

Государственное учреждение образования
“БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ”

Кафедра: Интеллектуальных информационных технологий
Дисциплина: Объектное моделирование интеллектуальных систем

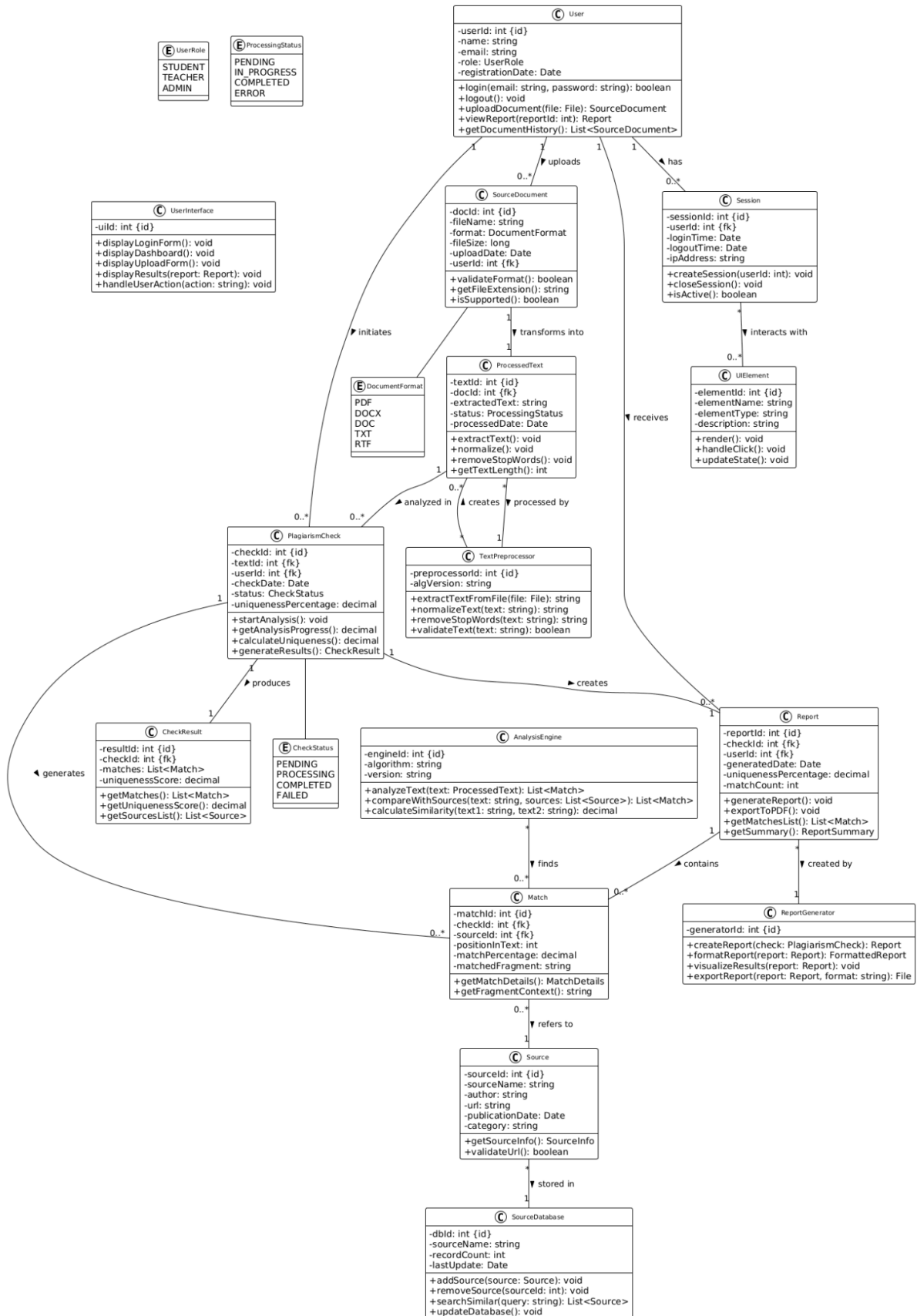
Отчет по лабораторной работе №4

Выполнил:
студент гр.321701
Семков А.А.

Проверил:
Зотов Н.В.

Минск 2025

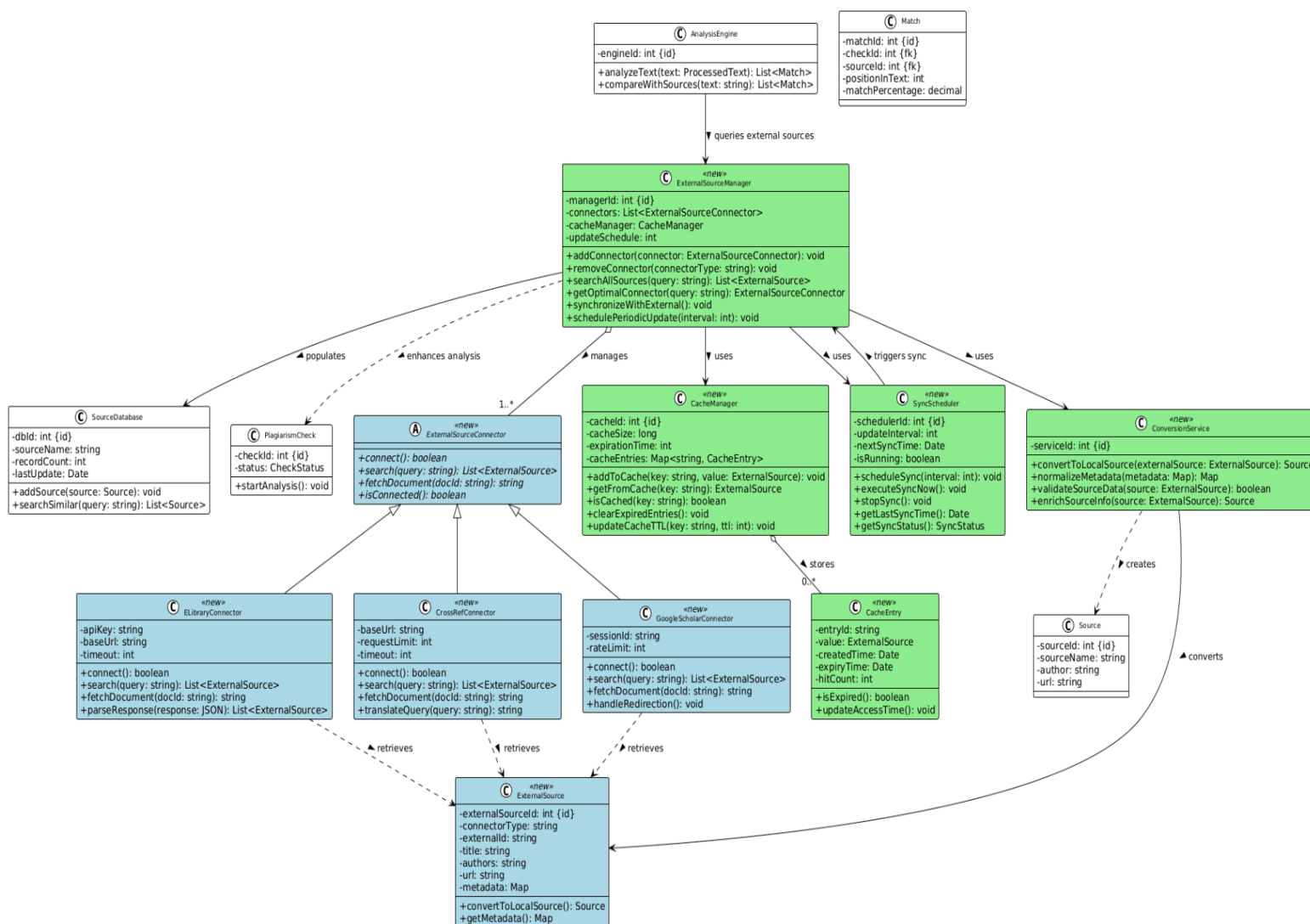
1. Базовая диаграмма классов:



2. Направления развития системы

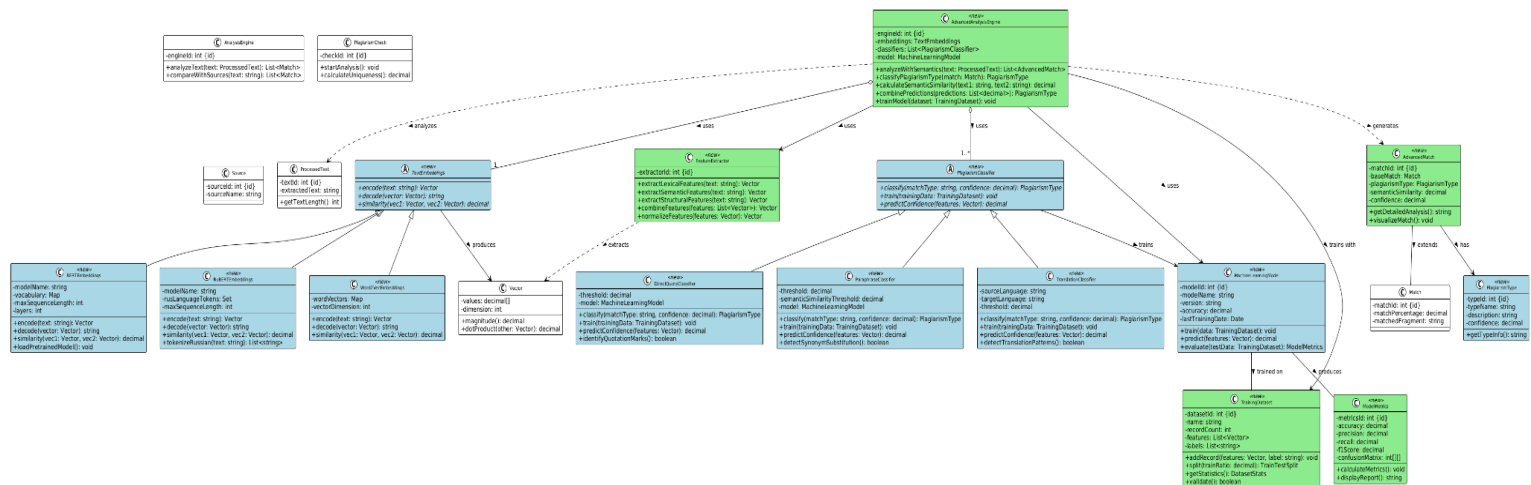
2.1. Интеграция с внешними источниками

Описание: Первое направление расширяет базовую систему подключением к внешним научным базам данных (eLibrary.ru, CrossRef, Google Scholar). Система использует паттерн Strategy для абстрактного интерфейса **ExternalSourceConnector**, позволяя добавлять новые источники без изменения основного кода. Применяется Open/Closed Principle: новые коннекторы создаются наследованием, а не модификацией. Компонент **ExternalSourceManager** управляет всеми источниками, **CacheManager** оптимизирует поиск кэшированием результатов, а **SyncScheduler** автоматически обновляет базу источников каждые 24 часа. Это позволяет системе получать актуальные данные о научных публикациях и повышает точность обнаружения плагиата.



2.2. Машинное обучение и обработка текста (NLP)

Описание: Второе направление внедряет нейросетевые модели обработки естественного языка для более точной классификации типов плагиата. Система использует три модели встраивания текста (**TextEmbeddings**): Google BERT для английского текста, RuBERT для русскоязычного контента и Word2Vec для классического подхода. На основе встраиваний работают три классификатора (**PlagiarismClassifier**): **DirectQuoteClassifier** обнаруживает дословные цитирования (80%+ совпадение), **ParaphraseClassifier** выявляет перефразирование через анализ синонимов и изменений структуры, **TranslationClassifier** определяет переводные заимствования. Компонент **AdvancedAnalysisEngine** комбинирует предсказания всех классификаторов для получения итогового типа плагиата с оценкой уверенности. Принцип Open/Closed соблюдается: новые модели встраивания и классификаторы добавляются через наследование. Преимущество: высокая точность (90%+), способность распознавать сложные случаи перефразирования и перевода. Недостаток: требует предварительного обучения на 10k+ проверенных работ.



2.3. Аудит и контроль действий пользователей

Описание: Третье направление добавляет полную систему аудита и мониторинга для защиты от злоупотреблений. Система логирует все действия пользователей через абстрактный класс **AuditEvent** с конкретными реализациями (**DocumentUploadEvent**, **PlagiarismCheckEvent**, **ReportAccessEvent**, **SettingsChangeEvent**, **DataExportEvent**). Компонент **AnomalyDetectionEngine** в реальном времени анализирует события через несколько детекторов: **BulkUploadDetector** выявляет массовую загрузку документов (более N за час), **UnusualAccessPatternDetector** выявляет доступ в необычное время или нестандартное поведение, **PrivilegeAbuseDetector** проверяет соответствие действий пользователя его ролевым полномочиям. При обнаружении аномалии генерируется **AnomalyAlert** с уровнем серьёзности. Компонент **ReportGenerator** создаёт дневные, недельные и месячные отчёты, а **NotificationService** с поддержкой Email и Slack уведомляет администраторов о критических событиях. Применяется Open/Closed Principle: новые типы событий, детекторы и каналы уведомлений добавляются наследованием. Это обеспечивает прозрачность и безопасность, позволяя администраторам контролировать использование системы и предотвращать злоупотребления.

