

Tecnologías de la Web Semántica: cómo funcionan y cómo interoperan

Raúl García-Castro

Ontology Engineering Group, Departamento de Inteligencia Artificial.
Facultad de Informática, Universidad Politécnica de Madrid, Spain
rgarcia@fi.upm.es

Resumen La Web Semántica es una extensión de la Web actual en la que a la información disponible se le otorga una semántica bien definida. Actualmente existe un problema de interoperabilidad entre las herramientas de la Web Semántica debido a que las herramientas existentes no son capaces de intercambiar conocimiento, lo que conlleva un retraso en la adopción de estas herramientas en el mundo empresarial. En este artículo se presenta el benchmarking como herramienta de mejora del software y se muestran dos actividades que tienen como objetivo mejorar la interoperabilidad de las herramientas de desarrollo de ontologías utilizando como lenguajes de intercambio RDF(S) y OWL.

1. Introducción

La Web ha cambiado profundamente la forma en la que las personas se comunican, negocian y realizan el trabajo diario, al tener acceso a infinidad de recursos, en diferentes formatos y en diferentes idiomas. Todos estos factores han contribuido al éxito de la Web, y a la vez han originado, paradójicamente, uno de sus principales problemas, el exceso de información.

La Web Semántica es una extensión de la Web actual en la que a la información disponible se le otorga una semántica bien definida, permitiendo que aplicaciones heterogéneas descubran y utilicen la información presente en la Web. El mecanismo utilizado para representar el conocimiento en la Web Semántica son las ontologías.

Alrededor de la Web Semántica han surgido numerosos tipos de aplicaciones para crear y procesar este conocimiento, cada una con sus propios modelos de conocimiento y sus técnicas de inferencia. Pero para que la Web Semántica sea una realidad, es necesario que todas estas herramientas interoperen entre ellas, siendo capaces de intercambiar información y de utilizar esta información intercambiada.

El consorcio de la World Wide Web (en adelante W3C) ha producido dos recomendaciones de lenguajes para representar recursos y ontologías en la Web, RDF y OWL respectivamente. Siendo estos lenguajes el medio principal de intercambio de información en la Web Semántica.

El problema actual es que las herramientas existentes no son capaces de intercambiar ontologías utilizando estos lenguajes, existiendo pérdidas de conocimiento en dichos intercambios, y sus usuarios no conocen hasta qué punto pueden intercambiar ontologías entre dos herramientas ni los efectos secundarios de dicho intercambio. Este problema ocasiona un retraso en la adopción de estas herramientas en el mundo empresarial.

Con el objetivo de reforzar la transferencia tecnológica de las herramientas de la Web Semántica, en la Red de Excelencia europea Knowledge Web¹ se realizan una serie de actividades de benchmarking sobre estas herramientas. En este artículo se resumen los progresos de dos de estas actividades que tienen como objetivo mejorar la interoperabilidad de las herramientas de desarrollo de ontologías utilizando como lenguajes de intercambio RDF(S) y OWL.

Este artículo está estructurado de la siguiente manera. La Sección 2 proporciona una visión general de la Web Semántica, las ontologías y la tecnología de la Web Semántica. La Sección 3 plantea el problema de la interoperabilidad en la Web Semántica y la Sección 4 presenta el benchmarking como un proceso de mejora de la tecnología y una metodología de benchmarking de software. Las Secciones 5 y 6 muestran dos actividades de benchmarking que se han llevado a cabo en Knowledge Web para mejorar la interoperabilidad de las herramientas de desarrollo de ontologías. Finalmente, la Sección 7 presenta algunas conclusiones que se pueden derivar de este artículo.

2. La Web Semántica

La Web actual está basada en HTML (HyperText Markup Language) [1], lenguaje de marcado que se utiliza para especificar cómo debe mostrarse la información a las personas en la Web.

En el año 1996, el W3C comenzó el desarrollo del lenguaje de marcado XML (eXtensible Markup Language) [2] que separa la visualización del contenido. XML se diseñó para resolver algunas limitaciones del HTML, permitiendo introducir etiquetas que describen la estructura del contenido presente en el documento o página Web. No obstante, XML no proporciona semántica a dichas etiquetas, con lo cual los programas software tienen dificultades en comprender el significado de una etiqueta codificada en XML.

La siguiente generación de la Web, conocida como la Web Semántica [3] pretende aliviar este problema al describir los recursos disponibles en la Web de forma que sean comprensibles por una máquina, con independencia del idioma en el que se encuentren escritos. Se concibe así la Web Semántica como una extensión de la Web actual en la que a la información disponible se le otorga (anota o marca con) una semántica bien definida (proporcionada por las ontologías), permitiendo que aplicaciones heterogéneas descubran y utilicen la información presente en la Web.

Surgen así nuevos lenguajes de marcado que tienen sus orígenes en los paradigmas de representación del conocimiento de la Inteligencia Artificial, y que

¹ <http://knowledgeweb.semanticweb.org/>

se construyen como extensión a XML. Dichos lenguajes permiten la inserción del vocabulario procedente de una ontología dentro de los documentos Web. Los lenguajes más representativos son los recomendados por el W3C y son:

- **RDF** (Resource Description Framework) [4], desarrollado por el W3C como un lenguaje basado en redes semánticas que describe recursos.
- **RDF Schema** [5], que define una extensión semántica de RDF, proporcionando mecanismos para describir grupos de recursos y relaciones entre dichos recursos.
- **OWL** (Ontology Web Language) [6], creado por el W3C como una extensión del lenguaje RDF Schema proporcionando vocabulario adicional y una semántica formal.

2.1. Ontologías

Una ontología [7] es una especificación formal y explícita de una conceptualización compartida. La conceptualización describe un modelo abstracto de un fenómeno que ocurre en el mundo al identificar los conceptos importantes de ese fenómeno. *Explícito* significa que los conceptos, relaciones, funciones, axiomas que se utilizan para describir el modelo y sus restricciones de uso están claramente definidos. El término *formal* describe el hecho de que la Ontología tiene que ser ejecutable por la computadora. Y el término *compartida* refleja la noción de que una ontología recoge conocimiento consensuado, es decir, que no es privativo de una persona sino que está aceptado por todo un grupo.

Las ontologías se construyen principalmente utilizando técnicas de representación del conocimiento basadas en marcos, lógica de primer orden y lógica descriptiva. Existen dos tipos de ontologías: ontologías ligeras (Lightweight ontologies) y ontologías pesadas (Heavyweight ontologies). Las ontologías ligeras utilizan como componentes de modelado clases, relaciones taxonómicas y no taxonómicas y atributos. Las ontologías pesadas utilizan los componentes anteriores además de axiomas.

2.2. Tecnología de la Web Semántica

Dentro del área de la Web Semántica, se han desarrollado distintas herramientas y plataformas que proporcionan servicios para crear y procesar elementos semánticos. La literatura distingue los siguientes tipos [8]:

- Herramientas de desarrollo de ontologías. Este grupo incluye herramientas y paquetes integrados que pueden ser utilizados para construir una nueva ontología a partir de cero. Complementando las funcionalidades básicas de edición y navegación, estas herramientas proveen normalmente soporte para documentar las ontologías, exportar e importarlas a, o desde, diferentes formatos y lenguajes, editarlas gráficamente, gestionar bibliotecas de ontologías, etc.

- Herramientas de evaluación de ontologías. Son utilizadas para evaluar (aspectos de consistencia, corrección y redundancia) el contenido de las ontologías.
- Herramientas de alineamiento y fusión de ontologías. Estas herramientas son utilizadas para resolver el problema de fusión y alineamiento de diferentes ontologías que pertenecen al mismo dominio.
- Herramientas de anotación basadas en ontologías. Con estas herramientas el usuario puede insertar instancias de conceptos y de relaciones en ontologías y mantener (semi)automáticamente la anotación basada en ontologías de páginas Web.
- Herramientas de consulta de ontologías y motores de inferencia. Éstas permiten la consulta de ontologías de manera sencilla y llevan a cabo inferencias con ellas.
- Herramientas de aprendizaje de ontologías. Estas herramientas pueden obtener ontologías desde textos en lenguaje natural, así como desde fuentes semi-estructuradas y bases de datos, utilizando para ello técnicas tanto del campo de aprendizaje automático como del procesamiento de lenguaje natural.

3. Interoperabilidad en la Web Semántica

Como ya se ha visto en la sección anterior, existe una gran variedad de tecnología dentro de la Web Semántica que permite desarrollar y manejar ontologías. Con gran frecuencia, el desarrollador de ontologías se plantea, al menos, las dos preguntas siguientes antes de reutilizar una ontología ya construida:

- ¿Es posible que una herramienta intercambie o utilice contenidos de ontologías disponibles en otra herramienta?
- ¿Qué componentes se podrían traducir a un lenguaje de intercambio sin perder conocimientos en la traducción?

Cada lenguaje de representación de conocimientos y cada herramienta de ontologías tiene su propio modelo de conocimientos y sus técnicas de inferencia asociadas. Por ello, antes de realizar la traducción entre dos lenguajes o intercambiar ontologías entre dos herramientas es necesario analizar y comparar las características de los modelos de conocimientos y de los razonadores de cada uno de ellos.

Los lenguajes RDF(S) y OWL, recomendados por el W3C, son los formatos más utilizados para el intercambio de ontologías entre plataformas y herramientas de construcción de ontologías, en el contexto de la Web Semántica. Sin embargo, el hecho de que las herramientas dispongan de traductores a estos lenguajes no quiere decir que estas herramientas puedan intercambiar sus ontologías con éxito, como se demuestra en [9].

En el SIG3 Enterprise-Standard Ontology Environments de la red temática OntoWeb² se realizaron experimentos que demuestran que las ontologías

² <http://www.ontoweb.org/>

sólo pueden ser intercambiadas si se modifica y adapta manualmente el código RDF(S) (u OWL) generado por cada herramienta al formato RDF(S) (u OWL) que acepta la herramienta destino. Asimismo, hay pérdidas de conocimientos en los casos en que la exportación e importación son satisfactorias, y no se preservan los conocimientos en transformaciones cíclicas. Igualmente es difícil determinar con exactitud cuáles son las pérdidas de conocimientos que se han producido, dado que los criterios utilizados en la traducción no son explícitos.

Por otro lado, la mayor parte de los servicios de traducción entre lenguajes y, o, herramientas existentes no tienen en cuenta distintos niveles de traducción (codificación, léxico, sintáctico, semántico y pragmático). Consecuentemente, las decisiones de traducción que implementan no son fáciles de entender, y los sistemas son difíciles de mantener.

4. Benchmarking de la Tecnología de la Web Semántica

La tecnología de la Web Semántica ha mejorado considerablemente desde que sus primeras herramientas fueron desarrolladas en los años noventa y, aunque ha sido generalmente utilizada en laboratorios de investigación, en los últimos años las empresas han empezado a interesarse por la tecnología de la Web Semántica y por las aplicaciones que utilizan dicha tecnología.

Para lograr consolidar la tecnología de la Web Semántica, tanto en el mundo industrial como en el académico, es necesario que dicha tecnología alcance un nivel de madurez en el que pueda cumplir con los requisitos de calidad que la industria requiere. Por lo tanto, la tecnología de la Web Semántica necesita por un lado ser evaluada exhaustivamente proporcionando resultados objetivos y por otro lado obtener una mejora masiva en su calidad.

En las últimas décadas, la palabra benchmarking ha tenido una gran relevancia dentro de la comunidad de la gestión empresarial. Una de las definiciones más aceptadas fue proporcionada por Spendolini [10], quien definió benchmarking como un proceso continuo y sistemático para evaluar los productos, servicios y procesos de trabajo de organizaciones que son reconocidas como las que representan las mejores prácticas con el objetivo de la mejora organizativa. La comunidad de Ingeniería del Software no tiene tampoco una definición común del término benchmarking. Algunos autores como Kitchenham [11] consideraron benchmarking como un método de evaluación de software. Para ella, benchmarking es el proceso de ejecutar un número de pruebas estándar con distintos métodos y herramientas, y evaluar el rendimiento relativo de las herramientas en las pruebas. Otros autores, como Wohlin et al. [12] adoptaron la definición de benchmarking de la comunidad empresarial, definiendo benchmarking como un proceso de mejora continua que consiste en ser el mejor de los mejores a través de la comparación de procesos similares en diferentes contextos.

Partiendo de las ideas de mejora continua y de búsqueda de mejores prácticas de la comunidad de la gestión empresarial, hemos desarrollado una metodología que considera el benchmarking de software como un proceso de mejora continua en vez de como una actividad puntual. Esta metodología hace hincapié en com-

parar el software mediante evaluaciones y en realizar el benchmarking siguiendo un procedimiento sistemático.

4.1. Metodología de Benchmarking de Software

Esta sección resume la metodología de benchmarking desarrollada por el autor, la cual proporciona una serie de directrices para realizar benchmarking sobre software. Esta metodología ha sido desarrollada a partir de otras metodologías de benchmarking procedentes de la comunidad empresarial, la Ingeniería Software Experimental y la Medición de Software. Una descripción detallada de la misma puede encontrarse en [13].

La Figura 1 muestra las fases principales de la metodología de benchmarking de software. Cada una de las iteraciones de la metodología se compone de tres fases (Planificación, Experimentación y Mejora) y finaliza con una tarea de recalibrado.

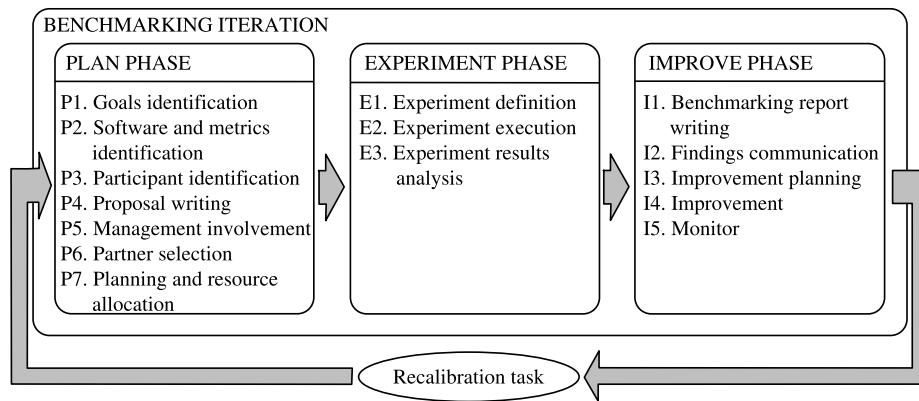


Figura 1. Metodología de benchmarking de software [13]

- **Fase de Planificación.** Su objetivo principal es obtener un documento con la propuesta de benchmarking detallada. Este documento se usará como referencia a lo largo del benchmarking, y debe incluir toda la información relevante acerca del mismo: sus objetivos, beneficios y costes; el software (y sus funcionalidades) que será evaluado; las métricas que se utilizarán para evaluar dichas funcionalidades; y las personas involucradas en el benchmarking. Las últimas tareas de esta fase consisten en buscar otras organizaciones que quieran participar en el benchmarking con otro software con el fin de llegar a un acuerdo en la propuesta de benchmarking (tanto dentro de la organización como con las otras organizaciones) y en la planificación del benchmarking.

- **Fase de Experimentación.** En esta fase, las organizaciones deben definir y ejecutar los experimentos de evaluación para cada software que participe en el benchmarking. Los resultados de la evaluación deben ser recopilados y analizados, determinando las prácticas que conducen a los mismos e identificando qué prácticas son las mejores prácticas.
- **Fase de Mejora.** La primera tarea de esta fase comprende la escritura del informe sobre el benchmarking, el cual debe incluir: un resumen del proceso seguido, los resultados y las conclusiones de la experimentación, recomendaciones para mejorar el software y las mejores prácticas encontradas en la experimentación. Los resultados del benchmarking deben ser comunicados a las organizaciones que participan en el mismo y finalmente, en varios ciclos de mejora, los desarrolladores del software deben realizar los cambios necesarios para mejorarlo y monitorizar esta mejora.

La tarea de recalibrado se realiza al final de cada iteración. En ella no se mejora la herramienta si no el mismo proceso de benchmarking, utilizando las lecciones aprendidas al llevar a cabo las fases del mismo.

4.2. Benchmarking en Knowledge Web

Knowledge Web es una Red de Excelencia europea cuya misión es la de reforzar a la industria europea y a los proveedores de servicios en el área de la Web Semántica. Dicho proyecto concentra sus esfuerzos en la transferencia de la tecnología de la Web Semántica a la industria, además de mejorar la investigación y la educación en dicho área.

Knowledge Web promueve varias actividades de benchmarking de la tecnología de la Web Semántica. Entre las herramientas que se contemplan para estas actividades se encuentran las herramientas de desarrollo de ontologías, las herramientas de alineación de ontologías, las herramientas de anotación basadas en ontologías y los razonadores.

Para llevar a cabo estas actividades, se creó un método sistemático de evaluación de esta tecnología que permita una mejora continuada de la misma, presentado en la sección anterior.

Las dos siguientes secciones presentan dos actividades de benchmarking de herramientas de desarrollo de ontologías llevadas a cabo en Knowledge Web. Otro benchmarking que está teniendo lugar y que no está cubierto en este artículo es el de las herramientas de alineación de ontologías. Más información sobre esta actividad se puede encontrar en la web³.

5. RDF(S) Interoperability Benchmarking

La metodología de benchmarking de software presentada en la Sección 4.1 ha sido utilizada para realizar el benchmarking de la interoperabilidad de las

³ <http://oaei.ontologymatching.org/>

herramientas de desarrollo de ontologías utilizando como lenguaje de intercambio RDF(S).

Esta actividad de benchmarking ha sido organizada dentro de Knowledge Web y la participación en la misma está abierta a cualquier organización. Además, toda la información relevante acerca del benchmarking se encuentra disponible en una página web⁴.

El objetivo del benchmarking consiste en mejorar la interoperabilidad de las herramientas de desarrollo de ontologías mediante la evaluación de los importadores y exportadores a RDF(S) que dichas herramientas poseen, y así identificar los componentes ontológicos (clases, atributos, relaciones, etc.) exportados e importados correctamente.

La interoperabilidad de las herramientas de desarrollo de ontologías utilizando RDF(S) como lenguaje de intercambio requiere que los importadores y exportadores desde/a RDF(S) de las herramientas funcionen correctamente. Por lo tanto, la experimentación comprende tres fases consecutivas:

- **Fase de acuerdo.** El primer paso es llegar a un acuerdo en la definición de los conjuntos de pruebas, los cuales son comunes a todas las herramientas, ya que la calidad de dichos conjuntos de pruebas es esencial para obtener buenos resultados en el benchmarking.
- **Fase de evaluación 1.** Los importadores y exportadores desde/a RDF(S) de las herramientas de desarrollo de ontologías son evaluados con las versiones definitivas de los conjuntos de pruebas.
- **Fase de evaluación 2.** La segunda fase de evaluación cubre la evaluación del intercambio de ontologías entre todas las herramientas de desarrollo de ontologías.

Siete herramientas están participando en el benchmarking. De éstas, cuatro son herramientas de desarrollo de ontologías: KAON⁵, OntoStudio⁶, Protégé⁷ utilizando su motor RDF y WebODE⁸; y tres son repositorios de RDF: Corese⁹, Jena¹⁰ y Sesame¹¹.

Un resumen de esta actividad de benchmarking, una explicación de cómo se definieron los conjuntos de pruebas, y los resultados detallados para las herramientas que han participado en el mismo pueden consultarse en [14].

⁴ http://knowledgeweb.semanticweb.org/benchmarking_interoperability/

⁵ <http://kaon.semanticweb.org/>

⁶ <http://www.ontostudio.de/>

⁷ <http://protege.stanford.edu/>

⁸ <http://webode.dia.fi.upm.es/WebODEWeb/index.html>

⁹ <http://www-sop.inria.fr/acacia/soft/corese/>

¹⁰ <http://jena.sourceforge.net/>

¹¹ <http://www.openrdf.org/>

6. OWL Interoperability Benchmarking

En la actualidad, dentro de Knowledge Web se está realizando una segunda actividad de benchmarking de la interoperabilidad de las herramientas de desarrollo de ontologías.

La diferencia entre estas dos actividades radica en el lenguaje de intercambio. Mientras que en el anterior benchmarking se utilizaba RDF(S) como lenguaje de intercambio, esta vez el lenguaje utilizado es OWL, el lenguaje de representación de ontologías recomendado por el W3C.

Esta actividad de benchmarking también sigue la metodología presentada en la Sección 4.1, y se está desarrollando siguiendo la misma aproximación que en el benchmarking de la interoperabilidad utilizando RDF(S) como lenguaje de intercambio.

Al igual que en el anterior benchmarking, la participación en el mismo está abierta a cualquier organización, estando toda la información acerca del mismo disponible en una página web¹².

Actualmente, el benchmarking se encuentra en la fase de definición de los conjuntos de pruebas que serán usados en la experimentación, y se han definido una serie de ontologías en OWL Lite para evaluar la importación de las herramientas.

Una de las conclusiones que se alcanzó al finalizar la primera iteración del benchmarking utilizando RDF(S) como lenguaje de intercambio es que la ejecución manual de los casos de prueba es muy costosa y puede dar lugar a fallos humanos. Por lo tanto, en este benchmarking el objetivo es realizar las pruebas de una manera automática.

7. Conclusiones

Este artículo presenta una descripción general de lo que son la Web Semántica y las ontologías, y muestra algunos ejemplos de la tecnología que se usa en la misma.

Además, se plantea la necesidad de evaluar y de realizar benchmarking de la tecnología de la Web Semántica. Y para ello se describe una metodología de benchmarking de software que puede ser utilizada para evaluar y mejorar esta tecnología.

Uno de los beneficios del benchmarking es que es una actividad realizada por miembros de una comunidad ya sea de investigación o industrial. Gracias a esto, los expertos de dicha comunidad participan activamente y los beneficios obtenidos al finalizar el benchmarking tienen un efecto en toda la comunidad.

No obstante, el benchmarking no es la solución para todos los casos. Una tarea preliminar al benchmarking sería la de evaluar si el benchmarking es la aproximación correcta o no, siendo éste útil cuando los objetivos son mejorar el software y obtener las mejores prácticas realizadas por otros.

¹² http://knowledgeweb.semanticweb.org/benchmarking_interoperability/owl/

Agradecimientos

Este trabajo está parcialmente financiado por una beca FPI del Ministerio de Educación y Ciencia español (BES-2005-8024), por el proyecto IST Knowledge Web (IST-2004-507482) y por el proyecto CICYT Infraestructura tecnológica de servicios semánticos para la web semántica (TIN2004-02660).

Referencias

1. Raggett, D., Hors, A.L., Jacobs, I.: HTML 4.01 Specification. W3C Recommendation (1999)
2. Bray, T., Paoli, J., Sperberg-McQueen, C., Maler, E., Yergeau, F.: Extensible markup language (xml) 1.0 (fourth edition) w3c recommendation (2006)
3. Berners-Lee, T.: Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web. HarperCollins, New York, NY (1999)
4. Manola, F., Miller, E.: RDF Primer. W3C Recommendation 10 February 2004 (2004)
5. Brickley, D., Guha, R.V. (editors): RDF Vocabulary Description Language 1.0: RDF Schema. W3C Recommendation 10 February 2004 (2004)
6. Smith, M.K., Welty, C., McGuinness, D.L. (editors): OWL Web Ontology Language Guide. W3C Recommendation 10 February 2004 (2004)
7. Studer, R., Benjamins, V., Fensel, D.: Knowledge engineering: Principles and methods. IEEE Transactions on Data and Knowledge Engineering 25 (1998) 161–197
8. OntoWeb: Ontoweb deliverable 1.3: A survey on ontology tools. Technical report, IST OntoWeb Thematic Network (2002)
9. Corcho, O.: A layered declarative approach to ontology translation with knowledge preservation. Volume 116 of Frontiers in Artificial Intelligence and its Applications. IOS Press (2005)
10. Spendolini, M.: The Benchmarking Book. AMACOM, New York, NY (1992)
11. Kitchenham, B.: DESMET: A method for evaluating software engineering methods and tools. Technical Report TR96-09, Department of Computer Science, University of Keele, Staffordshire, UK (1996)
12. Wohlin, C., Aurum, A., Petersson, H., Shull, F., Ciolkowski, M.: Software inspection benchmarking - a qualitative and quantitative comparative opportunity. In: Proceedings of 8th International Software Metrics Symposium. (2002) 118–130
13. García-Castro, R., Maynard, D., Wache, H., Foxvog, D., González-Cabero, R.: D2.1.4 Specification of a methodology, general criteria and benchmark suites for benchmarking ontology tools. Technical report, Knowledge Web (2004)
14. García-Castro, R., Sure, Y., Zondler, M., Corby, O., Prieto-González, J., Bontas, E.P., Nixon, L., Mochol, M.: D1.2.2.1.1 Benchmarking the interoperability of ontology development tools using RDF(S) as interchange language. Technical report, Knowledge Web (2006)