КОНТРОЛЬНАЯ РАБОТА

по курсу

**«Корпоративные информационные системы»**

**Лингвистическое и правовое обеспечение КИС.**

**Лингвистическое обеспечение КИС**

**Лингвистическое обеспечение (ЛО)** представляет собой совокупность научно–технических терминов и других языковых средств, используемых в информационных системах, а также правил формализации естественного языка, включающих в себя методы сжатия и раскрытия текстовой информации для повышения эффективности автоматизированной обработки информации.

Основным средством описания информационной базы и информационной потребности служат **информационно–поисковые языки**, относящиеся к классу искусственных языков. Помимо таких строго формализованных с точки зрения семантики и синтаксиса средств, в качестве дополнительных широко применяются **терминологические структуры**различного назначения, имеющие как линейную, так и нелинейную (иерархическую, сетевую) организацию. Состав лингвистического обеспечения (ЛО) информационных систем может быть представлен следующей схемой (рис. 1):



Рис. 1. Состав лингвистического обеспечения информационных систем

В предложенной схеме информационно–поисковые языки служат как средства выражения смыслового содержания документов и информационной потребности пользователя, языки манипулирования как попытки обобщения средств управления данными, терминологические структуры как моделей данных, с одной стороны, и понятийных систем, выражаемых средствами искусственного языка с естественной лексикой – с другой. Далее приведем характеристики выделенных компонентов ЛО ИС.

Для изучения принципов и методов построения и использования компонентов ЛО предварительно определим базовые понятия.

**Язык**– это знаковая система любой физической природы, выполняющая познавательную и коммуникативную функцию в процессе человеческой деятельности.

**Искусственный язык**(в отличие от естественного, представляющего собой средство общения и выражения мысли и неоднозначного по своей природе) – это специализированный язык, основное назначение которого состоит в устранении многозначности слов естественного языка и всего того, что характеризует эмоции и отношение к различным предметам. В искусственном языке должны выражаться лишь объективные характеристики предметов, их связей и соотношений.

Многозначность слов естественного языка и служащая в некотором роде показателем развитости языка, становится препятствием в случае использования в системе хранения и обработки информации. В связи с этим в ИС применяются искусственные языки, специально сконструированные для формулировки основного смыслового содержания информационной базы и информационной потребности с целью последующего их сопоставления. К таким языкам в первую очередь относятся **информационно–поисковые языки (ИПЯ)**,обеспечивающие компактную, строго алгоритмизированную запись содержания документов и запросов в информационно–поисковой системе. ИПЯ можно определить как специализированную семантическую систему, состоящую из алфавита, правил образования (грамматики) и правил интерпретации (семантики).

**Алфавит**– это любая конечная совокупность знаков (букв, цифр и т.п.), используемых в ИПЯ.

Выделяют **морфологические и синтаксические правила** образования (построения) **терминов**– слов языка*.* Морфологические правила определяют процедуру построения терминов ИПЯ из его морфем, а синтаксические – процедуру построения предложений (фраз) из этих терминов. Синтаксические правила – обязательный элемент любого ИПЯ. В некоторых ИПЯ для соединения терминов в предложения (фразы) применяются специальные лексические средства.

Последний элемент ИПЯ, если его рассматривать как специализированный абстрактный язык это **правила интерпретации**, т.е. правила перевода терминов и предложений (фраз) ИПЯ на соответствующий естественный язык. Эти правила задаются, например, в в виде двуязычных словарей, в которых каждому термину (лексической единице) ставится в соответствие определенное слово или выражение естественного языка, и наоборот. В такой словарь включаются также все символы, применяемые в данном ИПЯ для соединения терминов в предложения (фразы). Кроме того, правила интерпретации для ИПЯ, как и правил построения, формулируются на естественном языке в специальных инструкциях, методиках и т. д.

Рассмотрим типологию ИПЯ по способности к выражению смыслового содержания документов, как структурных единиц информационной базы. Опираясь на лексику, грамматику и синтаксис, выделим два основных типа ИПЯ:

– языки классификационного типа (классификация);

– языки дескрипторного типа (индексирование).

**Классификация**, как средство описания содержания документа, представляет собой процесс соотнесения содержания документов с понятиями, зафиксированными в заранее составленных систематических схемах. Основная цель классификации – приписать каждый документ классу, или, иначе – приписать каждому документу имя класса, формируя тем самым множества сообщений для обработки и поиска.

Языки дескрипторного типа поддерживают процесс **индексирования**, который заключается в формировании описания документа как совокупности дескрипторов (синонимов, обозначающих одно понятие), выбираемых из заранее созданных словарей понятий, либо из текстов документов.

Следующий компонент ЛО ИС: **языки описания данных**, предназначены для описания данных на разных уровнях абстракции: внешнем, логическом и внутреннем. Считается, что языки описания данных на логическом (концептуальном) и внутреннем уровнях независимые и разные. Однако в большинстве промышленных СУБД языки не делится на два отдельных языка описания логической и физической организации данных, а существует единый язык, которая еще называется языком описания схем. В известных и широко используемых на практике СУБД семьи dBASE применяется единый язык описания данных. Он предназначен для представления данных на логическом и физическом уровнях. Этот язык имеет свой синтаксис: например, имя файла не должно превышать восьми символов, а имя поля - десяти; при этом каждое имя может начинаться с буквы, поля календарной даты обозначаются символом D (DATA), символьные поля — С (CHARACTER), числовые — N (NUMERIC), логические — L (LOGICAL), примечаний — М (MEMO).

Описание всех имен, типов и размеров полей сохраняется в памяти вместе с данными; эти структуры в случае необходимости можно просмотреть и исправить. Если логический и физический уровни отделены, то в состав СУБД может входить язык описания сохранения данных. В некоторых СУБД используется еще язык описания подсхем, который нужен для описания части БД, которая отражает информационные потребности отдельного пользователя или прикладной программы. В составе СУБД типа dBASE такой язык не используется.

Язык описания данных на внешнем уровне используется для описания требований пользователей и прикладных программ и создания инфологической модели БД. Этот язык не имеет ничего общего с языками программирования. Так, языковым средством, которое используются для моделирования, является обычный естественный язык или его подмножество, а также язык графов и матриц.

**Языки манипулирования данными** используются для обработки данных, их преобразований и написания программ. DML может быть базовым или автономным.

Базовый язык DML — это один из традиционных языков программирования (BASIC, C, FORTRAN и др.). Системы, которые используют базовый язык, называют **открытыми**. Использование базовых языков как языков описания данных сужает круг лиц, которые могут непосредственно обращаться к БД, поскольку для этого нужно знать язык программирования. В таких случаях для упрощения общения конечных пользователей с БД предполагается язык ведения диалога, который значительно проще для овладения, чем язык программирования.

Автономный язык DML — это собственный язык СУБД, который дает возможность выполнять различные операции с данными. Системы с собственным языком называют **закрытыми**.

Теперь рассмотрим еще один компонен ЛО ИС: **терминологические структуры**. В большинстве информационных систем помимо ИПЯ на этапах индексирования и поиска документов применяются различные средства, имеющие лингвистическую природу, например, тематические рубрикаторы, тезаурусы, словари как информативных, так и неинформативных лексических единиц, словари синонимов, словари словосочетаний и т.п. Организационная типология терминологических структур, приведенная на рис. 2, тесно связана с типологией по семантическому признаку. С точки зрения семантики словоупотребления терминологические структуры могут быть разделены на **семантически упорядоченные** и **семантически неупорядоченные**. При этом семантически неупорядоченные терминологические структуры всегда имеют **линейную организацию**, а семантически упорядоченные – **иерархическую или сетевую организацию**.



Рис. 2. Типология терминологических структур по семантическому признаку

Семантически упорядоченные терминологические структуры отражают оба типа связей, которые могут существовать между отдельными терминами – парадигматические и синтагматические. **Парадигматические связи** характеризуют различные виды отношений – отношения иерархии, ассоциативные отношения и отношения эквивалентности. **Синтагматические связи** показывают логические отношения между понятиями (рис. 2).

Основными представителями сетевых терминологических структур являются **тезаурусы***.* Все понятия естественного языка, служащие для описания окружающего мира, входят во всеобщий тезаурус мира, отражающий весь универсум знаний. Такой тезаурус представляет собой список понятий, выраженных на естественном языке, с обозначением отношения между ними.

Всеобщий тезаурус можно подразделить на частные тезаурусы путем выделения совокупности однородных понятий по их иерархическому уровню или путем выделения понятий, которыми можно описать какую–либо специфическую часть мира. Таким образом, на основе всеобщего тезауруса можно составить бесконечное множество тезаурусов по различным областям знаний, по отдельным проблемам и задачам.

Тезаурус может быть представлен как семантическая сеть, в которой понятия связаны регулярными и устойчивыми семантическими отношениями – иерархическими (например, род–вид, целое–часть), ассоциативными, а также отношениями эквивалентности. При этом отдельное понятие определенной области знаний в тезаурусе представлено словом или словосочетанием, соотносящимся с другими словами и словосочетаниями и образующим вместе с ними замкнутую систему. Иерархические отношения в тезаурусе представляют собой классификацию, основанную на словах естественного языка, а не на абстрактных категориях, поэтому нарушается правильная структура дерева – один и тот же термин может иметь несколько «родителей» – вышестоящих терминов на предыдущем уровне. Например, в Тезаурусе по информатике словосочетание *Автоматизированная обработка информации* имеет два вышестоящих родителя: *Автоматизированная обработка* и *Обработка информации,* а слово *Буквы* – целых три родителя: *алфавиты, символы, буквенно–цифровая информация*. Тезаурус, отображая возможные семантические связи терминов, представленных в базах данных, является идеальным лексическим инструментом информационно–поисковых систем, с помощью которого можно найти необходимую лексику для составления запросов или их модификации с целью достижения наилучших показателей эффективности поиска.

**Иерархические классификационные структуры.**К таким структурам относятся различные **рубрикаторы** и **классификаторы**, фиксирующие подчинение терминов в определенной предметной области.

На рисунке 3 приведен фрагмент Рубрикатора ВИНИТИ для заглавной рубрики «201 Информатика». Рубрикатор ВИНИТИ является локальным (отраслевым) по отношению к Государственному рубрикатору и отличается большей детализацией рубрик.



Рис. 3. Фрагмент Рубрикатора ВИНИТИ

**Словарь синонимов.** Словарь синонимов, который для каждого входа словаря определяет одну или больше синонимичных категорий, также с точки зрения своей структуры может быть отнесен к иерархической организации терминов. Такие словари широко используются при индексировании, а также позволяют искать не только по запрошенному слову, но и по его синонимам. Ниже приведен фрагмент словаря синонимов для области «Информатика»:

…

ЭФФЕКТИВНОСТЬ ПОИСКА

информационная эффективность

техническая эффективность

эффективность информационного

ЮРИДИЧЕСКАЯ ДЕЯТЕЛЬНОСТЬ

юридическая практика

ЮРИСПРУДЕНЦИЯ

право

правоведение

юридические аспекты

К **линейным терминологическим структурам** относятся линейные словари различного назначения, обычно упорядоченные по лексикографическому принципу. С точки зрения своего участия в процессах индексирования документов и запросов такие словари делятся на **положительные** и **отрицательные**. Положительные словари объединяют лексику, которую можно использовать в процессе индексирования. Отрицательные словари содержат лексику, запрещенную для использования при индексировании.

**Морфологические словари** содержат основы слов, аффиксы, суффиксы и окончания. Такие словари могут быть использованы, с одной стороны, для нормализации поисковых образов документов, а с другой – для нормализации лексики поисковых запросов. Грамматический строй естественных языков нередко расходится со структурой логической мышления, и поэтому при поиске информации необходимо полностью или частично исключить влияние аффиксов и окончаний слов естественных языков. Для этого можно предусмотреть наращивание документов всеми потенциально возможными словоформами, которые можно составлять, например, на базе основ слов, первоначально содержащихся в документах. Наличие в паре «документ – запрос» словоформ, совпадающих с точностью до общности их корней, в результате такого наращивания может привести к появлению в документе словоформы, полностью совпадающей со словоформой, имеющейся в запросе. Такое наращивание снимало бы различие употреблений словоформ в документах и запросах. Другой технологический вариант, позволяющий снимать различие употреблений словоформ, состоит в использовании **кодирования слов**. Сущность метода автоматического кодирования слов с помощью наперед заданных словарей аффиксов и окончаний заключается в автоматической проверке на наличие в словах естественных языков элементов, вошедших в наперед заданные (составленные экспертами– лингвистами) словари аффиксов и окончаний, и отсечении их, если они имеются. От качества составления словарей аффиксов и окончаний в значительной мере зависит качество автоматического кодирования слов естественных языков, а, следовательно, и функциональная эффективность ИПС в целом. Ошибки могут быть следствием такого алгоритма, когда после включения очередной морфемы в словарь, она отсекается из всех слов естественно–языкового употребления в базе данных, не зависимо от того, является ли для конкретно рассматриваемого словаря морфемой или частью корня.

**Словарь словосочетаний**. Такой словарь используется для определения наиболее часто встречающихся устойчивых комбинаций слов. Словарь словосочетаний повышает эффективность анализа содержания, выделяя для идентификации содержания однозначные словосочетания вместо множества в общем случае однозначных (например, пара отдельных терминов «программа» и «язык» является менее определенной, чем словосочетание «язык программирования»).

Лингвистической особенностью словаря является то, что термины – одиночные слова зачастую не выражают никакого смысла, являясь только составной частью словосочетания.

Основываясь на том, что наиболее информативными терминами являются термины–словосочетания, наиболее правомерно использовать именно их для составления поискового запроса.

**Частотный словарь.**Частотный словарь – перечень дескрипторов и ключевых слов. Термины располагаются в алфавитном порядке, либо в порядке убывания (возрастания) частоты использования их в информационном массиве. Частотная характеристика термина показывает количество документов информационного массива, в которых термин встретился хотя бы один раз. Частота встречаемости ориентирует пользователя в лексике информационного массива с точки зрения включения какого–либо термина в поисковый запрос. Рассмотрим, например, фрагмент частотного словаря ретроспективной реферативной базы данных «Инфрматик» (1986–2002 г):

51 ИНФОРМАЦИОННАЯ ГРАМОТНОСТЬ

1 ИНФОРМАЦИОННАЯ ГРАНИЦА ВСЕЛЕННОЙ

1 ИНФОРМАЦИОННАЯ ДЕМОКРАТИЯ

**Стоп–словарь (словарь отрицаний)**.Словарь отрицаний («стоп–слов») содержит термины, которые признаны не информативными для данной предметной области. Использование их запрещается для индексирования содержания документов. Например, термины «исследование», «вопросы», «требования», «проблемы» и др. являются политематическими и удаляются из поисковых образов документов и запросов. Словарь стоп–слов может использоваться как при построении частотных словарей, так и при разборе выражения информационной потребности на ИПЯ. Запрещенные термины не заносятся в словарь. Таким образом, неинформативные термины автоматически исключаются из поискового процесса.

**Правовое обеспечение КИС**

**Правовое обеспечение**– совокупность правовых норм, определяющих создание, юридический статус и функционирование информационных систем,  регламентирующих порядок получения, преобразования и использования информации***.***

Различные информационные системы, а особенно Интернет как одно из самых масштабных проявлений развития информационных технологий, предоставили огромные возможности доступа к информации. В то же время информационные системы выступили катализатором многих негативных явлений (пиратство, кража информации, денежных средств и т.д.). Поэтому правовое регулирование данных систем приобретает все большую остроту. Здесь важны роль и место не только международного сообщества, но государственных органов в регулировании деятельности по созданию систем телекоммуникаций и обеспечения защиты информации. В настоящее время эти функции возложены на Министерство связи и информатизации Республики Беларусь.

Одной из наиболее важных проблем, связанных с использованием информационных систем, является нарушение авторских прав на произведения, к которым можно получить доступ. Легкость копирования, почти невозможный контроль со стороны авторов создают условия для массовых нарушений авторского права. Закон «Об авторском праве и смежных правах» создает в Беларуси правовые условия для охраны прав авторов и правообладателей.

Исходные положения правового обеспечения процессов информатизации в Беларуси определены концепцией государственной политики в области информатизации, одобренной Указом Президента Республики Беларусь от 6 апреля 1999 г. № 195, а также содержатся в Законе «Об информации, информатизации и защите информации». Основной задачей, реализуемой в рамках данного направления, является создание правовой базы для решения проблем, связанных с:

      реализацией конституционного права граждан на информацию - обеспечением им равных прав на получение информации из всех общедоступных информационных систем;

      производством и распространением информационных ресурсов и управлением ими;

      формированием рынка информационной продукции и информационных услуг;

      созданием и использованием документов в электронной цифровой форме;

      использованием Интернет-технологий для информационного взаимодействия и обмена информацией;

      определением правового статуса информационных ресурсов, в том числе и ресурсов Интернет;

      формированием информационно-телекоммуникационной инфраструктуры в республике и управлением ею;

      обеспечением информационной безопасности государства, юридических и физических лиц, защитой персональных данных и созданием систем противодействия компьютерным преступлениям;

      управлением, координацией и контролем деятельности в области внедрения компьютерных систем и средств телекоммуникаций;

      организацией внутреннего и внешнего информационного взаимодействия, межгосударственным обменом информацией;

      подготовкой, повышением квалификации кадров и организацией подготовки специалистов в системе высших учебных заведений страны;

      развитием и законодательным закреплением понятийного аппарата в сфере информационно-коммуникационных технологий;

      развитием и совершенствованием правовой информатизации как правового, организационного, социально-экономического и научно-технического процесса обеспечения потребностей государственных органов, юридических и физических лиц в правовой информации на базе информационных систем и сетей.

Так как правовое регулирование сферы информационных систем является достаточно новым явлением для Беларуси важно строить его в соответствии с наилучшими традициями и опытом западных стран.

Значительное место в международных документах уделяется правовым проблемам, связанным с деятельностью коммерческих онлайновых компаний и Интернета. Эти проблемы можно разделить на следующие: регулирование содержания (вредное и незаконное), соблюдение авторских и смежных прав в условиях технически легкого копирования любой информации представленной в цифровом виде, вопросы формирования киберэкономики (электронные деньги, реклама, маркетинг, электронные публикации, электронные контракты, налог на передачу информации), информационная безопасность, понимаемая в широком смысле, как безопасность жизненно важных для общества систем управления транспортом, войсками, хозяйством крупных городов и т. п.

Компьютерные сети типа Интернета или сетей, обеспечивающих онлайновые услуги, теоретически могут использоваться в преступных целях. Широкий набор правонарушений в данной области можно разделить на два больших класса: преступления, направленные на сети и системы обработки информации, и преступления, в которых сети используются как каналы связи.

В первую категорию попадают «компьютерные» преступления, связанные с несанкционированным доступом, изменением или разрушением данных, пользованием услугами. Они уже довольно хорошо известны; имеется разработанная юридическая база для борьбы с ними, изложенная в уголовном законодательстве.

Ко второй относятся преступления, связанные главным образом с «выражением мнения» – показом насилия, расовой дискриминации, порнографией. Поскольку Интернет – всемирное средство для передачи текстов, изображений и звуков, он идеально подходит для совершения таких правонарушений. Большая часть информации в Интернете используется в развлекательных и деловых целях, причем совершенно легально. Однако, как и другие технологии коммуникаций, особенно на первоначальных стадиях развития, Интернет несет много потенциально вредного или незаконного содержания, может быть использован как средство осуществления незаконной деятельности. Эти нарушения и злоупотребления, связанные с Интернетом, имеют разнообразный характер и связаны с защитой:

• национальной безопасности (инструкции по изготовлению взрывчатых устройств, производству наркотиков, террористической деятельности);

• несовершеннолетних (оскорбительные формы маркетинга, насилие и порнография);

• человеческого достоинства (расовая дискриминация и расистские оскорбления);

• информации (злонамеренное хакерство);

• тайн личной жизни (несанкционированный доступ к персональным данным, электронные оскорбления);

• репутации («навешивание ярлыков», незаконная сравнительная реклама);

• интеллектуальной собственности (несанкционированное распространение защищенных авторским правом работ, например, программного обеспечения, музыки и т. п.).

В современных национальных системах правового регулирования обработки и использования персональных данных существуют четыре компонента защиты персональных данных:

– генеральный (базовый) подход к защите персональных данных;

– секторный (отраслевой) подход;

– орган или система органов защиты данных;

– корпоративные средства защиты.

**Генеральный** заключался в стремлении к созданию единого и всеобъемлющего закона о защите сферы частной жизни и был связан с попытками теоретически обосновать некое «всеобщее и абсолютное право на невмешательство в частную жизнь».

**Секторный** (отраслевой) состоит в создании специализированных законов либо для каждого типа посягательств на сферу частной жизни, либо для каждой отрасли или сектора человеческой деятельности, служащей потенциальным источником угроз для права человека на невмешательство в его частную жизнь (например, для почты и средств связи, бюро кредитной информации, средств массовой информации и рекламной сферы, частных детективов, для компьютерных банков данных).

В подавляющем большинстве стран современные национальные системы правового регулирования обработки и использования персональных данных применяют так называемый **смешанный принцип**, объединяющий некоторые аспекты генерального и отраслевого подходов. Национальное законодательство в сфере защиты данных, как правило, состоит из базового или, как их еще называют, системообразующего, комплекса секторных законов, обеспечивающих защиту персональных данных в секторах (отраслях) человеческой деятельности, несущих потенциальную угрозу для права субъекта персональных данных на невмешательство в его частную жизнь.

Активным регулирующим компонентом современных систем защиты персональных данных является **национальный орган (или система органов)** защиты данных, наделенный регистрационно–разрешительньми, контрольно–надзорными, а также арбит­ражными, экспертными и методологическими функциями.

И последний компонент таких систем – **корпоративные средства защиты**, которые часто называют средствами саморегулирования.

Существуют различные пути принятия компаниями или отраслями мер саморегулирования в области защиты персональных данных:

– введение внутренних руководящих указаний или принципов;

– принятие кодексов практики или поведения:

– учреждение должности специального ответственного.

Эти меры защиты персональных данных могут составлять часть целой серии внутренних процедур, которая, в свою очередь, представ­ляет собой часть внутреннего менеджмента или руководящих принци­пов безопасности некой корпорации или отрасли человеческой деятельности. Из всех предпринимаемых мер все более популярными становятся кодексы поведения или практики. Помимо всего прочего, они могут предоставлять полезные средства для обеспечения прозрачности отраслевой или корпоративной политики. На практике разница между такими кодексами и внутренними руководящими принципами (или иными отраслевыми или корпора­тивными подзаконными актами) может быть невелика.

В области защиты данных форма, содержание и основная направ­ленность усилий кодексов практики или поведения не являются сколько–нибудь единообразными, что сильно затрудняет их формальную оценку и сравнение. Не существует никакого универсального опреде­ления кодексов практики или поведения. Например, Голландский законодательный акт 1988 г. о защите данных определяет кодексы поведения как «некий набор правил или рекомендаций по отноше­нию к файлам персональных данных, принятый одной или более организациями, являющимися репрезентативными для того сектора (отрасли), который они охватывают в интересах защиты неприкосновенности частной жизни».

Одним из ключевых элементом любого кодекса практики должен быть его добровольный характер, поскольку любой такой кодекс является средством саморегулирования, используемым некой отраслью, компанией, корпорацией или профессиональной ассоциацией для достижения определенного уровня защиты данных. Одна из значимых сторон любого кодекса состоит в том, что члены отрасли, корпорации или ассоциации могут самостоятельно определять, какие принципы защиты данных им стоит переводить в работоспособные положения своего кодекса практики или поведения. Если кодекс просто провозглашает широкие принципы защиты данных, но не предлагает мер для соблюдения этих принципов, то он не является средством защиты данных.

И, наконец, следует отметить, что кодексы поведения или практики являются обычно инструментами частного сектора. Этому способствуют несколько причин. Во–первых, регулирование защиты данных публичного сектора обычно осуществляется на основании правил, уста­новленных внутренними инструкциями. Во–вторых, в большинстве стран защита данных частного сектора остается сравнительно нерегулируемой, что предоставляет отраслям и корпо­рациям возможности для саморегулирования, что приемле­мо и для международных отраслевых ассоциаций, таких, как Международная ассоциация воздушного транспорта (IАТА), находя­щаяся вне сферы правового регулирования какой–либо страны.

Однако существуют и кодексы публичного сектора, например, Кодекс ассоциации начальников полиции в Великобри­тании или Кодекс католических начальных школ в Голландии.

Итак, есть две категории кодексов: кодексы, которые не имеют никакой законодательной поддержки, и кодексы, которые пользуются той или иной формой законодательной поддержки. Большинство существующих кодексов попадает в первую категорию, и лишь немногие (по крайней мере, на данной стадии развития) – во вторую. Большинство «незаконодательных» кодексов пришло из стран общего права, в которых (за исключением Великобритании) отсутствует исчерпывающее общее законодательство о защите данных в частном секторе.

Кодексы поведения могут быть весьма гибкими инструментами для внедрения закона в отрасли и секторы экономики. Кодексы поведения предоставляют определенным отраслям возможность продемонстрировать реальную заботу о вопросах защиты права на невмешательство в частную жизнь. Однако регулирование при помощи кодексов может быть ограничено условиями конкуренции и другими аспектами некоего конкретного сектора или отрасли. Кодексы поведения могут усложнить или запутать правовые рамки, субъекты данных не всегда осведомлены о статусе конкретного кодекса поведения. Как следует из изложенного, наметилась важная тенденция к легитимизации корпоративных и профессиональных кодексов практики/поведения, возведения их в статус «отраслевых» законов.

Другим фактором, повышающим роль кодексов практи­ки/поведения, стало стремление международных организаций исполь­зовать такие кодексы в совокупности с новым правовым средством – прямыми договорами между экспортером и импортером данных с обусловленными стандартами защиты передаваемых данных для преодоления трудностей при обмене данными между странами с разными уровнями правовой защиты персональных данных.

**34. Обеспечение достоверности информации в КИС. Программные методы контроля учетной информации**.

**Программные методы контроля учетной информации** представляют собой программное обеспечение, специально предназначенное для выполнения функций защиты информации.

Программные комплексы защиты реализуют максимальное число защитных механизмов:

• идентификация и аутентификация пользователей;

• разграничение доступа к файлам, каталогам, дискам;

• контроль целостности программных средств и информации;

• возможность создания функционально замкнутой среды пользователя;

• защита процесса загрузки операционной системы (ОС);

• блокировка ПЭВМ на время отсутствия пользователя;

• криптографическое преобразование информации;

• регистрация событий;

• очистка памяти

• антивирусные средства защиты

**Аутентификация пользователей** – процесс достоверной идентификации отождествления пользователя, процесса или устройства, логических и физических объектов сети для различных уровней сетевого управления. **Идентификация пользователей** – присвоение каждому пользователю персонального идентификатора: имени, кода, пароля и т.д. Программные системы защиты в качестве идентификатора используют, как правило, только пароль. Пароль может быть перехвачен резидентными программами двух видов. Программы первого вида перехватывают прерывания от клавиатуры, записывают символы в специальный файл, а затем передают управление операционной системой. После перехвата установленного числа символов программа удаляется из оперативной памяти (ОП). Программы другого вида выполняются вместо штатных программ считывания пароля. Такие программы первыми получают управление и имитируют для пользователя работу со штатной программой проверки пароля. Они запоминают пароль, имитируют ошибку ввода пароля и передают управление штатной программе парольной идентификации. Отказ при первом наборе пароля пользователь воспринимает как сбой системы или свою ошибку и осуществляет повторный набор пароля, который должен завершиться допуском его к работе. При перехвате пароля в обоих случаях пользователь не почувствует, что его пароль скомпрометирован. Для получения возможности перехвата паролей злоумышленник должен изменить программную структуру системы. В некоторых программных системах защиты для повышения достоверности аутентификации используются съемные магнитные диски, на которых записывается идентификатор пользователя.

Значительно сложнее обойти блок идентификации и аутентификации в аппаратно–программных системах защиты от несканкционированного доступа к информации. В таких системах используются электронные идентификаторы, чаще всего – Touch Memory.

Для каждого пользователя устанавливаются его полномочия (**разграничение доступа**) в отношении файлов, каталогов, логических дисков. Элементы, в отношении которых пользователю запрещены любые действия, становятся «невидимыми» для него, т. е. они не отображаются на экране монитора при просмотре содержимого внешних запоминающих устройств. Для пользователей может устанавливаться запрет на использование таких устройств, как накопители на съемных носителях, печатающие устройства. Эти ограничения позволяют предотвращать реализацию угроз, связанных с попытками несанкционированного копирования и ввода информации, изучения системы защиты.

В наиболее совершенных системах реализован механизм **контроля целостности файлов** с использованием хэш–функции Причем существуют системы, в которых контрольная характеристика хранится не только в ПЭВМ, но и в автономном ПЗУ (постоянном запоминающем устройстрве) пользователя. Постоянное запоминающее устройство, как правило, входит в состав карты или жетона, используемого для идентификации пользователя. Так в системе «Аккорд–4» хэш–функции вычисляются для контролируемых файлов и хранятся в специальном файле в ПЭВМ, а хэш–функция, вычисляемая для специального файла, хранится в Touch Memory.

После завершения работы на ПЭВМ осуществляется запись контрольных характеристик файлов на карту или жетон пользователя. При входе в систему осуществляется считывание контрольных характеристик из ПЗУ карты или жетона и сравнение их с характеристиками, вычисленными по контролируемым файлам.

Для того, чтобы изменение файлов осталось незамеченным, злоумышленнику необходимо изменить контрольные характеристики как в ПЭВМ, так и на карте или жетоне, что практически невозможно при условии выполнения пользователем простых правил.

Очень эффективным механизмом борьбы с несанкционированным доступом к информации является создание **функционально–замкнутых сред пользователей**. Суть его состоит в следующем. Для каждого пользователя создается меню, в которое попадает пользователь после загрузки ОС. В нем указываются программы, к выполнению которых допущен пользователь. Пользователь может выполнить любую из программ из меню. После выполнения программы пользователь снова попадает в меню. Если эти программы не имеют возможностей инициировать выполнение других программ, а также предусмотрена корректная обработка ошибок, сбоев и отказов, то пользователь не сможет выйти за рамки установленной замкнутой функциональной среды. Такой режим работы вполне осуществим во многих АСУ (автоматизированных системах упроавления).

**Защита процесса загрузки ОС** предполагает осуществление загрузки именно штатной ОС и исключение вмешательства в ее структуру на этапе загрузки. Для обеспечения такой защиты на аппаратном или программном уровне блокируется работа всех ВЗУ, за исключением того, на котором установлен носитель со штатной ОС. Если загрузка осуществляется со съемных носителей информации, то до начала загрузки необходимо удостовериться в том, что установлен носитель со штатной ОС. Такой контроль может быть осуществлен программой, записанной в ПЗУ ЭВМ. Способы контроля могут быть разными: от контроля идентификатора до сравнения хэш–функций. Загрузка с несъемного носителя информации все же является предпочтительнее.

Процесс загрузки ОС должен исключать возможность вмешательства до полного завершения загрузки, пока не будут работать все механизмы системы защиты. В КС достаточно блокировать на время загрузки ОС все устройства ввода информации и каналы связи.

При организации многопользовательского режима часто возникает необходимость на непродолжительное время отлучиться от рабочего места, либо передать ЭВМ другому пользователю. На **время отсутствия пользователя** необходимо блокировать работу ЭВМ. В этих случаях очень удобно использовать электронные идентификаторы, которые при работе должны постоянно находиться в приемном устройстве блока идентификации ЭВМ. При изъятии идентификатора гасится экран монитора и блокируются устройства управления. При предъявлении идентификатора, который использовался при доступе к ЭВМ, осуществляется разблокировка, и работа может быть продолжена. При смене пользователей целесообразно производить ее без выключения ЭВМ. Для этого необходим аппаратно–программный или программный механизм корректной смены полномочий. Если предыдущий пользователь корректно завершил работу, то новый пользователь получает доступ со своими полномочиями после успешного завершения процедуры аутентификации.

Одним из наиболее эффективных методов разграничения доступа является **криптографическое преобразование информации**, которое основано на использовании методов шифрования (кодирования) информации. Этот метод является универсальным. Он защищает информацию от изучения, внедрения программных закладок, делает операцию копирования бессмысленной. При шифровании данные обрабатываются специальной шифрующей программой, что не дает злоумышленникам восстановить данные из зашифрованного кода. Для этого требуется ключ (обычно это длинная бинарная последовательность), при помощи которого и просходит шифрование и дешифрование текста. Криптосистемы с секретными ключами используют один и тот же ключ для шифрования и дешифрования. Криптосистемы с открытым ключом используют секретный ключ для расшифровки, и открытый ключ – для шифрования. В системах с открытым ключом кто угодно может послать секретное сообщение, но прочитать его сможет лишь тот, коту оно предназначено.

Для своевременного пресечения несанкционированных действий в отношении информации, а также для контроля за соблюдением установленных правил субъектами доступа, необходимо обеспечить **регистрацию событий**, связанных с защитой информации. Степень подробности фиксируемой информации может изменяться и обычно определяется администратором системы защиты. Информация накапливается на ВЗУ (внешних запоминающих устройствах). Доступ к ней имеет только администратор системы защиты.

Важно обеспечивать **стирание информации в ОП** и в рабочих областях ВЗУ. В ОП размещается вся обрабатываемая информация, причем, в открытом виде. Если после завершения работы пользователя не осуществить очистку рабочих областей памяти всех уровней, то к ней может быть осуществлен несанкционированный доступ.

В настоящее время наиболее острую проблему безопасности (даже в тех системах, где не требуется сохранять секретную информацию, и в домашних компьютерах) составляют вирусы[[1]](#footnote-1)\*. Исследования международного института по компьютерной безопасности (ISCA) показали, что вирусы являются причиной более 40% случаев потери данных в информационных системах. Поэтому использование эффективных **антивирусных средств защиты** является критическим для безопасности корпоративной информационной системы.

Основным средством защиты от вирусов служит **архивирование***.* Другие методы заменить его не могут, хотя и повышают общий уровень защиты. Архивирование необходимо делать ежедневно. Архивирование заключается в создании копий используемых файлов и систематическом обновлении изменяемых файлов. Это дает возможность не только экономить место на специальных архивных дисках, но и объединять группы совместно используемых файлов в один архивный файл, в результате чего гораздо легче разбираться в общем архиве файлов. Наиболее распространенные программы архивирования: *zip* и *rar.*

Файлы рекомендуется периодически копировать на специальную дискету. Их **резервирование** важно не только для защиты от вирусов, но и для страховки на случай аварийных ситуаций или чьих–то действий, в том числе собственных ошибок.

В современных **антивирусных программах** используются различные методики обнаружения вирусов: сканирование сигнатур (для борьбы с вирусами, использующими неизменный код), проверка целостности (путём создания и использования базы контрольных сумм файлов), эвристические методы (анализ программы по выявлению таких действий, как форматирование жёсткого диска), полиморфный анализ (в специальной защищённой области), анализ на наличие макровирусов (они распространяются, например, с файлами MS Word, MS Excel, MS Access). Наряду с этими средствами ряд пакетов содержат дополнительные функции защиты от почтовых вирусов и вирус–модулей ActiveX и Java–аплетов. Некоторые антивирусные пакеты (например, Panda Antivirus Platinum и PC–cillin) блокируют доступ компьютера к подозрительным Web–страницам.

Определенным недостатком всех антивирусных пакетов является то, что они сильно загружают систему при работе. Это может проявляться в замедлении работы компьютера, в особенности, если часть модулей активизируются по умолчанию. Избежать большой загрузки процессора и оперативной памяти поможет отключение ряда не слишком необходимых функций сложных мониторингов, оставляя лишь самые простые (например, антивирусный монитор).

Среди антивирусных программных продуктов можно отметить, прежде всего, пакеты: Norton Antivirus (Symantec), Virus Scan (McAfee), Dr.Solomon AV Toolkit (S&S Intl.), AntiVirus (IBM), InocuLAN (Computer Associates) и AntiViral Toolkit Pro (Лаборатория Касперского). Данные программные продукты отвечают требованиям ICSA, отслеживая 300 наиболее распространённых и хотя бы 9 из каждых 10 остальных вирусов. В той или иной степени данные антивирусные программы обладают функциями проверки на вирусы и удаления их в реальном времени, отключения заражённых рабочих станций от сети, определения источника заражения, проверки сжатых файлов в режимах сканирования и реального времени. Кроме того, эти программы позволяют проводить удаленное сканирование ПК с Windows NT и ведут единый журнал событий, a Norton Antivirus и InocuLAN позволяют также обновлять клиентские программы с сервера и устанавливать уровень загрузки процессора на сервере при фильтрации проходящих сетевых пакетов.

Новые версии пакета Norton AntiVirus содержат ряд передовых технологий: автообновление антивирусного механизма для определения новых типов вирусов (при обновлении антивирусных баз пользователь получает обновление самой программы, это позволяет существенно сэкономить расходы, время, способствует обеспечению моментальной защиты против новых вирусов раньше, чем они успеют нанести ущерб), технологию Striker32 (обнаружение вирусов и восстанавление данных, поврежденных сложными полиморфными вирусами), эвристическую технологию Bloodhound (для определения неизвестных макровирусов с помощью эвристической технологии и восстановления данных, поврежденных макровирусами), функцию автозащиты (непрерывная фоновая защита файлов, загружаемых из Интернета, локальной сети, вложеных в электронную почту файлов, с гибких и жестких дисков, CD–ROM и сетевых дисков). Кроме того, Norton AntiVirus проверяет и восстановливает файлы в архивах путем обнаружения вирусов, не позволяя им распаковаться и задействоваться (для архиваторов ZIP, LZH/LHA, ARJ, MIME/ UU, CAB, PKLite и LZEXE).

По проведенным специалистами исследованиям антивирусных сканеров и резидентных перехватчиков (мониторов) для ОС Windows NT (Windows 2000), были определены лучшие программные продукты. Ими оказались AVP («Лаборатория Касперского»), NAI McAfee VirusScan, Symantec Norton AntiVirus, CA InnoculatelT, Dr Web («ДиалогНаука»), Norman Virus Control, Command AntiVirus и Sophos Anti–Virus. Результаты тестирования антивирусных сканеров для Windows NT приведены в таблице 1.

##### Таблица 1

Лучшие антивирусные сканеры для Windows NT

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Антивирусная программа** | **Доля обнаруженных вирусов, %** | | | | |
| **загрузочные вирусы** | **файловые вирусы** | **макро–комаидные вирусы** | **полиморфные вирусы** | **стандартные вирусы** |
| AVP (Лаборатория Касперского) | 100 | 100 | 100 | 100 | 100 |
| NAI McAfee Virusscan | 100 | 99,9 | 99,9 | 100 | 100 |
| Symantec Norton AntiVirus | 97,3 | 100 | 99,4 | 93,9 | 99,7 |
| С A InnoculatelT | 100 | 100 | 99,7 | 96,9 | 99,9 |
| Dr Web (Диалог Наука) | 100 | 99,1 | 99,3 | 99,8 | 99,7 |
| Norman Virus Control | 100 | 100 | 99,8 | 96,9 | 100 |
| Command AntiVirus | 100 | 99,4 | 99,8 | 98,0 | 100 |
| Sophos Anti–Virus | 100 | 97,9 | 98,2 | 96.9 | 99,5 |

Источник:В. Митин. Лучшие антивирусы лета 1999 г. // PCWeek/RE, 11 сентября 1999 г.

Ряд антивирусных пакетов обладает возможностью проверки различных типов ресурсов Интернет до того, как они пройдут через почтовый шлюз и попадут в корпоративную сеть (HTTP, FTP, SMTP). При этом конфигурация проверки поддерживает различные сценарии для внутреннего и внешнего траффика, направляя защиту от вирусов туда, где она необходима. Проводится проверка и восстанавление сообщений электронной почты, вложенных в нее архивов ZIP и MIME–кодированных файлов. Такие системы не требует установки и конфигурации программ на клиентских машинах, пользователи могут не отключать антивирусную защиту, так как прокси–сканером не задействуется.

**Литература.**

1. Бакланова О.Е. Информационные системы: Учебно–практическое пособие. – М.: Изд. центр ЕАОИ. 2007. – 270 с.

2. Голицына О.Л, Максимов Н.В. Информационные системы / Московский международный институт эконометрики, информатики, финансов и права. – М.: 2004. – 329 с.

3. Государственная программа информатизации Республики Беларусь на 2003-2005гг. и на перспективу до 2010 года «Электронная Беларусь».

4. Информационные системы в экономике: учебник. / Под ред. Г.А. Титоренко. – 2–е изд., перераб. и доп. – М.: ЮНИТИ–ДИАНА, 2008. – 463 с.

5. Управление проектами с Primavera: Учебное пособие / Под ред. В.В.Трофимова. – СПб.:Изд–во СПбГУЭФ, 2006.–216с.

6. Филатов С. Правовое обеспечение информационных систем в России и в мире // «Право и жизнь». – №88(11). – 2006.

1. \* *Компьютерный вирус —* это специально написанная небольшая по размерам программа, которая может «приписывать» себя к другим программам (т. е. «заражать» их), а также выполнять различные нежелательные действия на компьютере (например, портить файлы или таблицы размещения файлов на диске, «засорять» оперативную память и т. д.). [↑](#footnote-ref-1)