ИНСТРУКЦИЯ: Сохранение проекта Anonymizer на GitHub

Дата создания: 21 октября 2025 г.

Версия: 1.0.0

Статус: BlockBuilder с поддержкой header/footer извлечения завершен

# Этап 1: Подготовка репозитория

## 1.1 Проверка текущего состояния Git

Выполните в PowerShell:

cd C:\Projects\Anonymizer  
git status

## 1.2 Инициализация Git (если нужно)

Если Git не инициализирован:

git init

## 1.3 Создание .gitignore файла

Создайте файл .gitignore со следующим содержимым:

# Python  
\_\_pycache\_\_/  
\*.py[cod]  
\*$py.class  
\*.so  
.Python  
env/  
venv/  
venv\_\*/  
ENV/  
env.bak/  
venv.bak/  
  
# FastAPI  
.env  
.env.local  
.env.production  
  
# IDE  
.vscode/  
.idea/  
\*.swp  
\*.swo  
\*~  
  
# OS  
.DS\_Store  
Thumbs.db  
  
# Logs  
\*.log  
logs/  
  
# Test results  
test\_result.json  
test\_\*.json  
  
# Temporary files  
temp/  
tmp/  
\*.tmp  
  
# Virtual environments  
venv\_gateway/  
venv\_orchestrator/  
venv\_unified\_document\_service/  
venv\_nlp\_service/  
venv\_rule\_engine/  
venv\_frontend/

# Этап 2: Подготовка коммита

## 2.1 Добавление файлов в индекс

Добавляем все изменения (кроме файлов из .gitignore):

git add .

## 2.2 Создание коммита

Создаем коммит с описанием изменений:

git commit -m "feat: Complete BlockBuilder with header/footer extraction for anonymization  
  
- Implement robust DOCX parsing with SDT support  
- Add header/footer text extraction with XPath  
- Include text normalization for consistent matching  
- Add sensitive\_matches metadata for anonymization  
- Support all 6 microservices with virtual environments  
- Complete startup scripts for Windows environment  
- Fix critical issue with missing header data in blocks"

# Этап 3: Создание GitHub репозитория

## 3.1 Создание README.md

Создайте файл README.md с описанием проекта (см. содержимое ниже)

## 3.2 Подключение к GitHub

Замените [YOUR\_USERNAME] на ваш GitHub username:

git remote add origin https://github.com/[YOUR\_USERNAME]/anonymizer.git

## 3.3 Пуш в GitHub

Отправляем код в GitHub:

git branch -M main  
git push -u origin main

# Этап 4: Создание релиза

## 4.1 Создание тега версии

Создаем тег версии для релиза:

git tag -a v1.0.0 -m "Release v1.0.0: Complete BlockBuilder with header extraction"  
git push origin v1.0.0

# Структура проекта

Anonymizer/  
├── frontend/ # Streamlit frontend (порт 8501)  
├── gateway/ # API Gateway (порт 8000)  
├── orchestrator/ # Orchestrator service (порт 8002)  
├── unified\_document\_service/ # Document processing (порт 8001)  
│ ├── app/  
│ │ ├── main.py  
│ │ ├── block\_builder.py # КРИТИЧНЫЙ: извлечение header/footer  
│ │ └── document\_io.py  
│ ├── test\_docs/ # Тестовые документы  
│ └── test\_block\_builder.py # Тесты  
├── nlp\_service/ # NLP processing (порт 8003)  
├── rule\_engine/ # Rule engine (порт 8004)  
├── start\_\*.bat # Скрипты запуска сервисов  
├── venv\_\*/ # Virtual environments  
└── \*.md # Документация

# Важные замечания

🔴 КРИТИЧНО: BlockBuilder исправлен для анонимизации

• Извлечение header/footer данных через SDT элементы

• Нормализация текста (неразрывные пробелы)

• Метаданные applies\_to: "section"

• Поддержка всех типов колонтитулов

✅ Готово к продакшн:

• 6 микросервисов с виртуальными окружениями

• Windows .bat скрипты для запуска

• Полное извлечение чувствительных данных

• Тесты и валидация BlockBuilder