On Implementing Relational Databases on DNA strands

István Katsányi

In the last decade molecular biology has become the fastest growing discipline in the world. Some of the results are widely known, let us only mention the major breakthroughs in the Human Genome Project and nanotechnology. The progress made possible the birth of a new branch of science, that is called molecular computing (or DNA computing). Leonard M. Adleman published a paper [1] in 1994, which later become the foundation–stone of this new subject. In his article Adleman demonstrates how can a classic NP–complete problem: the problem of searching for a Hamiltonian path in a directed graph can be solved in polynomial time using the techniques of molecular biology and DNA strands. He outlines the great opportunity in the large computing power, and the extremely compact data storage. In a test tube there can be performed as much as 10^{16} operations in a second. That is much more than current supercomputers can execute. In a litre of water the DNA strands can encode 10^8 terabytes, and we can perform associative searches on the data in constant time.

In the past years many papers dealing with the computing power of DNA were published. However, only a few article studied the possibility of data storage and processing (see e.g. [2], [3]). Recently two papers ([4], [5]) described methods that yielded in an operation that closely resembles to the join operation of relational algebra. In spite of this no one has extensively studied the potentialities of the usage of molecular computing in the field of RA.

This work describes the bases of the implementation of relational databases in test tubes, using an abstract model of molecular computing. It specifies the representation of columns, tuples, tables, and databases, and the execution program of the relational operations. We examine the efficiency of each operation and compare them to traditional methods. We investigate the possibilities of practical usage of the proposed model as well as the bounds of it.

Keywords: molecular computing, theory of computing, relational database

References

[1] Leonard M. Adleman. Molecular computation of solutions to combinatorial problems. Science, 266:1021–1024, November 11, 1994.

[2] Eric B. Baum. Building an associative memory vastly larger than the brain. Science, 268:583–585, April 28, 1995.

[3] John H. Reif, T. H. LaBean, M. Pirrug, V. S. Rana, B. Guo, C. Kingsford, and G. S. Wickham. Experimental construction of a very large scale DNA database with associative search capability. In DNA Computing, 7th international Workshop on DNA-Based Computers, DNA 2001, Tampa, U.S.A., 10–13 June 2001, pages 241–250. University of South Florida, 2001.

[4] John H. Reif. Parallel molecular computation: Models and simulations. Algorithmica, 1998. Special issue on Computational Biology.

[5] Masanori Arita, Masami Hagiya, and Akira Suyama. Joining and rotating data with molecules. In IEEE International Conference on Evolutionary Computation, April 13–16, 1997.