

# Using linked data to classify web documents

Dominic Fripp

*Aslib proceedings*, Vol. 62, n. 6, 2010

El artículo establece una relación entre el esquema tradicional de clasificación facetada y la Web semántica en el entorno de los datos enlazados (linked data). Las ideas más recientes para la clasificación de documentos usan un método semántico que pretende hacer recuperable la información partiendo de un enfoque automatizado. En el origen de estos proyectos están las ideas con las que Ranganathan creó su esquema de clasificación, centrado en los conceptos y en la construcción de relaciones entre ellos, que tiene un claro paralelismo en el ámbito de la construcción de ontologías y en la formación de áreas temáticas. En la terminología Ranganathan los conceptos se pueden leer como facetas, inspirándose en el juego del mecano argumentó que cualquier materia sin importar su complejidad se podía construir a partir del mismo conjunto de componentes básicos. Como este esquema de clasificación, examinando los documentos en términos de sus componentes temáticos, están los datos enlazados. La mayoría de la información en la Web está diseñada para el consumo humano, la metodología de la Web semántica desarrolla lenguajes para expresar la información de una forma procesable por una máquina. La adición de metadatos al documento hace que el ordenador pueda comprenderlo. Esa comprensión de la máquina viene determinada por el uso de conceptos claramente definidos en lenguaje máquina. En el mundo de los datos enlazados, estos conceptos son alimentados por RDF triples y URIs. En ese sentido, Calais es un proyecto de Thomson Reuters diseñado para llevar a cabo una lectura mecánica de la Web. Su proyecto Semantic Proxy tiene como objetivo traducir el contenido o cualquier URL en la Web en su representación semántica en RDF, HTML o Microformatos. Para la visualización de los metadatos Christian Hirsch desarrolló en 2008 Thinkpedia, que usa la información semántica producida por Semantic Proxy y se alimenta a través del software thinkmap para crear un gráfico visual de los metadatos del documento.

Resumen elaborado por: Natividad Escavias