

# A Question-Answering System Using Argumentation

Emanuela Moreale and Maria Vargas-Vera

Knowledge Media Institute, The Open University,  
Walton Hall, Milton Keynes MK7 6AA, England  
{E.Moreale, M.Vargas-Vera}@open.ac.uk  
<http://kmi.open.ac.uk/>

**Abstract.** This paper presents a novel approach to question answering: the use of argumentation techniques. Our question answering system deals with argumentation in student essays: it sees an essay as an answer to a question and gauges its quality on the basis of the argumentation found in it. Thus, the system looks for expected types of argumentation in essays (i.e. the expectation is that the kind of argumentation in an essay is correlated to the type of question). Another key feature of our work is our proposed categorisation for argumentation in student essays, as opposed to categorisation of argumentation in research papers, where - unlike the case of student essays - it is relatively well-known which kind of argumentation can be found in specific sections.

## 1 Introduction

A new line of research in Question Answering is the use of knowledge in question answering (roadmap, Symposium on New Directions on Question Answering, Stanford University, spring 2003). This knowledge – which might be encoded in ontologies – would, in our view, enhance the question answering process. Such a research direction has been already taken by the AQUA project [1][2] at the Open University, England. AQUA makes extensive use of knowledge (captured in an ontology) in several parts of the question answering process, such as in query reformulation and in its similarity algorithm (assessing similarity between name of relations in the query and in the knowledge base). Currently, AQUA is coupled with the AKT reference ontology <sup>1</sup>, but it will eventually be able to handle several different ontologies.

This paper proposes a somewhat different approach to question answering: here we use argumentation for finding answers in the specific domain of student essays. This means that specific categories of argumentation depend on the type of question. Current work is on how argumentation could be complemented with a reasoning system which will be able to decide on action plan in case an answer is not found. We

---

<sup>1</sup> The AKT reference ontology contains classes and instances of people, organizations, research areas, publications, technologies and events. (<http://akt.open.ac.uk/ocml/domains/akt-support-ontology/>)

also make use of knowledge in advising students about missing categories in the essays just like in Expert Systems.

Our test bed is a set of postgraduate student essays – a type of free or non-structured text – and corresponding essay questions. Therefore, in our domain, argumentation cannot be found in specific sections of the text (as in research papers).

The first contribution of this paper is the use of argumentation techniques in the question answering problem, as opposed to conventional approaches to question answering such as information retrieval. Our second contribution is our argumentation categorisation for free text (student essays). This is loosely based on research in argumentation in academic papers, but omits categories that are not applicable to this domain. More details on our categorisation can be found in section 1 and [3].

The paper is organised as follows: section 2 presents the question answering process model. Section 3 discusses the research background on argumentation schemas in papers and argument modelling and then introduces our essay metadiscourse categorisation in the context of the reviewed background. Section 4 reports on our annotation categories and essay questions. Section 5 describes our testbed and actual matching of argumentation with essay questions. Section 6 reports preliminary results and indicates future work. Finally, section 7 draws our conclusion.

## 2 Question Answering Process Model

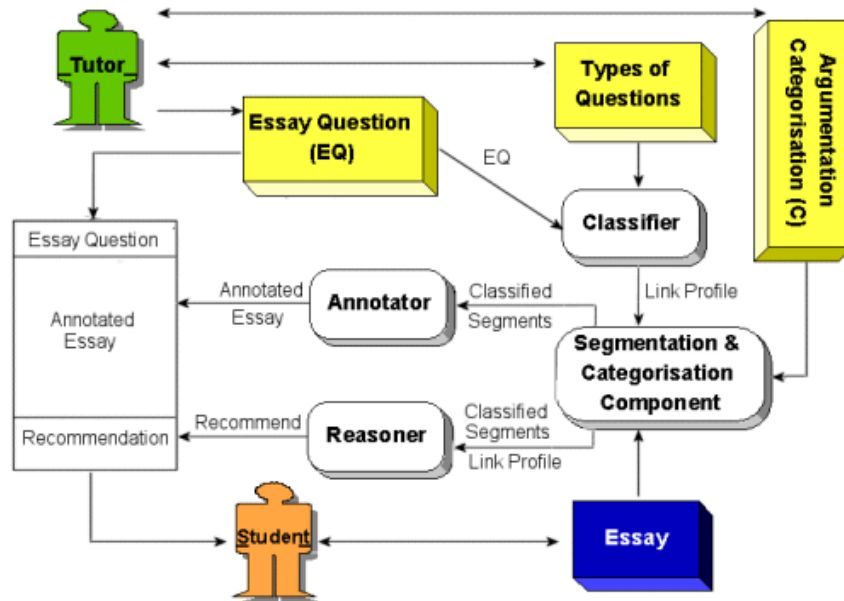
The proposed architecture of our system comprises (Figure 1): interface, classifier, segmentation and categorisation component, annotation and reasoner modules.

- The interface (box on the left in Figure 1) is a window menu interface in which essay question, annotated essay and recommendation are presented.
- The classifier (or query classification module) classifies queries as belonging to one of the types defined in our system.
- The segmentation and categorisation component has two functionalities: it obtains segments of student essays by using a library of cue phrases and patterns and it classifies the segments with respect to one of our categories.
- The annotation module annotates relevant phrases as belonging to one of our defined categories. These annotations are saved as semantic tags. Future implementation may use machine learning for learning cue phrases.
- The reasoner is an expert system that will reason about categories found in a student essays.

Our question answering tool, the *Student Essay Viewer (SEV)*, highlights instances of our argumentation categories in an essay, so as to give a visual representation of argumentation within an essay, in a shallow version of “making thinking visible”[4][5]. The intuition is that essays with considerably more “highlighting” contain more argumentation (and actual “content”) and therefore attract higher grades.

*SEV* can be used by tutors during assessment: they may refer to its automatic counts indicator, citation highlighting or simply use it to quickly gauge the amount and distribution of argumentation cues across an essay. *SEV* can also provide formative feedback to students. Thus, if students running *SEV* on their own essay see that little argumentation is found, they are well advised to “revise” their essay before

submission. An improvement in the essay (more background and reasoned argumentation) should result in more highlighting, which may increase motivation in some students.



**Fig. 1.** Argumentation-based Question-Answering Model. The yellow boxes represent tutor-provided information and the blue box is the student's input to the system (an essay). Finally, the white boxes represent the components proper of our architecture and of our tool, *SEV*.

### 3 Argument Modelling in Papers

Relevant research background spans from articles on argumentation in research papers to knowledge representation tools supporting the construction of rhetorical arguments. An important strand of research has focused on paper structure, producing metadiscourse taxonomies applicable to research papers. In his CARS model, Swales [6] synthesised his findings that papers present three moves: authors first establish a territory (by claiming centrality, making topic generalisations and reviewing items of previous research), then they establish a niche (by counter-claiming, indicating a gap or question-raising) and finally they occupy this niche (by outlining purpose, announcing present research and principal findings and indicating paper structure). Although his analysis targeted only the introductory part of an academic research paper, his model has nevertheless been influential. For instance, Teufel [7] extended Swales's CARS model by adding new moves to cover the other sections. They classify sentences into background, other, own, aim, textual, contrast and basic

categories. The authors claim that this methodology could be used in automatic text summarisation, since the latter requires finding important sentences in a source text by determining their most likely argument role. Their experiments showed that the annotation schema can be successfully applied by human annotators, with little training.

Hyland [8] distinguishes between textual and interpersonal metadiscourse in academic texts. The former refers to devices allowing the recovery of the writer's intention by explicitly establishing preferred interpretations; they also help form a coherent text by relating propositions to each other and to other texts. Textual metadiscourse includes logical connectives (in addition, but, therefore etc), frame markers (e.g. finally, to repeat, our aim here, endophoric markers (noted above, see Fig 2, table 1, below), evidentials (According to X, Y states) and code glosses (namely, e.g., in other words, such as). Interpersonal metadiscourse, instead, expresses the writer's persona by alerting the reader to the author's perspective to both the information and the readers themselves. Categories of interpersonal metadiscourse are hedges (might, perhaps, it is possible), emphatics (in fact, definitely, it is clear, obvious), attitude markers (Surprisingly, I agree), relational markers (Frankly, note that, you can see) and person markers (I, we, me, mine, our).

Another interesting source is ScholOnto, an Open University project aiming to model arguments in academic papers and devise an ontology for scholarly discourse[9]. As part of their project, they developed *ClaiMaker*, a tool for browsing and editing claims. These are classified as general (e.g. is about, uses, applies, improves on), problem-related (e.g. addresses, solves), evidence (supports or challenges), taxonomic, similarity (or difference) and causal. *ClaiMaker* is meant for academic research papers, whereas we want an argumentation categorisation for student essays.

### 3.1 Our approach to Argumentation on Student Essays

As a first step in our research, we identified candidate categories of argumentation in student essays through a preliminary manual analysis of essay texts. Some categories were influenced by *ClaiMaker* and the other categorisations seen above.

Our bottom-up approach initially yielded the following argumentation categories: definition, comparison, general, critical thinking, reporting, viewpoint, problem, evidence, causal, taxonomic, content/expected and connectors. Some categories have sub-categories (e.g. connectors comprises topic introduction, inference, contrast, additive, support, reformulation and summative subcategories of connectors).

A review of this schema prompted us to reduce the number of categories (cognitive overload, clearer visualisation). We thus grouped related categories and turned them into subcategories of a new category (e.g. evidence, causal and taxonomic became subcategories of the new "link" category) or modified categories ("viewpoint" merged into "positioning", the new name for "critical thinking"). Our revised categorisation also sees comparison as part of definition, because we often define a concept by comparing it with others. The outcome of the rationalisation process is the following student essay categorisation: definition, reporting, positioning, strategy, problem, link, content/expected, connectors and general (Table 1).

Compared to Teufel’s schema, ours lacks an AIM category: this is because all student essays have the implicit aim of answering the essay question. Similarly, we do not distinguish between OTHER and OWN (knowledge shared by author in other papers and this paper respectively), as it is not relevant to student essays. On the other hand, our content/expected category has no counterpart in the other categorisations, since it is a student essay-specific category comprising cue phrases identifying content that the tutor expects to find in the essay. Overall, however, there are remarkable similarities across these categorisations (for a comparison, see [3]).

**Table 1.** Our Taxonomy for Argumentation in Student Essays

Category	Description	Cue phrases (examples)
DEFINITION	Items relating to the definition of a term. Often towards the beginning. IS_ABOUT, COMPARISONS	is about, concerns, refers to, definition; is the same; is similar /analogous to;
REPORTING	Sentences describing other research in neutral way	“X discusses”, “Y suggests”, “Z warns”
POSITIONING	Sentences critiquing other research; VIEWPOINTS	“I accept”, “I am unhappy with”, “personally”;
STRATEGY	Explicit statements about the method or the textual section structure of the essay	“I will attempt to”, “in section 2”
PROBLEM	Sentences indicating a gap or inconsistency, question-raising, counter-claiming	“There are difficulties”, “is problematic”, “limitations”
LINK	Statements indicating how categories of concepts relate to others: TAXONOMIC, EVIDENCE, CAUSAL	“subclass of”, “example of”, “would seem to confirm”, “has caused”
CONTENT/ EXPECTED	Any concept that the tutor expects students to mention in their essay. Tutor-editable	Essay-dependent
CONNECTORS	Links between propositions may serve different purposes (topic introduction, support, inference, additive, parallel, summative, contrast, reformulation)	“With regard to”, “As to”, “Therefore”, “In fact”, “In addition”, “Overall”, “However”, “In short”
GENERAL	Generic association links	“is related to”

## 4 Annotation Categories and Essay Questions

Query classification gives information about the kind of answer our system should expect. The classification phase involves processing the query to identify the category of answer that the user is seeking. This step is performed using the information obtained during the segmentation of the sentence. During segmentation of the sentence the system finds nouns, verbs, prepositions and adjectives. The categories of possible answers, which are listed below, extend the universal categorisation used in traditional question answering systems (by adding to the six categories: what, who, when, which, why and where). Our analysis of the essay questions in our testbed (see Table 2 for questions and Section 5 for testbed) showed that they were answered by essays with different “link profiles” (see Table 3).

**Table 2.** Examples of Essay Questions

	<i>Assignment</i>	<i>Example</i>
1. Summary + How and Why Opinion about X	Ass 1, part 2	“In the light of Otto Peter’s ideas... say how each type can or cannot serve these ideas and why”
	Ass 2, part 1	- “Who do you think should define the learners’ needs in distance education?”
	Ass 4, part 3	- “State and define your views on the questions of whether the research is adequately addressing what you regard to be the important questions or debates”
3. Describe + Discuss	Ass 2, part 2	“Imagine you are student and your teacher has a strong leaning towards the technical-vocational orientation. Describe and discuss your experiences, using concepts and examples from text book 1.”
	Ass 4, part 2	“Define and discuss any cultural factors you observe in relation to each of these questions”
4. Give example of X and Critique X	Ass 4, part 1	“Provide examples of web links covering a wide range of choose aspects of open and distance education and write a short critique of each.”

The basic idea is that, depending on the essay question, we expect to find a different “distribution” of links in the essay themselves. For instance, a question asking for a “summary” is usually answered by an essay containing many “reporting” links. Table 3 matches essay questions with our essay metadiscourse categories (Table 2). We ran a statistical analysis of links and question types and our findings are presented in section 5.

**Table 3.** Examples of Essay Questions and Expected Links

<i>Example of Question</i>	<i>Links Expected to be Important in Essay</i>
1. Summary of X + How and Why	Essays answering such questions have a high number of reporting, positioning, expected and contrast links.
2. Opinion about X	Essay has a high number of background, expected names, positioning links.
3. Describe and Discuss	These essays feature a high number of support and positioning links. In assignment 2, part 2, there was a low number of reporting links, as students were asked to describe a hypothetical situation; however, this may not always be the case.
4. Give an example of X and Critique X	Here, analysis and summative connector links are higher than “is about” and “contrast” links.

## 5 Test Bed

Our testbed consisted of 193 anonymised essays (belonging to 4 different assignments), with corresponding essay questions. The essays were anonymised versions of actual essays submitted by students as part of a Masters Course at the

Open University. The essays were marked by three experienced tutors, whose comments we also consulted.

We chose student essays as our domain, as essays are typically less structured than other types of documents, including academic research papers, and therefore more difficult to work with. If our approach works with student essays, it will be almost guaranteed to work also with more structured types of texts. Also, while good results have been achieved in the area of student essays classification and assessment with statistical methods, these methods have no semantics and do not provide useful feedback, which is of course particularly important in the area of student essays.

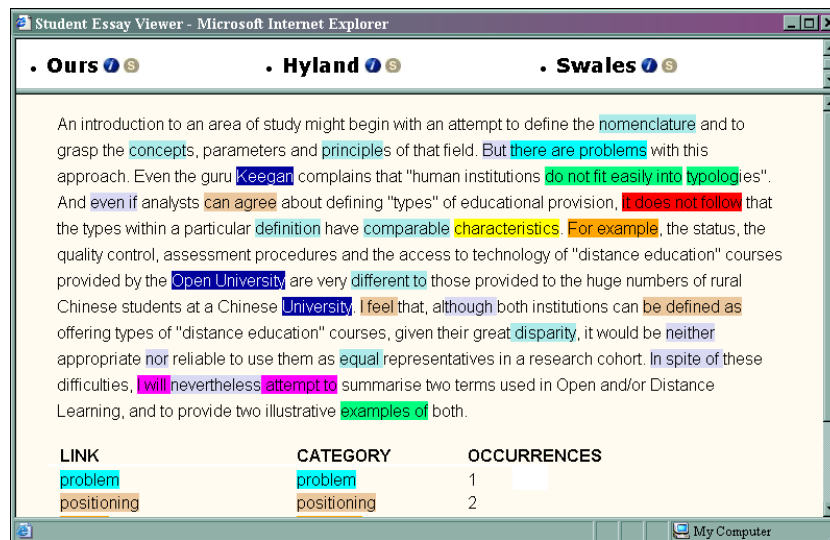


Fig. 2. Student Essay Viewer (SEV) showing all annotations in an essay

Having devised an argumentation schema for essays, we decided to implement a system to visualise argumentation in student essays. The resulting *SEV* is a tool that – while being as easy to use as a webpage – can be a time-saving tool for both tutors and students to use, thanks to its quick visualisation of argumentation in an essay. However, students may particularly benefit from a question analysis tool: this could analyse and classify essay questions with respect to the type of argumentation required in the essay, thus allowing alerting students to missing (or lacking) categories of argumentation. This tool would be very useful in a formative context and would certainly get the students to stop and think about whether what they are writing is answering the question, rather than simply waffling on, as they are sometimes known to do. *SEV* is a question-answering tool, in that it tries to help create a satisfactory answer to a question and can alert the user if such satisfaction is not achieved. At this moment in time, *SEV* works at the phrase level, but we are hoping to move on to longer linguistic units (e.g. sentences and/or paragraphs) soon.

We therefore determined what “link profiles” (Tables 2, 3) could reasonably be expected in a satisfactory essay written for assignments 1 and 2 and then performed a

statistical analysis on the data in our possession to verify these hypotheses and find out the specific kind of argumentation that *SVE* should be looking for each type of question. The results of this analysis are summarised in Table 4.

**Table 4.** Expected and Actual Argumentation links in Assignments 1 and 2

ID	Expected	Results	Analysis
Ass 1 Part 1	many reporting links	<ul style="list-style-type: none"> <li>- reporting links count significant (r=0.730; N=12; p&lt;0.01)</li> <li>- positioning links count is not               <ul style="list-style-type: none"> <li>- total link count significant: r=0.624; N=12; p&lt;0.05</li> <li>- F(1,10)= 6.385; p&lt;0.05</li> </ul> </li> </ul>	Both Spearman correlation and ANOVA F-statistic seems to support our expectations: reporting links are more important than positioning links in this type of essay.
Ass 1 Part 2a	high number of reporting, positioning and expected links.	<ul style="list-style-type: none"> <li>- reporting more important than positioning</li> <li>- statistical significance for "specific reporting links":               <ul style="list-style-type: none"> <li>- a) "Peters" r=0.744;n=12;p&lt;0.01</li> <li>- b) "Peters+industrial+ODE" r=0.717;n=12;p&lt;0.01</li> <li>- F(1,14)=6.524; p&lt;0.05</li> </ul> </li> </ul>	Some students, while including sufficient reporting /expected links, managed to wander off topic (and hence their grade was not high). Better grades achieved by essays that stayed "on topic" ("specific reporting" links)
Ass 1 Part 2b		<ul style="list-style-type: none"> <li>- significant correlation between score and specific reporting links:               <ul style="list-style-type: none"> <li>- r=0.526;n=15;p&lt;0.05</li> <li>- (r=0.586 if we ignore references to "Holmberg")</li> </ul> </li> <li>- no statistical significance for generic reporting or positioning links</li> <li>- expected not significant</li> </ul>	Many students wandered off topic (discussed around Holmberg / expected stuff but not enough on guided didactic conversation or GDC). Hence, only reliable indicator is specific reporting links.
Ass 2 Part 1	positioning links important	<ul style="list-style-type: none"> <li>- positioning links show a significant correlation with score: r=0.538;n=20;P,0.05</li> </ul>	When background is not "at the forefront" in an essay question, positioning tends to be the determinant link type.
Ass 2 Part 2	-reporting (especially reporting on Schön)	<ul style="list-style-type: none"> <li>- reporting links (generic): Spearman's Rho: 0.467; n=20; p&lt;0.05;</li> <li>-specific reporting links r=0.541; n=20; p&lt;0.05;</li> <li>- word count: r=0.639;n=20;p&lt;0.01</li> </ul>	Reporting links are important in this kind of essay, particularly links directly connected to the question (students sometimes tended to wander off topic). Word count is important, again, as this is the last part in Ass2 and some students overran their target in part 1.

While positioning links are determinant in Assignment 2 part 1, overall, the importance of reporting links is apparent: after all, essays at graduate and post-graduate level nearly always – to some extent – require showing that one has "done



the reading”. Where reporting links were not significantly correlated with grade, it seems to be because students wandered off topic (e.g. talked about Holmberg and his ideas at length, but neglected to spend most of their time and words on guided didactic conversation, which is what the question specifically asked about). This suggests that – in order to detect if an essay is answering the question (as opposed to going off topic) – our tool should make use of both a “generic” reporting link category and a more specific one (“specific reporting links” in Table 4), with instances derived from query classification techniques (such as sentence segmentation) applied to the essay query. Examples of cues used for “specific reporting” in Ass 1 part 2a were: Peters, industrial and ODE (Open & Distance Learning).

## 6 Results and Future Work

Our main contribution is the application of argumentation techniques to question answering. A second contribution of this paper is our student essay metadiscourse schema, which we have compared and contrasted with categorisations in the research paper domain (Section 1). We have analysed links between argumentation in essays and score to determine if an essay is answering the question (Section 4) and gauge its overall quality. We found that the total number of links seems correlated with score, that positioning and background (expected + reporting) are the variables that generally contribute the most to score prediction and that the essay question is associated with the relative importance of different link types in an essay. We also found that “specific reporting links” are often needed to detect off-topic wanderings in student essays.

We have implemented an easy-to-use tool to visualise the highlighted categories used in an essay and shown how the essay viewer can be helpful to both tutors and students. The main benefit for tutors is quick visualisation of the type of argumentation and concentration of links, while students might use the essay viewer to get feedback about their essay, particularly about lacking categories.

In our investigation, we have used real data, actual essays written by postgraduate students as part of their course. We believe that the results reported here are encouraging in terms of the quality and robustness of our current implementation. However, there is clearly a lot more work needed to make this technology easy enough to use for tutors and students (who are neither experts in language technologies nor ‘power knowledge engineers’) to use. Future implementations of the student essay viewer could categorise longer linguistic units (e.g. sentences or paragraphs) and explain the reasons why a specific categorisation is assigned to them. These explanations might be displayed in pseudo-natural language.

Future work includes implementation of an “essay question analysis tool”. As this paper has shown, depending on the type of essay question asked, different types of argumentation are required to answer it and this is exactly where students tend to need the most help. The question analysis tool will help analyse the question, establish what type of argumentation is missing/lacking and will determine a set of “specific reporting links” for use to detect off-topic wanderings in essays. A reasoning system

could then explain why the student is not answering the question and a visualisation component like that of SEV would be provided to display argumentation in student essays.

## 7 Conclusion

This paper has shown how argumentation techniques could be used successfully for finding answers to specific categories of questions. Furthermore, it has also briefly described our generic metadiscourse annotation schema for student essays and its links to other schemas relating to argumentation in academic papers. An argumentation visualisation tool for student essays has been introduced that uses our essay annotation schema and a cue-based approach to detect argumentation.

Finally, the paper has explored some hypotheses as to how essay assessment and creation may be aided by the student essay viewer. In particular, thanks to its argumentation and question-answering approach, this tool may help students write essays that answer the essay question and give them formative feedback during their essay writing efforts.

## References

- [1] Vargas-Vera, M., Motta, E., Domingue, J.: AQUA: An Ontology-Driven Question Answering System. AAAI Spring Symposium, New Directions in Question Answering, Stanford University (2003a)
- [2] Vargas-Vera, M., Motta, E., Domingue, J.: An Ontology-Driven Question Answering System (AQUA), KMI-TR-129 (2003b)
- [3] Moreale, E. and Vargas-Vera, M.: Genre Analysis and the Automated Extraction of Arguments from Student Essays, *7<sup>th</sup> Int. Computer-Assisted Assessment Conference CAA-2003*, Loughborough (2003)
- [4] Bell, P.: Using Argument Representations to Make Thinking Visible for Individuals and Groups. In R Hall, N Miyake, & N Enyedy (Eds), *Proceedings of CSCL '97: The Second International Conference on Computer Support for Collaborative Learning*, Toronto: University of Toronto Press . (1997) 10–19
- [5] Sharples, M. and O'Malley, C.: *A Framework for the Design of a Writer's Assistant. Artificial Intelligence and Human Learning: Intelligent Computer-Aided Instruction*. J Self, Chapman and Hall Ltd (1988)
- [6] Swales, J.M.: *Genre Analysis*. Cambridge University Press (1990)
- [7] Teufel, S., Carletta, J. and Moens, M.: An Annotation Schema for Discourse-Level Argumentation in Research Articles, *Proceedings of EACL'99* (1999) 110–117
- [8] Hyland, K.: Persuasion and Context: The Pragmatics of Academic Metadiscourse, *Journal of Pragmatics* 30 (1998) 437–455
- [9] Buckingham Shum, S., Uren, V., Li, G., Domingue, J., Motta, E.: Visualizing Internetworked Argumentation, in *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making*. Paul A Kirschner, Simon J. Buckingham Shum and Chad S. Carr (Eds), Springer-Verlag: London (2002)  
<http://www.VisualizingArgumentation.info>