# CauseNet: Towards a Causality Graph Extracted from the Web

Stefan Heindorf
Paderborn University
heindorf@uni-paderborn.de

Yan Scholten
Technical University of Munich
yan.scholten@tum.de

Henning Wachsmuth
Paderborn University
henningw@uni-paderborn.de

Axel-Cyrille Ngonga Ngomo
Paderborn University
axel.ngonga@uni-paderborn.de

Martin Potthast
Leipzig University
martin.potthast@uni-leipzig.de

## ABSTRACT

Causal knowledge is seen as one of the key ingredients to advance artificial intelligence. Yet, few knowledge bases comprise causal knowledge to date, possibly due to significant efforts required for validation. Notwithstanding this challenge, we compile CauseNet,[1] a large-scale knowledge base of *claimed* causal relations between causal concepts. By extraction from different semi- and unstructured web sources, we collect more than 11 million causal relations with an estimated extraction precision of 83% and construct the first large-scale and open-domain causality graph. We analyze the graph to gain insights about causal beliefs expressed on the web and we demonstrate its benefits in basic causal question answering. Future work may use the graph for causal reasoning, computational argumentation, multi-hop question answering, and more.

**ACM Reference Format:**
Stefan Heindorf, Yan Scholten, Henning Wachsmuth, Axel-Cyrille Ngonga Ngomo, and Martin Potthast. 2020. CauseNet: Towards a Causality Graph Extracted from the Web. In *The 29th ACM International Conference on Information and Knowledge Management (CIKM '20), October 19–23, 2020, Virtual Event, Ireland*. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3340531.3412763

## 1 INTRODUCTION

The nature of cause and effect is subject to inquiry in many scientific fields, including the natural sciences, philosophy, the social sciences and the humanities, and not least computer science and artificial intelligence. Since ancient times, theories of causation have been proposed and discussed, yet none have met with universal acceptance [17]. The basic statement "A causes B" for two events A and B can be interpreted under various causal models [16]. The unifying meta-model of causation of Judea Pearl distinguishes three levels of abstraction [31]: (1) association, i.e., the occurrence of A correlates with the occurrence B; (2) intervention, i.e., an action that affects A changes the probability of B; and (3) counterfactuals, i.e., if some action had been taken affecting A, it would have predictably changed the probability of B. The first level can be reached through passive observation, the second through active experimentation, and the third through understanding (e.g., a law of nature). Obtaining causal knowledge about the world is at the heart of the scientific method in particular, and learning in general.

[1]https://causenet.org


*CIKM '20, October 19–23, 2020, Virtual Event, Ireland*

ACM ISBN 978-1-4503-6859-9/20/10…$15.00
https://doi.org/10.1145/3340531.3412763

**Table 1: Overview of causal relations in knowledge bases.**

| Knowledge Base | Relations | Concepts | Acquisition | Source |
|---|---|---|---|---|
| CauseNet (83% prec.) | 11,609,890 | 12,186,310 | extracted | web+wikip. |
| CauseNet (96% prec.) | 197,806 | 80,223 | extracted | web+wikip. |
| Freebase | 128,766 | 52,487 | crowdsourced | volunteers |
| ConceptNet Multilingual | 114,308 | 57,561 | crowdsourced | volunteers |
| Wikidata | 95,335 | 88,233 | crowdsourced | volunteers |
| ConceptNet English | 21,485 | 16,432 | crowdsourced | volunteers |
| DBpedia Live | 8,025 | 7,691 | extracted | wikipedia |
| Berenberg and Bagrow [5] | 1,329 | 394 | crowdsourced | paid |
| YAGO 3 | 0 | 0 | extracted | wikipedia |

Outside of science, however, in everyday life, hardly anyone has sufficient scientific training to validate a claimed causal relation between two events. Still, virtually everyone (from as young an age as three [7]), learns causal relations as a matter of course. The sources of one's personal stock of causal beliefs are twofold: (1) learning from personal experience at all levels of Pearl's model, and, (2) learning from other people, i.e., by observing them, talking to them, or reading their texts. Neither of these sources is flawless. Even within the scientific literature, many claims of causal relations are not sufficiently justified by the evidence supplied to support them. People casually communicate their beliefs about causal relations with and without supplying justification, based on guesswork and speculation. Consequently, only a subset of one's causal beliefs can be called causal *knowledge*, as in: justified true beliefs.

Compiling an actual causal knowledge base requires expert council for any involved domain, so that doing so at scale is a Herculean task. This is perhaps the reason why not many causal knowledge bases have been compiled to date (see Table 1 for an overview). The intermediate step of compiling a knowledge base of *claimed* causal relations appears more feasible. Similar to how humans obtain causal beliefs, causal relations can be extracted at scale from natural language texts on the web. Even though such a knowledge base might not be suitable for all use cases, (1) it could serve as a testing ground for corresponding algorithms, (2) may be part of an information system, e.g., for question answering, and (3) could provide invaluable insights into the state of our society's causal beliefs at large, enabling applications in the computational social sciences and artificial intelligence, e.g., to assess the acceptance of certain views or to reason in automated decision making.

In this paper, we construct, analyze, and apply CauseNet, a large-scale knowledge base of claimed causal relations extracted from the web. We implement a number of information extraction algorithms for causal relations and tune them to high precision in order to collect over 11 million causal relations from the ClueWeb12 corpus as well as from Wikipedia. Going beyond previous work, we reconcile the extracted causal concepts to a coherent knowledge graph. The causal relations and the resulting causal network are subject to a number of analyses to explore the data, its origins, and its properties. Finally, we demonstrate the potential and practical applicability of CauseNet in a question answering experiment.

In what follows, we briefly review related work (Section 2). In Section 3, we define causality graphs and describe how we extract causal relations from natural language texts. Section 4 gives key statistics about our graph. Section 5 evaluates the graph in terms of precision and recall as part of a manual analysis, and demonstrates its application for causal question answering. Section 6 discusses strengths and weaknesses of our approach.

## 2 RELATED WORK

*Causality Graphs.* To our knowledge, no dedicated causality graph is available to date. Perhaps most related to our work, Berenberg and Bagrow [5] crowdsourced a small causality graph which is not publicly available. While causality graphs can be seen as special cases of knowledge graphs, existing knowledge graphs contain few causal relations and hardly suffice to answer common causal questions. Our graph contributes orders of magnitude more causal relations. Knowledge graphs can be categorized by their construction methodology, degree of normalization, and domain specificity: Wikidata [38] and ConceptNet [37] are constructed manually by the crowd, DBpedia [3] and YAGO [24] automatically from Wikipedia. We adopt an automatic construction approach for better scalability. While Wikidata, DBpedia, and YAGO are strongly normalized with a strict notion of entities, WordNet [26], BabelNet [29], and ConceptNet [37] are weakly normalized with nodes representing lexemes rather than entities. Lacking a controlled set of causal entities and avoiding information loss in the normalization process, we adopt weak normalization. All the aforementioned knowledge graphs are open-domain, but there are also domain-specific ones, e.g., for genes [21] or news [33]. CauseNet is open-domain, capturing claimed causal relations on the web.

*Causal Relation Sources.* Hashimoto [13] extracts causal relations from Wikipedia; Radinsky et al. [33] extract causal relations from the New York Times, BBC, and WikiNews; Ittoo and Bouma [18] from domain-specific documents, such as customer complaints. Other causality extraction approaches such as Hendrickx et al. [15], Li et al. [23], and Li and Mao [22] do not extract real-world relations at all and stick to small-scale benchmarking datasets. To the best of our knowledge, web crawls, such as the ClueWeb crawls, have not been considered for causality extraction so far. Semi-structured sources for causal relations, such as Wikipedia infoboxes and lists which yield relations with a high precision, have been largely neglected, too (except for infoboxes by Hashimoto [13]). Although Wikipedia's infoboxes are one of the main sources of knowledge for DBpedia [3] and YAGO [24], neither of them contains many causal relations, possibly because most of this information was added to Wikipedia only recently.

*Causal Relation Extraction.* Existing works for extracting relations from natural language texts follow two paradigms: bootstrapping and supervised machine learning. The relation extraction approaches Snowball [1] and DIRPE [6] are bootstrapping approaches, which form the basis of our approach for causal relations: Starting with a small number of seed pairs of entities in a desired relation, sentences are identified that contain those entities. Analyzing these sentences for linguistic patterns expressing the relation, new entities are identified enabling iterative repetition. Supervised machine learning either requires fully annotated training texts, which are currently unavailable for causal relations and expensive to create, or distant supervision sources: Mintz et al. [27] propose a generic distant supervision training approach to train a classifier based on entities and relations from an existing knowledge base. However, in a pilot experiment, we found causal knowledge to be too sparse and too specific in existing knowledge bases (e.g., more than 80% of Wikidata's causal relations refer to people's causes of death).

Extraction approaches can also be distinguished by their learning mechanism, namely, surface patterns [11, 20], dependency trees [18, 19], and neural networks [10, 22, 23]. While surface patterns are based on literal strings between entities, dependency trees consider the linguistic dependencies of tokens (e.g., as generated by the Stanford NLP parser). When trained from scratch, neural networks require large amounts of training data, which are currently unavailable; hence, we employ dependency trees.

Hassanzadeh et al. [14] extract causal relations from news articles and Hashimoto [13] develops an approach to extract causal relations between Wikipedia entities (described by pages). Neither of them make their extracted causal relations available and both provide little insights on them. Li and Mao [22], Li et al. [23] extract *causal* relations with neural network architectures, such as BiLSTMs and CNNs. However, they train and evaluate their approach only on small benchmarking datasets, e.g., data from a SemEval challenge on relation classification [15]. In contrast, we extract causal relations at scale from the web. Moreover, the SemEval dataset was created for relation classification *given* two entities, not for identifying the entities themselves. Each entity annotation consists of a single token only, and while Li et al. [23] suggest better ones, their dataset is not publicly available. We thus compile new training data for causal entity recognition on the web.

*Causal Question Answering.* Girju [11] were one of the first to extract causal relations for question answering. However, their approach uses a small set of 50 questions only. Sharp et al. [35] use a larger set of 3,031 questions that were semi-automatically selected from Yahoo Answers with linguistic patterns. We consider search engine questions with a focus on causal questions instead by mining them from Microsoft's MS MARCO question answering dataset [4], which comprises over one million questions from Bing's search logs. Its high-quality ground truth answers were obtained by human crowdworkers, who were presented with text passages obtained by Bing's passage retrieval system, and who wrote new answers in their own words. In cases where the provided passages contained no answer, annotators marked the question as unanswerable. While Hassanzadeh et al. [14] aim to answer binary causal questions, they artificially transform causal relations to binary causal questions instead of answering real questions, e.g., by search engine users.

## 3 CAUSALITY GRAPH CONSTRUCTION

For the construction of CauseNet, two sources of causal relations are tapped, namely the web in general in the form of a large crawl, and Wikipedia in particular. This section describes our construction pipeline with regard to these two sources, and how they were merged to form a coherent causal knowledge graph. Our graph and source code is publicly available under permissive CC BY-SA 4.0 and MIT licenses.[2]

### 3.1 Operationalization

Causal relations are modeled as relations between causal concepts in the form of lexemes, i.e., words and noun phrases: Two causal concepts $C$ and $E$, where $C$ takes the role of a cause and $E$ that of an effect, are claimed to be causally related if $C$ *may have* or *might have had* a causal effect on $E$, according to some web sources. The causal effect between $C$ and $E$ can be claimed to be direct or indirect, and we model causal relations independent of any conditions required to activate the causality. Modeling causality as a directed relation between concepts enables the construction of a structured knowledge graph $\mathcal{K}$, which stores knowledge about the world as subject-predicate-object triples $(s, p, o) \in \mathcal{K}$. At present, only the predicate $p = mayCause$ is employed, which allows for an easy integration into existing knowledge graphs. For each relation, we store comprehensive provenance data.

### 3.2 Causal Relation Extraction from the Web

To extract causal relations from the web at scale, we analyze the ClueWeb12 web crawl, which comprises about 733,019,372 English web pages crawled between February and May 2012.[3] We chose this crawl over more recent ones, such as the Common Crawl, since it is heavily used as part of many TREC evaluation tracks, thus fostering synergies in future research. Our extraction approach is precision-oriented, and given the scale requirements of the ClueWeb12, we resort to a minimally supervised bootstrapping approach [1, 6, 18] that only requires a few initial training samples.

The bootstrapping process starts with a small set of well-known causal relations as seeds (A→B stands for "A causes B"):

| | | | |
|---|---|---|---|
| smoking→cancer; | earthquake→tsunami; | rainfall→flooding; | rain→flood; |
| radiation→cancer; | dehydration→death; | poison→death; | HIV→AIDS |

Using the seeds, sentences in a corpus of text are sought that contain both the cause and effect of a given seed. From these sentences, linguistic patterns are mined that capture how causal relations are expressed in natural language. The linguistic patterns enable the extraction of previously unseen causal relations from the text corpus. This process is iteratively performed two times, augmenting the set of seed relations with the new ones found after each iteration.

Not all newfound patterns and causal relations qualify as additional seeds for the next iteration. They are selected based on a *support* criterion (adapted from [1]): Pattern support is measured by the number of unique seeds that led to its identification. Similarly, seed support is measured by the number of unique patterns that extract the seed's causal relation. The support criterion relies on the redundancy of causal relations on the web: The more unique ways a causal relation is expressed in, the higher its support. While support is not the only heuristic described in the literature [12], our pilot studies showed it to work well and more robust at a large-scale than other heuristics, such as *confidence* [1] and *reliability* [18, 30]. For our high-recall version of CauseNet (83% precision), we include all extracted relations; for our high-precision version (96% precision), the support is at least 2 (cf., Section 5).

**Table 2: Top 10 patterns of causal relations: cause/N and effect/N refer to nouns within causal concepts, joined in a sentence fulfilling the respective pattern's dependencies.**

| Linguistic Pattern | | | | | | Relations |
|---|---|---|---|---|---|---|
| Cause dependency | | Token/POS | Effect dependency | | | \|E\| |
| cause/N | -nsubj | cause/VB | +dobj | effect/N | | 904,385 |
| cause/N | -nmod:with | associated/VBN | -acl | effect/N | | 892,908 |
| cause/N | -nsubj | lead/VB | +nmod:to | effect/N | | 783,860 |
| cause/N | -nsubj | led/VBD | +nmod:to | effect/N | | 724,978 |
| cause/N | -nsubjpass | associated/VBN | +nmod:with | effect/N | | 692,666 |
| cause/N | -nmod:by | caused/VBN | -acl | effect/N | | 598,639 |
| cause/N | -nsubj | result/VB | +nmod:in | effect/N | | 552,352 |
| cause/N | -nsubj | causes/VBZ | +dobj | effect/N | | 496,426 |
| cause/N | -nsubj | leads/VBZ | +nmod:to | effect/N | | 491,340 |
| cause/N | -nsubj | resulted/VBD | +nmod:in | effect/N | | 473,298 |

The linguistic patterns are mined based on the shortest path between nouns signifying cause and effect within the dependency graph of a sentence [8, 9, 18], which was implemented using the so-called enhanced dependency graphs from the Stanford NLP Parser [34]. To obtain linguistic patterns for causality extraction with high precision, we apply two bootstrapping iterations to Wikipedia. In each iteration, the number of selected seeds is increased by 10, the number of patterns by 25, merging them with those obtained in the previous iteration. For example, merging the 25 patterns after the first iteration with the 50 patterns after the second iteration yields 53 unique patterns. The top 10 patterns are shown in Table 2. In a pilot study, we experimented with varying numbers of iterations, patterns, and seeds. Increasing the number of iterations increases recall at the cost of precision, whereas varying the number of seeds and patterns had little effect and yielded similar results. Having obtained the 53 linguistic patterns from Wikipedia, we apply them to the entire ClueWeb12 crawl. Typical indicators for causal statements include "cause(d)", "lead(s)/led", "result(ed)", and "associated with".

The linguistic patterns presume each causal concept (i.e., each cause and effect) to contain a noun. As a concept can be composed of multiple words (e.g., "global warming", "human activity", or "lack of exercise"), we determine the exact start and end of a causal concept in a sentence using the following concept spotter: First, the sentences are tokenized and part-of-speech-tagged using the Stanford Parser [32]. Modeling concept spotting as a sequence tagging problem similar to Li et al. [23], we tag each token with an *inside, outside, beginning* tag using the IOB2 format, where the first token of a causal concept is tagged as *beginning*, other tokens of the same concept as *inside*, and other words as *outside*. As a sequence tagger, we use the state-of-the-art BiLSTM-CRF model based on Flair embeddings [2].

[2] https://github.com/causenet-org/CIKM-20 https://doi.org/10.5281/zenodo.3876154
[3] https://lemurproject.org/clueweb12/

To train the sequence tagger, we heuristically identified Wikipedia articles that potentially contain causal relations, reviewed their sentences, and tagged them manually: By searching for causal headings in their table of contents (e.g., “cause(s)”, “effect(s)”, “risk factor(s)”, “symptom(s)”, and “signs and symptoms”), we found 8,974 articles with causal headings from which we randomly sampled 100. Of these, 87 contained 832 overall sentences with at least one cause and one effect. These sentences express a total 1,572 causal relations between causal concepts, suggesting that some sentences contain multiple causal relations. Using 80% of the data for training, 10% for development, and 10% as test set, we achieved an $F_1$-score of 0.65 with 0.62 precision and 0.67 recall on the test set. We extract causal concepts from all sentences retrieved from the ClueWeb12 using the linguistic patterns. In cases the tagger failed, we disregarded the causal concepts.

## 3.3 Causal Relation Extraction from Wikipedia

During our investigation of Wikipedia to bootstrap causal relation extraction on the web, we found that Wikipedia itself comprises as of yet untapped sources of causal relations: infoboxes and listings.

*Infoboxes.* An infobox in a Wikipedia article presents information as key-value pairs with keys being predefined by templates. For example, the infobox about “lung cancer” contains a key “risk factors” with value “tobacco smoking, genetic factors, [...]” and is derived from the template “medical condition”.[4] Although infoboxes have been successfully used as a source for knowledge graph construction, particularly for DBpedia [3] and YAGO [24], the many causal relations in infoboxes are not reflected in either of them: YAGO’s ontology entirely lacks causal properties and even DBpedia Live [28] misses many causal relations as of August 2020.

To reconstruct causal relations as subject-predicate-object triples from infoboxes, their titles serve as subject (or their articles’ titles in case an infobox has none), their causality-related infobox keys (e.g., “cause”, “causes”, “symptoms” and “risks”) serve as unified predicate *mayCause* (or its inverse), and their respective values are used as objects. While the infobox keys are a controlled vocabulary, their values are arbitrary unstructured text and frequently contain lists of causal concepts (for instance, consider the aforementioned example from the infobox about lung cancer). To spot individual causal concepts from these lists, we retrain the aforementioned sequence tagger for infobox values. Trained on 179 randomly selected and manually annotated infobox values, the sequence tagger achieves an $F_1$-score of 0.95 with 0.95 precision and 0.96 recall. We further determine a relation’s direction based on the infobox and its predicate. For example, extracting “tobacco smoking” from the infobox about lung cancer, we define the relation (tobacco smoking, mayCause, lung cancer), because the “risks” predicate is inverse to *mayCause*. To reduce noise, we exclude keys from templates which are used in less than 10 articles. Moreover, keys from six infobox templates are omitted that express causal relations only indirectly related to the topic of an article.[5]

[4] https://en.wikipedia.org/wiki/Lung_cancer

[5] For example, given an article about a person and a key “cause of death”, neither holds that any cause of death causes the person, nor that the person causes a cause of death. The causality rather refers to “death” than the “person”.

**Table 3: (a) Overview of CauseNet and its subgraphs due to ClueWeb12’s pages (CW pages) and Wikipedia’s articles’ plain texts (WP texts), lists, and infoboxes (ibxs.), where $V$ denotes the set of nodes in the form of causal concepts, and $E$ the set of edges in the form of causal relations between the concepts. (b) Overlap of causal relations between the subgraphs. (c) Overview of the sources of causal relations.**

(a)

| CauseNet | $\|V\|$ | $\|E\|$ |
|---|---|---|
| Joint graph | 12,186,310 | 11,609,890 |
| CW pages | 11,368,371 | 10,872,313 |
| WP texts | 1,070,686 | 793,593 |
| WP lists | 8,295 | 10,612 |
| WP infoboxes | 7,201 | 7,880 |

(b)

| $E_1$ | $E_2$ | $\|E_1 \cap E_2\|$ |
|---|---|---|
| pages | texts | 72,937 |
| pages | ibxs. | 939 |
| pages | lists | 506 |
| texts | ibxs. | 209 |
| lists | ibxs. | 93 |
| texts | lists | 70 |

(c)

| Source | Pages / Articles |
|---|---|
| CW pages | 733,019,372 |
| w/ causality | 12,111,758 |
| WP articles | 5,208,098 |
| w/ caus. in texts | 427,893 |
| w/ caus. in ibxs. | 2,725 |
| w/ caus. in lists | 1,194 |

*Listings.* Besides infoboxes, Wikipedia articles often contain itemized lists that represent causal relations. Because lists are made of only one kind of causal concept (i.e., either causes or effects), the respective other concept needs to be derived from the remainder of an article. We do so based on the hierarchical structure of an article and its title: extracting causal relations from lists appearing within one of the special top-level sections “cause(s)”, “symptom(s)”, “signs and symptoms”, and “risk factor(s)”,[6] and assuming the missing concept to be an article’s title. Like for infoboxes, we determine a relation’s direction with a simple hand-coded lookup rule based on the section title. Since each list item itself can contain a (comma-separated) list of causal concepts, we again retrain our sequence tagger for them, focusing on short list items that do not constitute full sentences[7] and that do not start with the word “see”, which points to a special kind of list used directly after section headings to link to articles with more details. Trained on 358 randomly selected and manually annotated list values, the sequence tagger achieves an $F_1$-score of 0.92 with 0.92 precision and 0.91 recall. List items consisting of full sentences are processed as described in Section 3.2.

## 3.4 Causality Graph Induction

To induce a coherent causality graph from the aforementioned three sets of causal relations, the causal concepts must be reconciled: First, all causal concepts are post-processed by discarding determiners, coordinating conjunctions, personal and possessive pronouns, and punctuation at both ends of a causal concept. Then, concepts are merged into a single graph node if their lower-case representations are equal. The resulting causality graph consists of unique causal relations between normalized concepts similar to ConceptNet [37].

For each relation in the causality graph, we store provenance information about its source: If the relation was extracted from a Wikipedia or ClueWeb12 sentence, we document the sentence and the path pattern that extracted the relation as well as a URL to the original source. Source information further allows us to adjust precision and recall of the causality graph, for example by filtering the graph for different sources and patterns.

[6] We exclude lists in the “effect(s)” sections. They are often used for sound effects rather than for causal relations.

[7] This is determined by the verb POS tags “VB”, “VBD”, “VBG”, “VBN”, “VBP”, “VBZ”.

**Table 4: Source analysis: The upper row shows the top sources of causal relations from ClueWeb12 in terms of (a) hostname, (b) domain, (c) top-level domain, and (d) DMOZ category, ordered by number of causal relations. The lower row shows sources of causal relations in Wikipedia: (e) infobox templates, and articles with causal relations in (f) infoboxes, (g) lists, and (h) texts.**

(a)

| Hostname | Category | $\|E\|$ |
|---|---|---|
| sdbonline.org | Science | 26,517 |
| bionewsonline.com | Science | 25,212 |
| jci.org | Science | 16,081 |
| sec.gov | Regional | 13,907 |
| plosone.org | Science | 12,722 |
| molvis.org | Science | 9,544 |
| neurotransmitter.net | Reference | 8,842 |
| diseaseinformation.info | Reference | 8,829 |
| leninist.biz | Reference | 8,033 |
| lansbury.bwh.harvard.edu | Science | 7,828 |

(b)

| Domain | Category | Subdomains | $\|E\|$ |
|---|---|---|---|
| researchtoday.net | Science | 302 | 125,728 |
| wordpress.com | Society | 6,835 | 91,230 |
| typepad.com | Society | 5,687 | 72,357 |
| hubpages.com | Society | 4,370 | 40,473 |
| nih.gov | Regional | 368 | 40,280 |
| deviantart.com | Arts | 20,365 | 40,064 |
| about.com | Reference | 828 | 36,363 |
| tripod.com | Society | 1,877 | 31,131 |
| sdbonline.org | Science | 1 | 26,517 |
| bionewsonline.com | Science | 1 | 25,212 |

(c)

| TLD | $\|E\|$ |
|---|---|
| .com | 5,597,297 |
| .org | 2,590,683 |
| .net | 793,937 |
| .edu | 766,731 |
| .gov | 320,263 |
| .co.uk | 229,834 |
| .ca | 185,661 |
| .info | 138,519 |
| .org.uk | 111,697 |
| .ac.uk | 85,304 |

(d)

| Category | Domains | $\|E\|$ |
|---|---|---|
| Science | 121 | 296,330 |
| Reference | 118 | 240,033 |
| Health | 84 | 147,851 |
| Society | 80 | 129,058 |
| Regional | 34 | 76,754 |
| Business | 21 | 43,900 |
| News | 11 | 33,906 |
| Computers | 18 | 27,319 |
| Shopping | 9 | 14,078 |
| Arts | 3 | 8,017 |

(e)

| Infobox Template | Articles | $\|E\|$ |
|---|---|---|
| medical condition (new) | 820 | 4,923 |
| civil conflict | 579 | 1,339 |
| rail accident | 452 | 530 |
| event | 380 | 495 |
| wildfire | 257 | 306 |
| news event | 146 | 170 |
| oil spill | 35 | 36 |
| military conflict | 23 | 32 |
| birth control | 13 | 26 |
| bus accident | 20 | 23 |

(f)

| Wikipedia article title (ibxs.) | $\|E\|$ |
|---|---|
| 2013 Romanian protests (…) | 23 |
| Shock (circulatory) | 19 |
| Breast cancer | 18 |
| Constipation | 17 |
| Intracerebral hemorrhage | 17 |
| Protests against Donald Trump | 17 |
| Heat stroke | 16 |
| Scombroid food poisoning | 16 |
| Acute lymphoblastic leukemia | 15 |
| Bowel obstruction | 15 |

(g)

| Wikipedia article title (lists) | $\|E\|$ |
|---|---|
| Flushing (physiology) | 58 |
| Mast cell activation syndrome | 56 |
| Coarse facial features | 50 |
| Hypotonia | 47 |
| Autistic catatonia | 46 |
| Livedo reticularis | 46 |
| Pallor | 43 |
| Delayed puberty | 42 |
| Eosinophilic myocarditis | 42 |
| Intraparenchymal hemorrhage | 42 |

(h)

| Wikipedia article title (texts) | $\|E\|$ |
|---|---|
| Effects of global warming on human health | 98 |
| Hepatitis | 79 |
| Horse colic | 77 |
| Safety of electronic cigarettes | 72 |
| Nutritional neuroscience | 71 |
| Causes of cancer pain | 70 |
| Dog health | 69 |
| Long-term effects of alcohol consumption | 69 |
| Famine | 67 |
| Progeroid syndromes | 60 |

# 4 ANALYSIS OF CAUSENET

The joint causality graph CauseNet consists of 11,609,890 causal relations (Table 3a). To the best of our knowledge, it represents the largest general-purpose causality graph to date. This section reports first results of a cursory exploratory analysis to investigate how and what kinds of causal relations are discussed on the web, and what is the structure of the causality graph built from them.

## 4.1 Source Analysis

Most of CauseNet's relations originate from natural language texts, 10,872,313 from the ClueWeb12, and 793,593 from the plain texts of Wikipedia articles (Table 3a). A smaller fraction of 10,612 and 7,880 originates from Wikipedia lists and infoboxes, respectively. Table 3b shows the overlap of the causality graphs from these sources. Compared to the respective subgraphs' overall sizes, their overlaps are relatively small. Despite the fact that the same linguistic patterns were applied to extract causal relations from the ClueWeb and the plain text of Wikipedia's articles, only about 73,000 causal relations out of 11 million from ClueWeb overlap with the 0.8 million from Wikipedia. Similarly, the overlap between infoboxes, lists, and texts is small. This suggests that the different sources of causal relations complement each other. Causal relations are found in about 1.7% of all pages from ClueWeb12, and 8.2% of Wikipedia's articles (Table 3c), where the higher prevalence of causal relations in Wikipedia can be explained by the fact that it is an encyclopedia.

*ClueWeb12.* The 11 million causal relations of CauseNet that originate from ClueWeb12 are distributed across 12 million web pages. To get an idea of the types of pages in which causal relations can be found, we grouped them by hostname, domain name, top-level domain, and DMOZ category. Table 4 shows the respective top 10 of each. Causal relations appear on 842,698 hosts with 18 relations per host on average. Among the top 10 hostnames (Table 4a), all but two sites cover biological or medical information. The remainder are sec.gov and leninist.biz. When grouping by domain, causal relations appear on 635,861 domains with an average of 22 relations per domain. Table 4b shows that particularly blog providers are among the top sources. Grouping by top-level domain (TLD), causal relations are found on 1,181 TLDs at an average of 10,206 relations per TLD. Table 4c shows that ".edu" and ".gov" are among the top sources, following the three most frequently used TLDs.

For a better understanding of the abstract topics of the websites from which causal relations have been extracted, the top 500 hostnames as per their contribution of causal relations were manually categorized into the 10 top-level categories of DMOZ, a widely-studied web directory [36]. Table 4d shows that the "science", "reference", and "health" categories contribute the largest amounts of causal relations. Causal relations can also often be found on law-related sites, here categorized as "society". Altogether, the primary sources of causal relations appear to be more often scientific, educational, health-related, and law-related websites.

*Wikipedia.* Wikipedia represents the single most important website for causality on the web: 793,593 causal relations were extracted from 427,893 Wikipedia articles (Tables 3a and 3c).[8] Moreover, 2,725 articles contained infoboxes and 1,194 lists with causal information. Although infoboxes with causal relations are more prevalent than lists with such relations, the causal graph induced by lists is larger. Regarding infoboxes, 10 templates feature causal information (Table 4e). The top template "infobox medical conditions (new)" describes causes, symptoms, and risk factors of medical conditions and represents 4,923 causal relations from 820 articles. In addition to medical causes, infoboxes often describe causes of civil and military conflicts as well as technical and natural disasters. Tables 4f-h show the Wikipedia articles with the most causal

[8]Wikipedia was omitted from the ClueWeb12 in favor of a direct dump.

Table 5: **Graph Analysis: The upper row shows the 10 most central nodes according to degree centrality (Cent.) in subgraphs extracted from (a) ClueWeb12 sentences, and Wikipedia (b) infoboxes, (c) lists, and (d) texts. The lower row shows 10 paths of (e) length 1, (f) 2 and (g) 3, ordered by rounded path support (Sup.), i.e., the geometric mean of edge support. We do not show paths with cycles and we only present node-disjoint paths to give a broader overview.**

(a)

| Concept | Out | In | Cent. |
|---|---|---|---|
| problems | 8,077 | 64,355 | 0.006 |
| death | 4,485 | 44,144 | 0.004 |
| damage | 3,890 | 28,301 | 0.003 |
| pain | 3,668 | 23,046 | 0.002 |
| disease | 11,198 | 15,175 | 0.002 |
| injury | 6,681 | 14,733 | 0.002 |
| stress | 10,114 | 9,077 | 0.002 |
| changes | 9,155 | 9,459 | 0.002 |
| problem | 2,608 | 15,975 | 0.002 |
| symptoms | 2,720 | 14,415 | 0.002 |

(b)

| Concept | Out | In | Cent. |
|---|---|---|---|
| unknown | 117 | 0 | 0.016 |
| fever | 4 | 103 | 0.015 |
| lightning | 80 | 0 | 0.011 |
| family history | 73 | 0 | 0.010 |
| under investigation | 68 | 0 | 0.009 |
| vomiting | 1 | 56 | 0.008 |
| obesity | 49 | 3 | 0.007 |
| shortness of breath | 0 | 52 | 0.007 |
| arson | 47 | 0 | 0.007 |
| smoking | 46 | 0 | 0.006 |

(c)

| Concept | Out | In | Cent. |
|---|---|---|---|
| fatigue | 3 | 66 | 0.008 |
| nausea | 0 | 68 | 0.008 |
| vomiting | 1 | 59 | 0.007 |
| flushing (physiology) | 0 | 58 | 0.007 |
| mast cell activation syndrome | 56 | 0 | 0.007 |
| fever | 5 | 50 | 0.007 |
| hypotonia | 1 | 53 | 0.007 |
| tachycardia | 4 | 50 | 0.007 |
| coarse facial features | 0 | 50 | 0.006 |
| anxiety | 35 | 14 | 0.006 |

(d)

| Concept | Out | In | Cent. |
|---|---|---|---|
| death | 689 | 4,054 | 0.004 |
| problems | 412 | 2,519 | 0.003 |
| damage | 340 | 1,953 | 0.002 |
| controversy | 427 | 1,729 | 0.002 |
| disease | 831 | 1,019 | 0.002 |
| events | 1,625 | 220 | 0.002 |
| accident | 778 | 1,064 | 0.002 |
| incident | 1,465 | 309 | 0.002 |
| deaths | 113 | 1,361 | 0.001 |
| success | 811 | 632 | 0.001 |

(e)

| Cause | Effect | Support |
|---|---|---|
| accident | death | 38 |
| drought | famine | 31 |
| injury | pain | 31 |
| disease | deaths | 30 |
| smoking | lung cancer | 30 |
| stress | illness | 30 |
| depression | suicide | 28 |
| anxiety | insomnia | 27 |
| bacteria | infection | 27 |
| diarrhea | dehydration | 27 |

(f)

| Cause | Mediator | Effect | Support |
|---|---|---|---|
| stress | illness | death | 33 |
| accident | injury | pain | 31 |
| exposure | disease | deaths | 28 |
| bacteria | infection | inflammation | 26 |
| obesity | diabetes | blindness | 24 |
| anxiety | depression | suicide | 24 |
| global warming | drought | famine | 24 |
| diarrhea | dehydration | headaches | 23 |
| lightning | fire | damage | 22 |
| negligence | injuries | disability | 21 |

(g)

| Cause | Mediator 1 | Mediator 2 | Effect | Support |
|---|---|---|---|---|
| negligence | accident | injury | death | 29 |
| bacteria | infection | disease | deaths | 27 |
| inflammation | pain | depression | suicide | 26 |
| fear | stress | illness | disability | 23 |
| greenhouse gases | global warming | drought | famine | 23 |
| lack of exercise | obesity | diabetes | blindness | 23 |
| lightning | fire | damage | cancer | 20 |
| virus | diarrhea | dehydration | headaches | 20 |
| anemia | fatigue | accidents | injuries | 19 |
| alcohol | problems | anxiety | insomnia | 19 |

relations extracted from texts, lists, and infoboxes, revealing topic differences: Causality in infoboxes refers to medical conditions and political protests; lists focus on medical conditions; article texts are more diverse, discussing societal issues such as global warming, medical conditions, safety of electrical cigarettes, as well as famine.

## 4.2 Graph Analysis

To gain first insights into the nature of large causality graphs, we analyze centrality, known unknowns, and causal paths.

*Central Causal Concepts.* To determine the most central concepts, we assess their degree centrality, i.e., the number of their incident (incoming and outgoing) edges normalized by the number of nodes in the graph except the node itself. The higher the degree centrality, the more likely causal information "flows" through the concept in question. Table 5 overviews the most central causal concepts with respect to the subgraphs from (a) ClueWeb texts, and Wikipedia (b) infoboxes, (c) lists, and (d) texts. The central concepts largely differ between subgraphs, except for "death", "problems", and "damage", which are shared among Wikipedia texts and ClueWeb12 pages. This variety is due to the aforementioned complementary nature of these sources of causal relations. Nevertheless, a topic shared among the central concepts of all subgraphs is medicine, e.g., "nausea", "fatigue", "death", and "pain". But causal relations from the web also include "economic inequality", "corruption", "problems", "success", "war", "storm", and "lightning".

The degree centrality gives further insights into the roles of concepts in causal relations: If the out-degree of a node is greater than its in-degree, the concept is more often used as cause, and vice versa. For example, the node "nausea" in the list graph only appears as effect. Most nodes in the infobox graph act either as cause or as effect. Only "fever", "vomiting", and "obesity" act as both. Such clear usage gets blurred, the larger the causality graphs are. However, some concepts still tend to be used more as cause or effect; for instance, "death" is rarely used as cause of something.

*Known Unknowns.* For targeted knowledge acquisition, knowing what is not, as of yet, known is important: In this regard, infoboxes frequently report known unknowns using the value "unknown", which even represents the most central node in the infobox subgraph. Spot checks show that known unknowns are hardly expressed in lists and texts. While our extraction pipeline focuses on noun phrases, we found that known unknowns in texts are often expressed in terms of adjectives. We leave it to future work to extract known unknowns from texts, too.

*Causal Relations and Causal Paths.* The causal relations in CauseNet originate from many different web pages, whereas inducing a graph among them creates causal paths none of which have been explicitly claimed in any of the individual sources. To gain further insights into these paths, Tables 5e-g show paths of length 1-3. Paths of length 1 are ordered by their support, and paths of lengths 2 and 3 by the average support of their individual relations as per their geometric mean. We present only node-disjoint paths, neglecting cycles, for a broader overview. Apparently, medical paths dominate. Moreover, most causal paths refer to negative events, including accidents, diseases, or death. These findings might corroborate and explain previous studies on "Cyberchondria" [39], where it was found that people searching for medical symptoms on the web often assume the worst. Despite the apparent dominance of negative connotations, our causality graph also comprises examples of positive causal relations, such as *love→happiness* and *teamwork→success*. Future work could utilize our causality graph to analyze not only the sentiment of causal relations or causal paths but also the influence of causality on sentiment in sentences.

## 5 EVALUATION AND APPLICATION

To evaluate our causality graph, we estimate its precision manually, and its recall within a question-answering benchmark application.

### 5.1 Estimating Precision Manually

We assess the precision of the causal relation extraction as follows: From each combination of CauseNet subgraph and support level shown in Table 6a, 100 causal relations are sampled for manual review. For texts, we consider a causal relation correctly extracted, if the relation is correctly extracted from at least two of three different, randomly selected sentences. If less than three different sentences are available, we consider only one randomly selected sentence to avoid draws. For infoboxes and lists, we manually double check the extracted relation with its single source. For relations with support 1, we achieve a precision between 0.74 and 0.79 for ClueWeb12 and Wikipedia. A support of 2 increases precision to over 0.95, more support yields a precision of 0.97 or more. Thus, increasing support allows for trading recall for precision. For lists and infoboxes, a generally high precision between 0.96 and 0.97 is achieved.

### 5.2 Estimating Recall via Question Answering

We were wondering whether the scale of causal relations compiled would allow for answering real-world causal questions, and how many. To address this question, we relied on the MS MARCO machine reading comprehension dataset [4]. Among its one million questions submitted to Bing, many causal ones can be found, plus answers as an independent ground truth. We focus on the subset of basic binary causal questions of the form "can A cause B", where A and B are causal concepts.[9] For instance, the question "Can global warming cause an ice age?" is then answered by looking up the relation (global warming, mayCause, ice age). Out of 4,287 such questions in MS MARCO, 2,169 have a ground truth answer. To link the questions' causal concepts to the available knowledge bases, we employ exact string matching for CauseNet and ConceptNet, DBpedia Spotlight [25] for DBpedia, and DBpedia's interwiki-links for Wikidata. For the latter two, only 1,423 and 1,331 questions could be linked, respectively.

Table 6b shows the results for our causality graph and its subgraphs compared to ConceptNet, DBpedia, and Wikidata. Out of 1829 questions answered with "Yes" in the ground truth, CauseNet (based primarily on its ClueWeb12 subgraph) provides the same answer in 487 cases, while contradicting the ground truth answer "No" in 53 out of 340 cases. This yields a "Yes"-precision and recall of 0.9 and 0.27, respectively. The performance of all other subgraphs and knowledge bases is negligible in comparison. These results show that our sample of causal relations extracted from the ClueWeb12 recall a quarter of those present in the much larger Bing crawl, since MS MARCO's ground truth answers indicate whether retrieved documents answered the question as judged by a crowd worker.

[9] Variants of the pattern replace "can" with "do/does/did/is/are/will/would", and "cause" with "causes/caused/causing". For simplicity, we exclude complex causal concepts with coordinating conjunctions, prepositions, and subordinating conjunctions, to-prepositions, Wh-determiners, Wh-adverbs, and Wh-pronouns as determined with the Stanford part-of-speech tagger [32], and then discard leading determiners and possessive pronouns of the remainder. The MS MARCO answers are reduced to "yes" or "no", excluding questions with multiple contradictory answers.

**Table 6: CauseNet evaluation: (a) Manual precision estimation; (b) automatic recall estimation via question answering.**

(a)

| CauseNet | Sup. | Relations | Prec. |
|---|---|---|---|
| Complete | 1-38 | 11,609,890 | 0.83 |
| High-Prec. | 2-38 | 197,806 | 0.96 |
| CW pages | 1 | 10,700,845 | 0.74 |
| CW pages | 2 | 116,912 | 0.95 |
| CW pages | 3-37 | 54,556 | 0.97 |
| WP texts | 1 | 789,381 | 0.79 |
| WP texts | 2 | 3,237 | 0.96 |
| WP texts | 3-21 | 975 | 0.98 |
| WP lists | n/a | 10,612 | 0.97 |
| WP ibxs. | n/a | 7,880 | 0.96 |

(b)

| Truth: | Yes | | No | | Total |
|---|---|---|---|---|---|
| Prediction: | Yes | No | Yes | No | |
| CauseNet | 487 | 1,342 | 53 | 287 | 2,169 |
| CW pages | 480 | 1,349 | 52 | 288 | 2,169 |
| WP texts | 51 | 1,778 | 8 | 332 | 2,169 |
| WP lists | 8 | 1,821 | 0 | 340 | 2,169 |
| WP ibxs. | 9 | 1,820 | 1 | 339 | 2,169 |
| ConceptNet | 1 | 1,828 | 0 | 340 | 2,169 |
| DBpedia Live | 11 | 1,179 | 2 | 231 | 1,423 |
| Wikidata | 10 | 1,098 | 0 | 223 | 1,331 |

## 6 DISCUSSION AND LIMITATIONS

*Causal Relation Extraction.* We extract causal relations expressed within single sentences with 53 linguistic patterns. Besides, causal relations might be expressed (1) implicitly rather than explicitly, (2) across sentences, (3) differently across domains, or (4) using negation. In this regard, our collection is biased towards domains where causal relations are discussed explicitly and concisely, such as scientific, health, and law-related topics. But even within science, causal relations are not always made explicit: For example, the Wikipedia article on "Free Fall" never explicitly mentions that gravity causes objects to fall. Similarly, if causal relations are expressed across sentences, coreferences might need to be resolved first. While we excluded sentences containing the common negations "no", "not", "doesn't" and "didn't", more work is required to properly process sentences like "The film features scientists and others who are skeptical that global warming is caused by human activity." Likewise, explicit statements of the absence of causality are of interest.

*Modeling Causality.* Our causal relations comprise causes and effects in the form of noun phrases, similar to ConceptNet and WordNet. An alternative approach might be to model causality in terms of entities, as is done by Wikidata and DBpedia. Of course, entity linkers could be employed to link our concepts to knowledge base entities. However, entities alone are insufficient at present. For example, the entities "artery" and "stroke" cannot express the fact that only a *blocked* artery can cause a stroke and data models are required that take conditional information into account.

We model causality with a single predicate *mayCause* allowing data consumers to easily use our graph. A more fine-grained distinction with respect to different kinds of causation such as "producing", "increasing" or "decreasing" are conceivable. Also, temporal aspects of causation could be modeled for applications like news event prediction [33]. Incorporating the probability of a cause having an effect would benefit causal reasoning applications.

Causal relations can be expressed on different levels of abstraction, omitting intermediate steps in a causal chain. For instance, stress might not be a direct cause of cancer, but might make people smoke more. Smoking, in turn, might not be a direct cause of cancer, either, but gene mutations triggered by the interaction of smoke with lung cells. Capturing causal relations at different levels of abstraction may be done within a hierarchical graph.

*Confidence in Causal Relations.* We extract causal relations from web sources and our graph reflects how (parts of) our society (currently) thinks about causality. This includes claims that HIV is caused by homosexuality or that autism is caused by vaccination. We include detailed provenance information for all causal relations. Just as today's search engines and question answering systems deliver results without guarantees, they are still useful at large.

# 7 CONCLUSION

The quest to unravel causality is as old as human civilization itself. Ever since ancient times, humanity has tried, and will continue to try to learn as much about the environment as possible in order to predict its behavior and to affect change, big and small, in desirable directions. If an artificial intelligence with any form of agency accompanies us in the future, it will have to learn causal relations, too, or else be the unwitting subject to the whims of its environment. Faced with a quest of such epic proportions, we stand awestruck by what has already been accomplished, and how much still remains to be done. Although collecting all beliefs and separating fact from fiction is still far beyond current technological capabilities, we expect that further unraveling causality will be supported by the current stock of causal beliefs. In this paper, we have started to scale the technology that already works. With CauseNet, we provide a causality graph with about 200,000 relations at 96% precision, along with a high-recall variant with 11.6M relations at 83% precision. While we have exemplified its benefit for causal question answering, future work may utilize it in the context of explainable AI, automated decision making, and much more.

# ACKNOWLEDGMENTS

The authors gratefully acknowledge the funding of this project by computing time provided by the Paderborn Center for Parallel Computing ($PC^2$).

# REFERENCES


- [1] Eugene Agichtein and Luis Gravano. 2000. *Snowball*: Extracting Relations from Large Plain-Text Collections. In *ACM DL*. ACM, 85–94.
- [2] Alan Akbik, Duncan Blythe, and Roland Vollgraf. 2018. Contextual String Embeddings for Sequence Labeling. In *COLING*. ACL, 1638–1649.
- [3] Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary G. Ives. 2007. DBpedia: A Nucleus for a Web of Open Data. In *ISWC/ASWC*, Vol. 4825. Springer, 722–735.
- [4] Payal Bajaj, Daniel Campos, Nick Craswell, Li Deng, Jianfeng Gao, Xiaodong Liu, Rangan Majumder, Andrew McNamara, Bhaskar Mitra, Tri Nguyen, Mir Rosenberg, Xia Song, Alina Stoica, Saurabh Tiwary, and Tong Wang. 2018. MS MARCO: A Human Generated MAchine Reading COmprehension Dataset. *CoRR* abs/1611.09268v3 (2018).
- [5] Daniel Berenberg and James P. Bagrow. 2018. Efficient Crowd Exploration of Large Networks: The Case of Causal Attribution. *PACMHCI* 2, CSCW (2018), 24:1–24:25.
- [6] Sergey Brin. 1998. Extracting Patterns and Relations from the World Wide Web. In *WebDB (LNCS, Vol. 1590)*. Springer, 172–183.
- [7] Merry Bullock and Rochel Gelman. 1979. Preschool Children's Assumptions about Cause and Effect: Temporal Ordering. *Child Development* 50, 1 (1979), 89–96.
- [8] Razvan C. Bunescu and Raymond J. Mooney. 2005. A Shortest Path Dependency Kernel for Relation Extraction. In *HLT/EMNLP*. ACL, 724–731.
- [9] Aron Culotta and Jeffrey S. Sorensen. 2004. Dependency Tree Kernels for Relation Extraction. In *ACL*. ACL, 423–429.
- [10] Tirthankar Dasgupta, Rupsa Saha, Lipika Dey, and Abir Naskar. 2018. Automatic Extraction of Causal Relations from Text using Linguistically Informed Deep Neural Networks. In *SIGDIAL Conference*. ACL, 306–316.
- [11] Roxana Girju. 2003. Automatic Detection of Causal Relations for Question Answering. In *Proceedings of the ACL 2003 workshop on Multilingual summarization and question answering-Volume 12*. ACL, 76–83.
- [12] Sonal Gupta and Christopher D. Manning. 2014. Improved Pattern Learning for Bootstrapped Entity Extraction. In *CoNLL*. ACL, 98–108.
- [13] Chikara Hashimoto. 2019. Weakly Supervised Multilingual Causality Extraction from Wikipedia. In *EMNLP/IJCNLP*. ACL, 2986–2997.
- [14] Oktie Hassanzadeh, Debarun Bhattacharjya, Mark Feblowitz, Kavitha Srinivas, Michael Perrone, Shirin Sohrabi, and Michael Katz. 2019. Answering Binary Causal Questions Through Large-Scale Text Mining: An Evaluation Using Cause-Effect Pairs from Human Experts. In *IJCAI*. 5003–5009.
- [15] Iris Hendrickx, Su Nam Kim, Zornitsa Kozareva, Preslav Nakov, Diarmuid Ó Séaghdha, Sebastian Padó, Marco Pennacchiotti, Lorenza Romano, and Stan Szpakowicz. 2010. SemEval-2010 Task 8: Multi-Way Classification of Semantic Relations between Pairs of Nominals. In *SemEval@ACL*. ACL, 33–38.
- [16] Christopher Hitchcock. 2020. Causal Models. In *The Stanford Encyclopedia of Philosophy* (summer 2020 ed.), Edward N. Zalta (Ed.). Metaphysics Research Lab, Stanford University.
- [17] Menno Hulswit. 2004. A Short History of Causation. *SEED Journal (Semiotics, Evolution, Energy, and Development)* 4, 3 (2004), 16–42.
- [18] Ashwin Ittoo and Gosse Bouma. 2013. Minimally-Supervised Learning of Domain-specific Causal Relations Using an Open-Domain Corpus as Knowledge Base. *Data Knowl. Eng.* 88 (2013), 142–163.
- [19] Sophia Katrenko and Pieter W. Adriaans. 2006. Learning Relations from Biomedical Corpora Using Dependency Trees. In *KDECB (LNCS, Vol. 4366)*. Springer, 61–80.
- [20] Zornitsa Kozareva. 2012. Cause-Effect Relation Learning. In *TextGraphs@ACL*. ACL, 39–43.
- [21] Volodymyr Kuleshov, Jialin Ding, Christopher Vo, Braden Hancock, Alexander Ratner, Yang Li, Christopher Ré, Serafim Batzoglou, and Michael Snyder. 2019. A Machine-compiled Database of Genome-wide Association Studies. *Nature communications* 10, 1 (2019), 3341.
- [22] Pengfei Li and Kezhi Mao. 2019. Knowledge-oriented Convolutional Neural Network for Causal Relation Extraction from Natural Language Texts. *Expert Syst. Appl.* 115 (2019), 512–523.
- [23] Zhaoning Li, Qi Li, Xiaotian Zou, and Jiangtao Ren. 2019. Causality Extraction based on Self-Attentive BiLSTM-CRF with Transferred Embeddings. *CoRR* abs/1904.07629 (2019).
- [24] Farzaneh Mahdisoltani, Joanna Biega, and Fabian M. Suchanek. 2015. YAGO3: A Knowledge Base from Multilingual Wikipedias. In *CIDR*.
- [25] Pablo N. Mendes, Max Jakob, Andrés García-Silva, and Christian Bizer. 2011. DBpedia Spotlight: Shedding Light on the Web of Documents. In *I-SEMANTICS*. ACM, 1–8.
- [26] George A. Miller. 1995. WordNet: A Lexical Database for English. *Communications of the ACM* 38, 11 (1995), 39–41.
- [27] Mike Mintz, Steven Bills, Rion Snow, and Daniel Jurafsky. 2009. Distant Supervision for Relation Extraction Without Labeled Data. In *ACL/IJCNLP*. ACL, 1003–1011.
- [28] Mohamed Morsey, Jens Lehmann, Sören Auer, Claus Stadler, and Sebastian Hellmann. 2012. DBpedia and the Live Extraction of Structured Data from Wikipedia. *Program* 46, 2 (2012), 157–181.
- [29] Roberto Navigli and Simone Paolo Ponzetto. 2010. BabelNet: Building a Very Large Multilingual Semantic Network. In *ACL*. ACL, 216–225.
- [30] Patrick Pantel and Marco Pennacchiotti. 2006. Espresso: Leveraging Generic Patterns for Automatically Harvesting Semantic Relations. In *ACL*. ACL.
- [31] Judea Pearl. 2000. *Causality: Models, Reasoning, and Inference.* Cambridge University Press, USA.
- [32] Peng Qi, Timothy Dozat, Yuhao Zhang, and Christopher D. Manning. 2018. Universal Dependency Parsing from Scratch. In *CoNLL Shared Task*. ACL, 160–170.
- [33] Kira Radinsky, Sagie Davidovich, and Shaul Markovitch. 2012. Learning Causality for News Events Prediction. In *WWW*. ACM, 909–918.
- [34] Sebastian Schuster and Christopher D. Manning. 2016. Enhanced English Universal Dependencies: An Improved Representation for Natural Language Understanding Tasks. In *LREC*. ELRA.
- [35] Rebecca Sharp, Mihai Surdeanu, Peter Jansen, Peter Clark, and Michael Hammond. 2016. Creating Causal Embeddings for Question Answering with Minimal Supervision. In *EMNLP*. ACL, 138–148.
- [36] Ashadullah Shawon, Syed Tauhid Zuhori, Firoz Mahmud, and Md Jamil-Ur Rahman. 2018. Website Classification Using Word Based Multiple N-Gram Models and Random Search Oriented Feature Parameters. In *ICCIT*. IEEE, 1–6.
- [37] Robyn Speer, Joshua Chin, and Catherine Havasi. 2017. ConceptNet 5.5: An Open Multilingual Graph of General Knowledge. In *AAAI*. 4444–4451.
- [38] Denny Vrandecic and Markus Krötzsch. 2014. Wikidata: A Free Collaborative Knowledgebase. *Communications of the ACM* 57, 10 (2014), 78–85.
- [39] Ryen W. White and Eric Horvitz. 2009. Cyberchondria: Studies of the Escalation of Medical Concerns in Web Search. *ACM Transactions on Information Systems* 27, 4 (2009), 23:1–23:37.