REFERENCES

[1] Bates, E. T., & Wiest, L. R. (2004). Impact of Personalization of Mathematical Word Problems on Student Performance. Mathematics Educator, 14(2), 17-26.

[2] Cummins, D. D., Kintsch, W., Reusser, K., & Weimer, R. (1988). The role of understanding in solving word problems. Cognitive psychology, 20(4), 405-438.

[3] (2013, May 20). Retrieved February 24, 2017, from http://www.moe.gov.lk/english/images/publications/Annual_performance_Report2013/Annual_performance_report_e.pdf, Annual Performance Report.

[4] Bennett, R. E. (2011). Automated scoring of constructed-response literacy and mathematics items. White paper, Educational Testing Service.

[5] Lan, A. S., Vats, D., Waters, A. E., & Baraniuk, R. G. (2015, March). Mathematical language processing: Automatic grading and feedback for open response mathematical questions. In Proceedings of the Second (2015) ACM Conference on Learning@ Scale (pp. 167-176). ACM.

[6] Badger, M. (2013). Problem-solving in undergraduate mathematics and computer aided assessment (Doctoral dissertation, University of Birmingham).

[7] (2010, April 04). Retrieved February 24, 2017, from http://www.open.ac.uk/openmarkexamples/numeric-response, Numerical Response.

[8] Fife, J. H. (2011). Automated scoring of CBAL mathematics tasks with m-rater. Research Memorandum, ETS RM-11-12.

[9] Fife, j. H. (2013). Automated scoring of Mathematics tasks in the common core era: enhancements to m-rater in support of CBAL Mathematics and the common core assessments. ETS research report series, 2013(2).

[10] Hilton III, J. L., Gaudet, D., Clark, P., Robinson, J., & Wiley, D. (2013). The adoption of open educational resources by one community college math department. The International Review of Research in Open and Distributed Learning, 14(4).

[11] Campbell, J. I. (1994). Architectures for numerical cognition. Cognition, 53(1), 1-44.

[12] Luneta, K., & Makonye, P. J. (2010). Learner Errors and Misconceptions in Elementary Analysis: A Case Study of a Grade 12 Class in South Africa. Acta Didactica Napocensia, 3(3), 35-46.

[13] Veloo, A., Krishnasamy, H. N., & Abdullah, W. S. W. (2015). Types of Student Errors in Mathematical Symbols, Graphs and Problem-Solving. Asian Social Science, 11(15), 324.

[14] Erabadda B., "Automatic identification of errors in multi-step answers to algebra questions" in Proceedings of the 17th IEEE International Conference on Advanced Learning Technologies (ICALT), 2017.(Accepted)

[15] Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. Discourse processes, 25(2-3), 259-284.

[16] Li, Y., McLean, D., Bandar, Z. A., O'shea, J. D., & Crockett, K. (2006). Sentence similarity based on semantic nets and corpus statistics. IEEE transactions on knowledge and data engineering, 18(8), 1138-1150.

[17] Zhao, J., Zhu, T. T., & Lan, M. (2014). Ecnu: One stone two birds: Ensemble of heterogenous measures for semantic relatedness and textual entailment. Proceedings of the SemEval, 271-277.

[18] Klai, S., Kolokolnikov, T., & Van den Bergh, N. (2000). Using Maple and the web to grade mathematics tests. In Advanced Learning Technologies, 2000. IWALT 2000. Proceedings. International Workshop on (pp. 89-92). IEEE.

[19] Ritter, S., Anderson, J. R., Koedinger, K. R., & Corbett, A. (2007). Cognitive Tutor: Applied research in mathematics education. Psychonomic bulletin & review, 14(2), 249-255.

[20] Melis, E., & Siekmann, J. (2004, June). Activemath: An intelligent tutoring system for mathematics. In International Conference on Artificial Intelligence and Soft Computing (pp. 91-101). Springer Berlin Heidelberg.

[21] Funk, N. (2000). JEP-Java Math Expression Parser. Online document.

[22] Joyner, D., Čertík, O., Meurer, A., & Granger, B. E. (2012). Open source computer algebra systems: SymPy. ACM Communications in Computer Algebra, 45(3/4), 225-234.

[23] Gomaa, W. H., & Fahmy, A. A. (2012). Short answer grading using string similarity and corpus-based similarity. International Journal of Advanced Computer Science and Applications (IJACSA), 3(11).

[24] Hale, M. M. (1998). A comparison of WordNet and Roget's taxonomy for measuring semantic similarity. arXiv preprint cmp-lg/9809003.

[25] Resnik, P. (1995). Using information content to evaluate semantic similarity in a taxonomy. arXiv preprint cmp-lg/9511007.

[26] Wu, Z., & Palmer, M. (1994, June). Verbs semantics and lexical selection. In Proceedings of the 32nd annual meeting on Association for Computational Linguistics (pp. 133-138). Association for Computational Linguistics.

[27] Tayal, M. A., Raghuwanshi, M. M., & Malik, L. (2014, December). Word net based Method for Determining Semantic Sentence Similarity through various Word Senses. In Proceedings of the First Joint Conference on Lexical and Computational Semantics.

[28] Mohler, M., & Mihalcea, R. (2009, March). Text-to-text semantic similarity for automatic short answer grading. In Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (pp. 567-575). Association for Computational Linguistics.

[29] Hirst, G., & St-Onge, D. (1998). Lexical chains as representations of context for the detection and correction of malapropisms. WordNet: An electronic lexical database, 305, 305-332.

[30] Jiang, J. J., & Conrath, D. W. (1997). Semantic similarity based on corpus statistics and lexical taxonomy. arXiv preprint cmp-lg/9709008.

[31] Leacock, C., & Chodorow, M. (1998). Combining local context and WordNet sense similarity for word sense identification. WordNet, an Electronic Lexical Database.

[32] Lesk, M. (1986, June). Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone. In Proceedings of the 5th annual international conference on Systems documentation (pp. 24-26). ACM.

[33] Lin, D. (1998, July). An information-theoretic definition of similarity. In ICML (Vol. 98, No. 1998, pp. 296-304).

[34] Mohler, M., Bunescu, R., & Mihalcea, R. (2011, June). Learning to grade short answer questions using semantic similarity measures and dependency graph alignments. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1 (pp. 752-762). Association for Computational Linguistics.

[35] Alves, A. O., Ferrugento, A., Lourenço, M., & Rodrigues, F. (2014). Asap: Automatic semantic alignment for phrases. SemEval 2014, 104.

[36] Bestgen, Y. (2014). CECL: a new baseline and a non-compositional approach for the Sick benchmark. SemEval 2014, 160.

[37] Agirre, E., Banea, C., Cardie, C., Cer, D., Diab, M., Gonzalez-Agirre, A., ... & Wiebe, J. (2014, August). Semeval-2014 task 10: Multilingual semantic textual similarity. In Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014) (pp. 81-91). Association for Computational Linguistics Dublin, Ireland.

[38] Marelli, M., Bentivogli, L., Baroni, M., Bernardi, R., Menini, S., & Zamparelli, R. (2014). Semeval-2014 task 1: Evaluation of compositional distributional semantic models on full sentences through semantic relatedness and textual entailment. SemEval-2014.

[39] Boehme, R. F., Fairweather, P. G., Farooq, U., Lam, D., & Singley, K. (2008). U.S. Patent Application No. 12/051,347.

[40] Li, G., Li, L., Su, W., & Zhao, Y. (2006). Design and implementation of MAML. In Proceedings of ATCM.

[41] Ambekar, D. (2015). Evaluation of essays using incremental training for Maximizing Human-Machine agreement (Doctoral dissertation, Indian Institute of Technology, Bombay).

[42] Livne, N. L., Livne, O. E., & Wight, C. A. (2007). Can automated scoring surpass hand grading of students' constructed responses and error patterns in mathematics. MERLOT Journal of Online Learning and Teaching, 3(3), 295-306.

[43] Pellegrino, J. W., & Goldman, S. R. (1987). Information processing and elementary mathematics. Journal of Learning Disabilities, 20(1), 23-32.

[44] Norman, D. A. (1981). Categorization of action slips. Psychological review, 88(1), 1.

[45] Feng, M., Heffernan, N., & Koedinger, K. (2010). Student modeling in an intelligent tutoring system. Intelligent tutoring systems in e-learning environments: Design, implementation and evaluation, 208-236.

[46] Arroyo, I., Beal, C., Murray, T., Walles, R., & Woolf, B. P. (2004, August). Web-based intelligent multimedia tutoring for high stakes achievement tests. In

International Conference on Intelligent Tutoring Systems (pp. 468-477). Springer Berlin Heidelberg.

[47] (2014). Retrieved February 24, 2017, from http://mathspring.org/, University of Massachusetts

[49] Welgama, V., Herath, D. L., Liyanage, C., Udalamatta, N., Weerasinghe, R., & Jayawardana, T. (2011). Towards a sinhala wordnet. In Proceedings of the Conference on Human Language Technology for Development.

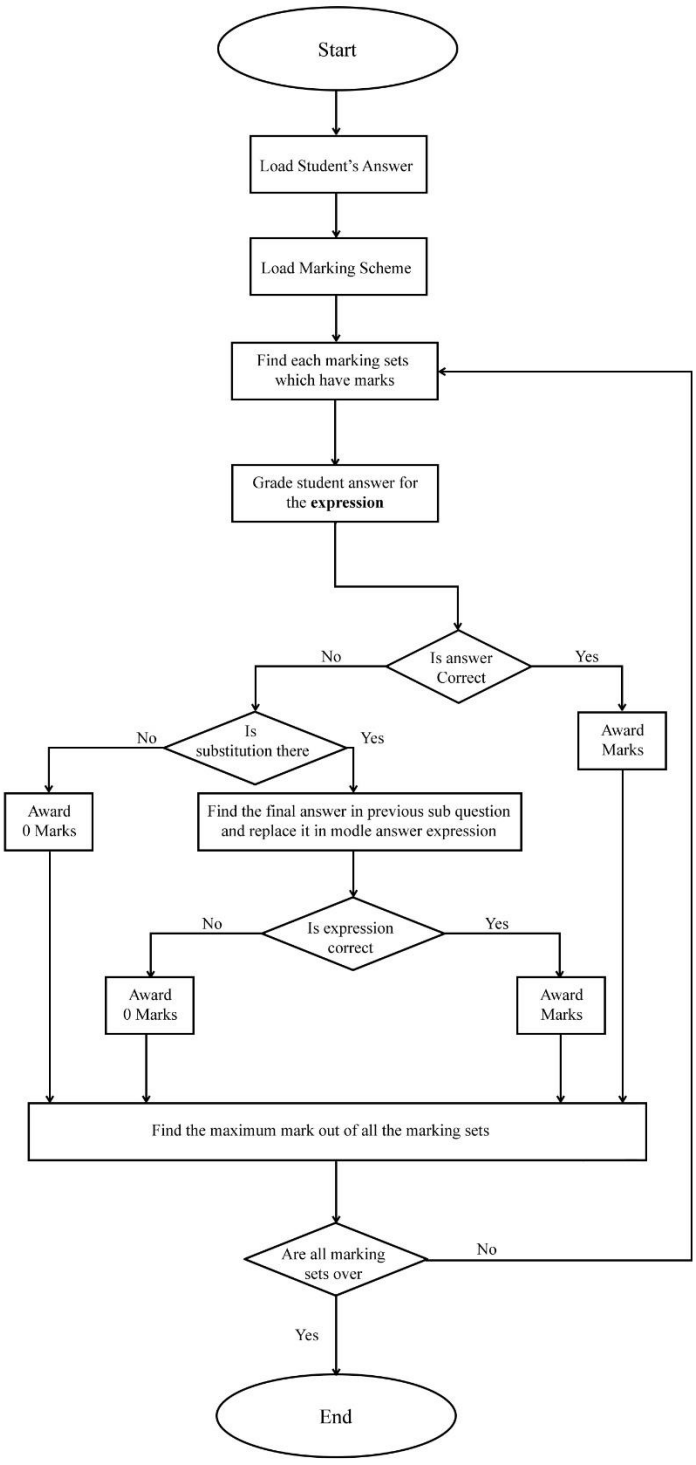
[50] Wijesiri, I., Gallage, M., Gunathilaka, B., Lakjeewa, M., Wimalasuriya, D. C., Dias, G., ... & De Silva, N. (2014). Building a WordNet for Sinhala. Volume editors, 100.

[51] Corley, C., & Mihalcea, R. (2005, June). Measuring the semantic similarity of texts. In Proceedings of the ACL workshop on empirical modeling of semantic equivalence and entailment (pp. 13-18). Association for Computational Linguistics.

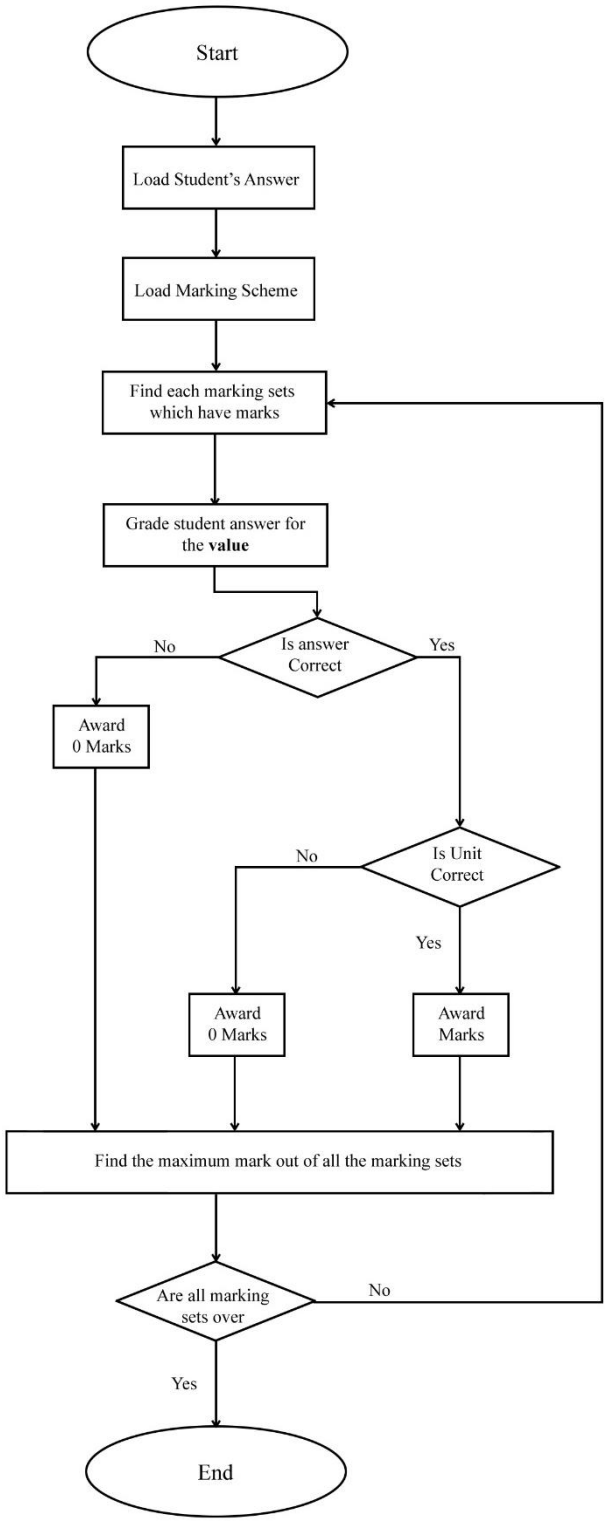
[52] Budanitsky, A., & Hirst, G. (2001, June). Semantic distance in WordNet: An experimental, application-oriented evaluation of five measures. In Workshop on WordNet and other lexical resources (Vol. 2, pp. 2-2).

[53] Gupta, R., Béchara, H., El Maarouf, I., & Orasan, C. (2014, August). UoW: NLP techniques developed at the University of Wolverhampton for Semantic Similarity and Textual Entailment. In 8th Int. Workshop on Semantic Evaluation (SemEval'14) (pp. 785-789).

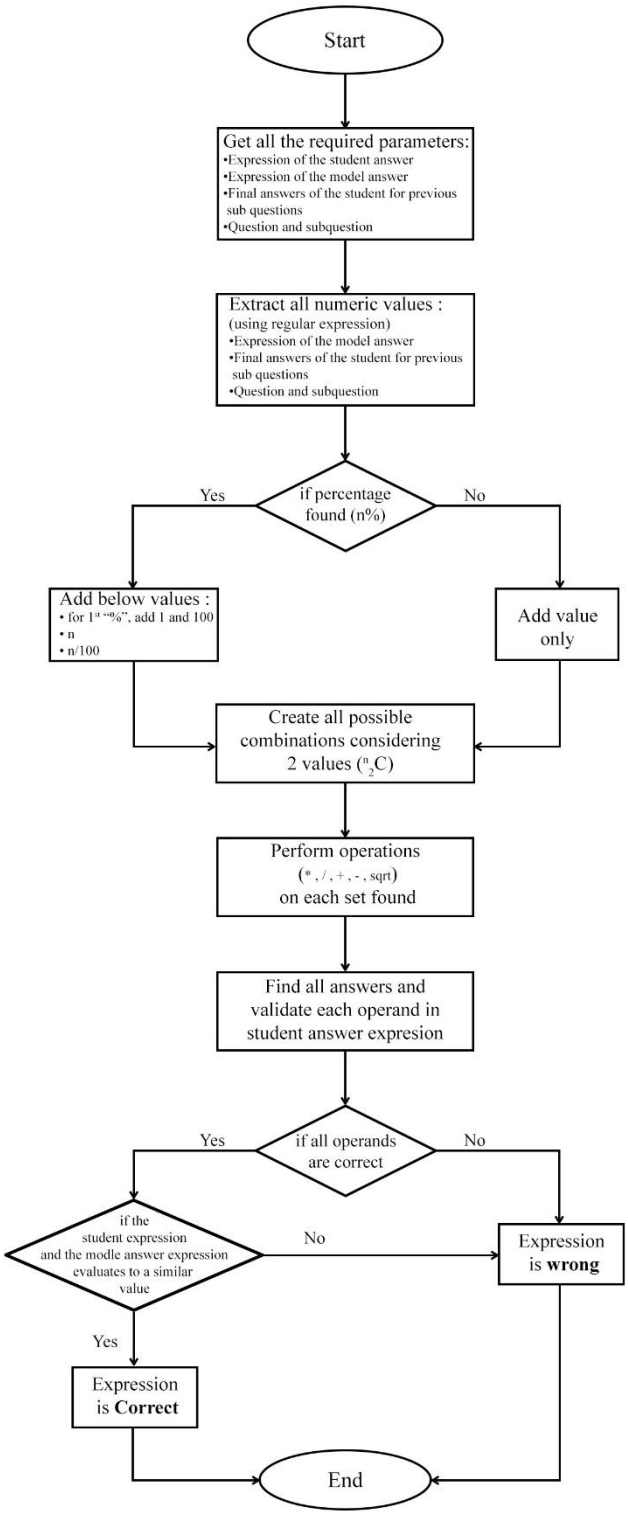
Appendix A: Flow chart for answer expression evaluation



Appendix B: Flow chart for final answer evaluation



Appendix C: Flow chart for Expression validation



Appendix D: Computer Representation of the Marking Rubric, Question, Unit Categorization and the Answer (XML Documents)

```xml
<?xml version="1.0" encoding="UTF-8"?>
<rubric>
    <type>InterestCalculation</type>
    <sectionid>1</sectionid>
    <question id="q01" totalMarks="10">
        <sub_question id="q01-01" totalMarks="1" isProof="true">
            <mark_set id="1">
                <data required="true" isFinal="false" marks="1">
                    <phrase>Total amount spent</phrase>
                    <math_expression unitCompulsory="false" >
                        <mn>600000</mn>
                        <mo>*</mo>
                        <mn>30</mn>
                        <mo>/</mo>
                        <mn>100</mn>
                    </math_expression>
                </data>
                <data required="true" isFinal="true" penalty="1">
                    <value unitCompulsory="true">
                        <preunit>s01-01</preunit>
                        <val>180000</val>
                    </value>
                </data>
            </mark_set>
        </sub_question>
        <sub_question id="q01-02" totalMarks="2" isProof="true">
        <sub_question id="q01-03" totalMarks="2" isProof="false">
        <sub_question id="q01-04" totalMarks="2" isProof="false">
        <sub_question id="q01-05" totalMarks="3" isProof="false">
    </question>
</rubric>
```

Figure D.1 XML representations of rubric document

```xml
<?xml version="1.0" encoding="UTF-8"?>
<units>
    <type>Units</type>
    <local_units id="1">
        <type>Mathematics</type>
        <section id="s01" >
            <name>Currency</name>
            <unit_set id="s01-01">
                <data>රුපියල්</data>
                <data>රු</data>
                <data>Rupee</data>
                <data>Rs</data>
                <data>LKR</data>
                <data>Sri Lankan Rupee</data>
            </unit_set>
            <unit_set id="s01-02">
        </section>
        <section id="s02" >
    </local_units>
</units>
```

Figure D.2 XML representations of unit categorization

```
<?xml version="1.0" encoding="UTF-8"?>
<answer_script>
    <type>AnswerScript</type>
    <student id="RU/E/2010/070">
    <question id="q01" >
        <manualGradingMarks>10</manualGradingMarks>
        <sub_question id="q01-01" >
        <sub_question id="q01-02" >
        <sub_question id="q01-03" >
        <sub_question id="q01-04" >
        <sub_question id="q01-05" >
            <manualGradingMarks_sub>3</manualGradingMarks_sub>
            <answer>Total amount at the end of the second year=420000*1.08^2</answer>
            <answer>=৳489888</answer>
        </sub_question>
    </question>
</answer_script>
```

Figure D.3 XML representations of student answer

Appendix E: Example word problems in Sinhala language

පුද්ගලයෙක් තමා සතු රු .420000 ස්ථාවර ගිණුමක තැන්පත් කරයි. වාර්ෂික වැල් පොලී අනුපාතිකය 8% ක් නම් මුල් වසර දෙක අවසානයේ දී ගිණුමෙහි ඇති මුළු මුදල ගණනය කරන්න.

පලමු වසර අවසානයේ පොලිය	= 420000*8/100
	= Rs.33600
දෙවන වසර අවසානයේ පොලිය	= (420000 + 33600)*8/100
	= Rs.36288
දෙවන වසර අවසානයේ මුළු මුදල	= 420000 + 33600 + 36288
	= Rs.489888

Figure 1 Numerical word problem and its sample answer

පතුලේ අරය r cm ද උස 12cm ද වන තුනී කුහර කේතුවක් සම්පූර්ණයෙන්ම ජලයෙන් පුරවා පැත්තක දිග 8cm වන සමචතුරස්‍රාකාර පතුලක් සහිත භාජනයකට වත් කරනු ලැබේ. කේතුවේ ජල පරිමාව r ඇසුරින් සොයන්න.

කේතුවේ ජල පරිමාව	= 1/3*π*r^2*h
	=1/3*π*r^2*12
	=4*π*r^2

Figure 2 Algebraic word problem and its sample answer

Appendix F: Sample word set in Sinhala-Sinhala lexical database