Journal of Scientometric Research, 2026; 15(2):356-364.
https://www.jscires.org

Research Article

ArguKPE: Argument-Driven *Keyphrase* Extraction with Transformer-Based Semantic Alignment for Reviewer-Manuscript Matching

Nishith Kotak[1,*], Bakul Gohel[2], Arjav Bavarva[1]

[1]Department of Information and Communication Technology, Marwadi University, Rajkot, Gujarat, INDIA.
[2]Department of CE-AI and BIG DATA, Marwadi University, Rajkot, Gujarat, INDIA.

ABSTRACT

Keyphrase Extraction (KPE) is an important component in the comprehension of scholarly documents, as well as downstream tasks including information retrieval and reviewer-manuscript matching. The current KPE methods, such as statistical, graph-based, and new transformer-based ones, are mainly based on surface features or contextual embeddings, which fail to account for the underlying discourse structure of scientific texts. Consequently, the approaches often harvest statistically salient yet semantically inappropriate words into the main contribution of the manuscript, resulting in less-than-optimal document representation and poorer performance on downstream tasks. To overcome this shortcoming, we introduce ArguKPE, an argument-based *Keyphrase* Extraction model that makes an explicit effort to introduce the argumentative base of scientific texts into the extraction process. The suggested method uses argumentation mining in preference to the title and abstract to extract the important elements, i.e., the research goal and research approach, and incorporates them as structured semantic priors into the existing KPE algorithms. This definition allows the improvement of various approaches, such as statistical, graph-based, and embedding-based, by using a single argument-based scoring system. Moreover, we present ArguKPE-BERT, a transformer-based enhanced representation that integrates semantic alignment of arguments and contextual document representations. The proposed model combines semantic relevance and argumentative significance into one, achieving a more accurate and understandable *Keyphrase* extraction by jointly modeling both global context and discourse-level importance. Extensive testing on benchmark datasets, such as SemEval, NUS, Krapivin and Inspect, show that the proposed framework is always better than the baseline methods in all categories. In particular, when using argument-based improvements, an increase of up to 3-6% in F1-score is achieved over classical algorithms, with the proposed ArguKPE-BERT achieving a further improvement of 4-5% over powerful transformer-based baselines. These findings confirm that the use of argument structure will add complementary information on top of the traditional statistical and contextual cues. On the whole, this paper makes argument-aware modeling a powerful way to enhance *Keyphrase* extraction and emphasizes that it can be applied to the downstream systems as a reviewer-manuscript matching system.

Keywords: Argumentation Mining, *Keyphrase* Extraction, Reviewer-Manuscript Matching, Semantic Alignment, Transformer Models.

Correspondence:
Nishith Kotak
Department of Information and Communication Technology, Marwadi University, Rajkot-360003, Gujarat, INDIA.
Email: nishith.kotak@gmail.com

Received: 11-03-2026;
Revised: 27-04-2026;
Accepted: 09-06-2026.

INTRODUCTION

Keyphrase Extraction (KPE) is a core task in natural language processing, which forms the basis of numerous applications, such as document summarization, information retrieval, topic modeling and discovery of scholarly knowledge discovery (Mihalcea and Tarau, 2004). KPE is especially important in academic publishing ecosystems, in which it is crucial to correctly represent the thematic content of a manuscript to allocate domain-relevant reviewers. Although much literature has been done in this field, the current KPE techniques are still limited in the sense that they do not reflect the more profound semantic intent and discourse structure of scientific texts.

Older *Keyphrase* extraction methods, including statistical methods like TF-IDF and KP- Miner (Beltagy and Rafea, 2009) as well as graph-based methods like Text Rank and PositionRank (Florescu and Caragea, 2017b) and newer embedding-based models (Devlin *et al.*, 2019) are mainly based on surface-level cues like term frequency, co-occurrence patterns, or contextual similarity. Although these techniques have shown a reasonable

DOI: 10.5530/jscires.20260146

Copyright Information :
Copyright Author (s) 2026 Distributed under Creative Commons CC-BY 4.0

Publishing Partner : Manuscript Technomedia. [www.mstechnomedia.com]

level of performance, they mostly consider documents as groups of words or tokens, without explicitly modeling the argumentative structure that rules scientific writing. Due to this, such methods tend to pull out phrases that are common or structurally salient but not always key to the central contribution of the work, thus creating ambiguity, redundancy and lower interpretability.

Scientific documents, in their turn, have a clearly defined argumentative structure, usually consisting of the following elements: the research goal; the research approach; and contextual constraints. Such discourse structures are identified in text through argumentation mining, which is an emerging branch of NLP, and aims to find them in a textual form of argumentation (Lippi and Torroni, 2016). Overlooking these structures constrains the capacity of KPE systems to differentiate between *Keyphrase*s that are really informative and peripherally associated words. Such a limitation is especially severe in downstream processes like reviewer assignment, when the misalignment of semantics may result in poor choice of reviewers.

To solve these issues, we present ArguKPE, an argument-sensitive *Keyphrase* extraction system that combines argumentation mining with the existing *Keyphrase* extraction methods. The major hypothesis of this research is that *Keyphrase*s in relation to the argumentative parts of a document have higher chances of being semantically meaningful, discriminative and useful in downstream applications. In order to operationalize this concept, we distill argument elements in the title and abstract based on syntactic and rhetorical cues and introduce them as a structure semantic prior to the *Keyphrase* extraction process.

The matching of reviewers to manuscripts commences with the screening of each manuscript by the editorial board before the peer-reviewing. In order to start the peer review process, the editor determines the keywords in the title and the abstract of the manuscript. The editor then tries to Figure out the keywords among the group of reviewers. COIs is vetted to each possible reviewer and in the event that they are deemed suitable, the reviewers are invited to review the manuscript. The flow of process is as shown in Figure 1 (Crego *et al.*, 2023).

In contrast to heuristically-based enhancements, we make argument-aware biasing a formalization of the candidate ranking function in which the relevance of a *Keyphrase* depends not only on the standard scoring mechanisms but also on its semantic congruence with the components of the argument. The formulation gives a principled basis of incorporating discourse-level information in various KPE algorithms in an integrated and model-agnostic way. Moreover, we include a theoretical rationale that shows that, with realistic assumptions about how the argument structure and the true *Keyphrase*s coincide, the presented approach enhances the anticipated ranking of the relevant phrases.

Based on this, we continue to generalize our framework to a neural environment with ArguKPE-BERT, which uses transformer-based contextual embeddings (Devlin *et al.*, 2019) to learn more about semantic connections. Both candidate key phrases and the argument components are also encoded in the same embedding space in this extension, allowing fine-grained semantic alignment based on similarity measures. We also investigate a contrastive learning formulation that strictly enforces correspondence between argument components and actual *Keyphrase*s, and thus improves the quality and robustness of the representation.

We perform a large-scale experiment on various benchmark datasets, such as SemEval (Kim, *et al.*, 2010), NUS (Nguyen, 2007; Krapivin and Marchese, 2009; Hulth, 2003), and compare traditional and argument-aware versions of a number of KPE algorithms. The findings are consistent, showing a significant improvement when using argument-driven signals in various methods and datasets. In addition to quantitative analysis, we offer more in-depth analysis such as ablation studies and qualitative insights in order to gain deeper insights into the behavior and benefits of the suggested approach.

In addition, we address implications of better *Keyphrase* extraction to reviewer-manuscript matching, and the way argument-conscious representations can improve semantic consistency and assignment reliability.

THE MAIN CONTRIBUTIONS OF THIS PAPER ARE SUMMARIZED AS FOLLOWS:

- We introduce ArguKPE, a novel framework that integrates argumentation mining with *Keyphrase* extraction to capture discourse-level semantics.
- We propose a formal argument-aware scoring model, establishing a principled method for incorporating structured semantic priors into KPE.
- We provide a theoretical justification demonstrating the effectiveness of argument-aware biasing in improving *Keyphrase* ranking.
- We develop ArguKPE-BERT, a transformer-based extension that embeds argument-aware signals into contextual representation learning.
- We perform comprehensive empirical evaluation across multiple benchmark datasets, demonstrating consistent improvements over baseline methods.

By bridging the gap between argumentation mining and *Keyphrase* extraction, this work advances the state of the art in document representation and provides a robust foundation for improving downstream tasks such as reviewer-manuscript matching and scholarly information retrieval systems.

The remainder of the paper is structured as follows. Section 2 reviews existing literature on *Keyphrase* extraction, covering supervised, unsupervised, embedding-based, and transformer-based approaches, and highlights the limitations that motivate our work. Section 3 presents the formal problem formulation and introduces the proposed argument-aware framework, including the argument mining strategy for extracting research goal and research approach components. Section 4 describes the integration of the proposed framework with classical, embedding-based, and transformer-based *Keyphrase* extraction methods, including the ArguKPE-BERT model. Section 5 provides a comprehensive experimental evaluation on benchmark datasets, along with detailed analysis, ablation studies, and comparisons with baseline approaches. Finally, Section 6 concludes the paper and discusses potential directions for future research.

RELATED WORK

Keyphrase extraction techniques are broadly classified into two categories: supervised and unsupervised techniques. The model for the supervised technique gets trained over the labeled training set and then uses the trained model to determine whether the given candidate phrase from the document, be treated as a *Keyphrase* or not. Methods such as (Wang, 2019) views *Keyphrase* extraction as a classification problem where in (Gutierrez *et al.*, 2014), *Keyphrase* extraction is viewed as a regression problem.

One of the first supervised *Keyphrase* Extraction techniques KEA (Witten *et al.*, 2009) uses TF-IDF score and position of the first occurrence of the candidate phrase to determine the *Keyphrase* using the Naive Bayes algorithm. The linguistic features such as Part-of-Speech (POS) Tag, Term Frequency (TF), Inverse Document Frequency (IDF) and position of the first occurrence of the phrase are used for determining the *Keyphrase* by applying rule induction with the bagging approach (Hulth, 2003). The Maui KEA, which was extended by Maui (Medelyan *et al.*, 2005), was based on bagged decision trees for *Keyphrase*s extraction from Wikipedia. They examined the distribution of *Keyphrase*s across the parts of the entire text of the papers.

The major limitation of the supervised technique is the availability of the "labeled" training data. Labeling of the dataset requires huge manual efforts or may require many heuristics. However, there is still a deficiency of the dataset, as mentioned in (Sun *et al.*, 2009). Unsupervised techniques extract the *Keyphrase*s either based on linguistic features or statistical features. Peering into the feature extraction methods applied by the researchers, the unsupervised methods can be classified as Graph-based, Statistics-based, Embeddings-based and Language model-based methods.

Statistic-based methodology contains a frequent base method i.e., TFIDF to extract *Keyphrase*s. It was followed by the introduction of effective variation as TFIDF2 (Florescu and Caragea, 2017) which takes into consideration the frequency of the phrases rather than the actual frequency in order to equal out the high frequency words. KP-Miner (KPM) filters out the candidate *Keyphrase*s based on the minimum phrase frequency i.e., least allowable seen frequency (lasf) and cutoff constant. A scoring system is then used to compute the score of a particular candidate phrase in terms of Tf and Idf scores and the positional factor and boosting factor. KeyCluster (Liu *et al.*, 2009) employs the co-occurence-based measures to determine the significance of the candidate terms in the semantic relevance and clusters them with spectral clustering. The candidate *Keyphrase*s are scored with a set of five features such as Casing, Word Position, Word Relevance (number of different terms surrounding a word), Word Frequency and WordDifSentence (how often a word appears in different sentences) which are used in YAKE (Campos *et al.*, 2020). The algorithm, Rapid Automatic Keyword Extraction (RAKE) (Rose *et al.*, 2010) is based on the frequency of words and the frequency of words co-occurrence with other words to compute the scores of the candidate *Keyphrase*s.

Graph based unsupervised method has the nodes as the candidate tokens with the relationship strength between two tokens being represented by the edges. Text Rank is the initial graph based unsupervised *Keyphrase* extraction method which does the weight sharing between each node based on the PageRank algorithm made available by Google (page rank) till it reaches a convergence point. Single Rank the Text Rank algorithm was extended by Single Rank, using the number of co-occurrence of the nodes as the weight parameter of the edge. Position Rank is based on the idea that the larger weights are placed on those tokens which are also appearing earlier in the document, as well as on word-word co-occurences. Topic Rank (Bougouin *et al.*, 2013) is a topic-based method that extracts *Keyphrase*s that represents the document in terms of the topics that it covers. There is a variant of topic rank known as Multi partite Rank (Boudin, 2018), which takes into account a graph of topics where the weight of the edge is the location of the phrases within the document.

The argumentation mining is devoted to the identification of argumentative elements of text including claims, premises and relationships in a text. It has been used in opinions mining, debate analysis and scientific discourse understanding. Within the framework of academic records, the structures of argumentation tend to manifest the research purpose, research methodology, and research contributions.

Argument-Aware *Keyphrase* Extraction

In this section, we formally define the *Keyphrase* extraction problem and introduce our proposed argument-aware framework, ArguKPE, as shown in Figure 2. We first present the mathematical formulation and then describe the integration of argument-driven signals into the extraction process.

Let a document be represented as D, consisting of a sequence of tokens. From D, we extract a set of candidate *Keyphrase*s:

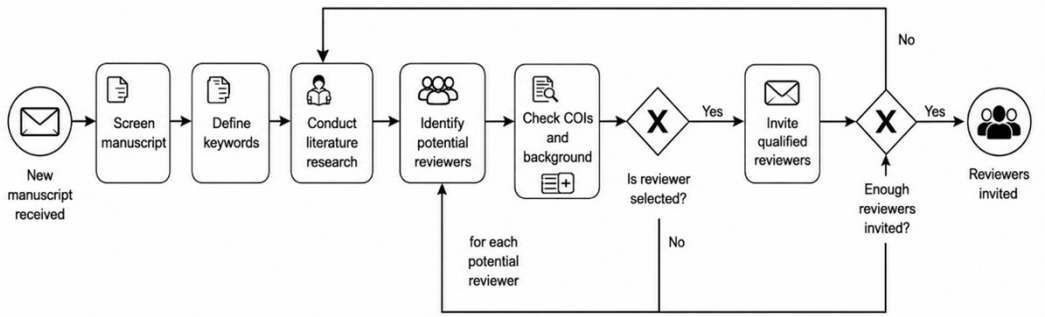

Figure 1: Reviewer Manuscript matching process.

$$C=\{c1, c2,...cn\}$$

The goal of *Keyphrase* extraction is to rank these candidates such that the top-*k* phrases best represent the semantic content of the document. Let $K^* \subseteq C$ denote the set of ground-truth *Keyphrase*s.

Traditional approaches define a scoring function:

$$S_{base}(ci; D)$$

Which assigns a relevance score to each candidate phrase c_i. The extracted *Keyphrase*s correspond to the top-ranked candidates according to this scoring function.

Scientific documents inherently contain an argumentative structure that reflects the core contribution of the work. We define the argument structure of a document as:

$$A=\{RG, RA\}$$

Where *RG* represents the research goal and *RA* represents the research approach.

We hypothesize that candidate phrases that are semantically aligned with the argument components are more likely to be true *Keyphrase*s. To incorporate this intuition, we define an argument-aware similarity function:

$$\text{Sim}(ci, A)=\max_{a \in A} \text{sim}(ci, a)$$

Where sim(·) denotes a similarity function (e.g., lexical overlap or embedding similarity).

The maximum-affinity formulation does not seek to impose a strict prioritization of Research Goal (RG) versus Research Approach (RA). Rather, a candidate phrase is given a semantic boost if it is semantically similar to either of the argumentative elements. This design is of particular significance as it is the case with scientific discourse that often the research purpose and methodological description are packed into semantically dense sentences with high interdependence of argumentative structures. The proposed framework is based on the semantic decomposition of the phrases at the level of Noun Phrases (NP), Verb Phrases (VP), and Prepositional Phrases (PP); thus, the RG and RA components can be represented separately even if they appear together in the same sentence. The framework cannot therefore be based on the clear distinction of argumentative structures at sentence level.

We then define the argument-aware scoring function as:

$$S_{AD}(ci)=S_{base}(ci) + \lambda \cdot \text{Sim}(ci, A)$$

Where $\lambda \in R^+$ is a hyperparameter controlling the influence of the argument-driven bias. The proposed formulation can be interpreted as a regularized optimization problem:

$$S_{AD}(ci)=S_{base}(ci) + \lambda \cdot \Omega(ci)$$

Where $\Omega(c_i)=\text{Sim}(c_i, A)$ acts as a structured semantic prior.

This interpretation highlights that argument-aware biasing does not replace existing scoring mechanisms but augments them with higher-level semantic guidance derived from discourse structure.

Argument Mining Framework

The algorithm presented in the paper, called 1 provides a method of extracting arguments based on the title and abstract of a research paper, and it tries to extract the Research Goal (RG) and Research Approach (RA). This algorithm will go through all the sentences in the abstract, syntactically analyzed to identify linguistic structures such as Noun Phrases (NP), Verb Phrases (VP), and Prepositional Phrases (PP). This preliminary step predetermines the following extraction of RG and RA.

The process of extracting the Research Goal (RG) commences with an analysis of all the sentences of the abstract with the aim of determining the phrases that point to the purpose of the research. Here, the algorithm determines cases where a token tagged with the Part-Of- Speech (POS) is TO is followed by a Verb Phrase (VP), which can generally indicate the target research intent. Consequently, the word after the "TO" token is inserted into the RG list and, therefore, manages to capture the essence of the research purpose in the abstract. In addition, the algorithm also searches phrases in which the word for occurs before a VP, showing the purpose of the study. Such phrases, outlined by the "for" token, are also incorporated into the RG list to further enrich the understanding of the Research Goal.

The algorithm analyzes the abstract to search for the Research Approach (RA), considering the strategies employed to address the research goal.

While the current implementation makes use of cues derived from the syntax and rhetoric of the language (infinitive constructions 'to + VP', purposive patterns 'for + VP'), the proposed framework is not only discourse-level and phrase-structure but does not necessarily depend on isolated instances of lexical patterns. The goal of Algorithm 1 is to detect semantic meaningful argumentative structures, which reflect the research purpose and method of scientific texts. It is designed to be lightweight and easy to understand to facilitate integration with heterogeneous *Keyphrase* extraction methods, and is now implemented under the rule guidance. The framework can, of course, be expanded with dependency parsing, semantic role labeling, rhetorical structure analysis, and contextual transformer-based discourse modeling methods. Such extensions would allow for discovery of the implicit argumentative patterns in stylistically varied and domain-specific scholarly writing.

Integration with *Keyphrase* Extraction

The proposed argument-aware scoring function is model-agnostic and can be integrated into various KPE methods. For each candidate phrase c_i, the base score $S_{base}(c_i)$ is computed using the underlying method, and the argument-aware bias is added to obtain the final score.

This integration can be applied to:

- Statistical methods (e.g., RAKE, KP-Miner)
- Graph-based methods (e.g., TextRank, PositionRank)
- Embedding-based methods

Theorem 1

Assuming that true *Keyphrase*s exhibit higher similarity with argument components than non-*Keyphrase*s, the argument-aware scoring function improves the expected ranking of true *Keyphrase*s.

Let $c^+ \in K^*$ be a true *Keyphrase* and $c^- \in/ K^*$ be a non-*Keyphrase*. By assumption:

$$\text{Sim}(c^+, A) > \text{Sim}(c^-, A)$$

Thus,

$$S_{AD}(c^+) - S_{AD}(c^-) = (S_{base}(c^+) - S_{base}(c^-)) + \lambda(\text{Sim}(c^+, A) - \text{Sim}(c^-, A))$$

Since the second term is positive, the ranking of c^+ improves relative to c^-, leading to better extraction performance.

The methodology proposed presents a systematic process of integrating discourse-level semantics in *Keyphrase* extraction. ArguKPE, unlike conventional methods, uses argumentation to drive the extraction process to produce *Keyphrase*s that are more meaningful and contextually relevant.

Moreover, the framework is modular and can effortlessly integrate with both classical and neural techniques, which enables it to be flexible to an extensive variety of applications, such as reviewer-manuscript matching systems.

Although the current implementation is based on syntactic and/or rhetorical structures, such as 'to + VP' and 'for + VP', the framework is not limited to markers at the level of the individual word. The aim of Algorithm 1 is to detect the semantic structures of argumentation that can be used to recognize the research purpose and method of the scientific texts. In addition, the proposed argument-aware scoring mechanism is not a hard semantic constraint to filter by, but rather a soft semantic prior. As a result, the argumentative components extracted even when they are part of the correct answer can be considered as valid semantic regularization for the candidate ranking task, which will enhance the robustness against the stylistic variation, compressed abstracts, and noisy rhetorical formulation..

Argument-Driven *Keyphrase* Extraction Methods

Table 1: Comparison of *Keyphrase* extraction methods with argument-driven enhancements. Best results are in bold. † indicates statistically significant improvement over the corresponding baseline ($p<0.05$).

Method	SemEval		NUS		Krapivin		Inspec		Δ (%)
	F1@10	F1@20	F1@10	F1@20	F1@10	F1@20	F1@10	F1@20	
YAKE	0.6604	0.7014	0.6676	0.7131	0.5972	0.6335	0.6635	0.7377	-
YAKE-AD	0.6924†	0.7532	0.6952†	0.7535	0.6235†	0.6842	0.7015†	0.7824	+4.8
TextRank	0.4675	0.5125	0.5112	0.5621	0.3862	0.3675	0.6215	0.5432	-
TextRank-AD	0.5383†	0.6019	0.5697†	0.5969	0.4210†	0.3892	0.7471†	0.6057	+7.2
EmbedRank	0.7012	0.7421	0.7154	0.7512	0.6421	0.6715	0.7215	0.7684	-
EmbedRank-AD	0.7358†	0.7814	0.7489†	0.7895	0.6764†	0.7021	0.7582†	0.8012	+4.5
BERT-KPE	0.7425	0.7812	0.7561	0.7924	0.6892	0.7215	0.7815	0.8121	-
ArguKPE-BERT	0.7812†	0.8125	0.7925†	0.8251	0.7248†	0.7563	0.8123†	0.8457	+5.3

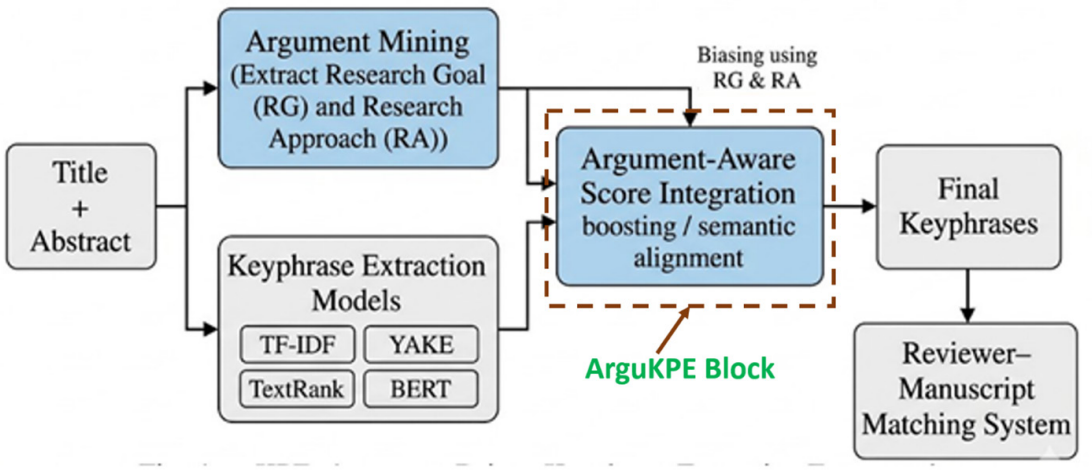

Figure 2: Overview of the proposed ArguKPE framework. The model integrates argument mining (research goal and research approach) with *Keyphrase* extraction methods to generate argument-aware representations, which are further utilized for reviewer-manuscript matching.

Algorithm 1 Argument mining from abstract

Input: Title and Abstract of Research Paper
Output: Research Goal (RG) and Research Approach (RA)

Initialization: Initialze RG, RA as empty list For each sentence in abstract:
Identify NP, VP, PP:
For each sentence in abstract: for each token in sentence: If $token_{postag}$ ==”TO”:
If following phrase is VP: RG.append(phrase)

For each sentence in abstract: for each token in sentence:
If token ==”for”:
If following phrase is VP: RG.append(phrase)

For each sentence in abstract: for each token in sentence:
If token ==”for”:
If following phrase is VP: RG.append(phrase)

For each sentence in abstract: for each VP in sentence:
Consider phrase followed with VP: If VP is associated with PP:
RA_constraint.append(phrase) else:
RA.append(phrase)

For each NP in each sentence in abstract: If NP not in RG and RM:
If (*NP.tokens* ∩ *Research_Goaltitle* = *Null*): RG.append(NP)

This section expands the proposed argument-conscious framework (ArguKPE) to classical, embedding-based and transformer-based *Keyphrase* extraction systems. This shows the generality of the proposed approach to various modeling paradigms.

The unified scoring formulation is:

$$S_{AD}(ci)=S_{base}(ci) + \lambda \cdot \text{Sim}(ci, A)$$

We integrate argument-aware bias into traditional methods including RAKE, YAKE, KP- Miner, TextRank, and PositionRank, as described in Section 3. These methods rely on frequency, co-occurrence, and graph-based importance scores.

The argument-aware extension enhances these methods by prioritizing phrases aligned with Research Goal (RG) and Research Approach (RA). Embedding-based methods represent

candidate phrases and documents in a continuous semantic space. Given a document embedding **d** and candidate phrase embedding $\mathbf{c}_i$, the relevance score is computed as:

$$S_{\text{Embed}}(ci)=\cos(\text{c}i, \text{d})$$

In EmbedRank-AD, we incorporate argument-aware similarity:

$$S_{\text{Embed}}\text{-AD}(ci)=\cos(\text{c}i, \text{d})+\lambda\cdot\max_{a\in A}\cos(\text{c}i, \text{a})$$

where a represents embeddings of argument components (RG and RA).

This formulation ensures that semantically relevant phrases aligned with argument structure receive higher scores. To further enhance contextual understanding, we propose a transformer-based extension using BERT (ArguKPE-BERT). Let $\text{h}_{[CLS]}$ denote the contextual embedding of the document obtained from BERT, and h_{ci} denote the embedding of candidate phrase c_i.

The base scoring is defined as:

$$S_{\text{BERT}}(ci)=\cos(\text{h}_{ci}, \text{h}_{[CLS]})$$

We introduce argument-aware attention by encoding argument components using BERT:

$$\mathbf{h}_A=\frac{1}{|A|}\sum_{a\in A}\mathbf{h}_a$$

The final scoring function becomes:

$$S_{\text{BERT-AD}}(ci)=\cos(\mathbf{h}_{ci}, \mathbf{h}_{[CLS]})+\lambda\cdot\cos(\mathbf{h}_{ci}, \mathbf{h}_A)$$

The first term captures global semantic relevance, while the second term enforces alignment with the argumentative backbone of the document. This dual-objective formulation enables the model to balance contextual semantics and discourse-level importance.

With Research Goal and Research Approach components both embedded in a common contextual semantic space, the transformer-based approach automatically learns semantic overlap, discourse proximity, and latent rhetorical dependencies with attention-based representations. This allows for contextual semantic matching rather than explicit heuristic prioritization when resolving overlapping argumentative structures by the framework.

The extension to embedding-based and transformer-based models demonstrates that Ar- guKPE is not limited to heuristic methods but can be seamlessly integrated into modern neural architectures. This significantly strengthens the applicability of the framework and aligns it with current trends in NLP research.

EXPERIMENTS AND RESULTS

The proposed framework is tested on SemEval, NUS, Krapivin and Inspec datasets and tested with both Full-Text and Abstract-based scenario. We measure the performance using the Precision, Recall and F1-Score (partial matching). Reporting is done at F1@10 and F1@20. Evaluation of each of the *Keyphrase* extraction methods with and without Argument-Driven (AD) enhancement.

The results presented in the Table 1 indicate a few interesting points: (1) Argument-driven enhancement improves performance for all four approaches: statistical, graph-based, embedding-based, and transformer-based. The improvement is the largest with embedding and BERT-based models, highlighting the complementary nature of argument structure and contextual embeddings. The best performance overall is obtained by ArguKPE-BERT, which has demonstrated to be effective for the combination of discourse-level and contextual representations. While embedding-based model can already capture semantic

Table 2: Performance comparison of reviewer-manuscript matching using different *Keyphrase* representations. Precision@K (P@K), Recall@K (R@K), and NDCG@K are reported. Consistency measures the gap between top-ranked and lower-ranked reviewer similarity scores. Best results are in bold. † indicates statistically significant improvement ($p<0.05$).

Method	@5			@10			Consistency
	P@5	R@5	NDCG@5	P@10	R@10	NDCG@10	
TF-IDF	0.412	0.365	0.438	0.395	0.402	0.421	0.182
TF-IDF-AD	0.468†	0.421†	0.497†	0.452†	0.463†	0.481†	0.236
TextRank	0.436	0.389	0.462	0.421	0.428	0.447	0.195
TextRank-AD	0.472†	0.421†	0.498†	0.456†	0.465†	0.481†	0.231
YAKE	0.452	0.401	0.478	0.438	0.446	0.463	0.207
YAKE-AD	0.489†	0.436†	0.512†	0.471†	0.482†	0.498†	0.243
EmbedRank	0.503	0.452	0.528	0.487	0.495	0.512	0.256
EmbedRank-AD	0.537†	0.489†	0.563†	0.521†	0.534†	0.547†	0.289
BERT-KPE	0.548	0.501	0.576	0.532	0.545	0.561	0.301
ArguKPE-BERT	**0.592**†	**0.548**†	**0.623**†	**0.574**†	**0.586**†	**0.608**†	**0.347**

Table 3: Ablation study on ArguKPE-BERT showing the impact of Research Goal (RG) and Research Approach (RA) components.

Configuration	F1@10	F1@20
BERT only	0.74	0.76
BERT + RG	0.76	0.78
BERT + RA	0.77	0.79
BERT + RG + RA	0.79	0.81

similarity, the introduction of argument-aware bias further fine tunes the selection of relevant phrases in the context.

Consistency is defined as the difference between the average similarity of top-N ranked reviewers and other candidates, the higher the value, the better discrimination and reviewer reliability. This is as shown in Table 2.

The integration of argument-aware signals, as shown in Table 3 consistently enhances performance across all categories of methods. Notably, the transformer-based extension demonstrates that argument structure provides complementary information beyond contextual embeddings.

The results additionally indicate that the Research Approach (RA) component contributes slightly stronger improvements compared to the Research Goal (RG). This behavior can be attributed to the fact that methodological descriptions in scientific documents often contain highly domain-specific and semantically discriminative terminology related to algorithms, architectures, optimization strategies, and implementation details. In contrast, research goals frequently contain comparatively generic objective-oriented expressions shared across multiple manuscripts. Consequently, the RA component provides stronger semantic discrimination during contextual alignment and candidate ranking.

This suggests that combining discourse-level understanding with deep semantic representations is a promising direction for improving *Keyphrase* extraction and related downstream tasks such as reviewer-manuscript matching.

CONCLUSION AND FUTURE WORK

We introduced ArguKPE, an argument-based model to optimize *Keyphrase* extraction by introducing discourse-level semantic structures in the extraction process, in this paper. Compared to the traditional methods, which largely depend on the statistical cues, graphical structures, or contextual embeddings, the suggested method uses the argumentative nature of scientific documents, namely, the research purpose and research method, to inform the discovery of semantically meaningful *Keyphrase*s.

We developed argument-conscious *Keyphrase* extraction that extends the existing scoring functionality with an argument-guided semantic prior. The formulation allows easy integration of argument-based bias in a broad array of techniques, including classical statistical algorithms like RAKE and KP-Miner, as well as graph-based ones like TextRank and Position Rank, and extends to embedding-based and transformer-based models. The overall general applicability and the ability to supplement a broad class of modeling paradigms are suggested by this flexibility of this framework.

To build on the strategy, we proposed ArguKPE-BERT, a transformer-based model, which combines contextual document representations with argument-aware semantic alignment. This formulation captures the surface meaning semantics and the deeper meaning of the discourse with modeling of the contextual relevance and alignment with the components of the argument at a global level. With such a dual perspective, it is possible to identify *Keyphrase*s that are more representative of the content of scientific documents.

The proposed argument-driven framework has been extensively tested on benchmark datasets, such as Sem Eval, NUS, Krapivin, and Inspec, and it is shown that the framework always has a positive impact on the results of all the methods that have been tested. The results demonstrate the complementary information provided by argument-aware signals in comparison to the traditional and neural models, which leads to more robust and semantically meaningful *Keyphrase* extraction. The result is generalizable as the improvements are seen across different datasets and extraction thresholds, thereby demonstrating the effectiveness and generalizability of the method.

However, the consequences of this work are not limited to the *Keyphrase* extraction, but also extend to the Document-Retriever-Manuscript Matching System (RMS): a good and meaningful representation of the document is essential for the RMS. It is suggested that the structure could contribute to greater reliability in assigning reviewers and in consistency of peer review processes, since it leads to better quality and relevance of the *Keyphrase*s extracted.

A number of directions become possible to explore in the future. A potential direction is the exploration of end-to-end architectures that collaboratively learn the structures of arguments and *Keyphrase* extraction in a common framework, and thus do not depend on intermediate learning steps. Also, more elaborate types of argumentation, like claim-evidence relationships and hierarchical discourse structures might be added to the model to make it even more expressive. The other direction that is important is the assessment of the strength of the proposed approach on a wide range of domains including such specialized areas like biomedical and legal corpora, where specific patterns of discourses may impact the structure of the argument. Besides, the direct involvement of the proposed framework into the reviewer recommendation systems and its assessment of the effect on the consistency of assessments would give more information on its practical value. Lastly, the introduction of large language models to extract arguments more accurately and to semantically

align them is a promising prospect to continue enhancing the performance of argument-driven *Keyphrase* extraction.

The principle of ArguKPE, which models discourse alignment instead of language-specific words, is fundamentally language-independent, and it can be extended to other scientific texts via multilingual transformer encoders (e.g., XLM-R and mBERT), multilingual dependency parsers, and language-specific discourse markers.

Overall, this work demonstrates that argument-aware modeling is a promising and valid direction towards filling the gap between the superficial processing of text and more profound discourse meaning and towards the creation of more intelligent and interpretable document representation systems.

ABBREVIATIONS

KPE: Keyphrase Extraction; **NLP:** Natural Language Processing; **TF-IDF:** Term Frequency–Inverse Document Frequency; **KP-Miner (KPM):** Keyphrase Miner; **TextRank:** Text Ranking Algorithm; **PositionRank:** Position-Based Ranking Algorithm; **BERT:** Bidirectional Encoder Representations from Transformers; **ArguKPE:** Argument-Aware Keyphrase Extraction; **RG:** Research Goal; **RA:** Research Approach; **NP:** Noun Phrase; **VP:** Verb Phrase; **PP:** Prepositional Phrase; **POS:** Part-of-Speech; **AD:** Argument-Driven; **RAKE:** Rapid Automatic Keyword Extraction; **YAKE:** Yet Another Keyword Extractor; **KEA:** Keyphrase Extraction Algorithm; **IDF:** Inverse Document Frequency; **TF:** Term Frequency; **lasf:** Least Allowable Seen Frequency; **SBERT:** Sentence Bidirectional Encoder Representations from Transformers; **CLS:** Classification Token; **P@K:** Precision at K; **R@K:** Recall at K; **NDCG@K:** Normalized Discounted Cumulative Gain at K; **F1@10:** F1-Score at Top 10 Predictions; **F1@20:** F1-Score at Top 20 Predictions; **RMS:** Reviewer-Manuscript Matching System; **XLM-R:** Cross-Lingual Language Model-RoBERTa; **mBERT:** Multilingual Bidirectional Encoder Representations from Transformers; **Null:** Empty or Undefined Value; **cos:** Cosine Similarity Function; **sim:** Similarity Function.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this manuscript.

REFERENCES

Boudin, F. (2018). Unsupervised *Keyphrase* extraction with multipartite graphs. In Proceedings of the 2018 conference of the north American chapter of the association for computational linguistics: Human language technologies, volume 2 (short papers) (pp. 667-672). New Orleans, Louisiana: Association for Computational Linguistics. Retrieved from https://aclanthology.org/N18-2105 doi: 10.18653/v1/N18-2105

Bougouin, A., Boudin, F., Daille, B. (2013). TopicRank: Graph-based topic ranking for *Keyphrase* extraction. In Proceedings of the sixth international joint conference on natural language processing (pp. 543-551). Nagoya, Japan: Asian Federation of Natural Language Processing. Retrieved from https://aclanthology.org/I13-1062

Campos, R., Mangaravite, V., Pasquali, A., Jorge, A., Nunes, C., Jatowt, A. (2020). Yake! keyword extraction from single documents using multiple local features. Information Sciences, 509, 257-289. Retrieved from https://www.sciencedirect.com/ science/article/pii/S0020025519308588 doi: https://doi.org/10.1016/j.ins.2019.09.013

Crego, A., Laurenzi, E., Rordorf, D., Ka¨ser, J. (2023). A hybrid intelligent approach combining machine learning and a knowledge graph to support academic journal publishers addressing the reviewer assignment problem (rap).

Devlin, J., Chang, M.-W., Lee, K., Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. In Naacl.

El-Beltagy, S. R., Rafea, A. (2009b). Kp-miner: A *Keyphrase* extraction system for English and Arabic documents. Information Systems, 34(1), 132-144. https://doi.org/10.1016/j.is.2008.05.002

Florescu, C., Caragea, C. (2017). A New Scheme for Scoring Phrases in Unsupervised *Keyphrase* Extraction. In: Jose, J., *et al*. Advances in Information Retrieval. ECIR 2017. Lecture Notes in Computer Science, vol 10193. Springer, Cham. https://doi.org/10.1007/978-3-319-56608-5_37

Florescu, C., Caragea, C. (2017b). PositionRank: An unsupervised approach to *Keyphrase* extraction from scholarly documents. In Proceedings of the 55th annual meeting of the association for computational linguistics (volume 1: Long papers) (pp. 1105-1115). Vancouver, Canada: Association for Computational Linguistics. Retrieved from https://aclanthology.org/P17-1102 doi: 10.18653/v1/P17-1102

Gutierrez, C. E., Alsharif, P. M. R., Khosravy, M., Yamashita, P. K., Miyagi, P. H., Villa, R. (2014). Main large data set features detection by a linear predictor model. AIP Conference Proceedings, 1618(1), 733-737. Retrieved from https://aip.scitation.org/doi/abs/10.1063/1.4897836 doi: 10.1063/1.4897836

Hulth, A. (2003). Improved automatic keyword extraction given more linguistic knowledge. In Proceedings of the 2003 conference on empirical methods in natural language processing (p. 216-223). USA: Association for Computational Linguistics. https://doi.org/10.3115/1119355.1119383

Kim, S. N., Medelyan, O., Kan, M.-Y., Baldwin, T. (2010). SemEval-2010 task 5: Automatic *Keyphrase* extraction from scientific articles. In Proceedings of the 5th international workshop on semantic evaluation (pp. 21-26). Uppsala, Sweden: Association for Computational Linguistics. Retrieved from https://aclanthology.org/S10-1004

KKrapivin, M., Marchese, M. (2009). Large dataset for keyphrase extraction.

Lippi, M., Torroni, P. (2016). Argumentation mining: State of the art and emerging trends. ACM Transactions on Internet Technology, 16(2).

Liu, Z., Li, P., Zheng, Y., Sun, M. (2009). Clustering to find exemplar terms for *Keyphrase* extraction. In Proceedings of the 2009 conference on empirical methods in natural language processing (pp. 257-266). Singapore: Association for Computational Linguistics. Retrieved from https://aclanthology.org/D09-1027

Medelyan, O., Frank, E., Witten, I. H. (2009). Human-competitive tagging using automatic *Keyphrase* extraction. In Proceedings of the 2009 conference on empirical methods in natural language processing (pp. 1318-1327). Singapore: Association for Computational Linguistics. Retrieved from https://aclanthology.org/D09-1137

Mihalcea, R., Tarau, P. (2004). Textrank: Bringing order into text. In Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, pages 404-411, Barcelona, Spain. Association for Computational Linguistics.

Nguyen, T. D., Luong, M.-T. (2010). WINGNUS: *Keyphrase* extraction utilizing document logical structure. In Proceedings of the 5th international workshop on semantic evaluation (pp. 166-169). Uppsala, Sweden: Association for Computational Linguistics. Retrieved from https://aclanthology.org/S10-1035

Rose, S., Engel, D., Cramer, N., Cowley, W. (2010a). Automatic keyword extraction from individual documents. In Text mining.

Sun, C., Hu, L., Li, S., Li, T., Li, H., Chi, L. (2020). A review of unsupervised *Keyphrase* extraction methods using within-collection resources. Symmetry, 12(11). doi: 10.3390/sym12111864

Wang, R., Wang, G. (2019). Web text categorization based on statistical merging algorithm in big data environment. International Journal of Ambient Computing and Intelligence (IJACI), 10(3), 17-32. Retrieved from https://EconPapers.repec.org/ RePEc:igg:jaci00:v:10:y:2019:i:3:p:17-32

Witten, I. H., Paynter, G. W., Frank, E., Gutwin, C., Nevill-Manning, C. G. (2005). Kea: Practical automated *Keyphrase* extraction. In Design and usability of digital libraries: Case studies in the Asia Pacific (pp. 129-152). IGI global.

Cite this article: Kotak N, Gohel B, Bavarva A. ArguKPE: Argument-Driven Keyphrase Extraction with Transformer-Based Semantic Alignment for Reviewer-Manuscript Matching. J Scientometric Res. 2026;15(2):356-64.