



Sistemas Distribuidos I (75.74)

Practica de Arquitecturas Paralelas *Multicomputing*

Web Crawler. Shopping Cart.

Docentes

- Pablo D. Roca
- Ezequiel Torres Feyuk

Agenda



- *Web Crawler*
- *Shopping Cart*



Web Crawler | Requerimientos Funcionales

- Dada una URL, obtener el HTML y analizarlo.
- Detectar nuevas URLs e identificar archivos multimedia, documentos y otras páginas Web.
- Descargar los archivos multimedia (.jpg, .png, .mp3, .ogg, etc)
- Descargar los documentos (.pdf, .odt, etc)
- Agendar el procesamiento de las nuevas páginas Web encontradas
- Detener el análisis de nuevos links luego de N anidaciones.



Web Crawler | Requerimientos No Funcionales

- Las descargas y el análisis se deben ejecutar en paralelo.
- Realizar un monitoreo constante en un archivo de texto con cantidades de:
 - URLs analizadas
 - Archivos descargados por extensión
 - Hilos en cada uno de los siguientes estados:
 - Obteniendo HTML
 - Procesando HTML
 - Descargando recurso

Agenda



- *Web Crawler*
- *Shopping Cart*



Shopping Cart | Requerimientos Funcionales

- Realizar un sistema que acepte pedidos de compra por productos.
- Los pedidos cuentan con usuario, tipo de producto y cantidad.
- El sistema debe mantener el stock por tipo de producto
- Además de los pedidos de compras el sistema debe aceptar:
 - Modificaciones de stock (administradores)
 - Consultas del estado de los pedidos (compradores)
 - Notificación de la entrega al comprador (empleados)
- Se debe controlar el estado de cada pedido pudiendo ser:
 - Recibido, Aceptado, Rechazado, Entregado



Shopping Cart | Requerimientos No Funcionales

- El catálogo de productos es reducido (10 aprox.) pero se espera un gran volumen de pedidos.
- Todo pedido de compra debe ser almacenado como una entrada de Log ordenada con fines de auditoría.
- Los pedidos pueden contener unidades de distintos tipos. Los pedidos son aceptados o rechazados en su totalidad.
- La aplicación debe estar preparada para una gran cantidad de usuarios concurrentes.
- El monitoreo de los flujos de información es clave para garantizar la performance del sistema.