

# Recuperación de Información en Internet

---

Localización y acceso a la información  
Instituto de Ciencias de la Educación  
Jesús Tramullas  
Dpto. CC. Documentación  
<http://tramullas.com/lai>

# RI tradicional / RI en Internet

---

- Necesidad de diferentes enfoques.
- Diferencia entre contextos abiertos y cerrados.
- Homogeneidad frente a heterogeneidad.
- Diferentes principios y niveles de descripción de contenidos.
- Tendencia al “do it yourself”.
- Diferentes generaciones de herramientas.

# Niveles de descripción

---

## ■ RI tradicional:

- Documentos similares.
- Control terminológico
- Interfaz homogéneo de interrogación
- Lenguajes documentales
- Principio de autoridad

## ■ RI en Internet:

- Documentos diferentes
- Carencia de control terminológico.
- Interfaces diversos
- Intentos de descripción de contenidos
- ¿ Autoridad ?

# Metadatos y Dublin Core

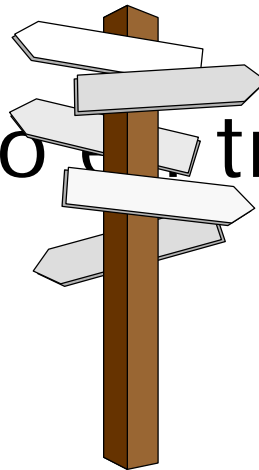
---

- Etiquetas en HTML para descripción del contenido del documento: `<META>`
- Normas para descripción del contenido: Dublin Core.
- La descripción se encomienda al responsable del documento.
- Diversos “usos” de la etiqueta `<META>`

# Enfoques de RI en Internet

---

- Enfoque basado en motores de búsqueda.
- Enfoque basado en agentes personales.
- Enfoque basado en trabajo individual.
- Enfoque basado en trabajo cooperativo.



# Motores de búsqueda, 1

---

- Bases de datos centralizadas.
- Creación de la base de datos: uso de indizadores o spiders independientes.
- Diferentes niveles de cobertura del contenido del WWW.
- Diferentes niveles y/o procesos de indización.
- Diferentes algoritmos de recuperación.
- Respuestas en formato HTML..



# Motores de búsqueda, 2

---

- Diferentes interfaces de usuario.
- Diferentes interfaces de interrogación.
- Problema creado por la ponderación de términos y la pertinencia.
- Indizadores:
  - Estándares para exclusión de robots.
  - Niveles de profundidad en la indexación.
  - Técnicas de “engaño” para spiders.
  - ¿Ética para spiders?

# Intermediarios de recuperación

---

- Actúan como interfaz único a múltiples motores.
- Adecúan la expresión a los diferentes motores.
- Remiten la consulta a los motores.
- Reciben la respuesta y la remiten al usuario, directamente, o tras realizar un primer filtrado de resultados.

# Agentes personales, 1

---

- Aplican los principios de los agentes de software.
- Se sitúan en el lado del usuario.
- Se les delegan las tareas.
- Actúan en segundo plano.
- Fase 1: consultan
- Fase 2: filtran
- Fase 3: recuperan
- Fase 4: organizan
- Fase 5: ¿copian el documento original?
- Fase 6: ¿mantienen actualizada la información?

# Agentes personales, 2

---

- Tipos de agentes:
  - De correo/news.
  - Para web.
  - De filtrado.
  - De recuperación.
  - De mantenimiento.
- Un concepto clave es la idea de **“cooperación”**.

