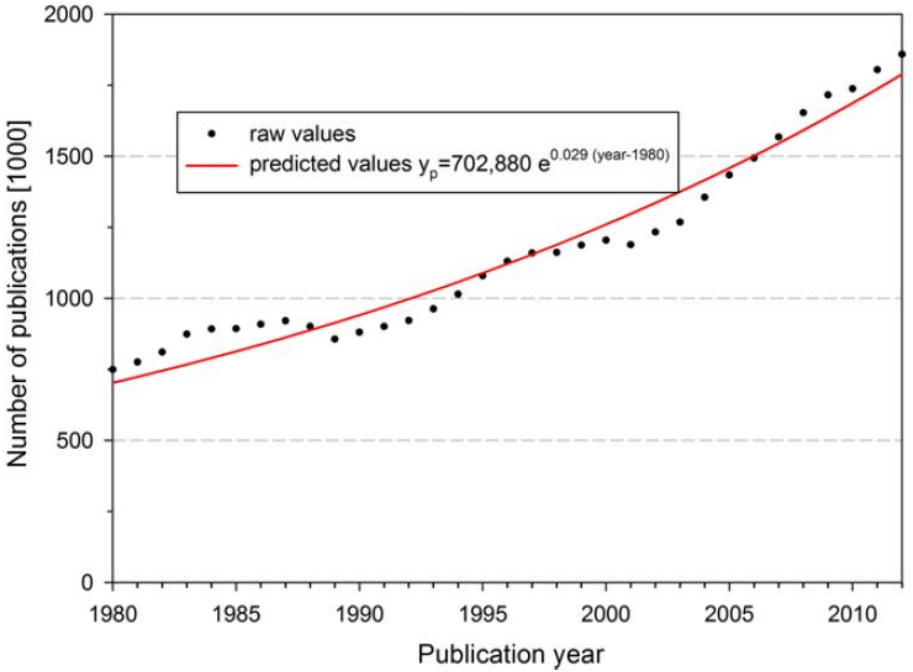
UNIVERSITY OF MANNHEIM

HOCHSCHULE DER MEDIEN

ArguminSci
A Tool for Analyzing Argumentation and Rhetorical Aspects in Scientific Writing

Anne Lauscher, Goran Glavaš and Kai Eckert@ArgMining 2018



The exponential growth of scientific output from 1980 to 2012

(Bornmann and Lutz, 2015)

Computational models are already in place for many rhetorical analysis tasks ...

- citation context analysis (e.g., Jha et al., 2017)
- discourse analysis (e.g., Teufel et al., 1999; Liakata et al., 2010)
- ...

Computational models are already in place for many rhetorical analysis tasks ...

- citation context analysis (e.g., Jha et al., 2017)
- discourse analysis (e.g., Teufel et al., 1999; Liakata et al., 2010)
- ...

... and downstream applications.

- Summarization (e.g., Cohan and Goharian, 2015)
- Research trend prediction (e.g., McKeown et al., 2016)
- Semantometrics (Herrmannova and Knoth, 2016)
- ...

Scientific publications are inherently argumentative
(Gilbert, 1976)

„tools of persuasion“
(Gilbert, 1977)

Carefully composed of different rhetorical layers (*„Scitorics“*)

”In general, our OMR preserves the high frequency content of the motion quite well, since inverse rate control is directed by Jacobian values.”

"In general, our OMR preserves the high frequency content of the motion quite well [claim], since inverse rate control is directed by Jacobian values [data]."

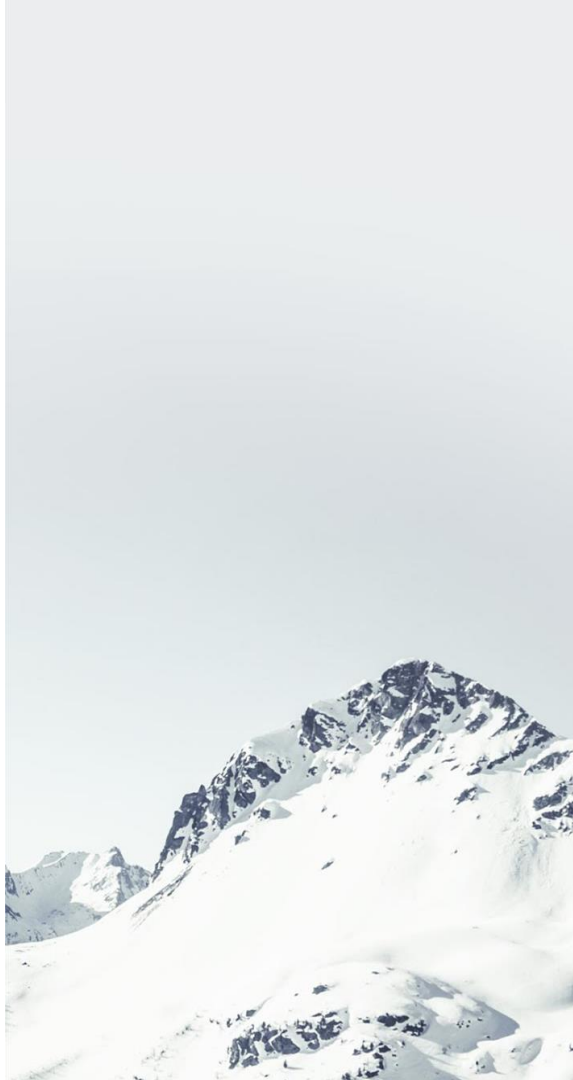
"In general, our OMR preserves the high frequency content of the motion quite well [claim], since inverse rate control is directed by Jacobian values [data]."

- Subjective Aspect: advantage
- Discourse Role: outcome
- Summary Relevance: relevant (Fisas et al., 2016)

ArguminSci aims to support a holistic analysis of scientific publications in terms of scitorics

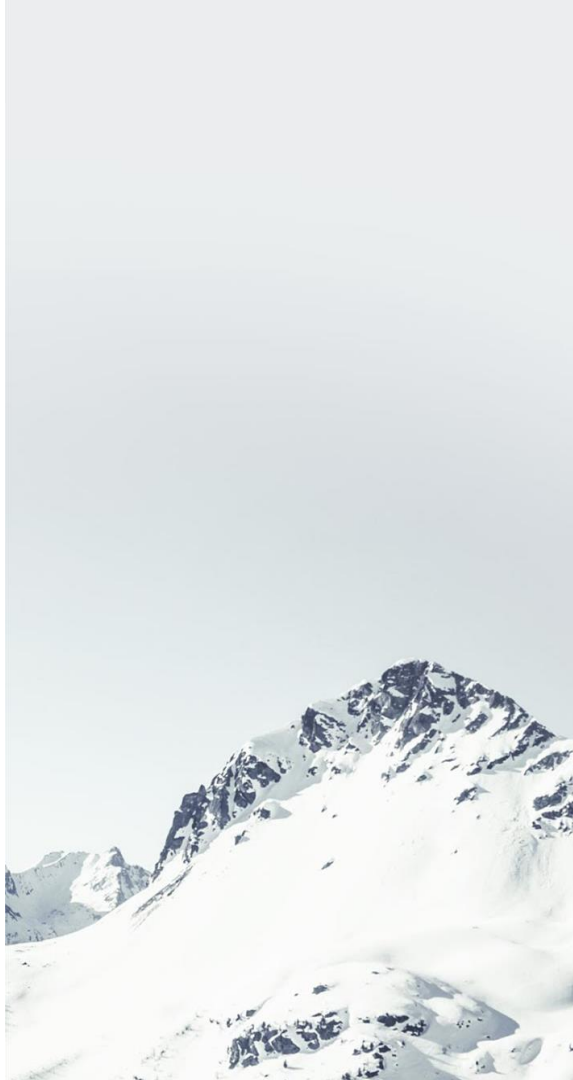
ArguminSci

1. Motivation
2. System Overview
3. Conclusion



ArguminSci

1. Motivation
2. **System Overview**
    - Annotation Tasks and Data Set
    - Annotation Models
    - Interfaces
3. Conclusion



System Overview:
Annotation Tasks and Data Set

Annotation Tasks

Discourse Role Classification
Background, Challenge, Approach, Future Work, Outcome, Unspecified

Subjective Aspect Classification
Advantage, Disadvantage, Novelty, Common Practice, Limitations, None

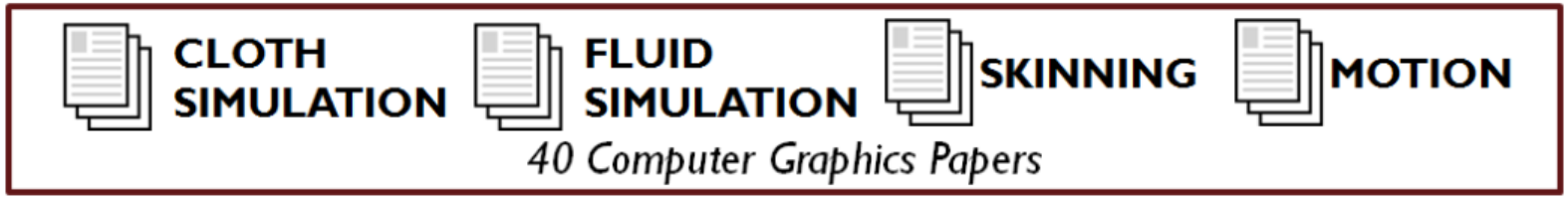
Summary Relevance Classification
Totally irrelevant, Should not appear, May appear, Relevant, Very relevant, None

Citation Context Identification
B-Citation Context, I-Citation Context, Outside

Argument Component Identification
B-I-O annotation scheme with three types of argumentative components:
Own claim, Background claim, and Data

Sentence-level Classification

Token-level Sequence-tagging



Dr. Inventor Corpus (Fisas et al., 2016)

Scientific discourse roles
Background, Challenge, Approach, Future Work, Outcome

Subjective aspects and novelty classes
Advantage, Disadvantage, Novelty, Common Practice, Limitations

Summary relevance grading + Summaries
Totally irrelevant, should not appear, may appear, relevant, very relevant

Citation purpose
Criticism, Comparison, Basis, Use, Substantiation, Neutral

Sentence-level annotations

Token-level annotations

Extension of the corpus with fine-grained argumentative structures

(Lauscher et al. 2018, derived from Toulmin, 2003; Dung 1995; Bench-Capon, 1998)

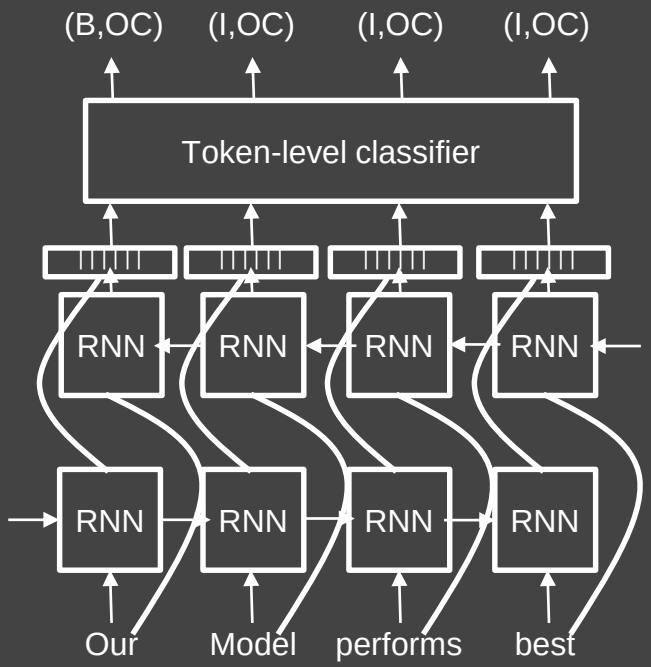
Background Claim	An argumentative statement in question related to the background of the presented work, such as common practices in the field or related studies.
Own Claim	An argumentative statement in question directly related to the author's own work.
Data	A fact that serves as evidence in favor or against a claim.

"SSD is widely adopted in games, virtual reality, and other realtime applications due to its ease of implementation and low cost of computing."

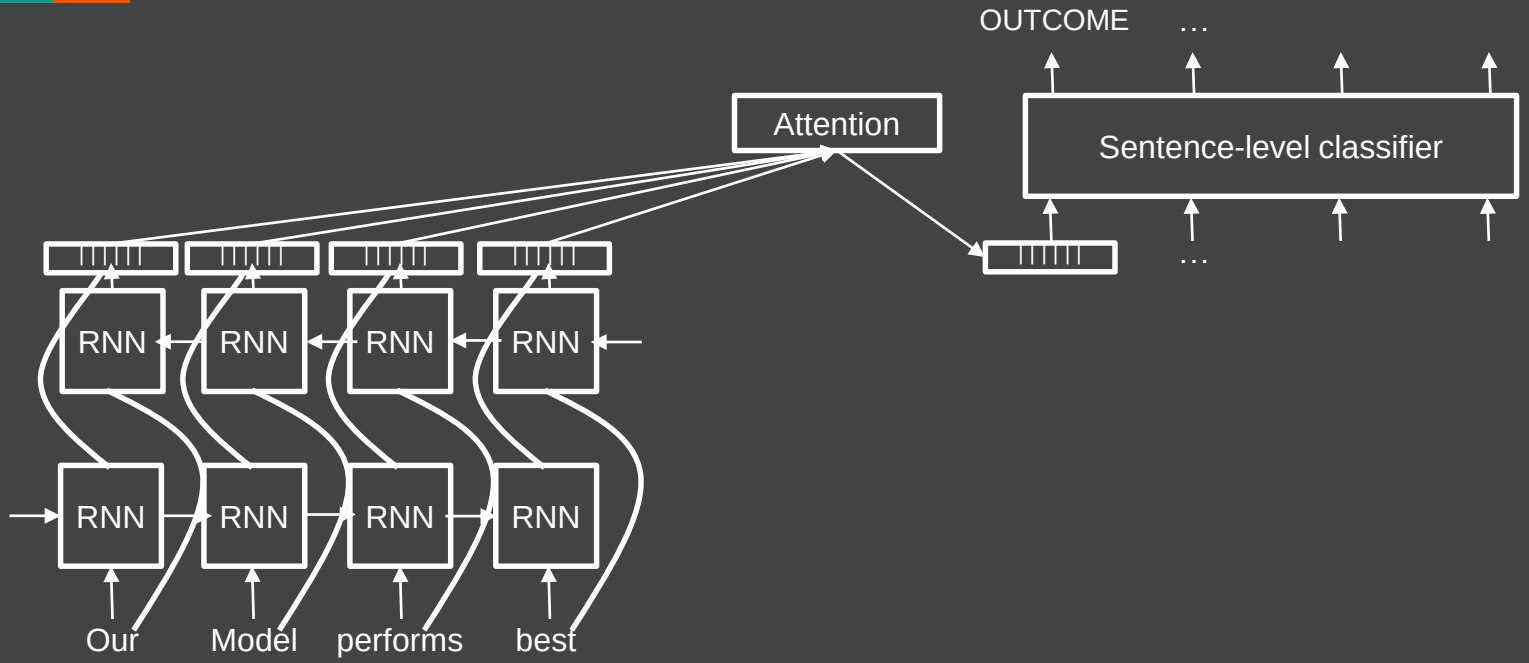
System Overview: Annotation Models

Model Architecture
Token-level tasks

Given a sequence of inputs **x**,
assign a sequence of tags **y**.



Model Architecture
Sentence-level tasks



Model Performances

Granularity	Task	F1 (%)
Token-level	Argument Component Identification	43.8
	Citation Context Identification	47.0
Sentence-level	Discourse Role Classification	42.7
	Subjective Aspect Classification	18.8
	Summary Relevance Classification	33.5

Evaluated on a held-out test set (2874 sentences)

Model Performances

Granularity	Task	F1 (%)
Token-level	Argument Component Identification	43.8
	Citation Context Identification	47.0
Sentence-level	Discourse Role Classification	42.7
	Subjective Aspect Classification	18.8
	Summary Relevance Classification	33.5

Evaluated on a held-out test set (2874 sentences)

Models can be exchanged

System Overview: ArguminSci's Interfaces

System Overview: ArguminSci's Interfaces

- Command Line Interface
- RESTful Application Programming Interface
- Web Application

```
C:\Users\anlausch\PycharmProjects\ArgDemo>python arguminsci.py -h
usage: arguminsci.py [-h] [--argumentation] [--discourse] [--aspect]
                     [--citation] [--summary]
                     inputfile outputfolder

Analyze Argumentation and Rhetorical Aspects in Scientific Writing.

positional arguments:
  inputfile         The name of the textual file containing the input text.
  outputfolder      The name of the output folder where the output should be
                    stored.

optional arguments:
  -h, --help        show this help message and exit
  --argumentation   Extract argument components.
  --discourse       Analyze discourse roles.
  --aspect          Analyze subjective aspects.
  --citation        Extract citation contexts.
  --summary         Assign summary relevance.
```

ArguminSci

Analyze Argumentation and Rhetorical Aspects in Scientific Writing.

Just copy your text in the input field below and submit.

Argument Components | Discourse Categories | Subjective Aspects | Citation Contexts | Summary Relevance

this paper aims to produce fluid simulations with a high degree of spatial adaptivity . we desire to enable a simulator to focus its computational resources on the visually interesting regions of a fluid flow , while remaining computationally efficient and avoiding common artifacts due to a spatially adaptive pressure solve . previous approaches have made great strides towards this goal , but they often exhibit visual artifacts , a lack of computational robustness , or an unacceptably hefty computational expense . the groundbreaking work of Losasso et al . [2004] introduced an octree for spatial adaptivity , but it suffers from spurious flows at T-junctions . finite volume methods [batty et al . 2010] repair these spatial artifacts at the expense of solving a significantly larger system of equations and sacrificing computational stability near poorly-shaped elements . furthermore , many existing methods still are not truly spatially adaptive in the sense that their computational complexity is still tied to a uniform grid or spatial parameter

Own Claim

Background Claim

Data

Insert scientific text here..

ArguminSci

Analyze Argumentation and Rhetorical Aspects in Scientific Writing.

Just copy your text in the input field below and submit.

Argument Components | Discourse Categories | Subjective Aspects | Citation Contexts | Summary Relevance

this paper aims to produce fluid simulations with a high degree of spatial adaptivity . we desire to enable a simulator to focus its computational resources on the visually interesting regions of a fluid flow , while remaining computationally efficient and avoiding common artifacts due to a spatially adaptive pressure solve . previous approaches have made great strides towards this goal , but they often exhibit visual artifacts , a lack of computational robustness , or an unacceptably hefty computational expense . the groundbreaking work of Losasso et al . [2004] introduced an octree for spatial adaptivity , but it suffers from spurious flows at T-junctions . finite volume methods [batty et al . 2010] repair these spatial artifacts at the expense of solving a significantly larger system of equations and sacrificing computational stability near poorly-shaped elements . furthermore , many existing methods still are not truly spatially adaptive in the sense that their computational complexity is still tied to a uniform grid or spatial parameter

Background
Challenge
Approach
Outcome
Future Work

Insert scientific text here..

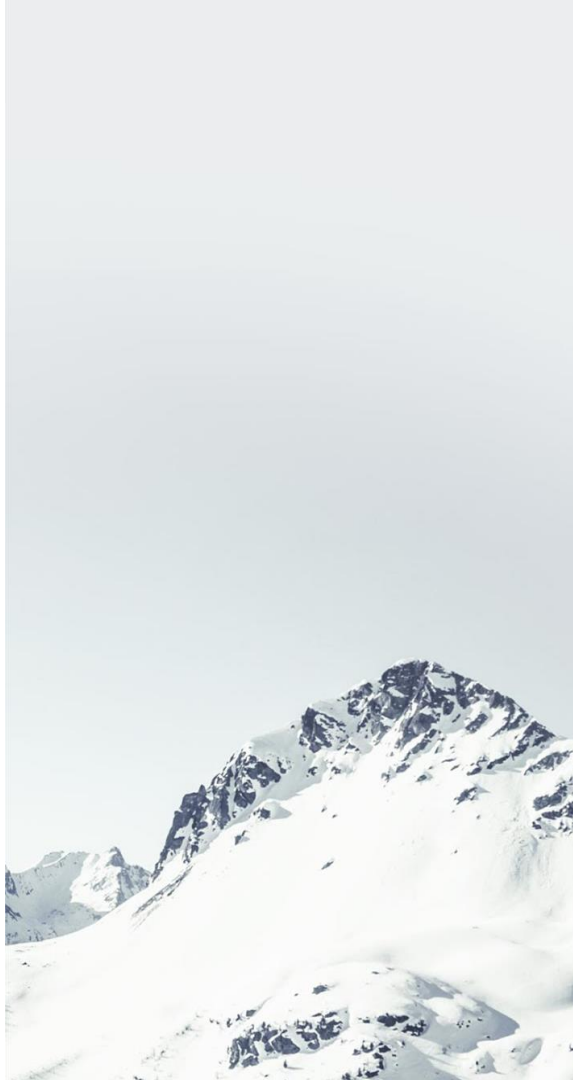
POST localhost:8000/predict?api_mode=True&text=Our model performs best

Pretty Raw Preview JSON


```
{
    "aspect": [],
    "summary": [],
    "discourse": [],
    "argumentation": [
        [
            [
                "our",
                "Token_Label.BEGIN_OWN_CLAIM"
            ],
            [
                "model",
                "Token_Label.INSIDE_OWN_CLAIM"
            ],
            [
                "performs",
                "Token_Label.INSIDE_OWN_CLAIM"
            ],
            [
                "best",
                "Token_Label.INSIDE_OWN_CLAIM"
            ],
            [
                ".",
                "Token_Label.OUTSIDE"
```

ArguminSci

1. Motivation
2. System Overview
3. **Conclusion**



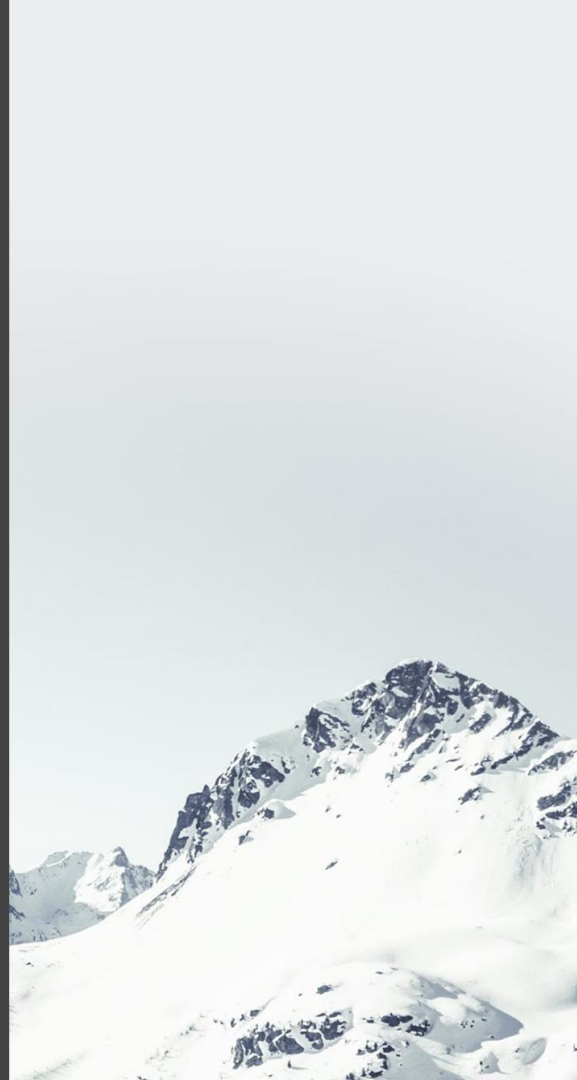
The rhetorical aspects of scientific writing should be studied holistically in order to understand a publication, i.e. a scientific argument, as a whole

ArguminSci illustrates this idea by providing multiple rhetorical analysis perspectives

The rhetorical aspects of scientific writing should be studied holistically in order to understand a publication, i.e. a scientific argument, as a whole

ArguminSci illustrates this idea by providing multiple rhetorical analysis perspectives

FW: Expose training phase, extend with other annotation layers and schemes



The rhetorical aspects of scientific writing should be studied holistically in order to understand a publication, i.e. a scientific argument, as a whole

ArguminSci illustrates this idea by providing multiple rhetorical analysis perspectives

FW: Expose training phase, extend with other annotation layers and schemes

https://github.com/anlausch/ArguminSci

http://data.dws.informatik.uni-mannheim.de/arguminsci/

Thank you

References

T. J. Bench-Capon, “Specification and implementation of Toulmin dialogue game,” in *Proceedings of JURIX*, 1998, vol. 98, pp. 5–20.

A. Cohan and N. Goharian, „Scientific article summarization using citation-context and article's discourse structure“. *arXiv preprint arXiv:1704.06619*, 2017.

P.H. Dung, "On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games." *Artificial intelligence* vol. 77, no. 2, pp. 321-357, 1995.

S. Eger, J. Daxenberger, and I. Gurevych, “Neural End-to-End Learning for Computational Argumentation Mining,” in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Vancouver, Canada, 2017, pp. 11–22.

B. Fisas, H. Saggion, and F. Ronzano, “On the Discoursive Structure of Computer Graphics Research Papers.,” in *LAW@ NAACL-HLT*, 2015, pp. 42–51.

References

B. Fisas, F. Ronzano, and H. Saggion, “A Multi-Layered Annotated Corpus of Scientific Papers.,” in *LREC*, 2016.

G. Nigel Gilbert, “The transformation of research findings into scientific knowledge”, *Social Studies of Science*, vol. 6, no. 3-4, pp. 281–306, 1976..

G. Nigel Gilbert, “Referencing as persuasion,” *Social Studies of Science*, vol. 7, no. 1, pp. 113–122, 1977.

D Herrmannova and P Knoth, “Semantometrics: Towards fulltext-based research evaluation“, in *Proceedings of the Joint Conference on Digital Libraries (JCDL)*, IEEE/ACM, 2016, pp. 235-236.

R. Jha, A. A. Jbara, V. Qazvinian, and D.R. Radev, “NLP-driven citation analysis for scientometrics.“, in *Natural Language Engineering*, 2017, vol. 23., no. 1, pp. 93-130.

References

C. Kirschner, J. Eckle-Kohler, and I. Gurevych, “Linking the Thoughts: Analysis of Argumentation Structures in Scientific Publications,” in *Proceedings of the 2nd Workshop on Argumentation Mining held in conjunction with the 2015 Conference of the North American Chapter of the Association for Computational Linguistics – Human Language Technologies (NAACL HLT 2015)*, 2015, pp. 1–11.

M. Liakata, S. Saha, S. Dobnik, C. Batchelor, and D. Rebholz-Schuhmann, “Automatic recognition of conceptualization zones in scientific articles and two life science applications,” *Bioinformatics*, vol. 28, no. 7, pp. 991–1000, Apr. 2012.

A. Lauscher, G. Glavaš, S. P. Ponzetto, and K. Eckert, “Investigating convolutional networks and domain-specific embeddings for semantic classification of citations,” in *Proceedings of WOSP 2017*, Toronto, 2017a, vol. tba, p. tba.

A. Lauscher, G. Glavaš, and K. Eckert, “University of Mannheim @ CLSciSumm-17: Citation-Based Summarization of Scientific Articles Using Semantic Textual Similarity,” in *2nd Joint Workshop on Bibliometric-enhanced Information Retrieval and Natural Language Processing for Digital Libraries 2017 [?]*, tba, 2017b, p. tba.

References

S. Teufel and M. Moens, “Summarizing Scientific Articles: Experiments with Relevance and Rhetorical Status,” *Comput. Linguist.*, vol. 28, no. 4, pp. 409–445, Dec. 2002.

S. Teufel, A. Siddharthan, and C. Batchelor, “Towards discipline-independent argumentative zoning: evidence from chemistry and computational linguistics,” in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3-Volume 3*, 2009, pp. 1493–1502.

S. Teufel, J. Carletta, and M. Moens, “An annotation scheme for discourse-level argumentation in research articles,” in *Proceedings of the ninth conference on European chapter of the Association for Computational Linguistics*, 1999, pp. 110–117.

K. McKeown *et al.*, “Predicting the impact of scientific concepts using full-text features,” *J Assn Inf Sci Tec*, vol. 67, no. 11, pp. 2684–2696, Nov. 2016.

S. E. Toulmin, *The Uses of Argument*. Cambridge University Press, 2003.