

Propuesta: WIRE – Web Information REtrieval.

Carlos Castillo

7 de mayo de 2003

Resumen

La siguiente propuesta considera el desarrollo de una *suite* de software para recuperación de información en la Web.

1 Introducción

El Centro de Investigación de la Web (www.ciw.cl) busca impulsar, entre otras, el área de recuperación de información en la Web. Muchos alumnos del DCC realizan durante sus memorias o tesis, software o componentes de software para recuperación de información.

El proyecto **WIRE** (Web Information REtrieval) busca potenciar estos desarrollos, integrándolos en una arquitectura común.

2 Objetivos

Coordinar la construcción de un conjunto de programas para recuperación de información en la Web, para realizar experimentación.

Esta coordinación, permitirá el desarrollo rápido de prototipos, al contar con un entorno de aplicación en que varias tareas comunes y estructuras de datos ya están implementadas.

Este conjunto de programas estarán disponibles bajo licencia GPL.

3 Componentes

El software incluye componentes de almacenamiento e indexación y componentes de procesamiento.

Las componentes de almacenamiento e indexación incluyen estructuras de datos como índices de textos, índices de enlaces, índices de estructuras XML.

Las componentes de procesamiento incluyen recolección, conversión o adquisición de datos y consultas o ranking.

4 Requisitos

El proyecto requiere ciertos requisitos tanto de sus participantes como del software que producen.

En el caso de los participantes, se buscan alumnos de ciclo terminal, interesados en el área, con buenas calificaciones y que integren habilidades de diseño e implementación.

En el caso del software, se busca que este sea adecuado para la tarea encomendada, en este caso, para experimentación. Son requisitos fundamentales la velocidad y eficiencia y la robustez.

Es indispensable que cada módulo se integre bien con el resto de las componentes de WIRE; y que no contenga software cuyo licenciamiento sea incompatible con GPL.