

RECJUDGE: SISTEMA DE RECOMENDACIONES PARA JURADOS ONLINE DE PROGRAMACIÓN

RECJUDGE: SYSTEM OF RECOMMENDATIONS FOR ONLINE JURIES OF PROGRAMMING

Autor: Raciél Yera Toledo

Institución: Universidad de Ciego de Ávila Máximo Gómez Báez

Correo electrónico: ryera@unica.cu

RESUMEN

El movimiento de los concursos de programación se ha visto por estos años inmerso en un período de intensa actividad en Cuba. Sin embargo, este auge aún no ha dado todos los resultados esperados y la principal causa es la ausencia de una adecuada preparación de estudiantes y entrenadores que no son capaces de asimilar el flujo de información que sobreviene hacia ellos. En el presente trabajo se concibe una aplicación que integrado a un jurado online de programación y basada en la técnica de filtrado colaborativo, intenta solucionar esta problemática al generar automáticamente paquetes de actividades a realizar por los estudiantes, basados en su rendimiento y progreso actual, siendo resuelto este problema desde el enfoque del problema general de la recomendación. Además, se realiza un análisis comparativo con otras herramientas semejantes actualmente disponibles en la web.

Palabras Clave: Filtrado colaborativo, Jurado online, Sistemas de recomendación.

ABSTRACT

The movement of the programming competitions one has seen for these years inmerso in a period of intense activity in Cuba. However, this peak has not still given all the prospective fruits, and the main cause is the absence of an appropriate preparation of students and trainers that are not able to assimilate the flow of information that happens toward them. Presently work is conceived

an application that, integrated to an online jury of programming and based on the technique of collaborative filtrate, he/she tries to solve this problem when automatically generating packages of activities to carry out for the students, based on its yield and current progress, being resolved this problem from the focus of the general problem of the recommendation. Also, he/she is carried out a comparative analysis at the moment with other similar tools available in the web.

Keyword: Collaborative filtrate, Online jury, Recommendation systems.

INTRODUCCIÓN

El presente trabajo coincide con la reciente celebración, en octubre del 2009, por primera vez en Cuba como una de las subsedes, de la competencia eliminatoria regional para estudiantes universitarios con vistas a la final mundial de la ACM-ICPC.

La participación en concursos de programación constituye un excelente complemento al proceso de formación de los estudiantes en las carreras relacionadas directamente con la computación. En estos, comúnmente se tiene que resolver un conjunto de problemas complejos, con los que se desarrollan el razonamiento lógico y pensamiento algorítmico, sirviendo también de experiencia en la futura vida como profesional. Además, contribuyen a fortalecer los valores de laboriosidad y trabajo colaborativo, tan necesarios en los egresados.

A pesar de que en la última década Cuba ha comenzado a ocupar un lugar importante en el área latinoamericana en estos concursos y de contarse con herramientas de apoyo dentro de las cuales se destacan los jurados online. Se puede afirmar que aún, en la mayoría de los niveles, se dista mucho de realizar una preparación eficiente para este tipo de competencias. Los resultados obtenidos se deben principalmente a esfuerzos autodidactas de los propios estudiantes y en muchos casos, la tendencia general es a perderse en un mar de teoría y ejercicios no esenciales ni ordenados lógicamente, deviniendo en malos resultados no acordes a su dedicación y a su tiempo de estudio. Todo

esto provoca que exista un notable desbalance entre el resultado y las posibilidades de cada uno de los estudiantes.

Este problema de desbordamiento de información no es exclusivo del área de los concursos de programación. A nivel global, y tras el florecimiento de la www, a muchos usuarios les resulta difícil obtener la mejor información de la red. Como afirmara Jeffrey M. O'Brien, la web está abandonando la era de la búsqueda, y entrando en la del descubrimiento, en la que algo que ni siquiera el usuario tiene la menor idea de su existencia, debe fluir hacia él.

Una de las soluciones que se están comenzando a utilizar buscando eliminar esta problemática, es el desarrollo de sistemas de recomendación. Y como parte de estos, particularmente la aplicación de la técnica conocida como filtrado colaborativo. En el presente trabajo se concibe una herramienta que pretende guiar al estudiante a través del contenido disponible en un jurado online de programación, tratando de establecer una ruta ideal para el aprovechamiento y asimilación óptima, siendo este además el objetivo general de este tipo de sistemas aplicados al e-learning.

El resto del documento está organizado de la siguiente manera: el primer punto del desarrollo realiza un breve recorrido por los sistemas de recomendación utilizando filtrado colaborativo desde un enfoque meramente teórico, el segundo describe la solución propuesta, así como la forma en que esta se acoplaría a las herramientas disponibles. Mientras que el tercer punto menciona algunos trabajos semejantes al actual, para terminar con las conclusiones y trabajos futuros.

MATERIALES Y MÉTODOS

Sistemas de Recomendación y Filtrado Colaborativo

Los sistemas de recomendación automatizan o soportan parte del proceso de recomendación. Además, forman parte de un tipo específico de técnica de filtro de información, los cuales presentan distintos tipos de temas o ítems de información (películas, música, libros, noticias, imágenes, entre otros.) a los usuarios, basándose en la predicción del ranking o ponderación que el usuario le daría a un ítem que el sistema aún no ha considerado.

El problema de la recomendación puede ser formulado de la siguiente manera:

Sea C un conjunto de usuarios, y S el conjunto de todos los ítems posibles a ser recomendados y considerando además la función $u(c,s)$, que mide la utilidad del ítem s al usuario c , se desea, para cada usuario c en C , encontrar qué ítem s' en S que maximice la utilidad del usuario. Más formalmente:

Este tipo de sistemas emergieron como un área independiente de investigación a mediados de la década de los 90. A pesar de que de una forma u otra habían sido abordados anteriormente de forma indirecta desde diferentes enfoques tales como la aproximación, la recuperación de información y las teorías del pronóstico. Esta línea de investigación alcanzó el papel protagónico que actualmente tiene, sólo en el momento en que los investigadores se enfocaron en el problema de la recomendación desde el punto de vista de la relación entre usuarios e ítems.

Los sistemas de recomendación usualmente se clasifican teniendo en cuenta la forma en que son hechas las recomendaciones. Entre las principales categorías se encuentran:

Recomendaciones basadas en el contenido: en los que al usuario se le recomiendan ítems basados en los ítems que este ha preferido en el pasado.

Recomendaciones colaborativas: en los que al usuario se le recomiendan ítems que han sido preferidos y/o seleccionados por usuarios similares en el pasado.

Enfoques híbridos: que son los métodos que combinan las dos técnicas anteriores.

Particularmente la recomendación basada en filtrado colaborativo, continuando con la definición formal dada más arriba, consiste en estimar la mencionada utilidad $u(c,s)$ del ítem s para el usuario c , basada en las utilidades $u(c_j, s)$, asignadas a los ítems s por aquellos usuarios c_j en C , que son similares al usuario c .

Para la resolución eficiente de este problema han sido aplicadas disímiles técnicas de inteligencia artificial, en las que se pueden destacar enfoques de clustering, redes bayesianas, aprendizaje basado en instancias, sistemas expertos, y también algoritmos genéticos.

Entre las principales aplicaciones prácticas de la teoría antes expuesta, actualmente están la recomendación en tiendas online, el filtrado de noticias, las recomendaciones musicales, de libros y de películas, la búsqueda de personas afines en las comunidades; derivando la mayoría de estas, del campo del *e-Commerce*. Como sistemas específicos que utilizan el filtrado colaborativo como técnica de recomendación se pueden mencionar varias redes sociales como Facebook, y otros con objetivos más puntuales como GroupLens y Jester, para la recomendación de filmes y de chistes, respectivamente.

Unido a la consolidada utilización en el *e-Commerce* y tópicos afines, poco a poco han ido ganando presencia los sistemas de recomendación en las aplicaciones de *e-Learning*. Muchos autores han coincidido en que con ayuda de sistemas de recomendación se puede incrementar la seriedad, eficiencia y calidad del proceso de aprendizaje que las aplicaciones de *e-Learning* clásicas soportan. El trabajo que se presenta puede ser mostrado como otro ejemplo de estos sistemas.

Perfil Tecnológico

El perfil tecnológico de la aplicación desarrollada estuvo conformado por las siguientes tecnologías:

Lenguaje de programación Java

Descrito como un lenguaje simple, orientado a objetos, distribuido, interpretado, robusto, seguro, de arquitectura neutra, portable, de altas prestaciones, multitarea y dinámico. Se ha desarrollado unido a internet y es por ello que es ampliamente utilizado en el desarrollo de sistemas distribuidos, como complejas aplicaciones de gestión.

Entre las razones que justifican su utilización están que es un lenguaje cuya línea de expansión y desarrollo se corresponde con la política trazada por la Universidad de las Ciencias Informáticas (UCI) y por el país. Que la mayoría de las aplicaciones a las que se integrará la solución que se desarrolla, están implementadas en Java, por lo que, al asumirlo como lenguaje, se facilitaría la comunicación con estas y que los servidores en los que se desplegará la

solución final actualmente son servidores de aplicaciones Java, siendo costosa la migración a otra tecnología.

Gestor de bases de datos PostgreSQL

Sistema de gestión de alto desempeño y utilización, en correspondencia con las líneas tecnológicas dictadas por la UCI. Se destaca por ser, de los SGBD seleccionables, el único que tiene todas sus versiones liberadas bajo licencias libres.

Arquitectura en tres capas

Para la implementación de la arquitectura en tres capas de la aplicación se concibió el uso de tres tecnologías punteras dentro de la plataforma Java 2 Enterprise Edition: JSP, Spring e Hibernate.

En la capa de presentación se opta por la tecnología nativa de Java, que es Java Server Pages (JSP), surgida casi desde los mismos inicios del lenguaje, altamente validada, totalmente compatible con los servidores de aplicación más utilizados y de fácil integración con tecnologías más actuales como AJAX.

Como parte de la capa de la lógica del negocio se decide la utilización el framework Spring por permitir implementar varias filosofías consideradas necesarias dentro una aplicación web, independientemente de su tamaño. Entre estas se pueden citar el soporte al patrón Inyección de dependencias y a la Programación Orientada a Aspectos. Por encima de todo esto, las principales razones de la selección son la ligereza, la transparencia y el ser no intrusivo.

Para la implementación del acceso a datos en la aplicación, se selecciona el framework Hibernate, entorno de trabajo que tiene como objetivo facilitar la persistencia de objetos Java en BD relacionales. Es considerado por gran margen el framework más popular de persistencia de Java y uno de los más populares en sentido general.

Otras herramientas

Para el desarrollo de la aplicación se utilizó como herramienta CASE el Visual Paradigm for UML 6.4 Enterprise Edition y el IDE RedHat Developer Studio 2.0

RC1, herramientas de amplio éxito en aplicaciones con perfil tecnológico semejante al actual.

Finalmente, en la construcción y despliegue se decide utilizar el servidor de aplicaciones Apache Tomcat 6.0.14, una de las variantes más utilizadas a nivel global para las aplicaciones con el perfil tecnológico anteriormente presentado.

RESULTADOS Y DISCUSIÓN

Concepción de la herramienta de recomendación

Propuesta de la herramienta

Desde hace dos años, en la facultad 8 de la UCI y disponible para todo el país, se encuentra funcionando un calificador automático que contiene una amplia colección de ejercicios a resolver, clasificados por temas. Este recibe como entrada un envío o interacción, que consiste en el código-solución a problemas planteados y tras probar con juegos de datos predefinidos, emite un veredicto acerca de la correctitud del programa enviado.

El veredicto puede tomar, entre otros, los valores de problema aceptado, si el programa resuelve de manera correcta todos los juegos de datos; tiempo límite excedido, si su tiempo de corrida es mayor que el predeterminado por el calificador; error en tiempo de ejecución, si aparece un error a la hora de procesar los datos; y respuesta incorrecta si la salida del programa no se corresponde con la salida esperada.

Actualmente este jurado constituye la herramienta central en la preparación para concursos y en las prácticas como parte de las asignaturas de programación. La aplicación que se propone se integra a este y puede ser utilizada por el estudiante como punto de apoyo en el sentido de recibir sugerencias respecto a los problemas a resolver.

El flujo principal viene dado en que al entrar inicialmente el estudiante en contacto con esta, se lleva a cabo una comunicación a través de servicios web con el jurado online y con la información que obtiene de esta interacción, se determinan los elementos a ser recomendados, siendo mostrados al usuario a través de una interfaz amigable. Finalmente, el usuario realiza en el jurado online las actividades sugeridas, almacenándose el desempeño en la base de

datos del jurado online, siendo posteriormente replicado hacia la base de datos del sistema recomendador.

Construcción

La construcción de un sistema de recomendación aplicado al e-learning se puede dividir en cinco pasos fundamentales: recolección de los datos, ETL (extracción, transformación y carga), generación del modelo, configuración, y administración del servicio. En el caso particular del sistema actual, existe una correspondencia entre cada una de estas etapas y determinadas funcionalidades, considerándose que la realización del proceso mencionado debe ser constantemente cíclico, iniciado la ejecución de la primera actividad una vez que se termine con la última, en la iteración anterior.

Recolección de los datos

En el caso del sistema de recomendación en propuesta, resulta evidente que una buena parte de los datos iniciales que requiere el sistema están ya contenidos en el jurado online de programación, particularmente los que relacionan los usuarios con los problemas propuestos, localizándose en tablas como users, solution y problem.

Por otra parte, para la captura de los datos no inherentes al jurado online y sí necesarios para la aplicación en cuestión, como son principalmente las preferencias de cada usuario y las peculiaridades de cada problema, se conciben en respectivos casos de uso.

ETL

Una vez tenida a disposición toda la información, sólo se traslada para la base de datos de la herramienta la necesaria para satisfacer en cada momento los requerimientos del usuario. Asimismo, se implementa un sistema de replicación para que a medida que este vaya progresando en el jurado online, la herramienta vaya reconociendo los nuevos cambios de estado del usuario y de esta forma realice en todo momento las recomendaciones acorde a su situación actual.

Generación del modelo

Resulta imposible realizar el proceso de recomendación tomando como referencias los datos planos del modelo relacional de la base de datos de la

aplicación en cuestión. Para esto, es necesario generar un nuevo modelo, el cual va a estar intrínsecamente relacionado con los algoritmos de recomendación a utilizar. En estos momentos se profundiza en el desarrollo de este modelo, no obstante, se realiza una exposición bastante detallada del estado del trabajo, y hacia dónde se dirige la concepción del mismo.

Configuración

En el sistema actual, este paso es indudablemente el de menor peso como parte del proceso completo. En la etapa de configuración, los diferentes modelos y algoritmos se despliegan a través de diferentes estrategias de recomendación, para suministrar diferentes servicios. Particularmente en el caso actual, inicialmente se ofrecerá un único servicio, que es el de la sugerencia de problemas. Se hace a su vez referencia al algoritmo a utilizar, basado en técnicas de filtrado colaborativo.

Suministración del servicio

La suministración del servicio, que en pocas palabras lo constituye la recomendación de ejercicios a resolver, acordes al perfil actual del usuario, se decidió implementar a través de un caso de uso en el que se plasma toda la teoría descrita anteriormente.

Este caso de uso contiene tres flujos principales: el de «Problemas a resolver», que constituye el principal y no sólo del caso de uso sino de la aplicación completa, puesto que en este se mostrarán los problemas sobre los cuales el usuario actual tiene mayor posibilidad de éxito. Otra sección no menos importante lo constituye la de Realizar recomendación, donde se le brindará a este la posibilidad de resaltar, dentro de todos los problemas que ya ha resuelto, aquellos que él considere de fácil y provechosa resolución, para todos los usuarios de perfil semejante al suyo. Finalmente, se decidió la inclusión de un ShoutBox, para facilitar la comunicación entre usuarios semejantes.

Así, con el uso de este sistema de filtrado colaborativo, el estudiante obtiene una aproximación mucho más específica de las actividades que debe realizar para obtener un aprovechamiento máximo de los recursos disponibles en el jurado online.

Trabajos Relacionados

Desde el surgimiento de las herramientas automáticas de calificación, hasta la actualidad, no se han desarrollado muchas aplicaciones semejantes a la que aquí presentamos. No obstante, las pocas disponibles indican su utilidad y necesidad para la comunidad de programadores en general, tanto de mayor como de menor nivel.

Entre las aplicaciones de recomendación más representativas están el ACM Problem Grading, de Sebastian Urbaniak, el Next2Solve, de Igor Naverniouk, y el Problem classification, en el Sphere Online Judge. A continuación se hará un breve análisis de cada uno de ellos.

ACM Problem Grading.

Esta herramienta, así como el Next2Solve, fueron implementadas en apoyo al jurado online de la Universidad de Valladolid (Uva Online Judge), el que constituye uno de los de mayor relevancia, entre todos los disponibles en la www. La aplicación se limita a proporcionar un listado de problemas a resolver, ordenado por un score de simplicidad que se calcula a partir de la cantidad de submisiones de cada problema, y su porcentaje de soluciones aceptadas. No se enfoca en un usuario en particular, y todos los criterios que utiliza son completamente estáticos y a la vez desconocidos.

Next2Solve

La novedad de esta aplicación viene dada en que es la primera que asume el camino de concentrarse en los usuarios específicos, dando un listado personalizado de actividades a resolver para cada usuario. No obstante, el propio autor, Igor Naverniouk, admite que para el desarrollo de la misma reutilizó técnicas aplicadas anteriormente por Urbaniak en su *ACM Problem Grading*, lo que en adición a su simplicidad, hace pensar en que presenta las mismas limitaciones de poca flexibilidad y estaticidad de los datos.

Actualmente dicha aplicación se encuentra fuera de servicio, al depender del script de Urbaniak que desde hace tiempo también tampoco se encuentra disponible.

SPOJ Problem Classification

Esta herramienta aparece adjunta al jurado online de la Universidad Tecnológica de Gdansk, en Polonia, probablemente el más importante después del Uva Online Judge. A diferencia de las anteriores, en vez de enfocarse en el usuario, se enfoca en los ejercicios, al permitir el etiquetado de estos, de forma anónima. A través de dicho etiquetado, a cada problema se le asocian determinados temas, pudiéndose posteriormente realizar un filtrado por diferentes criterios, obteniéndose como resultado los ejercicios correspondientes.

A pesar de constituir un importante paso de avance en este tipo de aplicaciones, el hecho de realizarse la mayoría de las funcionalidades por parte del usuario de forma anónima, constituye una limitante importante, pues al no poder indicar un paquete específico de actividades a realizar por usuario, imposibilita un mejor aprovechamiento por parte de este.

Tras analizar tres de los ejemplos más destacados en el campo de la recomendación aplicada a los jurados online, se puede afirmar que estos se encuentran aún en un estado rudimentario y bastante primitivo, no a la par del desarrollo actual de muchas otras tecnologías de e-Learning. Esto resalta el impacto y la relevancia que pudiera tener el trabajo que se presenta, en el cual se intenta combinar de manera efectiva los metadatos asociados a los usuarios con los asociados a los problemas, desde un enfoque del problema general de la recomendación.

CONCLUSIONES

En la era de la web 2.0, la educación se vuelve cada vez más interactiva y los sistemas de e-Learning son los encargados de dar el paso al frente, ante las cada vez más exigentes necesidades de los educandos. Siguiendo estos principios, se ha presentado un sistema de recomendación basado en técnicas de filtrado colaborativo, para la recomendación hacia los estudiantes, de los principales recursos disponibles en un jurado online de programación. Como parte de los anexos se muestran algunas vistas de la aplicación obtenida.

Todas las pruebas que se le han realizado hasta el momento a la aplicación, enfocadas tanto al software en sí como a la interacción con el estudiante, han resultado exitosas. No obstante, aún se dista mucho de conocer el grado de efectividad del mismo y su influencia real sobre el estudiante, encaminándose hacia esto posibles futuras investigaciones. El software se encuentra actualmente desplegado y disponible para la comunidad universitaria, encaminándose los esfuerzos a la integración con el jurado oficial de la Sub-Región para el Caribe de la ACM-ICPC.

BIBLIOGRAFÍA CONSULTADA

- GARCÍA SALCINES, C.: *Sistema recomendador colaborativo usando minería de datos distribuida para la mejora continua de cursos e-learning, IEEE-RITA*. 2008.
- ADOMAVICIUS, G. Y TUZHILIN, A.: *Recommendation Technologies: Survey of Current Methods and Possible Extensions*, Stern School of Business. 2004.
- TAN, H.; GUO, J. Y LI, Y.: *E-learning Recommendation System, International Conference on Computer Science and Software Engineering*. 2008.
- HERLOCKER, J. Y GROUPLENS: *Applying Collaborative Filtering to UseNet News., Communications of the ACM*. 1997.
- O' BRIEN, J.M.: *The Race to Create a 'Smart' gGoogle, Fortune*. 2006.
- TAEK-HUN, Y.: *An effective Recommendation Algorithm for Improving Prediction Quality, Lecture Notes on Artificial Intelligence*. 2006.
- TORVEEN, H. Y BEYOND RECOMMENDER SYSTEMS: *Helping People Help Each Other, HCI in the New Millennium*. 2001.
- VELEZ-LANG, O. Y SANTOS, C.: *Sistemas recomendadores: Un enfoque desde los algoritmos genéticos., Industrial data*. 2006.
- YERA TOLEDO, R.: *Concepción de una estrategia para un sistema adaptativo de preparación de estudiantes en los concursos de programación, COMPUMAT*. 2009.
- Galán Nieto, G.: *Filtrado colaborativo y sistemas de recomendación, IRC*. 2007.
- Ya Tang, T. y McCalla, G.: *Smart Recommendations for an Evolving e-Learning System, International Journal on E-Learning*. 2005.

Junco Vázquez, H.: *Un jurado online, una nueva forma de ejercitación y evaluación en las asignaturas de Programación, UCIENCIA. 2007.*