

Annotating Very Large Arguments

Kamila GORSKA^{a,1}, Wassiliki SISKOU^b and Chris REED^a

^aCentre for Argument Technology, University of Dundee, UK

^bUniversity of Konstanz, Germany

Abstract. We present a method of annotating very large arguments, with the use of IMC-Tool. IMC-Tool aids in creating long-distance argument structure relations, by providing a simple annotation tool, and integration software to synthesise argument annotations at scale.

Keywords. argument annotation, inter-map correspondence, large arguments

Annotating very large arguments presents particular problems, which can be naively tackled by simply dividing into sub-tasks – rather than try to analyse 5,000 words of argumentation as a whole, instead split into 20 sub-tasks of 250 words each. For environments in which arguments can be artificially constrained, (such as kialo.com and debategraph.org, for example) this solution can suffice. In general, however, the problem is that synthesising the solutions to the sub-tasks is a major challenge in itself. IMC-Tool offers a solution to this problem by providing a method to add *inter-map correspondence*, or IMC, to the process of argument analysis, and specifically to argument analysis conducted using IAT (Inference Anchoring Theory) [1].

IMC-Tool consists of multiple components; an *annotation tracker spreadsheet* which stores id numbers of annotated argument maps, an *imc spreadsheet* which is used by the annotators for the IMC argument annotation, the *extractNodes script* which populates the imc spreadsheet with node details for use for IMC annotation, and a *createIMCMap script* which turns the identified relations into an IAT map uploaded to AIFdb. Figure 1 shows a diagram demonstrating how these components work together.

Firstly, the initial annotation is split into sub-tasks and annotated using OVA+ [2]. Each map is uploaded to AIFdb [3] and a unique AIFdb map id for each map is stored in the annotation tracker spreadsheet. To begin the IMC procedure, by using the annotation tracker spreadsheet, the *extractNodes script* extracts all required node details for the task, by requesting each map from AIFdb. The map is parsed and the data is used to populate the imc spreadsheet with the content of each node.

The annotator carries out the IMC annotation by selecting the source and target nodes in the IMC spreadsheet. Subsequently, the target node's details are appended to the source node's row, containing the node's excerpt number, locution id and content. The analyst can specify the identified structure of the relation from a drop down list, containing relation types such as inference, conflict and rephrase, and their certainty that the identified relation is correct. The annotation is then verified by another annotator, who will either accept or reject the implementation. An example of the IMC spreadsheet containing IMC annotations is shown in Figure 1.

¹Corresponding Author. E-mail:k.gorska@dundee.ac.uk

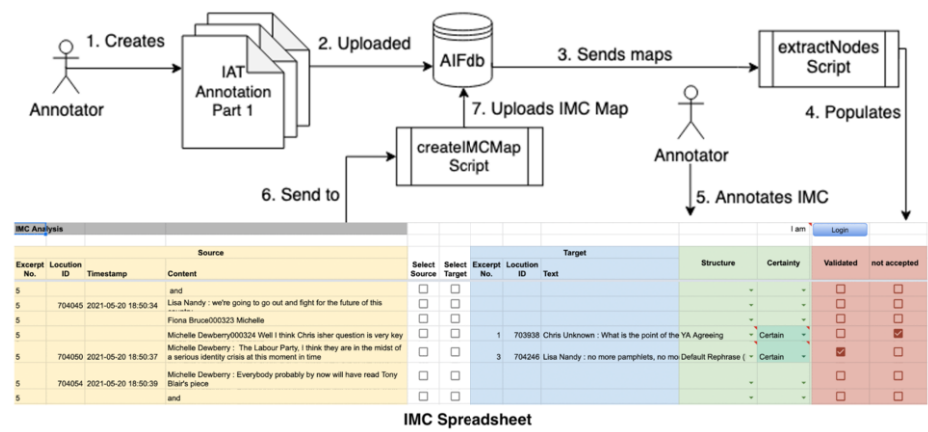


Figure 1. Model of the IMC process using IMC-Tool.

Once the IMC annotation is complete and all relations between parts have been implemented, the createIMCMap script is used to build the complete IAT structure of all identified relations. The source and target locutions are matched to the corresponding location nodes in the annotated maps in AIFdb. The relation type identified by the annotator is automatically implemented using IAT structure, and once complete for all rows, a map of the complete IMC annotation is uploaded to AIFdb. When all individual maps and the map of identified relations between them are added to one corpus, AIFdb resolves all duplication of nodes, resulting in a complete IAT analysis of the full text.

The IMC procedure described was applied to a 280,000-word corpus of 30 episodes of IAT-annotated topical debate, QT30 [4]. The IMC analysis was completed in the context of near real-time analysis by 4-6 annotators in around 90-100 minutes per hour-long debate. In the QT30 corpus, IMC relations make up 14% of all argumentative relations. Out of all IMC relations, there is a predominance of rephrases (64% compared to 22% and 13% of inferences and conflicts respectively), which is also a higher proportion than the 43% found in non-IMC relations. This suggests that in order to understand non-local phenomena, we must first understand rephrase. Another area of future work is evaluating the inter-annotator agreement of the IMC procedure.

References

[1] Reed, C.A. and Budzynska, K. (2011) "How dialogues create arguments" in van Eemeren, F.H. et al. (eds) Proceedings of the 7th Conference of the International Society for the Study of Argumentation (ISSA 2010), Sic Sat, Amsterdam
[2] Janier M, Lawrence J, Reed C. OVA+: an Argument Analysis Interface. InComputational Models of Argument 2014 (pp. 463-464). IOS Press.
[3] Lawrence J, Bex F, Reed C, Snaith M. AIFdb: Infrastructure for the argument web. InComputational Models of Argument 2012 (pp. 515-516). IOS Press.
[4] A. Hautli-Janisz, Z. Kikteva, W. Siskou, K. Gorska, R. Becker, and C. Reed. QT30: A Corpus of Argument and Conflict in Broadcast Debate. In Proceedings of the Language Resources and Evaluation Conference, Marseille, France, Jun. 2022, pp. 3291—3300.