**Universidad del Cauca**

**Facultad de Ingeniería Electrónica y Telecomunicaciones**

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

**Seminario de Investigación**

**Modelo de Agrupación de Procesos de Negocio Basado en Representación Multimodal**

**Conferencista: MSc Ing Hugo Armando Ordóñez, estudiante de Doctorado en Ingeniería Telemática**

**Co-Relator: Dr. Juan Carlos Corrales**

**Protocolante: Liseth Viviana Campo A., estudiante de Maestría**

**Fecha:** 30 de mayo del 2014

**Hora inicio:** 11:05 a. m.

**Hora fin:** 11:54 a. m.

**Lugar:** Salón de posgrado, IPET, Universidad del Cauca, Popayán

**Asistentes:**

Dr. Juan Carlos Corrales, coordinador del seminario

Ing. Hugo Armando Ordoñez, estudiante de Doctorado, relator

Estudiantes de Maestría y Doctorado en Ingeniería Telemática

Estudiantes de pregrado de la FIET

**Orden del día:**

1. Presentación a cargo del relator
2. Intervención del co-relator
3. Discusión
4. Conclusión

**Desarrollo**

1. **Presentación a cargo del relator**

El ingeniero Hugo Ordoñez indica que en la presentación pretende presentar el avance de la tesis doctoral con la dirección del doctor Juan Carlos Corrales y codirección del doctor Carlos Alberto Cobos. La agenda incluye el contexto general, problema, objetivos, actividades, metodología y resultados actuales.

En el contexto se tiene que un modelo de proceso de negocio representa unas eventualidades que están enmarcadas en una organización: En la organización se relacionan los socios que en el modelo son representados como roles, cómo es la participación o el flujo que ellos conciben dentro del proceso de negocio. Lo procesos de negocios son modulares y extensibles, es decir que permiten crear nuevos procesos de negocios partiendo de uno. El contexto se concentra en la búsqueda de modelos de procesos de negocio para lo cual existen cuatro técnicas que incluyen: lingüística (análisis del texto) y la semántica que complementa la lingüística para dar sentido al contenido del texto. Otro tipo de técnica es el análisis de la estructura o la distribución de las tareas dentro de un modelo de proceso de negocio y la última técnica, analiza el comportamiento (identifica cuales caminos se siguieron). Cualquier técnica se puede generalizar como tomar un modelo de proceso de negocio o parte de él y encontrar similares que sean susceptibles a cambios y/o adecuaciones a los requerimientos de una organización determinada.

El contexto abarca el concepto de clustering el cual es un proceso de agrupación por algún tipo de afinidad que se encuentra en la información. En el caso particular en el proceso de negocio se comparte información que sería detectada y agrupada por alguna característica (textual, estructural o una combinación).

El otro foco del contexto es la búsqueda multimodal. Se refiere a la representación de información en varios tipos de formatos y realizar la búsqueda sobre uno o sobre un conjunto de estos. El resultado de la búsqueda se coloca en una estructura para desfragmentarla según algún tipo de característica y luego recuperar la información. Esta búsqueda se complementa con tipos de información diferentes. El expositor describe como ejemplo la información que se analiza a partir de imágenes y texto, que al combinar el análisis de las características independientes de los dos formatos, es posible obtener una representación del objeto de estudio más detallada, ampliar el número de consultas sobre la búsqueda y por tanto tener resultados más precisos.

Para el planteamiento del problema el estudiante plantea las características que deben tener los procesos de negocio que en la actualidad están orientados a ejecutarse en una arquitectura orientada a servicios (SOA), la cual es modular, reutilizable y abierta. Bajo esta arquitectura se plantean requisitos para el diseño de un modelo de negocio entre ellos la reproducibilidad, escalabilidad, estructuración (definir roles) y documentación de las dependencias de una organización. Todas estas características toman relevancia ya que el fin del proceso de negocio es la trasferencia del conocimiento a una organización en todos sus niveles.

En el diseño de escalabilidad de un proceso de negocio surge la minería de procesos de negocio, esta permite analizar en los repositorios registros históricos para descubrir el proceso de negocio que está ligado al análisis de lo que se ejecutó y puede ser representado por un flujo de roles, permite chequear los flujos, modificar los componentes y extenderlo para nuevas funcionalidades.

En los trabajos más representativos en la búsqueda de procesos de negocio se encuentran situados en el tipo lingüístico, que generalmente extraen las características de un proceso de negocio, ubicar en un repositorio los procesos de negocios que más se asemejen y representarlos en una la lista de resultados. Las brechas que resalta el Ingeniero Hugo Ordoñez en estos trabajos son: solo consideran información textual y en la estructura y el comportamiento no se contempla el tipo de actividades.

En la búsqueda de procesos de negocio hay técnicas que aplican solo dos reglas de asociación (técnica de minería de datos) para analizar las actividades que posiblemente fueron ejecutadas dentro del proceso de negocio. En estas propuestas se identifican las siguientes brechas: siempre se realiza sobre ejecuciones previas, solo realiza búsqueda estructural (análisis de nodos, no en la lingüística) y el tiempo de ejecución es elevado. Otras técnicas de búsqueda se basan en algoritmos genéticos, estos se basan en tomas las características del proceso de negocio y con una serie de algoritmos heurísticos determinan los procesos de negocio que se asemejen más al proceso en consulta. Las limitaciones que se encontraron es que siempre trabajan con distancia ponderada basada en heurística, inician leyendo archivos log (para recuperarlo debe haber sido ejecutado), se debe calcular más variables lo que implica más demanda de tiempo. Por otro lado está la búsqueda por clustering, las cuales agrupan los hallazgos por alguna semejanza. Sus brechas se remarcan porque no tienen en cuenta el tipo de actividades no sus relaciones y el proceso de categorización solo ocurre una vez.

El estudio de las técnicas y la literatura llevaron a la siguiente pregunta de investigación: ¿Cómo formar grupos o familias de procesos de negocio mediante una búsqueda multimodal que permita a los modeladores de procesos de negocio obtener resultados más relevantes que los reportados con sistemas actuales?

Para responder a la pregunta de investigación se plantea los siguientes objetivos:

General: Proponer un modelo de agrupación de procesos de negocio basado en una representación multimodal usando codebooks[[1]](#footnote-1) complementados con características textuales buscando obtener resultados más relevantes a las búsquedas realizadas por los desarrolladores.

Específicos:

* Definir un diccionario de codebooks modelados en procesos de negocio con BPMN.

*Actividades asociadas:* Desarrollar un estado del arte de la minería de procesos de negocio, caracterizar componentes conceptuales, definir la estructura del codebook y los conceptos para este, definir una representación de N-Gramas más extendida y evaluar los resultados con una base de datos y cuál de las representaciones permitirá obtener la mejor búsqueda.

* Establecer un modelo de representación multimodal que incorpore el diccionario previamente definido y las características textuales de los modelos de procesos de negocio.

*Actividades asociadas:* Para realizar la normalización de los componentes secuenciales del proceso de negocio unido con las características textuales de cada proceso para ampliar el espacio de búsqueda, se realizó un estado del arte detallado sobre búsqueda multimodal, caracterización del comportamiento de los procesos de negocio, definición de un modelo de representación del texto de cada uno de los procesos de negocio, definición de una forma de indexación multimodal, definición de un umbral de precisión, creación de un método que haga uso del índice del análisis multimodal y de esta manera recuperar los datos representados en el índice de búsqueda.

* Proponer un algoritmo que agrupe los procesos de negocio basado en el modelo de representación multimodal.

*Actividades asociadas:* Con un modelo de ponderación conceptual es posible identificar los procesos de negocio que son similares al proceso de consulta y luego se generaría una lista de resultados que pueden ser agrupados por las características más compartidas (Aporte del Doctor Leandro Krug). Para desarrollar este objetivo se realizó una pasantía en Brasil con la asesoría del doctor Leandro Krug en la Universidad Federal de Rio Grande del Sur. Detalladamente se elaboró un estado del arte sobre clustering de procesos de negocio, se analizó los algoritmos de clustering aplicables a procesos de negocio, se modificó el algoritmo aportado por el doctor Leandro, se construyó un mecanismo que aplica el algoritmo que agrupa los resultados entregados por el modelo de búsqueda multimodal, se realizó una formación de búsqueda automática de procesos de negocio con las medidas de silueta, cohesión y acoplamiento y se evaluó la formación manual de grupos de procesos de negocio.

* Medir la calidad entregados por el modelo a través de evaluaciones experimentales con medidas tales como: precisión y medida, en colecciones cerradas de procesos de negocio y experimentos con usuarios.

*Actividades asociadas*: Análisis de la herramienta prototipo, definición de las medidas de precisión y rendimiento al tiempo, evaluación del modelo propuesto con usuarios, se construyó un repositorio cerrado (uno de los mayores aportes) y se evaluó la formación de grupos.

La metodología con la que se desarrolló el trabajo de investigación se planteó en base a la propuesta por un profesional de ingeniería, el doctor Serrano. Se constituye de dos fases. La primera es acerca de la investigación documental, que incluye la fase preparatoria, la fase descriptiva y la fase de construcción teórica global, posteriormente se plantea el modelo de construcción de soluciones, haciendo un estudio de pre-factibilidad, la formulación del proyecto, la ejecución y la validación.

Desde el principio de la propuesta se planteó con los tutores escribir artículos por cada objetivo con el fin de hacerlos medibles. De aquí los siguientes resultados: Para el objetivo uno se escribió un artículo titulado “Modelo multimodal para búsqueda de procesos de negocio” en la revista *Information Science (Estados Unidos),* la cual se encuentra en categoría A1 en la clasificación de Colciencias. Adicionalmente existe un artículo que está en revisión del doctor Leandro titulado “Mejoramiento de la recuperación de procesos de negocio usando caracterización y búsqueda multimodal”, en este se incluye los objetivos uno y dos anteriormente nombrados. Este documento se pretende enviar a gestión y colección de procesos de negocio 2014, categoría A2 en Colciencias.

Para el objetivo dos, se tiene el artículo titulado “Modelo de descubrimiento de procesos de negocio basado en trazas de ejecución”, esta fue una tesis de pregrado dirigida por el estudiante de doctorado. Fue publicado en el año 2012 como un capítulo de un libro de la universidad del Alkala. Adicionalmente se generó un método de indexación, llamado “Indexación de procesos de negocio basado en similitud de casos de ejecución” fue publicado en 2014 (ACM).

Para el objetivo tres, se consideró un evento en cuba, *convención científica de ingeniería,* en el quefue aceptado un artículo pero nunca se tuvo respuesta de la revista. Ante esto se publicó un artículo (complementación de la tesis) titulado “Agrupación de modelos de procesos de negocio basado en representación multimodal y lógica difusa[[2]](#footnote-2)”, este será enviado a la revista *Information Systems,* es categoría A2 en Colciencias. Y como complemento se tiene el artículo “Una propuesta para la recuperación procesos de negocio basado en búsqueda multimodal clustering”, este fue enviado a la revista IEEE América Latina, categoría A2 en Colciencias.

Para la evaluación de la propuesta aquí planteada se creó un repositorio cerrado, con una metodología para evaluar motores de búsqueda de procesos de negocio. En esta evaluación se contaron con 59 usuarios, se desarrolló una plataforma para que los usuarios evaluaran la búsqueda y se les permitiera formar grupos manualmente.

El expositor muestra la herramienta y añade que se publicó la construcción del modelo del repositorio de procesos de negocio sobre una metodología colaborativa. Este fue publicado en IEC 6 2014 (Springer).

1. **Intervención del Co-Relator**

En este punto el doctor Juan Carlos aclara que se muestra solo el 80% de avance en el trabajo de doctorado ya que hay varios papers sometidos y solo se está esperando respuestas para iniciar el proceso de sustentación. Agrega también que este trabajo es de los últimos registros de su trabajo de Doctorado y que este trabajo fue presentado a la convocatoria de la vicerrectoría de investigaciones con lo cual se pudo soportar viajes y matriculas.

Se le da también un reconocimiento al apoyo recibido por el profesor Carlos Cobos del departamento de Sistemas y el doctor Leandro Krug de Brasil.

El doctor Juan Carlos motiva a los estudiantes de posgrado, indicando que en la actualidad hay una gran demanda de doctores, como es el caso del Ingeniero Hugo Ordoñez que trabaja actualmente con la Universidad de San Buenaventura. Además resalta que las revistas en las que se publique deben ser relevantes como Atis, IEC 6, IEEE, etc. Finalmente aclara que la tesis tiene posibilidades de ser honorifica dependiendo de la defensa del estudiante.

1. **Discusión**

Las respuestas aquí presentadas son dadas por el Ingeniero Hugo Ordoñez.

El Ingeniero Camilo Corrales: ¿Para elegir el algoritmo de clustering solo se utilizaron tres métricas?

Rta/ *“Se utilizaron seis. El algoritmo de lógica difusa fue seleccionado porque son algoritmos orientados a cadenas de texto, aunque se pueden definir clases. Las clases exigen un procesamiento para establecer que clases se van a definir por cada uno de los elementos. En este algoritmo se agrupa sin necesidad de hacer clases en el contenido y lo que permite la establecer la medida de similitud del profesor Leandro es inferir que tan similares son dos ítems dentro de una lista de resultados con base al número de términos y de su posición”.*

El doctor Juan Carlos Corrales: ¿Qué algoritmos evaluaste?

Rta/ *“Se evaluó LINGO, STC, KAMILS, STAR…En total fueron siete”*

El Ingeniero Camilo Corrales: ¿En las métricas para seleccionar se observó que tan representativo era el proceso?

Rta/ *“Que tan cercanos. Es decir la cohesión y el acoplamiento. La cohesión se orienta a observar si tengo cinco elementos en un grupo quiero saber que tan parecidos son ellos y el acoplamiento analiza que tan distribuidos están los grupos. La silueta analiza que tan dispersos están los componentes entre ellos”.*

El Ingeniero Camilo Corrales: ¿Los cluster se definen manualmente?

Rta/ *“Es automático. Se obtiene de la matriz de elementos por documento o por proceso de negocio, se establece entonces un índice de la matriz, que contiene los modelos”.*

El Ingeniero Camilo Corrales: ¿Cuántos procesos de negocio se utilizaron?

Rta/ *“Se tomaron 100 procesos de negocio modelados con WSDL y se llevaron a BPMN. De esos 100 que están en la representación visual se llevaron a la representación formal en XPDL (mapa del proceso de negocio), con lo cual es posible leerlo”*

1. **Conclusiones**

El coordinador del seminario da fin a la discusión.

Se termina la sesión.

1. El expositor aclara: “Los codebooks son definidos o modelados a través de unidades básicas estándar de comportamiento secuencial que están representadas dentro de los procesos de negocio ” [↑](#footnote-ref-1)
2. Es una modificación del trabajo del doctor Leandro Krug, que incluye etiquetas de los procesos de negocio con base al contenido [↑](#footnote-ref-2)