

# ПРОГРАММНО-АЛГОРИТМИЧЕСКИЙ КОМПЛЕКС ДЛЯ МНОГОМЕРНОГО АНАЛИЗА МИКРОЧИПОВЫХ ДАННЫХ

**Ефимов В.М.**<sup>1,2</sup>, Штайгер И.А.<sup>3</sup>, Полунин Д.А.<sup>3</sup>, Агбаш И.А.<sup>3</sup>,  
Сулопаров Д.С.<sup>3</sup>

<sup>1</sup> ИЦиГ СО РАН, Новосибирск

<sup>2</sup> ТГУ, Томск

<sup>3</sup> НГУ, Новосибирск

Несмотря на огромное количество специализированных программных средств, предназначенных для анализа микрочиповых данных, в том числе и многомерного, на практике ими не слишком удобно пользоваться, не говоря уже о том, что часть из них является коммерческими. Основная проблема – закрытая реализация, не позволяющая в полной мере реализовывать нужную пользователю обработку. Универсальные средства, такие как R, имеют слишком тяжеловесный интерфейс. Поэтому разработка простого и удобного программного комплекса для анализа микрочиповых данных, доступного для пользователя-биолога без математического или технического образования, представляется весьма актуальной.

Пользователю нужна программная система, обслуживающая реальные практические запросы и сочетающая свойства статистических пакетов и систем управления базами данных. От статистических пакетов, как правило, требуется постоянное преобразование данных, включая их структуру, и порождение новых данных и структур. От систем управления базами данных – точная идентификация исходной информации для хранения, поиска, выборки и замены или корректировки, не меняющей ее структуру. Очевидно, что эти задачи противоречивы и компромисс может быть достигнут только за счет потери эффективности на задачах, типичных для какой-либо одной из названных областей. Однако удобство использования одной системы вместо нескольких с постоянными переключениями от формата одной к формату другой и неуклонно растущая производительность вычислительных средств, по нашему мнению, делают такой компромисс рентабельным.

Разрабатываемая система будет представлять головную программу-диспетчер и комплекс независимо реализованных исполняемых модулей. Для повышения надежности работы предусматривается только пакетная обработка. Предполагается реализовать различные алгоритмы многомерного анализа, адаптированные к специфике данных – небольшое число (несколько десятков или сотен) объектов, очень большое (несколько десятков тысяч) признаков, включая неметрическое многомерное шкалирование, многомерный PLS-анализ и нейронные сети.