DEVELOPMENT OF INTERACTIVE MULTIMEDIA APPLICATION THE TALE TELLER USING SPEECH RECOGNITION METHOD

Aldia Rahman Mahmudi[1], Hanhan Maulana[2]

[1,2]Informatics Engineering – Indonesian Computer University
Jl. Dipatiukur 112-114 Bandung 40132
E-mail : aldiarahman8@gmail.com[1], hanhan@email.unikom.ac.id[2]

ABSTRACT

Storytelling is one of the teaching and learning activities conducted in TK Swadaya. Storytelling activities depend on the ability of the teacher to convey the story, unfortunately that not all teachers have the ability to tell good stories, which results in students not paying more attention. Storytelling activities currently do not have interactive teaching aids and only use a picture book as a media that is shown in front of students. Based on these problems, it is proposed an application that can present a more interactive and interesting story with speech recognition controls. Speech recognition control is expected to attract interest in student interaction to be active in storytelling activities. Speech recognition technology is implemented using Google Speech to Text API and a desktop-based system. The results of the study were in the form of developing interactive multimedia applications for storytelling activities using the speech recognition method. After testing, the results show that the functionality that the application has can run well and the application is assessed to be able to help the teacher in delivering stories and as an interactive visual media in storytelling activities.

Keywords : Storytelling, Interactive Media, Speech Recognition, Desktop Application, Google Speech to Text API.

1. PRELIMINARY

Storytelling is an oldest branch of literature. Humans from birth have the nature to tell something that is experienced to explain something and provide understanding [1]. One common storytelling media is fairy tales. Fairy tales have existed for a long time and are told from generation to generation, irrational but have the benefit of a moral message. Fairy tales are often told to children because they affect various aspects of development, especially cognitive [2]. TK Swadaya is an educational institution that focuses on early childhood education and one of the TK Swadaya learning activities is storytelling. Storytelling activities are intended to stimulate active language children by responding to answers based on knowledge from the story.

Based on observations to 84 students of TK Swadaya and interviews with the Principal of TK Swadaya, it was found that students liked fairy tales but had problems in the way the teacher were delivered. Submission of stories depends on the ability of the teacher, unfortunately that not all teachers have the ability to tell good stories so that the response and interest of students decreases. The next obstacle in storytelling activities is the lack of media or teaching aids. For early childhood learning is playing and playing is learning, when there are media or props that can be played, children can listen more and actively interact because they are given the opportunity to participate in the story.

There are several studies that have discussed the influence of storytelling on improving language skills and interactive multimedia applications developed for learning media. Storytelling is considered to have a significant influence on early childhood development, especially in language and cognitive abilities [1], storytelling in the form of digital media helps convey stories more fun and interactive [3] and the addition of application control types with the use of speech recognition technology can attract children to interact [4].

Based on the description above, we need a media that can present storytelling activities in a more interactive and interesting way. The solution to help storytelling activities is by applying it to the form of a game with speech recognition control to interact and control objects in the story so that children want to actively interact and participate in the story. Speech recognition is implemented using Google API, because it has good accuracy with the lowest error rate of 9 percent [5]. The topic will be discussed in this study and entitled "Development of Interactive Multimedia Application The Tale Teller Using Speech Recognition Method" which aims to assist teachers in delivering material and provide interactive media or teaching aids that can attract students to interact in storytelling activities.

2. THEORITICAL BASIS

The theoretical basis contains a basic explanation of the theory associated with the research conducted. The theoretical basis to be discussed is about multimedia, storytelling, speech recognition, Google

Speech to Text API and system testing in the form of accuracy testing and black box.

2.1. Multimedia

Multimedia is the main concept that includes this research and applications that will be built, or rather multimedia in the form of a computer application. Multimedia is the use of applications from various elements such as text, audio, images, animations, videos and all forms of interactivity [6].

2.2. Storytelling

Storytelling is a skill of narrating stories in the form of poetry or prose, led by one person in front of the listener directly where the story is narrated by way of being told or sung, using musical accompaniment, pictures, or other accompaniment that can be learned orally, both through the source of writing or through recording sources [1].

2.3. Speech Recognition

Speech recognition is a process for converting a form of conversation into appropriate words or entities through certain algorithms that are implemented in a computer device. Speech recognition has four main components, namely signal processing and feature extraction, acoustic models of input sound, searching hypotheses from keyword correlations and language models of recognizable words [7].

2.4. Google Speech to Text API

Google Speech to Text can translate audio into text in an easy-to-use API. Google Speech to Text can recognize up to 120 types of languages and can process streaming audio in real-time [8]. Google Speech to Text API menggunakan algoritma Deep Learning Neural Network yaitu Recurrent Neural Network (RNN). The concept of RNN is processing information in stages, by doing the same tasks on each element in a sequence, then processing the output that refers to previous computing [9].

2.5. Accuracy Testing

Accuracy testing is one way to measure the performance of speech recognition. In this accuracy, there are several respondents. Everyone gets the opportunity to test all keywords in the application. Calculation of accuracy, is used to determine how much success of the application is built, where the benchmark of success is the suitability of the command with the output [10]. The following is the calculation formula for accuracy used can be seen in Figure 1.

$$Akurasi = \frac{Jumlah\ kata\ benar}{Jumlah\ kata\ yang\ diuji}\ x\ 100\%$$

Figure 1

2.6. Black Box Testing

Black Box Testing focuses on the functionality of the software. The tester defines a set of input conditions and tests the functional program. Testing is done to find things like incorrect or non-existent functions, interface errors, data structure errors and database access, performance and initialization errors or terminations [11].

3. RESEARCH METHODS

The research method used is descriptive research method, because this study aims to explain in detail the stages of the research carried out based on facts and phenomena obtained from the data that were extracted and the research that has been done before [12].

The flow of research conducted in this study includes data collection, problem identification. analysis and design, software development, research testing and drawing conclusions. The following is a description of the research flow that can be seen in Figure 2.

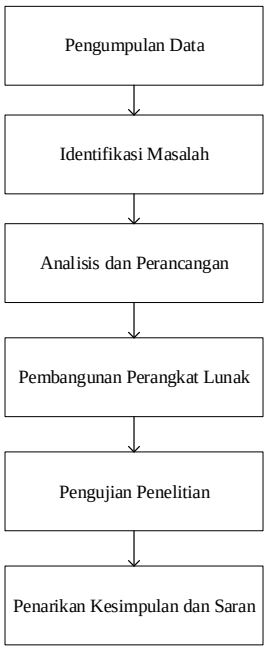


Figure 2

The software development method used in this study is the Multimedia Development Life Cycle (MDLC) method which consists of 6 stages namely concept, design, material collection, manufacture, testing and distribution [13]. The description of the MDLC flow can be seen in Figure 3.

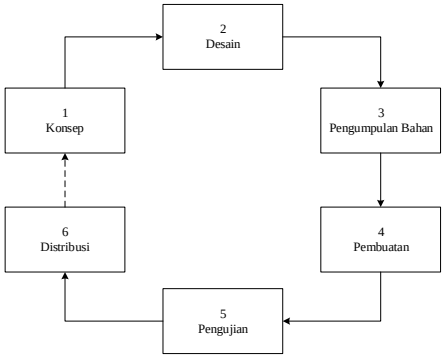


Figure 3

a. Concept
   The concept is based on the purpose and who is the user of the software that is built, where in this study the purpose of the user is an early childhood guided by teaching staff.
b. Design
   Designing software structure specifications, appearance and material requirements in the form of materials and information for software built.
c. Material Collecting
   Collection of material that is in accordance with the needs of the software being built.
d. Assembly
   Combining all multimedia objects and materials collected based on the concepts and designs that have been made.
e. Testing
   Testing of system functionality built.
f. Distribution
   The software will be stored in storage media and distributed to end users, namely teachers at TK Swadaya and evaluation.

4. RESULTS AND DISCUSSION

The results and discussion describe the analysis, design, implementation and testing of the system being built.

4.1. Problem Analysis

Submission of story material depends on the individual abilities of each teacher, unfortunately not all lecturers have the ability to tell stories well and result in the majority of children not even paying attention, playing alone or chatting. This can be seen from the majority of children who do not respond to the interaction conducted by the teacher, students also cannot conclude or retell the stories that have been told.

Storytelling activities were conveyed so far only using auxiliary media in the form of one picture story book that was shown to students. The use of media or teaching aids in early childhood teaching and learning activities can help attract children's interest in listening to the material, this is because early childhood learning is playing and playing is learning. Media or props that can be played can help children listen more and actively interact because they are given the opportunity to participate or be involved in the activities carried out.

4.2. System Overview

The system to be built is an interactive multimedia application for storytelling activities in digital media, in this application there are 8 story titles and various forms of interaction between the user and the system are added in an effort to add interesting interactivity. Speech recognition is applied in most uses of this application to control and interact with objects in the application, where the application output is in the form of animated visualization of existing objects.

4.2.1. System Architecture

System architecture describes how the components in the system interact. In the application to be built, the instructor becomes a guide for students in storytelling activities. The teacher reads the story narrative and the system will receive voice input through the microphone, voice input in the form of story narratives will be processed by the system using Google Speech to Text API and get the written text of the story that was read. From the text, identification of keywords will be used to control the movement and calling of objects or characters in the story. The system architecture built can be seen in Figure 4.

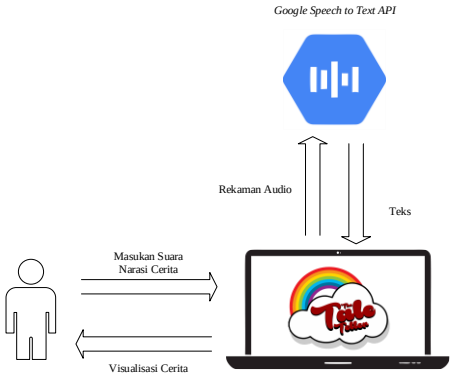


Figure 4

4.2.2. Speech Recognition Analysis

In the speech recognition process there are several main processes, starting from the input received, namely the story narration read out by the sound recording process, the recording results will be processed by the Google Speech to Text API to produce identification text and then identify keywords and control objects so that they are obtained story visualization output. Speech recognition stages can be seen in Figure 5.

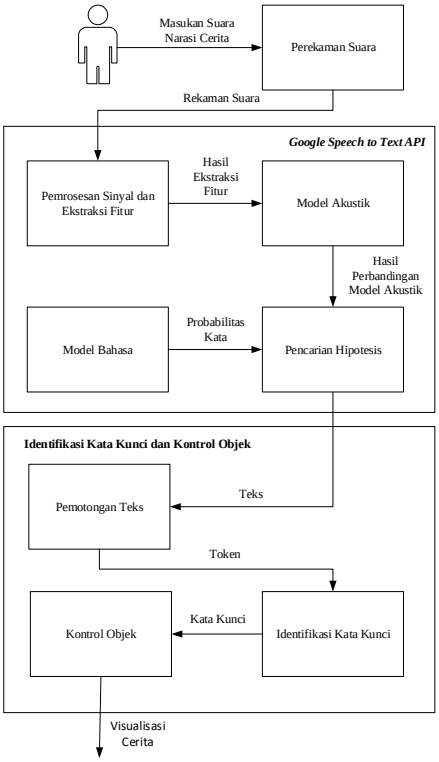


Figure 5

4.3. System Modeling

System modeling aims to make specifications, visualize and build the framework of the system.

4.3.1. Use Case Diagram

Use case diagrams are used to describe interactions between users and systems. Use Case Diagram of the application built can be seen in Figure 6.

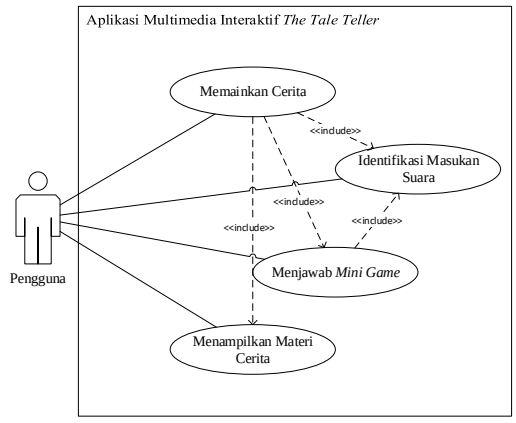


Figure 6

The use case diagram above shows there is an actor that is a user, who uses an application and there are several use cases which are the functionality that the system has. The use case definition can be seen in Table 1.

Table 1

No.	Use Case	Explanation
UC-01	Memainkan Cerita	The process for choosing stories and starting to play stories.
UC-02	Identifikasi Masukan Suara	The process of identifying object control texts and keywords based on the narrative voice input that the user reads.
UC-03	Menjawab Mini Game	The process for answering mini games in the form of quizzes is found in and at the end of the story.
UC-04	Menampilkan Materi Cerita	The process for displaying a summary of the material contained in the story.

4.3.2. Class Diagram

Class diagrams are the main core of a system that describes the different types of objects and classes that the system has. Class diagrams also show relations between classes in the system. Class diagrams of applications that are built can be seen in Figure 7.

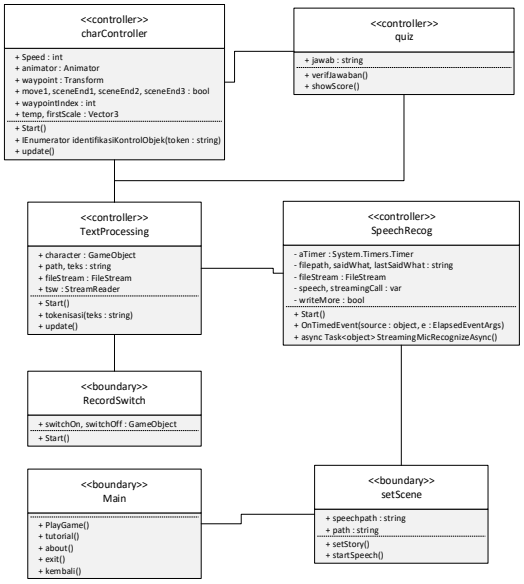


Figure 7

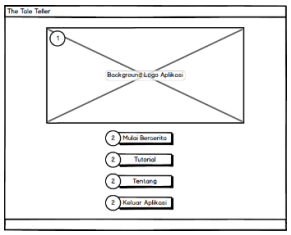
4.4. Design and Implementation

Design and implementation describes the main design of the system and the results of implementation of the system.

4.4.1. Interface Design

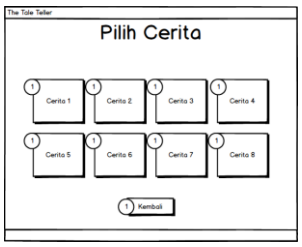
The interface design is a rough description of the application display design that will be built. The following is the design of the menu utama page which can be seen in Table 2.

Table 2

T01	
	**Explanation :** 1. Background 2. Button **Navigation :** a. Press the Mulai Bercerita button to go to T02. b. Press the Tutorial button to go to T06. c. Press the Tentang button to go to T07.
Form name : Menu Utama Size : 1280 x 800 pixels Orientation : Landscape Font type : BadaBoom BB	

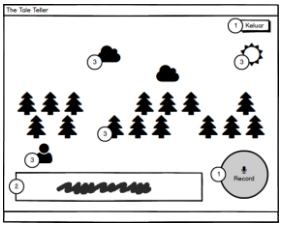
The following is the design of the menu pilih cerita page which can be seen in Table 3.

Table 3

T02	
	**Explanation :** 1. Button **Navigation :** a. Press the Cerita 1 button to go to T03. b. Press the Cerita 2 button to go to T03. c. Press the Cerita 3 button to go to T03. d. Press the Cerita 4 button to go to T03. e. Press the Cerita 5 button to go to T03.
	f. Press the Cerita 6 button to go to T03. g. Press the Cerita 7 button to go to T03. h. Press the Cerita 8 button to go to T03. i. Press the Kembali button to go to T01.
Form Name: Menu Pilih Cerita Size : 1280 x 800 pixels Orientation : Landscape Font type : BadaBoom BB	

The following is the design of the halaman main which can be seen in Table 4.

Table 4

T03	
	**Explanation :** 1. Button 2. Subtitle 3. Animated Object **Navigation :** a. Press the Keluar button to go to T02. b. If the user has completed the Halaman Main it will go to T04.
Form Name : Halaman Main Size : 1280 x 800 pixels Orientation : Landscape Font type : BadaBoom BB	

4.4.2. Semantic Network

Semantic network describes the flow and relationship of the interface design that has been created. The semantic network of applications built can be seen in Figure 8.

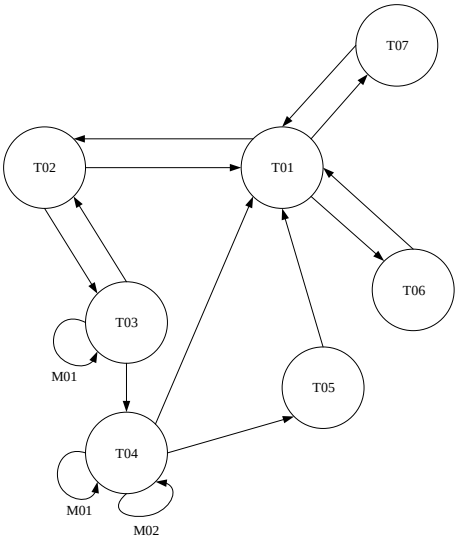


Figure 8

4.4.3. Interface Implementation

Interface implementation is the result of the implementation of the design that has been made. The following are the results of the implementation of the menu utama interface where there are 4 choices of menus, starting from mulai bercerita, tutorial, tentang aplikasi and keluar aplikasi, can be seen in Figure 9.

Figure 9

Following are the results of the implementation of the interface from the pilih cerita menu, there are 8 choices of stories that can be seen in Figure 10.

Figure 10

The following is the result of the implementation of the interface display from main where users can play selected stories which can be seen in Figure 11.

Figure 11

4.5. Testing and Results of Research Trials

Testing is the stage to find errors or shortcomings of the system built that was obtained from a series of stages of testing carried out.

4.5.1. Speech Recognition Testing

Speech recognition testing will be carried out by 8 respondents using a list of 27 keywords that come from 1 story that is read by the way each user reads the story provided using a laptop microphone. Tests are carried out in 2 environmental conditions, quiet environments in empty classrooms and crowded environments in classrooms containing students. The following are the results of the noise level measurements of the two environments which can be seen in Table 5.

Table 5

Environment	Noise Level (Average)
Quiet Class Room	26 dB
Crowded Class Room	66 dB

The following are the results of testing carried out by each user in a quiet classroom environment with an average noise level of 26 dB and a crowded environment with an average noise level of 66 dB, can be seen in Table 6.

Table 6

Number of Words Tested	Number of Tests for Each Word	Noise Level	Average Accuracy (%)
27	8	26 dB	92,3
		66 dB	49,54

Based on the results of speech recognition testing, it is known that in low noise environments, the level of accuracy of word recognition is very good with an average of 92.3%, but in tests in high noise environments, there is a high decrease in accuracy where word recognition accuracy on average only 49.54%.

4.5.2. Keyword Testing

Keyword processing testing is a test performed on keyword sequence validation. Testing is done by reading out the sequence of keywords to test the logic of the word sequence processing validation process

that is in the application, where the keyword functions as a control of the animation movement and scene transfer. The keyword testing scenario can be seen in Table 7.

Table 7

Scene	Keyword Sequence	Testing Plan
1	singa, murung, makan, cari	Displays the animation of the story when the word sequence matches and moves to scene 2 when all words have been read.
2	tikus, main	Displays the animation of the story when the word sequence matches and moves to scene 3 when all words have been read.
3	taring, melompat, menggeng gam	Displays the animation of the story when the word sequence matches and moves to scene 4 when all words have been read.
4	takut, menangis, masuk, mulut	Displays the animation of the story when the word sequence matches and moves to scene 5 when all words have been read.
5	berani, melihat, daging, kenyang	Displays the animation of the story when the word sequence matches and moves to scene 6 when all words have been read.

Based on the results of keyword testing, it can be concluded that the keyword sequence processing validation process contained in the system built has been running well and functioning in accordance with the expected output.

4.5.3. Black Box Testing

Black Box testing aims to test the functionality of a newly built system. Following are the black box testing scenarios of the applications built, can be seen in Table 8.

Table 8

No.	Functional Tested	Testing Plan
1.	Memainkan Cerita	Objects and animated stories that appear in accordance with the selected story.
2.	Identifikasi Masukan Suara	Translating the narrative voice input into text, displaying objects and animations in the story based on the text input sound recognition results.
3.	Memainkan Mini Game	Showing mini games according to the story being played and checking answers.
4.	Menampilkan Materi Cerita	Display story material according to the selected story.

Based on the results of black box testing it can be concluded that all the functionality contained in the system built can run well according to the expected functional requirements.

4.5.4. Interview Testing

Testing interviews aims to determine the user's response to the system that has been built by conducting interviews so that conclusions can be made, whether the system that has been built is in accordance with the objectives of the study. The following is a list of questions submitted to the principal and teaching staff of TK Swadaya based on the initial objectives of the study, can be seen in Table 9.

Table 9

Research Purposes	Questions
1. Helping teachers in delivering story material for storytelling activities.	1. Is the delivery of story content in this application easy to understand? 2. Is the appearance of this application interesting? 3. Is this application easy to use? 4. Can this application be an alternative in delivering story material?
2. Providing media or interactive teaching aids that can attract students to actively interact.	5. Can this application present the visualization of the story correctly? 6. Can the application of speech recognition in this application increase interest in student interaction? 7. Can this storytelling activity be more interesting with this application?

Based on the results of testing interviews with user responses, it can be concluded that the system that has been built has fulfilled the initial objectives of the study.

5. CONCLUSION

Based on the final results of the analysis, design, implementation and testing of the research conducted, it can be concluded that the interactive multimedia applications of the tale teller using the speech recognition method that is built can fulfill the objectives of the research, which can help teachers in delivering story material and can be tools interactive display to attract interest in student interaction in storytelling activities.

The following are suggestions for further research and development of applications that have been built.

a. Add more story content and animation in the application
b. Addition of noise suppressor to reduce noise and improve recognition accuracy in crowded environments.
c. Design improvements so that they can be made better and more attractive.

BIBLIOGRAPHY

[1] Y. I. Firyati, E. Y. Haenillah, and Sasmiati, "Story Telling Meningkatkan Perkembangan Bahasa Anak Usia Dini," *J. Pendidik. Anak*, vol. 2, no. 2, 2016.

[2] P. P. Ardini, "Pengaruh Dongeng dan Komunikasi Terhadap Perkembangan Moral Anak Usia 7-8 Tahun," *J. Pendidik. Anak*, vol. 1, no. 2, pp. 44–58, 2012.

[3] D. M.P., I. R. Mutiaz, and A. B. Sriwarno, "Transformasi Cara Penuturan Cerita dalam Bentuk Aplikasi Digital Storytelling sebagai Media Pengenalan Kembali Legenda 'Calon Arang,'" *Nirmana*, vol. 15, no. 1, pp. 1–12, 2013.

[4] H. A. Ericksoon, I. Kuswardayan, and N. Suciati, "Rancang Bangun Game Berhitung Spaceship dengan Pengendali Suara Menggunakan Speech Recognition Plugin pada Unity," *J. Tek. ITS*, vol. 5, no. 2, pp. A620–A624, 2016.

[5] V. Këpuska, "Comparing Speech Recognition Systems (Microsoft API, Google API And CMU Sphinx)," *Int. J. Eng. Res. Appl.*, vol. 07, no. 03, pp. 20–24, 2017.

[6] Z. N. Li and M. S. Drew, *Fundamentals of Multimedia*. Pearson Education Inc., 2004.

[7] J. Li, L. Deng, R. Haeb-Umbach, and Y. Gong, *Robust automatic speech recognition: A bridge to practical applications*. Academic Press, 2015.

[8] Admin, "Google Cloud Speech-to-Text." [Online]. Available: https://cloud.google.com/speech-to-text/. [Accessed: 28-Feb-2019].

[9] H. Sak, A. Senior, and F. Beaufays, "Long short-term memory recurrent neural network architectures for large scale acoustic modeling," in *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2014.

[10] A. D. Andriana, "Perangkat Lunak Untuk Membuka Aplikasi Pada Komputer Dengan Perintah Suara Menggunakan Metode Mel Frequency Cepstrum Coeficients," *Ilm. Komput. dan Inform.*, vol. 2, no. 1, pp. 21–26, 2013.

[11] M. E. Khan, "Different Approach to Blackbox Testing Technique for Finding Error," *Int. J. Softw. Eng. Appl.*, vol. 2, no. 4, pp. 31–40, 2011.

[12] M. Nazir, *Metode penelitian*, 10th ed. Bogor: Ghalia Indonesia, 2014.

[13] I. Binanto, *Multimedia Digital Dasar Teori dan Pengembangannya*. Yogyakarta: ANDI OFFSET, 2010.