**Universidad del Cauca**

**Instituto de postgrados en Ingeniería Electrónica y Telecomunicaciones**

**Programas de Maestría y Doctorado en Ingeniería Telemática**

**Seminario de Investigación**

***Titulo de la relatoría:***

**Towards Automated Biomedical Ontology Harmonization**

**Relator: MSc. Gustavo Andrés Uribe Gómez, estudiante de Doctorado**

**Co-relator: PhD. Diego Mauricio López**

**Protocolante: Ing. Carlos Felipe Estrada Solano, estudiante de Doctorado**

**Fecha:** 10 de abril de 2015

**Hora de Inicio:** 10:00 a.m.

**Hora de Finalización:** 11:00 a.m.

**Lugar:** Universidad del Cauca, Popayán

**Asistentes:**

PhD. Juan Carlos Corrales (Coordinador del seminario)

PhD. Diego Mauricio López (Co-relator)

MSc. Gustavo Andrés Uribe Gómez (Relator)

Estudiantes de Maestría y Doctorado en Telemática (Universidad del Cauca)

Estudiantes de Pregrado de la FIET (Universidad del Cauca)

**Orden del día:**

1. Presentación a cargo del relator.
2. Discusión.

**Desarrollo:**

1. **Presentación a cargo del relator:**

El MSc. Gustavo Uribe inicia la presentación de su artículo de investigación mencionando los co-autores: el PhD. Diego Mauricio López y el PhD. Bernd Blobel, quienes son su director y co-director del proyecto de doctorado, respectivamente. Aclara que el PhD. Bernd Blobel es de la Universidad de Regensburg, Alemania.

**Problema**

El MSc. Gustavo Uribe explica que el presente artículo, y por ende su trabajo de doctorado, abarca el problema de la falta de colaboración e interoperabilidad entre diferentes actores del sistema, quienes poseen un objetivo común. Siguiente, expone, como ejemplo de problema de interoperabilidad, la historia de la Torre de Babel: cada actor tenía su propio idioma y no se podían interpretar entre sí, por lo cual no se pudo construir una torre que llegara al cielo (objetivo común). A partir de esta historia, extiende que este tipo de problemas existen en varios ámbitos de la vida cotidiana, en donde varios actores que manejan su propio conocimiento y habilidades deben colaborar entre sí; por ejemplo, en el área de la salud es bastante común este problema, puesto que se presenta dificultad en la comunicación entre varios médicos con distintas especialidades.

Para finalizar el problema, expone la pregunta de investigación: ¿cómo soportar la colaboración y la interoperabilidad entre los actores utilizando sistemas computacionales?

**Propuesta**

El MSc. Gustavo Uribe comenta que en su trabajo plantean una serie de pasos que se deben resolver para soportar la colaboración:

Paso 1: ¿Cómo colaborar uno a uno? El relator presenta el Ciclo de Colaboración: *(i)* los actores tienen un objetivo común, el cual está relacionado en los datos; *(ii)* los datos provienen del entorno o del mismo proceso realizado, y deben ser interpretados o analizados para obtener la información; *(iii)* a partir de la información se toman decisiones para realizar la acción apropiada; *(iv)* las acciones permiten lograr el objetivo planteado con la colaboración; y *(v)* el motor es el conocimiento.

Seguido, explica que si dos actores manejan el mismo conocimiento, entonces un sistema informático solamente transmite los datos entre ellos. De esta forma, como manejan el mismo conocimiento, pueden interpretar los datos y lograr las acciones apropiadas. Esta interoperabilidad es de tipo técnica, estructural y sintáctica.

Luego expone un caso más complejo al anterior: cuando los actores no manejan el mismo conocimiento, el sistema (además de transmitir los datos) también necesita interpretar los datos para convertirlos en información, permitiendo que los actores puedan interpretar correctamente los datos que no corresponden a su conocimiento. A lo anterior se le denomina interoperabilidad semántica.

Adicionalmente, comenta que si uno de los actores no puede realizar la acción apropiada basándose en la información, el sistema debe proveer el mecanismo para realizar dicha acción, lo cual se conoce como interoperabilidad organizacional.

Para finalizar el paso 1, el expositor resalta que un sistema informático que soporta los anteriores tipos de interoperabilidad tiene la capacidad para lograr la interoperabilidad entre dos actores. Concluye mencionando que hasta el momento existen muchos trabajos alrededor del nivel técnico, estructural y sintáctico, pero que poco se ha logrado en el nivel semántico y organizacional.

Paso 2: ¿Cómo colaborar con cualquiera? El MSc. Gustavo Uribe indica que abordar una colaboración entre muchos actores por medio de la solución anterior, causaría que el sistema sea muy costoso y difícil de manejar. Por lo anterior, explica que se proponen estándares para disminuir dicha complejidad, los cuales se deben definir a diferentes niveles, tales como a nivel de protocolos, estructural, de conocimiento, de tecnologías, etc. De esta forma, resuelve que se evita procesar las individualidades de cada actor, reduciendo la complejidad del sistema.

Paso 3: ¿Cómo lidiar con diferentes estándares? El relator resalta el problema de que el sistema no lidia con un único estándar (por ejemplo, no se tiene un solo idioma). Luego, comenta que a nivel de estándares un estándar se traslapa con otro, es decir, que aunque un estándar define ciertos aspectos y otro define otros, varias veces existen aspectos en común. Por lo tanto, el sistema debe inclinarse a uno de los estándares o soportar ambos. En el mismo sentido, expresa que el sistema también debe ser compatible con las versiones previas del mismo estándar.

Con base en lo anterior, el expositor plantea que para soportar la colaboración se necesita de sistemas adaptativos, flexibles e inteligentes, definiéndolos como sistemas que se pueden adaptar a diferentes estándares y que manejan de forma inteligente el conocimiento utilizado en la colaboración. Después, aclara que podrían existir sistemas que soporten la colaboración sin ser adaptativos, flexibles e inteligentes, pero que sería muy costoso incluir nuevos actores (la idea es incluir nuevos actores de forma fácil).

Paso 4: ¿Cómo crear sistemas inteligentes? El relator expresa el requerimiento de un conocimiento acordado, no uno estadístico o difuso, sino un conocimiento que puede ser formalizado. Siguiente, resalta que las ontologías son el mecanismo más usado para formalizar el conocimiento, con el fin de que el sistema pueda hacer inferencias y cumplir con el Ciclo de Colaboración. No obstante, también menciona que las ontologías son acuerdos que tienen los mismos problemas que los estándares.

Paso 5: ¿Cómo lograr una armonización ontológica entre ontologías solapadas?

Paso 5.1: el MSc. Gustavo Uribe expone cuatro (4) mecanismos para la armonización de ontologías, los cuales se realizan a priori (antes de iniciar el funcionamiento del sistema y que se realice la colaboración): *(i)* fusión, es la más sencilla y mezcla los conceptos y clases, ignorando traslapes y los posibles conflictos entre conceptos; *(ii)* alineación, depura y define los conflictos complicados a manejar desde una ontología, para retirarlos y crear unos nuevos; *(iii)* emparejamiento, define un nuevo conocimiento a partir de múltiples conocimientos, definiendo una nueva ontología que muestra las equivalencias entre las ontologías fuente; y *(iv)* mapeo, es el más complejo y no solamente muestra la equivalencia entre conceptos, sino que además permite saber si un concepto es más general o específico que otro, o si un concepto es parte de otro.

Paso 5.2: el relator muestra las herramientas de análisis actuales más relevantes que soportan los mecanismos presentados, resaltando algunas de sus distintas falencias, por ejemplo: *(i)* los algoritmos no tienen en cuenta la parte arquitectónica, la cual permite seguir un modelo genérico de componentes para detallar dos (2) sistemas como una serie de componentes; y *(ii)* el emparejamiento usado es básico, a nivel lingüístico y terminológico, ignorando completamente el significado del concepto, reduciendo la precisión y añadiendo mayor dependencia a la estructura de la ontología. Dichas ontologías son heterogéneas, lo cual conduce a obtener malos resultados. En este punto, manifiesta que solamente algunos trabajos estudian la parte semántica, pero que son muy limitados e ignoran completamente la lógica descriptiva.

Paso 5.2: el expositor presenta los resultados de las pruebas en los algoritmos, aclarando que el algoritmo de mejor precisión es NodeNames, ya que destruye la arquitectura de la ontología y únicamente mantiene los términos para realizar la comparación. No obstante, explica que se destruye la semántica, puesto que la estructura de la ontología provee algo de semántica.

Por otra parte, expresa que aunque el emparejamiento a nivel terminológico es fácil de implementar, ignora el significado de los conceptos, obteniendo baja precisión y una incorrecta toma de decisiones. El relator considera que un primer paso para solucionar este problema es incluir el uso de las ontologías de dominio, para que las ontologías de aplicación (conocimiento de cada actor representado en la ontología) no se encuentren aisladas. Con base en lo último, explica la importancia de comenzar a estandarizar el proceso de definición de las ontologías para lograr una armonización entre ontologías de dominio y de aplicación, es decir, para que los conceptos definidos en la ontología de aplicación se relacionen con la ontología de dominio. De esta forma, comenta que se utiliza la inferencia para hallar las equivalencias entre los conceptos (con base en la lógica descriptiva) y lograr un proceso de armonización automático.

Seguido, el MSc. Gustavo Uribe aclara que lo anterior es limitado a un único dominio. Por ejemplo, el dominio médico, compuesto por médicos generales y especialistas; los últimos manejando términos que el primero no. El sistema encuentra las relaciones, permitiendo que los médicos generales puedan entender los conceptos que manejan los especialistas.

Posteriormente, plantea qué pasaría en dominios no relacionados. Por ejemplo, el dominio médico y el dominio de políticas; este último especifica las políticas de seguridad de pacientes, de seguridad de información, de manejo de recursos humanos, de manejo de espacio, etc. Ambos son dominios que no se encuentran relacionados en conocimiento, pero que si lo están en el sistema, puesto que los dos se deben definir para que funcione un sistema informático de salud. El relator destaca que para esta armonización se necesita una abstracción mayor, para lo cual definió ontologías de nivel superior, las cuales describen conceptos muy generales.

El relator manifiesta el uso del modelo genérico de componentes para describir cualquier sistema, en cuyo caso son los sistemas de colaboración. El expositor detalla dicho modelo con base en tres (3) perspectivas: *(i)* dominios, para definirlo en múltiples dominios; *(ii)* partes y componentes, para describirlo desde la más general a la más específica; y *(iii)* vistas, para dividir el desarrollo del sistema de software en varias vistas. De esta forma, expresa en términos generales que primero se debe describir el sistema de colaboración, luego representar todo el conocimiento manejado y al final comenzar con el desarrollo de software.

Finalmente, plantea la siguiente hipótesis: desarrollar un sistema utilizando los pasos expuestos, permitiría desarrollar el Ciclo de Colaboración.

**2-**  **Discusión**

El MSc. Camilo Corrales pregunta acerca de cómo garantizar que el sistema propuesto es una herramienta inteligente, lo cual se expuso en el paso cuatro (4) de la presentación. Aclara que pregunta lo anterior porque un experto en el área de inteligencia artificial puede cuestionar las pruebas de que el sistema es inteligente, por ejemplo el Test de Turing, el cual define si una herramienta es o no inteligente. Finalmente, sugiere que la palabra “inteligente” es muy comprometedora y aconseja modificarla o acotar mejor lo que está definiendo.

El MSc. Gustavo Uribe responde que hay muchas formas de hacer un sistema inteligente. Es este caso, se propone un sistema basado en conocimiento que en su código no tiene programado lo que debe realizar, sino que utiliza ontologías para interpretar datos, obtener la información y tomar decisiones para realizar acciones. Agrega que una de las ramas de la inteligencia artificial son las ontologías, por lo tanto un sistema basado en conocimiento escrito en ontologías se puede considerar como una herramienta inteligente.

El PhD. Juan Carlos Corrales solicita detallar más la parte de las pruebas. Es decir, qué tipos de pruebas se realizaron y por qué ese tipo de pruebas.

El MSc. Gustavo Uribe señala que se realizaron dichas medidas porque su interés era la precisión, es decir, verificar si es posible automatizar el proceso de armonizar el uso de dichas herramientas en el proceso de encontrar las equivalencias entre las ontologías y la colaboración. No obstante, añade, que los resultados mostraron baja precisión, razón por la cual no fue posible reutilizar lo existente, puesto que su trabajo requiere que las ontologías sean relacionadas. Por último, expone que se consideraron otras medidas, pero estas últimas no mostraron valores significativos.

El MSc. Diego Durán pide una explicación más amplia acerca de lo que hace la propuesta, incluyendo qué tipo de información inferida provee el sistema, un ejemplo de este tipo de información y para qué sirve esta información.

El MSc. Gustavo Uribe comunica que un ejemplo sencillo es el sistema y el conocimiento para la diabetes. Un sistema personal para manejar la diabetes que se conecta con una historia clínica electrónica, en el cual pueden colaborar tres actores: médico general, médico especialista y paciente, los cuales poseen conocimiento diferente. Por ejemplo, el paciente no entiende los resultados de un examen bacterial, valores que si interesan a los médicos. De esta forma, el sistema solución debe ser capaz de presentar la información de tal manera que pueda ser interpretado correctamente tanto por el paciente como por los médicos, y ocultar la información que no es relevante para algún tipo de actor. En este sentido, adiciona, que el sistema debe adaptarse a cada actor y a su conocimiento en muchos niveles. Por ejemplo, diferentes modelos de datos como conocimiento de bajo nivel de información, tales como el manejo del estándar HL7 y OpenHR.

El PhD. Oscar Caicedo pregunta por el principal aporte presentado.

El MSc. Gustavo Uribe responde que la serie de pasos y la realización de las pruebas, resaltando que este tipo de pruebas no existían, puesto que la mayoría de ellas se hacen para la parte terminológica. El relator manifiesta que muchos estudios proponen algo parecido, pero utilizando algoritmos de terminologías basados en lingüística, los cuales mostraron resultados poco satisfactorios para el área de la diabetes. Luego, el MSc. Gustavo Uribe explica que la estructura de la ontología es muy fuerte en los resultados, razón por la cual dichos algoritmos que trabajan a nivel terminológico no permitirán obtener la colaboración deseada. De esta forma, añade que se propone que los sistemas se conecten y se puedan construir siguiendo la metodología presentada, considerando la parte arquitectónica y los diferentes dominios para obtener un sistema totalmente interoperable.

El MSc. Mario Solarte consulta si el principal resultado es la ontología mediadora o el descubrimiento del conocimiento a través de la ontología.

El MSc. Gustavo Uribe aclara que la ontología mediadora se obtuvo desde el trabajo de maestría, pero ese no es el resultado del sistema ni del artículo. Explica que la ontología mediadora es el insumo del Ciclo de Colaboración, para que el conocimiento no sea solamente del dominio, sino también de los actores.

El PhD. Oscar Caicedo comenta que es difícil captar el aporte presentado, puesto que se presenta un problema y rápidamente se profundiza en los detalles de cada paso, por lo que se pierde el concepto general de la presentación. Por otra parte, adiciona que no se observa claramente la relación entre la propuesta y las pruebas de algoritmos realizadas. Aclara que el enlace debe existir en el artículo, pero que en la presentación no se conectó. Por lo tanto, sugiere mejorar el guión de la presentación.

El PhD. Juan Carlos Corrales complementa mencionando que desde su punto de vista faltó motivar más el trabajo, y a partir de la motivación tener una percepción general del problema que se quiere solucionar. Por ejemplo, si se motiva por el lado del dominio de la diabetes es posible entender mejor el problema que se quiere solucionar. El PhD. Corrales añade que la propuesta es muy buena, pero que en el transcurso de la presentación se desengancha y se pierde, tal como las pruebas mostradas. Explica que aunque ya es un trabajo presentado y aprobado, se exigen presentaciones “*top*” en el seminario, especialmente por el público heterogéneo del mismo. Con base en lo anterior, sugiere realizar una presentación que oscile entre la visión general y la especialidad, incluyendo explicaciones con ejemplos, ya que son la forma más sencilla para darse a entender al público.

Finalmente, el PhD. Oscar Caicedo recomienda modificar el término “*decisionmaking*” por “*making-decision*”.

Se termina la sesión.