

Quién

Irene prado

Qué

Dando cobertura a Google Search Appliance mediante productos Open: Apache Nutch y SOLR

Por qué

Tras el corte de servicio de Google de su search appliance, se requería unificar varias páginas webs de distintos dominios (intranet, extranet, subdominios, ...) , tecnologías (HTML, JSP, ASP), naturalezas (con trabajo de SEO, sin SEO, con iframes, flash...) y tipos de documentos (html, doc, pdf, ...) en un mismo buscador con capacidad avanzadas, para el usuario del portal corporativo.

“Una solución para gobernarlos a todos. Una solución para encontrarlos, una solución para atraerlos a todos y atarlos en las tiniebla... digo pantalla de búsqueda avanzada.”

Cuándo

Apache Nutch es una solución de crawling construida sobre Java que nos puede servir para extraer información “pseudo-homogénea” de diferentes orígenes / semillas, y mediante un esquema de conversión, converger su información en un sistema de índices, como puede ser SOLR

Links

<http://nutch.apache.org/> <http://lucene.apache.org/solr/> <https://www.zylk.net/en/web-2-0/blog/-/blogs/como-configurar-nutch-1-13-para-que-use-solrcloud-6-6-0>