

**Annotation guidelines**  
**Argumentative units and relations in the SciDTB corpus**  
**v1.3 - February 2020**

1. Introduction .....	2
Argument mining .....	2
Tasks.....	2
2. Annotation of arguments.....	3
Annotation schema: Argumentative units and relations.....	3
Most frequent relations.....	10
Conflict resolution for types of units .....	11
Annotation process.....	11
3. Assignment of quality scores .....	18

## 1. Introduction

### Argument mining

The analysis of argumentation has its roots in philosophy and different theories have been proposed to characterize argumentative structures in texts. In the context of this work we consider an argument to be formed by argumentative units (or components) such as *claims* and evidence (also referred to as *premises*), linked by directed relations like *support* (in which the premise is intended to convince the reader about the validity of an assertion) or *attack* (in which the premise intends to undercut the validity of an assertion). Fig. 1 shows an argument with its components and relations from a scientific article in the domain of educational research.<sup>1</sup>

The automatic identification of arguments, its components and relations in texts is called *argument mining*. Being able to automatically identify not only what is being said by the author of a text but also the reasons provided to persuade the reader about the validity of the assertions made in it it can be useful in multiple applications, ranging from a fine-grained analysis of opinions to the generation of abstractive summaries of texts.

In this work we are interested in the identification of argumentative structures in scientific texts (in the domain of natural language processing / computational linguistics) and, based on these elements, to characterize the argumentative quality of the texts according to two specific dimensions: their logical quality in terms of their *cogency* or argumentative strength, and their rhetorical quality in terms of their *persuasive effect*.

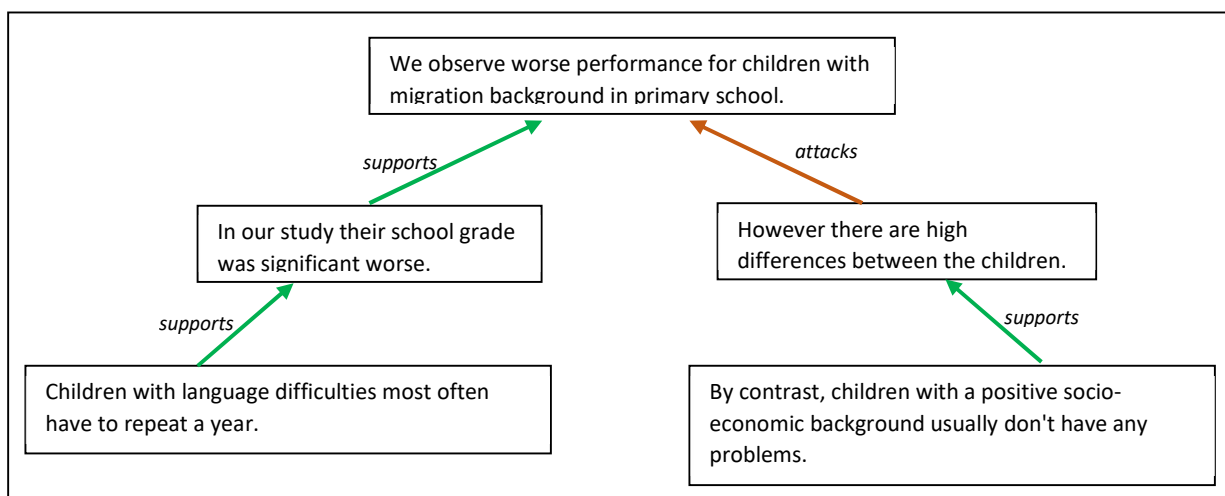


Fig 1. Argumentative units and relations.

### Tasks

These guidelines describe two **tasks**:

1. The identification of argumentative units and relations in abstracts of papers included in the Discourse Dependency TreeBank for Scientific Abstracts (SciDTB);<sup>2</sup>
2. The assignment of quality scores to the annotated abstracts.

<sup>1</sup> Example from: Kirschner, Christian, Judith Eckle-Kohler, and Iryna Gurevych. "Linking the thoughts: Analysis of argumentation structures in scientific publications." In Proceedings of the 2nd Workshop on Argumentation Mining, pp. 1-11. 2015.

<sup>2</sup> <https://www.aclweb.org/anthology/P18-2071/>

## 2. Annotation of arguments

### Annotation schema: Argumentative units and relations

The annotation schema includes the following types of units and relations:

Unit	Description
<i>proposal</i>	Proposed solutions/approaches
<i>proposal-implementation</i>	Specific methods proposed to achieve solutions/implement approaches
<i>observation</i>	Data obtained from experiments
<i>result</i>	Direct interpretation of observed data
<i>conclusion</i>	High-level interpretation/generalization of results
<i>means</i>	Methods/processes/tools used that are not a relevant component of the proposed solution (e.g. used in evaluations)
<i>result-means</i>	One unit that includes both results and the means by which they were obtained
<i>motivation-problem</i>	Existing/known problem/limitation addressed by the proposed solution/approach
<i>motivation-hypothesis</i>	New potential ideas/paths for addressing known problems/limitations
<i>motivation-background</i>	Known information to support relevance/appropriateness of proposed solution/approach
<i>information-additional</i>	Additional contextual information (concepts/definitions/examples) that cannot be considered as motivation of the proposed solution

Table 1. Types of units

Relation	Description
<i>support</i>	Reasons/evidence for the validity/relevance/appropriateness of the elements mentioned in the parent unit
<i>elaboration</i>	More detailed information about the issues (proposals, etc.) referred to in the parent unit
<i>sequence</i>	Successive steps in a process
<i>by-means</i>	A combination of support and elaboration used to link a means unit with its parent (in general, a unit of type result or result-means)
<i>info-required</i>	Information needed to understand the parent unit
<i>info-optional</i>	Additional (not essential) information used to contextualize the parent unit

Table 2. Types of relations

## I. Units

### ***proposal***

These units are used to identify contributions of the paper as proposed by the authors, including new solutions and/or approaches to existing needs / problems.

Example:

- *We present a novel approach to improve word alignment for statistical machine translation ( SMT ) . (D14-1016)*

### ***proposal-implementation***

These units are used to describe, with a greater level of detail than units of type *proposal*, how / by which means the identified problem / need is addressed in order to achieve the solution proposed by the authors.

These units describe tools, methods, resources that are relevant enough as to be considered part of the authors' contributions .

Examples:

- *We observe , identify , and detect naturally occurring signals of interestingness in click transitions on the Web between source and target documents , which we collect from commercial Web browser logs . (D14-1002)*
- *The model is entirely unsupervised - preferences are learned from unannotated corpus data . (D14-1004)*

Please note that in the cases where the method is the main object of the investigation and it is described on a high level, the unit should be assigned the label *proposal*, not *proposal-implementation*. For instance:

- *In this paper , we use a weighted vote method to transform discontinuous word alignment to continuous alignment , which enables SMT systems extract more phrase pairs . (D14-1016)*

### ***observation***

These units describe observed experimental results (data) obtained by the authors in their experiments / through their proposed methods.

Example:

- *Our method produces a gain of +1.68 BLEU on NIST OpenMT04 for the phrase-based system , and a gain of +1.28 BLEU on NIST OpenMT06 for the hierarchical phrase-based system . (D14-1016)*

### ***result***

These units describe an interpretation of observed results. The actual observations that lead to these interpretations can be (and many times, are) omitted.

Example:

- *Experimental results show statistically significant improvements of BLEU score in both cases over the baseline systems . (D14-1016)*

Note that, in some cases, the information of the data (observation) and its interpretation (result) is contained in one sentence which cannot be separated. For instance:

- *Our results show that the various methods linking the two tasks are not significantly different from one another , although they perform better than the isolated prediction method by 0.5-1.5 % in the F1 score .* (D14-1013)

In this case, the whole unit should be labeled as *result*.

### **means**

These units are used to describe means (processes / tools / algorithms / corpora) used in different phases of the research (including the implementation of the proposed experiments and their evaluations).

Unlike units with type *proposal-implementation*, these have a more marginal relevance in terms of the contributions of the described work. In general, it could be considered that other means could have been chosen by the authors without making a significative difference.

Example:

- *We conducted experiments on two standard benchmarks : Chinese PropBank and English PropBank .* (D14-1041)

### **result-means**

It is frequently the case that results and the means by which they were obtained are described in the same sentence – or even EDUs. In these cases the units should be annotated as *result-means*.

Examples:

- *Results on the Switchboard disfluency tagged corpus show utterance-final accuracy on a par with state-of-the-art incremental repair detection methods , but with better incremental accuracy , faster time-to-detection and less computational overhead .* (D14-1009)
- *The model's performance is evaluated on a pseudo-disambiguation task , on which it is shown to achieve state of the art performance .* (D14-1004)

If concrete means (tools / methods / resources) are not mentioned, but generic expressions are used (such as “*Experimental results show...*”) the unit should be labeled as result, as it is not one of the main communicative functions to describe the specific means by which the results where obtained.

### **conclusion**

These units are used to introduce high-level assertions made by the authors in relation of the merits of their contributions. They should, in general, originate from an interpretation of the evidence obtained in the work - although these results might not be included explicitly in the abstract and it might not be clear how the authors achieved those conclusions.

Example:

- *This transfer learning approach brings a clear performance gain over features based on the traditional bag-of-visual-word approach .* (D14-1005)

### ***motivation-problem***

These units are used to describe problems / limitations of existing solutions, which contribute to justify the need / relevance / appropriateness of the proposed research.

Example:

- *However , fundamental problems on effectively incorporating the word embedding features within the framework of linear models remain . (D14-1012)*

### ***motivation-hypothesis***

These units are used to explicitly state the hypothesis that the authors want to corroborate through the described work. They are not frequent in the SciDTB corpus.

Example:

- *Combining the two tasks can potentially improve the efficiency of the overall pipeline system and reduce error propagation . (D14-1013)*

### ***motivation-background***

These units are used to introduce background information that contributes to support the relevance / appropriateness of the research described in the paper.

Example:

- *Recent work has shown success in using continuous word embeddings learned from unlabeled data as features to improve supervised NLP systems , which is regarded as a simple semi-supervised learning mechanism . (D14-1012)*

In a few cases, the motivation is included in a sentence describing the proposal.<sup>3</sup> For instance (D14-1004):

- *Inspired by recent advances of neural network models for NLP applications , we propose a neural network model that learns to discriminate between felicitous and infelicitous arguments for a particular predicate .*

In this case the whole unit should be labeled as *proposal*.

### ***information-additional***

These units introduce concepts / definitions / examples that contribute to better understand the described research. Unlike the units of type motivation-\*, they do not introduce a motivation for the described research or, in other words, they are not used to persuade the reader about the relevance / appropriateness of the proposed solution.

Examples:

- *The basic idea of partial-label learning is to optimize a cost function that marginalizes the probability mass in the constrained space that encodes this knowledge . (D14-1010)*
- *The structure of argumentation consists of several components ( i.e. claims and premises ) that are connected with argumentative relations . (D14-1006)*

---

<sup>3</sup> This can occur for units of type *motivation-background*, *motivation-problem* or *motivation-hypothesis*.

## Second type

There might be sentences that contain information corresponding to more than one type. Sentences can therefore be annotated with a second type of unit. Please weight the relevance of the different types of information contained in the sentence to make a decision with respect to the main and secondary types.

Example (D14-1057):

- *We find that all the models we compare are effective for verb clustering ; the best-performing model uses syntactic information to induce nouns classes from unlabelled data in an unsupervised manner*

Main type: *result*; Second type: *proposal-implementation*

- *Whereas in the open domain system , features derived from WordNet show only slight improvements , we show that its counterpart for the medical domain ( UMLS ) shows a significant additional benefit when used for feature generation .*

Main type: *result*; Second type: *motivation-background*

## Main claim

When adding / modifying a unit corresponding to a claim (*proposal* or *conclusion*) the *claim type* attribute is made available. It is used to identify the main contribution of the research paper.

Only one argumentative unit should be marked as the main claim.<sup>4</sup> In order to identify it, one could formulate the following question: If I were to chose only one unit to describe the most significant contribution of this work, which would it be?

In general, the main claim is the unit where the proposal / approach described in the paper is introduced. There might be a few cases, though, where this information is too general: it basically establishes the area of work / family of problems where the current research can be framed and does not allow the reader to identify what is specific about the proposed solution / approach. In these cases, a more specific unit should be selected, if possible. If a specific unit cannot be identified, the general one should be marked as the "main" claim.

Example (D14-1004):

1. *This paper investigates the use of neural networks for the acquisition of selectional preferences .*
2. *Inspired by recent advances of neural network models for NLP applications , we propose a neural network model that learns to discriminate between felicitous and infelicitous arguments for a particular predicate .*

In this case, unit 2 should be marked as the main claim, since it describes more specifically the main contributions / proposal of the research, while unit 1 is used to introduce the area in which the research is framed. If unit 2 were not present, unit 1 should be marked as the main claim.

---

<sup>4</sup> Even if this is currently not limited by the annotation tool.

## II. Relations

As mentioned, we are interested not only in identifying the components of the arguments used in the abstracts but also their structure. This is done by means of directed relations that indicate how the AUs are connected from an argumentative (and, in some cases, discursive) perspective.

### **support**

Units of type *proposal*, for instance, convey implicit claims with respect to the relevance / validity / appropriateness of the proposed method or solution. In other cases, such as *conclusion*, claims are expressed more explicitly.

Implicit or explicit claims contained in argumentative units can be supported by other units which provide reasons / evidence to convince the reader about the validity of the claim. These units are considered as premises which can, in turn, be supported by additional premises.

The *support* relation links premises (*motivation-\**, *result*, *conclusion*, *observation*), to claims (*proposal*, *conclusion*) or other premises (*result*). More specifically, this type of relation is established between units of types:

*motivation-\** -> *proposal*  
*conclusion* -> *proposal*  
*result* -> *proposal*  
*result* -> *conclusion*  
*observation* -> *result*  
*observation* -> *proposal*

### **elaboration**

This relation is used to link a unit with another one that provides more detailed / specific information about a particular issue. The direction of the relation goes from the more specific unit to the more general one.

Unlike support, *elaboration* relations do not provide reasons that contribute to affirm the relevance / validity / appropriateness of the elements considered in the parent unit.

These relations are used to link together two units of the same type (in general, *proposal*) or a unit of type *proposal* and the *proposal-implementation* unit that describes more specifically the proposed means to implement the solution. Due to their brevity, in abstracts it is less frequent to find *elaboration* relations between units with other types (such as *result* or *motivation-background*), although they can be present.

Example:

D14-1004

1. *Inspired by recent advances of neural network models for NLP applications , we propose a neural network model that learns to discriminate between felicitous and infelicitous arguments for a particular predicate . (proposal)*
2. *The model is entirely unsupervised - preferences are learned from unannotated corpus data . (proposal-implementation)*

In this case, a relation [2] –*elaboration*–> [1] can be established.

### **sequence**

This relation is not very frequent and it is used solely to link together units that contain information about successive steps in a process.



It can be established between two units of types that describe processes: *proposal-implementation* (most frequent) or *means*.

The direction of the relation goes in the direction opposite to how the steps are performed (from the second to first, third to the second, etc.).

Examples:

D14-1006

1. *First , we identify the components of arguments using multiclass classification . (proposal-implementation)*
2. *Second , we classify a pair of argument components as either support or non-support for identifying the structure of argumentative discourse . (proposal-implementation)*

D14-1038

1. *ReNoun creates a seed set of training data by using specialized patterns and requiring that the facts mention an attribute in the ontology . (proposal-implementation)*
2. *ReNoun then generalizes from this seed set to produce a much larger set of extractions that are then scored . (proposal-implementation)*

Please note that there are some cases in which the authors use terms as "first", "second", etc. in order to structure their discourse but they are not necessarily describing sequential processes. For instance, consider the following fragment (D14-1044):

1. *We present two improvements to the use of such large corpora to augment KB inference . (proposal)*
2. *First , we present a new technique for combining KB relations and surface text into a single graph representation that is much more compact than graphs used in prior work . (proposal)*
3. *Second , we describe how to incorporate vector space similarity into random walk inference over KBs , reducing the feature sparsity inherent in using surface text . (proposal)*

In this case, there is not a sequential relation between units [2] and [3]. Instead, they are both linked by *elaboration* relations to [1].

### ***by-means***

This relation is used to link a unit that describes *means* used to achieve results described in the parent unit.

It can therefore be considered as a combination of *elaboration* (in the sense tha it provides more information about the implementation of an issue mentioned in the parent unit) and *support* (in the sense that it explains how the authors arrived at the conclusions/interpretations contained in the parent unit).

The source of the relation is the unit of type *means*.

Example (D14-1004):

1. *We conducted experiments on two standard benchmarks : Chinese PropBank and English PropBank . (means)*
2. *The experimental results show that our approach can significantly improve SRL performance , especially in Chinese PropBank . (result)*

In this case, a relation [1] –by-means--> [2] can be established.

Please note that units that describe the specific way in which a *proposal* is implemented, which are of type *proposal-implementation*, should be linked to the *proposal* unit by an *elaboration* relation, and not a *by-means* one, as described above.

### **info-required**

This relation is used to link a source unit (in general, of type *motivation-background*) that is necessary to understand/contextualize a target unit (in general, of type *motivation-problem*). These units can form a chain of motivations.

Example (D14-1016):

1. *Conventional word alignment methods allow discontinuous alignment , meaning that a source ( or target ) word links to several target ( or source ) words whose positions are discontinuous . (motivation-background)*
2. *However , we cannot extract phrase pairs from this kind of alignments as they break the alignment consistency constraint . (motivation-problem)*

In this case, a relation [1] –*info-required*--> [2] can be established.

Please note that, even if units of type *motivation-\** can also be considered to explain/contextualize units of type *proposal*, in this case the relation that should be used is *support*.

### **info-optional**

This relation is used to link units of type *information-additional* to units of type *proposal* or *motivation-\** in which the definition / example / additional information is used.

Example (D14-1006):

1. *In this paper , we present a novel approach for identifying argumentative discourse structures in persuasive essays . (proposal)*
2. *The structure of argumentation consists of several components ( i.e. claims and premises ) that are connected with argumentative relations . (information-additional)*

In this case, a relation [2] –*info-optional*--> [1] can be established.

### **Most frequent relations**

Types of source unit	Most frequent relations / Types of target units
<i>proposal</i>	<i>elaboration (proposal)</i>
<i>proposal-implementation</i>	<i>elaboration (proposal*)</i> ; <i>sequence (proposal-implementation)</i>
<i>observation</i>	<i>support (result*)</i>
<i>result</i>	<i>support (proposal/conclusion)</i> ; <i>elaboration (result*)</i>
<i>conclusion</i>	<i>support (proposal)</i>
<i>means</i>	<i>by-means (result*)</i>
<i>result-means</i>	<i>support (proposal/conclusion)</i> ; <i>elaboration (result*)</i>
<i>motivation-problem</i>	<i>support (proposal/conclusion)</i>
<i>motivation-hypothesis</i>	<i>support (proposal)</i>
<i>motivation-background</i>	<i>info-required (motivation-problem)</i> ; <i>support (proposal/conclusion)</i>
<i>information-additional</i>	<i>info-optional (proposal/ motivation*)</i>

## Conflict resolution for types of units

Types of units		Criteria
<i>information-additional</i>	<i>motivation-background</i>	If the unit is linked to a <i>motivation-problem</i> unit by means of an <i>info-required</i> relation it should be labeled as <i>motivation-background</i> . If the information contained in the unit could be omitted without a significative loss of argumentative strength, it should be labeled as <i>information-additional</i> .
<i>proposal</i>	<i>proposal-implementation</i>	When very specific/concrete details are provided, the units should be labeled as <i>proposal-implementation</i> . Please also consider the verbs used. Action-oriented verbs (“use”, “identify”, “observe”, “collect”, “train”, “learn” etc.) tend to describe implementations, while verbs such as “propose”, “present”, “introduce” tend to identify proposals at a higher level. <sup>5</sup> If you cannot make a decision based on the language and the level of detailed included in the unit, label it as <i>proposal</i> .
<i>means</i>	<i>proposal-implementation</i>	Units that are linked to units of type <i>proposal</i> should be labeled as <i>proposal-implementation</i> , while units linked to units of type <i>result</i> should in general be labeled as <i>means</i> .
<i>result</i>	<i>conclusion</i>	The decision should be made based on the level of assertiveness of the language used. Terms such as “it is shown” / “experimental results show” tend to identify units of type <i>result</i> . If it is not possible to decide, label them as <i>conclusion</i> .

## Annotation process

The annotation is done by means of an online tool for the annotation of argumentative graphs developed on top of the GraPAT tool<sup>6</sup> and tailored to the specificities of this task.

It is available at <http://scipub-taln.upf.edu/grapat/GraPAT>

You will be provided with abstracts from the SciDTB corpus which have previously been segmented into elementary discourse units (EDUs).

Your task is:

1. To identify which EDUs form an argumentative unit (AU);
2. To assign a type to each AU according to the annotation schema described in this document;
3. To link the AUs in order to build an argumentative tree. One AU is linked to another by means of the argumentative relations (ARs) described in this document.

When the process is completed you will have created an **argumentative graph** in which the nodes are the AUs and the edges are the identified relations.

We consider the following **contrains**:

1. The argumentative graph has to be a **tree**: only AUs at the bottom of the tree do not have incoming edges and only one AU at the top of the tree does not have outgoing edges;
2. Each EDU has to be included in one and only one AU.
3. An **argumentative unit** should span over one or more sentences. This means that the smallest annotation unit is a sentence.

<sup>5</sup> This is in general but not always the case. Consider, for instance, D14-1009: “STIR uses information-theoretic measures...”. STIR is the solution being proposed, so in this case the unit should be labeled as *proposal*.

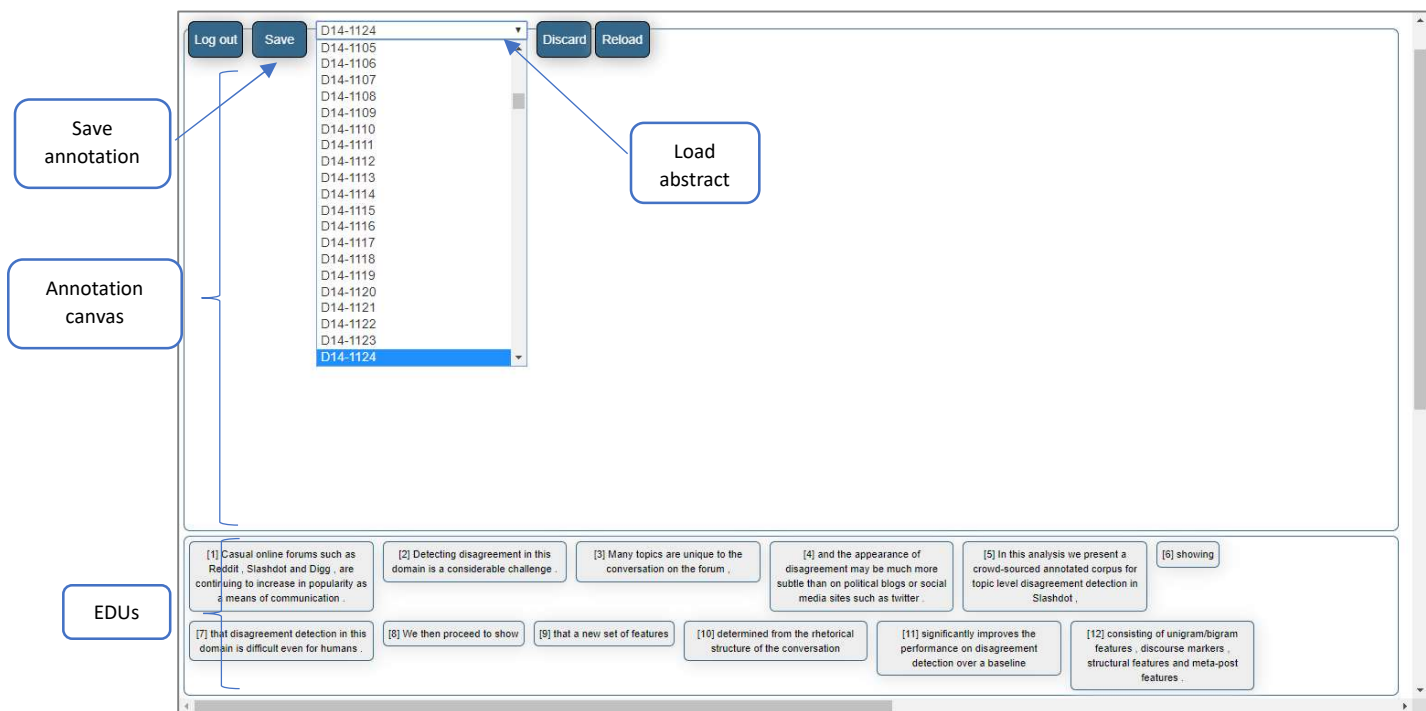
<sup>6</sup> <https://www.aclweb.org/anthology/L14-1636/>

## Steps

1. Log in into the system with the provided username and password;
2. Select a new abstract to annotate based on its ACL Anthology identifier;
3. Create new (empty) nodes for each AU and assign types to them;
4. Populate the newly created AUs with the EDUs contained in them;
5. Link the AUs by means of the directed relations;
6. When there are no more EDUs to use and no AUs without being linked to other AUs, save the annotation.

We now describe these steps in more detail.

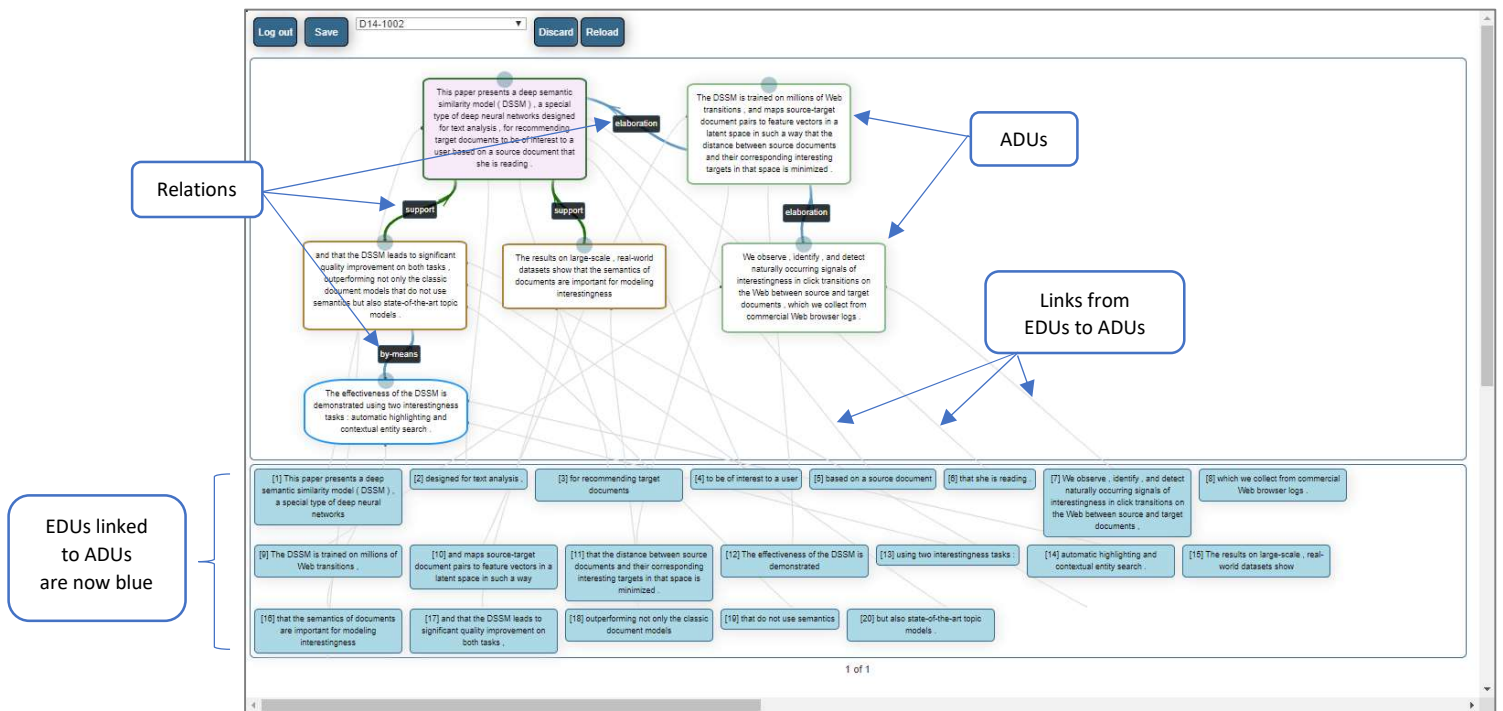
When you log in into the annotation tool you first have the option of selecting an abstract to annotate by means of a drop-down list of ACL Anthology identifiers.



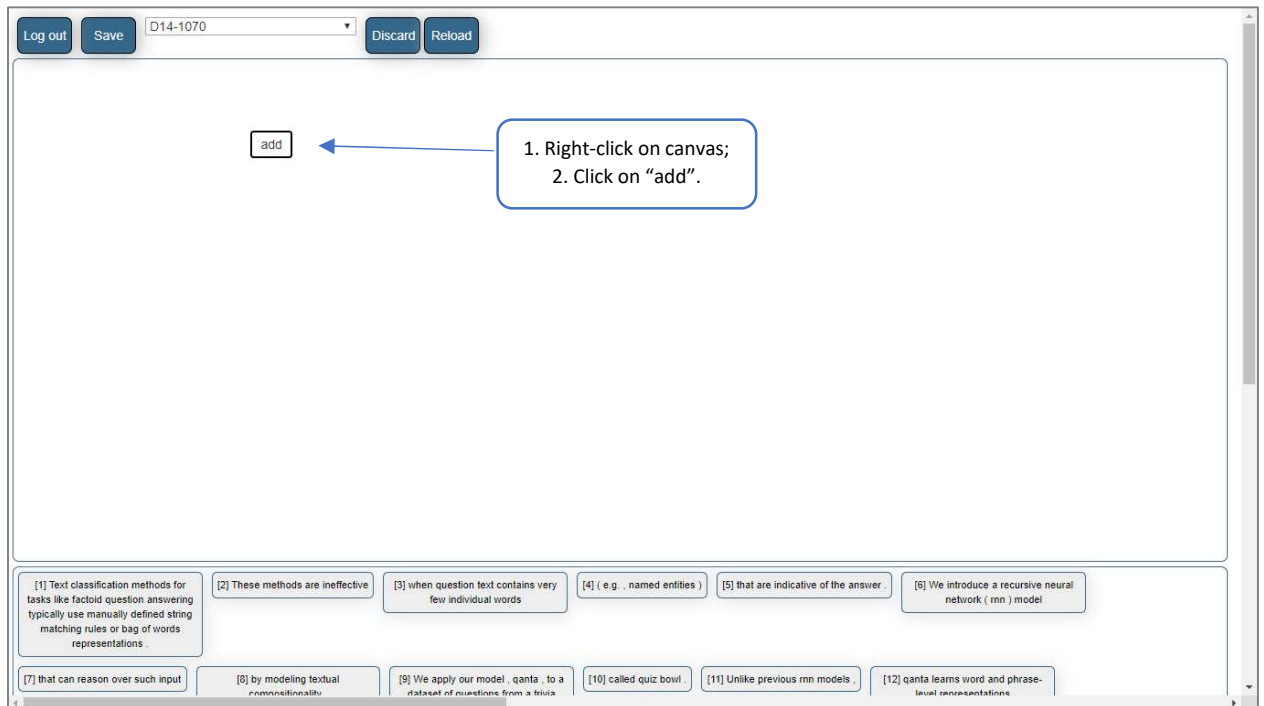
Once an abstract is loaded, the argumentative units and relations are displayed inside a rectangular area at the top of the page ("annotation canvas") and the original elementary discourse units (EDUs) are displayed at the bottom of the page, each one in a separate box.

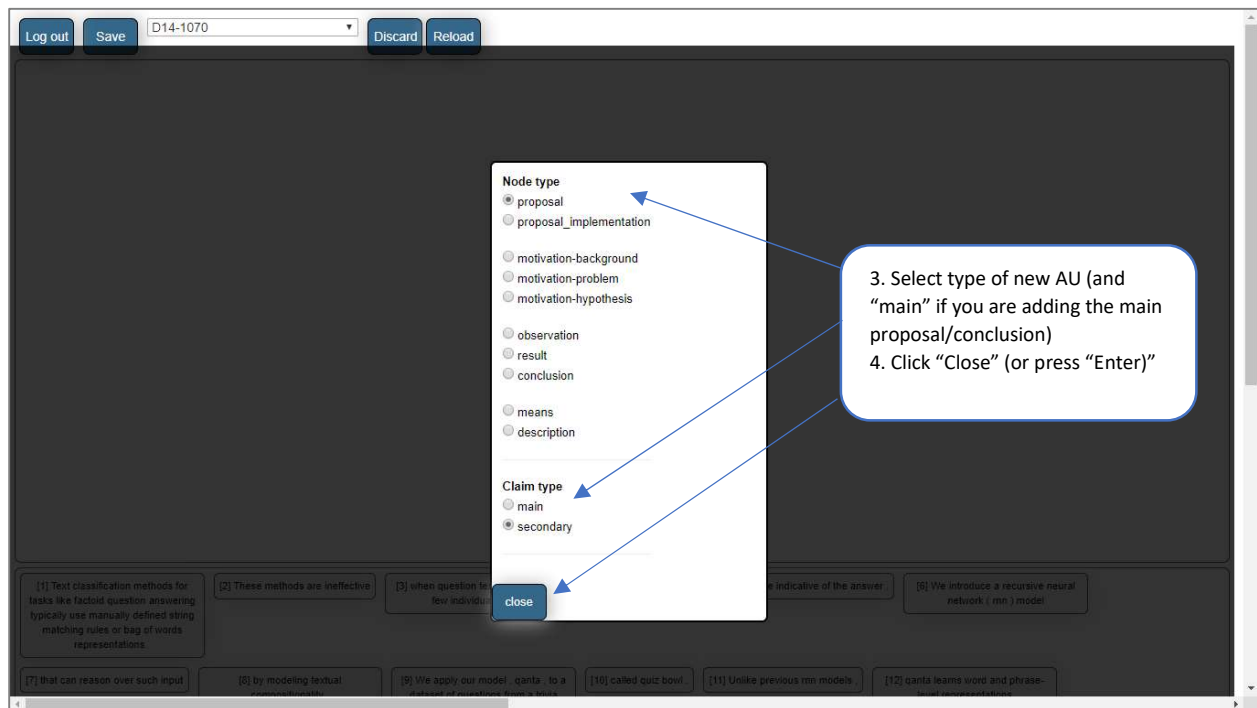
The first time an abstract is loaded it will contain no annotations, so the "annotation canvas" will be empty. Through the annotation process you will generate the *argumentative tree* in this area.

The nodes of this graph will be the *argumentative units* (containing one or more EDUs) and the edges will be the *relations* identified between them. When the annotation is completed, your canvas should look like:

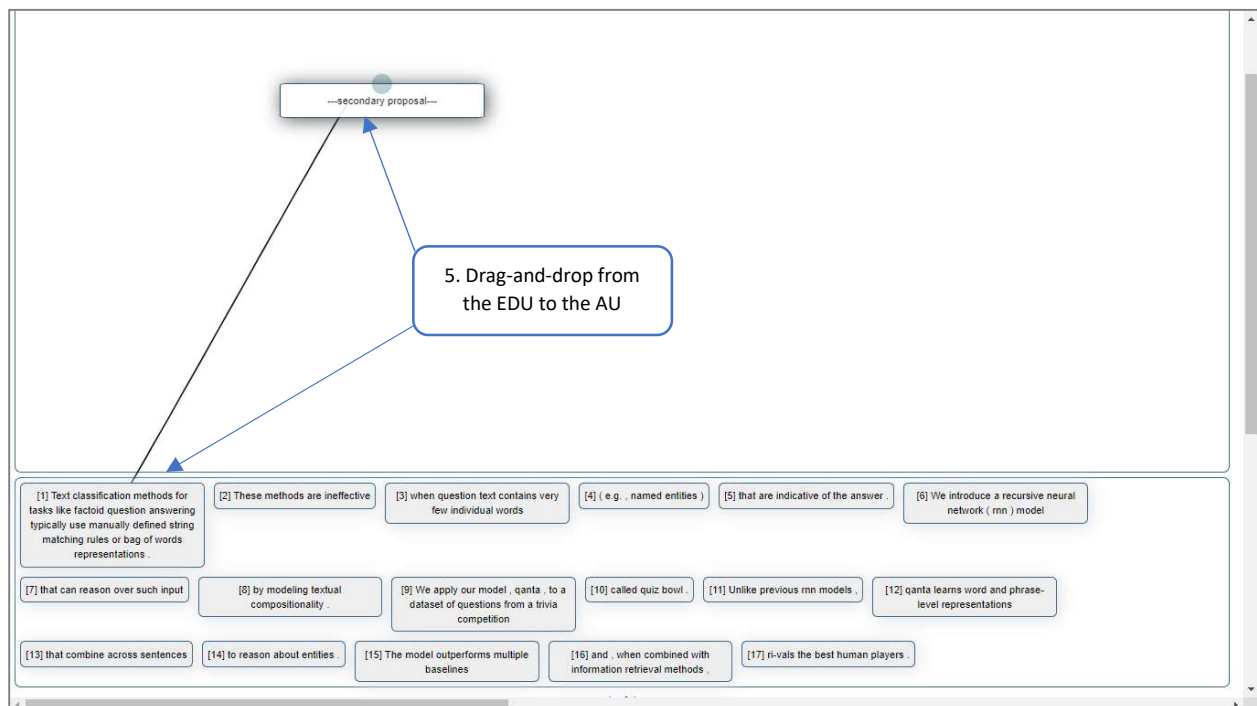


In order to add a new AU node, you have to right-click on the canvas – and then click on the “add” button. This will make a dialog box to pop-up, where you can select the type of the AU being added.

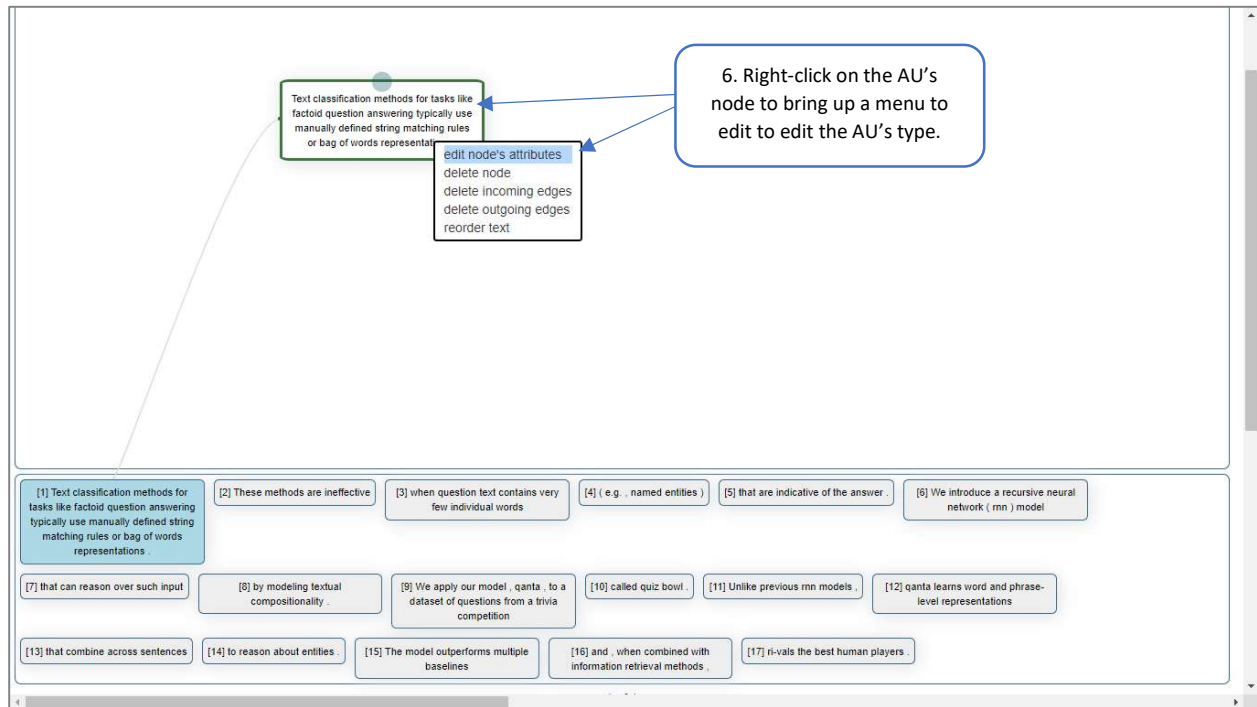




To add an EDU to the recently created AU, click on the EDUs' box and drag-and-drop the mouse from it to the AU. This will make the text of the EDU to be added to the AU. The color of the EU changes to blue to indicate that it's being used in some ADU (in the end all the EDUs should be blue).

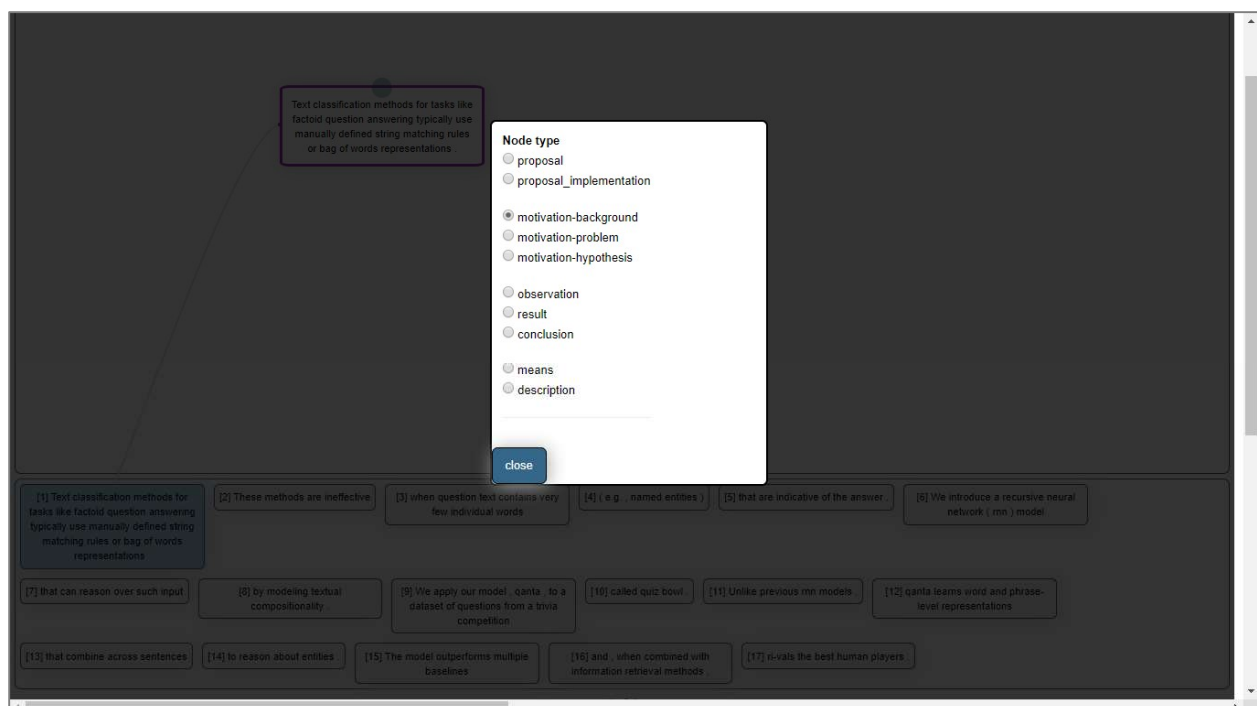


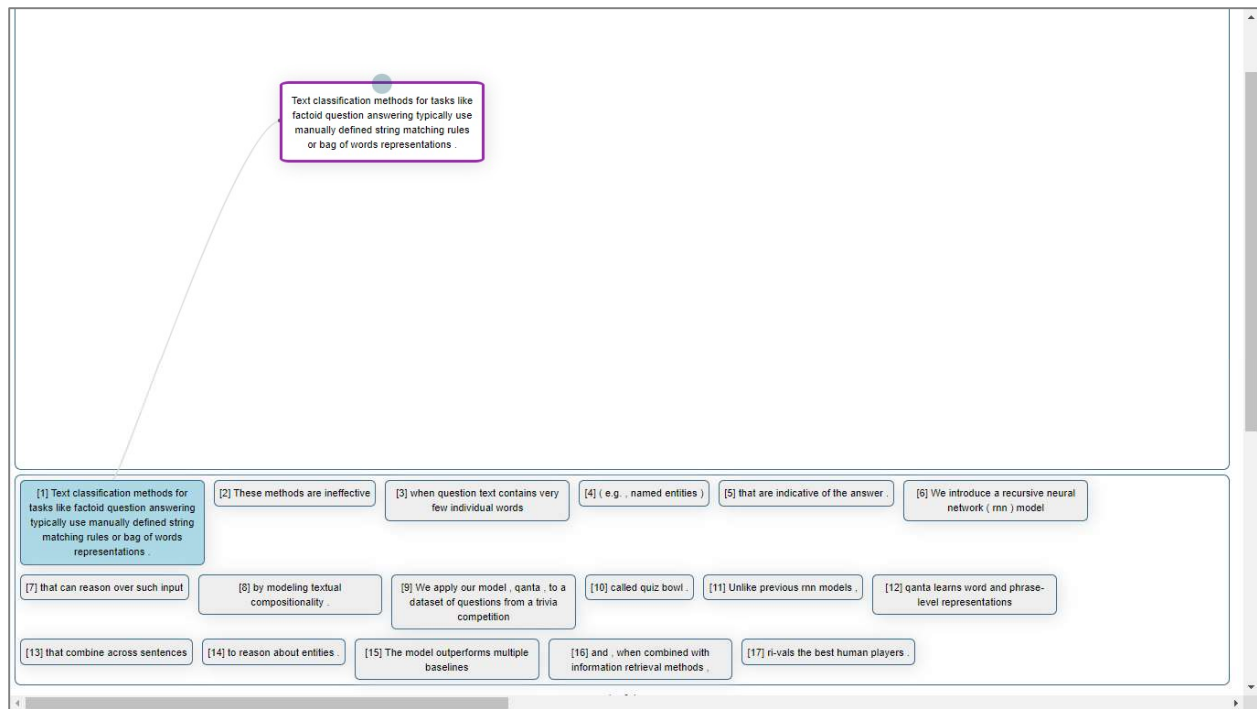
You can then right-click on the AU's node in order to edit it – for instance, to change its type or indicate that it corresponds to the main claim.



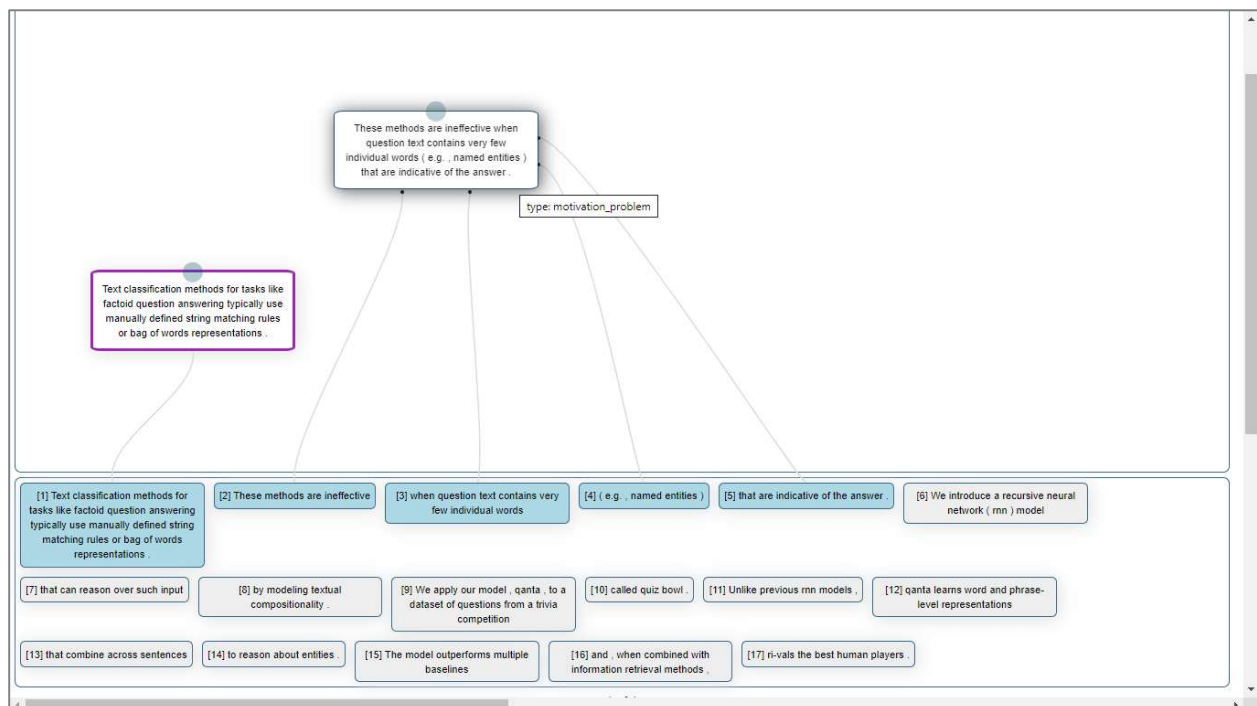
In this example, we change the type from *proposal* to *motivation-background* (the different types available in the annotation schema are explained below).

You will see that the border of the AU's node changes to indicate the AU's type. If you position the mouse over the AU's node the type of the AU will be displayed.



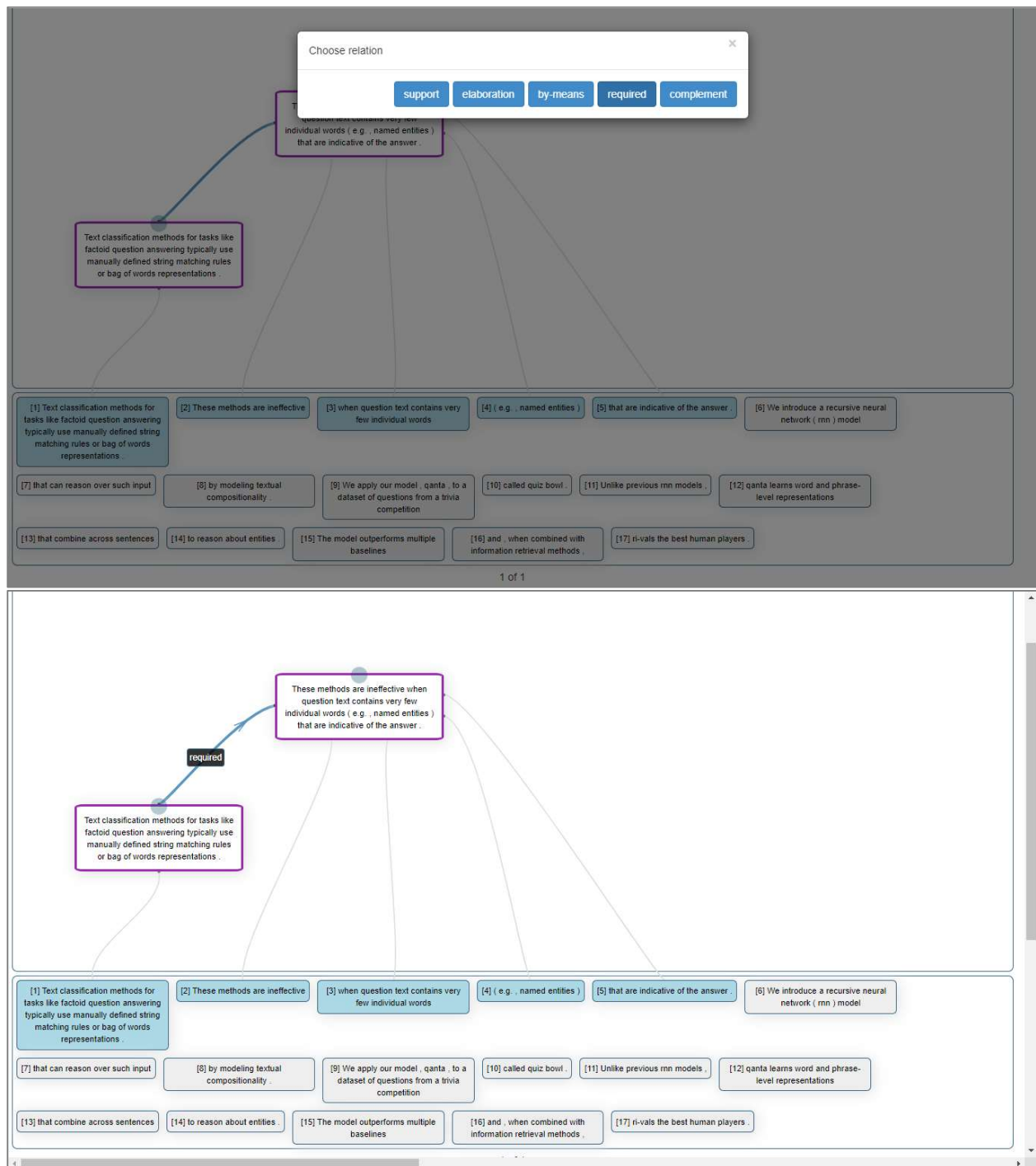


In the same way you can continue creating AUs and linking EDUs to them. In this example we add a new AU of type motivation-problem and link the corresponding EDUs to it.





In order to add a new relation edge between two existing AU nodes, you have to click on the grey circle on top of the source AU and drag-and-drop the cursor in the node of the target AU. A pop-up dialogue box will appear so you can indicate the type of the relation. In this case, we will add a *info-required* relation from the first node (*motivation-background*) to the second (*motivation-problem*).



You can save your annotations at any point and come back to them later to complete or modify them.

If you leave the page without saving your annotations, , a confirmation dialog box will appear. If you confirm it, the changes will be lost and it will not be possible to recover them.

### 3. Assignment of quality scores

In the process of identifying argumentative units and relations you will acquire a clear idea of the *argumentative strength* of the text. For instance, conclusions that are not supported by evidence, and lack of motivation for the proposed work might evidence a poor argumentative quality of the abstract. This is valuable information which we ask you to record.

After you complete the process of annotation of each abstract, we ask you to assign a score (from 1 to 3, being 1 the lowest score) to the two argumentative quality dimensions described below.

You will be provided a spreadsheet with the ACL Anthology identifier of the paper corresponding the abstract and, in each column, a quality dimension that you are asked to fill.

#### Quality dimensions

Dimension	Question to consider
Clarity	For the reasonably well-prepared reader, is it clear what was done and why?
Arrangement	Does the structure of the text contribute to persuade the reader about the value of the contributions?
Sufficiency	Are the premises (motivation, evidence) provided enough to justify the proposed solution/approach?
Relevance	Does the evidence provided contribute to convince the reader that the proposed approach/solution is effective/appropriate?