

Projeto YARA

Documento de Levantamento de Requisitos

Data: 03 de Setembro de 2025

Versão: Não versionado;

Status: Concluído

1. Introdução

Este documento detalha os resultados da fase de levantamento de requisitos para o projeto YARA. As informações aqui contidas foram compiladas a partir do documento original do projeto e de uma pesquisa simulada para identificar as dificuldades e necessidades comuns de pesquisadores que utilizam a ferramenta QIIME 2. O objetivo deste documento é guiar as fases subsequentes de design, desenvolvimento e validação do agente conversacional.

2. Simulação de Feedback de Pesquisadores (Compilado da Web)

Para guiar a definição de requisitos, foi realizado um levantamento de dúvidas e dificuldades comuns publicadas em fóruns científicos e tutoriais sobre QIIME 2. O perfil de usuário simulado é um pesquisador da área de ecologia microbiana, com conhecimento iniciante a intermediário em bioinformática.

Dificuldades e Dúvidas Principais:

- **Diversidade Alfa:** A maior dúvida é sobre **qual métrica estatística utilizar** (Shannon, Simpson, etc.) para cada tipo de estudo e **como interpretar a significância estatística** dos resultados entre diferentes grupos de amostras.
- **Diversidade Beta (Gráficos PCoA):** Considerada uma das maiores fontes de confusão. As perguntas recorrentes são: "O que os eixos do gráfico representam?", "Como posso afirmar que meus grupos estão estatisticamente separados?" e "Como reportar os resultados do teste PERMANOVA em uma publicação?".
- **Gráficos de Taxonomia:** A dificuldade reside na visualização de dados muito diversos e na interpretação de abundâncias relativas, especialmente quando há muitos táxons "não classificados" (*unassigned*).
- **Curvas de Rarefação:** A principal incerteza é sobre como avaliar se a curva atingiu um "platô", indicando que a profundidade de sequenciamento foi suficiente para capturar a diversidade da amostra.

Necessidades e Expectativas para o YARA:

- **Interação Direta:** Desejo de poder fazer perguntas diretas como: "Qual amostra é mais diversa?" ou "Gere uma explicação para este gráfico".
- **Exportação para Publicação:** Necessidade de exportar gráficos em alta resolução (PNG, SVG, PDF) e tabelas em formatos editáveis (TSV, Excel).
- **Automação de Texto:** A funcionalidade de gerar automaticamente um parágrafo descritivo para um resultado é vista como de altíssimo valor para acelerar a escrita de artigos científicos.
- **Funcionalidades Adicionais:** Sugestões incluem alertas sobre possíveis problemas na análise, sugestões de próximos passos estatísticos e um glossário interativo para termos técnicos.

3. Requisitos Funcionais (O que o sistema deve fazer)

- **RF01: Interação por Linguagem Natural:** O sistema deve ser capaz de processar e entender perguntas de usuários escritas em linguagem natural (português e inglês).
- **RF02: Interpretação de Diversidade Alfa:** O agente deve explicar métricas de diversidade alfa (como Shannon, Simpson, Observed Features) presentes nos resultados do QIIME 2.
- **RF03: Interpretação de Diversidade Beta:** O agente deve interpretar análises de diversidade beta (como PCoA e UniFrac), explicando o que os gráficos e os testes estatísticos associados significam.
- **RF04: Análise de Composição Taxonômica:** O sistema deve gerar, exibir e descrever gráficos de barras de composição taxonômica.
- **RF05: Esclarecimento sobre Rarefação:** O agente deve ser capaz de explicar curvas de rarefação e ajudar o usuário a entender se a profundidade de amostragem foi adequada.
- **RF06: Geração de Relatórios:** O sistema deve gerar relatórios técnicos automatizados, formatados e compreensíveis a partir dos resultados.
- **RF07: Exportação de Dados:** O usuário deve poder solicitar e receber tabelas e gráficos em diversos formatos (PNG, PDF, SVG, HTML, TSV).
- **RF08: Alertas de Erros Comuns:** O sistema deve identificar e alertar o usuário sobre possíveis problemas ou erros comuns nas análises (ex: classificações taxonômicas truncadas, baixa cobertura de sequenciamento).
- **RF09: Geração de Texto Explicativo:** O agente deve ser capaz de gerar parágrafos de texto que descrevam e interpretem um gráfico ou resultado estatístico.
- **RF10: Sugestões de Análise:** Com base em um resultado inicial, o agente deve ser capaz de sugerir próximos passos ou análises estatísticas complementares.
- **RF11: Interface com Elementos Interativos:** O agente deve usar botões, menus e outros elementos interativos para facilitar a navegação.

4. Requisitos Não Funcionais (Como o sistema deve ser)

- **RNF01: Plataforma de Acesso:** O sistema deve ser acessível via uma interface de chat, especificamente através do Telegram.
- **RNF02: Usabilidade:** A ferramenta deve ter um alto nível de usabilidade, a ser validado com métodos padronizados como o System Usability Scale (SUS).
- **RNF03: Código Aberto:** O projeto será desenvolvido com o framework open source Rasa e o código-fonte final será disponibilizado com uma licença livre.
- **RNF04: Idioma:** A interface do agente deve ser bilíngue, suportando interações em português e inglês.
- **RNF05: Auditabilidade e Transparência:** A arquitetura do sistema deve ser transparente e auditável, garantindo a soberania dos dados e a confiança nos resultados gerados.
- **RNF06: Modularidade:** A arquitetura deve ser dividida em módulos (NLU, Core, Actions Server) para facilitar a manutenção e futuras expansões.
- **RNF07: Interoperabilidade:** O sistema deve ser construído em Python e ter integração direta com bibliotecas científicas essenciais (pandas, matplotlib, seaborn, biom-format, etc.).