

## Método de análisis semántico basado en WordNet para la extracción de información en mapas conceptuales

Wenny Hojas-Mazo, Alfredo Simón-Cuevas, Manuel de la Iglesia-Campos

Universidad Tecnológica de La Habana José Antonio Echeverría, Cujae, La Habana, Cuba

{whojas, asimon, miglesia}@ceis.cujae.edu.cu

**Resumen.** El mapa conceptual es un tipo de representación de conocimiento basada en grafo, donde el conocimiento se expresa en lenguaje natural a través de conceptos y relaciones entre ellos. Este se puede construir manualmente o automáticamente a partir de un texto, propiciando la obtención de repositorios de conocimiento de gran valor para la gestión de conocimiento y el análisis de textos. En este trabajo, se propone un método de análisis semántico a ser aplicado en el procesamiento de consultas sobre un repositorio de mapas conceptuales. El método se basa en la extensión semántica de conceptos usando WordNet y un algoritmo de desambiguación, e incluye reglas que guían los procesos de búsqueda e integración de información. Mediante el desarrollo de un caso de estudio en el ámbito de la ingeniería ontológica se ejemplifica la ejecución del método y se muestran sus beneficios para extraer información en este tipo de repositorios.

**Palabras clave:** Mapas conceptuales, extracción de información, desambiguación, WordNet.

### Method of Semantic Analysis Based on WordNet for Information Extraction in Concept Maps

**Abstract.** Concept map is a graph-based knowledge representation, where the knowledge is expressed in natural language through concepts and relationships among them. This can be built manually or automatically from texts, propitiating the obtaining of knowledge repositories with great value for the knowledge management and the texts analysis. In this work, a method of semantic analysis to be applied in the processing of queries on a concept maps repository is proposed. The method is based on the semantic extension of concepts using WordNet and a disambiguation algorithm, and several rules for guiding the search and integration of information processes are included. Through the development of a study case in the context of ontology engineering the execution of the method was exemplified and its benefits for the extraction of information in this type of repositories were shown.

**Keywords.** Concept maps, information extraction, disambiguation, WordNet.

## 1. Introducción

Los Mapas Conceptuales (MC) constituyen simultáneamente, un método para captar lo más significativo de un tema y un recurso esquemático para representar un conjunto de significados conceptuales mediante una estructura de proposiciones [1]. En los MC el conocimiento es expresado en lenguaje natural y estructurado en forma de grafo, donde los nodos representan conceptos y estos se relacionan mediante arcos dirigidos y etiquetados por una frase de enlace formando proposiciones. El MC surgen en el área de la pedagogía, pero su uso se ha extendido a otras áreas tales como: la gestión del conocimiento [2], la ingeniería ontológica [2, 3], y el análisis de textos [4].

Los MC generalmente son construidos manualmente, con la asistencia de herramientas como *CmapTools* [5], pero también se reportan propuestas para generarlos de manera automática a partir de textos [6]. Estas herramientas y métodos propician la creación de repositorios de MC (RMC), los que son reconocidos como modelos de conocimiento [5], cuando el conocimiento representado está asociado a un dominio específico. El conocimiento almacenado en los RMC puede resultar de gran valor en los diferentes contextos de aplicación de los MC, por lo que el desarrollo de soluciones que propicien el incremento de su aprovechamiento constituye un aspecto de gran interés en este ámbito. En este sentido, mejorar los mecanismos de consulta sobre este tipo de base de conocimiento es una de las metas a alcanzar para lograr ese propósito. La mayoría de las propuestas actuales que reportan mecanismos de consultas sobre RMC se centran en obtener conceptos, proposiciones, MC o estructuras proposicionales frecuentes [4, 5, 7, 8, 9, 10]. Por otra parte, en [2] se reporta CMQL (*Concept Maps Query Language*) como una propuesta más abarcadora para consultar un RMC. En CMQL se formalizan un conjunto de operaciones de consultas que facilitan obtener diferentes vistas del conocimiento almacenado en un RMC, e incluye mecanismos de integración de información como parte del procesamiento interno de cada consulta y una de sus bondades. En estas propuestas la búsqueda de información en el RMC y el procesamiento interno del conocimiento representado se lleva a cabo mediante el análisis de los conceptos a nivel sintáctico, y no a nivel semántico, donde se tratan las posibles ambigüedades que puedan existir en ellos. Esto constituye una limitación para lograr un mayor aprovechamiento del conocimiento almacenado en un RMC, ya que puede provocar, por ejemplo, que no se obtenga información potencialmente útil sobre un determinado concepto debido a que no se encuentra explícitamente en el RMC, existiendo posibles conceptos que sean sinónimos. Todo lo anterior conduce a la necesidad de incorporar a los mecanismos de consultas sobre RMC soluciones que posibiliten realizar análisis semántico a nivel de conceptos, teniendo en cuenta como un elemento clave la reducción de ambigüedades. La presente investigación aborda esta problemática.

En este trabajo, se propone un método de análisis semántico dirigido a mejorar los resultados de la extracción de información a partir de consultas sobre un RMC. El método está basado en un mecanismo de extensión semántica de los conceptos del RMC a partir de *WordNet* [11], así como en la definición de un conjunto de reglas que guían la búsqueda de información e integración de información en el RMC. La extensión semántica de términos ha sido una técnica muy tratada en el ámbito de la recuperación de información para mejorar los resultados en ese proceso, pero generalmente se ha

aplicado en los términos de la consulta [12]. Esta técnica se aplica en el método propuesto sobre los conceptos incluidos en el RMC, ya que de esta manera también se contribuye a mejorar los resultados de la integración de información que se ejecuta como parte del procesamiento de las consultas. El método incluye además la aplicación de un algoritmo de desambiguación, con el objetivo de resolver las ambigüedades existentes en los conceptos y al mismo tiempo hacer más efectivo el uso de *WordNet*; siendo esta una de sus contribuciones respecto al estado del arte. La desambiguación del sentido de las palabras ha sido un tema muy abordado en ámbito de los textos, pero no así en el contexto de los MC donde se reportan pocas soluciones [13, 14], las cuales presentan algunas limitaciones. En este caso, se aplica una variante del algoritmo reportado en [14], en la cual las heurísticas se utilizan de forma combinada y no secuencialmente, con lo que se mejoran los resultados respecto a [13, 14]. El método propuesto ha sido implementado en CMQL, teniendo en cuenta el alcance de sus prestaciones para extraer información en un RMC. Mediante el desarrollo de un caso de estudio, enmarcado en el área de la ingeniería ontológica, se ejemplifica la ejecución del método, y se muestra su utilidad en la extracción de información sobre la conceptualización de una ontología.

El resto del trabajo se organiza según se describe a continuación. En la sección 2 se analizan los trabajos relacionados con la problemática tratada. En la sección 3 se presenta y describe el método de análisis semántico propuesto. En la sección 4 se describe el desarrollo de la aplicación de la propuesta en un caso de estudio y se analizan los resultados. Las conclusiones y líneas de trabajo futuro se exponen en la sección 5.

## 2. Trabajos relacionados

La problemática de la extracción de información en RMC se ha enfocado en la extracción de conceptos [7], proposiciones [5] o MC [8, 9], mediante procesos de búsquedas ejecutados a partir consultas definidas sobre uno o varios conceptos. En [10] se explora la viabilidad del análisis de MC a partir de la extracción de conceptos y submapas frecuentes en un RMC; reportándose algo similar en [4]. En [2] se propone el lenguaje de consulta CMQL, en el que se formalizan un conjunto de operaciones de consulta, combinando elementos de la teoría de grafos y de conjuntos, para extraer información del RMC desde diferentes perspectivas.

En esta propuesta se formalizan cuatro tipos de consultas: *unión*, *intersección*, *submap* (o *proyección*) y *extensión* [2]. En estas consultas es necesario especificar un espacio de búsqueda, conformado por MC incluidos en el RMC, y en el caso de las dos últimas, también se deben especificar conceptos de interés que guiarán la búsqueda de información a extraer. En el procesamiento interno de cada consulta se incluye un mecanismo que posibilita integrar la información contenida en el espacio de búsqueda, a partir de la identificación de conceptos que puedan ser unificados. El resultado de cada consulta de CMQL se representa en forma de MC, y este puede ser almacenado o no dentro del RMC como un nuevo MC.

Los procesos de búsqueda de conceptos en RMC que se reportan en las propuestas estudiadas, se basan fundamentalmente en la identificación conceptos incluidos en el

RMC que sean sintácticamente equivalentes a los definidos en la consulta. En el proceso de integración de información propuesto en CMQL, para determinar si dos conceptos son unificados, también se aplica este tipo de análisis sintáctico. En estas propuestas no se tiene en cuenta la posible ambigüedad presente en los conceptos incluidos en el RMC. Esto constituye una limitación para la extracción de información en RMC, ya que reduce las posibilidades de obtener información asociada a conceptos que no son sintácticamente equivalentes pero que son semánticamente similares, con respecto a los incluidos en la consulta. También el no tratar esa ambigüedad puede conducir a resultados no apropiados en el proceso de integración de información, por ejemplo: la integración de conceptos con diferentes significados. La aplicación de algoritmos de desambiguación sobre los conceptos representados en un MC puede ser considerada como parte importante de la solución a esta problemática.

La desambiguación del sentido de las palabras, conocido como *WSD (Word Sense Disambiguation)* es una problemática ampliamente abordada en el ámbito de los textos. Sin embargo, son pocas las soluciones reportadas a este problema en MC [13, 14] y no se ha reportado la aplicación en MC de un algoritmo de desambiguación diseñado para textos, debido fundamentalmente a las diferencias estructurales existentes entre ellos. Un MC suele constituir un resumen de un tema determinado, aunque puede representar conceptos de diferentes dominios, lo que dificulta el proceso de desambiguación, ya sea por la poca información contextual que se puede obtener sobre una palabra o concepto, o por la diversidad de dominios representados a través de sus conceptos. Un aspecto favorable es que las relaciones entre los conceptos están explícitamente representadas, lo que no ocurre en el caso de los textos y ha sido un aspecto muy bien aprovechado en las soluciones reportadas [13, 14].

En [13] se explota la topología del MC para determinar el *synset* en *WordNet* [11] que desambigua una palabra dentro de un concepto, mediante un análisis de similitud entre el contexto en *WordNet* de cada uno de los *synset* posibles (considerando solo relaciones de *hiperonimia*) y el contexto en el que se encuentra la palabra en el MC. En [14] también se incluye el análisis contextual, como una de las heurísticas para determinar el sentido de los conceptos, pero considerando también relaciones de *hiponimia*, *meronimia/holonimia*. También se incluyen los análisis de dominio (en correspondencia con [15]) y glosa como otras heurísticas. En esta última propuesta, el sentido se determina a partir del resultado obtenido por cada heurística, en un proceso de ejecución secuencial [14], sin considerar los beneficios que puede proporcionar la combinación de los resultado obtenidos por cada heurística.

### 3. Método de análisis semántico

El método incluye un mecanismo de extensión semántica de los conceptos representados en un RMC, y la definición de un conjunto de reglas que guían los procesos de búsqueda e integración de información que se llevan a cabo en el procesamiento de las consultas, específicamente de CMQL. La extensión semántica ha sido concebida con el uso de *WordNet*, y soportada en la aplicación de un algoritmo de desambiguación de conceptos inspirado en [14]. Las reglas definidas utilizan la información semántica con

la que han sido extendidos los conceptos del RMC para aumentar las capacidades del mecanismo de consulta en la identificación de información relevante a extraer, y al mismo tiempo mejorar la efectividad en el proceso de integración de información. El método puede ser aplicado para el procesamiento de MC en idioma inglés y español, requiriéndose solo el uso de la versión de *WordNet* correspondiente al idioma empleado en los MC. El RMC que se utilice como punto de partida para la ejecución del método puede incluir MC elaborados manualmente y/o generados automáticamente a partir de textos.

En la Figura 1 se esquematiza el flujo de trabajo del método propuesto instrumentado en CMQL, siendo  $Q$  una consulta definida formalmente por la tripleta  $(I, E, t)$ , donde  $I$  es el conjunto de conceptos sobre los que interesa extraer información,  $E$  el conjunto de MC del RMC que conforman el espacio de búsqueda,  $t$  el tipo de consulta ( $u$ : unión,  $i$ : intersección,  $p$ : proyección,  $e$ : extensión),  $E_{ext}$  es el conjunto de MC de  $E$  cuyos conceptos se han extendido semánticamente y  $E_{ext}^*$  es un MC generado automáticamente en el que se integra el conocimiento representado en los MC incluidos en  $E$ .  $E_{ext}$  y  $E_{ext}^*$  constituyen representaciones intermedias resultantes del pre-procesamiento que se realiza sobre el espacio de búsqueda.

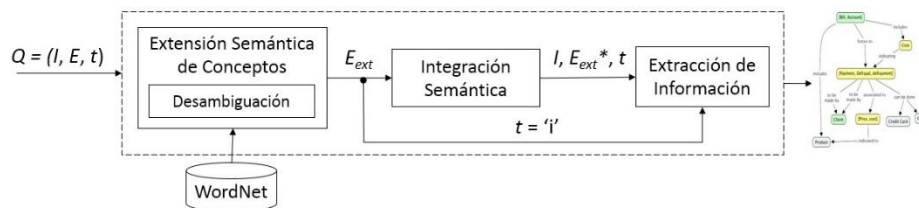


Fig. 1. Flujo de trabajo del método de análisis semántico propuesto.

### 3.1. Extensión semántica de conceptos

La extensión semántica de conceptos se define como el proceso de asociarle a un concepto información referente a otros términos, a partir de la identificación de relaciones de sinonimia entre ellos. En este proceso *WordNet* es utilizado como base de datos de sentidos y el mismo se aplica a todos los conceptos representados en los MC del RMC que conforman el espacio de búsqueda de la consulta. Inicialmente, se identifican y recuperan de *WordNet* todos los *synset* en los cuales están presentes los conceptos a ser extendidos. A partir de esta información los conceptos son clasificados en: ambiguos (aparece en más de un *synset*), no ambiguo (aparece en un solo *synset*) y desconocidos (no se identifican *synset* en los que está presente).

Seguidamente, se ejecuta un algoritmo de desambiguación para determinar el sentido más apropiado para cada concepto ambiguo, según el contexto en el que está siendo usado dentro del MC. En este caso, se decidió utilizar una versión mejorada del algoritmo reportado en [14], donde el sentido del concepto se determina mediante la combinación de los resultados obtenidos por cada una de las heurísticas usando una votación ponderada [16]; en correspondencia con estrategias propuestas en [17]. En experimentos parciales realizados se comprobó que esta nueva versión mejora los resultados

de [13, 14], obteniéndose valores de precisión y cobertura superiores al 80% y al 90%, respectivamente, tanto en MC en español, como en inglés. Luego de ejecutado este algoritmo, se actualiza la lista de conceptos ambiguos y no ambiguos, ya que es posible que no todos los conceptos ambiguos puedan ser desambiguados. Los conceptos no desambiguados tendrían asociados más de un *synset*, en correspondencia con el resultado obtenido por el algoritmo de desambiguación. Al final, todos los conceptos presentes en *WordNet* son extendidos información asociada a los *synset* identificados en el proceso de desambiguación. Esta información consta de: identificador del *synset* y la lista de sinónimos.

### 3.2. Integración semántica del espacio de búsqueda

En este proceso se utiliza la información resultante de la extensión de conceptos para identificar posibles estructuras proposicionales que puedan ser integradas a partir de la unificación de conceptos representados en diferentes MC del espacio de búsqueda y se ejecuta como parte del procesamiento interno de cada una de las consultas de CMQL. La unificación de dos conceptos se lleva a cabo básicamente a partir de la identificación de una relación de sinonimia entre ellos y siguiendo un conjunto de reglas definidas para este propósito, las que a continuación se presentan.

Sean,

$CNA$ : el conjunto de conceptos no ambiguos;

$CA$ : el conjunto de conceptos ambiguos;

$S(c)$ : el conjunto de *synset* ( $s$ ) asociados a un concepto  $c_i$ ;

$c_1$  y  $c_2$ : dos conceptos incluidos en diferentes MC del espacio de búsqueda;

Entonces  $c_1$  y  $c_2$  son unificados si:

R1:  $(c_1, c_2 \in CNA) \wedge (S(c_1) = S(c_2))$ ;  $o$

R2:  $(c_1, c_2 \in CA) \wedge (\exists s'/s' \in S(c_1) \wedge s' \in S(c_2))$ ;  $o$

R3:  $((c_1 \in CNA \wedge c_2 \in CA) \vee (c_1 \in CA \wedge c_2 \in CNA)) \wedge (\exists s'/s' \in S(c_1) \wedge s' \in S(c_2))$ .

La etiqueta correspondiente al concepto unificado que se mostrará como resultado de la consulta se construye concatenando las etiquetas de los conceptos originales, separadas cada una por una coma ',' y encerrándolas entre corchetes '[' ]' en caso de ser más de una. Adicionalmente, se determina el *synset* que finalmente queda asociado a ese concepto unificado mediante las siguientes reglas:

R4: Si se dispara R1, entonces se asocia el mismo *synset* de los conceptos;

R5: Si se dispara R2, entonces se asocia el *synset* común a ambos conceptos;

R6: Si se dispara R3, entonces se asocia el *synset* correspondiente al  $c_i \in CNA$ .

### 3.3. Extracción de información

Este proceso se centra fundamentalmente en la evaluación de la consulta  $Q$ , teniendo en cuenta su tipo y los elementos que la definen. En la especificación de la consulta *unión* no se definen conceptos de interés (por tanto,  $I = \{\}$ ), y se obtiene como resultado un nuevo MC que representa una vista integrada de todos los MC que componen en el espacio de búsqueda, obteniéndose  $E_{ext}^*$ . Este tipo de consulta se puede ejecutar de

manera directa por un usuario, pero también es usada como paso intermedio en la ejecución de otras consultas. En la consulta *intersección* tampoco son definidos conceptos de interés, y tiene el propósito de identificar y extraer aquellos conceptos y proposiciones que son comunes a todos o a un subconjunto de los MC incluidos en  $E_{ext}$ .

Las consultas *proyección* y *extensión*, están concebidas para extraer del RMC estructuras proposiciones que muestren información relacionada con los conceptos de interés de la consulta ( $I$ ). Por tanto, su procesamiento resulta ser más compleja porque en el proceso búsqueda hay que determinar a partir de qué conceptos (de los incluidos en  $E_{ext}$ ) es que se extrae la información requerida. La identificación de los conceptos de interés en  $E_{ext}$  se lleva a cabo mediante las reglas que a continuación se presentan, las cuales son comprobadas en el mismo orden en el que han sido numeradas.

Sea  $a$  un concepto incluido en la consulta tal que  $a \in I$ ,  $b$  un concepto representado en algún  $mc_i$  tal que  $mc_i \in E_{ext}$ ,  $Pl(c_i)$  la lista de palabras que componen la etiqueta de un concepto  $c_i$ , y  $si(c_i)$  el conjunto de sinónimos asociados a  $c_i$ .

Entonces,  $b$  es identificado como concepto de interés si:

R1:  $a$  y  $b$  son sintácticamente equivalentes; o

R2:  $a \in Pl(b)$ ; o

R3:  $a \in si(b)$ ;

#### 4. Aplicación en caso de estudio

El caso de estudio se enmarca en el área de la ingeniería ontológica, en la cual se investiga sobre principios, métodos, y herramientas para la construcción y mantenimiento de ontologías [18]. La obtención de una descripción informal o conceptualización del conocimiento que se quiere formalizar en la ontología, usualmente representada mediante conceptos y relaciones entre ellos, constituye una tarea común en las etapas tempranas de la mayoría de las metodologías existentes para la construcción de ontologías.

En este contexto, la utilidad de los MC para capturar y representar ese tipo de conceptualizaciones ha sido reconocida por varios autores [2, 3]. El MC facilita la captura del conocimiento que poseen los expertos del dominio, por la flexibilidad que brindan para representar el conocimiento y porque constituyen una herramienta muy intuitiva para las personas. El MC también ayuda al ingeniero de conocimiento a identificar los conceptos más significativos del dominio y los diferentes tipos de relaciones que se pueden establecer entre ellos en ese contexto.

En las Figuras 2 y 3 se muestran dos MC en idioma inglés, nombrados 'Cuenta' y 'Pago' respectivamente, elaborados manualmente usando *CmapTools* y que constituyen fragmentos de una conceptualización del dominio de gestión hotelera, construida como resultado de la captura de conocimiento de expertos. La obtención de esta conceptualización constituyó uno de los primeros pasos llevados a cabo en la construcción de una ontología terminológica que sería empleada para la extensión semántica de consultas en un sistema de recuperación de información. La ontología no solo debía incluir conceptos asociados a ese dominio específico, los cuales se representarían como clases, sino también debía incluir sinónimos asociados a dichos conceptos.

Los objetivos del desarrollo de este caso de estudio están dirigidos a ejemplificar la aplicación del método propuesto y a mostrar su utilidad para la extracción de información sobre los MC que se muestran en las Figuras 2 y 3, como parte del análisis que puede realizar el ingeniero de conocimiento para la obtención de la ontología. En este sentido, se describe el procesamiento que se lleva a cabo como parte de la ejecución de consultas que responden a los siguientes requisitos informacionales del ingeniero de conocimiento:

1. Obtener una vista global de la conceptualización capturada;
2. Obtener información sobre los conceptos específicos 'defrayment' y 'cost'.

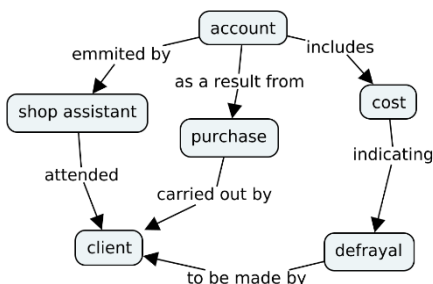


Fig. 2. Mapa conceptual 'Cuenta'.

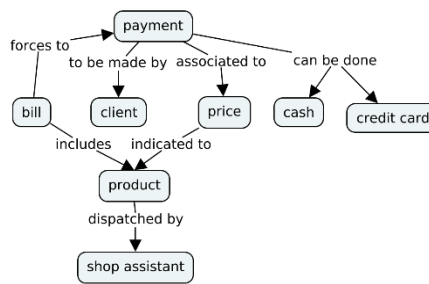


Fig. 3. Mapa conceptual 'Pago'.

A través de CMQL, el primer requisito se puede satisfacer mediante la ejecución de la consulta *unión*, cuya formalización sería  $UM(\{Cuenta, Pago\})$ , y el segundo requisito se puede satisfacer ejecutando una consulta *proyección*, cuya formalización sería  $SM^{+1}(\{Cueta, Pago\}, \{'defrayment', 'cost'\})$ , según [2].

El primer paso que se ejecuta en cada consulta, como parte del método, es la extensión semántica de los conceptos representados en los MC 'Cuenta' y 'Pago'. En este proceso inicialmente se identifican los *synset* en *WordNet* asociados a cada concepto y luego se trata de eliminar las ambigüedades con el algoritmo de desambiguación propuesto, cuyos resultados se muestran en las Tablas 1 y 2. Se puede apreciar en estas tablas el alto grado de ambigüedad de los conceptos, teniendo en cuenta que como promedio la cantidad de *synsets* de *WordNet* asociados a dichos conceptos en resultó ser de 5 *synsets* en ambos MC.

Tabla 1. Resultados del algoritmo de desambiguación en 'Cuenta'.

Conceptos	Synsets	Sentidos identificados
account	13	06516955 - Economy - {bill, <b>account</b> , invoice}
purchase	5	00079018 - Economy - { <b>purchase</b> }
shop assistant	1	10548227 - Commerce - {salesclerk, shop_clerk, clerk, <b>shop_assistant</b> }
client	3	09984659 - Commerce - {customer, <b>client</b> }
cost	5	05163807 - Factotum - {price, <b>cost</b> , toll}
defrayal	3	01120448 - Economy - {payment, <b>defrayal</b> , defrayment}



La cobertura del algoritmo de desambiguación en ambos MC fue del 100 %, e igual resultado de precisión se obtuvo en ‘Pago’. En el caso de ‘Cuenta’ la precisión fue del 83 % ya que uno de los conceptos no se logró desambiguar satisfactoriamente, específicamente el concepto ‘cost’. Estos resultados se pueden considerar prometedores, teniendo en cuenta que los MC son pequeños, por tanto, con poca información contextual a tener en cuenta en el análisis, y con alto grado de ambigüedad de los conceptos representados.

Tabla 2. Resultados del algoritmo de desambiguación en ‘Pago’.

Conceptos	Synsets	Sentidos identificados
payment	3	01120448 - Economy - { <b>payment</b> , defrayal, defrayment}
bill	13	06516955 - Economy - { <b>bill</b> , account, invoice}
credit card	1	13376012 - Banking - { <b>credit_card</b> , charge_card, charge_plate, plastic}
cash	4	13386614 - Money - { <b>cash</b> , hard_cash, hard_currency}
price	9	05145118 - Money - {monetary_value, <b>price</b> , cost}
client	3	09984659 - Commerce - {customer, <b>client</b> }
shop assistant	1	10548227 - Commerce - {salesclerk, shop_clerk, clerk, <b>shop_assistant</b> }
product	6	03748886 - Commerce - {merchandise, ware, <b>product</b> }

Finalmente, como resultado de la extensión semántica cada uno de los conceptos representados en los dos MC fueron extendidos con el identificador del *synset* identificados y los términos (sinónimos) incluidos en ellos, y que se muestran en las tablas anteriores. Luego de obtener los MC extendidos ( $E_{ext}$ ), se lleva a cabo el proceso de integración semánticas, como parte también de la ejecución cada consulta. Con el objetivo de mostrar y analizar la contribución del método para la extracción de información sobre los MC seleccionados, se decidió ejecutar cada consulta siguiendo dos variantes: (1) sin usar el método de análisis semántico y (2) usando el método análisis semántico. Los resultados de la consulta *unión* (en sus dos variantes) se muestran en las Figuras 4 y 5, y los de la consulta *proyección* se muestran en las Figuras 6 y 7. En las Figuras 5 y 7 también se pueden apreciar resultados de la unificación de conceptos.

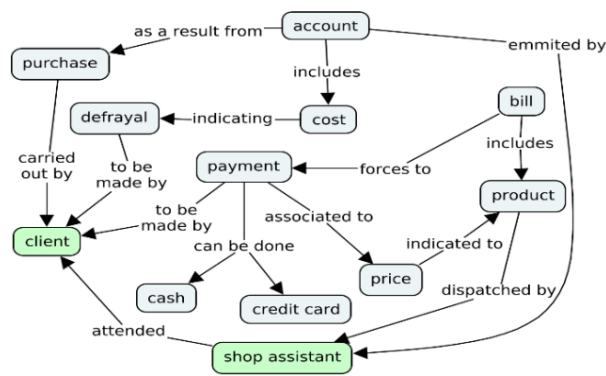


Fig. 4. Resultado de la *unión* sin usar el método.

En la Figura 4 se aprecia que ambos MC se integraron solo a partir de los conceptos 'client' y 'shop assistance', quedando representados como nodos diferentes los conceptos 'account' y 'bill' que tienen el mismo sentido. Esto se resuelve con la aplicación del método propuesto, como se muestra en la Figura 5. En esa figura se puede apreciar como son unificados aquellos conceptos con igual significado, según los resultados de la desambiguación, tal es el caso de: [defrayal, payment], que se corresponden a conceptos sintácticamente diferentes, así como 'client' y 'shop assistance', que se corresponden a conceptos sintácticamente equivalentes.

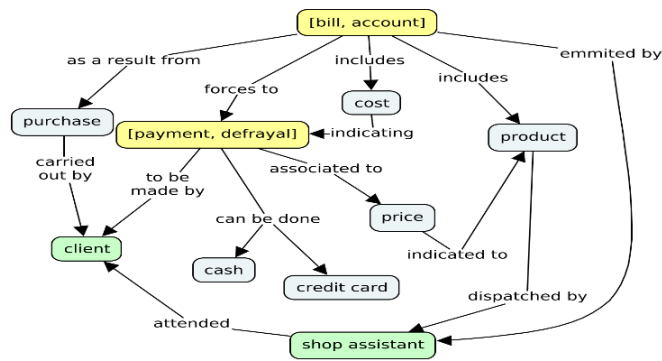


Fig. 5. Resultado de la consulta unión usando el método.

Este resultado muestra que con la aplicación del método se alcanza una mayor precisión con respecto a la información extraída del RMC, respecto a propuestas que no tratan la semántica. Además, también aporta beneficios para el ingeniero de conocimiento, a la hora de llevar a cabo la etapa de formalización de la ontología. Por ejemplo, el resultado mostrado en la Figura 5 facilita identificar grupos de conceptos semánticamente similares (conceptos unificados) que deben ser codificados como una única clase en la ontología, y suministra información sobre posibles sinónimos de dicha clase.

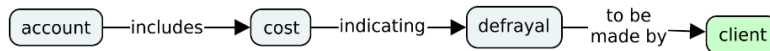


Fig. 6. Resultado de la proyección sobre los conceptos 'defrayment' y 'cost' sin usar el método.

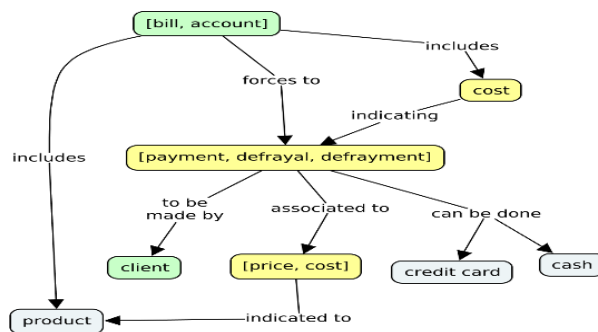


Fig. 7. Resultado de la proyección sobre los conceptos 'defrayment' y 'cost' usando el método.

En la Figura 7 se puede apreciar como aumenta la cantidad de información (conceptos y proposiciones) extraída y vinculada a los conceptos de la consulta al aplicar el método propuesto, respecto a lo mostrado en la Figura 6; siendo esta una de sus contribuciones. A partir de la aplicación del método fue posible extraer información sobre el concepto *'defrayment'*, no estando representado explícitamente en ninguno de los dos MC, ya que se identificó que existían conceptos que tenían una relación de sinonimia con ese término. Por otra parte, este resultado también es de utilidad para el ingeniero de conocimiento, ya que mediante la consulta realizada puede identificar si un nuevo término del dominio, en este caso *'defrayment'*, ya se encuentra representado en la conceptualización, y a partir de ello tomar decisiones sobre como incluirlo en la ontología, ya sea como una nueva clase o como un sinónimo de una clase existente.

## 5. Conclusiones y trabajo futuro

El análisis de la semántica asociada a los conceptos representados en un MC es esencial para su procesamiento computacional, ya que esta se encuentra implícita. En el trabajo se ha presentado un método que posibilita dotar a los mecanismos de consulta sobre RMC, y en particular a CMQL, de capacidades para realizar este tipo de análisis, sobre la base de asociar a los conceptos incluidos en el espacio de búsqueda información semántica recuperada de *WordNet*. Se aplica un algoritmo de desambiguación para reducir las ambigüedades que puedan tener algunos de esos conceptos. La información semántica asociada a los conceptos se utiliza en la definición de un conjunto de reglas que guían los procesos de búsqueda e integración de información que se ejecutan como parte de una consulta. La combinación de estos elementos permitió incrementar las capacidades de las consultas de CMQL para encontrar información potencialmente útil en el RMC, y contribuyó a mejorar la precisión en la integración de la información dentro del RMC. El desarrollo del caso de estudio posibilitó ejemplificar el funcionamiento del método, y sus resultados evidenciaron los beneficios que ofrece para la extracción de información en un RMC, específicamente en el análisis y procesamiento de la conceptualización que sirve de base para la construcción de una ontología. Se comprobó que el método posibilita incrementar de la cantidad de información a obtener a partir de una consulta y la efectividad en la integración de información.

En el futuro, se mejorarán las propuestas de extensión semántica de conceptos e integración de información definidas, aportando una solución para cuando los conceptos no estén incluidos en *WordNet*, lo que suele ocurrir cuando representan nombres de entidades o son de dominio específico. También se trabajará en el diseño de un entorno de evaluación experimental, actualmente no identificado en la bibliografía consultada, que posibiliten medir de una manera más objetiva los resultados que se obtienen con el método, y establecer líneas de comparación con propuestas similares.

## Referencias

1. Novak, J. D., y Gowin, D. B.: *Learning How to Learn*. Cambridge University Press (1984)

2. Liu, S. H., y Lee, G. G.: Using a concept map knowledge management system to enhance the learning of biology, *Comput. Educ.* 68, 105–116 (2013)
3. Simón, A., L. Ceccaroni, L., Rosete, A., Suárez, A., y Victoria, R.: A support to formalize a conceptualization from a concept maps repository. In: *Proc. of CMC'08*, 68–75 (2008)
4. Rizzi, R., y Parente de Oliveira, J. M.: Concept maps as the first step in an ontology construction method. *Information Systems*, 38, 771–783 (2013)
5. Rodríguez, L., Hojas, W., y Simón, A.: Método para la identificación de submapas frecuentes en modelos de conocimiento. *CICCI'16*, La Habana, Cuba (2016)
6. Cañas, A. J., Hill, G., Carff, R., Niranjana, S., Lott, J., Eskridge, T. C., Gómez, G., Arroyo, M., y Carvajal, R.: Cmaptools: a knowledge modeling and sharing environment. In: *Proc. of CMC'04*, 1, 125–13 (2004)
7. Zubrinic, K., Kalpic, D., y Milicevic, M.: The automatic creation of concept maps from documents written using morphologically rich languages. *Expert Systems with Applications*, 39 (16), 12709–12718 (2012)
8. Leake, D., Maguitman, A., Reichherzer, T., Cañas, A. J., Carvalho, M., Arguedas, M., y Eskridge, T.: Googling from a concept map: Towards automatic concept map-based query formation. In: *Proc. of CMC'04*, 1, 409–416 (2004)
9. Cañas, A. J., Leake, D. B., y Maguitman, A. G.: Combining Concept Mapping with CBR: Towards Experience-Based Support for Knowledge Modeling. In: *Proc. of FLAIRS Conference*, AAAI Press, 286–290 (2001)
10. Eskridge, T. C., Granados, A., y Cañas, A. J.: Ranking concept map retrieval in the Cmap-Tools network. In: *Proc. of CMC'06*, 1, 7–484 (2006)
11. Yoo, J. S., y Cho, M.-H.: Mining Concept Maps to Understand University Students' Learning. *International Educational Data Mining Society*, ERIC, 1, 19–21 (2012)
12. Miller, G., y Fellbaum, C.: *WordNet: An Electronic Lexical Database*. The MIT Press: Cambridge, MA (1998)
13. Carpineto, C., y Romano, G. A.: Survey of Automatic Query Expansion in Information Retrieval. *ACM Comput. Surv.* 1 (44), 1–50 (2012)
14. Cañas, A., Valerio, A., Lalinde, J., Carvalho, M., y Arguedas M.: Using WordNet for Word Sense Disambiguation to Support Concept Map Construction. *LNCS*, Springer, 2857, 350–359 (2003)
15. Simón, A., Ceccaroni, L., Rosete, A., Suárez, A., y de la Iglesia, M.: A concept sense disambiguation algorithm for concept maps. In: *In Proc. of CMC'08*, 14–21 (2008)
16. Bentivogli, L.; Forner, P.; Magnini, B.; y Pianta, E.: Revising WordNet Domains Hierarchy: Semantics, Coverage, and Balancing. In: *Proc. of COLING 2004 Workshop on Multilingual Linguistic Resources*, 101–108 (2004)
17. Klein, D., Toutanova, K., Ilhan, H., Kamvar, S., y Manning, C.: Combining heterogeneous classifiers for word-sense disambiguation. In: *Proc. of the ACL-02*. 8, 74–80 (2002)
18. Navigli, R.: Word Sense Disambiguation: A Survey. *ACM Computing Surveys*, 41(2), 1–69 (2009)
19. Sure, Y., Staab, S., Studer, R.: *Ontology Engineering Methodology*. In: Staab S., Studer R. (Eds.), *Handbook on Ontologies*, Springer-Verlag (2009)