

Agente Inteligente para la identificación automática de perfiles de usuarios de turismo

Silvana Aciar, Laura Aballay, Leandro Castro, Franco Cortez,
Cesar Collazos, Susana Ruiz, Cintia Ferrarini, Maria
Romagnano

Instituto de Informática IdeI, Facultad de Ciencias Exactas, Físicas
y Naturales, Universidad Nacional de San Juan
Ignacio de la Roza y Meglioli. Complejo CUIM. UNSJ

Tel 264-265101

saciar, laballay{ @iinfo.unsj.edu.ar }castroleandro6@gmail.com

Resumen

El dominio de turismo es un ambiente muy dinámico, los productos (paquetes turísticos) cambian continuamente, como así también los intereses de los clientes (turistas). Por ejemplo, las preferencias sobre un tipo de lectura son más estables y constantes que las preferencias sobre las visitas turísticas. Las diferentes fuentes de información para poder predecir preferencias más completas, como así también la consideración de múltiples variables para poder realizar las recomendaciones, son los temas centrales a los que apunta el proyecto. En este aspecto el proyecto permitirá el desarrollo e implementación de innovadoras tecnologías de software aplicables al turismo regional que ayudará a mejorar la efectividad de los sitios web locales en cuanto a la atracción de turistas. La incorporación de diferentes fuentes de información del turista y los métodos y técnicas para su recuperación incrementaría la precisión de las recomendaciones. Recomendaciones más precisas incrementa la satisfacción del turista, ocasionando la fidelidad del turista e incrementando la popularidad del sitio web permitiendo una más amplia difusión del turismo local y promoción del patrimonio local.

Como así también permitirá definir paquetes turísticos en función de la caracterización del perfil dinámico del turista que se obtendrá a partir de la interacción del turista con los sitios web

Palabras clave: Sistemas Recomendadores, Sistemas Multi-agentes, web mining.

Contexto

El proyecto, aprobado por evaluación externa y financiado por la UNSJ -Cod E904-, se encuentra inserto en el marco de las líneas de investigación del Gabinete de Ingeniería de Software del Instituto de Informática. Además se encuentra vinculado a cátedras de las carreras de Licenciatura en Ciencias de la Computación y Licenciatura en Sistemas de Información, de la FCEF N de la Universidad Nacional de San Juan.

Introducción

La penetración masiva de Internet en los hogares sanjuaninos hace posible que los usuarios tengan acceso a la gran cantidad de información disponible en la web. Esto provoca que en muchas ocasiones se produzca la paradoja de no saber qué escoger a pesar de que a nuestro alrededor se acumula información que podría ser de nuestro interés. Esta situación también se observa en el sector turístico. Hoy en día, la fuente de información de un turista es Internet. Actualmente los sitios web con información turística presentan la misma información a millares de usuarios que entran en el sitio, información estándar sobre rutas turísticas con links a otras páginas web sobre

lugares, transportes, alojamiento y actividades, y al final desistir en su búsqueda por la cantidad de tiempo que le lleva visitar cada página, o quedarse con una opción que quizás no sea la mejor para su perfil. Es por eso que el problema ya no es de no tener información, si no, qué escoger entre un número inmanejable de alternativas disponibles.

Como solución a este problema están en continuo refinamiento los algoritmos de los motores de búsqueda tales como Google, Yahoo y Bing, que son los más utilizados en el mercado hoy en día. Pero estos no dejan de ser herramientas de búsqueda masiva. Para evitar esta sobrecarga de información algunos sitios web, como Amazon2, han adoptado herramientas de recomendación para sugerirles cuál alternativa es la que mejor se adecua a sus necesidades de entre miles que presenta un buscador. Los “Sistemas Recomendadores” [1][2] son ampliamente estudiados en Inteligencia Artificial aplicando técnicas de esta disciplina para predecir los intereses y necesidades de los usuarios para poder presentarle la información que más necesite. En Sistemas Recomendadores existen numerosos trabajos relacionados a los métodos y técnicas para filtrar la información relevante a los usuarios basado en Árboles de Decisión, Clasificador Bayesiano, Sistemas Basado en Reglas con restricciones, Reglas Difusas, etc. [3] [4][5][6][7]. Como así también existen numerosos trabajos relacionados a la creación del perfil del usuario en el cual se basa el filtrado de la información. Mucha investigación se está realizando para analizar la información que el usuario deja de forma implícita al navegar por el sitio web y con ella crear su perfil. La historia de navegación de los usuarios es la fuente más utilizada para obtener preferencias e intereses. Los usuarios son clasificados dentro de patrones dependiendo de su historial de navegación; técnicas de minería de datos son utilizadas para la clasificación tales como reglas de asociación, árboles de decisión y reglas bayesianas [8][9][10]. Estas técnicas incorporan información sobre la visita a la página. Datos de tiempo de visita, frecuencia y orden de visita son utilizados para obtener el interés del usuario. Más recientemente, agentes de software y ontologías son utilizadas para incorporar más información del usuario y no tan solo información de su navegación, para conseguir personalización efectiva. Las ontologías poseen información de los intereses del usuario como así

también información del contexto [11][12][13]. Los agentes de software son generalmente utilizados en la creación de interfaces amigables de interacción con el usuario y para el manejo de grandes cantidades de información a filtrar y sintetizar. También los agentes de software son usados para poblar las ontologías [14][15][16][17]. La información que utilizan las diferentes aportaciones nombradas en el apartado anterior es generalmente provista de forma estructurada, es decir tiene una estructura que es fácil de leer por los métodos que se están utilizando actualmente. Por ejemplo se utilizan los clicks que realiza un usuario en su navegación por un sitio web para poder predecir su comportamiento, o se pide al usuario que se registre en el sitio e ingrese información de él, o se utiliza una ontología previamente definida con los intereses del usuario. Pero existe otra fuente de información que es más valiosa y es el texto que el propio usuario web escribe libremente al dejar un comentario u opinión sobre productos, experiencias y hechos. Esta clase de información es muy popular en turismo, donde los usuarios dejan su opinión principalmente sobre el alojamiento al que fueron. La incorporación de este tipo de información está comenzando a ser estudiada incorporando técnicas de text mining [18][19] para mejorar los perfiles de usuarios y así incrementar la precisión de las recomendaciones.

Las aportaciones mencionadas anteriormente, son aportaciones contexto-dependiente, donde el contexto o dominio es reducido, y utilizan una o dos fuentes de información, pero turismo es la integración de diferentes dominios; los cuales son dinámicos y que requiere la aplicación de técnicas casi en tiempo real y los intereses del usuario cambian continuamente y depende de muchas otras variables, muchas técnicas mencionadas en el estado del arte deben ser refinadas y adaptadas para la creación dinámica y casi en tiempo real de los perfiles de los turistas. La aportación de este proyecto es un avance al estado del arte utilizando agentes de software, cuyas técnicas de creación automática de perfiles de usuarios incorporarán no solo información estructurada como historial de navegación, información del contexto e información de usuarios similares, sino también información no estructurada (texto), esto permitiría conocer mucho mejor al usuario ofreciéndole la mejor alternativa para él de un conjunto numeroso de alternativas disponibles.

Líneas de investigación y desarrollo

En este proyecto se realizará investigación sobre el potencial de diferentes técnicas y herramientas como sistemas recomendadores, aprendizaje automático, web mining y agentes de software para apoyar la automatización de las tareas de recuperación de información de los usuarios de un sitio web turístico. La investigación y desarrollo que se propone en este proyecto nos llevará a cumplir el objetivo principal que es promover el turismo regional por medio de un sistema inteligente que brinde información turística personalizada.

Específicamente se han definido las siguientes actividades de investigación y desarrollo.

1. Definición de un Agente Inteligente que a partir de diferentes fuentes de información, disponibles en Internet, permita obtener automáticamente el perfil de un usuario de turismo.
2. Investigar y aplicar técnicas de aprendizaje automático para obtener el comportamiento del usuario de acuerdo a su historial de navegación por el sitio web.
3. Definir métodos para la obtención de información del turista que se encuentra en fuentes de información no estructurada (texto), como opiniones y comentarios que realizó sobre lugares turísticos que visitó.
4. Definir métodos para la obtención de información de usuarios similares e incorporación de dicha información en el perfil del usuario.

Resultados y Objetivos

Resultados obtenidos y resultados futuros:

La clave de este proyecto reside en la definición de un Agente Inteligente que a partir de diferentes fuentes de información, disponible en Internet, permita obtener automáticamente el perfil de un usuario de turismo. Para llevar a cabo este objetivo partimos de la premisa que más conocimiento del turista permitirá elaborar mejores perfiles del mismo.

Para obtener más conocimiento se deberían consultar diferentes fuentes de información del turista. Para adquirir esta información automáticamente se necesita de un programa

autónomo (Agente Inteligente) para seleccionar, filtrar e integrar dicha información y crear el perfil del turista. Se llevaron a cabo diferentes métodos y estudios para lograr este objetivo. Los resultados obtenidos hasta el momento son:

1. Estudio bibliográfico específicamente en modelado de usuarios, la información utilizada para crear los perfiles, las técnicas, métodos y metodologías existentes. Se estudiarán las variables y parámetros que intervienen directamente e indirectamente en la creación de los perfiles. Análisis de deficiencias en las soluciones actuales.
2. Estudio y análisis de las capacidades de los Agentes de Software en cuanto a la recolección y filtrado de información de usuarios en un sitio web. Este estudio incluirá las aportaciones realizadas en trabajos anteriores por el grupo de investigación sobre Agentes de software y sus capacidades de inteligencia, aprendizaje, búsqueda de información, comportamiento, razonamiento, etc.
3. Aplicación de técnicas de aprendizaje automático para obtener el comportamiento del usuario de acuerdo a su historial de navegación por el sitio web. Con la identificación de hábitos de navegación en un sitio web se conseguirá entender el comportamiento de navegación del usuario para adaptar las recomendaciones a sus necesidades.
4. Definición de mecanismos para la obtención de información del usuario que se encuentra en fuentes de información no estructurada (texto) como las opiniones y comentarios que realizó sobre sitios turísticos que visitó. Con la explosión de la Web 2.0 la mayoría de los sitios web de turismo poseen un blog, o fórum o un espacio donde brindan a los turistas la oportunidad para compartir experiencia y opiniones tanto negativas como positivas sobre su experiencia en los sitios turísticos, alojamiento, transporte, etc.
5. Definir métodos para la obtención de información de usuarios similares e incorporación de dicha información en el perfil del usuario. Se parte de la premisa que usuarios con perfiles similares tienen la misma preferencia sobre lugares turísticos. Lo cual permite la automatización del proceso "boca-a-boca" donde los turistas recomiendan lugares a otros con preferencias similares a ellos. Para

establecer la similitud entre los perfiles de turistas se debería realizar:

Resultados previstos a obtener en el transcurso de 2012.

1. Integración y filtrado de intereses y preferencias del turista desde las tres fuentes de información: historial de navegación, similitud entre usuarios y de la información textual que introduce en un sitio web. Para ello se incorporarán las técnicas más viables de obtención de preferencias e intereses de los turistas estudiadas en las tareas anteriores en el agente de software para la creación del perfil del turista.
2. Se deberán definir las métricas de evaluación y las pruebas a realizar con usuarios finales para demostrar las hipótesis que dieron origen al desarrollo de la investigación, como así también demostrar la fiabilidad de los resultados obtenidos a lo largo del desarrollo de la investigación. En esta instancia es importante la colaboración de los expertos en Ingeniería del Software, integrantes del grupo colaborador del proyecto. Se realizará un experimento controlado con usuarios (turistas) reales durante un periodo de tres meses aproximadamente
3. Difusión y preparación para la explotación y uso de los resultados de la investigación. Consiste en la preparación de artículos para enviar a congresos y revistas nacionales e internacionales para lograr la aceptación de los resultados de investigación en la comunidad científica.

Actividades realizadas durante el primer año del proyecto

En el transcurso del primer año del proyecto se realizaron varias actividades en las cuales están involucrados miembros del proyecto:

- Capacitación:

Los integrantes de este proyecto han realizado diferentes cursos de capacitación como lo son:

- Inteligencia Artificial Distribuida – Dictado por la Dra Beatriz López de la Universitat de Girona, España.
- Curso de Posgrado HCI (Interacción Humano Computador) – Dictado por el professor Cesar Collazos , Universidad del Cauca , Colombia.
- Curso Programación Java- Proyecto "Cursos Interuniversitarios para el intercambio del

profesora Patricia Jaques, Universidad UNISINOS, Brasil.

- Cursos específicos al programa de doctorado de Ingeniería con orientación informática que realizan dos alumnas de doctorado involucrados en este proyecto.

- Seminarios

La Dra. Silvana Aciar ha dictado diferentes seminarios referentes a temas que se están investigando en el proyecto en las siguientes universidades:

- Universidad UNISINOS, Brasil. Agosto 2011.
- Universidad Libre, Bogotá Colombia. Octubre 2011.
- Universidad Privada Antenor Orrego, Trujillo, Perú. Noviembre 2011
- Universidad de Paris, Paris, Francia. Diciembre 2011.
- Universidad de Montevideo, Uruguay. Febrero 2012.

Publicaciones:

- Leandro Castro, Silvana Aciar: Personalización de sitios web turísticos locales utilizando Sistemas Recomendadores- 1ºPrototipo. VIII CISIISTR – Noviembre 2011, TRUJILLO-PERU
- Silvana Aciar, Leandro Castro, Cesar Collazos. Recommender Systems and Personalization. CAVA 2011- Noviembre, 2011. Bogotá, Colombia
- IHC+CLIHC 2011: “Modelo Colaborativo y Distribuido para Enseñanza de la Usabilidad”.L. Aballay, C Collazos y M Herrera

Formación de Recursos Humanos

Tesis de grado finalizadas:

Leandro Castro y Laura Aballay, integrantes del proyecto han elaborado su tesis de grado de la carrera de Licenciatura en la FCEFYN de la UNSJ.

Tesis de grado en curso:

Un estudiante adscripto al proyecto está elaborando su tesis de grado en el marco de este proyecto prevista a finalizar en julio 2012, Franco Cortez.

Tesis de maestria:

Dos integrantes que finalizaron la tesis de grado en este proyecto comenzarán la maestria en el marco de este proyecto en abril 2012:

Leandro Castro: Directores Sivana Aciar UNSJ – Argentina, Eliseo Reategui UFRGS – Brasil

Laura Aballay: Directores Cesar Collazos UNC – Colombia, Silvana Aciar UNSJ-Argentina

Tesis de doctorado:

Dos estudiantes de doctorado están realizando la tesis doctoral en temáticas relacionada al proyecto bajo la dirección de la Dra. Silvana Aciar. Maria Romagnano y Cintia Ferrarini

Posdoctorado:

Silvana Aciar está realizando un posdoctorado del CONICET en temas relacionados directamente con este proyecto. Como así también fue aprobada su entrada a la carrera del CONICET en noviembre de 2011.

Referencias

- [1] Adomavicius, G.. 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).
- [2] Felfernig, A. and Burke, R. Constraint-based recommender systems: technologies and research issues. In *Proceedings of the 10th international Conference on Electronic Commerce (Innsbruck, Austria, August 19 - 22, 2008)*. ICEC '08, vol. 342:1–10. (2008).
- [3] Mahmood, T. and Ricci, F. 2009. Improving recommender systems with adaptive conversational strategies. In *Proceedings of the 20th ACM Conference on Hypertext and Hypermedia (Torino, Italy, June 29 - July 01, 2009)*. HT '09. ACM, New York, NY, 73-82.
- [4] Chung, R., Sundaram, D., and Srinivasan, A. 2007. Integrated personal recommender systems. In *Proceedings of the Ninth international Conference on Electronic Commerce (Minneapolis, MN, USA, August 19 - 22, 2007)*. ICEC '07, vol. 258. ACM, New York, NY, 65-74.
- [5] Nnadi, N. J. 2009. Applying relevant set correlation clustering to multi-criteria recommender systems. In *Proceedings of the Third ACM Conference on Recommender Systems (New York, New York, USA, October 23 - 25, 2009)*. RecSys '09. ACM, New York, NY, 401-404.
- [6] Umyarov, A. and Tuzhilin, A. 2009. Improving rating estimation in recommender systems using aggregation- and variance-based hierarchical models. In *Proceedings of the Third ACM Conference on Recommender Systems (New York, New York, USA, October 23 - 25, 2009)*. RecSys '09. ACM, New York, NY, 37-44.
- [7] Herlocker, J. L., Konstan, J. A., Terveen, L. G., and Riedl, J. T. 2004. Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst.* 22, 1 (Jan. 2004).
- [8] Adomavicius, G. and Tuzhilin, A., Using Data Mining Methods to Build Customer Profiles, *Computer*, v.34 n.2, p.74-82, February 2001.
- [9] S. T.Wu, Y.Li, Y.Xu, B.Pham, and P.Chen, "Automatic Pattern-Taxonomy Extraction for Web Mining," presented at the IEEE/WIC/ACM International Conference on Web Intelligence, Beijing, China, 2004.
- [10] Bresferean V. P. (2007) "Analysis and predictions on student's behavior using decision trees in WEKA environment". *IEEE proceedings of the 29th International Conference on Information Technology Interfaces* , ITI 2007.
- [11] Sieg, A., Mobasher, B., and Burke, R. 2007. Web search personalization with ontological user profiles. In *Proceedings of the Sixteenth ACM Conference on Conference on information and Knowledge Management (Lisbon, Portugal, November 06 - 10, 2007)*. CIKM '07. ACM, New York, NY, 525-534.
- [12] He, S. and Fang, M. 2008. Ontological User Profiling on Personalized Recommendation in e-Commerce. In *Proceedings of the 2008 IEEE international Conference on E-Business Engineering (October 22 - 24, 2008)*. ICEBE. IEEE Computer Society, Washington, DC, 585-589.
- [13] Zhou, X., Wu, S., Li, Y., Xu, Y., Lau, R. Y., and Bruza, P. D. 2006. Utilizing Search Intent in Topic Ontology-Based User Profile for Web Mining. In *Proceedings of the 2006 IEEE/WIC/ACM international Conference on Web intelligence (December 18 - 22, 2006)*. Web Intelligence. IEEE Computer Society, Washington, DC, 558-564.
- [14] Pan, J., Zhang, B., Wang, S., Wu, G., and Wei, D. 2007. Ontology Based User Profiling in Personalized Information Service Agent. In *Proceedings of the 7th IEEE international Conference on Computer and information Technology (October 16 - 19, 2007)*. CIT. IEEE Computer Society, Washington, DC, 1089-1093.
- [15] Chu, W. and Park, S. 2009. Personalized recommendation on dynamic content using predictive bilinear models. In *Proceedings of the 18th international Conference on World Wide Web (Madrid, Spain, April 20 - 24, 2009)*. WWW '09. ACM, New York, NY, 691-700.
- [16] Ghosh, R. and Dekhil, M. 2009. Discovering user profiles. In *Proceedings of the 18th international Conference on World Wide Web (Madrid, Spain, April 20 - 24, 2009)*. WWW '09. ACM, New York, NY, 1233-1234
- [17] Ahn, J., Brusilovsky, P., Grady, J., He, D., and Syn, S. Y. 2007. Open user profiles for adaptive news systems: help or harm?. In *Proceedings of the 16th international Conference on World Wide Web (Banff, Alberta, Canada, May 08 - 12, 2007)*. WWW '07. ACM, New York, NY, 11-20.
- [18] Liu, B.. *Web Data Mining Exploring Hyperlinks, Contents and Usage Data*. Springer-Verlag, New York.(2006)
- [19] Aciar, S., Zhang, D., Simoff, S., and Debenham, J. 2007. Informed Recommender: Basing Recommendations on Consumer Product Reviews. *IEEE Intelligent Systems* 22, 3 39-47 (May. 2007)