Extraction of Problem Events from Web Documents to Construct Cause-Effect Loop

Chaveevan Pechsiri[1] and Renu Sukharomana[2]
[1]*College of Innovative of Technology and Engineering, Dhurakij Pundit University,Bangkok,Thailand*
[2]*Institute for Social and Economic Sciences, TRF, Bangkok, Thailand*
chaveevan.pec@dpu.ac.th

***Abstract*— This research aims to extract problem events, particularly cause-effect concept pair series with explanations by several simple sentences with causative/effect concepts, from web documents of drug addiction. The extracted problem events are used to construct cause-effect loop which benefits for the problem analysis in the solving system. The research has three problems; how to determine the cause/effect event concepts expressed by verb phrases having a problem of the overlap between causative-verb concepts and effect-verb concepts, how to determine the series of cause-effect concept pairs with the causative/effect concept boundary consideration, and how to determine the feedback-loop of cause-effect concept pair series. Therefore, we apply the event rate to solve the overlap problem. We then propose using N-WordCo to determine the cause-effect concept pair series and also use a cue-word set to solve the feedback-loop. The research results provide the high precision of the problem event extraction from the documents.**

***Index Terms*—Cause-Effect Series; N-WordCo; Cause-Effect Loop.**

I. INTRODUCTION

The objective of this paper is to extract problem events with concepts, especially cause-effect concept pairs as event series, from drug addiction documents downloaded from hospitals' healthcare web-boards (i.e., http://haamor.com/ which is a non-government-organization website). The problem events of the drug addiction are increasing concern to people because they worry about the crime and violence that is associated with drugs. They also worry that drugs are becoming more widespread and are becoming increasingly easy for children to use. Therefore, the research concerns on determining and extracting the problem events represented by a cause-effect loop (which links between causative-concept event nodes and effect-concept event nodes into a loop similar to a causal-loop diagram [1] without the positive/negative identifications on links) from texts to enhance the preliminary problem analysis of the solving system. Where the problem-event expression as the series of the cause-effect concept pairs (which are the cause-effect relation type) are explained by several EDUs (each EDU is an Elementary Discourse Unit expression defined as a simple sentence or a clause, [2]) as shown in Example 1(Figure 1). Figure 1: the EDU10-EDU9 association in step4 is another effect of Step3. The Step2 through Step4 occurrences can be represented by the cause-effect loop as shown in Figure 2 having a feedback-loop variable as 'using drug' (EDU11)

This research emphasizes only the verb phrase expression because the problem events of the research mostly are based on several consequences of events expressed by the EDUs'verb phrases. The EDU expression has the Thai linguistic patterns (as shown in Figure 3) after stemming words and the stop word removal.

Example1: EDU1: "*พ่อแม่ทะเลาะกันทุกวัน/Parents fight each other in every day.*"
EDU2: "*จน [พวกเขา]แตกแยก/ until [they] separate.*
EDU3: "*ทำให้เด็กรู้สึกไม่อยากอยู่บ้าน/ Cause the teenager don't want to stay home.*"
EDU4: "*[เด็ก]อยู่กับเพื่อนมีปัญหาเดียวกัน/ [He] stay with friends having the same problem.*"
EDU5: "*และ [เด็ก] รู้สึกเครียด/ and [he] feels stress .*"
EDU6: "*ทำให้[เด็ก]เริ่มใช้ยาเสพติดเพื่อแก้ปัญหา/ Cause [him] to start using drug for solving problems.*"
EDU7: "*สารเสพติดมีผลต่อสมอง/ The drug has an affect to the brain.*"
EDU8:"*เด็กเริ่มมีปัญหากับการเรียนในชั้น/He starts to have the problem of studying in the class.*"
EDU9: "*ต่อมา[เด็ก]มีอาการกระวนกระวาย/ Then [he] have the impatient symptom.*"
EDU10: "*เพราะ [เด็ก]ต้องการใช้ยาอีก/ because [he] craves for using the drug again.*"
EDU11: "*และ เมื่อ[เด็ก]เสพ [ยา] / and when [he] use [drug].*" ……………………………
(where [..] means ellipsis.)
Example1 can be expressed as the series of the cause-effect concept pairs as follow
Step1. (EDU1∧EDU2): Cause → (EDU3∧EDU4∧EDU5): Effect - of the family problem.
Step2. (EDU3∧EDU4∧EDU5): Cause → (EDU6): Effect - of the family problem..
Step3. (EDU6): Cause → (EDU7∧EDU8): Effect - of the drug-use symptoms.
Step4. (EDU10): Cause → (EDU9): Effect- of the craving symptoms for the drug.

Figure 1: Example of Problem Event Occurrences on Documents

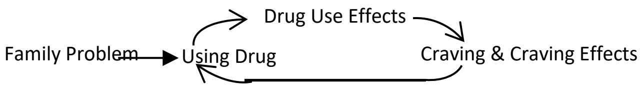

Figure 2: Addictive Cause-Effect

EDU → NP1 VP | VP
VP → Verb NP2 | Verb adv | Verb AdvPhrase$_{dose}$
Verb→ Preverb Verb | V_{weak}-noun2| V_{weak}-noun2 Verb| V_{strong}| V_{strong} Verb
NP1 → pronoun | Noun1 | Noun1 modify | Noun2 | Noun2 modify
NP2→ Noun2 | Noun2 modify
modify →Adj| Adj modify| Noun1 modify |Noun2 modify
V_{weak}→ {'เป็น/*be*', 'มี/*have*', 'ใช้/*use*', 'นำ/*take*', 'เอา/*get*', 'รู้สึก/*feel*'}
V_{strong}→{'ว่างงาน/*be-jobless*','ยากจน/*be-poor*','ชักชวน/*induce*','ทะเลาะ /*quarrel,fight*', 'แยก/*separate*',...,'ดื่ม,กิน,เสพ/*consume*','ใช้/*use*','ฉีด/*inject*','สูดดม/*sniff*',...,'ออกฤทธิ์/*activate*','กระตุ้น/*urge*','ตื่นตัว/*be-awakened-to*','หวาดระแวง/*be-mistrustful*','ชัก/*convulse*','เสียสติ,บ้า/*be-insane*','คลุ้มคลั่ง/*be-manic-depression*','หมดสติ/*lose-consciousness*','เสื่อม/*deteriorate*','เสียชีวิต/*die*','หัวเราะ /*laugh*','เคลิบเคลิ้ม/*be-absent-minded*','กดประสาท/*be-sedative*','ลด/*reduce*',...,'อยาก,ต้องการ/*crave*','ติด/*be-addicted-to*','ขาด/*withdraw*',...,'กระวนกระวาย/*be-nervous*','วิตกกังวล/*be-anxious*','ทำร้าย/*harm*','เครียด/*be-stressed-out*','หงุดหงิด/*fidget*','ก้าวร้าว/*be-aggressive*','อ่อนเพลีย/*be-weak*','ซึมเศร้า/*sadden*,…}
Noun1→{' ','เด็ก,วัยรุ่น/*youth,teenager*','พ่อแม่/*parents*','ครอบครัว/*family*',…}
Noun2→{' ','ยา/*drug*','อาการ/*symptom*','ประสาท/*nerve*','สมอง/*brain*','จิตใจ/*mental*', 'หัวใจ/*heart*','หลอน/*hallucination* ',…}
Adv→{'อย่างแรง/*intensely*','ซ้ำ/*repeatly*'..}; Adj→{'สูง/*high*','ต่ำ/*low*'..};
Preverb→{'ไม่/*not*'..}

where NP1 and NP2,are noun phrases. VP is a verb phrase. V_{strong} is a strong verb concept set. V_{weak} is a weak verb concept set. Adv is an adverb concept set. Adj is the adjective concept set.

Figure 3: Thai Linguistic Expression after Stemming Words and Stop Word Removal

In Figure 3, V_{strong} consists of the causative verb concept set,V_{sc}, and the effect verb concept set ,V_{se}, ($V_{strong}= V_{sc} \cup V_{se}$). V_{weak} requires more information, i.e., V_{weak}-Noun2, to have either the cause-event concept or the effect-event concept. As Regard to Example1 on Figure 1, the problem events expressed by verb phrases can be presented by the following general cause-effect series expression.

VP_{EDUc} = an EDU's verb phrase with a causative concept.
VP_{EDUe} = an EDU's verb phrase with an effect concept.
CEi= a cause-effect concept pair which consists of a vector of $VP_{EDUc\text{-}ia}$ and a vector of $VP_{EDUe\text{-}ib}$; i=1,2,..n; a =1,2,.. α ; b=1,2,..$\beta/\gamma/\varphi$
CE1:$\langle VP_{EDUc\text{-}11}VP_{EDUc\text{-}12}\ldots VP_{EDUc\text{-}1\alpha}\rangle$as Cause → $\langle VP_{EDUe\text{-}11}VP_{EDUe\text{-}12}\ldots VP_{EDUe\text{-}1\beta}\rangle$as Effect
CE2:$\langle VP_{EDUe\text{-}11}\ldots VP_{EDUe+1\beta}\rangle$as PartialImplicit/Implicit Cause→$\langle VP_{EDUe\text{-}21}..VP_{EDUe\text{-}2\gamma}\rangle$as Effect
CE3:$\langle VP_{EDUe\text{-}21}\ldots VP_{EDUe\text{-}2\gamma}\rangle$as PartialImplicit/Implicit Cause→...........

There are several techniques [3]-[8] having been applied for determining the cause-effect/ causality/causal relation from texts (see section II). However, the Thai documents have several specific characteristics, such as zero anaphora or the implicit noun phrase, without a word and sentence delimiters, and etc. All of these characteristics are involved in three main problems (see section III). The first problem is how to determine the cause/effect event concepts expressed by verb phrases having a problem with the overlap between causative-verb concepts and effect-verb concepts. The second problem is how to determine the series of cause-effect concept pairs with the causative/effect concept boundary consideration. And the third problem is how to determine the feedback-loop of the cause-effect loop. According to these problems, we need to develop a framework which combines machine learning and the linguistic phenomena to learn the several EDUs of the cause-effect expressions on the downloaded documents. Therefore, we apply the experimental event rate [9] between two event-concept occurrences to solve the verb overlap problem. We collect N-WordCo (is a word co-occurrence with N words) with causative/effect concepts having N-WordCo size learned by Naïve Bayes (NB) [10] from verb phrases after stemming words and eliminating stop words. We then propose using collected N-WordCo expressions to solve the cause-effect concept pair series. We also use the cue-word set or the loop cue-word set to determine the feedback-loop.

Our research is separated into 5 sections. In section II, related work is summarized. Problems in extracting series of cause-effect concept pairs as the problem events from texts are described in section III, and section IV shows our framework of the problem event extraction system. In section V, we evaluate and conclude our proposed model.

II. Related Works

Several strategies, [3]-[8], have been proposed to determine the cause-effect relation from texts without the cause-effect series consideration except [8]. [3] proposed decision tree learning the causal relation from a sentence based on the lexico syntactic pattern (NP1 causal-verb NP2). [4] used cue-phrase and the statistical approach to NP-pair probabilities to solve the causal relation occurrence within two EDUs. [5] applied verb-pair rules and machine learning techniques to extract the causality occurrence within several effect EDUs. There are more research works based on the lexico syntactic pattern with the causal concept as in [6] proposed the Restricted Hidden Naïve Bayes model to learn and extract the causality from the English documents. [6]'s learning features include contextual, syntactic, position, and connective features. [7] applied the rule-based, Support Vector Machine and the temporal reasoning to extract the causal relation on a complex sentence or two simple sentences from English documents. [8] made causal chains by adding the causal chains (obtained from latent topics) to the causal chains obtained from word matching. [8]'s model is based on noun features. However, most of the previous works on the cause-effect relation are based on noun/NP features existing on one/two sentences without the boundary consideration except [5] whereas our work has several NP ellipses occurring on documents. And there are few works on cause-effect series extraction from texts.

III. Problems in Extracting Series of Cause-Effect Concept Pairs

There are three problems, how to identify VP_{EDUc} and VP_{EDUe} , how to determine the cause-effect concept pair series with the cause/effect boundary consideration, and how to determine the feedback-loop.

A. How to Identify VP_{EDUc} and VP_{EDUe}

Regard to the session I, V_{strong} can be used to identify VP_{EDUc} and VP_{EDUe}.

V_{sc}={'ว่างงาน/*be-jobless*', 'ยากจน/*be-poor*', 'แยก/*separate*', 'เครียด/*be-stressed-out*', 'ดื่ม,กิน,เสพ/*consume*', 'ใช้/*use*', 'ฉีด/*inject*', 'อยาก,ต้องการ/*crave*', 'ขาด/*withdraw*' , 'ติด/*be-addicted-to*',…}
V_{se}={'กิน,เสพ/*consume*','ใช้/*use*','ฉีด/*inject*',อยาก,ต้องการ/*crave*','ขาด/*withdraw*','กระตุ้น/*urge*', 'หวาดระแวง/*be-mistrustful*','ทำร้าย/*harm*','เคลิบเคลิ้ม/*be-absent-minded*','กระวนกระวาย/*be-nervous*',..}

However, some V_{sc} and V_{se} elements cannot be used to identify VP_{EDUc} and VP_{EDUe} respectively because of $V_{sc}\cap V_{se} \neq \emptyset$. Moreover, using V_{weak} to identify VP_{EDUc}/VP_{EDUe} has a problem of how to determine the number of followed words, i.e. Noun2.., for providing the causative/effect concept.

B. How to Determine VP_{EDUc} /VP_{EDUe} Boundary

Pattern1
$VP_{EDUc\text{-}i1}..VP_{EDUc\text{-}i\alpha}VP_{EDUe\text{-}i1}VP_{EDUe\text{-}i2}..VP_{EDUe\text{-}i\beta}VP_{EDUc\text{-}(i+1)1}VP_{EDUe\text{-}(i+1)1}VP_{EDUe\text{-}(i+1)2}..VP_{EDUc\text{-}(i+1)\gamma}..$
< ----cause--------- >< -----------effect------------ ><--cause-->< -----------effect---------------------- >
< ------------------- CE*i* ---------------------------- >< ------------- CE*i*+1------------------------------- >
Pattern2
$VP_{EDUc\text{-}i1}..VP_{EDUc\text{-}i\alpha}VP_{EDUe\text{-}i1}..VP_{EDUe\text{-}i\beta}VP_{EDUe\text{-}(i+1)1}..VP_{EDUe\text{-}i\delta}VP_{EDUc\text{-}(i+1)1}VP_{EDUe\text{-}(i+1)(\delta+1)}..VP_{EDUc\text{-}(i+1)(\delta+\gamma)}.$
< ------ cause ------>< ------ effect ----- >< --------- effect ------>< --cause-->< -------------effect ------------ >
< ------------------ CE*i* ------------------- >< ----------------------- CE*i*+1 -- >
Pattern3
$VP_{EDUc\text{-}i1}..VP_{EDU\text{-}1}VP_{EDU\text{-}2}..VP_{EDU\text{-}\eta}VP_{EDUe\text{-}i1}VP_{EDUe\text{-}i2}.....$
<-cause-> < ------ effect --------- >
< ----CE*i* -- -------------------------- >
$VP_{EDUc\text{-}i1}..VP_{EDUc\text{-}i\alpha}VP_{EDUe\text{-}i1}\ VP_{EDU\text{-}1}VP_{EDU\text{-}2}..VP_{EDU\text{-}\eta}VP_{EDUe\text{-}i2}.....\ VP_{EDUe\text{-}i\beta}..$
< ------ cause-------><-effect-- ------- effect ---------->
< ------------------ CE*i* -------- -------------------------->

Figure 4: Patterns of Cause-Effect Concept Pair Series

There are several patterns of the cause-effect concept pair series expression as CE*i* and CE*i*+1on the documents, as shown in Figure 4. Pattern2:V_{se} cannot solve two adjacent-VP_{EDUe} boundaries of CE*i* and CE*i*+1. Pattern3: there are non-VP_{EDUc}/non-VP_{EDUe} occurrences within the VP_{EDUc}/VP_{EDUe} boundary or between CE*i* and CE*i*+1.

According to III.*A* and III.*B* problems, we apply the experimental event rate (or "*Event Rate, ER, is a measure of how often a particular statistical event, i.e. response to a drug, occurs within the experimental group*") (https://en.wikipedia.org/wiki/Odds_ratio) [9] to solve the verb overlap problem. We use ER to measure the frequencies of the V_{strong} occurrences and V_{weak}-Noun2 occurrences on the corpus as causative concepts and as effect concepts for the verb categorization into three verb groups/sets, a root-cause group (VRC), an inter-cause/effect group(VCorE), and an effect group(VE) as follow.

$$\text{ER-of-}v_{s\text{-}c} = theNumberOf\ v_{s\text{-}c} / theNumberOf\ v_{s\text{-}c} + theNumberOf\ v_{s\text{-}e} \quad (1)$$

$$\text{ER-of-}v_{s\text{-}e} = theNumberOf\ v_{s\text{-}e} / theNumberOf\ v_{s\text{-}c} + theNumberOf\ v_{s\text{-}e} \quad (2)$$

$v_{s\text{-}c}$ is a v_s with a causative concept; $v_{s\text{-}e}$ is a v_s with an effect concept; where ($v_s \in V_{strong}$)

$$\text{ER-of-}v_{w\text{-}c}w_1 = theNumberOf\ v_{w\text{-}c}w_1 / (theNumberOf\ v_{w\text{-}c}w_1 + theNumberOf\ v_{w\text{-}e}w_1) \quad (3)$$

$$\text{ER-of-}v_{w\text{-}e}w_1 = theNumberOf\ v_{w\text{-}e}w_1 / (theNumberOf\ v_{w\text{-}c}w_1 + theNumberOf\ v_{w\text{-}e}w_1) \quad (4)$$

$v_{w\text{-}c}w_1$ is a 2-WordCo occurrence with a causative concept; $v_{w\text{-}e}w_1$ is a 2-WordCo occurrence with an effect concept; where ($v_w \in V_{weak}$; $w_1 \in$ Noun2; v_w and w_1 are adjacent)

From Equation (1)-(4), the Verb set can be categorized by ER value into three verb group as follow.

If ER-of-$v_{s\text{-}c} \geq 0.9$ or ER-of-$v_{w\text{-}c}w_1 \geq 0.9$ then
$v_{s\text{-}c} \in$ VRC or $v_{w\text{-}c}w_1 \in$ VRC respectively
ElseIf ER-of-$v_{s\text{-}e} \geq 0.9$ or ER-of-$v_{w\text{-}e}w_1 \geq 0.9$ then
$v_{s\text{-}e} \in$ VE or $v_{w\text{-}e}w_1 \in$ VE respectively
Else $v_s \in$ VCorE or $v_w w_1 \in$ VCorE respectively.

We also determine a set of two-related events between a VE element and a VCorE element, i.e. ER of 'stress'- 'crave' occurrence as EffectOfCrave ≥ 0.9. We then apply NB to learn N-WordCo boundary/size with concepts based on three verb groups after stemming words and eliminating stop words from $VP_{EDU\text{-}i}$ of the documents. The collected N-WordCo occurrences are used for solving the VP_{EDUc}/VP_{EDUe} boundary.

C. How to Determine Feedback-Loop

With regard to the extracted cause-effect concept pair series from one document, it is necessary to determine whether there is the feedback-loop occurrence which implies to the addiction. Therefore, we apply a loop cue-word set (CW={ '*อีก/again*' '*เพิ่ม/more*'}) along with the following causative verb concepts set for feedback-loop determination ($V_{sc\text{-}loop}$={ '*ดื่ม,กิน,เสพ/consume*', '*ใช้/use*', '*ฉีด/inject*'}).

IV. A FRAMEWORK FOR PROBLEM EVENT EXTRACTION

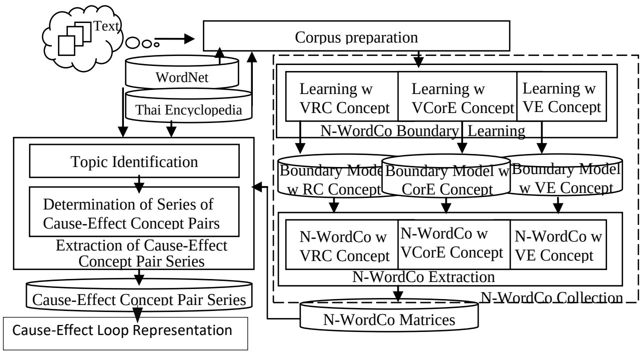

Figure 5: System Overview

There are four steps in extracting the problem events, Corpus Preparation, N-WordCo Collection, and Problem Event Extraction and Cause-Effect Loop Representation as shown in Figure 5.

A. Corpus Preparation

This step is to prepare an EDU corpus from the addiction-problem documents downloaded from hospitals web-boards. The step involves using Thai word segmentation tools [11] and Named Entity recognition [12]. After the word segmentation is achieved, EDU Segmentation [13] is then operated to provide a 2500 EDUs' corpus. The corpus included stemming words and the stop word removal is separated into 3 parts; a 1000-EDUs' part for corpus studying,i.e. ER, and learning the N-WordCo size/boundary with the causative/effect concepts. The next1000-EDUs'part is for the N-WordCo extraction. The last500-EDUs'part is for extracting the problem events. Then, we semi-automatically annotate N-WordCo concepts of VRC, VCorE, and VE on the corpus (Figure 6). All N-WordCo concepts are referred to WordNet(http://word-net.princeton.edu/) after the Thai-English translation by Lexitron (http://longdo.com)

"ยาอี……………
วัยรุ่นใช้ยาเสพติดระยะแรก [EDU1] เพราะ[เขา] อยากลองสิ่งใหม่ๆ [EDU2] เมื่อ[เขา]ได้เสพยาระยะต่อมา [EDU3] ยาเสพติดจะออกฤทธิ์ต่อสมอง [EDU4] ทำให้[เขา]มีสมาธิในการเรียนลดลง [EDU5] .."

"Ectasy………………
A teenager uses an addictive drug at first . [EDU1] **because [he] wants to try the new thing.** [EDU2] **When [he]…"**

<Topic_name Entity-concept= Ectasy/drug>ยาอี</Topic_name>…………
<EDU1 ><NP1 concept= teenager/person>วัยรุ่น/ncn </NP1>
<**VP Type=effect Group=VCorE**> <**N-Word-Co N=2** words **concept**= '**use drug**'>
< w*1*: setType='verb-weak' ; concept= 'use' boundary ='**yes**'>ใช้ </w*1*>
< w*2*: setType='Noun2' ; concept= 'drug' boundary ='**yes**'>ยาเสพติด </w*2*>
< w*3*: setType='Adv' ; concept='initial' boundary ='no'>ระยะแรก</w*3*></N-Word-Co>
</VP> </EDU1>
<EDU2><Conj concept=because>เพราะ</Conj><NP1 concept= teenager /person>ϕ</NP1>
<**VP Type=cause Group=VRC**><**N-Word-Co N=2** words **concept= 'want to try'**>
< w*1*: setType='cause-verb' ; concept= 'want' boundary ='**yes**'>อยาก</ w*1*>
< w*2*: setType='cause-verb' ; concept= 'try' boundary ='**yes**'>ลอง</ w*2*>
< w*3*: setType='Noun2' ; concept= 'thing' boundary ='no'>สิ่ง</w*3*>
< w*4*: setType='Adj' ; concept= 'new' boundary ='no'>ใหม่</w*4*></N-Word-Co></VP
</EDU2>**……**

The N-Word-Co tag is the word boundary tag of each N-Word-Co expression. The w*i* tag is the word-*i* tag where *i*=1,2,..,*num*. The **[..] symbol** or ϕ means **ellipsis** (Zero Anaphora)

Figure 6: Examples of cause-effect concept pair series annotation

B. N-WordCo Collection

This step starts with N-WordCo$_{g\text{-}i}$ boundary learning. Each annotated VP of each verb group (VRC orV_g as g=1,VCorE orV_g as g=2,VE orV_g as g=3) from the corpus preparation is used as a word feature vector ($W_{g\text{-}i}$) of N-WordCo$_{g\text{-}i}$ on$VP_{EDU\text{-}i}$ based on V_g. $W_{g\text{-}i}$is collected into a matrix vector,W_g, for N-WordCo$_{g\text{-}i}$ boundary learning.

$W_{g\text{-}i}$={$w_{g\text{-}i1}$, $w_{g\text{-}i2}$….$w_{g\text{-}ik}$ Φ/ non-Φ} as a word feature vector of N-WordCo$_{g\text{-}i}$ where 'Φ' and non-Φ are a causative concept and a non-causative concept if g=1, a cause-or-effect concept and a non cause-or-effect concept if g=2, and an effect concept and a non-effect concept if g=3 respectively; existing in EDU1,EDU2…EDU*n*.

N-WordCo$_{g\text{-}i}$=$w_{g\text{-}i1}$+$w_{g\text{-}i2}$+..+$w_{g\text{-}ik}$ (where $w_{g\text{-}i1} \in Verb_{strong} \cup Verb_{weak}$ as <u>a starting word of N-WordCo</u>; i=1,2,..n; j=2,3,..,k; g=1,2,3) on $VP_{EDU\text{-}i}$ (a verb phrase of EDU*i*).

W_g = {$W_{g\text{-}i}$} where i=1,2,..n ; Word$_g$ = { $w_{g\text{-}1}$, $w_{g\text{-}2}$….$w_{g\text{-}z}$} collected from $W_{g\text{-}i}$ elements.

With regards to W_g, after the learning corpus has been annotated V_g concepts and N-WordCo$_{g\text{-}i}$ boundary occurrences, we determine the Φ and non-Φ probabilities of $w_{g\text{-}ij}$ and $w_{g\text{-}i(j+1)}$ features by a slide window size of two consecutive words on $VP_{EDU\text{-}i}$ with the one-sliding-word distance by using Weka (http://www.cs.wakato.ac.nz/ml/weka/). Both Φ and non-Φ probabilities of $w_{g\text{-}ij}$ and $w_{g\text{-}i(j+1)}$ from each verb group are the N-WordCo$_{g\text{-}i}$ boundary model for solving the N-WordCo$_{g\text{-}i}$ size by NB, Equation (5), on the testing corpus. For testing-corpus's $VP_{EDU\text{-}i}$, if $(w_{g\text{-}i1} \in V_{strong} \wedge w_{g\text{-}i1} \in V_g) \vee (w_{g\text{-}i1} \in V_{weak} \wedge w_{g\text{-}i1}+w_{g\text{-}i2} \in V_g)$, N-WordCo$_{g\text{-}i}$ starting is occurred with Φconcept. The N-WordCo$_{g\text{-}i}$ boundary is then determined by Equation (5) with the Φ and non-Φ probabilities of $w_{g\text{-}ij}$ and $w_{g\text{-}i(j+1)}$ to determine the consecutive words on $VP_{EDU\text{-}i}$ with a slide window size of two words and having the one-sliding-word distance. As soon as the class 0 (non-Φ) is determined, the N-WordCo$_{g\text{-}i}$ boundary is ended. The extracted N-wordCo$_{g\text{-}i}$ occurrences with Φ are collected into three N-WordCo$_g$ Matrices (NWC_g) ; NWC_1:VRC-base, NWC_2:VCorE-base, NWC_3:VE-base.

$$NWordCoBoundaryClass = \arg\max_{class \in Class} P(class \mid w_{g-ij}, w_{g-i(j+1)}) = \arg\max_{class \in Class} P(w_{g-ij} \mid class) P(w_{g-i(j+1)} \mid class) P(class) \quad (5)$$

$$\begin{aligned} &\textit{where } w_{g-ij} \in Word_g \; ; \; w_{g-i(j+1)} \in Word_g \; ; \\ &\textit{and } W_{g-i} \textit{ is a word\_}\Phi\textit{concept vector of } VP_{EDU-i} \; ; \\ &i = \{1,2,..n\}; \; j = \{1,2,..k\}; \\ &\textit{if } g = 1, \textit{ then } \Phi \textit{ is based on } VRC; \\ &\textit{if } g = 2, \textit{ then } \Phi \textit{ is based on } VCorE; \\ &\textit{if } g = 3, \textit{ then } \Phi \textit{ is based on } VE. \end{aligned}$$

C. Extractions of Cause-Effect Concept Pair Series

This step is to extract the problem events from the testing corpus after a document topic name has been identified by WordNet, and Lexitron. Then, we mark N-WordCo of EDU*i* ($NWC_{EDU\text{-}i}$) having *cw* and $v_{sc\text{-}loop}$ (where $cw \in CW$; $v_{sc\text{-}loop} \in V_{sc\text{-}loop}$) as $^*NWC_{EDU\text{-}i}$. The VP_{EDUc} and VP_{EDUe} identification of the cause/effect-event-concept occurrences in the series is solved by Similarity Score [14] as MaxSimScore, Equation (6), through MaxMaxSimScore, Equation (7), between the testing-corpus's N-WordCo and the candidate N-WordCo element (NWCcandidate) from NWC_g(g=1,2,3; $NWC_1 \cap NWC_2 \cap NWC_3 = \emptyset$) (Figure 7).

$$MaxSimScore = ArgMaxSimilarity_{t=1}^{Cardinality} \left[\frac{|\, NWCcorpus \cap NWCcandidate_t \,|}{\sqrt{|\, NWCcorpus \,| \times |\, NWCcandidate_t \,|}} \right] \quad (6)$$

$$\begin{aligned} &\textit{where Cardinality is the number of } N-WordCo \textit{ elements of the } N-WordCo \textit{ Concept set or } NWC_g \; ; \; g = 1,2,3 \\ &NWCcandidate \textit{ is a candidate } N-WordCo \textit{ element of the } N-WordCo \textit{ Concept set or } NWC_g \\ &NWCcorpus \textit{ is an } N-WordCo \textit{ of EDU from the testing corpus} \end{aligned}$$

$$MaxMaxSimScore = \underset{class \in Class}{ArgMax} \, (MaxSimScore1, MaxSimScore2, MaxSimScore3) \quad (7)$$

$$\begin{aligned} &\textit{where } MaxSimScore1 \textit{ is } MaxSimScore \textit{ between } NWCcorpus \textit{ and } NWCcandidate_t \textit{ from } NWC1 \\ &MaxSimScore2 \textit{ is } MaxSimScore \textit{ between } NWCcorpus \textit{ and } NWCcandidate_t \textit{ from } NWC2 \\ &MaxSimScore3 \textit{ is } MaxSimScore \textit{ between } NWCcorpus \textit{ and } NWCcandidate_t \textit{ from } NWC3 \\ &Class = \{'root-cause', 'cause/effect', 'effect'\} \end{aligned}$$

```
Assume that each EDU is represented by(NP1 VP)after stemming words & eliminating stop words.
L is a list of EDU;   Problem-Events are the output expressed by verb phrases as   a cause-effect concept pair series (CES) based on three verb groups: VRC or Vg as g=1,   VCorE or Vg as g=2 , VE or Vg as g=3;
VP_EDU-i (a verb phrase of EDUi) is an input of the testing corpus where VP_EDU-i contains  NWC_EDU-i (N-WordCo of EDUi );   canNWC_g is a candidate N-WordCo set based on V_g.
PROBLEM_EVENTS_EXTRACTION
{i=1; RCG←∅; COEG←∅; EG←∅; COE←∅; temp←∅;
  C←∅; E←∅; CES←∅; CE←∅; flagE=0;
 While ( NWC_EDU-i.w1∈V_strong ∪ V_weak) ∧ i ≤ (Length[L]) do
  {1 If (MaxMaxSimScore(NWC_EDU-i,canNWC1,canNWC2,canNWC3)>0.9)∧(class ='root-cause')
    {2 If  RCG=∅        /* Determination of  N-WordCo basedon VRC Group
       { If  COE≠∅  then  {E← E+COE;  COE=∅  };
         RCG←RCG ∪ NWC_EDU-i }}2
     If(MaxMaxSimScore(NWC_EDU-i,canNWC1,canNWC2,canNWC3)>0.9)∧(class='cause/effect')
     {2  COEG ← NWC_EDU-i ;      /* Determination of N-WordCo basedon VCorE Group
       If  RCG=∅ ∧ C=∅ ∧ E=∅ ∧ temp ≠ NWC_EDU-i
           COE← COEG  ;                    */ It may be C1 or E1
       If  RCG≠∅ ∧E=∅ ∧ temp ≠ NWC_EDU-i
         E← COEG  ;                        */ RC1+CoE1..
       If  RCG≠∅ ∧E≠∅ ∧ temp=NWC_EDU-i ∧flagE=0   */for CE1=RC1+CoE1; C2(CoE2)…
         {CE←(RCG+E) ; CES←CES+CE; C←COEG;  E←∅;  RCG←∅ } ;
       If  RCG≠∅ ∧E≠∅ ∧ temp≠NWC_EDU-i ∧flagE=0    */ for RC1+CoE1+CoE1..
         E← E+COEG  ;
       If (RCG≠∅ ∧E≠∅∧ temp≠NWC_EDU-i ∧ flagE=1  */ for RC1+CoE1+..E1; C2(CoE2)..
         {CE←(RCG+E); CES←CES+CE; C←COEG; E←∅; RCG←∅; flagE=0}
       If (RCG=∅∧C≠∅)∧E≠∅∧temp≠NWC_EDU-i ∧flagE=1  */for C1+CoE1+..E1;C2(CoE2)..
         {CE←(C+E); CES←CES+CE; C←COEG; E←∅; flagE=0 }
       If (RCG=∅∨C=∅)∧E≠∅ ∧temp≠NWC_EDU-i∧flagE=2      */for  E2+ C2(CoE2)….
         {C← COEG ; flagE=0 }
       temp= NWC_EDU-i   }2
     If (MaxMaxSimScore(NWC_EDU-i,canNWC1,canNWC2,canNWC3)>0.9)∧(class ='effect')
      {2 EG←NWC_EDU-i ; temp←∅;   /*Determination of  N-WordCo basedon VE Group)
      If COE≠∅ ∧ C=∅ then {C←COE; COE=∅; E←E+EG };
      If (RCG≠∅∨C ≠∅) ∧ E≠∅ ∧(EG.verb∉EffectOfCravings)
        {E←E+EG ;  flagE=1 }
        Else-If RCG=∅∧ C=∅∧ E≠∅∧(EG.verb ∈ EffectOfCravings)
            {E←E+EG ; flagE=2 }
          Else-If RCG=∅∧ C≠∅∧ E≠∅∧(EG.verb ∈ EffectOfCravings)
              {E←E+EG ; flagE=2 }
            Else-If RCG≠∅ ∧ E≠∅∧ (EG.verb ∈ EffectOfCravings)
                                  */for RC1+..E1+E2ofC2
                {CE←(RCG+E);CES←CES+CE;E←EG;RCG←∅; flagE=2}
              Else-If C≠∅∧ E≠∅∧(EG.verb∈EffectOfCravings)
                                  */for C1+..E1+E2ofC2
                  {CE←(C+E); CES←CES+CE; E←EG; C←∅; flagE=2}
               Else-If RCG≠∅ ∧E=∅∧(EG.verb∈EffectOfCravings)
                                     */for RC1+E1
                     {E←EG; flagE=0 } }2
    i++   }1
   If  RCG≠∅∧ E≠∅ then CES←RCG+E;  If C≠∅∧ E≠∅ then  {CE←C+E; CES←CES+CE}
 }Return   CES
```

Figure 7: Cause-Effect Concept Pair Series Extraction Algorithm

D. Cause-Effect Loop Representation

According to each extracted cause-effect concept pair series, the * mark on N-WordCo or $^*NWC_{EDU\text{-}i}$ is searched on each cause-effect concept pair series. If $^*NWC_{EDU\text{-}i}$ is found, the MaxSimScore_ForFeedBackLoop is determined between $^*NWC_{EDU\text{-}i}$ and all previous N-WordCo occurrences from EDU*i*-1 down to EDU1 as shown in Equation (8).

$$MaxSimScore\_ForFeedBackLoop = ArgMaxSimilarity_{l=1}^{num} \left[\frac{|\, NWCstar \cap NWCprevious_l \,|}{\sqrt{|\, NWCstar \,| \times |\, NWCprevious_l \,|}} \right] \quad (8)$$

where num is the number of N-WordCo occurrences before $^*NWC_{EDU\text{-}i}$; NWCstar is $^*NWC_{EDU\text{-}i}$ *; and NWCprevious is the N-WordCo element of all previous N-WordCo occurrences from EDUi-1 down to EDU1.*

The N-WordCo element that has the MaxSimScore_ForFeedBackLoop is the starting node of the cause-effect loop and is connected to N-WordCo of EDU*i*-1 as the feedback-loop.

V. EVALUATION AND CONCLUSION

There are two evaluations of the proposed research, the N-WordCo extraction on 1000EDUs' testing corpus and the cause-effect concept pair series extraction on 500 EDUs' testing corpus. Both evaluations are based on the precisions and the recalls which are evaluated by three expert judgments with max win voting. The precisions of extracting N-

$WordCo_{g\text{-}i}$ on $VP_{EDU\text{-}i}$ (where g=1 or VRC, g=2, or VCorE, g=3 or VE) are 0.875, 0.861, and 0.848 with the recalls of 0.79, 0.78, and 0.73 respectively. The precision of the cause-effect concept pair series extraction is 0.88 with the 0.81 recall. The reason of having low recalls of both evaluations is that there are some effect event occurrences expressed by NP1 related to VP, i.e. EDUi: (‘*การเคลื่อนไหว/Movement*’)/NP (‘*ช้าลง/decelerate*’)/VP , instead of by VPs only i.e. EDUi+1 ((‘*มี/have*’)/Verb (‘*การเคลื่อนไหว/Movement*’)/NP (‘*ช้า/be slow*’)/adv)VP. The correctness of the cause-effect loop construction is 90% where the error occurs from the noun ellipsis, i.e. “*ใช้/use เพิ่ม/more*” (“Use [drug] more”). Hence, the research contributes the methodology to determine cause-effect concept pair series as the cause-effect loop for finding the root cause and the addiction occurrence. Finally, the research results as the extracted problem events, especially represented by the cause-effect loop, hold a benefit for the problem analysis to control the loop in the solving system through mobile media devices regardless to anywhere and anytime to enhance the problem analysis.

Acknowledgment

This work was supported in part by the Thailand Research Fund under Grant MU-SSIRB:2014/092(B1).

References

[1] H.De Pinho, “Systems Tools for Complex Health Systems:A Guide to Creating Causal Loop Diagrams” Mailman School of Public Health, Columbia University,New York (2015)

[2] L.Carlson, D.Marcu, M.E.Okurowski, “Building a Discourse-Tagged Corpus in the Framework of Rhetorical Structure Theory,” *Current and New Directions in Discourse and Dialogue*, vol. 22, pp. 85-112 (2003)

[3] R.Girju, “Automatic detection of causal relations for question answering,” Proc. 41st annual meeting of the assoc. for computational linguistics, workshop on multilingual summarization and question answering-Machine learning and beyond, Japan (2003)

[4] D S Chang and K S. Choi, “Causal relation extraction using cue phrase and lexical pair prob-abilities,” *In Proc. IJCNLP*, Hainan Island, China, Mar. 22-24, 2004, pp.61-70.

[5] C.Pechsiri and R.Piriyakul, “Explanation knowledge graph construction through causality extraction from texts*,” Journal of Computer Science and Technology*, Vo.25, No 5 , 2010, pp.1055-1070.

[6] S.Zhao, T.Liu, S.Zhao, Y.Chen, and J-Y. Nie, “Event Causality Extraction Based on Connectives Analysis,” *Neurocomputing 173* ,2016, pp.1943–1950.

[7] P.Mirza, and S.Tonelli, “CATENA: CAusal and TEmporal relation extraction from NAtuaral language texts. Proc.of COLING 2016, Osaka, Japan, December 11-17 2016, pp.64-75.

[8] H.Sawamaru and I.Kobayashi, “An Approach to Extraction of Causal Chain among Events in Multiple Documents,” *SCIS-ISIS 2012*, Kobe Japan, Nov20-24, pp.1104-1108.

[9] J G Smith, A J. Duncan, *Elementary Statistics and Applica-tions: Fundamentals of the Theory of Statistics*. Mc Graw-Hill Book Company Inc., New York, London, 1944.

[10] T.M.Mitchell, *Machine Learning*. The McGraw-HillCo.Inc., and MIT Press, Singapore ,1997.

[11] S.Sudprasert and A.Kawtrakul, “Thai Word Segmentation based on Global and Local Unsu-pervised Learning,” *NCSEC’2003 Proceedings*, Thailand, pp.1-8.

[12] H.Chanlekha and A.Kawtrakul, “Thai Named Entity Extraction by incorporating Maximum Entropy Model with Simple Heuristic Information,” *IJCNL’2004 Proc.* , pp. 1-7.

[13] J .Chareonsuk, T. Sukvakree, A.Kawtrakul, “Elementary Discourse unit Segmentation for Thai using Discourse Cue and Syntactic Information,” *NCSEC Proc*.,Thailand,2005, pp. 85-90.

[14] S.Biggins, S. Mohammed, S.Oakley, “University of Sheffield: Two Approaches to Semantic Text Similarity,” *In: Proc.1stJoint Conf. on Lexical & Computational Semantics*, Canada , 2012, pp655-661.