Online Multilingual Translation of Technical Service Reports over the World Wide Web

S. Liu, S.C. Hui, S. Foo and P.C. Leong
School of Applied Science, Nanyang Technological University
Nanyang Avenue, Singapore 639798

Abstract

In a typical customer service support cycle of a multinational manufacturing company, technical service reports are used to record customer reported machine problems and a corresponding set of proposed remedies or suggestions taken to rectify the problems. These reports are stored in English in a customer service database. Service engineers and customers are able to access the database and retrieve the technical service reports via the World Wide Web. In order to support a worldwide customer base to display the technical reports in the customers' own language, a provision of online multilingual translation of technical reports from English to other languages is necessary but yet avoiding the need to create other separate native language-based customer service databases.

This paper discusses an approach to the online multilingual translation of technical service reports. As a result of the technical vocabulary of the contents, grammatical and syntactical differences between English and other languages, a simple word-for-word translation approach is not feasible since it will almost certainly yield incomprehensible or useless results. Thus, the approach taken consists of two main processes, namely, Technical Report Analysis and Online Translation. Technical Report Analysis analyses the existing customer service reports to generate simplified common syntactic structures from which translation tables are created to map the English component to the corresponding language counterparts. During Online Translation, customer service reports are retrieved, translated online and displayed on the Web browser. The approach has been applied successfully to both the English-to-Japanese and English-to-Chinese translation and it can be extended to other languages as well.

Keywords: Multilingual online translation, report analysis, World Wide Web
1. INTRODUCTION

Technical service reports are used to record reported machine problems and proposed remedies or suggestions taken to rectify these problems during a customer service support cycle in a multinational manufacturing company that supplies insertion and surface mount machines. As the reporting is subjected to the service engineers' vocabulary, no standard format or vocabulary is used to record these service reports. These service reports are stored in a customer service database for future customer support purposes. Each service report contains the customer information and service details. Service details are stored in two fields: fault-condition and checkpoint. The fault-condition field contains customer's description about their faulty machines. The checkpoint field indicates the suggested actions or services to be carried out to repair the machine into normal condition based on the occurred fault-conditions given by the customer. Service engineers and customers are able to access the database and retrieve the technical service reports via the World Wide Web (or Web).

The service reports serve as a record of the problems as well as provide ‘knowledge’ into tackling future similar problems for its worldwide customers and service support engineers. However, as the service reports are stored in English, they are retrieved and displayed in English accordingly. This often poses problems for many of their worldwide customers including their Chinese and Japanese customers who find difficulty in using the system. Therefore, in order to support its worldwide customers to tap the expertise provided by the system, a provision of online multilingual translation of service reports from English to other languages is necessary but yet avoiding the need to create other separate native language-based customer service databases. As customer information is static and easy to understand in English, only the service details of the fault-condition and checkpoint fields of the technical service reports are translated.

In this paper, we propose an approach to the online multilingual translation of these technical service reports based on a table look-up mechanism. In this research, we focus on the English-to-Japanese and English-to-Chinese translation. The paper is organised as follows. Section 2 presents the proposed approach on the multilingual translation of customer service reports. An overview of the system is given in section 3. Section 4 discusses the multilingual translation process that consists of two main steps: Technical Report Analysis and Online Translation. Section 5 discusses the Technical Report Analysis. The Online Translation is presented in section 6. Finally, conclusions are given in section 7.

2. PROPOSED APPROACH

Machine translation techniques [1-6] have been traditionally applied to English-to-Japanese and English-to-Chinese translation. Recently, machine translation systems for English-to-Japanese [7, 8, 9] have been developed for Internet or Web users. The machine translation systems translate English information on the Web from a Web browser into Japanese with the original layout. By executing the translation program while receiving HTML documents, the translation systems give the users the impression of real-time translation. Japanese users can thus browse the Internet as if all home pages were written in Japanese.

However, machine translation techniques are unsuitable for translating service
reports because of its technical nature and non-standard grammatical syntax. Therefore, we propose another method that uses translation tables for English-to-Japanese and English-to-Chinese translation of technical service reports. The approach starts by analysing all existing service reports to identify common structures for its content. Based on the new structures, translation tables can be created. Finally, translation can be carried out dynamically and translation results will be displayed directly on the Web browser’s user interface.

With the common structures, all service engineers can subsequently adopt them to input their future service reports. However, to make this method effective, the proposed approach will limit the vocabulary currently used in the service reports. Otherwise, the translation tables will be too big and the performance of the system will be degraded. As the common structures will be derived from the existing service reports that were created by service engineers previously, the vocabulary and sentence structures identified will be similar to the existing reports and be familiar to most service engineers. Therefore, the final structures would not be too restrictive and should pose no problems for service engineers to use and follow in their future input. In addition, with this approach, deep-level analysis of syntax and semantics between the English and other languages can be avoided.

3. SYSTEM OVERVIEW

In this research, a Web-based system is developed to support online retrieval of technical reports from the customer service database. In addition, multilingual translation capability is also incorporated into the system. Figure 1 shows the system overview. The system is developed on a Windows NT environment. The Netscape Enterprise Server 3.0 [10] is used as the Hypertext Transfer Protocol (HTTP)[11] server. The Retrieval Engine and Multilingual Translation Program are written in C and can be executed and linked to the Enterprise Server via Common Gateway Interface (CGI) [12]. The Microsoft Access
database management system is used to store the customer service database. The CGI programs communicate with the Access database system through Open Database Connectivity (ODBC) [13] to insert, delete or update of the customer service database. Hypertext Mark-up Language (HTML) [14] is used to create the user interface as Web pages to accept user retrieval query. The results of the query will be formatted and then displayed.

As shown in Figure 1, customers can access the customer service database via any Web browser such as Netscape's Navigator [15] or Microsoft's Internet Explorer [16]. The customer can interact with the system in many ways. One common method is to define fault-conditions on a user retrieval form. The user retrieval form is a user interface that allows a user to select the fault-conditions for their machines. After the user specified the type of machine and the corresponding fault-conditions, the system searches the corresponding checkpoints from the customer service database through the Retrieval Engine. The retrieved checkpoints are then passed to the client Web browser for display through the Internet. If multilingual translation is enabled, the fault-conditions and checkpoints will be translated into the specified language code using the translation tables by the Multilingual Translation. The translated language code will then be displayed on the client Web browser using multilingual display software such as TwinBridge [17] or WinMass [18].

4. MULTILINGUAL TRANSLATION

The multilingual translation process consists of two main processes as shown in Figure 2. The first process is Technical Report Analysis, while the second process is Online Translation. Technical Report Analysis analyses existing customer service
reports, and in particular the fault-condition and checkpoint fields, to generate common syntactic structures. Then, translation tables of each component of the common structures will be created which will map the English component to the corresponding language counterparts. As discussed before, this research will focus on English-to-Japanese and English-to-Chinese translation. Therefore, the corresponding Japanese and Chinese translation tables are created.

During Online Translation, customer service reports will be retrieved based on the customer's input. The corresponding fault-conditions and checkpoints will be parsed to the common structures identified by the Technical Report Analysis process. Translation tables will then be used to generate the specified language counterparts (i.e. Japanese or Chinese) of the retrieved fault-conditions or checkpoints. Finally, the customer service reports will be displayed in Japanese/Chinese through a multilingual display software on the Web browser.

5. TECHNICAL REPORT ANALYSIS

As fault-conditions and checkpoints are translated during the Online Translation, all the existing fault-conditions and checkpoints have been analysed in order to understand their internal structures. As the two processes are very similar, we will focus on discussing the checkpoint analysis process to illustrate the underlying concepts.

5.1 Checkpoint Analysis

The purpose of checkpoint analysis is to identify a common structure for checkpoints from all existing checkpoints stored in the customer service database. With this structure, translation tables can be generated. Before discussing the checkpoint analysis process, some general features of the existing checkpoints are observed and described as follows:

- Most checkpoints possess the following structure of:
  
  \[ \text{VERB} + \text{OBJECT} \]

  where OBJECT may be nouns or noun-phrases. Examples of this structure include:

  1. \( \text{ADJUST} \ \text{SERVO DRIVER} \rightarrow (\text{ADJUST} + \text{SERVO DRIVER}) \)
  2. \( \text{CHECK MACHINE INITIAL SETTING} \rightarrow (\text{CHECK} + \text{MACHINE INITIAL SETTING}) \)
  3. \( \text{CHECK THE BINARY LEVEL SETTING} \rightarrow (\text{CHECK} + \text{THE BINARY LEVEL SETTING}) \)
  4. \( \text{RETEACH} \ \text{MARK DATA} \rightarrow (\text{RETEACH} + \text{MARK DATA}) \)
  5. \( \text{CHANGE POWER SOURCE BOX} \rightarrow (\text{CHANGE} + \text{POWER SOURCE BOX}) \)

- For some checkpoints, they have the structure of:

  \[ \text{VERB} + \text{OBJECT} + \text{CONDITION (ADVERB, NOUN or ADJECTIVE)} \]

  where CONDITION indicates the state of the OBJECT. For example:
(1) TURN POWER OFF ONCE -> (TURN + POWER + OFF ONCE)

(2) CHECK WIRES FOR CONTINUITY -> (CHECK + WIRES + FOR CONTINUITY)

(3) CONFIRM SENSOR, DIRTY OR DAMAGED -> (CONFIRM + SENSOR + DIRTY OR DAMAGED)

(4) ATTEMPT RECOGNITION AGAIN -> (ATTEMPT + RECOGNITION + AGAIN)

- A few checkpoints have a structure containing only a VERB. For example:
  
  (1) EXIT
  
  (2) RESET
  
  (3) RETRY

- Some checkpoints may have complex structures. Most of these checkpoints can be further separated into one or more individual checkpoints. Examples of these checkpoints include:

  (1) TURN OFF THE POWER SOURCE AND CHECK THE CYCLE TIMER, ANGLE SETTING AND WIRING

  (2) IN THE PRODUCT MANAGEMENT INFORMATION, THIS IS INCLUDED IN COMPONENT RECOGNITION ERROR

  (3) TURN OFF THE POWER AND CHECK IF THE RECOGNITION BOARD IS SET IN THE CONTROLLER CORRECTLY

  (4) FOR ORIGIN RETURN, IT MUST NOT TAKE MORE THAN 20 SECONDS

- Some checkpoints may contain reference information which are enclosed by brackets. For example:

  (1) CONFIRM THE NC CARD (LAM 00205)

  (2) CHECK THE WIRING CONNECTION OF THE ENCODER (MOTOR -> DRIVER -> NC CARD)

  (3) CONFIRM THE NC CARD (783-203)

  (4) CONFIRM THE ROMS IN THE MAIN CPU CARD. (LAM 00001)

These observations will form some of the heuristic concepts used for the checkpoint analysis process.
Figure 3. Checkpoint Analysis Process

The idea behind the checkpoint analysis process is as follows. Given a set of checkpoints, the process will perform a sequence of steps based on some heuristic concepts discussed above to eliminate or remove the existing checkpoints (or components of existing checkpoints) until no more or the smallest set of checkpoints remains. In this way, the structure of the checkpoints can be identified from the heuristic concepts used.

Figure 3 illustrates the checkpoint analysis process. A total of 4851 checkpoints have been analysed. The analysis process is carried out as follows. Firstly, all repeated checkpoints are eliminated. Subsequently, the reference information, which can form part of the common structure, but need not be translated is identified and extracted. Next, non-alphabetical characters are removed since they do not need to be translated and can be displayed in its original form during translation. Removing them from existing checkpoints will not affect the analysis of report structure. As some checkpoints have complex structures which contain multiple checkpoints information connected using logical operators such as "AND", "OR" or "TO", these checkpoints will be identified and then separated into individual single checkpoints.

The next step is to identify conditions within checkpoints. The conditions will be
used to form part of the common structure. Stop words and technical words in short form will be removed during analysis. However, in the display of the English text, the original form of fault conditions and checkpoints in the customer database is preserved so that stop words and technical short form words will be displayed in its original form. In the Japanese or Chinese translated text, stop words do not exist while technical short form words are displayed in the original form. Finally, verb and verb phrases, and noun and noun phrases are identified which will form the core information in the common report structure. The details of each process is elaborated further:

**Eliminate Repeated Checkpoints**

This process removes all repeated checkpoints from the existing checkpoints to generate a set of unique checkpoints for analysis. This process is repeated after each stage of analysis to ensure that the smallest unique set of checkpoints is used for the next stage of analysis.

**Identify References**

References that are enclosed by brackets usually contain reference information. The format of reference information is usually in alphanumeric word format that need not be translated. During translation, reference information can be displayed in its original format. Of course, if there is a need to translate this reference information, it still can be done by creating another translation table based on existing references. But as the format of reference is quite unstructured, the reference would not be translated in this research.

**Remove Non-Alphabetical Characters**

We found that there are many non-alphabetical characters in the existing checkpoints. These include digital characters ("0" - "9"), "#", ",", ",", etc. These characters are not essential for translation as the same set of characters is available in Chinese. So during translation, these characters can be displayed in its original form. These characters include:

```
"=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" "=" 
```

Another two characters "," and ":" which are used to separate words is replaced by a space " " during checkpoint analysis. There is no change in meaning as the space used in the existing checkpoints represents a word separator.

**Separate Complex Checkpoints**

Some of the checkpoints contain logical operators such as ",", "AND", "OR" and "TO". These checkpoints need to be separated into individual single checkpoint. This is straight-forward if the checkpoints use logical connectors such as "AND", "OR" and "TO". However, the period connector "," can have two meanings. First, it can be part of a word. Second, it can also be a logical separator for checkpoints. In the first case, if it appears in front of a digit, it can be part of a number; if it is between two alphabetical characters, it may be part of a word, e.g. "L.STOP". In the second case, if it appears in front of a character space " " and a verb after the space, then it is a logical separator. The "," should then be interpreted accordingly. An example is given as follows:
Checkpoint: CHECK COMPONENT PICK UP POSITION. MAKE SURE IT IS PICK AT THE CENTER OF THE COMPONENT

can be separated into

Checkpoint 1: CHECK COMPONENT PICK UP POSITION

Checkpoint 2: MAKE SURE IT IS PICK AT THE CENTER OF THE COMPONENT

If the checkpoint contains "TO" and after it followed by a verb, then we can separate it into two different checkpoints. For example:

Checkpoint: CORRECT THE NC DATA AFTER CHECKING THE NC PROGRAM EDITING DISPLAY TO MAKE SURE Z NOS ARE WITHIN THE Z Z RANGE

can be separated into

Checkpoint 1: CORRECT THE NC DATA AFTER CHECKING THE NC PROGRAM EDITING DISPLAY

Checkpoint 2: MAKE SURE Z NOS ARE WITHIN THE Z Z RANGE

If the checkpoint contains "AND" and is subsequently followed by a verb, then we can separate it into two different checkpoints. For example:

Checkpoint: CHECK THE MACHINE INITIAL SETTINGS AND MAKE SURE THE ORIGIN OFFSET X DATA IS CORRECT

can be separated into

Checkpoint 1: CHECK THE MACHINE INITIAL SETTINGS

Checkpoint 2: MAKE SURE THE ORIGIN OFFSET X DATA IS CORRECT

At the end of this stage of analysis, the number of single checkpoints will increase as a result of this separation.

**Identify Conditions within Checkpoints**

Some checkpoints contain the condition part that is used to describe the main body. Checkpoints containing keywords such as "BETWEEN", "IS", "WAS", "ARE" and "BE" will have the condition parts. Once these keywords are identified, we can separate the conditions from the main body accordingly. For example:

(1) ENSURE AIR PRESSURE IS SATISFACTORY

can be separated into

ENSURE AIR PRESSURE----SATISFACTORY

(2) CONFIRM FEEDER UNITS ARE IN GOOD CONDITION

can be separated into

CONFIRM FEEDER UNITS ---- IN GOOD CONDITION

(3) ENSURE THE LIGHT IMAGE CAN BE CAPTURE BY THE CAMERA
can be separated into

ENSURE LIGHT IMAGE--- CAPTURE BY THE CAMERA

(4) CHECK THE CONNECTION BETWEEN THE LOADER CONVEYOR VERTICAL TRANSFER VALVE AND
THE OPTICAL OUTPUT MODULE

can be separated into

CHECK CONNECTION---BETWEEN THE LOADER CONVEYOR VERTICAL TRANSFER VALVE AND
THE OPTICAL OUTPUT MODULE

Remove Stop Words and Short Form Words

As stop words (such as "THE", "IS", "OF", etc.) are not used for translation, these
words can be removed from the checkpoints. A stop-list is used to remove these stop
words. Moreover, there are many technical words in short form such as "NC", "SC",
"I/O", etc. These words have their own technical meaning, but they do not need to be
translated. In this process, these short form words can be treated as stop-list words. By
using the look-up tables for stop-list and short form word list, they can be removed from
the existing checkpoints during the checkpoint analysis. However, the short form words
will be retained in its original form during translation.

Identify Verb and Verb phrases

At this stage, most checkpoints' format will be "VERB [PHRASE] + NOUN
[PHRASE]". In most checkpoints, the first word of each checkpoint represents a verb. A
total of 194 distinct verbs are identified. In addition, there are cases where verb phrases
occur. For example, it is found that "SURE" always follows "MAKE" in all the
checkpoints, and hence "MAKE SURE" is a verb phrase. The phrase "MAKE SURE"
will be treated as a single verb. Another example is "WIPE OFF" which is also a verb
phrase. Other common examples of verbs used in the checkpoints are ALLIGN, APPLY,
ACTIVATE, ADJUST, CLOSE, CONFIRM, CONNECT, CHECK, PULL, PUSH, etc.

However, there are still some existing checkpoints without containing a verb in
their first position. As it only occupies a small percentage (about 4%), these checkpoints
will be ignored in this analysis.

In addition, it was found that there are many verbs and verb phrases that are
synonyms, such as CONFIRM, MAKE SURE, VERIFY, etc. As a result, it becomes
possible to reduce the translation look-up table by standardising the verb and verb
phrases by eliminating all synonyms. This will allow the VERB set to be reduced to the
absolute necessary minimum.

Identify Noun and noun phrases

After the verb or verb phrases are trimmed off from the checkpoints with verb in
the first position, almost all of the remaining checkpoints contain nouns or noun phrases.
As most of the noun phrases are still quite complicated, further analysis must be carried
out to get the noun phrases as simple as possible. There are also many situations such as
"BETWEEN THE LEFT SAFETY SENSOR AND THE OPTICAL INPUT MODULE"
which are composed of noun phrases. They can be separated into simple noun phrases.
For example:

Checkpoint: CHECK THE CONNECTION BETWEEN THE LOADER CONVEYOR VERTICAL TRANSFER VALVE AND THE OPTICAL OUTPUT MODULE

gives two noun phrases:

"LOADER CONVEYOR VERTICAL TRANSFER VALVE"

and

"THE OPTICAL OUTPUT MODULE"

This conversion is carried out in the form of:

BETWEEN + NOUN(OR NOUN PHRASE) + AND + NOUN(OR NOUN PHRASE)

Moreover, the noun or noun phrases may contain stop words and short form words. As in the previous process, stop words are not used for translation and hence removed from further analysis. Likewise, technical words in short form have their own technical meaning but do not need to be translated. They are also removed from further analysis. However, these short form words will be retained in its original form during translation. At this stage, a total of 2486 different noun and noun phrases were extracted. Some of the most frequently appeared nouns are AXIS, CARD, WIRING, CONTROLLER, CUTTER, NOZZLE, GAP, MOTOR, TABLE, MACHINES, etc.

**Final Structure**

From the 4,851 checkpoints analysed, we found that approximately 95% of the checkpoints have the following common structure:

VERB [PHRASE] [+ NOUN [PHRASE]] [+ CONDITION] [+ REFERENCE]

With the common structure, a checkpoint translation table can be created to store the English component and the corresponding language (Japanese or Chinese) component. However, about 5% of the existing checkpoints do not follow the common structure. This is mainly due to the fact that there is no standard format and vocabulary to record the service reports. The remaining 5% checkpoints are required to be converted and re-written according to the common structure before they can be translated.

### 5.2 Fault-Condition Analysis

Similarly, the fault-condition analysis process has been carried out to analyse the existing fault-conditions. A total of 9,319 fault-conditions are analysed from the customer service database. As a result, a common structure for most of the fault-conditions (92%) has been derived as follows:

SUBJECT [+ VERB [PHRASE]] + STATE (ADVERB, ADJECTIVE or NOUN)

### 6. ONLINE TRANSLATION

Customers or service engineers interact with the system via client Web browsers. If online multilingual translation is enabled, the retrieved fault-conditions and checkpoints will be parsed in the same way as the fault-condition analysis and checkpoint
analysis processes into the common structures discussed in the Technical Report Analysis process. Based on the translation tables, the Japanese/Chinese counterparts of the fault-conditions and checkpoints are looked-up and sent to the client Web browser for display.

In the client machine, multilingual display software such as WinMass is initially started to run as a background process. When the translated fault-conditions and checkpoints are passed to the client Web browser, the corresponding Japanese/Chinese information will be displayed. In Figure 4, when the customer specifies a fault-condition, the corresponding original checkpoints in English are retrieved from the customer service database and displayed. In Figures 5 and 6, the corresponding checkpoints are translated into Japanese and Chinese accordingly following the translation process of Figure 2 when the online translation of Japanese and Chinese is enabled. It is evident from Figures 5 and 6 that stop words in the original English text are eliminated while short form technical words are retained in the translated text.
7. CONCLUSION

This paper discusses an approach to online multilingual translation of customer service reports over the World Wide Web environment. The research focuses on both online English-to-Japanese and English-to-Chinese translation. The approach consists of two processes: Technical Report Analysis and Online Translation. Technical Report Analysis is an off-line process that analyses existing service reports to identify common

Figure 5. Online Japanese Translation of the Checkpoints

Figure 6. Online Chinese Translation of the Checkpoints
structures for its fault-conditions and checkpoints. The Online Translation translates retrieved fault-conditions and checkpoints into Japanese/Chinese code, transfer them to the client and display the fault-conditions and checkpoints in Japanese/Chinese on a Web browser with the aid of the multilingual display software WinMass. This approach is effective as it does not need to have deep-level understanding of syntax and semantics between the languages. In addition, the approach can also be extended to other translation if the translation tables are created for other languages.

REFERENCES


