

II INTERNATIONAL SCIENTIFIC CONVENTION  
“II ICCUCLV 2019”

JUNE 23<sup>th</sup> – 30<sup>th</sup>, 2019  
CAYOS DE VILLA CLARA. CUBA.



II CONFERENCIA INTERNACIONAL DE PROCESAMIENTO DE  
LA INFORMACIÓN (CIPI 2019)

**DecisionRuleMiner: Descubrimiento de reglas de decisión a partir de  
descripciones textuales**

*DecisionRuleMiner: Discovering decision rules from textual descriptions*

Ana Laura Lara-Pérez<sup>1</sup>, Leticia Arco<sup>1,2</sup>, Gonzalo Nápoles<sup>2</sup>,  
Frank Vanhoenshoven<sup>2</sup>, Koen Vanhoof<sup>2</sup>

1- Universidad Central “Marta Abreu” de Las Villas, Cuba. Correos: lara@uclv.cu,  
leticiaarco@gmail.com

2-Universidad de Hasselt, Bélgica. Correos: {frank.vanhoenshoven, gonzalo.napoles,  
koen.vanhoof}@uhasselt.be

**Resumen:** El Modelado y Notación de Decisiones (DMN) es un estándar que ofrece una notación para modelar decisiones lógicas y dependencias entre éstas y elementos de datos; por ello, se ha convertido en uno de los más utilizados en el mundo empresarial actualmente. La modelación de decisiones es una tarea compleja y costosa; sin embargo, no existen aún propuestas que permitan automatizar este proceso a partir de descripciones textuales siguiendo el estándar DMN. Por eso, este trabajo se enfoca en la obtención de reglas de decisión a partir de descripciones textuales basándose en la aplicación de técnicas de procesamiento del lenguaje natural (NLP). Como resultado de esta investigación se presenta el software DecisionRuleMiner que permite la extracción de reglas de decisión mediante el empleo de técnicas de NLP, siguiendo un modelo y procedimiento general que describen cómo extraer los elementos de las reglas. La validación realizada muestra que la mayoría de las reglas extraídas a partir de las oraciones seleccionadas de 15 fuentes diferentes, se correspondieron con las reglas obtenidas por un modelador experto en DMN.

Contact Information  
[convencionuclv@uclv.cu](mailto:convencionuclv@uclv.cu)  
[www.uclv.edu.cu](http://www.uclv.edu.cu)

II INTERNATIONAL SCIENTIFIC CONVENTION  
“II ICCUCLV 2019”

JUNE 23<sup>th</sup> – 30<sup>th</sup>, 2019  
CAYOS DE VILLA CLARA. CUBA.



**Abstract:** *Decision Modeling and Notation (DMN) is a standard that provides notation for modeling logical decisions and dependencies between these decisions and data elements; therefore, it has become one of the most used in the business world today. Decision modeling is a complex and expensive task; however, there are still no proposals to automate this process based on textual descriptions following the DMN standard. For that reason, this work focused on obtaining decision rules from textual descriptions based on the application of natural language processing techniques (NLP). As a result of this research we present the DecisionRuleMiner software that allows the extraction of decision rules through the use of NLP techniques, following a conceptual model and a general procedure that describes how to extract elements from the rules. The validation shows that most of the extracted rules from the selected sentences from 15 different sources corresponds to the obtained rules by an expert modeler in DMN.*

**Palabras claves:** Modelado y Notación de Decisiones, procesamiento del lenguaje natural, reglas de decisión.

**Keywords:** *Decision Model and Notation, natural language processing, decision rules.*

## 1. Introducción

Hace algunos años las estrategias de negocio de las empresas se orientaban al cumplimiento de sus objetivos. En la actualidad, están enfocadas en la prestación de servicios, surgiendo los procesos de negocio con el objetivo de buscar la excelencia de la calidad a través de la satisfacción del cliente, mediante la reducción de costos y la optimización de los recursos (Corradini et al., 2018; De Smedt, De Weerd, Serral, & Vanthienen, 2018; Kuss, Leopold, van der Aa, Stuckenschmidt, & Reijers, 2018; Satyal, Weber, Paik, Di Ciccio, & Mendling, 2018). Con el objetivo de estandarizar la modelación de procesos de negocio surge el Modelado y Notación de Procesos de Negocio (*Business Process Model and Notation; BPMN*). BPMN es un estándar legible y entendible soportado por el Grupo de Gestión de Objetos (*Object Management Group;*

Contact Information  
[convencionuclv@uclv.cu](mailto:convencionuclv@uclv.cu)  
[www.uclv.edu.cu](http://www.uclv.edu.cu)

II INTERNATIONAL SCIENTIFIC CONVENTION  
“II ICCUCLV 2019”

JUNE 23<sup>th</sup> – 30<sup>th</sup>, 2019  
CAYOS DE VILLA CLARA. CUBA.



OMG). BPMN no es muy adecuado para modelar decisiones (Silver, 2016), de ahí que, para suplir esta necesidad, haya surgido el Modelado y Notación de Decisiones (*Decision Model and Notation*; DMN) como un estándar reciente de OMG.

BPMN no modela las decisiones lógicas en detalles. De ahí que DMN lo complemente con una notación para modelar decisiones lógicas y dependencias entre decisiones y elementos de datos. De esta forma, DMN permite a los usuarios controlar sus procesos y decisiones organizacionales de una manera más eficiente y efectiva. Los especialistas del campo de la modelación de decisiones han tenido en cuenta que es una tarea compleja y costosa, a partir de los factores que hacen este proceso complicado, entre ellos: la necesidad de capacitación de personal experto, el consumo de tiempo para modelar y el necesario pre-procesamiento de los datos para la obtención de los requisitos de negocio de cada empresa.

Se han realizado algunos trabajos que intentan automatizar la modelación de procesos de negocio a partir de las descripciones textuales debido a su complejidad, entre ellos se destacan las propuestas de (Caporale, 2016; Fabian Friedrich, 2010). Sin embargo, al ser DMN un estándar relativamente nuevo, no aparece reportado aún en la literatura algún trabajo que permita la modelación automática siguiendo este estándar. Lo cual motiva la realización de este trabajo, donde se presenta el sistema DecisionRuleMiner que permite obtener reglas de decisión a partir de descripciones textuales de tipo IF-THEN basándose en la aplicación de técnicas de procesamiento del lenguaje natural y siguiendo un modelo general diseñado a tales efectos.

Este artículo está organizado de la siguiente manera. Los principales enfoques para la extracción de reglas desde textos y las tareas de procesamiento del lenguaje natural se presentan en la Sección 2. La Sección 3 describe el principal resultado de esta investigación, el software DecisionRuleMiner que permite la generación automática de reglas de decisión y la validación realizada a partir de las reglas descritas por el modelador experto. Para concluir, se proporcionan las observaciones finales y las futuras direcciones de investigación.

Contact Information  
[convencionuclv@uclv.cu](mailto:convencionuclv@uclv.cu)  
[www.uclv.edu.cu](http://www.uclv.edu.cu)

II INTERNATIONAL SCIENTIFIC CONVENTION  
“II ICCUCLV 2019”

JUNE 23<sup>th</sup> – 30<sup>th</sup>, 2019  
CAYOS DE VILLA CLARA. CUBA.



## 2. Metodología

Existen varios trabajos que permiten extraer patrones desde textos. Aunque ellos no se focalizan en DMN, si ofrecen enfoques generales aplicables al contexto de la modelación de decisiones. De ahí que en esta sección se analizan las ventajas y desventajas de los trabajos precedentes, así como las principales tareas del Procesamiento del Lenguaje Natural utilizadas a tales efectos.

### 2.1 Enfoques para la extracción de reglas desde textos

Se han publicado varios trabajos dirigidos a la extracción de patrones, relaciones y reglas a partir de textos en diversos dominios (Califf & Mooney, 2010; Ciravegna, 2010; Huffman, 2010; Moldovan, 2010; Riloff, 2010; Soderland, 2010). Las desventajas más significativas de estas propuestas se relacionan a continuación. Algunas requieren reglas manualmente definidas y listas de palabras clave también manualmente seleccionadas, o una revisión humana de las reglas propuestas. Otras identifican el campo sintáctico, como el sujeto o el objeto directo, que contiene la frase objetivo, pero no identifican la frase objetivo en sí. La mayoría de los enfoques no explotan los principales beneficios del NLP, ya que utilizan un número limitado de tipos de relaciones entre los términos; la mayoría de ellos dependen del dominio y necesitan información adicional, como ontologías, diagramas o modelos; y algunos de ellos requieren interacción humana. No obstante, estas propuestas sugieren cómo enfrentar la extracción automática de reglas de decisión a partir de descripciones textuales, ya que es posible identificar qué puede ser reutilizado de las experiencias previas y aplicarlo a la extracción de reglas de decisión. La principal experiencia que podemos obtener de estos trabajos relacionados es que los mismos se sustentan en los árboles sintácticos y en las relaciones de dependencias.

Específicamente en el área de la modelación automática de procesos de negocios se han publicado varias propuestas (Borges Ferreira, Thom, & Fantinato, 2017; Caporale, 2016; Epure, Martin-Rodilla, Hug, Deneckere, & Salinesi, 2015; F. Friedrich, Mendling, & Puhmann, n.d.; Ghose, Koliadis, & Chueng, 2007; Gonçalves, Santoro, &

Contact Information  
[convencionuclv@uclv.cu](mailto:convencionuclv@uclv.cu)  
[www.uclv.edu.cu](http://www.uclv.edu.cu)

II INTERNATIONAL SCIENTIFIC CONVENTION  
“II ICCUCLV 2019”

JUNE 23<sup>th</sup> – 30<sup>th</sup>, 2019  
CAYOS DE VILLA CLARA. CUBA.



Baião, n.d.; Riefer, Ternis, & Thaler, 2016; Silver, 2016; Sinha & Paradkar, 2010). Estos artículos siguen un análisis sintáctico y semántico del texto de entrada, en algunos casos restringen el problema al uso de un lenguaje controlado o requieren la acción de los expertos humanos.

## 2.2 Tareas del procesamiento del lenguaje natural para la extracción de reglas

Cuando se aplica NLP, es muy importante determinar cuáles tareas son necesarias para extraer los patrones deseados desde las descripciones de decisiones textuales. La delimitación en oraciones tiene especial relación con el inicio y fin de las reglas a obtener. Debido a que muchas palabras son especialmente comunes y pueden servir de múltiples partes de la oración, en esta investigación resultan de interés la etiquetación de las partes de cada oración previamente delimitada. El análisis sintáctico (tanto el análisis de dependencias como el constituyente) juega un rol fundamental en la extracción de patrones desde textos, como se concluyó en la sección 2.1. El análisis de dependencias se enfoca en las relaciones entre palabras de una oración, mientras que el análisis constituyente se enfoca en construir el árbol sintáctico usando una gramática libre de contexto probabilística. Además, el reconocimiento de entidades nombradas (*Named Entity Recognition*; NER) juega un rol fundamental, ya que permite determinar qué elementos se asignan a nombres propios, tales como personas, ubicaciones u organizaciones.

## 3. Resultados y discusión

Esta sección está dirigida a la presentación y descripción de DecisionRuleMiner, software que permite el descubrimiento de reglas de decisión a partir de descripciones de la forma IF-THEN. Se describen las funcionalidades de la aplicación, los requisitos y su manejo. Finalmente se ilustra la calidad de las reglas obtenidas por DecisionRuleMiner a partir de descripciones de la forma IF-THEN extraídas de diversas fuentes.

Contact Information  
[convencionuclv@uclv.cu](mailto:convencionuclv@uclv.cu)  
[www.uclv.edu.cu](http://www.uclv.edu.cu)

### 3.1 DecisionRuleMiner

El software DecisionRuleMiner permite la extracción de reglas de decisión mediante el empleo de técnicas de NLP, siguiendo un modelo conceptual y un procedimiento general que describen cómo extraer los elementos de las reglas. DecisionRuleMiner se concibió para dotar a los investigadores y desarrolladores en el campo de la modelación de procesos de negocio de una herramienta que posibilite la obtención automática de reglas de decisión a partir de descripciones textuales de la forma IF-THEN. Esta aplicación es capaz de recibir un conjunto de decisiones de la forma IF-THEN en formato texto y devolver los componentes del antecedente y consecuente de las reglas de decisión que se generan de cada oración analizada, como se muestra en la Figura 1.

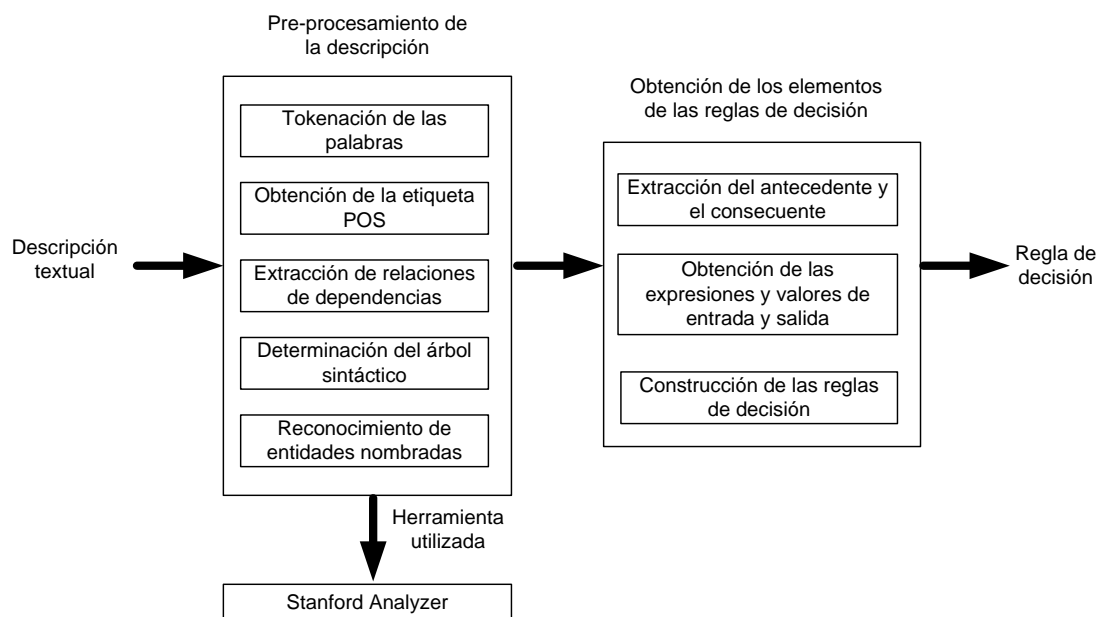


Figura 1. Esquema general de DecisionRuleMiner. Fuente: Elaboración propia.

DecisionRuleMiner fue desarrollado completamente en JAVA, característica que lo convierte en multiplataforma. Un elemento importante que tuvimos en cuenta al

II INTERNATIONAL SCIENTIFIC CONVENTION  
“II ICCUCLV 2019”

JUNE 23<sup>th</sup> – 30<sup>th</sup>, 2019  
CAYOS DE VILLA CLARA. CUBA.



seleccionar Java es que este lenguaje permite una fácil integración con el Stanford Analyzer<sup>1</sup>.

Este software consta de los módulos pre-procesamiento de la descripción y la obtención de los elementos de las reglas de decisión. El software transita por dos módulos definidos, el primer módulo realizar la tokenización de las palabras e identifica varios elementos que caracterizan cada oración. Finalmente, el segundo módulo permite la extracción de los antecedentes y consecuentes de las reglas de decisión a partir de cada oración analizada.

Stanford Analyzer fue seleccionado para garantizar las principales tareas del NLP que se requieren en el primer módulo de DecisionRuleMiner, ya que es una herramienta eficaz y flexible para el desarrollo de aplicaciones que requieran un pre-procesamiento previo siguiendo técnicas de NLP. El segundo módulo se desglosa en dos tareas fundamentales: la identificación de los pares antecedente-consecuente, y la extracción de las variables, los valores de entrada y de salida. Estas dos tareas se realizan de manera combinada, ya que la identificación de los pares antecedente-consecuente conlleva a la extracción de las variables, valores de entrada y salida. En el segundo módulo se utilizó una función general, como se muestra en la Figura 2, para la extracción de las reglas de decisión a partir de una descripción textual. *Extract\_rules* permite detectar todos los componentes del antecedente y consecuente, donde cada componente del antecedente consiste en una expresión de entrada y valor de entrada, mientras que cada componente del consecuente consiste en un nombre de salida y un valor de salida.

El primer paso a realizar de la función general definida en el módulo es la creación de la regla, en el segundo paso mediante la función *Extract\_antecedent\_rule*, como se muestra en la Figura 2, se obtiene el primer antecedente de la regla, que consiste en el valor de entrada y la expresión de entrada. La función *Extract\_first\_input\_entry* es la encargada de buscar la dependencia **mark** y a partir de ahí localizar las partes del primer elemento detectado del antecedente. Las funciones *Complete\_input\_entry* y

---

<sup>1</sup> [https://jar-download.com/?search\\_box=stanford-parser](https://jar-download.com/?search_box=stanford-parser)

II INTERNATIONAL SCIENTIFIC CONVENTION  
“II ICCUCLV 2019”

JUNE 23<sup>th</sup> – 30<sup>th</sup>, 2019  
CAYOS DE VILLA CLARA. CUBA.



*Complete\_input\_expression* solo son utilizadas para completar los valores y las expresiones de entrada: sin embargo, la función *Detect\_input\_expression* es esencial para la obtención de los elementos del antecedente, esta utiliza las dependencias **nsubjpass** y **nsubj** para detectar el enlace entre los valores y las expresiones de entrada. La función *Extract\_first\_consequent\_output\_entry* recibe como entrada el valor de entrada encontrado en el antecedente, busca la dependencia **advcl** que contiene a este valor y finalmente devuelve el elemento dependiente de valor de entrada según **advcl**. A partir de ese valor de salida y analizando las dependencias, es posible obtener todos los componentes del consecuente, de forma similar a como se realiza con el antecedente.

```
Extract_rules  
Input: Dependencies corresponding to an antecedent-consequent sentence.  
Output: The principal rule corresponding to the antecedent-consequent sentence and additional rules  
  
1. rule := Create rule  
2. extract_antecedent := Extract_antecedent_rule (dependencies)  
3. rule.antecedent_component_list.Add (extract_antecedent)  
4. more_antecedent := Extract_more_antecedent_rule(dependencies)  
5. rule.antecedent_component_list.Add (more_antecedent)  
6. extract_consequent := Extract_consequent_rule (dependencies)  
7. rule.consequent_component_list.Add (extract_consequent)  
8. more_consequent := Extract_more_consequent_rule(dependencies)  
9. rule.consequent_component_list.Add (more_consequent)  
10. additional_rules :=  
    Create_additional_rules(dependencies,extract_antecedent,extract_  
    consequent)
```

Figura 2. Función general para la extracción de antecedentes y consecuentes. Fuente: Elaboración propia.

Finalmente, las reglas adicionales pueden ser generadas por la función *Create\_additional\_rules*. Esta función busca cada dependencia **conj:or**. Aquí también es necesario tener en cuenta la etiqueta POS asignada a las palabras relacionadas por **conj:or**, porque **conj:or** podría estar conectando dos entradas de la misma expresión de entrada o el mismo nombre de salida, o podría estar conectando dos expresiones de entrada diferentes o dos nombres de salida diferentes. Por lo tanto, las etiquetas POS y

II INTERNATIONAL SCIENTIFIC CONVENTION  
“II ICCUCLV 2019”

JUNE 23<sup>th</sup> – 30<sup>th</sup>, 2019  
CAYOS DE VILLA CLARA. CUBA.



la dependencia **conj:or** son muy útiles para detectar otros componentes del antecedente y el consecuente, y así, generar nuevas reglas.

El sistema procesa solo documentos en idioma inglés; no obstante, el procesamiento de documentos en otros idiomas se puede realizar cambiando los modelos cargados por el Stanford Analyzer.

La salida de este software es la tabla de decisión correspondiente a una descripción, la cual está conformada por un conjunto de reglas de decisión siguiendo la notación DMN por cada oración de entrada. Una regla de decisión está formada por un antecedente y un consecuente, a su vez un antecedente consiste de una lista de pares entrada-valor y un consecuente consiste de una lista de pares salida-valor. En este trabajo se asume que cada oración describe una decisión, no obstante, en algunos casos se necesita más de una oración para describir una decisión.

DecisionRuleMiner ayuda al usuario en el procesamiento de un conjunto de procesos de negocio obteniendo como resultado las reglas de decisión de manera automática, totalmente no supervisada y sin requerir conocimiento a priori del dominio de aplicación. Además, ofrece otras funcionalidades, como la de mostrar el listado de tokens, sus etiquetas POS, el árbol sintáctico y las relaciones de dependencia.

### 3.2 Resultados experimentales

Para validar el sistema DecisionRuleMiner se seleccionaron oraciones en inglés de la forma IF-THEN que describen decisiones procedentes de diversas fuentes, principalmente de códigos de conducta de empresas<sup>2</sup>. La Tabla 1 muestra la cantidad de oraciones analizadas por cada tipo IF-THEN identificado, donde antecedente o consecuente simple indica que solo está compuesto por un elemento, y antecedente o

---

<sup>2</sup> <https://abc.xyz/investor/other/google-code-of-conduct.html>  
[https://www.apple.com/anzsea/supplier-responsibility/pdfs/supplier\\_code\\_of\\_conduct.pdf](https://www.apple.com/anzsea/supplier-responsibility/pdfs/supplier_code_of_conduct.pdf)  
<http://www.beiersdorf.com/investors/corporate-governance/code-of-conduct>  
<http://www.atlantapd.org/Home/>  
[http://www.citigroup.com/citi/investor/data/codeconduct\\_en.pdf](http://www.citigroup.com/citi/investor/data/codeconduct_en.pdf)

II INTERNATIONAL SCIENTIFIC CONVENTION  
“II ICCUCLV 2019”

JUNE 23<sup>th</sup> – 30<sup>th</sup>, 2019  
CAYOS DE VILLA CLARA. CUBA.



consecuente complejo indica que está compuesto por dos o más elementos relacionados con conjunciones o disyunciones.

Tabla 1: Cantidad de oraciones a analizar por cada tipo IF-THEN identificado. Fuente: Elaboración propia.

<b>Tipos de oraciones</b>	<b>Cantidad</b>
A) Oraciones IF-THEN con antecedente y consecuente simples.	10
B) Oraciones IF-THEN con antecedente complejo y consecuente simple.	3
C) Oraciones IF-THEN con antecedente simple y consecuente complejo.	5
D) Oraciones IF-THEN con antecedente y consecuente complejo.	2
E) Oraciones THEN-IF.	5

No se ha encontrado en la literatura alguna medida que permita cuantificar cuál es el nivel de coincidencia entre las reglas obtenidas automáticamente y aquellas formuladas por un modelador experto en DMN. Por ello, y con el objetivo de validar la propuesta presentada en esta investigación, hemos definido nueve criterios que permiten cuantificar cuánto se corresponden ambas soluciones:

- (1) La coincidencia en la cantidad de reglas generadas dada una descripción textual.
- (2) La cantidad de elementos del antecedente que coinciden, entre el total de elementos del antecedente.
- (3) La cantidad de elementos del consecuente que coinciden, entre el total de elementos del consecuente.
- (4) La cantidad de términos involucrados en los nombres de entrada que coinciden, entre la cantidad total de términos utilizados en los nombres de entrada.
- (5) La cantidad de términos involucrados en los valores de entrada que coinciden, entre la cantidad total de términos utilizados en los valores de entrada.
- (6) La cantidad de términos involucrados en nombres de salida que coinciden, entre la cantidad total de términos utilizados en los nombres de salida.
- (7) La cantidad de términos involucrados en los valores de salida que coinciden, entre la cantidad total de términos utilizados en los valores de salida.

Contact Information  
[convencionuclv@uclv.cu](mailto:convencionuclv@uclv.cu)  
[www.uclv.edu.cu](http://www.uclv.edu.cu)

II INTERNATIONAL SCIENTIFIC CONVENTION  
“II ICCUCLV 2019”

JUNE 23<sup>th</sup> – 30<sup>th</sup>, 2019  
CAYOS DE VILLA CLARA. CUBA.



- (8) La cantidad de términos involucrados en los nombres y valores de entrada que coinciden, entre la cantidad total de términos utilizados en los nombres y valores de entrada.
- (9) La cantidad de términos involucrados en los nombres y valores de salida que coinciden, entre la cantidad total de términos utilizados en los nombres y valores de salida.

En la Tabla 2 se muestra cómo se comportaron estos indicadores para las 25 decisiones analizadas por DecisionRuleMiner y por un modelador experto<sup>3</sup>. Es posible apreciar que según los tres primeros criterios existe casi coincidencia total, ya que estos criterios son los más generales. Los criterios del cuatro al siete son muy específicos, de ahí que generen indicadores más bajos de calidad, especialmente en la identificación de los valores del antecedente de las oraciones tipo E. La coincidencia alcanzada según los criterios ocho y nueve es aceptable, ilustrando que de manera general los antecedentes y consecuentes son bastante parecidos al comparar lo real con lo inferido.

Tabla 2: Comportamiento de los indicadores de correspondencia entre las reglas obtenidas por DecisionRuleMiner y aquellas obtenidas por el modelador experto. Fuente: Elaboración propia.

<b>Criterios</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>	<b>8</b>	<b>9</b>
Tipo A	1,000	1,000	1,000	0,700	0,900	0,521	0,650	0,822	0,590
Tipo B	1,000	1,000	1,000	0,440	0,500	0,667	0,667	0,543	0,667
Tipo C	1,000	1,000	0,700	0,566	0,720	0,413	0,400	0,768	0,448
Tipo D	1,000	1,000	1,000	0,335	0,750	0,475	0,585	0,635	0,625
Tipo F	1,000	0,900	1,000	0,330	0,266	0,616	0,550	0,388	0,633

### 3.3 Ejemplos ilustrativos

A continuación, se cuantifica el comportamiento de los nueve criterios definidos para la validación respecto a tres oraciones de tipo B, como se muestra en la Figura 3.

El mejor desempeño del sistema para este tipo de oraciones se alcanza con la tercera oración: “If both Personal and Work Experience Information are included in the CV, then the CV is considered valid.”. Es posible apreciar a continuación que la solución del modelador y la obtenida por DecisionRuleMiner son muy parecidas.

<sup>3</sup> Frank Vanhoenshoven, Laboratorio de Informática Empresarial, Universidad de Hasselt, Bélgica

II INTERNATIONAL SCIENTIFIC CONVENTION  
“II ICCUCLV 2019”

JUNE 23<sup>th</sup> – 30<sup>th</sup>, 2019  
CAYOS DE VILLA CLARA. CUBA.



Regla obtenida por el modelador:

IF “Personal Information” = “is defined” AND “Work Information” = “is defined”  
THEN “CV Validity” = “valid”.

Regla generada por DecisionRuleMiner:

IF “Personal Information\_CV” = “are included” AND “Work Experience Information\_CV” = “are included” THEN “CV\_valid” = “is considered”.

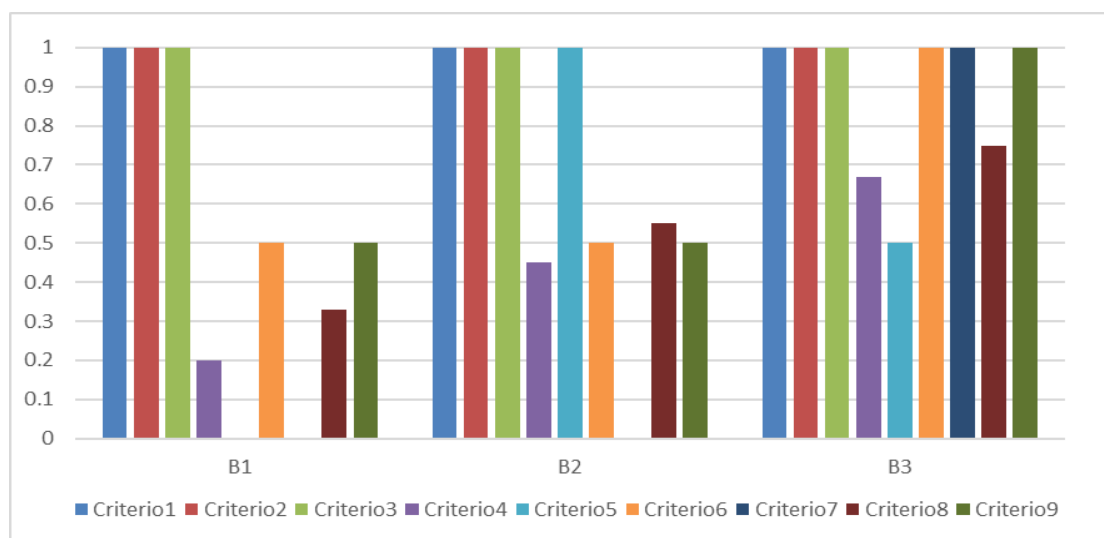


Figura 3: Representación de los criterios en un gráfico de barras para las oraciones de tipo B. Fuente: Elaboración propia.

Al generar la regla correspondiente a la oración “If the student provides insufficient documents, the registration cannot take place and the process ends without registration.” se aprecia aparentemente baja coincidencia entre el sistema y el experto. Esto se debe esencialmente a que existen algunas descripciones que son redundantes al formalizar los procesos y el experto es capaz de analizar semánticamente la oración y generalizar. Tales situaciones constituyen aún un reto en la modelación automática.

Regla obtenida por el modelador:

IF “Documents” = “INSUFFICIENT” THEN “Registration” = “REJECTED”.

Regla generada por DecisionRuleMiner:

II INTERNATIONAL SCIENTIFIC CONVENTION  
“II ICCUCLV 2019”

JUNE 23<sup>th</sup> – 30<sup>th</sup>, 2019  
CAYOS DE VILLA CLARA. CUBA.



IF “Student\_Insufficient documents” = “provides” THEN “registration\_place” = “not take” AND “process\_without registration” = “ends”.

Se aprecia en apariencia baja coincidencia entre el sistema y el experto. Esto es ocasionado esencialmente a que existen algunas descripciones que son redundantes al formalizar los procesos y el experto es capaz de analizar semánticamente la oración y generalizarla. Tal cuestión constituye un reto en la modelación automática.

#### 4. Conclusiones

Como resultado de esta investigación se desarrolló el software DecisionRuleMiner, que ofrece a los modeladores de decisiones la obtención automática de reglas de decisión a partir de descripciones textuales de la forma IF-THEN, transitando por dos módulos que permiten el pre-procesamiento de la descripción hasta la extracción de los elementos de las reglas de decisión. DecisionRuleMiner ayuda a la modelación de decisiones según el estándar DMN, mediante un diseño flexible y generalizable que permite interactuar con el Stanford Analyzer para la aplicación de las técnicas de NLP. Existe alta coincidencia semántica entre las reglas obtenidas por el experto y las generadas por el sistema, obteniéndose la mayor coincidencia para las oraciones tipo A; sin embargo, en algunos casos, siguiendo los criterios definidos, la coincidencia no es tan alta debido a las libertades que se toma el experto o el nivel de especificidad del sistema en la generación de las reglas.

Algunos retos para el trabajo futuro lo constituyen: valorar con los expertos en DMN la inclusión o no de determinados complementos en las oraciones que pudieran hacer más o menos específicas las reglas generadas y diseñar nuevos algoritmos que a partir de los resultados del Stanford Analyzer sean capaces de identificar reglas cuando las oraciones se tornan más complejas.

#### 5. Referencias bibliográficas

Borges Ferreira, R. C., Thom, L. H., & Fantinato, M. (2017). A semi-automatic approach to identify business process elements in natural language texts. In *Proceedings*

II INTERNATIONAL SCIENTIFIC CONVENTION  
“II ICCUCLV 2019”

JUNE 23<sup>th</sup> – 30<sup>th</sup>, 2019  
CAYOS DE VILLA CLARA. CUBA.



- of the 19th International Conference on Enterprise Information Systems* (pp. 250–261). Porto, Portugal.
- Califf, E., & Mooney, J. (2010). Relational learning of pattern - match rules for information extraction. *Computational Linguistics*, 9–15. <https://doi.org/10.1162/153244304322972685>
- Caporale, T. (2016). A tool for natural language oriented business process modeling. In C. Hochreiner & S. Schulte (Eds.), *Proceedings - 8th ZEUS Workshop* (Vol. 1562, pp. 49–52). Vienna, Austria.
- Ciravegna, F. (2010). Adaptive information extraction from text by rule induction and generalisation. *Natural Language Engineering*, 10, 145–165. <https://doi.org/10.1017/S135132490400333X>
- Corradini, F., Ferrari, A., Fornari, F., Gnesi, S., Polini, A., Re, B., & Spagnolo, G. O. (2018). A guidelines framework for understandable bpmn models. *Data & Knowledge Engineering*, 113, 129–154.
- De Smedt, J., De Weerd, J., Serral, E., & Vanthienen, J. (2018). Discovering hidden dependencies in constraint-based declarative process models for improving understandability. *Information Systems*, 74, 40–52.
- Epure, E. V., Martin-Rodilla, P., Hug, C., Deneckere, R., & Salinesi, C. (2015). Automatic process model discovery from textual methodologies: An archaeology case study. *Proceedings - International Conference on Research Challenges in Information Science*, (June), 19–30. <https://doi.org/10.1109/RCIS.2015.7128860>
- Friedrich, F. (2010). *Automated Generation of Business Process Models from Natural Language Input*. Humboldt-Universität zu Berlin. Retrieved from <http://www.mendling.com/2010-Friedrich-Master-Thesis-HU-Berlin.pdf>
- Friedrich, F., Mendling, J., & Puhmann, F. (n.d.). Advanced Information Systems Engineering. In *23rd International Conference on Advanced Information Systems Engineering (CAiSE 2011)* (pp. 482–496). London, UK: Proceedings, Vol. 6741 of Lecture Notes in Computer Science, Springer, 2011.
- Ghose, A., Koliadis, G., & Chueng, A. (2007). Conceptual Modeling - ER. In *26th*

Contact Information  
[convencionuclv@uclv.cu](mailto:convencionuclv@uclv.cu)  
[www.uclv.edu.cu](http://www.uclv.edu.cu)

II INTERNATIONAL SCIENTIFIC CONVENTION  
“II ICCUCLV 2019”

JUNE 23<sup>th</sup> – 30<sup>th</sup>, 2019  
CAYOS DE VILLA CLARA. CUBA.



*International Conference on Conceptual Modeling* (November 5, Vol. 4801, pp. 391–406). Auckland, New Zealand: Lecture Notes in Computer Science.

Gonçalves, J. C., Santoro, F. M., & Baião, F. A. (n.d.). Business Process Mining from Group Stories Business Process Mining from Group Stories. In *Proceedings of the 13th International Conference on Computer Supported Cooperative Work in Design (CSCWD 2009)* (pp. 161–169). <https://doi.org/10.1109/CSCWD.2009.4968052>

Huffman, S. (2010). Learning information extraction patterns from examples. *Connectionist Statistical and Symbolic Approaches to Learning for Natural Language Processing*, 1040, 246–260. Retrieved from <http://www.springerlink.com/index/752683888P304522.pdf>

Kuss, E., Leopold, H., van der Aa, H., Stuckenschmidt, H., & Reijers, H. A. (2018). A probabilistic evaluation procedure for process model matching techniques. *Data & Knowledge Engineering*. <https://doi.org/10.1016/j.datak.2018.04.008>.

Moldovan, D. I. (2010). Acquisition of Linguistic Patterns for Knowledge-Based Information Extraction. *IEEE Transactions on Knowledge and Data Engineering*, 7(5), 713–724. <https://doi.org/10.1109/69.469825>

Riefer, M., Ternis, S. F., & Thaler, T. (2016). Mining Process Models from Natural Language Text: A State-of-the-Art Analysis. V. Nissen, D. Stelzer, S. Straßburger, D. Fischer (Eds.), *Proceedings of the Multikonferenz Wirtschaftsinformatik (MKWI 2016)*, (Springer), 1–12.

Riloff, E. (2010). Automatically constructing a dictionary for information extraction tasks. *Proceedings of the National Conference on Artificial Intelligence*, 811. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.41.708&rep=rep1&type=pdf>

Satyral, S., Weber, I., Paik, H. -y., Di Ciccio, C., & Mendling, J. (2018). Business process improvement with the ab-bpm methodology. *Information Systems*, 1–15.

Silver, B. (2016). *DMN Method & Style The practitioner’s guide to decision modeling with business rules*. Altadena, CA: Cody-Cassidy Press.

Sinha, A., & Paradkar, A. (2010). Use Cases to Process Specifications in Business

Contact Information  
[convencionuclv@uclv.cu](mailto:convencionuclv@uclv.cu)  
[www.uclv.edu.cu](http://www.uclv.edu.cu)

**II INTERNATIONAL SCIENTIFIC CONVENTION  
“II ICCUCLV 2019”**

**JUNE 23<sup>th</sup> – 30<sup>th</sup>, 2019  
CAYOS DE VILLA CLARA. CUBA.**



Process Modeling Notation. <https://doi.org/10.1109/ICWS.2010.105>

Soderland, S. G. (2010). *Learning text analysis rules for domain-specific Natural Language Processing*. University of Massachusetts.

Contact Information  
[convencionuclv@uclv.cu](mailto:convencionuclv@uclv.cu)  
[www.uclv.edu.cu](http://www.uclv.edu.cu)