

WriteRec Annotation Guidelines

V2.2 // 4.07.2016

Ariella Levine, Clay Riley, Patricia Whitlock

Introduction to the Task

Current book recommendation systems are predicated on features external to the books themselves, making predictions about books a user might enjoy without regard to the contents of the books themselves. These simple systems cluster books that are often purchased together, or that users often rate similarly. They cannot account for the most compelling motivator for a user to actually become interested in the book: its blurb, or plot summary.

This project aims to create a book recommendation model that accounts for the characters and their relationships as they are explored in plot summaries. As a proof of concept, this report is limited to a specific genre, “Heroic Fantasy”, which is a sub-genre of fantasy which chronicles the tales of heroes in imaginary lands.

“Frequently, the protagonist is reluctant to be a champion, and/or is of low or humble origin, may have royal ancestors or parents but does not know it. Though events are usually beyond their control, they are thrust into positions of great responsibility where their mettle is tested in a number of spiritual and physical challenges.”

https://en.wikipedia.org/wiki/Heroic_fantasy

Your task will be to identify the characters and their relationships as stated in the plot summary of a given book. The annotations that you create will be built using the standoff annotation tool Multi-document Annotation Environment and used in a machine learning algorithm to group books on the basis of the similarities between their characters’ conflicts and friendships. You will be directed to specific book entries in book review website GoodReads in order to obtain the plot summaries over which your annotation will be written.

Resources for Annotators

The Multidocument Annotation Environment version 2.0, or MAE v2.0, is a graphical interface for adding annotations over a text. Annotations produced in MAE can be saved in several ways, including XML files. MAE can be obtained at the following location:

<https://github.com/keighrim/mae-annotation/wiki/getting-started>

You will be provided with a package containing these guidelines, a .dtd file to build your annotations in MAE, and other documents that you will need to complete your annotation.

It will be useful to have a basic understanding of English parts of speech, although you should not need an in-depth meta-knowledge of English grammar. You will need to be able to separately consider the information that is explicitly given and that which you are able to derive from implications in the text. Both are important to this task, but they serve different purposes.

Annotation Instructions

General Rules

Throughout the annotation process, you must follow these basic guidelines:

1. Do not tag within direct quotes from the book.
This information is not considered to be part of the book's plot. However, it may contain information that you may find useful in classifying a character's type or subtype.
2. Ignore information that is not part of the book's narrative.
Information about the publishing of the book, the author, or the series that the book is a part of are all irrelevant to its plot.
3. When tagging a span of text, do not include whitespace around the relevant words or punctuation that is stylistic instead of lexical.
It is important that all annotators select the same span for a given tag. Be careful not to include outer spaces, commas, periods, or the like, unless one of these is part of the actual word(s) being tagged, like the period after an abbreviation.
4. Do not use personal knowledge about the story to make decisions.
If you have read a plot that you are annotating, you must try to avoid injecting your own knowledge about the story and rely only on the summary given.

Entity Identification

Tag characters, groups of characters, and important objects that are explicitly named in the summary. Each entity will be tagged **only once**, no matter how many times it is referenced in the text. To decide what to tag for each entity, follow the sequence below. Once an item in the sequence has been found for that entity, tag accordingly and start the sequence over for the next entity.

If an entity is implied, but never explicitly written about, it won't get any tag.

1. Select exactly one portion of the text that identifies each entity, following these ordered rules to select the correct textual span:
 - a. Select the **first full-name mention** of the entity (including any immediately adjacent **capitalized titles**, words indicating their **possessors**, or **definite designations** like "the" or "this", **descriptions starting with "of"**, or **definite appositives**) if one exists;
e.g. **[Frodo Baggins]**, a hobbit of the Shire
e.g. the corrupted **[Forest of Flin]**

- b. Otherwise, select the **first named mention** of the entity (including any immediately adjacent **capitalized titles**, words indicating their **possessors**, or **definite designations** like “the” or “this”, **descriptions starting with “of”**, or **definite appositives**) if one exists;
 - e.g. *[Joel] wants to be a Rithmatist*
 - e.g. *Beside her is [Brand], a young warrior*
- c. Otherwise, select the **first capitalized title** (including any other immediately adjacent **capitalized titles**, words indicating their **possessors**, or **definite designations** like “the” or “this”, **descriptions starting with “of”**, or **definite appositives**) if one exists;
 - e.g. *but [the King] had other plans*
 - e.g. *humanity’s only defense against [the Wild Chalklings]*
- d. Otherwise, select the **first nominal descriptor** of the entity (including any immediately adjacent **capitalized titles**, words indicating their **possessors**, or **definite designations** like “the” or “this”, **descriptions starting with “of”**, or **definite appositives**) if one exists;
 - e.g. *three young [survivors of vicious demon attacks]*
 - e.g. *a merciless [army] will cross the water*
- e. Otherwise, select the **first pronominal or pronoun mention** of the entity (including any immediately adjacent **capitalized titles**, words indicating their **possessors**, or **definite designations** like “the” or “this”, **descriptions starting with “of”**, or **definite appositives**) if one exists;
 - e.g. *[THEY] KILLED THE KING*
 - e.g. *to destroy [all] that was good in Middle-earth*

Groups of entities should be tagged, but only if they do not consist **exclusively** of a group of entities that are tagged individually elsewhere in the text. Nations, realms, classes, peoples and the like should be labeled as groups.

- e.g. *[Hobbits] are a very domestic people*
(A newly named group is tagged)
- e.g. *[Sam] and [Frodo] must pass through the mines; will the hobbits survive?*
(A group consisting entirely of two tagged characters is not tagged)
- e.g. *the Dark Lord will unleash his wrath upon [all of Middle-earth].*
(A new pronoun-identified group of nations with an of-phrase is tagged)
- e.g. *burn with a consuming hatred of [humanity]*
(A newly named peoplehood is tagged)

Metonymy is when a word referring to an object, place, or attribute is used to stand in for a person or group of people: for example, using “the White House” to refer to the

administration of the President of the United States. In this task, you may encounter metonymy when the name of a country is used to refer to the people of that country. If you suspect that this is the case, carefully consider whether or not the phrase “the people of [country]” could be **substituted** for the name of the country. If the text still makes sense, **tag the name of the country** (and any of the relevant adjacent descriptors given above).

2. Mark the **number** of the selected entity. This may be one of:
 - a. **Groups** are an entity consisting of more than one person or item. If a tagged entity is plural, it should be marked as a group.
 - b. **Individuals** are single people or items. This is the default value for an entity.
3. Choose a **type and subtype** for the selected entity. These classifications must be inferred from the information in the plot summary, not from outside sources. Entities may be one of:
 - a. **Protagonists**: the main character(s) of the story with whom the reader is expected to identify, and their allies. They are usually, but not exclusively, identifiable as the agents of some abstract “good” or “justice” in a given story. Protagonists must be further identified as one of the following:
 - i. **Underdog Hero(in)es** are primary protagonists who are explicitly described as having characteristics that make them unlikely to be especially courageous or heroic. These are main characters who are typically physically small, young, or weak, and are described as overcoming significant obstacles to be seen as a worthy opponent to the antagonists, or as a valuable member of the protagonists’ team.
 - ii. **Traitors** are protagonists, usually not the main protagonist, who are explicitly described as being former antagonists who have joined forces with the protagonists for some reason.
 - iii. **Antihero(in)es** are protagonists who are explicitly described with negative ethical characteristics or properties. Their sense of morality is often murky and their actions may appear to be self-serving, but they are ultimately either the book’s main character or allied with the main character(s). They are often similar to traitors, but are not described in the summary as having formerly been on the other side of the main conflict. A typical antihero might be described as a thief, a jailbird, or mercenary.

- iv. **Other Hero(in)es** are primary protagonists who do not fall into the above two types. They may be under-described in the summary, or they may just be a well-rounded and powerful character with a strong moral character.
 - v. **Sidekicks** are usually secondary protagonists who are less glorious or powerful than the other main characters, whom they often serve. They are similar to underdog heroes, but are not the focus of the story.
 - vi. **Mentors** are usually secondary protagonists who train, teach, or guide the other (more primary) protagonists. They are often described as old, wise, or well-studied, and provide information and advice to support the other protagonists.
 - vii. **Other** protagonists are any allies of the other protagonists who are not described in enough detail to assign them one of the other subtype labels. These are usually not the primary characters.
- b. **Antagonists:** the character(s) of the story whose actions or intents place them in opposition to the protagonist(s). They are usually, but not exclusively, identifiable as the agents of some abstract “evil” in a given story, or are clearly described in terms of their greed for physical or political power. Antagonists must be further identified as one of the following:
- i. **Boss Bad Guys** are the primary antagonists that direct the actions of the main bad guys. The defining feature of a boss is that their own actions are not clearly seen, and they seem to operate mostly by way of instructions to other antagonists. They are often described as shadowy, intelligent, or masterminds.
 - ii. **Traitors** are antagonists who are explicitly described as being former protagonists who have joined forces with the antagonists for some reason.
 - iii. **Main Bad Guys** are the primary antagonists of the protagonists. They are the direct source of most of the conflicts facing the protagonists. They are typically very powerful or intelligent, but may not be the highest-ranking or most powerful member of the antagonists. They may be very similar to traitors, but are not explicitly described as having switched sides of the story’s main conflict.
 - iv. **Underlings** are usually secondary antagonists who are subordinates of one or more of the other antagonists. They are often weaker or less

intelligent than the others, but are rigidly defined as taking orders from them, regardless.

- v. **Others** are any allies of the above antagonists who are not described in enough detail to assign them one of the other subtype labels.
- c. **Objects**: artifacts, or other non-living structures or items that are explicitly identified in the summary that have some special purpose or power relevant to the narrative. Objects are represented by capitalized names or descriptions using concepts that do not exist in real-world, everyday English (like “his [orcish scepter]”) in the summary.
- d. **Others**: entities who do not clearly fit into one of the above categories. These may include characters who do not appear to clearly be on the side of either the protagonists or the antagonists; those who are explicitly described as neutral to the plot’s main conflict; or large groups such as kingdoms that typically cannot be cleanly labeled as entirely good or bad.
- e. **Unknowns**: entities that are unable to be assigned into one of the above categories. Their allegiance cannot be determined due to a lack of information in the summary.

Relation Trigger Identification:

Relations may include situations that place two or more characters in opposition or alignment with one another. Relations can also include events that represent the personal achievements or failings of a single character. Tag the conflicts, relationships, and interactions among characters following the steps below, keeping each textual span minimal by including only a single word.

The single word chosen should be the word that most fully encapsulates the relation’s meaning, usually a main verb, a participle, noun modifiers or predicates showing a relationship, and the like.

- ★ Sometimes a situation will not have occurred within the timeline of the story’s plot. It may be in the future, or it may only happen if something else occurs first. Triggers that represent **events in the future, intentions, or things that will happen if some other condition is satisfied** should be tagged. Such conditions should not be tagged.

Follow these steps:

1. Tag any verbs, descriptions, and other events that represent **interactions or relationships between characters**.
2. Tag any verbs, descriptions, and other events representing **internal or external struggles, victories, and failures** by characters against their environment or their own

(potential) shortcomings.

3. Mark each trigger as **belonging to one of four types**:
 - a. **Romantic** triggers indicate any romantic relation between characters.
 - b. **Familial** triggers indicate any familial relation between characters.
 - c. **Conflict** triggers indicate a conflict or struggle between characters.
 - d. **Other** triggers indicate a non-romantic, non-conflict, and non-familial relation between characters.

Relation trigger spans may overlap with the entity spans defined before in cases where a descriptive word is used both to show a relation and an identity (like “[his [lover]] had other plans”). Each one of these will be used as a trigger for one or more of the relations among characters that will be defined next.

Relation Linking

To show a relation among characters first create a Relation link tag. This is done by taking the following steps:

1. **Create a link tag** with no span selected. You will need at least as many links as trigger tags.
2. Select the trigger that **defines the relation**. A trigger may be used for more than one relation link, but a single link may only have one trigger.
3. Select the entity **from which the relation is caused or directed** and label it “From”. This is typically the agent or “doer” of a verb, or the entity in the relation that is implied to cause the relation to exist (either intentionally or unintentionally), or the character in the relationship that exerts an effort, or the source of the relation. Many relations will be perfectly symmetric; in these cases, the From entity should be the **first** one mentioned.
4. Select the entity **to which the relation is directed or given** and label it “To”. This is typically the patient or object of an action, or the receiver of something, or the passive member of a relationship, or the target or endpoint of a motion or action. These entities experience the action that defines the relation. Many relations will be perfectly symmetric; in these cases, the To entity should be the **second** one mentioned.

The To entity may be the same as the From entity, especially when the trigger represents **a struggle or action that shows personal internal growth, victory over the environment, or interaction with no particular partner**. See the FAQ for

examples.

5. In the link's menu, **choose the beneficiaries** of the relation. This is the linked entity or entities for whom the relation shows a positive or beneficial effect, or who gain something from the relation, or whose role in the relation shows some type of victory. Only this relation's linked parties should be considered when assigning this value. The beneficiaries may be one of:
 - a. **From**, the member of the relation who was identified earlier as the source of the relation.
 - b. **To**, the member of the relation who was identified earlier as the target of the relation.
 - c. **Both** the source and the target of the relation. The relation shows a mutually positive event or state of affairs in which both members grow or win.
 - d. **Neither** the source nor the target of the relation. The relation is neutral, described in too little detail to determine the beneficiaries, or simply doesn't help anyone.
6. In the link's menu, **choose the harmed members** of the relation. This is the linked entity or entities for whom the relation shows a negative or harmful effect, or who lose something due to the relation, or whose role in the relation show some type of defeat. Only this relation's linked parties should be considered when assigning this value. The harmed members may be one of:
 - a. **From**, the member of the relation who was identified earlier as the source of the relation.
 - b. **To**, the member of the relation who was identified earlier as the target of the relation.
 - c. **Both** the source and the target of the relation. The relation shows a mutually negative event or state of affairs in which both members suffer or lose.
 - d. **Neither** the source and the target of the relation. The relation is neutral, described in too little detail to determine the harmed parties, or simply doesn't hurt anyone.

There may be multiple relations identified by a single trigger. In cases where a group of characters should be at the endpoints of the link, **a separate link should be created for each pairing.**

Frequently-Asked Questions

- Q. *How do I know whether two characters are both heroes or a hero and a sidekick?*
- A. If the text does not make the characters' relative hierarchy in the group clear, then you may rely on the rule of thumb that sidekicks will typically receive less exposition within the plot.
- Q. *What is the difference between a traitor and an antihero? A traitor and any other form of antagonist?*
- A. Traitor classifications should be prioritized over other potential subtype classifications. They are unique in that their allegiance has shifted over time. This shift is a major generator of conflicts both between them and their new allies, who frequently mistrust them, and between them and their former allies, who may wish to bring them back to their side of the main conflict. As such, this is an important subtype to categorize for our purposes.
- Q. *Do I need to tag generic classes like "humankind", "teachers", "anyone", or "everyone"?*
- A. Yes. If any people or groups of people are mentioned in the plot, we'd like to know about them.
- Q. *How do I tag non-existent or unrealistic entities?*
- A. Follow the normal guidelines for these entities, which will usually be groups. These may appear as "no one", "none", "nobody who's anybody", etc.
- Q. *How do I know if a place is taggable as an object or not?*
- A. In general, don't tag places. Objects are things explicitly identified in the summary that have some special purpose or power relevant to the narrative, and these may sometimes be places. If the plot summary does not state that the object is special, magical, or otherwise needed by some character for some reason, then it should not be tagged. An altar where a demon can be summoned is an object, while the altar someone was left at on their wedding day is not.
- Q. *What if a relation describes an event or state of affairs between two parties, but benefits a third party?*
- A. If the benefit is implied, but not stated, then ignore any unlinked beneficiaries. If the benefit is explicitly stated, then an additional relation should be created between the benefitting parties and the participants in the relational event. The same goes for harmed parties.
- Q. *When does an object need to be tagged?*
- A. Tagged objects should ideally be plot devices (things with significant power or of interest to the characters and their relationships), rather than everyday tools used by the characters. Objects used in an unusual way should be tagged (e.g. a ring that turns the

wearer invisible, a sword that is a key) while objects used in a natural way should not (e.g. a sword that is used to stab things).

Q. *How do I know if someone is harmed or benefits in a relation?*

A. Someone is “harmed” in a relation if it introduces some difficulty, conflict, hardship, offense, or strain on them. Someone “benefits” if the relation seems to make some task or goal easier for them, or if it gets them something that they seem to want. A romantic relation typically benefits both parties, unless it is made clear that this isn’t the case. A conflict relation typically harms one or both parties.

Q. *I tagged another story in this series that gave evidence that character X should be subtype Y. Can I use that knowledge since it came from the annotation task?*

A. No. Please rely only on the individual summary you are currently working on!

Q. *When does a relation link a character to themselves?*

A. This type of link is used when a character does some action that primarily affects or pertains to themselves, elements and creatures of their environment, or someone who is left unstated. A difficult mental or physical task, choice, or challenge that does not directly affect another tagged entity falls into this category.