数式処理による入試問題への挑戦
～ロボットは東大に入れるか～

Challenge of Passing Mathematics Entrance Exam with Computer Algebra

● 岩根秀直　● 穴井宏和

あらまし

今日，ビッグデータが注目される背景には，いかにうまくデータを解析し，新たな価値を創出できるかという大きな期待がある。データを効率的に解析し，新たな知見を抽出し価値創造するために，自然言語処理，画像認識，音声認識，機械学習などの人工知能の要素技術への関心が再度高まっている。人工知能の発展に期待する知的タスクは増大する一方で，全てのタスクにおいて必要十分なビッグデータが手に入るわけではなく，スモールデータしか存在しない場合でもうまく処理するために，既存の人工知能の技術に新たなロジックを融合することが必要となる。そのような試みとして，自然言語で与えられた入試問題に対して自動解答する人工知能を開発し，東京大学の入試を突破することを目標に掲げた人工頭脳プロジェクト「ロボットは東大に入れるか」を紹介する。富士通研究所は，数学入試問題解法に対する強力なツールである数式処理，および限量記号消去法を長年研究しており，数学チームへ参画している。

本稿では，数学入試問題自動解法の概要を示した後，その技術課題について述べる。また，実際の入試問題や入試模試を用いた現状の解答システムの評価についても紹介する。

Abstract

Today's increasing interest in big data is backed by an aspiration for better ways of analyzing data to create new value. This creation of value with new knowledge gained through efficient data analysis is a source of renewed interest in component technologies of artificial intelligence (AI), such as natural language processing, image recognition and machine learning. While the intelligent tasks to advance AI continue to grow, big data is insufficient for some of them. In order to process data correctly even in cases where there is only a small volume of data available, new logic needs to be integrated into the existing AI technology. As an attempt to achieve this, there is a project to develop AI that is capable of automatically solving university entrance exam problems given in natural language, with a goal to achieve a level that exceeds the threshold required for admission to the University of Tokyo (Todai); thus the project is named "Todai Robot Project - Can a Robot Pass the University of Tokyo Entrance Exam?" Having long engaged in researching computer algebra and quantifier elimination (QE), powerful tools that can be applied to solve mathematics problems of entrance exams, Fujitsu Laboratories is taking part in the project's math team. This paper outlines an automatic solution scheme that deals with a university entrance exam in mathematics, and then it explains the technical challenges thereof. It also describes evaluations of the current system in terms of solving problems using real and mock exams.

ま　え　が　き

国立情報学研究所は，人工頭脳プロジェクト「ロボットは東大に入れるか」（以下，東ロボ）を2011年に開始した。東ロボでは，自然言語で記述された入試問題に対して自動解答する人工知能を開発し，2016年までに大学入試センター試験で高得点をマークし，2021年に東京大学入試を突破することを目標に掲げている。この開発では，これまで蓄積されてきた様々な人工知能の要素技術の精度を高めて再統合し，情報技術分野の未来価値創成につなげるとともに，人間の思考に関する包括的な理解を深めることを目指している。[(1), (2)]

東ロボでは，教科ごとのチーム体制を採り，共通する基盤技術について連携しながら研究を推進している。富士通研究所では，高度な数学問題を解くために必要となる数式処理・計算機代数を長年研究してきており，[(3)] この技術をベースに，コンピュータで数学問題を解くためのプログラム（ソルバー）を中心に貢献すべく，2012年度から東ロボの数学チームに参画している。

これまでも人工知能の分野では，自然言語で記述された数学問題に対する自動解答器の開発は行われてきた。分野を限定し，決められた手順で公式を適用して問題を解くテンプレートを利用した手法が主流であったが，入試問題の適用範囲を広げるためには複数のテンプレートの開発が必要になること，および手順に合致しない問題には対応できない問題がある。一方で，汎用的な手法は，自然言語処理による問題文の解析が困難であることと，ソルバーの計算量が増加する問題があった。しかし，自然言語処理の要素技術の発展による深い解析が実現可能になりつつあること，また数式処理技術およびコンピュータの性能の発展により，限量記号消去法が実用的になっていることで，入試問題において適用範囲の広いソルバーの開発が現実的になってきていた。しかし，実際に自然言語処理との接合を行うと，冗長な式が含まれるなどの影響で，現在の数式処理技術では現実的な時間で解けないことが確認された。

本稿では，開発中の自動解答器の概要とソルバーにおける計算量の課題解決方法について紹介する。

限量記号消去法による入試問題解法

数式処理は，入力された数式に対してコンピュータ上で代数的な記号演算を行い，数式を出力する。多くの計算で，浮動小数点数ではなく任意多倍長の整数または有理数を用い，誤差のない結果を返す。例えば，多項式の最大公約因子や因数分解などの計算ができる。数式処理計算を実現する商用の数式処理システムではMapleやMathematica，フリーではRisa/Asirなどがある。数学入試問題では誤差のない計算が要求されるため，数式処理システムの活用は有効な手段であると考えられる。数式処理技術の中でも，実閉体（RCF：Real Closed Field）上の限量記号消去（QE：Quantifier Elimination）は，入試問題解法の強力なツールの一つである。

QEは，限量記号がついた一階述語論理式を入力として，等価で限量記号のない論理式を出力するアルゴリズムである。例えば，$\exists x\,(x^2+bx+c=0)$ に対してRCF上のQE（RCF-QE）を適用すると，それと等価で限量記号がついた変数xのない論理式，つまり2次方程式$x^2+bx+c=0$を満たす実数xが存在するbとcの条件である$b^2-4c\geq 0$が得られる。

一階述語論理式は，限量記号である∀(全称記号)，∃(存在記号)，有理数係数の多項式の等式・不等式からなる原子論理式，∧(かつ)・∨(または)・¬(否定）などの論理演算子から成る。一階述語論理は広い記述能力を持ち，最適化問題などの重要な応用問題を統一的に取り扱うことが可能であるため，計算機科学や各種の理工学分野の研究者がRCF-QEを活用するようになっている。富士通研究所でもRCF-QEの研究を続けてきており，社内のものづくりにおける最適化などにも適用している。

現在公開されているQEツールは，フリーウェアでは，QEPCADとREDUCE上のREDLOGパッケージ，商用ソフトウェアではMathematicaなどに実装されている。筆者らの研究グループは，Maple上で動作するQEパッケージSyNRAC[(4)]を開発中であり，フリーで公開している。QEアルゴリズムとその応用に関する詳細は，参考文献(3)を参照されたい。

RCF-QEアルゴリズムの存在は「実数xの値（範囲）を求めよ」というタイプの問題に対して，一階のRCFの言語による問題の表現が得られれば，原理的にはxの値を全て見つけられることを意味している。つまり，QEで問題を解く場合には，ソルバーに合わせた定式化が必要なく，問題文をそのまま一階述語論理式の形に変換すれば良いため，自然言語処理との接合において都合が良い。また，RCFの理論によって表現できる問題は，入試問題においては代数および幾何のかなりの部分を含む。適用範囲が広い点からも，RCF-QEによるアプローチは非常に適したものとなっている。

具体的な例を用いて，QEによる解法を紹介する。

> tが実数全体を動くとき，xyz空間内の点（$t+2$, $t+2, t$）がつくる直線をlとする。3点O$(0, 0, 0)$，A′（2，1，0），B′（1，2，0）を通り，中心をC（a，b，c）とする球面Sが直線lと共有点をもつとき，a，b，cの満たす条件を求めよ。
>
> （北海道大学入学試験問題2011理系[3]（2））

球面Sの半径をrとすると，問題を表す一階述語論理式は以下のようになる。

$$\exists r \exists t \left(\begin{array}{c} (0-a)^2+(0-b)^2+(0-c)^2=r^2 \wedge \\ (2-a)^2+(1-b)^2+(0-c)^2=r^2 \wedge \\ (1-a)^2+(2-b)^2+(0-c)^2=r^2 \wedge \\ (t+2-a)^2+(t+2-b)^2+(t-c)^2=r^2 \end{array} \right)$$

上記の式は，球面が点O，A′，B′および（$t+2$，$t+2$，t）を通る実数tとrが存在することを表している。この一階述語論理式にRCF-QEを適用すると，a，b，cが満たすべき条件を表す以下の論理式が数秒で得られる。

$$6a=5 \wedge 6b=5 \wedge (3c \leq 1 \vee 13 \leq 3c)$$

このように，入試問題を一階述語論理式で表現できれば，QEにより正確に解くことができる。

RCF-QEは上記のように非常に強力なツールであるが，その計算量の大きさが問題である。これまでに見つかっている最も効率の良いアルゴリズムでも，変数の数に対して2重指数で，現実的な時間で解けるのは5変数程度が限界である。ただし，数式処理研究者の調査により，入試問題に限れば，多くの問題がRCF-QEで解ける規模であることが確認されている。

コンピュータが数学入試問題を解く

コンピュータが数学入試問題を解くということは，人間にとって理解しやすい自然言語や数式で表現された問題文をコンピュータが実行可能な形式に変換し，ソルバーで問題を解くということである。

現在開発中の自動解答器「東ロボくん」の基本的な処理は，**図-1**に示すように大きく三つのステップから構成される。[(5), (6)]

（1）言語解析・意味合成

自然言語や数式で表現された問題文を理解する。具体的には，問題のテキストを曖昧性のない形式的な意味表現へ翻訳する。

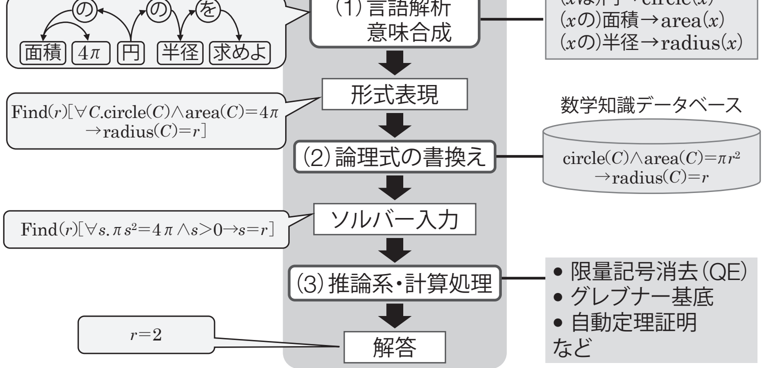

図-1　東ロボくんの数学問題求解フロー

(2) 論理式の書換え

数学知識データベースを利用して，問題の意味表現を実際に自動演繹が可能な形，つまりソルバーが処理できる形式に変換する。

(3) 推論系・計算処理

数式処理・自動演繹ソルバーで答えを求める。

自動解答するための最初のステップでは，自然言語処理技術により，自然言語で書かれた問題を論理による表現へと翻訳する。翻訳先の言語として，その表現力の広さからZermelo-Fraenkel（ZF）集合論の大幅な保存拡大を選択した。一方，ZF集合論の表現のままでは実用的な推論技術に直接接続できない欠点を補うため，同値性を保ついくつかの書換え規則を繰り返し適用することによって，自動推論が可能な問題表現を探すように設計している。ZF集合論による問題の表現に対し，書換えが成功したときは，得られた問題表現に応じた推論器が呼び出される。そこで，主に利用している推論器の一つが前章で述べたRCF-QEである。

自然言語処理との接合における課題

自動解答器「東ロボくん」が自然言語を処理して，ソルバーの入力を構築する流れについて紹介する。

富士通研究所は，主にソルバー後半の計算処理の開発に携わっており，自然言語処理による最初のステップ「言語解析・意味合成」の説明は割愛する。詳細は参考文献（4），（5）を参照されたい。前節の北海道大学の入試問題に対する自然言語処理の結果として，おおよそ以下のようなZFの式が得られる。

$$\mathrm{Find}(a,b,c)\left[\begin{array}{c}\exists l\exists u\exists v\exists S\exists R(\\ l=\mathrm{line}(u,v)\wedge\\ \forall p(p\in l\leftrightarrow\exists t(p=(t+2,t+2,t)))\wedge\\ S=\mathrm{sphere}((a,b,c),R)\wedge\\ (0,0,0)\in S\wedge\\ (2,1,0)\in S\wedge\\ (1,2,0)\in S\wedge\\ \exists q(\mathrm{intersect}(S,l,q)))\end{array}\right]$$

ここでFind(a，b，c)[φ(a，b，c)] は，自由変数としてa，b，cを含む論理式 φ(a，b，c）が定める値の組（a，b，c）の範囲を求めよ，という指令を意味する。

また，関数line(u，v）は定点vを通り方向ベクトルuを持つ直線，関数sphere((a，b，c)，R）は中心が（a，b，c）で半径がRである球面を表し，述語intersect(S，l，q）は，球面Sと直線lが点qを共有することを表す。

2番目のステップ「論理式の書換え」では，数式処理ソルバーを実行するための入力を構築する。この指令を実行して（a，b，c）の範囲を求めるためには，Findの引数である論理式φ(a，b，c）から限量記号を消去し，a，b，cの関係を表す簡単化された式を得れば良い。QEを利用してこれを行うためには，まずφ(a，b，c）と等価で，実多項式の間の等式・不等式のみからなる一階の論理式を得る必要がある。

このためには，以下のような処理を行う。

(1) line，intersectなどの関数・述語記号を消去する。

(2) 上記の例のSやu，（0，0，0）など，実数以外のオブジェクトを表す変数や項を消去する。

(3) 上記の例では現れていないが，自然言語処理の結果として導出されるZFの式には，一般的な関数や集合を表すために高階の項（ラムダ式）や変数が用いられる場合がある。これらを消去し，一階の式を得る。

(4) 同様に上記の例にはないが，微分や積分といった操作，囲まれた領域の面積など複雑な関数・述語は，数式処理アルゴリズムを用いてそれらを評価することで消去する。

上記の（1），（2）については，式の操作による等価変換に加えて，述語・関数の関係の定義，例えば，

$$\begin{array}{c}\forall a\forall b\forall c\forall R\forall u_x\forall u_y\forall u_z\forall v_x\forall v_y\forall v_z\forall q_x\forall q_y\forall q_z(\\ \mathrm{intersect}(\mathrm{sphere}((a,b,c),R),\\ \mathrm{line}((u_x,u_y,u_z),(v_x,v_y,v_z)),(q_x,q_y,q_z))\\ \leftrightarrow\left(\begin{array}{c}(q_x-a)^2+(q_y-b)^2+(q_z-c)^2=R^2\wedge\\ \exists t(tu_x+v_x=q_x\wedge tu_y+v_y=q_y\wedge tu_z+v_z=q_z)\end{array}\right))\end{array}$$

による書換えを行う。現在，数学知識データベースとして，数学入試問題解法に必要な述語・関数を3000個程度作成している。上記の操作をZFの式に適用すると，**図-2**のような問題文と等価な一階述語論理式が得られる。

最後のステップ「推論系・計算処理」では，構築した一階述語論理式に対しQEを実行する。しかし，前節で構築した論理式に比べ，東ロボくん

$$
\begin{aligned}
&\exists a\exists b\exists c\exists u_x\exists u_y\exists u_z(((\neg(u_x=0))\vee(\neg(u_y=0))\vee(\neg(u_z=0)))\wedge(\exists v_x\exists v_y\exists v_z(\\
&(\exists R((\exists t((tu_x+v_x-a)^2+(tu_y+v_y-b)^2+(tu_z+v_z-c)^2=R^2))\wedge(0<R)\wedge(a^2+b^2+c^2=R^2)\\
&\wedge((1-a)^2+(2-b)^2+c^2=R^2)\wedge((2-a)^2+(1-b)^2+c^2=R^2)))\wedge(\exists v_{lx}\exists v_{ly}((\exists v_{lz}((0=u_y(v_z-v_{lz})-u_z(v_y-v_{ly}))\wedge\\
&(0=u_z(v_x-v_{lx})-u_x(v_z-v_{lz}))))\wedge(0=u_x(v_y-v_{ly})-u_y(v_x-v_{lx}))))\wedge(\forall p_x\forall p_y\forall p_z(((p_x=p_z+2)\\
&\wedge(p_y=p_z+2))\vee\quad(\forall t_2((\neg(p_x=t_2u_x+v_x))\vee(\neg(p_y=t_2u_y+v_y))\vee(\neg(p_z=t_2u_z+v_z))))))\\
&\wedge(\forall p_t(\exists t_3((p_t=t_3u_z+v_z)\wedge(p_t+2=t_3u_x+v_x)\wedge(p_t+2=t_3u_y+v_y))))\\
&))\wedge(\exists u_{lx}\exists u_{ly}((\exists u_{lz}(((\neg(u_{lx}=0))\vee(\neg(u_{ly}=0))\vee(\neg(u_{lz}=0)))\\
&\wedge(0=u_yu_{lz}-u_zu_{ly})\wedge(0=u_zu_{lx}-u_xu_{lz})))\wedge(0=u_xu_{ly}-u_yu_{lx}))))
\end{aligned}
$$

図-2　東ロボくんが構築した論理式

が構築した論理式の規模は非常に大きくなっている（**表-1**）。自然言語処理によって導かれる問題文の論理式は，人による立式に比べ冗長な表現になるが，多くのQEツールでは，あらかじめ冗長な条件は人が取り除くことが期待されている。それは，QEの計算量が大きく，少しでも簡単な表現でなければ解けないことが多いことと，人には不要と簡単に分かる式でも，それを取り除くのは非常に大きな計算量を要する処理であるからである。実際，東ロボくんが構築した図-2の論理式は，Mathematicaなど既存のQEツールでは1時間以上かけても解くことができなかった。そのため，より簡単な式への立式方法やソルバーの計算速度効率化の工夫が大きな課題の一つである。次章で紹介する工夫により，現在では上記の入試問題に対して，東ロボくんが構築した冗長な論理式でも解けるようになっている。

表-1　人が構築した論理式と，東ロボくんが構築した論理式の比較（北大2011理［3］（2））

	人間	東ロボくん
変数の数	5	22
限量記号の数	2	19
原子論理式の数	4	22

課　題　解　決

数学入試問題と東ロボくんが構築する冗長な論理式は，次のような共通した特徴を持つ。

- 語間の等価性を表す式が含まれるため，等式の条件が多い
- 本質的には解ける規模の問題である
- 最終的な解答は必ずシンプルである

前に述べたように，入試問題はRCF-QEで解ける規模の問題であり，本来QEツールとしてはそのような問題は冗長な表現であっても解くべき問題である。東ロボくんが構築する冗長な論理式に対応するため，上記の性質を利用した以下の改善を行った。[(7)–(9)]

（1）与えられた問題を部分問題に分割
（2）分割された問題を解く順序を決定
（3）定式化のヒューリスティクスを利用した論理式の簡単化
（4）途中で得られた結果を利用した論理式の簡単化

（1）では，分配法則などを利用して部分問題に分割する。これは，消去する変数の次数が線形の場合など特別な入力にのみ適用可能で，高速な専用QE手法の適用の可能性を増やすとともに，問題の規模を小さくすることで計算時間の削減を実現する。ただし，分割数が指数的に増えてしまうことがあるため，しきい値を設定した上で，計算時間の削減の見込みの高い部分から分割するようにしている。

（2）では，入力の一階述語論理式に含まれる原子論理式の数や変数の数などを利用して，問題の解きやすさを表す評価値を設定し，分割した問題を整列する。

（3）では，計算量を削減するために論理式を簡単化する。例えば，先の北大の入試問題において，半径をrと置いたが，この問題では半径そのものを利用していない。したがって，半径の2乗を変数に置くことで，次数の低い問題に帰着できる。ほかに，幾何の問題では，ある点を原点においても一般性を失わないことを利用して変数の数を削減できる場合がある。このように，入力と等価で簡単な問題に帰着させる変換を行う。

（4）では，（2）により計算が容易な部分から解いた結果を利用して簡単化することで，規模の大

表-2　旧帝大の入試問題に対する結果
（入力：人手で作成したZFの式）

	完答/RCF/全体	完答率（%）RCF	完答率（%）全体
北海道	18/32/72	56.3	25.0
東北	24/42/80	57.1	30.0
東京	15/29/80	51.7	18.8
名古屋	15/28/65	53.6	23.1
京都	29/38/88	76.3	33.0
大阪	21/24/64	87.5	32.8
九州	18/35/96	51.4	18.8
合計	140/228/545	61.4	25.7

きな問題を解ける可能性を増やす。

そのほか，適用範囲を増やすため，曲線の囲まれた部分の面積を求める関数の実装や，RCF-QEの問題に帰着できる三角関数の問題への対応などを行っている。

計算実験結果

1999年から2013年までの理系と文系の奇数年度に行われた旧帝大の数学入試問題について，ZFの式を人手で構築し，現在のシステムの評価を行った（**表-2**）。問題全体の中で，RCF-QEに帰着できたもの，更に完答できたものをカウントし，問題全体に対する完答率とRCF-QEに帰着できた問題に対する完答率を比較している。

大学ごとにばらつきはあるが，全体の約1/3強がRCF-QEの問題に帰着できることが分かった。そのうち6割が現在のシステムで解けるようになっており，平均して大問が1〜2問解けている。もちろん，これらの傾向は年度ごとにもばらつきがあり，例えば京大2011年の理系は6問中5問がRCFに帰着でき，そのうち4問に完答でき（以下，これを4/5/6と表す），文系は3/4/4であった。同様に，東大2011年の理系は3/3/6，文系は2/4/4と現在でも合格に十分な成果を上げている。一方で，東大2013年の理系は0/3/6，文系は0/2/4と完答がなく，まだシステムとして不十分な状況である。

現時点での性能評価を行うことを目的として，2013年11月と2014年10月の2回，代々木ゼミナールの模試に挑戦した。ここでは，自然言語で書かれた問題文を入力としている。数学チームでは，全国センター模試（数IA，数IIB）と東大入試プレ（文系，理系）に取り組んだ。問題テキストの言語処理の一部で人の介入を許したが，東大入試プレでは2013年と2014年に偏差値約60，55をそれぞれ獲得した。一方で，センター模試では偏差値が50弱であった（**表-3**）。

表-3　代々木ゼミナール模試受験結果（入力：アノテーション付の自然言語で記述された問題文）

	得点			偏差値	
	2013	2014	配点	2013	2014
東大理系	40	36	120	61.2	55.7
東大文系	40	32	80	59.4	54.1
センター IA	57	40	100	51.9	46.9
センター IIB	41	55	100	47.2	47.2

現在のシステムで解ける問題は，QEが可能なRCFで表現できるタイプの問題にほぼ限られている。このため，センター試験において東大合格に要求される満点に近い点数を取るには，RCF-QEに基づく解法以外の対応が必須である。

人の介在（問題文に対する言語アノテーションを意味合成に利用，および不足していた辞書項目の追加）を許したことで，この結果は最終的なEnd-to-Endの自動解答システムの性能に対する現時点での上限見積もりと考えるべき結果ではある。しかし，東大模試問題に対する結果では，受験者全体の平均点を既に上回っている点は特筆すべきと考える。

む　す　び

数式処理システムを含む多くの数学ソフトウェアでは，同じ問題を解く場合でも，実装した人や中のアルゴリズムに精通した人が利用しなければ解けないことがある。この事実は，アルゴリズムには興味はなく，単に問題を解きたいだけのユーザーの利用に対する妨げの一つになっている。機械的に構築された冗長な式に対しても効率的に解けるようになることは，QEおよび数学ソフトウェアの普及において重要な課題の一つだと考えている。

今後，国立情報学研究所と共同で，意味解析における曖昧性解消の高度化，QE手続きの更なる効率化，決定手続きが存在しないタイプの問題に対する発見的解法の蓄積や，初等関数を含む問題の

うち特定のタイプのものに対する系統的解法の実現などに取り組み，自動解法の高度化を図っていく予定である。東ロボくんが東大に合格できるまでの技術が確立されれば，文章を理解し，要約するなど多くのホワイトカラーによる知的作業を機械に置き換えられるようになると期待される。

参考文献

(1) N.H. Arai et al.：Mathematics by Machine. Proceedings of 39th International Symposium on Symbolic and Algebraic Computation，p.1-8，2014.

(2) 新井紀子：ロボットは東大に入れるか．イースト・プレス，2014.

(3) 穴井宏和ほか：QEの計算アルゴリズムとその応用—数式処理による最適化．東京大学出版会，2011.

(4) 富士通研究所：SyNRAC. http://jp.fujitsu.com/group/labs/techinfo/freeware/synrac/

(5) T. Matsuzaki et al.：The most uncreative examinee：a first step toward wide coverage natural language math problem solving. Proceedings of 28th Conference on Artificial Intelligence，p.1098-1104，2014.

(6) 松崎拓也ほか：大学入試過去問による数学問題解答システムの評価と課題分析．人工知能学会全国大会，2014.

(7) H. Iwane et al.：Automated Natural Language Geometry Math Problem Solving by Real Quantifier Elimination. Proceedings of the 10th International Workshop on Automated Deduction，p.75-84，2014.

(8) 岩根秀直：Formula Simplification for Real Quantifier Elimination．京都大学数理解析研究所講究録 1927，p.77-88，2014.

(9) 岩根秀直ほか：数式処理による入試数学問題の解法と言語処理との接合における課題．人工知能学会全国大会，2013.

著者紹介

岩根秀直（いわね　ひでなお）
知識情報処理研究所
Big Intelligence プロジェクト 所属
現在，数式処理，人工知能の理論・応用に関する研究・開発に従事。

穴井宏和（あない　ひろかず）
知識情報処理研究所
Big Intelligence プロジェクト 所属
現在，数式処理，最適化，人工知能の理論・応用に関する研究・開発に従事。