

**Новосибирский
государственный
университет**

**Российская Академия Наук
Сибирское отделение
Институт вычислительных технологий**



Проблемы интеграции ресурсов

Ю.И. Шокин, А.М. Федотов, О.Л. Жижимов

Информационные технологии

- **Проблема доступа к информации является одной из основных проблем, возникающих в современной человеческой деятельности.**
- **Любой производственный или научный процесс порождает огромные объемы данных, и работать с ними становится все сложнее по мере того, как гигабайты данных превращаются в терабайты. Количество данных когда-нибудь превысит способность компьютеров их обрабатывать, поэтому необходимы новые инструментальные средства и алгоритмы для анализа этих данных.**

Информационные технологии

*Кабы схемку иль чертеж,
Мы б затеяли вертеж,
Ну а так - ищи сколь хочешь,
Черта лысого найдешь!*

Л. А. Филатов. «Про Федота-
стрельца, удалого молодца»

Вместе с тем предъявляются серьезные требования к обеспечению прозрачного доступа и долговременной сохранности «информации». А в результате вопросы «что хранить?», «как хранить?» и «как найти?» остаются самыми существенными: без ответа на них все остальные теряют актуальность

Нынешнюю технологическую революцию характеризует не центральная роль знаний и технологий, а применение знаний и информации к генерированию знаний и созданием систем, обрабатывающих информацию и осуществляющих передачу «информации».

Если бы комплименты были правдой, это были бы не комплименты, а информация.

Кретья Патачкувна "Моя кибернетика», в книге «Мысли людей великих, средних и пса Фафика»

«Есть правила для выбора решения, но нет правил для выбора этих правил»

Энон

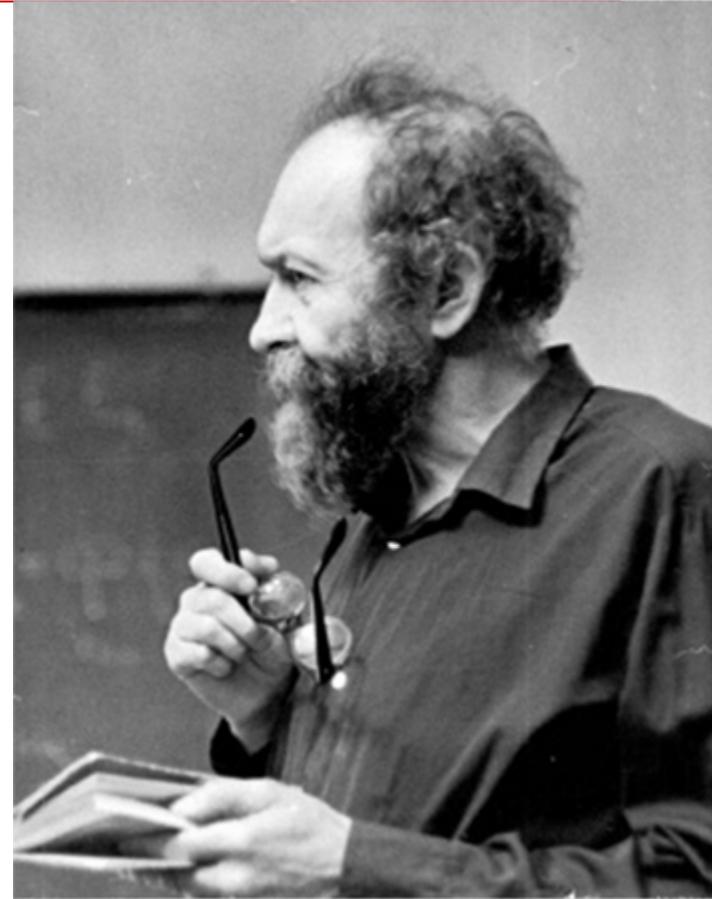
- **Развитие глобальных информационных и вычислительных сетей сегодня ведет к изменению фундаментальных парадигм работы с информационными ресурсами.**
- **Развитие распределенных ресурсов, создание инфраструктуры для их интеграции в единую информационную систему, обеспечивающую прозрачный доступ к распределенной информации – жизненная необходимость.**

Информационные модели

- Современные вычислительная техника и информационные технологии предоставляют исследователю мощный аппарат для «манипулирования данными», а не информацией. Данные, переведенные в электронную форму, приобретают новое качество, обеспечивая им более широкое распространение и эффективное использование.
- Однако применение информационных технологий должно основываться на использовании различных *моделей* (феноменологических, информационных, математических и др.).

Информационные модели

- Как неоднократно отмечал А.А.Ляпунов: *"нет модели - нет информации"*.
- Перефразируя А.А.Ляпунова следует отметить, что *"конечная цель всей работы, связанной с применением информационных технологий - является **понимание** того или иного явления, а не получение каких-либо чисел или картинок"*.



Информационные технологии

- *Скажите, пожалуйста, куда мне отсюда идти?*
- *Это во многом зависит от того, куда ты хочешь прийти, – ответил Кот.*
- *Да мне почти всё равно, – начала Алиса.*
- *Тогда всё равно, куда идти, – сказал Кот.*
- *Лишь бы попасть куда-нибудь, – пояснила Алиса.*
- *Не беспокойся, куда-нибудь ты обязательно попадёшь,*
- *сказал Кот,*
- *конечно, если не остановишься на полпути.*

Льюис Керролл, «Алиса в стране чудес»

-
- В современном ИТ-мире все большее значение, помимо традиционной производительности, стали приобретать так называемые нефункциональные требования (высокая доступность, масштабируемость, высокая степень безопасности систем).
 - Встречной тенденцией является перенос все большего и большего объема коммерческих и научных вычислений на открытые программно-аппаратные платформы. Фактически, мы являемся свидетелями своего рода революции. Такого рода технологический сдвиг в свою очередь определяет спрос на те навыки, которыми должны обладать ИТ-специалисты, чтобы быть востребованными на современном рынке труда.

Когда мы пытаемся вытащить что-нибудь одно, оказывается, что оно связано со всем остальным.

Закон Муира

-
- Неожиданно для многих аналитиков, на пересечении названных выше тенденций оказались технологии «облачных вычислений».
 - С одной стороны: «золотой стандарт» в области безопасности, надежности, масштабируемости и доступности.
 - С другой стороны: уникальные технологии виртуализации.
 - Расширение аутсортинга по предоставлению ИТ
 - Серверная платформа «майнфрейм»

Предназначение информационных систем



*Если вам все равно, где вы
находитесь, значит вы не
заблудились*

- **Организация хранения информации (организация хранилищ, поддержка систем хранения данных)**
- **Управление информацией (добавление, модернизация, изменение данных)**
- **Управление доступом к информации (контроль исполнения правил регламентации доступа к данным), идентификация данных**
- **Поиск информации**
- **Извлечение информации и предоставление ее пользователю в необходимом ему виде**
- **Визуализация информации в соответствии требованиями пользователя**

Предназначение информационных систем



Большинство людей не получают того, чего хотят, а всё потому, что сами не знают чего хотят.

Как исполнять желания, если их нет?

Дж.Б.Пристли. Тридцать первое июня.

-
- **Цель работы информационных систем – обеспечение конечного пользователя необходимой информацией.**
 - **Пользователя, как правило, не интересует, как устроена технологическая «кухня» информационной системы.**
 - **Более того, чем меньше эта «кухня» пользователю заметна, тем лучше построена та или иная информационная система.**
 - **Следует думать, что использование проверенных временем и практикой типовых решений позволит обеспечить эту технологическую «прозрачность».**

*Сложные проблемы всегда имеют простые,
легкие для понимания неправильные решения*

Закон Х.Л.Менкина

- **модели и стандарты представления информации и метайнформации;**
- **автоматическая классификация информации;**
- **доступ к распределенным и разнородным коллекциям (интероперабельность, масштабируемость, обнаружение релевантной информации, интеграция метайнформации);**
- **интерфейсы пользователей, визуализация и анализ данных,**
- **вопросы интеллектуальной собственности;**

-
- **анализ и обработка естественного языка, изображений, видео- и аудиоданных;**
 - **многоязыковый доступ к данным и обслуживание данных на нескольких языках;**
 - **мобильные технологии и агенты;**
 - **методы и средства поиска, каталогизация, индексирование, поддержка целостности и непротиворечивости коллекций, безопасность и защита информации.**

Машина должна работать, человек - думать.

Постоянно действующими факторами при формировании единого (виртуального) информационного пространства являются:

- **иерархичность информационных систем и ресурсов;**
- **разнородность ресурсов и программно-технических сред, объединяемых в едином сетевом операционном пространстве;**
- **распределенность элементов информационной инфраструктуры.**



Требования

Требовать и эффективности, и гибкости от одной и той же программы - все равно, что искать очаровательную и скромную жену... по-видимому, нам следует остановиться на чем-то одном из двух.

Фредерик Брукс-младший

Требование: использование технологий открытых систем

Создание распределенного гетерогенного информационного пространства, объединяющего разнородные ресурсы, предполагает удовлетворение следующих основных требований к информационным системам:

- способность систем функционировать в условиях информационной и реализационной неоднородности, распределенности и автономности информационных ресурсов ИС;**
- обеспечение интероперабельности, повторного использования неоднородных информационных ресурсов в разнообразных применениях;**

Требование: использование технологий открытых систем

- **возможность объединения систем в более сложные, интегрированные образования, основанные на интероперабельном взаимодействии компонентов;**
- **осуществлении миграции унаследованных систем в новые системы, соответствующие новым требованиям и технологиям при сохранении их интероперабельности;**
- **обеспечение более длительного жизненного цикла систем.**

Фундаментальные принципы модели ОС

- Использование подхода объектного моделирования при определении требований к системе
- Использование при определении требований различных, но взаимоувязанных точек зрения на систему
- Описание инфраструктуры системы таким способом, который обеспечивает для системных приложений прозрачность распределённой архитектуры (делает её незаметной для пользователя)
- Использование рамочных соглашений для поддержания системного соответствия

Под основными свойствами **открытых систем** понимаются:

- *переносимость* и *переиспользуемость* программного обеспечения, данных, моделей и **опыта**;
- *масштабируемость* как свойство сохранения работоспособности системы ИТ в условиях варьирования значений параметров, определяющих технические и ресурсные характеристики системы и/или поддерживающей среды.

Технологии ОТКРЫТЫХ СИСТЕМ

Если вам непонятно какое-то слово в техническом тексте, не обращайтесь на него внимания.

Текст полностью сохраняет смысл и без него.

Купер

Под основными свойствами **открытых систем** понимаются:

- *интероперабельность*, т.е. возможность взаимодействия компонентов распределенной системы посредством обмена информацией и ее совместного использования;
- Открытость систем достигается на основе стандартизации их структуры и поведения, наблюдаемого на границах систем или их интерфейсах.

Два определения интероперабельности

«Классическое»

Интероперабельность

- способность двух или более систем обмениваться информацией и правильно использовать её

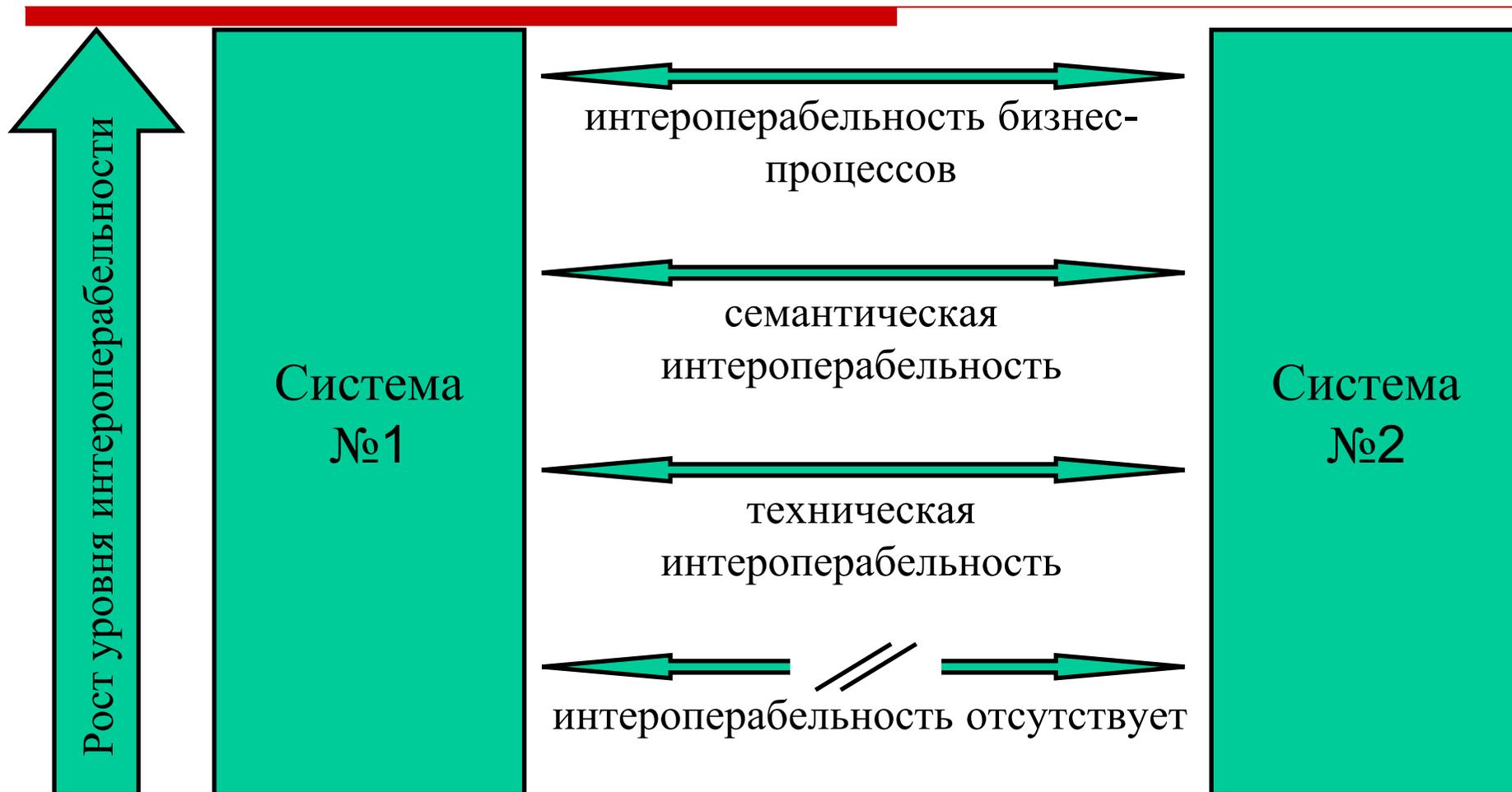
«Необходимое»

Интероперабельность

- способность различных систем и **организаций** к совместной работе

Кроме «технического» смысла появляется более широкий смысл, включающий социальные, политические и организационные факторы

Уровни итероперабельности



Открытая распределенная обработка

- ТОС представляет ряд стандартов, позволяющих реализовать преимущества услуг распределенной обработки информации в среде неоднородных ресурсов ИТ и нескольких организационных областях.
- Эти стандарты направлены на ограничение спецификаций систем и обеспечение для них инфраструктуры, которая снимает трудности, унаследованные от проектирования и программирования распределенных систем.

Открытая распределенная обработка

- **Распределенные системы важны вследствие растущей потребности взаимодействия систем обработки информации. Эта необходимость возникает в связи с такими организационными тенденциями, как, например, уменьшение размеров организаций, что требует информационного обмена как между группами внутри организации, так и между взаимодействующими организациями.**

Открытая распределенная обработка

- **Достижения технологии делают такой обмен возможным, отвечая на новые требования развитием сетей информационных услуг и персональных рабочих станций и допуская конструкции приложений, распределенных по большой конфигурации взаимосвязанных систем.**

Характеристики ОРО

Ключевые характеристики распределенной системы:

- **Удаленность** – компоненты распределенной системы могут быть широко распределены в пространстве; взаимодействие может быть локальным или удаленным;
- **Конкуренция** – любой компонент распределенной системы может работать параллельно с любыми другими компонентами;
- **Отсутствие глобального состояния** – глобальное состояние распределенной системы не может быть точно определено;
- **Частичные отказы** – любой компонент распределенной системы может отказать независимо от любых других компонентов;

Характеристики ОРО

Ключевые характеристики распределенной системы:

- *Асинхронность* – коммуникационная деятельность и обработка не управляются едиными глобальными часами. Нельзя считать, что взаимосвязанные изменения в распределенной системе происходят как единый акт;
- *Неоднородность* – нет гарантий того, что компоненты распределенной системы построены по единой технологии, а набор различных технологий определенно будет меняться со временем. Неоднородность проявляется в разных местах: в технических средствах, операционных системах, коммуникационных сетях и протоколах, языках программирования, приложениях и т.п.;

Характеристики ОРО

Ключевые характеристики распределенной системы:

- **Автономность** – распределенная система может быть поделена между несколькими автономными уполномоченными по административному управлению и контролю, без единого центра контроля. Степень автономности определяет, насколько ресурсы обработки и соответствующие устройства (принтеры, запоминающие устройства, графические дисплеи, аудиоустройства и т.п.) находятся под контролем отдельных организационных единиц;
- **Эволюция** – в течение своего жизненного цикла распределенная система столкнется со многими изменениями, которые мотивируются техническим прогрессом, обеспечивающим лучшую производительность за лучшую цену, стратегическими решениями о новых целях, новыми типами приложений;
- **Мобильность** – источники информации, узлы обработки и пользователи могут перемещаться физически.

*Послушайте, ребята,
Что вам расскажет дед.
Земля наша богата,
Порядка в ней лишь нет.*



А. К. Толстой

- Для выполнения приведенных требований необходимо создание развитой инфраструктуры для представления и обмена **метаданными (данными о ресурсах)**, без которой невозможно формирование единого информационного пространства.
- Как правило, в основе интеграции метаданных лежит выработка стандарта на формат для представления метаданных, одновременно с унификацией нормативно-справочной информации (профиля информационных систем).

ПРОБЛЕМЫ



Если за две минуты омлет еще не готов, у заказчика два выбора - подождать или съесть его сырым.

Фредерик Брукс-младший

- **Отметим, что большинство «владельцев» информационных и вычислительных ресурсов, к числу которых относятся и учреждения СО РАН, формируют информационные ресурсы, исходя из принципа их приватности и ориентированности на внутреннее использование.**
- **В результате большие объемы формально публичной информации труднодоступны или недоступны внешним потребителям, что, в частности, сдерживает мультидисциплинарные исследования и может приводить к курьезным ситуациям. Нередко легче получить научный результат заново, чем узнать о его наличии и получить к нему доступ.**

-
- **Схожие проблемы существуют и в отношении вычислительных ресурсов, которые, вследствие отсутствия надлежащих сервисов, являются труднодоступными для большинства потенциальных пользователей и поэтому часто остаются недогруженными, в то время как потенциальным потребителям приходится довольствоваться своими локальными ресурсами, плохо приспособленными для проведения вычислений.**

Я не знаю причины, по которой нам не следует этого делать, но, возможно, позже мы придумаем какую-нибудь.

Марк Дэвисон

- **Таким образом, помимо проблем интеграции первичных информационно-вычислительных ресурсов в единое информационно-вычислительное пространство, стоит проблема разработки основных принципов организации таких ресурсов с целью их эффективного использования.**

*...Из окна чудесный вид на будущий стадион,
вот только бульдозеры окончат работу
и будет осушено болото.*

Дж.Б.Пристли

- **обеспечивают механизмы интеграции информационных ресурсов из разных источников сведениями о свойствах этих ресурсов;**
- **являются источниками сведений о свойствах и содержании информационных ресурсов для механизмов управления данными в информационных системах;**
- **являются источником информации, необходимой для осуществления реинжиниринга информационных систем;**
- **обеспечивают представление сведений о системе, ее информационных и других ресурсах для различных приложений и пользователей системы.**

Как только проект окончательно принят, он становится устаревшим в смысле своих концепций. Настойчивость руководителя может определить график выполнения задания, но не в состоянии определить срок его действительного завершения.

Фредерик Брукс-младший

Центры обработки данных

Когда кажется, что все уже работает, все объединено в систему - вам еще осталось работы на четыре месяца.

Чарльз Портман

Как уже отмечалось, наиболее экономически адекватным и востребованным решением является создание Центров обработки данных (ЦОД), позволяющих аккумулировать мощные вычислительные ресурсы и системы хранения информации, а также резко сократить затраты на обслуживающий персонал и сервисные услуги.

Центры обработки данных

- ЦОД могут предоставлять информационно-вычислительные ресурсы как непосредственно через развитую систему сервисов, так и в среде «облачных вычислений».

Когда на руках выигрышные карты, следует играть честно.

Оскар Уайльд

Центры обработки данных

- **Суть облачных вычислений состоит в том, чтобы поместить имеющиеся ресурсы в виртуальное «вычислительное облако», так что доступ к ним можно было бы осуществлять из любого места по мере необходимости.**
 - **Например, к услугам «облака» может обратиться удаленный сервер института в пиковый период или рабочая станция в лаборатории, на которой запустили научное приложение, требующее серьезных вычислительных ресурсов.**
-

Центры обработки данных

- **Инфраструктура вычислительного облака обеспечивает гибкое маневрирование ресурсами и их оптимальной загрузки и детального учета объема потребляемых услуг, предоставляемых ЦОД.**

В поиски повышения надежности будут вкладываться средства до тех пор, пока они не превысят величину убытков от неизбежных ошибок или пока кто-нибудь не потребует, что- бы была сделана хоть какая-то полезная работа.

Джилб

Типовой информационный центр это:

- **центр аккумуляции информационных ресурсов, обеспечивающий их централизованное хранение и оперативную обработку**
- **центр регистрации распределенных информационных ресурсов**
- **единая точка доступа к распределенным информационным ресурсам**
- **центр администрирования информационных ресурсов**
- **центр сбора и обработки статистики использования информационных ресурсов**
- **центр тестирования программного обеспечения для функционирования комплекса**

Основные функции организации информационных систем и базовые технологии интеграции ресурсов



<i>Local</i>	<i>организация хранения информации - организация хранилищ, поддержка систем хранения данных управление информацией - добавление, модернизация, изменение данных</i>
<i>LDAP</i>	<i>управление доступом к информации – аутентификация пользователей, контроль исполнения правил регламентации доступа к данным и идентификация данных</i>
<i>Z39.50</i>	<i>поиск информации извлечение информации и предоставление ее пользователю в необходимом ему виде</i>
<i>WWW</i>	<i>визуализация информации в соответствии требованиями пользователя</i>

Единая политика доступа к ресурсам

- **предоставление унифицированных пользовательских интерфейсов для доступа к однотипным ресурсам;**
- **возможность организации сквозного поиска информации в распределенных однотипных ресурсах;**
- **поддержка единых правил ограничения доступа для всех ресурсов, контроля пользователей и однократной аутентификации;**
- **предоставление пользователям информации о номенклатуре актуальных информационных ресурсов и информации о каждом из них;**
- **сбор статистической информации о работе всех подсистем распределенной системы.**

Как бы кропотливо и тщательно вы ни готовили выборку, вам всегда могут сказать, что она неправильна и неприменима к данной проблеме.

Эдингтон

- На сегодняшний день технология LDAP – одна из немногих промышленных технологий, используемых для централизованно-распределенного хранения информации и управления ей в глобальном масштабе
- Технологии LDAP основаны на построении и управлении распределенной информацией, упорядоченной в иерархическую древовидную структуру, и определяют правила сетевого доступа к этой информации
- Технологии LDAP позволяют достаточно просто создать распределенную иерархическую информационную систему с разграничением зон ответственности по различным параметрам корпоративного каталога

