Faculty Members and Labs in Department of Computer Science

Professor   Masami Hagiya

The world is full of computation!
From Anshin-Anzen (reliable and safe) to Ayashii-Abunai (enigmatic and adventurous)

While our laboratory has been working on formal verification of software and protocol (Anshin-Anzen), we are more and more interested in computational processes hidden in natural phenomena and conducting research with the ultimate goal of constructing artificial information processing systems employing those phenomena (Ayashii-Abunai).

More concretely, with background in formal logic, our laboratory is seeking for new computational models and developing methods and tools for their analysis, verification and synthesis. In particular, we are investigating potential of computation not by conventional electronic computers but by natural phenomena including physical, chemical and biological ones (natural computation or unconventional computation), and conducting research on molecular computing, quantum computing, cellular computing, etc., and their applications (e.g., molecular computing to molecular robotics).

http://hagi.is.s.u-tokyo.ac.jp/members/hagiya.html   ← Hagiya's home page (in English)
http://hagi.is.s.u-tokyo.ac.jp/rigakuru.pdf   ← Introduction to the lab (in Japanese)

Recent research projects:
* Model checking distributed software   * Testing IoT protocol
* Verification of quantum protocol   * Analysis of quantum entanglement
* Simulation of biomolecular reactions   * Cellular Automata by a reaction-diffusion system

Our laboratory is also working on bioinformatics researches by a group centered on **Assist. Prof. Tsukasa Fukunaga**. We are studying various research topics from classical bioinformatics problems, such as genome sequence analysis and gene function prediction, to novel research fields such as the application of bioinformatics to ecology or ethology.

Professor   Hiroshi Imai

Our laboratory aims to introduce new computational models such as quantum computers with new quantum algorithms, to develop effective algorithms for fundamental and new research field such as huge network, game theory, and to analyze complexities to represent difficulty of computing in sophisticated problems. We undertake a variety of studies ranging from fundamental theories to applied fields. Presently the issues we study include:
(1) development of geometric algorithms in lower and higher dimensions; algorithmics to break the state-of-the-art barriers against fundamental important problems such as matrix multiplication and large graph problems.
(2) optimization over discrete system such as matroids, oriented matroids and their BDD as combinatorial model of geometric as well as algebraic structures.
(3) expansion into complexity theory, including Kolmogorov complexity, randomized computation and complexity over the reals; we also study time and space complexities on various computational models such as communication complexity and (quantum) interactive proof system.
(4) computer-scientific approach to recreational as well as economics games; we analyze algorithms and complexities on various subjects from GO to assignment problem in internet and a network communication.
(5) studies on quantum computing and quantum cryptography; in particular, we have been investigating the following topics, fundamental theory for realizing quantum computers, quantum algorithms, quantum complexity theory to analyze computational ability of quantum computers, quantum information theory for quantum cryptography, and theoretical analysis of quantum information such as quantum entanglement.
   http://www-imai.is.s.u-tokyo.ac.jp/

Professor Reiji Suda

Our laboratory researches on (1) parallel and high performance computing and (2) numerical algorithms, aiming to higher speed, higher precision and higher reliability in large scale scientific computations.

One target in parallel processing is *GPU computing*, where graphic processors are used for general purpose computing to get higher performance. Algorithms should be implemented so to fully exploit its parallelism, as GPU has hundreds or thousands of cores in it. We investigate also use of multiple (hundreds of) GPUs in parallel.

Another target is very large scale supercomputers. One of our approach for forthcoming supercomputers is *communication avoiding algorithms.* They are designed to reduce the amount of inter-processor communication, which will be a significant cost in a large scale parallel computer with several millions of processors.

Automatic performance tuning or *autotuning* is a mechanism that chooses the best implementation after several trial executions on the real machine. We are investigating mathematical methods to model the performance from the settings of tunable parameters, and to optimize tunable parameters efficiently, based on Bayesian statistics.

We have worked on *fast numerical algorithms* with lower computational complexities. With the increase of scale and complexity of numerical computations, double precision sometimes fails to provide enough precision, and algorithms of *high precision and/or high numerical stability* are needed.

http://sudalab.is.s.u-tokyo.ac.jp/~reiji/lab-e.html

Professor Naoki Kobayashi

Our research group is studying theoretical foundations for software and their applications such as program verification. On the one hand, an increasing number of important systems such as transportation systems, medical devices, and banking systems are now controlled by computer software, and a bug of such software can cause a serious disaster. On the other hand, such software is becoming more and more complex and larger, and it is difficult to maintain the quality of software by using traditional software engineering technologies like testing. In view of these situations, we aim to improve the reliability and efficiency of software by developing automated techniques for program verification and transformation based on rigorous mathematical methods. To achieve the goal, we also need to study and advance many research topics in theoretical computer science, such as type theory, formal languages and automata, and automated theorem proving. It is a pleasure of our research to find out that deep mathematical results, which initially seem to be only of theoretical interests, are actually quite useful for the practically-motivated research mentioned above.
See http://www-kb.is.s.u-tokyo.ac.jp for details.
Recent research topics include:
1. Higher-order model checking: This is an extension of traditional model checking that has been successfully applied to system verification. We have recently constructed the first higher-order model checker in the world.
2. Automated program verification: By applying the higher-order model checking mentioned above, we are developing fully-automated program verification tools for programming languages like ML and Java.
3. Data compression: String and tree data can be compressed in the form of programs that generate them. The higher-order model checking above can be used to transform such compressed data without decompression.
4. Type theory and program semantics: We are studying intersection type systems and game semantics as foundations for program verification and transformation.
5. Protocol verification: Cryptographic communication protocols are used, for example in Internet shopping, for safely exchanging confidential data. We are developing a method for automated verification of such protocols.

Professor Takeo Igarashi

(1) User Interface: We are working on user interfaces for information appliances ranging from personal computers, smartphone, robots, and self-driving automobiles. We not only develop techniques to improve efficiency but also propose new ways of interaction.
- User interfaces for machine learning and artificial intelligence. We focus on the preparation of training data and interactive learning.
- Interaction techniques for novel appliances such as smartphone, smart watches, AI speakers and smart glasses.
- Interaction techniques for giving directions and controlling real-world systems such as robots and self-driving automobiles.
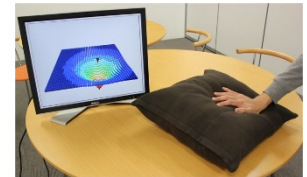
 (2) Graphics: We mainly work on interactive shape modeling. We work on content creation for digital media such as movies and games. We also work on interaction techniques for digital fabrication using 3D printers and laser cutters.
- Content creation such as 3D modeling and 2D animation using recent technologies such as sketching and machine learning. We also work on interaction techniques for medical imaging.
- Shape modeling for real world objects such as musical instruments, clothes, and toys leveraging real-time physical simulation.
- Novel techniques for 3D scanning and 3D printing to support personal fabrication.
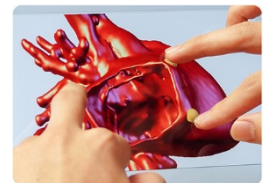
We can provide opportunity for students to collaborate with research groups in other countries and productions. http://www-ui.is.s.u-tokyo.ac.jp/

Associate Professor Noboru Kunihiro

Our laboratory is working on various research subjects concerning on cryptographic theory. Cryptographic technology is essential to secure communication through an insecure channel. To use cryptosystems securely, we must thoroughly verify whether they are secure. Based on security evaluation, proposing highly functional cryptographic encryption schemes and protocols is also an important research topic. Specifically, we deal with the following topics.
(1) Algorithms for solving mathematical problems that are the basis of security: Many cryptosystems are constructed based on mathematical problems which are difficult to solve. A prime factorization problem and a discrete logarithm problem are typical examples. To evaluate the security, understanding the accurate hardness of these problems and proposing effective algorithms are important research topics. Furthermore, we also study how the implementation of large-scale quantum computers will affect cryptographic technology.
(2) Security evaluation under the condition that partial information is leaked: Cryptographic schemes that are theoretically secure can be broken because of improper implementation or by physical observation. For example, it is known that when an error occurs randomly in the secret key of the RSA cryptosystem, the whole key can be recovered by employing the key redundancy. We use complexity theory and number theoretic algorithms as tools for security evaluation. In addition to theoretical analysis, security is evaluated by numerical experiments.
(3) Proposals of highly functional encryption schemes and protocols: Highly functional encryption schemes and protocols become essential toward highly information-oriented society. Homomorphic encryption and searchable encryption are such examples. We aim to propose efficient schemes for application to the real world. We also try to propose secure protocols that satisfy social requirements by combining cryptographic schemes.

Associate Professor   Yoshihide Yoshimoto

　One of the major motivations to invent electronic computers was the application to science and technology. After the invention, the performance of the computers improved dramatically with the exponential development of semiconductor technology: Moore's law. Computational science, which advances science with computation has benefited greatly from the development. Nowadays, however, because the limitation of semiconductor technology is coming up to the surface, the complexity of computer systems such as parallelization etc. is so increasing that both computational science and computer science have to cooperate once again for the further progress. This laboratory was set up under the above background in August 2014 to perform education and research which connect computational science and computer science. (http://www.cp.is.s.u-tokyo.ac.jp/)

　Yoshimoto himself is specialized in solid-state physics which elucidates the properties of materials such as semiconductors, metals, dielectrics, and magnets. Especially, he is specialized in the first-principles electronic structure calculations which accurately simulates the quantum mechanics of electrons which dominates most of the properties of materials. He has developed and open to the public a program package xTAPP for this purpose. The right-hand side figures are examples of visualization of an electronic structure calculation. (http://xtapp.cp.is.s.u-tokyo.ac.jp/)
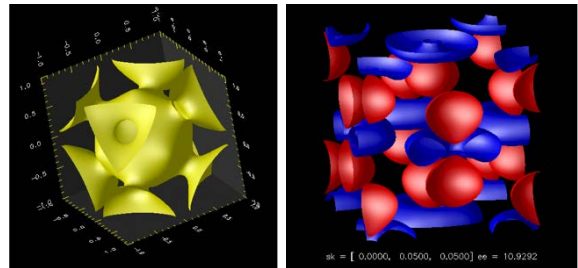
　From the electronic structure calculation as a hometown, the aim of the laboratory set up between computational science and computer science is as follows:
1. Perform mutual exchange between wide range of fields in computational science by re-interpretation of the methods developed individually in each field from viewpoints of computer science.
2. Understand needs of computational science from viewpoints of computer science and propose more essential solutions.

Associate Professor   Shinpei Kato

Computing Platform Lab. (PFLab)

Our research interest includes computing platforms facilitating disruptive innovations. PFLab was founded in April 2016. Topics of special interest includes, but not limited to, the following:
➢ OS kernels for more than hundreds of cores on a chip.
➢ Real-time 3D data processing for peta-scale real-world data.
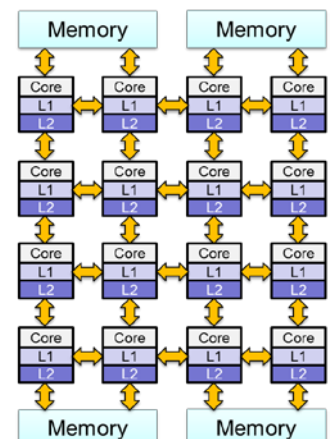➢ System on a chip technology for mobile robots and AI.
10x performance and 0.01x power – this is what we want to achieve by our computing platform. We also study artificial intelligence (AI), robotics, sensing, and cloud computing for emerging real-world applications, such as autonomous vehicles and high-precision 3D maps. PFLab is in collaboration with foreign top-tier schools and automotive/electronics/technology/startup companies.
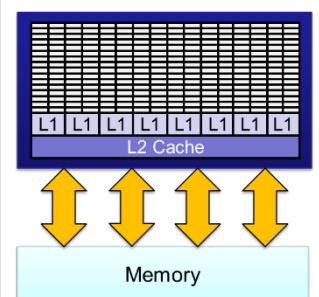
NoC-based many cores
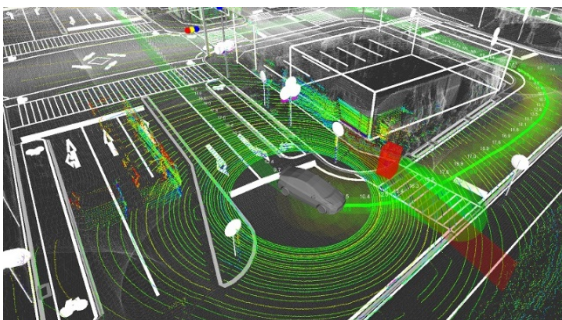
3D data processing
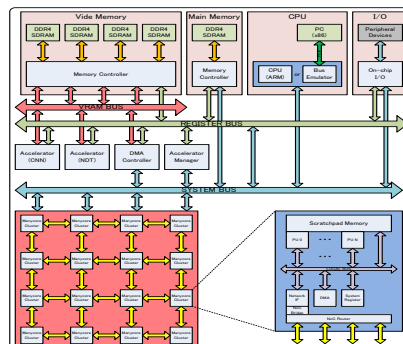
System on a chip for AI

GPU-featured many cores

Professor   Masashi Sugiyama   & Lecturer   Junya Honda

   With the dramatic performance improvement of information and communication technology, intelligent information processing that was done only by humans is becoming possible also by computers. Under the theme of "how intelligent can computers be?", Our laboratory is working on various research topics related to intelligent data analysis, called **machine learning**, in the field of artificial intelligence.

(1) **Construction of Learning Theory**

   Generalization is the ability to cope with unknown situations, and is indispensable for computers to behave intelligently. We are theoretically investigating the mechanism of acquiring the generalization capability based mainly on probability and statistics.

(2) **Development of Learning Algorithms**

   Machine learning involves various subjects such as supervised learning (learning from input-output paired data), unsupervised learning (learning from input-only data), and reinforcement learning (learning through interaction with an environment). We are developing practical and theoretically motivated machine learning algorithms.

(3) **Application of Machine Learning Technologies to Real-World**

  Growth and spread of the Internet and sensor technologies allow us to collect a huge amount of data in engineering and fundamental sciences such as documents, audio, images, movies, e-commerce, electric power, medicine, and biology. We are collaborating with industry partners and applying state-of-the-art machine learning technologies to solving real-world challenging problems.

http://www.ms.k.u-tokyo.ac.jp/

Lecturer   Issei Sato
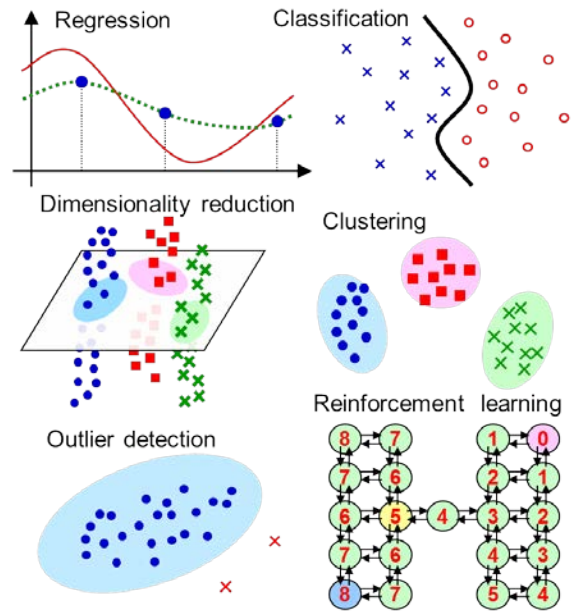
"Heading for statistical machine learning as social infrastructure"

Statistical machine learning is a technology to extract automatically the rules, which enable a computer to perform intelligent process, from a large quantity of data. Machine learning is playing an important roles on various scenes in real society, such as face recognition system, which almost all the smart phones of modern models equipped with, recommendation system in on-line shopping sites, car and vehicle automatic driving system to gather attention recently. Our laboratory is working on the research topics as follows.

(1) **Mathematical modeling**   It is required in statistical machine learning to construct mathematical models corresponding to the properties of data and problems to be solved.   Our laboratory is working on the researches of statistical modeling with a latent variable that renders hidden properties of data. For example, when recommending items according to user's preference through purchase history, user's preference is a hidden property, because it does not appear explicitly. In such cases, user's preference can be represented as latent variable in statistical modeling.

(2) **Learning algorithm**   After determining a mathematical model, we have to estimate the parameters of the model from data. We study learning algorithms from a large amount of data based on Bayesian estimation and stochastic optimization.

(3) **Machine learning to support scientific experiments**   Researchers repeat the try-and-error of experimental design, analysis of test results and setting experiments. If this process can be supported or automated by a computer, it must be a great contribution to the progress of the overall fields of sciences. We study a machine learning technology to construct the environment where experimental design is conducted by researchers and a computer takes try and error of experiment.

(4) **Application in social context**   We apply machine learning to social problems. For example, we have developed a computer-assisted system for medical image diagnosis under joint research with the university of Tokyo hospital.

Professor Kenta Nakai & Lecturer Ashwini Patil
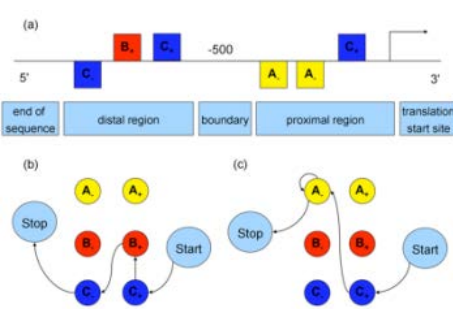
Our laboratory focuses on bioinformatics, like three other labs in the Shirokanedai campus, where a huge super-computer system is available. Prof. Nakai's fundamental research interest is to elucidate how the biological information is encoded as 1-dimensional character strings (sequences). It is well established that such 'information for life' is basically encoded as the chemical structure of a set of DNA molecules, called the genome, which is kept within each cell. Its necessary part is copied into another kind of molecules, RNAs, which are then used to synthesize various proteins for various purposes. Thus, there can be three levels for our study (DNA, RNA, and protein). Currently, we are mainly interested in understanding how the regulatory information for reading only necessary parts is encoded in the genome DNA. The attached figure shows our model of the structure of regulatory regions deduced from a set of genes that are read out in the same cells. The model is probabilistically represented using the Markov chain. Once such a model is made, we can predict unknown genes that are regulated in the same manner and such predictions would be verified experimentally.

Although we belong to the computer science department, our research projects are rather biology-oriented. Nevertheless, students with any background would be welcome if they are interested in bioinformatics. In fact, another feature of our laboratory is that the majority of the members are international students/staffs. In principle, we would like to support each student's spontaneous motivation, and each research project will thus proceed independently. We will, however, try to ensure that these projects will become stimulating for each other. Please visit http://fais.hgc.jp for details.

Professor Seiya Imoto

Currently, the cost of whole-genome sequencing analysis for an individual runs about thousand USD. This cost should be decreased to hundred USD within several years. At that time, it is obvious that almost all of us can have own whole-genome sequence information. Under this perspective, it is necessary to develop methods that can use genomic big data of several millions people's whole-genome sequence information for the prediction and prevention of diseases and improving our health. Based on genomic big data including whole-genome, transcriptome, epigenome and meta-genome, and time-series information of health and medical records, we address this problem by utilizing statistical data analysis technologies and supercomputing under the collaboration with researchers in companies and universities.

We are pursuing the following researches (but not limited):

Research theme 1: Development of sequence data analysis methods

We develop Bayesian models for analyzing genomic regions related to immune response including human leukocyte antigen (HLA), T-cell receptor repertoire and meta-genome, and for analyzing cancer genome to detect somatic mutations more accurately. Especially, in the International Cancer Genome Consortium, we use whole-genome sequence data of about 3,000 cancer patients to investigate the relationship between genomic mutations and immune system.

Research theme 2: Development of mathematical modeling techniques for systems biology

Probabilistic graphical model that can handle several tens of thousands molecules (genes and non-coding RNAs) are developed for predicting molecular networks. Using the network information, we develop statistical methods that can predict the efficacy of anti-cancer drug for each of cancer patients.

Research theme 3: Clinical sequence

From the analysis of cancer genome, we will detect genomic mutations of several thousands to tens of thousands. We next detect causal genomic mutations for cancer from the list of genomic mutations. So, we need to interpret the genomic information. Although the interpretation of genomic information is based on the published literature, if we focus on cancer research, several hundreds of thousand research papers are published per year. It is obvious that comprehensive utilization of such a huge amount of information is beyond human capacity. We investigate to utilize artificial intelligence to solve this problem.

Associate Professor Tetsuo Shibuya

The whole human genome was completely sequenced in 2003, after extraordinary efforts of many researchers under tremendous cost. On the other hand, the DNA sequences are different individual by individual, and the difference is known to effect disease susceptibility, drug sensitivity, etc. But revealing such unknown variations was very difficult due to costs and time constraints until very recently. The situation has changed in the last few years by developments of "next generation sequencers" or "next-next generation sequencers" which dramatically reduce the costs and time for sequencing. They enable us to sequence individuals' whole genomes, a genome of each cell, genomes of various organisms, etc. Furthermore, we can comprehensively obtain additional data such as epigenetic information. But the amount of obtained data in these research is very huge. They increase sometimes in a pace faster than the Moore's law. Moreover, these data require consideration of various additional factors like privacy when analyzed. They make it very difficult to apply traditional analytic methods to them. We need a paradigm shift of information science for bio-medical research in the big-data era.

We aim to make impacts on bio-medical sciences through development of information-scientific technologies, *e.g.*, research on high-speed and/or highly functional searching/analyzing technologies for bio-medical big data. We also aim to develop big-data technologies not restricted to bio-medical data, as such technologies are required everywhere these days. Our research field is very interdisciplinary. We aim to contribute not only to the basic research of related fields, but also to the state-of-the-art applications of these fields.

Home page: http://shibuyalab.hgc.jp

Associate Professor Rui Yamaguchi

The aim of our laboratory is to "hack" cancer systems and to "control" them based on cancer systems data such as genomes, cancer genomes, epigenomes, transcriptomes. By the advancement of sequencing technology, we can now sequence and analyze individual genomes (variations), cancer genomes (mutations), epigenomes (modifications), and transcriptome (gene expressions). The speed of technology development for sequencing is very rapid.

It is increasingly clear that cancer is a very complex disease that occurs from accumulation of multiple genetic and epigenetic changes in individuals who carry different genetic backgrounds and have suffered from distinct carcinogen exposures. Therefore, elucidation of cancer biology as a system appears to be extremely important to advance knowledge of the disease, not only from the viewpoint of basic research, but also for future developments of novel diagnostic and therapeutic methods. With mathematical modeling and data analysis using supercomputer, detailed investigations of cancer using a "systems biology" approach have made it possible to gain insight into the molecular mechanisms of system dysregulation in a yet unexplored scale that occur in cancer cells.
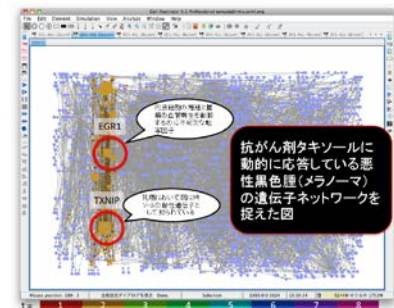
Research theme 1: Integrative systems understanding of cancer for diagnosis, therapy and prevention
Research theme 2: Mathematical statistics and applications to biomedical research
Research theme 3: Mathematical modeling and simulation for personalized medicine

Lab page: http://dnagarden.hgc.jp   Systems Cancer Project: http://neosystemscancer.hgc.jp
Priority Issue on Post-K computer: http://en.postk.hgc.jp

Professor Toshiyuki Nakata

Applying Cyber Physical systems (IoT) to real world social problems is expected to solve a number of serious hazards such as aging society, food shortage, energy crisis, and decrease in the labor population. On the other hand, current cyber physical systems are usually applied to relatively small problems and have not been applied to large scale real world problems.

Our research focuses on large data processing technologies which can be the basis for building large scale high performance computing systems which would make social scale cyber physical systems a reality. Some of the potential research themes are shown below.

- ➤ Research on efficient methods for big data processing. This research focuses on big data processing using large scale big data processing platforms methods such as Hadoop / Spark. Efficient processing using parallel/processing technologies, GPGPU's and fast I/O technologies.
- ➤ Research on efficient processing of AI / Machine Learning using the above platforms.
- ➤ Research on architecture and methodologies for efficient cooperation between the cloud systems and the edge nodes.
- ➤ Research on linking Open Data processing and big data processing
- ➤ Research on real world modelling and simulation
- ➤ Research on big data processing using real world data in collaboration with other laboratories
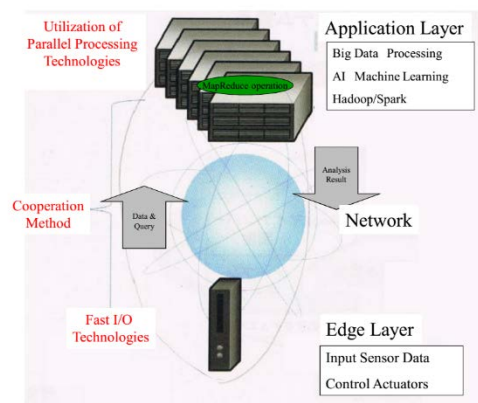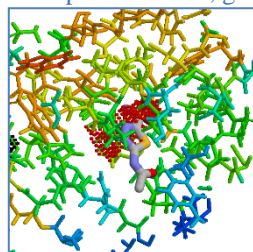
Professor Kentaro Shimizu

We conduct extensive bioinformatics studies at different levels ranging from protein structure and function analysis to genome and transcriptome analysis using a wide variety of methods, such as machine learning, statistical analysis, and molecular simulation. Below are some of the areas of our research:
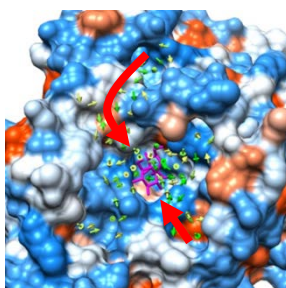
- High-resolution protein structure prediction
- Prediction of binding sites between proteins and other molecules (proteins, nucleic acids, sugar chains, lipids, metals, low-molecular-weight compounds, etc.)
- Docking prediction (prediction of complex structures in the bound state)
- Various functional predictions

In addition, by analyzing the dynamics of proteins, we conduct studies in the following areas: process of binding of molecules and other physicochemical properties such as free energy, interatomic interactions in vivo, and simulation of the process of protein folding. Recently, we are involved in research associated with the analysis and prediction of various structural features of intrinsically disordered regions and membrane proteins, which are important for protein function. In collaborative investigations with experimental researchers, we are engaged in agricultural bio research, which is related to medicine, food, and environment, such as enzyme modification, drug development, functional foods, and environmental purification. http://www.bi.a.u-tokyo.ac.jp/
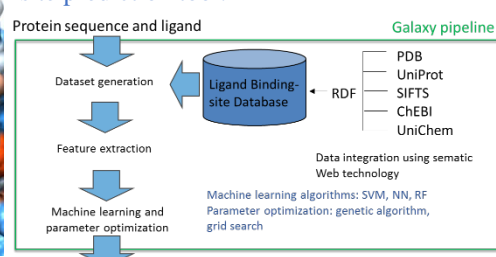
Protein–ligand-binding site prediction (red color is the predicted site, gray denotes the actual ligand)

Automatic generation of a ligand-binding site prediction tool.

Coarse-grained molecular dynamics simulation of protein–ligand-binding (Analysis of the movement of ligands into a pocket along the protein surface)

Professor　Akihiko Takano

(1) Algebra of Programming: To establish a scientific method for program construction, the rich collection of software parts with correctness proofs and the flexible way to combine those parts into working software with required function and efficiency are crucial.　Algebraic foundation of program fusion, partial evaluation, and data transformation are studied in this perspective.

(2) Informatics of Association: How to vitalize our association using vast collections of digital information is investigated. Association computation is formulated as a computational basis for efficient evaluation of similarity among documents or words. GETA is a scalable and efficient implementation of this computation. The GUIs for creative interaction with digital information based on association are studied. Here are some examples of the information services we launched as our research products.

- Webcat Plus (http://webcatplus.nii.ac.jp/)
- Cultural Heritage Online (http://bunka.nii.ac.jp/)
- Shinsho-Map (http://shinshomap.info/)
- IMAGINE Book Search (http://imagine.bookmap.info/)

Professor　Akiko Aizawa

Our laboratory's research is focused on natural language technologies to assist human intelligent activities. Our major challenges include the following subjects in text and media studies that are based on machine learning including deep learning, statistical modeling and analysis, or annotation and corpus analysis.

(1) Text mining: document structure analysis and information analysis; information identification and entity disambiguation; automatic construction of linguistic resources

(2) Machine reading comprehension: Exploiting the semantic structure of natural language; extracting knowledge and information from natural language

(3) Modeling human language activities :Measuring and analyzing language activities in text media

Our recent research topics are as follows: analysis of reading comprehension tasks, sentence compression and text summarization, analysis and utilization of document structure, retrieval and understanding of mathematical expressions, analyzing the behavior of readers by using gaze tracking methods, real-scale information identification, and information recommendation and English writing assistance for academic researchers. We also welcome new research topic proposals related to natural language processing and information retrieval.

Our research facilities include variety of information resources and a large-scale computation platform at NII. Students are encouraged to act as independent researchers/engineers by being allowed to join seminars and discussion groups with interdisciplinary and international researchers and to participate in related joint research activities at NII.

http://www-al.nii.ac.jp/en/

Professor Yusuke Miyao

Natural Language Processing, Computational Linguistics, and Human Cognition

We aim the goal of making computers understand natural language. People communicate without much thinking about the process of choosing words. To make computers do the same, however, we have to know clearly how natural languages work. That is to say, our research is to unfold the mechanism of natural language, which will eventually lead to the understanding of how people recognize the world. We study step by step how people use words and go on living as people learn the movements of the universe and works of life.

Specifically, we perform research on fundamental theories of natural language and language processing technologies, such as syntactic/semantic parsing and recognizing textual entailment, as well as their applications, including grounding language into images and non-verbal data, text data analysis over specialized domains such as financial or academic technical papers, machine translation, question answering and dialogue systems.

Refer to our homepage for the details of the research: https://mynlp.github.io/

Syntactic/semantic parsing using
Combinatory Categorial Grammar