Leveraging graph-based semantic annotation for the identification of cause-effect relations

by Susetyo Bagas Bhaskoro

Leveraging graph-based semantic annotation for the identification of cause-effect relations

Susetyo Bagas Bhaskoro[1], Inkreswari Retno Hardini[2]
[1]Department of Manufacturing Automation and Mechatronics Engineering Technology,
Politeknik Manufaktur Bandung, Indonesia
[2]School of Electrical Engineering and Informatics, Institut Teknologi Bandung, Indonesia

Article Info	ABSTRACT
	This research is related to language article in Indonesia that discuss about causality relationship research used as public health surveillance information monitoring system. Utilization of this research is suitability of feature selection, phrase annotation, paragraph annotation, medical element annotation and graph-based semantic annotation. Evaluation of system performance is done by intrinsic approach using the naive bayes multinomial method. The results obtained sequentially for recall, precision and f-measure are 0.924, 0.905, and 0.910.

Corresponding Author:

Susetyo Bagas Bhaskoro,
Department of Manufacturing Automation and Mechatronics Engineering Technology,
Politeknik Manufaktur Bandung,
Kanayakan Street, No. 21, Bandung, Indonesia.
Email: bagas@polman-bandung.ac.id

1. INTRODUCTION

Research on natural language processing is increasing because it is influenced by the availability of information that always increases. One example is natural language processing in the health domain whose source of information is not only derived from the patient's medical record, but through community writing participation shared using internet media [1]. Not a few practitioners, communities and even health service units seek medical information and utilize information on the internet for certain purposes. One of the characteristics of information seeking in the health domain is that the search for health information will end when information on cause and effect has been found that discusses the root causes of a medical condition.

$$f(x, y) = (x) \rightarrow (y); \text{ or } f(x, y) = if (x) \text{ then } (y) \tag{1}$$

Meanwhile, having a large collection of medical information from various sources brings challenges for practitioners and the public to find the information they need in a fast time. Getting new knowledge from medical articles automatically becomes a separate problem because the sentence usually written on medical articles has the characteristics of multiple sentences [2], namely a long article that usually has several paragraphs, the explanation consists of background then presents the problem and provides solutions to the core part of the article and ends with conclusions or messages that are considered important [3, 4]. There is a possibility that in an article, the position of the causal relationship pattern allows to be in a different paragraph, meaning the sentence because it can be written at the beginning of paragraphs and sentences due to being explained at the end of the paragraph, and vice versa.

$$f(x, y, z) = \begin{array}{l} x \text{ is the potential for development } y, \\ x \text{ cause } z, \text{ and } y \text{ is the main cause of } z \end{array} \qquad (2)$$

The purpose of this research is to optimize the use of annotations to do a summary of an online medical article in which it is able to display the meaning of cause and meaning as a result of which the two meanings have a mutually explaining relationship. Overall, this study aims to contribute as follows: (a) proposed phrase annotations used to identify implicit meanings in a sentence; (b) proposed paragraph annotations are used to carry out the core classification of the discussion in each paragraph in the medical article; (c) proposed annotations of medical elements are used to transform natural language into medical annotations; (d) proposed graph-based semantic annotations are used to identify cause-effect relationships through a template-based.

2. RESEARCH METHODOLOGY

The method used as a reference is Atkinson's research [5]. The difference with the current research lies in the summary section of the text, the transformation of medical elements, semantic annotations and paragraph annotations shown in Figure 1 (blocks given dark background colors). The current research is using an online medical article dataset for public health surveillance. The aim is to display information on causes and consequences, as well as the mitigation process from the occurrence of these health cases. The challenge in the current research is to determine the medical elements that match the characteristics of the text in Indonesian medical articles.
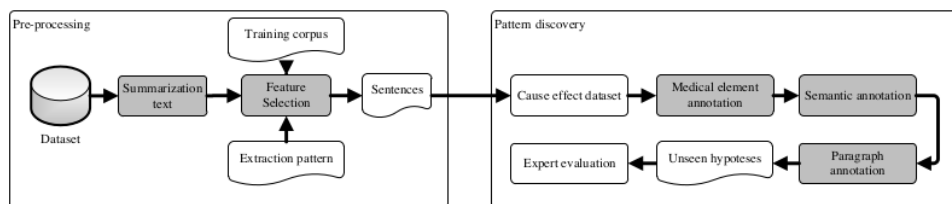

Figure 1. Process approach for the identification of causal relationship in multiple sentences

2.1. Summary of extractive text

The principle of using text summaries in the current research is to complete the stages of feature selection needed by Atkinson [5], in addition summary techniques are used to narrow the problem search space and the characteristic used in the text summary phase [6].

2.2. Transforming natural languages into medical element annotations

Annotation is a technique used to give notes to each sentence. This solution is used to find important sentences discussed in a medical text. The fact is that the results of annotations actually have a lot of connection with other annotations that form new meanings [7]. Here is an annotation comparison proposed by several other researchers. Atkinson [5] proposed annotation are substance, effect, symptom, disease, and body part. Mihăilă [7] proposed annotation are drug, physical stimulation, symptom, inhibition, and diagnosis. Byrd [8] utilizing the information occurs on Twitter for public health surveillance with annotations used are sentiment analysis and geocoding. Liang Wu [9] using twitter to find out information about the reactions of drugs consumed with the annotations used are drug name, brand name, prescribe for, effect. Yanging Ji [10] utilizing health articles by using annotations drug name, ICD Code, and symptom. Yepes [11] used annotations for surveillance problems. The proposed annotations are disease, symptoms, pharmacological and location.

The obstacle of the several annotations that have been proposed is that it has not been able to display a dependency relationship that explains the meaning of cause and effect, therefore in the current study the semantic annotation is proposed to relate the meaning of each annotation that has a connection pattern with other annotations according to medical information needs. Table 1 is the proposal to develop natural language into the LPpAJSFPePnGOD medical element by considering the needs of the dataset taken from online articles, namely handling in the form of statements from resource persons, facilities, and recovery that have been carried out as indicators of mitigation.

Table 1. Proposed `LPpAJSFPePnGOD` for medical element annotations

Medical Entities	Description
LOKASI (L) In English LOCATION	Used to extract location and place names about events or events being discussed The rule is `Location (x, y)` → `probability of location`.
POPULASI PENYAKIT (Pp) In English DISEASE POPULATION	Used to perform extraction issues and problems of the disease. Get the problem issues with the approach of the method of biological named entity recognition. The rule is `Disease population (x, y)` → `biological probability named entity recognition (biological NER)`.
AKIBAT (A) In English EFFECT	Used to perform the extraction impact by the subject as specific disease populations. The rule is `Effect (x, y)` → `disease_population (x), location (y)`.
JUMLAH KEJADIAN (J) In English SUM OF EVENTS	Used to perform the extraction of statistical count in numeral from the population of certain diseases. The rule is `Sum_of_events (x, y)` → `effect (x, y)`.
SEBAB (S) In English CAUSE	Used to perform the extraction of the main causes. The method for obtaining a cause is to calculate the word emergence statistics on online articles and classification of certain disease populations. The rules are `Cause_1 (x, y)` → `disease_population (x, y)`; `Cause_2 (x, y)` → `sum_of_events (x, y)`; `Cause_3 (x, y)` → `sum_of_events(x), location(y)`.
FASILITAS (F) In English FACILITY	Used to name extraction institutions that provide treatment based on the consequences of certain disease populations. The rules are `Facility_1 (x, y)` → `disease_population (x), sum_of_events(y)`; `Facility_2 (x, y)` → `effect(x, y)`; `Facility_3 (x, y)` → `cause_1(x, y)`.
PEMULIHAN (Pe) In English RECOVERY	Used to solution extraction prepared and carried out from facilities in certain locations. The rules are `Restitution_1 (x, y)` → `facility_1 (x, y)`; `Restitution_2 (x, y)` → `facility(x), location(y)`.
PENANGGUNG JAWAB (Pn) In English RESPONSIBLE PERSON	Used to subject extraction which is responsible for certain disease population problems and recovery carried out by facilities in certain locations. The rules are `Responsible_person_1 (x,y)` → `facility_1 (x,y)`; `Responsible_person_2 (x,y)` → `cause_3 (x,y)`; `responsible_person (x,y)` → `effect(x), facility(y)`
GEJALA (G) In English SYMPTOMS	Used to extraction part of the body are diseased or in pain. The rules are `Symptoms (x,y)` → `disease_population (x,y)`; `Symptoms (x,y)` → `disease_population (x,y)`
OBAT (O) In English DRUGS	Used to information extraction drug or vaccine, as one method of recovery. The rule is `Drugs (x,y)` → `cause_1(x), recovery(y)`
DIAGNOSA (D) In English DIAGNOSE	Used to extraction of information related to medical statement about the disease. The rule is `diagnosis (x,y)` → `disease_population (x), symptoms(y)`

2.3. Paragraph annotation

Paragraph annotations are needed because the important messages conveyed in online medical articles are in unpredictable paragraphs or each paragraph discusses the different core stories. Therefore the main task of paragraph annotations is to produce the essence of word categories on medical content consisting of core paragraphs, supporting paragraphs and conclusion paragraphs. List the paragraph position annotations as found in Table 2. If the core story of each paragraph has been obtained, then the pattern of the graph pair from the paragraph classification can be determined based on the rules of pattern in Figure 2.
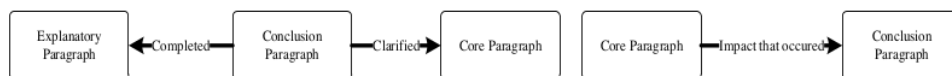

Figure 2. Connection of meaning based on annotations of paragraph positions

Table 2. Proposed paragraph annotation

Paragraph Annotation	Characteristics

Core Paragraph	Sentences that have frequency of occurrence more than the other sentences and describe the event being discussed; The core sentence in it has a meaningful effect; The sentence describes the fact of an event discussed. Elements related to the core paragraph as follows: Amount in numbers: the number of events, evidence of the incident; effects; pharmacy; diagnosis; Signs in the body related to the symptoms. Techniques for use calculating the frequency of the number of words and do a ranking based on a threshold value of the article.
	
	`Loop article, sum of[]`
	`{element_paragraf_core,`
	`positions_paragraph_core = n`
	`n = n+1}`
	`End loop`

Explanatory Paragraph	Sentences that are identical to information on the cause of an event; Sentence describing the subject discussed in the article; Sentence that defines the name of the disease that are discussed in this article; Sentence that describes the description of the place. The related of elements to explanatory paragraphs are: Location: country, city, district, address; Disease; cause; subject: person, object; Comparison sentence. The technique for using is calculate the occurrences frequency of words number and rank based on the threshold value of the article:
	
	`Loop article, sum of[]`
	`{element_paragraph_core,`
	`position_paragraph_core = n`
	`n = n+1}`
	`End loop`

Conclusion Paragraph	Sentence that describes the mitigation of the ongoing activity; The sentence that defines the facilities discussed in the article. The related of Elements to the conclusion paragraph are: Mitigation; Facilities provided. The technique for using is calculate the occurrences frequency of words number and rank based on the threshold value of the article:
	
	`Loop article, sum of[]`
	`{element_paragraph_core,`
	`position_paragraph_core = n`
	`n = n+1}`
	`End loop`

2.4. Feature selection

Selection of features tailored to the characteristics of the text in complex medical articles and the meaning of medical sentences. Analysis of the pre-processing features used in medical articles as in the current research is better not to use this pre-processing stage, compared to using it because the results are ambiguous [12]. Figure 3 is the pre-processing stage in question.
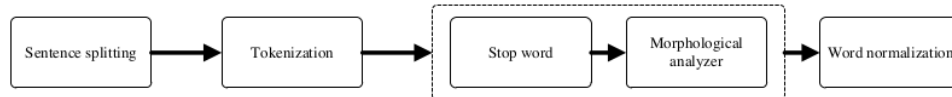

Figure 3. Stages of pre-processing

The first stage is sentence splitting, which is to separate sentences until the delimiter '.' But not all delimiter signs are the final marker of a sentence. For example, like 'dr.' The word is not the end of a sentence, but a profession title. The second stage is tokenization, which is to separate each word from a sentence. Examples of problems such as the sentence `<rumah sakit in English is hospital>`. The sentence is one term not `<home>` and `<sick>` or `<care>`, as well as `<Ministry of Health>`. The sentence is one term not `<ministry>` and `<health>`.

The last step is the morphological analyzer, which converts a word into a basic word. An example is `<immunized>`. The word has a suffix in bahasa language `(di-)`, `(sasi-)` and the basic word is `<immune>`. Immunized has the meaning of being given a vaccine for the immune system, while the immune system is the immune system. The analysis of the other feature selection is `n-gram` as seen in the following equation:

$$P(k_1, k_2, \ldots, k_n) = P(k_1)P(k_2|k_1) \ldots P(k_n \mid k_1 \ldots k_{n-1}) \qquad (3)$$

Leveraging graph-based semantic annotation for the identification of... (Susetyo Bagas Bhaskoro)

The use of n-gram in this research uses trigram, and the equation is as follows:

$$P(k_n|k_1,\ldots,k_{n-1}) = P(k_n \mid k_{n-2}\ldots k_{n-1}) \tag{4}$$

The previous equation can be written as follows:

$$P(k_1,k_2,\ldots,k_n) = P(k_1)P(k_2|k_1)\prod_{i=3}^{n} P(k_i \mid k_{i-2}\ldots k_{i-1}) \tag{5}$$

The next step is to adopt the use of semantic analysis as done by Byrd [8] into phrase annotations to identify positive or negative meanings. Opinion analysis uses analysis of `positive sentences (+)` to represent the meaning of causes, and analysis of `negative sentences (-)` to represent the meaning of consequences. So, the syntactic results obtained are as follows:

```
In Bahasa                                    In English
<Demam Berdarah Dengeu (DBD)> / NN           <Dengue Hemorrhagic Fever (DHF)> / NN
kembali / SC                                 return / SC
<memakan korban jiwa> / VB + (NEG)           <casualties> / VB + (NEG)
di / IN                                      in / IN
Kota Pekanbaru. / NNP                        Pekanbaru City. / NNP
Seorang bocah perempuan / NNP                A girl / NNP
berusia / CD                                 old / CD
tujuh tahun / CD                             seven years / CD
bernama / DT                                 named / DT
Nursabrina / NNP                             Nursabrina / NNP
<meninggal dunia> / VB + (NEG)               <dies> / VB + (NEG)
akibat / SC                                  effect / SC
<gigitan nyamuk Aides Aigepty>. / VB + (NEG) <Aides Aigepty's mosquito bites>. / VB + (NEG)
```

The results of the exploration of the sentence above show the sentence `<taking casualties> / VB + (NEG)` and `<death> / VB + (NEG)` having negative meanings.

2.5. Semantic annotation

Semantic annotation technique is a technique that expresses the connectedness of each word in an article so as to produce new information that can be understood by its meaning. This step is to adopt the use of semantic annotation analysis as done by several researchers [13-16]. The current research uses a hybrid method, namely a combination of pattern-based and machine-based learning to build semantic relationships and get interpretations that are appropriate to medical needs. Each node is defined as having an association with the other annotation node. Although the image looks separate, but each annotation node will look for other annotation nodes that have the same information.

2.6. Analysis of usage annotation

The analyses needed to design the tests that will be carried out research currently carried out implemented in the health surveillance system are:
- Understand the use of online medical articles to support public health surveillance.
- Determine feature selection that is most suitable for use in the characteristics of complex texts from online medical articles.
- Transform natural language into annotations of medical elements.
- Propose the transformation of medical texts into paragraph annotations: core paragraphs, supporting paragraphs and conclusion paragraphs;
- Testing for the dependency sentences → medical element annotations (`LPpAJSFPePnGOD`) and graph-based of semantic annotation;
- Testing system performance evaluation that will be compared with other studies.

3. TESTING

The tests are divided into three discussions, namely `tagging (verb (VB) + noun (NN)) - phrase annotation, multi feature selection and dependency relation`.

3.1. Parsing `(VB + NN) + phrase annotation + opinion analysis`

The discussion of the test begins with the analysis of opinions using `positive (+)` sentences to represent the meaning `<cause>`, and `negative (-)` sentences to represent the meaning `<effect>`. Giving

a positive class label if in the sentence `<verb>` and `<noun>` have a positive meaning, while labeling a negative class if in that sentence `<verb>` and `<noun>` have negative meanings. The experiment conducted a comparison of several classification techniques [17], namely (a) testing using lexical; (b) using the naïve bayes method; and (c) using the SVM method. Table 3 is the result of a comparison of explicit and implicit data modeling analysis to identify patterns of causal relationships with other researchers, as follows:

Table 3. Comparison of testing identification of cause relations

Researcher	Feature proposed	Accuracy
Ittoo [18]	`<verb> Wikipedia + Lexical.`	0.650
Rink [19]	`Word+POS+Parse graphs + word+POS+Parse+Stem`	0.381
Riaz [3]	`Tagging Verb + Noun`	0.807
The proposed model (current study)	`Tagging Verb + Noun + Tagging phrase + LB`	0.650
	`Tagging Verb + Noun + Tagging phrase + NB`	0.660
	`Tagging Verb + Noun + Tagging phrase + SVM`	0.890

3.2. Multi-feature selection

This test uses a summary technique to narrow down the problem search space, then combines with the semantic pattern modeling method and the proposed annotation to find important sentences that have a meaningful cause-effect relationship. The following is the pseudo code for the feature selection, weighting and n-gram used in the summary system. The next test is to combine `summary + annotation of medical elements + phrase tagging + opinion analysis`. The aim is to get a summary that is more in line with the information search needs of the medical domain and has a pattern of causal relationships. Because this research is a subjective classifier, the learning model must be adjusted based on the individual preferences of the labeler or an expert and the label results may vary. Therefore, research activities explore features by means of classification. The following is one example of a medical article taken from online media and used as an experimental dataset in the current research. In this study using a dataset of `500` medical articles online and when extracted it had sentences of `10,176` sentences. The method used as a comparison for the classification of `summaries + annotations of medical elements + phrase tagging + analysis of opinion` is naïve Bayes.

The current research is able to answer from previous studies when the output of more than one entity then the meaning becomes ambiguous as mentioned in Khoo's research [20–22] and Park [23]. This study even though the results of medical elements obtained more than one entity, then the technique is by doing a ranking to get the medical element into one entity using the calculation of the frequency of occurrence of words. The results obtained using a `summary approach + annotation of medical elements + phrase tagging + opinion analysis` to find important information in medical text articles that have causal meanings as found in Table 4.

Pseudo code combination of feature selection, n-Best and weighting

```
1   input      : document as d, feature selection as  fs weighting  as w
2   output     : summary
    Query      : {Title, noun, statistics on the number of occurrences of words, length of
3   words, statistics on the number of occurrences of words and nouns, number of
    occurrences of words and titles}
4   Weighting (w)        : {tf; tf-idf; tf-idf-df}
5   Parameter values: (0.4;0.6;0.7;0.8)
6       di ← get document i
7       fs ← choosing query
8       w ← weighting fittest
9   Query combination ← get output values fs and w
10  for each (document fittest != null)
11      Sentence to i (si) = the sentence originating from the document i
12  end for each
13  if(query combination fittest != null )
14  si = each (sentence i) compared to query
15  word = word detection from si compared to query
16  statistic = get a statistic from (sentence i)
17      w=tf (si to sn)  x idf (si to sn)
18   mmr (si)=π x Sim_1 (si,d)- (1-π)x Sim_2    (si)
19  end if
20  summary ← if the sentence i is the value is greater than the threshold, then
    the sentence i is the summary result
```

$$tf\,(t,d) = \frac{N(t)}{\sum_{i=1}^{T} N(i)} \tag{6}$$

Table 4. Comparison for testing the identification of cause-effect relations

Researcher	Features	Recall	Precision	f1
Riaz [24]	Tagging Verb + Noun	0.807	–	–
Mihăilă [7]	BioCause: drug, physical stimulation, symptom, inhibition, dan diagnosis.	–	–	0.873
Yepes [11]	Proposed annotations are disease symptoms, pharmacological and location. disease, symptoms, pharmacological dan location.	0.214	0.938	–
The proposed model (current study)	Summary + annotations medical elements + phrase tagging + opinion analysis	0.924	0.905	0.910

3.3. Dependency relation

At this stage the discussion adds a paragraph annotation and semantic annotation approach to display new knowledge, so that readers can easily understand the summary results in a medical text article that displays important information and has a meaningful cause-effect relationship [25]. The technique for dependency relations is a combination of rule-based and statistical based. Rule-based techniques are used when the annotation resulting from the classification of paragraph sentences in medical articles must be paired with the classification of other paragraph sentences as shown in Figure 2. Interrelation of meaning based on paragraph position annotations and Figure 4. The following is a rule-based pseudo code.
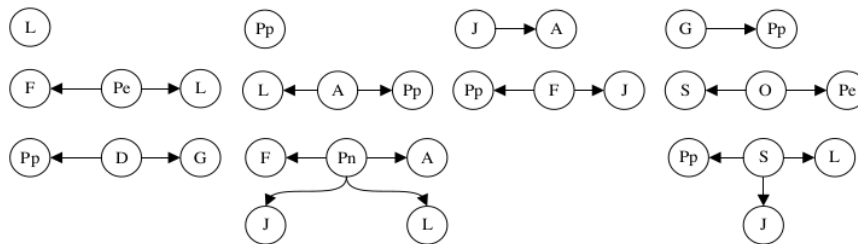

Figure 4. Graph-based of semantic annotation

Pseudo code-based word linkage

```
Cause pattern identification:
Cause entity == [Disease population]
The word position is in the explanatory paragraph and core paragraph
Cause is words that have confidence (association) = 0.2 from the ratio of all words (in the
form of [noun] + [noun]).
Words positions there is in the explanatory paragraph and core paragraph.

Effect pattern identification
effect entity = [words has a negative meaning]
effect entity = words has a negative meaning + nominal + medical element [sum of events]
effect entity = words has a negative meaning + medical meaning [facilities]
effect entity = words has a negative meaning and confidence value (association) = 0.2
Words positions there is in the explanatory paragraph and core paragraph.
```

Machine learning-based are used when conducting annotation classification of medical elements and paragraph annotations. The following is a statistical pseudo code:

Pseudo code word-based statistical connectivity

```
Training data
Insert two sentences that have a relationship and each other's has a class
classification results
    1.  sentences A = [cause]
    2.  sentences B = [facility]
    3.  sentences C = [effect]
    4.  sentences D = [sum of event]
    5.  sentences E = [responsible person]
    6.  sentences F = [population disease]
    7.  sentences G = [effect]
    8.  sentences A and sentences G = [causal]
```

Pseudo code word-based statistical connectivity

```
Getting words connectivity by several techniques
Technique A
Using paragraph classes
Explanatory paragraph - core paragraph
Explanatory paragraph - supporting paragraph
Core paragraph - explanatory paragraph
Finding a derivative word or meaning of the same words that often appear
The word is repeated at the end of the sentence
Technique B
Medical element of entity = words has confidence value (association) = 0.2
```

The technique used in the current research is to build new knowledge by combining rule-based and statistical-based dependency relations to display summaries of important information aided by the formation of new knowledge patterns as shown in Table 5.

Table 5. Knowledge discovery

Rule	Pattern
R1	Di [LOKASI] bagian dari [LOKASI] Knowledge pattern in English: At [LOCATION] part of [LOCATION]
R2	Terdapat kejadian sejumlah [JUMLAH KEJADIAN] dari [POPULASI PENYAKIT] Knowledge pattern in English: Sum of events [SUM OF EVENTS] from [DISEASE POPULATION]
R3	Mengakibatkan [AKIBAT] Knowledge pattern in English: Result in [EFFECT]
R4	[AKIBAT] tersebut dikarenakan [SEBAB] dan terjangkit [POPULASI PENYAKIT] Knowledge pattern in English: [EFFECT] because of [CAUSE] and infected [DISEASE POPULATION]
R5	Menurut [PENANGGUNG JAWAB] dari [LOKASI], penanganannya dengan cara [PEMULIHAN] yang diberikan oleh [FASILITAS] Knowledge pattern in English: By [RESPONSIBLE PERSON] from [LOCATION], handling by means [RECOVERY] given by [FACILITIES]
R6	Di [LOKASI] bagian dari [LOKASI] Knowledge pattern in English: At [LOCATION] part of [LOCATION]
R7	Terdapat kejadian sejumlah [JUMLAH KEJADIAN] dari [POPULASI PENYAKIT] Knowledge pattern in English: Sum of events [SUM OF EVENTS] from [DISEASE POPULATION]
R8	Mengakibatkan [AKIBAT] di bagian [GEJALA] yang merupakan bagian dari [GEJALA] Knowledge pattern in English: Result in [EFFECT] in the [SYMPTOM] which is part of [SYMPTOM]
R9	[AKIBAT] tersebut dikarenakan [SEBAB] dan terjangkit [POPULASI PENYAKIT] Knowledge pattern in English: [EFFECT] because of [CAUSE] and infected [DISEASE POPULATION]
R10	Menurut [PENANGGUNG JAWAB] dari [LOKASI], penanganannya dengan cara [PEMULIHAN] yang diberikan oleh [FASILITAS] Knowledge pattern in English: By [RESPONSIBLE PERSON] from [LOCATION], handling by means [RECOVERY] given by [FACILITIES]
R11	Obat yang dikonsumsi [OBAT] karena didiagnosa yang diberikan ahli adalah [DIAGNOSA] Knowledge pattern in English: Medicine consumed [DRUG] because being diagnosed by an expert is [DIAGNOSIS]

The output from the mapping of new knowledge formation patterns with the results of the classification of paragraph annotations in summary form looks like the following. In this research take an example of data testing using online articles [26]:

```
Di <Desa Sigedong Kabupaten Temanggung> terdapat seorang meninggal dunia karena
Diare. Sebagai penanggulangannya Dinkes Temanggung telah mendirikan Posko dan
melakukan sosialisasi untuk menerapkan hidup bersih dan sehat.
In English:
In Sigedong Village, Temanggung Regency, one person died of Diarrhea. As a response,
the Temanggung Health Office has established a Command Post and conducted
socialization to implement a clean and healthy life.
```

The testing scenarios carried out for the stages of dependency relations are more focused on extrinsic evaluation which is divided into two categories, namely in Table 6 which is used for testing based on the suitability of the summary results, appropriate summary outputs, and inappropriate outputs. Remarks for Table 6: (a) summary by an expert; (b) summary by the system; (c) the appropriate output (d) the output is not appropriate. Based on Table 6 seen in point (a), an expert determines the output that should be generated by the system, then in point (b) is the output generated by the system using a combination of summary features + annotations of medical elements + phrase tagging + opinion analysis. In point (c) is a comparison between the output determined by an expert, with the output produced by the system. Pattern (d) is a non-matching comparison of output between points (a) and points (b). The results of the average success of the system to produce output in accordance with an expert's decision are 0.857. Figure 5 is a comparison of the performance output produced by the system with a decision previously made by an expert. The results obtained on the system are said to be good when approaching the value of '1', meaning that the output on the system has a degree of similarity to the decisions made by an expert.

Meanwhile, Table 7 is used for evaluating an expert's evaluation of the outputs generated from the system integration semantic template method with a health surveillance system. Based on Table 7 the results obtained from extrinsic evaluation are 0.720 with the decision to agree. The assessment given by an expert into a system between 0-100. The decision agreed to illustrate that the summary method + annotation of medical elements + phrase tagging + opinion analysis, the system is able to produce results that resemble the way an expert does inference.

Table 6. Testing annotation fitness

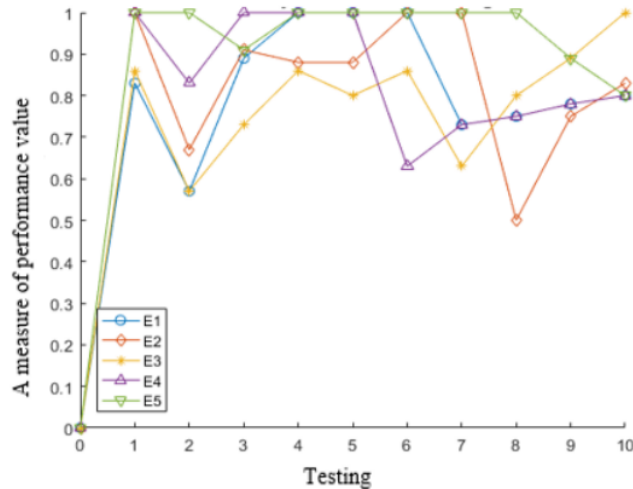| #	E1				E2				E3				E4				E5			
	a	b	c	d	a	b	c	d	a	b	c	d	a	b	c	d	a	b	c	d
1	6	13	5	1	8	13	8	0	7	13	6	1	5	13	5	0	9	13	9	0
2	7	7	4	3	6	7	4	2	7	7	4	3	6	7	5	1	7	7	7	0
3	9	21	8	1	11	21	10	1	11	21	8	3	12	21	12	0	11	21	10	1
4	6	11	6	0	8	11	7	1	7	11	6	1	9	11	9	0	7	11	7	0
5	6	14	6	0	8	14	7	1	5	14	4	1	10	14	10	0	8	14	8	0
6	5	11	5	0	6	11	6	0	7	11	6	1	8	11	5	3	7	11	7	0
7	11	8	8	3	8	8	8	0	8	8	5	3	11	8	8	3	8	8	8	0
8	4	5	3	1	4	5	2	2	5	5	4	1	4	5	3	1	4	5	4	0
9	9	10	7	2	8	10	6	2	9	10	8	1	9	10	7	2	9	10	8	1
10	5	7	4	1	6	7	5	1	6	7	6	0	5	7	4	1	5	7	4	1

Figure 5. Comparison of system performance output with decisions of an expert on a system

Table 7. Assessment of an expert assessment

Testing	E1	E2	E3	E4	E5
1	0.80	0.70	0.70	0.80	0.70
2	0.80	0.70	0.80	0.80	0.80
3	0.70	0.80	0.80	0.70	0.70
4	0.70	0.80	0.80	0.70	0.80
5	0.80	0.80	0.80	0.70	0.70
6	0.70	0.70	0.80	0.70	0.70
7	0.70	0.70	0.70	0.70	0.70
8	0.80	0.80	0.80	0.80	0.70
9	0.80	0.80	0.80	0.80	0.80
10	0.70	0.80	0.80	0.80	0.80

4. CONCLUSION

Generated semantic template method that has been successfully implemented on health domain is public health surveillance system. During implementation, there are several proposals have been generated, i.e. natural language transformation into medical element annotation (LPpAJSFPePnGOD) to identify causal relationship pattern, paragraph annotation pattern to classify sentence position on medical article paragraph and build a semantic relationship pattern for semantic annotation.

REFERENCES

[1]   S. Akbar, L. Slaughter, and Ø. Nytrøø, "Collecting health related text from patient health writings," in *The 2nd International Conference on Computer and Automation Engineering*, vol. 1, pp. 15-19, 2010.

[2]   C. Hidey and K. Mckeown, "Identifying Causal Relations Using Parallel Wikipedia Articles," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pp. 1424–1433, 2016.

[3]   M. Riaz and R. Girju, "Recognizing Causality in Verb-Noun Pairs via Noun and Verb Semantics," in *EACL 2014 Workshop on Computational Approaches to Causality in Language (CAtoCL)*, no. 1, pp. 48-57, 2014.

[4]   K. Hanawa, A. Sasaki, N. Okazaki, and K. Inui, "A Crowdsourcing Approach for Annotating Causal Relation Instances in Wikipedia," in *The 31st Pacific Asia Conference on Language, Information and Computation*, pp. 336-345, 2016.

[5]   J. Atkinson and A. Rivas, "Discovering novel causal patterns from biomedical natural-language texts using Bayesian nets," *IEEE Trans. Inf. Technol. Biomed*, vol. 12, no. 6, pp. 714–722, 2008.

[6]   S. H. S. Susetyo Bagas Bhaskoro, Saiful Akbar, "Enhancing Performance in Medical Articles Summarization with Multi-feature Selection," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 8, no. 4, pp. 2299-2309, 2018.

[7]   C. Mihăilă, T. Ohta, S. Pyysalo, and S. Ananiadou, "BioCause: Annotating and analysing causality in the biomedical domain," *BMC Bioinformatics*, vol. 14, no. 2, pp. 1-18, 2013.

[8]   K. Byrd, A. Mansurov, and O. Baysal, "Mining Twitter data for influenza detection and surveillance," *Proc. Int. Work. Softw. Eng. Healthc. Syst. - SEHS '16*, pp. 43–49, 2016.

[9]   L. Wu, T.-S. Moh, and N. Khuri, "Twitter opinion mining for adverse drug reactions," *2015 IEEE Int. Conf. Big Data (Big Data)*, pp. 1570-1574, 2015.

[10] Y. Ji *et al.*, "A potential causal association mining algorithm for screening adverse drug reactions in postmarketing surveillance," *IEEE Trans. Inf. Technol. Biomed*, vol. 15, no. 3, pp. 428-437, 2011.

[11] A. J. Yepes, A. MacKinlay, and B. Han, "Investigating Public Health Surveillance using Twitter," *Proc. 2015 Work. Biomed. Nat. Lang. Process. (BioNLP 2015), Beijing, China, July 30, 2015*, no. BioNLP, pp. 164–170, 2015.

[12] S. B. Bhaskoro, S. Akbar, and S. H. Supangkat, "Identification of Causal Pattern using Opinion Analysis in Indonesian Medical Texts," in *2015 International Conference on Information Technology Systems and Innovation (ICITSI)*, pp. 1–7, 2015.

[13] S.-L. Hsieh, W.-Y. Chang, C.-H. Chen, and Y.-C. Weng, "Semantic similarity measures in the biomedical domain by leveraging a web search engine.," *IEEE J. Biomed. Heal. informatics*, vol. 17, no. 4, pp. 853-861, 2013.

[14] N. Mostafazadeh, A. Grealish, N. Chambers, J. Allen, and L. Vanderwende, "CaTeRS: Causal and Temporal Relation Scheme for Semantic Annotation of Event Structures," *Proceedings of the 4th Workshop on Events: Definition, Detection, Coreference, and Representation*, pp. 51–61, 2016.

[15] A. Kogilavani and P. Balasubramanie, "Sentence Annotation based Enhanced Semantic Summary Generation from Multiple Documents," *Journal, Am. Sci. Appl. Publ. Sci.*, vol. 9, no. 7, pp. 1063–1070, 2012.

[16] K. Okoye, A. R. H. Tawil, U. Naeem, R. Bashroush, and E. Lamine, "A semantic rule-based approach supported by process mining for personalised adaptive learning," *Procedia Comput. Sci*, vol. 37, pp. 203–210, 2014.

[17] S. B. Bhaskoro, S. Akbar, and S. H. Supangkat, "Extracting Important Sentences for Public Health Surveillance Information from Indonesian Medical Articles," in *The International Conference on ICT for Smart Society (ICISS)*, pp. 1-7, 2017.

[18] A. Ittoo and G. Bouma, "Extracting Explicit and Implicit Causal Relations from sparse, Domains-Specific Texts," in *NLDB'11 Proceedings of the 16th international conference on Natural language processing and information systems*, pp. 52–63, 2011.

[19] B. Rink, C. A. Bejan, and S. Harabagiu, "Learning Textual Graph Patterns to Detect Causal Event Relations," in *the Twenty-Third International Florida Artificial Intelligence Research Society Conference (FLAIRS 2010) Learning*, pp. 265–270, 2010.

[20] C. Khoo, S. Chan, Y. Niu, and A. Ang, "A method for extracting causal knowledge from textual databases," *Singapore J. Libr. Inf. Manag.*, vol. 28, pp. 48–63, 1999.

[21] C. Khoo, S. Chan, and Y. Niu, "Extracting causal knowledge from a medical database using graphical patterns," *Proc. 38th Annu. Meet. Assoc. Comput. Linguist.*, pp. 336–343, 2000.

[22] C. Khoo, "Automatic Extraction of Cause-Effect Information From Newspaper Text Without Knowledge-Based Inferencing," *Literary & Linguistic Computing*, vol. 13, no. 4, pp. 177–186, 1998.

[23] W. Park and M. Abraham, "Acquisition of Causal Information From Medical Documents in French for Experts," in *23rd Annual EMBS international Conference*, pp. 3848–3851, 2001.

[24] M. Riaz and R. Girju, "In-depth Exploitation of Noun and Verb Semantics to Identify Causation in Verb-Noun Pairs," in *The 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pp. 161–170, 2014.

[25] T. N. De Silva, X. Zhibo, Z. Rui, and M. Kezhi, "Causal Relation Identification Using Convolutional Neural Networks and Knowledge Based Features," *International Journal of Computer and Systems Engineering*, vol. 11, no. 6, pp. 691–696, 2017.

[26] "Desa Sigedong Temanggung KLB Diare, 1 Korban Meninggal Dunia," okezone.com, 2017. [Online]. Available: https://lifestyle.okezone.com/read/2017/08/09/481/1752519/desa-sigedong-temanggung-klb-diare-1-korban-meninggal-dunia [Accessed 12 Feb. 2018].

Leveraging graph-based semantic annotation for the identification of cause-effect relations

Conference on ICT For Smart Society (ICISS), 2017
Publication

6   Lecture Notes in Computer Science, 2015.
    Publication   <1%

7   hdl.handle.net
    Internet Source   <1%

8   www.degruyter.com
    Internet Source   <1%

9   docplayer.net
    Internet Source   <1%

10 doria.fi
    Internet Source   <1%

11 repositorio.udd.cl
    Internet Source   <1%

12 Alisha Das, A. Shukla, Roopa Manjunatha, Ejaz
    Aslam Lodhi. "IoT based Solid Waste
    Segregation using Relative Humidity Values",   <1%
    2021 Third International Conference on
    Intelligent Communication Technologies and
    Virtual Mobile Networks (ICICV), 2021
    Publication

13 Lecture Notes in Computer Science, 2011.
    Publication   <1%

14 publikasi.mercubuana.ac.id
    Internet Source   <1%

Exclude quotes   Off   Exclude matches   Off

Exclude bibliography   On

Leveraging graph-based semantic annotation for the identification of cause-effect relations