



Developing a Why–How Question Answering system on community web boards with a causality graph including procedural knowledge



C. Pechsiri ^{a,*}, R. Piriyaikul ^b

^a Department of Information Technology, Dhurakij Pundit University, Bangkok, Thailand

^b Department of Computer Science, Ramkhamhaeng University, Bangkok, Thailand

ARTICLE INFO

Article history:

Received 12 June 2015

Received in revised form

15 January 2016

Accepted 19 January 2016

Available online 22 January 2016

Keywords:

Why-Q

How-Q

Visualized answer

Integrated causality graph

ABSTRACT

The research aims to develop an automatic Question Answering system, in particular *Why* and *How* questions, on community web-boards to support ordinary people in preliminary diagnosis and problem solving, such as plant disease problems. The research includes two main problems: *Why* and *How* question identification and *Why* and *How* answer determination, where *Why* and *How* questions are based on explanations. Therefore, the research applies machine learning techniques for question type identification. We also propose an integrated causality graph with extracted procedural knowledge from text to determine the visualized answers based on the information retrieval technique. The experiment shows the Question Answering system can achieve answers at Rank 1 with 91.1% and 88.9% correctness for *Why* questions and *How* questions, respectively.

© 2016 China Agricultural University. Production and hosting by Elsevier B.V. All rights reserved.

1. Introduction

In the online community, most people prefer to post their problems or queries on a certain thread on their community's web page and then wait for times ranging from a few minutes to several days to receive the answers and recommendations made by the problem-solving experts on the web page. However, it is time consuming for people to wait for the answers. In a rural community, there are inexperienced farmers and others who know how to use information technology but lack experience in other areas, e.g. agriculture, health-care, etc. For example, on community web-boards, people with an illness try to explain their disease symptoms by asking a *Why*

question (*Why-Q*) type, asking for reasons, and/or a *How* question (*How-Q*) type, asking for a problem solving approach. However, the speed of response to questions depends on the question domain, the chat room type of a certain web-board, the web-board domain, etc. Most plant disease questions receive responses within a week through web-boards. While waiting, an automatic *Why–How* Question–Answering (QA) system could be developed to provide a preliminary diagnosis including possible solutions before or during an epidemic. Therefore, this research aims to develop a *Why* and *How* QA system based on questions that require explanation of problems, especially plant-disease symptoms, on a certain web board. The corresponding answers are the visualized as causality graphs [1] integrated with procedural knowledge extracted from texts for the preliminary diagnosis and problem solving of plant disease symptoms. There are several types of *How* question [2] e.g. Causality *How-Q* (which is used to determine the causes of a certain event: “*How did*

* Corresponding author. Tel.: +66 2 954 7300.

E-mail address: itdpu@hotmail.com (C. Pechsiri).

Peer review under the responsibility of China Agricultural University.

<http://dx.doi.org/10.1016/j.inpa.2016.01.002>

2214-3173 © 2016 China Agricultural University. Production and hosting by Elsevier B.V. All rights reserved.

John die?”), Instrumental How-Q (which is used to learn about instruments as in “How is *couscous* eaten in Morocco?”, answer: “by hand”), and Instructional How-Q (which corresponds to an organized set of instructions designed to reach a goal: “How do you change a car wheel?”), etc. However, How-Q in this research is Instructional How-Q, which emphasizes the organized instruction set for problem solving which depends on the cause of the problems/symptoms. The Why and How questions with explanations are expressed in the form of Elementary Discourse Units (where each EDU is defined as a simple sentence or a clause, [3]) with the following question patterns (called ‘Qpattern’) through the community web board.

- Qpattern-1: EDU_{ct-1} EDU_{ct-2} ... EDU_{ct-n} EDU_q
- Qpattern-2: EDU_{ct-1} EDU_{ct-2} ... EDU_{ct-n} EDU_q EDU_{ct-(n+1)}
- Qpattern-3: EDU_q EDU_{ct-1} EDU_{ct-2} ... EDU_{ct-n}

where:

EDU_q is a question EDU containing a question word (*qw*) as shown in the following linguistic pattern of a Thai-question EDU.

- EDU_q → Qword NP1 V NP2 | Qword NP1 V | NP1 V NP2 Qword | NP1 V Qword | V NP2 Qword | V Qword
- V → v_q | pre-verb v_q
- v_q → v_{q-Strong} | v_{q-weak} w_{info}
- pre-verb → ‘จะ/will’ ‘ต้อง/must’
- v_{q-Strong} → ‘ทำ/solve’ ‘แก้/solve’ ‘แสดง/express’ ‘เกิดจาก/be caused by’ ‘แห้ง/dry’ ‘ร่วง/come off’ ‘แคะแกรน/stunt’ ‘หึง/change shape’ ...
- v_{q-weak} → ‘เป็น/be’ ‘มี/have’
- w_{info} → ‘อาการ/symptom’ ‘แผล/mark’ ‘สี/color’ ‘เพราะ/reason’ ‘สาเหตุ/cause’ ‘ผลลัพธ์/result’ ...
- Qword → {‘ทำไม/Why’ ‘อย่างไร/How’ ‘อะไร/What’ ‘แสดงวิธี/Show method’}

(where Qword is a question-word set and *qw* ∈ Qword; v_q is a verb concept expressed on EDU_q; NP1 and NP2 are noun phrases.)

EDU_{ct-a} is a content EDU expressing a content of EDU_q, where a = 1, 2, ..., n or n + 1. n is an integer number and is greater than 0. EDU_{ct-a} has the following Thai linguistic pattern.

- EDU_{ct-a} → NP1 VP
- VP → v_{ct-a} NP2 | v_{ct-a} | v_{ct-a} AdjectivePhrase | pre-verb v_{ct-a} NP2 | pre-verb v_{ct-a} | pre-verb v_{ct-a} AdjectivePhrase

(where v_{ct-a} is a causative verb concept (v_c) or an effect verb concept (v_e) as shown in Table 1 (v_c ∈ V_c; v_e ∈ V_e; V_c and V_e are a causative verb concept set and an effect verb concept set, respectively)).

Moreover, the Thai documents have several specific characteristics, such as zero anaphora or implicit noun phrases, without word delimiters, without sentence delimiters (e.g. without a question mark), etc as shown in Fig. 1.

All of these characteristics are involved in determining the question type and its answer in the Why-How QA system of this research based on Qpattern, which contains several EDUs as explanations. It attempts to determine the answer with Qpattern, whilst previous QA researches, especially on Why-How QA systems, were based on one or two EDUs. It also attempts to answer a How-Q which expresses only the sequence of events of the effect/symptom EDUs without mention of their cause. In this research, the How-Q expression results in diagnosing the effect/symptom events before determining the solution whereas previous How-Q researches are based on direct instruction guidelines or an event description graph without including problem/symptom diagnosis.

Table 1 – List of V_c and V_e provided by [1].

Verb type		Surface form	Conceptual class
V _c (Causative-Verb Concept set)	Strong Verb	ดูด/suck, ตูดกิน/suck. กิน/eat, กัด/bite, ทำลาย/destroy, กำจัด/eliminate, ฆ่า/kill, หัก/break,	consume/destroy destroy
	Weak Verb + Noun or Information	เป็น + โรค/be + disease, ได้รับ + เชื้อโรค/get + pathogen, ...	getDisease getPathogen ...
V _e (Effect-Verb Concept set)	Strong Verb	หึง/shrink, งอ/bend, บิด/twist, โคลงง/curl	be_abnormal_shape
		แห้ง/dry, ไหม้/blast, เหี่ยว/wilt	dry/be_symptom lose_water/be_symptom
	Weak Verb + Noun or Information	แคะแกรน/stunt เป็น + จุด/be + spot, เป็น + ขีด/be + scratch, เป็น + แผล/be + lesion มี + จุด/have + spot, มี + ขีด/have + scratch, มี + แผล/have + lesion มี + สี + น้ำตาลไหม้/have + color + dark brown ...	stunt/be_symptom be_spot_mark/be_symptom, be_scratch_mark/be_symptom be_mark/be_symptom have_spot_mark/have_symptom have_scratch_mark/have_symptom have_mark/have_symptom have_brown_color/have_symptom ...

Download English Version:

<https://daneshyari.com/en/article/4493311>

Download Persian Version:

<https://daneshyari.com/article/4493311>

[Daneshyari.com](https://daneshyari.com)