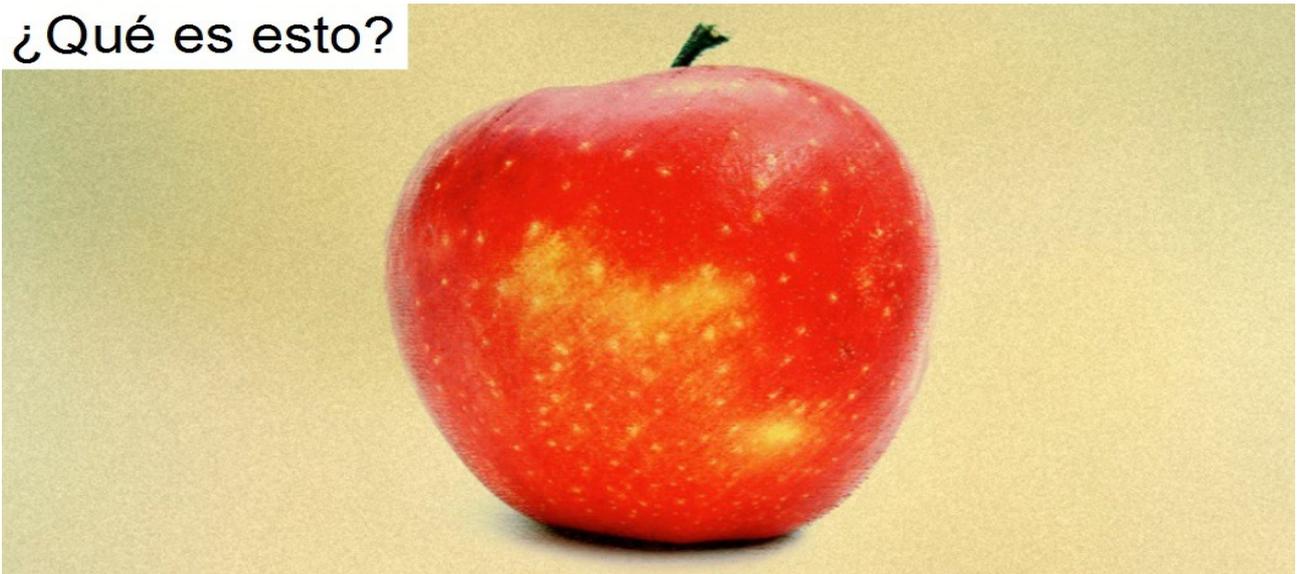


¿Cómo nombrar los datos en la web?

Para construir la Web de Datos es necesario poder nombrar los datos del grafo de conocimientos de forma única para poder identificarlos. Si queremos poder integrar datos que provienen de grafos distintos es necesario que la identificación sea global, utilizando por ejemplo URLs. Pero es necesario distinguir entre una página web que describe un recurso en la web y una URL que es utilizada para identificar al recurso.

Si preguntamos:

¿Qué es esto?



y observamos que una respuesta válida es:



Lo que esperaríamos como respuesta correcta sería "Esto es la foto de una manzana" pero no tendríamos problema en aceptar que para ciertos fines nos resulta válida la respuesta: "Esto es una manzana". No podremos comerla pero podremos describir su color, su forma y razonar con esta

abstracción. De todas formas es importante representarla con un modelo donde ambos conceptos: la representación que asumimos de la manzana y la foto de la manzana no se confundan. Se debe distinguir que la foto de la manzana está siendo usada como sustituto del objeto del mundo real (ver Figura 5). De esto se trata la utilización de URLs y de IRIs para identificar las representaciones de los recursos del mundo real en la web.

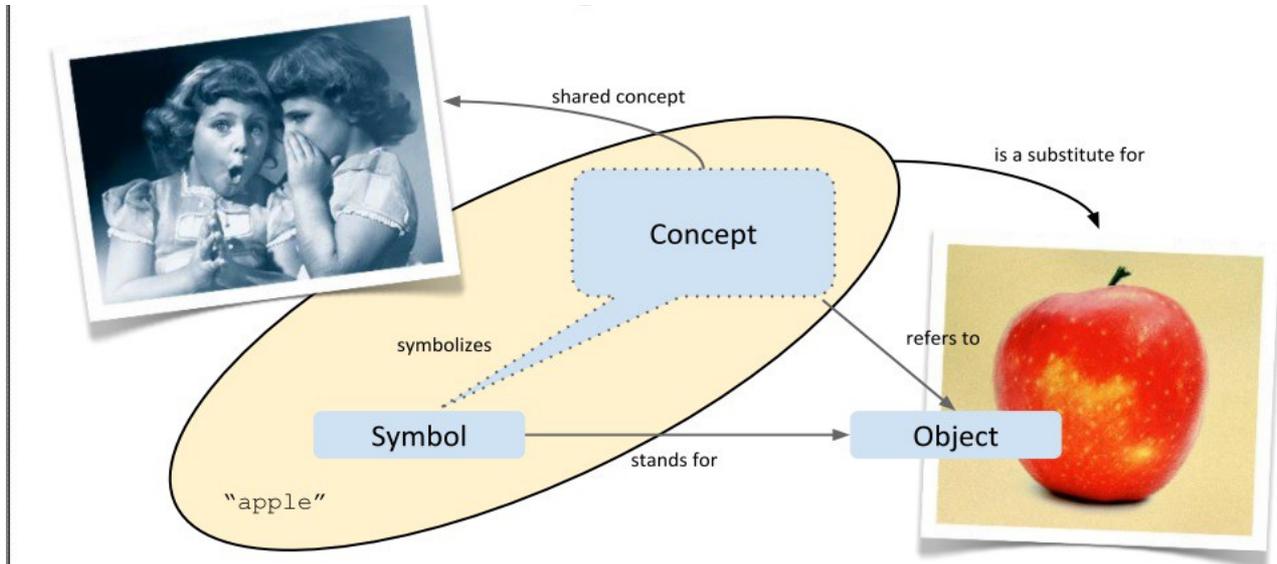


Figura 5. Uso del triángulo semiótico para nombrar objetos en la web. (Tomado de presentación de Dr. Harald Sack – FIZ Karlsruhe.)

En conclusión podemos sustituir al objeto del mundo real por una representación, como en el ejemplo anterior de la manzana por una fotografía de la manzana. Si llevamos esto al contexto de la web, cambiamos la fotografía de la manzana por un sitio web donde describimos lo que es una manzana. De esta forma ante cualquier duda de lo que es una manzana podemos referirnos a este sitio web que describe a una manzana. Pero para facilitar la integración de las representaciones del concepto manzana que tenemos en la web deberíamos poder nombrarlos de una forma uniforme, aparecen así en acción las [URI \(Uniform Resource Identifier\)](#).

En varios dominios encontramos Identificadores Uniformes de Recursos (URI), por ejemplo para los libros se usa el ISBN, el DOI para los artículos científicos, para los autores de artículos científicos se usa el ORCID y Alpha-2 para los países, entre muchos otros.

Los [IRI \(Identificadores de Recursos Internacionalizados\)](#) son una generalización de los URIs. Mientras que URI sólo admite codificación ASCII, IRI es totalmente compatible con caracteres internacionales. Por ejemplo, <http://example.com/Ñam> es una IRI, pero no una URI, debido al uso de "Ñ". Utilizando el símbolo de porcentaje podemos escribir de forma codificada la Ñ de la siguiente manera: <http://example.com/%C3%91am> – transformando así una IRI en una URI (pero reduce la legibilidad). Apesar que las IRI son más legibles para las personas, también pueden generar problemas por confusiones al usar caracteres similares por ejemplo usar ã por a.

RDF utiliza IRIs para nombrar a sus recursos y a sus propiedades

En la representación RDF de la Wikidata por ejemplo, <https://www.wikidata.org/wiki/Q1335> es la IRI que refiere a una página web que se puede cargar en un navegador que proporciona metadatos legibles por humanos sobre Montevideo capital del Uruguay, mientras que la IRI <https://www.wikidata.org/entity/Q1335> se refiere a la ciudad misma.

Al usar IRI con HTTP para identificar las entidades del grafo, cuando se busca el IRI (a través de HTTP), el servidor web puede devolver (o redirigir a) una descripción de esa entidad en formatos como RDF. Esta permite además que los grafos RDF se vinculen con entidades relacionadas descritas en grafos RDF externos a través de la Web, dando lugar a datos vinculados o también llamados enlazados (Linked Data). Aunque los IRI de HTTP ofrecen un mecanismo flexible y poderoso para emitir identificadores globales en la Web, no son necesariamente persistentes: los sitios web pueden salir de línea, los recursos descritos en una ubicación determinada pueden cambiar, etc. Para mejorar la persistencia de dichos identificadores, los servicios de URL persistente (PURL) ofrecen redireccionamientos desde un servidor central a una ubicación particular, donde el PURL se puede redirigir a una nueva ubicación si fuera necesario, cambiando la dirección de un documento sin cambiar su identificador. La persistencia de los IRI de HTTP se mejoran mediante el uso de espacios de nombres definidos a través de los servicios PURL.
