ISSN 1883-1974 **No. 32, January 2010** (This English language edition of NII Today corresponds to No.46 of the Japanese edition)

NII SPECIAL Mathematics and Logic as the Cornerstone

Solving Issues in the Real World with Mathematics — Mathematics and Logic and Their Applications Mathematics / Logic and Informatics

> Algorithms to Fill in the Gap of the Genome! Why Aspire to Mathematics at the NII?





Ken-ichi Kawarabayashi

Professor, Principles of Informatics Research Division, NII



NII Interview: Ken-ichi Kawarabayashi+Mariko Takahashi

Solving Issues in the Real World with Mathematics Mathematics and Logic and Their Applications

Takahashi: Your main research interest is graph theory. I first heard this word in 1990 when Japanese high school students participated in the Mathematical Olympiad for the first time. I recognized it as a genre that is the source of many exam problems.

Kawarabayashi: Yes, problems are often taken from graph theory (*1) or discrete mathematics (*2) or combinatorics (*3).

Takahashi: Whenever I ask what graph theory is, no professor explains it to me in a single word. They say "it is geometry of point and edge and line," or "it is unicursal arithmetic," and so on, and it never rings a bell.

Kawarabayashi: Probably so. But the real world is full of phenomena that can be expressed with a graph, road networks for example. When I was a student I only had mathematical interests, but after getting my doctorate I spent a year at Princeton University where I came to understand that graph theory is widely used for things I knew nothing about, so I came to the direction of theoretical computer science. Phenomena that occur in the real world are quite interesting.

Takahashi: Specifically?

Kawarabayashi: For example, the mechanism for traffic jams. When there is an accident somewhere, traffic jams extend widely, thinly and on a large scale. How to analyze this "widely and thinly" is quite a difficult problem, one that is often discussed in mathematical circles.

Takahashi: Jamology is trendy now,

NII SPECIAL

Mathematics and Logic as the Cornerstone

(*1) Graph theory is a branch of mathematics that studies the properties of graphs consisting of a set of vertices and a set of edges. It has wide applications for computer data structures and algorithms.

(*2) Discrete mathematics refers to a branch of mathematics that studies discrete (in other words, non-continuous) objects. Combinatorics and graph theory are at the heart of the field.

(*3) Combinatorics is a branch of mathematics that studies bodies formed of (usually finite) objects that fulfill certain conditions. It is closely related to the principal branches of mathematics such as number theory, abstract algebra, geometry and analysis.

isn't it?

Kawarabayashi: Jamology is a little different. Jamology studies the reasons for traffic congestion. I think that talking about the impact of a traffic jam after it has occurred is something completely different. For example, when the shortest route is displayed on a car navigation system, the calculations must be quick with a tolerable degree of accuracy and speed, adjusting for accidents etc. with each passing moment. No matter how accurate, calculations that take an hour are meaningless. How quickly and how accurately to do it are perfect problems for logic.

Takahashi: How do you choose the problems?

Kawarabayashi: I go to the top overseas conferences in the field of computer science and I stock up on talk. In a 20 minute presentation, the first half is devoted to explain why someone is working on a problem and how it applies in the real world. Graduate students overseas are particularly good at such explanations. The top conferences on computer science are interesting, they attract bigname mathematicians as well as people who are involved with the practical aspects. A lot of systems people and coding people attend as well so it is incredibly stimulating.

Takahashi: Japan does not organize such conferences.

Kawarabayashi: To begin with, there is not much interaction between the mathematical sciences, computer sci-

Mariko Takahashi

Staff Writer, Tokyo Head Office, The Asahi Shimbun



ences and informatics. Besides, when modeling a real problem, we need to put our heads together with people who are working on the practical aspects and this is where cooperation is still a little lacking. Japanese mathematics has a set of values that views pure mathematics as "pedigreed" and I think practical applications have been looked down on in the course of its history. But the people, who is now at the top of the international union of mathematics, is working on discrete mathematics and various practical

applications, and he has a lot of influence on the computer sciences. I think the time is changing to a remarkable degree.

Takahashi: Japan is a little removed from world trends. Kawarabayashi: Very far re-

moved. It is often said that even the salaries are higher for professors of applied mathematics at the top universities in the United States.

Takahashi: So the United States is still creating the global trends?

Kawarabayashi: Israel is fairly strong in computer sciences. There are also some amazing people in India. When I was at Princeton University, it seemed like one third of the department were Indians, one third Chinese and the rest where other people.

Takahashi: What about the Japanese? Kawarabayashi: Not visible at all. Nobody is toughing it out at all.

Takahashi: Recently, even the hardware of Japanese computers has been suspect. In the past people used to say that the hardware was good but the software was hopeless. This is linked to weaknesses in the theoretical and fundamental areas of information

science, isn't it?

Kawarabayashi: There is a very strong link. A long time ago, I visited for a show time at Microsoft Research, but the people in the theory group had almost no responsibilities. It was enough to be working on your own research. However, when other people worked on developing software and something strange happened, if it was a logic problem, everybody came together to find the solution. Recently, somebody from AT&T (the American telecommunications company) said that they work



in the same way. This is clearly why people who are excellent at theory earn much higher salaries than a professor of the mathematical science. I think Japan must also create an environment like this.

Takahashi: Because Yukio Hatoyama has become Prime Minister, some light might shine on mathematical programming...

Kawarabayashi: During the recession in the 1990s, the United States increased spending on mathematics. Fundamentally, theory is only a matter of personnel cost so it is not that expensive. As for Hatoyama's discipline of operations research, it deals with, for example, being able to save billions of dollars in airfares by successfully scheduling major league tournaments. A range of conditions, such as half each of home and away games, or that a nine-game series at home is no

good, are fed into a computer which runs the calculations. It is the kind of thing that students at Harvard University or MIT think up and are awarded scholarships for.

Takahashi: Perhaps you could receive a percentage of the money saved?

Kawarabayashi: You're right. Even 10% would be enough to put up a whole building. I am sure that professional baseball in Japan and the J-League could save hundreds of millions of yen.

Takahashi: When they hear examples

like that, people reconsider their opinions of theory.

Kawarabayashi: For that, we need excellent young talent to enter this field. That is why I think it is important that a research institute like ours does well and that it is widely

known.

🔇 A Word from the Interviewer

When you hear that Kawarabayashi earned his doctorate after only one year of Ph.D. studies at Keio University, you realize that he has a natural ability. Mathematics is a world where people in their 40s and 50s have held sway, but informatics is a fast-moving field where people in their 30s must play the lead role. Indeed, to win Japan a world ranking in information is a big job and a lot of hopes are pinned on Kawarabayashi, who became professor at age 34. To change the Galapagos tendencies of the mathematical world in Japan, to change the way society views mathematicians, and to increase the number of mathematicians working at corporations. The future of Japan depends on bringing this to fruition.

Mathematics and Logic as the Cornerstone

Mathematics / Logic and Informatics

Whether delving into the enigma of the stars in the universe, or predicting the path of a typhoon, or maneuvering a robot, unless mathematics is present at the foundation, it is not possible to solve problems. However, mathematics is a discipline that began with the curiosity and the desire of human beings to learn, and mathematicians say that, similar to art, it possesses 'a beauty' that appeals to the emotions of human beings. We listen to a discussion about the attraction of doing research in mathematics and logic, from the significance of pure mathematics to mathematics that contributes to the world and, above all, mathematics that is applied to informatics.



Ken Satoh Professor, Principles of Informatics Research Division, NII

Mathematical logic and mathematical engineering supporting the everyday world

Satoh: To start with, please tell us about your field of research, Professor Tatsuta.

Tatsuta: My field of research is mathematical logic. It is the discipline that analyzes the way human beings think about mathematics by objectifying and precisely describing the sequence of thoughts. It is a discipline with an extremely long history dating back to human pursuits in ancient Greece, but it has come to attract more attention with its application to computers for the past several decades. Above all, it is now an indispensable discipline in many fields including hardware circuits, hardware and software verification, software specification description, designing programming languages, and reasoning by means of artificial intelligence.

For example, software verification uses mathematical proof to confirm whether or not the software is functioning correctly. There are bound to be mistakes because the only way to produce software even now is for a person to write it manually. If there are mistakes in the software, the PC freezes, online banking systems stop working, or we lose control of social infrastructure such as airplanes or nuclear power, so that there may be serious disruptions to everyday life. All these things are caused by 'bugs,' or parts of the programs where the programmer unwittingly wrote in mistakes.

The normal approach is to launch a test program and if it more or less works, the software is run in practice. However, serious accidents may also occur if there is an omission in the test and it misses the bug. This is also called the software crisis. Consequently, in order to produce software that is completely free of bugs, a guarantee that the software will work in any scenario are required through mathematical verification.

However, as matters now stand, this method is seldom adopted because to mathematically verify software requires a lot of effort and expense. Nonetheless, since software plays an extremely important role in supporting social infrastructure, all software will probably be mathematically verified in the future. A new kind of mathematics, program theory is a boundlessly interesting and in-depth discipline with the capability to unravel the true nature of computation inherent in software.

Satoh: For example, these days, a lot of computers are installed in cars and verification on their correct behavior is turning into a major issue. Moreover, with cars, it is necessary to exercise judgment and control in realtime and instantaneously. It seems that mathematical logic would have an extremely important role to play in such situations as well.

Tatsuta: That's true. In addition, mathematical logic is also useful for computer education. For example, it plays a role in fostering the foundation for the way that software engineers think. Software is written logically and executed logically and by studying mathematical logic, engineers can acquire the basic grounding they need.

Satoh: I see. Now, over to you, Professor Hayami. **Hayami:** My field of research is mathematical en-

TLCA List of Open Problems

http://tlca.di.unito.it/opltlca/ Updated July 11, 2008

Problem # 20 [SOLVED]

Submitted by Mariangiola Dezani-Ciancaglini Statement. Type theoretic characterisation of hereditary permutations

According to [Bergstra and Klop, 1980, Barendregt, 1984] (Definition 21.2.9) a closed lambda-term M is a hereditary permutation if the Böhm tree BT(M) has the following properties:

Date: 2006

- the label ⊥ does not occur in BT(M); each variable occurs exactly once
- the head variable is the first abstracted variable; all other variables occur at one level lower than their abstractions

problem is to find a type assignment system in which all and only the hereditary permu-ons get types of a fixed shape. any other sets of A-terms have been characterised in this way, see for example [Kurata, 2002, ani-Ciancaglini et al., 2005] and the references there.

Solution: Makoto Tatsuta proved [Tatsuta, 2008] that the set of all hereditary per

- [Barendregt, 1984] Barendregt, H. (1984). The Lambda Calculus. Its Syntax and Semantics. North-Holland, second edition. ergstra and Klop, 1980] Bergstra, J. A. and Klop, J. W. (1980). Invertible terms in the lambda calculus Theoretical Computer Science, 11:19-37.
- [Dezani-Ciancaglini et al., 2005] Dezani-Ciancaglini, M., Honsell, F., and Motohama, Y. (2005). Composi-tional characterization of Asterns using intersection types. *Theoretical Commuter Science*, 340(3):459–495.
- [Kurata, 2002] Kurata, T. (2002). Intersection and singleton type assignment characterizing finite Böhm-trees. Information and Computation, 178(1):1–11.
- ta, 2008] Tatsuta, M. (2008). Types for hereditary permutators. In Proceedings of the nual IEEE Symposium on Logic in Computer Science, LICS 2008, 24-27 June 2008, P A. http://research.mi.ac.jp/TechReports/07-010E.html.

Fig. 1 Solution to Problem No. 20. The International Society for Program Theory published a list of 22 unsolved problems in the field of mathematics. Professor Tatsuta tackled the problem No. 20, 'Provide a type theory that characterizes the lambda calculus corresponding to bijection, and gave an elegant solution as shown here.

gineering. In a nutshell, it is a discipline that first converts a problem into mathematical language to find the solution and work out the methodology for the solution when there is a problem to be solved. My specialty within the field is numerical computation. I develop the basic technology for using computers to simulate phenomena, in nature, engineering, social phenomena, etc.

For example, for simulations for weather forecasts, including typhoons, or airplane design, we have to solve equations that express fluid phenomena. They are the mathematical models of fluids. So, we solve these equations, but if we are dealing with complex phenomena and complex shapes, they cannot be solved analytically by hand, so we use computers to find approximate solutions. This is where a process called 'discretization' is needed. Discretization converts and approximates continuous or analog equations into discrete equations that computers can handle. In addition, in order to improve the accuracy of the approximation of the discretized equation, for example, when dealing with three dimensions, we need to discretize the space using an enormous mesh, and the number of variables handled by the computer also increases. Consequently, it is necessary to develop efficient numerical methods (algorithms) to derive an answer using the limited memory and computation speed of the computer.

Mathematics is also necessary in order to guarantee and analyze how the numerical methods that have been developed generally behave, in what situations they function properly, the degree of accuracy in the answer, and the speed with which they approach (converge to) the answer.

Thus, mathematics is necessary for everything including modeling the phenomena, discretization, developing and analyzing algorithms. In short, if using computers to simulate various phenomena is one of the missions of informatics, I think you can understand that mathematics is necessary in every situation.

Satoh: Modeling phenomena and discretization, developing and analyzing algorithms - each of these deal with how to convert the phenomena of the real world into mathematics and to present it in a way that is easy to understand.

Hayami: That's right. The mission is to get as close to the real world as possible and in order to increase accuracy, models, computational methods and computers are required to constantly evolve.

The joy of solving mathematical challenges.

Satoh: Specifically, what kind of research are you doing?

Tatsuta: I am working on program theory and mathematical logic. The former involves mathematical logic for software verification as I already mentioned, while the latter is a discipline that dates back to the times of the ancient Greeks. One recent achievement was two years ago when I found the solution to problem No. 20 on a list of 22 open problems compiled by an international society of program theory, 'Provide a model theory defining the lambda calculus corresponding to bijection.' (Fig. 1). To paraphrase, the query says, 'Define abstract programs that



Makoto Tatsuta Professor, Principles of Informatics Research Division,NII



Ken Hayami Professor, Principles of Informatics Research Division,NII

Mathematics and Logic as the Cornerstone

express one-to-one correspondence using analysis of data types.' As a result I presented a proof that 'it does not exist.' It was the first of the 22 open problems to be solved, contributing not only to the development of software theory but it is also a major accomplishment in the sense of the search for truth.

Hayami: I am doing research in numerical methods for solving large-scale simultaneous linear equations and least squares problems on the computer. Large-scale simultaneous linear equations appear when discretizing the equations for fluid as I mentioned earlier. For example, if the space is discretized in to 100 meshes for each direction, the whole space is discretized into a million mesh points and simultaneous equations with one million unknowns is created. The issue is how to solve this efficiently.

The least squares method was originally developed by the 19th century mathematician Gauss in order to predict and measure the orbits of heavenly bodies. (Refer to 2007 NII Public Lecture 'Mathematics that Counts in Society') Its



Fig. 2 Comparison of the BA method and the conventional method.

Developed the BA method, a new iterative method for large-scale least squares problems. When the BA method is used, it is possible to find the least squares solution with far fewer iterations compared to the conventional method.

applications are wide-ranging and include statistical analysis of large amounts of data produced by experiments, image processing, signal processing and control.

When you try to solve large-scale least squares problems on a computer, the 'iterative method' which improves the solution by repeated computation is important. At present, we are developing methods which can solve even large-scale and ill-conditional problems at high speeds, and we are analyzing the behavior of the methods mathematically. My graduate students come up with one new idea after another, and I am enjoying doing research together with them.

Satoh: For the average person, both themes are mostly unfamiliar and difficult but it is clear that mathematics and logic are indispensable to the progress of science and technology.

The value of mathematics is in curiosity and emotion.

Satoh: On the other hand, mathematics is not only something that is useful to society in specific terms. In pure mathematics, above all, the emphasis is on the quest for truth.

Tatsuta: Looking back at the history of mathematics, there are many examples of how theorems and laws that are important to the quest for truth have, after several decades or several centuries, turned out to be useful in the real world. Conversely, research is not done having in mind from the outset how it can be useful.

I believe that there is a threefold significance in mathematics and logic and the most important is curiosity. The significance of mathematics is to enrich the mind by responding to the human desire of wanting to know. Astronomy, for instance, is a discipline that was initially driven by the desire to discover what lies beyond the beautiful stars. In a word, the meaning of mathematics and logic is similar to that of literature.



Fig. 3 Example of applying the BA method to least squares problems. It can be used to interpolate topographical data as shown in the figure. (Created by Professor Hiromichi Hashizume, Information Systems Architecture Science Research Division, NII)

For example, Fermat's theorem (*No integers n>2, x,y,z>0 exist such that $x^n+y^n=z^n$), which many gifted mathematicians continued to challenge for many years, or the four-color theorem stating that any map can be colored using four colors, or questions about large prime numbers, or the diffidult problem about the nonexistence of programs that can simulate all programs, i.e., the universal Turing machine – the source for all these lie in curiosity. I believe that literature and the arts exist because people are moved when they sense their value, and the same is true for mathematics and logic. The solution to the difficult problem I mentioned before is certainly poignant.

Satoh: Certainly, when I look at a good proof, I think that a great job is done and I feel the beauty of it. This is a sensation that scientists and mathematicians share and mathematicians are able to tackle the so-called difficult problems body and soul because they genuinely sense the fun and the excitement.

Hayami: Interestingly, it is also a fact that the more a theory inspires and excites curiosity and the more mathematicians instinctively feel it is beautiful, the more useful the results will be to society.

Tatsuta: Regrettably, since the general public find it difficult to understand the curiosity and emotion as the significance of mathematics, at present there tends to be an over-emphasis on the significance of being useful to society. In fact, there is no doubt that mathematics and logic are indispensable to the development of science and technology. However, research in mathematics and logic takes an incredibly long span of time. Fermat's theorem was solved in 1994 after 360 years, and the four-color theorem in 1976 after 120 years. Moreover in many cases, it is even much later that it becomes evident whether it is useful or not in concrete terms. Logic, which has been studied since the ancient Greece, is now indispensable to computers, the four-color theorem is useful for optimizing a range of computations, while large prime numbers are useful in the field of coding, but it took several decades and even several centuries for this to happen.

Satoh: It is not just mathematics, you can say the same thing for physics. Newton's low of universal gravitation was not discovered with the intent of making it useful in our daily lives, but it led to the discovery of to Einstein's theory of relativity from which the development of nuclear energy was started.

Hayami: Of course, theories sometimes arise from the desire to solve a problem, so there is no need to regard the fact that it is useful to the development of science and technology negatively.

Incidentally, Gauss concentrated on research in number theory as a young man, but later turned to astronomy, devising the method of least squares, and was catapulted to fame for predicting the emergence of a comet. Further, he used the method of least squares in the survey of land and producing maps. Tackling the problems in survey, he triggered the birth of differential geometry. In a later generation, differential geometry was used to construct Einstein's theory of general relativity. In this way, science is structured like a chain where theories that are useful to the real world, stimulate curiosity and give rise to new theories.

Tatsuta: If I may add a word, it may be that only the mathematician's instinct for truth and conscience can provide the right research directions for mathematics and logic...

Satoh: If it is only a matter of usefulness, the job is done when the current problem is solved, but since there is a generality to a discipline as deep as mathematics, there is potential for a range of applications. This again creates new theories, which may become tremendously useful in unexpected areas.

Thank you for taking your time today.

(Written by Madoka Tainaka)

Informatics & Genetics

Algorithms to Fill in the Gap of the Genome!

Little by little, the Genome Project has identified the entire gene sequence of living organisms. They say that the genome of many life forms has already been unraveled but in actual fact the genomes are still gaps of sequences that have not been accurately clarified. New algorithms are attempting to fill in the unidentified gaps.In addition, a new approach to biological research is emerging with these algorithms.



Takeaki Uno Associate Professor, Principles of Informatics Research Division, NII

Attempting to create fusion research in new areas, the Research Organization of Information and Systems organized a 'Cross-talk for Young Researchers' at Ito in Shizuoka Prefecture in December 2006. In search of a field of application for a similarity analysis program he had written, Takeaki Uno, Associate Professor at the Principles of Informatics Research Division. NII, gave a presentation on how fast we can find similarities in the human and mouse genome sequences in an extremely short time. Meanwhile, Dr. Juzoh Umemori, specially appointed researcher at the National Institute of Genetics, participated in search of an analytical method for the mouse genome sequencing that he was researching. "This is it!" he thought and started talking to Prof. Uno. The two of them hit it off and talked until midnight about applications for the program.

Due to its potential, their joint research was accepted as a project at the Transdisciplinary Research Integration Center, marking the start of three years of fusion research.

Searching for things that computers can do

The first task was to accurately unravel chromosome 13 in the mouse genome sequence. This chromosome has an area named *Genic1*. Dr. Umemori had researched this area (Fig.) since he determined that the cause for nerve abnormalities (photo) that occurred when



cross-breeding particular families of mice was in Genic1. However, the research was hindered by a lack of progress in unraveling the genome sequence because there were many repeated sequences making analysis of the area difficult.

Saying that it was his job to find the areas where computers can play an active role, Prof. Uno started to think about where in the genome sequence definition there was scope for a computer to take an active role. The genome is made up of a series of 4 types of base pairs called AGCT. With current technology. it is only possible to read the sequence by a technique of cutting it into chunks and reading 1000 base pairs at a time. By reading the numerous fragments in a sequence and studying how they overlap, it is possible to unravel the original sequence. It is as if researchers were assembling the pieces of a puzzle. Computer programs are needed to find out how the large numbers of fragments fit together. Prof. Uno, whose specialty is the 'algorithm research' that designs the calculation methods that determine the performance of the program, proposed a new 'high-speed homology search algorithm.'

One of the new features of Prof. Uno's program is that it is able to calculate similarity for 30 base pairs (Fig. left). Until now, most programs have discovered overlapping areas by gaining a foothold where approximately 10 base pairs correspond exactly. Since 4 types



Fig. Comparison of sequences of mouse genome fragments (left) and the base pair sequence in mouse chromosome 13

Fig. left shows the result of using the "high speed homology search algorithm" to calculate similarities in two genome fragments. It is clear that there is overlap in the shaded area of the two genome fragments that correspond to the clean white area in Fig. right. By determining the order of the genome fragments in this way, it is also possible to understand the base sequence of mouse chromosome 13.



Juzoh Umemori Specially appointed researcher, National Institute of Genetics



Tsuyoshi Koide Associate Professor, National Institute of Genetics



Photo Normal mouse (right) and genetically incompatible mouse (left)

If you crossbreed a Korean mouse with a Bulgarian mouse, some mice will have a genetic incompatibility and be taken with a fit of trembling. Dr. Umemori ascertained that the cause is in the *Genic1* area.

of base pairs and 10 characters can be lined up in 4^{10} ways = approximately one million, it is possible to increase the calculation speed by narrowing down the list of overlapping areas to 1 one-millionth. However, when dealing with genome sequences with more than 10 billion base pairs, there are at least 10,000 or more areas of correspondence even if the list is narrowed down to 1 one-millionth. In reality, these calculations are not possible. When using Prof. Uno's program, there are approximately 4^{30} types of sequences and there is very little chance that sequences that are not the same will be equated with one another.

Another important feature of the program is that it takes into account errors in the genome fragment sequence that occur in read processing. The correct sequencing for *Genic1* was unraveled (Fig. right) by using the algorithm "even if 1 or 2 base pairs among 30 do not correspond, the sequence can be regarded as the same."

The genome starts to look different

At present, the fusion research is entering a new stage where the program is providing hints for biological research. If you use the 'high-speed homology search algorithm,' it is easy to investigate what kind of repeated sequences exist with approximately what frequency in the genome. It will probably be possible to acquire new knowledge about the repeated sequences, which were thought difficult even to discover in the past.

Prof. Uno says that he only provided the tools. "Even if the computer discovers a list of repeated sequences, it is, in the final count, the human being who makes the judgment. That's why Dr. Umemori with his knowledge of biology is incredible." Dr. Umemori, on the other hand, says, "I have started to think that these repeated sequences may be highly significant to the genome structure. This is the result of the influence of the ideas of Prof. Uno, who views the genome as simply rows of base pairs." A completely new way of looking at things has emerged by fusing two different fields.

Cooperating for their respective goals

Referring to the frame of mind for advancing fusion research, Prof. Uno says, "Naturally there are discrepancies. That is why we have to keep talking until we understand each other." To begin with, Dr. Umemori was interested in the biological knowledge that could be obtained by using the 'high-speed homology search algorithm" while Prof. Uno was interested in finding out how far the program he had developed could go in terms of application scope. It is tough for two researchers with different expertise and goals to work together. However, they have obtained unprecedented research results precisely because they compliment one another in areas where the other falls short.

"When you are developing programs, there is a tendency to avoid complex real problems and to escape into hypothetical problems that are easy to calculate. I think that with this fusion research I have been able to expose myself to realistic research that is actually useful as opposed to such information research," says Prof. Uno. On this point, he has been highly commended with an award for excellence at an international conference for data mining. Tsuyoshi Koide, Associate Professor at the National Institute of Genetics, who has cooperated with both researchers recalls, "It was also challenging for me to use the new technique of information to try to solve biology problems that had not been solved in the past."

This fusion research project will come full term in March 2010. However, both Prof. Uno and Dr. Umemori agree that there are still things they would like to do. There will be more days with the two of them discussing the data before them. (Written by Akiko Ikeda)

Why Aspire to Mathematics at the NII?

Where is the real thrill in studying mathematics, the foundation for science and, above all, the foundation for informatics? Sho Yamashita and Keiichi Morikuni, who are in the second year of the Five year Ph.D. course in Informatics at the Graduate University for Advanced Studies (SOKENDAI), and Mahmudul Faisal Al Ameen, who is in the third year of the Three year Ph.D. course at the same university, talk about what motivated them to aim for mathematics and why it is interesting.

----Please tell me what prompted you to study for your current major.

Yamashita: Originally I studied humanities. I had an interest in personal communication but I was also drawn to mathematics, and when I was thinking about studying both, I came across the field of mathematical linguistics. Associate Professor Kanazawa had been my supervisor since I was a university student (Chiba University, Faculty of Letters, Division of Behavioral Science) and as my interest in the discipline grew, I thought I would like to continue learning from him and that is why I am studying here. My research is on building mathematical models that can simulate the various linguistic phenomena that people use.

Morikuni: As an undergraduate, my major was in acoustics (Kyushu University, School of Design, Department of Acoustic Design). My research for my graduation thesis was on the application of numerical computation techniques for the purpose of mathematical prediction of acoustic fields and design of them.Since this triggered a growing interest in numerical computation, I found my current supervisor and set my sights on this major.

Faisal: I have come from Bangladesh where I studied for a bachelor's degree in computer science at Darul Ihsan University and master's in East West University. Professor Tatsuta is my supervisor and I am now researching on the techniques for verifying program design and the separation logic that forms the basis for this. I have also studied lambda calculus and first order predicate logic here as a foundation of my current work. I think that informatics is the foundation for modeling and analyzing all artifacts and natural systems. Today, computers are gradually gaining entrance in society, supporting our lifestyles in various ways. We are operating a wide variety of programs and there is a demand for reliable algorithms and techniques for checking the correctness of a program to support the designs.

In Bangladesh, informatics and computer science research is still in an embryonic state, but it is a field to emphasize in the future. These are also reasons why I want to do research.

Mathematics is necessary to underpin logic.

— Please describe your research for us.

Yamashita: The natural language that people speak has rules for word order and sentence formation, which we refer to as 'formal grammar' when expressed in mathematical terms. For

example, when we say in natural language, "Katsuo ate the fish that Sazae cooked," we have a structure with one sentence embedded within another. The basic query is whether it is possible to model this kind of ordinary linguistic phenomena by means of formal grammar. This query is answered by means of the mathematical properties of formal grammar. My research is about investigating these sorts of mathematical properties.

In programming language, there is a software called a compiler that converts programs to a machine language that the computer can execute. This is where the results of formal grammar are used and this is an area where the research results can be applied.

Morikuni: My research is closely re-

lated to the simultaneous linear equations we study in junior high school. In the simultaneous equations that you learn in junior high, the number of equations is the same as that of unknowns, and the equation can be solved by calculating the remaining unknown after manipulating at the most two or three variables to eliminate unknowns.

At present, I am researching solu-



tions for simultaneous equations where the number of unknowns and equations do not correspond, and the so-called the least squares problem. In case of several tens of thousands of equations, the amount of calculation grows massive with the method of eliminating the unknown, but if we can gradually get close to the unknown by repeating a particular algorithm, we can find the solution with less calculation. The method is not empirical but must be underpinned with the correct mathematical support. In practical terms, it is applicable to predicting avalanches based on the status of snow observations, or for image processing that reconstructs the original image from observed data

Faisal: When you are in the middle of

Keiichi Morikuni Advisor: Professor Ken Hayami



Mahmudul Faisal Al Ameen Advisor: Professor Makoto Tatsuta



Sho Yamashita Advisor: Associate Professor Makoto Kanazawa

developing software, it is extremely difficult to verify whether or not you are really going in the direction of your goal. My research utilizes structured logical systems to try to verify in advance that partially programmed pieces will work according to expectations and accurately once they have been assembled into the complete software. At present, there are limitations to the class of programs that can be



verified, but I am aiming for a method that can be applied to more complex systems on a larger class. Even if you have succeeded in constructing a logical system, you need to prove that the system has integrity, and that is where the challenge lies.

The fun of giving shape to an idea

----When do you feel the fun in re-search?

Yamashita: Even though it has been called a social science in the past, there were many areas in linguistics that were not science. By using mathematics, it is possible to sound out areas that have been left untouched, and that is where appeal of research lies. Nonetheless, the framework for this kind of research has been gradually assembled for half a century or more with research advancing only very slowly, so it is a big challenge.

Morikuni: For me, the most interesting thing about research is when you try out one of those ideas that pop up and it works. In the real world, some arguments ask what use there is in studying mathematics, but it is the

> foundation for a lot of research, and it is great to feel that you are actually doing something useful.

> **Faisal:** In mathematics, to show a particular property or proving a new theorem. It' s like solving a game and it is very exciting.

> — Have you had a chance to present research results or to engage in external activities?

> **Morikuni:** I have been fortunate enough to have oppor-

tunities to make oral presentations at academic meetings and conferences in Japan and overseas, and to visit researchers and professors at other research institutes to undertake research and presentation. Since I have been able to get correct results with calculation times that are shorter than past techniques, I am presenting my work and I plan to bring it all together in my Master's dissertation as a comprehensive survey.

Yamashita: I have not yet presented at a full-scale academic meeting, only at NII Logic Seminar. My master's dissertation will be the first opportunity to bring together some authoritative results.

It wasn't an academic meeting, but I participated in a summer school in France last summer. Since Europe is comparatively active in researching formal grammar, it was very stimulating for me. I also felt a keen need to polish my English skills.

Faisal: With my other research fellows, I have presented my research results in an NII technical report. It was also accepted by SEFM (Software Engineering and Formal Methods), an international conference. Professor Tatsuta is the first author but if they give me the opportunity, it will be my first presentation.

Wanting to contribute as a researcher

— Please tell us about your future ambitions.

Yamashita: Formal grammar is a minor field but after completing my doctorate, I plan to study with the intention of becoming a researcher. The first big incentive was to want to find out how language exists in the brain. To get at the ultimate truth will probably be difficult in my own lifetime...

Morikuni: I think I would like to continue to study more, and to use programs while constantly adding new analyses. I hope that the results of my research will be accepted by many people for practical situations.

Faisal: After I have earned my doctorate, I will return to Bangladesh and become a university professor to teach the next generation. Compared to Europe, North American and Japan, the research infrastructure in Bangladesh is insufficient. It is my dream to prepare a more advanced research environment and to contribute to my own country in this field.

— Thank you for being here today. (Written by Asako Tsukasaki)

Structures of Information Processing Hidden in Games

Kunihiko Sadakane

Associate Professor, Principles of Informatics Research Division, NII

Imagine a game like this one. There are a lot of numbered boxes and balls in a room. You can put one ball in each box. At the start of the game, all the boxes are empty. Standing by separately, m players are

called into the room in random order. The player lifts the lid off one box and if there is a ball inside, he can remove it, if the box is empty, he can put a ball in it or he can do nothing at all. However, at the end he must without fail replace the lid. This action is repeated t number of times.

A victory of the game is decided as follows. If every player guesses whether or not he will be the last person to be called, they "win," but all of them "lose" if even one of them misses. Seemingly, it is a game that cannot be won, but

with some clever decisions about how to insert the balls, there is a way that you can win for sure. The players can confer in advance and decide how to put the balls in the boxes. In this regard, what can you do to make t number of times of opening the boxes as small as possible?

What can you do to be sure you will win?

If you know what number player you are, there is one method of sure victory. For this, you can use the boxes and balls to represent a counter. Namely, regarding a box containing a ball as 1, and a box containing no ball as 0, you can use the boxes and balls to represent a binary counter. Done like this, each player can look at the present counter value and raise it by 1. To increase the value of the counter by 1, open the box that corresponds to the lowest binary digit, and if there is no ball in it put one in or if there is, take it out. If there is a ball in the box, you go one digit up so all you need to do is to repeat the same with the next digit up. When the counter value is m, you know that you will be the last person. Since the counter value is between 1 and m, the binary can be expressed as log m bit and

you are sure to win if you open the box t=log m times.

This is seemingly the optimal method since log m bit is an absolute requirement for expressing the numbers from 1 to m. However, there is another method of certain victory with a smaller t. If you think carefully, in this game, it is enough for each player to know whether they are last or not. In short, there is no need to know exactly what number you are. In this regard, the method of using a counter is redundant, and a more suitable method suggests itself.

In this game, the number of boxes that each player can open is limited but how many boxes you use is irrelevant. If you think like this, you can actually reduce t. The method for putting the balls in the boxes is more complex, but t = 5 log log m is a sure method to win.

Quick use of information decides who wins and who loses

In this game, if we equate the boxes with computer memory where a box with a ball inside represents 1 and one with no ball inside represents 0, the number of boxes corresponds to the amount of memory (number of bits) needed to represent certain information (whether you are last or not). The number of times t of opening the boxes corresponds to the number of times needed to access the memory, or access time, in order to obtain the information. Being able to extract the information quickly ties in with a better way of winning. It looks just like a game but it runs deep.

NII Today No.32, January 2010 (This English language edition of NII Today corresponds to No.46 of the Japanese edition) Published by: National Institute of Informatics, Research Organization of Information and Systems Address: National Center of Sciences 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430 Chief editor: Yoh'ichi Tohkura Cover illustration: Makoto Komori Photography: Shuichi Yuri Design: Kotaro Suzuki Production: Sci-Tech Communications Inc. Contact: Publicity and Dissemination Team, Planning and Promotion Strategy Department TEL:+81-3-4212-2135 FAX:+81-3-4212-2150 e-mail: kouhou@nii.ac.jp/





Weaving Information into Knowledge