# CHAPTER 3 Research Methodology



## SOURCES OF DATA

Physics is a language, in particular the language of a certain kind of worldview (Brown, 2013). Physics word problems can be acquired easily with the use internet sources with accessible websites and forums. However, the researchers will use published textbooks to obtain English physics word problems about electricity, forces, kinematics, work and energy, etc. *See figure 1.4.* These published textbooks are either from elementary physics or high school level physics.

Because of the reason that the study aims to comprehend and solve physics problems with no boundaries in the word problem structure, the physics word problem can be in any structure as long as the problem is within the given domain of physics and the problem is textually represented.

Building a dynamic system that interprets any given word problem in a certain domain will require the researchers to use quota sampling and purposive sampling techniques. Quota sampling refers to selection with controls, ensuring that specified numbers (quotas) are obtained from each specified population subgroup (Elder, 2009). The researchers will collect domain-specific word problems from various published physics books in 100 questions possible. Every Physics domain selected in this study needs a large amount of training and test data so the researchers will use purposive sampling technique to ensure the quality of the data being selected due to the visible impact of this texts when the training started.

## 3.3.1.2 DEVELOPMENT DETAILS

The researchers used Java programming language to develop the system. The database will be built in MySQL for storing dictionary, tokens and their corresponding POS tags and Named Entity it belongs, and physics information. The proponents will use Stanford NLP Group’s tools such as POS Tagger and Named Entity Recognizer to acquire the necessary information of the tokens to make it readable for the RNN to train the word problem. The researchers will use Eclipse as a programming and testing platform for development process. The development of the tool is best partnered with Agile development model, where every module starts with planning, analysis, designing, developing then testing, after this, if there is necessary changes seen in the testing, it will start in planning, this will continue until the module is ready for deployment and the tool is completely done. Also, using this development model will assure the quality of the tool and provide great communication among the proponents. The use of Recurrent Neural Network will require training stage for weights and biases of the network, we will use TensorFlowTM to train the said neural network.

**3.3.2 RESEARCH INSTRUMENT**

The researchers used this experiment paper to show the statistical treatment of data result in every test. This paper will be used for measuring the accuracy in recognizing the given figures and their entities in the word problem, required variable. This paper also will be used to measure the performance of formulating the necessary formula for the problem and how the system will solve the problem systematically, and showing the final answer with respect to the unit being asked.