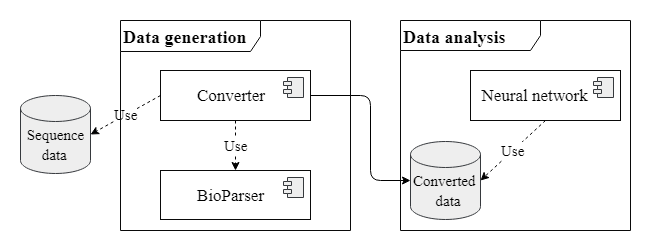
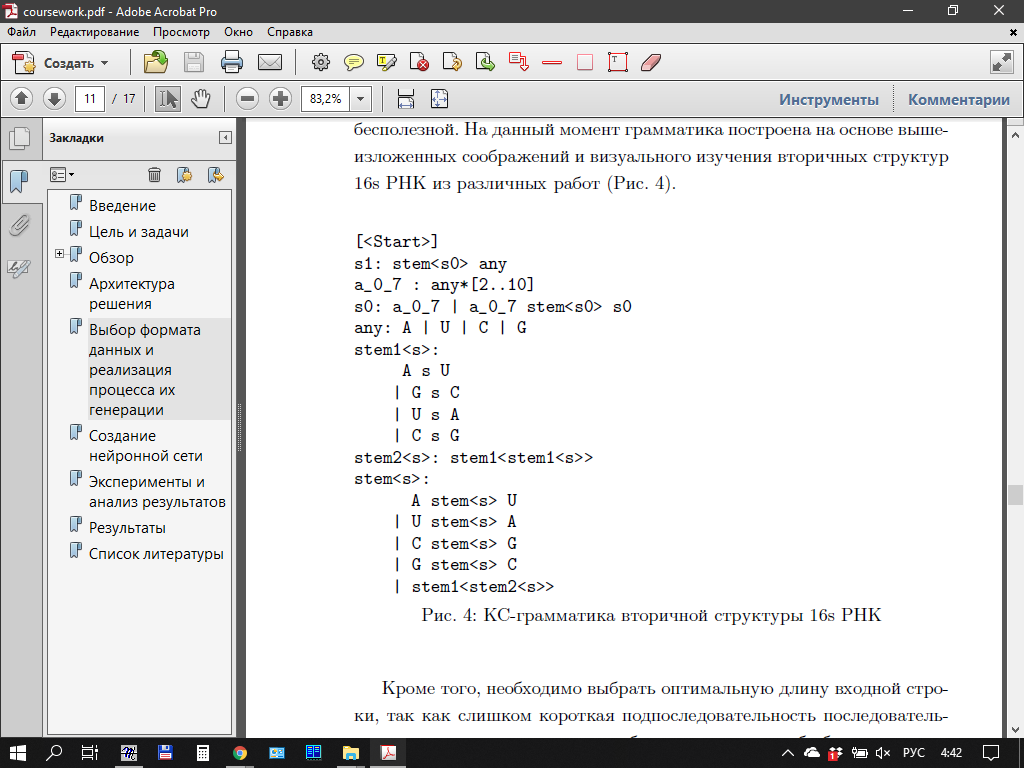
**Что есть сейчас (распознавание):**

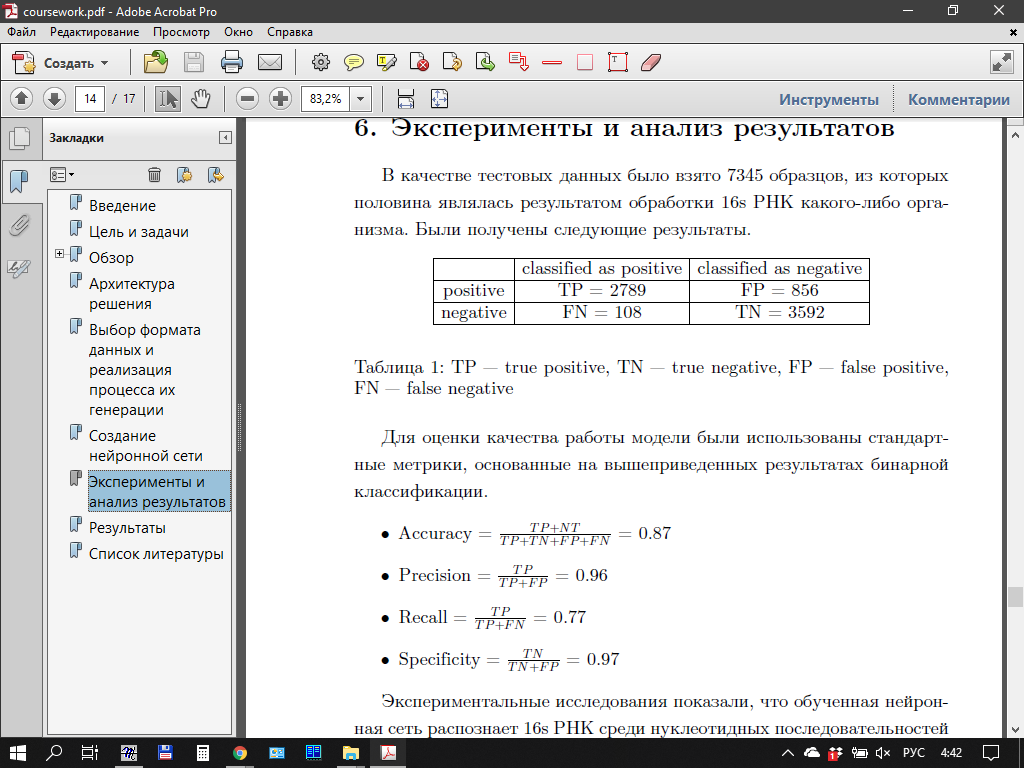
1. Архитектура решения:

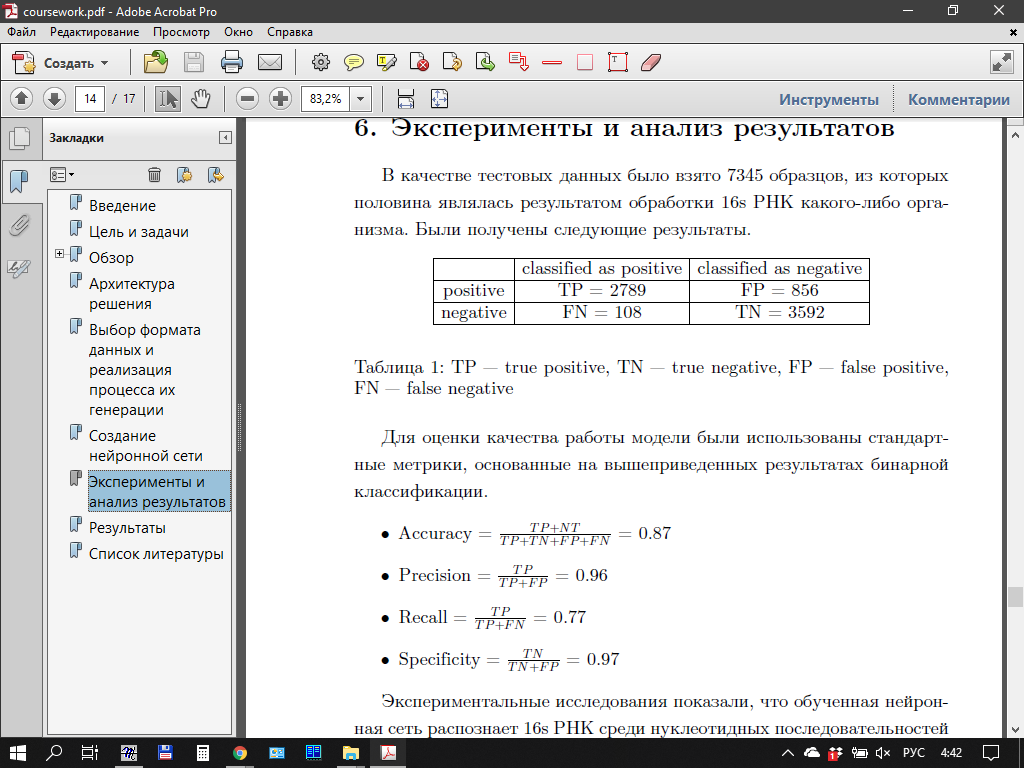


1. Sequence data из бд SIlVA (подпоследовательности длины 512)
2. Грамматика:



1. Нейронная сеть в файле Model1





**Планы (классификация):**

1. Последовательности из GreenGenes, только бактерии с полной классификацией
2. Добавить в метаданные информацию из дерева жизни
3. Взять только те организмы, для которых в базе есть 100 и более цепочек. Всего 143 вида, достаточно равномерно распределенных по общему дереву жизни
4. Длина подцепочки?
5. Сбалансированная выборка: чем меньше цепочек есть в базе, тем меньше шаг
6. Улучшить парсер и грамматику
7. Скрипт для подбора параметров модели нейросети и обучения
8. На выходе -- вероятностная модель

Жалкая попытка архитектуры решения:

