Visual Analysis of Argumentation in Essays

Dora Kiesel, Patrick Riehmann, Henning Wachsmuth, Benno Stein, and Bernd Froehlich

Abstract—This paper presents a visual analytics system for exploring, analyzing and comparing argument structures in essay corpora. We provide an overview of the corpus by a list of ArguLines which represent the argument units of each essay by a sequence of glyphs. Each glyph encodes the stance, the depth and the relative position of an argument unit. The overview can be ordered in various ways to reveal patterns and outliers. Subsets of essays can be selected and analyzed in detail using the Argument Unit Occurrence Tree which aggregates the argument structures using hierarchical histograms. This hierarchical view facilitates the estimation of statistics and trends concerning the progression of the argumentation in the essays. It also provides insights into the commonalities and differences between selected subsets. The text view is the necessary textual basis to verify conclusions from the other views and the annotation process. Linking the views and interaction techniques for visual filtering, studying the evolution of stance within a subset of essays and scrutinizing the order of argumentative units enable a deep analysis of essay corpora. Our expert reviews confirmed the utility of the system and revealed detailed and previously unknown information about the argumentation in our sample corpus.

Index Terms—Information Visualization, Text Analysis, User Interfaces, Visual Analytics, Argumentation Visualization, Glyph-based Techniques, Text and Document Data, Tree-based Visualization, Coordinated and Multiple Views, Close and Distant Reading

1 Introduction

Argumentation is key to human communication and writing, and subject to research not only of (digital) humanists and linguists but also of computer scientists for a long time. An argumentation consists of a complex structure of statements that give reason to accept or reject other statements [36]. The study of argumentation requires in-depth insights into various structural aspects, e.g. hierarchical, stance, or sequence relations, of not only one individual argumentative text but a whole corpus thereof. While humanist and linguist researchers are often interested in the cultural or topic-based differences within these aspects, computational linguists might use them to improve the automatic detection, extraction and analysis of arguments from argumentative texts. To be successful in those endeavors, however, all researchers need to gain a deep understanding of typical and common argumentative structures as well as their differences in a corpus.

So far, visualization research focused mostly on the depiction of individual argument structures via node link diagrams (e.g. *gIBIS* [6], *Belvedere* [39] or *Rationale* [45]) which often lack sequence information. Little work has been done for analyzing text collections (e.g. Wachsmuth et al. [48]) and only on a basic level: Even seemingly simple tasks, like determining the fraction of texts that contain a specific argument structure, are almost infeasible using such tools.

To fill this gap and support researchers in the study of argumentation, we designed, implemented, and evaluated a visual analytics system that allows the examination and comparison of structural aspects across multiple argumentative texts. It employs two novel coordinated visualizations: Our *ArguLines* summarize each individual argument structure in a glyph-based, space-saving manner whereas the *Argument Unit Occurrence Tree (AUOT)* overlays the argument structures of multiple essays by using extended hierarchical histograms. Elaborate queries for corpus-wide filtering of argumentative structures are enabled by direct interaction with the visual elements.

The development followed a user-centered design process that involved experts in all stages.

Our work is motivated by discussions and interviews with digital humanists and computer linguists who were dissatisfied with their possibilities to analyze text corpora with respect to argumentation practices.

- *D. Kiesel, P. Riehmann, B. Stein and B. Froehlich are with the Bauhaus-Universität Weimar. E-mail: <first>.<last>@uni-weimar.de*
- *H. Wachsmuth is with the Paderborn University. E-mail: henningw@upb.de*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

Digital humanists explained to us that their research includes questions such as what are the differences between the dominant argument structures of two sets of texts or which fraction of texts contains a certain argument structure. Computational linguists, on the other hand, wanted to detain the part of the structure that tends to be the same in all texts – e.g. to support finding argument structures in an unseen text – and to compare automatic structure annotations with those of human curators to assess the quality of the classification. Digital humanists also noted that it is helpful in the classroom to show students examples of student essays and good as well as debatable arguments in a visual way.

To support these and other tasks, our system for the analysis of argumentative text collections – more specifically essays – contributes the following novel visualization and interaction techniques:

- The ArguLines which represent each text by a sequence of glyphs encoding its argument units, the unit's argument depth, stance and order of appearance (Fig. 1, right). The compact design implemented as a variant of small multiples allows the user to quickly compare individual structures with each other and assess typical patterns as well as anomalies.

- Comparing structures in an entire corpus is enabled by the Argument Unit Occurrence Tree (AUOT) (Fig. 1, top left) which aggregates the argument structures of multiple essays into hierarchical histograms. It supports the user in determining the proportions of text that contain specific argument structures and makes it possible to examine the differences between sets of structures.

- The implemented interactive visual filtering system allows users to filter the text corpus for texts containing one or more relevant argument structures by interacting with the respective graphical representatives in a visualization.

Our reviews with external experts confirmed the utility of the individual visualizations and their coordinated interactions. This assessment was impressively substantiated by various new findings on the corpus, which were previously unknown even to experts who have conducted studies on machine learning with it.

2 Argumentation Model and Corpora

Various methods to model the characteristics and relations of an argument have been proposed [10, 21, 35, 40, 41, 44, 50]. We follow the simple and precise definition of Sinnott-Armstrong and Fogelin [36] extended by the approach of Stab and Gurevych [38]. An argument consists of a single statement and a set of other statements that give

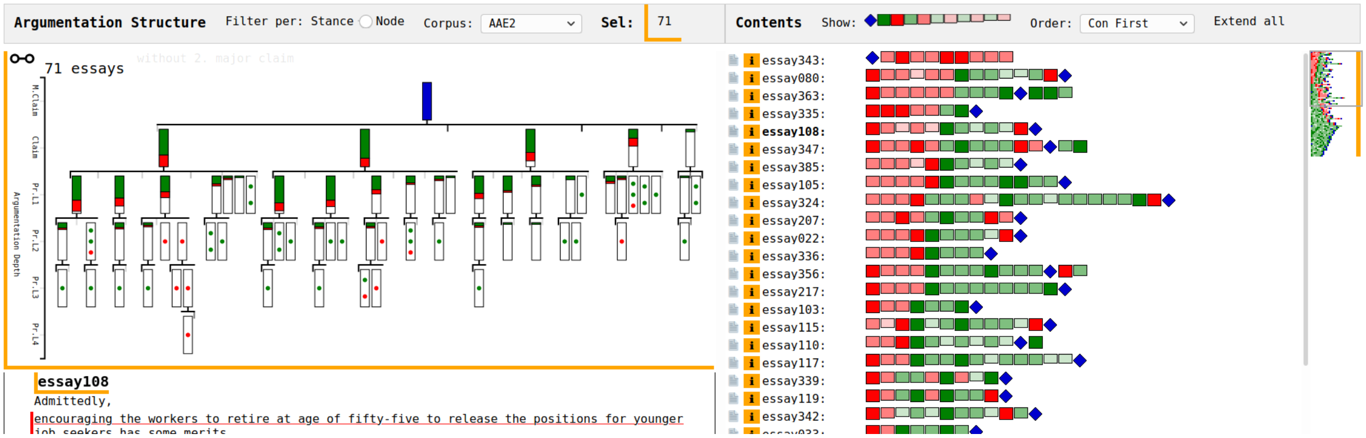


Fig. 1. An extract from the visual analysis system for essay corpora: On the right, the ArguLines view represents each document's structure as an ArguLine in overview and detail (see Sect. 4 for more details). Currently, all documents of the filtered dataset are selected (highlighted by orange markings in the list). The filter excludes all essays with more than one major claim. Essay 108 has been opened in text view on the bottom left. The structure view called AUOT (argument unit occurrence tree) on the top left provides an aggregation of all essay structures currently selected for further analysis of argumentation patterns. Each node of the AUOT shows the fraction of essays with a major claim at the node's position in blue or the fraction of pro and con claims or premises in green and red.
It is clearly visible that all essays in the selection have at least two claims (with con fractions between a quarter and a fifth), most of them three, less than half four and very few even five claims. Below that, partially filled nodes are displayed indicating more diverse structures on the premises level.

Major Claim:
Rick will get seriously sunburnt.
- **Claim:**
  People with fair skin, red hair and freckles usually get sunburnt easily.
  - **Premise lvl1:**
    Those people have little melanin in their skin.
    - **Premise lvl2:**
      Melanin protects against sunburn.
  - **Premise lvl1:**
    Rick's parents both have fair skin, red hair and freckles, but they never seem to get sunburnt however much they sit outside.
- **Claim:**
  Rick has fair skin, red hair and freckles, and he sun-bathed all day.

Fig. 2. A sample argument following the argumentation scheme used in this paper. The black lines symbolize the hierarchical relations between the argument units; The labels follow the overall color scheme (Sec. 4): green ⇒ pro unit, red ⇒ con unit, blue ⇒ major claim.

reason to accept or reject the first one. All statements are referred to as *argument units*. *Pro* units support their parent unit; *con* units attack it.

The concatenation of arguments results in an argumentation with an inherently hierarchical structure where the argument units form the nodes and the support/attack relations form the edges. To facilitate the communication with linguistic researchers, we follow the nomenclature of Stab and Gurevych [38]. The root units are called *major claims* and define both the topic and stance of the essay. Their children are called *claims* and serve as starting points for single arguments arguing for or against a major claim; Further units usually contain examples, evidence or facts in favor or against their parent and are called *premises*. Since premises are allowed to relate to other premises, the depth level of a premise needs to be indicated by a level suffix if an unambiguous identification is required in the analysis (e.g. premise lvl2) (see Fig. 2). In simple arguments, the major claim is stated only once. In essays, however, it is not uncommon to repeat or rephrase the major claim during the conclusion such that a hierarchy may possess multiple root nodes that are each connected to all claims of the text.

Only few medium sized, argument-annotated corpora exist: Microtexts [30] (112 texts), Argument Annotated User-Generated Web Discourse [13] (340 texts), or AraucariaDB [33] (662 argument maps). The corpus we focus on is the Argument Annotated Essays 2 (AAE2) corpus [38] of 402 essays written by English students and their prompts from *essayforum.com*. The texts are manually annotated with argument units (major claim, claim or premise), claims' stances (pro, con) and relations between claims and premises (attacking, supporting). We chose the AAE2 since it is the richest with respect to argument structure information. Additionally, we could acquire the original annotations for both steps of the corpus creation process: In the text annotation step, the text passages containing an argument unit are marked and already classified to be a major claim, claim or premise, without taking the stance into account. If multiple annotators are involved, the annotated text chunks are then unified to one common consensus. In the association step, the stances and pro/con relationships are added to the found units. These again are unified to become part of the final corpus.

3 Related Work

Argument structure visualization is a special form of text visualization (see Kucher and Kerren [18] for a survey). Previous work on visualizing argument structures specialized in visualizing, analyzing and synthesizing single argument structures. Most of these run under the name of *Argument Mapping* and utilize node-link diagrams to depict the argumentation: *gIBIS* [6], *ArgVis* [16], *Belvedere* [39], *DebateGraph* [7], and *Dicode* [43] serve as mind maps to support an ongoing discussion. *Araucaria* [33], *ArgueApply* [32], *Argunet* [1], *Rationale* [45], and *Truthmapping* [42] allow for analysis and evaluation of an argument. *Dialectic Map* [28], *SEAS* [23], and *VUE* [2] facilitate the decision making process during a discussion by automatic analysis and visualization. While effective for the tasks they have been created for, they do not support comparisons between multiple texts or consider the order of units in the argument. Wachsmuth et al. [48] provided a static, accumulated argument structure visualization for comparing sets of texts. However, it is very hard to read for the non-expert user and lacks clarity due to clutter and overplotting. Thus, we provide interactive and new simplified visual representations for accumulated argument graphs and to compare the structures of argumentative texts.

The problem of visualizing multiple argument structures can be interpreted as the problem of visualizing multiple trees. In that area, the *DAViewer* [51] facilitates the comparison of the discourse trees created by different discourse parsers using interactive dendrograms. Similarly, Bremm et al. [4] and Munzner et al. [27] find and display similarities and differences in phylogenetic trees. Instead of using small multiples of the full tree like the *DAViewer* and the works of Bremm et al. [4] and Munzner et al. [27], we introduce small multiples of glyph-based tree summarizations accompanied by an accumulated view for detailed comparisons, since Graham and Kennedy [12], who surveyed different strategies of tree visualizations, concluded that agglomeration is the most suitable for comparative tasks. Viewed from another angle, parts

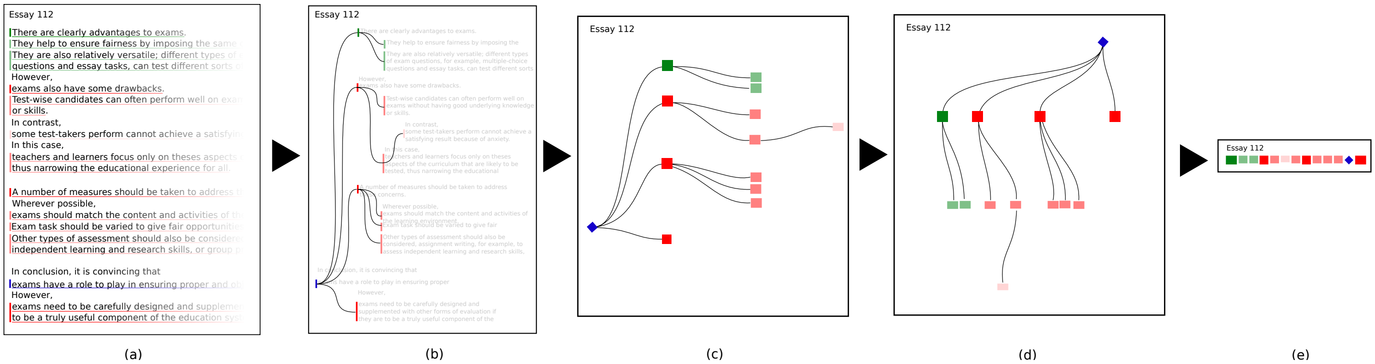


Fig. 3. Construction of an ArguLine. (a) The annotation of a given text contains the identified text chunks as well as their stance and type (pro or con; major claim, claim, or premise). (b) The relations between the text chunks specify a hierarchical argument structure. (c) The text chunks are transferred into glyphs that encode the stance (pro or con) and the type of argument unit (major claim , claim , premise lvl1 , premise lvl>1). (d) Following the mental text flow model of our experts and the orientation of the commonly used node-link diagrams, the graph is rotated; now depicting the text flow horizontally and the depth vertically. In the last step, the tree is vertically compressed into a line while preserving the order of the units within the text.

of an argument structure can also be seen as a sequence of events. Monroe et al. [26] developed a system for sequential pattern analysis in medical records and Cappers et al. [5] explored the possibilities of visualizing event sequences containing multivariate data. Riehmann et al. [34] show the sequence of plagiarism events in a text as DiffLine. All approaches are close to the ArguLine overview visualization, but – like the ArguLines – are not able to represent the hierarchical nature of argument structures in detail. Therefore, our proposed system implements a coordinated multiview approach to present all aspects of the data without overloading a single visualization.

Wachsmuth et al. [46, 47, 49] developed a visualization of the sentiment flow in hotel reviews. Similar to the argument flow in the ArguLines and structure view, it shows the changes in sentiment throughout an argumentation, but fails to present them in a way that allows the comparison of a whole corpus of texts.

Spoken argumentation is analyzed by the VisArgue group of the Konstanz University. They specialize in visualizing transcribed discussions, showing the topical development and deliberative quality of a discussion over time [9, 11], finding patterns in a discussion [15], or summarizing a discussion in a mind map [8]. Outside this group, *Conversation Clusters* [3] summarizes discussions in topical word clouds, and *GroupMeter* [22] supports students to develop argumentation skills. Group discussions, however, often contain multiple intertwined threads with separate argument structures. To develop a first argument structure analysis system, we concentrated on the single-threaded, monological argumentation in essays.

Stance visualization has been studied extensively on other kinds of texts, foremost social media. StanceXplore [25] allows the study of stances in social media over multiple attributes, like time, space, topic and content similarity. *MultiConVis* [14] shows the timeline and stance distribution of discussions in internet forums. DoSVis [19] enables research on stance in longer documents. An extensive overview is given by Kucher et al. [20]. Unlike current stance visualizations that focus on the demography of writers, topics or temporal development, our system allows the study of stance and its role within arguments.

4 Argument Structure Analysis System

Our system for the visual analysis of argument structures is designed to support (computational) linguists and digital humanists in analyzing the argument structure of a corpus of essays. We started our development by deriving an initial list of required functionalities for our system in initial interviews with domain experts about their usual tasks:

Task 1 The task our experts conduct mostly is *finding common structural and temporal patterns* within a given corpus, since it is the most crucial part in studying argumentation. It is also highly relevant in building better heuristics for automatic structure detection. Students might benefit by developing more sophisticated argumentation skills.

Task 2 The second most important task in argumentation studies is *comparing the structures of sets of text* with each other for revealing e.g. the differences between the typical argument structures of authors with different cultural backgrounds. Students might also benefit from visually exploring the differences of well written and debatable essays.

Task 3 Support for *finding texts containing specific or potentially uncommon structures* is crucial for verifying hypotheses about the corpus (e.g. do texts beginning with a con argument argue against the given prompt?) or comparing defined subsets of essays (e.g. all essays with one major claim against all with two or three major claims).

Task 4 To generate meaningful results, the quality of the corpus is of utmost importance. Therefore, means to *detect outliers* and *compare the original annotations* are needed.

Task 5 A system supporting scholarly work needs a way to *read/extract suitable text passages for evaluation or as prove by example.*

Currently none of these tasks are easily manageable using regular node-link diagrams. Especially comparing argument structures is difficult since the layout of the diagrams is not unified and the node sizes depend on the length of the embedded argument unit's text. So, even similar structures tend to look different in traditional tools.

To overcome this deficiency, we developed a system consisting of three interactively coordinated views (see Fig. 1): (1) The ArguLines view on the right summarizes the entire corpus in a Overview and Detail list of ArguLines which represent an essay as a sequence of glyphs, one for each argument unit. The list can be resorted to *find common temporal patterns* or *detect outliers*. Additionally, it provides detail on individually selected essays on demand, e.g. on their *original annotations*. (2) The structure view on the top left allows the analysis of the argument structures of the texts selected in the overview list using the hierarchical histograms of the Argument Unit Occurrence Tree (AUOT). It is used to *find the most common structural patterns* and to *compare the structures of sets of text*. (3) The document view on the bottom left provides means to *read/extract suitable text passages* connected to the abstractions of the other views. Each visualization can be used as a filter to reduce the corpus to *texts containing specific structures*. The ArguLines view is the pivotal visualization in our system since it is used in common tasks which first scrutinize the list of essays for potentially relevant texts and select subsets for closer inspection in the structure and document view. Therefore, the layout was chosen such that the ArguLines have enough space to be used to maximum effect. The ArguLines exploit the whole height of the screen and, since the ArguLines' overview also functions as a scroll bar, is placed on the right side of the screen. The other two views have been arranged such that the structure view typically used in an earlier step of the analysis is on top.

All visualizations follow a system-wide color design to differentiate between different argument units as well as selected subsets. The color scheme for the argument units follows the European standard: green for pro units, red for con units. The major claims – as the units that express the essence of the essay – are separately marked in blue. The lightness is mapped to the depth level within the hierarchical argument structure: the fainter, the deeper. This scheme was chosen as conceptually, the deeper a unit is within the hierarchy, the less influence it has on the overall argumentation. Therefore, it should be less prominent in the visualizations. Unfortunately, this pro/con model is not compatible with the most common form of color blindness (red/green). To cover these cases, the color scheme is customizable via a configuration file. The colors to denote selected subsets were chosen such that they are easily distinguishable among themselves and do not intersect with the set of colors for the argument units.

4.1 Overview: Corpus Summarization with ArguLines

The overview is conceived as list of all essays in a corpus that gives an impression of its scope and of the frequency of certain temporal patterns in the argument structures it contains. Detail is available on demand. Each essay is represented by an ArguLine which abstracts the argument units from the concrete text and compresses the hierarchical structures into a single line which provides a compact impression of the text flow and depth within the argument structure.

An *ArguLine* is constructed by replacing each argument unit with a representative that encodes both its stance and depth level (see Figure 3). Similar to common argument maps, claims and premises translate into rectangles, but with a height related to the depth level and a color that encodes both the stance of the unit in hue (pro ⇒ green, con ⇒ red) and the depth level in lightness (the deeper, the lighter the color). Major claims are treated specially since they compose the most important units in an essay defining the topic and stance of the whole text. In many essays, they figure more than once, mostly within the introduction and/or conclusion. To have them pop out among other units, major claims are redundantly encoded as blue diamonds instead of rectangles. This way, the number and positions of major claims can be retrieved at a glance just as well as the number of overall argument units, the distribution of pro and con units, and the distribution of depth levels. The visualization does not show non-argumentative text pieces, since they do not convey any structurally relevant information. If needed, the information can be shown on demand using the length-proportional detail view of the ArguLine.

The ArguLines reveal overall tendencies, like the blue major claim at the start and end of most essays (see Fig. 4), and subsets of texts with similar structure: Some texts do not contain any con units, some scatter them, and some present them as one block. From these subsets and tendencies, samples of interest can easily be identified and selected for detailed examination either with the other visualizations (Sect. 4.2, Sect. 4.3) or the ArguLine's own detail views: A tree view to provide the full argument structure of a single text, a length-proportional view to analyze the actual positional distribution and lengths of the argument units, as well as an annotation view to review the steps and decisions made during corpus creation if the data is available. The tree view (Fig. 5, b) is brought forward on clicking the ArguLine's info icon. On activation, the ArguLine's glyphs move vertically to their respective depths within the argument hierarchy. Additionally, lines appear to explicitly define the structural relations between the argument units. The explicit argument tree strengthens the mental model of the analyst and resolves uncertainties introduced through the reduction of the full structure into an ArguLine. The length-proportional view (Fig. 5, c, top line) is activated in a similar way for a single ArguLine or via a

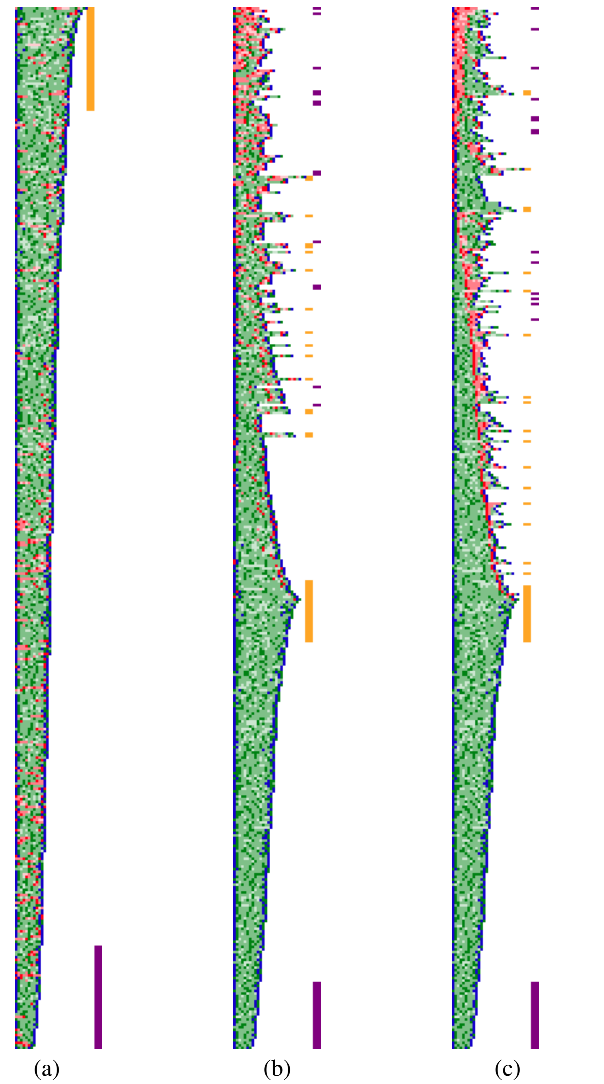


Fig. 4. Different orderings of the list of minimized ArguLines in a typical three-step analysis of the AAE2: (a) ordered by the number of argument units. In (a), the selection of the 40 longest (orange) and 40 shortest (purple) texts was conducted. (b) ordering by the number of con arguments in the text shows long essays (orange) are more seldom purely pro than short ones (purple). (c) ordered by the position of the con units shows short essays (purple) place con units earlier in the text.

control element for the whole corpus at once. The glyphs move to their relative positions in the text and stretch or compress to match the length of the text chunk. Additionally, gray rectangles are included to show non-argumentative text. The positions and lengths are encoded relative to the length of the longest essay showing the length-proportional view, enabling comparisons between texts. Including the unit's length and non-argumentative text revealed that there is often a non-argumentative introductory text and hardly any non-argumentative passages between argument units. It also showed that premises are mostly longer than claims or major claims. The annotation view (Fig. 5, c) shows the originally annotated text chunks of all annotators – in case of the AAE2 corpus: three – in both annotation steps if that data is available. The visual encoding is equivalent to the encoding in the length-proportional view. The annotations of each annotator are shown as one line for each step. The annotation view provides valuable information about the consensus process of the annotators; annotations in a text with many disagreements might be less reliable. The example in Fig. 5 shows that annotators 2 and 3 disagreed on the type of most units while agreeing perfectly on their boundaries during the text annotation step.

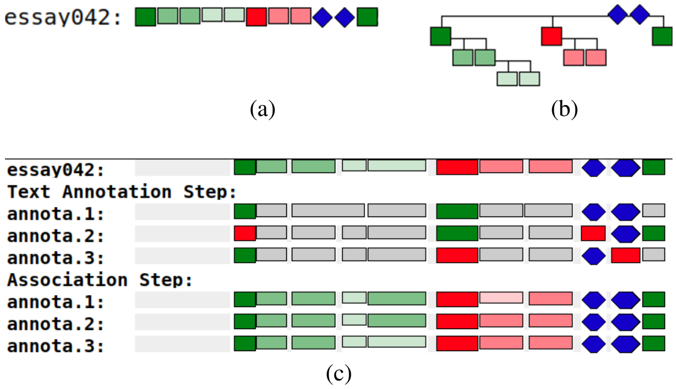


Fig. 5. The (a) regular ArguLine for essay 042, its (b) tree view and its (c) annotation view (including the length-proportional view in the first line) showing segment lengths, text positions as well as the choices of all annotators when marking up the argument structure during the text annotation step and the association step. During the text annotation step, annotator 2 and 3 are in consensus about the positions and lengths of the argument units, but never agree in type. The color code follows the overall color scheme, adding gray for units that do not have a stance assigned yet and light gray for non-argumentative text.

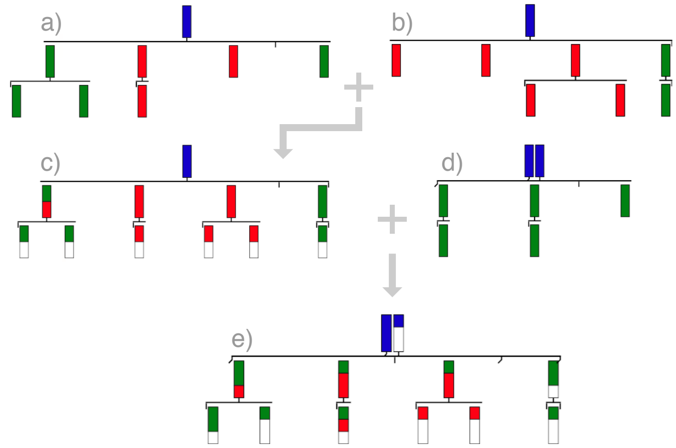


Fig. 6. Construction of the Argument Unit Occurrence Tree (AOUT). The aggregation of the two argument structures a and b by merging positionally equivalent nodes that render the number of overlapping pro and con units at the respective position as one-bar histogram (c). Adding d extends the aggregation resulting in an AOUT (e) surveying the argument structures of 3 essays.

In the association step, the relations are mostly agreed upon; only two disagreements concerning the depth of the fifth and seventh argument unit are visible.

Reordering the list allows for visual clustering of structurally similar texts and the detection of patterns. Fig. 4 shows a typical three-step analysis of the whole AAE2 corpus: in (a) the number of argument units is key to the ordering. In this view also, the selection of the 40 longest (orange selection) and 40 shortest (purple selection) texts has been conducted. The corpus can be reordered by the number of con arguments in the text (b). More than half of the short essays (purple) do not contain any con units, while the fraction of purely pro essays is much smaller in long texts (orange). Interestingly, the texts with the most con units are, nevertheless, within the shortest texts (purple). When reordering by the position of con units within the argumentation(c), the list shows con units show up earlier in short essays (purple).

4.2 Structure View: Aggregation with AUOT

Argument structure views support the researcher in detecting common structural and sequential patterns, analyzing texts containing specific substructures, and comparing subsets of essays against each other. Until now, researchers had to laboriously compare the node-link diagrams from each relevant argument structure with one another to accomplish any of these tasks. So, a view that can aggregate a subset of argument structures across multiple essays would simplify this task considerably.

Early efforts followed the initial suggestion of one of our internal experts to overlay the logical graph structures of all essays visually into a single aggregated graph visualization while maintaining the full sequential position information of each argument unit (not depicted but in principle an improved version of the visualization shown in Fig. 12). However, it led to misinterpretations and did not allow to derive reasonable conclusions for the structural analysis (e.g. for ascertaining the characteristics of a set of texts such as the typical number of major claims/claims/premises, the changes in stance between subsequent claims, etc.).

Thus, we developed the *Argument Unit Occurrence Tree (AUOT)* which uses the new concept of hierarchical histograms and focuses on the hierarchical organization of the aggregated argument structures (see Fig. 6). The histogram of each node in the AUOT, displayed as a rectangular box, indicates the fraction of essays within the selected set of essays that have a node at that respective structural position. The colors of the bars represent the fraction of pro (green) and con (red) units while the white part represents the essays that do not have a node at that particular position. If a bar gets too small to be readable, it is replaced by likewise colored circles, which each represent a text of the corpus and are called countables in the remainder of the paper. To emphasize the hierarchical structure, siblings are connected by a dark horizontal line. Small vertical stubs indicate the location of the parent node in relation to the child nodes and the gray level indicates the probability of that location. Fig. 7 shows two subsets of essays as AUOTs that were previously selected in the ArguLines overview and can now be used to compare a set's structural characteristics. The most prominent difference between the sets is the pro/con ratios of the claims. In the upper, orange subset, all essays start with a con argument (the bar of the first claim is completely filled in red), while the lower, purple subset shows a higher probability of a con claim later in the text (from the third claim on, the red bar is larger than the green one).

The stubs only hint at the parent's most common position. A more detailed analysis of a parent node's position in the texts in relation to its children can be revealed on demand. Right-clicking on a node leads to its replacement by (multiple) slimmer subnodes at the actually existing positions of the node in relation to its children (Fig. 8).

The cross-count ribbons enable the analysis of sequential patterns of changes in stance. They can be drawn between any two nodes in the AUOT. Fig. 9 analyzes the similarity of the stance ratio between the second and third claims in the upper selected subset. One might expect that pro units would be followed by pro units in the next claim, and con units would be followed by con units. Spanning the cross-count ribbons between the pro part of the second claim and the third claim shows that actually, about one fifth of the pro units become con units and a small fraction does not have a third claim at all. Both subnodes and cross-count ribbons have been implemented to only appear on demand since they convey supplementary information that would clutter the visualization otherwise.

4.3 Document View: Text Evaluation

The document view (Fig. 10) allows the user to check for the validity of found patterns in the other views, pick examples for different subsets of essays in the corpus, and verify the annotations of the argument units. It shows the full text of one or more essays on demand at the bottom left of the screen (Fig. 1). Within the text, all argument units are highlighted with an underline and a vertical bar on the left hand side. The color scheme matches the colors used with the ArguLines. The color of the bars encodes the stance of the unit towards the overall stance of the essay (major claim ⇒ blue, pro ⇒ green, con ⇒ red) as well as the hierarchy level (the lighter the deeper in the hierarchy). The visualization can be extended to reveal the original annotations,

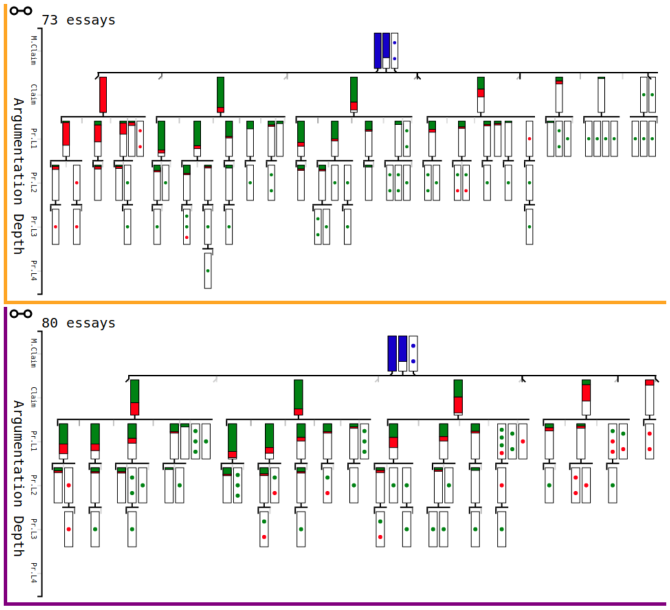
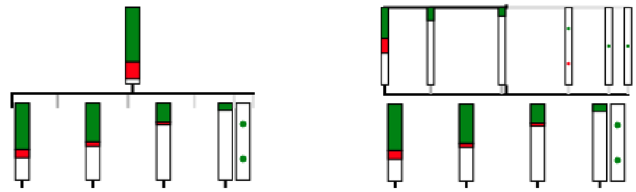
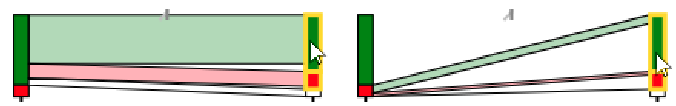


Fig. 7. Comparing 73 essays beginning with con arguments (top) with 80 essays ending with con arguments (bottom) using the AUOT visualization. All essays have at least one major claim (the first node at the top level is fully filled with a blue bar), about three quarters of the essays have a second major claim and two essays have a third one (there are two blue circles in the third node). In the upper AUOT, there is one essay that has 7 claims, while in the lower AUOT, essays have only up to 5 claims.



(a) (b)

Fig. 8. Detail view of the third claim and its children in Fig. 7 (top) (a) showing the normal node and (b) analyzing the probabilities of its positions by replacing it with multiple, slimmer subnodes for each position: the claim occurs most often before all its children but may occur at any other position as well (with far smaller probability).



(a) (b)

Fig. 9. Detail view of the second and third claim in Fig. 7 (top) analyzing the evolution in stance between the two nodes with cross-count ribbons: (a) about a fifth of the pro claims turn into con claims and (b) more than two thirds of the con claims turn into pro claims. The proportion of second claims that do not have a third claim (the white band) is very small (about 5% of the selection) and stems mostly from pro claims (a).

as well (Fig. 10). Besides the vertical bar showing the agreed stance and depth of an argument unit, vertical bars reveal the annotators' markings during both steps of the annotation process. The leftmost cluster of lines shows the choices of the text annotation step and the second cluster shows the annotations of the association step. The text lines are wrapped such that the information provided by a vertical bar on the left applies to the entire corresponding text line. Important characteristics of the original text such as the placement of paragraphs and text flow are still preserved. To lead the user's attention to the interesting annotation cases, we do not show annotation data in sections in which all annotators concurred with each other.

4.4 Visual Queries, Filtering and Linking

One idea of our experts was to shape queries visually to find similar essays corpus-wide. Their discussions mainly revolved around two types of queries: queries regarding the argument's hierarchical structure (e.g. "I want to see all essays of a certain complexity/depth and width", "Show me all essays with at least 2 major claims") and queries regarding the text flow (e.g. "I need all texts starting with a con claim"). Our visualizations are ideally suited for these types of visual queries since they exactly match our two types of visualizations. The AUOT expresses aggregated hierarchical structure, hence, its elements can be used to link and filter essays based on type, stance and structural position. The ArguLines show sequential data, thus allowing for type-, stance-, and sequential position-based queries. An ArguLine expanded to tree view features a mixture of both. As a result, we implemented a consistent set of filtering and linking techniques that acknowledges the different types of encoded information (hierarchical structure and text flow) in our coordinated multiview visualization.

Hovering over an argument unit in the ArguLine view, for example, highlights all matching units at the same sequential position in the text, having the same type (and optionally also the same stance) in all other ArguLines. At the same time, the AUOT and text view highlight the positions of the same argument units in the hierarchical structure and opened texts respectively. On the other side, hovering directly over a particular node (or optionally the pro or con stance inside this node) in the AUOT reveals all argument units sharing the same path through the tree (and optionally the stance of the hovered node) in the ArguLines which highlights the hovered unit's sequential position individually in every ArguLine (Fig. 11) and the corresponding text parts in the text view. Both highlighting modes affect the ArguLines overview which slightly fades all non-matching units so that the number of essays containing a matching unit can be estimated at a glance. To quickly find the very ArguLine that corresponds to a text or connect the text of an argument unit to its structural position, hovering over any of text units in the text view will highlight that particular unit in its ArguLine as well as in the AUOT.

The highlight on hover also serves as a preview to filtering (which is invoked by clicking an argument unit's representative), generally removing all non-matching ArguLines and morphing the AUOT to display only the subset of matching essays. Multiple filters can be applied on top of each other to further narrow down the set of essays. Each filtering step can be undone any time, by clicking on the filtering argument unit's representative (marked by a thicker black outline). An inverse filter (removing all essays that contain the particular argument unit) and a selection-specific filter (filtering only within a selected subset of texts) complete the implemented analytics toolset.

The visual filtering system allows for shaping complex queries by simply clicking the structural elements that should or should not be part of the subset. Using comparable filtering and querying systems like SQL, finding documents that figure e.g. only one major claim and a con claim at first text position would require a complex SQL statement involving several GROUP BY and WHERE clauses while using the AUOT and ArguLines only 3 clicks are necessary: one on the second major claim (while holding the control key to activate the inverse filter), one to show the ArguLines tree view, and one on the con claim at the first position in the tree view.

5 Expert Reviews and Domain Findings

The developments took place based on a user-centered approach, in which two local experts evaluated the status quo at regular intervals and provided feedback for further development. After completion of the main functionalities, we additionally asked three external experts, who were familiar with the used corpora, to evaluate our system. Subsection 5.1 describes the feedback about the system itself and subsection 5.2 provides an overview of the most relevant linguistic findings revealed in these sessions.

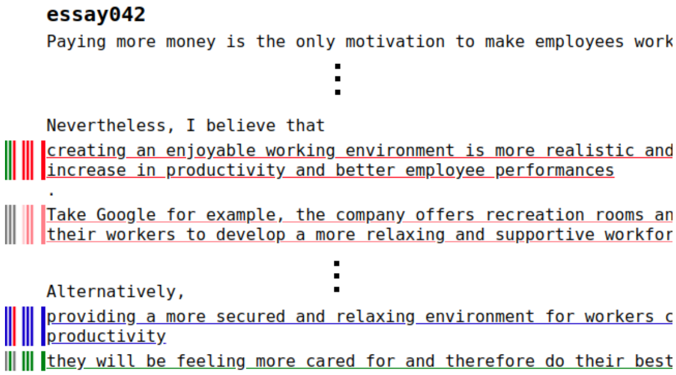


Fig. 10. The document view of essay042 with some text cut off at the right side. Within the text, argument units are underlined. On the left, vertical bars show the markup of the three annotators in the two annotation steps and the resulting consensus. In this example, disagreement among annotators occurred with respect to stance and depth of argument units.

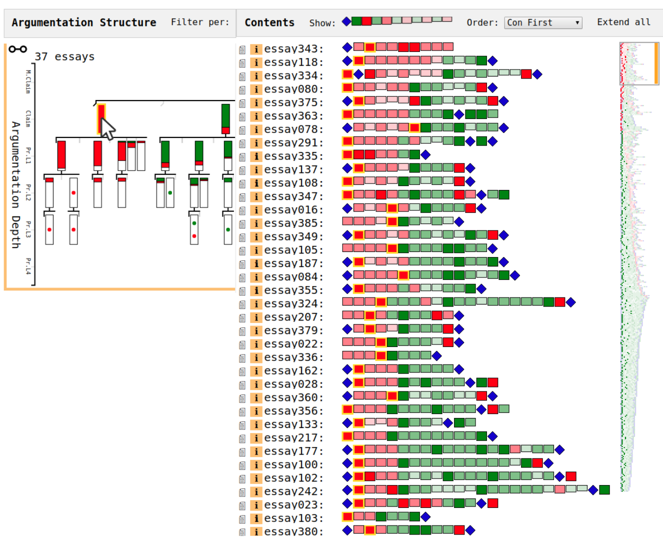


Fig. 11. Hovering over a node in the AUOT reveals all argument units sharing the same path through the tree in the ArguLines.

5.1 User-centered Design and External Reviews

For the user-centered approach, we conducted regular reviews with two local experts during development (at least five times with each expert). Both are very experienced researchers working as computer scientists in computational linguistics and text retrieval with more than 15 respectively 10 years of experience.

After the dissatisfaction regarding early prototypes showing aggregated structures (not depicted but in principle improved versions of the visualization shown in Fig. 12) the introduction of the ArguLines was a turning point in the third round of discussions – for our experts as well as for ourselves. The experts stated that the ArguLines are far better suited for the task of getting an overview of a corpus. There was even a revision that arranged the Argulines in a Focus and Context approach, but we decided for the Overview and Detail layout in the end since it was preferred by the experts. However, the ArguLine Overview and Detail could not reveal all relations in a corpus that our experts were interested in, e.g. whether connections exist across different argument substructures. For instance, the graph visualization by Wachsmuth et al. [48] for the AAE2 corpus indicated that premises were locally clustered around their claims and the premises of an argument did not cross-reference the claim or premise of another argument. However, the legibility of this graph drawing severely suffers from clutter and overplotting. In order to investigate such relations better, we developed the AUOT visualization (Section 4.2).

For evaluating the final system, we conducted expert reviews with three independent experts: a computer scientist working with the group that compiled the AAE2 corpus [38], a linguist who was involved in gathering the Microtexts corpus [30] and a second very experienced linguist also familiar with both corpora. All of them are familiar with computer-assisted research and use basic visualizations in their daily work. During the reviews, the visuals of each part of the interface were explained separately and directly demonstrated with sample tasks that they are typically faced with in their research, like confirming or contradicting hypotheses they had about the corpus. The gained insights were discussed before moving on to the next feature. To a certain extent, the responses were similar. All of them were very enthusiastic about looking at all essays and their argument structures in a corpus at once using the ArguLine overview. They used the different sorting options extensively (Section 4.1) and were fond of the patterns and argumentation strategies (Task 1 in Section 4) that were revealed as a result. One expert stated: “This is really great, I can immediately see the distribution of pro and con-arguments.”, one expert declared enthusiastically, “You can see very well how different people argue. Some keep themselves short, try to be succinct, and yet illuminate both sides, others like to hear themselves talk and write endlessly only about their own point of view.” The corpus overview also allowed the expert to visually filter the texts based on the shown structural patterns (Task 3), hence he continued: “As a person with little time, I can immediately see which texts I would read. I would leave out the long, one-sided ones from the beginning.” meaning long texts without any con arguments at all; a large fraction of the entire corpus. Another expert stated: “I would primarily use those [the ArguLines]. I have the sequence and with the icons the argumentational depth as well.” One wished for being able not only to sort but also to group the ArguLine list by entering his own criteria.

The pro/con colors were obvious to all of them. Two were very fond of mapping the depth of the tree nodes as color shades and thus revealing the level of importance of an argument unit. The two linguists also immediately understood the potential of the AUOT for comparing individual subsets of the corpus against each other with respect to pro/con patterns. Here, one suggested being capable of generating subsets by providing a criterion (see sorting above) instead of selecting them individually in the list. For the computer scientist, the pixel-based overview was more revealing since he realized immediately how large the fraction of essays in the AAE2 corpus really was that did not contain a single con argument. For investigating this further, we immediately had a closer look at some of the wordings in the texts guided by the argument unit linking provided across the visualizations. All experts considered it essential to have direct and instant access to the text of the argument units.

Overall, they stated that our system would facilitate the process of deriving and verifying hypotheses. For instance, the experts told us that the usual way of checking hypotheses in linguistics involved a lot of tedious manual work. They would look at individual texts or, if available, so-called RST trees (Rhetorical Structure Theory by Mann [24]). In both cases, they would have to deal with multiple independent application windows while we provide linked views in our system. Furthermore, one of them suggested that there is a large potential in extending our ArguLine concept for providing an overview of all RST trees [29] of a corpus at once. This expert also would like to use the system for teaching linguistics and to show students the particular features of the Microtexts corpus.

Even though the five final expert reviews confirm the utility of our system, we are aware that humanists who are not used to computer-assisted research and visualization systems may need a longer training phase than our experienced researchers. However, the suggestion by one of our external experts to use the tool for teaching linguistics shows that the expert felt that it should be quickly learnable.

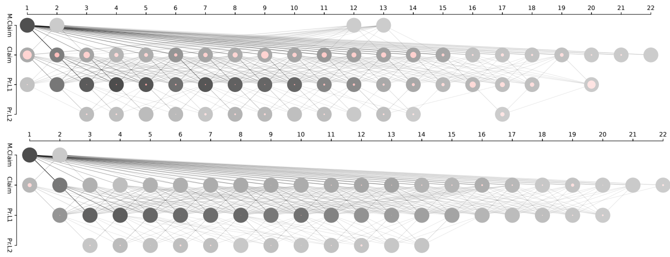


Fig. 12. Comparing structures of texts with myside bias (bottom) with those without bias (top) with the early visualization concept used in Wachsmuth et al. [48]. In this version, argument trees are laid out in their sequential order (identical to the tree view of the ArguLine) and plotted on top of each other, aggregating units of the same position and depth. The lightness of the circles representing the nodes indicates the number of units in the aggregation; the darker the more units. Some nodes have light red inner circles whose area compared to the overall area of the node depicts the ratio of con units at this position. Nodes with very few occurrences are removed.

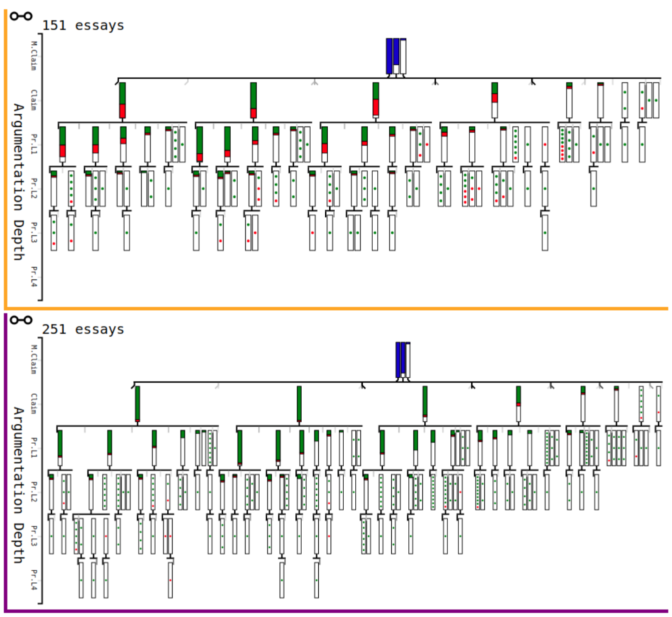


Fig. 13. Comparing the same structures as in Fig. 12 (with myside bias at the bottom, without myside bias on top) with the proposed AUOT. The argument's hierarchy is far better visible, the number of each type of unit (major claims, claims, premises) countable and the frequency of each unit position interpretable. However, positional information is less present in this visualization.

5.2 Linguistic Findings

Using our system resulted in several intriguing insights about and into the AAE2 corpus [38] that were previously undiscovered by our experts – despite their thorough investigation and knowledge of the corpus.

With the computer scientists who used the initial version of the ArguLines, we observed different patterns (Task 1) in the sequences and the argument structure of the essays in the corpus. A more thorough investigation using the AUOT's filtering techniques revealed e.g. that 18% of the essays have 1 major claim, 78% have 2 and 4% have 3; 57% contain at least one con unit and 43% do not; the pro/con ratio is exactly balanced for only 2.5% of the essays, whereas 92.5% lean towards pro and 5% towards con; about 75% of the essays start with a major claim, 10% with a con argument unit and 15% start with pro. The first claim, however, is pro in 82% of the corpus, and con only in the remaining 18%. None of these subsets with their particular properties became apparent in prior analyses of the corpus.

By analyzing the pro/con ratio of essays (Task 1), together with the expert who was involved in compiling the AAE2, we found that 173 of the 402 essays contain heavy cases of myside bias [31] and do not contain any con argument units at all. Cases of less heavy myside bias can be studied using the categorizations of Stab and Gurevych [37] and Wachsmuth et al. [48], where an essay has myside bias if it does not have a con claim in its main body (meaning without the first and last paragraph). Comparing the cases of myside bias according to that categorization to all essays without myside bias (Fig. 13) reveals a similar general structure: both sets always have at least one major claim, at least two claims and the odds of having a second/third major claim or third/fourth/fifth/etc. claim are about the same. Identical to the original conclusions drawn from the original visualization (Fig. 12) by Wachsmuth et al. [48], the visualization suggests the first major claim to be usually stated before its claims in both sets. However, the AUOT allows the additional insight that an essay may contain multiple major claims and how likely it is to encounter a second and third one. Furthermore, Wachsmuth et al. [48] found that should the major claim not be in first position, the text is likely to start with a con argument; the AUOT, however, allows for statements about the actual probabilities: about 3/4 if the major claim is placed after the first claim, about 1/3 when it is after the second, and about 2/3 when after the third. Furthermore, the ArguLine list clearly shows that the con sections of most texts cluster either at the start or end of the essay in the vicinity of a major claim.

An analysis of the subsets with respect to the number of major claims (Task 1&2) was also conducted with one of the linguistics experts. It revealed that essays with only one major claim place it towards the end of the essay summarizing the point of view of the writer. The preferred position of the first major claim – it may be repeated up to two times throughout the text – across the entire corpus is the very beginning to introduce the stance of the author along with the topic of discussion. The interplay between the argument units in the ArguLines and their textual counterparts explored in the expert reviews lead to the insight that the repetition of major claims was in most cases not done by simple rephrasing as one might expect. Instead, the restatement of a major claim is typically more general, more specific or raises another but strongly related point, e.g. "[...], we cannot ignore its [meaning the introduction of machines and technology] negative effects" and "[...], the problems posed by machines should draw more attention, as well".

Investigation of the order of claim stances during the argumentation (Task 1) allowed interesting insights as well: Using the cross-count ribbons between subsequent claims, we found a surprisingly high probability for a change in stance between them (pro→con: $\sim$20%, con→pro: $\sim$80% between first and second claim, pro→con: $\sim$50%, con→pro: $\sim$50% between second and third claim, etc.). This pattern of changing stances, as one of the linguists explained, is called *thesis-antithesis pattern*. Furthermore, the last claim of an argumentation often presents a con stance. As it seems, the students liked to close with a weak con argument to further elevate the major claim.

The detailed visualization of the annotation process (Task 4) with the same linguist revealed the attention the authors of the corpus paid to annotator disagreements. With three annotators, one would expect to see a majority vote after each annotation step. Instead, in more than half of the cases, the vote of one annotator was further used or even a completely new structure was derived through discussions with all three annotators and became the agreed one. In general, we found a number of annotation inconsistencies during software development and expert reviews. One case that would have been difficult to find if it was not for the ArguLines and their sorting features (here *pro/con ratio*) was an essay containing only con arguments without any pro argument at all (). One linguistics expert decided to examine the origins of this outlier and took a closer look at the corresponding essay (Task 5). Without the need to read the entire essay, the explicit highlighting of the argument units (see Figure 10) in the text view and the interactive linking to their counterparts in the ArguLine was crucial in revealing that from the second claim on the stance annotations of this essay were incorrect.

6 DISCUSSION

One major design challenge, we were faced with was the orientation of the graph view of an essay in the ArguLines and the AUOT. While we, as text visualization researchers, wanted to lay out the nodes representing argument units from top to bottom and the argument hierarchy from left to right following the idea of a table of contents or indented tree, our experts in computational argumentation insisted on laying out the argument units from left to right following the reading direction and the argument hierarchy from top to bottom to encode argument depth following the layout of the commonly used node-link diagrams. While it meant giving up the possibility to align the actual text with the corresponding visual elements in the graph visualization and adding another step to the construction of both the AUOT and the ArguLines (Fig. 3), we could not convince our experts and implemented our visualizations following their mental model.

We used a corpus consisting of 402 essays which is the largest corpus available that has a detailed markup and categorization of argument structures and contains even details about the annotation process. There are other corpora that provide nearly the same annotation detail (see Section 3), but they are even smaller – probably due to the annotation effort needed to create high-quality argument structure annotations. While the used corpus is somewhat a sweet spot for our visualization techniques, our system is clearly limited with respect to the size of the corpus, the total number of argument units per essay, and the maximal depth of the argument trees. For much larger corpora, the scrollable overview of the ArguLines poses a limitation while the AUOT could easily aggregate many more than 402 essays. However, if the corpus is very inhomogeneous, many nodes could be only sparsely populated. Very long essays with many argument units will lead to quite small nodes in the AOUT and are also a challenge for the ArguLine visualization. However, interactively limiting the depth of the argument units and pruning or replacing subtrees with appropriate glyphs could make the system quite scalable.

There was some discussion about the countables that were appreciated and used by two of our experts since they were very interested in outliers with an unconventional argument structure, but disliked by the third who focused more on patterns occurring in larger subsets of essays. His point was comparing the countables of one node with the bars inside another node is barely possible, as it would require scaling the bars to numbers by using the size of the selected subset of the corpus. However, the countables could be just ignored, turned off or the respective nodes pruned since those nodes would appear to be empty otherwise. In general, using different encodings in a single visualization has certainly usability issues but our experts immediately understood the concept and at two of the three made good use of it.

Unfortunately, the countables do no scale well to a larger number of essays. They are at their best within the subitizing range [17] within which the number of items is known at a glance without conscious counting; so, up to 4 pro and 4 con countables per node. Beyond that range, the number of countables is limited by the space restrictions of the containing node. Depending on the maximum depth of the depicted argument structures, a maximum of about 20 countables can be displayed without becoming visually too close to the bar representation. While being sufficient for the AOUT visualization of a subset of the 402 essays of our corpus, it does not scale to much larger corpora. A possible solution to the problem might be the introduction of aggregated countables that represent more than one essay at once. However, having three or more different visual encodings will require some cognitive effort to derive and compare the corresponding quantities.

7 CONCLUSION AND FUTURE WORK

We presented a visual analytics system for exploring, analyzing and comparing argument structures in essay corpora. The ArguLines list serves as an overview providing a short structural summary of both the individual documents and the entire corpus. Different orderings of the list ease the identification and analysis of subsets and distributions of argument units inside the corpus. These subsets can be analyzed in detail in the aggregated structure view using the Argument Unit Occurrence Tree. Cross-count ribbons revealing the evolution of stance between argument units, subnodes indicating the relative positions of claims with respect to their premises and hierarchical histograms enable detailed analysis of the differences and similarities of different subsets of documents. The document view provides examples of scholarly work and the necessary textual basis to verify conclusions from the other views as well as the annotation process. Our expert reviews revealed detailed and unknown information about the argument structures of the essays including the use of major claims and their relation to supporting and attacking claims and premises, cases of myside bias, the use of the thesis-antithesis pattern and the detection of outliers.

Using our system, we discussed the analysis of structure, expressed by hierarchical relations, the distribution of pro and con arguments and their order in a document. This already profound analysis and discussion can be further extended. One possible direction could be comparing the typical argument structures of authors with different cultural backgrounds. Does a typical Chinese argumentation look different than an American or Russian? In which way do they differ and why? Similarly, one can study argument structures under the aspect of different topics. Do certain topics entail a certain argumentation strategy? Which topics do, which do not?

In order to allow a deeper analysis of the semantics within an argumentation, the system can also be extended to not only depict stance, but also the function an argument unit has within the argumentation: examples, conclusions, restatements, pieces of evidence, counter-argument, etc. The extended classification would allow studying effective strategies and common fallacies. Encoding established schemes, e.g. arguments from examples, expert opinion or evidence, as glyphs to summarize argument structures of long, rich argumentative texts and to examine them from a different perspective is also an exciting challenge. For a quicker assessment of diverging opinions of annotators during the annotation or the confidence level of an automatic structure classifier, the ArguLines could be extended to show uncertainty information, e.g. by distorting, blurring or morphing the glyphs accordingly.

The application of the system is not restricted to the study of argument structures for (computational) linguistics. One promising direction is visual tutoring in the humanities or in journalism in order to teach the different styles of argumentation, their pros and cons. With future improvements of automatic argumentation recognition, it would be even possible to import the essays of students at an instant and comparing them with best practices in class.

Potential further use cases came up during the expert review sessions, such as visually evaluating classification results by selecting correctly and wrongly classified samples as two separate subsets in our system. This way, the classification process gets more transparent and can be corrected to yield better results. Another suggestion was compiling special purpose corpora from samples of existing corpora. The analyst would set some properties necessary for the planned research – such as: "there must be a counter argument for each argument" or "the text needs to include at least three claims that each have at least two premises" – to create a new corpus from all matching documents. We believe that these and other use cases will drive the further development of our system for visual argumentation analysis in essay corpora.

ACKNOWLEDGMENTS

This work was supported in part by the German Federal Ministry of Education and Research (BMBF) under grant 03PSIPT5A (project Provenance Analytics) as well as by the Thuringian Ministry of Economy, Science and Digital Society in form of a scholarship.

REFERENCES

[1] Argunet. http://www.argunet.org/, 2018.

[2] P. Baroni, M. Romano, F. Toni, M. Aurisicchio, and G. Bertanza. Automatic evaluation of design alternatives with quantitative argumentation. *Argument & Computation*, 6(1):24–49, 2015. doi: 10.1080/19462166.2014.1001791

[3] T. Bergstrom and K. Karahalios. Conversation clusters: Grouping conversation topics through human-computer dialog. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pp. 2349–2352. New York, NY, USA, 2009. doi: 10.1145/1518701.1519060

[4] S. Bremm, T. von Landesberger, M. Heß, T. Schreck, P. Weil, and K. Hamacherk. Interactive visual comparison of multiple trees. In *IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 31–40. IEEE, 2011. doi: 10.1109/VAST.2011.6102439

[5] B. C. Cappers and J. J. van Wijk. Exploring multivariate event sequences using rules, aggregations, and selections. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):532–541, 2017. doi: 10.1109/TVCG.2017.2745278

[6] J. Conklin and M. L. Begeman. gibis: A hypertext tool for exploratory policy discussion. *ACM Transactions on Information Systems*, 6(4):303–331, 1988. doi: 10.1145/58566.59297

[7] Debategraph. `https://debategraph.org`, Accessed 2018.

[8] M. El-Assady, V. Gold, C. Acevedo, C. Collins, and D. Keim. Contovi: Multi-party conversation exploration using topic-space views. *Computer Graphics Forum*, 35(3):431–440, 2016. doi: 10.1111/cgf.12919

[9] M. El-Assady, R. Sevastjanova, B. Gipp, D. Keim, and C. Collins. Nerex: Named-entity relationship exploration in multi-party conversations. *Computer Graphics Forum*, 36(3):213–225, 2017. doi: 10.1111/cgf.13181

[10] J. B. Freeman. *Argument Structure: Representation and Theory*, vol. 18 of *Argumentation Library*. Springer, 2011. doi: 10.1007/978-94-007-0357-5

[11] V. Gold, C. Rohrdantz, and M. El-Assady. Exploratory Text Analysis using Lexical Episode Plots. In *Eurographics Conference on Visualization (EuroVis)*, pp. 85–89, 2015. doi: 10.2312/eurovisshort.20151130

[12] M. Graham and J. Kennedy. A survey of multiple tree visualisation. *Information Visualization*, 9(4):235–252, 2010. doi: 10.1057/ivs.2009.29

[13] I. Habernal and I. Gurevych. Argumentation mining in user-generated web discourse. *Computational Linguistics*, 43(1):125–179, 2017. doi: 10.1162/COLI_a_00276

[14] E. Hoque and G. Carenini. Multiconvis: A visual text analytics system for exploring a collection of online conversations. In *Proc. of the 21st International Conference on Intelligent User Interfaces*, IUI '16, pp. 96–107. New York, NY, USA, 2016. doi: 10.1145/2856767.2856782

[15] W. Jentner, M. El-Assady, B. Gipp, and D. A. Keim. Feature Alignment for the Analysis of Verbatim Text Transcripts. *EuroVis Workshop on Visual Analytics (EuroVA)*, pp. 13–17, 2017. doi: 10.2312/eurova.20171113

[16] A. Karamanou, N. Loutas, and K. Tarabanis. Argvis: structuring political deliberations using innovative visualisation technologies. *Electronic Participation*, pp. 87–98, 2011. doi: 10.1007/978-3-642-23333-3_8

[17] E. L. Kaufman, M. W. Lord, T. W. Reese, and J. Volkmann. The discrimination of visual number. *The American journal of psychology*, 62(4):498–525, 1949. doi: 10.2307/1418556

[18] K. Kucher and A. Kerren. Text visualization techniques: Taxonomy, visual survey, and community insights. In *2015 IEEE Pacific Visualization Symposium (PacificVis)*, vol. 1, pp. 117–121, 2015. doi: 10.1109/PACIFICVIS.2015.7156366

[19] K. Kucher, C. Paradis, and A. Kerren. Dosvis: Document stance visualization. In *Proc. of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 2: IVAPP*, pp. 168–175, 2018. doi: 10.5220/0006539101680175

[20] K. Kucher, C. Paradis, and A. Kerren. The state of the art in sentiment visualization. *Computer Graphics Forum*, 37(1):71–96, 2018. doi: 10.1111/cgf.13217

[21] W. Kunz and H. W. J. Rittel. *Issues as elements of information systems*. Berkley: Institute of Urban and Regional Development, University of California, 1970.

[22] G. Leshed, J. T. Hancock, D. Cosley, P. L. McLeod, and G. Gay. Feedback for guiding reflection on teamwork practices. In *Proc. of the 2007 International ACM Conference on Supporting Group Work*, GROUP '07, pp. 217–220, 2007. doi: 10.1145/1316624.1316655

[23] J. D. Lowrance, I. W. Harrison, and A. C. Rodriguez. Structured argumentation for analysis. In *Proc. of the 12th International Conference on Systems Research, Informatics, and Cybernetics*, pp. 47–57, 2000. doi: 10.1080/19462166.2013.869764

[24] W. C. Mann and S. A. Thompson. Rhetorical structure theory: Toward a functional theory of text organization. *Text & Talk*, 8(3):243 – 281, 1988. doi: 10.1515/text.1.1988.8.3.243

[25] R. M. Martins, V. Simaki, K. Kucher, C. Paradis, and A. Kerren. Stancexplore : Visualization for the interactive exploration of stance in social media. In *Proc. of 2nd Workshop on Visualization for the Digital Humanities*, 2017.

[26] M. Monroe, R. Lan, H. Lee, C. Plaisant, and B. Shneiderman. Temporal event sequence simplification. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2227–2236, 2013. doi: 10.1109/TVCG.2013.200

[27] T. Munzner, F. Guimbretière, S. Tasiran, L. Zhang, and Y. Zhou. TreeJuxtaposer: scalable tree comparison using Focus+Context with guaranteed visibility. *ACM Transactions on Graphics (TOG)*, 22:453–462, 2003. doi: 10.1145/1201775.882291

[28] H. Niu. *Pedagogical efficacy of argument visualization tools*. PhD thesis, Simon Fraser University, Faculty of Education, 2016.

[29] M. O'Donnell. Rsttool 2.4: a markup tool for rhetorical structure theory. In *Proc. of the 6th European Workshop on Natural Language Generation*, pp. 253–256, 2000. doi: 10.3115/1118253.1118290

[30] A. Peldszus and M. Stede. An annotated corpus of argumentative microtexts. In *Proc. of the 1st European Conference on Argumentation*. College Publications, 2016.

[31] S. Plous. *The psychology of judgment and decision making*. Mcgraw-Hill Book Company, 1993.

[32] J. Pührer. Argueapply: A mobile app for argumentation. In *Proc. of 14th International Conference on Logic Programming and Nonmonotonic Reasoning*, pp. 250–262. Springer, 2017. doi: 10.1007/978-3-319-61660-5_23

[33] C. Reed and G. Rowe. Araucaria: Software for argument analysis, diagramming and representation. *International Journal on Artificial Intelligence Tools*, 13(04):961–979, 2004. doi: 10.1142/S0218213004001922

[34] P. Riehmann, M. Potthast, B. Stein, and B. Froehlich. Visual assessment of alleged plagiarism cases. *Computer Graphics Forum*, 34(3):61–70, 2015. doi: 10.1111/cgf.12618

[35] J. R. Searle. *Speech Acts*. Cambridge University Press, 1969.

[36] W. Sinnott-Armstrong and R. J. Fogelin. *Understanding Arguments: An Introduction to Informal Logic*. Cengage Learning, 9 ed., 2013.

[37] C. Stab and I. Gurevych. Recognizing the absence of opposing arguments in persuasive essays. In *Proc. of the Third Workshop on Argument Mining (ArgMining2016)*, pp. 113–118, 2016. doi: 10.18653/v1/W16-2813

[38] C. Stab and I. Gurevych. Parsing argumentation structure in persuasive essays. *Computational Linguistics*, 43(3):619–659, 2017. doi: 10.1162/COLI_a_00295

[39] D. Suthers, A. Weiner, J. Connelly, and M. Paolucci. Belvedere: Engaging students in critical discussion of science and public policy issues. In *Proc. of the 7th World Conference on Artificial Intelligence in Education*, pp. 266–273, 1995.

[40] S. A. Thompson and W. C. Mann. Rhetorical structure theory. *IPRA Papers in Pragmatics*, 1(1):79–105, 1987. doi: 10.1075/iprapip.1.1.03tho

[41] S. Toulmin. The uses of argument. *Cambridge: Cambridge UP*, 1958.

[42] Truthmapping. `https://www.truthmapping.com`, Accessed 2020.

[43] M. Tzagarakis and N. Karacapilidis. On the exploitation of semantic types in the visualization of complex argumentative discourses. In *Proc. of the 2nd International Workshop on Intelligent Exploration of Semantic Data*, p. 3. ACM, 2013. doi: 10.1145/2462197.2462200

[44] F. H. van Eemeren and R. Grootendorst. *A Systematic Theory of Argumentation: The Pragma-Dialectical Approach*. Cambridge University Press, 2004.

[45] T. Van Gelder. The rationale for rationale™. *Law, Probability & Risk*, 6(1-4):23–42, 2007. doi: 10.1093/lpr/mgm032

[46] H. Wachsmuth. *Text Analysis Pipelines—Towards Ad-hoc Large-scale Text Mining*, vol. 9383 of *Lecture Notes in Computer Science*. Springer, 2015. doi: 10.1007/978-3-319-25741-9

[47] H. Wachsmuth, J. Kiesel, and B. Stein. Sentiment Flow—A General Model of Web Review Argumentation. In *Proc. of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP 2015)*, pp. 601–611, 2015. doi: 10.18653/v1/D15-1072

[48] H. Wachsmuth, G. D. S. Martino, D. Kiesel, and B. Stein. The Impact of Modeling Overall Argumentation with Tree Kernels. In *Proc. of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP 17)*, pp. 2369–2379, 2017. doi: 10.18653/v1/D17-1253

[49] H. Wachsmuth, M. Trenkmann, B. Stein, and G. Engels. Modeling Review Argumentation for Robust Sentiment Analysis. In *Proc. of the 25th International Conference on Computational Linguistics (COLING 2014)*, pp. 553–564, 2014.

[50] D. Walton. *Argumentation Schemes for Presumptive Reasoning*. L. Erlbaum Associates, 1996.

[51] J. Zhao, F. Chevalier, C. Collins, and R. Balakrishnan. Facilitating discourse analysis with interactive visualization. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2639–2648, 2012. doi: 10.1109/TVCG.2012.226