

***Universidad Agraria de la Habana***

***“Fructuoso Rodríguez Pérez”***

**Facultad de Ciencias Técnicas**



**Título**

**“Modelación matemática para optimizar la estructura de los cursos en  
las plataformas e-learning”**

**Tesis presentada en opción al grado científico de Máster en  
Biomatemática**

**Autor: Ing. Amaury Machín Armenteros**

**Tutor: Dr.C Yasser Vázquez Alfonso**

**Mayabeque, 2016**

## **Agradecimientos**

*Son muchas las personas que han contribuido en la realización de esta Tesis.*

*Mis agradecimientos especiales para:*

- ✓ *Mis padres y hermano por su apoyo incondicional en cada momento.*
- ✓ *Mi tutor Dr.C Yasser Vázquez Alfonso por la profesionalidad, rigor científico y sobre todo por lo mucho que me ha enseñado.*
- ✓ *Nuestra Jefe de Departamento MSc. Mercedes Albelo Martín, por sus oportunas exigencias.*
- ✓ *A todo el claustro de profesores de la maestría en Biomatemática.*

*A todos, MUCHAS GRACIAS.*



## **Resumen**

Esta investigación se realizó en la Universidad Agraria de la Habana “Fructuoso Rodríguez Pérez” (UNAH) ubicada en la provincia de Mayabeque. Tuvo como principal objetivo modelar matemáticamente el comportamiento de los usuarios en las plataformas educativas para la optimización de la estructura de los cursos a partir de la matriz de adyacencia y bajo los supuestos que a mayor tiempo de permanencia en una página de la misma sesión, mayor interés en ella, y que la frecuencia de uso de las interrelaciones indica una mayor utilidad de las mismas para acceder hacia la información deseada. Haciendo uso de Minería Web se extrajeron los datos necesarios para aplicar el modelo a los logs del curso Estructura de Datos II de la plataforma Moodle. El modelo propuesto permite optimizar y personalizar la estructura de los cursos en las plataformas educativas, se logra encontrar estructuras de navegación más eficientes desde el punto de la navegabilidad y usabilidad, mejorando en concreto la estructura de los cursos de las plataformas educativas en la UNAH.

# Índice principal

|                                                                                    | Pág.      |
|------------------------------------------------------------------------------------|-----------|
| <b>Introducción .....</b>                                                          | <b>2</b>  |
| <b>Capítulo 1: Revisión bibliográfica .....</b>                                    | <b>6</b>  |
| 1.1 Introducción .....                                                             | 6         |
| 1.2 Modelación matemática. Aplicación en la web .....                              | 7         |
| 1.3 Minería Web. Clasificaciones y características.....                            | 9         |
| 1.4 Técnicas usadas en los sistemas e-learning.....                                | 13        |
| 1.3 Algoritmos genéticos. Características. ....                                    | 17        |
| 1.3.1 Algoritmos Genéticos (AG) .....                                              | 17        |
| 1.3.2 Origen y definición.....                                                     | 18        |
| 1.3.2 Fundamentos Biológicos .....                                                 | 19        |
| 1.3.3 Codificación de algoritmo .....                                              | 21        |
| 1.4 Operadores genéticos.....                                                      | 23        |
| 1.5 Aplicaciones.....                                                              | 29        |
| <b>Capítulo 2: Materiales y métodos .....</b>                                      | <b>32</b> |
| 2.1. Métodos empleados en la investigación .....                                   | 32        |
| 2.2 Datos web .....                                                                | 32        |
| 2.3 Proceso de sesionalización.....                                                | 33        |
| 2.4 Diseño del algoritmo.....                                                      | 34        |
| 2.5 Parámetros del algoritmo.....                                                  | 35        |
| 2.6 Fitness .....                                                                  | 38        |
| 2.6.1 Cálculos previos (fórmulas) .....                                            | 38        |
| <b>Capítulo 3. Resultados y discusión .....</b>                                    | <b>43</b> |
| 3.1 Introducción .....                                                             | 43        |
| 3.2 Descripción del sitio web .....                                                | 43        |
| 3.3 Aplicación del algoritmo .....                                                 | 45        |
| 3.4 Propuesta de acciones para la aplicación del modelo matemático propuesto. .... | 50        |
| <b>Conclusiones .....</b>                                                          | <b>53</b> |
| <b>Recomendaciones .....</b>                                                       | <b>54</b> |
| <b>REFERENCIAS BIBLIOGRÁFICAS.....</b>                                             | <b>55</b> |

## Introducción

El desarrollo de las nuevas tecnologías de la educación y la comunicación han hecho posible la utilización de Internet y más concretamente la WWW (World Wide Web) en la educación a distancia, dando lugar a la denominada Educación basada en Web o *e-learning* (Horton, 2000).

En el sentido literal *e-learning* significa aprendizaje electrónico, es decir, el aprendizaje producido a través de un medio tecnológico-digital. Rosenberg define *e-learning* como el uso de las tecnologías basadas en Internet para proporcionar un amplio despliegue de soluciones que impactan en el aprendizaje y el rendimiento (Rosenberg, 2001, Rosenberg, 2005). Otros autores como Clark y Mayer definen este concepto como la enseñanza transmitida desde un ordenador a través de los medios DVD-ROM, Internet, o intranet con las siguientes características (Clark and Mayer, 2011):

- ✓ Incluye contenido relevante para alcanzar los objetivos de aprendizaje.
- ✓ Utiliza métodos de enseñanza como ejemplos y ejercicios para ayudar en el aprendizaje.
- ✓ Utiliza elementos de comunicación como palabras e imágenes para transmitir el contenido y los métodos.
- ✓ Puede ser enseñanza dirigida por el profesor (*e-learning* síncrono) o diseñada para auto-aprendizaje (*e-learning* asíncrono).
- ✓ Sirve de ayuda a los estudiantes para conseguir sus objetivos educacionales.

En los últimos años, *e-learning* supone una nueva concepción de la educación mediante la incorporación de nuevas tecnologías al proceso de enseñanza y aprendizaje. Esta plataforma educativa proporciona tanto a docentes como alumnos una importante variedad de recursos, lo que en muchas ocasiones esto suele ser una desventaja. Esto hace que el alumno se encuentre desorientando frente a la elección de recursos diseñados. Este problema requiere una reestructuración de cómo el docente ha diseñado la asignatura, con el objetivo de que ese diseño se adapte de mejor manera a las características e intereses de los alumnos (Corso and Alfaro, 2007).

El uso de ambientes virtuales basados en la web, como herramienta de apoyo al proceso educativo, se ha expandido en los últimos años, gracias al desarrollo de nuevos enfoques y modalidades de educación, como el *e-learning* y el *b-learning*. Estas nuevas formas de utilización de la tecnología como apoyo al proceso de enseñanza aprendizaje, han incentivado la aparición de software especializado en proveer los mecanismos que facilitan el desarrollo del trabajo autónomo del estudiante, su interacción con el docente, y el uso de herramientas colaborativas y de comunicación para el trabajo grupal (Aponte et al., 2012).

La introducción de estas tecnologías ha originado una proliferación de experiencias en los países de América Latina y el Caribe, que amerita un mayor y mejor conocimiento de sus características, con miras al mejoramiento de la calidad de la educación superior impartida a través de estas nuevas modalidades de trabajo académico.

Para contribuir a satisfacer estas necesidades, a resolver esta problemática y mejorar la calidad de la educación superior en esta región a través de esas tecnologías, el Instituto Internacional de la Unesco para la Educación Superior en América Latina y el Caribe (IESALC) emprendió en el año 2002 un proyecto sobre la educación superior virtual y a distancia en esta región, el cual se insertó en su Programa Marco denominado Observatorio Sobre Educación Superior en América Latina y el Caribe (Silvio, 2004).

En Cuba, varias de sus universidades han creado diferentes centros de estudios para la investigación en TIC para la Educación Superior, tales como la Facultad de Educación a Distancia de la Universidad de La Habana, el Centro de Referencia en Educación Avanzada (CREA), donde se desarrolló un Modelo de Universidad Virtual CUJAE, la Facultad de Educación a Distancia de la Universidad Central de Las Villas, donde se desarrolló el Sistema de Enseñanza Personalizada a Distancia (SEPAD), el Centro de Estudio de Educación Superior Agraria (CEESA) de la Universidad Agraria de La Habana (Vázquez, 2008).

En la Universidad Agraria de La Habana "Fructuoso Rodríguez Pérez" (UNAH) desde hace varios años se viene trabajando en la aplicación de las TIC a la Educación Superior. Expresiones de esta es la creación de un Laboratorio de Tecnología Educativa (LATED), donde se desarrolla una amplia variedad de software educativos. Pero a pesar de los grandes esfuerzos realizados por crear un

sistema que permita estudiar el comportamiento de los estudiantes mediante el cual el profesor pueda brindarle al estudiante lo que realmente necesita para potenciar sus conocimientos con el empleo de este tipo de tecnologías, aún no se cuenta con un sistema que permita analizar las preferencias y técnicas que usan los estudiantes para estudiar, lo cual posibilitaría perfeccionar la calidad educativa, ofreciendo a los alumnos un servicio más personalizado acorde con sus necesidades e intereses reales y de esta forma facilitar el aprendizaje mediante este tipo de plataformas en la universidad.

Esta situación problemática conlleva a plantear el siguiente **problema científico**: ¿Cómo establecer un modelo matemático que posibilite la optimización de la estructura de los cursos en las plataformas *e-learning* de la UNAH?

**Hipótesis:** Con el empleo de la modelación matemática de los diferentes componentes del *log* mediante los algoritmos genéticos (AG) es posible la optimización de la estructura de los cursos en las plataformas *e-learning* de la UNAH.

El **objeto de investigación** es la modelación matemática relacionada con las plataformas *e-learning* y su **campo de acción** modelo matemático para la optimización de la estructura de los cursos en las plataformas *e-learning* de la UNAH.

El quehacer que se desprende de la problemática antes planteada, hace que se plantee el siguiente **objetivo general**: Desarrollar un modelo matemático para la optimización de la estructura de los cursos en las plataformas *e-learning* de la UNAH.

Para el logro del objetivo expresado anteriormente se plantearon los siguientes **objetivos específicos**:

- Estudiar los referentes teóricos que sustenta la modelación matemática en el análisis de los sistemas *e-learning*.
- Implementar el modelo matemático que posibilite la optimización y personalización de los cursos de las plataformas *e-learning*.

El **Aporte Práctico** se sustenta en el diseño de un modelo matemático que posibilita la optimización de la estructura de los cursos en las plataformas *e-learning* de la UNAH.

## Capítulo 1: Revisión bibliográfica

### 1.1 Introducción

Los sistemas educativos tienen como objetivo principal guiar personalmente a los estudiantes durante su proceso de aprendizaje adaptando los contenidos y la guía de navegación ofrecida, a las características personales y necesidades de cada usuario (Brusilovsky, 1998).

Se busca fortalecer los procesos cognitivos individuales, pero sobre todo con la construcción de nuevos diálogos para la creación de procesos de aprendizaje que trasciendan la idea de la individualidad, en lo que se ha denominado aprendizaje colaborativo (Cardona, 2010).

En los últimos años, el aprendizaje electrónico o conocido también por el anglicismo, *e-learning*, ha generado una nueva concepción de la educación mediante la incorporación de nuevas tecnologías al proceso de enseñanza y aprendizaje. El *e-learning* ha posibilitado que los estudiantes puedan comunicarse y colaborar con otros estudiantes y docentes (profesores, tutores, mentores, entre otros), de forma síncrona o asíncrona, sin limitaciones espacio-temporales. En otras palabras, el *e-learning* se comporta como una modalidad de aprendizaje dentro de la educación a distancia en la que se utilizan las redes de datos como medios (internet, intranets, entre otros), las herramientas o aplicaciones hipertextuales como soporte (correo electrónico, web, chat, entre otros) y los contenidos y/o unidades de aprendizaje en línea como materiales formativos (desde simples imágenes, audio, video, documentos, entre otros, hasta complejas producciones multimedia, así como contenidos contruidos de forma colaborativa), entre otros (García *et al.*, 2008).

Estas plataformas educativas suministran tanto a docentes como alumnos una importante variedad de recursos complementarios a las asignaturas y cursos, lo que en varias ocasiones esto suele constituir una desventaja. Provocando en ocasiones que el alumno se encuentre desorientando frente a la elección de los recursos diseñados. Este problema requiere una reestructuración de cómo el docente ha diseñado la asignatura, con el objetivo que dicho diseño se adapte mejor a las características e intereses de los alumnos.

Desde el punto instructivo del *e-learning*, las actividades de aprendizaje buscan estimular al estudiante en el proceso formativo, y deberían ofrecerse aquellas que mejor se adapten a su estilo, intentando que el estudiante satisfaga sus expectativas personales o propias de educación. En relación a ello, el Diseño Centrado en el Usuario (Vilar, 2010) es una filosofía de diseño, y un proceso donde las necesidades, requerimientos y limitaciones del usuario final, constituyen el centro de cada etapa del proceso de diseño. Involucrando al usuario en cada fase del proceso, se intenta garantizar que el producto final responda a sus necesidades y características, de este modo, se les posibilitaría una experiencia positiva de aprendizaje.

Con el objetivo de desarrollar una herramienta que les permita a los profesores estudiar las necesidades reales de los estudiantes y tenerlas en cuenta para la creación de sus cursos en las plataformas *e-learning* a continuación se exponen una serie de definiciones y herramientas que se tuvieron en cuenta para este propósito.

## **1.2 Modelación matemática. Aplicación en la web**

La modelación matemática y la simulación por computadora son herramientas de mucha utilidad en los diferentes campos del saber humano. En la actualidad, la modelación es un instrumento muy común en el estudio de sistemas y el desarrollo acelerado de las técnicas de computación que han permitido este desarrollo. La modelación ha adquirido un carácter de método científico general, enriquecido por las conquistas de las matemáticas, la cibernética y el enfoque sistémico en la investigación científica, y ha penetrado en las diferentes esferas de la actividad cognoscitiva y transformadora del hombre.

Un modelo matemático es el reflejo de un proceso o sus partes, en forma de relaciones o formulaciones algebraicas. Según (Ortiz, 2000), el modelo es el resultado de estas formulaciones, mientras que la metodología constituye la modelación.

La modelación según (Sheptulin, 1984) se define como la reproducción de determinadas propiedades y relaciones del objeto investigado en otro objeto especialmente creado (modelo) con el fin de su estudio detallado.

En cuanto a la definición de Modelación Matemática diferentes autores la han definido, tal es el caso de (Brown, 1987, Buslenko, 1988, Ambrosano, 1990, Torres, 1995, Wand et al., 1995, Fischer, 1995) sin embargo se considera como la definición más precisa la propuesta por (Buslenko, 1988) que plantea:

*"es la descripción abstracta y formal del objeto cuyo estudio es posible por medio de Modelos Matemáticos".*

En ciencias aplicadas, un modelo matemático es uno de los tipos de modelos científicos que emplea algún tipo de formulismo matemático para expresar relaciones, proposiciones sustantivas de hechos, variables, parámetros, entidades y relaciones entre variables y/o entidades u operaciones, para estudiar comportamientos de sistemas complejos ante situaciones difíciles de observar en la realidad.

Los modelos matemáticos son modelos formales que utilizan el lenguaje de las matemáticas para describir un sistema, expresando parámetros, variables, relaciones (Luo and Rudy, 1994, Bassanezi and Biembengut, 1997).

El lenguaje matemático no se limita a la expresión de números y operadores aritméticos que los relacionan. Así por ejemplo la teoría de grafos, ampliamente utilizada en aplicaciones prácticas, es un "*subconjunto*" de la más general teoría de conjuntos (García and Maheut, 2012).

Según (García and Maheut, 2012) los modelos matemáticos se pueden clasificar de múltiples maneras. A continuación se describen algunas que se consideran relevantes.

Los modelos pueden ser estáticos o dinámicos. Un modelo estático no tiene en cuenta el tiempo, mientras que los modelos dinámicos sí. Los modelos dinámicos se suelen representar con ecuaciones en diferencias o ecuaciones diferenciales.

Los modelos pueden ser lineales o no-lineales. Si todos los operadores de un modelo son lineales el modelo es lineal, si al menos uno es no-lineal el modelo es no-lineal. Aunque hay excepciones, los modelos lineales son mucho más fáciles de manejar que los modelos no-lineales. En general los modelos no-lineales pueden ser linealizados, pero entonces, es posible, que se estén perdiendo aspectos relevantes del problema.

Un modelo puede ser determinista o estocástico. Un modelo determinista es aquel en que cada conjunto de variables en un estado está definido por los parámetros del

modelo y por los estados anteriores. Un modelo determinista se comporta siempre igual para un conjunto de parámetros de entrada. En un modelo estocástico las variables de estado se representan por distribuciones de probabilidad, y por tanto el modelo es capaz de recoger aleatoriedad o incertidumbre.

Los modelos matemáticos son de amplia utilidad en la web el ejemplo más popular es el utilizado por el buscador de internet Google en su famoso algoritmo para rankear páginas web *PageRank* (Velásquez and Palade, 2008). Otros ejemplos de esta aplicación son en el algoritmo para rankear páginas web Hyperlink-Induced Topic Search (HITS) (Chakrabarti et al., 1999), en el análisis de redes sociales, en sistemas de recomendación (Sanjuán et al., 2009), en la minería web, entre otros.

### **1.3 Minería Web. Clasificaciones y características**

En este contexto lo ideal y deseable sería disponer de algún método o herramienta que permita analizar el comportamiento y preferencias del usuario y defina una base de conocimientos que se encuentre estructurada y gestionada debidamente para la ágil y correcta recuperación de información. Teniendo en cuenta lo planteado anteriormente se realizó una investigación con el objetivo de identificar herramientas o métodos basados en la web que facilitaran remediar las necesidades planteadas. Un estudio de la Minería Web, arrojó que esta es definida como el proceso global de descubrir información o conocimiento potencialmente útil y previamente desconocido a partir de los datos de la web (Etzioni, 1996). Además, se enfoca en la aplicación de técnicas para el descubrimiento de patrones en la web, por lo cual esta herramienta podría orientar y ayudar a conocer como el alumno interactúa con las plataformas *e-learning*.

La explotación de la información que se encuentra en la web se puede realizar teniendo en cuenta diferentes puntos de vista. Se puede analizar a través del contenido que podemos extraer o encontrar en la web; este punto de vista es enfocado principalmente en la extracción de conocimiento sobre el contenido de documentos y es llamado *Minería de Contenido*. Otra manera de inferir conocimiento es analizando la organización de la web o de las relaciones entre los enlaces: esta forma de descubrir conocimiento de la estructura de la web es llamado Minería Web de Estructura. En cambio el proceso de extracción de patrones interesantes de la

información de navegación o del tráfico del usuario en la web es llamado Minería de Uso Web (Heughes, 2007).

El proceso de la Minería Web, lo podemos definir formalmente como el uso de técnicas de minería de datos para descubrir y extraer información automáticamente desde el *World Wide Web* (Saloña et al., 2001).

La minería web es una metodología de recuperación de información que utiliza herramientas de la minería de datos. Se enfoca en la aplicación de técnicas para el descubrimiento de patrones en la Web, esta herramienta nos podría orientar y ayudar a conocer como el alumno interactúa con la plataforma de educación virtual y en el caso de enfrentarse con alguna/s dificultades identificar cuáles son esos factores. (Srivastava et al., 2000).

Atendiendo a los aspectos que se desean investigar, la minería web se clasifica en tres ramas:

**Minería de contenido web:** Es un proceso automático que va más allá de la extracción de palabras claves. Se centra en los datos reales que contienen las páginas, es decir el contenido, estos datos consisten generalmente en textos y gráficos. Se puede obtener datos acerca de la forma de escribir, que elementos o recursos son más atractivos para el usuario, si la catalogación que usamos sirve para mejorar un ranking, si los temas que se tratan interesan o no.

**Minería de estructura web:** En esta variante de *web mining*, la información a emplear es aquella que describe la organización del contenido de la página web. La información de la estructura interna de una página incluye por ejemplo el arreglo de las varias etiquetas HTML o de XML dentro de esta. Esto se puede representar como una estructura de árbol en donde una etiqueta (HTML) se convierte en la raíz de este. La clase principal de información en la estructura interna de una página son los “*hyper-links*” o enlaces que conectan una página con otra. El tipo de información acerca de si los usuarios encuentran la información, si la estructura del sitio es demasiado ancha o demasiado profunda, si los elementos están colocados en los lugares adecuados dentro de la página, si la navegación se entiende, cuáles son las secciones menos visitadas y su relación con el lugar que ocupan en la página central.

**Minería de uso web:** La minería de uso en la web consiste en la aplicación de técnicas de minería de datos para descubrir los patrones de uso de la información

web con el objetivo de entender y satisfacer las necesidades de los usuarios (Srivastava et al., 2000).

Estas técnicas de minería de datos ya se han aplicado con éxito en sistemas de comercio electrónico o *e-commerce*, para comprender el comportamiento de clientes en línea de sistemas de comercio electrónico y poder incrementar las ventas (Mobasher et al., 2000). Estas herramientas inteligentes utilizan técnicas de extracción de conocimiento o minería de datos para descubrir información útil para poder mejorar el sistema. Aunque los métodos de descubrimiento de información utilizados en ambas áreas (*e-commerce* y *e-learning*) son similares, los objetivos finales tienen matices totalmente diferentes debido a que en *e-commerce* el objetivo es guiar a los clientes durante la compra para maximizarla, mientras que en *e-learning* el objetivo es guiar a los estudiantes durante su aprendizaje para maximizarlo. Por lo tanto, cada uno tiene unas características específicas que requieren de un tratamiento diferente dentro del problema de minería de Web (Romero et al., 2009b).

Las principales aplicaciones de las técnicas de minería de datos en educación, son los sistemas de personalización, sistemas recomendadores, sistemas de modificación, sistemas de detección de irregularidades, entre otros, la mayoría de estos sistemas se basan en modelos matemáticos. debido a sus capacidades para: el descubrimiento de patrones de navegación regulares e irregulares, realización de clasificaciones de alumnos y de los contenidos, construcción adaptativa de planes de enseñanza, descubrimiento de relaciones entre actividades, diagnóstico incremental de los estudiantes, entre otras (Romero et al., 2009b).

Según (Romero et al., 2009b) la aplicación de técnicas de minería de datos en educación se puede ver desde dos puntos de vista u orientaciones distintas:

- ✓ **Orientado hacia los autores.** Con el objetivo de ayudar a los profesores y/o autores de los sistemas de *e-learning* para que puedan mejorar el funcionamiento o rendimiento de estos sistemas a partir de la información de utilización de los alumnos. Sus principales aplicaciones son: obtener una mayor realimentación de la enseñanza, conocer más sobre como los estudiantes aprenden en el web, evaluar a los estudiantes por sus patrones

de navegación, reestructurar los contenidos el sitio web para personalizar los cursos, clasificar a los estudiantes en grupos, entre otros.

- ✓ **Orientado hacia los estudiantes.** Con el objetivo de ayudar o realizar recomendaciones a los alumnos durante su interacción con el sistema de *e-learning* para poder mejorar su aprendizaje. Sus principales aplicaciones son: sugerir buenas experiencias de aprendizaje a los estudiantes, adaptación del curso según el progreso del aprendiz, ayudar a los estudiantes dando sugerencias y atajos, recomendar caminos más cortos y personalizados, entre otros.

Es importante indicar que aunque el área de la minería de datos aplicada a educación es muy reciente, ya cuenta con un número importante de investigadores, y muestra de ello son las múltiples contribuciones publicadas en diferentes congresos internacionales (ICCE, ICALT, ITS, Elearn, PAKDD, GECCO, UM, AH, WISE, ISDE, etc.) y revistas (IJEL, IEEE Education, UMUAI, etc.), y que su número se está incrementando enormemente cada año, indicando la gran importancia que está teniendo y el interés que está despertando. También indicar que aunque la mayoría de las investigaciones se están realizando dentro de las Universidades, ya se están comenzando a interesarse algunas empresas (SPSS, TEMIS, *Giunti Interactive Labs*, etc.) dado idea de su potencial aplicación comercial (Romero et al., 2009b).

Al realizar una navegación por la web, los usuarios dejan huellas digitales, tales como direcciones de Protocolo de Internet o dirección IP (*Internet Protocol* por sus siglas en inglés), navegador empleado, cookies, url de los sitios visitados, entre otros datos que los servidores almacenan automáticamente en una bitácora de accesos (*logs*). El análisis de ficheros *logs* de los servidores web puede proporcionar variedad de información valiosa que permitiría mejorar la estructura de un sitio web, así como permite estudiar el comportamiento de los usuarios al visitar estos sitios con el objetivo principal de crear una navegación más efectiva y un acceso más eficiente.

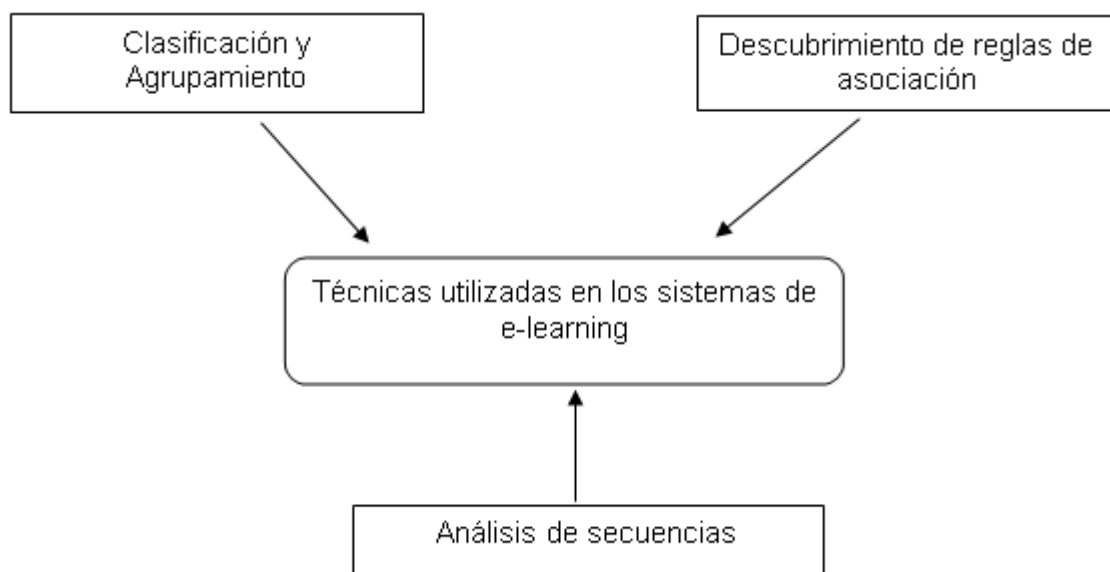
La minería de uso en la web consiste en la aplicación de técnicas de minería de datos para descubrir los patrones de uso de la información web con el objetivo de entender y satisfacer las necesidades de los usuarios (Srivastava et al., 2000).

Las principales técnicas usadas son de análisis estadístico, reglas de asociación, modelos matemáticos para el estudio del comportamiento de los usuarios, agrupación de ítems, clasificación y modelos de dependencia aplicadas a las bitácoras de los servidores web y secuencia de páginas visitadas para realizar transacciones.

Los aspectos en los que se enfoca están orientados a descubrir el uso de las páginas web, direcciones IP, referencias a páginas, la fecha y la hora de accesos, datos que proporcionan la información demográfica sobre los usuarios del Sitio Web. La fuente de datos típica en la minería de uso web son: *logs* de acceso en servidores, *proxies*, agentes y cookies.

#### 1.4 Técnicas usadas en los sistemas e-learning

Las técnicas más utilizadas en la minería de datos aplicada a los sistemas de *e-learning* son: clasificación y agrupamiento, descubrimiento de reglas de asociación, y análisis de secuencias (Romero et al., 2009b). Ver Figura 1.1.



**Figura 1.1:** Técnicas más usadas en la minería de datos aplicada a los sistemas *e-learning*.

Fuente: Elaboración propia

Las técnicas de clasificación y agrupamiento o *clustering* consisten según (Arabie et al., 1996) en la habilidad intelectual para ordenar o dividir fenómenos complejos (descritos por conjuntos de objetos descritos por datos altamente dimensionales) en pequeños y comprensibles unidades o clases que permiten un mejor control o comprensión de la información. Su aplicación a sistemas de *e-learning* permite

agrupar a los usuarios por su comportamiento de navegación, agrupar a las páginas por su contenido, tipo o acceso páginas y agrupar los comportamientos de navegación similares.

La utilización de técnicas de agrupamiento es utilizada por (Tang and McCalla, 2003) para formar *clusters* o grupos de usuarios basándose en su comportamiento de navegación. Utilizan un algoritmo de *clustering* basado en largas secuencias generalizadas. También proponen incluir un sistema recomendador inteligente dentro de un sistema de aprendizaje basado en web evolutivo capaz de adaptarse no sólo a sus usuarios sino también a la Web abierta. El sistema puede encontrar contenidos relevantes en la web y puede personalizar y adaptar sus contenidos basándose en observaciones del sistema y por las propias valoraciones acumuladas dadas por los estudiantes.

Otro trabajo que también emplea agrupamiento es el realizado por (Talavera and Gaudioso, 2004) que analizan los datos obtenidos de cursos basados en sistemas *e-learning* y utilizan técnicas de *clustering* similares al modelo probabilístico de *Naive Bayes* para descubrir patrones que reflejan comportamientos de los usuarios. Su objetivo es utilizar la minería de datos para dar soporte a la tutoría en comunidades de aprendizaje virtual. La utilización conjunta de *clustering* con otras técnicas como secuenciación es realizada por (Mor and Minguillón, 2004) para analizar el comportamiento de navegación de los usuarios para la personalización de los sistemas *e-learning*. Utilizan *clustering* de estudiantes para intentar extender las capacidades de secuenciación de algunos sistemas estándares de manejo de aprendizaje como *Sharable Content Object Reference Model* (SCORM) para incluir el concepto de itinerario recomendado.

Los autores (Hämäläinen et al., 2004) proponen un modelo híbrido que combina técnicas de minería de datos y de aprendizaje de máquinas para la construcción de una red bayesiana para describir el proceso de aprendizaje de los estudiantes. Su objetivo es clasificar a los estudiantes para poder ofrecerles diferentes guías dependiendo de sus habilidades y otras características. Esta tarea se realiza con la categorización y *clustering* de los estudiantes dependiendo de sus habilidades o conocimiento. El trabajo realizado por (Luan, 2002) utiliza técnicas de minería de datos en educación superior y propone la utilización conjunta de predicción y *clustering* dentro de una herramienta de soporte de decisiones.

Las reglas de asociación (Agrawal et al., 1993) descubren relaciones entre atributos de un conjunto de datos que superan unos determinados umbrales. Su aplicación más típica ha sido en los sistemas de comercio electrónico para informar sobre las preferencias de compra de los clientes. Su aplicación a sistemas *e-learning* permite de descubrir relaciones o asociaciones entre distintas páginas Web visitadas.

Los pioneros en la utilización de las técnicas de minería web en sistemas de *e-learning* son (Zaiane and Luo, 2001) y actualmente propone utilizar agentes recomendadores (Zaiane, 2002) para recomendar actividades de aprendizaje en línea o atajos en un curso web basándose en los historiales de acceso y mejorar el proceso de aprendizaje en línea. Concretamente utiliza minería de reglas de asociación para entrenar al agente recomendador y construir un modelo que representa el comportamiento de acceso o asociaciones entre actividades de aprendizaje en línea.

Un trabajo que analiza los ficheros *log* de entornos de aprendizaje web utilizando técnicas de minería de reglas de asociación y filtrado colaborativo, es el realizado por (Wang, 2002) para descubrir patrones de navegación útiles y proponer un modelo de navegación. El modelo de navegación consiste en dos tipos de relaciones: relaciones de asociación y relaciones de secuencia entre documentos.

La utilización de métodos heurístico como regresión lineal en combinación con reglas de asociación es propuesta por (Freyberger et al., 2004) para buscar buenos modelos de transferencia de aprendizaje de estudiantes a partir de los ficheros *logs* de interacción de sistemas tutores inteligentes. El modelo de transferencia relaciona preguntas con los componentes del conocimiento necesarios para responderlas correctamente. Las reglas de asociación guían el proceso de búsqueda para encontrar modelos de transferencia que predicen el éxito de los estudiantes.

(Freyberger et al., 2004) proponen la utilización de reglas de asociación difusas para descubrir relaciones entre patrones de comportamiento de los estudiantes, incluyendo el tiempo de acceso, números de páginas leídas, preguntas contestadas, mensajes leídos y enviados, etc. Mediante monitorización y análisis el sistema indicará mediante umbrales de alarma los resultados a los estudiantes y al instructor. Un trabajo más orientado a los sistemas hipermedia adaptativos es el de Romero y Ventura (Romero et al., 2004) que utilizan minería de reglas como técnica de descubrimiento de información útil para los autores de este tipo de cursos con el

objetivo de poder realizar mejoras tanto del contenido, como de la estructura de los cursos y de su adaptación. Concretamente proponen la utilización de algoritmos evolutivos multiobjetivo para el descubrimiento de relaciones importantes a partir de los datos de utilización (tiempos de acceso a páginas, aciertos y fallos, conocimiento) de los estudiantes. Otro trabajo que también emplea algoritmos evolutivos es realizado por (Minaei-Bidgoli et al., 2003) en el cual realizan un análisis de asociación para predecir el rendimiento de los estudiantes. Utilizan *clustering* de recursos web valorados y descubrimiento de reglas de asociación interesantes mediante algoritmos genéticos para optimización de minería de datos (Minaei-Bidgoli and Punch, 2003) con el objetivo es clasificar a los estudiantes para predecir su clasificación final basándose en las características extraídas de los ficheros *logs*.

El análisis de secuencias o secuencia de patrones (Agrawal and Srikant, 1995) es una técnica de minería de datos que descubre secuencias dentro de un conjunto de datos. Al igual que las reglas de asociación, también se han aplicado en sistemas de comercio electrónico para descubrir secuencias de acciones de los clientes. Su aplicación a sistemas de *e-learning* permite analizar secuencias de páginas visitadas durante una sesión o en distintas sesiones de un mismo usuario. A continuación se describen algunos trabajos de aplicación de minería de datos en *e-learning* que utilizan esta técnica.

El análisis de patrones de navegación en entornos de aprendizaje basado en web es utilizado por (Marquardt et al., 2004) dentro de una herramienta de minería de utilización web para el análisis de patrones y preprocesado de datos de utilización de entornos de aprendizaje basados en web. Las técnicas de descubrimiento de patrones utilizadas son asociación y secuencia, y la secuencia de patrones describen accesos a páginas relaciones en un orden específico.

Otra propuesta de análisis de caminos es el realizado por (Ha et al., 2000) que utilizan el análisis de caminos transversales de páginas web para la personalización de la enseñanza y la asociación de páginas web para la estructuración de conocimiento. El descubrimiento de caminos se realiza desde dos puntos de vista: caminos individuales y caminos agregados que incluye el proceso de *clustering* utilizando mapas autoorganizativos de Kohonen.

El análisis de los patrones de comportamiento es realizado por (Pahl, 2004) para el análisis de la interacción en sistemas *e-learning*. Para ello utiliza técnicas de minería

Web como clasificación, patrones secuenciales y series temporales para el análisis y la evaluación de las interacciones de los estudiantes con los contenidos de los sistemas de *e-learning*.

La minería de la información sobre el comportamiento del aprendizaje en entornos basados en SCORM es el realizado por (Su et al., 2006) para ayudar a los profesores a comprender los motivos del buen aprendizaje o no de los alumnos. Utilizan primero técnicas de minería de secuencias de patrones, posteriormente agrupan a los estudiantes mediante *clustering* y finalmente generan un árbol de actividades personalizadas.

Todas estas técnicas aplicadas sobre Web 2.0 dan soporte a la creación de contenidos a gran escala que, en la mayoría de las ocasiones, son almacenados y pocas veces se utilizan por otros usuarios. Es en este punto donde el concepto de Web 3.0 y su relación con la Inteligencia Artificial juega un papel fundamental.

El concepto que soporta la Web 3.0 está basado en aplicaciones web capaces de aportar un valor añadido al usuario, proporcionándole información precisa y fidedigna. La idea de conocimiento en cualquier momento y lugar resulta de extrema importancia en la actualidad, máxime con la proliferación de diferentes dispositivos fijos y móviles (Gonçalves et al., 2015).

Existen investigadores que consideran que el conocimiento que puede ser procesado de forma automática será la base para que las herramientas Web 3.0 usen e interpreten la información, beneficiando a los autores de contenidos y a los educadores. Existirán plataformas de *e-learning* 3.0 que se adaptarán a cada uno de los alumnos de forma particular (Yang, 2009).

### **1.3 Algoritmos genéticos. Características.**

#### **1.3.1 Algoritmos Genéticos (AG)**

*...”Como de cada especie nacen muchos más individuos de los que pueden sobrevivir, y como, en consecuencia, hay una lucha por la vida, que se repite frecuentemente, se sigue que todo ser, si varía, por débilmente que sea, de algún modo provechoso para él bajo las complejas y a veces variables condiciones de la vida, tendrá mayor probabilidad de sobrevivir y, de ser así, será naturalmente seleccionado. Según el poderoso principio de la herencia, toda variedad*

*seleccionada tenderá a propagar su nueva y modificada forma”...(Darwin and de Zulueta, 2009).*

Los algoritmos genéticos (Sivanandam and Deepa, 2007) pertenecen a la familia de modelos computacionales inspirados en la evolución. Estos algoritmos codifican una solución potencial de un problema específico como un cromosoma contenedor de la estructura de datos (características), para luego, aplicar operadores de selección, cruce y mutación con el fin de preservar la mejor información. Así se construye la mejor solución, es decir, el individuo más adaptado al medio. Estos algoritmos suelen ser vistos para fines de optimización restringidos a ciertas áreas de investigación, sin embargo, el rango de problemas en los que pueden ser aplicados es mucho más amplio.

### **1.3.2 Origen y definición**

La computación evolutiva (Gestal et al., 2010), basa sus prácticas en los principios de Darwin elaborando fundamentos, aplicaciones y técnicas heurísticas. Se clasifica como uno de los campos de investigación en *Soft Computing*.

La computación evolucionaria agrupa una serie de técnicas conocidas como algoritmos evolutivos, entre ellos: algoritmos genéticos, programación evolucionaria, estrategias evolutivas y programación genética. En este marco, los AG son métodos adaptativos. Su definición formal viene dada por David Goldberg, consolidador de este mecanismo en su obra en 1989: *“Los Algoritmos genéticos son algoritmos de búsqueda basados en la mecánica de selección natural y de la genética natural. Combinan la supervivencia del más apto entre estructuras de secuencias con un intercambio de información estructurado, aunque aleatorizado, para constituir así un algoritmo de búsqueda que tenga algo de las genialidades de las búsquedas humanas”* (Golberg, 1989).

Una definición más actual es la desarrolla por John Koza en su obra en 1992: *“Un algoritmo genético es un algoritmo matemático que transforma un conjunto de individuos o población (colección de objetos matemáticos representado un individuo), cada uno de los cuales tiene asociado un valor de adaptación, en una nueva población (la siguiente generación) utilizando una serie de operadores basados en los principios darwinianos de supervivencia del más adaptado”* (Koza, 1992).

### 1.3.2 Fundamentos Biológicos

Los fundamentos biológicos que sustentan el desarrollo de los algoritmos genéticos, tienen relación con los conceptos que participan en todo el proceso de evolución y que han sido llevados de manera analógica hacia la creación del algoritmo.

Los AG comprenden todo el campo de soluciones potenciales que corresponden a los individuos, donde aquellos con mejor adaptación al medio, es decir, con mejor valor matemático que responda al problema cumpliendo las restricciones, son las que tendrán mayor probabilidad de reproducirse y dejar una descendencia que contenga parte de las mejores soluciones. Así, se va expandiendo el espacio de búsqueda en base a procesos no determinísticos en cada generación. Aquellos aspectos o cambios que favorecen la competitividad son preservados, y aquellos aspectos que debilitan su adaptación son eliminados. Estas características favorables o desfavorables, se almacenan y controlan desde unas unidades llamadas genes, que a su vez se agrupan formando conjuntos llamados cromosomas (Gallego et al., 2002).

Aunque no se sabe a ciencia cierta qué información específica codifica cada unidad genética en un individuo, se plantean los hechos aceptados por la comunidad científica sobre la teoría de la evolución (Gestal et al., 2010):

- ✓ La evolución es un proceso que opera fundamentalmente sobre los cromosomas.
- ✓ El proceso de selección natural señala que aquellos individuos más adaptados al medio se reproducen más que aquellos menos aptos.
- ✓ En la reproducción es donde tiene lugar la evolución.

Para profundizar, se lleva a cabo un orden de los conceptos biológicos raíces en el desarrollo del presente documento con su definición científica (Solomon, 2001) y correspondiente analogía algorítmica:

**Gen:** Unidad indivisible de información hereditaria, que interviene en el desempeño de una función que al final afectará alguna característica del organismo. En AG, los genes corresponderán al conjunto de Bits que determinan un parámetro (característica).

**Alelo:** Formas alternativas de un mismo gen que regulan las variaciones de las mismas características determinando el estado recesivo o dominante de información genética. En términos algorítmicos, será la unidad elemental que regirá la variación de un mismo estado de la representación del individuo. Corresponde a cada bit de información binaria tomando valores 0 o 1.

**Cromosoma:** Estructuras del interior del núcleo celular que contiene todas las unidades hereditarias, llamadas genes. Corresponderá por ende, en analogía, a la cadena de valores de todos los parámetros que caracterizan un individuo.

**Genotipo:** Construcción genética completa del organismo, toda la información indispensable en su composición. Por ello, en AG corresponderá al conjunto de características que conforman a un individuo (solución posible) por completo.

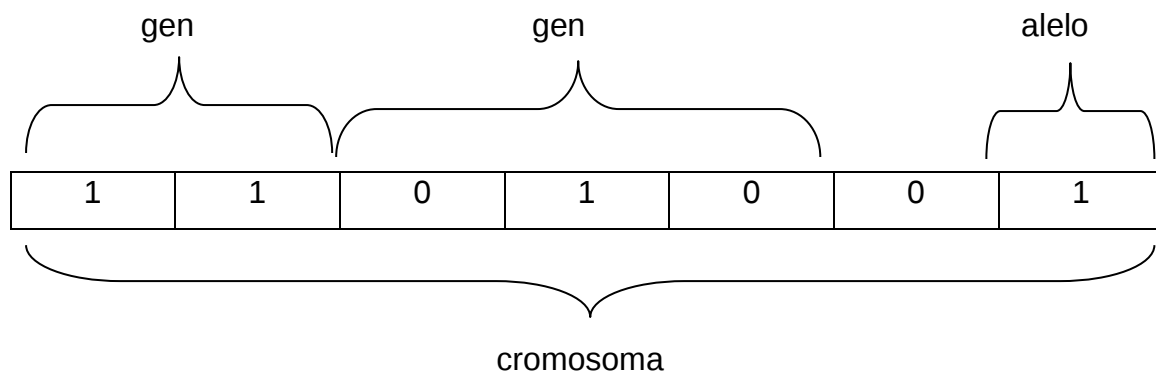
**Fenotipo:** Expresión física de los genes de un organismo, lo “visible” del genotipo, es decir, en analogía, su valor como solución potencial del problema, la adaptación (función *fitness*).

**Individuo:** Organismo vivo que interactúa con el medio y su grupo de semejantes más cercanos, con la principal característica y capacidad de reproducir su especie. En un AG, corresponderá a la representación de las soluciones potenciales del problema a resolver.

**Población:** Conjunto de individuos, el cual ha de representar toda la variedad posible de soluciones, en otras palabras, corresponde a la colección de soluciones candidatas.

**Ambiente:** Corresponde a los factores externos con que interactúa el individuo, condicionando en cuanto a su rigurosidad las adaptaciones necesarias a lo largo del tiempo. En relación al nivel algorítmico, se verá representado por las restricciones y condiciones que ha de cumplir para satisfacer la problemática a resolver.

La Figura 1.2 muestra la representación binaria de un individuo genético.



**Figura 1.2:** Representación binaria de un individuo genético.

Fuente: Elaboración propia

### 1.3.3 Codificación de algoritmo

La clave para un buen desarrollo del algoritmo viene dado por una adecuada representación de los individuos, es decir, como se verán representadas las soluciones candidatas en término de una cadena de valores binarios, enteros u otros, además del por qué de dichos valores. Por otro lado, gran relevancia tiene la definición de la función objetivo, que indique que tanto mejor es la nueva solución hija con respecto a sus progenitoras y también en relación a los demás individuos de la población.

Por mucho tiempo, y desde los inicios, la principal **representación** de los AG ha sido del tipo binario. Dichos números representan en su conjunto, en términos simples, la presencia o no de una o más características, y pese a que puede complicarse mucho la representación binaria, de una u otra forma siempre es posible. Ante dicha complejidad no menor, son necesarias y útiles otras representaciones de acuerdo al problema a resolver.

#### ➤ Representación binaria

Cada alelo tiene sólo dos opciones, ser 0 o 1, tal como muestra la Figura 1.3.

|   |   |   |   |   |   |
|---|---|---|---|---|---|
| 1 | 1 | 0 | 0 | 0 | 1 |
|---|---|---|---|---|---|

**Figura 1.3:** Representación binaria.

Fuente: Elaboración propia

#### ➤ Representación entera

En este caso (Figura 1.4) cada alelo es un valor entero, depositando información a niveles cuantitativos (dar significados propios a cada número).

|   |   |   |   |   |   |
|---|---|---|---|---|---|
| 5 | 7 | 9 | 3 | 1 | 2 |
|---|---|---|---|---|---|

**Figura 1.4:** Representación entera.

Fuente: Elaboración propia

#### ➤ **Representación real**

Cada alelo es un valor real, tal como se indica en la Figura 1.5, donde se abarca un aún mayor espacio cuantitativo, aunque agrega una complejidad adicional, pero que dependerá de las características del problema.

|   |      |     |      |     |     |
|---|------|-----|------|-----|-----|
| 5 | -7,8 | 6,4 | -4,8 | 1,2 | 2,3 |
|---|------|-----|------|-----|-----|

**Figura 1.5:** Representación real

Fuente: Elaboración propia

Gracias a los procedimientos del sistema binario es posible llevar a nomenclatura de ceros y unos tanto números enteros como racionales, pero esto depende de las características del problema a resolver para no exceder en complejidad la representación.

La función de evaluación o ***fitness***, corresponde a la medida de ajuste y adaptación de los individuos a las restricciones del problema, un mayor valor en su función de evaluación indica mejor adaptación a las características del medio, y con eso el ser más cercano al valor óptimo.

Para resolver un problema de optimización, es necesario minimizar o maximizar una función objetivo, ésta corresponderá entonces a la función *fitness*. Al igual que en mecanismos de búsquedas de respuestas óptimas, para poder castigar la función ante el incumplimiento de las restricciones, se procede con lo que se denomina **la penalización**, es decir, dar un valor arbitrariamente muy alto al *fitness* si se está buscando minimizar por ejemplo. En la literatura por otro lado, es posible distinguir cuatro tipos de *fitness* (Koza, 1992):

#### ➤ **Fitness puro $fp(i,t)$**

El valor de bondad de un individuo se mide en base a la diferencia entre su valor deseado  $e$  y su valor obtenido  $r$ .

➤ **Fitness estandarizado  $fe(i,t)$**

Trabajo con valor absoluto del *fitness* anterior para evitar la dificultad para identificar si se está ante un proceso de maximización o minimización.

➤ **Fitness ajustado  $fa(i,t)$**

Proceder con el ajuste estableciendo el cociente  $1/(1+ fe(i,t))$ . Siempre tomará valores entre 0 y 1, donde mientras más cercano esté el *fitness* ajustado a la unidad, mayor será su adaptación.

➤ **Fitness normalizado  $fn(i,t)$**

Se introduce ahora la bondad de la solución en relación a las demás soluciones (población restante), representado por el cociente entre el *fitness* ajustado y la sumatoria total de los *fitness* ajustados.

#### **1.4 Operadores genéticos**

Los siguientes operadores, son aquellas etapas claves en la construcción, desarrollo y convergencia del algoritmo dependiendo de la representación que tenga cada individuo y la adaptación ad hoc al problema a resolver.

##### **Población inicial**

Dado que se trabaja sobre poblaciones, para comenzar se ha de poseer evidentemente una población inicial. Para generarla se pueden emplear diversas técnicas como (Gallego et al., 2002):

- Generación aleatoria de cada valor de los alelos de los individuos.
- Uso de una función ávida (codiciosa en cuanto a su valoración y representación)
- Generar alguna parte de cada individuo y luego aplicar búsqueda local.
- Registro de soluciones reales o conocidas.

##### **Principio de reproducción**

Los algoritmos de reemplazo, los cuales indican como insertar la nueva población generada por algunas de las técnicas anteriores en la nueva generación: (Gestal et al., 2010, Golberg, 1989).

- **Reemplazo de los padres:** Se eliminan los padres, ingresan los hijos.
- **Reemplazo de individuos similares:** Un individuo de la población hija reemplazará a un individuo previo con ajuste similar, dentro de algún rango definido.
- **Reemplazo de los peores individuos:** Se escoge en general al 10% peor de los individuos de la población actual para ser reemplazados por la descendencia.
- **Reemplazo aleatorio:** aleatoriamente se seleccionan los individuos a eliminar.

En estos casos, para pasar a la siguiente generación, se deben alcanzar cierto número de selecciones, cruces y mutaciones, los cuales dependen de las probabilidades respectivas de que se concreten las tasas de dichos operadores.

El operador de reproducción además, necesita establecer la diferencia entre cruce y copia, el primero como operador que representa la reproducción sexual, de donde se obtienen individuos (en general 2 hijos por cada 2 padres) con características compartidas de los padres, mientras que el segundo representa la reproducción asexual, donde un subconjunto de individuos de la población actual, pasa a ser parte, sin modificaciones, de la población de la generación siguiente.

### **Selección**

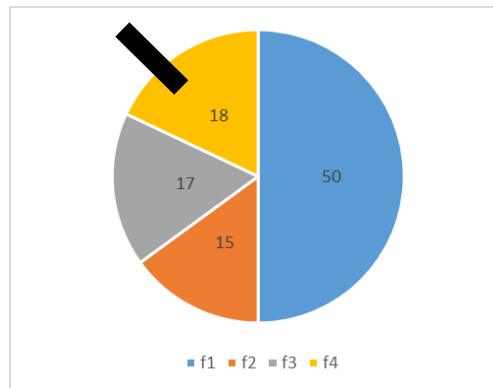
La idea de la selección, es señalar cuáles serían los individuos más propensos a cruzarse, con el fin de dar más posibilidades a aquellos con mejor *fitness*. La idea no es prescindir de los individuos con bajos *fitness* pues igualmente tienen posibilidades de ser seleccionados y con ello, capacidad de ofrecer información útil. El número de individuos a seleccionar, depende del parámetro de tamaño de la población y de aquel que indica cuántos hijos se obtienen a partir de los padres. Hay que tener claro que el operador de selección no genera nuevos individuos, sino que determina cuáles dentro de la población actual dejarán descendencia.

Para llevar a cabo la selección, hay diversas técnicas en la literatura (Golberg, 1989, Holland, 1992, Sivanandam and Deepa, 2007), entre ellas selección proporcional, ranking y torneo.

- **Selección proporcional** o método de la ruleta, desarrollado en un principio por Holland en 1975, y ampliamente analizado por otros autores como Brindle en su Tesis Doctoral en 1981 y (Golberg, 1989). La probabilidad de selección

$p_i$ , para el individuo  $i$  es igual a el valor del *fitness* del individuo  $i$  dividido entre la sumatoria del valor de *fitness* de los individuos desde  $j=1$  hasta el tamaño de la población  $n$ .

La representación gráfica de este sistema es una ruleta dividida en  $n$  áreas cada una de ellas proporcional a la calidad de cada individuo de la población. La selección consistirá rotar  $n$  veces la ruleta y la zona donde caiga el marcador indicará el individuo seleccionado para reproducirse (Figura 1.5).



**Figura 1.5:** Método de la ruleta.

Fuente: Elaboración propia

De esta forma, los individuos con mayor calidad relativa tendrán mayores probabilidades de ser seleccionados para la población de padres, y además más de una vez. Por lo que se presenta el problema de superindividuos y, por lo tanto, de convergencia prematura. Para evitarlo se desarrollaron los siguientes métodos.

- ✓ **Selección estocástica por restos**, que asigna directamente a cada individuo un número de copias en la población de padres igual a  $[p_i \times n]$ , redondeando al menor entero más cercano, y con los restos, se aplica el método proporcional para obtener las copias vacantes. También fue desarrollado por Brindle, en cuyo estudio consideró las posibilidades de aceptar individuos repetidos o realizar un muestreo sin sustitución (Pérez, 2010).
- ✓ **Selección estocástica universal**, similar al método de la ruleta, pero en lugar de tener una ruleta con un único marcador, se tienen  $n$  marcadores igualmente distanciados, donde  $n$  es el número de selecciones que deseamos efectuar y mediante una tirada se selecciona a los  $n$  individuos.

#### ➤ **Ranking**

Con el fin de corregir la convergencia prematura por la presencia de superindividuos (Mor and Minguillón, 2004) anteriormente indicada, es que se

desarrolla la selección vía método de Ranking. Éste corresponde a ordenar la población en cuanto a su valor de adaptación. El más habitual es al ranking lineal.

➤ **Torneo**

Este método en general contempla la selección aleatoria de dos factores, un par de individuos y un número entre 0 y 1. Entonces, teniendo a los dos individuos, si el factor 2 es mayor que un nivel dado (comúnmente 75%) se selecciona dentro del par de individuos, el individuo con mejor adaptación, si no es así, se selecciona aquel con peor *fitness*. Por la sencillez enunciada, cabe destacar que éste método tiene un bajo costo computacional, dado que no requiere de algoritmos de muestreo adicionales, y por ello, no se basa en valores esperados.

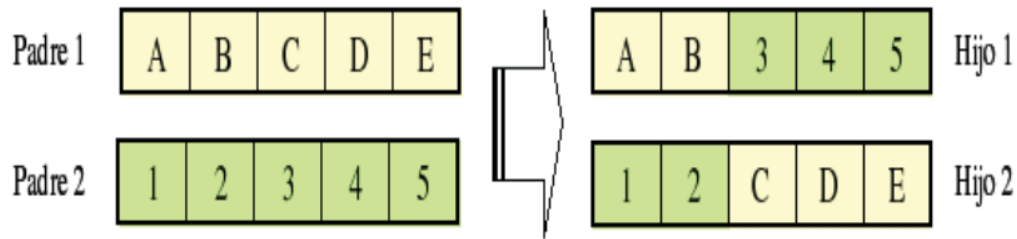
Una medida para evaluar el desempeño de los métodos de selección es su presión selectiva (Valencia, 1997), la cual se define como *“el tiempo requerido por el mejor individuo para llenar la población con copias de sí mismo (takeover time) cuando no actúa otro operador genético”*. Para valores estándar de sus parámetros, los métodos de selección ordenados de mejor a menor presión selectiva son torneo, ranking y selección proporcional.

➤ **Cruce**

Una vez ya realizada la selección, y con eso la expansión en el espacio de búsqueda hacia los individuos potencialmente mejores, es hora de explorar y trabajar la información almacenada, combinándola adecuadamente, con la idea de aprovechar la alta adaptación de los padres y con ello, las ventajas de dicha combinación. La forma de cruce determina los distintos tipos existentes, dependiendo de la cantidad y distribución de los puntos de corte.

➤ **Cruce de un punto**

Método de cruce más sencillo, donde se selecciona una de las posiciones intermedias de la cadena de representación de las soluciones y se intercambian los genes, formando dos (por lo general) nuevos individuos, como se observa en la Figura 1.6.

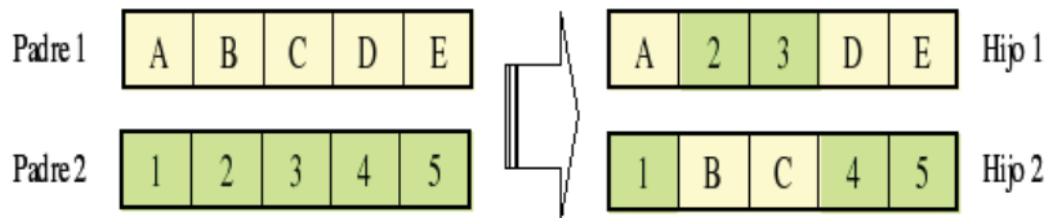


**Figura 1.6:** Cruce de un punto.

Fuente: (Gestal, 2008).

➤ **Cruce de dos puntos**

Pequeña modificación en cuanto al tipo de cruce anterior donde ahora, tal como se indica en la Figura 1.7, son dos puntos de intercambio de información.



**Figura 1.7:** Cruce de dos puntos.

Fuente: (Gestal, 2008).

➤ **Cruce de n puntos**

Extensión de los dos anteriores con “n” puntos de cruce.

➤ **Cruce Uniforme/Binomial**

Se realiza un test aleatorio intermedio para decidir cuál será el progenitor que aporte su información a la descendencia. Dicho test intermedio puede ser representado por un nuevo individuo con valores binarios, donde 0 indica que la descendencia tendrá la característica del padre 1 en dicha posición y 1, que poseerá la característica del padre 2. El inconveniente de este método es que por su construcción es posible obtener un solo hijo.

El operador de cruce opera con probabilidad  $p_c$ , lo que permite que haya posibilidad de parejas de padres sin cruce, manteniéndose ellos mismos en la siguiente generación. De esta manera algoritmo, explora un amplio espacio de soluciones y mediante la iteración de operadores de selección y recombinación, va convergiendo hacia las regiones con mayores grados de adaptación.

## **Mutación**

Si con la selección se expande el espacio de búsqueda y con el cruce se combina la información previamente seleccionada, con la mutación se busca no perder la diversidad posible y no converger a un óptimo local, pues, el cambiar cierto número de genes de una solución existente, se está fomentando la variabilidad dentro de una población. Cabe destacar que la tasa de mutación  $p$  debe ser pequeña, para no tener demasiada diferencia con los padres. Con este operador, se evita la pérdida de diversidad producto de bits que han convergido a un cierto valor para toda la población, lo que no es recuperado por el operador de cruce.

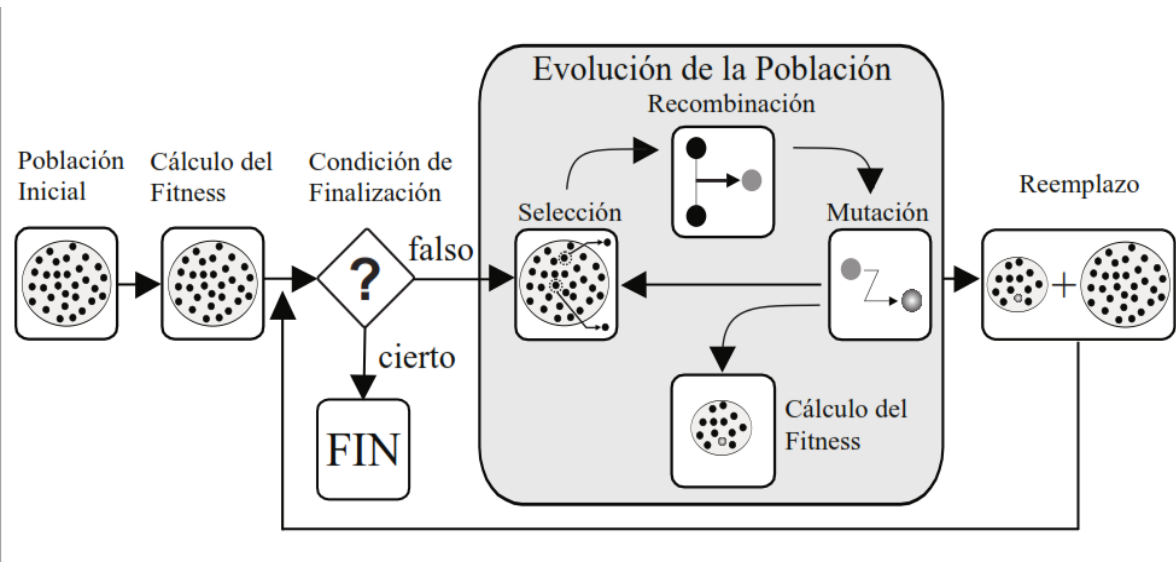
## **Criterios de detención**

Los operadores anteriores se van aplicando hasta que no se cumpla algún criterio de detención previamente definido. Los mecanismos para detener la simulación de más y más generaciones, viene dado por alguna de las siguientes (o un subconjunto de más de dos opciones) alternativas comúnmente aplicadas:

- La diferencia entre los mejores *fitness* de las poblaciones sucesivas es menor que un pequeño  $\varepsilon$  definido inicialmente.
- Un 95% de la población de soluciones potenciales posee el mismo valor para un gen, con lo que se dice que la población ha convergido.
- Se alcanzó el número de generaciones máximas especificadas
- Se ha superado el tiempo máximo de ejecución predefinido

Una vez detenida la funcionalidad del algoritmo, la solución será el ajuste del mejor individuo en la última generación.

En la Figura 1.8 se puede observar el funcionamiento de un AG típico.



**Figura 1.8:** Funcionamiento de un AG típico.

Fuente: (Dorransoro, 2007)

## 1.5 Aplicaciones

Los algoritmos genéticos y otros algoritmos como *Simulated Annealing* y *Direct Search*, son usados para enfrentar problemas difíciles de resolver por técnicas de optimización tradicionales, incluyendo problemas no bien definidos o difícilmente modelados matemáticamente, además, son usados cuando la función objetivo es discontinua, fuertemente no lineal, estocásticos o tiene poco confiables o indefinidas derivadas.

Con el fin de fomentar la rapidez y potencial computacional de los AG simples, nacen los AG paralelos. En este contexto se encuentran los **algoritmos de nichos paralelos** (Valencia, 1997), los cuales obtienen más de una solución cuando no hay solo un valor óptimo, como es el caso de la optimización de funciones con múltiples máximos donde un algoritmo genético tradicional solo convergería a un individuo. Por otro lado, es posible ejecutar algoritmos genéticos en paralelo en distintas computadoras, donde una de ellas coordinará el intercambio, procesamiento, evaluación y limpieza de la población de cada una, esto se le ha denominado **máquina paralela** (Romero and Parada, 2009).

El *web mining* está creando nuevos desafíos hacia los diferentes componentes y tareas del mismo, sobre todo ante el hecho de que la cantidad de información en la Web se está incrementando y cambiando rápidamente sin control (Bandyopadhyay and Pal, 2007). Con el descubrimiento y análisis de la información útil gracias al

diseño, búsqueda y obtención de algoritmos que pueden encontrarla, surge la necesidad de incorporar inteligencia artificial dentro de herramientas Web. Así, los AG y sus características han demostrado interesantes resultados en diversas áreas del *web mining*. Por ejemplo, en búsqueda y recuperación de información, se presenta GSearch (Crestani and Pasi, 2013) como mecanismo de búsqueda basada en AG, también GMiner (Picarougne et al., 2002) como una máquina de búsqueda complementaria a las estándar y GCrawler (Shokouhi et al., 2005), con el objeto de optimizar la búsqueda vía *crawlers* mejorando el rendimiento de la misma. En optimización de consultas, se presenta un AG para construir perfiles de usuarios, que permite monitorear el comportamiento de navegación usando representación de documentos y consultas vía vectores que permitan representarlas (Shokouhi et al., 2005). Otra aplicación relevante de los AG en *web mining* tiene que ver con la representación de documentos, con el fin de derivar en la mejor de ellas en base a sets de términos indexados (Gordon, 1988), como también en relación a la minería distribuida, donde GEMGA (*Gene Expression Messy Genetic Algorithm*) (Kargupta and Buescher, 1996) es el algoritmo de búsqueda evolucionaria paralela que trata los problemas de búsqueda de patrones en un medio ambiente donde tanto los datos como los recursos computacionales son distribuidos.

Cabe destacar además, las investigaciones en relación a la estructura web y el estudio del comportamiento del consumidor, aspectos relacionados directamente con la investigación del presente documento. Aquí se destacan cuatro fuentes fundamentales. En (Said et al., 2002) se presenta un simulador del comportamiento del consumidor en general, para simular los efectos de las estrategias de marketing en un contexto de mercado competitivo. Ello fue realizado al implementar un modelo para obtener consumidores virtuales de un mercado real, ajustando las características de ellos mediante algoritmos genéticos y simulación multi-agente. Por otro lado, una combinación de múltiples clasificadores (*majority voting*, método bayesiano, BKS, *borda count* y redes neuronales) para la predicción del comportamiento de compra de un consumidor en los *e-commerces* llevados a un AG de modo de obtener las mejores combinaciones, se reflejan en (Kim et al., 2003). Como tercera fuente en (Marghny and Ali, 2005) se plantea, una visión de la búsqueda web como un problema de optimización introduciendo en un AG una población de páginas para obtener aquellas más interesantes para el usuario.

Finalmente, en (Lin and Liu, 2008) se busca construir un sitio que proporciona a sus usuarios la información deseada con la menor cantidad de enlaces posibles, analizando la estructura y contenido del mismo, es decir, construir un modelo para optimizar la estructura Web para una navegación más efectiva, usando un AG para precisar los enlaces que cumplen con el objetivo.

En definitiva la aplicación de AG en el dominio del *web mining* provee una perspectiva y dirección potencial para futuras investigaciones (Bandyopadhyay and Pal, 2007) sobre todo con miras hacia los nuevos desafíos de proponer modelos de optimización dinámica para hacer frente a las complejas características de la Web (el dinamismo en la información solicitada, enlaces existentes, estructura de un sitio, contenidos, entre otros. (Lin and Liu, 2008)) hacia la obtención sitios web adaptativos (Velásquez and Palade, 2008) de comportamiento óptimo.

## Capítulo 2: Materiales y métodos

### 2.1. Métodos empleados en la investigación

En la investigación se utilizaron los métodos teóricos siguientes:

- **Histórico lógico** se realizó una reseña acerca de las características y utilización de los *e-learning* y los AG, así como de las principales técnicas utilizadas y aplicaciones respectivamente.
- **Análisis y síntesis:** consistió en un estudio pormenorizado de las características de los archivos *log* en los sistemas web, arribando a los elementos esenciales para su empleo en un modelo matemático que describa el comportamiento de los usuarios y optimice la estructura de los cursos.
- **Enfoque de sistema** se emplea para el estudio de los diferentes componentes de la Minería Web.
- **Inducción-deducción:** se emplea para la creación de un modelo matemático a partir de los *log* permitiendo el análisis del comportamiento de los usuarios y la optimización de la estructura de los cursos en las plataformas *e-learning*.

Como método empírico se utilizara:

- **Sondeo o levantamiento:** se utilizó la técnica análisis documental para la recopilación de datos e información necesarios para el estudio del comportamiento de los usuarios y la optimización de la estructura de los cursos.

El método matemático que se utiliza es:

- **Modelación matemática:** se modeló a través de ecuaciones matemática utilizando el algoritmo genético para el estudio del comportamiento de los usuarios y la optimización de la estructura de los cursos en las plataformas *e-learning* de la UNAH.

### 2.2 Datos web

Con el fin de anticiparse a la toma de decisiones, el análisis y entendimiento del comportamiento del usuario cobra una importancia fundamental, donde se hace necesario seguir en detalle sus acciones, las cuales podrían incluir toda la

trayectoria de páginas visitadas, el tiempo gastado en cada una de ellas, los productos escogidos, cantidad de transiciones antes de llegar a concretar una compra, etc. (Velásquez and Palade, 2008). Así, con el hecho de navegar en la Web se van acumulando una serie de datos que proporcionarán información relevante: los registros web *log* (Ortega-Priego and Aguillo, 2009). Ellos contienen todo el recorrido de los usuarios, descargas asociadas (páginas, objetos multimedia, archivos, entre otros) y tiempo invertido. Estos se generan cuando se realiza la lectura del código HTML de cada página, que llega y se interpreta por el navegador, registrando cada objeto entregado. Es importante indicar que aunque solo una página sea visitada, más de un objeto puede ser registrado (Velásquez and Palade, 2008), entre ellos una alta cantidad de objetos sin información relevante (banner publicitarios, íconos, etc.), aunque dicha relevancia dependerá del estudio que se quiera realizar, por lo que el filtrar y registrar previamente los datos de interés se vuelve una tarea clave.

En la Figura 2.1 se muestra un ejemplo de *log* de la plataforma Moodle en la UNAH.

| Fecha              | Dirección IP | Nombre                     | Acción      | Información                                      |  |  |
|--------------------|--------------|----------------------------|-------------|--------------------------------------------------|--|--|
| 2015 mayo 28 15:48 | 10.4.41.11   | Elio Minel del Pozo Nunez  | course view | EVEA Agronomía                                   |  |  |
| 2015 mayo 28 15:45 | 10.4.41.11   | Elio Minel del Pozo Nunez  | course view | Química General                                  |  |  |
| 2015 mayo 28 15:45 | 10.4.41.11   | Elio Minel del Pozo Nunez  | course view | Química Inorgánica y Analítica                   |  |  |
| 2015 mayo 28 15:44 | 10.4.41.11   | Elio Minel del Pozo Nunez  | course view | Problemas Sociales de la Ciencia y la Tecnología |  |  |
| 2015 mayo 28 15:44 | 10.4.41.11   | Elio Minel del Pozo Nunez  | course view | Economía Política del Capitalismo                |  |  |
| 2015 mayo 28 15:39 | 10.4.41.11   | Elio Minel del Pozo Nunez  | course view | Ecología y Elementos de Agroecología             |  |  |
| 2015 mayo 11 10:29 | 10.4.132.143 | Arlenys Cruz Mendez        | course view | Ecología y Elementos de Agroecología             |  |  |
| 2015 abril 23 4:20 | 10.4.40.4    | Arlenys Cruz Mendez        | course view | Ecología y Elementos de Agroecología             |  |  |
| 2015 abril 23 4:20 | 10.4.40.4    | Arlenys Cruz Mendez        | course view | Ecología y Elementos de Agroecología             |  |  |
| 2015 abril 8 7:27  | 10.4.15.6    | Alejandro Emilio Ramos Rod | course view | Proyecto agrícola                                |  |  |
| 2015 abril 8 7:26  | 10.4.15.6    | Alejandro Emilio Ramos Rod | course view | EVEA Agronomía                                   |  |  |
| 2015 abril 6 4:07  | 10.4.40.2    | Hector Rodriguez Morell    | course view | EVEA Agronomía                                   |  |  |
| 2015 abril 2 9:01  | 10.4.128.121 | Valentin Garcia Castillo   | course view | Práctica Agrícola I                              |  |  |

**Figura 2.1:** Ejemplo de *log* de Moodle.

Fuente: Elaboración propia

## 2.3 Proceso de sesionalización

Una sesión de usuario está definida por una secuencia de accesos temporales a un sitio particular de la web por el usuario (Menascé et al., 1999). Cada sesión de usuario es una representación lógica de una visita de un usuario a algún sitio web usualmente dentro algún intervalo de tiempo.

Entonces podemos definir una sesión de usuario *S* como el conjunto de páginas *P* visitadas:

$$P = \{p_1, p_2, p_3, \dots, p_n\}$$

Teniendo los registros *web log* se hace necesario ahora estudiar las rutas de cada usuario individualizando sus sesiones. A aquél proceso de reconstrucción de caminos tomados se le llama **sesionalización** (Velásquez and Palade, 2008). El propósito del mismo, es encontrar las sesiones reales de cada usuario, por ello se han propuesto heurísticas para reconstruirlas desde los *web log* para asociar los registros a una sesión única durante un período de tiempo. Las técnicas ocupadas pueden ser clasificadas según la estrategia que siguen, ya sea **proactiva** o **reactiva**: Para la realización de este trabajo se utilizará una estrategia proactiva conocida como Cookies o simplemente la inscripción que realiza el usuario en el sitio web.

## 2.4 Diseño del algoritmo

Si una compañía desea permanecer competitiva en el mercado digital, necesita un sitio web que ofrezca la información en concreto que los usuarios están buscando, de manera simple y accesible. Sin embargo, la contingencia indica que en muchos casos la estructura del sitio no ayuda a los usuarios a encontrar la información deseada, incluso cuando ésta si existe (Velásquez and Palade, 2008). Ante eso, y con el fin de determinar la mejor estructura de las plataformas *e-learning*, es que desde ahora se describe un modelo matemático que mediante el estudio del comportamiento de los usuario (gracias a registros *web log*) permita determinar el mejor grafo para una navegación más eficiente (maximizando la utilidad de los enlaces existentes en base a los principios de usabilidad) en el sitio web.

Considerando los datos de entrada que se observan en la Tabla 2.1, cabe señalar preliminarmente los supuestos y observaciones propios de la extracción de datos y operación de los mismos.

**Tabla 2.1:** Datos de entrada del algoritmo.

| Datos                                                |
|------------------------------------------------------|
| Número total de páginas                              |
| Sesiones de usuario                                  |
| Matriz de adyacencia = existencia de enlaces o links |
| Rutas de las páginas visitadas                       |
| Tiempo de permanencia en cada página                 |

Fuente: Elaboración propia.

## Supuestos y Observaciones

1. A mayor tiempo de permanencia en una página de una misma sesión, mayor interés tiene el usuario en ella (Velásquez and Palade, 2008).
2. Sesiones de largo 1 no se consideran.
3. No se considera la comparación de secuencias totales (clúster de sesiones (Velásquez and Palade, 2008)), es decir, se consideran las relaciones entre las de páginas en cuanto a frecuencias de uso.
4. No se considera la existencia de más de un enlace en una misma página  $i$  saliendo a otra  $j$ , se tomará como un solo camino presente entre dichas páginas.
5. Los cálculos son contruidos en base a los registros de las visitas al sitio considerando cada transición realizada, es decir, se considera el comportamiento de los usuarios para la construcción de sus sesiones, teniendo implícito por lo tanto un cálculo basado en el uso.

## 2.5 Parámetros del algoritmo

Los parámetros son aquellos valores de entrada que no cambian a lo largo de la ejecución del algoritmo y que dependen en gran medida de las etapas a implementar. En el caso de los AG se hace necesario definir:

### Tamaño de la los individuos:

En este caso, corresponde a un vector fila de largo  $m \times m$  columnas, donde  $m$  es el número total de páginas del sitio. Sus valores binarios internos representan la existencia o no de enlaces entre páginas.

### Tamaño de la población:

Indica cuántos individuos habrá en cada generación. Desde el punto de vista algorítmico, tiene que ver con el espacio de las soluciones posibles a estudiar.

### Criterio de detención:

Son aquellos criterios por los cuales se establece que el algoritmo termine. Estos pueden ser fijando un número máximo de iteraciones que el algoritmo realice (número máximo de generaciones), indicando un tiempo límite de ejecución del mismo precisando un *fitness* límite (se detiene el algoritmo al encontrar valores menores o iguales a dicho *fitness*), detallando una función de tolerancia que considere el cambio acumulado en el valor del *fitness* a través de las generaciones,

o bien, considerando intervalos de tiempo máximo. De esta manera, si no hay mejoras en la función objetivo, se detiene el algoritmo.

Los restantes parámetros corresponden a los operadores genéticos propios de los AG, los cuales dependerán del estudio que se desea realizar.

**Fitness scaling:**

Especificación de la escala a realizar entre todos los individuos de modo de compararlos en base a dicha escala.

**Selección:**

Elección de los padres de la próxima generación, basado en su valor *fitness* escalado.

**Reproducción:**

Forma en que se crearán los hijos, especificando qué individuos tendrán sobrevivencia garantizada, qué porcentaje serán producidos vía cruce y qué porcentaje vía mutación.

**Mutación:**

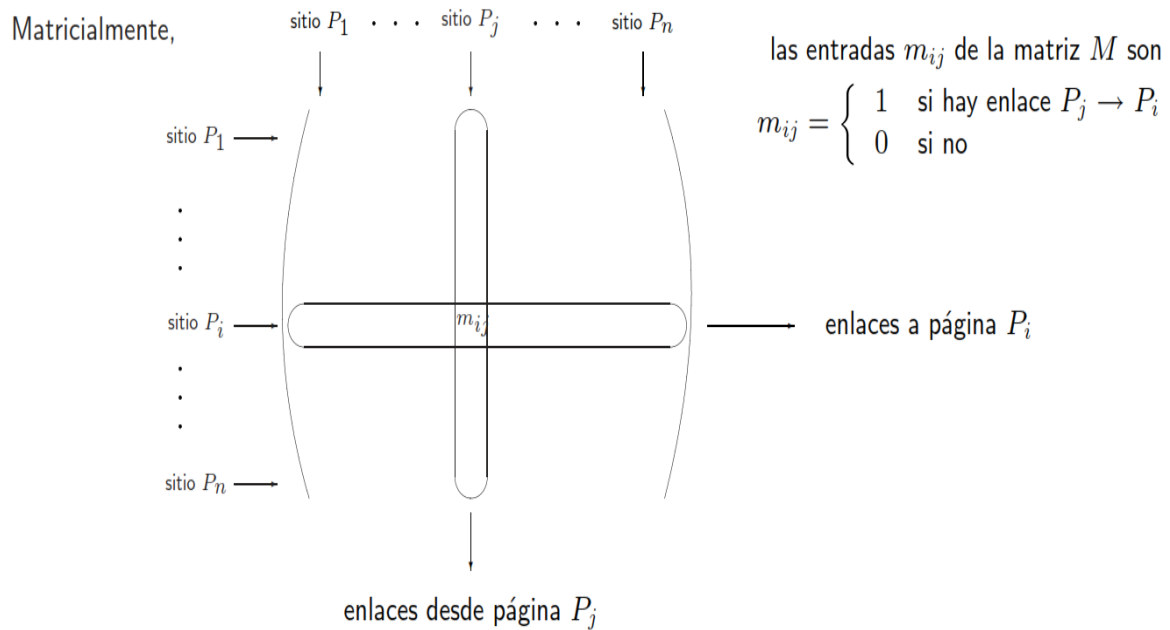
Porcentaje de cambio aleatorio en los individuos de la población, el cual provee diversidad genética y explorar un espacio más amplio.

**Cruce:**

Forma de combinación de los padres, para formar un nuevo individuo.

## **2.6 Representación del individuo**

Con el objetivo de encontrar una mejor estructura web de un sitio, se establece como individuo genético la matriz de adyacencia  $M$ , matriz binaria contenedora de la existencia de enlaces como se muestra en la Figura 2.2.



**Figura 2.2:** Matriz de adyacencia.

Fuente: (Fernández, 2009).

Con el fin de construir un individuo binario lineal para aplicar el AG se procede:

### Largo de individuo

Corresponde al total de datos contenidos en la matriz estructura, es decir, el individuo lineal tendrá un largo  $L$  igual a  $m \times m$ , siendo  $m$  el número total de páginas del sitio.

### Linealización

Se llamará **linealización** al proceso por el cual se construirá el arreglo lineal (cadena) de representación del individuo genético mediante la matriz estructura (se linealiza la matriz concatenando las filas en un solo gran vector), donde el primer bit corresponde a la existencia o no de un enlace entre la página 1 y 1, segundo bit: existencia de enlace entre 1 y 2, tercer bit: existencia de enlace entre 1 y 3, y así sucesivamente. La existencia es indicada en términos binarios: 1= si hay enlace, 0=no hay enlace (ver Tabla 2.2). Por lo que para un máximo de tres páginas la representación de un individuo genético sería como se muestra en la Figura 3.3:

**Tabla 2.2:** Matriz de adyacencia de  $m=3$  páginas

| Páginas | 1 | 2 | 3 |
|---------|---|---|---|
| 1       | 0 | 1 | 1 |

|   |   |   |   |
|---|---|---|---|
| 2 | 1 | 0 | 0 |
| 3 | 1 | 1 | 0 |

Fuente: Elaboración propia

| c1  | c2  | c3  | c4  | c5  | c6  | c7  | c8  | c9  |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1→1 | 1→2 | 1→3 | 2→1 | 2→2 | 2→3 | 3→1 | 3→2 | 3→3 |
| 0   | 1   | 1   | 1   | 0   | 0   | 1   | 1   | 0   |

**Figura 3.3:** Ejemplo de representación de un individuo genético con m=3 páginas (la primera fila corresponde a los bit, la segunda a los enlaces y la tercera al individuo).

Fuente: Elaboración propia

## 2.6 Fitness

La función *fitness*, tal como se indicó en la sección 1.3.1 Algoritmos Genéticos, es clave para el correcto desarrollo, convergencia y ajuste de los datos reales (Marghny and Ali, 2005), por lo que una clara definición es indispensable con el fin de obtener lo deseado. Antes de comenzar, se realizan cálculos previos de acuerdo al modelo a desarrollar para luego establecer la función de adaptación de los mismos en búsqueda del individuo óptimo.

La idea será construir las frecuencias de uso de los enlace, considerando las rutas seguidas para ir de un nodo a otro, independiente de la cantidad de páginas visitadas entre ellas.

### 2.6.1 Cálculos previos (fórmulas)

La cantidad de veces que se pase desde una página a otra es independiente del número de pasos requeridos para ello, se agrupa en un solo gran conteo realizado sobre todas las sesiones, determinando la frecuencia de llegada a la página  $j$  desde el nodo  $i$ , tal como se observa en la Ecuación 2.1.

$$f_{arr}(i,j) = \frac{\text{Cantidad de pasadas de } i \text{ a } j}{\text{Cantidad de pasadas de } i \text{ a cualquier página}}$$

**Ecuación 2.1:** Frecuencia de llegada.

$f_{arr}(i,j)$  corresponde a la fracción entre el número de veces de pasar de  $i$  a  $j$  con el total de veces de uso de enlaces que salen desde  $i$ , por ende, corresponde a la

probabilidad condicional de acceder desde  $i$  a  $j$ , es decir, la probabilidad estacionaria (Roose, 2000) de uso del enlace o bien, la probabilidad de transición entre  $i$  y  $j$ .

Por otro lado, se necesita calcular la frecuencia de estadía, que corresponde al factor tiempo condicional de visita a la página de destino, como se muestra en la Ecuación 2.2.

$$f_t(i,j) = \frac{\text{Factor de tiempo en } j \text{ al acceder desde } i}{\text{Sumatoria del factor de tiempo de las páginas visitadas desde } i}$$

**Ecuación 2.2:** Frecuencia de estadía.

Donde Factor tiempo (Ecuación 2.3) corresponde al tiempo neto (en segundos) en la página  $i$  en la sesión  $S$ , dividido por el tiempo total de la sesión  $S$  (sumatoria sobre el tiempo de todas las páginas que pertenecen a dicha sesión). Lo anterior bajo el supuesto que “el tiempo gastado en una página es proporcional al interés que el visitante tiene en su contenido” (Velásquez et al., 2004).

$$\text{Factor tiempo}(i) = \frac{ts(i)}{\sum_{i \in S} ts(i)}$$

**Ecuación 2.3:** Factor tiempo.

Donde:

$ts(i)$  = Tiempo de permanencia (en segundos) de la página  $i$  en la sesión ( $S$ ).

Por último se calcula la frecuencia de utilidad de las páginas  $i$  y  $j$  como se muestra en la Ecuación 2.4.

$$f_u(i,j) = f_u(i) + f_u(j)$$

**Ecuación 2.4:** Frecuencia de utilidad.

Donde:

$$f_u(i) = f_u(j) = \frac{\text{Cantidad de sesiones en las que se ha visitado la página } i \text{ o } j}{\text{Cantidad de sesiones totales}}$$

**Ecuación 2.5:** Frecuencia de la páginas  $i$  y  $j$ .

Con estos tres valores, se procede a realizar el cálculo de los pesos potenciales (Ecuación 2.6).

$$\omega(i,j) = f_{arr}(i,j) + f_t(i,j) + f_u(i,j)$$

**Ecuación 2.6:** Pesos potenciales.

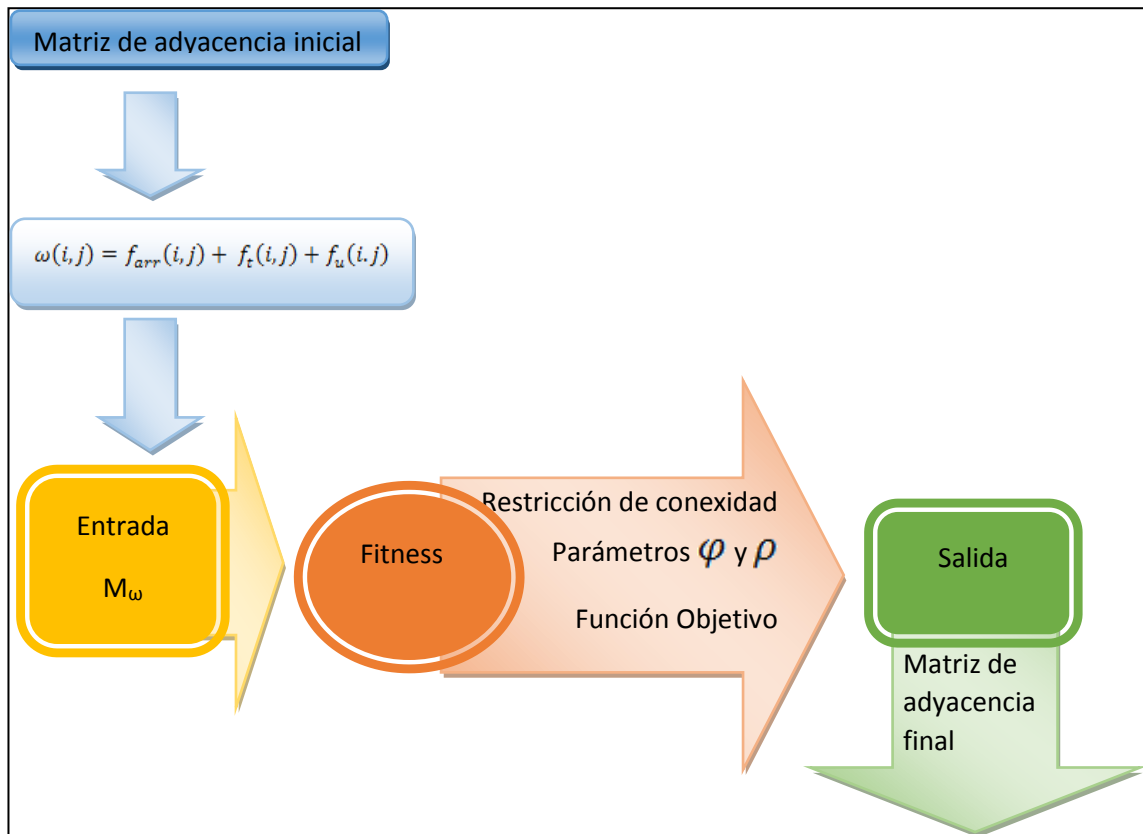
Se puede desprender que numéricamente se dispondrá de los pesos para la mayoría de las relaciones entre todas las páginas, sin embargo, habrá pares de páginas con pesos nulos. Esto significa que ninguna de las sesiones accedió desde  $i$  hacia  $j$ , ni siquiera como página de inicio y fin (máxima distancia entre ellas). Ante ello, dicho enlace tendrá peso nulo siempre, pues por la gran cantidad de datos desde los que se realizaron los cálculos, se puede establecer que su uso no sucede (ley de los grandes números (Roose, 2000)).

### 2.6.2 Etapas del algoritmo

En la Figura 2.3 se pueden observar las principales etapas para la aplicación del algoritmo genético.

A continuación se exponen algunos aspectos que son necesarios para la aplicación de estas etapas:

- Para determinar la conexidad del grafo se debe llevar la matriz cuadrada de los individuos a matriz laplaciana (Jeraldo, 2001).
- Para calcular el parámetro  $\varphi$  es necesario tener en cuenta el valor de la cantidad de enlaces totales de la matriz inicial y para calcular  $\rho$  se debe de tener en cuenta el valor de la función beneficio de la matriz de adyacencia inicial. Estos valores estarán en un rango aproximadamente de -10 con respecto al valor inicial y serán definidos por el usuario.
- Finalmente se procede a calcular el valor de la función *fitness*.



**Figura 2.3:** Etapas del algoritmo AG.

Fuente: Elaboración propia.

## Entrada

Esta etapa inicial consiste en aplicar la Ecuación 2.6 de modo de determinar el peso.  $M_\omega$  es la matriz que contiene el peso de los enlaces entre dos pares de páginas, es decir, el hecho de poseer un valor más alto indica que hay más interés en la comunicación entre ambas páginas.

## Función objetivo o Fitness del algoritmo AG

Como se ha visto en sesiones anteriores de este trabajo la función de aptitud (*fitness*) es la que permite valorar la aptitud de los individuos y debe tomar siempre valores positivos.

En el presente trabajo la función de evaluación que se utilizará para evaluar a la nueva generación de individuos genéticos está definida por las siguientes ecuaciones:

$$beneficio = \sum R * M_{\omega}$$

**Ecuación 2.7:** Ecuación para el cálculo del beneficio.

$$enlaces\ totales = \sum R$$

**Ecuación 2.8:** Ecuación para el cálculo de la cantidad de enlaces.

Donde:

R: matriz cuadrada del individuo lineal.

M<sub>ω</sub>: matriz de los pesos potenciales.

Finalmente se obtiene la función de evaluación de los individuos o *fitness* como se muestra en la Ecuación 2.9.

$$G = \varphi * beneficio - \rho * enlaces\ totales * C_{vp}$$

**Ecuación 2.9:** Función de evaluación o *fitness*.

Donde:

C<sub>vp</sub>: Cantidad de vectores propios de la matriz laplaciana.

$\varphi$  y  $\rho$ : parámetros que se calculan a partir de los valores de beneficio y enlaces totales de la matriz de adyacencia inicial respectivamente.

## **Capítulo 3. Resultados y discusión**

En este capítulo se describen cuestiones relacionadas con la implementación del modelo expuesto en el epígrafe 2.6. Además se abordan aspectos considerados en la aplicación que influyen de manera directa en los resultados.

### **3.1 Introducción**

Los sistemas educativos tienen como objetivo principal guiar personalmente a los estudiantes durante su proceso de aprendizaje adaptando los contenidos y la guía de navegación ofrecida, a las características personales y necesidades de cada usuario (Brusilovsky, 1998).

En los entornos de aprendizaje virtuales, es de gran importancia obtener información acerca del comportamiento de los usuarios y entender como los estudiantes utilizan el entorno de aprendizaje y cómo navegan por los materiales educativos. Entender el comportamiento de los usuarios proporciona una información de gran utilidad (Hilbert and Redmiles, 2000). Posibilitado el avance en el diseño y desarrollo de soluciones de *e-learning*.

De este modo y con el fin de mejorar la estructura de enlaces de los cursos de estas plataformas mediante el procesamiento los registros de visita del mismo o *log* y el modelo propuesto, se procede con la aplicación en una plataforma *e-learning* real.

### **3.2 Descripción del sitio web**

Para la aplicación del modelo matemático se utilizó la información de los registros *log* del curso Estructura de Datos II (ver Figura 3.1), perteneciente a la carrera de Ingeniería en informática.

## 2 Tema II: El Tipo de Dato Abstracto Grafo



En matemáticas y ciencias de la computación, un **grafo** (del griego *grafos*: dibujo, imagen) es un conjunto de objetos llamados vértices o nodos unidos por enlaces llamados aristas o arcos, que permiten representar relaciones binarias entre elementos de un conjunto.

-  10. Conferencia 3. Grafos. Definiciones y representación
-  10. Conferencia 3. Presentación electrónica
-  11. Clase Práctica 8 . Aplicación de las definiciones del TDA Grafo
-  12. Clase práctica 9. Solución de problemas utilizando el TDA Grafo, con la representación por lista de adyacencia
-  13. Clase práctica 10. Aplicación de las definiciones del TDA utilizando la representación por matriz de adyacencia.
-  14. Clase práctica 11. Solución de problemas utilizando el TDA Grafo, con la representación por matriz de adyacencia
-  15. Conferencia 4. Recorridos para grafos
-  15. Conferencia 4. Recorridos para grafos
-  16. Clase práctica 12. Implementación algoritmo BPA. Lista de adyacencia
-  17. Clase práctica 13. Implementación algoritmo BPA. Matriz de adyacencia
-  18. Clase Practica 14. Implementacion BPP - Lista de adyacencia

**Figura 3.1:** Imagen del curso Estructura de Datos II.

Fuente: Plataforma Moodle

Este curso se encuentra publicado en la plataforma Moodle (*Modular Object Oriented Developemental Learning Environment*) en la Universidad Agraria de la Habana, la dirección principal del sitio es <http://evea.unah.edu.cu/>.

Con el objetivo de aplicar el algoritmo se tomó una muestra de las 9 primeras páginas visitadas del curso, las cuales se codificaron de la siguiente manera:

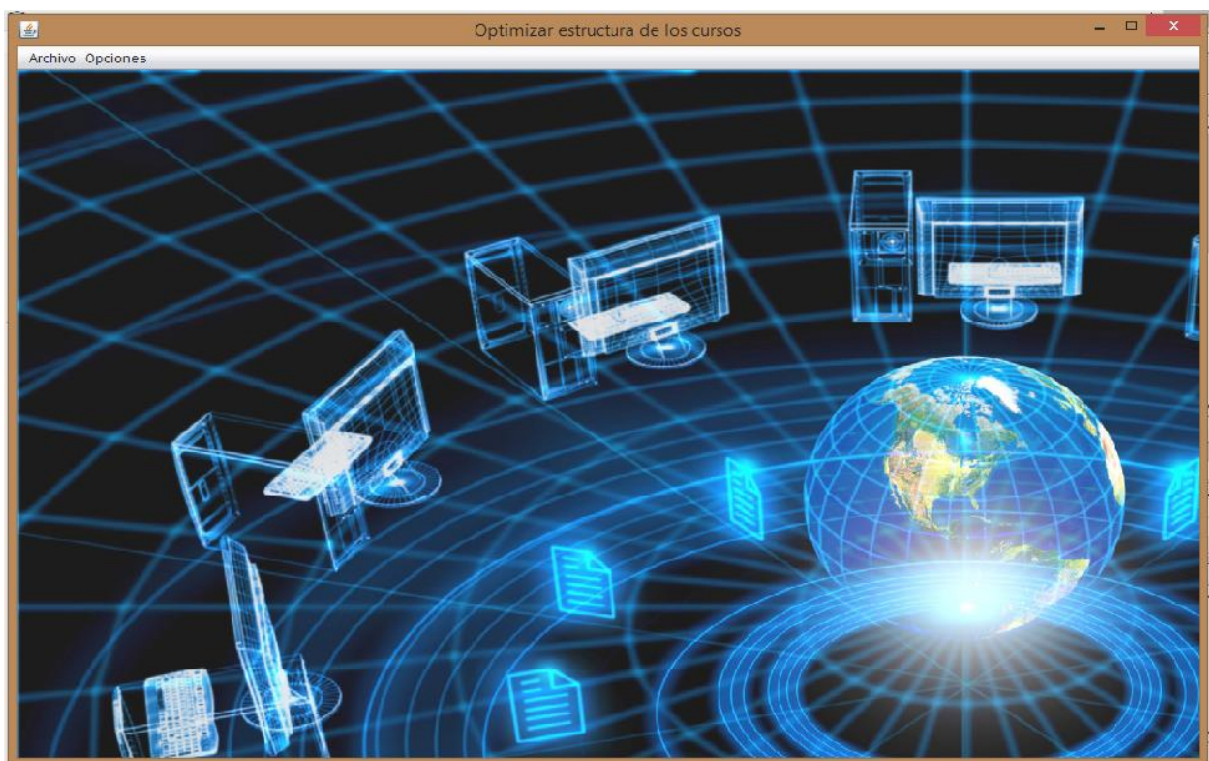
1. Home
2. Árboles binarios
3. Conferencia 2. Árboles Binarios de Búsqueda
4. Conferencia 3. Grafos. Definiciones y representación
5. Conferencia 4. Recorridos para grafos
6. Materiales Complementarios/Recorridos para grafos
7. Recursividad
8. Data Structures and Algorithms
9. Clase práctica 5. Aplicación las definiciones del árbol binario de búsqueda.

### 3.3 Aplicación del modelo matemático

En la Figura 3.3 se muestra la relación entre las diferentes etapas para la aplicación del modelo matemático.

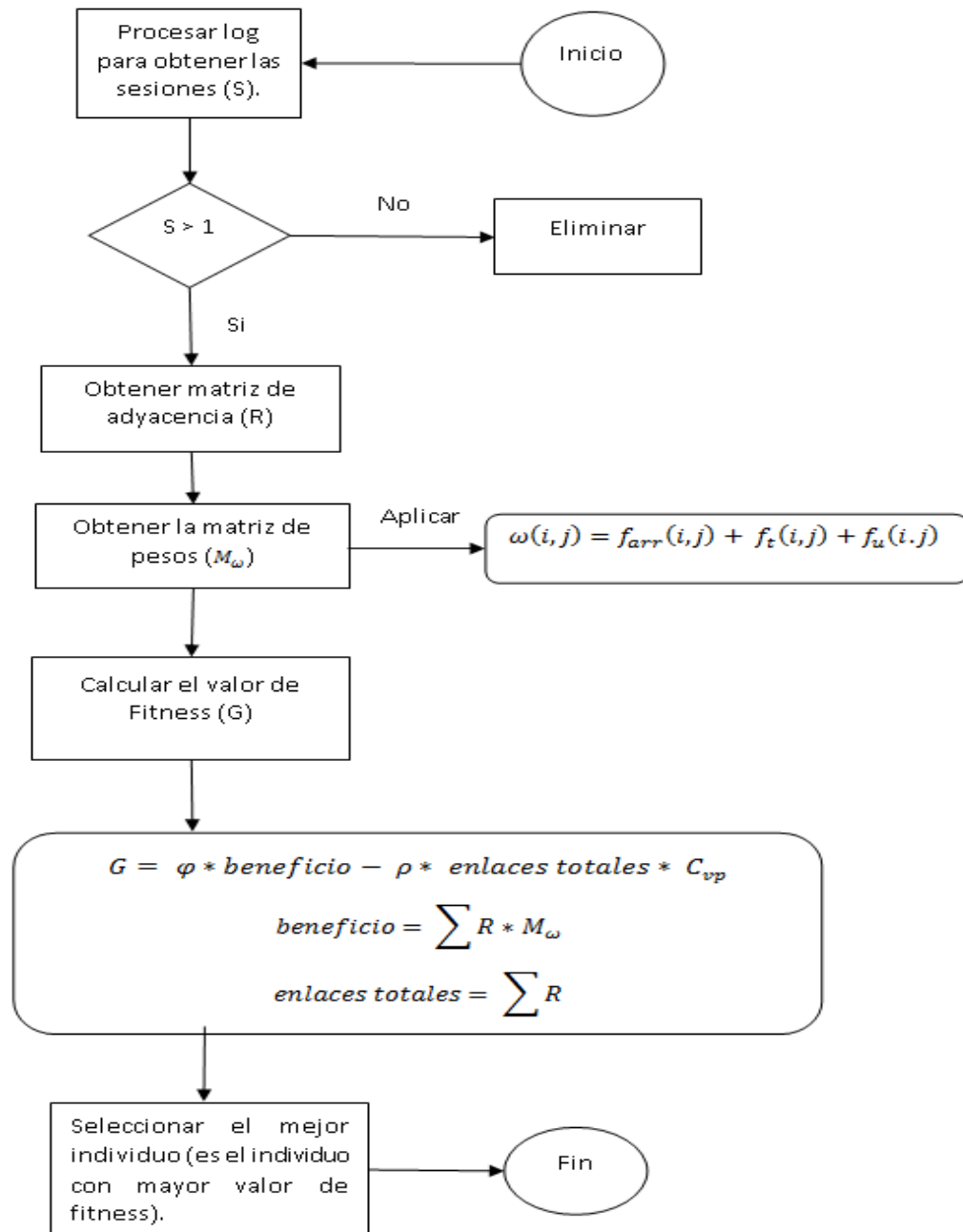
Con el objetivo de realizar todo este proceso automáticamente, se utilizó un programa informático utilizando el lenguaje Java a través del NetBeans IDE 8.0 (Public, 2015)(ver Figura 3.2) que procesa el fichero *log* y extrae los datos necesarios, generando la matriz de adyacencia (ver Tabla 3.1) y su grafo correspondiente (se obtuvo utilizando la librería *jgrapht-0.9.1* (Naveh and Contributors, 2015) desarrollada en java ver Figura 3.4).

El programa permite aplicar el modelo matemático a través de un AG simple.



**Figura 3.2:** Interfaz principal del programa informático.

Fuente: Elaboración propia



**Figura 3.3:** Relación entre las diferentes etapas para la aplicación del modelo matemático.

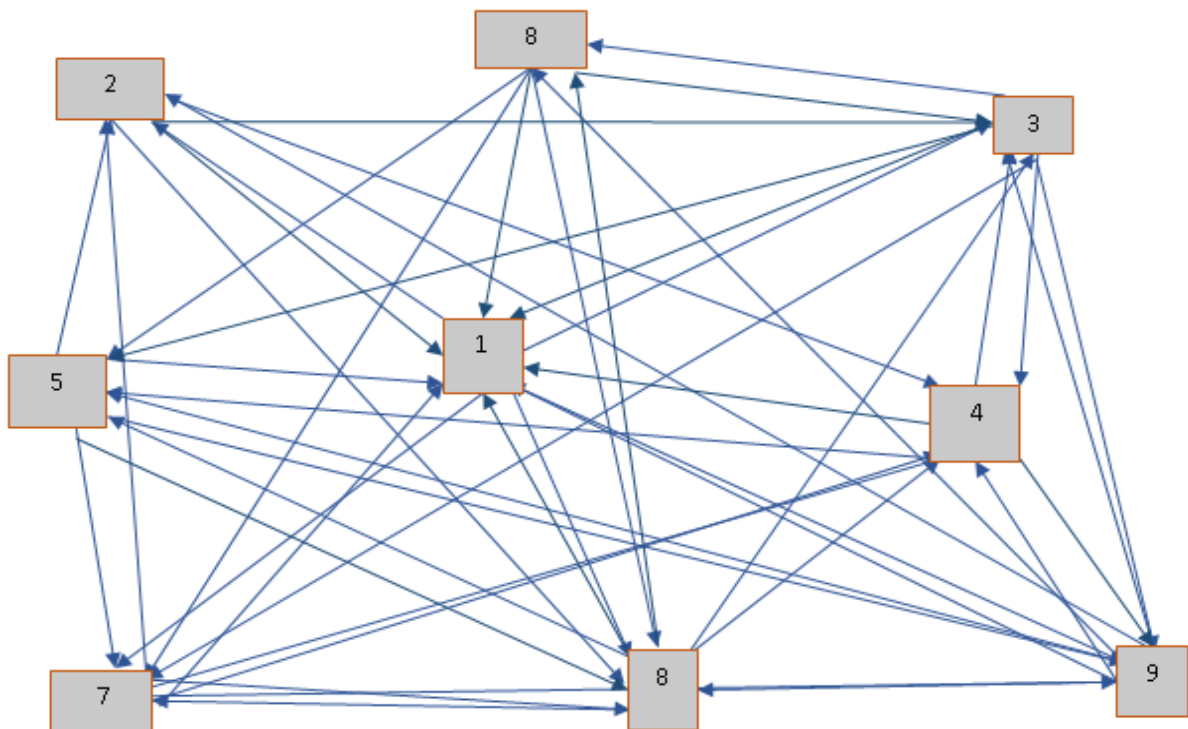
Fuente: Elaboración propia

**Tabla 3.1: Ejemplo de la matriz de adyacencia original.**

| Páginas | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---------|---|---|---|---|---|---|---|---|---|
| 1       | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1 |
| 2       | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 |

|   |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|
| 3 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 |
| 4 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 |
| 5 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 |
| 6 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 |
| 7 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 |
| 8 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 |
| 9 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 |

Fuente: Elaboración propia



**Figura 3.4:** Grafo de la matriz inicial.

Fuente: Grafo resultante de la matriz de adyacencia inicial usando la librería jgrapht-0.9.1.

Una vez obtenida la matriz de adyacencia inicial, se linealiza la matriz concatenando las filas en un solo gran vector. Después se realizan una serie de corridas de prueba a modo de análisis de sensibilidad de los parámetros a utilizar. El objetivo es determinar la combinación de parámetros que otorga un mejor resultado dado por

valores numéricos en la suma final de los pesos existentes. Los parámetros seleccionados para este experimento se pueden observar en la Tabla 3.2.

**Tabla 3.2:** Parámetros del algoritmo AG.

| Parámetro                 | Método    |
|---------------------------|-----------|
| Escala del <i>Fitness</i> | Rankin    |
| Selección                 | Ruleta    |
| Fracción de Cruce         | 0,8       |
| Cruce                     | Disperso  |
| Mutación                  | Gaussiana |

Fuente: Elaboración propia

La escala del *fitness* se realiza mediante el método *Rankin*, el cual clasifica los puntajes de cada individuo basado en un ranking global de ellos. Una selección vía ruleta, simula una rueda con el área de cada segmento proporcional al *fitness*. La fracción de cruce por otro lado, especifica la fracción de la próxima generación que será producida vía cruce. El proceso de cruce es del tipo disperso, es decir, crea un vector binario el cual selecciona los genes respectivos del padre si su valor es 1, y los genes de la madre si su valor es 0, combinando los genes para formar el hijo. Finalmente, la mutación gaussiana cambia un número aleatorio a cada vector de entrada, cuyo gen afectado es tomado desde una distribución gaussiana centrada en cero.

Los otros operadores claves en la ejecución del algoritmo, tienen que ver con los criterios de detención, que son aquellos que determinan su interacción al momento de encontrar un buen ajuste dado por ciertos parámetros. Ellos corresponden a un número máximo de generaciones, a un factor de tolerancia que indique el grado de diferencia entre el mejor individuo detectado entre las distintas generaciones y a un conteo de las mismas. El primero se fija en 10.000, el segundo se toma como nulo, y la cantidad de generaciones en que el mejor individuo detectado se mantenga (no hay otro individuo que tenga mejor *fitness*) se fija en 1000. Con éste último valor se está diciendo que si el mejor individuo se mantiene como mejor individuo por sucesivas generaciones, su calidad de óptimo es cada vez mayor. Estos parámetros pueden variar y están en dependencia de cada problema, es decir estos valores

varían de un problema a otro hasta encontrar los que mejor se ajustan (los que mejor resultado devuelvan).

La configuración de este modelo requiere ajustar los valores de las constantes  $\varphi$  y  $\rho$ , lo que se hace calculando sus valores mediante la Ecuación 2.6 y la Ecuación 2.7 respectivamente de la matriz inicial. Así, se busca hacer comparable ambos términos, de modo que el modelo no busque los extremos: minimizar la cantidad de enlaces o bien maximizarla de manera independiente; lo que resultaría en la matriz nula o unitaria respectivamente. Así, los cálculos en la matriz inicial resultaron ser una cantidad de enlaces iniciales de 50 y un beneficio de 1,02. Por lo que se estudia empíricamente estos valores disminuyendo o manteniendo el valor inicial. Finalmente se fijó  $\varphi= 50$  y  $\rho=0,3$  al proporcionar una mejor matriz resultante y adecuada convergencia.

La Tabla 3.3, indica los resultados del mejor individuo resultante del modelo (es el individuo que mayor valor de *fitness* después de 1000 generaciones (condición de parada del AG) sin mejorar este valor). La nueva adyacencia contempla 45 enlaces. El valor de la función objetivo fue incrementado y se redujo la cantidad de enlaces con respecto a la matriz inicial.

**Tabla 3.3:** Ejemplo de la tabla de adyacencia final.

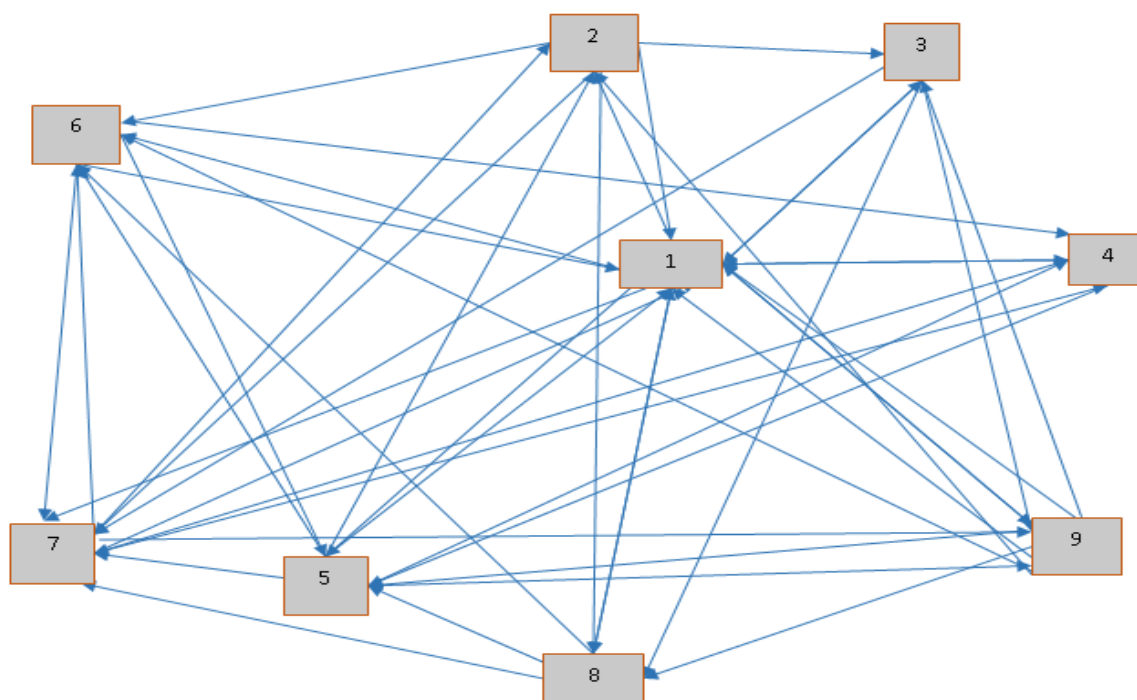
| Páginas | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---------|---|---|---|---|---|---|---|---|---|
| 1       | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2       | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 |
| 3       | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| 4       | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| 5       | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 |
| 6       | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 |
| 7       | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 |
| 8       | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 0 |

|   |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|
| 9 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 |
|---|---|---|---|---|---|---|---|---|---|

Fuente: Elaboración propia

El grafo final que representa al individuo con mayor valor de la función objetivo o *fitness* (mejor individuo) después de cumplirse las condiciones de parada, y por ende, la nueva estructura web propuesta por el modelo, puede observarse en la Figura 3.5.

Resaltan por su importancia (valor del peso) los enlaces desde y hacia la página principal Home y los enlaces entre las páginas 2 (Árboles binarios) y 3 (Árboles Binarios de Búsqueda).



**Figura 3.5:** El grafo final que representa el mejor individuo.

Fuente: Grafo resultante de la matriz de adyacencia final usando la librería *jgrapht-0.9.1*.

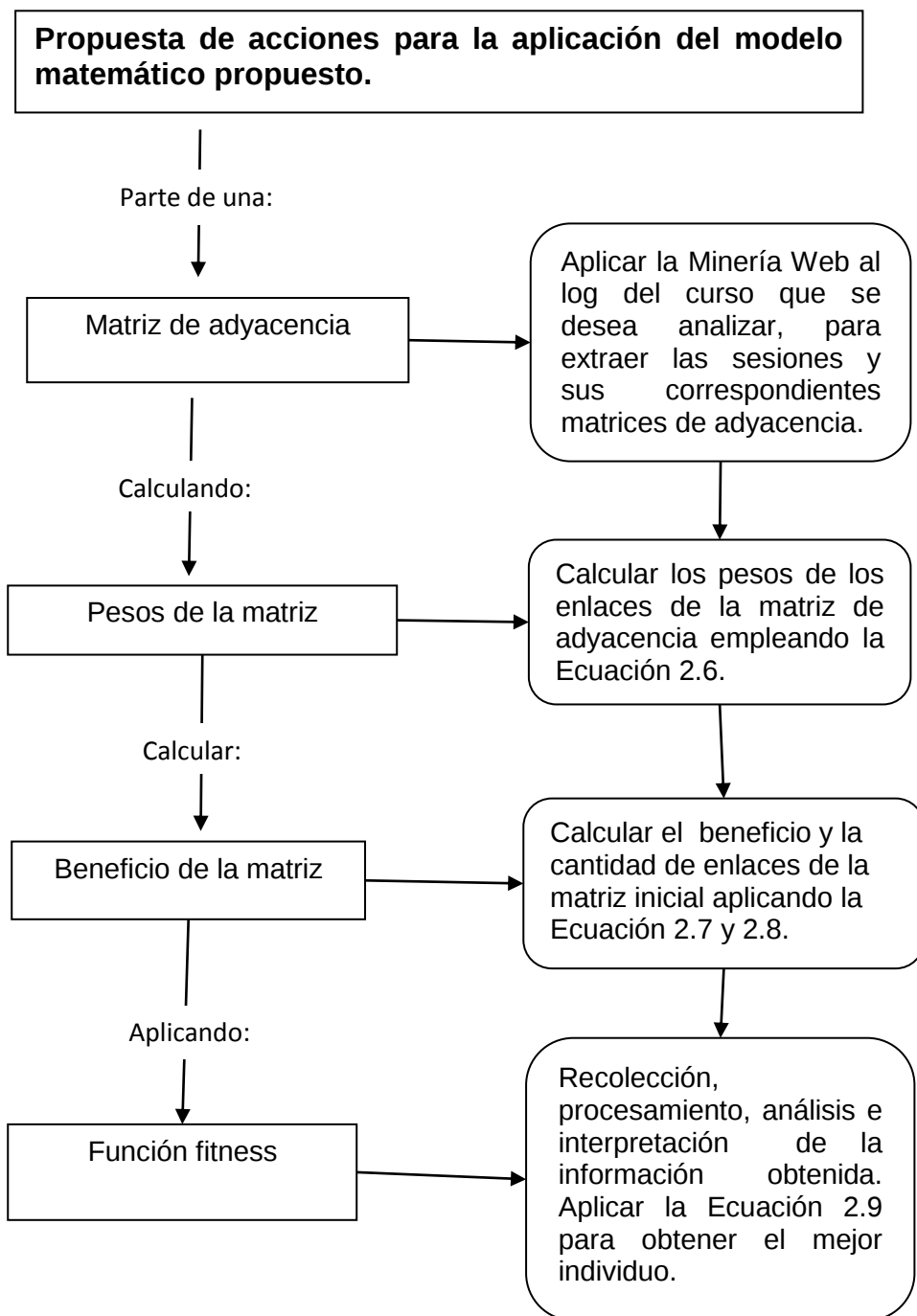
La nueva estructura propuesta permite la optimización de los cursos de las plataformas e-learning a través de la información almacenada en los ficheros *log*, lo cual complementa los estudios realizados por los autores: (Zaiane and Luo, 2001), (Romero et al., 2009a), (Lorena, 2010), (Aponte et al., 2012), (Ballesteros et al., 2013) y (Chakurkar and Adiga, 2015).

### **3.4 Propuesta de acciones para la aplicación del modelo matemático propuesto.**

Integrando todos los resultados obtenidos en esta investigación, se realiza la siguiente propuesta de acciones para optimizar y personalizar los cursos de las plataformas *e-learning* mediante la información almacenada en los *log* de estas plataformas.

En la Figura 3.5 se muestra la propuesta de acciones para la aplicación del modelo matemático, en esta se incluyen los siguientes pasos:

1. Aplicar la Minería Web al *log* del curso que se desea analizar, para extraer las sesiones y sus correspondientes matrices de adyacencia.
2. Calcular los pesos de los enlaces de la matriz de adyacencia empleando la Ecuación 2.6.
3. Calcular el beneficio y la cantidad de enlaces de la matriz inicial aplicando la Ecuación 2.7 y 2.8.
4. Recolección, procesamiento, análisis e interpretación de la información obtenida. Aplicar la Ecuación 2.9 para obtener el mejor individuo.



**Figura 3.5:** Propuesta de acciones para la aplicación del modelo matemático propuesto.

Fuente: Elaboración propia

## Conclusiones

Se ha realizado un estudio de la modelación matemática, particularizando en el análisis de la Minería de Uso Web con el objetivo de darle solución a las desventajas que se generan en torno del empleo por parte de los alumnos de la plataforma *e-learning*.

Se ha fundamentado y demostrado la importancia de la Minería Web, permitiendo extraer de los archivos *log* de las plataformas *e-learning* la información necesaria para la aplicación del modelo matemático, a partir de la matriz de adyacencia del *log* y la utilidad o peso de cada enlace.

El modelo matemático propuesto permite optimizar y personalizar la estructura de los cursos en las plataformas *e-learning*, considerando el supuesto de que a mayor tiempo de permanencia en una página de la misma sesión, mayor interés en ella, y que la frecuencia de uso de las interrelaciones indica una mayor utilidad de las mismas para acceder hacia la información deseada, se logró encontrar estructuras de navegación más eficientes desde el punto de la navegabilidad y usabilidad, mejorando en concreto la estructura de los cursos de las plataformas *e-learning* en la UNAH.

## **Recomendaciones**

El trabajo adicional como consecuencia del aquí expuesto debe enfocarse en primera instancia a la implementación concreta de los cambios en la estructura de los cursos, estudiando el uso real de los usuarios en un considerable período de tiempo, con porcentajes de mejora, frecuencias de acceso.

Estudiar nuevas restricciones que pueden ser analizadas e incorporadas al modelo, como por ejemplo tener en cuenta el número de enlaces entrantes o salientes desde las páginas.

Incorporar métodos que permitan el estudio más dinámico, de modo que se pueda alcanzar una configuración del sitio web del tipo adaptativa, personalizada y que contenga sugerencias mientras se navega.

Proponer la aplicación, en los centros de educación superior para de esta forma obtener retroalimentación a partir de las sugerencias que realicen los usuarios de la herramienta e ir incorporando al modelo las solicitudes realizadas por estos.

Incorporar los resultados y experiencias de esta investigación a la enseñanza de pregrado y postgrado, así como a la actividad que se desarrolla en el marco de la línea de investigación sobre la modelación matemática, de forma tal que los actuales y futuros profesionales se apropien de los fundamentos de este enfoque en su quehacer profesional y pueda servir para otras investigaciones que se realicen en este campo.

## REFERENCIAS BIBLIOGRÁFICAS

- AGRAWAL, R., IMIELIŃSKI, T. & SWAMI, A. Mining association rules between sets of items in large databases. *ACM SIGMOD Record*, 1993. ACM, 207-216.
- AGRAWAL, R. & SRIKANT, R. Mining sequential patterns. *Data Engineering*, 1995. Proceedings of the Eleventh International Conference on, 1995. IEEE, 3-14.
- AMBROSANO, G. M. 1990. Models and simulation and their application to livestock breeding. 9-23.
- APONTE, F. A., HOYOS, J. G. & MONSALVE, J. A. 2012. Minería de usabilidad aplicada a plataformas virtuales de aprendizaje.
- ARABIE, P., HUBERT, L. J. & DE SOETE, G. 1996. *Clustering and classification*, World Scientific.
- BALLESTEROS, A., SÁNCHEZ-GUZMÁN, D. & GARCÍA, R. S. 2013. Minería de datos educativa: Una herramienta para la investigación de patrones de aprendizaje sobre un contexto educativo.
- BANDYOPADHYAY, S. & PAL, S. K. 2007. *Classification and learning using genetic algorithms: applications in bioinformatics and web intelligence*, Springer Science & Business Media.
- BASSANEZI, R. & BIEMBENGUT, M. 1997. Modelación matemática: Una antigua forma de investigación un nuevo método de enseñanza. *Números. Revista de didáctica de las matemáticas*, 32, 13-25.
- BROWN, J. 1987. "Global sustainability: toward definition".
- BRUSILOVSKY, P. Adaptive educational systems on the world-wide-web: A review of available technologies. *Proceedings of Workshop "WWW-Based Tutoring"* at 4th International Conference on Intelligent Tutoring Systems (ITS'98), San Antonio, TX, 1998.
- BUSLENKO, N. P. 1988. Modelación de sistemas complejos. 400.
- CARDONA, J. J. C. 2010. Los aprendizajes colaborativos como estrategia para los procesos de construcción de conocimiento. *Revista Educación y desarrollo social*, 4, 87-103.
- CLARK, R. C. & MAYER, R. E. 2011. *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*, John Wiley & Sons.
- CORSO, C. L. & ALFARO, S. L. 2007. MINERIA DE USO WEB: PRESENTACIÓN DE CASO PRÁCTICO PARA EL MEJORAMIENTO DE LA EDUCACIÓN, BASADA EN EL USO DE PLATAFORMAS VIRTUALES.
- CRESTANI, F. & PASI, G. 2013. *Soft Computing in Information Retrieval: techniques and applications*, Physica.
- CHAKRABARTI, S., DOM, B. E., KUMAR, S. R., RAGHAVAN, P., RAJAGOPALAN, S., TOMKINS, A., GIBSON, D. & KLEINBERG, J. 1999. Mining the Web's link structure. *Computer*, 32, 60-67.
- CHAKURKAR, M. & ADIGA, D. 2015. A Web Mining Approach for Personalized E-Learning System. (*IJACSA*) *International Journal of Advanced Computer Science and Applications*, 5.
- DARWIN, C. & DE ZULUETA, A. 2009. *El origen de las especies por medio de la selección natural*, Editorial CSIC-CSIC Press.
- DORRONSORO, B. 2007. *Diseño e implementación de algoritmos genéticos celulares para problemas complejos*. PhD thesis, University of Málaga.
- ETZIONI, O. 1996. The World-Wide Web: quagmire or gold mine? *Communications of the ACM*, 39, 65-68.
- FERNÁNDEZ, P. 2009. El secreto de Google y el Álgebra lineal. *Boletín SEMA*.
- FISCHER, K. 1995. Ecological aspects of meat production. *Mitteilungsblatt der Bundesanstalt fuer Fleischforschung, Kulmbach (Germany)*.

- FREYBERGER, J. E., HEFFERNAN, N. & RUIZ, C. 2004. *Using association rules to guide a search for best fitting transfer models of student learning*. Citeseer.
- GALLEGO, M. L., GONZÁLEZ, A. T. & NUÑEZ, T. G. 2002. Algoritmos genéticos para la resolución de problemas de Programación por Metas Entera. Aplicación a la Economía de la Educación. Rafael Caballero Fernández. Julián Molina Luque.
- GARCÍA, J. P. S. & MAHEUT, J. 2012. Modelos y Métodos de Investigación de Operaciones. Procedimientos para Pensar.
- GARCÍA, S. E., ROMERO, C., VENTURA, S. & DE-CASTRO-LOZANO, C. 2008. Sistema recomendador colaborativo usando minería de datos distribuida para la mejora continua de cursos e-learning. *IEEE-RITA*, 3, 19-30.
- GESTAL, M. 2008. Introducción a los algoritmos genéticos, Notas de Investigación, Universidad de la Coruña, Departamento de Tecnologías de la Información y las Comunicaciones, 16 p.
- GESTAL, M., RIVERO, D., RABUÑAL, J. R., DORADO, J. & PAZOS, A. 2010. *Introducción a los algoritmos genéticos y la programación genética*, Universidade da Coruña, Servicio de Publicacións.
- GOLBERG, D. E. 1989. Genetic algorithms in search, optimization, and machine learning. *Addison wesley*, 1989.
- GONÇALVES, S., CARNEIRO, D., FDEZ-RIVEROLA, F. & NOVAIS, P. 2015. Análisis del comportamiento en entornos de e-learning. *Revista de Educación Mediática y TIC*.
- GORDON, M. 1988. Probabilistic and genetic algorithms in document retrieval. *Communications of the ACM*, 31, 1208-1218.
- HA, S. H., BAE, S. M. & PARK, S. C. Web mining for distance education. Management of Innovation and Technology, 2000. ICMIT 2000. Proceedings of the 2000 IEEE International Conference on, 2000. IEEE, 715-719.
- HÄMÄLÄINEN, W., SUHONEN, J., SUTINEN, E. & TOIVONEN, H. Data mining in personalizing distance education courses. World conference on open learning and distance education, 2004. 16.
- HEUGHES, V. 2007. MineríaWeb de Uso y Perfiles de Usuario: Aplicaciones con Lógica Difusa.
- HILBERT, D. M. & REDMILES, D. F. 2000. Extracting usability information from user interface events. *ACM Computing Surveys (CSUR)*, 32, 384-421.
- HOLLAND, J. H. 1992. *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*, MIT press.
- HORTON, W. K. 2000. Designing web-based training: how to teach anyone anything anywhere anytime. *New York*.
- JERALDO, O. L. R. 2001. LA MATRIZ LAPLACIANA Y LA CONECTIVIDAD ALGEBRAICA DE UN GRAFO.
- KARGUPTA, H. & BUESCHER, K. The gene expression messy genetic algorithm for financial applications. Computational Intelligence for Financial Engineering, 1996., Proceedings of the IEEE/IAFE 1996 Conference on, 1996. IEEE, 155-161.
- KIM, E., KIM, W. & LEE, Y. 2003. Combination of multiple classifiers for the customer's purchase behavior prediction. *Decision Support Systems*, 34, 167-175.
- KOZA, J. R. 1992. *Genetic programming: on the programming of computers by means of natural selection*, MIT press.
- LIN, W.-L. & LIU, Y.-Z. A Novel Website Structure Optimization Model for More Effective Web Navigation. *wkdd*, 2008. 36-41.
- LORENA, I. C. C. 2010. Minería de uso Web: Presentación de Caso práctico para el mejoramiento de la educación, basada en el uso de plataformas virtuales.
- LUAN, J. 2002. Data Mining and Knowledge Management in Higher Education-Potential Applications.

- LUO, C.-H. & RUDY, Y. 1994. A dynamic model of the cardiac ventricular action potential. II. Afterdepolarizations, triggered activity, and potentiation. *Circulation Research*, 74, 1097-1113.
- MARGHNY, M. & ALI, A. Web mining based on genetic algorithm. AIML 05 Conference, 2005. 19-21.
- MARQUARDT, C. G., BECKER, K. & RUIZ, D. D. A pre-processing tool for Web usage mining in the distance education domain. Database Engineering and Applications Symposium, 2004. IDEAS'04. Proceedings. International, 2004. IEEE, 78-87.
- MENASCÉ, D. A., ALMEIDA, V. A., FONSECA, R. & MENDES, M. A. A methodology for workload characterization of e-commerce sites. Proceedings of the 1st ACM conference on Electronic commerce, 1999. ACM, 119-128.
- MINAEI-BIDGOLI, B., KASHY, D. A., KORTEMEYER, G. & PUNCH, W. F. Predicting student performance: an application of data mining methods with an educational web-based system. Frontiers in education, 2003. FIE 2003 33rd annual, 2003. IEEE, T2A-13.
- MINAEI-BIDGOLI, B. & PUNCH, W. F. Using genetic algorithms for data mining optimization in an educational web-based system. Genetic and Evolutionary Computation—GECCO 2003, 2003. Springer, 2252-2263.
- MOBASHER, B., COOLEY, R. & SRIVASTAVA, J. 2000. Automatic personalization based on web usage mining. *Communications of the ACM*, 43, 142-151.
- MOR, E. & MINGUILLÓN, J. E-learning personalization based on itineraries and long-term navigational behavior. Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters, 2004. ACM, 264-265.
- NAVEH, B. & CONTRIBUTORS. 2015. *JGraphT* [Online]. Available: <http://jgrapht.org/> [Accessed 8 de octubre 2015].
- ORTEGA-PRIEGO, J.-L. & AGUILLO, I. F. 2009. Minería del uso de webs. *El profesional de la información*, 18, 20-26.
- ORTIZ, J. 2000. *Modelación y simulación matemática del reciclaje de N, P y K en sistemas de pastoreo vacuno en Cuba*. Tesis Dr. Cs. Instituto de Ciencia Animal. La Habana, Cuba.
- PAHL, C. 2004. Data mining technology for the evaluation of learning content interaction. *International Journal on E-learning*, 3, 47.
- PÉREZ, E. 2010. Guía para recién llegados a los ALGORITMOS GENÉTICOS.
- PICAROUGNE, F., MONMARCHÉ, N., OLIVER, A. & VENTURINI, G. Web mining with a genetic algorithm. Eleventh International World Wide Web Conference, 2002. 131-69.
- PUBLIC, G. G. 2015. *Netbeans IDE 8.0* [Online]. [Accessed 6 de enero 2015].
- ROMERO, C., VENTURA, S. & DE BRA, P. 2004. Knowledge discovery with genetic programming for providing feedback to courseware authors. *User Modeling and User-Adapted Interaction*, 14, 425-464.
- ROMERO, C., VENTURA, S., ZAFRA, A. & DE BRA, P. 2009a. Applying Web usage mining for personalizing hyperlinks in Web-based adaptive educational systems.
- ROMERO, F. & PARADA, V. Una máquina genética paralela para resolución de problemas de optimización combinatoria. VIII Congreso Chileno de Investigación Operativa, 2009.
- ROMERO, M. C., VENTURA, S. S. & HERVÁS, M. C. 2009b. Estado actual de la aplicación de la minería de datos a los sistemas de enseñanza basada en web.
- ROOSE, L. 2000. Probabilidades y estadística para ingenieros y científicos.
- ROSENBERG, M. J. 2001. *E-learning: Strategies for delivering knowledge in the digital age*, McGraw-Hill New York.
- ROSENBERG, M. J. 2005. The Future of e-Learning. ASTD's Learning Circuits Webzine.
- SAID, L. B., BOURON, T. & DROGOUL, A. Agent-based interaction analysis of consumer behavior. Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 1, 2002. ACM, 184-190.

- SALOÑA, A., CANO, C. J. B., GÓMEZ, Ó. G. & CAMPOS, I. S. 2001. Informe sobre el estado de la teleeducación en España. *Asociación de Usuarios de Internet. Madrid*, 2.
- SANJUÁN, O., TORRES, E., CASTÁN, H., GONZALEZ, R., PELAYO, C. & RODRIGUEZ, L. 2009. Viabilidad de la aplicación de Sistemas de Recomendación a entornos de e-learning. *Universidad de Oviedo, España*.
- SHEPTULIN, A. P. 1984. *El método dialéctico de conocimiento*, Editorial Cartago.
- SHOKOUHI, M., CHUBAK, P. & RAEESY, Z. Enhancing focused crawling with genetic algorithms. *Information Technology: Coding and Computing*, 2005. ITCC 2005. International Conference on, 2005. IEEE, 503-508.
- SILVIO, J. 2004. Tendencias de la educación superior virtual en América Latina y el Caribe. *La educación superior virtual en América Latina y el Caribe*, 15-39.
- SIVANANDAM, S. & DEEPA, S. 2007. *Introduction to genetic algorithms*, Springer Science & Business Media.
- SOLOMON, E. 2001. *Biología*. 5ª Edición Editorial Mc Graw-Hill Interamericana. México.
- SRIVASTAVA, J., COOLEY, R., DESHPANDE, M. & TAN, P.-N. 2000. Web usage mining: Discovery and applications of usage patterns from web data. *ACM SIGKDD Explorations Newsletter*, 1, 12-23.
- SU, J.-M., TSENG, S.-S., WANG, W., WENG, J.-F., YANG, J.-T. D. & TSAI, W.-N. 2006. Learning portfolio analysis and mining for SCORM compliant environment. *Educational Technology & Society*, 9, 262-275.
- TALAVERA, L. & GAUDIOSO, E. Mining student data to characterize similar behavior groups in unstructured collaboration spaces. *Workshop on artificial intelligence in CSCL. 16th European conference on artificial intelligence*, 2004. 17-23.
- TANG, T. Y. & MCCALLA, G. Smart recommendation for an evolving e-learning system. *Workshop on Technologies for Electronic Documents for Supporting Learning, AIED*, 2003.
- TORRES, V. 1995. Contribución a la modelación de un área de pastoreo. 5to Seminario Científico Internacional. *XXX Aniversario Instituto de Ciencia Animal, La Habana, Cuba*.
- VALENCIA, E. Optimización mediante algoritmos genéticos. *Anales del Instituto de Ingenieros de Chile*, 1997. 83-92.
- VÁZQUEZ, Y. A. 2008. *Sistema de Objetos de Aprendizaje para la enseñanzade la Matemática Básica en la carrera Ingeniería Agronómica*. Master, Universidad Agraria de la Habana.
- VELÁSQUEZ, J. D. & PALADE, V. Adaptive web sites: A knowledge extraction from web data approach. *Proceedings of the 2008 conference on Adaptive Web Sites: A Knowledge Extraction from Web Data Approach*, 2008. IOS Press, 1-272.
- VELÁSQUEZ, J. D., YASUDA, H. & WEBER, R. 2004. A new similarity measure to understand visitor behavior in a web site. *IEICE TRANSACTIONS on Information and Systems*, 87, 389-396.
- VILAR, P. 2010. Designing the User Interface: Strategies for Effective Human-Computer Interaction. *Journal of the American Society for Information Science and Technology*, 61, 1073-1074.
- WAND, Y., MONARCHI, D. E., PARSONS, J. & WOO, C. C. 1995. Theoretical foundations for conceptual modelling in information systems development. *Decision Support Systems*, 15, 285-304.
- WANG, F.-H. 2002. On analysis and modeling of student browsing behavior in web-based asynchronous learning environments. *Advances in Web-Based Learning*. Springer.
- YANG, H. H. 2009. *Handbook of Research on Practices and Outcomes in E-Learning: Issues and Trends: Issues and Trends*, IGI Global.
- ZAÍANE, O. R. Building a recommender agent for e-learning systems. *Computers in Education*, 2002. Proceedings. International Conference on, 2002. IEEE, 55-59.

ZAÍANE, O. R. & LUO, J. Web usage mining for a better web-based learning environment. Proceedings of conference on advanced technology for education, 2001. 60-64.