$ (Simetría)
 - $d(A, A) = 0$
 - $d(A, B) \geq 0$ ($d(A, B) = 0$ sii $A = B$)
 - $d(A, B) \leq d(A, C) + d(B, C)$ (Desigualdad Triangular)

◆ Métrica de Minkowski

$$d(Q, C) \equiv \sqrt[p]{\sum_{i=1}^n (q_i - c_i)^p}$$

Si $p = 1$ Manhattan (Rectilínea, City Block)

$p = 2$ Euclidea

$p = \infty$ Máximo

◆ Distancia de edición

$$d(\text{anemia}, \text{anestesia}) = 4$$

Consultas por similitud

Dado $X \subseteq D$ en un espacio métrico $M (D,d)$

se pueden definir dos tipos básicos de consulta por similitud para una consulta $q \in D$

- Range query

$$\{x \in X; d(q,x) \leq r\}$$

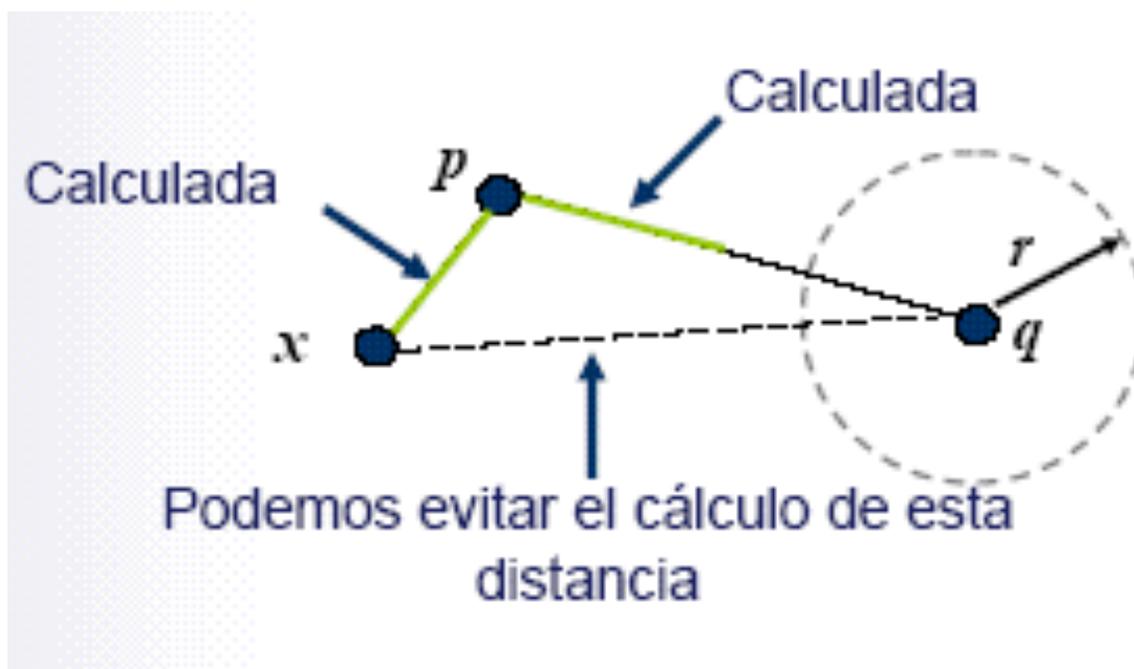
- Nearest neighbours query

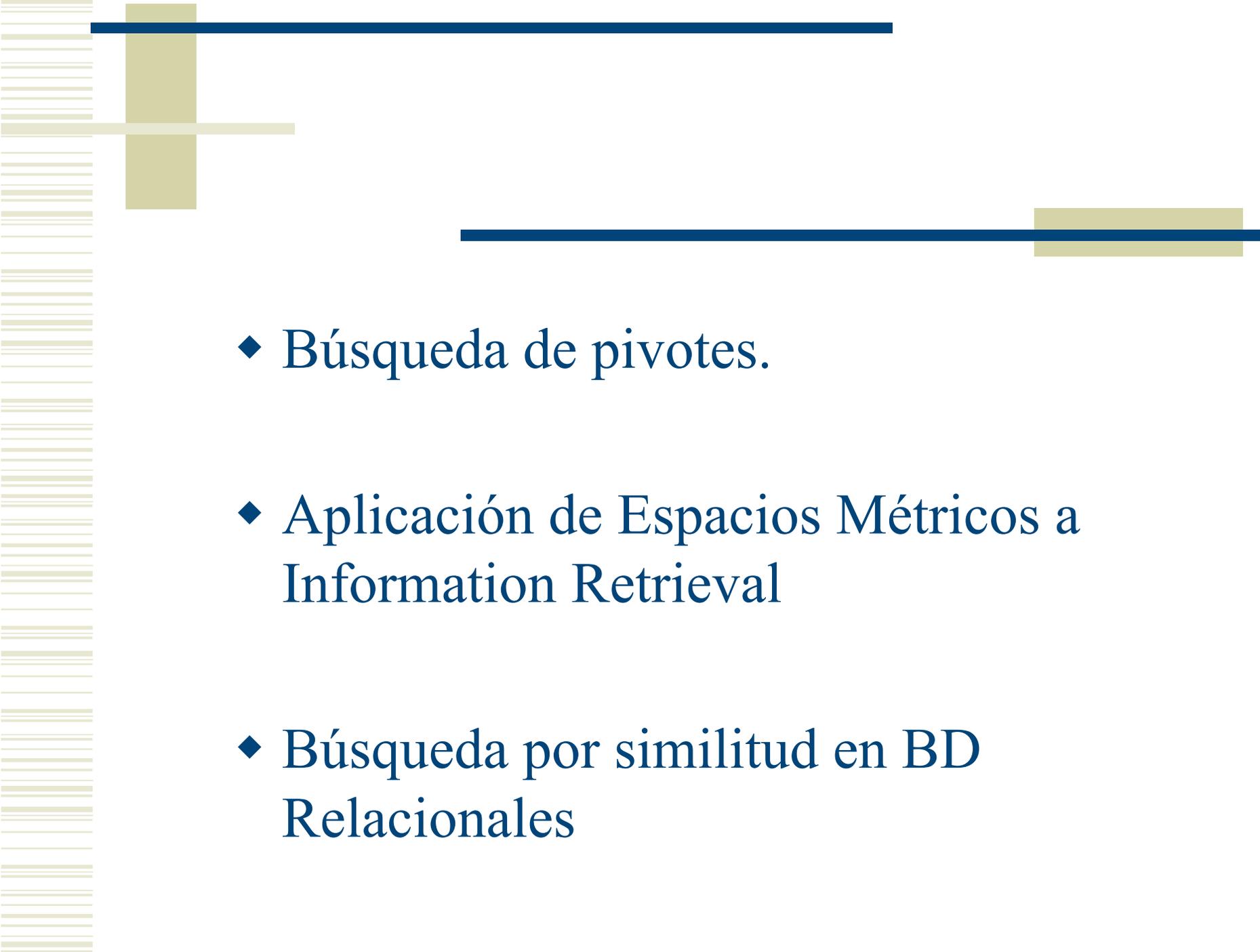
- ◆ el vecino más cercano
- ◆ los k vecinos más cercanos.

- 
- ◆ Evaluar d tiene un coste computacional elevado. **Comparar** la consulta con **toda la base de datos** es muy **costoso**.
 - ◆ Se utilizan **índices** sobre la BD para evitar la comparación de la consulta con todos los objetos de la base de datos.

La desigualdad triangular, base de los algoritmos de indexación.

$$\forall x, y, z \in U, d(x, y) \leq d(x, z) + d(z, y)$$



- 
- ◆ Búsqueda de pivotes.
 - ◆ Aplicación de Espacios Métricos a Information Retrieval
 - ◆ Búsqueda por similitud en BD Relacionales

- 
- ◆ A veces **no interesa el match exacto**.
 - ◆ Pueden interesar los **elementos** más **similares** o cercanos al elemento dado en la consulta.
 - ◆ Mostrar las **tuplas** resultantes **ordenadas según la cercanía** a la consulta.
 - ◆ Para la **formalización** del modelo propuesto se trabaja con:
 - valores **difusos**,
 - existencia de **condiciones ausentes**,
 - asignación de **pesos** a las condiciones presentes.
 - ◆ **Aplicación:** Recuperación personalizada de e-cursos



Investigación - Proyectos actuales

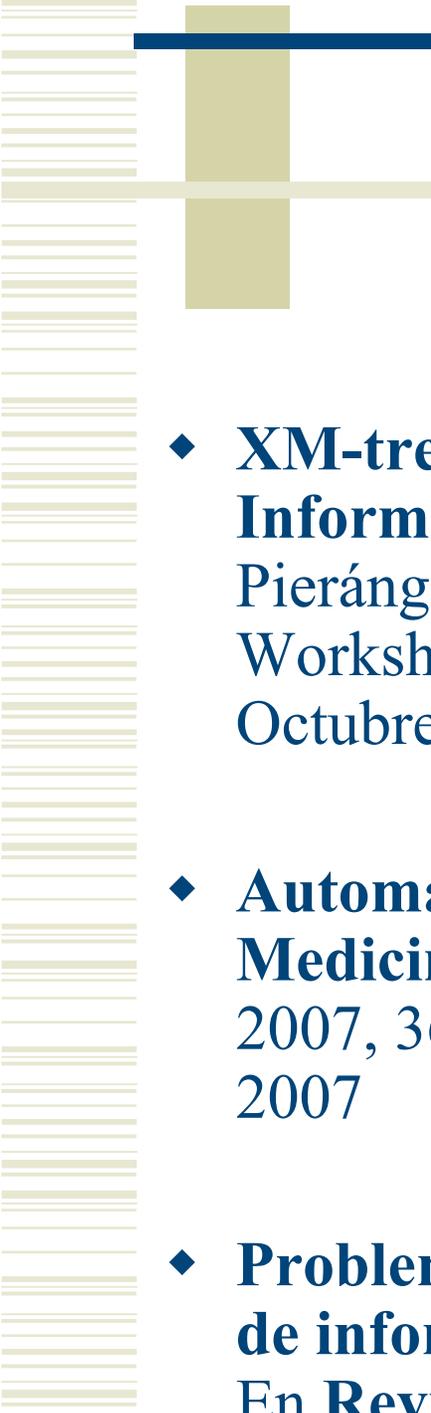
- ◆ **Búsqueda en Bases de Datos de Texto.** Directora: Claudia Deco. FCEIA - UNR - 2007-2010
- ◆ **INFOSUR: Investigación y Desarrollo.** Directora: Dra. Zulema Solana. Facultad de Humanidades y Artes-UNR- 2005-2007
- ◆ **EduCa: Red de Educación con Calidad Cultural.** Proyecto conjunto Uruguay, Argentina y Brasil. Fondo Regional para la Innovación Digital en América Latina y el Caribe (FRIDA) 2004-2006.
- ◆ **Búsqueda asistida de evidencia clínica en medicina.** FCEIA – UNR - 2004-2006.



Vinculación con otros grupos de investigación

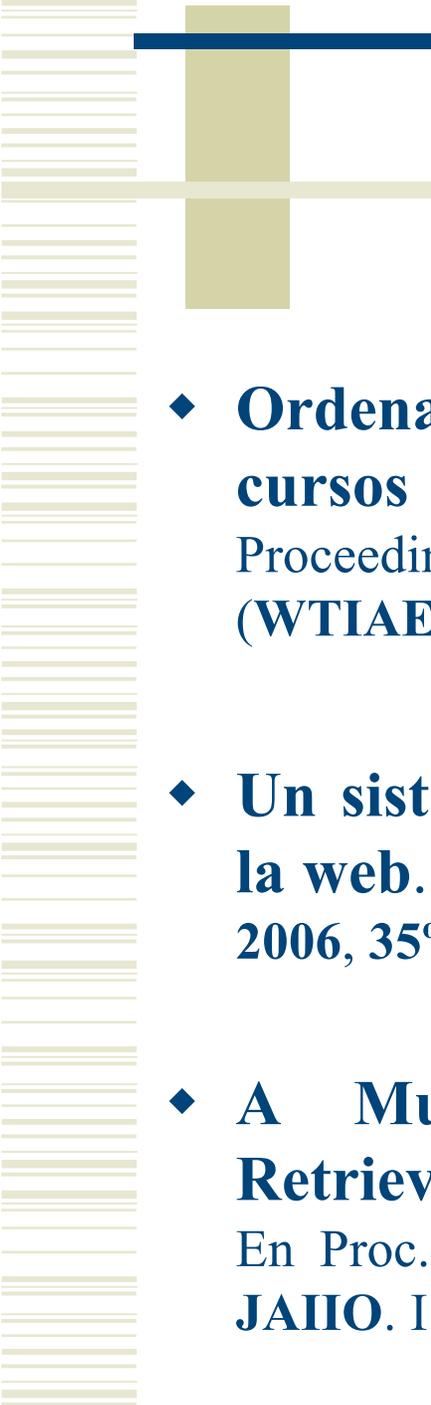
- ◆ Concepción de Sistemas de Información, UdelaR, Uruguay
- ◆ Facultad de Humanidades y Artes, UNR
- ◆ Universidad Nacional de San Luis

- ◆ Departamento de investigación institucional, UCA
- ◆ Proyecto de investigación, UCSE
- ◆ Red RITOS 2, Cytel



Resultados

- ◆ **XM-tree, un nuevo índice para Recuperación de Información en la Web.** Claudia Deco, Guillermo Pierángeli, Cristina Bender, Nora Reyes. En Proceedings Workshop de Bases de Datos. **CACIC 2007.** Corrientes. Octubre 2007.
- ◆ **Automatización en la Búsqueda de Respuestas en Medicina.** C Bender, A Dallosta, C Deco. En Proc. SIS 2007, 36° **JAIIO.** ISSN 1850 2822. Mar del Plata. Agosto 2007
- ◆ **Problemas de la traducción de la consulta en la búsqueda de información multilingüe.** C. Deco, C. Bender, M. Chiari. En **Revista Infosur.** UNR. 2007

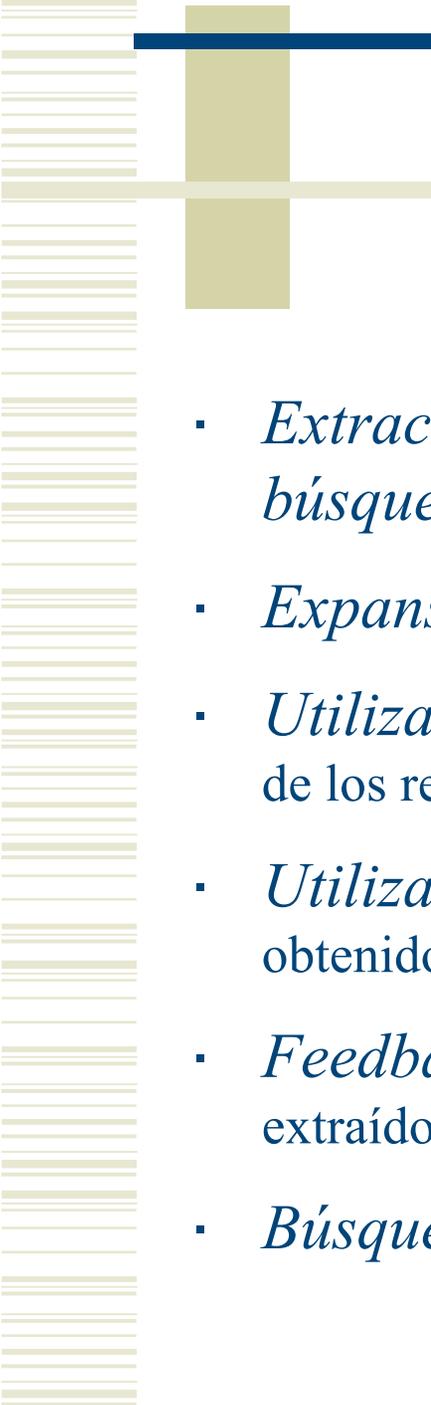
- 
- ◆ **Ordenación de tuplas para la búsqueda de múltiples e-cursos similares.** Bender C, Deco C, Bernini MB, Asás M, Motz R. En Proceedings V Workshop de Tecnología Informática Aplicada en Educación (WTIAE), CACIC 2006. San Luis, octubre de 2006.
 - ◆ **Un sistema de búsqueda asistida de información médica en la web.** C Bender, C Deco, J Plüss, A Dallosta, ML Ramírez. En Proc. SIS 2006, 35° JAIIO. ISSN 1850 2822. pp 19-28. Mendoza. Septiembre 2006
 - ◆ **A Multiagent Approach To Educational Resources Retrieval.** Ana Casali, Claudia Deco, Cristina Bender and Regina Motz. En Proc. Workshop on Artificial Intelligence for Education WAIFE, 35° JAIIO. ISSN 1850 2784. pp 35-41. Mendoza. Septiembre 2006.

- 
- ◆ **A multiagent platform for educational resources retrieval driven by cultural aspects.** C. Bender, C. Deco, A. Casali, R. Motz, J. Guzmán. En Proc. I Congr.de Tecnología en Educación y Educación en Tecnología. **TE&ET**. ISBN 950-340373-1. pp 1-9. La Plata. Agosto **2006**.
 - ◆ **Un sistema para mejorar la recuperación de información médica en la web mediante la expansión semiautomática de la consulta.** C Deco, C Bender, J Plüss, A Dallosta, ML Ramírez. En Revista **Informática y Salud**, Sociedad Española de Informática y Salud. Nro. 57, Junio **2006**, pp 91-97. ISSN 1579-8070. España
 - ◆ **Applying ontologies to educational resources retrieval driven by cultural aspects.** R Motz, J Guzmán, C Deco and C Bender. Journal of Computer Science & Technology. ISSN 1666-6038. **JCS&T** Vol 5, N° 4, pp 279-284, December **2005**.

- 
- ◆ **Semantic Refinement for Web Information Retrieval.** C Deco, C Bender, J Saer, M Chiari, R Motz. Proceedings Third Latin American Web Congress **La Web 2005.** IEEE Press.: 106-110.
 - ◆ **Capítulo 4: Expansión de consultas utilizando recursos lingüísticos para mejorar la recuperación de información en la web.** C. Deco, C. Bender, J. Saer y M. Chiari. En *Desarrollo, implementación y utilización de modelos para el procesamiento automático de textos.* UNCuyo: 35-46. ISBN: 987-575-019-0. **2005**

- 
- ◆ **Combining techniques for the classification of web pages resulting from a query.** Bender, C., Perlo, L., Deco, C., Motz, R. **Workshop Chileno de Bases de Datos, XII Jornadas Chilenas de Computación.** ISBN 956-7021-18-X. Arica, Chile, **2004**
 - ◆ **Refinamiento Semántico para Recuperación de Información desde la Web.** Motz, R., Deco, C., Bender, C., Saer, J., Chiari, M. **Proceedings Workshops on Artificial Intelligence, Iberamia 2004,** pp 172-179. ISBN: 968-863-786-6. Puebla, México, noviembre de **2004.**
 - ◆ **Búsqueda asistida de evidencia clínica en medicina en Internet utilizando tecnologías middleware.** Plüss, J; Bender, C; Deco, C; et al.. **Revista Informática y Salud de la Sociedad Española de Informática y Salud.** Nro. 47, Abril **2004,** pp 76-81. ISSN 1579-8070.

- 
-
-
- ◆ **La clasificación en la carga de Web Data Warehouses.** R. Motz, C. Deco, C. Bender, C. Manzano, L. Perló, E. Ruiz, A. von Fürth. **Jornadas Chilenas de Computación. II Workshop de Bases de Datos.** Chillán, Chile, noviembre 2003. ISBN 956-7813-27-2.
 - ◆ **Arquitectura de un asistente para la recuperación semántica de referencias bibliográficas en la web.** Motz R., Deco C., Bender C.. Anales de las 32 Jornadas Argentinas de Informática e Investigación Operativa (*JAIIO*). ISSN 1666 1141. Buenos Aires, septiembre de 2003.
 - ◆ **Utilización de ontologías y tesauros para mejorar la recuperación de la información de la web en el área salud.** Bender C., Deco C., Motz R. **IX Jornadas Iberoamericanas de Informática.** Cartagena de Indias, Colombia, agosto de 2003.



Temas abiertos en búsqueda

- *Extracción automática de conceptos para la estrategia de búsqueda*, por ejemplo a partir de una fuente de datos.
- *Expansión multilingual*. Experimentación con EuroWordNet.
- *Utilización de perfiles de usuario*. Permitiría la selección automática de los recursos lingüísticos adecuados.
- *Utilización de ontologías con axiomas*. Incorporar conceptos obtenidos a través de la inferencia.
- *Feedback de relevancia*. Mejorar la estrategia incorporando conceptos extraídos de docs marcados como relevantes por el usuario.
- *Búsqueda en espacios métricos*.