



SISTEMA DE RECOMENDACIÓN Y PLANIFICACIÓN TURÍSTICA DE LA CIUDAD DE VALENCIA VÍA WEB

Autor: Cesar Augusto Guzmán Álvarez

Directores: Dra. Inmaculada García

Dra. Laura Sebastia Tarín

Departamento de Sistemas Informáticos y Computación

UNIVERSIDAD POLITÉCNICA DE VALENCIA

Valencia, España

ENERO 2009

A mis padres.

Tabla de Contenidos

Tabla de Contenidos	v
Lista de Tablas	vii
Lista de Figuras	viii
Agradecimientos	x
Resumen	xiii
Abstract	xv
Resum	xvii
1. Introducción	1
1.1. Introducción	1
1.2. Motivación y objetivos	3
1.3. Organización de la memoria	5
2. Estado del Arte	7
2.1. Antecedentes	7
2.2. Sistemas Recomendadores	8
2.2.1. Ventajas e Inconvenientes de los Sistemas Recomendadores	10
2.2.2. Definición Formal	10
2.3. Perfil del Usuario	11
2.3.1. APML	12
2.4. Técnicas de Recomendación	13
2.4.1. Técnicas de Recomendación Básicas	13
2.4.2. Problemas de las Técnicas de Recomendación	15
2.4.3. Técnicas de Recomendación Híbridas	19
2.5. Refinamiento de Resultados: Sistemas de Recomendación Conversacionales	22

2.6. Sistemas de recomendación de turismo y ocio	25
3. E-Turismo: SR de rutas turísticas planificadas en la ciudad de Valencia	27
3.1. Introducción	27
3.2. E-Turismo	28
3.3. Arquitectura de e-Turismo	31
3.3.1. Representación del conocimiento en e-Turismo	35
3.3.2. Núcleo Generalista del Sistema Recomendador (GRSK)	40
3.3.3. Núcleo Planificador (PK)	49
3.3.4. Núcleo Gestor de la Base de Datos (DBMK)	54
3.4. Conclusiones	56
4. Evaluación y Resultados	59
4.1. Introducción	59
4.2. Obtención de datos de prueba	61
4.2.1. Cuestionarios	62
4.2.2. Carga de información	63
4.3. Resultados	65
4.4. Conclusiones	68
5. Conclusiones	71
5.1. Introducción	71
5.2. Contribuciones de este trabajo	71
5.3. Beneficios de este trabajo	72
5.4. Futuros desarrollos	73
5.5. Publicaciones y artículos relacionados	73
Bibliografía	75

Índice de tablas

2.1. Ejemplos de SR que utilizan cada técnica de recomendación.	16
2.2. Técnicas de recomendación [16].	17
2.3. Comparación de las estrategias de feedback [53].	24
3.1. Actividades en nuestro problema de planificación.	52

Índice de figuras

2.1. Técnicas de recomendación [17]	18
2.2. Recomendador híbrido ponderado.	20
2.3. Recomendador híbrido mezclado.	20
2.4. Recomendador híbrido alternado.	21
2.5. Recomendador híbrido en cascada.	21
2.6. Recomendador híbrido de combinación de características.	21
2.7. Recomendador híbrido de aumento de características.	22
2.8. Recomendador híbrido meta-level.	22
3.1. Formulario principal de e-Turismo.	28
3.2. Preferencias específicas.	29
3.3. Proceso de puntuaciones.	30
3.4. Interfaz e-Turismo.	31
3.5. Obtener perfil del usuario.	32
3.6. Perfil del usuario en archivo XML.	32
3.7. Sistema e-Turismo.	33
3.8. Lista de actividades recomendadas al usuario.	34
3.9. Lista de actividades finales.	35
3.10. Lista de actividades confirmadas.	36
3.11. Taxonomía de e-Turismo.	37
3.12. Horario de actividad.	38
3.13. Relaciones del GRSK	42
3.14. Arquitectura del GRSK	43
3.15. Agenda turística.	54

3.16. Núcleo Gestor de la Base de datos.	55
4.1. Datos personales.	62
4.2. Visitas previas realizadas.	63
4.3. Frecuencia de actividades.	63
4.4. Preferencias generales.	64
4.5. Escenario de pruebas.	65
4.6. Resultado de la evaluación aplicada a un máximo de 25 recomen- daciones.	66
4.7. Comparación de resultados con un máximo de 25 recomendaciones sin feedback.	67
4.8. Comparación de resultados con un máximo de 25 recomendaciones con feedback.	68

Agradecimientos

A mis directores de Tesis, la Dra. Inmaculada Garcia Garcia y la Dra. Laura Sebastia Tarín, y a la Dra. Eva Onaindía de la Rivaherrera por haberme proporcionado la oportunidad de ser su alumno de tesis, por leer pacientemente cada capítulo de este trabajo, y por aportar incansablemente mejoras al mismo.

Mil gracias a mis padres Esperanza y Carlos y a mis hermanas Carlos y Andrés, por su confianza, apoyo y amor incondicional. Son mi referencia y ejemplo, son mi motivo de superación.

A mi pareja y compañera sentimental Mayra, por convertirse en un motivo de lucha y superación para seguir adelante cada día de la vida.

Valencia, España
Enero, 2008

Cesar Augusto Guzmán Álvarez

Agradecimientos a otras instituciones o grupo.

Esperamos que los resultados prácticos de esta Tesis de Máster sean de utilidad y permitan llegar a establecer un estándar en los Sistemas Recomendadores enfocados en el turismo no solo de la Comunidad Valenciana sino también para cualquier Ciudad o País. Sería una muestra más de la necesaria relación Universidad-Sector Turístico y un factor de éxito y relevancia de los trabajos aquí desarrollados.

Resumen

Los problemas de recomendación son complejos y variados ya que exigen un elevado conocimiento de los gustos y preferencias del usuario para proporcionar recomendaciones que le resulten satisfactorias. Igualmente son considerados de gran interés tanto en el aspecto científico como aplicado. Por ello, el desarrollo de técnicas de recomendación eficientes y adaptables en diferentes dominios de aplicación es el objetivo de muchos trabajos de investigación.

Entre los problemas más típicos se encuentran la recomendación de libros o música y también la recomendación de lugares turísticos, aunque la mayor parte del trabajo en este campo se centra en la recomendación de paquetes turísticos. En concreto el trabajo a consistido en la creación de un sistema de recomendación y planificación de rutas turísticas en una ciudad.

Para el desarrollo de este sistema, una vez planteado el problema, se analizaron las técnicas de recomendación en diferentes publicaciones que se relacionan con el tipo de dominio estudiado.

El presente trabajo consiste en la creación de un Sistema Recomendador que utiliza las técnicas de recomendación básicas más apropiadas al dominio y las une utilizando una técnica de recomendación híbrida mezclada. Detalla también un núcleo planificador para la organización de las recomendaciones, y la elaboración de una agenda turística teniendo en cuenta las restricciones del entorno. Además se incluyen pequeños módulos que estudian y organizan la información de los usuarios y un módulo independiente que interactúa con la base de datos. Igualmente se independiza el Sistema Recomendador de la interfaz, haciendo posible su utilización en cualquier interfaz o aplicación de dominio.

Con el objeto de evaluar el sistema, se evaluaron diversas situaciones del problema, todas de casos reales. Se seleccionaron diferentes tipos de usuario, cada uno con preferencias y gustos diferentes que permiten deducir la eficacia del Sistema Recomendador implementado.

El proyecto propone también futuros trabajos de ampliación. Como por ejemplo,

ampliaciones relacionadas con recomendaciones para grupos y su adaptación para la integración con agentes.

El trabajo destaca aportaciones no sólo respecto a los servicios turísticos online, sino también respecto a cualquier servicio recomendador, que suponen contribuciones tangibles en el campo de los sistemas recomendadores.

Abstract

Make and evaluate recommendations for a particular application domain is considered as a complex and varied problem because it requires much knowledge about the user's tastes and preferences. The problem of providing satisfactory recommendations is also very interesting from the scientific and applied point of view. For these reasons, the development of efficient and adaptable recommendation techniques for different domains is the purpose of many research works.

Among the most typical problems, we can find recommendation systems for books or music, and recommendations for tourism too. This work is focused on the latter domain, and, particularly, it introduces a recommendation and a planning system for computing tourist routes in a city.

In order to develop this system, once the problem was analyzed, some of the basic recommendation techniques used in the tourist domain were studied. In this work we propose a module that analyzes and combines some of these basic techniques, thereby obtaining a general hybrid technique. It also details the main characteristics of a planning kernel which is in charge of organizing the obtained recommendations. The planner computes a tourist schedule taking into account some restrictions from the environment. The system contains some small modules to study and organize the user's information, and an independent module to interact with the database. Moreover, the recommender system is independent from the interface, making it possible to use with other interfaces.

In order to evaluate the system, we designed different problem situations, taking information from real users. We selected different types of users, with different preferences and tastes, to check the effectiveness of the recommender system.

The project also proposes future research works as the use of multi-agent systems to perform recommendation for groups of people.

It is important to highlight that this recommender system can be used not only for an online tourist service but also for any other application domain that can be defined through a taxonomy.

Resum

Els problemes de recomanació són complexos i variats ja que exigeixen un gran coneixement dels gustos i preferències de l'usuari per a proporcionar recomanacions que resulten satisfactòries. Igualment són considerats de gran interès en l'aspecte tant científic com aplicat. Per això, el desenvolupament de tècniques de recomanació eficients i adaptables en diferents dominis d'aplicació és l'objectiu de molts treballs d'investigació.

Entre els problemes més típics es troben la recomanació de llibres o música i també la recomanació de llocs turístics. En aquest últim domini està centrat aquest treball, concretament s'aporta un sistema de recomanació i planificació de rutes turístiques en una ciutat.

Per al desenvolupament d'aquest sistema, una vegada plantejat el problema, s'analitzaren les tècniques bàsiques de recomanació en diferents publicacions que es relacionen amb el tipus de domini estudiat.

El present treball proposa un mòdul que considera diverses d'aquestes tècniques bàsiques mitjançant una tècnica híbrida general. Detalla també un nucli planificador per a l'organització de les recomanacions, i l'elaboració d'una agenda turística tenint en compte les restriccions de l'entorn. A més s'inclouen xicotets mòduls que estudien i organitzen la informació dels usuaris i un mòdul independent que interactua amb la base de dades. Igualment s'independitza el sistema recomanador de la interfície, fent possible el seu acoplament amb qualsevol interfície o aplicació de domini.

Amb l'objectiu d'avaluar el sistema, es computaren i avaluaren diverses situacions del problema, totes de casos reals. Es seleccionaren diferents tipus d'usuaris, cadascun amb preferències i gustos diferents que permeten deduir l'eficàcia del sistema recomanador implementat.

El projecte proposa també futurs treballs d'ampliació, com per exemple, ampliacions relacionades amb recomanacions per a grups i la seua adaptació per a la integració d'agents.

El treball destaca aportacions no sols respecte als serveis turístics online, sinó també davant qualsevol servei recomanador, que suposen contribucions tangibles al camp dels sistemes recomanadors.

Capítulo 1

Introducción

1.1. Introducción

El sector turístico se ha convertido, progresivamente, en una de las principales actividades generadoras de riqueza en la Comunidad Valenciana (CV), no sólo por su contribución directa a la misma, sino por el efecto expansivo que ha ejercido sobre un gran conjunto de actividades afines. Se estima que el turismo representa una participación del 11 % del Producto Interior Bruto (PIB) de la Comunidad Valenciana, que se ha convertido, con más de 22 millones de turistas al año, en la segunda región española receptora de turismo interior y en la quinta de turismo procedente del exterior [21]. La ciudad de Valencia es uno de los principales destinos turísticos dentro de la CV, y recibió en el 2007 alrededor de 2 millones de viajeros, creciendo un 18 % [21]. El desarrollo del turismo en esta ciudad durante la última década ha sido mayor que en cualquier otro lugar en España, siendo Valencia en estos momentos uno de los ocho destinos españoles más visitados. En la actualidad, la atracción más popular es la Ciudad de las Artes y las Ciencias, uno de los complejos culturales más importantes de Europa, que atrae cada año a 3.5 millones de visitantes. Además, eventos como la Copa América, que se celebró en 2007, atrajeron a 1 millón de turistas nacionales y extranjeros (desde 2004). Durante la regata final de la Copa América, el 68.4 % del total de turistas eran extranjeros. En 2008 se celebró la primera competición en el circuito de Fórmula 1 en la ciudad de Valencia. Se estima que el evento congregó unas 120.000 personas, según una investigación que recoge un estudio realizado por el Instituto Valenciano de Investigaciones Económicas (IVIE).

Cada vez y con más frecuencia las personas se apoyan en los avances tecnológicos en el momento de realizar un viaje. Esto se debe, fundamentalmente, al uso cotidiano que se hace hoy por hoy de la tecnología como Internet. Existen numerosas instituciones y empresas que ofrecen diversa información turística acerca de la Comunidad Valenciana como, por ejemplo: La Agencia Valenciana de Turismo [88], Turismo Valencia Convention Bureau [87], Lanetro [46] (se puede consultar el callejero de la ciudad, información sobre eventos, restaurantes, hoteles, etc), Páginas Amarillas [4] (ofrece además del callejero y otra información útil, un servicio de consulta sobre restaurantes y hoteles vía teléfono móvil, entre otros).

La información turística que ofrecen estos servicios se caracteriza principalmente por los siguientes aspectos:

- Se muestra una información muy general y estática, es decir, se ofrece sistemáticamente la misma información a todos los usuarios, independientemente del perfil de éstos.
- Se muestran grandes volúmenes de información que el usuario debe procesar posteriormente para seleccionar aquellos elementos que pueden ser de su interés.

Generalmente, el usuario o viajero parte de un desconocimiento del lugar a visitar y de los sitios que pueden ser de interés por su valor artístico, social o lúdico. Además, cada persona puede “ adoptar ” un perfil distinto cuando viaja, de manera que se pueden distinguir distintos perfiles como, por ejemplo, el viajero cultural, el viajero gastronómico, el viajero en familia, etc. Se puede disponer de una gran cantidad de datos e información sobre un lugar determinado, pero el usuario querrá obtener información principalmente de aquellos elementos que corresponden a su perfil o interés particular.

Cada uno de estos perfiles determina los distintos lugares a visitar o los distintos modos de planificar una ruta. Por ejemplo, el viajero gastronómico situará sus preferencias culinarias en primer lugar, es decir, el restaurante en el que desea comer, para luego organizar los lugares o monumentos a visitar alrededor del mismo. Por otro lado, el usuario que viaja con niños evitará visitar museos durante un largo tiempo e incluirá en su viaje la visita a algún jardín o parque de atracciones.

Este proyecto va dirigido a dos tipos de público:

1. Turistas nacionales o extranjeros que deseen conocer la ciudad de Valencia, sobre todo si desean conocer aspectos de la ciudad que no son los que habitualmente se promocionan.
2. Ciudadanos de Valencia o de los alrededores que deseen realizar visitas a la ciudad y quieren ser aconsejados sobre lugares a visitar o descubrir lugares que se adapten a sus preferencias.

En definitiva, el objetivo del proyecto de Máster es desarrollar un sistema de recomendación y planificación turística de la ciudad de Valencia que brinde una guía de actividades personalizadas, permitiendo a los usuarios obtener datos de lugares a visitar como si se los recomendase una persona que conoce bien al usuario y al lugar a visitar (similar al boca a boca).

Así, se desarrollará una herramienta personalizada a través de un Sistema de Información Web que recomiende rutas turísticas o lugares a visitar en la ciudad de Valencia adaptadas a las preferencias propias de cada usuario. Se dispondrá de una página web que recogerá información sobre el usuario, el sistema la analizará y determinará su perfil, de modo que se le proporcione una visita adecuada al mismo.

1.2. Motivación y objetivos

Como se ha introducido en el apartado anterior, los sistemas de recomendación cada vez son más utilizados en multitud de dominios. Por ello, son tema de investigación de diversos proyectos en el área de la Inteligencia Artificial. Concretamente, tal y como se ampliará en el capítulo 2, existen varios proyectos centrados en el dominio de turismo. Sin embargo, la mayoría de estos trabajos se centran en la recomendación de destinos de viajes.

El problema que se plantea en este trabajo es de recomendación de rutas turísticas dentro de la ciudad de Valencia adaptadas al perfil del usuario. Ofrecer una recomendación turística en un Servicio Web de las actividades que se pueden realizar en una determinada ciudad responde a una necesidad real, ya que la oferta de dichas actividades cada vez es más amplia y resulta más difícil visitar todo lo que nos puede interesar. Además, la planificación de las actividades escogidas, exige tener en cuenta diversos aspectos como los horarios, las distancias, etc.

Así, este problema plantea dos aspectos bien diferenciados:

1. Obtener actividades más satisfactorias para el usuario, es decir. las que mejor se adaptan a su perfil.
2. Planificar las actividades obtenidas de acuerdo a ciertas restricciones como los horarios de apertura y cierre, la distancia geográfica entre las actividades, el tiempo disponible del usuario, etc.

El objetivo principal es desarrollar un Sistema de Información Web personalizado para recomendar rutas turísticas o lugares a visitar en la ciudad de Valencia adaptadas a las preferencias propias de cada usuario. Para conseguir nuestro objetivo principal se desarrollaron los siguientes objetivos específicos:

- Definir formalmente el problema de recomendación de lugares turísticos, considerando la taxonomía, preferencias y restricciones que subyacen en el dominio de turismo, así como las características principales de cada uno de los lugares.
- Analizar las técnicas de recomendación turísticas y de planificación, revisando su adecuación y / o aplicabilidad al problema.
- Establecer las ventajas e inconvenientes de los Sistemas Recomendadores existentes, analizando su viabilidad y adecuación al problema planteado, permitiendo justificar la necesidad de mejorar y desarrollar nuevas técnicas más adecuadas para la resolución de dicho problema.
- Desarrollar un Sistema Recomendador que combine las técnicas básicas de recomendación mas apropiadas para la resolución del problema de recomendación de visitas turísticas. Se diseñará este sistema para que sea lo más independiente posible del dominio de recomendación y, al mismo tiempo, permita fácilmente incluir nuevas técnicas de recomendación.
- Identificar y definir el problema de planificación de la agenda turística en el que se incorporan nuevas restricciones a las contempladas en los problemas genéricos de planificación, como por ejemplo, horarios de apertura y cierre, limitación del tiempo disponible, etc.

- Evaluar el Sistema Recomendador desarrollado, diseñando casos de prueba con instancias a priori del problema a través de varias encuestas realizadas a diferentes usuarios.

Es importante comentar que este trabajo se centra, fundamentalmente, en el aspecto de la recomendación.

1.3. Organización de la memoria

Este trabajo se organiza como sigue:

- Capítulo 2. Estado del arte. Este capítulo resume el estado del arte de los sistemas de recomendación y cuáles son las técnicas básicas empleadas. Asimismo, hace un barrido de los sistemas de recomendación centrados en el dominio de turismo.
- Capítulo 3. e-Turismo: SR de rutas turísticas planificadas en la ciudad de Valencia. Este capítulo presenta e-Turismo, un sistema de recomendación de rutas turísticas para la ciudad de Valencia. Así, se detalla la funcionalidad e importancia de e-Turismo para el usuario, la representación del conocimiento y de cada uno de los módulos que componen e-Turismo, prestando especial atención al aspecto de recomendación.
- Capítulo 4. Evaluación y Resultados. Este capítulo expone los experimentos realizados para evaluar el comportamiento del núcleo recomendador de e-Turismo.
- Capítulo 5. Conclusiones. En este capítulo, se resumen las aportaciones más relevantes de este trabajo y se indican algunas líneas abiertas por las que puede continuarse este trabajo.

Capítulo 2

Estado del Arte

2.1. Antecedentes

Internet ofrece actualmente un volumen de información tal, que la selección de información realmente interesante para el usuario supone una tarea tediosa y compleja. El usuario, a menudo, no es capaz de visualizar toda la información que se le ofrece, por lo que podría existir información de gran interés para él pero se pierde entre el resto de datos. Los primeros sistemas de ayuda en la búsqueda de información fueron:

- **Motores de búsqueda** (como los buscadores de Internet, Google [33] o Yahoo [91]): Son sistemas informáticos que indexan archivos almacenados en servidores web. Las búsquedas se hacen con palabras clave o con árboles de jerarquías por temas; el resultado de la búsqueda es un listado de URLs en los que se mencionan temas relacionados con las palabras claves buscadas. Habitualmente, gran parte de la información que ofrecen los motores de búsqueda no resulta interesante para el usuario.
- **Sistemas de ayuda**: Enseñan al usuario a usar el programa que acompañan, describiéndolo y explicando la forma de uso. Por ejemplo, Microsoft incluye un sistema de ayuda que recoge información sobre las características del usuario y sus acciones, el estado del programa, y las palabras introducidas en una consulta del usuario, para calcular la probabilidad de que el usuario se encuentre necesitando ayuda sobre algún tema. Estos sistemas suelen enlazar con páginas web que completan la información ofrecida.

- **Sistemas de filtrado [11] y recuperación de la información [43]:** El objetivo de estos sistemas es obtener la información que es relevante para el usuario, minimizando el número de elementos irrelevantes.

Los sistemas de filtrado de información permiten eliminar gran cantidad de información no deseada, pero lo más útil sería que estuvieran dotados de la capacidad de adecuarse a las preferencias del usuario, mediante técnicas de aprendizaje automático. Se trata de evolucionar hacia la creación de servicios web dotados de Inteligencia Artificial, que brinden al usuario el contenido en el cual está interesado, en lugar de enviarle un paquete de información para que tenga que encontrar lo que busca.

Antes de definir el concepto de Sistema Recomendador, hay que diferenciar entre sistemas **adaptables** y **adaptativos** [34, 60]. En el primer sistema, la **adaptación** la decide el usuario, eligiendo los valores de ciertos parámetros. Y en los sistemas **adaptativos**, la adaptabilidad se presenta como una característica autónoma del sistema, sin intervención del usuario. Ambas características pueden coexistir en un mismo sistema sin que esto afecte el rendimiento del mismo. Al contrario, aumenta la efectividad y eficacia del sistema.

2.2. Sistemas Recomendadores

Un **Sistema de Recomendación (SR)** [68, 76] es un tipo específico de filtro de información adaptativo; es una técnica que trata de presentar al usuario, únicamente, información sobre la que esté interesado.

El eje central de un SR es el término **personalización** [1]. El objetivo de la personalización es proveer a los usuarios con lo que necesitan sin necesidad de preguntárselo explícitamente. El sistema debe inferir lo que requiere el usuario basándose en datos que conoce acerca de él y en datos de otros usuarios similares (información que les interesa a personas similares a él) o en datos de su entorno (perfil demográfico). En definitiva, un SR ofrece la posibilidad de personalizar la información en internet por medio de un filtrado, de manera que muestre una cantidad de información que el usuario pueda manejar, que se adapte a sus necesidades y gustos y sobre todo que sea de su interés.

El resultado de la recomendación es un conjunto de elementos (también llamados ítems o características) que se muestran al usuario de la manera en la que se haya definido la interfaz.

En [1] se define el proceso de personalización en tres etapas:

1. **Recopilar información del usuario:** Entender a los usuarios, recopilando información sobre ellos y transformarla en conocimiento. Se almacena en forma de perfiles de usuario.
2. **Recomendar elementos:** Ofrecer al usuario ítems basados en el conocimiento adquirido.
3. **Grado de satisfacción:** Medir el impacto de la personalización determinando el grado de satisfacción del usuario con los ítems recomendados. Esta etapa sirve de **feedback** a la primera etapa, pues permite aumentar el conocimiento sobre el usuario.

Los SR aparecen en la década de los 90. El interés en el campo ha sido enorme tanto a nivel académico como a nivel de la industria [2]. El primer SR es *Tapestry* [32] (donde aparece por primera vez el término filtrado colaborativo). Se han creado multitud de servicios web que utilizan SR para personalizar la información que contienen, como por ejemplo: Amazon [5], Audioscrobbler de Last.fm [48], Like-i-like [48], MovieLens: SR de cine [57], MyStrands [84], Photoree [65], entre otros (Ver tabla 2.1). En [76] se pueden encontrar más ejemplos de aplicaciones de los SR. Algunas empresas han incorporado SR en sus servicios comerciales.

Los SR son un campo en auge, sin embargo queda mucho camino por recorrer [2], incluyendo técnicas que mejoren la representación del comportamiento del usuario y la información sobre los ítems a recomendar, mejoras en los algoritmos de modelado de la recomendación, puntuaciones de elementos multicriterio, aplicación de técnicas de computación evolutiva...

El campo de los SR tiene enlace con multitud de campos, entre ellos: ciencia cognitiva (cognitive science) [73], teoría de la aproximación (approximation theory) [66], recuperación de la información (information retrieval) [43, 74], teorías de predicción (forecasting theories) [8], gestión comercial (management science) [58] y marketing (choice modeling in marketing) [50]. Fue en los 90 cuando los investigadores comenzaron a centrarse en las estructuras de ratios propias de los SR, es

decir, la formulación más simple de un SR se reduce al problema de estimar ratios para los elementos aún no vistos por el usuario, pues una vez se ha realizado la estimación se puede recomendar al usuario los elementos con mayor ratio estimado.

2.2.1. Ventajas e Inconvenientes de los Sistemas Recomendadores

La incorporación de los SR en las páginas Web permite personalizar la utilización del servicio, ofreciendo al usuario únicamente la información que éste necesita o en la que está interesado. En [34] se comentan ampliamente los beneficios de los SR, como por ejemplo, la eliminación de información innecesaria para el usuario gracias a sus preferencias generales.

Pero en la actualidad, los SR introducen ciertos problemas sociales [68], como guiar las recomendaciones a intereses económicos, bien porque el recomendador ofrece mayoritariamente elementos recomendados de un tipo en concreto, guiado por ciertos intereses, o porque se registren en el sistema usuarios que puntúan favorablemente recomendaciones de una empresa, en perjuicio de otra, de forma interesada. La solución a este problema depende del nivel de confianza que el usuario deposite en la página Web.

Otro problema importante es la falta de privacidad que supone que diferentes sistemas obtengan información sobre las preferencias de los usuarios. Diversos estudios tratan de solucionar o aliviar este problema [89].

2.2.2. Definición Formal

La definición formal de un SR puede encontrarse en [38, 67, 79]. La cual se define de la siguiente manera, sea:

U Conjunto de Usuarios.

I Conjunto de posibles ítems (ó elementos) a recomendar.

u Usuario sobre el cual se realiza la recomendación.

i Ítem para el que se quiere generar la predicción de la preferencia de u.

fu Función de utilidad que mide la utilidad de i para u. Habitualmente es el ratio.

De una manera formal, un SR se puede formular con las siguientes ecuaciones:

$$fu : U \times I \longrightarrow R \quad (2.2.1)$$

$$\forall u \in U, \exists i' \in I \mid i'_u = \underset{i \in I}{\text{máx}}(fu(u, i)), \quad (2.2.2)$$

La función de utilidad suele ser la puntuación dada por el usuario al elemento (aunque también podría ser cualquier función heurística), que luego será utilizada como **feedback** [52] en el SR.

2.3. Perfil del Usuario

Los SR trabajan partiendo de la información que se conoce sobre el usuario, habitualmente almacenada en el **perfil de usuario**. Cuantos más datos se conozcan sobre el usuario (enlaces a los que visita, preferencias generales y/o específicas, gustos musicales, turísticos, etc.), cuanta más información se tenga de su perfil, más acertada será la recomendación ofrecida. El perfil debe contener la información necesaria para que se puedan realizar recomendaciones al usuario, siempre se debe evitar que la tarea para solicitar dicha información no sea demasiado tediosa para el usuario.

Inicialmente se solicitan al usuario una serie de datos (email, gustos generales, edad, etc.) que se almacenan en su perfil. Posteriormente, el sistema debe enriquecer su perfil con la información que va conociendo del usuario a medida que éste utiliza la aplicación (**feedback**).

El perfil de usuario suele generar dos tipos de información:

- **Modelo de las preferencias del usuario:** Descripción de los tipos de ítems en los que está interesado el usuario. La representación más común es una función que para cada ítem predice el interés de ese ítem para el usuario (función de utilidad).

Habitualmente, los ítems o características almacenadas en el perfil tienen asociado un valor (ratio, puntuación o ranking) que permite conocer el grado de interés del usuario en esa característica en concreto, también puede aparecer información sobre si el usuario ha puntuado el ítem positiva o negativamente.

Toda esta información es dinámica, se modifica durante las sucesivas recomendaciones, añadiendo nuevas características, modificando ratios o marcando elementos como obsoletos.

- **Histórico de interacción del usuario con el SR.** Puede consistir en el conjunto de ítems en los que ha estado interesado el usuario o las peticiones que el usuario ha realizado al SR.

2.3.1. APML

APML [6, 39] (Attention Profiling Mark-up Language) o Perfil de atención es un estándar abierto basado en XML [36] para encapsular el perfil de usuario. Permite a los usuarios que su perfil se comparta entre diversos servicios web (que podrán utilizar este perfil para realizar recomendaciones). Cada sitio web añadirá al perfil los nuevos datos obtenidos del usuario. La idea es comprimir todos los perfiles de usuario de los sitios web visitados en un formato portable que contenga los gustos del usuario.

Cuando se hace uso de un servicio o aplicación que soporta APML, el perfil de atención se coloca en un archivo XML fácil de compartir, que luego se puede importar a otros servicios que soporten este estándar.

Esta forma de compartir información sobre los usuarios en los sitios web tiene el peligro para los usuarios de la falta de privacidad sobre sus datos, aunque también le reporta beneficios, pues permite filtrar gran cantidad de información no deseada.

Los beneficios de usar el perfil de atención APML son varios:

- **Información dirigida:** si se comparte el archivo APML se puede hacer que los servicios y sitios que se visiten sugieran contenido que van con los gustos e intereses del usuario, en lugar de servir información genérica y común para todo el mundo.
- **Publicidad sin molestias:** a lo sumo, cualquier publicidad que se reciba, estará asimismo personalizada según las cosas que son útiles o de interés.
- **Control:** el usuario controla el perfil de atención, APML da la opción de escoger con quién compartir los datos.

APML establece cuatro derechos fundamentales del usuario:

- Propiedad: el usuario es dueño de sus datos de atención y puede almacenarlos donde desee (tiene *control*).
- Movilidad: el usuario puede mover de manera segura su atención donde quiera y cuando quiera (tiene la capacidad de *transferir* su atención).
- Economía: el usuario puede prestar atención a quien quiera y recibir un pago a cambio (su perfil de atención tiene un *valor*).
- Transparencia: el usuario puede ver exactamente cómo se usa su Perfil de Atención (puede *decidir* en quien confiar).

APML provee la forma en la que el usuario puede tener control sobre la información almacenada en su perfil, para usarla cuando crea conveniente y para sus propios fines y beneficios. Al mismo tiempo, los SR pueden tener un cuadro más claro de lo que podría interesarle al usuario, ahorrándole la molestia de rellenar la información común en cada sitio web visitado.

Actualmente está siendo utilizado por algunos sitios web y su interés va en aumento. Sigue la tendencia más amplia de evolucionar hacia una Web Semántica [78], más inteligente. La web semántica es una extensión de World Wide Web donde los contenidos web se expresan mediante un lenguaje que puedan usarlo tanto los usuarios como las herramientas web, permitiendo a personas y máquinas compartir información.

Algunos servicios que usan actualmente APML son: Dandelife [20], Engagd [26], Particls [61], Cluztr [19], Bloglines [13] o Newsgator [59].

2.4. Técnicas de Recomendación

2.4.1. Técnicas de Recomendación Básicas

Los SR trabajan con [16] datos base o iniciales, que es la información previa al proceso de recomendación. Datos de entrada, los proporciona el usuario al sistema para realizar la recomendación. Un algoritmo para combinar los datos base y los

datos de entrada capaz de proporcionar sugerencias. En este apartado se muestran los distintos algoritmos que proporcionan la recomendación.

Se distinguen seis técnicas principales de recomendación:

1. **Recomendación demográfica** (demographic) [30]: Se clasifica al usuario basándose en sus atributos personales y se realizan recomendaciones según el grupo demográfico al que pertenece. El grupo demográfico al que pertenece el usuario se determina de la investigación en el campo del marketing. La ventaja de estos SR es que no requieren que el usuario puntúe ítems (ratios) ni requiere información de otros usuarios. Es la técnica más simple, pero también es la menos precisa. Su gran ventaja es que siempre es capaz de producir una recomendación.
2. **Recomendación colaborativa** (collaborative) [12, 77]: Es la técnica más común. Recomienda ítems en función de los gustos de los usuarios similares al actual. Son usuarios similares los que tienen gustos parecidos, es decir, que han puntuado favorablemente ítems que el usuario actual también ha puntuado favorablemente. Esta técnica requiere información de un número elevado de usuarios para que la recomendación sea precisa. La principal ventaja de esta técnica es que es independiente de la representación de los ítems a recomendar y trabaja bien con objetos complejos [16]. También presenta la ventaja de que los elementos que recomienda no siempre son del tipo de los ya vistos por el usuario (introduce elementos novedosos en la recomendación). La recomendación colaborativa **item-to-item** busca elementos similares en vez de usuarios similares. El perfil de usuario en estos sistemas suele ser un vector de ítems con los ratios correspondientes a cada ítem.
3. **Recomendación basada en contenido** (content-based): Es una ampliación del filtrado de información [11]. Se definen los ítems por sus características y se recomienda al usuario ítems con características similares a los que ha puntuado favorablemente en el pasado. El problema de esta técnica es que siempre se recomiendan ítems similares a los ya vistos por el usuario, por lo que no se introduce ninguna novedad en lo que se le ofrece. La recomendación basada en contenido **item-to-item** busca ítems similares al actual.

4. **Recomendación basada en conocimiento** (knowledge-based) [85]: Dispone de información acerca de cómo un ítem satisface una necesidad del usuario y establece la relación entre una necesidad y una recomendación. Sugiere productos basados en la inferencia sobre las necesidades y preferencias del usuario. El modelo de usuario es una estructura que soporta esta inferencia, que puede basarse en: queries, casos (razonamiento basado en casos), métricas de similitud para realizar el matching o parte de una ontología.
5. **Recomendación basada en utilidad** (utility-based): realiza sugerencias basadas en el cálculo de la utilidad de cada posible ítem a recomendar para el usuario. Calcula el valor de la función de utilidad y elige los ítems que optimicen el valor de la función. El problema de esta técnica es la definición de la función de utilidad [35], ésta no puede ser ni muy sencilla ni muy compleja. Compara la necesidad del usuario con el conjunto de opciones disponibles, mediante la función de utilidad de cada ítem para el usuario. En algunos casos esta función es su perfil.
6. **Recomendación basada en casos** (case-based reasoning) [81]: Utilizan la información sobre la resolución de problemas (casos) previos para la resolución del caso actual. Tanto los SR basados en conocimiento como los SR basados en utilidad pueden utilizar técnicas de Razonamiento Basado en Casos (CBR) [42]. Por tanto, los SR basados en casos pueden verse como un subtipo de ambos SR, según el SR base del que parta. Los SR basados en utilidad utilizan como función de utilidad las técnicas con las que se resolvieron casos parecidos.

La tabla 2.1 muestra ejemplos de cada tipo de SR.

La figura 2.1 muestra las fuentes de datos de cada una de estas técnicas de recomendación. La tabla 2.2 muestra un resumen de las técnicas de recomendación [16] (utilizando la nomenclatura del apartado 2.2.2).

2.4.2. Problemas de las Técnicas de Recomendación

Los SR pueden presentar los siguientes problemas, dependiendo del tipo de SR [16]:

- **Ramp-up:** Este problema presenta, en realidad, dos vertientes:

Técnica de recomendación	Ejemplos
Demográfica	Grundy [73] y Lifestyle Finder [44]
Colaborativa	GroupLens/NetPerceptions [67], Ringo/FireFly [79], Tapestry [32], Recommender [38], MovieLens [57]
Basada en contenido	NewsWeeder [47], Personal WebWatcher [55], InfoFinder [45], Letizia [49]
Basada en conocimiento	Google [15], Entrée [16]
Basada en utilidad	JR.com [41]

Tabla 2.1: Ejemplos de SR que utilizan cada técnica de recomendación.

- **Problema del nuevo usuario o start-up problem:** Cuando un usuario se registra en el sistema debe puntuar ratios suficientes para poder ser recomendado con ciertas garantías. Habitualmente el número de ratios necesarios suele ser elevado. Las técnicas de recomendación se suelen basar en la puntuación que el usuario otorga a los ítems, por ello si el usuario es nuevo es probable que tenga pocos elementos puntuados y por tanto pocos elementos de comparación. Este problema aparece en SR basados en contenido y SR colaborativos.
- **Problema del nuevo elemento:** Hasta que un elemento no ha sido puntuado por un número suficiente de usuarios no es recomendado. Un nuevo ítem es probable que tenga pocas puntuaciones de los usuarios y por ello será difícil que ese ítem sea recomendado. Cuando una persona puntúa un ítem por primera vez no obtiene mucho beneficio de ello [9], ya que no permite inferir nueva información. Este problema aparece en SR colaborativos.
- **Sparsity o dispersión de ratios:** Los ratios del usuario actual pueden no coincidir con los ratios de otros usuarios y, por tanto, puede haber pocos usuarios con los que comparar o pocos elementos con los que buscar similitudes. Este problema aparece en SR basados en contenido y colaborativos. Un usuario con gustos comunes tiene mayor probabilidad de que se acierte en la recomendación ofrecida.

Técnica	Datos base	Entrada	Proceso
Demográfica	Información demográfica sobre U y sus ratios de los ítems en I.	Información demográfica sobre u.	Identificar usuarios que sean demográficamente similares a u y extrapolar sus ratios a i.
Colaborativa	Ratios de U de los ítems en I.	Ratios de u a los ítems en I.	Identificar los usuarios en U similares a u y extrapolar sus ratios para i.
Basada en contenido	Características de los ítems en I.	Ratios que ha dado u a los ítems en I.	Generar una clasificación que se ajuste a los ratios del comportamiento de u y usarlos sobre i.
Basada en conocimiento	Características de los ítems en I. Conocimiento de cómo esos ítems encajan con las necesidades del usuario.	Descripción de las necesidades o intereses de u.	Inferir un ajuste entre i y las necesidades de u.
Basada en utilidad	Características de los ítems en I.	Función de utilidad de los ítems en I que describen las preferencias de u.	Aplicar la función a los elementos que determinan los ratios de i.

Tabla 2.2: Técnicas de recomendación [16].

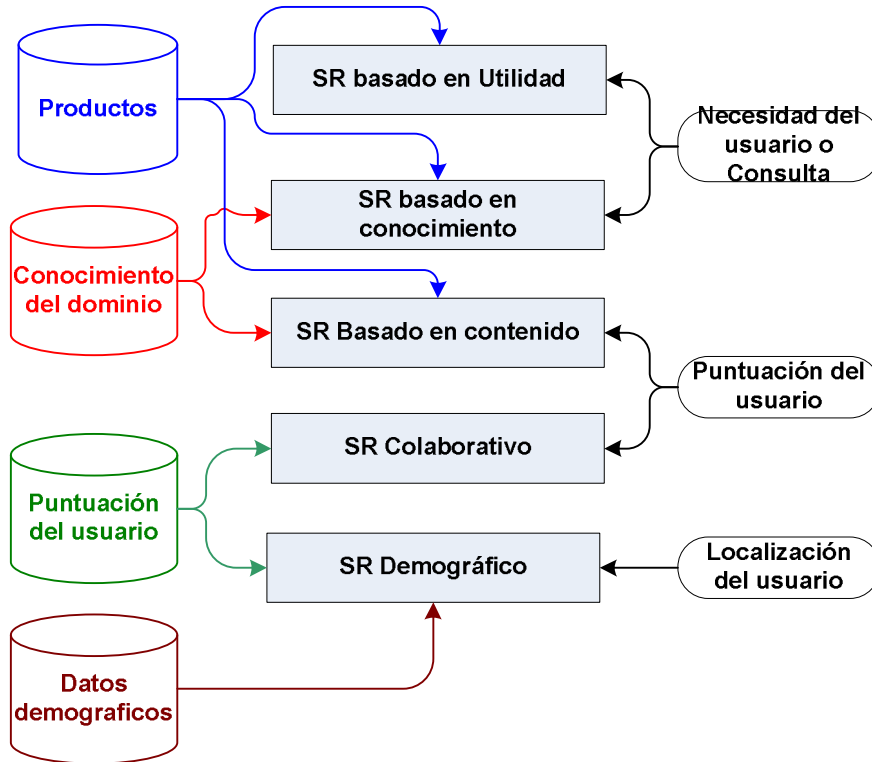


Figura 2.1: Técnicas de recomendación [17]

- **Recopilación de información demográfica** [16]: Los SR demográficos requieren que el usuario introduzca información personal para que se puedan realizar recomendaciones basadas en su perfil demográfico. Puede resultar tedioso para el usuario proporcionar esta información. Sin embargo, cuanto mayor sea la información proporcionada al sistema, de mayor calidad será la recomendación ofrecida. Además, puede que el usuario no esté dispuesto a proporcionar al sistema información (por miedo a la falta de privacidad o poca confianza en el servicio) que podría resultar valiosa a la hora de realizar la recomendación. Una forma de tratar este problema es determinar el número óptimo de ratios que se debe solicitar al usuario (puede formularse como un problema de optimización) o utilizar técnicas no intrusivas.
- **Efecto portafolio (portfolio effect)** [16]: Ciertos SR no pueden sugerir al usuario un ítem demasiado parecido a otro que ya ha visto antes (no se puede

sugerir que compre un libro que ya tiene, en una nueva edición, o en una encuadernación diferente,...), en cambio en otros sistemas la similitud sí que será un criterio para sugerir el elemento al usuario (una noticia parecida a otra que ya ha visto anteriormente).

- **Falta o exceso de resultados de la recomendación:** Los SR asumen que los requerimientos y restricciones del usuario van a poder satisfacerse utilizando las técnicas de recomendación adecuadas, de manera que siempre se va a poder presentar al usuario una recomendación [54]. Sin embargo el proceso de recomendación puede fallar, no produciendo ninguna recomendación que satisfaga los requerimientos del usuario. También puede ocurrir que el proceso de recomendación produzca demasiados resultados, con lo que la recomendación ofrecida no es tan útil. Esto se cumplía en el apartado .
- **Serendipity** [69]: Algunos SR ofrecen al usuario únicamente ítems similares a los ya vistos anteriormente, asumiendo que son los que el usuario desea ver. Sería deseable ofrecerle cosas nuevas que le puedan gustar. La falta de novedades es un problema, sobre todo, en los SR basados en contenido y demográficos, que los SR colaborativos no suelen tener.

2.4.3. Técnicas de Recomendación Híbridas

Como hemos visto, todas las técnicas básicas de recomendación tienen puntos fuertes y débiles. Por ejemplo, la recomendación demográfica requiere de pocos datos para realizar la recomendación, pero las recomendaciones obtenidas no son muy precisas. La recomendación colaborativa requiere gran cantidad de usuarios (con puntuaciones en los mismos ítems que el usuario actual) para ser efectiva. Para solucionar estos problemas se usan los SR híbridos [16], que combinan varias técnicas de recomendación intentando que cada una de las técnicas evite los problemas de las otras. Habitualmente los SR híbridos combinan dos técnicas de recomendación. Estas se pueden combinar de distintas formas [17]:

1. **Recomendador híbrido ponderado (weighted)** (ver figura 2.2): Se obtienen los resultados de las técnicas de recomendación, se clasifican (se combinan linealmente, se unen o intersectan) y se ponderan para obtener los resultados finales.

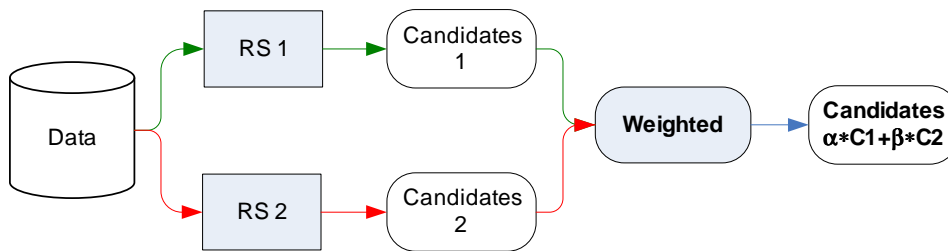


Figura 2.2: Recomendador híbrido ponderado.

2. **Recomendador híbrido mezclado (mixed)** (ver figura 2.3): Se procesan conjuntamente diversas técnicas de recomendación.

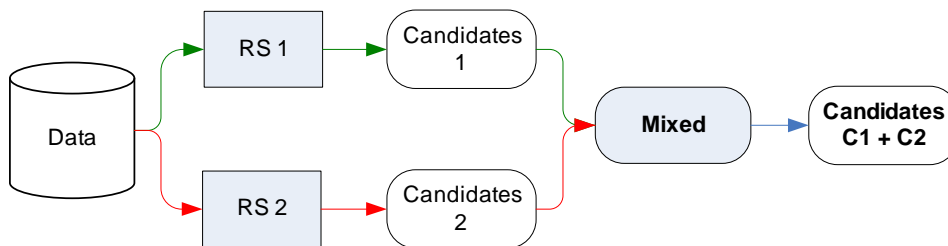


Figura 2.3: Recomendador híbrido mezclado.

3. **Recomendador híbrido alternado (switching)** (ver figura 2.4): El sistema utiliza algún criterio para alternar las recomendaciones de cada técnica.
4. **Recomendador híbrido en cascada (cascade)** (ver figura 2.5): Un recomendador refina la recomendación dada por otro. En una etapa previa un recomendador produce una recomendación que en una segunda etapa ajusta otro recomendador. La información utilizada por el segundo recomendador es únicamente la salida del primer recomendador.
5. **Recomendador híbrido de combinación de características (feature**

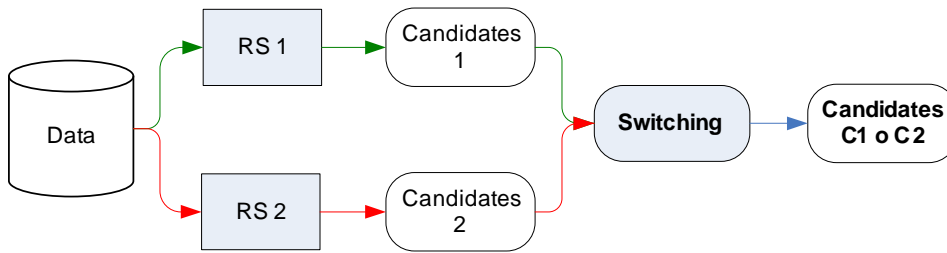


Figura 2.4: Recomendador híbrido alternado.

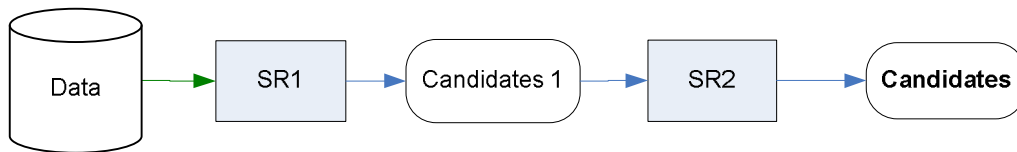


Figura 2.5: Recomendador híbrido en cascada.

combination) (ver figura 2.6): La información obtenida de una o más técnicas de recomendación (habitualmente colaborativa) se utiliza como una característica adicional asociada a los datos. Sobre estas características, combinadas con la información general, se utiliza otra técnica de recomendación.

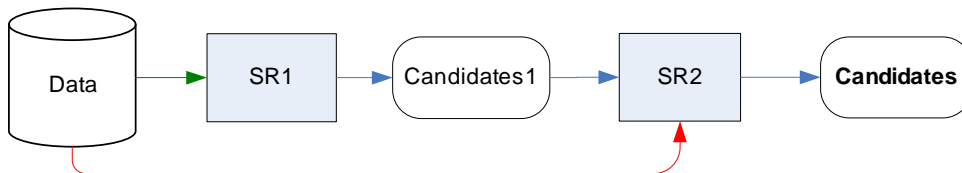


Figura 2.6: Recomendador híbrido de combinación de características.

6. **Recomendador híbrido de aumento de características (feature augmentation)** (ver figura 2.7): La información utilizada por el segundo recomendador es la información general más la salida del primer recomendador. Se usa el resultado del primer recomendador para aumentar las características de entrada al segundo recomendador.
7. **Recomendador híbrido meta-level** (ver figura 2.8): Utiliza el modelo completo generado por una técnica de recomendación como entrada a otra. El

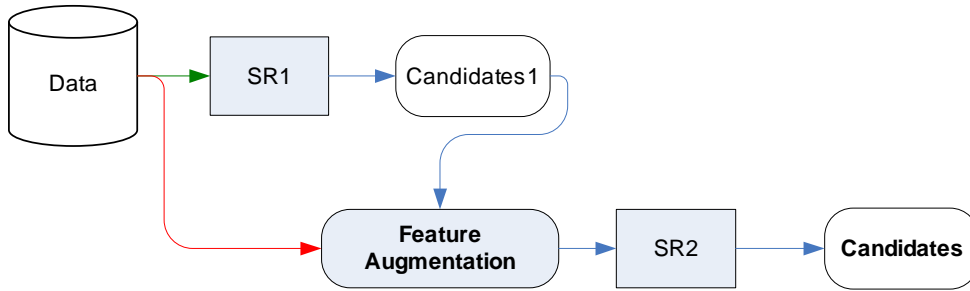


Figura 2.7: Recomendador híbrido de aumento de características.

primer SR genera un modelo que usa el segundo SR.

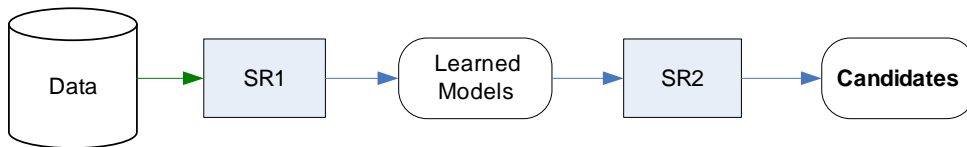


Figura 2.8: Recomendador híbrido meta-level.

En [16] se muestra una tabla con todas las posibles técnicas de recomendación utilizadas por cada tipo de recomendador híbrido. También se muestran aquellas combinaciones no posibles.

2.5. Refinamiento de Resultados: Sistemas de Recomendación Conversacionales

Un problema de la búsqueda de información es que proporcione pocos o demasiados resultados. Este problema se intenta solucionar de una forma diferente con los SR conversacionales [54]. Por ejemplo, cuando una recomendación falla (no produce resultados) se inicia una interacción con el usuario (conversación) para refinar las peticiones. Este procedimiento también se usa cuando hay un número excesivo de resultados, iniciando un diálogo con el usuario que permite aumentar las restricciones. El diálogo se basa en métodos de selección de características (feature-selection methods).

En los SR vistos hasta ahora, el usuario solicita una recomendación al sistema el cual, basándose en los datos aportados en su perfil (o en datos de otros usuarios) proporciona una recomendación. En cambio, un sistema conversacional debe tener en cuenta los datos aportados por el usuario durante toda la conversación. Además, permite que el usuario exprese sus necesidades, preferencias o restricciones sobre cada ítem en concreto.

Una diferencia entre los filtros de información y los SR es la posibilidad de capturar el feedback del usuario durante el proceso de recomendación [52]. Las técnicas de captura del feedback se pueden clasificar según:

- **Tipo de feedback:**
 - **Navigation by asking o feature-level** [80]: Se le solicita al usuario información sobre cada una de las características (p.e. seleccionar el tipo de habitación de hotel: doble o individual). Después se le pueden solicitar datos complementarios sobre lo seleccionado (p.e. si ha elegido doble, preguntar una o dos camas). Cada nueva respuesta permite que el sistema elimine elementos a recomendar (elementos irrelevantes). Requiere un esfuerzo considerable por parte del usuario, porque debe indicar que características serán nuevos criterios de búsqueda. Este proceso puede ser tedioso para el usuario [80]. En otras ocasiones el usuario no tendrá una visión clara de lo que quiere, por lo que no podrá dar una respuesta. En [83] se estudia que características podrían ser las más discriminantes, con el objeto de solicitar la menor cantidad de datos posible.
 - **Navigation by proposing (browsing) o case-level** [80]: Solicita al usuario que puntúe las recomendaciones ofrecidas como relevantes o no relevantes. No requiere demasiado esfuerzo por parte del usuario.
- **Coste para el usuario:** Viene representado por la complejidad en la forma que debe puntuar el usuario las recomendaciones. Por ejemplo, cuando se le solicita al usuario que puntué de forma negativa o positiva, se dice que es de bajo coste, pero el usuario debe rellenar formularios extensos para realizar la puntuación, se dice que es de alto coste.

Se distinguen cuatro **estrategias** de feedback:

Feedback	Coste	Ambigüedad	Experiencia	Interfaz	Tipo
Value elicitation	Alto	Bajo	Alto	Alto	N. asking
Critiquing	Medio	Alto	Medio	Bajo	N. asking + N. proposing
Ratings	Medio	Medio	Medio	Bajo	N. proposing
Preference	Bajo	Alto	Bajo	Bajo	N. proposing

Tabla 2.3: Comparación de las estrategias de feedback [53].

- **Value elicitation:** Es de tipo navigation by asking puro, de alto coste. Ejemplos: ExpertClerk [80] y NaCoDaE [3].
- **Tweaking o critiquing:** Es una combinación de navigation by asking y navigation by proposing de coste medio. Se le propone al usuario secuencias de casos posibles y el usuario elige. Ejemplo: FindMe [18].
- **Ratings-based:** Navigation by proposing de coste medio/alto. Se le solicita al usuario que puntúe propuestas de recomendación. Los valores de puntuación están en un cierto rango. Ejemplos: Ptv [82], Fab [10] y Turas [51].
- **Preference-based:** navigation by proposing de coste bajo. Se le solicita al usuario que puntúe propuestas de recomendación (positiva o negativamente). Ejemplos: Amazon.com [5] y Google [33].

En la tabla 2.3 se muestra una comparación de las diferentes técnicas de feedback. La comparación se basa en cuatro aspectos: el coste para el usuario, la ambigüedad inherente al feedback, el nivel de conocimiento sobre los ítems, el tipo de interfaz necesario para capturar el feedback y el tipo de técnica de captura.

2.6. Sistemas de recomendación de turismo y ocio

Este apartado se centra en el estudio de los sistemas de recomendación para el dominio de ocio y turismo.

El turismo es una actividad que está íntimamente conectada con las preferencias e intereses personales. Es por ello, que muchas aplicaciones web de viajes, ocio y turismo incorporan SR. Intentan con ello simular la interacción con un agente de viajes humano [23].

Algunos ejemplos de servicios web de turismo que incorporan SR son: Intrigue [7], DieToRecs (SR basado en casos) [24, 29], Triplehop's TripMatcher (SR basado en contenido) [23], ITR [70] o Trip@dvice [71, 86]. Existe abundante bibliografía sobre la necesidad de utilizar SR en servicios de turismo y ocio, y sobre experiencia de integración de SR en estos servicios.

Más escasos son los servicios creados para facilitar la creación de rutas turísticas dentro de una misma ciudad. Deep Map [22] es un proyecto para el desarrollo de guías turísticas para web y móviles. Dentro de este proyecto se enmarca el subproyecto WebGuide [90, 31], que permite obtener itinerarios para realizar visitas guiadas por la ciudad de Heidelberg, teniendo en cuenta intereses y preferencias del usuario. El sistema identifica puntos geográficos de interés y calcula la ruta basándose en información geográfica, puntos de interés turístico, medios de transporte seleccionados, preferencias y restricciones definidas por el usuario.

No es fácil aplicar técnicas de recomendación a los servicios de ocio y turismo. Algunos de los problemas que presentan son:

- La técnica fundamental de recomendación, la recomendación colaborativa, es difícilmente aplicable a la mayor parte de estos sistemas [28]. Para que pueda aplicarse, los usuarios deben puntuar muchos elementos y otros usuarios también deben haberlo hecho. Sin embargo, no son actividades de uso tan frecuente como para que el usuario puntúe suficientes elementos. Algunos sistemas utilizan técnicas conversacionales para solucionar este problema [40, 72].
- Recomendaciones para grupos: Habitualmente se viaja en grupo, por lo que hay que tener en cuenta las preferencias de varios usuarios [7]. La recomendación debe adecuarse a los gustos de la mayoría.

- La recomendación ofrecida al usuario no depende únicamente de los gustos de otros usuarios o de los suyos propios, sino que se debe tener en cuenta la información de entorno [23]: Distancia entre los puntos, posibilidades de transporte, época del año, horarios de apertura, etc.

A pesar de los problemas que presenta la utilización de recomendadores en servicios de ocio y turismo, hay gran interés en el tema y actualmente numerosos grupos siguen investigando en este campo. En concreto, el eCTRL (eCommerce and Tourism Research Laboratory) [25] liderado por Francesco Ricci, con proyectos (algunos de ellos utilizados en servicios comerciales) como: Trip@dvice, Harmonise, ETD Project o Harmo-TEN.

Capítulo 3

E-Turismo: SR de rutas turísticas planificadas en la ciudad de Valencia

3.1. Introducción

Internet es un medio de comunicación que alberga un sin número de información. Muchos usuarios utilizan a través de este medio buscadores para encontrar información turística en el momento de planificar un viaje o una salida. Los buscadores tradicionales les ofrecen a los usuarios una gran cantidad de información que difícilmente puede ser asimilada por ellos.

Un Sistema Recomendador de rutas turísticas planificadas ayuda al usuario a buscar y filtrar toda esa información turística, seleccionando la que es interesante para él. Partiendo de su **perfil de usuario**, el SR brinda información relevante y personalizada de lo que el usuario busca en el momento.

En este capítulo se presenta e-Turismo, un SR de rutas turísticas planificadas en la ciudad de Valencia. El capítulo se encuentra dividido en 4 secciones, que detallan la funcionalidad e importancia de e-Turismo para el usuario, una explicación detallada de la representación del conocimiento y de cada uno de los módulos: el Núcleo Generalista del Sistema Recomendador (Generalist Recommender System Kernel - GRSK), el Núcleo Planificador (Planning Kernel - PK) y el Núcleo Gestor de la Base de Datos (Database Manager Kernel - DBMK), que forman parte de e-Turismo. Además de una última sección en la que se explican cada una de las conclusiones

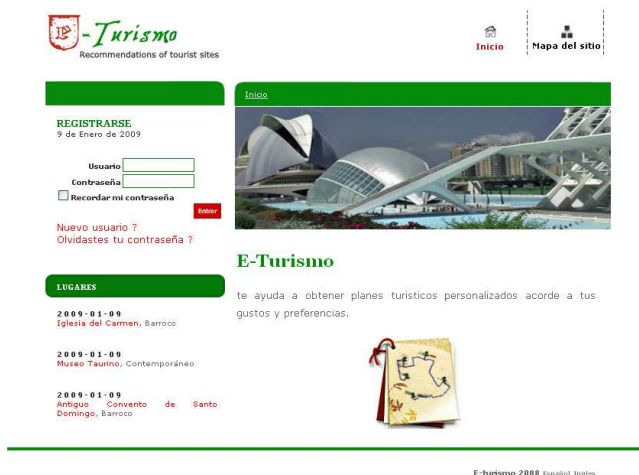


Figura 3.1: Formulario principal de e-Turismo.

obtenidas en el presente capítulo.

Es importante destacar que este trabajo se centra, fundamentalmente, en el aspecto de recomendación. Sin embargo, también se ha incluido una posible técnica muy sencilla de resolución del problema de planificación.

3.2. E-Turismo

El SR de rutas turísticas planificadas en la ciudad de Valencia, denominado e-Turismo (ver figura 3.1), tiene como objetivo principal calcular un plan turístico o de ocio para un usuario, teniendo en cuenta sus preferencias y la información de cuándo se llevará a cabo la visita y la duración de la misma. El sistema no resuelve el problema de viajar desde un lugar específico a otro, sino que se centra en recomendar una lista de las actividades que un turista puede realizar en la ciudad de Valencia. Asimismo, considera los calendarios y distancias entre las actividades con el fin de calcular un itinerario de ocio y turismo acorde al día. E-Turismo permite al usuario planificar nuevas visitas a la ciudad de Valencia, manteniendo siempre un historial de sus visitas anteriores.

Inicialmente el usuario debe registrarse en el sistema con sus datos personales y preferencias generales. Con esta información el sistema construye el **perfil del**

The screenshot shows the 'Current Visit' form on the E-Turismo website. The form is titled 'Current Visit' and includes a navigation bar at the top with links for Home, User Profile, Current Visit, Visited sites, and Recommendation. The form is divided into several sections:

- REMEMBER:** Get the best recommendations of sites to visit in the city. Includes a 'Get Perfil' button and a 'Logout' button.
- Current Visit:** A section where the user can specify their visit preferences. It includes a text input field for the user's name (cgaz) and a prompt to enter as much information as possible to get better recommendations.
- Would you like to make a particular visit today?:** A section with a list of preferences and checkboxes:
 - Styles and periods
 - Shopping
 - Museums
 - Open Spaces
 - Religious architecture
 - Defensive architecture
 - Civil Architecture
 - Gourmet
 - Sports
 - Monuments
 - Leisure
 - Health & SPA
 - Events
 - Children
 - None
- Visit Time:** A section with three input fields:
 - Date: 12/01/2009
 - Start Time: 15:19
 - End Time: 19:19
- Buttons:** Three buttons at the bottom: 'Not thanks', 'Recommendation', and 'Save'.

At the bottom of the page, there is a 'Process of Recommendation' section with a progress bar showing the current step: 'User profile - Current Visit Preferences - Recommendation - Visit confirmation - Visit Planning'. The footer includes the text 'E-turismo 2008 Spanish English'.

Figura 3.2: Preferencias específicas.

usuario, que se va actualizando con el historial de sus visitas anteriores.

Para planificar una nueva visita el usuario define, con ayuda de sus preferencias generales, unas preferencias específicas (ver figura 3.2) y unos datos básicos de la **visita actual**, como la duración, es decir, tiempo del que dispone el usuario, si va con niños o solo, etc. A partir de esta información el sistema muestra al usuario una **lista de actividades recomendadas**. Esta lista de actividades recomendadas son las que se consideran más interesantes para el usuario.

El usuario debe seleccionar de esta lista de actividades, aquellas en las que está o no interesado (ver figura 3.8 más adelante) y por consiguiente desea visitar (**seleccionar lugares a visitar**). Con estas restricciones y la duración de la visita (seleccionadas anteriormente) el sistema determina las actividades que le interesan al usuario y cuándo puede realizarlas, **planificando de este modo la visita** a los lugares turísticos seleccionados.

E-Turismo también cuenta con un proceso de retroalimentación sencillo e intuitivo para el usuario. El sistema le informa al usuario sobre las últimas actividades recomendadas y que no han sido puntuadas por él (ver figura 3.3). Una vez el usuario puntúa dichas actividades el sistema utiliza esta información para su

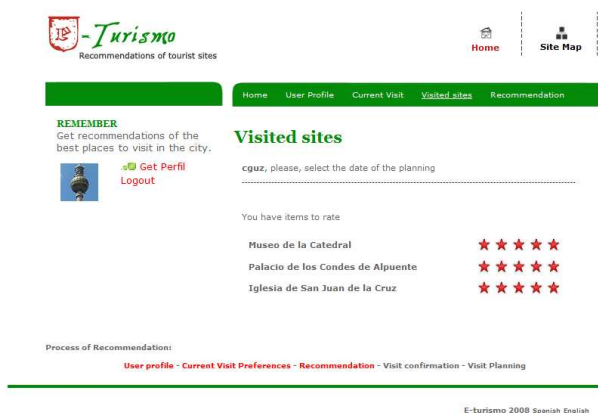


Figura 3.3: Proceso de puntuaciones.

retroalimentación (ver apartado 3.3.2).

La información a mostrar en e-Turismo se distribuyó de la mejor forma, para brindarle al usuario una Interfaz Web sencilla y fácil de usar. Se diseñó una interfaz común o estandarizada para todos los formularios. En la capa de presentación de e-Turismo la información es totalmente dinámica, por ejemplo, las preferencias del usuario se toman de la BD, es decir, si se agrega o modifica alguna preferencia general ésta será dinámicamente actualizada en la interfaz sin tener que modificar el código de la misma (ver figura 3.4).

Cabe destacar que e-Turismo está diseñado para funcionar independientemente del tipo de interfaz a utilizar, en esta tesina de Máster se desarrolló una interfaz Web. También se podía haber desarrollado una interfaz de escritorio o de móvil.

E-Turismo también le brinda la posibilidad al usuario de exportar su perfil de usuario (ver figura 3.5) utilizando el estándar APML (ver subsección 2.3.1). El sistema coloca el perfil del usuario en un archivo XML (ver figura 3.6) fácil de compartir, que luego el usuario puede importar en otros servicios que soporten este estándar. Actualmente en este archivo se guarda información de los datos de registro y preferencias generales del usuario.

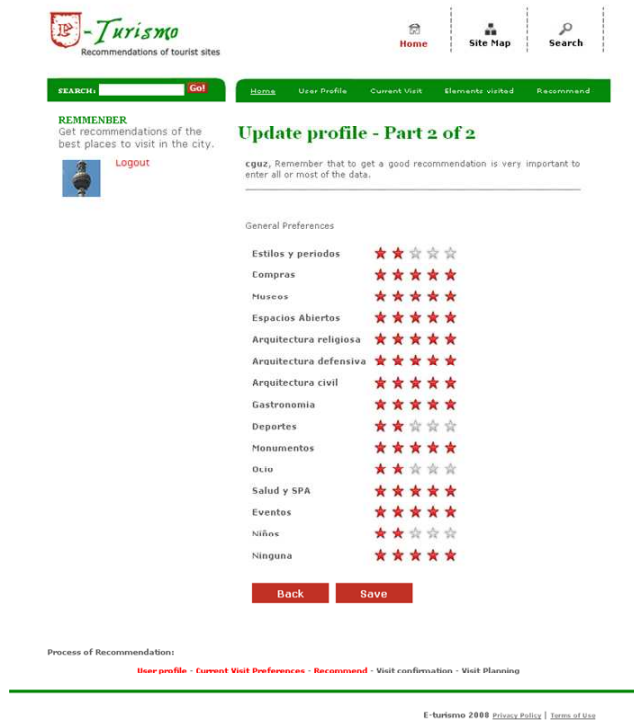


Figura 3.4: Interfaz e-Turismo.

3.3. Arquitectura de e-Turismo

E-Turismo se compone de tres subsistemas (ver figura 3.7): El Núcleo Generalista del Sistema Recomendador (Generalist Recommender System Kernel - GRSK), el Núcleo Planificador (Planning Kernel - PK) y el Núcleo Gestor de la Base de Datos (DBMK). GRSK es un módulo recomendador de propósito general (base fundamental de esta tesina) mientras que el resto de subsistemas dependen de la aplicación (turismo en la ciudad de Valencia). Estos subsistemas son coordinados por un cuarto módulo: el subsistema de control.

La primera vez que un usuario utiliza e-Turismo, debe registrarse en el sistema con sus datos personales y preferencias generales, dando lugar al primer paso, **construir el perfil de usuario**. Con esta información el sistema construye un primer perfil del usuario que se actualizará de acuerdo con la puntuación dada a los lugares



Figura 3.5: Obtener perfil del usuario.

```

- <Profile name="Turista">
+ <ImplicitData></ImplicitData>
- <ExplicitData>
  <Concepts>
  - <Sources>
    <Source key="http://158.42.184.205/Styles and periods" value="4,00" name="Styles and periods" type="application/rss+xml"/>
    <Source key="http://158.42.184.205/Shopping" value="10,00" name="Shopping" type="application/rss+xml"/>
    <Source key="http://158.42.184.205/Museums" value="10,00" name="Museums" type="application/rss+xml"/>
    <Source key="http://158.42.184.205/Open Spaces" value="10,00" name="Open Spaces" type="application/rss+xml"/>
    <Source key="http://158.42.184.205/Religious architecture" value="10,00" name="Religious architecture" type="application/rss+xml"/>
    <Source key="http://158.42.184.205/Defensive architecture" value="10,00" name="Defensive architecture" type="application/rss+xml"/>
    <Source key="http://158.42.184.205/Civil Architecture" value="10,00" name="Civil Architecture" type="application/rss+xml"/>
    <Source key="http://158.42.184.205/Gourmet" value="10,00" name="Gourmet" type="application/rss+xml"/>
    <Source key="http://158.42.184.205/Sports" value="4,00" name="Sports" type="application/rss+xml"/>
    <Source key="http://158.42.184.205/Monuments" value="10,00" name="Monuments" type="application/rss+xml"/>
    <Source key="http://158.42.184.205/Leisure" value="4,00" name="Leisure" type="application/rss+xml"/>
    <Source key="http://158.42.184.205/Health & SPA" value="10,00" name="Health & SPA" type="application/rss+xml"/>
    <Source key="http://158.42.184.205/Events" value="10,00" name="Events" type="application/rss+xml"/>
    <Source key="http://158.42.184.205/Children" value="2,00" name="Children" type="application/rss+xml"/>
    <Source key="http://158.42.184.205/None" value="10,00" name="None" type="application/rss+xml"/>
  </Sources>
  </ExplicitData>
</Profile>

```

Figura 3.6: Perfil del usuario en archivo XML.

que el usuario visita (**feedback**). Por tanto, los lugares que han sido visitados deberán ser puntuados por el usuario (les asigna un porcentaje de aceptación de 0 a 100).

Además de estas preferencias generales, el usuario puede definir unas preferencias específicas, particulares para las nuevas visitas, por si desea hacer algo diferente a lo habitual. También el sistema solicita los datos de la visita: fecha de la visita, tiempo disponible, si va solo o con niños, etc.

El subsistema encargado de construir el perfil del usuario es el de **Control**, es el núcleo organizativo de e-Turismo. Funciona como intermediario entre la interfaz Web y los demás subsistemas. Este subsistema inicia la ejecución de los demás subsistemas y centraliza el intercambio de información. Esto incluye, convertir los datos del usuario en el formato apropiado para la consulta, mostrar al usuario la lista de actividades recomendadas, registrar las actividades que son finalmente seleccionadas por el usuario, y, posteriormente, mostrar al usuario el plan programado para la visita turística.

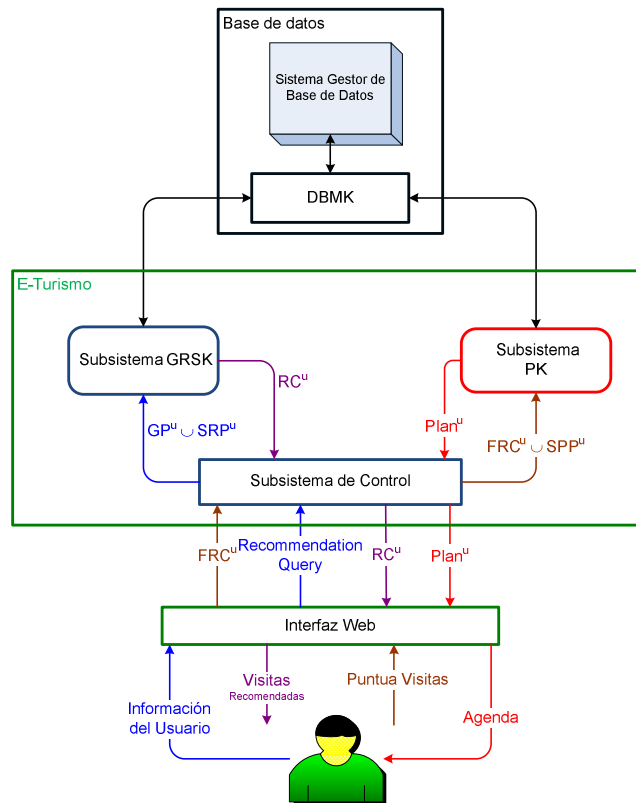


Figura 3.7: Sistema e-Turismo.

El segundo paso es generar una **lista de actividades recomendadas al usuario**. Esta lista la realiza el subsistema GRSK, cuya entrada son las preferencias generales y específicas del usuario. GRSK calcula las recomendaciones de acuerdo con el perfil del usuario actual y de otros usuarios (dependiendo de la técnica de recomendación). La recomendación es la lista de lugares propuestos para visitar. En la sección 3.3.2 se explica el funcionamiento del GRSK.

En la figura 3.8 se muestra un ejemplo de una lista de actividades recomendadas para una visita particular. A cada lugar se le asocia una prioridad y GRSK selecciona las de mayor prioridad. Se le presenta al usuario una lista de recomendaciones, y a partir de esta lista, el usuario selecciona las actividades en las que realmente está interesado (ver figura 3.8), las que no seleccione serán descartadas en la planificación turística definitiva.

The screenshot shows the 'E-Turismo' website interface. At the top, there is a logo for 'E-Turismo' with the tagline 'Recommendations of tourist sites'. Navigation links for 'Home', 'Site Map', and 'Search' are visible. Below the navigation is a search bar and a menu with 'Home', 'User Profile', 'Current Visit', 'Elements visited', and 'Recommend'. A 'REMEMBER' section on the left contains a user profile picture and a 'Logout' button. The main content area is titled 'Recommended' and displays a list of 12 recommended places to visit, each with a checkbox, a name, a category, and a 'Ver mapa en google' link. The sites listed are:

- Museo de la Catedral (Museum, Religious)
- Paseo de la Alameda (OpenSpace, Botanical Garden)
- Jardines de Viveros (OpenSpace, Botanical Garden)
- Jardin Botánico (OpenSpace, Botanical Garden)
- Palacio del Almirante (Building, Gothic)
- Palacio de la Generalitat (Building, Gothic)
- Reales Atarazanas (Building, Gothic)
- Torre del Miguelete (Building, Gothic)
- Palacio de los Condes de Alpuente (Building, Gothic)
- Mercado de Colón (Building, Markets)

At the bottom, a 'Process of Recommendation' section shows a progress bar with steps: 'User profile', 'Current Visit Preferences', 'Recommend', 'Visit confirmation', and 'Visit Planning'. The 'Recommend' step is currently active. The footer contains 'E-turismo 2009', 'Política Privada', and 'Términos de Uso'.

Figura 3.8: Lista de actividades recomendadas al usuario.

El tercer paso del proceso consiste en **planificar la visita a los lugares turísticos seleccionados**. En esta fase, el sistema determina los horarios de las actividades de acuerdo con las restricciones de tiempo del usuario y los horarios de los lugares seleccionados. El subsistema a cargo de esta fase es el Núcleo de Planificación (PK). La entrada a PK es el conjunto de actividades seleccionadas por el usuario, las actividades no seleccionadas o marcadas como indiferentes, y otras preferencias necesarias para la planificación, como la disponibilidad de tiempo. El resultado es un plan turístico que contiene las actividades que llevará a cabo el usuario, junto con el horario de inicio y la duración estimada de cada actividad. En la sección 3.3.3 se explica cómo se calcula este plan.

El paso final es el **proceso de puntuación de los lugares visitados por parte del usuario o retroalimentación del SR (feedback)**. Cuando el usuario inicia una nueva sesión en el sistema, se le pedirá que puntúe los lugares o actividades

The screenshot shows the e-Turismo website interface. At the top, there is a logo for 'e-Tourism' with the tagline 'Recommendations of tourist sites'. Navigation links include 'Home', 'Site Map', and 'Search'. A search bar is present with a 'Go!' button. Below the search bar, a 'WELCOME' message is displayed for user 'john' on May 18, 2008, with links for 'User Profile' and 'Logout'. A navigation menu includes 'Home', 'Current Visit', 'Elements visited', 'Rate', and 'Recommend'. A large image of the Valencia Opera House is shown. Below this, a section titled 'Plan recommended for john !' displays a list of activities with time slots and descriptions.

Time	Place
12:00 - 12:10	Go from Astoria Hotel to Lonja
12:10 - 13:10	Visit Lonja
13:10 - 14:10	Have Lunch
14:40 - 15:00	Go from Lonja to Valencia Port
15:00 - 16:30	Visit Valencia Port
16:30 - 16:45	Go from Valencia Port to Hemisferic
16:45 - 18:00	Visit Hemisferic

At the bottom of the page, there is a footer with links for 'E-turismo 2008 Privacy Policy | Terms of Use', 'Home', 'Showtimes', 'Events', 'Restaurants', and 'Theaters'.

Figura 3.9: Lista de actividades finales.

realizadas en el último programa turístico recomendado. La información obtenida de estas puntuaciones se utiliza para mejorar el proceso de recomendación y la ontología de los lugares turísticos de Valencia.

3.3.1. Representación del conocimiento en e-Turismo

En esta sección se describe la información particular del dominio (lugares turísticos de Valencia) requerido por e-Turismo para dar una recomendación.

E-Turismo usa una taxonomía para representar y manipular la información de las preferencias del usuario, y catalogar la lista de actividades a recomendar. En la siguiente subsección se define la taxonomía utilizada en e-Turismo.

Además de categorizar las actividades y a los usuarios, e-Turismo necesita más información que será utilizada por el subsistema de planificación, por ejemplo, la duración de las actividades o la hora de apertura de edificios. Esta información ayuda al subsistema de planificación a decidir el calendario de actividades de acuerdo con el tiempo disponible del usuario (ver sección 3.3.3). Además, se representa el mapa de la ciudad donde las actividades se llevarán a cabo, calculando la distancia entre



Figura 3.10: Lista de actividades confirmadas.

dos lugares dados (ver sección 3.3.3).

Una ventaja de tener conocimiento explícito acerca de las actividades que se realizan en una ciudad es que la recomendación puede realizarse más rápidamente. Por otro lado, la información puede quedarse obsoleta y necesitará ser actualizada periódicamente. Con este fin se utilizarían en un futuro procesos para adquirir nueva información de los servicios de Internet.

Taxonomía

El comportamiento de e-Turismo se basa en el uso de una taxonomía para representar las preferencias del usuario y los lugares ha recomendar. Las entidades en una taxonomía se organizan en una estructura jerárquica conectadas a través de una relación “es-una”, donde los niveles de clasificación más específicos corresponden a niveles inferiores. En la taxonomía de e-Turismo, las entidades representan **conceptos** comunes de un dominio de turismo, como son, los estilos arquitectónicos o tipos de edificios. Los nodos hojas de la taxonomía son actividades recomendadas a los usuarios, es decir, el concepto más específico.

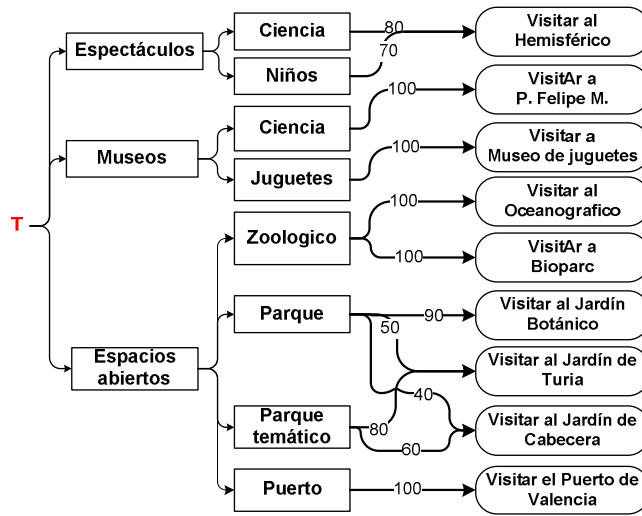


Figura 3.11: Taxonomía de e-Turismo.

Más formalmente:

Definición 1: La taxonomía T es un grafo (C, E) , donde C es el conjunto de nodos del grafo, los cuales representan conceptos de la taxonomía y E es el conjunto de flechas que conectan un concepto con sus sucesores. Se distinguen dos tipos de flechas: $E_{c-c} = (C_j, C_k)$, que vincula un concepto C_j con un concepto sucesor C_k ; y $E_{c-i} = (C_j, i, r_j^i)$, que vincula un concepto C_j con una actividad i a un grado de interés r_j^i .

La figura 3.11 muestra parte de la taxonomía de e-Turismo. Los nodos hojas del árbol son las actividades que se pueden realizar en la ciudad de Valencia, por ejemplo, visitar el lugar indicado por la actividad. Las flechas que conectan una actividad a un concepto están asociadas a un valor que indica el grado de interés de la actividad (un lugar en la taxonomía de e-Turismo) en relación al concepto, por ejemplo, en la figura 3.11 la actividad “Visitar el Puerto de Valencia” está asociada a un valor de interés del 100% al concepto “Puerto”, indicando que pertenece totalmente a dicho concepto. Una actividad puede estar clasificada en más de un concepto de la taxonomía. En la misma figura 3.11, la actividad “Visitar el Hemisférico” se encuentra asociada a dos conceptos “Ciencia” y “Niños” con un valor de interés de 80% y 70% respectivamente.

Como se indicó anteriormente, los valores de las flechas que conectan una actividad con su concepto más específico, muestran el grado de interés que tiene ese lugar. Por ejemplo, tanto la “Iglesia de San Juan” como la “Catedral de Valencia” están clasificadas como “Iglesias”, pero el interés de la visita a la “Catedral de Valencia” es mayor que a la “Iglesia de San Juan” (para un usuario interesado en “Iglesias”).

Cada nodo hoja representa un lugar a recomendar (ver figura 3.11), que en el caso particular de la taxonomía de e-Turismo, contiene el nombre, una breve descripción, una foto del lugar y su ubicación (dirección). También es necesario registrar los horarios de apertura de los lugares y edificios (que pueden ser diferentes dependiendo de la temporada o el día de la semana) y una duración estándar de cada actividad (que puede ser personalizada para un usuario específico). Esta información será utilizada por el PK (ver figura 3.12).

Visitar el Oceanográfico		
Abierto:	Julio, agosto: 10:00 – 24:00 Otros: 10:00 – 20:00	
Dirección:	Camino de las Moreras s/n. 46013 Valencia	
Duración Estandar:	Clasificación del usuario	Duración
	1	1h 30'
	2	2h 15m
	3	4h 15m
	4	6h 45m

Figura 3.12: Horario de actividad.

Además, una actividad o lugar $i \in I$ está definida por una lista preferencias F^i , que representan la entrada de cada flecha en los nodos hojas.

Por ejemplo, en la taxonomía de la figura 3.11, si $i =$ “Visitar el Jardín de Turia” entonces F^i podría ser fijado a $\{(\text{“Parque”}, 50), (\text{“Parque temático”}, 80)\}$. Por otra parte, si $i' =$ “Visitar el Jardín Botánico” y este jardín es considerado más interesante que el “Jardín de Turia” entonces $F^{i'}$, tendrá preferencias con mayor

grado de interés, es decir, {"Parque", 90}. El grado de interés puede ser actualizado dinámicamente a través de puntuaciones (**feedback**) de los usuarios.

Además, los lugares tienen asociado un valor numérico AC^i (contador de aceptación) este valor indica cuántas veces ha sido visitado este lugar, en el momento de ser recomendado (representa lo popular que es el lugar i para los usuarios).

Información acerca de la ciudad

Almacenamos el mapa de la ciudad que comprende todas las calles de la ciudad con los siguientes atributos: nombre, distrito, secciones, longitud de cada sección y las coordenadas geográficas. Por otra parte, también se representan las intersecciones de las diferentes secciones de cada calle. Esta información será utilizada por el PK para calcular las distancias entre las actividades.

Perfil del usuario

El perfil de un usuario u está definido en la base de datos de e-Turismo con los siguientes datos. En primer lugar, los datos personales y demográficos como edad, sexo, familia o país. En segundo lugar, las preferencias generales del usuario, denotadas por GP^u , contiene todas las preferencias en las que el usuario u está interesado. Más formalmente: $GP^u = \{ (c_n, r_n) : 1 \leq n \leq |C| \}$.

El perfil del usuario en e-Turismo también contiene información acerca de la interacción histórica del usuario con el SR, es decir, el conjunto de lugares que le han sido recomendados, ha visitado y ha puntuado y su grado de satisfacción con la recomendación.

Petición de Recomendación (Recommendation Query - RQ)

Cada vez que el usuario entra en el sistema para una nueva visita puede introducir unas preferencias específicas para la visita actual, si lo desea, que pueden ser diferentes a sus preferencias generales (y que se almacenaran dentro de una RQ). La RQ contiene el conjunto de preferencias específicas, formalmente denominado SP^u . SP^u se divide en preferencias de recomendación específicas $SRP^u = \{ (c_n, r_n) : 1 \leq n \leq |C| \}$ y preferencias de planificación específicas $SPP^u = \langle fecha, (Ts;$

T_e), dur_{comida} , dur_{cena} , $ubicación\ del\ usuario$ }, donde $fecha$ indica el día exacto en el que se realizará la visita, (T_s, T_e) representan el tiempo disponible del usuario, dur_{comida} y dur_{cena} representan el tiempo reservado para la comida y la cena, respectivamente (si el usuario quiere incluir las comidas), y $ubicación\ del\ usuario$ es la ubicación geográfica actual del usuario.

3.3.2. Núcleo Generalista del Sistema Recomendador (GRSK)

La tarea principal del GRSK es generar una lista de actividades a recomendar para el usuario. GRSK utiliza una *taxonomía* para generar la lista de ítems a recomendar. Se ha diseñado para ser generalista, es decir, independiente de la lista de ítems a recomendar. GRSK puede trabajar con cualquier dominio de aplicación siempre y cuando los datos del nuevo dominio se puedan representar en una *taxonomía* específica. GRSK puede utilizar cualquier tipo de ítems, en nuestro caso los ítems son actividades, que representan lugares turísticos.

GRSK utiliza una **técnica de recomendación híbrida mezclada** que combina las siguientes técnicas básicas de recomendación: demográfica, basada en contenido; y dos filtrados de información: basado en preferencias generales y basado en preferencias específicas (de la visita actual). Hemos definido un módulo independiente para cada técnica básica de recomendación y un módulo que controla la técnica de recomendación híbrida. GRSK ha sido diseñado para añadir fácilmente nuevas técnicas de recomendación básicas o híbridas simplemente con el desarrollo de nuevos módulos de recomendación. Las recomendaciones generadas por cada módulo básico son independientes de los demás módulos.

Definiciones básicas

Antes de explicar el comportamiento del GRSK, introduciremos algunos conceptos básicos que servirán de ayuda para futuras explicaciones. Un problema de recomendación puede formularse de la siguiente manera [1]. Sea U el conjunto de todos los usuarios e I el conjunto de todas las posibles actividades que pueden ser recomendadas. Entonces se tiene que, para cada usuario $u \in U$, se quiere elegir una actividad $i \in I$ que maximice la satisfacción del usuario. Con el fin de seleccionar un

elemento i , GRSK utiliza la taxonomía de las entidades para representar la clasificación de todas las actividades en I y los perfiles de los usuarios en U . En particular, se creó una estructura, llamada *preferencias*, para describir estos datos.

Definición 2: Sea T el conjunto de todos los conceptos en la taxonomía. Una preferencia es una pareja de la forma (f_n^i, r_n^i) , donde $f_n^i \in T$ es un concepto definido previamente en la taxonomía, y $r_n^i \in [k_1, k_2]$ es el grado de satisfacción de f_n a la actividad i donde $k_1, k_2 \in \mathbb{R}$.

Perfil del usuario en el GRSK

En los SR, un usuario $u \in U$ está definido por un perfil que incluye [63]: datos personales y demográficos del usuario, sus preferencias y un histórico de interacciones del usuario en el SR. GRSK define un perfil estándar para un determinado usuario u (ver sección 3.3.1).

Definición 3: Se denota RC^u a la lista de actividades recomendadas a los usuarios u . Se trata de un conjunto de tuplas de la forma (i, f_n^i) , donde $i \in I$ es una actividad recomendada y $f_n^i \in T$ es la preferencia en la que la actividad i ha sido recomendada.

Cuando se le recomienda una actividad al usuario, él puede realizar una acción para llevar a cabo dicha actividad. Por ejemplo, en el dominio de turismo, el usuario puede visitar un museo, o en una aplicación de recomendación de libros, el usuario puede comprar un libro. Cuando el usuario inicie sesión nuevamente en el sistema para una nueva consulta, se le pedirá que puntúe las actividades anteriormente recomendadas. El usuario debe especificar las actividades que ha visitado y el grado de satisfacción.

Definición 4: Dada una actividad recomendada i tal que $(i, f_n^i) \in RC^u$, se dice que ha sido aceptada por el usuario cuando el usuario indica que ha realizado una acción para llevar a cabo dicha actividad e informa su grado de satisfacción. Una actividad aceptada es descrita por la tupla (i, f_n^i, ur^i) , donde $ur^i \in [k_1, k_2]$ es la calificación dada por el usuario u a la actividad i . Se denota A^u al conjunto de actividades aceptadas por el usuario u .

Restricciones

Los módulos del GRSK no trabajan directamente con las actividades especificadas en I . En lugar de ello, a partir de las preferencias generales y específicas del usuario, se infiere un conjunto de restricciones, que se utilizarán para seleccionar las actividades recomendadas al usuario. Más formalmente:

Definición 5: Una restricción c es una preferencia de la forma (f_n^c, r_n^c) . GRSK gestiona dos tipos de restricciones: CP^u , las restricciones positivas, que denotan conceptos que deben ser cumplidos por las actividades recomendadas, y CN^u , las restricciones negativas, que denotan conceptos que no deben ser cumplidos por las actividades recomendadas.

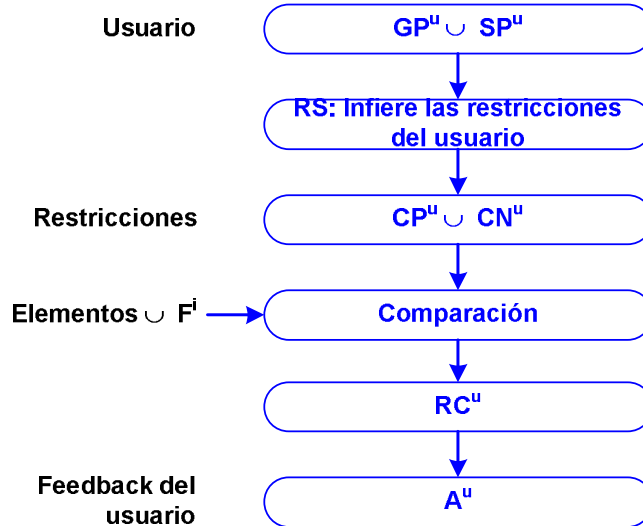


Figura 3.13: Relaciones del GRSK

El comportamiento del GRSK se puede resumir fácilmente por medio de relaciones entre las definiciones anteriormente declaradas (ver figura 3.13). Un usuario se describe mediante unos datos personales y una serie de preferencias (generales GP^u y específicas SP^u) que serán utilizadas por el SR para inferir un conjunto de restricciones positivas CP^u y negativas CN^u . Las actividades que coincidan con estas restricciones (se clasifican en la ontología) serán recomendadas (RC^u) y, por consiguiente, podrán ser aceptadas por el usuario (A^u).

Arquitectura del GRSK

La figura 3.14 muestra la arquitectura del GRSK. Los principales módulos son:

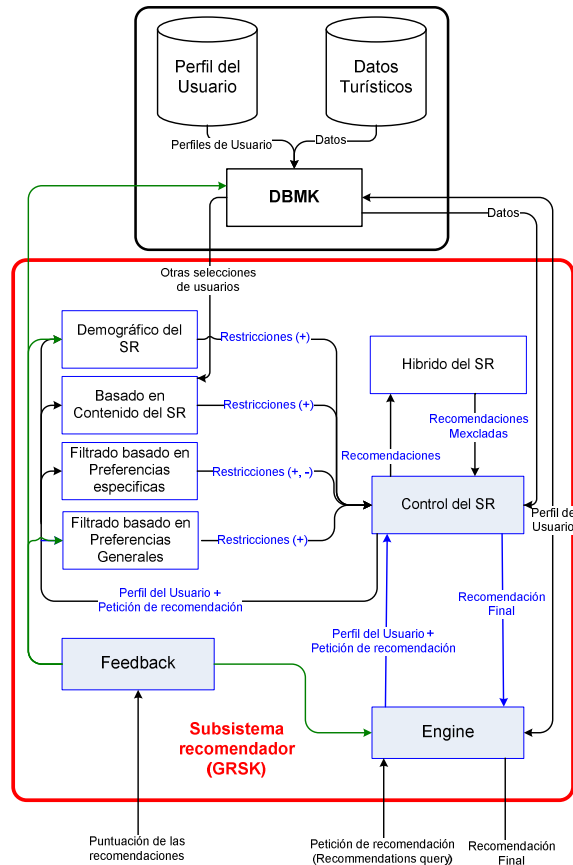


Figura 3.14: Arquitectura del GRSK

- El **Engine** es el módulo principal del GRSK. Es el encargado de gestionar la RQ y de actualizar los perfiles de los usuarios.
- El módulo de **Control del SR** administra todo lo referente al proceso de recomendación. Su primera tarea es calcular el conjunto de restricciones para cada módulo de recomendación. En este momento, GRSK contiene cuatro módulos básicos de recomendación: SR demográfico, SR basado en contenidos y los filtrados de información: basados en preferencias generales y preferencias

específicas. Se han optado estas técnicas, por que juntas son capaces de producir recomendaciones sin importar la cantidad de información que se tenga del usuario; pero como se explicó anteriormente, nuevas técnicas de recomendación (ó módulos) pueden ser agregados fácilmente a GRSK cuando sea necesario aplicarlo a otros dominios. Cada uno de los módulos del SR puede obtener un conjunto diferente de restricciones o un grado diferente de adecuación para la misma *preferencia*.

- El módulo de **retroalimentación del usuario (feedback)** se encarga de procesar el **feedback** del usuario acerca de las recomendaciones dadas, con el fin de mejorar: su perfil, la clasificación de los lugares y la información asociada a las restricciones calculadas en los módulos del SR.

Módulos de recomendación

El **módulo de recomendación demográfica** se basa en la **técnica de recomendación demográfica (TRD)** que clasifica al usuario u en una categoría demográfica de acuerdo con las características de su perfil y de sus preferencias generales (GP^u). Esta técnica asocia una lista de preferencias a un tipo de usuario. Por ejemplo, una persona con niños estará clasificada en una categoría diferente que un jubilado, porque tendrán preferencias diferentes. Esta clasificación implica asignar un conjunto de restricciones positivas a u : $CP^u = (f_n^i, r_n^i) : 1 \leq n \leq |T|$. Por ejemplo, si un usuario está clasificado como una “Persona con niños”, el sistema considera, entre otras cosas, las siguientes características para realizar la recomendación de las actividades: $\{(\text{“Zoológico-Acuario”}, 100), (\text{“Parques temáticos”}, 80), \text{etc.}\}$. Optamos por una TRD, ya que es capaz de proporcionar una recomendación cuando el usuario es nuevo en el sistema, evitando el problema del “Nuevo Usuario”. Además, puede recomendar actividades de categorías diferentes de otras previamente recomendadas (introduciendo prioridad en la recomendación).

El **módulo de recomendación basado en contenido** utiliza la **técnica basada en contenido (TBC)**. Recomienda una serie de actividades teniendo en cuenta las preferencias de las actividades puntuadas positivamente por el usuario en el pasado. Recomienda actividades similares a las que el usuario ya ha visitado. El objetivo de utilizar esta técnica de recomendación es aumentar la satisfacción de los usuarios

recomendándoles actividades que se asumen que le son de interés. Por ejemplo, si el usuario ha visitado y puntuado positivamente “Jardines”, el sistema incluiría en la recomendación otros “Jardines” que se encuentren en la ciudad de Valencia. La TBC calcula un conjunto de restricciones positivas CP^u en la forma habitual (f_n^i, r_n^i) . r_n^c que se calcula como:

$$r_n^c = \frac{\sum_{\forall i \in A^u} ur^i * r_n^i}{|A^u|}$$

Recordemos que r_n^i es el grado de adecuación del concepto f_n^i para la actividad i , es decir, $(f_n^i, r_n^i) \in F^i$.

El **módulo de filtrado de preferencias generales** utiliza la **técnica de filtrado de información de preferencias generales del usuario** (TGP^u), es una técnica de filtrado de información que trabaja con las preferencias generales (GP^u), especificadas por el usuario en su perfil. En este módulo se calculan una serie de restricciones, CP^u , de la siguiente manera:

$$CP^u = \{(f_n^g, r_n^g) \in GP^u\}$$

El **módulo de filtrado de preferencias específicas** está asociado a la **técnica de filtrado de información de las preferencias específicas del usuario** (TSP^u), la cual es una técnica de filtrado de información que trabaja con las preferencias específicas del usuario (SP^u). Básicamente, se analizan y almacenan las SP^u que pueden diferir de las GP^u . Esta técnica permite al usuario definir algunas preferencias diferentes a sus GP^u sin modificarlas. Por ejemplo, un usuario ha definido la “Arquitectura Gotica” como una SP^u para la visita actual y, por esta razón, el sistema recomienda visitar la “Torre del Miguelete” y el “Palacio de los Condes de Alpuente” (ver figura 3.8). En este módulo se calculan dos tipos de restricciones, CP^u y CN^u , de la siguiente manera:

$$CP^u = \{(f_n^s, r_n^s) \in SP^u : (\nexists (f_n^g, r_n^g) \in GP^u / f_n^g = f_n^s) \vee (\exists (f_n^g, r_n^g) \in GP^u / f_n^g = f_n^s \wedge r_n^s \geq r_n^g)\}$$

$$CN^u = \{(f_n^s, r_n^s) \in SP^u \mid \exists (f_n^g, r_n^g) \in GP^u / r_n^s < r_n^g\}$$

Una vez que cada SR ha calculado las correspondientes restricciones (CP^u y CN^u), el **módulo de control del SR** convierte las restricciones en actividades a recomendar (accediendo a la base de datos a través del DBMK), además el conjunto de preferencias negativas permiten eliminar actividades recomendadas que el usuario no desea que se le recomienden en la visita actual. Una actividad i coincide con las restricciones CP^u y CN^u si satisface una restricción positiva, no satisface ninguna restricción negativa y no ha sido puntuada por el usuario, es decir, si cumple las tres condiciones siguientes:

1. $\exists(f_n^i, r_n^i) \in F^i, \exists(f_n^c, r_n^c) \in CP^u / f_n^c = f_n^i$
2. $\forall(f_n^i, r_n^i) \in F^i, \neg \exists(f_n^c, r_n^c) \in CN^u / f_n^c = f_n^i$
3. $\neg \exists(i', f_n^{i'}, ur^{i'}) \in A^u / i = i'$

Estas listas de actividades son luego procesadas por el **módulo híbrido del SR**, que aplica la **técnica de recomendación híbrida mezclada (TRHM)**. En primer lugar, calcula la prioridad de cada uno de los elementos de las listas. Dada una actividad i y una restricción positiva en la cual ha sido seleccionada (f_n^c, r_n^c), la prioridad Pr^i se calcula de la siguiente manera:

$$Pr^i = \frac{AC^i}{\sum_{\forall i \in I} AC^i} * (\kappa_2 - \kappa_1) + r_{SR}^i + r_{taxonomia}^i$$

Donde AC^i es el contador de aceptación de la actividad i , r_{SR}^i es el grado de interés de la actividad i obtenidos a partir de la TR y $r_{taxonomia}^i$ es el grado de interés de la actividad i bajo el concepto de la taxonomía.

La TRHM combina los elementos de cada TR para obtener una lista única de elementos recomendados, que se ordena de acuerdo a la prioridad calculada (de mayor o menor). En caso de que un elemento aparezca en más de una lista (es decir, que ha sido seleccionado por varias TR), sólo se considera la aparición con la prioridad más alta y las demás las descarta. La TRHM selecciona las N mejores recomendaciones, que son el conjunto de elementos recomendados al usuario u (RC^u). Cada actividad recomendada en RC^u es representada por una tupla de la forma $[i, Pr^i]$.

El **módulo feedback** permite la retroalimentación del sistema. Cuando a un usuario se le han recomendado una lista de actividades, e inicia una nueva sesión en

el sistema, este módulo solicita al usuario que puntúe la última lista de recomendaciones, indicando si visitó o no el lugar y la puntuación que le otorga. Con estos valores se realimenta el sistema. Esta estrategia se denomina *ratings-based* [20] (ver sección 2.5). La información obtenida de estas puntuaciones, denominadas A^u , son luego procesadas. Los lugares rechazados no son considerados en este momento.

Como se explicará más adelante, la adecuación del perfil del usuario y de cada lugar puntuado se hace cuando un usuario puntúa un lugar, mientras que la adecuación del conjunto de lugares (I) y las restricciones se lleva a cabo periódicamente, porque es necesario tener en cuenta, en la retroalimentación, información de muchos usuarios.

Adecuación del perfil de usuario

La adecuación del perfil de usuario consiste en analizar si una preferencia general GP^u debe ser ajustada o no y en qué grado. Por ejemplo, si el perfil de un usuario u indica que le gusta visitar “Iglesias”, pero cuando el sistema le recomienda visitar la “Catedral de Valencia”, el usuario le asigna una puntuación baja a esta actividad, entonces el grado de adecuación de “Iglesias” debería ser disminuido en el perfil del usuario.

Más formalmente, sea $\delta_{up} \in \mathfrak{R}$ un umbral definido en GRSK, y que es establecido cuando hay una diferencia significativa entre la calificación de un usuario y el grado de adecuación de la correspondiente preferencia en su perfil. Dada una actividad i puntuada por el usuario u tal que $(i, f_n^i, ur^i) \in A^u$ y la correspondiente preferencia general del usuario $(f_n^g, r_n^g) \in GP^u$ tal que $f_n^i = f_n^g$, el nuevo valor de r_n^g será calculado de la siguiente manera:

$$r_n^g = \begin{cases} \max(r_n^g + \varepsilon_{up}, \kappa_2) & \text{si } ur^i > r_n^g + \delta_{up} \\ \min(r_n^g - \varepsilon_{up}, \kappa_1) & \text{si } ur^i + \delta_{up} < r_n^g \\ r_n^g & \text{Otro caso} \end{cases}$$

donde ε_{up} es el que determina el incremento o decremento del grado de adecuación de una función y se encuentra definido en GRSK.

Actualización de la lista de actividades

La retroalimentación del usuario es también utilizada para adecuar los lugares en I . En primer lugar, el contador de aceptación de un lugar i (AC^i) es incrementado

cada vez que un usuario puntúa un lugar i . Por otro lado, la retroalimentación del usuario es utilizada para ajustar las preferencias asignadas a un lugar. Por ejemplo, supongamos una actividad i asociada a dos conceptos f_n^i y f_m^i ; la actividad i obtiene altas puntuaciones cuando es recomendada a causa del concepto f_n^i y obtiene bajas puntuaciones cuando la recomendación se basa en f_m^i . En este caso, el grado de adecuación (r_m^i) de f_m^i debe ser disminuido. Esta actualización no sólo se basa en la opinión de un único usuario, sino que también tiene en cuenta el conjunto de los usuarios U .

Dado un lugar i y una preferencia (f_n^i, r_n^i) , R_n^i será definido como el conjunto de todas las puntuaciones de i cuando éste ha sido seleccionado atendiendo al concepto f_n^i , es decir, $R_n^i = \{(i, f_n^i, ur^i) \in \bigcap_{\forall u \in U} A^u\}$. Entonces, la media de las puntuaciones del usuario en R_n^i es: $avg_n^i = \frac{\sum ur^i}{|R_n^i|}$. El nuevo valor para r_n^i se calcula de la siguiente manera:

$$r_n^i = \begin{cases} \text{máx}(r_n^i + \varepsilon_{up}, \kappa_2) & \text{si } avg_n^i > r_n^i + \delta_{up} \\ \text{mín}(r_n^i - \varepsilon_{up}, \kappa_1) & \text{si } avg_n^i + \delta_{up} < r_n^i \\ r_n^i & \text{Otro caso} \end{cases}$$

Adecuación de las restricciones

Cuando se le asigna una puntuación baja a un lugar recomendado, esto puede indicar que las GP^u deben ser actualizadas o que este lugar no está bien clasificado en la taxonomía. Sin embargo, también podría revelar que el proceso de recomendación no es lo suficientemente preciso.

La TBC y el filtrado de información basado en preferencias se benefician (indirectamente) de la adecuación del perfil de usuario y la actualización de la lista de actividades, ya que utilizan esta información (actualizada) para sus procesos. Sin embargo, el TRD merece un análisis particular.

La TRD trabaja con un conjunto de categorías que definen el conjunto de restricciones que corresponden a cada tipo de usuario. Este conjunto de restricciones pueden ser actualizadas dinámicamente a través del **feedback** realizado por los usuarios. Por ejemplo, si la categoría demográfica “Personas con niños” ha sido inicialmente asignada a la preferencia “Artes”, pero muchos usuarios en esta categoría

dan una baja puntuación a las actividades que coincide con esta preferencia, entonces el grado de adecuación de “Artes” debe disminuir en esta categoría. Esto se lleva a cabo por un método similar al utilizado para actualizar la lista de actividades.

3.3.3. Núcleo Planificador (PK)

Una vez que GRSK ha calculado el conjunto de actividades recomendadas RC^u , éstas se le muestran al usuario y él debe seleccionar o rechazar las actividades que le interesen. Las actividades no seleccionadas son consideradas como indiferentes. Esta lista de actividades filtradas por el usuario se envían al módulo PK junto con sus preferencias específicas de planificación. El módulo de planificación analiza esta información y construye un problema de planificación adaptado al usuario cuya solución será la agenda turística (ver figura 3.9).

Construcción del problema de planificación

El módulo de planificación gestiona tres grupos de datos diferentes:

(i) Las preferencias específicas de planificación del usuario, en adelante SPP^u (véase el módulo de filtrado de preferencias específicas en la sección 3.3.2).

(ii) El filtrado de actividades recomendadas, en adelante FRC^u , que es una lista de tuplas de la forma $\langle i, Pr^i, si^i \rangle$, donde Pr^i es la prioridad calculada por GRSK para una actividad i y si^i es un valor en el conjunto {seleccionado; rechazado; no seleccionado} que indica si el usuario ha seleccionado la actividad i , la ha rechazado o no tiene preferencia por ella (cuando no las selecciona).

(iii) La información sobre cada actividad i en FRC^u , es una tupla de la forma $\langle i, abierto_i, cerrado_i, localizacion_i \rangle$. $abierto_i$ y $cerrado_i$ indican el horario de apertura y cierre, respectivamente, de una actividad i (teniendo en cuenta la fecha de la visita) y $localizacion_i$ es la dirección del lugar donde la actividad se lleva a cabo. Estos valores se extraen de la información del lugar (ver figura 3.12).

Todos estos datos son analizados y adecuadamente combinados para construir el problema de planificación adaptado al usuario. Es importante tener en cuenta que no todas las actividades en FRC^u serán incluidas en el plan, ya que la planificación depende de la disponibilidad del usuario, sus restricciones temporales y restricciones de tiempo en el entorno (es decir, el horario de atención de los lugares). Por lo

tanto, el módulo de planificación debe seleccionar las actividades a incluir en el plan. Asimismo, se planea a futuro la posibilidad de que el usuario pueda cambiar el punto de partida en su visita y pasar de un lugar inicial a otro, calculando de manera automática el tiempo y la distancia entre las nuevas actividades.

El resultado del problema de planificación se puede especificar [79] como $\langle acciones, inicio, objetivo \rangle$. *acciones* es el conjunto de todas las posibles acciones en el dominio, y en nuestro caso contiene dos conjuntos diferentes: el conjunto A de actividades seleccionadas o no seleccionadas y el conjunto M de acciones de movimiento para pasar de una actividad a otra. Cada acción $a \in acciones$ está definida por $a = (dur^a, u^a)$. Para cada actividad $A_j \in A$, la duración dur^{A_j} es un valor obtenido según el tipo de usuario. Es decir, cada usuario es clasificado en una categoría diferente que determina la duración de cada actividad. Por ejemplo, la figura 3.12 muestra cuatro duraciones diferentes, una para cada tipo de usuario. Por otra parte, la utilidad de u^{A_j} de una actividad A_j es un valor que se calcula de la siguiente manera:

$$u^{A_j} = \begin{cases} Pr^{A_j} & \text{if } si^{A_j} = \textit{seleccionado} \\ Pr^{A_j} * \alpha & \text{if } si^{A_j} = \textit{indiferente} \end{cases}$$

donde $\alpha \in [0, 1]$ es un parámetro para ponderar la importancia relativa de las diferentes actividades.

Adicionalmente, añadimos al conjunto A tres acciones básicas:

(i) $A_0 = (0, \infty)$ para representar la acción inicial “comenzar en la ubicación inicial”.

(ii) $A_{|A|+1} = (dur_{comida}, u_{comida})$ y $A_{|A|+2} = (dur_{cena}, u_{cena})$, para representar las acciones “comer” y “cenar”. Si el usuario quiere incluir las comidas en el plan, u_{comida} y u_{cena} se incluyen en ∞ y la duración de ambas acciones se especifican en el SPP^u . En caso contrario, la utilidad de estas acciones se establece en 0. Además, construimos una tupla de la forma $\langle comida, abierto_{comida}, cerrado_{comida}, localizacion_{comida} \rangle$, donde $abierto_{comida}$, $cerrado_{comida}$ son establecidas como el típico inicio / fin de la hora de comida en la ciudad para visitar y, en aras de la simplicidad, consideramos que la ubicación de estas actividades coinciden con la ubicación de la última actividad realizada antes de la comida o la cena.

M es el conjunto de acciones de movimiento, donde cada $m_{j,k} \in M$ representa la acción de trasladarse desde el lugar donde la actividad A_j se realiza, al lugar donde

la actividad A_k es realizada. La duración $dur^{m_{j,k}}$ de una acción $m_{j,k}$ es calculada teniendo en cuenta la distancia entre las actividades A_j y A_k . Para calcular esta distancia, mantenemos en la base de datos el mapa de la ciudad que comprende todas las calles y las intersecciones de las diferentes secciones de cada calle, que dará el camino para ir de A_j a A_k . La utilidad $u^{m_{j,k}}$ es tratada como una penalización, para evitar emplear mucho tiempo en movimientos no deseados y se calcula como $u^{m_{j,k}} = -dur^{m_{j,k}} * \beta$, donde $\beta \in \Re$ es un valor calculado de acuerdo a la utilidad de las actividades seleccionadas.

inicio es el estado inicial del problema de planificación, que contiene la disponibilidad de tiempo del usuario (Ts, Te) y la apertura en horas de cada actividad ($abierto_{A_j}, cerrado_{A_j}$), mientras que el *objetivo* principal es optimizar las funciones de optimización del plan y del tiempo.

Resolución del problema de planificación

Dada las preferencias de planificación del usuario, un conjunto A de todas las actividades que pueden ser realizadas (incluidas las tres acciones básicas) y el conjunto M de acciones de movimientos, formulamos el resultado del problema de planificación como un *Problema de Satisfacción de Restricciones* (CSP) con las siguientes variables:

- $\forall j, k \in [0, |A|] : a_j, m_{j,k} \in [0, 1]$ indica cualquier actividad A_j (resp. $m_{j,k}$) que es aceptada en el plan. $(|A| + 1)^2 - |A|$ son las variables que se necesitan para representar todos los posibles movimientos.
- $\forall j \in [1, |A|] : ts_j, te_j \in [abierto_j, cerrado_j]$, que denotan la hora de comienzo y fin de la actividad j . El dominio de esas variables está determinado por los horarios de atención de los diferentes lugares.

A modo de ejemplo, consideramos las actividades de la tabla 3.1 como aquellas seleccionadas por el usuario. Además, resume la información necesaria para cada actividad seleccionada o no seleccionada y para cada acción básica. Por lo tanto, el correspondiente CSP necesitará las variables de a0 a a7 y 56 variables para representar los movimientos entre los lugares correspondientes.

Las restricciones del CSP son las siguientes:

	Actividad	prioridad	apertura horas	dur	utilidad ($\alpha = 0.5$)
0	Localización inicial	∞	-	0'	∞
1	El Oceanográfico	270	10-24	4h	270
2	Puerto de Valencia	240	8-24	1h30'	240
3	El Hemisférico	225	16-21	1h15'	225
4	Lonja	214	10-15	1h	107
5	BioParc	210	10-21	2h15'	105
6	Comidad	∞	13-15	1h30'	∞
7	Cena	0	0-24	0'	0

Tabla 3.1: Actividades en nuestro problema de planificación.

El tiempo de inicio y final para cada actividad debe coincidir con la duración especificada:

$$te_j = ts_j + dur_j \quad \forall j \in [1, |A|]$$

Por ejemplo, la restricción $te_1 = ts_1 + 4$ establece la hora de inicio y finalización de la actividad 1.

Las actividades y los movimientos en el plan no deben colapsar. Por esta razón, definimos las siguientes restricciones: la *primera restricción* dice, si la actividad i se realiza antes que la actividad j , la hora de inicio de j debe ser superior a la hora de finalización de i más el tiempo necesario para moverse de i a j ; la *segunda restricción* establece la misma restricción en el caso que la actividad j se realice antes de la actividad i , y la *última restricción* indica que ambas situaciones no pueden ocurrir en el mismo plan.

$$\forall i, j \in [1, |A|] / i < j:$$

$$\begin{cases} te_i + dur_{i,j} \leq ts_j + (1 - m_{i,j}) * 24 \\ te_j + dur_{j,i} \leq ts_i + (1 - m_{j,i}) * 24 \\ m_{i,j} + m_{j,i} \leq 1 \end{cases}$$

Por ejemplo, las siguientes restricciones indican que las actividades 1 y 2 no pueden colapsarse, asumiendo que se necesiten 15 minutos para ir del Oceanográfico al Puerto de Valencia:

$$\begin{cases} te_1 + 15 \leq ts_2 + (1 - m_{1,2}) * 24 \\ te_2 + 15 \leq ts_1 + (1 - m_{2,1}) * 24 \\ m_{1,2} + m_{2,1} \leq 1 \end{cases}$$

La duración total de las actividades y los movimientos no pueden exceder la disponibilidad del usuario:

$$Te - Ts \geq \sum_{\forall j \in [1, |A|]} a_j * dur_j + \sum_{\forall j, k \in [1, |A|]} m_{j,k} * dur_{j,k}$$

Con el fin de dar al usuario la oportunidad de elegir el plan que prefiere, consideramos dos funciones de optimización: la primera maximiza la utilidad de todo el plan, mientras que la segunda está dirigida a maximizar el tiempo empleado en las actividades preferidas. Estas funciones de optimización se definen como:

$$\begin{aligned} \text{máx } U &= \sum_{\forall i \in [0, |A|]} a_i * u_i + \sum_{\forall i, j \in [0, |A|]} m_{i,j} * u_{i,j} \\ \text{máx } T &= \sum_{\forall i \in [0, |A|]} a_i * dur_i * u_i - \sum_{\forall i, j \in [0, |A|]} m_{i,j} * dur_{i,j} \end{aligned}$$

En nuestro caso, cuando el CSP se resuelve con la función de optimización de U , obtenemos un plan con más actividades porque asigna un valor de utilidad más alto. En este caso, el plan consiste en $a_0 \rightarrow a_4 \rightarrow a_2 \rightarrow a_3$. La agenda correspondiente se muestra en la figura 3.15. Por otra parte, cuando se utiliza la función de optimización T , el plan contiene un menor número de acciones porque se prefieren las acciones de mayor duración. En este caso, el plan consiste en $a_0 \rightarrow a_1$.

Agenda turística

Cuando el sistema resuelve estos dos CSPs, obtiene (al menos) dos planes que contienen un subconjunto de las actividades FRC^u junto con el horario de visita de las actividades. Este plan se muestra como una agenda turística. Un plan está definido como una tupla de la forma: $\langle fecha, (a, ts_a, te_a)^* \rangle$, donde $fecha$ es la fecha exacta de la visita, a se refiere a la actividad programada, y ts_a y te_a son la hora de inicio y finalización de cada actividad, respectivamente.

The screenshot shows the E-Tourism website interface. At the top, there is a logo for 'E-Tourism' with the tagline 'Recommendations of tourist sites'. Navigation links include 'Home', 'Site Map', and 'Search'. A search bar is present with a 'Go!' button. The user is logged in as 'john', with options for 'User Profile' and 'Logout'. A navigation menu includes 'Home', 'Current Visit', 'Elements visited', 'Rate', and 'Recommend'. A large image of the Valencia Convention Center is displayed. Below it, a section titled 'Plan recommended for john !' lists a schedule of activities:

Time	Place
12:00 - 12:10	Go from Astoria Hotel to Lonja
12:10 - 13:10	Visit Lonja
13:10 - 14:10	Have Lunch
14:40 - 15:00	Go from Lonja to Valencia Port
15:00 - 16:30	Visit Valencia Port
16:30 - 16:45	Go from Valencia Port to Hemisferic
16:45 - 18:00	Visit Hemisferic

At the bottom, there is a footer with 'E-turismo 2008 Privacy Policy | Terms of Use' and navigation links for 'Home', 'Showtimes', 'Events', 'Restaurants', and 'Theaters'.

Figura 3.15: Agenda turística.

3.3.4. Núcleo Gestor de la Base de Datos (DBMK)

E-Turismo cuenta con un módulo (DBMK) que sirve de intermediario entre la Base de datos y los subsistemas GRSK y PK. DBMK está diseñado para ser utilizado con cualquier Sistema Gestor de Base de Datos (SGBD). Está compuesto por tres módulos: Conexión, solicitud de datos y solicitud de peticiones (RTDatabase). El módulo de **Conexión** se encarga de conectarse con la Base de datos (BD). Se deben configurar previamente los datos de acceso. El módulo de **solicitud de datos** realiza las peticiones que solicita RTDatabase, obteniendo una matriz con los datos solicitados. Este módulo también puede funcionar con procedimientos almacenados. Los procedimientos almacenados son pequeños programas guardados y ejecutados físicamente en la BD, que se utilizan para simplificar la creación y el mantenimiento de los programas involucrados. Tienen como ventaja que disminuyen la sobrecarga resultante de comunicar grandes cantidades de datos. El módulo de solicitud de peticiones solicita las consultas al módulo de solicitud de datos y el resultado obtenido lo envía a los subsistemas anteriormente mencionados (ver figura 3.16). Se utilizó este sistema modular para facilitar futuras correcciones en el código.

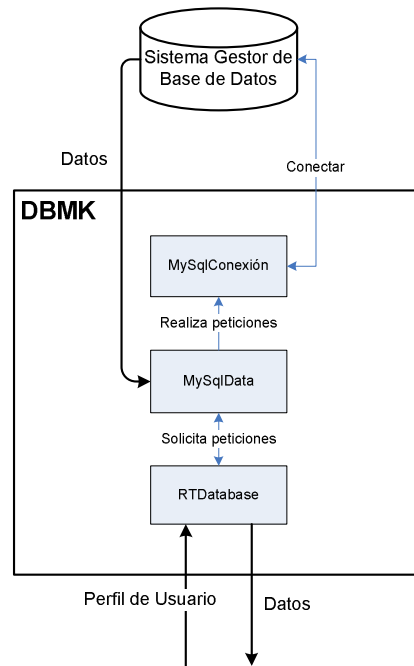


Figura 3.16: Núcleo Gestor de la Base de datos.

Los datos que recibe DBMK están organizados de la siguiente manera: $\langle \text{Tabla}, \text{Datos}, \text{Campos}, \text{Acción}, \text{Valores} \rangle$, donde $\text{Tabla}(T)$, $\text{Datos}(D)$, $\text{Campos}(C)$ y $\text{Valores}(V) \in \text{Esquema de la Base de Datos (EBD)}$ y dependen de la $\text{Acción}(Ac) \in \text{String condicional}$, que se lleve a cabo.

El String condicional, especifica que tipo de consulta se desea hacer (“select”, “insert”, “update”, “delete”). DBMK, dependiendo del Ac , ejecuta una acción contra el SGBD; el resultado de esta acción se retorna al subsistema que la haya solicitado.

Más formalmente, el sistema viene representado de la siguiente manera:

$$DBMK = \begin{cases} \langle T, C, D, V \rangle & \text{si } Ac = \text{"select"} \\ \langle T, C = D, V \rangle & \text{si } Ac = \text{"insert"} \\ \langle T, C, D, V \rangle & \text{si } Ac = \text{"update"} \\ \langle T, C, V \rangle & \text{si } Ac = \text{"delete"} \end{cases}$$

Donde T representa la tabla en la que se realizan las acciones, C y D representan los campos en la BD sobre los cuales se realizan las condiciones y consultas y V es

una tupla con los valores de los campos cuando se realiza una Acción condicional.

DBMK está diseñado para ser flexible en cuanto al esquema de tablas utilizado y principalmente para liberar al desarrollador del manejo manual de los datos que resultan de la ejecución de sentencias SQL, manteniendo la portabilidad entre los diferentes SGBD con un pequeño incremento en el tiempo de ejecución. Además permite independizar GRSK del dominio de e-Turismo.

3.4. Conclusiones

Hoy en día, más y más personas utilizan en sus viajes los Servicios Web para obtener información de sus vuelos y lugares a visitar. Estos sistemas ofrecen al usuario gran cantidad de información, que, a menudo no se ajusta a sus preferencias. Por ello estos sistemas han ido incorporando SR que acotan la información presentada al usuario e intentan que ésta sea lo más adecuada posible al usuario. Sin embargo estos servicios se centran en la planificación de viajes y no ofrecen un plan turístico para visitar lugares de una ciudad.

E-Turismo es un Servicio Web que genera recomendaciones personalizadas de planes turísticos en la ciudad de Valencia (España). Se pretende que sea un servicio para turistas locales, que ayude a encontrar las mejores actividades para realizar en la ciudad de Valencia. E-Turismo realiza recomendaciones de acuerdo a las preferencias del usuario, su clasificación demográfica, sus lugares visitados, las preferencias de su visita actual y los datos obtenidos de otros usuarios. El SR muestra al usuario un plan turístico con las actividades recomendadas, que reflejan las preferencias del usuario. En la elaboración de este plant turístico de las actividades recomendadas, el sistema tiene en cuenta la distancia geográfica entre dichas actividades, los horarios de apertura y cierre, etc, para ajustar el plan al tiempo disponible del usuario.

El núcleo de e-Turismo, GRSK, fue diseñado para ser independiente de los otros módulos, los cuales dependen del dominio. GRSK es un núcleo de recomendación híbrido y generalista que puede ser usado en cualquier dominio de aplicación que puede ser definido a través de una taxonomía. Adicionalmente, GRSK tiene un diseño modular que lo hace independiente de las técnicas de recomendación básicas o híbridas utilizadas, permitiendo la incorporación de nuevas técnicas de recomendación fácilmente.

Actualmente, GRSK combina técnicas de recomendación demográficas y basadas en contenido así como filtrados por preferencias generales y específicas que, como se mostrará en el siguiente capítulo, han permitido obtener recomendaciones de buena calidad.

Un elemento clave en e-Turismo es la retroalimentación por parte del usuario o *Feedback*. Este proceso es necesario para adquirir información del usuario aprendiendo más sobre sus gustos y preferencias y además, esta información permite ajustar de manera automática el núcleo GRSK para futuras recomendaciones.

Capítulo 4

Evaluación y Resultados

4.1. Introducción

Parte del proceso de personalización consiste en conocer si las recomendaciones se ajustan a los gustos del usuario. Este capítulo expone los experimentos realizados para evaluar el comportamiento del GRSK, ya que esta tesina se centra fundamentalmente en el núcleo de recomendación y por ello es el objetivo de la evaluación.

Para evaluar un SR [75] existen varias métricas, entre las más importantes tenemos las siguientes [37, 56, 92]:

- **Ratio de conversión** (conversion rate) [69]: Mide el porcentaje de usuarios que realmente llevan a cabo la recomendación ofrecida (por ejemplo: compran el producto).
- **Cobertura** (coverage) [37]: Mide el porcentaje de ítems para los cuales el SR es capaz de realizar predicciones.
- **Precisión** (user satisfaction) [14, 62]: Es una de las métricas más extendidas. Refleja el grado de satisfacción o insatisfacción del usuario con un elemento recomendado, es decir, mide el porcentaje de ítems recomendados que son de alta relevancia para el usuario. Esta métrica no es capaz de medir aspectos complejos de la personalización. Y por ello, se han desarrollado otras métricas [64] (consumer lifetime value, loyalty value, purchase and consumption experience, return on customer). Si se detecta que la recomendación no está funcionando

bien, hay que entender las causas (colección pobre de datos de usuario, perfiles de usuario poco precisos, técnicas de matching inapropiadas o entrega de contenido poco efectiva). Estos problemas se detectan y corrigen mediante un mecanismo de feedback [1].

Básicamente, existen tres medidas de precisión [27]:

- Medidas de **precisión de la predicción**, que miden la diferencia entre el valor del rating del usuario y el valor del rating que predijo el sistema para dicho usuario:

$$\text{Mean Absolute Error} = \sum_{i=1}^N |p_i - r_i| / N$$

Donde p_i es el rating calculado y r_i es el rating del usuario para el ítem i , dados N ítems.

- Medidas de **precisión de la clasificación**, que calculan unos ratios entre los ítems relevantes que han sido seleccionados o no. En este caso, no se calcula un rating sino que se calcula si un ítem será relevante o no para el usuario. Utiliza dos medidas conjuntamente [69]:

- Precisión: Número de ítems relevantes recomendados respecto al número total de ítems recomendados.

$$\text{Precisión} = N_{rs} / N_s$$

- Recall: Número de ítems relevantes recomendados respecto al número de ítems relevantes totales.

$$\text{Recall} = N_{rs} / N_r$$

Donde N_r es el número de ítems que el usuario ha clasificado como relevantes, N_{rs} es el número de ítems relevantes seleccionados por el sistema y N_s es el número total de ítems seleccionados por el sistema.

- Medidas de **precisión del orden de selección**, que miden las diferencias entre el ranking de los ítems que realiza el usuario y el calculado por el sistema.

Para evaluar el comportamiento del GRSK, se escogió, en primer lugar, una métrica de precisión, ya que ésta proporciona mayor información acerca de la satisfacción del usuario sobre la recomendación ofrecida. Entre las métricas de precisión, se escogió la métrica de **precisión de la clasificación**, que como se mencionó anteriormente, consiste en el cálculo de unos ratios entre los lugares relevantes que han sido seleccionados o no por el usuario. Esto es debido a que GRSK no predice un rating para un lugar, sino que decide si un determinado ítem será relevante o no para el usuario, teniendo en cuenta sus datos demográficos, preferencias generales y datos históricos. Y si lo considera relevante, lo recomendará.

Dicha evaluación puede realizarse con datos generados aleatoriamente o con datos reales. En [27] se explican las ventajas e inconvenientes de utilizar datos generados aleatoriamente. Se recomienda el uso de éstos sólo al principio del proyecto y en casos muy limitados. Los datos generados aleatoriamente presentan un gran riesgo, porque pueden generar mejores resultados en un algoritmo que en otro. Por esto la mejor forma de evaluar un SR es utilizando datos reales.

4.2. Obtención de datos de prueba

Como principal problema para realizar las pruebas se encontró que el sistema no contaba con perfiles de usuarios reales (datos personales, preferencias generales, etc.). Para no realizar la evaluación sobre datos aleatorios, resultó conveniente simular la utilización del sistema por parte del usuario realizando encuestas a usuarios reales. Los datos obtenidos se procesaron e ingresaron en la base de datos. Con estos datos se hicieron las pruebas necesarias para verificar el correcto funcionamiento del GRSK. Ésta resultó ser la forma más rápida y eficaz para obtener el perfil del usuario.

Concretamente, se realizaron encuestas a 60 usuarios diferentes. La finalidad de las encuestas era obtener el perfil de cada uno de los usuarios. De los datos obtenidos, los más relevantes fueron los datos personales y los referentes a las visitas previas realizadas a diferentes lugares turísticos de la ciudad de Valencia.

Formulario de Datos Personales

Nombre

Edad

Sexo

e-Mail

Hijos

Edad del hijo mayor

Edad del hijo menor

Profesión

Datos sobre preferencias al realizar una visita turística en una ciudad

Forma de desplazarse entre las actividades:

- Transporte público (metro, autobús)
- Taxi
- Vehículo particular
- Caminando

Me gusta visitar cuantas más cosas mejor

Prefiero tomarme las cosas con calma, ya volveré en otra ocasión

Selecciona el perfil o perfiles de viajero con los que más te identificas (máximo 3)

- Viajero cultural
- Viajero de negocios
- Viajero urbano
- Turista
- Viajero familiar
- Viajero de playa
- Viajero de compras
- Viajero de salud y relax
- Viajero gastronómico
- Viajero de ocio nocturno

Figura 4.1: Datos personales.

4.2.1. Cuestionarios

Los cuestionarios utilizados para las encuestas se pueden ver en las figuras 4.1 - 4.4.

En el cuestionario de la figura 4.1 se obtienen los datos personales del usuario, como el nombre, la edad, el estado civil, etc. Como punto importante, en este cuestionario, el usuario puede seleccionar con qué perfil turístico se identifica más (cultural, de negocios, familiar, etc.). Esto ayudará posteriormente (en un futuro trabajo) a mejorar la definición de los perfiles demográficos.

En las encuestas, el usuario puede indicar las visitas que previamente ha realizado (ver figura 4.2), señalando los lugares que conoce, si los ha visitado o no y si le interesan. Esta información sirve para calcular el contador de aceptación AC de un lugar determinado (ver sección 4.2.2). Recordemos que el AC representa el total de veces que un lugar ha sido visitado por el conjunto de usuarios U .

También se solicita al usuario que indique con qué frecuencia realiza ciertas

Lugar turístico	Visitado
Antiguo Convento de Santo Domingo	No lo conozco
Archivo Histórico Municipal (Palacio de Cervelló)	No lo he visitado
Ateneo Mercantil	No lo he visitado, pero me gustaría
Ayuntamiento	No lo he visitado, pero no me interesa
Baños del Almirante	Lo he visitado
Basilica de la Virgen	Lo he visitado y me gusta
Biblioteca Histórica Universitaria	Lo he visitado y no me gusta
Biblioteca Municipal - Crucero del Antiguo Hospital	Lo he visitado
Bioparc	Lo he visitado
Campo de Golf del Saler	No lo he visitado, pero me gustaría
Casa de la America's Cup	Lo he visitado

Figura 4.2: Visitas previas realizadas.

¿Con que frecuencia realizas estas actividades?

Cine	Semanal
Teatro	Nunca
Comer / cenar en restaurantes	Semanal
Conciertos	Semanal
Eventos deportivos	Nunca

Cine

Elige los tres géneros de cine que sueles ver por orden de preferencia

Romántica
Aventuras
Drama

Figura 4.3: Frecuencia de actividades.

actividades como ir al cine o al teatro, comer en restaurantes, etc. (ver figura 4.3).

Además, el usuario puede indicar sus gustos ante ciertas actividades (ver figura 4.4), con un valor de interés, denominado $U_r(f) \in \mathfrak{R} [1, 100]$. Este valor, junto con las visitas previas realizadas, ayuda a definir los valores de interés de cada lugar con respecto a los conceptos de la taxonomía, como se explicará en la siguiente sección.

4.2.2. Carga de información

La información obtenida con los cuestionarios se volcó a cada una de las tablas correspondientes de la BD, relacionadas con los datos personales, las preferencias generales, el historial de lugares recomendados, etc. Los datos personales (ver figura 4.1) se utilizan para generar la información demográfica del usuario. El historial de lugares recomendados (ver figura 4.2) formará parte de su historial real, y será utilizado como feedback en el sistema. Además, es utilizado para calcular la puntuación

Deportes		4
Deportes nauticos		4
Eventos deportivos		4
Futbol		4
Baloncesto		5
Otros deportes		2

Museos		5
Arte		1
Ciencia y tecnologia		2
Historia		8
Religión		3
Ciencias naturales		10
Arqueología		3
Historia y cultura local		4
Artesanía		5
Militar		3
Otros museos		6

Figura 4.4: Preferencias generales.

que le da el usuario a cada lugar visitado, denominada $U_r(i)$. También se almacenaron las preferencias generales de cada uno de los usuarios con la puntuación otorgada a cada característica, denominada $U_r(f)$ (ver figura 4.4).

Para realizar las pruebas, cuyo resultado se muestra en la sección 4.3, se dividió el total de usuarios n almacenados en la BD en dos subconjuntos: un conjunto de entrenamiento (Training Set) de 50 usuarios y otro de pruebas (Test Set) con 10 usuarios (ver figura 4.5).

Con el Training Set se determinó un contador de aceptación AC , que representa el conteo de visitas de un lugar determinado, y un valor de interés V para cada característica f de un lugar i , dependiendo si el usuario u ha visitado el lugar o no. Por tanto, asumiremos que, dado un elemento i asociado a una característica f , el valor de interés de i con respecto a f se encuentra determinado por:

$$V(f,i) = \sum_{U_r=1}^{n_t} U_r(i) * U_r(f) / n_t$$

donde n_t representa el total de usuarios del Training Set, $U_r(i)$ es la puntuación que le da el usuario a un lugar determinado y $U_r(f)$ es la puntuación que le otorga el usuario a una característica determinada y $V(f,i)$ que representa el valor de interés, suponiendo que se ha recomendado el lugar i debido a la característica f . Como se está utilizando como base información no extraída del proceso de recomendación, no se sabe qué característica ha originado dicha recomendación, y por tanto se realiza la operación para todas las características de cada lugar.

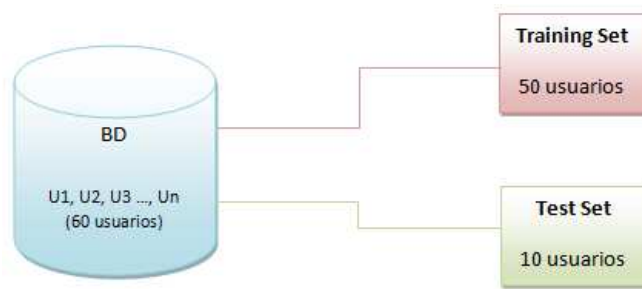


Figura 4.5: Escenario de pruebas.

El valor de interés V obtenido debe ser normalizado en relación al número de usuarios que han puntuado positivamente un lugar i , denominado $T_{ur}(i)$ y la puntuación máxima que puede obtener un lugar, denominada $P_{Max} \in \mathfrak{R} [1, 100]$, cuando es puntuado positivamente por todos los usuarios (n_t). La normalización se realiza con un factor de ponderación FP .

Más formalmente sería:

$$V'(f,i) = V(f,i) * FP$$

De donde $FP \in \mathfrak{R} [0, 1]$ y viene determinado por $T_{ur}(i)/n_t$.

Este factor de ponderación permite acercar los resultados un poco más a la realidad, teniendo en cuenta que no es lo mismo que un lugar i lo puntúen 3 personas a que lo puntúen 20 personas. El valor de interés debe ser más bajo cuando lo puntúan 3 personas, sin importar lo positiva o negativa que sea la puntuación.

4.3. Resultados

Tal y como se comentó anteriormente, dado un conjunto U de n usuarios, se dividieron en dos subconjuntos. Con el Training Set se entrenó el sistema y con el Test Set se realizaron y evaluaron las pruebas que simulan la utilización del sistema. En la figura 4.5 se muestra el escenario en el que se realizaron las pruebas.

Para cada uno de los usuarios del Test Set, se simulaban dos situaciones: cuando se trata de un usuario nuevo (*sin feedback*) y cuando se trata de un usuario que ya ha utilizado el sistema (*con la mitad del feedback*). El proceso es el siguiente: se simula

el registro de cada usuario asociado al Test Set en el sistema, ingresando la mínima información conocida del usuario (datos personales e información demográfica). Seguidamente el sistema genera una recomendación, una selección S para el usuario. Con la intersección entre S y los lugares que realmente ha visitado el usuario, denominados R (obtenidos en el cuestionario de la figura 4.2 “Visitas previas realizadas”), obtuvimos el conjunto de lugares relevantes y seleccionados RS .

Si denominamos $N_s = |S|$, $N_r = |R|$ y $N_{rs} = |R \cap S|$, se halla la **precision** y el **recall**. La **precisión** mide la probabilidad de que un elemento recomendado sea relevante para el usuario mientras que el **recall** mide la probabilidad de que un elemento relevante sea recomendado.

Este proceso se realizó primeramente sin la información de las “Visitas previas realizadas” cargada en la base de datos, en adelante *sin feedback* (es decir, con un usuario nuevo), y después con la mitad aleatoria de esta información ingresada en la base de datos, en adelante *con feedback* (es decir, un usuario que ya ha utilizado el sistema).

Para evaluar el SR, se efectuaron las pruebas con un número máximo de recomendaciones generadas por el sistema de 25.

Los resultados obtenidos fueron los siguientes:

	Sin feedback		Con la mitad del feedback	
	Precisión	Recall	Precision	Recall
usuario1	88%	31%	88%	31%
usuario2	92%	27%	88%	26%
usuario3	16%	15%	32%	30%
usuario4	48%	29%	64%	38%
usuario5	44%	20%	60%	28%
usuario6	68%	27%	72%	29%
usuario7	40%	32%	52%	42%
usuario8	56%	25%	68%	30%
usuario9	0%	0%	0%	0%
usuario10	64%	31%	60%	29%
Media	52%	24%	58%	28%

25 recomendaciones

Figura 4.6: Resultado de la evaluación aplicada a un máximo de 25 recomendaciones.

Al ejecutar el SR e-Turismo *sin feedback* (desconocimiento total de las visitas previas realizadas por el usuario), éste realiza recomendaciones menos efectivas en

comparación con la ejecución del SR *con feedback* (ver figura 4.6), en la que obtuvo una media de precisión del 58 %, mejorando en un 6 % las recomendaciones ofrecidas. Aunque en algunos usuarios (usuario2 y usuario10) la precisión disminuye, las recomendaciones ofrecidas siguen siendo buenas ya que la precisión no disminuye de forma drástica. Esta disminución se produce porque el sistema nunca recomendará lugares ya visitados por el usuario (es decir, en su historial); en este caso han coincidido varios lugares en la recomendación con lugares visitados previamente. Esto se puede observar en la figura 4.6.

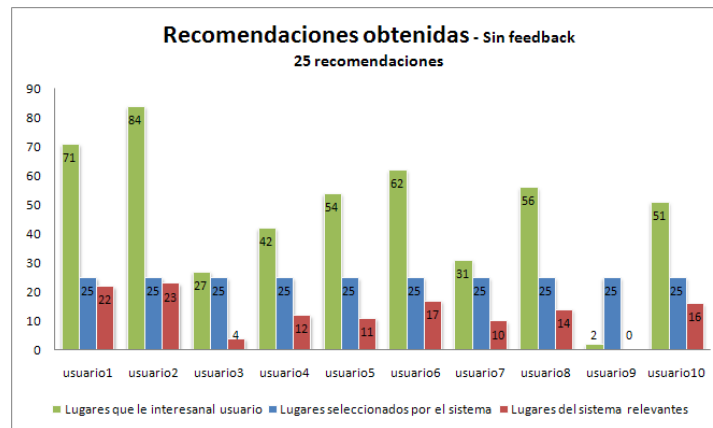


Figura 4.7: Comparación de resultados con un máximo de 25 recomendaciones sin feedback.

Como caso anómalo, se encontró que un usuario había puntuado en las encuestas solo 2 de 140 lugares, dando lugar a un $Nr=2$. Con este valor tan bajo de Nr al sistema le resulta imposible obtener una recomendación con ítems que sean relevantes para el usuario. Por este motivo la **precisión** y el **recall** presentan en ese usuario el valor cero (ver figuras 4.7 - 4.8). Cabe destacar, que si se calcula la media de la **precisión** sin este usuario (usuario9), la **precisión** estaría entre un 62 y 64 %, lo que permite concluir que el sistema en general presenta muy buenas recomendaciones.

En cuanto al **recall**, cabe mencionar que los valores obtenidos están alrededor del 24 % *sin feedback* y del 28 % *con feedback*. Esto implica que, de nuevo, la información del usuario permite recomendar elementos que el usuario ha marcado como relevantes.

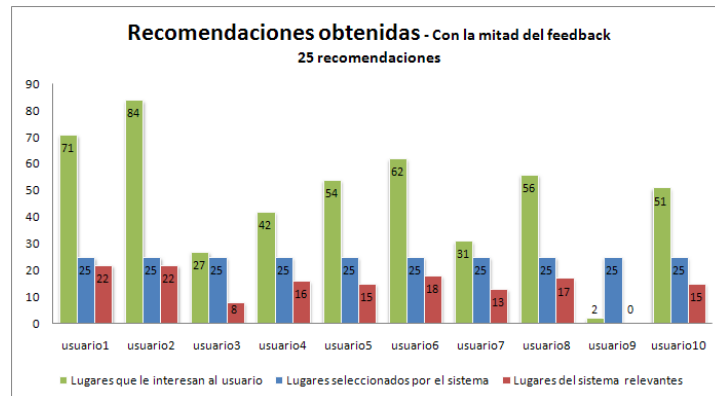


Figura 4.8: Comparación de resultados con un máximo de 25 recomendaciones con feedback.

La razón fundamental por la que e-Turismo *con feedback* obtiene recomendaciones más relevantes para el usuario, es que en ese momento el SR cuenta con información puntuada por el usuario (**feedback**). Esta información funciona como un conocimiento previo a la recomendación, permitiéndole al SR aprender de la experiencia del usuario antes de generar la recomendación. Esto se corrobora también al observar que para los usuarios que puntuaron más lugares, el sistema obtiene mejores recomendaciones.

4.4. Conclusiones

En esta sección se llevó a cabo la evaluación del SR e-Turismo propuesto en el capítulo 3. La evaluación ha consistido en una comparación de las recomendaciones obtenidas en la resolución del problema de turismo mediante, las técnicas básicas de recomendación incluidas en GRSK.

Tras analizar los resultados obtenidos con información recogida a través de encuestas, se deduce que, en general, es una buena opción mezclar las diferentes técnicas básicas para resolver el problema mediante una técnica de recomendación híbrida, sobre todo cuando se pretende trabajar con dominios de gran complejidad, bien por la gran cantidad de información a manejar o por la complejidad de las técnicas utilizadas.

A partir de los resultados obtenidos durante la evaluación del SR e-Turismo desarrollado en este trabajo, se resaltaron las siguientes propiedades que presenta este Sistema Recomendador:

- Evita el problema del “Nuevo usuario”. El módulo GRSK desarrollado es capaz de recomendar al menos una lista de actividades adecuadas al usuario, aún cuando éste es nuevo en el sistema.
- Resulta ser eficiente. En una ejecución, GRSK proporciona recomendaciones que resultan interesantes al usuario según sus preferencias.
- Es apropiado para casos reales. La información de los usuarios de prueba que se utilizaron, se recolecto por medio de encuestas. Esto hace que la información sea tomada de usuarios reales. Un gran número de recomendaciones resultaron factibles para cada uno de los usuarios en distintas situaciones, lo que permite concluir que las técnicas híbridas de recomendación son métodos eficientes para la resolución del problema.

En resumen, podemos concluir que GRSK obtiene recomendaciones para los usuarios gracias al uso de la técnica híbrida, que permite combinar diversas técnicas básicas.

Capítulo 5

Conclusiones

5.1. Introducción

En este capítulo se resumen las principales aportaciones realizadas en esta tesina en relación a los objetivos específicos planteados inicialmente.

Además se destacan las principales contribuciones de este trabajo. Se explican cada uno de los posibles futuros desarrollos que se pueden realizar para ampliar la funcionalidad de e-Turismo. También se indican las publicaciones a las que han dado lugar los resultados de este trabajo.

5.2. Contribuciones de este trabajo

Este trabajo ha contribuido a definir un problema de recomendación de lugares turísticos, considerando una taxonomía para representar el dominio de turismo y las características principales de cada uno de los lugares.

Además, se han analizado diferentes técnicas de recomendación, revisando la forma más adecuada de aplicarlas al problema. Así, se han establecido las ventajas e inconvenientes que tienen los Sistemas Recomendadores existentes, estudiando la viabilidad de aplicar una técnica o otra en el problema planteado.

Además el presente trabajo ha definido las principales características de un problema de planificación turística, dando una visión de cómo incorporar nuevas restricciones a las contempladas en los problemas genéricos de planificación, como por ejemplo, horarios de atención, limitación del tiempo disponible, etc.

5.3. Beneficios de este trabajo

Uno de los objetivos al desarrollar este proyecto es que sea utilizado por todas aquellas personas que deseen visitar la ciudad de Valencia y a los propios ciudadanos que deseen conocerla en profundidad o simplemente realizar actividades culturales.

También beneficiaría a empresas e instituciones que ofrecen servicios turísticos en la ciudad de Valencia. Muchas de estas empresas disponen de una página web a través de la cual se pueden consultar rutas turísticas totalmente estática y general, mostrando siempre la misma información a usuarios muy distintos. Es el usuario quien debe seleccionar, de toda la información disponible, únicamente aquella que le interesa.

Así, el principal beneficio que aporta este proyecto es la posibilidad de generar visitas turísticas adaptadas al perfil del usuario, simplificando la tarea de planificar un viaje o un día de turismo. El sistema puede incorporar información muy valiosa para el usuario. Por ejemplo, puede potenciar la visita a lugares interesantes pero poco conocidos. También podría sugerir la visita a ciertos lugares en horas de menor aglomeración, o a exposiciones o actividades temporales en lugares ya visitados. En definitiva, el sistema permite incorporar conocimiento experto sobre la ciudad, diseñando visitas totalmente personalizadas y guiadas, tal y como sucede cuando se visita una ciudad de la mano de un habitante del lugar. Además el hecho de mostrarse como un servicio web permite que sea accesible desde cualquier parte del mundo llegando a la mayoría del público.

Por tanto, podríamos concluir que la utilización de este sistema permitiría mejorar, los servicios turísticos online que actualmente se proporcionan. Esto reportaría importantes beneficios en el sector social, cultural y económico no sólo de la ciudad de Valencia sino de cualquier ciudad que lo adoptara. Además el mismo sistema de recomendación utilizado en este caso para recomendar lugares turísticos en una ciudad, puede utilizarse para la recomendación de cualquier otro tipo de ítems que puedan ser representados mediante una taxonomía (como películas, libros, etc.). Este ha sido uno de los objetivos principales del diseño: la independencia de la aplicación.

5.4. Futuros desarrollos

Se pueden deducir posibles trabajos futuros en las siguientes líneas:

- **Incorporar actividades temporales.** Actualmente se realizan las recomendaciones con actividades permanentes como por ejemplo, “Museo de las Artes y las Ciencias”, “Palacio del congreso”, etc. Sin que se tengan en cuenta eventos temporales como un concierto, eventos deportivos (Formula 1, partidos de fútbol, etc), etc.

El objetivo futuro consiste en construir recomendaciones en las cuales se tenga en cuenta las actividades de tipo temporales, de forma que se obtengan recomendaciones más satisfactorias para el usuario.

- **Recomendaciones para grupos.** Basadas en la negociación de agentes, en la que un grupo de usuarios, a través de la negociación entre agentes, puedan ponerse de acuerdo de las actividades a realizar durante su viaje turístico a la ciudad de Valencia. Actualmente GRSK no contempla este punto, y resultaría interesante implementarlo a futuro.
- **Análisis más exhaustivo del rendimiento del sistema.** Por un lado, se podrían analizar otras técnicas híbridas para compararlas con las que se ha utilizado en este trabajo y por otro, se podrían mejorar los perfiles demográficos, ya que a través de estudios basados en el marketing, se podrían adaptar más fielmente las restricciones de los perfiles demográficos.

5.5. Publicaciones y artículos relacionados

Los resultados del presente proyecto han dado lugar a las siguientes publicaciones:

- **e-Tourism: a tourist recommendation and planning application.** Laura Sebastián, Inma Garcia, Eva Onaindía, Cesar Guzmán. Ponencia en International Conference on Tools with Artificial Intelligence (ICTAI 2008) Dayton, Ohio (USA). (CORE:B, Computer Science Conference Ranking:0.74).

- **e-Tourism: a tourist recommendation and planning application.** Laura Sebastián, Inma Garcia, Eva Onaindía, Cesar Guzmán. Aceptado para su publicación durante 2009 en el International Journal on Artificial Intelligence Tools (IJAIT). (JCR: tercer tercio).

Bibliografía

- [1] Adomavicius G., Tuzhilin A. Personalization technologies: a process-oriented perspective. *Communications of the ACM*, 48(10):83–90, 2005.
- [2] Adomavicius G., Tuzhilin A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6):734–749, 2005.
- [3] Aha D.W., Breslow L., Muñoz-Avila H. Conversational case-based reasoning. *Applied Intelligence*, 14:9–32, 2000.
- [4] Paginas Amarillas. <http://www.paginasamarillas.es/>.
- [5] Amazon. <http://www.amazon.com/>.
- [6] Apml. <http://www.apml.org/>.
- [7] Ardissono L., Goy A., Petrone G., Segnan M., Torasso P. Intrigue: personalized recommendation of tourist attractions for desktop and handset devices. *Applied AI, Special Issue on Artificial Intelligence for Cultural Heritage and Digital Libraries*, 17(8-9):687–714, 2003.
- [8] Armstrong J.S. *Principles of Forecasting. A Handbook for Researchers and Practitioners*. Kluwer Academic, 2001.
- [9] Avery C., Zeckhauser R. Recommender systems for evaluating computer messages. *Communications of the ACM*, 40(3):88–89, 1997.
- [10] Balabanovic M., Shoham Y. Fab: Content-based collaborative recommender. *Communications of the ACM*, 40(3):66–72, 1997.

- [11] Belkin N.J., Croft W.B. Information filtering and information retrieval: Two sides of the same coin? *Communications of the ACM*, 35(12):28–38, 1992.
- [12] Billsus D., Pazzani M. User modeling for adaptive news access. *User Modeling and User-Adapted Interaction*, 10(2-3):147–180, 2000.
- [13] BlogLines. <http://www.bloglines.com/>.
- [14] Breese J., Heckerman D., Kadie C. Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, pages 43–52, 1998.
- [15] Brin S., Page L. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1-7):107–117, 1998.
- [16] Burke R. Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, 12(4):331–370, 2002.
- [17] Burke R. *The Adaptive Web*, chapter Hybrid web recommender systems, pages 377–408. Springer Berlin / Heidelberg, 2007.
- [18] Burke R., Hammond K., Young B. The findme approach to assisted browsing. *IEEE Expert*, 12(4):32–40, 1997.
- [19] Cluztr. <http://www.cluztr.com/>.
- [20] Dandelife. <http://www.dandelife.com/>.
- [21] Agencia Valenciana de Turismo. Servicio de estudios y planificación.
- [22] DeepMap. Deep map: Intelligent, mobile, multi-media and full of knowledge (project homepage). european media laboratory. <http://www.eml.org/english/research/deepmap/deepmap.html>, 2001.
- [23] Delgado J., Davidson R. Knowledge bases and user profiling in travel and hospitality recommender systems. In *Proceedings of the Enter 2002 Conference*, pages 1–16. Springer Verlag, 2002.
- [24] dietorecs. <http://dietorecs.itc.it/technicalapproach.html>.

- [25] ectrl. <http://ectrl.itc.it:8080/home/index.jsp>.
- [26] Engagd. <http://www.engagd.com/>.
- [27] Herlocker et al. Evaluating collaborative filtering recommender systems. *Information Retrieval*, 2008.
- [28] Felfernig A., Gordea S., Jannach D., Teppan E., Zanker M. A short survey of recommendation technologies in travel and tourism. *OEGAI Journal*, 25(7):17–22, 2007.
- [29] Fesenmaier D.R., Ricci F., Schaumlechner E., Wober K., Zanella C. Dietorecs: Travel advisory for multiple decision styles. In *Information and Communication Technologies in Tourism 2003: Proceedings of the International Conference in Lausanne*, pages 232–241, 2003.
- [30] Fink J., Kobsa A. A review and analysis of commercial user modeling servers for personalization on the world wide web. *User Modeling and User-Adapted Interaction*, 10(10):209–249, 2000.
- [31] Fink J., Kobsa A. User modeling for personalized city tours. *Artificial Intelligence Review*, 18(1):33–74, 2002.
- [32] Goldberg D., Nichols D., Oki B.M., Ferry D. Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12):61–70, 1992.
- [33] google. <http://www.google.com/>.
- [34] Goy A., Ardisson L., Petrone G. *The Adaptive Web*, chapter Personalization in e-commerce applications, pages 485–520. Springer Berlin / Heidelberg, 2007.
- [35] Guttman R.H., Moukas A.G., Maes P. Agent-mediated electronic commerce: A survey. *Knowledge Engineering Review*, 13(2):147–159, 1998.
- [36] Harold E.R. *Effective XML*. Addison-Wesley, 2004.
- [37] Herlocker J.L., Konstan J.A., Terveen L.G., Riedl J.T. Evaluating collaborative filtering recommender systems. *ACM Trans. Information Systems*, 22(1):5–53, 2004.

- [38] Hill W., Stead L., Rosenstein M., Furnas G. Recommending and evaluating choices in a virtual community of use. In *Conference Proceedings on Human Factors in Computing Systems (CHI)*, pages 194–201, 1995.
- [39] Hoekstra M. Basics of attention profiling through apml. <http://www.cleverclogs.org/2007/10/basics-of-atten.html>, 2007.
- [40] Jannach D., Zanker M., Jessenitschnig M., Seidler O. *Information and Communication Technologies in Tourism 2007*, chapter Developing a conversational travel advisor with Advisor Suite, pages 43–52. Springer, 2007.
- [41] jrcom. <http://www.jr.com/>.
- [42] Kolodner J. *Case-Based Reasoning*. Morgan Kaufmann, 1993.
- [43] Kowalski G.J., Maybury M.T. Information storage and retrieval systems. theory and implementation. *Kuwer Academic*, 2001.
- [44] Krulwich B. Lifestyle finder: Intelligent user profiling using large-scale demographic data. *Artificial Intelligence Magazine*, 18(2):37–45, 1997.
- [45] Krulwich B., Burkey C. Learning user information interests through extraction of semantically significant phrases. In *Proceedings of the AAAI Spring Symposium on Machine Learning in Information Access*, 1996.
- [46] Valencia Lanetro. <http://valencia.lanetro.com/>.
- [47] Lang K. Newsweeder: Learning to filter netnews. In *Proceedings of the 12th International Conference on Machine Learning*, 1995.
- [48] LastFM. <http://www.lastfm.es/>.
- [49] Lieberman H. Letizia: An agent that assists web browsing. In *Proceedings of the 14th International Joint Conference in Artificial Intelligence*, pages 924–929, 1995.
- [50] Lilien G.L., Kotler P., Moorthy K.S. *Marketing Models*. Prentice Hall, 1992.
- [51] McGinty L., Smyth B. Collaborative case-based reasoning: applications in personalised route planning. In editors D. Aha, I. Watson, editor, *Proceedings*

- of the *International Conference on Case-Based Reasoning (ICCBR-01)*, pages 362–376. Springer-Verlag, 2001.
- [52] McGinty L., Smyth B. Comparison-based recommendation. In Springer Verlag, editor, *Advances in Case-Based Reasoning, Proceedings of the 6th European Conference on Case Based Reasoning (ECCBR 2002)*, pages 575–589, 2002.
- [53] McGinty L., Smyth B. Improving the performance of recommender systems that use critiquing. In *Intelligent Techniques for Web Personalization, IJCAI 2003 Workshop, ITWP 2003*, pages 114–132, 2003.
- [54] Mirzadeh N., Ricci F. Cooperative query rewriting for decision making support and recommender systems. *Applied Artificial Intelligence*, 21(10):895–932, 2007.
- [55] Mladenic D. Personal web watcher: Implementation and design. Technical report, Technical Report IJSDP-7472, Department of Intelligent Systems, J. Stefan Institute, Slovenia, 1996.
- [56] Mooney R.J., Roy L. Content-based book recommending using learning for text categorization. In *Workshop on Recommender Systems: Algorithms and Evaluation (SIGIR'99)*, 1999.
- [57] movielens. <http://www.movielens.org/login>.
- [58] Murthi B.P.S., Sarkar S. The role of the management sciences in research on personalization. *Management Science*, 49(10):1344–1362, 2003.
- [59] newsgator. <http://www.newsgator.com/>.
- [60] Opperman R. *Adaptive user support - Ergonomic design of manually and automatically adaptable software*. Lawrence Erlbaum Associates, 1994.
- [61] Particls. <http://www.particls.com/>.
- [62] Pazzani M.J. A framework for collaborative, content-based and demographic filtering. *Artificial Intelligence Review*, 13:393–408, 1999.
- [63] Pazzani M.J., Billsus D. *The Adaptive Web*, chapter Content-based recommendation systems, pages 325–341. Springer Berlin / Heidelberg, 2007.

- [64] Peppers D., Rogers M. *Managing Customer Relationships: A Strategic Framework*. John Wiley and Sons, Inc., 2004.
- [65] Photoree. <http://www.photoree.com/photos/start>.
- [66] Powell M.J.D. *Approximation Theory and Methods*. Cambridge Univ. Press, 1981.
- [67] Resnick P., Iacovou N., Sushak M., Bergstrom P., Riedl J. GroupLens: An open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 Computer Supported Collaborative Work Conference*, 1994.
- [68] Resnick P., Varian H. Recommender systems. *Communications of the ACM*, 40(3), 1997.
- [69] Ricci F. Advanced topics in information systems. <http://www.inf.unibz.it/ricci/ATIS/index.html>, 2008.
- [70] Ricci F., Arslan B., Mirzadeh N., Venturini A. Itr: A case-based travel advisory system. In *Proceedings of the 6th European Conference on Advances in Case-Based Reasoning, Lecture Notes In Computer Science*, volume 2416, pages 613–627, 2002.
- [71] Ricci F., Cavada D., Nguyen Q.N. Integrating travel planning and on-tour support in a case-based recommender system. In *Proceedings of the Workshop on Mobile Tourism Systems (in conjunction with Mobile HCI'02)*, 2002.
- [72] Ricci F., Nguyen Q.N. Critique-based mobile recommender systems. *OEGAI Journal*, 24(4), 2005.
- [73] Rich E. User modeling via stereotypes. *Cognitive Science*, 3:329–354, 1979.
- [74] Salton G. *Automatic Text Processing*. Addison-Wesley, 1989.
- [75] Sarwar B.M., Karypis G., Konstan J.A., Riedl J. Analysis of recommendation algorithms for e-commerce. In *ACM Conference on Electronic Commerce*, pages 158–167, 2000.
- [76] Schafer J.B., Konstan J.A., Riedl J. E-commerce recommendation applications. *Data Mining and Knowledge Discovery*, 5(1/2):115–153, 2001.

- [77] Schwab I., Kobsa A., Koychev I. Learning user interests through positive examples using content analysis and collaborative filtering. Technical report, Fraunhofer Institute for Applied Information Technology, Germany, 2001.
- [78] Shadbolt N., Hall W., Berners-Lee T. The semantic web revisited. *IEEE Intelligent Systems*, 2006.
- [79] Shardanand U., Maes P. Social information filtering: Algorithms for automating word of mouth. In *Proceedings of CHI*, pages 210–217, 1995.
- [80] Shimazu H. Expertclerk: Navigating shoppers buying process with the combination of asking and proposing. In editor Nebel, B., editor, *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence, IJCAI 2001*, pages 1443–1448. Morgan Kaufmann, 2001.
- [81] Smyth B. *The Adaptive Web*, chapter Case-based recommendation, pages 342–376. Springer Berlin / Heidelberg, 2007.
- [82] Smyth B., Cotter P. A personalized tv listings service for the digital tv age. *Knowledge-Based Systems*, 13:53–59, 2000.
- [83] Smyth B., Cunningham P. A comparison of incremental case-based reasoning and inductive learning. In *Proceedings of the Second European Workshop on Case-Based Reasoning (EWCBR-94)*. Springer, 1994.
- [84] My Strands. <http://www.mystrands.com/>.
- [85] Towle B., Quinn C. Knowledge based recommender systems using explicit user models. Technical report, Knowledge-Based Electronic Markets, Papers from the AAAI Workshop, AAAI Technical Report WS-00-04, 2000.
- [86] tripadvice. <http://tripadvice.itc.it:8080/>.
- [87] Turista Valencia. <http://www.turisvalencia.es/>.
- [88] Comunitat Valenciana. <http://www.comunitatvalenciana.com/>.
- [89] Wang Y., Kobsa A. Respecting users' individual privacy constraints in web personalization. In *The 11th International Conference on User Modeling*. Springer-Verlag, 2007.

- [90] WebGuide. Webguide: A city guide for the internet. european media laboratory.
<http://www.eml.org/english/research/deepmap/deepgis/webguide.html>, 2001.
- [91] yahoo. <http://www.yahoo.com/>.
- [92] Yang Y., Padmanabhan B. On evaluating online personalization. In *Proc. Workshop Information Technology and Systems*, pages 35–41, 2001.