



МИНИСТЕРСТВО ПРОСВЕЩЕНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ ПЕДАГОГИЧЕСКИЙ
УНИВЕРСИТЕТ им. А. И. ГЕРЦЕНА»

**ИНСТИТУТ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И
ТЕХНОЛОГИЧЕСКОГО ОБРАЗОВАНИЯ**
Кафедра информационных технологий и электронного обучения

АНАЛИЗ РАЗЛИЧНЫХ ИСТОЧНИКОВ

ПО ТЕМЕ «БИОИНФОРМАТИКА»
(научно-исследовательская работа)

по направлению “09.03.01 – Информатика и вычислительная техника”
(профиль: “Технологии разработки программного обеспечения”)

Зав. кафедрой ИТиЭО д.п.н., проф.

(Власова Е.З.)

Руководитель: доцент кафедры ИТиЭО

(Власов Д. В)

Студент 4 курса

(Матвиив К.А.)

С развитием информационных технологий появилось множество удобств самого разного характера. Не обошли стороной эти удобства и научные исследования. Совершенно неудивительно, что и в биологии появился и нашёл применение целый огромный раздел науки — биоинформатика. Хотя следует отметить, что это и произошло совсем недавно. Так, датой начала использования биоинформатики в биологии можно назвать 1970 год. Тогда Полина Хогевер (нидерландский биолог) и Бен Хеспер ввели термин «биоинформатика», определив его как изучение информационных процессов в биотических системах.

Итак в районе 70-х годов прошлого века ЭВМ стали активно применяться в биологии: к этому времени возросла их память, увеличилась скорость операций, уменьшились размеры. И к тому же накопилось большое количество экспериментальных данных по биологии, требующих осмысления и обработки. Например, уже к 2003 г. объединенными усилиями ученых многих стран был в общих чертах прочитан геном человека. Таким образом, на стыке XX-го и XXI-го веков появилась бурно развивающаяся область биомедицинской науки биоинформатика, которой к настоящему времени насчитывается уже около 30 лет. Биоинформатика обязана своим появлением накоплению обширных экспериментальных данных. Особенно заметно это накопление стало проходить в 70-х гг. XX-го века. Информации, получаемой в биологических экспериментах, было значительно больше, чем возможности человека к запоминанию фактов и их анализу. Возникла необходимость хранения все быстрее увеличивающегося объема информации. Первые несколько сотен расшифрованных последовательностей белков были опубликованы в виде книги-атласа. Однако уже в начале 70-х г. число расшифрованных последовательностей возросло настолько, что из-за их объема оказались невозможными публикации в виде книг. Стало понятно, что нужны специальные программы для сравнения последовательностей, поскольку мозг человека не справляется с анализом такой информации. В 90-е годы происходит расцвет геномики. К настоящему времени расшифрованы полные геномные последовательности человека, мыши, цыпленка, лягушки, отдельных видов рыб, круглых червей, нескольких сотен вирусов и бактерий и т.д. Прочтение генома бактерии – это теперь задача, посильная для группы из 2–3 исследователей за время, меньшее, чем один год. Геном человека составляет около 3-х миллиардов букв, что эквивалентно 15000 книжных томов. А значение факта его «прочтения» для биологов сравнимо с открытием Д.И. Менделеевым периодического закона для химиков. Поэтому для обработки такой обширной биологической информации на помощь пришли компьютерные технологии. Первый алгоритм выравнивания генных последовательностей был предложен еще в 1970 г.

Источников по выбранной теме в интернете на данный момент представлено очень большое количество, так как тема сейчас как никогда актуальна. Первым делом, вдаваясь в биоинформатику стоит ознакомиться с литературой, раскрывающей основы, такие, как например теория эволюции, ведь очень важно уметь быстро вникать в то, о чём же говорят эти непонятные коллабораторы-биологи, и понимать, насколько эта тема разработана и насколько тяжело в ней будет разбираться. Справиться с этим помогут книги, вроде «Эволюция. Триумф идеи», «Внутренняя рыба. История человеческого тела с древнейших времён до наших дней», «Вселенная внутри нас. Что общего у камней, планет и людей», «Приспособиться и выжить! ДНК как летопись эволюции» и «Происхождение видов».

Большое количество методической литературы говорит о том, что проводится очень много исследований, пишется большое количество статей. Так же пользуется спросом и научно-популярный контент, касающийся этой темы. Актуальность биоинформатики

заключается в том, что развитие этого метода с очень большой вероятностью приведет к серьезному прогрессу в решении множества прикладных биологических и медицинских проблем. Так же не стоит забывать и про шагающую по планете информатизацию всего и вся, так же упомянутую выше. Она конечно же не обошла и сферу биологических исследований, что повлияло на актуальность биоинформатики, по сути породив данный гигантский раздел науки при помощи синтеза биологии, статистики, математики, инженерии, computer science, химии и биохимии. Данный факт хорошо отражен в емкой статье «Биоинформатика – это наука, или всё же метод?», расположенной в интернет-источнике «Хабр».

Рассмотреть область работы ученых, продвигающихся в этой сфере на более конкретном примере можно проанализировав проекты научно-исследовательской группы Санкт-Петербургского института биоинформатики.

Выделяются три основных направления исследований:

- Изучение генетической архитектуры сложных признаков человека и интеграция данных полногеномных исследований на уровне совокупности признаков (фенома)

Применяются различные статистические и сетевые методы к публичным наборам данных, такими как UK Biobank, для поиска ключевых генов и молекулярных процессов, лежащих в основе проявления множества сложных признаков. Исследования по этой теме поддержаны стипендиальной программой Systems Biology Fellowship.

- Исследования эволюционных аспектов вариативности генома человека.

Используются публичные массивы геномных и экзомных данных для оценки консервативности различных регионов генома и распределения вариаций внутри них.

- Применение методов биоинформатического анализа данных для решения задач медицинской генетики и клинической медицины в России.

Разрабатываются методы анализа данных NGS в медицинской генетике, а также исследуем генетическую структуру российской популяции и представленность патогенных изменений в геномах россиян.

К сожалению, несмотря на большое разнообразие информации по теме пока что данная отрасль не так хорошо освоена русской научной общественностью. Конечно, перспективы развития области в России не плохие, ведь уже сейчас повсеместно проводятся различные конференции, к примеру конференция 2016 года в Томске.

Несговорова Галина Петровна в своей статье – «Биоинформатика: Пути развития и перспективы» хорошо раскрывает проблему того, что в России под биоинформатикой понимают достаточно узкую область, связанную в первую очередь с обработкой геномных данных, и изучение структур биологических молекул.

Публикация под названием большие данные в биоинформатике говорит нам о том, что современная биоинформатика столкнулась не только с проблемой больших данных, но и с огромным разнообразием методов обработки и представления, одновременным существованием различных программных средств и форматов данных. Обсуждаются пути

решения возникших проблем, в частности путем использования наработок работы с большими данными из других областей современной жизни, таких как сетевой анализ и анализ деловых данных. Новые системы управления базами данных, отличные от реляционных, помогут решить проблему хранения больших данных и обеспечения приемлемого времени выполнения поисковых запросов. Новые технологии программирования, такие, как обобщенное программирование и визуальное программирование призваны решить проблему разнообразия форматов геномных данных и обеспечить возможность оперативного создания собственных скриптов для обработки данных.

Примеры анализа иностранных источников: Nature-Inspired Computing Paradigms in Systems, Anomaly Detection And Complex Event Processing Over Iot Data Streams. Показывают, что мире к ней относят и огромное количество других областей, связанных с обработкой любых биологических данных с использованием информационных технологий. Потому в биоинформатику сегодня приходят физики, математики, химики, люди из смежных областей биологии, программисты в широком смысле слова. При этом профессиональные знания именно в области биологии для биоинформатики очень важны.

На основании всего вышесказанного можно сделать вывод, о том, что источники есть, ведутся различные дискуссии, исполняются проекты, но, к сожалению, данная область пока не так развита в нашей стране, она довольно новаторская и многообещающая, что в очередной раз подчеркивает ее актуальность и необходимость большего углубления в исследования.

Литература

1. Биотехнология, биоинформатика и геномика растений и микроорганизмов : материалы конференции. — Томск : ТГУ, 2016. — 140 с. — ISBN 978-5-94621-539-8. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/92007> (дата обращения: 23.12.2022).
2. Афонников Дмитрий Аркадьевич, Иванисенко Владимир Александрович Биоинформатика: метод во главе угла // Наука из первых рук. 2013. №1 (49). URL: <https://cyberleninka.ru/article/n/bioinformatika-metod-vo-glave-ugla> (дата обращения: 23.12.2022).
3. Несговорова Галина Петровна Биоинформатика: Пути развития и перспективы // Сер. "Конструирование и оптимизация программ" Новосибирск, 2012. URL: https://www.iis.nsk.su/files/articles/sbor_kas_21_nesgovorova_bio.pdf (дата обращения: 23.12.2022).
4. Биоинформатика – это наука, или всё же метод? [Электронный ресурс] // Хабр . URL: <https://habr.com/ru/company/ruvds/blog/650449/> (дата обращения: 23.12.2022).
5. Большие данные в биоинформатике. [Электронный ресурс] // Публикации ВШЭ. URL: <https://publications.hse.ru/articles/203153578> (дата обращения: 23.12.2022).
6. Онлайн-лекторий и гостевые лекции. [Электронный ресурс] // Сайт института биоинформатики. URL: <https://bioinf.me/guests> (дата обращения: 23.12.2022).
7. Schneider Anomaly Detection And Complex Event Processing Over Iot Data Streams // Elsevier science, 2021. URL: <https://medpublishing.ru/catalog/id/129006> (дата обращения: 23.12.2022).
8. Б. Альбертс, Д. Брей, Дж. Льюис, М. Рэфф, К. Робертс, Дж. Уотсон: Молекулярная биология клетки // «Мир» 2010. URL: http://biology.org.ua/files/lib/Alberts_et_al_vol1.pdf (дата обращения: 23.12.2022).
9. К. Циммер «Эволюция. Триумф идеи» // Альпина нон-фикшн, 2015. URL: <https://www.labirint.ru/books/306404/> (дата обращения: 23.12.2022).
10. Н. Шубин «Внутренняя рыба. История человеческого тела с древнейших времён до наших дней» // Corpus, 2021. URL: <https://www.labirint.ru/books/808107/> (дата обращения: 23.12.2022).
11. Н. Шубин «Вселенная внутри нас. Что общего у камней, планет и людей»// Corpus, 2013. URL: <https://www.labirint.ru/books/417109/> (дата обращения: 23.12.2022).
12. Ш. Кэрролл «Приспособиться и выжить! ДНК как летопись эволюции» // Corpus, 2015. URL: <https://www.labirint.ru/books/486073/> (дата обращения: 23.12.2022).
13. Стефанов, В. Е. Биоинформатика : учебник для вузов / В. Е. Стефанов, А. А. Тулуб, Г. Р. Мавропуло-Столяренко. — Москва : Издательство Юрайт, 2023. — 252 с. — (Высшее образование). — ISBN 978-5-534-00860-9. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/511736> (дата обращения: 25.12.2022).
14. А. Леск Введение в биоинформатику // Лаборатория базовых знаний, 2009. URL: <https://www.livelib.ru/book/1000416615-vvedenie-v-bioinformatiku> (дата обращения: 23.12.2022).
15. Часовских Наталия Юрьевна: Биоинформатика. Учебник // ГЭОТАР-Медиа, 2020. URL: <https://www.labirint.ru/books/743227/> (дата обращения: 23.12.2022).
16. Часовских, Н. Ю. Биоинформатика : учебно-методическое пособие / Н. Ю. Часовских. — Томск : СибГМУ, 2015. — 109 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/105971> (дата обращения: 25.12.2022). — Режим доступа: для авториз. пользователей.

17. M. Mellal, M. Petch. Nature-Inspired Computing Paradigms in Systems // Elsevier Science 2021. URL: <https://medpublishing.ru/catalog/id/129088> (дата обращения: 23.12.2022).
18. С. Игнасимуту «Основы биоинформатики»// R&C , 2007. URL: https://market.yandex.ru/product--s-ignasimutu-osnovy-bioinformatiki/1781687370?was_redir=2&cpa=1 (дата обращения: 23.12.2022).
19. А. Павлов Экология и биоинформатика // НИА-Природа, 2007. URL: <http://www.ecocommunity.ru/material.php?id=78> (дата обращения: 23.12.2022).
20. Биотехнология, биоинформатика и геномика растений и микроорганизмов : материалы Всероссийской молодежной научной конференции с международным участием : материалы конференции / под редакцией О. В. Карначук. — Томск : ТГУ, 2016. — 140 с. — ISBN 978-5-94621-539-8. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/80248> (дата обращения: 25.12.2022). — Режим доступа: для авториз. пользователей.
21. Ч. Дарвин. «Происхождение видов» // АН СССР, Москва, 1939. URL: http://www.chronos.msu.ru/old/RREPORTS/darvin_proishozhdenie_vidov.pdf (дата обращения: 25.12.2022).