Reducing expression size using rule-based integration

D.J. Jeffrey and A.D. Rich

[1] Department of Applied Mathematics, The University of Western Ontario, London, Canada. email: djeffrey@uwo.ca

[2] 62-3614 Loli'i Way, Kamuela, Hawaii, USA. email: Albert_Rich@msn.com

Abstract. This paper describes continuing progress on the development of a repository of transformation rules relevant to indefinite integration. The methodology, however, is not restricted to integration. Several optimization goals are being pursued, including achieving the best form for the output, reducing the size of the repository while retaining its scope, and minimizing the number of steps required for the evaluation process. New optimizations for expression size are presented.

1 Introduction

The methods of integration can be conveniently divided into several categories.

- Look-up tables. These are collections or databases, such as [4], which try to list all possible integrals, each in a general form. Many special cases are also listed separately.
- Rule-based rewriting. The databases used are smaller than those for the look-up tables. They contain rules for transforming a given integral into one or more simpler integrals, together with rules for completing the evaluation in terms of known functions.
- Algorithmic methods. Under this heading, we include Risch integration, Rothstein-Trager-Rioboo integration, and others, which require extended computations.

A table of reduction rules can serve more roles than merely the database for an evaluation system; it can also serve as a repository for mathematical knowledge. Each rule can be annotated with information on its derivation, with references to the literature, and so on. An evaluation system can display transformations as they are used, for the information of users.

Here, we consider the repository of transformation rules for indefinite integrals that is described in [5, 6]. We shall refer to it by the acronym Rubi: RUle-Based Integrator. We review the general state of the repository and then focus on particular aspects, namely, its efficiency, and the selection of output forms. Procedures have been written in Mathematica to implement the evaluation of integrals using the repository, and these procedures have been the basis of testing and comparisons.

The role of rule-based approaches, and how they should complement algorithmic methods, can be a subject of debate. For example, Fateman wrote, as part of a review of the system MATHEMATICA [2]

> "Yet the evidence of the past several decades casts strong doubt on the idea that an efficient version of mathematical knowledge can be imparted to a symbolic system *primarily* by rule-transformations on trees."
> **Richard Fateman (1992)**

Owing to poor implementations, rule-based systems have a reputation for being inefficient and plagued by endless loops. This paper, however, describes the crafting of a rule-based repository (RUBI) that is compact, efficient, transparent and modular. We shall not address the combining of RUBI with algorithmic approaches, as would be required to arrive at a full integration system, but concentrate on the constructing of a database of knowledge, with examples of how it performs in practice.

It must be emphasized again that what is not being described is a scheme for table look-up. Such schemes were described, for example, in [3]. The approach there was to consider data structures and search techniques which would allow them to encode all the entries in reference books such as [1]. Adopting this approach for integration — or *a fortiori* for all simplification — would result in huge databases which would be unwieldy to maintain, debug and utilize. The set of rules described here is relatively compact, verifiable and efficient.

2 Basic details of system

Here, we give a brief account of the RUBI system. At the time of testing, it consisted of 1377 reduction rules. Each rule is an entry in the database and consists of the following fields.

- Conditions under which the reduction rule is applied. These conditions result either from requirements for the validity of the transformation, or from requirements that the transformation be a reduction, meaning a step towards evaluation of the integral.
- The transformation from one expression to another.
- Comments recording the source of the rule (usually a reference to one or more standard reference books) or other useful information.

It should be noted that programming constructs, such as loops or branching statements are never used. Examples of these rules are given below in section 4 (without the comments).

The total size of the database (including comments) was 554 Kb. This is an uncompressed text file. About one third of the file consists of comment text. Procedures using the pattern-matching functions of MATHEMATICA were written to apply the database to the test problems, and no attempt is made to measure the sizes of subsystems of MATHEMATICA used.

The construction and selection of the rules is based on the principle of mutual exclusivity. For a database of reduction rules to be properly defined, at most one of the rules can be applicable to any given expression. Mutual exclusivity is critical to ensuring that rules can be added, removed or modified without affecting the other rules. Such stand-alone, order-independent rules make it possible to build a rule-based repository of knowledge incrementally and as a collaborative effort.

3 Performance Comparison with Other Systems

In order to provide quantitative evidence of the benefits of rule-based integration, we present a comparison of the performance of various computer algebra systems on a test suite containing 7927 problems. The performance measure is based on the validity and simplicity of the expressions returned. Other performance measures, such as speed, have been measured, but direct comparisons can at present be made only with MATHEMATICA, and so here the emphasis is on expression size, until a variety of platforms can be compared for speed[3]. We note in passing, however, that smaller expression sizes will also contribute to speed advantages.

The expression given for each integral was checked against the simplest form, obtained from published integral tables, or from integration by hand. For each problem, the integration result was differentiated by the system being tested, the derivative subtracted from the integrand, and the system asked to test whether the result was zero. Each test yielded one of the following 4 judgements:

- **Optimal**: Correct and close to the best form.
  Example: $\displaystyle\int \frac{5x^4\,\mathrm{d}x}{(1+x)^6} = \frac{x^5}{(1+x)^5}$.
- **Messy**: Correct, but the expression is overly large. E.g.
  $\displaystyle\int \frac{5x^4\,\mathrm{d}x}{(1+x)^6} = -\frac{1}{(1+x)^5} - \frac{5}{(1+x)} - \frac{10}{(1+x)^3} + \frac{5}{(1+x)^4} + \frac{10}{(1+x)^2}$.
  Note that the optimal and messy results differ by a constant, and the optimal form cannot be obtained by simplification of the messy.
- **Inconclusive**: No result was obtained in 60 seconds, or the result could not be verified, usually because the output was so large that the simplifier failed while attempting to differentiate and reduce to zero.
- **Invalid**: The difference between the derivative and integrand was not zero.

The performances on the test suite of MAPLE, MATHEMATICA and the present rule-based system RUBI are presented in the tables below. Since RUBI

[3] MATHEMATICA has been used to implement RUBI and a comparison with its built-in `Integrate` command shows RUBI to be faster on the test suite by a factor of 10. The test suite has been ported to MAPLE, but the different syntax for pattern matching has so far prevented RUBI from being ported. Therefore only the output forms can be compared.

was developed using the test suite, its good performance is to be expected, but even so, the favourable comparison with the other systems remains valid.

Although the test suite of 7927 problems is large, the problems themselves are all part of mainstream calculus, and therefore even the small rate of 3% invalid results for the large commercial systems is disappointing, to say the least. However, for the purposes of this paper, the emphasis of the main benefit of Rubi in this comparison lies in the simpler form of the results. Since the main advantage lies in the simplicity of the results, we concentrate here on how Rubi achieves its results, by presenting two case studies.

4 First Case Study: Alternative Strategies

The first study concerns an optimization to reduce output size. In order to obtain quantitative measures for expression size, utility functions have been written in Mathematica and Maple that count the number of nodes in the internal tree representation of a particular expression. Although there are variations in the internal representations of expressions, the functions provide comparable measures in the two systems.

We consider the problem of evaluating the integral

$$\int \frac{x^m \,\mathrm{d}x}{(a+bx)^{12}} \;, \tag{1}$$

for different values of $m \in \mathbb{Z}$. This is a special case of the more general problem

$$I(a,b,c,d,m,n) = \int (a+bx)^m (c+dx)^n \,\mathrm{d}x \;, \tag{2}$$

where $m, n \in \mathbb{Z}$, and $a, b, c, d \in \mathbb{C}$.

Our aim is to minimize the number of terms in the expression for the integral. As a starting point, we can use the standard integrators in Mathematica and Maple to evaluate the integral, and plot the expression sizes of the results as functions of m. Figures 1 and 2 show the expression counts for the two systems.

We have extracted below from Rubi the 9 transformation rules applying to the integral class (2). Each rule is presented in the following form:
N: Necessary conditions for mathematical validity.
T: The transformation rule $A \to B$.
S: Simplification conditions to ensure the transformation yields a simplification.

All rules require that a, b, c, d, m, n do not contain x, and that $b \neq 0$. The rules are:

1. **T**: $\displaystyle\int \frac{dx}{a+bx} \to \frac{\ln(a+bx)}{b}$.
2. **N**: $m+1 \neq 0$
   T: $\displaystyle\int (a+bx)^m dx \to \frac{(a+bx)^{m+1}}{(m+1)b}$.

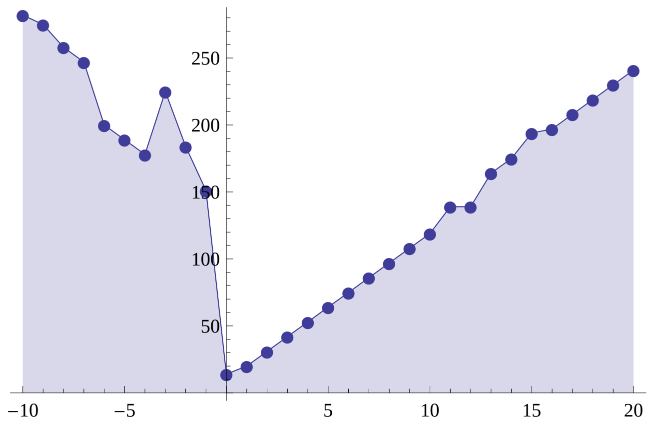

Fig. 1. The node count for expressions returned by MATHEMATICA 7 for the integral in (1). The horizontal axis shows values of the exponent m, while the vertical axis shows the node count for the corresponding expression for the integral.

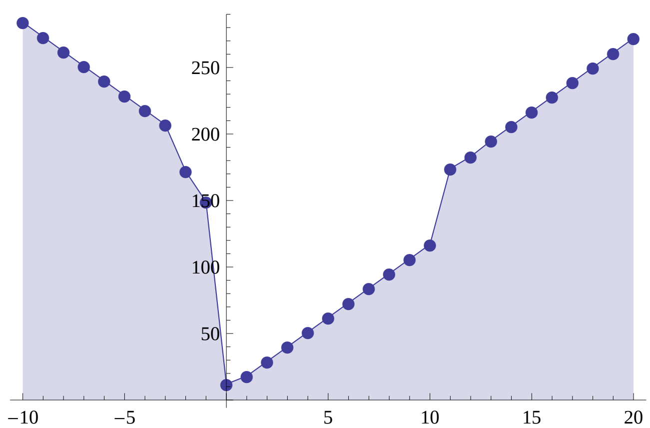

Fig. 2. The node count for expressions returned by MAPLE 13 for the integral in (1). The horizontal axis shows values of the exponent m, while the vertical axis shows the node count for the corresponding expression for the integral.

3. **N**: $bc - ad = 0,\ m + n + 1 = 0$
   T: $\int (a + bx)^m (c + dx)^n \,\mathrm{d}x \to (a + bx)^{m+1}(c + dx)^n \ln(a + bx)/b$.
4. **N**: $bc - ad = 0,\ m + n + 1 \neq 0$
   T: $\int (a + bx)^m (c + dx)^n \,\mathrm{d}x \to \dfrac{(a + bx)^{m+1}(c + dx)^n}{b(m + n + 1)}$.
5. **N**: $bc - ad \neq 0$
   T: $\int (a + bx)^{-1}(c + dx)^{-1} \,\mathrm{d}x \to \dfrac{\ln(a + bx) - \ln(c + dx)}{bc - ad}$.
6. **N**: $bc - ad \neq 0,\ m + n + 2 = 0,\ n \neq -1$
   T: $\int (a + bx)^m (c + dx)^n \,\mathrm{d}x \to -\dfrac{(a + bx)^{m+1}(c + dx)^{n+1}}{(n + 1)(bc - ad)}$.
7. **N**: $m + n + 1 = 0,\ m > 0,\ bc - ad \neq 0$
   T: $$\int (a + bx)^m (c + dx)^n \,\mathrm{d}x \to -\frac{(a + bx)^m}{dm(c + dx)^m} + \frac{b}{d}\int (a + bx)^{m-1}(c + dx)^{-m} \,\mathrm{d}x\ .$$
8. **N**: $bc - ad \neq 0,\ m + n + 1 \neq 0,\ n > 0$
   T: $$\int (a + bx)^m (c + dx)^n \,\mathrm{d}x \to \frac{(a + bx)^{m+1}(c + dx)^n}{b(m + n + 1)} + \frac{n(bc - ad)}{b(m + n + 1)}\int (a + bx)^m (c + dx)^{n-1} \,\mathrm{d}x\ .$$
   S: $(2n + m + 1 < 0 \vee m + n + 1 > 0) \wedge (m < 0 \vee n \leq m)$
9. **N**: $bc - ad \neq 0,\ n + 1 \neq 0$
   T: $$\int (a + bx)^m (c + dx)^n \,\mathrm{d}x \to -\frac{(a + bx)^{m+1}(c + dx)^{n+1}}{(n + 1)(bc - ad)} + \frac{(m + n + 2)b}{(bc - ad)(n + 1)}\int (a + bx)^m (c + dx)^{n+1} \,\mathrm{d}x\ .$$
   S: $n < -1,\ m < 0 \vee 2m + n + 1 \geq 0$.

We wish to show how these rules are optimized relative to other possible sets of rules. Specifically, we shall compare these rules with a set in which the simplification conditions in rules 8 and 9 are modified. We start, however, with remarks on the rules as presented.

4.1 Remarks

1. An alternative strategy to the set of transformations shown here would be to define rules for the simpler integrand $x^m(a+bx)^n$, and then use the linear substitution $u = c+dx$ to transform expressions of the form $(a+bx)^m(c+dx)^n$ into the simpler form. This strategy was explored, but we discovered that several more rules are required when starting from the non-symmetrical form $x^m(a+bx)^n$ than when starting with the symmetrical $(a+bx)^m(c+dx)^n$. This is because two versions each of rules 7, 8 and 9 had to be given depending upon whether the exponent of the monomial or the linear factor had to be incremented or decremented. This subtle, but important, point shows that sometimes defining more general rules leads to a simpler repository.

2. It should be noted that rule 6 is in fact a special case of rule 9. It is included because it is convenient to have an explicitly non-recursive entry.
3. Rules 8 and 9 respectively increment and decrement one of the exponents of the integrand. Unlike the other rules, it is not always obvious which of these two rules should be applied to a given integrand in order to minimize the number of steps required to integrate it. This choice is the subject of our optimization.

5 Integration strategies

The rules stated above describe a complete strategy for integration of the given class of integrals. The strategy is not unique, however, and other strategies might be more efficient. We therefore describe two other strategies and compare them with the preferred strategy.

5.1 Preliminary strategy 1

We replace rule 8 with a rule 8a, in which the simplification conditions are removed. Thus we have

8a. **N**: $bc - ad \neq 0$, $m + n + 1 \neq 0$, $n > 0$

T: $$\int (a+bx)^m (c+dx)^n \,\mathrm{d}x \to \frac{(a+bx)^{m+1}(c+dx)^n}{b(m+n+1)} + \frac{n(bc-ad)}{b(m+n+1)} \int (a+bx)^m (c+dx)^{n-1} \,\mathrm{d}x \ .$$

The effect of removing the restrictions is that all integrals will be reduced until one of the exponents becomes zero, at which point rules 1 to 6 will terminate the reduction. When this strategy is applied to the test case (1), the sizes of the results are as shown in figure 3.

The dip for the case $m = 10$ is important. For this case, rule 6 provides a direct one-step integration to a very compact form:

$$\int \frac{x^{10}\,\mathrm{d}x}{(1+x)^{12}} = \frac{x^{11}}{11(1+x)^{11}} \ .$$

This possibility is not noticed by the standard integrators of Mathematica and Maple, as can be seen in figures 1 and 2.

5.2 Preliminary strategy 2

We now remove the restrictions from rule 9, and place it above rule 8. Thus the rule becomes

9a **N**: $bc - ad \neq 0$, $n + 1 \neq 0$

T: $$\int (a+bx)^m (c+dx)^n \,\mathrm{d}x \to -\frac{(a+bx)^{m+1}(c+dx)^{n+1}}{(n+1)(bc-ad)} + \frac{(m+n+2)b}{(bc-ad)(n+1)} \int (a+bx)^m (c+dx)^{n+1} \,\mathrm{d}x$$

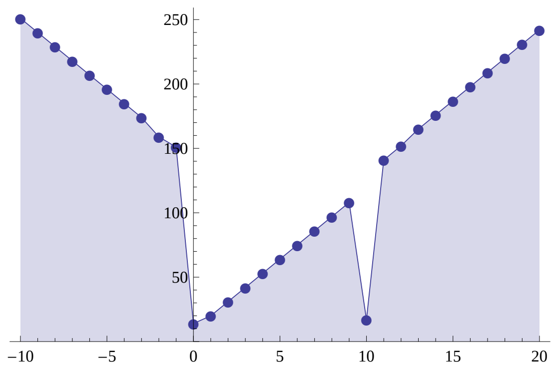

Fig. 3. The node count for expressions returned by the first alternative integration strategy for the integral in (1). The horizontal axis shows values of the exponent m, while the vertical axis shows the node count for the corresponding expression for the integral.

The effect of this is to increase one negative exponent until rule 6 can be applied. The resulting statistics on the size of integral expressions is shown in figure 4.

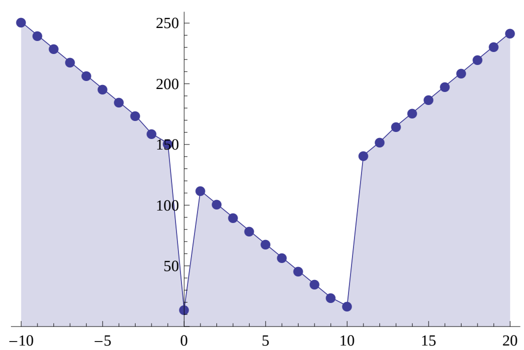

Fig. 4. The node count for expressions returned by the second alternative integration strategy for the integral in (1). The horizontal axis shows values of the exponent m, while the vertical axis shows the node count for the corresponding expression for the integral.

The dip at $m = 0$ is a result of rule 2 being applied before the general rules.

5.3 An optimal strategy

Clearly, one can obtain smaller expression sizes if one can switch between the two strategies just tested. This is what is done in rules 8 and 9 as presented. For the test case, the two points $m = 10$ and $m = 0$ are targets and for $m \le 5$ the integrands are moved towards $m = 0$, while for $m > 5$ they are moved towards $m = 10$. The generalization to other powers is shown in rules 8 and 9. The resulting expression sizes are shown in figure 5.

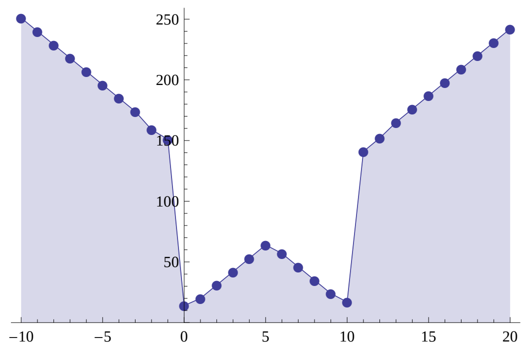

Fig. 5. The node count for expressions returned by the optimal integration strategy for the integral in (1). The horizontal axis shows values of the exponent m, while the vertical axis shows the node count for the corresponding expression for the integral.

5.4 Comparison with other methods

An obvious algorithmic approach to integral (1) is to expand the fraction using partial fractions, and then integrate each term. This gives results similar to those found using Maple and Mathematica. One of the advantages of Rubi is that such special cases can be identified and taken advantage of. One of the useful services that computer-algebra systems can offer mathematicians is the identification of special cases. The general algorithms preferred by Maple and Mathematica can succeed on large problems which Rubi is not yet capable of tackling. However, for smaller problems, where special cases might exist, Rubi is to be preferred.

6 Second Case Study: Two-part reduction

The second case study involves the class of integrals

$$J(m,n,p) = \int x^m (a+bx)^n (c+dx)^p \, \mathrm{d}x \ , \tag{3}$$

where the dependence of J on a, b, c, d has been suppressed, since we shall focus on the powers. We require $m \in \mathbb{Z}$, and $n, p \in \mathbb{Q}$. The aim is to reduce the integral in (3) to integrals with known solutions. In this case, the problems with known solutions are $J(0, N, P)$ and $J(M, N, N)$. It is straightforward to derive the equality

$$J(m, n, p) = (1/b)J(m-1, n+1, p) - (a/b)J(m-1, n, p) \tag{4}$$

An obvious strategy for $m > 0$ is to use this relation to reduce all integrals to the form $J(0, N, p)$. Thus, using the above conventions for describing a reduction rule, the rule reads

11. **T**: $J(m, n, p) \to \frac{1}{b}J(m-1, n+1, p) - (a/b)J(m-1, n, p)$
    S: $m \in \mathbb{Z}$, $m > 0$, $n, p \in \mathbb{Q}$, $n - p < 0$

At this point the algorithmically oriented person jumps to a composite rule by applying (4) m times to obtain

$$J(m, n, p) = \frac{1}{b^m} \sum_{k=0}^{m} \binom{m}{k} (-a)^k J(0, n+k, p) \ . \tag{5}$$

This, however, falls again into the trap that awaits grand algorithmic, or general-formula based, approaches. There are many special-case simplifications that will be skipped over by (5). Because the formula is derived for generic n, p, it can have no special behaviour for special cases. For example, if there exists k such that $n + k = p$ and $k < m$, then some terms can be removed from the sum and simplified separately, using the special case $J(m-k, n+k, p) = J(M, p, p)$. One of the differences between different computer systems is the extent to which they attempt intermediate simplifications. Using a step-based series of transformations (as Rubi does) each intermediate result can be tested for simplification before continuing.

For the case $m < 0$, we rewrite (4) as

$$J(m, n, p) = (1/a)J(m, n+1, p) - (b/a)J(m+1, n, p) \tag{6}$$

Applying this reduction k times, we would obtain

$$J(m, n, p) = \frac{1}{a^k} \sum_{i=0}^{k} \binom{k}{i} (-b)^i J(m+i, n+k-i, p) \tag{7}$$

The terms in the sum can be evaluated whenever $m + i = 0$ or $n + k - i = p$. Clearly, the latter condition requires that initially $n - p \in \mathbb{N}$. Therefore, the integral $J(m, n, p)$ will be evaluated after at most $\max(n - p, m)$ steps. As in the $m > 0$ case, however, it is better to apply the reduction stepwise in order to obtain the maximum benefit from intermediate, special case, simplifications.

As an example of the above rules, we present the same integral calculated by Rubi, by Mathematica and by Maple. First, Rubi:

$$\int \frac{\sqrt{2+3x}\,\mathrm{d}x}{x^2(5-x)^{3/2}} = \frac{2\sqrt{2+3x}}{25\sqrt{5-x}} - \frac{\sqrt{5-x}\sqrt{2+3x}}{25x} - \frac{21}{25\sqrt{10}} \operatorname{arctanh} \frac{\sqrt{10-2x}}{\sqrt{10+15x}}$$

Next, Mathematica:

$$= \frac{1}{500}\left(\frac{20(-5+3x)\sqrt{2+3x}}{x\sqrt{5-x}} + 21\sqrt{10}\ln\left(21\sqrt{10}x\right)\right.$$
$$\left. -21\sqrt{10}\ln\left(50\left(20+13x+2\sqrt{10}\sqrt{5-x}\sqrt{2+3x}\right)\right)\right)$$

Finally, Maple:

$$= -\frac{1}{500}\left(21\sqrt{10}\,\mathrm{arctanh}\left(1/20\frac{(20+13x)\sqrt{10}}{\sqrt{10+13x-3x^2}}\right)x^2\right.$$
$$-105\sqrt{10}\mathrm{arctanh}\left(1/20\frac{(20+13x)\sqrt{10}}{\sqrt{10+13x-3\,x^2}}\right)x + 60x\sqrt{10+13\,x-3x^2}$$
$$\left.-100\sqrt{10+13x-3x^2}\right)\sqrt{5-x}\sqrt{2+3x}\,(-5+x)^{-1}\,\frac{1}{\sqrt{10+13x-3x^2}}x^{-1}$$

There is a disadvantage, however, to stepwise application of the above reduction, a disadvantage well known in other contexts. This is the repeated evaluation of the same integral during recursive calls. The standard example of this effect is the recursive evaluation of Fibonacci numbers. This is paralleled in applications of (4) and (6). This effect was one reason that Maple introduced its `option remember` early in its development. The important additional feature present here, that is not present in the Fibonacci example, is the possibility of different simplification options directing the computation to simpler results.

7 Concluding remarks

In [5], a number of advantages of rule-based simplification were listed. These included (see reference for details).

- Human and machine readable.
- Able to show simplification steps.
- Facilitates program development.
- Platform independent.
- White box transparency.
- Fosters community development.
- An active repository.

In this paper we have shown that an additional advantage of rule-based evaluation, illustrated in the integration context, is greater simplicity of results. Finally, we wish to point out that the integration repository described here has been published on the web [6], and is available for viewing and testing by all interested people.

References

1. Abramowitz, M. & Stegun, I., *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. US Government Printing Office, 1964. 10th Printing December 1972.
2. Fateman, Richard J., A Review of Mathematica, J. Symb. Computation, **13**(5), 1992.
3. Einwohner, T.H. & Fateman, Richard J., *Searching techniques for integral tables*, Proceedings ISSAC '95, pp 133-139, ACM Press, 1995.
4. Gradshteyn, I.S. & Ryzhik, I.M., Table of Integrals, Series and Products. Academic Press 1965.
5. Rich, Albert D. & Jeffrey, David J., *A knowledge repository for indefinite integration based on transformation rules.* In: Intelligent Computer Mathematics, edited by J. Carette, L. Dixon, C.S. Coen, S.M. Watt. LNAI 5625, pp 480–485, Springer 2009.
6. Rich, Albert D. *Indefinite Integration Rules*, `www.apmaths.uwo.ca/RuleBasedMathematics`

Test Items		Rubi: Rule-based Integrator			
Integrand	**Number**	**Optimal**	**Messy**	**Inconc.**	**Invalid**
Rational	1426	1424	1	1	0
Algebraic	1494	1483	8	3	0
Exponential	456	452	0	4	0
Logarithmic	669	667	0	2	0
Trigonometric	1805	1794	8	3	0
Hyperbolic	1386	1379	6	1	0
Inverse trig	283	281	0	2	0
Inverse hyperbolic	342	335	2	5	0
Special functions	66	66	0	0	0
Percentages		99.4%	0.3%	0.3%	0%

Table 1. The integration test suite, with the numbers of problems broken down in categories. The performance of the Rule-based Integrator (Rubi) is given using measures described in the text.

Test Items		Maple			
Integrand	**Number**	**Optimal**	**Messy**	**Inconc.**	**Invalid**
Rational	1426	1176	249	0	1
Algebraic	1494	1126	277	45	46
Exponential	456	351	63	37	5
Logarithmic	669	284	161	194	30
Trigonometric	1805	1054	619	83	49
Hyperbolic	1386	521	641	181	43
Inverse trig	283	206	64	5	8
Inverse hyperbolic	342	159	96	55	32
Special functions	66	38	1	25	2
Percentages		62.0%	27.4%	7.9%	2.7%

Table 2. The performance of Maple on the test suite, using measures described in the text.

Test Items		Mathematica			
Integrand	**Number**	**Optimal**	**Messy**	**Inconc.**	**Invalid**
Rational	1426	1239	187	0	0
Algebraic	1494	1228	246	18	2
Exponential	456	406	32	12	6
Logarithmic	669	581	84	4	0
Trigonometric	1805	1212	573	3	17
Hyperbolic	1386	911	464	6	5
Inverse trig	283	211	62	10	0
Inverse hyperbolic	342	198	140	3	1
Special functions	66	53	9	4	0
Percentages		76.2%	22.7%	0.8%	0.4%

Table 3. The performance of Mathematica on the test suite, using measures described in the text.