

UNIVERSIDAD DE LAS CIENCIAS INFORMÁTICA
Centro internacional de posgrado

**Herramienta computacional para la caracterización de los estudiantes en las
universidades públicas de Angola a partir del rendimiento académico**

Trabajo final presentado en opción al título de
Máster en Informática Aplicada

Autor: Lic. Nadilson Camilo Sangunga Eduardo
Tutores: Dra. Natalia Sánchez Martínez
Dr C. Adolfo Días Sardiñas

Ciudad de La Habana, Diciembre año 2015

La imaginación es más importante que el conocimiento

Albert Einstein

AGRADECIMIENTOS

Yo sé que un día tendré que agradecer, quiero empezar agradeciendo a Dios que me ha dado fuerzas para seguir adelante en todo que hago.

A mis tutores Dra. Natalia Sánchez Martínez y Dr C. Adolfo Díaz Sardiñas, quiero darles las gracias por hacerme enamorar del área de reconocimiento de patrones, por su paciencia y apoyo en el desarrollo de esta investigación.

Al Dr. Mateus Padoca Calado, a este hombre no lo veo solo como un coordinador de la carrera de computación sino como el hombre que ha hecho la diferencia en mi vida como académica, agradezco a él por proporcionarme este reto.

Agradezco a todos los profesores de la maestría en informática aplicada, al centro internacional de posgrado por su atención.

Al ministerio de enseñanza superior de Angola por esta oportunidad, a la universidad Agostinho Neto, a Facultad de ciencias y al departamento de matemática.

A mis compañero de lucha, de la carrera, de la maestría José, Joaquim, Victor, Ladislau, Edson, por su compañerismo y apoyo en todos los momentos que fue posible.

A mi amigo Alfredo Trujillo por su bondad y sus ideas de conocer más el mundo.

A mi familia que me ha apoyado siempre, a mis amigos.

A todos ustedes quiero dar mis más profundo agradecimiento.

DEDICATORIA

Por amarme, cuidarme, y enseñarme mucho de la vida yo quiero dedicar este trabajo de todo corazón y mucho más a mis padres, Artur Eduardo y Alice Clara Sangunga.

Quiero también dedicar este trabajo a mis hermanos Tagilde, Isaac y Martinho.

A mi compañero Randy, que me ha probado desde niño que aquí es donde pertenezco, hermano por conocerte en 2009 he llegado hasta aquí.

En memoria de mi compañero Jeronimo Hortencio Chilulo, que siempre ha depositado su creencia en mí y en Randy.

Quiero también dedicar este trabajo a mi querida abuela Amelia, quisiera yo haber podido decirte que te amo mucho mi corazón.

DECLARACIÓN JURADA DE AUTORÍA

Declaro por este medio que yo Nadilson Camilo Sangunga Eduardo del registro de inscripción de nacimiento>, con carné de identidad 91101700906, soy el autor principal del trabajo final de maestría < sistema inteligente basado en los algoritmos conceptuales para la toma de decisión en las universidades públicas de Angola, desarrollada como parte de la Maestría en Informática Aplicada y que autorizo a la Universidad de las Ciencias Informáticas a hacer uso de la misma en su beneficio, así como los derechos patrimoniales con carácter exclusivo.

Y para que así conste, firmo la presente declaración jurada de autoría en Ciudad de La Habana a los 11 días del mes de diciembre del año 2015.

Maestrante firma

RESUMEN

El estudio del análisis del rendimiento académico es un reto en el mundo universitario de hoy en día. Varios son los trabajos científicos encontrados en la literatura consultada sobre estrategias y herramienta computacionales con el objetivo de apoyar la toma de decisiones para elevar la eficiencia académica de los estudiantes.

En este trabajo se propone un nuevo modelo informático basado en la extracción de patrones relacionados con el rendimiento académico de los estudiantes de las universidades públicas de Angola. Utilizando técnicas de reconocimiento lógico combinatorio de patrones, específicamente los algoritmos de agrupamiento conceptual.

El modelo se implementa a través de la herramienta computacional SADUP para el análisis de los datos de los estudiantes de las universidades públicas de Angola. Finalmente se presenta la evaluación del modelo que muestra la conformidad con el mismo.

Palabras claves: Toma de decisión, reconocimiento lógico combinatorio de patrones, algoritmos conceptuales, rendimiento académico.

Abstract

The study's analysis of academic performance is a challenge in academia today. There are several scientific studies found in the literature on strategies and computational tools in order to support the decision to raise the academic performance of students.

This paper presents a new computer model based on extraction related to academic performance of students in public universities of Angola it intends patterns. Using techniques of combinatorial logic pattern recognition, specifically conceptual clustering algorithms.

The model is implemented through the computational tool SADUP for the analysis of data from students at public universities in Angola. Finally, assessment model showing compliance with it is presented.

Keywords: *Decision making, pattern recognition approach combinatorial logic, algorithms conceptual, academic performance.*

Índice de ilustraciones

Ilustración 1 Inteligencia artificial vs inteligencia computacional.....	19
Ilustración 2 Estructura de un sistema inteligente	21
Ilustración 3 - Árbol de objetos.....	29
Ilustración 4 - Taxonomía de los algoritmos conceptuales, adaptado de.....	30
Ilustración 5 - Algoritmo de agrupamiento conceptual LC	32
Ilustración 6 - Proceso del algoritmo RGC	33
Ilustración 7 - Algoritmo de agrupamiento RGC propuesto por Aurora Pons, 2002.....	34
Ilustración 8 - Modelo Computacional	40
Ilustración 9 - Tipo de variables	41
Ilustración 10 - Matriz inicial.....	42
Ilustración 11 - Matriz de Aprendizaje	43
Ilustración 12-Modelo conceptual de la base de datos.....	46
Ilustración 13 - Modelo de datos	49
Ilustración 14 - arquitectura MVC.....	50
Ilustración 15 - Ventana de autenticación del usuario.....	51
Ilustración 16 - Ventana principal del software	51
Ilustración 17 - Importación de los datos de estudiantes.....	52
Ilustración 18 - Datos de entrada de los estudiantes.....	54
Ilustración 19 - Reporte principal del caso A.....	55
Ilustración 20 - Reporte relacionado a un grupo del caso A.....	56
Ilustración 21 - Reporte de grupo relacionado al caso A.....	57
Ilustración 22 - Reporte principal del caso B.....	57
Ilustración 23 - Reporte relacionado a un grupo del caso B.....	58
Ilustración 24 - Reporte principal del caso C.....	58

Índice de tablas

<i>Tabla 1 - Reconocimiento de patrones en las ciencias biomédicas</i>	25
<i>Tabla 2 - Reconocimiento de patrones en las Geociencias</i>	26
<i>Tabla 3 - Reconocimiento de patrones en la Criminalística</i>	26
<i>Tabla 4 - Reconocimiento de patrones en la Sociología</i>	26
<i>Tabla 5 - Reconocimiento de patrones en la Pedagogía</i>	26
Tabla 6 - Comparación de los algoritmos de agrupamiento conceptuales	35
Tabla 7 - Rasgos propuestos y sus respectivos dominios.....	41
<i>Tabla 8 - Herramientas seleccionadas para el entorno de desarrollo</i>	46
Tabla 9 - Requisitos Funcionales	47
Tabla 10 - Requisitos no Funcionales.....	48
Tabla 11 - Descripciones del modelo de datos	49
Tabla 12 - Facultades y carreras de los estudiantes.....	54
Tabla 13 - Tabla de resultados del caso A.....	56
Tabla 14 - Leyenda del rendimiento.....	56
Tabla 15 - Resultados del caso C.....	59
Tabla 16 -Evaluación de los caso de estudio	59
Tabla 17 - Resultado experimentales I	62
Tabla 18 - Resultados experimentales II	62
Tabla 19 - Resultados experimentales III.....	63

Índice de contenido

Introducción.....	11
Objetivo de la investigación.....	15
Tareas de investigación.....	16
Métodos científicos empleados.....	16
Capítulo-I “Fundamentación teórica”.....	18
1.2 Sistemas inteligentes.....	18
1.2.1 Definición de los sistemas inteligentes.....	18
1.2.2 Historia de los sistemas inteligentes.....	19
1.2.3 Estructura de los sistemas inteligentes.....	20
1.2.3 Aplicación de los sistemas inteligentes.....	22
1.3 Técnicas de inteligencia artificial.....	22
1.3.1 Reconocimiento de patrones.....	23
1.3.1.1 Definición.....	24
1.3.1.2 Enfoques en el reconocimiento de patrones.....	24
1.3.1.3 Problemas en el reconocimiento de patrones.....	25
1.3.1.4 Aplicación del reconocimiento de patrones.....	25
1.3.1.5 Algoritmos de agrupamiento conceptual.....	26
1.3.1.5.1 Teoría de algoritmos conceptuales.....	28
1.3.1.5.2 Conceptos básicos.....	31
1.3.1.5.3 Algoritmo LC-conceptual.....	32
1.3.1.5.4 Algoritmo RGC.....	33
1.3.1.6 Teoría de los testores.....	35
1.4 Sistemas para el proceso de toma de decisión.....	36
1.4.1. Definición de los sistemas para la toma de decisión.....	36
1.4.2 Características de los sistemas para la toma de decisión.....	36
1.4.3 Sistema para la toma de decisión en la educación.....	38
Conclusiones parciales.....	38
Capítulo-II “Desarrollo de la propuesta”.....	40
2.1. Modelo computacional y/o algoritmo.....	40
2.1.1. Recolección de los datos.....	40
2.1.2. Estructuración de los datos.....	41
2.1.3. Determinación extensional.....	42

2.1.4. Determinación intencional	43
2.1.5. Presentación de los resultados.....	45
2.2. Implementación del sistema	45
2.2.1 Entorno de desarrollo	46
2.2.2. Requisitos del sistema	47
2.2.3. Diseño del sistema.....	48
2.2.3.1. Diseño de la base de datos.....	48
2.2.3.2. Diseño arquitectónico.....	49
2.2.4. Interfaz gráfica	50
Conclusiones parciales	52
3.1. Validación de los resultados	53
Caso A “Estudiantes del año lectivo 2007 “	55
Caso B Estudiantes del año 2008	57
Caso C estudiantes del año 2011.....	58
3.1.1. Evaluación de los resultados “ <i>Rand Statistic</i> “	59
3.2. Medidas de validación basada en la teoría de los conjuntos aproximados	59
3.2.1. Medidas de inferencia clásicas de la teoría de los conjuntos aproximados	60
3.2.2. Resultados experimentales	61
3.3. Conclusiones parciales.....	63
Conclusiones y recomendaciones.....	63
Referencias bibliográficas.....	65

Introducción

Un creciente número de herramientas han soportado la creación y diseminación de la información provocando su proliferación arrolladora. Mientras que la información está continuamente creciendo, la habilidad de los humanos para procesarla y asimilarla permanece constante. Por otro lado, la información en sí misma no supone ninguna ventaja, su sistematización, incorporación y utilización son los elementos que aportan su valor añadido: el conocimiento. Por tanto, es necesario crear sistemas que generen conocimiento. Así, es posible asegurar el uso productivo de la información y guiar una toma de decisiones óptima, dando lugar a la Gestión del Conocimiento (GC) (Arcos 2009).

El conocimiento se puede gestionar de diversas formas y hacerlo requiere de la integración de varias áreas del saber: el descubrimiento de conocimiento en bases de datos, la minería de datos y de textos. Dentro de éstos, el agrupamiento y los procesos post-agrupamiento permiten organizar la información, determinar información relevante y crear nuevo conocimiento a partir de la información disponible.

El agrupamiento es una tarea del aprendizaje no supervisado que tiene como objetivo descomponer el conjunto de datos, de forma tal que los objetos que pertenecen al mismo grupo sean tan similares como sea posible y los objetos que pertenecen a grupos diferentes sean tan disimilares como sea posible. El análisis de grupos es una herramienta para descubrir una estructura previamente oculta en los datos, asumiendo que existe un agrupamiento natural o cierto en ellos. Sin embargo, la asignación de los objetos a las clases y la descripción de esas clases son desconocidas (Arco 2009).

No sólo es necesario obtener un buen agrupamiento, el post-agrupamiento es indispensable para caracterizar los grupos obtenidos, sobre todo cuando el agrupamiento ayuda a la gestión de la información. La valoración de los grupos obtenidos se puede realizar mediante su validación, etiquetamiento o alguna forma de caracterización. Los investigadores le han dedicado menor interés al post-agrupamiento, una de las causas es que aún quedan problemas sin resolver en el proceso de agrupamiento.

Por otra parte, las universidades hoy día en el mundo están estudiando nuevas estrategias para mejorar la calidad del proceso de enseñanza universitaria, por ejemplo, el espacio de convergencia europea de la educación está sugiriendo el diseño de estrategias para la mejora de la calidad en el proceso de enseñanza. Como se ha comentado en, *“reflexionar sobre todos los elementos que la evaluación del rendimiento del alumnado proporciona se convierte en un mecanismo claro para la mejora de la calidad del proceso educativo”*.

En la universidad Politécnica de Madrid se ha realizado un estudio para analizar el rendimiento del primer ciclo en todas sus escuelas, ellos se basaron en el rendimiento de una asignatura que es un valor numérico constituido por coeficientes asociados al tiempo que lleva aprobar una materia, la nota de aprobado, la sumatoria del rendimiento de las asignaturas de dicho curso y la sumatoria del rendimiento de las asignaturas en un año (Tirados, 1993).

En la universidad Nacional del Nordeste en la Argentina en la facultad de Ciencias exactas y naturales y agrimensura (FACENA-UNNE) existe un grupo de docentes que investiga esta temática y tienen publicados sus resultados utilizando modelos predictivos y técnicas de minería de datos para la identificación de factores asociados al rendimiento académico de alumnos universitarios, en su artículo se basan en las siguientes variables:

- 1) Resultados del diagnóstico de conocimientos matemáticos previos;
- 2) Condiciones socioeconómicas de los alumnos de las distintas carreras;
- 3) datos obtenidos de encuesta directa a los alumnos de primer año.

Para la formulación y ajuste de dicho modelo de predicción se utilizaron técnicas de minería de datos clásicas y métodos simbólicos o inteligentes, evaluando el desempeño en la predicción del rendimiento académico de los alumnos. Donde dicha solución de forma general, permitió analizar el perfil socioeconómico y educativo de los estudiantes, se ha analizado el rendimiento académico de los alumnos de las carreras de grado de la FACENA, tales informaciones pueden servir para el proceso de ayuda a la toma de decisiones en el proceso de enseñanza universitaria. (Porcel, Dapozo, & López, 2010).

Análisis del rendimiento académico en las carreras de informática de la Universidad Politécnica de Valencia (UPV) aplicando técnicas de minería de datos, en este trabajo se relacionó el rendimiento académico con las características socioeconómicas y académicas de los estudiantes, que se obtienen en el momento de sus matrículas y que se recogen de la base de datos de la Universidad. De acuerdo a las calificaciones obtenidas y convocatorias

utilizadas ellos definieron un indicador del rendimiento para cada estudiante (R. Alcover, y otros).

En este estudio se utilizaron técnicas de minería de datos, para encontrar el nivel de condicionamiento entre dicho rendimiento y características como el nivel de conocimiento de entrada de los estudiantes. El desarrollo de la investigación, se enfocó en el objetivo de obtener una herramienta, para la acción tutorial que puede apoyarse en las predicciones de los modelos que se obtienen para encauzar sus recomendaciones y encuadrar las expectativas y el esfuerzo necesario para cada alumno, lógicamente dentro de la cautela habitual a la hora de tratar modelos inferidos a partir de datos.

En el caso de la Universidad Nacional de Tucumán (UNT) Argentina, ellos desarrollaron una Metodología para realizar el seguimiento académico de alumnos universitarios, de acuerdo a los datos existentes en la institución lo aplicaron en las primeras cohortes de la carrera de Ingeniería Biomédica de la Facultad de Ciencias Exactas y Tecnología de la UNT. Dicho estudio fue utilizado como forma de responder diferentes preguntas durante el proceso de acreditación de la carrera, para proponer formas de disminuir la deserción en los primeros años y para mejorar las acciones de orientación vocacional.

El sistema universitario público angolano (SUPA) al igual que todo sistema universitario tiene dentro de sus principales retos elevar el rendimiento académico de los estudiantes universitarios.

Hoy en día el número de estudiantes angoleños que ingresan al SUPA se ha incrementado considerablemente. Sin embargo el rendimiento académico de los estudiantes es bajo, que resulta en diversas situaciones como; permanencia prolongada de los estudiantes en las carreras, alumnos que desisten de las carreras, consecuentemente la eficiencia académica aún no alcanza los niveles deseados.

La mayoría de las instituciones públicas en Angola no cuentan con el proceso académico automatizado lo que limita poder realizar un análisis objetivo de los factores que están influyendo directamente en el bajo rendimiento académico de las mismas. Y por ende no se pueden tomar estrategias y decisiones a partir de una caracterización real de los estudiantes.

Basado en la necesidad de elaborar un sistema informático que genere conocimiento, a partir de la información con que se cuenta en estas instituciones, de apoyo a la toma de decisiones para elevar el rendimiento académico. Así como, las características distintivas entre los métodos de agrupamiento y la problemática planteada son factores a estudiar que conllevaron

a la utilización de los algoritmos de agrupamiento conceptual del reconocimiento lógico combinatorio de patrones.

En la literatura científica consultada se han propuesto una variedad de métodos de agrupamiento. Los métodos se clasifican siguiendo varios criterios: tipo de los datos de entrada del algoritmo, criterios para definir la similitud entre los objetos, conceptos en los cuales se basa el análisis y forma de representación de los datos, entre otros. Una clasificación general distingue dos tipos: aquellos que forman particiones y los jerárquicos.

Los métodos de aprendizaje y clasificación pueden ser organizados, atendiendo a su naturaleza, en métodos estadísticos, modelos o algoritmos matemáticos para el reconocimiento de patrones, estrategias basadas en árboles de decisión, sistemas basados en el conocimiento, entre otros.

Por su parte, la casi totalidad de los modelos de algoritmos para la solución de problemas de clasificación no supervisada ofrecen estructuraciones de los espacios sobre los que se aplican en forma extensional, es decir, determinan qué objetos están en un cierto agrupamiento, o con qué grado pertenecen a un agrupamiento dado (modelos difusos, es decir, basados en la Teoría de los Subconjuntos Difusos). En otras palabras, dan una estructuración extensional de los espacios.

El modelo conceptual de Michalski constituye una excepción. Este modelo de algoritmos pretende determinar, no sólo cuál es la integración de sus agrupamientos sino además, qué propiedades cumplen los objetos que pertenecen a un mismo agrupamiento. Se puede afirmar que los algoritmos de Michalski son los primeros en dar una estructuración conceptual del espacio (Ruiz-Shulcloper 2013).

Los algoritmos de agrupamiento conceptual se pueden dividir en dos grandes grupos: los algoritmos incrementales y los no incrementales. Los algoritmos incrementales basan su funcionamiento en la adaptación de los agrupamientos (o conceptos) con los nuevos objetos que se le van presentando, es decir, cada vez que llega un nuevo objeto mediante una cierta estrategia éste es clasificado en los agrupamientos ya existentes o se crean nuevos agrupamientos. Por otro lado los algoritmos no incrementales estructuran una muestra de objetos sin presuponer que éstos llegan de uno en uno.

El algoritmo LC-Conceptual (Martínez-Trinidad 2001) está basado en los conceptos de la clasificación no supervisada en el enfoque lógico-combinatorio (Ruiz-Shulcloper 2013) y

retoma algunas ideas propuestas por Michalski para generar conceptos, interpretables por los especialistas, en términos del conjunto de rasgos original.

Para la conformación de los conceptos el algoritmo LC-Conceptual utiliza los testores y/o testores típicos. Los rasgos que componen los testores son rasgos relevantes con un alto carácter diferenciante y fuerza predictora; los cuales además son una alternativa del enfoque lógico combinatorio para el cálculo de la importancia de los mismos. Este elemento influye de forma determinante el descubrimiento del conocimiento o caracterización del rendimiento académico del universo a estudiar.

La anterior situación justifica el planteamiento del siguiente **problema científico**: ¿Será posible desarrollar una herramienta computacional que facilite el análisis del comportamiento del rendimiento académico de los estudiantes utilizando métodos de agrupamiento conceptual del reconocimiento lógico combinatorio de patrones?

Por tanto, la presente investigación tiene como **objeto de estudio** el proceso de descubrimiento de conocimiento a partir de grandes volúmenes de información en instituciones académicas.

Objetivo de la investigación

Desarrollar una herramienta computacional que facilite el descubrimiento de conocimiento o caracterización de conjuntos de estudiantes según su rendimiento académico en las instituciones públicas de Angola utilizando el agrupamiento conceptual del reconocimiento lógico combinatorio de patrones.

Se enmarca como **campo de acción** la caracterización del rendimiento académico de poblaciones de estudiantes universitarios en un universo específico utilizando el agrupamiento conceptual.

Se tiene como **idea a defender** de la presente investigación:

Si se desarrolla una herramienta computacional para extraer patrones o conceptos a partir del rendimiento académico de los estudiantes universitarios, permitirá brindar a los especialistas información, a tener en cuenta, en la toma de decisión con el objetivo de elevar la eficiencia académica en las instituciones públicas de Angola.

Tareas de investigación

Para dar cumplimiento al objetivo, se plantearon las siguientes tareas de investigación:

- Elaborar el marco teórico a partir del estado del arte sobre el tema de investigación.
- Estudiar los algoritmos de reconocimiento lógico combinatorio de patrones para la clasificación no supervisada.
- Diseñar el modelo a implementar utilizando la idea básica del algoritmo de agrupamiento conceptual LC - conceptual.
- Implementar computacionalmente el modelo propuesto.
- Evaluar el modelo propuesto con datos similares a los almacenados en las bases de datos históricas de las facultades de las universidades públicas de Angola.

La **novedad científica** del presente trabajo consiste en que se desarrolla un modelo de agrupamiento conceptual del reconocimiento lógico combinatorio de patrones que permite encontrar las características distintivas de conjuntos de estudiantes según su rendimiento académico, que puede servir de apoyo a la toma de decisiones en el proceso de formación de los estudiantes.

El **valor práctico** del trabajo se fundamenta en la herramienta computacional SADUP (*Sistema de Apoyo a la toma de Decisión en las Universidades Públicas*) para el análisis de los datos, que facilita a los especialistas el análisis del comportamiento de grupos de estudiantes según su rendimiento académico, lo cual puede tenerse en cuenta para la toma de decisiones en aras de incrementar el rendimiento académico.

Métodos científicos empleados

La presente investigación consta el método científico fundamentalmente empleado es el analítico-sintético. El análisis permite determinar las características del proceso de extracción de conocimiento o patrones en sistemas informáticos con este fin y la búsqueda de algoritmos de agrupamiento no supervisado que en problemas con características similares son eficiente. La síntesis permite integrar los algoritmos de agrupamiento conceptuales al problema de extracción de conceptos o características distintivas en el problema en cuestión. La tesis está conformada por tres capítulos.

En el primero se abordan las principales características de las herramientas computacionales desarrolladas en el campo de la educación con el objetivo de realizar análisis sobre el rendimiento académico de los estudiantes universitarios. Así como, los algoritmos de agrupamiento del reconocimiento lógico combinatorio de patrones.

El segundo capítulo describe el modelo computacional fundamentado en la idea básica del algoritmo de agrupamiento conceptual LC-conceptual implementado en la herramienta SADUP.

El tercer y último capítulo contiene los resultados obtenidos en la caracterización conceptual de dos bases de datos que contienen información sobre el rendimiento académico de estudiantes universitarios de las instituciones públicas de Angola.

Por último, se establecen las conclusiones, se emiten recomendaciones relacionadas con la continuidad de la investigación y se relacionan las referencias bibliográficas.

Capitulo-I “Fundamentación teórica”

En el presente capítulo se presenta un estudio sobre los sistemas inteligentes de apoyo al proceso de toma de decisión. Se presentaron además las técnicas de inteligencia artificial comúnmente utilizadas hoy día.

Se hace referencia a los principales elementos relacionados con objeto de estudio de la investigación que se centra mayormente en:

- ✓ Sistemas inteligentes y sus aplicaciones
- ✓ Sistemas para la toma de decisión
- ✓ Reconocimiento de patrones mediante el enfoque lógico combinatorio y algoritmos conceptuales.

1.1 Sistemas inteligentes

1.1.1 Definición de los sistemas inteligentes

La definición de **Sistemas Inteligentes (SI)** es un problema difícil y un tema de gran debate, en la perspectiva de la computación, los sistemas inteligentes son sistemas que simulan aspectos de la inteligencia humana, que pueden ser caracterizados por la flexibilidad, adaptabilidad, capacidad de aprendizaje y habilidad de gestionar una información (Krishnakumar, 2003).

(Ng, 2003) En su libro “*Intelligent Systems: fusion, tracking and control* 2Edition“plantea cuestiones cómo, ¿qué es considerado un sistema inteligente? ¿Una máquina que interactúa con un hombre puede ser clasificada inteligente?

Él define un sistema inteligente como cualquier sistema que procesa la información con eficiencia computacional, donde combina uno o más algoritmos inteligentes (en que los algoritmos pueden ser: bioinspirados, modelos matemáticos, algoritmos estadísticos, etc).

Independientemente de la definición, no hay duda de que la inteligencia artificial es la base esencial para desarrollar un sistema inteligente según (Russel & Norvig, 1995).

El Prof. (Zadeh, 1994) planteó un nuevo concepto para *machine intelligence* (máquinas inteligentes), separando por *Hard computing* y *Soft computing* (Ilustración. 1).

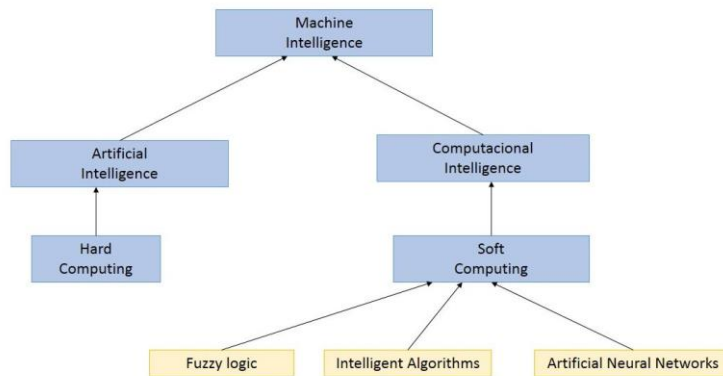


Ilustración 1 Inteligencia artificial vs inteligencia computacional

Hard computing es orientado al concepto de análisis y diseño de los procesos físicos y sistemas, son basados en lógica binaria, análisis numérico teoría de la probabilidad, ecuaciones diferenciales, análisis de funciones, teoría de la aproximación.

Soft computing es orientado al concepto de análisis y diseño de los sistemas inteligentes. Donde ellos se basan en técnicas de inteligencia artificial como la lógica difusa, las redes artificiales neuronales, los algoritmos genéticos, *machine Intelligent*, la minería de datos y otros.

Según **Prof. Lotfi A. Zadeh soft computing** es el consorcio de metodologías que permiten desarrollar sistemas inteligentes, de acuerdo a las habilidades humanas de tomar decisiones inteligentes en un determinado ambiente complejo.

1.2.2 Historia de los sistemas inteligentes

Al principio de la segunda guerra mundial (1939-1945), donde la tecnología disponible no era tan eficiente de manera tal que los intentos resultaban prematuros, desde allí se enmarcó el principio de la construcción de las máquinas inteligentes. Con el diseño de computadores analógicos pensados para controlar cañones o para la navegación.

Muchos de los investigadores observaron entonces que tenía una semejanza el funcionamiento de estos dispositivos de control y los sistemas reguladores de los seres vivos. De este modo, combinando las nuevas teorías sobre la realimentación, los avances de la electrónica de la posguerra y los conocimientos disponibles sobre los sistemas nerviosos de los seres vivos, se construyeron máquinas capaces de responder y de aprender como los animales. Ejemplos clásicos desarrollados en los años cincuenta son las tortugas de **W. Grey Walter**, que exhibían comportamientos sociales, o La Bestia de **J. Hopkins**, que guiada por un sonar y un ojo fotoeléctrico era capaz de encontrar un enchufe para alimentarse.

En el año 1950 se desarrollaron los primeros programas por Claud Shannon y Alan Turing, que permitían a un computador digital razonar y jugar ajedrez.

A. Newell, H. Simon y J. Shaw en el año 1957, presentaron el Teórico Lógico, el primer programa capaz de razonar sobre temas arbitrarios. Hacia 1960 John McCarthy acuña el término inteligencia artificial o IA, para definir los métodos algorítmicos capaces de hacer pensar a los ordenadores. En 1965 **Marvin Minsky, Newell y Simon** habían creado programas de IA que demostraban teoremas de geometría. Los desarrollos citados fueron de gran importancia, pero solamente eran capaces de resolver aquellos problemas para los que habían sido construidos. Sin embargo, los resultados eran tan alentadores que a finales de los sesenta los investigadores llegaron a pensar que en una década se conseguiría construir una máquina realmente inteligente.

La IA se ha desarrollado de una forma que en los años setenta con la introducción de los **sistemas expertos**, complejos programas de computador en los que se codifica el conocimiento de expertos en una materia muy concreta (concesión de créditos, diagnóstico de enfermedades, etc.), en forma de reglas de decisión. Después de pasar un cuarto de siglo los computadores pasaron a ser mil veces más potentes que los computadores de la época de los pioneros de la IA, sin embargo, en general, no resultan mucho más inteligentes. El problema radica en que al binomio lógica booleana-máquina **Von Neumann** sobre el que se asienta la IA, pese a su gran potencia, presenta problemas a la hora de abordar ciertas tareas, como aquellas denominadas del mundo real, donde la información que se presenta es masiva, imprecisa y distorsionada. Para abordar este tipo de tareas, desde hace unos años (década de los ochenta, principalmente) se han vuelto a retomar (o han surgido) una serie de paradigmas de cómputo alternativos, como las redes neuronales, los sistemas borrosos, algoritmos genéticos o la computación evolutiva, de los cuales los dos primeros quizás sean los más relevantes y empleados (Banda, 2011).

1.2.3 Estructura de los sistemas inteligentes

El principio de la generalidad dice que un sistema al tener una estructura lógica básica necesaria para solucionar un problema, entonces, en principio puede usarse la potencia del computador para llegar hacia la solución.

Hoy día los **Sistemas Expertos (SE)** o Sistemas Basado en Conocimiento (SBC) son bases para el desarrollar de los sistemas inteligentes, la idea de construcción es basada en dos conceptos:

- a) La adquisición, la representación y el uso de conocimiento especializado.

b) El reconocimiento de que los sistemas, además puedan poseer atributos propios de experto humano como :

*Capacidad de adquirir nuevo conocimiento.

*Capacidad para justificar sus conclusiones.

*Capacidad de hacer preguntas y explicar por qué las hacen.

*Capacidad conversacional

El concepto dado, aún puede causar confusión ya que muchos sistemas basados en programas convencionales podrían ser incorrectamente categorizados como sistemas basados en conocimiento. Esta inconsistencia puede ser aclarada, sobre la base de tres conceptos fundamentales que distinguen a los sistemas basados en conocimiento de los programas algorítmicos convencionales y de los programas generales basados en búsqueda:

- a) La separación del conocimiento y el modo en que es usado.
- b) La naturaleza del conocimiento empleado (heurística antes que algorítmica).
- c) El uso de conocimiento específico de un determinado dominio.

Los sistemas basados en caso presentan las siguientes características principales:

- a) Fácil modificación del conocimiento.
- b) Respuestas coherentes.
- c) Disponibilidad casi completa.
- d) Conservación del conocimiento.
- e) Capacidad de resolver problemas disponiendo de información incompleta.
- f) Capacidad de explicar los resultados y la forma de obtenerlos.

Los principales problemas asociados a ellos son:

- a) Las soluciones no siempre son las mejores.
- b) Conocimiento limitado frente al dominio de un experto.
- c) Carecen del sentido común o criterio que puede tener un experto.
- d) Es difícil extraer todo el conocimiento que maneja un experto.

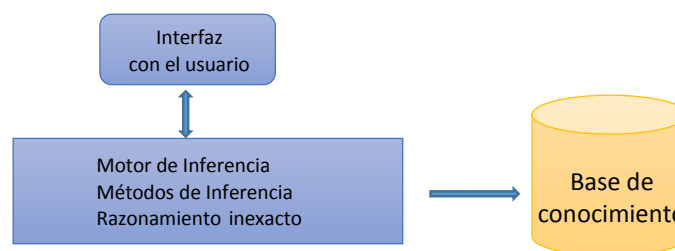


Ilustración 2 Estructura de un sistema inteligente

Interfaz con el usuario, es lo que posibilita a los expertos, de una forma interactiva y amigable, introducir en la base de casos todo el conocimiento y las experiencias acumuladas sobre el tema en cuestión. También es el componente a través del cual el usuario solicita al razonador que haga la búsqueda que requiere.

El motor de inferencia (MI), contiene un interpretador que decide cómo aplicar las reglas para inferir el nuevo conocimiento; y un administrador que decide el orden en el cual las reglas serán aplicadas. En otras palabras, selecciona, decide, interpreta y aplica conocimiento de la base de conocimiento con el fin de obtener soluciones.

Bases de conocimiento (BC), un conjunto de reglas (operaciones bien definidas) y una estrategia de control. La importancia de identificar un apropiado nivel de descripciones ayuda mucho porque un sistema inteligente puede consistir de muchas BC, operaciones y controles interactuando en una compleja forma computacional.

1.2.3 Aplicación de los sistemas inteligentes

Los sistemas inteligentes se aplican en una gran cantidad de campos y/o áreas, entre ellas pueden destacarse las siguientes:

1. Militar | Informática | Telecomunicaciones
2. Química | Derecho | Aeronáutica
3. Geología | Arqueología | Agricultura
4. Medicina | Industria | Finanzas y Gestión

Estos programas presentan la característica de trabajar con grandes cantidades de informaciones, que es uno de los grandes problemas a que se puede enfrentar un analista humano en el proceso de toma de decisión, los sistemas inteligentes analizan toda la información incluyendo los detalles más pequeños, todo este proceso de forma muy rápida, de manera tal que sea posible aportar una decisión más sólida.

1.3 Técnicas de inteligencia artificial

Hoy día las técnicas basadas en inteligencia artificial son más empleadas que desarrolladas, debido a la gran factibilidad de resolución de problemas. De estas técnicas, las más utilizadas son :

- ✓ Machine Learning
- ✓ Lógica difusa
- ✓ Redes neuronales artificiales
- ✓ Aprendizaje Automático

- ✓ Agentes inteligentes
- ✓ Algoritmos genéticos
- ✓ Estrategias evolutivas
- ✓ Técnicas de representación de conocimiento
- ✓ Redes semánticas
- ✓ Frames
- ✓ Visión artificial
- ✓ Audición artificial
- ✓ Lingüística computacional
- ✓ Procesamiento de lenguaje natural
- ✓ Minería de datos
- ✓ Reconocimiento de patrones

Entre ellas las más empleadas son la **lógica difusa, redes neuronales, algoritmos genéticos, agentes inteligentes, minería de datos y reconocimiento de patrones**, en nuestra investigación se abordara con más profundidad el reconocimiento de patrones, debido al hecho de que es la que se ha seleccionado para el desarrollo de la aplicación.

1.3.1 Reconocimiento de patrones

Reconocer patrones es una habilidad extremadamente desarrollada por los seres humanos y algunos animales. El ser humano tiene la facilidad de reconocer rostros, voces, caligrafías, y hasta el estado de humor de las personas. Parte de los animales también presentan estas características.

Se comprende como **patrones**, propiedades o características que ayudan al agrupamiento de objetos semejantes dentro de una determinada clase o categoría, mediante el criterio de entrada que le permite extraer los rasgos más relevantes (Tou & Gonzales, 1981). Al conjunto de características o propiedades comunes relacionado al objeto de estudio se le puede llamar clase.

El estudio del reconocimiento de patrones según (Tou & Gonzales, 1981) puede ser dividido en dos categorías básicas:

1. Estudios de los seres humanos y organismos vivos con el objetivo de establecer los modos, que les permiten desarrollar sus capacidades de reconocer patrones.
2. El desarrollo de teorías o técnicas con el objetivo de construir máquinas o dispositivos capaces de presentar características parecidas a las de los seres humanos en el reconocimiento de patrones.

1.3.1.1 Definición

De acuerdo a (Ruiz-Schulcloper, Guzmán Arenas, & Martínez, 1999) reconocimiento **de patrones** se puede definir como todo aquello relacionado con la clasificación de objetos y fenómenos y con la determinación de los factores que inciden en los mismos.

1.3.1.2 Enfoques en el reconocimiento de patrones

Para resolver problemas de reconocimiento de patrones han surgido varios enfoques como:

- ✓ Reconocimiento estadístico de patrones
- ✓ Reconocimiento sintáctico de patrones
- ✓ Redes neuronales
- ✓ Reconocimiento lógico combinatorio

Reconocimiento estadístico de patrones: Este enfoque se basa en la teoría de probabilidad y estadística y se supone que tiene un conjunto de medidas numéricas con distribución de probabilidad conocida o estimable, y a partir de ella se hace el reconocimiento (Tou & Gonzales, 1981).

Reconocimiento sintáctico de patrones: Este enfoque se basa en los objetos que no están descritos como vectores de atributos, por ejemplo objetos descritos como: grafos, cadenas de símbolos, etc. Normalmente el objetivo es encontrar las relaciones estructurales que guardan los objetos de estudios (Fu, 1982) (Gonzales & Thomason, 1978).

Redes neuronales: Este enfoque se basa en resolver los problemas de reconocimiento de patrones mediante redes neuronales artificiales, las cuales pueden ser entrenadas para dar una cierta respuesta cuando se les presentan determinados valores numéricos en sus entradas. De este modo una red neuronal artificial puede dar una respuesta similar cuando se presenta una entrada parecida a las que se usaron para entrenarla (Ripley, 1982).

Reconocimiento lógico combinatorio: A lo largo del trabajo se usar este enfoque, este constituye una alternativa a los enfoques antes mencionados y se basa en la idea de que el modelado del problema debe ser lo más cercano posible a la realidad del mismo, sin hacer suposiciones que no estén fundamentadas. Uno de los aspectos esenciales del enfoque es que los atributos utilizados para describir a los objetos de estudio deben ser tratados cuidadosamente para no realizar operaciones que resulten antinaturales respecto al problema que están representando. Este tratamiento cuidadoso permite trabajar con atributos cualitativos y cuantitativos e incluso con ausencia de información (Ruiz-Schulcloper, Guzmán Arenas, & Martínez, 1999).

1.3.1.3 Problemas en el reconocimiento de patrones

En el reconocimiento de patrones uno de los problemas más abordados es la **selección de variables**, a partir de un conjunto de variables pertenecientes a los objetos en estudio. Esta selección tiene como objetivo identificar dentro del conjunto de variables las más relevantes, este proceso ocurre debido al hecho de mejorar el proceso de reconocimiento de patrones, de esta forma se puede disminuir el tiempo y aumentar la calidad de los resultados. Sin embargo este proceso conlleva aún otros estudios.

La **clasificación supervisada**, consiste en clasificar nuevos objetos a partir de una previa información de los objetos a clasificar. Por otra parte tenemos el **agrupamiento conceptual** generalmente llamado de **clasificación no supervisada**, que consiste en dado un conjunto de objetos encontrar las clases que agrupan dichos objetos, en la presente investigación se utilizará el enfoque de los algoritmos conceptuales.

1.3.1.4 Aplicación del reconocimiento de patrones

Hoy día las técnicas de reconocimiento de patrones son aplicadas en casi cualquier área del conocimiento, Según (Ruiz-Schulcloper, Guzmán Arenas, & Martínez, 1999) la aplicación del reconocimiento de patrones es más frecuente en las ciencias poco formalizadas¹, como la geología, geofísica, medicina, psicología, criminología, etc, en las cuales es difícil que los expertos del área construyan algoritmos para resolver determinados problemas.

El reconocimiento de patrones presenta una gran ventaja debido al hecho de que puede ser entrenado con casos conocidos (casos estos que presentan experiencia del experto). A continuación se presentan algunas aplicaciones del reconocimiento de patrones en dichas ciencias.

Número	Nombre
1	Determinación de síndromes de enfermedades
2	Diagnóstico diferencial de enfermedades
3	Pronóstico de complicaciones posoperatorias
4	Pronóstico de la calidad de vida al nacer
5	Determinación de la terapéutica adecuada para un paciente dado y muchos otros

Tabla 1 - Reconocimiento de patrones en las ciencias biomédicas

¹ Según (Ruiz-schulcloper,1999) Ciencias que no existe una regla que se resuelva todos los problemas , ciencias poco formalizadas (Geociencias, medicina, pedagogía, psicología, criminalística y otros)

Número	Nombre
1	Pronóstico de perspectiva de minerales
2	Pronóstico de magnitud máxima de terremotos
3	Pronóstico de ocurrencia de terremotos
4	Pronóstico meteorológicos y muchos otros

Tabla 2 - Reconocimiento de patrones en las Geociencias

Número	Nombre
1	Determinación de los factores que influyen en el surgimiento de actividades delictivas
2	Tipificación de lugares propensos a ser asaltados
3	Determinación del modus operandi de grupos delictivos y muchos otros

Tabla 3 - Reconocimiento de patrones en la Criminalística

Número	Nombre
1	Pronóstico del surgimiento de actividades revolucionarias
2	Pronóstico de líderes
3	Caracterización de grupos poblacionales
4	Caracterización de la dinámica poblacional en diferentes zonas geográficas y muchos otros

Tabla 4 - Reconocimiento de patrones en la Sociología

Número	Nombre
1	Estratificación de grupos de estudiantes a partir de sus habilidades para ciertas asignaturas
2	Determinación de programas de capacitación y superación diferenciada en colectivos de estudiantes
3	Diagnóstico precoz de estudiante sobresaliente (pronóstico de genios) y muchos otros

Tabla 5 - Reconocimiento de patrones en la Pedagogía

1.3.1.5 Algoritmos de agrupamiento conceptual

La clasificación no supervisada o agrupamiento conceptual, como muchas bibliografías lo referencian, normalmente es conocido como el proceso de agrupar un conjunto de objetos, donde cada grupo presenta objetos que son más semejantes entre sí que a los objetos de otros grupos, de acuerdo con un criterio de selección entre ellos.

En la presente investigación se ha seleccionado los algoritmos conceptuales, debido el hecho que muchos algoritmos solo se limitan a agrupar los objetos, y muchas veces se necesita además de los grupos, sus conceptos, como por ejemplo en la caracterización de los estudiantes, pero este problema se deja a cargo de los expertos. **R. S. Michalski** a

finales de los años 70 y principios de los 80, introdujo un conjunto de ideas que dieron origen al agrupamiento conceptual y a la familia de algoritmos CLUSTER. Su objetivo fue presentar al usuario un listado de los elementos (llamado de concepto) relacionando a cada grupo obtenido en la clasificación; el número de grupos también es un problema a resolver, la clasificación no supervisada presenta dos tipos fundamentales de problemas que son **libre** en caso de que se desconozca el número de grupos a formar y **restringido** en el caso de que sea solo un dato del problema.

Los algoritmos conceptuales se basan en la Teoría Clásica de los conjuntos, donde el concepto de conjunto es primario. Por lo que hay dos conceptos, uno es denominado **extensional** (que consiste en presentar una lista relacionada a los objetos de un conjunto) y otro **intencional** (consiste en presentar la propiedad que caracteriza a los objetos de un conjunto), los algoritmos clásicos utilizan el primer enfoque mientras los algoritmos conceptuales utilizan el segundo enfoque.

La clasificación no supervisada según (Michalski R. S., 1981) plantea diferentes problemas, que son:

1. Determinar las características de la proximidad que se debe utilizar para agrupar.
2. Encontrar una buena medida de proximidad para resolver problemas en particular.

Generalmente la proximidad entre objetos se asume como una medida en un espacio multidimensional que opera sobre un conjunto fijo de rasgos que caracterizan a los objetos.

La mayoría de las medidas de proximidad empleadas consideran a todos los atributos con igual importancia, por lo que no hacen distinción entre atributos que pueden ser más o menos relevantes para el problema a resolver. Son escasos los mecanismos para seleccionar y evaluar atributos en el proceso de generación de los grupos, o sea, casi no se tienen en cuenta las consideraciones que emplean los métodos humanos para agrupar objetos. Observaciones de cómo las personas agrupan objetos indican que existe una tendencia a seleccionar uno o unos cuantos atributos relevantes y agrupar los objetos teniendo en cuenta solo estos atributos (Michalski R. S., 1981).

El aporte que ofrecen los algoritmos conceptuales es sin duda de una gran utilidad práctica. Debido al hecho de no limitarse a apenas darle al usuario el listado de los objetos agrupados, sino también presentar un conjunto de características distintas.

1.3.1.5.1 Teoría de algoritmos conceptuales

Se considera un conjunto de objetos $O = \{O_1, O_2, \dots, O_n\}$, $R = \{r_1, r_2, \dots, r_l\}$ un conjunto de rasgos que describe un objeto. Comprendiéndose como representación de objeto i -ésimo O_i el l -uplo $x_i = (r_1(O_i), \dots, r_l(O_i)) = (x_{i1}, \dots, x_{il}) \in \Omega = \Omega_1 \times \dots \times \Omega_l$, donde $r_j(O_i) = x_{ij} \in \Omega_j$ es el valor que toma el rasgo r_j en el objeto O_i . El conjunto Ω_j es el conjunto de los valores admisibles del rasgo r_j , el conjunto $X = \{X_1, X_2, \dots, X_n\}$ es la representación de conjunto de objeto.

Dentro de la representación del universo de objeto, se conoce como función **de semejanza** Γ , como una función que se aplica a un par de objeto O_i, O_j obteniendo como resultado un valor real que indica qué tan semejante son dichos objetos. Formalmente se puede definir como $\Gamma : \mathbb{O} \times \mathbb{O} \rightarrow \mathbb{R}$.

Existen dos tipos de funciones de semejanza, que son:

1. Similitud: se considera que cuando se obtiene un valor mayor $\Gamma(O_i, O_j)$, más semejante son los objetos.
2. Disimilitud: en este caso se considera por medio de un valor menos, mientras menos es el valor $\Gamma(O_i, O_j)$, más semejantes son los objetos.

Según (Ruiz-Schulcloper y otros, 1999) esta definición está dada de la forma más general posible, hasta ahora no existe un consenso de la misma, existen varios puntos de vista, entre ellos a veces asumen el cumplimiento de determinadas propiedades como por ejemplo: la simetría, la no negatividad, etc., sin tener en cuenta realmente el tipo de dato sobre el que se va a aplicar.

El espacio de representación inicial de los objetos es considerado como el producto cartesiano Ω , y se asume como la naturaleza de las variables o tributos como cualquiera:

1. Univaluada (cualitativa nominal u ordinal, cuantitativa)
2. No Univaluada (conjunto o intervalo) o estructurada (tipo árbol o grafo, etc)

En el problema de la clasificación no supervisada, el objetivo es encontrar una estructura de un conjunto de descripciones de objetos en el espacio de representación, es decir hallar una familia $P = \{C_1, \dots, C_m\}$, donde $C_i \subseteq \mathbb{O}$, $C_i \neq \emptyset$ tal que $\bigcup_{i=1}^m C_i = \mathbb{O}$, hay veces que es necesario $C_i \cap C_j = \emptyset$, $i \neq j$ (partición). En una primera instancia el resultado obtenido depende de la selección del propio espacio de representación inicial, al mismo tiempo de una función de semejanza Γ y de un criterio de agrupamiento Π , así mismo se observa la forma que vamos utilizar Γ .

El agrupamiento de cada objeto para su determinado grupo de objetos se hace según un criterio de agrupamiento todo eso por medio de una función de semejanza, estos dos conceptos son muy importantes en el problema de la clasificación no supervisada, y su determinación debe estar relacionada en el conocimiento del problema particular que se está modelando (Ruiz-Schulcloper, et al., 1999).

Se dice que $NU \subseteq O, NU \neq \emptyset$ es una componente conexa si se cumple:

1. $\forall o_i, o_j \in NU, o_i \neq o_j, \exists o_{i_1}, \dots, o_{i_q} \in NU \left[o_i = o_{i_1} \wedge o_j = o_{i_q} \wedge \forall p = 1, \dots, q - 1 \Gamma(o_p, o_{p+1}) \geq \beta_o \right]$.
2. $\forall o_i \in O \left[o_j \in NU \wedge \Gamma(o_i, o_j) \geq \beta_o \right] \Rightarrow o_i \in NU$
3. *Todo elemento β_o - aislado es un componente conexo (degenerado)*

Sea $NU \subseteq O, NU \neq \emptyset$, es un conjunto compacto si se cumple que [1]:

1. $\forall o_j \in O \left[o_i \in NU \wedge \max\{\Gamma(o_l, o_t)\} = \Gamma(o_i, o_j) \geq \beta_o \right] \Rightarrow o_j \in NU$
2. $\left[\max\{\Gamma(o_p, o_i)\} = \Gamma(o_p, o_t) \geq \beta_o \wedge o_t \in NU \right] \Rightarrow o_p \in NU$
3. *|NU| es mínimo*
4. *Todo elemento β_o - aislado constituye un conjunto compacto (degenerado)*

En las componentes conexas, la condición 1 dice que todos los objetos de NU tienen en NU a sus objetos más β_o -semejantes; La condición 2 dice que no existe fuera de NU un objeto cuyo objeto más β_o -semejante esté en NU ; La condición 3 dice que NU es el conjunto más pequeño que cumple las condiciones 1 y 2.

El criterio compacto, al igual que el criterio de las componentes conexas de una partición, esta partición es además única para cada conjunto de datos dados.

De las componentes conexas y conjuntos compactos, se cumple: Toda componente conexa es la unión finita de un conjunto compacto.

Ejemplo:

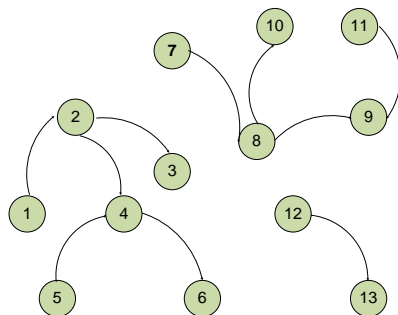


Ilustración 3 - Árbol de objetos

Componente conexa:

NU1 = {1,2,3,4,5,6,7,8,9,10,11} NU2 = {12,13}

Conjunto compacto:

NU1 = {1,2,3,4,5,6} ; NU2 = {7,8,9,10,11}; NU3 = {13,16,17,18}

El problema de la estructuración conceptual sobre \mathbb{O} consiste en hallar los conceptos que describen a cada grupo correspondiente a una estructuración y representarlos de una forma natural e interpretables por los humanos. Para realizar esta tarea se define como función de semejanza Γ la cohesividad conceptual (Michalski & R.E.).

Ellos se pueden clasificar en dos grandes grupos, que son:

- 1. Algoritmos incrementales
- 2. Algoritmos no incrementales

Algoritmos incrementales el funcionamiento de estos algoritmos depende de los grupos ya existentes, en relación a los nuevos objetos, es decir, siempre que se recibe un objeto nuevo, este es clasificado según una determinada estrategia, se incorpora en uno de los grupos ya existentes o se crea un nuevo grupo para este objeto. Generalmente estos algoritmos se utilizan en el caso en que el conjunto de objetos a estructurar no está completamente dado. Ejemplos de algoritmos de este tipo son: UNIMEM, COBWEB y Galois.

Algoritmos no incrementales estos algoritmos trabajan con una estructura de datos completos, es decir del conjunto que se recibe no se van a incorporar nuevos objetos. Ejemplos de estos algoritmos son: CLUSTER/PAF, CLUSTER/2, y CLUSTER/S, WITT, K-Means conceptual y LC'conceptual (Ruiz-Schulcloper, et al., 1999).

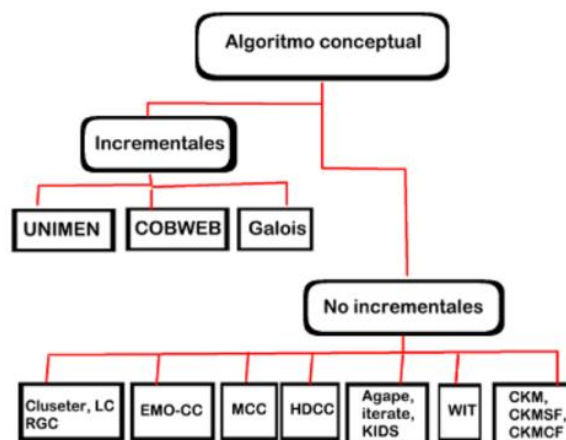


Ilustración 4 - Taxonomía de los algoritmos conceptuales, adaptado de (Guerra-Gandón, Alejandro; Vegas-Pons, Sandro; Ruiz-Shulcloper, José, 2012)

1.3.1.5.2 Conceptos básicos

Los conceptos señalados abajo son dados según (Ruiz-Shulcloper, et al., 1999) y (Suárez , et al., 2014):

1. Selector: sea $r_j \in R$ un rasgo, $s_j \subseteq D$ un subconjunto del dominio de valores asociado al rasgo r , y $\#$, un operador de conjunto del conjunto $\{\#, \neq, \geq, >, <, \leq, =\}$. Un con selector es una expresión del tipo $[r\#s_j]$.

2. Testores: Es un conjunto de rasgos que discriminan objetos de diferentes clases.

3. L-complejo: se entiende como l-complejo al producto lógico de dos o más selectores. Un objeto satisface a un l-complejo o l-complejo cubre al objeto, si el objeto satisface los selectores de lo l-complejo. Al conjunto de todos objetos que satisfacen a un l-complejo se llaman s-complejo, generalmente la descripción de un l-complejo se define como de objetos que satisfacen los l-complejos. Se llama l-complejo maximal, dada una propiedad P, si no existe otro l-complejo l' con la propiedad P tal que $l \subset l'$.

4. Refunion: un conjunto de l-complejo $L = \{l_1, l_2, \dots, l_k\}$, se llama refunion de L, expresado por LU (L), siempre que haya una transformación del conjunto L en l-complejo l que tiene la dispersión mínima entre todos los l-complejos que cubren al mismo conjunto de objetos, se puede entender como LU (L).

5. Estrella: Un conjunto de objetos $\mathbf{o} \in O$ y $F \subseteq O$, la estrella de \mathbf{o} con respecto F se expresa por $G(\mathbf{o}, F)$, es el s-complejo maximal respecto a la inclusión y que no cubren a los objetos de F.

6. GEN: es un operador cuyo su objetivo es simplificar generalizar un determinado conjunto de l-complejos, este operador es proveniente del estudio realizado en aprendizaje inductivo, el operador utiliza un conjunto de regla que depende de tipo de rasgos de cada selector de dicho l-complejo.

7. Matriz de aprendizaje : Sea M un conjunto de objetos admisibles x_1, \dots, x_n rasgos que caracterizan a los elementos de M y cuyos conjuntos de valores admisibles son M_1, \dots, M_n sobre los que se han definido criterio de comparación C_1, \dots, C_n respectivamente; K_1, \dots, K_r es un cubrimiento finito de subconjunto propio de M; dada una información estándar verdadera $l_0(K'_1, \dots, K_r)$ de las clases K_1, \dots, K_r donde $K'_i \subset K_i$, $i = 1, \dots, r$.

8. Matriz de diferencia: Sea T_{mnr} un testor y $\Gamma = (c_1, \dots, c_n)$ una medida de semejanza entre objetos definido en términos de los criterios de comparación booleano C_i , $i = 1, \dots, n$. Llamaremos matriz de diferencia de T_{mnr} a la matriz booleana M_{nm} .

9. Matriz básica: Dada una matriz de comparación M llamaremos matriz básica a la matriz MB formada exclusivamente por las filas básicas de M .

1.3.1.5.3 Algoritmo LC-conceptual

LC-conceptual es un algoritmo conceptual para datos incompletos y mezclados, no incremental propuesto por (Martínez-Trinidad, 2000), es basado en la teoría de los grafos, el algoritmo retoma algunas ideas de Michalski para generar los conceptos.

Se basa en dos etapas, la primera etapa es la determinación intencional, esta consiste en hallar el agrupamiento de los objetos en diferentes clases mediante su semejanza, en la etapa intencional, del conjunto de clases de objetos se generan los conceptos de cada clase.

Sea un conjunto de objetos $O = \{O_1, O_2, \dots, O_n\}$, donde cada objeto conserva un conjunto de rasgos (Booleano, multi-valuados, lingüístico, numérico, etc) hasta aquellos que no se conoce su valor, y una función de semejanza $\Gamma(O_i, O_j)$ que determina el grado de semejanza entre dos objetos perteneciente a la colección de objetos (Guerra-Gandón, Alejandro; Vegas-Pons, Sandro; Ruiz-chulcloper, José, 2012).

Después de cumplir la determinación extensional de cada grupo, en la segunda etapa LC describe intencionalmente a cada grupo, generando los conceptos, definiendo un l -complejo de la familia CLUSTER (Mischalski, 1983), y con operador refunión y método GEN perteneciente a la misma familia CLUSTER, LC propone describir un grupo G , por los l -complejos constituidos por el conjunto de los testores típicos.

A continuación se presenta una breve descripción del algoritmo en pseudocódigo, donde se puede verificar paso por paso.

Paso 1 : selección de los parámetros : $C_i, i=1, \dots, n, \Gamma, \Pi$ y W
Paso 2 : Obtención de los grupos K_1, \dots, K_c de la MI aplicando Π
Paso 3: Para cada grupo $K_i, i=1, \dots, c$, haz

i) Calcular la estrella $G_\tau(k_i \setminus k_1, \dots, k_{i-1}, k_{i+1}, \dots, k_c)$ aplicando RU .
ii) Aplicando GEN a todos l -complejos obtenido arriba y construir y generalizar los l -complejos $\alpha_1, \dots, \alpha_q$
iii) Los conceptos caracterizado por los grupos k_i en términos τ es:

$$P_i(\alpha, v, \dots, v \alpha q) \wedge \neg(\alpha_{0_1} v, \dots, \alpha_{0_x});$$

donde $\alpha_{0_j}, j = 1, \dots, t$ elemental en el l -complejo asociados a los objetos $O_1, \dots, O_t \in MI \setminus k_i$ que satisfacen a los conceptos $\alpha, v, \dots, \alpha q$

Ilustración 5 - Algoritmo de agrupamiento conceptual LC propuesto por Martínez-Trinidad, 2000

1.3.1.5.4 Algoritmo RGC

El algoritmo RGC, es un algoritmo conceptual no incremental propuesto por Aurora Pons en el año 2002, tal como el algoritmo LC, considerando un conjunto de objetos y sus características, su objetivo es encontrar un grupo de clases de objetos y sus determinados conceptos en el espacio inicial.

El algoritmo presenta dos etapas principales, estas que son:

1. **Determinación intencional:** creación de los grupos usando una medida de similitud entre objetos y un criterio de agrupamiento para generar la estructuración, donde se puede emplear cualquier algoritmo de clasificación no supervisada.
2. **Determinación extensional:** creación y generalización de los conceptos de cada grupo obtenido en la primera etapa.

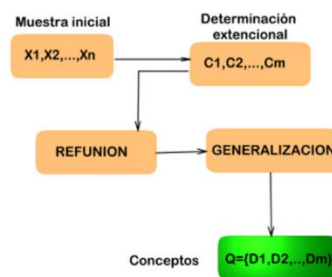


Ilustración 6 - Proceso del algoritmo RGC, adaptado de (Guerra-Gandón, Alejandro; Vegas-Pons, Sandro; Ruiz-shulcloper, José, 2012)

Para que sea procesada, en la segunda etapa, primero se selecciona un sistema de conjunto de apoyo que deberá responder a las necesidades del problema. Para cada conjunto de apoyo y cada grupo se construirán los conceptos (I-complejos) que los caracterizan utilizando el operador de reunión extendida (RUE) los I-complejos obtenidos mediante el operador RUE cumplen la propiedad de ser excluyentes para cada grupo. Posteriormente estos I-complejos son simplificados y generalizados aplicando el operador GEN en cada variable de dichos I-complejos.

El operador GEN aumenta la dispersión de los I-complejos. A pesar de que al aplicar el operador GEN en cada variable del I-complejo se exige que no pierda su condición de ser excluyente para el grupo, al unir todas las generalizaciones de las variables en el I-complejo generalizado, puede perder esta propiedad. Es por eso, que una vez generalizados los I-complejos es necesario determinar los objetos de los restantes grupos que los satisfacen y construir los conceptos. El conjunto de I-complejos construido mediante el operador RUE pudiera ser diferente si se cambia el orden de presentación de los objetos. Esto no es

importante, pues el objetivo es encontrar un concepto que caracteriza al grupo y no un único concepto. Además, resulta natural que un grupo no posea una única propiedad que lo caracteriza. La posibilidad de encontrar varias propiedades que caracterizan a un mismo grupo puede resultar, incluso, interesante para el especialista en el sentido de poder analizar los resultados desde distintos puntos de vista. El algoritmo retoma las ideas de Michalski y realiza extensiones de varios operadores propuestos en los algoritmos CLUSTER para obtener descripciones de los grupos. Es un método para encontrar los conceptos de los grupos independientemente de la forma en que se construyeron los mismos, ya que propone utilizar cualquier algoritmo en la fase de agregación (Guerra-Gandón, Vegas-Pons, & Ruiz-shulcloper, 2012).

El algoritmo **RGC** y **LC** son los únicos algoritmos conceptuales encontrado en la presente bibliografía que se adecuan al problema de la investigación, debido al que esto dos algoritmos están diseñado para los datos incompletos y mezclados, qué es el caso del estudio de análisis del rendimiento académico de los estudiantes, que cuenta con varios tipos de variables.

A continuación se presenta una breve descripción del algoritmo en pseudocódigo, donde se puede verificar paso por paso.

Paso 1 : selección de los parámetros : $C_i, i=1, \dots, n, \Gamma, \Pi$ y W

Paso 2 : Obtención de los grupos K_1, \dots, K_c de la MI aplicando Π

Paso 3: Para cada grupo $K_i, i=1, \dots, c$, haz

i) Calcular la estrella $G_\tau(k_i \setminus k_1, \dots, k_{i-1}, k_{i+1}, \dots, k_c)$ aplicando RUE.

ii) Aplicando GEN a todos l-complejos obtenido arriba y construir y generalizar los l-complejos $\alpha_1, \dots, \alpha_q$

iii) Los conceptos caracterizado por los grupos k_i en términos τ es:

$$P_i(\alpha, v, \dots, v \alpha q) \wedge \neg(\alpha_{0_1} v, \dots, \alpha_{0_x});$$

donde $\alpha_{0_j}, j = 1, \dots, t$ elemental en el l-complejo asociados a los objetos

$O_1, \dots, O_t \in MI \setminus k_i$ *que satisfacen a los conceptos $\alpha, v, \dots, \alpha q$*

Ilustración 7 - Algoritmo de agrupamiento RGC propuesto por Aurora Pons, 2002

El tiempo de ejecución del algoritmo depende del tipo de rasgo que describe los objetos, del criterio de comparación, de la medida de similitud y del criterio de agrupación. A continuación se presenta una tabla de comparación del RGC y los algoritmos más representativos de agrupamiento conceptual, se analizó la potencialidad de los algoritmos mediante unos rasgos, entre los más significativos tenemos, **características** (enseña si todos los objetos pertenecientes a los grupos satisfacen los conceptos), **excluyente** (si ninguno de los objetos satisface los conceptos de otro grupo diferente), **clasificación** (si las descripciones son obtenidas de una determinación extensional de los grupos).

Algoritmo	Características	Excluyente	clasificación	Claridad
K-MEANS	No	No	Sí	Poco
LC	No	Sí	Sí	Sí
COBWEB	No evaluado	No evaluado	No	No
LINNEO	No evaluado	No evaluado	No	No
RGC	Sí	Sí	Sí	Sí

Tabla 6 - Comparación de los algoritmos de agrupamiento conceptuales

1.3.1.6 Teoría de los testores

En las descripciones de los objetos en término de sus rasgos, es posible observar que generalmente no son necesario todos ellos para diferenciar una clase de otras y clasificarlas correctamente. Así mismo surge la necesidad de encontrar qué subconjunto de los rasgos son suficiente para caracterizar las clases. Para determinar cuáles rasgos son los que clasifican mejor los objetos, no es una tarea fácil, hoy día existen muchos algoritmos que se dedican al problema de selección de variables, como el algoritmo BT, LEX, CC, Ancora y otros. En este trabajo se aplicó una medida de importancia de los rasgos, generalmente cuándo el número de testores típico es grande cada agrupamiento podrá presentar una gran cantidad de conceptos, lo que se propone siempre utilizar un criterio para seleccionar los mejores testores típicos para construir los conceptos.

El concepto de testor se puede entender como un “subconjunto de rasgos del n-uplo informacional X_1, \dots, X_s , ($s \leq n$) que cumple la condición de que no existen descripciones (parciales) semejantes de objetos en clases diferentes” (Ruiz-Shulcloper, et al., 1999). Su longitud s es menor o igual a la cantidad de rasgos que describían al objeto antes de la selección.

Análogamente, se denomina **testor típico** a la “combinación mínima de rasgos que tipifican o caracterizan una clase de objetos, diferenciándolas de las demás clases” (Ruiz-Shulcloper, et al., 1999); la eliminación de cualquier rasgo del testor elimina su condición de tipicidad; es decir, la propiedad que distingue a los testores típicos de los testores es que cada rasgo resulta imprescindible para mantener las diferencias entre las clases.

Algunos de estos rasgos aparecen en muchas combinaciones seleccionadas como testores típicos, resultando difícil prescindir de ellos, su importancia aumenta cuando se trata de testores típicos de menor longitud. Sobre la base de esta idea, Zhuravliov formuló **SU** definición de **peso informacional de un rasgo** considerando dos magnitudes: la frecuencia de aparición del rasgo en la familia de testores típicos y la longitud de los testores típicos en los cuales aparece el rasgo. De esta manera la importancia informacional de cada rasgo se

encuentra en función de la influencia del mismo en la calidad de clasificación de un cierto algoritmo (Ruiz-Shulcloper, et al., 1999).

Utilizando esta definición se han resuelto una serie de problemas interesantes tales como : la comparación de la estructura geológica de yacimiento minerales (Konstantinov), el pronóstico de objetos y fenómenos geológicos (Dmitriev), el diagnóstico diferencial de la esquizofrenia y enfermedades orgánicas del cerebro (Volkov), clasificación de materiales sociológicos (Zogoroiko), determinación de la carga subjetiva en obreros de una fábrica de calzado (Douglas-la-peña), pronóstico de tormentas ionosféricas (Lopez-Reyes), pronóstico de magnitud máxima sísmica (Ruiz) y otros muchos trabajos realizados en diferentes países.

1.4 Sistemas para el proceso de toma de decisión

1.4.1. Definición de los sistemas para la toma de decisión

Los **sistemas de soporte a las decisiones (DSS** por sus siglas en inglés *Decision Support System*) pueden adoptar muchas formas diferentes. En general, podemos decir que un DSS es un sistema informático utilizado para servir de apoyo, más que automatizar, el proceso de toma de decisiones. La decisión es una elección entre alternativas basadas en estimaciones de los valores de esas alternativas. El apoyo a una decisión significa ayudar a las personas que trabajan solas o en grupo a reunir inteligencia, generar alternativas y tomar decisiones. Apoyar el proceso de toma de decisión implica el apoyo a la estimación, la evaluación y/o la comparación de alternativas. En la práctica, las referencias a DSS suelen ser referencias a aplicaciones informáticas que realizan una función de apoyo (Alter, 1980).

1.4.2 Características de los sistemas para la toma de decisión

Los sistemas para la toma de decisión están dirigidos a los problemas no estructurados (semiestructurado), estos sistemas dan más soporte a los tomadores de decisión. Algunos de ellos ayudan permitiendo el acceso a la información importante en el momento oportuno, otros contienen modelos explícitos que proveen de una estructura para decisiones particulares. Ellos son esencialmente sistemas interactivos (hombre-máquina); la creencia en la utilidad de un funcionamiento interactivo en los procesos de toma de decisiones se basa en parte en la idea de que hay muchos problemas no estructurados o semiestructurados que se pueden manejar más efectivamente a través de un proceso de solución de problemas interactivo, es decir un proceso de búsqueda en el cual una o más personas (los que tienen que tomar la decisión) junto con la computadora exploran un espacio de problemas buscando la mejor conceptualización posible del problema y la mejor acción a ejecutar para resolverlo (Haettenschwiler, 1999).

Considerando que la toma de decisión está presente en innumerables situaciones los DSS tienen muy diversos campos de aplicación. Entre ellos resultan de mucho interés los negocios, el diseño de ingeniería, la medicina, las estrategias militares, la política, y el manejo de recursos en general. Especial significación tienen en el primer y último campo mencionado por la trascendencia económica de las decisiones. Un DSS es más un servicio que un producto. Como el problema puede solamente ser parcialmente estructurado, y el que tiene que tomar la decisión incrementa su comprensión del mismo y sus necesidades con el tiempo, el DSS tiene que constantemente crecer y evolucionar en la medida que el usuario se adapta y aprende.

Al igual que ocurre con la definición, no existe una taxonomía universalmente aceptada para los DSS. Diferentes autores proponen diferentes clasificaciones. Utilizando la relación con el usuario como criterio, (Haettenschwiler, 1999) distingue entre:

- **DSS pasivo.**- Es un sistema de ayudas para el proceso de toma de decisiones, pero que no puede llevar a cabo una decisión explícita, sugerencias o soluciones.
- **DSS activo.**- Puede llevar a cabo dicha decisión, sugerencias o soluciones.
- **DSS cooperativo.**- Permite al encargado de la toma de decisiones (o a sus asesores) modificar, completar o perfeccionar las sugerencias de decisión proporcionadas por el sistema, antes de enviar de vuelta al sistema para su validación. El nuevo sistema mejora, completa y precisa las sugerencias del tomador de la decisión y las envía de vuelta a su estado para su validación. Entonces, todo el proceso comienza de nuevo, hasta que se genera una solución consolidada.

Según (Sprague & E. D. Carlson, 1982) identifican tres componentes básicos que son explicados con más detalles por (Haag, Cummings, McCubbrey, Pinsonneault, & Donovan, 2000):

- **El sistema de gestión de base de datos.**- Almacena información de diversos orígenes, puede proceder de los repositorios de datos de una organización tradicional, de fuentes externas (como Internet), o del personal (de ideas y experiencias de los usuarios individuales).
- **El sistema gestor de modelos.**- Se ocupa de las representaciones de los acontecimientos, hechos o situaciones utilizando varios tipos de modelos (dos ejemplos serían modelos de optimización y modelos de búsqueda-objetivo).
- **El sistema gestor y generador de diálogos.**- Se trata de la interfaz de usuario; es, por supuesto, el componente que permite a un usuario interactuar con el sistema.

1.4.3 Sistema para la toma de decisión en la educación

Actualmente en las universidades y centros de enseñanza se trabaja en la implementación de herramientas o soluciones que les ayuden en el logro de sus tareas. Alcanzar dichas metas, en muchas ocasiones es necesario tomar decisiones correctas o buenas, pero sin duda eso constituye un proceso complejo que involucra responsabilidades y riesgos, debido al hecho de que es necesario apostar a las nuevas tecnologías. Hoy día las universidades cuentan con tecnología que les permite automatizar sus procesos, en estos procesos conlleva el poder de tomar decisiones para mejorar los resultados y logros obtenidos, de ahí estos sistemas sirven como medio para lograr estos objetivos.

En los sistemas basados en conocimiento para la educación, el nuevo conocimiento descubierto muestra de forma clasificada la información solicitada y sus descripciones, de manera tal que se tenga un conocimiento más profundo del que existe en el universo estudiantil y sirve como referencia e influencia para tomar decisiones estratégicas en la mejora continua de la calidad de la enseñanza y por supuesto en el alcance de mejores rendimientos académicos en los estudiantes. **Romero y Ventura (2007)**, afirman que el descubrimiento del conocimiento en la educación puede ser tomado de diferentes actores (orientado a los estudiantes, orientado a los educadores, orientado a los gestores de la educación) de acuerdo a los intereses de cada uno.

Como se ha comentado en el transcurso del trabajo, muchos estudios han sido desarrollados para descubrir conocimiento dentro del proceso de enseñanza universitaria, de manera tal que se tenga un mejor conocimiento del proceso y ayude a tomar decisiones estratégicas.

En el caso de la situación actual del proceso universitario público en Angola, se tiene en muchas instituciones superiores el proceso académico automatizado, semiautomatizado o no automatizado, lo que dificulta el análisis de los datos estudiantiles. El desarrollo de una solución que colecte los datos de varias fuentes y ofrezca a los tomadores de decisión una visión más amplia de la que tienen hoy, mostrando no solo el estado actual sino también apuntando hacia la solución.

Conclusiones parciales

La aplicación de técnicas inteligentes sobre los sistemas de toma de decisión permite que estos sistemas puedan ser de gran importancia para el proceso de toma de decisión en la educación, no solo para los especialistas del área sino también para los no especialistas del área.

La utilización de los algoritmos conceptuales de datos incompletos y mezclados, son factibles para la presente investigación, de acuerdo con lo que se desea, no solo el agrupamiento de los estudiantes, sino también extraer patrones relacionados al rendimiento académico que caracterizan cada grupo. Lo que se propone podrá facilitar un análisis de los datos de una manera fácil y que sea comprensible.

Definir los elementos claves permitió desarrollar una solución que contenga aspectos fundamentales para que los tomadores de decisiones tengan una herramienta que sea capaz de apuntar hacia la solución.

Capitulo-II “Desarrollo de la propuesta”

En el presente capítulo se describe la solución obtenida, la descripción se basa en la implementación de una herramienta computacional para la extracción de patrones de los estudiantes de las universidades públicas de Angola a partir del rendimiento, la herramienta es basada en los algoritmos de conceptuales de agrupamiento conceptual compacto, para datos incompletos y mezclados. Se describe cada proceso del algoritmo, los requisitos de software identificados, el diseño del sistema y la tecnología utilizada.

2.1. Modelo computacional y/o algoritmo

El empleo de las técnicas de inteligencia artificial en los sistemas informáticos da lugar a los sistemas inteligentes, estos sistemas han servido como medio de solución de diversos problemas en muchas áreas, en el modelo propuesto se hace referencia al enfoque lógico combinatorio, el modelo presentado abajo consta de 5 etapas, a continuación se hace una descripción de cada etapa del modelo.

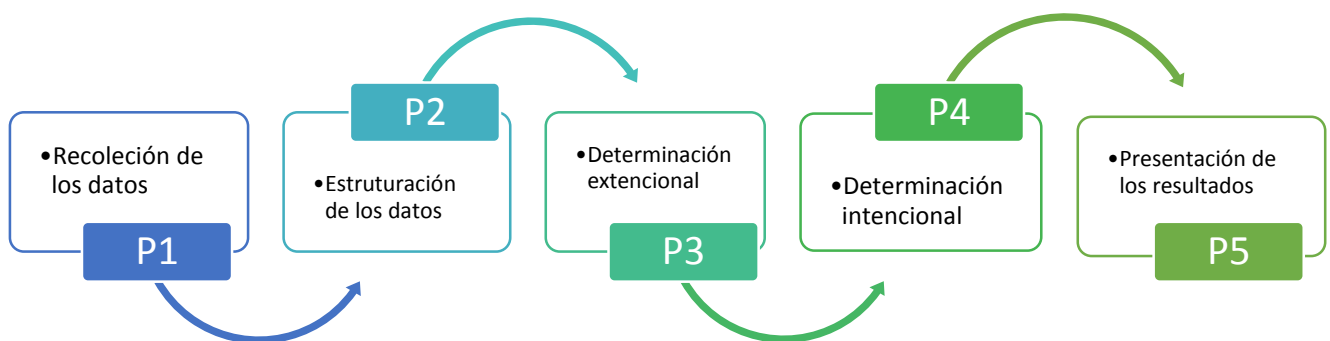


Ilustración 8 - Modelo Computacional

2.1.1. Recolección de los datos

Antes de hablar de la recolección de datos hace falta referenciar cómo se hizo la elección de los rasgos para el análisis, en este estudio se procedió de la siguiente forma, se han seleccionado varias bases de datos de diferentes instituciones superior públicas angolanas, entre ellas se identificaron los rasgos comunes referentes al rendimiento académico de los estudiantes, datos sociales y otros, de estos rasgos se identificó un conjunto compuesto por 14 rasgos. Este proceso se realizó debido al hecho de que la solución sirva a cualquier institución superior pública angolana.

La etapa de recolección es la etapa responsable del proceso de obtención de los datos de los estudiantes, en la figura que se muestra debajo se pueden ver los tipos de datos que son recogidos en las universidades.

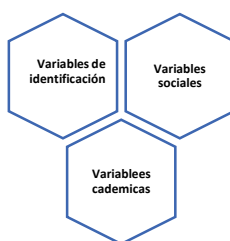


Ilustración 9 - Tipo de variables

Estos datos se pueden ver en la tabla que sigue, ellos son cargados a partir de un archivo de datos (XML, EXCEL, CSV).

N	Rasgos	Dominio	Observación
1	Habilitación literaria	Numérico	Identifica la habilitación literaria del estudiante
2	Escuela de precedencia	Texto	Nombre de la escuela que frecuentó antes de ingresar en la universidad
3	Carrera de precedencia	Texto	Carrera del estudio de formación básica
4	Año de conclusion	Numérico	Año de conclusión de la formación básica
5	Profesión	Texto	Profesión del estudiante, si es solo estudiante o trabajador
6	Edad	Numérico	Edad del estudiante
7	Dirección	Texto	Dirección de la residencia del estudiante
8	Carrera	Texto	Carrera que el estudiante está inscrito
9	N.asignaturas	Texto	Número de asignaturas inscritas en dicho año académico
10	N.asignatura_aprobadas	Numérico	Número de asignaturas aprobadas hasta el año académico actual
11	Tipos. Exámenes	Texto	Tipos de exámenes finales aprobados por estudiantes
12	Periodo	Texto	Período en que el estudiante frecuenta los estudios
13	Nota del examen de ingreso	Numérico	Clasificación del examen de ingreso a la universidad
13	Año de estudiante	Númerico	Nivel académico del estudiante
14	Sexo	Texto	Género del estudiante
15	Promedio	Númerico	Promedio del estudiante

Tabla 7 - Rasgos propuestos y sus respectivos dominios

2.1.2. Estructuración de los datos

La estructuración de los datos consiste en hallar un patrón de formato de datos general, es decir los datos recogidos pueden presentar formatos diferentes por ejemplo en el caso de describir el género de un estudiante puede ser que haya datos como masculino, m, femenino, f, en este caso es necesario encontrar una forma general de representar tales datos, en el caso del promedio y la nota de ingreso se representa en clases con un determinado rango,

en la edad se sigue el mismo criterio, esta etapa es una etapa crucial para el proceso, en esta etapa se encuentra el formato general para todos los datos.

2.1.3. Determinación extensional

Esta etapa es la primera etapa de los algoritmos conceptuales, que consiste por medio de una matriz inicial (m_i), compuesta por objetos (m) en término de sus rasgos (n), como se puede observar en la figura 1, hallar una matriz de aprendizaje, que es compuesta de objetos agrupados en diferentes clases.

		x_1, \dots, x_n
O_1		$x_1(O_1), \dots, x_n(O_1)$
.		.
.		.
O_m		$x_1(O_m), \dots, x_n(O_m)$

Ilustración 10 - Matriz inicial

La matriz inicial se presenta en término de 15 rasgos, desde rasgos cualitativos, cuantitativos y otros, el algoritmo presentado debajo describe la etapa extensional (agrupamiento de los objetos).

Algoritmo 1 Agrupamiento no supervisado

Entrada: Matriz inicial (M_i)

Salida: Matriz de aprendizaje (MA)

1. Constuir la matriz de semejanza aplicando una $\Gamma(O_i, O_j)$
2. Calcular el umbral de semejanza β_0 (seleccionar una de las variantes):

2.1.* seleccionar un valor de 0 hasta 1

2.2.*

$$\beta_0 = \left(\frac{2}{m(m-1)} \right) \sum_{i=1}^m \sum_{j=1}^m \Gamma(O_i, O_j)$$

2.2.*

$$\beta_0 = \left(\frac{1}{m} \right) \sum_{j=1}^m \max\{\Gamma(O_i, O_j)\}$$

3. Agrupar los objetos utilizando un criterio de agrupamiento
-

2.1.4. Determinación intencional

Esta etapa es la segunda etapa de los algoritmos conceptuales, consiste en a partir de una matriz de aprendizaje (MA), que es una matriz donde los objetos están agrupados en clases, como se presenta en la figura 2, de esta misma matriz se generan los conceptos.

En la generación de los conceptos, el primero paso es hallar el conjunto de testores típicos donde se procedió por medio del cálculo de medida de relevancia, después de hallar el conjunto de los testores típicos que se pueden considerar el conjunto de apoyo se generan los conceptos de acuerdo el algoritmo LC.

	x_1, \dots, x_n
O_1	$x_1(O_1), \dots, x_n(O_1)$
\vdots	\vdots
O_m	$x_1(O_m), \dots, x_n(O_m)$

Ilustración 11 - Matriz de Aprendizaje

Algoritmo 2.1 Selección de variables * cálculo de los testores

Entrada: Matriz aprendizaje (MA)

Salida: Matriz básica (MB)

1. Construir la matriz de diferencia (simétrica) aplicando una función de disimilaridad (seleccionar una de las variantes)

$$1.1^* \beta_0 = \begin{cases} 0 & \text{si } X_s(O_i) = X_s(O_j) \vee X_s(O_i) = ? \\ & \vee X_s(O_j) = ? \\ 1 & \text{en otro caso} \end{cases}$$

$$1.2^* \beta_0 = \begin{cases} 0 & \text{si } |X_k(O_i) - X_k(O_j)| \leq \varepsilon \\ 1 & \text{en otro caso} \end{cases}$$

2. Construir la matriz básica /* matriz compuesta por apenas filas básicas (no tienen superfilas)*/

En este algoritmo se ha seleccionado la primera función de disimilitud, se obtuvo la matriz básica compuesta por filas básicas de la matriz de diferencia, el algoritmo 3 que se muestra debajo, es responsable de encontrar los testores típicos según un criterio de medida de relevancia.

Algoritmo 2.2 Selección de variables * cálculo de los mejores testores

Entrada: Matriz básica (MB)

Salida: Testores típicos (TT)

1. Seleccionar el conjunto de testores
2. Calcular el peso de los rasgos presente en el conjunto de los testores

$$P(X_i) = \frac{|T(X_i)|}{|T|}$$

3. Calcular la longitud de los mismo testores

$$L(x_i) = \frac{\sum_{t \in |T(X_i)|} t^{-1}}{|T(X_i)|}$$

$|T|$ conjunto de todos los testores

$|T(X_i)|$ conjunto de todos testores que contiene (X_i)

4. Calcular la medida de relevancia

$$\varepsilon(x_i) = \alpha P(X_i) + \beta L(X_i)$$

se consideró α, β igual 0,5

5. Para cada testor calcular la suma de la medida de relevancia (magnitud)

$$\psi_i(t_i) = \sum_{n=1}^{|t_i|} \varepsilon(x_i)$$

6. Seleccionar el conjunto de testor de mayor magnitud
-

Algoritmo 2.3 generación de los conceptos LC (Martínez-Trinidad,1999)

Entrada: Testores típicos (TT)

Salida: conceptos generalizado

1. Para cada conjunto de apoyo (TT)

1.1. Calcular la estrella de cada clase $G_t = (k_i \setminus k_1, \dots, k_{i-1}, k_{i+1}, \dots, k_c)$ aplicando el operador RU

2. Aplicar el operador GEN a todos l-complejos obtenido y construir la generalización de los l-complejos $\alpha_1, \dots, \alpha_q$
-

En nuestro caso nos quedó más factible el uso del LC debido a la complejidad computacional presentada por el operador RUE de RGC, donde del el algoritmo LC se utilizó un agrupamiento compacto. Con estos grupos compactos, como es matemáticamente probado (Structuralization of universes, 2000), presentan mayor semejanza entre objetos de mismo grupo, así los conceptos generados son más factibles, es decir una vez que se genera los conceptos de cada grupo hay más probabilidad que cubra todos objetos debido al que los grupos generados contienen objetos que presentan el máximo índice de semejanza entre ellos.

2.1.5. Presentación de los resultados

Los conceptos generados por los algoritmos de agrupamiento conceptual, generalmente son muy extensos y difíciles de comprender para los tomadores de decisión, en esta etapa se traducen estos resultados hallados en un dominio más cercano al entendimiento de estos especialistas, desarrollándose un *dashboard*², que presenta estos resultados de una forma más perceptible.

2.2. Implementación del sistema

En respuesta al problema planteado en la investigación, en esta sección se presenta el diseño y la implementación del sistema.

Teniendo en cuenta la metodología de desarrollo, se ha seleccionado la metodología ágil XP (*extreme Programming*), siendo XP una metodología ágil, permite desarrollar softwares en poco tiempo y además estas metodologías generalmente son orientadas a pequeños proyectos aunque simplifican el control del proceso de desarrollo de software, ellas garantizan las prácticas de calidad del producto.

XP se basa en las relaciones cliente y equipo de desarrollo, la comunicación entre todos los que participan en el desarrollo es sumamente importante, esta metodología como todas las metodologías ágiles es muy flexible a los cambios de requisitos, en general son adecuadas para el desarrollo de softwares que están sujetos a constantes cambios de requisitos (Canós, José H.; Letelier, Patricio; Penadés, M^a Carmen ;).

² Los dashboard's son aplicaciones de business Intelligence (inteligencia de negocio) que le permiten a una organización visualizar la información más importante para monitorear, analizar y administrar el desempeño del negocio de manera más efectiva. Deben ser vistos más que como una pantalla con gráficas atractivas, como un sistema de información que ayuda a las organizaciones a optimizar su desempeño y alcanzar sus objetivos estratégicos.

A continuación se presenta el diseño del software y otros detalles sobre la implantación.

2.2.1 Entorno de desarrollo

En el desarrollo del software es de gran importancia la selección adecuada de las tecnologías y herramientas a utilizar según el entorno y el tipo de aplicación a desarrollar. A continuación se presenta una tabla que contiene la información de la tecnología seleccionada para desarrollar la herramienta web.

Entorno de desarrollo integrado (IDE)	Netbeans IDE 8.0
Lenguaje	PHP
Plugins	Kendo UI, PhpExcelReader, XMIRReader, db2PHP 1.7.0
Gestor de base de datos SGBD	MySql 5.6.16
Servidor web	Apache 2.4.9

Tabla 8 - Herramientas seleccionadas para el entorno de desarrollo

2.2.1 Modelo conceptual

Según (Marqués, 2011) el modelo conceptual presenta un esquema de una determinada información, independientemente de cualquier consideración física. El objetivo del modelo es:

1. La perspectiva que cada usuario tiene de los datos
2. Naturaleza de los datos, independientemente de su representación física
3. El uso de los datos a través de áreas funcionales.

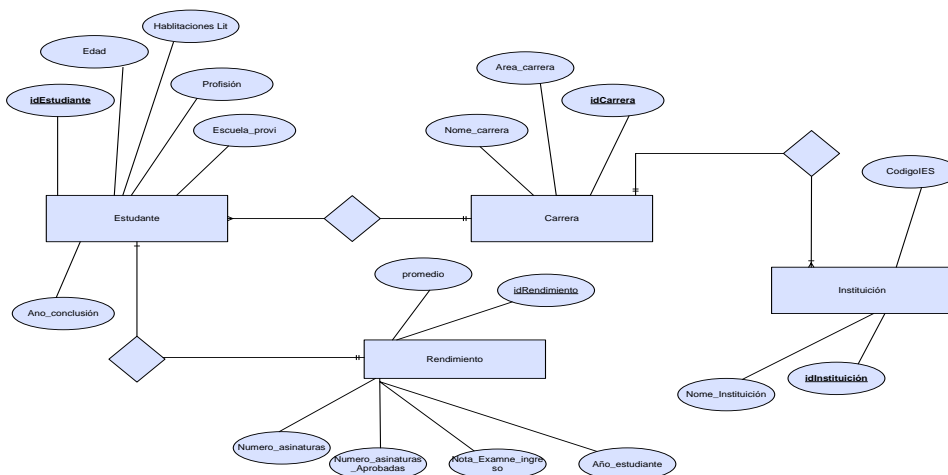


Ilustración 12-Modelo conceptual de la base de datos

2.2.2. Requisitos del sistema

El requisito consiste en la definición documentada de una propiedad o comportamiento expresado por un cliente que debe cumplir un sistema, ellos se pueden clasificar en requisitos funcionales y no funcionales.

Requisitos funcionales son tareas o actividades que deben ser realizadas por un determinado sistema sin ninguna restricción física, generalmente en el sistema ellos especifican las entradas y salidas.

Requisitos no funcionales suelen ser una propiedad del sistema, por ejemplo: tecnologías de desarrollo, requerimientos para instalación del sistema, mantenibilidad, estos requisitos sí consideran las restricciones físicas. A continuación se presenta un listado de los requisitos funcionales y no funcionales.

Requisitos funcionales	
1	Acceso al sistema
2	Gestionar datos del utilizador Insertar, modificar, consultar, eliminar datos de utilizador
3	Gestionar datos de los estudiantes Importar, consultar, eliminar datos de los estudiantes
4	Crear un proceso de análisis Preparar los datos Generar la matriz inicial Generar la matriz de aprendizaje Generar la matriz básica Generar los testores típicos Generar los conceptos
5	Presentar los resultados Exportar informes Ver informes

Tabla 9 - Requisitos Funcionales

Requisitos no funcionales	
1	Requisitos de interface externa La aplicación debe presentar un diseño de los campos, mensajes de error, texto de una forma legible
2	Requisitos de usabilidad El lenguaje empleado debe ser fácil para que sea entendible por los usuarios. Al nombrar los términos debe existir un estándar para que sea de fácil comprensión

Las interfaces deben prevenir de los errores con opciones bien definidas

3 Requisitos de hardware

La aplicación debe ser accedida desde sistemas operativos : Windows, Mac OSX, Linux

Requerimientos mínimos de hardware: 1 GB RAM, 9 GB de HDD, 500 MHz

4 Requisitos de software

El cliente debe tener un browser o navegador :

Internet explorer, Mozilla Firefox 32.0.2, Chrome 40.0.2214.91, safari

El cliente debe tener el plugin Adobe reader instalado

El servidor debe tener instalado MySQL 5.6.16, Apache 2.0 Handler

Tabla 10 - Requisitos no Funcionales

2.2.3. Diseño del sistema

En ingeniería de software el diseño de software es un proceso crucial en la etapa de desarrollo del software, en él se presenta un alto nivel de abstracción, un nivel en el que se pueda comprender directamente el objetivo del sistema específico y según unos requisitos más detallados de comportamiento funcionales y de datos a medida que ocurren las iteraciones de diseño (Pressman, 1997).

2.2.3.1. Diseño de la base de datos

Los sistemas de base de datos proporcionan un cierto nivel de abstracción de los datos, ellos ocultan características sobre almacenamiento físico que la mayor parte de los utilizadores no necesitan saber. Los modelos de datos son aquellos que ofrecen dicha abstracción a través de su jerarquía de niveles. Un modelo de datos es una descripción de la estructura de una base de datos (Marqués, 2011).

En (Pressman, 2010) se comenta que los modelos de datos son usados para describir la representación lógica y física de la información en el sistema, ellos se pueden crear a partir de la ingeniería inversa de un sistema de base de datos ya existente o también pueden ser creados a partir de un modelo de diseño.

A continuación se presenta el modelo de datos del sistema y sus descripciones presentadas en tablas.

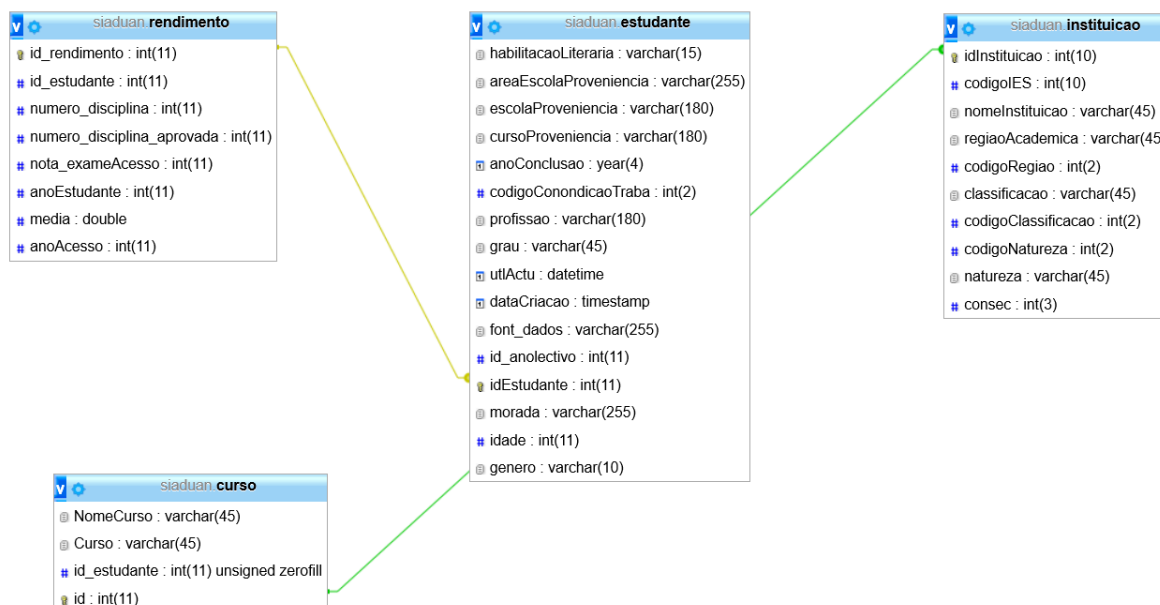


Ilustración 13 - Modelo de datos

Tabla	Descripciones
Estudiante	En esta tabla se almacenan datos personales de los estudiantes
Rendimiento	En la tabla rendimiento se almacenan datos sobre el rendimiento de los estudiantes
Institución	En esta tabla se almacenan datos de la institución a la que pertenece el estudiante
Carrera o curso	En esta tabla se almacenan datos de la carrera que el estudiante está cursando

Tabla 11 - Descripciones del modelo de datos

2.2.3.2. Diseño arquitectónico

En el diseño arquitectónico se define la relación entre los elementos más importantes del software, los patrones de diseño pueden usarse para cumplir con los requisitos definidos por el sistema (Pressman, 2010).

Los patrones definen la estructura general de un software, definen las relaciones entre el subsistema y los componentes del software y las reglas para especificar las relaciones entre elementos (Pressman, 2010).

2.2.3.2.1. Patrón MVC

El patrón MVC tuvo su inicio en los años ochenta con una primera implementación del popular lenguaje *Smalltalk*, este patrón ha sido uno de los más importantes e influyentes en

la historia de la programación, actualmente sigue siendo un patrón ampliamente utilizado. Este patrón se basa en separar el código responsable de la representación de los datos en pantalla, del código encargado de la ejecución de la lógica de negocio. Para ellos el patrón divide la capa de presentación en tres tipos de objetos básicos: modelo, vista y controlador. El flujo de comunicación que hay entre estos objetos es la ventaja que ofrece el MVC, como se puede observar en la figura de abajo (Llorente, César de las Torres; Castro, Unai Zorrilla; Nelson, Calvarro Javier; Ramos Barroso, Miguel Ángel;, 2010).

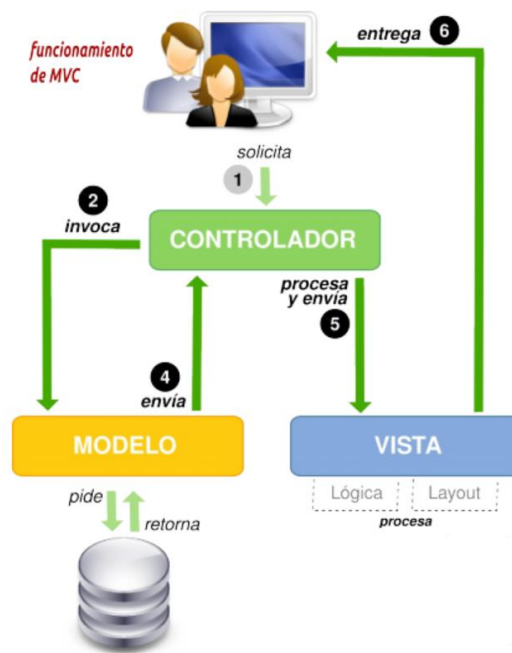


Ilustración 14 - arquitectura MVC

Generalmente la vista representa los datos del modelo y solicita acciones de un controlador en respuesta a las acciones del usuario, el controlador es responsable por recoger las acciones del usuario, interactúa con el modelo y devuelve una determinada vista en respuesta. La mayoría de las implementaciones del MVC, los tres objetos pueden relacionarse directamente el uno con el otro en otras implementaciones el controlador es responsable de determinar que vista mostrar (Llorente, César de las Torres; Castro, Unai Zorrilla; Nelson, Calvarro Javier; Ramos Barroso, Miguel Ángel;, 2010).

2.2.4. Interfaz gráfica

La interfaz de un software, es una aplicación donde el utilizador se comunica, está relacionada con la arquitectura y con la capa lógica de negocios. El diseño gráfico o la interfaz gráfica incluyen una serie de objetos que permiten al usuario hacer su trabajo de una manera fácil.

El software presenta un conjunto de colores cambiabile, para que el propio utilizador escoja el color que se adecue más a él, cuenta con una serie de menús distribuido de una forma para que le sea fácil al utilizador acceder a estos, además cuenta con una pantalla que se adecua al tamaño de muchos navegadores. A continuación se muestran unas imágenes del software, la autenticación del utilizador, importación de los datos y el proceso de cargado o iniciado.

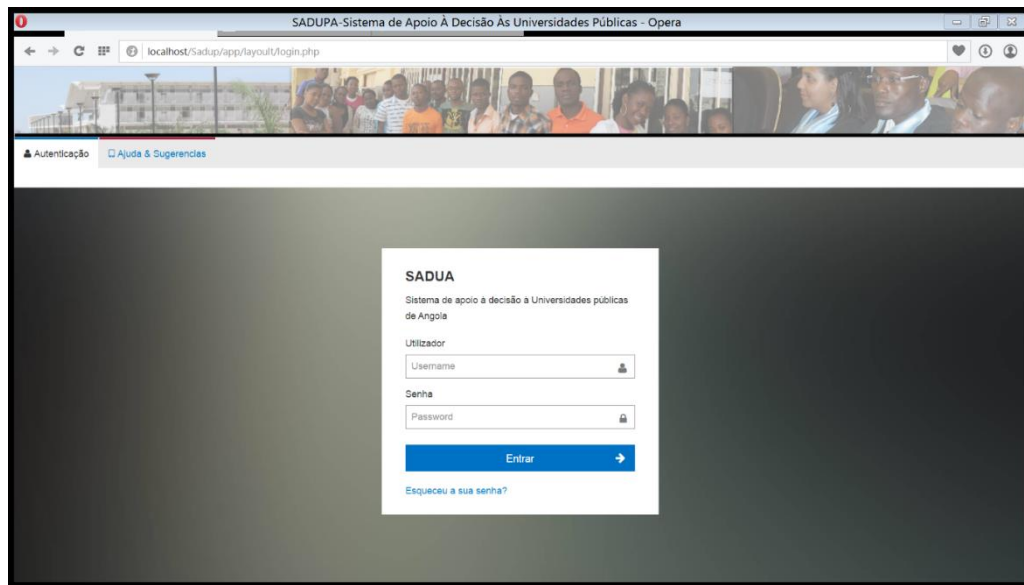


Ilustración 15 - Ventana de autenticação del utilizador

Esta es la ventana donde cada usuario con sus credenciales (utilizador y contraseña) se autentica y así logra acceder al ambiente de trabajo de la herramienta. A continuación en la **ilustración 16** e **ilustración 17** se muestra el ambiente de trabajo, como es hecha la importación de los datos y como es visualizado cada proceso de un determinado año lectivo.

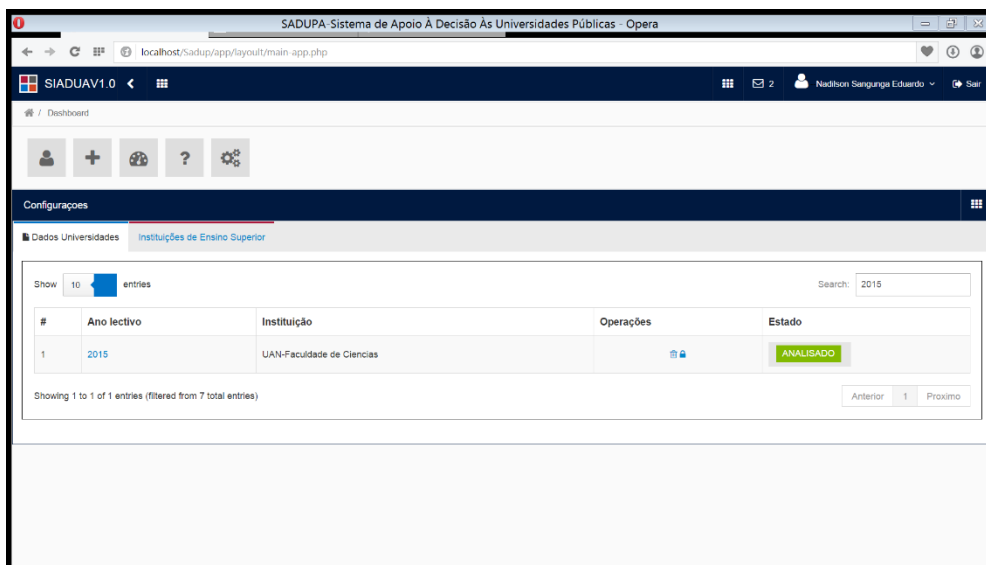


Ilustración 16 - Ventana principal del software

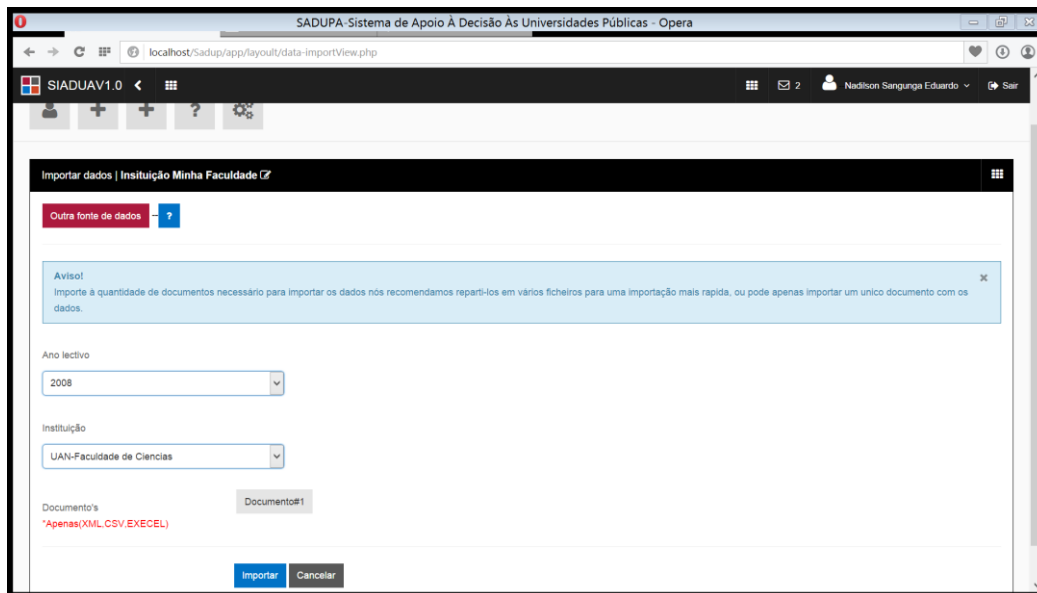


Ilustración 17 - Importación de los datos de estudiantes

Conclusiones parciales

El sistema descrito anteriormente tiene como centro, el descubrimiento de información a partir del rendimiento académico de los estudiantes universitarios, información esta que caracteriza los estudiantes, lo que constituye una herramienta para el apoyo al proceso de toma de decisiones. En este capítulo se describió:

- Desarrollo de un modelo y/o algoritmo basado en los algoritmos conceptuales con el enfoque de una herramienta para la extracción de patrones relacionado al rendimiento académico de los estudiantes en la universidades públicas de Angola;
- Modelo conceptual para el acercamiento al dominio del problema relacionado con la investigación;
- El diseño de la aplicación como principal entrada a la implementación de la propuesta.

III-Capítulo “Validación de la propuesta”

La validación de una investigación es de gran importancia, siendo que es una forma de garantizar que se cumpla con los requerimientos necesarios.

En este capítulo se lleva a cabo la validación de la solución propuesta, haciendo un estudio de los resultados obtenidos por el sistema después de haber sido probado con varios datos generados de la facultad de ciencias y Facultad de Letras, ambos pertenecientes a la universidad Agostinho Neto.

3.1. Validación de los resultados

En los algoritmos conceptuales se pueden obtener resultados diferentes, sin embargo es siempre una buena práctica la evaluación de los resultados, de manera tal que sea posible tener una idea del grado de eficacia de los resultados obtenidos.

Para Pal y Bezdek en 1995 (1) la validación de los algoritmos de agrupamiento depende mucho de la calidad del agrupamiento, lo que se relaciona con el criterio utilizado para agrupar.

Según (citar) para medir la calidad de los algoritmos de agrupamiento generalmente se hace por una medida o índice. Estos índices de validación se pueden dividir en dos grupos:

- ✓ Medidas internas: evalúan la propiedad del algoritmo como compacidad³ y separación de los objetos en los agrupamientos construidos basándose en la similaridad entre ellos.
- ✓ Medidas externas: evalúan el resultado del agrupamiento respecto a qué tanto se acerca un resultado ideal.

Se espera que estas medidas sean objetivas y que no tengan preferencia por el algoritmo utilizado en particular.

Los resultados preliminares de esta investigación se validaron utilizando una de las medidas externas más utilizadas actualmente, el índice **Rand Statistic**(R).

$$R = \frac{SS+SD}{M}$$

M: Es la cantidad de pares de objetos posible en datos.

SS: La cantidad de pares de objetos de la misma clase que pertenecen al mismo grupo.

³ El objeto de cada grupo debe ser lo más cercano posible a los objetos de sus respectivos grupos (4).

SD: La cantidad de pares de objetos que no pertenecen a la misma clase y que tampoco pertenecen al mismo grupo.

El objetivo es tener un índice (R) que esté en el intervalo [0, 1], los valores cercanos a 1 indican que el resultado del algoritmo de agrupamiento es correcto.

En las secciones a continuación se presentan varios casos de análisis de los datos de estudiantes, y la validación de los resultados utilizando el *Rand Statistic*.

Se generaron datos aleatorios de los estudiantes de la facultad de letras y facultad de ciencias de la universidad de Angola de años lectivos diferentes.

HL-Habilitações literárias	EP-Escola de província	C-Curso	AC-Ano conclusã	P-Profissã	I-idade	M-Morada
11	Puniv Central	LÃ-nguas e Literaturas Angolanas	2002	Regular	33	da Ingomboto
12	PIR	LÃ-nguas e Literaturas Angolanas	2002	Regular	29	da Maianga
12	IMIL	LÃ-nguas e Literaturas Angolanas	2002	Trabalhador	34	do Rangel
12	INEF	LÃ-nguas e Literaturas Angolanas	2002	Trabalhador	21	do Soyo
12	Escola do Ensino SecundÃrio do II Ciclo Kapolo II	LÃ-nguas e Literaturas Angolanas	2002	Trabalhador	33	da Gabela
11	Puniv do Cazenga	LÃ-nguas e Literaturas Angolanas	2002	Trabalhador	27	de Calandula
11	PuniÃ-v 11 de Novembro do Moxico	LÃ-nguas e Literaturas Angolanas	2002	Trabalhador	30	de Viana
11	PIR	LÃ-nguas e Literaturas Angolanas	2003	Regular	22	do Bailundo
12	INE 22 de Novembro	LÃ-nguas e Literaturas Angolanas	2003	Regular	28	do Cazenga
12	MagistÃrio PrimÃrio de Luanda	LÃ-nguas e Literaturas Angolanas	2003	Regular	27	do Cazengo
11	Puniv 21 de Janeiro - FAN	LÃ-nguas e Literaturas Angolanas	2003	Regular	26	do Cubal
11	Puniv do Cazenga	LÃ-nguas e Literaturas Angolanas	2003	Regular	32	do Kilamba - I
12	INE - Garcia Neto	LÃ-nguas e Literaturas Angolanas	2003	Regular	26	do Luena
12	SeminÃrio PropedÃutico Aquidiocesano do Huambo	LÃ-nguas e Literaturas Angolanas	2003	Regular	22	do Menongue
12	Papa JoÃo Paulo II	LÃ-nguas e Literaturas Angolanas	2003	Regular	24	do Seles
12	Escola de FormaÃÃo de Professores de CacuanÃ	LÃ-nguas e Literaturas Angolanas	2003	Regular	29	da Ingomboto
12	Instituto Normal de EducaÃÃo AntÃnio Jacinto	LÃ-nguas e Literaturas Angolanas	2003	Regular	24	da Maianga
12	Complexo Escolar 28 de Agosto	LÃ-nguas e Literaturas Angolanas	2003	Regular	30	da Maianga
12	Papa JoÃo Paulo II	LÃ-nguas e Literaturas Angolanas	2003	Regular	35	de Viana
12	Puniv 21 de Janeiro - FAN	LÃ-nguas e Literaturas Angolanas	2003	Regular	29	do Bolongongo
12	Complexo Escolar 28 de Agosto	LÃ-nguas e Literaturas Angolanas	2003	Regular	27	do Huambo
12	Aida Lara	LÃ-nguas e Literaturas Angolanas	2003	Regular	25	do Kilamba - I
12	Escola de Professores do Futuro	LÃ-nguas e Literaturas Angolanas	2003	Trabalhador	35	do Libolo
11	Puniv do Cazenga	LÃ-nguas e Literaturas Angolanas	2003	Trabalhador	7	do Lobito

Ilustración 18 - Datos de entrada de los estudiantes

Esta es la plantilla diseñada de acuerdo a los requerimientos del sistema para la entrada de los datos. Después de importar los datos de la planilla al Sistema, el sistema de acuerdo al utilizador sigue con las otras operaciones hasta terminar el proceso.

Facultad de Ciencias	Facultad de Letras
Ciencias de la computaci3n	Lengua y literatura Angola
Matemática	Lengua y literatura Inglesa
Física	Lengua y literatura Portuguesa
Química	Lengua y literatura Francesa
Biología	Secretariado Administrativo
Geofísica	Filosofía
Geología	

Tabla 12 - Facultades y carreras de los estudiantes

Caso A “Estudiantes del año lectivo 2007 “

En este caso, se presentaron un total de 467 estudiantes de las carreras de la facultad de letras de la universidad Agostinho Neto.

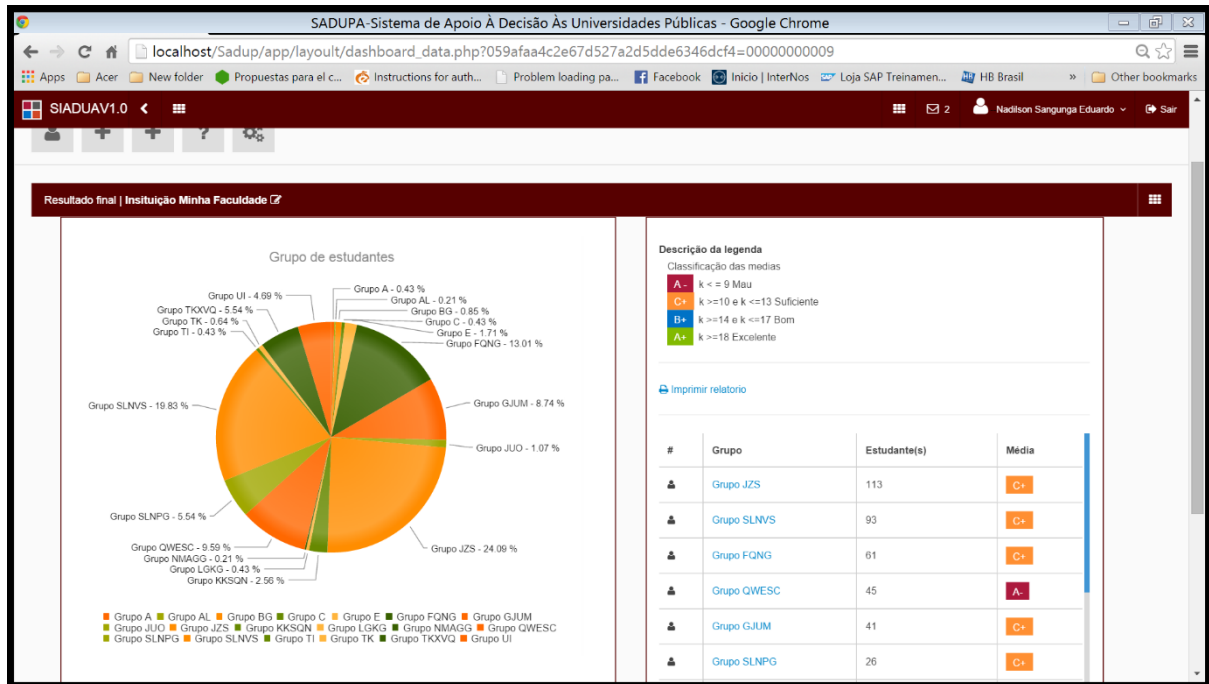


Ilustración 19 - Reporte principal del caso A

En esta ilustración 3, se presenta un *dashboard* que contiene toda la información después del análisis hecho por el sistema, el *dashboard* presenta la información de manera que sea fácil de interpretar por los no especialistas. A continuación se presenta una tabla con las informaciones del estudio detalladas.

Grupo	Estudiante(S)	Rendimiento
JZS	113	C+
SLNVS	93	C+
FQNG	61	C+
QWESC	45	A-
GJUM	41	C+
TKXVQ	26	A-
SLNPG	26	C+
UI	22	C+
KKSQN	12	C+
E	8	B+
JUO	5	C+
BG	4	B+

TK	3	C+
C	2	B+
LGKG	2	C+
TI	2	C+
A	2	B+
AL	1	B+
NMAGG	1	C+
Total de grupos	19	

Tabla 13 - Tabla de resultados del caso A

El promedio presenta las siguientes descripciones:





Leyenda	Descripción
	K <= 9 Malo
	K >= 10 y k <= 13 Suficiente
	K >= 14 y k <= 17 Equilibrado
	K >= 18 y k <= 20 Excelente

Tabla 14 - Leyenda del rendimiento

De los 19 grupos clasificados, cada grupo presenta sus características. El grupo JZS con 113 estudiantes, presenta las siguientes características.

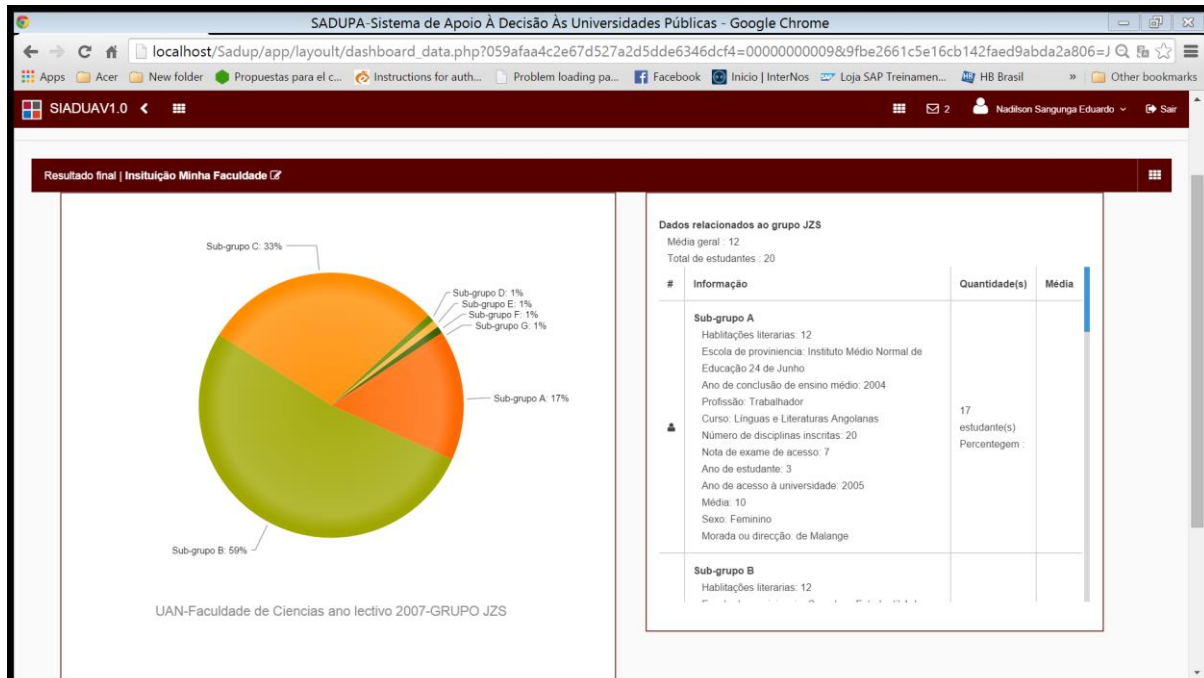


Ilustración 20 - Reporte relacionado a un grupo del caso A

El grupo QWESC tiene 45 estudiantes con un promedio de A-. Lo que significa que estos estudiantes tienen un promedio de 9, considerado malo. Estos estudiantes presentan las siguientes características.

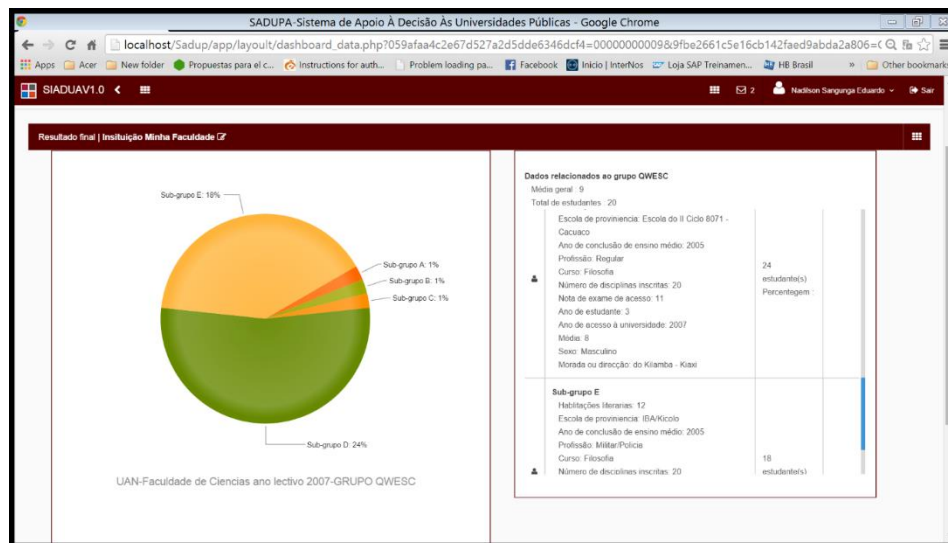


Ilustración 21 - Reporte de grupo relacionado al caso A

En el dashboard aún se puede verificar una segunda clasificación que se le hace solo por el promedio, es decir además de tener el promedio del grupo aún se tiene el porcentaje de los estudiantes por cada promedio.

Caso B Estudiantes del año 2008

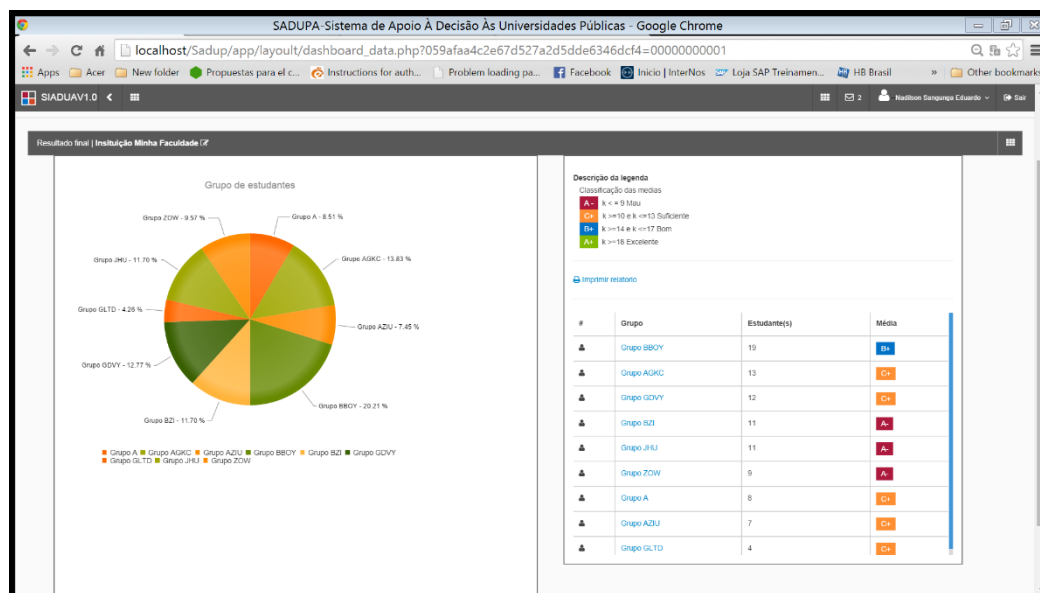


Ilustración 22 - Reporte principal del caso B

En este caso de estudio, se presentó un total de 94 estudiantes de las carreras de la facultad de ciencias de la Universidad Agotinho Neto.

Este análisis presenta una cantidad reducida de estudiantes, como se puede verificar en el *dashboard*, en la ilustración 23 se pueden verificar las características del grupo con mejor promedio y mayor cantidad de estudiantes.

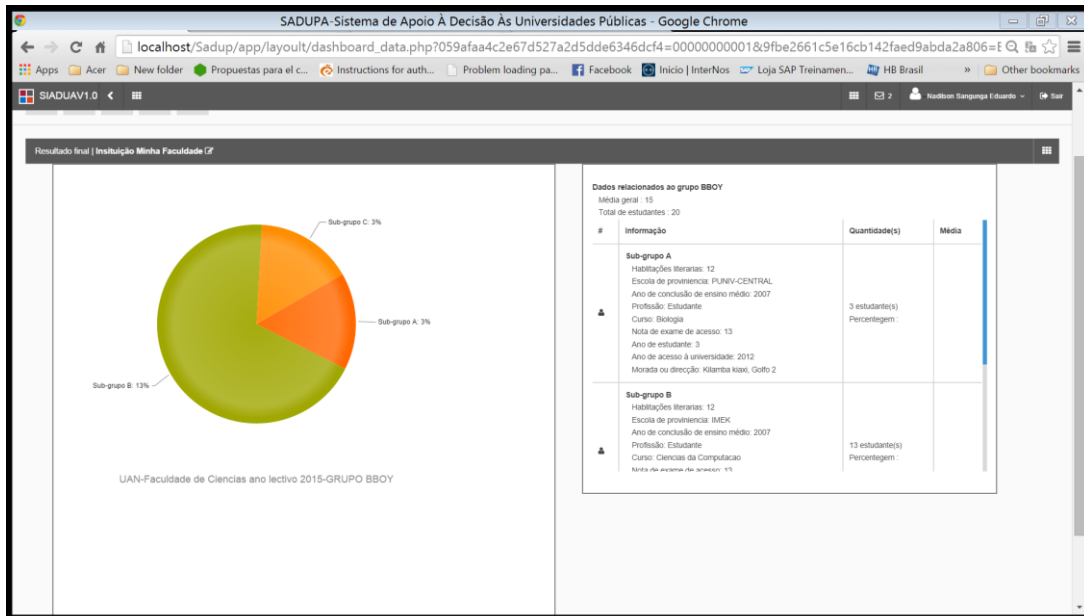


Ilustración 23 - Reporte relacionado a un grupo del caso B

Caso C estudiantes del año 2011

En este caso se analizaron 491 estudiantes de las carreras de la facultad de letras de la universidad Agostinho Neto, donde se han identificado 22 grupos y sus respectivas características y promedios por cada año. A continuación se pueden observar los resultados del análisis en las ilustraciones.

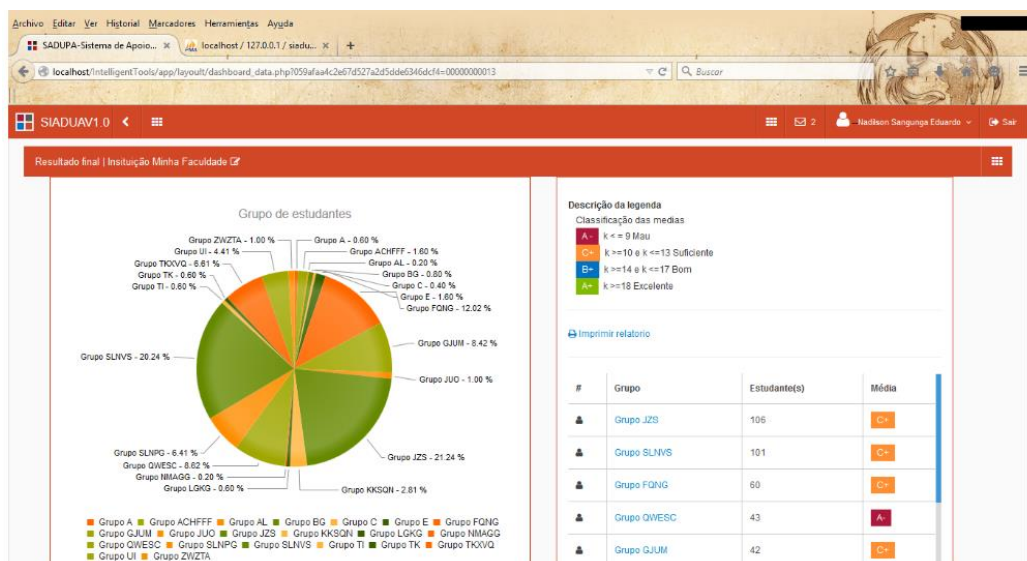


Ilustración 24 - Reporte principal del caso C

Nuevamente se presenta el *dashboard* principal del análisis, con informaciones relacionadas a cada grupo y sus determinados promedios.

Grupo	Estudiantes	Rendimiento
JZS	106	C+
SLNVS	101	C+
FQNG	60	C+
QWESC	43	A-
GJUM	42	C+
TKXVQ	33	A-
SLNPG	32	C+
UI	22	C+
KKSQN	14	C+
ACHFFF	8	C+
E	8	B+
JUO	5	C+
ZWZTA	4	A-
BG	3	B+
TK	3	C+
TI	3	C+
A	2	B+
C	1	B+
LGKG	1	C+

Tabla 15 - Resultados del caso C

3.1.1. Evaluación de los resultados “*Rand Statistic*”

Después de haber presentado los resultados surge la necesidad de evaluar los resultados. A continuación se presenta una tabla con los resultados de la evaluación de cada caso de estudio.

Caso	Cantidad	Índice(RI)
A	467 estudiantes	0,995
B	94 estudiantes	1
C	499 estudiantes	0,999

Tabla 16 -Evaluación de los caso de estudio

Como se aprecia en la tabla 16, el resultado producido por el algoritmo es eficaz en los tres casos de estudios según el RI. Estos resultados muestran que el sistema presenta resultados aceptables.

3.2. Medidas de validación basada en la teoría de los conjuntos aproximados

En esta sección se muestra el uso de la teoría de los conjuntos aproximados para validar los conceptos generados por el algoritmo conceptual. De acuerdo la idea de

(González, Sánchez, & Lorenzo, 2012) la teoría de los conjuntos aproximados no compite con la teoría de los agrupamiento conceptual, sino que se completan.

La teoría de los conjuntos aproximados (*Rough Sets Theory*) fue introducido por Z. Pawlak em 1982 como un modelo matemático para tratar problemas de incertidumbre (Roberto, 2006) . Se basa en aproximar cualquier concepto, un subconjunto duro al dominio, un ejemplo, es una clase en un problema de clasificación supervisada por un par de conjuntos exactos llamados aproximación inferior y aproximación superior del concepto. Con la RST (*Rough Sets Theory*) es posible tratar los datos mezclados (cualitativos, cuantitativo), y no se requiere eliminar las inconsistencias previas al análisis, respecto a la información de salida puede ser utilizada para encontrar la relevancia de los atributos (Parsons, 1996).

Según (Skowron, 1999) un sistema de información, es un conjuntos de atributos $A = \{a_1, a_2, \dots, a_n\}$ y un conjunto U no vacío llamado universo de ejemplo (objetos, entidades, situaciones o estado) descrito usando los atributos a ; al para (U, A) se le denomina sistema de información. Si a cada elemento de U se le agrega un nuevo atributo d llamado decisión, indicado la decisión tomada en este estado o situación entonces se obtiene un sistema de decisión $(U \cup \{d\})$, donde $d \notin A$.

Definición inseparabilidad: A cada subconjunto de atributos B de A $B \subseteq A$ está asociada una relación binaria de inseparabilidad denotada por R , la cual es el conjunto de pares de objetos que son inseparables uno de otros por esa relación (Skowron, 1999).

Una relación de inseparabilidad que sea definida a partir de formar subconjuntos de elementos de U que tienen igual valor para un subconjunto de atributos B de A , $B \subseteq A$, es una relación de equivalencia. Los conceptos de un subconjunto $X \subseteq U$. Estos conceptos fueron originalmente introducido con referencia a una relación de inseparabilidad R sea R una relación binaria definida sobre U la cual representa la inseparabilidad, se dice que $R(X)$ significa $R(X) = \{y \in U : y R x\}$. En la RST clásica, R es definida como $R \subseteq U \times U$ que es reflexiva, simétrica, transitiva.

3.2.1. Medidas de inferencia clásicas de la teoría de los conjuntos aproximados

La teoría de los conjuntos aproximados ofrece algunas medidas para analizar los sistemas de información.

Definición Precisión de aproximación (Caballeiro, Bello, Arco, & Cárdenas, 2010) : un conjunto aproximado X puede ser caracterizado numéricamente por el coeficiente llamado precisión de la aproximación $|x|$ denotada por la cardinalidad x , $x \neq \emptyset$.

$$\alpha(x) = \frac{|R'(x)|}{|R'^*(x)|}$$

Obviamente $0 \leq \alpha(x) \leq 1$. Si $\alpha(x) = 1$, x es duro (exacto), si $\alpha(x) < 1$, x es aproximado (vago, inexacto), siempre respecto al conjunto de atributos considerados.

Definición calidad de aproximación (Caballeiro, Bello, Arco, & Cárdenas, 2010) : se representa por la siguiente $\gamma(x) = \frac{|R'^*(x)|}{|x|}$

Expresa la proporción de objetos que pueden ser correctamente clasificados en las clases X. Además $0 \leq \alpha(x) \leq \gamma(x) \leq 1$, y $\gamma(x) = 0$ si y solo si $\alpha(x) = 0$, mientras $\gamma(x) = 1$ si y solo si $\alpha(x) = 1$. Considerando que X_1, \dots, X_l son las clases del sistema de decisión se define la medida.

Definición calidad de clasificación (Caballeiro, Bello, Arco, & Cárdenas, 2010) : Este coeficiente describe la inexactitud de clasificaciones aproximadas:

$$\Gamma = \frac{\sum_{i=1}^l |R'(x_i)|}{|U|}$$

La medida calidad de la clasificación expresa la proporción de objetos que puedan estar correctamente clasificados en el sistema. Si ese coeficiente es igual a 1, entonces el sistema de decisión es consistente, en otro caso es inconsistente.

3.2.2. Resultados experimentales

Estos resultados están de acuerdo a la segunda parte del modelo computacional, que es la parte intencional, es decir, es la parte responsable por la generación de los conceptos. Con el mismo estudio de caso de las secciones anteriores, se calcularon las tres medidas clásicas de la teoría de cálculo aproximado para validar los resultados obtenidos de la etapa intencional, los cuales se pueden observar en las tablas que se muestran a continuación.

Caso A					
Grupos	Cantidad en grupos	Conjunto aproximación inferior	$\alpha(x) = \frac{ R'(x) }{ R'^*(x) }$	$\gamma(x) = \frac{ R'^*(x) }{ x }$	$\Gamma = \frac{\sum_{i=1}^l R'(x_i) }{ U }$
1	113	110	0,98		
2	93	88	0,95		
3	61	57	0,93		
4	45	42	0,93		
5	41	36	0,88		

6	26	20	0,77
7	26	9	0,35
8	22	18	0,81
9	12	9	0,75
10	8	5	0,63
11	5	4	0,8
12	4	4	1
12	3	3	1
14	2	2	1
15	2	2	1
16	2	2	1
17	1	1	1
18	1	1	1
	467	413	0,88

Tabla 17 - Resultado experimentales I

Caso B					
Grupos	Cantidad en grupos	Conjunto aproximación inferior	$\alpha(x) = \frac{ R'(x) }{ R''(x) }$	$\gamma(x) = \frac{ R' * (x) }{ x }$	$\Gamma = \frac{\sum_{i=1}^i R'(x_i) }{ U }$
1	19	16	0,84		
2	13	11	0,69		
3	12	9	0,58		
4	11	8	0,73		
5	11	7	0,64		
6	9	6	0,67		
7	8	5	0,63		
8	7	4	0,58		
9	4	4	0,5		
	94	70			0,74

Tabla 18 - Resultados experimentales II

Caso C					
Grupos	Cantidad en grupos	Conjunto aproximación inferior	$\alpha(x) = \frac{ R'(x) }{ R''(x) }$	$\gamma(x) = \frac{ R' * (x) }{ x }$	$\Gamma = \frac{\sum_{i=1}^i R'(x_i) }{ U }$
1	106	103	0,97		
2	101	97	0,96		
3	60	58	0,96		
4	43	40	0,93		
5	42	38	0,90		
6	33	30	0,90		
7	32	31	0,96		
8	22	20	0,90		

9	14	10	0,71
10	8	7	0,87
11	8	6	0,75
12	5	4	0,8
12	4	4	1
14	3	2	0,66
15	3	2	0,66
16	3	3	1
17	2	2	1
18	1	2	1
19	1	1	
	491	459	0,93

Tabla 19 - Resultados experimentales III

Los resultados presentados de acuerdo la validación utilizando la teoría de los conjuntos aproximados, presentan un promedio de 85% lo que valida los resultados obtenidos por la aplicación. Es necesario destacar que esta validación se hizo de acuerdo los conceptos que caracterizan cada elemento del agrupamiento.

Conclusiones parciales

El presente capítulo, tienen como objetivo probar la aplicación con datos reales, además de procesar la información para el análisis del rendimiento académico de los estudiantes en las universidades públicas de Angola.

El *dashboard* presentado para visualizar la información que se obtiene del proceso de análisis a partir de los datos generados, proveen al tomador de decisiones medios necesarios para caracterizar, desde diferentes puntos de vista, el rendimiento de los estudiantes. Así como las variables que están asociadas al rendimiento.

Los resultados arrojados en la aplicación de la técnica de evaluación por medidas externas, utilizando el índice *RI* y por la teoría de los conjuntos aproximados, fueron satisfactorios; lo que demuestra la confiabilidad de la solución propuesta.

Conclusiones y recomendaciones

Los resultados obtenidos en esta investigación permiten concluir que:

1. Se tiene un estudio del estado del arte de las principales herramientas informáticas utilizadas en problemas similares al objetivo de esta investigación.

2. Un modelo informático que integre el cálculo de los testores y la idea básica de los algoritmos conceptuales favorece la caracterización o selección de patrones para el posterior análisis y estudio del comportamiento académico de los estudiantes.
3. La implementación computacional de la herramienta SADUP sirve de apoyo para la toma de decisiones en el análisis del rendimiento académico.
4. La validación experimental realizada permite recomendar la utilización de la herramienta computacional SADUP con estos fines en cualquier institución pública de Angola.

Derivadas del estudio realizado, así como de las conclusiones de este trabajo se recomienda:

1. Implementar la opción del cálculo de los conceptos utilizando la variante de los testores típicos por clases.
2. Implementar el algoritmo de agrupamiento conceptual RGC y realizar un estudio comparativo de los resultados.
3. Realizar la validación del modelo con datos reales y bases de grandes volúmenes de datos.

Referencias bibliográficas

1. *Structuralization of universes*. **Martínez-Trinidad, José F., Shulcloper, José Ruiz y Cortés, Manuel S. Lazo**. 2000, ReserachGate.
2. *Algoritmo conceptual para el agrupamiento conceptual de objetos*. **Suárez, Airel Pérez; Pagola, José E. Medina**; ISSN 2072-6260, s.l. : CENATV, Abril de 2014, Serie azul-reporte técnico de minería de datos.
3. **Ruiz-Schulcloper, José , Guzmán Arenas, Adolfo y Martínez, J.Francisco**. *Enfoque lógico al Reconocimiento de patrones*. Mexico : IPN, 1999.
4. *Journal of Intelligence Systems*. **Haldaki, Maria, Batistakis, Yannis y Vazirgiannis, Michalis**. Netherlands : Kluwer Academic Publisher, 2001.
5. *Algoritmo de agrupamiento conceptuales: un estado de arte*. **Guerra-Gandón, Alejandro; Vegas-Pons, Sandro; Ruiz-chulcloper, José**. La Habana : s.n., Junio de 2012, Reporte Técnico de Reconocimiento de Patrones-Serie Azul.
6. **Canós, José H.; Letelier , Patricio; Penadés, M^a Carmen ;.** *Métodologías Ágiles en el Desarrollo de Software*. Valencia : DSIC -Universidad Politécnica de Valencia.
7. **Llorente, César de las Torres; Castro, Unai Zorrilla; Nelson, Calvarro Javier; Ramos Barroso, Miguel Ángel;** *Microsoft-architecture - Guia de arquitectura N-capas orientada al dominio con.NET 4.0 1 Edición, Microsoft Ibérica S.R.L*. España : Krasis Press, 2010. ISBN:978-84-936696-3-8.
8. *Un concepto de testor para cualquier función de analogía con imagen en un conjunto totalmente ordenado*. **La Habna : Revista INTEGRACIÓN Universidad Industrial de Santander Escuela de Matemáticas, 1996, Vols. Vol. 14, No 1, p. 75-83.**
9. *Estudio sobre algoritmos de agrupamiento jerárquico con funciones de semejanza no simétricas*. **Espinosa, Dustin L y Ruiz-Shulcloper, José**. La Habana : s.n., 2010, Vol. RNPS No.2142. ISSN 2072-6287.
10. **Franco-Arcega, Anilu, Sánchez-Díaz, Gullermo y Ruiz-Schulcloper, José** . *ADT : A decision tree algorithm based on concepts. center of Technologies Research on of Applications of Havana Cuba*. INTERNATIONAL SYMPOSIUM ON ROBOTICS AND AUTOMATIONS, 2006.
11. **Kniberg, Henrik**. *SCRUM Y XP DESDE LAS TRINCHERAS Cómo hacemos Scrum*. Estados Unidos de América : C4Media, 2007. ISBN: 978-1-4303-2264-1.

12. Marqués, Mercedes. *Bases de datos*. España : Publicacions de la Universitat Jaume I. Servei de Comunicació i Publicacions Campus del Riu Sec. Edifici Rectorat i Serveis Centrals. 12071 Castelló de la Plana, 2011. ISBN: 978-84-693-0146-3.
13. Martínez-Trinidad, José Francisco. *RESUMEN DE TESIS DOCTORAL-Herramienta para la Estructuración Conceptual de Espacio*. Mexico : Centro de investigación y computación, 2001. Vol. Vol. 4 No.3 . ISSN 1405-5546.
14. *Knowledge acquisition through conceptual clustering : A theoretical framework and an algorithm for partitioning data into conjunctive concepts*. Michalski, R.S. Report No. UICDS-R-80-1026, California, USA : International Journal of Policy Analysis and Information System, 1980, Vol. Vol 4 No.3 1980.
15. *Automated construction of classifications conceptual clustering versus numerical taxonomy*. Michalski, Ryszard S y Steep, Robert. s.l. : IEEE Trans. On Pattern Analysis and Machine Learning , 1983, Vols. Volume PAMI-5, No.4.
16. Michalski, Ryszard S y Mitchell, T. *Machine Learning : An Artificial Intelligence Approach*. California : TIOGA publishing, 1983.
17. Alvarez, Leonardo , Pico Peña, Ramón y Cotilla, Mario. Clasificación no supervisada por métodos lógico-combinatorio problema de zonación sísmica. 1995.
18. Bahit, Eugenia . *POO y MVC en PHP*.
19. Banda, Hugo A. *Inteligencia Artificial principios y aplicaciones*. Quito : s.n., 2011.
20. Estayno, Marcelo, Dapozo, Gladys y Cuenca Pletch Liliana, Greiner, Cristina. *MODELOS Y MÉTRICAS PARA EVALUAR CALIDAD DE SOFTWARE* . Argentina : Departamento de Informática. Facultad de Ingeniería. Universidad Nacional de Lomas de Zamora .
21. Fu, K. -S. *Syntatic Pattern Recognition and Neural Net works*. s.l. : Cambrige University Press, 1982.
22. Gonzales, R.C y Thomason, M.G. *Syntatic Pettern Recognition : An introduction*. s.l. : Prentice-Hall, 1978.
23. *Reconocimiento de Patrones- Algoritmo de agrupamiento conceptuales : un estado del arte*. Guerra-Gandón, Alejandro , Vegas-Pons, Sandro y Ruiz-shulcloper, José. 2012, Centro de aplicaciones tecnológicas avanzada- CENATEV.
24. Haag, y otros. *Management Information Systems: For The Information Age*. s.l. : McGraw-Hill Ryerson Limited: 136-140. ISBN 0-07-281947-2, 2000.

25. *Neues anwenderfreundliches Konzept der Entscheidungsunterstützung. Gutes Entscheiden in Wirtschaft, Politik und Gesellschaft in Wirtschaft, Politik und Gesellschaft. Zurich, vdf . Haettenschwiler, P. 1999, Hochschulverlag AG: 189-208.*
26. *On Validation Techniques.* Haldaki, Maria, Batistakis, Yannis y Vazirgiannis, Michalis. [ed.] Kluwer Academic Publisher. Netherlands : s.n., 2001, Journal of Intelligent Systems, págs. 107-43.
27. Krishnakumar, K. *Intelligent systems for aerospace engineering – an overview.* NASA Technical Report, Document ID: : s.n., 2003.
28. Martínez, Irene Olaya , Martínez-Trinidad, José Francisco y Carrasco-Ochoa, Jesús Ariel . *RESUMEN DE TESIS DOCTORAL-Algoritmos conceptuales basado en semillas.* Mexico : s.n., 2007. ISSN 1405-5546.
29. MARTINEZ-TRINIDAD, JOSÉ FRANCISCO , SANCHEZ, GUILLERMO y RUGERIO, BONIFACIO. Genetic Algorithm to Compute Fuzzy FS-Testors.
30. McCall, J.A. *Factors in Software Quality Vols I, II, III.*
31. Menzies. T.. Applications of Abduction #1: Intelligent Decision Support Systems. *Intelligent Decision Support Systems.* s.l., Monash University : In Proc. Of the Melbourne Workshop on Intelligent Decision Support, 1996.
32. *Concept-based clustering versus Numerical Taxonomy in IEEE Transaction on Pattern Analysis and Machine Intelligence.* Michalski, R. S. 1981, IEEE.
33. *Learning from Observation: Conceptual Clustering, in Machine Learning: An artificial intelligence approach 1983. p. 331-363.* Michalski, R.S. y R.E. , Stepp.
34. Miguel Fernando , González Pinzón y Sanabria», Juan Sebastián González . *Aplicación del estándar ISO/IEC 9126-3 en el modelo de datos conceptual entidad-relación.* Colombia : Revista Facultad de Ingeniería, UPTC, Julio - Diciembre de 2013, Vol 22, No. 35.pp. 113-125 , 2013.
35. *Visual Cluster Validate for phototype generator cluster models.* Pal, N.R y Hathaway , R.J. 1995, Pattern Recognition letters.
36. Palacio , Juan . *Flexibilidad con Scrum principios y diseños e implementación de campos Scrum.* s.l. : SafeCreative, 2008.
37. Pons Porrata , Aurora . *TESIS DOCTORAL -DESARROLLO DE ALGORITMOS PARA LA ESTRUCTURACIÓN DINÁMICA DE INFORMACIÓN Y SU APLICACIÓN A LA DETECCIÓN DE SUCESOS.* Castellón : s.n., 2004.

38. Porrata, Aurora Pons. TESIS DOCTORAL -DESARROLLO DE ALGORITMOS PARA LA ESTRUCTURACIÓN DINÁMICA DE INFORMACIÓN Y SU APLICACIÓN A LA DETECCIÓN DE SUCESOS. Castellón : s.n., 2004.
39. Pressman, Roger S. *Ingeniería del Software un enfoque práctico 5 Edición*. s.l. : McGraw Hill , 1997.
40. —. *Software Engineering : A practitioner's Approach 7th Edition*. s.l. : MacGraw-Hill, 2010.
41. Ripley, B.D. *Pattern Recognition and Neural Net Work*. s.l. : Cambridge University Press, 1982.
42. Ruiz-Chulcloper, José , Carrasco-Ochoa, Ariel Jesús y Martínez-Trinidad , José Francisco. Reconocimiento de patrones conceptos y metodología. SEREI AZUL, 2013.
43. Ruiz-Shulcloper , José . *Logical Combinatorial Pattern Recognition* . Tennessee, Knoxville, USA : s.n.
44. Russel, Stuart y Norvig, Peter. *Artificial intelligence A Modern Approach*. New jersey : Prentice Hall, 1995.
45. *Modelos baseado em agrupamento conceptual para apoio á decisão no processo de formação de estudantes em universidades públicas de Angola*. Sangunga Eduardo, Nadilson Camilo. Habana : UCIENCIA 2014, 2014. ISBN 978-959-286-026-1.
46. *Sistema inteligente para apoio a decisão apartir do rendimento academicos dos estudantes de licenciatura das universidades publicas de Angola*. Sangunga Eduardo, Nadilson Camilo. 2014, Congresso TUS 2014, Republica dominicana.
47. Sprague, R. H. y E. D. Carlson. *Building effective decision support systems*. Englewood Cliffs : N.J., Prentice-Hall ISBN 0-13-086215-0, 1982.
48. Suárez , Airel Perez y Pagola, José E. Medina. Algoritmos para agrupamiento conceptual de Objetos. 2014, Vol. SERIES GRIS.
49. Tou , J.T y Gonzales, R.C. *Pattern Recognition principles*. s.l. : Addison-Wesley, 1981.
50. *Structuralization of universes*. Trinidad, José Fco. Martínez. 2000, ResearchGate.
51. Villuendas Rey, Yenny , Franco Acerga, Anilú y Ruiz Schuloper, José. Métodos de construcción de árboles de clasificación. *SERIE AZUL*. 2008, Vol. RNPS No.2142, ISSN 2072-6287.

52. *The CLUSTER3 system for goal-oriented conceptual clustering: method and preliminary results.* W. D, Seeman y Michalski, R. S. 80, George Mason University, USA : WIT Press, 2006, Vol. Vol 37. ISSN 1743-3517.

53. Zadeh, Lofti .A. "*Fuzzy Logic and Soft Computing: Issues, Contentions and Perspectives,*" In: *Proc. of IIZUKA'94: Third Int.Conf. on Fuzzy Logic, Neural Nets and Soft Computing.* Japan : lizuka, 1994.