

Estimación de la similitud semántica de tareas entre procesos de negocio de telecomunicaciones

Semantic similarity estimation of tasks between telecommunications business processes

Cálculo da similaridade semântica de tarefas entre processos de negócio de telecomunicações

Leandro Ordóñez Ante¹, Adriana X. Bastidas Narváez² y Juan Carlos Corrales³

Recepción: 22-may-2011/Modificación: 26-abr-2012/Aceptación: 30-abr-2012
Se aceptan comentarios y/o discusiones al artículo

Resumen

Como una medida para mejorar la competitividad en el mercado de las telecomunicaciones, actualmente las empresas del sector crean nuevos servicios de valor agregado, con el fin de ampliar su portafolio de servicios y garantizar bien sea la permanencia de sus clientes o ampliar el número de suscriptores. Estos nuevos servicios deben estar soportados en los procesos de negocio definidos por el Operador Telecomunicaciones, los cuales están compuestos por tareas de operación, gestión, mantenimiento y soporte. Habitualmente los Arquitectos Telco reutilizan dichas tareas con el objetivo de optimizar los recursos de la empresa y garantizar la rápida recuperación de la inversión, amortizando

¹ Magíster(C) en Ingeniería Telemática, leandro@unicauca.edu.co, investigador, Grupo de Ingeniería Telemática Universidad del Cauca, Popayán Colombia.

² Ingeniera en Electrónica y Telecomunicaciones, abastidas@unicauca.edu.co, investigadora, Grupo de Ingeniería Telemática Universidad del Cauca, Popayán Colombia.

³ Doctor en Ciencias de la Computación, jcorral@unicauca.edu.co, profesor titular, Coordinador Grupo de Ingeniería Telemática Universidad del Cauca, Popayán Colombia.

en el menor tiempo posible los gastos de creación y despliegue del nuevo servicio. La reutilización de tareas Telco presenta limitaciones en cuanto a la agilidad en la selección, ya que normalmente existen cientos de tareas y se requiere de la intervención de personal técnico que lleve a cabo las funciones de recuperación, a partir de la interpretación subjetiva del proceso de negocio a implementar. Existen diferentes enfoques que pretenden automatizar la selección de recursos pero generalmente están centrados en el emparejamiento semántico de los conceptos que describen sus interfaces de acceso (entradas y salidas); sin embargo, se ha demostrado que en la aplicación de estas técnicas es omitida información relevante contenida en otros atributos, tales como los identificadores. Por esta razón, el presente trabajo propone un mecanismo para determinar la similitud semántica entre tareas que conforman procesos de negocio de telecomunicaciones. Dicho mecanismo considera dos perspectivas: la inferencia sobre la funcionalidad de las tareas especificada en los identificadores y el análisis de cobertura de sus entradas y salidas.

Palabras claves: procesos de negocio, telecomunicaciones, tareas, emparejamiento semántico, similitud semántica, ontología.

Abstract

As a measure to improve competitiveness in the telecommunications market, currently companies in the sector create new value added services, in order to extend its services portfolio and to ensure either the retention of its customers or increase the number of its subscribers. These new services must rely on business processes defined by the Telecommunications Service Provider, which are composed of operation, management, maintenance and support tasks. Generally, Telco Architects reuse those tasks in order to optimize enterprise resource and to ensure prompt return on investment, amortizing over the shortest possible time the outgoings due to creation and deployment of the new service. The reuse of Telco tasks involves constraints regarding the speed in selection, since usually, there are hundreds of tasks, and it requires the intervention of technical staff to carry out the recovery operations, based on their subjective interpretation of the business process to be implemented. There exist different approaches to automate the resources selection, generally focused on the semantic matching of concepts that describe their access interfaces (inputs and outputs); however, is shown that the application of these techniques omits relevant information contained in other attributes, such as identifiers. For this reason, this paper proposes a mechanism to determine the semantic similarity of tasks that make up telecommunications business processes, considering two perspectives: the inference on the tasks functionality specified in identifiers, and coverage analysis of inputs and outputs.

Key words: business processes, telecommunications, tasks, semantic match-making, semantic similarity, ontology.

Resumo

Como medida para melhorar a competitividade no mercado das telecomunicações, atualmente as empresas deste sector criam novos serviços de valor acrescentado, a fim de expandir seu portfólio de serviços e garantir a permanência dos clientes e aumentar o número de subscritores. Esses novos serviços devem ser apoiados em processos de negócio definidos pela Operadora de Telecomunicações, que são feitos das tarefas de operação, gestão, manutenção e suporte. Geralmente, Arquitetos de Telecomunicações reutilizam essas tarefas, a fim de otimizar os recursos da empresa e garantir um rápido retorno sobre o investimento, amortizado no menor tempo possível, os custos de criação e implantação do novo serviço. A reutilização das tarefas de Telecomunicações tem limitações em termos de flexibilidade na sua seleção, dado que normalmente há centenas de tarefas que exige a intervenção do pessoal especializado para realizar as funções de recuperação, a partir da interpretação subjetiva do processo de negócio que se deseja criar. Existem diversas abordagens que procuram automatizar a seleção de recursos, mas comumente são centradas na correspondência semântica de conceitos que descrevem suas interfaces de acesso (entradas e saídas); no entanto, na aplicação dessas técnicas se omite informação relevante contida em outros atributos, como os identificadores. Por esta razão, este trabalho propõe um mecanismo para determinar a similaridade semântica entre as tarefas que compõem os processos de negócio de telecomunicações. Este mecanismo considera duas perspectivas: a inferência sobre a funcionalidade das tarefas especificadas nos identificadores e a análise de cobertura de suas entradas e saídas.

Palavras chaves: processos de negócio, telecomunicações, tarefas, correspondência semântica, similaridade semântica, ontologia.

1 Introducción

El fenómeno de la globalización en el mercado de las telecomunicaciones ha marcado una tendencia hacia el establecimiento de alianzas entre empresas e instituciones del sector. Lo anterior, ha requerido que los operadores de telecomunicaciones sometan sus procesos de negocio a actividades de análisis, gestión y reingeniería con el fin de favorecer su interoperabilidad. Esto ha promovido la generación de mecanismos estándares de definición y descripción de los procesos [1], los cuales permiten un entendimiento común de los objetivos del negocio en todos los niveles de la empresa, y favorecen la automatización de las actividades de descubrimiento, recuperación y reutilización de componentes existentes, independizándolas de la interpretación subjetiva del personal técnico [2].

Al interior de un operador de telecomunicaciones, un proceso de negocio representa de manera formal, el trabajo que se lleva a cabo para proveer soporte a un requisito de la organización o al alcance de un objetivo de negocio. Así por ejemplo, puede articularse un proceso de negocio con el conjunto de operaciones involucradas en la facturación de un servicio determinado, o con las actividades ejecutadas para dar trámite a las peticiones de los clientes. En el dominio específico de las telecomunicaciones, un referente obligado para la definición de los procesos de negocio, lo constituye el trabajo desarrollado en el marco de la iniciativa de estandarización del Telemanagement Forum denominada Sistemas de Operaciones y Software de Nueva Generación (NGOSS, *New Generation Operations Systems and Software*). La NGOSS ha establecido un conjunto de modelos para la gestión de los procesos de negocio, la reingeniería de procesos y la integración de aplicaciones empresariales, propias del sector de las telecomunicaciones [3], entre los cuales se destacan el Mapa de Operaciones de Telecomunicaciones Mejorado (eTOM, *enhanced Telecom Operations Map*), y el Modelo de Información/Datos Compartidos (SID, *Shared Information/Data Model*). En conjunto estos modelos configuran una guía para la definición estandarizada de procesos de negocio, habilitando de esta manera su interoperabilidad y facilitando su reutilización.

La reutilización para un Operador de Telecomunicaciones cobra importancia en el desarrollo de capacidades para crear y desplegar nuevos servicios de valor agregado con un bajo tiempo de salida al mercado (*Time-to-Market*) [4]. En otras palabras, adelantarse a la competencia en el lanzamiento de un servicio, con el fin de generar altos beneficios para amortizar rápidamente los gastos de creación y despliegue. En este sentido, un arquitecto Telco debería contar con mecanismos adecuados para la reutilización de tareas, a partir del amplio y creciente portafolio ofrecido gracias a la definición estándar de los procesos de negocio por parte de los Operadores de Telecomunicaciones.

El presente trabajo describe un enfoque orientado hacia la recuperación y reutilización de tareas, las cuales son los elementos esenciales que constituyen procesos de negocio de telecomunicaciones. El principal aporte está centrado en definir un mecanismo de comparación semántica, para determinar si una tarea publicada satisface la funcionalidad de una tarea solicitada, a partir de la aplicación de medidas de similitud semántica sobre los conceptos que enriquecen sus atributos más relevantes (identificador o nombre, entradas y salidas).

Considerando que en la actualidad no hay suficiente documentación acerca de mecanismos de comparación semántica de tareas entre procesos de negocio, en el presente trabajo se realizó un estudio de las principales aproximaciones de investigación relacionadas con el descubrimiento de servicios web, teniendo en cuenta que estos últimos describen su funcionalidad por medio de atributos semejantes a los definidos para las tareas. De esta manera, las técnicas aplicadas en el descubrimiento de servicios web constituyen una base para la formulación del mecanismo descrito en el presente artículo.

Algunos de las técnicas existentes para el descubrimiento automático de servicios web se basan en interfaces, y están orientadas hacia la utilización de criterios sintácticos para la comparación de palabras clave que describen los atributos del servicio [5, 6]. El proceso de descubrimiento soportado en este tipo de comparación no tiene en cuenta la semántica de la solicitud del servicio respecto a las funcionalidades esperadas, en consecuencia, dicha técnica tiene alta probabilidad de recuperar servicios ambiguos que no corresponden directamente con el criterio de búsqueda. Por esta razón, la mayoría de trabajos de investigación actualmente promueven el uso de inferencia semántica sobre las descripciones de los servicios web, las cuales deben soportarse en lenguajes que incorporen componentes semánticos, con el fin de minimizar en lo posible la intervención humana en la recuperación de servicios que satisfagan adecuadamente los requerimientos del cliente.

En general, los enfoques actuales hacen énfasis en mecanismos de emparejamiento semántico para automatizar el descubrimiento de servicios web, empleando inferencia lógica para identificar relaciones de equivalencia o sub-sunción entre los parámetros que definen las descripciones de una solicitud de servicio y aquellos que describen un servicio suministrado por un proveedor (servicio publicado). En este sentido es posible identificar los tres enfoques más representativos que permiten abordar el proceso de emparejamiento de servicios web:

- *Enfoque de Emparejamiento basado en Relaciones Terminológicas*: consiste en comparar las descripciones de los servicios, utilizando sistemas de léxico como Wordnet o Thesaurus, contrastando los conceptos de la solicitud con los del servicio publicado para identificar relaciones terminológicas, por ejemplo de sinonimia, especialización/generalización y composición [5].

- *Enfoque de Emparejamiento basado en Categorías*: evalúa la relación existente entre los parámetros definidos en los documentos descriptores del servicio publicado y solicitado, tales como los conceptos que enriquecen sus entradas y salidas, con el fin de determinar la correspondencia entre ambos y clasificarla dentro de cinco categorías denominadas grados o niveles de correspondencia (Exacto, Contenido, Contenedor, Intersección y Nulo) [7, 8].
- *Enfoque de Emparejamiento basado en la Clasificación (Ranking) de los servicios*: este enfoque sugiere que la clasificación de la correspondencia basada en categorías no es suficiente para determinar el grado de relación entre los servicios comparados. Propone entonces cuantificar esta relación a partir del cálculo de la similitud semántica¹ entre las entradas/salidas del servicio solicitado y las entradas/salidas de los servicios publicados, de manera que pueda realizarse una clasificación o ranking de servicios de acuerdo con dicho valor de similitud [9, 10, 11, 12]. Existen muchas formas de definir matemáticamente esta relación, una de las más utilizadas consiste en medir la longitud del trayecto entre los dos conceptos comparados, asignando pesos a las aristas del camino que los separa, los cuales dependen del número de nodos secundarios presentes en la jerarquía de la ontología de dominio a la cual pertenecen (profundidad) [12].

A partir de la exploración realizada a través de los tres enfoques de emparejamiento identificados, y considerando la orientación de la presente propuesta, dirigida hacia la utilización de ontologías de dominio de telecomunicaciones; se resolvió aprovechar las ventajas propias del segundo y tercer enfoque (enfoque basado en categorías y enfoque basado en el ranking de los servicios) para determinar la similitud entre tareas de procesos de negocio de telecomunicaciones. Lo anterior teniendo en cuenta que el enfoque basado en relaciones terminológicas, si bien ofrece una alta flexibilidad, en algunos casos puede provocar que los resultados obtenidos del proceso de emparejamiento incluyan servicios que no satisfagan eficientemente el requerimiento del cliente, lo cual traduce en una medida de precisión limitada y en un bajo índice de

¹La similitud semántica hace referencia a la proximidad que existe entre dos conceptos dentro de una ontología.

recuperación (*recall*, en inglés) [5]. Por otra parte, el segundo enfoque mejora la precisión del proceso de emparejamiento respecto al primero, en tanto está soportado en la inferencia sobre ontologías de dominio. Sin embargo, la sola clasificación de la correspondencia en categorías, no permite establecer cuál de los servicios pertenecientes a una misma categoría satisface mejor la solicitud. Este inconveniente puede abordarse mediante el enfoque basado en la clasificación de los servicios, el cual permite determinar cuantitativamente la capacidad de un servicio publicado para satisfacer la funcionalidad de un servicio de consulta.

De esta manera, se definió que es adecuado en primera instancia aplicar un algoritmo de emparejamiento basado en categorías, que permita caracterizar cualitativamente la correspondencia existente entre los atributos que definen la funcionalidad de las tareas comparadas (tarea publicada y tarea de consulta). Posteriormente, es necesario cuantificar estas relaciones para realizar una clasificación de tareas publicadas, que permita reflejar su capacidad para satisfacer el requerimiento definido por la tarea de consulta, mediante la estimación de similitudes semánticas entre estas.

La revisión bibliográfica desarrollada en esta sección muestra que, si bien actualmente la semántica favorece la automatización del descubrimiento de servicios web, la mayoría de los trabajos que abordan esta aproximación, sólo tiene en cuenta el emparejamiento semántico de sus entradas y salidas, omitiendo la información relevante respecto a la funcionalidad, contenida en otros atributos del servicio como su identificador o nombre. Lo anterior se constituye en una limitación que puede ser abordada a partir de la metodología descrita en la siguiente sección.

2 Metodología

La metodología propuesta para llevar a cabo la estimación de la similitud semántica de tareas que conforman procesos de negocio de telecomunicaciones se deriva de la adaptación de algunos mecanismos comúnmente empleados en procesos de descubrimiento y recuperación de servicios web.

Como requisito fundamental para asegurar la fiabilidad en los resultados de similitud entre las tareas de procesos de negocio, es indispensable garantizar una correcta asignación de su enriquecimiento, en otras palabras, es ne-

cesario que los conceptos ontológicos sean asociados fielmente a los atributos que definen la funcionalidad de las tareas (identificadores, entradas y salidas). En este sentido, es necesario emplear ontologías de dominio, que proporcionen un vocabulario común dentro del área específica de interés, las cuales en lo posible, deben estar fundamentadas en modelos ampliamente reconocidos. En el caso particular del sector de las telecomunicaciones, una alternativa adecuada, tal como se argumenta en [13], son las propuestas de la mencionada iniciativa de estandarización NGOSS: los modelos eTOM [14] y SID [15]. Estos modelos constituyen la base de conocimiento sobre la cual se concibieron las ontologías SeTOM y SSID [16], adoptadas como soporte semántico por el mecanismo propuesto en este artículo.

El modelo eTOM provee un marco de referencia que ordena de forma jerárquica y coherente los conceptos que constituyen el conocimiento relativo a los procesos, subprocessos y actividades que se llevan a cabo dentro de un proveedor de servicios de Telecomunicaciones, el cual facilita y soporta las relaciones operador-a-operador, operador-a-cliente y operador-a-socio/proveedor.

La ontología creada a partir del conocimiento dispuesto en eTOM, se denomina SeTOM (Semantic eTOM) y modela los más de trescientos conceptos contenidos en este estándar, los cuales definen en detalle las áreas funcionales y actividades que se desarrollan en un operador de Telecomunicaciones (Por ejemplo: Autorizar Crédito, Crear Factura, Analizar calidad del servicio, entre otros) [17]. La figura 1, ilustra parte de la estructura de la ontología SeTOM, correspondiente a algunos conceptos de nivel superior. En esta figura las líneas de color azul indican conceptos del primer nivel de la ontología, mientras que las líneas rojas determinan relaciones de herencia que se originan desde conceptos de nivel superior hacia sus correspondientes sub-conceptos.

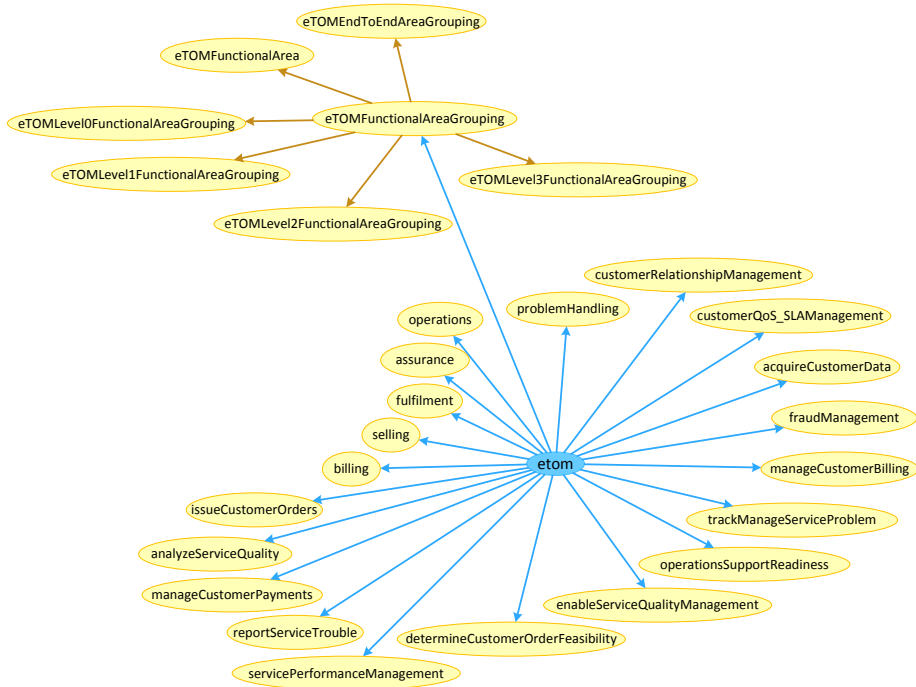


Figura 1: Sección del árbol de conceptos de la Ontología SetOM desplegada en la herramienta WSMO Studio.

Por su parte, el modelo SID captura los conceptos y principios necesarios para la definición de un modelo de información compartida, así como también los elementos del negocio del dominio de los operadores de telecomunicaciones (conocidos en SID como “Entidades”), y sus atributos asociados. Una entidad de negocio es un objeto de interés en las empresas tales como cliente, producto, servicios o red, mientras que sus atributos son características que describen la entidad. Estas definiciones en conjunto proporcionan una perspectiva orientada hacia el negocio, tanto de los datos como de la información necesaria para el funcionamiento de la empresa.

La ontología SSID (Semantic SID) organiza y modela los más de mil conceptos y relaciones definidos en el modelo SID, los cuales cubren términos como cliente, producto, servicio, recurso, proveedor, etc. [17].

En la figura 2 se presentan algunos de los conceptos que configuran esta ontología, los cuales hacen parte de los primeros cuatro niveles de la taxonomía conceptual definida en SSID.

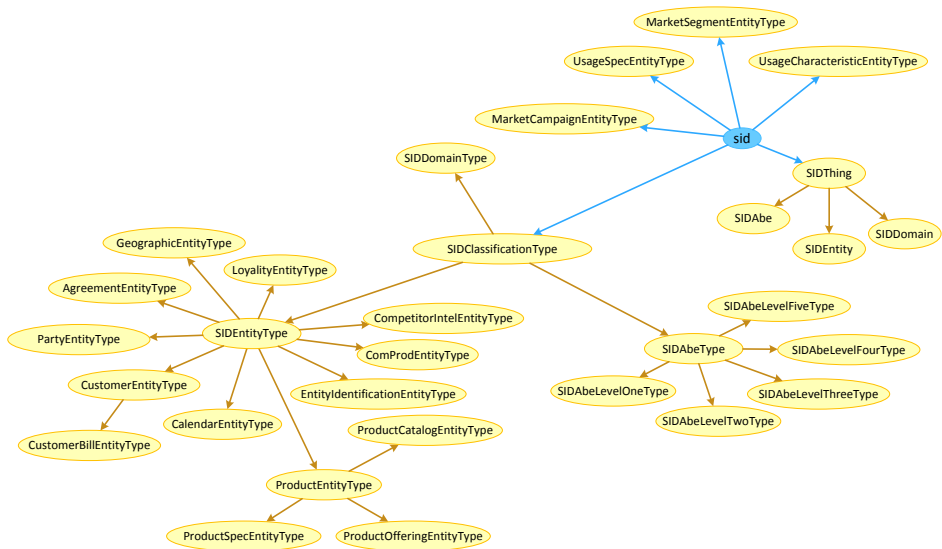


Figura 2: Sección del árbol de conceptos de la Ontología SSID desplegada en la herramienta WSMO Studio.

La metodología propuesta en éste artículo, se soporta en las ontologías descritas, y dispone de los siguientes pasos para determinar la similitud semántica de tareas:

Paso 1: se clasifica el grado de correspondencia entre los conceptos del enriquecimiento de las entradas/salidas de las tareas comparadas de acuerdo a las siguientes categorías:

- **Exacto:** Cuando los dos términos del enriquecimiento comparados pertenecen al mismo concepto en la ontología SSID.
- **Contenido:** Cuando el concepto del enriquecimiento de la tarea publicada pertenece al conjunto de super-conceptos del enriquecimiento de la tarea de consulta, y la separación entre conceptos es menor a dos saltos

en la ontología de dominio.

- **Contenedor:** Cuando el concepto del enriquecimiento de la tarea de consulta pertenece al conjunto de super-conceptos del enriquecimiento de la tarea publicada, y la separación entre conceptos es menor a cuatro saltos en la ontología de dominio.
- **Intersección:** Cuando los conceptos del enriquecimiento semántico de las tareas de consulta y publicada tienen super-conceptos comunes, y la separación entre conceptos es menor a cuatro saltos en la ontología de dominio.
- **Nulo:** Cuando la separación entre conceptos es mayor a cuatro saltos en la ontología de dominio (una distancia semántica de cuatro implica una disimilitud semántica considerable, teniendo en cuenta que la ontología de SID tiene diez niveles de jerarquía).

Algoritmo 1 Determinar el Grado de Correspondencia entre parejas de entradas/salidas

Entrada: concepto de la tarea de consulta (C_Q) y concepto de la tarea publicada (C_T)

Salida: *gradoDeCorrespondencia* entre C_Q y C_T .

```
1: si  $C_Q = C_T$  entonces
2:   devolver gradoDeCorrespondencia  $\leftarrow exacto$ 
3: si no
4:    $SD \leftarrow obtenerDistanciaSemantica(C_Q, C_T)$ 
5:   si  $SD < 4$  entonces
6:      $SC_Q \leftarrow obtenerSuperConceptos(C_Q)$  { $SC_Q$  es el conjunto de super-
       conceptos de  $C_Q$ }
7:      $SC_T \leftarrow obtenerSuperConceptos(C_T)$  { $SC_T$  es el conjunto de super-
       conceptos de  $C_T$ }
8:     si  $C_T \subseteq SC_Q$  and  $SD < 2$  entonces
9:       devolver gradoDeCorrespondencia  $\leftarrow contenido$ 
10:    si no, si  $C_Q \subseteq SC_T$  entonces
11:      devolver gradoDeCorrespondencia  $\leftarrow contenedor$ 
12:    si no
13:      devolver gradoDeCorrespondencia  $\leftarrow interseccion$ 
14:    fin si
15:  si no
16:    devolver gradoDeCorrespondencia  $\leftarrow nulo$ 
17:  fin si
18: fin si
```

Tomando como referencia estas categorías fue diseñado un algoritmo para clasificar la relación entre parejas de entradas/salidas de las tareas comparadas (Algoritmo 1). De acuerdo con este algoritmo, cuando el concepto de la tarea de consulta (C_Q) sea igual al concepto de la tarea publicada (C_T) el grado de correspondencia es establecido como *exacto*. De no cumplirse la condición anterior, es necesario calcular la distancia semántica (SD) entre los dos conceptos C_Q y C_T . Luego, con base en este valor, se determina si el grado de correspondencia es nulo ($SD > 4$).

Si ninguna de las condiciones anteriores se satisface es necesario recuperar el conjunto de super-conceptos de C_Q y C_T (denominados SC_Q y SC_T res-

pectivamente). Con estos conjuntos, el algoritmo determina si el grado de correspondencia es *contenido* o *contenedor*, evaluando las condiciones previamente descritas. Finalmente, el grado de correspondencia se clasifica como *intersección*, de no cumplirse las condiciones para las categorías anteriores (*exacto*, *contenido*, *contenedor* o *nulo*).

En el algoritmo anterior, la función *obtenerDistanciaSemantica* (línea 4) hace referencia al procedimiento que permite establecer la distancia entre los conceptos que enriquecen las entradas y salidas de las tareas comparadas (los cuales están dispuestos dentro de la jerarquía de diez niveles definida en la ontología SSID). El procedimiento implementado en esta función es el siguiente:

Inicialmente, se calcula la *Distancia por Salto* (D_{salto}) en cada uno de los niveles de la ontología que separan a los dos conceptos comparados, aplicando la ecuación (1):

$$D_{salto} = 1 + \frac{1}{2^{depth}} \quad (1)$$

depth es el número de saltos en la ontología que hay desde el *concepto raíz* hasta el *concepto objetivo*. D_{salto} es calculada para cada concepto presente en el trayecto entre el concepto del enriquecimiento de la tarea de consulta y el concepto del enriquecimiento de la tarea publicada. Finalmente, se calcula la Distancia Semántica (SD) entre los conceptos comparados mediante la sumatoria de los valores de D_{salto} asignados a cada uno de los trayectos que los separa:

$$SD = \sum D_{salto} \quad (2)$$

La figura 3 presenta el procedimiento para calcular la distancia semántica entre dos conceptos (A y D) pertenecientes al enriquecimiento de dos tareas.

Paso 2: una vez establecidas la distancia semántica y la categoría en que se clasifica la correspondencia entre el enriquecimiento de las entradas/salidas de las tareas comparadas, es necesario ponderar estas relaciones para determinar su similitud semántica. Para esto se ha dispuesto emplear la ecuación (3):

$$SS = \frac{1}{\rho SD + 1} \quad (3)$$

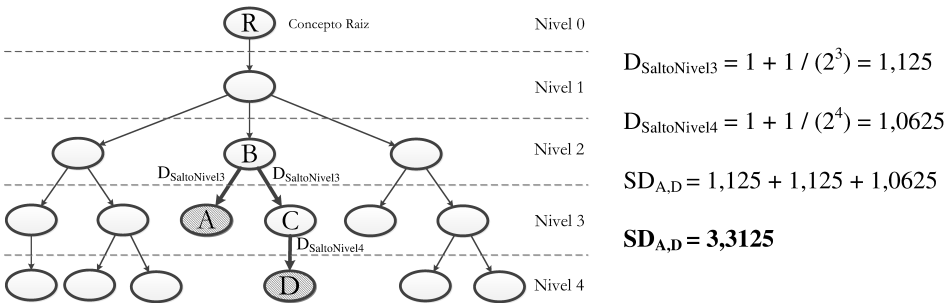


Figura 3: Cálculo de la Distancia Semántica sobre SSID.

El factor ρ en la ecuación (3) permite condicionar el impacto de la distancia semántica en el valor de la similitud, dependiendo del grado de correspondencia que hay entre los dos conceptos comparados. Así, entre mayor sea la correspondencia entre los conceptos, menor será la influencia de la distancia semántica y por tanto mayor su similitud. Teniendo esto en cuenta, los valores de ρ fueron asignados de la siguiente manera²: si el grado de correspondencia es *exacto* el valor de ρ es 0, si es *contenido* ρ es 0,2, si es *contenedor* ρ es 0,6, si es *intersección* ρ es 0,8 y si es *nulo* ρ es 1.

El Algoritmo 2 describe formalmente el mecanismo para calcular la similitud semántica entre conceptos del enriquecimiento. Éste algoritmo recibe como entradas los dos conceptos a comparar (C_Q y C_T), y la distancia semántica entre ellos (SD). En primer lugar, evalúa si el grado de correspondencia es exacto, caso en el cual el factor ρ es nulo (0) y la similitud semántica (SS) es equivalente a 1. De lo contrario, el algoritmo determina grado de correspondencia entre C_Q y C_T (MD). Luego, de acuerdo con el resultado obtenido en MD , ajusta el parámetro ρ que interviene en el cálculo de la similitud semántica junto con SD .

²El valor de ρ se definió en la fase de pruebas de un prototipo software, que implementa la metodología de comparación descrita en esta sección del artículo.

Algoritmo 2 Determinar la Similitud Semántica entre Conceptos

Entrada: concepto de la tarea de consulta (C_Q), concepto de la tarea publicada (C_T) y la distancia semántica (SD) entre C_Q y C_T

Salida: similitud semántica SS entre C_Q y C_T .

```

1:  $\rho \leftarrow 0$  { $\rho$  es un factor que depende del grado de correspondencia}
2:  $MD \leftarrow obtenerGradoDeCorrespondencia(C_Q, C_T)$  { $MD$  es el grado de correspondencia entre  $C_Q$  y  $C_T$ }
3: si  $MD \neq exacto$  entonces
4:   si  $MD = contenido$  entonces
5:      $\rho \leftarrow 0,2$ 
6:   si no, si  $MD = contenedor$  entonces
7:      $\rho \leftarrow 0,6$ 
8:   si no, si  $MD = interseccion$  entonces
9:      $\rho \leftarrow 0,8$ 
10:  si no
11:     $\rho \leftarrow 1$ 
12:  fin si
13: fin si
14:  $SS \leftarrow \frac{1}{\rho SD + 1}$ 
15: devolver  $SS$ 

```

Paso 3: Dado que las tareas pueden contener varias entradas y varias salidas, los dos pasos anteriores deben aplicarse para estimar la similitud entre cada una de las entradas/salidas de las tareas publicada y de consulta. A partir de los resultados obtenidos, se calcula la similitud general de entradas/salidas de las tareas comparadas.

El procedimiento para llevar a cabo este cálculo es una adaptación del mecanismo propuesto en [18], la cual consiste en un análisis de cobertura (de las entradas o salidas) para determinar la tarea publicada (T_T) más similar a una tarea de consulta (T_Q), teniendo en cuenta las siguientes condiciones:

1. T_T debe compartir y, en lo posible, no exceder las entradas de T_Q .
2. T_T debe compartir el mayor número posible de salidas con T_Q .

A partir de la evaluación de las condiciones anteriores, se define el siguiente procedimiento para determinar la similitud general de entradas (también válido para el caso de las salidas, considerando algunas variaciones):

Los resultados de similitud entre parejas de entradas de la tarea de consulta y la tarea publicada (calculados según el Algoritmo 2), son ordenados en una matriz de similitud, cuyas filas corresponden a las entradas de la tarea de consulta y las columnas a las entradas de la tarea publicada. Un análisis realizado sobre esta matriz permite determinar las *similitudes máximas* entre parejas de entradas de las dos tareas comparadas, las cuales hacen referencia a los mayores valores de similitud que pueden obtenerse entre todas las posibles combinaciones de parejas de entradas (una entrada de la tarea de consulta y otra de la publicada), garantizando que cada entrada de la tarea de consulta esté relacionada con una única entrada de la tarea publicada y viceversa.

De esta manera, definiendo S como la matriz de similitud de entradas de dimensión $m \times n$ (donde m hace referencia a las entradas de la tarea de consulta y n a las de la tarea publicada), se determina el conjunto de *similitudes máximas* como la agrupación de similitudes $S_{i,j}$ que cumplen con la siguiente propiedad:

$$S_{i,j} : \forall S_{k,j} (k = 1, 2, \dots, m), S_{i,j} > S_{k,j} \wedge \forall S_{i,l} (l = 1, 2, \dots, n), S_{i,j} > S_{i,l}$$

Las entradas que no hacen parte de las relaciones establecidas en el anterior conjunto de similitudes se denominan entradas perdidas (estas se presentan cuando m excede a n o viceversa). A partir de la especificación del conjunto de *similitudes máximas* y teniendo en cuenta las condiciones del análisis de cobertura descritas previamente, se define la Similitud Semántica entre entradas o salidas (*SIOS*) como:

$$SIOS = \frac{\sum S_{i,j}}{\min(m, n) + \beta |m - n|} \quad (4)$$

donde:

- $\sum S_{i,j}$ representa la sumatoria de las similitudes máximas
- $\min(m, n)$ es el máximo número de parejas de entradas que puede establecerse entre las dos tareas comparadas
- $\beta |m - n|$ es un parámetro que permite condicionar el valor de la similitud cuando existen entradas perdidas.

De esta manera, cuando la tarea de consulta tiene más entradas que la tarea publicada, (cumpliéndose la condición 1) β toma el valor de 0,8, de lo

contrario, significa que la tarea publicada excede el número de entradas de la tarea de consulta, entonces, β es equivalente a 1.

El procedimiento descrito, se formaliza en el Algoritmo 3, en el cual inicialmente son extraídos los conceptos del enriquecimiento de entradas de la tarea de consulta y de la tarea publicada para almacenarlos en los arreglos IE_Q e IE_T respectivamente. Posteriormente, se obtiene la matriz de similitudes de entradas ($ISM[m][n]$), a partir de la comparación de los arreglos mencionados, para luego determinar las máximas similitudes (MSA). Obtenida esta información, el algoritmo aplica la ecuación (4) para calcular la similitud semántica de entradas.

Algoritmo 3 Determinar la Similitud Semántica de Entradas

Entrada: tarea de consulta (T_Q) y tarea publicada (T_T)

Salida: similitud semántica de entradas (SIS) entre T_Q y T_T .

```

1:  $I_Q \leftarrow obtenerEntradas(T_Q)$ 
2:  $m \leftarrow dimension(I_Q)$ 
3:  $I_T \leftarrow obtenerEntradas(T_T)$ 
4:  $n \leftarrow dimension(I_T)$  { $I_Q$  es el arreglo de entradas de la tarea de consulta,
   e  $I_T$  es el arreglo de entradas de la tarea publicada}
5:  $IE_Q[m] \leftarrow obtenerEnriquecimientoDeEntradas(T_Q)$ 
6:  $IE_T[n] \leftarrow obtenerEnriquecimientoDeEntradas(T_T)$ 
7: si  $I_Q = \emptyset$  and  $I_T = \emptyset$  entonces
8:   devolver  $SIS \leftarrow 1$ 
9: si no, si  $I_Q = \emptyset$  or  $I_T = \emptyset$  entonces
10:  devolver  $SIS \leftarrow 1$ 
11: si no
12:   $ISM[m][n] \leftarrow obtenerMatrizSimilitudEntradas(IE_Q[m], IE_T[n])$ 
    { $ISM[m][n]$  es la matriz de similitudes entre entradas}
13:   $MSA \leftarrow obtenerArregloDeSimilitudesMaximas(ISM)$  { $MSA$  es
    el arreglo de similitudes máximas}
14:   $MSS \leftarrow \sum_i MSA[i]$ 
15:  si  $m > n$  entonces
16:    devolver  $SIS \leftarrow \frac{MSS}{n+0,8|m-n|}$ 
17:  si no
18:    devolver  $SIS \leftarrow \frac{MSS}{n}$ 
19:  fin si
20: fin si

```

El mecanismo para calcular la similitud semántica de salidas entre las dos tareas comparadas es semejante al procedimiento descrito en el algoritmo 3. La única variación de este procedimiento respecto al anterior consiste en modificar el valor de β para cumplir la condición 2 del análisis de cobertura, de esta manera, cuando la tarea publicada excede el número de salidas de la tarea de consulta, el valor de β es 0,8, en caso contrario β se ajusta en 1; esto permite condicionar el impacto de las salidas perdidas en el valor de similitud.

Paso 4: Determinar la distancia semántica y el grado de correspondencia entre identificadores de las tareas, requiere de procesos diferentes a los utilizados para las entradas y salidas, dado que la ontología de dominio utilizada para enriquecer dichos identificadores (SeTOM) no está organizada en niveles de jerarquía. En lugar de esto, SeTOM proporciona un conjunto de axiomas que permiten inferir sobre relaciones existentes entre los conceptos de la ontología, las cuales no son explícitas en la estructura de la misma. Así por ejemplo, estos axiomas, proveen información respecto a la distribución de las actividades de un Operador de Telecomunicaciones (asociadas cada una a un concepto de SeTOM) en las diferentes áreas funcionales (niveles) del modelo eTOM.

La distancia semántica entre conceptos pertenecientes a la ontología SeTOM puede calcularse mediante el análisis de los elementos que conforman los conceptos del nivel 2 del modelo eTOM. De la misma manera, los conceptos que hacen parte de este nivel proporcionan una visión intermedia de los procedimientos llevados a cabo al interior del Operador. Este nivel se encuentra por debajo de la perspectiva gerencial de los procesos otorgada por el nivel 1, y por encima del grado de detalle de las actividades que los componen, el cual está dado por el nivel 3.

De esta manera, se determinó que la cantidad de componentes de nivel 3 de eTOM, que integran un componente de nivel 2, da cuenta de la distribución del trabajo llevado a cabo en éste último. En este sentido, puede relacionarse la distancia que separa un concepto componente de un concepto contenedor como el número de sub-operaciones (conceptos componentes) necesarias para llevar a cabo la operación representada por el concepto contenedor. De esta manera, entre menor sea el número de sub-operaciones que conforma un concepto contenedor, menor será la distancia semántica entre este último y sus conceptos componentes. Así mismo, la distancia entre conceptos componentes depende de su cantidad, en otras palabras, entre mayor sea el número

de elementos que componen un concepto de nivel superior, menor será su dependencia funcional y por tanto mayor su distancia semántica.

Teniendo en cuenta lo anterior, el algoritmo propuesto que permite estimar la distancia semántica entre conceptos del enriquecimiento de identificadores (Algoritmo 4), comienza por evaluar si el concepto del enriquecimiento del identificador de la tarea de consulta (Id_Q) es diferente al concepto de la tarea publicada (Id_T). El no cumplimiento de esta condición implica que la distancia semántica sea nula (0).

Algoritmo 4 Calcular Distancia Semántica entre parejas de identificadores

Entrada: concepto del identificador de la tarea de consulta (Id_Q) y concepto del identificador de la tarea publicada (Id_T)

Salida: distancia semántica (SD) entre Id_Q y Id_T .

```

1:  $SD \leftarrow 0$ 
2: si  $Id_Q \neq Id_T$  entonces
3:    $LId_Q \leftarrow obtenerNivel(Id_Q)$ 
4:    $LId_T \leftarrow obtenerNivel(Id_T)$ 
5:   si ( $LId_Q \neq 2$  and  $LId_Q \neq 3$ ) or ( $LId_T \neq 2$  and  $LId_T \neq 3$ ) entonces
6:      $SD \leftarrow \infty$ 
7:   si no
8:      $SLC_Q \leftarrow obtenerConceptoDeSegundoNivel(Id_Q)$  { $SLC_Q$  es el concepto de Segundo nivel, al cual pertenece  $I_Q$ }
9:      $SLC_T \leftarrow obtenerConceptoDeSegundoNivel(Id_T)$  { $SLC_T$  es el concepto de Segundo nivel, al cual pertenece  $I_T$ }
10:    si  $SLC_Q = SLC_T$  entonces
11:       $SD \leftarrow obtenerNumeroDeComponentes(SLC_Q)$ 
12:    si no
13:       $SD \leftarrow \infty$ 
14:    fin si
15:  fin si
16: fin si
17: devolver  $SD$ 

```

Si Id_Q y Id_T son distintos, el algoritmo procede a encontrar el nivel de eTOM al cual pertenecen, utilizando los axiomas de descomposición y agrupamiento de la ontología (estos niveles se denotan como LId_Q y LId_T). Pos-

teriormente, se verifica que tanto Id_Q como Id_T pertenezcan a los niveles 2 o 3 de la misma área funcional, en cuyo caso, la Distancia Semántica (SD) entre ellos sería equivalente al número de componentes de nivel 3 que integran el de nivel 2. De no cumplirse las anteriores condiciones, la distancia semántica es fijada en un valor lo suficientemente grande para reflejar que no hay semejanza entre los dos conceptos comparados.

Paso 5: Teniendo en cuenta el segundo enfoque de emparejamiento y el análisis realizado a los niveles 2 y 3 del modelo eTOM, se definieron 5 categorías para clasificar la correspondencia entre identificadores:

- **Exacto:** Cuando los dos términos del enriquecimiento comparados pertenecen al mismo concepto en la ontología SeTOM.
- **Contenido:** Cuando se cumplen las siguientes condiciones:
 1. El concepto del enriquecimiento de la tarea de consulta pertenece al nivel 3 del modelo eTOM.
 2. El concepto del enriquecimiento de la tarea publicada pertenece al nivel 2 del modelo eTOM.
 3. El concepto del enriquecimiento de la tarea de consulta hace parte del conjunto de componentes que conforman el concepto del enriquecimiento de la tarea publicada.
- **Contenedor:** Cuando se cumplen las siguientes condiciones:
 1. El enriquecimiento de la tarea publicada pertenece al nivel 3 de eTOM.
 2. El enriquecimiento de la tarea de consulta pertenece al nivel 2 de eTOM.
 3. El enriquecimiento de la tarea publicada hace parte del conjunto de componentes que conforman el enriquecimiento de la tarea de consulta.
- **Intersección:** Cuando los conceptos del enriquecimiento semántico de las tareas de consulta y publicada son de nivel 3, y hacen parte del mismo componente de nivel 2.
- **Nulo:** Cuando ninguna de las condiciones anteriores se satisface.

Tomando como referencia estas categorías, fue diseñado un algoritmo para clasificar el grado de correspondencia entre parejas de identificadores de tareas comparadas (Algoritmo 5).

Algoritmo 5 Calcular Distancia Semántica entre parejas de identificadores

Entrada: concepto del identificador de la tarea de consulta (Id_Q) y concepto del identificador de la tarea publicada (Id_T)

Salida: *gradoDeCorrespondencia* entre Id_Q y Id_T .

```

1: si  $Id_Q = Id_T$  entonces
2:   devolver gradoDeCorrespondencia  $\leftarrow exact$ 
3: si no
4:    $LId_Q \leftarrow obtenerNivel(Id_Q)$ 
5:    $LId_T \leftarrow obtenerNivel(Id_T)$ 
6:   si ( $LId_Q \neq 2$  and  $LId_Q \neq 3$ ) or ( $LId_T \neq 2$  and  $LId_T \neq 3$ ) entonces
7:     devolver gradoDeCorrespondencia  $\leftarrow nulo$ 
8:   si no
9:      $SLC_Q \leftarrow obtenerConceptoDeSegundoNivel(Id_Q)$  { $SLC_Q$  es el concepto de Segundo nivel, al cual pertenece  $I_Q$ }
10:     $SLC_T \leftarrow obtenerConceptoDeSegundoNivel(Id_T)$  { $SLC_T$  es el concepto de Segundo nivel, al cual pertenece  $I_T$ }
11:    si  $SLC_Q = SLC_T$  entonces
12:      si  $LId_Q = 3$  and  $LId_T = 3$  entonces
13:        devolver gradoDeCorrespondencia  $\leftarrow interseccion$ 
14:      si no, si  $LId_T = 2$  entonces
15:        devolver gradoDeCorrespondencia  $\leftarrow contenido$ 
16:      si no
17:        devolver gradoDeCorrespondencia  $\leftarrow contenedor$ 
18:      fin si
19:    si no
20:      devolver gradoDeCorrespondencia  $\leftarrow nulo$ 
21:    fin si
22:  fin si
23: fin si

```

Según este algoritmo, cuando el concepto de la tarea de consulta (Id_Q) es igual al concepto de la tarea publicada (Id_T) el grado de correspondencia es establecido como *exacto*. De lo contrario se verifica que tanto Id_Q como Id_T pertenezcan a los niveles 2 o 3 de la misma área funcional. Al cumplirse esta condición, el algoritmo evalúa si tanto el concepto de la tarea de consulta (Id_Q) como el de la tarea publicada (Id_T) pertenecen al nivel 3, caso en el cual

la correspondencia se establece como *intersección*. De no ser así el algoritmo evalúa las condiciones para definir el grado de correspondencia como *contenido* o *contenedor*. Finalmente, el grado de correspondencia es *nulo* si no se satisface ninguna de las condiciones anteriores.

Paso 6: La Similitud Semántica de identificadores entre las tareas publicada y de consulta, se calcula mediante el Algoritmo 2.

Paso 7: Por último, una vez obtenidos los valores de similitud de identificadores, entradas y salidas mediante los algoritmos explicados anteriormente, se procede a calcular la similitud semántica general (OSS) entre las dos tareas comparadas, a través de una suma ponderada de dichos resultados:

$$OSS = IdSym \cdot wId + InSym \cdot wIn + OutSym \cdot wOut \quad (5)$$

donde:

- *IdSym* es la similitud semántica de Identificadores.
- *wId* es el factor de contribución de la similitud de identificadores en la similitud semántica general de las tareas.
- *InSym* es la similitud semántica de entradas.
- *wIn* es el porcentaje que determina la contribución de la similitud de entradas en la similitud semántica general de las tareas.
- *OutSym* es la similitud semántica de salidas.
- *wOut* es el porcentaje que determina la contribución de la similitud de salidas en la similitud semántica general de las tareas.

Los valores de *wId*, *wIn* y *wOut* son ajustados al criterio del usuario. Sin embargo, de acuerdo con los resultados obtenidos en el proceso de evaluación, al cual fue sometido el sistema que implementa la metodología propuesta, se identificaron unos pesos sugeridos (para los cuales se alcanza el mejor rendimiento). Éstos son: *wId* = 0,45; *wIn* = 0,35; *wOut* = 0,20.

La metodología para calcular la similitud semántica entre dos tareas de procesos de negocio de telecomunicaciones descrito en esta sección, puede aplicarse para la comparación entre las tareas del proceso de negocio de consulta y las tareas de los procesos disponibles en el operador. De esta manera, mediante los valores de similitud es posible establecer un *ranking* (en español: clasificación) de tareas publicadas para cada una de las tareas de consulta, dependiendo de la capacidad que tienen las primeras para cumplir los requerimientos de la solicitud.

3 Evaluación

Previo a la evaluación de la metodología propuesta, fue necesario el modelado de un conjunto de procesos de negocio de telecomunicaciones, utilizando BPMO (*Business Process Modelling Ontology*) [19], una notación semántica basada en el estándar BPMN (*Business Process Management Notation*) [20]. Los procesos modelados están organizados en 5 dominios diferentes, de acuerdo con su ubicación en el modelo eTOM, y representan en conjunto 97 tareas (tareas publicadas), las cuales fueron semánticamente enriquecidas, siguiendo el enfoque descrito en [13].

Posteriormente, 10 de las 97 tareas publicadas, fueron definidas como tareas de consulta. A continuación, se compararon las tareas de consulta con las tareas publicadas, mediante un prototipo que implementa la metodología introducida en la sección anterior. De esta manera, con la base de resultados producto de estas comparaciones se generó el *benchmark del prototipo*, el cual define un punto de referencia, para determinar el desempeño relativo del sistema desarrollado [21], respecto a un benchmark que reúne los resultados generados por expertos del dominio, al realizar las mismas comparaciones ejecutadas por el prototipo.

Fue necesario entonces, el desarrollo de una herramienta que permitiera realizar comparaciones manuales entre tareas, a partir de la calificación de la similitud entre sus atributos. Esta herramienta se puso a disposición de un grupo de ocho (8) personas con amplio conocimiento del dominio de los procesos de telecomunicaciones, con el objetivo de construir el *benchmark de referencia* para evaluar el desempeño de la metodología propuesta.

En la herramienta desarrollada se realizaron en total 2910 comparaciones entre parejas de tareas (publicada y de consulta), garantizando que cada una de las 970 posibles parejas (10 tareas de consulta \times 97 tareas publicadas) fuera comparada por al menos 3 de los evaluadores. En consecuencia, al final del proceso de evaluación se obtuvo 3 resultados de similitud para cada par de tareas, los cuales fueron promediados en cada caso, conformando de esta manera la base de 970 resultados de similitud que constituyen el benchmark de referencia.

Una vez generado el benchmark de referencia se compararon sus resultados con aquellos registrados en el benchmark del prototipo, a partir de un análisis basado en la aplicación de medidas estadísticas, empleadas en la

caracterización del desempeño de sistemas de recuperación de información. Estas medidas se denominan $precision(p)$, $recall(r)$, $overall(o)$ [22], $top-k\ precision(P_k)$ y $top-p\ precision(P_p)$ [23].

4 Resultados

La figura 4 presenta los resultados obtenidos de la aplicación de las medidas de $precision$, $recall$ y $overall$ para diferentes valores de similitud semántica ($umbral$ en la gráfica).

De acuerdo con estos resultados, el desempeño de la metodología es en general adecuado, ya que gran parte de los valores de similitud entre tareas obtenidos, coinciden con los resultados registrados en el benchmark de referencia. Esto se refleja en el valor de la medida de precision, el cual es superior al 60 % y es en promedio del 90 %, para todo el rango de valores de similitud semántica. El mejor desempeño del mecanismo se obtiene para similitudes superiores a 3,5 de acuerdo con la medida de overall, la cual determina una alta correspondencia entre ambos benchmark para este intervalo.

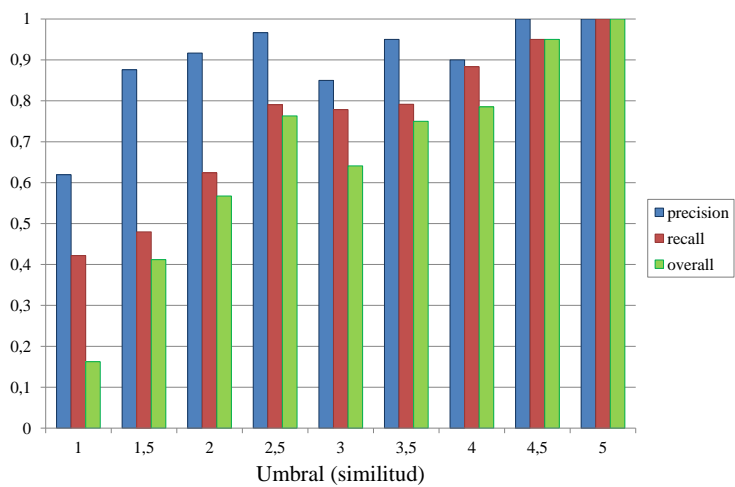


Figura 4: Medidas de Desempeño: Precision, Recall y Overall para el Sistema.

Por otra parte, el desempeño más pobre del sistema se presenta en los valores bajos de similitud (valores inferiores a 2), rango en el que es evidente la dificultad que presenta la metodología para estimar la similitud semántica de tareas disimiles. En este intervalo, si bien la mayor parte de las tareas clasificadas en el ranking del prototipo³, hacen parte igualmente del ranking de referencia⁴ (*precision* aceptable), también son omitidas tareas que, de acuerdo con este último, deberían clasificarse (bajo *recall*). Este comportamiento se justifica en que, para valores bajos de similitud, el mecanismo de comparación semántica evalúa relaciones conceptuales débiles, las cuales pueden resultarle ambiguas, subestimándolas o sobreestimándolas respecto al criterio de un evaluador experto.

Las dos últimas medidas de desempeño: *top-k precision* y *top-p precision*, permiten evaluar la calidad del procedimiento de clasificación de las tareas más relevantes (dispuestas por orden de similitud) para satisfacer la tarea de consulta. Para estas medidas se tuvo en cuenta los siguientes escenarios de prueba:

1. **Dominio vs. Dominio:** evaluación aplicada sobre las comparaciones realizadas entre tareas de consulta y tareas publicadas, pertenecientes a procesos de negocio del mismo dominio.
2. **Dominio vs. Todos los Dominios:** evaluación aplicada sobre todas las comparaciones realizadas (sin discriminar el dominio al cual pertenecen las tareas comparadas).

La precisión del sistema, medida en términos del porcentaje de tareas relevantes recuperadas en los primeros k niveles del ranking del prototipo, puede observarse en la figura 5.

De acuerdo con los resultados presentados en esta gráfica, la medida de precisión disminuye conforme aumenta el valor de k, y alcanza finalmente un valor mínimo, alrededor del cual se mantiene en los niveles restantes. El mejor comportamiento de la metodología se obtiene para las condiciones establecidas en el escenario 1, donde la precisión del sistema es considerablemente alta entre los primeros 3 niveles del ranking, siendo ideal (100 %) para el primer nivel y manteniéndose por encima del 60 % hasta el doceavo.

³Ranking de tareas generado a partir del benchmark del prototipo.

⁴Ranking de tareas generado a partir del benchmark de referencia.

Para el escenario 2 de pruebas, la precisión es inferior pero significativamente favorable, puesto que al igual que en el primer escenario, alcanza un valor máximo del 100 % en el primer nivel del ranking, el cual disminuye hasta el 70 % en el nivel 3 y se mantiene superior al 50 % en los niveles restantes. De esta manera, el desempeño del sistema respecto a la medida de *top-k precision* es óptimo para valores de k inferiores a 3 y es en general aceptable para los demás niveles del ranking.

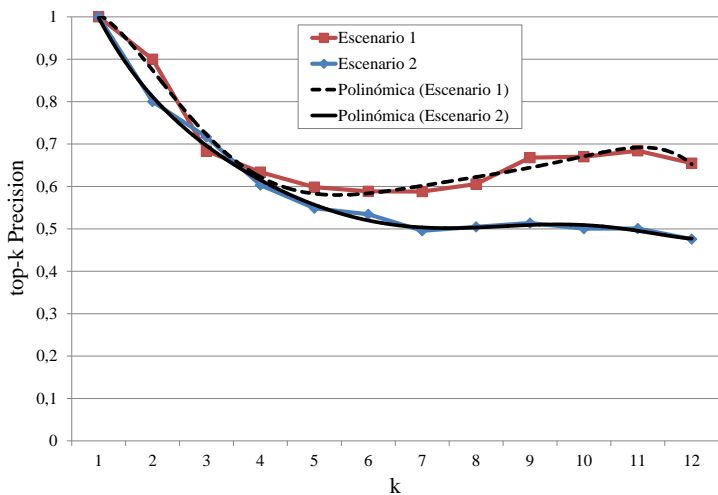


Figura 5: Gráfica de top-k precision vs. k.

La medida de desempeño *top-p precision*, es más exigente que la anterior, ya que evalúa la capacidad del sistema para posicionar adecuadamente las tareas publicadas en el ranking del prototipo. Los resultados de la aplicación de esta medida sobre el sistema implementado se ilustran en la figura 6. En esta gráfica se observa como las curvas de *top-p precision* (para ambos escenarios de pruebas) tienen un comportamiento muy similar: las dos curvas alcanzan un valor máximo de precisión del 100 % en el primer nivel, el cual decrece en forma continua conforme el número k de niveles del ranking aumenta.

El primer escenario de pruebas presenta una precisión ligeramente superior respecto al segundo, lo cual se atribuye al filtrado de tareas por dominio. Es

as3 como, puede caracterizarse el desempe1o del sistema (en relaci3n con la medida de *top-p precision*) como 3ptimo en los 3 primeros niveles del ranking, en los cuales la precisi3n es superior al 60 % y aceptable en hasta el nivel 4, donde la precisi3n decrece hasta el 50 %.

El comportamiento del sistema proyectado en esta 3ltima medida de desempe1o, se justifica en el car3cter sem3ntico del mecanismo de comparaci3n. Es as3 como, en determinados casos el sistema puede llegar a estimar relaciones de similitud entre tareas, las cuales no son inmediatamente evidentes para una persona com3n e incluso para un evaluador experto, debido a que los t3rminos que identifican los atributos de dichas tareas difieren significativamente desde el punto de vista sint3ctico. De esta manera, el sistema puede clasificar en niveles primarios del ranking, tareas que sem3nticamente son muy semejantes, pero que sint3cticamente no lo son tanto.

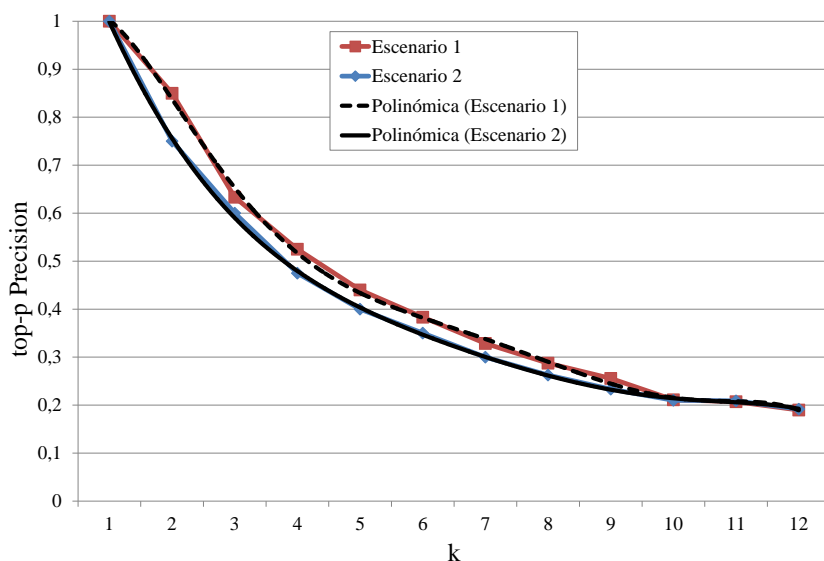


Figura 6: Gr3fica de top-p precision vs. k.

5 Conclusiones

La propuesta presentada en este artículo, aborda la formulación de una metodología de comparación semántica, para habilitar la reutilización de las tareas que componen procesos empresariales de telecomunicaciones, la cual facilita la ejecución de actividades de gestión y reingeniería orientadas hacia la creación y el despliegue de nuevos servicios de valor agregado, al interior un Operador Telco.

Para definir la metodología descrita, fue necesario realizar un estudio del estado actual de los trabajos de investigación relacionados con el descubrimiento de servicios web, haciendo especial énfasis en los algoritmos de emparejamiento, distancia y similitud semánticos aplicados sobre los conceptos (pertenecientes a ontologías de dominio) que enriquecen sus descripciones. Lo anterior considerando, en primer lugar, la escasa documentación respecto a mecanismos de comparación semántica de tareas pertenecientes a procesos de negocio de telecomunicaciones, y en segundo lugar, la correspondencia existente entre los principales atributos que describen la funcionalidad de las tareas y los parámetros definidos para los servicios web (identificador o nombre, entradas y salidas).

La revisión bibliográfica realizada permitió verificar que, si bien los enfoques actuales consideran la semántica presente en la descripción de las capacidades de los servicios web, en la mayoría de los casos sólo tienen en cuenta el emparejamiento semántico de sus entradas y salidas, omitiendo la información relevante contenida en su identificador o nombre. En virtud de lo expuesto, la metodología propuesta en este artículo define un mecanismo de comparación semántica de tareas, basado en el análisis de los conceptos que enriquecen tanto sus entradas y salidas como el identificador de las mismas.

Los resultados de la evaluación realizada sobre la metodología de comparación de tareas, determinaron (de acuerdo con las medidas de *Precision*, *Recall*, *Overall*, *top-k precision* y *top-p precision*) que el desempeño general del sistema desarrollado es satisfactorio. Se demostró de esta manera, que la especificación de procesos de negocio soportada en la utilización de una notación estándar y un vocabulario común (provisto por ontologías de dominio), efectivamente permite disminuir la ambigüedad en la selección de tareas y obtener resultados pertinentes, en relación con el criterio de un evaluador competente en el dominio. Sin embargo, la fiabilidad de los resultados obtenidos de las

comparaciones ejecutadas por la metodología propuesta, están sujetos al adecuado enriquecimiento semántico de las tareas, es decir, debe asegurarse que los conceptos ontológicos asociados a sus identificadores, entradas y salidas, reflejen efectivamente el significado de dichos atributos.

Finalmente, en mérito de lo planteado en este artículo, pueden sugerirse algunas líneas de investigación futuras, relacionadas con la adaptación y aplicación de la metodología propuesta sobre otros dominios de conocimiento, la aproximación hacia la utilización de múltiples ontologías de dominio en el procedimiento de comparación (mediante la implementación mediadores de ontologías) y la consideración de otros atributos de las tareas, en la estimación de su similitud semántica.

Agradecimientos

Los autores agradecen a la Universidad del Cauca, al Departamento Administrativo de Ciencia, Tecnología e Innovación - Colciencias y al Proyecto TelComp2.0 por financiar y soportar esta investigación.

Referencias

- [1] Z. Yan, E. Cimpian, M. Mazzara, M. Zaremba. *D1.1 BPMO Requirements Analysis and Design–SemBiz Deliverable*.
<http://www.sembiz.org/attach/D1.1.pdf>, Noviembre de 2009.
Referenciado en 67
- [2] N. Pérez, H. Muñoz, D. Marcos, J. Martínez. *Gestión de Procesos de Negocio Semánticos–Telefónica Investigación y Desarrollo*.
<http://www.ip-super.org/res/Papers/SBPM-TelecomID2007.pdf>,
Enero de 2009. Referenciado en 67
- [3] J. Fleck. *Overview of the Structure of the NGOSS Architecture*–White Paper, Hewlett Packard, May 2003. Referenciado en 68
- [4] D. Moro, D. Lozano, M. Macias. *WIMS 2.0: Enabling Telecom Networks Assets in the Future Internet of Services*, ISBN 978-3-540-89896-2. Springer Berlin / Heidelberg, 2008. Referenciado en 68
- [5] D. Bianchini, V. De Antonellis, B. Pernici, P. Plebani.
Ontology-based Methodology for e-Service discovery. Information Systems, ISSN 0306–4379, **31**(4), 361–380 (2006). Referenciado en 69, 71

- [6] A. Budanitsky, G. Hirst. *Evaluating WordNet-based Measures of Lexical Semantic Relatedness*. Computational Linguistics, ISSN 0891-2017, **32**(1), 13–47 (2006). Referenciado en 69
- [7] L. Li, I. Horrocks. *A software framework for matchmaking based on semantic web technology*. Proceedings of the 12th international conference on World Wide Web, 2003. WWW '03. ISBN 1-58113-680-3, Budapest, Hungary, 331–339 (2003). Referenciado en 70
- [8] M. Paolucci, T. Kawamura, T. Payne, K. Sycara. *Semantic Matching of Web Services Capabilities*. Proceedings of the First International Semantic Web Conference. ISWC 2002. ISBN 3-540-43760-6, Sardinia, Italy, 333–347 (2002). Referenciado en 70
- [9] B. Xu, P. Zhang, J. Li, W. Yang. *A Semantic Matchmaker for Ranking Web Services*. Journal of Computer Science and Technology, ISSN 1000-9000, **21**(4), 574–581 (2006). Referenciado en 70
- [10] C. d'Amato. *Similarity-based Learning Methods for the Semantic Web*. PhD Thesis, University of Bari, 2007. Referenciado en 70
- [11] D. Inkpen, A. Désilets. *Semantic Similarity for Detecting Recognition Errors in Automatic Speech Transcripts*. Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing. HLT '05. ACMID 1220582, Vancouver, British Columbia, Canada, 49–56 (2005). Referenciado en 70
- [12] H. Al-Mubaid, H. Nguyen. *A Cluster-based Approach for Semantic Similarity in the Biomedical Domain*. Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE. ISBN 1-4244-0032-5, New York, USA, 2713–2717 (2006). Referenciado en 70
- [13] D. de-Francisco, P. Grenon. *Enhancing Telecommunication Business Process Representation and Integration with Ontologised Industry Standards*. Proceedings of the 4th International Workshop on Semantic Business Process Management. SBPM '09. ISBN 978-1-60558-513-0, Heraklion, Greece, 59–66 (2009). Referenciado en 72, 87
- [14] TeleManagement Forum–eTOM. *Enhanced Telecom Operations Map (eTOM) - The business process framework*. <http://www.billingcollege.com/upload/M.3050.1.pdf>, Septiembre de 2009. Referenciado en 72
- [15] TeleManagement Forum–eTOM. *Information Framework (SID) In Depth*. <http://www.tmforum.org/InformationFramework/6647/home.html>, Septiembre de 2009. Referenciado en 72

- [16] A. Bastidas, L. Ordóñez, J. Corrales, C. Figueroa. *Comparación Semántica de Tareas entre Procesos de Negocio de Telecomunicaciones*. V Seminario De Tecnologías emergentes en telecomunicaciones y telemática. TET2010. ISBN 978-958-732-050-3, Popayán, Colomba, 25–30 (2010). Referenciado en 72
- [17] J. Martinez, N. Pérez. *YATOSP: Marco de referencia semántico para el sector Telco*. Telecom I+D Conference. Bilbao, Spain, October 29–31, (2008). Referenciado en 72, 73
- [18] B. Benatallah, M. Hacid, C. Rey, F. Toumani. *Request Rewriting-based Web Service Discovery*. Proceedings of the Second International Semantic Web Conference. ISWC 2003. ISBN 978-3-540-20362-9, Sanibel Island, FL, USA, 242–257 (2003). Referenciado en 79
- [19] E. Komazec, S. Lintner, D. Blamauer, C. Evenson, M. Cimpian. *D1.3 Business Process Modeling Ontology BPMO–SemBiz Deliverable*. <http://www.sembiz.org/attach/D1.3.pdf>, Agosto de 2009. Referenciado en 87
- [20] Object Management Group. *Business Process Modeling Notation–OMG Available Specification v1.1*. <http://www.omg.org/spec/BPMN/1.1/PDF>, Junio de 2009. Referenciado en 87
- [21] R. Chillarege. *Software Testing Best Practices–Center for Software Engineering, IBM Research, IBM Technical Report 1999*. <http://goo.gl/Z0d48>, Marzo de 2010. Referenciado en 87
- [22] M. Yatskevich. *Preliminary Evaluation of Schema Matching Systems–Technical Report, Department of Information and Communication Technology, University of Trento, 2003*. <http://eprints.biblio.unitn.it/490/1/028.pdf>, Febrero de 2010. Referenciado en 88
- [23] J. Lu, S. Wang, J. Wang. *An Experiment on the Matching and Reuse of XML Schemas*. Proceedings of the 5th international conference on Web Engineering. ICWE'05. ISBN 978-3-540-27996-9, Sydney, Australia, 273–284 (2005). Referenciado en 88