Know-Why Extraction from Textual Data for Supporting What Question

Chaveevan Pechsiri
Dept. of Information
Technology,
DhurakijPundit University,
Bangkok, Thailand
itdpu@hotmail.com

Phunthara Sroison
Dept. of Information
Technology,
DhurakijPundit University,
Bangkok, Thailand
phunthara@it.dpu.ac.th

U. Janviriyasopak
Eastern Industry Co.ltd.
Bangkok, Thailand
uraiwanjan@hotmail.com

Abstract

This research aims to automatically extract Know-Why from documents on the website to contribute knowledge sources to support the question-answering system, especially What-Question, for disease treatment. This paper is concerned about extracting Know-Why based on multiple EDUs (Elementary Discourse Units). There are two problems in extracting Know-Why: an identification problem and an effect boundary determination problem. We propose using Naïve Bayes with three verb features, a causative-verb-phrase concept set, a supporting causative verb set, and the effect-verb-phrase concept set. The Know-Why extraction results show the success rate of 85.5% precision and 79.8% recall.

1   Introduction

Automatically Know -Why extraction is essential for providing the rational knowledge source, to the society through question answering system, especially in herbal medicines when assisting the locals to understand more about herbs. According to Jana Trnkova and Wolfgang Theilmann (2004) Know-Why is the knowing of the reason of why something is the way it is. Therefore, Know-Why has to involve the causal relation which is "an irreflexive, transitive and asymmetrical" relation that contains the properties of "productivity (effect is 'produced' by the cause) and locality (it obeys the markov

condition, for model A$\rightarrow$ B $\rightarrow$ C, if there is no B, then A does not cause C)"(Lemeire J. et al. (2004)). Wolff P. (2007) stated that the causal relation can be decomposed into 2 major approaches, the dependency model and the physicalist models. The dependency model can be represented by using statistical dependency model whereas in recent physicalist models are based on the concepts of force dynamic models consisting of 2 force entities in certain events; the agonist and the antagonist (Talmy, 2000). Later, the agonist form (Wolff P., 2007) can be viewed as the 'effect' and the antagonist as the 'cause'. According to Talmy (2000), if there is a situation where the antagonist is stronger, which can be expressed as 'event X happens because of event Y'(Y contains the antagonist.), it is a form of causation. Moreover, the causal relation can pivot on the distinction between causality and causation (Lehmann J. et al, 2004) whereas causality is 'a law-like relation between cause events and effect events' and causation is 'the actual causal relation that holds between individual events'. For example:
"Because a bird sings a song at a window, The rock is thrown at the window."
Causality: "An object vibrates. An object moves."
Causation: "A bird sings. The rock is thrown"
This research focuses only on 'causal relation' to provide both 'causality' for extracting Know-Why from the herbal medicine domain and 'causation' for answering What-question, since what questions contain ambiguities (Girju R. and Moldovan D., 2002) for example:
Know-Why: "*ใบกระเพราใช้เป็นยาขับลม แก้คลื่นไส้ แก้ปวดท้อง /A basil leaf is used as a medicine releasing gas. [The leaf] stops nausea. [The leaf] stops paining the abdomen.*" (where the [..] symbol means ellipsis.)

Know-Why concept: "*A herb organ is used as being a carminative drug. [The organ] is anti nausea, [The organ] is anti stomachache.*"
Question: "ใช้สมุนไพรอะไรแก้คลื่นไส้/*What herb is used for stopping nausea*?" From this example, '*A basil leaf is used as a medicine releasing gas*' is the causation and the concept is the causality.

There are various forms of causal-relation expression such as in the form of intra-NP, inter-NP, and inter-sentence (Chang and Choi,2004). According to our research, we separated this relation into 2 main forms based on the elementary discourse unit (EDU) as defined by (Carlson et al., 2003) as a simple sentence or clause. We defined the intra-causal EDU as an expression within one simple EDU being equivalent to either the intra-NP form or the inter-NP form (Chang and Choi,2004). The inter-causal EDU is defined as an expression within more than one simple EDU which is equivalent to the inter-sentences of Chang and Choi (2004). However, this paper works on only the inter-causal EDU extraction because some cause-effect relation from the herbal web sites are expressed in the form of the EDU containing an EDU-like name entity with the causative action followed by some effect EDUs.

Several techniques (Marcu and Echihabi,2002; Torisawa 2003; Inui and et al.,2004; Pechsiri and Kawtrakul, 2007) have been used to extract cause-effect knowledge varying from two adjacent sentences to multiple sentences. Our work aimed at mining and extracting Know-Why from Thai documents of herbal medicines. Thai has several specific characteristics, such as the existence of sentence-like name entity, zero anaphora or the implicit noun phrase. All of these characteristics are involved in the two main problems of Know-Why extraction: the first problem is how to identify the interesting causality events expressed by an EDU- like name entity from documents, and the second one is how to identify the effect boundary, where The problem of implicit delimiter of the boundary is involved. From all of these problems, we needed to develop a framework which combineed Language Processing and the machine learning technique as Naïve Bayes to learn features of three verb sets, a causative concept verb set, a supporting causative verb set, and an effect concept verb set, for solving those problems.

In conclusion, unlike other methods (Marcu and Echihabi ,2002; Torisawa 2003; Inui and et al.,2004) where the emphasis is based on two adjacent sentences, this paper is based on multiple EDU extraction. Our research was separated into 5 sections. In section 2, related work was summarized. Problems in causality mining from Thai documents will be described in section 3 and in section 4 our framework for causality extraction was explained. In section 5, we evaluated and concluded our proposed model.

2 Related Work

Several strategies such as those done by Marcu and Echihabi ,2002, Torisawa(2003), Inui and et al.(2004), and Pechsiri and Kawtrakul (2007) have been proposed to extract and discover knowledge from the textual data.

Marcu and Echihabi (2002) presented the unsupervised approach to recognize the discourse relations by using word pair probabilities between two adjacent sentences for classifying the rhetorical relations, such as Contrast, Cause-Explanation, Condition, and Elaboration, between two adjacent sentences by using Naïve Bayes classifier to the BLIPP corpus (Charniak, 2000). They determined the word pairs in the cartesian product from the sentence pairs connected with or without discourse marker or connective marker , i.e. 'because' 'but' 'then', to classify the causal relation from other rhetorical relations. The result showed an accuracy of 75% of inter-sentence causality extraction from the corpus size of more than a million sentences for learning whereas our corpus size is 3000 sentences for learning. Therefore, our approach is the supervised approach with the statistical method because our corpus size is small.

Inui's work (Inui and et al.,2004) proposed a method of extraction and classification of causal knowledge. The method of extraction was accomplished under two adjacent sentences by using explicit connective markers; e.g. "because" "since" "if..then" "as the result" etc.. SVM was used for the classification process in (Inui and et al.,2004). Four types of causal relations are studied, including the following: cause, precondition, mean, effect relations. Inui's work's precision is high: 90% but the recall is low: 30%, because of unresolved anaphora. However, in our work, we extract multiple EDUs with some implicit discourse markers.

Torisawa(2003)' s work in extracting the verb phrase pair from the news corpus worked on the assumption that if two events share a common participant (is specified by a noun) then the two events are likely to have a logical relation as causal relation. For example "A man drank

liquor and was intoxicated by the liquor."(a common participant is 'liquor'). However, this assumption can not be applied in our research because most of our causality expression does not share a common participant; e. g. "ขิงใช้เป็นยา ระบาย แก้ท้องผูก/*Ginger is used as being laxative medicine. [The ginger] stops constipation.*

Pechsiri and Kawtrakul (2007), proposed verb-pair rules learned by two different machine learning techniques (NB and SVM) to extract causality with multiple EDUs of a causative unit and multiple EDUs of an effect unit with the problems of the discourse marker ambiguity and the implicit discourse marker. This verb-pair rule has been represented by the following equation (1) (Pechsiri and Kawtrakul, 2007) where V_c is the causative verb concept set, V_e is the effect verb concept set , C is the Boolean variables of causality and non-causality, and a causative verb concept (v_c , where $v_c \in V_c$) and an effect verb concept (v_e , where $v_e \in V_e$) are referred to WordNet (http://wordnet.princeton.edu/) and the predefined plant disease information from Department of Agriculture (http://www.doa.go.th/).

CausalityFunction: $V_c \wedge V_e \rightarrow C$ (1)

They also proposed using V_c and V_e to solve the boundary of the causative unit and using the Centering theory along with V_e to solve the boundary of the effect unit. The outcomes of their research were the verb-pair rule, V_c, V_e, and the multiple EDUs of causality (extracted from textual data) was at their highest precision of 89% and their highest recall of 76%. The correctness of the causality-boundary determination is 88% on average. However, our causative unit consisted of only one EDU containing an EDU-like name entity as a cause, and this EDU was followed by several effect EDUs.

In our current work, we aimed at extracting the Know-Why in Natural Language description instead of visualizing only associations of concepts, by applying both language processing and learning technique by Naïve Bayes to identify the causality expression.

3 Problem of Know-Why Extraction

To extract the cause-effect expressions, there are two main problems that must be solved. The first problem is how to identify interesting cause-effect relations from the documents. The second problem is how to determine the effect boundary.

There is also the problem of implicit noun phrase.

3.1 Causality Identification

The problem involves the word level and the sentence level. For the word level, the medicinal name entity may express in the form of a sentence like name entity or an EDU- like name entity which explains the medicinal action as the causative action of medicine, and medical characteristic. The problem of this level is how to identify the causative name entity. For example:

a) "ใบกระเพรา/***A basil leaf*** ใช้เป็น/***is used as*** ยา/ ***medicine*** ขับ/***releases*** ลม/***gas***"

where 'a medicine releases gas' is an EDU-like name entity with the causative action, 'release'.

b) "แก่นกระแจะ/***Nicolson stem*** ใช้ทำ/***is used for making*** ยา/***medicine*** ดอง/***soaks in*** เหล้า/ ***liquor***"

where 'a medicine soaks in liquor' is an EDU-like name entity with the characteristic of medicine being preserved in the alcohol.

The above examples, a) and b), contain an EDU-like name entity which is a cause in a) and a non cause in b).

For the sentence level, the EDU containing an EDU-like name entity with the causative action may be followed by an effect EDU(s) to form the cause-effect or causality relation between the EDU like name entity and that following EDU(s). For example:

Causality

EDU1 "ตะไคร้หอม/***Lemon grass*** ใช้เป็น/***is used as*** ยา/***medicine*** บีบ/***contracts*** ,มดลูก/***a uterus***" (where 'a medicine contracts a uterus.' is the EDU-like name entity with concept of 'the medicine causes uterus to contract'.)

EDU2 "***[The plant]*** ขับ/***discharges*** ประจำเดือน/ ***period.***" (=The plant discharges period.)

Non causality

EDU1 "ใบกระเพรา/***A basil leaf*** ใช้เป็น/***is used as*** ยา/***medicine*** ขับ/***releases*** ลม/***gas***." (where 'a medicine releases gas' is the causative EDU-like name entity.)

EDU2 "*[the basil leaf]*รักษา/*relieves* แผล/*ulcer* ใน/*in*กระเพาะอาหาร/*stomach.*"(= [The basil leaf relives ulcer in a stomach.)

Where in this example, EDU 1 is the cause and EDU2 is the effect

3.2 Effect Boundary Determination

There are two problems of an implicit effect boundary cue and the effect EDU containing interrupts.

3.2.1 *Implicit Effect Boundary Cue*
Some cause-effect relations from the herbal web sites are expressed in the form of the EDU containing an EDU like name entity with the causative action followed by some effect EDUs without any cue of ending effect boundary, e.g. "และ/and". For example:

EDU1 "ใบกระเพรา*A basil leaf* ใช้เป็น/*is used as* ยา/*medicine* ขับ/*releases* ลม/*gas*" (=A basil leaf is used as a medicine releasing gas.)

EDU2 "*[The basil leaf]* แก้/*stops* คลื่นใส้/ *nauseate.*" (=The basil leaf stop being nausea.)

EDU3 "*[And the leaf]* แก้/*stops* ปวด/*pain* ท้อง/ *abdomen*." (= [And the leaf] stops paining abdomen.)

Where in this example, EDU 1 is the cause and EDU 2 & EDU3 are the effects. EDU 2 and EDU3 help us to determine the boundary.

3.2.2 *Effect EDU Containing Interrupts*
There are some effect EDUs containing interrupts as shown in the following example:

EDU1 "หอมแดง/*A red onion* ใช้เป็น/*is used as* ยา/ *medicine* ถ่าย /*be laxztive*" (=A red onion is used as a laxative medicine.)

EDU2 "*[And the red onion]* แก้/*stops* ท้องผูก/ *being constipation*" (= [And the red onion] stops being constipation.)

EDU3 "*[The red onion]* ขับ/*discharges* ปัสสาวะ /*urine.*" (= [The red onion] discharges urine.)

EDU4 "*[The red onion makes a patient]* เจริญ อาหาร/*be appetite*." (= [The red onion] makes a patient] be appetite.)

Where the EDU-like name entity in EDU1 is a cause with EDU2 and EDU4 as its effects. The EDU3 is an interrupt. Although EDU3 is the effect of red onions, but EDU 3 is not the effect of laxatives.
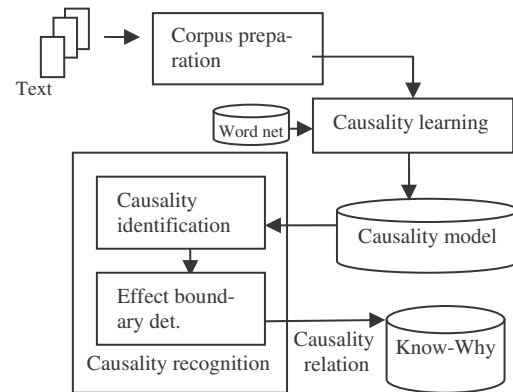
4 A Framework for Know-Why Extraction

Figure 1. System Overview

There are three steps in our framework. First is the corpus preparation step followed by causality learning, and causality recognition steps (as shown in figure 1).

4.1 Corpus Preparation

There are two steps of pre-annotation and Causality annotation.

4.1.1 *Pre-annotation*
This step is the preparation of the corpus in the form of EDU from the text. The step involves using Thai word segmentation tools to solve a boundary of a Thai word and tagging its part of speech (Sudprasert and Kawtrakul, 2003). This process includes Name entity (Chanlekha and Kawtrakul, 2004), and word-formation recognition (Pengphom, et al 2002) to solve the boundary of Thai Name entity and Noun phrase.

After the word segmentation is achieved, EDU segmentation is dealt with. According to Charoensuk et al. (2005), this process segments plain text into units of EDUs by using the rule based and the machine learning technique of C4.5 (Mitchell T.M., 1997). These generated EDUs will be kept as an EDU corpus. This corpus will contain 4500 EDUs and will be separated into 2 parts, one part is 3500 EDUs for causality learning and the other part of 1000 EDUs for causality recognition and extraction.

4.1.2 *Causality Annotation*
Due to the problems in the causality identification, verbs from three EDUs (with one EDU as an EDU-like name entity) in the EDU corpus are used in this process to learn for extracting causality. Word ambiguity will be solved through the

finding of word concepts from Wordnet. Since Thai Wordnet does not exist, we need to translate from Thai to English, using Lexitron (the Thai-English dictionary)(http://lexitron.nectec.or.th/), before using Wordnet(http://wordnet.princeton. edu/obtain). In this process, we manually annotate the causality EDUs by annotating the EDU containing the causative EDU-like name entity as the causative EDU. We annotate a verb phrase in the causative EDU-like name entity to be a causative-verb-phrase concept (referred to Wordnet). The verb from EDU which contains the causative EDU-like name entity is annotated with a concept and we call this verb as 'supporting causative verb'. We also annotate the effect-verb-phrase concept(referred to Wordnet and http://www.ars-grin.gov/duke/ethnobot.html) from effect EDUs following the EDU containing the causative EDU-like name, as shown in Figure 2)

4.2 Causality Learning

The aim of this step was to learn cause-effect relation between causative events and effect events from annotating an EDU corpus.

4.2.1 *Feature Extraction*

All annotated verb features from the previous step are extracted into database table (in Table 1) including surface forms of verb features along with their concepts used for probability determination in the next step.

```
<C id=1>
<EDU type =cause>
  <NP1 concept=a herb organ>ใบโหระพาA basil leaf</NP1>
  <VS concept=use#1>ใช้เป็นis used as</VS>
   <EDU-Like-NE >
     <NP2 concept=drug>ยาmedicine</NP2>
     <CVC concept= be carminative/ eliminate gas from a body>
        ขับreleases ลมgas
     </CVC>
   </EDU-Like-NE>
</EDU>
</C>
<R id=1>
<EDU type=effect>
   <EVC concept= stop nausea/ be anti nausea>แก้stops คลื่นไส้ nauseate.
   </EVC>
</EDU>
<EDU type=effect>
   <EVC concept=stop paining an abdomen/ relieve abdominal pain>แก้/
     stops ปวด/pain ท้อง/ abdomen
   </EVC>
</EDU>
</R>

EDU= EDU, EDU-Like-NE= EDU-like name entity tag,
C=cause tag, R=result or effect tag, VS= supporting verb tag ,
CVC=causative verb concept tag, EVC=effect verb concept tag
NP1 NP2= noun phrase tag
```

Figure 2. Causality Annotation Tag

NP1	NP1 Concept	Vs	Vs Concept	VPc	VPc concept	VPe	VPe Concept	Class
กระแจะ/ Naringi crenulata	herb	ใช้เป็น	use as	ดับพิษ/ cure poison	be-antipyretic	แก้ปวด เมื่อย	relieve muscle pain	n
บัวบก/ Asiatic Pennyworth	herb leaf	ใช้เป็น	use as	ทา ภายนอก/ apply externally	apply topically	รักษาแผล	heal wound	y
หอมแดง/ red onion	herb	เป็น	is	ถ่าย/ excrete	be-lexative	แก้ท้องผูก	stop being constipation	y
หอมแดง/ red onion	herb	เป็น	is	ถ่าย/ excrete	be-lexative	ขับ ปัสสาวะ	discharge urine	n
หอมแดง/ red onion	herb	เป็น	is	ถ่าย/ excrete	be-lexative	เจริญ อาหาร	be appetite	y
ขมิ้นชัน/ curcumin	herb	เป็น	is	ฆ่าเชื้อ/ antiseptic	be-antiseptic	รักษา โรค ผิวหนัง	cure skin disease	y
มะแว้ง/ Soianum indicum Linn	herb	ทำเป็น	make as	ลดน้ำตาล ในเลือด/ reduce blood sugar	balance blood sugar level	แก้ไอ	stop coughing	n
กระเพรา / Basil	herb leaf	ใช้เป็น	use as	ขับลม/ release gas	be carminative	แก้คลื่นไส้	relieve nausea	y
กระเพรา / Basil	herb leaf	ใช้เป็น	use as	ขับลม/ release gas	be carminative	แก้ปวด ท้อง	stop paining an abdomen	y
ขิง / ginger	herb	ใช้เป็น	use as	ขับลม/ release gas	be carminative	แก้คลื่นไส้	relieve nausea	y
มะกูด / bergamot leaf	herb leaf	ใช้เป็น	use as	ขับลม/ release gas	be carminative	แก้คลื่นไส้	relieve nausea	y
...	...	...	...	...	...	...	...	...

Table 1. The extracted features from the annotated corpus

21

Vs concept	causality	non causality
ใช้+เป็น/use as	0.27619	0.290323
Be	0.561905	0.612903
ทำ+เป็น/make as	0.009524	0.032258
ใช้+ทำ+เป็น/use for making as	0.066667	0.053763
…	…	…
VPc concept	**causality**	**non causality**
'ขับ+ลม/release-gas'	0.371901	0.192661
'แก้+ไอ/anti coughing'	0.024793	0.045872
'ทา/apply'	0.140496	0.009174
'ขม/be-bitter'	0.041322	0.009174
'ขับ+ปัสสาวะ/discharge-urine'	0.057851	0.06422
'ขับ+เสมหะ/be expectorant'	0.041322	0.06422
'บีบ+มดลูก/contract uterus/oxytocic'	0.041322	0.027523
'รักษา+เบาหวาน/be antidiabetic'	0.008264	0.027523
…	…	…
VPe concept	**causality**	**non causality**
'แก้+ปวด+ท้อง/stop-stomachach/relieve abdominal pain'	0.035714	0.007813
'แก้+คลื่นใส้/stop-naucea/be anti nausea'	0.035714	0.007813
'แก้+ท้องอืดท้องเฟ้อ/stop-flatulence/relieve indigestion'	0.15	0.007813
'แก้+ลมพิษ/stop-rash/ be antiurticaria'	0.035714	0.023438
'ลดไข้/reduce-fever'	0.021429	0.039063
'ขับ+รก/eliminate-placenta'	0.007143	0.054688
'เจริญ+อาหาร/increase appetite'	0.092857	0.007813
'ขับ+เหงื่อ/release-sweat/be diaphoretic'	0.007143	0.070313

Table 2. Show probability of V_s concept, VP_c concept and VP_e concept

4.2.2 *Probability Determination*
After we had obtained the extracted verb features, we then determined the probability of causal and non causal from the occurrences of the cartesian products of three verb feature concepts , shown in Table2, by using Weka which is a software tool for machine learning (http://www.cs. waikato.ac.nz/ml/weka/).

4.3 Causality Recognition and Extraction

The objective of this step was to recognize and extract the cause-effect relation from the testing EDU corpus. In order to start the causality recognition process, Naïve Bayes Classifer shown in equation (2) is applied with the feature probabilities in Table 2, where EDUs class is determined

by class1 (causality EDUs) and class0 (non causality EDUs).

$$EDUclass = \underset{class \in Class}{\arg\max} \ P(class|v_s, vp_c, vp_e)$$
$$= \underset{class \in Class}{\arg\max} \ P(v_s|class)P(vp_c|class)P(vp_e|class)P(class) \quad (2)$$

$v_s \in V_s$ where V_s is a Supporting Causative Verb concept set
$vp_c \in VP_c$ where VP_c is a Causative VerbPhrase concept set
$vp_e \in VP_e$ where VP_e is a Effect VerbPhrase concept set

Therefore, Causality Recognition can be separated into 2 steps: causality identification and effect boundary determination.

4.3.1 *Causality Identification*
This step was to determine the interesting locations that are cause-effect relations by searching any EDU which consists of a verb matching to a verb in the supporting causative concept set, V_s, and an EDU-like name entity containing a causative-verb-phrase concept as vp_c (where $vp_c \in VP_c$).

4.3.2 *Effect Boundary Determination*
The effect EDU and the effect boundary were determined at the same time by checking all sequence EDUs right after the EDU containing vp_c in the EDU-like name entity. If a verb phrase from the sequence of checked EDUs is not in VP_e, the possible effect boundary is end. After the possible boundary is determined, v_{s_inEDU1}, vp_{c_inEDU1} and $vp_{e\_inEDU2}..vp_{e\_inEDUn}$ (where n>2) will be used to determine the causality class from the Naïve Bayes Classifier equation (2) as shown in Figure 3. The actual effect boundary is determined from the last class1 in the sequence of $EDU_2.. EDU_n$.

Furthermore, where the implicit noun phrase occurs as the subject of the current EDU, this has to be solved in this step by using the heuristic rule which is that the noun phrase as a subject of the previous EDU will be the subject of the current EDU.

Assume that each EDU is represented by (np vp)
L is a list of EDU
VP$_C$ is a causative-verb-phrase concept set, VP$_E$ /VPe is a effect-verb-phrase concept set
V$_S$ is a supporting causative verb concept set
CAUSALITY_EXTRACTION (L, V$_C$, V$_E$, Vs)

```
1       i ← 1, R ← ∅
2       while i ≤ length[L] do
3       begin_while1
4           CA ← ∅, EC ← ∅
5           if (vp_i ∈ V_S) ∧ (vp_{i_in_NE} ∈ VP_C) then
6            begin_if
                CA ← CA ∪ {i}, i ← i + 1    /*CA is causa
                                              tive EDU
7                while (vp_i ∈ VP_E) do
8                  begin_while2
9           res ←
              arg max   P(v_s |c)P(vp_c |c)P(vp_e |c)P(c)
              c∈( yes,no)
10               if res=yes
11                 EC ← EC ∪ {i},   /*EC is effect
                                             EDU
12                 i ← i + 1
13               end_while2
14            endif
15          if res = yes ∧ CA <>∅  then
16              R = R ∪ { (CA,EC) }
17      end_while1
18       return R
```

Figure3. Show Causality Extraction algorithm for the EDU containing the causative EDU-like name entity, and followed by multiple effect EDUs .

5   Evaluation and Conclusion

The Thai corpora used to evaluate the proposed causality extraction algorithm consist of about 1,000 EDUs collected from several herbal web sites. The evaluation of the causality extraction performance of this research methodology is expressed in terms of the precision and the recall as shown below, where R is the causality relation:

$$Precision = \frac{\#of\ samples\ correctly\ extracted\ as\ R}{\#of\ all\ samples\ output\ as\ being\ R} \quad (3)$$

$$Recall = \frac{\#of\ samples\ correctly\ extracted\ as\ R}{\#of\ all\ samples\ holding\ the\ target\ relation\ R} \quad (4)$$

The results of precision and recall are evaluated by three expert judgments with max win voting. The precision of the extracted causality 85.5% with 79.8% recall. The correctness of our effect boundary determination by these expert judgments is 86%. These research results can be increased if we use a larger corpus. However, our methodology will be very beneficial for con-

tribute the causality knowledge for supporting What-question with the concept of causal relation from a web page by inference method of backward chaining, for example:

Extracted causality: "ใบโหระพาใช้เป็นยาขับลม แก้คลื่นไส้ แก้ปวดท้อง /A basil leaf is used for a gas released medicine. [The leaf] stops nausea. [The leaf] stop stomachache." ……………

The above extracted causality can be represented by the following predication.

a) ∀x   be_herb(x) ^ be_herb_medicine(y) ^ be_carminative (y) ^ use_as(x,y)→ stop(x, z) ^ be_nausea(z)

b) ∀x   be_herb(x) ^ be_herb_medicine(y) ^ be_carminative (y) ^use_as(x,y)→ stop(x, z) ^ be_abdominal pain(z)

Where x ∈ X,{'ใบกระเพรา/basil leaf' 'ขิง/ginger' 'พริกไทย/black pepper' 'ใบมะกรูด/bergamot leaf'..}, and X is the extracted NP1 set from EDUs containing the causative EDU-like name entities and being followed by the effect EDUs , e.g. (stop(x, z) ^ be_nausea(z)), (stop(x, z) ^ be_stomach ache(z)).

Question: "ใช้ /**use** สมุนไพร /**herb** อะไร /**what** แก้/**stop** คลื่นไส้/**nausea** (What kind of herb is used for stop nausea?)

The backward chaining from the above question and the extracted causality in a) is shown in the following
stop(x, z) ^ be_nausea(z) →be_herb(x) ^ be_herb_medicine(y) ^ be_carminative (y) ^ use_as(x,y)

where x is 'ใบกระเพรา/basil leaf', 'ขิง/ginger', 'พริกไทย/black pepper', or 'ใบมะกรูด/bergamot leaf'

References

Carlson L., Marcu D., and Okurowski M. E. 2003. Building a Discourse-Tagged Corpus in the Framework of Rhetorical Structure Theory. In Current Directions in Discourse and Dialogue. pp.85-112.

Chanlekha H. and Kawtrakul A. 2004. Thai Named Entity Extraction by incorporating Maximum Entropy Model with Simple Heuristic Information. IJCNLP' 2004.

Chareonsuk J ., Sukvakree T., and Kawtrakul A. 2005. Elementary Discourse unit Segmentation for

Thai using Discourse Cue and Syntactic Information. NCSEC 2005.

Chang D.S. and Choi K.S. 2004. Causal Relation Extraction Using Cue Phrase and Lexical Pair Probabilities. IJCNLP. pp. 61 – 70.

Charniak, E. 2000. A maximum-entropy-inspired parser. Proc. of NAACL, pp.132-130.

Girju R. and Moldovan D. 2002. Mining answers for causation questions. AAAI Symposium on Mining Answers from Texts and Knowledge Bases.

Inui T., Inui K., and Matsumoto Y. 2004. Acquiring causal knowledge from text using the connective markers. Journal of the information processing society of Japan 45(3), pp. 919-993.

Lemeire, J., S. Maes and E. Dirkx. 2004. Causal Models for Parallel Performance Analysis. Fourth PA3CT-Symposium, Edegem, Belgium, September.

Marcu D. and Echihabi A. 2002. An Unsupervised Approach to Recognizing Discourse Relations. in Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics Conference. pp. 368 – 375.

Pechsiri C., Kawtrakul A. and Piriyakul R. 2005. Mining Causality Knowledge From Text for Question Answering System. IEICE Transactions on Information and Systems, Vol.E90-D, No.10 :1523-1533.

Pengphon N., Kawtrakul A., and Suktarachan M. 2002. Word Formation Approach to Noun Phrase Analysis for Thai. SNLP.

Mitchell T.M. 1997. Machine Learning. The McGraw-Hill Companies Inc. and MIT Press, Singapore.

Sudprasert S. and Kawtrakul A. 2003. Thai Word Segmentation based on Global and Local Unsupervised Learning. NCSEC'2003.

Talmy, L. 2000. Toward a Cognitive Semantics Concept Structuring Systems – Vol. 1. The MIT Press.

Torisawa K. 2003. Automatic Extraction of Commonsense Inference Rules from Corpora. In Proc. Of The 9th Annual Meeting of The Association for Natural Language Proceeding. pp. 318-321.

Trnkova, Jana, Wolfgang Theilmann. 2004. Authoring processes for Advanced Learning Strategies. Telecooperation Research Group,TU Darmstadt, and SAP Research, CEC Karlsruhe. Germany.

Wolff, P. 2007. Representing Causation. Journal of experimental psychology: General 2007 Vol. 136 No.1 82-111. USA.