Evaluating the Inverted Pyramid Structure through Automatic 5W1H Extraction and Summarization

Brian Keith Norambuena*
Virginia Tech
Blacksburg, Virginia, United States
briankeithn@vt.edu

Michael Horning
Virginia Tech
Blacksburg, Virginia, United States
mhorning@vt.edu

Tanushree Mitra
Virginia Tech
Blacksburg, Virginia, United States
tmitra@vt.edu

ABSTRACT

The inverted pyramid is a basic structure of news reporting used by journalists to convey information and it is considered a key element of objectivity in news reporting. In this article, we propose the Inverted Pyramid Scoring method to evaluate how well a news article follows the inverted pyramid structure using main event descriptors (5W1H) extraction and news summarization. We evaluate our proposed method on a proprietary data set of Associated Press news articles across breaking and non-breaking news spanning two topics—political and business. Our results show that the method works at distinguishing the structural differences between breaking and non-breaking news. In particular, our results confirm that breaking news articles are more likely to follow the inverted pyramid structure.

CCS CONCEPTS

• **Computing methodologies** → *Information extraction*; • **Applied computing** → *Publishing*.

KEYWORDS

natural language processing, computational journalism, inverted pyramid, 5W1H extraction

ACM Reference Format:
Brian Keith Norambuena, Michael Horning, and Tanushree Mitra. 2020. Evaluating the Inverted Pyramid Structure through Automatic 5W1H Extraction and Summarization. In *Proceedings of Computation + Journalism Symposium (C+J 2020).* ACM, New York, NY, USA, 7 pages.

1 INTRODUCTION

The inverted pyramid structure—a system of news writing that arranges facts in descending order of importance—has been a cornerstone of journalism since the late 19th century [8]. This style of structuring information emphasizes fact-based reporting and neutrality—two of the key components asserting objectivity in journalistic writing [16]. These elements are also particularly important for hard news reports that require timely reporting and are characterized by high news value (e.g., breaking news stories on political topics). Moreover, scholars have found that an inverted pyramid information structure is a distinctive feature of real journalistic reports; whereas fake news stories often rely on opinion-based reporting and at times are written in structurally ambiguous ways [19]. Thus, determining how well a news story fits the inverted pyramid arrangement could be useful in determining whether the report follows journalistic standards. This paper introduces one such measure for structural analysis of news writing—the Inverted Pyramid Score (IPS).

Prior work on computationally analyzing the inverted pyramid structure includes Zhang and Liu's visual and statistical exploration using rhetorical structure theory [24] and Dai et al.'s classifier-based approach to detect structure in news articles using various lexical, syntactic and semantic features [4]. While these approaches exploit the rhetorical and syntactic structures in news reporting, they do not leverage the two distinctive elements of the inverted pyramid reporting style—summarizing and compressing the most newsworthy aspects of the story at the very start [8] Thus, we design our IPS scoring method by leveraging these two key elements and comprises of the following two components:

(1) **Main event descriptor locations**: The most newsworthy aspects of a story (i.e., the 5W1H questions: who, what, when, where, why and how) are compressed in the opening paragraphs (see Figure 1). The answers to the 5W1H questions describe the main event of the news article.
(2) **Summary similarity**: The opening paragraphs summarize the story in such a way that it is possible to cut out the last paragraphs without losing key information [14]. Thus, any summary of an inverted pyramid news article should be similar to the Opening Paragraphs (OP), composed of the headline, lead, and second paragraph (see Figure 1). We consider the 2nd paragraph in addition to the headline and the lead because it may include some key details, such as the answers to 5W1H questions [8].

We validate and show the effectiveness of our IPS method on a proprietary data set of Associated Press' News articles, comprising news from December 2016 to December 2017. Our experiments demonstrate that our proposed method is capable of evaluating the inverted pyramid structure and, on average, can distinguish the structural differences between breaking and non-breaking news. This is the first step towards evaluating objectivity by the use of the inverted pyramid structure.

2 BACKGROUND AND RELATED WORK

We divide this section into three parts. First, we delve into the different news structures used in written journalism. Next, we focus on the two components that form the core of our IPS calculation, namely 5W1H extraction and text summarization.

2.1 Structure of News Articles

There are many ways to structure a news article. We describe the four most common structures found in the literature [4, 8, 18]. In the *Inverted Pyramid* structure, the article presents content in descending order of importance with key events placed first and

*Also with Universidad Católica del Norte, Department of Computing & Systems Engineering.

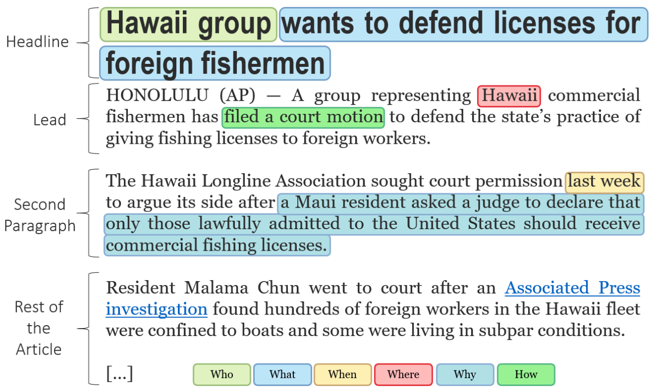

Figure 1: News article [13] showing the Opening Paragraphs (OP) and the rest of the article. The highlighted phrases show the answers to the 5W1H questions that define a news event: *What* happened? *Who* is involved? *When* and *Where* did it happen? *Why* and *How* did it happen?

additional details discussed later. In the *Kabob*, the article starts with an anecdote to capture the reader's attention, then introduces the key events and main story, followed by a general discussion with more details. In the *Martini Glass*, the article relies on narrative chronology beginning with a summary of the main event. The article starts with the inverted pyramid structure but then transitions into a narrative story following a chronological order. Lastly, in the *Narrative* structure, the article presents a chronological sequence of events with more details than usual news articles.

Each one of these structures has distinctive features that distinguish them from one another, in particular, writing style and presentation order [4]. While the last three structures are useful for some news, none enjoy the popularity of the inverted pyramid, the most common structure in written journalism. Furthermore, the inverted pyramid is considered a fundamental pillar of objectivity [16] as well as a key feature of professional journalistic news reporting as opposed to fake news [20]. Hence, for the purposes of this study, we focus on analyzing the inverted pyramid structure, instead of attempting to classify among all possible journalistic structures.

2.2 Main Event Descriptor Extraction

There are many methods to extract main event descriptors (i.e., the 5W1H answers). We present some approaches in this subsection.

Most works are purely based on rules that leverage lexical, syntactic and semantic information to obtain answer candidates. *Verb-based approaches* work by identifying the main action in a sentence or text, represented by a verb [17]. Once the method has identified the main verb, it extracts the arguments associated with it to find the main event descriptors (e.g., the subject of the verb). *Semantic role labeling* identifies semantic predicates at a sentence level [3]. Then, it identifies syntactic components through shallow parsing and assigns them a semantic role in the predicate. This method leverages syntactic relationships to identify text semantics. *Machine learning classifiers* can be trained and used to extract main event descriptors [23]. In particular, they can predict the arguments of the main predicate of a sentence (i.e., the 5W1H answers). However, due to the scarcity of annotated data sets [6] there is less work on methods that exploit annotations to improve their results.

Finally, we highlight the recent works of Hamborg et al. [6, 7] which present the development of an open-source system for 5W1H extraction (`Giveme5W1H`) along with a gold-standard data set. We have used their extraction and scoring methods as a guide for our own 5W1H extraction system.

2.3 News Summarization

News summarization is an extensively studied application of natural language processing. It comprises of two main approaches: abstractive and extractive. While abstractive methods rephrase and compress the original text to create the summary, extractive methods select key sentences from the text to build the summary [10].

For our work, we consider using an extractive summarization algorithm. In particular, we use TextRank, an algorithm that has been successfully used in multiple applications [2], is domain-independent and does not require deep linguistic knowledge [15].

Finally, news summarization is closely related to 5W1H extraction, since answers to the main event descriptor questions can be used to provide an explicit summary of the main event [7]. In essence, both summaries and main event descriptors are performing the same task: they distill the article into a simpler representation. In the context of the inverted pyramid structure, the results from both tasks should always be related to the beginning of the article.

3 DATA COLLECTION AND ANNOTATION

We start by describing the dataset used, present the subset employed for the IPS evaluation and describe the annotated sample used for evaluating the main event descriptor extraction (5W1H).

3.1 AP News Data

Our work is based on a proprietary data set from the Associated Press News (AP News) spanning a full year of news articles from December 2016 to December 2017, a total of 65,535 articles. The Associated Press was instrumental in creating the inverted pyramid structure [22] and continues to use it for reporting. Hence, the AP news data is ideal for testing our inverted pyramid scoring method.

Each news article in our dataset includes information about the topics of the news, referred to as subject tags or categories, spanning from serious topics like *Business* to lightweight topics, such as *Entertainment*. Following Bakshy et al's [1] hard-soft news classification scheme, the following AP News categories are likely to be hard news: *Science*, *Politics*, *Business*, *Health*, and *Weather*. Hard news is characterized by a high level of newsworthiness or news value and require timely publication (e.g., politics or business news). Whereas, soft news has a low level of informational value (e.g., entertainment news) and does not need immediate publication [11]. The AP data also contains a label indicating whether a news article is *Breaking* or *Non-Breaking* news. Breaking news articles usually follow the inverted pyramid structure and we expect them to have a higher IPS compared to non-breaking news articles. Thus, if our scoring method works, it should be able to score breaking news with significantly higher IPS value than non-breaking new articles.

What genre of news should we evaluate with our scoring method? We decided to test our scoring method on "hard breaking news" since the inverted pyramid is seen as the distinctive feature of hard news reports [21]. Furthermore, we are interested in applying IPS as a means of establishing journalistic standards and as a method to contrast it with fake news. Hence, focusing on hard news makes sense as they are often subjected to misinformation.

We first filtered articles with missing fields (e.g., countries, tags), non-English articles, and retained those that were tagged as "United States". We focused on a single month to reduce variability in the news articles and to minimize computational costs. In particular, we chose November 2017 as it is the most recent month with a high number of articles (5,045 articles, compared to December 2017 which only has 945 articles). For this first phase of our study, we focus on the two most frequent hard news categories in our data set: *Politics* and *Business*. Our final sample comprised 1,529 articles.

3.2 Annotations

One of the key steps in building our scoring method is extracting answers to the 5W1H main event descriptor questions from an article. How well does our 5W1H extraction work? To answer, we extracted a random sample of 30 breaking news from our data (15 political and 15 business articles) and obtained annotations for the 5W1H answers from experts trained in writing journalistic articles following the inverted pyramid. Specifically, we asked senior journalism students to assess our 5W1H answers who are extensively trained in using the inverted pyramid and writing 5W1H answers.

Journalism students were provided a questionnaire with a 3-point Likert scale to evaluate the descriptors. Additionally, if a descriptor was not present in the article, students could mark it as N/A. Figure 5 in the appendix shows an example question. We received six evaluations for each article, totaling 180 annotations for our sample of 30 articles. For each descriptor per article, we averaged the expert-assigned scores. Next, we averaged these results again over all articles to get the final evaluation for each question.

4 INVERTED PYRAMID SCORE

Here we present the Inverted Pyramid Score and its two main components: 1) main descriptor locations and 2) summary similarity. The IPS answers the following question: how well does a news article follow journalistic standards? And, in particular, how well does an article fit the inverted pyramid structure?

Figure 4 in the appendix presents a general overview of our method. To find the IPS, we first apply standard preprocessing steps on our dataset. Next, we compute a score for the **main descriptor locations** and the **summary similarity**. Both scores are computed with respect to the opening paragraphs comprising the headline, lead, and 2nd paragraph. Finally, we compute the final IPS as the weighted average of the two component scores. We assigned a higher weight to main descriptor locations since the 5W1H questions guide the writing of the inverted pyramid [5].

Data Preprocessing. We start by applying standard preprocessing techniques, such as tokenization, sentence splitting, part-of-speech tagging, dependency parsing and named entity recognition. Additionally, we use `neuralcoref` to handle coreference resolution. As an example of coreference resolution, consider Figure 1. "Hawaii group" is the answer to WHO, but this entity is also mentioned in the text as a "group representing Hawaii commercial fishermen." This is a coreference because they refer to the same entity.

4.1 Main Descriptor Locations

The first component of the IPS models how well the article captures the main event descriptors. In particular, we first check whether all the main event descriptors (i.e., 5W1H answers) are present in the article. Our 5W1H main event extraction comprises two steps: extracting all possible candidates for 5W1H and scoring candidates to find the best match. Next, we ensure that the main event descriptors appear early in the text. We assign a score based on the descriptor's location in the article, penalizing those appearing below the headline sentence.

4.1.1 Extracting main event descriptors. We build our main event descriptor extraction module by following the generic 5W1H architecture described by Hamborg et al., specifically their implementation of `Giveme5W1H`. We extend their implementation by including additional rules and refining the candidate scoring mechanism.

Extracting possible candidates for 5W1H

What & Who. We find all sentences of the form NP-VP-NP (e.g., *[The cat]*$_{\text{NP}}$ *[quickly climbed]*$_{\text{VP}}$ *[the apple tree]*$_{\text{NP}}$). Usually, the answer to WHO is contained in the first NP of these structures and the answer to WHAT is contained in the VP-NP part. We also exclude candidates that contain attribution verbs in the VP (e.g., *said, told*). Since attributions usually only offer supporting information, they are unlikely to contain the answer to WHAT and WHO.

When. To extract WHEN candidates, we parse regular dates and relative dates. We also check for dates that escaped automatic date parsing (e.g., "Christmas weekend"). We handle them by adding manual rules. Furthermore, we identify additional time noun phrases using a dictionary of time nouns [12].

Where. For WHERE candidates, we find all named entities that are tagged as location. We geolocate them using the OpenCage API.

Why. We search for three elements to extract WHY candidates: adverbs that express causal relationships, causal conjunctions, NP-VP-NP structures with causal verbs and auxiliary verbs that can be used for showing causes but aren't specific enough (e.g., "to be," as in "the airplane failure was a mechanical issue.").

How. We extract sentences that use one of the copulative conjunctions (usually the phrase after the conjunction is the HOW). We also find NP-VP-NP phrases that have adverbs and adjectives (since these modifiers can reflect the answer to HOW).

Candidate scoring to find the best match

After extracting the potential candidates, we score and rank them to get the final answer for each 5W1H. We designed all scores to be between 0 and 1. We score candidates based on a combination of the following criteria: position, type, frequency, precision, length, and other candidate-specific scoring criteria.

Position score. For all 5W1H questions, we assign a high score when candidates are found early in the text. For occurrences in the first sentence of the document (or headline), we assign a score of 1. For occurrences in subsequent sentences, the score follows an exponential decay, decreasing with an increase in position, p. Specifically, $S_{\text{p}}(C) = e^{(-dp)}$ with decay coefficient, $d > 0$. To illustrate, let's refer to Figure 1. Considering logarithmic decay, $d = \log(2)$, we divide the score by half whenever we move farther away from the headline. Thus, the WHO candidate ($p = 0$) will be scored with 1 and the WHEN candidate with 0.25 ($p = 2$).

Type score. Scoring based on candidate type, such as proper or common noun, date or time, etc., depends on the 5W1H question being answered. For WHO, it refers to whether the candidate is a named entity (i.e., a proper noun). For example, if the extracted candidate for WHO is a named entity, we score it as 1, otherwise it is scored as 0. For WHEN, *type* refers to whether the candidate is a proper date or a vague expression. For WHERE, it refers to the type of location (e.g., geopolitical entities, geographical locations, man-made structures, or organizations which can be used to refer to places in some cases). For WHY and HOW, we score candidates based on whether it is expressed through an NP-VP-NP pattern or conjunction or a combination of both.

Frequency Score. For all questions, except WHY and HOW, we rank-score candidates by their frequency of occurrence in the article. The highest frequency candidate is scored as 1. If the candidate is a named entity, we count all its coreferences, otherwise, we simply count the raw occurrences. For example, consider "Hawaii" and "United States" as WHERE candidates for the article in Figure 1. If we only consider the parts shown in Figure 1, then the article mentions the first candidate four times and the second candidate only once. We normalize the counts by the highest frequency and assign a score of 1 to "Hawaii" and 1/4 to "United States."

Precision and Length Score. For WHERE and WHEN, we consider the **Precision** of the candidate. For example, a date with an exact time is ranked higher than a vague phrase like "election time" and "London" is ranked higher than "UK" because it is a more precise location. For WHY and HOW, we consider the **Length** of the candidate. We prefer longer explanations for the cause and method. To implement this, we count the number of words in the candidate and divide by the maximum count in all candidates. Moreover, we add a redundancy penalty if the candidate repeats the answer to WHAT or if we get the same answer for WHY and HOW.

Other Scoring Criteria. For WHEN, we also score candidates by *distance to publication date*, preferring dates closer to the publication date. For WHERE, we score candidates by *clustering*. We assign a higher score if a candidate is close to the other candidates. For example, if most locations are in Germany, then we would assign less score to a random location in Japan. For HOW, we score candidates by *modifier frequency*, which counts the number of adverbs and adjectives used by the candidate.

4.1.2 Location Scoring of Main Event Descriptors. We assign the location scores for each main event descriptor using the following criteria: if an article follows an inverted pyramid structure, it should provide answers to the 5W1H questions in the OP (see Figure 1). Thus, if we find the answers there, we assign a high IPS. While the headline and lead are usually one sentence long each [8], the 2nd paragraph can have at most three sentences. We found this maximum length by analyzing breaking news articles in the data set. Hence, for the purposes of our estimation, we consider the OP to be the first 5 sentences of an article. We give a full score if all 5W1H descriptors are contained in the OP. Otherwise, we apply an exponential penalty by location of each descriptor. More formally, considering the headline index to be 0, for each descriptor D,

$$\text{LocScore}(\text{D}) = \begin{cases} 2^{4-\max(4, \text{Location}(\text{D}))} & \text{if answer found} \\ 0 & \text{if answer not found.} \end{cases}$$

Finally, we obtain a weighted average of all the location scores. Since HOW and WHY are not necessarily present, and even humans may have problems extracting them, we assign them a lower weight than the other descriptors.

4.2 Summarization

The second component of the IPS models how well an article is summarized by the OP. By definition, an article following an inverted pyramid structure must be summarizable by removing everything except the OP—the headline, lead, and 2nd paragraph. Note how in Figure 1 the OP contains all relevant information about the news story. Hence, our generated summary should be similar to the OP. Thus, we implement our summary similarity module by comparing the summary of the full article with the OP. First, we summarize the full article using an extractive summarization algorithm—TextRank. TextRank ranks an article by the most important sentences and then uses those to build the summary. Next, we compare the full article summary and the OP by comparing the language representations of the two. In particular, we do this using `Spacy` and their pre-trained `en_core_web_lg` model. This model uses GloVe vectors and it was trained with a multi-task CNN on blogs, news, and comments [9]. We average all the word vectors contained in a text to get its final representation. Finally, we compute the summarization score using the Cosine similarity distance between the vector representations of the OP and the summary.

5 RESULTS AND DISCUSSION

Here we show our main findings and discussions. We begin by presenting the evaluation of our main event descriptors extractor. Next, we report the results on the November 2017 AP News articles, showing the IPS distributions for breaking and non-breaking news.

5.1 5W1H Extraction

Table 1 shows the evaluation results of our 5W1H method. We find that our extractor is capable of obtaining the right answers for the basic 4W with 78% accuracy on average. Out of the four basic descriptors, our method systematically extracted better results for WHERE in this data set. This could be attributed in turn to the date-line being explicitly included in AP News articles.

However, for the full main event descriptors we only achieve 67% average accuracy. This reduction in accuracy makes sense considering the inherent difficulty of extracting the causes and methods from news articles. Even though the accuracy for WHY and HOW is still low compared to the other questions, our method is on par with the state-of-the-art.

As a baseline for comparison, `Giveme5W1H` gets 0.73 accuracy for all descriptors and 0.82 for the basic 4W on a BBC news data set [6]. However, it is hard to draw a direct comparison because of differences in the background of the annotators (journalism students vs IT students) and of data sets (AP News vs BBC).

Question	Business	Politics	*Total*
Who	0.74 ± 0.06	0.77 ± 0.07	0.76 ± 0.04
What	0.79 ± 0.05	0.73 ± 0.06	0.76 ± 0.04
When	0.71 ± 0.05	0.83 ± 0.05	0.77 ± 0.04
Where	0.87 ± 0.04	0.81 ± 0.04	0.84 ± 0.03
Why	0.42 ± 0.08	0.51 ± 0.06	0.46 ± 0.05
How	0.46 ± 0.07	0.42 ± 0.07	0.44 ± 0.05
Avg (Total)	0.66 ± 0.08	0.68 ± 0.07	0.67 ± 0.07
Avg (4W)	0.78 ± 0.04	0.78 ± 0.02	0.78 ± 0.02

Table 1: 5W1H evaluation results for the breaking news in the AP data set by subject category (± standard errors).

5.2 Inverted Pyramid Score

After testing the main event descriptor extractor on the previous sample we turn to the main task. Using our full data set, we compute the IPS of each article and show its distribution and basic statistics in Figure 2. In general, a higher IPS means that the articles adhere more to the inverted pyramid structure. Thus, these results match our intuition that breaking news usually follows the inverted pyramid structure. Non-breaking news shows more structural variety, as evidenced by their higher standard deviation and lower IPS.

While our IPS method gets intuitively correct results on breaking and non-breaking news, there might be other factors that affect whether a news article is written using an inverted pyramid structure or something else. In particular, an important element is the writing style used in the article. Inverted pyramid news will likely follow an expository writing style rather than a narrative writing style [4]. Consequently, we could add a new component to our scoring method that accounted for writing style differences.

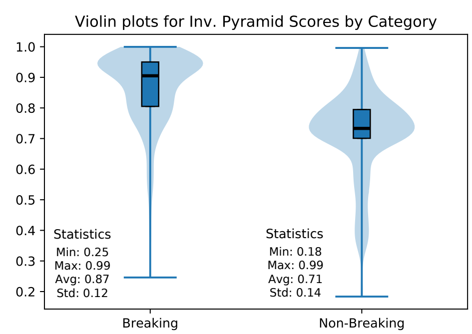

Figure 2: Plot of the IPS distribution and basic statistics for breaking and non-breaking news. On average, breaking news has a higher IPS than non-breaking news.

6 CONCLUSION

We have presented our work on evaluating the inverted pyramid structure using 5W1H extraction and summarization. Our analyses of results show that the method works well, allowing us to distinguish between breaking and non-breaking news articles.

In terms of improving our method, future work includes making improvements to the 5W1H extractor and using state-of-the-art summarization schemes tailored for news articles. In terms of potential applications, we plan on using this work to evaluate the different structures of news articles, not only restricting ourselves to the inverted pyramid. The current implementation could also be used, with additional features and descriptors, to provide a classification tool for breaking and non-breaking news.

Finally, our long-term goal is to evaluate how fake news sources structure their articles, as well as comparing them to mainstream outlets. By finding these structural differences we hope to elucidate how fake news articles differ from regular news.

ACKNOWLEDGMENTS

This work was partially funded by CONICYT PCFHA / DOCTORADO EXTRANJERO BECAS CHILE/2019 - 72200105.

REFERENCES

[1] Eytan Bakshy, Solomon Messing, and Lada A. Adamic. 2015. Exposure to ideologically diverse news and opinion on Facebook. *Science* 348, 6239 (2015), 1130–1132.

[2] Federico Barrios, Federico López, Luis Argerich, and Rosita Wachenchauzer. 2015. Variations of the Similarity Function of TextRank for Automated Summarization. In *Argentine Symposium on Artificial Intelligence (ASAI 2015) (Rosario, 2015)*.

[3] Kunal Chakma and Amitava Das. 2018. A 5w1h based annotation scheme for semantic role labeling of English tweets. *Computación y Sistemas* 22, 3 (2018), 747—-755.

[4] Zeyu Dai, Himanshu Taneja, and Ruihong Huang. 2018. Fine-grained Structure-based News Genre Categorization. In *Proc. of the Workshop Events and Stories in the News 2018*. Association for Computational Linguistics, Santa Fe, New Mexico, U.S.A, 61–67.

[5] Delia Gavriliu. 2012. From the Print Press to Online Press: Constraints and Liberties of the Journalistic Discourse. *Procedia - Social and Behavioral Sciences* 63 (2012), 263 – 270. The 4th Edition of the International Conf.: Paradigms of the Ideological Discourse 2012.

[6] Felix Hamborg, Corinna Breitinger, and Bela Gipp. 2019. Giveme5W1H: A Universal System for Extracting Main Events from News Articles. arXiv:cs.CL/1909.02766

[7] Felix Hamborg, Soeren Lachnit, Moritz Schubotz, Thomas Hepp, and Bela Gipp. 2018. Giveme5W: Main Event Retrieval from News Articles by Extraction of the Five Journalistic W Questions. In *Transforming Digital Worlds*. Springer International Publishing, Cham, 356–366.

[8] Tim Harrower. 2010. *Inside reporting*. Vol. 310. McGraw-Hill Education, 1221 Avenue of the Americas, New York, NY 10020.

[9] Matthew Honnibal and Ines Montani. 2017. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. (2017).

[10] Jagadish S Kallimani, KG Srinivasa, and B Eswara Reddy. 2012. Summarizing news paper articles: experiments with ontology-based, customized, extractive text summary and word scoring. *Cybernetics and Information Technologies* 12, 2 (2012), 34–50.

[11] Sam N. Lehman-Wilzig and Michal Seletzky. 2010. Hard news, soft news, 'general' news: The necessity and utility of an intermediate classification. *Journalism* 11, 1 (2010), 37–56.

[12] Michaela Mahlberg. 2005. *English general nouns: A corpus theoretical approach*. Vol. 20. John Benjamins Publishing.

[13] Audrey McAvoy. [n.d.]. Hawaii group wants to defend licenses for foreign fishermen. *Associated Press News* ([n. d.]). https://apnews.com/00962efef9e941ee8563cd24855c1fb8

[14] Melvin Mencher and Wendy P Shilton. 2011. *News reporting and writing*. Brown & Benchmark Publishers.

[15] Rada Mihalcea. 2004. Graph-based ranking algorithms for sentence extraction, applied to text summarization. In *Proc. of the ACL Interactive Poster and Demonstration Sessions*. 170–173.

[16] David TZ Mindich. 2000. *Just the facts: How "objectivity" came to define American journalism*. NYU Press.

[17] Kristen Parton, Kathleen R McKeown, Bob Coyne, Mona T Diab, Ralph Grishman, Dilek Hakkani-Tür, Mary Harper, Heng Ji, Wei Yun Ma, Adam Meyers, et al. 2009. Who, What, When, Where, Why? Comparing Multiple Approaches to the Cross-Lingual 5W Task. In *Proc. of the Joint Conf. of the 47th Annual Meeting of the ACL and the 4th International Joint Conf. on Natural Language Processing of the AFNLP*. Association for Computational Linguistics, Suntec, Singapore, 423–431.

[18] Horst Pöttker. 2003. News and its communicative quality: the inverted pyramid—when and why did it appear? *Journalism Studies* 4, 4 (2003), 501–511.

[19] Jaakko Salo. 2019. A Genre Analytical Comparison of Real and Fake News. (2019).

[20] Pamela J. Shoemaker. 2017. *News Values: Reciprocal Effects on Journalists and Journalism.* American Cancer Society, 1–9.

[21] Elizabeth A. Thomson, Peter R. R. White, and Philip Kitley. 2008. “Objectivity” and “Hard News” reporting across cultures. *Journalism Studies* 9, 2 (2008), 212–228.

[22] Patrick Walters. 2017. Beyond the inverted pyramid: Teaching the writing and all-formats coverage of planned and unplanned breaking news. *Teaching Journalism & Mass Communication* 7, 2 (2017), 9–22.

[23] Sibel Yaman, Dilek Hakkani-Tür, Gokhan Tur, Ralph Grishman, Mary Harper, Kathleen R McKeown, Adam Meyers, and Kartavya Sharma. 2009. Classification-based strategies for combining multiple 5-w question answering systems. In *Tenth Annual Conf. of the International Speech Communication Association.* INTERSPEECH-2009, 2703–2706.

[24] Hongxin Zhang and Haitao Liu. 2016. Visualizing structural “inverted pyramids” in English news discourse across levels. *Text & Talk* 36, 1 (2016), 89–110.

APPENDIX: FIGURES

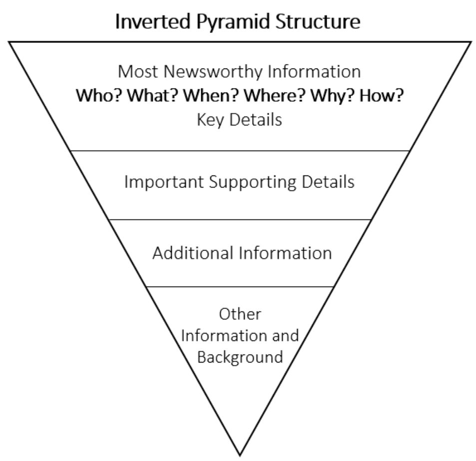

Figure 3: Diagram describing the inverted pyramid structure. The information is shown in descending order of importance, with the key details at the top.

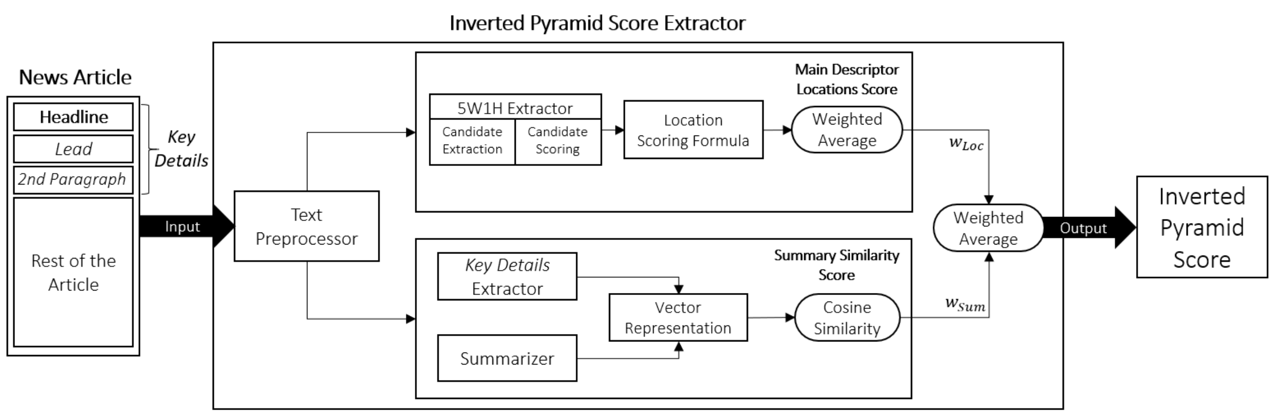

Figure 4: Diagram showing the computation of the Inverted Pyramid Score (IPS) of a news article. The first step is preprocessing, then we compute the *main event descriptor locations* score and the *summary similarity* score. For the main descriptor locations score we extract the position of the main event descriptors (5W1H answers), then we get the weighted average of these scores. For the summary similarity score, we get the similarity between an extractive summary of the article and the key details. We get the final IPS using the weighted average of the previous scores.

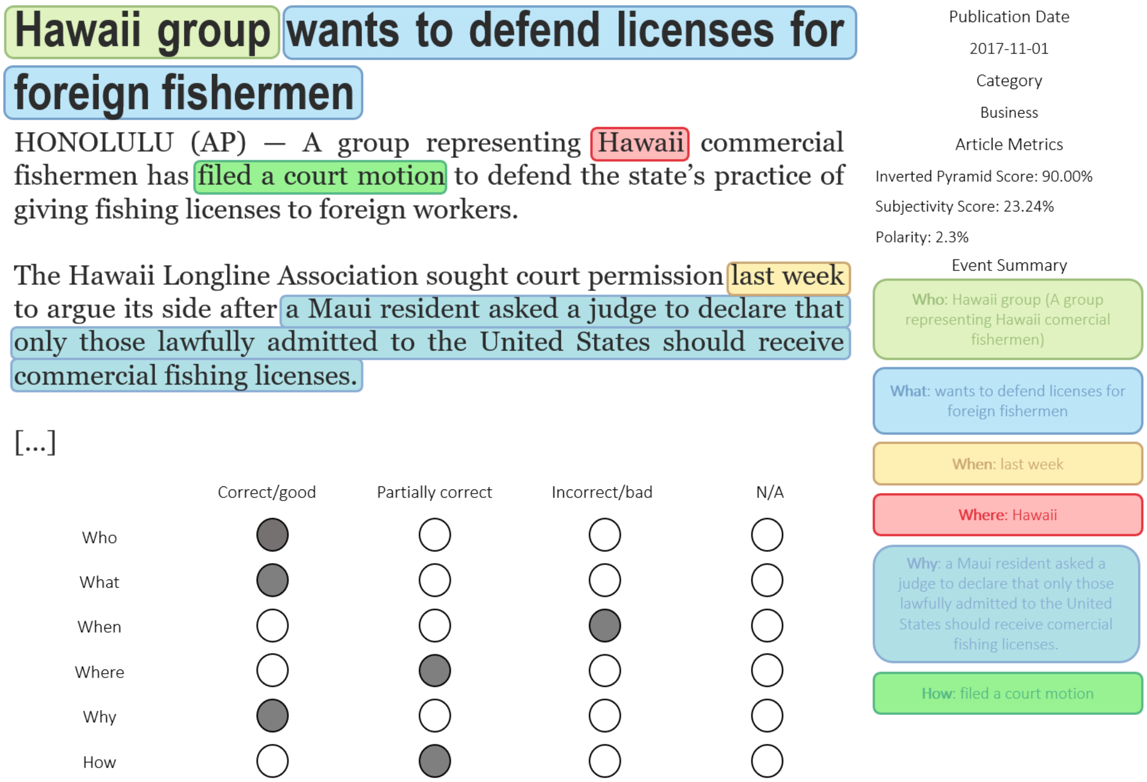

Figure 5: Example of a question from the evaluation questionnaire. The questionnaire showed the full article with highlighted answers, some additional information about the article on the top left, and the answers below that. The students had to select one answer for each 5W1H question after reading the article and the proposed answers.