

## AN ARCHITECTURE FOR LEARNING TO REDUCE THE RISK OF FAILURE OF WEB SERVICES FOR ITS COMPOSITION

## UNA ARQUITECTURA DE APRENDIZAJE PARA DISMINUIR EL RIESGO DE FALLA DE SERVICIOS WEB DURANTE SU COMPOSICIÓN

Ing. Byron Enrique Portilla Rosero, MSc. Jaime Alberto Guzmán Luna

**SINTELWEB: Grupo de Investigación “Sistemas Inteligentes en la Web”**

Escuela de Sistemas, Universidad Nacional de Colombia – Sede Medellín.

Cll. 65, No. 80 -223, Bloq. M8A, Medellín, Antioquia, Colombia.

E-mail: {beportillar, jaguzman}@unal.edu.co

**Abstract:** This paper describe machine learning architecture to management failure during Web services composition. We illustrate two important aspects: time and security. For this reason, this architecture uses learning machine models based on relational decision trees that allow through experience; obtain relevant information on the performance of services in order to be evaluating in future compositions. This information will decrease the failure risk caused by Web services.

**Keywords:** Semantic Web services, services composition, management failure risk, decision trees.

**Resumen:** Este documento describe una arquitectura de aprendizaje máquinas para el manejo de fallas generadas durante el proceso de composición de servicios Web, actuando principalmente bajo dos aspectos fundamentales como son: el tiempo y la seguridad. Para ello, esta arquitectura utiliza modelos de aprendizaje de máquinas basados en árboles de decisión relacionales, que permiten a través de la experiencia, obtener la información pertinente del comportamiento de los servicios de manera que pueda ser evaluada y categorizada con el fin que en futuras composiciones, se disminuya el riesgo de falla presentado por los servicios Web.

**Palabras clave:** Servicios Web Semánticos, composición de servicios, manejo de fallas, árboles de decisión.

### 1. INTRODUCCIÓN

El mundo de la Web se caracteriza por ser un entorno dinámico, donde la información que allí se encuentra, es manejada a través de múltiples sistemas de software definidos como servicios web, los cuales permiten el intercambio de datos y su funcionalidad en la red. Debido a este tipo de ambiente, la información está sujeta a cambios ocasionados por un sin número de causas afectando directamente al conjunto de sistemas que lo comparte. Esas causas pueden ser ocasionadas por

el mismo servicio como es el caso de accesos a datos privados, mala entrega de la información debido a datos incompatibles o bien, fallos relacionados con la programación del sistema; es decir, encontrar errores de programación o errores en el diseño de estos. Asimismo, se puede encontrar problemas en cuanto a tiempo requerido para desarrollar una o varias operaciones de manera simultáneas por el sistema ocasionando retardos en la toma de decisiones al buscar una respuesta en particular. Por otro lado, se encuentran presentes los errores del entorno, esto es debido al

concepto de redes donde los problemas de caídas de servidores, fallos en la energía, sobrecarga de redes, accesos prohibidos en protocolos de seguridad, evitan el buen desempeño de los servicios Web, generando la no satisfacción de requerimientos necesarios para la consecución de una solución, ya sea por parte del usuario o de un servicio que requiere parte de información para realizar otra serie de procesos.

Para realizar una buena composición de servicios, no necesariamente se requiere hacer un estudio minucioso sobre los procesos de evaluación de cada servicio y su entorno. Sin embargo, se hace de gran utilidad el uso de prácticas que permitan identificar cuáles fueron las fallas o fracasos de una composición y para ello en la literatura, se ha identificado una serie de parámetros no funcionales que representan características poco utilizadas en algunos procesos de composición pero que revelan detalles que soportan el uso de un determinado servicio en un espacio determinado de un plan de composición. Estos parámetros definidos en la literatura como QoS pueden verse desde dos perspectivas distintas: en primer lugar, estos pueden ser como la consecución de objetivos adicionales sin ninguna influencia o interés o, en segundo lugar estos parámetros están atados a la identificación de datos ocultos para el usuario pero no para el sistema; los cuales, hacen posible adquirir una información relevante en la selección de un servicio en la composición de estos. (Kahn et al. 2002).

Este artículo, está estructurado de la siguiente manera: en la sección 2, se presenta una definición y categorización de los principales criterios evaluados en el entorno de los servicios Web; en la sección 3, se definen los criterios evaluados por la arquitectura de aprendizaje propuesta; en la sección 4, se describe la arquitectura de aprendizaje basada en los criterios expuestos en la sección inmediatamente anterior y finalmente en la sección 5 se presentan las conclusiones y trabajos futuros.

## 2. CONCEPTUALIZACIÓN DE LOS CRITERIOS DE CALIDAD

Los trabajos revisados en la literatura han identificado un conjunto de criterios que describen algunas características entorno a representar la calidad de los servicios Web y como a partir de ésta, en algunos casos como (Kokash, 2007) ha sido de suma importancia para la identificación de fallas ocurridos en los entornos Web.

En trabajos como el de (Ran, 2003) en el cual se presenta un conjunto de parámetros a través de los cuales, genera un modelo de descubrimiento de servicios Web. Se aclara la importancia del uso de criterios de calidad debido a que este modelo es una clara representación de la satisfacción y confiabilidad de los consumidores, sobre la calidad del servicio que sea descubierto. Esto se supone daría confianza a los consumidores de servicios Web, acerca de la calidad del servicio que están a punto de invocar. Entre los criterios evaluados se encuentra: la escalabilidad, la capacidad, la rentabilidad, la fiabilidad, la disponibilidad, la robustez, la flexibilidad, el manejo de excepciones y la precisión.

Asimismo, en (Cardoso et al., 2002), se define un conjunto de criterios de calidad utilizados para identificar flujos de información que determinan la calidad. Estos criterios son el costo, el tiempo, la disponibilidad y la reputación.

Los criterios anteriormente presentados han sido de gran ayuda en la consecución de información adicional de los servicios la cual, ha sido fuente determinante a la hora de realizar una selección de servicios. Asimismo, han sido unos de los más estudiados en la literatura. Sin embargo, existen otros criterios que también pueden servir como herramientas determinantes en la selección de servicios y procesos de composición, y que no han sido muy utilizados. Tal es el caso de criterios orientados a la seguridad. Para (Ran, 2003), la seguridad mide la fiabilidad de los mecanismos de seguridad y aplicación. De manera que su utilización, permita adquirir características adicionales de los servicios, que los criterios mencionados anteriormente no lo podrían alcanzar.

Dentro de este criterio se encuentra una serie de definiciones.

**Autenticación:** define el funcionamiento de los servicios de autenticación determinando quién puede acceder a servicios y datos.

**Autorización:** funcionamiento de los servicios de autorización de manera que solo los autorizados puedan acceder a servicios protegidos.

**Confidencialidad:** define el tratamiento de los datos de manera que solo puedan ser accedidos y modificados por servicios autorizados.

**Rendición de cuentas o Responsabilidad:** el proveedor del servicio es el responsable del funcionamiento y manejo del servicio.

**Trazabilidad y auditoría:** hace un seguimiento del historial de un servicio cuando una solicitud de servicios fue hecha.

**Cifrado de datos:** mide el funcionamiento de la encriptación de los datos de un servicio.

**No Refutación:** define la decisión de negar un servicio o solicitar los datos después de los hechos.

Los criterios expuestos en los anteriores trabajos han sido definidos en función de los servicios; esto, es viendo al servicio como un todo, de manera que sus cálculos determinan la calidad global de este. Sin embargo, la calidad también puede estar ligada a la información asociada a la veracidad de los servicios y no a sus comportamientos. Para ello se ha identificado una serie de características que pueden ser determinantes al momento de asociar una respuesta.

En (Wilkinson et al., 1997) (Bennett et al, 1996), se estudian los criterios de calidad en función de los datos, acceso y usabilidad, identificación de Recursos y Documentación, Identificación del Autor y autoría, Estructura de la Información y Diseño, Relevancia y Alcance del Contenido, Validez del Contenido, Precisión y Balance de contenidos, Navegación dentro del documento, Calidad de los Enlaces, Estético.

Estos aspectos, permiten controlar la calidad de los datos que contienen los servicios de manera que se alcance una mejor satisfacción por parte del usuario final.

Un aspecto de suma importancia dentro de este análisis, es la identificación de fallas que pueden presentar los servicios entorno a los criterios anteriormente definidos. Estos fallos, determinan el riesgo de una composición. Por lo tanto, el riesgo es determinante en estos procesos.

Según (Kokash, 2007), el riesgo es definido como una probabilidad de una amenaza multiplicada por la respectiva magnitud de su impacto. Según (Roy, 2004), los sistemas del mundo real presentan un factor de riesgo el cual puede ser calculado a través de la suma de de los riesgos por cada amenaza particular.

Para (Verdon y McGraw, 2004) y (Freimut et al., 2001) el objetivo del manejo de riesgos ayuda a los diseñadores a gestionar proyectos de software dentro de los plazos y límites presupuestarios.

### 3. SELECCIÓN DE LOS CRITERIOS

Para la selección de los criterios aplicables en entornos de composición y aprendizaje de servicios Web, se tuvo en cuenta el nivel de impacto y su categorización dentro del estado de riesgos. Es decir, la influencia de la información generada por el criterio en los procesos de composición y el nivel de riesgo que generaría.

Para ello, se consideró la clasificación criterios de (Hwang et al., 2007) y la clasificación de riesgos de (Kokash, 2007) teniendo en cuenta el nivel de acción del criterio en función del comportamiento del servicio. La tabla 1. Describe los criterios a evaluar.

#### Análisis de Criterios

**Relevancia De La Información:** este criterio evalúa la importancia que tienen los datos de los servicios en función de optimizar el tiempo utilizado en los procesos de composición. Es decir, disminuir el riesgo de falla al momento de instanciar los servicios de manera que se adquiera aquella información estrictamente necesaria para alcanzar el correcto funcionamiento del servicio. Ecuación 1.

$$R_{inf} = \min((C(1,0)) \in SW) \text{ tal que } (WS) \text{ satisfy } E(WS) \quad (1)$$

Su estructura se presenta en la figura 1.

Tabla 1. Selección de los criterios a evaluados

Criterio	Categoría	Nivel	Riesgo
Autorización: acceso a datos en función de la seguridad.	Seguridad	Servicio	Cambios en las especificaciones Comportamientos inesperados de los servicios
Relevancia y Alcance del Contenido	Fidelidad	Información	Comportamientos inesperados de los servicios Cambios en las especificaciones

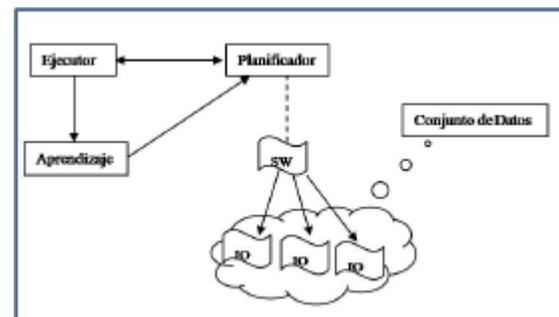


Fig. 1: Criterio relevancia de información

La evaluación de la información presume la menor cantidad de combinaciones datos de entrada y salida (I,O) que satisfagan correctamente la ejecución del servicio Web de manera que cumpla todas las características requeridas utilizando menos requerimientos computacionales o información innecesaria (marcos de datos).

**Autorización:** definida como acceso restringido a servicios protegidos, de manera que no se puede acceder a ningún dato contenido en el servicio. Este criterio se ejecuta a nivel del servicio pero asociado directamente a los ítems internos; esto se debe a que el servicio debe responder a un conjunto de mensajes *soap* los cuales, se encuentran intervenidos por protocolos *ws-security*.

La poca profundidad que se ha dado a este tema, hace posible realizar un trabajo donde se pueda identificar aquellas fallas que ocurren en función de la ejecución de los servicios. En la web, la infinidad de servicios que pueden ser utilizados en una composición, pueden estar supeditados al uso de políticas de seguridad, privacidad, calidad, entre otras, lo que dificulta el acceso a la indagación de los datos que en ellos recae. La estructura de evaluación se presenta en la figura 2. Aquí se evalúa la petición y/o respuesta del servicio o hacia el servicio determinando el correcto acceso o la violación de accesibilidad a la información. Para ello, se identifica la optimización de los datos, definiendo si el servicio puede o no soportar y retornar los datos que el ejecutor requiere para la ejecución normal del servicio.

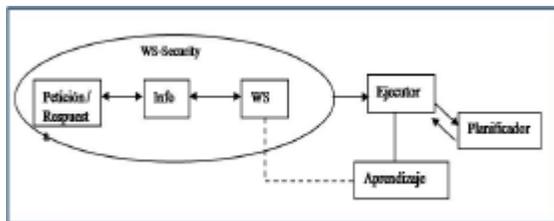


Fig. 2: Captura del criterio autorización

El aprendizaje de este criterio está asociado a la identificación de aquellos servicios que consigan el conjunto de efectos solicitados pero que a su vez puedan ser accedidos y manipulados si es el caso. Ecuación 2.

$$Acc = \text{satisf}\{eff\} \forall eff \in SW \text{ tal que } \{eff\} \neq ws.\text{security} \quad (2)$$

Donde la satisfacción de los efectos por parte del servicio web (SW), puedan ser utilizados durante el proceso de ejecución y soportar la manipulación de ellos durante todo el procesos.

### 3.1 Algoritmo De Aprendizaje

Uno de los algoritmos encontrados en la literatura que se adaptan a modelos de planificación orientados a la composición de servicios son los árboles de decisión. Estos, son una técnica de aprendizaje de máquinas utilizada para la selección de una respuesta adecuada dentro de un conjunto de posibilidades. Utiliza el algoritmo Top-Down Induction of Decision Trees (TDIDT) (Quinlan, 1986). Esta técnica construye árboles de decisión dividiendo los ejemplos entrenados de acuerdo a los valores de un atributo seleccionado el cual, minimiza el valor de la varianza según la predicción de la variable. Los árboles de decisión relacional son una actualización de la lógica de primer orden de los árboles de decisión. La diferencia radica en el uso de árboles relacionales los cuales son implementados con lógica de predicados. (Blockeel, H. y De Raedt, 1998). Describiendo cada árbol como un conjunto de hechos lógicos.

La diferencia principal de este algoritmo de aprendizaje con respecto a otros es el uso de reglas de producción, donde el principal objetivo es lograr una interacción entre el usuario y la máquina de manera que exista un complemento de las partes y así estudiar los casos evaluados.

Este aprendizaje usa experiencias y preconceptos del pasado para mejorar la habilidad de actuar en el futuro. En la comunidad de la planificación IA, el aprendizaje de máquinas es visto como un poderoso mecanismo que dota a este tipo de agente de gran autonomía y flexibilidad, con el fin de compensar el conocimiento incompleto que el diseñador le brinda acerca del mundo que el agente enfrentará (Jiménez, 2009).

## 4. DESCRIPCIÓN DE LA ARQUITECTURA DE APRENDIZAJE

Con base en lo mencionado en la sección anterior, se plantea una arquitectura de aprendizaje que apoye la composición de servicios Web a través del control de riesgos que los servicios pueden generar durante el proceso de ejecución de los mismos, teniendo en cuenta sus parámetros de calidad y algunas características del entorno Web. Para ello, la arquitectura plantea un conjunto de elementos que soportan los datos generados durante la ejecución de los servicios Web, así como la proliferación de estos, en el contexto de identificación, validación, formulación y almacena-

miento de los mismos y a partir de estos procesos, obtener los datos requeridos que caracterizan a cada uno de los servicios Web evaluados. Es decir, se obtiene aquellos datos considerados como relevantes en la construcción de información que permita identificar los valores de error producidos por los servicios Web y/o su entorno.

En la figura 3, se aprecia los elementos utilizados por la arquitectura de aprendizaje. Aquí se observa los elementos de identificación o ingreso de datos, el proceso de validación de los mismos, el módulo de formulación-aprendizaje y el almacenamiento.

Con esta arquitectura, se pretende aprender a partir de una serie de ejecuciones, el comportamiento de los servicios Web teniendo en cuenta el monitoreo de sus salidas en función de sus entradas. Es decir, reemplazar aquellos datos que sean determinados como fallos y mejorar la composición de los servicios. Para ello, se integran los conocimientos de los QoS soportados por la composición de servicios (Jurisica et al, 1998), (Zeng et al, 2003) y sobre los cuales actúa la arquitectura de aprendizaje, evaluando principalmente aquella información que haga posible determinar los riesgos de un servicio a través de su relación en la Web y su interacción con otros servicios.

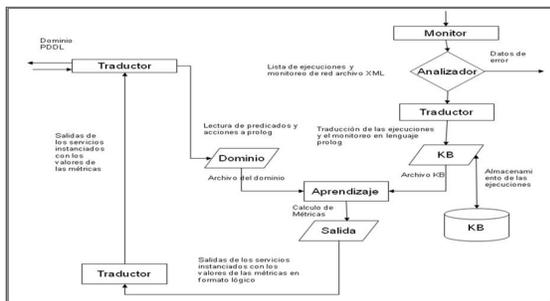


Fig. 3: Arquitectura de Aprendizaje

La arquitectura cuenta con un módulo de aprendizaje, el cual se apoya en el aprendizaje inductivo, (árboles de decisión). Estos, permiten a partir de una serie de ejemplos y utilizando decisiones secuenciales basada en el uso de resultados y probabilidades asociadas, guiar los datos que evalúan a cada servicio Web. El algoritmo utilizado para tal fin es C4.5; este algoritmo, es una extensión mejorada del ID3, el cual genera un árbol de decisión a partir de los datos mediante particiones realizadas recursivamente utilizando la estrategia (depth-firts). (Quinlan et al, 1993).

La arquitectura de aprendizaje consta de un conjunto de elementos que interactúan entre sí, a partir de la obtención de información proveniente del mundo, y de información sobre el dominio en el que se está trabajando.

#### 4.1 Adquisición de la Base de Conocimientos

En primer lugar está el proceso de obtención de la información a través del ejecutor de servicios. Este se encarga de obtener toda la información real de los servicios y su influencia por el entorno. Con el fin de instanciar esos datos relevantes en la composición de servicios, la arquitectura incorpora un monitor localizado dentro del ejecutor de servicios, que permita hacer un seguimiento paso a paso a la ejecución de los servicios y donde haga posible analizar las fallas que se producen en un determinado proceso en el tiempo. Por lo tanto, el monitor es el encargado de abstraer aquella información considerada principal para el desarrollo del proceso de aprendizaje.

La información resultante de este proceso es un documento XML, el cual es enviado a un traductor para ser evaluado por la arquitectura de aprendizaje.

Sin embargo, antes de esta etapa, un analizador es el encargado de verificar la consistencia de este documento y determinar si su contenido está acorde a las ejecuciones de los servicios.

Una vez que los datos son capturados, pasan por un módulo de traducción. Este elemento es el encargado de leer la codificación del monitor y trasladarla a un lenguaje de codificación lógica. El traductor lee un archivo XML de donde obtiene las sentencias de respuesta de cada servicio así como el estado de la red. Esta información es codificada al lenguaje prolog, generando un archivo de lectura rápida que representa la base de conocimientos de la arquitectura de aprendizaje.

#### 4.2 Adquisición de la Información del Dominio

Con el fin de obtener la información del dominio requerido para inducir el aprendizaje, la arquitectura requiere de un archivo donde se describe el dominio que es asumido por los servicios a utilizar. Este archivo, es un documento xspddl por lo tanto, se hace necesaria la utilización de un segundo traductor. En este caso el traductor codifica las expresiones descritas en xspddl en instancias de inducción a un lenguaje prolog, generando un nuevo archivo que representa las especificaciones de inducción utilizadas por la

herramienta de aprendizaje para la construcción de relaciones lógicas y realizar el proceso de aprendizaje.

#### 4.3 Aprendizaje

La parte más importante de la arquitectura se encuentra dentro de este módulo. Es aquí donde los datos capturados en los procesos anteriormente descritos, serán evaluados para obtener la información necesaria que permita determinar el factor de riesgo de falla de un servicio Web, en un determinado espacio en el tiempo.

La evaluación consiste en determinar patrones de ejecución de los servicios donde se evalúan los criterios descritos anteriormente y sobre los cuales se aplica el aprendizaje.

Este módulo recibe dos archivos: la información del dominio y la información de la base de conocimientos. En el primer archivo se encuentra la definición de las acciones, que determina cual será el objetivo a predecir; igualmente, se encuentra la definición de los predicados; los cuales, determinan los patrones de inducción. En el segundo archivo se encuentra los datos generados durante la ejecución del servicio. Esta es la base de conocimiento sobre la cual se infieren los patrones que tuvo ese servicio al ser ejecutado y así determinar su comportamiento.

#### 4.4 Salida del Modelo

La información generada durante el aprendizaje, hace posible medir la calidad de servicio, identificando los costos asignados a sus salidas.

Por lo tanto, se requiere de un último traductor que interprete la información obtenida durante el proceso de aprendizaje, y la traslade al documento del dominio; este dominio será utilizado por un planificador en la generación de nuevas composiciones.

### 5. TRABAJOS RELACIONADOS

Dentro de la composición de servicios Web un conjunto de técnicas entre las cuales se encuentra la planificación en IA, síntesis de programas, cálculo de situaciones, máquinas de estado finito entre otras, han sido estudiadas y ejecutadas. Sin embargo, aunque no se puede indicar cuál de estas soluciones prevalece sobre las otras, los trabajos basados en las técnicas de planificación IA, han

mostrado ser lo suficientemente adecuados desde que los servicios Web han sido considerados en (Mithun, 2004) y (Sirin et al, 2004), como sinónimos de las acciones en planificación.

Teniendo en cuenta lo anterior, se ha identificado el uso de técnicas de aprendizaje para aprender una o varias características de calidad de acciones de planificación (QoS de servicios Web). A continuación se presentan algunos trabajos relacionados a esta temática.

En primer lugar, el trabajo de (Jiménez, 2007), se enmarca en función de proveer a planificadores la información de la ejecución de las acciones para alcanzar planes robustos. Es decir, la utilización de una métrica que determine la funcionalidad de un plan, haciendo referencia a la identificación de las acciones de éxito.

Para ello, define tres fases, la primera consiste en captura de la información; esto es obtener a través de una serie de ejecuciones, el conjunto de información que representa el comportamiento de una acción en el mundo real. La segunda fase, consiste en la inducción de la información encontrada, más específicamente los patrones de inducción (aquellas características que determina el éxito de la ejecución) e implementa la programación lógica inductiva, mecanismo de aprendizaje de máquinas que le permite generar patrones de aprendizaje. Finalmente, utiliza dichos patrones en la generación del plan con robustez. En este caso se definen dos posibilidades para calcular la robustez de un plan. La primera es la utilización de efectos condicionales probabilísticos para determinar la robustez de un plan, y la segunda es a través del uso de los costos de planificación a través de condiciones de costos. Esta última se determina a través de la fragilidad (verificar la robustez de un plan), la cual, permite transformar la maximización del producto de las probabilidades de éxito de acción a lo largo del plan, en una minimización de la suma de los costos de fragilidad. Se calcula como  $-\log(p)$  donde  $p$  es la probabilidad.

Otro trabajo es el de (Rosa y Jiménez, 2008) donde define políticas (*policy*) esto, en función de la calidad. Según (Fernández et al, 2009), una política es un mapeo entre los estados del mundo y la mejor acción a ser ejecutada, con el fin de lograr una serie de objetivos. Una política general se trata de un mapeo de las instancias del problema de un dominio dado; es decir, las diversas combinaciones de estado inicial y objetivos, en la mejor acción

para ser ejecutada, con el fin de alcanzar los objetivos. Esta es capaz de resolver cualquier instancia de un posible problema. Para este caso, el aprendizaje se realiza a través de árboles de decisión mediante tres pasos: 1, Generación de los ejemplos para el aprendizaje. Es decir, todo el background (una serie de problemas entrenados), 2. Clasificación de las acciones. Esto es obtener una clasificación de los mejores operadores que representen el contexto en función de los ejemplos entrenados. Finalmente los enlaces de clasificación obtienen los mejores, de acuerdo a los ejemplos entrenados.

El trabajo de (Ming et al, 2008) presenta un esquema de selección de servicios para ayudar al servicio solicitante a escoger los servicios web a través de sus características no funcionales, utilizando el aprendizaje ubicuo. Describe como características de QoS la ejecución del tiempo de envío y respuesta, la dependencia en función de la disponibilidad y el costo de ejecución. Hacen que la composición pase a ser una optimización combinatoria que puede ser definida en problemas de programación matemática.

En (Cardoso et al, 2004), se describe un modelo que permite predecir la calidad de los servicios para flujos de trabajo, basados en atributos de servicios atómicos en la calidad de servicios. Aclara que se puede compensar la deficiencia de la composición, si existen muchos servicios con funcionalidad compatible. La calidad de los servicios en los flujos de trabajo, representan las características cuantitativas y cualitativas de una aplicación de flujo de trabajo necesario para lograr un conjunto de necesidades iniciales. Los caracteres cuantitativos pueden ser evaluados en términos de medidas concretas, tales como: flujo de trabajo de tiempo de ejecución, coste, entre otros y, las características cualitativas especifican los servicios esperados ofrecidos por el sistema, como la seguridad y los mecanismos de tolerancia a fallos. QoS debe considerarse como un aspecto integral de los flujos de trabajo, por lo que deben integrarse con las especificaciones de flujo de trabajo. Finalmente, (Carman y Knoblock, 2007), expone un aprendizaje de definiciones semánticas a través de procesos de inducción. Las definiciones son dadas en el sistema o aprendidas previamente; esto, mediante el uso de ejemplos entrenados. La característica que evalúa es el uso único de las entradas y salidas donde se identifican los tipos de datos que se requieren en la ejecución, y cuál es el tipo de dato que se espera.

## 6. CONCLUSIONES Y TRABAJOS FUTUROS

En este trabajo se presentó una arquitectura de aprendizaje basada en árboles de decisión la cual, permite obtener la información crítica durante la ejecución de los servicios Web, con el fin de aprender su desempeño y optimizar el proceso de composición de servicios a través de la generación de criterios que evalúan el comportamiento de cada servicio determinando su funcionalidad al momento de su ejecución.

Asimismo, se hizo una descripción de los criterios utilizados por la arquitectura y los componentes que la constituyen, haciendo énfasis en la obtención de la base de conocimientos y la descripción del aprendizaje orientada a la utilización de penalidades sobre la ejecución de los servicios.

Actualmente se realiza la versión inicial de la arquitectura bajo la plataforma Linux la cual será manipulada con la ayuda de la herramienta de planificación INDYGO (Guzmán y Ovalle, 2007) utilizando inicialmente el dominio de compras virtuales (SHOPPING) creado dentro del mismo proyecto.

## AGRADECIMIENTOS

El presente trabajo está apoyado parcialmente por el proyecto de investigación de tesis de maestría de la Escuela de Sistemas de la Universidad Nacional sede Medellín "Modelo Basado en Aprendizaje de Máquinas para el Manejo de Riesgo de Falla Durante la Composición de Servicios Web" noviembre 2008, tesis asociada al grupo de investigación SINTELWEB.

## REFERENCIAS

- Bennett, L., Wilkinson, G. Y Oliver, K. (1996). The development and validation of instruments to assess the quality of Internet information: A progress report. Paper Presented at The Annual Convention of the Association for the Advancement of Computing in Education (AACE), Ed-Media 96, Boston Massachusetts.
- Blockeel, H. and De Raedt, L.(1998),. Topdown Induction of firstorder Logical Decision Trees. In, Artificial Intelligence.
- Cardoso Jorge , Miller John , Sheth Amit And Arnold Jonathan. (2002). Modeling Quality of

- Service for Workflows and Web Service Processes Technical Report# 02-002 v2, LSDIS Lab, Computer Science, University of Georgia.
- Cardoso, J., Sheth, A., Miller, J., Arnold, J., Kochut, K. (2004). Quality of service for workflows and web service processes, *Journal of Web Semantics*, Vol. 1, No. 3, pp. 281–308.
- Carman Mark James And Knoblock Craig. (2007). A. Learning Semantic Descriptions of Web Information, In: *International Joint Conference on Artificial Intelligence*.
- Fernández Susana, Jiménez Sergio And Rosa Tomás De La. (2009.). Improving Automated Planning with Machine Learning. *Handbook of Research on Machine Learning Applications and Trends*.
- Freimut, B., Hartkopf, S., Kaiser, P., Kontio, J., Kobitzsch, W. (2001). An industrial case study of implementing software risk management. In: *ESEC/FSE*, ACM Press.
- Hwang San-Yih, Wanhaojun G, Tang Jian, Srivastava Jaideep. (2007). A probabilistic approach to modeling and estimating the qos of web-services-based workflows *Information Sciences: an International Journal*.
- Jimenez Sergio. (2007). Learning actions success patterns from execution. In *Doctoral Conotium of the International Conference on Automated Planning & Scheduling, ICAPS'07*. Providence, Rhode Island, USA.
- Jimenez Sergio. (2009). *Planning & Learning Under Uncertainty*. Phd Thesis (Pending for submission). Departamento de Informática. Universidad Carlos III de Madrid.
- Jurisica I., A. Brian. Nixon. (1998). Building Quality into Cases Based Reasoning Systems, *Lecture Notes in Computer Science*, 1413, 363-380.
- Guzmán J., Ovalle D. (2007). Un Modelo de Planificación Incremental para Servicios Web Semánticos, *Revista Avances en Sistemas Informáticos*, 4, 3, 131-140.
- Kahn Beverly K., Strong Diane M. Y Wang Richard. (2002). Information Quality Benchmarks: Product and Service Performance. En: *Communications of the ACM*. Vol. 45 No. 4ve.
- Kokash Natalia. (2007). Risk Management for Service-Oriented Systems *Lecture Notes in Computer Science*.
- Ming Huang Fu, Lan Ci-Wei Stephen J.H. Yang (2008). QoS-based Learning Services Composition for Ubiquitous Learning. *IEEE International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing*.
- Mithun S. (2004). Automatic Composition and Invocation of Semantic Web Services, Tesis para Mag, University of Maryland.
- Quinlan, J. R. (1986). Induction of decision trees. In, *Machine Learning* (v.1 pp. 81106).
- Quinlan J. Ross and Morgan Kaufmann. (1993). C4.5: Programs for Machine Learning , 235-240.
- Ran Sh. (2003). A Model for Web Services Discovery With qos, *ACM sigecom Exchanges*, 4(1), 1–10.
- Rosa Tomás De La And Jiménez Sergio. (2008). Roller: A Lookahead Planner Guided by Relational Decision Trees. *International Planning Competition*.
- Roy, G.G. (2004). A risk management framework for software engineering practice. In: *Australian Software Engineering Conference (ASWEC)*, IEEE Computer Society.
- Sirin Evren , Bijan Parsia, Dan Wu, James Hendler And Dana Nau. (2004). HTN Planning for Web Service Composition Using SHOP2, *Journal of Web Semantics*, 1(4). 377-396.
- Verdon, D., Mcgraw, G. (2004). Risk analysis in software design. *IEEE Security and Privacy*.
- Wilkinson Gene L., Lisa T. Bennett, And Kevin M. Oliver. (2007). Evaluating the Quality of Internet Information Sources: Consolidated Listing of Evaluation Criteria and Quality Indicators. Department of Instructional Technology, University of Georgia, Athens, GA.
- Zeng L., Benatallah B., Dumas M., Kalagnanam J., Sheng Q.Z. (2003). Quality Driven Web Services Composition. In *Twelfth International Conference of WWW*, May 20-24, Budapest.