

Universidad Central “Marta Abreu” de Las Villas
Facultad de Matemática, Física y Computación



Trabajo para optar por título de
Máster en Ciencia de la Computación

TÍTULO

Sistema recomendador de instalaciones turísticas, usando un sistema
experto híbrido.

Autor

Lic. Danilo Valdés Ramírez

Tutora

Dra. María Matilde García Lorenzo

2013

AGRADECIMIENTOS

El trabajo aquí presentado ha sido resultado de la colaboración de un grupo de compañeros a quienes agradezco infinitamente su ayuda. En primer lugar a la Dra. María Matilde García Lorenzo, tutora de este trabajo, quien ha colaborado constantemente en su desarrollo. A los Ingenieros Igor Shólojov, Reyniel Ramírez y Roberto Jiménez por su participación permanente en el proyecto. A los compañeros de INFOTUR y a los especialistas de las agencias Thomas Cook, Nolitur, Transat y Cubatur en Jardines del Rey.

RESUMEN

En el Caribe actual la competitividad en el mercado del turismo es muy fuerte. Sin embargo, algunas empresas cubanas no acumulan tantos años de experiencia ni tienen presencia en suficientes mercados internacionales. De ahí la necesidad de utilizar métodos efectivos para la toma de decisiones, vinculados a las Tecnologías de la Información y las Comunicaciones (TIC) y lograr la presencia de las ofertas de los destinos cubanos en todo el mundo.

En este trabajo se presenta un sistema capaz de ofrecer al cliente una lista de instalaciones hoteleras, hecha a partir de las características y preferencias del usuario. Para ello se tienen en cuenta las decisiones tomadas por los clientes anteriores con características afines. Se valoran los sistemas similares existentes a nivel mundial, conocidos como sistemas recomendadores. La tecnología empleada para desarrollar el sistema recomendador para el turismo cubano es el Razonamiento Basado en Casos (RBC). Se analiza el problema del diseño de la base de conocimientos de manera que se alcance la rapidez necesaria en el sistema orientado a la WEB. Para dar solución a esta dificultad se diseña un sistema híbrido que combina RBC con árboles de decisión, permitiendo la organización jerárquica de los casos.

ABSTRACT

In the current Caribbean competitiveness in the tourism market is very strong. However, some Cuban companies do not accumulate many years of experience and have a presence in many international markets. Hence the need for effective methods for decision making related to Information Technology and Communications and ensures the presence of the offers of the Cuban destinations worldwide.

This paper presents a system capable of giving the customer a list of holiday destinations obtained from the characteristics and preferences of the user, using the decisions taken by previous customers with similar characteristics. In the paper it's also valorized the similar system identified in the world as recommender system. The technology used to develop the recommender system for Cuban tourism is the Case Based Reasoning (CBR). Also discusses the problem of design of the knowledge base, so as to reach the required speed in the web-oriented system. To solve this problem we propose a hybrid system that combines CBR with decision trees, allowing hierarchical organization of cases.

Contenido

INTRODUCCIÓN	5
DESARROLLO.....	11
Capítulo 1. Estado del arte sobre el uso de la Inteligencia Artificial en la WEB y el turismo.	11
1.1 El uso de las técnicas de la IA en la WEB y en el turismo.	11
1.1.1 Técnicas de la IA usadas en sistemas WEB.	12
1.1.2 Técnicas de la IA más usadas en el turismo.	15
1.2 Estado histórico y actual de los sistemas expertos orientados a la WEB y en el turismo. Análisis histórico lógico con tendencias actuales.	17
1.3 Situación del uso de los sistemas expertos orientados a la WEB en el turismo en Cuba.	19
1.4 Técnicas de la IA que permiten dar solución al problema.	22
1.4.1 Técnicas de Optimización	22
1.4.2 Técnicas de filtrado de datos para la solución de problemas similares.	22
1.4.3 Técnicas de Sistemas Multiagentes (SMA) similares.	23
1.4.4 Técnicas de Sistemas Basados en el Conocimiento (SBC).	24
1.5 Conclusiones del capítulo	25
Capítulo 2 Descripción del problema de la selección de instalaciones hoteleras en Cuba.	26
2.1 Descripción del problema de la selección de instalaciones hoteleras a partir de las características y gustos de los clientes.	26
2.2 Necesidad de crear el sistema de selección de instalaciones hoteleras cubanas orientado a la WEB.	28
2.3 Planteamiento del negocio actual.	31
2.4 Selección y justificación de los paradigmas de la IA usados.	32
2.5 Diseño de la base de conocimientos.	32
2.5.1 Estructura jerárquica de la base de conocimientos.	33
2.5.2 Cálculo de la importancia de los rasgos.	36
2.5.3 Criterios de comparación por rasgos y función de similitud.	38
2.6 Conclusiones del capítulo.	40
Capítulo 3 Sistema basado en casos, recomendador para instalaciones turísticas cubanas.	41
3.1 Elaboración de los módulos de recuperación, adaptación y aprendizaje.	41
3.1.1 Recuperación y adaptación en el sistema basado en casos.	41
3.1.2 Aprendizaje en el sistema basado en casos.	42
3.2 Implementación del sistema en el portal Cuba.travel	44
3.2.1 Herramientas y tecnologías usadas.	44
3.2.2 Descripción estructural del sistema.	45
3.2.3 Forma de interactuar con el sistema recomendador.	49
3.3 Validación y verificación del sistema.	51
3.4 Conclusiones del capítulo	55
CONCLUSIONES	56
RECOMENDACIONES	57
REFERENCIAS BIBLIOGRÁFICAS	58
Anexos.	68

INTRODUCCIÓN

Cuba surgió como destino turístico desde antes del triunfo revolucionario, cuando todavía era dominada por los Estados Unidos de América. Alrededor de los años 50 se crearon los primeros hoteles ubicados en La Habana, entre ellos estaban el Havana Riviera, Capri y el Havana Hilton. Todos estaban llenos de casinos y los principales clientes eran norteamericanos, los cuales querían desarrollar un turismo poco saludable en la Isla (1). Cuba formaba parte en ese entonces del circuito de las cuatro ciudades más importantes del mundo en el turismo: París, Londres, New York, La Habana.

Con el triunfo revolucionario se comenzaron a desarrollar otras áreas de la economía y el turismo fue disminuyendo su importancia en el país hasta llegar a niveles pequeños. A principio de la década del 90, durante el período especial, se comenzó a desarrollar nuevamente el turismo internacional. Esta vez se incursiona en modalidades que beneficiaban más a las personas, cambiando radicalmente de 1959 a la fecha. Se pasa de un turismo de casinos y burdeles, que beneficiaba más a los propietarios de estos lugares, a convertirse en uno de los renglones principales de la economía nacional. Aunque el turismo internacional siempre trae consigo riesgos, su aporte a la economía resulta decisivo.

Desde los inicios de los años 90, año en que comenzó a revitalizarse el turismo en Cuba, hasta la actualidad, ha existido una gran evolución en las Tecnologías de la Información y las Comunicaciones a nivel mundial. Esto ha favorecido la promoción, permitiendo su desarrollo hasta niveles nunca antes alcanzados.

Las ciencias de la computación y la informática han brindado soporte al desarrollo de la tecnología y han aportado soluciones novedosas al tratamiento de la información. De manera general, el uso de estas en las organizaciones empresariales y en la toma de decisiones ha permitido alcanzar nuevos horizontes en la competitividad internacional en muchas áreas y especialmente en el turismo. En particular, la aparición de sistemas informáticos WEB que posibilitan el control de la información a distancia, ha sido un pilar determinante para estos logros.

La programación WEB cada vez alcanza niveles superiores. Se pueden desarrollar aplicaciones WEB, que son ejecutadas remotamente desde un servidor, tan fuertes en los cálculos como las aplicaciones llamadas de escritorio, que son ejecutadas localmente. Esto ha dado lugar a sistemas informáticos complejos para las organizaciones empresariales orientados a la WEB. Con ellos se incrementan las posibilidades de acceso a la información desde cualquier lugar. Además, permiten la publicación de la información compartimentada mediante sistemas de autenticación, por lo que son muy usados en la promoción y el manejo de información.

La industria turística mundial ha incorporado paulatinamente las nuevas tecnologías desarrolladas para "acercar" cada vez más a las personas a través de Internet y los dispositivos móviles. En el uso de estas tecnologías ha jugado un papel importante la Inteligencia Artificial (IA) como disciplina dentro de las ciencias de la computación. Los distintos métodos y paradigmas de la IA permiten encontrar soluciones en espacios de búsquedas relativamente grandes en los que los métodos secuenciales requieren un tiempo excesivamente grande.

Artículos publicados por revistas especializadas en el turismo ponen de manifiesto la necesidad del uso de técnicas computacionales y de las comunicaciones en este sector. En el trabajo citado en (2) el Centro de Estudios del Turismo de la Universidad de Oriente presenta un trabajo periodístico sobre la cadena Barceló Hotels. Se expone un sistema desarrollado para apoyar la promoción y las reservaciones a través de la WEB. El Dr. Martín Fernández, decano de la Facultad de Turismo de la Universidad de La Habana en el año 2008, plantea en una conferencia magistral, la necesidad de utilizar más los sistemas informáticos para la promoción y el control de la planta hotelera cubana (3). Expone que, aun cuando se encuentra en desarrollo, es insuficiente el papel que juega la informática dentro del turismo en Cuba. La Revista de Investigación en Turismo y Desarrollo Local presenta en (4) un sistema informático para el uso de los clientes turísticos en Argentina a través de la WEB. Este sistema sugiere al cliente el lugar a visitar a partir de un conjunto de datos aportados por él. También en Europa se ha visto desde hace algunos años la necesidad planteada anteriormente.

Un trabajo presentado en el VII Taller de Turismo y Tecnologías de la Información y las Comunicaciones en España, estudia el estado en que se encontraba el uso de los sistemas informáticos inteligentes en la ciudad de Madrid (5). El estudio, realizado en forma de diagnóstico, arriba a resultados en los que se demuestra la intención de algunos de los hoteles y cadenas hoteleras de introducir sistemas informáticos que hagan uso de la IA para mejorar sus ofertas.

En una breve descripción de la historia de la IA se puede plantear que, los primeros intentos de solucionar problemas se basaron en resolver los mismos problemas que resuelve el cerebro humano, todas las soluciones en un mismo sistema. Esta idea fracasó, lo que provocó una depresión del uso de las técnicas de IA. Luego, la especialización en problemas más concretos trajo de nuevo a la moda estas técnicas y se comenzaron a diseñar sistemas específicos. La incorporación de conocimiento específico de los expertos en diferentes materias a sistemas computacionales permitió desarrollar Sistemas Basados en el Conocimiento (SBC) (6). La experticia de los hombres en un dominio determinado no constituyó la única fuente de conocimientos para los SBC, sino que se usaron otras fuentes. De esta manera se utilizaron las probabilidades de ocurrencias de hechos, experiencias pasadas, errores cometidos anteriormente, entre otras. Las ciencias de la computación se enfrentaron entonces al problema de representar ese conocimiento computacionalmente. Por tal motivo aparecieron diversos tipos de SBC (7): sistemas basados en reglas (SBR), sistemas basados en marcos (SBF), sistemas basados en redes semánticas, sistemas basados en casos (SBCA), sistemas basados en probabilidades (SBP) y redes expertas. En la actualidad los SBC se han expandido, de manera que son usados en disímiles áreas de la ciencia, la técnica, los servicios y la industria.

La industria turística no fue una de las áreas en las que más se usaron los SBC desde sus inicios. Con la expansión y desarrollo del turismo fue que se comenzaron a descubrir problemas que requieren de su uso. El incremento de la información disponible sobre los clientes y el desarrollo de estudios de

mercados¹, unido al aumento de servicios y habitaciones, provocaron que se comenzara a necesitar el uso de técnicas de la IA para analizar el gran número de variables e información.

Un estudio empírico realizado sobre los sistemas informáticos publicados en Internet que promocionan o permiten el manejo de instalaciones hoteleras, aportó datos que justifican la realización de esta investigación.

De los sistemas informáticos consultados en la WEB, 16 en total, el 67% se dedican a la realización de reservaciones on-line y brindar información sobre los destinos de Cuba. Además permiten la búsqueda de paquetes turísticos y la localización de disponibilidades en hoteles. Estos parten de que el cliente conoce la instalación hotelera a la que desea ir. El 80% pertenecen a empresas con capital extranjero que venden paquetes turísticos a Cuba. Los sistemas netamente cubanos como www.havanatur.cu, www.infotur.cu y www.cubatravel.cu solo se dedican a la promoción y no permiten la selección de un destino dentro de Cuba a partir de las características del cliente. Ninguno de los sistemas informáticos consultados permite la personalización de la información que se le muestra al cliente. Sin embargo, la Oficina Nacional de Información Turística (INFOTUR) ha planteado su interés por la incorporación de un sistema informático con estas características.

En el estudio mencionado se valoraron varios portales que promocionan el turismo cubano (anexo 1). Otros sistemas informáticos, publicados en la WEB, dedicados a diferentes países usan técnicas novedosas para el desarrollo de la promoción del turismo u otros negocios. Entre ellos se encuentran el portal Smartourism de Granada en España, el portal WEB oficial de Canarias o la tienda de libros on-line Amazon (10) y (70). El sistema desarrollado en Argentina y presentado en (4) permite la selección del destino turístico a visitar según los deseos del cliente, sin necesidad de leer toda la información sobre los destinos. En estos sistemas informáticos se aprecia el uso de técnicas y métodos de la IA que permiten personalizar los servicios y la información que se brinda a los clientes.

Unido al desarrollo vertiginoso de la Internet y los sistemas informáticos on-line, se expandió la publicidad por esta vía. En fechas actuales la carga publicitaria en Internet cubre un alto por ciento de la información que se publica. Sin embargo, los conceptos de la publicidad tradicional ya no son aplicables al mercado online global. Ahora se presta atención a la comunicación de marca² en lugar de la publicidad, en tanto el marketing ha entrado en una nueva era luego de atravesar varias fases en su evolución. Los canales tecnológicos a través de los cuales los operadores comunican su marca también están cambiando. Hoy la atención se centra en las redes sociales (8). La industria turística se ha planteado dirigir la carga publicitaria de manera que se personalice al cliente la información que necesita. Esta no es una idea propia del turismo, sino que es un aspecto general del uso de Internet, de manera que varios investigadores y centros se dedican a esta temática. Por ejemplo, Google presentó en el año

¹ Los estudios de mercado en el turismo son descripciones profundas de las características, historia, preferencias y comportamiento de los clientes de determinados países o zonas geográficas. Permiten a los hoteleros personalizar el servicio.

² La comunicación de marca es la evolución natural del marketing en Internet. Se trata de una relación más cercana con el cliente personalizando servicios y productos haciendo énfasis en los productores.

2010 el producto adSense³ (9) que publica información relacionada con las palabras buscadas por el usuario. Amazon presenta al usuario que se interesa por un libro otros libros, a partir de la compra realizada por clientes anteriores (10). El portal WEB oficial de Canarias personaliza los contenidos que se muestran al usuario a partir de sus características, con el marcado de etiquetas el usuario va registrando su historial, que es analizado por el portal para sugerirle nuevos contenidos. República Dominicana (11) posee un portal para la elaboración de paquetes turísticos vía Internet, este portal no recurre a las técnicas de la IA sino que confía en la experiencia de sus expertos humanos que elaboran los paquetes y responden las solicitudes. Este proceso manual demora alrededor de un día en dar respuesta a la solicitud del cliente lo que puede significar la pérdida de clientes y de dinero.

A pesar de existir algunos sistemas informáticos que permiten personalizar ofertas turísticas o de compra de productos en general, es muy común encontrar en las redes sociales solicitudes de consejos para comprar cierto producto. También se pueden ver mensajes solicitando consejos para hospedarse en un hotel en cierta ciudad. En la actualidad los clientes turísticos de Cuba deciden la instalación a visitar a partir del consejo dado por un amigo (12), la oferta hecha por una AAVV o un Touroperador (TTOO), o luego de leer la publicidad de cada una de las instalaciones en Internet. Teniendo en cuenta que Cuba posee más de 400 instalaciones hoteleras distribuidas en todas las provincias, con varias modalidades de turismo y con diversos servicios, se torna sumamente difícil encontrar expertos capaces de seleccionar la instalación ideal para cualquier cliente. También resulta complejo para un turista leer toda la publicidad para seleccionar el destino a visitar.

A partir del análisis realizado anteriormente y las dificultades planteadas en los sistemas informáticos que promocionan el turismo cubano, se plantea el problema a resolver. ¿Cómo asistir la selección de la instalación hotelera cubana a visitar, a través de Internet, teniendo en cuenta las características y los gustos de los clientes?

Para resolver este problema se plantea como objetivo general de la investigación. Desarrollar un sistema informático inteligente para la selección de la instalación hotelera a visitar en Cuba vía Internet, a partir de las características del cliente turístico.

Para un mejor desarrollo de la investigación y alcanzar eficazmente el objetivo general propuesto se plantean los siguientes objetivos específicos:

1. Analizar el desarrollo histórico y tendencial de las investigaciones en el campo de la IA y en el uso de esta en los problemas del turismo y en la promoción en Internet.
2. Modelar computacionalmente las características de los clientes turísticos de interés para la selección de un destino en Cuba, teniendo en cuenta la opinión de los expertos, los estudios de mercado y las informaciones acumuladas por las empresas.
3. Diseñar e implementar un sistema informático inteligente para la selección de instalaciones turísticas en Cuba vía Internet.

³ Google adSense es un producto informático de Google que coloca información publicitaria en páginas relacionadas con los temas que se divulgan, muestra a los usuarios la publicidad relacionada con los temas buscados en el buscador.

4. Validar el funcionamiento del sistema a partir de los criterios de validación de sistemas computacionales y de la IA.

Una mejor comprensión del problema planteado se logra a partir de las preguntas de investigación relacionadas a continuación:

1. ¿Es posible y viable la construcción de un sistema informático inteligente para asistir la elección de destinos cubanos acorde a las exigencias y preferencias de los clientes?
2. ¿Qué técnica de la IA o combinación de ellas, permite la creación de un sistema informático inteligente para la selección de instalaciones hoteleras cubanas vía Internet?
3. ¿Cuáles son las características de los clientes que determinan la selección de un destino turístico en Cuba?
4. ¿Es posible modelar las características y los gustos de los clientes turísticos y representarlos matemáticamente?
5. ¿Cómo validar la eficiencia del sistema que se construya?

Estas preguntas dan origen a un grupo de tareas que guían el desarrollo de la investigación.

1. Valoración de las técnicas de la IA usadas en la construcción de sistemas informáticos inteligentes para la promoción en Internet y para el turismo, evaluando su viabilidad para Cuba.
2. Selección de la técnica de la IA o combinación de ellas que permita desarrollar el sistema informático para la selección de instalaciones hoteleras en Cuba vía Internet.
3. Selección de las características de los clientes a tener en cuenta al escoger una instalación hotelera.
4. Determinación del dominio y la forma de comparar los rasgos que caracterizan a los clientes.
5. Diseño e implementación del sistema informático para la selección de instalaciones hoteleras vía Internet.
6. Validación y verificación del sistema informático para la selección de instalaciones hoteleras.

La tesis cuenta con tres capítulos en los que se presenta el trabajo realizado para la investigación y desarrollo del sistema informático. Cuenta además con una sección de anexos donde se recogieron un conjunto de imágenes, tablas y diagramas que aportan mayor información. La introducción, ubicada en el inicio, describe el problema a resolver, los objetivos y las tareas de investigación. Las conclusiones y recomendaciones, al final, muestran los principales resultados alcanzados y los aspectos a mejorar que constituyen nuevos problemas a resolver.

En el capítulo uno se hace un análisis del estado del arte en la aplicación de técnicas de la IA en el turismo. Se valoran un conjunto de sistemas informáticos similares. Se hace un diagnóstico de la situación de este tipo de sistemas en Cuba y del uso de las técnicas de la IA en la industria turística y en la WEB. Finalmente se desarrolla un análisis crítico de las posibles soluciones al problema apuntando a la solución dada por el autor.

El capítulo dos describe el problema con mayor profundidad. Se presentan algunos diagramas del negocio actual, representando la forma en que un cliente decide la instalación hotelera cubana a visitar.

Se plantea una solución al problema a través del uso de un sistema híbrido que combina dos paradigmas de la IA. Por último se describe la estructura de algunos de sus módulos.

El capítulo tres está orientado a la implementación del sistema recomendador. En él se describen las principales funcionalidades y los individuos que interactúan con el sistema a través de diagramas de casos de uso y de actores. Se exponen las soluciones dadas al diseño computacional del sistema recomendador para su funcionamiento en la WEB. Se plantean diagramas de colaboración de las funcionalidades principales y una descripción de cómo interactuar con él. Por último se presenta un epígrafe con la validación y verificación del sistema recomendador.

DESARROLLO

Capítulo 1. Estado del arte sobre el uso de la Inteligencia Artificial en la WEB y el turismo.

La IA como área dentro de las ciencias de la computación ha tenido una evolución histórica muy vinculada a la solución de problemas prácticos (7). De esta manera, su unión a otras disciplinas de la ciencia, la técnica, la industria y los servicios, ha permitido el desarrollo conjunto de las mismas. En este capítulo se presenta una valoración del estado en que se encuentra el uso de las técnicas de la IA en la programación de sistemas informáticos WEB y para el turismo. Se analizan las tendencias históricas de las tres disciplinas, haciendo énfasis en la IA. Se estudian soluciones dadas al problema planteado en la introducción. Y se expone un diagnóstico de la situación actual del problema en Cuba y el mundo.

1.1 El uso de las técnicas de la IA en la WEB y en el turismo.

La información generada en Internet ha sufrido un crecimiento vertiginoso. Se plantea que solo alrededor del 30% de la información presente en Internet es visible (14), a pesar de existir la posibilidad de acceder a ella si se conoce la dirección exacta. El problema radica en que las posibilidades de organizar o estructurar toda la información y sus direcciones en la WEB se tornan prácticamente imposibles. En mayo de 2001, el creador de la red global, Tim Berners-Lee, publicó con otros dos investigadores, James Hendler y Ora Lassila un artículo en la revista Scientific American titulado “The Semantic WEB” (15). En dicho artículo se introduce la extensión de la red de redes. Entre las características de la WEB Semántica se encuentran que los procesadores puedan evaluar la información automáticamente. Esta WEB da la posibilidad de que sea el propio dispositivo de hardware quien se encargue de localizar la información. Sin embargo, aún es idílica la WEB Semántica. Varios grupos de científicos se dedican a su desarrollo, por ejemplo en América Latina grupos de Chile y México (14). Pero la necesidad de modificar las estructuras actuales de la programación WEB y de los ficheros hace que el proceso sea lento, difícil y costoso. En la figura 1 se presenta el esquema por capas que se ha planteado para la WEB Semántica (16). El uso de ontologías de vocabulario para dotar de significado a las aplicaciones WEB eleva la complejidad en su desarrollo.

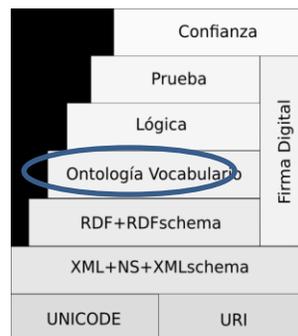


Figura 1 Arquitectura en capa de la WEB Semántica.

Una ontología de vocabulario es: “Un instrumento que define los términos básicos y relaciones a partir del vocabulario de un área de conocimiento, así como las reglas de combinación de estos términos y relaciones para definir extensiones a un vocabulario” (17). Gruber además la define como: “una especificación formal y explícita de una conceptualización compartida” (18).

Las ontologías varían según el contexto en que son usadas las mismas palabras, por lo que el significado de estas puede ser diferente en contextos distintos. Por ejemplo: ratón en las ciencias biológicas es un animal, mamífero, cuadrúpedo, roedor, de tamaño pequeño y transmisor de la enfermedad conocida como leptospirosis. Sin embargo, en las ciencias informáticas, ratón es un dispositivo de hardware externo que permite el control del cursor, puede ser conectado por diversos tipos de puertos y puede presentar otras funcionalidades añadidas.

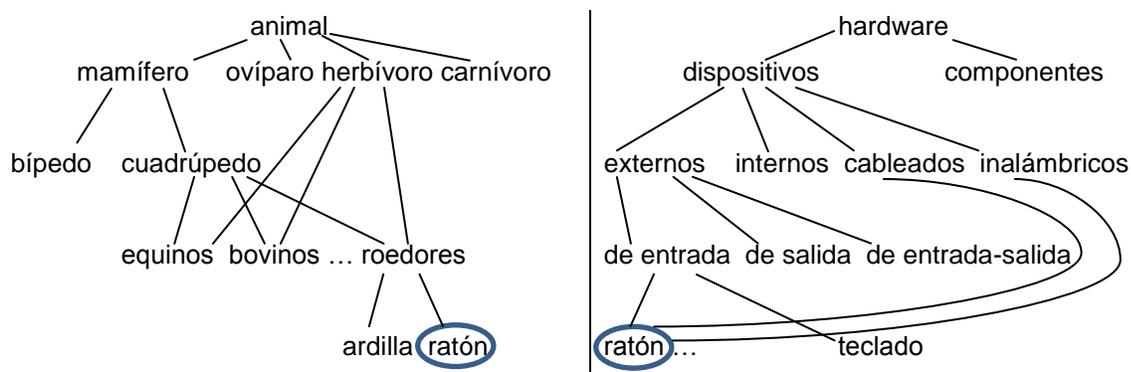


Figura 2 Ejemplos de significados diferentes para palabras iguales.

Estas diferencias hacen que sea necesario definir diferentes ontologías de vocabulario y discernir entre una u otra para la recuperación y búsqueda de información, de acuerdo con el contexto en el que se está usando la palabra o frase. Debido a las dificultades que impone la implementación de aplicaciones WEB Semánticas se recurre en ocasiones a la IA para dotar a los portales o los sistemas informáticos WEB de opciones para estructurar y recuperar la información de manera rápida y personalizada.

Antonio Fumero y Genís Roca, en el libro WEB 2.0, presentan un análisis en el que abordan la cronología de la WEB realizando un recorrido por las etapas pasadas y con cierta osadía dan una predicción de ciertas etapas en la próxima evolución de la WEB (19). La figura 3 muestra un gráfico con estas tendencias. Estos autores y otros (14) y (16) coinciden en la necesidad de incluir técnicas de la IA para llegar a darle sentido semántico a la WEB.

1.1.1 Técnicas de la IA usadas en sistemas WEB.

Uno de los usos que se da a las técnicas de la IA en la WEB es en la recuperación de la información. Como se planteó anteriormente, también puede ser usada en la estructuración de la información de manera semántica, así como en buscadores semánticos que permiten la recuperación de información

más específica. Algunos de estos buscadores semánticos son: Swoogle⁴ y Kartoo⁵. En ellos se hace la búsqueda de la información a partir de ontologías de vocabularios definidas.

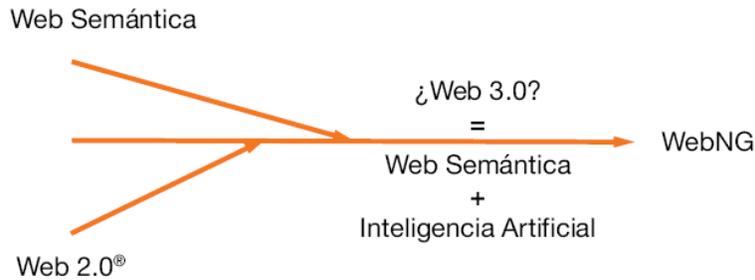


Figura 3 Evolución hacia la WEB Nueva Generación en una etapa de convergencia

Sin embargo, esta no es la única técnica de la IA usada en la WEB. Para la recuperación de información e incrementar los índices de los buscadores tradicionales, han sido usados Agentes Inteligentes y Sistemas Multiagentes (SMA) (20) (21) (22).

Junto al desarrollo de la Minería de Datos como disciplina dentro de la IA, ha aparecido la Minería WEB. Su surgimiento viene dado por la necesidad de usar técnicas no tradicionales en la extracción de conocimiento de la WEB, debido a la forma heterogénea en que se muestra la información en Internet. En (23) se aborda el tema y se plantea esta dificultad. En este sentido se han desarrollado múltiples investigaciones sobre la adquisición de conocimiento de la WEB. Entre algunas de ellas se encuentra el trabajo “MINERÍA DE USO WEB: Metodología para Mejorar la Funcionalidad de Plataformas Virtuales de Aprendizaje” (24) desarrollado en la Universidad de las Ciencias Informáticas de Cuba. “Consultando conjuntos de datos heterogéneos en la WEB: desafíos, aproximaciones y rumbos” (25) desarrollado en el Instituto de investigaciones digitales de empresas, centro de investigación de la Universidad Nacional de Irlanda. En resumen, al buscar en algunas bases de datos internacionales de revistas científicas la frase “minería web” o “web data mining”, aparecen varios resultados. La tabla 1 muestra un resumen.

Tabla 1 Resultados de Bases de Datos de Revistas Científicas a la frase “minería web”.

Base de Datos	Resultados	Dirección URL
CiteSeer	229741	http://citeseerx.ist.psu.edu/
redALyC	371	http://www.redalyc.org
EBSCO	362	http://web.ebscohost.com
Scielo	13	http://search.scielo.org/
DOAJ	11	http://www.doaj.org/

Por su parte los buscadores WEB aportan otros resultados. Google muestra más de 15 millones y Bing más de 25 mil resultados.

⁴ www.swoogle.com

⁵ www.kartoo.com

Además de los usos ya mencionados que se han dado a la IA en la WEB, la evolución de las potencialidades de los lenguajes de programación WEB como Java, JavaScript, ASP, JSP, PHP, J2EE, XML, AJAX, HTML5 y CSS3; han permitido que puedan ser más utilizadas las técnicas de la IA en la WEB. A través de las potencialidades de estos lenguajes y técnicas, se han construido sistemas expertos e incluso redes neuronales que funcionan a través de la WEB. Por ejemplo, en 2007 fue publicado el trabajo titulado “Una herramienta orientada a la WEB para la enseñanza de los conceptos de redes neuronales”. En este se presenta una herramienta orientada a la WEB para la enseñanza del funcionamiento de las redes neuronales artificiales; desarrollado en Java, permite la creación y entrenamiento de redes neuronales artificiales de distintos tipos y ver su funcionamiento (26).

En la revista de la Asociación Americana para la Inteligencia Artificial (AAAI), se publicó un sistema online para recomendar a las familias de bajos ingresos las escuelas públicas a usar en una ciudad de Estados Unidos. Este sistema fue construido usando el lenguaje PHP. Implementa un sistema de clasificación lineal basado en probabilidades y trabaja a través de la WEB. El sistema SmartChoice publicó su primera versión en 2008 para más de 50 usuarios que lo usaron como prueba. Este sistema es clasificado por los autores como un sistema recomendador. Los sistemas recomendadores se pusieron de moda a partir de los años 90 (27) y aun son muy usados en la actualidad. Los sistemas recomendadores a través de la WEB son, en general, sistemas que indican los productos que le puedan interesar a los clientes, los más adecuados de acuerdo a sus preferencias y ofrecen una información que facilita el proceso de decisión (28).

En el Simposio Internacional de Procesamiento de Información del año 2008 fue presentado “Diseño de un sistema experto orientado a la WEB para el diagnóstico en tiempo real de la hipertensión”. Este sistema, publicado en (29), utiliza reglas definidas a partir del conocimiento de los expertos en medicina, para dar un diagnóstico sobre el estado de pacientes hipertensos a través de la WEB. Haciendo uso de un SE basado en reglas se define el diagnóstico y el tratamiento al paciente que se puede encontrar distante de un hospital y se le entrega impreso a través de dispositivos conocidos como kioscos médicos. Los kioscos médicos son lugares donde el paciente puede acceder para una teleconsulta.

Otro ejemplo de recomendador se puede ver en la Primera Conferencia Internacional sobre Avances en Bases de Datos, Conocimiento y Aplicaciones de datos. En esta conferencia del año 2009 se presentó “Sistema Experto médico orientado a la WEB con una heurística para el autoentrenamiento del algoritmo de inducción de reglas”. Este sistema, orientado a la WEB, implementa un SE médico. Realiza un autoentrenamiento a partir de un algoritmo de inducción de reglas. Los datos introducidos por el personal médico son usados para el aprendizaje del sistema. El sistema es entrenado usando un algoritmo de heurística híbrido para la inducción de reglas de clasificación que fueron previamente construidas (30).

En julio de 2011 se publicó en la Revista de investigación en turismo y desarrollo local, un sistema basado en reglas dirigido a la WEB para visitantes en Argentina (4). El sistema realiza un conjunto de preguntas al cliente turístico para formar los hechos que, luego son usados en la inferencia que se realiza a través de reglas predefinidas. Es capaz de sugerir al cliente la región turística aconsejable para

sus expectativas. En este contexto el sistema es fuerte, debido a que las regiones turísticas son poco cambiantes. No así para seleccionar instalaciones o circuitos turísticos, en cuyo caso se debe tener en cuenta el crecimiento de la planta hotelera y las modificaciones que esta sufre en períodos de tiempo relativamente cortos. Este es otro sistema que clasifica dentro de los recomendadores e introduce el uso de la IA en el turismo.

1.1.2 Técnicas de la IA más usadas en el turismo.

La tabla 2, obtenida de (31), muestra las áreas del turismo potenciales para la aplicación de las ciencias de la computación. Este resumen es planteado por investigadores del Centro de Competencia de Comercio Electrónico de Austria y la Universidad de Trento en Italia. A partir de las áreas planteadas por estos investigadores y el análisis de los sistemas informáticos dedicados al turismo, se puede plantear que existen tres vertientes en las que se hace mayor uso de las técnicas de la IA.

- Sistemas de control de instalaciones hoteleras.
- Toma de decisiones e Inteligencia de Negocios (BI).
- Sistemas recomendadores para clientes.

Tabla 2. Áreas potenciales para la aplicación de las ciencias de la computación en el turismo.

Sistemas de reservación e información	Sistemas de reservación distribuidos.
Sistemas de administración e información	Internet y servicios WEB
Sistemas distribuidos en redes de banda ancha	Multimedia y realidad virtual
Diseño de interfaces de usuarios gráficas, modelado de usuarios y estudios de aceptación.	Sistemas expertos, sistemas basados en el conocimiento, agentes inteligentes y presentación inteligente de información.
Sistemas de toma de decisiones, minería de datos y descubrimiento del conocimiento	Arquitectura de sistemas, ingeniería de software y diseño de base de datos
Servicios móviles	Comunidad, Servicios basados en contextos y localización.
Modelado empresarial	Comercio electrónico y marketing
Ingeniería de negocios	Temas de estandarización.
Estadísticas, optimización y simulación.	Marketing inverso.
Capacidad y gestión del rendimiento	Control de la calidad.
Métodos de marketing apoyados en la computación	Aspectos de administración, organizacionales, de negocios y económicos. Ciencias de la administración
Aspectos legales y sociales de la industria turística en viajes y turismo.	Nuevos intermediarios y la desintermediación.
La creación de redes dinámicas y empresas virtuales	Marcos de trabajo y ontologías.

La complejidad de las instalaciones hoteleras actuales y la diversidad de servicios, tipos de habitaciones y necesidades de los clientes, han llevado a que los empresarios necesiten sistemas informáticos que les permitan manejar todas las posibles combinaciones de procesos y servicios. En España en el año 2008 se realizó una investigación exploratoria cuyos resultados fueron presentados en el VII Congreso Turismo y Tecnologías de la Información y las Comunicaciones (Turitec 2008). En esta investigación se estudió la posibilidad de usar sistemas expertos en los 20 hoteles de categoría 5 estrellas de la ciudad de Madrid. Los hoteles entre 10 y 50 años de antigüedad, deseaban en su mayoría la implantación de un

sistema de este tipo. Así como los más grandes, con más de 200 habitaciones, también lo deseaban (32).

Otro ejemplo es anunciado en enero de 2012 en España, con la puesta en marcha del sistema iGreen por parte de ITH y Euroconsult. Este se considera el primer sistema de IA para la gestión y optimización del gasto energético en hoteles (33). El sistema usa para su funcionamiento controladores lógicos programables (PLC) y permite el control energético de una instalación hotelera, gas, agua y electricidad, pero los fabricantes no explican las técnicas de la IA usadas.

Desde la vertiente de los sistemas para la toma de decisiones y la inteligencia de negocios también se han desarrollado varias investigaciones aplicadas al turismo. De estas investigaciones han surgido productos como el *Oracle Business Intelligence Standard One*. Este sistema, desarrollado por la empresa Oracle⁶, fue usado por la empresa española DHARMA INGENIERIA como base para un sistema de Inteligencia de Negocios (BI) en instalaciones hoteleras. La implantación incluyó entre sus fases el análisis de los informes que el cliente solicitaba, diseño de un *Data Mart* con características acordes a la consolidación de información necesaria, instalación y puesta en marcha del producto de BI y la implementación de una interfaz WEB que permitía acceso vía Intranet e Internet (34). La investigación desarrollada para este producto pone de manifiesto el uso de los almacenes de datos y la minería de datos en la toma de decisiones en el turismo.

Otro tanto se plantea en el trabajo de un colectivo de autores de la *Universität Klagenfurt* de Austria y el Instituto Europeo de Investigaciones del Turismo (ETOUR) de Suecia (35). En este trabajo se hace una propuesta de uso de varias técnicas de minería en la WEB para el turismo electrónico (eTourism). Entre ellas se menciona el uso de BI, minería de estructura WEB⁷, minería de contenido WEB⁸ y minería de uso WEB⁹. A partir de estas técnicas se propone un flujo de trabajo para la clusterización de los usuarios de acuerdo a los datos obtenidos sobre ellos y utilizar esta información en la toma de decisiones para la reestructuración de las ofertas turísticas.

Sin embargo, los sistemas recomendadores para clientes han tenido un mayor desarrollo a nivel mundial. Un sistema recomendador no necesariamente debe usar técnicas de la IA. Por ejemplo, los primeros usaron el filtrado de datos para mostrar a los clientes información más relacionada a sus intereses (36). El uso de la IA permitió el análisis de mayor cantidad de variables y por tanto la mejora de las propuestas.

Un grupo de sistemas recomendadores han intentado resolver el problema de la planificación de circuitos turísticos. Se trata de recorrer varias instalaciones, ciudades y/o destinos de manera que se ahorre tiempo, dinero y se maximice el placer obtenido por el cliente. Este problema ha tenido diversos enfoques para su solución. Algunos trabajos lo han modelado como un problema de optimización y han

⁶ Oracle es la empresa líder en software de base de datos. Desarrolla tecnologías y ha adquirido las mejores compañías en su clase en los últimos años.

⁷ Minería sobre la estructura de las URLs de consultas de los usuarios en la WEB.

⁸ Minería sobre los contenidos de las páginas en la WEB.

⁹ Minería sobre el uso de la WEB y los logs de los usuarios.

usado técnicas de la programación lineal (37). Además ha sido resuelto a través del uso de un modelo de optimización por colonias de hormigas ACO (38). Otros autores han modelado el problema para su solución a través de SBC (39).

Otro enfoque para el desarrollo de sistemas recomendadores parte del lugar donde se encuentra el cliente y las actividades turísticas que puede desarrollar a partir de allí. De esta manera se han realizado investigaciones que hacen uso de los SMA para hacer propuestas teniendo en cuenta las ofertas más cercanas y atractivas para el cliente (40), (41) y (42). Por último, existen un conjunto de sistemas recomendadores que parten de las experiencias de clientes anteriores para realizar ofertas a los nuevos clientes. Con este punto de vista se han desarrollado trabajos como el Sistema Adaptativo Multi-Agente de Planificación (43) o el sistema *Intelligent Travel Planning* (44).

1.2 Estado histórico y actual de los sistemas expertos orientados a la WEB y en el turismo. Análisis histórico lógico con tendencias actuales.

La aparición y desarrollo de los SBC data de la década del 70 del siglo pasado. Entre los más reconocidos el MYCIN y el PROSPECTOR con fines médicos y de descubrimiento de yacimientos respectivamente (45) (7). En el quinquenio del 80 al 85 se crearon diversas empresas dedicadas a la producción de SBC. Luego en la década del 90, con la miniaturización de los componentes y la incorporación a la vida doméstica de equipos más modernos, comienzan a invadir los SBC la vida común. Aparecen para diagnosticar fallas en equipos o incluso corregirlas, entre otras funciones.

Una definición de SBC dada en (6) plantea: Un SBC es un sistema computarizado que usa conocimiento sobre un dominio para arribar a una solución de un problema de ese dominio. Esta solución es esencialmente la misma que la obtenida por una persona experimentada en el dominio del problema cuando se enfrenta al mismo problema pero que presentan tres rasgos que los caracterizan y son:

- La separación del conocimiento de cómo es usado (distinción entre conocimiento y estrategia de control).
- El uso de conocimiento muy específico del dominio.
- Naturaleza heurística, en lugar de algorítmica, del conocimiento empleado.

Otra definición similar es dada en (46).

Los SBC se clasifican por la forma en que estructuran el conocimiento y la forma en que realizan la inferencia. Una clasificación de los SBC es dada en (7) y (6).

- Sistemas basados en reglas (SBR).
- Sistemas basados en frames (SBF).
- Sistemas basados en redes Semánticas.
- Sistemas basados en casos (SBCA).
- Sistemas basados en probabilidades (SBP).
- Redes expertas.

Dentro de los SBC que se han implementado para la industria turística una de las formas de representación del conocimiento (FRC) usada son las reglas.

Los SBR han sido usados por varios investigadores para desarrollar aplicaciones que permitan solucionar problemas a través de la WEB. Entre algunos de ellos están (47) y (48), en este último se hace uso de un algoritmo ID3¹⁰ para la generación de reglas que permiten la clasificación a través de la WEB. El recorrido por un árbol ID3 es más rápido que el análisis de todas las reglas, por lo que se alcanzan respuestas más rápidas, condición necesaria en los entornos WEB.

Otro ejemplo en América Latina lo constituye “Implementación en la Universidad Veracruzana de un SE para el apoyo en la solución de problemas en un sistema de educación en línea” (49). En este trabajo se representa el conocimiento a través de ontologías desarrolladas con Protégé¹¹. Además el SBC es construido con el Shell nombrado *Java Expert System Shell* que permite interactuar con Java y Protégé.

Orientados a la WEB se han desarrollado varios sistemas recomendadores para el turismo que hacen uso de SBR. Algunos de ellos fueron mencionados ya en el epígrafe anterior y están publicados en (4), (28) y (43).

Otra solución de la IA usada en problemas del turismo son los Sistemas Basados en Casos (87), (88), (89), (91), los cuales tienen como forma de representar el conocimiento los propios casos. Sus antecedentes están ubicados en el comportamiento humano de recordar experiencias pasadas para resolver nuevos problemas.

Según Kolodner: “El RBC es un tipo especial de razonamiento análogo. Una experiencia previa puede sugerir una solución para un nuevo problema o una manera de interpretar una situación, puede avisar sobre la forma en que debe ser tratado un problema o mostrar los efectos potenciales de una solución predicha.” (50)

Otras definiciones de RBC:

- Razonamiento basado en casos es [...] razonar mediante recordar. - Leake, 1996 (51).
- Un razonador basado en casos resuelve nuevos problemas mediante la adaptación de soluciones que fueron usadas para viejos problemas. - Riesbeck & Schank, 1989 (51).
- Razonamiento basado en casos es una aproximación reciente a la resolución de problemas y el aprendizaje [...] - Aamodt & Plaza, 1994 (52), (76).

Los SBCA basan su funcionamiento en recuperar los casos más similares, adaptarlos para dar solución al nuevo problema, revisar la respuesta dada y por último retener el conocimiento mediante el mecanismo de aprendizaje automatizado. Estas características lo hacen ideal para su uso en sistemas en los que el tiempo de respuesta no es una limitante. Por este motivo, varios de los SBCA son desarrollados para su funcionamiento en ambiente de escritorio, no tanto así en entorno WEB.

¹⁰ Induction Decision Tree.

¹¹ Editor de ontologías y bases de conocimiento basado en clases y sus atributos que permite una organización y modificación sencilla del conocimiento mediante un lenguaje visual.

Durante el desarrollo de la investigación fueron encontrados algunos trabajos que hacen uso del RBC para sugerir opciones a los clientes turísticos (87), (88), (89), (91). Varios de estos sistemas son conocidos como sistema recomendadores o planificadores de viajes turísticos. El RBC ha sido combinado con otras técnicas de la IA como sistemas multiagentes (91), lógica difusa (89), conjuntos borrosos (90), entre otras. Dada la característica del RBC de utilizar conocimiento previo para resolver nuevas situaciones, es usado en los sistemas recomendadores para proponer guías de viajes a usuarios con requerimientos similares a otros previamente analizados.

1.3 Situación del uso de los sistemas expertos orientados a la WEB en el turismo en Cuba.

En la bibliografía consultada se plantea con gran fuerza la necesidad de desarrollar la informática en el campo de los servicios turísticos no sólo en Cuba, sino en el mundo entero. Con el objetivo principal de lograr un mejor acercamiento a las expectativas de los clientes y un servicio de mayor calidad, se plantea la necesidad de usar las técnicas actuales de publicidad.

Sobre las características del turismo en Cuba, sus perspectivas y desarrollo, en una conferencia dada por el Dr. Ramón Martín Fernández, decano en la fecha citada de la Facultad de Turismo de la Universidad de la Habana plantea: “Por otra parte, se deberá eliminar el insuficiente conocimiento de la demanda e incrementar notablemente la hasta hoy insuficiente aplicación de la ciencia y la tecnología. Especialmente en las aplicaciones informáticas a la comercialización y a la gestión. También se nota la falta de sistema de gestión integrada en los destinos específicos de Cuba” (3).

En el año 2009 la Universidad de Oriente publicó en un artículo sobre un encuentro de los principales directores de la empresa Barceló en Argentina: “...Otra de las novedades que se comentaron a los profesionales durante los encuentros fue el modo de acceso al sistema online de la firma, tanto para obtener información como para efectuar reservas en forma automática, usando un nombre de usuario y contraseña. En una primera fase, se van a poder reservar noches de estadía en todos los hoteles de la cadena, y en una segunda instancia se avanzará en un portal de Vacaciones Barceló en que la oferta se va a ampliar a paquetes de servicios que incluirán pasajes aéreos, traslados, y productos de otros proveedores con quienes la firma ya realizó la integración XML. Esta última etapa está en el período de prueba y muy próximamente será lanzada oficialmente al mercado” (2).

Este ejemplo pone de manifiesto la intención ya marcada de desarrollar aplicaciones de la WEB 2.0 para su uso en los sistemas de reservaciones en el turismo. Sin embargo, se plantea para cualquier destino de la cadena Barceló y no para Cuba. Además no presupone el uso de la IA en la selección de los destinos acorde al cliente.

Durante el proceso de investigación se realizaron entrevistas a algunos especialistas del turismo de diferentes agencias turísticas y centros de trabajo relacionados con el turismo en el polo Jardines del Rey. Los expertos entrevistados son representantes de la agencia turoperadora Thomas Cook (una de las más representativas en Jardines del Rey) y guías de la agencia de viajes Cubatur. Además se aplicó una encuesta a representantes y guías de las agencias Cubanacán, Transat, Hola y Sun, Sunwing y

especialistas de calidad en instalaciones hoteleras. Todos cumplen el requisito de llevar más de 5 años trabajando con clientes turísticos de diversos países y tener amplia experiencia en el sector de los servicios turísticos. Ninguno de los entrevistados conocía de la existencia de sistemas informáticos capaces de apoyar la selección de una instalación hotelera en Cuba a partir de las preferencias de los clientes.

En el año 2010 La Oficina Nacional de Información Turística de Cuba (INFOTUR) inició un proyecto de desarrollo conjunto con la Universidad de las Ciencias Informáticas (UCI). Específicamente fue firmado con la facultad “Félix Francisco Borrero Lavadí” de Ciego de Ávila. El proyecto tiene como fin el desarrollo de portales WEB para la promoción de los destinos turísticos de Cuba. La figura 4 plantea un esquema que esboza la estructura general de estos portales. Estos portales se encuentran conectados entre sí a través de hipervínculos. Sin embargo, en su concepción no se incluyó un sistema recomendador para los clientes. Por lo que, para seleccionar su instalación hotelera, o incluso su destino favorito, deben hacer una búsqueda y procesamiento de la información publicada en todos los portales.

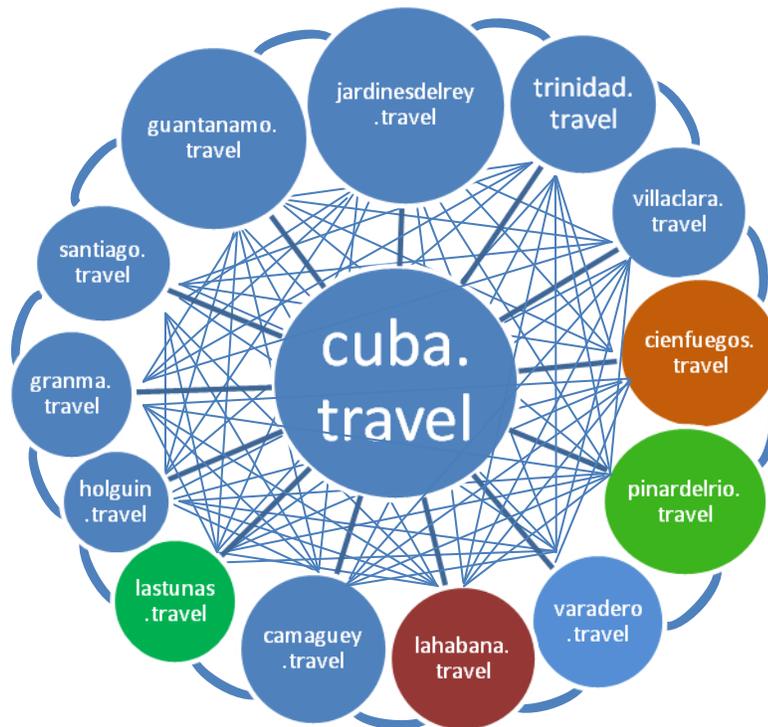


Figura 4 Esquema de los portales WEB para destinos turísticos de Cuba

A partir de los ejemplos planteados anteriormente surge el reto de incrementar las investigaciones en las ciencias de la computación y la informática para dar solución a problemas de la industria turística cubana. Esto se evidencia además, en los lineamientos de la política económica y social aprobados en el VI Congreso del Partido Comunista de Cuba. Entre algunos de estos lineamientos se encuentran:

135. Definir una política tecnológica que contribuya a reorientar el desarrollo industrial, y que comprenda el control de las tecnologías existentes en el país; a fin de promover su modernización sistemática

atendiendo a la eficiencia energética, eficacia productiva e impacto ambiental, y que contribuya a elevar la soberanía tecnológica en ramas estratégicas. Considerar al importar tecnologías, la capacidad del país para asimilarlas y satisfacer los servicios que demanden, incluida la fabricación de piezas de repuesto, el aseguramiento metrológico y la normalización.

258. Perfeccionar las formas de comercialización utilizando las tecnologías más avanzadas de la información y las comunicaciones.

261. Perfeccionar la comunicación promocional a nivel institucional y empresarial, precisando la administración de los recursos y su asignación por mercados y técnicas, con la utilización de tecnología de avanzada.

268. Lograr una gestión más eficiente y eficaz de las oficinas y empresas en el exterior, con el objetivo de captar e incrementar nuevas modalidades de turismo y segmentos de mercados de altos ingresos. (54)

En ellos se plantea también la necesidad de poner la ciencia en Cuba en función de la solución de problemas prácticos. Queda plasmada en los lineamientos 258, 261 y 268 la intención gubernamental del uso de tecnologías de la información y las comunicaciones en función de la promoción y gestión hotelera en Cuba.

A pesar de las necesidades planteadas, en la investigación desarrollada en bases de datos de revistas científicas internacionales, revistas y memorias de eventos nacionales, se conoció solo un sistema recomendador publicado para el turismo en Cuba. El trabajo fue publicado en las memorias de la VI Conferencia UCIENCIA (38), se encarga de diseñar circuitos turísticos en Cuba, pero aún no se encuentra implementado. Este sistema además, es diseñado para planificar circuitos, o sea recorridos con pernoctaciones por varios hoteles, lo que lo hace diferente del problema planteado para esta investigación.

A diferencia de Cuba, a nivel mundial si se han desarrollado desde hace algunos años varios sistemas recomendadores para el turismo. Algunos de estos sistemas ya fueron mencionados anteriormente y se encuentran publicados en (4), (28), (36), (41) y (43), entre otros. Los sistemas encontrados y ya referenciados en epígrafes anteriores, hacen uso de SMA, modelos de optimización o SBR. A pesar de que estos resuelven los problemas para los que fueron diseñados, no se ajustan a las exigencias del problema actual. En el problema actual se cuenta con ejemplos de visitas de clientes anteriores a Cuba. Se cuenta además con expertos en el trabajo en AAVV y TTOO. Por otro lado se tienen como exigencias la respuesta rápida que debe dar el sistema por el entorno WEB para el que se diseña, la naturaleza de los datos a tener en cuenta para el análisis y el tipo de respuesta que se debe dar al cliente. Estas exigencias apuntan inicialmente hacia la utilización de un SBC, específicamente al uso del RBC.

Existen varios sistemas híbridos que usan RBC unida a otra técnica de la IA (55, 61, 62, 63, 89, 90). Estos han permitido que dificultades de unas técnicas sean resueltas con las ventajas de otras. Ejemplos de estos sistemas híbridos pueden ser encontrados al unir RBC con redes neuronales artificiales (55). Otras variantes combinan RBC con reglas de producción (62) y en ocasiones se manejan rasgos

borrosos (56, 86, 90). Por otra parte los algoritmos genéticos han favorecido la selección o ponderación de rasgos y por tanto auxilian en la conformación de los casos que constituirán la Base de Casos del SE (61).

Por último, aunque pueden existir otras combinaciones de técnicas de la IA, se hace referencia a la unión entre RBC con los árboles de decisión. Los árboles de decisión poseen una estructura que permite de manera más rápida la recuperación de la información. Estos a su vez permiten el agrupamiento de casos en clases lo que puede ser usado en el RBC para organizar la base de conocimientos. Esto trae consigo una reducción en el tiempo empleado en el acceso y la recuperación de casos semejantes. Su uso se evidencia en (63) y en el trabajo presentado en (86).

Los árboles de decisión, de manera general, aportan beneficios que pueden resolver dificultades del RBC, relacionadas con los algoritmos para la recuperación de casos y las estructuras para la organización de las bases de conocimientos.

1.4 Técnicas de la IA que permiten dar solución al problema.

1.4.1 Técnicas de Optimización

Previamente fueron valorados algunos trabajos que dan solución a problemas relacionados con el problema actual. Conocidos como sistemas recomendadores, varios sistemas han sido desarrollados para apoyar la selección de productos y servicios a través de la WEB. Estos sistemas modelan problemas de la vida cotidiana matemáticamente para ser solucionados a través de algoritmos computacionales. Una de las variantes computacionales usadas para dar solución al problema de la elaboración de paquetes turísticos es la optimización. Varios trabajos e investigaciones se han desarrollado, entre ellos se encuentran los citados en (37) y (38). Estos dos sistemas usan diferentes técnicas de optimización. En el caso del primero se usa un modelo de optimización basado en la búsqueda Tabú, una metaheurística usada en la IA. El segundo es desarrollado usando optimización por colonia de hormigas (ACO). En ambos casos se trata de maximizar los beneficios que recibe el cliente.

El problema resuelto usando técnicas de optimización consiste en la selección de un circuito turístico óptimo para el agrado del cliente. O sea, recorrer varias instalaciones en varios días, de manera que se maximice la satisfacción del cliente y se minimicen diversos criterios. Los criterios a minimizar en ocasiones son el dinero empleado, los kilómetros recorridos, los días de estancias, entre otros. Por ejemplo, el trabajo presentado en (37) minimiza el precio del transporte, el precio de las actividades y el tiempo empleado para desarrollar las actividades. En el anexo 2 se puede observar una secuencia de imágenes del sistema Plnr de Barcelona (68).

1.4.2 Técnicas de filtrado de datos para la solución de problemas similares.

Para mostrar al cliente en la WEB los temas más sugerentes de acuerdo a su comportamiento en portales promocionales del turismo, ha sido usado el filtrado de información. La guía turística Tiddler utiliza este tipo de técnica para mostrar al cliente la información ajustada a sus preferencias (36). Otro

ejemplo se puede ver en el portal de turismo de Granada, España. Este aplica el filtrado de datos en la planificación de visitas turísticas. Está vinculado a un portal de promoción llamado Smartourism y es capaz de generar un plan de visita turística personalizado de manera automática como lo hace un guía turístico (69).

El filtrado de información es una solución al problema de búsqueda, en una o varias bases de datos, de la información y su publicación siguiendo algunas reglas. Algunas de los métodos usados para el filtrado de datos son:

- Filtrado simple.
- Reglas de asociación.
- Filtrado colaborativo.
- Filtrado por contenido.
- Perfilado.

El filtrado simple consiste en la definición de grupos de usuarios, a los que se asocian vistas particulares de la aplicación. Estos grupos de usuarios pueden estar predefinidos, o crearse en base a determinados atributos del usuario o del contexto. Las reglas de asociación parten de la definición de criterios que, cuando son cumplidos por un determinado usuario y opcionalmente ante la ocurrencia de un determinado evento, desencadenan una acción. El filtrado por contenido determina la pertenencia o no de los objetos a cada una de dichas categorías y se muestran aquellos que entran en el grupo de categorías de interés de cada usuario. El filtrado colaborativo se basa en la actividad de otros usuarios de comportamiento similar al actual, para determinar aspectos como cuál será la próxima acción que el usuario realice en el sistema o en qué productos puede estar interesado. El perfilado parte de un identificador proporcionado por el usuario de manera implícita y permite la personalización de la aplicación en base a la actividad del usuario en otras aplicaciones WEB (70).

1.4.3 Técnicas de Sistemas Multiagentes (SMA).

En la industria turística se han usado los SMA con éxito. Varios de los sistemas desarrollados permiten resolver problemas de planificación de viajes turísticos y recomendación de lugares a partir de la posición actual de un cliente. El trabajo "Sistema turístico multiagente: una aplicación de agentes en la industria turística" presenta un sistema informático para dispositivos móviles que permite realizar la planificación de un día de turismo o escoger un lugar a visitar, por ejemplo un restaurante, a partir de la posición actual (71). En ese trabajo se implementan 4 agentes inteligentes, BrokerAgent, UserAgent, SightAgent y PlanAgent. De ellos, BrokerAgent es el agente encargado de la comunicación entre los demás, UserAgent es el encargado de interactuar con el cliente a través de dispositivos móviles, luego el SightAgent se encarga de manipular la información sobre los lugares turísticos. Por último el PlanAgent permite la obtención de una planificación para el día.

Un trabajo similar fue desarrollado en la Universidad de London. El sistema llamado CRUMPET implementa 4 agentes con características similares (72). Fueron usadas otras técnicas de programación, pero en esencia el sistema desarrollado posee características similares.

Años antes, en 1998, se publicó el trabajo "Un kiosco en Internet basado en un sistema multiagente". Este trabajo resuelve el problema de ofrecer una guía de asistencia al turista. Permite seleccionar destinos o instalaciones a visitar a partir del lugar donde se encuentra el cliente. Hace uso de tres agentes inteligentes *ClientAgent*, *InformationAgent* y *NameserverAgent* (41). Los ejemplos muestran que los SMA permiten dar solución al problema de selección de un destino o una oferta, partiendo de un conjunto de características.

1.4.4 Técnicas de Sistemas Basados en el Conocimiento (SBC).

En el turismo existen varias áreas donde los SBC pueden usarse. En particular, los sistemas recomendadores han hecho uso de esta técnica. Algunos ejemplos de SBC como recomendadores del turismo han sido ya mencionados en (4), (5) y (39).

En la investigación desarrollada por el autor no se han encontrado ejemplos que hagan uso de redes neuronales artificiales (RNA) en la solución del problema planteado. Sin embargo, la creación de una RNA puede ser útil ya sea como sistema clasificador o como selector de los rasgos. Modelos como el de Hopfield (73), IAC (74) Hammig o MLP (74) pudieran ser factibles.

Otro tipo SBC que se ha utilizado para desarrollar sistemas recomendadores son los SBR. En los SBR la base de conocimientos está constituida por un conjunto de reglas. El trabajo publicado en (4) es un ejemplo del uso que se ha dado a este tipo de sistema. En el mismo, se utilizan reglas definidas a través de un Shell conocido como *Java Expert System Builder*. El sistema comienza preguntando al cliente un conjunto de informaciones que son usadas como hechos en las reglas de inferencia. Finalmente se sugiere al cliente un destino argentino a visitar, de acuerdo a las características y preferencias descritas. Para incluir nuevo conocimiento se requiere adicionar nuevas reglas, lo que puede provocar el crecimiento de la base de conocimiento y por consiguiente algunos problemas de inconsistencia.

El RBC ha sido planteado como un método de la IA útil cuando es difícil formular reglas pero existen casos disponibles. Es útil también cuando las reglas requieren mayor información de la que se dispone o usarlas es costoso por la cantidad excesivamente grande y se tienen casos disponibles (6).

Algunos autores han descubierto las posibilidades que brinda el RBC para la construcción de sistemas recomendadores en el turismo. Su uso se ha orientado hacia la creación de planes de viajes basados en experiencias anteriores. Las características de los sistemas recomendadores ha exigido la combinación del RBC con otras técnicas, para incrementar la velocidad en la respuesta dada al cliente a través de la WEB. Uno de los usos que se le ha dado ha sido en combinación con sistemas multiagentes (20).

No fueron encontrados sistemas recomendadores para el turismo basados en casos desarrollados en Cuba. Sin embargo, a nivel mundial han sido diseñados combinando el RBC con otras técnicas (87),

(88), (89). Además, la existencia de sistemas que hacen uso del RBC en otras áreas del saber, sugiere la posibilidad de su uso en el turismo.

1.5 Conclusiones del capítulo

A partir del análisis desarrollado en el estudio del estado del arte se puede concluir que:

Los sistemas recomendadores, a pesar de ser desarrollados con diversas técnicas computacionales, han encontrado en los paradigmas de la IA mejores soluciones para ajustar las ofertas a las diversas preferencias de los clientes. En su desarrollo se debe partir de una correcta definición del problema a resolver, definiendo el objetivo a alcanzar y los datos con que se cuenta.

El RBC, como paradigma de la IA basado en experiencias anteriores para inferir soluciones nuevas, constituye una solución para el desarrollo de un sistema recomendador. Si se tiene en cuenta que un alto por ciento de los clientes que visitan Cuba lo hacen a partir de sugerencias hechas por amigos o familiares, entonces los turistas anteriores pueden ser tomados como experiencias para resolver nuevos problemas.

Un sistema recomendador para el turismo que funcione a través de la WEB debe responder las solicitudes de manera rápida. Por lo anterior, se hace necesario concebir soluciones híbridas.

Capítulo 2 Descripción del problema de la selección de instalaciones hoteleras en Cuba.

En este capítulo se describe el problema de la selección de instalaciones hoteleras a partir de las características y gustos de los clientes. Se plantea la modelación de la solución al problema y se expone la importancia para la promoción turística de Cuba. Por último, se presenta el diseño de un sistema híbrido basado en casos que conjuga el uso de árboles de decisión.

2.1 Descripción del problema de la selección de instalaciones hoteleras a partir de las características y gustos de los clientes.

Varios de los países ubicados en la zona geográfica del Caribe tienen como renglón económico fundamental el turismo. Tras años de experiencia en este renglón cuentan con la presencia en sus países de varias empresas transnacionales que dominan el mercado internacional. De esta manera garantizan que sus productos sean promocionados en todo el mundo. Sin embargo, las empresas cubanas no acumulan tantos años de experiencia ni tienen presencia en tantos mercados internacionales. En la tabla 3 se expone la presencia de las principales cadenas hoteleras en el mundo y Cuba en el año 2010.

Tabla 3 Presencia de Cadenas hoteleras en Cuba y el mundo

Cadena hotelera internacional	País	Cantidad Instalaciones
Sol Meliá	Cuba	24
	Alemania	21
	España	150
	Costa Rica	1
	Puerto Rico	1
	República Dominicana	4
Iberostar	Cuba	3
	México	2
	República Dominicana	3
	Jamaica	1
Tryp	Cuba	3
	España	55
	Brasil	11
	Portugal	7

A pesar de que con estas cadenas Cuba posee gran cantidad de instalaciones, las diez cadenas hoteleras más importantes del mundo en el año 2009 (InterContinental, Cendant, Wyndham, Marriott Internacional, Accor, Hilton Corporation, Choice, BestWestern, Starwood Hotels & Resorts) (67) no poseen instalaciones en Cuba.

Unido a la fuerte competencia en el turismo se plantea el reto internacional de desarrollar aplicaciones WEB que sean capaces de compartir información y conocimiento colectivos. Este conocimiento debe ser obtenido a partir de los propios usuarios de Internet.

El proceso general de realizar una reservación en un destino turístico presupone en la mayoría de los casos que ya el cliente conoce el lugar al que desea ir. A partir de esta información se puede realizar *online* o en oficinas de reservaciones de algunos países. En ocasiones el cliente puede llegar a estas oficinas de Agencias de Viaje (AAVV) o Turoperadores (TTOO) y solicitar a los expertos sugerencias sobre el destino a escoger. De otra manera, el cliente puede leer la descripción de los lugares del país que desea visitar publicada en Internet y escogerlo a partir de sus preferencias. El proceso es representado en diferentes notaciones en las figuras 5 y 6.

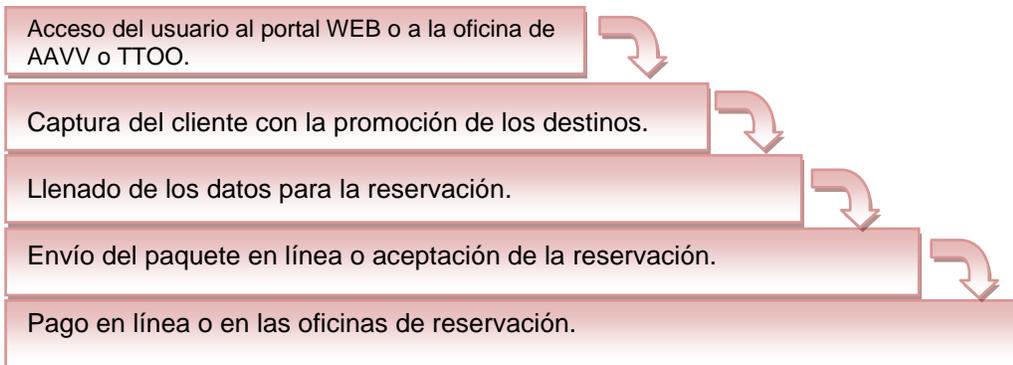


Figura 5 Representación del proceso de reservación turística en lenguaje natural.

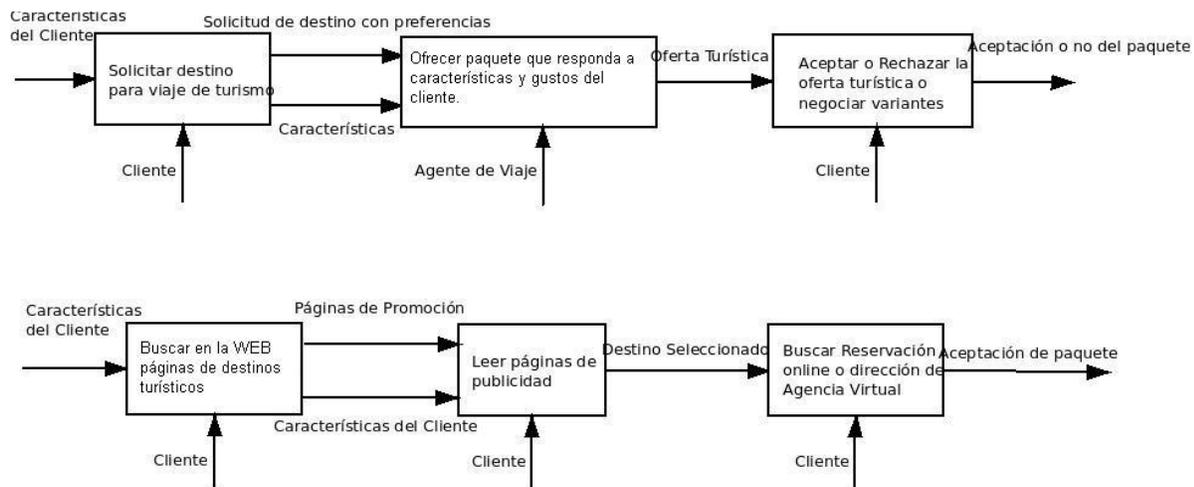


Figura 6 Representación del proceso de reservación turística en notación IDEF.

En los últimos años llegan a Cuba más de 2,5 millones de visitantes. La mayoría de ellos vienen con el propósito de disfrutar de vacaciones, aunque también se incluyen otros motivos como negocios, salud, excursiones, científicos, etc. El anexo 3 muestra los datos de arribos entre los años 2005 y 2010 en miles de clientes.

Combinando la cantidad de motivos por los que los clientes visitan Cuba, con los principales países emisores y la cantidad de instalaciones hoteleras, se obtiene una idea empírica de la dimensión del problema para definir la instalación a sugerirle a un cliente en particular. Si se adicionan otras características como la edad, los precios, la fecha y las preferencias de los clientes, entonces el

problema adquiere mayor dimensión. Los anexos 4 y 5 muestran tablas con los datos de los principales países emisores de turistas y la cantidad de instalaciones hoteleras por tipo.

Matemáticamente el problema a analizar por un agente de viajes es encontrar la combinación de variables para un cliente en particular. Es decir, la instalación hotelera a partir de: el país de origen, los motivos, las preferencias de actividades, precios, la fecha, entre otros aspectos.

Si se suponen los siguientes datos, algunos tomados de la Oficina Nacional de Estadísticas (ONE):

- Instalaciones: 447.
- Motivos: 5 (considerando ocio-recreo, negocios, religiosos, salud, investigaciones).
- Países: 26 (considerando solo los países de mayor emisión de turistas a Cuba).
- Actividades: 16 (considerando, caza, deportes náuticos, arte-artesanía-ciudad, pesca, buceo, excursiones-senderismo, playa, salud, negocios,...).
- Fechas: 3 (considerando las temporadas del turismo, alta, baja y media).
- Precios: 5 (considerando los términos más usados en el turismo, lujoso, todo incluido, media pensión, medio, económico).

A partir del factor entre las posibles combinaciones generadas, un agente de viajes necesita valorar 13.946,400 opciones. Más de 13 millones de posibles variantes para sugerir un destino.

Desde el punto de vista computacional la búsqueda de soluciones en este espacio puede ser engorrosa. A pesar de que las computadoras actuales pueden realizar millones de operaciones por segundo, la complejidad de realizar búsquedas en bases de datos puede provocar lentitud en el sistema, lo cual no es deseado en entornos WEB.

2.2 Necesidad de crear el sistema de selección de instalaciones hoteleras cubanas orientado a la WEB.

El reto a los especialistas de las ciencias de la computación ha sido lanzado por los estudiosos del turismo en Cuba. Ramón Martínez en (3) y el Centro de Información y Documentación Turística (CIDTUR) en (2) coinciden en la necesidad de usar más las TIC con sistemas que ayuden a la personalización de los servicios. Esta no es una necesidad única del turismo, sino que la presentan varias empresas en Cuba. Sin embargo, por las características de los clientes que visitan Cuba se convierte en una muy buena opción la publicidad a través de Internet. La figura 7 muestra los principales países emisores de turistas a Cuba en los últimos años. En la figura 8 se muestran los niveles de acceso a Internet por cada región del planeta. A través del análisis de estas dos gráficas se puede observar que las zonas de Norteamérica y Europa coinciden en tener entre los más altos niveles de acceso a Internet, 71% y 43% respectivamente y ser las de mayores emisiones de turistas a Cuba, 40% y 20% respectivamente.

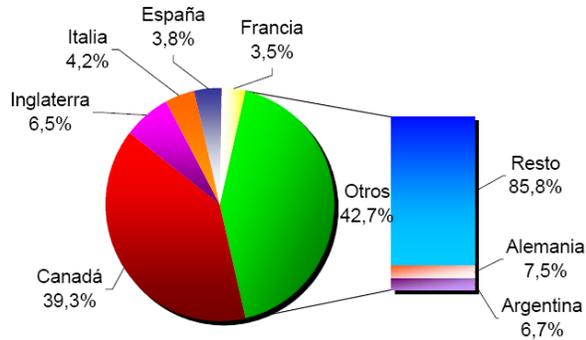


Figura. 7 Principales emisores de turistas a Cuba, enero-agosto 2011.
Fuente: Oficina Nacional de Estadísticas.

Área Geográfica	Población Mundial	Porcentaje
África	14.2%	4.7 %
Asia	56.5%	13.7 %
Europa	12.1%	43.4 %
Medio Oriente	2.9 %	17.4 %
Norte América	5.1 %	71.1 %
Latinoamérica y el Caribe	8.6 %	22.2 %
Oceanía	0.5 %	57.1 %

Figura. 8 Porcentaje de acceso a Internet por continentes.
Fuente: Organización mundial del turismo OMT 2009.

El informe de enero de 2011 del CIDTUR plantea: "... los nativos digitales¹² buscan tanto distinción como personalización, y las encuentran en las comunidades cibernéticas mediante el uso de seudónimos y avatares". Hay diversos sitios que satisfacen esta necesidad y permiten a los usuarios personalizar los servicios en función de cómo usan Internet. El propio informe plantea como tendencias para el turismo de las nuevas generaciones, que las agencias y proveedores de viajes deberán:

- Entender y reconocer: archivar y guardar toda la información y datos de los clientes a fin de ayudarles a administrar sus diferentes identidades (trabajo, esparcimiento, solos, con familia, con amigos, etc.).
- Ser proactivos ante una petición relacionada con viajes: ya no ser simplemente un proveedor de servicio que satisface un requerimiento específico, sino presentarse de una manera más amplia como un socio de viajes, una organización de consultoría que ayuda al cliente a elegir en materia de viajes.
- Filtrar el contenido y ofrecer una respuesta personalizada ante cualquier petición del cliente: desempeñar el rol de recomendador y proporcionar una explicación de las recomendaciones con el fin de cultivar la confianza y lealtad del cliente.
- Desarrollar una red de expertos en viajes, también entre personal de agencias de viajes, que permita suministrar y actualizar el contenido evaluado de esta manera.
- Crear asesores de viajes virtuales para atender las consultas de los clientes 24 horas/7 días. (68)

Otra razón por la que resulta interesante resolver el problema previamente planteado surge de los recursos humanos y los niveles de experticia que estos llegan a alcanzar. Lograr un buen agente de viajes es un proceso de años. Incluso son muy valorados en las principales AAVV o TTOO, los trabajadores con más de 5 años en el empleo, porque se considera que llegan a conocer al cliente y las características de los destinos, logrando un mejor servicio. Sin embargo, cuando estos empleados

¹² Personas nacidas a partir de la década del 90 del siglo pasado. Desde su nacimiento están en contacto con equipos electrónicos en la vida cotidiana.

cambian de empresa o de empleo, por lo general, queda un vacío que demora algún tiempo volver a llenar. Dicho en términos de la IA, se pierde la experticia lograda por esa persona y la empresa deja de poseer ese conocimiento, lo que contrarresta la inteligencia de la organización.

Una tercera razón se presenta a partir de la estrategia de mostrar a los internautas la información más ajustada a su búsqueda. Este es uno de los aspectos en los que se investiga en la actualidad, fundamentalmente las empresas que desarrollan buscadores de propósito general en Internet.

Otra razón importante para dar solución al problema es la integración del sistema con otras aplicaciones informáticas de reservaciones online que ya existen. Este sistema acerca las reservaciones online a la realidad de las oficinas de reservaciones de las AAVV y TTOO. En estas se establece un diálogo entre el agente de viajes y el cliente mediante el cual el agente aconseja al cliente los mejores destinos para sus deseos.

Un software de asistencia a la selección de destinos turísticos en Cuba genera gran cantidad de información relacionada con los gustos y preferencias por países, edades, temporadas, etc. Esta información, analizada usando técnicas de minería de datos y otras técnicas de la IA, genera conocimiento necesario para la toma de decisiones de las empresas. Con lo que se pueden realizar ofertas más agradables basadas en este conocimiento. Es importante señalar que esta información se obtiene sin conocer ningún dato sensible del cliente. La información y el conocimiento que se genera de ella, parten del comportamiento general y no de las individualidades.

La imagen de la figura 9 resume las funcionalidades de un sistema integral de reservaciones. En ella se plantea la posibilidad de obtener conocimiento a partir de la información, enlazar el sistema experto con otro de reservaciones online y mantener contacto con los clientes permanentes a partir del establecimiento de relaciones más cercanas entre las instalaciones y sus clientes habituales.

Las encuestas realizadas por algunos centros de investigación del turismo plantean que varios de los turistas que llegan a Cuba lo hacen a partir de una sugerencia de un amigo o conocido. Esto puede ser visto como una ventaja, teniendo en cuenta que, si las instalaciones cubanas son recomendadas es por la calidad de sus servicios. Sin embargo, no lo es desde el punto de vista de la cantidad de personas a las que llegan estas sugerencias. Por otro lado de acuerdo con Euromonitor International, en 2015 el 50% de los viajeros europeos empleará su smartphone para localizar información turística o hacer reservas a través de la WEB. Lo que da a lugar a que los sistemas recomendadores asuman el papel de amigo a través de Internet (68).

Los portales para el turismo .travel mencionados anteriormente, son desarrollados a partir de la inclusión de colores y formas llamativas y siguiendo el concepto de acumular la mayor cantidad de información que necesita el cliente al alcance de un clic. Estas son buenas prácticas para la programación WEB pero no son suficientes en la era de los nativos digitales. Es necesario que la información sea más personalizada y responda más rápidamente a las necesidades que el cliente busca. Por lo que, como parte del proyecto convenido entre INFOTUR y la Facultad Regional de la UCI en Ciego de Ávila, se

incluye el sistema propuesto como solución al problema de selección de instalaciones hoteleras en Cuba a través de la WEB.



Figura 9 Esquema general de un sistema integrado de reservaciones y atención a clientes.

2.3 Planteamiento del negocio actual.

En la actualidad existen dos variantes fundamentales para que un cliente potencial pueda realizar la elección de una instalación turística cubana, estas son, dirigirse a una agencia de viajes o visitar un portal WEB de promoción turística en Internet. En el primer caso, el cliente solicita un viaje, el agente de viajes pregunta por sus preferencias, el cliente da sus datos, el agente brinda una o varias propuestas según sus gustos, negocian ambos y al final el cliente decide. En el segundo caso el cliente localiza un portal WEB en la red donde encuentra información publicitaria, lee toda esa información hasta hallar el destino adecuado, luego se dirige a una agencia de viajes para reservar, o lo hace en el portal si posee el servicio de reservación en línea.

A partir de las vías planteadas para la selección se representa el negocio mediante dos diagramas. Los diagramas representados son los de actividades y de casos de uso para el negocio. Para la variante en la que el cliente hace uso de los portales promocionales el cliente constituye un único actor del negocio. El actor realiza las acciones de localizar el portal, leer la información y tomar la decisión sobre el destino a visitar. Los anexos 6.a y 6.b muestran los diagramas de actividades y casos de uso para esta variante respectivamente.

En la segunda variante el cliente se presenta en las oficinas de las AAVV distribuidas por el mundo y solicita información o sugerencias para su viaje. En estos lugares es atendido por agentes de viajes con

experiencia que hacen ofertas de acuerdo a las expectativas y deseos de los clientes. De esta forma y a través del diálogo, se va refinando la oferta hasta que el cliente decide el destino a visitar y en ocasiones las actividades a desarrollar durante el viaje. Los anexos 6.c y 6.d muestran los diagramas de actividades y casos de uso para esta variante respectivamente.

2.4 Selección y justificación de los paradigmas de la IA usados.

Con el análisis de los paradigmas de la IA abordados anteriormente se justifica la utilización de un sistema basado en casos (SBCA) a partir de:

- La necesidad de distribuir el conocimiento de los expertos en todos los países donde existen posibles clientes turistas.
- La posibilidad de atender varios clientes a la vez; posibilidad de extender el trabajo de los turoperadores y agentes de viajes 24h al día.
- Permite la separación del sistema en dos partes, máquina de inferencia y base de conocimientos por lo que pueden ser validadas las dos partes con diferentes pruebas.
- La alta complejidad de construir reglas para este tipo de problema.
- Existencia reducida de expertos y por lo general dominan las características de pocos destinos.
- Existencia de casos anteriores (turistas que visitaron antes Cuba y que están caracterizados).
- La necesidad de retener el nuevo conocimiento generado de manera automática.
- La alta velocidad con la que debe operar el sistema en la WEB.
- La posibilidad de justificar al cliente los destinos mostrados.
- El fácil diseño de los casos que provoca mayor comodidad para la representación del conocimiento y mayor entendimiento para los especialistas que operan el sistema.

Los SBCA resuelven un nuevo problema a partir de: recuperar uno o varios casos previos de la base de conocimiento similares al nuevo caso, reusar de manera positiva o inversa, revisar la solución obtenida y retener el nuevo conocimiento incorporándolo a la base de conocimientos (75). Es por ello que el RBC es conocido por el ciclo de las 4R para resolver problemas.

A pesar de las ventajas que ofrece la utilización de un SBCA en la solución, la alta velocidad con la que debe operar el sistema puede verse comprometida con un mal diseño de la base de conocimientos o de los módulos del SBCA. El diseño de la base de conocimientos y de los módulos de acceso, recuperación y adaptación se convierten en pasos críticos.

2.5 Diseño de la base de conocimientos.

Un caso es un problema resuelto. Para el problema definido, un caso está formado por un conjunto de características que representan a un cliente turístico que visita Cuba así como la instalación visitada por él. Las características que describen al cliente son los rasgos o atributos predictores y la instalación visitada el rasgo objetivo.

Los métodos para obtener la información de la base de conocimientos parten de dos vías. Una a través de la información acumulada en las oficinas de reservaciones de diferentes países y de Cuba y la otra de la experticia de los trabajadores de estas áreas. Los rasgos fueron obtenidos a través de una encuesta desarrollada a expertos del turismo (anexo 7). En esta encuesta fueron presentados a los expertos 12 rasgos predictivos, obtenidos de estudios de mercados (76), (77), (78) y de entrevistas con otros expertos. A partir de esta información se obtuvieron 11 rasgos predictores y un rasgo objetivo y los valores de importancia para cada rasgo predictor evaluado entre 0 y 1 (anexo 8). La tabla 4 muestra los rasgos predictores definidos a partir del criterio de los expertos.

Tabla 4. Dominios y funciones de comparación de los rasgos.

Rasgo	Dominio	Función	Tipo
País de origen	Texto	Booleana	Predictor
Edad	[16-32, 33-60. 60>>]	Booleana	Predictor
Sexo	[masc, fem]	Booleana	Predictor
Ha venido otras veces	Booleano	Booleana	Predictor
Compañía en el viaje	[familia, amigos, socios, solo]	Booleana	Predictor
Precio	[lujoso, todo incluido, medio, económico]	Booleana	Predictor
Actividades de turismo	[salud, ocio, negocios, científicos, religiosos, deportivos-culturales, caza, pesca, buceo]	Comparación por partes	Predictor
Motivos del viaje	Sol y playa, naturaleza, ciudad, salud.	Booleana	Predictor
Hemisferio	[norte, sur]	Booleana	Predictor
Fecha	Fecha	Con umbral	Predictor
Vuelo directo del país al destino	Booleano	Booleana	Predictor
Presencia de turoperadores en su país	Booleano	Booleana	Predictor
Destino	[todas las instalaciones de Cuba]	-----	Objetivo

Los expertos encuestados son representantes de la agencias turoperadoras Thomas Cook, Cubatur, Cubanacán, Transat, Hola y Sun, Sunwing, así como especialistas de calidad en instalaciones hoteleras. En total 15 expertos fueron encuestados. Todos cumplen el requisito de llevar más de 5 años trabajando con clientes turísticos de diversos países y tener amplia experiencia en el sector de los servicios turísticos. Las AAVV mencionadas están entre las de mayor actividad en Cuba.

2.5.1 Estructura jerárquica de la base de conocimientos.

Para el diseño de la base de conocimientos se tiene en cuenta la velocidad con que debe operar el sistema, las características de los datos a organizar y el entorno para el que se construye el sistema.

Las bases de conocimiento en los SBCA han sido organizadas de diversas formas (6, 83, 86). Uno de los modelos desarrollados es la memoria dinámica, propuesto por Schank en 1982 (80). En este modelo los casos se organizan en una estructura jerárquica usando índices y episodios generalizadores. Tiene como dificultad que un mismo caso pueda aparecer bajo diversos episodios generalizadores lo que provoca un crecimiento de la base. En el problema actual, la cantidad de casos y por ende el crecimiento de la base son altos, por lo que no es ventajoso el uso del modelo de memoria dinámica.

Otros modelos para organizar la base de conocimientos son:

- El modelo de categorías, características y ejemplos (83).
- Estructura plana o lineal (6).
- Árboles de decisión (86).

Los árboles de decisión forman una estructura jerárquica, lo que permite la organización de la información atendiendo a criterios determinados. La información es estructurada por niveles o capas y para acceder a ella se realiza un recorrido por el árbol, lo que beneficia en velocidad la recuperación de casos semejantes.

Las características de los clientes turísticos de Cuba son modeladas como datos nominales. La tabla 5 muestra los dominios para cada uno de los rasgos. Teniendo en cuenta el entorno WEB en el que debe funcionar el sistema y la rapidez con que deben recuperarse los casos similares, se estructura la base de conocimientos en un árbol de decisión. La información almacenada en una base de datos relacional es usada para construir la base de conocimientos en forma de un árbol de decisión usando el algoritmo ID3 (65) debido a que todos los rasgos toman valores discretos. La estructura jerárquica de los casos se puede describir mediante la figura 10. Este árbol parte del cálculo de la ganancia de información de cada uno de los rasgos. Existen dos tipos de nodos, nodos rasgo y nodos valor. Los nodos rasgo son únicos en cada nivel y de ellos se ramifican tantos nodos valor como elementos de su dominio hayan en la base de conocimientos. En cada nivel se escoge el nodo de mayor ganancia de información. Se crea un nuevo nodo valor por cada elemento del dominio del rasgo existente en la base de casos y se guarda, en este nodo valor, la base con todos los casos que cumplan las condiciones hasta ese nivel.

La ecuación de ganancia de información usada es la planteada por Quinlan en (65):

$$GI = Ent_0 - \sum_{i=1}^{NP} \left(\frac{n_i}{n_0} * Ent_i \right) \quad (II)$$

Donde:

- Ent_0 = entropía del nodo antes de la división.
- Ent_i = entropía de cada partición (fórmula general para dos o más particiones).
- NP = número de particiones.
- n_i = número de instancias pertenecientes a la partición i .
- n_0 = número total de instancias del nodo.

Los nodos rasgo poseen el nombre del rasgo y una lista de nodos valor. Los nodos valor poseen un elemento del dominio del rasgo, un cubrimiento de la base de conocimientos y un apuntador al rasgo con mayor ganancia de información en ese cubrimiento. La raíz del árbol posee la base de conocimientos completa extraída de la base de datos relacional. Luego en la medida que se construye el árbol, cada nodo de tipo valor tiene un cubrimiento de la base de casos con los que cumplen las condiciones de igualdad hasta ese nivel. De esta manera, el recorrido por el árbol permite acceder a un cubrimiento de la base de casos reducido a la cantidad de casos que cumplan las condiciones de similitud al nuevo

problema a resolver. La principal dificultad de esta organización jerárquica es la redundancia de casos. Esta repetición de información puede ser una dificultad dependiendo de la capacidad de almacenamiento en memoria del hardware. Sin embargo, para una base de 1500 casos las pruebas realizadas demostraron que el tamaño de la base de casos está alrededor de los 900 kilobytes.

Otra variante para construir la estructura jerárquica de los casos siguiendo el mismo principio es manteniendo el cubrimiento de la base de casos solo en los nodos hoja. De esta forma la información no es replicada por lo que no hay redundancia de casos y se resuelve la dificultad anterior. Sin embargo, el recorrido para acceder a los casos semejantes de la estructura jerárquica es más complejo. Cuando no exista un camino que describa exactamente el nuevo problema en la estructura jerárquica de casos, el algoritmo para acceder a los casos debe reconstruir la base a partir de los subconjuntos guardados en varias hojas.

Finalmente se decide usar la primera variante debido a que, se prioriza la velocidad en el acceso a los casos semejantes y las capacidades de almacenamiento del hardware dedicado para el sistema son altas.

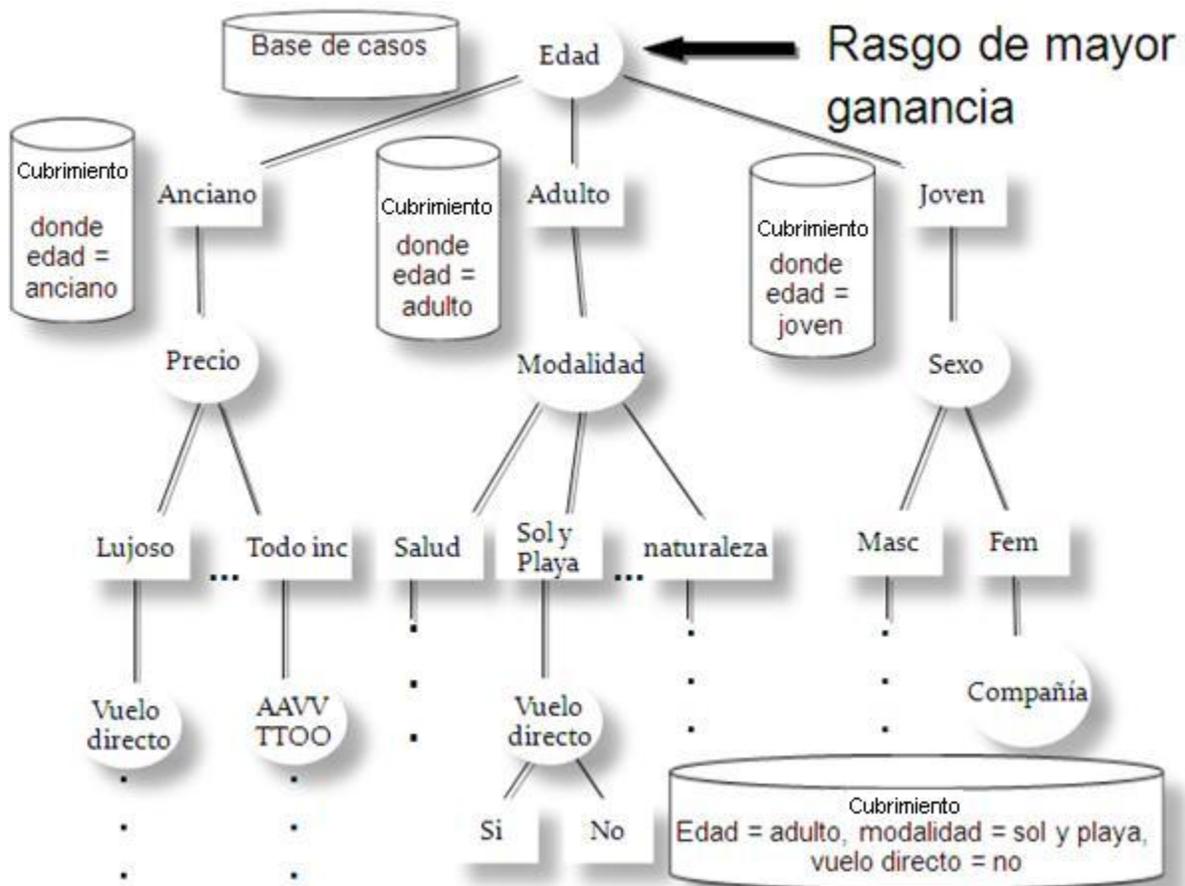


Figura 10 Estructura jerárquica en forma de árbol para la base de casos.

2.5.2 Cálculo de la importancia de los rasgos.

Cada rasgo tiene un nivel de importancia dado por los expertos y tabulado. Este nivel de importancia se calcula a partir de los datos y la opinión de los expertos. Para su determinación mediante expertos se asignaron cuatro niveles de importancia alto, medio, bajo y no interviene en la decisión. A partir de la cantidad de votos dados por cada uno de los expertos en la encuesta, se asignó un valor numérico de importancia entre cero y uno. La tabla muestra los valores de relevancia dados por los expertos así como el coeficiente calculado. La última columna muestra el coeficiente normalizado entre cero y uno.

Los coeficientes normalizados muestran bajos niveles de importancia en la decisión para los rasgos hemisferio, compañía durante el viaje y sexo del cliente. Estos constituyen rasgos posibles a eliminar de los usados para la inferencia.

Tabla 5. Coeficientes de importancia de los rasgos dados por los expertos.

Rasgo	Alto	Medio	Bajo	No interviene	Total	Coeficiente Final
Vuelo directo al destino elegido	13	2	0	0	15	0.1290
Motivos del viaje	12	2	1	0	15	0.1203
Precio	11	4	0	0	15	0.1169
Presencia de turoperadores en su país	11	2	2	0	15	0.1116
Modalidad de turismo	8	6	1	0	15	0.0961
País de origen	9	4	2	0	15	0.0995
Edad	8	4	3	0	15	0.0907
Si ha venido otras veces	5	7	1	2	15	0.0712
Fecha	5	4	3	3	15	0.0625
Hemisferio	0	13	2	0	15	0.0450
Viene acompañado o no	2	6	4	3	15	0.0417
Sexo	0	3	8	4	15	0.0155
Valor numérico de importancia del voto	0.7	0.25	0.05	0	-	1.0000

La opinión de los expertos es uno de los métodos para determinar los coeficientes de importancia para los rasgos. Sin embargo, existen otros métodos que parten de los datos (7):

- Funciones de evaluación dinámica (C. Stanfill, 1987).
- Análisis dimensional (E.L. Rissland y K.P. Ashley, 1998).
- Preferencias heurísticas (J. Kolodner, 1988).
- Grafos de Similaridad (H. Bunke y B. T. Messner, 1993).
- Pesos de una red neuronal artificial (M.M. García y R. Bello, 1997).

Para corregir los coeficientes calculados a partir de la opinión de los expertos, se hace uso de un selector de rasgos de la herramienta WEKA, que obtiene para cada rasgo una ponderación del mismo. Este evaluador calcula la ganancia de información en los rasgos utilizando

“weka.attributeselection.InfoGainAttributeEval”. Los valores de los coeficientes obtenidos se relacionan a continuación (tabla 6):

Tabla 6. Coeficientes de los rasgos calculados mediante WEKA.

No	Nombre	Ganancia de Información
8	Fecha	0.288451
11	Modalidades	0.233837
2	País	0.131729
7	Motivo	0.089359
1	Precio	0.063377
3	Edad	0.06329
4	Compañía	0.061579
9	VueloDirecto	0.028293
6	Repitente	0.016048
10	Turoperadores	0.01247
5	Sexo	0.011567

Existen diferencias en el valor de importancia de los rasgos obtenidos a partir del criterio de expertos y del calculado mediante el preprocesamiento de los datos usando el evaluador de ganancia de información del WEKA. Todos los rasgos, excepto sexo, cambian su orden en el nivel de importancia. La fecha, que para los expertos es poco significativa, según los datos es lo más importante. La existencia de vuelo directo del país de origen al destino turístico y la existencia de oficinas de TTOO o AAVV, que son muy relevantes para los expertos, según los datos son poco significativos. A pesar de las diferencias, ambos análisis demuestran que el rasgo sexo es prácticamente insignificante. Además, entre los seis primeros rasgos se encuentran, motivos del viaje, precio, país de origen, modalidades de turismo y precio; cuatro de los seis primeros coinciden en ambos análisis. Los dos rasgos que no coinciden entre los seis primeros son la fecha y la existencia de vuelo directo. En el caso de la fecha, se trata de un rasgo con una gran cantidad de valores en su dominio, lo que provoca que su ganancia de información sea mayor al de los demás. Por tal motivo la fecha aparece como primer rasgo según los datos obtenidos del WEKA. Sin embargo, en el sistema la fecha es comparada mediante una función que establece rangos de +/- 15 días para la similitud. De tal forma que una fecha es semejante a todas las que están hasta 15 días después y antes que ella. Esto reduce la cantidad de valores en su dominio. El rasgo existencia de oficinas de TTOO o AAVV en el país de origen posee dos valores, es un rasgo booleano, lo que provoca que su ganancia de información sea pequeña, por tener solo dos valores posibles. Aun cuando su ganancia de información es pequeña el rasgo es muy importante de acuerdo con el criterio de los expertos.

Para definir los coeficientes a usar en el sistema se calcula la media entre la opinión de los expertos y los resultados obtenidos de los datos. La media permite equilibrar los resultados obtenidos por ambos criterios, lo que evita que los valores falseados prevailezcan. Por ejemplo, la fecha en el método de determinación de los coeficientes por ganancia de información, está falseado por la gran cantidad de

valores en su dominio; la modalidad de turismo según el criterio de expertos está muy por debajo de su verdadero lugar, esto puede ser debido a que no se logró explicar bien en la encuesta el significado de modalidades (anexo 7).

2.5.3 Criterios de comparación por rasgos y función de similitud.

La construcción de una función de comparación debe ser trabajo conjunto entre los expertos y el ingeniero del conocimiento. Los expertos son personas con experiencia en el área para la que se construye el sistema. El ingeniero de conocimientos es el profesional de la ciencia de la computación que se encarga del diseño del SBC. Para la construcción de un SBCA se encarga además de la selección de rasgos, sus funciones y dominios. En este caso en particular, para desarrollar la función de comparación, no se tuvo en cuenta la existencia de datos difusos, aunque los rasgos edad y país de origen pudieran ser tratados como borrosos. Las empresas turísticas cubanas y los centros de investigación del turismo, realizan estudios de mercados por países, para personalizar sus ofertas a los gustos de los clientes (77, 78, 79). De esta manera han sido determinadas las características y preferencias de los canadienses, quebecuas, rusos, etcétera. Los estudios no son únicamente para países sino para regiones específicas. Tal es el caso de los habitantes de Quebec que por la cantidad de clientes que visitan Cuba, se han realizado estudios de mercados para esa zona. Por lo general, los estudios se hacen atendiendo a divisiones políticas de los territorios. Sin embargo, ¿cuán distinto es un argentino de un uruguayo, en cuanto a las preferencias turísticas? Por el momento no se cuenta con estudios que respondan esta pregunta. Pero en el futuro, con la existencia de estos estudios, se pueden solapar regiones dentro de distintos países según sus preferencias turísticas, por lo que sería necesario tratar el rasgo país como borroso. La edad es un rasgo que ha sido tratado como borroso en varios ejemplos y trabajos de investigación, desde el punto de vista turístico, no existe una edad exacta en la que los gustos varían, lo que justifica el uso de conjuntos borrosos (90). Debido a que no se cuenta con estudios que favorezcan la modelación difusa se tratan mediante una función booleana.

La función de similitud usada para la recuperación de los casos es una suma ponderada de criterios de comparación por rasgos.

$$\beta (P, R_j) = \sum_{i=1}^n W_i \cdot \delta_i(P_i, R_{j,i}) \quad (III)$$

La función δ puede ser diferente para cada uno de los rasgos y se implementa teniendo en cuenta el dominio del rasgo. En el problema actual se usan criterios de comparación booleanos. En la tabla 5 se presentan cada uno de los rasgos, con sus dominios y la función de comparación.

Se adicionan dos rasgos a cada caso que son: selected y candidate. La inclusión de los dos campos responde al criterio natural de que es más recomendado lo que es más conocido. Cada vez que un caso es recuperado entre los más similares se incrementa en uno el campo candidate. Cuando una de las

instalaciones mostradas es escogida por el cliente, se incrementa en uno el campo *selected* del caso recuperado al que pertenecía esta instalación, si la similitud entre el nuevo caso y el seleccionado por el usuario es mayor a 0.99. La función de comparación dada en III se modifica atendiendo a estos rasgos según se muestra en IV. En la función de comparación se incorpora la razón entre las veces que ha sido seleccionado y las veces que ha sido candidato, multiplicado por un coeficiente que es menor que el menor de los coeficientes de los rasgos. A partir de los análisis realizados sobre los coeficientes de importancia para los rasgos, el rasgo sexo resultó ser el menos importante con un valor medio de 0.0135335. Se toma entonces, como coeficiente para la razón entre *selected* y *candidate* $\mu = 0.01$, un valor menor que el menor de los coeficientes de los rasgos pero que permita premiar a las instalaciones más seleccionadas anteriormente por los clientes.

$$\beta (P, R_j) = \frac{(\sum_{i=1}^n W_i * \delta_i(P_i, R_{j,i})) + \mu * \frac{S_j}{C_j}}{(1 + \mu)} \quad (IV)$$

Donde:

W_i : es el coeficiente de importancia en la decisión del rasgo i en el caso j recuperado de la base.

δ_i : es la función de comparación para el rasgo i .

$\frac{S_j}{C_j}$: es la razón entre *selected* y *candidate* del caso j .

μ : es el coeficiente de penalización para beneficiar a los casos más conocidos $\mu < w_i \forall i: 1 \leq i \leq n$.

P_i : es el rasgo i para el nuevo problema.

$R_{j,i}$: es el rasgo i para el j -ésimo caso recuperado de la base.

n : es el total de rasgos.

Un caso puede tener un valor de semejanza para esta función entre 0 y 1 a partir de la normalización que se realiza de la ecuación. Los casos que mayor similitud tienen, incluyendo el coeficiente de uso que se les ha dado, son organizados descendientemente y de esta forma son mostrados como primeros los casos más similares. En principio, todos los casos tienen los campos *selected* y *candidate* en 1. Sus valores se modifican mediante el mecanismo de autoaprendizaje del sistema.

La base de casos accedida en la estructura jerárquica de casos ya posee un grupo de rasgos semejantes a los del nuevo problema. Estos rasgos no son analizados nuevamente para la recuperación por la ecuación de similitud, sino que se comparan con sus respectivas funciones, solo los que puedan ser diferentes.

Para determinar los valores del rasgo objetivo que se muestran al cliente se hace uso de la ecuación de similitud. Los valores del rasgo objetivo son las instalaciones hoteleras de los casos recuperados. La recuperación de los casos semejantes se hace en dos pasos. El primero es el acceso a los casos almacenados en memoria y el segundo la recuperación de los semejantes. El acceso a estos datos se

realiza mediante un recorrido por la estructura jerárquica en la que se almacenan los casos. El recorrido se mantiene hasta que se terminan los rasgos, no se encuentra el valor para un rasgo o existe solo un único valor para el rasgo en el cubrimiento de la base accedido en un nodo. Luego del acceso a los casos se realiza la recuperación mediante la aplicación de la ecuación IV. La ecuación permite ordenar los casos por similitud y los valores del rasgo objetivo de estos casos son presentados al cliente para su selección.

2.6 Conclusiones del capítulo.

El sistema recomendador para el turismo cubano se construye haciendo uso de un SBCA, donde cada caso representa un cliente turístico que ya visitó Cuba y está escrito en función de 11 rasgos predictores discretos y un rasgo objetivo, que es la instalación visitada por el cliente.

La base de casos almacena inicialmente 1500 casos, el crecimiento y refinamiento se logra como resultado de la interacción con los clientes en la WEB.

La organización de la base de casos responde a una estructura jerárquica donde cada nivel del árbol es obtenido usando el algoritmo ID3 y se corresponde con el atributo de mayor ganancia de información. El árbol consta de 11 niveles y cada nodo contiene un cubrimiento de la base de casos que tienen evaluados los rasgos analizados a cada nivel con igual valor.

El acceso de los casos se realiza a través de un recorrido por el árbol y la recuperación de los casos más similares usando una ecuación de similitud, basada en la suma ponderada de criterios de comparación booleanos.

Capítulo 3 Sistema basado en casos, recomendador para instalaciones turísticas cubanas.

Para dar solución al problema se combinan las técnicas de la IA y la publicidad turística para obtener un producto orientado a la WEB. Como resultado de la combinación de estas disciplinas se obtiene un recomendador híbrido que permite sugerir a los clientes turísticos las instalaciones hoteleras de Cuba que más se ajustan a sus necesidades y preferencias, a través de Internet. En este capítulo se muestra el proceso de construcción de cada uno de los componentes del sistema.

3.1 Elaboración de los módulos de recuperación, adaptación y aprendizaje.

3.1.1 Recuperación y adaptación en el sistema basado en casos.

La recuperación de la información en un SBCA es un paso determinante en la solución a un problema. La calidad de la información de una base de casos puede verse afectada si la información recuperada no es la más precisa o ajustada al problema. De esta manera, se manifiesta una estrecha relación entre el diseño de la base de casos y el algoritmo de recuperación de información.

Por otro lado, se tiene como aspecto relevante, que el sistema se concibió para trabajar en un entorno WEB y este tipo de aplicaciones debe atender gran cantidad de solicitudes y dar respuestas rápidas. Por tal razón, el algoritmo de recuperación de información no puede convertirse en un cuello de botella para el sistema.

La recuperación de los casos semejantes se hace a partir de dos módulos. El módulo de acceso recorre el árbol y localiza los casos más semejantes. El módulo de recuperación los ordena mediante una función de similitud.

La base de casos se organiza siguiendo una estructura jerárquica en forma de árbol, donde cada nivel del árbol está en correspondencia con los atributos de mayor ganancia de información. El recorrido para acceder a los casos semejantes parte de la raíz desplazándose, en cada nivel, al nodo cuyo valor sea similar al valor del rasgo en el nuevo caso. Así sucesivamente, hasta llegar a un nivel donde no exista un valor similar al del nuevo caso, exista un solo valor para un rasgo o se terminen los rasgos. Al llegar a la condición de parada se accede al cubrimiento de la base de casos presente en ese nodo del árbol.

Los casos accedidos en esta base son recuperados usando una función de similitud y presentados al cliente. El rasgo fecha es comparado mediante una función booleana con un umbral. El umbral es seleccionado por el ingeniero de conocimientos y puede ser modificado a través de la interfaz de administración. El umbral para comparar las fechas utilizado es de +/- 15 días. De este modo se elimina una cantidad de valores diferentes para el rasgo fecha, lo que evita que sea beneficiado en el cálculo de la ganancia de información. La figura 11 muestra detalles del recorrido por el árbol.

Luego de recuperar los casos semejantes, los SBCA realizan la adaptación de estos para resolver el nuevo problema. No necesariamente debe existir un problema idéntico al actual para solucionarlo, sino que de las experiencias pasadas pueden ser combinadas varias soluciones o partes de ellas.

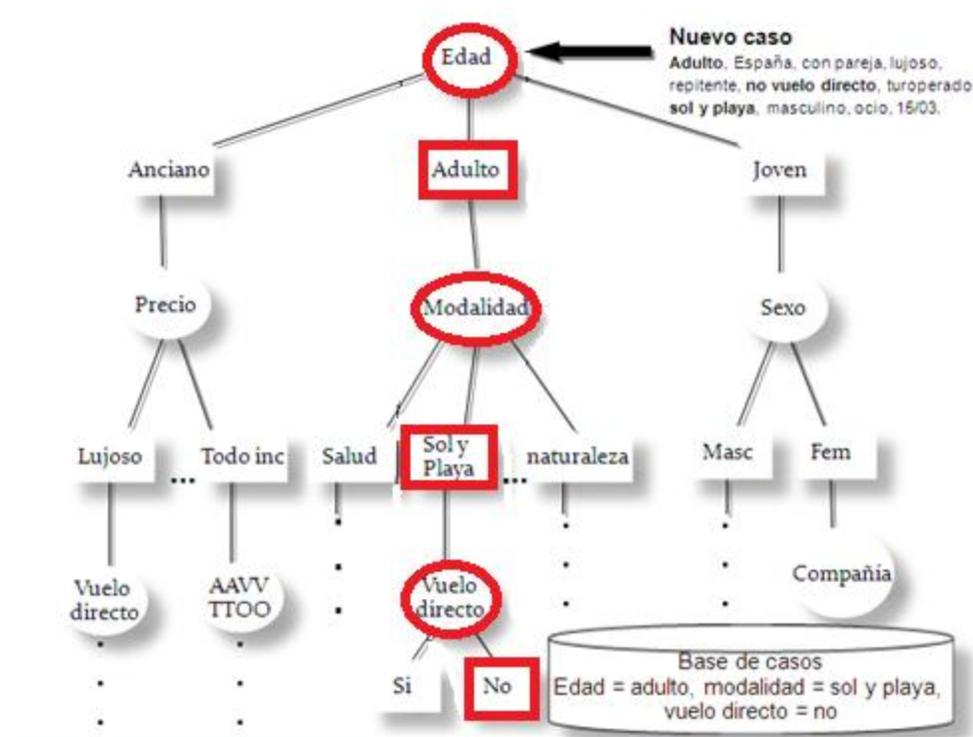


Figura 11 Recorrido por el árbol para recuperar los casos semejantes

En el sistema no se realiza adaptación de casos, se transfiere la solución vieja al nuevo problema directamente según (6) y (7). El sistema no brinda una solución única ni exacta al problema, sino que muestra una lista de los valores del rasgo objetivo correspondientes a los casos más semejantes recuperados, así se reduce el espacio de decisiones del cliente. A partir de este punto se deja al usuario realizar la revisión de la solución. Es el usuario quien escoge el destino que sea de mayor agrado para él de la lista que se le presenta. En la base de casos son accedidos todos los casos semejantes mediante el recorrido por el árbol. Estos son recuperados a través de la función de similitud y se muestran al cliente las 20 instalaciones pertenecientes a los casos más similares. Si el número de instalaciones recuperadas no llega a 20 se muestra una menor cantidad. Si dos casos recuperados tienen el mismo rasgo objetivo, o sea la misma instalación, son contados como uno solo y se continúa con el siguiente caso hasta completar la cantidad. Los casos recuperados con el mismo nivel de similitud se ordenan por aparición en la base de casos. Sin embargo, con la puesta en marcha del sistema los atributos selected y candidate van variando, por lo que se obtienen valores de similitud diferentes para cada caso, aun cuando todos sus atributos sean semejantes. Luego es el cliente quien determina la instalación que más le guste, realizando en este paso la revisión del caso.

3.1.2 Aprendizaje en el sistema basado en casos.

Los SBCA y los árboles de decisión usan fundamentalmente aprendizaje supervisado. Este tiene como meta el procesamiento de un conjunto de datos cuyos atributos han sido divididos en dos grupos. Estos dos grupos muchas veces son llamados predictores y objetivos. El aprendizaje es supervisado debido a

que la distinción entre atributos predictores y objetivos es hecha por el investigador o alguna agencia externa (84), (85). En el problema en cuestión, la selección de atributos predictores y objetivos fue realizada por el investigador, a partir del criterio de los expertos y la información obtenida.

La incorporación de un nuevo caso permite refinar la respuesta y mejorar la rapidez con que se da solución a un nuevo problema. Ya anteriormente se abordó que en este sistema se da como salida al cliente los valores del rasgo objetivo del grupo de casos más semejantes. De estos valores, el cliente selecciona uno que termina de conformar el nuevo caso. Así, el cliente aporta su inteligencia a la base de conocimientos. El nuevo caso se analiza por el sistema y se incorpora, si no tiene similitud superior o igual a 0.99 con el caso cuyo rasgo objetivo fue seleccionado por el cliente, a sus atributos *selected* y *candidate* se les asignan valor 1. La similitud superior o igual a 0.99 significa que el caso es semejante en todos los rasgos, excepto en la razón entre *selected* y *candidate*, de ser así se incrementa en uno el valor de *selected* y no se incorpora el nuevo caso a la base de datos, debido a que ya existe. El rasgo *selected* cuenta las veces que el rasgo objetivo del caso ha sido seleccionado de entre los candidatos. Mientras el rasgo *candidate*, cuenta las veces que ha sido candidato para ser seleccionado, o sea las veces que se ha mostrado a algún cliente la instalación de ese caso. Ambos rasgos presentan como dominio números naturales por lo que no incrementan la complejidad del sistema. Sin embargo, con su uso se premian las instalaciones más visitadas y no se incorporan los casos que ya están.

Los valores *selected* y *candidate* permiten además que el sistema se modifique ante las variaciones de los gustos de los clientes. Con el tiempo, los clientes pueden modificar sus preferencias por una instalación u otra a partir de que reciban mejores atenciones u otras razones. Dos casos con atributos iguales pueden tener distintos rasgos objetivos, esto se debe a que un mismo cliente, o clientes similares, pueden visitar diferentes lugares. Debido a que el valor de *selected* se va incrementando en la medida que una instalación sea más escogida por lo clientes, entonces si en cierto período de tiempo los clientes prefieren una instalación, esto es reflejado por un incremento del valor *selected* en uno de los casos. Finalmente tiene mayor similitud entre dos casos con atributos iguales, el de mayor valor en la razón entre *selected* y *candidate* y por lo tanto es mostrado primero al cliente.

Los casos que se incorporan como parte del autoaprendizaje se añaden a la base de datos, a partir de la cual se crea la estructura jerárquica de casos. Esto hace que el sistema sea más rápido, al realizar una consulta en una base de datos, sin necesidad de recorrer el árbol para encontrar el nodo donde insertar el nuevo caso. El árbol se reconstruye a partir de la información en la base de datos. De esta forma se incluyen los casos nuevos, se actualizan los valores de *selected* y *candidate* y se reorganizan los casos. El sistema permite que el administrador planifique la periodicidad y la hora en la que se realiza el proceso de actualización de la base de casos. El tiempo aconsejado y usado por defecto es diariamente a las 4:00am. El administrador puede modificar la planificación de manera que se reconstruya la base de casos semanalmente o mensualmente.

La figura 12 muestra un diagrama del algoritmo para retener los casos nuevos y el conocimiento aportado por el cliente. En el diagrama se incluye la reconstrucción del árbol debido a que este paso es el que finaliza el proceso de incorporación de conocimiento al sistema.

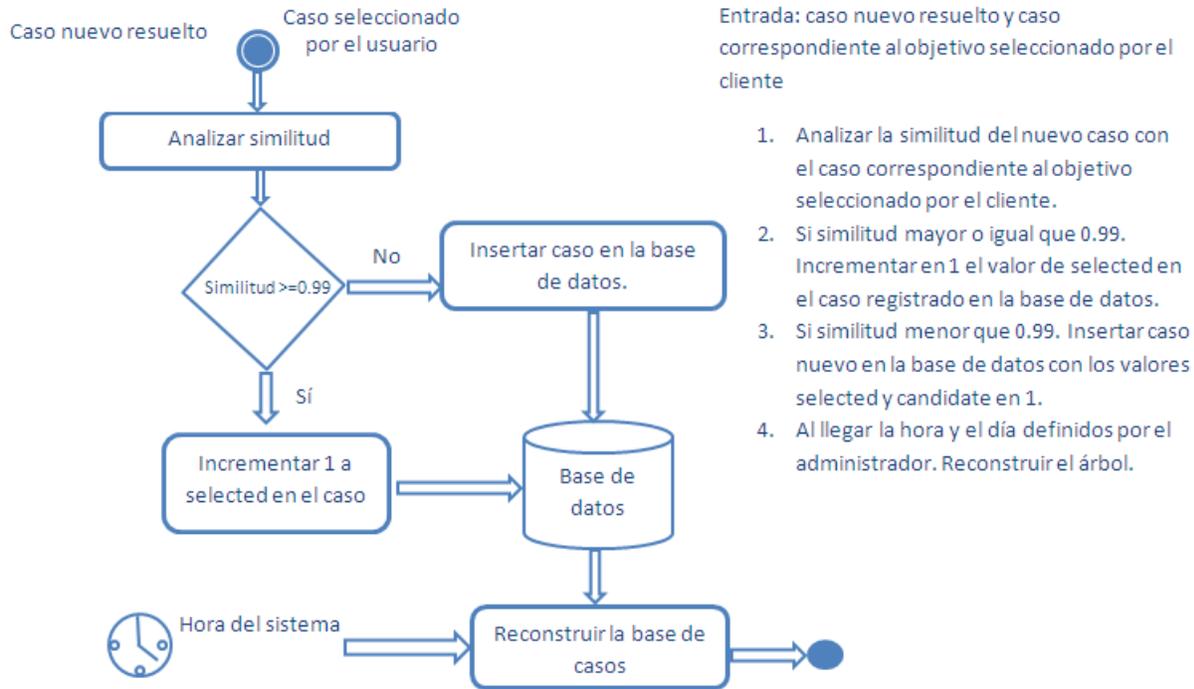


Figura 12 Diagrama de flujo del algoritmo de autoaprendizaje.

3.2 Implementación del sistema en el portal Cuba.travel

Diseñado para funcionar en un entorno WEB y empotrado en un portal construido en PHP, el sistema recomendador para el turismo es construido desde cero sin hacer uso de los *Shells* existentes para este propósito. Existen herramientas computacionales que permiten desarrollar un SBCA de manera eficiente y rápida. Algunos *Shells*, como *Java Expert System Builder* permiten la integración con el lenguaje Java. Durante la investigación desarrollada para seleccionar la arquitectura del sistema no se encontró ningún *Shell* para la construcción de un SBCA que permitiera la integración con PHP y su montaje sobre un servidor Apache, por lo que se utilizan varias herramientas.

3.2.1 Herramientas y tecnologías usadas.

Anteriormente fue planteada como característica del sistema su funcionamiento dentro del portal WEB Cuba.travel. Este portal fue desarrollado sobre la base de un Sistema Manejador de Contenidos (CMS) conocido como Joomla. Por tal motivo y para optimizar el funcionamiento del SBCA en el servidor, este se construye usando las mismas herramientas.

Joomla versión 2.5, es un software que permite gestionar con facilidad cualquier sitio WEB o aplicación en línea. Es una herramienta de código abierto, multiplataforma y libre de utilizarlo. Está escrito en el

lenguaje de programación PHP en su versión 5.0. Es desarrollado y mantenido por una comunidad de usuarios. Se compone de una plataforma y extensiones (componentes, módulos, *plugins*). Existen varios tipos de extensiones, cada una capaz de llevar a cabo tareas específicas. Algunas extensiones se desarrollan como parte de la comunidad de Joomla, otras vienen con la aplicación. Joomla hace uso del servidor de bases de datos MySQL versión 5.1 y del servidor de aplicaciones WEB Apache versión 2.0. Además hace uso del patrón de diseño Modelo-Vista-Controlador, por lo que son organizadas las clases en cada una de estas capas.

Para la verificación del funcionamiento del SBCA y el preprocesamiento de los datos son usadas otras herramientas. Entre ellas se encuentra el sistema *Waikato Environment for Knowledge Acquisition* (WEKA) en su versión 3.6.0 y otras para realizar pruebas de rendimiento del sistema.

3.2.2 Descripción estructural del sistema.

El CMS Joomla permite agregar características o funcionalidades a un determinado proyecto mediante la instalación de extensiones. Teniendo el conocimiento requerido se pueden desarrollar extensiones acorde a las necesidades del proyecto. Para el funcionamiento del SBCA son desarrolladas tres extensiones de Joomla, las tres son creadas como componentes. La figura 13 muestra el diagrama de componentes para el componente de Joomla con el que interactúa el cliente. El diagrama del sistema recomendador completo se muestra en el anexo 13. En él se describen los componentes fundamentales del sistema, tres de ellos son ingeniero, administrador y cliente, los otros dos componentes fundamentales lo constituyen la base de datos y la base de casos. En los tres primeros se hace uso del patrón modelo-vista-controlador, de manera que se accede a los datos a través del modelo. El paquete controlador del cliente contiene dos importantes componentes del SBCA, estos son la máquina de inferencias y el mecanismo de retención de casos o autoaprendizaje.

El primero de los componentes es nombrado en Joomla `com_cliente`. Constituye la interfaz para los clientes interactuar con el SBCA (anexos 9 y 10). Este componente posee además la implementación de la máquina de inferencias y el mecanismo de autoaprendizaje usados para el SBCA.

El segundo componente es nombrado `com_ingeniero`. Constituye la herramienta de trabajo para el ingeniero de conocimientos. Permite manipular los casos y los rasgos y actualizar sus coeficientes de importancia; también permite seleccionar la ecuación de comparación de entre tres posibles, booleana, borrosa y con umbral. La función booleana toma valor uno cuando el atributo en el nuevo problema y en el caso recuperado son iguales y cero en caso contrario (figura 14 a). La función borrosa no se usa aun en esta versión, sin embargo, se incluye en el sistema para su utilización en versiones posteriores en los rasgos país y edad. La función de comparación con umbral es similar a la booleana, aunque permite una diferencia de un valor ϵ definido por el ingeniero de conocimientos (figura 14 b). La función de comparación con umbral es usada en el sistema con el rasgo fecha y $\epsilon = 15$ días, aunque puede ser variado por el ingeniero de conocimientos. El componente `com_ingeniero` permite además actualizar datos en los casos (anexo 11).

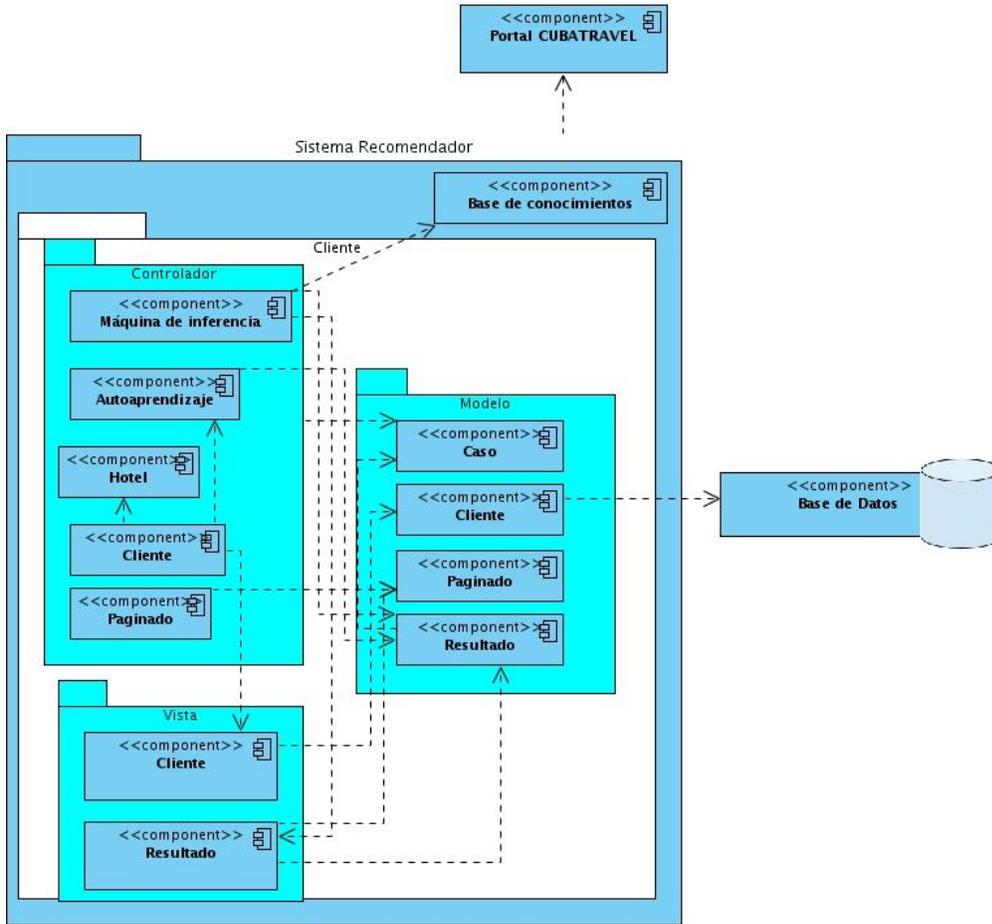


Figura 13 Diagrama del componente cliente del sistema recomendador.

$$a) \delta_i(\alpha, \beta) = \begin{cases} 1 \rightarrow \alpha = \beta \\ 0 \rightarrow \alpha \neq \beta \end{cases}$$

$$b) \delta_i(\alpha, \beta) = \begin{cases} 1 \rightarrow |\alpha - \beta| \leq \varepsilon \\ 0 \rightarrow |\alpha - \beta| > \varepsilon \end{cases}$$

Figura 14 Funciones de comparación booleana, con umbral.

El tercero es el componente com_administrador. Este componente es la interfaz de trabajo para el administrador del sistema. Permite añadir información a la base de datos necesaria para el funcionamiento del SBCA. Por ejemplo, actualizar los vuelos entre ciudades cubanas y de otros países. Actualizar la existencia de oficinas de TTOO y de AAVV en otros países. Actualizar el intervalo de tiempo en que es construido el árbol de decisión. En general, tareas de administración del sistema (anexo 12). En resumen, el sistema informático para la selección de instalaciones hoteleras en Cuba posee tres usuarios, el cliente, el ingeniero de conocimientos y el administrador. Dos de ellos pueden autenticarse y

realizar tareas de administración, el cliente solo puede interactuar con el sistema y aportar su conocimiento.

Cada uno de los componentes con que cuenta el sistema informático construido responde a las necesidades de un actor del sistema. Los actores son el ingeniero de conocimientos, el administrador y el cliente. El cuarto actor del sistema es el tiempo, debido a que cada cierto período de tiempo, definido por el administrador, se reconstruye el árbol de decisión a partir de la base de datos. La tabla 7 muestra los actores con sus descripciones.

Tabla 7. Actores del sistema recomendador para el turismo en Cuba.

Actores del Sistema	Descripción
Ingeniero de conocimientos	Es el encargado de administrar o gestionar toda la información referente a los rasgos y casos.
Administrador del sistema	Este es el encargado de gestionar la información referente a los países, hoteles, modalidades, vuelos directos, turoperadores y las entidades derivadas de estas.
Cliente	Interactúa con la interfaz de usuario principal, inicializa la construcción de un caso nuevo y el proceso de inferencia para un nuevo problema. Ordena además, el autoaprendizaje a partir de la selección de un destino Solicita información o explicación sobre el resultado mostrado.
Tiempo	El tiempo es tomado como un actor debido a que las funcionalidades de la construcción del árbol ID3 se hacen cada cierto periodo de tiempo definido por el administrador, por tanto es el tiempo el que inicializa esta acción.

Los casos de uso del sistema definidos para cada uno de los usuarios del sistema se muestran en la figura 15. Cada uno de los tres usuarios tiene acceso restringido, de manera que solo puede realizar las acciones autorizadas para su nivel. Las acciones autorizadas para cada usuario se relacionan con el rol que desempeñan. El ingeniero de conocimientos puede gestionar (añadir, eliminar o modificar) los casos y los rasgos definidos para el sistema. El administrador del sistema es el encargado de gestionar las informaciones necesarias para completar los casos que no se le preguntan al cliente. Entre estas informaciones se encuentran: vuelos directos entre el país de origen y los destinos, presencia de TTOO o AAVV en el país de origen, datos de los hoteles, entre otras. El administrador del sistema además define el tiempo en el que debe reconstruirse la base de casos, organizada jerárquicamente. Por último el cliente puede construir un nuevo caso, solicitar más información sobre el destino o sobre la inferencia. En la figura 16 se muestra el diagrama de flujo de datos para el sistema recomendador para el turismo en Cuba. En este diagrama se describe el proceso de construcción de la base de conocimientos a partir de un evento generado por el administrador o en una fecha planificada. Se muestra además como proceso principal la construcción de un caso nuevo, seguido de la clasificación de este caso mediante la recuperación de casos semejantes y la aplicación de la función de semejanza para ordenarlos y por último el aprendizaje del sistema, basado en el nuevo caso.

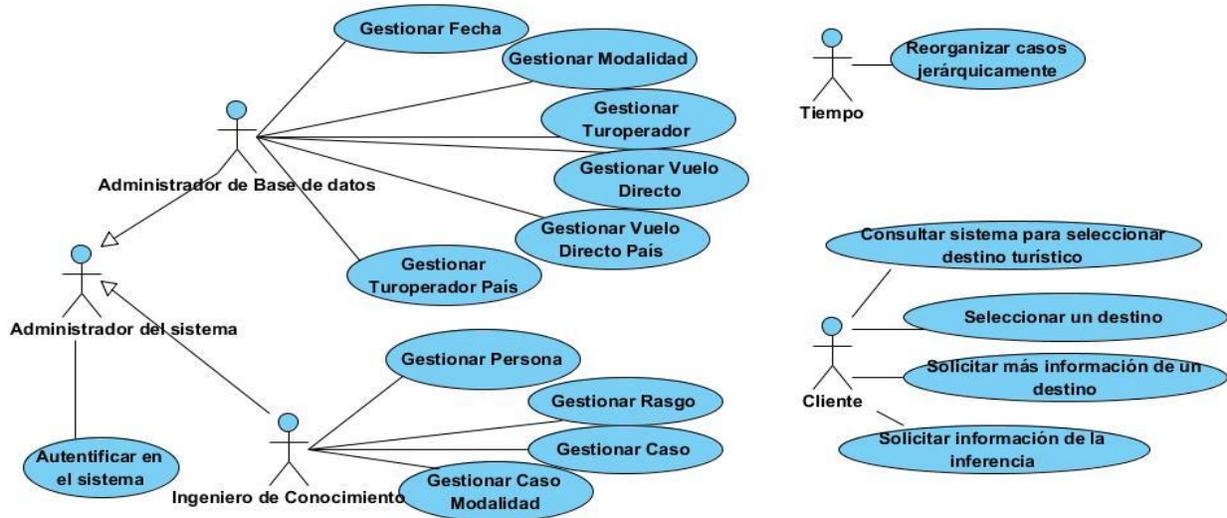


Figura 15 Diagrama de casos de uso del sistema.

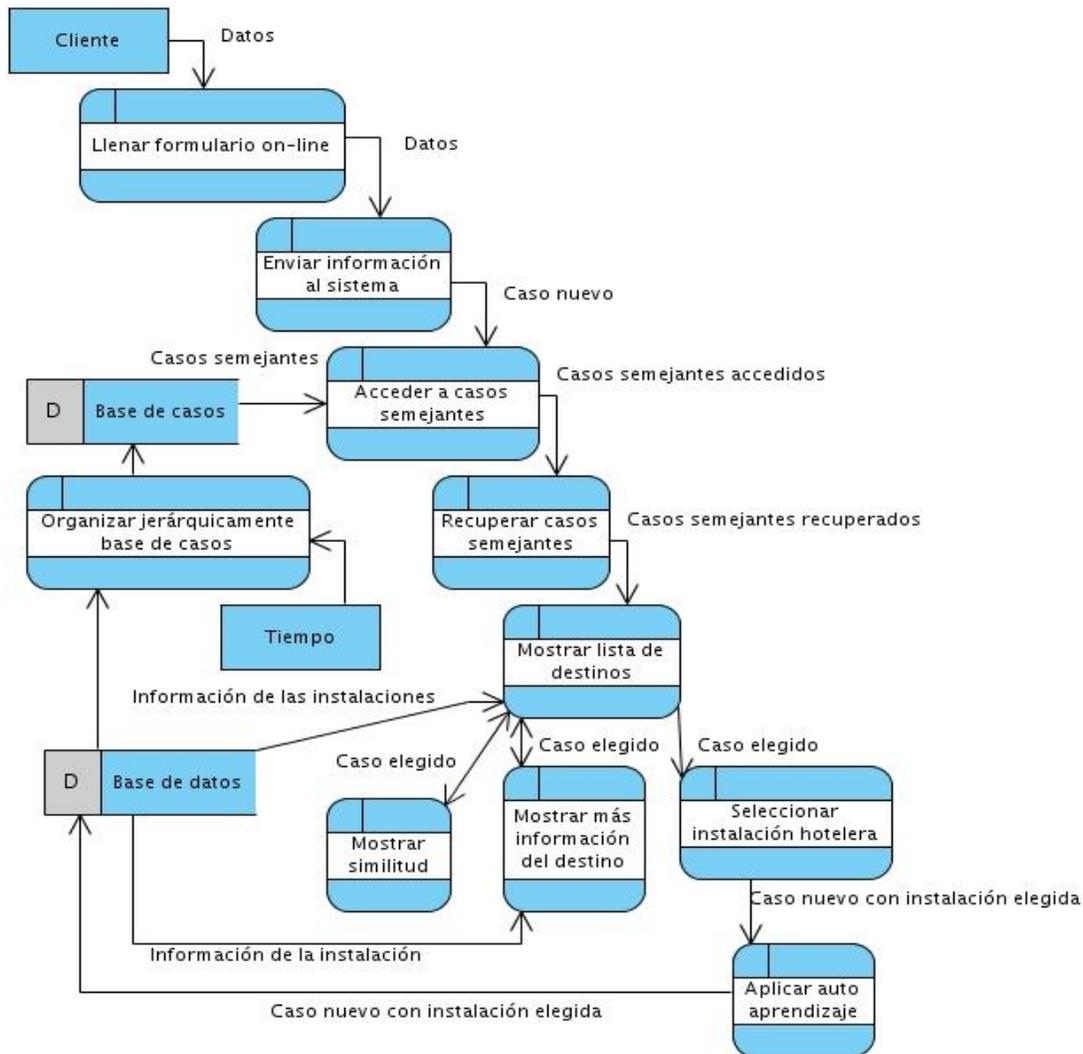


Figura 16 Diagrama de flujo de datos del sistema recomendador

Para mayor comprensión y como parte de la documentación del sistema construido, se muestra en el anexo 14 el diagrama de clases con estereotipos WEB del sistema recomendador para el turismo en Cuba. En este diagrama aparecen las páginas clientes, páginas servidores, los servidores y las clases controladoras para el sistema. Este diagrama corresponde al componente cliente que inicia las acciones principales del sistema.

En el anexo 15 se muestran algunos de los diagramas de colaboración para las principales funcionalidades del sistema. Los diagramas mostrados corresponden a enviar formulario, recuperar casos semejantes, calcular semejanza, aceptar destino, mostrar más información sobre un destino, cancelar destino y ordenar autoaprendizaje.

3.2.3 Forma de interactuar con el sistema recomendador.

El sistema recomendador para el turismo desarrollado brinda tres interfaces de usuarios fundamentales, para el cliente, para el ingeniero de conocimientos y para el administrador del sistema. Los privilegios para cada usuario se garantizan a través del sistema de autenticación del núcleo del CMS Joomla.

Para utilizar el sistema recomendador para el turismo en Cuba desde el punto de vista del cliente se necesita acceder al portal WEB cuba.travel donde está instalado. En la barra de menú del portal se encuentra un acceso rápido nombrado “Experto en destinos”. Este se presiona para acceder a la vista del sistema (anexo 9). En el panel central del portal WEB aparece un formulario en el que se solicitan los datos del cliente para formar el nuevo problema a clasificar. Varios de los datos que son usados en el problema se deducen de otros y no se preguntan. Por ejemplo, a partir del país se deduce si hay vuelo directo o no hasta el destino turístico. El usuario debe llenar el formulario con todos los datos que le solicitan.

Datos Personales

Edad: 33-50 Sexo: M

País: CANADA Desea viajar con: Pareja

Precio: Medio Fecha: 2012-06-22

Motivos del viaje: Negocio Ha venido a Cuba?

¿Qué le gustaría hacer en Cuba?

ARTE	ARTESANIA	CAZA	CIUDAD	DEPORTE	DISCOTECA	EXCURSIONES
<input type="checkbox"/> incluir	<input type="checkbox"/> incluir	<input checked="" type="checkbox"/> incluir	<input checked="" type="checkbox"/> incluir	<input type="checkbox"/> incluir	<input checked="" type="checkbox"/> incluir	<input type="checkbox"/> incluir
HISTORIA	NATURALEZA	NEGOCIOS	PESCA	SALUD	BUCEO	NAUTICA
<input type="checkbox"/> incluir	<input type="checkbox"/> incluir	<input checked="" type="checkbox"/> incluir	<input type="checkbox"/> incluir	<input type="checkbox"/> incluir	<input type="checkbox"/> incluir	<input type="checkbox"/> incluir
PLAYA	RELIGION					
<input type="checkbox"/> incluir	<input type="checkbox"/> incluir					

Enviar

Figura 17 Formulario de entrada de datos para el cliente.

En el formulario se determina el país del cliente a partir de la dirección IP desde donde se conectó el usuario. Sin embargo, se permite modificar este dato para las situaciones en las que un cliente se conecte desde un país diferente al suyo. Una vez completados los datos el usuario debe presionar el botón enviar para realizar la inferencia.

Resultado de la inferencia.

Al enviar el formulario se muestran los datos de las instalaciones turísticas visitadas por clientes anteriores con características similares al actual (anexo 10). En esta interfaz se muestran las instalaciones hoteleras ordenadas por similitud en grupos de cinco, para comodidad de los usuarios. Hacia el final se localiza un paginado que permite avanzar a los próximos cinco hoteles o regresar a los anteriores. Puede que para un cliente aparezcan cinco destinos o menos, en ese caso no aparece el paginado.



Figura 18 Paginado de la interfaz de resultados.

Opciones: Ver más y similitud

Para cada instalación hotelera en la interfaz principal se muestra una foto, el nombre y una pequeña descripción. El cliente puede ver una descripción más amplia (figura 19), la cadena a la que pertenece la instalación, los servicios y cómo llegar a ella, presionando la opción Ver más.

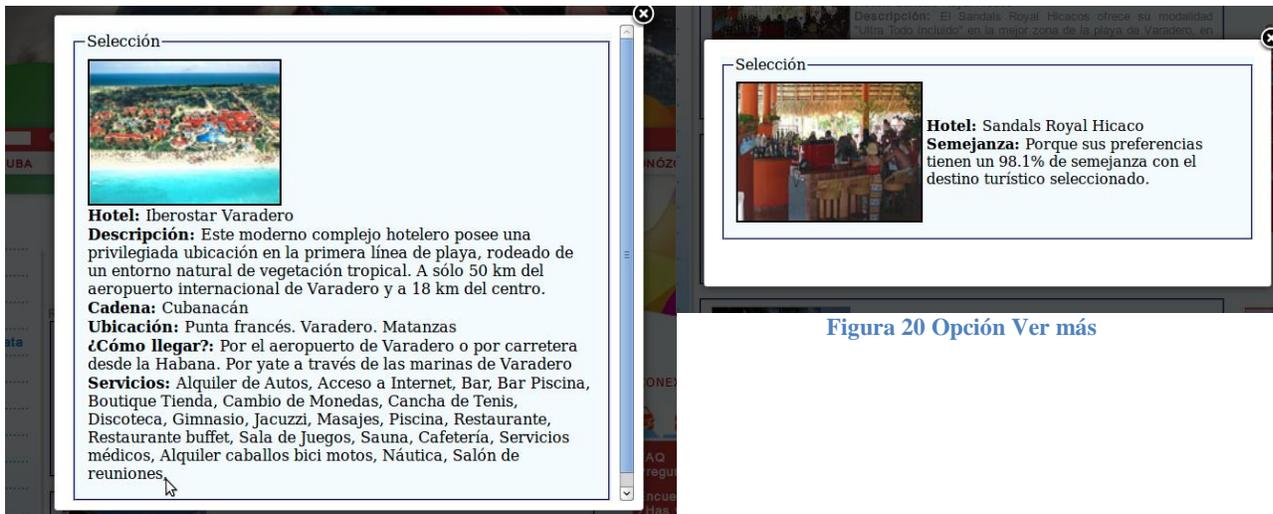


Figura 19 Opción Ver más



Figura 20 Opción Ver más

El cliente puede conocer la similitud de sus características con las del cliente que seleccionó ese destino, presionando la opción “¿Por qué?” En esta opción se muestra la similitud, entre el nuevo problema y el caso correspondiente a esa instalación, expresada en porcentaje (figura 20).

Interfaz para el administrador del sistema

El administrador es un actor que gestiona la información general del sistema recomendador. Se diferencia del ingeniero de conocimientos en que no manipula los casos ni los rasgos. Es el encargado de mantener actualizada la información necesaria para el funcionamiento correcto del sistema. Por ejemplo, para añadir una nueva instalación el administrador accede a la interfaz correspondiente y selecciona la opción “nuevo”, llena la información de la instalación y la añade (anexo 12 b). De igual manera para eliminar o modificar una instalación existente debe seleccionarlo en la lista (anexo 12 a).

El administrador es el encargado de planificar el momento en que se debe actualizar la estructura jerárquica de la base de casos. Para ello selecciona la opción “Construir árbol” e introduce el criterio para actualizar (anexo 12 c). Se puede reconstruir en el momento o planificarlo para que se realice todos los días, semanal o mensualmente en un horario determinado. Por defecto el árbol es reconstruido todos los días a las 4:00am.

3.3 Validación y verificación del sistema.

El sistema recomendador para la selección de instalaciones turísticas a visitar en Cuba incluye entre sus requerimientos el funcionamiento en el entorno WEB. Esto implica que el tiempo de respuesta ante las solicitudes sea rápido. Una de las razones que provoca un tiempo de respuesta lento en las aplicaciones es la alta complejidad computacional de los algoritmos de búsqueda y recuperación de la información. Para el análisis de la complejidad temporal del sistema se tienen en cuenta los algoritmos que tienen interacción con el cliente. Ellos son: el recorrido por la estructura jerárquica de casos para acceder a la base de casos semejantes a un nuevo problema y la recuperación mediante el cálculo de la semejanza de cada uno de los casos accedidos con el nuevo. Estos algoritmos trabajan de forma seriada y ninguno de ellos está anidado en otro, por lo que se puede plantear que:

Complejidad del sistema = max (Complejidad de recorrer el árbol, Complejidad de calcular la semejanza).

$$(V)$$

A partir del análisis del algoritmo para acceder los casos semejantes del árbol en la organización jerárquica de los casos se determina su complejidad como $O(n*m)$, donde n es la cantidad de rasgos y m es la cantidad de valores para los rasgos. En el peor de los casos el recorrido debe llegar hasta el último nivel del árbol. Teniendo en cuenta esto, se puede asignar el valor a n del máximo de rasgos a comparar, por lo que quedaría $O(11*m)$. Esto da como resultado final una complejidad lineal en función de m . La variable m toma su mayor valor para el rasgo “modalidad”. En este rasgo existen 16 posibles valores y las combinaciones entre ellos para distintas cantidades de modalidades. Calculando las combinaciones de las 16 posibles modalidades para valores en 1, 2, 3 hasta 16, m toma valor 78405.

$$\text{Total de Modalidades} = \sum_{i=1}^{k/2} C(k, i) = \sum_{i=1}^{k/2} \frac{k!}{(k-i)!i!} = 78405$$

Donde:

K: Cantidad de combinaciones igual a 16

(VI)

Finalmente la complejidad del algoritmo se puede expresar como $O(11 \cdot 78405) = O(862455)$ que es una complejidad constante.

El algoritmo para el cálculo de la semejanza de cada uno de los casos accedidos realiza en primer lugar un recorrido por todo el arreglo de casos accedidos calculando la semejanza con el nuevo problema. Este recorrido da una complejidad lineal $O(m \cdot p)$ donde p es la cantidad de casos recuperados y m la cantidad de valores para los rasgos. El valor de m puede ser sustituido por el mayor posible, 78405, quedando entonces la complejidad de este paso solo en función de p . Luego de finalizar este paso se realiza un ordenamiento de la lista de casos para organizarlos en función de la semejanza y presentarlos al cliente. Este ordenamiento presenta una complejidad acotada por $O(p \cdot \log_2 p)$. Finalmente la complejidad se puede determinar por el máximo de los dos, obteniendo como complejidad $\max((p \cdot \log_2 p), (78405))$. La base de casos actual contiene 1500 casos. Sustituyendo este valor se obtiene una complejidad de $O(1500 \cdot \log_2 1500) = O(15825)$, menor $O(78405)$ pero en ambos casos constante.

Sustituyendo los valores en V se obtiene:

$$\text{Complejidad del sistema} = O(862455) + O(78405) = O(862455) = O(n \cdot m). \quad \text{(VII)}$$

Por lo que la complejidad temporal no constituye una dificultad para las consultas en la WEB.

Existe un tercer algoritmo que permite la construcción del árbol de decisión cada cierto periodo de tiempo. Por su complejidad se programa para realizarse como una tarea del sistema informático, en los horarios de menor demanda. El algoritmo parte del cálculo del rasgo de mayor ganancia de información. Para ello se recorren todos los rasgos y los casos, lo que aporta una complejidad $O(n \cdot p)$ con p casos recuperados y n cantidad de rasgos. Luego se construyen los nodos hijos con sus bases de casos y se repite el proceso recursivamente hasta que no queden más rasgos, hasta que la base de casos de un nodo determinado tenga un solo caso, o todos los casos en un nodo pertenezcan a una misma clase. Este proceso recursivo posee la mayor complejidad temporal. Su ejecución está en función de la cantidad de valores para los rasgos y de la cantidad de rasgos. La complejidad es exponencial $O(n^m) \cdot O(m \cdot p)$. Este algoritmo no afecta la ejecución de consultas en el sistema por parte de los clientes por lo que no se tiene en cuenta en la determinación de la complejidad.

En una computadora con recursos de procesamiento bajos la actualización de la estructura jerárquica de casos en forma de árbol consume 34 segundos. Este tiempo es muy grande para una consulta a través de la WEB pero pequeño para una actualización.

Como resultado del análisis se puede apreciar que la mayor complejidad del sistema no depende de la cantidad de casos sino de la cantidad de rasgos y valores posibles. Esta complejidad justifica el uso de

una estructura jerárquica de casos que sea actualizada cada cierto período de tiempo. Si además se tiene en cuenta que el algoritmo para construir el árbol se ejecuta solo cada cierto tiempo y en los horarios de menor carga del sistema, entonces se puede plantear que la complejidad de los algoritmos no afecta el desempeño del sistema.

Debido a las limitaciones de almacenamiento que aun presentan las computadoras modernas se analiza además la complejidad espacial del sistema. La mayor dificultad en este sentido aparece en el árbol de decisión. Este árbol posee en cada nodo un cubrimiento de la base de casos con los cumplen las condiciones hasta ese punto. La base de casos actual con 1500 casos, estructurados por niveles en forma de árbol, consume 948 kilobytes de memoria. El crecimiento de la base, caso a caso, permite establecer una equivalencia y suponer que para 500000 casos la memoria usada es de 309 megabytes. Este crecimiento del espacio usado está dado fundamentalmente por la redundancia de casos, lo que puede ser modificado colocando los subconjuntos o cubrimientos de la base de casos solo en los nodos hojas. Sin embargo, esta repetición permite que el acceso a la información sea más rápido y no sea necesario acceder a los casos semejantes de diferentes nodos hojas haciendo el proceso más lento. Teniendo en cuenta además, que el entorno de servidores en los que se ejecuta el sistema tiene una capacidad de almacenamiento en memoria externa e interna muy alta, esto no constituye un problema para el sistema.

Una consulta al sistema en una computadora con las características descritas y funcionando en una red a 100Mbps demora aproximadamente 1,7 segundos. El tiempo se considera alto para un sistema con los requisitos actuales, sin embargo, el entorno real en el que funcionará el sistema recomendador tiene prestaciones muy superiores. Por esta razón se presupone un buen rendimiento.

Para la validación de la calidad de la selección de los rasgos se analizan los datos en la herramienta WEKA. Para su análisis se usa el WEKA Explorer, se diseña una evaluación de los rasgos con el evaluador "weka.attributeSelection.InfoGainAttributeEval". Esta evaluación muestra que los rasgos analizados se ordenan como se muestra en la tabla 8. Los datos mostrados ya están normalizados entre 0 y 1.

A partir de la comparación entre las tablas 5 y 8, que muestran los valores de significación de los rasgos en el cálculo de la semejanza dados por los datos y los expertos respectivamente, se puede apreciar que el sexo tiene una relevancia mínima en ambos casos. Por tal motivo el rasgo sexo puede ser eliminado. La eliminación de un rasgo favorece al sistema recomendador debido a que es una información menos a preguntar al cliente y a analizar por el sistema.

Otra selección de atributos hecha con la herramienta WEKA aporta resultados más cercanos a la realidad. Usando el evaluador "weka.attributeSelection.GainRatioAttributeEval" se calculan los niveles de significación de los rasgos a partir de la ecuación usada para construir un árbol con el algoritmo C4.5 (82), (83). En el anexo 16 se pueden apreciar los resultados del uso de los dos evaluadores. Con estos resultados se puede proponer el uso de un árbol C4.5 para organizar jerárquicamente los casos. A pesar de que todos los rasgos tienen dominios nominales la diferencia en la cantidad de valores posibles entre

ellos hace que unos sean beneficiados sobre otros. El algoritmo C4.5 resuelve esta dificultad conocida como sobreajuste a los datos. Los resultados mostrados en el anexo 16 ilustran que el orden de los rasgos en la estructura jerárquica de casos es diferente para ambos algoritmos, siendo más eficiente el algoritmo C4.5.

Tabla 8. Valores de Ganancia de Información Calculados en el WEKA 3.6

No	Nombre	Ganancia de Información	Ganancia de Información sin rasgo Sexo
8	Fecha	0.288451	0.291827
11	Modalidades	0.233837	0.236573
2	País	0.131729	0.133271
7	Motivo	0.089359	0.090405
1	Precio	0.063377	0.064118
3	Edad	0.06329	0.06403
4	Compañía	0.061579	0.0623
9	VueloDirecto	0.028293	0.028624
6	Repitente	0.016048	0.016236
10	Turoperadores	0.01247	0.012616
5	Sexo	0.011567	0

Desde otro punto de vista se puede apreciar que existen diferencias entre los valores asignados por los expertos y los aportados por los casos. Por tal motivo se decide usar la media entre los coeficientes obtenidos por los datos y los expertos.

Otros aspectos a tener en cuenta para la validación y verificación del sistema permiten plantear que la interfaz para los clientes es fácil de usar, sin requerir entrenamiento extra. No así las interfaces para el administrador y el ingeniero de conocimientos, que si deben dominar el trabajo con el CMS Joomla. La interfaz principal para los clientes se encuentra en dos idiomas, Español e Inglés, no presenta términos desconocidos para los usuarios y permite que los clientes encuentren la similitud con otros usuarios anteriores así como más información de la instalación turística. Una vez aceptado un destino determinado el sistema se enlaza con otro on-line que permite la reservación. El módulo de autoaprendizaje incorporado, permite capturar la información de cada cliente y convertirla en conocimiento. El sistema también combina imágenes con textos para dar una información completa acerca del destino a visitar.

Para la evaluación posterior del sistema se definieron como indicadores de fiabilidad, tres métricas que aportan criterios sobre su evaluación.

1. Cantidad de ofertas aceptadas/cantidad de ofertas rechazadas = **Efectividad**.
2. Cantidad de ofertas aceptadas/cantidad de ofertas hechas = razón de **Impactos**.
3. Cantidad de ofertas rechazadas/cantidad de ofertas hechas = razón de **Rechazos**.

Por último se debe tener en cuenta la curva de crecimiento que tiene el sistema, en la medida que aumente su publicidad si decrece la cantidad de visitas puede ser consecuencia de malos resultados en la clasificación, por lo que se debe revisar el sistema. No todas las visitas al inicio de la puesta en marcha son para realizar reservaciones por lo que los indicadores pueden estar falseados.

3.4 Conclusiones del capítulo

En este capítulo se presentó el diseño e implementación del sistema recomendador para el turismo en Cuba, haciendo uso de un sistema basado en casos híbrido combinado con árboles de decisión, con tres interfaces adaptadas a los casos de uso, garantizando seguridad y persistencia del conocimiento.

Se probó la validez y efectividad del sistema atendiendo a la velocidad de respuesta ante las consultas, la capacidad de almacenamiento de casos y de actualizar el conocimiento almacenado.

Es posible mejorar el desempeño del sistema a partir del autoaprendizaje de los casos resueltos, los cuales son almacenados temporalmente y en períodos de pocas solicitudes se modifica la estructura jerárquica que conforma la base de casos, además se puede actualizar sistemáticamente la base de datos mejorando la información que se tiene sobre vuelos, instalaciones hoteleras y presencia de AAVV y TTOO en otros países.

El sistema se implementó en forma de extensiones para el CMS Joomla, lo que permite su reutilización en otras aplicaciones similares.

CONCLUSIONES

Los objetivos propuestos al inicio del trabajo se cumplen satisfactoriamente y como resultados se puede concluir lo siguiente:

Se realizó el análisis del desarrollo histórico y tendencial de las investigaciones en el campo de la IA y en el uso de esta en los problemas del turismo y en la promoción en Internet, encontrándose un número significativo de investigaciones asociadas al uso de técnicas de la IA para los sistemas recomendadores con tendencias al uso de sistemas multiagentes (SMA), reglas de producción y razonamiento basado en casos (RBC), declarándose ventajas en sus usos.

Se determinaron mediante expertos las principales características a modelar computacionalmente de los clientes turísticos, resultando 11 rasgos predictores y un rasgo objetivo, organizándose en forma de casos cada cliente y usando un árbol de decisión para jerarquizar la base de casos con el objetivo de lograr un acceso más rápido.

Se diseñó e implementó el sistema informático inteligente para la selección de instalaciones turísticas en Cuba vía Internet, obteniéndose como resultado un sistema recomendador híbrido que hace uso de dos técnicas de la IA, RBC y árboles de decisión, estos últimos utilizados para proponer la organización jerárquica de los casos. Este sistema en fase de prueba, se encuentra incorporado al portal promocional del turismo cuba.travel y se cuenta con avales de INFOTUR (anexo 17) que plantean la satisfacción con los resultados.

Se evaluó el sistema a partir del análisis de la complejidad temporal y espacial, velocidad de respuesta del sistema en un entorno de trabajo, la calidad de los coeficientes de relevancias calculados y la eficiencia de la estructura jerárquica para el acceso a los casos, mostrándose niveles de eficacia aceptables o satisfactorios. Usando una base de casos de medio millón de casos la memoria usada es de 300 megabytes. La velocidad de respuesta ante una consulta no supera los dos segundos. Cinco de los seis rasgos más importantes coinciden en la opinión de los expertos y el análisis realizado sobre los datos usando el evaluador de rasgos "weka.attributeSelection.InfoGainAttributeEval"

RECOMENDACIONES

Como resultado de la investigación desarrollada, fueron apareciendo nuevas ideas que, por su complejidad, originaron temas para próximas investigaciones. Además en el proceso de validación del sistema aparecieron nuevas propuestas que ahora se exponen como recomendaciones para futuros trabajos, entre las que se pueden citar:

- Utilizar en la estructura jerárquica para la organización de los casos, la fórmula de ratio de ganancia de información (82) para determinar el atributo de mayor relevancia en cada nivel.
- Implementar otras variantes para calcular los coeficientes de importancia de los rasgos y usarlas en el sistema.
- Evaluar el uso de otras formas de organización de los casos utilizadas en otras investigaciones, por ejemplo redes bayesianas o árboles ordenados por los valores de correlación entre atributos predictores y objetivos (92).
- Investigar la existencia de estudios sociales sobre el turismo para convertir en borrosos los atributos edad y país de origen y usar los conjuntos borrosos en la función de similitud (90).
- Evaluar la eficiencia del sistema usando las métricas elaboradas: efectividad, razón de impactos y razón de rechazos.

REFERENCIAS BIBLIOGRÁFICAS

1. *Actualidad y perspectivas del turismo en el Caribe Insular*. **H., Ayala**. La Habana: Centro de Estudios Turísticos, Universidad de la Habana, 2004, Vols. 1, 2004.
2. **Centro de Estudios Turísticos**. Centro de Estudios del turismo. "“Creciendo en Latinoamérica”". [Online] Universidad de Oriente, 06 08, 2009. [Cited: 06 08, 2009.] http://www.cetur.uo.edu.cu/index.php?option=com_content&task=view&id=74&Itemid=68.
3. "“Síntesis del Turismo en Cuba”", *Síntesis de la Conferencia Magistral en el Aula Magna de la Universidad de la Habana con motivo del 50º Aniversario de la Revolución*. **Fernández., Ramón Martín**. La Habana, Cuba: Universidad de la Habana. <http://www.uh.cu/sitios/turismo/?q=node/>, 19 de diciembre de 2008..
4. "“Un sistema experto para asistir decisiones turísticas, diseño de un prototipo basado en WEB”". **C., Primorac**. Argentina: Revista de investigación en turismo y desarrollo local, (julio/julho 2011), Vols. 4, Nº 10.
5. *La aplicación de sistemas expertos para el aprendizaje organizacional en los hoteles de cinco estrellas de la Comunidad de Madrid*. **Dr. D. Jose Miguel Rodríguez Antón, Dra. María Soledad Celemín Pedroche, Dra. M. Mar Alonso Almeida, Dr. D. Luis Rubio Andrada**. Madrid: VII Congreso "“Turismo y Tecnologías de la Información y las Comunicaciones” Turitec 2008, 2008, Vols. 1, 2008.
6. **D., Gálvez Lío**. *Sistemas Basados en el Conocimiento: una introducción*. Santa Clara, Cuba: Grupo de Inteligencia Artificial, Departamento de Ciencia de la Computación, Universidad Central "“Marta Abreu” de Las Villas., 2006.
7. **Bello Pérez R, Valdivia García Z, García Lorenzo M M, Reynoso Lobato A**. *Aplicaciones de la Inteligencia Artificial*. Núcleo Los Belenes 45000 Zapopan. Jalisco: Universidad de Guadalajara. Centro Universitario de Ciencias Económico Administrativas. ISBN: 970-27-0177-5. Cap. 5 pág. 108-109.
8. *PUBLICACIÓN TRIMESTRAL DE LA REVISTA VISIÓN CIDTUR*. **HOTELERO, PANORAMA**. La Habana, Cuba: CIDTUR. Trocadero No. 55 esq. Prado, Hab. Vieja. Tel. 861-1691. e-mail: cidtur@eaeht.tur.cu, web: <http://cidtur.eaeht.tur.cu>, marzo 2011, Vols. Año 13, No. 1. RNPS 0494, ISSN 1727-1495.
9. Google.com. *Google adSense*. [Online] Google, 2010. [Cited: 03 20, 2012.] <http://www.google.com>.

10. **Amazon.** Amazon.com. [Online] Amazon. [Cited: 03 20, 2012.] <http://www.amazon.com>.
11. *Sistema de reservaciones Online.* [Online] Travels Solution, Costa Rica 2011, 05 10, 2011. [Cited: 01 28, 2012.] http://www.costaricaviajes.com/costa_rica_vacaciones/reservaciones.htm.
12. *Evolución y Tendencias de la Industria Turística.* **Álvarez C, et col.** La Habana, Cuba: Centro de Información y Documentación Turísticas. ISSN 1727-1460 / RNPS 0493, marzo, 2011, Vols. Vol 14, No 1.
13. **Grau R, et col.** “*Metodología de la investigación*”. Biblioteca Central Universidad de Ibagué. Ibagué -Tolima, Colombia: Fondo Editorial Coruniversitaria. ISBN: 958-8028-10-8. Noviembre de 2004.
14. *OBTENCIÓN DE LA COMPONENTE SEMÁNTICA EN TEXTOS.* **Arévalo, Dr. Iván López.** La Habana: VI conferencia Científica UCIENCIA, CINVESTAV - México. 20 Febrero 2012. 978-959.286-019-3.
15. *The Semantic Web.* **Berners Lee T, Handler J, Lassila O.** New York, United State of America: Scientific American 2001, 2001. 284(5).
16. **Yusniel Hidalgo Delgado, Rafael Rodríguez Puente.** LA WEB SEMÁNTICA: UNA BREVE REVISIÓN. *Memorias de la VI Conferencia Científica UCIENCIA.* bienal, marzo, 2012, ISBN 978-959-286-019-3.
17. **R., Neches.** *Enabling technology for knowledge sharing.* s.l: AI Magazine, 1991. 12(3):36-56.
18. **Gruber., T. R.** *A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition.* s.l.: 5(2), pp. 199-220., 1993.
19. **Gennis Roca, Antonio Fumero.** *WEB 2.0.* Madrid, España: Fundación Orange. Licencia Creative Commons de tipo “Reconocimiento-NoComercial-SinObraDerivada”, 2009. Depósito Legal: M-15478-2007.
20. **Weiss, Gerhard.** *Multiagent Systems A Modern Approach to Distributed Modern Approach to Artificial Intellig.* Massachusetts Institute of Technology. Cambridge London, England: The MIT Press p 449-451, 1999. ISBN 0-262-23203-0.
21. **Jorge Giraldo, Alber Ledesma, Hozcar López, Francisco Javier, Jaime Guzmán.** *Vivaldi: Una Arquitectura Multiagente para la Composición de Servicios Web Semánticos en ambientes de información incompleta.* Colombia: Universidad Nacional de Colombia, 2006.

22. "INTELLIGENT AGENTS ON THE WEB: A REVIEW ". **Mohammad S, et col** s.l.: IEEE Copublished by the IEEE CS and the AIP., July-August 2004, Vol. 04. 1521-9615/04/.
23. **ZhaoHui Tang, Jamie MacLennan**. *Data Mining with SQL Server 2005*. Indianapolis, USA: Wiley Publishing Inc, 2005. ISBN-13: 978-0-471-46261-3.
24. **Palma, Julio Amado López**. *MINERÍA DE USO WEB: METODOLOGÍA PARA MEJORAR LA FUNCIONALIDAD DE PLATAFORMAS VIRTUALES DE APRENDIZAJE*. La Habana, Cuba: Memorias de la VI Conferencia Científica UCIENCIA, marzo, 2012. ISBN 978-959-286-019-3.
25. "Querying Heterogeneous Datasets on the Linked Data Web: Challenges, Approaches, and Trends". **Andre Freitas, Edward Curry, Joao Gabriel Oliveira, Sean O'Riain**. no. 1, pp. 24-33, s.l: IEEE Internet Computing, Jan./Feb. 2012, Vol. vol. 16. doi:10.1109/MIC.2011.141.
26. *A WEB-Based Tool For Teaching Neural Network Concepts*. **AYBARS UGUR, AHMET CUMHUR KINACI**. Universitesi Bilgisayar Muhendisligi Bolumu Bornova, Izmir35100, Turkey: Wiley Periodicals, Inc. Comput Appl Eng Educ, 2007, Vol. 18. 449-457.
27. *SmartChoice: An Online Recommender System to Support Low-Income Families in Public School Choice*. **Wilson, David C, et col**. 2, USA: AAAI Magazine, Summer Issue 2009, Vol. 30. 0738-4602-1995.
28. "SISTEMA DE AYUDA AL TURISTA. Modelo para la planificación de un viaje personalizado". **Fernández, Beatriz Rodríguez Díaz. Rafael Caballero**. Universidad de Málaga - España: Estudios y Perspectivas en Turismo., 2012, Vols. Volumen 21 pp.108 – 125.
29. *Design of the Intelligent Real-Time Hypertensive Diagnosis Expert System Based on Web*. **Dai Sheng-hui, Gui Ying, Zhu Xue-qin**. USA: International Symposiums on Information Processing, May 23-May 25 2008. ISBN: 978-0-7695-3151-9.
30. *Web Based Medical Expert System with a Self Training Heuristic Rule Induction Algorithm*. **Ivan Chorbev, Dragan Mihajlov, Ilija Jolevski**. 200, Gosier, Guadeloupe, France: 9 First International Conference on Advances in Databases, Knowledge, and Data Applications, March 01-March 06 2009, Vols. dbkda, pp.143-148. 978-0-7695-3550-0.
31. **Markus Gratzer, Werner Winiwarter y Hannes Werthner**. State of the Art in eTourism. Viena, Austria: E-Commerce Competence Center – EC3 1220 Vienna., 2002.

32. *La aplicación de sistemas expertos para el aprendizaje organizacional en los hoteles de cinco estrellas de la Comunidad de Madrid*. **Rodríguez J.M. et col.** Madrid, España: VII Congreso Turismo y Tecnologías de la Información y las Comunicaciones Turitec, 2008.
33. **ITH, Instituto Tecnológico Hotelero**. www.ithotelero.com. [Online] Instituto Tecnológico Hotelero España, 01 11, 2012. [Cited: 05 25, 2012.] <http://www.ithotelero.com/ith-y-euroconsult-probar%C3%A1n-igreen-el-primer-sistema-de-inteligencia-artificial-para-la-gesti%C3%B3n-y-opt>.
34. **INGENIERÍA, DHARMA**. *Implantación Business Intelligence en Hotel Wellington*. MADRID, ESPAÑA: DHARMA INGENIERIA - Monte Esquinza, 24 - 3º dcha. - 28010, 2009. Tel.: 91 185 22 40 - Fax: 91 391 33 24.
35. **Pitman, A., Zanker M., Fuchs M., Lexhagen M.** Web Usage Mining in Tourism - A Query Term Analysis and Clustering Approach. *U. Gretzel at al. (Eds.): Information and Communication Technologies in Tourism*. Proceedings of the ENTER Conference 2010, 2010, Vols. Springer, 2010, pp. 393-403.
36. **Colineau, N. y Wan, S.** Mobile delivery of customised information using Natural Language Generation. *s.l: Monitor*, 2001. Vol. 26 (3), 27-31.
37. *SISTEMA DE AYUDA AL TURISTA: Modelo para la planificación de un viaje personalizado*. **Rafael Caballero Fernández, Beatriz Rodríguez Díaz**. pp.108 – 125, Madrid: Estudios y Perspectivas en Turismo, 2012, Vol. 12.
38. *HERRAMIENTA PARA LA GENERACIÓN AUTOMÁTICA DE CIRCUITOS TURÍSTICOS ORIENTADA AL COMERCIO ELECTRÓNICO*. **Ismael Alpizar Gonzalez, Erlys de la Nuez Rodríguez**. La Habana: Memorias de la VI Conferencia Científica Uciencia, 2012. 978-959-286-019-3.
39. **Vansteenwegen, P. Souffriau, W. Berghe, G. V. y Oudheusden, D. V.** “The city trip planner: An expert system for tourist”. *s.l: Expert Systems with Applications*, 2011. Vols. 38: 6540-6546.
40. **Kramer, R. Modsching, M. Schulze, J. y Ten Hagen, K.** “Context-aware adaptation in a mobile tour guide. *Dey, A.; Kokinov, B.; Leake, D. & Turner, R. (Eds.): Modeling and using context*. Springer, Berlín.: s.n., 2005. Vols. pp. 210-224.
41. *A Multi-Agent Based Tourism Kiosk on Internet*. **Chris Yeung, Pang-Fei Tung, Jerome Yen**. Hong Kong: IEEE Magazine p 452--461, 1998. 1060-3425/98.

42. **Schmidt-Belz, B. Laamanen, H. Poslad, S. y Zipf, A.** Location-based mobile tourist services - first user experiences. *Springer Computer Science*. New York: Frew, A. (Ed.): Information and communication technologies in tourism, 2003.
43. *SAMAP: An user-oriented adaptive system for planning tourist visits*. **Castillo, L. Armengol, E. Onaindía, E. Sebastián, L. González-Boticario, J. Rodríguez, A. Fernández, S. Arias, J. D. y Borrajo, D.** ESPAÑA: Expert Systems with Applications, 2008, Vol. 34. 1318–1332.
44. *Intelligent travel planning: A MultiAgent Planning System to solve Web problems in the e-tourism domain*. **Camacho, D. Borrajo, D. y Molina, J. M.** ESPAÑA: Autonomous Agents and Multi-Agent, 2001, Vol. 4. 387-392.
45. **Empresarial, Grupo Alianza.** Revista Virtual Grupo Alianza Empresarial. *INTELIGENCIA ARTIFICIAL, BIOCHIPS, MYCIN, SISTEMAS EXPERTOS, REDES NEURONALES*. [Online] Grupo Alianza Empresarial México, 10 2010. [Cited: 02 12, 2011.]
<http://www.grupoalianzaempresarial.com/revistavirtual/inteligenciaartificial.htm>.
46. **Castillo E, et col.** “*Sistemas expertos y modelos de redes probabilísticos*”. Santander, España.: Universidad de Cantabria 39005, Cap 1, pag 3.
47. **Vanessa FERNÁNDEZ, Argenis ROJAS, Esmeralda RAMOS, Haydemar NÚÑEZ, Marcel CASTRO, Claudia CRESSA.** *Método Commonkads para el Desarrollo de un Sistema Experto en Ambiente Web para la Identificación de Especies de Insectos Acuáticos*. Caracas, Venezuela: Laboratorio de Inteligencia Artificial Centro de Ingeniería de Software y Sistemas Escuela de Computación. Universidad Central de Venezuela Caracas, Venezuela, 2010.
48. *Uso de un sistema experto en ambiente web para facilitar la identificación de géneros de bacilos gramnegativos no fermentadores de la glucosa*. **Ivan Flores, Haydemar Núñez, Esmeralda Ramos, Juana Vitelli-Flores, Vidal Rodríguez Lemoine.** Caracas, Venezuela: Revista de la Sociedad Venezolana de Microbiología, 2007, Vols. 27 p 90-94.
49. **Mestizo, Sonia Lilia.** *Implementación en la UV de un Sistema Experto para el apoyo en la solución de problemas en un sistema de educación en línea*. Veracruz, México: Dirección General de Tecnología de Información. Universidad Veracruzana, 2012.
50. **Kolodner J, et col.** “*Case Based Learning AIDS*”. Georgia, USA: Georgia Institute of Technology, Chapter 32.
51. **I., Watson.** “*Applying Case-Based Reasoning: Techniques for Enterprise Systems*”. s.l.: Morgan Kaufmann, 1997.

52. *“Applying Knowledge Management: Techniques for Building Corporate Memories.”*. s.l.: Morgan Kaufmann., 2003.
53. **Yanet Rodríguez Sarabia, María Matilde García Lorenzo, Carlos Morell Pérez, Rafael Bello Pérez, Ricardo Grau Avalo, Rafael Falcón Martínez, Mabel González Castellanos, Mayelis Espinosa García, Adonis Aguirre Roque, Haydee Leyva Castellano, Maybel Martín P.***ConFuCiuS, Sistema Basado en Casos Borroso Conexionista y sus aplicaciones*. Santa Clara, Cuba: XVI Forum de Ciencia y Técnica, 2008.
54. **Cuba, Partido Comunista de.** Lineamientos de la política económica y social del Partido y la Revolución. LaHabana, Cuba: <http://www.cubadebate.cu>, 18, abril 2011.
55. *Case Based Refinement of Knowledge Based Neural Network*. **Gennady Agree, Irena Koprinska**. p 221-226, Bulgaria: In Proceedings of the International Conference; Intelligent Systems: A Semiotic Perspective, 1996, Vol. volume II: Applied Semiotics.
56. *Fuzzy and Neural Hybrid Expert Systems*. **Motohisa Cunabashi, Akira Maeda**. Tokio, Japon: IEEE Expert Magazine, 1995, Vol. Agosto. 0885-9000.
57. *Explicación Basada en Casos utilizando Conjuntos Borrosos para un Sistema Experto Conexionista*. **Yanet Rodríguez Sarabia, Maria Matilde García Lorenzo, Rafael E. Bello Pérez, Jesús R. Falcón Martínez**.La Habana, Cuba: Convención INFORMATICA 2005. VIII Congreso de Nuevas Tecnologías y Aplicaciones Informáticas, 2005.
58. *Sistema experto para el diagnóstico del embarazo ectópico*. **Yanet Rodríguez Sarabia, Karina L. Fernández Sánchez, Maria M García Lorenzo, Zoila Zenaida García Valdivia**. La Habana: 1er Simpósium Cubano de Inteligencia Artificial. Convención Informática 2004, 2004.
59. *Conversational Case Based Reasoning*. **David W Aha, Leonard A Breslow, Héctor Muñoz-Ávila**. 1, Bosoton: Applied Intelligence. Kluwer Academic Publisher Boston. Manufactured in Netherland, 1999, Vol. 25.
60. *Genetic Algorithms for Open Shop Scheduling and Re-Scheduling*. **Sushil Louis, Zhijie Xu**.Reno, USA: ISCA 11th Intl. Conf. on Computers and their Applications, 1996, Vols. p 99-102.
61. *Selecting and weighting features using a genetic algorithm in a case-based reasoning approach to personnel rostering*. **Gareth Beddoe, Sanja Petrovic**.Nottingham, United Kingdom: European Journal of Operational Research, April 2004.

62. *Integrating Case-Based and Rule-Based Reasoning to Meet Multiple Design Constraints*. **C. R. Marling, G. J. Petot, L. S. Sterling, Eleanor Eckstein**. 3, Ohio, USA: Computational Intelligence, 1999, Vol. 15. p 308--332.
63. *Enhancing the effectiveness of interactive case-based reasoning with clustering and decision forests*. **Qiang Yang, Jing Wu**. Netherland: Applied Intelligence Kluwer Academic Publisher., 2001, Vol. 14. p 49--64.
64. *Aprendizaje de árboles de decisión*. **UNSL., Departamento de Informática**. San Luis, Argentina: Curso: Aprendizaje Automático y Minería de Datos. Universidad Nacional de San Luis (UNSL), Octubre 2006.
65. *Induction of Decision Trees*. **Quinlan, J. R.** p 81-86, Boston: Kluwer Academic Publishers, Boston - Manufactured in The Netherlands, 1986.
66. *Tema 10. Árboles de Clasificación*. **Pedro Larrañaga, Iñaki Inza, Abdelmalik Moujahid**. País Vasco: Departamento de Ciencias de la Computación e Inteligencia Artificial. Universidad del País Vasco. Euskal Herriko Unibertsitatea.
67. **C., Alvarez**. Las diez mayores cadenas hoteleras del mundo aceleran su crecimiento. [Online] 11 2009. http://www.hosteltur.com/noticias/35168_diez-mayores-cadenas-hoteleras-mundo-aceleran-su-crecimiento.html.
68. *Panorama turístico internacional en más de cinco décadas: 1950-2005*. **J., Pichardo**. La Habana: Visión CIDTUR., 2006. RNPS 0494, ISSN 1727-1495.
69. **Goldberg, Imri**. plnnr.com/city/barcelona/. *Plnnr: Your personal tour guide*. [Online] 06 2012. <http://plnnr.com>.
70. **Onmyplan**. Granadatur. *Smartourism Granada*. [Online] GRANADATUR. Ayuntamiento de Granada, 10 2011. <http://www.grnadatur.com/en/city-guide/visit-granada-we-plan-your-trip-to-granada/>.
71. *Personalización de Aplicaciones en OO-H*. **Cristina Cachero, Irene Garrigos, and Jaime Gómez**. Alicante España: Universidad de Alicante. España, 2012.
72. *MultiAgent Tourism System: An Agent Application on the Tourism Industry*. **Juan Sebastian Lopez, Fabian Andres Bustos**. Ribeirao Preto, Brasil: Proceedings of the International Joint Conference IBERAMIA/SBIA/SBRN 2006 - 1st Workshop on Industrial Applications of Distributed Intelligent Systems (INADIS'2006)., 2006. ISBN 85-87837-11-7.

73. *CRUMPET: CREATION OF USER-FRIENDLY MOBILE SERVICES PERSONALISED FOR TOURISM*. **Stefan Poslad, Heimo Laamanen, Rainer Malaka, Achim Nick, Phil Buckle, Alexander Zipf**. London, United Kingdom: Queen Mary, University of London, United Kingdom. European Media Laboratory GmbH, Germany, 2002.
74. **James A. Freeman, David M. Skapura**. *Neural Networks Algorithms, Applications and Programming Techniques*. Huston, United States of America: Addison Wesley, 1991. 0-201-51376-5.
75. **Enrique Castillo, José Manuel Gutiérrez y Ali S. Hadi**. *Sistemas Expertos y Modelos de Redes Probabilísticas*. Universidad de Santander, España: s.n., 1997. 39005.
76. *Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches*. **A. Aamodt, E. Plaza**. pp. 39-59., s.l: AI Communications. IOS Press, (1994), Vol. Vol. 7: 1.
77. *ESTUDIO DE MERCADO DE LA DEMANDA TURÍSTICA*. **SERVICIO NACIONAL DE TURISMO**. Santiago de Chile, CHILE: SERVICIO NACIONAL DE TURISMO, DEPARTAMENTO DE PROMOCION, 2008.
78. **Prieto, Rosendo Bermúdez**. Aproximación al perfil de los turistas chinos y rusos. Varadero, Cuba: Escuela de Hotelería y Turismo, Varadero, 2010.
79. **Aloida Rodríguez Alvarez y Benigno Pérez López**. Estudio de mercado, Canadá. La Habana, Cuba: Centro de Información y Documentación Turísticas (CIDTUR), 2011.
80. *Dynamic memory; a theory of reminding and learning in computers and people*. **Schank, R**. Cambridge: CambridgeUniversity Press, 1982.
81. **Bareiss, Ray**. PROTOS; a unified approach to concept representation, classification and learning. *Ph.D. Dissertation, Technical Report*. Texas, USA: University of Texas at Austin, Dep. of Computer Sciences., 1988. AI88-83.
82. **Quinlan, J Ross**. C4.5: Programs for Machine Learning. California, USA: Morgan Kaufmann, San Mateo, California, 1993.
83. **Ron Kohavi and J Ross Quinlan**. Decision Tree Discovery. San Mateo, California, USA: Data Mining Blue Martini Software 2600 Campus., 1999. CA 94403.
84. **William J. Raynor, Jr**. *The International Dictionary of Artificial Intelligence*. New York, ISBN: 0-8144-0444-8: Glenlake Publishing Company. Ltd., 1999.

85. “*Conferencia Aprendizaje con árboles de decisión*”. **Morales Bueno R, Ramos Jiménez G.** Málaga, España: Dpto. Lenguajes y Ciencias de la Computación, Escuela Técnica Superior de Ingeniería Informática, Universidad de Málaga, Programa Doctoral., 2007.
86. *Enhancing the Effectiveness of Interactive Case-Based Reasoning*. **Qian Yang, Jing Wu.** 49–64, Toronto and Burnaby, Canadá: Applied Intelligence. Kluwer Academic Publishers. Manufactured in The Netherlands., 2001, Vol. 14.
87. **Frédéric Peuret.** *Case Based Travel Agent*. A dissertation submitted to the University of Dublin, Trinity College, in partial fulfilment of the requirements for the degree of Master of Science in Computer Science. Department of Computer Science, Trinity College Dublin, September 1999.
88. **Bergmann R, et col.** *Case-Based Reasoning Applied to Planning*. Reporte de Investigación culminada. University of Kaiserslautern, Germany, 1998.
89. **García-Crespo A, et col.** *Sem-Fit: A semantic based expert system to provide recommendations in the tourism domain*. Computer Science Department, Universidad Carlos III de Madrid, Av. Universidad 30, Leganés 28911, Madrid, España.
90. **Pérez, Carlos A Morell.** *Extensiones al razonamiento basado en casos para su aplicación en la planificación de procesos*. Santa Clara : Centro de Estudios Informáticos. Universidad Central "Marta Abreu" de Las Villas, 2005. Tesis en opción al grado de Doctor en Ciencias Técnicas.
91. **Camacho D, Borrajo D, Molina J.** *Intelligent Travel Planning: A MultiAgent Planning System to solve WEB Problems in the e-Tourism domain*. Autonomous Agent in Multiagent system, vol 4, 387-392. Kluwer Academic Publisher. Netherlands 2001.
92. **Gutiérrez, I.** *Un Modelo para la Toma de Decisiones usando Razonamiento Basado en Casos en condiciones de Incertidumbre*. Departamento de Ciencias de la Computación. Santa Clara, UNIVERSIDAD CENTRAL “MARTA ABREU” DE LAS VILLAS. **PhD:** 100. 2003. Cuba.

Anexos.

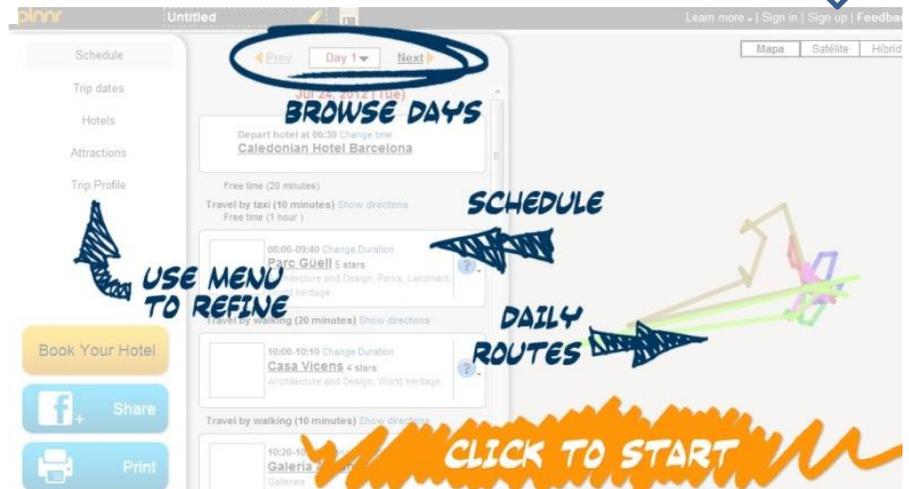
Anexo 1. Listado de portales WEB que promocionan el turismo cubano analizados en el trabajo.

Portal WEB	Descripción
www.conocecuba.com.mx	Portal de promoción el turismo cubano hospedado en México.
www.solwayscuba.com	Portal oficial de la agencia de viajes de Solways con negocios en Cuba.
www.cubanparadises.com	Constituye el portal WEB oficial de la Corporación Turística del Caribe COTURCA S.A. con casa matriz en República Dominicana.
www.guiacuba.info	Portal de promoción el turismo cubano hospedado en México.
www.aircanada.com	Portal oficial de la aerolínea Air Canadá con promoción al turismo cubano hospedado en Canadá.
www.flightsincubareservations.com	Portales de promoción el turismo cubano hospedados en Canadá.
www.guajira.com.ar	Perteneciente a Agencia de Viaje con el mismo nombre y con negocios en Cuba.
www.albatour.es	Perteneciente a Agencia de Viaje con el mismo nombre y con negocios en Cuba.
www.westjet.com	Perteneciente a Agencia de Viaje con el mismo nombre y con negocios en Cuba.
www.sunwing.com	Perteneciente a Agencia de Viaje con el mismo nombre y con negocios en Cuba.
www.nolitours.com	Perteneciente a Agencia de Viaje con el mismo nombre y con negocios en Cuba.
www.dtcuba.com	Portal de la empresa Vacacionar Travels que permite la realización de reservaciones y pago on-line.
www.infotur.cu	Portal de la Oficina de Información Turística.
www.cubatravel.cu	Sitio oficial del turismo cubano.
www.havanatur.cu	Portal de la agencia cubana Havanatur.

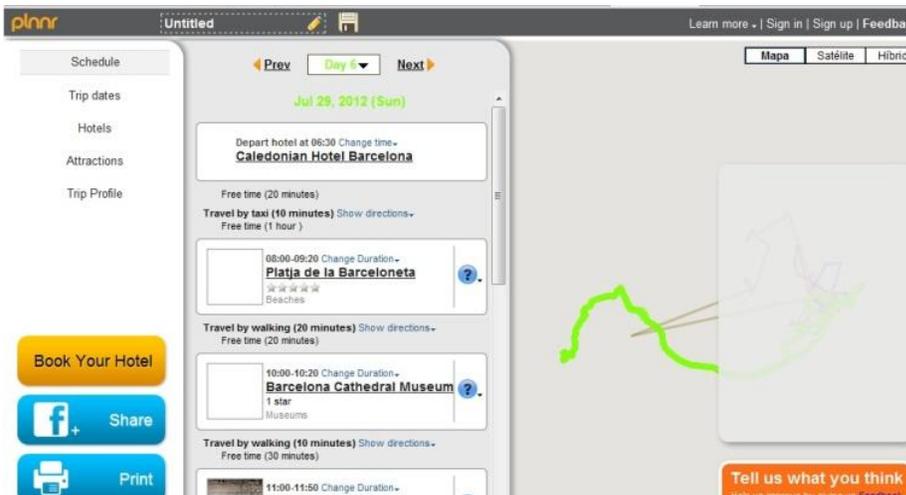
Anexo 2. Secuencia de imágenes de un planificador de viajes para Barcelona y otras ciudades del mundo.



Paso 1 Ajuste de características del viaje



Paso 2 Instrucciones para navegar por las ofertas.



Paso 3 Navegar para chequear las ofertas hechas por el sistema, guardar la planificación, reservar hotel.

Anexo 3. Tabla de arribo de turistas por motivo de visita a Cuba.

15.1 - Series de base sobre el turismo / Basic series on Cuba's tourist industry

	Miles					
CONCEPTO	2005	2006	2007	2008	2009	2010
Llegadas de turistas por motivo de visita						
Vacaciones, ocio y recreo	1.982	1.931	1.989	2.190	2.289	2.397
Negocios y motivos profesionales	12	13	12	12	12	12
Otros	267	206	118	114	104	98

Fuente: Oficina Nacional de Estadísticas (miles de visitantes)

Anexo 4. Tabla de arribo de turistas por regiones.

15.1 - Series de base sobre el turismo / Basic series on Cuba's tourist industry

	Miles					
CONCEPTO	2005	2006	2007	2008	2009	2010
Llegadas de visitantes por región						
África	6	7	7	8	9	9
América	1.216	1.149	1.173	1.380	1.536	1.664
Europa	1.048	1.014	924	909	838	810
Asia Oriental y el Pacífico	42	44	42	44	41	41
Asia Meridional	5	5	4	5	4	6
Oriente Medio	2	2	2	2	2	2
Total	2.319	2.221	2.152	2.348	2.430	2.532

Fuente: Oficina Nacional de Estadísticas (miles de visitantes)

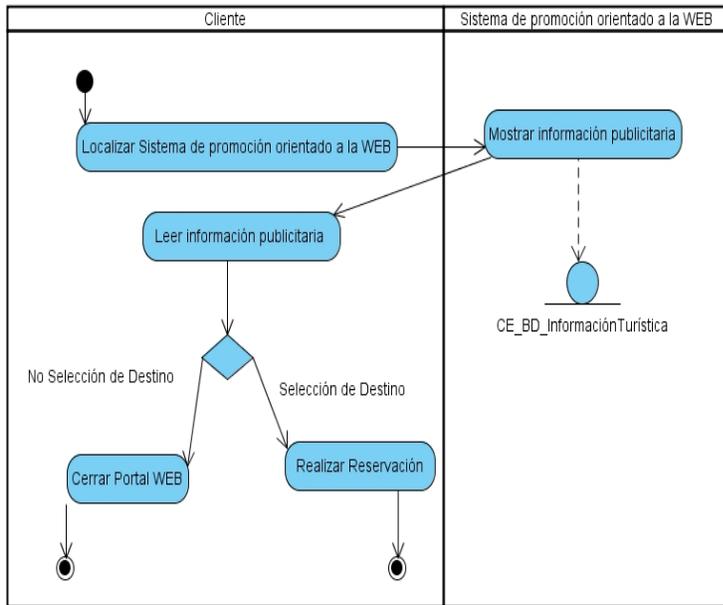
Anexo 5. Tabla de cantidad de instalaciones hoteleras por tipo.

15.9 - Capacidad de los hoteles y otros establecimientos seleccionados *Hotels and other selected resorts accommodation rates*

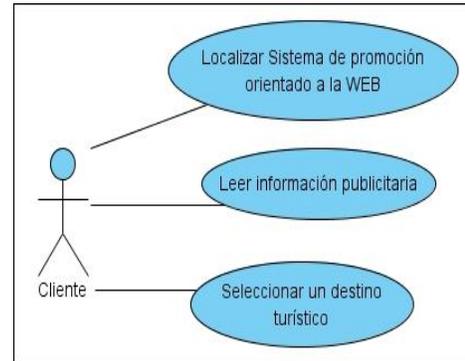
CONCEPTO	Unidad					
	2005	2006	2007	2008	2009	2010
Número de establecimientos						
Hoteles y otros establecimientos	464	451	449	427	438	447
Hoteles	371	374	371	360	371	380
5 Estrellas	24	28	26	27	29	29
4 Estrellas	70	79	83	78	81	86
3 Estrellas	80	89	92	95	97	108
2 Estrellas	70	67	65	62	68	68
1 Estrella	127	111	105	98	96	89
Otros establecimientos	93	77	78	67	67	67
Moteles	79	63	64	54	54	54
Hoteles-Apartamentos ^(a)	2	2	2	2	2	2
De ello: Aparthoteles	1	1	1	1	1	1
Hostal	12	12	12	11	11	11

Fuente: Oficina Nacional de Estadísticas.

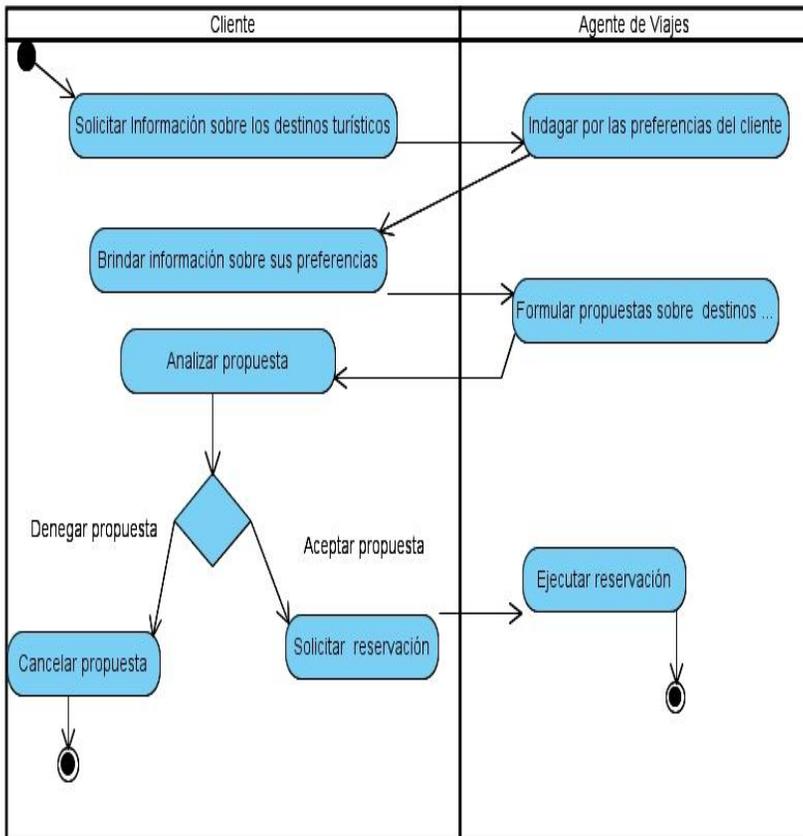
Anexo 6. Diagramas de actividades y casos de uso del negocio para las variantes de selección de destino a través de AAVV y portales WEB.



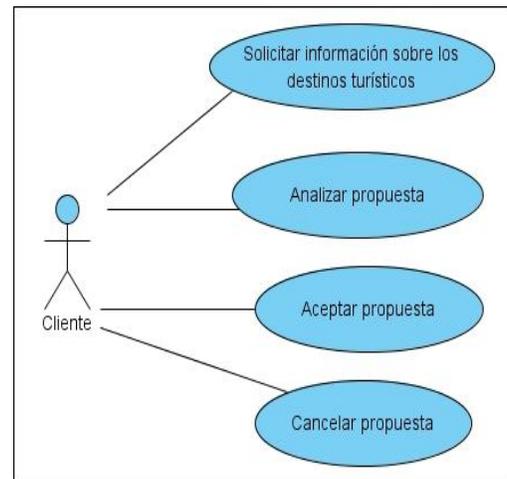
a. Diagrama de actividades para portales WEB.



b. Diagrama de casos de uso para portales WEB.



c. Diagrama de actividades para las AAVV.



d. Diagrama de casos de uso para las AAVV.

Anexo 7. Pregunta 1 de la encuesta hecha a los expertos en turismo.

Encuesta para el Sistema de promoción turística automatizado para la WEB.

La información aquí recolectada se utilizará para desarrollar un sistema computacional que permita sugerir a un cliente turístico desde internet el destino cubano ideal para él según sus características y preferencias.

Solicitamos su atención por unos minutos para responder las siguientes preguntas:

1. Marque con una (X) las características de los turistas que determinan que vayan a un destino u otro del archipiélago cubano. (Ver ejemplo)

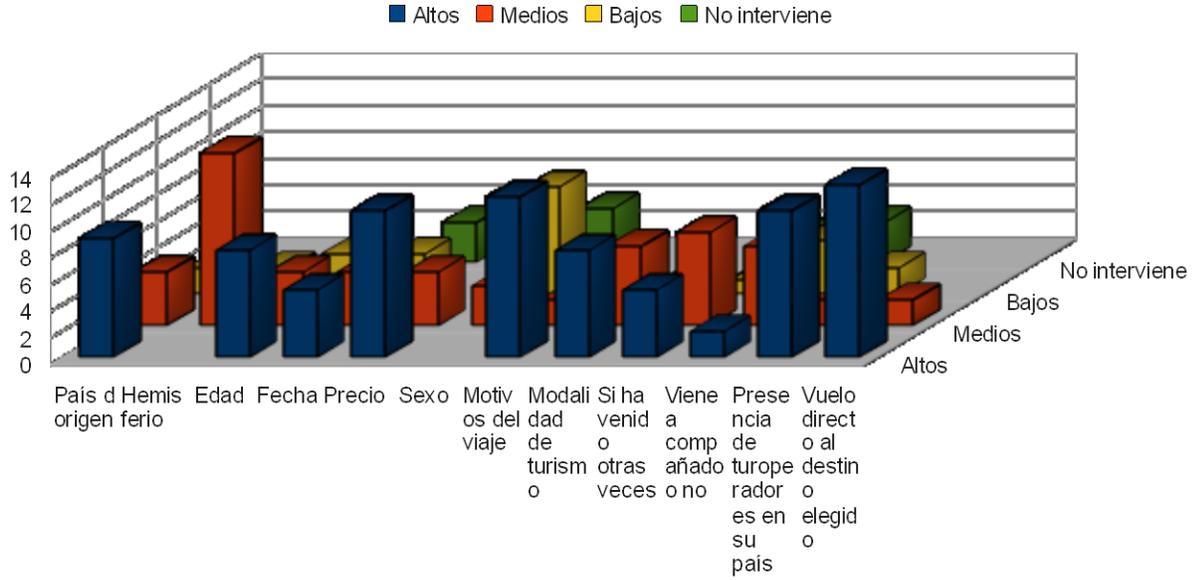
Rasgo	Nivel de importancia en la decisión de ir a un destino u otro			No interviene en la decisión
	Alto	Medio	Bajo	
País de origen				
Hemisferio				
Edad				
Fecha				
Precio				
Sexo				
Motivos del viaje				
Modalidad de turismo				
Si ha venido otras veces				
Viene acompañado				
Presencia de turoperadores en su país				
Vuelo directo de su país al destino elegido				

Si considera otras características las puede añadir en los espacios en blanco

Ejemplo (Se conoce que las personas de más de 60 años prefieren los destinos tranquilos de sol y playa, por lo que prefieren los cayos que la ciudad de la Habana. En este caso la edad es un rasgo que determina la decisión de un cliente).

Anexo 8. Tabulación de los datos de la pregunta 1 de la encuesta hecha a expertos del turismo.

Criterio de los expertos sobre el nivel de decisión de cada rasgo



Anexo 9. Interfaz inicial del componente cliente.

videos

Buscar... **Experto en destinos** vie. 13. jul. 2012

SOBRE CUBA ALOJAMIENTO QUÉ HACER EVENTOS OFICINAS EN EL EXTERIOR CONÓZCANOS

Destinos

- Pinar del Río
- La Habana
- Varadero
- Península de Zapata
- Cayo Largo
- Villa Clara
- Cienfuegos
- Trinidad de Cuba
- Jardines del Rey
- Santa Lucía
- Camaguey
- Las Tunas
- Holguín
- Granma
- Santiago de Cuba
- Guantánamo

Datos Personales

Edad: Sexo:

País: Desea viajar con:

Precio: Fecha:

Motivos del viaje: ¿Ha venido a Cuba?

¿Qué le gustaría hacer en Cuba?

ARTE <input type="checkbox"/> incluir	ARTESANIA <input type="checkbox"/> incluir	CAZA <input type="checkbox"/> incluir	CIUDAD <input type="checkbox"/> incluir	DEPORTE <input type="checkbox"/> incluir	DISCOTECA <input type="checkbox"/> incluir	EXCLUSIONES <input type="checkbox"/> incluir
HISTORIA <input type="checkbox"/> incluir	NATURALEZA <input type="checkbox"/> incluir	NEGOCIOS <input type="checkbox"/> incluir	PESCA <input type="checkbox"/> incluir	SALUD <input type="checkbox"/> incluir	BUCEO <input type="checkbox"/> incluir	NAUTICA <input type="checkbox"/> incluir
PLAYA <input type="checkbox"/> incluir	RELIGION <input type="checkbox"/> incluir					

CONEXIONES ON LINE

- FAQ Preguntas Frecuentes
- Encuesta ¿Has visitado a Cuba?
- Galería ¿Has visto Cuba?

FITCuba 2011
Havana, Cuba

FITCuba 2011
Havana, Cuba

Enviar

Anexo 10. Interfaz de resultados del componente cliente.

- Jardines del Rey
- Santa Lucía
- Camaguey
- Las Tunas
- Holguín
- Granma
- Santiago de Cuba
- Guantánamo

Ver Más ¿Por qué?

**Hotel: Meliá Las Antillas**
Descripción: Hacia el final de la playa Varadero, en una zona de reconocida belleza natural. Recomendado para turistas individuales, parejas y familias con jóvenes mayores de 16 años.

Ver Más ¿Por qué?

**Hotel: Meliá Cayo Santa María**
Descripción: En un paradisíaco islote de la cayería norte cubana se levanta el hotel Meliá Cayo Santa María, instalación armónicamente insertada en un polo turístico dotado de hermosas playas...

Ver Más ¿Por qué?

**Hotel: Meliá Las Dunas**
Descripción: El hotel Meliá Las Dunas se localiza en Cayo Santa María, uno de los cayos pertenecientes al archipiélago Jardines del Rey. Su ancha franja de arena blanca, sus aguas cristalinas...

Ver Más ¿Por qué?

**Hotel: Barceló Cayo Santa María**
Descripción: Se encuentra situado en una zona de bellas playas de arena blanca bañadas por las claras aguas color turquesa de Cuba. En este hotel podrá disfrutar de un completo programa todo incluido...

Ver Más ¿Por qué?

FAQ
Preguntas Frecuentes
Encuesta
¿Has visitado a Cuba?
Galería
¿Has visto Cuba?




Anexo 11. Interfaces del componente ingeniero

Ingeniería del Conocimiento New Delete

Rasgos | **Casos** | Personas | Ayuda

Filtrar por Precio

<input type="checkbox"/>	Pais	Id Persona	Destinos	Precio	Fecha	Candidato	Seleccionado
<input type="checkbox"/>	CANADA	176	Brisas Guardalavaca	todo_incluido	2011-10-10	2	1
<input type="checkbox"/>	CANADA	171	Tryp Cayo Coco	todo_incluido	2012-02-10	1	1
<input type="checkbox"/>	CANADA	175	IberoStar Daiquirí	todo_incluido	2012-12-28	1	1
<input type="checkbox"/>	CANADA	172	Meliá Cayo Coco	todo_incluido	2012-02-12	7	1
<input type="checkbox"/>	CANADA	174	Blau Colonial C Coco	todo_incluido	2012-12-22	5	1
<input type="checkbox"/>	CANADA	170	Villa Don Lino	todo_incluido	2011-10-12	1	1

a. Listado de casos

Ingeniería del Conocimiento New Save & Close Close

Datos Generales del Caso

Pais: GERMANY

Destino: Alegro Varadero

Fecha: 2012-01-01

Precio: Todo_incluido

Datos de la Persona

Edad: joven

Sexo: M

Repitente: Si

Compañia: Familia

Motivos del Viaje: Salud

Seleccione las Modalidades

<input type="checkbox"/> ARTE	<input type="checkbox"/> ARTESANIA	<input type="checkbox"/> CAZA	<input type="checkbox"/> CIUDAD	<input type="checkbox"/> DEPORTE	<input type="checkbox"/> DISCOTECA	<input type="checkbox"/> EXCURSIONES	<input type="checkbox"/> HISTORIA	<input type="checkbox"/> NATURALIZA
<input type="checkbox"/> NEGOCIOS	<input type="checkbox"/> PESCA	<input type="checkbox"/> SALUD	<input type="checkbox"/> BUCEO	<input type="checkbox"/> NAUTICA	<input type="checkbox"/> PLAYA	<input type="checkbox"/> RELIGION		

b. Entrar nuevo caso o modificar existente.

Ingeniería del Conocimiento New Delete

Rasgos | Casos | Personas | Ayuda

<input type="checkbox"/>	Nombre	Dominio	Coefficiente	Función	Épsilon
<input type="checkbox"/>	fecha	fecha	0.0625	umbral	15
<input type="checkbox"/>	vuelodirecto	vuelodirecto	0.129	booleana	0
<input type="checkbox"/>	turoperador	turoperador	0.1116	booleana	0
<input type="checkbox"/>	tipomodalidad	sad	0.0961	booleana	0
<input type="checkbox"/>	compania	sad	0.0417	booleana	0
<input type="checkbox"/>	motivoviaje	sad	0.1203	booleana	0
<input type="checkbox"/>	sexo	sexo	0.0155	booleana	0
<input type="checkbox"/>	precio	lujoso medio mediapension todoincluido economico	0.1169	booleana	0
<input type="checkbox"/>	edad	joven, adulto, anciano	0.0907	booleana	0
<input type="checkbox"/>	nombrepais	Texto	0.0995	booleana	0
<input type="checkbox"/>	repitente	booleano	0.0712	booleana	0

c. Modificar, eliminar o añadir rasgo.

Anexo 12. Interfaces del componente administrador.


Administración de la Base de Datos

New
 Delete
 Edit

Hoteles
Paises
Turoperadores
Fechas
Aereopuertos
Vuelos Directos-Paises
Turoperadores-Paises
Modalidades

Hoteles
Servicios
Paises
Construir Árbol
Ayuda

Filtrar por Hoteles Buscar Limpiar

	Foto	Nombre del Hotel	Cadena	Servicios	Ubicación	¿Cómo Llegar?
<input type="checkbox"/>		Habana Riviera	Gran Caribe	Alquiler de Autos, Acceso a Internet, Bar, Bar Piscina, Boutique Tienda, Cambio de Monedas, Masajes, Piscina, Restaurante, Restaurante buffet, Buró de turismo, Cafetería, Servicios médicos, Salón de reuniones,	Calle Paseo y Malecón, Vedado. Plaza de la Revolución. La Habana	Por avión, por yate o por crucero.
<input type="checkbox"/>		Hotel Kohly	Gaviota	Alquiler de Autos, Acceso a Internet, Bar, Boutique Tienda, Cancha de Tenis, Gimnasio, Masajes, Piscina, Restaurante, Restaurante buffet, Sala de Juegos, Sauna, Buró de turismo, Cafetería, Servicios médicos,	Ave. 49-A y 36, Reparto Kohly. Playa. La Habana.	Por avión, por yate o por crucero.
<input type="checkbox"/>		Las Terrazas	Islazul	Alquiler de Autos, Acceso a Internet, Bar, Bar Piscina, Boutique Tienda, Cambio de Monedas, Masajes, Piscina, Restaurante, Sala de Juegos, Buró de turismo, Servicios médicos, Alquiler caballos bici motos,	Ave. Las Terrazas e/ 10 y Rotonda, Playa Santa María del Mar. La Habana del Este. La Habana	Por avión, por yate o por crucero.
<input type="checkbox"/>		Hotel Lido	Islazul	Alquiler de Autos, Acceso a Internet, Bar, Cambio de Monedas, Restaurante, Buró de turismo,	Calle Consulado No. 210 e/ Animas y Trocadero. Centro Habana. La Habana.	Por avión, por yate o por crucero.
<input type="checkbox"/>		Hotel Lincoln	Islazul	Bar, Bar Piscina, Cambio de Monedas, Discoteca, Restaurante, Servicios médicos,	Calle Virtudes No. 164 esq. a Gallano. Centro Habana. La Habana	Por avión, por yate o por crucero.

a. Listado de instalaciones.


Administración de la Base de Datos

Save & Close
 Close

Imagen

Foto del Hotel: Seleccionar Limpiar

Nombre del Hotel:

Cadena:

Descripción:

Ubicación:

¿Cómo llegar?:

Seleccione los Servicios

Joomla!® is free software released under the GNU General Public License.

b. Añadir una instalación.

78



Administración de la Base de Datos

Detalles

Frecuencia: Seleccione ▾

Día: Seleccione ▾

hora:

Actualizar Árbol

c. Entrar criterio para la construcción del árbol dinámicamente

COM_LOGIN_JOOMLA_ADMINISTRATION_LOGIN

COM_LOGIN_VALID

User Name	<input type="text" value="ingeniero"/>
Password	<input type="password" value="..."/>
Language	Default ▾
<input type="button" value="Log in"/>	

COM_LOGIN_RETURN_TO_SITE_HOME_PAGE

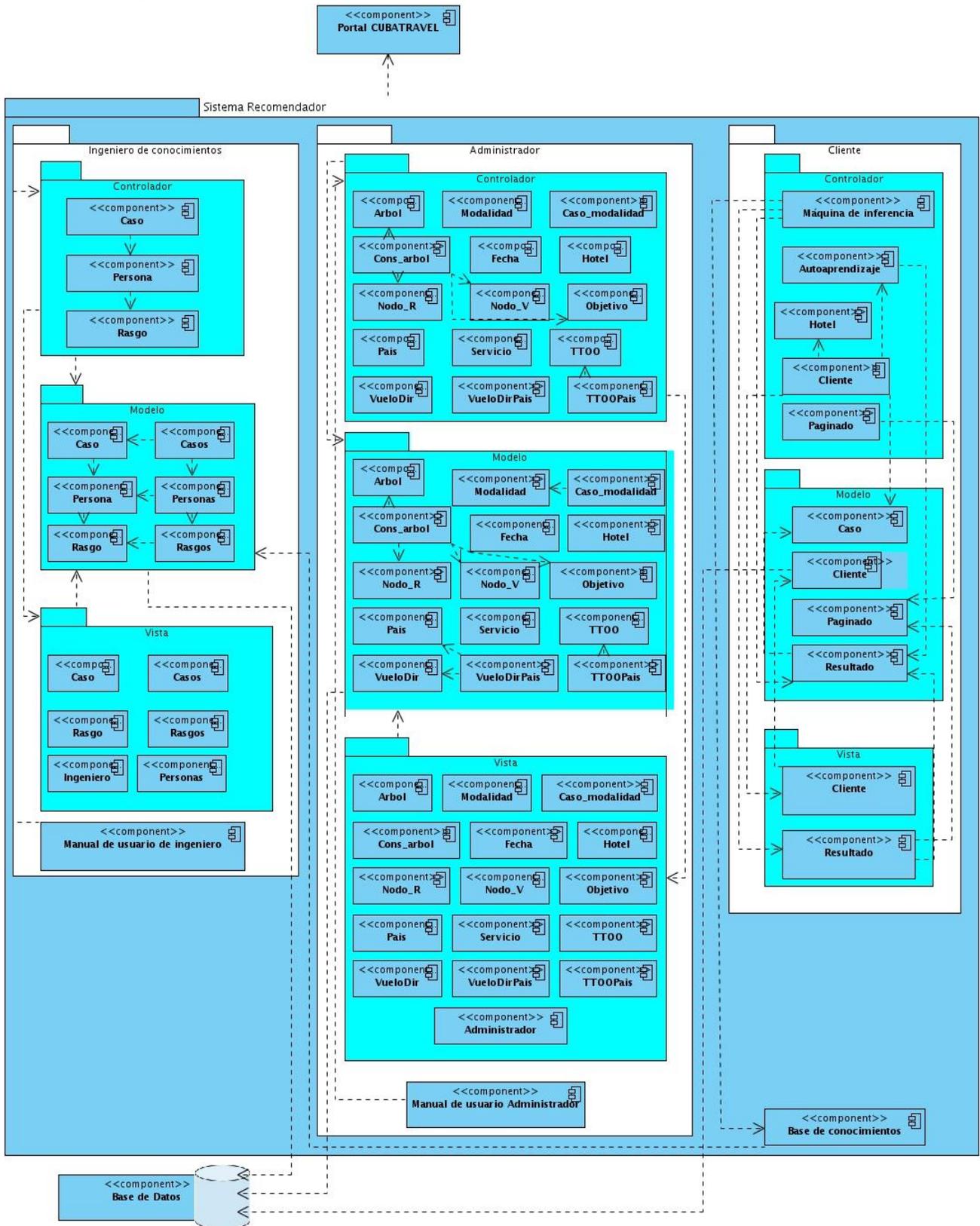


d. Interfaz de autenticación en Joomla.

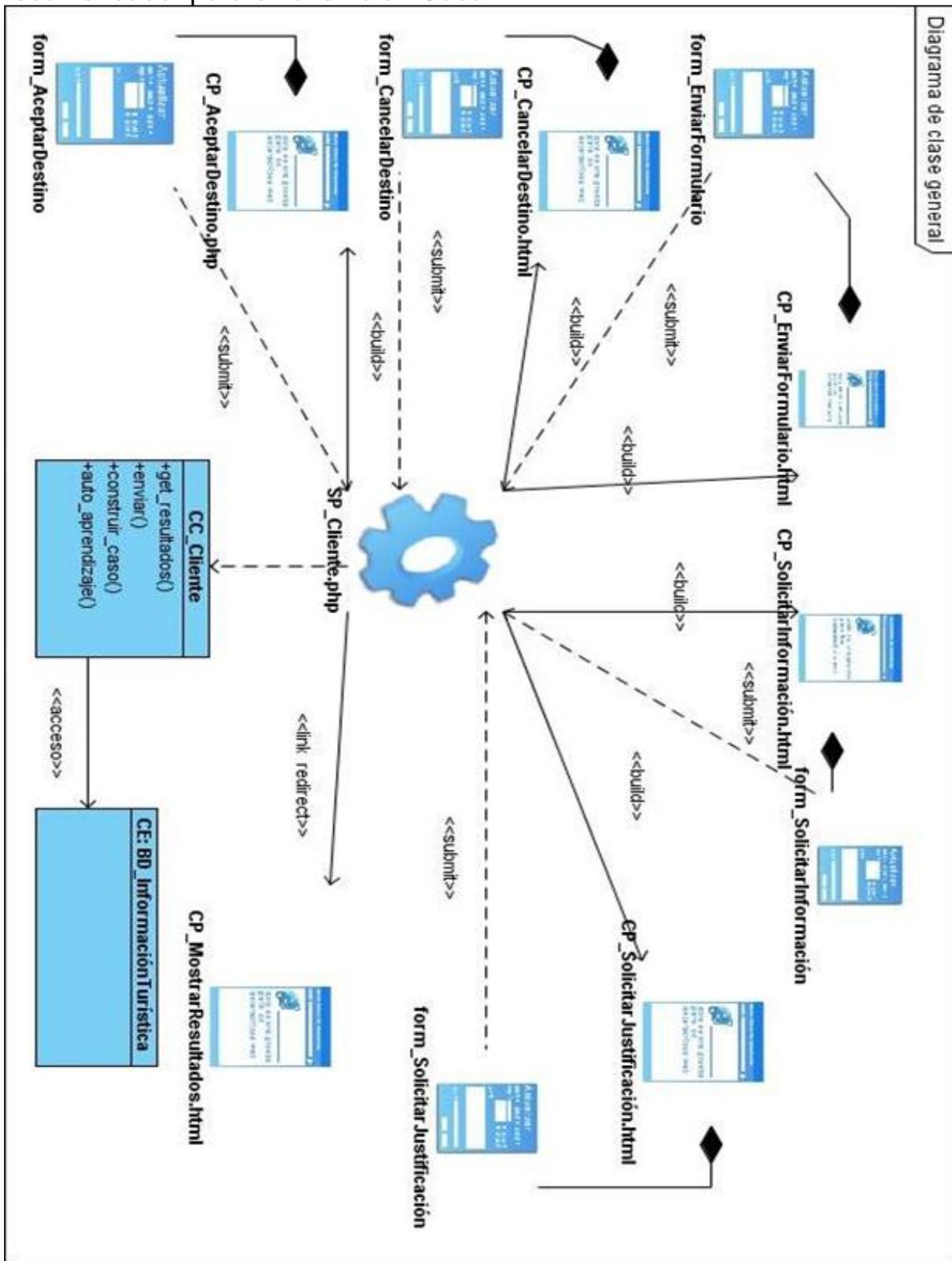
Site	Users	Menus	Content	Components	MOD_STATUS_USERS
				2012-05-24-01-Permissions	Help
<div style="display: flex; justify-content: space-between;"> <div> <p>▼ Enlaces rápidos</p> <div style="display: grid; grid-template-columns: 1fr 1fr; gap: 10px;"> <div> Add New Article</div> <div> Article Manager</div> <div> Media Manager</div> </div> </div> <div style="border: 1px solid gray; padding: 5px;"> <ul style="list-style-type: none"> Akeeba Backup Anuncios BS MyJspace Buscar busqueda Canales electrónicos com-adminbd com-ingeniero comservicioinfotur </div> </div>					

e. Menú componentes/com-ingeniero para el acceso a la interfaz del ingeniero de conocimientos.

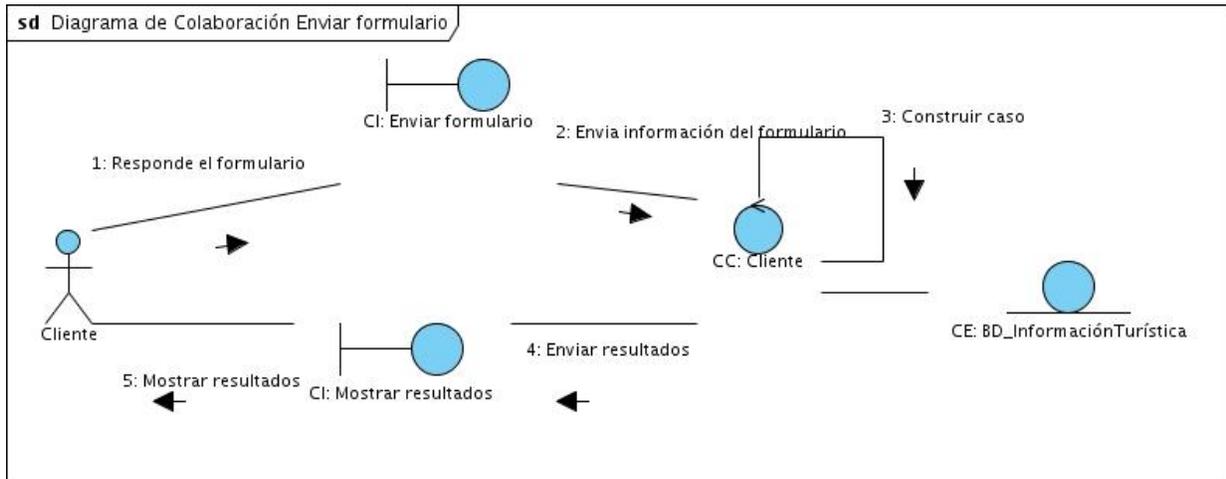
Anexo 13. Diagrama de componentes del sistema recomendador para el turismo en Cuba.



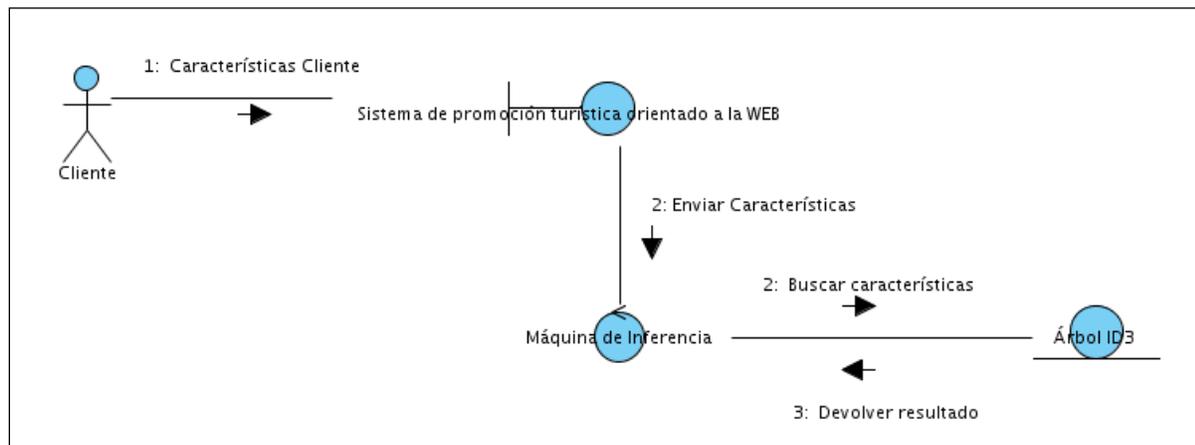
Anexo 14. Diagrama de clases con estereotipos WEB del sistema recomendador para el turismo en Cuba.



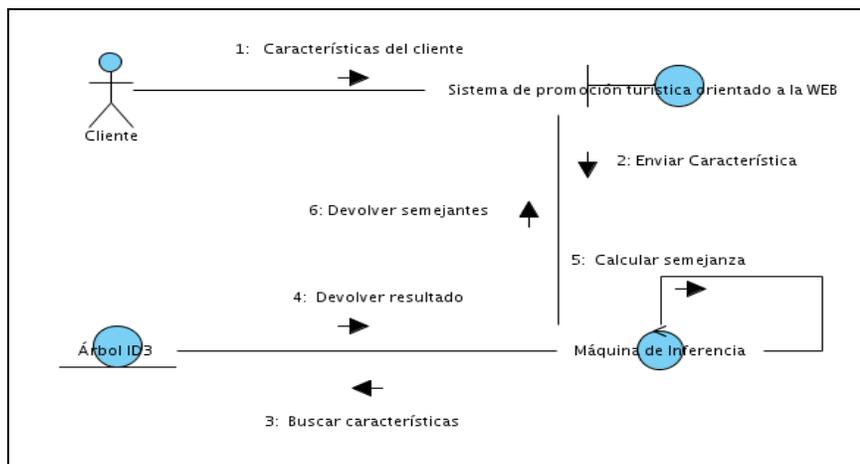
Anexo 15. Diagramas de colaboración del sistema recomendador para el turismo en Cuba.



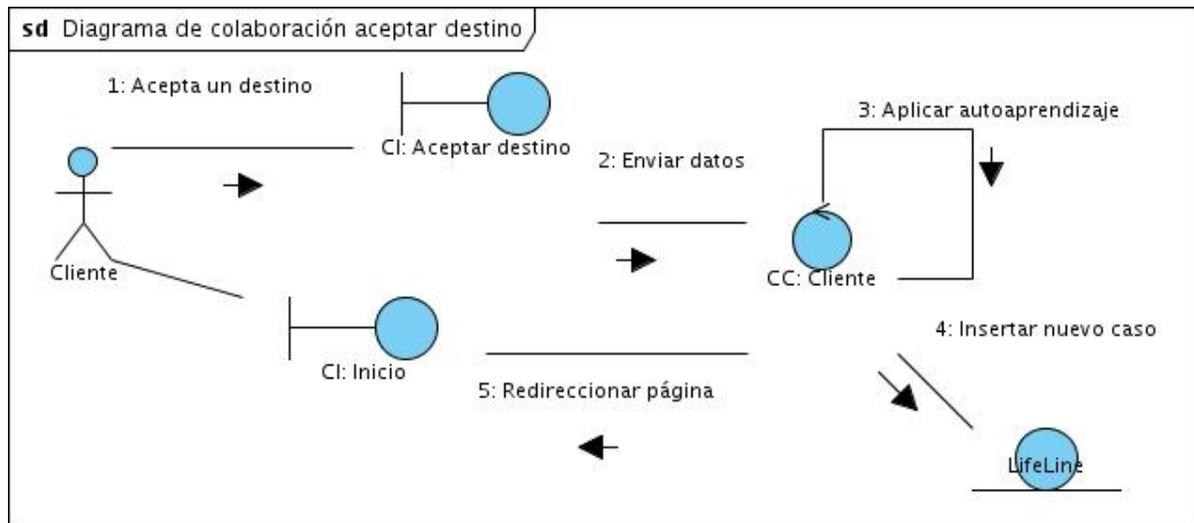
a. Diagrama de colaboración enviar formulario



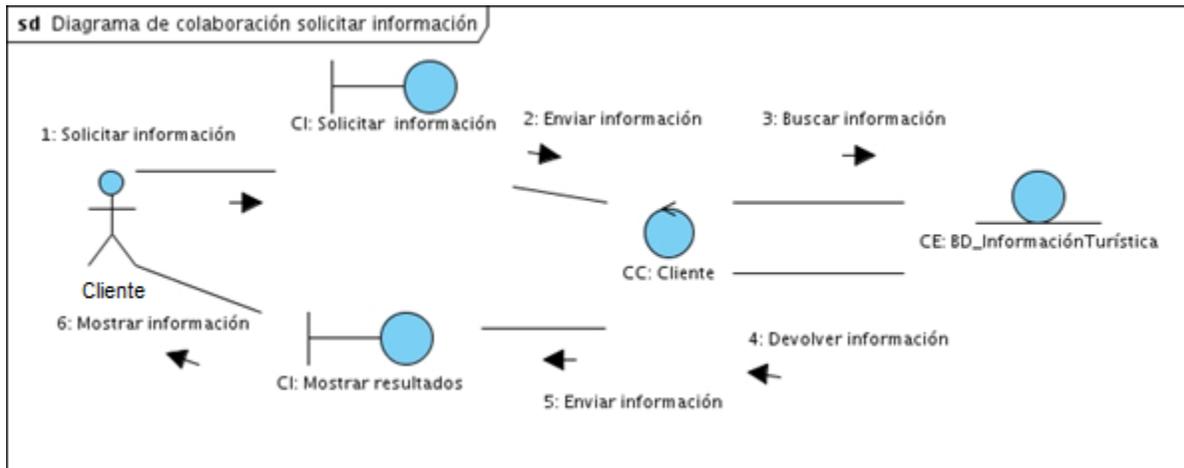
b. Diagrama de colaboración recuperar casos semejantes



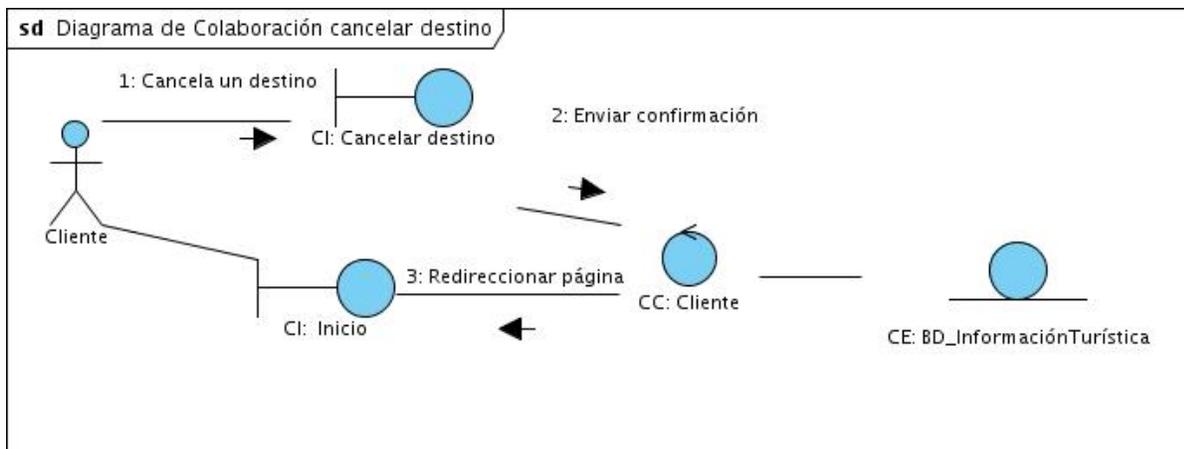
c. Diagrama de colaboración calcular semejanza



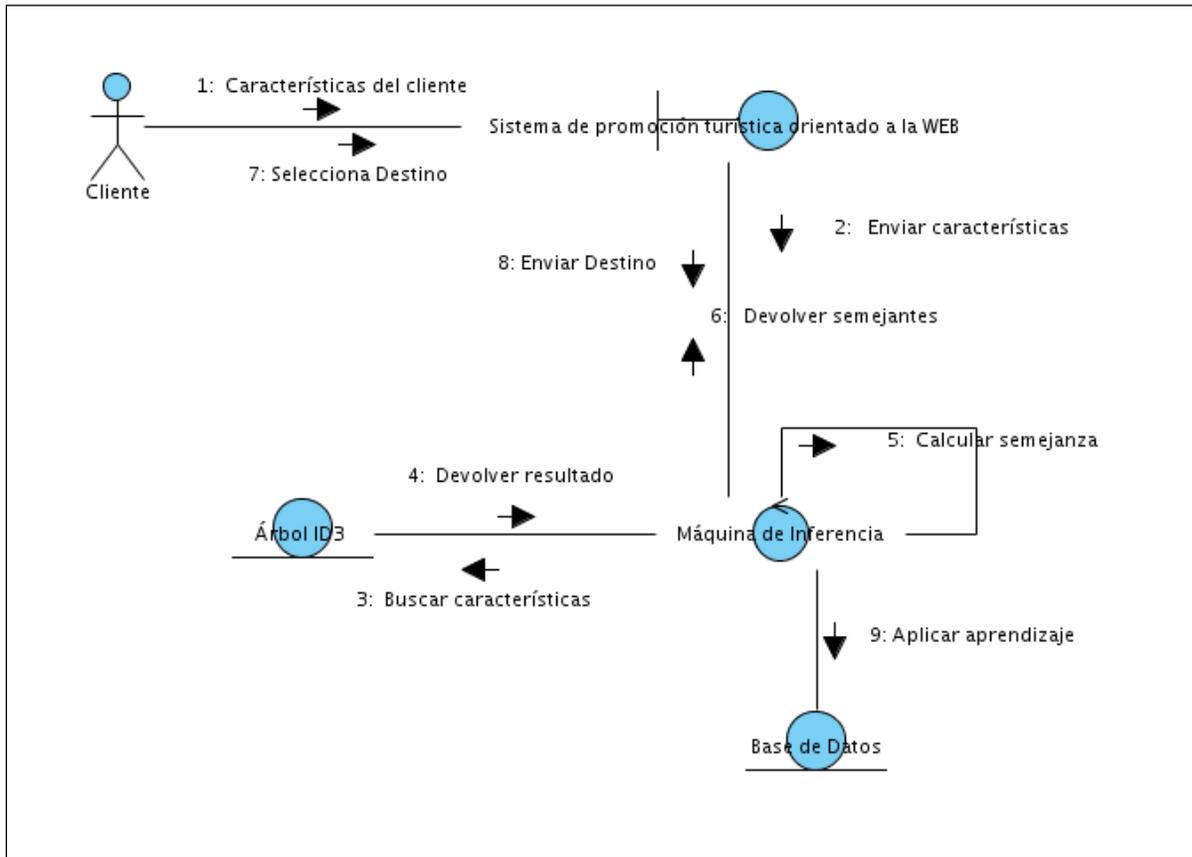
d. Diagrama de colaboración aceptar destino.



e. Diagrama de colaboración solicitar información sobre un destino.



f. Diagrama de colaboración cancelar destino



g. Diagrama de colaboración ordenar autoaprendizaje

Anexo 16. Resultados de dos evaluadores de atributos de la herramienta WEKA.

Evaluador: weka.attributeSelection.InfoGainAttributeEval	Evaluador: weka.attributeSelection.GainRatioAttributeEval
<pre> ==== Run information ==== Search: weka.attributeSelection.Ranker -T - 1.7976931348623157E308 -N -1 Relation: base_casos2- weka.filters.unsupervised.attribute.Reorder- R1,2,3,4,5,6,7,8,9,10,12,13,14,11 Instances: 1526 Attributes: 12 Precio Pais Edad Compania Sexo Repitente Motivo Fecha VueloDirecto Turoperadores Modalidades Objetivo Evaluation mode: evaluate on all training data ==== Attribute Selection on all input data ==== Search Method: Attribute ranking. Attribute Evaluator (supervised, Class (nominal): 12 Objetivo): Information Gain Ranking Filter Ranked attributes: 3.3216 8 Fecha 2.6927 11 Modalidades 1.5169 2 Pais 1.029 7 Motivo 0.7298 1 Precio 0.7288 3 Edad 0.7091 4 Compania 0.3258 9 VueloDirecto 0.1848 6 Repitente 0.1436 10 Turoperadores 0.1332 5 Sexo Selectedattributes: 8,11,2,7,1,3,4,9,6,10,5 : 11 </pre>	<pre> ==== Run information ==== Search: weka.attributeSelection.Ranker -T - 1.7976931348623157E308 -N -1 Relation: base_casos2- weka.filters.unsupervised.attribute.Reorder- R1,2,3,4,5,6,7,8,9,10,12,13,14,11 Instances: 1526 Attributes: 12 Precio Pais Edad Compania Sexo Repitente Motivo Fecha VueloDirecto Turoperadores Modalidades Objetivo Evaluation mode: evaluate on all training data ==== Attribute Selection on all input data ==== Search Method: Attribute ranking. Attribute Evaluator (supervised, Class (nominal): 12 Objetivo): Gain Ratio feature evaluator Ranked attributes: 0.875 7 Motivo 0.749 11 Modalidades 0.717 1 Precio 0.579 3 Edad 0.512 8 Fecha 0.468 2 Pais 0.41 9 VueloDirecto 0.372 10 Turoperadores 0.311 4 Compania 0.19 6 Repitente 0.147 5 Sexo Selectedattributes: 7,11,1,3,8,2,9,10,4,6,5 : 11 </pre>

Anexo 17. Avaluos del proyecto de desarrollo conjunto UCI-INFOTUR para la promoción del turismo cubano.



La Habana, 30 de abril del 2012
"Año 54 de la Revolución"

Grisel Valdés
Decana de la Facultad Regional de la
UCI de Ciego de Ávila

Estimada Decana:

Aprovecho la ocasión de que un grupo de Especialistas del Ministerio participarán con Ustedes en el análisis de los trabajos finales de curso que, en esta ocasión, están dedicados a la Promoción de Cuba como destino turístico, para expresarles nuestro Reconocimiento por los portales elaborados, que ya están en red y por las nuevas versiones que saldrán en los próximos meses.

El turismo es uno de los sectores de la economía nacional que determina en el crecimiento del PIB y, lo más importante, en el bienestar de la población por cuanto capta una cantidad importante de recursos, permite el movimiento y crea fuentes de empleo.

En la estrategia de comunicación para llegar a 2 900 000 visitantes en el 2012 y a 3 100 000 en el 2013, está la utilización de la "Nuevas Tecnologías de la Información y las Comunicaciones" en sus diversas aplicaciones. Los portales en Internet, que le dan visibilidad al destino Cuba en los mercados emisores, son uno de los elementos que determinan en la comunicación de nuestro producto turístico. Por ello valoramos como muy favorable el trabajo y contribución de la Facultad.

Coincide con que los principales mercados emisores de turistas hacia Cuba son países desarrollados. También tenemos que competir con otros destinos que disponen de recursos técnicos muy desarrollados. Es una necesidad estar al nivel del mundo. Según me informan, las nuevas versiones recogen prestaciones más avanzadas y en el próximo curso se enfocarán en llevar los portales a Internet 2.0.

Nos satisface destacar que con la terminación y puesta en explotación de los portales, estamos dando cumplimiento a los lineamientos 256 y 261 de la política económica y social del Partido.

Fraternalmente,


Rosa Adela Mejías Jiménez
RS.200/1012



La Habana, 30 de abril del 2012
"Año 54 del Triunfo de la Revolución"

Grisel Valdés
Decana de la Facultad de la UCI de Ciego de Ávila

La Oficina Nacional de Información Turística Infotur, unidad presupuestada del Ministerio de Turismo, encargada de ofrecer información del destino Cuba, mantiene por más de dos años un trabajo conjunto para desarrollar portales de turismo en Internet, con la Facultad Regional de la UCI de Ciego de Ávila.

El objetivo central es darle visibilidad en el exterior a la oferta turística de los territorios y contribuir con la soberanía tecnológica de la creación de sitios web.

En este periodo se crearon seis portales de los principales destinos específicos aprobados por Calisoft: La Habana, Varadero, Jardines del Rey, Holguín, Pinar del Río y Camagüey. Los portales han tenido buenos resultados en su repercusión.

En la actualidad se trabaja en la segunda versión tomando como base la experiencia del funcionamiento de los primeros. Se incorporan nuevas prestaciones que permiten ampliar la información al visitante. En esta etapa se termina el resto de los destinos y se renuevan los otros. En el mes de julio tendremos a todas los territorios con los sitios concluidos.

La Facultad también está enfrascada en otros proyectos: el portal oficial de Cuba en Internet, Cubatravel.cu, y en un sitio de fotografías para entregarse en el mes de julio.

Los sitios se han programado en software libre sobre la base de Joomla aportándole nuevas prestaciones específicas, basadas en las necesidades de los territorios y en otras nacionales. Hoy la Facultad ha contribuido a enriquecer el software libre en Cuba.

La investigación para la creación de estos trabajos ha estado presente para la búsqueda de soluciones. Se han estudiado los

principales sitios de turismo de la competencia y otros que tiene que ver con la información. Los temas de posicionamiento y análisis de la utilización de las redes sociales también han sido objeto de estudio

Se elabora un análisis económico de factibilidad que da los elementos sólidos para seguir esta estrategia.

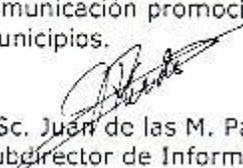
Desde el punto de vista docente la creación de estos sitios ha servido para crear metodologías para los diferentes procesos. Como culminación de esta etapa, los portales y algunos de sus componentes se convertirán en temas de tesis que, sin duda, serán un aval para la consolidación del área de desarrollo.

Para el nuevo curso se prevé llevar los sitios a Internet 2.0, hacer la propuesta para otros sitios de las instituciones turísticas, adentrarse en los sitios de comercialización, así como continuar estudiando el posicionamiento y las redes sociales.

El trabajo de las multimedia con diferentes formatos y usos será otro tema de trabajo.

De manera particular el trabajo que se realiza en la UCI de Ciego de Ávila responde a los siguientes lineamientos del Congreso del Partido:

131, que se refiere a la industria del software y a la informatización de la sociedad; 223, que alude a la elevación de la soberanía tecnológica en el desarrollo de la infraestructura de telecomunicaciones; 225, que alude a la captación de divisa; 257, que permite aumentar la competitividad de Cuba en los mercados; 258, que se refiere a perfeccionar las formas de comercialización utilizando las tecnologías más avanzadas de la información y las comunicaciones; 261, que identifica el perfeccionamiento de la comunicación promocional; 264, referido al desarrollo de los municipios.


MSc. Juan de las M. Pardo Cruz
Subdirector de Información

