Libres volume 29, issue 2, pages 99-108 (2020)

Automatic Extraction of Causal Chains from Text

Aliaksandr Huminski
Institute of High Performance Computing, Singapore
huminskia@ihpc.a-star.edu.sg

Ng Yan Bin
Institute of High Performance Computing, Singapore
ng_yan_bin@ihpc.a-star.edu.sg

ABSTRACT

Background. Automatic extraction of causal chains is valuable for discovering previously unknown and hidden connections between events. However, there is only a handful of works devoted to automatic extraction of causal chains from text.
Objective. To develop a method for automatic extraction of causal chains from text.
Method. A new approach based on linguistic templates is suggested for causal chain extraction. It is domain-independent, not restricted to extraction from single sentences and unfolded on big data. For implementation, a sequence of four modules was deployed. These are verb restriction, part-of-speech tagging, extracting causal relations, and unification and matching events.
Results. 14,821 causal chains (with length=2) have been extracted from 100,000 English Wikipedia articles.
Contributions. The extracted causal chains can contribute to developing commonsense knowledge bases, reasoning resources, problem-solving, and generally in discovering previously unknown relationships between entities/events.

INTRODUCTION

In recent years, automatic extraction of causal chains has become increasingly important to discover previously unknown relationships between entities or events. This type of knowledge could be very valuable (de Silva, Zhibo, Rui, & Kezhi, 2017). Causal chains are particularly useful in medicine and biology, where they can be applied to finding unknown connections between symptoms, diseases, and their drugs (Khoo, Chan, & Niu, 2000). By analyzing chains of causal implication within the medical literature, new hypotheses for causes of rare diseases have been discovered. Some of those have received supporting experimental evidence (Swanson, 1987; Swanson & Smalheiser, 1997). The chain of causation can also benefit problem-solving systems.

Nevertheless, to the best of our knowledge, there are only a handful of works devoted to automatic extraction of causal chains from text (Asghar, 2016). Few works determine the causal chain based on the template *NP1* causal-verb *NP2* (Sawamaru & Kobayashi, 2012). A number of cases were observed where a single sentence contains two

Published by WKW School of Communication & Information & NTU Libraries, Nanyang Technological University

different causal assertions, chained together. To handle multiple instances of causality present in the same sentence, it is split into sub-sentences (Hendrickx et al., 2009).

There is some research on extracting causal chains in narrow domains, for example, to determine the chain of causation of teen drug addiction from the documents for enhancing the warning system on the social Web (Pechsiri & Sukharomana, 2017), or to reveal the connection between initial information about the incident and its causes in aviation investigation reports, based on a graph-based text representation that captures both the structure and the content of the report (Sizov & Ozturk, 2013).

We present an approach to causal chain extraction that is domain-independent, not restricted to single sentences, and unfolded on big data.

CAUSAL CHAIN VERSUS CAUSAL RELATION

A causal chain is defined as a sequence of causal relations that lead to some final effect. So, we have something like*: event-1 causes event-2 causes event-3*, etc. In other words, a causal chain is a sequence of events related by causality.

Causal chains extracted from the text are not just a summation of particular chains but a causal chain net, where several events *{Ai}* can cause the same event *B* which, in turn, can cause several events *{Cj}*. For example, the event B *open the door* can be caused by events *{Ai}* of pushing, pulling or kicking the door; while events *{Cj}* such as *entering the room*, or *getting fresh air* can be effects of the opening the door.

PROBLEM OF CAUSAL CHAIN EXTRACTION

Extraction of causal chain assumes that in a cause-effect relation, the effect can be a follow-up cause. It suggests that the extraction pattern for cause and effect should be the same.

The traditional explicit syntactic patterns for the detection and extraction of causal relations being focused on 2-member causal extraction (Bethard, Corvey, Klingenstein, & Martin, 2008; Khoo, Kornfilt, Oddy, & Myaeng, 1998; Luo, Zhu, Hwang, & Wang, 2016; O'Gorman, Wright-Bettner, & Palmer, 2016) cannot be used directly for this purpose since they either have different sub-patterns for cause and effect (resultative constructions) or implicit patterns such as, causal links or causal cues; if-then conditionals: or adverbs/adjectives where cause and effect do not have any patterns at all.

The difficulties with causal chain extraction can be illustrated in the following example from (Khoo et al., 1998). Let's assume the following causal relation was extracted: *It was raining heavily and because of this, the car failed to brake in time*. There are two events here: event A (*It was raining heavily*) and event B (*the car failed to brake in time*). Now, to extract a chain, we need the event B, as an effect in cause-effect relation with the event A, to consider as a cause for the next step in chain causality. So, we need to find an event C that will be caused by the event B. The problem is that it is very unlikely for the event B to be described by the same seven words (*the car failed to brake in time*) in both A *causes* B and B *causes* C. The problem is how to represent event B for matching.

METHOD OF CAUSAL CHAIN EXTRACTION

Causal chain extraction A *causes* B *causes* C assumes (1) finding explicit linguistic patterns (clues) of causality between A&B and B&C, and (2) finding event B which is the same in both causal relations: A *causes* B and B *causes* C. To satisfy both of the above assumptions, we make another two assumptions:

1) the simplest linguistic patterns indicating causality between events are based on *to/by*
2) the simplest syntactical structure for event representation and matching is *V+NP/Pro*, where *V* is a verb, *NP* is a noun phrase, *Pro* is a pronoun[1].

Based on the assumptions, the following two linguistic templates are used for causal chain extraction, where *V1* is a verb representing a cause-event and *V2* a verb representing the effect-event:

V1+NP1+to+V2+NP2
Example: *stabbed the guy to kill him; change his name to obtain ownership*

V2+NP2+by+V1+NP1
Example: *kill the guy by stabbing him; changed his name by dropping the prefix*

Two linguistic templates create the unification in a causal chain because *V2* can be considered as a representative of a cause the same way as *V1* but for the further step in causality.

It is important to note that the patterns should contain at least one affected object otherwise the causal link will be too general and abstract.

Similar patterns were suggested in VerbOcean (Chklovski & Patel, 2004) for enablement; for example, ***X****ed* * *by* ***Y****ing*, where * matches any single word. However, VerbOcean does not include a noun after a verb ***X*** in a causal pattern between verbs. The principal difference in our approach is that *opened the door by, opened the bottle by* and *opened the book by* assume a different ***Y***.

Extracting causal relations from text using linguistic templates was also explored in (Luo et al., 2016). However, the causal events considered in their work are limited by a single-term representation (for example, typhoon, sunrise, bacteria) and are linked into causal relations, not causal chains. Our work is more general in the sense that an event is described as a verb and a noun phrase, and causal relations are extended into causal chains.

IMPLEMENTATION

The flow chart in Figure 1 shows the general approach of causal chains extraction from text. The process is divided into two main steps. In the first step, causal relations are found by matching pre-defined linguistic templates. In the second step, causal chains are constructed by joining the relations using the process of unification and matching. The details of the approach are below:

- *Raw data*
  English Wikipedia articles are used as raw data. 100,000 articles were selected randomly for automatic extraction of causal chains.
- *Verb restriction*
  There are some phrases such as *allow workers to liberate themselves* or *have the power to remove the Head* that match the templates but are not causal relations. It happens

[1] For simplification, the structure will be reduced further to V+NP.

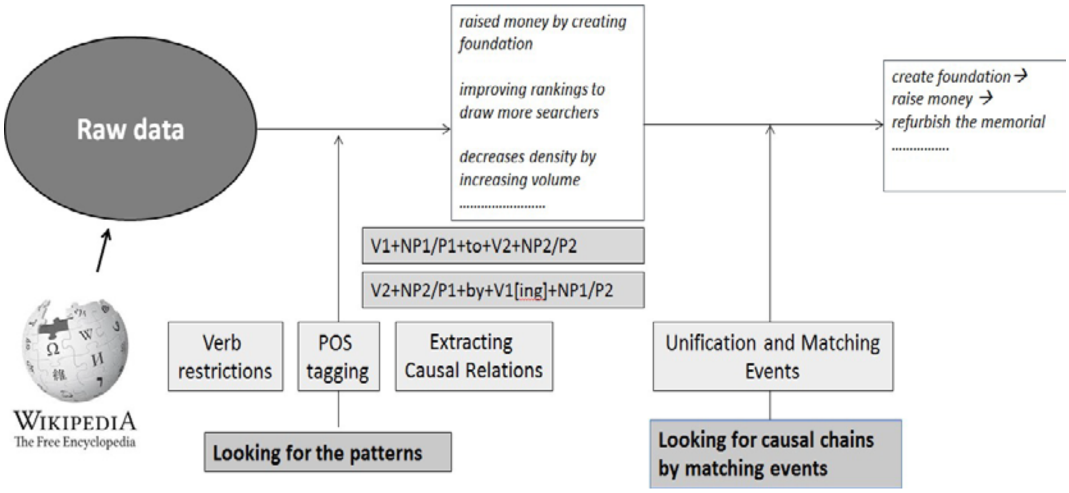

Figure 1. Flowchart of causal chains extraction from text

because of a modal verb (*allow, have*) that governs the other verb in the phrase. To eliminate the problem, this type of verbs is excluded (for example *appear, be, begin, consider)* from the extraction. In total, 85 modal verbs were excluded. This way, the number of false positives is minimized.

- *POS tagging*
  This is performed to allow for POS pattern matching in the next step. The POS tagger used is Averaged Perceptron (Collins, 2002) and the tagset is the Penn Treebank tagset (Santorini, 1990).

- *Extracting causal relations*
  After POS tagging is done, the causal relations are extracted by matching the two linguistic templates:
  V1+NP1+to+V2+NP2
  V2+NP2+by+V1+NP1

- *Unification and matching events*
  The procedure used to construct causal chains is as follows: for every causal relation, find a second causal relation such that the cause event of the second relation can be unified with the effect event of the first relation. In our current work, only causal chains consisting of two relations are constructed. Longer causal chains can be built in a similar manner and might be explored in future work. In order to unify the events, exact string matching is insufficient and will miss out many causal chains. The main reason is due to the fact that the same event can be represented by different phrasings. The event unification procedure only considers nouns, pronouns and verbs in the causal relations, ignoring other parts of speech that are not considered significant in event matching. Verbs of different inflections but of the same infinitive form are also considered as the same verb. In order to unify the events, exact string matching is insufficient and will miss out many causal chains. The main reason is the same event can be represented by different phrasings. The event unification procedure only considers nouns, pronouns and verbs in the causal relations, ignoring other parts of speech that are not considered significant in event matching. Verbs of different inflections but of the same infinitive form are also considered as the same verb.

RESULTS

Examples of the extracted causal chains are given in Table 1. The dependencies between the number of articles, number of sentences involved, causal relations extracted and causal chains extracted are shown in Table 2.

The graph in Figure 2 shows almost linear relationship between the number of causal relations extracted and the number of sentences. The graph in Figure 3 shows the relationship between the number of causal chains and the number of sentences. It follows an approximate quadratic dependency which is as expected since the number of causal chains is proportional to the number of pairs of causal relations.

As mentioned earlier, the extracted causal chains do not follow a linear structure. It is a net since an event can be caused by many events and can cause many events. For example, using the *by*-template for the event *commit suicide*, the following causes were extracted: *by dashing her head, by jumping off, by taking overdose, by cutting his throat, by throwing herself, by falling on sword, by gassing herself, by hanging herself, by tying a shoelace, by inhaling the exhaust gas, by leaping, by opening his veins, by slashing his, own throat, by stabbing*, etc. In turn, using the *to*-template for the same event as a cause we got the following effects: *avoid capture, defend the honour, avoid assimilation*, etc. Combination of the set of extracted causal relations with itself for the same event as an effect (1st causal relation) and a cause (2nd causal relation) accordingly allows making Cartesian multiplication for getting all possible sequences of causal events as separate chains.

Table 1. Examples of causal chains from Wikipedia

s/n	1st relation (A -> B)	2nd relation (B -> C)
1	avoid this situation by sleeping on the same schedule	avoids a situation to avoid unwanted emotions
2	kill herself by slitting her wrists	killing herself to save her husband
3	kill herself by slitting her wrists	killed herself to avoid capture
4	increasing stability by reducing size	stabilizes an emulsion by increasing its kinetic stability
5	pass legislation by negotiating with Colston	passed legislation to limit colonial trade
6	pass legislation by negotiating with Colston	pass remedial legislation to overrule the Manitoba Act
7	saved money by using competitive bidding	save his money to buy books
8	raised money by playing charity matches	raise money to fight Oregon Ballot Measure
9	transmits a signal by converting adenosine triphosphate	transmits a signal to adenylyl cyclase
10	accelerate the exhaust rearwards to provide thrust	provides the thrust to propel the spacecraft
11	accepted an offer to play six shows	played two shows to raise money
12	accepted an offer to take charge	took charge to revitalize the company
13	accepted corporate sponsorship to raise funds	raised funds to refurbish the memorial
14	accepted corporate sponsorship to raise funds	raises funds to benefit hospitalized children
15	accepted corporate sponsorship to raise funds	raises funds to provide restorative surgery
16	attempted a last-ditch stock sale to raise money	raised money to build a Presbyterian Church
17	borrowed money to build the dam	building a hydroelectric dam to flood Hetch Valley
18	borrowed money to build the dam	built a dam to catch the Temecula Creek water
19	built this ornate temple to raise money	raise money to fight Alzheimer
20	cancelled Franco-British plans to send troops	sent troops to stop the rebellion
21	cancelled Franco-British plans to send troops	send troops to rescue Lorca

Table 2. Causal chains versus Wikipedia articles

Articles	Sentences	Causal Relations	Causal Chains (Length 2)
1000	116394	758	4
5000	576568	3693	123
7500	889689	5718	249
10000	1178324	7546	534
15000	1760224	11312	1106
25000	2547205	16514	2632
40000	3747222	23942	5816
50000	4297336	26541	7351
60000	4778287	27689	8186
70000	5165499	28261	8692
80000	5561878	28990	9599
90000	6109638	31752	11701
100000	6773307	35703	14821

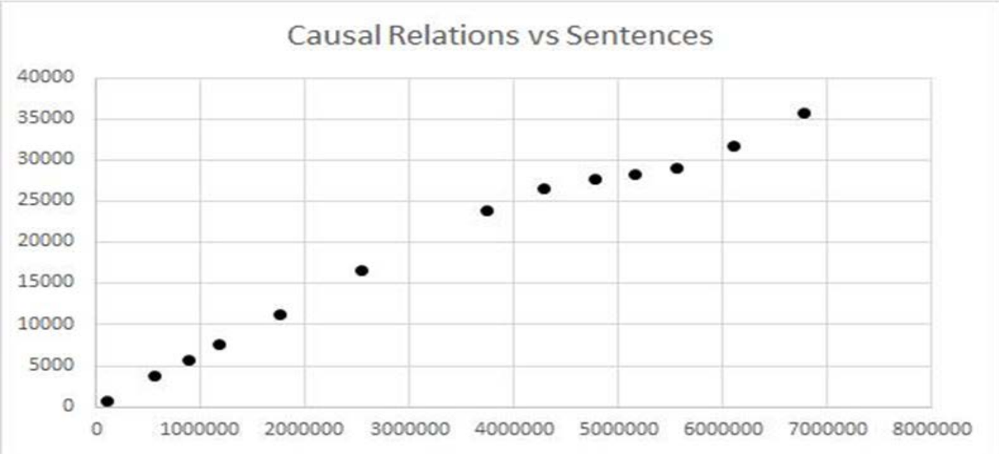

Figure 2. Relation between the number of causal relations and the number of sentences

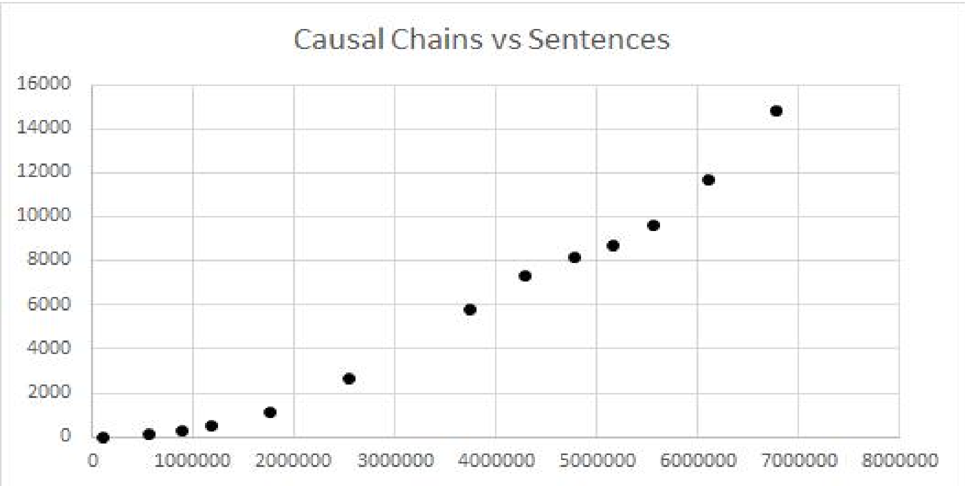

Figure 3. Relation between the number of chains and the number of sentences

Table 3. Method of aggregate evaluation

Deviation	Judges no.	E.g. of evaluation	Calculation
Max.2 points	5	5-3-4-4-5	Average: (5+3+4+4+5=4.2)
Max. 1 point	4	1-3-4-4-3	Average: (3+4+4+3=3.5)
0 points	3	5-2-2-2-4	No. that occurs three times: 2

Table 4. Chains before and after aggregate evaluation

No. of judges	No. of causal chains in the sample for judge evaluation	No. of causal chains in the sample after aggregate evaluation
5	100	70

Table 5. Evaluation for each type of deviation

Deviation	No. of aggregate causal chains	Average evaluation
Max. 2 points	40	2.93
Max. 1 point	45	2.69
0 points	45	2.78
All types	70	3.02

Table 6. Distribution of the scores

Score	No. of causal chains	Example of causal chains	
		First causal relation	Second causal relation
≥4	29	committed suicide by shooting herself	committed suicide to avoid capture
3--4	10	raised the money by asking York County	raised money to buy new instruments
<3	31	shelved plans to build the bridge	built a bridge to connect Parkersburg

EVALUATION

We got 14,821 causal chains (with length=2) from 100,000 English Wikipedia articles. With the naked eyes it is clear that most of the bad chains was caused by two reasons: (1) bad POS tagging (for example: *gives birth to twin sons* where *twin* is recognized as the verb; *reduce redundant work by taking advantage* with incomplete *NP*) and (2) bad verbs to be used for event representation (the verb *remain* in *remains the only foreign-born driver to win the race).*

Extraction of causal chains is a new task and there are yet no systematic evaluation measures. Our evaluation was based on a sample of 100 causal chains randomly taken from the ones we extracted.

Due to restrictions on event extraction *(V+NP)* and causal chain extraction (patterns with *by* and *to* only) we cannot extract causal chains comprehensively. As a result of that, recall of the extraction is not appropriate. Nevertheless, we do not make any restrictions on the verbs (except a very obvious one: *see ch.5 for details*) and our method assumes involvement the whole set of English verbs. The data we used might create some illusion that only causal chains related with human actions were extracted. In reality, it was caused by the nature of the data chosen (Wikipedia), not by the method applied to it.

The precision (effectiveness) of causal chain extraction was evaluated by five human judges. They were asked to assign each chain a number from 1 (very bad chain) to 5 (very good chain). They were instructed with the task formulation and the definitions of event, causal relations and causal chain.

After scoring by five judges, the method of aggregate evaluation was applied. If the deviation in evaluation between all five judges was no more than two points, the average was calculated. If the deviation between four judges was no more than one point, the average was calculated. If there was no deviation between three judges, their evaluation was accepted. If a score falls into more than one category (for example, 5-3-3-4-4 falls in two categories: five judges with deviation of two points and four judges with deviation of one point), the average will be calculated for each category and the maximum of those values is taken. In all other cases (for example, in the case of 5-3-2-4-1), the judge's evaluations were not counted, since they were too variable. See Table 3 for details.

Table 4 shows the number of causal chains (among 100 randomly chosen) that passed the aggregate evaluation.

Table 5 shows the number of aggregate chains for each type of deviation with the corresponding average mark. The table also shows the final evaluation for all types which has a deviation with average mark of 3.02. Based on that, one can conclude that the method we used allows extracting the chains that were estimated in average as *not bad*.

Table 6 provides the distribution of the scores (among 70 cases in total) with average evaluation of 4 and above, 3-4, and lowers than 3. Examples of chains for each cluster are provided. Our observations show the cluster with average evaluation lower than 3 contain lots of relations with bad POS tagging.

POTENTIAL APPLICATIONS AND FURTHER WORK

From the first attempt of method deployment that includes four modules—verb restriction, POS tagging, extracting causal relations, and unification and matching events, we can conclude that the method is rather effective. It can be tuned further for making evaluation results better. The only module which is the weakest point in the whole process—POS tagging, we cannot improve ourselves since we took it as ready-to-use from outside resources.

The causal chains that were derived from Wikipedia are universal in a sense that they are not related to any specific area. It is domain-independent since we use the whole set of English verbs for event extraction. Causal chains form a net that can be used in developing commonsense knowledge base, reasoning resources, and generally in discovering previously unknown relationships between entities/events. In particular, it can help when an event prediction or decision making is needed. For example, having *refurbishing the memorial* as a final goal to be accomplished, we can make decision how to do that by making a choice between multiple chains leading to the goal:

refurbishing the memorial by raising money;
raising money by creating the foundation OR
raising money by selling accessories OR
raising money by running the race, etc.

We hope to improve the current results by adding synonyms and by increasing 100,000 articles to 5.7 million in English Wikipedia. One of the primary goals for our future work will be to develop a formal procedure for evaluation.

REFERENCES

Asghar, N. (2016). Automatic extraction of causal relations from natural language texts: A comprehensive survey. *arXiv 1605.0789v1*. https://arxiv.org/abs/1605.07895

Bethard, S., Corvey, W., Klingenstein, S., & Martin, J.H. (2008). Building a corpus of temporal-causal structure. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC-2008), Marrokech Morocco, 28-30* May 2008 (pp. 908-915). http://www.lrec-conf.org/proceedings/lrec2008/pdf/229_paper.pdf

Chklovski, T., & Patel, P. (2004). VerbOcean: Mining the Web for fine-grained semantic verb relations. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2004), 25-26 Jul 2004, Barcelona, Spain* (pp. 33-44). Association for Computational Linguistics. https://www.aclweb.org/anthology/W04-3205.pdf

Collins, M. (2002). Discriminative training methods for hidden Markov models: Theory and experiments with perceptron algorithms. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2002), Philadelphia, USA, 6-7 July 2002* (pp. 1-8). https://doi.org/10.3115/1118693.1118694

de Silva, T.N., Zhibo, X., Rui, Z., & Kezhi, M. (2017). Causal relation identification using convolutional neural networks and knowledge based features. *World Academy of Science, Engineering and Technology International Journal of Computer and Systems Engineering*, *11*(6), 696-701.

Hendrickx, I., Kim, S.N., Zornitsa, Nakov, P., Seaghdha, D.O., Pado, S.,…Pennacchiotti, M., (2009). Semeval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals. In *Proceedings of the 5th International Workshop on Semantic Evaluations (ACL 2010), Uppsala, Sweden, 15-16 July 2010* (pp. 33-38). Association for Computational Linguistics. https://www.aclweb.org/anthology/S10-1006.pdf

Khoo, C.S.G., Chan, S., & Niu, Y. (2000). Extracting causal knowledge from a medical database using graphical patterns. In *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics (ACL 2000), Hong Kong, 3-6 October 2000* (pp. 336-343). Association for Computational Linguistics. https://doi.org/10.3115/1075218.1075261

Khoo, C.S.G., Kornfilt, J., Oddy, R.N., & Myaeng, S.H. (1998). Automatic extraction of cause-effect information from newspaper text without knowledge-based inferencing. *Literary and Linguistic Computing*, *13*(4), 177-186.

Luo, Z., Sha, Y., Zhu, K.Q., Hwang, S-W., & Wang, Z. (2016). Commonsense causal reasoning between short texts. In *Proceedings of 15th International Conference on the Principles of Knowledge Representation and Reasoning (KR2016), 25-26* April 2016, *Cape Town, South Africa* (pp. 421-430). AAAI Press. https://www.microsoft.com/en-us/research/wp-content/uploads/2016/07/12818-57567-1-PB.pdf

O'Gorman, T., Wright-Bettner, K., & Palmer, M. (2016). Richer event description: Integrating event coreference with temporal, causal and bridging annotation. In *Proceedings of 2nd Workshop on Computing News Storylines, (CNS 2016), 5*

November 2016, Austin, Texus, USA (pp. 44-56). https://www.aclweb.org/anthology/W16-5706.pdf

Pechsiri, C., & Sukharomana, R. (2017). Chain of causation determination from texts. *International Journal of Machine Learning and Computing*, *7*(5), 94-95. http://www.ijmlc.org/vol7/627-C44.pdf

Santorini, B. (1990). Part-of-speech tagging guidelines for the Penn Treebank Project (Technical Report No. MS-CIS-90-47). University of Pennsylvania, Department of Computer and Information Science. https://repository.upenn.edu/cgi/viewcontent.cgi?article=1603&context=cis_reports

Sawamaru, H., & Kobayashi, I. (2012). An approach to extraction of causal chain among events in multiple documents. In *Proceedings of the 6th International Conference on Soft Computing and Intelligent Systems and the 13th International Symposium on Advanced Intelligent Systems, 20-24 November 2012, Kobe, Japan* (pp. 1104-1108). https://doi.org/10.1109/SCIS-ISIS.2012.6505070

Sizov, G., & Ozturk, P. (2013). Automatic extraction of reasoning chains from textual reports. In *Proceedings of the TextGraphs-8 Workshop (at ACL 2013),18 October 2013, Seattle, Washington* (pp. 61-69). Association for Computational Linguistics. https://www.aclweb.org/anthology/W13-5009.pdf

Swanson, D.R., & Smalheiser, N.R. (1997). An interactive system for finding complementary literatures: A stimulus to scientific discovery. *Artificial Intelligence*, *91*(2), 183–203.

Swanson, D.R. (1987). Two medical literatures that are logically but not bibliographically connected. *Journal of the American Society for Information Science*, *38*(4), 228–233.