# Mathematics by Machine

[Invited Talk Paper]

Noriko H. Arai
National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, Japan
arai@nii.ac.jp

Takuya Matsuzaki
National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, Japan
takuya-matsuzaki@nii.ac.jp

Hidenao Iwane
National Institute of Informatics /
Fujitsu Laboratories Ltd
2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, Japan
iwane@jp.fujitsu.com

Hirokazu Anai
Fujitsu Laboratories Ltd / Kyushu University
4-1-1 Kamikodanaka Nakahara-ku,
Kawasaki, Japan
anai@jp.fujitsu.com

## ABSTRACT

When David Hilbert started so called "Hilbert's program" (formalization of mathematics) in the early 20th century to give a solid foundation to mathematics, he unintentionally introduced the possibility of automatization of mathematics. Theoretically, the possibility was denied by Gödel's incompleteness theorem. However, an interesting issue remains: is "mundane mathematics" automatizable? We are developing a system that solves a wide range of math problems written in natural language, as a part of the Todai Robot Project, an AI challenge to pass the university entrance examination. We give an overview and report on the progress of our project, and the theoretical and methodological difficulties to be overcome.

## Categories and Subject Descriptors

F.4.1 [**Mathematical Logic and Formal Languages**]: Mathematical Logic; G.4 [**Mathematics of computing**]: Mathematical Software; I.1.2 [**Symbolic and Algebraic Manipulation**]: Algorithms; I.2.7 [**Artificial Intelligence**]: Natural Language Processing

## General Terms

Algorithm, Experimentation

## Keywords

mathematical logic, quantifier elimination, natural language processing

## 1. INTRODUCTION

When at the beginning of the twentieth century Hilbert and others attempted the formalization of mathematics, they defined it as a game, in which, starting from a number of initial assumptions, or axioms, one derives other propositions using only the rules of inference. Many mathematicians have felt a resistance to this definition, but when we *retrospectively* examine the mathematics actually published in the area of so-called pure mathematics, it comes as a shock to see that the theorems and their proofs can actually be formalized as the mathematical game.

The objective of formalism was to avoid the crisis to which mathematics had been brought by the succession of paradoxes discovered in the late nineteenth and early twentieth centuries, and it was not that the formalists thought of mathematics itself as a game of symbol processing. Mathematicians carry on their mathematical activities, thinking at the level of 'semantics,' such as the properties of numbers or the shape of geometric objects, and making intuitive deductions that are often described as "revelation." Hilbert and his associates of course knew it. But as a result of the formalist approach that they introduced, an unsettling possibility came to light. This was the possibility that mathematics could be carried out without 'semantics,' by means of a machine that operates on axioms and inference rules, but producing the same mathematical outcomes as human mathematicians.

The work *Principia Mathematica*, written over the years from 1910 to 1913 by Whitehead and Russell showed how the theorems of set theory and elementary real number theory could be expressed formally [14], which suddenly brought this possibility closer to reality. At this stage, however, the possibility of mechanical execution of mathematics was merely a matter for philosophical debate. In 1936, by quite different approaches, both Church [2] and Turing [13] produced formalizations of "What it means to compute," and with the development in the 1940s of practical computers based on this model, the possibility of "mathematics by machine" got into the realm of engineering.

Our research team has been working since 2011 on the mechanical solving of university entrance exam problems [7][8]. This paper gives a survey of this work, and also considers the meaning of "Mathematics by machine."


*ISSAC '14,* July 23 - 25, 2014, Kobe, Japan

ACM 978-1-4503-2501-1/14/07 ...$15.00.
http://dx.doi.org/10.1145/2608628.2627488 .

$$\left(a^2 = ba + c \land \forall\epsilon > 0\exists\delta > 0\forall h\left(0 < |h| < \delta \rightarrow \left|\frac{(a+h)^2 - b(a+h) - c}{h}\right| < \epsilon\right)\right)$$

**Figure 1: "$y = ax + b$ is tangent to $y = x^2$ at $(a, a^2)$" in epsilon-delta notation**

## 2. WHAT DOES IT MEAN TO SOLVE A MATHEMATICS PROBLEM?

We use the expression "Solve a mathematics problem" quite naturally. It does not only mean to find a proof of a theorem, as in "Wiles has solved Fermat's Last Theorem," but can also refer to mathematical activities other than theorem proving, as in "Did you solve all yesterday's homework problems?" But what does it mean to "solve a mathematics problem"? What has to happen for a problem to be "solved"?

Consider a problem such as the following:

PROBLEM 2.1. *Let $C$ be a parabola in the xy-plane with the equation $y = x^2$. Find the equation of the tangent to $C$ at the point $(a, a^2)$ on $C$.*

When we interpret the wording of the problem as plain language, we can broadly summarize it as saying: "the problem is to *calculate* the *object* which is the equation of the tangent at $(a, a^2)$ to another object, $y = x^2$, named $C$."

An *object* is something to which we can direct an action. We could also say that an object is anything that we can refer to as "this" or "that." For example, "the apple I bought the other day" and "the intersection of the straight line $l$ with the parabola $C$" are objects. But what about "apple" or "straight line"? A statement such like "Apples are red" suggests that the words "apple" and "red" do not refer to specific objects. If we add wording to look more deeply at this statement we have: "If the object $a$ is an apple, then $a$ has the property of being red." From a normal mathematical perspective, $a$ will be seen as an object, but "apple" and "red" will be seen not as objects but as concepts. It is not, however, obvious where to draw the line between objects and concepts. For example, consider "the intersection of the straight line $l$ and the parabola $C$"; when there is only one intersection point, it is surely an object. But is it still an object if there is no intersection? What happens if there is more than one intersection? Would it be more appropriate to handle "intersection" as a concept (unary relation)? Is even the straight line $l$ an object, or should this too be a concept? There are no absolute criteria to determine what is an object and what is a concept; the same thing appears to be an object in one context but a concept in another.

There is an approach to overcome this philosophical problem: axiomatic set theory.

In *Zermelo-Fraenkel axiomatic set theory (ZF)*, the object is a set, and in principle all mathematics can be expressed within ZF.[1] This means that in ZF there is no distinction between objects and concepts, and an object is simultaneously a concept. In the problem cited above, $C$, $y = x^2$, and $(a, a^2)$ are all terms of ZF. Further, "the equation of the tangent at the point $(a, a^2)$ on the parabola $y = x^2$" is also a term. The solution to this problem, that is, "$y = 2ax - a^2$" is itself a term, too.

It thus appears that the word, "Find," in the problem statement actually means to rewrite the ZF term "the equation of the tangent at the point $(a, a^2)$ on the parabola $y = x^2$" into its equivalent form "$y = 2ax - a^2$".

When should we stop this rewriting process? For example, suppose someone provides as an answer: "The equation of the tangent at the point $(a, a^2)$ on the parabola $y = x^2$." This is trivially equivalent to the ZF term representing the problem, but would not be accepted as a correct answer. Given that there are an unlimited number of terms equivalent to a direct translation of the problem, what is the condition for an equivalent term to be a "solution" to the problem?

We can find some hints about it by looking at the propositions that are actually regarded as "solutions" to some problems by a human. To take the example of a university entrance exam, except in proof problems, the answers are in the form of equations and inequalities, such as $y = 2ax - a^2$ or $a_1 + \cdots + a_n > 0$. For a difficult problem the answer may be separated into cases, but these can also be expressed with combinations of propositional connectives, such like $(x < 0 \rightarrow a = 3) \land (x \geq 0 \rightarrow a = 5)$. Note that there is no *quantifier* $(\forall, \exists)$ in the answers. Moreover, this vocabulary is largely restricted to the four arithmetic operators $(+, -, \times, \div)$ and the basic binary relations $(=, <)$ together with symbolic use of elementary functions such as trigonometrical and exponential functions. If we refer to this vocabulary as the vocabulary of elementary mathematics, then it would not be far wrong to say that to "solve" the kind of mathematics problem we are familiar with, such as those in a university entrance exam, is first to convert the problem statement into an equivalent proposition in the vocabulary of elementary mathematics and then to eliminate quantifiers from it.

Let's think about the problem, "Find the equation for the tangent to $y = x^2$ at $(a, a^2)$." What formula of elementary mathematics can we rewrite this to?

First we notice that this problem is not a proposition. A proposition must be declarative, with the form "Something has some property," while the sentence above is imperative. Similarly, a question in interrogative form, such as "What relation holds between X and Y?" also cannot be expressed as a proposition. Thus a question in the form of "Find ..." cannot be (directly) represented as a proposition, which has been a central figure in formalism.

Let $y = bx + c$ be the equation of the tangent to be found. We can then rewrite the problem as follows:

PROBLEM 2.2. *Let $C$ be the parabola $y = x^2$ in the xy-plane. The equation of the tangent to $C$ at the point $(a, a^2)$ on $C$ is $y = bx + c$.*

This statement is at last in the form of a proposition. We can further express the notion of "tangent" using epsilon-delta notation and translate the problem statement simplis-

[1]This really is "in principle," and attempting to embed ordinary mathematics within ZF would be a nightmare.

$$
\dfrac{\text{When}}{\mathsf{S/S/S} : \lambda P.\lambda Q.P \to Q} \quad \dfrac{\dfrac{\text{the centers of } C_1 \text{ and } C_2}{\mathsf{S/(S\backslash NP)} : \lambda P.\exists x \exists x_1 \exists x_2 (x = [x_1, x_2] \wedge x_1 = \text{center\_of}(C_1) \wedge x_2 = \text{center\_of}(C_2) \wedge Px)} \quad \dfrac{\text{coincide}}{\mathsf{S\backslash NP} : \lambda x.\text{coincide}(x)}}{\mathsf{S} : \exists x \exists x_1 \exists x_2 (x = [x_1, x_2] \wedge x_1 = \text{center\_of}(C_1) \wedge x_2 = \text{center\_of}(C_2) \wedge \text{coincide}(x))} >
$$

$$
> \quad \mathsf{S/S} : \lambda Q.(\exists x \exists x_1 \exists x_2 (x = [x_1, x_2] \wedge x_1 = \text{center\_of}(C_1) \wedge x_2 = \text{center\_of}(C_2) \wedge \text{coincide}(x))) \to Q
$$

**Figure 2: A part of CCG derivation tree**

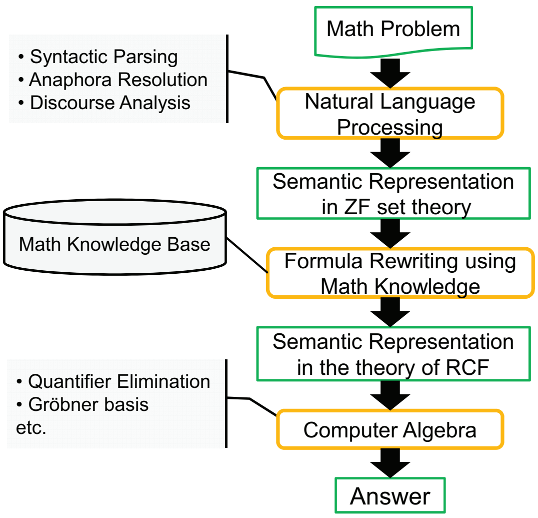


**Figure 3: Process of solving a mathematics problem**

tically to an expression such as one in Fig. 1.

This proposition includes several quantifiers. If we can transform this expression while preserving equivalence until there is no quantifier at all, then it will be a "solution."

To summarize, our hypothesis is that to solve a mathematics problem is to translate the problem statement into a ZF proposition, further translate it into a proposition of elementary mathematics, and then eliminate quantifiers. The flow diagram (Fig. 3) illustrates this.

In the flow diagram, the step labeled "Natural language processing" translates the problem statement in natural language to a ZF proposition. The translation is done by firstly analyzing the syntactic structure of the sentences. The ZF proposition is then composed along the syntactic structures of the sentences, by combining the semantic representation of words that are defined in the semantic lexicon. Fig. 2 shows a part of the semantic composition process for a phrase "When the centers of $C_1$ and $C_2$ coincide" (the derivation of "the centers of $C_1$ and $C_2$" is omitted), in which the syntactic structure is analyzed based on a grammar formalism called Combinatory Categorial Grammar (CCG) [10] and the semantic representation is composed through lambda calculus.

The result of this stage is no more than a direct translation of the problem statement, which gives almost no clue for solving the problem. To actually solve it, we further need to paraphrase the ZF proposition to another proposition in a local theory such as the theory of real numbers or the theory of natural numbers, in which various mathematical tools have been developed. This is the step labeled "Formula rewriting using math knowledge." In that step, a ZF proposition is rewritten using the definitions of the predicates and functions (e.g., center_of$(\cdot)$ and coincide$([\cdot,\cdot])$ in Fig. 2) stored in a database of defining axioms ("math knowledge base"), in addition to generic equivalence-preserving transformations of logical formulas.

Only when the rewriting is successful can the work of mathematics as we normally consider it begin; we can finally utilize the traditional technologies for automated math, such as the theory and the practice of computer algebra and automatic theorem proving.

The problem of " mathematics by machine" can thus be restated as the question of whether these processes can be fully automated, or, in other words, reduced to algorithms and executed in a realistic amount of time.

## 3. CAN A MATHEMATICS PROBLEM BE READ UNAMBIGUOUSLY?

The task of translating the problem statement to a proposition of ZF falls on the area of artificial intelligence research known as *natural language processing*. We tend to think of "translation" as a rather mechanical process that can be done by using the rules of syntax and a dictionary. The following examples of translation from Japanese to English however illustrate that the problem is not so simple. [2]

| | | | |
|---|---|---|---|
| *Watashi-wa* | *Kyoto-to* | *Tokyo-ni* | *itta.* |
| I-TOPIC | Kyoto-WITH | Tokyo-TO | went. |

'I went to Kyoto and Tokyo.'

| | | | |
|---|---|---|---|
| *Watashi-wa* | *Kyota-to* | *Tokyo-ni* | *itta.* |
| I-TOPIC | Kyota-WITH | Tokyo-TO | went. |

'I went to Tokyo with Kyota.'

If we look at the original sentences as character strings, the only difference is the last letter of the names, Kyoto and Kyota. However, this greatly changes the meaning: in the first example Kyoto and Tokyo are seen as a conjoined noun phrase, while in the second we read Kyota as a person accompanying the speaker. Thus the different attachment of the particle *to* completely changes the meaning, as shown by the English translations.

When there is more than one possible interpretation, in natural language processing this is referred to as *ambiguity*. According to a standard Japanese dictionary the particle *to* has at least six different usages. This suggests that we cannot obtain a correct translation only using a word dictionary together with the rules of grammar.

The current state-of-art machine translation, Google Translate for example, mistranslates the second sentence to "I

[2]Kyoto and Tokyo are cities in Japan; Kyota is a fairly common first name. The hyphenated words are postpositions indicating the grammatical or semantic relation within the sentence. The particle *to* can be attached to the noun phrase to the right, forming a conjunction (A-and-B), or can attach to the final verb, indicating that its noun phrase accompanies the topic of the sentence.

went to Kyota and Tokyo," where the particle *to* is misinterpreted as conjoining noun phrases. This example shows that, in principle, there is no way to obtain an accurate translation using only local information such as the words and phrases of the text. In addition to the ambiguity among the usages of the particles such as *to*, demonstrative pronouns (such as *sono* in Japanese) often have more than one possible referents and there are many polysemous words and phrases, for which we need to select a correct one from their possible meanings. We thus encounter such ambiguity everywhere in attempting to process language automatically.

It might be thought that ambiguity is present because the meaning of natural speech is vague, whereas a mathematical statement would avoid such problems. We might hope so, but sadly our hopes are betrayed.

- There is a real number $y$ for an arbitrary positive real number $x$ such that $y = \sqrt{x}$ holds.

The natural reading of this statement is surely:

$$\forall x(x > 0 \to \exists y(y = \sqrt{x})).$$

However, there are other possible readings. For example:

$$\exists y(\forall x(x > 0 \to y = \sqrt{x})).$$

Everyone would reject such a reading, since this proposition is false. However, we have to be careful here. We may talk about the truth or falsity of a proposition (written in natural language) only after determining its reading; hence, we cannot assume the truth/falsity of a proposition for determining its reading. Otherwise, we have to abandon the traditional concept of truth and the proof.

There has long been awareness of the ambiguities associated with the scopes of quantifiers and negations, and various strategies for dealing with this have been proposed. Unfortunately, this is not the only source of ambiguity. Our investigations alone have found more than thirty ways of using the particle *to*. The idea that being a statement of mathematics would be sufficient to avoid ambiguity turns out to have been a false hope.

In the system for solving mathematics problems that we are developing, we have adopted the somewhat cheeky approach of assembling all possible readings, then seeking for one that provides a plausible solution, and regarding this as the correct reading. Of course we acknowledge that such an indiscriminate approach can never guarantee the correctness of the reading. It would however be difficult to logically explain the difference between the cheeky approach and the rejection by a human of the readings that make a proposition false.

What is perhaps more serious than the "ethical" problem is that this method of exhaustive listing of the possible readings runs into the problem of exponentially increasing degrees of ambiguity.

## 4. WHAT DOES THIS PROPOSITION TALK ABOUT?

When we say this is a project to solve university entrance exam problems automatically, many people, including mathematicians look quizzically, saying "Isn't automatic solving of entrance exam problems rather easy?" This somewhat naive response misses an important point (even considering the lack of appreciation of the difficulties of natural language processing). This is the question of whether it is even possible to determine from the text of a proposition just what it talks about. In other words, the question is whether it is possible by looking at the problem statement to determine the formal system required to solve the problem.

The following is a quotation from the Introduction section of *Mathematical Logic* by Toshiyasu Arai [1]:

> In principle, there is no mathematics which cannot be expressed as a formally provable formula in first order predicate logic — this would be "(a part of) the Definition" of mathematics. Even stated conservatively, we can say from experience that all mathematics can ultimately be formalized in first order predicate logic.

Here the term "first order predicate logic" refers to the first order predicate logic of ZF set theory. And "ultimately" means that substantial paraphrasing may be required. However, when we look again at ZF we find it an extremely restricted theory. The naive set theory allows a set to be constructed with a description such as "All elements having such-and-such a property," but ZF doesn't. This is because such indiscriminate constructions would immediately run into Russell's paradox. Only something obtained from existing sets in a limited manner can be a set. By the very nature of the construction, at most countably infinite different properties can exist. This leads to a curious problem: for example, an uncountable set such as the reals is itself allowed by ZF (since ZF has the axiom of infinity and the axiom of the empty set), but a description of the distinguishing properties of many individual reals is in principle impossible.

Whether we like it or not, we have no other means than language (or symbols) to share our thought with others. Even if our mathematical intuition tells us that a proposition is true, this must retrospectively be proved as "a series of symbols." The result should thus remain within the range of ZF. Hence, whether we call it mathematical intuition or telepathy, the limit of symbolic communication is the limit to mathematics – this can be said to be the ideology of formalism that rescued mathematics from a crisis.

Now even if we assume that we have made translations into the first order predicate logic of ZF, there is no guarantee, anywhere, that this can be written in the vocabulary of some local axiom sets of the mathematics, such as the theory of elementary geometry. It is a classical result that elementary geometry can be embedded in the theory of real closed field (RCF). However, given "a problem of generic geometric objects," it is immediately unclear what vocabulary and axioms we can assume.

Let us consider a proposition such as the following:

THEOREM 4.1. *If a closed, non-self-intersecting loop lies in a plane, then the loop divides the plane into two regions.*

This is an informal presentation of the famous Jordan curve theorem. Bolzano was the first to recognize the difficulty of this problem, in the 19th century, and it took almost a century to finally determine the language and axioms required to express this problem. This is exactly how long it took to solve this problem.

Today it is possible to create a semi-automatic mechanical proof of the Jordan curve theorem [5]. However, this is ONLY after determining for the problem: "What does this proposition talk about?" To provide a mechanical proof after the Jordan curve theorem has already been formalized, and to determine beforehand a formal expression to allow the problem statement to be solved, are completely different problems.

Since a deep problem such as the Jordan curve theorem is not likely to appear in a university entrance exam, the reader might be tempted to assume that for an ordinary problem there would be a well-known (formal) theory in which it could be embedded. But what can we make of the following examples?

1. Let $O$ be a circle of radius 1 centered on the origin. Given points $A$ and $B$ on the circumference of $O$, find the point on the $x$-axis equidistant from $A$ and $B$.
2. Let $O$ be a circle of radius 1 centered on the origin. Find a point $A$ on the $x$-axis such that the distance from point $A$ to the origin is equal to the length of the circumference of $O$.

At the word level, the vocabulary of these two examples is almost the same. However, the first statement can be expressed in the language of RCF, while the second cannot, because the phrase "length of the circumference" is not representable with the vocabulary of RCF. In other words, the vocabulary set is not sufficient to automatically determine the formal system in which we shall think about the problem.

For the second example, adding $\pi$ as an undefined constant to the vocabulary of the RCF to create an "algebraic extension," RCF[$\pi$], of RCF allows the problem to be solved on the face of it. Even so, it is not possible to prove the truth or falsity of the following proposition:

$$3 < \pi < 4.$$

Adding another axiom to RCF[$\pi$] would make this problem solvable.[3] However many axioms we add, in an extension of first-order RCF we will never be able to obtain an expression for the precise position of $\pi$ on a number line since $\pi$ is a transcendental number.

Additionally, when we consider the meaning of expressions such as, "the area enclosed by two curves" or "dividing a disk into parts and combining them so as not to overlap" we have a very deep problem. The idea that ordinary mathematics can be naturally embedded to some (natural and well-known) formal theory is no more than an illusion.

Here we show our working definition of "the area enclosed by a set of curves." We first define regions enclosed by curves.

Definition 4.1. *A* region *is defined as an open nonempty connected set of* $\mathbb{R}^k$.

Definition 4.2. *Let $F$ be a finite set of $k$-variate functions and $c$ be a region in $\mathbb{R}^k$. If the sign of each function in $F$ does not change over $c$, $c$ is* $F$-invariant. *A partition $\mathcal{D}$ of $\mathbb{R}^k$ is an* $F$-invariant decomposition *if each element of $\mathcal{D}$ is a region and the region is $F$-invariant.*

[3] A brute force solution, of course, is to add the proposition itself as an axiom!

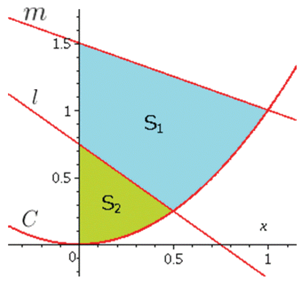


**Figure 4: A problem in a Tohoku Univ. entrance exam**

Definition 4.3. *Let $F$ be a finite set of $k$-variate functions, $\mathcal{D}$ be an $F$-invariant decomposition of $\mathbb{R}^k$, and $S$ be a union of a subset of $\mathcal{D}$ where $S$ is a region and $F$-invariant, and for all region $c'$ of $\mathcal{D}$, either $c'$ is a subset of $S$, or union of $S$ and $c'$ is not a region and $F$-invariant. If $S$ is bounded, we define $S$ is a* region enclosed by $F$.

Following the above definition, we have developed a Maple command where a set of polynomial curves is allowed as an input. Although our definition works for most cases, we have found a few problems that do not fit it. The following problem is taken from a Tohoku University entrance exam:

Problem 4.1. *Let $l$ be a line $y = -x + 3/4$, $m$ be a line $y = -x/2 + 3/2$, and $C$ be a curve $y = x^2$. Compute the area enclosed by $l$, $m$, and $C$ where $x \geq 0$.*

Fig. 4 shows the graphs of $l$, $m$, $C$, and $x = 0$, and regions enclosed by them. According to our definition, both $S_1$ and $S_2$ are 'enclosed by $l$, $m$, $C$, and $x = 0$' and the answer shall be the sum of their areas. The test writer however seems to require the area of $S_1$ only [4], because we do not need $l$ if the intended answer is the area of $S_1 \cup S_2$. We thus need to further tune the definitions to fit better to our intuitive grasp of 'the area enclosed by a set of curves' such that, for instance, $S_2$ is not counted as enclosed by $l$, $m$, $C$, and $x = 0$.

The axiomatic framework required to solve a particular problem as discussed above is a particular case of the notion of *frame* in artificial intelligence. In general, the frame for a problem is only ever determined after solving the problem, by looking back at the tools used for the solution, and not by determining the frame instantly from the presentation of the problem. It is not necessary to pick on a particular case such as the Jordan curve theorem, as the phenomenon will be familiar to anyone who has ever written a dissertation in mathematics. Definitions are not created in advance of the proof of a theorem, but are put together after the theorem has been proved. A finished paper has no sign of the to-and-fro that occurred before the definitions were finalized. This is a courtesy to the reader, and neither would it be allowed by mathematical esthetics. Mathematics dissertations (and textbooks) are thus written in the format in which definitions are given "as they should have been from the beginning," and theorems are proved based on these definitions. Hilbert and the other early formalists were outstanding mathematicians of their day, and were well aware

[4] The answer is not publicized.

of this inner reality. Yet they pretended (or more precisely, determined to pretend) that this format is the real mathematics. It was from this "double disguise" that was born formalism. The twisted attitude arises from the establishment of formalism itself, and therefore cannot be resolved within the framework of formalism, for a fundamental reason. This problem is an inescapable one for all fields associated with artificial intelligence, and of course the field of automatic problem solving is one of them.

As an intellectual endeavor, automated theorem proving hides deep inside it the twist concerning formalism but puts the philosophical problems about it into a cold storage. Instead, it pursues the possibility of statistically determining the frame of a problem from its statement and how far the proof can be automated once the frame is decided.

Within the framework of artificial intelligence, in addition to the above fundamental questions, there is also a heavy emphasis on the relative performance of an intellectual machine "compared to an average human."

The following problem caused a stir when it appeared in a Tokyo University entrance exam:

PROBLEM 4.2. *Show that* $\pi$ *is greater than* 3.05.

The reason this caused a fuss was that average students taking the exam found it quite hard, and the reason for this was that many of them were unable to grasp *what the problem was about.* In other words, the choice of an axiom set is a difficulty not only for a machine but also for the human examinees.

Within the research framework of artificial intelligence, hence, a rational choice is to limit oneself to those problems for which the axiom set, or the frame of the problems, can be determined from the vocabulary appearing explicitly in the problem statement.

# 5. DECIDABLE AXIOM SETS

Let us suppose that the problem statement has been successfully formalized as a proposition in ZF, and it has been determined what the problem talks about from the vocabulary of the problem statement, and that the problem has been converted to an expression in the first order predicate logic of a local axiom set. We have at last reached the start point, where we can consider whether it is possible to solve the problem automatically.

As described in Section 2, the question "Can a problem be solved automatically?" is almost the same as "Can the quantifiers be eliminated from this logical expression?" Here we are faced with a high wall. For an axiom set $T$, if there is a quantifier elimination algorithm, then we can regard $T$ as decidable.[5] To say that $T$ is *decidable* means that for any proposition $P$ (not including free variables) of $T$, there is an algorithm which determines whether $P$ is true or false in $T$. Extremely few decidable axiom sets are meaningful. For example, Gödel's incompleteness theorem and subsequent results in proof theory show that even a very restricted subset of the theory of natural numbers (Peano arithmetic) is undecidable [4][9]. However, fortunately it turns out that the theory of real closed field (with elementary geometry embedded in it) is decidable!

[5]For an algebraically closed field there is also a quantifier elimination algorithm, but this is not decidable until the characteristic has been determined.

**Table 1: Results on Univ. Tokyo Mock Test**

| Science Course (full score: 120) | | | |
|---|---|---|---|
| Test set | Translation | | Human |
| | Manual | Semi-Auto | Avg. |
| 2013 Jul | 40 | 40 | 21.8 |
| 2012 Nov | 40 | 40 | 32.5 |
| 2012 Jul | 25 | 18 | 29.8 |

| Humanities Course (full score: 80) | | | |
|---|---|---|---|
| Test set | Translation | | Human |
| | Manual | Semi-Auto | Avg. |
| 2013 Jul | 40 | 40 | 24.9 |
| 2012 Nov | 20 | 8 | 25.7 |
| 2012 Jul | 40 | 40 | 30.9 |

The RCF quantifier elimination (RCF-QE) algorithm was first developed in the 1930s by Tarski [12].[6] His name will no doubt be familiar to many readers from the truly mysterious theorem known as the Banach-Tarski theorem. The RCF-QE algorithm by Tarski is so extremely inefficient that it was never implemented on a computer. However, in the 1970s, Collins proposed an efficient algorithm using Cylindrical Algebraic Decomposition (CAD), which enabled implementation of RCF-QE on a real machine [3]. Quantifier elimination by CAD is based on Sturm's theorem on counting the number of different real roots of a polynomial, and is currently implemented in Mathematica [11], Maple (SyNRAC [6]), and other computer algebra software.

We evaluated a prototype problem solving system on the University of Tokyo entrance exam mock tests held by one of the largest cram schools in Japan (Yoyogi Seminar). In the prototype system, a natural language processing module is combined with an implementation of RCF quantifier elimination (SyNRAC) [6]. The problems in the mock tests are roughly at the same difficulty level as the real entrance exams of the University of Tokyo. There are two types of tests, one is for future applicants for the Univ. of Tokyo science courses and the other is for humanities courses. Both types of the mock tests are sat by thousands of test takers.

We run the system both on manual translation of the problems and those semi-automatically derived from annotated texts. The annotation on the text includes linguistic annotations such as the syntactic structures of the sentences as well as a small additional dictionary for those words in the problem that were missing in the main dictionary at the time of the experiment. These additional information was used as a surrogate of a part of the natural language processing module currently under development. Table 1 shows the results on the six latest test sets that were available to us at the time of writing. The successfully solved problems included those on 2D and 3D geometry, linear algebra, and calculus. Fig. 5 shows one of the problems that were successfully solved by the system, on which human average score was 2.1 points out of the full score of 20. The prototype system has a simple output module that generates human-readable answer sheet. Fig. 6 shows the generated answer sheet for the problem in Fig. 5.[7]

[6]The result was not published until after WWII in 1951.

[7]It was manually translated to English from the system's

PROBLEM 5.1. *Let $a$ be a positive real constant. Suppose that two real numbers $x$ and $y$ satisfy*

$$\frac{1}{2} \leq x \leq 1 \text{ and } a \leq y \leq 2a.$$

*Under these conditions, find the minimum value of*

$$F = \frac{y}{x} + \frac{x}{y} - xy.$$

**Figure 5: A problem in Univ. Tokyo mock test (2013 Jul; Science course)**

Let

$$x_{gen12}$$

be the real number we want to determine.
The condition given in the question is equivalent to the following first order formula:

$$(0 < a \wedge a \leq y \wedge y \leq 2a \wedge \exists x_{.0.0}(\exists y_{.0.0}(a \leq y_{.0.0} \wedge y_{.0.0} \leq 2a \wedge y_{.0.0}(-x_{.0.0}) + \frac{y_{.0.0}}{x_{.0.0}} + \frac{x_{.0.0}}{y_{.0.0}} = x_{gen12}) \wedge \frac{1}{2} \leq x_{.0.0} \wedge x_{.0.0} \leq 1) \wedge 0 < a \wedge (\forall y_{.0}(\forall x_{.0}(\frac{1}{2} > x_{.0} \vee x_{.0} > 1 \vee x_{gen12} \leq y_{.0}(-x_{.0}) + \frac{y_{.0}}{x_{.0}} + \frac{x_{.0}}{y_{.0}}) \vee a > y_{.0} \vee y_{.0} > 2a) \vee 0 \geq a))$$

Since this formula is in the language of real closed field, by the Tarski-Seidenberg theorem, it is possible to find a quantifier-free formula that is equivalent to the above formula. In fact, we have the following formula by using Tarski's quantifier-elimination algorithm (transformation process omitted):

$$((0 < a \wedge a \leq \frac{1}{2\sqrt{5}} \wedge x_{gen12} = \frac{12a^2+1}{4a} \wedge a \leq y \wedge y \leq 2a) \vee (\frac{1}{2\sqrt{5}} < a \wedge a \leq \frac{1}{2\sqrt{2}} \wedge x_{gen12} = \sqrt{4 - 16a^2} \wedge a \leq y \wedge y \leq 2a) \vee (a > \frac{1}{2\sqrt{2}} \wedge x_{gen12} = \frac{1}{2a} \wedge a \leq y \wedge y \leq 2a))$$

By solving this, we have:

$$\begin{array}{lll} x_{gen12} = \frac{12a^2+1}{4a} & \text{if} & (0 < a \wedge a \leq \frac{1}{2\sqrt{5}} \wedge a \leq y \wedge y \leq 2a) \\ x_{gen12} = 2\sqrt{1 - 4a^2} & \text{if} & (\frac{1}{2\sqrt{5}} < a \wedge a \leq \frac{1}{2\sqrt{2}} \wedge a \leq y \wedge y \leq 2a) \\ x_{gen12} = \frac{1}{2a} & \text{if} & (a > \frac{2}{2\sqrt{2}} \wedge a \leq y \wedge y \leq 2a). \end{array}$$

**Figure 6: Answer sheet produced by the prototype system**

Overall, the results are encouraging. In four out of the six test sets, the scores attained by the prototype system matched the averages of human test takers. We should note that the results are optimistic estimates of the performance of a true end-to-end system wherein the linguistic annotations on the text are automatically produced by natural language processing. The results however show that the overall architecture of our system is a reasonable choice at least as a first step toward a true end-to-end problem solver.

The promising experimental results owe much to the intensive research in last decades that has pursued practical algorithms of RCF-QE. Even though the efficiency is vastly better than that of Tarski's original algorithm, the calculation time required for CAD is doubly exponential in the number of variables $n$ in the proposition supplied. The practical limit to obtain a solution would be at most five variables. For example, it would not be possible to solve the following problem, even with a supercomputer.

output in Japanese.

PROBLEM 5.2. *A convex quadrilateral lies in a plane. Find the point from which the sum of the distances to the vertices is minimum.*

The RCF-QE algorithm can in principle be improved to doubly exponential of the number of alternating quantifiers in the given proposition. While this may seem a subtle change, it is extremely important: in almost all theorems found in mathematics up to now (excluding somewhat artificial cases found in mathematical logic), the number of alternating quantifiers is at most four. This is probably the limit of complexity for a problem that can be grasped by the human mind. Given the discovery of an optimum quantifier elimination algorithm, it would not be at all surprising for an automatic system to be able to solve all of the elementary geometry and real number problems within human grasp.

# 6. CONCLUSION

Wittgenstein once asked:

> What is left over if I subtract the fact that my arm goes up from the fact that I raise my arm? [15]

Let us think similarly. What is left over if we subtract mathematics done by a machine from mathematics done by a human?

This is the sort of question that leaves mathematicians with an uncomfortable feeling.

It is the authors' confident belief that mathematics will not fall to the machine. Yet, both chess and shogi have already been conquered. The simple faith in Hilbert's day that mathematics continue in the human realm is progressively problematic for us living in the 21st century.

Alternatively, we could ask this: if we subtract the mathematics done by machine from the mathematics which we can teach to a human, is there anything left?

This question is more violent than the one above, and is not easy to confidently answer in the affirmative. As any mathematician involved in mathematics education is painfully aware, up to the present the subjects we have been relatively successful at teaching students with textbooks and lessons are multiplication tables, differentiation and integration, matrix manipulation, and so on, but these are just the subjects we can easily teach a machine.

This is why, in the authors' opinion, we must ask this violent question. This is the only methodology for scientifically asserting the existence of what remains.

# 7. REFERENCES


[1] T. Arai. *Mathematical Logic*. Iwanami Shoten, 2011. (in Japanese).

[2] A. Church. A note on the Entscheidungsproblem. *Journal of Symbolic Logic*, 1:40–41, 1936.

[3] G. E. Collins. Quantifier elimination for real closed fields by cylindrical algebraic decomposition. In *Automata Theory and Formal Languages 2nd GI Conference Kaiserslautern*, volume 33 of *Lecture Notes in Computer Science*, pages 134–183. Springer-Verlag, 1975.

[4] K. Gödel. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik*, 38(1):173–198, 1931.

[5] T. C. Hales. The Jordan curve theorem, formally and informally. *The American Mathematical Monthly*, 114(10):882–894, 2007.

[6] H. Iwane, H. Yanami, H. Anai, and K. Yokoyama. An effective implementation of symbolic-numeric cylindrical algebraic decomposition for quantifier elimination. *Theoretical Computer Science*, 479:43–69, 2013.

[7] T. Matsuzaki, H. Iwane, H. Anai, and N. Arai. The complexity of math problems – linguistic, or computational? In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pages 73–81, 2013.

[8] T. Matsuzaki, H. Iwane, H. Anai, and N. H. Arai. The most uncreative examinee: a first step toward wide coverage natural language math problem solving. In *Proceedings of the 28th AAAI Conference on Artificial Intelligence*, 2014. (to appear).

[9] J. Robinson. Definability and decision problems in arithmetic. *The Journal of Symbolic Logic*, 14(2):98–114, 1949.

[10] M. Steedman. *The Syntactic Process*. Bradford Books. Mit Press, 2001.

[11] A. W. Strzeboński. Cylindrical algebraic decomposition using validated numerics. *Journal of Symbolic Computation*, 41(9):1021–1038, 2006.

[12] A. Tarski. *A Decision Method for Elementary Algebra and Geometry*. University of California Press, Berkeley, 1951.

[13] A. Turing. On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, 42:230–265, 1936.

[14] A. N. Whitehead and B. A. W. Russell. *Principia Mathematica*. Cambridge Univ. Press, Cambridge, 1910, 1912, and 1913.

[15] L. Wittgenstein. *Philosophical Investigations / Philosophische Untersuchungen*. Oxford: Basil Blackwell, 1953.