Giveme5W: Main Event Retrieval from News Articles by Extraction of the Five Journalistic W Questions

Felix Hamborg[(✉)], Soeren Lachnit, Moritz Schubotz, Thomas Hepp, and Bela Gipp

University of Konstanz, Konstanz, Germany
{felix.hamborg,soeren.lachnit,moritz.schubotz,
thomas.hepp,bela.gipp}@uni-konstanz.de

Abstract. Extraction of event descriptors from news articles is a commonly required task for various tasks, such as clustering related articles, summarization, and news aggregation. Due to the lack of generally usable and publicly available methods optimized for news, many researchers must redundantly implement such methods for their project. Answers to the five journalistic W questions (5Ws) describe the main event of a news article, i.e., who did what, when, where, and why. The main contribution of this paper is Giveme5W, the first open-source, syntax-based 5W extraction system for news articles. The system retrieves an article's main event by extracting phrases that answer the journalistic 5Ws. In an evaluation with three assessors and 60 articles, we find that the extraction precision of 5W phrases is $p = 0.7$.

Keywords: News event detection · 5W extraction · 5W question answering

1 Introduction and Background

Extraction of a news article's main event is a fundamental analysis task required for a broad spectrum of use cases. For instance, news aggregators, such as Google News, must identify the main event to cluster related articles, i.e., articles reporting on the same event [5, 15]. News summarization extracts an article's main event to enable users to quickly see what multiple articles are reporting on [16, 25]. Other disciplines also analyze the events of articles, e.g., in so called frame analyses researchers from the social sciences identify how media reports on certain events [31].

Though *main event extraction from news* is a fundamental task in news analysis [16, 27], no method is publicly available that extracts *explicit descriptors* of the main event. We define explicit event descriptors as properties that occur in a text that is describing an event, e.g., text phrases in a news article that enable a news consumer to understand what the article is reporting on. Explicit descriptors could be used by various news analysis tasks, including all of the previously mentioned news analysis tasks, e.g., clustering, summarization, and frame analysis. State-of-the-art methods that extract events from articles suffer from three main shortcomings. Most approaches either (1) detect events only *implicitly* or are (2) *highly specialized* for the extraction of

task-specific event properties. Some approaches extract explicit event descriptors, but (3) are *not publicly available*.

Approaches of the first category detect events only implicitly, e.g., they find groups of textually similar articles by employing topic modeling or other clustering methods [32]. Some approaches afterward compute cluster labels that describe what is common to the group of related articles, typically the shared event or topic [2, 16, 27]. However, none of these approaches extract descriptors of a single article's main event to enable further analysis using these descriptors. The second category of approaches is highly specialized on task-specific event properties, such as the number of dead or injured people for crisis monitoring [32] or the number of protestors in demonstrations [26]. Approaches of the third category extract explicit event descriptors but are not publicly available [29, 34–36].

These shortcomings result in two disadvantages to the research community. First, researchers need to redundantly perform work for a task that can be well addressed with state-of-the-art techniques, due to the non-availability of suitable implementations. Second, non-optimal accuracy of produced results, since for many projects the extraction of explicit event descriptors is only a necessary task but not their actual contribution.

The main objective of our research is to devise an automated method that extracts the main event of a single news article. To address the three main shortcomings of state-of-the-art methods, our method needs to *extract explicit main event descriptors* that are *usable* by later tasks in the analysis workflow. The approach must also be *publicly available* and *reliably extract* the main event descriptors by exploiting the characteristics of news articles.

Journalists typically answer the five journalistic W-questions (5W), i.e., *who* did *what, when, where,* and *why*, within the first few sentences of an article to quickly inform readers of the main event. Figure 1 shows an excerpt of an article reporting on a terrorist attack in Afghanistan [1]. The highlighted phrases represent 5W main event properties. Due to their descriptiveness of the main event, we focus our research on the extraction of the journalistic 5Ws. Extraction of event-describing phrases also allows later analysis tasks to use common natural language processing (NLP) methods, such as

Taliban attacks German consulate in northern Afghan city of Mazar-i-Sharif with truck bomb

> *The death toll from a powerful Taliban truck bombing at the German consulate in Afghanistan's Mazar-i-Sharif city rose to at least six Friday, with more than 100 others wounded in a major militant assault.*

> The Taliban said the bombing late Thursday, which tore a massive crater in the road and overturned cars, was a "revenge attack" for US air strikes this month in the volatile province of Kunduz that left 32 civilians dead. […]

Fig. 1. News article [1] with title (bold), lead paragraph (italic), and first of remaining paragraphs. Highlighted phrases represent the 5W event properties (who did what, when, where, and why).

TF-IDF and cosine similarity including named entity recognition (NER) [12] to assess the similarity of two events.

Section 2 discusses 5W extraction methods that retrieve the main event from news articles. Section 3 presents Giveme5W, the first open-source 5W extraction system. The system achieves high extraction precision, is available under an Apache 2 license, and through its modular design can be efficiently tailored by other researchers to their needs. Section 4 describes our evaluation, and discusses the performance of Giveme5W with respect to related approaches. Section 5 discusses future work.

2 Extraction of Journalistic 5Ws from News Articles

This section gives a brief overview of 5W extraction methods in the news domain. The task is closely related to closed-domain question answering, which is why some authors call their approaches *5W question answering* (QA) systems. Systems for 5W QA on news texts typically perform three tasks to determine the article's main event: (1) *preprocessing*, (2) *phrase extraction*, and (3) *candidate scoring* [34, 35]. The input data to QA systems is usually text, such as a full article including headline, lead paragraph, and main text [30], or a single sentence, e.g., in news ticker format [36]. Other systems use automatic speech recognition (ASR) to convert broad casts into text [35]. The outcomes of the process are five phrases, one for each of the 5W, which together represent the main event of a given news text, as exemplarily highlighted in Fig. 1. The *preprocessing task (1)* performs sentence splitting, tokenizes them, and often applies further NLP methods, including part-of-speech (POS) tagging, coreference resolution [30], NER [12], parsing [24], or semantic role labeling (SRL) [8].

For the *phrase extraction task (2)* various strategies are available. Most systems use manually created linguistic rules to extract phrase candidates from the preprocessed text [21, 30, 35]. Noun phrases (NP) yield candidates for "who", while sibling verb phrases (VP) are candidates for "what" [30]. Other systems use NER to only retrieve phrases that contain named entities, e.g., a person or an organization [12]. Others approaches use SRL to identify the agent ("who") performing the action ("what") and location- and temporal information ("where" and "when") [36]. Determining the reason ("why") can even be difficult for humans because often the reason is only described implicitly, if at all [13]. The applied methods range from simple approaches, e.g., looking for explicit markers of causal relations [21], such as "because", to complex approaches, e.g., training machine learning (ML) methods on annotated corpora [4]. The clear majority of research has focused on explicit causal relations, while only few approaches address implicit causal relations, which also achieve lower precision than methods for explicit causes [6].

The *candidate scoring task (3)* estimates the best answer for each 5W question. The reviewed 5W QA systems provide only few details on their scoring. Typical heuristics include: shortness of a candidate, as longer candidates may contain too many irrelevant details [30], "who" candidates that contain an NE, and active speech [35]. More complex methods are discussed in various linguistic publications, and involve supervised ML [19, 36]. Yaman et al. use three independent subsystems to extract 5W

answers [36]. A trained SVM then decides which subsystem is "correct" using features, such as the agreement among subsystems, or the number of non-null answers per subsystem.

While the evaluations of the reviewed papers generally indicate sufficient quality to be usable for news event extraction, e.g., the system from [36] achieved $F_1 = 0.85$ on the Darpa corpus from 2009, they lack comparability for two reasons: (1) There is no gold standard for journalistic 5W QA on news; even worse, evaluation data sets of previous papers are no longer available publicly [29, 35, 36]. (2) Previous papers use different quality measures, such as precision and recall [11] or error rates [35].

3   Giveme5W: System Description

Giveme5W is an open-source main event retrieval system for news articles that addresses the objectives we defined in Sect. 1. The system extracts 5W phrases that describe the generally usable properties of news events, i.e., who did what, when, where, and why. This section describes the processing pipeline of Giveme5W as shown in Fig. 2. Giveme5W can be accessed by other software as a Python library and via a RESTful API. Due to its modularity, researchers can efficiently adapt or replace components, e.g., use a parser tailored to characteristics their data or adapt the scoring functions if their articles cover only a specific topic, such as finance.
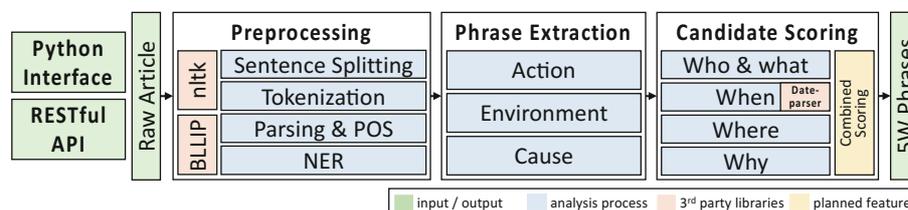

Fig. 2. Shown is the three-tasks analysis pipeline as it preprocesses a news text, finds candidate phrases for each of the 5W questions, and scores these.

3.1   Preprocessing of News Articles

Giveme5W can work with any combination of the following input types, where at least one must be provided: (1) *headline*, (2) *lead paragraph*, and (3) *main text*. If more than one type is given, Giveme5W appends them to one document, but keeps track of the individual types for later candidate scoring. Optionally, the article's publishing date can be provided, which helps Giveme5W to parse relative dates, such as "today at 12 am". Giveme5W integrates with the news crawler and extractor *news-please* [17].

During preprocessing, we use the Python NLP toolkit nltk [7] for sentence splitting, tokenization, and NER (with the trained seven-class model from Stanford NER [12]). For POS-tagging and full-text parsing we use the BLLIP parser [9]. To parse dates, we use parsedatetime [28]. For all libraries, we use the default settings for English.

3.2 Phrase Extraction

Giveme5W performs three independent extraction chains to extract the article's main event: (1) the *action* chain extracts phrases to the journalistic "who" and "what" questions, (2) *environment* for "when" and "where", and (3) *cause* for "why".

The *action extractor* identifies who did what in the article's main event, analyzing named entities (NE) and POS-tags. First, we look for any NE that was identified as a person or organization during preprocessing (cf. [12, 30]). We merge adjacent tokens of the same type within one NP to phrases (agent merge range $r_a = 1$ token), and add them to a list of "who"-candidates. We also add a sentence's first NP to the list if it contains any noun (NN*)[1] or personal pronoun (PRP) (cf. [30]). For each "who"-candidate, we take the VP that is the next right sibling in the parse tree as the corresponding "what"-candidate (cf. [7]).

The *environment extractor* identifies the temporal and local context of the event. Therefore, we look for NE classified as a location, date, time, or a combined datetime (cf. [36]). Similarly to "who"-candidates we merge tokens to phrases, using a temporal range $r_t = 2$ and locality range $r_l = 2$. This is necessary to handle phrases that do not purely consist of NE tokens, such as "Friday, 5th".

The *cause extractor* looks for linguistic features indicating a causal relation. The combined method consists of two subtasks, one analyzing POS-patterns, the other tokens. First, we recursively traverse the parse-tree to find the POS-pattern NP-VP-NP, where often the last NP is a cause [13]. We then check if a pattern contains an action verb, such as "allow" or "result", by using the list of verbs from [21]. If an action verb is used, the last NP of the POS-pattern from above is added to the list of cause candidates. The second subtask looks for cause indicating adverbs (RB) [3], such as "therefore", and causal conjunctional phrases [3], such as "because" or "consequence of".

3.3 Candidate Scoring

The last analysis task is to determine the best candidate of each 5W question. To score "*who*"-candidates we define three goals: the candidate shall occur in the article (1) *early* (following the inverse pyramid concept [10]) and (2) *often* (a frequently occurring candidate more likely refers to the main event), and (3) contain an *NE* (in news the actors involved in events are often NEs, e.g., politicians). The resulting scoring formula is $s_{who}(c) = w_0(d - p(c)) + w_1 f(c) + w_2 NE(c)$, where the weights $w_0 = w_1 = w_2 = 1$ (cf. [30, 35]), d the document length measured in sentences, $p(c)$ the position measured in sentences of candidate c within the document, $f(c)$ the frequency of phrases similar to c in the document, and $NE(c) = 1$ if c contains a NE, else 0 (cf. [12]).

To measure $f(c)$ we initially counted only exact matches, but we achieved better results with a simple *distance measure* for which we compute the normalized Levenshtein distance lev_{ij} between any candidate pair $c_i c_j$ of the same 5W question and increase the frequency of both c_i and c_j if $lev_{ij} < t_w$, where t_w is defined for each

---

[1] We use the POS-tag abbreviations from the Penn Treebank Project [33].

question w. We achieve the best results with $t_{\text{who}} = 0.5$. Due to the strong relation between agent and action, we rank the VPs according to the scores of their NPs. Hence, the most likely VP is the sibling in the parse tree of the most likely NP: $s_{\text{what}} = s_{\text{who}}$.

We score *temporal* candidates according to three goals: (1) occur *early* in the document, (2) *accuracy* (the more accurate, the better, i.e., instances including date and time are preferred over only date over only time), and (3) *parsable* to a datetime object [28]. Hence, $s_{\text{when}}(c) = w_0 \frac{d-p(c)}{d} + w_1 \text{DT}(c) + w_2 \text{TM}(c) + w_3 \text{TP}(c)$, where $w_0 = 10$, $w_1 = w_2 = 1$, $w_3 = 5$, $\text{DT}(c) = 1$ if c is a date instance, else 0, $\text{TM}(c) = 1$ if c is a time instance, 0.8 if c is a date instance, in which an adjacent time instance was merged, 0 else. $\text{TP}(c) = 1$ if c can be parsed into a datetime object, else 0.

The scoring of *location* candidates follows two simple goals: the candidate shall occur (1) *early* and (2) *often* in the document. $s_{\text{where}}(c) = w_0(d - p(c)) + w_1 f(c)$, where $w_0 = w_1 = 1$. The distance threshold to find similar candidates is $t_{\text{where}} = 0.6$. Section 4 describes how we plan to improve the location scoring.

Scoring *causal* candidates turned out to be challenging, since it often requires semantic interpretation of the text and simple heuristic may fail [13]. We define two objectives: (1) occur *early* in the document, and (2) the *causal type*. $s_{\text{why}}(c) = w_0 \frac{d-p(c)}{d} + w_1 \text{CT}(c)$, where $w_0 = w_1 = 1$, and $\text{TC}(c) = 1$ if c is a bi-clausal phrase, 0.6 if it starts with a causal RB, and 0.3 else (cf. [21, 22]).

3.4 Output

The highlighted phrases in Fig. 1 are the highest scored candidates extracted by Giveme5W for each of the 5W event properties of the sample article. If requested by the user, Giveme5W enriches the returned phrases with additional information that the system needed to extract for its own analysis. The additional information types for each token are its POS-tags, syntactical role within the sentence, which was extracted using parsing, and NE type if applicable. Enriching the tokens with this information increases the efficiency of the overall analysis workflow in which Giveme5W may be embedded since later analysis tasks can reuse the information.

Giveme5W also enriches "when"-phrases by attempting to parse them into datetime objects. For instance, Giveme5W resolves the "when"-phrase "late Thursday" from Fig. 1 by checking it against the article's publishing date, Friday, November 11, 2016. The resulting datetime object represents 18:00 on November 10, 2016.

4   Evaluation and Discussion

We performed a survey with three assessors (graduate IT students). We created an evaluation dataset by randomly sampling 60 articles (12 for each category) from the BBC corpus described in [14]. Instructions to recreate the dataset are available in the project's repository (see Sect. 5). The BBC corpus consists of 2,225 articles in the categories business (*Bus*), entertainment (*Ent*), politics (*Pol*), sport (*Spo*), and tech (*Tec*).

We presented all articles (one at a time) to each participant. After reading an article, we showed them Giveme5W's answers. We asked them to judge the relevance of each

answer on a 3-point scale: *non-relevant* (if an answer contains no relevant information, score $s = 0$), *partially relevant* (if part of the answer is relevant or information is missing, $s = 0.5$), and *relevant* (if the answer is completely relevant without missing information, $s = 1$).

Table 1 shows the *mean average generalized precision* (MAgP), a precision score suitable in multi-graded relevance assessments [20]. The MAgP over all categories and questions was 0.7. Excluding the "why"-question, which also the assessors most often disagreed on (discussed later and in Sect. 5), the overall MAgP was 0.76.

Table 1. ICR and generalized precision of Giveme5W.

Property	ICR	Bus	Ent	Pol	Spo	Tec	Avg.
Who	.87	.92	.94	.84	.74	.74	.87
What	.90	.90	.91	.81	.87	.70	.84
When	.76	.77	.77	.65	.51	.87	.72
Where	.75	.73	.52	.76	.56	.52	.62
Why	.63	.48	.62	.42	.44	.34	.46
Avg.	.78	.76	.75	.70	.66	.63	**.70**

Compared to the fraction of "correct" answers by the best system in [29], Giveme5W achieves a 0.05 higher MAgP. The best system in [36] achieves a precision of 0.89, which is 0.19 higher than our MAgP and surprisingly even better than the ICR of our assessors. However, comparing the performance of Giveme5W with other systems is not straightforward for several reasons: other systems were tested on non-disclosed datasets [29, 35, 36], were translated from other languages [29], or used different evaluation measures, such as error rates [35] or binary relevance assessments [36], which are both not optimal because of the non-binary relevance of 5W answers (cf. [20]). Finally, none of the related systems have been made publicly available, which was the primary motivation for our research as described in Sect. 1. For this reason, comparing the evaluation results of our system and related work was not possible.

Using the *intercoder reliability* (ICR) as a very rough approximation of the best possible precision that could be achieved (cf. [18]), we conclude that Giveme5W comes very close to the current optimum ($ICR = 0.78$, $MAgP = 0.7$).

We found that different forms of journalistic presentation in the five news categories led to different QA performance. Business and entertainment articles, which yielded the best performance, mostly reported on single events, while the sports and tech articles, on which our system performed slightly weaker, contained more non-event coverage, e.g., background reports or announcements.

Before we conducted the survey, we conducted a pre-survey to verify sufficient agreement among the assessors. We let the assessors rate ten articles and measured the overall ICR of the assessors' ratings using the average pairwise percentage agreement. We also let users fill in a questionnaire, asking how they understood the rating task. The pre-survey yielded an $ICR_{pre} = 0.65$. We found that some questions, specifically the "why"-question, required further explanation so that we added examples and clarified the assessment rules in the tutorial section of our survey application.

The ICR was 0.78 in the final survey, which is sufficiently high to accept the assessments (cf. [23]). While assessors often agreed on "who" and "what", they agreed less often on "when" and "where" (see Table 1). Similarly to Parton et al. [29], we found that lower ICR for "when" and "where" were caused by erroneous extractions of the "who" and "what" question, which in turn also yielded wrong answers for the remaining questions. "Why" had the lowest ICR, which is primarily because most articles do not contain explicit causal statements reasoning the event (see also Sect. 5). This increases the likelihood that assessors inferred different causes or none, and hence rated Giveme5W's answers discrepantly (see Sect. 5).

5   Future Work

We plan to investigate three ideas, from which all 5W-questions may benefit: (1) *coreference resolution* and (2) *semantic distance measure*, which will both allow Giveme5W to better assess the main agent (including the main action), and potentially also the cause. We plan to use WordNet or Wikidata to measure how two candidates are semantically related, and we will replace the currently used Levenshtein distance, which cannot handle synonyms. (3) Introduce *combined scoring* (see Fig. 2), which uses features of other Ws to score one W. For instance, if the top candidates for "who" and "what" are located at the beginning of the article, "when" and "where" candidates that are likewise at the beginning should receive a higher rating than others further down in the article. In our dataset, we found that this idea would particularly improve the performance of "where" and "why".

We also plan to improve the individual 5W extractors and scorers. For "where"-extraction we will replace the current accuracy estimation with a method that uses reverse geocoding, and prefer locations, e.g., a restaurant, over small regions, e.g., San Francisco, over larger regions, e.g., California, since the former are more accurate. The poor performance and rather low ICR of "why" require further investigation, especially when compared to evaluations of other systems, which have higher ICR and better performance. Some evaluations are biased, e.g., the dataset used in [36] was specifically designed for 5W QA. Such datasets may contain more explicit causal phrases than our randomly sampled articles that often only implicitly describe the cause. We plan to use the sophisticated list of rules suggested in [22] to further improve our cause extraction. We also plan to add an extractor for "how"-phrases (cf. [30, 34]).

Finally, we think that the creation of a gold standard dataset containing articles with manually annotated 5W phrases will help to advance research on main event retrieval from articles.

6   Conclusion

The main contribution of this paper is the first open-source system for retrieving the main event from news articles. The system, coined *Giveme5W*, extracts phrases answering the five journalistic W-questions (5W), i.e., *who* did *what, when, where,* and *why*. Giveme5W also enriches the phrases with POS-tags, named entity types, and

parsing information. The system uses syntactic and domain-specific rules to extract and score phrase candidates for each 5W question. In a pilot evaluation, Giveme5W achieved an overall, mean average generalized precision of 0.70, with the extraction of "who" and "what" performing best. "Where" and "why" performed more poorly, which was likely due to our use of real-world news articles, which often only imply the causes. We plan to use coreference resolution and a semantic distance measure to improve our extraction performance. Since answering the 5W questions is at the core of any news article, this task is being analyzed using different approaches by many projects and fields of research. We hope that redundantly performed work can be avoided in the future with Giveme5W as the first open-source and freely available 5W extraction system.

The code of Giveme5W and the evaluation dataset used in this paper are available under an Apache 2 license at: https://github.com/fhamborg/Giveme5W.

References

1. Agence France-Presse: Taliban attacks German consulate in Northern Afghan city of Mazar-i-Sharif with truck bomb. The Telegraph (2016)
2. Allan, J., et al.: 1998 Topic detection and tracking pilot study: final report. In: Proceedings of the DARPA Broadcast News Transcription and Understanding Workshop, pp. 194–218 (1998)
3. Altenberg, B.: Causal linking in spoken and written English. Studia Linguistica **38**(1), 20–69 (1984)
4. Asghar, N.: Automatic extraction of causal relations from natural language texts: a comprehensive survey. arXiv preprint arXiv:1605.07895 (2016)
5. Best, C., et al.: Europe media monitor (2005)
6. Bethard, S., Martin, J.H.: Learning semantic links from a corpus of parallel temporal and causal relations. In: Proceedings of the 46th Annual Meeting of the ACL on Human Language Technologies, pp. 177–180 (2008)
7. Bird, S., et al.: Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. O'Reilly Media, Inc., Sebastopol (2009)
8. Carreras, X., Màrquez, L.: Introduction to the CoNLL-2005 shared task: semantic role labeling. In: Proceedings of the Ninth Conference on Computational Natural Language, pp. 152–164 (2005)
9. Charniak, E., Johnson, M.: Coarse-to-fine n-best parsing and MaxEnt discriminative reranking. In: Proceedings of the 43rd Annual Meeting on ACL, pp. 173–180 (2005)
10. Christian, D., et al.: The Associated Press Stylebook and Briefing on Media Law. Associated Press, New York (2014)
11. Das, A., Bandyaopadhyay, S., Gambäck, B.: The 5W structure for sentiment summarization-visualization-tracking. In: Gelbukh, A. (ed.) CICLing 2012. LNCS, vol. 7181, pp. 540–555. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-28604-9_44
12. Finkel, J.R., et al.: Incorporating non-local information into information extraction systems by gibbs sampling. In: Proceedings of the 43rd Annual Meeting on ACL, pp. 363–370 (2005)
13. Girju, R.: Automatic detection of causal relations for question answering. In: Proceedings of the ACL 2003 Workshop on Multilingual Summarization and Question Answering, vol. 12, pp. 76–83 (2003)

14. Greene, D., Cunningham, P.: Practical solutions to the problem of diagonal dominance in kernel document clustering. In: Proceedings of the 23rd International Conference on Machine Learning, pp. 377–384 (2006)

15. Hamborg, F., et al.: Identification and analysis of media bias in news articles. In: Proceedings of the 15th International Symposium of Information Science (2017)

16. Hamborg, F., et al.: Matrix-based news aggregation: exploring different news perspectives. In: Proceedings of the ACM/IEEE Joint Conference on Digital Libraries, p. 10 (2017)

17. Hamborg, F., et al.: news-please: A generic news crawler and extractor. In: Proceedings of the 15th International Symposium of Information Science, pp. 218–223 (2017)

18. Hripcsak, G., Rothschild, A.S.: Agreement, the F-measure, and reliability in information retrieval. J. Am. Med. Inform. Assoc. **12**(3), 296–298 (2005)

19. Jurafsky, D.: Speech and Language Processing. Pearson Education India, New Delhi (2000)

20. Kekäläinen, J., Järvelin, K.: Using graded relevance assessments in IR evaluation. J. Am. Soc. Inform. Sci. Technol. **53**(13), 1120–1129 (2002)

21. Khoo, C.S.G., et al.: Automatic extraction of cause-effect information from newspaper text without knowledge-based inferencing. Lit. Linguist. Comput. **13**(4), 177–186 (1998)

22. Khoo, C.S.G.: Automatic identification of causal relations in text and their use for improving precision in information retrieval (1995)

23. Landis, J.R., Koch, G.G.: The measurement of observer agreement for categorical data. Biometrics **33**(1), 159–174 (1977)

24. Manning, C.D., et al.: Foundations of Statistical Natural Language Processing. MIT Press, Cambridge (1999)

25. McKeown, K.R., et al.: Tracking and summarizing news on a daily basis with Columbia's Newsblaster. In: Proceedings of the 2nd International Conference on Human Language Technology Research, pp. 280–285 (2002)

26. Oliver, P.E., Maney, G.M.: Political processes and local newspaper coverage of protest events: from selection bias to triadic interactions. Am. J. Sociol. **106**(2), 463–505 (2000)

27. Park, S., et al. NewsCube: delivering multiple aspects of news to mitigate media bias. In: Proceedings of SIGCHI 2009 Conference on Human Factors in Computing Systems, pp. 443–453 (2009)

28. parsedatetime - Parse human-readable date/time strings. https://github.com/bear/parsedatetime. Accessed 21 Aug 2017

29. Parton, K., et al.: Who, what, when, where, why?: comparing multiple approaches to the cross-lingual 5W task. In: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, vol. 1, pp. 423–431 (2009)

30. Sharma, S., et al.: News event extraction using 5W1H approach & its analysis. Int. J. Sci. Eng. Res. – IJSER **4**(5), 2064–2067 (2013)

31. Stemler, S.: An overview of content analysis. Pract. Assess. Res. Eval. **7**(17), 137–146 (2001)

32. Tanev, H., Piskorski, J., Atkinson, M.: Real-time news event extraction for global crisis monitoring. In: Kapetanios, E., Sugumaran, V., Spiliopoulou, M. (eds.) NLDB 2008. LNCS, vol. 5039, pp. 207–218. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-69858-6_21

33. Taylor, A., et al.: The Penn treebank: an overview. In: Abeillé, A. (ed.) Treebanks. TLTB, vol. 20, pp. 5–22. Springer, Dordrecht (2003). https://doi.org/10.1007/978-94-010-0201-1_1

34. Wang, W., et al.: Chinese news event 5W1H elements extraction using semantic role labeling. In: 2010 Third International Symposium on Information Processing (ISIP), pp. 484–489 (2010)
35. Yaman, S., et al.: Classification-based strategies for combining multiple 5-W question answering systems. In: INTERSPEECH, pp. 2703–2706 (2009)
36. Yaman, S., et al.: Combining semantic and syntactic information sources for 5-W question answering. In: INTERSPEECH, pp. 2707–2710 (2009)