


 ФЕДЕРАЛЬНАЯ СЛУЖБА
 ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ПАТЕНТУ

(21)(22) Заявка: 2011116188/08, 26.09.2009

(24) Дата начала отсчета срока действия патента:
26.09.2009

Приоритет(ы):

(30) Конвенционный приоритет:
23.10.2008 US 12/257,091

(43) Дата публикации заявки: 27.10.2012 Бюл. № 30

(45) Опубликовано: 10.03.2014 Бюл. № 7

(56) Список документов, цитированных в отчете о
поиске: US 2007/0283125 A1, 06.12.2007. US
2006/0253682 A1, 09.11.2006. US 2007/0283124
A1, 06.12.2007. US 6970991 B2, 29.11.2005. US
2007/0283115 A1, 06.12.2007. RU 2285286 C2,
10.10.2006.(85) Дата начала рассмотрения заявки РСТ на
национальной фазе: 22.04.2011(86) Заявка РСТ:
US 2009/058511 (26.09.2009)(87) Публикация заявки РСТ:
WO 2010/047918 (29.04.2010)

Адрес для переписки:

129090, Москва, ул. Б. Спасская, 25, стр.3,
ООО "Юридическая фирма Городисский и
Партнеры"

(72) Автор(ы):

ФОЛЦ Форрест К. (US),
КАТЛЕР Дэвид Н. (US)

(73) Патентообладатель(и):

МАЙКРОСОФТ КОРПОРЕЙШН (US)

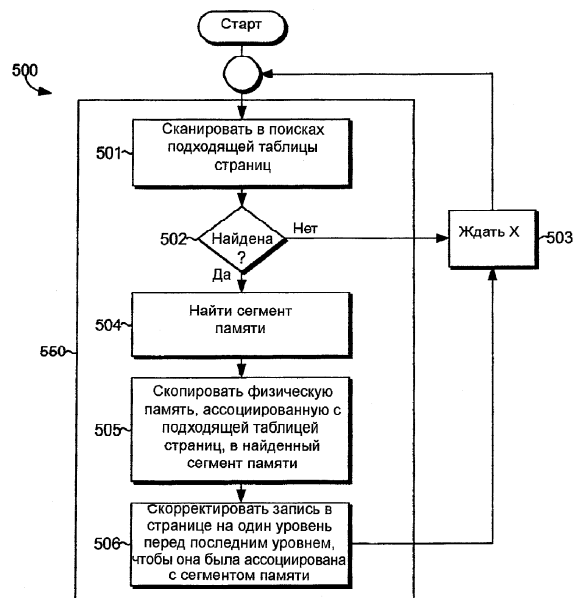
(54) ГИБКОЕ УВЕЛИЧЕНИЕ СТРАНИЦ ПАМЯТИ

(57) Реферат:

Изобретение относится к средствам преобразования множества малых страниц, ассоциированных с одним или более процессами, в большую страницу. Технический результат заключается в повышении эффективности вычислительной системы. Таблицы страниц на последнем уровне иерархической системы таблиц страниц сканируются в поисках подходящих таблиц страниц. Подходящие таблицы страниц преобразуются в большие страницы памяти,

имеющие запись таблицы страниц на один уровень перед последним уровнем иерархической системы таблиц страниц, скорректированную таким образом, чтобы она была ассоциирована с вновь созданной большой страницей. После приема уведомления, что большая страница должна быть преобразована в таблицу страниц, создается новая таблица страниц. Каждая запись в новой таблице страниц ассоциирована с малым сегментом памяти в большой странице, и запись в таблице страниц на один

уровень перед последним уровнем в иерархической системе таблиц страниц скорректирована, чтобы быть ассоциированной с новой таблицей страниц. 3 н. и 17 з.п. ф-лы, 7 ил.



Фиг.5



FEDERAL SERVICE
FOR INTELLECTUAL PROPERTY

(12) **ABSTRACT OF INVENTION**(21)(22) Application: **2011116188/08, 26.09.2009**(24) Effective date for property rights:
26.09.2009

Priority:

(30) Convention priority:
23.10.2008 US 12/257,091(43) Application published: **27.10.2012 Bull. 30**(45) Date of publication: **10.03.2014 Bull. 7**(85) Commencement of national phase: **22.04.2011**(86) PCT application:
US 2009/058511 (26.09.2009)(87) PCT publication:
WO 2010/047918 (29.04.2010)

Mail address:

**129090, Moskva, ul. B. Spasskaja, 25, str.3, OOO
"Juridicheskaja firma Gorodisskij i Partnery"**

(72) Inventor(s):

**FOLTs Forrest K. (US),
KATLER Dehvid N. (US)**

(73) Proprietor(s):

MAJKROSOFT KORPOREJShN (US)(54) **FLEXIBLE ENLARGEMENT OF MEMORY PAGES**

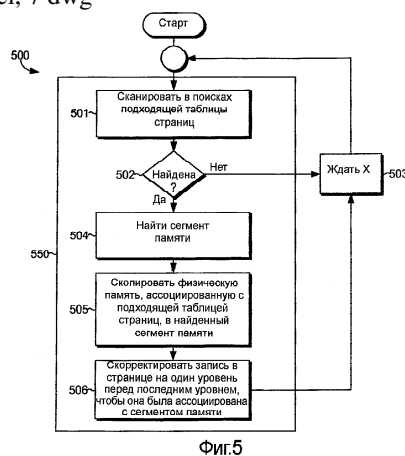
(57) Abstract:

FIELD: information technology.

SUBSTANCE: page tables at the last level of a hierarchical page table system are scanned in search of suitable page tables. The suitable page tables are converted to large memory pages, having a page table entry at a level before the last level of the hierarchical page table system adjusted to be associated with the newly created large page. Upon receiving a notification that a large page is to be converted into a page table, a new page table is created. Each entry in the new page table is associated with a small segment of memory in the large page and an entry in a page table at one level before the last level in a hierarchical page table system is adjusted to be associated with the new page

table.

EFFECT: high efficiency of the computer system.
20 cl, 7 dwg



Для выполнения процессов на вычислительных устройствах часто необходимы данные, которые будут использоваться в вычислениях. Эти данные обычно хранятся операционной системой в памяти, такой как RAM (запоминающее устройство с произвольным доступом). Эта память разбивается на части, называемые страницами. Каждая страница связана с уникальным адресом. Когда процессу требуются данные, к данным обращаются по их уникальному адресу, и адрес используется для поиска физического расположения страницы для возвращения данных. Один из распространенных способов такой трансляции адреса в физическое расположение выполняется посредством обхода иерархии таблицы переадресации страниц. В таких иерархиях поступаются размером адресуемых страниц памяти в пользу количества уровней иерархии. Однако размер страницы памяти также определяет, насколько эффективно используется пространство памяти, при этом большие страницы памяти являются менее эффективными. Поэтому существует прямое соотношение между эффективностью пространства (обусловленной размером страниц памяти) и эффективностью времени трансляции (обусловленной количеством страниц в иерархии таблиц переадресации страниц).

Дополнительный фактор определения эффективности системы таблиц переадресации страниц заключается в потребностях процессов. Если процессам обычно требуются большие объемы данных, то большие страницы могут на самом деле быть эффективными с точки зрения использования памяти. Однако если процессам обычно требуются небольшие объемы данных, тогда меньшие страницы будут более эффективными. Так как обычно в системе работают процессы и того, и другого типа, то способ динамической поддержки обоих приведет к большей эффективности. Поддержка больших страниц памяти операционной системой так же не столь надежна в вычислительных системах, как поддержка страниц памяти меньшего размера. Это приводит к дополнительной проблеме в использовании больших страниц памяти.

СУЩНОСТЬ ИЗОБРЕТЕНИЯ

Это изложение сущности изобретения приведено с целью представить в упрощенной форме выбор понятий, которые описаны далее в подробном описании. Это изложение сущности изобретения не направлено на идентификацию основных особенностей или важных особенностей заявленного объекта и также не направлено на использование при определении объема заявленной сущности изобретения.

Варианты осуществления настоящего изобретения относятся к сканированию последнего уровня в иерархии таблиц страниц с целью нахождения записей таблиц страниц (PTE) для преобразования в отображения больших страниц. Как только возможные (потенциальные) PTE найдены, эти PTE преобразуются в большие страницы памяти посредством нахождения больших непрерывных сегментов физической памяти, переноса данных, ассоциированных со всеми PTE в возможной странице таблицы страниц, в найденный сегмент памяти, и корректировка PTE в странице таблицы страниц на один уровень перед последним уровнем иерархии таблиц страниц таким образом, чтобы она была ассоциирована с вновь созданной большой страницей памяти. В некоторых вариантах осуществления при получении уведомления, указывающего большую страницу, которая должна быть преобразована обратно в малые страницы, создается новая страница таблицы страниц. Каждая PTE в новой странице таблиц страниц связана с малым сегментом большой страницы, и PTE в таблице страниц на один уровень перед последним уровнем иерархии таблиц страниц корректируется таким образом, чтобы она была ассоциирована с новой страницей

таблицы страниц.

КРАТКОЕ ОПИСАНИЕ ЧЕРТЕЖЕЙ

Настоящее изобретение подробно описано ниже со ссылками на приложенные чертежи, на которых:

5 На Фиг.1 изображена блок-схема иллюстративного вычислительного устройства, пригодного для использования в реализации настоящего изобретения;

на Фиг.2 представлена диаграмма типичной схемы распределения ячеек памяти, которая используется операционными системами и пользовательскими процессами;

10 на Фиг.3 изображена иллюстративная связь между таблицей переадресации страниц и физической памятью;

на Фиг.4 изображена иллюстративная иерархическая система таблиц страниц;

на Фиг.5 представлена блок-схема, показывающая способ поиска возможных таблиц страниц для преобразования в большие страницы, и выполнение преобразования;

на Фиг.6 показана блок-схема, показывающая способ приема уведомления, что большая страница должна быть преобразована в таблицу страниц, ассоциированную с малыми страницами, и выполнение преобразования, и

20 на Фиг.7 представлена блок-схема, показывающая способ приема уведомления, что большая страница памяти должна быть преобразована в таблицу страниц, ассоциированную с малыми страницами, и приема тайм-аута, указывающего, что пришло время сканировать возможные таблицы страниц для преобразования в большие страницы.

ПОДРОБНОЕ ОПИСАНИЕ

25 Сущность настоящего изобретения здесь описана с детализацией для удовлетворения нормативных требований. Однако описание само по себе не направлено на ограничение объема этого патента. Напротив, авторы предполагают, что заявленная сущность изобретения может также быть осуществлена другими способами, включающими в себя другие этапы или комбинации этапов, похожих на те, что описаны в данном документе, в сочетании с другими существующими или будущими технологиями. Более того, хотя термины «шаг» и/или «блок» здесь могут быть использованы для обозначения различных элементов используемых методов, эти термины не должны интерпретироваться как подразумевающие какой-либо конкретный порядок различных этапов, раскрываемых здесь, если только порядок отдельных этапов явно не описан.

30 Варианты осуществления настоящего изобретения направлены на гибкое нахождение групп РТЕ, которые могут быть преобразованы в большие страницы, и выполнение преобразования. Дополнительно, как только страница таблицы страниц преобразована в большую страницу, в качестве реакции на уведомление от операционной системы может быть выполнена обратная операция.

45 В соответствии с некоторыми вариантами осуществления настоящего изобретения подсистема памяти вычислительного устройства управляет ресурсами совместно используемой памяти. Данные, требующиеся для вычислений одному или нескольким процессам, хранятся в ресурсах совместно используемой памяти. Как правило, процессы, выполняемые на вычислительных устройствах, не осведомлены о физическом местоположении данных. Вместо этого, эти процессы представлены с адресным пространством, отображающим адреса на физические местоположения в памяти. Один или несколько процессов, выполняемых на вычислительном устройстве, используют адрес для обращения к данным, требуемым для вычислений. Подсистема

памяти вычислительного устройства выполняет трансляцию из адреса в физическое местоположение, выполняя преобразования адресов.

В современных вычислительных устройствах физическая память разделена на сегменты, называемые страницами. Эти страницы представляют минимальный объем
5 данных, который может быть представлен иерархией таблиц страниц. Таблицы страниц используются подсистемой памяти вычислительного устройства для отображения виртуальных адресов на физические местоположения в памяти.

Существует множество возможных структур для систем таблиц страниц; однако
10 наиболее распространенные отображения адресов на физические местоположения памяти используют множественные, иерархические преобразования таблицы переадресации страниц, которые подробно описаны ниже. Эти иерархии допускают фиксированные размеры адресов (как правило, измеряемые в битах) для адресации больших объемов физической памяти. Такие поиски по иерархической таблице
15 требуют нескольких обращений к памяти для нахождения физической страницы, ассоциированной с данным виртуальным адресом. Чем больше уровней в иерархической системе таблиц страниц, тем более дорогостоящими являются операции доступа к памяти с точки зрения времени трансляции из адреса в физическую память.

Однако существует компромисс между количеством уровней в иерархии таблиц
20 страниц и размером страницы памяти. Меньшее количество уровней в иерархии таблиц страниц означает больший размер страницы. Таким образом, для приложений, использующих небольшие сегменты данных, небольшие размеры страниц, и поэтому более глубокие иерархии, позволяют уменьшить излишние траты памяти. Однако для
25 приложений, использующих большие объемы данных, большой размер страниц уменьшит количество обращений к таблице страниц, требуемых для нахождения необходимых данных, и, таким образом, увеличит эффективность поиска.

Когда отдельная часть данных более не требуется, или к ней не было обращений в
30 течение порогового периода времени, общим для подсистем памяти является сохранение этой части данных на диск, освобождая память для данных, которые необходимы чаще, или необходимы в настоящий момент. Этот процесс называется выгрузкой памяти. Однако многие подсистемы памяти могут выгружать только
35 некоторый фиксированный размер страниц. Поэтому любой механизм, создающий страницы больше этого фиксированного размера, должен иметь возможность разбивать большие страницы на множество страниц меньшего размера в случае, если будет необходимо выгрузить часть большой страницы. Существует множество других дополнительных ситуаций, в которых необходимо разбить большую страницу на
40 страницы меньшего размера посредством подсистемы памяти.

Соответственно, вариант осуществления изобретения направлен на
машиночитаемые носители хранения данных, воплощающие используемые
компьютером инструкции для выполнения способа преобразования множества малых
страниц, ассоциированных с одним или несколькими процессами, выполняемыми на
45 вычислительном устройстве, в большую страницу. Каждая из страниц связана с элементом таблицы страниц из иерархической системы таблиц страниц, содержащей по меньшей мере два уровня таблиц страниц. Способ включает сканирование последнего уровня иерархической системы таблиц переадресации страниц в поиске
возможных РТЕ, которые являются таблицами страниц, содержащими по меньшей
50 мере пороговое количество записей, ассоциированных со страницами. Способ далее находит физически непрерывный сегмент памяти, достаточно большой для хранения каждой из страниц, ассоциированных с записями в подходящей таблице страниц, и

копирует сегменты памяти из каждой страницы в найденный сегмент памяти. Способ корректирует запись таблицы страниц в таблице страниц на один уровень перед последним уровнем иерархической системы таблиц страниц так, чтобы она была связана с вновь созданной большой страницей.

В соответствии с другими вариантами осуществления изобретение направлено на машиночитаемые носители, хранящие выполнимые компьютером инструкции, реализующие способ преобразования большой страницы в множество маленьких страниц, ассоциированных с одним или несколькими процессами, выполняемыми на вычислительной системе. Каждая из страниц ассоциирована с записью таблицы страниц в иерархической системе таблиц страниц. Способ включает получение уведомления операционной системы, указывающего, что большая страница должна быть преобразована в группу малых страниц. После получения уведомления создается новая таблица страниц, и элементы новой таблицы страниц ассоциируются с малыми сегментами большой страницы. Способ включает корректировку записи из таблицы страниц на один уровень перед последним уровнем иерархической системы таблиц страниц таким образом, чтобы она была ассоциирована с новой таблицей страниц.

В соответствии с еще одним вариантом осуществления изобретение направлено на машиночитаемые носители, хранящие выполнимые компьютером инструкции, реализующие способ сканирования последнего уровня иерархической системы таблиц страниц, содержащей по меньшей мере два уровня таблиц страниц, в каждом из множества адресных пространств, ассоциированных с одним или несколькими процессами, выполняемыми в вычислительной системе. Это сканирование включает попытку идентифицировать подходящие таблицы страниц, которые являются таблицами страниц, для которых каждая запись ассоциирована с одним или несколькими сегментами физической памяти. Способ также включает поиск сегмента памяти, состоящего из смежных сегментов физической памяти, достаточно больших для хранения каждого из множества сегментов физической памяти, ассоциированных со всеми записями в подходящей таблице страниц, и копирование этих сегментов физической памяти во вновь найденный сегмент памяти. Способ освобождает сегмент памяти, содержащий подходящую таблицу страниц и корректирует запись таблицы страниц в таблице страниц на один уровень перед последним уровнем в иерархической системе таблиц страниц, которая была ассоциирована с подходящей таблицей страниц таким образом, чтобы она была ассоциирована с вновь найденным сегментом памяти, называемым большой страницей. Способ также включает получение указания от подсистемы памяти, неспособной выгружать большие страницы, которое указывает, что один или несколько сегментов большой страницы должны быть выгружены. Способ также включает создание новой таблицы страниц, где каждая запись новой таблицы страниц ассоциирована с сегментом большой страницы, содержащей сегмент или сегменты, которые должны быть выгружены. Способ также включает корректирование записи таблицы страниц в таблице страниц на один уровень перед последним уровнем иерархической системы таблиц страниц, которая была ранее ассоциирована с большой страницей, чтобы она была ассоциирована с новой таблицей страниц.

После кратко описанного обзора вариантов осуществления настоящего изобретения ниже, с целью предоставления общего контекста для различных аспектов настоящего изобретения, описана иллюстративная операционная среда, в которой варианты осуществления настоящего изобретения могут быть реализованы. На Фиг. 1 показана иллюстративная операционная среда для реализации вариантов

осуществления настоящего изобретения, обозначенная в целом как вычислительное устройство 100. Вычислительное устройство 100 является примером подходящего вычислительного окружения и не предназначено для наложения какого-либо ограничения на объем использования или функциональность изобретения. Также, вычислительное устройство 100 не должно быть интерпретировано как имеющее какие-либо зависимости или требования относительно любого показанного компонента или их комбинации.

Изобретение может быть описано в общем контексте машинного кода или готовых к машинному использованию инструкций, включая выполняемые компьютером инструкции, такие как программные модули, исполняемые компьютером или другим устройством, таким как карманный компьютер или другое портативное устройство. Как правило, программные модули включают подпрограммы, программы, объекты, компоненты, структуры данных и т.д., относящиеся к коду, выполняющему конкретные задачи или реализующие конкретные абстрактные типы данных. Изобретение может применяться во многих конфигурациях систем, включая портативные устройства, бытовую электронику, компьютеры общего назначения, более специализированные вычислительные устройства и т.д. Изобретение может также применяться в распределенных вычислительных окружениях, в которых задачи выполняются устройствами дистанционной обработки данных, которые связаны через коммуникационную сеть.

Согласно Фиг. 1, вычислительное устройство 100 включает шину 110, которая явно или неявно связывает следующие устройства: память 112, один или несколько процессоров 114, один или несколько внешних компонентов хранения данных 116, порты ввода-вывода (I/O) 118, компоненты ввода 120, компоненты вывода 121, и иллюстративный источник электропитания 122. Шина 110 представляет собой то, что может быть одной или несколькими шинами (например, шина адресов, шина данных, или их комбинация). Хотя различные блоки на Фиг. 1 показаны для ясности линиями, в реальности разграничение различных компонентов не так ясно, и, метафорически, для большей точности линии могут быть серыми или нечеткими. Например, многие процессоры имеют память. Понятно, что такова специфика данной области техники и что диаграмма на Фиг.1 является всего лишь иллюстрацией вычислительного устройства, которое может быть использовано в связи с одним или несколькими вариантами осуществления настоящего изобретения. Не делается различий между такими категориями, как «рабочая станция», «сервер», «ноутбук», «портативное устройство» и т.д., так как все они предусмотрены в рамках Фиг.1 и ссылки на «вычислительное устройство».

Вычислительное устройство 100, как правило, включает различные машиночитаемые носители данных. Машиночитаемые носители данных могут быть любыми доступными носителями данных, к которым можно осуществлять доступ с вычислительного устройства 100, и включают энергозависимые и энергонезависимые носители, съемные и несъемные носители. В качестве примера, но не ограничения, машиночитаемые носители могут содержать компьютерные устройства хранения данных и коммуникационную среду. Компьютерное устройство хранения данных включает как энергозависимые, так и энергонезависимые, съемные и несъемные носители, выполненные любым способом или технологией для хранения данных, таких как машиночитаемые инструкции, структуры данных, программные модули и другие данные. Компьютерные устройства хранения данных включают, но не ограничены: RAM, ROM, EEPROM, флэш-память или другие технологии памяти, CD-

ROM, универсальный цифровой диск DVD или другие оптические диски, магнитные кассеты, магнитная лента, магнитные дисковые носители или другие магнитные устройства хранения данных, или любые другие носители, которые могут быть использованы для хранения желаемых данных и к которым может быть осуществлен доступ с вычислительного устройства 100.

Память 112 включает компьютерное устройство хранения данных в форме энергозависимой памяти. Иллюстративные аппаратные устройства включают в себя твердотельную память, например RAM. Внешнее запоминающее устройство 116 включает в себя компьютерное устройство хранения данных в форме энергонезависимой памяти. Память может быть съемной, несъемной или комбинацией последних. Иллюстративные аппаратные устройства включают в себя твердотельную память, жесткие диски, оптические диски и т.д. Вычислительное устройство 100 содержит один или несколько процессоров, которые читают данные из нескольких объектов, таких как память 112, внешнее запоминающее устройство 116, или компоненты ввода 120. Компоненты вывода 121 представляют отображение данных пользователю или другим устройствам. Типичные компоненты вывода включают в себя дисплей, динамик, компонент печати, вибрирующий компонент и т.д.

Порты ввода-вывода 118 позволяют логически связать вычислительное устройство 100 с другими устройствами, включая компоненты ввода 120 и компоненты вывода 121, часть из которых могут быть встроенными. Типичные компоненты включают в себя микрофон, джойстик, игровой контроллер, спутниковую антенну, сканер, принтер, беспроводное устройство и т.д.

В соответствии с вариантом осуществления изобретения, вычислительное устройство 100 может использоваться в качестве гипервизора, который является платформой виртуализации, абстрагирующей физические компоненты вычислительного устройства 100, такие как компоненты ввода 120 и память 112, от операционной системы или системы, выполняемой на вычислительном устройстве 100. Такой гипервизор позволяет запускать несколько операционных систем на одном вычислительном устройстве 100 посредством такого абстрагирования, позволяющего каждой независимой операционной системе иметь доступ к своей виртуальной машине. В гипервизорных вычислительных устройствах издержки, связанные с перемещением иерархий таблиц страниц, еще больше, и выигрыш от использования больших страниц даже больше, чем в системах, в которых запущена единственная операционная система, которая имеет прямой доступ к компонентам вычислительного устройства 100.

Согласно Фиг.2, физическая память 200, такая как RAM, разделена на некоторое количество секций. В соответствии с некоторыми вариантами осуществления изобретения, память разделена на два основных раздела - пространство памяти операционной системы 201 и пользовательское пространство памяти 202. Подсистема памяти операционной системы, выполняемой на вычислительном устройстве, управляет физической памятью 200, позволяя приложениями пользователя использовать части пользовательского пространства памяти 202. Однако приложения могут не иметь доступа к смежным расположениям памяти. Согласно Фиг.2, в соответствии с некоторыми вариантами осуществления настоящего изобретения, пользовательское пространство памяти 202 разделено на две страницы (обозначенные четырехугольниками), которые распределены между двумя гипотетическими приложениями с целью иллюстрации, но не ограничения: пространство приложения 1 обозначено посредством x (например, сегмент памяти 203), а пространство

приложения 2 обозначено посредством / (например, сегмент памяти 204). Свободные страницы не заполнены на диаграмме (например, сегмент памяти 205). Пространство памяти операционной системы 201 может быть использовано для разных целей, одной из которых является хранение таблиц страниц 206, содержащих отображение
 5 адресного пространства на физическую память. Для приложений эти отображения ассоциируют страницы в пользовательском пространстве памяти 202, где хранятся данные, с адресами.

Как показано на Фиг.3, в соответствии с вариантом осуществления настоящего
 10 изобретения, таблица страниц 301 включает записи 303, каждая из которых ассоциирована с конкретной страницей в пользовательском пространстве памяти, хранимой в физической памяти 304. Записи 303 в таблице 301 страниц не обязательно могут быть ассоциированы со смежными страницами 304 физической памяти.

На Фиг.4, в соответствии с различными вариантами осуществления настоящего
 15 изобретения, адреса 401 представлены строками битов. Эти адреса отображаются посредством иерархической системы 402 таблиц страниц. В качестве примера, но не ограничения, рассмотрим 48-битную адресную схему 401 и четырехуровневую иерархическую систему таблиц 402 страниц. 48-битный адрес 401 разделен на пять
 20 секций. Первые девять бит 403 используются в качестве указателя в первую таблицу 404 страниц. Запись, находящаяся в первой таблице 404 страниц посредством первых девяти бит 403 адреса 401, ассоциирована с сегментом памяти 422, хранящим вторую таблицу 406 страниц. Вторые девять бит 405 служат указателем во вторую таблицу 406 страниц. Запись, находящаяся во второй таблице 406 страниц,
 25 ассоциирована с сегментом памяти 422, содержащим третью таблицу 408 страниц. Третьи девять бит 407 адреса 401 служат указателем в третью таблицу 408 страниц. Запись, находящаяся в третьей таблице 408 страниц, ассоциирована с сегментом памяти 423, содержащим четвертую таблицу 410 страниц. Четвертые девять битов 409
 30 адреса 401 служат указателем в четвертую таблицу 410 страниц. Запись, находящаяся в четвертой таблице 410 страниц, ассоциирована с сегментом памяти 424 в пользовательском пространстве памяти, содержащем страницу 412. Последние двенадцать бит 411 адреса 411 служат указателем в страницу 412. Сегмент памяти в
 35 странице 412 по указателю, заданному последними двенадцатью битами 412, является данными, на которые ссылается адрес 401. Как видно, имеется как минимум один доступ к памяти на поиск по таблице страниц в процессе поиска данных, адресуемых посредством иерархической системы таблиц страниц.

Специалистам в данной области должно быть понятно, что конкретный размер
 40 адреса, количество таблиц страниц, количество уровней в иерархической системе таблиц страниц и размер страниц могут варьироваться. Исключительно в качестве примера, но не ограничения, размеры страниц могут быть 4 КБ, 2 МБ или 1 ГБ. Размер адреса может быть, например, в пределах от 32 бит до 64 бит. В примере по Фиг.4, каждая таблица страниц содержит 512 записей (2^9), а каждая страница - это 4
 45 КБ (2^{12}). Нахождение данных на странице памяти занимает четыре поиска по таблице страниц. Если бы все данные, ассоциированные со всеми 512 записями в таблице страниц, были бы объединены в единую страницу, то полученная страница (называемая большой страницей) была бы размером 2 МБ, и для нахождения в
 50 иерархической системе таблиц страниц требовалось бы всего три поиска по таблице страниц.

На Фиг.5 представлена блок-схема, которая иллюстрирует способ 500 для поиска подходящей таблицы страниц, чтобы преобразовать в большую страницу, и таким

образом преобразование таблицы страниц (блок 550 содержит этапы способа без части тайм-аута, показанной в блоке 503, все обсуждается ниже). Показанный в блоке 501 последний уровень иерархической системы таблиц страниц сканируется в

поисках подходящей таблицы страниц для преобразования в большую страницу.

Например, последний уровень иерархии таблиц страниц из Фиг.4, в которой существует четвертая таблица 410 страниц, может сканироваться в поисках подходящей таблицы страниц. Специалисту в данной области должно быть понятно, что для определения того, является ли таблица страниц подходящей для

преобразования в большую страницу, может быть использован широкий спектр критериев. В качестве примера, но не ограничения, такие критерии могут включать поиск полностью заполненной таблицы страниц, или поиск таблицы страниц, заполненной пороговым количеством записей. Полностью заполненная таблица

страниц - это такая таблица страниц, в которой все записи ассоциированы с расположениями в физической памяти. В соответствии с одним вариантом осуществления изобретения, такое сканирование предполагает сканирование каждой из таблиц страниц, связанных с записями в таблице страниц на один уровень перед последним уровнем, и проверку найденных таблиц страниц последнего уровня, чтобы определить, являются ли они полностью заполненными таблицами страниц.

Специалисту в данной области должно быть понятно, что существует множество способов определения порога, включая, но не ограничиваясь указанным, процент записей, ассоциированных с физическими ячейками памяти, или общее число записей, ассоциированных с физическими ячейками памяти.

Посредством сканирования последнего уровня иерархической системы таблиц страниц (например, уровень, на котором расположена таблица 410 страниц на Фиг.4), могут быть идентифицированы одна или несколько подходящих таблиц страниц (см. блок 502). Если не найдено ни одной подходящей таблицы страниц, то выполняется временная задержка 503 перед очередным выполнением сканирования в блоке 501. Эта временная задержка 503 является параметром, который может настраиваться программистом, системным администратором, пользователем или кем-либо еще с определенным уровнем доступа к системе. Если, однако, подходящая таблица страниц была найдена, то находится сегмент непрерывной памяти, достаточно большой для хранения данных, ассоциированных с каждой записью в подходящей таблице страниц, как показано в блоке 504.

В вариантах осуществления поиск сегмента памяти требует сканирования физической памяти в поисках достаточного количества смежных сегментов памяти для хранения всех записей, ассоциированных с подходящей таблицей страниц.

Напомним, что таблица страниц может не иметь соседних записей, ассоциированных со смежными сегментами физической памяти. Однако, когда записи в подходящей таблице страниц преобразуются в большую страницу, они должны храниться в порядке записей в таблице страниц, с которой они ассоциированы. В соответствии с одним вариантом осуществления изобретения нахождение сегмента памяти - это просто вопрос сканирования физической памяти и поиска большого непрерывного сегмента памяти (например, 2 МБ). В некоторых вариантах осуществления это сканирование может выполняться посредством сканирования базы данных опорных номеров страниц, содержащей состояния всех физических страниц в системе.

Дополнительно, большой непрерывный сегмент памяти может быть ограничен таким образом, чтобы он начинался с predetermined границы байта. В качестве примера, но не ограничения, рассматривая пример, приведенный выше, при использовании 512

4 КБ страниц малого размера для комбинирования в большую страницу размером 2 МБ, предопределенная граница байта может быть 2 МБ границей байта. Специалисту в данной области должно быть понятно, что могут использоваться многие другие значения предопределенной границы байта. Согласно другому варианту осуществления изобретения, если может быть найдено недостаточное число смежных сегментов памяти, тогда активируется подпрограмма управления памятью, которая создает большой непрерывный сегмент памяти путем перемещения сохраненных данных в свободные сегменты из конкретного местоположения в памяти и

5

10

корректирования соответствующих им записей таблицы страниц. Таким образом создается большой непрерывный сегмент памяти для использования при преобразовании таблицы больших страниц.

Как только непрерывный сегмент памяти достаточного размера был найден или создан, все физические сегменты памяти, ассоциированные с записями в подходящей

15

таблице страниц, копируются надлежащим образом в найденный сегмент памяти, как показано в блоке 505. В одном варианте осуществления настоящего изобретения, после того, как физические сегменты памяти скопированы в найденный сегмент, исходное расположение физической памяти освобождается. В другом варианте

20

осуществления изобретения, исходные расположения памяти каждого из сегментов памяти, ассоциированных с каждой из записей подходящей таблицы страниц, также оставляют свои копии данных.

Как показано в блоке 506, запись таблицы страниц на один уровень перед последним уровнем иерархической системы таблиц страниц (например, таблицы 408

25

страниц на Фиг.4) ассоциирована с новой большой страницей. В одном варианте осуществления изобретения преобразованная таблица страниц освобождается, и запись таблицы страниц на один уровень перед последним уровнем иерархической системы таблиц страниц, которая была ассоциирована с освобожденной таблицей

30

страниц, корректируется таким образом, чтобы быть ассоциированной с новой большой страницей. После преобразования кандидатов в большие страницы, существует временная задержка в блоке 503 перед запуском нового сканирования в поисках новой подходящей таблицы страниц. Эта временная задержка в блоке 503 является параметром, который может настраиваться программистом, системным

35

администратором, пользователем или кем-либо еще, имеющим соответствующий доступ к системе.

На Фиг.6 представлена блок-схема, которая иллюстрирует способ 600 для преобразования большой страницы в записи таблицы страниц, связанных с

40

множеством страниц меньшего размера. В соответствии с одним вариантом осуществления настоящего изобретения, уведомление операционной системы получается в блоке 601, указывающее, что большая страница должна быть преобразована в малые страницы. Специалисту в данной области должно быть

45

понятно, что существует множество событий, которые могут активировать такие уведомления. Только в качестве примера, но не ограничения, такие события включают сегмент большой страницы, запланированный к выгрузке на диск системой с операционной системой, неспособной выгружать большие страницы, и записи таблицы страниц, ассоциированные с памятью, принадлежащей пространству памяти

50

приложения, которое было уничтожено.

После получения уведомления, указывающего, что большая страница должна быть преобразована, создается новая таблица страниц, как показано в блоке 602. В соответствии с одним вариантом осуществления изобретения это создание

предполагает выделение памяти в пространстве памяти операционной системы для новой таблицы. Как только таблица страниц создана, каждая запись новой таблицы страниц ассоциируется с одним сегментом меньшего размера большой страницы в блоке 603, пока все сегменты большой страницы не будут ассоциированы с некоторой записью в новой таблице страниц. Продолжая пример из Фиг.4, каждая из 512 записей таблицы страниц должна быть ассоциирована с одним 4 КБ сегментом большой страницы.

Наконец, запись таблицы страниц на один уровень перед последним уровнем иерархической системы таблиц страниц (например, уровень, на котором находится таблица страниц 408 из Фиг.4) корректируется таким образом, чтобы она была ассоциирована с новой таблицей страниц, как показано в блоке 604. В соответствии с одним вариантом осуществления изобретения записью из таблицы страниц на один уровень перед последним уровнем иерархической системы таблиц страниц, ассоциированной с новой таблицей страниц, была запись, прежде ассоциированная с большой страницей.

В соответствии с другим вариантом осуществления изобретения, на Фиг.7 представлен способ 700 преобразования таблицы страниц в большую страницу и преобразования больших страниц в таблицы страниц, ассоциированные с множеством страниц малого размера. Сначала способ предполагает ожидание события, как показано в блоке 701. В качестве примера, но не ограничения, событием может быть либо истечение времени ожидания (тайм-аут), либо уведомление операционной системы. Специалисту в данной области должно быть понятно, что существует множество других событий, которые могут активировать тот или другой тип преобразования. Как только событие происходит, принимается решение. Если событием был тайм-аут, указывающий что время задержки истекло 702, предпринимается попытка преобразования таблицы страниц из последнего уровня иерархической системы таблиц страниц в большую страницу, например в соответствии со способом 550 из Фиг. 5. Эта временная задержка является параметром, который может настраиваться программистом, системным администратором, пользователем или кем-либо еще, имеющим соответствующий допуск к системе. Если событием является уведомление операционной системы 702, тогда большая страница преобразуется в таблицу страниц, содержащую записи, указывающие на страницы меньшего размера, например, в соответствии со способом 600 на Фиг. 6. После завершения либо способа, пытающегося преобразовать таблицу страниц в большую страницу, либо способа преобразования большой страницы в таблицу страниц с записями, ассоциированными с множеством страниц меньшего размера, снова начинается период ожидания в блоке 701. Этот период ожидания снова истекает либо при получении другого уведомления операционной системы, либо при истечении временной задержки.

Множество различных конфигураций различных описанных компонентов, равно как и не показанных компонентов, возможны без отклонения от сущности и объема настоящего изобретения. Варианты осуществления настоящего изобретения были описаны с намерением проиллюстрировать, но не ограничить. Альтернативные варианты осуществления, которые не отклоняются от своей сферы, будут очевидны для специалистов в данной области. Специалист может разработать альтернативные средства для реализации вышеупомянутых улучшений, не отклоняясь от объема настоящего изобретения.

Следует понимать, что некоторые признаки и подкомбинации являются полезными

и могут быть реализованы без обращения к другим признакам и подкомбинациям и рассматриваются в рамках формулы изобретения. Не все этапы, перечисленные в различных чертежах, должны выполняться в конкретном описанном порядке.

5

Формула изобретения

1. Один или более машиночитаемых носителей, на которых хранятся выполняемые компьютером инструкции для выполнения способа преобразования множества малых страниц, ассоциированных с одним или более процессами, выполняемыми в
10 вычислительной системе, в большую страницу, причем каждая из множества малых страниц ассоциирована с одной из множества записей таблицы страниц из иерархической системы таблиц страниц, содержащей по меньшей мере два уровня таблиц страниц, причем способ включает в себя:

15 сканирование последнего уровня иерархической системы страниц в поиске таблицы страниц, в которой каждая из по меньшей мере порогового количества множества записей ассоциирована с одной из множества страниц, приводящее к идентификации подходящей таблицы страниц;

20 нахождение сегмента памяти, составленного из множества смежных сегментов физической памяти, достаточно больших для хранения каждого из множества сегментов физической памяти, ассоциированных со всеми из множества записей подходящей таблицы страниц;

25 копирование каждого из множества сегментов физической памяти, ассоциированных со всеми из множества записей подходящей таблицы страниц, в сегмент памяти, составленный из множества смежных сегментов физической памяти, и корректировку записи таблицы страниц в таблице страниц на один уровень перед последним уровнем иерархической системы таблиц страниц таким образом, чтобы она была ассоциирована с сегментом памяти, составленным из множества смежных
30 сегментов физической памяти.

2. Носитель по п.1, в котором сканирование последнего уровня иерархической системы таблиц страниц включает в себя выборочное сканирование каждого из множества адресных пространств, ассоциированных с одним или несколькими процессами, выполняемыми в вычислительной системе.

35 3. Носитель по п.1, в котором пороговое количество множества записей составляет общее число записей во всем множестве записей.

40 4. Носитель по п.1, в котором каждая из множества записей подходящей таблицы страниц ассоциирована с одним из одного или более процессов, выполняемых в вычислительной системе.

5. Носитель по п.1, в котором нахождение сегмента памяти, состоящего из множества смежных сегментов физической памяти, включает в себя копирование данных из первого расположения вблизи области физической памяти во второе расположение, удаленное от указанной области, для создания множества
45 смежных сегментов физической памяти, достаточно больших для хранения каждого из множества сегментов физической памяти, ассоциированных со всеми из множества записей в подходящей таблице страниц.

50 6. Носитель по п.1, в котором сегмент памяти, состоящий из множества смежных сегментов физической памяти, находится на предопределенной границе байта.

7. Носитель по п.1, в котором копирование каждого из множества сегментов физической памяти, ассоциированных со всеми из множества записей подходящей таблицы страниц, также включает в себя освобождение указанного множества

сегментов физической памяти после того, как они были скопированы.

8. Носитель по п.1, в котором упомянутая запись таблицы страниц в таблице страниц на один уровень перед последним уровнем в иерархической системе таблиц страниц была ранее ассоциирована с подходящей таблицей страниц.

9. Носитель по п.1, в котором корректировка записи таблицы переадресации страниц также включает в себя освобождение сегмента памяти, содержащего подходящую таблицу страниц.

10. Один или более машиночитаемых носителей, на которых хранятся исполняемые компьютером инструкции для выполнения способа, включающего в себя:

сканирование последнего уровня иерархической системы таблиц страниц, содержащей по меньшей мере два уровня таблиц страниц, в каждом из множества адресных пространств, ассоциированных с одним или более процессами, выполняемыми в вычислительной системе, в поисках таблицы страниц, в которой каждая из по меньшей мере порогового числа множества записей ассоциирована с одним или более сегментами физической памяти, приводящее к идентификации подходящей таблицы страниц;

нахождение сегмента памяти, состоящего из множества смежных сегментов физической памяти, достаточно больших для хранения каждого из множества сегментов физической памяти, ассоциированных со всеми из множества записей подходящей таблицы страниц;

копирование каждого из множества сегментов физической памяти, ассоциированных со всеми из множества записей подходящей таблицы страницы, в сегмент памяти, состоящий из множества смежных сегментов физической памяти;

освобождение сегмента памяти, содержащего подходящую таблицу страниц;

корректирование записи таблицы страниц в таблице страниц на один уровень перед последним уровнем иерархической системы таблиц страниц, причем упомянутая запись таблицы страниц была прежде ассоциирована с подходящей таблицей страниц, чтобы она была ассоциирована с сегментом памяти, состоящим из множества смежных сегментов физической памяти;

прием указания от подсистемы памяти, неспособной выгружать большие страницы, указывающего, что сегмент большой страницы должен быть выгружен;

создание новой таблицы страниц, где каждая запись в упомянутой новой таблице страниц ассоциирована с сегментом большой страницы, и

корректирование записи таблицы страниц в таблице страниц на один уровень перед последним уровнем иерархической системы таблиц страниц, чтобы она была ассоциирована с новой таблицей страниц, причем упомянутая запись таблицы страниц в таблице страниц на один уровень перед последним уровнем иерархической системы таблиц страниц была ранее ассоциирована с большой страницей.

11. Носитель по п.10, в котором иерархическая система таблиц страниц содержит 4 уровня.

12. Носитель по п.10, в котором каждая малая страница имеет размер 4 КБ.

13. Носитель по п.10, в котором каждая большая страница имеет размер 2 МБ.

14. Носитель по п.10, в котором сегмент памяти, состоящий из множества смежных сегментов физической памяти, находится на предопределенной границе байта.

15. Носитель по п.14, в котором предопределенная граница байта соответствует 2 МБ.

16. Носитель по п.10, в котором таблица страниц иерархической системы таблиц страниц адресуется в соответствии с 64-битной архитектурой.

17. Вычислительное устройство, содержащее процессор, сконфигурированный для выполнения способа преобразования множества малых страниц, ассоциированных с одним или более процессами, в большую страницу, причем каждая из множества малых страниц ассоциирована с одной из множества записей таблицы страниц из иерархической системы таблиц страниц, содержащей по меньшей мере два уровня таблиц страниц, причем способ включает в себя:

сканирование последнего уровня иерархической системы страниц в поиске таблицы страниц, в которой каждая из по меньшей мере порогового количества множества записей ассоциирована с одной из множества страниц, приводящее к идентификации подходящей таблицы страниц;

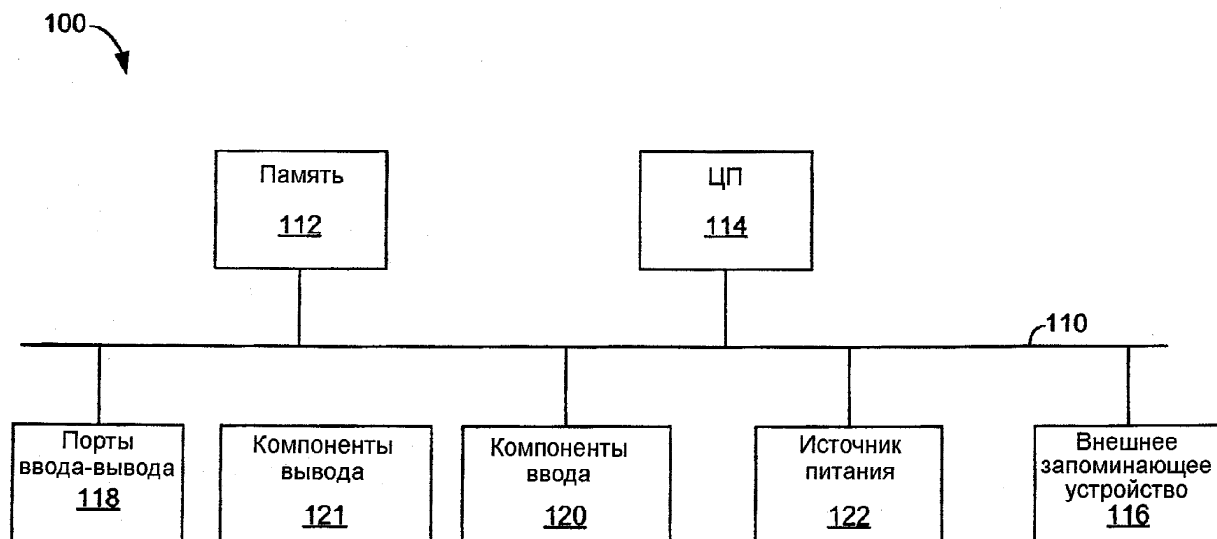
нахождение сегмента памяти, составленного из множества смежных сегментов физической памяти, достаточно больших для хранения каждого из множества сегментов физической памяти, ассоциированных со всеми из множества записей подходящей таблицы страниц;

копирование каждого из множества сегментов физической памяти, ассоциированных со всеми из множества записей подходящей таблицы страниц, в сегмент памяти, составленный из множества смежных сегментов физической памяти, и корректировку записи таблицы страниц в таблице страниц на один уровень перед последним уровнем иерархической системы таблиц страниц, чтобы она была ассоциирована с сегментом памяти, составленным из множества смежных сегментов физической памяти.

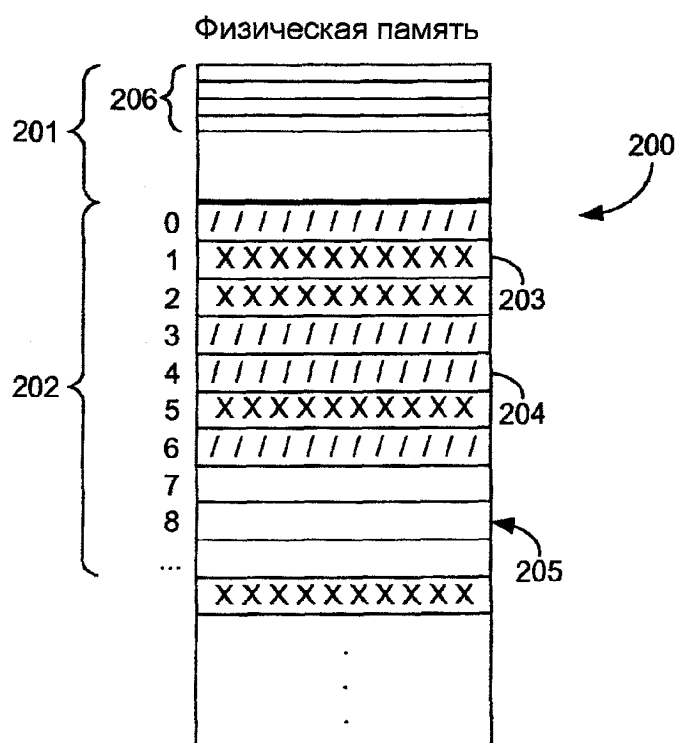
18. Устройство по п.17, в котором пороговое количество множества записей составляет общее число записей во всем множестве записей.

19. Устройство по п.17, в котором сегмент памяти, состоящий из множества смежных сегментов физической памяти, находится на предопределенной границе байта.

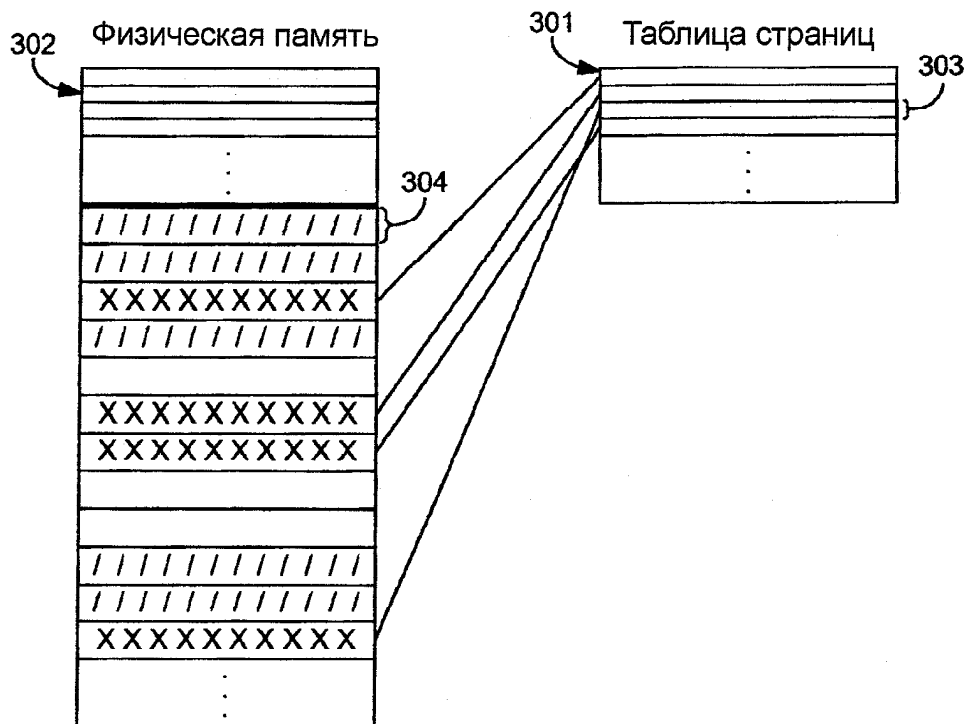
20. Устройство по п.17, в котором копирование каждого из множества сегментов физической памяти, ассоциированных со всеми из множества записей подходящей таблицы страниц, также включает в себя освобождение указанного множества сегментов физической памяти после того, как они были скопированы.



Фиг.1

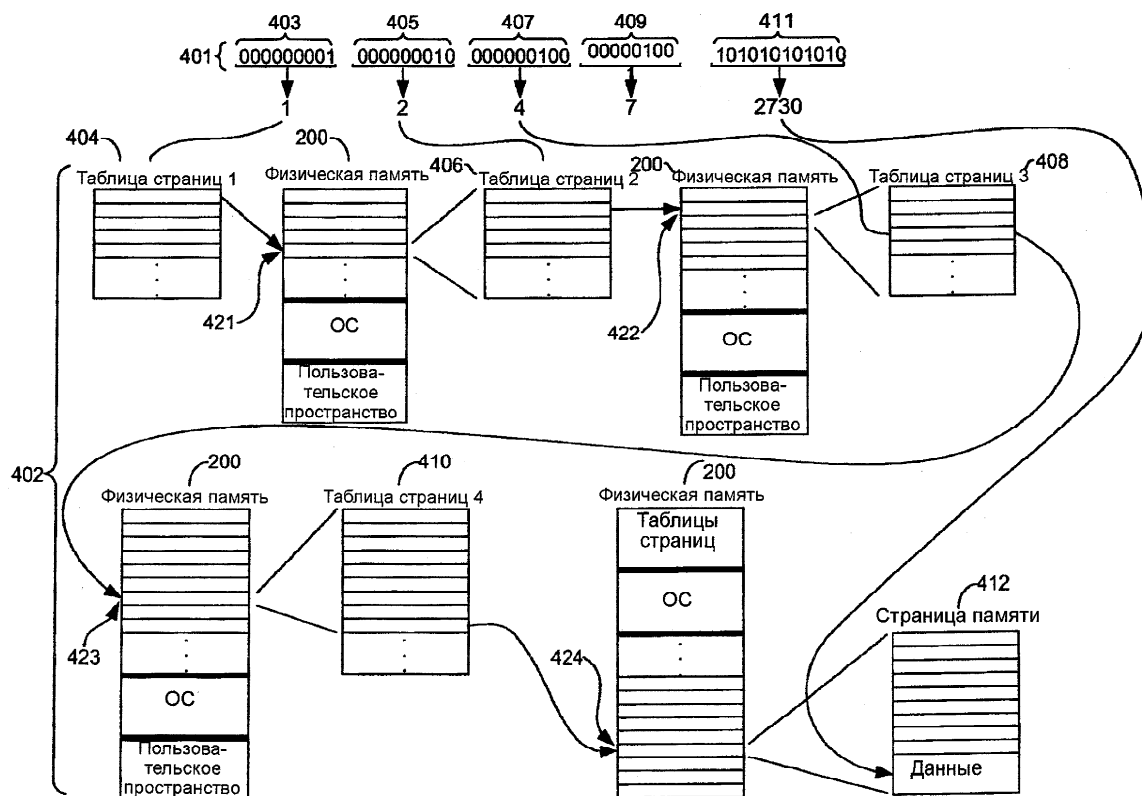


Фиг.2

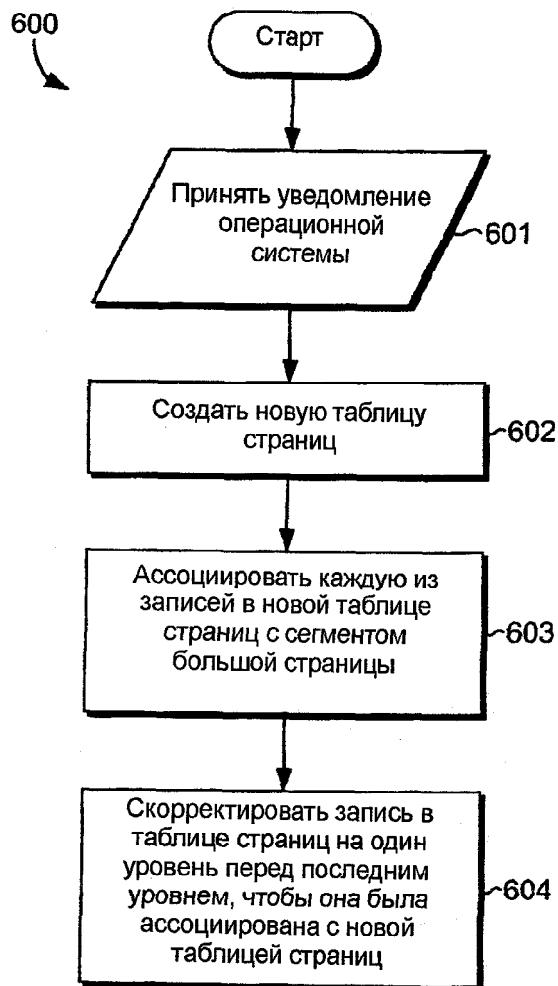


Фиг.3

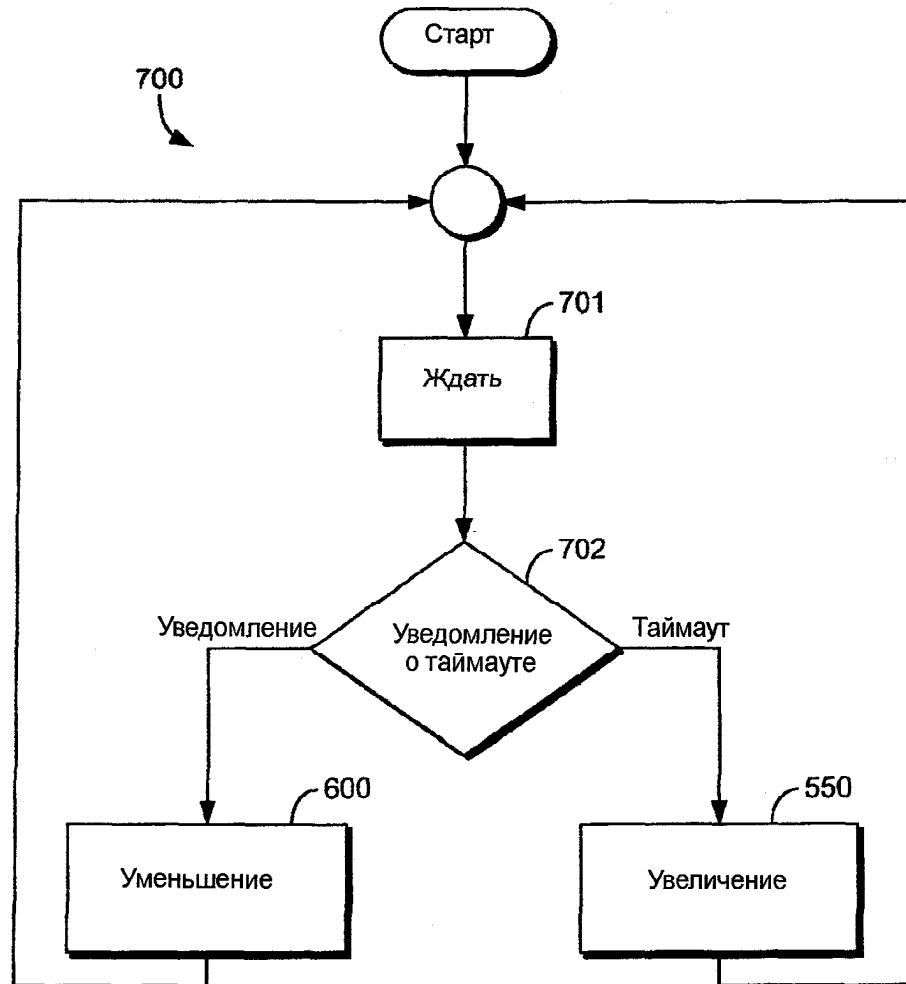
Иерархия таблиц страниц



Фиг.4



Фиг.6



Фиг.7