ViewDesigner: A Tool of Designing Views on Data in Discovery System HypothesisCreator

Osamu Maruyama[1]
maruyama@ims.u-tokyo.ac.jp

Yoshinori Tamada[2]
tamada@hunter.ss.u-tokai.ac.jp

Satoshi Matsumoto[2]
matumoto@ss.u-tokai.ac.jp

Satoru Miyano[1]
miyano@ims.u-tokyo.ac.jp

[1] Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan
[2] Department of Mathematical Sciences, Tokai University, 1117 Kitakaname, Hiratsuka, Kanagawa 259-1292, Japan

Keywords: view design, viewscope, hypothesis creation, discovery system, sequence analysis

1 HypothesisCreator

Knowledge discovery from complete genome data demands more discovery-oriented computational methods than homology search or keyword search. Currently, over 26 complete genomes have been determined, and 104 prokaryotic and 32 eukaryotic genome sequencing projects are in progress [3]. However, software tools to strongly support the discovery process of genomic researches have not established yet.

HypothesisCreator is a multi-strategy discovery-oriented knowledge discovery system, which allows users (i.e., domain experts) to design views on data [1, 2]. A view on data means a particular way to interpret the data. In the scientific literature, devising new view on data is a key to discovery. To deal with views in a systematic way, the concept of "viewscope" is formulated in [1, 2]. Informally, a *viewscope* V is defined as a pair of a polynomial-time algorithm ψ, called the *interpreter*, and a set P whose elements are called *viewpoints* of V. For data x and a viewpoint $p \in P$, the interpreter ψ returns a value $\psi(x, p)$ as an evaluation of the data x with respect to ψ and p. A *view* on data is a viewscope $V = (\psi, P)$ with a fixed viewpoint, i.e., $|P| = 1$.

One of the features of HypothesisCreator is the automatic generation of viewscopes in the process of search for good views along with hypothesis generation. In the current version of the system, clustering programs and a decision tree generator are available as hypothesis generators. Another is to realize human intervention in the discovery process by designing viewscopes by experts. ViewDesigner is a component of HypothesisCreator, which manages viewscopes and provides an interface between the system and users to design and operate viewscopes. In this abstract, we introduce a new version of the component, which is designed to provide environment for users to design viewscopes intuitively.

2 ViewDesigner

ViewDesigner assists users in designing viewscopes on domain data like DNA sequences, amino acid sequences, gene expression profile data and so on. When we design region-specific viewscopes like approximate string matching viewscopes on upstream regions, arbitrary regions easily can be defined by drawing arrows along with a biosequence in a drawing canvas (see Fig. 1). The locations of such regions are described as the interval of one or two particular axes. The origins of these axes

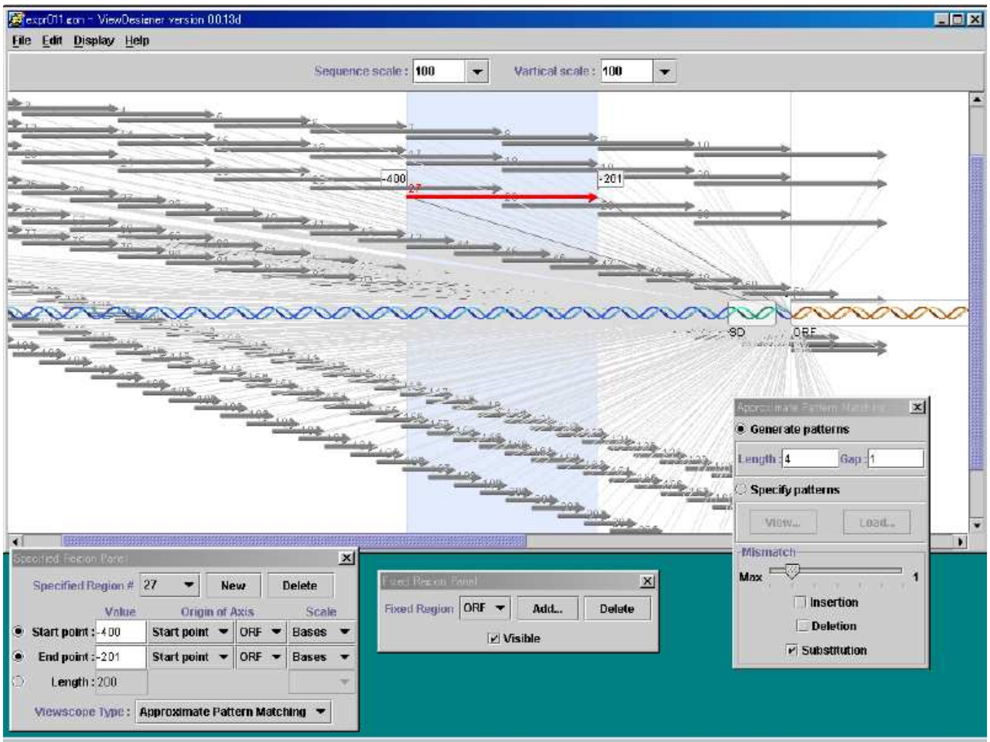

Figure 1: ViewDesigner

are determined by a biologically established sites like open reading frames, regions corresponding to the Shine-Dalgarno sequence, various distinctive binding sites, etc. Any region-specific viewscopes, like approximate string matching, occurrence possibilities of bases and residues, regular expressions matching, and PROSITE pattern matching, can be assigned to such regions expressed as arrows. By designing viewscopes reflecting a creative thought like an intuition, an assumption, and a specific knowledge of the data, the designed viewscopes turn to be a method of verifying the validity of such a thought on the data in collaboration with a selected hypothesis generation algorithm. Furthermore, ViewDesigner is designed and implemented to handle new biological sites on biosequences and additional viewscopes by dynamically loading such data.

ViewDesigner can be considered as a tool dealing with creative thought of experts, which can be employed as a viewscope, in the process of hypothesis generation of HypothesisCreator. To realize easy access to ViewDesigner, the system is implemented like a conventional drawing tool. We think that ViewDesigner is developed into a user interface strongly assisting experts in the discovery process of genomic researches.

References

[1] Maruyama, O., Uchida, T., Shoudai, T. and Miyano, S., Toward Genomic Hypothesis Creator: View designer for discovery, *Proc. 1st International Conference on Discovery Science (Lecture Notes in Artificial Intelligence vol.1532)*, Springer-Verlag, 105–116, 1998.

[2] Maruyama, O., Uchida, T., Sim, K.L. and Miyano, S., Designing views in HypothesisCreator: System for assisting in discovery, *Proc. 2nd International Conference on Discovery Science (Lecture Notes in Artificial Intelligence vol.1721)*, Springer-Verlag, 115–127, 1999.

[3] http://geta.life.uiuc.edu/˜nikos/genomes.html.