

MEMORIA DE ACTIVIDADES REALIZADAS

1. Identificación de las prácticas

Nombre del alumno: Adriano José Martín Gutiérrez

Empresa: Grupo de Sistemas Inteligentes. Departamento de Ingeniería de Sistemas Telemáticos, Universidad Politécnica de Madrid

Nombre del Tutor: Carlos Ángel Iglesias Fernández

Nombre y Departamento del Supervisor Académico: Carlos Ángel Iglesias Fernández, Departamento de Ingeniería de Sistemas Telemáticos.

Modalidad: 4. Beca en la Escuela, en el marco de un contrato o convenio.

2. Descripción de las prácticas

2.1. Objetivos técnicos

Las prácticas realizadas se han centrado en el marco de agentes de inteligentes y más concretamente en los agentes de recomendación.

El agente de recomendación que se ha desarrollado se corresponde con la parte servidor de una aplicación para la plataforma Google TV. Este agente ofrece al usuario que use el servicio los contenidos disponibles en la web que realmente le interesen. Para ello, el agente se apoya en un conjunto de valoraciones previas con las que el usuario ha puntuado contenidos que haya visionado anteriormente. Con estas valoraciones se elaborará un perfil de usuario y se buscarán a usuarios de perfiles parecidos dentro del servicio, mediante la combinación de distintas métricas de recomendación.

La combinación de los distintos algoritmos de recomendación existentes en este tipo de agentes junto con la repercusión social observable en las distintas redes sociales dará como resultado una puntuación para cada contenido. Esta puntuación será el indicador de cuánto le puede llegar agradar el contenido al usuario. Por lo que finalmente se le presentan al usuario a través de la interfaz del servicio aquellos contenidos no visionados y con mayor puntuación.

El proyecto ha buscado contribuir a la experiencia de usuario, ofreciéndole una interfaz intuitiva y en la que se fomenta su participación. Además, la interfaz pretende facilitar al usuario el uso del servicio haciendo transparente toda la parte desarrollada por el servidor y siendo de un manejo lo más simple posible.

En etapas posteriores a este período transcurrido se extenderá el mercado de este servicio a tabletas y smartphones para ampliar su cuota de usuarios.

2.2. Objetivos formativos

El principal objetivo formativo de estas prácticas, ha sido el de familiarizarse y obtener competencias en la utilización de las tecnologías específicas para los objetivos técnicos mencionados. En concreto, la práctica y el dominio de algoritmos de recomendación que han sido utilizados por la parte servidor de la aplicación. Entre los algoritmos desarrollados destacan los siguientes: algoritmos de filtrado colaborativo, algoritmos de recomendación social y algoritmos genéticos.

La práctica también ha servido para una mayor especialización en las tecnologías subyacentes. En concreto, se ha invertido gran cantidad de tiempo en: programar en lenguaje Java, en el manejo de estructuras de datos en lenguaje JSON, así como en el estudio del proyecto Mahout desarrollado por Apache.

Menos intensivamente, se han utilizado tecnologías “scrappers” para extraer los contenidos del conjunto de fuentes seleccionadas previamente. Los “scrappers” extraen asimismo información de carácter social sobre los contenidos a partir de las más conocidas redes sociales para evaluar su repercusión.

También se ha destacado la importancia de los procedimientos a seguir en el desarrollo del sistema software. Desde la recogida de requisitos y el diseño de la arquitectura hasta la implementación final, puntos esenciales para el desarrollo de proyectos software de cierta complejidad.

En una última consideración se pondrá en práctica la formación en la elaboración de documentación con respecto al proyecto, tanto de forma previa como a posteriori. Resaltando la importancia de recoger las características del servicio tanto en documentos como en entregables.

3. Actividades Realizadas

Durante el periodo de prácticas, se ha abarcado el desarrollo de la plataforma de acceso a contenidos digitales de noticias a través de televisión interactiva conocido como Noticias TVi. Más concretamente en este período de prácticas se ha abarcado la creación de servicio hasta su versión beta.

En estas semanas, hemos abarcado la etapa inicial del proyecto en la que nos hemos centrado en el análisis y primer trabajo con las tecnologías que se utilizarán en el sistema, así como de definir las funcionalidades y requisitos que cubrirá el mismo.

En primer lugar, se han definido el alcance y los objetivos del proyecto en diversas reuniones de equipo. Tras la definición formal del mismo y el acuerdo de los objetivos anteriormente descritos, se decidió centrarse en la aplicación de las tecnologías de agentes de recomendación para aconsejar contenidos en función de una base de datos de preferencias de los usuarios y el historial de cliente.

Así mismo, se acordó para el escenario de la aplicación que el sistema pudiese obtener información de las distintas redes sociales con el fin de ofrecer un servicio más completo.

Personalmente mi tarea ha consistido en el desarrollo de la parte servidor del servicio, es decir, la implementación de los algoritmos de recomendación estudiados, las primeras pruebas y la realización de los informes correspondientes.

A continuación se pasa a detallar más detenidamente los aspectos que se han cubierto durante el desarrollo de estas prácticas:

3.1. Puesta en común de requisitos, decisión de tecnologías a utilizar y preparación de documentos preliminares con definición del proyecto (semanas 1 y 2).

Al comienzo del proyecto, se realizaron diversas reuniones de equipo para establecer las definiciones formales de las características y objetivos del proyecto. Se acordaron los términos anteriormente presentados, así como una primera lista de tecnologías de recomendación a estudiar para su posterior utilización en el proyecto.

Como documentación adicional, se definieron las interfaces de las que constaría el sistema. En concreto, se ha comenzado por una implementación para Android en Google Tv, siendo posible posteriormente la implementación de un cliente en dispositivos móviles o tabletas. Para esta tarea, se han realizado diversos wireframes donde se recoge una descripción preliminar de las vistas y funcionalidades de la interfaz, que sirven además como especificación para la posterior implementación de las tecnologías.

3.2. Definición de arquitectura y reparto de responsabilidades en el equipo (semana 3).

Tras definir el alcance y las características del proyecto, se realizó una división modular de las diversas funcionalidades para definir la arquitectura del sistema. Los distintos subsistemas se definieron teniendo siempre en mente la principal idea del proyecto: ofrecer una interfaz intuitiva y sencilla que permite al usuario ver en cada momento el conjunto de contenidos que él desea observar entre los que están disponibles en ese momento o por los cuales se sienta realmente atraído, fomentando además su participación.

El reparto de trabajo en el equipo resultó convenientemente sencillo, debido a la característica eminentemente modular del sistema.

3.3. Investigación de tecnologías para las partes asignadas de forma individual. Descripción de estado del arte y planificación de desarrollo del subsistema (semanas de la 4 a la 9).

En esta etapa del proyecto, se han investigado diversas tecnologías consideradas apropiadas para la implementación de los diversos subsistemas anteriormente descritos. En concreto, el trabajo ha consistido en el estudio de distintos algoritmos o sistemas de recomendación entre los que destaca:

- **Mahout**, proyecto proporcionado por Apache y que nos proporciona un conjunto de librerías y métodos que nos permite realizar fácilmente procesos de filtrado de contenidos, para la generación de recomendaciones. Estas se basan en el uso del algoritmo de filtrado colaborativo.
El algoritmo de filtrado colaborativo, funciona internamente, recomendando nuevos contenidos para un usuario basándose en sus elecciones anteriores y elecciones de gente con similar historial de valoraciones a los contenidos.
- **Algoritmos genéticos**. Inspirados en la evolución biológica y su base genético-molecular. Estos algoritmos hacen evolucionar una población de contenidos someténdola a acciones aleatorias semejantes a las que actúan en la evolución biológica. Lo que nos permitirá ofrecer al cliente de la aplicación los contenidos más óptimos para su visionado.
- **API Twitter**, dada su componente social y la gran propagación de las noticias a través de esta plataforma, Twitter es una herramienta muy útil para comprobar la popularidad de una noticia en la red.
- **API Facebook**, desde sus inicios Facebook ha sido una herramienta que muestra a distintas personas y la relación entre ellas. Con el tiempo esta idea se ha ido desarrollando y se ha aumentado este grafo social, de tal forma que actualmente incluye a los usuarios y las conexiones que tienen con lo que les importa.

3.4. Puesta en común de trabajo anterior y planificación de siguientes etapas (semana 10).

En esta etapa se realizó una evaluación de los resultados obtenidos en la etapa anterior, para poder así realizar los cambios necesarios y la planificación adecuada de las tareas siguientes.

3.5. Implementación de subsistemas (semanas de las 11 a la 17).

En esta etapa, se comenzó a implementar las funcionalidades de los subsistemas requeridos tras el estudio de las tecnologías a utilizar realizado en las etapas anteriores.

Para una presentación preliminar de la funcionalidad del sistema, se realizó un mockup del servicio que presentaba las interfaces y funcionalidades básicas del sistema.

La implementación de las tecnologías se han llevado a cabo de la siguiente manera:

- **Mahout**. El algoritmo de filtrado colaborativo, funciona internamente, recomendando nuevos contenidos para un usuario basándose en sus elecciones anteriores y elecciones de gente con similar historial de valoraciones a los contenidos. Para ello nuestro sistema recoge la satisfacción obtenida por el usuario mediante un sistema de marcado que le permite señalar como favorito aquel contenido cuyo visionado le haya gustado especialmente.

Cuando ya tenemos la suficiente información el algoritmo utiliza el coeficiente de correlación de Pearson, que nos devuelve un peso (valoración) estimado del grado de satisfacción que alcanzará el cliente al visionar un determinado contenido.

En nuestro caso, y debido a que en teoría tendremos una base de datos con un gran número de clientes. Para el cálculo de este coeficiente utilizaremos el subconjunto de usuarios más parecidos a aquel que realiza la petición, lo que conlleva a una mejora tanto de la eficiencia como de la efectividad.

Para poder llevar a cabo lo anteriormente descrito será necesario obtener a priori el perfil de usuario que esté usando la aplicación y la información relativa a cada contenido, respecto a las valoraciones recibidas y de que clientes.

- **Algoritmos genéticos.** Para llevar a cabo el uso de esta tecnología, nuestro sistema se ayuda de una herramienta denominada JGAP (Java Genetic Algorithms Package) que provee mecánicas genéticas básicas que pueden combinarse para aplicar principios evolutivos a la resolución de problemas. Inicialmente son necesarios dos conjuntos de datos para conseguir un buen funcionamiento por parte del algoritmo: la oferta global de contenidos y el conjunto de valoraciones estimadas para el usuario a los contenidos disponibles.

A partir de esto el algoritmo genético genera una población inicial de individuos tomados de los contenidos de forma aleatoria. El espacio de soluciones del problema se conforma entonces por todos los individuos posibles, que son todas las combinaciones posibles de contenidos. Siendo el objetivo encontrar en estos conjuntos aquel individuo que representa la solución óptima del problema, en términos evolutivos aquel que sobrevive al final.

- Se han utilizado la **API de Twitter y Facebook** para completar para hallar que los contenidos más populares en un momento determinado.

Finalmente y cuando se termine de integrar completamente el servicio deberá tener las siguientes funcionalidades: recomendación de contenidos basados en similitud de usuarios, recomendación social, parrilla (conjunto de contenidos recomendables), contenidos nuevos, contenidos populares y gestión de preferencias.

3.6. Integración de subsistemas y pruebas (semanas 18 y 19).

Durante esta etapa, además de continuar con la implementación de los subsistemas (interfaz y servidor), se comenzó a evaluar la integración de los mismos en el sistema global.

Igualmente, se realizaron pruebas unitarias sobre los subsistemas desarrollados y generales sobre las funcionalidades deseadas con usuarios de prueba, hasta el punto en el que el nivel de implementación lo permitía.

3.7. Evaluación beta de los resultados (semana 20).

Durante esta semana se realizó una evaluación exhaustiva con usuarios de la versión beta del proyecto y se analizaron los resultados para observar que se cumplieran correctamente las funcionalidades del sistema.

Durante el transcurso del desarrollo del servicio se han ido igualmente elaborando informes regulares sobre la elaboración del proyecto.

El proyecto se podrá extender más allá de la semana número 20, mejorándolo tras la experiencia de los primeros usuarios beta con la inclusión de nuevas fuentes de contenidos e incluyendo nuevos mecanismo de recomendación.

4. Métodos de trabajo, medios materiales y herramientas

El método de trabajo en equipo se ha basado en la planificación de responsabilidades y tareas mediante la utilización de sistema de organización de proyectos Pivotal Tracker. Así mismo para compartir el código y las distintas versiones del proyecto, tanto de la parte cliente como la del servidor, se ha usado un repositorio github al que tienen acceso todos los colaboradores.

Por otro lado el grupo de sistemas inteligentes, ha puesto a mi disposición todas las herramientas necesarias para el desarrollo del servicio. El laboratorio dispone de un ordenador personal con un sistema Linux, las herramientas necesarias de programación Java (usándose en este caso Eclipse), una amplia colección de libros y toda la información disponible de forma gratuita internet entre los que se destacan los artículos de investigación.

Todas las tecnologías utilizadas durante las prácticas son libres y de código abierto, por lo no ha sido necesario el uso de licencias.