



UNIVERSIDAD DE CONCEPCIÓN  
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS

# APLICACIÓN DE SISTEMAS DE RECOMENDACIÓN A JUEGOS DE MESA MODERNOS

**Por: Daniel Eduardo Suazo Pavez**

Tesis presentada a la Facultad de Ciencias Físicas y Matemáticas de la  
Universidad de Concepción para optar al título profesional de Ingeniero  
Civil Matemático

Marzo 2024  
Concepción, Chile

**Profesor Guía: Sebastián Niklitschek-Soto**

© 2024, Daniel Eduardo Suazo Pavez

Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica del documento

A Carolina, el amor de mi vida. Y también a Obi y Tuki, los otros amores de mi vida.

## AGRADECIMIENTOS

Quiero comenzar dando las gracias a Sebastián, como mentor y amigo, has sido siempre una figura que me inspira a seguir adelante y superarme cada día.

Agradecer también a todos los profesores que fueron parte de mi proceso formativo, por su tiempo, buena disposición y las oportunidades que me han ofrecido a lo largo de mis años como estudiante universitario.

Quiero agradecer también a mis padres Manuel y Dina, por llevarme de la mano en cada etapa de mi formación y a mis hermanos Manuel y Catalina por su apoyo incondicional en los momentos importante de la vida.

Gracias también a todos los voluntarios que ayudaron en la evaluación de los métodos de recomendación, por su tiempo y buena voluntad.

Finalmente, pero no menos importante, a mi esposa Carolina, compañera de vida, por ser esa persona que siempre encuentra la palabra de aliento adecuada para lograr levantarme de nuevo.

A todos, muchas gracias.

## Resumen

En la actualidad, existe una sobrecarga de información y opciones, tanto en entornos digitales, como en redes sociales, como también a la hora de utilizar nuestro tiempo libre y elegir un hobby al cual dedicarle tiempo. En la búsqueda de aliviar esta carga o intentar darle solución a este problema, aprovechando las capacidades crecientes de la computación de las últimas décadas, y las posibilidades de captura de información mediante el uso constante de dispositivos electrónicos, se han creado diversas técnicas de recomendación apuntadas a satisfacer de mejor manera las necesidades de las personas, primero, en su papel de usuarios, facilitando la toma de decisiones e intentando maximizar su satisfacción, como de los proveedores de servicios que buscan maximizar su retorno, al generar lealtad o mantener a las mismas personas, ahora, como clientes. El abanico de industrias donde pueden ser utilizados sistemas de recomendación abarca prácticamente todo el espectro de actividades humanas donde los usuarios deben tomar decisiones sobre productos a consumir, desde el comercio electrónico, las redes sociales, el mercado del entretenimiento u otros hobbies, como lo es el de los juegos de mesa modernos. Con esto, el objetivo de la presente memoria de título consiste en explorar y evaluar la aplicación de sistemas de recomendación, utilizando tanto métodos basados en el filtrado colaborativo como basados en contenido, además de proponer una estrategia híbrida para abordar las recomendaciones a usuarios interesados en conocer nuevos títulos, en el mundo de los juegos de mesa modernos.

**Keywords** – recommendation systems, collaborative filtering, content-based, word embedding, board games

# Índice general

<b>AGRADECIMIENTOS</b>	<b>i</b>
<b>Resumen</b>	<b>ii</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Objetivos . . . . .	2
1.1.1. Objetivo General . . . . .	2
1.1.2. Objetivos Específicos . . . . .	2
<b>2. Planteamiento del Problema</b>	<b>3</b>
2.1. Juegos de mesa . . . . .	3
2.2. Juegos de mesa modernos . . . . .	4
2.2.1. El mercado de los juegos de mesa modernos . . . . .	5
2.2.2. Boardgamegeek (BGG) . . . . .	7
2.2.3. Recomendaciones en el contexto de los Juegos de Mesa Modernos . . . . .	9
<b>3. Marco Teórico</b>	<b>11</b>
3.1. Conceptos Preliminares . . . . .	11
3.1.1. Sistemas de recomendación . . . . .	11
3.1.1.1. Sistemas basados en Filtrado Colaborativo . . . . .	12
3.1.1.2. Sistemas basados en Contenido . . . . .	13
3.1.1.3. Recomendadores Híbridos . . . . .	14
3.1.1.4. Objetivos de los Sistemas de Recomendación . . . . .	15
3.1.1.5. Principales problemas al aplicar Sistemas de Recomendación . . . . .	17
3.1.2. Juegos de Mesa Modernos . . . . .	19
3.1.3. Métricas de Evaluación . . . . .	22
<b>4. Metodología</b>	<b>25</b>
4.1. Obtención y Procesamiento de los datos . . . . .	25
4.1.1. Boardgamegeek API . . . . .	25
4.1.1.1. Definición de API . . . . .	25
4.1.1.2. BGG XML API2 . . . . .	26
4.1.2. Procesamiento y Almacenamiento de datos . . . . .	32

4.1.2.1.	Procesamiento de los datos de juegos . . . . .	32
4.1.2.1.1.	Número ideal de jugadores: . . . . .	34
4.1.2.2.	Procesamiento de los datos de ratings . . . . .	36
4.1.2.3.	Almacenamiento de los datos . . . . .	36
4.2.	Sistemas de Recomendación . . . . .	39
4.2.1.	Método basado en Filtrado Colaborativo . . . . .	40
4.2.1.1.	Entradas globales . . . . .	40
4.2.1.2.	Generación de Recomendaciones . . . . .	41
4.2.2.	Método basado en Contenido . . . . .	43
4.2.2.1.	Entradas globales . . . . .	45
4.2.2.2.	Generación de Recomendaciones . . . . .	47
4.3.	Implementación . . . . .	47
<b>5.</b>	<b>Análisis</b>	<b>49</b>
<b>6.</b>	<b>Resultados y Discusión</b>	<b>52</b>
6.1.	Definición de los Experimentos . . . . .	52
6.1.1.	Experimento 01: Conjunto Test BGG . . . . .	53
6.1.2.	Experimento 02: Encuesta Expertos . . . . .	54
6.2.	Resultados . . . . .	55
6.2.1.	Experimento 01 . . . . .	55
6.2.2.	Experimento 02 . . . . .	57
6.3.	Costo Computacional . . . . .	58
<b>7.</b>	<b>Conclusión</b>	<b>60</b>
	<b>Referencias</b>	<b>63</b>
	<b>Apéndices</b>	<b>66</b>
<b>A.</b>		<b>66</b>
A1.	Ejemplos de llamadas a API BGG . . . . .	66
A1.1.	Consulta Datos Juego . . . . .	66
A1.2.	Consulta Datos Rating . . . . .	70
A2.	Ejemplo de Almacenamiento intermedio - Datos Juego . . . . .	78

# Índice de cuadros

2.2.1. Guía para asignar ratings, desde BGG. . . . .	9
4.1.1. Resumen de parámetros modificables en solicitudes a la BGG API	27
4.1.2. Ejemplo de entrada Algoritmo 1 . . . . .	31
4.1.3. Información base de un juego, desde API BGG . . . . .	33
4.1.4. Información en Tags de juegos, desde API BGG. . . . .	34
4.1.5. Resultados encuesta número de jugadores, juego Indonesia . . . . .	35
4.1.6. Cálculo de mejores números de jugadores, juego Indonesia. . . . .	36
4.1.7. Ejemplo de ratings extraídos y tabulados, desde API BGG . . . . .	36
4.3.1. Características de la Máquina y librerías de Python utilizadas. . . . .	48
5.0.1. Números Resumen Extracción datos desde BGG . . . . .	50
6.1.1. Ejemplo de Formulario Experimento 02 . . . . .	54
6.2.1. Resultados Experimento 01 . . . . .	56
6.2.2. Resultados Experimento 02 . . . . .	57
6.3.1. Resumen Costo Computacional . . . . .	59

# Índice de figuras

2.2.1.Juegos Publicados entre 1980 y 2022. . . . .	6
2.2.2.Pantalla de inicio del sitio boardgamegeek.com. . . . .	7
2.2.3.Extracto de la página dedicada a “Roads & Boats”. . . . .	8
2.2.4.Extracto del ranking actual de juegos en la BGG. . . . .	8
3.1.1.Taxonomía de Sistemas de Recomendación Híbridos . . . . .	16
4.1.1.Esquema Base de datos Inicial . . . . .	38
4.1.2.Esquema Base de datos con Post-Proceso de los datos . . . . .	39
5.0.1.Revisión Ratings por Juego, datos desde BGG . . . . .	50
5.0.2.Revisión Ratings por Usuario, datos desde BGG . . . . .	51
6.2.1.Resultados Experimento 01 - Métrica 01 . . . . .	56
6.2.2.Resultados Experimento 01 - Métrica 02 . . . . .	56
6.2.3.Resultados Experimento 02 . . . . .	57

# Capítulo 1

## Introducción

En la actualidad, se tiene acceso a una enorme cantidad de información gracias a la evolución de la tecnología y la presencia de Internet. Con el surgimiento de plataformas como Netflix y Spotify, que nos permiten disfrutar de una amplia variedad de productos y entretenimiento basados en el reconocimiento de nuestros gustos y preferencias, éstas mismas plataformas han crecido de la mano de herramientas más sofisticadas de recomendación.

La primera implementación de sistemas de recomendación a gran escala, fue desarrollada por Amazon en la década de los 90 ([Smith and Linden, 2017](#)), a partir de ese momento, los sistemas de recomendación se han vuelto una herramienta cada vez más común en las plataformas digitales. Su popularidad se debe, en gran medida, a su capacidad para ayudar a los usuarios a encontrar productos y servicios que satisfagan sus necesidades y gustos personales.

Los sistemas de recomendación están basados principalmente en las evaluaciones o interacciones de los usuarios con los productos que consumen. Por otro lado, pueden además incluir información adicional, tanto de los usuarios como de los ítems en cuestión.

Existe una amplitud de usos de recomendadores en la vida diaria. Un ejemplo cotidiano se puede encontrar en las redes sociales (Instagram, TikTok), distribuidores de contenido (Youtube, Netflix) o tiendas en línea (Amazon, Falabella). Sin embargo, aún existen áreas en las que no se han integrado un completamente a los sistemas.

Así ocurre con el hobby de los juegos de mesa modernos, un mercado en crecimiento donde no se ha explorado profundamente la generación de recomendaciones personalizadas a los usuarios, dejando la decisión de miles de compras a la experiencia de los vendedores.

Esta problemática surge en un sector donde las recomendaciones vienen principalmente del boca a boca o de buscar información de forma personal. Esto puede funcionar para un grupo de entusiastas del hobby, pero al recibir recomendaciones de, por ejemplo, vendedores de una tienda especializada, se es muy propenso a sesgos personales o comerciales, pudiendo generar insatisfacción al usuario en su búsqueda o derivar en una elección inadecuada.

Dado que los algoritmos de recomendación conceptualmente son desarrollados para dar solución a los problemas antes mencionados a una gran escala, el objetivo de esta tesis se centra en proponer un algoritmo de recomendación en el contexto de juegos de mesa modernos, además de proponer una estrategia de evaluación para los resultados obtenidos basada en la recopilación de datos adicionales con entusiastas del hobby.

## **1.1. Objetivos**

### **1.1.1. Objetivo General**

Comparar los resultados de aplicar diferentes métodos de recomendación para juegos de mesa modernos.

### **1.1.2. Objetivos Específicos**

- Generar un dataset de productos, usuarios y de evaluaciones de éstos últimos a los productos, en el contexto de los juegos de mesa modernos.
- Implementar y evaluar técnicas de recomendación basados solo en calificaciones hechas por los usuarios.
- Implementar y evaluar técnicas de recomendación basadas en contenido.
- Comparar los resultados de aplicar ambas estrategias de recomendación.

## Capítulo 2

# Planteamiento del Problema

### 2.1. Juegos de mesa

Juego de mesa es la clasificación que se da a los juegos que usualmente cuentan con un tablero y fichas, que pueden ser jugados sobre una mesa o superficie plana. Estos usualmente están pensados para múltiples participantes o jugadores, pero existen además algunos para practicar en solitario. Se han encontrado piezas que se consideran parte de un juego de mesa, con hasta 5000 años de antigüedad <sup>1</sup> en un yacimiento funerario del asentamiento Başur Höyük, en Turquía. Destacados juegos de mesa son el Ajedrez o el Go, juegos abstractos con mucha historia y practicados a lo largo de todo el mundo. Más reciente es la creación del juego Monopoly <sup>2</sup> en el 1935, una reimplementación de The Landlord's Game <sup>3</sup>, juego que originalmente tenía como fin el demostrar los efectos de la especulación (en este caso, inmobiliaria) y sus consecuencias, pero que en Monopoly toma la forma de una lucha entre magnates inmobiliarios por el dominio de una ciudad. Otros juegos destacables pueden ser Risk <sup>4</sup> (juego de diplomacia y guerras), o juegos de cartas populares en Chile, basados tanto en una baraja de naipe inglés (como Carioca <sup>5</sup>), o una baraja de naipe español, como la Escoba <sup>6</sup>.

Una breve línea de tiempo con eventos relevantes del último siglo relativos al tema

<sup>1</sup><https://www.nbcnews.com/sciencemain/oldest-known-gaming-tokens-dug-bronze-age-turkish-graves-6c10920354>

<sup>2</sup><https://boardgamegeek.com/boardgame/1406/monopoly>

<sup>3</sup><https://boardgamegeek.com/boardgame/29316/the-landlords-game>

<sup>4</sup><https://boardgamegeek.com/boardgame/181/risk>

<sup>5</sup>[https://es.wikipedia.org/wiki/Carioca\\_\(juego\)](https://es.wikipedia.org/wiki/Carioca_(juego))

<sup>6</sup>[https://es.wikipedia.org/wiki/Escoba\\_del\\_15](https://es.wikipedia.org/wiki/Escoba_del_15)

puede encontrarse en el siguiente [enlace](#) o en el siguiente [artículo](#).

## 2.2. Juegos de mesa modernos

¿Qué diferencia a los juegos de mesa modernos de los mencionados anteriormente? A pesar de que pueden existir diferentes opiniones al respecto, puntos en común que se pueden encontrar son los siguientes:

- **Variedad en las mecánicas:** Juegos como **Snakes and Ladders** o **Ludo** se basaban principalmente en la mecánica de *Roll & Move* (lanzar un dado y moverse una cantidad de casillas), ahora es posible encontrar un abanico muy amplio de estas, donde algunas destacables pueden ser: *Worker Placement*, *Area Control*, *Deck Building*, *Drafting*, entre muchas otras. Más detalle sobre estos términos en la sección 3.1.2. Esto permite que la actividad de jugar pueda ser mucho más variada y alcanzar a un público más amplio.
- **Multitud de Temáticas:** Debido a la cantidad de juegos publicados cada año, es posible encontrarse juegos ambientados en Ciencia Ficción, Fantasía, Carreras, Apocalipsis Zombie, incluso construir y hacer crecer tu propia cadeana de comida rápida (como es el caso de **Food Chain Magnate**<sup>7</sup>). La oferta es tremendamente extensa y tanto editoriales como diseñadores buscan ideas nuevas para implementar todo el tiempo.
- **No eliminación de jugadores:** Juegos como **Monopoly** o **Risk** presentan la posibilidad de que un jugador quede eliminado de la partida de forma temprana (antes del final de ésta). Ésta situación puede generar una mala experiencia para el jugador eliminado. Actualmente es posible encontrar juegos que incorporan esto pero la gran mayoría de títulos publicados la evitan e incluso presentan mecanismos para mantener a los jugadores dentro de la partida (es decir, dentro de la contienda por la victoria en juegos competitivos) de forma activa. Por ejemplo, en **Power Grid**<sup>8</sup>, un juego que trata sobre el mercado y la distribución de energía eléctrica, los jugadores con menos alcance en el mapa (lo que conlleva menos ingresos), tienen acceso prioritario a la compra de los recursos para alimentar sus plantas de energía, a precios menores.

---

<sup>7</sup><https://boardgamegeek.com/boardgame/175914/food-chain-magnate>

<sup>8</sup><https://boardgamegeek.com/boardgame/2651/power-grid>

- **Juegos "de diseñador"**: Tanto en jugabilidad, como en calidad y cantidad de componentes, hay títulos para todos los gustos, existen nombres reconocidos que en la práctica representan una marca en si misma, cada uno apuntando a un público objetivo diferente. Para juegos grandes con mucha producción y miniaturas, puede mencionarse al autor **Eric Lang**<sup>9</sup>, con juegos como **Blood Rage**<sup>10</sup> o **Rising Sun**<sup>11</sup>. Al otro lado del espectro de diseño, puede mencionarse a **Uwe Rosenberg**<sup>12</sup> como un gran representante de los juegos de gestión de recursos, con un estilo de diseño preponderantemente iterativo, revisitando diseños previos y construyendo de esta manera experiencias muy distintas de juegos. De este diseñador se pueden destacar **A Feast for Odin**<sup>13</sup> y **Agricola**<sup>14</sup>, manteniéndose en el número 1 del ranking global de la BGG desde el 18 de agosto de 2008 al 28 de febrero del 2010<sup>15</sup>.

### 2.2.1. El mercado de los juegos de mesa modernos

El mercado de los juegos de mesa modernos ha presentado un notable crecimiento en los últimos años (ver figura 2.2.1), esto puede deberse a múltiples factores, en los que se encuentran: la influencia de grandes títulos que han llamado la atención de un público más amplio, como Catan (publicado el año 1995) o Carcassonne (publicado el año 1998), la proliferación de comunidades en línea para el hobby, el aumento en plataformas online para jugar una creciente cantidad de títulos y la aparición y el crecimiento del sitio **boardgamegeek** (en adelante, abreviado **BGG**), donde las personas pueden buscar y acceder a información de los títulos disponibles. Esta página web almacena la mayor cantidad de información disponible online sobre juegos de mesa en el mundo y funciona como una guía tanto para jugadores novatos como experimentados. Otros posibles motivos más recientes del crecimiento de este mercado se pueden encontrar en el aumento de canales de difusión del hobby y contenido especializado dedicado (reseñas, *podcasts* y tutoriales en Youtube, reseñas cortas y fotos en Instagram, *podcast* en Spotify y

<sup>9</sup><https://boardgamegeek.com/boardgamedesigner/1533/eric-m-lang>

<sup>10</sup><https://boardgamegeek.com/boardgame/170216/blood-rage>

<sup>11</sup><https://boardgamegeek.com/boardgame/205896/rising-sun>

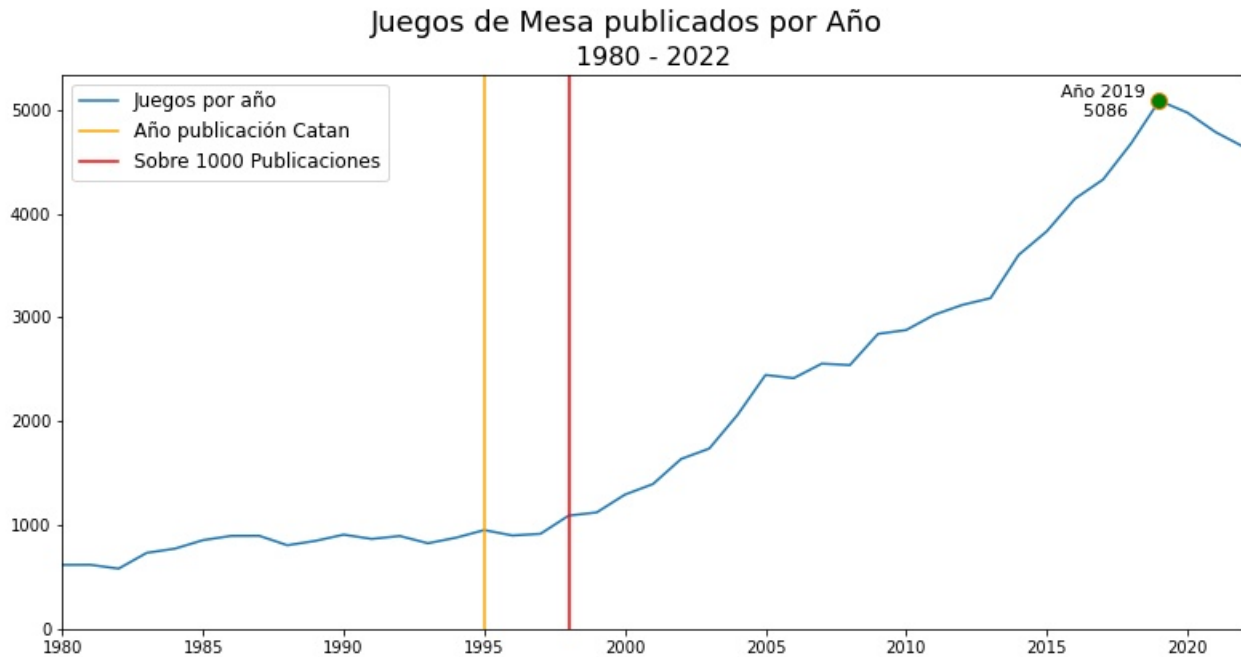
<sup>12</sup>

<sup>13</sup><https://boardgamegeek.com/boardgame/177736/a-feast-for-odin>

<sup>14</sup><https://boardgamegeek.com/boardgame/31260/agricola>

<sup>15</sup><https://boardgamegeek.com/geeklist/256388/bgg-number-ones-history-definitive-ongoing-list-wi>

otras plataformas), el explosivo fenómeno del *Crowdfunding* durante la última década <sup>16</sup>. Por último, mencionar que a nivel local existen más de 60 tiendas especializadas <sup>17</sup>, lo que permite alcanzar a un público cada vez más amplio.



**Figura 2.2.1:** Juegos Publicados entre 1980 y 2022.

Como se mencionó previamente puede verse en la figura 2.2.1, la evolución del hobby en términos de juegos nuevos por año cambió notoriamente alrededor del año de publicación de Catan, siendo que antes se mantuvo constante bajo los 1000 títulos por año por al menos 15 años, en menos de 10 ya alcanzaba los 2500 (año 2005), alcanzando su *peak* en el año 2019, con más de 5000 juegos publicados. Aunque es posible que tanto la crisis de COVID-19 como la *Shipping Crisis*<sup>18</sup> post-pandemia hayan afectado la dinámica vista pre-2019 en términos de crecimiento, sigue siendo notable el tamaño que ha alcanzado este mercado, que se presupuesta alcance los \$USD 14b (14 billones o 14.000 millones de dolares)<sup>19</sup> en 2024.

<sup>16</sup><https://www.polygon.com/tabletop-games/23552590/kickstarter-top-10-largest-most-backed-board-games-campaigns-2022>

<sup>17</sup><https://juegos.tabula.cl>

<sup>18</sup><https://time.com/6096497/board-games-shipping-crisis/>

<sup>19</sup><https://wordsrated.com/board-game-market-statistics/>

### 2.2.2. Boardgamegeek (BGG)

El sitio web Boardgamegeek (abreviado como BGG) ya se ha mencionado antes como un lugar donde encontrar información relativa a juegos de mesa. A continuación, una descripción más detallada, especificando su relevancia para el trabajo descrito en este documento. Boardgamegeek es un foro online fundado en el año 2000 por Scott Alden y Derk Solko, dedicado para los amantes de los juegos de mesa, que contiene reseñas, imágenes, y todo tipo de contenido relativo al hobby. Además, cuenta con una base de datos de acceso público con más de 125.000 títulos distintos.

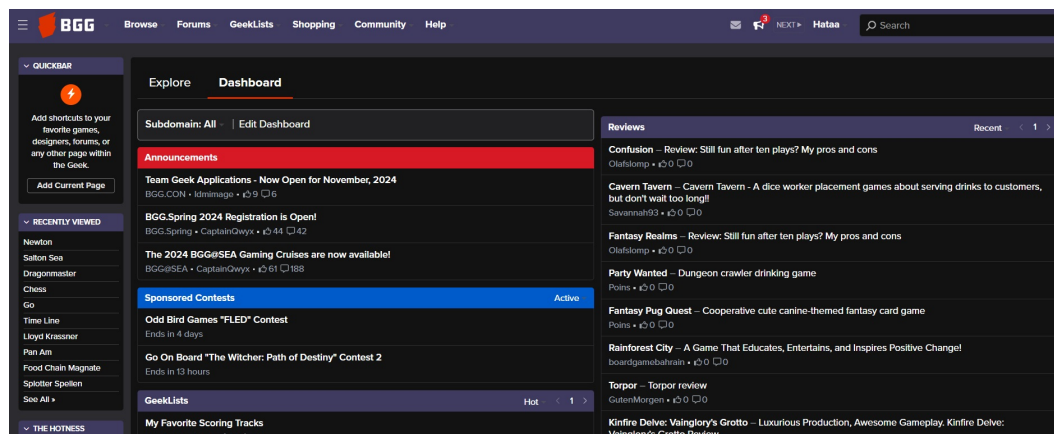


Figura 2.2.2: Pantalla de inicio del sitio boardgamegeek.com.

En este sitio, podemos encontrar una página dedicada a cada juego ingresado en su base de datos, donde podremos encontrar su descripción, información, reviews/críticas, un sub-foro con hilos de preguntas o comentarios iniciados por los usuarios, entre otras cosas.

**8.3** Roads & Boats: 20th Anniversary Edition (2019)  
Give geese paper and see what they discover. 🗨️  
487 Ratings & 155 Comments - GeekBuddy Analysis

1-6 Players  
Community: 1-4 — Best: 2

120–240 Min  
Playing Time

Age: 14+  
Community: 10+

Weight: 4.27 / 5  
\*Complexity\* Rating

Designer: Jeroen Doumen, Joris Wiersinga  
Artist: Tamara Jannink  
Publisher: Splotter Spellen  
See Full Credits

My rating - Edit ★★★★★★ 10

Buy a Copy In Collection (Own) Log Play 16 60 Subscribe

Overview Ratings Forums Images Videos Files Stats Versions Expansions My Games Shopping More

**Description**

For the 20th anniversary of *Roads & Boats*, the game is being reprinted for the fifth time. This edition will include the *&Cetera* expansion, so the game can be played with up to six players and you will get all the additional modules of *&Cetera* included in the game. There are no other changes to the game.

In *Roads & Boats*, each player builds a civilization over a long period of time, as in many other games. Unlike most games, however, the emphasis is not on warfare, population growth, city or statebuilding, but on logistics, or rather on transport. Each player starts the game with three donkeys, a pile of wooden boards, a number of stones, and two geese. With these few resources, you try to build such diverse things as woodcutters, roads, boats, mines, a stock exchange... but beware! There is no concept of territory in this game! You cannot own land, nor buildings, so the things you build can be used by any other player...  
The original game can be played by two to four players, but it can also be used as a one-person puzzle.

CLASSIFICATION  
Type: Uncategorized  
Category: City Building, Civilization, Economic

Figura 2.2.3: Extracto de la página dedicada a “Roads & Boats”.

Además, dentro de las funcionalidades más utilizadas es el de agrupar las valoraciones realizadas por todos los usuarios y ordenar las entradas por el rating final calculado, generándose un ranking con el top de los juegos mejor evaluados por la comunidad.

Board Game Rank	Title	Your Rating	Geek Rating	Avg Rating	Num Voters	Status	Your Plays	Shop
1	<b>Brass: Birmingham</b> (2018) Build networks, grow industries, and navigate the world of the Industrial Revolution.	8.5 Feb 2023	8.417	8.60	43948		Plays: 5	List: \$79.99 Amazon: \$69.95
2	<b>Pandemic: Legacy: Season 1</b> (2015) Mutating diseases are spreading around the world - can your team save humanity?	N/A	8.384	8.53	52795			Amazon: \$71.99
3	<b>Gloomhaven</b> (2017) Vanquish monsters with strategic cardplay. Fulfill your quest to leave your legacy!	N/A	8.364	8.60	61326			
4	<b>Ark Nova</b> (2021) Plan and build a modern, scientifically managed zoo to support conservation projects.	N/A	8.333	8.54	40615		Plays: 1	List: \$74.95 Amazon: \$55.00
5	<b>Twilight Imperium: Fourth Edition</b> (2017) Build an intergalactic empire through trade, research, conquest and grand politics.	N/A	8.243	8.61	23066		Plays: 3	Amazon: \$131.99

Figura 2.2.4: Extracto del ranking actual de juegos en la BGG.

Cabe mencionar que los valores contenidos en su escala de ratings se encuentran entre el 1 y el 10 e incluyen hasta tres decimales, correspondiendo el 1 a la calificación más baja o negativa y el 10 a la más alta o positiva. Entendiendo que estas notas o evaluaciones son subjetivas, la BGG entrega a la comunidad una guía a la hora de ingresar evaluaciones, resumida en la siguiente tabla<sup>20</sup>:

<sup>20</sup><https://boardgamegeek.com/wiki/page/ratings>

<b>Nota</b>	<b>Texto Original</b>	<b>Traducción</b>
10	Outstanding. Always want to play and expect this will never change.	Sobresaliente. Siempre dispuesto a jugarlo y espero que esto nunca cambie.
9	Excellent Game. Always want to play it.	Juego Excelente. Siempre con ganas de jugarlo.
8	Very good game. I like to play. Probably I'll suggest it and will never turn down a game.	Muy buen juego. Disfruto jugarlo. Probablemente lo recomendaría y nunca rechazaría una partida.
7	Good game, usually willing to play	Buen juego, normalmente aceptaría una partida
6	Ok game, some fun or challenge at least, will play sporadically if in the right mood.	Juego OK, algo divertido o desafiante, lo jugaría esporádicamente si estoy de buen humor.
5	Average game, slightly boring, take it or leave it.	Juego promedio, un poco aburrido, tómalo o déjalo.
4	Not so good, it doesn't get me but could be talked into it on occasion.	No muy bueno, no me gusta pero puedo ser convencido de jugarlo ocasionalmente.
3	Likely won't play this again although could be convinced, Bad.	Probablemente nunca lo jugaría de nuevo, eventualmente puedo ser convencido de lo contrario. Malo.
2	Extremely annoying game, won't play this ever again.	Extremadamente molesto, espero no volver a jugarlo de nuevo
1	Defies description of a game. You won't catch me dead playing this. Clearly broken.	Desafía la descripción de Juego. No me pillarán muerto jugando a esto. Definitivamente roto.

**Cuadro 2.2.1:** Guía para asignar ratings, desde BGG.

### 2.2.3. Recomendaciones en el contexto de los Juegos de Mesa Modernos

Tanto para jugadores nuevos buscando introducirse en este mundo como para jugadores experimentados buscando nuevos títulos para jugar o adquirir, la oferta de productos disponibles puede ser abrumadora. Como se mencionó anteriormente, en Chile existen alrededor de 60 tiendas especializadas dedicadas al hobby, con una oferta de más de 1.000 productos distintos a disposición de los entusiastas (según la información recopilada por el sitio <https://juegos.tabula.cl>).

Para resolver esta situación, una posibilidad es la de visitar o ponerse en contacto con una tienda en particular y buscar recomendaciones de los vendedores de ésta.

En la misma línea, otra alternativa, y una de las prácticas más comunes en el hobby, son las recomendaciones de boca en boca, es decir, solicitar ayuda de amigos o conocidos con más experiencia a la hora de elegir el próximo juego a conseguir. Una tercera opción es la de informarse de forma individual, accediendo a reviews online de otros jugadores o personalidades del mundo lúdico en redes sociales (Youtube, Instagram, Boardgamegeek, etc.), buscando refinar la búsqueda. Éste último método puede llegar a ser extremadamente laborioso y consumir gran cantidad de tiempo.

Sin embargo, todas las opciones mencionadas anteriormente pueden sufrir de problemas de sesgo o favoritismo de parte de las personas entregando la recomendación, ya que estas consideran su experiencia y están limitadas a su exposición a los juegos (es decir, consideran un subconjunto acotado del espectro disponible), además de sus preferencias particulares, las que no necesariamente pueden ser compartidas por el que busca la recomendación, existiendo la posibilidad de generar insatisfacción en el proceso. Una forma de mitigar estos problemas es ampliar la escala en la que son utilizados estos mecanismos, lo que en general, da lugar a un fenómeno del tipo “Sabiduría de la multitud” (Galton, 1907), (Yi et al., 2012), esta suerte de eliminación de sesgo al agregar diversas opiniones, es lo que ha estado a la base de la inspiración de los algoritmos de recomendación.

# Capítulo 3

## Marco Teórico

### 3.1. Conceptos Preliminares

Los sistemas de recomendación surgen como una posible solución a la difícil tarea que supone el tomar una decisión sobre el consumo o adquisición de productos en la era del Internet. Éstos ayudan a los usuarios a encontrar productos o experiencias que se adecuen a sus preferencias y necesidades (Resnick and Varian, 1997).

Los sistemas de recomendación se encuentran ampliamente implantados en las plataformas de consumo de las que se ha hablado anteriormente, debido a su demostrada utilidad (Kumar et al., 2021).

Ante esto, a lo largo del capítulo 3 se abordarán los diferentes sistemas de recomendación, los beneficios que conlleva el uso de estos, métricas utilizadas para su evaluación, entre otros conceptos importantes.

#### 3.1.1. Sistemas de recomendación

Un sistema de recomendación es una herramienta inteligente que personaliza la recomendación de productos a un usuario en particular Jannach et al. (2010). Su objetivo principal es encontrar las relaciones o dependencias que existen entre las interacciones de los usuarios y los productos. Esta relación se puede representar a través de una función que toma como entrada el usuario y los productos y devuelve una recomendación personalizada (Zafarani et al., 2014). Al encontrar estas relaciones, los sistemas de recomendación pueden hacer predicciones sobre

cómo los usuarios valorarán nuevos productos o qué productos serán consumidos por qué usuarios en el futuro (Bobadilla et al., 2013), (Varga, 2019).

Son múltiples los sistemas de recomendación existentes. Todos ellos guían a los usuarios hacia productos que pueden ser de su interés. Es así como los diferentes sistemas de recomendación mejoran la experiencia del usuario y le ayudan a seleccionar de entre la gran cantidad de opciones. A continuación se enumeran los tipos de sistemas de recomendación más relevantes en la literatura (Ricci et al., 2015), (Burke, 2002).

#### 3.1.1.1. Sistemas basados en Filtrado Colaborativo

Los sistemas basados en el filtrado colaborativo basan su funcionamiento en las interacciones de usuarios y productos, buscando relaciones y filtrando desde el uso de información de gran cantidad de usuarios. Existen dos tipos de recomendaciones basados en el filtrado colaborativo: los métodos basados en memoria y en modelos.

- **Los métodos basados en memoria o basados en vecinos** tienen en cuenta las relaciones entre los productos y/o entre los usuarios, es por esto que pueden subdividirse tanto en métodos basados en productos como métodos basados en usuarios.
  - **Métodos basados en productos**, estos predicen el rating de un producto o si éste es relevante para un usuario objetivo, teniendo en cuenta la valorización de productos vecinos a este.
  - **Métodos basados en usuarios**, predicen el rating de un producto o la relevancia de estos para un usuario objetivo, definiendo usuarios similares (**vecinos**) y tomando en cuenta las valorizaciones de éstos a los posibles productos a recomendar.
- **Métodos basados en modelos**, Aquí se engloban métodos que utilicen técnicas de aprendizaje automático (*machine learning*) y minería de datos (*data mining*) como modelos predictivos. Estos son más complejos de implementar, sin embargo, pueden conseguir excelentes resultados (Kumar et al., 2021).

### 3.1.1.2. Sistemas basados en Contenido

Los sistemas de recomendación basados en contenido toman en cuenta principalmente los atributos de los productos, además de incorporar los perfiles de los usuarios. Esto se realiza tomando como entradas las descripciones, palabras claves o etiquetas que definen principalmente propiedades de los productos (Ruiz Iniesta, 2014).

En el caso de perfilar a los usuarios, es necesario tener información relevante de éstos con respecto a sus preferencias, recogidas de interacciones previas, las cuales pueden ser anexadas a las propiedades de los productos (Aggarwal, 2016).

Un ejemplo claro son los recomendadores utilizados en plataformas de streaming como **Netflix o Spotify**, las cuales, a través de la búsqueda de la propiedad de un producto como lo puede ser una categoría del estilo “comedia, románticos, acción, etc.”, el tiempo que los usuarios visualizan ciertos tipos de contenido, entre otros detalles, pueden guardar las preferencias del usuario, siendo capaces de realizar recomendaciones de productos similares no vistos y que tienen propiedades en común. Una de las principales ventajas de estos sistemas es que es posible que puedan depender únicamente de información (descripción, etiquetas) de los productos, lo que permite una rápida incorporación en una etapa inicial (Park and Tuzhilin, 2008).

Sin embargo, estos sistemas pueden mostrar problemas de rendimiento cuando un usuario es nuevo en el sistema y no existen interacciones previas (Problema denominado *Cold Start*, (Ricci et al., 2015)). Muchas plataformas utilizan un **perfilamiento inicial** para usuarios nuevos, de manera de generar un mínimo de información para guiar sus recomendaciones.

Otra dificultad que pueden surgir de la implementación de estos sistemas, es la falta de diversidad en las recomendaciones.

Según la literatura, al comparar ambos sistemas (basados en filtrado colaborativo contra basados en contenido), los primeros poseen mejores resultados (Isinkaye et al., 2015).

### 3.1.1.3. Recomendadores Híbridos

Los sistemas de recomendación híbridos surgen para resolver los problemas que poseen los recomendadores antes mencionados al ser implementados de forma individual, combinando diversas técnicas para mitigar los inconvenientes posibles (Kumar et al., 2021), (Aggarwal, 2016).

Existen tres posibles diseños de sistemas de recomendación híbridos: *Ensemble Design* o Diseño en Conjunto, *Monolithic Design* o Diseño Monolítico y finalmente *Mixed Systems* o Sistemas Mixtos. Esta clasificación se realiza en función de las entradas utilizadas, las salidas obtenidas y como los métodos de recomendación que actúan dentro del sistema se relacionan (Cano and Morisio, 2017).

***Ensemble design* o diseño en conjunto** Combina recomendadores aplicados sobre una única entrada, para obtener una única salida. Dentro de este tipo de recomendador podemos encontrar dos tipos de estructuras;

- **Diseño paralelo:**
  - **Ponderación (Weighted):** Combina las puntuaciones para un producto de cada método recomendador, definiendo pesos distinto a esta puntuación y del resultado, elegir el producto con la puntuación más alta de esta combinación.
  - **Conmutación (Switching):** Emplea distintos sistemas de recomendación según las necesidades del problema a enfrentar, definiendo criterios para aplicar uno u otro dependiendo de las circunstancias.
- **Diseño secuencial:** Como su nombre lo indica, este diseño estructura los diferentes sistemas individuales de forma secuencial, donde en cada etapa el modelo recibe las entradas originales además de las salidas de los pasos anteriores, combinando todo para entregar una salida final (Aggarwal, 2016). Entre las formas posibles de integrar secuencialmente los métodos, se encuentran:
  - **Inclusión de características (Feature Augmentation):** Aquí las salidas de cada etapa se incorporan como características y son integradas a las entradas de las etapas posteriores.

- **Cascada (Cascade):** Para esta técnica se toma directamente las salidas de una etapa como las entradas de la siguiente.

**Diseño Monolítico** A diferencia de los diseños vistos anteriormente, el diseño monolítico utiliza recomendadores sobre diversas entradas con el fin de obtener una sola salida. Aquí encontramos:

- **Combinación de características (Feature Combination):** Se combinan todos los datos utilizados por los distintos métodos recomendadores para generar una única entrada, siendo ésta utilizada por el método recomendador seleccionado.
- **Meta-Level):** En este caso, el modelo utilizado por un sistema se utiliza como entrada de otro. Un ejemplo de esto sería el utilizar un método basado en contenido para encontrar grupos de usuarios similares, para luego, en conjunto con un sistema basado en filtrado colaborativo, generar las predicciones o salidas.

**Diseño Mixto** Aquí, múltiples recomendaciones de diferentes sistemas son presentadas al mismo tiempo. Ya que no existe una combinación de estas salidas, estos métodos no son de tipo *Ensemble*. Usualmente se utilizan cuando el objetivo de la recomendación es una composición de diversos elementos como un conjunto en su totalidad (Aggarwal, 2016).

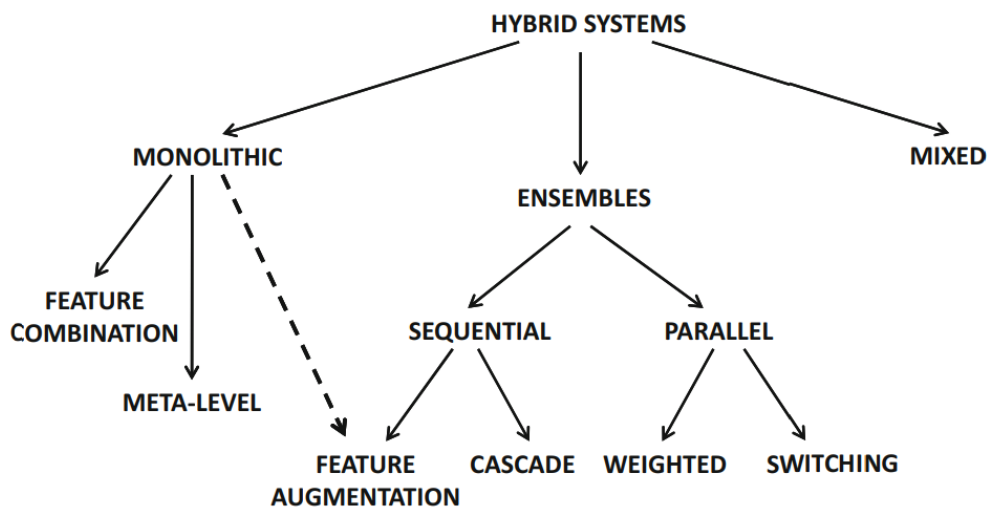
A continuación, se presenta una figura resumen de la taxonomía de los métodos híbridos, presentada en (Aggarwal, 2016):

#### 3.1.1.4. Objetivos de los Sistemas de Recomendación

Existen diferentes definiciones para los objetivos posibles de un sistema de recomendación. En (Ricci et al., 2015), los autores declaran que el objetivo de un sistema de recomendación depende de las distintas perspectivas de las partes involucradas, siendo de particular importancia los usuarios y los proveedores de servicios.

Ejemplos de estos objetivos desde la perspectiva de los proveedores de servicio serían:

- Incrementar la tasa conversión de clientes/usuarios, es decir, aumentar la



**Figura 3.1.1:** Taxonomía de Sistemas de Recomendación Híbridos

tasa de usuarios nuevos o casuales que permanecen como clientes a largo plazo.

- Generar diversidad en las elecciones de los usuarios, para que no se estancuen en elegir siempre los mismos ítems;
- Satisfacer, “enganchar” y entender al usuario. El usuario debe considerar las recomendaciones apropiadas e interesantes. Esto, implícitamente, acerca más y más al usuario al proveedor.

Un buen ejemplo de como ejecutar esto último, puede encontrarse en el servicio que provee **Netflix**: La plataforma va un paso más allá y no solo muestra las recomendaciones, sino que también muestra las razones específicas detrás de la recomendación, generando algo llamado “*Personalization Awareness*” (Amatriain and Basilico, 2012). Los autores de (Pu et al., 2016) sugiere que esto además incrementa la confianza percibida por el usuario.

En (Aggarwal, 2016), se dice que existen principalmente dos formas con las cuales se logra el objetivo de una recomendación:

- El primero es el de predecir el valor del rating para una combinación usuario-ítem.
- El segundo no se expresa en términos de ratings, sino que solo se entrega una lista de top-k ítems. Esta forma es también llamada *Top-K Recommendation Task* o *Top-N Recommendation Task* (Cremonesi et al., 2010). Esto se

logra definiendo una forma de puntuar y ordenar los elementos posibles a recomendar, por ejemplo, con métricas de similitud a elementos previamente evaluados por el usuario (más detalle en la sección 3.1.3).

**Top-K Recommendation Task** En sistemas de recomendación de top-k ítems, un valor adecuado para k debe ser seleccionado. Esto no solo depende de la aplicación en particular, sino que debe considerar la usabilidad. En (Pu et al., 2016) se recomienda listas de largo entre 5 y 20 ítems, esto se resume en la siguiente frase:

“Desplegar más productos y ordenarlos de una forma natural lleva naturalmente a aumentar la sensación de confianza y control de los usuario.”

Sin embargo, simplemente aumentar el tamaño de la lista de recomendaciones puede mejorar el *recall* pero reducirá la precisión de las sugerencias además de ir en contra del objetivo de simplificar el proceso de elección del usuario (Shani and Gunawardana, 2011).

Algunos ejemplos de aplicaciones de este estilo son:

- IMDb.com presenta 12 películas, distribuidas en 2 páginas.
- Amazon.com despliega diferente cantidad de ítems dependiendo de la resolución de pantalla e información del navegador.
- Google.com también despliega una cantidad diferente de anuncios dependiendo de la resolución de pantalla.

Se considera para esta propuesta definir como objetivo de los sistemas de recomendación a trabajar una lista de top-k recomendaciones, dado que, en la práctica dentro del hobby de los Juegos de Mesa Modernos, ésta es la práctica más común a la hora de entregar recomendaciones.

### 3.1.1.5. Principales problemas al aplicar Sistemas de Recomendación

Como ya fue mencionado anteriormente, existen diversos problemas identificados en la literatura que pueden surgir al aplicar las distintas técnicas de recomendación mencionadas. A continuación, un listado de las más comunes y de mayor relevancia para este proyecto (Su and Khoshgoftaar, 2009):

- **Sparsity:** Afecta principalmente al Filtrado Colaborativo, al utilizar las evaluaciones de los usuarios de los diferentes productos. Ésto ocurre cuando la cantidad de productos es mucho mayor que la que un usuario puede consumir y eventualmente evaluar o que pueda generarse un rating asociado a esta relación, generando una matriz de relaciones usuario-producto *sparse* o dispersa, con gran cantidad de ceros en sus coeficientes  $r_{ij}$ , siendo éste la evaluación sobre el producto  $i$  del usuario  $j$ . Ya que para este método, es necesario calcular similitudes entre representaciones vectoriales de ítems o de usuarios, puede ocurrir que disminuya la calidad de las recomendaciones al no existir suficiente solapamiento entre los diferentes elementos disponibles.
- **Cold Start:** Ocurre principalmente con nuevos usuarios y productos nuevos, al no poseer información previa de estos cargada en el sistema.. Este problema es muy común en los sistemas de puntuación de productos. (Lika et al., 2014)
- **Sinonimia:** Ocurre cuando productos similares o iguales presentan nombres diferentes y el sistema los maneja por separado sin asociación.
- **Outside the box:** Este problema aparece al no poder recomendar de buena manera productos adecuados debido a que no se tiene productos suficientemente similares a los puntuados por el usuario, generando una sobre-especialización del modelo.
- **Shilling Attack:** Principalmente presente en sistemas donde los ratings son “públicos”, por ejemplo, evaluaciones de videos en **Youtube** o interacciones en redes sociales en general. Se denomina *Shilling Attack* cuando usuarios o grupos se organizan de forma maliciosa para agregar una cantidad anómala de notas positivas o negativas en algún producto, afectando las recomendaciones basadas en esta información. En el contexto de este proyecto, existen casos de juegos, aún sin salida al mercado, con campañas de financiamiento colectivo en *Kickstarter*, donde personas han evaluado éstos sin existir la posibilidad de haberlo experimentado previamente<sup>1</sup>.
- **Oveja Gris:** Muy común en sistemas de recomendación de filtrado colaborativo. Usuarios o grupos de usuarios con gustos particulares que desentonan con los alcances del perfilamiento implementado. Esto produce

<sup>1</sup><https://boardgamegeek.com/wiki/page/Shilling>

dificultades al entregar buenas recomendaciones para estos usuarios, además de empeorar las recomendaciones entregadas a los demás usuarios del sistema. En el caso de los juegos de mesa modernos, esto puede ocurrir, por ejemplo, desde la perspectiva de los usuarios, cuando sus preferencias se encuentran concentradas en un nicho demasiado específico o pequeño. En este caso podemos mencionar los juegos económicos de la familia [18xx](#), principalmente enfocados en manejar compañías ferroviarias mediante la compra de acciones o *shares*, con el objetivo de maximizar los retornos personales. Son juegos complejos con duración aproximada comenzando usualmente desde las 3 horas (por ejemplo, [18Chesapeake](#), un título introductorio del género), llegando a 9 o más horas para ciertos títulos, como es el caso de [1817](#). Existen comunidades muy involucradas con este tipo de juegos, pero debido a sus características, son limitadas las recomendaciones entregables fuera de la familia a usuarios cuyo foco principal sean este tipo de títulos.

### 3.1.2. Juegos de Mesa Modernos

Los juegos de mesa modernos pueden ser categorizados por multitud de parámetros o términos. En un primer nivel pueden considerarse las **Mecánicas** y la **Temática**, en un segundo nivel podemos considerar otros conceptos, como lo son: Cantidad de jugadores posibles (además del número óptimo de jugadores), Tiempo de juego o duración de una partida, Nivel de Interacción, Complejidad, entre otros.

A continuación, se detallan estos conceptos en mayor detalle:

- **Mecánicas:** Cada juego tiene su propio conjunto de reglas, sistemas e interacciones que controlan como éste es jugado, definiendo las posibles interacciones entre jugador-juego y jugador-jugador. Un breve listado de las más comunes es:
  - a) *Roll/Spin and Move:* Aquí se lanzan dados o algún otro elemento que entregue resultados aleatorios, definiendo cuantos espacios el jugador puede mover sus piezas. Ejemplos: Monopoly (1935), Formula D (2008), Deep Sea Adventure (2014), Camel Up (2018).
  - b) *Tile Placement:* Involucra el posicionamiento de piezas, usualmente de forma adyacente a otras previamente colocadas, para la obtención de puntos de victoria o gatillar acciones, considerando tanto la ubicación

espacial como otras propiedades, como color, tipo o categoría de las piezas, etc. Ejemplos: Carcassonne (2000), Age of Steam (2002), Castles of Burgundy (2011), Cascadia (2021).

- c) *Worker Placement*: También conocida como Posicionamiento de Trabajadores, involucra a los jugadores manejando una cantidad de piezas, en general limitada y usualmente con forma humana (llamados *meebles*<sup>2</sup>), para gatillar acciones de un conjunto disponible para todos los jugadores, usualmente en orden de turno y uno a la vez. Es una especie de “*Draft de Acciones*” (selección de acciones desde un conjunto limitado compartido entre jugadores), donde las piezas utilizadas temáticamente representan trabajadores haciendo las acciones correspondientes. Ejemplos: Agricola (2007), Tzolk’in (2012), Targi (2012), The Gallerist (2015).
  - d) *Auction/Bidding* (Subastas): Esta mecánica involucra pujar o apostar una cantidad de recursos (usualmente dinero del juego), donde los jugadores compiten por obtener un bien o recurso presentado por el juego, en busca de mejorar su posición actual o la posibilidad de mejorar sus acciones futuras. Ejemplos: For Sale (1997), Power Grid (2004), Indonesia (2005), Five Tribes (2014).
- **Temática:** Ambientación elegida para el juego. Se puede categorizar el universo de juegos de mesa modernos principalmente en 3 categorías:
- a) Juegos donde la temática es central a éste (juegos **Temáticos**).
  - b) Juegos donde el tema está presente en favor de las mecánicas, presentado como un medio de comunicar éstas y las reglas, más que como punto central del juego.
  - c) Juegos donde ésta es prácticamente nula o inexistente (juegos Abstractos).

Algunos temas utilizados de forma frecuente son: Ciencia Ficción (viajes al espacio, vida extraterrestre), Medieval, Apocalipsis Zombie, Antigüedad (Grecia, Roma), Económicos (el foco está en manejo de empresas o compra/venta de productos), entre muchos otros. Además, existen

---

<sup>2</sup><https://boardgamegeek.com/thread/119726/why-are-they-called-meebles>

juegos basados en alguna *IP* (*Intellectual Property/Propiedad Intelectual*) preexistente (existe gran cantidad de juegos basados en **Star Wars** o **El Señor de los Anillos**).

- **Cantidad de participantes:** En general los juegos presentan un rango de cantidad de jugadores (por ejemplo, 1-4 jugadores, 3-6 jugadores, etc.), pero en otros casos pueden estar diseñados para un número específico (1 jugador o 2 jugadores serían los más comunes).
- **Tiempo de juego:** Dentro de la información que siempre se encuentra presente en la caja de los juegos de mesa, se encuentra un estimado de la duración de éste, puede ser un rango en tiempo total o en tiempo por jugador. Algunos ejemplos serían: Juegos con tiempo no mayor a 30 minutos, duración entre 30 y 90 minutos, duración de 2+ horas o 30 minutos por jugador.
- **Nivel de Interacción:** Diferentes mecánicas y sistemas de reglas permiten un rango de acción que en mayor o menor medida habilita a los jugadores interactuar directamente con los recursos y acciones de los demás participantes. Este rango puede ir desde la Interacción directa, con posibilidades de atacar/destruir/robar/influenciar directamente a otros jugadores y sus recursos, con diferentes niveles, hasta una indirecta, donde cada jugador maneja de forma principalmente independiente sus recursos, sin mayores oportunidades de injerencia sobre lo obtenido por los demás jugadores. Se puede llegar al extremo de que la interacción sea inexistente, donde ninguna acción de un jugador influye en las posibilidades de los demás.
- **Complejidad:** Se puede resumir en la pregunta “¿Qué tan difícil de jugar/aprender es?”. A pesar de ser difícil de estandarizar y definir en una escala universal, existen esfuerzos comunitarios en esta dirección<sup>3</sup>. Por ejemplo, en BGG existe un escala de 5 puntos (entre el 1 y el 5, desde *Light* (Liviano) a *Heavy* (Pesado), en ese orden), bajo el concepto de *Weight* (Peso), basado en la compilación de opiniones de los usuarios del sitio.
- **Categoría:** BGG nos presenta amplia lista de 84 categorías<sup>4</sup>, donde se agrupan grupos de juegos con características similares. Se pueden considerar

<sup>3</sup><https://boardgamegeek.com/wiki/page/Weight>

<sup>4</sup><https://boardgamegeek.com/wiki/page/Category>

categorías particularmente importantes las siguientes:

- *Party Games*: Rápidos y casuales, para una gran cantidad de jugadores.
- *Children's Games*: Orientados a niños, pueden tener un objetivo educacional.
- *Abstract Games*: Sin temática y usualmente para 2 jugadores, como el Ajedrez o Go
- *Strategy Games*: En estos títulos, las decisiones de los jugadores son preponderantes a la hora de definir al ganador, con un efecto reducido del azar.
- *War Games*: Recreación detallada de eventos bélicos históricos.

Estas categorías permiten agrupar a grandes rasgos, de manera rápida y simple el posible público objetivo del juego en particular.

- **Edad Sugerida**: Dentro de los datos disponibles en la caja de cada título, puede encontrarse una edad mínima recomendada para los jugadores. Esto puede estar asociado tanto a temáticas del juego como a la cantidad o dificultad de las decisiones a tomar durante una partida. Este parámetro puede ser particularmente útil a la hora de generar recomendaciones en juegos para niños, donde editoriales como **Haba**<sup>5</sup> publican juegos para distintas edades, desde los 2 años en adelante.

### 3.1.3. Métricas de Evaluación

De acuerdo a lo mencionado en la sección 3.1.1.4, se define como objetivo el *Top-K Recommender Task*, es decir, recomendar una lista de los K ítems más apropiados para el usuario.

La evaluación de los sistemas de recomendación siempre ha sido desafiante. De hecho, durante años no existió un consenso claro en la comunidad científica respecto a los experimentos o métricas adecuadas para este fin. Adoptaremos para este trabajo, lo propuesto en (Hernández and Gaudioso, 2008).

El conjunto de datos original será dividido en dos conjuntos disjuntos, los conjuntos de entrenamiento y prueba. El primero, será utilizado para calibrar o ajustar

---

<sup>5</sup><https://boardgamegeek.com/boardgamepublisher/384/haba>

los parámetros de los modelos construidos, y el segundo, será utilizado para su evaluación. Es importante destacar que los sistemas implementados retornarán una lista  $R$  de  $K$  elementos, y ésta será contrastada con una lista  $V$  de  $K$  elementos definida previamente para cada usuario en el conjunto de prueba.

Tal y como fue comentado en la sección 3.1.1.5, este tipo de problemas sufre de la condición de “Sparsity” lo que genera una complejidad adicional a la hora de generar la evaluación de las propuestas. Dado el amplio universo de juegos existentes, la falta de tiempo por parte de los aficionados, la oferta disponible en los mercados locales e incluso su predisposición a evaluar un título en particular en algún momento del tiempo, podría darse, con una probabilidad alta, la situación de que un usuario no haya evaluado un juego que en la realidad pudo gustarle mucho. Es por esto que proponemos un esquema de evaluación en dos etapas, las que se detallan a continuación.

En una primera etapa, se considerará un éxito cada vez que las listas  $R$  y  $V$  compartan  $M \leq K$  o más elementos en común. En este caso, la métrica relevante será la tasa de aciertos, dada por:

$$\text{Tasa\_aciertos} = \frac{\#\text{exitos}}{\#\text{totales}}$$

En una segunda etapa, se seleccionará del conjunto de test un listado de  $E$  expertos que corresponden a usuarios cuyo conocimiento excede ampliamente a los títulos evaluados por los mismos. Para ellos, se realizará una encuesta que buscará evaluar su percepción de la pertinencia de cada recomendación. Con los resultados de dicha encuesta, se evaluará la tasa de recomendaciones pertinentes, dada por:

$$\text{Tasa\_pertinentes} = \frac{\#\text{recomendaciones relevantes}}{\#\text{recomendaciones totales}}$$

.

Note que una tasa igual a 1, implicaría que los usuarios en el conjunto de test han jugado  $M = K$  de los juegos incluidos en la recomendación, lo que quita valor a la recomendación dado que el objetivo original es generar contenido novedoso y potencialmente interesante. Adicionalmente, recordar al lector que no es el objetivo del algoritmo el predecir qué juegos han sido jugados/calificados por el usuario. La segunda parte del proceso de evaluación nos permitirá eludir algunas de las

limitaciones de la primera.

# Capítulo 4

## Metodología

En este capítulo se detalla el proceso completo de ejecución del proyecto, desde la perspectiva de las herramientas utilizadas y el flujo empleado desde la obtención de los datos, su procesamiento y las técnicas de recomendación utilizadas.

### 4.1. Obtención y Procesamiento de los datos

Ya se ha mencionado la importancia del sitio **Boardgamegeek**, tanto para el mercado de los juegos de mesa como para el hobby en sí mismo. En esta sección se detallará como funciona su API como herramienta de extracción de datos, su utilización, algunos ejemplos y el procesamiento posterior de los datos obtenidos, con tal de generar las entradas necesarias para las técnicas de recomendación a implementar.

#### 4.1.1. Boardgamegeek API

##### 4.1.1.1. Definición de API

API<sup>1</sup>, por su sigla en inglés, *Application Programming Interface*. En este contexto, *Application* se refiere a cualquier software con alguna función específica e *Interface* como un contrato de servicio entre dos aplicaciones. Este contrato define el como éstas se comunican entre ellas utilizando solicitudes (*requests*) y respuestas a éstas (*responses*). Su documentación contiene la información de como los desarrolladores

---

<sup>1</sup><https://aws.amazon.com/what-is/api>

deben de, en nuestro caso, estructurar estas solicitudes y como interpretar las respuestas.

Existen múltiples categorías de APIs, pero nos centraremos específicamente en las **REST APIs**. Éstas son las más populares y flexibles encontradas en la web actualmente. REST significa Representational State Transfer. Define una serie de funciones del tipo GET, PUT, DELETE, etc. que los clientes pueden utilizar para acceder a información contenida en el Servidor. Clientes y Servidores intercambian estos datos utilizando el protocolo HTTP.

La principal característica de las REST APIs es el concepto de *statelessness*. *Statelessness* significa que el servidor no almacena información del cliente entre solicitudes. Las solicitudes realizadas por el cliente son similares a ingresar una URL a un navegador para ingresar a un sitio web, y las respuestas son devueltas en texto plano. De esta forma, el cliente envía una solicitud (*request*) al servidor, el servidor utiliza ésta para activar de forma interna los procesos necesarios, para finalmente devolver como respuesta (*response*) los datos solicitados.

#### 4.1.1.2. BGG XML API2

Boardgamegeek disponibiliza la información contenida en el sitio mediante la última versión de su API, [BGG XML API2](#).

Los términos y condiciones de su uso se encuentran detallados en el siguiente [link](#). La información obtenida mediante el uso de esta API se encuentra disponible de manera no exclusiva y de forma gratuita, para usos no comerciales.

Mediante el uso de esta API, un desarrollador puede obtener información de los diferentes elementos almacenados en la base de datos de la BGG. En esta base de datos, cada elemento posible tiene la denominación *thing*. Para el alcance de este proyecto, las llamadas a la API siempre serán sobre elementos de tipo (*type*) *boardgame*.

Como se definió anteriormente, las solicitudes o llamadas realizadas mediante la API de la BGG tienen una forma similar a escribir un URL. En este caso, la URL base de una solicitud tiene la forma <https://boardgamegeek.com/xmlapi2/>. Además, para consultar por información de elementos específicos, debe agregarse lo siguiente al final de la solicitud: [thing?parameters](#), donde los parámetros a ingresar se encuentran resumidos en la tabla 4.1.1.

Parámetro	Descripción
id=NNN	Especifica el o los ids de los <i>things</i> a solicitar. Para solicitar múltiples, estos pueden ser entregados como una lista separada por comas.
type=THINGTYPE	Filtra los resultados solo al (o los) THINGTYPE(s) indicado(s).
stats=1	Entrega rankings y estadísticas sobre ratings para los ítems solicitados.
ratingcomments=1	Entrega el listado completo de ratings y comentarios realizados sobre el o los ítems ingresados. Estos son entregados de forma descendente desde el más alto disponible. Por defecto solo entrega los primeros NNN (con NNN el valor ingresado en <code>pagesize</code> ).
page=NNN	Por defecto, tiene valor 1. Permite obtener un subconjunto diferente de los ratings y comentarios.
pagesize=NNN	Define la cantidad de registros a devolver para los campos que tienen páginas (en este caso, los ratings y comentarios). Valor entre 10 y 100.

**Cuadro 4.1.1:** Resumen de parámetros modificables en solicitudes a la BGG API

Para ingresar múltiples parámetros en la misma solicitud, éstos se separan mediante el uso del símbolo `&`. Algunos ejemplos de utilización son:

1. Extraer información sobre un juego en particular (por ejemplo Indonesia, id 19777), incluyendo estadísticas pero sin ratings ni comentarios:

[https://boardgamegeek.com/xmlapi2/thing?](https://boardgamegeek.com/xmlapi2/thing?id=19777)

[id=19777](https://boardgamegeek.com/xmlapi2/thing?id=19777)

[&type=boardgame](https://boardgamegeek.com/xmlapi2/thing?id=19777)

[&stats=1](https://boardgamegeek.com/xmlapi2/thing?id=19777)

2. Extraer información de múltiples juegos (por ejemplo Catan, Carcassonne y Ticket to Ride, con ids 13, 822 y 9209, respectivamente), sin las estadísticas de ratings ni rankings y fijando la página de ratings y comentarios a obtener como la número 5: [https://boardgamegeek.com/xmlapi2/thing?](https://boardgamegeek.com/xmlapi2/thing?id=13,822,9209)

[id=13,822,9209](https://boardgamegeek.com/xmlapi2/thing?id=13,822,9209)

[&type=boardgame](https://boardgamegeek.com/xmlapi2/thing?id=13,822,9209)

[&ratingcomments=1](https://boardgamegeek.com/xmlapi2/thing?id=13,822,9209)

[&page=5](https://boardgamegeek.com/xmlapi2/thing?id=13,822,9209)

Los resultados entregados por la API están en formato XML (Extensible Markup

Language) y son devueltas como texto plano. Ejemplos de estas llamadas a la API se encuentran en el Apéndice [A1](#).

Para la obtención de la información de todos los ítems (juegos de mesa) disponibles en la base de datos de la BGG, el proceso será dividido en tres etapas:

- Identificar y obtener los ids de los ítems a extraer.
- Extraer la información de los ítems identificados.
- Extraer los ratings y comentarios hechos por los usuarios.

**Identificar y obtener los ids de los ítems a extraer:** Para este propósito, la BGG nos presenta un archivo en formato *csv* (comma-separated values) en el siguiente [link](#), que incluye:

- Id
- Nombre
- Año de publicación
- Múltiples rankings
- Otros Rankings
- Rating promedio ajustado
- Rating promedio
- Número de ratings

Los rankings incluidos corresponden a las categorías:

- Global
- Abstract
- Customizable
- Children's Games
- Family
- Party
- Strategy

- Thematic
- Wargames

mencionadas previamente en la sección 3.1.2.

**Extraer la información de los ítems identificados:** Considerando la información obtenida en el paso anterior y siguiendo el ejemplo 1 de la sección 4.1.1.2, es posible extraer información más detallada para cada juego. Como ya se vió en la sección 4.1.1.2, es posible realizar solicitudes por múltiples ítems. Como el listado actual de juegos de mesa disponible es cercano a las 150 000 entradas, es importante definir un buen  $N$  para la cantidad de ítems solicitados por *request*, considerando un buen balance entre cantidad de información por *request* y el tiempo de respuesta por *request*.

La información que es posible obtener en esta etapa puede ser separada en dos grandes categorías: Campos específicos y listas. Los campos específicos son detalles de información con un valor unitario, dentro de los que se encuentran:

- Id
- Nombre
- Descripción
- Año de publicación
- Mínimo de jugadores permitidos
- Máximo de jugadores permitidos
- Número de ratings
- Rating promedio ajustado
- Rating promedio
- Peso promedio
- Tiempo de juego esperado
- Edad mínima sugerida por la editorial

Por otro lado, también es posible encontrar datos que, debido a su naturaleza, son entregados en forma de listas, donde encontramos:

- Encuestas
- Tags

Las encuestas disponibles son abiertas al público y se encuentran disponibles en la página específica de cada juego. Éstas cuentan con la cantidad de votos por cada opción y consideran la siguiente información:

- Cantidad de jugadores sugerida
  - Best/Mejor número.
  - Recommended/Recomendado.
  - Not Recommended/No recomendado.
- Edad mínima sugerida
  - Edades 2, 3, 5, 6, 7, 8, 10, 12, 14, 18, 18 y 21+.
- Dependencia del idioma
  - “No necessary in-game text” / “Sin texto necesario durante el juego”.
  - “Some necessary text - easily memorized or small crib sheet” / “ Algo de texto necesario - fácilmente memorizable o pequeña hoja de ayuda necesaria.”
  - “Moderate in-game text - needs crib sheet or paste ups” / “Cantidad moderada de texto durante la partida - Necesita hojas de ayuda o pequeñas modificaciones”.
  - “Extensive use of text - massive conversion needed to be playable” / “Amplio uso de texto durante la partida - Necesita gran cantidad de modificaciones para ser jugable en otro idioma”.
  - “Unplayable in another language” / “Imposible de jugar en otro idioma”.

De estos datos, serán utilizados para generar *features* las encuestas de *Cantidad de jugadores sugerida* y *Dependencia del idioma*.

**Extraer los ratings y comentarios hechos por los usuarios:** Habiendo obtenido todos los datos pertinentes a los ítems a procesar con los pasos anteriores, se tiene identificados los ids de cada uno y además la cantidad de ratings existentes

para estos en la base de datos de la BGG. Con esto, podemos definir la estructura de las consultas a realizar a la API de manera de obtener todos los ratings y comentarios disponibles.

Cada juego cuenta con una cantidad particular de ratings, y si bien, es posible incluir una lista de juegos por solicitud a la API, dentro de sus limitaciones se encuentra que, al ingresar múltiples ids en una consulta y especificar el número de página, el servidor entrega los resultados del mismo número de página para cada ítem solicitado.

Luego, considerando como entrada el listado de ítems con su respectiva cantidad de ratings ( $n\_ratings$ ), ordenadas de forma descendente por este número de ratings, considerando solo ítems con número de ratings mayor a 0 y además, un tamaño de página ( $page\_size$ ) de 100 (el valor máximo posible), definimos una tabla con la siguiente estructura:

<b>game_id</b>	<b>n_ratings</b>	<b>n_pages</b>	<b>last_page_extracted</b>
13	124967	1250	0
822	124571	1246	0
30549	123421	1235	0
68549	102846	1029	0
⋮	⋮	⋮	⋮
707	99	1	0
4325	99	1	0

**Cuadro 4.1.2:** Ejemplo de entrada Algoritmo 1

Donde  $n\_pages$  es la aproximación por exceso de la cantidad  $n\_ratings/page\_size$  y  $last\_page\_extracted$  es inicializada en 0. Con lo anterior, y definiendo a la tabla anterior filtrada con el criterio  $last\_page\_extracted < n\_pages$  como *remaining*, se propone el siguiente algoritmo para generar la lista de consultas necesarias a realizar al servidor:

---

**Algorithm 1** Generar lista de requests de ratings API BGG

---

```

Require:  $\text{len}(\textit{remaining}) > 0$ 
 $n\_remaning \leftarrow \text{len}(\textit{remaining})$ 
 $page\_size \leftarrow 100$ 
 $current\_page \leftarrow 1$ 
 $current\_list \leftarrow \{\}$ 
 $list\_of\_tuples \leftarrow \{\}$ 
while  $n\_remaning > 0$  do
   $valid\_ids \leftarrow \{id \mid id \in \textit{remaining} \text{ if } \textit{last\_page\_extracted}(id) < n\_pages(id)\}$ 
  if  $\text{len}(valid\_ids) > 0$  then
    while  $(\text{len}(valid\_ids) > 0 \ \& \ \text{len}(current\_list) \leq page\_size)$  do
       $current\_item \leftarrow valid\_ids[1]$ 
      add  $current\_item$  to  $current\_list$ 
      remove  $current\_item$  from  $valid\_ids$ 
       $\textit{last\_page\_extracted}(current\_item) \leftarrow current\_page$ 
    end while
    add  $(current\_page, current\_list)$  to  $list\_of\_tuples$ 
     $current\_list \leftarrow \{\}$ 
  else
     $current\_page \leftarrow current\_page + 1$ 
  end if
  update  $\textit{remaining}$ 
   $n\_remaning \leftarrow \text{len}(\textit{remaining})$ 
end while

```

---

Con lo anterior, se obtiene una lista de elementos de la forma (page\_number, list\_of\_items), con la cual es posible generar las URLs para las consultas a la API de la BGG, siguiendo el ejemplo 2 de la sección 4.1.1.2.

### 4.1.2. Procesamiento y Almacenamiento de datos

De acuerdo a lo detallado en la sección anterior, al realizar la extracción de la información necesaria desde la base de datos de la BGG, será necesario procesar dos tipos distintos de respuesta desde el servidor: Información de los ítems e Información de los ratings generados por los usuarios.

#### 4.1.2.1. Procesamiento de los datos de juegos

**Información base de un juego:** Como fue detallado en la sección 4.1.1.2, la mayoría de los campos extraídos con la llamada del formato ejemplo 1 pueden ser procesados directamente como texto o siendo transformados a formato numérico.

El nombre indicado será, por lo general, en inglés, aunque existe la posibilidad de que su idioma original no sea éste. Además, existen algunos campos indicados por la editorial (description, minplayers, maxplayers, minplaytime, maxplaytime) y otros calculados por la plataforma BGG, de forma interna o mediante encuestas a los usuarios (playingtime, minage, ratings, numcomments, average, bayesaverage, averageweight).

Con estos, se puede generar una tabla con la siguiente estructura de columnas:

Columna	Descripción	Ejemplo
bgg_id	id del ítem en BGG	19777
name	Nombre principal del juego	Indonesia
thumbnail	URL con la imagen de portada del juego	<a href="#">link</a>
description	Descripción del juego	Judging by the smirk on the face of the Sultan of Solo...
yearpublished	Año de publicación	2005
minplayers	Mínimo de jugadores indicado	2 jugadores
maxplayers	Máximo de jugadores indicado	5 jugadores
playingtime	Tiempo de juego estimado	210 minutos
minplaytime	Tiempo de juego mínimo indicado	180 minutos
maxplaytime	Tiempo de juego máximo indicado	240 minutos
minage	Edad mínima indicada	14+ años
ratings	Cantidad de ratings	4631 ratings
numcomments	Cantidad de ratings con comentario	1381 comentarios
average	Promedio de los ratings	7.885
bayesaverage	Promedio ajustado de los ratings	7.163
averageweight	Peso o dificultad estimada	3.99

**Cuadro 4.1.3:** Información base de un juego, desde API BGG

**Información de Tags de un juego:** El otro subconjunto principal de información disponible de los juegos está en formato de *tags* o etiquetas, que quedan resumidas en la tabla siguiente:

Tag	Descripción	Ejemplo (Juego/Tag)
boardgamecategory	Agrupar juegos por características o temas similares	Antiquity: City Building, Economic, Medieval.
boardgamemechanic	Término que refiere a un aspecto funcional de un juego	Roads & Boats: Pick-up and Deliver, Modular Board.
boardgamefamily	Grupo de juegos que comparten un concepto: temática, periodo histórico, deporte, etc. Existen otras posibilidades (juego y sus expansiones, serie de juegos del mismo diseñador, etc.)	Ascension: Game: Ascension Deck Building. GIPF, TZAAR, DVONN: Series: GIPF Project.
boardgamedesigner	Diseñador del juego.	The Gallerist: Vital Lacerda.
boardgameartist	Encargado del arte y/o diseño gráfico del juego	Lisboa: Ian O'Toole.
boardgamepublisher	Editoriales que han publicado alguna versión de juego	Indonesia: Editorial Splotter Spell.
boardgameimplementation	Flag o marca si otro título es una reimplementación de éste	King of the Elves reimplementa Elfenland.
boardgameexpansion	Ids de expansiones del juego	Age of Steam: Age of Expansion #1: England & Ireland .
boardgameaccessory	Accesorios o ítems adicionales relacionados al título	Bonfire: Bonfire: Folded Space Insert
boardgamecompilation	Flag o marca que indica si el título es una compilación de juegos o un juego y sus expansiones	Village: Big Box. Agrupa el juego Village y sus expansiones Inn, Port, entre otras
boardgameintegration	Etiqueta que indica con que otros títulos puede integrarse para formar otro juego	Gemischtes Doppel 1, 2, 3 y 4.

**Cuadro 4.1.4:** Información en Tags de juegos, desde API BGG.

**4.1.2.1.1. Número ideal de jugadores:** Por último, el último dato a procesar directamente de lo extraído a nivel de juego es el número ideal de participantes para un título determinado. Como se detalló en la sección 4.1.1.2, dentro de la respuesta a la llamada de la API podemos encontrar los resultados de la encuesta de cada título a diferentes números de participantes, en tres niveles: *Best/Recommended/Not Recommended*.

Nos será útil más adelante el calcular tanto el número ideal o mejor posible de jugadores, como también cantidad de participantes “buenas”, es decir, que no sean el ideal pero que la comunidad en su conjunto las considere también para

una buena experiencia. Por ejemplo, para el juego **Indonesia**, la comunidad ha votado 104 (un 77% de un total de 135 votos) que la mejor experiencia es a 4 jugadores (además de 30 votos como recomendado a este número). Por otro lado, con 72 de 118 votos y 69 de 124 votos es recomendado también a 3 y 5 jugadores. Se utilizarán dos métricas simples definidas de la siguiente manera para comparar un título:

$$best\_score_n = 2 * best\_votes_n + 1 * recommended\_votes_n$$

y

$$good\_score_n = 1 * best\_votes_n + 1 * recommended\_votes_n$$

Es decir, para encontrar el puntaje para encontrar el mejor número de jugadores, sea  $n$  ésta cantidad, se le otorga 2 puntos por voto en la categoría *Best* y 1 voto por la categoría *Recommended*. Por otro lado, para encontrar un buen número de jugadores, simplemente sumamos ambos votos. Finalmente, para el *good\_score*, si éste es mayor al 50% del total de votos para ese número de jugadores, se considerará como un buen número. Continuando así con el ejemplo, aplicado esto al mismo juego anterior **Indonesia**, inicialmente tenemos la siguiente tabla de resultados de la encuesta para número de jugadores:

Player Count	Best	Recommended	Not Recommended	Vote Count
<b>1</b>	1.4% (1)	0.0% (0)	<b>98.6% (73)</b>	74
<b>2</b>	4.3% (4)	17.0% (16)	<b>78.7% (74)</b>	94
<b>3</b>	22.0% (26)	<b>61.0% (72)</b>	16.9% (20)	118
<b>4</b>	<b>77.0% (104)</b>	22.2% (30)	0.7% (1)	135
<b>5</b>	25.8% (32)	<b>55.6% (69)</b>	18.5% (23)	124
<b>More than 5</b>	1.7% (1)	3.3% (2)	<b>95.0% (57)</b>	60
			<b>Total voters</b>	144

**Cuadro 4.1.5:** Resultados encuesta número de jugadores, juego Indonesia

Considerando que la editorial indica que el juego considera entre 2 y 5 jugadores y calculando la métrica mencionada anteriormente, se obtiene:

Player Count	Best votes	Recommended votes	Total votes	Best	Good	Good (%)
2	4	16	94	24	20	21.3 %
3	26	72	118	124	98	83.1 %
4	104	30	135	<b>238</b>	134	<b>99.3 %</b>
5	32	69	124	133	101	<b>81.5 %</b>

**Cuadro 4.1.6:** Cálculo de mejores números de jugadores, juego Indonesia.

Como muestra la tabla, el mejor número de jugadores para el juego **Indonesia** sería 4, además de contar con buena reputación a 3 y 5 jugadores.

#### 4.1.2.2. Procesamiento de los datos de ratings

Para el caso de la extracción de los ratings, como podemos ver en el ejemplo del Apéndice A1, la respuesta del servidor al consultar por éstos es más fácil de procesar. En esta respuesta, podremos encontrar, por ítem y página de ratings solicitados, un listado con  $N$  (en este caso, 100) ratings, con información del nombre de usuario, la nota respectiva, y, de existir, un comentario como texto. Vale la pena comentar que, dado que la BGG es una plataforma web usada a nivel global, aunque principalmente los comentarios sean en inglés, es posible encontrarlos en diferentes idiomas.

Con esto, es posible transformar la respuesta en bruto desde cada consulta, en una tabla similar a la siguiente:

bgg_game_id	bgg_game_name	bgg_user_name	rating	comment
102794	Caverna: The Cave Farmers	geschichte	10	
102794	Caverna: The Cave Farmers	vinashu	10	Reservado para o Frank, mas o Cass e o Renato...
104780	Hoppe Reiter	Teppolainen	6.0	A rather simple dice rolling game with fairly...
127798	Express 01	drbert	4.0	

**Cuadro 4.1.7:** Ejemplo de ratings extraídos y tabulados, desde API BGG

#### 4.1.2.3. Almacenamiento de los datos

En la sección anterior se detalló el formato en el que los datos serán recibidos al consultar a la API de BGG. Habiendo analizado el formato que tendrán estos, lo que sigue es detallar el como serán almacenados para la ejecución del proyecto. Se

definen con objetivos principales para esta etapa, tanto la rapidez a la hora de consultar los datos almacenados, como la facilidad de inserción de datos nuevos. Por último, el factor económico puede ser relevante a la hora de elegir alternativas.

Como primera etapa, se define el almacenar todas las consultas en formato raw o texto plano, de manera de tener la posibilidad de repetir el proceso de limpieza, procesamiento en inserción en el sistema a elegir. Esto viene a ser el similar a levantar un *Data Lake* donde guardar toda la información en su forma más pura, o lo más cercano a lo obtenido desde su fuente, en este caso, las diversas consultas realizadas a la API de BGG. Para el almacenamiento final y de acuerdo a las consideraciones mencionadas, se define utilizar como DBMS (Database Management System) a **SQLite**, un DBMS relacional, open-source y escrito en el lenguaje C. Está basado en archivos, no necesita mantener activo un servidor y es similar a PostgreSQL. Es ligero y fácil de implementar, con el que es posible trabajar desde diversas herramientas como Python (Lenguaje de Programación) y DBeaver (herramienta de gestión de bases de datos relacionales), utilizadas en este proyecto.

Las consultas serán almacenadas en archivos de formato JSON (JavaScript Object Notation), opción popular para almacenar datos semi-estructurados. Para las consultas de información de los juegos, los archivos tienen el formato siguiente:

```
1 {
2   "list_ids": [1,2,3,4,5],
3   "request_str": "http://www.boardgamegeek.com/xmlapi2/thing?id=1,2
4     ,3,4,5&type=boardgame&stats=1",
5   "request_raw": "Ver Apendice"
6 }
```

Para las consultas de ratings, el formato raw será el siguiente:

```
1 {
2   "request_id": 0,
3   "request_url": "http://www.boardgamegeek.com/xmlapi2/thing?id=13,
4     822,30549,68448,167791&type=boardgame&ratingcomments=1&page=1
5     ",
6   "request_raw": "Ver Apendice"
7 }
```

A continuación, y en un paso intermedio, siguiendo la lógica presentada en las secciones 4.1.2.1 y 4.1.2.2, se decide guardar la información de juegos en archivos JSON y la información de ratings directamente en archivos con formato tabular, como csv o parquet. Un ejemplo para la información de juegos en formato JSON se encuentra en el Apéndice.

Finalmente, se definen los siguientes esquemas para la base de datos implementada en SQLite, donde en el primero se muestran las tablas donde se almacenarán los datos de los juegos, de acuerdo a la distinción hecha en la sección 4.1.2.1, usando las tablas *game* para los datos base, *link\_raw* para los datos en forma de *tags* y finalmente los ratings, como se indicó en la sección 4.1.2.2.

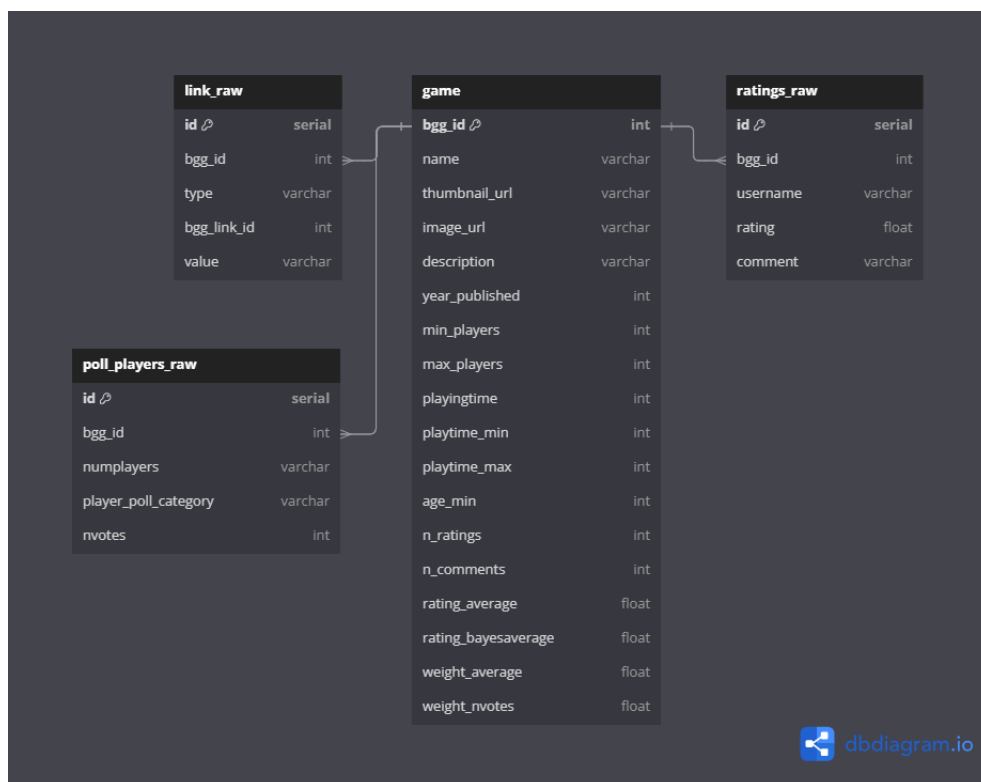
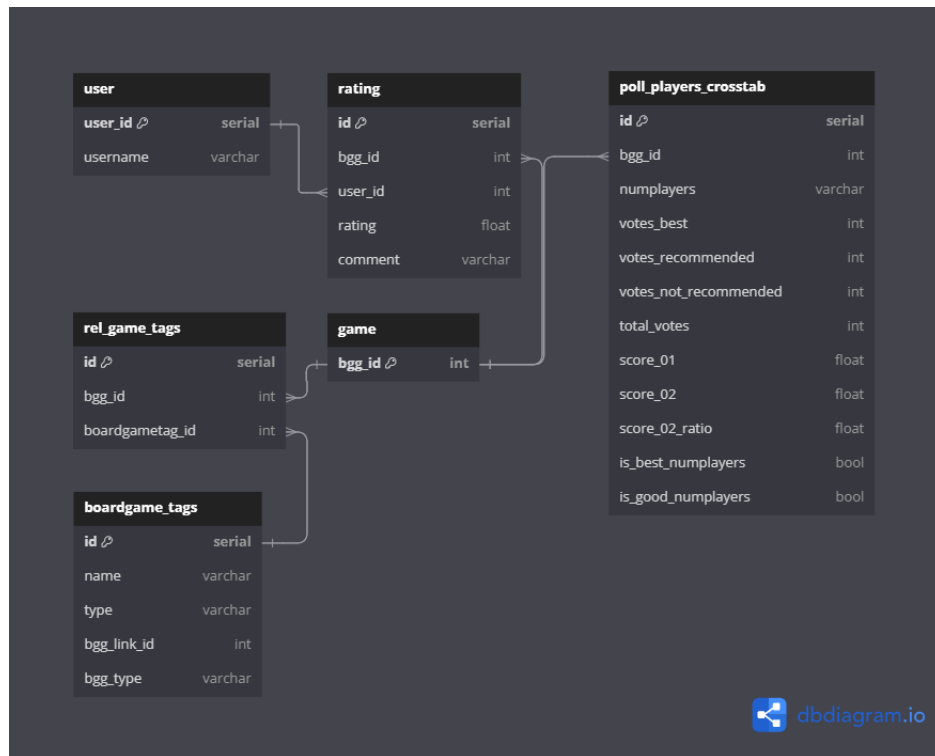


Figura 4.1.1: Esquema Base de datos Inicial

Al generar los ids de los usuarios (ya que estos no se extraen directamente desde la BGG, solo puede obtenerse el nombre de usuario), procesar la encuesta de mejor número de jugadores y categorizar los tags, se generan las siguientes tablas:



**Figura 4.1.2:** Esquema Base de datos con Post-Proceso de los datos

Siguiendo estas definiciones, será posible acceder de manera rápida y eficiente a toda la información necesaria para alimentar a los sistemas de recomendación a implementar.

## 4.2. Sistemas de Recomendación

De entre los métodos descritos en la sección 3.1.1, se han seleccionado los siguientes como base del trabajo de este proyecto:

- Método basado en Filtrado Colaborativo basado en memoria, específicamente, en productos.
- Método basado en Contenido.

A continuación, se detallará en mayor profundidad la manera en que estos métodos serán definidos en la práctica y el procesamiento específico de los datos para generar las entradas de cada uno.

### 4.2.1. Método basado en Filtrado Colaborativo

Para este método, se implementará una versión *naive*, utilizando directamente la información de ratings (transacciones usuario-productos), almacenada en la base de datos, en particular, en la tabla **rating**, sin considerar la existencia de comentarios.

#### 4.2.1.1. Entradas globales

Siguiendo la recomendación de (Ellmeier, 2018), para cada usuario se calculará el promedio de los ratings que ha asignado y esta media se restará de las evaluaciones ingresadas en la plataforma, centrando así todos estos ratings en el valor 0. Con esto, se obtendrá una tabla con la siguiente estructura:

user_id	bgg_id	rating	user_mean	rating_adjusted
1	1	8	7.5	0.5
1	2	6.3	7.5	-1.2
1	3	7	7.5	-0.5
⋮	⋮	⋮	⋮	⋮
1000	5	7	6.3	0.7
1000	123	9.5	6.3	3.2

Aquí, los campos se definen de la siguiente manera:

- **user\_id**: Identificador único en base de datos, representando cada usuario encontrado en la extracción realizada.
- **bgg\_id**: Identificador único proveniente de forma directa desde BGG de cada juego.
- **rating**: Rating ingresado por el usuario. Rango de valores entre 1 y 10.
- **user\_mean**: Media de los ratings ingresado por cada usuario.
- **rating\_adjusted**: Rating ajustado para centrar en 0 todas las evaluaciones. Definido como  $\text{rating} - \text{user\_mean}$ .

Con esta información, utilizando los ratings ajustados, es posible generar la matriz  $R$  de ratings, mostrada en la ecuación 4.2.1, de dimensiones  $m \times n$ , con  $m$  usuarios y  $n$  productos. Cada coeficiente  $r_{i,j}$  representa el rating ajustado del usuario  $i$

sobre el producto  $j$ . Será importante calcular el parámetro *sparsity level* ( $SL$ ) de la matriz  $R$ , definido en (Sarwar et al., 2000) como  $SL = 1 - \frac{\text{nonzero entries}}{\text{total entries}}$ .

$$R_{m,n} = \begin{pmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,n} \\ r_{2,1} & r_{2,2} & \cdots & r_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m,1} & r_{m,2} & \cdots & r_{m,n} \end{pmatrix} \quad (4.2.1)$$

Como medida de similaridad entre ítems, se utilizará el *cosine similarity*. Definida como el ángulo entre dos ítems, con valores entre -1 y +1 (disimilitud a similitud, respectivamente). Esta métrica fue utilizada en (Sarwar et al., 2000) para métodos de filtrado colaborativo basado en productos, sobre el *dataset* MovieLens (recopilación de ratings de películas). Matemáticamente, el *cosine similarity* entre dos vectores  $u$  y  $v$ , se define como el producto escalar entre éstos, dividido por el producto de los módulos de ambos vectores (4.2.2).

$$\text{cosineSim}(u, v) = \frac{\vec{u} \cdot \vec{v}}{\|\vec{u}\| \cdot \|\vec{v}\|} \quad (4.2.2)$$

#### 4.2.1.2. Generación de Recomendaciones

Teniendo la matriz  $R$  como base, se define el procedimiento de generación de Top-K recomendaciones para un usuario de la manera siguiente:

- **Entradas Usuario:**

- Listado de  $p$  ítems o productos valorados positivamente por el usuario,  $p_{items}$
- Listado de  $n$  ítems o productos valorados negativamente por el usuario,  $n_{items}$ .

- **Procedimiento:**

- Definir el listado de posibles recomendaciones, eliminando del total de ítems los ingresados por el usuario (positivos y negativos),  $r_{items}$
- Calcular el *cosine similarity* de cada posible recomendación con los listados positivos y negativos.

- Calcular el puntaje positivo para cada ítem  $u$  de las posibles recomendaciones,  $p_{score}$ , como

$$p_{score}(u) = \sum_{j=1}^p \text{cosineSim}(u, v_j), v_j \in p_{items}$$

- Calcular el puntaje negativo para cada ítem  $u$  de las posibles recomendaciones,  $n_{score}(u)$ , como

$$n_{score} = \sum_{j=1}^n \text{cosineSim}(u, v_j), v_j \in n_{items}$$

- Calcular el puntaje final de recomendación  $r_{score}$  para cada ítem, como

$$r_{score}(u) = p_{score}(u) - n_{score}(u)$$

- Ordenar los ítems  $u$  de forma descendente por  $r_{score}$ .

■ **Salidas Usuario:**

- Filtrar y entregar el Top-K de ítems de los resultados calculados.

**Posibles Variantes:** Siguiendo la metodología descrita en (Ellmeier, 2018), se propone la implementación de dos variantes adicionales para el procedimiento descrito anteriormente:

- Para reducir la influencia de los ítems entregados por el usuario como negativos o no de su preferencia, se propone ponderar el  $n_{score}$  para cada ítem  $u$ , de la siguiente forma:

$$\hat{n}_{score}(u) = n_{score}(u) * \left(1 - \frac{\text{rank}_R(u)}{|R|}\right)$$

donde  $R$  es el conjunto de posibles recomendaciones, es decir, los ítems considerados por el sistema de recomendación, al descartar los que están siendo utilizados como entrada y

$$\text{rank}_R(u)$$

definido como el ranking relativo global de la BGG dentro del conjunto  $R$  (obtenido durante la extracción de los datos).

- Para contrarrestar posibles tendencias sistemáticas a la hora de generar ratings por parte de los usuarios en la plataforma de BGG, se propone generar la matriz  $R$  ajustando los valores a ingresar de la siguiente manera:

$$\hat{r}_{i,j} = r_{i,j} - (\mu + b_i + b_j)$$

donde  $r_{i,j}$  es el rating original ingresado por el usuario,  $\mu$  la media global de todos los ratings ingresados para todos los juegos por todos los usuarios,  $b_i$  el sesgo o *bias* del usuario con respecto a la medio de todos ellos, es decir,  $b_i = \mu_i - \mu_u$ , con  $\mu_i$  la media de ratings del usuario y  $\mu_u$  la media de los usuarios, y  $b_j$  definido de manera análoga, como el sesgo para el ítem  $j$  y su diferencia con el rating promedio de todos los ítems considerados. Los términos  $b_i$  y  $b_j$  son utilizados aquí para ajustar los ratings de acuerdo al comportamiento de los usuarios (por ejemplo, se castigan los ratings del usuario en el caso de que estos sean "generosos", es decir, que las notas ingresadas sean superiores con respecto a la media esperada global de los usuarios). De la misma manera, juegos mejor considerados por la comunidad se ven beneficiados por el ajuste.

Esta segunda variante fue utilizada originalmente en (Koren, 2008) para definir un *baseline* sobre el cual comparar los ratings ingresados por cada usuario, atendiendo a sesgos generados por la evaluación de productos por parte de los usuarios y también efectos posibles sobre las evaluaciones de ciertos ítems.

### 4.2.2. Método basado en Contenido

Para este método, se generará una representación de los juegos estudiados, utilizando tanto la información base de los juegos, almacenada en la tabla **game** de la base de datos, como las etiquetas o *tags*, almacenados en las tablas **boardgame\_tags**, **rel\_game\_tags** y **poll\_players\_crosstab**. Para el procesamiento de la descripción de cada juego, se utilizará un *Vector Space Model*, introducido en 1988 por Salton (Salton, 1988), donde se toma un conjunto de documentos de texto y se llevan a una representación vectorial de éstos, donde

cada componente está relacionado a un término o concepto clave, llamado *n-gram*.

Para determinar estos conceptos, existen diferentes formas de medir su relevancia. Sea  $D$  el conjunto de documentos a utilizar, sea  $f_d(j)$  la frecuencia o cantidad de apariciones de un término  $j$  en un determinado documento  $d$ ,  $T_d$  el conjunto de términos distintos en un documento  $d$  dado, se define el *Term Frequency* (TF) de un término  $j$  en un documento  $d$ , como la razón entre la frecuencia de  $j$  en  $d$  y la máxima frecuencia de los términos en  $d$ , es decir, entre los elementos en  $T_d$  (ecuación 4.2.3). Este puede no representar fielmente la relevancia de un término particular como representativo del documento, ya que, en este proyecto, por ejemplo, es posible que se encuentre muchas veces el término *board game* o “juego de mesa”, pero que esto no sea significativo. Así, también puede definirse otra medida, llamada *Inverse Document Frequency*, con su fórmula en la ecuación 4.2.4. Este valor será alto si un concepto aparece pocas veces en un documento y será bajo si aparece en muchos documentos. Se define tomando a  $j$  como el término específico, sea  $N$  la cantidad de documentos total y  $DF_j$  la cantidad de documentos donde el término  $j$  aparece.

$$TF_d(j) = \frac{f_d(j)}{\max_{t \in T_d} f_d(t)} \quad (4.2.3)$$

$$IDF_D(j) = \log \left( \frac{N}{DF_j} \right) \quad (4.2.4)$$

La ecuación 4.2.5 combina ambos números,  $TF$  e  $IDF$ , generando una métrica nueva. Este valor será alto si el término  $j$  aparece gran cantidad de veces en pocos documentos y será bajo si aparece muy poco (bajo  $TF$ ) o si es mencionado en demasiados documentos (bajo  $IDF$ ).

$$TF-IDF_{d,D}(j) = TF_d(j) \times IDF_D(j) \quad (4.2.5)$$

Teniendo la representación vectorial de los documentos, es necesario definir una medida de similitud entre estos, de la misma manera que se definió en la sección 4.2.1. Los autores de (Pazzani and Billsus, 2007) recomiendan utilizar el *cosine similarity* a otras medidas como la distancia euclideana, ya que reconoce de mejor manera tópicos similares.

La vectorización de los documentos está compuesta por una serie de pasos, donde primero se remueven *stop words*, palabras que no son relevantes para el significado de un texto (conjunciones, preposiciones, etc.). Además, es necesario definir el largo máximo de los conceptos o *tokens* a utilizar, donde estos pueden tener largo entre 1 y  $n$  (de aquí el nombre de *n-gram*). Es posible tanto ocupar solamente los documentos en cuestión u ocupar conjuntos pre-cargados de conceptos, como por ejemplo, algún *dataset* del proyecto **GloVe** (Pennington et al., 2014), un algoritmo de aprendizaje automático no supervisado utilizado para encontrar representaciones vectoriales de palabras, elaborados sobre distintos conjuntos de entrenamientos, como por ejemplo textos de Wikipedia, Twitter o *webscraping* de la web.

#### 4.2.2.1. Entradas globales

Para este método, se considerarán los siguientes datos como entrada, indicando además el procesamiento adicional utilizado:

##### Datos Base

- Descripción del juego:
  - Limpieza y lematización del texto de la descripción mediante la herramienta **spaCy**.
  - Detección de palabras clave o *tokens* utilizando la métrica TF-IDF, removiendo *stopwords* del idioma inglés.
- Tiempo de juego:
  - Promedio entre tiempo mínimo y máximo esperado
  - Para tiempos mayores a las 6 horas, reemplazar por 6 horas.
- *Weight*, peso o dificultad del juego.
- Rating promedio y promedio ajustado global.

**Datos desde etiquetas o *tags*:** Dentro de las etiquetas encontradas, éstas tienen diferentes categorías, como fue descrito en la sección 4.1.2.1. Debido a que éstas etiquetas y categorías han sido generadas por la comunidad y han ido evolucionando con el tiempo, existen posibles solapamientos entre etiquetas para

las distintas categorías a utilizar. A continuación se detalla el procesamiento de las etiquetas y posibles recategorizaciones de éstas.

- Diseñadores:
  - Se extrae desde las etiquetas de tipo **boardgamedesigner** a la totalidad de diseñadores con juegos publicados y entradas ingresadas en BGG.
  - Se identifica a los diseñadores una cantidad de juegos publicados sobre un valor umbral mínimo cada uno con una etiqueta individual.
  - Se agrupa a los diseñadores con menos del mismo valor umbral, añadiendo la etiqueta “Diseñador con poca experiencia”.
- Editoriales:
  - Se realiza un tratamiento similar al de las etiquetas de Diseñadores, definiendo un umbral mínimo y agrupando las editoriales con menos títulos.
- Temática:
  - Primero, se extrae desde las etiquetas de tipo **boardgamefamily**, las que tienen la forma “Theme: {temática}”.
  - Por otro lado, desde las etiquetas de tipo **boardgamecategory**, utilizando conocimiento previo del contexto, puede definirse de forma manual etiquetas a agregar al listado completo de Temáticas.
- Mecánicas:
  - Primero, se extrae desde las etiquetas de tipo **boardgamemechanic**.
  - Se definen familias para agrupar grupos de mecánicas que tienen un nivel de detalle demasiado alto, por ejemplo, existen 12 entradas distintas para subastas/*auction*, que se considera pueden reducirse a una etiqueta.
  - Se agregan desde las etiquetas de tipo **boardgamecategory** un grupo de etiquetas adicional.
- Familias:

- Se extrae desde las etiquetas de tipo **boardgamemecategory** las siguientes: *Wargame*, *Party Game*, *Abstract Strategy* y *Children's Game*.
- Subcategorías:
  - Se extrae desde las etiquetas de tipo **boardgamemefamily** las que tienen la forma “Category: {categoría}”.
- Categorías:
  - Después de remover las etiquetas de tipo **boardgamecategory** utilizadas para otros grupos de etiquetas, se procede a filtrar las restantes.

Se ha categorizado estas etiquetas para poder asignarles ponderaciones distintas diferentes por grupo, siguiendo las recomendaciones de pesos utilizada en [Wadkinds \(2022\)](#).

Finalmente, se puede generar una Matriz  $C$ , similar a la matriz  $R$ , donde se ubicarán los juegos en las filas y los *features* generados en las columnas, obteniendo una representación vectorial para cada ítem.

#### 4.2.2.2. Generación de Recomendaciones

Para la generación de recomendaciones, teniendo en cuenta que lo que se obtuvo en la sección anterior es también una representación vectorial de los ítems a evaluar, se seguirá el mismo procedimiento indicado en la sección 4.2.1.2, tomando como entradas un listado de ítems evaluados positivamente y otro listado de ítems evaluados negativamente por el usuario, para finalmente entregar como salida el Top-K de juegos a recomendar, utilizando el método basado en contenido.

### 4.3. Implementación

Como ya ha sido mencionado previamente, la fuente principal de información para este proyecto ha sido la plataforma BGG, teniendo acceso a ella mediante la API disponibilizada por ésta. Todo la extracción, limpieza y procesamiento de los datos, creación y poblado de la base de datos, implementación y posterior análisis fue realizado en una máquina personal, utilizando el lenguaje de programación **Python**. **Python** es un lenguaje de programación interpretado de alto nivel, de

código abierto y uno de los lenguajes más populares en la actualidad.

En la tabla 4.3.1 se resumen las librerías de Python utilizadas, además de las características generales de la máquina utilizada.

<b>OS</b>	Microsoft Windows 11 Home
<b>Python</b>	Version 3.7.11
<b>Libraries</b>	bs4 (4.12.2) xmldict (0.13.0) matplotlib (3.4.3) numpy (1.21.2) pandas (1.3.4) pyarrow (12.0.1) scipy (1.7.3) scikit-learn (0.23.2) sqlite3 (3.45.1) sqlalchemy (1.3.13)
<b>Sqlite DB</b>	Version 3.45.1
<b>CPU</b>	Intel i5-9300H, 2.40GHz, 4 Cores
<b>RAM</b>	32GB

**Cuadro 4.3.1:** Características de la Máquina y librerías de Python utilizadas.

El código utilizado se encuentra organizado en archivos de tipo *jupyter notebook* (.ipynb) y otros archivos de Python (.py), además del archivo binario que cuenta como base de datos, denominado “bgg.db”. Todo el código se encuentra en almacenado en un repositorio de **Github** (plataforma de desarrollo de software que permite crear, almacenar y mantener código) personal, a disposición de los interesados.

# Capítulo 5

## Análisis

En este capítulo se realizará un análisis exploratorio de los datos extraídos y organizados en la forma descrito en la sección 4.1.2.3, con el objetivo de contextualizar mejor el objeto de estudio, además de aportar a la toma de decisiones con respecto a los mismos, definiendo umbrales razonables para la selección de usuarios y juegos a considerar (por ejemplo, definir un mínimo de ratings asociados a cada usuario y/o juego).

A la fecha de la última extracción de los datos (2024-02-08), se han encontrado 116 215 juegos, con 88 567 títulos con al menos un rating ingresado. Con respecto a los usuarios, ya que estos son obtenidos directamente desde los ratings extraídos, se tienen 521 911 usuarios, todos con al menos un rating ingresado (el aproximado a la fecha de extracción total de usuarios registrados es de aproximadamente 720 000<sup>1</sup>).

Por otro lado, la cantidad de ratings extraídos fue de 24 547 733. donde el usuario con más ratings ha registrado 13 990 y el juego con más ratings es **Catan**<sup>2</sup>, sumando una cantidad de 124 228 registros. Finalmente, y solo como referencia, ya que éstos no serán utilizados en este proyecto, se obtuvo un total de comentarios registrados, con el juego más comentado siendo nuevamente **Catan**, con 21 873 registros, seguido de cerca por **Carcassonne**<sup>3</sup>, con 21 527 comentarios. A continuación, un cuadro resumen con los números mencionados:

---

<sup>1</sup><https://boardgamegeek.com/users>

<sup>2</sup><https://boardgamegeek.com/boardgame/13/catan>

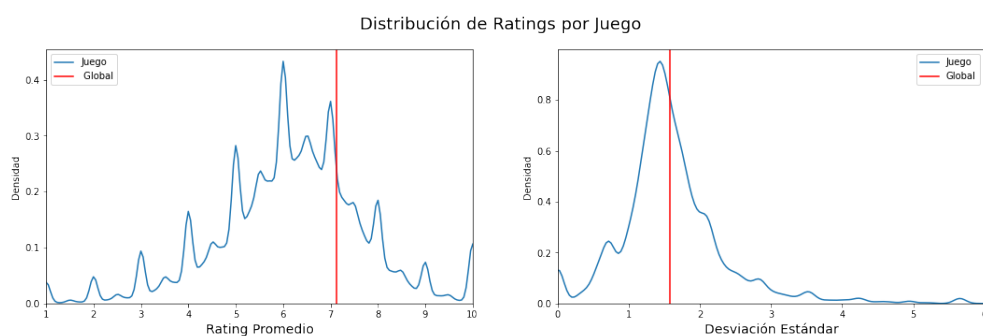
<sup>3</sup><https://boardgamegeek.com/boardgame/822/carcassonne>

Total de Juegos	116 215
Total de Juegos con Ratings ( $n \geq 1$ )	88 567
Total de Usuarios (estimado)	720 000
Total de Usuarios con Ratings ( $n \geq 1$ )	521 911
Total de Ratings Registrados	24 547 733
Total de Ratings Juego con mayor cantidad (Catan)	124 228
Total de Comentarios Juego con mayor cantidad (Catan)	17 957

**Cuadro 5.0.1:** Números Resumen Extracción datos desde BGG

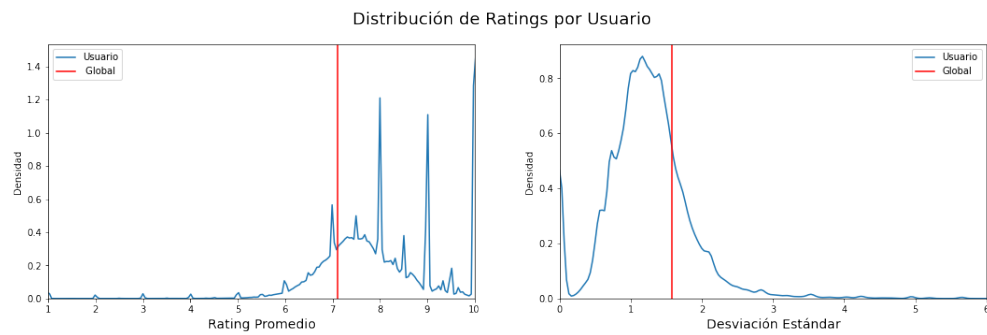
Como referencia de otros intentos previos en la literatura de trabajo con datos de la BGG, en [Aranda et al. \(2007\)](#), en el año 2007, se obtuvo registros asociados a 30 000 juegos y cerca de 40 000 usuarios con al menos un rating. Por otro lado, en [Ellmeier \(2018\)](#), en el año 2018, se extrajo información de un total de 88 567 juegos, 197 979 usuarios y 10 712 438 ratings.

Analizando la distribución de ratings para los juegos, nos encontramos que el rating promedio es de 7.12, una mediana de 7.0 y una distribución estándar promedio de 1.58 puntos. En la figura 5.0.1 se muestra un resumen de esta información. Son notables los picos en los números enteros del gráfico de ratings promedio, esto es debido a que el **83 %** de los ratings ingresados corresponde a cifras cerradas.



**Figura 5.0.1:** Revisión Ratings por Juego, datos desde BGG

Por otro lado, al estudiar la distribución de ratings desde la perspectiva de los usuarios, nos encontramos con una media de 7.64, una mediana de 7.60 y una desviación estándar de 0.88. Cabe notar que la cantidad de usuarios con solo un rating ingresado es de 111 194 (el 21.3 % del total), además del número de usuarios con más de un rating ingresado pero que presenta una desviación estándar de 0, que es de 22 611 (el 4.33 %).



**Figura 5.0.2:** Revisión Ratings por Usuario, datos desde BGG

Además, considerando los 24 547 733 ratings extraídos, correspondientes a 88 567 juegos e ingresados por 512 911 usuarios, se tiene un *Sparsity* de 0,9995.

# Capítulo 6

## Resultados y Discusión

A continuación, se presentan los resultados de los experimentos realizados durante este proyecto.

### 6.1. Definición de los Experimentos

Tomando en consideración lo expuesto en el capítulo anterior, para la aplicación y evaluación de los métodos implementados, se tomarán las siguientes definiciones para el conjunto de datos:

#### Juegos:

- Considerar juegos con un mínimo de 30 ratings registrados como parte el conjunto de de posibles ítems a recomendar.

Con esto, se tiene un universo de 26 246 juegos disponibles para los sistemas.

#### Recomendaciones:

- Para un usuario en particular, tomando en consideración su rating promedio personal, se define, para los datos disponibles de éste, ítems **positivos** y **negativos** de la siguiente manera: Si un ítem evaluado por el usuario tiene un rating mayor o igual a su media, se considera como positivo y en caso contrario, como negativo. Así, una recomendación se considerará **adecuada** si para el usuario en particular el ítem recomendado es **positivo**.

**Usuarios:** Se definen dos criterios diferentes para conjuntos de usuarios a utilizar:

- Usuarios a utilizar para la matriz de ratings del Sistema basado en Filtrado Colaborativo.
- Usuarios a considerar como parte del conjunto de test para el experimento 01.

Para el primer conjunto de usuarios, se define como primeras restricciones: el contar con una cantidad mayor a 10 ratings emitidos, una media de evaluaciones del usuario entre 5,5 y 8,5 (valor dentro del rango: **media global  $\pm$  1 desviación estándar global**, con valores 7,1 y 1,5, respectivamente) y desviación estándar para el usuario mayor a 0,5, dentro de los juegos considerados en el estudio. Con este primer criterio, se obtiene una población total de 322 099 usuarios.

El segundo criterio utilizado fue el siguiente: Contar con una cantidad de ítems positivos mayor o igual a 15 y una cantidad de ítems negativos mayor o igual a 5. Esto fue motivado por el objetivo de generar una lista de Top-10 recomendaciones y considerando un mínimo de 10 ítems positivos y 5 ítems negativos de entrada para cada usuario. Con este filtro, se obtuvieron 293 451 usuarios. Se realizó un muestreo aleatorio para seleccionar los usuarios de test, obteniendo así 41 298 usuarios. Dejando fuera estos usuarios del conjunto de entrenamiento, se obtuvo finalmente 252 153 usuarios para el entrenamiento, con un total de 22 750 642 ratings válidos.

### 6.1.1. Experimento 01: Conjunto Test BGG

Para este experimento se utilizó el conjunto de usuarios definido en la sección anterior de 41 298 usuarios. Para cada usuario, primero se seleccionó una muestra aleatoria de 10 ítems dentro de sus ítems positivos. Este conjunto será el utilizado para comparación con las recomendaciones realizadas por ambos métodos. A continuación, se define como entrada para el usuario todos sus otros ítems positivos, además de sus 5 ítems negativos peor evaluados.

### 6.1.2. Experimento 02: Encuesta Expertos

Para este experimento, se hizo un llamado a 100 voluntarios en distintas comunidades del hobby de los juegos de mesa, principalmente en Chile, recibiendo respuesta de 34. En términos de la captura de información, puede separarse a los participantes en dos conjuntos:

- Usuarios con cuenta registrada en la plataforma BGG y que cuentan con una cantidad mínima de ratings ingresados en ésta.
- Usuarios que no cumplen lo anterior, a los cuales se les pidió elaborar una lista manualmente de juegos.

Para el primer caso, al realizar la extracción desde la BGG, ya se obtuvo la información necesaria de estos usuarios para generar una lista de recomendaciones personalizada. Para el segundo caso, se definió entregar una lista con evaluaciones en forma de notas/ratings (siguiendo la escala de recomendación de la BGG 4.1.2.2) o directamente con la categoría positivo/negativo ya definida, de forma manual. Como ejemplo, un fragmento uno de los formularios recibidos:

<b>bgg_game</b>	<b>bgg_id</b>	<b>rating</b>	<b>value</b>
Age of War	155695	6.5	negative
Agricola	31260	8	positive
Anachrony	185343	7	positive
Ascension: Deckbuilding Game	69789	5	negative
Azul	230802	7	positive

**Cuadro 6.1.1:** Ejemplo de Formulario Experimento 02

En ambos casos, utilizando toda la información disponible, se generó una lista de 10 recomendaciones con cada método implementado, para que posteriormente, cada voluntario evalué las recomendaciones entregadas utilizando las siguientes categorías:

- **Adecuada:** He jugado el título y me gusta o conozco información de éste y estoy interesado en jugarlo/adquirirlo.
- **Sin Información:** No tengo información del título en cuestión.
- **Inadecuada:** He jugado o conozco el título y no me interesa comprarlo o volver a jugarlo.

Las respuestas fueron entregadas y recopiladas en una planilla excel. Para la totalidad de los formularios entregados, se revisó y validó de forma manual las respuestas ingresadas por los voluntarios.

Finalmente, se considera importante hacer notar que el objetivo al realizar este segundo experimento, que, si bien, en magnitud no es comparable al primero, es el de complementar los resultados del primero, aprovechando la activa comunidad del hobby en el país y que al realizarse de forma manual y directa con cada voluntario, se puede obtener información sobre relaciones usuario-ítem que pueden existir pero no estar ingresadas en la base de datos de la BGG.

## 6.2. Resultados

### 6.2.1. Experimento 01

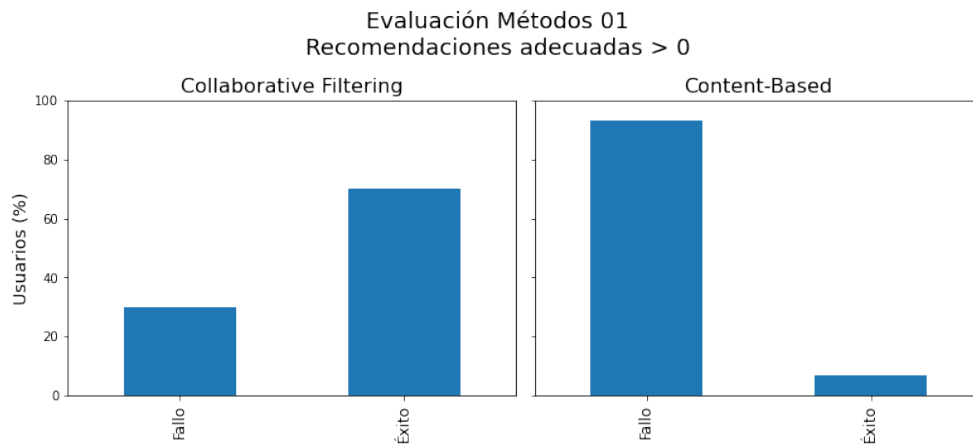
A continuación, se presentan los resultados obtenidos en la ejecución del experimento 01, usando los datos de los usuarios de test seleccionados. Se generaron listas de recomendaciones de tamaño 10 con cada sistema (basado en filtrado colaborativo y basado en contenido) y para cada usuario, para luego comparar con el listado de ítems positivos apartado previamente cada uno. Para este experimento en particular utilizaremos lo ya expuesto en la sección 3.1.3 añadiendo una métrica adicional, esto es, nos centraremos en las dos métricas siguientes:

- Métrica 01: Si la lista de recomendaciones contiene alguno de los elementos apartados previamente, es decir, al menos una recomendación adecuada, se le atribuye un éxito, en caso contrario, un fallo. Esta métrica corresponde a la métrica “Tasa\_aciertos” definida en la sección 3.1.3.
- Métrica 02: Si la lista de recomendaciones no contiene ningún ítem negativo asociado al usuario (no utilizados como entradas para generar recomendaciones del usuario), se le atribuye un éxito, en caso contrario, un fallo. Esta segunda métrica nos permitirá cuantificar recomendaciones no pertinentes.

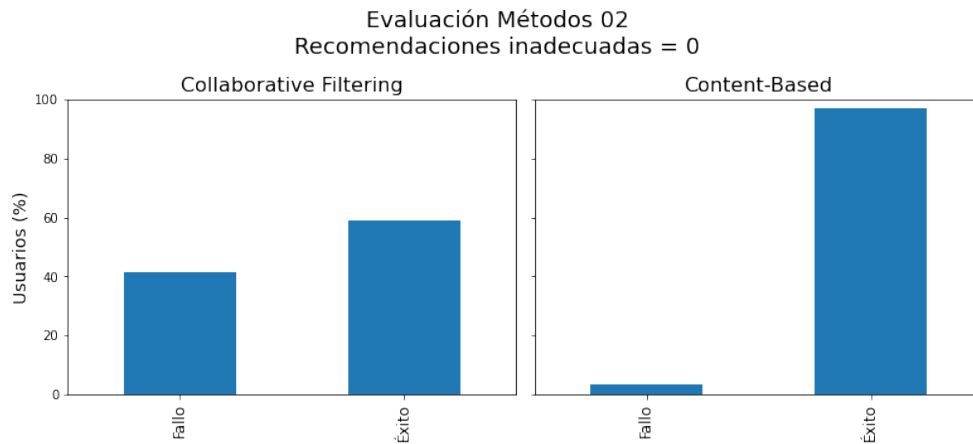
En la tabla 6.2.1 y los gráficos 6.2.1 y 6.2.2 podemos ver el resumen del desempeño de ambos métodos para ambas métricas descritas:

Modelo	Métrica	N usuarios	%
Collaborative Filtering	01	4126	<b>70.1</b>
Collaborative Filtering	02	3462	58.8
Content Based	01	407	6.9
Content Based	02	5703	<b>96.8</b>

**Cuadro 6.2.1:** Resultados Experimento 01



**Figura 6.2.1:** Resultados Experimento 01 - Métrica 01



**Figura 6.2.2:** Resultados Experimento 01 - Métrica 02

Aquí podemos ver que el modelo basado en Filtrado Colaborativo supera con creces al basado en contenido en términos de la métrica 01, con un 70,1% contra un 6,9% de éxito, lo que indica que el primer método genera mejores recomendaciones **adecuadas** sobre el conjunto de prueba. Notemos adicionalmente que esta estrategia requiere tener un mayor conocimiento del usuario, y por ende,

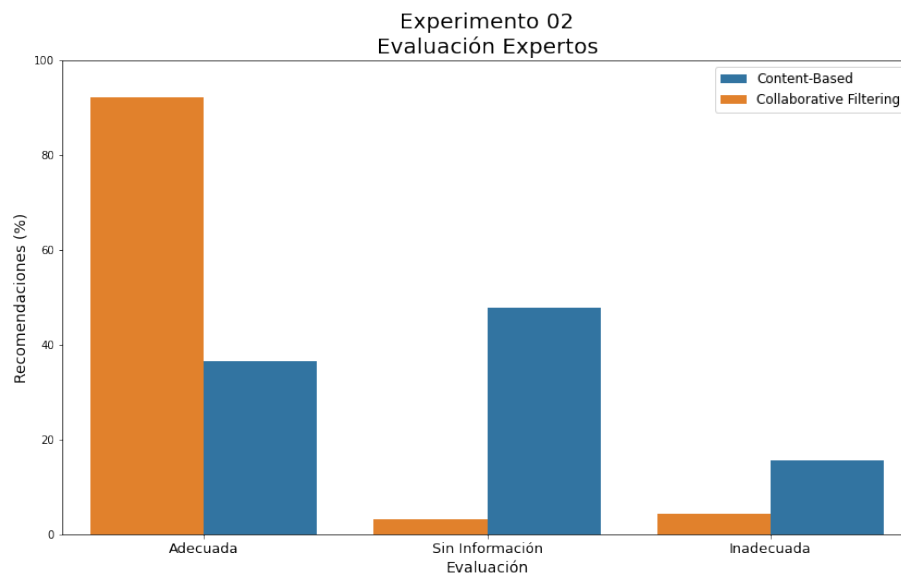
los resultados muestran que es más pertinente para este caso. Por otro lado, en términos de la métrica 02, el método basado en contenido entrega recomendaciones **inadecuadas** en una proporción mucho menor a su contraparte, con un 3,2% contra un 41,2% del Filtrado Colaborativo, lo que lo convierte en una buena alternativa para el caso en que la información conocida del usuario es limitada.

### 6.2.2. Experimento 02

A continuación, los resultados obtenidos en el experimento 02. Como se mencionó en la sección anterior, los ítems recomendados a los expertos en este experimento se encuentran completamente fuera de la información obtenida previamente para ellos, luego, para su evaluación, utilizaremos el *recall* como métrica. Se presentan los resultados en la tabla 6.2.2 y el gráfico 6.2.3:

Modelo	Adecuada (%)	Sin Información (%)	Inadecuada (%)
Collaborative Filtering	92.22	3.33	4.44
Content Based	36.67	47.78	15.56

**Cuadro 6.2.2:** Resultados Experimento 02



**Figura 6.2.3:** Resultados Experimento 02

Aquí, vemos que el sistema basado en *Collaborative Filtering* obtuvo muy buenos

resultados, obteniendo un 92,22 % de recomendaciones **adecuadas** y tan solo un 4,44 % de recomendaciones **inadecuadas**. Con respecto al método Content-Based, podemos ver que entrega un gran cantidad de recomendaciones desconocidas para los usuarios, alcanzando un 47,78 % en esta categoría. Notamos que para ambos algoritmos, la relevancia de las recomendaciones mejora respecto a los valores obtenidos para ambas métricas en el experimento 1 debido a que el conocimiento experto permite compensar la escasez de evaluaciones.

### 6.3. Costo Computacional

Debido a que la ejecución de ambas metodologías es similar, el costo total de generar recomendaciones se puede resumir en los siguientes componentes:

- **Espacio en memoria y tiempo de carga Matriz Sparse**

Para el método basado en filtrado colaborativo, se consideraron 252 153 usuarios y 26 246 ítems, almacenadas en un archivo en formato .npz de 267 MB (formato de librería numpy para guardar *arrays* de forma comprimida). Por otro lado, para el método basado en contenido, se generaron 10 343 *features*, para la misma cantidad de ítems, pudiendo ser almacenado en un archivo .npz de 95 MB. Los tiempos de carga desde este formato para ambos es menor a 1 segundo.

- **Tiempo de cálculo y espacio utilizado por la Matriz de *Cosine Similarities***

Para ambos métodos, luego de realizar la carga de las matrices *sparse*, es necesario calcular matrices de **cosine similarities** entre ítems. Los tiempos de cálculo para el método por filtrado colaborativo y el método basado en contenido fueron de 45 segundos vs 6 minutos y 30 segundos, respectivamente. Por otro lado, tener cargadas estas matrices en RAM o al almacenar directamente en disco duro utiliza una gran cantidad de espacio con respecto al archivo original mencionado en el paso anterior, ocupando 4,8 GB y 5,4 GB, para el filtrado colaborativo y el método basado en contenido, respectivamente.

- **Tiempo de ejecución para una lista de recomendaciones a un usuario**

Al tener cargada las matrices correspondientes y teniendo los listados evaluados del usuario a recibir recomendaciones, los métodos por filtrado colaborativo y basado en contenido tienen velocidades de 10 iteraciones por segundo y 8 iteraciones por segundo, respectivamente.

Toda la información anterior se presenta resumida en la siguiente tabla:

<b>Item Costo</b>	<b>Filtrado Colaborativo</b>	<b>Basado en Contenido</b>
Almacenamiento matriz sparse	267 MB	95 MB
Cálculo matriz cosine similarities	45 segundos	390 segundos
Almacenamiento matriz cosine similarities	4,8 GB	5,4 GB
Velocidad para generar recomendaciones	10,5 it/s	8,1 it/s

**Cuadro 6.3.1:** Resumen Costo Computacional

# Capítulo 7

## Conclusión

A lo largo de este proyecto se han presentado, explorado e implementado diferentes métodos de recomendación. En primer lugar se definió un origen de datos (la plataforma BGG), una vía de obtención de estos (la API de BGG) y mecanismos para la ejecución de esta extracción (mediante scripts automáticos en el lenguaje Python). A continuación se seleccionaron y definieron las herramientas adecuadas para llevar a cabo el proceso de ingestión y almacenamiento de los datos, utilizando conocimiento del negocio o contexto y tomando en consideración tanto las particularidades de éste como los demás objetivos del proyecto. Se realizó un análisis de los datos obtenidos y se tomaron decisiones con respecto a su tratamiento en pos de alimentar los sistemas de recomendación a implementar con datos de calidad. Finalmente se implementaron dos de las técnicas de recomendación más utilizadas en la actualidad, siendo éstas evaluadas en dos contextos distintos.

Se logró estandarizar un procedimiento de extracción, limpieza, proceso y almacenamiento de datos en el contexto de los juegos de mesa modernos, considerando diversos formatos de datos y diversas técnicas de procesamiento, principalmente de *Data Cleaning*, utilizando además de técnicas de Procesamiento de Lenguaje Natural (*NLP*) a la hora de generar representaciones vectoriales de los ítems.

Se implementaron dos estrategias de recomendación con enfoques distintos en sus entradas pero homólogas a la hora de generar recomendaciones, lo que permitió comparar directamente ambas en términos de métricas de precisión, pero además, en función del costo computacional de cada una, donde, el sistema

---

basado en Filtrado Colaborativo obtuvo mejores resultados de precisión, menores tiempos de carga y cálculo de la matriz de similitudes por coseno con respecto a su contraparte (45 segundos vs 6 minutos y 30 segundos, respectivamente), y generación de recomendaciones de manera más rápida (10,5 iteraciones por segundo vs 8 iteraciones por segundo, respectivamente). Estos resultados se condicen con la literatura, donde en (Isinkaye et al., 2015) y (Ellmeier, 2018) se indica que en este contexto el Filtrado Colaborativo obtiene mejores resultados.

Se propusieron dos estrategias de evaluación para los resultados, una estándar Hernández and Gaudioso (2008) y otra basada en una revisión por expertos. Esto permitió refinar las métricas obtenidas para dilucidar la pertinencia de recomendaciones afectadas por el fenómeno de “Sparsity”. En particular, el Filtrado Colaborativo subió desde un 70,1 % con la primera estrategia a un 92,22 % de recomendaciones **adecuadas** en la revisión con expertos, y el mismo fenómeno se observó para el Basado en Contenido, que subió de un 6,9 % a un 37 %.

Considerando la naturaleza del hobby de los juegos de mesa modernos y la preponderancia en este de las reseñas o reviews escritas o en formato de video publicadas por personalidades del nicho, para mejorar el desempeño de las recomendaciones basadas en contenido, es el de ampliar la cantidad de texto incluida en la determinación de sus componentes, tomando, por ejemplo, reseñas escritas y/o comentarios de los usuarios (ambos ejemplos disponibles en la BGG), ya que estos pueden ser ricos en información que no es posible encontrar en la descripción de éstos entregada por la editorial.

Por otro lado, y de manera general, se sugiere el revisar el tratamiento utilizado para re-implementaciones o versiones distintas de un mismo juego (títulos distintos presentes en la BGG con identificadores distintos pero que, en gran medida, son prácticamente idénticos o muy similares), evaluando si es necesario o beneficioso el unificar estos ítems a la hora de alimentar los sistemas de recomendación.

Por otro lado, siguiendo la línea de los Métodos de Filtrado Colaborativo, se propone incorporar información adicional de los usuarios, como por ejemplo, sus colecciones o *wishlist* (juegos que tienen “en la mira”), perfilando a los usuarios para mejorar las recomendaciones basado en esta información adicional.

Finalmente, entendiendo el hobby de los juegos de mesa modernos como una actividad que principalmente se realiza bajo diferentes condiciones, la versión

final del sistema de recomendación propuesto se beneficiaría de considerar la incorporación de post-filtros a la metodología de recomendación, utilizando criterios específicos del juego como lo son: mejor número de jugadores, temática, dificultad de éste o duración de una partida, pero además información externa, como disponibilidad o costo en el mercado o tiendas locales, aplicando también criterios de accesibilidad a la hora de entregar recomendaciones de calidad.

# Bibliografía

- Aggarwal, C. C. (2016). *Recommender systems: Textbook*. Springer.
- Amatriain, X. and Basilico, J. (2012). Netflix recommendations: Beyond the 5 stars (part 1). <https://medium.com/netflix-techblog/netflix-recommendations-beyond-the-5-stars-part-1-55838468f429>.
- Aranda, J., Givoni, I. E., Handcock, J., and Tarlow, D. (2007). An online social network-based recommendation system. *Computer Science*.
- Bobadilla, J., Ortega, F., Hernando, A., and Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-Based Systems*.
- Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User Model User-Adap Inter*.
- Cremonesi, P., Yehuda, K., and Turri, R. (2010). Performance of recommender algorithms on top-n recommendation tasks. *RecSys '10: Proceedings of the fourth ACM conference on Recommender systems*.
- Ellmeier, D. (2018). Implementing and evaluating a recommendation framework for boardgames.
- Galton, F. (1907). Vox populi. *Nature*.
- Hernández, F. and Gaudioso, E. (2008). Evaluation of recommender systems: A new approach. *Expert Systems with Applications*.
- Isinkaye, F., Folajimi, Y., and Ojokoh, B. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal*.
- Jannach, D., Zanker, M., Felfernig, A., and Friedrich, G. (2010). *Recommender systems: an introduction*. Cambridge University Press.
- Koren, Y. (2008). Factorization meets the neighborhood: a multifaceted collaborative filtering model. *KDD '08: Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*.
- Kumar, P. P., Vairachilai, S., Potluri, S., and Mohanty, S. N. (2021). *Recommender Systems: Algorithms and Applications*. Springer.
- Lika, B., Kolomvatsos, K., and Hadjiefthymiades, S. (2014). Facing the cold start problem in recommender systems. *Expert Systems with Applications*.

- Park, Y. J. and Tuzhilin, A. (2008). The long tail of recommender systems and how to leverage it. *Proceedings of the 2008 ACM conference on Recommender systems*.
- Pazzani, M. J. and Billsus, D. (2007). Content-based recommendation systems. *The adaptive Web*.
- Pennington, J., Socher, R., and Manning, C. D. (2014). Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543.
- Pu, P., Faltings, B., Chen, L., and Zhang, L. (2016). Usability guidelines for product recommenders based on example critiquing research. *En: Ricci, F., Rokach, L., Shapira, B., Kantor, P. (eds) Springer, Boston*.
- Resnick, P. and Varian, H. (1997). Recommender systems. *Communications of the ACM*.
- Ricci, F., Rokach, L., and Shapira, B. (2015). *Recommender Systems Handbook*. Springer, Boston, MA.
- Ruiz Iniesta, A. (2014). *Estrategias de recomendación basadas en conocimiento para la localización personalizada de recursos en repositorios educativos*. PhD thesis, Universidad Complutense de Madrid.
- Salton, G. (1988). Automatic text processing. *Addison-Wesley Longman Publishing Co., Inc. isbn: 0-2:1-1227-8*.
- Sarwar, B., Karypis, G., Konstan, J. A., and Riedl, J. (2000). Analysis of recommendation algorithms for e-commerce. *Proceedings of the 2nd ACM conference on Electronic commerce*.
- Shani and Gunawardana (2011). Evaluating recommendation systems. *En: Recommender Systems Handbook*.
- Smith, B. and Linden, G. (2017). Two decades of recommender systems at amazon.com. *IEEE Internet Computing*.
- Su, X. and Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. *Advances in Artificial Intelligence*.
- Varga, E. (2019). *Practical Data Science with Python 3*. Apress.
- Wadkinds, J. (2022). Boardgamegeek recommender system. <https://github.com/threnjen/boardgamegeek/tree/main>.
- Yi, S., Steyvers, M., Lee, M., and Dry, M. (2012). The wisdom of the crowd in combinatorial problems. *Cognitive Science*.
- Zafarani, R., Abbasi, M. A., and Liu, H. (2014). *Social media mining: an introduction*. Cambridge University Press.

---

Çano, E. and Morisio (2017). Hybrid recommender systems: A systematic literature review. *Intelligent Data Analysis*.

# Apéndice A

## A1. Ejemplos de llamadas a API BGG

### A1.1. Consulta Datos Juego

A continuación, un ejemplo de la respuesta de la API de BGG al solicitar datos y estadísticas de un juego en particular, con el request enviado al servidor y la response devuelta por éste:

Request:

```
https://boardgamegeek.com/xmlapi2/thing?
id=19777
&type=boardgame
&stats=1
```

Response:

```
1 <?xml version="1.0" encoding="utf-8"?><items termsofuse="https://boardgamegeek.
   com/xmlapi/termsofuse">
2 <item type="boardgame" id="19777">
3 <thumbnail>https://cf.geekdo-images.com/dMKwaRT7SYvukMuz6A8cxQ__thumb/img/
   RCh0FVF5BT9-ATms0Wc_6hs2z5I=/fit-in/200x150/filters:strip_icc()/pic93485.
   jpg</thumbnail>
4 <image>https://cf.geekdo-images.com/dMKwaRT7SYvukMuz6A8cxQ__original/img/
   uJxqaJuxawHp28Wmvjv4JyUp0ac=/0x0/filters:format(jpeg)/pic93485.jpg</image>
5 <name type="primary" sortindex="1" value="Indonesia"/>
6 <description>Judging by the smirk on the face of the Sultan of Solo, the
   businessman from North Sumatra has just made a mistake. And a costly one
   at that, too. He has paid way too much for merging his rubber company with
   the Sultan&rsquo;s extensive rubber plantations: there are no ships in
```

the area to transport the rubber to the booming cities of Java. And the Sultan is now dripping in cash &mdash; he can bribe city authorities so that his ships will gain preferential access to those ports where they are competing. Or he might invest in building an oil imperium. But wait &mdash; what&rsquo;s that young fellow over there up to? Buying all the shipping lines? That might change the outlook &hellip; considerably, in fact &hellip; let&rsquo;s see&hellip;&#10;&#10;Indonesia is a game in which two to five players build up an economy, trying to acquire the most money. Players acquire production companies, which produce goods (rice, spices, microwaveable meals, rubber, and oil), and shipping companies, which deliver goods to cities. As cities receive goods, they grow, increasing their demands. Production companies earn money for each good delivered to a city, up to the city's capacity, but they must pay shipping companies for the distance traveled, even if they end up losing money. Players can research advantages, like greater shipping capacity or the ability to merge companies, possibly stealing ownership of lucrative plantations or shipping routes by buying out other players.&#10;&#10;Players keep their money hidden, trying to accumulate the most by the end of the third era. When only one type of a company is left, an era ends. New eras bring new cities, companies, and types of goods. At the end of the third era, the player with the most money wins. However, money earned from the last round of operations (selling/shipping goods) counts double, so it can be important to control the timing.&#10;&#10;</description>

```

7 <yearpublished value="2005"/>
8 <minplayers value="2"/>
9 <maxplayers value="5"/>
10 <poll name="suggested_numplayers" title="User Suggested Number of Players"
    totalvotes="144">
11 <results numplayers="1">
12 <result value="Best" numvotes="1"/>
13 <result value="Recommended" numvotes="0"/>
14 <result value="Not Recommended" numvotes="73"/>
15 </results>
16 <results numplayers="2">
17 <result value="Best" numvotes="4"/>
18 <result value="Recommended" numvotes="16"/>
19 <result value="Not Recommended" numvotes="74"/>
20 </results>
21 <results numplayers="3">
22 <result value="Best" numvotes="26"/>
23 <result value="Recommended" numvotes="72"/>
24 <result value="Not Recommended" numvotes="20"/>

```

```
25 </results>
26 <results numplayers="4">
27 <result value="Best" numvotes="104"/>
28 <result value="Recommended" numvotes="30"/>
29 <result value="Not Recommended" numvotes="1"/>
30 </results>
31 <results numplayers="5">
32 <result value="Best" numvotes="32"/>
33 <result value="Recommended" numvotes="69"/>
34 <result value="Not Recommended" numvotes="23"/>
35 </results>
36 <results numplayers="5+">
37 <result value="Best" numvotes="1"/>
38 <result value="Recommended" numvotes="2"/>
39 <result value="Not Recommended" numvotes="57"/>
40 </results>
41 </poll>
42 <playingtime value="240"/>
43 <minplaytime value="180"/>
44 <maxplaytime value="240"/>
45 <minage value="14"/>
46 <poll name="suggested_playerage" title="User Suggested Player Age" totalvotes=
    "27">
47 <results>
48 <result value="2" numvotes="0"/>
49 <result value="3" numvotes="0"/>
50 <result value="4" numvotes="0"/>
51 <result value="5" numvotes="0"/>
52 <result value="6" numvotes="0"/>
53 <result value="8" numvotes="0"/>
54 <result value="10" numvotes="3"/>
55 <result value="12" numvotes="2"/>
56 <result value="14" numvotes="9"/>
57 <result value="16" numvotes="13"/>
58 <result value="18" numvotes="0"/>
59 <result value="21 and up" numvotes="0"/>
60 </results>
61 </poll>
62 <poll name="language_dependence" title="Language Dependence" totalvotes="33">
63 <results>
64 <result level="1" value="No necessary in-game text" numvotes="15"/>
65 <result level="2" value="Some necessary text - easily memorized or small crib
```

```
        sheet" numvotes="17"/>
66 <result level="3" value="Moderate in-game text - needs crib sheet or paste ups
    " numvotes="1"/>
67 <result level="4" value="Extensive use of text - massive conversion needed to
    be playable" numvotes="0"/>
68 <result level="5" value="Unplayable in another language" numvotes="0"/>
69 </results>
70 </poll>
71 <link type="boardgamecategory" id="1021" value="Economic"/>
72 <link type="boardgamecategory" id="1088" value="Industry / Manufacturing"/>
73 <link type="boardgamecategory" id="1011" value="Transportation"/>
74 <link type="boardgamemechanic" id="2012" value="Auction/Bidding"/>
75 <link type="boardgamemechanic" id="2007" value="Pick-up and Deliver"/>
76 <link type="boardgamemechanic" id="2849" value="Tech Trees / Tech Tracks"/>
77 <link type="boardgamefamily" id="73592" value="Components: Gems/Crystals"/>
78 <link type="boardgamefamily" id="10615" value="Country: Indonesia"/>
79 <link type="boardgamefamily" id="74252" value="Digital Implementations:
    SlothNinja"/>
80 <link type="boardgamefamily" id="73596" value="Digital Implementations:
    TableTop Simulator Mod (TTS)"/>
81 <link type="boardgameaccessory" id="274799" value="Indonesia: Cities Upgrade
    Set"/>
82 <link type="boardgamedesigner" id="270" value="Jeroen Doumen"/>
83 <link type="boardgamedesigner" id="407" value="Joris Wiersinga"/>
84 <link type="boardgameartist" id="12603" value="Jeroen Kesselaar"/>
85 <link type="boardgameartist" id="12602" value="Ynze Moedt"/>
86 <link type="boardgamepublisher" id="30552" value="One Moment Games"/>
87 <link type="boardgamepublisher" id="140" value="Spotter Spellen"/>
88 <statistics page="1">
89 <ratings>
90 <usersrated value="4631"/>
91 <average value="7.88454"/>
92 <bayesaverage value="7.16257"/>
93 <ranks>
94 <rank type="subtype" id="1" name="boardgame" friendlyname="Board Game Rank"
    value="308" bayesaverage="7.16257"/>
95 <rank type="family" id="5497" name="strategygames" friendlyname="Strategy Game
    Rank" value="170" bayesaverage="7.3343"/>
96 </ranks>
97 <stddev value="1.52162"/>
98 <median value="0"/>
99 <owned value="5175"/>
```

```

100 <trading value="56"/>
101 <wanting value="878"/>
102 <wishing value="2645"/>
103 <numcomments value="1381"/>
104 <numweights value="443"/>
105 <averageweight value="3.9865"/>
106 </ratings>
107 </statistics>
108 </item>
109 </items>

```

**Listing A.1:** Ejemplo API Response, Game Info.

## A1.2. Consulta Datos Rating

A continuación, un ejemplo de la respuesta de la API de BGG al solicitar las ratings y comentarios de un par de juegos, con el request enviado al servidor y la response devuelta por éste:

Request:

```

https://boardgamegeek.com/xmlapi2/thing?
id=19777
&type=boardgame
&ratingcomments=1
&page=5

```

Response:

```

1 <?xml version="1.0" encoding="utf-8"?><items termsofuse="https://boardgamegeek.
  com/xmlapi/termsofuse">
2 <item type="boardgame" id="19777">
3 <thumbnail>https://cf.geekdo-images.com/dMKwaRT7SYvukMuz6A8cxQ__thumb/img/
  RCh0FVF5BT9-ATms0Wc_6hs2z5I=/fit-in/200x150/filters:strip_icc()/pic93485.
  jpg</thumbnail>
4 <image>https://cf.geekdo-images.com/dMKwaRT7SYvukMuz6A8cxQ__original/img/
  ujxqaJuxawHp28WMvjv4JyUp0ac=/0x0/filters:format(jpeg)/pic93485.jpg</image>
5 <name type="primary" sortindex="1" value="Indonesia"/>
6 <description>Judging by the smirk on the face of the Sultan of Solo, the
  businessman from North Sumatra has just made a mistake. And a costly one
  at that, too. He has paid way too much for merging his rubber company with
  the Sultan&rsquo;s extensive rubber plantations: there are no ships in

```

the area to transport the rubber to the booming cities of Java. And the Sultan is now dripping in cash &mdash; he can bribe city authorities so that his ships will gain preferential access to those ports where they are competing. Or he might invest in building an oil imperium. But wait &mdash; what&rsquo;s that young fellow over there up to? Buying all the shipping lines? That might change the outlook &hellip; considerably, in fact &hellip; let&rsquo;s see&hellip;&#10;&#10;Indonesia is a game in which two to five players build up an economy, trying to acquire the most money. Players acquire production companies, which produce goods (rice, spices, microwaveable meals, rubber, and oil), and shipping companies, which deliver goods to cities. As cities receive goods, they grow, increasing their demands. Production companies earn money for each good delivered to a city, up to the city's capacity, but they must pay shipping companies for the distance traveled, even if they end up losing money. Players can research advantages, like greater shipping capacity or the ability to merge companies, possibly stealing ownership of lucrative plantations or shipping routes by buying out other players.&#10;&#10;Players keep their money hidden, trying to accumulate the most by the end of the third era. When only one type of a company is left, an era ends. New eras bring new cities, companies, and types of goods. At the end of the third era, the player with the most money wins. However, money earned from the last round of operations (selling/shipping goods) counts double, so it can be important to control the timing.&#10;&#10;</description>

```

7 <yearpublished value="2005"/>
8 <minplayers value="2"/>
9 <maxplayers value="5"/>
10 <poll name="suggested_numplayers" title="User Suggested Number of Players"
    totalvotes="144">
11 <results numplayers="1">
12 <result value="Best" numvotes="1"/>
13 <result value="Recommended" numvotes="0"/>
14 <result value="Not Recommended" numvotes="73"/>
15 </results>
16 <results numplayers="2">
17 <result value="Best" numvotes="4"/>
18 <result value="Recommended" numvotes="16"/>
19 <result value="Not Recommended" numvotes="74"/>
20 </results>
21 <results numplayers="3">
22 <result value="Best" numvotes="26"/>
23 <result value="Recommended" numvotes="72"/>
24 <result value="Not Recommended" numvotes="20"/>

```

```
25 </results>
26 <results numplayers="4">
27 <result value="Best" numvotes="104"/>
28 <result value="Recommended" numvotes="30"/>
29 <result value="Not Recommended" numvotes="1"/>
30 </results>
31 <results numplayers="5">
32 <result value="Best" numvotes="32"/>
33 <result value="Recommended" numvotes="69"/>
34 <result value="Not Recommended" numvotes="23"/>
35 </results>
36 <results numplayers="5+">
37 <result value="Best" numvotes="1"/>
38 <result value="Recommended" numvotes="2"/>
39 <result value="Not Recommended" numvotes="57"/>
40 </results>
41 </poll>
42 <playingtime value="240"/>
43 <minplaytime value="180"/>
44 <maxplaytime value="240"/>
45 <minage value="14"/>
46 <poll name="suggested_playerage" title="User Suggested Player Age" totalvotes=
    "27">
47 <results>
48 <result value="2" numvotes="0"/>
49 <result value="3" numvotes="0"/>
50 <result value="4" numvotes="0"/>
51 <result value="5" numvotes="0"/>
52 <result value="6" numvotes="0"/>
53 <result value="8" numvotes="0"/>
54 <result value="10" numvotes="3"/>
55 <result value="12" numvotes="2"/>
56 <result value="14" numvotes="9"/>
57 <result value="16" numvotes="13"/>
58 <result value="18" numvotes="0"/>
59 <result value="21 and up" numvotes="0"/>
60 </results>
61 </poll>
62 <poll name="language_dependence" title="Language Dependence" totalvotes="33">
63 <results>
64 <result level="1" value="No necessary in-game text" numvotes="15"/>
65 <result level="2" value="Some necessary text - easily memorized or small crib
```

```
        sheet" numvotes="17"/>
66 <result level="3" value="Moderate in-game text - needs crib sheet or paste ups
    " numvotes="1"/>
67 <result level="4" value="Extensive use of text - massive conversion needed to
    be playable" numvotes="0"/>
68 <result level="5" value="Unplayable in another language" numvotes="0"/>
69 </results>
70 </poll>
71 <link type="boardgamecategory" id="1021" value="Economic"/>
72 <link type="boardgamecategory" id="1088" value="Industry / Manufacturing"/>
73 <link type="boardgamecategory" id="1011" value="Transportation"/>
74 <link type="boardgamemechanic" id="2012" value="Auction/Bidding"/>
75 <link type="boardgamemechanic" id="2007" value="Pick-up and Deliver"/>
76 <link type="boardgamemechanic" id="2849" value="Tech Trees / Tech Tracks"/>
77 <link type="boardgamefamily" id="73592" value="Components: Gems/Crystals"/>
78 <link type="boardgamefamily" id="10615" value="Country: Indonesia"/>
79 <link type="boardgamefamily" id="74252" value="Digital Implementations:
    SlothNinja"/>
80 <link type="boardgamefamily" id="73596" value="Digital Implementations:
    TableTop Simulator Mod (TTS)"/>
81 <link type="boardgameaccessory" id="274799" value="Indonesia: Cities Upgrade
    Set"/>
82 <link type="boardgamedesigner" id="270" value="Jeroen Doumen"/>
83 <link type="boardgamedesigner" id="407" value="Joris Wiersinga"/>
84 <link type="boardgameartist" id="12603" value="Jeroen Kesselaar"/>
85 <link type="boardgameartist" id="12602" value="Ynze Moedt"/>
86 <link type="boardgamepublisher" id="30552" value="One Moment Games"/>
87 <link type="boardgamepublisher" id="140" value="Spotter Spellen"/>
88 <comments page="5" totalitems="4670">
89 <comment username="JK75" rating="10" value="" />
90 <comment username="Tyler Mayes" rating="10" value="I've owned this for a while
    now (since maybe January) but only finally got it to the table. Holy cow
    is it good. We did mess up a couple of rules that made it longer than it
    should have been and impacted some of the endgame but overall I really
    enjoyed it. The game is long but fully engaging the entire time. Its a
    solid economic game that has a big focus on being able to analyze the
    current game state and adapt. I did not intend to go into shipping in my
    game however I had built a lot of wealth with production and I saw an
    opportunity to use that to steal control of the biggest shipping lines in
    the game. This huge cost was paid off over the next few rounds. Also if
    you like auction games this is one of the best. The merger auctions are
    super intense. I believe this is because there are so few assets in play
```

```
and when a merger occurs it is consolidating those into even fewer. So the
value of the merged company is high, but you don't want to get stuck
overpaying. Can't wait to play this one more."/>
91 <comment username="BeardonBoards" rating="10" value="-Printed Map -
    Meeplesource Upgrades"/>
92 <comment username="KnightFanDan" rating="10" value=""/>
93 <comment username="Mattoise" rating="10" value=""/>
94 <comment username="trezemauro" rating="10" value=""/>
95 <comment username="grschmitkons" rating="10" value=""/>
96 <comment username="RonCarlson01" rating="10" value=""/>
97 <comment username="Tlalox" rating="10" value=""/>
98 <comment username="Tania94" rating="10" value=""/>
99 <comment username="B457i" rating="10" value="If you like complex long lasting
    strategy games Indonesia is exactly what you will like. It gets even more
    fun when you play it with the same people over and over, because everybody
    has to adopt strategies. Furthermore, I can recommend to make alliances,
    with other players. This can make it even more exciting."/>
100 <comment username="TheIrreverend" rating="10" value=""/>
101 <comment username="Planesio" rating="10" value=""/>
102 <comment username="LeftHookLarry" rating="10" value=""/>
103 <comment username="Benitopia" rating="10" value="Mine is a "deluxe custom"
    version with much bigger and better board, cards etc."/>
104 <comment username="G3ck092" rating="10" value="The best economic game I've
    played so far. The focus being purely on what companies you want to open,
    what companies to merge and how to expand makes the rules easy to learn,
    and the deep strategy easier to think about. Every time I think about
    playing another economic game I ask myself "why not just play Indonesia
    instead?""/>
105 <comment username="ELTAUSER" rating="10" value=""/>
106 <comment username="brandtb" rating="10" value=""/>
107 <comment username="juanjo1990" rating="10" value="Ubc. 3C Quizzes con FCM el
    mejor Splotter. Yo lo pondra por encima mecnicamente, pero la pifia en la
    produccion es imperdonable. Edit: Se suma al podio The Great Zimbabwe"/>
108 <comment username="ipser" rating="10" value=""/>
109 <comment username="VictoryRobins" rating="10" value=""/>
110 <comment username="Citery" rating="10" value="This is a sobering reminder of
    all the great, singular designs that may have been lost to the maw of 18xx.
    Here is a game that clearly borrowed heavily from the framework, but is
    so adventurous and self-assured as to stand firmly in its own category.
    The merger system gives Indonesia a dynamism that I have found in few
    other economic games; companies change hands constantly as mergers are
    resolved for equally offensive and defensive reasons, and player positions
```

become delightfully fluid. The parasitic, catch-22 relationship between producers and shippers is fascinating, and never seems to play out exactly the same way twice. Im a little sad I havent seen more of this asymmetric interdependency in other economic games. I will admit that out of the box, Indonesia is (while undeniably beautiful) completely unusable as a game object. The map is two sizes too small, the wooden pieces and cardboard ships are two sizes too big, and tracking city demand with a pile of chits is untenable. Most of the components need straight up replacing, and that 's a big ask for a game this expensive."/>

111 <comment username="obo\_bb" rating="10" value="Fantastic design. Merger and resulting from it takeover mechanics are here really unique. Yes, you can lost your company, but you can use these money to takeover somebody else enterprise. Final round has doubled income and the winner is the player with most money, not with most and best companies! Yes, it cause a lot of interesting situations. Typical Splotter - very simple rules, deep gameplay and interaction. PnP Version - fan's maps and goods management implementation is better that in inaccessible original - I recomment Luke Weisman variant."/>

112 <comment username="jambudvipa" rating="10" value="Possible PnP new board to make. So much potential - every option is invaluable - long term strategies develop through tactical reactions to other players choices as much as your own. Triggering mergers can reward you with ownership of companies that you previously lost in auctions. "/>

113 <comment username="villiman" rating="10" value=""/>

114 <comment username="fcboyd" rating="10" value=""/>

115 <comment username="mintiful" rating="10" value=""/>

116 <comment username="TheBardGamer" rating="10" value=""/>

117 <comment username="11xor6" rating="10" value=""/>

118 <comment username="Rankus" rating="10" value=""/>

119 <comment username="odie1973ro" rating="10" value="I am the KING of INDONESIA ....a wait a minute. I forgot to expand my cargo Hull....NO..... =@ Can't ship, and BANKRUPT... AWESOME GAME"/>

120 <comment username="Gerth13" rating="10" value=""/>

121 <comment username="fteplin" rating="10" value=""/>

122 <comment username="Cheesauce" rating="10" value=""/>

123 <comment username="boboyo" rating="10" value=""/>

124 <comment username="jamcco4" rating="10" value=""/>

125 <comment username="massibull" rating="10" value=""/>

126 <comment username="SushiMooshi" rating="10" value=""/>

127 <comment username="ManchasBta" rating="10" value=""/>

128 <comment username="ducknuts" rating="10" value=""/>

129 <comment username="TalesWithinWheels" rating="10" value="Incredible game of

```
    economic manipulation. An absolute favorite and one I see myself returning
    to for the rest of my life. The tricks you can pull off in this are so
    satisfying."/>
130 <comment username="tosza1111" rating="10" value=""/>
131 <comment username="quartz54" rating="10" value=""/>
132 <comment username="spiderM9" rating="10" value=""/>
133 <comment username="instinctive" rating="10" value="had to make my own!"/>
134 <comment username="solidblade" rating="10" value=""/>
135 <comment username="cowboyfringant" rating="10" value=""/>
136 <comment username="arceushui" rating="10" value=""/>
137 <comment username="Hzzon_73" rating="10" value=""/>
138 <comment username="dragonseto" rating="10" value=""/>
139 <comment username="Kalii" rating="10" value=""/>
140 <comment username="KoRnfr34k" rating="10" value=""/>
141 <comment username="faizali" rating="10" value=""/>
142 <comment username="Ghellestad" rating="10" value=""/>
143 <comment username="NeonFire" rating="10" value=""/>
144 <comment username="Mandurang" rating="10" value=""/>
145 <comment username="Avicheeta" rating="10" value=""/>
146 <comment username="Jonathan York" rating="10" value=""/>
147 <comment username="Spitlock" rating="10" value=""/>
148 <comment username="geniusss" rating="10" value=""/>
149 <comment username="oRd3s" rating="10" value=""/>
150 <comment username="Knacks" rating="10" value=""/>
151 <comment username="Viniciuszen" rating="10" value=""/>
152 <comment username="rbaba12" rating="10" value=""/>
153 <comment username="Agente42" rating="10" value=""/>
154 <comment username="ajaykrishnan123" rating="10" value=""/>
155 <comment username="digiclam" rating="10" value=""/>
156 <comment username="ihtiadrenko" rating="10" value=""/>
157 <comment username="EthanolCataclysm" rating="10" value=""/>
158 <comment username="LoneWolfVC" rating="10" value=""/>
159 <comment username="z4zinter" rating="10" value=""/>
160 <comment username="extatix" rating="9.9999" value="This game blew me away. It'
    d be a solid 10 if not for the components. And that font."/>
161 <comment username="JohnnyOffice" rating="9.9" value="Great game. With the
    right players, this is a 10! Great auction mechanic, the research tracks
    are perfect for the game (you want to do everything but really can't) and
    who can hate microwave meals?!?"/>
162 <comment username="baughmm" rating="9.9" value="Almost a 10! If I could play
    it more than once a year it might be!"/>
163 <comment username="larczz" rating="9.9" value=""/>
```

```
164 <comment username="Khrysto" rating="9.9" value=""/>
165 <comment username="Thecostcutter" rating="9.9" value=""/>
166 <comment username="codewizpt" rating="9.81" value="Kallax: D4"/>
167 <comment username="fs1973" rating="9.8" value=""/>
168 <comment username="randomaccess" rating="9.8" value="Components are ok. But
    excellent play for a heavy game. Two players makes for a very heavy game,
    compared to more players."/>
169 <comment username="Grayfox IT" rating="9.8" value=""/>
170 <comment username="zeamadevarza" rating="9.8" value=""/>
171 <comment username="s83m" rating="9.8" value=""/>
172 <comment username="eleanti" rating="9.8" value=""/>
173 <comment username="IlReb" rating="9.8" value=""/>
174 <comment username="1alexanderhuemer2" rating="9.8" value="I loooove the
    oversized wooden goods ;))"/>
175 <comment username="Sepsy" rating="9.8" value=""/>
176 <comment username="Atus" rating="9.8" value=""/>
177 <comment username="Ice X" rating="9.75" value=""/>
178 <comment username="Mymil" rating="9.75" value=""/>
179 <comment username="rbayu2" rating="9.75" value=""/>
180 <comment username="0bi_Ita" rating="9.71" value=""/>
181 <comment username="The Prophet" rating="9.7" value="Intense economic
    experience. To merge or not to merge... That! is the question. 2-5 players
    "/>
182 <comment username="fokos" rating="9.7" value="Really Great! It might be too
    slow but we had been playing for 4-5 hours and we didn't even realise it!
    An economic game with tones of things to admire,the mergers, the
    difference between companies,the great timing that must be in everything,
    really eager to play it again!"/>
183 <comment username="Hadgoth" rating="9.7" value="PnP"/>
184 <comment username="LeBao" rating="9.7" value=""/>
185 <comment username="Brave Sir Robin" rating="9.6" value="Not for everyone, due
    to the screw-your-neighbor merge mechanic. If you like economic games, it'
    s hard to beat this one!"/>
186 <comment username="lukedenby" rating="9.6" value="Even with the math I love
    the acquisitions and mergers of companies. One of the most interesting
    scenarios in gaming. I wish I could get this to the table more often but
    it's such a time commitment."/>
187 <comment username="naarnold" rating="9.6" value=""/>
188 <comment username="Yllien" rating="9.6" value=""/>
189 </comments>
190 </item>
```

191 </items>

**Listing A.2:** Ejemplo API Response, Ratings Info.

## A2. Ejemplo de Almacenamiento intermedio - Datos Juego

A continuación, un ejemplo del formato de almacenamiento intermedio definido para la información de los juegos (tomando solo un ítem, cada archivo contiene la información de 200 ítems):

```
1 {
2   "name": "Die Macher",
3   "alternative_names": [
4     "\u5fb7\u56fd\u5927\u9009",
5     "\ub514 \ub9c8\uud5c8"
6   ],
7   "thumbnail": "https://cf.geekdo-images.com/rpwCZajYLD940
  NWwP3SRoA__thumb/img/YT6svCVsWqLrDitcMEtyazVktbQ=/fit-in
  /200x150/filters:strip_icc()/pic4718279.jpg",
8   "image": "https://cf.geekdo-images.com/rpwCZajYLD940NWwP3
  SRoA__original/img/yR0aoBVKNrAmmCuBeSzQnMf1LYg=/0x0/
  filters:format(jpeg)/pic4718279.jpg",
9   "description": "Die Macher is a game about seven sequential
  political races in different regions of Germany.
  Players are in charge of national political parties, and
  must manage limited resources to help their party to
  victory. The winning party will have the most victory
  points after all the regional elections. There are four
  different ways of scoring victory points. First, each
  regional election can supply one to eighty victory
  points, depending on the size of the region and how well
  your party does in it. Second, if a party wins a
  regional election and has some media influence in the
  region, then the party will receive some media-control
  victory points. Third, each party has a national party
```

```
membership which will grow as the game progresses and
this will supply a fair number of victory points. Lastly
, parties score some victory points if their party
platform matches the national opinions at the end of the
game. The 1986 edition featured four parties
from the old West Germany and supported 3-4 players. The
1997 edition supports up to five players in the re-
united Germany and updated several features of the rules
as well. The 2006 edition also supports up to five
players and adds a shorter five-round variant and
additional rules updates by the original designer.
;";
10 "yearpublished": "1986",
11 "minplayers": "3",
12 "maxplayers": "5",
13 "playingtime": "240",
14 "minplaytime": "240",
15 "maxplaytime": "240",
16 "minage": "14",
17 "link": {
18     "boardgameartist": {
19         "928": "Bernd Brunnhofer",
20         "12517": "Marcus Gschwendtner",
21         "4959": "Harald Lieske"
22     },
23     "boardgamefamily": {
24         "10643": "Country: Germany",
25         "81575": "Digital Implementations: VASSAL",
26         "34116": "Political: Elections",
27         "91": "Series: Classic Line (Valley Games)"
28     },
29     "boardgamedesigner": {
30         "1": "Karl-Heinz Schmiel"
31     },
32     "boardgamepublisher": {
```

```
33         "133": "Hans im Gl\u00fcck",
34         "2": "Moskito Spiele",
35         "24883": "Ediciones MasQueOca",
36         "2726": "Portal Games",
37         "15108": "Spielworxx",
38         "39249": "sternenschimmermeer",
39         "11652": "Stronghold Games",
40         "5382": "Valley Games, Inc.",
41         "8147": "YOKA Games"
42     },
43     "boardgamemechanic": {
44         "2916": "Alliances",
45         "2080": "Area Majority / Influence",
46         "2012": "Auction/Bidding",
47         "2072": "Dice Rolling",
48         "2040": "Hand Management",
49         "2020": "Simultaneous Action Selection"
50     },
51     "boardgamecategory": {
52         "1021": "Economic",
53         "1026": "Negotiation",
54         "1001": "Political"
55     }
56 },
57 "poll": [
58     {
59         "@name": "suggested_numplayers",
60         "@title": "User Suggested Number of Players",
61         "@totalvotes": "136",
62         "results": [
63             {
64                 "@numplayers": "1",
65                 "result": [
66                     {
67                         "@value": "Best",
```

```
68         "@numvotes": "0"
69     },
70     {
71         "@value": "Recommended",
72         "@numvotes": "1"
73     },
74     {
75         "@value": "Not Recommended",
76         "@numvotes": "86"
77     }
78 ]
79 },
80 {
81     "@numplayers": "2",
82     "result": [
83         {
84             "@value": "Best",
85             "@numvotes": "0"
86         },
87         {
88             "@value": "Recommended",
89             "@numvotes": "1"
90         },
91         {
92             "@value": "Not Recommended",
93             "@numvotes": "88"
94         }
95     ]
96 },
97 {
98     "@numplayers": "3",
99     "result": [
100         {
101             "@value": "Best",
102             "@numvotes": "2"
```

```
103         },
104         {
105             "@value": "Recommended",
106             "@numvotes": "26"
107         },
108         {
109             "@value": "Not Recommended",
110             "@numvotes": "76"
111         }
112     ]
113 },
114 {
115     "@numplayers": "4",
116     "result": [
117         {
118             "@value": "Best",
119             "@numvotes": "25"
120         },
121         {
122             "@value": "Recommended",
123             "@numvotes": "86"
124         },
125         {
126             "@value": "Not Recommended",
127             "@numvotes": "9"
128         }
129     ]
130 },
131 {
132     "@numplayers": "5",
133     "result": [
134         {
135             "@value": "Best",
136             "@numvotes": "116"
137         },
```

```
138         {
139             "@value": "Recommended",
140             "@numvotes": "11"
141         },
142         {
143             "@value": "Not Recommended",
144             "@numvotes": "2"
145         }
146     ]
147 },
148 {
149     "@numplayers": "5+",
150     "result": [
151         {
152             "@value": "Best",
153             "@numvotes": "1"
154         },
155         {
156             "@value": "Recommended",
157             "@numvotes": "0"
158         },
159         {
160             "@value": "Not Recommended",
161             "@numvotes": "62"
162         }
163     ]
164 }
165 ]
166 },
167 {
168     "@name": "suggested_playerage",
169     "@title": "User Suggested Player Age",
170     "@totalvotes": "31",
171     "results": {
172         "result": [
```

```
173     {
174         "@value": "2",
175         "@numvotes": "0"
176     },
177     {
178         "@value": "3",
179         "@numvotes": "0"
180     },
181     {
182         "@value": "4",
183         "@numvotes": "0"
184     },
185     {
186         "@value": "5",
187         "@numvotes": "0"
188     },
189     {
190         "@value": "6",
191         "@numvotes": "0"
192     },
193     {
194         "@value": "8",
195         "@numvotes": "0"
196     },
197     {
198         "@value": "10",
199         "@numvotes": "0"
200     },
201     {
202         "@value": "12",
203         "@numvotes": "6"
204     },
205     {
206         "@value": "14",
207         "@numvotes": "18"
```

```
208         },
209         {
210             "@value": "16",
211             "@numvotes": "4"
212         },
213         {
214             "@value": "18",
215             "@numvotes": "2"
216         },
217         {
218             "@value": "21 and up",
219             "@numvotes": "1"
220         }
221     ]
222 }
223 },
224 {
225     "@name": "language_dependence",
226     "@title": "Language Dependence",
227     "@totalvotes": "49",
228     "results": {
229         "result": [
230             {
231                 "@level": "26",
232                 "@value": "No necessary in-game text",
233                 "@numvotes": "37"
234             },
235             {
236                 "@level": "27",
237                 "@value": "Some necessary text - easily
                memorized or small crib sheet",
238                 "@numvotes": "5"
239             },
240             {
241                 "@level": "28",
```

```
242         "@value": "Moderate in-game text - needs
243             crib sheet or paste ups",
244     },
245     {
246         "@level": "29",
247         "@value": "Extensive use of text -
248             massive conversion needed to be
249             playable",
250         "@numvotes": "0"
251     },
252     {
253         "@level": "30",
254         "@value": "Unplayable in another language
255             ",
256         "@numvotes": "0"
257     }
258 ],
259 "statistics": {
260     "rating": {
261         "ratings": "5759",
262         "numcomments": "2094",
263         "average": "7.59802",
264         "bayesaverage": "7.04421",
265         "stddev": "1.56509",
266         "median": "0"
267     },
268     "weight": {
269         "averageweight": "4.315",
270         "numweights": "781"
271     },
272     "collection": {
```

```
273         "owned": "8105",
274         "trading": "274",
275         "wanting": "509",
276         "wishing": "2136"
277     },
278     "ranking": {
279         "ranks": [
280             {
281                 "@type": "subtype",
282                 "@id": "1",
283                 "@name": "boardgame",
284                 "@friendlyname": "Board Game Rank",
285                 "@value": "409",
286                 "@bayesaverage": "7.04421"
287             },
288             {
289                 "@type": "family",
290                 "@id": "5497",
291                 "@name": "strategygames",
292                 "@friendlyname": "Strategy Game Rank",
293                 "@value": "237",
294                 "@bayesaverage": "7.17912"
295             }
296         ]
297     }
298 }
299 }
```

**Listing A.3:** Ejemplo JSON File, Game Info.