

Poblado automático de ontologías de perfiles académicos a partir de textos en español

José A. Reyes-Ortiz, Maricela Bravo, Oscar Herrera y Alejandro Gudiño

Universidad Autónoma Metropolitana, Unidad Azcapotzalco,
Departamento de Sistemas, Distrito Federal,
México

{jaro, mcbc, oha}@correo.azc.uam.mx, al209202106@alumnos.azc.uam.mx

Resumen. Este artículo propone un enfoque para el poblado automático de ontologías de perfiles académicos a partir de los textos, expedientes curriculares y resúmenes, de publicaciones científicas en español. El enfoque utiliza reglas semánticas y marcadores lingüísticos para extraer los individuos de clase, relaciones y valores de propiedad. Una evaluación ha sido realizada con un conjunto de individuos *gold standard*, en términos de precisión y exhaustividad para el poblado de elementos ontológicos.

Palabras clave: poblado automático de ontologías, reglas semánticas, marcadores lingüísticos, procesamiento de lenguaje natural (PLN).

1. Introducción

En los últimos años, ha surgido la necesidad de procesar la información de manera automática debido al crecimiento acelerado de la información electrónica disponible en Internet, empresas, organizaciones y repositorios en general. Para lograr este procesamiento ha sido necesario representar la información de tal manera que sea procesable con computadoras. Esta representación de conocimiento puede realizarse con ontologías, las cuales en los últimos años han ganado importancia [1].

Según [2], las ontologías poseen características significativas que las posicionan en una de las formas de representación más utilizadas inicialmente en la Web, y ahora en diversos medios electrónicos. Sus componentes básicos son: *conceptos*, *relaciones (propiedades de objeto y propiedades de dato)*, *funciones*, *individuos de clase*, *relaciones entre individuos*, *valores de las propiedades* y *axiomas*. La construcción de ontologías puede realizarse de manera manual pero esto ocasiona diversos problemas de costo y tiempo. Como una alternativa a esto surge el aprendizaje automático de ontologías a partir de textos cuyo objetivo es identificar los elementos ontológicos de manera automática o semiautomática [3].

En el contexto del aprendizaje de elementos ontológicos, el poblado de ontologías hace referencia a extraer y representar, de manera automática o semiautomática, los *individuos de clase*, *relaciones entre los individuos* y *valores de las propiedades*. En la

literatura, existen enfoques ligeros para el poblado de ontologías basados en modelos estadísticos-sintácticos ([4, 5, 6, 7]) y enfoques semánticos basados en reglas ([8, 9, 10]), estos no satisfacen el poblado ya que desatienden el conocimiento lingüístico profundo de los textos.

Por lo tanto, en este artículo nos centramos la *extracción de individuos*, sus *relaciones* y los valores de las *propiedades de datos* para la representación semántica de publicaciones científicas a partir de textos de artículos en español. El enfoque propuesto se basa en reglas semánticas y marcadores lingüísticos que capturan la *información semántica* de los textos. La idea de proponer un enfoque para el español se debe, en gran medida, a la carencia de enfoques, herramientas o métodos de poblado de ontologías para este idioma.

El resto del artículo está organizado de la siguiente manera. En la Sección 2 se presenta las bases sobre el poblado de ontologías a partir de textos y el conjunto de tareas que intervienen. La Sección 3 expone un acercamiento al estado de arte de los trabajos relacionados al poblado de ontologías, particularmente, de dominio de los perfiles académicos o publicaciones científicas. El enfoque propuesto en este trabajo se muestra en la Sección 4, el cual consta de tres etapas: el etiquetado de los textos, la extracción de individuos basadas en reglas semánticas y el poblado de individuos en las ontologías. La Sección 5 presenta la experimentación y los resultados obtenidos en el enfoque propuesto. Finalmente, las conclusiones y trabajos futuros son presentados en la Sección 6.

2. Poblado de ontologías a partir de textos

Las ontologías han ganado gran popularidad en la comunidad científica al convertirse en una alternativa viable para la representación de conocimiento [1]. Este conocimiento se puede encontrar en los textos y necesita ser representado en una estructura que sea procesable por computadoras con la finalidad de lograr su manejo automático. Por ello, se plantea construir ontologías a partir de textos de manera cuidadosa para expresar con claridad y precisión el conocimiento. Éste es un proceso crítico, el cual puede realizarse de manera manual pero representa una tarea tediosa, costosa y que consume mucho tiempo [1].

En efecto, el proceso de construcción de ontologías se puede realizar de manera semiautomática o automática con *aprendizaje de ontologías*, tarea que consiste en construirlas a partir de un conjunto de datos iniciales, las cuales pueden ser un *corpus* de textos en lenguaje natural [1]. La idea consiste en partir de un conjunto de textos y relacionar los segmentos del texto (categorías gramaticales) con entidades ontológicas. Esta relación puede ser rígida y poco flexible; sin embargo, existen las excepciones, en las cuales intervienen técnicas de Procesamiento de Lenguaje Natural (PLN) para lograr una buena representación de los textos mediante ontologías.

El aprendizaje de ontologías incluye en una serie de tareas ordenadas [3]:

1. Adquisición de terminología relevante.
2. Adquisición de sinónimos.
3. Formación de conceptos.
4. Organización jerárquica de los elementos.

5. Aprendizaje de relaciones, propiedades, atributos, junto con su respectivo rango y dominio.
6. Organización jerárquica de las relaciones.
7. Instanciación de los axiomas del esquema.
8. Definición de los axiomas arbitrarios.

Este trabajo se centra en el *poblado de ontologías*, el cual corresponde a la tarea que [3] define como instanciación de los axiomas. El *poblado de ontologías* hace referencia al descubrimiento, a partir de los textos, de individuos de clases, relaciones entre los individuos y su representación en el modelo ontológico. Diversos trabajos han sido propuestos en este ámbito [11], [5] y [12] por ello se ha dedicado la Sección 3 de este trabajo a revisar el estado del arte relacionado al poblado de ontologías, particularmente del dominio académico y de perfiles profesionales.

3. Trabajos relacionados

El poblado de ontologías a partir de textos en forma automática o semiautomática es un área que ha sido considerada por la comunidad de investigadores, un reto actual para el *Procesamiento de Lenguaje Natural* [3]. Por ello, en la literatura existe una amplia diversidad de enfoques sobre el poblado de ontologías a partir de textos y, particularmente, considerando ontologías de perfiles académicos.

En el aspecto del poblado de ontologías de diversos dominios, se han identificado características valiosas que han servido para organizar los enfoques, tales como: el idioma de los textos de entrada, poblados de elementos ontológicos, el tipo de enfoque de PLN utilizado para extraer el conocimiento y el nivel de análisis aplicado a los textos.

El idioma de los textos de entrada es una característica que define, entre otras cosas, el tipo de enfoque a aplicar y el nivel de análisis. Esto debido a la existencia de una diferencia de complejidad en las estructuras sintácticas y semánticas de los idiomas. De esta manera, existen enfoques que consideran al inglés como su dominio de los documentos de entrada como en [4, 6, 7, 8, 9, 10, 13, 14, 15, 16 y 17], el idioma francés considerado en los trabajos presentados en [18], enfoques que han sido propuestos para el español [19, 20, 21] y enfoques que son independientes del idioma en sus textos de entrada como en [22, 23].

Los elementos ontológicos que son poblados, a partir de textos, por los diversos enfoques son: *instanciación de clases*, *descubrir y representar individuos* [5, 6, 11, 20 y 23]; poblado de propiedades de objetos como en [4, 24]; poblado de valores de propiedades de dato [4, 8, 11]; finalmente, el poblado de relaciones entre individuos [4, 7, 9, 10, 11, 21].

El tipo de enfoque y nivel de análisis de PLN aplicado a los textos de entrada con la finalidad de extraer el conocimiento para el poblado de las ontologías depende del idioma del dominio. En este ámbito, se han identificado trabajos que utilizan un análisis ligero de los textos con un enfoque estadístico como en [4]; enfoques basados en dependencias sintácticas con un análisis superficial han sido propuestos en [5, 6, 7];

semántico o heurístico basado en reglas [8, 9, 10, 11, 20, 21, 23] y un enfoque lingüístico con un análisis ligero de los textos ha sido presentado en [24].

Por otro lado, el poblado automático de ontologías a partir de textos, en el dominio de perfiles profesionales o académicos, ha sido propuesto por trabajos como en [25] que presenta un marco de trabajo para la extensión semántica de documentos PDF escritos en LATEX, mediante la creación de metadatos para publicaciones científicas. Este marco de trabajo permite a los autores la creación de metadatos mientras se desarrolla el proceso de escritura de documentos; el trabajo presentado en [26] muestra un métodos para la ingeniería ontológica y poblado de ontologías a partir de textos, el cual fue probado con textos sobre experiencias académicas y de proyectos de investigación; el sistema descrito en [27] utiliza técnicas de extracción de información y aprendizaje automático para el reconocimiento y clasificación de eventos a partir de artículos de noticias electrónicas, las cuales describen la vida académica en el área de conocimiento multimedia, en particular, eventos como la adjudicación de proyectos, investigaciones y visitas; el trabajo expuesto en [28] ha presentado una herramienta para el poblado automatizado de ontologías a partir de textos utilizando extractores, es decir patrones, escritos por humanos con la finalidad de capturar el conocimiento de los textos. La herramienta ha sido utilizada en el dominio académico, donde se han poblado, entre otras clases, *Publicaciones*, *Revistas*, *Conferencias* y *Libros*, además de las relaciones *responsabilidad_de*, *publicado_en* y *autor_listado_en*; finalmente, el trabajo presentado en [29] propone una descripción ontológica del dominio de las ciencias computacionales y la aplicación de un enfoque de aprendizaje supervisado para la extracción de individuos a partir de artículos anotados manualmente.

El poblado automático de ontologías a partir de textos exige la intervención de técnicas de procesamiento de lenguaje natural. Además hemos identificado una desatención en la literatura para proponer enfoques de poblado de ontologías para el dominio académico en español. Por lo tanto, en este artículo se considera el idioma español como dominio de los textos de entrada, reglas semánticas basadas en marcadores lingüísticos, mediante un análisis profundo de los textos, con la finalidad de lograr el poblado automático de ontologías de perfiles académicos. El enfoque propuesto considera el poblado de individuos de clase, propiedades de objeto, propiedades de dato y relaciones entre los individuos.

4. Enfoque propuesto

El enfoque que se propone en este artículo para el poblado de ontologías de perfiles académicos, a partir del texto de los resúmenes, se centra en descubrir individuos de publicaciones y sus propiedades relacionadas con investigadores. A partir de los expedientes curriculares, se descubren las *propiedades de dato* de las publicaciones como título, autores listados, editor y año de la publicación. El contenido del artículo, particularmente el resumen, también es procesado para obtener información detallada de la publicación científica, tal como: solución aportada, evidencia de la solución y propósito del método.

El enfoque propuesto se compone de tres etapas: 1. Etiquetado de los textos, 2. Extracción de elementos ontológicos (*individuos*) basada en reglas semánticas y 3. El poblado de las ontologías.

4.1. Etiquetado de los textos

En la primera etapa, el texto de los expedientes curriculares y los resúmenes de las publicaciones científicas se divide en oraciones, y después en palabras para su etiquetado morfológico. Las tareas estándar de la segmentación de las palabras y el etiquetado morfológico se realizan con la interfaz de programación de aplicaciones proporcionada en el marco de desarrollo para el procesamiento de lenguaje natural llamado *GATE* [30].

El marco de desarrollo *GATE* proporciona los componentes necesarios para realizar la segmentación del texto en oraciones (*Sentence_Splitter*) y palabras (*Tokenizer*), además del etiquetado morfológico (*POS Tagger*). Estos componentes son fácilmente ensamblados para lograr una aplicación más compleja basada en tuberías, donde se agrega el componente de extracción de información.

El etiquetado de partes de la oración se encarga de asignar una categoría gramatical a cada palabra. Para esta tarea se recurre al componente llamado *POSTagger* en el cual, para el caso de textos en español, utiliza el etiquetador *Spanish TreeTagger* [31].

4.2. Extracción de información

La segunda etapa consiste en realizar la extracción de individuos de las publicaciones, sus relaciones y sus propiedades de dato a partir de los textos etiquetados.

Para esta tarea de extracción de información, se utiliza el componente de *GATE* llamado *Jape Transducer*, el cual se encarga de compilar y ejecutar un conjunto de reglas basadas en la gramática *JAPE* (*Java Annotation Pattern Engine*) [32]. Un conjunto de 27 reglas semánticas fueron construidas y codificadas en la gramática *JAPE* con la finalidad de extraer los individuos, sus *propiedades de objeto* y sus *propiedades de dato*.

```
Rule: Propósito
(
  ({SpaceToken.kind == space}
    ({Token.string == "para"})
    {SpaceToken.kind == space}
  )
  (({Token.pos == VLinf})
    (NOUNPHRASE)
  ):prop
)--> :prop.Propósito = {rule = "Propósito", text =:prop@string}
```

Fig. 1. Regla en JAPE para la extracción de la relación *tiene_propósito*.

Los individuos de las publicaciones se encuentran relacionados mediante las siguientes propiedades de objeto: *publicada_en*, *es_listado_en* y *aporta_solución*. Además, se crearon reglas semánticas basadas en marcadores lingüísticos para la extracción de los valores de las siguientes propiedades de datos: *tiene_título*, *tiene_año_publicación*, *tiene_propósito*, *tiene_evidencia*. En la Figura 1 se muestra una regla semántica, codificada en la gramática *JAPE*, creada para la extracción del propósito de una publicación a partir del resumen (texto) del artículo científico.

Los individuos relacionados, mediante las propiedades de objeto llamadas *aporta_solución*, *tiene_propósito* y *tiene_evidencia*, se extraen mediante reglas semánticas basadas en marcadores lingüísticos. Estos marcadores ayudan a señalar la presencia de una relación entre segmentos del texto, en nuestro caso, apoyan la señalización de relaciones entre individuos y valores de propiedades de dato. En la Tabla 1 se muestra un conjunto marcadores lingüísticos para estas relaciones y propiedades.

Tabla 1. Marcadores lingüísticos para la extracción de relaciones.

Relación	Marcadores lingüísticos
Solución	<i>en este artículo, nuestra aportación</i>
Propósito	<i>para + verbo infinitivo, con la finalidad de, con el fin de</i>
Evidencia	<i>ya que, claramente</i>

Los individuos sobre las publicaciones, sus relaciones y las propiedades que fueron extraídas a partir de los textos se representan en un modelo ontológico, como individuos, el cual se describe en la sección 4.3.

4.3. Poblado ontológico

En la tercera etapa se diseñó un modelo ontológico para representar la información extraída de los expedientes curriculares y resúmenes de publicaciones como *individuos de clase*, junto con sus propiedades de objeto y datos.

El modelo está codificado en el Lenguaje Ontológico para la Web (OWL 2)¹ y se encuentra constituido por las siguientes clases: *Publicación*, la cual representa la descripción semántica de un artículo científico con sus relaciones correspondientes; la clase *Método* que describe semánticamente la forma de solución del problema planteado en la publicación; la clase *Autor* que representa la persona o lista de personas que colaboran en la publicación; la clase *Editor* que se utiliza para relacionar la revista, memoria o actas donde se publica el trabajo científico. En la Figura 2 se muestra este modelo ontológico con sus clases, propiedades de objeto y datos.

¹ <http://www.w3.org/TR/owl2-overview/>

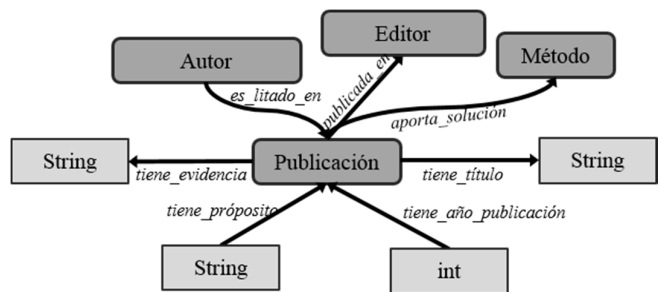


Fig. 2. Esquema ontológico para la representación semántica de publicaciones científicas.

El modelo ontológico también considera relaciones de tipo de dato, las cuales se utilizan para asignar valores a las propiedades de un individuo. De esta manera, en el modelo ontológico propuesto (Figura 2), también se consideran las siguientes propiedades: *tiene_título* y *tiene_año_publicación* que contienen el título y año de la publicación con sus valores respectivos; la relación de tipo de dato *tiene_propósito* que se encarga de relacionar la intención que existe detrás del método propuesto con las publicaciones científicas; finalmente la relación *tiene_evidencia* que contienen información destinada a aumentar la confianza del lector sobre el método y el propósito de la publicación.

La axiomatización del modelo ontológico se origina en la clase *Publicación*, en la cual los individuos extraídos son representados junto con las propiedades de objeto y datos. De esta manera, el modelo está compuesto por axiomas de clase y de relaciones, la Figura 3 muestra la instanciación de una publicación (*hasIndividual*) titulada: POBLADO_AUTOMÁTICO_DE_ONTOLOGÍAS_DE_PERFILES_ACADÉMICOS_A_PARTIR_DE_TEXTOS_EN_ESPAÑOL, la cual tiene una relación mediante la propiedad de objeto *aporta_solución* con el método REGLAS_SEMÁNTICAS-MARCADORES_LINGÜÍSTICOS y tiene como propósito al texto EXTRAER-REPRESENTAR_LAS_INSTANCIAS_EN_UN_MODELO_ONTOLÓGICO_DE_PUBLICACIONES_CIENTÍFICAS.

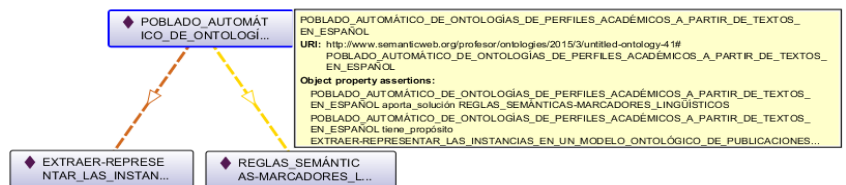


Fig. 3. Axiomatización de un individuo en la case *Publicación*.

En la Sección 5 se muestra la experimentación y los resultados obtenidos de la aplicación del enfoque con sus tres etapas, cuyo objetivo final es el poblado automático de ontologías a partir de textos.

5. Evaluación y resultados

La evaluación del enfoque propuesto fue realizada contra un conjunto *de individuos de clase, relaciones entre individuos* (propiedades de objeto) y valores de las *propiedades de dato*, las cuales fueron identificadas y representadas por expertos en el dominio científico-académico. Este conjunto incluye la identificación y representación manual de 568 individuos, los cuales se encuentran distribuidos de la siguiente manera: 337 individuos relacionados con la publicación y 231 individuos que están relacionados con el contenido del resumen de los artículos.

El procedimiento de evaluación consiste en comparar los elementos extraídos de manera automática con los individuos identificados y validados por los expertos. Las métricas de evaluación *precisión* y *exhaustividad* [33] fueron utilizadas con la finalidad de cuantificar los individuos identificados correctamente por el enfoque propuesto. En este contexto, se adaptan y definen las siguientes medidas: la *precisión* (P) que es el coeficiente entre el número de individuos relevantes identificados y el total de individuos extraídos, como se muestra en (1); y la *exhaustividad* (E) que es el coeficiente entre el número de individuos relevantes identificados y el número de individuos que deben ser extraídos, la cual se muestra en (2).

$$P = \frac{\# \text{ Individuos relevantes poblados}}{\# \text{ Total de individuos poblados}} \quad (1)$$

$$E = \frac{\# \text{ Individuos relevantes poblados}}{\# \text{ Individuos que deberían ser poblados}} \quad (2)$$

Una media armónica que combina los valores de *precisión* y *exhaustividad* llamada *medida F*, la cual se muestra en (3).

$$\text{medida } F = \frac{2 * P * E}{P + E} \quad (3)$$

La Tabla 2 muestra los resultados del enfoque propuesto en términos de precisión, exhaustividad y medida F para la tarea de poblado automático de los 568 individuos de clase y relaciones (propiedad de datos y relaciones entre individuos).

Tabla 2. Resultados del poblado de la ontología de publicaciones científicas con instancias de clase y relaciones entre individuos.

Tipo de individuo	Precisión	Exhaustividad	Medida F
Publicación	0.894	0.931	0.912
Resumen	0.794	0.887	0.837
Promedio	0.844	0.909	0.875

La Tabla 3 presenta los resultados de los enfoques y propuestas presentadas en [4, 20, 24, 27, 34]. Ellos presentan propuestas para el descubrimiento y representación de instancias de clases, propiedades y relaciones entre las instancias. Los resultados están

expuestos en términos de un promedio de la medida F de los tipos de axiomas ontológicos que cada uno descubre.

Tabla 3. Resultados del poblado ontológico con individuos de clase, propiedades y relaciones entre individuos.

	Medida F
[4]	0.833
[20]	0.873
[24]	0.818
[27]	0.589
[33]	0.765
Nuestra propuesta	0.875

Los resultados demuestran una efectividad de nuestro enfoque propuesto, el cual logra una ligera ventaja de 0.002 sobre el trabajo presentado en [20]. Cabe mencionar que nuestro enfoque utiliza texto libre para la extracción de relaciones entre individuos a partir de los resúmenes de artículos científicos. Sin embargo, aun cuando nuestro enfoque se basa en textos semiestructurados, sólo para la extracción de propiedades de las publicaciones, el método utilizado puede ayudar a expertos del dominio a representar información sobre publicaciones científicas.

Los resultados del poblado de ontología de publicaciones científicas han demostrado que las reglas somáticas logran extraer información concerniente a las publicaciones como título, editor donde se publica, la lista de autores listados y el año de publicación. Mientras tanto, los marcadores lingüísticos son capaces de señalar la información de solución, propósito y evidencia a partir de los resúmenes.

6. Conclusiones

En este artículo se ha presentado un enfoque basado en reglas semánticas y marcadores lingüísticos, con la finalidad de identificar y representar individuos de publicaciones científicas a partir de expedientes curriculares y resúmenes de artículos científicos en español.

El enfoque propuesto ha sido evaluado por un conjunto *gold standard* de 568 individuos que han sido identificados y representados por expertos. Los resultados han mostrado un mejor desempeño en el poblado de información propia de la publicación al poblar el 91.2% de los individuos. Sin embargo, en el poblado de individuos extraídos a partir del resumen de la publicación se logra un desempeño promedio del 83.7%. El bajo desempeño logrado en la precisión, apenas un 79.4%, del poblado de individuos que provienen del resumen se debe a problemas del lenguaje natural como la polisemia. Este fenómeno del lenguaje está presente en los marcadores lingüísticos utilizados para la extracción, por ello un tratamiento de este fenómeno es necesario y deja una oportunidad para un trabajo futuro.

Los resultados de la evaluación del enfoque propuesto son prometedores y mejores (por encima del promedio) que los obtenidos por los trabajos relacionados del estado del arte. Todos ellos enfocados en la extracción y representación (poblado) de individuos de clases, propiedades y relaciones en un modelo ontológico.

Las principales contribuciones de este trabajo son (1) un modelo de representación semántica de publicaciones científicas, (2) un conjunto de reglas basadas en marcadores lingüísticos para extracción de información sobre publicaciones científicas y (3) un enfoque de poblado automático de ontologías del dominio académico para el idioma español.

La información extraída y representada en el modelo ontológico resulta de gran utilidad para tareas como la recuperación de información en el dominio académico sobre autores, año y editor, aportación, evidencia y propósito de las publicaciones científicas. Como trabajo futuro, se puede utilizar esta información por un sistema de pregunta-respuesta para conceder información precisa a las peticiones de los usuarios.

Agradecimientos. Este trabajo ha sido financiado por el Programa para el Desarrollo Profesional Docente (PRODEP) con el número de proyecto UAM-A-PTC-34. Agradecemos el apoyo otorgado por la Universidad Autónoma Metropolitana Unidad Azcapotzalco y el SNI-CONACyT.

Referencias

1. Maedche, A.: *Ontology learning for the semantic web*. Springer Science & Business Media, Massachusetts, USA (2002)
2. Gruber, T. R.: *Toward principles for the design of ontologies used for knowledge sharing?*. *International journal of human-computer studies*, vol. 43(5), pp. 907–928 (1995)
3. Cimiano, P.: *Ontology Learning and Population from Text: Algorithms, Evaluation and Applications*. Springer, New York, USA (2006)
4. Faria, C., Serra, I., Girardi, R.: *A domain-independent process for automatic ontology population from text*. *Science of Computer Programming*, vol. 95, pp. 26–43 (2014)
5. Oliveira, H., Lima, R., Gomes, J., Ferreira, R., Freitas, F., Costa E.: *A confidence-weighted metric for unsupervised ontology population from web texts*. *Database and Expert Systems Applications*, pp. 176–190 (2012)
6. Shen, W., Wang, J., Lou, P., Wang, M.: *A graph-based approach for ontology population with named entities*. In: *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, pp. 345–354 (2012)
7. Buitelaar, P., Eigner, T.: *Topic Extraction from Scientific Literature for Competency Management*. In: *The 7th International Semantic Web Conference, Karlsruhe, Germany, (2008)*
8. Cui, G., Lu, Q., Li, W., Chen, Y.: *Automatic Acquisition of Attributes for Ontology Construction*. *Proceedings of the 22nd International Conference on Computer Processing of Oriental Languages, Language Technology for the Knowledge-based Economy, LNAI*, vol. 5459, pp. 248–259 (2009)
9. Blomqvist, E.: *Ontocase - A pattern-based ontology construction approach*. In: *Proceedings of On The Move to meaningful internet systems, ODBASE - The 6th International*

- Conference on Ontologies, DataBases, and Applications of Semantics, Trento, Italy, pp. 971–988 (2007)
10. Dahab, M. Y., Hassan, H. A., Rafea, A.: TextOntoEx: Automatic Ontology Construction from Natural English Text. *Expert System with Applications: An International Journal*, vol. 34(2), pp. 1474–1480 (2008)
 11. Draicchio, F., Gangemi, A., Presutti, V., Nuzzolese, A. G.: Fred: From natural language text to RDF and owl in one click. In: *The Semantic Web: ESWC 2013 Satellite Events Springer Berlin Heidelberg*, pp. 263–267 (2013)
 12. Wong, W.: Discovering Lightweight Ontologies using the Web. In *Proceedings of the 9th Postgraduate Electrical Engineering & Computing Symposium, Perth, Australia* (2008)
 13. Fortuna, B., Lavrac, N., Velardi, P.: Advancing Topic Ontology through Term Extraction. In: *Proceedings of the 10th Pacific Rim International Conference on Artificial Intelligence, Hanoi, Vietnam*, pp. 626–635 (2008)
 14. Cerbah, F.: Mining the Content of Relational Databases to Learn Ontologies with Deeper Taxonomies. In: *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, pp. 553–557 (2008)
 15. Celjuska, D., Vargas-Vera, M.: Ontosophie: A Semi-Automatic System for Ontology Population from Text. In: *Proceedings of the 3rd International Conference on Natural Language Processing* (2004)
 16. Waard, A., Buitelaar, P., Eigner, T.: Identifying the Epistemic Values of Discourse Segments in Biology Texts. *International Workshop on Computational Semantics, Eindhoven, Netherlands*, pp. 351–354 (2009)
 17. Morante, R., Van-Asch, V., Daelemans, W.: A memory-based learning approach to event extraction in biomedical texts. In: *Proceedings of the BioNLP 2009 Workshop Companion Volume for Shared Task*, pp. 59–67 (2009)
 18. Aussenac-Gilles, N., Despres, S., Szulman, S.: The TERMINAE method and platform for ontology engineering from texts. In *Bridging the Gap between Text and Knowledge-Selected Contributions to Ontology Learning and Population from Text*, pp. 199–223 (2008)
 19. Sánchez, S., Llorens, J., Morato, J., Hurtado, J.: Extracción Automática de Relaciones Semánticas. En: *Memorias de la 2da. Conferencia Iberoamericana en Sistemas, Cibernética e Informática*, pp. 265–268 (2005)
 20. Ruíz-Martínez, J.M.: Ontology Population: An Application For The E-Tourism Domain. *International Journal of Innovative Computing, Information and Control*, vol. 7 (11), pp. 6115–6134 (2011)
 21. Lacasta, J., Lopez-Pellicer, F. J., Florczyk, A., Zarazaga-Soria, F. J., Noguera-Iso, J.: Population of a spatio-temporal knowledge base for jurisdictional domains. *International Journal of Geographical Information Science*, vol. 28 (9), pp. 1964–1987 (2014)
 22. Valencia, R.: Un Entorno para la Extracción Incremental de Conocimiento desde Texto en Lenguaje Natural. Tesis doctoral de la Universidad de Murcia, España (2005)
 23. Llorens, H., Navarro, B., Saquete, E.: Detección de Expresiones Temporales TimeML en Catalán mediante Roles Semánticos y Redes Semánticas. *Revista de Procesamiento del Lenguaje Natural*, vol. 43, pp. 13–21 (2009)
 24. Faria, C., Girardi, R. Novais, P.: Using domain specific generated rules for automatic ontology population. In: *The International Conference on Intelligent Systems Design and Applications, Koshi, India*, pp. 297–302 (2012)
 25. Groza, T., Handschuh, S., Möller, K., Decker, S.: SALT-Semantically Annotated LaTeX for Scientific Publications. In: *The Semantic Web: Research and Applications*, pp. 518–532 (2007)

26. Aussenac-Gilles N., Jacques M. P.: Designing and Evaluating Patterns for Relation Acquisition from Texts with Caméléon. *Terminology, Pattern-Based approaches to Semantic Relations*, vol. 14 (1), pp. 45–73 (2008)
27. Vargas-Vera, M., Celjuska, D.: Event recognition on news stories and semi-automatic population of an ontology. In: *Proceedings of the 2004 IEEE/WIC/ACM International Conference on Web Intelligence*, pp. 615–618 (2004)
28. Aleman-Meza, B., Halaschek, C., Sheth, A. P., Arpinar, I. B., Sannapareddy, G.: SWETO: Large-scale semantic web test-bed. In: *16th International Conference on Software Engineering and Knowledge Engineering*, Banff, Canada (2004)
29. Kröll, M., Klampfl, S., Kern, R.: Towards a Marketplace for the Scientific Community: Accessing Knowledge from the Computer Science Domain. *D-Lib Magazine*, vol. 20(11), pp. 10 (2014)
30. Cunningham, H., Maynard, D., Bontcheva, K.: *Text processing with GATE*. Gateway Press, California, USA (2011)
31. Schmid, H.: Probabilistic part-of-speech tagging using decision trees. In: *Proceedings of the international conference on new methods in language processing*, vol. 12, pp. 44–49 (1994)
32. Cunningham, H., Maynard, D., Tablan, V.: *JAPE: a Java Annotation Patterns Engine (Second Edition)*. Technical Report, University of Sheffield, Department of Computer Science, United Kingdom (2000)
33. Baeza-Yates, R., Ribeiro-Neto, B.: *Modern information retrieval*, Association for Computing Machinery press, New York, USA (1999)
34. Navigli, R., Velardi, P.: From glossaries to ontologies: Extracting semantic structure from textual definitions. *Ontology Learning and Population: Bridging the Gap between Text and Knowledge*, pp. 71–87 (2008)