Applying Unified Modeling Language (UML)-Use Case Diagram into a Developing Prototype Malay Grammar Checker System

Suhaimi Ab. Rahman

College of Information Technology (CoIT)
Universiti Tenaga Nasional (UNITEN),
KM7, Jalan Kajang-Puchong,
43009 Kajang,
Selangor, MALAYSIA
smie@uniten.edu.my

ABSTRACT

A grammar checker for Malay is a system that will develop to do automatic checking for incorrect grammatical structure in Malay sentences. At present, most of the task for checking grammar in Malay sentences has done manually. There is no other available commercial product/system on the market for Malay Grammar Checker, particularly well adapted toward the requirements of teaching and learning this language in a primary and secondary school. The developing potential prototype of Malay grammar checker was planned in earlier phases with considering this system is very essential and useful used as problem solvers detect errors by giving designate results for an incorrect grammar in Malay sentence. To support this research work, this paper provides a description of analysis and design using Unified Modeling Language (UML) – Use Case diagram as described in (Craig larman,1999; John W.Satzinger et al.,2005) which is contemplated as a standard modeling notation framework for implementing object oriented development concepts in developing a system. The UML – Use Case diagram can insist on a precise estimation of the project and understand system requirements, specifically in managing projects or modules for all phases of the system development life cycle.

Keywords
Malay Grammar Checker, Use Case Diagram, Part Of Speech (POS), Malay Grammar Rules base

1.0 INTRODUCTION

The Malay language is spoken by over 300 million people and considered the fourth largest language group as discussed in (Asian Regional Exchange for New Alternatives (ARENA), 2007). According to Dewan Bahasa dan Pustaka (DBP)[1], there are roughly 118 institutions throughout the world teaching and conducting research on Malay language.

There is a need of electronic language tools to carry out extensive research on Malay language to match increased research activity in universities particularly from Asean countries, France, Britain, Netherlands and Australia where Malay is taught and researched. Currently, no commercialising tools released to do automatic checking in a Malay Grammar language sentence. As of this issue, we have taken an initiative to form a research group to work on this Research and Development (R&D) project, which is principally concentrating on the following objectives:

- To do a comparative study on the workable similar product in the growing market.
- To act comprehensively research on acceptable methods or techniques that might be applied into our Malay Grammar language checker.
- To design and develop our own prototype localized Malay Grammar checker system that can formally recommend corrections.
- To evaluate and test the reliability of prototype system developed.

Further to our discussion, we will organize this paper as follows: Section 2 shortly discusses on system development methodology. Section 3 briefly describes our system overview. Section 4 demonstrates system analysis and design using Unified Modeling Language (UML)-Use Case diagram. Finally, section 5 concludes the paper and future works.

2.0 SYSTEM DEVELOPMENT METHODOLOGY

Unified Process (UP) is a system development methodology that will apply into our system development. As explained in (Jacobson I. et al., 1999; Jacobson I. et al., 1997) the concept of Unified Process is derived from an iterative approach. This approach will complete the progress modules by well executing in several iterations including analysis, design, implementation and finally applies them into the

[1] Dewan Bahasa dan Pustaka (Malay for The Institute of Language and Literature) (abbreviated DBP) is the government body responsible for coordinating the use of the Malay language in Malaysia and

operational environment. Each iteration is inside four phases from UP. The four phases are known as inception, elaboration, construction and transition explained in further details below.

2.1 Inception Phase

As described in (Rational Unified Process: Inception phase, 2003), the primary objectives of the Inception phase include:

- Establishing the project's software scope and boundary conditions.
- Estimating the overall cost and schedule for the entire project.
- Estimating potential risks.
- Preparing the supporting environment for the project.

As referred in our initial report, all the basic requirements needed for developing a Malay grammar checker system is collected wholly by considering through all the objectives as pointed out above.

2.2 Elaboration Phase

As deliberated in (Rational Unified Process: Elaboration phase, 2003), the goal of the elaboration phase is to baseline with architecture of the system. It is providing a stable basis for the substance of the design and implementation work in the construction phase. Essential activities take into:

- Refining the Vision.
- Creating and base lining detailed iteration plans for the construction phase.
- Refining the development case and putting in place the development environment.
- Refining the architecture and selecting components.

In the elaboration phase, a framework of system architecture is built relating with the scope, size, risk, and novelty of the project.

In our data requirement activities, we have formed the data from children's storybooks. All the words composed were done manually. The ratified words then will be analysed and allotted with Part-Of-Speech (POS), such as ***kata nama*** (noun), ***kata kerja*** (verb), ***kata adjektif*** (adjective) and ***kata tugas*** (adverb). To address POS of words, we have referred in (Abdullah Hassan, 2004; Abdullah Hassan et.al., 2004; Abdullah Hassan, 2007) on how to suit POS and rules for each word in Malay sentence. Even so, the Linguist effort also needed to substantiate the right in Malay grammar structure before inserting into a database.

2.3 Construction Phase

At present, we did not pleasantly reach this phase. As outlined in (Rational Unified Process: Construction phase, 2003), the goal of the construction phase is refining the persisting requirements and perfecting the development of the system based upon the baseline architecture. The principal activities of the construction phase consist of:

- Resource management, control and process optimization.
- Overall evaluation of product releases.

All related resources are required and would appropriately be overseen to make sure the implementation of the system is running on planned.

2.4 Transition Phase

To begin with, we have also planned overall architecture for our deployment process. The purpose of this phase is to ensure the software developed is attainable for running. As described in (Rational Unified Process: Transition phase, 2003) relevant activities in this phase are presented as follows:

- Execution deployment plans.
- Thoroughly testing a product at the development site.
- Setting up a product release.
- Acquiring user comment.
- Refined the product based on feedback.

All these activities have to work out much earlier at the preface of iterations within UP process.

3. OVERVIEW OF MALAY GRAMMAR CHECKER SYSTEM

Further to our discussion in Malay Grammar Checker system, we will conveniently divide this discussion interested in three main categories:

3.1 Malay Grammar Checker System Flow Process

The system comprises of three main components containing database, Malay Grammar checker engine and user interface. Figure 1 indicates the relationships surrounded by all these components.

Input sentences for the system are Malay Language. All the input sentences are required to pass through a pre-processing module to filter unrecognized symbols that might be inner the text. After the pre-processing completed, the system will move inside to a Malay grammar checker engine module in deciding a correct grammatically in an input sentence. If the grammar structure complies with the rules stored in a database, then the system will immediately notify that the grammar structure is correct. Otherwise, the system will tentatively recommend a correct grammar suitably with the rules taken out from a Malay grammar rule database. A grammar guideline is also provided to give ideas on the alterations that will be impressively made. In the end of the process, pre-post editing task can be done to get better accurate grammar structure. This effort is definitely required in order to obtain a correct result for upcoming grammar checking process.

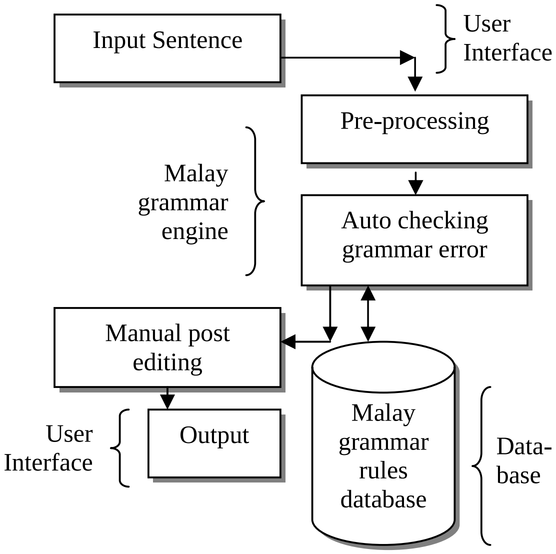

Figure 1: System components

3.2 Database

Database is specifically designed to provide an organized more effective mechanism for storing, managing and retrieving information from a large collection of data. The running of a computer system can impetuously decide on data based on input from the user. Our Malay Grammar Checkers system can absolutely assimilate with a MySQL database and SQL (Structure Query Language) script.

At present, our database consists of 500 Malay POS-words and 16 rules. Table 1 depicts a typical example of compound word categorizations structure obtaining from a Malay grammar database.

Table 1: Word categorization table

No	Lemma word	POS	Description
1	keluar	KK	Kata Kerja
2	bunga	KN	Kata Nama
3	cantik	KA	Kata Adjektif
4	mengejar	KK	Kata Kerja
5	di	KS	Kata Sendi

3.3 Malay Grammar Checker Engine

Malay grammar checker engine will develop to ensure that a grammar structure of an input sentence is correct. To work up this engine, we finally decided to use JAVA programming language. This programming language is originally designed for the system developer to create a software application easily. The software applications come up with this language can be running with a single computer or in a distributed network. The basic concept of an object-oriented programming model also will turn on inside our coding.

4. DESIGNING A USE CASE DIAGRAM

Use case is a powerful tool to describe a functionality of a system. Use case diagram help to model structure of a system, which can assist to identify and recognize an information wish on the system.

Pertaining with our system, beneath is a proper description on the significant of using Use Case diagrams that have been formed into our Malay Grammar Checker System.

4.1 Use Case Diagram: Construction Data

Use Case diagram in Figure 2 shows the task performs by user and linguist. The list of the tasks such as registering a new word, manipulating word, creatively constructing grammar rules and browsing word should be assimilated together in preparing and constructing data for Malay Grammar Checker System.

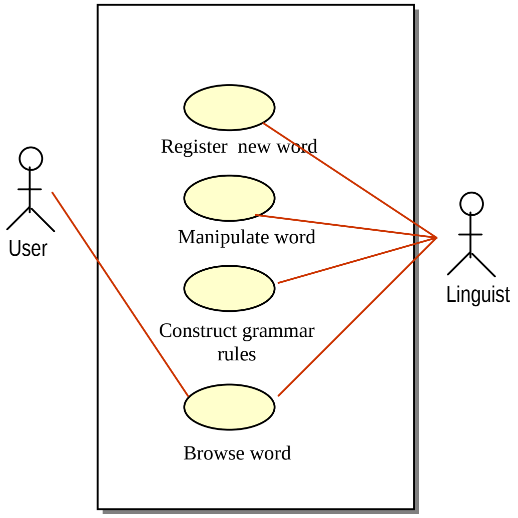

Figure 2: Use case diagram for construction data

i.) A Use Case "Register New Word" - This use case provides a capability to add recent words in Malay Grammar database.
ii) A Use Case "Manipulate Word"- This use case specifies the features to manage words stored in a database.
iii) A Use Case "Construct Grammar Rule"- This use case definitely wishes linguist to confirm and formulate rules for the word categories and grammatical orders.
iv) A Use Case "Browse Word"- This use case enables user and linguist to leaf through words alongside morphological and lexical information.

4.2 Use Case Diagram: Grammar checker

A use case diagram in Figure 3 depicts seven Use Cases used by user and linguist for easily accomplish the process of checking grammar in Malay sentence.

i) A Use Case "Make Query" - This use case endures user or linguist to insert one sentence into a text box provided and correspondent the submit button to complete a grammar checking process.
ii) A Use Case "Check Query" – This use case will satisfy automatic thoroughly checking grammatical errors in an input sentence.
iii) A Use Case "View Error" - To sight unpredictably error appeared after a Malay Grammar checker perfected the process of checking grammatical errors in an input sentence.
iv) A Use Case "View Solution" - This use case offers a view solution to get well correct word(s) for the incorrect grammar.
v) A Use Case "Error correction" - Improvement was made in a panel editor according to the guides given.
vi) A Use Case "Update File" - To continuously update and store all the acceptable grammatical structure into a database.

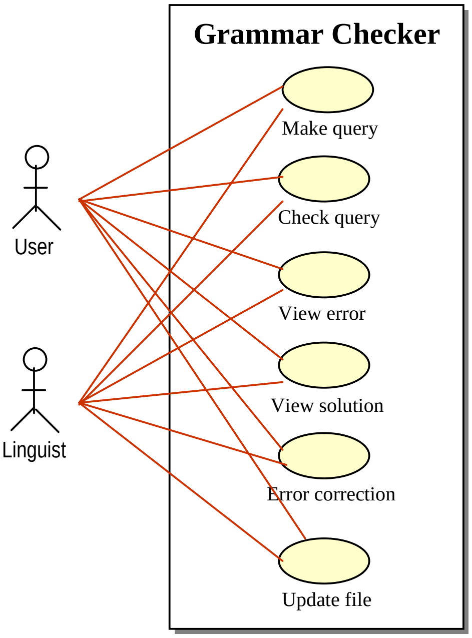

Figure 3 : Use case diagram for grammar checker

4.3 Use Case Diagram: System Administration

This Use Case diagram provides functionality for a system administrator to professionally manage user information, accessibility and viewing report. The use cases and actor involved shown in Figure 4.

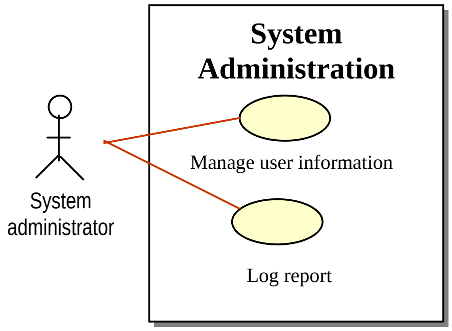

Figure 4: Use case diagram for system administration

i) A Use Case "Manage User Information" - This use case provides substance of operational activity for system maintenance (e.g. maintaining username and password, user profiles, etc.).
ii) A Use Case "Log Report" – This use case routinely keep the matters associating with the regularly tasks from the system.

5. CONCLUSION AND FUTURE WORK

The major contribution of this research is providing an underlying research on a Malay Grammar Checker system. The basic design system flow of the process substantiating with a suitable UML-Use Case diagram is really worthwhile to get further understanding on the requirements in developing the system. This growing research project will offer a lot greater potential research areas such as involved in data analysis, system implementation, system testing and deployment.

Currently, the subsequent work in this research project is forming additional data for the system. Contributing from linguists to ratify the rules is important before a system developer writes a code for certifying the rules constructed.

In developing this system, we are also planning to implement generic algorithms for Malay Grammar checker engine. This technique is prominent in accomplishing the data exchange integration standard and gives an advantage for integration with distinct types of other languages.

As pointed out earlier, we are planning to growth the number of data in our database by tempting substantial contribution linguists to verify rules and data for our system.

REFERENCES

Abdullah Hassan. (2004). *Tatabahasa Bahasa Melayu: morfologi dan sintaksis untuk guru dan pelajar*, PTS Publications & Distributors Sdn.Bhd, Pahang, Malaysia.

Abdullah Hassan, et.al. (2004). *Sintaksis: Siri pengajaran dan pembelajaran Bahasa Melayu*, ,PTS Publications & Distributors Sdn.Bhd, Pahang, Malaysia.

Abdullah Hassan. (2007). *Linguistik am: Siri pengajaran dan pembelajaran Bahasa Melayu*, PTS Publications & Distributors Sdn.Bhd, Pahang, Malaysia.

Asian Regional Exchange for New Alternatives (ARENA) (2007). [Online]. Available:http:// www.arenaonline.org /content/view/ 305 /151/

Craig larman. (1998). *Applying UML and Patterns: An introduction to object-oriented analysis and design*: Prentice Hall PTR, New Jersey.

Jacobson, I., Booch,G., Rumbaugh,J. (1999). *The Unified Software Development Process*: Addison-Wesley, Boston.

Jacobson, I., Griss, M., Jonsson, P. (1997). *Software Reuse: Architecture, Process and Organization for Business Succes*: Addison Wesley, Boston.

John W.Satzinger, Robert B. Jackson, Stephen D.Burd. (2005). *Object oriented analysis and design with the unified process*: Thomson ^{TM} learning, Boston.

Rational Unified Process: Construction phase. (2003). [Online]. Available: http:// rup.hops-fp6.org/process/ itrwkfls /iwf_lic.htm

Rational Unified Process: Elaboration phase. (2003). [Online]. Available: http://www.yoopeedoo.org/ upedu/ process /itrwkfls/ iwf_iie.htm

Rational Unified Process: Inception phase. (2003). [Online]. Available: http :// rup.hops-fp6.org/process/ itrwkfls / iwf_iii.htm

Rational Unified Process: Transition phase. (2003). [Online]. Available:http:// rup.hops-fp6.org /process /itrwkfls/ iwf_lit.htm