A Model-Based Framework for News Content Analysis

Fazle Rabbi[1][a], Bahareh Fatemi[1][b], Yngve Lamo[2][c] and Andreas L. Opdahl[1][d]
[1]*Information Science and Media Studies, University of Bergen, Norway*
[2]*Department of Computer science, Electrical Engineering and Mathematical Sciences, Western Norway University of Applied Sciences, Bergen, Norway*

Keywords: Category Theory, Content Analysis, Model-Based Framework, Knowledge Graph, Natural Language Processing, Computational Journalism.

Abstract: News articles are published all over the world to cover important events. Journalists need to keep track of ongoing events in a fair and accountable manner and analyze them for newsworthiness. It requires an enormous amount of time and effort for journalists to process information coming from mainstream news media, social media from all over the world, as well as policy and law circulated by governments and international organizations. News articles published by different news providers and reporters may also be subjective due to the influence of reporters' backgrounds, world views and opinions. In today's journalistic practice there is a lack of computational methods to support journalists to investigate fairness and monitor and analyze massive information streams. In this paper we present a model-based approach to analyze the perspectives of news publishers and monitor the progression of news events from various perspectives. The key concepts in the news domain such as the news events and their contextual information is represented across various dimensions in a knowledge graph. We presented a multi dimensional and comparative news event analysis method for analyzing news article variants and for uncovering underlying storylines. To show the applicability of the proposed method in real life, we also demonstrate a running example. The utilization of a model-based approach ensures the adaptability of our proposed method for representing a wide array of domain concepts within the news domain.

1 INTRODUCTION

In every human community, individuals bring news to one another. News has an important role in humankind and journalists are involved in carrying out the task in a professional way. While reporting about real life events through news articles, journalists turn facts into stories and analyses that engage an audience (Schudson, 2020). While crafting news, good journalists put reality first and they follow the core principles of ethical journalism (EJN, 2023), which include *trust and accuracy*; *independence*; *fairness and impartiality*; *humanity*; *accountability*. However there is no bias-free journalism (Schudson, 2020) in reality. The problem of bias in media has been an important topic and it requires sophisticated techniques to analyze the media bias in a systematic way. Journalists also need to keep track of the ongoing events all around the world and analyze the events carefully as they need to inform their audience about the changing world. Since there is an abundance of news articles being published all over the world by several news media outlets, journalists would benefit from techniques for systematically analyzing events from news publications. Sociologists, historians, political scientists, information scientists are involved in gathering information from news articles and extract insightful information. In this paper we present a model-based framework that employs a diverse range of models to represent knowledge from news articles and uses computational methods for the analysis of news events. The framework integrates the following components:

- state-of-the-art natural language processing technique for parsing content from news articles;
- a multi dimensional meta-model allowing data to be arranged into hierarchical groups and a knowledge graph schema for structuring event related

[a] https://orcid.org/0000-0001-5626-0598
[b] https://orcid.org/0000-0002-8944-5051
[c] https://orcid.org/0000-0001-9196-1779
[d] https://orcid.org/0000-0002-3141-1385

Rabbi, F., Fatemi, B., Lamo, Y. and Opdahl, A.
A Model-Based Framework for News Content Analysis.
DOI: 10.5220/0012306800003645

In *Proceedings of the 12th International Conference on Model-Based Software and Systems Engineering (MODELSWARD 2024)*, pages 99-107
ISBN: 978-989-758-682-8; ISSN: 2184-4348

information;

- a content comparison method based on category theory; and
- a statistical analysis method for analyzing news article variants.

The knowledge graph represents news events with relevant information e.g., source article, publication date, involved persons, involved countries, and type of event. We annotate news events with IPTC (International Press Telecommunications Council) Media Topics. IPTC is a global standardization organization that provides metadata standards for the news industry. The terms in the IPTC Media Topics are represented in a hierarchical structure which allows us to extract news events from different abstraction levels. By combining different attributes and relationships of news events along with the domain ontology in IPTC Media Topics, the framework allows users to extract different views of news events from a knowledge graph. The framework integrates a computational model based on category theory which allows us to analyze news events at a higher abstraction level, for example, to compare and categorize events and to analyze flow of progression of events. We present novel application areas of category theory for analyzing events stored in a knowledge graph. We assume the reader is comfortable with the basics of category theory (Barr and Wells, 1990).

In section 2, we present a method for extracting structured information about news events from news articles using large language models (LLMs). We present a running example while describing the proposed method. In section 3, we present a model-based framework for content analysis. In section 4, we provide a discussion about the proposed method and provide a comparison with existing works.

2 HARVESTING NEWS EVENTS KNOWLEDGE GRAPH WITH A PRE-TRAINED LLMs

Harvesting news events into a knowledge graph is an important topic and it has been investigated in several other projects to support various tasks in the news domain. Opdahl et al. (Opdahl et al., 2022) provides a review of using semantic knowledge graphs in news production, distribution, and consumption, emphasizing their potential for integrating heterogeneous information in the news industry. The Global Database of Events, Language, and Tone (GDELT) is a Google-sponsored project that monitors news media from all over the world and provides a real-time update of events in every 15 minutes (Gde, 2023).

Rospocher et al. present a method to automatically build Event-Centric Knowledge Graphs from news articles using NLP techniques, such as Entity Linking and Semantic Role Labeling (Rospocher et al., 2016). Liu et al. introduces a domain-specific knowledge graph called the "news graph" that incorporates collaborative relations between entities and topic context information for news recommendations (Liu et al., 2019).

Harvesting news events into a knowledge graph has been studied by Berven et al. in (Berven et al., 2020) where they presented a knowledge graph platform for a newsroom. They propose an event detection technique that identifies potentially newsworthy events from clusters of news items according to named entities, topics, and location.

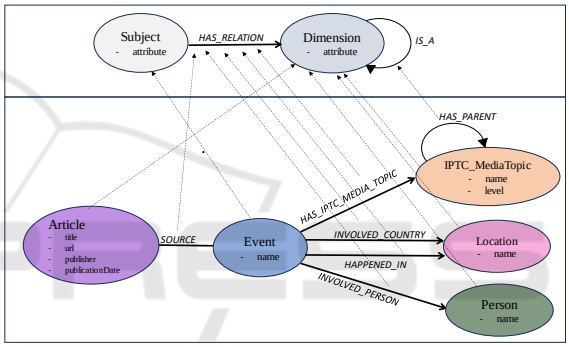

Figure 1: Dimensional meta-model (top) and Knowledge Graph Schema (bottom) for structuring event related information.

In our proposed technique, we take input from GDELT in every 15 minutes. The input includes web addresses to news article texts. The news article texts are parsed for analysis using a pre-trained LLMs. We use GPT-Turbo 3.5 for extracting information from news article texts and harvest news event related information. To structure the information about news events we propose to use a dimensional meta-model (Figure 1 top) which allows storing events with contexts along various dimensions in a hierarchical model. The bottom of Figure 1 shows a knowledge graph schema for structuring an event and its contextual information such as event location, event type and involved countries. The knowledge graph is also enriched with IPTC Media Topics. The knowledge graph allows us to access the hierarchical information from the IPTC Media Topics ontology by traversing over *:HAS_PARENT* relationships. A Neo4j graph database has been used to store news events and their relationships with other entities. The information model is centered around *Event* and it

also allows us to preserve the epistemic view of individual publishers. For example, if two publishers publish 2 news articles about a certain event, we will be storing 2 instances of *Event* (along with their contextual information) in our knowledge graph.

Table 1 illustrates the prompt we have used to extract structured information from news article texts. The temperature is set to 0 to limit the creativity of the LLMs. The prompt includes an instruction about producing output in JSON format. It also includes instructions to classify events using IPTC Media topic names. However, the GPT 3.5 Turbo model generates slightly different names from what we have in the knowledge graph. For example, in our knowledge graph we have 'arts, culture and entertainment' but the output of the prompt may include the following name: 'Cultural, Arts and Entertainment'. It is therefore necessary to perform a similarity analysis of the media topic names. To find out the most similar topic name we calculated cosine-similarity using Python and the Spacy library.

Table 1: Prompt for extracting event related information.

prompt = "'*Write the name of the event, type of the event, involved person, involved countries and the location of the event from the following news. Use IPTC media topic name while writing values for 'Event type'. Write full name while mentioning involved persons and locations. Write only name of persons if they are known. No need to include any unknown person. Also do not need to write the designation or position of the persons. While returning the location, mention the country where the event took place. While returning the iptc media topic names, please return the output for which you are significantly confident about. If there are more values, include all of them in comma separated format. Format your answer as a JSON object with the following key-values:*
{"Event": "event-name",
"Event Type": "iptc-media-topic-name",
"Involved Countries": "country-name",
"Location of Event": "country-name",
"Involved-Person": "Person-name", } "'
response = openai.ChatCompletion.create(
model="gpt-3.5-turbo",
messages=[
"role": "system", "content": prompt,
"role": "user", "content": articleText],
temperature=0, max_tokens=256, top_p=1,
frequency_penalty=0, presence_penalty=0)

{ 'Event': "Closure of Niger's Airspace",
'Event Type': 'Civil Unrest',
'Involved Countries': 'Niger, United Kingdom, South Africa',
'Location of Event': 'Niger',
'Involved-Person': 'President Mohamed Bazoum, General Abdourahmane Tchiani' }

The proposed method in this paper is demonstrated with a running example which includes a knowledge graph of news events about *Niger* and *Gabon* extracted from the news articles published by 6 media outlets (*aljazeera.com, theguardian.com, reuters.com, independent.co.uk, nytimes.com, washingtontimes.com*) from July 28th, 2023 to September 2nd, 2023. The knowledge graph consists of news events in *Niger* and *Gabon* about two coups that took place during the above-mentioned period.

3 MODEL-BASED FRAMEWORK FOR CONTENT ANALYSIS

We propose a new model-based framework for news content analysis that includes techniques for multi dimensional comparative analysis. The framework allows us to analyze different perspectives on news contents; progression of events from a variety of abstraction levels with various perspectives. The framework allows the user to select an appropriate dimension and abstraction level. For instance, a user might be interested in comparing the perspectives of different publishers over a certain period of time or the progression of events at a certain level of abstraction. The knowledge graph includes events and their contextual information along various hierarchically organized dimensions. For example, the IPTC Media topic ontology organizes topic names in a hierarchy. The highest level of abstraction in the IPTC Media topic ontology (i.e., level 1 of the ontology) includes 17 media topic names. The selection of dimension and abstraction level is used for extracting information from the knowledge graph. The information is then used for comparative analysis. The analysis results are used for extracting patterns of variants. We propose a semi-automated approach where humans are involved in the process of variant analysis. Figure 2 illustrates the model-based framework which employs models for representing computational methods for the analysis of news events. Graph patterns are used to specify search criteria. We propose to use categorial operations to perform comparative analysis over the search results (i.e., subgraphs). Category theory allows us to deal with abstract structures and relationships between them. It allows us to study the news content from high levels of abstraction and thereby enables us to gain deeper insights into media contents. In this paper we focus on the perspective comparison, progression of events, and variant analysis. The model-based framework is adaptive to new dimensions with more contextual information, for example numbers of casualties, sentiments, proximity, news angles, etc.

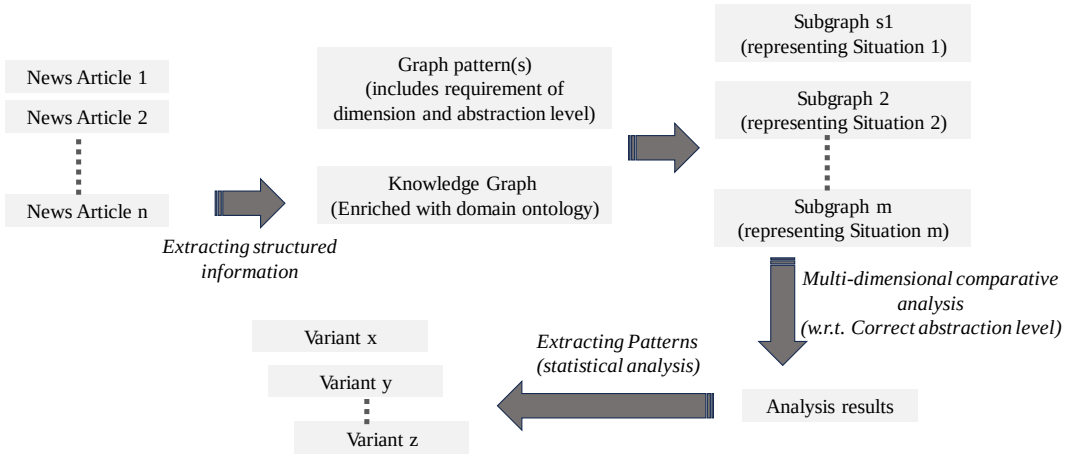

Figure 2: Model-based framework for multi dimensional comparative analysis of news contents.

3.1 Perspective Comparison

We store contextual information about news events, such as event location, event type, involved countries and individuals, in the knowledge graph. When reporting on specific events and their subsequent developments, different publishers may have reported them differently. In our proposed method, we compare the perspectives across various dimensions of these events. For instance, we examine the types of events that were reported by different publishers during a specific time period while they were covering a particular event and its subsequent development.

To compare the perspectives of different publishers, we propose employing category theory operations, including pullback and commutative diagrams. Figure 3 gives an overview of the proposed method for perspective analysis. All the news article related information in the graph database is represented as I in the figure. S_1 and S_2 represent the reports from two different publishers. S_1 and S_2 can be computed by querying the graph database. Cypher queries (Cyp, 2023) may be used to extract the fragments of graphs (i.e., subgraphs) from I which represents the local perspective of individual publishers. For instance, we may be interested to know the extent to which the media topics used by different publishers match and differ while they report about some events in their published news articles. The figure shows pullback object C which is computed from the following two morphisms: $S_1 \xrightarrow{m_1} I$ and $S_2 \xrightarrow{m_2} I$. From the pullback object, we can figure out the perspectives of different publishers as shown in Figure 3 by object D_1 and D_2.

Here we explain the proposed method with a running example. We compute the perspectives from two publishers about their news stories that cover

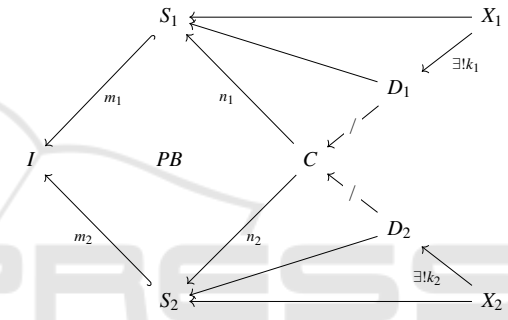

Figure 3: Pullback object (C) computes the commonality between S_1 and S_2; D_1 and D_2 objects are used to compute the dissimilarities between S_1 and S_2.

events in *Niger* from July 28th to September 2nd. We use Cypher queries to get data from the Neo4j graph database. Cypher queries can be expressed as graph patterns which include variables. The results of these queries would be subgraphs of the whole graph database. One can compute the pullback object by writing a program using a general purpose programming language (e.g., Python using the Neo4j library) but in this paper we present a Cypher query (Figure 4) which computes the pullback object of two subgraphs from a graph database by combining two Cypher queries as shown above. We ensure that the diagram commutes by specifying $t1 = t2$ as a condition in the query. Since the two subgraphs S_1 and S_2 include only nodes of type $IPTC\_MediaTopic$, we include $IPTC\_MediaTopic$ nodes in the result pullback object. Figure 4 shows a cypher query expression to compute the pullback object of $S_1 \rightarrow I$ and $S_2 \rightarrow I$.

The perspectives of the publishers are computed from the difference of the subgraphs S_1 and S_2 with the pullback object. Here we have demonstrated the perspective analysis with respect to IPTC media top-

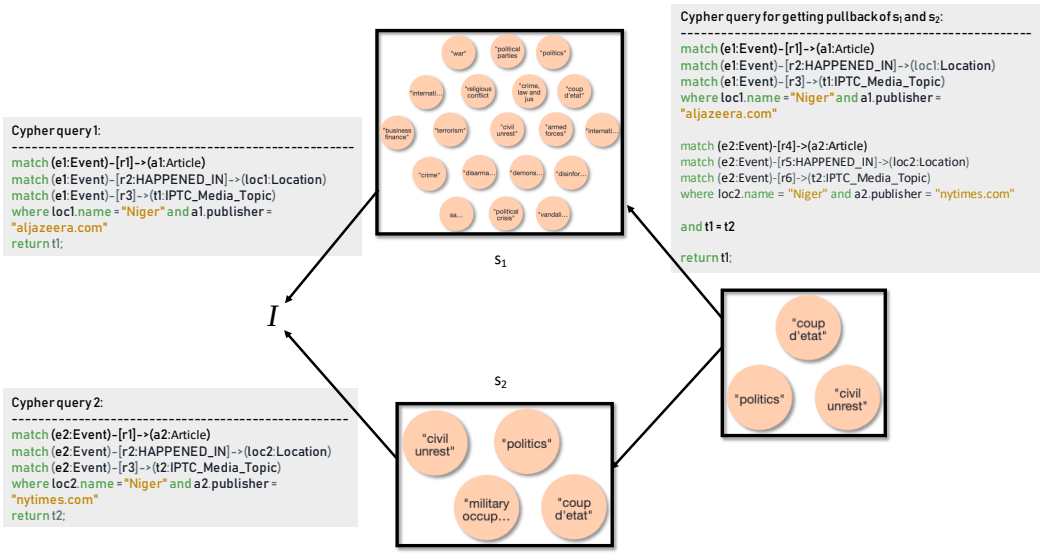

Figure 4: Computing pullback with Cypher query.

ics but the other dimensions can also be used for perspective analysis.

3.2 Analyzing the Progression of Events

To analyze the progression of events in computational journalism is a complex task as there is an abundance of information. Numerous publishers from all around the world publish about ongoing events. There is a lack of tool support in computational journalism to keep record of the events and systematically analyze them to extract insightful information about the progression of events. We propose (1) to use features such as names, locations and IPTC topics to group news articles covering stories about closely related topics and, then, (2) to use category theory to analyze the progression of events by means of analyzing contents in news articles. We reuse the concept presented in Figure 3 where we adapt S_1 and S_2 with a selection of events capturing situations from $time_{x1} - time_{y1}$ and $time_{x2} - time_{y2}$ respectively. From S_1 and S_2 we systematically compare the evolution of events from $time_{x1} - time_{y1}$ to $time_{x2} - time_{y2}$. For example, S_1 and S_2 may represent the IPTC media topics being used to cover the news events about *Niger* from July 31st-August 6th and August 7th-August 13th respectively. From these subgraphs we compute the emerging IPTC media topics in the reports published during August 7th-August 13th. This comparative analysis allows journalists to get an overview of the progression of events.

The progression of events can be represented as a transformation of IPTC media topics being covered by the publishers. Let us consider that

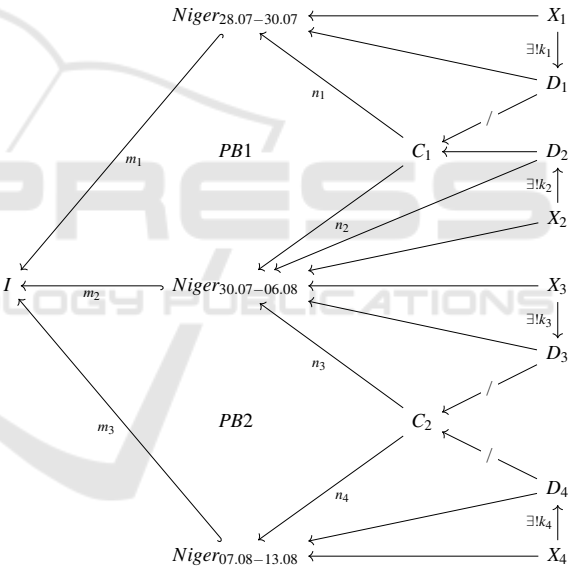

Figure 5: Capturing the progression of events with pullback operation.

in Figure 5, $Niger_{28.07-30.08}$, $Niger_{31.08-06.08}$ and $Niger_{07.08-13.08}$ are representing the IPTC media topics being used to cover the news events in *Niger* for periods July 28th-July 30th, July 31st-August 6th and August 7th-August 13th respectively. The pullback object C_1 and C_2 represents the commonality of the events (with respect to IPTC media topics) in $Niger_{28.07-30.08}$, $Niger_{31.08-06.08}$ and $Niger_{31.08-06.08}$, $Niger_{07.08-13.08}$ respectively. The object D_1 would capture the media topics being removed from the reporting during July 30th-August 6th; D_2 would capture the media topics being newly

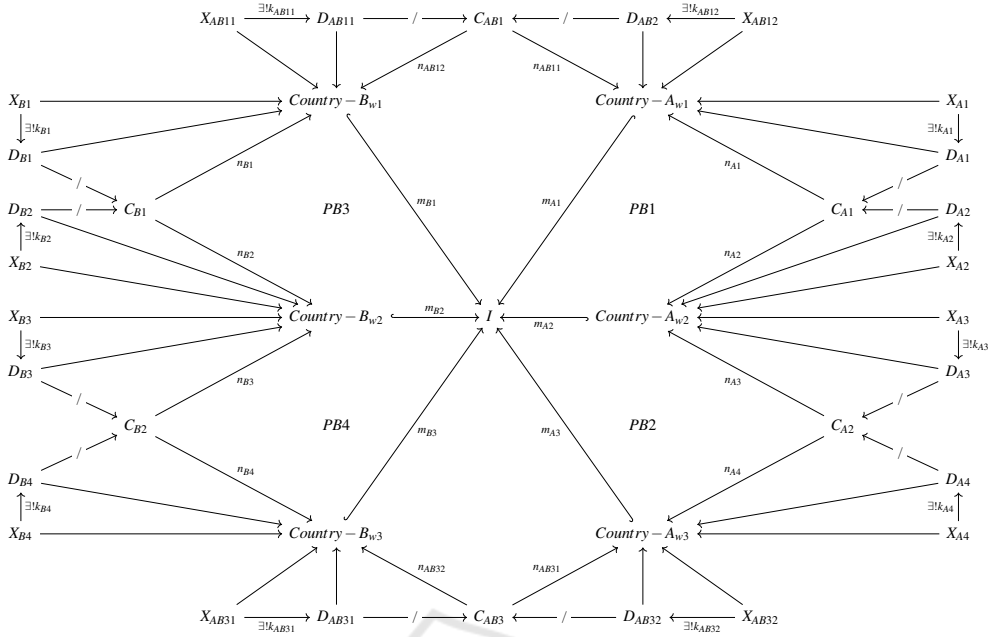

Figure 6: Comparison of progression of events.

added during July 30th-August 6th. Similarly, D_3 would capture the media topics being removed from the reporting during August 7th-August 13th and D_4 would capture the media topics being added during August 7th-August 13th.

Similar categorical operations can also be used to analyze the progression of events in two different countries. Let us consider that we want to analyze the weekly progression of events in *Niger* and *Gabon* since the coup started in those two countries. Figure 6 illustrates a computational model for such analysis. The pullback object C_{AB1} is the commonality between the progression of events in the two countries $Country-A$ and $Country-B$ in the first week where $Country-A_{w1}$ and $Country-B_{w1}$ represents contextual information of events (such as IPTC media topics or involved countries or individuals) reported in the first week. For brevity we did not show C_{AB2} (pullback object between $Country-A_{w2}$ and $Country-B_{w2}$) in the diagram. Common development between the two countries progression can be found from the pullback objects $C_{AB1}, C_{AB2}, C_{AB3}$, etc.

Figure 7 illustrates a computation model for the comparison of progression of events at a higher level of abstraction. $\alpha_1, \alpha_2, \beta_1, \beta_2$ represents contextual information of events specified at a certain abstraction level j; In our running example we only have a hierarchical data model for IPTC Media topics, therefore, all the IPTC Media topics in $\alpha_1, \alpha_2, \beta_1, \beta_2$ are at level j in the IPTC Media topic ontology. $\alpha_1', \alpha_2', \beta 1', \beta 2'$ represents contextual information of events specified at a higher level of abstraction. The pullback objects $C_{\alpha\beta i}$ (where $i = 1,2$) in the bottom layer represent the commonality of the progression of events. The arrows between layers represent graph homomorphisms between corresponding elements from lower to higher levels of abstraction in the knowledge graph I.

Theorem: For any non-empty pullback object $C_{\alpha\beta i}$ (where $i = 1,2$) at level j, the corresponding pullback objects $C'_{\alpha\beta i}$ at level $k < j$ is non-empty.

Proof Sketch: Consider a non-empty pullback object $C_{\alpha\beta i}$ (where $i = 1,2$) at level j; this would require at least one element $n_a \in \alpha_i$ and one element $n_b \in \beta_i$ where n_a and n_b are mapped to the same element in the knowledge graph. If n'_a (with level k) is a parent of n_a, and n'_b (with level k) is a parent of n_b, then n'_a and n'_b must also map to the same element in the knowledge graph. The pullback objects $C'_{\alpha\beta i}$ should at least contain an element that maps to n'_a and n'_b and therefore cannot be empty.

3.3 Variant Analysis

In this section we present a technique for variant analysis over the computation results from section.3.2. We present an application of statistical analysis method for detecting news article variants. In section 3.2 we presented techniques to retrieve data from a knowledge graph across various dimensions and on various abstraction levels. This selection of data from knowledge graphs are used for identifying variants by applying statistical methods. In this section we present *Exploratory data analysis* for identi-

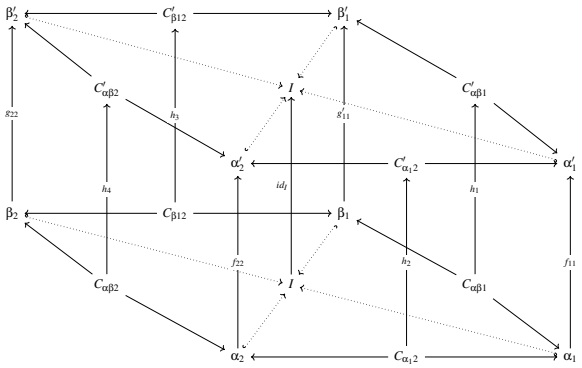

Figure 7: Comparison of progression of events at a higher level of abstraction.

fying trends in time and space and use them for variant analysis.

In order to identify trends in reporting across different topics, we need to select a dimension and abstraction level and extract data from the knowledge graph. Suppose we would like to identify trends of publishers reporting about *civil unrest* in *Niger* from August 1st, 2023-August 20th, 2023, we retrieve the events from the knowledge graph that matches with the *civil unrest* IPTC media topic. The results are therefore used for statistical analysis e.g., frequency distribution and for visualization of trends in a timeline. Visualizing the events in a timeline allows us to depict types of events being reported by different publishers and their engagement in reporting throughout a selected period of time.

Figure 8 highlights the duration of engagement of individual publishers among *al jazeera.com*, *theguardian.com*, *reuters.com*, *independent.co.uk*, *nytimes.com*, *washingtontimes.com*, *cnn.com* for their reporting about *civil unrest* in *Niger*. The background in the figure indicates the $co-limit$ (i.e., a categorical representation of union) of all the events from these publishers about *civil unrest* in *Niger*. From the figure we can extract variants e.g., *independent.co.uk* and *washingtontime.com*'s similarity during the time of publishing about *civil unrest* in *Niger*. However, one might be interested to explore the dataset for identifying trends in other dimensions e.g., the involvement of certain countries in a conflict. Such requirements can be adapted by the proposed method as we can retrieve events that are about any kind of conflict and that involve any countries. We exploit the use of ontological hierarchies for the retrieval of events at the correct abstraction level. For instance, we can identify common trends in the involvement of foreign countries in coups that have taken place in African nations.

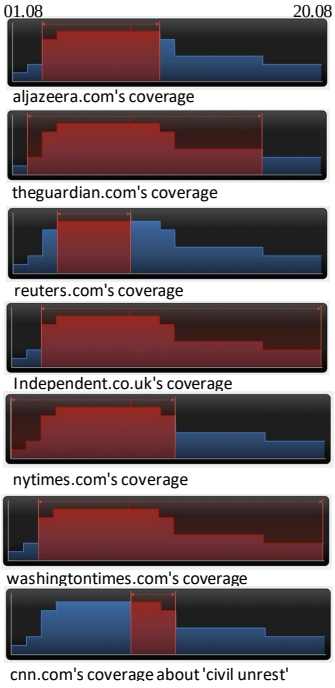

Figure 8: Timeframe showing the engagement of news publishers in reporting about *civil unrest* in *Niger*.

4 DISCUSSION AND FUTURE WORK

The proposed method allows us to analyze the perspective of publishers across different dimensions and abstraction levels, and we have presented how perspective of publishers covering the types of events can be captured. However, there are many other aspects that might be important to capture such as presentations, opinions, etc. In the landscape of news content analysis, various systems such as GDELT(Leetaru and Schrodt, 2013) have been developed for identifying and organizing news events from vast data streams in structured formats. While GDELT efficiently aggregates and quantitatively analyzes vast volumes of news data, offering an overview of the dynamics within the media landscape, a new approach is needed to enable researchers to dive deeper into individual news events.

We presented a model-based framework for content analysis that deviates from traditional news analysis methods that mostly rely on text mining and semantic technologies (Leban et al., 2014; Rudnik et al., 2019). Our proposed method introduces a comprehensive framework that holds the potential to address critical challenges within the media domain. One limitation in the previous research lies in the limited

ability to effectively compare news items with one another. Our model fills this gap by offering a robust mechanism for comparative analysis. As a result, our model empowers users to explore and solve open problems in the field of media with a holistic approach, leading to enhanced insights and deeper understanding of the complex media landscape.

In this paper, our primary focus has been on the analysis of various reports pertaining to a specific event, particularly in terms of perspectives. By focusing into the perspective of reports, we aim to uncover the nuances encapsulated within the media discourse surrounding the event. We can furthermore include the intricacies of reporting angles, tones, and the framing of articles, enriching our understanding of news narratives. Additionally, we have employed a systematic approach to track the evolution and progression of these events over time which provides valuable insights into how events unfold and transform over time, enriching our understanding of their dynamics and implications.

Large language models (LLMs) have demonstrated exceptional performance in specific language-related tasks. However, they also fall short in delivering the structured approach and transparency necessary for conducting in-depth multi-dimensional analyses. Our proposed framework, on the other hand, provides a holistic structure for exploring news, ensuring transparency and facilitating a deeper understanding of news content from various dimensions and abstractions. Moreover, our approach distinguishes itself by offering a high level of abstraction combined with the flexibility for users to select different dimensions for exploration. In contrast to LLMs, our approach goes beyond natural language understanding to incorporate statistical analysis, enriching our capacity to uncover nuanced patterns and insights in news content.

While we have presented some analysis technique using category theory, there is much more to explore and develop in this field. We believe that the integration of generative AI and category theory can contribute to the evolution of journalism in the digital age, fostering transparency, accountability, and enriched news content for both journalists and readers. Particularly, our approach has the capacity to assist in tasks that involve the comparison of news items. For instance, it can be particularly useful in multilingual news comparison, where it can facilitate cross-cultural analysis of news events by overcoming language barriers. Moreover, our model can play a valuable role in fact-checking and verification, aiding in the assessment of news source credibility. Additionally, it is well-suited for bias and framing analysis, enabling the exploration of different perspectives presented in the media. In (Fatemi et al., 2023) we enhanced an existing automated journalism framework by incorporating an awareness of fairness concerns. The integration of a comparative analysis technique into automated journalism processes would be useful for systematically evaluating bias and ensuring the fairness of automatically generated content.

ACKNOWLEDGEMENTS

This research is funded by SFI MediaFutures partners and the Research Council of Norway (grant number 309339).

REFERENCES

(2023). Cypher query language. https://neo4j.com/developer/cypher/. Accessed: 2023-09-25.

(2023). Ethical journalism network. https://ethicaljournalismnetwork.org/who-we-are. Accessed: 2023-09-26.

(2023). GDELT. https://www.gdeltproject.org/data.html. Accessed: 2023-09-12.

Barr, M. and Wells, C. (1990). *Category Theory for Computing Science*. Prentice-Hall, Inc., USA.

Berven, A., Christensen, O. A., Moldeklev, S., Opdahl, A. L., and Villanger, K. J. (2020). A knowledge-graph platform for newsrooms. *Computers in Industry*, 123:103321.

Fatemi, B., Rabbi, F., and Tessem, B. (2023). Fairness in automated data journalism systems. *NIKT: Norsk IKT-konferanse for forskning og utdanning*.

Leban, G., Fortuna, B., Brank, J., and Grobelnik, M. (2014). Event registry: Learning about world events from news. In *Proceedings of the 23rd International Conference on World Wide Web*, page 107–110. ACM.

Leetaru, K. and Schrodt, P. A. (2013). Gdelt: Global data on events, location, and tone, 1979–2012. In *ISA annual convention*, volume 2, pages 1–49. Citeseer.

Liu, D., Bai, T., Lian, J., Zhao, X., Sun, G., Wen, J.-R., and Xie, X. (2019). News graph: An enhanced knowledge graph for news recommendation. In *KaRS@ CIKM*, pages 1–7.

Opdahl, A. L., Al-Moslmi, T., Dang-Nguyen, D.-T., Gallofré Ocaña, M., Tessem, B., and Veres, C. (2022). Semantic knowledge graphs for the news: A review. *ACM Computing Surveys*, 55(7):1–38.

Rospocher, M., Van Erp, M., Vossen, P., Fokkens, A., Aldabe, I., Rigau, G., Soroa, A., Ploeger, T., and Bogaard, T. (2016). Building event-centric knowledge graphs from news. *Journal of Web Semantics*, 37:132–151.

Rudnik, C., Ehrhart, T., Ferret, O., Teyssou, D., Troncy, R., and Tannier, X. (2019). Searching news articles using an event knowledge graph leveraged by wikidata. In

Companion proceedings of the 2019 world wide web conference, pages 1232–1239.

Schudson, M. (2020). *Journalism: why it matters*. Polity Press, United Kingdom.