Una propuesta para la adaptación del razonamiento semántico a los recursos disponibles

Carlos Bobed[1,2], Fernando Bobillo[1,2], and Eduardo Mena[1,2]

[1] Universidad de Zaragoza, Zaragoza
[2] Instituto Universitario de Investigación en Ingeniería de Aragón (I3A), Zaragoza
{cbobed,fbobillo,emena}@unizar.es

Resumen Presentamos una arquitectura para que sistemas inteligentes basados en representación del conocimiento y razonamiento se adapten dinámicamente a los recursos disponibles, sin fallar por falta de memoria o sufrir tiempos de respuesta excesivos. La idea básica (resumen de [1]) consiste en predecir el uso de recursos y adaptar la cantidad de conocimiento a los recursos disponibles, obteniendo un subconjunto del original sobre el que razonar. La propuesta se ha concretado a razonadores semánticos, se ha implementado en un prototipo y se ha evaluado empíricamente mostrando resultados prometedores.

Keywords: Ontologías · Razonamiento semántico · Modularización.

1. Introducción

Muchas aplicaciones reales de la Inteligencia Artificial requieren la representación y la gestión de conocimiento para llevar a cabo sus cometidos. Las tecnologías semánticas (o tecnologías de la Web Semántica) y, en particular, las ontologías son la principal opción en la actualidad para ello. Una ontología es una especificación formal de una conceptualización compartida y permite representar el conocimiento relevante en un cierto dominio de interés, haciendo explícito el significado (la semántica) del conocimiento y favoreciendo que las máquinas puedan explotarlo. Una ventaja fundamental de las ontologías es la posibilidad de usar *razonadores semánticos*, que son aplicaciones software que permiten inferir conocimiento implícito que se deduce necesariamente de la ontología, permitiendo responder preguntas sobre una ontología como recuperar las instancias de una clase, clasificar su taxonomía de clases, etc.

Si bien existen algunos razonadores semánticos específicamente diseñados para dispositivos con pocos recursos [5], los razonadores actuales no tienen en cuenta los recursos disponibles en el entorno de ejecución, entendiendo el término *recurso* de una manera amplia que incluye el hardware (capacidad de la CPU, memoria, batería, conectividad, etc.) y las preferencias del usuario (tiempo máximo de respuesta, nivel de privacidad, etc.), etc. Esto hace que no pocas veces la tarea de razonamiento acabe con errores de falta de memoria o se tenga que abortar por un tiempo excesivo de ejecución. Esta necesidad de recursos depende de la tarea específica de razonamiento a resolver y del tamaño y complejidad semántica de la ontología de entrada.

El objetivo de este artículo es discutir un enfoque para la adaptación del razonamiento semántico a los recursos disponibles, originalmente publicado en septiembre de

2024 [1]. La idea básica consiste en predecir el uso de recursos y en adaptar la ontología para adecuarla a los recursos disponibles, obteniendo una subontología de la original.

2. Razonamiento adaptado a recursos

Arquitectura general. Como primera aportación, se propone una arquitectura general basada en un módulo de Evaluación de Restricciones y un módulo de Extracción de Conocimiento. En primer lugar, el módulo de Evaluación de Restricciones recibe la ontología y evalúa el cumplimiento de una serie de restricciones (e.g., tiempo máximo de razonamiento) según las características de la ontología y los recursos del dispositivo. Si se predice que la ontología es adecuada, se envía directamente al razonador semántico. En otro caso, se transfiere al módulo de Extracción de Conocimiento, que adapta el conocimiento según las restricciones. En concreto, se calcula una subontología de la ontología de entrada, intentando priorizar la conservación del conocimiento más importante para minimizar la pérdida semántica. Esta subontología se devuelve al Módulo de Evaluación de Restricciones para su evaluación, repitiendo el proceso tantas veces como sea necesario, hasta obtener una subontología que se considere manejable por el razonador semántico. Si la subontología está vacía, significa que no fue posible cumplir con todas las restricciones. En otro caso, se ejecuta la tarea de razonamiento sobre la ontología obtenida.

Instanciación. Esta arquitectura general se concreta como se aprecia en la Figura 1. Por supuesto, no es la única instanciación posible. Como puede verse, en cada fase se enumeran diferentes posibles opciones que podrían aplicarse. Además, se tiene en cuenta un módulo de Predicción de los Recursos necesarios, en el que se basaría el módulo de Evaluación de Restricciones para su decisión. Tanto el módulo de Evaluación de Restricciones como el de Extracción de Conocimiento son muy generales y consideran diferentes criterios posibles para evaluar y para crear la subontología, respectivamente. Debemos mencionar que aunque existen trabajos previos que permiten identificar los elementos más importantes y calcular fragmentos de una ontología, no tienen en cuenta los recursos disponibles y deben adaptarse a nuestro nuevo escenario.

Prototipo. Como prueba de concepto, se implementa un prototipo que predice el tiempo de razonamiento semántico en ordenadores de sobremesa, reusando estrategias previas [4]. Para generar la subontología, se parte de una signatura inicial, se añaden los axiomas de subclases y se continúa añadiendo axiomas, siguiendo un método de búsqueda en anchura, hasta reducir un 25% en cada iteración (también se realizaron pruebas con un 12,5%). Soporta cualquier razonador semántico compatible con la OWL API.

Pérdida semántica. Un aspecto clave de nuestra propuesta es que, dado que estamos descartando axiomas de la ontología para adaptarnos a los recursos, debemos ser capaces de estimar cuánto conocimiento estamos perdiendo realmente. Es decir, no basta con conocer el número de axiomas, sino las inferencias que dejarían de poder realizarse. Para ello, adaptamos unas medidas propuestas en el ámbito del alineamiento de ontologías [3] para definir varias medidas de exhaustividad semántica (*semantic recall*). La idea intuitiva es calcular la proporción entre el número de posibles consecuencias

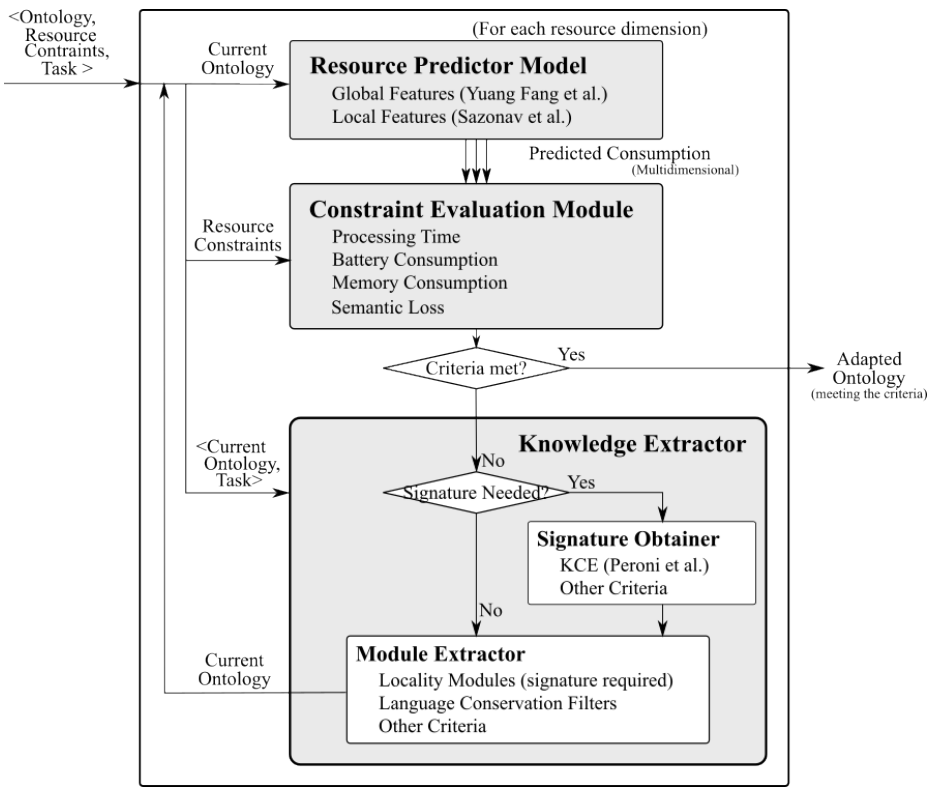


Figura 1. Posible implementación de nuestra arquitectura general.

que se pueden obtener de la subontología adaptada con respecto a la ontología original. La formulación es más compleja, pues restringe las posibles consecuencias (que serían infinitas) y tiene en cuenta la signatura inicial. Además, definimos una medida de solapamiento semántico (*semantic overlap*) que permite comparar varias subontologías adaptadas. De momento, estas medidas solo nos permiten evaluar empíricamente la pérdida semántica pero, en el futuro, esta pérdida semántica podría ser un criterio más a considerar a la hora de decidir qué o cuánto conocimiento descartar.

Evaluación. Para demostrar la viabilidad de nuestra propuesta, se realizaron varios tipos de experimentos. En primer lugar, puesto que la reducción del conocimiento depende del coste estimado, se evaluó la predicción del tiempo de ejecución en ordenadores de sobremesa. A partir de 143 características de las ontologías de entrada, se usaron bosques aleatorios (*random forests*) para predecir el tiempo de razonamiento en un conjuntos de datos con más de 10.000 ontologías. Se analizaron varias medidas (R^2, MAE y RMSE) y se evaluó una técnica para aumentar los datos, con resultados satisfactorios.

En segundo lugar, se evaluó la capacidad de adaptar las ontologías, trabajando con matrices de confusión que tenían en una dimensión los números *estimados* tanto de las ontologías que consideramos adecuadas como de las que pensamos que requieren

una adaptación, y en la otra dimensión los números *reales* de ambos casos. El peor caso son los falsos positivos (ontologías que se predicen como adecuadas pero en la práctica no lo son), pero fueron solo un 0.38% de las ontologías. En cambio, el prototipo fue capaz de reducir el 14.46% de ontologías que se predijo correctamente que consumirían demasiados recursos a un 1.90%.

En tercer lugar, se evaluó empíricamente la pérdida semántica de nuestra técnica de adaptación de ontologías, obteniendo resultados satisfactorios tanto con respecto a otros métodos previos para modularizar ontologías [2] como con respecto a las 1.892 ontologías que se adaptaron reduciendo su tamaño.

Por último, se ilustró su utilidad con algunas ontologías particulares. Por ejemplo, al tratar de razonar con la ontología médica GALEN, de 20.1 MB, en un PC con CPU Intel Core i7-6700K@4.00 GHz y 24GB de memoria RAM, el razonador HermiT dio un error de falta de memoria; por ello, algunos investigadores han extraído manualmente ciertos fragmentos de esta ontología. Con nuestro prototipo, este proceso se automatiza, permitiendo razonar con GALEN sin provocar errores.

3. Discusión y trabajo futuro

Esta propuesta es la primera aportación al razonamiento semántico adaptable a los recursos disponibles. Creemos que este artículo muestra claramente el potencial de esta línea de investigación, pero también que aún queda un largo camino por recorrer. Por ejemplo, se necesitan experimentos en dispositivos móviles (con recursos mucho más limitados que los de sobremesa) para confirmar empíricamente que la predicción es adecuada. Además, nuestro prototipo podría ampliarse de varias maneras: utilizando estrategias de predicción más complejas (pero eficientes), mejorando las técnicas identificación de los elementos más importantes de la ontología o decidiendo si realizar un razonamiento local o delegarlo (total o parcialmente) en otro dispositivo.

Agradecimientos. Trabajo financiado por los proyectos de I+D+i PID2024-159530OB-I00 (MCIN/AEI 10.13039/501100011033) y UZ2024-IyA-02 (Universidad de Zaragoza).

Referencias

1. Bobed, C., Bobillo, F., Jiménez-Ruiz, E., Mena, E., Pan, J.Z.: Praedixi, redegi, cogitavi: Adaptive knowledge for resource-aware semantic reasoning. Expert Systems with Applications **250**, 123838 (2024). https://doi.org/10.1016/j.eswa.2024.123838
2. Cuenca Grau, B., Horrocks, I., Kazakov, Y., Sattler, U.: Modular reuse of ontologies: Theory and practice. Journal of Artificial Intelligence Research **31**, 273–318 (2008)
3. Euzenat, J.: Semantic precision and recall for ontology alignment evaluation. In: Proceedings of the 20th International Joint Conference on Artifical Intelligence (IJCAI 2007). pp. 348–353 (2007)
4. Pan, J.Z., Bobed, C., Guclu, I., Bobillo, F., Kollingbaum, M.J., Mena, E., Li, Y.F.: Predicting reasoner performance on ABox intensive OWL 2 EL ontologies. International Journal on Semantic Web and Information Systems **14**(1), 1–30 (2018)
5. Ruta, M., Scioscia, F., Bilenchi, I., Gramegna, F., Loseto, G., Ieva, S., Pinto, A.: A multiplatform reasoning engine for the semantic web of everything. Journal of Web Semantics **73**, 100709 (2022)