



**Information
Retrieval
Group**

**Parallel
Computing
Group**



[Escenario]

- **Grupo de Computación Paralela**

*... El grupo de Computación Paralela nació en **1998** con el objetivo de trabajar en métodos numéricos desde una perspectiva paralela ...*

- **Grupo de Recuperación de Información**

*... El grupo de Recuperación de Información nace en el año **2001** con la finalidad de desarrollar y mejorar alguna de las fases de la construcción de un SRI ...*

- **La fusión**

ORIP TIN2004-05920

[Escenario]



- **Cluster de Modelización de Mieres**

*... Proyectos de Infraestructura Científica 2005-2006
Ministerio de Educación y Ciencia ...*

[La eliminación de *Neville*]

- Sea A no singular de orden n , entonces

$$A = \bar{A}_1 \rightarrow A_1 \rightarrow \bar{A}_2 \rightarrow A_2 \rightarrow \cdots \rightarrow \bar{A}_n = A_n = U$$

$$A_t = \begin{pmatrix} a_{11}^{(t)} & a_{12}^{(t)} & \cdots & \cdots & \cdots & \cdots & a_{1n}^{(t)} \\ 0 & a_{22}^{(t)} & \cdots & \cdots & \cdots & \cdots & a_{2n}^{(t)} \\ \vdots & 0 & \ddots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \ddots & a_{t-1,t-1}^{(t)} & a_{t-1,t}^{(t)} & \cdots & a_{t-1,n}^{(t)} \\ \vdots & \vdots & \vdots & 0 & a_{tt}^{(t)} & \cdots & a_{tn}^{(t)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & a_{nt}^{(t)} & \cdots & a_{nn}^{(t)} \end{pmatrix}$$

[La eliminación de *Neville*]

- Si A es Totalmente Positiva

$$A = A_1 \rightarrow A_2 \rightarrow \cdots \rightarrow A_n = U \longrightarrow T_{Neville}(n;1) < T_{Gauss}(n;1)$$

- Aplicaciones

- Sistemas Mecánicos: análisis de vibraciones
- Tomografía
- Diseño Geométrico Asistido por Computador
- Análisis Estocástico
- Combinatoria, Teoría de Grafos
- ...

[La eliminación de *Neville*]

■ Desde la perspectiva de la CP

- Unidimensionales y Bidimensionales
- Distribuciones por bloques, cíclicos, cíclico por bloques
- *Neville* por Bloques, explotación de la caché
- Análisis de tamaño de bloque óptimo
- Tiempo de ejecución, Eficiencia, Escalabilidad,...
- ...

■ Algunos Resultados

- Eficiencia muy próxima a uno en algunos supuestos
- Mejor que Gauss en otros
- Variantes escalables

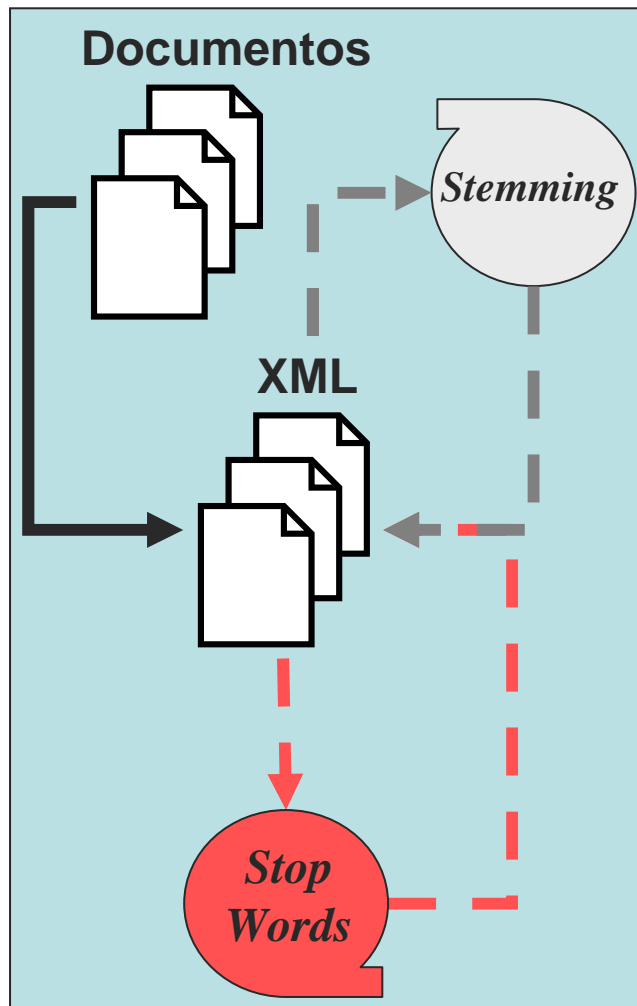
[Presente y Futuro del]

- **En la eliminación de *Neville***
 - Profundizar en Bloques (BLAS, etc.)
 - Obtención de la expresión general

- **Soporte del Sub-Grupo en IR**
 - *SERAPAD* ¿TIN 2004-2007?

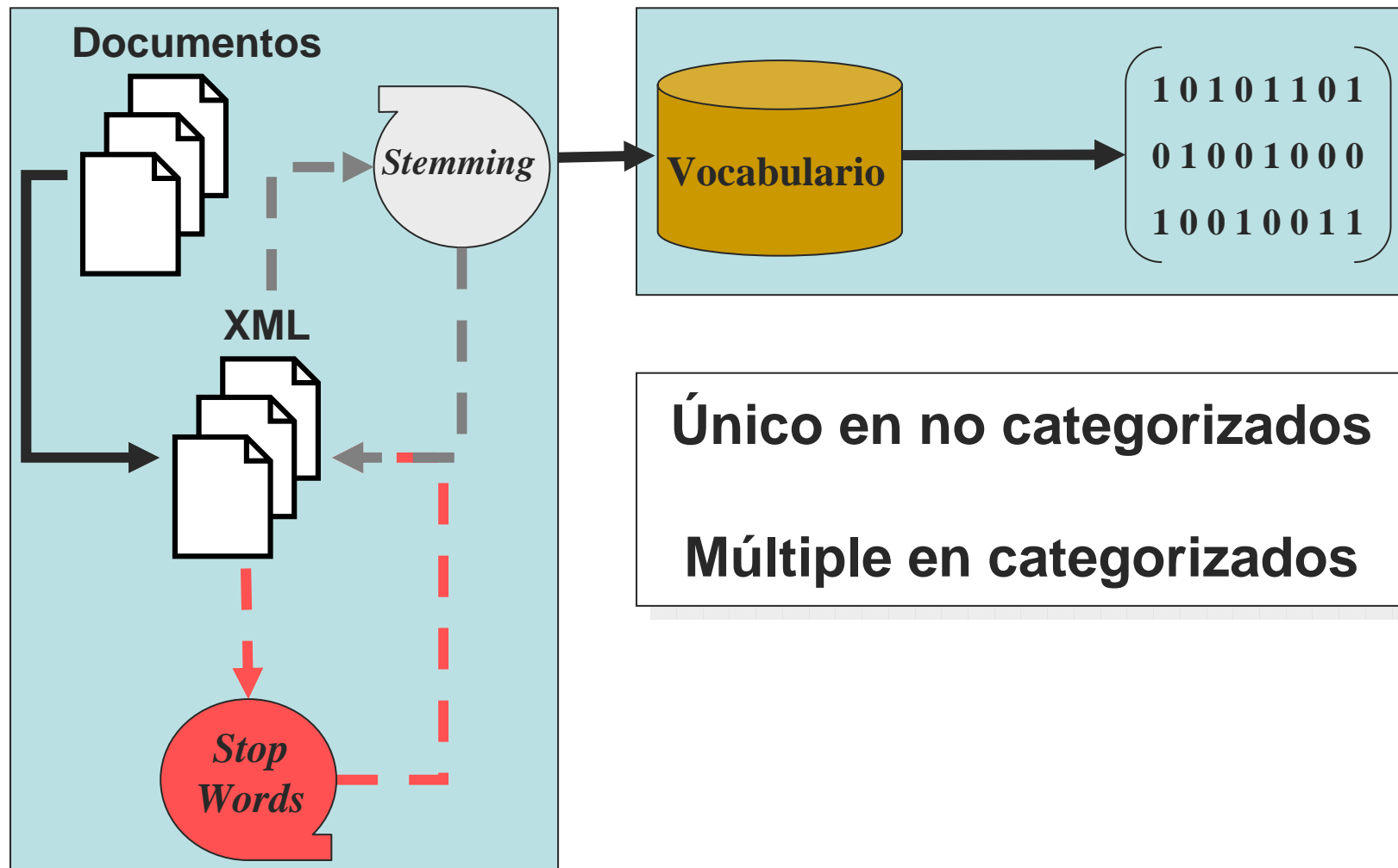
- **Búsqueda de Aplicaciones**
 - Departamentos de Física, Construcción y Química
 - *IFCA*: Grupo de Física Estadística y No lineal
 - Paralelización de algoritmos de Fracturas en Sólidos*
 - ...

[IR El Problema]

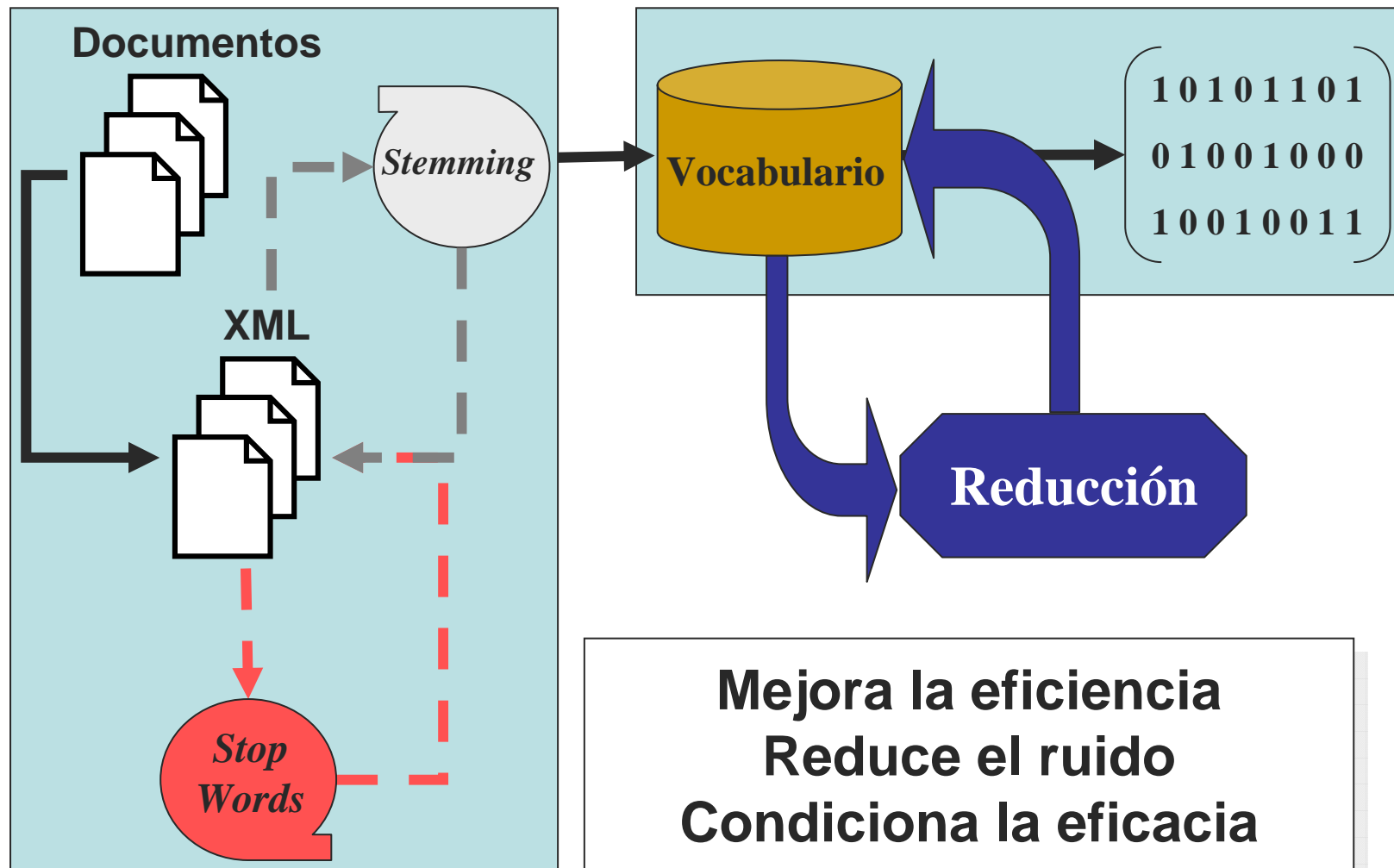


Proceso estándar

IR El Problema



IR El Problema



[IR Categorización]

■ Objetivo

- Asigna categorías a documentos nuevos

■ Problemas

- Elevada dimensionalidad
- Solapamiento entre categorías

■ Métodos

- Reducción de dimensionalidad (**selección**/extracción)
- Binarización del problema (*one-against-all*)
- *Support Vector Machines*

[IR Categorización]

■ Selección (*Feature Selection*)

+ Eficacia



- Filtrado: Medidas de Relevancia
- *Wrappers*
 - Genéticos
 - Híbridos
 - ...
- Set Cover
- Redundancia / Relevancia
- ...

- Eficacia

- Paralelismo



+ Paralelismo

[IR No Categorizados]

■ Objetivo

- Recuperar documentos relevantes

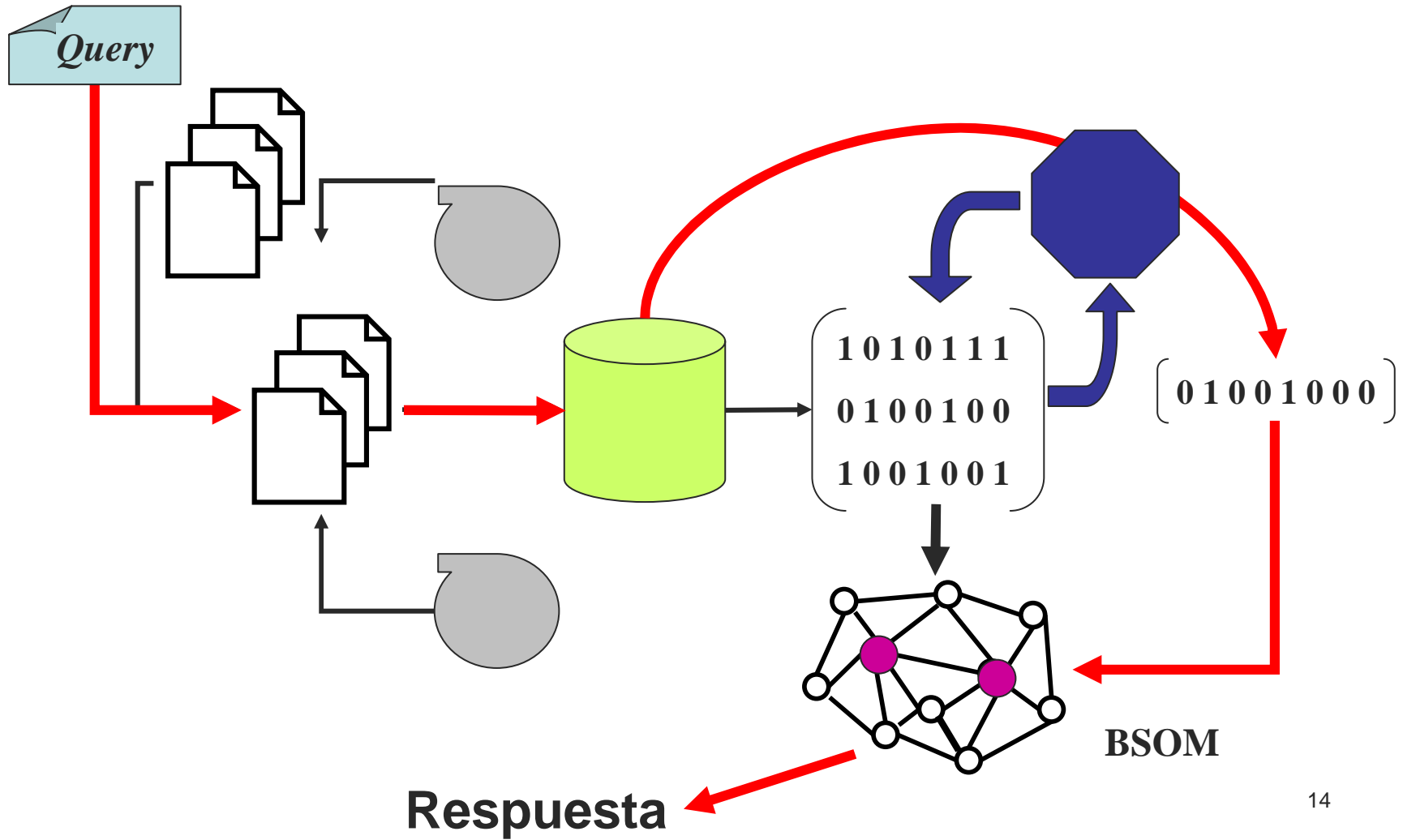
■ Problemas

- Elevada dimensionalidad
- Muchos métodos de filtrado no son aplicables

■ Métodos

- Reducción de dimensionalidad (selección/**extracción**)
- Distancia coseno (producto escalar)
- *Clustering (K-means, BSOM, GNG, GCS,...)*

[IR No Categorizados: PirSOM]



[IR No Categorizados: *PirSOM*]

■ Reducción / Proyección

- *Latent Semantic Indexing (LSI)*
 - Rendimiento y reducciones elevadas
 - Coste computacional alto
- Proyección Aleatoria (*PA*)
 - Bajo rendimiento
 - Coste computacional menor

■ Combinados

- *MR + LSI*
- ...

[Presente y Futuro del IIG]

■ Selección de atributos

- Problemas similares a categorización de textos
 - *Microarrays, spam,...*
 - *Clasificación jerárquica*
- Combinación de métodos (comités)

■ Extracción de atributos

- Métodos matriciales y otros
 - *NMF, QR, PCA, CCA, Isomap, ...*
- Problemas afines
 - *Sistemas de recomendación, ...*

[*Cluster de Mieres*]

- **Características**

- Heterogéneo
- Factor de crecimiento a corto plazo elevado
- Posibilidad de integración en alguna **GRID**

- **Ha supuesto**

- Mucho esfuerzo y algo de conocimiento

- **Idea**

- *¿Transferencia Tecnológica?*