Fakulti Teknologi Maklumat dan Sains Kuantitatif

UNIVERSITI TEKNOLOGI MARA

JURNAL TEKNOLOGI MAKLUMAT DAN SAINS KUANTITATIF

Jilid 9, Bil. 1, 2007

ISSN 1823-0822

Universiti Teknologi MARA

Fakulti Teknologi Maklumat dan Sains Kuantitatif ISSN: 1823-0822 Jilid 9, Bil 1, 2007

UNIVERSITI TEKNOLOGI MARA

JURNAL TEKNOLOGI MAKLUMAT DAN SAINS KUANTITATIF

Kandungan **Muka Surat**

Universiti Teknologi MARA

DARI MEJA PERUNDING

Assalamualaikum w.b.k dan salam sejahtera kepada semua pencinta ilmu. Alhamdulillah, dapat kita terbitkan Jurnal Teknologi Maklumat dan Sains Kuantitatif Jilid 9, Bil. 1, 2007 ini. Saya merakamkan berbanyak-banyak terima kasih kepada mantan dekan, Prof.Madya Dr Adnan Ahmad yang memberikan galakan untuk penerbitan walaupun kadangkala susah untuk mendaptkan penulis-penulis yang berminat. Bermula dari 1 hb.Disember, 2007 fakulti telah diterajui oleh dekan yang baru, iaitu Prof. Dr Zainab Abu Bakar. Moga-moga dengan dekan yang baru, mudah-mudahan kita akan dapat suntikan semangat yang baru dan masing-masing baik dari fakulti di UiTM mahu pun dari IPTA yang lain berlumba-lumba untuk terus menulis .

Melalui penulisan dan bacaan dapat kita menambahkan ilmu pengetahuan kita. Kemajuan dan peningkatan tamadun maanusia adalah juga melalui penyebaran ilmu. Di era Teknologi Maklumat dan Komunikasi (ICT) dan globalisasi ini, penyebaran ilmu boleh dibuat secara pantas dengan melayari internet atau pun pmelalui blog-blog ilmiah. Namun begitu, hasil ilmuwan melalui karya penulisan dalam bentuk, jurnal, proceeding mahu pun laporan projek adalah masih releven dan penting . Malahan dewasa ini, penyebaran ilmu di pusat-pusat pendidikan tinggi mahu pun organisasi penyelidikan, kaedah ini masih jadi pilihan utama.

Saya berharap semua penulis-penulis semasa dan yang akan datang tetap gigih untuk menulis supaya karya kita dapat dimanfaatkan oleh semua pihak yang cintakan ilmu.Untuk para pendidik, hasil karya anda akan menjadi satu rangsangan untuk siswa-siswi muda kita mendaptkan ilham untuk giat dalam penulisan.

Terima kasih.

Ketua Penyunting.

Prof. Dr. Mohd Sahar Sawiran.

Automated Marking of Linear Algebraic Equation Using n-Gram Method

[1] Arsmah Ibrahim, [2]Zainab A Bakar, [3]Nuru'l–'Izzah Othman, [4]Nur Fuzaina Ismail

[1,2,3,4]Fakulti Teknologi Maklumat dan Sains Kuantiatif,
Universiti Teknologi MARA
40450 Shah Alam, SELANGOR
Mel-e:[1]arsmah@tmsk.uitm.edu.my, [2]zainab@tmsk.uitm.edu.my, [3]nurul@tmsk.uitm.edu.my, fuzaina@tmsk.uitm.edu.my[4]

Abstract:

This paper demonstrates the basic feasibility of using n-gram string matching method to grade on a set of four linear algebraic equations. The degree of correctness between the worked-out solutions submitted by respondents and the solutions schemes provided by the examiner is measured using Dice coefficient. Experiments have been conducted on 180 respondents for each question. The scores obtained by n-gram method are compared with the manually marked scores. The n-gram method is proven to be suitable when applied to these four linear algebraic equations as its performance is satisfactory and the results suggest feasibility. It is also realized that n-gram method has the advantage of producing uniformity and consistency.

Keywords: *n-gram, algebraic equations, Dice coefficient, manual marking*

1 Introduction

Automated marking is a marking mechanism that is processed by a computer via a computational scheme while manual marking is the mechanism processed by human beings based on a prepared solution scheme. Marking of mathematics tutorials manually is indeed laborious and time consuming. Thus, efforts in producing an efficient automated marking scheme are actively researched area. Although many computer-aided assessment packages have emerged, majority of them utilize the multiple-choice questions, filling in the blanks, and keying in end answer formats. Thus, the challenge of producing an efficient algorithm to grade symbolic texts in free-format is still wide open.

This paper concentrates on automated marking of the free-form or open-type format where each line is checked by an efficient computational means. The n-gram is well reported in document retrieval systems such as spelling corrections, word variants, text retrieval, string or pattern matching, hypertext, document retrieval, natural language, and information retrieval.

2 n-Gram Method

The n-gram method assumes that strings whose structures are highly similar have a high probability of having the same meaning. The similarity between two strings are measured by comparing unique n-gram from both strings which can be expressed as $n - gram(x \cap y)$. The number of common n-gram between both strings indicates the degree of similarity. A high value of common n-gram indicates a high degree of similarity (Angell et al. 1983; Owolabi & McGregor 1988).

Marking simple algebraic equations computationally via an algorithm that adopts the n-gram method with the string similarity approach has shown promising results (Arsmah & Zainab, 2003). Several experiments have been carried out on simple algebraic equation whereby the Dice coefficient is used in evaluating one term against another. Dice coefficients are calculated to represent the degree of accuracy of the respondents' solutions. Dice coefficient is mathematically expressed as

$$d_{i,j} = \frac{2\ common\_token(x_i \cap y_j)}{token(x_i) + token(y_j)} \qquad 1 \le i \le m,\ 1 \le j \le n \qquad [1]$$

where x_i is the *i-th* row of the solution scheme's string and y_j the *j-th* row of the respondent's solution scheme string where *i* and *j* are positive integers.
The best Dice coefficient D_j is chosen

$$D_j = \max_{1 \le i \le m} d_{i,j} \qquad [2]$$

Then the average Dice coefficient is calculated as

$$Average\ Dice\ coefficient = \frac{\sum_{j=1}^{n} D_j}{n} \qquad [3]$$

Line-by-line marking of simple algebraic equations by adopting the procedures above has shown promising results.

3 Test Collection

Four types of data are required: a set of linear algebraic equations; a solution scheme for automated marking; a set of worked-out solutions; and a solution scheme for manual marking.

A tutorial consisting of samples of linear algebraic equations must be designed in free-form or open format. The sample questions must be designed in increasing complexity. As this is the first attempt, very simple linear algebraic equations are set. Second, a solution scheme to the questions designed must be obtained for this will be required in the similarity measure. Third, a substantial number of worked-out solutions must be obtained for each question. These will be used as data for the automated marking algorithm and manual marking. The scores obtained will be compared. Manual graded scores will be used as the benchmark.

3.1 Sample Questions

Four simple questions in solving linear algebraic equations are designed for a tutorial class. Each question involves solving equations of the form $ax + b = 0$ where a and b are nonzero real numbers. Questions 1 – 3 designed take the simplest form of the linear algebraic equation where Question 1 involves two terms only, Question 2 involves three terms and Question 3 involves four terms. Question 4 however apparently is not linear but becomes linear when cross multiplied. Here it is assumed that the respondents would apply the basic rules in algebra of real numbers in solving the equation.

Question 1: $2x = 10$

Question 2: $3x - 15 = 9$

Question 3: $5x + 4 = 10 - 3x$

Question 4: Solve $\frac{x - 4}{x} = 3$

3.2 Solution Schemes

A mathematician prepares three sets of solution schemes. Scheme 1 consists of one typical solution for each question. Scheme 2 considers various possible ways of solving the four equations. The order of the solutions for both schemes is not regarded since every line of the solution will undergo repeated comparison as the search of similarity process progresses.

4 Implementation

Seven experiments of different variations are carried out ranging from the simplest to modified versions. The variation is based on the token type used, the n-gram values, the order of tokens and the solution schemes used for the Dice coefficient measure. Polya's method is employed for the manual marking.

4.1 Experiment 1

In Experiment 1, token of the first variation is used. The value of the n-gram is fixed to be '1'. The order of the tokens is **maintained**. All coefficients, symbols and variables including the repeated ones are considered as individual tokens. In other words, duplicate n-grams are **not removed**. They are treated as strings and are evaluated against solution Scheme 1. This is illustrated in Figure 1. For each question, the degree of accuracy or correctness between the worked-out solutions submitted by respondents and the solution Scheme 1 is measured using Dice coefficient.

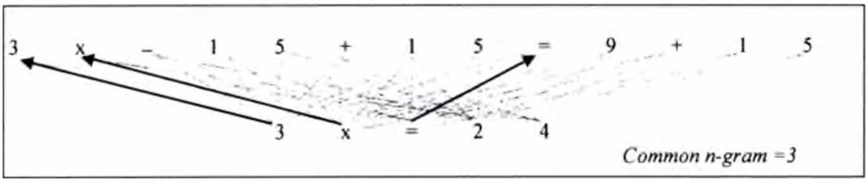

Figure 1: Illustration of Experiment 1

4.2 Experiment 2

In Experiment 2, token of variation 1 is also adopted. As in Experiment 1 the value of the n-gram is fixed to be '1'. However, the order of the tokens is not maintained but **sorted** in ascending order while duplicate n-grams are **removed**. The remaining tokens are treated as strings and are evaluated against solution Scheme 1 too. Figure 2 illustrates how the first lines of the worked-out solution and the solution scheme are sorted before the best match search scheme is done. Similarly, for each question the degree of accuracy or correctness between the worked-out solutions submitted by respondents and the solution Scheme1 is measured using Dice coefficient.

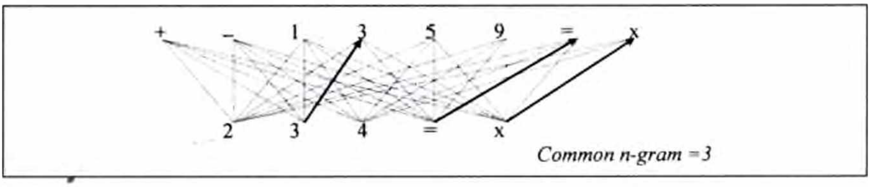

Figure 2: Illustration of Experiment 2

4.3 Experiment 3

In Experiment 3, varied values of n according to the size of each term inclusive of the '+' or '-' sign preceding it, is used. In these experiments, tokens are compared instead of characters and Scheme 1 solution is still used in evaluating the tokens. This is illustrated in the Figure 3.

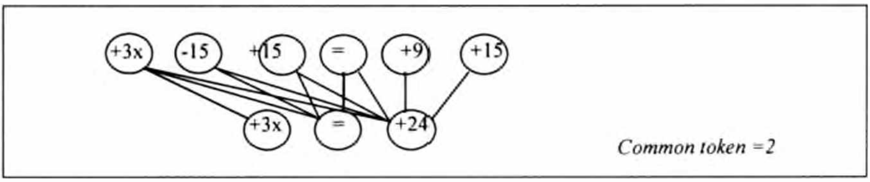

Figure 3: Illustration of Experiment 3

Similarly, the degree of correctness between the worked-out solutions submitted by respondents and the solution Scheme 1 used is measured using the Dice coefficient.

4.4 Experiment 4

Experiment 4 is carried out using varied values of *n*, according to the size of each term inclusive of the '+' or '-'signs preceding it. This experiment repeats Experiment 3 but solution Scheme 2 is used instead of solution Scheme 1 in evaluating the similarity of the tokens. The degree of correctness between the worked-out solutions submitted by respondents and the solution Scheme 2 used is measured using Dice coefficient.

4.5 Experiment 5

After performing Experiment 3 and Experiment 4 using tokens of the first variation, it is found that the degree of correctness is inaccurate. This is because tokens that are located at the wrong locations in the equation are counted in the Dice coefficient thus giving a higher value. This is illustrated in the following examples.

Example 1: (+6 / +3) +3 = x ⇔ (6/3) + 3 = 5

Example 2: +6 / (+3 +3) = x ⇔ 6/(3 +3) = 1

Example 1 and Example 2 have the same set of tokens but arranged in different order. The Dice coefficient of these equations is thus equals to 1.00. This implies perfect match. However this is not true in mathematical context because the original equations are different. This means that the order of characters or symbols in a mathematical equation must be preserved. So in Experiment 5, tokens of the second variation is employed (Figure 4). This experiment is similar to Experiment 4 except for the token variation.

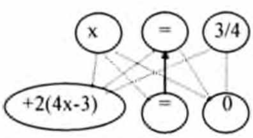

Figure 4: Illustration for Application of Token of Second Variation

4.6 Experiment 6

After performing Experiment 5 it is observed that if two tokens are only different by the arithmetic sign '+' or '-', then these tokens will not be counted (Figure 5).

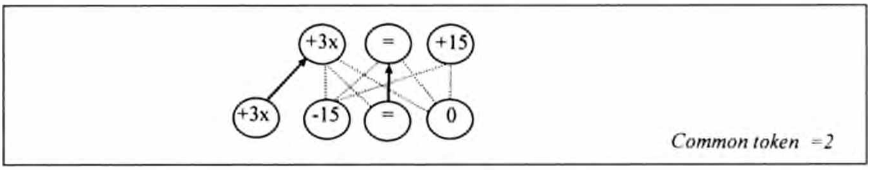

Figure 5: Illustration for Tokens Differed by '+' and '-' Symbol

If these tokens are placed on the left hand side of the '=' sign and closed under the algebraic laws that encompass properties such as commutative, associative, distributive, identity, inverse, and quotients (Flaunders & Price, 1975), only then these tokens will be counted in the Dice coefficient. In Figure 5 only two tokens are matched but after shifting all tokens to the left hand side of the '=' as in Figure 6, three tokens are matched.

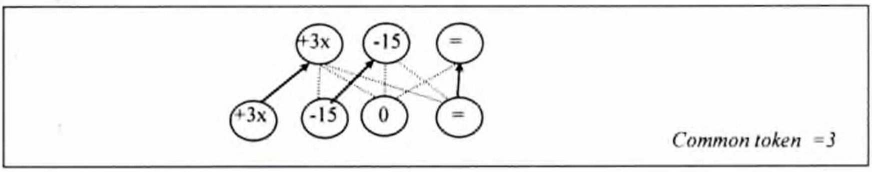

Figure 6: Tokens Differed by '+' or '–' Shifted to LHS

In Experiment 6, both the worked-out solutions and the prepared solution Scheme 2 are transformed into tokens of second variation. All the right hand side tokens are shifted to the left hand side of the '=' sign (Figure 6). In so doing all tokens with '+' will be changed to '-', and vice versa. In this way the chance of each respondent getting a better score is preserved.

4.7 Experiment 7

According to Zipf Law (1949) the most occurring term is considered to be common and removing it leaves the rest of the terms to be significant. Thus in Experiment 7 tokens '=' are removed and the remaining tokens of worked-out solutions and Scheme 2 are evaluated (Figure 7).

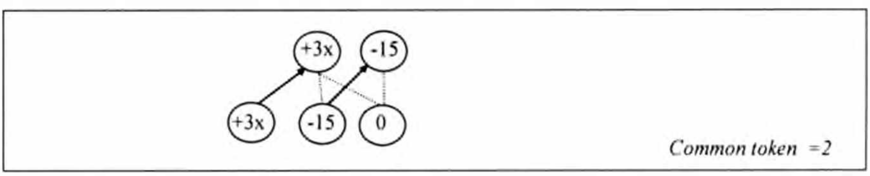

Figure 7: Common Symbol '=' Removed

4.8 Manual Marking

Manual marking is the conventional process of marking and giving scores to students' worked-out solution by hand. Five experienced Mathematics lecturers of UiTM are selected to examine the test collection gathered. As consistency is important in any assessment the examiners sat together to examine the sample questions to produce a common marking scheme. The concept implemented in the marking scheme designed is based on Polya's problem solving theory (1957). Polya put forward a four-point approach to learning i.e:

i. Understanding the problem.
ii. Devising a plan to solve the problem.
iii. Carrying out the plan.
iv. Checking the result.

Table 1: Manual Grading Solution Scheme

No	Solutions	Marks	Polya's Criteria			
			Understanding Problem	Devising Plan	Executing Plan	Checking Result
1	$2x = 10$ $x = 5$	 1	Reflected from answer given	• Determine x	• Divide RHS and LHS by 2	$x = 5$
2	$3x - 15 = 9$ $3x = 24$ $x = 8$	 1 1	Reflected from equation $3x = 24$	• Isolate x terms on LHS • Determine x	• Add 15 to RHS • Divide RHS and LHS by 3	$x = 8$
3	$5x + 4 = 10 - 3x$ $5x + 3x = 10 - 4$ $8x = 6$ $x = 6/8$ $x = 3/4$	 1 1 1 1	Reflected from equation $5x + 3x = 10 - 4$	• Collect all x terms to LHS • Determine x • Simplify answer	• Add 3x to 5x on LHS and minus 4 to RHS • Divide RHS and LHS by 8 • Divide deno and nume on RHS by 2	$x = 3/4$
4	$\frac{x - 4}{x} = 3$ $x - 4 = 3x$ $-2x = 4$ $x = -2$	 1 1 1	Reflected from equation $x - 4 = 3x$	• Remove denominator on LHS • Collect all x terms to LHS • Determine x	• Multiply RHS by x • Minus 3x to LHS and add 4 to RHS • Divide LHS and RHS by −2	$x = -2$
	TOTAL	10				

Based on these criteria a total of 10 marks is distributed among the 4 questions (Table 1). As the solution of Question 1 is very simple and consists of only one arithmetic operation, only 1 mark is allocated to the accurate answer. The examiners agree that 1 mark is conclusive to show the four criteria. As for Question 2, 2 marks are allocated. 1 mark is for the correct formation of the equation that reflects understanding and correct planning while the other 1 mark is allocated to the accurate answer. The solution of Question 3 basically involves four arithmetic operations to arrive to the answer, so 4 marks are allocated to it. 1 mark is for correct formation of the equation that follows the question, 1 mark for the algebraic manipulation to simplify the equation, 1 mark for the answer that follows and 1 mark for an accurate simplified answer. In Question 4, 3 marks are allocated. 1 mark is for the formation of the succeeding equation that reflects understanding and correct planning. The next 1 mark is for carrying out the plan correctly and 1 mark is allocated for the accurate answer. In all cases the checking of result criteria is achieved by substituting the variable with the answers obtained. It can be observed that the manual grading solution scheme is a subset of the automated solution scheme (Table 1).

All the marks obtained from each respondent of each question are scaled between 0.0 – 1.0. The main reason for doing this is to provide a uniform representation of the marks similar to Dice Coefficient, 1.00 being 100% similar or correct while 0.0 implies no match or totally incorrect.

5 Results and Discussion

It can be seen from Table 2, that employing solution Scheme 2 and the token concept in Experiments 4, 5, 6 and 7 can successfully identify correct worked-out solutions. Here the Dice coefficients are all equal to 1.0000, which imply perfect match are found for all the 4 worked-out solutions. The performance for both the automated and manually graded correct worked-out solutions, are exactly the same. All scaled marks are also equal to 1.0000 that implies that the worked-out solutions by respondents are correct according to the solution scheme prepared by the examiner. Thus it can be deduced that the solution scheme used in manual grading prepared by the examiners is a subset of the solution scheme used in the automated grading.

Table 2: Case Correct Worked-out Solution

	Question	1	2	3	4
	Respondent	180	180	141	39
Automated Marking	Expt 1	0.7519	0.7593	0.8034	0.9030
	Expt 2	0.8128	0.8229	0.9067	0.9231
	Expt 3	0.5470	0.6147	0.7333	0.7094
	Expt 4	1.0000	1.0000	1.0000	1.0000
	Expt 5	1.0000	1.0000	1.0000	1.0000
	Expt 6	1.0000	1.0000	1.0000	1.0000
	Expt 7	1.0000	1.0000	1.0000	1.0000
Manual Marking	Exmr 1	1.0000	1.0000	1.0000	1.0000
	Exmr 2	1.0000	1.0000	1.0000	1.0000
	Exmr 3	1.0000	1.0000	1.0000	1.0000
	Exmr 4	1.0000	1.0000	1.0000	1.0000
	Exmr 5	1.0000	1.0000	1.0000	1.0000

In the case of incorrect worked-out solutions the value of the Dice coefficients are reduced upon employing solution Scheme 2 and the first token concept. However when the token of the second variation is used and the most frequently occurring tokens are removed (Experiment 7) the Dice coefficient is reduced further. This implies that worked-out solutions that are either totally correct or totally incorrect are successfully identified (Table 3). There exists a significant difference in performance of the automated and the manually graded. Here it can be observed that all examiners gave zeroes (0.0000) for all the wrong worked out solutions. Unfortunately the Dice coefficients of the automated graded worked-out solutions do not achieve this! Between Experiments 4, 5, 6 and 7, Experiment 7 seems to have the lowest Dice coefficient, which is 0.4000. However this value is still rather high for an incorrect solution. This implies that further refinement must be made to the tokens and the automated solution scheme. Nevertheless for Experiment 5, Experiment 6, and Experiment 7 it can be conjectured that any Dice coefficients values that are less than or equal 0.7083, 0.6818, and 0.5833 respectively, can be replaced by zero (0.0000). This is because they are supposed to imply wrong worked-out solutions. Overall it seems like the appropriate threshold to give zero mark for any worked-out solution is anything not more than 0.6 (≤ 0.6000).

Table 3: Case Incorrect Worked-out Solution

	Question	1	2	3	4
	Respondent	173	50	50	173
Automated Marking	Expt 1	0.6190	0.6000	0.6830	0.8157
	Expt 2	0.6970	0.6940	0.7247	0.8542
	Expt 3	0.5833	0.5917	0.5667	0.5530
	Expt 4	0.7333	0.7417	0.6667	0.6652
	Expt 5	0.6667	0.7083	0.6250	0.6667
	Expt 6	0.6667	0.6583	0.5833	0.6818
	Expt 7	0.5000	0.4375	0.4000	0.5833
Manual Marking	Exmr 1	0.0000	0.0000	0.0000	0.0000
	Exmr 2	0.0000	0.0000	0.0000	0.0000
	Exmr 3	0.0000	0.0000	0.0000	0.0000
	Exmr 4	0.0000	0.0000	0.0000	0.0000
	Exmr 5	0.0000	0.0000	0.0000	0.0000

Removing the '=' further gives an even better representation of partially correct worked-out solutions (Table 4). The Dice coefficient in Experiments 5, 6, and 7 of Question 1 Respondent 1 is 0.8333, 0.8333, and 0.7500 respectively implying partial credits are given, whereas all examiners give zero mark to Question 1 of this respondent. This is due to the fact that the manually graded solution scheme allocates only 1 mark for this problem. The mark is allocated for an accurate answer only. The examiners assume that it is such a simple problem that nobody should get it wrong! In the original script an error exists in the second line where instead of writing it as '*x=10/2*' the respondent wrote it as '*2x=10/2*'. However the end answer is still correct. For that all the examiners give zero (0.0000) mark following the manual solution scheme while the automated grading give partial marks for the similarity with '*2x/2=10/2*' of the automated solution scheme. Thus in this case it appears that the automated grading scheme is more sensitive than the manual grading scheme.

For Question 2 Respondent 151, Examiners 2, 4, and 5 are generous and give full marks while Examiners 1 and 3 are strict and give zero (0.0000). Upon checking the original answer scripts it is found that the fault is only unpaired parenthesis. The respondent missed the closing parenthesis in the expression but calculated the answer correctly and accurately! Here human judgment takes place. Apparently it implies that Examiners 1 and 3 strictly insist perfect symbolic writing unlike Examiners 2, 4 and 5. So Dice coefficients that are greater than 0.8000 in value from Experiments 5, 6 and 7 can be considered valid.

For partially correct worked-out solutions of Question 3 Respondent 86 Examiners 4 and 5 give the same marks of 0.7500. It can be observed that the Dice coefficient from Experiment 7 perfectly matches the marks of the examiners. Upon checking the original scripts it is found that the mistakes done by the respondent is inaccurate simplification that ended up with a wrong end answer.

For partially correct worked-out solutions of Question 4 of Respondent 72, all examiners give equal partial marks. However the automated graded scheme gives partial marks of higher numerical value. In other words they do not match perfectly. Upon checking the original script it is found that the error that is made by the respondent is leaving out the '-' sign at the end answer. For this mistake 1 mark is reduced from a total of 3 marks in the manual grading scheme. It can be said that the Dice coefficient of 0.8000 gives a lower penalty relative to the penalty given by the examiners. Thus this value can also be considered valid.

Table 4: Case Partially Correct Worked-out Solution

	Question	1	2	3	4
	Respondent	1	151	86	72
Automated Marking	Expt 1	0.9375	0.8333	0.9000	0.9314
	Expt 2	1.0000	0.9375	0.9000	0.9532
	Expt 3	0.9167	0.7500	0.9000	0.8267
	Expt 4	0.9167	0.9286	0.9000	0.9314
	Expt 5	0.8333	0.8650	0.8333	0.8500
	Expt 6	0.8333	0.8750	0.8333	0.8500
	Expt 7	0.7500	0.8333	0.7500	0.7900
Manual Marking	Exmr 1	0.0000	0.0000	1.0000	0.6667
	Exmr 2	0.0000	0.0000	1.0000	0.6667
	Exmr 3	0.0000	0.0000	1.0000	0.6667
	Exmr 4	0.0000	1.0000	0.7500	0.6667
	Exmr 5	0.0000	1.0000	0.7500	0.6667

After comparing results from Experiments 5, 6, and 7 with the manually graded worked out solution for totally correct, totally wrong, and partially correct worked out solutions, it clearly shows the importance of defining what a token must consists of. In addition, the worked-out solution scheme should have all possible listing of solutions. Worked-out solutions of average Dice coefficient 1.0000 are considered correct and get maximum score. Partially correct worked-out solutions are given scores according to the respective Dice coefficients. While worked out solutions that get less than 0.6000 are given zero mark.

6 Conclusion

From the results obtained it can be concluded that n-gram string matching method can be used successfully for online marking of these 4 linear equations for free-form or open-type worked-out solutions that are checked line by line. Tokens of varied n-gram values must be carefully grouped to avoid wrong worked-out solutions that produce a high Dice coefficient. Results obtained using token of second variation manage to segregate distinctively the correct and the wrong worked-out solutions. Solutions schemes must be prepared extensively and include all possible locations of terms either on the left hand side or right hand side of the '='. The size of a token can be further improved.

It is also observed that the manually graded method shows inconsistency in the marking due to human judgment as demonstrated in the partially correct worked-out solutions particularly for Question 1 and 2. It is suggested that manual marking scheme should not be end-answer oriented. Marks must be allocated for the devising and carrying out plan criteria.

References

Angell, R.C., Freund, G.E. & Willett, P. 1983. *Automatic Spelling Correction Using a Trigram Similarity Measure.* Information Processing and Management 19. 255-261.

Flaunders, H. & Price, J. 1975. Algebra. Academic Press Inc.

Owolabi, O. & McGregor, D.R. 1988. *Fast Approximate String Matching. Software-Practice and Experience* 18(4). 383-293.

Polya. 1957. "*How to Solve It*", 2nd ed., Princeton University Press, ISBN 0-691-08097-6

Arsmah Ibrahim & Zainab Abu Bakar. 2003. *Experimenting n-gram Method on Linear Algebraic Equations for Online Grading.* Proceedings of the International Conference on Research Education in Mathematics, 2-4 April 2003

Zipf, H.P. 1949. *Human Behavior and The Principle of Least Effort.* Cambridge, Massachussetts: Addison Wesley.