

## Q6 — Recommender Systems (12 points)

### Setup

Book	Length	Genre	Author	Total Ratings
A	72k	Sci-fi	A1	100
B	108k	Sci-fi	A1	1000
C	383k	Fantasy	A2	600
D	40k	Mystery	A3	40
E	52k	Mystery	A3	8
F	12k	Horror	A4	200

**User U1:** Longtime user interested in **long fantasy novels**. **Observation:** System R recommended **Book B** to U1.

### Part 1 (4 points) — Which System(s) Could R Be?

User-user CF and/or Item-item CF

System	Could it be R?	Reasoning
Content-based	No	U1's profile = {long, fantasy}. Book C (383k, Fantasy) is the best match. Book B (108k, Sci-fi) doesn't match on genre or relative length.
User-user CF	Yes	Finds users with similar rating patterns to U1. If those users liked B, it gets recommended — CF is blind to content features.
Item-item CF	Yes	Finds books with similar rating vectors to U1's liked books. B could have similar rating patterns to U1's favorites despite different genre.

**Additional signal:** B has 1000 ratings (most popular), which collaborative filtering methods tend to surface due to popularity bias.

## Part 2 (4 points) — Content-Based Rec for New Sci-Fi User U2

Either Book A or Book B

U2 is a **new user** who wants science fiction. A content-based system matches the user profile (genre = Sci-fi) against item features:

- **Book A:** 72k, Sci-fi, A1 → genre matches ✓
- **Book B:** 108k, Sci-fi, A1 → genre matches ✓
- All other books: wrong genre ✗

Since U2 has no rating history, the system can only filter by the stated preference. Both A and B are equally valid sci-fi recommendations. (No further information distinguishes them for this user.)

---

## Part 3 (4 points) — Genre-Diverse Top-3 via Item-Item CF

**Goal:** Use item-item CF to recommend top 3 books while incorporating genre diversity.

**Approach:** Run item-item CF with  $k \gg 3$  to produce a full ranked list of predicted books. Then apply a genre-diversity selection:

1. Walk down the ranked list from highest to lowest predicted rating.
2. Select each book only if its genre hasn't already filled its "quota" (e.g., at most 1 book per genre).
3. Stop after selecting 3 books.

This ensures the final 3 books come from different genres while still respecting the CF ranking within each genre.

**Alternative valid approaches:**

- Run item-item CF separately for each genre (sci-fi, fantasy, mystery, horror) and take the top-1 from each genre, then pick the best 3 across genres.
- After generating the full CF ranking, apply a re-ranking step that optimizes a combined objective of predicted rating + genre diversity.

**Key constraint:** The answer must be based on item-item CF — not switching to a different recommendation paradigm.