

# Sistema de recomendaciones literarias basadas en un perfil social



E. Sánchez

Máster Universitario de Aplicaciones Multimedia

Universitat Oberta de Catalunya (UOC)

Una propuesta de investigación para el

*Trabajo de Fin de Máster*

Junio 2017

---

# Índice

<b>Lista de Figuras</b>	<b>v</b>
<b>1 Introducción</b>	<b>1</b>
1.1 Justificación . . . . .	1
1.2 Objetivo principal de investigación . . . . .	3
1.3 Objetivos . . . . .	4
1.4 Metodología de investigación . . . . .	5
1.4.1 Estrategía de investigación . . . . .	5
1.4.2 Técnicas de generación de datos . . . . .	6
1.4.3 Evaluación . . . . .	7
1.5 Plan de investigación . . . . .	8
<b>2 Estado del Arte</b>	<b>11</b>
2.1 Definición, origen y clasificación de redes sociales . . . . .	11
2.2 Tipología de redes sociales . . . . .	12
2.3 Principales redes sociales en la actualidad . . . . .	13
2.4 Investigación en el ámbito de las redes sociales . . . . .	15
2.4.1 Utilizando las redes sociales para generar valor . . . . .	15
2.4.2 Análisis y procesamiento de datos . . . . .	16
2.4.3 Añadiendo otras fuentes de información para generar valor . . . . .	17
2.5 Sistemas de recomendación en el ámbito empresarial . . . . .	19
2.6 Seis Grados de Separación . . . . .	21
<b>3 Propuesta</b>	<b>23</b>
3.1 Nuevo sistema de recomendación . . . . .	23
3.1.1 Entrada A: Perfil Social . . . . .	26

## ÍNDICE

---

3.1.1.1	Perfil . . . . .	29
3.1.1.2	Conexiones . . . . .	33
3.1.1.3	Actividad . . . . .	36
3.1.2	Entrada B: Base de Datos del ISBN . . . . .	41
3.1.3	Integración con el ISBNdb . . . . .	43
3.1.4	Estrategia para emitir recomendaciones . . . . .	45
3.1.5	Salida del sistema de recomendación . . . . .	47
3.2	Cuestionario para captar usuarios . . . . .	54
3.3	Encuesta para evaluar el sistema . . . . .	57
<b>4</b>	<b>Conclusiones</b>	<b>61</b>
<b>5</b>	<b>Apéndice</b>	<b>65</b>
5.1	Integración con Twitter . . . . .	65
5.2	Integración con Facebook . . . . .	66
5.3	Integración con ISBNdB . . . . .	67
	<b>Bibliografía</b>	<b>69</b>

# Lista de Figuras

1.1	Diagrama de Gantt. . . . .	10
3.1	Diagrama de Flujo del Sistema de Recomendación. . . . .	25
3.2	Diagrama de Bloques de la Entrada A del Sistema de Recomendación. . . . .	29
3.3	Diagrama de la Salida del Recomendación. . . . .	53

## LISTA DE FIGURAS

---

# 1

## Introducción

### 1.1 Justificación

En Agosto de 2015 un billón de personas se conectó a Facebook en un sólo día (1). El número equivale a una séptima parte de la población del planeta Tierra. Es un dato impresionante y muy revelador, que muestra el grado de penetración de las redes sociales en la sociedad actual. Cada uno de esos usuarios entró en la red social para compartir información, publicar una opinión sobre un tema o simplemente dar a conocer un gusto o afición. De este modo, esos usuarios generan cada día miles de millones de datos que pueden ser utilizados desde un punto de vista científico para generar valor, de forma similar a como explica Jennifer Rowley (2) con la Pirámide del Conocimiento o Jerarquía DIKW. Las redes sociales almacenan esa gran cantidad de datos sobre sus usuarios, los cuales pueden ser procesados convenientemente para generar información acerca de los gustos o aficiones de éstos. La información generada a partir de esos datos puede ser analizada hasta cumplir el objetivo marcado, mediante el cual será posible conocer el comportamiento de los usuarios de esas redes sociales, y lo que es más importante, poder llegar a determinar si una afición gusta a un usuario o a otro y conocer las características de la misma. Finalmente, si se utiliza ese conocimiento desde un punto de vista científico junto a otras fuentes de información, se podrá alcanzar un grado de sabiduría que permita aportar algún beneficio a esos usuarios, y todo ello gracias a las redes sociales.

Este proceso, como se verá en el capítulo sobre el estado del arte de esta investigación, ha permitido a científicos e investigadores llevar a cabo estudios que ponen de

## 1. INTRODUCCIÓN

---

manifiesto la utilidad de las redes sociales en la sociedad actual para generar valor. Y esta es la principal justificación para llevar a cabo esta investigación, utilizar las redes sociales a las que día tras día acceden millones de usuarios para ofrecer a éstos un beneficio.

Llevado al caso concreto de esta investigación y utilizando nuevamente la Pirámide del Conocimiento, en este caso los datos se obtendrán de las principales redes sociales existentes en la actualidad : Facebook, Twitter o LinkedIn. La información generada será obtenida tras procesar esos datos y se centrará en los gustos o aficiones de los usuarios de esas redes sociales. Analizando esta información, será posible conocer qué usuarios son aficionados a la literatura y sobre qué temas, descartando a otros usuarios que no tienen esta afición. Finalmente, con ese conocimiento, y utilizando una base de datos pública como es el ISBN (3) se podrá alcanzar la sabiduría necesaria mediante la cual será posible recomendar a esos usuarios obras o autores que coincidan con sus gustos. Como se comentaba al principio, esto permitirá generar un beneficio para los usuarios de las redes sociales.

Sin duda, esta es la principal justificación para emprender esta investigación, aunque no es la única. Al comienzo de este apartado se comentó el hito alcanzado por Facebook en Agosto de 2015, es un dato reciente que ha ido en aumento y que hace ver la importancia de las redes sociales en la actualidad, sin embargo, éstas han sido objeto de estudio desde hace muchos años por parte de numerosos científicos, antes incluso de que se dieran a conocer tal y como son ahora. Hubo numerosos investigadores que estudiaron las redes sociales desde un punto de vista puramente científico sin la pretensión de obtener valor a partir de ellas pero con el importante objetivo de desgranar su estructura, como están conformadas. Los primeros científicos que justificaron su trabajo de esta forma, lo hicieron mucho antes del nacimiento de Facebook, pero sentaron las bases para investigaciones posteriores. En 1930, el escritor húngaro Frigyes Karinthy escribió el cuento Chains (4) donde formula la hipótesis de los Seis Grados de Separación, muy ligada a la definición de red social. Esta hipótesis fue tema central de investigación de varios científicos, como el psicólogo americano Stanley Milgram (5) con su experimento del mundo pequeño o el sociólogo Duncan Watts (6). Estos investigadores no conocían las redes sociales como son en la actualidad, pero intentaron demostrar esta hipótesis, afirmando que cualquier persona en el planeta puede estar conectada con cualquier otro individuo a través de una cadena de conocidos que no tiene más de cinco

## 1.2 Objetivo principal de investigación

---

intermediarios, seis si incluimos a la propia persona. Es una definición que se acerca mucho a lo que hoy se conoce como Facebook o Twitter. De este modo, esta hipótesis ayuda a visualizar a una red social como un conjunto de nodos interrelacionados entre sí. Estos nodos se pueden visualizar como usuarios, y cada uno de ellos almacena una gran cantidad de información relativa a su actividad dentro de la red. A partir de esta teoría es posible formular muchas preguntas, más allá de si es cierto que entre un nodo y otro existen como máximo otros 5 nodos, también es cuestionable si esa separación depende del contexto o información que almacena cada usuario. Hace algunos años era muy difícil poder contestar a estas preguntas, sin embargo, en la actualidad existen herramientas que, como se verá en el capítulo sobre el estado del arte, han sido utilizadas por la comunidad científica para intentar encontrar respuestas. Por tanto, esta es otra de las justificaciones para emprender esta investigación, y quizás la que desde un punto de vista más científico pretende aportar conocimiento para entender mejor las redes sociales, al igual que otros muchos científicos lo hicieron en su momento, pudiendo aprovechar las herramientas que existen en la actualidad. Los científicos citados anteriormente aportaron su conocimiento en un momento en el que no existían Facebook, Twitter o LinkedIn para realizar investigación a través de sus datos, y quizás el nacimiento de estas y otras redes sociales se deba a esos investigadores. Aportar ahora conocimiento con esta investigación sobre como funcionan las redes sociales permitirá en un futuro que otros investigadores con recursos más avanzados puedan llevar a cabo su investigación y quizás sirva para que las redes sociales evolucionen más aún.

En definitiva, las dos principales justificaciones para emprender esta investigación son, por un lado profundizar desde un punto de vista científico en el concepto y naturaleza de red social, independientemente del tipo, con el objetivo de aportar una base científica que pueda servir a otros investigadores, y en concreto a esta investigación, para generar valor a la sociedad actual dentro de este ámbito.

## 1.2 Objetivo principal de investigación

Dado el perfil social de un usuario, el cual estará formado por las distintas cuentas que dicho usuario tiene en las principales redes sociales, el objetivo principal de la investigación será aportar valor en el ámbito de las redes sociales mediante un sistema que recopile los datos generados por el perfil social de ese usuario, los analice para

## 1. INTRODUCCIÓN

---

encontrar sus gustos o aficiones y sea capaz de determinar si dicho usuario es aficionado a la literatura. Si lo es, el sistema generará un conjunto de recomendaciones sobre autores y obras literarias cuya temática sea del agrado del usuario. Para ello, durante la elaboración del estado del arte, se analizarán los mejores sistemas de recomendación centrados en este ámbito con el fin de poder aportar mejoras sobre el resultado final que se le ofrecerá al usuario.

### 1.3 Objetivos

Teniendo en cuenta el objetivo principal de la investigación, formulado en el apartado anterior, se relacionan a continuación los objetivos que se pretenden alcanzar:

- Aportar una visión única de los datos que las distintas redes sociales almacenan sobre un usuario. Para ello será necesario analizar y explotar los mecanismos de integración que ofrece cada red social teniendo como objetivo simplificar esta integración múltiple mediante una solución que permita extraer información sobre el perfil social de un usuario con independencia de las redes sociales donde éste tenga actividad.
- Representar un determinado hobby o afición mediante la información catalogada en una base de datos pública, para lo cual será necesario analizar y explotar los servicios que se dispongan para permitir obtener datos sobre un autor u obra en concreto.
- Implementar un sistema con dos entradas y una salida. Las dos entradas, representadas en los dos objetivos anteriores, serán por un lado toda la información obtenida a partir de los datos del perfil social del usuario y por otro lado la información de la base de datos que represente el hobby o afición, en este caso la literatura mediante el ISBN. La salida del sistema dependerá de la primera de las entradas, es decir, si el usuario no es aficionado a la literatura, se determinará en la salida, mientras que si la literatura se encuentra dentro de sus gustos o aficiones, la salida del sistema será el conjunto de recomendaciones o sugerencias sobre autores u obras que pueden interesar al usuario.

Por último, en el apartado introductorio se citó la hipótesis de los Seis Grados de Separación, clave en el nacimiento del término red social, tal y como se conoce en la actualidad. Como parte del proyecto de investigación, se plantea probar esta hipótesis utilizando para ello el sistema que se va a desarrollar, objetivo principal de la investigación. Por tanto, se pretende dentro de un contexto determinado, que en este caso será el de usuarios aficionados a la literatura, probar el número de grados de separación entre unos y otros para poder demostrar esta hipótesis. Esto permitirá comprobar si se han alcanzado los objetivos anteriormente descritos, que pasan porque el sistema sea capaz de crear el perfil social de un usuario, con independencia de las redes sociales que lo conforman, representando una afición, como es la literatura, mediante el conjunto de información obtenido gracias a una base de datos pública, en este caso el ISBN.

## 1.4 Metodología de investigación

### 1.4.1 Estrategia de investigación

Como se ha comentado en apartados anteriores, el objetivo principal de la investigación se basa en desarrollar un sistema de recomendación que aporte un beneficio a los usuarios de redes sociales, con independencia del tipo o naturaleza de éstas. Para ello se analizarán los principales sistemas de recomendación, revisando las estrategias que utilizaron para emitir las recomendaciones, los algoritmos usados y las fuentes de datos empleadas, no sólo de aquellas propuestas de investigación que tenían un objetivo más puramente científico e incluso docente, sino también de otros trabajos llevados a cabo en empresas, que apostaron y siguen apostando por estos sistemas para mejorar sus modelos de negocio.

Por tanto, dado que se trata de generar valor mediante la creación de un producto de TI y tal y como recomienda Oates en su libro *Researching Information Systems and Computing* (7), se utilizará una estrategia de Diseño y Creación que permitirá cumplir con el objetivo más importante que se pretende alcanzar. Se llevará a cabo mediante una primera fase de concienciación del problema que se desea solucionar, en este caso más bien una necesidad, centrada en el avance dentro del ámbito de las redes sociales. A continuación, se pasará a una fase más creativa donde, de manera general, se aportarán las líneas más importantes mediante las cuales se pretende cubrir

## 1. INTRODUCCIÓN

---

la necesidad detectada en la fase anterior. Posteriormente se pasará a fases más propias de un ciclo de desarrollo de TI: análisis, desarrollo, evaluación y conclusiones.

En resumen y como se verá más adelante, la aplicación de esta estrategia será clave durante el proyecto de investigación y permitirá cumplir con los objetivos anteriormente descritos. No sólo consistirá en desarrollar el sistema de recomendación sino en evaluar su correcto desempeño mediante la fase de evaluación, que dará paso a una última fase donde se analizarán los datos obtenidos tras evaluar el nuevo sistema con el fin de determinar, entre otros, cuáles son las líneas de futuro de la investigación.

### 1.4.2 Técnicas de generación de datos

En el apartado anterior se ha descrito la estrategia que se utilizará para cumplir con los objetivos que se pretenden alcanzar en el proyecto de investigación, y que estará basada en el desarrollo del nuevo sistema y su posterior fase de evaluación que permita comprobar si se cumplen los requisitos y especificaciones definidos en las primeras fases de la estrategia elegida para aplicar en esta investigación. Como indica Oates (7), la estrategia de Diseño y Creación normalmente viene acompañada del uso de técnicas de generación de datos, y el caso de este proyecto de investigación no es una excepción. En concreto estas técnicas, que se detallarán a continuación, se aplicarán en la fase de evaluación para obtener los datos necesarios que permitan probar que el sistema funciona correctamente, porque como no puede ser de otro modo, será necesario conseguir usuarios reales que deseen participar en el proyecto de investigación y que cumplan con el perfil necesario para ello. Asimismo, la experiencia de estos usuarios será muy valiosa a la hora de extraer conclusiones que permitan medir el desempeño del nuevo sistema de recomendación para determinar si se han cumplido los objetivos que se pretendían alcanzar y definir posibles puntos de mejora. El procedimiento para poder conseguir los datos de usuarios que deseen participar en el proyecto, se basará en la siguiente técnica de generación de datos:

- **Cuestionario.** A través de un email enviado a una lista de contactos, se lanzará un cuestionario ubicado en una página web, que también estará disponible para cualquier usuario con acceso a Internet, desde donde se presentará la investigación en el contexto de las redes sociales con el objetivo de encontrar usuarios que deseen

participar en la fase de evaluación de este proyecto de investigación. Para ello, los usuarios deberán cumplir con los siguientes requisitos:

- Consentimiento para participar en el proyecto de investigación.
- Poseer un perfil en al menos una de las principales redes sociales.
- Permitir el acceso del nuevo sistema al perfil social del usuario, formado por una o varias redes sociales, con el único objetivo de poder extraer todos los datos que se dispongan, relacionados con el usuario, para poder probar el desarrollo efectuado durante una de las fases de la estrategia de Diseño y Creación.
- Destacar que no será necesario que el usuario sea aficionado a la literatura ya que precisamente lo que se desea evaluar, entre otros casos de prueba, es que el nuevo sistema sea capaz de determinar si la literatura se encuentra dentro de las aficiones de un determinado usuario.

A los usuarios que cumplan los requisitos anteriores, y que por tanto participen en el proyecto de investigación, se les solicitará, una vez que éste finalice, feedback a cerca de los resultados obtenidos, utilizando para ello la siguiente técnica de generación de datos:

- **Encuesta.** Mediante esta técnica se podrá solicitar al usuario feedback en relación a su participación en el proyecto de investigación mediante una encuesta cuyas preguntas intentarán obtener una valoración del usuario sobre los resultados o beneficios obtenidos gracias al nuevo sistema y las recomendaciones o sugerencias generadas por éste. Los resultados de esta encuesta serán de gran ayuda para definir las líneas de futuro de la investigación.

### 1.4.3 Evaluación

Como se ha comentado en puntos anteriores, en este período enmarcado dentro de la estrategia de Diseño y Creación se podrá determinar si el desarrollo que ha sido llevado a cabo cumple con los requisitos que se definieron en las primeras fases del ciclo de la estrategia. Por tanto, esta fase no servirá únicamente para realizar tests unitarios que verifiquen el correcto desempeño del nuevo sistema, según los casos de uso definidos durante la fase de análisis, sino para probar éste a través de una aplicación multimedia

## 1. INTRODUCCIÓN

---

disponible para los usuarios que deseen participar en el proyecto de investigación. Para ello, se utilizarán las dos técnicas de generación de datos comentadas en el punto anterior: cuestionario y encuesta, y por tanto, en el transcurso de esta tarea se deberá abordar el diseño y ejecución de cada una de estas técnicas, siguiendo las pautas ya comentadas.

### 1.5 Plan de investigación

En este apartado se definirá el plan de proyecto mediante el desglose de las principales fases que lo componen, cada una de las cuales irá acompañada de un número de tareas determinado. Destacar que todas las tareas están ordenadas cronológicamente en orden ascendente, tal y como se puede comprobar en el diagrama de Gantt de la figura 1.1:

1. **Propuesta de Investigación.** Se llevará a cabo durante el primer año del proyecto de investigación y seguirá la estructura del documento actual, con el objetivo de demostrar que la investigación sigue las líneas adecuadas.
2. **Diseño y Creación.** Se trata de la fase más importante dentro del plan de investigación ya que es la que engloba el desarrollo del nuevo sistema con el que se pretende alcanzar los objetivos marcados, utilizando para ello una estrategia de Diseño y Creación. En esta fase se incluyen las tareas de análisis, diseño, implementación, evaluación y conclusiones.
3. **Artículo Científico.** Aunque no se trata de una fase como tal dentro del plan de la investigación si supondrá el alcance de un hito dentro del proyecto y como tal se refleja en este plan de proyecto. Se podrá llevar a cabo gracias al conocimiento adquirido en la fase anterior con la utilización de la estrategia de Diseño y Creación. Se propone la redacción de dos artículos:
  - Un artículo que profundice en la hipótesis de los Seis Grados de Separación con los datos obtenidos en la fase de evaluación de la estrategia de Diseño y Creación y gracias al análisis cuantitativo realizado a partir de esos datos, pueda determinar el máximo número de nodos que separan a los usuarios que participen en la evaluación del nuevo sistema a través de la aplicación multimedia desarrollada a tal efecto.

- Otro artículo, de base científica y muy ligado a una de las justificaciones comentadas en el apartado introductorio de esta propuesta de investigación, donde se detalle la estructura de una red social, sea cual sea su tipo, y como ésta organiza su información en conjuntos de nodos interrelacionados. Todo ello, explicado a través del denominado perfil social, que engloba todas las cuentas que un usuario tiene en una red social, y que se abstrae de la definición de cada una de ellas.
4. **Redacción de la Memoria.** Con el proyecto de investigación alcanzando sus fases finales, esta fase abarcará la redacción de la memoria del proyecto de investigación, utilizando para ello toda la información obtenida en fases anteriores.
  5. **Defensa de la Tesis.** Será la fase que de por concluido el proyecto de investigación y consistirá en presentar el trabajo realizado durante todas las fases anteriores.
  6. **Actualizar Estado del Arte.** No se trata de la última fase del proyecto sino de una fase que estará vigente durante toda la investigación, desde el comienzo, cuando se prepare la propuesta de investigación con la elaboración del estado del arte, el cual se irá actualizando a medida que avance el proyecto con toda las fases y tareas que lo componen.

Teniendo en cuenta las fases y tareas anteriores que se han descrito, se ha elaborado un diagrama de Gantt que como punto de partida realiza las siguientes suposiciones:

- Duración: 4 años.
- Fecha de Inicio: 01/18.
- Fecha de Fin: 12/21.

De este modo, en la siguiente figura se muestra el desglose del proyecto en las fases y tareas comentadas anteriormente y planificadas en el tiempo en base a días de esfuerzo y la representación de las mismas en un diagrama de Gantt.

# 1. INTRODUCCIÓN

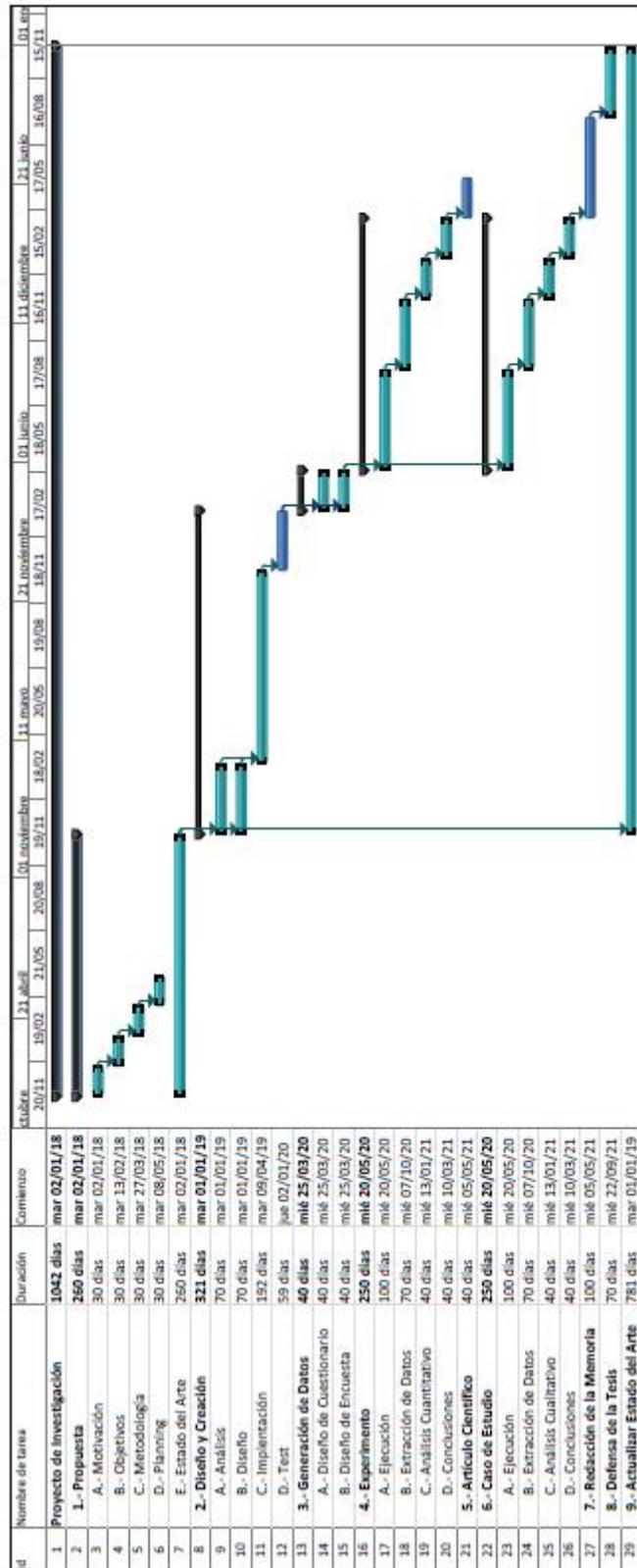


Figura 1.1: Diagrama de Gantt.

## 2

# Estado del Arte

## 2.1 Definición, origen y clasificación de redes sociales

Una red social es un servicio dentro de la web que permite a cualquier individuo llevar a cabo las siguientes acciones:

1. Construir un perfil público o semi-público dentro de un sistema acotado.
2. Articular una lista de otros usuarios con los que comparte una conexión.
3. Visualizar y rastrear la lista de conexiones de otros usuarios.

La definición anterior, formulada por los investigadores M. D. Boyd y N. B. Ellison (8), se acerca al fenómeno de red social ligado a la sociedad actual y representado por Facebook, Twitter y otras redes sociales muy conocidas. Sin embargo, para encontrar una definición sobre el término red social que no esté asociada a las ciencias de la computación, es necesario retroceder en el tiempo hasta finales del siglo XIX, tal y como detalla Linton Freeman (9), profesor e investigador del Departamento de Sociología de la Universidad de Northwestern en su libro titulado *El desarrollo del análisis de redes: Un estudio de sociología de la ciencia*. En este trabajo se describe una red social como una estructura compuesta por un conjunto de individuos que están relacionados de acuerdo a algún criterio. Es una definición que permite entender la verdadera naturaleza de una red social e incluso comprender su objetivo principal: relacionar usuarios en base a uno o varios criterios. Por tanto, esta última definición se aleja un poco del software y en concreto de las aplicaciones web con las que tanto se relaciona al término red social en los últimos años y que, sin embargo, le acerca a los trabajos de los primeros

## 2. ESTADO DEL ARTE

---

investigadores, que sin los avances actuales, sentaron las bases de lo que hoy conocemos como red social. De esta forma, Freeman centra el origen del término red social en los trabajos de Jacob Levy Moreno en 1930, a quién se le atribuye la creación de sociogramas para explicar las relaciones interpersonales. La explicación sobre como se estructuran las redes sociales actuales en conjuntos de nodos interrelacionados está en las obras de Moreno y otros investigadores y como se verá más adelante, los científicos que han profundizado en las redes sociales dentro de sus trabajos de investigación han tenido muy en cuenta estos trabajos y otros similares llevados a cabo en la primera mitad del siglo XX.

### 2.2 Tipología de redes sociales

No existe unanimidad a la hora de clasificar los distintos tipos de redes sociales, ni es posible encontrar una única clasificación dentro de las publicaciones sobre redes sociales de la comunidad científica. Sin embargo, la clasificación de redes sociales publicada por el estudio del Observatorio Nacional de las Telecomunicaciones y de la Sociedad de la Información de España (10) es la de mayor relevancia y además está presente en la mayor parte de la literatura que versa sobre este ámbito. Este estudio establece una tipología de redes sociales basadas en dos grandes grupos:

- **Redes Sociales Directas.** Formadas por conjuntos de usuarios que interactúan entre ellos. A su vez se pueden clasificar en otros tipos según su finalidad, con las redes de ocio y de profesionales, según su funcionamiento, con las redes de contenido, basadas en perfiles y redes de microblogging, según su grado de apertura, con las redes públicas y privadas y según su nivel de integración con las redes verticales y horizontales.
- **Redes Sociales Indirectas.** Son precursoras de lo que hoy en día se conoce como red social y en ellas, los usuarios no tienen un perfil accesible para cualquier otro usuario. El ejemplo más claro para definir las son los blogs o foros.

Dado que las redes sociales indirectas están cada vez más en desuso existe un gran número de referencias donde se toma como estándar de facto la clasificación de redes sociales según su nivel de integración en los dos tipos que se han comentado anteriormente: redes sociales verticales y redes sociales horizontales.

### 2.3 Principales redes sociales en la actualidad

En este caso la investigación se llevará a cabo con independencia del tipo de red social, ya sea directa o indirecta, horizontal o vertical, puesto que el objetivo principal que se pretende alcanzar se basa en lograr un avance en el ámbito de las redes sociales, sin poner foco en ninguna de ellas. Sin embargo, para conseguir esto, será necesario revisar las principales redes sociales utilizando para ello una clasificación de redes sociales de finales de 2015 (11), que está basada en el número de usuarios activos de cada red social. En este caso, y para no desvirtuar este estudio, se han excluido las aplicaciones de mensajería como WhatsApp o Line, consideradas en algunos casos como redes sociales, aunque éstas no ofrezcan mecanismos de integración que permitan explotar sus contenidos e incluso por su propia naturaleza, posean sistemas de privacidad que los alejan de cualquiera de las redes sociales actuales, sean del tipo que sean. A continuación se detallan las principales redes sociales que serán tomadas en consideración durante el proyecto de investigación:

1. **Facebook.** En el caso de Facebook (12), su entorno de integración para desarrolladores ofrece multitud de recursos que permite a compañías u organizaciones la doble comunicación con esta red social, tanto para explotar los contenidos que se almacenan sobre millones de usuarios mediante su Graph API (13) como utilizar los denominados plugins sociales para que los propios contenidos de la empresa interactúen con Facebook incluso a través de aplicaciones desarrolladas a medida que puedan ser integradas en la propia red social. En cuanto al proceso de integración de un tercero con Facebook (14), que puede ser extrapolable a cualquier otra de las redes sociales estudiadas, ya que se comportan de forma similar, se inicia con la obtención de credenciales en la red social mediante el protocolo OAuth (15). Una vez que se han obtenido las autorizaciones necesarias, la comunicación con la red social se realiza mediante un API (Application Programming Interface), normalmente basado en servicios REST (Representational State Transfer), y el intercambio de datos se lleva a cabo utilizando el formato JSON (JavaScript Object Notation).
2. **Instagram.** Esta red social ofrece, como en el caso de Facebook, un entorno para desarrolladores (16) con diversas herramientas que permiten la integración

## 2. ESTADO DEL ARTE

---

con los contenidos de esta red social, no sólo sobre las imágenes sino también incluyendo los metadatos asociados a ellas. Por ello, son varias las publicaciones que se pueden encontrar (17) donde se analizan estas herramientas de integración disponibles para que cualquier desarrollador pueda utilizar los contenidos en su propia aplicación.

3. **Twitter.** De un modo similar, Twitter (18) permite interactuar con su API (19) para explotar los contenidos que almacena esta red social por parte de un tercero. Al igual que la mayoría de redes sociales, las herramientas de integración son independientes de la aplicación externa que las va a explotar, por lo que ésta podrá estar desarrollada en cualquier plataforma, lo cual permite un alto grado de interoperabilidad.
4. **Pinterest.** Aunque se trata de una red social con menor recorrido que las anteriores, ha despertado el interés de investigadores por las herramientas de integración que ofrece (20) y que permiten, de un modo similar al de Instagram, analizar y procesar las millones de imágenes que publican sus usuarios y la información asociada a ellas (21).
5. **Linkedin.** Esta red social es un claro ejemplo de red profesional, según la tipología analizada en el punto anterior. Por ello, la información que contiene y que puede ser analizada con herramientas similares a las de los casos anteriores (22) ha despertado el interés de algunos investigadores (23) con el fin de desgranar el perfil profesional de un usuario y sus conexiones con otros individuos dentro del entorno empresarial.
6. **Google Plus.** De todas las redes sociales tratadas en puntos anteriores, esta es la que menor interés despierta en otros investigadores, motivado en gran parte por el menor volumen de usuarios que maneja en relación a otras redes sociales. Aún así, presenta un gran número de usuarios activos, como se puede observar en la clasificación citada anteriormente, motivado por el empuje del buscador Google, lo cual le sigue haciendo merecedora de entrar en este ranking. Aún así, si existen algunas referencias (24) en las que se ha utilizado las herramientas de integración de esta red social (25) para desgranar su estructura, que por otro lado presenta formas muy similares a las de los casos anteriormente citados.

### 2.4 Investigación en el ámbito de las redes sociales

#### 2.4.1 Utilizando las redes sociales para generar valor

Las referencias anteriores tienen un objetivo más didáctico, centrado en realizar un análisis completo de todas las posibilidades que ofrecen las herramientas de integración de las distintas redes sociales e incluso desgranar su estructura y como almacenan la información generada por sus usuarios. Sin embargo, existen otros artículos científicos que a través de estrategias de Diseño y Creación, Experimento y Caso de Estudio pretenden generar valor tomando como referencia y principal fuente de datos las redes sociales. Se trata de casos cuyos objetivos están más próximos a los que se pretenden alcanzar con esta investigación. De este modo, con Instagram (26) es posible publicar fotos y aplicar determinados filtros para modificar la apariencia original de cada imagen. Mediante la aplicación de estos filtros, los usuarios son capaces de crear un estilo que está relacionado con diferentes rasgos de su personalidad, tal y como determina un estudio en el que participaron 113 usuarios y 22.398 imágenes almacenadas en sus cuentas de Instagram. Para llevar a cabo este caso de estudio fue necesario que los usuarios permitieran el acceso a la aplicación externa desarrollada para la investigación, la cual utilizó el API de Instagram (16). Se pueden observar algunas similitudes con el caso de estudio que se plantea para esta investigación con el objetivo de probar el correcto funcionamiento del sistema de recomendación, sin embargo, atendiendo a los datos que se aportan en el caso de referencia parece difícil extraer conclusiones que puedan ser generalizadas con un rango de actuación tan bajo y limitado a una única red social. Dado que todas las redes sociales parten del mismo concepto, basado en los estudios de Moreno y otros, mencionados por Freeman y que han sido comentados al comienzo de este capítulo y ya que todas ofrecen herramientas de integración basadas en el mismo concepto tecnológico (REST API y OAuth) parece tener sentido abordar el proyecto de investigación con un punto de vista único en cuanto a la manera de operar con los entornos de integración de las redes sociales, independientemente del tipo de cada una de ellas. Si bien es cierto, que el caso planteado con Instagram ofrece pautas interesantes en cuanto al proceso para analizar y explotar datos de una red social en concreto.

No es el único caso relacionado con una investigación científica donde el objetivo principal es conseguir un avance en el ámbito de las redes sociales. En este caso,

## 2. ESTADO DEL ARTE

---

utilizando los mecanismos de integración que ofrece Pinterest (27), se llevó a cabo un estudio para tratar de entender como los usuarios gestionaban y compartían las acciones llevadas a cabo en esta red social, también llamados pines. Nuevamente encontramos un interesante estudio que pretende analizar el comportamiento de los usuarios que interactúan con una determinada red social intentado extraer conclusiones relacionadas con el comportamiento y características de los usuarios de Pinterest. Sin embargo, volvemos a encontrar una investigación en la que el foco se centra única y exclusivamente en una determinada red social, sin que se compare con otra u otras, pudiendo en ese hipotético caso extraer conclusiones más generalizadas y permitiendo generar un beneficio que pueda ser válido para el ámbito de las redes sociales con total independencia del tipo o naturaleza de red social.

Por último, destacar una referencia (28) que se basa en la red social más importante en la actualidad y que ofrece similitudes con el propósito de esta investigación en cuanto al hecho de utilizar las redes sociales para generar recomendaciones a un determinado usuario en función de su perfil social. De este modo, este experimento emite recomendaciones a un determinado usuario sobre los grupos de Facebook que pueden ser de su agrado. Sin embargo, al igual que en los casos anteriores, el perfil social utilizado y que representa al usuario, únicamente utiliza una red social y por tanto, se echa en falta una visión global de las redes sociales, que permita utilizar datos de todas ellas, o al menos de las más importantes, para que el valor generado o en el caso de la investigación, la salida del sistema de recomendación, pueda ser más precisa, con un porcentaje de acierto en cuanto a las recomendaciones emitidas mayor que el anunciado en la investigación anterior, que se anuncia próximo al 73%.

### 2.4.2 Análisis y procesamiento de datos

En la época del Big Data y en una sociedad donde cada vez se acumulan más y más datos, las redes sociales no podían ser una excepción y quizás son uno de los más claros ejemplos de aplicaciones que cada día generan millones de datos con la actividad de sus usuarios, que acceden para compartir una foto, publicar un comentario o simplemente para conectar con otro usuario. Todas estas acciones generan datos que son almacenados y que pueden ser procesados y analizados para obtener valor. Sin duda este es uno de los grandes retos para muchos investigadores que han profundizado en este ámbito con el fin de ser capaces de procesar un gran volumen de información con el fin de obtener

## 2.4 Investigación en el ámbito de las redes sociales

---

algún beneficio. De este modo, tomando como referencia las publicaciones de millones de usuarios en Twitter (29) y realizando un análisis cualitativo de la información fue posible extraer conclusiones a cerca del estado de salud de los usuarios e incluso ofrecer con un alto grado de detalle qué enfermedades son las más comunes entre los usuarios, sobre cuáles de ellas no tienen inconveniente en hablar o comentar y por el contrario, que otras enfermedades prefieren no hacer públicas. En esta misma línea, y de nuevo utilizando la red social Twitter, el proyecto *Tweet My Street* (30) pretender analizar y procesar masivamente los datos de usuarios posicionados geográficamente en un mismo lugar: un barrio de Newcastle (UK). El objetivo es extraer conclusiones a cerca del comportamiento en una red social de usuarios que comparten una misma ubicación. Este es un claro ejemplo de lo comentado anteriormente, es decir, una investigación donde el análisis y procesamiento de datos juega un papel crucial y el volumen de información a procesar es muy elevado. Por último, para finalizar este apartado cabe destacar este experimento (31) más cercano a esta propuesta de investigación en cuanto al propósito final. Mediante el análisis y procesamiento de los millones de datos que almacena LinkedIn a cerca de los gustos e intereses de sus usuarios, se centra este proyecto de investigación que pretender analizar y procesar esa gran cantidad de información con el fin de ser capaces de emitir recomendaciones a esos usuarios sobre trabajos de investigación académicos que puedan resultar de su interés. Nuevamente, en este caso, se observa un experimento científico donde el foco se centra en los datos de una única red social, limitando los resultados que podrían ser mucho más precisos si se contemplaran el resto de redes sociales en los que los usuarios del experimento tienen actividad. Como se puede comprobar, son millones los datos que existen en las redes sociales, y por tanto, múltiples las formas de generar conocimiento a partir de esos datos, y a su vez, crear valor.

### 2.4.3 Añadiendo otras fuentes de información para generar valor

En los casos anteriores se detallan propuestas de investigación que buscan generar valor en el ámbito de las redes sociales, utilizando para ello los millones de datos que almacenan sobre sus usuarios. Existe un paso más, en el que los investigadores buscan generar valor añadiendo otras fuentes de información, de manera similar a como se plantea en esta propuesta de investigación, donde se utilizará la base de datos del ISBN para generar un conjunto de recomendaciones sobre autores y obras literarias a aquellos

## 2. ESTADO DEL ARTE

---

usuarios que tengan este hobby o afición. Similar al ISBN encontramos el IMDB (32), base de datos pública que almacena autores y obras cinematográficas. Sobre ella, se encuentran varias propuestas de investigación. De este modo, este estudio (33) utiliza el IMDB a través de un subconjunto de datos para estudiar las colaboraciones entre actores y de este modo analizar la topología de red que puede ser conformada, estableciendo analogías con otras redes. Además, concluye que el uso de esta base de datos puede ofrecer múltiples ventajas en el estudio de las interacciones sociales. El ISBN, al igual que el IMDB, parece una fuente de datos idónea para que el sistema de recomendación pueda tomar los datos necesarios para lanzar las recomendaciones y sugerencias de autores y obras. En este caso, al tratarse de una base de datos pública se facilita el análisis y la explotación de toda la información almacenada, tal y como se puede comprobar en el minucioso estudio llevado a cabo en la referencia que se ha comentado anteriormente. Con los datos de las dos entradas del sistema, es decir, redes sociales y bases de datos públicas que representen una afición, el software que se desarrollará durante la investigación deberá ser capaz de generar una salida o resultado en forma de recomendaciones o sugerencias para usuarios cuya actividad social ha permitido generar la entrada referente a redes sociales. Sobre esta base, existen varias referencias que explotan datos de ambas entradas, es el caso de este estudio (34), donde se predice la clasificación de las películas almacenadas en el IMDB. Para ello, se utilizan como entradas, por un lado la información en redes sociales, en este caso tweets de Twitter y comentarios de YouTube y por otro lado, los datos sobre películas almacenados en el IMDB. La salida en este caso será una predicción sobre la valoración que tendrá una determinada película en el IMDB. En este caso de estudio, a diferencia de otros comentados con anterioridad, si se contempla más de una red social, Twitter y Facebook, incluso en las conclusiones se definen unas líneas de futuro que pasan por incluir más redes sociales. Es destacable que la investigación llevada a cabo en este caso de estudio se aleja del objetivo de ofrecer al usuario un beneficio tomando como referencia las redes sociales, que bien podría pasar por ofrecer recomendaciones sobre actores o películas del IMDB a los usuarios de Twitter y Youtube en base a sus tweets y comentarios respectivamente. Otro estudio (35), que a diferencia del anterior, si genera un beneficio claro para la comunidad, plantea como entradas los datos e información almacenados en Twitter y la DBLP (36) respectivamente, esta última se trata de una base de datos pública que almacena datos sobre obras y autores relacionados con las ciencias de la

## 2.5 Sistemas de recomendación en el ámbito empresarial

---

computación. El objetivo principal pasa por conseguir encontrar coincidencias entre el perfil de un científico en Twitter y en la DBLP, donde estarían relacionadas todas sus obras. En este caso además, se plantea como línea de futuro poder valorar al científico no sólo por su producción científica, que puede encontrarse en esta base de datos, sino también por sus aportaciones en Twitter. Nuevamente se echa en falta poder encontrar una línea de futuro que pase por definir el objetivo de lograr la equivalencia entre el científico de la DBLP y su perfil social, con independencia de la red social, ya sea Twitter, Facebook o cualquier otra, e incluso que la valoración del mismo se base no sólo en sus aportaciones en Twitter, sino también en sus comentarios en Facebook o el tipo de fotos que publique en Instagram, por poner un ejemplo. Las anteriores referencias presentan experimentos o casos de estudio donde se utiliza una base de datos pública, ya sea DBLP o IMDB como fuente de datos. En este estudio (37) se plantea un experimento que utiliza ambas, en pruebas independientes, siguiendo una propuesta similar a la planteada en esta investigación, en cuanto a la salida o beneficio que se pretende alcanzar, con la generación de recomendaciones sobre autores de estas bases de datos, aunque en este caso la recomendación sería únicamente sobre un autor en base a otro diferente, pero que presenta cierta relación con el anterior. Para ello, en el experimento que se plantea, las citadas bases de datos pasan a ser conjuntos de nodos, cada uno de los cuáles sería un autor, de manera análoga a como se estructura una red social con usuarios. Sin embargo, aunque el título de la publicación es bastante prometedor, deja como línea de futuro probar el sistema desarrollado con una verdadera red social, como Facebook, y en ningún caso se plantean generar dos entradas, es decir, utilizar los datos en redes sociales de un determinado usuario para recomendarle un autor de ciencias de la computación o cine, tomando para ello los datos de la DBLP o el IMDB.

## 2.5 Sistemas de recomendación en el ámbito empresarial

Esta propuesta de investigación plantea un sistema de recomendación basado en el perfil social de un usuario, de manera que sea posible generar recomendaciones sobre autores y obras del ISBN que puedan llegar a ser del agrado de éste. Fuera del ámbito de las propuestas y trabajos que plantean el uso de sistemas de recomendación con un objetivo ms centrado en hacer investigación o con un carácter docente, se ha podido comprobar que el uso de estos sistemas en el mundo empresarial ha sido recurrente a

## 2. ESTADO DEL ARTE

---

lo largo de los años para que grandes empresas mejoren la experiencia de sus clientes. Existen muchos ejemplos, pero quizás el más popular sea el de Amazon, cuyo sistema de recomendación ha sido ampliamente reconocido. En 2003, un grupo de ingenieros de Amazon (38) publicó los detalles sobre su sistema de recomendación basado en el filtrado colaborativo, el cual utiliza técnicas en las que intervienen diversos agentes y fuentes de datos. En el caso del sistema de recomendación de Amazon se utiliza el filtrado colaborativo pero aplicando algunas mejoras en lo que Amazon definió como "item-to-item collaborative filtering". Este sistema asocia cada producto comprado por un usuario con una lista de productos similares, que se obtiene en función de los elementos que hayan sido adquiridos en un mismo pedido, añadidos a un carrito de la compra o almacenados en una lista de deseos. Con esta asociación de productos, el sistema es capaz de emitir recomendaciones a usuarios con una tasa de acierto muy elevada.

Pero Amazon no ha sido la única gran empresa en utilizar estos sistemas de recomendación para mejorar la experiencia del cliente y su modelo de negocio, YouTube (39) utiliza lo que puede denominarse como filtrado basado en contenido para emitir recomendaciones que están basadas en el conocimiento que albergan en sus sistemas sobre los vídeos que el usuario ha valorado previamente, de manera que se le recomendarán vídeos similares que les puedan llegar a gustar. FilmAffinity o Netflix, por ejemplo, utilizan un filtrado colaborativo más puro, consistente en averiguar que usuarios son similares entre sí y en base a esto, emitir recomendaciones sobre aquellos ítems que no han sido valorados todavía por un determinado usuario pero sí por otros usuarios similares con una valoración alta. Existen otros filtrados demográficos, que emiten recomendaciones únicamente tomando como referencia algunas características del usuario como son la edad, el sexo, la situación geográfica, etcétera. Finalmente, el filtrado híbrido se presenta como una combinación de los anteriores, donde podría englobarse el sistema de Amazon, que como se ha comentado, combina varias técnicas y se apoya en la gran cantidad de información que almacena Amazon sobre sus usuarios. Estableciendo algunas similitudes con esta propuesta de investigación, esa gran cantidad de información que Amazon guarda sobre sus usuarios y que es de gran utilidad para que su sistema de recomendación obtenga un gran desempeño, el sistema de recomendación propuesta la obtendría del perfil social del usuario compuesto por toda la información que las principales redes sociales albergan sobre éste.

### 2.6 Seis Grados de Separación

Finalmente, para dar por concluido este capítulo es necesario volver al punto introductorio de esta propuesta donde se comentaban las principales justificaciones para emprender esta investigación, una de las cuales se basa en aportar un mayor conocimiento de las redes sociales desde un punto de vista científico, con el objetivo de poder seguir avanzando en este ámbito. Para poder realizar esta aportación, es importante conocer el origen de las redes sociales y, como se comentaba al comienzo del primer capítulo, conocer la hipótesis de los Seis Grados de Separación, uno de los orígenes del término red social que se conoce en la actualidad. Esta hipótesis ha despertado el interés de la comunidad científica, ya sea para realizar una aproximación que permita demostrar su veracidad o para aplicarla en un contexto científico y de este modo poder llevar a cabo un determinado experimento o caso de estudio. Es el caso de este estudio (40), que además utiliza de nuevo la DBLP como fuente principal de datos. De esta forma, utilizando toda esta información se pretende analizar a los autores de esta comunidad científica estableciendo una analogía con la hipótesis de los Seis Grados de Separación, para determinar los nodos que definen la distancia entre un autor y otro. A modo de conclusión determinan que la separación entre dos autores de la DBLP es pequeña, sin embargo no llegan a precisar con exactitud la distancia entre dos nodos dentro del caso de estudio que conforman los autores de la DBLP, y quizás esta falta de precisión se deba a que se trata de un artículo escrito en el año 2005. En el caso de este proyecto de investigación, que si puede hacer uso de toda la tecnología existente en la actualidad así como todas las herramientas que ofrecen las redes sociales y que se han comentado a lo largo de este capítulo, si se pretende utilizar la hipótesis de los Seis Grados de Separación para probar el correcto desempeño del sistema que se va a desarrollar como objetivo principal de esta investigación.

Desde luego existen referencias más actualizadas que aunque siguen la misma base y fundamentos que los de la investigación citada anteriormente, si pueden hacer uso de tecnologías más modernas, de forma similar a como se planteará en este proyecto de investigación. De este modo, Facebook (41) desarrolló un software con el que posteriormente planteó un caso de estudio donde determinó que entre dos personas existía una separación pequeña que iba disminuyendo fundamentalmente debido al hecho de estar inmersos en una sociedad global. Además, cuantificó esa separación en 3,57. Otro

## 2. ESTADO DEL ARTE

---

estudio utilizó Twitter (42) como red social de referencia, donde se llevó a cabo un experimento que determinó que el grado de separación entre dos personas era de 4,59. Sin embargo, estos casos únicamente utilizaron una única red social para elaborar el estudio y demostrar la hipótesis de los Seis Grados de Separación, por lo que únicamente contaron con los datos de usuarios de Facebook y Twitter respectivamente. Para este proyecto de investigación, se plantea utilizar el sistema de recomendación que presenta una entrada con el perfil social del usuario compuesto por las principales redes sociales. De este modo se podrá demostrar la hipótesis con datos de diferentes redes sociales lo que podrá cuantificar los nodos entre usuarios y por tanto definir la separación entre ellos con mayor precisión.

Como se ha podido comprobar a lo largo de este capítulo, la comunidad científica muestra un gran interés por las posibilidades que ofrecen las redes sociales, bien para fines únicamente didácticos pero también para alcanzar otros objetivos más ambiciosos centrados en obtener valor a partir de los datos almacenados en ellas. En muchos casos, los estudios se han centrado únicamente en una única red social, mientras que en esta investigación se pretende dar una visión única de todas las redes sociales para obtener un beneficio para los usuarios, con independencia de que éstos tenga un perfil social conformado por una o varias redes sociales, aunque lógicamente cuanto mayor cantidad de datos sea posible recoger sobre el usuario, los resultados podrán ser más precisos. Para alcanzar este objetivo, se utilizará el amplio catálogo de herramientas que las redes sociales ponen a disposición de terceros para explotar los datos que día a día generan millones de usuarios.

# 3

## Propuesta

### 3.1 Nuevo sistema de recomendación

En el capítulo del estado del arte de esta investigación se han analizado las principales propuestas llevadas a cabo por otros investigadores en el ámbito de los sistemas de recomendación y su posible conexión con las redes sociales para generar valor. Además se revisaron algunas propuestas fuera de la comunidad científica, en un entorno empresarial, donde cada vez más empresas apuestan por estos sistemas para mejorar su modelo de negocio. Esta revisión es clave para analizar los puntos de mejora en los que el sistema de recomendación que se propone en esta investigación debe incidir para marcar una evolución dentro de este ámbito.

Por ello, analizando previamente los principales sistemas de recomendación y las distintas aplicaciones donde se pueden encontrar, se llevará a cabo una estrategia de Diseño y Creación para implementar un nuevo sistema que emita sugerencias y recomendaciones a un usuario, en base a los datos de su perfil social y a una determinada afición, representada por una base de datos pública, en este caso la literatura mediante los datos que ofrece el ISBN.

El algoritmo que se implementará tendrá dos entradas y una única salida. Una de las entradas serán los datos de las redes sociales donde el usuario tenga actividad, para lo cual habrá que utilizar las herramientas de integración que ofrecen las principales redes sociales y cuya aplicación en casos reales se ha tratado en las referencias comentadas en el capítulo anterior. La otra entrada de datos será resultado del análisis y explotación de una base de datos que representa una afición en concreto, en el caso

### 3. PROPUESTA

---

de este proyecto de investigación, la literatura mediante la base de datos de autores y obras que ofrece el ISBN. Finalmente, la salida del algoritmo vendrá representada por un conjunto de recomendaciones sobre autores y obras para el usuario cuyo perfil social se ha definido en una de las entradas. Hay que tener en cuenta además, que si el usuario no está interesado en la literatura, el sistema lo detectará en su salida no emitiendo ninguna recomendación o sugerencia. En el siguiente diagrama de flujo 3.3 se puede ver representado un esquema general del funcionamiento del sistema de recomendación que se desarrollará durante la investigación, con sus dos entradas (A y B) y su única salida.

### 3.1 Nuevo sistema de recomendación

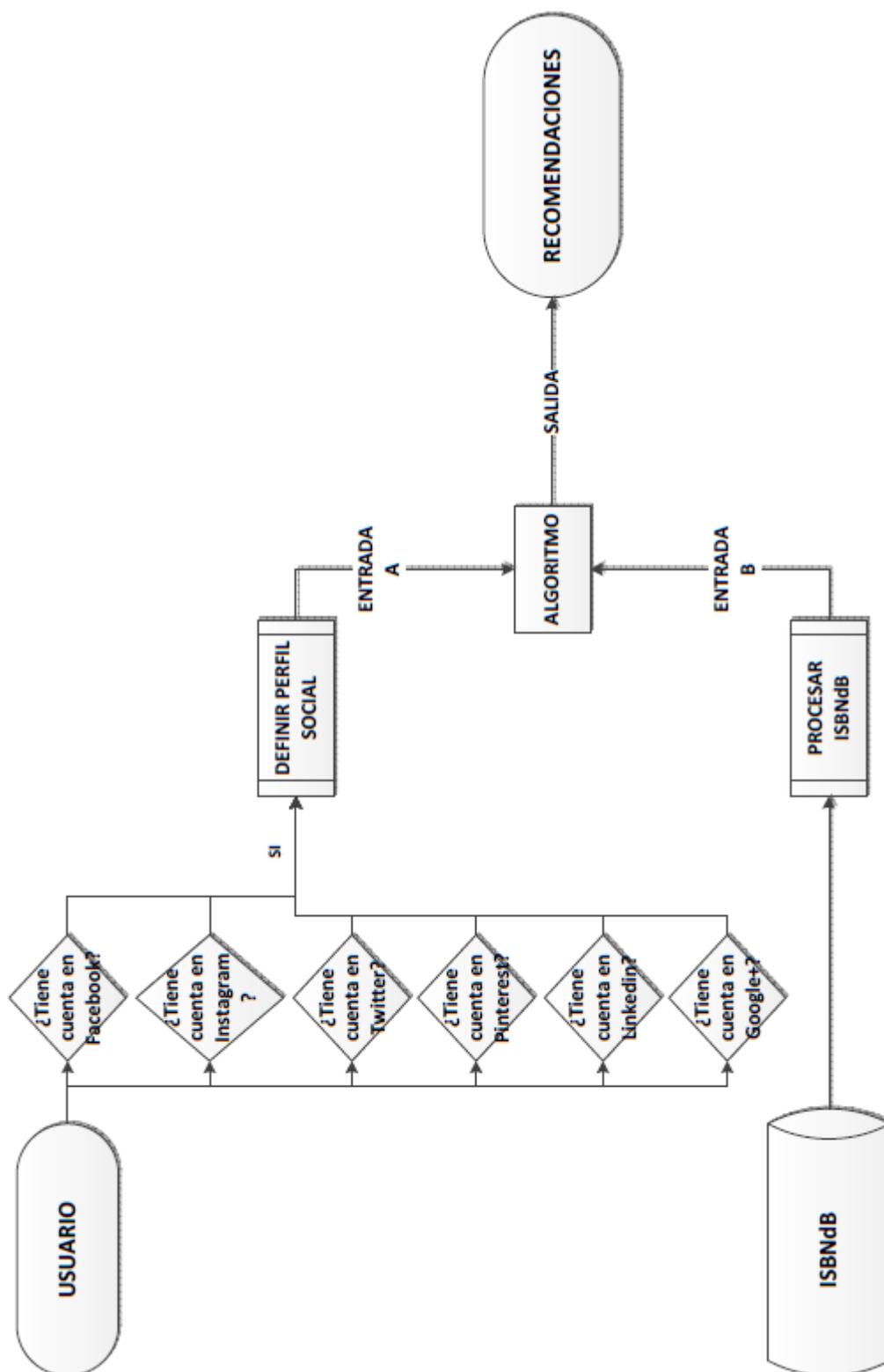


Figura 3.1: Diagrama de Flujo del Sistema de Recomendación.

### 3. PROPUESTA

---

#### 3.1.1 Entrada A: Perfil Social

Como se ha comentado en el apartado anterior, la entrada A del nuevo algoritmo será el perfil social del usuario. Este perfil estará definido por las cuentas que el usuario tenga en las principales redes sociales:

- **Facebook.** El API (43) de esta red social permite recuperar el listado de amigos de un usuario, los grupos de los que es miembro y su actividad social, incluyendo likes, comentarios y publicaciones realizadas. Asimismo es posible recuperar los datos de su perfil en Facebook, que incluyen los temas sobre los que muestra interés e incluso su usuario en Instagram, en caso de tenerlo, facilitando al sistema la conexión con esta otra red social, objeto de la investigación. Toda esta información resultará muy útil para poner en práctica la estrategia que seguirá el sistema para emitir recomendaciones, tal y como se detallará en próximos apartados.
- **Instagram.** Su API (44) permitirá recuperar los datos del usuario, sus conexiones con otros usuarios, su actividad social conformada por likes, comentarios y publicaciones y un aspecto clave para el sistema de recomendación: los tags con los que clasificará sus publicaciones en esta red social.
- **Twitter.** De manera similar a los casos anteriores, el API (45) de Twitter permite recuperar cualquier metadato del tweet de un usuario, que permitirá al sistema clasificarlo, así como información más general sobre el usuario, como conexiones con otros usuarios, listas de visualización de *tweets* o información sobre el perfil del usuario.
- **Pinterest.** Siguiendo la línea de las redes sociales anteriores, su API (46) permite recuperar cualquier información relacionada con un usuario de esta red social.
- **Linkedin.** El API (47) de esta red social ofrece información muy detallada sobre el perfil del usuario, quizás la que más detalle ofrece en comparación con otras redes sociales, en parte por la importancia de estos datos para la interacción del usuario con esta red social. Esta información será de gran utilidad para que el sistema de recomendación pueda clasificar al usuario y a otros similares que estén interconectados con éste, con el objetivo de poder emitir recomendaciones que sean de su agrado.

- **Google Plus.** Quizás esta sea la red social de menor importancia, pero su relación con el gigante Google y su buscador, permiten que éste API (48) pueda ofrecer información muy valiosa sobre sus usuarios, que sea de gran ayuda para el sistema de recomendación.

Como se ha podido comprobar, la mayoría de las redes sociales descritas ofrecen información muy importante sobre un determinado usuario centrándose de manera general en dos aspectos bien diferenciados: su perfil, que facilitará al sistema la tarea de averiguar si el usuario es aficionado a la literatura y sobre qué géneros, y su actividad social, donde en la mayoría de redes sociales, fundamentalmente en Twitter e Instagram, ésta vendrá clasificada mediante tags que facilitarán la tarea al sistema para determinar qué temáticas son las preferidas para un usuario en concreto.

La razón de utilizar estas seis redes sociales y no otras, como se comentó en el capítulo del estado del arte, se basa en que éstas son las que tienen un mayor número de usuarios activos, y por tanto, más datos será posible recopilar y más probabilidades tendrá el sistema de encontrar información sobre el usuario, pieza clave para que la salida sea la esperada en cada caso. Durante el capítulo del estado del arte se han aportado múltiples referencias sobre otros estudios llevados a cabo por investigadores que pretendían generar valor mediante la utilización de redes sociales, gracias a los datos que almacenan sobre sus usuarios. Sin embargo, la mayoría de estos estudios se centran en una única red social, o a lo sumo incorporan las dos principales redes sociales, pero en todos los casos el experimento o caso de estudio realizado permanece muy ligado a las características de una determinada red social: los *tweets* de Twitter o las conexiones de LinkedIn por citar sólo dos ejemplos. Por tanto, éste sería uno de los elementos diferenciadores de la investigación con respecto a otras similares, ya que se pretende alcanzar una independencia total de los datos que conforman el perfil social del usuario con respecto a las redes sociales de las cuales se han obtenido esos datos. De esta forma, el sistema de recomendación que se propone no trabajará con las características propias de cada red social sino que se desarrollará una capa de abstracción que será utilizada por el sistema y ésta a su vez será la que se integre con las seis redes sociales, para obtener datos suficientes como para definir el perfil social del usuario. Por tanto, el sistema trabajará con independencia de cada red social manejando la información almacenada en el perfil social, quedando definida de esta forma la entrada A del sistema

### 3. PROPUESTA

---

de recomendación. Esta entrada estará conformada por tres bloques de información claramente diferenciados:

- **Perfil.** Son los datos que definen los rasgos principales de la persona que hay detrás del usuario. Son fundamentalmente datos demográficos tales como edad, sexo, aficiones, etcétera. Estos datos serán de gran ayuda para que la aplicación pueda agrupar correctamente al usuario junto a otros usuarios, de tal forma que sea más fácil ofrecerle recomendaciones similares a las que han tenido éxito con otros usuarios de una edad similar o que compartían una misma afición.
- **Conexiones.** Son los datos que definen las conexiones entre un usuario y otros. Son fundamentales para la generación de recomendaciones ya que, como se verá más adelante, se parte de la premisa de que un conjunto de recomendaciones que tienen éxito con un determinado usuario, también lo tendrán con otro usuario conectado al anterior, si ambos tiene algún nexo común como puede ser una afición.
- **Actividad.** En este bloque se engloba el gran volumen de datos que generan las redes sociales para un determinado usuario. Se trata de toda la información que éste genera con su actividad social: likes, artículos compartidos, *tweets*, *pins*, etcétera. Será una información que permitirá determinar como van evolucionando los gustos y aficiones de un determinado usuario para que las recomendaciones sean lo más precisas posibles.

A continuación se muestra un diagrama de bloques que refleja de manera resumida la información de las redes sociales que será tenida en cuenta dentro de la entrada A del sistema para obtener las evidencias con las que se emitirán las recomendaciones. Destacar que en negrita aparecen los bloques de información que tendrán más peso para el sistema de recomendación y en gris los elementos que no serán tenidos en cuenta en una primera fase del sistema. En los siguientes apartados se explicará con detalle cada uno de los bloques de información que permiten conformar la entrada A del sistema, imprescindible para que éste pueda cumplir con el objetivo de emitir conjuntos de recomendaciones sobre obras y autores del ISBN.



**Figure 3.2:** Diagrama de Bloques de la Entrada A del Sistema de Recomendación.

#### 3.1.1.1 Perfil

Como se ha comentado, este bloque de información engloba las características principales que definen a la persona que hay detrás de un determinado usuario de una red social. El valor añadido que ofrece esta propuesta de investigación se basa en que dicho perfil no estará definido únicamente por la información de una red social, normalmente Facebook o Twitter, como ocurre con otras investigaciones que se analizaron en el capítulo del estado del arte. En este caso, el sistema de recomendación que se plantea, a través de su entrada A compondrá el perfil social del usuario con la información que contengan sobre el mismo las principales redes sociales.

A continuación se detalla la integración de la aplicación con el API de Facebook (43), red social que ofrece la mayor información en este bloque de datos:

### 3. PROPUESTA

---

```
GET/v2.9/1603098691?fields=age_range , birthday , devices ,education ,
favorite_athletes , favorite_teams , gender , hometown , inspirational_people ,
interested_in , languages , location , meeting_for , political , religion , sports ,
work , books , games , movies , music , television
{
  "age_range": {
    "min": 21
  },
  "birthday": "05/14/1981",
  "devices": [
    {
      "hardware": "iPhone",
      "os": "iOS"
    },
    {
      "hardware": "iPad",
      "os": "iOS"
    }
  ],
  "education": [
    {
      "school": {
        "id": "108587789173044",
        "name": "Bernardo Balbuena"
      },
      "type": "High School",
      "id": "10204052869154358"
    },
    {
      "school": {
        "id": "114840225195101",
        "name": "UCLM Cuenca"
      },
      "type": "College",
      "id": "3575292633316"
    }
  ],
  "favorite_athletes": [
    {
      "id": "320863761326472",
      "name": "Valent Sanjuan"
    }
  ],
  "gender": "male",
  "hometown": {
    "id": "115185408492679",
    "name": "VALDEPEAS"
  },
  "location": {
    "id": "106504859386230",
    "name": "Madrid, Spain"
  },
  "id": "1603098691"
}
```

### 3.1 Nuevo sistema de recomendación

---

En el ejemplo anterior se observa como la aplicación se comunica con Facebook mediante una petición GET donde se informan todos los atributos para los que se desea obtener información dado un determinado usuario, previamente validado por la red social siguiendo el protocolo OAuth (15). Los atributos solicitados en la petición son todos los que se consideran que pueden resultar de utilidad para que la aplicación emita recomendaciones de calidad. A su vez, Facebook devolverá informados los atributos que tengan valor cumplimentado en su sistema para el usuario en cuestión, no devolviendo datos en caso contrario, tal y como se puede observar en el ejemplo anterior.

De un valor similar al aportado por Facebook, se encuentra Google Plus, cuya integración con el sistema se muestra a continuación, siguiendo un procedimiento similar al descrito para Facebook. Google Plus aprovecha la influencia del buscador de Google para que un gran porcentaje de sus usuarios presenten perfil en esta red social, lo cual la convierte junto a Facebook en las dos redes sociales con mayor importancia para el sistema de recomendación en el contexto de este bloque de información. De nuevo, la comunicación se realiza mediante una petición GET al API de Google Plus (25), validando previamente al usuario para el que se van a emitir las recomendaciones con el protocolo OAuth. En este caso, en la petición se solicita toda la información del perfil social del usuario sin posibilidad de filtrar atributos como se comentó en el caso de Facebook.

### 3. PROPUESTA

---

```
GET https://www.googleapis.com/plus/v1/people/me
{
  "kind": "plus#person",
  "etag": "\"Sh4n9u6EtD24TM0RmWv7jTXojqc/q5hWk_21HWbpcTnuViaCTLS70M\"",
  "gender": "male",
  "emails": [
    {
      "value": "emisagil@gmail.com",
      "type": "account"
    }
  ],
  "objectType": "person",
  "id": "110572664145415971579",
  "displayName": "Emiliano S nchez Gil",
  "name": {
    "familyName": "S nchez Gil",
    "givenName": "Emiliano"
  },
  "url": "https://plus.google.com/110572664145415971579",
  "image": {
    "url": "https://lh3.googleusercontent.com/-XdUIqdMkCWA/AAAAAAAAAAI/AAAAAAAAAA/4252rschv5M/photo.jpg?sz=50",
    "isDefault": true
  },
  "isPlusUser": true,
  "language": "es",
  "ageRange": {
    "min": 21
  },
  "circledByCount": 10,
  "verified": false
}
```

El resto de redes sociales no aportarán mayor valor del que ofrece Facebook y en menor medida Google Plus. En la mayoría de los casos únicamente servirán para refrendar que la información aportada por Facebook tiene valor, o llegado el caso poco probable, completarán algún atributo básico como la edad, no informado en Facebook por el usuario. La única red social que puede aportar información relevante para el perfil del usuario es LinkedIn por tratarse de una red profesional y por tanto, por aportar una serie de atributos que definen una serie de características útiles para la generación de recomendaciones por parte de la aplicación. A continuación se detalla la integración de la aplicación con el API de LinkedIn (47):

## 3.1 Nuevo sistema de recomendación

```
GET https://api.linkedin.com/v1/people/~?format=json
{
  "firstName": "Emiliano",
  "headline": "IT Service Manager en Mutua Madrileña",
  "id": "p3M5G-9vM4",
  "lastName": "Sánchez Gil",
  "siteStandardProfileRequest": {"url": "https://www.linkedin.com/profile/view?id=AAoAAAojdOcbgJvJ2aBDHr0P9JbgZIxMui80_n0&authType=name&authToken=VmcV&trk=api*a3227641*s3301901*"}
}
```

En el ejemplo anterior se observa como la aplicación se comunica con el API de LinkedIn (22) mediante una petición GET donde se solicita información sobre la trayectoria profesional del usuario de esta red social, previamente validado siguiendo el protocolo OAuth, de manera similar al caso de Facebook y Google Plus. La red social devolverá toda la información que albergue sobre dicho usuario.

Con todos los atributos del usuario obtenidos de las distintas redes sociales, la aplicación compondrá un perfil único que será uno de los tres puntos que tomará como base el sistema para emitir las recomendaciones sobre autores y obras literarias.

### 3.1.1.2 Conexiones

Las conexiones que genera un usuario en una red social aportan información valiosa para que un sistema de recomendación, basado en el filtrado colaborativo, emita nuevas recomendaciones teniendo en cuenta las que ya ha emitido a otros usuarios con un nexo común con el usuario actual que están siendo evaluado por el sistema. De este modo, si un sistema de recomendación emite un conjunto de recomendaciones para un determinado usuario, si a continuación entra otro usuario a la aplicación, el sistema evaluará si está conectado con el usuario anterior u otro ya evaluado para determinar si están conectados ya que de ser así las posibilidades de que el conjunto de recomendaciones emitidos en su momento tenga éxito, aumenta exponencialmente.

Como se verá a continuación, una de las grandes ventajas de las redes sociales son las relaciones que se establecen entre sus usuarios, y en este punto, todas aportan valor. A diferencia del bloque anterior, donde el perfil de Facebook es claramente más rico en información que en las otras redes sociales, las conexiones son igualmente valiosas en todas y cada una de las redes sociales, aunque quizás, por antigüedad, importancia y

### 3. PROPUESTA

---

estructura, las listas de conexiones de Facebook y Twitter sobre sus usuarios, son las más valiosas. A continuación se muestra como se integrará la aplicación con Twitter:

```
GET https://api.twitter.com/1.1/friends/ids.json
{"ids":[69962890,222598557,4796898335,2657046746,5832612,247509848,
715460515514941440,746039889704136704,21826657,190595327,297516932,
2457016706,216352300,253017888,773362776,9473582,103815290,4179826343,
1627741537,411618368,2882859263,156950068,125593849,193870850,184993308,
269996602,2377724724,57330551,538942623,14460582,351945610,512325153,
166772945,3071319568,243832023,259079045,324958945,308403230,364528766,
238334213,27266797,186982547,568453360,15052516,126036804,188077169,
120695807,835568149,80251660,191084621,135159056,31395914,319285743,
1265278585,252235954,183735793,981045312,115319935,127550968,85563781,
2541845341,273591193,2336454198,102646663,974117197,260305007,20444254,
2165120203,155068442,373718524,17507250,95521447,283090686,103982389,
165867955,488731085,23126091,1897670472,21829681,1706828990,340187674,
151090430,1579800343,835955214,9031882,74518740,365879592,88324102,
25319414,19923144,20346956,26442916,423868716,597718220,14498713,
293415952,1229534000,263353954,14340617,123332609,255674587,36709098,
535707261,932429090,1222060448,277418664,407775884,471218008,1147824680,
201905142,243608028,69011293,220047267,119938261,282064204,197496299,
456728111,2188291,148720585,15217919,376066038,62887164,7996082,
13426312,773357198,155579083,110850885,151123001,420301755,14872237,
245792560,481930317,174284325,278694865,443680219,126575175,203590680,
6640722,39509236,25477286,45534541,241147024,74992460,2853351,86074234,
27908927,507672047,226519339,465463432,473227811,155196498,11761292,
18200400,271909348,317727341,17305888,485511296,41412970,119405353,
31451582,34222270,155941371,16236094,16218180,105830515,235910394,
249196966,143836233,375163947,35993422,214109893,246492635,29708196,
28371887,105656768,65369125,23746926,211476024,267274202,294010646,
344634424,177603526,114728026,138088942,40047903,72811133,308973177,
189678284,32342975,99518813,74154806,104282000,163559269,138358313,
183387449,105451020],"next_cursor":0,"next_cursor_str":"0",
"previous_cursor":0,"previous_cursor_str":"0"}
GET https://api.twitter.com/1.1/friendships/lookup.json?user_id=1265278585%
2C252235954%2C183735793%2C981045312%2C115319935%2C127550968%2C85563781
[{"name":"We Like L.A.,"screen_name":"WeLikeLA","id":1265278585,
" id_str":"1265278585","connections":["following"]}, {"name":"Lucas_locura",
"screen_name":"lucasheras","id":252235954," id_str":"252235954",
"connections":["following"]}, {"name":"Matoses","screen_name":"Matoses",
" id":183735793," id_str":"183735793","connections":["following"]},
{"name":"MAD Running Company","screen_name":"madridrunningco",
" id":981045312," id_str":"981045312","connections":["following"]},
{"name":"Fernando Ruiz","screen_name":"fernan_ruiz","id":115319935,
" id_str":"115319935","connections":["following"]}, {"name":"I\u00f1a
D\u00e9daz-Guerra","screen_name":"InakoDiazGuerra","id":127550968,
" id_str":"127550968","connections":["following"]}, {"name":"Alma Obreg\u00f3n",
"screen_name":"alma_cupcakes","id":85563781," id_str":"85563781",
"connections":["following"]}]
```

### 3.1 Nuevo sistema de recomendación

---

Como se ha podido comprobar, el API de Twitter (19) obliga a dos llamadas para obtener los usuarios conectados con un usuario previamente validado mediante OAuth. En este caso, primero se recuperan los identificadores de los usuarios conectados y posteriormente se recupera información más detallada, incluyendo el identificador de cada conexión para poder buscar recurrentemente otras conexiones.

En este bloque de información hay que tener en cuenta como estructura cada red social las conexiones de sus usuarios ya que no todas lo hacen igual, lo cual añade complejidad al sistema. En los siguientes puntos se detalla como estructuran las redes sociales las conexiones de sus usuarios, haciendo especial hincapié en la información que será útil para el sistema de recomendación:

- **Facebook.** Esta red social presenta multiples formas de estructurar las conexiones de sus usuarios: listas de amigos y familiares, usuarios suscritos a grupos o eventos, etcétera. En este caso, se tendrán en cuenta dos tipos de conexiones para que el sistema de recomendación emita sus recomendaciones: la lista de amigos del usuario objeto de evaluación y la naturaleza de los grupos o eventos a los que dicho usuario se ha suscrito.
- **Instagram.** Se organizan las conexiones en dos grupos: *Seguidores* y *Seguidos por*, de forma que el primer bloque representa los usuarios que han tenido la iniciativa de conectar con el usuario para el que se desea emitir las recomendaciones, mientras que el segundo bloque representa los usuarios con los que ha conectado el usuario en cuestión.
- **Twitter.** Se estructuran las conexiones en base a *Amigos* y *Seguidores*, con un tratamiento bien diferenciado por el API de esta red social. Twitter entiende que los *Amigos* son los usuarios que sigue un determinado usuario, mientras que los *Seguidores* son los usuarios que siguen a un determinado usuario. A efectos del sistema de recomendación, los *Amigos* se tendrán en cuenta para buscar las preferencias de un determinado usuario, mientras que los *Seguidores* se tendrán en cuenta para buscar relaciones próximas con el objetivo de aprovechar conjuntos de recomendaciones ya realizados.
- **Linkedin.** En este caso, el API de Linkedin ofrece grandes dificultades para explorar las conexiones de sus usuarios por lo que el sistema de recomendación

### 3. PROPUESTA

---

no tendrá en cuenta esa información en una primera fase de la aplicación y se planteará su uso en una línea de futuro de la investigación.

- **Pinterest.** De manera similar a los casos anteriores, las conexiones se estructuran en *Seguidores* y *Seguidos por*, de manera que el primer bloque representa los usuarios que siguen al usuario para el que se desea emitir las recomendaciones, mientras que el segundo bloque representa los usuarios con los que ha conectado el usuario en cuestión.
- **Google Plus.** En este caso las conexiones de un usuario se organizan de una manera muy sencilla a través de *Círculos*, donde se agrupan los usuarios con los que el usuario para el que se desean emitir las recomendaciones ha conectado. A su vez, existirán otros *Círculos* donde otros usuarios han incluido al usuario en cuestión. Como ocurre con el resto de casos, los *Círculos* que el propio usuario ha creado serán tenidos en cuenta por el sistema de recomendación para establecer relaciones con las que aprovechar conjuntos de recomendaciones sobre autores y obras literarias ya emitidas a otros usuarios.

Con todas las conexiones recuperadas de las distintas redes sociales, la aplicación generará un conjunto de relaciones globales para un usuario en concreto, y servirá para que las recomendaciones que se emitan en base a la actividad social de un usuario, o a su perfil, sean tenidas en cuenta para sus conexiones, ya que probablemente puedan llegar a resultar útiles. De la misma forma, se buscarán conjuntos de recomendaciones ya emitidas sobre las conexiones de un usuario para evaluar si éstas podrán tener éxito con el usuario en cuestión. Asimismo, este bloque de información servirá para que el sistema explore la hipótesis de los Seis Grados de Separación para testear cuantos nodos separarán a dos usuarios, teniendo en cuenta sus conexiones en redes sociales.

#### 3.1.1.3 Actividad

Este es el bloque de información más relevante y quizás el que mejor ejemplifique el gran valor de las redes sociales cuando son aplicadas en sistemas de recomendación. Se trata de información de gran valor, muy dinámica y que puede reflejar con gran precisión las preferencias de un determinado usuario, vitales para ofrecer recomendaciones, pero a su vez de gran complejidad ya que cada red social estructura esta información de manera muy diferente. Para el objetivo de esta investigación se evalúa toda la información

### 3.1 Nuevo sistema de recomendación

---

que ofrecen las redes sociales sobre la actividad de sus usuarios, sin embargo, no toda será útil para el objetivo de emitir recomendaciones de calidad sobre autores y obras literarias. Es importante destacar que se debe evolucionar el sistema de recomendación a través de pequeñas fases que incorporen cada vez más aspectos de este bloque de información, en lugar de tener en cuenta toda la información sobre la actividad social del usuario en una primera fase ya que ésta podría generar ruido e introducir resultados negativos que repercutirían en una baja calidad del conjunto de recomendaciones emitidos por el sistema.

A continuación se detalla la información que será tenida en cuenta por el sistema para emitir las recomendaciones en una primera fase de la aplicación, dejando para las siguientes fases y líneas de futuro de la investigación otras acciones a tener en cuenta de la lista de la actividad del usuario en cada red social.

- **Facebook.** De entre todas las posibles opciones que ofrece el API de Facebook para recuperar la información de la actividad social del usuario, se evaluará si en la lista de likes del usuario para el que se van a emitir recomendaciones existe alguna página o grupo de Facebook, y en ese caso, se analizará su naturaleza para ver la actividad o el sector de interés que cubren. Este dato supondrá una evidencia para que el sistema lo tenga en cuenta a la hora de emitir recomendaciones. De esta forma, si por ejemplo el usuario genera un like sobre la página de la *Liga de Fútbol Profesional*, el sistema podrá determinar que es aficionado al fútbol y tenerlo en cuenta a la hora de emitir sus recomendaciones. A continuación se muestra la integración de la aplicación con el API de Facebook para recuperar los últimos likes del usuario:

### 3. PROPUESTA

---

```
GET /v2.9/1603098691?fields=id,name,likes
{
  "id": "1603098691",
  "name": "Emiliano S nchez Gil",
  "likes": {
    "data": [
      {
        "name": "Saleta Castro",
        "id": "135278313171912",
        "created_time": "2017-05-01T08:41:49+0000"
      },
      {
        "name": "Racion de NBA",
        "id": "153772351383456",
        "created_time": "2015-08-29T20:56:26+0000"
      },
      {
        "name": "Valent Sanjuan",
        "id": "320863761326472",
        "created_time": "2015-06-28T19:24:59+0000"
      },
      {
        "name": "Play English",
        "id": "489192447817809",
        "created_time": "2015-03-05T20:56:10+0000"
      },
      {
        "name": "Hoy es s bado",
        "id": "529358017080814",
        "created_time": "2012-09-02T14:32:40+0000"
      }
    ],
    "paging": {
      "cursors": {
        "before": "MTM1Mjc4MzEzMTcxOTEy",
        "after": "NTI5MzU4MDE3MDgwODE0"
      }
    }
  }
}
```

En los datos devueltos por el API de Facebook que muestran los likes del usuario que se ha validado mediante OAuth, se puede observar como, entre otros, se recuperan los identificadores de los perfiles de un programa de radio sobre la NBA o perfiles de varios deportistas. Estos datos pueden ser claves para que el sistema genere el conjunto de recomendaciones sobre el usuario en cuestión.



### 3. PROPUESTA

---

docentes dentro del ámbito de la investigación, y por tanto no comerciales, puedan explorar la actividad social de sus usuarios. Por este motivo y al no tratarse de una de las principales redes sociales, no se tendrá en cuenta en una primera fase de la aplicación y se evaluará su posible incorporación al sistema en una línea de futuro de la investigación.

- **Pinterest.** Se analizará la naturaleza y características principales de los *boards* del usuario y los *pins* que los contienen. En Pinterest los *boards* son paneles donde el usuario puede englobar imágenes sobre una temática clara. Por ejemplo, un usuario puede crear un panel con *pins* sobre moda, obtenidos a través de las fotos publicadas por revistas y agencias publicitarias. Esto supone una información valiosa para el sistema de recomendación a la hora de generar recomendaciones de calidad para el usuario. A continuación se muestra la integración de la aplicación con el API de Pinterest (20) para recuperar los *boards* del usuario:

```
GET https://api.pinterest.com/v1/me/\textit{boards}/?access_token=XXX
&fields=id%2Cname%2Curl
root:{} 1 item
data:[] 1 item
0:{} 3 items
url:https://www.pinterest.com/emisagil/nba/
id:161988986536208512
name:NBA
```

En los datos devueltos por el API de Pinterest se muestran los *boards* cuyo autor es el usuario para el que se van a emitir las recomendaciones, y que previamente ha sido validado mediante el protocolo OAuth. Se puede observar como uno de los *boards* tiene como título *NBA* e incluye un identificador a través del cual se pueden consultar más detalles sobre la naturaleza de este panel. Esta información genera una evidencia clara para que el sistema emita un conjunto de recomendaciones que pueda ser del agrado del usuario.

- **Google Plus.** En este caso se considera que la actividad social generada por esta red social no es de gran utilidad para cumplir con el objetivo de la investigación en una primera fase y únicamente se aprovechará la información del perfil, con un nivel de detalle similar al de Facebook. Esta red social, por sus características

### 3.1 Nuevo sistema de recomendación

---

y su relación con el buscador de Google contiene perfiles de un gran número de usuarios e información sobre sus conexiones. Sin embargo, su uso no está tan extendido como el resto de redes sociales por lo que la actividad generada por sus usuarios es muy limitada y su consideración para el sistema que va a emitir las recomendaciones se descarta en una primera fase y se plantea su incorporación en alguna de las líneas de futuro de la investigación.

Hay que tener en cuenta que la actividad social del usuario genera mucha y variada información, además de la que será objeto de análisis por parte del sistema, hay otros muchos puntos posibles a tener en cuenta para el análisis como los comentarios del usuario en las distintas redes sociales, las acciones para compartir un determinado link o artículo, el análisis de albums fotográficos o vídeos, etcétera. Todos estos bloques de información son válidos para el sistema de recomendación y pueden aportar valor, pero a su vez introducen una gran complejidad y pueden llegar a generar ruido si no se introducen paulatinamente en el sistema. Por ello, se ha decidido comenzar por los puntos anteriores y el resto de opciones posibles se plantearán como líneas de futuro para mejorar el sistema de recomendación.

#### 3.1.2 Entrada B: Base de Datos del ISBN

En el capítulo del estado del arte se mencionaron varias referencias a proyectos de investigación donde además de utilizar las redes sociales como fuente de información, se añadía un actor más, en este caso una base de datos pública que aporta otra fuente de información para generar valor, incluso en algunos casos de forma similar al sistema planteado en esta investigación, que se basa en la generación de recomendaciones o sugerencias a usuarios sobre una temática determinada. Como se ha podido comprobar, las principales bases de datos utilizadas en el ámbito de la investigación son la DBLP, que almacena datos sobre obras y autores relacionados con las ciencias de la computación y sobre todo el IMDB, que contiene un amplio catálogo de películas, actores y directores. Sin embargo, prácticamente no se ha encontrado ninguna referencia relevante sobre el ISBN y más aún, ninguna investigación que haya utilizado esta fuente de datos junto a las redes sociales, con el objetivo de generar recomendaciones sobre autores y obras literarias a los usuarios de esas redes sociales. Por contra, las grandes empresas son las que más han apostado por este tipo de sistemas de recomendación, junto con aquellos

### 3. PROPUESTA

---

que manejan recursos multimedia, puesto que es donde mayor valor se puede generar para el usuario y el camino más rápido para optimizar sus modelos de negocio en un entorno cada vez más digital. Las principales empresas de este sector, con Amazon a la cabeza, se han dado cuenta de ello, y se esfuerzan a diario en mejorar sus sistemas de recomendación, pieza clave en la relación que establecen con sus clientes.

Por tanto, este será otro elemento diferenciador con respecto a otras investigaciones, la utilización de una base de datos poco utilizada en este ámbito, el ISBN, como fuente de información para realizar investigación y de este modo publicar todo el conocimiento que permita a otros investigadores la utilización de esta base de datos como medio para cumplir los objetivos definidos en sus proyectos de investigación. Por tanto, la entrada B del algoritmo será la base de datos del ISBN, cuya información se estructura en cinco grandes bloques:

- **Libros.** La búsqueda de libros dentro de la base de datos se puede realizar por multitud de atributos entre los que destacan:
  - *ISBN.*
  - *Título.*
  - *Clave,* con la que se identifica al libro en la base de datos del ISBN.
  - *Combinada,* que permite buscar con varios atributos a la vez, como son el título, el autor y/o la editorial.
  - *Completa,* que permite buscar en todos los atributos de la base de datos del ISBN: título, autor, editorial, resumen, notas, premios, etcétera.
- **Categorías.** Esta búsqueda está más limitada que en el caso de los libros y se limita al nombre de la categoría, a la clave con la que se identifica la categoría en la base de datos del ISBN o a la clave de la categoría de primer nivel.
- **Temas.** Del mismo modo que para el caso de las categorías, la búsqueda se limita al nombre del tema, a la clave con la que se identifica el tema en la base de datos del ISBN o a la clave del tema de primer nivel. Tanto este punto como el anterior, serán los que de manera más directa podrán relacionarse con los tags que clasifican la información que interesa a un usuario en su perfil en red social, y la actividad que éste desarrolla dentro de la misma, siendo un procedimiento

### 3.1 Nuevo sistema de recomendación

---

clave para el sistema poder relacionar ambos aspectos de las dos entradas para generar una salida con un porcentaje de éxito lo más elevado posible.

- **Autores.** En este caso la búsqueda únicamente se puede realizar por nombre de autor o por la clave con la que se identifica al autor en la base de datos del ISBN.
- **Editoriales.** Al igual que ocurre con los autores, la búsqueda únicamente se puede realizar por el nombre de la editorial o por la clave con la que se identifica a la editorial en la base de datos del ISBN.

El algoritmo define la entrada B a través de la cual recibirá la información de Libros, Categorías, Temas, Autores y Editoriales, datos más que suficientes para poder emitir el conjunto de recomendaciones sobre autores y obras que puedan ser del agrado del usuario cuyo perfil social se define a través de la entrada A del algoritmo. Destacar que, de un modo similar al descrito para las redes sociales, las herramientas de integración que ofrece la base de datos del ISBN permiten mediante un API REST con intercambio de datos en formato JSON, realizar consultas sobre su catálogo compuesto por casi ocho millones de libros y más de dos millones de autores, según estadísticas publicadas en su propio sitio web (3). Para cumplir con el objetivo de la investigación, la entrada B tendrá que encontrar autores y obras literarias basándose en las evidencias encontradas a través de la entrada A de la aplicación, conformada por los datos del perfil social del usuario: perfil, conexiones y actividad en las redes sociales donde éste tenga presencia. Por ello, en la mayoría de los casos la búsqueda al API del ISBN se realizará utilizando el tipo completo que se ha comentado antes ya que permite realizar una búsqueda amplia sobre autores, obras, editoriales, etcétera que cumplan con los criterios de búsqueda obtenidos a partir del perfil social del usuario. Únicamente en los casos claros en los que la aplicación determine con gran precisión que el usuario tiene afición por un determinado autor u obra, se realizará una búsqueda más precisa, utilizando algunas de las otras modalidades que el API del ISBN ofrece.

#### 3.1.3 Integración con el ISBNdb

Existen varias base de datos relacionadas con la literatura que permiten integrar con aplicaciones de terceros, ofreciendo APIs que permiten consultar la información que albergan en sus sistemas sobre autores y obras literarias en base a criterios de búsqueda

### 3. PROPUESTA

---

como son el título de la obra, el nombre del autor u otros que combinan varios criterios a la vez. Como se ha comentado, por el volumen que ofrecen (según su propia página web (3) casi 8 millones de autores y 3 millones de obras) se utilizará la base de datos del ISBNdb. A continuación se detalla la integración de la aplicación con el API que ofrece esta base de datos online:

```
GET http://isbndb.com/api/books.xml?access_key=BBS565DT&index1=full&
value1=nba,michael%20jordan
<ISBNdb server_time="2017-04-27T20:02:18Z">
<BookList total_results="180068" page_size="10" page_number="1" shown_results
="10"><BookData book_id="michael_jordan_a47" isbn="1561733997" isbn13=
"9781561733996"><Title>Michael Jordan</Title>
<TitleLong>Michael Jordan: MVP and NBA Champ</TitleLong>
<AuthorsText>Bob Sakamoto,</AuthorsText>
<PublisherText publisher_id="publications_international">Publications
International</PublisherText></BookData>
<BookData book_id="ultimate_jordan" isbn="0790784629" isbn13="9780790784625">
<Title>Ultimate Jordan</Title>
<TitleLong>Ultimate Jordan</TitleLong>
<AuthorsText>NBA Entertainment.</AuthorsText>
<PublisherText publisher_id="warner_home_video">Burbank, CA; Warner Home
Video</PublisherText>
</BookData>
<BookData book_id="michael_jordan_a49" isbn="1562940856" isbn13=
"9781562940850"><Title>Michael Jordan</Title>
<TitleLong>Michael Jordan (Millbrook Sports World)</TitleLong>
<AuthorsText>Bill Gutman,</AuthorsText>
<PublisherText publisher_id="millbrook_press">Millbrook Press</PublisherText>
</BookData>
<BookData book_id="the_official_nba_basketball_encyclopedia_a02" isbn=
"0385501307" isbn13="9780385501309">
<Title>The Official NBA encyclopedia</Title><TitleLong/>
<AuthorsText>
foreword by Michael Jordan; introduction by David J. Stern; edited by
Jan Hubbard</AuthorsText>
<PublisherText publisher_id="doubleday">New York : Doubleday, c2000.
</PublisherText></BookData>
<BookData book_id="michael_jordans_50_greatest_games" isbn="0806520302"
isbn13="9780806520308">
<Title>Michael Jordan's 50 greatest games</Title>
<TitleLong>Michael Jordan's 50 greatest games: from his NCAA championship
to six NBA titles</TitleLong>
<AuthorsText>Bob Condor</AuthorsText>
<PublisherText publisher_id="carol_pub_group">Secaucus, N.J. : Carol Pub.
Group, c1998.</PublisherText>
</BookData></BookList></ISBNdb>
```

## 3.1 Nuevo sistema de recomendación

---

Los datos anteriores muestran los resultados obtenidos en formato XML tras realizar la petición GET al API de la base de datos del ISBN para encontrar autores y obras literarias que contengan las palabras *NBA* y *Michael Jordan*, que representan, por un lado, a una organización deportiva y por otro lado a un deportista. Como se puede comprobar los resultados obtenidos son bastantes fiables, por lo que se puede determinar que si la entrada A del sistema genera evidencias claras, existen grandes probabilidades de que las recomendaciones generadas cumplan con el objetivo de satisfacer al usuario. Como se ha comentado, el ISBNdb permite realizar consultas a través de su API de diversas formas, utilizando algunas de las colecciones de datos que ofrecen: libros, autores, editores, temáticas, categorías y precios. Para el objetivo de esta investigación, el sistema de recomendación utilizará las siguientes colecciones de datos:

- **Libros**, utilizando los dos endpoints que ofrecen: *book* y *books*.
- **Autores**, utilizando los dos endpoints que ofrecen: *author* y *authors*.

Además, en la práctica totalidad de los casos, el sistema buscará, como ya se ha comentado, en las dos colecciones de datos, utilizando el índice completo y en un endpoint o en otro, dependiendo del número de potenciales recomendaciones que se pretendan obtener.

### 3.1.4 Estrategia para emitir recomendaciones

La pieza clave que marcará el diseño del sistema de recomendación propuesto será la información obtenida a través de la entrada A, es decir, todos los datos del perfil social conformado por la información que las redes sociales albergan sobre un determinado usuario. Este punto será vital para que el sistema de recomendación ofrezca una salida óptima y en este sentido seguirá una línea similar a la del sistema de recomendación de Amazon, cuyo principal valor es la gran cantidad de información que alberga en su sistema sobre sus usuarios. Con toda esta información el primer paso será organizarla y clasificarla para poder agrupar usuarios en torno a gustos e intereses similares sobre obras y autores literarios. Por cada usuario cuyo perfil social aparezca representado en la entrada A del sistema de recomendación no sólo se almacenará la información de dicho usuario sino la de todas sus relaciones con otros usuarios, información que se obtendrá de los contactos que dicho usuario tenga en cada una de las redes sociales

### 3. PROPUESTA

---

donde tenga presencia. Se presupone que en las redes sociales los usuarios establecen relaciones con otros usuarios en parte porque comparten un nexo de unión: pueden ser compañeros de trabajo, aficionados al baloncesto o por el contrario, un usuario sigue a otro porque es fan de éste. En este punto se puede comprobar que la literatura es una buena opción para ser presentada como entrada B del sistema de recomendación puesto que presenta multitud de géneros con los que cubrir cualquier tipo de conexión entre usuarios, independiente de si el usuario lee más o menos, aunque, por supuesto, cuanto más aficionado sea a la literatura un usuario mayor probabilidad de éxito tendrá el sistema de recomendación. Siguiendo con los ejemplos anteriores, si dado un usuario, éste presenta una relación con otros usuarios de una misma empresa informática, se presupone que trabajan en la misma empresa y que son aficionados o están interesados en libros de programación por lo que ésta será una buena aproximación para que el sistema emita recomendaciones de autores y obras literarias sobre lenguajes de programación. Del mismo modo, si un usuario es aficionado a equipos de la NBA, se presupone que este usuario y la gran mayoría de sus contactos pueden ser aficionados al baloncesto, y por tanto el sistema podrá emitir recomendaciones sobre autores y obras literarias que cubran este género. Finalmente, si un usuario sigue la cuenta en varias redes sociales de Steven Spielberg, el sistema de recomendación podrá emitir recomendaciones sobre biografías de este director de cine u otras obras que traten sobre él.

Por tanto, el sistema de investigación propuesto utilizará un filtrado colaborativo basado en la semejanza entre usuarios ya que para generar las recomendaciones se buscan usuarios con intereses similares. El principal problema de utilizar esta técnica es que ofrece un desempeño más bajo ante millones de usuarios, sin embargo, en el caso que se propone, se utilizará con el perfil social del usuario (entrada A) compuesto por el propio usuario y sus conexiones con otros usuarios en todas las redes sociales donde éste tenga presencia. Hablamos por tanto de un número mucho más limitado de usuarios por lo que el desempeño del sistema utilizando estas técnicas será mucho mayor. En este caso, otra de las grandes ventajas de utilizar las redes sociales como entrada A del sistema de recomendación es que no será necesario utilizar ningún algoritmo para clasificar usuarios similares, como puede ser el algoritmo del vecino más próximo, utilizado habitualmente en este tipo de casos, ya que se aprovechará la red de contactos de cada usuario en las redes sociales, donde en la mayoría de ellas, la calidad de la clasificación será aún mayor, ya que estos contactos suelen estar clasificados con mayor

### 3.1 Nuevo sistema de recomendación

---

nivel de detalle, por ejemplo en Twitter con sus listas que clasifican las conexiones de un usuario en función del nexa de unión: cuentas de usuarios sobre tecnología, baloncesto, cine, o Facebook con sus listas de amigos y familiares, por citar sólo algunos ejemplos reales que se pueden llegar a encontrar y que facilitan enormemente la tarea al sistema para emitir recomendaciones más precisas en base a la semejanza entre usuarios, para, de esta manera, poder emitir las mismas recomendaciones a dos usuarios semejantes.

Otro aspecto que marcará la clasificación de un determinado usuario será la información de su actividad social: de sus publicaciones, normalmente clasificadas mediante tags, de sus likes sobre determinadas publicaciones, que a su vez también estarán clasificadas y de los temas sobre los que muestre interés, normalmente catalogados en su perfil social, tal y como se ha descrito en el apartado donde se describe la entrada A del sistema. Con todos estos datos, el sistema debe ser capaz de determinar que, por ejemplo, un usuario es aficionado a la novela policiaca, bien porque lo describa de este modo en su perfil social, o porque presente un like sobre la última novela de Jeff Lindsay o publique un comentario sobre la serie Dexter, catalogándolo con un tag que mejore su clasificación y facilite la tarea al sistema de recomendación. A su vez, si una conexión de dicho usuario comparte este comentario, será una evidencia de que ambos usuarios comparten los mismos gustos, con lo que las posibilidades crecen exponencialmente para que el sistema de recomendación sea capaz de emitir sugerencias con gran probabilidad de éxito y por tanto este aspecto será clave en la estrategia a seguir por el sistema de recomendación propuesto: la información de las redes sociales siempre está perfectamente catalogada y bien relacionada con uno o varios de sus usuarios.

#### 3.1.5 Salida del sistema de recomendación

Teniendo en cuenta la entrada A del sistema de recomendación compuesto por la información en redes sociales del usuario, la entrada B definida mediante la información de autores y obras del ISBN, y siguiendo la estrategia detallada en el apartado anterior, el sistema de recomendación será capaz de emitir un conjunto de recomendaciones de autores y obras literarias que puedan ser del agrado del usuario. Como se ha comentado en apartados anteriores, el factor clave para que el sistema cumpla con el objetivo marcado se basa en que la entrada A genere el mínimo de evidencias necesarias para que sea posible encontrar resultados en la base de datos del ISBN. Dicha entrada presenta

### 3. PROPUESTA

---

tres grandes bloques de información: perfil, conexiones y actividad, de forma que el sistema emitirá recomendaciones basadas en tres grandes titulares:

- **Recomendaciones basadas en el perfil del usuario.** Fundamentalmente estarán basadas en el perfil del usuario en redes sociales, cuya información, como se ha podido ver, se nutre principalmente de Facebook y Google Plus. En las redes sociales el usuario define sus rasgos y las principales características que lo definen, y en base a éstas el sistema emitirá recomendaciones sobre autores y obras literarias.
- **Recomendaciones basadas en las conexiones del usuario.** Se basan en otra recomendaciones ya emitidas a conexiones del usuario y en nuevas recomendaciones basadas en las personas a las que sigue el usuario. En el primer caso, si no existieran, únicamente se tendría en cuenta el segundo caso, y por tanto se tomarían como base las personas que sigue el usuario, y no tanto a sus amigos, para buscar evidencias en la base de datos del ISBN y emitir recomendaciones. De esta forma, si ninguna conexión del usuario catalogada como amigo presenta recomendaciones efectuadas por el sistema, se analizarán las conexiones que representan las personas a las que sigue el usuario para buscar evidencias con las que cruzar información en la base de datos del ISBN en busca de recomendaciones. Por ejemplo, si el usuario sigue al jugador de baloncesto *Lebron James*, el sistema podrá buscar en el ISBN esta evidencia y encontrar el libro *Lebron James (The World's Greatest Athletes)* de *John Walters*, que debe ser tenido en cuenta para emitir una recomendación.
- **Recomendaciones basadas en lo que le gusta al usuario.** Este último bloque de recomendaciones estará basado en la información más actualizada que genera el usuario en las redes sociales, y no es otra que su actividad social. Tal y como se comentó en apartados anteriores, se analizarán los likes sobre páginas de grupos y/o eventos de Facebook, las publicaciones favoritas del usuario en Twitter, los metadatos de las imágenes que publique en Instagram o los paneles que comparta en Pinterest. Con toda esta información el sistema sera capaz de generar evidencias con las que poder emitir conjuntos de recomendaciones para el usuario en cuestión.

### 3.1 Nuevo sistema de recomendación

A continuación se detalla un ejemplo real sobre como se desempeña el sistema de recomendación cuando a su entrada A llega el usuario *Emiliano Sánchez Gil* con perfil en las seis redes sociales objeto de la investigación. Este usuario es el que ha sido utilizado en los apartados anteriores para representar como se integraba la aplicación con cada red social. El ejemplo muestra la salida del sistema a través de su API, cuyos datos serán interpretados por la aplicación para ser mostrados al usuario. El usuario será identificado a través de OAuth, para que también éste pueda otorgar su permiso para que la aplicación acceda a sus cuentas en redes sociales con el objetivo de obtener la información necesaria para generar las evidencias que sirvan al sistema para emitir las recomendaciones:

```
GET https://api.socialrecommendation.com/v1/~?format=json
{
  "size":"2"
  "user":{
    "size":"0", "books":[]
  },
  "connections":{
    "size":"1", "books":[
      {
        "title":" Maraton Puedes Hacerlo",
        "author":" Jeff Galloway, Inma Estany (Translator)"
      }
    ]
  },
  "activity":{
    "size":"1", "books":[
      {
        "title":"NBA basketball: an official fan's guide",
        "author":"Mark Vancil"
      }
    ]
  }
}
```

Los datos devueltos por el API se estructuran en tres grandes bloques: *user*, que incluye las recomendaciones basadas en el perfil del usuario, *connections*, que incluye las recomendaciones basadas en las conexiones del usuario y *activity*, que incluye las recomendaciones basadas en la actividad del usuario en redes sociales. Para el global de las recomendaciones, se incluye un atributo que indica cuantos resultados ha devuelto el sistema y para cada uno de los bloques, se incluye a su vez un nuevo atributo

### 3. PROPUESTA

---

que indica cuantas recomendaciones de cada apartado ha sido capaz de generar el sistema. Teniendo en cuenta esta estructura, para el caso anterior, el sistema devuelve dos recomendaciones para el usuario definido a través de la entrada A. Se explica a continuación como ha procedido el sistema para emitir los tres tipos de recomendaciones:

- El caso representado ejemplifica uno de los problemas que se puede encontrar el sistema a la hora de emitir recomendaciones: el perfil del usuario en redes sociales es pobre en cuanto a información cumplimentada por éste. Los pocos atributos informados por el usuario: colegio, universidad o lugar de trabajo, no son lo suficientemente fiables como para que el sistema los tenga en cuenta como evidencias para emitir recomendaciones. En este caso, es preferible que la aplicación no devuelva resultados en este bloque de información, antes que hacerlo con alguna recomendación que no tenga éxito con el usuario. Por último, destacar un aspecto relevante relacionado con uno de los atributos del perfil que si está cumplimentado por el usuario: los deportistas favoritos. En este caso, se incluyen dos registros, *Saleta Castro* y *Valentí Sanjuan*, que parecen evidencias claras para realizar búsquedas en la base de datos del ISBN. El sistema no encuentra resultados, sin embargo, contrastando con otras fuentes, se confirma que el segundo registro presenta varias obras literarias de las que es autor. Este punto plantea la necesidad de estudiar la conveniencia de incorporar otras fuentes de datos para la entrada B del sistema, como puede ser el API de Google Books (49).
- El sistema evaluó las conexiones del usuario en redes sociales. En primer lugar, se buscaron conjuntos de recomendaciones ya emitidas para las conexiones formadas por los amigos del usuario, y lógicamente no se encontraron resultados. Por ello, se analizaron las conexiones formadas por las personas a las que seguía el usuario. En este caso, se encontraron varias evidencias, entre otras, que el usuario seguía recientemente la página en Facebook del Maratón de Bilbao por lo que se generó una recomendación que cumplía con las características de esta página.
- Finalmente, con el último bloque el sistema evaluó la actividad social del usuario y encontró varias evidencias con un nexo común: un *retweet* sobre un comentario de un jugador de baloncesto, un *board* sobre la *NBA* y un *like* a un programa de radio de la *NBA*. Por todo ello, el sistema emitió una recomendación sobre un libro de baloncesto de la *NBA*.

### 3.1 Nuevo sistema de recomendación

Sin embargo, el sistema de recomendación que se plantea en esta propuesta de investigación también tiene como objetivo emitir recomendaciones sobre autores y obras literarias en función de las evidencias encontradas en el perfil social del usuario sobre literatura, y no únicamente basadas en aficiones del usuario como puedan ser la música o el deporte, lo cual si cabe, añade un punto más de complejidad al sistema. A continuación se detalla un ejemplo real y se explica posteriormente como procede el sistema para emitir estas recomendaciones:

```
GET https://api.socialrecommendation.com/v1/~?format=json
{
  "size": "3"
  "user": {
    "size": "0", "books": []
  }
  "connections": {
    "size": "2", "books": [
      {
        "title": "Ketchup Clouds",
        "author": "Annabel Pitcher"
      },
      {
        "title": "My sister lives on the mantelpiece",
        "author": "Annabel Pitcher"
      }
    ]
  },
  "activity": {
    "size": "1", "books": [
      {
        "title": "Hollywood",
        "author": "Bukowski"
      }
    ]
  }
}
```

Nuevamente, los datos devueltos por el API se estructuran en tres grandes bloques: *user*, que incluye las recomendaciones basadas en el perfil del usuario, *connections*, que incluye las recomendaciones basadas en las conexiones del usuario y *activity*, que incluye las recomendaciones basadas en la actividad del usuario en redes sociales. Siguiendo esta estructura, el sistema devuelve tres recomendaciones para el usuario definido a través de la entrada A. Se explica a continuación como ha procedido el sistema para emitir los tres tipos de recomendaciones:

### 3. PROPUESTA

---

- De nuevo para este bloque de recomendaciones el sistema no devuelve ningún resultado, ya que a través del perfil en redes sociales, es difícil poder obtener evidencias con las que emitir recomendaciones. De hecho, únicamente el perfil de Facebook especifica un bloque de información donde el usuario puede incluir los libros que ha leído o quiere leer. Si este bloque de datos tuviera información del usuario, sería muy sencillo para el sistema emitir recomendaciones basadas en otras obras literarias del mismo autor para el que el usuario ya la leído algún libro. Sin embargo, no es común que este apartado de información tenga datos del usuario, como ocurre con este caso real, y por tanto, lo normal es que el sistema no devuelva recomendaciones.
- El sistema evaluó las conexiones del usuario en redes sociales. Como ocurrió con el caso anterior, se buscaron conjuntos de recomendaciones ya emitidas para las conexiones formadas por los amigos del usuario, y lógicamente no se encontraron resultados. Por ello, se analizaron las conexiones formadas por las personas a las que seguía el usuario. En este caso, se encontró una evidencia en uno de los usuarios a los que seguía el usuario en Twitter, ya que se trataba de *Annabel Pitcher*, autora de varios libros, por lo que el sistema lo tuvo en cuenta para emitir recomendaciones relacionadas con ésta. El sistema analiza el perfil del usuario para determinar si se trata de un escritor/a, siendo esta tarea menos compleja si se analizan los perfiles de Facebook o Google Plus, donde éstos aparecen categorizados como *autor*. En Twitter, como ocurre con este caso, el sistema debe analizar las líneas que el propio usuario informa en su perfil y buscar palabras clave (*author, autor, escritor, writer...*). En este caso, la información del perfil de Annabel Pitcher mostraba el siguiente texto: *Award-winning author of My Sister Lives on the Mantelpiece, Ketchup Clouds and Silence Is Goldfish. Wife. Mum.*, siendo su nombre de usuario *@APitcherAuthor*, información suficiente para que el sistema obtuviera las evidencias con las que emitir las recomendaciones.
- Finalmente, con el último bloque el sistema evaluó la actividad social del usuario, con especial énfasis en las palabras marcadas como *hashtag*, elemento que apareció en Twitter pero que se ha hecho prácticamente común a todas las redes sociales, en parte gracias a las aplicaciones utilizadas por algunos usuarios y que permiten lanzar comentarios en varias redes de forma simultánea. En el caso del ejemplo

### 3.1 Nuevo sistema de recomendación

se encontró el mismo texto en Twitter e Instagram con el hashtag *bukowski*. Partiendo de esta evidencia, el sistema encontró resultados a través de la entrada B y por tanto pudo generar la recomendación que se muestra en el ejemplo.

Para entender mejor los ejemplos anteriores, a continuación se muestra un diagrama que, de manera esquemática, representa como procede el algoritmo para encontrar las evidencias en la entrada A, realizar la búsqueda a través de la entrada B, y finalmente generar las recomendaciones en la salida del sistema.

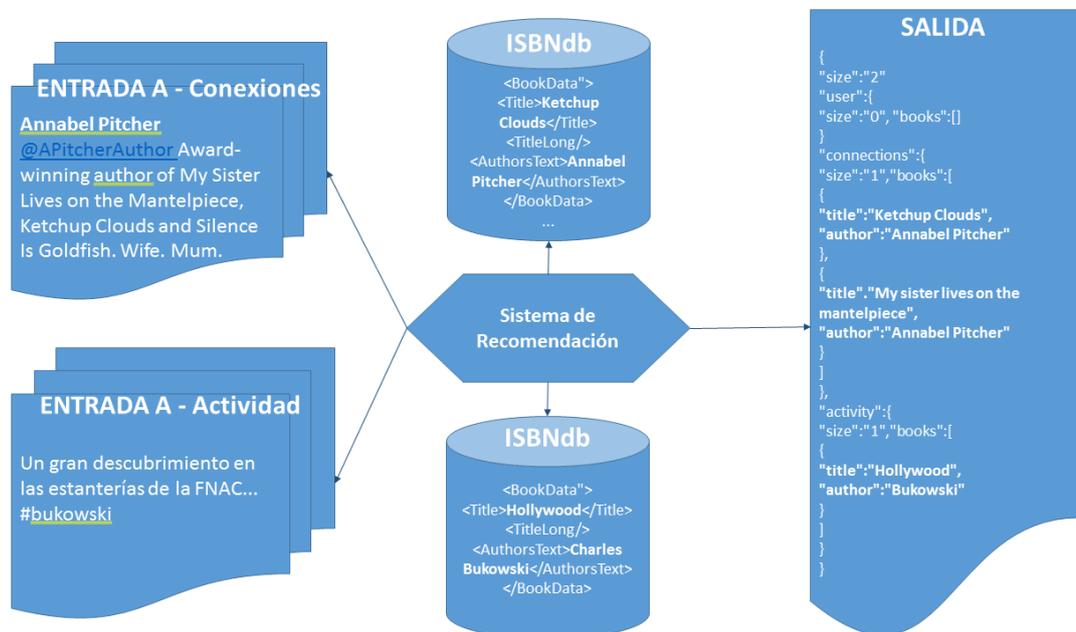


Figure 3.3: Diagrama de la Salida del Recomendación.

Por último, destacar que una de las principales líneas de futuro de la investigación debe ser evolucionar el sistema de recomendación de manera que éste cruce la información obtenida a través de cada uno de los tres grandes bloques de datos generados en la entrada A del sistema, es decir, cruce la información del perfil con sus conexiones, las conexiones con la actividad y el perfil con la actividad, para tener la certeza de que las evidencias generadas permitirán garantizar el éxito de las recomendaciones emitidas. Con este cruce de información se podrá potenciar la información de uno de los bloques cuando ésta por si sola no ofrezca muchas garantías. Por ejemplo, en el caso expuesto con anterioridad el perfil no ofrecía mucha información cumplimentada por el usuario,

### **3. PROPUESTA**

---

pero si incluía algunos atributos como el colegio, la universidad o el lugar de trabajo. Si éstos hubieran sido cruzados con la información de la actividad social del usuario quizás se podría haber determinado que el usuario no sólo trabaja en el Departamento de Sistemas de Información de una empresa de seguros, sino que además es aficionado a la informática, por todas las publicaciones que realiza en las redes sociales. Con este cruce de información se podrían haber buscado autores y obras literarias sobre temas de informática de interés para el usuario. De esta forma, el valor de la información de cada uno de los bloques se multiplica para generar recomendaciones de mayor calidad y con mayores probabilidades de éxito.

#### **3.2 Cuestionario para captar usuarios**

Tal y como se comentaba en el primer capítulo de esta propuesta de investigación, se diseñará un cuestionario para atraer usuarios con los que poder probar el nuevo sistema de recomendación en un entorno lo más cercano posible a la realidad. El procedimiento a seguir se basará en enviar un email donde se presentará la investigación y se invitará al destinatario a participar en la misma. Para ello, deberá acceder a una página web donde se explicarán los objetivos de esta investigación, el desarrollo llevado a cabo y se mostrará el cuestionario que deberá cumplimentar el usuario interesado en participar en la investigación. Las preguntas que lo compondrán irán destinadas a segmentar el usuario que ha accedido para determinar si es válido para participar en la investigación y en caso afirmativo darle acceso a la misma. A continuación se detallará la composición del cuestionario que se presentará en la citada página web.

## 3.2 Cuestionario para captar usuarios

### RECOMENDACIONES LITERARIAS BASADAS EN TU PERFIL SOCIAL

Las redes sociales son uno de los grandes fenómenos de nuestro tiempo, nos ayudan a estar informados, conectados con otros e incluso nos ayudan a aprovechar todas las posibilidades que nos ofrece Internet. Aunque asociamos red social a las herramientas que hoy en día nos ofrecen Facebook o Twitter, hace ya muchos años que varios científicos e investigadores comenzaron a explorar a grupos de individuos u organizaciones relacionados por un criterio común, es lo que en sus orígenes se conoció como red social. Desde entonces esos trabajos de investigación han posibilitado llegar hasta nuestros días donde las herramientas que conocemos son parte de nuestra actividad diaria.

Si has llegado hasta aquí tienes la posibilidad de colaborar con un proyecto de investigación que pretende explorar nuevas posibilidades que ofrecen las redes sociales. ¿Cómo lo haremos?. Queremos conocerte a través de tu actividad en las redes sociales y ofrecerte recomendaciones sobre libros y autores que pueden ser de tu interés. Creemos que a través de las redes sociales podemos conocer lo suficiente sobre tí como para poder recomendarte un libro o un autor con una alta probabilidad de que te guste. Tan sólo debes contestar a las siguientes preguntas:

1. Selecciona en cuál de las siguientes redes sociales estás registrado:

- |                                     |                                       |                                 |
|-------------------------------------|---------------------------------------|---------------------------------|
| <input type="checkbox"/> Facebook.  | <input type="checkbox"/> Pinterest.   | <input type="checkbox"/> Otras. |
| <input type="checkbox"/> Instagram. | <input type="checkbox"/> LinkedIn.    |                                 |
| <input type="checkbox"/> Twitter.   | <input type="checkbox"/> Google Plus. |                                 |

2. ¿Con qué frecuencia accedes a las redes sociales?.

- |   |   |
|---|---|
| <input type="radio"/> Todos los días varias veces.    | <input type="radio"/> Todos los meses varias veces.         |
| <input type="radio"/> Una vez al día.                 | <input type="radio"/> Una vez al mes.                       |
| <input type="radio"/> Todas las semanas varias veces. | <input type="radio"/> No recuerdo la última vez que accedí. |
| <input type="radio"/> Una vez a la semana.            |   |

3. ¿Cuántos libros lees al año?.

- |                           |                           |                             |                                  |
|---------------------------|---------------------------|-----------------------------|----------------------------------|
| <input type="radio"/> 0-2 | <input type="radio"/> 4-6 | <input type="radio"/> 8-10  | <input type="radio"/> Más de 12. |
| <input type="radio"/> 2-4 | <input type="radio"/> 6-8 | <input type="radio"/> 10-12 |                                  |

### SEIS GRADOS DE SEPARACIÓN

En 1930 el escritor húngaro Frigyes Karinthy escribió el cuento Chains donde habló por primera vez de la hipótesis de los Seis Grados de Separación, que afirma que cualquier persona en la Tierra puede estar conectado a cualquier otra persona del planeta a través de una cadena de conocidos que no tiene más de cinco intermediarios. Este tema ha sido asunto central de investigación de numerosos científicos e investigadores, incluso las principales redes sociales realizaron estudios al respecto: Facebook determinó que el máximo grado de separación era de 3,57 y Twitter de 4,59. ¿Quieres ser partícipe de un estudio que permita determinar el máximo grado de separación entre tú, aficionado a la literatura, y otros aficionados a la literatura?.

- Sí.  No.

Si deseas participar en esta investigación y colaborar en la mejora y evolución de las redes sociales, sacándole el máximo partido a todas las posibilidades que nos ofrecen, tan sólo debes aceptar el siguiente consentimiento para que el sistema de recomendación sea capaz de acceder a tu perfil en redes sociales y de este modo, a través de tu actividad, sea capaz de recomendarte tu próxima lectura. ¡¡Gracias!!.

- Sí.  No.

### 3. PROPUESTA

---

En las líneas anteriores se ha presentado el detalle del cuestionario que se utilizará para introducir la investigación e invitar a todo aquel que lo desee a participar en la misma, probando el correcto desempeño del sistema de recomendación tanto para emitir recomendaciones sobre autores u obras literarias presentes en el ISBN como también para probar la base del mismo, intentando demostrar la hipótesis de los Seis Grados de Separación siempre dentro del contexto de la investigación. A continuación se detallan algunos puntos relacionados con el cuestionario anterior, que se deben tener en cuenta:

- El cuestionario introduce la investigación e intenta atraer usuarios, que serán la base para poder probar con un caso real si el sistema de recomendación ofrece los resultados esperados.
- Los dos puntos principales del usuario sobre los que se desea obtener información a través del cuestionario son:
  - Si realmente el usuario es aficionado a la literatura, y en este caso, la generación de valor en forma de recomendaciones puede llegar a serle de utilidad. Este punto servirá para averiguar si la literatura es una buena base como entrada B del sistema de recomendación o sería conveniente plantear una línea de futuro con el cine o recursos multimedia en general.
  - Si tiene actividad en redes sociales, ya que si es usuario esporádico o únicamente utiliza una red social, las probabilidades de obtener los datos suficientes para que el sistema genere recomendaciones adecuadas se reduce considerablemente.
- El cuestionario debe dejar claro que no resulta necesario realizar preguntas personales tales como la edad o el sexo del usuario, ya que podrían echar atrás a alguien interesado en la investigación pero que no desea cumplimentar un cuestionario que incluya muchas preguntas. Además estas preguntas no son necesarias si el usuario acepta participar en la investigación y da su consentimiento para que el sistema acceda a sus cuentas en redes sociales, puesto que será posible recuperar la información que existe en cada perfil de red social.

### 3.3 Encuesta para evaluar el sistema

Dentro de la investigación, tan importante será desarrollar un sistema que emita recomendaciones de calidad, como establecer las pautas necesarias para medir que dichas recomendaciones han generado el efecto deseado y por tanto han sido del agrado del usuario al que han sido dirigidas. De esta forma, como se comentó en el primer capítulo de esta propuesta de investigación, se diseñará un sistema de evaluación para poder detectar si la aplicación genera recomendaciones de calidad e intentar generar mejoras evolutivas del sistema que conformen las líneas de futuro de la investigación, siempre con el objetivo de mejorar el desempeño del mismo.

La base de este sistema de recomendación será diseñada a través de una técnica de generación de datos: la encuesta. De este modo, para todos los usuarios que participen en la investigación y una vez que hayan visto las recomendaciones sobre autores y obras literarias que el sistema les haya propuesto en base a su perfil social, se lanzará una invitación para participar en una encuesta, cuyo objetivo principal será valorar el conjunto de recomendaciones que han recibido, así como proponerles que inviten a otros amigos a participar en la investigación. A continuación se presenta la encuesta, que tendrá tantos puntos como recomendaciones haya recibido el usuario del sistema, de forma que pueda valorar cada una de ellas de manera independiente:

### 3. PROPUESTA

---

#### VALORA TUS RECOMENDACIONES

¡Has terminado!. Gracias por participar en esta investigación y hacer posible los resultados que acabas de recibir. Tan sólo te pedimos un último esfuerzo para que cumplimentes la siguiente encuesta. No tardarás más de 5 minutos y a nosotros nos servirá de mucho. Queremos mejorar el sistema de recomendación y para ello, tu opinión es muy valiosa. Además queremos aprovechar las recomendaciones emitidas para ofrecerlas a otros usuarios similares a tí... Sin embargo, sólo queremos hacerlo si éstas te han resultado útiles. ¡Gracias por tu tiempo!.

1. Valora la recomendación (el 1 refleja que la recomendación no te ha servido para nada y el 5 indica que la recomendación te ha encantado): *Annabel Pitcher*

1.                       2.                       3.                       4.                       5.

Si la recomendación no te ha gustado, indica el motivo:

- No me gusta el escritor.                       No conozco al escritor.  
 He leído todos sus libros.  
 El escritor no tiene libros en mi idioma.                       Otros.

Si por el contrario la recomendación te ha encantado, dínos el por qué:

- Es mi escritor favorito.                       Es un escritor de un estilo similar a otros que me gustan.  
 Tenía ganas de leer un libro de este escritor.  
 Ya me lo habían recomendado.                       Otros

2. Valora la recomendación (el 1 refleja que la recomendación no te ha servido para nada y el 5 indica que la recomendación te ha encantado): *Ketchup Clouds*

1.                       2.                       3.                       4.                       5.

Si la recomendación no te ha gustado, indica el motivo:

- No me gusta el libro.                       Es un libro que empecé y lo dejé a medias.  
 Ya lo había leído.  
 El libro está en un idioma que no conozco.                       Otros.

Si por el contrario la recomendación te ha encantado, dínos el por qué:

- Es mi libro favorito.                       Es una libro escrito por un autor que me gusta.  
 Tenía ganas de leer este libro.  
 Ya me lo habían recomendado.                       Otros

¡Esto ha sido todo!. Si quieres que tus amigos prueben este sistema y reciban nuevas recomendaciones o algunas similares a las tuyas, por favor, déjanos su email y nos pondremos en contacto con ellos. ¡Muchas gracias!.

### 3.3 Encuesta para evaluar el sistema

---

En las líneas anteriores se ha presentado el detalle de la encuesta que se utilizará para evaluar los resultados obtenidos por el sistema de recomendación con el fin de mejorar el desempeño del mismo a través de la definición de líneas de futuro de la investigación. A continuación se detallan algunos puntos relacionados con el cuestionario anterior, que se deben tener en cuenta:

- La encuesta valora tanto las recomendaciones de autores como de obras literarias, ya que es posible que un autor sea del agrado del usuario pero la obra no, y viceversa.
- Se pide al usuario que sugiera a otros amigos la utilización del sistema para generar recomendaciones, de manera que, algunas de las que se han emitido para dicho usuario podrían ser propuestas para sus amigos. En este punto entra en juego el sistema de evaluación como elemento diferencial, no sólo para definir las líneas de evaluación del sistema, sino para tener en cuenta las valoraciones de una recomendación a la hora de utilizarlas con otro usuario, de forma que, por ejemplo, un libro escrito en un idioma desconocido, no debe ser tenido en cuenta como recomendación para otro usuario. Y de la misma forma, por ejemplo, un libro que ya había sido recomendado a un usuario (no a través del sistema) puede ser una buena evidencia para reutilizar dicha recomendación con otros amigos del usuario.
- La encuesta pretende diferenciar las recomendaciones que no han sido de calidad, por ejemplo por el idioma, de aquellas que no han dado el resultado esperado con un determinado usuario pero que sí podrían funcionar adecuadamente con otro usuario, por ejemplo aquellas que dicho usuario ha valorado negativamente por ser un obra que ya había leído, por poner un ejemplo en concreto.

### **3. PROPUESTA**

---

## 4

# Conclusiones

En esta propuesta de investigación se han detallado los objetivos principales que se pretenden alcanzar y las técnicas y estrategias a seguir para conseguirlo. Además, se ha elaborado el estado del arte para conocer otras propuestas realizadas dentro del mismo ámbito por otros investigadores y científicos. Finalmente, se detalló la propuesta de investigación basada en un sistema de recomendación, explicando las dos entradas del sistema: perfil social del usuario y base de datos de autores y obras literarias, así como la salida del mismo, compuesta por un conjunto de recomendaciones sobre autores y obras literarias que deberían ser del agrado del usuario. Con este trabajo realizado, a continuación se detallan las principales conclusiones a las que se han llegado durante la elaboración de esta propuesta de investigación:

- Como ya se comentó en la explicación de los casos reales de desempeño del nuevo sistema de recomendación, parece que la elección del API del ISBNdb (3) como fuente de información para la entrada B del sistema no es la opción más adecuada a pesar de ofrecer un catálogo muy amplio, puesto que los resultados que ofrecen en algunos casos muestran libros y escritores desactualizados o que no incluyen las últimas novedades, aspecto relevante dado que la otra fuente de datos muestra información muy actualizada, ya que trabaja, por ejemplo, con la información generada a través de la actividad social más reciente de un determinado usuario. Además otro aspecto relevante es que esta base de datos se encuentra en inglés, lo que puede ser un problema a la hora de generar conjuntos de recomendaciones a usuarios que no dominen esta lengua. Por todo ello, se plantea como tarea explorar otras bases de datos importantes como las que ofrecen Google Books

#### 4. CONCLUSIONES

---

(49), Good Reads (50) o Amazon WS (51), con su opción para búsqueda de libros con el ISBN, aunque parece basada en el API que ofrece el ISBNdb.

- En cuanto a la entrada A del sistema de recomendación, conformada por el perfil social del usuario basado en las principales redes sociales: Facebook, Instagram, Twitter, Pinterest, LinkedIn y Google Plus, en algunos casos ya se descarta la información de LinkedIn o Google Plus al ser poco relevante y se plantea explorar más descartes para simplificar la entrada de información al sistema, fundamentalmente centrándose en dos de las principales redes sociales: Facebook y Twitter, donde están la mayoría de usuarios potenciales de la aplicación y valorar la utilización de la información de otras redes sociales como los metadatos de las imágenes de Instagram o los *boards* de Pinterest sólo en los casos donde realmente vayan a aportar valor. En definitiva, si estas redes sociales no tiene un gran uso por parte del usuario es preferible descartarlas para no introducir ruido al sistema de recomendación y centrarse únicamente en aquellas redes donde el usuario tiene mucha presencia, y éstas normalmente serán Twitter y Facebook.
- Uno de los puntos más importantes de la propuesta de investigación será la estrategia a seguir para valorar si las recomendaciones emitidas ha cumplido con el objetivo de agradar al usuario. Para ello, la pieza básica a utilizar será una encuesta que se lanzará a los usuarios que ya hayan recibido sus recomendaciones. Esta técnica de generación de datos tendrá un doble cometido: por un lado establecer una valoración de las recomendaciones emitidas, y por otro lado, crecer en número de usuarios del sistema, siempre a través de conexiones de otros usuarios que ya lo han utilizado, aspecto clave para poder reutilizar recomendaciones, teniendo muy en cuenta la valoración ya emitida sobre éstas. Estos puntos son muy importantes dentro de una estrategia de filtrado colaborativo, tal y como se comentó en apartados anteriores.
- La hipótesis de los Seis Grados de Separación planteada en el primer capítulo como una de las justificaciones para emprender esta investigación quizás no debería ser un objetivo de la misma para centrarse en lo verdaderamente importante: emitir recomendaciones de calidad para el usuario en base a su perfil social. Sí puede ser una justificación para utilizar las redes sociales como elemento de valor en el

---

ámbito de los sistemas de recomendación pero no debería ser un aspecto clave a explorar dentro del transcurso de la investigación. Se puede plantear como objetivo, una vez que el sistema esté rodado, utilizar los elementos de integración desarrollados para obtener la información de las conexiones de un usuario en redes sociales, con el fin de probar la hipótesis, generando un artículo de investigación que presente los resultados obtenidos.

#### 4. CONCLUSIONES

---

## 5

# Apéndice

En este capítulo se incluyen algunos de los fragmentos del código fuente que se ha desarrollado para la implementación del nuevo sistema de recomendación utilizado tanto para realizar la integración con las distintas redes sociales, y de este modo conformar el perfil social del usuario para la entrada A del sistema, como para integrar con la base de datos del ISBN y así generar los datos necesarios para la entrada B. Para la implementación se ha utilizado el lenguaje de programación Scala (52) y el framework Play (53) asociado a dicho lenguaje.

### 5.1 Integración con Twitter

A continuación se muestra un bloque de código que permite la integración con Twitter para recuperar la lista de tweets favoritos de un determinado usuario cuyas credenciales son previamente validadas mediante OAuth. En este caso se utiliza como librería de apoyo Twitter4j (54), desarrollada en Java, la cual facilita la integración con esta red social.

## 5. APÉNDICE

---

```
object Application extends Controller {

  def index = Action {
    val cb = new ConfigurationBuilder()
    cb.setDebugEnabled(true)
      .setOAuthConsumerKey("eiSB6r7lZCm1lu78Is0UIQ")
      .setOAuthConsumerSecret("Awvaztigu7xpKX97vYwtmDZdOwUOQVGbIWDYg")
      .setOAuthAccessToken("266732707-gGqlSex06ye4DQi5edYF1KE42K7wdvKuk9X")
      .setOAuthAccessTokenSecret("DU43byw580QjMw2ucExBIgTXP5ztpWUunIM8FzGs")

    val twitterFactory = new TwitterFactory(cb.build())
    val twitter = twitterFactory.getInstance()

    val favorites = twitter.getFavorites()
    val iteratorFavorites = favorites.iterator()

    var buffer = ""
    while (iteratorFavorites.hasNext()) {
      val favorite = iteratorFavorites.next()
      buffer+="Usuario: " + favorite.getUser().getName() +
        " - Tweet:" + favorite.getText() + "\n"
    }

    Ok(buffer)
  }
}
```

Como ejemplo, se ha mostrado un controlador de Play que representa por pantalla cada uno de los tweets marcados como favoritos por parte del usuario, así como el nombre del usuario que escribió el tweet.

### 5.2 Integración con Facebook

De forma similar a lo que se ha mostrado en el apartado anterior con Twitter, en este caso se muestra un bloque de código que permite la integración con Facebook para recuperar la información del perfil del usuario que previamente ha permitido el acceso de la aplicación para consultar la información que esta red social almacena sobre dicho usuario. En este caso se utiliza como librería de apoyo Facebook4j (55), desarrollada en Java, que facilita la integración con esta red social.

```
object Application extends Controller {  
  
  def index = Action {  
  
    val configurationBuilder = new ConfigurationBuilder();  
    configurationBuilder.setDebugEnabled(true)  
                        .setOAuthAppId("258121334292831")  
                        .setOAuthAppSecret("5fd008bfc8d0e9db092b2fb042cfb")  
                        .setOAuthPermissions("email , publish_stream")  
                        .setOAuthAccessToken("KXtN5TVczZC9YurGFqVTtEIZD")  
  
    val facebookFactory = new FacebookFactory(configurationBuilder.build())  
    val facebook = facebookFactory.getInstance()  
    val user = facebook.getMe()  
  
    Ok(user.getName + " - " + user.getBirthday)  
  }  
}
```

Como ejemplo, se ha mostrado un controlador de Play que representa por pantalla el nombre y apellidos del usuario y su fecha de nacimiento, datos que junto a otros, se almacenan en el perfil que dicho usuario tiene en Facebook.

## 5.3 Integración con ISBNdB

El siguiente bloque de código ilustra el procedimiento llevado a cabo para integrar con la base de datos del ISBN, en este caso buscando todos los libros de la autora *Annabel Pitcher*. Para facilitar la integración, de manera similar a cómo se ha seguido con Twitter y Facebook se utiliza una librería de terceros (56) implementada en Java, que facilita la integración con esta base de datos.

## 5. APÉNDICE

---

```
object Application extends Controller {  
  
  def index = Action {  
  
    val booksProxy = BooksProxy.getInstance()  
    booksProxy.setDeveloperKey("BBS565DT1321")  
    val books = booksProxy.getBooks("Annabel Pitcher")  
    val iteratorBooks = books.iterator()  
  
    var buffer = ""  
    while (iteratorBooks.hasNext()) {  
      val book = iteratorBooks.next()  
      buffer+="Titulo: " + book.getTitle() + "\n"  
    }  
  
    Ok(buffer)  
  }  
}
```

Como ejemplo, se ha mostrado un controlador de Play que representa por pantalla el título de cada uno de los libros que contiene la base de datos del ISBN y que cumplen con la condición de que la autora sea *Annabel Pitcher*.

# Bibliografía

- [1] **What's in a billion? Facebook users hit milestone in 1 day.** *New Orleans CityBusiness*. 1
- [2] J. ROWLEY. **The wisdom hierarchy: representations of the DIKW hierarchy.** *Journal of Information Science*, **33**(2), 2007. 1
- [3] ISBNDB. **International Standard Book Number Database.** <http://isbndb.com>). 2, 43, 44, 61
- [4] F. KARINTHY. **Chains.** *Everything is Different*. 2
- [5] S. MILGRAM. **The small world problem.** *Psychology Today*, **2**. 2
- [6] D. WATTS. **Six Degrees: The Science of a Connected Age.** *Psychology Today*. 2
- [7] B. J. OATES. **Researching Information Systems and Computing.** 5, 6
- [8] N. B. ELLISON D. M. BOYD. **Social Network Sites: Definition, History and Scholarship.** *Journal of Computer-Mediated Communication*, **13**(1):210–230, 2007. 11
- [9] L. C. FREEMAN. **The development of social network analysis: a study in the sociology of science.** *Empirical Press of Computer-Mediated Communication*. 11
- [10] **Las Redes Sociales en Internet.** *Observatorio Nacional de las Telecomunicaciones y de la SI*. 12
- [11] CIO. **15 social networks with the most active users in 2015.** <http://www.cio.com/article/3014362/social-networking/15-social-networks-with-the-most-active-users-in-2015.html>. 13
- [12] F. THIESING F. MULLER. **Social networking APIs for companies: An example of using the Facebook API for companies.** *2011 International Conference on Computational Aspects of Social Networks (CASoN)*. 13
- [13] FACEBOOK. **API Graph.** <https://developers.facebook.com/docs/graph-api/using-graph-api>. 13
- [14] N. V. FLOR. **Technology Corner: Automated data extraction using Facebook.** *The Journal of Digital Forensics*, **7**(2):149–164, 2012. 13
- [15] OAUTH. **OAuth 2.0.** <http://oauth.net>. 13, 31
- [16] INSTAGRAM. **Instagram API Platform.** <https://www.instagram.com/developer>. 13, 15
- [17] W. TSAI R. PENG, D. SUN. **Success factors in mobile social networking application development: case study of Instagram.** *In Proceedings of the 29th Annual ACM Symposium on Applied Computing*. 14
- [18] R. ABOUKHALIL. **Using the Twitter API to mine the Twitterverse, XRDS Crossroads.** *The ACM Magazine for Students*, **19**(4):52–55, 2013. 14
- [19] TWITTER. **REST APIs.** <https://dev.twitter.com/rest/public>. 14, 35
- [20] PINTEREST. **Pinterest API.** <https://developers.pinterest.com>. 14, 40
- [21] C. HALL M. ZARRO. **Exploring Social Curation.** *D-Lib Magazine*, **18**. 14
- [22] LINKEDIN. **REST API.** <https://developer.linkedin.com>. 14, 33
- [23] D. BRADBURY. **Data mining with LinkedIn.** *Computer Fraud & Security*, **10**(5), 2011. 14
- [24] A. KAMEAS K. TOGIAS. **An Ontology-based Representation of the Google+ API.** *The Third International Conference on Models and Ontology-based Design of Protocols, Architectures and Services MOPAS*. 14
- [25] GOOGLE PLUS. **Google+ API.** <https://developers.google.com/+web/api/rest>. 14, 31
- [26] M. TKALCIC B. FERWERDA, M. SCHEDL. **Using Instagram picture features to predict users personality+ API.** *Multimedia Modeling*, **9516**. 15
- [27] B. CHUN T. KWON H. KIM Y. CHOI J. HAN, D. CHOI. **Collecting, organizing, and sharing pins in Pinterest: interest-driven or social-drive?** *ACM SIGMETRICS Performance Evaluation Review*, **42**(1):15–27, 2014. 16
- [28] R. DANTU E. BAATARJAV, S. PHITHAKKITNUKON. **Group Recommendation System for Facebook.** *Springer Berlin Heidelberg*. 16
- [29] S. T. ROSENBLOOM B. MALIN Z. YIN, D. FABBRI. **A scalable framework to detect personal health mentions on Twitter.** *Journal of medical Internet research*, **17**(6), 2015. 17
- [30] R. RICHARDSON M. TURNER P. WATSON P. MISSIER G. MEARNES, R. SIMMONDS. **Tweet My Street: A Cross-Disciplinary Collaboration for the Analysis of Local Twitter Data, Recommendation System for Facebook.** *Future Internet*, **6**(2):378–396, 2014. 17
- [31] G. SEMERARO F. NARDUCCI C. MUSTO P. LOPS, M. GEMMIS. **Leveraging the linkedin social network data for extracting content-based user profiles.** *In Proceedings of the fifth ACM conference on Recommender systems*. 17
- [32] IMDB. **Internet Movie Database.** <http://www.imdb.com>. 18
- [33] J. S. ANDRADE JR. H. A. MAKSE L. K. GALLOS, F. Q. POUTIGUAR. **IMDB Network revisited: unveiling fractal and modular properties from a typical small-world network.** *PLoS One*, **8**(6), 2013. 18

## BIBLIOGRAFÍA

---

- [34] M. TSAGKIAS M. DE RIJKE A. OGHINA, M. BREUSS. **Predicting IMDB Movie Ratings Using Social Media**. *Advances in Information Retrieval*, **7224**. 18
- [35] D. HIEMSTRA M. NESHATI H. BEIGY, E. ASGARI. **Integration of scientific and social networks**. *World Wide Web 09/2014*, **17**(5):1051–1079, 2014. 18
- [36] DBLP. **Digital Bibliography & Library Project**. <http://dblp.uni-trier.de>. 18
- [37] T. WENINGER J. HAN Z. YIN, M. GUPTA. **A unified framework for link recommendation using random walks**. *2010 International Conference on Advances in Social Networks Analysis and Mining*. 19
- [38] J. YORK G. LINDEN, B. SMITH. **Amazon.com Recommendations: Item-to-Item Collaborative Filtering**. *IEEE Internet Computing*. 20
- [39] J. LIU P. NANDY T. VAN VLEET U. GARGI D. SAMPATH J. DAVIDSON, B. LIEBALD. **The YouTube video recommendation system**. *In Proceedings of the fourth ACM conference on Recommender systems*. 20
- [40] D. LEE E. ELMACIOGLU. **On six degrees of separation in DBLP-DB and more**. *ACM SIGMOD Record*, **34**(2), 2005. 21
- [41] H. HUSNA. **Six degrees of separation? More like 3.57, says Facebook**. *The Christian Science Monitor*. 21
- [42] T. SUZUMURA M. WATANABE. **How social network is evolving? A preliminary study on billion-scale twitter network**. *WWW12 Companion Proceedings of the 22nd International Conference on World Wide Web*. 22
- [43] FACEBOOK. **Graph API**. <https://developers.facebook.com/>. 26, 29
- [44] INSTAGRAM. **Instagram API**. <https://www.instagram.com/developer/>. 26
- [45] TWITTER. **REST APIs**. <https://dev.twitter.com/rest/public>. 26
- [46] PINTEREST. **Pinterest API**. <https://developers.pinterest.com/>. 26
- [47] LINKEDIN. **Linkedin API**. <https://developer.linkedin.com/>. 26, 32
- [48] GOOGLE PLUS. **Google+ API**. <https://developers.google.com/+/web/api/rest/>. 27
- [49] GOOGLE BOOKS. **Google Books API**. <https://developers.google.com/books/>. 50, 62
- [50] GOOD READS. **Good Reads API**. <https://www.goodreads.com/api>. 62
- [51] AMAZON WEB SERVICES. **Amazon Web Services**. <https://aws.amazon.com/es/blogs/aws>. 62
- [52] SCALA. **Scala**. <http://docs.scala-lang.org/index.html>. 65
- [53] PLAY. **Play Framework**. <https://www.playframework.com/>. 65
- [54] TWITTER4J. **Twitter4j**. <http://twitter4j.org/en/index.html>. 65
- [55] FACEBOOK4J. **Facebook4j**. <http://facebook4j.github.io/en/index.html>. 66
- [56] JAVA LIBRARY FOR INTERACTING WITH THE ISBNDB REST ENDPOINTS. **Java Library for interacting with the isbndb REST endpoints**. <https://github.com/scholnicks/isbndb>. 67