Where's the Why? In Search of Chains of Causes for Query Events

Suchana Datta[1], Derek Greene[1], Debasis Ganguly[2], Dwaipayan Roy[3], and Mandar Mitra[4]

[1] School of Computer Science, University College Dublin
[2] IBM Research Europe, Dublin
[3] Indian Institute of Science, Education and Research, Kolkata
[4] Indian Statistical Institute, Kolkata
suchana.datta@ucdconnect.ie, derek.greene@ucd.ie,
debasis.ganguly1@ie.ibm.com, dwaipayan.roy@gmail.com, mandar@isical.ac.in

Abstract. Traditional information retrieval systems are primarily focused on finding topically-relevant documents, which are descriptive of a particular query concept. However, when working with sources such as collections of news articles, a user might often want to identify not only those documents which describe a news event, but also documents which explain the chain of events which potentially led to that event occurring. These associations might be complex, involving a number of causal factors. Motivated by this information need, we formulate the task of *causal information retrieval*. We provide a literature survey on causality-related research, and explain how the proposed task differs from standard retrieval problems. We then empirically investigate the ability of popular retrieval methods to successfully retrieve causally-relevant documents. Our results demonstrate that the performance of traditional methods are not upto the mark for this task, and that causal information retrieval remains an open challenge which is worthy of further research.

1 Introduction

Faced with any situation or event, it is a fundamental part of human nature to ask 'why?' and 'how?', as we attempt to understand the context in which we find ourselves. The same can be said when we seek to analyze any complex nature of events in modern society. As a concrete example, we may want to understand 'why was the UAE-Israel peace accord signed?' so that we can analyze its after-effects. Consequently, we often try to map events in the form of *cause-effect* relations. Over the years, the study of cause-effect relations has focused on uncovering the inter-relationships among different phenomena in terms of cause and effect [3]. Sometimes these associations are immediately evident to us, such as smoking *causes* lung cancer. However, these associations often can be rather complex, involving a combination of a number of causal factors that might have led to an observed event, together with a number of further precursory components that might have triggered events present in these causal factors in a recursive fashion. In the example above, the instant causal factors might include Israel's settlement plan or Trump's diplomatic strategy [4]. However, if we look further for foregoing causes of *Israel's settlement plan*, certain factors such as, acquiring global recognition, improving relations with middle east etc. might be notable. Literature emphasizes that in most situations there will be no definitive rules around how cause-effect relations should be structured [15,30]. It is rather difficult to explicitly enlist a list of causes (in the form of short text segments) for these complex cause-effect relationships. Rather, these causal factors are spread across a number of multi-topical documents. In that sense, perhaps it is better to present this information to a user leaving him the task of subjectively figuring out the potential causes.

Traditional search systems concentrate on matching terms between documents and a user query. However, this might not cover the situation where a user's search is intended to reveal the causes which led to specific event. In this paper, we investigate this gap in the information retrieval literature, by addressing two associated research questions:

- **RQ-1:** Is a new research paradigm required to address the requirements of identifying causally-relevant information (i.e., *causal information retrieval*)?
- **RQ-2:** Is a traditional search system adequate for the requirements?

To address these key questions, in Section 2 we provide a detailed literature survey on the causality research to date. In Section 3 we explore the emergence and the challenges of *causal information retrieval* task and conduct few experiments to investigate whether or not these models can meet the requirements of that task. We conclude in Section 4 with suggestions for further research in this area.

2 Literature Review

Identifying the inherent nature of cause-effect relations from text has been explored in multiple ways, although largely in the context of textual entailment [6]. However, we are interested in capturing document-level causal information, rather than working at the sentence level. In this section, we provide a high level overview of various existing approaches designed to capture cause-effect associations, which will help us to frame the problem of causal information retrieval.

Causal Relation Extraction. With the increasing popularity of deep neural architectures, the study of causation is now more based around counterfactuals (i.e., what might have happened?). But initially causality was more closely related to identifying semantic relations between a cause and an effect [30,33]. While sentence-level entailment has been harnessed to capture causal characteristics [17], other authors have investigated causal relations between two queries [32] which eventually has lead to the idea of using event pairs [5]. Later, the authors in [9] attempted to establish causality within texts by predicting event causality, i.e causality between event pairs, (e.g. 'police arrested him' *because* 'he killed someone'). Nonetheless, these approaches are concerned with sentential cause-effect relation extraction, whereas we investigate on causality spanned across a document collection for a given query.

Graph-based approaches. Graphs provide a convenient way to visualize cause-effect relations. While authors in [25] proposed a non-parametric graph-based framework to trace causal inferences, other works [8,24] used Directed Acyclic Graph (DAG) to represent causal relations and later focus shifted to Bayesian Network [36]. On the other hand, the authors in [31] focused on solving event-pair causality relations, encoded in text (e.g. we *'recognized'* the problem and *'took'* care of it), with graphs. Thus graph pattern based techniques primarily focus on identification of event pairs from text and study their patterns with probabilistic measures. However, for our task, selecting candidate events from a larger set of events that are likely to be related to the query event is the primary challenge, as causal events might not hold any direct relation with the query.

Causal Knowledge Bases. Research on causality that made use of domain-independent knowledge was first introduced in the late 1990s and continues today. As knowledge-based causality developed gradually, researchers attempted to explore automatic causal relation acquisition (specifically common cause-effect propositions) [18] and exploit semantic property of predicates [16] which efficiently find contradictory pairs (e.g. 'destroy cancer' $\perp$ 'develop cancer'). The knowledge-base pattern approach was extended in [37], where a set of patterns was initially used to create a network of causes and effects, leading to a relational embedding method.

Document Classification. Causality has also been shown to be relevant in document classification, where the relationship between features and classes is often complex. Paul [23] sought to answer the question of 'which term features *cause* documents to have the class labels that they do?', and developed a *propensity score matching* technique for selecting important features. The work in [34] considered the causal inference task as a classification problem, and using logistic regression, they illustrated how to analyze causality a variety of datasets. The authors took into account factors such as missing data and measurement errors, which often hinders downstream causal analysis.

Future Scenario Generation/Prediction. Contingency discourse problems in NLP, specifically new event prediction, consider causal relation extraction from text data as being particularly challenging [27]. The authors in [26] initiated this research with the automatic compilation and generalization of a sequence of events from different web corpora. However, other researchers argue that in order to address causality, either two of the events in the consecutive sentences must hold an inter-sentential contingent relation [29] or there should be a pre-trained event-causality chaining database generated from web data [14]. Therefore, future scenario prediction problems require prior event knowledge, which is unlikely in our case as users may have no prior knowledge about the plausible causes of a query event.

Question-Answering. The NLP literature highlights that question-answering (QA) systems exploit the inherent nature of causality by disambiguating the pervasive nature of causal relations [10] which aids to identify *inter* and *intra*-sentential causal links between terms and clauses to answer 'why' questions [22]. Lately, a decision support system [19] was proposed to foresee the consequences of queries like, '*Should I join the military*?' or '*Should I move to California*?'. another group of researchers focused on a new variant of QA, referred to as common sense causality identification [12,11]. This causality variant helped to disambiguate discourse relations and reasoning with sentence proximity by making use of knowledge-bases. Thus, QA approaches involve either lexical or syntactic patterns generation; or morphological features extraction between cause and effect. Therefore, this does not fit into tasks where causal documents are unlikely to have any definite pattern with the query event.

Deep Causal Relations. Since 2018, causality has been incorporated in classical CNN models [21], and has also been used to furnish a general abstraction over deep unsupervised learning methods [28]. Work in [13] focused on the salient concepts extracted from a target CNN network, which further helped to estimate the information captured by activations in the target network. Conversely, the authors in [20] propose the use of knowledge-based CNN to identify causal relations from natural language text.

3 Causal Information Retrieval

The techniques described in Section 2 consider causal relations either at the sentence level or within a single document. In certain cases, these methods involve prior knowledge about causal events, while in other cases they require some predefined lexical, syntactic or morphological relations. However, these techniques do not cover more nuanced causes and effects in larger document collections, such as those we hope to capture with retrieval models.To address the research questions introduced in Section 1, we propose a theoretical model of causality from an IR perspective. We propose an associated workflow, and we then investigate to what extent the requirements of causal search diverge from those of topical search. We do this by analyzing the performance of different standard retrieval models on a benchmark dataset with causal annotations.

3.1 Why do we need a Causal Retrieval Model?

In practice, information retrieval tasks are addressed by making use of term overlaps between a query and documents, where the notion of relevance varies depending on the task specifications. As an example of this, consider the query 'American military officers at Abu Ghraib prison accused', and a set of sample top-ranked document excerpts for this query (see Table 1). Now, if the task is to retrieve documents that are related to the topic itself, then any document highlighting an accusation against US military officers, offensive treatment towards detainees, leaked pictures of their torture, steps taken by US government

	Query - Accused American military officers in Abu Ghraib prison
Topical **RelDoc: 1**	The US is investigating a series of allegations of abuse, including sexual humiliation, of prisoners by the US military in Iraqs Abu Ghraib jail... The first American military intelligence soldier to be court-martialled over the Abu Ghraib abuse scandal was sentenced today to eight months in jail...
RelDoc: 2	The torture in Abu Ghraib prison reflects the breakdown in the chain of command in the US military... ...abuse is everywhere routine. One cornerstone of this new US policy seems to be to outsource the task of interrogating....where torture is routine like Syria or Egypt...
Causal **RelDoc: 1**	....a female US soldier dragging an Iraqi detainee on the prison floor like a dog on a leash, one end of which is shown tied to the mans neck... ....one detainee handcuffed to a bunk bed in Baghdads Abu Ghraib prison, his arms pulled so wide apart that his back is arched...
RelDoc: 2	....they were savagely beaten and repeatedly humiliated by American soldiers working on the night shift at Tier 1A in Abu Ghraib during the holy month of Ramazan,.... ...they were pressed to denounce Islam or were force-fed pork and liquor...They forced us to walk like dogs on our hands and knees...hitting us hard on our face and chest...

Table 1: Document excerpts taken from the FIRE collection [1], for a query seeking information on accusations related to Abu Ghraib prison.

etc. is considered as relevant. As such, four of the documents in Table 1 might be deemed relevant and retrieval using term overlap suffices the task. On the other hand, if the task shifts to identifying causally-relevant documents recursively (i.e. $query_{event} \leftarrow cause_{event} \leftarrow cause_{event} \leftarrow ...$) for the same query, the notion of relevance would now be concentrated on *'why US military officers are accused'* and the chain of further precursory causal events. In that case, reports on officers' torture stories, detainees statements accusing officers, evidence published on newspapers etc. are likely to meet the requirements of the task at this level (say, $level_i$) and for next level onward (i.e., $level_{i+1}$), we would be finding further prevalent causes given the effect event at $level_i$. Thus, only two of the documents in Table 1, labelled as 'causal', appear to be causally relevant to the aforementioned query. Now the question arises if term overlap between query and documents is adequate to meet up with this current task specifications or it requires different ideologies which we investigate in the later part of this paper.

Moreover, events that are eventually reported by news media are often triggered by a series of causes spread over an extended period of time. Consequently, making the initial query more specific by adding cause-related keywords, such as 'American military officers accusation causes (or reasons)' etc., and then using a traditional IR system is unlikely to retrieve relevant information, since details regarding the causes of the event might not be explicitly reported in news articles. However, such causality-specific information could be discovered by analyzing a number of documents and associating the latent relationships between their terms, along with the series of triggering causes.

3.2 Model Architecture

For a causal retrieval model, we assume the user is searching for cause-related information and there exists some agent or system to assist the user. Given a query event $\mathcal{Q} = \{q_0, q_1, .., q_n\}$, the user seeks documents containing causal information related to the query, and the search is performed over a fixed document collection C. The causal retrieval model aims to present causally-connected information in a recursive fashion. That is, given an event, it finds possible causes for that even, and given those causes (i.e. additional events), the system then finds what might have caused those successively (see Figure 1). Here each succession represents one level. We now formally describe this process.

We assume that in a n term query $\mathcal{Q}$, a small text snippet (i.e. sequence of terms) would be considered as the potential causal query (i.e. effect event) which we refer as initial query event Q. Therefore, Q can be represented as the 0^{th} event at level-0 (i.e. no retrieval is performed yet), which we denote as $D^0_{(0,0)}$. At the next level (i.e. level-1), given the query $D^0_{(0,0)}$, the system displays a set of top ranked k documents to the user, denoted $\mathcal{D}^{(1)} = \{D_1, D_2, ..., D_k\}$. Here each document D_j can be further fragmented into short text segments that might be a potential event having preceding causes. Thus, we constitute $D_i = \{D^1_{(j,1)}, D^1_{(j,2)}, ..., D^1_{(j,i)}, ..., D^1_{(j,n(D^1_j))}\}$, where $D^1_{(j,i)}$ denotes the i^{th} event identified at level-1 from the document retrieved at j^{th} rank. Assume that at level-1, the text segment $D^1_{(j,i)}$ is recognized as a potential event which has precursory chain of causes. Consequently, $D^1_{(j,i)}$ will act as query at level-1 and retrieve a further set of k causally-relevant documents, which will be treated as level-2. In this situation, the effect query event $D^1_{(j,i)}$ could be displayed to the user as hyperlink, which could expand to another new set of ranked documents once it is clicked by the user. As shown in Figure 1, the candidate effect event $D^1_{(j,i)}$ is considered as root of the sub-tree and it further expands to an immediate level with a new

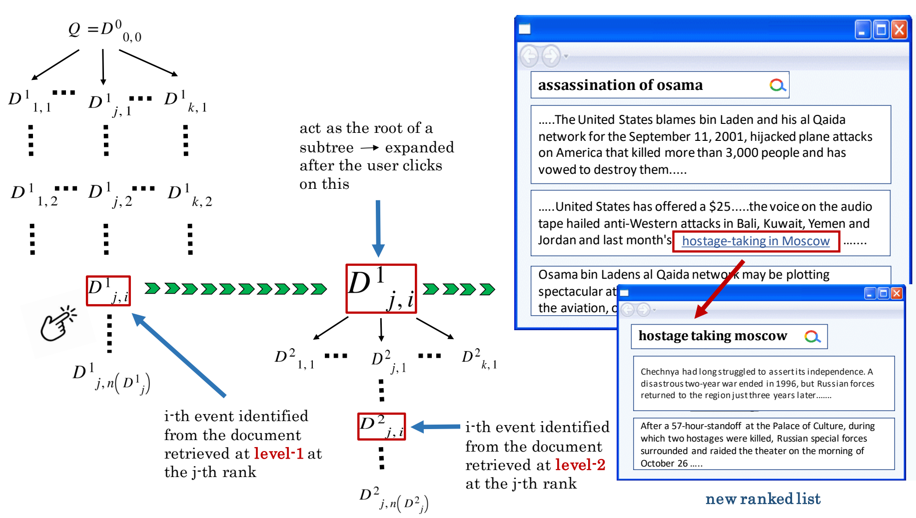
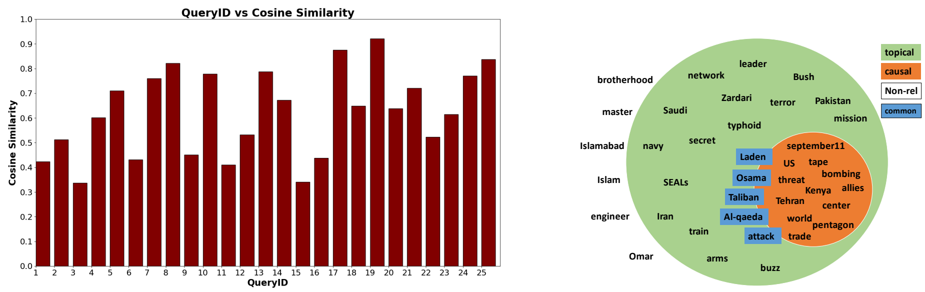


Fig. 1: Workflow of a user's experience in an interactive causality search interface.



(a) Cosine similarities between topical and causal documents.

(b) Term associations related to bin Laden's assassination.

Fig. 2: Per-query topical-causal relations, in terms of relevant documents, and terms in the top ranked-documents.

ranked list of documents $\mathcal{D}^{(2)} = \{D_1, D_2, ..., D_k\}$. Thus, we again fragment each document D_j into short text segments, and identify potential effect event for the next level of retrieval and the process continues recursively.

Evidently, at each level of this process, the main challenge involves retrieving the top-ranked causally-relevant document pertaining to the event. Therefore, in the next section we investigate the problem analytically to find the answer to our second research question – is a traditional search system adequate for the requirements of the causal information retrieval task?

3.3 Problem Investigation

Recently, the authors in [7] claimed that, for a given query event, the two sets of relevant documents (topical and causal) will have only a partial term overlap. With the help of a pseudo-relevance feedback technique, they made use of high term sampling probabilities for terms that are infrequent in the pseudo-relevant document set to identify causal documents However, prioritizing infrequent terms might always not helpful, especially in cases where the query is quite broad, such as 'Assassination of Osama bin Laden'. We illustrate this situation in Figure 2b, where it is clear that many terms, such as, Bush, Iran, SEALs, and typhoid are quite infrequent. However, these terms might not lead us to the actual causes of the event.

Therefore, to investigate the nature of causally-relevant documents and how they are coupled with that of topical one, in this paper we conduct a number of experiments on the open dataset proposed in the shared task [2]. The collection consists of $303,291$ news articles collected from Telegraph India[5]. Also, they provide 25 query topics which have a causal information need and annotated relevance judgements, each related to a different news event. We measure the cosine similarity between the two associated relevance judgement sets (topical

[5] https://www.telegraphindia.com

and causal) based on their term associations, as depicted in Figure 2a. We observe that news events which might have been triggered by multiple causes, such as *Assassination of Osama bin Laden* (topic-1) or involve prominent figures or organizations that are often reported in news articles, such as *Maharashtra chief minister resigned* (topic-3), have poor similarity between both set of documents. This reflects the fact that the causal results for this event have a small term overlap with the topical set. In contrast, the similarity value increases substantially if events have either a smaller number of causal factors, such as *Carphone Warehouse terminated deal with Channel 4* (topic-19), or are related to less significant entities, for example *Court blocks Facebook in Pakistan*. Such cases exhibit considerable term overlap, which we validate with retrieval experiments later in this paper. Furthermore, we explore this association with a couple of experiments and discuss our observations in the following subsections.

3.4 Experimental Setup

Since we aim to investigate the notion of causal relevance for query events, we analyze the performance of a number of standard retrieval models, in order to obtain an insight into whether these models can address the requirements of causality. Firstly, we employ a retrieval framework with the BM25 ranking function to see if query term overlaps with the document could capture causes or not. We named this method 'BM25' as reported in Table 2. Next, we evaluated how classical language retrieval models, specifically a linear smoothed language model performed with: (*i*) Jelinek-Mercer smoothing; (*ii*) Dirichlet smoothing [35]. We refer to these methods as 'LM-JM' and 'LM-DIR' respectively.

It is evident that there are specific representative terms for each query event which result in the difference between its corresponding topical and causal document sets. Usually query narrations are good resources for those representative terms as they clearly express information need for the associated task. Therefore, the next method that we investigate is 'BM25-TN' (i.e. search using Title along with Narration and rank by BM25), where we use *topic narrations* as queries, which in turn leads us to a causally-relevant document set. Based on the intuition that terms close to the query event in an N-dimensional word vector space might be useful to capture causes, we examine whether query reformulation with *word2vec* word vectors can capture causality. We make use of a pre-trained model, built on the Telegraph collection described previously, to help us to learn query-term associations. Once trained, this model can recommend related terms that are similar to the query terms, which might potentially be causally relevant. Thus we selected m nearby candidate terms for expanding the query to identify causal documents from the target collection, ranking them using BM25 (referred to as 'BM25-W2V').

Finally, we explored the method 'BM25-CS' (Causality Specific), where we make the query more specific to the causal information need. We consider that a user might build queries including one or more causality-indicative terms. For instance, 'Assassination of Osama bin Laden *causes* (or *reasons*)' might sound more reasonable than 'Assassination of Osama bin Laden', if the search intention is to to find the causes of the event. Therefore, we made use of a subset of 25 synonyms for the term 'cause' to formulate more causality-specified queries on which to search. This set includes terms such as: {*induce, lead, produce, provoke, compel, elicit, evoke, incite, introduce, kickoff, kindle, motivate, reason*}.

Parameter Settings. The parameters associated with BM25, specifically k_1 (used for term frequency scaling) and b (term frequency normalization by document length), were varied in range of $[0.1, 1.5]$ and $[0.1, 0.9]$ respectively in steps of 0.1. We also tuned λ for the method LM-JM in the range $[0.1, 0.9]$ (varied in steps of 0.1), and μ for LM-DIR in $[500, 2000]$ (varied in steps of 100). Additionally, we varied the number of candidate expansion terms chosen by BM25-W2V from 50 to 200, varying in steps of 10. Table 2 illustrates the optimal results achieved by optimizing parameters using grid search.

3.5 Observations

From our results we make a number of observations. Firstly, it is clear from Table 2 that, irrespective of examined model architecture, the performance of traditional retrieval algorithm drops considerably as it attempts to find causal information, in comparison with topical search. Secondly, BM25 improves recall marginally over linearly smoothed language models. However, Dirichlet-smoothed LM appears to be as efficient as BM25 in terms of precision. Thirdly, as discussed in Section 3.4, topic narrations are expected to lead us to the causal chain of any query event and should deviate the search from topical relevance to causal. In practice, BM25-TN proves to be competent in terms of capturing more cause-related information than topical in the retrieved relevant set (i.e. increased recall), which is our primary intention. Fourthly, it is evident that blindly formulating any query that itself mentions the search intention (i.e BM25-CS), or expanding a query with terms that are closely associated in the vector space of the target collection (i.e. BM25-W2V), is not adequate to harness the search scope; rather it might deviate the search intention from the actual topic to a large extent by adding noise.

	Topical				**Causal**			
	MAP	**Recall**	**NDCG**	**P@5**	**MAP**	**Recall**	**NDCG**	**P@5**
BM25	0.6400	0.9125	0.8181	0.9440	0.4690	0.7846	0.7581	0.5840
LM-JM	0.6410	0.8917	0.8148	0.9520	0.4423	0.7825	0.7411	0.5360
LM-DIR	0.6304	0.8846	0.8133	0.9040	0.4635	0.7817	0.7542	0.5840
BM25-TN	0.5774	0.8130	0.8062	0.9200	0.5272	0.9310	0.8043	0.7600
BM25-W2V	0.5390	0.7627	0.7691	0.9131	0.4410	0.6900	0.7382	0.5273
BM25-CS	0.2149	0.4829	0.4805	0.5200	0.1803	0.6170	0.4806	0.3120

Table 2: Retrieval effectiveness of various standard retrieval models, using standard retrieval evaluation metrics.

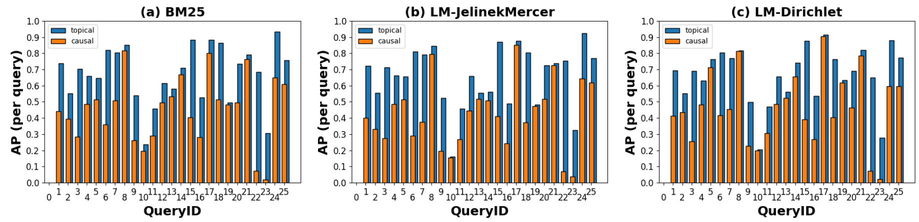


Fig. 3: Comparison of AP scores per query for standard retrieval models.

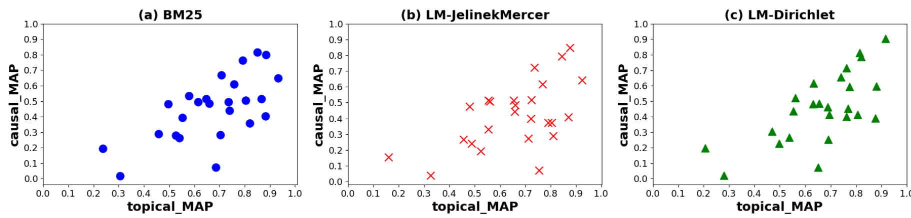


Fig. 4: Distribution of MAP scores per query for classical retrieval models.

To obtain a better understanding of document associations, we plot per-query MAP histograms for both topical and causal relevance for three of the standard retrieval frameworks (see Figure 3). Also, we show the topical-causal MAP distributions for each of the 25 queries in Figure 4. In Section 3.3, we argued that cosine similarity values between topical and causal set of documents are influenced by; (*i*) the number of causal factors (inversely proportional); (*ii*) whether the query has any association with familiar entities (holds inverse relation). The results show that the MAP values obtained for sets of topics justify this argument. For example, **topic-6**: Babri Masjid demolition case against *Advani* (Indian Politician), **topic-22**: *Lalu Prasad Yadav* (Minister of Indian Parliament and was accused for multiple scams) convicted etc. achieved lower MAP for causality task as compared to topical. Conversely, for cases, such as **topic-8**: Court blocks facebook in Pakistan (single cause query and no important entity), **topic-21**: Praveen Mahajan accused (non-public figure) etc. traditional models performed well in terms of causality.

4 Conclusion

Causal retrieval is important in situations where a user's search is focused on finding the plausible causes of an event mentioned in the search query. For instance, when a user wishes to investigate the chain of preceding occurrences in the context of event-driven news. In this paper, we have presented a high-level literature survey on causality, covering the last three decades. We have observed that there is a gap in the literature in terms of research on causality search. In an effort to mitigate this gap, we have formally defined the problem of *causal information retrieval*, and explained how it differs from traditional topical search. Furthermore, we have conducted experiments which demonstrate that traditional methods from the information retrieval literature, which are focused on topical relevance, provide limited utility in finding causally-relevant documents. This re-enforces the view that causal information retrieval remains an open challenge which is worthy of further research in the IR community.

Taking this into account, we have proposed an architecture for a *recursive* causal retrieval model that will help users to perform in-depth exploration in terms of causality pertaining to a news event, and the chain of causes which led to that event. Therefore, implementing the recursive model, conducting comprehensive offline experiments to evaluate it, and performing an extensive user study will form the most important future extensions of our work.

Acknowledgement. This work was supported by Science Foundation Ireland (SFI) under Grant Number SFI/12/RC/2289_P2.

References

1. Forum for Information Retrieval Evaluation. `http://fire.irsi.res.in/fire/2020/home`
2. Causality-driven Adhoc Information Retrieval. `https://cair-miners.github.io/CAIR-2020-website/` (2020)
3. Asghar, N.: Automatic extraction of causal relations from natural language texts:A comprehensive survey. CoRR **abs/1605.07895** (2016)
4. BBC Middle East editor: Five reasons why Israel's peace deals with the UAE and Bahrain matter. `https://www.bbc.com/news/world-middle-east-54151712` (2020), online; accessed 14 September 2020
5. Beamer, B., Girju, R.: Using a bigram event model to predict causal potential. p. 430–441. Proc. CICLing'09 (2009)
6. Blanco, E., et al.: Causal relation extraction. In: LREC 2008 (2008)
7. Datta, S., et al.: Retrieving potential causes from a query event. p. 1689–1692. Proc. SIGIR'20 (2020)
8. Dawid, A.P.: Beware of the dag! p. 59–86. Proc. COA'08 (2008)
9. Do, Q., et al.: Minimally supervised event causality identification. In: Proc. EMNLP'11. pp. 294–303 (2011)
10. Girju, R.: Automatic detection of causal relations for question answering. p. 76–83. Proc. MultiSumQA'03 (2003)
11. Gordon, A., et al.: SemEval-2012 task 7: Choice of plausible alternatives: An evaluation of commonsense causal reasoning. In: (SemEval 2012). pp. 394–398 (2012)
12. Gordon, A.S., Bejan, C., Sagae, K.: Commonsense causal reasoning using millions of personal stories. In: Proc. AAAI'11 (2011)
13. Harradon, M., et al.: Causal learning and explanation of deep neural networks via autoencoded activations. ArXiv **abs/1802.00541** (2018)
14. Hashimoto, C., et al.: Toward future scenario generation: Extracting event causality exploiting semantic relation, context, and association features. In: Proc. 52nd Annual Meeting of the ACL. pp. 987–997 (2014)

15. Hashimoto, C., et al.: Generating event causality hypotheses through semantic relations. p. 2396–2403. Proc. AAAI'15 (2015)
16. Hashimoto, C., et al.: Excitatory or inhibitory: A new semantic orientation extracts contradiction and causality from the web. In: Proc. EMNLP-CoNLL'12. pp. 619–630 (2012)
17. Inui, T., Okumura, M.: Investigating the characteristics of causal relations in Japanese text. In: Proc. Workshop on Frontiers in Corpus Annotations II: Pie in the Sky. pp. 37–44 (2005)
18. Kaplan, R.M., Berry-Rogghe, G.: Knowledge-based acquisition of causal relationships in text. Knowledge Acquisition **3**(3), 317–337 (1991)
19. Kiciman, E., Thelin, J.: Answering what if, should i, and other expectation exploration queries using causal inference over longitudinal data. In: Proc. DESIRES'18. pp. 9–15 (2018)
20. Li, P., Mao, K.: Knowledge-oriented convolutional neural network for causal relation extraction from natural language texts. Expert Systems with Applications **115**, 512–523 (2019)
21. Narendra, T., et al.: Explaining deep learning models using causal inference. ArXiv **abs/1811.04376** (2018)
22. Oh, J.H., Torisawa, K., Hashimoto, C., Sano, M., De Saeger, S., Ohtake, K.: Why-question answering using intra- and inter-sentential causal relations. In: Proc. 51st Annual Meeting of the ACL (2013)
23. Paul, M.J.: Feature selection as causal inference: Experiments with text classification. In: Proc. (CoNLL 2017). pp. 163–172 (2017)
24. Pearl, J., Paz, A.: Graphoids: Graph-based logic for reasoning about relevance relations or when would x tell you more about y if you already know z? In: Proc. ECAI'86 (1986)
25. Pearl, J.: Causal diagrams for empirical research. Biometrika **82**(4), 669–688 (1995)
26. Radinsky, K., Horvitz, E.: Mining the web to predict future events. p. 255–264. Proc. WSDM'13 (2013)
27. Radinsky, K., et al.: Learning causality for news events prediction. p. 909–918. Proc. WWW'12 (2012)
28. Raina, R., Madhavan, A., Ng, A.Y.: Large-scale deep unsupervised learning using graphics processors. In: Proc. ICML'09. p. 873–880 (2009)
29. Riaz, M., Girju, R.: Another look at causality: Discovering scenario-specific contingency relationships with no supervision. p. 361–368. Proc. ICSC'10 (2010)
30. Riaz, M., Girju, R.: In-depth exploitation of noun and verb semantics to identify causation in verb-noun pairs. In: Proc. SIGDIAL'14. pp. 161–170 (2014)
31. Rink, B., Bejan, C., Harabagiu, S.M.: Learning textual graph patterns to detect causal event relations. In: Proc. FLAIRS'11 (2010)
32. Sun, Y., et al.: Causal relation of queries from temporal logs. p. 1141–1142. Proc. WWW'07 (2007)
33. Tanaka, S., et al.: Acquiring and generalizing causal inference rules from deverbal noun constructions. In: Proc. COLING'12. pp. 1209–1218 (2012)
34. Wood-Doughty, Z., et al.: Challenges of using text classifiers for causal inference. In: Proc. EMNLP'18. pp. 4586–4598 (2018)
35. Zhai, C., Lafferty, J.: A study of smoothing methods for language models applied to ad hoc information retrieval. p. 334–342. Proc. SIGIR'01 (2001)
36. Zhang, J.: Causal reasoning with ancestral graphs. J. Machine Learning Research **9**, 1437–1474 (2008)
37. Zhao, S., et al.: Constructing and embedding abstract event causality networks from text snippets. pp. 335–344. Proc. WSDM'17 (2017)