

**Generación automática de instancias del modelo  
cognitivo-afectivo (COGAF) partiendo de ejemplos  
textuales usando Procesamiento de Lenguaje  
Natural (PLN)**

**Estudiante:**

Alejandro Velásquez Arango

avelas56@eafit.edu.co

**Director:**

José Lisandro Aguilar Castro

Área de Computación y Analítica

jlaguilarc@eafit.edu.co

**UNIVERSIDAD EAFIT**

**MEDELLÍN**

**2023**

# Resumen

Los juegos serios son una herramienta muy útil para incentivar la educación y el entrenamiento en diversas áreas del conocimiento. Crear juegos serios que mejoren de manera eficaz las funciones cognitivas y las emociones que influencia una capacidad determinada puede ser un gran desafío. Para esto el modelo Cognitivo-Afectivo (CO-GAF) presenta una conceptualización de los componentes necesarios para el diseño de juegos serios. Sin embargo, la instanciación de estos componentes actualmente se hace de manera manual a partir de la descripción textual de un caso de estudio que se pretenda simular a través de un juego serio. Este trabajo tiene como objetivo generar automáticamente instancias del modelo COGAF siendo definido como una ontología, aplicando un sistema de población automática de ontología y aplicando técnicas de Procesamiento de Lenguaje Natural (PLN) a partir de textos de casos de estudios para facilitar y acelerar el proceso de diseño de juegos serios.

# 1. Planteamiento del problema

Los juegos serios son aquellos que han sido diseñados con un objetivo principal diferente al de la mera diversión. Normalmente estos juegos tienen el objetivo de entrenar o enseñar al usuario de algún tema en particular y resolver problemas del mundo real. Los juegos serios han ido ganando popularidad en distintos sectores debido a la efectividad que tienen para mejorar la experiencia del aprendizaje. Esto se logra con la integración de simulaciones, mecánicas de juego y estrategias pedagógicas que mejoran la experiencia del usuario y mitiga posibles problemas de memoria, atención y concentración.

Sin embargo, el diseño y desarrollo de estos juegos serios es una tarea complicada que requiere de un acercamiento metodológico. Para esto, se ha presentado un modelo cognitivo-afectivo (COGAF) que permita conceptualizar el diseño de los juegos serios, identificando diferentes componentes y las relaciones entre la cognición, la emoción y las mecánicas del juego. Este modelo permite a desarrolladores de videojuegos implementar actividades y mecánicas correctas que entrenen y estimulen las funciones cognitivas que se requieran mejorar. A partir de casos estudios se pueden generar instancias de los componentes del modelo para estudiar de manera sistemática las funciones cognitivas y emociones particulares involucradas en los eventos y mecánicas de un juego.

Actualmente, la instanciación de los componentes del modelo cognitivo-afecto se hace manualmente a partir de textos que detallen los casos de estudios que pueden ser modelados. Un ejemplo en particular es el caso de un accidente minero por explosión redactado por la Agencia Nacional de Minería (ANM) de Colombia, descrito en un documento textual en donde se detalla el evento ocurrido, las causas que generaron el accidente, las consecuencias y las lecciones aprendidas para evitar que ocurran de nuevo siniestros similares. Del texto se pueden determinar instancias de los componentes del modelo, como se ve en la figura 2

La determinación de las instancias de los componentes del modelo COGAF de un caso de estudio particular puede tomar bastante tiempo y de conocimiento experto. De esto se deriva la necesidad de poder realizar este proceso de instanciación de manera automática para acelerar este proceso. A partir de técnicas de Procesamiento de Lenguaje Natural (PLN) es posible desarrollar un modelo de Inteligencia Artificial que sea capaz de identificar las instancias a partir de un texto similar al caso de estudio mencionado anteriormente.

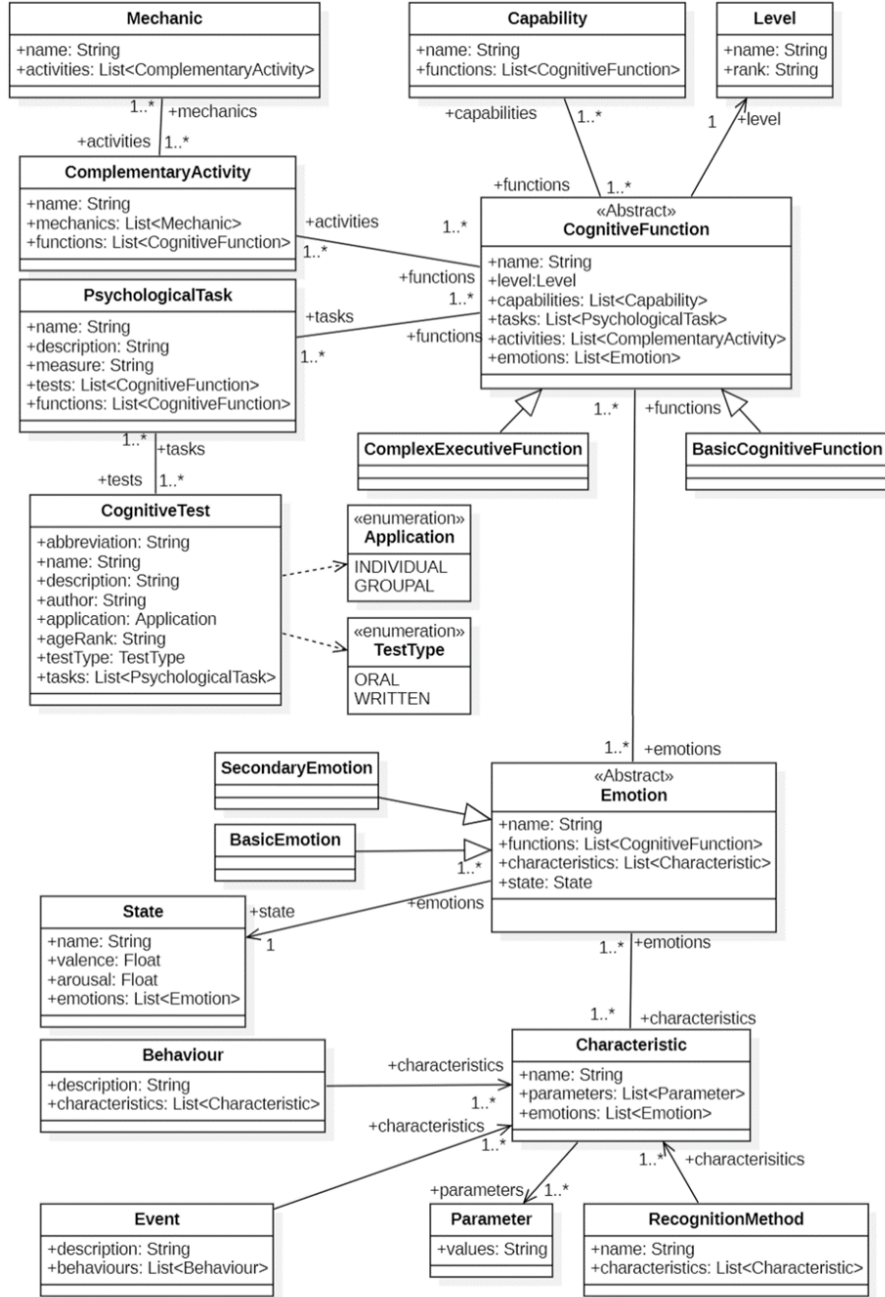


Figura 1: Modelo Cognitivo-Afectivo (COGAF)



## 2. Justificación

5

como parte del entrenamiento de los trabajadores. Los juegos serios son una poderosa herramienta para el mejoramiento de las funciones cognitivas y emociones implicadas en las labores y potenciales accidentes que puedan enfrentarse. La posibilidad instanciar automáticamente el modelo COGAF aceleraría considerablemente el desarrollo de estos juegos serios, abaratando costos y acortando tiempos, permitiendo que los juegos serios puedan ser producidos en mayor escala.

## **3. Objetivos**

### **3.1. Objetivo General**

Generar automáticamente instancias del modelo cognitivo-afectivo partiendo de textos y aplicando Procesamiento del Lenguaje Natural (PLN).

### **3.2. Objetivos Específicos**

- Recopilar textos de lecciones aprendidas de accidentes ocurridos en distintos sectores de la industria para formar un corpus
- Definir el modelo COGAF como una ontología legible por computadora
- Procesar, limpiar y preparar los textos para que puedan ser procesados por computadora
- Definir la métrica de evaluación para el entrenamiento y selección de modelos
- Entrenar modelos de PLN que instancien automáticamente los componentes del modelo COGAF con un texto de entrada
- Evaluar los modelos generados a partir de la métrica definida para seleccionar el modelo con mejor desempeño

## 4. Marco Teórico

### 4.1. Procesamiento de Lenguaje Natural

El Procesamiento del Lenguaje Natural es el campo de la inteligencia artificial que se interesa en poder otorgarle a una máquina la capacidad de leer, comprender y derivar significado de textos escritos, que son datos de tipo no estructurado

AMPLIAR SECCIÓN

### 4.2. Ontologías

En ciencias de la computación, una ontología es una representación formal del conocimiento que define las propiedades, relaciones y entidades de los conceptos de un dominio de discurso en particular [1]. Una ontología provee una forma estructurada y explícita de representar el conocimiento, lo cual permite que el conocimiento pueda ser leído y procesado por máquina.

En otras palabras, una ontología puede ser vista como un contenedor en donde el conocimiento de un dominio determinado está especificado formalmente de tal manera que sea legible por computadora y pueda ser almacenado, compartido y utilizado [2].

La capacidad de las ontologías de conceptualizar el conocimiento de manera formal y estructurada permite que encuentre aplicaciones en diversos campos. Entre los ejemplos más notables se encuentra el desarrollo de la Web Semántica, los sistemas de información en salud para definir relaciones entre enfermedades, pacientes y medicamentos para mejorar la interoperabilidad entre sistemas de información, y las ontologías biomolecular y biomédicas para representar información sobre moléculas, genes y proteínas, entre otros tipos de entidades biológicas, con el fin de entender el comportamiento e interacciones de estas entidades [3]. Las ontologías pueden ser aplicadas en sistemas de recuperación de información, chatbots y sistemas de soporte de decisiones. CITAR

SEGUIR DESARROLLANDO IDEAS SOBRE APLICACIONES REALES

### 4.3. Enriquecimiento y Población de Ontologías

Los componentes de una ontología pueden ser clasificados en dos tipos: TBox y ABox. Los componentes TBox hace referencia a los conceptos y relaciones de la ontología. Son las abstracciones que sirven para representar objetos del mundo real de acuerdo a ciertas propiedades bien definidas. Por ejemplo, el concepto estudiante, que representa a una persona que posee ciertas cualificaciones que lo hacen competente en un área del conocimiento dicta clases sobre ese tema particular, pertenece al conjunto TBox. Igualmente las relaciones abstractas entre conceptos también son componentes de tipo TBox. Por otro lado, los componentes ABox son instanciaciones de los conceptos y relaciones definidos. Es decir, son ejemplos reales de los componentes de tipo TBox. Por ejemplo, "Tim Berners Lee.<sup>es</sup> una instanciación del concepto profesor y por tanto pertenece a ABox. Estos componentes deben cumplir con las propiedades y relaciones definidos por TBox [2].

Extender los conjuntos TBox y ABox conlleva a dos tipos de extensión de las ontologías. El aumento del tamaño de TBox, es decir, el proceso de cambiar la estructura de la ontología añadiendo nuevos conceptos y relaciones, se conoce como enriquecimiento, mientras que la extensión del conjunto ABox, que es la adición de nuevas instancias de los componentes de TBox, se conoce como población de ontologías.

Las ontologías pueden ser construidas, enriquecidas y pobladas manualmente a partir de conocimiento experto y de fuentes de información, pero este proceso es altamente costoso y propenso a errores, por lo que existen sistemas automáticos y semi-automáticos para la elaboración de ontologías. Estos sistemas se pueden dividir en tres grupos según las tecnologías que implementan: Sistemas basados en reglas lexico-sintácticas, sistemas estocásticos y sistemas de Machine Learning

## 5. Metodología

Para poder cumplir con los objetivos planteados, la metodología propuesta para el desarrollo del proyecto es CRISP-DM, la cual es extensivamente utilizada en la industria para la realización de proyectos relacionados con ciencia de datos e inteligencia artificial. La metodología CRISP-DM se divide en 6 etapas: Entendimiento del negocio,



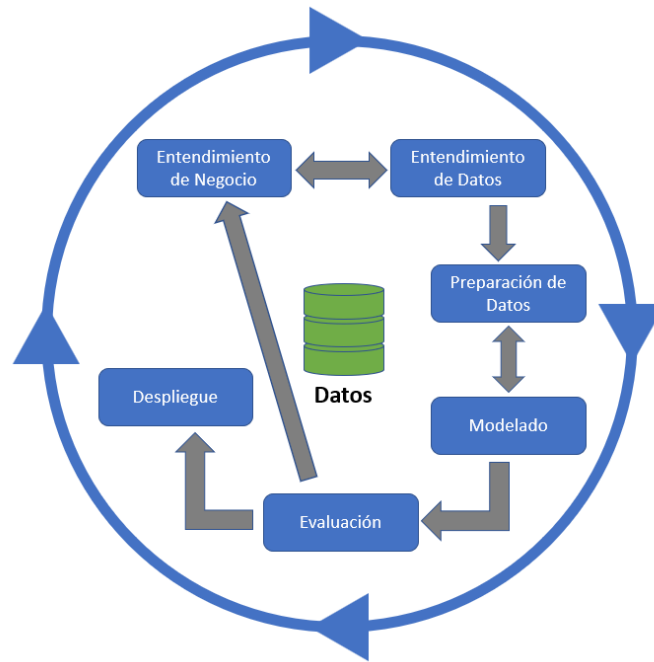


Figura 3: Esquema ilustrativo de las etapas de la metodología CRISP-DM

comprensión de los datos, preparación de datos, modelado, evaluación y despliegue. El flujo de estas etapas se ilustra en la figura 3.

- **Entendimiento del negocio:**
- **Comprensión de los datos:**
- **Preparación de los datos:**
- **Modelado:**
- **Evaluación**

## 6. Productos Esperados

Para la finalización de este proyecto se espera entregar un informe escrito exponiendo los detalles de la teoría, implementación y resultados del proyecto al igual que un repositorio que contenga el modelo generado y todo el código fuente que fue escrito para generarlo.

## 7. Plan de Gestión de Datos

GESTION DE DATOS

## 8. Aspectos Éticos

ASPECTOS ETICOS

## Referencias

- [1] M. Uschold, M. Gruninger, Ontologies: principles, methods and applications, The Knowledge Engineering Review 11 (1996) 93–136. doi:10.1017/S0269888900007797.  
URL [https://www.cambridge.org/core/product/identifier/S0269888900007797/type/journal\\_article](https://www.cambridge.org/core/product/identifier/S0269888900007797/type/journal_article)
- [2] M. Lubani, S. A. M. Noah, R. Mahmud, Ontology population: Approaches and design aspects, Journal of Information Science 45 (2019) 502–515. doi:10.1177/0165551518801819.  
URL <http://journals.sagepub.com/doi/10.1177/0165551518801819>
- [3] A. Ayadi, A. Samet, F. D. B. D. Beuvron, C. Zanni-Merk, Ontology population with deep learning-based nlp: a case study on the biomolecular network ontology ontology population with deep learning-based nlp: a case study on the biomolecular sciencedirect-nc-nd license (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) peer-review under responsibility of kes international. ontology population with deep learning-based nlp: a case study on the biomolecular network ontology 159 (2019) 572–581. doi:10.1016/j.procs.2019.09.212.  
URL [www.sciencedirect.com](http://www.sciencedirect.com)[www.sciencedirect.com](http://www.sciencedirect.com)