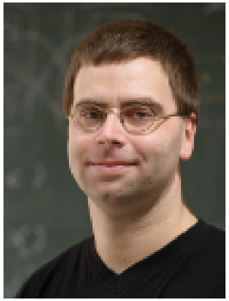
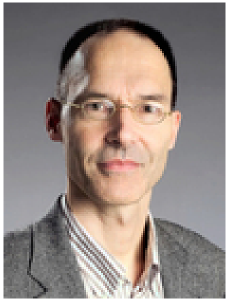
Mining Health-related Cause–Effect Statements with High Precision at Large Scale

October 12–17, 2022


Ferdinand Schlatt*


Dieter Bettin‡


Matthias Hagen*


Benno Stein†


Martin Potthast§

*Martin-Luther-Universität Halle-Wittenberg †Bauhaus-Universität Weimar
‡Westfälische Wilhelms-Universität Münster §Universität Leipzig

www.webis.de

COLING 2022 29th International Conference on Computational Linguistics (COLING 2022)

Health-related Cause–Effect Statements on the Web

The web is abundant with health-related cause–effect statements

Health-related Cause–Effect Statements on the Web

The web is abundant with health-related cause–effect statements

MedlinePlus
Trusted Health Information for You

“Some ear disorders can result in hearing disorders and deafness.”

Health-related Cause–Effect Statements on the Web

The web is abundant with health-related cause–effect statements

MedlinePlus
Trusted Health Information for You

“Some ear disorders can result in hearing disorders and deafness.”

$15.70

“Magnesium deficiency may be a major cause of fatal cardiac arrhythmia, hypertension . . . ”

Health-related Cause–Effect Statements on the Web

The web is abundant with health-related cause–effect statements

MedlinePlus
Trusted Health Information for You

“Some ear disorders can result in hearing disorders and deafness.”

$15.70

“Magnesium deficiency may be a major cause of fatal cardiac arrhythmia, hypertension . . . ”

“The weak strength of Mars causes blood infections, high/low BP, obstruction in physical growth, muscle cramps, . . . ”

Health-related Cause–Effect Statements on the Web

The web is abundant with health-related cause–effect statements

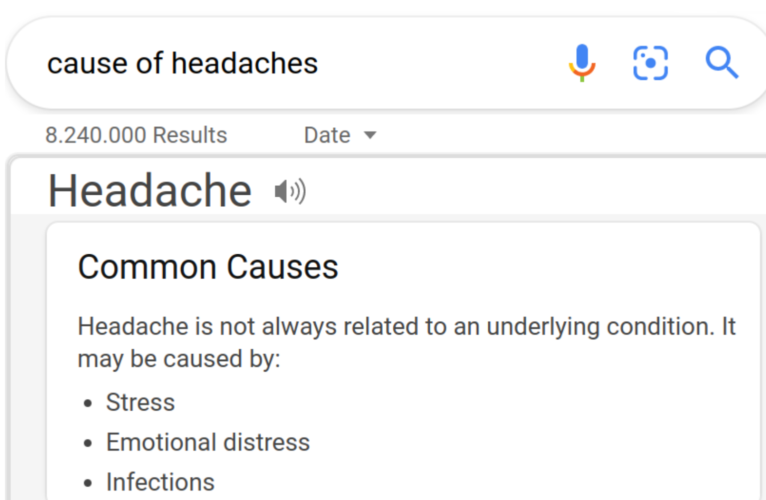
MedlinePlus
Trusted Health Information for You

“Some ear disorders can result in hearing disorders and deafness.”

$15.70

“Magnesium deficiency may be a major cause of fatal cardiac arrhythmia, hypertension . . . ”

“The weak strength of Mars causes blood infections, high/low BP, obstruction in physical growth, muscle cramps, . . . ”



Cause–Effect Statement Extraction

State of the Art

CauseNet is a graph of more than 11M causal relations extracted from the ClueWeb12 web crawl [Heindorf et al., CIKM'20]

Cause–Effect Statement Extraction

State of the Art

CauseNet is a graph of more than 11M causal relations extracted from the ClueWeb12 web crawl [Heindorf et al., CIKM'20]

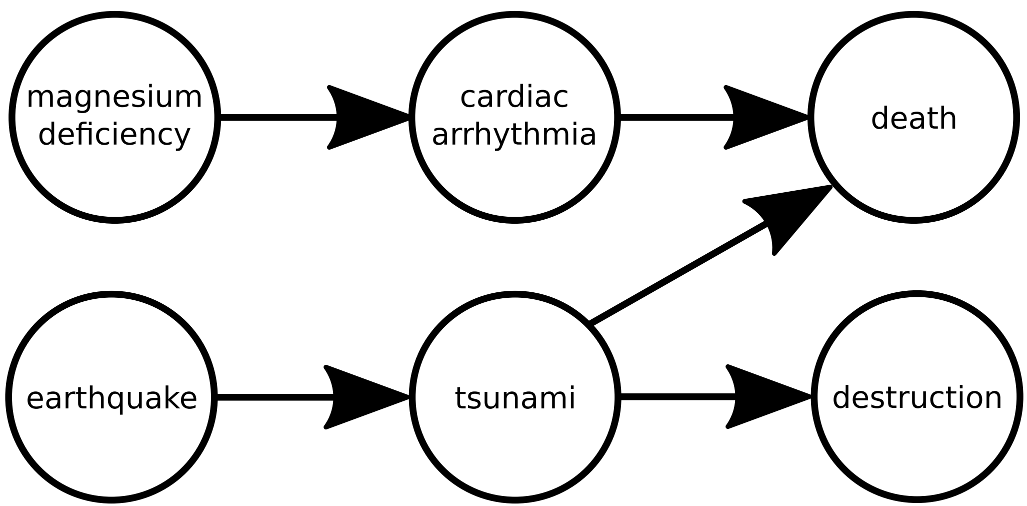
- "*Magnesium deficiency* may **be a cause** of *cardiac arrhythmia*."
- "*Cardiac arrhythmia* may **lead to** *death*."
- "The *tsunami* **was caused** by an *earthquake*."
- "The *tsunami* **resulted in** *death* and *destruction*."

Cause–Effect Statement Extraction

State of the Art

CauseNet is a graph of more than 11M causal relations extracted from the ClueWeb12 web crawl [Heindorf et al., CIKM'20]

- "*Magnesium deficiency* may **be a cause** of *cardiac arrhythmia*."
- "*Cardiac arrhythmia* may **lead to** *death*."
- "The *tsunami* **was caused** by an *earthquake*."
- "The *tsunami* **resulted in** *death* and *destruction*."

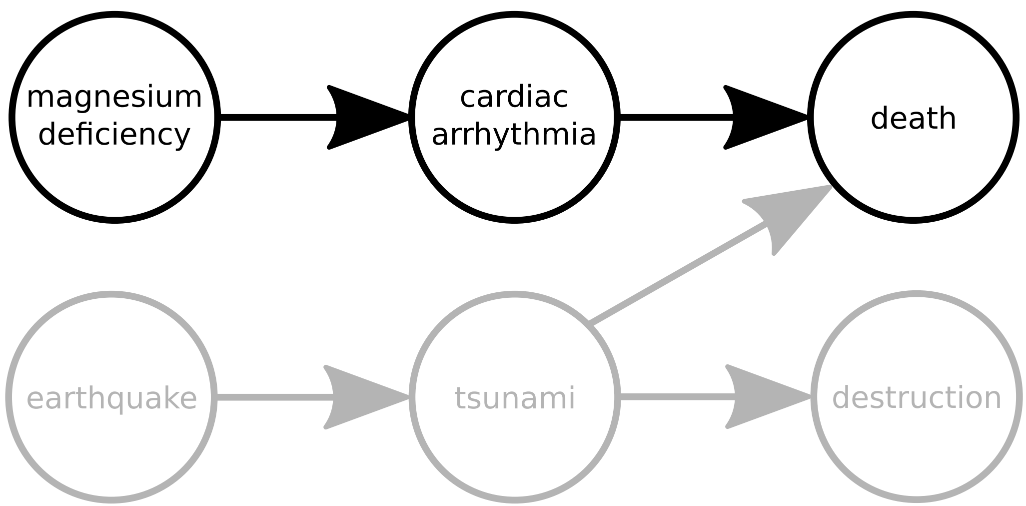


Cause–Effect Statement Extraction

State of the Art

CauseNet is a graph of more than 11M causal relations extracted from the ClueWeb12 web crawl [Heindorf et al., CIKM'20]

- "*Magnesium deficiency* may **be a cause** of *cardiac arrhythmia*."
- "*Cardiac arrhythmia* may **lead to** *death*."
- "The *tsunami* **was caused** by an *earthquake*."
- "The *tsunami* **resulted in** *death* and *destruction*."



Phrase Health Relatedness Assessment

Termhood Scores

Task: Classify whether a phrase is health-related

- “Her death was a result of cancer.”
- “Virgo, cancer, and mercury are associated with food and nourishment.”

Phrase Health Relatedness Assessment

Termhood Scores

Task: Classify whether a phrase is health-related

- "Her death was a result of cancer."
- "Virgo, cancer, and mercury are associated with food and nourishment."

Termhood scores determine the degree to which a phrase is specific for a domain

Phrase	Health-related Corpus (PubMed)	Contrastive Corpus (Wikipedia)
actor	5,590	**539,180**
carcinoma	**987,164**	7,410
diagnosis	**1,851,514**	34,218
study	**10,630,098**	508,740
the	**200,926,211**	196,374,618
ward	47,099	**186,811**

Termhood Scores

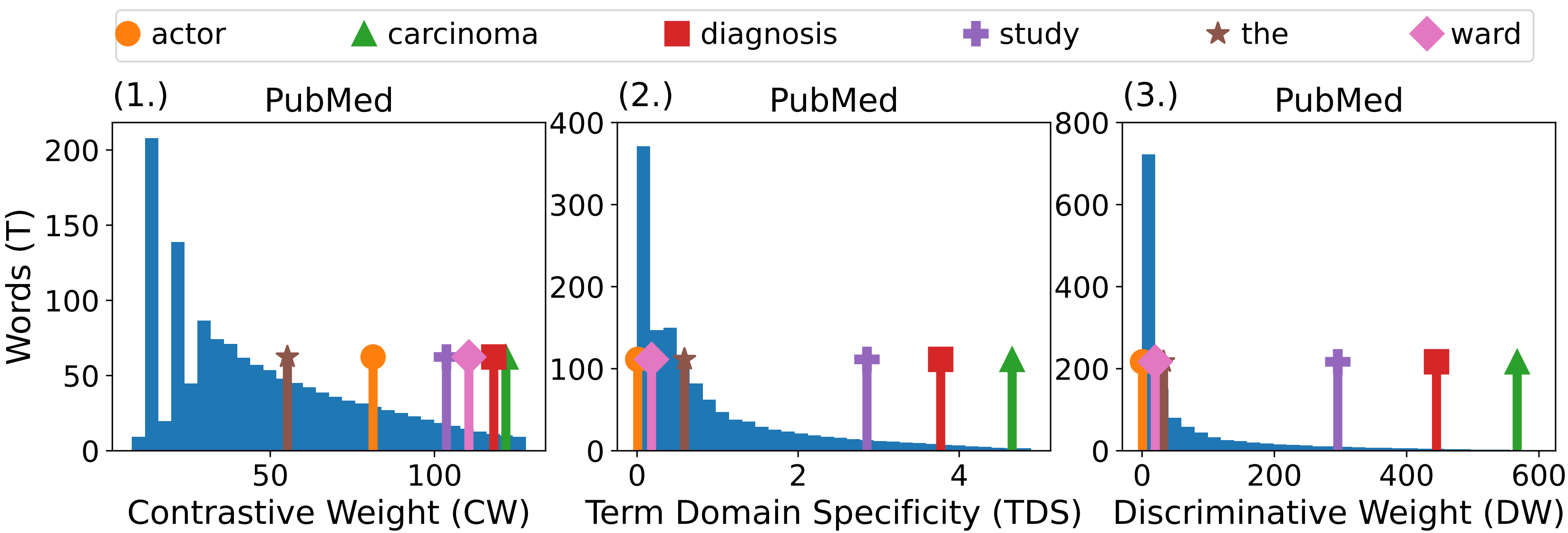
Specifically three scores were considered in this work

1. Contrastive Weight [Basili et al., TIA'01]
2. Term Domain Specificity [Park et al., INTERSPEECH'08]
3. Discriminative Weight [Wong et al., AusDM'07]

Termhood Scores

Specifically three scores were considered in this work

1. Contrastive Weight [Basili et al., TIA'01]
2. Term Domain Specificity [Park et al., INTERSPEECH'08]
3. Discriminative Weight [Wong et al., AusDM'07]



Evaluation

Annotated Datasets

We evaluate our termhood scores on two manually labelled subsets of CauseNet

Dataset	Type	Train Size	Test Size
CauseNet-P-Phrase	High Precision	800	200
CauseNet-F-Phrase	Full	800	200

Evaluation

Comparison Approaches

We evaluate our termhood scores on two manually labelled subsets of CauseNet

Dataset	**Type**	**Train Size**	**Test Size**
CauseNet-P-Phrase	High Precision	800	200
CauseNet-F-Phrase	Full	800	200

and compare them to four medical entity linkers

1. cTakes [Savova et al., Journal of the American Medical Informatics Association (2010)]
2. MetaMap [Aronson et al., AMIA'01]
3. QuickUMLS [Soldaini et al., MedIR'16]
4. SciSpacy [Neumann et al., BioNLP@ACL'19]

as well as three fine-tuned BERT models

1. Base BERT [Devlin et al., NAACL'19]
2. SciBERT [Beltagy et al., BioNLP@ACL'19]
3. PubMedBERT [Gu et al., ACM Transactions on Computing for Healthcare (2022)]

Evaluation

Effectiveness Results

Optimizing for precision (>0.9) or Matthews correlation coefficient

Main Takeaways

	Approach	R	M
P-Phr.	Entity Linker	–	0.52 †
	BERT	**0.89**	**0.83**
	Termhood	**0.89**	0.79
F-Phr.	Entity Linker	0.05 †	0.54 †
	BERT	**0.82**	**0.82**
	Termhood	0.74	0.79

Evaluation

Effectiveness Results

Optimizing for precision (>0.9) or Matthews correlation coefficient

Main Takeaways

1. Entity linkers significantly† worse . . .

	Approach	R	M
P-Phr.	Entity Linker	–	0.52†
	BERT	**0.89**	**0.83**
	Termhood	**0.89**	0.79
F-Phr.	Entity Linker	0.05†	0.54†
	BERT	**0.82**	**0.82**
	Termhood	0.74	0.79

Evaluation

Effectiveness Results

Optimizing for precision (>0.9) or Matthews correlation coefficient

Main Takeaways

1. Entity linkers significantly† worse especially in high-precision scenarios

	Approach	R	M
P-Phr.	Entity Linker	$\boxed{–}$	0.52 †
	BERT	**0.89**	**0.83**
	Termhood	**0.89**	0.79
F-Phr.	Entity Linker	$\boxed{0.05}$ †	0.54 †
	BERT	**0.82**	**0.82**
	Termhood	0.74	0.79

Evaluation

Effectiveness Results

Optimizing for precision (>0.9) or Matthews correlation coefficient

Main Takeaways

1. Entity linkers significantly[†] worse especially in high-precision scenarios
2. BERT best effectiveness in all scenarios

	Approach	R	M
P-Phr.	Entity Linker	–	0.52 †
	BERT	$\boxed{\mathbf{0.89}}$	$\boxed{\mathbf{0.83}}$
	Termhood	**0.89**	0.79
F-Phr.	Entity Linker	0.05 †	0.54 †
	BERT	$\boxed{\mathbf{0.82}}$	$\boxed{\mathbf{0.82}}$
	Termhood	0.74	0.79

Evaluation

Effectiveness Results

Optimizing for precision (>0.9) or Matthews correlation coefficient

Main Takeaways

1. Entity linkers significantly† worse especially in high-precision scenarios
2. BERT best effectiveness in all scenarios
3. Termhood scores close second . . .

	Approach	R	M
P-Phr.	Entity Linker	–	0.52 †
	BERT	$\boxed{\textbf{0.89}}$	$\boxed{\textbf{0.83}}$
	Termhood	$\boxed{\textbf{0.89}}$	$\boxed{0.79}$
F-Phr.	Entity Linker	0.05 †	0.54 †
	BERT	**0.82**	$\boxed{\textbf{0.82}}$
	Termhood	0.74	$\boxed{0.79}$

Evaluation

Effectiveness Results

Optimizing for precision (>0.9) or Matthews correlation coefficient

Main Takeaways

1. Entity linkers significantly† worse especially in high-precision scenarios
2. BERT best effectiveness in all scenarios
3. Termhood scores close second but 108x times faster

	Approach	R	M
P-Phr.	Entity Linker	–	0.52 †
	BERT	**0.89**	**0.83**
	Termhood	**0.89**	0.79
F-Phr.	Entity Linker	0.05 †	0.54 †
	BERT	**0.82**	**0.82**
	Termhood	0.74	0.79

Webis Health CauseNet 2022

Using the termhood scores, we create a resource of health-related cause–effect statements extracted from the web

https://github.com/webis-de/COLING-22

Subset	Statements	Sentences	P	R
Precision (P)	103,273	1,259,339	0.93	0.89
Precision (MCC)	112,707	1,340,873	0.89	0.93
Full (P)	2,201,071	5,680,635	1.00	0.74
Full (MCC)	3,206,964	7,842,464	0.78	0.90

Webis Health CauseNet 2022

Using the termhood scores, we create a resource of health-related cause–effect statements extracted from the web

https://github.com/webis-de/COLING-22

Subset	Statements	Sentences	P	R
Precision (P)	103,273	1,259,339	0.93	0.89
Precision (MCC)	112,707	1,340,873	0.89	0.93
Full (P)	2,201,071	5,680,635	1.00	0.74
Full (MCC)	3,206,964	7,842,464	0.78	0.90

Important to note: it contains *claimed* statements

- ‘stress’ ➜ ‘insomnia’
- ‘incorrect placement of jupiter’ ➜ ‘diabetes’

Summary

Contributions

Generalized termhood-based method to effectively and efficiently assess the health-relatedness of phrases

- Performs as good or not significantly worse than BERT
- Substantially faster than BERT, making application at web scale possible

Creation and release of Webis Health CauseNet 2022

- 7.8M health-related cause–effect sentences
- Useful for health-sociological analyses, health-related web search, and more

Code and data available at:

`https://github.com/webis-de/COLING-22`

Summary
Contributions

Generalized termhood-based method to effectively and efficiently assess the health-relatedness of phrases

- Performs as good or not significantly worse than BERT
- Substantially faster than BERT, making application at web scale possible

Creation and release of Webis Health CauseNet 2022

- 7.8M health-related cause–effect sentences
- Useful for health-sociological analyses, health-related web search, and more

Code and data available at:

https://github.com/webis-de/COLING-22

Thank you!

Evaluation

Runtime Results

Runtime comparison classifying the phrase datasets

Main Takeaways

1. Termhood scores substantially faster than all other approaches
2. Large variance in entity linkers and comparably slow

Approach	**ms**	$\times$
cTakes	119.68	0.5
MetaMap	49.64	1.2
QuickUMLS	7.23	8.3
ScispaCy	16.38	3.7
PubMedBERT	60.19	1.0
Termhood	**0.56**	**107.5**