

# A platform for crowdsourcing corpora for argumentative dialogue

Catalina Popescu  
Imperial College London

Oana Cocarascu  
Imperial College London  
oc511@ic.ac.uk

Francesca Toni  
Imperial College London  
f.toni@ic.ac.uk

## ABSTRACT

One problem that Argument Mining (AM) is facing is the difficulty of obtaining suitable annotated corpora. We propose a web-based platform, BookSafari, that allows crowdsourcing of annotated corpora for *relation-based AM* from users providing reviews for books and exchanging opinions about these reviews to facilitate argumentative dialogue. The annotations amount to pairwise argumentative relations of *attack* and *support* between opinions and between opinions and reviews. As a result of the annotations, reviews and opinions form structured debates which can be understood as bipolar argumentation frameworks. The platform also empowers annotations of the same pairs by multiple annotators and can support different measures of inter-annotator agreement and corpora selection.

## CCS CONCEPTS

• **Software and its engineering** Application specific development environments;

## KEYWORDS

Argument Mining

## 1 INTRODUCTION

Argument Mining (AM) is a relatively new research area which involves the automatic detection in text of arguments, argument components, and relations between arguments (see [12] for an overview). AM is a complex task for humans and machines alike because of the lack of clear argumentative structures in free natural language text.

One of the hurdles AM is facing is the difficulty of obtaining suitable annotated corpora to train supervised machine learning models to distinguish argumentative and non-argumentative text and to predict argument components and relations between arguments. This is caused by the fact that this annotation task is generally difficult for humans (e.g. see [8] for an overview of some efforts in annotation). Annotations are usually obtained by some form of tailored crowdsourcing but are relatively small and/or have a low inter-annotator agreement. Alternative, more engaging techniques for obtaining larger and better corpora may be beneficial, as advocated by [9]. We propose a web-based platform that can support the creation, maintenance and continuous growth of annotated corpora for *relation-based AM* [3, 7, 14], focusing on identifying arguments in text as well as dialectical relations of *attack* and *support* between these arguments. This form of AM can be used to extract, from text, *Bipolar Argumentation Frameworks (BAFs)* [5], which in turn can be used to support further tasks (e.g. analysing debates [2] or spotting deceptive reviews [6]).

Our proposed web-based platform, *BookSafari*, allows users to provide reviews for books and to debate the reviews in a structured manner, by providing opinions for or against reviews and other

opinions. We choose a *product review platform* as it is well-known that users are able to easily engage with this kind of platform, as with popular such platforms for example for movie reviews (e.g. IMDb, Rotten Tomatoes), hotel reviews (e.g. booking.com, TripAdvisor), and product reviews (e.g. Amazon). We choose a platform for *book reviews* as, whilst there are platforms for book reviews (e.g. Goodreads), they do not reach the same level of popularity as their equivalent for other products.

BookSafari combines the capabilities of other book review platforms with those of discussion fora, allowing users to engage in debates about books. Dialectical relationships between reviews and opinions and between opinions expressed in these debates are stored as pairs so as to be usable as part of corpora for relation-based AM. Furthermore, users can express their opinions about relations identified by other users, as part of debates: these further opinions result in multiple annotations, that can be used to strengthen or weaken the membership of pairs in corpora extracted from BookSafari, depending on selected measures of inter-annotator agreement.

The platform allows visualisations of opinions for books by means of interactive trees with nodes as supporting/attacking arguments for/against other arguments which can be seen as explanations as to why the book is good or not. The annotated corpora can be used to build systems capable of engaging in a dialogue with users about books by expressing whether they liked a particular book or not and giving the reasons why by agreeing or disagreeing with arguments put forward by others.

The paper is organised as follows. In Section 2 we give essential background in relation-based AM, web-generated corpora creation for AM and debate fora. In Section 3 we describe BookSafari. In Section 4 we describe how BookSafari allows for multiple annotations of the same pairs. In Section 5 we describe how we have bootstrapped BookSafari using sentiment analysis on existing datasets. In Section 6 we conclude, discussing in particular challenges and future directions.

## 2 BACKGROUND

A comprehensive review on argumentation corpora was conducted by [8], which reports on an extensive set of works that created new corpora along with the argument model used, the domain it was conducted on, the size of the corpus, and a measure for the inter-annotator agreement, where existent. Most studies were conducted on reviews (i.e. hotel, car, camera), political and legal documents, social media contents (i.e. editorials, blog posts, tweets), essays and wikipedia pages. Most of the datasets available are relatively small (i.e. 500 Wikipedia pages, 8K sentences, 7K tweets).

In existing works, data creation and annotation rely mostly on crowdsourcing with Amazon Mechanical Turk. An alternative to this



Figure 1: Example BAF generated in BookSafari.

approach is to acquire data through serious games. In the argumentation landscape, [9] propose Argotario, a serious game that focuses on data acquisition and annotation of fallacious arguments.

Relation-based AM [3, 7, 14] aims at solving a 3-class prediction problem, whereby pairs of sentences are classified as being in an attack, support, or neither attack nor support relation. It is motivated by the desire to mine Bipolar Argumentation Frameworks (BAFs) [5] from text. BAFs are triples  $\langle AR, attacks, supports \rangle$  where  $attacks \subseteq AR \times AR$  and  $supports \subseteq AR \times AR$  are two relations over  $AR$ , a given set of abstract entities (the *arguments*). Existing corpora for this form of AM (e.g. [4, 14]) are quite small.

Several debate platforms exist. Most relevant to our work in this paper is Kialo<sup>1</sup>, that allows discussions to be visualised in the form of interactive trees with nodes as supporting/attacking arguments for/against other arguments. Kialo also allows arguments to be rated based on the impact they have on their parent argument (in the tree) and to give comments to ask for clarifications about arguments or to make suggestions. Our tree visualisations (see Section 4) are inspired by the ones in Kialo.

### 3 PLATFORM OVERVIEW

BookSafari is a client-server web-based application that runs in all modern browsers and works on other devices such as smartphones and tablets. Figure 1 shows an example of a debate generated in BookSafari. This corresponds to a BAF, with  $AR$  the set of nodes and links from a node of level  $L$  to its parent of level  $L - 1$  either an attack or a support, depending on the emoticon. Arguments in Level 1 include reviews for the book (at level 0) and represent argument supporting (thumb-up) or attacking (thumb-down) the default argument “The book *name of book* is worth reading”. From the view point of relation-based AM, arguments in Level  $L$  represent the children whereas arguments in Level  $L - 1$  represent the parents in the child-parent attack/support relation.

BookSafari features include: (i) finding new books (ii) browsing through reviews (iii) giving reviews (iv) engaging in debates with other users on books and their reviews (v) visualizing the debate flow in a tree-like structure with BAF at its core (vi) evaluating

others’ annotations of links between reviews and opinions expressed in the debates, and between opinions (vii) extracting corpora for relation-based AM.

BookSafari allows for two types of users: regular and admin, having different capabilities and uses of the above features. Regular users are able to create an account through the login page. On registration, the user is shown a disclaimer “By contributing to the debate in BookSafari you agree to your opinions being used for argument mining research purposes”. Each user has access to all functionalities and features of the book reviews platform. The user is able to personalize his/her profile page as well as to view other users’ profiles. The “Book page” is where reviewing, debating and annotations take place. Additionally, a regular user has access to the about page and his/her personalized home page. Admin users are envisaged as researchers interested in AM. Along with regular user capabilities, admin users have access to the admin view of the platform which comprises the “User” and “Post” table views. The most important service the admin has access to is the exporting feature, which allows data to be downloaded, along with the annotations, within a corpus.

#### 3.1 Views

There are four views available in BookSafari: the (regular) user view, the admin view, the book view, and the debate view.

Each regular user has the following features for his/her profile: (i) login page (ii) customizable profile page (iii) ability to have an avatar. The admin view is designed to allow researchers to access the BookSafari database. It offers views to the User and Post tables within the database. The Post table can be used to filter based on parent or child in the parent-child relation or to filter results based on the level the argument is in the debate. The admin users can also sort each column lexicographically by clicking on it. After applying all the desired filters, the admin users can export the dataset to a csv (comma-separated values) file. The request is sent back to the administrator of the platform via email. The administrator receives a token of seven days during which he/she is allowed to grant admin privileges to the user who requested credentials. If the administrator grants access to the user, then the user receives an email with instructions about how to access the admin account.

The book view is the most important part of the application, as it is where argumentation happens by reviewing and agreeing or disagreeing with the reviews or other opinions. The book view comprises information about the book: title, description, genre, followers, reviews score (doughnut chart on the left side) along with the percentages of positive vs negative reviews and a list of reviews and comments ordered by time of posting. In order to facilitate the view of the debate, the comments are displayed as a boxes-within-boxes (comment inside a comment) type of structure, similarly to the design used in other popular websites such as Reddit or microblogs. Adding reviews for a book can be done straightforwardly from the book page. After clicking one of the like/dislike buttons, a form will appear with the title “I liked the book because...” as seen in Figure 2a or “I did not like the book because...” as seen in Figure 2b.

The argumentation annotation is enabled through the attack comments as seen in Figure 3a and support comments as seen in Figure 3b. The same design as for adding reviews was followed, except

<sup>1</sup><https://www.kialo.com/>

Figure 2 shows two forms for adding reviews. Form (a) is for 'Like' and form (b) is for 'Dislike'. Both forms have a header with a green thumbs-up icon and a red thumbs-down icon. Form (a) has a green input field and a 'Submit' button. Form (b) has a red input field and a 'Submit' button.

Figure 2: Adding reviews for a book.

for the agree/disagree icons. A different format was preferred (i.e. smiley/sad face) to reflect the difference between like-agree and dislike-disagree functionalities. This section represents the core of the annotation as it allows the formation of child-parent arguments along with their argumentative relation annotation. For instance, if one clicks the agree button (i.e. smiley face) then he/she will be adding a support argument of the comment he/she is agreeing with.

Figure 3 shows two forms for agreeing or disagreeing with reviews. Form (a) is for 'Disagree' and form (b) is for 'Agree'. Both forms have a header with a green smiley face icon and a red sad face icon. Form (a) has a green input field and a 'Submit' button. Form (b) has a red input field and a 'Submit' button.

Figure 3: Agreeing or disagreeing with reviews of a book.

The debate view is what makes BookSafari stand out from existing book reviews platforms and discussion fora. It offers a comprehensive view of the debate flow of a discussion and aims to facilitate the visualization of a topic's pro and con standpoints.

The tree structure describes the dynamics of discussion of one review by showing its disagreeing comments on the left side (coloured in red) and its agreeing comments on the right side (coloured in green). These reviews may in turn have agreeing/disagreeing comments that are shown in the next level of the tree. Hovering over a node in the tree will provide a partial view of the comment at that node. In addition, the whole discussion flow can be observed under the tree in the form of a table-like system, identical to the one in the book view as seen in Figure 4. The advantage of this view is that it offers the possibility of following the discussion path. Figure

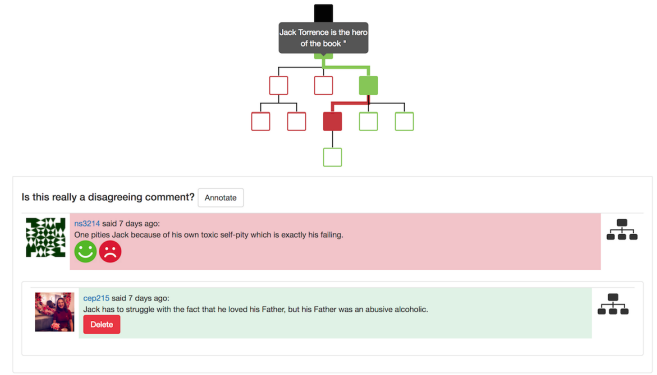


Figure 4: View of the tree when node is clicked.

4 shows what happens when a node from the penultimate level is clicked. The user is offered a view of the path until that particular comment (i.e. the selected comment is an attack to the agreement of a positive review). This is particularly helpful when the debate flow is complex and a user is interested in a specific argument, but at the same time wants to understand the context for it (i.e. by looking at the path it takes from the first review).

## 4 MULTIPLE ANNOTATIONS

For the annotation aspects of the platform (namely allowing users to provide feedback on existing annotations resulting from the debates) we wanted them to be engaging and visible but at the same time not to interfere with the other functionalities of the platform. The re-annotation section can be found in any view page that contains posts and comments (book view and debate view). BookSafari allows a user to contribute to the multiple annotation task by re-labelling an attack or support relationship from a review (to a Level 0 argument) or from a comment (to an argument of the preceding level).

Above each comment not already annotated by the current user, this user can find the question: "Is this really an agreeing/disagreeing comment?" as shown in Figure 5 or, for top level reviews: "Is this really a positive/negative review?". This question is followed by a button labelled "Annotate". After clicking this button, a small pop-up window is opened. For instance, if a user is trying to annotate a pair of arguments, the pop-up window will show the parent and child arguments and the user can then choose whether he/she agrees with the annotation (*Yes*), does not agree with it (*No*) or is not sure (*Not sure*), possibly because the arguments are not explicit. These annotations add to the original annotation coming from the user who authored comments/reviews.

## 5 BOOTSTRAPPING

Users tend to be more interested in using a web application that shows some level of maturity as well as popularity among other users. In order to make the platform appealing to users and thus create new corpora, we bootstrapped the platform with a relatively large set of data for books along with their labelled reviews.

We identified the following requirements, in order of priority, for the dataset chosen for bootstrapping: (i) large number of books

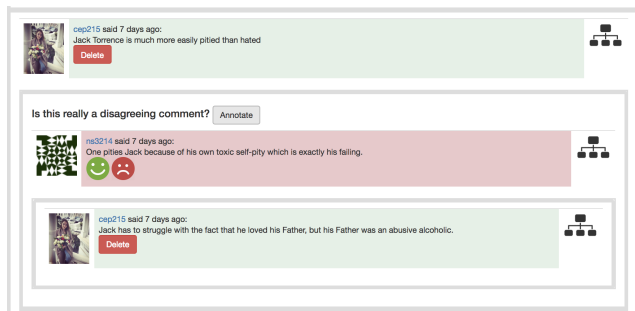


Figure 5: Annotation view.

(ii) large number of reviews (iii) detailed information about books  
(iv) easily accessible (v) easy to process.

We opted for the datasets used in [11, 13] restricted to books of the following genres: Business & Money, Literature & Fiction, Mystery, Thriller & Suspense, Romance, Science Fiction & Fantasy and Self-Help. These (restricted) datasets fulfil the given requirements. Bootstrapping was done in order to populate the top level of the tree (as shown in Figure 1) with positive and negative reviews for a book. This task can be split in two subtasks: (i) populate the database with reviews from various datasets (ii) detect the polarity of a review. For the polarity detection step, we initially opted to assign a polarity (positive or negative) to each review based on the score yielded by sentiment analysis tools such as the one developed by [10], integrated in the NLTK library [1]. We have randomly extracted a sample of 100 reviews for different books and manually verified them against the algorithm, getting an accuracy of 67%.

We have considered two other approaches: (i) splitting the reviews into sentences and calculating the average polarity of the reviews (ii) dividing reviews into blocks of positive and negative sentences. However, none of these methods improved the accuracy and we have decided to use the polarity score of the whole review.

We have also made use of the information coming from the Amazon reviews dataset, in particular the overall score of the review, ranging between 1 and 5, with 1 being the lowest score and 5 the highest score. If the overall value did not relate to the sentiment polarity value then we discarded the review and did not include it in the database. The relatedness was calculated as follows: overall value: 1-2 to negative, 3 to neutral, 4-5 to positive, and sentiment value (compound): < 0 to negative, 0 to neutral, > 0 to positive. Since our focus is on the supporting and attacking arguments, we did not include any neutral reviews in the database. In case of conflicts between the overall value and the sentiment value, we decided that the reviews should not be included in the database. The number of reviews gathered was sufficiently large to permit such cuts.

The annotation feature (see Section 4) allows the users to express any objections with regards to the initial annotation. The results from the annotation feedback are stored in the database and can also be examined from the admin view.

## 6 CONCLUSION

We proposed a web-based platform, BookSafari, that allows for the creation of annotated corpora, highly needed in Argument Mining

tasks. The corpora obtained represent Bipolar Argumentation Frameworks with two types of argumentative relations: *attack* and *support*. Our platform tries to fill a gap in popular websites dedicated to book reviews, similarly to already well-known platforms dedicated to movies (e.g. IMDb, Rotten Tomatoes), hotel reviews (e.g. booking.com, TripAdvisor), product reviews (e.g. Amazon). BookSafari enhances the capabilities of book review platforms with the ones from discussion fora, allowing users to engage in debates.

BookSafari allows visualisations in the form of table-like system and tree structures. The latter describes the dynamics of discussion for each review by showing its disagreeing comments and its agreeing comments. The agreeing/disagreeing comments of each review are shown in the next level of the tree. This view offers the possibility of following the discussion path. These tree-like structures can be seen as explanations for how good a book is. The annotated corpora can be used in building systems capable of engaging in a dialogue with users about books by putting forward arguments for/against the book or other arguments.

Preliminary evaluation after one week of trial launch allowed to gather approximately 500 pairs of arguments. This is encouraging and shows that BookSafari is an engaging platform and can be used for collecting data. More evaluation is needed to determine if we can obtain a large enough annotated dataset, highly needed in Argument Mining tasks.

## REFERENCES

- [1] Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural Language Processing with Python* (1st ed.). O'Reilly Media, Inc.
- [2] Elena Cabrio and Serena Villata. 2013. A natural language bipolar argumentation approach to support users in online debate interactions†. *Argument & Computation* 4, 3 (2013), 209–230. <https://doi.org/10.1080/19462166.2013.862303>
- [3] Lucas Carstens and Francesca Toni. 2015. Towards relation-based Argumentation Mining. In *Proceedings of the 2nd Workshop on Argumentation Mining*. 29–34.
- [4] Lucas Carstens and Francesca Toni. 2017. Using Argumentation to Improve Classification in Natural Language Problems. *ACM Transactions on Internet Technology* 17, 3 (2017), 30:1–30:23.
- [5] Claudette Cayrol and Marie-Christine Lagasque-Schiex. 2005. On the Acceptability of Arguments in Bipolar Argumentation Frameworks. In *Proceedings of the 8th European conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*. Springer Berlin Heidelberg, 378–389.
- [6] Oana Cocarascu and Francesca Toni. 2016. Detecting deceptive reviews using Argumentation. In *Proceedings of the 1st International Workshop on AI for Privacy and Security, PrAISe@ECAI*. ACM, 9:1–9:8.
- [7] Oana Cocarascu and Francesca Toni. 2017. Identifying attack and support argumentative relations using deep learning. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017*. 1385–1390.
- [8] Ivan Habernal and Iryna Gurevych. 2017. Argumentation Mining in User-Generated Web Discourse. *Computational Linguistics* 43, 1 (2017), 125–179.
- [9] Ivan Habernal, Raffael Hannemann, Christian Pollak, Christopher Klammer, Patrick Pauli, and Iryna Gurevych. 2017. Argotario: Computational Argumentation Meets Serious Games. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP, System Demonstrations*. 7–12.
- [10] Clayton J. Hutto and Eric Gilbert. 2014. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. In *Proceedings of the Eighth International Conference on Weblogs and Social Media, ICWSM*.
- [11] Brian Kenji Iwana and Seichi Uchida. 2016. Judging a Book By its Cover. *CoRR* abs/1610.09204 (2016). <http://arxiv.org/abs/1610.09204>
- [12] Marco Lippi and Paolo Torroni. 2016. Argumentation Mining: State of the Art and Emerging Trends. *ACM Transactions on Internet Technology* 16, 2 (2016), 10.
- [13] Julian J. McAuley, Christopher Targett, Qinfeng Shi, and Anton van den Hengel. 2015. Image-Based Recommendations on Styles and Substitutes. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 43–52.
- [14] Stefano Menini, Elena Cabrio, Sara Tonelli, and Serena Villata. 2018. Never Retreat, Never Retract: Argumentation Analysis for Political Speeches. In *AAAI*.