

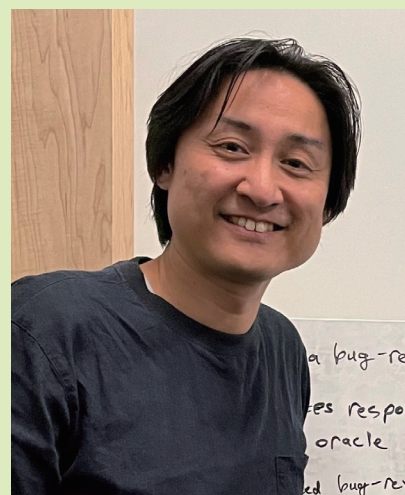
Software Digital Twins for Reducing Software Waste and Achieving a Sustainable IT Society

Shinobu Saito
Distinguished Researcher,
NTT Computer and Data Science
Laboratories

Abstract

The global information and technology services market for 2022 was reported to be 1.3 trillion U.S. dollars (about 150 trillion yen), and it's been pointed out that the massive amount of software produced by this market demand, while bringing great value to our daily lives, has also led to a variety of problems. There are expectations that an approach that not only seeks to increase the amount of software (production) but to also reduce that amount (reduction) could help solve those problems. In this interview, we asked NTT Distinguished Researcher Shinobu Saito about his research on “Software Digital Twins.”

Keywords: software waste, Software Digital Twins, Software 3Rs



Achieving an ideal environment for users and developers through Software 3Rs (Reduce, Reuse, and Recycle)

—Dr. Saito, what kind of technology is “Software Digital Twins”?

“Software Digital Twins” that I am now researching is a technology that reproduces the internal conditions of software in a digital world on the computer to discover and predict locations (programs) where software should be revised or deleted. It's been pointed out that the increase in software that is not used at all called “software waste” is a major problem that has come to encircle software in recent years. As it turns out, software, which was originally produced with the aim of making the world a better place, is

also giving rise to a variety of problems. These problems are called “big code” or “software bloat.” The former means that software that has become too big is difficult to maintain by a limited number of people (developers). In this regard, it is predicted that the amount of software will eventually exceed the range in which programmers are capable of managing it. In overseas surveys, more than half of all programmers said “The amount of program code that one must deal with has increased by more than 100 times compared with that 10 years earlier.” The latter, meanwhile, means that software that has become overly abundant (bloated) is consuming system resources such as central processing units and memory on a large scale. This results in excessive power consumption and carbon dioxide emissions, which has a negative impact on the global environment.

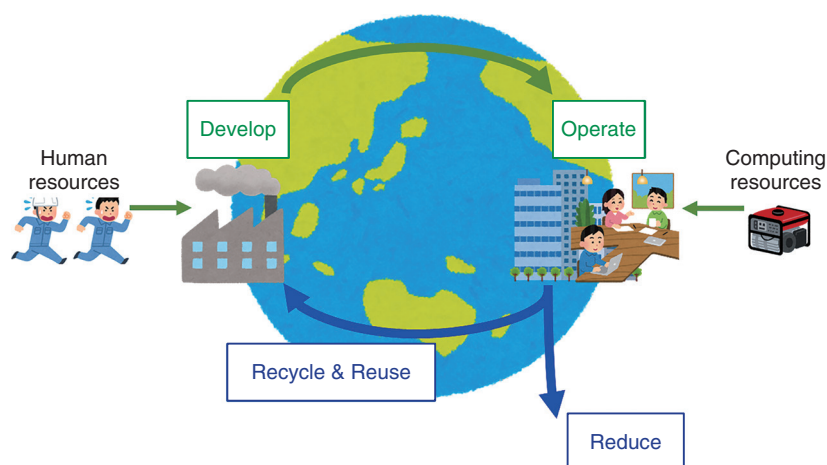


Fig. 1. Vision of Software 3Rs.

After joining NTT DATA Corporation, I myself thought for many years “What can be done to make it easier for developers to create good software for users?” Several years ago, however, I came across two articles in a business magazine titled “Software not used by companies and ignored by users is waste” and “One-third of an average company’s information technology (IT) budget is consumed as software waste,” both of which shocked me. Indeed, it has also been reported that more than half of the functions in a developed system are rarely used. I therefore began to feel “There is the possibility that my research in the end will actually have a negative impact on people and the Earth,” and I began to think of an approach that could break through this situation.

Then, one day while thinking about this, when I was helping my child in elementary school with summer-vacation homework, I noticed that the concept of the 3Rs in relation to environmental problems was introduced in the textbook. The 3Rs refer to Reduce, Reuse, and Recycle. Seeing this, I thought “Shouldn’t this concept also be applicable to IT society that is constantly producing software?” I considered that, from here on, “developing” software that can enrich everyone’s life is necessary, but it is also important for people and the environment to “eliminate” software that has been overly developed and become excessive. The concept born out of these thoughts is “Software 3Rs” (Fig. 1), which promotes not only the development of software but also the reduction of unnecessary software and the reusing of existing software. In this way, I believe that we can achieve a sustainable IT society for both users and

developers of software. Last year, I presented this concept at an international conference on software engineering and received comments like “very interesting” from participating researchers.

—In particular, what kind of approach are you researching?

I am currently researching Software Digital Twins as an approach to implementing the Software 3Rs concept. This idea came to me from research on digital twins that reproduce data collected from real-world objects such as cities and human bodies in a digital world on the computer. For example, since it is difficult to directly observe an actually functioning brain or heart, research is progressing on using the digital twin of a human body to express such organs in digital space with the aim of investigating the causes of disease, predicting the occurrences of disease, and even preventing disease. Software, as well, consists of an aggregation of programs, and directly observing the internal conditions of working software is said to be a difficult objective. As a result, correctly grasping “where the frequently used programs are” and conversely “where unused programs are” can be extremely difficult. Consequently, my aim with Software Digital Twins is to discover and predict locations (programs) that are unnecessary and non-urgent or in need of improvement by reproducing with high accuracy the internal conditions of “black-box” software and making that content “fully visible” (Fig. 2).

This research is closely related to Digital Twin

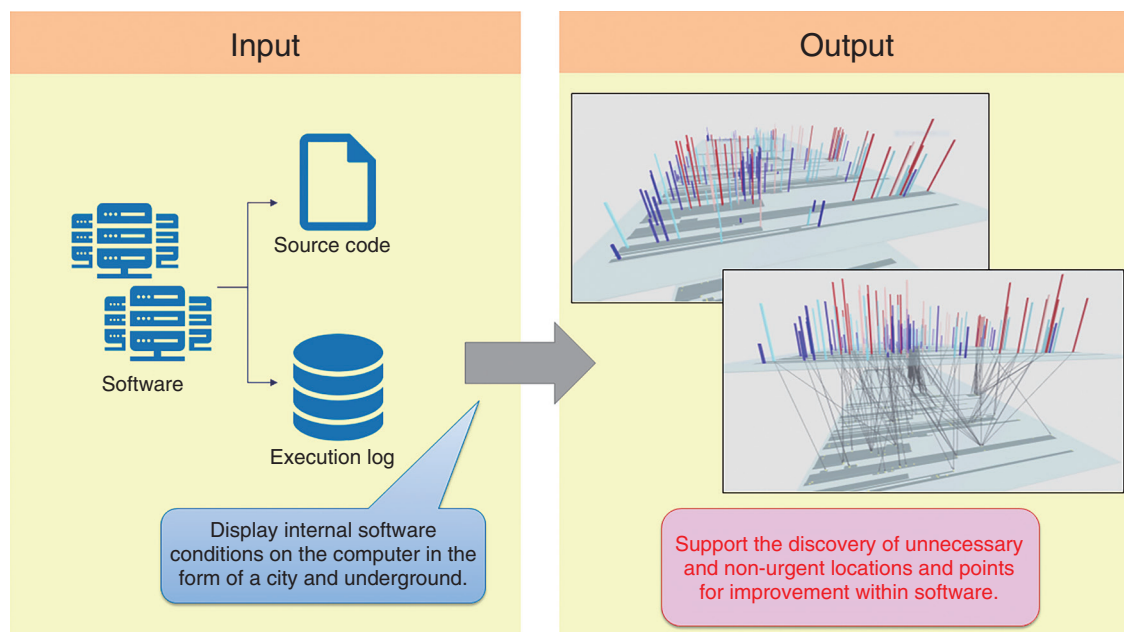


Fig. 2. Overview of Software Digital Twins.

Computing (DTC), a major pillar of the Innovative Optical and Wireless Network (IOWN) concept proposed by NTT. At present, studies of DTC are proceeding in areas targeting the real world such as urban management and medical treatment. In contrast to this approach that attempts to reproduce the real world in the cyber world, my research attempts to reproduce a cyber world in another cyber world, that is, to reproduce the interior of software running on a computer on another computer. Here, reproducing a cyber entity as another cyber entity may appear at first glance to be somewhat strange, but I believe that this approach can enable humans such as software architects to accurately understand the internal execution conditions of software for the first time and facilitate predictions and simulations.

Main data used for accurately reproducing the execution conditions of software are of two types: “source code” and “execution log” that records program operation. In the latter, the collection of the execution log presents a major hurdle to making advances in research and achieving practical implementation. The ideal method would be to record the operation of “all locations (programs)” inside software during “all periods” in which the software is executed. This, however, would require that software engineers create a new program to record that operation at all locations. In addition, attempting to collect

the execution log in all periods of execution would consume an excessive amount of computing resources, which in itself would have a negative impact on people and the environment, thereby defeating our original purpose.

Given the above, the technique that I am studying in my current research will estimate and predict execution conditions throughout the software from execution log data obtained in limited periods at limited locations within the software. My aim here is to establish a technique that applies fields such as statistical analysis and machine learning through joint research with universities.

Continue with “solitary” and “interaction” activities to ensure that research does not end with basic research

—Please tell us about your research vision going forward.

One research achievement in Software Digital Twins to come out so far is ProcessCity, which is technology for visualizing the usage frequency of each function in a business system and the operation of the entire system in a three-dimensional manner. We have already provided this technology to several NTT operating companies, and at NTT R&D Forum

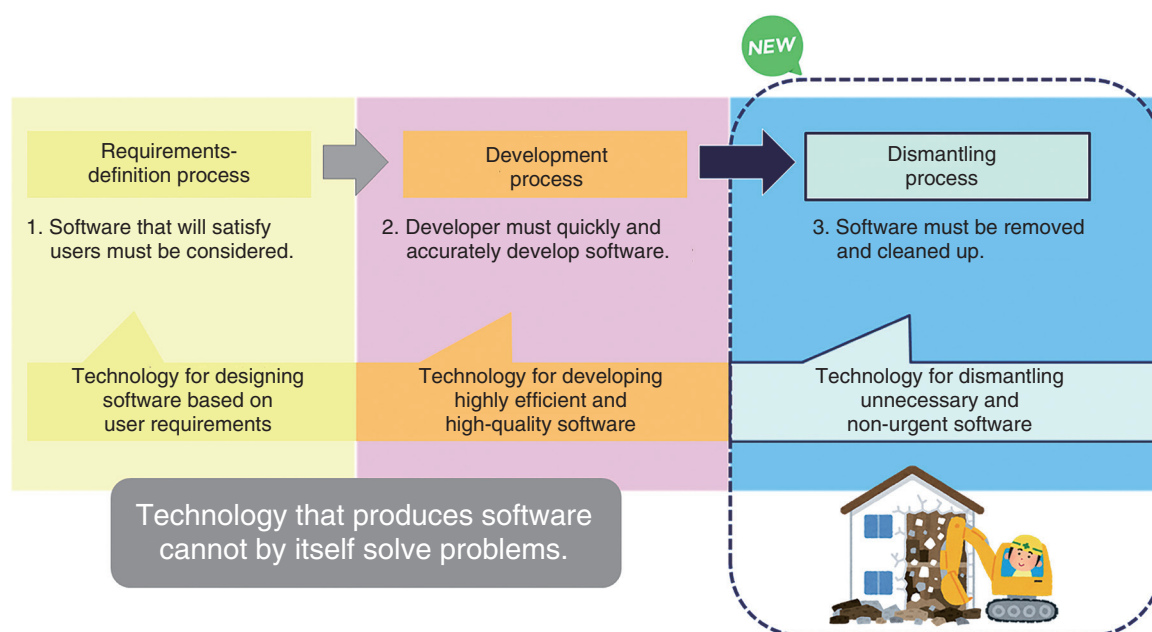


Fig. 3. Dismantling process in software development.

2022, we held a joint exhibit and a press release was issued with one of those companies. I also published a paper with members of another operating company to which this technology had been provided, and we were honored to receive a best paper award at the largest software-engineering international conference in the Asia-Pacific region held in 2022.

At present, it is said that the cost of maintaining aging and complex old systems called “legacy systems” has risen to 60–70% of a company’s IT budget. In addition, there have been cases in which surveys conducted for discovering system functions that could be scaled down or discontinued took months to complete. Consequently, the first thing that could be achieved by using the results of this research is a shortening of the survey period and improvement in the accuracy of the survey itself thereby reducing the labor required in the survey step by 20–30% and eventually to as much as 10% of the current amount.

It was reported that the global IT services market for 2022 was 1.3 trillion U.S. dollars (about 150 trillion yen), and according to the business magazine that I mentioned earlier, it’s been calculated that about one-third of those IT services, or 50 trillion yen, is being wasted. Of course, it would be difficult for NTT by itself to reduce such software waste throughout the world, but I would nevertheless like to create technology that could help reduce such a huge

amount of waste on the scale of the national budget and to connect this effort to providing great value to NTT’s corporate customers.

Additionally, on looking at the IT services industry going forward, I anticipate that an industry for the demolition of software in a safe and secure manner will probably come to be formed. In this sense, if producing software is a “construction industry,” then removing (deleting) software is a “demolition industry.” By making further progress in my research, my ultimate goal is to form a new industry and create new business opportunities and to make the software dismantling business one of NTT’s core business areas (**Fig. 3**).

—What is your impression of NTT laboratories?

As the name implies, NTT Computer and Data Science Laboratories that I belong to conducts research related to computers and data, and in this capacity, it is involved in a wide variety of research across all sorts of businesses and industries. NTT Group companies include both companies that develop software and companies that use software, which is why I believe that NTT laboratories are especially strong in both user and development fields. The scope of stakeholders is broad, and I am given opportunities for exchanging opinions with group companies from the

standpoint of using software, so I think I am blessed to be in such an environment that helps me make real progress in my research.

Moreover, on meeting with people from companies outside the NTT Group, there are many occasions in which I can feel the name value of NTT. For example, when attending an international conference for the first time 10 years ago, I knew nobody there, but on walking by myself to the venue after receiving my name tag at the reception desk, someone walking next to me noticed my name tag and said in a very friendly manner: “Ah, NTT! Do you know Mr. XX?” This person is a very famous professor in the software engineering community, and thanks to this chance happening, I still talk with him and discuss research whenever we meet up at an international conference. Of course, the name value of NTT does not guarantee that research will always go well, but it certainly gives me a reassuring feeling whenever I go to an unfamiliar environment or place.

—Dr. Saito, please leave us with a message for researchers, students, and business partners.

As I conduct my research at NTT, I have made the words “there is no research that ends with basic research” into a guiding principle. I first learned about this concept on joining NTT DATA from researchers that had come to my department from NTT laboratories. What this meant was that “Even basic research that seems unlikely to have an immediate application must be meaningful in the sense of being useful in some way. If it has no meaning (if no meaning can be discussed), it is not basic research,

and in fact, it is not research in the first place.” I keep these words in mind all the time, and since research is funded, I make an effort to present “the meaning of research” in terms of how those research results will someday be useful.

At around the same time, I remember well the words of a senior colleague, who said, “Research, in the end, is a solitary endeavor.” Research requires originality, so in some respects, I think “solitary” is correct here. Looking back at my own research when I was at the stage of thinking about a certain basic concept, there were many times when I could not verbalize it well and just kept worrying about it by myself. However, at the stage of actually implementing that concept, I found that it was necessary to gain the cooperation of many people. NTT provides an environment and the opportunities to interact with other people and other companies, so there are many uplifting (exciting) times that occur when acquiring new knowledge from outside one’s field and seeing a new world that one could not reach by oneself. Going forward, I plan to pursue my research with great feeling and conviction and to make up for what I am lacking by interacting with many of my fellow researchers with the hope that we can conduct research together.

■ Interviewee profile

Shinobu Saito received his M.S. degree from Keio University in 2001 and joined NTT DATA Corporation in the same year. He received his Ph.D. degree (engineering) from Keio University in 2007. He is engaged in the research of new software engineering for the digital era. He is a recipient of the 2022 IPSJ-CS Outstanding Achievement and Contribution Award and the APSEC2022 Best SEIP (Software Engineering in Practice) Paper Award among other awards. He was a visiting researcher at the Institute for Software Research (ISR), University of California at Irvine, from 2016 to 2018 and was a distinguished researcher at NTT Software Innovation Center from 2018 to 2021. He has been affiliated with NTT Computer and Data Science Laboratories since 2021.

