

Personal Statements / Inherent Motivations

| | | | | |
|---|--|---|-----------------------|--|
| Car Insurance Expand ▾ | Car Spec Comparisons + Price Expand ▾ | Computer Spec Comparisons + Price Expand ▾ | Bulk Image Translator | Book Finder <ul style="list-style-type: none">- Genre- Catalog- Library- Search |
|---|--|---|-----------------------|--|

I read a lot of books in the past. Every Wednesday, my mom will take me to the library, and I will always choose 6 or 8 books to take home and read, finishing them all by next week. Sadly, I don't remember all the titles of the books I read, and this frustrates me because some of the books were really fun and I want to read them again. I sometimes get lucky and through key word searching, I can find the book from Google, but sadly this isn't always the case. Nowadays I have been reading off of digitized novel sites, web novel sites, and web story sites. These sites are nice but their searching algorithm could be lacking. It's nice that they have a vast amount of genres and tags to help narrow my searches, but some sites have more tags and genres than others. Some of the stories are not properly tagged properly, either they miss obvious genres or tags, or they are under the wrong genres and tags. This can occur because there is no proper definition for some of the tags, or the site offers freedom through personal tag creation, which creates redundancy within the site. Also some sites are just way too limited in their amount of tags and genres available.

This paper will cover the architectural technologies necessary for a Book Finder service, which is a critical subset of modern digital library and publishing systems responsible for maximizing content discoverability and user engagement. In such a Book Finder operation, metadata supplied by publishers and librarians would be processed and curated with the expectation that there would be a precise and rapid retrieval upside to any user's query. You will see Book Finder operations inside of large retail companies, where they organize extensive digital catalogs, or in university-like engagements where patrons provide specific search parameters, and these organizations leverage billions or at minimum hundreds of millions of data points within catalog operations to predict a user's intent. This is typical across digital information services, wherein multiple machine learning services must coordinate with other more traditional relational library applications in order to predict user genre preferences and short-term reading interests.

Points of Pain / Deficiency Analysis

Bulk Image Translator

- Sign-in Required
- Not very user friendly interface
- Unclear data (Vague, Ambiguous)
- Ads
- Sponsors take priority (Bias)

Book Finder

- It's hard to find books based on what you remember
- Summary searching
- Either too broad and not narrow enough or can't find it based on your criteria
- Not enough books
 - Either digital (webnovels) or published
- Tags and genres can be misleading, wrong, or missing
 - **CHOOSING TO FOCUS ON THIS**
- AI can easily summarize data and use your words as reference to find what you are searching for
- Some websites stop bots from entering through cloudflare
 - Prevents easy access to their data (Data mining)
- Redundancy in titles
 - People name their books very similar to others
 - They like to describe books differently
 - Hard to find specific good sources

As I read more and more novels online, one thing I have noticed, and use frequently, was the tag system. The tag system is usually split into main tags (usually popular genres such as action, tragedy, mystery, etc.) and subtags which are more specific niches (such as reincarnation, level system, special abilities, weak to strong, etc.). The tag system helps readers such as myself find novels that better suit our taste. A flaw in this system is that there is no universal tagging system. One website may have one tag while another doesn't. Another one is the confusion regarding tags. The lack of some specific tags causes some authors to post their novels under tags that may not necessarily suit their novel themes. Some websites try to counteract this by adding definitions to their tags, which helps in reducing confusion for both the reader and the author. Another issue with the tagging system is when some websites allow for the users to create their own tags. This solves the issue with the lack of suitable tags but it also causes redundancies since the website usually does not screen out similar tags that should instead all be under one tag. One issue this affects a lot is fanfiction tags. When they add what media that they are writing fanfiction for, there are a lot of different names one could call the media, either through popular

acronyms of the show or maybe the full title of the show as well (Ex. [Title of Media] vs. [Media Title] vs. [ToM]. This can be frustrating to find specific novels that you would like to read since you need to think of all the different ways people may put their novels under which can be hard to account for.

Frequently, modern Book Finder systems must interface with multiple content providers, ranging from large digital library archives to independent retail catalog systems. This multiplies the data ingestion workload and increases the number of services needed by any given book discovery platform. Hence, in order to operate, a company must initiate dozens or even hundreds of services to communicate with each respective source, which may also be protected by barriers like Cloudflare that prevent easy data access (data mining). This creates complexity, in that every content feed can have errors, missing entries, misleading tags and genres, and other data points. Thus, the integrity of the data on every server is critical to providing value. Management cannot operate unless they pool and coalesce this concert of data in one place to address the pain point of hard-to-find books based on memory or criteria that are too broad or not narrow enough. This necessitates real-time information to power AI systems that can summarize data and use the user's words as a reference for summary searching, and requires management to receive a daily update on search performance. This is due to the large amount of user frustration at stake, the high-volume nature of daily search queries, and the need to properly organize a vast inventory of both digitally-published and traditionally-published books.

SWOT Modeling - For the Tagging System

(SWOT of what happens if problem is solved)

| | |
|---|---|
| Strength: It will be easier for readers to find novels. More tags. Less redundancy. | Weaknesses: Decreased views of some novels. Too many tags. |
| Opportunities: Filtering is much easier. Larger flow of readers. | Threats: Certain novels can get quickly buried. Oversaturation of readers (Pushy). |

If the pain point of tags in novel sites was solved, this would vastly improve the filtering, and searching function of the site. It would be much easier for me and others to find novels that suit their taste, to explore different niches that people write about. There are some threats that can arise from this improvement, however. One is related to a common issue with reviews and ratings. Reviews and ratings can subconsciously affect how a reader views a novel, they are most likely willing to not give a novel a chance if it has a lower rating, even if the content itself is really good and creative. Tags can function similarly. Some authors deliberately not include certain tags that they understand may turn off the general reader demographic, in order to draw them in and slowly acclimate them to their writing. This is a common tactic that although can be viewed as scummy, is also a great way to help expand a reader's viewpoint and tastes in novels. An improved tagging system may bury these novels that threaten to broaden a reader's horizons, all because filtering is much better. Another threat is the oversaturation of readers. Some authors like to specialize, targeting specific readers. They like to stay in the background, only catching the attention of readers that share the same view as the author. A better tagging system could lead to a new influx of readers to their novels, ruining their experience with writing. I have seen many authors that succumbed to the pressure of a massive flow of fame (due to whatever algorithm) that leads to the novel either going into hiatus or becoming abandoned completely. Despite this, I feel that the strengths outweigh the weaknesses if the tagging system is improved, more tags can still help to expand what readers are comfortable with. By knowing that such options exist, readers may feel curious to explore what those tags pertain to. Certain sites help to define the tags so when I see a tag that I don't know, I like to click it and explore the novels that are tied to it.

Critical Success Factors

1. Advanced Search with More Filtering Options
 - Allows readers to quickly and easily find novels
 - Promotes their interests, and the exploration of new ones
2. Tag Clarity (Specification and Defined Properly)
 - Eliminate tag confusion, redundancy, and inconsistency
3. Encourage the Viewing of New Novels
 - Encourages new reading tastes

No system is perfect, there will always be flaws. A system that works great for one may be horrible for another. This is natural since we are all different. I have experienced this in a multitude of websites where I find the tag system they provide horrifying, they are confusing, hard to navigate, and even lacking in content, but for others, the tags provided are the perfect amount (Mostly because they are satisfied with the site only providing the tags for popular genres and none of the specifics). It contains the types of filters that make them happy. What's hard about adding more tags to a system is that some tags eventually become buried. People forget about story archetypes, they move on to new fads, they construct their stories off of new ideas or twists of old ones. It would be nice to make readers go through the entire tag list, being able to see definitions and examples for each one in order to expand their palette and tastes but this isn't a feasible option. Not everyone is willing to do so. This means that novel tagging only gets the popular tags, the tags that the authors and readers are most comfortable and familiar with, causing people that only search for novels through niche tags to not have these novels in their radar at all, even if the content of the novel contains the niche they are craving. I feel a LLM could help with this. By setting definitions to each tag and allowing the addition of new ones. I would like the tags to be accurate to the content of the story, but to achieve this through human work is tedious and annoying. I don't want to sit down and read a whole story just to find the specific tags to label it. An AI could help to speed up this process since they are able to digest the entire content of the story incredibly quickly and add the appropriate and delete the inappropriate tags associated with the novel. By putting the correct tags to a novel, it will help the readers who search for novels through niche tags to finally get more accurate searching and finding of new material to read. Furthermore, readers that have smaller palettes are able to experience these niche tags through the recommendation of their favorite novels. Basically, if the reader likes a story, they are more likely to look at the different tags that the story contains, and search for similar novels with those tags. The addition of tags that authors may otherwise overlook due to a lack of information, will help the readers to finally learn about these new genres. It will also help the author to enrich their stories by focusing more on the different archetypes that the readers seem to enjoy. More novels, more user interactions, will all promote more content for me to read which is what I consider a critical success factor. Another thing is that the system needs to properly be able to understand how there can be redundancies in the tags. This can again be corrected by AI, and with proper definition of each tag, the chance of

redundancy can hopefully be decreased. This helps with a common issue I have which is not being able to find novels based on certain tags due to people listing certain genres under other names. This is all because of misinformation that leads to their own incorrect personal definition of a tag. This is common in fanfic genres where (like mentioned above), people name their favorite works of fiction (tv shows, movies, animes, books, etc.) differently depending on what they are comfortable with, which although I can understand, I think it is wrong. Instead the name should be strictly formatted into one name only (the full title so anyone can understand what the work the name is linked to) instead of acronyms or shorthanded or whatever way people do it. AI can differentiate between all the different types because they are able to absorb the content of the author's writing as well as their wider range of understanding for the different nuance ways people refer to different works which will be much more helpful than a human doing the sorting themselves. Anyway, basically my critical success factor is a system that can help to finally catalogue all the novels in a neat and tidy way, one with actual order and structure so that I can more easily find content to consume.

Please understand that since the issue and my passion all relate to novel sites and how they are structured (in terms of their tagging/genre system), these references pertain more to how books can be sorted (by genres) and less about the websites themselves. Those are more informal from blogs and stuff and I didn't know if it was ok to use that.

References

- Carman, N. (2009). *LibraryThing tags and Library of Congress subject headings: A comparison of science fiction and fantasy works*. ResearchGate.
https://www.researchgate.net/publication/43950692_LibraryThing_Tags_and_Library_of_Congress_SubjectHeadings_a_Comparison_of_Science_Fiction_andFantasy_Works
- Eang, C., & Lee, S. (2024, June 18). Improving the Accuracy and Effectiveness of Text Classification Based on the Integration of the Bert Model and a Recurrent Neural Network (RNN_Bert_Based). *Applied Sciences*, (Special Issue: Natural Language Processing: Novel Methods and Applications), 18. <https://doi.org/10.3390/app14188388>
- Ellis, L. (2022, November 29). *Genrify your Catalog, not your Collection!* Knowledge Quest.
<https://knowledgequest.aasl.org/genrify-your-catalog-not-your-collection/>
- Julian, A., Ramyadevi, R., & Ayshwarya, V.R. A. (2025, February 26). Enhancing Book Recommendations Through Genre Prediction. *Lecture Notes in Networks and Systems*, 1182(International Conference on Recent Trends in Machine Learning, IOT, Smart Cities & Applications), 355-367. https://doi.org/10.1007/978-981-97-8865-1_30
- Kuo, T.-L., Chiu, T.-W., Lin, T.-S., Wu, S.-Y., Huang, C.-W., & Chen, Y.-N. (2024, June 4). A Survey of Generative Information Retrieval. *arXiv*. <https://arxiv.org/html/2406.01197v2+>