



**UNIVERSIDAD DE LAS CIENCIAS INFORMÁTICAS**

**FACULTAD 1**

**DEPARTAMENTO DE TÉCNICAS DE PROGRAMACIÓN**

**TÉCNICA PARA LA RECOMENDACIÓN DE PRODUCTOS  
TURÍSTICOS EN CUBA A TRAVÉS DE UN SISTEMA DE  
INFORMACIÓN GEOGRÁFICA**

---

Tesis presentada en opción al título de Máster en Informática  
Aplicada

**Autor:** Ing. Jesse Daniel Cano Otero

**Tutor:** Dr.C. Yusnier Valle Martínez

**Co-Tutor:** Dr.C. Sergio Daniel Cano Ortiz

**La Habana, Julio del 2014**

## **AGRADECIMIENTOS**

Al Dr.C. Yusnier Valle Martínez por el apoyo en la culminación del trabajo.

Al Dr.C. Sergio Daniel Cano Ortiz por sus invaluable consejos.

A la M.Sc. Delly Lien González Hernández por sus utilísimas observaciones.

A todos los que de una forma u otra contribuyeron a la realización del trabajo.

## DEDICATORIA

A mi familia por hacerme quien soy.

A mis amigos por ayudarme a llegar hasta aquí.

A mi hija Daniella por ser la razón por la que hago todo.

## **DECLARACIÓN JURADA DE AUTORÍA**

Declaro por este medio que yo Jesse Daniel Cano Otero, con carné de identidad 84100126823, soy el autor principal del trabajo final de maestría Técnica para la recomendación de productos turísticos en cuba a través de un sistema de información geográfica, desarrollada como parte de la Maestría en Informática Aplicada y que autorizo a la Universidad de las Ciencias Informáticas a hacer uso de la misma en su beneficio, así como los derechos patrimoniales con carácter exclusivo.

Y para que así conste, firmo la presente declaración jurada de autoría en Ciudad de La Habana a los \_\_\_\_ días del mes de \_\_\_\_\_ del año 2014.

---

**Nombre y Firma del maestrante**

## RESUMEN

A medida que Internet fue tomando auge en la década del 90 se hizo común la utilización de las nuevas tecnologías de la información y las comunicaciones en la industria turística permitiendo el surgimiento del e-turismo. La información que, sobre productos turísticos, publicaban las organizaciones dedicadas al e-turismo en la web, para aprovechar las ventajas de las nuevas tecnologías, fue tomando proporciones gigantescas. Esto provocó que el cúmulo de información disponible se convirtiese en una desventaja para la eficiencia de estas organizaciones ya que los usuarios frecuentemente abandonan la búsqueda de información antes de encontrar un producto de su interés. La personalización del contenido a mostrar prominentemente a los usuarios se convirtió en un factor crítico para las organizaciones de e-turismo exitosas. Para lograr esta personalización se impone la utilización de los sistemas de recomendación ya que estos constituyen la vía más popular y con mejores resultados en este ámbito.

Las organizaciones que utilizan el e-turismo en Cuba han fallado en identificar la personalización como un factor importante y se limitan a listar toda la información sobre los productos que ofertan. Debido a la importancia que, para Cuba, tiene la industria turística es vital la utilización de un sistema de recomendación que logre decidir, eficazmente, el contenido que debe ser mostrado prominentemente a los usuarios de las organizaciones de e-turismo en Cuba.

En el presente trabajo se exponen los distintos enfoques de recomendación particularmente el enfoque colaborativo por constituir el más exitoso de todos. A partir de un análisis de las características del entorno turístico cubano, luego, se propone una técnica de recomendación para productos turísticos que tiene como base un motor de recomendación donde se implementa el enfoque colaborativo y a su vez se hace uso de un sistema de información geográfica que permite explotar la información contextual de cada usuario para lograr un aumento significativo de la cobertura del sistema.

## Índice

INTRODUCCIÓN.....	7
CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA.....	12
1.1 La industria del turismo en la actualidad.....	12
1.2 Actualidad del turismo en Cuba .....	13
1.3 Un nuevo panorama en la industria del turismo .....	16
1.4 Sistemas de Recomendación .....	17
1.5 Sistemas de Recomendación Colaborativos .....	23
1.5.1 Recomendación “Usuario - Usuario” .....	25
1.5.2 Recomendación “Ítem - Ítem” .....	27
1.5.3 Componentes de los métodos de recomendación colaborativa .....	28
1.5.4 Problemas de los sistemas de recomendación colaborativos.....	32
1.6 Uso de los GIS con los SRs de productos turísticos .....	32
1.7 Conclusiones Parciales .....	33
CAPÍTULO 2. PROPUESTA DE TÉCNICA .....	34
2.1 Estructura General.....	34
2.2 Interfaz Gráfica de Usuario .....	35
2.3 Selector de Recomendación .....	38
2.4 Motor de Recomendación .....	41
2.5 Lidiando con el arranque en frío.....	44
2.6 Conclusiones Parciales.....	44
CAPÍTULO 3. ANÁLISIS DE RESULTADOS .....	46
3.1 Uso de la Predicción .....	46
3.2 Exactitud de la Predicción .....	49
3.2 Conclusiones Parciales.....	53
CONCLUSIONES GENERALES .....	54
RECOMENDACIONES.....	55
REFERENCIAS BIBLIOGRÁFICAS .....	56

## Índice de Figuras

Figura 1. Llegada de Turistas a Cuba, 1925-1998.....	14
Figura 2. Evolución de los Ingresos por Turismo y del número de camas de hotel. ....	14
Figura 3. Representación del espacio de un problema asociado a un SRC. ....	25
Figura 4. Estructura general del sistema. ....	35
Figura 5. Funcionamiento del componente Interfaz Gráfica de Usuario. ....	38
Figura 6. Funcionamiento del componente Selector de Recomendación.....	41
Figura 7. Comparación de la Precisión entre motor de recomendación y el sistema implementado.....	48
Figura 8. Comparación del Recall entre motor de recomendación y el sistema implementado.....	49
Figura 9. Comparación de la métrica MAE entre el Motor de Recomendación y el sistema implementado.....	50
Figura 10. Comparación de la Cobertura entre el Motor de Recomendación y el sistema implementado.....	52

## Índice de Tablas

Tabla 1. Cantidad de establecimientos de alojamiento y cantidad de visitantes arribados al país. ....	15
Tabla 2. Cantidad de productos comercializados por algunas organizaciones de e-turismo cubanas y estructura organizativa del contenido. ....	17
Tabla 3. Datos de entrada para el cómputo de la Precisión y el Recall. ....	47
Tabla 4. Resultado medio de Precisión y Recall del Motor de Recomendación.....	48
Tabla 5. Resultado medio de Precisión y Recall del sistema implementado. ....	48
Tabla 6. Error medio absoluto del Motor de Recomendación.....	50
Tabla 7. Error medio absoluto del sistema implementado.....	50
Tabla 8. Por ciento de cobertura del espacio de predicción para el Motor de Recomendación.....	52
Tabla 9. Por ciento de cobertura del espacio de predicción para el sistema implementado. .....	52

## INTRODUCCIÓN

Luego de la caída del campo socialista, el gobierno revolucionario apostó por el desarrollo turístico en Cuba como vía de proveer un alivio a la economía. Este sector se desarrolló al punto de convertirse en uno de los motores de la economía cubana. Con el auge de internet en los noventa la industria turística mundial fue cambiando dando lugar al turismo electrónico (e-turismo). Varios estudios recientes indican que la Web se ha convertido en la principal fuente de información en el mundo occidental para las personas que buscan obtener información de los destinos turísticos cuando se encuentran planificando sus viajes (Jannach, Zanker et al. 2007). A pesar de este hecho, la mayoría de los sitios web cubanos de temática turística se limitan a brindar información de los distintos destinos y en algunos casos brindan la posibilidad de realizar reservaciones para algún que otro producto turístico. Es obvio que la cantidad de información que recibe el usuario cuando desea encontrar un producto turístico cubano en la web puede convertirse en un problema porque el usuario frecuentemente se ve abrumado por el cúmulo de información disponible y es probable que abandone el sitio web antes de encontrar el producto de su preferencia.

La personalización ha sido reconocida como un factor crítico para las empresas de e-turismo exitosas (Gao, Liu et al. 2010). El enfoque más exitoso para lidiar con el problema de la personalización es la utilización de los sistemas de recomendación (SRs) (Resnick and Varian 1997; Adomavicius and Tuzhilin 2005). Este enfoque provee técnicas para el filtrado de información usando el conocimiento sobre sus preferencias. Los SRs pueden brindar al visitante del sitio un conjunto de opciones, desde destinos, hoteles, museos y lugares de interés que se ajusten al perfil y presupuesto del visitante. Para la personalización del servicio en el e-turismo, los enfoques más populares son el uso de SRs con filtrado colaborativo y basados en el contenido.

Por otro lado, muchos SRs hacen uso de sistemas de información geográfica (GIS por sus siglas en inglés) como una vía de mejorar la presentación de las recomendaciones y, de esta manera, incrementar la interactividad del sistema con los usuarios. La utilización de GIS para el e-turismo es totalmente razonable dado a que la mayoría de la información turística tiene una estrecha relación con el espacio (Charou, Kabassi et al. 2010). Los

GIS, además, permiten la presentación de varios tipos de información y permiten que los usuarios sean capaces de determinar las distancias entre varios puntos de interés turístico, lo que es una ayuda al proceso de toma de decisiones, por ejemplo, a la hora de planificar un recorrido turístico.

Los SRs colaborativos funcionan mediante la recolección de valoraciones humanas (*ratings* en inglés) para los distintos ítems pertenecientes a un dominio dado y agrupando a las personas que tienen necesidades y/o preferencias similares (Adomavicius and Tuzhilin 2005).

Los SRs están dirigidos esencialmente a individuos carentes de suficiente experiencia personal para evaluar el potencialmente abrumador número de ítems alternativos que un sitio web puede ofrecer (Resnick and Varian 1997). Los SRs son, en la actualidad, un campo de investigación activo y están siendo utilizados exitosamente en varios sitios web de comercio electrónico, e-turismo y entretenimiento como Amazon, Netflix, YouTube, iTunes y Last.fm.

Uno de los SRs pioneros fue el sistema Tapestry, desarrollado por Xerox PARC, este sistema dio el primer paso hacia la recomendación de contenidos mediante la incorporación de las acciones y opiniones del usuario en el proceso (Goldberg, Nichols et al. 1992). Luego del surgimiento de Tapestry, otros investigadores empezaron a reconocer el potencial de explotar la actividad humana en la recomendación de contenidos. Maltz y Ehrlich (Maltz and Ehrlich 1995) desarrollaron un sistema de recomendación colaborativo que facilitó la recomendación de un documento leído por un usuario a otros usuarios que, a su juicio, debían leerlo también. Este tipo de sistemas tiene la limitación de que requieren que la comunidad de usuarios que interactúan con ellos se conozcan los unos a los otros. Esto dio paso a los sistemas de filtrado colaborativo automatizado (ACF por sus siglas en inglés), estos sistemas relevan a los usuarios de la carga de tener que recomendar los contenidos mediante el uso de bases de datos que almacenan los historiales de las opiniones de los usuarios, para luego comparar a los individuos con opiniones similares y de este modo efectuar una recomendación.

Los primeros sistemas ACF incluían a GroupLens (Konstan, Miller et al. 1997) en los servicios de noticias mediante "newsgroups", Ringo (Shardanand and Maes 1995) en música y Bellcore's Video Recommender (Hill, Stead et al. 1995) que recomendaba

películas. Cada uno de estos sistemas siguió un proceso de recolección de valoraciones de los usuarios, comparando, luego, las correlaciones entre pares de usuarios para identificar la vecindad de este en cuanto al gusto en un determinado tema.

Inicialmente los SRs colaborativos eran diseñados para proveer información explícita de los ítems a los usuarios. Por ejemplo, los usuarios visitaban los sitios web con el propósito de obtener recomendaciones a partir del ACF implementado. En la actualidad los SRs, utilizan los ACF de manera transparente para el usuario, por ejemplo, para adaptar el contenido a cada tipo específico de usuario mostrando qué ítem debe ser presentado prominentemente a este.

A pesar de la gran importancia que, para Cuba, tiene la industria del turismo las empresas que utilizan el e-turismo como vía de captar clientes saturan de información la Web de manera que los potenciales clientes se enfrentan al problema de determinar qué producto turístico consumir a partir de un monto abrumador de información. Esto invariablemente atenta contra la eficiencia con que operan estas empresas y contra el consumo de los productos turísticos que se ofertan. Esta situación puede ser mitigada con el uso de un método que permita predecir qué producto turístico puede ser de interés para un grupo determinado de clientes. En este contexto se puede hacer uso de un sistema de recomendación, pero, para esto, se deben utilizar técnicas que ayuden a caracterizar los distintos elementos que componen el entorno turístico cubano con el objetivo de proveer una recomendación de productos turísticos con exactitud y de manera precisa.

Teniendo en cuenta el significado que tiene la industria turística para Cuba y lo útil que resultaría una mejora en la presentación de la información en las actividades de e-turismo y debido al poco desarrollo que, al respecto, presentan las organizaciones que utilizan el e-turismo como vía de atraer clientes potenciales que consuman la amplia gama de productos turísticos cubanos, se plantea como **problema de la investigación**, ¿cómo seleccionar de manera eficaz el contenido a mostrar prominentemente a los usuarios de las organizaciones que utilizan el turismo electrónico en Cuba?

El **objeto de estudio** en el cual se enmarca el desarrollo de este trabajo, desde el punto de vista práctico y teórico para darle solución al problema identificado se define como la recomendación de elementos. Mientras que el **campo de acción** está dado por la recomendación de elementos en aplicaciones que ofrecen información sobre productos turísticos en Cuba.

Con el propósito de solucionar efectivamente el problema, se plantea **como objetivo general** desarrollar una técnica que permita definir y obtener un sistema de recomendación a partir de la definición de sus componentes principales, para seleccionar de manera eficaz el contenido que se muestra prominentemente a los usuarios de las organizaciones que utilizan el e-turismo en Cuba.

En el presente trabajo se procederá a proponer una solución a la situación problemática planteada partiendo de la siguiente **hipótesis**:

Si se obtiene una técnica para la definición de los principales componentes de un sistema de recomendación de productos turísticos se logrará una selección eficaz del contenido que se muestra prominentemente a los usuarios de las organizaciones de e-turismo en Cuba.

Como **tareas de investigación** se proponen las siguientes:

1. Identificación de las tendencias actuales para dar solución al problema planteado, a partir del estudio profundo de los referentes teóricos que preceden el presente trabajo.
2. Caracterización de los distintos enfoques y técnicas utilizadas en la definición sistemas de recomendación así como la presentación de los resultados haciendo uso de un sistema de información geográfica, que puedan servir de precedente para la investigación.
3. Desarrollo de una técnica que permita obtener un sistema de recomendación que utilice un sistema de información geográfica como medio de representar la información.
4. Evaluación de los resultados obtenidos haciendo uso de métricas de evaluación de los sistemas de recomendación, a partir del desarrollo de una aplicación de software que permita validar la efectividad de la técnica desarrollada.

La **contribución** fundamental de este trabajo consiste en la propuesta de una técnica para la obtención de SRs según la caracterización de los entornos turísticos para los que se quiera obtener los beneficios del filtrado colaborativo en aras de mejorar la personalización del servicio por parte de las empresas de e-turismo en Cuba.

Como **resultado** se espera:

Primero: Una caracterización profunda de los SRs, y sus principales enfoques y variantes.

Segundo: Una técnica para la conceptualización de SRs en entornos turísticos basados en GIS para la presentación de los resultados.

Tercero: Una aplicación de software que brindará recomendaciones sobre lugares turísticos y realizará la representación de los resultados sobre un mapa del entorno turístico que se seleccione.

### **Métodos Teóricos:**

Analítico–Sintético: A partir del análisis de los referentes teóricos y la bibliografía en cuestión se hará una síntesis de los elementos significativos que forman parte del dominio del problema, el objeto de estudio y las soluciones alternativas existentes.

Inductivo–Deductivo: En este método a partir de la hipótesis y siguiendo reglas lógicas de deducción se llega a la contribución principal de la investigación que posteriormente será sometida a verificaciones empíricas.

Histórico–Lógico: Para resolver el problema de la investigación será necesario el estudio y evolución de los conceptos presentes en la investigación hasta el momento actual, además del objetivo y campo de acción propuestos.

### **Métodos Empíricos:**

Medición: Este método se usará para obtener información numérica acerca de una propiedad o cualidad del objeto, en este es donde se comparan magnitudes medibles y conocidas.

### **Métodos Particulares:**

Entrevista: Se utilizó con la finalidad de buscar información sobre la situación existente. Permitted reunir un conjunto de características necesarias para poder dimensionar el problema de la investigación.

# CAPÍTULO 1. FUNDAMENTACIÓN TEÓRICA

Con el objetivo de facilitar la comprensión del alcance de la investigación, en el presente capítulo se definen un conjunto de conceptos asociados al dominio del problema planteado. Se explica qué son y cómo pueden ser usados los SRs a partir de la descripción de los distintos enfoques que en la actualidad definen el estado del arte. Por último se resaltan las ventajas de la utilización de los GIS en la representación de los resultados de una recomendación de productos turísticos.

## 1.1 La industria del turismo en la actualidad

El turismo es considerado la actividad económica más dinámica del siglo XX, es un fenómeno característico e irreversible de la actualidad, con especial importancia en el desarrollo de las sociedades, entre otras causas, por la gran cantidad de personas que participan en él, así como su amplia repercusión en lo ecológico, económico, social y cultural (Salinas and Mundet 2000).

El desarrollo del turismo a escala mundial y en el Caribe es una realidad y constituye el modo de vida principal para varios países en desarrollo debido a que sus efectos son aprovechados en beneficio de las economías.

La industria del turismo es una industria global y tiene características específicas:

- Los viajes y el turismo representan aproximadamente el 11 % del producto interno bruto (PIB) mundial (Werthner 2003).
- Según la organización mundial del turismo (OMT), en el año 2010 se sobrepasó el billón de llegadas internacionales a los distintos puertos del mundo.
- Representa una enorme industria “entrecruzada”, que incluye diversos sectores económicos asociados, tales como la cultura, el deporte o la agricultura, donde se han identificado más de 30 componentes industriales que sirven a los viajeros (Werthner 2003).
- Tiene una enorme importancia para el desarrollo regional, por ejemplo, en la Unión Europea existen alrededor de 1 300 000 millones de hoteles y restaurantes (9% del total de empresas)(Werthner 2003).

El turismo es un negocio que se basa en la información, el producto es un bien que depende de la confianza y un análisis, a priori, de sus cualidades es imposible. El

consumidor al ejecutar el proceso de toma de decisiones cuenta solo con un modelo o descripción del producto. Esta característica de los productos turísticos provoca que el costo de la búsqueda de información, al respecto, sea elevado y causa imperfecciones informacionales en el mercado.

### **1.2 Actualidad del turismo en Cuba**

El turismo no es una actividad nueva en Cuba. Al igual que en el mundo, tuvo su auge a partir de los años 50. En esa época, cuando el desarrollo turístico en la mayor de las Antillas estuvo muy ligado a la presencia de la mafia norteamericana en la Isla, Estados Unidos era el mercado principal, y el juego y la prostitución eran las principales ofertas de la Isla. Este turismo de ciudad condicionó el poco desarrollo del turismo natural en esa etapa. Con el triunfo de la Revolución comenzó la política norteamericana de bloqueo y se eliminó el turismo proveniente de Estados Unidos. A partir de 1959, el desarrollo de la economía estuvo dirigido a otros programas importantes del país, por lo cual el turismo era entonces fundamentalmente nacional, hecho que condicionó una estructura habitacional poco competitiva como producto internacional (MINTUR 2002).

En 1989 con la caída del campo socialista, que supuso una caída del 35% del PIB cubano, el gobierno se vio obligado a considerar al turismo como una alternativa importante para la recuperación económica del país. Esto conllevó un proceso de construcción de nuevas capacidades hoteleras, el desarrollo de infraestructura turística y la participación cada vez mayor del capital extranjero.

Indicadores turísticos actualizados muestran que si en 1987 Cuba recibió 29000 turistas, solo en 10 años logró multiplicar por 4 esa cifra (ver figura 1) y según la Oficina Nacional de Estadísticas (ONE) en 2011 se contabilizó la llegada de 2.716.317 visitantes extranjeros.

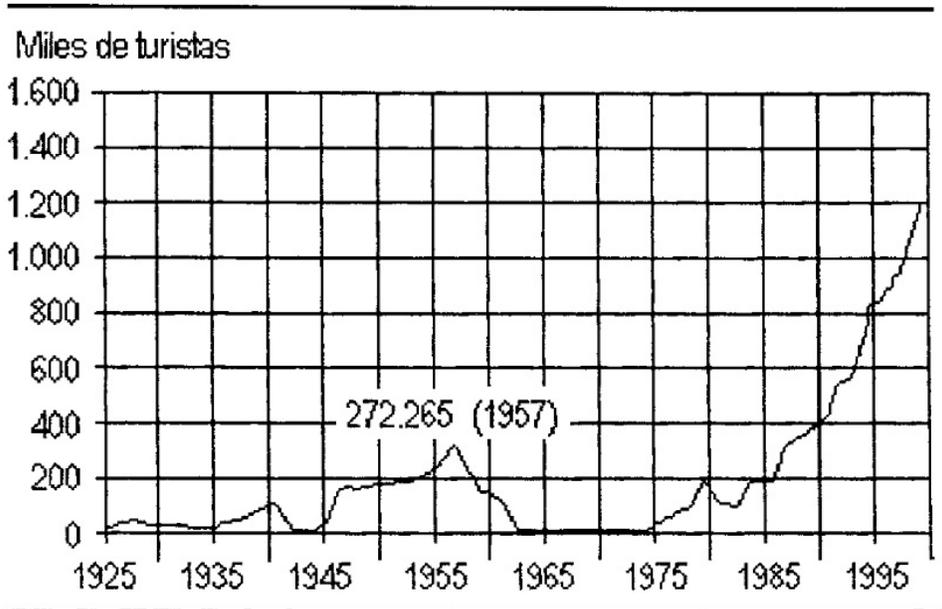


Figura 1. Llegada de Turistas a Cuba, 1925-1998. Tomado de Salinas, E.: Caribbean Tourism Organization.

Los ingresos en este mismo período pasaron de 165 millones de dólar en 1987 a 1500 millones en mil 1997, mientras el número de hoteles creció, en ese mismo período, aumentando la capacidad hotelera de 7.500 a 27.400 habitaciones repartidas en 179 hoteles (ver figura 2).

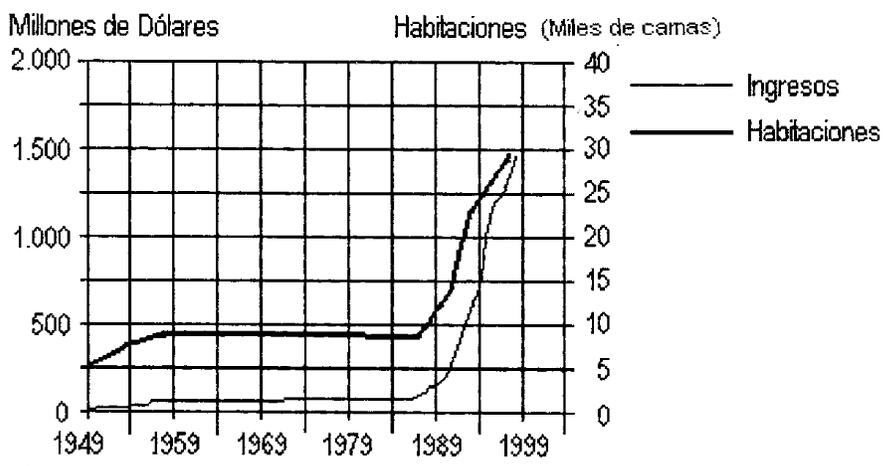


Figura 2. Evolución de los Ingresos por Turismo y del número de camas de hotel. Fuente: Ministerio de Turismo de Cuba.

Como estrategia de consolidación de la industria turística en Cuba, en 1994 se crea el Ministerio de Turismo, encargado del control, fiscalización y dirección de las actividades del sector. En 1996 el turismo se ubicó como la actividad económica más importante de Cuba (Salinas and Mundet 2000). En la tabla 1 se muestra la información referente a la cantidad de visitantes extranjeros arribados en el período 2006 – 2011 y la cantidad de establecimientos de alojamiento.

Tabla 1. Cantidad de establecimientos de alojamiento y cantidad de visitantes arribados al país. Elaboración propia a partir de la información del anuario estadístico de la ONE (2011).

	<b>2006</b>	<b>2007</b>	<b>2008</b>	<b>2009</b>	<b>2010</b>	<b>2011</b>
Hoteles y moteles	437	435	414	425	434	432
Villas turísticas	70	65	60	60	60	60
Casas y cabañas turísticas	38	38	33	32	38	40
Apartamentos turísticos	1	1	1	1	1	1
Apartahoteles	1	1	1	1	1	1
Hostales	12	12	11	11	11	11
Cotos de caza	7	7	7	7	7	7
Campamentos turísticos	1	1	1	1	1	1
Albergues	17	10	5	5	4	3
Otros	10	10	6	5	8	6
<b>Total de establecimientos</b>	<b>594</b>	<b>580</b>	<b>539</b>	<b>548</b>	<b>565</b>	<b>562</b>
<b>Total de visitantes arribados</b>	<b>2.220.567</b>	<b>2.152.221</b>	<b>2.348.340</b>	<b>2.429.809</b>	<b>2.531.745</b>	<b>2.716.317</b>

Como se puede observar en la tabla 1, el número de turistas arribados al país continúa en ascenso en el orden de los miles cada año, mientras el número de establecimientos permanece estable lo que constituye una característica significativa del entorno turístico cubano.

### **1.3 Un nuevo panorama en la industria del turismo**

El desarrollo de internet y la Web, facilita nuevas maneras de cumplir con las cambiantes necesidades de los consumidores de productos turísticos lo que ha provocado una informatización de la industria del turismo en aras de generar un valor agregado (Sweet 2001). Esto ha permitido la implementación de distintas estrategias al respecto:

- Incremento de la eficiencia y reducción de costos a través de la automatización de procesos dentro de la industria.
- Uso de los datos de los clientes y de las ventas para apoyar las actividades de marketing, por ejemplo: el uso de minería de datos para predecir el comportamiento de la industria.
- Combinación de productos y/o servicios para crear paquetes de productos más atractivos, por ejemplo: información en tiempo real para los clientes sobre qué consumir durante su viaje basándose en servicios Web y tecnologías móviles.

El e-turismo puede, entonces, ser definido como la aplicación de las tecnologías de la información (TI) en los procesos y cadenas de valores del turismo para lograr un aumento en la efectividad y eficiencia de las organizaciones dedicadas al área (Buhalis 2003). Desde este punto de vista, se puede afirmar que las TI pueden contribuir a disminuir el precio de la personalización, lo que permite realizar ofertas individuales a los clientes basadas en una personalización masiva.

Dada la dinámica del sector y el emergente, pero actualmente muy competitivo, mercado electrónico del turismo, casi todas las organizaciones presentes en la industria han implementado una estrategia que explota las TI e internet (Werthner 2003). Esto sucede porque la utilización de las TI ha permitido que los turistas tomen un rol más activo en la producción de productos turísticos porque ya no se conforman con productos estandarizados (Berka and Plößnig 2004).

En la actualidad el volumen de información que ofrecen las organizaciones que utilizan el e-turismo provoca que los viajeros necesiten ser aconsejados sobre los lugares a visitar. Generalmente, los turistas eran ayudados por un agente de viajes, una persona con los conocimientos necesarios para proveer ese consejo. Sin embargo, las recomendaciones provistas por ese agente tienen una marcada limitación como consecuencia de múltiples factores humanos como la falta de memoria, el conocimiento limitado acerca del mundo, los países, las ciudades y los productos turísticos que se ofertan, lo que limita la

capacidad de vincular los deseos o requisitos del turista en potencia con las opciones turísticas disponibles. Con mucha frecuencia la decisión final del consumidor depende de la recomendación hecha por el agente de viajes.

En Cuba son precisamente las agencias de viajes el tipo de organizaciones que usualmente apuestan por el e-turismo para aumentar la eficiencia en la comercialización de sus productos. Sin embargo la mayoría de estas organizaciones fallan en implementar un enfoque de personalización que le permita mejorar la manera en que sus usuarios pueden acceder a los productos de interés. La tabla 2 muestra, de algunas de las principales agencias de viajes de Cuba, la cantidad de productos que comercializan en sus respectivos sitios web para el destino La Habana, así como el enfoque organizativo que utilizan para la visualización del contenido que se muestra a los usuarios.

Tabla 2. Cantidad de productos comercializados por algunas organizaciones de e-turismo cubanas y estructura organizativa del contenido. Elaboración propia a partir de la información extraída de los sitios web de las organizaciones.

	<b>Cantidad de Productos</b>	<b>Estructura Organizativa</b>
Infotur	475	Orden alfabético
Cuba Travel	340	Ninguna
Havanatur	157	Ninguna
Grupo Gaviota	114	Orden alfabético

Como se puede apreciar en la tabla 2, las organizaciones cubanas, aunque esperan beneficiarse del resultado de la utilización del e-turismo y lo logran en el sentido de la visibilidad de la organización y sus productos a nivel mundial usando la web, conservan la deficiencia de ofrecer un número relativamente elevado de opciones sin diferenciarlas según las características particulares de cada usuario. El proceder de estas organizaciones, sin embargo, no puede verse desligado de las restricciones tecnológicas que tiene el país, por ejemplo: no se puede acceder a internet a través de la red móvil nacional; lo que constituye una limitación importante en el funcionamiento de las organizaciones que utilizan el e-turismo en el contexto cubano.

#### **1.4 Sistemas de Recomendación**

Los SRs son herramientas de software y técnicas que proveen sugerencias o recomendaciones acerca de los ítems que pueden ser útiles para un usuario (Resnick and Varian 1997; Mahmood and Ricci 2009). Desde este punto de vista se puede decir,

entonces, que los SRs son la respuesta técnica al hecho de que las personas frecuentemente se basan en la experiencia y recomendaciones de otras personas cuando se enfrentan a un nuevo campo del conocimiento, cuando no conocen con amplitud todos los hechos o cuando el conocimiento excede la cantidad de información que un humano puede manejar cognoscitivamente (Berka and Plöbñig 2004). Estas recomendaciones se relacionan a varios procesos de tomas de decisiones, por ejemplo: qué ítem comprar, qué canción escuchar o qué noticia leer.

El término “ítem” se usa, de manera general, para denotar qué es recomendado a los usuarios por el sistema. Los SRs normalmente se centran en un tipo de ítem específico (dígase canciones, noticias o libros) y de acuerdo a esto su diseño, su interfaz gráfica de usuario (GUI por sus siglas en inglés), y la técnica de recomendación se configura para proveer sugerencias útiles y efectivas para ese tipo específico de ítem (Desrosiers and Karypis 2011).

Luego, los SRs se dirigen principalmente a los individuos que carecen de la suficiente experiencia personal o la competencia para evaluar el potencialmente abrumador número de alternativas que un sitio Web, por ejemplo, puede ofrecer (Resnick and Varian 1997).

En su forma más simple, las recomendaciones se ofrecen como una lista ordenada de ítems. Al ejecutar este ordenamiento, los SRs tratan de predecir cuáles son los productos o servicios más convenientes para el usuario. Para lograr esto, computacionalmente, los SRs recolectan las preferencias de los usuarios, ya sea explícitamente, por ejemplo: solicitando a los usuarios valoraciones sobre los ítems, o infiriéndolas a partir de las acciones de los usuarios.

Los primeros SRs aplicaban algoritmos para relacionar las recomendaciones hechas por un grupo de usuarios con el objetivo de proveer una recomendación para un usuario activo. Las recomendaciones eran, entonces, ítems que habían gustado a usuarios similares (dígase usuarios con gustos similares). Este enfoque se denomina filtrado colaborativo y parte del razonamiento de que si el usuario activo tiene gustos y preferencias similares con otros usuarios, entonces, las valoraciones de esos usuarios deben ser relevantes y de interés para el usuario activo (Desrosiers and Karypis 2011).

El crecimiento explosivo y la variedad de la información disponible en la Web y la rápida introducción de nuevos servicios de e-comercio (compra de productos, comparación de

productos, subastas, etc.) frecuentemente abruma a los usuarios, lo que provoca que estos tomen decisiones equivocadas en varias ocasiones. Entonces, se entiende que una opción puede ser buena, pero muchas opciones no siempre es mejor. En la actualidad, los SRs han demostrado ser un medio valioso de lidiar con el problema del exceso de información (Desrosiers and Karypis 2011).

### *FUNCIONES DE LOS SISTEMAS DE RECOMENDACIÓN*

Los SRs juegan roles distintos para aquel que presta el servicio y para usuario del sistema. Por ejemplo, un sistema de recomendación de viajes, normalmente, es introducido por un intermediario en el negocio de los viajes o por una organización gestora de destinos para aumentar sus ganancias, vender más habitaciones de hoteles o para aumentar el número de turistas hacia un destino en particular (Ricci 2002). Sin embargo, la principal motivación de un usuario que accede a este tipo de sistemas es la de encontrar un hotel adecuado y opciones de eventos o atracciones interesantes mientras se encuentre en el destino turístico.

Existen varias razones por la que una organización pudiera poner en explotación un SR:

- Incrementar el número de ítems vendidos. Es la función más importante para los SRs con fines comerciales (Desrosiers and Karypis 2011). Esto se logra porque es probable que los ítems que son recomendados satisfagan las necesidades del usuario en una mayor medida que aquellos que se muestran sin recomendación alguna.
- Vender una mayor variedad de ítems. Otra función importante de los SRs es permitir a los usuarios encontrar y seleccionar ítems que, de otra manera, hubiesen sido difíciles de encontrar sin una recomendación precisa.
- Incrementar la satisfacción del usuario. Un SR bien diseñado puede mejorar la experiencia de usuario con el sitio Web o aplicación ya que el usuario encontrará las recomendaciones interesantes y relevantes.
- Incrementar la lealtad del usuario. Es probable que un usuario se mantenga leal a un sitio Web que, cuando es visitado, lo reconozca como un cliente frecuente y, en consecuencia, lo trate como un visitante valioso. Los SRs logran esto con relativa facilidad ya que personalizan el servicio que se presta al usuario a partir de las interacciones previas del usuario con ellos.

- Mejorar el entendimiento de las necesidades del usuario. Los SRs brindan descripciones de las preferencias de los usuarios, estas descripciones pueden ser recogidas explícitamente de la base de datos del sistema para ser analizadas o pueden realizarse automáticamente a partir de predicciones.

Los usuarios también pueden desear utilizar los SRs, si esto les sirve para mejorar la ejecución de sus tareas o la consecución de sus objetivos en un momento dado, algunas motivaciones son:

- Encontrar algunos ítems adecuados. Los usuarios pueden obtener una lista ordenada con algunos productos que probablemente sean de su gusto.
- Encontrar todos los ítems adecuados. Los usuarios pueden obtener una lista ordenada de todos los ítems que pueden satisfacer algunas de sus necesidades en aquellos casos en los que mostrar solo algunos ítems sea insuficiente.
- Anotación en contexto. Dado un contexto existente, por ejemplo: una lista de ítems, el usuario puede ver resaltados aquellos que pueden satisfacer mejor sus necesidades.
- Recomendar una secuencia. Los usuarios pueden obtener una secuencia de ítems que son adecuados para él como un todo.
- Recomendar un paquete. Los usuarios pueden obtener un grupo de ítems que, juntos, pueden ser interesantes para el usuario.
- Búsqueda simple. Frecuentemente el usuario navega a través de un catálogo sin ninguna intención inminente. En este caso los SRs ayudan al usuario a ver los ítems que probablemente entren dentro del espectro de interés del usuario.
- Encontrar un SR creíble. Algunos usuarios no confían en los SRs así que algunas aplicaciones brindan funcionalidades para permitir a los usuarios evaluar sus comportamientos.
- Mejorar el perfil. Los usuarios pueden indicarle al sistema qué les gusta o les desagrada para mejorar la experiencia de usuario.

El rol de los SRs dentro de un sistema informático puede ser bastante diverso. Esta diversidad provoca la necesidad de la explotación de un amplio rango de técnicas con el objetivo de producir las recomendaciones correctas.

## TÉCNICAS DE RECOMENDACIÓN

Con el objetivo de cumplir con su principal función, identificar los ítems útiles para el usuario, los SRs deben predecir los ítems con valor para ser recomendados. Luego, el sistema debe ser capaz de predecir la utilidad de estos o, al menos, comparar la utilidad de algunos ítems para, entonces, decidir qué ítems recomendar basándose en esa comparación. El paso de predecir puede no estar explícito en el algoritmo de recomendación pero es el rol que va a jugar cualquier SR.

El enfoque lógico del paso de predicción de los SRs, es que, en ausencia de información precisa acerca de las preferencias de un usuario específico, se puede asumir que los ítems más populares, los que gustan a una gran cantidad de usuarios, probablemente serán de utilidad para este usuario en particular, al menos es más probable que cualquier ítem seleccionado aleatoriamente. Por lo tanto, en este caso, se predice que la utilidad de estos ítems populares es razonablemente alta para ese usuario específico (Adomavicius and Tuzhilin 2005).

En el artículo: “Toward the next generation of recommender system” (Adomavicius and Tuzhilin 2005), se modela la utilidad de un ítem  $i$  para un usuario  $u$  como el valor de la función  $R(u,i)$ , esta es la manera en que se calcula la utilidad en los sistemas que utilizan el filtrado colaborativo teniendo en cuenta las valoraciones que los usuarios han dado a los distintos ítems. Luego, la tarea fundamental de los SRs es predecir el valor de  $R$  sobre pares de usuarios – ítems, llamando  $R'$  a esta predicción. Consecuentemente, habiendo calculado esta predicción para un usuario activo  $u$  sobre un conjunto de ítems:  $R'(u,i_1), \dots, R'(u,i_N)$ , el sistema recomendará los ítems  $i_{j_1}, \dots, i_{j_K}$  ( $K \leq N$ ) que tengan las utilidades más altas según la predicción. Normalmente  $K$  es un número pequeño, al menos mucho más pequeño que la cantidad de elementos presentes en el conjunto de datos que debe ser evaluado para predecir la utilidad.

Algunos SRs no estiman completamente la utilidad de un ítem antes de realizar una recomendación, sin embargo, estos pueden aplicar alguna heurística que permita decir que un ítem es útil para un usuario hipotéticamente. Esto se realiza típicamente en los sistemas basados en conocimiento. Estas predicciones se calculan con algún algoritmo para luego utilizar un conocimiento diverso sobre los usuarios, ítems, y sobre la misma función de utilidad (Burke 2007). Por ejemplo, el sistema puede asumir que la función de utilidad devuelve valores “*booleanos*” y por lo tanto solo podrá determinar si un ítem es útil

o no para un usuario. Consecuentemente, asumiendo que se tiene algún conocimiento acerca del usuario que está solicitando la recomendación, conocimiento sobre los ítems y sobre otros usuarios que han recibido recomendaciones, el sistema puede ponderar este conocimiento mediante un algoritmo para generar varias predicciones de utilidad (Burke 2007).

Es importante destacar que, en algunos casos, la utilidad de un ítem para un usuario depende de otras variables que se llaman “contextuales” genéricamente (Adomavicius, Sankaranarayanan et al. 2005). Por ejemplo, la utilidad de un ítem para un usuario puede estar influenciada por el conocimiento sobre el dominio del usuario o puede depender del momento en el que se solicita la recomendación, o el usuario puede estar interesado sólo en los ítems más cercanos a su posición actual. Esto provoca que las recomendaciones tengan que ser adaptadas a esos detalles adicionales y, como resultado, la tarea de estimar correctamente las recomendaciones a realizar se vuelve mucho más difícil.

La taxonomía propuesta en el artículo “Hybrid web recommender systems” (Burke 2007) se ha convertido en un clásico a la hora de distinguir los distintos tipos de SRs, en dicho artículo se proponen 6 enfoques de recomendación diferentes:

- Sistemas de recomendación basados en contenido: Este tipo de SR computan las recomendaciones basándose en las características de los ítems que el usuario ha valorado positivamente en el pasado.
- Sistemas de recomendación colaborativos (SRC): Estos sistemas usan las valoraciones de los usuarios para categorizarlos en grupos de acuerdo a la similitud de sus preferencias. Las recomendaciones se infieren teniendo en cuenta las valoraciones de los usuarios categorizados en un mismo grupo.
- Sistemas de recomendación demográficos: Estos categorizan a los usuarios en grupos demográficos basándose en atributos personales. Los usuarios, entonces, reciben recomendaciones de acuerdo a los grupos en los que se encuentran clasificados.
- Sistemas de recomendación basados en conocimiento: Estos sistemas infieren las recomendaciones basándose en las necesidades del usuario y en el conocimiento de las características de los ítems. Típicamente, computan las recomendaciones usando procesos de razonamiento basado en casos, por ejemplo: el usuario provee un ejemplo de ítem similar al que le interesa y el sistema, a partir de este

conocimiento, infiere un perfil de ítem que permita encontrar los mejores ítems en el espacio de búsqueda.

- Sistemas de recomendación basados en comunidades: Este tipo de SR sugiere los ítems basándose en las preferencias de las “amistades” del usuario en cuestión. Esta técnica sigue el epigrama “Dime con quién andas y te diré quién eres” (Ben-Shimon, Tsikinovsky et al. 2007; Arazy, Kumar et al. 2009).
- Sistemas de recomendación híbridos. Este tipo de sistema se basa en la combinación de cualquiera de las técnicas mencionadas anteriormente. Un sistema híbrido que combina las técnicas A y B trata de aprovechar las ventajas de A para solucionar las desventajas de B.

De acuerdo con varias investigaciones (Manouselis, Vuorikari et al. 2010; Redpath, Glass et al. 2010), la aplicación de los sistemas de recomendación colaborativos es el enfoque más exitoso para realizar recomendaciones.

### **1.5 Sistemas de Recomendación Colaborativos**

Como se había mencionado anteriormente los SRC computan las recomendaciones a partir de la clasificación de los usuarios en grupos con valoraciones similares. Luego, el problema de la recomendación puede ser definido como: la tarea de estimar la respuesta de un usuario para un ítem que es nuevo para él, basándose en la información histórica almacenada en el sistema. El tipo de respuesta usuario – ítem varía de una aplicación a otra y se puede categorizar de 3 maneras: escalar, binaria y unitaria (Desrosiers and Karypis 2011). Las respuestas escalares son numéricas (ejemplo: valoraciones de 1 a 5 estrellas) u ordinales (ejemplo: Muy de acuerdo, De acuerdo, Neutral, En desacuerdo, Muy en desacuerdo) siempre representando los posibles niveles de apreciación de los usuarios respecto a los ítems. Las respuestas binarias solo tienen dos posibles valores donde se capturan dos niveles de apreciación opuestos (ejemplo: Me gusta, No me gusta). Finalmente, las respuestas unitarias capturan la interacción de un usuario con un ítem (ejemplo: compra, acceso online, etc.) sin brindar una información explícita de la apreciación del usuario por el ítem en cuestión.

Las maneras en las que las respuestas del usuario son capturadas pueden ser distintas también. Una manera es solicitar explícitamente una valoración al usuario pero otra manera podría ser el almacenamiento del historial de compras o acceso de un usuario a un grupo de ítems (Konstan, Miller et al. 1997).

Se puede decir, entonces, que la idea central de los SRC es que la valoración de un usuario  $u$  para un nuevo ítem  $i$  será probablemente similar a la valoración de otro usuario  $v$  sobre  $i$  si  $u$  y  $v$  han valorado otros ítems de manera similar. El enfoque colaborativo resuelve alguna de las limitaciones del enfoque basado en contenido. Por ejemplo, los ítems para los cuales no se tiene información de sus características, o esta información resulta difícil de obtener, pueden ser recomendados a través de la retroalimentación provista por otros usuarios. Además, las recomendaciones obtenidas por un SRC se basan en la calidad de los ítems según el criterio de valoración de los pares del usuario activo en vez de basarse en el contenido, pues este puede ser un mal indicador de la calidad (Desrosiers and Karypis 2011). Finalmente el enfoque colaborativo puede recomendar ítems con un contenido diametralmente distinto siempre y cuando otros usuarios hayan mostrado interés por estos ítems.

Las recomendaciones en los SRC pueden ser obtenidas de 2 maneras: “usuario – usuario” o “ítem – ítem” (Hill, Stead et al. 1995; Deshpande and Karypis 2004). En los sistemas que se basan en la técnica “usuario - usuario” como GroupLens (Konstan, Miller et al. 1997), Bellcore’s video (Hill, Stead et al. 1995) y Ringo (Shardanand and Maes 1995), se evalúa el interés de un usuario  $u$  por un ítem  $i$  usando las valoraciones de otros usuarios, llamados vecinos, sobre  $i$ . Los vecinos son los usuarios que tienen patrones de valoración similares a  $u$ . En los sistemas que utilizan la técnica “ítem – ítem” (Deshpande and Karypis 2004), por el contrario, se predice la valoración de un usuario  $u$  para un ítem  $i$  basándose en las valoraciones de  $u$  sobre ítems similares a  $i$ . En este tipo de enfoque dos ítems son similares si varios usuarios del sistema los han evaluado de manera similar.

Entre las ventajas de utilizar un SRC se pueden mencionar:

- Simplicidad. Este método es intuitivo y relativamente fácil de implementar. En su forma más simple, solo un parámetro (el número de vecinos utilizados en la predicción) necesita ser ajustado (Desrosiers and Karypis 2011).
- Justificación de los resultados. Este método, además, provee una manera concisa e intuitiva de justificar las predicciones. Por ejemplo, en una recomendación “ítem - ítem”, la lista de ítems vecinos y las valoraciones dadas a estos ítems por los usuarios pueden ser presentados al usuario activo como una justificación de las recomendaciones realizadas. Esto puede ayudar a que el usuario entienda mejor la recomendación y su relevancia y puede servir de base para un sistema

interactivo en el que los usuarios pueden seleccionar los vecinos que desea sean tomados en cuenta con mayor peso al realizar una recomendación (Bell, Koren et al. 2007).

- Eficiencia. Uno de los puntos “fuertes” de este método es que no requieren ningún tipo de fases de “entrenamiento” (al contrario de sistemas que utilizan técnicas de inteligencia artificial) (Desrosiers and Karypis 2011).
- Estabilidad. Otra propiedad importante de los SRC es que no son afectados por la constante adición de usuarios, ítems y valoraciones (Desrosiers and Karypis 2011).

### 1.5.1 Recomendación “Usuario - Usuario”

Los SRC automatizan el concepto de “rumor”, donde se confía en la opinión de personas con similitud de pensamiento u otra fuente confiable para evaluar la relevancia de un ítem (película, libro, etc.). En la figura 3 se muestra la representación del espacio de un problema asociado a un SRC.

	The Matrix	Titanic	Die Hard	Forrest Gump	Wall-E
John	5	1		2	2
Lucy	1	5	2	5	5
Eric	2	?	3	5	4
Diane	4	3	5	3	

Figura 3. Representación del espacio de un problema asociado a un SRC. Tomado de: Ricci, F.: Recommender Systems Handbook.

### PREDICCIÓN “USUARIO - USUARIO”

En el método de predicción “usuario - usuario”, se predice la valoración  $r_{ui}$  de un usuario  $u$  para un ítem  $i$  usando las valoraciones dadas a  $i$  por los usuarios vecinos de  $u$ . Suponiendo que para cada usuario  $v$  distinto de  $u$  se tiene un valor  $w_{uv}$  que representa la similitud de las preferencias entre  $u$  y  $v$ , los  $k$  vecinos más cercanos ( $k$ -NN) de  $u$ , denotados por  $N(u)$ , son los  $k$  usuarios con la mayor similitud  $w_{uv}$  a  $u$ . Sin embargo, solo pueden ser usados en la predicción usuarios que hayan valorado el ítem  $i$ . Se puede

escribir este subconjunto de vecinos como  $N_i(u)$ . El rating  $r_{ui}$  puede ser estimado como la valoración promedio dada a  $i$  por estos vecinos (Desrosiers and Karypis 2011):

$$\text{Fórmula 1: } r'_{ui} = \frac{1}{|N_i(u)|} \sum_{v \in N_i(u)} r_{vi}$$

Un problema con la Fórmula 1, es que no tiene en cuenta el hecho de que los vecinos pueden tener niveles diferentes de similitud. Considerando la figura 3, si dos vecinos cercanos de Eric son Lucy y Diane, sería poco inteligente considerar con igual peso sus valoraciones sobre la película "Titanic" dado que los gustos de Lucy son mucho más cercanos a los de Eric que los de Diane. Una solución común a este problema es ponderando la contribución de cada vecino de acuerdo a su similitud con  $u$  de la siguiente manera (Desrosiers and Karypis 2011):

$$\text{Fórmula 2: } r'_{ui} = \frac{\sum_{v \in N_i(u)} w_{ui} r_{vi}}{\sum_{v \in N_i(u)} |w_{uv}|}$$

La Fórmula 2 tiene una falla importante: no considera el hecho de que varios usuarios pueden dar distintas valoraciones al cuantificar el mismo nivel de apreciación por un ítem. Este problema normalmente se resuelve normalizando las valoraciones  $r_{vi}$ , y la predicción quedaría como sigue (Desrosiers and Karypis 2011):

$$\text{Fórmula 3: } r'_{ui} = h^{-1} \left( \frac{\sum_{v \in N_i(u)} w_{uv} h(r_{vi})}{\sum_{v \in N_i(u)} |w_{uv}|} \right)$$

Nótese que la valoración predicha debe convertirse a la escala original es por esto que se introduce  $h^{-1}$  en la ecuación.

#### CLASIFICACIÓN "USUARIO - USUARIO"

La clasificación "usuario - usuario", encuentra la valoración más probable que daría un usuario  $u$  a un ítem  $i$ , teniendo el voto de los vecinos más cercanos de  $u$  sobre esta valoración. El voto  $v_{ir}$  dado por los  $k$ -NN de  $u$  para la valoración  $r$  que pertenece a  $S$  puede ser obtenido como la sumatoria de los coeficientes de similitud de los vecinos que han dado esa valoración  $r$  a  $i$  (Desrosiers and Karypis 2011):

$$\text{Fórmula 4: } v_{ir} = \sum_{v \in N_i(u)} \partial(r_{vi} = r) w_{uv}$$

Donde  $\partial(r_{vi} = r)$  es 1 si  $r_{vi} = r$  y 0 en caso contrario. Una vez que se realiza el cálculo para cada posible valoración, la predicción es simplemente el valor  $r$  para el cual  $v_{ir}$  toma el mayor valor.

Un método de clasificación que parte de la normalización de las valoraciones puede ser definido como (Desrosiers and Karypis 2011):

Fórmula 5:

$$r'_{ui} = h^{-1} \left( \underset{r \in S_r}{\operatorname{argmax}} \sum_{v \in N_i(u)} \partial(h(r_{vi}) = r) w_{uv} \right)$$

La opción de implementar el método de predicción o clasificación “usuario - usuario” depende, en gran medida, de la escala de valoración. Si el sistema utiliza una escala continua que va, por ejemplo, desde 1 a 10, es recomendable la utilización del método de predicción. Por el contrario si la escala de valoración contiene pocos valores discretos por ejemplo “Me gusta” y “No me gusta” o si los valores de la escala no pueden ser ordenados de una manera trivial, es recomendable la utilización del método de clasificación.

### 1.5.2 Recomendación “Ítem - Ítem”

Mientras los métodos “usuario - usuario” dependen de la opinión de usuarios similares al activo para predecir una valoración, los “ítem – ítem” se basan en las valoraciones dadas a ítems similares. En el ejemplo mostrado en la figura 3 Eric, determinaría si la película “Titanic” es “buena” para él considerando las películas que ya ha visto. Eric observa que las personas que han valorado “Titanic” han dado valoraciones similares a las películas “Forrest Gump” y “Wall-E”. Dado que Eric valoró positivamente estas dos películas puede concluir, entonces, que la película “Titanic” también obtendrá una valoración positiva por su parte (Desrosiers and Karypis 2011).

Esta idea puede ser formalizada como sigue: teniendo que  $N_u(i)$  son los ítems más similares a  $i$  de los valorados por el usuario, la predicción de la valoración que daría  $u$  a  $i$  se obtiene como el promedio ponderado de las valoraciones dadas por  $u$  a los ítems de que pertenecen a  $N_u(i)$  (Desrosiers and Karypis 2011):

$$\text{Fórmula 6: } r'_{ui} = \frac{\sum_{j \in N_u(i)} w_{ij} r_{uj}}{\sum_{j \in N_u(i)} |w_{ij}|}$$

De manera análoga a los métodos “usuario - usuario”, las diferencias entre las escalas de valoración individuales pueden ser consideradas normalizando las valoraciones:

$$\text{Fórmula 7: } r'_{ui} = h^{-1} \left( \frac{\sum_{j \in N_u(i)} w_{ij} h(r_{uj})}{\sum_{j \in N_u(i)} |w_{ij}|} \right)$$

También se puede definir un enfoque de clasificación “ítem - ítem”. En este caso los ítems  $j$  valorados por usuario  $u$  votan por la valoración que debe ser dada a un nuevo ítem  $i$ , y estos votos son ponderados por la similitud existente entre  $i$  y  $j$ . La versión normalizada de este enfoque puede ser expresada de la siguiente manera (Desrosiers and Karypis 2011):

Fórmula 8:

$$r'_{ui} = h^{-1} \left( \underset{r \in S}{\operatorname{argmax}} \sum_{v \in N_u(i)} \partial(h(r_{uj}) = r) w_{ij} \right)$$

### 1.5.3 Componentes de los métodos de recomendación colaborativa

Además de los atributos cruciales presentados anteriormente, se deben tener en cuenta tres consideraciones importante para la implementación de un SRC ya sea “usuario - usuario” o “ítem - ítem”, estas son: la normalización de las valoraciones, el cálculo de los coeficientes de similitud y la selección de los vecinos.

#### NORMALIZACIÓN DE LAS VALORACIONES

Dos de los esquemas de normalización más populares que han sido propuestos para convertir las valoraciones individuales hacia una escala más universal son los esquemas “*mean-centering*” y “*Z-score*”.

La idea de “*mean-centering*” es determinar si una valoración es positiva o negativa mediante la comparación con una valoración media. En las recomendaciones “usuario - usuario” una valoración  $r_{ui}$  es transformada a una normalizada  $h(r_{ui})$  mediante la substracción de  $r_{ui}$  al promedio  $\bar{r}_u$  de las valoraciones dadas por el usuario  $u$  (Breese, Heckerman et al. 1998):

$$\text{Fórmula 9: } h(r_{ui}) = r_{ui} - \bar{r}_u$$

Utilizando este esquema la predicción de la valoración  $r_{ui}$  en una recomendación “usuario – usuario” sería:

$$\text{Fórmula 10: } r'_{ui} = \bar{r}_u + \frac{\sum_{v \in N_i(u)} w_{uv}(r_{vi} - \bar{r}_v)}{\sum_{v \in N_i(u)} |w_{uv}|}$$

De la misma manera, la normalización usando “*mean-centering*” para “ítem - ítem” sería:

$$\text{Fórmula 11: } h(r_{ui}) = r_{ui} - \bar{r}_i$$

Donde  $\bar{r}_i$  sería el promedio de las valoraciones recibidas por  $i$ . Esta técnica de recomendación se usa con mayor frecuencia en recomendaciones “ítem – ítem” (Breese, Heckerman et al. 1998). Combinando las fórmulas 7 y 11,  $r_{ui}$  se predice:

$$\text{Fórmula 12: } r'_{ui} = \bar{r}_i + \frac{\sum_{j \in N_u(i)} w_{ij}(r_{uj} - \bar{r}_j)}{\sum_{j \in N_u(i)} |w_{ij}|}$$

En el esquema de normalización “*Z-score*” considera la dispersión de las escalas individuales de valoración. Para las recomendaciones “usuario - usuario” la normalización de la valoración  $r_{ui}$  divide el cálculo de la normalización provista por “*mean-centering*” entre la desviación estándar  $\delta_u$  de las valoraciones dadas por un usuario  $u$  (Herlocker, Konstan et al. 1999):

$$\text{Fórmula 13: } h(r_{ui}) = \frac{r_{ui} - \bar{r}_u}{\delta_u}$$

Una recomendación “usuario - usuario” utilizando este esquema de normalización se obtendría mediante la fórmula:

$$\text{Fórmula 14: } r'_{ui} = \bar{r}_u + \delta_u \frac{\sum_{v \in N_i(u)} w_{uv}(r_{vi} - \bar{r}_v) / \delta_u}{\sum_{v \in N_i(u)} |w_{uv}|}$$

Asimismo, la normalización “*Z-score*” para  $r_{ui}$  en los métodos “ítem - ítem” se obtiene dividiendo la normalización provista por “*mean-centering*” entre la desviación estándar  $\delta_u$  de las valoraciones dadas para el ítem  $i$  (Herlocker, Konstan et al. 1999):

$$\text{Fórmula 15: } h(r_{ui}) = \frac{r_{ui} - \bar{r}_i}{\delta_i}$$

La predicción “ítem - ítem” de la valoración  $r_{ui}$  sería entonces:

$$\text{Fórmula 16: } r'_{ui} = \bar{r}_i + \delta_i \frac{\sum_{j \in N_u(i)} w_{ij} (r_{uj} - \bar{r}_j) / \delta_j}{\sum_{j \in N_u(i)} |w_{ij}|}$$

Comparando los esquemas “*mean-centering*” y “*Z-score*”, se puede decir que el segundo provee el beneficio adicional de considerar la varianza de las recomendaciones provistas por usuarios o para ítems individualmente. Esto es particularmente útil si la escala de valoración tiene un rango amplio de valores discretos o si este rango es continuo (Howe and Forbes 2008). Teniendo en cuenta esto último, para seleccionar un esquema de normalización se debe valorar qué tan dispersos pueden estar las valoraciones de los usuarios en el contexto específico del SRC que se desea implementar.

### CÁLCULO DEL COEFICIENTE DE SIMILITUD

El coeficiente de similitud juega dos roles esenciales en los SRC: permite la selección de vecinos confiables para utilizar sus valoraciones en las predicciones y provee los medios para asignar, mayor o menor importancia a estos vecinos en la predicción. El cálculo de los coeficientes de similitud es uno de los aspectos críticos en la construcción de un SRC ya que puede tener un impacto significativo en el rendimiento y en la precisión de este (Desrosiers and Karypis 2011).

Una medida de la similitud entre dos objetos  $a$  y  $b$ , usada frecuentemente en la captura de información, consisten en la representación de esos objetos como dos vectores  $x_a$  y  $x_b$  para luego calcular la similitud del vector coseno (CV por sus siglas en inglés) resultante entre estos vectores (Lang 1995):

$$\text{Fórmula 17: } \text{COS}(x_a, x_b) = \frac{x_a x_b}{\|x_a\| \|x_b\|}$$

En el contexto de recomendación de ítems, esta medida puede ser utilizada para calcular las similitudes entre usuarios considerando al usuario  $u$  como un vector  $x_u$  donde  $x_{ui} = r_{ui}$  si el usuario  $u$  ha valorado el ítem  $i$  y 0 en caso contrario. Esta similitud entre dos usuarios  $u$  y  $v$  puede ser calculada como:

$$\text{Fórmula 18: } CV(u, v) = \text{COS}(x_u, x_v) = \frac{\sum_{i \in I_{uv}} r_{ui} r_{vi}}{\sqrt{\sum_{i \in I_u} r_{ui}^2 \sum_{j \in I_v} r_{vj}^2}}$$

Donde  $I_{uv}$  denota los ítems valorados por  $u$  y  $v$ . Un problema con esta medida es que no tienen en cuenta las diferencias en la media y la varianza de las valoraciones realizadas por  $u$  y  $v$ .

Otra medida popular que compara las valoraciones y donde los efectos de la media y la varianza han sido mitigados es el cálculo de la similitud utilizando la correlación de Pearson (PC por sus siglas en inglés):

$$\text{Fórmula 19: } PC(u, v) = \frac{\sum_{i \in I_{uv}} (r_{ui} - \bar{r}_u)(r_{vi} - \bar{r}_v)}{\sqrt{\sum_{u \in U_{ij}} (r_{ui} - \bar{r}_i)^2 \sum_{u \in U_{ij}} (r_{uj} - \bar{r}_j)^2}}$$

La misma idea puede ser utilizada para obtener la similitud entre dos ítems  $i$  y  $j$ , esta vez comparando las valoraciones hechas por usuarios que han valorado ambos ítems (Deshpande and Karypis 2004):

$$\text{Fórmula 20: } PC(i, j) = \frac{\sum_{u \in U_{ij}} (r_{ui} - \bar{r}_i)(r_{uj} - \bar{r}_j)}{\sqrt{\sum_{u \in U_{ij}} (r_{ui} - \bar{r}_i)^2 \sum_{u \in U_{ij}} (r_{uj} - \bar{r}_j)^2}}$$

Mientras el signo del coeficiente de similitud indica si la correlación es directa o inversa, su magnitud (que varía entre 0 y 1) representa la fuerza de la correlación.

Las diferencias entre las escalas de evaluación de los usuarios son, frecuentemente, más pronunciadas que las diferencias en las valoraciones dadas a los ítems individualmente. Por lo tanto, a la hora de calcular la similitud entre ítems, es apropiado comparar las valoraciones normalizadas por la media del usuario en vez compararlas normalizadas por la media de las valoraciones que ha recibido el ítem. La similitud del “Coseno Ajustado” (AC por sus siglas en inglés), es una modificación del cálculo de la similitud usando PC donde se efectúa la comparación utilizando valoraciones normalizadas mediante el uso de la media de valoraciones del usuario en cuestión (Desrosiers and Karypis 2011):

$$\text{Fórmula 21: } AC(i, j) = \frac{\sum_{u \in U_{ij}} (r_{ui} - \bar{r}_u)(r_{uj} - \bar{r}_u)}{\sqrt{\sum_{u \in U_{ij}} (r_{ui} - \bar{r}_u)^2 \sum_{u \in U_{ij}} (r_{uj} - \bar{r}_u)^2}}$$

Se ha demostrado que, en la mayoría de los casos, la similitud AC funciona mejor que la similitud PC para la predicción de valoraciones utilizando el método “Ítem - Ítem” (Desrosiers and Karypis 2011).

#### **1.5.4 Problemas de los sistemas de recomendación colaborativos**

A pesar del buen desempeño general de los SRC, estos presentan algunas debilidades o limitaciones:

- Problema de la oveja gris. Una oveja gris es aquel usuario que se encuentre en la frontera entre dos grupos de usuarios, estos usuarios típicamente no obtienen buenas recomendaciones.
- Conjunto de datos históricos. La adaptabilidad de los SRC debido al manejo de información histórica es una ventaja, pero puede convertirse en una desventaja cuando el conjunto de datos históricos es pequeño.
- Problema de arranque en frío. Este problema se presenta para usuarios e ítems. Cuando un nuevo usuario accede al sistema, este no tiene ninguna información acerca de él o esta información es escasa. Por lo tanto, como el sistema no puede compararlo con los restantes usuarios de la base de datos, entonces, no puede proveer una recomendación (Massa and Bhattacharjee 2004). Además, cuando un nuevo ítem se adiciona al sistema y ningún usuario lo ha valorado tampoco puede ser recomendado (Trujillo, Millan et al. 2007).

Para mitigar estos problemas se propone la hibridación con otro tipo de SR o la utilización de técnicas de recomendación triviales hasta que el SRC cuente con la información necesaria para ofrecer una recomendación (Said 2010).

#### **1.6 Uso de los GIS con los SRs de productos turísticos**

Un GIS es un sistema de software diseñado para permitir la entrada, el almacenamiento, la edición, la captura, el análisis y la salida de información geográficamente referenciada (DeMers 2008). Los GIS son herramientas con un alto valor para los servicios turísticos porque los productos que se ofrecen en este sector normalmente pueden ser ubicados en un mapa.

El diseño de la interacción y la GUI de un SR pueden afectar profundamente el proceso de toma de decisiones del usuario. La utilización de un GIS como vía de representar las recomendaciones de productos turísticos obtenidas por un SR, brinda la ventaja de permitir al usuario la visualización geográfica de las entidades e ítems recomendados de una manera intuitiva, permitiéndoles, entre otras cosas, evaluar la distancia que existe entre ítems de los que puede convertirse en un consumidor. Por lo que al pensar en la

implementación de un SR para productos turísticos, la utilización de un GIS para la presentación de los resultados se torna prácticamente obvia.

### **1.7 Conclusiones Parciales**

La utilización de SRs se ha convertido en una manera muy popular y extremadamente usada para facilitar el acceso a la información relevante contenida en un amplísimo universo de información presente en la web. En este contexto, la propuesta de implementación de un SRC ha sido la más exitosa cuando se quiere lograr este objetivo.

Debido a la importancia de la industria del turismo en Cuba, se debe considerar la aplicación de los SRs como manera de potenciar, entre otras cosas, los ingresos del sector. Para hacer esto es recomendable la utilización de un SRC siempre y cuando se mitigue el problema del arranque en frío, que puede ser logrado con facilidad mediante la aplicación de alguna técnica de recomendación trivial.

La utilización de los GIS posee una elevada utilidad en el sector ya que ayuda a elevar la calidad del proceso de toma de decisiones del cliente potencial mediante la ubicación espacial de los productos ofertados.

## **CAPÍTULO 2. PROPUESTA DE TÉCNICA**

El estudio realizado en el Capítulo 1 sirvió para desglosar los principales fundamentos teóricos que permitirán obtener una técnica para recomendar productos turísticos en Cuba. En lo que respecta a esta investigación la palabra técnica se refiere al conjunto de reglas que permitirán obtener un sistema capaz de predecir el contenido que debe ser mostrado prominentemente a los usuarios de los sitios de e-turismo cubanos.

En el presente capítulo se define como debe ser implementado un sistema para la recomendación de productos turísticos cubanos. Primero se describe una regla que define una estructura general para el sistema y luego se definen un conjunto de reglas para cada uno de los componentes de esta estructura teniendo en cuenta las características contextuales del entorno turístico en Cuba. Para cada componente se define el comportamiento que debe ser implementado así como sus responsabilidades, la información de entrada – salida y las relaciones que pudieran establecerse con otros componentes.

### **2.1 Estructura General**

Aunque fundamental, la manera en que se calculan las predicciones de valoración, de un usuario activo a uno o varios ítems, no es el único componente relevante para que un sistema sea capaz de efectuar una recomendación. La manera en que la información es representada así como la forma de retroalimentación del sistema también constituyen características importantes para lograr una recomendación. Además el análisis del contexto en el que se encuentra el usuario en un instante dado para decidir la forma en que se va a predecir las valoraciones es otro elemento crucial dentro del proceso de recomendación.

En la figura 4 se define la estructura general que debe seguir el sistema, donde se ilustran como componentes principales la Interfaz Gráfica de Usuario, el Selector de Recomendación y el Motor de Recomendación.

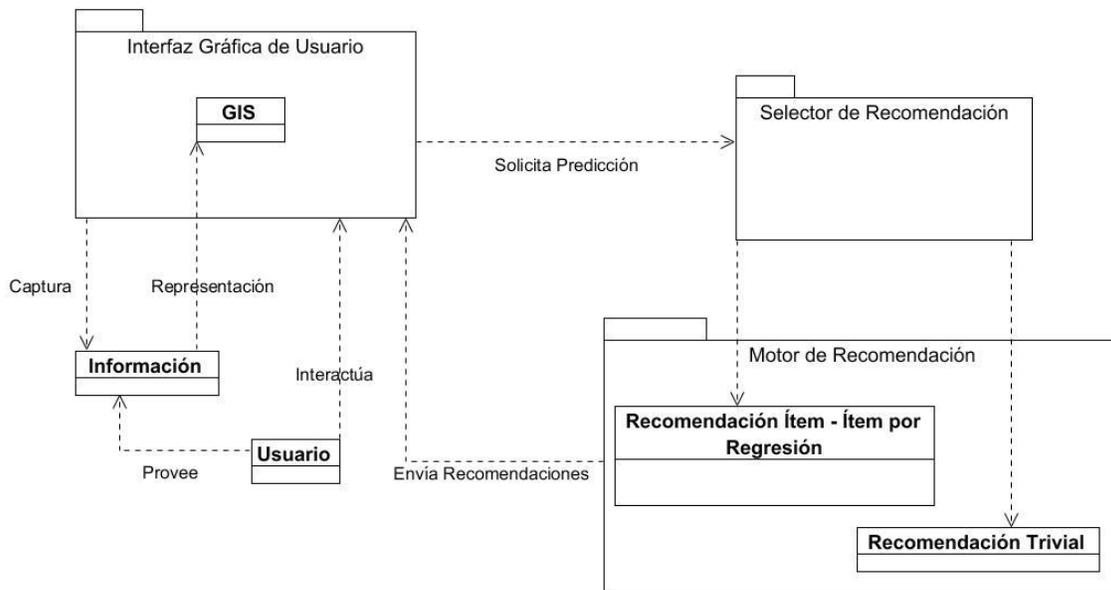


Figura 4. Estructura general del sistema.

## 2.2 Interfaz Gráfica de Usuario

La responsabilidad de este componente es la de garantizar la captura de la información que provee el usuario a partir de sus entradas de datos o de sus acciones sobre el sistema; además debe brindar una representación correcta de la información de manera que el usuario se sienta cómodo en su interacción con el sistema.

Luego para describir este componente se deben analizar por separado sus dos responsabilidades principales: la representación y captura de la información.

### *REPRESENTACIÓN DE LA INFORMACIÓN*

La información que debe ser representada es, básicamente, la referente a los productos turísticos que la organización que utilice el e-turismo desea ofertar a los usuarios. La manera en que esta información es presentada incide directamente en cómo los usuarios interactúan con el sistema así como el nivel de aceptación que estos le confieren.

Actualmente la mayoría de las principales organizaciones que utilizan el e-turismo en Cuba no van más allá de ofrecer a sus potenciales clientes los listados de todos los productos que ofrecen permitiendo la posibilidad de acceder, en algunos casos, a información detallada sobre estos. Esta forma de proceder no potencia un componente esencial de los productos turísticos: su ubicación geográfica. A la hora de explotar esta potencialidad de los productos turísticos se persigue sobre todo ubicar al cliente

espacialmente, esto se podría lograr incluyendo la dirección en los detalles del producto suponiendo que todos los clientes poseen un vasto conocimiento de la región geográfica en cuestión, lo que en la mayoría de los casos es una premisa falsa.

Para lograr el objetivo de ubicar al cliente en el espacio, entonces, sería muy útil vincular una referencia geográfica (longitud y latitud) a cada producto turístico para luego utilizar esta información en la ubicación sobre un mapa. De esta manera el usuario podría visualizar donde se encuentra el producto turístico de su interés y también se encontraría en condiciones de realizar análisis más profundos como evaluar visualmente la distancia entre dos o más productos que tentativamente pudiese estar interesado en consumir.

Para la representación de la información, luego, se propone utilizar un GIS debido a la característica inherente de los productos turísticos de poder ser ubicados geográficamente. El GIS debe mostrar un mapa de la región geográfica en cuestión donde se ubiquen los productos turísticos a ofertar y debe contar con funcionalidades básicas de zoom y paneo para permitir al usuario moverse libremente por el mapa.

Además, este componente debe representar la información referente a las recomendaciones que se le van a realizar al usuario, esta información constituye una entrada de este componente y se obtienen a partir del componente Motor de Recomendación. Se propone separar la visualización de las recomendaciones del mapa, es importante no efectuar recomendaciones sobre el mapa pues esto limitaría la variedad de productos con una buena predicción de valoración que pudieran ser recomendados cuando el usuario esté visualizando un fragmento de toda la región geográfica representada en el mapa.

### *CAPTURA DE INFORMACIÓN*

La información que debe ser capturada, dando cumplimiento a la segunda responsabilidad de este componente, es aquella referente a la opinión del usuario sobre algún ítem así como aquella asociada la descripción de la situación del usuario en el sistema. La valoración de los usuarios sobre los ítems que le son mostrados es la base del sistema de recomendación, estas valoraciones constituyen la información sobre la cual serán calculadas todas las predicciones que utilizará el sistema para decidir que ítems debe mostrar prominentemente a un usuario u otro.

Para el contexto de las organizaciones dedicadas al turismo en Cuba se propone la utilización de un método de captura de las valoraciones explícito y con una escala continua; esta manera debe implementarse de forma tal que no se interrumpa la navegación del usuario. Por la relación que guarda con la actividad turística y la manera en que son categorizados la mayoría de las instalaciones de alojamiento en Cuba se propone usar una escala de valoración de 1 a 5 esta escala podría ser presentada en forma de estrellas cuando el usuario esté visualizando la información detallada de algún ítem así este podría valorar la calidad de un servicio o producto de 1 a 5 estrellas sin ver afectada, además, su interacción con el sitio como lo sería si se le mostrara un formulario adicional solo para que la valoración sea provista.

Por otro lado, la información situacional o contextual del usuario en un instante de tiempo determinado es la base que utilizará el componente Selector de Recomendación para decidir qué tipo de predicción debe utilizar para obtener el listado de ítems que debe ser recomendado. En este componente debe implementarse la captura de la información relevante a ser usada por el selector de recomendación, esta información sería la referencia geográfica del centro del fragmento de la región geográfica que está visualizando el usuario, el nivel de zoom con que se está visualizando el mapa o si el usuario está visualizando o no los detalles de un producto turístico. La ubicación geográfica del centro de la porción de el mapa que está siendo visualizada y el nivel de zoom se utilizarán para completar, por cercanía, el conjunto de ítems que van a conformar la vecindad a partir de la cual se calculará la predicción de valoración para cada ítem, para luego obtener el listado de productos a recomendar; la información de si un usuario esta visualizando los detalles de un producto se utilizará para recomendar productos similares a este y, de esta forma, contribuir a mitigar el problema del arranque en frío. Es importante destacar que este tipo de información en particular no es persistente dado que describe una situación en un instante dado, esta información constituye una salida del componente Interfaz Gráfica de Usuario y entrada del componente Selector de Recomendación. En la figura 5 se muestra el funcionamiento general de este componente.

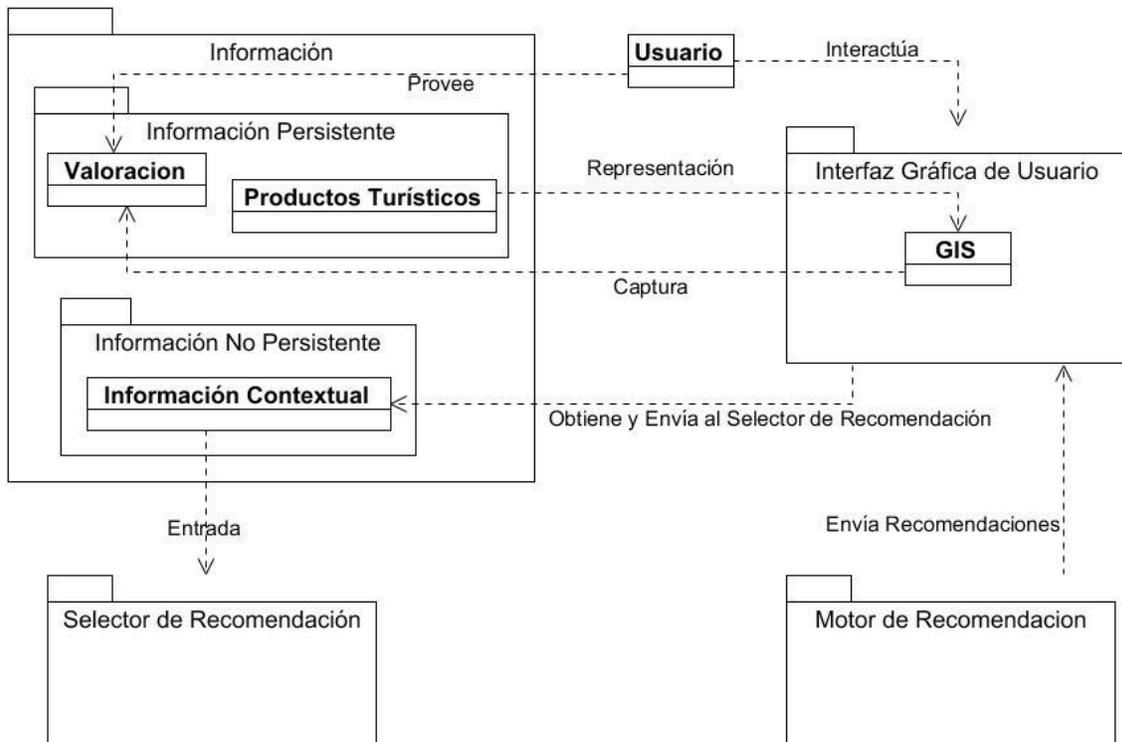


Figura 5. Funcionamiento del componente Interfaz Gráfica de Usuario.

### 2.3 Selector de Recomendación

Sin restar importancia a los demás componentes de la técnica, se debe dar una elevada relevancia a la implementación de este componente. La responsabilidad del selector de recomendación será la de analizar cada escenario en que se encuentre cada usuario en particular para decidir qué método de recomendación debe ser utilizado.

Es en este componente donde se realiza la mayor parte de la tarea correspondiente a mitigar el problema del arranque en frío. Como se había mencionado en el Capítulo 1 el problema del arranque en frío viene dado cuando no se tienen suficientes valoraciones de un usuario o sobre un ítem de manera que no se puede realizar una predicción efectiva de valoración para uno y el cálculo de la vecindad para el otro.

De manera general, la información que sirve de entrada a este componente es la información contextual del usuario activo que se había mencionado anteriormente, referencia geográfica del centro de la región que está siendo visualizada, nivel del zoom y si está siendo visualizada la descripción detallada de un producto. A esta información contextual se le debe agregar la información de la cantidad de valoraciones que ha realizado el usuario activo, esta información aunque es situacional (puede variar de un

instante de tiempo a otro) no la ofrece el componente Interfaz Gráfica de Usuario y debe ser obtenida a partir de los datos de las valoraciones persistidos por el sistema. La cantidad de valoraciones suficientes para ofrecer una predicción efectiva es un número que debe ser definido por aquel que se encuentre implementando la técnica pues depende de la situación particular (volumen de usuarios, ítems y valoraciones con que la organización debe trabajar). En cualquier caso, luego de definido el número, el trabajo con él es sencillo: si el usuario tiene las valoraciones necesarias se utiliza el método de recomendación ítem – ítem correspondiente, de lo contrario se debe utilizar un método de recomendación trivial.

Por otro lado, el dato de la referencia geográfica del centro de la región que está siendo visualizada debe ser utilizado, de ser necesario, para completar la vecindad de ítems que permitirá calcular una predicción de valoración. Por ejemplo, se puede definir una distancia  $X$  y, de no encontrarse los  $k$  ítems valorados por el usuario activo que son necesarios para predecir una valoración a cualquier ítem no valorado, se completaría el número  $k$  de ítems utilizando aquellos que se encuentren dentro de un radio  $X$  a partir del punto que marca el centro de la región que se está visualizando y de los que se obtengan una mayor predicción de valoración para el usuario activo. Aplicar esta restricción puede incidir, en la calidad del funcionamiento del sistema de recomendación que se obtenga. Esto se debe a que a la relevancia que tiene el componente geográfico de los productos turísticos, por ejemplo: para un usuario que se encuentre hospedado o tenga reservación en un hotel, existe la probabilidad de que esté más interesado en los productos turísticos relativamente cercanos al dicho hotel. Por esta razón, se asume que si el usuario está visualizando el mapa con un nivel de zoom que se presume elevado es porque ese usuario tiene un interés especial en esa porción del mapa y luego tendría sentido usar este completamiento con ítems que se encuentren en una distancia razonablemente corta. Una vez más, el número que se defina para  $X$  y para el nivel de zoom depende de la situación en particular en la que se encuentre el equipo de desarrollo cuando esté implementando la técnica.

La información de si se está visualizando los detalles de un producto sirve para decidir que se va a recomendar en caso que no se tenga un número de valoraciones suficientes para efectuar una recomendación. Si se está visualizando la descripción detallada de un producto se puede asumir que este producto está recibiendo un nivel de atención significativo y en consecuencia se deben recomendar los productos que se encuentren en

la vecindad inmediata (tengan un mayor valor coeficiente de similitud) de aquel que está siendo visualizado.

De manera general, en el cumplimiento de su responsabilidad, este componente, primero debe analizar si el usuario activo posee el número suficiente de valoraciones necesarias para realizar una recomendación. De no contar con las valoraciones suficientes se debe solicitar al Motor de Recomendación una recomendación trivial como lo sería el listado de ítems más populares (mejor valorados). En el caso de que el usuario activo no cuente con las valoraciones necesarias pero esté visualizando la información detallada de un ítem, se debe recomendar el listado con la vecindad inmediata de ese ítem. Por último, si el usuario cuenta con las valoraciones suficientes para obtener una recomendación el componente debe analizar si el usuario está haciendo énfasis en una región específica del mapa para solicitar al Motor de Recomendación el listado de ítems a recomendar usando o no el completamiento de vecindad a partir de la distancia del centro de la región que está siendo visualizada.

En la figura 6 se ilustra de manera general el comportamiento que debe ser implementado en este componente.

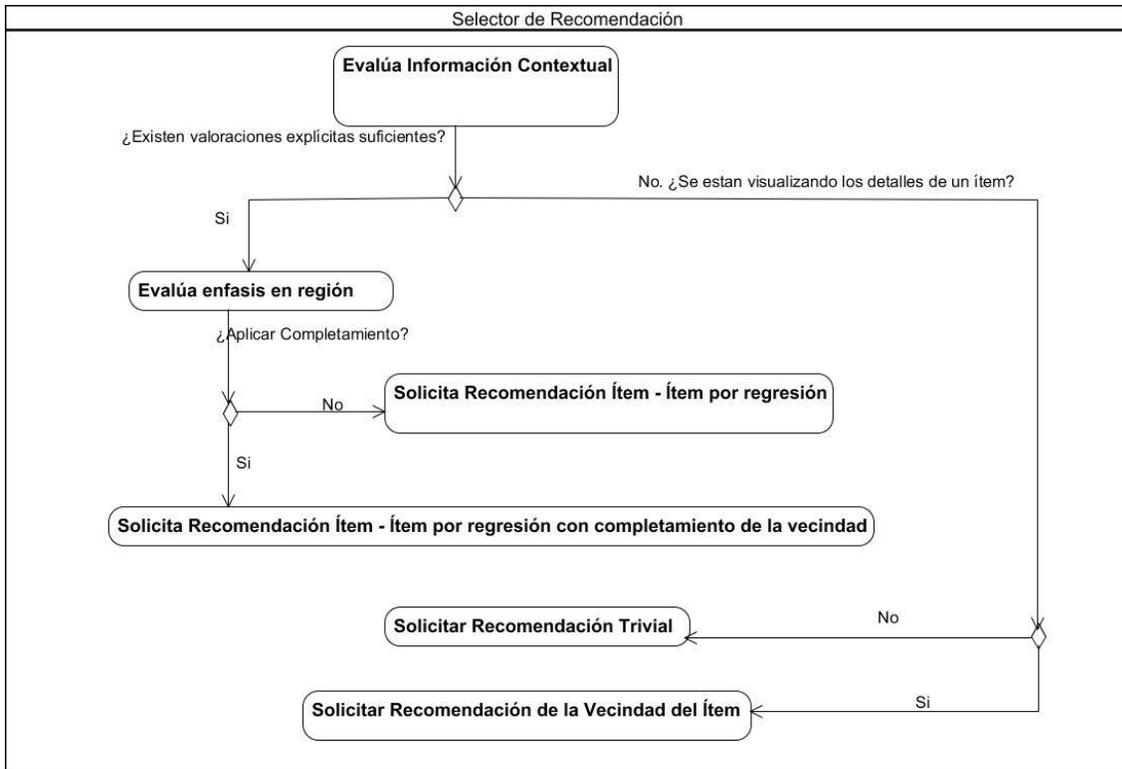


Figura 6. Funcionamiento del componente Selector de Recomendación.

## 2.4 Motor de Recomendación

La responsabilidad del Motor de Recomendación es la de obtener el listado de ítems que serán mostrados al usuario prominentemente. Como su nombre lo indica, este componente es el responsable de hacer que la herramienta obtenida sea capaz de recomendar los elementos que son, presumiblemente, de interés para el usuario. Es en este componente donde se debe implementar el comportamiento derivado del razonamiento matemático usado para predecir la valoración que un usuario activo daría a un ítem no valorado, para luego mostrar prominentemente el listado de ítems que posean una predicción de valoración más elevada.

### *ENFOQUE DE RECOMENDACIÓN*

Como se discutió en el Capítulo 1, se ha demostrado recientemente que el enfoque colaborativo es el más exitoso para efectuar recomendaciones al usuario. Principalmente porque ofrece varias y útiles ventajas y porque sus deficiencias pueden ser mitigadas con relativa facilidad. Usando este enfoque se persigue predecir las valoraciones, para los ítems de los cuales estas no se conocen, utilizando las valoraciones que los usuarios de la comunidad ya han dado. Siguiendo esta línea de pensamiento, existen dos métodos

para implementación del enfoque colaborativo: el “Usuario - Usuario” y el “Ítem – Ítem”. En el contexto turístico cubano, la población de clientes (usuarios) es usualmente más voluminosa e inestable que la de los productos disponibles para el consumo (ítems) por esto se propone la implementación del método “Ítem – Ítem”.

Al utilizar el método “Ítem – Ítem” las recomendaciones se obtendrían mediante el uso de la fórmula 7 que fue presentada en el Capítulo 1. Sin embargo, para la utilización de esta fórmula se debe calcular el coeficiente de similitud entre los ítems para determinar los k-NN que serán utilizados en la predicción así como para, también, ponderar las valoraciones que ha dado el usuario activo a los k-NN ítems. Para el cálculo de este coeficiente se debe utilizar la fórmula 21 que corresponde a la similitud AC. Además las valoraciones que se utilizan en las fórmula 7 deben ser normalizadas, para lo cual se deben utilizar la fórmulas 11 que es la correspondiente al cálculo de la predicción por regresión.

#### *FLUJO DE INFORMACIÓN*

Para cumplir con su responsabilidad, este componente debe tener disponible, como información de entrada, toda aquella asociada a las valoraciones otorgadas por los usuarios del sistema a los ítems disponibles. Esta información es la que se utilizará para el cálculo de las predicciones mediante el uso de las fórmulas definidas en el acápite anterior. Además, debe tener disponible el componente geográfico de cada uno de los ítems, en caso que deba efectuar un completamiento de la vecindad usando la distancia de estos respecto a un punto (x;y). Este punto constituye la única información que deberá proveer el Selector de Recomendación al Motor de Recomendación, y lo hará solo en el caso que se desee obtener una lista de recomendación aplicando esta restricción en particular. En cualquier caso, la información de salida que debe proveer este componente es una lista con los elementos que deben ser mostrados prominentemente (recomendados) al usuario que será obtenida a partir del tipo de recomendación que haya solicitado, en particular, el Selector de Recomendación

#### *CONSIDERACIONES EN LA IMPLEMENTACIÓN*

Durante el proceso de implementación del Motor de Recomendación se deben tener en cuenta algunas consideraciones que son relevantes. En primer lugar, para la obtención de los k-NN ítems al ítem para el cual se desea calcular la recomendación se debe tener el

coeficiente de similitud de cada ítem en particular respecto al resto de los ítems presentes en el dominio de la organización. La cantidad de coeficientes a calcular, entonces, para una población de 2000 ítems estaría cerca del millón, esto, junto a la frecuencia con que se debe utilizar esta información hace que sea recomendable calcularla con el sistema fuera de línea pues de calcularse en tiempo de ejecución el rendimiento del sistema puede verse afectado. Similarmente sucede con el cálculo de las predicciones, de manera general, se debe obtener la predicción de valoración para cada pareja usuario  $u$  a cada ítem  $i$  siempre y cuando  $u$  no haya valorado a  $i$ . Estas operaciones frecuentemente superan el millón y por lo tanto se recomienda también calcularlas con el sistema fuera de línea para no afectar el rendimiento. Se podrá obtener una predicción siempre y cuando se puedan encontrar  $k$  ítems de la vecindad de  $i$  que hayan sido valorados por  $u$ , cuando el Selector de Recomendación solicite una recomendación con completamiento, simplemente se debe completar el número  $k$  de ítems en la vecindad de  $i$ , en caso de que no se encuentren  $k$  ítems valorados por  $u$ , con los ítems cercanos geográficamente a  $i$  y de los cuales se obtenga una mayor predicción de la valoración.

Otro tema a considerar durante la implementación de este componente es la definición del valor que tendrá  $k$  en la búsqueda de los ítems más cercanos. Este valor se relaciona directamente con la precisión de la recomendación. Pero la definición del valor se debe buscar mediante la experimentación pues varía según la dispersión de los datos de las valoraciones de los usuarios del sistema al conjunto de ítems. Para  $k$  se cumple que: mientras mayor valor tenga, mayor será la precisión de la recomendación. Sin embargo, mientras mayor sea el valor de  $k$  el número de elementos para los cuales se puede obtener una predicción disminuye. En el Capítulo 3 se discute la precisión del sistema y la definición del valor de  $k$ .

Por último se encuentra la cuestión del máximo número de elementos que se devolverán en la lista de recomendación. Algunos sistemas suelen devolver un número fijo de elementos, pero esta manera de proceder puede conllevar a una sobre especialización del sistema, donde siempre se recomiendan el mismo tipo de elementos, dejando un conjunto de ítems fuera de visibilidad para el usuario. Una manera de resolver esto sería definiendo un umbral, para luego recomendar todos los elementos que tengan una predicción de valoración que supere este umbral. La definición del valor del umbral depende, desde luego, del contexto específico en el cual se encuentre la organización. Por ejemplo, si se define un umbral muy bajo, se devolverán listas con muchos

elementos, lo que atenta contra la usabilidad de la herramienta y contra el objetivo esencial del sistema que sería mostrar prominentemente los ítems que presumiblemente son de mayor interés para el usuario. Por esto se debe encontrar un punto medio en el que el umbral no permita que se devuelva un número razonablemente alto de elementos y tampoco permita que el usuario solo pueda acceder a un restringido grupo de estos dentro del dominio de la organización.

### **2.5 Lidiando con el arranque en frío**

Durante el Capítulo 1 se definió el problema del arranque en frío a la situación que se genera cuando un usuario o ítem es incorporado al sistema. En el caso del usuario nuevo no se cuenta con sus valoraciones a los ítems del sistema y como resultado no se podrán obtener recomendaciones para él. Por otro lado, cuando se incorpora un ítem nuevo al sistema no se cuenta con valoraciones de los usuarios para él y por lo tanto no podrá ser incluido en ninguna lista de recomendaciones.

Este problema, en la presente técnica se resuelve de varias maneras. Cuando el usuario no cuenta con ningún tipo de valoración en el sistema (ni implícita ni explícita) se debe solicitar al Motor de Recomendación que emita una recomendación trivial. Usualmente esta recomendación trivial constituye el listado de elementos mejor valorados (en promedio) del sistema. Además se propone recomendar los elementos dentro de la vecindad inmediata de un ítem  $i$  cuando el usuario esté visualizando los detalles de  $i$  ya que en este caso se puede asumir que el usuario está prestando un nivel de atención inusualmente mayor a  $i$ . Luego que se capturen un determinado número de valoraciones se podrán obtener predicciones y se sustituirá el método de recomendación según sea evaluado por el Selector de Recomendación. Para resolver el problema del ítem nuevo se propone mostrar permanentemente un listado de ítems que no han sido o tienen muy pocas valoraciones de la comunidad y así se garantiza la visibilidad de estos en el sistema para que puedan ser incluidos en las listas de recomendaciones luego de que sean valorados suficientemente.

### **2.6 Conclusiones Parciales**

Durante el proceso de obtención de una herramienta capaz de realizar recomendaciones se deben tener en cuenta los elementos del contexto que permitirán realizar una captura y representación pertinente de la información. Además se debe prestar atención al método de recomendación a utilizar en una situación u otra. En términos generales la estructura que propone la técnica se caracteriza, entre otras propiedades por:

- Definir cómo capturar y representar la información relevante que debe utilizar el sistema de recomendación que se desea obtener.
- Caracterizar el proceso de selección del método de recomendación que debe ser implementado.
- Definir los elementos que formarán parte del motor de recomendación encargado de predecir los ítems que presumiblemente serán de interés para el usuario.

Los mecanismos propuestos para la mitigación del problema del arranque en frío permitirán, en primer lugar, disminuir el tiempo en que un usuario o ítem se encontrará en esa situación y, en segundo lugar, que ese usuario o ítem reciba recomendaciones o sea visualizado respectivamente, lo que a la larga les permitirá salir de la situación de arranque en frío.

## CAPÍTULO 3. ANÁLISIS DE RESULTADOS

Una vez descrita la propuesta completamente se procede a analizar los resultados obtenidos con la aplicación de la misma, lo que constituye el objetivo del presente capítulo. Para probar estos resultados se implementó un Sistema de Recomendación de Productos Turísticos aplicando las reglas definidas en la técnica. Luego se pobló una base de datos con 4000 usuarios y 1000 productos turísticos para, entonces, generar aleatoriamente valoraciones de esos 4000 usuarios sobre los 1000 ítems.

Para analizar la información que brinda como resultado la herramienta implementada se utilizaron indicadores de la calidad de las recomendaciones obtenidas como lo son el Error Absoluto Medio (MAE por sus siglas en inglés), probabilidad de recomendaciones relevantes (Precisión), probabilidad de que los ítems relevantes sean recomendados (Recall) y el por ciento de ítems para los cuales se puede predecir una valoración (Cobertura).

En el presente capítulo se realizará una breve descripción de cada técnica de evaluación utilizada y luego se mostrarán los resultados obtenidos mediante el sistema a través del uso de varios valores de  $k$  para obtener los  $k$ -NN de cada ítem en cuestión.

### 3.1 Uso de la Predicción

Las métricas del uso de la predicción miden la frecuencia con la cual el sistema de recomendación toma decisiones correctas o incorrectas acerca de la utilidad de un ítem para el usuario (Herlocker, Konstan et al. 2004). Estas métricas no persiguen medir directamente la habilidad del sistema para predecir valoraciones. Por el contrario, este tipo de métricas se centran en determinar cuan útil pueden llegar a ser las recomendaciones del sistema para el usuario.

#### *PRESICIÓN Y RECALL*

La Precisión y el Recall son las métricas más usadas para la evaluación de sistemas de recuperación de información (Cleverdon, Mills et al. 1966). Con el auge de los sistemas de recomendación estas métricas se convirtieron también en populares para su evaluación (Sarwar, Karypis et al. 2000).

La Precisión y el Recall se computan a partir de una tabla de  $2 \times 2$  como se muestra en la Tabla 3. El conjunto de ítems debe ser dividido en dos clases: ítems relevantes o no

relevantes. En el caso de la técnica propuesta en el presente trabajo se considerarán relevantes para un usuario aquellos ítems que hayan obtenido una valoración de 4 o 5. Además se deben considerar los ítems que fueron recomendados (Seleccionados) al usuario, para el caso de la técnica se recomiendan todos aquellos que superen un umbral determinado. A su vez, en el ámbito de la implementación de la técnica, el umbral de predicción que debe superar un ítem para ser recomendado fue definido en 3.5.

Tabla 3. Datos de entrada para el cómputo de la Precisión y el Recall.

	Seleccionados	No Seleccionados	Total
Relevantes	$N_{rs}$	$N_{rn}$	$N_r$
Irrelevantes	$N_{is}$	$N_{in}$	$N_i$
Total	$N_s$	$N_n$	$N$

Luego, la precisión de las recomendaciones efectuadas a un usuario determinado, que representa la probabilidad de que un ítem recomendado sea relevante para el usuario, se obtiene como la división de la cantidad de ítems que fueron recomendados y son relevantes entre el total de ítems que fueron recomendados (Herlocker, Konstan et al. 2004):

$$\text{Fórmula 22: } P = \frac{N_{rs}}{N_s}$$

Por otro lado el Recall representa la probabilidad de que un ítem relevante sea recomendado. Esta métrica se calcula como la división de la cantidad de ítems que fueron recomendados y son relevantes entre el total de ítems que son relevantes para el usuario (Herlocker, Konstan et al. 2004):

$$\text{Fórmula 23: } R = \frac{N_{rs}}{N_r}$$

Aunque lo normal para un sistema de recomendación es predecir la valoración que le daría un usuario a cualquier ítem que este no haya valorado, a priori se puede observar que, para este tipo de prueba, se deben generar las predicciones de valoración para el conjunto de ítems valorados por cada usuario. Luego del cómputo de los valores de Precisión y el Recall para cada usuario, se obtuvo la media de estas métricas para el sistema usando solo el método de recomendación Ítem – Ítem por regresión, que fue implementado en el Motor de Recomendación, los resultados se muestran en la Tabla 4.

Tabla 4. Resultado medio de Precisión y Recall del Motor de Recomendación.

	k = 5	k = 10	k = 15
Precisión	0.8012	0.8194	0.8230
Recall	0.6043	0.6269	0.6305

Luego se procedió a efectuar los mismos cálculos pero para cuando se utiliza la técnica como un todo, donde la diferencia en la recomendación la marca la utilización del componente Selector de Recomendación perteneciente a la técnica propuesta. Los resultados arrojados se muestran en la Tabla 5.

Tabla 5. Resultado medio de Precisión y Recall del sistema implementado.

	k = 5	k = 10	k = 15
Precisión	0.7512	0.7694	0.7733
Recall	0.5043	0.5369	0.5505

Las figuras 7 y 8 muestran sendas gráficas comparativas para las métricas Precisión y Recall respectivamente del motor de recomendación y del sistema implementado como un todo.

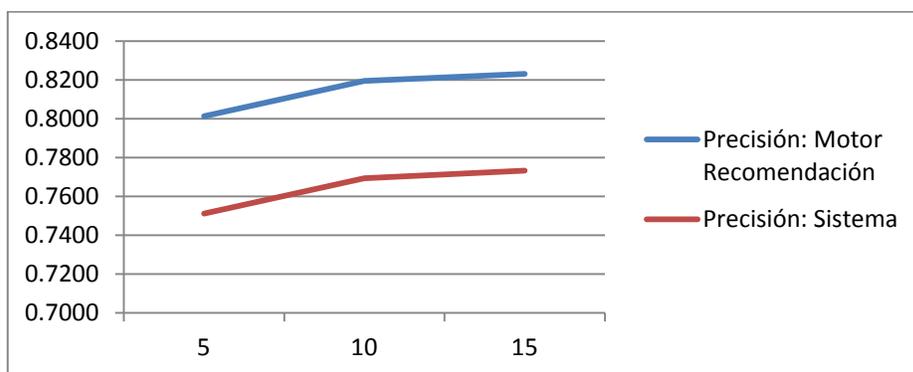


Figura 7. Comparación de la Precisión entre motor de recomendación y el sistema implementado.

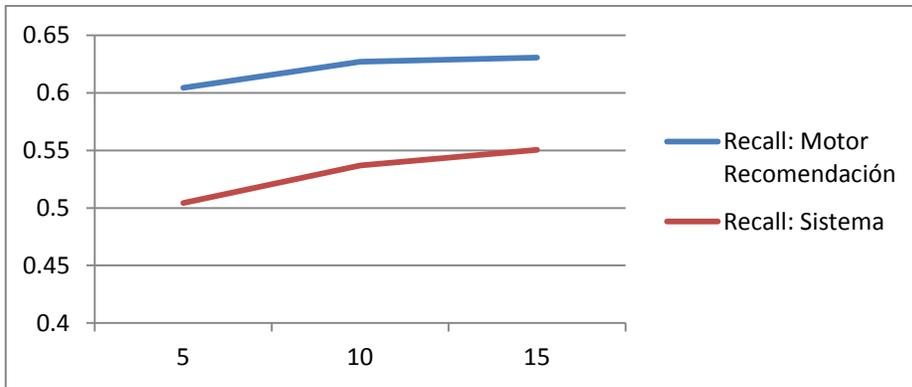


Figura 8. Comparación del Recall entre motor de recomendación y el sistema implementado.

Como se puede apreciar, los resultados arrojados para la Precisión aumentan según aumenta el valor de  $k$  y para todos los casos. La probabilidad de que un ítem recomendado por el sistema sea relevante para el usuario, supera el 80 % para las 3 variantes probadas en el caso del Motor de Recomendación en solitario y se mantiene sobre el 75 % cuando se utiliza el sistema como un todo. En el caso del Recall, la probabilidad de que un ítem relevante sea recomendado, también crece en relación directa con el valor de  $k$  y esta probabilidad se mantiene por encima del 60 % cuando se evalúa al Motor de Recomendación solamente y sobre el 50 % para la evaluación del sistema. Estos resultados se pueden catalogar de muy buenos para el sistema de recomendación obtenido pues permite afirmar que las recomendaciones obtenidas son probablemente útiles al usuario en cuestión. La pérdida que se observa en la Precisión y el Recall cuando se evalúa al sistema como un todo es completamente previsible a partir de completamiento de la vecindad propuesto en el Capítulo 2.

### 3.2 Exactitud de la Predicción

La exactitud de la predicción es, con amplia diferencia, la propiedad más discutida en la literatura asociada a los sistemas de recomendación (Shani and Gunawardana 2011). Al medir la eficacia de la predicción se está midiendo cuán acertada es una recomendación dada.

#### MAE

La técnica MAE mide la desviación promedio absoluta entre una valoración predicha y la valoración real dada por el usuario. Esta métrica de evaluación constituye una manera muy popular de evaluar el rendimiento de los sistemas de recomendación (Herlocker, Konstan et al. 2004):

$$\text{Fórmula 24: } |\bar{E}| = \frac{\sum_{i=1}^N |p_i - r_i|}{N}$$

Donde  $N$  representa la cantidad de ítems evaluados por un usuario  $u$  y donde  $p_i$  y  $r_i$  representan la predicción de valoración y la valoración otorgada realmente al ítem  $i$  por el usuario  $u$  respectivamente.

Al igual que para la Precisión y el Recall, para computar el valor de esta métrica se debe predecir la valoración para los ítems que el usuario ha valorado. En la Tabla 6 se muestra el error medio que se obtiene en las recomendaciones, para distintos valores de  $k$  en la búsqueda de los  $k$ -NN, del Motor de Recomendación.

Tabla 6. Error medio absoluto del Motor de Recomendación.

	k = 5	k = 10	k = 15
MAE	0.8193	0.7514	0.7384

Luego se realizó el cómputo del MAE para el sistema implementado. En la Tabla 7 se muestran los resultados arrojados como promedio.

Tabla 7. Error medio absoluto del sistema implementado.

	k = 5	k = 10	k = 15
MAE	0.8256	0.7748	0.7399

En la figura 9 se muestra la comparación resultante de la evaluación representada en las tablas 4 y 5.

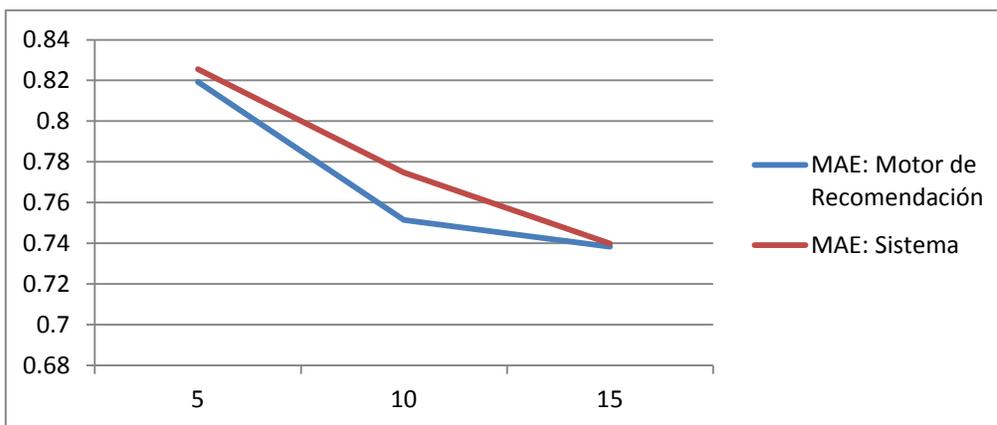


Figura 9. Comparación de la métrica MAE entre el Motor de Recomendación y el sistema implementado.

Como se puede observar para el resultado de la predicción obtenida tendrá una desviación media menor de una unidad en todos los casos lo que se puede considerar como un desempeño bueno. Las predicciones, entonces, estarán como media lo suficientemente cercanas a la realidad como para que correspondan con la valoración que, presumiblemente, les daría el usuario.

Por otro lado, se puede observar que a mayor valor de  $k$  se obtiene un menor error absoluto, esto viene determinado porque se utilizan más vecinos de un ítem cualquiera para predecir su valoración. Esto no significa necesariamente que un mayor valor de  $k$  sea recomendable, pues habría que observar su incidencia en otros indicadores. La diferencia resultante entre el Motor de Recomendación y el sistema como un todo es lo suficientemente pequeña como para considerarse despreciable.

### COBERTURA

La cobertura de un sistema de recomendación es una medida del dominio de ítems en el sistema sobre el cual este puede obtener predicciones o realizar recomendaciones. Los sistemas con baja cobertura pueden resultar poco útiles para los usuarios ya que estarían limitados en las decisiones en la que son capaces de proveer ayuda. La cobertura puede ser definida como el por ciento de ítems para los cuales se puede obtener una predicción (Herlocker, Konstan et al. 2004).

Para la implementación del sistema de recomendación obtenido mediante la aplicación de la técnica se definió que, de manera aleatoria, los usuarios realizaran de 0 a 300 valoraciones por cada ítem. Para el cálculo de la cobertura, el sistema debería ser capaz de generar un total de predicciones igual a la cantidad de ítems menos la cantidad de valoraciones que haya realizado cada usuario  $u$  que pertenece al conjunto de dominio  $U_n$  de usuarios que posee el sistema. Por lo que el por ciento de cobertura total del sistema vendría dado por:

$$\text{Formula 25: } PCT = \frac{\sum_{u \in U_n} \text{cantPredicciones}_u}{\sum_{u \in U_n} \text{cantItem} - \text{cantVal}_u} * 100$$

Durante el proceso de generación aleatoria de datos se generaron un total de 545.934 valoraciones lo que dejó un total de 3.454.066 posibles valoraciones por predecir. En la Tabla 8 se muestran los resultados que, para el por ciento de cobertura, arrojó el Motor de Recomendación.

Tabla 8. Por ciento de cobertura del espacio de predicción para el Motor de Recomendación.

	k = 5	k = 10	k = 15
% de Cobertura.	87.69 %	81.85 %	58.62 %

También se calculó la cobertura sobre el espacio de predicción para cuando se utiliza el sistema implementado como un todo. Los resultados se muestran en la tabla 9.

Tabla 9. Por ciento de cobertura del espacio de predicción para el sistema implementado.

	k = 5	k = 10	k = 15
% de Cobertura.	97.22 %	93.67 %	75.62 %

La figura 10 muestra la comparación entre los resultados plasmados en las tablas 6 y 7.

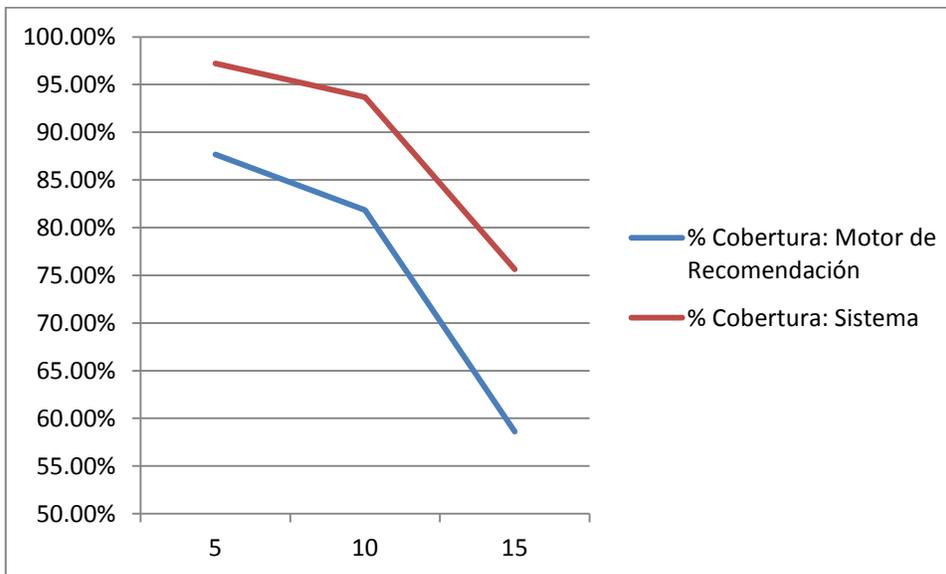


Figura 10. Comparación de la Cobertura entre el Motor de Recomendación y el sistema implementado.

Los resultados para esta métrica muestran una dramática disminución del porcentaje de cobertura para el espacio de predicción cuando  $k$  tiene por valor 15. Esto ocurre porque a medida que se incrementa el valor de  $k$  al sistema se le hace más difícil encontrar el número de ítems cercanos que hayan sido valorados por el usuario activo. En el caso de que el sistema no pueda encontrar la cantidad de vecinos cercanos definida por  $k$  no puede calcular una predicción y es en este caso cuando se realiza un completamiento de

la vecindad. Es por esto que se debe juzgar las ventajas y desventajas (error, precisión, recall versus cobertura) de fijar la *ken* un valor determinado. Es importante resaltar que el porcentaje que, para la cobertura, obtuvo el sistema implementado como un todo presenta un aumento significativo respecto al obtenido por el componente Motor de Recomendación por sí solo. Obsérvese también que para el caso particular del sistema implementado se obtiene una muy buena cobertura para  $k = 5$  y  $k = 10$ , por lo que sería recomendable utilizar 10 como el número de vecinos cercanos a encontrar para, de esta forma, obtener mejores resultados de precisión, recall y MAE mientras que se mantiene un buen porcentaje de cobertura.

### **3.2 Conclusiones Parciales**

Para la aplicación de la técnica propuesta se debe tener en cuenta, esencialmente, el contexto en el que el sistema va a ser implantado. En el presente capítulo se demostró, que usando la solución propuesta para obtener en un sistema de recomendación en el contexto adecuado se pueden obtener buenos resultados en las recomendaciones.

Los resultados arrojados por la implementación del componente Motor de Recomendación pueden ser catalogados de buenos, pues 8 de cada 10 elementos recomendados resultan relevantes para el usuario.

La utilización del componente Selector de Recomendación resulta positiva dentro del sistema implementado porque, aunque se evidencia una ligera disminución en la precisión de las recomendaciones, el sistema es capaz de incrementar su cobertura del espacio de predicción significativamente, lo que significa que se puede obtener una predicción de valoración para un número mayor de ítems que el componente Motor de Recomendación por sí solo. Esto se traduce en que se puede recomendar una mayor variedad de ítems presentes en el dominio.

Finalmente, los resultados obtenidos en la medición de las métricas tanto del uso como de la exactitud de la predicción permiten afirmar que, mediante la aplicación de la técnica propuesta, se puede obtener un sistema capaz de realizar recomendaciones eficaces de productos turísticos cubanos.

## CONCLUSIONES GENERALES

La importancia que, para Cuba, tiene la industria del turismo, hace de la personalización masiva de los contenidos que muestran las organizaciones de e-turismo cubanas una tarea necesaria.

Para el contexto turístico cubano es recomendable utilizar un enfoque de recomendación colaborativo Ítem – Ítem para realizar la tarea de decidir el contenido que debe ser mostrado prominentemente.

La explotación de las características contextuales derivadas de la utilización de un GIS permite incidir positivamente en la cobertura del sistema sobre el espacio de predicción sin hacer grandes sacrificios en la precisión con que se calculan las predicciones de valoración.

La técnica propuesta define las características y el comportamiento que debe ser implementado para obtener recomendaciones de productos turísticos cubanos a partir de las preferencias y la información contextual de cada usuario.

Las pruebas realizadas demuestran que para el contexto turístico descrito se obtienen resultados positivos en la exactitud y el uso de las predicciones de valoración calculadas mientras se incide positivamente en la cobertura del sistema implementado, lo que permite afirmar que la aplicación de la técnica se logra obtener de manera eficaz el contenido que debe ser mostrado prominentemente a los usuarios de las organizaciones de e-turismo cubanas.

## RECOMENDACIONES

- Desplegar la solución implementada para los experimentos o alguna otra desarrollada utilizando la técnica propuesta, en una organización de e-turismo con datos reales para, luego, medir el impacto de la técnica en la eficiencia de la organización y para ratificar los resultados obtenidos en las pruebas realizadas en un entorno experimental.
- Explorar vías de hibridación del enfoque de recomendación propuesto en aras de explotar las ventajas de otros enfoques y mitigar aun más las del enfoque colaborativo.
- Explorar nuevas vías de explotación de la información contextual según la empresa cubana de telecomunicaciones ETECSA vaya ampliando el espectro de servicios disponibles para la telefonía celular en Cuba para incidir positivamente en la precisión de las recomendaciones.

## REFERENCIAS BIBLIOGRÁFICAS

Adomavicius, G., R. Sankaranarayanan, et al. (2005). "Incorporating contextual information in recommender systems using a multidimensional approach." ACM Transactions on Information Systems (TOIS) **23**(1): 103-145.

Adomavicius, G. and A. Tuzhilin (2005). "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions." Knowledge and Data Engineering, IEEE Transactions on **17**(6): 734-749.

Arazy, O., N. Kumar, et al. (2009). "Improving Social Recommender Systems." IT professional **11**(4): 38-44.

Bell, R., Y. Koren, et al. (2007). Modeling relationships at multiple scales to improve accuracy of large recommender systems. Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM.

Ben-Shimon, D., A. Tsikinovsky, et al. (2007). Recommender system from personal social networks. Advances in Intelligent Web Mastering, Springer: 47-55.

Berka, T. and M. Plößnig (2004). "Designing recommender systems for tourism." Proceedings of ENTER 2004.

Breese, J. S., D. Heckerman, et al. (1998). Empirical analysis of predictive algorithms for collaborative filtering. Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence, Morgan Kaufmann Publishers Inc.

Buhalis, D. (2003). eTourism: Information technology for strategic tourism management, Pearson Education.

Burke, R. (2007). Hybrid web recommender systems. The adaptive web, Springer: 377-408.

Cleverdon, C. W., J. Mills, et al. (1966). "Factors determining the performance of indexing systems,(Volume 1: Design)." Cranfield: College of Aeronautics.

Charou, E., K. Kabassi, et al. (2010). Integrating Multimedia GIS Technologies in a Recommendation System for Geotourism. Multimedia Services in Intelligent Environments. G. Tsihrintzis and L. Jain, Springer Berlin Heidelberg. **3**: 63-74.

DeMers, M. N. (2008). Fundamentals of geographic information systems, John Wiley & Sons.

Deshpande, M. and G. Karypis (2004). "Item-based top-n recommendation algorithms." ACM Transactions on Information Systems (TOIS) **22**(1): 143-177.

Desrosiers, C. and G. Karypis (2011). A comprehensive survey of neighborhood-based recommendation methods. Recommender systems handbook, Springer: 107-144.

Gao, M., K. Liu, et al. (2010). "Personalisation in web computing and informatics: Theories, techniques, applications, and future research." Information Systems Frontiers **12**(5): 607-629.

Goldberg, D., D. Nichols, et al. (1992). "Using collaborative filtering to weave an information tapestry." Communications of the ACM **35**(12): 61-70.

Herlocker, J. L., J. A. Konstan, et al. (1999). An algorithmic framework for performing collaborative filtering. Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, ACM.

Herlocker, J. L., J. A. Konstan, et al. (2004). "Evaluating collaborative filtering recommender systems." ACM Transactions on Information Systems (TOIS) **22**(1): 5-53.

Hill, W., L. Stead, et al. (1995). Recommending and evaluating choices in a virtual community of use. Proceedings of the SIGCHI conference on Human factors in computing systems, ACM Press/Addison-Wesley Publishing Co.

Howe, A. E. and R. D. Forbes (2008). Re-considering neighborhood-based collaborative filtering parameters in the context of new data. Proceedings of the 17th ACM conference on Information and knowledge management, ACM.

Jannach, D., M. Zanker, et al. (2007). "Developing a conversational travel advisor with ADVISOR SUITE." Information and Communication Technologies in Tourism 2007: 43-52.

Konstan, J. A., B. N. Miller, et al. (1997). "GroupLens: applying collaborative filtering to Usenet news." Communications of the ACM **40**(3): 77-87.

Lang, K. (1995). Newsweeder: Learning to filter netnews. In Proceedings of the Twelfth International Conference on Machine Learning, Citeseer.

Mahmood, T. and F. Ricci (2009). Improving recommender systems with adaptive conversational strategies. Proceedings of the 20th ACM conference on Hypertext and hypermedia, ACM.

Maltz, D. and K. Ehrlich (1995). Pointing the way: active collaborative filtering. Proceedings of the SIGCHI conference on Human factors in computing systems, ACM Press/Addison-Wesley Publishing Co.

Manouselis, N., R. Vuorikari, et al. (2010). "Collaborative recommendation of e-learning resources: an experimental investigation." Journal of Computer Assisted Learning **26**(4): 227-242.

Massa, P. and B. Bhattacharjee (2004). Using trust in recommender systems: an experimental analysis. Trust Management, Springer: 221-235.

MINTUR (2002). "Ministerio de Turismo." Retrieved 15 de marzo, 2014, from [http://www.cubagob.cu/des\\_eco/turismo.htm](http://www.cubagob.cu/des_eco/turismo.htm).

Redpath, J., D. H. Glass, et al. (2010). Collaborative filtering: the aim of recommender systems and the significance of user ratings. Advances in Information Retrieval, Springer: 394-406.

Resnick, P. and H. R. Varian (1997). "Recommender systems." Communications of the ACM **40**(3): 56-58.

Ricci, F. (2002). "Travel recommender systems." IEEE Intelligent Systems **17**(6): 55-57.

Said, A. (2010). Identifying and utilizing contextual data in hybrid recommender systems. Proceedings of the fourth ACM conference on Recommender systems, ACM.

Salinas, E. and L. Mundet (2000). "El turismo en Cuba. Un análisis geográfico." El Turismo en Cuba, Geographicalia: 53-66.

Sarwar, B., G. Karypis, et al. (2000). Analysis of recommendation algorithms for e-commerce. Proceedings of the 2nd ACM conference on Electronic commerce, ACM.

Shani, G. and A. Gunawardana (2011). Evaluating recommendation systems. Recommender systems handbook, Springer: 257-297.

Shardanand, U. and P. Maes (1995). Social information filtering: algorithms for automating "word of mouth". Proceedings of the SIGCHI conference on Human factors in computing systems, ACM Press/Addison-Wesley Publishing Co.

Sweet, P. L. (2001). Designing Interactive Value Development: Perspectives and Strategies for High Precision Marketing, Lund University.

Trujillo, M., M. Millan, et al. (2007). A recommender system based on multi-features. Computational Science and Its Applications–ICCSA 2007, Springer: 370-382.

Werthner, H. (2003). Intelligent systems in travel and tourism. IJCAI.