Nutrición Hospitalaria

Nutr Hosp. 2012;27(2):323-332
ISSN 0212-1611 • CODEN NUHOEQ
S.V.R. 318

Artículo especial

Linked Data como herramienta en el ámbito de la nutrición

R. Míguez Pérez, J. M. Santos Gago, V. M. Alonso Rorís, L. M. Álvarez Sabucedo y F. A. Mikic Fonte

Departamento de Ingeniería Telemática. Universidad de Vigo. Vigo. España.

Resumen

En la actualidad existe una enorme cantidad de información en Internet que no puede ser interpretada y reutilizada por agentes software. Este hecho impone una importante limitación al potencial de las herramientas que es posible construir haciendo uso de los datos disponibles en la Web actual. Sin embargo, los recientes avances introducidos en el ámbito de la Web Semántica nos permiten crear una nueva generación de aplicaciones inteligentes capaces de ofrecer valor añadido al usuario. Este trabajo presenta los desafíos tecnológicos que es preciso abordar para, dentro del ámbito de la nutrición, partir de una o varias fuentes de datos convencionales y generar un repositorio web, basado en tecnologías semánticas, ligado con otras fuentes de datos públicas en Internet. Este enfoque permite que herramientas automáticas operen sobre esta información y presten nuevas funcionalidades altamente interesantes en el ámbito de la salud pública, como la generación automática de menús para escolares o asesores dietéticos inteligentes, entre otras. El presente artículo detalla el proceso para generar dicho soporte de información siguiendo las directrices de la iniciativa Linked Data e ilustra el uso de herramientas que sacan partido de este soporte para su adopción en otros casos de uso y entornos próximos.

(*Nutr Hosp.* 2012;27:323-332)

DOI:10.3305/nh.2012.27.2.5594

Palabras clave: *Nutrientes. Servicios dietéticos. World wide web. Semántica. Proceso automático de datos.*

LINKED DATA AS A TOOL IN THE NUTRITION DOMAIN

Abstract

Currently, there is a huge amount of information available on Internet that can neither be interpreted nor used by software agents. This fact poses a serious drawback to the potential of tools that deal with data on the current Web. Nevertheless, in recent times, advances in the domain of Semantic Web make possible the development of a new generation of smart applications capable of creating added-value services for the final user. This work shows the technical challenges that must be faced in the area of nutrition in order to transform one or several old-fashion sources of raw data into a web repository based on semantic technologies and linked with external and publicly available data on Internet. This approach makes possible for automatic tools to operate on the top of this information providing new functionalities highly interesting in the domain of public health, such as the automatic generation of menus for children or intelligent dietetic assistants, among others. This article explains the process to create such information support applying the guidelines of the Linked Data initiative and provides insights into the use of tools to make the most of this technology for its adoption in related use cases and environments.

(*Nutr Hosp.* 2012;27:323-332)

DOI:10.3305/nh.2012.27.2.5594

Key words: *Nutrients. Dietary services. World wide web. Semantics. Automatic data processing.*

Introducción

La revolución provocada por el fenómeno de la Web 2.0 ha hecho que ahora cualquier persona, independientemente de sus conocimientos técnicos, pueda publicar información en la Web. Como consecuencia, nos encontramos ante un escenario caracterizado por la masificación de contenidos en la Red, precisando el ser humano de algún tipo de intermediario "inteligente" capaz de extraer, procesar y localizar de forma autónoma la información requerida. La nueva "Web de los Datos", también conocida como "Web 3.0" o "Web Semántica", sienta los cimientos de este futuro inmediato, creando una red de nodos con información multidisciplinar que puede ser explorada por aplicaciones software sin necesidad de la intervención humana.

La iniciativa Linked Open Data (LOD)[1], impulsada por Tim Berners Lee (creador de la Web), define los mecanismos que dan forma a esta nueva Internet, en la que los datos ya no están cautivos en silos propietarios, sino que pueden ser libremente compartidos y reutilizados por agentes software. Un nodo de la red Linked Data se caracteriza porque, además de mantener los

Correspondencia: Luis Álvarez Sabucedo.
Departamento de Ingeniería Telemática.
Universidad de Vigo.
Vigo. España.
E-mail: lsabucedo@det.uvigo.es

Recibido: 7-XI-2011.
Aceptado: 22-XI-2011.

datos propios de su dominio de interés, define enlaces a otros con información relacionada o complementaria. Así, un nodo dedicado a la publicación de registros sobre medicamentos como DrugBank[2] almacena tanto datos propios (nombre comercial, fórmula química, modo de administración, etc) como referencias a fuentes externas, incluyendo posibles enfermedades objetivo (descritas en el nodo Diseasome[3]) o su utilización en pruebas clínicas (publicadas por LinkedCT[4]). La filosofía Linked Data permite que los nodos se beneficien del "efecto red", de forma que la adición de uno nuevo aporta valor al resto de integrantes de la red LOD. El área de las ciencias de la salud en general, y el campo de la nutrición en particular, son ámbitos en los que el potencial brindado por la Web Semántica no ha sido aun debidamente explotado. El nuevo enfoque propuesto por la iniciativa LOD pone a disposición del público general una extensa cantidad de información que agentes inteligentes, soportados por tecnologías semánticas, pueden reutilizar y aprovechar para crear soluciones de alto valor añadido para el usuario.

Este trabajo presenta los principales desafíos y guías de diseño que los desarrolladores deben afrontar a la hora de crear, publicar e incorporar un nodo con información nutricional a la red Linked Data. El apartado 2 resume brevemente la iniciativa LOD y las tecnologías que le dan soporte. En el apartado 3 se describen detalladamente los pasos necesarios para la construcción y publicación de uno de estos nodos, utilizándose como ejemplo práctico la información proporcionada por la "USDA Database for Nutrition Information"[5]. El apartado 4 muestra cómo es posible poner en valor la información publicada mediante el desarrollo de un asesor nutricional para el ámbito de la educación infantil. Finalmente, el apartado 5 presenta las principales conclusiones de este trabajo.

La iniciativa Linked Open Data

El modo en el que la información se publica en Internet ha sufrido una profunda evolución durante los últimos años. Desde las primeras publicaciones de datos en la Web a principios de los 90, basadas en textos estáticos, hasta los actuales modelos de publicación de la información basados en blogs, redes sociales y comunicación viral, se puede afirmar que han cambiado no solo las tecnologías sino también los paradigmas subyacentes. La Web actual se basa en un diseño sencillo, accesible e intuitivo para el ser humano. Este interpreta la información presentada en pantalla y accede a nuevos datos mediante una serie de hiperenlaces incluidos en el documento que está consultando. Resulta paradójico que la principal razón del éxito de la Web se haya convertido en uno de los principales desafíos a resolver a la hora de evolucionar cara a un nuevo modelo de acceso a la información en el que el ser humano, abrumado por la ingente cantidad de datos disponibles en la Red, precisa de un intermediario (un agente software) que se encargue de explorar la Web, descubrir y procesar los datos buscados y presentarlos finalmente en un entorno amigable. La "Web de los Documentos", accesible y entendible únicamente por el ser humano, se convierte así en la "Web de los Datos", accesible y entendible también por las máquinas.

Para hacer realidad la "Web de los Datos" es preciso definir un mecanismo que permita a los agentes software "interpretar" (manipular simbólicamente) la información disponible en Internet. La primera y más básica especificación en este sentido definida por el W3C es el modelo RDF[6]. Esta especificación define cómo representar sentencias o declaraciones sobre recursos "referenciables" en la web (prácticamente cualquier cosa, ya sea física o abstracta puede ser referenciada mediante URI). Una declaración RDF toma la forma de una tripla compuesta de un sujeto, un objeto, y un predicado que determina la relación que une sujeto y objeto.

Mientras que RDF garantiza la interoperabilidad sintáctica de los datos, queda por resolver el problema de la interoperabilidad semántica de los mismos. Para ello es preciso establecer un consenso sobre el significado concreto de los términos (nombre de conceptos y relaciones) que existen en un dominio particular. La Web Semántica dispone de un instrumento específico para realizar esta labor, la ontología, entendiendo como tal una "especificación explícita de una conceptualización"[7], que puede ser descrita formalmente mediante la especificación RDFS[8] o bien, si la potencia semántica de esta no es suficiente, mediante OWL[9], ambas especificaciones basadas en RDF definidas por el W3C.

Desde un punto de vista tecnológico existen básicamente dos estrategias para la publicación de la información en esta nueva Web. La primera, más continuista, pasa por enriquecer las páginas web existentes, expresadas en HTML, con anotaciones RDF (utilizando las directrices establecidas en la especificación RDFa[10]), que aportan contexto e información procesable por un ente automatizado al documento. Esta aproximación, aunque sencilla en su concepción, incrementa la complejidad en la creación y mantenimiento de las páginas web, por lo que en los últimos años ha ganado fuerza una iniciativa alternativa: Linked Opend Data. Se basa, a grandes rasgos, en la creación de nodos web con información expresada directamente en RDF, ligados entre sí, capaces de ofrecer una representación distinta de los contenidos según el tipo de usuario que la solicita. Cada objeto dentro de un nodo Linked Data cuenta con un nombre único, su URI, que nos permite referenciarlo de forma unívoca. Cuando una petición de información es recibida, el nodo analiza la fuente de la solicitud y, en función de ella, devuelve una versión del documento adaptada al tipo de usuario origen: (1) una página HTML, en caso de un ser humano accediendo mediante un navegador web; y (2) un documento RDF, en caso de un agente software.

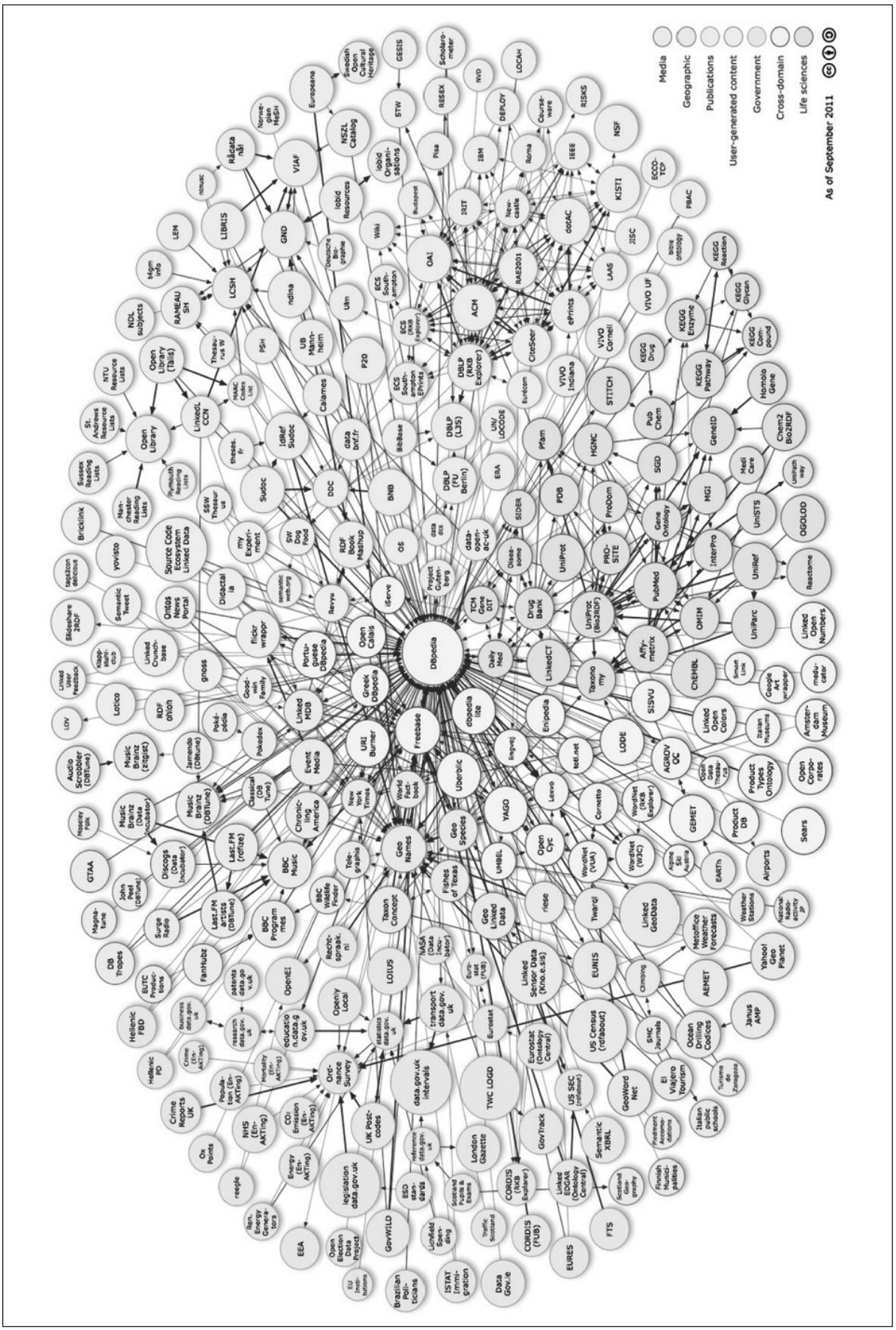

Fig. 1.—Nodos incluidos en septiembre de 2011 dentro de la iniciativa Linking Open Data.

La iniciativa Linked Data basa su funcionamiento en tecnologías y estándares ampliamente aceptados, cimentándose en 4 principios básicos:

1. Uso de URI para referenciar todo objeto de información.
2. Utilización del protocolo HTTP[11] para acceder a la información almacenada en las URI.
3. Descripción de los recursos de información mediante RDF y utilización del lenguaje de consultas SPARQL[12] para la búsqueda sobre estos repositorios.
4. Incluir enlaces a otras entidades mediante URI para potenciar el descubrimiento de nuevos elementos de información que puedan ser relevantes para el usuario.

Como resultado de la puesta en práctica de estos principios, es posible implementar aplicaciones genéricas capaces de operar sobre un espacio de datos universal, en otras palabras, tratar la Web como una única base de datos global. El auténtico potencial de este modelo reside en poder combinar datos procedentes de diferentes fuentes (otros nodos Linked Data) y, a partir de su integración, extraer nuevo conocimiento que nos permite resolver el problema particular al que nos enfrentamos. Conscientes de las enormes posibilidades de este enfoque, cada vez son más las empresas, instituciones y gobiernos que contribuyen en este esfuerzo haciendo públicos libremente sus datos en la Web, existiendo en la actualidad nodos con información de tipo enciclopédico, información biblioteconómica o del ámbito de las ciencias de la salud entre otros. La figura 1 muestra el estado actual de la red LOD, que en Septiembre de 2011[13] contaba ya con 295 nodos enlazados entre sí y más de 31 billones de triplas RDF. Un nodo central de esta red es la DBPedia[14], repositorio que mantiene la información generada de forma colaborativa por la comunidad a través del proyecto Wikipedia. Además de nodos con información multidisciplinar como la DBpedia, la LOD está poblada por otros repositorios más específicos. En el ámbito de las ciencias de la vida podemos destacar los nodos DrugBank, DailyMed o PubMed, y si bien es cierto que en la actualidad el porcentaje de información publicada en este campo no es muy elevado (9,6% del total), debemos reseñar el alto grado de interconexión que presentan con otras fuentes (38,06% del total de enlaces de la Red). A mayor grado de interconexión entre datos, mayor es la información potencial que un motor de inferencia puede extraer de los mismos, y por tanto, mayores las ventajas que la red LOD puede ofrecer en los ámbitos de la salud y la medicina. El proyecto del W3C: "Linking Open Drug Data"[15], cuyo objetivo es crear nodos ligados con información sobre medicamentos y pruebas clínicas, es un ejemplo claro del potencial de esta nueva Web en el que están involucradas las compañías farmacéuticas Eli Lilly, AstraZeneca y Jhonson & Johnson.

Creación y publicación de un nodo Linked Data

En este apartado se describe el proceso de creación de un nodo Linked Data orientado a la publicación de información nutricional ("interpretable" por agentes software) de productos alimenticios. En nuestro caso particular y como prueba de concepto, se ha decidido utilizar como fuente de información la USDA Database for Nutrition Information (USDA de ahora en adelante), reputada base de datos en este ámbito. Este proceso conlleva la consecución de una serie de pasos, los cuales se exponen brevemente a continuación:

Paso 1: Definir la terminología (ontología)

El primer paso consiste en definir la ontología que identifica los términos (en particular los nombres de los conceptos y de las relaciones) que se van a utilizar para describir la información a publicar. En la actualidad existen multitud de bases de datos y otro tipo de registros electrónicos que almacenan información nutricional de alimentos. Por tanto, para el desarrollo de la ontología es conveniente realizar un estudio previo de los esquemas utilizados en estos registros, basándonos en nuestro caso en la USDA. Por otro lado, las guías de buenas prácticas para la publicación de información en la Web de los Datos recomiendan que los términos utilizados para describir la información estén basados en vocabularios ya existentes en el ámbito de la Web Semántica y de la red Linked Data en particular, en especial aquellos de mayor difusión, con el fin de mejorar la interoperabilidad y la compatibilidad semántica de la información. Por este motivo, para la definición de nuestra ontología, hemos utilizado términos extraídos de los vocabularios Dublin Core[16] (esquema de metadatos de muy amplia difusión que define términos básicos para la descripción de recursos genéricos) y SKOS[17] (conjunto de términos que pueden ser empleados para la representación de tesauros, taxonomías y otros modelos de clasificación).

La figura 2 refleja los principales conceptos y relaciones identificados. Como se puede observar, las entidades fundamentales identificadas son *Food* (Alimento), *Nutrient* (Nutriente) y *NutrientAmount* (CantidadNutriente). Este último concepto representa una relación ternaria entre un alimento, un nutriente y la cantidad de ese nutriente disponible en 100 g del alimento. Además de estas 3 entidades básicas, se han incluido los conceptos *FoodGroup* y *LangualFactor*, que permiten, el primero, clasificar taxonómicamente los alimentos y, el segundo, catalogar los alimentos atendiendo a los elementos del tesauro LanguaL[18]. Por último, el concepto *TypicalMeasure* permite registrar el peso, en gramos, de medidas o porciones típicas de un determinado alimento (p. ej. de "una taza" de harina o de "una cucharada" de azúcar). Esta información facilita, entre otras tareas, el cálculo de los nutrientes de un determinado menú expresado en base a sus ingredientes utilizando medidas convencionales.

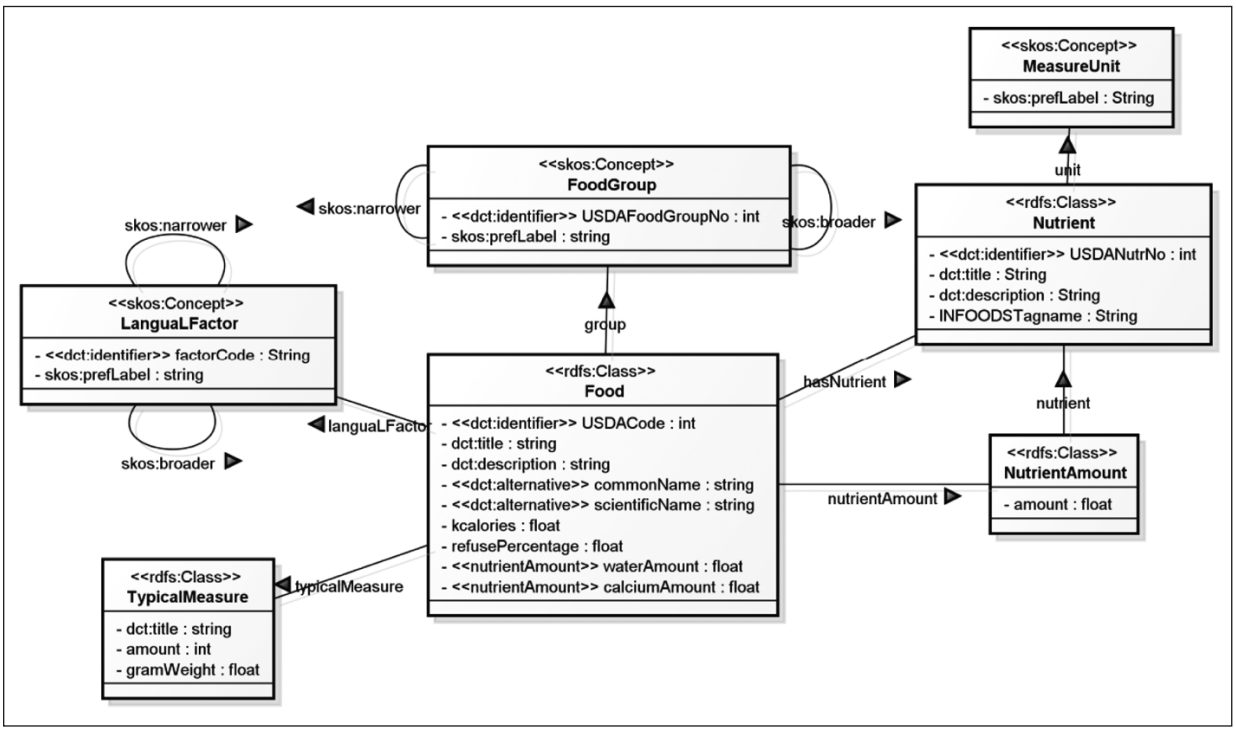

Fig. 2.—Vista parcial de la ontología para la descripción de información nutricional de alimentos.

Para cada uno de estos conceptos se han definido las propiedades o relaciones que permiten caracterizarlos (fig. 2). Así, por ejemplo, para el concepto *Food* se han definido, entre otras, propiedades como *dct:title* (nombre del alimento en diferentes idiomas), *dct:description* (descripción textual), *scientificName* y *commonName* (haciendo referencia, respectivamente, a su nombre científico y otros nombres alternativos de uso común), *kcalories* (valor energético de 100 g del alimento), etc. Cabe destacar que, para el concepto *Food*, se han provisto propiedades (como *calciumAmount*), que permiten registrar directamente la cantidad de un determinado nutriente (calcio en este ejemplo) disponible en 100 g de un determinado alimento. Si bien esto resulta redundante, puesto que la entidad *NutrientAmount* nos permite registrar esta información, se han mantenido ambas representaciones para facilitar el acceso a esta información a aquellas aplicaciones que no requieren de la potencia semántica inherente en la utilización del concepto *NutrientAmount* (más complejo de manipular). La definición de reglas lógicas, expresadas en SWRL[19], permiten que sea el propio sistema (a través de un motor de inferencia) el que realice la transformación de forma automática de una representación a la otra. Por ejemplo, la siguiente regla:

$$Food(?f) \wedge nutrientAmount(?f,?na) \wedge NutrientAmount(?na) \wedge nutrient(?na,usda:Calcium) \wedge amount(?na,?a) \rightarrow calciumAmount(?f,?a)$$

permite "calcular" la relación *calciumAmount* en base a la información recogida utilizando el concepto *NutrientAmount*.

Paso 2: Población

El segundo paso consiste en recopilar la información que se desea exponer a través del nodo Linked Data y registrarla en formato RDF haciendo uso de los términos identificados en la ontología. Para ello se ha creado un *script* semiautomatizado encargado de: (1) extraer la información de interés del catálogo de la USDA; (2) identificar potenciales conflictos; (3) expresar dicha información en forma de triplas RDF; y (4) almacenar los datos en un almacén RDF (en nuestro caso se ha utilizado el Virtuoso Universal Server[20]). Aunque el *script* funciona de manera autónoma, en ocasiones es precisa la colaboración de un experto para resolver posibles inconsistencias en los datos (p. ej. varias entradas en la base de datos de la USDA que hacen referencia a un mismo tipo de alimento). A través de un formulario web el experto selecciona, de entre la lista de posibles alternativas, la más adecuada.

Paso 3: Configuración del nodo

Linked Data no define un patrón concreto para asignar un nombre único (una URI) a los recursos, por lo que cada nodo puede seleccionar el esquema que considere más conveniente. En nuestro caso, el patrón seleccionado combina un prefijo común (el espacio de nombres propio del nodo), seguido de un sufijo que identifica la clase del recurso y su identificador USDA (fig. 3). En la actualidad, existen herramientas automatizadas que facilitan este proceso, habiéndose utilizado

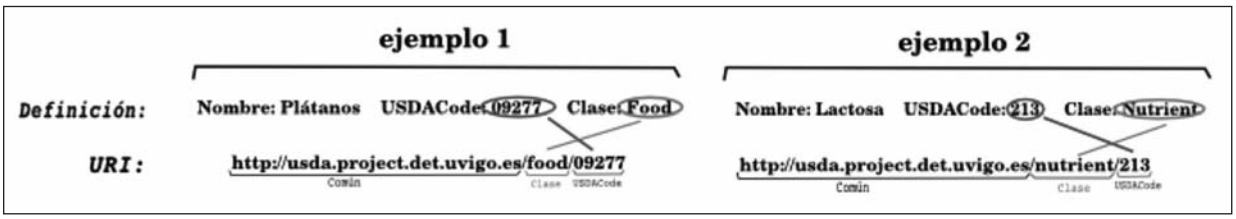

Fig. 3.—Esquema utilizado para la asignación de URI unívocas a cada uno de los registros extraídos de la base de datos de la USDA.

en este caso las herramientas de configuración incluidas en el propio Virtuoso.

Paso 4: Vinculación con otros nodos

Tras configurar nuestro nodo hemos alcanzado los tres primeros principios básicos de los cuatro establecidos en Linked Data, restando el establecimiento de relaciones con otras entidades. Uno de los procedimientos más comunes para realizar esta labor es el *record linkage*[21,22], proceso definido en la literatura especializada como la "identificación y relación de diferentes registros en fuentes de datos heterogéneas que hacen referencia al mismo objeto en el mundo real". A la hora de identificar estas relaciones de igualdad es posible utilizar diferentes técnicas de cálculo de similitud que miden el parecido entre dos registros a través de métricas de comparación sobre los valores de las propiedades que describen el mismo concepto (p. ej. nombre, descripción, etc.). A día de hoy este proceso es más un arte que una ciencia, debiendo ser el propio experto el que, en base al estudio de los modelos de datos, defina los heurísticos que permitan realizar la comparación entre la información almacenada en diversos nodos.

Basándonos en esta metodología es necesario seguir los siguientes pasos para completar todo el proceso: (1) identificar fuentes de datos externas que almacenan registros del mismo tipo que el nuestro; (2) estudiar sus modelos de datos para definir cómo abordar la comparación; (3) definir los heurísticos que nos permiten establecer el grado de similitud entre dos registros; y (4) resolver los casos conflictivos.

En el caso que nos ocupa, se seleccionaron como fuentes externas de interés dos grandes repositorios de carácter internacional: Agrovoc[23] y DBpedia. Agrovoc es un repositorio elaborado por la FAO con el mayor tesauro de agricultura del mundo, mientras que la DBpedia es un repositorio multidisciplinar compuesto por información extraída de la Wikipedia.

Tras analizar el modelo de datos de ambas fuentes y compararlo con el desarrollado, se detectó que para los registros de tipo *Food* y *Nutrient* resultaba factible encontrar similitudes con recursos en fuentes externas, aplicando mecanismos de comparación sintáctica sobre propiedades como el nombre del elemento, su descripción, grupo alimenticio y su nombre científico. Para ello, utilizamos un compendio de los heurísticos jaroSimilarity[24], basado en la métrica de distancia Jaro, Levenshtein[25], que mide el número de cambios necesarios para transformar una cadena en otra y qGramSimilarity[26], que mide el número de secuencias de letras por palabra compartidas por las cadenas. En la figura 4 se puede apreciar un ejemplo de las propiedades y registros comparados.

De entre las diferentes herramientas existentes que nos facilitan la configuración y ejecución de los heurísticos sobre los nodos, hemos utilizado el framework SILK[27]. Dado que estos procesos se ejecutan de forma automática, es posible obtener tanto falsos positivos como falsos negativos tras la ejecución de los algoritmos de comparación. Para minimizar en la medida de lo posible esta situación, se aplica como valor base para evaluar la similitud de los registros el calculado tras ponderar la salida de los heurísticos anteriores (*Sim*), definiendo dos umbrales *Uask* y *Uaut* que nos permiten definir cuándo un valor es automáticamente desechado o aceptado, o cuándo es preciso consultar con un experto:

Sim < Uask,las entidades no son iguales
Uask < Sim < Uaut, se debe preguntar a un experto
Uaut < Sim, las entidades se consideran iguales

Finalmente, los pares exitosos son relacionados mediante la propiedad *owl:sameAs* en una tripla RDF que es almacenada en nuestro nodo.

Puesta en valor del nodo: desarrollo de un asesor nutricional inteligente

La creación de la red Linked Data ha supuesto un importante impulso a la interoperabilidad y reutilización de los datos públicos disponibles en la web, dando lugar a una nueva generación de herramientas inteligentes capaces de rastrear, extraer y procesar esta información. Las aplicaciones LinkedData no se ven limitadas únicamente a su conjunto de datos local, sino que pueden recoger información de todos los nodos de la red siguiendo, de forma automatizada, los enlaces definidos en cada uno. Aquellos nodos con información multidisciplinar como Freebase[28] o la DBPedia actúan como puntos de interconexión entre diferentes dominios, constituyendo el pilar básico sobre el que se sustentan estas aplicaciones. Un ejemplo paradigmático de aplicación Linked Data es el portal BBC

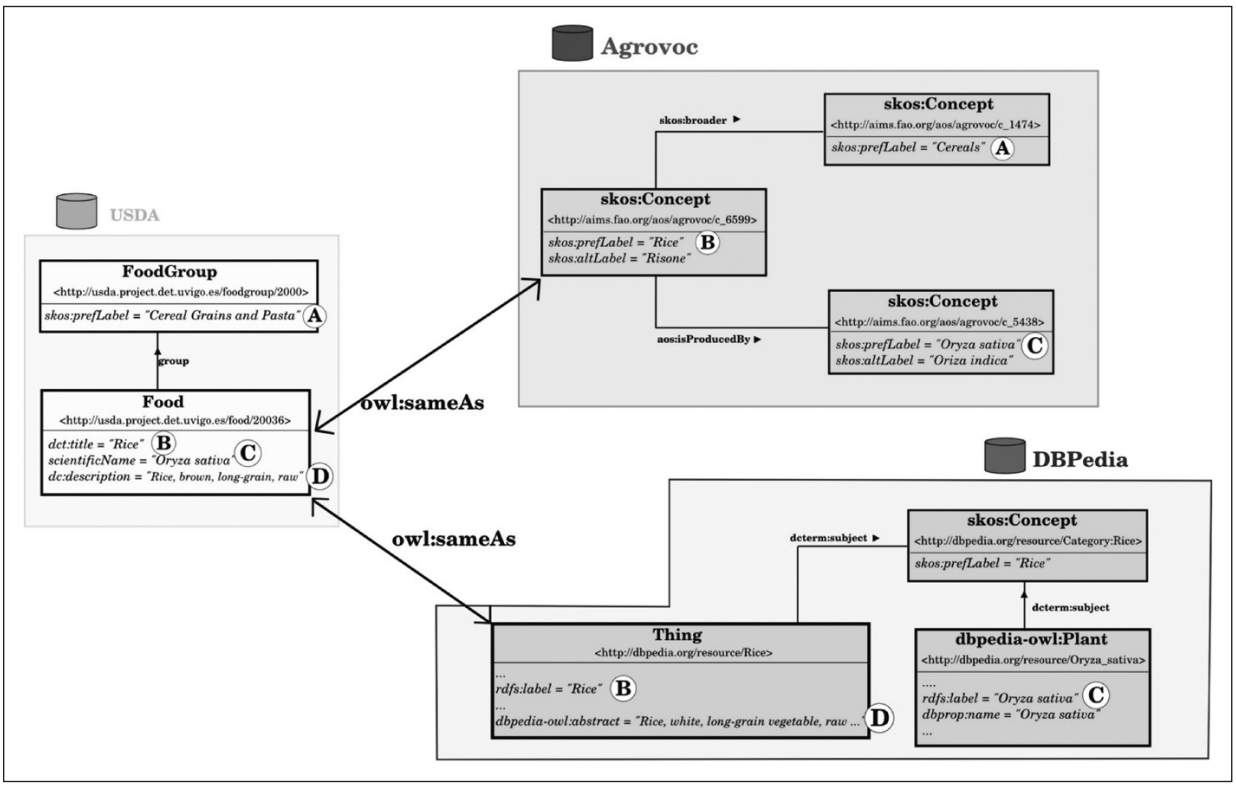

Fig. 4.—Ejemplo de las propiedades comparadas para el alimento Arroz en el proceso de encontrar registros referentes al mismo objeto entre los de nuestro repositorio (nodo USDA) y los de la DBPedia y Agrovoc.

Music[29], capaz de complementar la información ya disponible en la base de datos de la BBC con registros extraídos de otros nodos (DBpedia, Music Brainz[30], Echo Nest[31]) aportando datos adicionales como la biografía del artista, discografía, información sobre un álbum, recomendaciones de cantantes con estilos similares, etc.

En el caso que nos ocupa, se ha desarrollado una aplicación inteligente que permite poner en valor la información nutricional publicada por nuestro nodo Linked Data. La aplicación diseñada, centrada en el ámbito de la educación infantil, tiene un objetivo doble: (1) ofrecer ayuda a los centros a la hora de configurar el menú de la escuela y (2) ofrecer información y consejo a las familias para mejorar y complementar la dieta de sus hijos. En líneas generales, los servicios ofrecidos por el sistema son:

- Controlar la alimentación de aquellos niños sujetos a restricciones particulares (celíacos, diabéticos, etc.), ofreciendo menús alternativos aptos para ellos.
- Ofrecer información sobre los valores nutricionales de un menú y facilitar el seguimiento de dietas.
- Realizar recomendaciones de comidas complementarias para el hogar.

La interfaz de acceso al sistema está basada en web, habiendo sido especialmente diseñada pensando en su uso tanto desde el salón del hogar (a través de un televisor con conectividad a Internet) como a través de dispositivos móviles (smartphones, tabletPCs, etc.). Gracias a este diseño, el usuario puede acceder a la aplicación, por ejemplo, desde un iPhone, un tablet basado en Android o el navegador web incluido con la consola Nintendo Wii, obteniendo idéntica funcionalidad en cada caso. A la hora de desarrollar el asistente nutricional, se ha partido de un trabajo previo[32] en el que se describe en mayor detalle las funcionalidades clave de la plataforma base. Sobre este desarrollo se han modificado los componentes software de la plataforma, creándose una nueva infraestructura técnica capaz de soportar el descubrimiento y procesado de información no presente en la base de datos de la propia aplicación sino libremente accesible a través de la red Linked Data. Gracias a esto, la nueva aplicación es más fácil de desarrollar y mantener (pues la información es almacenada y mantenida por terceros), ofreciendo capacidades mejoradas al usuario al trabajar con un universo de datos mayor.

Desde un punto de vista técnico, la plataforma está basada en una capa de presentación de la información (responsable de adaptarla para garantizar su correcta visualización en distintos dispositivos) y una capa de negocio formada por un motor de inferencia alimentado por un explorador Linked Data.

La aplicación almacena en su base de datos local, conforme a un modelo semántico, los menús semanales de los centros. Tomando esta información como punto de partida, el explorador Linked Data integrado localiza,

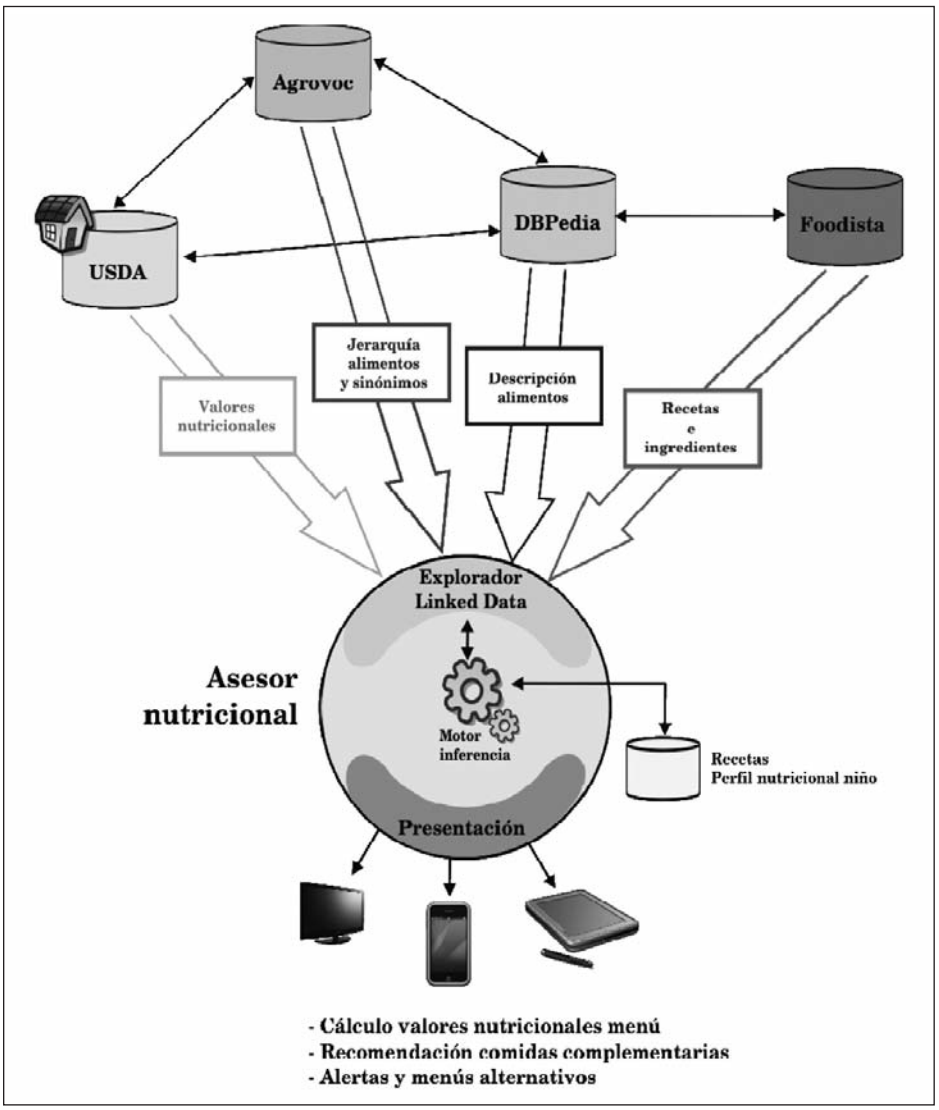

Fig. 5.—Arquitectura general del Asesor Nutricional desarrollado incluyendo las interacciones con las diferentes fuentes de información y repositorios Linked Data de los que hace uso.

mediante consultas SPARQL, las entradas correspondientes a sus ingredientes en el nodo, extrayendo información sobre su valor nutricional. De modo adicional, a través de la red de enlaces definida, el explorador identifica las equivalencias establecidas con recursos de Agrovoc (para identificar sinónimos y la posición relativa del ingrediente dentro de la jerarquía de alimentos) y la DBPedia, de donde se extrae información adicional sobre el ingrediente (descripción, preparaciones habituales, imagen identificativa, etc.) que puede mostrar al usuario en forma de ayuda contextual. En base a toda esta información, este módulo genera un nuevo modelo RDF que es utilizado como entrada de datos por parte del motor de inferencia (fig. 5).

A continuación un agente inteligente filtra y procesa los datos, generando un informe con la cantidad exacta de kilocalorías, proteínas, grasas y lípidos de cada uno de los platos y del menú diario en su conjunto, mostrándose de un modo simple y atractivo al usuario en forma de una escala de colores (fig. 6). Este tipo de operaciones se realizan sin que el usuario tenga consciencia de ellas, actuando el sistema como un asistente virtual.

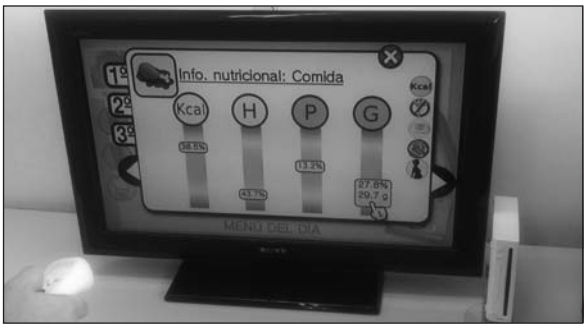

Fig. 6.—Asistente nutricional inteligente desarrollado para la gestión de los menús de las escuelas infantiles. Interfaz de usuario para la consola Nintendo Wii mostrando la información nutricional básica del menú escolar de un niño calculada de forma automática a partir de sus ingredientes.

Cuando es utilizada por una escuela, la herramienta, tras analizar el perfil nutricional de los alumnos, emite

una alerta en caso de detectar un menú que contenga un nutriente al que un niño es alérgico, ofreciendo una opción alternativa para cada caso (p. ej. sustituir la nata por nata con soja en el caso de alergia a la lactosa). De este proceso se encarga el motor de inferencia de la aplicación, cuya base de conocimiento almacena una serie de alimentos que pueden actuar como sustitutivos de otros. La definición de reglas semánticas también juega un papel vital a la hora de configurar menús complementarios para el niño. En base a la información nutricional de la dieta diaria del niño y un conjunto de reglas preconfiguradas, la aplicación es capaz de calcular el aporte nutricional extra que este necesita y, en base a ello, seleccionar una receta adecuada de entre las disponibles en su base de conocimiento u otros nodos Linked Data con información sobre recetas (p. ej. Foodista[33]).

La aplicación presentada muestra cómo la combinación de los principios fundamentales de la iniciativa Linked Data, la capacidad expresiva del lenguaje para definición de ontologías OWL2, y la potencia de los motores de inferencia (p.ej. Pellet, RacerPro, Fact+) nos ha permitido no sólo reutilizar información publicada en fuentes heterogéneas sino razonar sobre ella y ofrecer nuevos servicios que ofrecen un importante valor añadido al usuario final.

Conclusiones y líneas futuras

El presente trabajo muestra la creación de un nodo Linked Data, un repositorio de datos "interpretables" por aplicaciones software, para el acceso a información de carácter nutricional partiendo de modelos de datos preexistentes. Para ello se hace uso de una metodología reutilizable que se puede aplicar a nuevas fuentes del ámbito de la nutrición o reorientarla para focalizarla cara otros aspectos del conocimiento a representar. De esta manera queremos hacer hincapié en el proceso de generación más allá del propio nodo en sí. Además, sobre este, se ha desarrollado un agente inteligente que es capaz de realizar un procesado significativo de los contenidos para ofrecer al usuario final del sistema información útil para evaluar sus hábitos alimenticios y realizar recomendaciones.

Esta funcionalidad, el procesado autónomo por parte de agentes software para realizar razonamientos, nos sirve de justificación para propugnar la introducción de los modelos de publicación semántica frente a otros basados en documentos web o APIs propietarias. Es decir, el uso de información expresada haciendo uso de estas tecnologías habilita nuevos servicios no factibles en entornos "presemánticos". Este esquema se basa en el uso de estándares que dirigen y facilitan el procesado de los datos para que los propios agentes software los conviertan en información y puedan incorporarlos a sus esquemas de decisión en tiempo de ejecución, es decir, sin modificar el propio código desarrollado inicialmente.

Desde un punto de vista funcional, este trabajo muestra cómo con un esfuerzo inicial reducido es posible desarrollar un sistema autónomo que permite realizar tareas de gran valor sociosanitario. Conviene recordar que los problemas relacionados con una nutrición incorrecta en etapas tempranas tienen un alto impacto en la vida de los niños a largo plazo. Esta mala nutrición puede derivar en problemas como diabetes, problemas cardiovasculares, etc.

Dentro de la línea de uso de repositorios Linked Data, se deben poner en valor no sólo las soluciones ya existentes, sino también aquellas nuevas opciones posibles que tomen esta idea como base. De este modo, como líneas de trabajo de futuro interés se propone el establecimiento de nuevas relaciones con otros nodos ya existentes (o aún por crearse) que generen nuevas opciones de razonamiento o de descubrimiento de datos implícitos. Sobre este modelo de publicación de información basado en conocimiento surgen nuevas opciones dentro del dominio y marco propuesto por este trabajo. Entre ellas podemos destacar el desarrollo de asistentes inteligentes para la creación de recetas de cocina, recomendadores de dietas en función de perfiles personalizados con información médica, de hábitos o de restricciones adicionales derivadas de criterios personales o estacionales, entre otros. En el momento actual, las tareas de investigación se centran en la inclusión de información de carácter estadístico sobre la población que permita ampliar el ámbito de aplicación de esta propuesta al entorno epidemiológico y la realización de predicciones relacionadas con la salud pública gracias a la combinación de diferentes fuentes de información publicadas en Internet.

Agradecimientos

Este trabajo ha sido parcialmente financiado por la Xunta de Galicia mediante el proyecto "Análisis, Diseño y Desarrollo de Servicios Educativos para Televisión Digital. Aplicación en el ámbito de la Educación Infantil" (09SEC035322PR).

Referencias

1. Bizer C, Heath T, Berners-Lee T. Linked Data - The Story So Far. Int *J Semant Web Inf Syst* 2009; 5 (3): 1-22.
2. Drugbank Dataset [base de datos en Internet]. Freie Universität Berlin. Web-based Systems Group [citado 3 Nov 2011]. Disponible en: http://www4.wiwiss.fu-berlin.de/drugbank/
3. Diseasome Dataset [base de datos en Internet]. Freie Universität Berlin. Web-based Systems Group [citado 3 Nov 2011]. Disponible en: http://www4.wiwiss.fu-berlin.de/diseasome/
4. The Linked Clinical Trials [base de datos en Internet]. LinkedCT [citado 3 Nov 2011]. Disponible en: http://data.linkedct.org/
5. USDA National Nutrient Database for Standard Reference [base de datos en Internet]. United States Department of Agriculture [citado 3 Nov 2011]. Disponible en: http://www.nal.usda.gov/fnic/foodcomp/search/
6. Manola F, Miller E. RDF Primer. W3C Recommendation, 10 Febrero 2004. Disponible en: http://www.w3.org/TR/rdf-primer/
7. Gruber TR. Toward principles for the design of ontologies used for knowledge sharing. *Int J Hum Comput Stud* 1995; 43 (4-5): 907-28.

8. Brickley D, Guha RV. RDF Vocabulary Description Language 1.0: RDF Schema. W3C Recommendation, 10 Febrero 2004. Disponible en: http://www.w3.org/TR/rdf-schema/
9. Hitzler P, Krötzsch M, Parsia B, Patel-Schneider PF, Rudolph S. OWL 2 Web Ontology Language - Primer. W3C Recommendation, 27 Octubre 2009. Disponible en: http://www.w3.org/TR/owl2-primer/
10. Adida B, Birbeck M, McCarron S, Pemberton S. RDFa in XHTML: Syntax and Processing. A collection of attributes and processing rules for extending XHTML to support RDF. W3C Recommendation, 14 Octubre 2008. Disponible en: http://www.w3.org/TR/rdfa-syntax/
11. Fielding R, Gettys J, Mogul J, Frystyk H, Masinter L, Leach P et al. Hypertext Transfer Protocol – HTTP/1.1. The Internet Society RFC 2616 Revisión 1.8, 2004.
12. Prud'hommeaux E, Seaborne A. SPARQL Query Language for RD., W3C Recommendation, 15 Enero 2008. Disponible en: http://www.w3.org/TR/rdf-sparql-query/
13. Bizer C, Jentzsch A, Cyganiak R. State of the LOD Cloud [página en Internet]. Freie Universität Berlin [citado 3 Nov 2011]. Disponible en: http://www4.wiwiss.fu-berlin.de/lodcloud/state/
14. Bizer C, Lehmann J, Kobilarov G, Auer S, Becker C, Cyganiak R et al. DBpedia – A Crystallization Point for the Web of Data. *Web Semant* 2009; 7 (3): 154-65.
15. Jentzsch A, Hassanzadeh O, Bizer C, Andersson B, Stephens S. Enabling Tailored Therapeutics with Linked Data. Actas del WWW2009 Workshop on Linked Data on the Web; 20 April 2009, Madrid, España.
16. DCMI Usage Board. DCMI Metadata Terms. DCMI Recommendation, 11 Octubre 2010. Disponible en: http://dublincore.org/documents/dcmi-terms/
17. Miles A, Bechhofer S. SKOS Simple Knowledge Organization System Reference. W3C Recommendation, 18 Agosto 2009. Disponible en: http://www.w3.org/TR/skos-reference/
18. LanguaL™ - the International Framework for Food Description [página en Internet]. European LanguaL Technical Committee [citado 3 Nov 2011]. Disponible en: http://www.langual.org
19. Horrocks I, Patel-Schneider PF, Boley H, Tabet S, Grosof B, Dean M. SWRL: A Semantic Web Rule Language Combining OWL and RuleML. W3C Member Submission, 21 Mayo 2004. Disponible en: http://www.w3.org/Submission/SWRL/
20. Virtuoso Universal Server [página en Internet]. OpenLink Software [citado 3 Nov 2011]. Disponible en: http://virtuoso.openlinksw.com/
21. Heath T, Bizer C. Linked Data: Evolving the Web into a Global Data Space. Synthesis Lectures on the Semantic Web: Theory and Technology. Morgan & Claypool Publishers, 2011.
22. Winkler WE. Overview of Record Linkage and Current Research Directions. Technical Report. Washington, DC: US Bureau of Census, Statistical Research Division. 2006.
23. AGROVOC Linked Open Data (LOD) [base de datos en Internet]. Food and Agriculture Organization of the United Nations [citado 3 Nov 2011]. Disponible en: http://aims.fao.org/standards/agrovoc/linked-open-data
24. Jaro MA. Probabilistic linkage of large public health data files. *Stat Med* 1995; 14 (5-7): 491-8.
25. Levenshtein VI. Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady* 1966; 10: 707-10.
26. Ullmann JR. A Binary n-Gram Technique for Automatic Correction of Substitution, Deletion, Insertion, and Reversal Errors in Words. *The Computer Journal* 1977; 20 (2): 141-7.
27. Volz J, Bizer C, Gaedke M, Kobilarov G. Silk - A Link Discovery Framework for the Web of Data. Actas del WWW2009 Workshop on Linked Data on the Web; 20 April 2009, Madrid, España.
28. Freebase [página en Internet]. Google Inc. [citado 3 Nov 2011]. Disponible en: http://www.freebase.com/
29. BBC-Music [página en Internet]. BBC [citado 3 Nov 2011]. Disponible en: http://www.bbc.co.uk/music
30. MusicBrainz-The Open Music Encyclopedia [página en Internet]. The MetaBrainz Foundation [citado 3 Nov 2011]. Disponible en: http://musicbrainz.org/
31. The Intelligent Music Application Platform - The Echo Nest [página en Internet]. The Echo Nest Corporation [citado 3 Nov 2011] Disponible en: http://the.echonest.com/
32. Álvarez-Sabucedo L, Míguez-Pérez R, Santos-Gago JM, Alonso-Roris VM, Mikic F. Plataforma de e-servicios para educación e higiene nutricionales, orientada a la población infantil. *Salud Colec* 2011; 7 (1): 71-81.
33. The Foodista Dataset [base de datos en Internet]. Foodista, Inc. [citado 3 Nov 2011]. Disponible en: http://kasabi.com/dataset/foodista