



ОПИСАНИЕ ПОЛЕЗНОЙ МОДЕЛИ К ПАТЕНТУ

(21) 2025/0932.2

(22) 19.06.2025

(45) 12.09.2025, бюл. №37

(72) Киреева Анель Ахметовна; Урдабаев Марат Тагайбекович; Дигель Иван Евгеньевич

(73) Товарищество с ограниченной ответственностью «Университет Международного Бизнеса имени Кенжегали Сагадиева»

(74) Жангабылов Абай Токтарбекович

(56) US 20100106669 A1, 29.04.2010

(54) **ПРОГРАММНО-АППАРАТНЫЙ КОМПЛЕКС ДЛЯ ВЕРИФИЦИРУЕМОЙ ОБРАБОТКИ БИБЛИОМЕТРИЧЕСКИХ ДАННЫХ, РАНЖИРОВАНИЯ ИЗДАНИЙ И ГЕНЕРАЦИИ ИНТЕГРАЛЬНОГО РЕЙТИНГА**

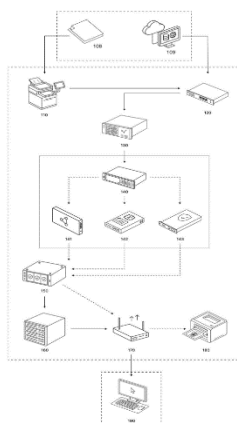
(57) Полезная модель относится к области аппаратных средств защищённой обработки данных, в частности, к распределённым вычислительным комплексам, реализованным в виде физически обособленных, но взаимосвязанных устройств, предназначенных для генерации верифицируемых аналитических показателей из множественных источников, включая как цифровые, так и физические (бумажные) носители.

Задачей, на решение которой направлена полезная модель, является создание распределённого аппаратно-программного комплекса, который преодолевает ограничения существующих

программных аналогов путём решения технических подзадач на аппаратном уровне.

Техническим результатом является создание комплекса, обеспечивающего не только повышенную точность рейтинга, но и его доказуемую достоверность и целостность за счёт архитектуры, реализующей многофакторный анализ с автоматической фильтрацией аномалий и выполнением критически важных вычислений в аппаратно-защищённой среде, что обеспечивает прозрачность, возможность внешней верификации и исключает риск манипуляций.

Предложен программно-аппаратный комплекс (ПАК) для верифицируемой обработки библиометрических данных, ранжирования изданий и генерации интегрального рейтинга, который включает станцию оцифровки и распознавания документов для обработки физических носителей, устройство приема данных, сервер обработки графов, кластер параллельной аналитики для генерации аналитических сигналов, верификационное устройство с поддержкой TEE, аппаратный шлюз API, представляющий собой устройство, принтер для печати защищённых аттестатов для выведения отчета на материальный носитель; центральный массив хранения данных, а также пользовательский терминал, который обращается к шлюзу API.



Полезная модель относится к области аппаратных средств защищённой обработки данных, в частности, к распределённым вычислительным комплексам, реализованным в виде физически обособленных, но взаимосвязанных устройств, предназначенных для генерации верифицируемых аналитических показателей из множественных источников, включая как цифровые, так и физические (бумажные) носители. Технически данный аппаратный комплекс решает задачу обеспечения целостности и доказуемости результатов сложных многофакторных вычислений, выполняемых на основе больших массивов данных. Она находит применение в сферах, где требуется не только результаты вычисления, но и аппаратно-гарантированное подтверждение того, что сам процесс вычисления был защищён от несанкционированного вмешательства, на примере обработки библиометрической информации для построения объективных и защищённых от манипуляций рейтингов.

Существующие системы для анализа научной деятельности, как правило, реализованы в виде программных приложений, работающих на серверах общего назначения. Эти системы основаны на метриках (таких как Impact Factor, SJR), которые в силу своей чисто программной архитектуры имеют существенные технические ограничения. Они обрабатывают преимущественно прямые, нефильТРованные показатели цитируемости, что делает их вычислительные процессы уязвимыми для манипуляций с данными, например, через аномальное самоцитирование.

Известна система ранжирования узлов в связанной базе данных, в которой используется сетевой анализ, например, на основе алгоритма PageRank. Данная система присваивает ранги важности узлам в связанной базе данных, например, в любой базе данных документов, содержащих цитаты, во всемирной паутине или в любой другой базе данных гипермедиа. Рейтинг документа рассчитывается на основе рейтингов цитирующих его документов. Кроме того, ранг документа рассчитывается на основе константы, представляющей вероятность того, что браузер, просматривающий базу данных, случайным образом перейдет к документу. Этот метод особенно полезен для повышения производительности результатов поиска в гипермедийных базах данных, таких как всемирная паутина, документы которой сильно различаются по качеству (US№6285999 B1, 04.09.2001г.).

Однако техническая реализация таких систем, как правило, не предусматривает механизмов для взвешивания связей в графе на основе показателей доверия к источникам или для учёта тематической релевантности, обрабатывая все связи как равнозначные. Это является архитектурным недостатком, снижающим достоверность конечного результата.

Наиболее близким аналогом является система поддержки принятия решений о подаче рукописи в журнал, целью которой является создание единой системы поддержки принятия решений о подаче

рукописей в журнал. Она включает в себя три основные подсистемы: систему фильтрации факторов принятия решений, систему поддержки принятия решений при подаче рукописей и систему проверки модели принятия решений (US№20100106669 A1, 29.04.2010г.).

Однако в данной системе отсутствует ключевой технический компонент: модуль для автоматизированной верификации достоверности исходных данных.

Кроме того, наиболее существенным техническим недостатком известных систем является отсутствие физических средств для обеспечения и верификации целостности самого вычислительного процесса. Поскольку критические вычисления выполняются в незащищённой среде на стандартном оборудовании, конечный пользователь получает результат вычислений, но не может получить технического, аппаратно-гарантированного доказательства того, что расчёт был произведён корректным, немодифицированным алгоритмом и на основе верифицированных данных. Таким образом, существующие системы функционируют как «чёрные ящики» на уровне программного обеспечения, не решая техническую проблему создания доказуемо достоверного и защищённого от манипуляций аналитического показателя, целостность которого можно было бы проверить на физическом уровне.

Задачей, на решение которой направлена полезная модель, является создание распределённого аппаратно-программного комплекса, который преодолевает ограничения существующих программных аналогов путём решения следующих технических подзадач на аппаратном уровне:

- создание многокомпонентного аппаратного конвейера обработки данных путем построения физической инфраструктуры, включающей специализированные устройства для каждого этапа обработки, что обеспечивает аппаратное разделение и оптимизацию вычислительных задач;
- интеграция в архитектуру комплекса специализированных аппаратных фильтров для повышения помехоустойчивости и надёжности вычислительного процесса за счет использования выделенного узла контроля качества данных, оптимизированный для автоматизированного обнаружения и коррекции аномалий в потоке входных данных;
- обеспечение доказуемой целостности вычислений на аппаратном уровне для устранения проблемы «чёрного ящика», присущей программным системам, в целях выполнения финального, наиболее критичного этапа расчёта внутри верификационного устройства, использующего аппаратно-защищённую среду (ТЭЕ). Это переносит гарантии целостности с уровня программного обеспечения на физический, аппаратный уровень процессора;
- реализация возможности физической и криптографической верификации результата для создания механизма, предоставляющего не только итоговый результат через аппаратный шлюз API, но

и криптографическое подтверждение (аттестат) того, что вычисление было выполнено корректно в защищённой аппаратной среде. Это обеспечивает возможность внешней верификации и, при использовании принтера, обеспечит создания материального доказательства целостности расчёта.

Техническим результатом является создание комплекса, обеспечивающего не только повышенную точность рейтинга, но и его высокую достоверность и целостность за счёт архитектуры, реализующей многофакторный анализ с автоматической фильтрацией аномалий и выполнением критически важных вычислений в аппаратно-защищённой среде, что обеспечивает прозрачность, возможность внешней верификации и исключает риск манипуляций.

Указанный технический результат достигается тем, что предложен программно-аппаратный комплекс (ПАК) для верифицируемой обработки библиометрических данных, ранжирования изданий и генерации интегрального рейтинга, который построен на взаимодействии специализированных физических устройств, обеспечивающих полный цикл обработки: от оцифровки бумажных носителей до предоставления криптографически заверенного итогового рейтинга. Таким образом, предлагаемый ПАК включает:

- станцию оцифровки и распознавания документов для обработки физических носителей (например, бумажных журналов), состоящую из высокоскоростного сканера документов, выделенного процессорно-вычислительного блока и терминала верификации;

- устройство приема данных, представляющее собой выделенный сетевой сервер, получающий данные от станции оцифровки и распознавания документов и внешних источников для передачи унифицированного потока данных далее по конвейеру;

- сервер обработки графов, оснащенный большим объемом оперативной памяти и быстрыми накопителями, преобразующий потоки данных в структуру взвешенного графа цитирования;

- кластер параллельной аналитики для генерации аналитических сигналов, состоящий из нескольких физических серверных узлов, параллельно обрабатывающие данные для генерации трех независимых показателей:

- узел сетевого анализа для вычисления показателя влияния цитирования (Citation Influence Score, CIS) посредством итеративного анализа структуры графа;

- узел контентного анализа для анализа текстов публикаций и вычисления показателя тематической близости (Quality-adjusted Similarity Score, QSS);

- узел контроля качества данных, функционирующий как аппаратный фильтр, анализируя граф на наличие нетипичных паттернов цитирования и формируя показатель целостности данных (Data Integrity Score, DIS);

- верификационное устройство с поддержкой TEE, реализованное в виде физического сервера, процессор которого поддерживает аппаратную

доверенную среду исполнения (Trusted Execution Environment, TEE);

- аппаратный шлюз API (Application Programming Interface), представляющий собой устройство, управляющее всеми внешними запросами, для предоставления верифицируемого результата и публикации итогового рейтинга QJR;

- принтер для печати защищённых аттестатов для выведения отчета на материальный носитель;

- центральный массив хранения данных, представляющий собой сетевую систему хранения данных (Storage Area Network, SAN или Network Attached Storage, NAS) с поддержкой RAID (Redundant Array of Independent Disks) для отказоустойчивости, а также

- пользовательский терминал, который обращается к шлюзу API.

Описание полезной модели со ссылками на фигуру:

На фиг.1 показана блок-схема ПАК с верифицируемой целостностью вычислений.

Далее варианты осуществления полезной модели будут подробно описаны со ссылками на прилагаемую фигуру.

Итак, согласно фиг.1, предлагаемый программно-аппаратный комплекс (ПАК) для верифицируемой обработки библиометрических данных, ранжирования изданий и генерации интегрального рейтинга включает:

- станцию оцифровки и распознавания документов (110) для обработки физических носителей (например, бумажных журналов), состоящую из высокоскоростного сканера документов, выделенного процессорно-вычислительного блока и терминала верификации;

- устройство приема данных (120), представляющее собой выделенный сетевой сервер, получающий данные от станции оцифровки и распознавания документов и внешних источников для передачи унифицированного потока данных далее по конвейеру;

- сервер обработки графов (130), оснащенный большим объемом оперативной памяти и быстрыми накопителями, преобразующий потоки данных в структуру взвешенного графа цитирования;

- кластер параллельной аналитики (140) для генерации аналитических сигналов, состоящий из нескольких физических серверных узлов, параллельно обрабатывающие данные для генерации трех независимых показателей:

- узел сетевого анализа (141) для вычисления показателя влияния цитирования (Citation Influence Score, CIS) посредством итеративного анализа структуры графа;

- узел контентного анализа (142) для анализа текстов публикаций и вычисления показателя тематической близости (Quality-adjusted Similarity Score, QSS);

- узел контроля качества данных (143), функционирующий как аппаратный фильтр, анализируя граф на наличие нетипичных паттернов цитирования и формируя показатель целостности данных (Data Integrity Score, DIS);

- верификационное устройство с поддержкой TEE (150), реализованное в виде физического сервера, процессор которого поддерживает аппаратную доверенную среду исполнения (Trusted Execution Environment, TEE);

- аппаратный шлюз API (170), представляющий собой устройство, управляющее всеми внешними запросами, для предоставления верифицируемого результата и публикации итогового рейтинга QJR (Quality Journal Rank);

- принтер для печати защищённых аттестатов (180) для выведения отчета на материальный носитель;

- центральный массив хранения данных (160), представляющий собой сетевую систему хранения данных (SAN или NAS) с поддержкой RAID для отказоустойчивости, а также

- пользовательский терминал (190), который обращается к шлюзу API (170).

Предлагаемый ПАК работает и используется следующим образом:

Реализация технического результата достигается за счет взаимодействия физических устройств. Каждый компонент ПАК представляет собой специализированное устройство, сконфигурированное для выполнения определенного этапа в конвейере верифицируемой обработки данных.

Архитектура и функционирование аппаратных компонентов:

1. Устройства ввода и структурирования данных:

Обработка начинается с устройств, предназначенных для сбора и первичной обработки информации из физических и цифровых источников.

Станция оцифровки и распознавания документов (110): Данное устройство представляет собой аппаратный пост для ввода данных с бумажных носителей.

Физически станция состоит из:

- Промышленного сканера документов: Высокоскоростное устройство с автоподатчиком, способное обрабатывать большие объемы печатной продукции.

- Выделенного вычислительного блока: Компактный компьютер, сопряженный со сканером, на котором исполняется программное обеспечение для оптического распознавания символов (Optical Character Recognition, OCR) и анализа структуры документа. Этот блок аппаратно оптимизирован для задач обработки изображений.

- Терминала верификации: Сенсорный монитор с клавиатурой, позволяющий оператору контролировать процесс распознавания и вносить ручные корректировки в извлеченные метаданные (название, авторы, журнал и т. д.) перед отправкой данных в систему.

Устройство приёма данных (120): это специализированный сетевой шлюз (appliance), выполненный в виде монтируемого в стойку сервера. Его техническая задача – агрегация данных. Устройство оснащено несколькими высокоскоростными сетевыми интерфейсами

(Network Interface Cards, NICs) для одновременного приема данных от станции оцифровки (110) и из внешних цифровых API (например, Web of Science, Scopus) по защищенным протоколам (Transport Layer Security, TLS). Аппаратный брандмауэр, встроенный в устройство, обеспечивает защиту периметра системы.

Сервер обработки графов (130): данные с устройства (120) поступают на этот сервер. Он сконфигурирован для интенсивных операций с графовыми базами данных (например, Neo4j). Его аппаратная конфигурация включает многоядерный процессор, значительный объем оперативной памяти (256 ГБ или более) и массив твердотельных накопителей (NVMe SSD), работающих в режиме RAID для обеспечения максимальной скорости чтения/записи, необходимой для построения и обновления взвешенного графа цитирований в реальном времени.

Кластер параллельной аналитики (140) используется для эффективного многофакторного анализа граф цитирования обрабатывается в кластере, который представляет собой единое шасси с несколькими независимыми серверными лезвиями (blades). Узлы кластера соединены между собой высокоскоростной внутренней шиной (например, InfiniBand) для минимизации задержек при обмене данными. Каждый узел сконфигурирован для выполнения своей задачи:

- Узел сетевого анализа (141): оптимизирован для итеративных вычислений. Выполняет анализ структуры графа для расчета показателя влияния цитирования (CIS).

- Узел контентного анализа (142): оснащен графическими ускорителями (Graphics Processing Unit, GPU) для ускорения вычислений, связанных с моделями обработки естественного языка (Natural Language Processing, NLP), и вычисляет показатель тематической близости (QSS).

- Узел контроля качества данных (143): использует процессор с расширенным набором инструкций для векторных вычислений для эффективного применения алгоритмов машинного обучения с целью выявления аномальных паттернов цитирования и расчета показателя целостности данных (DIS).

Верификационное устройство и защищённые вычисления (150): это ядро всей системы, обеспечивающее доказуемую целостность вычислений.

- Техническая основа: устройство (150) представляет собой физически защищенный сервер, центральный процессор которого (например, Intel® Xeon® с технологией SGX) поддерживает создание доверенной среды исполнения (Trusted Execution Environment, TEE). Эта среда создает защищенный анклав – изолированную область на кристалле самого процессора. Код и данные внутри анклава зашифрованы и недоступны для инспекции или модификации со стороны любого другого программного обеспечения, включая операционную систему хоста.

- Защищённые операции: внутри этого аппаратного анклава выполняются две критически важные операции: 1) расчет базового коэффициента доверия источника (Journal Credibility Index, JCI); 2) финальная агрегация показателей CIS, QSS и DIS в итоговый рейтинг QJR. Процесс является итеративным. Для первоначального построения графа сервер (130) может использовать ранее вычисленные значения JCI. В ходе каждого цикла обработки верификационное устройство (150) рассчитывает обновленные, верифицированные значения JCI, которые передаются обратно на сервер обработки графов (130) для уточнения весов графа в последующих циклах вычислений. Когда система запускается в самый первый раз, «ранее вычисленных» значений еще не существует. В этом случае Сервер обработки графов (130) использует начальное (стартовое) значение JCI. Как правило, это нейтральное значение (например, $JCI = 1.0$), которое присваивается всем источникам по умолчанию.

- Аппаратная аутентификация: для авторизации изменений в коде, исполняемом внутри анклава (например, при обновлении формулы расчета рейтинга), устройство (150) оснащено USB-портом для подключения физического ключа аутентификации. Операции администрирования могут быть выполнены только после предъявления такого ключа.

- Аттестация: устройство (150) способно генерировать криптографический отчет (аттестат), подписанный уникальным аппаратным ключом процессора. Этот отчет позволяет любой внешней стороне криптографически верифицировать, что вычисление было произведено подлинным, немодифицированным кодом внутри подлинного аппаратного анклава.

2. Устройства хранения, вывода и взаимодействия:

Центральный массив хранения данных (160) представляет собой сетевую систему хранения данных (SAN или NAS) с поддержкой RAID для отказоустойчивости. Здесь хранятся итоговые рейтинги QJR и другая служебная информация.

Аппаратный шлюз API (170) – физическое устройство, управляющее всеми внешними запросами к системе. Оно выполняет аппаратное ускорение шифрования (TLS offloading) и маршрутизацию запросов к массиву хранения данных (160) для получения данных. Шлюз также предоставляет доступ к криптографическим аттестатам, сгенерированным устройством (150).

Принтер для печати защищённых аттестатов (180): специализированное печатающее устройство, подключенное к шлюзу (170). Оно предназначено для вывода криптографических аттестатов на физический носитель, создавая материальное, защищенное от подделки доказательство целостности вычислений.

Пользовательский терминал (190) представляет собой любое устройство конечного пользователя (персональный компьютер, рабочая станция), которое через стандартный веб-браузер обращается к веб-интерфейсу, предоставляемому шлюзом API

(170), для визуализации рейтингов и их составляющих.

Работа предлагаемого ПАК:

Процесс обработки данных начинается с физических и цифровых источников. Бумажные носители (108) поступают на станцию оцифровки и распознавания документов (110), которая преобразует их в цифровой формат. Этот поток, вместе с данными из внешних цифровых источников (109) (например, Web of Science, Scopus), поступает на устройство приёма данных (120).

Унифицированный поток данных передается на сервер обработки графов (130), который формирует взвешенный граф цитирования. Далее граф служит основой для параллельной обработки в кластере параллельной аналитики (140). Кластер состоит из трёх независимых аппаратных узлов: узла сетевого анализа (141), который вычисляет показатель влияния цитирования (CIS); узла контентного анализа (142), который вычисляет показатель тематической близости (QSS); и узла контроля качества данных (143), который формирует показатель целостности данных (DIS).

Эти три показателя (CIS, QSS, DIS) передаются в качестве входных данных в центральный узел системы — верификационное устройство с поддержкой TEE (150). Именно внутри этого устройства (150), в изолированной и защищённой аппаратной среде, происходит финальная агрегация и расчёт итогового ранга (QJR).

Расчитанный ранг (QJR) сохраняется в центральном массиве хранения данных (160). Одновременно, устройство (150) генерирует криптографический аттестат, подтверждающий целостность вычислений. Этот аттестат может быть передан на принтер для печати защищённых аттестатов (180) для создания физической копии.

Доступ к итоговым рейтингам из хранилища (160) и к аттестату для верификации предоставляется через аппаратный шлюз API (170). Взаимодействие конечного пользователя с системой осуществляется через пользовательский терминал (190), который обращается к шлюзу API (170).

Таким образом, предлагаемый ПАК реализует полнофункциональный цикл обработки и верификации данных — от оцифровки и агрегации информации до получения итогового рейтинга и его подтверждения в защищённой среде.

Архитектура комплекса обеспечивает высокую степень надёжности, прозрачности и доказуемой достоверности результатов, что делает систему применимой в задачах, где критически важна объективность и верифицируемость вычислений.

ФОРМУЛА ПОЛЕЗНОЙ МОДЕЛИ

Программно-аппаратный комплекс для верифицируемой обработки библиометрических данных, ранжирования изданий и генерации интегрального рейтинга, который построен на взаимодействии устройств, обеспечивающих полный цикл обработки от оцифровки бумажных

носителей до предоставления криптографически заверенного итогового рейтинга, *отличающийся* тем, что ПАК дополнительно включает?

- станцию оцифровки и распознавания документов для обработки физических носителей, состоящую из высокоскоростного сканера документов, выделенного процессорно-вычислительного блока и терминала верификации;

- устройство приема данных, представляющее собой выделенный сетевой сервер, получающий данные от станции оцифровки и распознавания документов и внешних источников для передачи унифицированного потока данных далее по конвейеру;

- сервер обработки графов, оснащенный большим объемом оперативной памяти и быстрыми накопителями, преобразующий потоки данных в структуру взвешенного графа цитирования;

- кластер параллельной аналитики для генерации аналитических сигналов, состоящий из физических серверных узлов, параллельно обрабатывающие данные для генерации трех независимых показателей:

• узел сетевого анализа для вычисления показателя влияния цитирования (Citation Influence Score, CIS) посредством итеративного анализа структуры графа;

• узел контентного анализа для анализа текстов публикаций и вычисления показателя тематической близости (Quality-adjusted Similarity Score, QSS);

• узел контроля качества данных, функционирующий как аппаратный фильтр, анализируя граф на наличие нетипичных паттернов цитирования и формируя показатель целостности данных (Data Integrity Score, DIS);

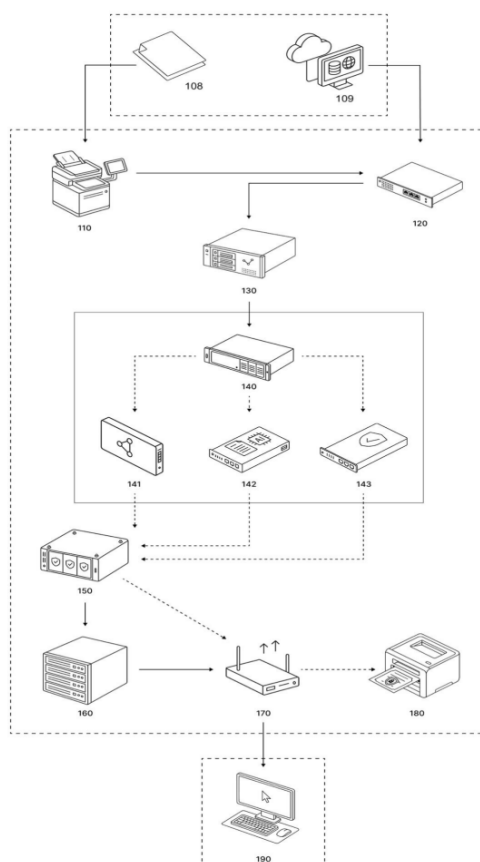
- верификационное устройство с поддержкой TEE, реализованное в виде физического сервера, процессор которого поддерживает аппаратную доверенную среду исполнения (Trusted Execution Environment, TEE);

- аппаратный шлюз API, представляющий собой устройство, управляющее всеми внешними запросами, для предоставления верифицируемого результата и публикации итогового рейтинга QJR;

- принтер для печати защищённых аттестатов для выведения отчета на материальный носитель;

- центральный массив хранения данных, представляющий собой сетевую систему хранения данных (SAN или NAS) с поддержкой RAID для отказоустойчивости, а также

- пользовательский терминал, который обращается к шлюзу API.



Верстка Д. Женысова
Корректор Г. Косанова