

Тестовое задание (Kotlin / Ktor) - внешние источники

Техническое задание

Общее описание

Существенная часть практической работы по анализу геномных данных связана с поиском отличий между генетическим материалом некоего образца и последовательностью референсного генома. Такие отличия называются генетическими вариантами, их наличие может свидетельствовать о изменении структуры белков (или функциональных РНК) организма, в свою очередь приводящих к изменениям фенотипа. Характеристика генетических вариантов с точки зрения возможности их влияния на фенотип называется аннотированием, а информация, которая появляется в результате такого процесса – аннотацией.

Существует масса баз данных, выступающих в качестве источников аннотации (например, dbSNP, Clinvar, gnomAD и проч.). На практике часто требуется оперативно получить аннотацию для некоторого генетического варианта (например, при анализе последовательности генома конкретного человека). Разные источники по-разному организуют доступ к данным аннотации. Например, некоторые предоставляют веб-интерфейсы для поиска генетических вариантов и выдачи данных аннотации по ним (здесь следует отметить, что в таком случае речи не идет о каком-либо публичном API, который позволял бы получать эти данные программно), а некоторые идут вместе со специальными утилитами, устанавливаемыми локально (то есть подразумевается, что файлы базы данных должны быть скачаны и доступны с клиентской машины). Так как в биоинформатике существует устойчивая тенденция к переходу к веб-сервисам, такой функционал было бы удобно предоставить в виде HTTP сервиса.

Цель

Создать REST HTTP сервис, предоставляющий информацию об аннотации некоторого генетического варианта в рамках данного источника аннотации (для текущего задания предлагается взять аннотации для вариантов из базы данных Clinvar).

Фактически по запросу по ключевым параметрам (например, http://example.com/info?rac=NC_000007.14&lap=55146654&rap=55146656&refkey=T) должны получить всю дополнительную информацию доступную в файле.

Входные данные

Запакованный текстовый файл со списков вариантов и аннотаций + tabix индекс к нему. Фрагмент файла:

Фрагмент файла

#RAC	LAP	RAP	REFKEY	VCF_ID	CLNSIG	CLNREVSTAT	CLNVC
NC_000001.11	925951	925953	A	1019397	Uncertain_significance	criteria_provided,_single	
NC_000001.11	925955	925957	T	1543320	Likely_benign	criteria_provided,_single_submitt	
NC_000001.11	925968	925970	T	1648427	Likely_benign	criteria_provided,_single_submitt	
NC_000001.11	925975	925977	C	1362713	Uncertain_significance	criteria_provided,_single	
NC_000001.11	925985	925987	T	1568423	Likely_benign	criteria_provided,_single_submitt	
NC_000001.11	926002	926004	T	1365270	Uncertain_significance	criteria_provided,_single	

Технические требования

- Сервис должен быть написан на языке Kotlin, используя Ktor (фреймворк для написания веб-сервисов).
- Сервис должен предоставлять аннотацию для переданного генетического варианта.
- Сервис должен обрабатывать и соответствующим образом информировать пользователя в случае, когда по каким-либо причинам запрос не удалось обработать успешно.
- Нужно учитывать, что файл-архив может быть большого размера (сотни гигабайт).
- Код должен быть доступен для удаленного просмотра и комментирования (github/gitlab и т.п.).
- Описание как сервис собрать и развернуть

Бонусное задание

- Создать docker образ и описать как его запускать