

Методология фиксации

Методология фиксации изменений (для приобщения к материалам)

1. Снапшоты. Система периодически скачивает публичные документы маркетплейсов по каноническим URL без обхода авторизации. Для каждого скачивания сохраняются сырые байты ответа сервера и вычисляется их хэш SHA-256 (raw_sha256).
2. Canonical-текст. Из документа извлекается основной текст, нормализуется (Unicode NFC, единые переводы строк, схлопывание пробелов, удаление заведомо волатильных элементов — счётчиков, дат генерации страницы). Хэш SHA-256 нормализованного текста (canonical_sha256) служит ключом версии документа.
3. Двухфазное подтверждение. Новая версия фиксируется только после двух последовательных скачиваний с одинаковым новым canonical_sha256 с интервалом 10–15 минут — защита от временных ошибок отдачи.
4. Hash chain (журнал). Каждый акт фиксации записывается в append-only журнал; каждая запись включает SHA-256-хэш предыдущей записи. Изменение любой записи задним числом делает невалидными все последующие записи цепочки.
5. Суточное якорение. Ежедневно по всем записям журнала за прошедшие сутки строится корень дерева Меркла, который якорится в блокчейне Bitcoin через протокол OpenTimestamps (квитанция .ots прилагается). Это доказывает, что данные существовали не позже даты соответствующего блока Bitcoin.
6. Независимые архивы. При каждой подтверждённой новой версии документ отправляется на архивацию в независимые сервисы Wayback Machine (web.archive.org) и archive.today; ссылки на копии прилагаются.
7. Время. Серверные часы синхронизируются по NTP (chrony). Метки времени в журнале — UTC; даты в отчёте приводятся по московскому времени (Europe/Moscow).