

ЭЛЕКТРОННЫЕ БИБЛИОТЕКИ В ОБРАЗОВАНИИ



Москва 2009

**А.Б. Антопольский, Т.С. Маркарова,
О.П. Крюкова, А.А. Харламов**

ЭЛЕКТРОННЫЕ БИБЛИОТЕКИ В ОБРАЗОВАНИИ

МОНОГРАФИЯ

**Под редакцией
О.П. Крюковой, А.А. Харламова**

Москва 2009

Содержание

Введение	3
Глава 1. Электронная библиотека как информационная база обеспечения образования	7
1.1. Опыт, проблемы и перспективы развития ЭБ	12
1.2. Методы комплектования электронных библиотек.....	27
1.3. ЭБ как информационно-поисковая система	37
Глава 2. Технология автоматической обработки текста для извлечения и представления знаний	44
2.1. Автоматическая обработка текстовой информации	44
2.2. Технология обработки и представления текстов	45
2.3. Психо-физиологическая основа данного подхода	48
2.4. Искусственная нейронная сеть, реализующая структурный анализ информации и ассоциативное обращение к ней.....	50
2.5. Искусственная нейронная сеть, реализующая перенормировку информации.....	54
2.6. Модель мира как две взаимосвязанные ассоциативные сети.....	55
2.7. Образная и текстовая модели мира	57
2.8. Приложения на основе технологии. Семейство программных продуктов TextAnalyst	57
2.9. Электронная книга.....	59
2.10. Персональная библиотека	60
Глава 3. Технология ручной обработки текстовой информации для извлечения и представления знаний в виде фреймовой модели.....	64
3.1. Теоретические основы ФМПЗ.....	65
3.2. Практическое применение ФМПЗ на материале подъязыков различных учебных дисциплин	69
Глава 4. Интегрированная лингводидактическая технология. Концепция	81
4.1. Интеллектуальный уровень	82
4.2. Теоретико-методический уровень.....	83
4.3. Технологический уровень	84
4.4. Педагогический сценарий работы с ИПЭБ.....	87
Заключение	92
Литература.....	94

Введение

Эта книга для широкого круга читателей: школьных учителей, преподавателей ВУЗов, администраторов и образовательных и библиотечных технологов, всех тех, кто сегодня, когда развитие мирового цивилизационного процесса связано с формированием информационного общества¹, занимается проблемами организации образования и культуры.

Россия также вступает на путь формирования информационного общества². Изменения в этой сфере происходят с большим отрывом от существующей педагогической практики. Учитель оказывается не готов к участию в этих процессах. В чем же проявляется эта неготовность?

Полную картину того, что делается в направлении развития информационного общества сегодня в развитых странах мы находим в программе НИР Европы. Вот вкратце история вопроса. Европейский Союз спонсирует передовые научные исследования проблем информационных и коммуникационных технологий обеспечения доступа, хранения и использования культурных и информационных ресурсов. Программа «Электронная культура» (англ. «Digital Culture» или сокращ. «DigiCult») управляется отделением «Культурное наследие и самостоятельная работа учащегося в компьютерной среде» (англ. «Cultural heritage and technology enhanced learning»). Данное отделение также организует исследования по использованию информационных и коммуникационных технологий (Information Communication Technologies, ICTs) для повышения их эффективности.

Два вышеуказанных направления составляют вместе общее стратегическое направление «Информационное общество и Информационная среда». Еще одним из направлений стратегического направления (Directorate-General) «Содержание электронных материалов и когнитивные системы»³. В сентябре 2002 года Совет Министров и Европейский Парламент приняли 6 рамочную программу Европейского сообщества – программу научных исследований и технологического развития, которая получила название Рамочная программа 6 или Framework Programme 6 или сокращ. FP6. Бюджет этой программы составил 17.5 млрд. евро до конца 2006 года. Одним из семи тематических приоритетов FP6 было направление «Технологии информационного общества» (Information Society Technologies или сокращ. IST) с соответствующим бюджетом 3.6 млрд. евро на пятилетний период.

Программа FP6 базировалась на концепции «Область европейских научных исследований» (European Research Area) и имела целью обеспечение поддержки деятельности и проектам, направленным на создание партнерских отношений между луч-

¹ Информационное общество отличается от общества, в котором доминируют традиционная промышленность и сфера услуг тем, что информация, знания, информационные услуги, и все отрасли, связанные с их производством и распределением (телекоммуникационная, компьютерная, телевизионная) растут более быстрыми темпами, являются источником новых рабочих мест, становятся доминирующими в экономическом развитии. Мелюхин И.С. Информационное общество и государство. <http://www.relcom.ru/Archive/1997/ComputerLaw/State.htm>

² Стратегия развития информационного общества в России (Утверждена Президентом Российской Федерации В. Путиным 7 февраля 2008г., № Пр-212) .

³ Информация по этим вопросам содержится на сайте <http://cordis.europa.eu/fp7/ict/telearn-digicult/>.

шими европейскими исследователями и исследовательскими центрами. Основной объем финансирования был сосредоточен на крупномасштабных проектах с интегрирующим европейскую научную деятельность потенциалом, с результатом, полученным совместными усилиями исследовательских и образовательных центров, концентрирующими научный потенциал, высокий научный профессионализм, знания, финансовые ресурсы, инфраструктуру.

Исследования, спонсируемые в рамках программы FP6, имела целью:

- улучшение качества услуг, которые люди получают, работая с различными культурными и научными ресурсами в электронной форме,
- разработку систем и инструментов, обеспечивающих доступность и длительное использование электронных ресурсов науки и культуры.

Были получены и другие результаты на фоне решения частных проблем отдельных проектов:

- складываются сообщества образовательных, исследовательских и административных центров, имеющие общее представление об актуальных проблемах технологий информационного общества и их решениях,
- это общее видение встраивается в учебные пособия, программы семинаров и тренингов для учителей, программы академической мобильности, новые административные (министерские) решения по развитию системы образования в Европе.

Миссией библиотечного проекта, как и проекта образовательного, по представлениям европейских специалистов⁴, должно стать следующее: «электронные библиотеки станут универсальными хранилищами и проводниками коммуникации будущего, обычными для каждого человека средствами доступа, анализа, оценки, трансформации и обмена любыми формами информации. Они будут незаменимыми средствами в повседневной личной и профессиональной жизни людей, доступными в любое время, в любом месте в легкой для пользователя мультимедийной, интерактивной компьютерной среде».

В итоге, можно сказать, что еще одним результатом является то, что складывается новая образовательная культура и парадигма развития информационного общества, построенная на симбиозе идей информатики, психологии, инженерии знаний, искусственного интеллекта, и, в значительной мере, лингвистики.

В свете всего сказанного зададим вопрос снова: к чему же сегодня не готов наш российский учитель? В чем требуется помощь?»⁵

Наш учитель не готов воспринимать информацию о подобных разработках, их результатах, поскольку они содержат новую терминологию новой образовательной парадигмы.

Нет знания и понимания как самой этой парадигмы, так и той технологической инновационной составляющей, с помощью которой можно повысить качество обучения или библиотечного обслуживания в компьютерной среде, что могло бы дать представление о научном прорыве выполненных исследований и их полезности для образовательной практики.

Нет готовности участвовать в этих процессах, хотя такая готовность и активное участие учителя необходимы для становления информационного общества и, в конечном итоге, реализации самих этих новых технологий информационного общества. Далее мы объясним почему. Этот фактор также обусловлен инновационной составляющей НИР как результата развития технологизации работы с текстом.

⁴ DELOS: a Network of Excellence on Digital Libraries. www.delos.info

⁵ Отметим здесь, что результаты и финансы Европейского НИРа на Россию пока активно не распространяются.

А в чем же состоит инновационная составляющая современных образовательных и библиотечных НИР, способная обеспечивать повышение качества услуг и автоматизацию и технологизацию процессов интеллектуального труда? С чем конкретно представляется необходимым знакомить учителя? Ответ на эти вопросы может быть дан как на материале европейских, так и отечественных исследований.

Историческое развитие процесса технологизации и автоматизации интеллектуального труда в целом привело к тому, что начали вырабатываться методы структурной обработки информации и построения нелинейных концептуальных (вербальных, понятийных) моделей знаний, дающих, с одной стороны, и при безмашинном применении, быстрое, полное, наглядное представление о квинтэссенции смысла предметной области и способах решения задач. С другой, они стали использоваться как материал для формализации знаний, поиска, машинной обработки. Эти нелинейные вербальные модели получили названия: семантические сети, фреймы, графы, и др. Стал очевиден новый метаязык человекомашинного интеллектуального взаимодействия, от наличия и использования которого и будет зависеть качество информационных продуктов, программ и сервисов в дальнейшем.

Суммируя сказанное, не будет преувеличением сказать, что с выдвиганием глобальной идеи построения информационного общества в мире начался процесс интенсивного поиска путей усовершенствования электронных услуг. И особый интерес и важность для повышения качества этих услуг придается сегодня решениям в области анализа смысловых аспектов текстовой информации, технологиям автоматизации работы с текстом, проблемам автоматического извлечения и представления знаний. Начала складываться новая лингвистическая когнитивная парадигма, в рамках которой мы, авторы данной книги, и ищем инновационные решения для развития образовательных технологий и технологии ЭБ.

Что конкретно рассматривается и предлагается в данной работе? Книга содержит три различные точки зрения на проблему организации работы с текстовой информацией, написанные специалистами в области библиотечного дела и развития электронных библиотек: президентом Национальной Ассоциации Электронных Библиотек А.Б. Антопольским и директором Государственной научно-педагогической библиотеки К.Д. Ушинского Т.С. Маркаровой, специалистом в области автоматической обработки текстовой информации для извлечения данных ТекстАналист – Харламовым А.А., и специалистом в области технологии лингводидактики Крюковой О.П. Мы рассматриваем эти точки зрения с позиций их различий, а также и как взаимодополнительные, аргументирующие и обеспечивающие возможность разработки образовательной технологии, построенной на единой трактовке знаний.

Ознакомившись с материалом книги, читатель увидит, что все эти специалисты представляют свой подход к моделированию знаний через систему ключевых слов и их иерархического упорядочивания: для библиотечной сферы это «рубрикатор» или «отраслевой тезаурус», для автоматической обработки текста и извлечения данных – это «квазинейронная сеть», для лингводидактического подхода обучения предметной области через ее язык – это «концептуальная модель» или «фрейм». Суммируя и несколько нивелируя различия, можно сказать, что это иерархическая понятийная⁶, но в то же время абстрактная и вероятностная модель знаний, которая строится на основе принципа родо-видовых отношений. Но при этом цели в использовании моде-

⁶ О подобных требованиях к структуре содержания говорится и в коллективной монографии Горбунова Е.И., Лобачева С.Л. и др. Открытое образование. Стандартизация описания информационных ресурсов. М.: Альфа, 2003.

лей знаний разные – для библиотеки – найти нужный документ, для системы извлечения данных – сформировать из ключевых слов смысловой портрет текста для быстрого понимания, для лингводидактики – разработать концептуальную модель, представляющую способ решения задачи или структуру предметной области или др. Комплексное и полифункциональное использование всех трех подходов дает возможность осмыслить целое под новым практическим углом зрения – как *проблему организации творческого взаимодействия преподавателя и учащегося в работе над текстовой информацией в условиях когнитивной лингвистической лаборатории, содержащей электронную библиотеку, возможности автоматизации обработки текстовой информации, а также эксплицитно (открыто) данные модели знаний, построенные на едином антропо- и компьютерно-сообразном принципе моделирования знаний через использование естественных понятий языка.*

Книга содержит как новейшие проектные, так и практически реализуемые в сегодняшних условиях решения. Знакомство с данной книгой поможет уже сегодня и сейчас по-другому решать проблемы организации учебного труда, найти рациональное применение новой техники, привести элементы творчества, экономии энергии, рационализма в свой труд.

Как читать данную книгу? Ее можно читать в различном порядке и с разными целями.

Глава 1 содержит проектные решения и идеи развития электронных библиотек в России. Ее прочтение помогает понять, что будет, аргументирует перспективный лингвистический поисковый информационный подход к определению знаний, как эти будущие решения могут повлиять на сегодняшние. Она написана Т.С. Маркаровой (разделы 1, а также 1.2; раздел 1.1. написан А.Б. Антопольским).

В главе 2 показано, как в результате использования системы ТексАналист можно получить извлеченными ключевые слова – заготовку для разработки модели знаний, как применить функциональные возможности этой системы для эффективной работы с текстовой информацией. Она написана Харламовым А.А.

В главах 3, 4 дается проектное предложение образовательной технологии организации обучения в электронной среде (Интегрированная лингводидактическая технология) и педагогическая концепция Интеллектуальной персональной библиотеки преподавателя учащегося. Они написаны Крюковой О.П., как и Введение и Заключение.

Почему мы ее предлагаем в дополнение к идее и проекту НЭБ, изложенному в главе 1 и 2?

Мы рассматриваем свое предложение и саму идею ИПЭБ как возможную практическую реализацию инициатив по проблеме становления «Новой школы», высказанных президентом Медведевым⁷ в послании Федеральному собранию. И если идти по этому пути и создавать такие ИПЭБ как ресурсные лингвистические центры, то в школе должны появиться места для самостоятельной коммуникативной работы преподавателей и учащихся, творчества, углубленного обучения по индивидуальной траектории, оснащенные программным инструментарием, автоматизирующим работу с текстовой информацией, поступающей как из внутренней библиотеки, так и из внешних. Идею такой библиотеки – лингвистической лаборатории, и технологию работы в ней мы и предлагаем в главе 4. Предлагается и программное обеспечение, рабочая версия которого имеется на сайте <http://www.analyst.ru>. Это означает, что преподаватель может начать строить свою ИПЭБ, совместно с другими коллегами, уже сегодня.

⁷ Послание Президента РФ Федеральному собранию от 5 ноября 2008 года <http://www.garant.ru/hotlaw/doc/124638.htm>.

Глава 1

Электронная библиотека как информационная база обеспечения образования

В настоящее время в России идет процесс формирования подходов к внедрению технологий информационного общества на федеральном, региональном и муниципальном уровнях.

В 2000 году Россия подписала Окинавскую хартию глобального информационного общества (Окинавская хартия глобального информационного общества // Развитие информационного общества в России: Сб. документов и материалов. СПб., 2001), в которой подчеркивается, что информационно-коммуникационные технологии (ИКТ) становятся жизненно важным стимулом развития мировой экономики и формулируется задача – превратить потенциальные возможности этих технологий в реальную силу.

Проработка концептуальных вопросов развития информационного общества становится весьма актуальной задачей также в связи с тем, что в России уже начинают осуществляться программы, направленные на развитие этих технологий. В первую очередь это Федеральная целевая программа (ФЦП) «Электронная Россия» (Федеральная целевая программа «Электронная Россия (2000–2010 годы)» // Развитие информационного общества в России. Т. 2. Концепции и программы: Сб. документов и материалов. СПб., 2001) и Национальная стратегия «Россия в информационном веке». Методологическими предпосылками к разработанным и разрабатываемым программам и проектам являются:

- Влияние развития сети Интернет и современных информационно-коммуникационных технологий на изменение системы социальной, научной и образовательной коммуникации;
- Трансформация социальных институтов в условиях развития информационного общества и распространения ИКТ;
- Развитие on-line сообществ, их взаимодействие между собой и традиционными структурами современного общества;
- Расширение зоны влияния электронной информации в научно-образовательной инфраструктуре современного мира;
- Решение Проблем сохранения научной и культурной идентичности в условиях глобализации, использование ИКТ для сохранения и развития национально-культурного и национально-научного достояния, формирование электронных коллекций и библиотек;
- Развитие дистанционного обучения, формирование образовательных консорциумов и системы открытого образования.

В рамках научно-технической программы Министерства образования и науки Российской Федерации «Создание системы открытого образования» Центром информационно-аналитического обеспечения системы дистанционного образования (ЦИАН) по итогам социологического мониторинга были обозначены следующие задачи исследования (Мониторинг и прогнозирование спроса населения России на образовательные услуги, предоставляемые с помощью технологий дистанционного обучения: отчет о НИР. НТП «Создание системы открытого образования» / Минобробразования РФ; ЦИАИ. М., 2003):

- Разработка методологии мониторинга и прогнозирования спроса на образовательные услуги;
- Исследование актуального состояния образовательных потребностей населения России;
- Изучение потребительских требований к условиям обучения;
- Определение перспективного ассортимента услуг дистанционного обучения в сфере базового профессионального образования, высшего профессионального, послевузовского и дополнительного образования;
- Исследование факторов влияния на спрос на услуги дистанционного обучения (материально-технических возможностей населения для потребления таких услуг, уровня информированности населения об этих услугах и др.);
- Количественная оценка спроса на отдельные виды услуг дистанционного обучения, расчет численности потенциального контингента потребителей услуг;
- Прогнозирование динамики спроса на услуги дистанционного обучения на среднесрочную перспективу.

Обобщая данные социологических опросов, исследователи пришли к выводу, что в российском обществе есть значительный психологический и образовательный ресурс для модернизационного рывка и перехода к информационному обществу.

Актуальные проблемы межкультурного и межнаучного взаимодействия в рамках нового информационного общества и информационного пространства на нынешнем витке развития требуют качественной научно-информационной поддержки. Для этого наряду с освоением высокотехнологичных информационных технологий необходимо учитывать еще и разнообразие культур наравне с богатейшим культурным и научным потенциалом конкретных участников информационной инфраструктуры. В условиях стремительно происходящих глубинных изменений в научно-информационной среде мира гуманистический подход к созданию научных, информационных и образовательных ресурсов становится все более ощутимой жизненной потребностью, которую контролирует и обеспечивает государство.

Впервые на международном уровне попытка регламентировать обязанности государств и международных организаций по подготовке человека к жизни в информационном обществе предпринята ЮНЕСКО в документе «Рекомендации о развитии и использовании многоязычия и всеобщем доступе к киберпространству». В этом документе сказано: «Государствам-членам и международным организациям следует содействовать распространению грамотности в области информационных и коммуникативных технологий (ИКТ). Решающее значение для информационного общества имеет развитие человеческого капитала, включая открытое, интегрированное и межкультурное образование в сочетании с обучением навыкам в области ИКТ». Это положение свидетельствует не столько о технологическом, сколько о гуманитарном подходе к решению проблемы подготовки человека к условиям информационной цивилизации.

Гуманитарный подход к проблеме отражает и одна из флагманских программ ЮНЕСКО – Программа «Информация для всех», знаменующая существенную революцию в осознании проблемы информационной культуры как гуманитарной. Доминантой новой мировой информационной политики становятся не только технологии и сама информация, а ее создатель и конкретный потребитель – человек.

Наряду с ЮНЕСКО разработку проблемы подготовки человека к жизни в информационном пространстве инициирует Международная федерация библиотечных ассоциаций (ИФЛА – International Federation of Library Associations and Institutions IFLA). На протяжении последних лет эта проблема активно обсуждалась на Генеральных конференциях и сессиях ИФЛА и РБА (Российская библиотечная ассоциация). Секция ИФЛА по информационной грамотности поставила задачу выявления стандартов информационной грамотности, сформировавшихся в разных странах и создания на этой основе международного стандарта.

В условиях современного общества библиотеки являются не только накопителями информационных источников и коллекций, но и проводниками в области новой информационной цивилизации, составными частями которой являются как информационная грамотность, так и информационная культура. Будучи в лице библиотекарей и библиографов создателями и хранителями информации, библиотеки призваны оказывать полномасштабную помощь потребителю информации – читателю, пользователю, абоненту, а, следовательно, наращивать информационно-ресурсную емкость, а также развивать свою и читательскую информационную грамотность и культуру.

Традиционно библиотеки являются важной составляющей образовательного процесса как в широком (общекультурном) смысле этого слова, так и в узком (учебном). Библиотеки становятся активными проводниками внедрения информационных технологий в отраслевые сферы, беря на себя задачу по созданию информационных ресурсов и мультимедиапродуктов. Деятельность библиотек оценивается в большей степени, чем раньше, по их способности выполнять свою коммуникативно-образовательную функцию, т.е. осуществлять целевую информационную поддержку научно-образовательной деятельности и удовлетворять информационные потребности специалистов различных областей науки и образования.

Одним из важных и приоритетных направлений научно-практической деятельности традиционных библиотек является формирование базовой инфраструктуры, развитие сервисов и ресурсов для создания электронных библиотек под своей эгидой. Крупные государственные библиотеки объединяются в консорциумы (Научная электронная библиотека eLibrary, НЭИКОН), ассоциации (ЭБНИТ – Электронные библиотеки и научно-информационные технологии, ЛИБНЕТ – Общероссийская информационно-библиотечная компьютерная сеть, ЭЛБИ – Ассоциация «Электронные библиотеки»). Специфика электронных библиотек состоит в том, что они, с одной стороны, выполняют функции традиционных библиотек, а с другой – функции, характерные для высокотехнологичных автоматизированных информационных систем, – организации и хранения собственных и удаленных электронных ресурсов и обеспечения on-line доступа к ним на основе компьютерных и Интернет-технологий.

Основная задача электронных библиотек – интеграция электронных информационно-образовательных ресурсов, эффективная навигация и обеспечение доступа к ним независимо от местоположения пользователя.

Информационно-образовательные ресурсы – это совокупность научно-педагогической, учебно-методической, хрестоматийной, нормативно-технической и инструктивной информации, к которой обеспечен доступ пользователей, дифференцированный по роли и месту участника образовательного процесса, а также возможность извлечения информационного ресурса в соответствии с личностными особенностями и предпочтениями пользователя.

Интеграция информационно-образовательных ресурсов предполагает их объединение с целью использования с помощью удобного и унифицированного пользовательского интерфейса к различной информации с сохранением ее свойств, особенностей представления и пользовательских возможностей манипулирования с ней. При этом объединение ресурсов не обязательно должно осуществляться физически – оно может быть виртуальным. Главное – оно должно обеспечивать пользователю восприятие доступной информации как единого информационного пространства. Все информационное пространство, доступное пользователю, должно быть представлено в виде совокупности самостоятельных объектов.

На сегодняшний день крупные библиотеки страны продолжают наращивать электронные информационно-ресурсные мощности. В области образовательных ресурсов, на наш взгляд, достаточно объемным и информативным национальным электронным образовательным ресурсом является сайт ГНУ ГНПБ им. К.Д. Ушинского РАО (www.gnpbu.ru). Государственное научное учреждение «Государственная научная педагогическая библиотека им. К.Д. Ушинского Российской академии образования», будучи научно-информационным центром Российской академии образования, де-факто, являясь государственным библиотечно-информационным центром образовательной отрасли в целом в России и одной из немногочисленных крупных педагогических библиотек в мире, обладает большим и уникальным отраслевым педагогическим фондом. По данным статистики на сайт поступает более 30 тысяч запросов в день из более чем 60 стран Европы, Америки, Азии и Австралии. Помимо общих сведений о Библиотеке (исторической справки, режима работы, правил пользования) и методических материалов на сайте представлены следующие информационно-электронные ресурсы.

1) Электронный каталог. Насчитывает более 270 тысяч библиографических записей информационных материалов по педагогике, психологии и смежным отраслям знаний, отражает фонд Библиотеки начиная с 1994 года. Поисковыми параметрами являются как значимые элементы библиографического описания, так и ключевые слова, содержащиеся в документах. Электронный каталог Библиотеки имеет локальную версию и web-версию. Использование web-версии электронного каталога дает возможность дистанционным пользователям получить сведения о наличии книг и журналов в фондах Библиотеки. В ближайшей перспективе намечено создание полного электронного каталога ГНПБ им. К.Д. Ушинского – ретроконверсия всего карточного каталога со дня основания Библиотеки (с 1925 года).

2) Полнотекстовая база авторефератов по педагогике. Ведется с 2000 года, насчитывает 3700 текстов авторефератов диссертаций и активно пополняется новыми поступлениями. Это наиболее востребованный информационный ресурс Библиотеки. Необходимость такой базы подтверждается многими исследователями, работающими в области педагогики и образования.

3) Аннотированный каталог «Образовательные ресурсы Интернет». Содержит 608 ссылок, отражающих информационно насыщенные легитимные сайты по педагогике и образованию, которые могут быть использованы в учебных и образовательных целях.

4) Электронный журнал «Образование: исследовано в мире». Функционирует под патронажем ГНПБ им. К.Д. Ушинского и Российской академии образования. Журнал содержит оригинальные тексты статей, монографий и диссертаций, передаваемых авторами в Библиотеку в электронном виде.

5) Русско-английский тезаурус по педагогике ГНПБ им. К.Д.Ушинского (На сайте расположена рекламная информация, краткая характеристика с небольшим фрагментом Тезауруса). Тезаурус разработан на основе Тезауруса ЮНЕСКО, ведется с 1993 года. Насчитывает более 5000 терминов и функционирует в качестве информационно-поискового языка вербального типа в ЭК ГНПБ.

6) База данных «Труды РАО». Отражает весь массив источников, поступающих в фонд Библиотеки из институтов Российской академии образования начиная с 1994 года. Обновляется один раз в месяц.

7) База данных «Новые поступления». Информировывает о поступлениях за последний месяц.

В соответствии с концепцией создания единого информационного образовательного пространства в РФ, а также в рамках национальной задачи формирования в России информационного общества, интеграция информационно-образовательных ресурсов направлена на достижение следующих целей.

1. Обеспечение доступа пользователей к документам из фондов традиционных библиотек, предоставление которого затруднено в силу ряда причин: удаленности пользователя от библиотеки, малой экзemplарности документа и т.д.

2. Обеспечение доступа к информации, существующей исключительно в электронной форме.

3. Целенаправленное информационное обеспечение пользователей путем формирования полнотекстовых баз данных и предоставления доступа к ним на основе Законодательства РФ об авторском праве.

Функциональные требования к интегрированным информационно-образовательным ресурсам заключаются: в обеспечении доступа к разнородным электронным документам в базе данных электронной библиотеки (ЭБ) из одной точки (в среде одного экрана) с помощью единого интерфейса, в единой поисковой среде, объединяющей разнородные электронные коллекции; в реализации новых форм библиотечно-информационного обслуживания пользователей; в предоставлении удобного, оперативного доступа к легитимным цифровым копиям оригиналов печатных изданий.

Прототип программных средств должен быть построен по прогрессивной технологии и включать в себя инструментальные средства рабочего информационного пространства, обеспечивающие интеграцию новых функциональных возможностей и расширение числа предлагаемых пользователям услуг. Программное обеспечение – основное инструментальное средство в интегрированной информационной системе – должно сочетать в себе как функциональность, присущую типичным СУБД (система управления базой данных) – поиск данных, создание и модификация данных, так и различные интерфейсы взаимодействия с другими специфическими возможностями, диктуемыми потребностями пользователей, а именно:

- работу с авторским набором данных;
- работу с метаданными;
- работу с распределенными наборами данных;
- работу с различными фильтрами и конверторами, обеспечивающими взаимодействие с гетерогенными распределенными электронными библиотеками;
- работу с разнообразными поисковыми средствами.

Таким образом, в современную эпоху при уже существующем информационном насыщении общества актуальна как проблема создания информационно-образовательного пространства, проблема широкого и грамотного использования информационных ресурсов в научных и образовательных целях, так и проблема обеспечения оперативного доступа к информационно-образовательным ресурсам.

Для успешного и эффективного использования информационно-образовательного пространства в условиях создаваемого информационного общества необходимо: обеспечение устойчивого взаимодействия образовательной и научной сфер в деле создания справочно-информационных систем и легитимных полнотекстовых баз данных; разработка совместных организационно-методических принципов и правил для облегчения взаимодействия на межведомственном уровне; определение подходов к созданию и использованию электронных информационно-образовательных ресурсов в интересах отечественной и зарубежной науки и образования; формирование информационной среды, обеспечивающей взаимодействие и оперативное обсуждение актуальных общих проблем; поддержка инициатив, помощь в нахождении партнеров для реализации социально-значимых и научно-обоснованных проектов.

1.1. Опыт, проблемы и перспективы развития ЭБ

Первые шаги по созданию электронных библиотек были сделаны за рубежом в начале 1980-х годов. В 1992 году конференция Национального научного фонда США положила начало использованию самого понятия «цифровая библиотека» в современном контексте. В 1990-х годах в США и странах Европы начинают активно разрабатываться программы развития электронных библиотек, а в 1995 году Европейская комиссия выдвигает создание электронной библиотеки в число приоритетов.

Более активно развиваются национальные программы, причем не только для текстовых электронных библиотек, но и в форме мультимедийных культурно-образовательных проектов – «Память Америки», «Память Испании», «Память мира» (Чешская республика), SCRAN (Шотландская сеть ресурсов по культурному наследию для поддержки образования, кумулирующая ресурсы музеев, архивов и библиотек) и многие другие. Отличительной чертой таких проектов является участие в их реализации большого числа различных организаций: библиотек, музеев, архивов, университетов.

В настоящее время идет активный процесс создания национальных электронных библиотек («Цифровая библиотека» Великобритании, «Немецкая библиотека» и др.). Преимуществом при оцифровывании пользуются редкие книги и рукописи. Другой организационной моделью кооперации электронных ресурсов являются консорциумы библиотек, создаваемые для корпоративных закупок и координированной оцифровки фондов. Сервер, содержащий информацию о каталогах и других данных, обычно располагается в одной из библиотек-участниц. Другие библиотеки получают к нему сетевой доступ. Примером такого консорциума служит проект DEBORA (Digital Access to Books of the Renaissance), который осуществляется в рамках программы Европейского Союза в области библиотек и обеспечивает доступ к электронным коллекциям книг эпохи Возрождения.

В конце прошлого века Библиотека Конгресса США начала реализацию национальной программы создания электронной библиотеки. Начиная с 1994 г., по инициативе NSF, DARPA и NASA в США была развернута исследовательская программа «Digital Libraries Initiative» (DLI) по электронным библиотекам. На второй стадии развития в

начале 1998 г. эти программы были объединены в единую межведомственную программу (**DLI – Phase 2**), в которой, кроме того, участвовали Национальная медицинская библиотека, Агентство по статистике США, Национальный гуманитарный фонд, Национальный архив США и другие федеральные агентства.

В университетах и в университетских библиотеках и издательствах США при финансовой поддержке со стороны государства, благотворительных фондов и корпораций осуществляются передовые исследования в области ЭБ, разрабатываются стандарты и создаются разного рода цифровые коллекции. Университетские электронные библиотеки, однако, предоставляют лишь ограниченный доступ – либо только своим сотрудникам и студентам, либо некоторому числу других университетов и культурных учреждений, получая в обмен доступ к их цифровым коллекциям.

Созданием и поддержкой ЭБ и других электронных продуктов занимается целый ряд коммерческих компаний, продающих как сами эти продукты на жестких носителях, так и лицензии на доступ к цифровой информации он-лайн. Так, компания «ProQuest» (<http://www.proquestcompany.com>), являющаяся одним из крупнейших провайдеров информационных услуг, лицензирует доступ к полным текстам тысяч периодических изданий, диссертаций, книг и других публикаций, включая обширные электронные ресурсы компании «Chadwyck-Healey», которая в 1999 г. стала подразделением «ProQuest».

В последние годы стали возникать электронные коллекции принципиально нового типа. Так, появилось понятие *веб-архив*, где объектом сохранения, описания и классификации выступают самые разнообразные Интернет-ресурсы: веб-страницы, целые сайты, архивы телеконференций и т. п.

Важной частью общегосударственной политики в области информации и информационных технологий США, Великобритании и других стран является развитая система учета информационных ресурсов и средств информационных технологий, создаваемых федеральными учреждениями, а также анализ эффективности используемых на эти цели бюджетных средств, и регулярная отчетность федеральных ведомств на всех уровнях.

Последние годы внушительная работа по созданию ЭБ ведется практически во всех странах Западной Европы, как на общеевропейском, так и на национальном и местном уровнях. Некоторые проекты охватывают предметные области (например, гуманитарные науки), другие – типы изданий (периодические издания, редкие книги или изображения), третьи сосредотачиваются на проблемах и задачах, связанных с технологией создания ЭБ (таких, как интеллектуальная собственность, цифровые технологии и управление базами данных). Ряд проектов по созданию и использованию электронных библиотек выполнялся в соответствии с Рамочными Программами Комиссии Европейских сообществ (КЕС), включая программу «Технологии информационного общества». Так цель программы Европейской Комиссии «Телематика для библиотек» – обеспечение доступа к знаниям, содержащимся в библиотеках ЕС. Программа охватывает такие темы, как сетевая работа (OSI, Web), каталогизация (OPAC), изображения, мультимедиа, авторское право, а также более ста проектов по проблемам цифровых библиотек.

Содержательные цели программы включают, в том числе, создание Европейской цифровой библиотеки по экономике с распределенным доступом к информационным ресурсам, хранящимся в различных библиотеках, создание виртуальной библиотеки

периодических изданий, обеспечивающей эффективный быстрый электронный доступ к относительно небольшим по объему документам (например, журнальным статьям), создание специализированной системы, посредством которой библиотеки, обслуживающие пользователей с ограниченным зрением, получают доступ к центральной базе данных и возможность скачивать информацию в специфических форматах. Новая инициатива Европейского союза в области электронных библиотек касается создания коллекций национальных книг в Интернете. Одним из важных направлений предполагается поддержка общих стандартов и методических решений.

Создаются архивы электронных препринтов по физике и математике, а также во другим наукам, в том числе, по гуманитарным наукам (психология, лингвистика и др.). Наряду с сайтами препринтов, есть тематические порталы, содержащие ссылки на самые разные Web-сайты.

В большом количестве стран уже реализованы (или находятся в различной стадии реализации) десятки проектов. Причем этот процесс, подкрепленный интеллектуальными и материальными ресурсами, имеет тенденцию к выходу на корпоративный уровень международных и национальных программ формирования и предоставления электронных коллекций. Как правило, в этих проектах принимают участие различные учреждения – библиотеки, университеты, иногда правительственные структуры.

Так *Национальная электронная библиотека США* («National Digital Library Program») – первый опыт систематизированной оцифровки наиболее ценных исторических материалов и обеспечения свободного доступа к ним конгрессменов, ученых, преподавателей, студентов, школьников.

Проект «Гутенберг» – следствие предоставления профессору Майклу С. Харту компьютерного времени стоимостью (по расценкам той поры) в 100 миллионов долларов США и предложения найти способ окупить эти огромные затраты. Список книг, изданных «гутенберговцами» до настоящего времени, поражает своим разнообразием и пестротой.

В *Великобритании* программа Британской библиотеки «Цифровая библиотека» ставит своей целью предоставление цифровых информационных услуг, основанных на фондах этой библиотеки, с последующим развитием возможностей доступа к фондам других крупнейших библиотек мира. ЭБ будет состоять из наиболее важных документов по разнообразным темам и включать в себя тексты, фото- и видеоизображения, звуковые файлы.

В *Японии* создано специальное Агентство по внедрению новых информационных технологий, в состав которого вошли специалисты из Национальной парламентской библиотеки, Национального центра НТИ, других информационных центров, а также из крупных фирм, действующих на рынке информационных технологий. Общая стоимость проектирования электронной «библиотеки XXI века» оценивается в 500 млн. долларов США. К его реализации привлечено девять крупных фирм, в том числе NEC, «Mitsubisi», «Fujitsu». В качестве пилотного подготовлен проект Электронной библиотеки Японии в провинции Кансай. В электронную форму уже преобразовано (главным образом в виде изображений) более 10 млн. страниц различных печатных изданий (книги, журналы, газеты, карты и др.).

В *Германии* действует «Global Info» – проект немецких цифровых библиотек, основанный на сотрудничестве с университетами, издательствами, книготорговцами, информационными центрами, научными обществами, академическими и научными

библиотеками. Его цели – **предоставление эффективного доступа к глобальной информации непосредственно с рабочего места ученого, стимулирование структурных изменений в информационных и коммуникационных процессах научного сообщества страны.**

Во **Франции** развивается проект «*Gallica*» (<http://gallica.bnf.fr>), предоставляющий удаленный доступ к оцифрованным печатным документам, фото- и видеоизображениям, аудиозаписям из фондов Национальной библиотеки Франции. Уже оцифровано более 30 млн. страниц (около трети из них относятся к XIX в.), однако пока доступ предоставлен только к 2 млн. страниц.

Опыт создания электронных библиотек в России достаточно богат и разнообразен, хотя их история насчитывает немногим более 10 лет.

Кажется вполне естественным, что традиционные библиотеки должны быть основными создателями электронных библиотек и «законодателями моды» в этой области. Действительно, именно в них накоплен огромный опыт работы с изданиями, документами, рукописями, именно они находятся ближе всего к этой информации, более того, они уже не одно десятилетие используют вычислительную технику. Наверное, так и будет со временем. Однако до сих пор практика показывала иное. Первые электронные библиотеки создавались либо по частной инициативе (библиотека М. Мошкова), либо в университетах при поддержке различных фондов (фонд Сороса и РФФИ в России).

В настоящее время одним из наиболее крупных и быстроразвивающихся проектов является «Электронная библиотека диссертаций РГБ». Наряду с ним существуют и другие: Библиотека Максима Мошкова, проект «Флогистон», Библиотека «Артефакт», проект «Моя библиотека», проект «Общий текст», Библиотека «Воронеж.Net», Библиотека «Психология на русском языке» и др.

С 2001 г. Российская национальная библиотека начала выборочный перевод изданий, хранящихся в ее фондах, в электронную форму, что положило начало созданию электронной библиотеки. Приоритетными с точки зрения вида материала для перевода в электронную форму являются: рукописные и архивные документы, уникальные изобразительные материалы, первопечатные и раскрашенные вручную карты, редкие книги и другие печатные материалы.

Крупный проект электронной библиотеки под названием «Научное наследие» начат консорциумом академических библиотек во главе с БЕН РАН. В рамках проекта предполагается оцифровать и предоставить в доступ труды российских ученых, оказавших существенное влияние на развитие науки за последние 200 лет.

На сайте Государственной публичной исторической библиотеки России (<http://www.shpl.ru>) представлена Национальная программа сохранения библиотечных фондов в РФ. Проект создания электронных коллекций «Книжные памятники по искусству» для информационного обеспечения научных исследований в области культуры реализует Российская государственная библиотека по искусству (<http://www.liart.ru>) при содействии Российского гуманитарного научного фонда. В его рамках предполагается создание электронных копий наиболее часто спрашиваемых изданий.

В настоящее время двумя национальными библиотеками страны развивается проект «Национальной электронной библиотеки», ставящий своей задачей создание распределенного репертуара электронных копий культурно и научно значимого наследия России. Принята концепция НЭБ, разработаны документы по формированию ее фондов, ведется работа по созданию программного обеспечения.

Электронные библиотеки в образовательных учреждениях. В рамках государственного сектора информационно-библиотечной сферы ведущей категорией учреждений по созданию электронных коллекций являются высшие учебные заведения. Университетские коллекции чаще всего ориентированы на поддержку системы открытого образования и включают в себя учебники, учебные и методические пособия, лекции и т.п. материалы. Вместе с тем в университетских электронных библиотеках все в большей степени представляется научная литература: ранее опубликованные книги и статьи по широкому спектру научно-технических направлений, материалы конференций, результаты научной деятельности сотрудников и студентов университетов, вузовские периодические издания. Общий объем накопленной в университетских электронных библиотеках информации исчисляется десятками тысяч книг, сотнями тысяч статей, большим количеством изображений и других материалов.

Университетские электронные библиотеки при фактическом отсутствии координации играют заметную роль в распространение учебной и научной информации. Подтверждением этого является достаточно большое число обращений. Ежедневное количество посетителей одной университетской электронной библиотеки лежит в достаточно широком диапазоне: от нескольких десятков до 12–15 тыс.

В рамках Федеральной целевой программы «Развитие единой образовательной информационной среды (2001-2005 г.г.)» была создана Центральная библиотека образовательных ресурсов (ЦБОР, <http://www.edulib.ru>), содержащая электронный фонд учебной литературы, с которыми можно вести работу через Интернет после регистрации. ЦБОР была призвана обеспечить накопление электронных версий учебников и учебных пособий, электронных изданий учебного назначения и предоставить авторизованный доступ к электронным ресурсам для пользователей системы открытого образования. В фондах библиотеки предполагалось хранить в электронном виде до 120 тысяч единиц учебно-справочной литературы, в основном рекомендованную Министерством образования и науки РФ. Количество оцифрованных учебников к началу 2005 г. достигло 12 тыс. Однако нерешенность организационных, правовых и финансовых вопросов привела в начале 2005 г. к приостановке деятельности данного портала и в настоящее время его судьба не определена.

Среди важных проектов в сфере образовательных электронных библиотек следует упомянуть электронную библиотеку, созданную ГНИИ ИТТ «Информика» в рамках информационной системы «Единое окно доступа к ресурсам образовательных порталов» (<http://window.edu.ru>). В библиотеке к концу 2006 г. было представлено около 12 тыс. документов (книг, учебных и методических пособий и др.). Все ресурсы снабжены метаописаниями, выполненными по стандарту LOM. Имеется развитая навигация с использованием Рубрикатора ГРНТИ и специальных классификаций сферы образования.

Лидерами в области создания электронных библиотек являются ведущие университеты страны. Около десятка электронных библиотек ведется в МГУ. В их числе электронная библиотека по химии (<http://www.chem.msu.ru/rus/elibrary/>), физике (<http://nuclphys.sinp.msu.ru/>), математике (<http://lib.mexmat.ru/>), истории (www.hist.msu.ru/ER/index.html) и другим научным направлениям.

Большую коллекцию (св. 3 тыс. электронных публикаций) представляет Библиотека учебной и научной литературы Российского гуманитарного Интернет-университета (Москва; <http://www.i-u.ru/biblio/default.aspx>), к которой ежедневно обращается св. 10 тыс. посетителей.

Все большую роль стали играть электронные библиотеки и полнотекстовые системы в среднем образовании, ориентированные, прежде всего на школьников, студентов лицеев и колледжей. В качестве примеров можно привести информационно-образовательный портал Тюменского государственного колледжа профессионально-педагогических технологий (http://www.tgc.ru/news/news_coll/), Библиотеку сайта «Российские кадеты» (<http://www.cadet.ru/index.php?page=library>), Электронную библиотеку «Книги ФМШ» школы им. А.Н. Колмогорова (<http://www.mathbook.ru/>). Большой интерес представляет электронная библиотека Московского центра непрерывного математического образования (<http://www.mccme.ru/>), где в настоящее время собрано около тысячи книг по «школьной математике», журнал «Квант», «Математическое просвещение» и другие, подчас трудно доступные для широкой публики, издания.

Общие оценки электронных библиотек России. Любые оценки, связанные с электронными библиотеками, в настоящее время будут весьма приблизительными. Причин этому, по крайней мере, три: высокая динамичность данной области деятельности, отсутствие четкого определения основных объектов, отсутствие до недавнего времени скоординированного сбора необходимых данных. Тем не менее, уже сейчас можно представить некоторые обобщенные оценки, основанные на результатах анализа отечественных электронных библиотек, полученных К.В. Вигурским и его коллегами (данные 2005 г)¹.

- Общее число русскоязычных ЭБ – около 1500.
- Не менее 5% ведется за рубежом, в том числе в Украине – 19, США – 8, Беларуси – 7, Израиле – 5.
- Средний срок жизни электронных библиотек 2–4 года.
- 25% ресурсов не обновляются более одного года.
- Не менее половины ресурсов содержит информацию, которую невозможно идентифицировать.
- Не менее 25% ресурсов содержат практически полностью заимствованную информацию.
- Около половины ЭБ создаются и ведутся «любителями».
- В большинстве ЭБ либо отсутствует вовсе, либо представлена некорректно идентификация источника электронной публикации.
- В значительной части ЭБ качество представления информации, а также качество состава и структуры информационного фонда находятся на низком уровне.

Следует отметить, что с 2006 г. работы по учету и анализу отечественных электронных библиотек, наконец, приобрели должную координацию. Эти работы ведутся под эгидой Российской ассоциации электронных библиотек (НП ЭЛБИ).

Разработка концепции электронной библиотеки. Создание электронных библиотек влечет за собой необходимость разрешения множества традиционных и вновь возникающих вопросов, обусловленных электронным документным пространством. Как справедливо замечают специалисты, в настоящее время работа с электронными изданиями ведется в условиях значительной – терминологической, типологической, правовой, технологической неопределенности, а сама идея электронной библиотеки базируется на еще неустоявшейся концепции, подлежащей систематическому пересмотру и обновлению.

¹ Вигурский К.В. Российские электронные библиотеки: состояние и проблемы // Материалы конференции «Информационные технологии в образовании». М., 2005. <http://ito.su/main.php?pid=26&fid=5966&PHPSESSID=4422817159f815c839644b5d8166f414>.

Проблемы формирования электронной библиотеки можно разделить на четыре группы:

- содержательные, связанные с определением состава фонда, контингентом пользователей, формулированием целей создания электронной библиотеки, ее функциональных возможностей;
- организационные, в том числе поиск партнеров и источников финансирования, вопросы управления ЭБ, особенно для распределенных систем;
- технологические, включая описание электронных документов, их учет, организацию их долговременного сохранения и т.д.;
- технические (например, методика перевода традиционных изданий или издательских оригинал-макетов в форматы электронной библиотеки; разработка средств контроля доступа и т.д.).

Эти аспекты отражаются в концепции электронной библиотеки в виде следующих разделов.

- Введения, в котором обосновывается необходимость создания электронной библиотеки.
- Формулирования цели и задач, которые будут решены посредством ее создания.
- Описания потенциальных пользователей, их особенностей, структуры информационных потребностей.
- Описания архитектуры электронной библиотеки.
- Формулирования функциональных требований.
- Определения принципов и источников комплектования фонда.
- Организации поискового аппарата.
- Определения статуса в структуре организации.
- Принципов взаимодействия с другими структурами и организациями в процессах создания и функционирования электронной библиотеки.
- Экономической модели функционирования электронной библиотеки.

Вне зависимости от того, является ли электронная библиотека локальной, или она выставлена в Интернет (с различными условиями доступа), ее создание должно быть направлено на достижение **основной цели**, которая видится в удовлетворении определенного сегмента информационных потребностей методами, наиболее отвечающими специфике этих информационных потребностей, на основе организованного массива электронных документов, сформированного по определенным критериям.

Посредством электронной библиотеки могут быть решены такие задачи, как:

- обеспечение более широкой доступности документов, предоставление которых читателям затруднено или ограничено (редких книг, фотоальбомов, рукописных книг, диссертаций, архивов и т. п.);
- организация фондов документов/изданий, существующих исключительно в электронной форме, их каталогизация и предоставление доступа;
- предоставление пользователям качественно новых возможностей работы с большими объемами электронных данных.

Среди основных функций электронной библиотеки могут быть названы:

- *информационная*, направленная на удовлетворение потребности в информации различных категорий пользователей по всем отраслям знаний, либо одной из предметных областей;

- *просветительская*, реализуемая, в том числе, за счет популяризации книг, манускриптов и других документов, относящихся к отечественной истории и культуре;
- *научно–исследовательская*, ориентированная на содействие глубокому изучению темы (предмета) научными работниками и специалистами высокого уровня подготовленности, в том числе за счет предоставления диссертаций и возможности качественно нового изучения рукописных материалов;
- *образовательная*, в рамках которой осуществляется поддержка как формального, так и неформального образования (путем предоставления не только учебного материала, но и необходимой дополнительной литературы в виде авторитетных монографий);
- *справочная*, позволяющая получать достоверные сведения, отраженные в документах определенного вида.

Для электронных библиотек характерным является важная роль интерфейса, отражающего и организующего основные функциональные возможности библиотеки: учет и регистрация читателей, комплектование, навигация и поиск, библиотечная статистика и др. Многие электронные библиотеки служат для привлечения пользователей к другим функциям портала, например, к электронной торговле, что также влияет на интерфейс электронной библиотеки. Электронные библиотеки также служат рекламной площадкой, что необходимо для обеспечения их финансирования. В таких случаях баннерная или контекстная реклама также является необходимым компонентом интерфейса.

В то время как традиционные библиотеки придерживаются принципа открытого и бесплатного доступа к фондам электронных публикаций, большинство создателей электронных библиотек используют многовариантные подходы – свободный доступ, доступ по регистрации, платный доступ.

При формировании электронных библиотек нельзя не принимать во внимание ее читательское назначение. Интернет–аудитория в целом характеризуется высоким образовательным уровнем и молодым возрастом. Быстро растет количество детей, имеющих необходимые навыки работы в компьютерной среде. В настоящее время социологические исследования не позволяют выявить однородные читательские группы с устойчивыми и в достаточной степени локализованными информационными потребностями. Таким образом, формирование электронных библиотек может опираться либо на известные их создателям целевые группы (студенты, специалисты в конкретных отраслях знания и т.д.), либо на экстраполяцию профессиональных представлений о функциях библиотеки в электронную среду.

Функциональные требования к электронной библиотеке должны ответить на ряд важных вопросов:

1. Включает ли библиотека разнородные или однородные электронные документы;
2. Будут ли созданы единые средства поиска по разнородным коллекциям;
3. Какие формы библиотечного и информационного обслуживания пользователей будут предоставлены в электронной библиотеке;
4. Как будет обеспечена сохранность фондов электронной библиотеки;
5. В зависимости от того, какие методы комплектования предполагается использовать, определяются важнейшие особенности электронной библиотеки. Например, если электронная библиотека будет формироваться путем оцифровки, то важно при-

нять решение об использовании технологии распознавания и, возможно, корректуры распознанных текстов.

Важную роль в концепции должны занимать правовые вопросы. Если предполагается перевод в цифровую форму изданий, охраняемых авторским правом, то в концепции должны быть определены принципы взаимоотношений со всеми правообладателями.

В случае если электронная библиотека создается несколькими организациями, в концепции должны быть прописаны механизмы их взаимодействия и основные функции, выполняемые в процессе работы.

Принципы, источники и организация комплектования фонда электронной библиотеки в концепции указываются в общем виде, поскольку более детально эти вопросы раскрываются в Положении об электронной библиотеке и Профиле ее комплектования. Аналогично определяются общие экономические принципы функционирования (составляющие основу Бизнес-плана).

Общие принципы проектирования электронной библиотеки. Выше было указано, что основной задачей электронной библиотеки является удовлетворение информационных потребностей определенной группы пользователей. Естественно, что для разных групп пользователей эти потребности не совпадают. Речь идет не только о тематическом и видовом содержании фонда, но также о способах представления информации в электронной библиотеке, ее достоверности, актуальности, полноте и других качественных характеристиках.

Также не совпадают потребности пользователей в функциональных возможностях электронной библиотеки: чем более широким является круг пользователей, тем проще должна быть функциональность системы, ее интерфейс. Для квалифицированных пользователей необходимо предлагать значительно большие возможности, чем для массового пользователя.

На принципы проектирования электронной библиотеки оказывает существенное влияние характер информационной деятельности организации, в которой создается электронная библиотека. Библиотеки, архивы, музеи, учебные заведения, научные учреждения, СМИ, издательства, электронные магазины имеют свои, сильно различающиеся требования к организации информационных процессов, которым должна удовлетворять проектируемая электронная библиотека. Эти требования, естественно минимизируются, если создается любительская электронная библиотека или электронная библиотека в рамках неформального сообщества. Самостоятельным классом электронных библиотек с этой точки зрения следует признать коммерческие электронные библиотеки, главным требованием к которым является рентабельность.

В общем, при проектировании электронной библиотеки следует рассматривать два класса требований, которые можно назвать пользовательскими и общесистемными.

Пользовательские требования определяют содержание фонда, его структуру, систему метаданных и функциональные возможности электронной библиотеки.

Общесистемные требования определяют общую структуру электронной библиотеки, технологию функционирования электронной библиотеки в рамках действующей организации с учетом ее задач и специфики, взаимодействия с другими организациями, порядок ее использования и администрирования.

Например, для библиотек и архивов дополнительной функцией электронной библиотеки может служить ее использование в качестве страхового фонда для обеспечения сохранности документов в иной форме хранения, чем традиционные документы.

И, конечно, для проектирования электронной библиотеки решающими условиями являются финансовые возможности и наличие квалифицированных разработчиков.

Требования к электронной библиотеке, которые необходимо учесть при ее концептуальном проектировании, будут рассмотрены с двух точек зрения – пользовательской и административной. Указанные требования, которые могут быть реализованы только в больших специализированных электронных библиотеках, далее перечисляются в наиболее полной форме.

Общие требования к пользовательским функциям и технологиям электронных библиотек. На концепцию электронной библиотеки, а также на ее программное обеспечение существенное влияние оказывают требования, связанные с представлением и возможностью использования информации.

Все информационное пространство электронной библиотеки, доступное пользователю, должно быть представлено в виде совокупности самостоятельных объектов. В качестве таковых во многих случаях выступают электронные документы. Электронные объекты могут представлять собой текстовые произведения, изображения, фонограммы, базы данных или их фрагменты, словарные статьи, подписи под рисунками, отдельные имена и т.д. Организация информационного пространства как совокупности объектов и однозначная идентификация последних необходимы для обеспечения эффективной навигации и выполнения некоторых видов информационных поисков.

Инструментом описания и идентификации выступают метаданные, в том числе, библиографические записи, поскольку основную часть фонда будут составлять обычные документы.

Электронные библиотеки должны поддерживать систему связей между объектами и между элементами этих объектов, в частности систему гипертекстовых и гипермедийных связей, которые могут быть различных типов. Они должны иметь достаточно мощный механизм поиска, позволяющий работать с лексикой естественных языков с учетом их грамматики – как минимум словоизменения.

Электронные библиотеки должны отображать пользователю информацию в виде максимально соответствующем исходному материалу.

Кроме того, должна быть предусмотрена возможность конструирования в системах сложных информационных объектов со своими характеристиками и идентификаторами. При этом новый, сложный объект обладает структурой, включающей в себя структуры входящих в него компонентов.

К основным функциональным возможностям, которые должны обеспечивать электронные библиотеки, следует отнести:

- Навигацию во всем доступном информационном пространстве – наглядное предоставление пользователю логической структуры информационного пространства, например дерева, организующей все информационные объекты системы, и средств работы с ней (частичное или полное раскрытие структуры, сокращение ее до заданных элементов и т.п.). Терминальными узлами структуры являются информационные объекты, которые пользователь может вызывать на экран непосредственно из этой структуры при работе с ней.

- Лексический поиск – информационный поиск по свободной лексике русского языка (как современного, так и дореформенного) и языков, использующих латинский алфавит. При этом поиск по лексике русского языка должен проводиться с учетом его

грамматических особенностей (изменения по падежам, наличие беглых гласных и т.п.). При составлении поисковых предписаний должно допускаться применение логических операторов (И, ИЛИ, И-НЕ) и контекстных операторов, учитывающих порядок следования операндов и расстояние между ними. Результат лексического поиска должен представлять собой совокупность информационных объектов, содержащих лексику поискового предписания (с точностью до критерия соответствия), которая в текстах должна быть выделена (например, подсвечена).

- Символьный поиск – разновидность лексического поиска, где в качестве лексической единицы выступает определенная последовательность допустимых символов, например, даты.

- Атрибутный поиск – информационный поиск объектов по значениям их характеристик. Для текстовых объектов к таким характеристикам могут относиться «автор», «название», «место издания», «дата издания» и т.п. Атрибутный поиск должен допускать использование логических операторов (И, ИЛИ, И-НЕ), а также возможность проведения поиска по значениям двух и более характеристик одновременно. Результат атрибутного поиска – совокупность информационных объектов, значения характеристик которых удовлетворяют условиям поискового предписания. Результаты поиска должны допускать различные сортировки.

- Просмотр содержания информационного объекта и его структуры: последовательный (например, страница за страницей) и выборочный (переход на любую заданную страницу или на любой элемент, отраженный в структуре). При этом структура и текст произведения должны быть синхронизированы – любое изменение положения в структуре вызывает соответствующее изменение положения в тексте и наоборот.

- Многооконный режим работы, позволяющий пользователю одновременно видеть на экране два окна (как минимум) с разным содержанием.

- Одновременный вывод на экран (в разных окнах) текстов двух разных произведений или двух экземпляров одного произведения.

- Манипулирование со структурой информационного объекта аналогичное работе со структурой всего информационного пространства.

- Поддержка аппарата гипертекстовых и гипермедийных связей, обеспечивающего пользователю оперативный переход от объекта или некоторого его элемента к другому взаимосвязанному с ним объекту или его элементу. Допускается существование связей типа «один-к-одному» и «один-ко-многим».

- Протоколирование сеанса работы пользователя с возможностью перехода в любое из ранее существовавших состояний системы.

- Выполнение пользователем настройки системы (задание определенной конфигурации окон, размеров шрифтов и др.).

- Установка закладок в тексте информационных объектов и возможность оперативного перехода на них.

- Экспорт информации из системы с указанием на источник (собственно ЭИ или ЭБ и информационный объект, из которого копируется информация).

- Списки должны допускать сортировку по значениям любого поля, определенного как ключевое.

Общесистемные требования к ЭБ. Обычно электронная библиотека формируется как единая база данных, состоящая из электронного каталога электронной библиотеки и фонда электронной библиотеки.

Необходимо определить, как будет организован электронный документ, состоящий из нескольких составляющих с различными типами данных, как будут связаны с другими материалами связями типа часть/целое или «последовательность». Электронный документ может состоять из страниц, глав, индексов, иллюстраций и других элементов, может физически сохраняться в виде нескольких файлов, содержащих текст и графику со связями между ними. При этом любой отдельный документ, состоящий из нескольких файлов, может быть «собран» в один файл и снабжен титульным листом. На каждый многостраничный документ, включаемый в электронную библиотеку, формируется новый титульный лист.

Одним из центральных вопросов проектирования является организация метаданных, включая информационно-поисковые языки атрибутного, классификационного и вербального типов. Выбор метаданных во многом определяет и функциональность электронной библиотеки и стоимость работ по ее формированию и поддержанию.

Выбор формата хранения документов. Для представления документов в ЭБ могут использоваться разные форматы, в том числе:

- формат PDF
- формат DeJa Vue;
- форматы DOC, TXT;
- форматы для изображений TIFF, JPEG;
- аудиоформаты, например, mp3;
- гипертекстовый язык разметки HTML;
- расширенный язык разметки текста XML; при этом для всех типов документов, включаемых в электронную библиотеку, определяются стандартные или специфические определители типа документа (DTD) в качестве грамматик, описывающих комплекс меток XML и их взаимосвязей, или схемы описания ресурса (RDF-схемы).

Могут быть также предусмотрены особые форматы для представления геоданных, трехмерных или анимационных объектов экспериментальных данных и других особых видов информации.

Выбор одного или нескольких форматов для хранения определяется в рамках концепции электронной библиотеки с учетом пользовательских и общесистемных требований.

Общесистемные требования к доступу к электронной библиотеке. Доступ к базе данных электронной библиотеки для пользователей Интернет должен быть ограничен документами, не охраняемыми авторским правом, включая произведения на которые истек срок охраны, а для охраняемых – теми, для которых имеется разрешение правообладателя на доведение произведения до всеобщего сведения. При размещении в электронной библиотеке аудио и видео документов следует иметь разрешение не только обладателей авторских прав, но и обладателей смежных прав (например, исполнителей) Для каждого документа и/или его составляющей определяются права доступа вплоть до запрещения предоставления его в Сети, в том числе в течение определенного периода времени.

Разработка положения о фонде электронной библиотеки и профиля комплектования. Формирование фонда электронной библиотеки должно опираться на два базовых документа, которыми являются «Положение о фонде электронной библиотеки» и «Профиль комплектования фондов электронной библиотеки».

В «Положении о фонде электронной библиотеки» отражаются базовые принципы формирования и организации фонда. Раздел «Общие положения» включает ссылки на документы, в соответствии с которыми создается электронная библиотека (законода-

тельные и другие нормативные акты, концепция электронной библиотеки, программа развития организации, государственный контракт, грант и т.д.). Далее необходимо указать статус фонда электронной библиотеки – предназначен ли он для долговременного хранения или является временной коллекцией; каковы условия доступа; является ли фонд электронной библиотеки самостоятельной единицей или входит в состав более общей системы фондов (например, фондов классической библиотеки в целом).

Одним из основных разделов «Положения» является формулирование принципов комплектования фонда, включающие вид комплектуемых документов (текстовые, графические и т.д.), их тематику (универсальная, тематическая, проблемно-ориентированная и т.д.), принятые ограничения (по хронологии, языку, назначению и другим признакам). При этом объектами комплектования не могут являться ресурсы, содержащие вредоносную информацию, определяемую в соответствии с действующим законодательством (содержащие порнографию, призывы к насилию, разжигающие расовую и национальную рознь и т.д.). В этом же разделе «Положения» определяется географическая и юридическая принадлежность комплектуемых источников, целевое читательское назначение (для определенного или неограниченного круга пользователей) и другие принципы, задаваемые критериями отбора электронных документов.

«Положение» определяет также правовые условия комплектования фонда электронной библиотеки. По правовому статусу или режиму использования объекты комплектования электронной библиотеки подразделяются на следующие категории:

– Ресурсы, на которые не распространяется российское законодательство об интеллектуальной собственности, например, официальные документы, включая правовые акты, стандарты, фольклорные произведения и др.

– Произведения, исключительные права на которые утрачены по сроку давности.

– Ресурсы, объявленные создателями и правообладателями общественным достоянием, не имеющим ограничений на распространение.

– Ресурсы, имущественные права на которые принадлежат государству.

– Ресурсы, имущественные права на которые принадлежат юридическим или физическим лицам.

– Ресурсы, охрана прав на которые передана специализированным коммерческим или общественным организациям, таким как РОМС (Российское общество по мультимедиа и цифровым сетям).

Отдельным разделом указываются источники (методы и технологии) комплектования, которые подробно изложены ниже.

Раздел «Положения», раскрывающий структуру фонда электронной библиотеки, должен содержать описание его основных частей:

– пользовательского и страхового фондов;

– открытой и закрытой частей (по уровням доступа);

– место расположения фонда электронной библиотеки (сервер, условия зеркалирования);

– используемые носители (сетевые ресурсы, документы на переносимых носителях).

В «Положении о фонде» прописываются основные направления и принципы взаимодействия с партнерскими организациями.

В разделе «Управление фондом электронной библиотеки» необходимо указать орган управления (или должностное лицо), устанавливаются его функции, ответственность, права и обязанности.

«Профиль комплектования фонда ЭБ». Данный документ определяет важнейшие видовые и содержательные параметры фонда электронной библиотеки, регулирует поступление в фонд электронной библиотеки документов, вне зависимости от источника их получения. Если речь идет о формировании фонда электронных документов в рамках классической библиотеки, то положения «Профиля комплектования фонда электронной библиотеки» приводятся в соответствие с общей политикой формирования фондов.

Электронные представления печатных изданий и материалов. Оригинальные электронные документы. В данном разделе рассматриваются электронные документы, имеющие в своей основе печатные издания или являющиеся полными аналогами таковых. В основном имеются в виду электронные документы, полученные путем оцифровки, а также оригинал-макеты печатных изданий. В том числе:

- Официальные;
- Научные материалы;
- Издания;
- Непубликуемые материалы (включая отчеты о НИР и НИОКР);
- Учебные и образовательные издания;
- Информационно-библиографические издания, каталоги;
- Энциклопедии, словари;
- Справочники;
- Популярные издания;
- Средства массовой информации;
- Путеводители;
- Картографические издания;
- Изобразительные издания;
- Фонодокументы;
- Кинофильмы;
- Патенты;
- Стандарты, нормативно-техническая документация постоянного действия;
- Рекламные издания, прейскуранты. Презентации;
- Литературно-художественные издания;
- Производственно-практические издания;
- Нотные издания;
- Периодические и продолжающиеся издания.

С точки зрения используемых форматов, наиболее перспективным представляется использование открытых стандартов, что обеспечивает независимость электронной библиотеки от фирм-поставщиков программного обеспечения. С другой стороны, использование распространенных фирменных стандартов, таких как pdf или doc, пока эффективнее с экономической точки зрения. Кроме того, для современных электронных библиотек критичным является различие символьных и графических форматов, что определяет структуру коллекций по функциональным возможностям разделов.

Отдельным вопросом является вопрос комплектования электронных библиотек информационными ресурсами, являющимися базами данных, поскольку они в общем случае не являются ни документами, ни массивами документов. Они подразделяются на:

- Полнотекстовые;
- Библиографические, реферативные;
- Каталоги ссылок;

- Адресно-справочные ;
- Базы данных показателей (фактографические, статистические);
- Мультимедийные ;
- Базы данных изображений:
- общие;
- массивы геоданных;
- проектно-конструкторские;
- Звуковые базы данных;
- комбинированные.

Базы данных должны быть проанализированы с точки зрения их содержания и функционального назначения. Отбираются те, которые соответствуют концепции электронной библиотеки, ее целевому читательскому назначению.

В программно-технологическом аспекте базы данных подлежат комплектованию, если их техническое решение позволяет осуществить перенос на сервер электронной библиотеки без существенных потерь.

Паспорт электронного фонда. Одним из механизмов управления процессами формирования фондов электронных документов может выступить Паспорт фонда, создаваемый уже на начальном этапе фондообразования. Мы предлагаем примерную схему, которая может лечь в основу паспортизации как электронного фонда в целом, так и составляющих его отдельных коллекций.

Таблица. Паспорт фонда электронных публикаций без индивидуального материального носителя

Название проекта/коллекции	<i>Заполняется даже, если речь идет о сканировании отдельных изданий</i>
Цель создания коллекции	<i>Снятие пользовательской нагрузки на оригинал, расширение ресурсной базы библиотеки и т.д.</i>
Состав коллекции	<i>Определяется как по содержанию, так и по видам документов</i>
Исполнители	<i>Учитываются все подразделения библиотеки, задействованные в работе и возможные внешние соисполнители; выделяются руководитель и главные ответственные</i>
Источник выявления документов	<i>Как правило, каталог библиотеки и библиографические указатели</i>
Форма обнародования полученных электронных документов	<i>Интернет, локальная сеть, CD-издание</i>
Условия доступа	<i>Платный (разовая плата, подписка и т.д.), бесплатный (свободный, по авторизации и др.)</i>
Предполагаемый объем	<i>Как по количеству документов, которые будут представлены в коллекции хотя бы на этапе перевода ее в пользовательский режим, так и по величине занимаемого дискового пространства</i>

Наличие программного и лингвистического обеспечения	<i>Графа заполняется специалистами в данной области</i>
Разработанность технологии оцифровки	<i>Технология оцифровки зависит от типов документов – книги, газеты, карты, фотографии и т.д.</i>
Оценка печатных изданий специалистами по сохранности фондов	<i>Издания, выбранные для сканирования должны быть максимально застрахованы от какого-либо ущерба.</i>
Оборудование	<i>Оценивается как имеющееся, так и необходимое для полномасштабной работы</i>
Финансирование	<i>Указывается имеющийся и необходимый объем, а так же источники и ритмичность поступления денежных средств</i>
Степень решенности правовых проблем	<i>Либо оцифровке подлежат издания, не находящиеся в зоне действия Закона РФ «Об авторском праве и смежных правах», либо имеются договора с правообладателями</i>
Стоимость проекта	<i>Включается цена оборудования, оцифровки, создание пользовательской копии нужного качества, оплата труда сотрудников, отчисления по авторскому праву и т.д.</i>

1.2. Методы комплектования электронных библиотек

Источники информации о российских ЭБ. При разработке модели комплектования ЭБ, естественно, нужно учитывать наличие уже существующих и доступных пользователям Интернета. В Интернете представлено достаточно много каталогов российских электронных библиотек. Специализированный каталог полнотекстовых коллекций ведется Российской ассоциацией электронных библиотек на сайте <http://reestr.elibra.ru> Другими источниками могут служить Государственный регистр баз и банков данных в НТЦ «Информрегистр» (<http://www.infoereg.ru/db/default.asp>), Навигационная система по информационным ресурсам науки и инноваций (<http://scireg.informika.ru>), а также Реестр проектов в области электронных библиотек (<http://www.elbib.ru>). Достаточно полный список сетевых полнотекстовых библиотек, содержащих литературу для образовательных целей, представлен в каталоге федерального портала «Российское образование» (<http://www.edu.ru>).

Некоторые сетевые адреса популярных электронных библиотек можно получить в Регистре полнотекстовых и справочных ресурсов Интернет (http://dc.rsl.ru/dc_bib.htm, разработчик Российская Государственная библиотека), а также из каталогов «Лучшие электронные библиотеки» сайта «Русский Журнал» (<http://www.russ.ru>) и Electronic Libraries Научной библиотеки МГУ (<http://www.lib.msu.ru:8080/authors-01a.htm>), Электронные каталоги российских библиотек – проект В.К. Степанова (<http://www.openweb.ru/>), каталог ссылок на сайте Фундаментальной библиотеки «Русская литература и фольклор» (<http://www.feb-web.ru>). Россия в зеркале www (<http://russia-in-www.narod.ru>). Большой список адресов бесплатных ЭБ преимущественно художественной литературы пред-

ставлен в разделе «Ссылки на электронные библиотеки» сайта «Zmiy – сканирование литературы» (<http://zmiy.by.ru>).

Всего по экспертным оценкам в Рунете функционирует не менее 2 тыс. электронных библиотек.

Технологии комплектования ЭБ. По отношению к электронным библиотекам методы, способы и технологии комплектования их фондов зачастую совпадают. Поэтому далее в тексте эти термины не различаются.

Комплектование электронных библиотек может осуществляться несколькими способами:

- Оцифровка документов, имеющих в распоряжении создателя;
- Получение электронных версий документа от автора или издателя;
- Заимствование документов, имеющих в свободном доступе в Интернет;
- Организация обмена с другими электронными библиотеками.
- Закупка законно распространяемых электронных изданий на переносимых носителях;
- Организация доступа к удаленным документам.

Электронная библиотека может выбрать один из названных способов, однако на практике происходит сочетание большинства из них. Каждый из методов имеет свои достоинства и недостатки, критичность которых можно оценить только с учетом специфики каждой конкретной электронной библиотеки.

Оцифровывание полиграфических изданий, наряду с комплектованием документами в электронном виде, является одной из основных технологий формирования электронной библиотеки. Гораздо менее развиты организация обмена между библиотеками. Организация доступа к удаленным ресурсам относится к комплектованию с достаточной степенью условности, поскольку с точки зрения традиционного библиотечного дела, комплектованием является процесс получения изданий непосредственно в фонд. Вопрос о том, насколько доступ к удаленным ресурсам (например, в рамках консорциума) является комплектованием, является спорным, и его решение еще не найдено.

Следует отметить, что оцифровка редких, особо ценных и ветхих книг должна проводиться на аппаратуре высокого класса в щадящем режиме. Причем в крупных библиотеках таких книг очень много, поэтому они стремятся организовать центры сканирования на своих территориях. Вместе с тем, существуют проекты (и наблюдается тенденция их роста), для реализации которых требуются специализированные организации, оказывающие услуги по оцифровке книг (конверсии в машиночитаемую форму). Таких организаций на отечественном рынке услуг по оцифровке немного, однако, среди них есть коллективы, обладающие опытом, технологиями и средствами для выполнения подобных работ. Ниже представлен список организаций, предлагающих услуги по формированию электронных копий книг:

Корпорация «Электронный Архив» (<http://www.elar.ru>) для широкого круга предприятий и организаций выполняет работы по массовому (поточному) сканированию документов и созданию электронных информационных ресурсов. Производственные мощности, уникальные технологии, конвейерная организация работ и штат высококвалифицированных операторов обеспечивают перевод в электронный вид нескольких миллионов документов в месяц. Корпорация уже обработала и передала заказчикам более 40 млн. электронных документов. Перечень проектов представлен на странице <http://www.elar.ru/projekt/>. Базовые расценки: 1,20 – 1,80 руб./стр. при поточном ска-

нировании формата А4; 3,00 – 4,00 руб./стр. при «книжном» сканировании формата А3. Стоимость сканирования зависит от качества сканируемого материала (бумага, калька, ветхость и проч.), качества и режимов сканирования (черно-белое, в градациях серого, цветное), типа используемого сканера (поточных для расшитых документов, книжных для нерасшитых документов, протяжных для чертежей и широкоформатных иллюстраций), количества страниц (от 200 тыс.).

ООО «ДиректМедиа Пабблишинг» (<http://www.directmedia.ru>) – дочернее предприятие берлинского электронного издательства «Directmedia Publishing GmbH» (<http://www.directmedia.de/>). Основные направления деятельности «Directmedia Publishing GmbH» – издание художественной и научной литературы, справочников, словарей и иллюстрированных изданий по искусству на CD-ROM. С 1997 г. Издательством выпущено более 150 названий CD-ROM общим объемом более 2,5 млн. страниц. «Directmedia Publishing GmbH» является одним из первых электронных издательств, ориентированных на высокие издательские стандарты электронной литературы. Базовая стоимость оцифровки 1 стр. текста формата А4 (без оптовых скидок) – 0,1 у.е. Гарантируемое качество работ может достигать до 1 ошибки на 20 стр.

Главный Информационно-Вычислительный Центр (ГИВЦ) Министерства Культуры РФ (<http://www.givc.ru/uslugi.html>) оказывает услуги по оцифровке книг.

Издательский дом «Равновесие» (<http://www.ravnovesie.com/>) с 1996 года занимается изданием электронных книг и предлагает услуги по оцифровке книжных изданий. Приоритетными направлениями для электронного издания являются экономика, бизнес, менеджмент, психология, педагогика, юридическая литература. Текущая базовая расценка по сканированию и распознаванию текста формата А3 – 1,5\$ за лист.

ЗАО «ПроСофт-М» (<http://www.prosoft-m.ru/>) – специализированный центр профессиональных решений в области массового сканирования и обработки бумажных документов, архивации информации, создания и оснащения корпоративных электронных архивов, центров обработки данных. ЗАО «ПроСофт-М» входит в состав корпорации «Электронный Архив». ЗАО «ПроСофт-М» предлагает для библиотек и музеев:

- Услуги по созданию высококачественных цифровых коллекций;
- Электронные книжные коллекции;
- Ретроспективная конверсия библиотечных каталогов;
- Ретроспективная конверсия музейных описей и инвентарных книг;
- Создание электронного иллюстрированного каталога музея.

Корпорация «Индустрия Интеллекта» (<http://www.atlas-center.org/atlas/service1.html>) оказывает услуги по переводу в цифровой формат библиотечных документов. Стоимость оцифровки одной страницы А4 формата составляет 0,14 центов. В услугу оцифровки входит: сканирование, распознавание, редактирование, корректировка и правка распознанного текста, разметка страниц и электронная доставка.

Фирма «Научное и научно-библиографическое обслуживание «Виртуальная библиотека аспиранта»» (<http://ukrdiser.com/>) оказывает услуги по оцифровке библиотечных источников, переводу их в электронную форму и доставке заказчику. «Виртуальная библиотека аспиранта» является украинским предприятием, которое работает и с российскими заказчиками. Стоимость оцифровки книги за 1 стр. – 0,36\$. Срок доставки – от 2 до 30 рабочих дней.

Фирма «СКАНМОС» (<http://www.scanmos.ru>). Основная услуга – сканирование текстов, документов, книг с доставкой по Москве. Сканирование 1 страницы формата А4 стоит 5-6 руб. Существует система скидок.

Заимствование документов, имеющих в свободном доступе в Интернет.

По данным исследователей компании Яндекс по состоянию на январь 2006 г. в российском Интернете имеется примерно 2,5 млрд уникальных документов, расположенных на почти 2 млн. серверах. Общее количество индексируемой информации (т.е. в основном текстовой) составляет свыше 17 терабайт².

Сюда не входит информация, находящаяся в так называемом «глубоком» Интернете, т.е. размещенная в базах данных, поиск в которых возможен по специальному запросу. Доступ в такие базы данных возможен платный или бесплатный, с регистрацией пользователя или без нее. Таких баз данных в Российском Интернете по экспертным оценкам имеется не менее 7 тыс.

Существуют и другие оценки. Поисковая система eboogle, специализирующаяся на поиске полных текстов документов (которые не совсем корректно называются в этой системе «книгами») в июне 2006 г. показывала общее число «книг» в Рунете около 650 тыс. (<http://www.ebdb.ru/list.aspx>)

Только в библиотеках, входящих в Ассоциацию библиотечных консорциумов АРБИКОН, имеется около 500 тыс. электронных версий полных текстов книг, журнальных статей и других документов.

Таким образом, в Интернет имеются значительные ресурсы, которые могут быть и реально являются источником комплектования электронных библиотек. Поэтому технология комплектования, основанная на заимствовании документов из Интернет, обычно рассматривается как наиболее очевидная и легко реализуемая. В то же время она имеет наряду с преимуществами и определенные недостатки, и трудности, которые рассматриваются ниже.

Правовые вопросы. Далеко не всегда документы, размещаемые в Интернет, снабжены юридически корректными указаниями на правообладателя, условия размещения, возможность копирования и использования в коммерческом или некоммерческом режиме. Поэтому если создатель ЭБ стремится обеспечить легитимность своей коллекции, ему необходимо решать все связанные с этим проблемы самостоятельно. Подробнее данные проблемы рассмотрены в главе 9 настоящего пособия.

Проблема качества. Значительная часть размещаемых в Интернет ресурсов представляется неудовлетворительной с точки зрения качества представления документа. Нередки случаи, когда книга выставляется без титульного листа, библиографического описания, даже издательского, когда неизвестно, какое издание произведения использовалось. Часты случаи появления электронных документов с пропуском отдельных фрагментов, с ошибками, возникающими при сканировании и т.д. Иногда создатель ресурса сознательно удаляет несущественные с его точки зрения фрагменты. Например, известно электронное издание Полного собрания сочинений А.С. Пушкина, из которого исключена переписка. Поэтому если создатель ЭБ предъявляет определенные требования к качеству включаемых в коллекцию документов, у него имеется альтернатива: отказаться от заимствования документа или производить достаточно трудоемкую работу по сверке документа.

Поиск и отбор. Документы в российском Интернете рассеяны, как было указано, по 2 млн. адресов. Существующие поисковые системы, а также каталоги ссылок не

² Сегалович И.В., Зеленков Ю.Г, Нагорнов Д.О. Методы сравнительного анализа современных поисковых систем и определения объема Рунета. Труды 8-й Всероссийской научной конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции»-RCDL 2006, Суздаль, 2006

обеспечивают ни полноту, ни точность поиска. Поэтому комплектование ЭБ на основе заимствования из Интернета требует дополнительных затрат на мониторинг Интернета, поиск и отбор ресурсов.

Дублирование. Документы в Интернете во множестве дублируются, причем дубли могут различаться как по формальным признакам (форматы, кодировки, разметка), так и по содержательным – например, различные издания одного произведения. Это заставляет разработчиков поисковых машин создавать различные механизмы для выявления уникальных документов. Однако создатель ЭБ не всегда может доверять этим средствам, особенно когда из множества дублей требуется выбрать документ, наиболее высокий по качеству или просто в наибольшей степени отвечающий требованиям данной ЭБ.

Форматы и стандарты. Документы в Интернете представлены в самых различных форматах, как графических, так и символьных, а также в различных кодировках символов. Преобразование в стандартный для данной ЭБ формат также требует определенных затрат. Особую проблему представляют наиболее ценные ресурсы с развитой функциональностью, которая обеспечивается различными программными приложениями. Таковы практически все интерактивные документы, например, учебники для дистанционного обучения, возможность копирования которых без нарушения функциональности ценных ресурсов может представлять значительные сложности. Это, в частности относится к различным графическим, анимационным, звуковым и другим вставкам в тексты. Основным критерием для отбора цифровых ресурсов в ЭБ из Интернет является возможность отделения данных от их представления, поскольку именно на представление данных оказывает наиболее сильное воздействие изменяющееся программное окружение. Идеальным случаем представляется, когда данные сохраняются на XML – подобном языке.

Гипертекстовые ссылки. При комплектовании документами из Интернета необходимо решить вопрос с гипертекстовыми ссылками, имеющимися в заимствуемых документах. При этом следует различать ссылки внутри документа, ссылки на другие документы, также включаемые в данную коллекцию и ссылки на ресурсы, не включаемые в коллекцию. Первые необходимо сохранять, вторые корректировать, заменяя внешние адреса на внутренние, относительно третьих решение нужно принимать исходя из общей концепции ЭБ.

Каталогизация. Одним из наиболее сложных и затратных процессов при комплектовании ЭБ путем заимствования является каталогизация, включающая заимствование или создание метаданных, классификационных индексов, а также поисковых образов на вербальных языках, если они применяются в заимствованных документах. Подробно данные проблемы рассмотрены в главе 8 настоящего пособия.

Организация доступа к удаленным документам. Альтернативой создания собственной электронной библиотеки для многих образовательных, информационных и библиотечных организаций может стать организация доступа пользователей к удаленным ресурсам. В мировой практике разработано множество вариантов удаленного доступа к электронным публикациям, как для отдельных библиотек, так и для библиотечных консорциумов, например, право доступа по требованию, право временного сохранения на сервере, копирования (на печать, на съемный носитель), право комплектования архивной копии и т.д. Эта форма информационного обслуживания активно развивается и в России.

Сохранение электронных документов. За рубежом данная проблема давно находится в сфере интересов профессионального сообщества. Этой проблеме специально посвящена Хартия ЮНЕСКО «О сохранности цифрового наследия» (ссылка), принятая Генеральной ассамблеей ЮНЕСКО в ноябре 2003 г. Основные положения Хартии базируются на понимании того, что быстрота эволюции цифрового мира нарушает порядок всех устоявшихся методов сохранения информации. Временные рамки для мероприятий по сохранению сужаются: меры по сохранению доступа к цифровым материалам приходится принимать уже на самом раннем этапе их существования, поскольку весьма реальна опасность утраты важнейших материалов, для которых были использованы ценные ресурсы. В этой связи насущное значение имеет признание странами своей ответственности за цифровое наследие и принятие мер по предотвращению их утраты.

Сохранение электронных документов в сравнении с печатными, приобретает новое качество – наряду с физической сохранностью материального носителя, необходимо обеспечить программную поддержку документов и наличие соответствующих аппаратных средств для их адекватного воспроизведения. Обеспечивая сохранность электронных документов, необходимо решать четыре группы задач: защита от вирусов; защита от несанкционированного доступа; предотвращение утраты данных и изменения (искажения) текста при каждом принципиальном обновлении программного обеспечения; сохранение самого материального носителя.

До включения в фонды изданий на машиночитаемых носителях библиотеки, сохраняя материальную основу издания, обеспечивали и возможность прочтения текстов. Однако жизненный цикл электронного издания (как и любого машиночитаемого документа) почти полностью зависит от степени адекватности форм его представления современным техническим возможностям. У электронных документов, подготовленных в оболочке со специфическим представлением содержания, жизненный цикл связан с жизнеспособностью самой оболочки. Если она не развивается и не поддерживается, то электронный документ сложно сохранить или восстановить без дополнительных усилий. Кроме того, возможности его использования также ограничены этой оболочкой.

В области сохранности электронных изданий можно выделить несколько проблемных точек или зон риска.

Недолговечность носителей информации (магнитные ленты, дискеты и отчасти оптические компакт-диски). Самыми ненадежными являются дискеты – при каждом их использовании головка дисководов соприкасается с магнитным слоем, намагниченность постепенно снижается, быстрее всего изнашивается тот участок, на котором размещено «оглавление», т.е. перечень файлов. Есть мнение, что содержание дискет надо копировать не реже, чем раз в 5 лет.

Более устойчивы жесткие диски компьютеров, чей ресурс по расчетам производителей составляет около 250 тысяч часов, т.е. примерно 28 лет. Однако этот оптимизм носит сугубо теоретический характер и проверке не поддается: на практике ни один жесткий диск столько не проработал. Но даже если он и проработает, то весь компьютер устареет гораздо раньше.

Сегодня самыми стабильными считаются оптические носители информации – CD-ROM, DVD, CD-R и т.д. Первоначально реклама утверждала, что они вообще вечны, однако в настоящее время данные колеблются от 2 – 3 лет до 100 – 500 лет, что объясняется отсутствием надежных прогностических методик. Срок гарантированного функционирования перезаписываемых дисков определяется фирмой «Кодак»

в пределах 20 лет при температуре 25 градусов Цельсия и 40% влажности, в темноте [427]. Понятно, что обозначенный срок для библиотек приемлемым не является. Но даже если диски сохраняются 100 лет, то возникает другая проблема – модернизация самих считывающих устройств может привести к невозможности прочтения электронных изданий, созданных для компьютеров предыдущего поколения.

Устаревание аппаратных средств и периферийных устройств. Наглядными примером служит попытка Дж. Лэнье (изобретатель термина «виртуальная реальность») выставить на выставке, посвященной истории мультимедийных средств, свою знаменитую компьютерную игру 1980–х годов «Лунная пыль», поскольку не удалось найти ни работающего компьютера «Коммодор–64» образца 1982–го года, ни нужного джойстика, ни подходящего видео–интерфейса. С такой же проблемой столкнулся К. Гудмен (попечитель Американского музея виртуальных образов), когда ему понадобился один из первых компьютерных компакт–дисков под названием «Как устроена ЭВМ». Диск был найден, а старого «Макинтоша» для работы найти не удалось.

И, наконец, необходимо принимать во внимание *стремительное появление новых машиночитаемых носителей*. Уже никто не пользуется перфокартами и дискетами 5,25 дюйма. Кстати, надо заметить, что данная проблема относится ко всем машиночитаемым документам – достаточно вспомнить валики Эдисона и виниловые пластинки.

Проблема сохранения электронных изданий сопрягается и с такими факторами, как *быстрая смена программного обеспечения, несовместимость некоторых платформ, устаревание форматов* (одновременно с появлением новых) и кодировок и т.д. Не зря среди программистов бытует шутка: цифровые данные хранятся либо 5 лет, либо вечно – в зависимости от того, что наступит раньше.

Еще одним аспектом сохранения интерактивных электронных изданий является решение проблемы стабилизации их содержания, поскольку цифровая форма позволяет вмешиваться в сам текст на всем протяжении жизненного цикла документа. Следовательно, необходимо создать систему сохранения эталонных текстов первоисточников на основе национальных хранилищ цифрового научного и культурного достояния страны.

В настоящее время специалисты апробируют несколько методик сохранения электронных документов:

1. Сохранение электронного документа в виде копии на носителе, доказавшем свою долговечность, например на бумаге или микрофильме. Однако таким образом можно сохранить лишь те электронные документы, которые содержат «плоский» текст, не критично связаны с другими объектами и не имеют специальной программной оболочки.

2. Перезапись на те же носители (оптические диски или жесткий диск компьютера). Эта технология сохраняет полную идентичность электронного документа, что делает ее одной из наиболее предпочтительных. Но здесь есть два сложных момента. Во–первых, пока трудно определить жизненный цикл электронного документа, и, следовательно, выработать оптимальную целесообразность сроков перезаписи. Во–вторых, по–прежнему остается опасность того, что прекратится поддержка исходных форматов, а значит, со временем возможна утрата ряда свойств сохраняемых документов.

3. Так называемая «миграция» в новую аппаратно–программную среду, которая призвана гарантировать полную воспроизводимость электронного документа. Ключевым моментом становится цена вопроса и определение сроков необходимой модернизации.

4. Эмуляция, подразумевающая более технологически сложное решение по сравнению с миграцией. Речь идет об имитации на новом витке развития компьютерных технологий старой программной среды. Специалисты полагают, что успешное применение этого метода возможно только при вычленении самих цифровых данных из программной оболочки и наличии полного описания цифрового объекта в его содержательно–программной целостности.

5. Сохранение компьютеров определенного поколения, на которых создавался электронный ресурс. Вероятно, это решение относится к разряду тупиковых, поскольку превращает библиотеку в склад малоиспользуемого оборудования. Да и срок работоспособности самих компьютеров не может быть настолько большим, что бы удовлетворять требованиям гарантированного долгосрочного сохранения электронных документов.

Одним из вариантов стратегии сохранности ЭБ может быть следующая:

- Создание базовой (депозитарной) электронной копии объекта с использованием открытых форматов данных с возможностью сложной разметки текста и характеризующихся четко определенным синтаксисом и семантикой.
- Автоматическое формирование на основе депозитарной копии необходимых представлений в форматах, удобных для работы конечного пользователя.
- Возможность итеративного возврата с целью модификации и развития депозитарной копии в процессе ее эксплуатации.
- Базовая электронная копия должна являться переносимым электронным документом, аппаратно и программно независимым, включать необходимую метainформацию и быть выполнена в форматах данных, обеспечивающих долговременное хранение. Базовая копия хранит всю имеющуюся информацию об объекте в удобной для последующей обработки форме.

Требования к программному обеспечению. Программное обеспечение для развитых электронных библиотек может иметь клиент-серверную архитектуру. Для большинства обычных электронных библиотек в качестве клиента используется стандартный браузер, а сама электронная библиотека поддерживается стандартным Интернет-сервером.

Основные функции ПО для развитой ЭБ могут быть сформулированы следующим образом:

- достижение интероперабельности разнородных информационных ресурсов в рамках ЭБ;
- обеспечение создания, поддержки и выбора баз данных, интерфейсов, структур данных и сервисов, обеспечивающих наилучший поиск и актуализацию данных в разнородных коллекциях электронных документов;
- участие в формировании и эксплуатации корпоративных информационных ресурсов;
- обеспечение защиты и авторизованного доступа к информации со статистикой о времени, проведенном пользователем в ЭБ и использованных документах;
- возможность передачи одного запроса пользователя одновременно в несколько коллекций, составляющих ЭБ, и обработку ответов от нескольких источников;
- визуализация информации, полученной в ответ на запрос, в формате, выбираемом пользователем из списка возможных;
- наличие лицензионных соглашений;

- учет при функционировании принятых стандартов представления, хранения и передачи информации;
- обеспечение адекватного использования выбранных или разработанных в рамках проекта метаданных;
- возможность автоматизированного расчета с пользователями, владельцами и авторами электронных ресурсов;
- минимальные требования к программному обеспечению рабочего места клиента (стандартное программное обеспечение клиента);
- независимость функционирования ЭБ от платформы на стороне клиента и независимость от используемой СУБД;
- возможность увеличения количества типов данных и подключаемых электронных коллекций;
- поддержка распределенных систем хранения информации;
- возможность обработки многоязычной текстовой информации с использованием оригинальной графики документов и метаданных; использование UNICODE.

Состав и содержание работ по созданию электронной библиотеки. На стадии реализации выполняется разработка электронной библиотеки, производится её тестирование на мощностях разработчика, подготавливается соответствующая документация. Стадия реализации электронной библиотеки может включать в себя несколько этапов. На каждом этапе разрабатывается определенный для данного этапа функционал. Ниже приводится описание этапов для реализации электронной библиотеки в Российской государственной библиотеке.

1 этап – создание хранилища документов. В ходе первого этапа должен быть реализован следующий функционал, реализующий загрузку и хранение данных в электронном фонде и электронном каталоге:

- Добавление нового электронного документа с созданием учетной записи электронного документа;
- Хранение электронных документов. Срок хранения электронного документа не ограничен. Хранение электронных документов должно осуществляться в файловой системе;
- Перемещение электронного документа в файловой системе с внесением изменений в учетную запись, обеспечение неизменности URL адреса электронного документа;
- Замещение электронного документа другим с внесением изменений в учетную запись;
- Импорт, экспорт и преобразование библиографических записей;
- Преобразование кодировки данных библиографической записи.

2 этап – создание системы доступа и показа документов, формирование системы сбора статистики. В ходе второго этапа должен быть реализован следующий функционал:

- Регистрация нового пользователя в системе;
- Регистрация администраторов электронного фонда;
- Изменение учетной записи пользователя;
- Назначение прав доступа для пользователя к коллекции электронных документов;
- Проверку прав доступа пользователя;
- Подключение/отключение электронного каталога к поисковому механизму;
- Добавление/изменение диапазона IP-адресов, с которых разрешен вход в Систему для абонентов электронной библиотеки;

- Проверку IP-адреса, с которого пользователь вошел в систему;
- Извлечение фрагмента электронного документа по запросу пользователя. Для электронных документов, хранящихся в формате PDF, DOC, XLS, имеющих разбиение на страницы, необходимо обеспечить извлечение постранично. Для графических изображений форматов JPEG, TIFF, GIF, PNG целесообразно обеспечить преобразование документа, с целью уменьшения размера изображения и текущего разрешения. Необходимо обеспечить возможность настройки качества графического изображения, передающегося для просмотра пользователю. Для электронных документов, хранящихся в других форматах, функция просмотра фрагмента документа не реализуется;
- Извлечение копии электронного документа для скачивания по запросу пользователя;
- Переход на список подборки, в которую включен рассматриваемый пользователем электронный документ;
- Формирование запроса на скачивание документа; В запросе необходимо предусмотреть указание способа предоставления документа (запись на CD, запись на Flash носитель, печать);
- Сбор статистической информации при обработке запросов пользователя и выполнении функций администрирования системы.

3 этап – создание поискового аппарата. В ходе третьего этапа в электронной библиотеке должен быть реализован следующий функционал, реализующий обработку поисковых запросов:

- Выбор формы поискового интерфейса (простой/расширенный);
- Настройка поисковых признаков, списка электронных каталогов к которым формируется запрос, формы предоставления результатов запроса;
- Формирование поискового запроса;
- Обеспечение вывода результатов поиска в табличной форме на экран и печать каталожных карточек с возможностью сортировки по столбцам таблицы. Ссылка на полный текст электронного документа должна включаться в форму вывода и подсвечиваться;
- Сохранение критериев поискового запроса;
- Добавление/удаление ссылки на электронный документ в избранное (корзину);
- Просмотр/редактирование содержимого избранного (корзины);
- Разработка полнотекстового индекса;
- Возможность настройки правил формирования критериев поискового запроса из полей библиографической записи;
- Поиск информации в электронном каталоге и в распознанных электронных документах.

4 этап – создание распределенной системы. В ходе четвертого этапа в системе должен быть реализован следующий функционал, реализующий управление распределенной информационной системой:

- Подключение/отключение электронной библиотеки;
- Добавление/удаление коллекции, доступной для совместного доступа в распределенной информационной системе;
- Развитие поисковой системы для поддержки распределенного поиска;
- Маршрутизация запросов на доступ к документам в удаленных электронных библиотеках;
- Формирование отчетов;
- Просмотр статистики выполненных работ, включая текущую стоимость сеанса;

- Сбор статистической информации при обработке запросов пользователя и выполнении функций администрирования в распределенной электронной библиотеке.

Направления взаимодействия создателей электронных библиотек и коллекций. Важнейшими направлениями взаимодействия в области создания электронных библиотек, являются правовое обеспечение электронных библиотек, координация формирования их фондов в «вертикальном» и «горизонтальном» аспектах, достижение совокупной функциональной полноты в рамках определенного сегмента электронного документного пространства.

Задача достижения взаимодействия и взаимоиспользования электронных библиотек предполагает четкое понимание неконкурентности всех субъектов электронного документного пространства, поскольку их целевое назначение имеет существенные различия.

Субъекты электронного документного пространства формируют электронные библиотеки различных типов в соответствии со своими задачами.

Инструментом координации формировании фондов электронных библиотек может быть соглашение о разграничении зон ответственности генераторов электронных библиотек. Взаимодействие носит «вертикальный» и «горизонтальный» характер. Вертикальный характер носит взаимодействие электронных библиотек с федеральными структурами. Для сферы электронных библиотек предлагались различные варианты такого федерального центра, в частности Национальная электронная библиотека, описанная выше, Федеральный информационный центр, проектируемый в рамках «Электронной России», Российская государственная библиотека и другие структуры. С учетом последних инициатив Президента, можно ожидать, что определенные функции по вертикальной координации электронных библиотек сможет выполнять Президентская библиотека им. Б. Н. Ельцина.

Для реализации горизонтальных связей предлагается концепция регионального ресурсного центра, рассмотренная ниже.

1.3. ЭБ как информационно-поисковая система

Новые информационные технологии и современная теория лингвистического обеспечения библиотечных процессов открывает широкие возможности для универсализации традиционных библиотечных классификационных систем, выступающих как в роли языков индексирования информационных материалов, так и формы представления знания в электронной базе данных. Широкое использование информационно-поисковых языков вербального типа (тезаурус, ключевые слова) в условиях автоматизации библиотечной технологии создает оптимальные возможности для комплексного применения информационно-поисковых языков символического типа (ББК, УДК, ГРНТИ, DDC) в индексировании информационных материалов. Совместимость классификационных систем символического типа обусловлена и обеспечивается полным или частичным тождеством между их лингвистическим содержанием на крупных иерархических уровнях. Такая совместимость однозначна на уровне классификационных разделов и дифференцирована на уровне их основных делений. При этом отправным или базовым может быть выбран любой из информационно-поисковых языков. В ГНПБ им. К.Д. Ушинского это Библиотечно-библиографическая классификация как наиболее соответствующая систематике педагогической науки.

При построении информационных систем, в процессе разработки информационно-поисковых языков возникает вопрос о статусе языковых единиц. При разработке принципов упорядочения и стандартизации отраслевой терминологии в центре внимания находится пара «понятие – термин» и, соответственно, «система понятий – система терминов». С проблемой «система понятий – система терминов» связан вопрос о полной или частичной эквивалентности терминов. В философии науки он составляет часть проблемы интертеоретических отношений. Интертеоретические отношения можно рассматривать как отношения сравнимости и переводимости языков науки внутри одной предметной области. Проблема переводимости языков науки в философии науки предстает как проблема соизмеримости (несоизмеримости) теорий, а в терминологии, или терминоведении как лингвистической дисциплины – как проблема унификации и координации терминов, построения ряда соответствий, где можно было бы указать, имеются ли между терминами отношения эквивалентности, включения, пересечения или исключения, подобные отношениям между теориями.

Предметом нашего рассмотрения является отраслевой тезаурус, вбирающий в себя понятийно-терминологический аппарат такой интегрированной области человеческого знания как педагогика (педагогическая наука).

Базовыми единицами тезауруса являются термины, когда речь идет о тезаурусе как о таковом, и дескрипторы, когда речь идет об информационно-поисковом тезаурусе. «Дескриптор – это слово (термин), которое выбирается из ряда синонимических терминов в качестве представителя этого ряда (предмета информационно-поискового запроса) в информационно-поисковых системах (информационно-поисковых тезаурусах). Перечень дескрипторов формирует поисковый образ документа» (Краткий словарь лингвистических терминов – М.: Русский язык, 1995, с. 31).

Мы пытаемся исходить из того положения, что дескриптор со всеми нюансами своих лингвистических характеристик (лексико-семантической вариативностью, синтаксической валентностью, парадигматическими и синтагматическими связями, наконец, своим семантическим потенциалом) почти равен термину. Наиболее релевантными признаками термина являются: стилистическая нейтральность, простота языковой формы, точность, понятность, краткость и нормативность. Другими словами, термин должен быть легко произносим, легко понимаем и удобен в обращении. Кроме того, термин должен обладать систематизирующими свойствами. Слово становится термином всякий раз, когда начинает обозначать научное понятие о предметах, явлениях, признаках, составляющих вместе с другими понятиями данной отрасли человеческой деятельности одну семантическую систему. Термин легко переходит вместе с научными понятиями из языка в язык и содействует международному научному общению. Термин призван обеспечивать оптимальное взаимопонимание в общении между специалистами.

Многие лингвисты предъявляют к термину такие требования, как однозначность и отсутствие синонимии. Следует заметить, что в большинстве случаев эти требования невыполнимы постольку, поскольку специфика термина лежит не только в сфере свойств лексической единицы, но и в сфере ее функций. Так как, термин – это основная единица специальной лексики, то на его классификационные характеристики накладываются систематизирующие и структурные признаки определенной области знания, отрасли науки. А в таком случае «сущностное проявление языкового знака (термина) связано с тем, что в его структуре сфокусировано отражение трех ипостасей: вещного

мира, языковой системы и языковой личности, выражение которых сообщает содержанию языкового знака комплексность и многоаспектность» (Данеш Ф.).

Некоторая многозначность и широкая синонимия термина-дескриптора, несмотря на внешнее противоречие с традиционной терминологией, только обогащает современную терминосистему, которая, в свою очередь, отражает интеграцию и взаимопроникновение различных областей знания и отраслей науки. Вот эта проблема интертеоретических («межнаучных», «междисциплинарных») отношений и заставляет обратиться к вопросу о синонимии в научной терминологии. В естественном языке синонимия является реализацией семантической гибкости языка, т.е. возможность разными способами выразить одно и то же или почти одно и то же содержание. В том осознанном процессе, которым является создание терминологии, синонимии, казалось бы, не должно быть места. Однако явления сходства по значению при различии в форме, и совпадения в форме при различии в значениях существуют в любой реально существующей терминологии. Как нам представляется, нужно различать внутриязыковую синонимию и частично эквивалентную лексику разных научных языков. Первая существует в пределах одного теоретического единства, когда на базе общих концептуальных представлений создается научный «обиход», где одно и то же понятие может быть выражено либо средствами разных естественных языков (Вин. пад.- аккузатив), либо наряду с общепринятой традицией, имеет номинацию, возникшую в индивидуальном языке ученого (при этом вовсе не обязательно создающего новые теории). Второе явление никак не может быть названо синонимией, как не могут быть названы синонимами слова из разных языков со сходными значениями (ср.: «варианты фонемы» у Реформатского и «аллофоны» у дескриптивистов).

Наиболее полно диалектические связи между языковыми знаками и реалиями экстралингвистической действительности раскрываются с помощью полевой модели системы языка. Поле в лингвистике становится одной из важнейших категорий, в которой системно и синтетически рассматриваются однородные по своему содержанию единицы; при этом описывается не только сам язык, но и система соответствующих понятий. Поле задается определенным смысловым содержанием, доминантой поля. В нем выделяется ядро (лексема-понятие или группа лексем-понятий), центр (классы основных понятий, реалем с их синонимическими, антонимическими и другими отношениями) и периферия (система смежных реалий, смежных полей – слов – понятий вторичной семантической функции).

Поле, представляя собой определенную семантическую структуру, отражает иерархию основных классов слов (реалем) и соотносительных языковых средств. Такая структура может быть описана с помощью родо-видовых отношений, отражающих предметно-понятийную классификацию элементов.

Современная интерпретация семантического поля опирается на синтагматические и парадигматические свойства единиц поля. Поэтому целесообразно, особенно учитывая коммуникативный характер языка, говорить о парадигматическом и синтагматическом измерении, где существенны не только положение слова-понятия в парадигме, но и его синтагматические (сочетаемые) свойства. Такая сложная система систем как язык может функционировать только на основе внутренней согласованности и синхронизации ее элементов.

Полевая модель языка легла в основу структуры информационно-поискового тезауруса по педагогике (Тезаурус ЮНЕСКО-МБП по образованию). Дескрипторы в тезаурусе размещены в восьми широких группах, определяемых как поля и охватываю-

щих наиболее широкую область педагогики и образования и смежных с ними отраслей науки и техники. Предметно-понятийные поля педагогического тезауруса описывают:

– социально-политический и общественный строй, в рамках которого осуществляется педагогический процесс (Поле 1. Контекст. Сюда такие дескрипторы, как: ПОЛИТИКА, ДЕМОГРАФИЯ, НАУКА, СЕМЬЯ и т.д.);

– область исследования и принципы управления образованием (Поле 2. Управление и исследования. Сюда входят такие дескрипторы, как: ПЛАНИРОВАНИЕ ОБРАЗОВАНИЯ, УПРАВЛЕНИЕ КАДРАМИ, РУКОВОДСТВО ШКОЛОЙ, СТАНДАРТЫ и т.д.);

– организацию процесса обучения (Поле 3. Обучение. Сюда входят такие дескрипторы, как: ПРОГРАММЫ ОБУЧЕНИЯ, УЧЕБНЫЙ ПЛАН, ПРОГРАММИРОВАННОЕ ОБУЧЕНИЕ, УЧЕБНО-ВОСПИТАТЕЛЬНАЯ СРЕДА и т.д.);

– объекты образовательного процесса – людей (Поле 4. Люди. Сюда входят такие дескрипторы, как: УЧАЩИЕСЯ, УЧИТЕЛЯ, КОНСУЛЬТАНТЫ, РЕЛИГИОЗНЫЕ ГРУППЫ и т.д.);

– развитие и учение с психологической и физиологической точек зрения (Поле 5. Развитие и учение. Сюда входят такие дескрипторы, как: МЫСЛИТЕЛЬНЫЕ ПРОЦЕССЫ, ЭМОЦИИ, ПОЛ, ФИЗИОЛОГИЯ и т.д.);

– учебно-предметное содержание процесса обучения и воспитания (Поле 6. Содержание. Сюда входят такие дескрипторы, как: БИОЛОГИЧЕСКИЕ НАУКИ, ВЫСШЕЕ ОБРАЗОВАНИЕ, ПЕДАГОГИЧЕСКОЕ ОБРАЗОВАНИЕ и т.д.);

– материальную базу процесса обучения и воспитания (Поле 7. Вещи. Сюда входят такие дескрипторы, как: РЕСУРСЫ, ШКОЛЬНЫЕ МАСТЕРСКИЕ, ЭВМ, НАГЛЯДНЫЕ МАТЕРИАЛЫ и т.д.);

– пространственно-временные рамки и этнические характеристики образовательного процесса (Поле 8. Место. Время. Пространство. Сюда входят такие дескрипторы, как: КОНТИНЕНТЫ, ОКЕАНИЯ, РУССКИЕ, ХРОНОЛОГИЧЕСКИЕ РАМКИ и т.д.).

Учитывая образовательное значение педагогической информации, целесообразно в процессе пополнения тезауруса новыми терминами руководствоваться следующими лингвистическими принципами:

- устойчивость языковой нормы, консерватизм языкового материала,
- распространенность языкового материала как принцип нормы,
- авторитет отраслевого и лингвистического источника как признак нормы,
- принцип функциональной и коммуникативной целесообразности,
- взаимообусловленность грамматических, лексических и семантических факторов,
- взаимосвязь парадигматических отношений и ассоциативных связей как основы построения фасетов и полей тезауруса,
- сочетаемость возможности (валентность) дескрипторов (терминов).

Лексическая наполненность и семантический потенциал логико-понятийных полей информационно-поискового тезауруса по педагогике, на наш взгляд, позволяют ему не только функционировать в качестве информационно-поискового языка, но и в равной мере представлять достаточно интегрированную и постоянно обновляющуюся педагогическую терминологию. Система отслеживания и распространения современной мировой педагогической терминологии позволяет достичь согласованности отечественной научной педагогической терминологии с наиболее распространенной в мире терминологической системой.

Говоря о системе терминов, мы подразумеваем сложную систему семантических отношений между ними. Описание терминосистем всегда связано с разработкой клас-

сификаций не только самих терминов, но и классификаций отношений. Традиционной является классификация (предложенная Е. Вюстером), в которой отношения делятся на онтологические и логические, а далее внутри каждой категории устанавливается своя иерархия: логические отношения определяются как отношения подобия, онтологические – как отношения смежности в пространстве и времени. Логическая иерархия выстроена с т.з. логики (математической, модальной и т.д.) реальных явлений, событий, отношений и вещей, существующей в реальном отрезке (фрагменте) времени и пространства. Онтологическая (историческая в иной терминологии) иерархия заложена в семантике самого термина и самодостаточна для выявления всех реально существующих или потенциальных связей и отношений между понятиями, событиями, явлениями, предметами, номинируемыми данным термином.

Глобальные таксономии представляют собой некоторые априорные схемы научного знания, которые могут накладываться на конкретную терминологию. Не менее ценными являются лингвистические эмпирические классификации семантических отношений, производимые при рассмотрении специальных текстов. При этом становятся очевидными определенные корреляции между этими видами классификаций. Можно сказать, что классификации показывают, как организуется и воплощается знание в семантической структуре терминологии.

В рамках языка науки на первый план выдвигается не номинативность термина имени, а функция выражения (репрезентации) целостного научного понятия, чей суммарный семантико-логический, понятийный объем значительно превышает объем семантики номинативной «оболочки» термилируемого понятия. Номинативная функция термина связанная в основном с именем существительным, может считаться главенствующей, если изучать термины изолированно от сферы функционирования и без учета понятийных и языковых связей и генезиса языка науки в целом.

Характерно, что наиболее подробные и полные классификации ориентированы на системы информационного поиска и явились итогом соединения философско-логических исследований о природе знания с прагматическими запросами информатики.

Итак, тезаурус представляет собой некоторую итоговую форму описания и представления систематизированного материала. Отраслевой тезаурус – это дискретный парадигматический инструмент со своим метаязыком, который может служить лингвистической моделью целостного знания, выраженного в научных текстах. Это словарь с концептуальным входом и фиксированными семантическими связями между его единицами. Функциональная значимость отраслевого тезауруса определяется тем, что он, репрезентируя терминосистему науки, одновременно является одним из наиболее оптимальных способов классификации, хранения и передачи специальной информации, средством создания поискового образа документов в электронных библиотечных информационно-поисковых системах. Концептуальный диапазон языка позволяет синтезировать информацию, сохраняя при этом потенциальную возможность этой информации к разложимости на составляющие.

Дескрипторы-термины, составляющие логико-понятийные поля тезауруса, являются одновременно вербальными элементами поискового образа документа в электронной базе данных. Документ – это в нашем случае педагогическая информация, заложенная в монографии, статьи и периодические издания, и подлежащая индексированию (идентификации). Индексирование – это процесс перевода содержания документов с естественного на информационно-поисковый язык и одна из разновид-

ностей аналитико-синтетической обработки или свертывания информации, что, в свою очередь, влечет за собой создание поискового образа документа. (См. Гендина Н.И. Лингвистическое обеспечение автоматизированных библиотечных систем. – Алма-Ата: Гылым, 1991, с. 136). Поисковый образ документа составляется из нескольких классификационных языков: ББК (библиотечно-библиографической классификации), ГРНТИ (государственного рубрикатора научно-технической информации), классификации Дьюи, тезауруса и словаря ключевых слов. Релевантность поискового образа документа достигается: а) согласованностью дескрипторной строки и синтагматической цепочки; б) соотношенностью лингвистической семантики с логической.

Рассмотрим Русско-английский Тезаурус по педагогике и образованию ГНПБ им. К.Д. Ушинского.

Тезаурус разработан на основе Тезауруса ЮНЕСКО-МБП по образованию. (UNESCO, 1991, 1998). Ведется в ГНПБ им. К.Д. Ушинского с 1993 года. Тезаурус применяется в качестве информационно-поискового языка лексического типа, используется в автоматизированных библиотечных системах и трансформируется в рабочую программу путем перемещения данных через буфер обмена операционной системы WINDOWS.

ТЕЗАУРУС постоянно пополняется русской педагогической терминологией с английским эквивалентом, насчитывает более пяти тысяч записей основного языка.

Термины (дескрипторы) расположены в восьми лексико-семантических полях, определяемых как «1. КОНТЕКСТ», «2. УПРАВЛЕНИЕ И ИССЛЕДОВАНИЯ», «3. ОБУЧЕНИЕ (ПРЕПОДАВАНИЕ)», «4. ЛЮДИ», «5. РАЗВИТИЕ И УЧЕНИЕ», «6. СОДЕРЖАНИЕ», «7. ВЕЩИ», «8. ОПРЕДЕЛИТЕЛИ МЕСТА И ВРЕМЕНИ».

Поля подразделяются на лексико-грамматические группы (фасеты). Фасеты состоят из терминов (дескрипторов), охватывающих связанные между собой понятия. Дескрипторы расположены в определенной иерархии и снабжены уточняющими элементами. Это – ограничительные пометы (аннотации), синонимы, входные термины, цифровые коды, ссылки, которые в целом и составляют словарную статью дескриптора.

Тематическая область ТЕЗАУРУСА отражает педагогическую науку и практику, а также сопредельные отрасли знания.

Структура ТЕЗАУРУСА подчинена лингвистическим законам парадигматики и синтагматики и представлена двумя списками: алфавитным (дескрипторы расположены в алфавитном порядке) и древовидным (дескрипторы расположены в иерархическом порядке).

Система предусматривает возможность поиска дескриптора по его фрагменту (пермутация).

Один и тот же дескриптор может входить в разные фасеты и может быть вышестоящим или нижестоящим по отношению к другим дескрипторам.

Структура ТЕЗАУРУСА дополнена 73 уточняющими подрубриками (модификаторами). Подрубрики предназначены для уточнения основных дескрипторов, отражающих учебную и методическую литературу.

ТЕЗАУРУС является репрезентатором международной педагогической терминологии, инструментом по индексированию информационных материалов и параметром вербального поискового элемента в ЭЛЕКТРОННОЙ БАЗЕ ДАННЫХ ГНПБ ИМ. К.Д. УШИНСКОГО.

Приведем пример фрагмента тезауруса.

ФАСЕТ

185 СТУПЕНИ ОБРАЗОВАНИЯ 185

СТУПЕНИ ОБРАЗОВАНИЯ 185.141

ОБРАЗОВАНИЕ СРЕДНЕЕ 185.287

ОБРАЗОВАНИЕ НЕПОЛНОЕ СРЕДНЕЕ 185.287.352

ОБРАЗОВАНИЕ ПОЛНОЕ СРЕДНЕЕ 185.287.556

ОБРАЗОВАНИЕ ДОШКОЛЬНОЕ 185.360

ОБРАЗОВАНИЕ ВЫСШЕЕ 185.546

ОБРАЗОВАНИЕ НЕПОЛНОЕ ВЫСШЕЕ 185.546.374

КРАТКОСРОЧНЫЙ КУРС ВЫСШЕГО ОБРАЗОВАНИЯ 185.546.748

ОБРАЗОВАНИЕ НАЧАЛЬНОЕ 185.653

ОБРАЗОВАНИЕ ПОСЛЕВУЗОВСКОЕ 185.779

Приведем пример словарной статьи.

ОБРАЗОВАНИЕ НЕПОЛНОЕ ВЫСШЕЕ (дескриптор)

LOWER HIGHER EDUCATION (английский эквивалент)

185.546.374 (цифровой код)

в Российской Федерации первый уровень высшего профессионального образования, осуществляемый высшим учебным заведением в течение первых двух лет обучения. По завершении этого срока студент имеет возможность продолжить высшее образование или получить диплом о неполном высшем образовании без итоговой аттестации (ограничительная помета, аннотация).

X: НЕПОЛНОЕ ВЫСШЕЕ ОБРАЗОВАНИЕ (синоним, входной термин)

Глава 2

Технология автоматической обработки текста для извлечения и представления знаний

В главе представлена технология автоматической обработки текстовой информации с применением формализма искусственных нейронных сетей. Показано, как на основе биологических сведений об обработке информации в мозге человека строится искусственная нейронная сеть, формализм которой позволяет реализовать обработку и представление информации, в том числе, текстовой, такие же, как у прототипа. Указанная нейронная сеть реализует ассоциативный доступ информации и, как следствие, структурный анализ информации. Показано, как в мозге человека реализуется модель мира – две взаимосвязанные ассоциативные сети – языковая и образная. Представлен способ автоматической обработки текстовой информации с использованием описанных представлений. Описана технология автоматической обработки текстовой информации TextAnalyst. Представлены приложения, реализованные на основе технологии: программные системы TextAnalyst, электронная книга, а также разрабатываемая в настоящее время персональная библиотека преподавателя/учащегося с ассоциативной навигацией и удаленным доступом.

2.1. Автоматическая обработка текстовой информации

В этом разделе представлено краткое описание реализации нейросетевой технологии применительно к автоматическому смысловому анализу текстовой информации [15, 16, 17] на примере семейства программных продуктов TextAnalyst, разработанного Научно-производственным инновационным центром «Микросистемы», г. Москва.

Существующие системы для анализа текстов используются для автоматического извлечения информации из текстов, и представления ее пользователям в сжатом и удобном виде. Основными функциями системы анализа текстов являются: извлечение индекса текста, автоматическое формирование множества рубрик – кластеризация множества текстов, отнесение текста к рубрике (классификация) и сравнение текстов, создание реферата, а также – формирование гипертекстового представления со средствами навигации по гипертекстовой структуре. Важнейшая из этих функций – индексация – осуществляется в разных системах по-разному. В основном используются два подхода. Статистический подход позволяет сформировать индекс на основе анализа частоты встречаемости слов в тексте, как в системе IBM Intelligent Miner for Text. Семантический подход использует в этих целях заранее сформированное семантическое представление в виде глобальной семантической сети, в том или ином представлении,

как в системе Экскалибур. В этом случае проекция входного текста на эту сеть и формирует индекс текста. Все остальные функции являются производными от функции индексации. В зависимости от качества ее реализации стоит и качество работы остальных функций.

Автоматизация извлечения знаний из текстов в современных гипертекстовых, информационно-поисковых и экспертных системах, средствами формализмов искусственных языков представления знаний, встречает ряд трудностей, связанных с разнообразием используемых методов, слабой автоматизацией процессов формирования базы знаний и необходимостью привлечения экспертов – лингвистов и инженеров по знаниям – для формирования базы знаний. Поэтому существующие системы являются весьма громоздкими структурами с большими семантическими сетями.

Представленная технология обработки текстовой информации основана на использовании структурных свойств языка и текста, которые могут быть выявлены с помощью статистического анализа, реализованного на основе описанного выше нейросетевого подхода. При этом автоматически формируется описание семантики предметной области текста, и реализуются функции организации текстовой базы в гипертекстовую структуру, автоматического реферирования, кластеризации и классификации текстов, а также функция смыслового поиска.

Нейросетевой подход является удобным инструментом для выявления структурных свойств текстовой информации. Использование указанного инструмента позволяет автоматически, на основе анализа статистики слов и их связей в тексте, реконструировать внутреннюю структуру текста.

Статистический анализ выявляет наиболее часто встречающиеся элементы текста: слова или устойчивые словосочетания. Важной особенностью используемого подхода, является возможность автоматически устанавливать взаимосвязи между выявленными элементами текста. При выявлении связей учитывается статистика попарного появления слов во фрагментах исследуемого материала. Далее статистические характеристики пересчитываются в семантические с помощью итерационных алгоритмов, учитывающих весовые характеристики взаимосвязанных элементов, а также учитывающих весовые характеристики связей. После пересчета статистических характеристик в семантические, понятия, которые мало соответствуют анализируемой предметной области, получают малый вес, а наиболее представительные наделяются высокими показателями. Полученная семантическая сеть позволяет производить различные виды анализа текстовой информации. Сеть отражает внутреннюю структуру текста, значимость выделенных понятий, а также, показывает степень связанности понятий в тексте. Такое представление текста получается полностью автоматически. Семантические веса элементов сети используются при выделении наиболее информативных участков текста. Использование ассоциативных связей элементов сети позволяет реализовать эффективную навигацию по тексту.

2.2. Технология обработки и представления текстов

На основе алгоритмов обработки текстовой информации, описанных выше, создана технология для анализа текстовой информации. Технология реализована как инструмент для автоматического формирования баз знаний на основе множества естественно-языковых текстов. Ядро системы выполнено как программный компонент (inproc server), соответствующий спецификации Component Object Model (COM) фирмы Microsoft.

Ядро системы реализует следующие функции. Нормализацию грамматических форм слов и вариаций словосочетаний. Автоматическое выделение базовых понятий текста (слов и словосочетаний) и их взаимосвязей с вычислением их относительной значимости. Формирование представления семантики текста (множества текстов) в форме ассоциативной (однородной) семантической сети.

В состав ядра системы, помимо блока первичной обработки, входят следующие блоки (см. рис. 1): лингвистический процессор, блок выделения понятий текста, блок формирования семантической сети, блок хранения семантической сети.

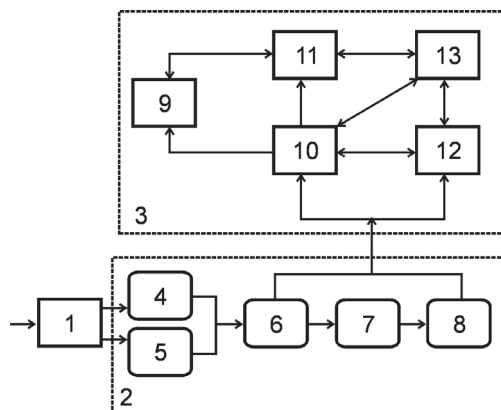


Рис. 1. Система анализа текстов содержит блок первичной обработки (1), лингвистический, и семантический процессоры. Лингвистический процессор (2) состоит из словарей: (4) слов разделителей, (5) служебных слов, (6) общеупотребимых слов, а также (7) флективных и (8) корневых морфем. Семантически процессор (3), в свою очередь, содержит: (9) блок отсылок в текст, (10) блок формирования семантической сети, (11) блок хранения семантической сети, (12) блок выделения понятий, и (13) блок управления.

Блок первичной обработки. Задачами этого блока являются извлечение текста из файла (входного потока данных) и подготовка его к обработке в лингвистическом процессоре. Подготовка текста заключается в очистке его от символов, неизвестных лингвистическому процессору, а также в корректной обработке таких единиц текста как аббревиатуры, инициалы, заголовки, адреса, номера, даты, указатели времени.

Лингвистический процессор. Лингвистический процессор осуществляет предобработку входного текста (последовательности символов в определенной кодировке) на основе априорных лингвистических знаний, общих для выбранного языка (в настоящий момент поддерживаются несколько европейских языков, помимо русского и английского), и выполняет следующие функции. Сегментацию предложений текста на основе знаков пунктуации и специальных грамматических слов, и их фильтрацию. Нормализацию слов и словосочетаний – фильтрацию флексий (окончаний) с сохранением только корневых основ. А также – фильтрацию в тексте семантически несущественных, вспомогательных слов: удаляются предлоги, числительные и самые общеупотребимые слова с широким значением. И, наконец, маркировку общеупотребимых слов.

Сегментация предложений позволяет разбить текст на участки, которые могут содержать терминологические словосочетания предметной области и избежать выделения неадекватных словосочетаний на стыках таких участков.

В результате предобработки семантически близкие словосочетания приводятся к одинаковой форме (нормализуются). Маркировка общеупотребимых слов необходима с целью исключения их выделения как самостоятельных терминов при дальнейшем анализе.

База общих языковых знаний лингвистического процессора содержит словари, по одному для реализации каждой из четырех функции: словарь слов-разделителей предложения, словарь вспомогательных слов, словарь флексий и словарь общеупотребимых слов.

Блок выделения понятий. Блок выделения ключевых понятий предметной области (слов и словосочетаний) создан на базе программной модели иерархических структур из нейронных сетей, и реализует алгоритмы автоматического формирования частотного словаря текста.

Число уровней в иерархической структуре определяет априорно заданную максимально допустимую длину понятия предметной области и равняется двадцати. На первом уровне иерархической структуры представлен словарь двухбуквенных специальных слов предметной области – слов, пропущенных через все фильтры лингвистического процессора, и не отнесенных к общеупотребимым, а также двухбуквенных сочетаний из слов этого словаря. Там же хранятся двухбуквенные слова общеупотребимой лексики, входящие в устойчивые словосочетания, и их начальные двухбуквенные фрагменты. На втором уровне иерархической структуры представлены словари трехбуквенных слов и сочетаний букв из словарей специальных и общеупотребимых слов, встреченных в тексте, в виде индексов элементов соответствующих словарей первого уровня, дополненных еще одной буквой. На последующих уровнях представление информации полностью однородно – хранятся индексы элементов хранения более низкого уровня, дополненные одной буквой.

В процессе формирования представления информации в иерархической структуре подсчитывается частота встречаемости каждого сочетания букв в соответствующих элементах. Частота слов (сочетаний букв, не имеющих продолжения на следующем уровне) используется для последующего анализа.

Сформированное таким образом представление лексики текста подвергается затем пороговому преобразованию по частоте встречаемости. Порог отражает степень детальности описания текста. В процессе статистического анализа в иерархической структуре выделяются устойчивые термины и терминологические словосочетания, которые служат далее в качестве элементов для построения семантической сети. При этом общеупотребимые слова, а также словосочетания, содержащие только общеупотребимые слова, опускаются.

Блок формирования семантической сети. Блок формирования семантической сети реализован как база данных, в которой представляются семантические связи понятий предметной области. Поскольку типы семантических связей [18] в системе не определяются, такие связи представляют собой просто ассоциативные связи.

В качестве критерия для определения наличия семантической связи между парой понятий используется частота их совместной встречаемости в одном предложении. Превышение частотой некоторого порога позволяет говорить о наличии между понятиями ассоциативной (семантической) связи, а совместные вхождения понятий в предложения с частотой меньше порога считаются просто случайными.

Элементы семантической (ассоциативной) сети и их связи имеют числовые характеристики, отражающие их относительный вес в данной предметной области – семантический вес. При достаточно представительном множестве текстов, описывающих предметную область, значения частот встречаемости понятий действительно отражают соответствующие семантические (субъективно оцениваемые) веса. Однако, для неболь-

ших обучающих выборок, в частности, при анализе отдельного текста, не все частотные характеристики соответствуют действительным семантическим весам – важности понятий в тексте. Для более точной оценки семантических весов понятий используются веса всех связанных с ними понятий, т.е. веса целого «семантического сгущения». В результате такого анализа наибольший вес приобретают понятия, обладающие мощными связями и находящиеся как бы в центре «семантических сгущений».

2.3. Психо-физиологическая основа данного подхода

В мозге человека сосуществуют два параллельно работающих взаимосвязанных представления модели мира, одно представление – образное многомодальное, другое – метаязыковое. Оба этих представления организованы в виде ассоциативных сетей, состоящих из вершин-событий с их взаимосвязями. Метаязыковое представление позволяет организовать ассоциативный доступ к нужной информации: обращение к понятию метаязыковой сети автоматически адресует к соответствующему образу образной сети. Образная сеть, в свою очередь, позволяет интерпретировать языковые сообщения в терминах адресуемых этими сообщениями образов образной сети.

Удобной для пользователя-читателя технологией обработки и представления текстовой информации была бы технология, которая бы извлекла из текста соответствующую этому тексту метаязыковую ассоциативную сеть ключевых понятий в их взаимосвязях. Легко воспринимаемая пользователем, такая сеть и в целом и по частям могла бы быть интерпретированной им в терминах его (пользователя) образной ассоциативной сети – образной модели мира. Одновременно, такая метаязыковая сеть ключевых понятий текста могла бы служить средством ассоциативной навигации по тексту, при наличии сформированной гипертекстовой структуры в составе собственно текста и извлеченной из нее метаязыковой (ассоциативной – семантической) сети. Такая технология существует и реализована в семействе программных продуктов для автоматического анализа неструктурированной текстовой информации «TextAnalyst». Рассмотрим некоторые нейроинформационные предпосылки возникновения этой технологии. Далее, познакомимся с нейросетевыми и информационными основами технологии. А также, рассмотрим собственно технологию обработки текстов, и программные продукты на основе этой технологии, как уже реализованные, так и разрабатываемые в настоящий момент.

Из работ по нейропсихологии известно [1, 2], что модальная информация в головном мозге человека хранится, в различных областях коры: зрительная – в затылочных долях, слуховая – в височных, соматосенсорная – ближе к лобным долям. В свою очередь, для хранения отдельных классов событий используются соответствующие отделы этих областей коры. Например, звуковые образы слов и их частей хранятся в полях левой височной доли (у правшей), зрительные сцены и их части – также в соответствующих полях височных долей, а также в затылочных долях. Информация о событиях, отображенных в разных модальностях, объединяется в единое целое – мультимодальный образ события – в теменной коре.

Это как бы статика модели мира: множество многомодальных образов, включающих в себя зрительные, слуховые, вкусовые, обонятельные и соматосенсорные компоненты, а также представление о вертикали. В правом полушарии они хранятся как индивидуальные многомодальные интегральные образы-синтагмы.

В левом полушарии хранится также многомодальная картина мира, но уже в парадигматическом представлении (в виде иерархии отдельных событий): на соответствующих уровнях иерархии представлены как бы словари событий различных уровней, например, сцен – на верхнем уровне – как схем соединения отдельных объектов более низкого уровня в сцену, объектов (уровнем ниже) как схем соединения элементов объектов, хранящихся еще ниже, в целые объекты, наконец, схем соединения элементарных представлений еще более низкого, по отношению к элементам объектов, уровня – в элементы объектов.

Это представление отличается от правополушарных образов наличием в нем вербальной слуховой компоненты (речевая информация обрабатывается только левым – у правшей – полушарием), которая позволяет именовать многомодальные образы с помощью элементов речи, как соответствующие им левополушарные, так и правополушарные. Вербальная слуховая компонента, таким образом, с одной стороны, входит в состав многомодального образа, с другой – выступает независимо в иерархии слуховых образов, опосредующих язык. Слуховая вербальная (языковая) иерархия включает в свой состав словари речевых образов от морфологического до синтаксического уровня, а также представления надъязыковых уровней: семантику и прагматику.

Элементы описанного выше статического хранилища мозга объединяются в динамические (развертываемые во времени) представления с использованием гиппокампа [3]. В одном из полей гиппокампа (поле CA₃) содержится информация о связях хранящихся в коре образов в более крупные образования – последовательности событий.

Рассмотрим более детально способы обработки и хранения информации в коре и гиппокампе. Для этого рассмотрим структурную единицу коры – так называемую макроколонку [4], состоящую, в основном, из пирамидных нейронов, а также структурные единицы гиппокампа – так называемые ламели поля CA₃, также содержащие в своем составе пирамидные нейроны (но уже гиппокампаальные).

Рассмотрим основную особенность пирамидных нейронов коры (в отличие от, например, гиппокампаальных пирамид): способность реализовать ассоциативное преобразование информационной последовательности в многомерное пространство. Покажем, как объединение таких нейронов автоматически извлекает из потока входной информации словари представлений различных уровней и различных модальностей. Покажем, как объединение таких нейронов можно использовать для структурной обработки информации – выявления связей слов словарей во входном потоке. Как на их основе формируются иерархические структуры, реализующие многомодальные представления. Такие структуры, дополненные иерархией, формирующей речевое языковое представление, реализуют многомодальную модель мира. Причем, в многомодальной модели события разных уровней поименованы словами (словосочетаниями). Покажем, как в такой иерархии возникают семантические (многомодальные – в правом полушарии и метаязыковые – в левом) представления, а на их основе, формируются прагматические представления.

Поскольку эти многомодальные и метаязыковые семантические представления почти изоморфны, анализ обработки текстовой информации, позволяет представить себе как реализуется обработка многомодальной информации. И, с другой стороны, использование визуализированных метаязыковых семантических представлений позволяет удобно эксплицировать картину мира, описанного в текстах, для ее представления пользователю, что и является конечной целью представляемой в настоящей главе технологии автоматической обработки текстовой информации.

2.4. Искусственная нейронная сеть, реализующая структурный анализ информации и ассоциативное обращение к ней

В основе представляемой нейросетевой технологии лежит ассоциативное преобразование информационной последовательности в многомерное пространство, которое обладает свойством ассоциативности обращения к информации. На этом свойстве в дальнейшем формируется структурный подход к обработке информации. Искусственная нейронная сеть, реализующая это ассоциативное преобразование, строится на основе модулей, которые можно считать моделью модулей коры, также реализующих ассоциативное преобразование информационной последовательности в многомерное пространство [5]. Эти модули коры называются макроколонками.

И.С. Беритов [4] считал, что основной функциональной единицей коры является модуль, состоящий из пирамид 3 слоя, объединенных между собой группой звездчатых клеток, который работает как единое целое под воздействием одного или нескольких афферентных волокон. Пирамиды 3 слоя коры, входящие в колонку, имеют управляющий апикальный дендрит, который добавляет фоновую импульсацию от горизонтальных клеток 1 слоя, и, тем самым, облегчает условия срабатывания нейрона. Это эквивалентно уменьшению порога нейрона. Как показали В.П. Бабминдра и Т.А. Брагина [6], во многих образованиях центральной нервной системы человека от спинного мозга до коры головного мозга имеется тип иннервации нейронов одним волокном, когда иннервирующее волокно распространяется от дистального участка дендрита к проксимальному. Там же показано существование так называемых триад – наличие между терминальными окончаниями афферентного волокна и дендритическим синапсом вставочного тормозного нейрона, что эквивалентно реализации тормозного синапса. Афференты, иннервирующие различные пирамидные нейроны колонки [6], могут объединяться электротонически, то есть входы всех входящих в структуру нейронов как бы объединены.

Далее можно показать, что в дендритах пирамидных нейронов коры реализуется временная обработка информации [7]. Известно [8], что наряду с электрически компактными нейронами, которые осуществляют функцию пространственной суммации сигналов, в центральной нервной системе имеются электрически некомпактные ней-

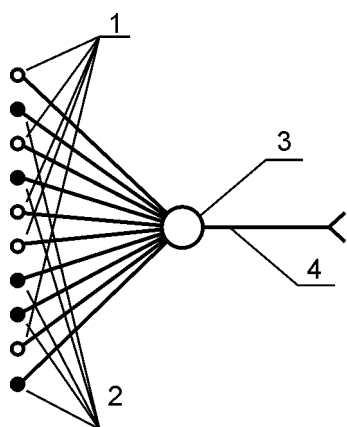


Рис. 2. Нейроподобный элемент с пространственной суммацией сигналов. Здесь 1 – возбуждающие и 2 – тормозные синапсы, 3 – тело клетки, где реализуется пороговое преобразование, 4 – выходная связь.

роны, то есть такие, время распространения потенциала по дендриту которых от места его появления на синаптическом контакте до тела клетки превышает временную константу мембраны в месте генерации спайка [8]. За счет электрической некомпактности возникает как бы задержка в распространении потенциала по дендриту в сторону тела клетки, которая тем больше, чем длиннее дендрит. Именно такие нейроны могут реализовать функцию временной суммации сигналов.

Электрически компактные нейроны моделируются нейроподобными элементами с пространственной суммацией сигналов, электрически некомпактные – с временной суммацией. Нейроподобный элемент с пространственной суммацией сигналов [9] представляет собой элемент, обладающий свойством алгебраически суммировать приходящие на него сигналы, а также находиться в состоянии покоя, или возбуждаться в случае превышения суммой порогового значения.

Добавление на входе нейрона с пространственной суммацией бинарного регистра сдвига превращает его в нейрон с временной суммацией. Его эквивалентное представление дано на рис. 3. Именно такой нейрон моделирует пирамидный нейрон 3-го слоя коры, входящий основной единицей в стандартный модуль обработки и хранения информации в коре – макроколону.

Модель нейрона с учетом временных характеристик последовательности сигналов позволяет формировать сети, обладающие свойством запоминания и воспроизведения динамической информации, а также способностью автоматического восстановления внутренней структуры входной информации (структурного анализа). Такие нейронные сети позволяют формировать поуровневые словари элементов событий, представленных во входной информации, а также восстанавливать структуру связей этих элементов. Модель нейрона с временной суммацией сигналов, помимо компонентов, имеющих в традиционном нейроне с пространственной суммацией сигналов, содержит так называемый обобщенный дендрит [8], представляющий собой цепочку суммирующих ячеек и задержек (см. рис. 3.).

Объединение всех синапсов обобщенного дендрита одним пресинаптическим афферентным волокном, которое распространяется в направлении от дистального конца дендрита к проксимальному, о чем говорилось выше, усиливает роль временной структуры последовательности импульсов (и их отсутствия) в тактированном времени. Воздействие последовательности, достигающей по афферентному волокну дендрита нейрона усиливается за счет суммирования.

Введение в состав обобщенного дендрита тормозных синапсов позволяет формировать комбинации синапсов, которые избирательно реагируют на разные временные последовательности импульсов и межимпульсных интервалов. В ближайшей к телу нейрона ячейке обобщенного дендрита накапливается максимальная сумма, если во фрагменте входной последовательности длиной n символов (где n – длина обобщенного дендрита) распределение импульсов и их отсутствия (единиц и нулей) совпадает с распределением возбуждающих и тормозных синапсов. Такой нейрон моделирует точку в n -мерном пространстве R^n , где n – длина регистра сдвига – обобщенного дендрита нейрона, а комбинация единиц и нулей является координатами точки. Нейроны с различными распределениями синапсов на дендритах можно избирательно возбуждать, подавая на них последовательности с разным распределением в них импульсов и межимпульсных интервалов.

Объединение множества из 2^n нейронов с разными адресными комбинациями:

$$b_0 = 00\dots 0;$$

$$b_1 = 00\dots 1;$$

...

$$b_{N-1} = 11\dots 1,$$

в единую структуру порождает модель n -мерного пространства R^n . В этом случае каждый отдельный нейрон моделирует одну из точек n -мерного сигнального пространства R^n . Эта структура позволяет отобразить любую последовательность A в последователь-

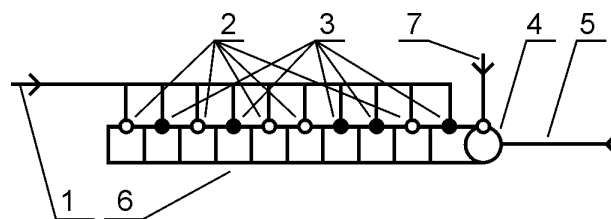


Рис. 3. Нейроподобный элемент с временной суммацией сигналов. Здесь 1 – вход, 2 – возбуждающие и 3 – тормозные синапсы, 4 – тело клетки, где реализуется пороговое преобразование, 5 – выходная связь, 6 – многоразрядный регистр сдвига, 7 – управляющая связь.

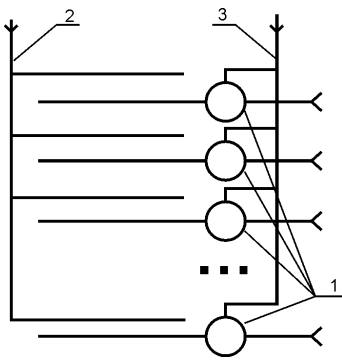


Рис. 4. Нейронный пучок. Здесь 1 – нейроны пучка, имеющие обобщенные дендриты с разными адресами от (000...0) до (111...1), 2 – общее афферентное волокно, 3 – управляющий вход.

ность точек – траекторию \dot{A} : $\dot{A} = F(A)$. Такое отображение F , является основой для реализации структурной обработки информации.

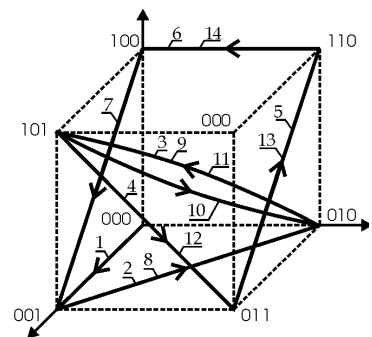
2^n нейроподобных элементов пучка с разными комбинациями синапсов (рис. 4) имеют адреса от (000...0) до (111...1). Здесь «0» соответствует тормозному синапсу, а «1» возбуждающему. Нейроны включены параллельно и каждый из них избирательно реагирует на свой n -членный фрагмент последовательности, соответствующий его адресу, поэтому любой фрагмент отобразится в один из нейроподобных элементов пучка, а вся последовательность – в последовательность сработавших нейроподобных элементов – траекторию (последовательность точек многомерного пространства).

Отображенная в многомерное пространство, моделируемое нейронным пучком, информационная последовательность может быть запомнена за счет изменения состояний нейронов, моделирующих соответствующие точки n -мерного пространства R^n , а впоследствии – воспроизведена. Для этого необходимо добавление элемента памяти к адресной части нейроподобного элемента (обобщенному дендриту).

Преобразование F обладает свойством ассоциативности обращения к точкам траектории \dot{A} ассоциацией по n -членному фрагменту последовательности A (то есть – по его содержанию): любые n символов как координаты сразу же адресуют нас к соответствующей точке траектории в многомерном пространстве.

Ассоциативность преобразования F позволяет сохранить топологию структуры преобразуемой информации: одинаковые фрагменты последовательности преобразуются в одну и ту же траекторию, разные – в разные траектории. Действительно, среди n -членных фрагментов информационной последовательности может встретиться уже ранее встречавшийся n -членный фрагмент, и траектория в этом случае пройдет через точку, уже принадлежащую ей, то есть пересечется с самой собой. Если во входной последовательности появляется фрагмент, уже встречавшийся ранее, ему соответствующий участок траектории целиком повторит ранее пройденный участок. Можно автоматически выявлять повторяющиеся фрагменты входных последовательностей, применяя пороговое преобразование к траекториям в многомерном пространстве. Участки траекторий, превышающие заданный порог по частоте встречаемости, противопоставляются участкам траекторий, встречающимся реже заданной частоты.

Действительно, если мы имеем класс последовательностей $\{A\}$, в которых в разных комбинациях встречаются подпоследовательности $\{B_i\}$ (например, буквы в словах), то, отображая последовательности класса $\{A\}$ в n -мерное пространство и применяя к ним пороговое преобразование, мы сформируем в многомерном пространстве множество фрагментов траекторий $\{\dot{B}_i\}$, соответствующее множеству последовательностей $\{B_i\}$ – словарь (букв, в предыдущем случае).



$A = (101100101011)$

Рис. 5. Пример представления траектории последовательности в n -мерном пространстве, где $n = 3$.

Сформированный словарь часто встречающихся событий может быть использован для детектирования старой информации в потоке новой. Для этого необходимо реализовать поглощение фрагментов входной последовательности A , соответствующих словам словаря, и пропускание новой, относительно словаря, содержащейся во входной последовательности, информации. При этом формируется так называемая синтаксическая последовательность или последовательность аббревиатур – C , характеризующая связи слов B словаря $\{B\}$ в последовательности A .

Таким образом, удастся устранить из входной последовательности \tilde{A} информацию, содержащуюся в словаре $\{\tilde{B}\}$. Тем самым создается предпосылка построения многоуровневой структуры для лингвистической (структурной) обработки входной информации. Синтаксическая последовательность C , содержащая только новую, по отношению к данному уровню, информацию, становится входной для следующего уровня. На следующем уровне, подобно описанному выше, из множества синтаксических последовательностей $\{C\}$ формируется словарь $\{\tilde{D}\}$ и множество синтаксических последовательностей следующего уровня $\{E\}$. То есть, мы имеем стандартный элемент многоуровневой иерархической структуры: такая обработка с выделением поуровневых словарей может происходить на всех уровнях. Словарь следующего уровня является, в этом случае, грамматикой для предыдущего уровня, так как его элементами, при соответствующем выборе размерностей пространств этих уровней, являются элементы связей слов предыдущего уровня.

Будем считать под распознаванием в терминах преобразования F процесс принятия решения о степени совпадения входной информации с ранее запомненной. Распознавание предполагает предшествовавший ему процесс обучения. В основе механизма распознавания лежит сравнение входной последовательности \tilde{A} и наиболее близкой ей, из запомненных в многомерном пространстве, последовательности A .

Рассмотрим формирование модели мира в терминах одной модальности. Рассмотрим иерархическую структуру процессов F обработки информации, имеющую вид, представленный на рис. 6, где на каждом уровне имеется ряд параллельно включенных процессов преобразования в многомерное пространство, связанных с процессами следующего уровня по типу «каждый-с-каждым».

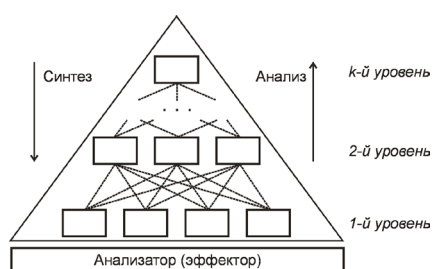


Рис. 6. Многоуровневая иерархическая структура из процессов обработки информации одной модальности, в которой на каждом уровне имеется множество параллельно включенных процессов, связанных с процессами следующего уровня по типу «каждый-с-каждым». Каждый процесс формирует подсловарь $\{\tilde{B}_i\}_{jkm}$. Здесь i – слово в подсловаре, j – номер подсловаря на уровне, k – номер уровня, а m – номер модальности.

События, происходящие в мире, отображающиеся в иерархическую структуру данной сенсорной или эффекторной модальности, формируют в этой иерархической структуре иерархию словарей, слова которых оказываются связанными между собой синтаксическими последовательностями. При этом синтаксические последовательности, сформированные в процессах предыдущего уровня, являются входными для формирования словарей следующего уровня. Эта иерархия словарей является моделью мира в терминах данной (m -ой) модальности.

Модель мира данной модальности представляет собой многократно вложенный суперграф \acute{M}_m . На верхнем (K -ом) уровне он объединяет все графы-слова \acute{B}_{ijk_1} словаря (подсловарей) верхнего уровня, а на всех более низких уровнях – включает в себя (в соответствующие места – по ассоциативному принципу) графы-слова \acute{B}_{ijk_1} словарей нижних уровней: $\acute{M}_m = \bigcup_{ij} \bigvee_{ijk < K} \acute{B}_{ijk_1}$. Здесь \bigcup – операция объединения, \bigvee – обозначает включение на свое место в слова словаря более высокого уровня. Это включение аналогично логическому сложению последовательности, соответствующей слову словаря нижнего уровня, с синтаксической последовательностью (вложению слова словаря в соответствующую купюру синтаксической последовательности).

Объединив несколько иерархических структур, соответствующих разным сенсорным и эффекторным модальностям со сформированными на них модальными моделями мира \acute{M}_m , мы получим объединенный суперграф – семантическую сеть: $\acute{M} = \bigcup_m \acute{M}_m$.

Суть структурной обработки можно продемонстрировать на процессе обработки языковой информации человеком [1, 2]. Язык, представленный в виде множества текстов, с помощью описанного выше преобразования подвергается статистическому анализу, в результате которого выявляются его словарные компоненты разных уровней. Опустим нижний (буквенный) уровень обработки информации. На следующем – морфологическом уровне – формируется словарь флективных морфем. Далее, на лексическом уровне, – словарь корневых основ слов и словосочетаний. На синтаксическом уровне формируется словарь синтаксем, представляющих собой флективную структуру синтаксических узлов с выколотыми корневыми основами. Наконец, на семантическом уровне формируется словарь попарной сочетаемости корневых основ, поскольку семантика языка описывается именно смысловой сочетаемостью [10] (которая легко преобразуется в ассоциативную сеть). В дальнейшем мы будем использовать именно эту ассоциативную сеть в качестве основного содержания понятия «семантическая сеть». Вершины ассоциативной сети раскрываются вниз словарями более низких уровней в виде иерархий, описанных выше.

Ассоциативная сеть, построенная из понятий метаязыкового представления в дальнейшем будет рассматриваться как левополушарная языковая семантическая сеть, а сеть, построенная из многомодальных образов – в качестве правополушарной образной сети. Между ними устанавливается изоморфизм, ставящий понятиям языковой сети образы многомодальной модели мира.

2.5. Искусственная нейронная сеть, реализующая перенормировку информации

В отличие от электрически некомпактных нейронов – пирамид коры, электрически компактные нейроны входят в состав другой структуры, которая реализует, в том числе, перевзвешивание входящей в нее информации. В мозге, помимо хранилища информации – коры – существует структура, ответственная за управление этим хранилищем – гиппокамп. Гиппокамп оценивает новизну поступающей в мозг информации и дает разрешение на ее запись в долговременную память [3]. Он формирует также оценку поступающей в него информации, а в результате, опосредует перенормировку весовых характеристик единиц хранения [12].

Перенормировка осуществляется в виде изменения весов связей в местах контакта возвратных коллатералей пирамид поля CA₃ с их апикальными дендритами.

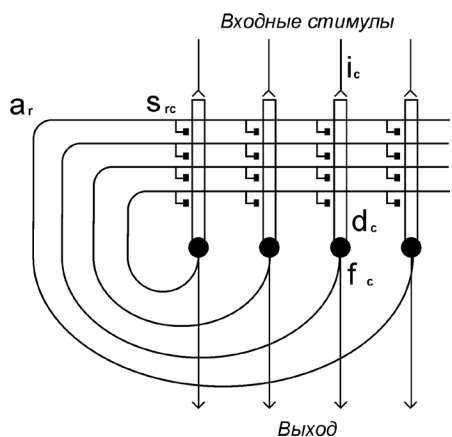


Рис. 7. Структура связей поля CA_3 . Автоассоциативная матричная память [12]. Дендриты d_c получают возвратные кооллатерали аксонов a_i , которые имеют синаптические контакты с другими нейронами популяции, модифицируемые по Хеббу s_c .

В качестве модели матрицы пирамид поля CA_3 с их дендритами, на которых оканчиваются их же возвратные колатерали, можно рассмотреть сеть Хопфилда [13].

Нейроподобные элементы в сети Хопфилда связаны по принципу каждый-с-каждым. Эта полносвязность делает сеть рекуррентной, так как выход каждого нейроподобного элемента ведет к входу всех остальных нейроподобных элементов. Рекуррентность позволяет сети переходить в устойчивые состояния при отсутствии входной информации. Если на вход сети поступает некоторый входной вектор, то сеть проходит некоторое число состояний, которые определяются весами связей ее элементов, и, в конце концов, приходит к устойчивому состоянию. Если ввести понятие функции энергии, то изменение состояний сети можно интерпретировать как сползание вниз точки,

определенной входным вектором, по склону некоторого энергетического рельефа под действием гравитации (минимизация энергии). Точка останавливается в одном из энергетических минимумов. При этом энергетический рельеф моделирует прикладную задачу (запоминаемый образ) в виде весов связей элементов.

На примере обработки текстовой информации это можно интерпретировать следующим образом: ассоциативная лингвистическая сеть текста, содержащая ключевые понятия, будучи обработанной сетью Хопфилда, меняет весовые характеристики понятий в соответствии с энергетическим рельефом текста.

2.6. Модель мира как две взаимосвязанные ассоциативные сети

Левополушарная модель мира формируется у человека под воздействием и при непосредственном участии речевого анализатора. А, значит, под воздействием и при непосредственном участии социума, так как язык используется в данном случае как средство формирования социализированной модели мира.

Наряду с социализированной моделью мира левого полушария у человека имеется строго индивидуальная модель мира в правом полушарии (при ее формировании она не испытывает прямого влияния вербальной информации). Но соответствующие элементы многомодальной модели левого полушария имеют ответные части в правом полушарии. Разница моделей заключается в том, что модель левого полушария, будучи определена вербальными эквивалентами, более социализирована и детерминирована, а модель правого полушария – индивидуальна и интегральна.

Обработка информации в правом и левом полушариях осуществляется одновременно и по схожему алгоритму, но обработка в правом полушарии работает с известной информацией, а обработка в левом полушарии включается только в случае появления незнакомой информации, когда в ход пускаются обобщенные схемы. В связи с этим, в правом полушарии формируются индивидуальные представления, описывающие конкретные ситуации. В коре левого полушария формируются похожие представления, но они усреднены по похожим объектам и сценам разных ситуаций. При этом формируются подобразы, которые объединяются в более сложные образы с помощью связей,

составляющих фреймовое представление. Подобразы отражают структуру индивидуальных объектов, являясь частями объектов или целыми объектами.

В основе образной модели лежит многомодальное представление. Представления образной модели складываются из информации различных модальностей. Представления образной модели формируются в первую очередь с опорой на кинестетическое и соматосенсорное представление, включающее в свой состав образ вертикали. Далее в образ включается зрительная и звуковая информация. На основе соматосенсорного, кинестетического, зрительного, слухового, обонятельного и вестибулярного представлений формируется многомодальное представление как объединение на верхнем уровне всех модальных информационных потоков в единое целое. Поскольку 90% информации, поступающей в мозг, имеет визуальную природу, условно можно представлять себе образное представление в виде зрительной картины.

Использование единого способа представления разномодальной информации позволяет объединить информацию нескольких модальностей в единое целое формированием многомодальной семантической сети. При этом, способ представления информации, в виде ассоциативной сети образов, един как при организации образной, так и языковой информации. Поэтому, понимание способа организации языковой сети позволит понять организацию многомодальной образной сети.

Модель мира, таким образом, представляется в виде двух взаимосвязанных сетей – иерархий – вербальной и многомодальной, каждая из которых состоит из связанных понятий-образов. Причем вербальная часть – граф-слово, соответствующее некоторому слову или морфеме, реализуется как композиция из более мелких единиц – фонем, сформированных на более низком уровне, а в фразу входят, наряду с единицами лексического уровня, и единицы синтаксического уровня. Аналогичным образом реализуется многомодальная семантическая сеть. На примере зрительного компонента многомодального представления эти понятия-образы содержат графы-слова элементов объектов, объектов и сцен (по уровням). Обе сети вместе являются связанным множеством понятий-образов: каждое из понятий вербальной сети оказывается связанным с соответствующим образом многомодальной образной сети.

В результате возможно следующее взаимодействие семантических представлений у человека. Любой вербальный материал – морфема, слово, словосочетание, фраза – имеет многомодальный эквивалент. Поэтому актуализация некоторой вербальной информации приводит к возбуждению определенного многомодального контекста, который и формирует точный смысл высказывания. При этом корневые основы слов (словосочетаний) формируют контекст в многомодальной модели мира в виде последовательности актуализированных объектов и действий (уровня представления объектов), что приводит к формированию некоторой динамики событий на уровне сцены. Фокусировка внимания на флективной информации приводит к уточнению топологии событий на сцене, а суффиксальной – к уточнению смысла. В многомодальной модели мира развертывается некоторое представление событий мира, описываемое вербальными компонентами. Правополушарная модель, актуализированная этим контекстом, накладывает на этот смысл свой индивидуальный отпечаток. И в обратную сторону: актуализация некоторой многомодальной модели (контекста) приводит к возбуждению определенных вербальных областей, эквивалентных смыслу модели (глубинно-семантической структуры), что может привести к порождению высказывания.

2.7. Образная и текстовая модели мира

В образной части модели мира (в правом полушарии у правшей) наряду с языковой зрительной компонентой сосуществует языковая зрительная компонента, представленная в виде отражения множества прочитанных текстов. Зрительная языковая (текстовая) модель мира отличается от вербальной слуховой модели мира левого полушария структурой. Вербальная языковая модель имеет парадигматическую многоуровневую структуру. В ней, как это было описано выше, события разной частоты встречаемости организованы в словари разных уровней. Элементы словарей почти равноправны и противопоставляются друг другу при выборе в зависимости от представляемого смысла (например, ехать, бежать, двигаться). Текстовая правополушарная модель имеет синтагматическую природу. События в ней сопоставлены по близости в тексте, по смыслу, по ассоциации.

Человек читающий отличается от человека не читающего наличием двух параллельно формирующихся в течение жизни и синхронно работающих моделей мира – образной и текстовой. Образная модель мира формируется у всех людей. Текстовая модель мира формируется только у людей читающих.

И та и другая модели мира имеют сетевую структуру. Образы в одной модели и понятия в другой модели связаны друг с другом в ассоциативную сеть. События, близкие по смыслу, находятся близко друг от друга. Обе модели синхронизированы: образам одной модели поставлены в соответствие понятия другой модели. Поэтому, обращение к одной модели (к ее части) вызывает актуализацию другой модели в соответствующей части.

Образная модель мира формируется в виде образов-сгущений элементов (подобразов), объединенных в единый образ [14], который, в свою очередь, может входить в образ более высокого уровня в качестве элемента.

Тексты обрабатываются правым полушарием, как зрительные образы [3]. Поэтому, текстовая модель мира формируется также в виде образов-сгущений элементов (подобразов), объединенных в единый образ, который также, в свою очередь, может входить в образ более высокого уровня в качестве элемента. Элементами могут быть события разного уровня: слова, устойчивые словосочетания, целые фразы, абзацы, тексты. Таким образом, на одном из уровней можно представить, что текстовая модель мира является сетью понятий – слов (словосочетаний), связанных между собой.

Понятия в текстовой модели не являются равнозначными. Понятия, которые связаны с большим числом других понятий, более весомы, поскольку они как бы стягивают на себя ситуацию, определяют, называют ее. Понятия текстовой модели оказываются связанными с обозначаемыми ими образами образной модели.

Формирование текстовой модели осуществляется поэтапно. Сначала формируются образы-сгущения – слова (понятия). Затем выявляются связи между понятиями. Таким образом формируется сеть понятий. Веса понятий время от времени пересчитываются, с учетом их связей с другими понятиями, при этом выявляются наиболее важные понятия.

2.8. Приложения на основе технологии. Семейство программных продуктов TextAnalyst

На основе результатов работы модуля индексации реализованы следующие функции обработки текстовой информации. Функция формирования гипертекстовой

структуры, навигации по тексту, формирования тематического дерева, реферирования текстов, автоматической кластеризации множества текстов, сравнения текстов (автоматической классификации текстов), и, наконец, функция формирования ответа на смысловой запрос пользователя – формирования тематического реферата.

После формирования семантической сети исходный текст, объединенный гиперссылками с семантической сетью, становится гипертекстовой структурой. Семантическая сеть в этом случае оказывается удобным средством навигации по тексту. Она позволяет исследовать основную структуру текста, переходя от понятия к понятию по ассоциативным связям. Пользуясь гиперссылками, пользователь может быстро найти множество предложений текста, содержащих эти понятия. С помощью тех же гиперссылок он может перейти от любого предложения непосредственно к его контексту в тексте. С этой же целью пользователь может пользоваться минимальным древовидным подграфом семантической сети – тематическим деревом. В нем оказываются иерархически представленными основные и соподчиненные понятия сети, причем понятия нижнего уровня объясняют содержание понятий более высокого уровня. Тематическим деревом также можно пользоваться для навигации по тексту – оно напоминает оглавление текста.

Семантическая сеть с числовыми характеристиками ее компонент – понятий и их связей – позволяет вычислить вес каждого предложения в тексте. Множество предложений текста, выбранных в порядке их появления в тексте, вес которых превысил некоторый пороговый уровень, можно считать рефератом текста.

Семантическая сеть исследуемого текста (или группы текстов) может быть разбита на подсети удалением из нее слабых связей. Каждая такая подсеть группируется вокруг некоторого понятия с максимальным весом в данной подсети. Это понятие обозначает тему части текста или отдельных текстов, которые оказываются сгруппированными в данной подсети. Такая автоматическая кластеризация позволяет разбить множество текстов на рубрики, а также визуализировать динамику развития этих рубрик во времени.

Используя числовые характеристики семантической сети, можно сравнивать сети двух текстов с точки зрения вычисления их пересечения (общей части). То есть можно сравнивать степень совпадения текстов по смыслу. Если в качестве одного из текстов берется целая рубрика, то имеется возможность оценить степень принадлежности исходного текста к данной рубрике, то есть автоматически классифицировать тексты.

Система для смыслового анализа текстов позволяет реализовать также смысловой поиск (сформировать тематический реферат). Функция смыслового поиска, основываясь на ассоциативном иерархическом представлении содержания информации в базе, функциях кластеризации и классификации, осуществляет выборку информации, соответствующей запросу пользователя, и структурирует ее в соответствии с близостью к запросу.

Смысловой поиск, используя ассоциации, способен выдавать пользователю информацию явно не указанную в тексте запроса, но связанную с ней по смыслу. Использование такого подхода ведет не к увеличению выдаваемой пользователю информации, а к ее тщательному отбору и анализу по главному критерию – смысловой близости к запросу.

2.9. Электронная книга

Понятие «электронная книга» (е-книга) в настоящий момент еще не вполне устоялось. Под электронной книгой сейчас понимают и просто текст книги в электронном виде, и хорошо структурированную базу данных – электронный учебник. Совершенно ясно, что чтение плоского текста с экрана – дело безнадёжное, если этот текст по объёму превышает две страницы. Формирование гипертекстовой страницы вручную – дело столь же неблагодарное. Другими словами, существует проблема подыскания подходящего инструмента для создания е-книги. Хороший инструмент должен удовлетворять определенным критериям.

Попытаемся понять основные принципы построения е-книги, а заодно и инструмента для ее создания, реализация которого позволила бы быстро создавать хорошие е-книги. Удобство гипертекстовой структуры для представления текста на экране компьютера не вызывает сомнений, по крайней мере – по сравнению с плоским текстом. Желательно расширить его автоматическим группированием материала по темам. А также – автоматической же реконструкцией тематической структуры текста. В дополнение к гипертекстовому представлению текста современные вычислительные средства предоставляют возможность его сопровождения другими мультимедийными модальностями: аудио и видео, а в последнее время – и тактильными, и соматосенсорными.

Для части читателей, по крайней мере, старшего возраста – которые учились читать по книжкам, привычное представление текста создает известный комфорт. Поэтому желательно сопровождать гипертекстовое представление обычным PDF форматом.

Кроме того, можно себе представить дополнительные возможности при создании е-книги. Такими дополнительными возможностями являются традиционное оформление книги, тестирование знаний (в случае электронного учебника), тестирование состояния пользователя. Действительно, принято иллюстрировать текст, а также сопровождать его таблицами, графиками, и просто украшать.

Современные вычислительные средства являются интерактивными. Поэтому современный учебник в электронном виде может быть активным: например, тестировать знания пользователя. Есть и другие дополнительные возможности, предоставляемые современными ИТ. Например, есть возможность тестировать физическое или психологическое состояние пользователя.

Нейросетевая технология для анализа неструктурированных текстов удовлетворяет большинству перечисленных принципов. Функциональность технологии позволяет автоматически сформировать индекс текста в виде перечня основных понятий и связей между ними. Формирует затем гипертекстовую структуру, в которой индекс является средством навигации по тексту, автоматически реферировать текст (можно также формировать реферат на заданную тему). И, наконец, позволяет кластеризовать множество текстов на рубрики.

Первый этап в создании е-книги в оболочке TextAnalyst – это формирование базы знаний, содержимое которой в дальнейшем будет представляться пользователю. Хотя обработка текста книги в оболочке TextAnalyst осуществляется автоматически – автоматически строится гипертекстовая структура текста и средство навигации по ней – тематическое дерево, требуются определенные усилия для приведения тематического дерева к наиболее удобному виду.

Так как исходный текст книги уже разбит на главы, гипертекстовая структура и тематическое дерево формируются для каждой главы в отдельности. Затем, автоматически сформированное тематическое дерево корректируется вручную: из него удаляются случайные темы, а грамматические формы слов приводятся к нужному виду. Если не пытаться добавить в структуру е-книги мультимедийной информации, дополнительного дизайна и дополнительного сервиса, можно считать, что после этого е-книга готова к использованию.

Подготовленная база готова для просмотра в оболочке **TextAnalyst**. В левом верхнем окне пользователь выбирает и открывает одну из глав книги. Затем, в том же окне, он может работать с тематическим деревом. Главная тема главы раскрывается содержащимися в ней подтемами. Каждая подтема также раскрывается вниз.

Каждой теме тематического дерева в правом окне ставится в соответствие множество предложений, содержащих данное понятие. Далее, из любого из этих предложений можно перейти непосредственно в текст книги в нижнее окно.

Такая ассоциативная навигация позволяет быстро познакомиться с содержанием книги на заданную глубину. Пороговые настройки позволяют изменять количество визуализируемого материала. При желании пользователь может получить в правом верхнем окне реферат выбранной главы, а воспользовавшись функцией смыслового поиска – получить реферат на заданную тему.

При желании, отдельные понятия тематического дерева можно снабдить ссылками на мультимедийные приложения, а также каждую ссылку в текст сопроводить параллельной страницей этой книги, например, в PDF формате.

Если е-книга представляет собой учебник, возможна организация активного тестирования знаний пользователя. Алгоритм тестирования может быть любым, например, следующим. Для выбранной темы тестируемому предлагается несколько предложений, имеющих максимальный смысловой вес в тексте. Эти предложения перемешиваются со случайно выбранными предложениями, взятыми из того же текста. Количество правильно определенных пользователем предложений по теме влияет на оценку.

2.10. Персональная библиотека

В настоящее время в лаборатории электронных образовательных ресурсов Центра общего среднего образования ФИРО на основе описанной технологии анализа текстов и с применением программного продукта **TextAnalyst SDK** разрабатывается персональная библиотека преподавателя/учащегося с ассоциативной навигацией и удаленным доступом. В ней предполагается использовать все возможности программы **TextAnalyst**, дополненные функциональностью удаленного доступа и возможностью аналитической работы с семантической сетью.

В составе библиотеки планируется наличие СУБД, текстового анализатора, графического интерфейса, средств навигации по тексту (гипертекстовой структуры). Функциональный дизайн библиотеки включает:

- сервер, в том числе:
 - неизменяемую часть библиотеки – электронный учебник (учебники), снабженный средствами формирования электронного учебника;
 - контент библиотеки, в том числе мультимедийный предполагается разрабатываемым пользователем;

- изменяемую часть – пополняемую текстовую выборку;
- средства навигации, в том числе: рубрикатор, а также ассоциативную сеть, построенную на пополняемой части с помощью текстового анализатора;
- средства автоматической классификации входных текстов;
- средства автоматической кластеризации;
- средства для построения семантических представлений, в том числе: семантической сети, а также фреймов ситуаций – шаблонов;
- графический интерфейс, в том числе: интерфейс семантических представлений и интерфейс управления библиотекой;
- web-интерфейс;
- средства организации метапоиска, в том числе: формирования сети запроса, а также средства ранжирования результатов поиска по степени близости к запросу;
- клиентская часть, в том числе:
 - средства удаленного доступа;
 - средства формирования неизменяемой части;
 - средства формирования изменяемой части;
 - средства формирования семантической сети.

Сервер

Неизменяемая часть библиотеки – собственно электронный учебник – представляет собой гипертекстовую структуру, формируемую автоматически на основе оглавления. Материал учебника раскладывается автором учебника по разделам оглавления и доступ к нему пользователя реализуется также с помощью оглавления.

Материал учебника содержит теоретическую часть, которая выдается пользователю в неизменном виде, тренировочные упражнения, которые предполагают активную работу пользователя, например, чтение, повторение, заполнение пропусков, и наконец, тестирующую часть, работа с которой также предполагает активные действия пользователя. Кроме текстового контента учебник может содержать аудио и видео приложения. Ссылки на которые назначаются и осуществляются автоматически.

Изменяемая часть библиотеки формируется пользователем. В простейшем случае это просто произвольный набор текстов. Этот набор текстов может быть обработан текстовым анализатором с целью формирования из него гипертекстового представления, в которое входят собственно тексты выборки, а также семантическая сеть этого множества текстов, которая позволяет пользователю осуществлять навигацию по текстовой выборке.

Средства навигации по текстовой выборке. Для удобства использования текстовой выборки пользователем вручную формируется рубрикатор, содержащий рубрики, по которым раскладываются тексты выборки. Сначала они раскладываются пользователем вручную. После того, как в рубриках появится некоторое количество текстов, дальнейшая раскладка текстов по рубрикам может осуществляться автоматически или полуматематически.

Если пользователь не желает формировать рубрикатор, возможны два решения. В первом случае формируется семантическая сеть всей выборки, которая в дальнейшем используется для навигации по этой выборке. Это же осуществляется в случае, если тек-

стовая выборка невелика по объему. Такая же гипертекстовая структура формируется в каждой листовой (оконечной) папке рубрикатора (на нижнем уровне). В другом случае рубрикатор (одноуровневый) формируется автоматически с помощью функции автоматической кластеризации, которая позволяет разделить текстовую выборку на подмножества в соответствии с тематической близостью текстов, входящих в подмножество. Та же функция может быть реализована полуавтоматически, когда пользователь утверждает решение, принятое системой автоматически.

Классификация входных текстов. В случае автоматической раскладки входящих в выборку текстов, для множества текстов, содержащихся в каждой рубрике, формируется семантическая сеть. Далее семантические сети рубрик используются для сравнения с семантической сетью входного текста с вычислением степени близости с сетью входного текста с сетями рубрик и принятием решения о его отнесении к одной из рубрик. При полуавтоматической классификации входного текста, принятое автоматически решение предлагается пользователю для утверждения.

Семантическая сеть – модель мира – формируется автоматически или полуавтоматически. В случае автоматического формирования, семантическая сеть текстовой выборки строится с помощью текстового анализатора. В случае полуавтоматического формирования, она может строиться вручную из отдельных звездочек. Может строиться полуавтоматически путем анализа текстов с последующим ручным редактированием полученных семантических сетей. Семантическая сеть нового текста может быть добавлена к семантической сети – модели мира. Семантические сети текстов хранятся вместе с текстами в базе, что позволяет как добавлять новые части к модели мира, так и удалять отдельные подсети из общей сети.

Помимо общей семантической сети – модели мира – и семантических сетей отдельных текстов и целых текстовых выборок, в базе могут храниться семантические сети отдельных ситуаций – фреймы. По структуре они могут отличаться от семантической сети – модели мира – включением в их состав помимо анализирующей компоненты – собственно семантической сети – также управляющей компоненты, реализующей некоторый алгоритм.

Также, в базе могут храниться так называемые семантические шаблоны: сети ситуаций, составляющих базовые знания. Эти сети могут использоваться как для расширения запроса при формировании запросов, так и для укрупнения-разукрупнения семантической сети при анализе текстов.

Инструмент для формирования семантических представлений. Семантические представления (сети и фреймы) могут формироваться вручную как комбинации звездочек, также формируемых вручную.

При формировании звездочки возможно использование шаблонов и фреймов, содержащих главное ключевое слово звездочки. При этом звездочка редактируется пользователем. При редактировании возможно не только удаление и переименование, но и добавление ассоциированных понятий.

При формировании фреймов и семантических сетей возможна их сборка из отредактированных звездочек, отредактированных шаблонов и отредактированных семантических сетей.

При формировании семантических представлений возможно использование логических функций на сетях: вычисление объединения, дополнения, пересечения.

Графический интерфейс предназначен для отображения семантических представлений, а также для отображения логических функций на семантических представлениях. Графический интерфейс также предназначен для формирования базы данных, управления базой данных, а также навигации по базе.

Web-интерфейс предназначен для работы с Интернет, в том числе, для получения информации из Интернет. Для этого используются стандартные поисковые машины.

Средства метапоиска предназначены для улучшения возможностей работы с Интернет. Если требуется сформировать расширенный запрос, формируется семантическая сеть запроса с использованием шаблонов или фрагментов семантической сети – модели мира.

Полученные из поисковой машины ссылки ранжируются по степени близости к запросу.

Клиентская часть

Удаленный доступ. Клиентская часть позволяет произвольному пользователю получить доступ к базе данных библиотеки с произвольного Web-браузера. В зависимости от ранга пользователя, он может использовать средства формирования неизменяемой части, изменяемой части, а также семантической сети.

Глава 3

Технология ручной обработки текстовой информации для извлечения и представления знаний в виде фреймовой модели

В настоящее время существует ряд направлений, в рамках которых ведется работа по разработке стандартов обобщенной модели представления знаний для управления знаниями в открытом образовании. Это стандарты **ISO, RDF, OWL и др.**, предназначенные для унификации подходов к решению проблемы единства информационного пространства в условиях электронного документооборота и наличия компьютерных систем, обеспечивающих обработку информации для различных целей унификации [1, 2]. Эти подходы регламентируют представление знаний для машинного применения, но не имеют отношения к моделям знаний для обучения, в которых заинтересованы преподаватели и учащиеся, т.к. моделирования знаний предметных областей и дисциплин есть нестандартная задача, ведь все предметные области специфичны, у них свой понятийный аппарат, аргументация, коммуникативные стратегии и др.

ФМПЗ – способ обработки знаний, привлекающий к себе внимание как информатиков, занимающихся разработкой автоматизированных обучающих систем, так и педагогов, использующих ФМПЗ эксплицитно, безмашинно. Это, в то же время, и по-разному трактуемый феномен. И в данном разделе мы даем важную для понимания проблемы ФМПЗ ретроспективу, аргументирующую его определение, а также роль и функции для предлагаемой нами технологии обучения.

Каковы общие структурные характеристики этого способа представления знаний? Предметная область описывается через фрейм- рамку (англ. frame – рамка) в виде системы понятий, отражающих сущностные признаки изучаемого объекта. Полученная система родовых понятий- языковая картина мира, состоит из концептов. Родовые понятия подвергаются концептуальному моделированию, т.е. структурируются методом деления понятия, известного из логики. Деление помогает выявить внутреннее содержание понятия, его значения или семы, которые и представляют собой вложенные понятия или концепты. Они соотносящиеся с родовыми понятиями как род-вид. Деление продолжается до тех пор, пока не достигает конца, когда уже больше невозможно его производить. Получаются единицы таксономического уровня- слова рядовые или лексемы. Деление родовых понятий помогает выявить их концептосферы. Методом деления получается вертикальный текст или гипертекст, который и является моделью знаний.

Это весьма грубый портрет явления. В качестве практического примера приведем общеизвестное: ФМПЗ реализовано эксплицитно как общепринятая система Windows-

простой способ работы с текстом, опосредованный системой искусственного интеллекта – вероятностной, понятийной, естественно-языковой, иерархической и по определению «фреймовой» моделью знаний. В данном примере ФМПЗ используется не только как модель знаний, но и как интерфейс системы, обеспечивающий быстрый доступ к знаниям.

3.1. Теоретические основы ФМПЗ

Каковы определения ФМПЗ и подходы к его реализации, применяемые различными специалистами?

Особенно интересным и важным для образования явилось то, как ФМПЗ стало использоваться в инженерии знаний и теории экспертных систем. Была предложена технология моделирования путем ручного извлечения знаний из текстового материала, также было предложено новое определение понятия «знания» как «основные закономерности предметной области, позволяющие человеку решать конкретные производственные, научные и другие задачи. Была предложена гибкая, ориентированная на человеко-машинное общение модель знаний, куда кроме самой модели были включены, факты, понятия, взаимосвязи, оценки, правила, эвристики, а также стратегии принятия решения в этой области» [3], все то, что может обеспечивать пользователю возможность решать сложную интеллектуальную задачу в общении с компьютерной системой, выступающей в качестве консультанта.

Были сформулированы требования к модели и моделированию знаний. При этом сложился ряд важных и устойчивых представлений о модели. Вот они:

1. Специалисты связали представление знаний о модели и саму модель с понятиями проблемная ситуация, задача, постановка [4].

2. Определилась структура модели фрейм – «иерархически организованная структура данных, предназначенная для представления знаний о некоторой стереотипной ситуации» [5].

3. Выделили модели фреймового типа как перспективные для разработки моделей предметных областей, для организации общения между человеком и автоматизированной системой. Понятие «структура» стало самостоятельным как универсальный способ организации знаний без содержания, где содержание стало заменяемым.

4. Выделили способы «представления» знаний – семантические сети, фрейм, граф.

5. Выдвинули понятие – «поле знаний» как полный комплекс элементов знаний, обеспечивающий правильное, аргументированное и верифицированное решение сложной задачи пользователем в диалоге с компьютером.

6. Сформулированы типы знаний – декларативные и процедурные как знания о том, «что» и знания о том, «как».

Применение ФМПЗ в данном подходе должно обеспечивать получение вывода о том, как то, что на входе в систему, или замысел, превращается в то, что на выходе – результат. Получение вывода должно получаться с помощью модели знаний. Таким образом, ФМПЗ получил определение средства решения задач.

Полуформальное представление знаний в ФМПЗ обладает уникальными дидактическими свойствами, очевидными эксплицитно. Это свойства, обеспечивающие внутреннюю наглядность ФМПЗ:

- внутренняя интерпретируемость, когда родовые понятия интерпретируются через видовые,

- структурированность, обусловленная последовательным делением и ветвлением,
- связность, непрерывность логических переходов и взаимосвязей понятий в системе,
- активность знаний, интерпретируемость их как процедурных по сумме признаков¹.

Дальнейшее развитие компьютерных технологий привело также к тому, что ФМПЗ стал использоваться и как универсальное средство нейросетевой организации и навигации информации.

В разработках компьютерных систем появился новый тип – системы data mining, системы, позволяющие извлекать из текстов ключевую для их интерпретации информацию. Это системы базирующиеся на так называемой «нейросетевой навигации» и результирующие работу в построении для пользователя моделей сжатого представления проблемной области в виде тематических или семантических деревьев – т.е. фреймов, а также рефератов и аннотаций. Появилась возможность автоматизировать работу по разработке фреймов, используя данный тип систем для извлечения тематических и семантических фреймов из текстовых корпусов, и, таким образом, быстрого автоматического создания смысловых портретов исследуемых текстов.

К системам datamining относится TextAnalyst – отечественная разработка, свойства которой и позволяют, по нашему мнению, оптимизировать и автоматизировать работу учащегося с текстовой информацией в компьютерной среде в том ключевом направлении, которое нами намечено – на основе фреймового представления знаний. Эта система берется за основу для разработки электронной оболочки ИЛДТ.

Эксплицитное представление знаний в виде ФМПЗ использовалось Крюковой О.П. [7] для организации обучения английскому языку. Разработанное ФМПЗ на основе языковой картины мира, выраженной в категориях Аристотеля, использовалось учащимися как проектировочная и ориентировочная основа подготовки учебных материалов, в том числе и для программ с открытым вводом, для организации аудиторной и самостоятельной работы, как проектировочная основа для разработки тестов, как знания для управления решением учебно-практических задач².

В настоящее время преподаватели различных дисциплин начали использовать ФМПЗ как способ открытого комплексного и концентрированного представления знаний в виде матриц, таблиц и др., применимые для управления знаниями в процессе решения учащимися различных предметных задач [8]. И это также говорит о применимости ФМПЗ для разработки учебных материалов, обладающих свойствами внутренней наглядности, предсказательным потенциалом, обеспечивающими управление знаниями.

В научной литературе по лингвистике и когнитивистике содержатся также различные описания и определения фрейма, указывающие на его связь с семантической памятью как организатора этого хранилища, и как средства извлечения из него необходимых данных. Вот высказывание лингвиста:

«Фрейм воссоздает идеальную картинку объекта или ситуации, которая служит своеобразной точкой отсчета для интерпретации непосредственно наблюдаемых, реальных ситуаций, с которыми человек имеет дело в действительности. Если обозначить фрейм как организацию хранимых в памяти человека представлений и организа-

¹ Понятия внутренней интерпретируемости, связности, структурированности, активности были выдвинуты Поспеловым Г.С. для систем искусственного интеллекта [6]. Мы обратили внимание на них в эксплицитном полуформальном использовании. – *Прим. авт.*

² О том, как ФМПЗ используется для проектирования курса, тестирования с полным покрытием дидактических целей см. [9].

цию процессов обработки и логического вывода оперирующих над этим хранилищем, то можно предположить, что фрейм представляет собой систему языковых выборов (лексических единиц, грамматических форм, категорий)» [10].

Когнитивная наука, которая длительное время занималась проблемой исследования роли различных «схем», структурных способов организации информации для различных целей изучения мышления, накопила и разные мнения и оценки их результатов. Так, с одной стороны, использование концептуальных моделей позитивно. Выдающуюся роль в их функционировании играют «относительно устойчивые, обобщенные структуры опыта, которые позволяют предвосхищать порядок развития событий, их содержание и внутреннюю связь, а также предвидеть изменения вида объектов и окружения при собственных действиях и локомоциях» [11]. Другими словами, невозможно отрицать, что концептуальные структуры или модели работают как модели вероятностные. И, более того, просто «организация материала, причем не обязательно иерархическая, может быть очень эффективной, вносящей вклад в улучшение воспроизведения» [12].

Фундаментальная роль концептов и концептуальных моделей подчеркивается в когнитивистике и в связи с методологией обучения с помощью естественных понятий, на которых ФМПЗ и строится. Э. Рош обратилась к анализу естественных понятий для выявления их роли в восприятии и действии. Она обратила внимание на некоторые, которые предпочтительнее, чем другие, являются хорошими, «базовыми» формами гештальт-психологии [13]. Работы Э. Рош повлияли на теорию семантической памяти.

Для данного исследования также важно и то, что можно использовать ФМПЗ – модель естественно-языковых понятий, как инструмент формирования семантической памяти с наполнением слотов иностранными словами, если сама модель универсальна, т.е. содержание слотов может настраиваться параллельно, в виде синонимических рядов лексики. Таким образом, модель можно настраивать на выбор языка.

Интересными и важными для нас представляются также определения фрейма как модели, отражающей типическое, стандартное. Об этом свидетельствует следующее: «Фрейм – это структура представления знаний, организованная вокруг некоторого понятия, которая в отличие от ассоциаций, содержит данные о существенном, типичном и возможном для этого понятия» [14].

В связи с этим в когнитивной психологии имеются и негативные высказывания об ФМПЗ. Так, в частности, утверждается также, что концептуальные модели носят произвольный характер, следовательно, это не научно. Укоры и критика раздаются и в связи с тем, что ФМПЗ – это попытка обуздать свободу мышления, навязывая несвободу. Это серьезные критические замечания, которые требуют осмысления.

По нашему мнению, по поводу произвольности следует сказать следующее. Концептуальные модели, являясь естественно языковыми, не могут сегодня не страдать произвольностью, особенно в тех случаях, когда в некоторых областях знаний отсутствует общепризнанный тезаурус. По-видимому, должен пройти некоторый период времени, и, конечно, должно быть приложено большое количество усилий для того, чтобы такой тезаурус мог появиться. Это вопрос согласования такого тезауруса силами специалистов. Сами же попытки строить естественно – языковую модель знаний в этом случае только обостряют проблему необходимости разработки такого тезауруса, показав наличие реальных когнитивных процессов решения задач, где появятся новые термины или будут задействованы какие-либо старые. Вопрос выработки единого тезауруса дисциплины должен быть поставлен на обсуждение в процессе разработки фреймов, их практического внедрения в обучение, в дискуссии всех заинтересованных специалистов.

Что касается свободы или несвободы, то в этой связи хочется просто подчеркнуть, что концептуальные модели служат цели передать учащемуся экспертные знания специалиста о способе решения задачи, ведь модель знаний это слепок знаний со специалиста высшей категории или же с целого ряда специалистов такого класса. Эти знания могут обогатить учащегося своим качеством, высокой аргументированностью, внутренней наглядностью. Сама работа с таким учебным материалом прививает высокую культуру мышления и показывает достойные образцы. Представляется, что о несвободе мышления не стоит заботиться. Необходимо в технологии учесть замечания о несвободе мышления и показать как новое внедряется в жизнь путем смены научной парадигмы и смены концептов языковой картины мира.

Для целей и задач данного исследования мы даем ФМПЗ следующую лингводидактическую характеристику.

Фрейм – искусственно создаваемая структура-модель метазнаний предметной области или знаний высокого уровня предметной области в виде системы абстрактных родовых понятий (они также могут называться семантической сетью или парадигмой, языковой картиной мира), отображает общее точное представление коллективного сознания о данном предмете и дает панорамное, но компактное и целостное видение сегмента миропонимания через его естественную языковую картину мира.

Последняя создается следующим путем. Каждое из понятий, входящих в парадигму, далее анализируется и моделируется методом деления понятия с целью выявления внутренней формы родового понятия, таким образом, выявляются понятия видовые. Дальнейшие этапы деления, если это требуется поставленной целью получения новых искомым знаний, достигается делением полученных понятий, переходом на уровень «подвид», и производится до тех пор, пока не будут выявлены искомые знания или рядовые единицы (единицы таксономического уровня). Это процесс построения иерархической модели (графа, концептуальной модели), целью которого является экспликация внутреннего содержания каждого родового концепта и построения полной модели знаний в виде концептосферы каждого понятия. Полная языковая картина мира и является суммой концептосфер.

Структурное представление об ФМПЗ объединяет в себе, таким образом, сумму понятий: языковую картину мира- родовые понятия, концептосферу- выявленную систему видовых понятий или слотов, и результат их раскрытия – искомой информации в виде словаря, текстов, структур, высказываний- предложений или др.

В рамках системы выделенные понятия находятся в определенных и специфических для предметной области логических отношениях: включения, причины и следствия, замены или др.

Полученный результат представляет собой древовидную или иерархически организованную структуру.

Исследования показали, что данная модель может применяться в обучении как инструмент управления этим процессом для следующих функций:

- управления работой преподавателя для точного проектирования содержания предметной области в виде моделей из естественных понятий, а. доступных не только для профессионалов, но и для учащихся, т.е. открытых знаний, б. пригодных для использования в проектировании компьютерных систем, в. пригодных для проектирования тестов с полным покрытием содержания. Это доказывает ее универсальность.

- управления работой учащихся для организации обучения творчеству – создания или синтеза нового текста,

- управления работой учащихся для ускоренного и правильного восприятия и понимания текста в процессе чтения или восприятия со слуха,
- управления знаниями для организации самостоятельной работы учащихся в плане разучивания текстовой информации.
- управления образовательным процессом, которое обеспечивает реализацию творческого, но жестко организованного поискового проблемного и одновременно проектного метода обучения в условиях смешанной – книжной и компьютерной информационной среды с множеством источников информации и многообразием форм ее представлений при активной самостоятельной и центральной роли учащегося как субъекта обучения.

При этом с помощью ФМПЗ может моделироваться:

- структура изучаемого объекта,
- способ организации деятельности,
- способ решения задачи,
- модели отношений между объектами, эволюция изучаемого объекта и др.

Подчеркнем, что данный способ моделирования является лингвистическим по своей природе. Его практическое применение связано с лингвистически трактуемым результатом.

Современные педагогические исследования базируются на идее трактовки фрейма как структурной модели мышления, а не ИИ, акцентируя ее когнитивную природу. Педагогическая информатика сделала громадный шаг в продвижении в практику компьютерного обучения идеи фрейма как стандартного решения для представления (описания) знаний для учащегося. Таким образом были показаны способы компактного, логически связанного лингвистического и естественного построения модели знаний в виде древовидной структуры (фрейма, сети, графа), которая могла бы стать той самой «схемой» Э. Канта как априорной формой познания при эксплицитном ее предъявлении человеку в процессе решения задач независимо от носителя. Это свойство фрейма как вторичного текста мы и считаем исключительно важным требованием для организации знаний для ИЛДТ.

3.2. Практическое применение ФМПЗ на материале подязыков различных учебных дисциплин

Приводимые ниже примеры были разработаны автором и апробированы в процессе обучения английскому языку в обучении различных аудиторий учащихся – школьников старших классов и студентов лингвистических, гуманитарных и технических специальностей.

3.2.1. Обучение географии в социальном аспекте

Тема обучения и исследования учащихся – современный город. Сначала постулируем систему понятий – парадигму, определяющую содержательную суть этого понятия. Отличительными признаками и понятиями, составляющими эту парадигму можно считать все то, что современный человек должен знать о своем городе, т.е. следующие концепты: политико-экономический статус города, его подсистемы истории, культурной жизни, транспорта, медицины, бизнеса, спорта, финансов и др.

Как будет выявляться концептосфера? Начнем с деления понятия: например, «транспорт». Мы будем делить это родовое понятие на виды транспорта: водный, воз-

душный, автотранспорт, железнодорожный транспорт. И, когда этот уровень деления исчерпан, начнем делить полученное на следующий подвид. Это потребует деления каждого из полученных концептов и выявления содержания их вложений. И концепт «воздушный транспорт» разделится на, возможно, гражданский и военный, и, далее, в свою очередь будет делиться на более мелкие составляющие. До какого предела? До получения единиц так называемого таксономического уровня, т.е. конкретных типов летательных аппаратов: боингов, ИЛов, ТУ и др. Разработка такой классификации потребует детального знания летательных устройств, возможно, их истории вопроса, чтобы четко отобрать. Таким образом, полученный с помощью классификации конечный продукт, – словарь единиц таксономического уровня (летательных устройств, судов, и др.) который может быть сделан руками учащихся, обеспечит для них и свое личное открытие нового в исследуемом объекте.

Но первичная классификация даст общее представление о транспортной системе и ее возможностях. А знание такой подсистемы на уровне города Казань или Воронеж потребует дополнительной работы и поисков. Тогда наполнение слотов будет ориентировано на другие цели и конкретным продуктом, полученным в результате обучения станет не только общая классификация, но и представление о современной транспортной системе применительно к реальным условиям. При таком подходе к обучению, использованию различных источников информации и, широко, Интерната, можно проводить и сопоставительные исследования городов- Москва и Нью-Йорк, изучать такие феномены как мегаполисы, например, Гонконг, сравнивать по отдельным параметрам, национальному компоненту, социальному компоненту, историческому и др. Широта мышления учащихся, их творческие возможности постижения нового и их микрооткрытия будут решающим образом определяться доступом к информации, быстротой чтения, доступом к компьютеру как средству исследования, носителю большого количества информации (иногда уже структурированной и готовой для заимствования и включения в собственный анализ).

Что может стать творческим итогом обучения? Это должны быть аналитические обзоры отдельных выделенных преподавателем тематик в рамках общей, заданной, в которых и будут отражены те творческие находки, которые сделают учащиеся. Преподаватель же должен при таком подходе продумывать наличие доступа к своей персональной библиотеке или библиотеке материалов для фонового чтения, материалов о предмете, где могут быть почерпнуты сведения, помимо Интернета. В целом же вся сеть ссылок на источники может включать систему: ИНТЕРНЕТ сайты, другие учебники, книги и журналы научного и научно- популярного характера, художественные произведения и др.

Широта и глубина охвата проблематики зависит от того, как преподаватель организует работу с коллективом учащихся в целом. Большая тема должна делиться на микро-темы и микро -коллективы учащихся – члены одного класса, должны работать в рамках своего творческого индивидуального проекта. А результат – их творческий аналитический обзор должен публиковаться в сумме с другими, на сайте школы или преподавателя, его блога или персональной библиотеки. Он, при грамотном подходе, может получить вид базы данных, квинтэссенции творческих поисков учащихся по новой тематике, опубликованной в Интернете и доступной для других коллективов учащихся, изучающих эти же темы или интересующиеся ими как дополнительными.

Инструментом для их общения могла бы стать и персональная библиотека общего доступа, организованная для комплекса школ или учебных заведений.

Это только один творческий результат работы. Второй, очевидный преподавателю, строящему фрейм реальный, идущий от существующей практики создания современных городов мира, (мы ведь исходили из фрейма гипотетического), будет уточнением его, преподавателя гипотетического фрейма, то, что он увидит на анализе различных примеров. Возможно это будет связано с такой ценностью как религия, жизнь этносов в рамках мегаполиса, языки и др. Фрейм получит свое развитие и как объективная данность, отражающая реальную практику.

3.2.2. Обучение иностранному языку (в сопоставительном аспекте на материале фонетики)

Тема – «Звуки речи». Здесь при систематизации для построения фрейма и попытки деления родового понятия «Звуки речи» открытия новых знаний для ученика могут начаться с самого начала. Дело в том, что если начать делить или структурировать содержание данного родового понятия, то может показаться на основании уже имеющихся знаний о фонетике русского языка, что система звуков в английском должна разделиться аналогично, на гласные и согласные. Именно это учащиеся чаще всего и заявляют, поскольку знают это на материале изучения родного языка. Но это не так в английском языке, где все звуки делятся не на две, а на три категории: гласные, согласные и полугласные или сонанты. Доказательством их статуса полугласных является в английском то, что они, также как и гласные, имеют в английском языке статус слогаобразующих.

Преподаватель, проводящий занятие как попытку построить фрейм – классификацию звуков речи задает подобный вопрос как провокацию для того, чтобы показать, как сам процесс деления родового понятия позволяет иной раз увидеть скрытые нетривиальные факты, если деление делается без пропусков, аккуратно, и ярко эксплицирует скрытую специфику. Получаемые учащимися знания в данном случае – это открытия индивидуального, но не коллективного сознания, поскольку вышеуказанная информация детально описана в учебниках фонетики.

Работа над фреймом позволяет корректировать неправильные представления учащихся.

Что является результатом такого подхода к организации обучения, когда производится корректировка классификации (или фрейма)? Уточнение теоретических представлений, выявление специфики казалось бы сходных ФМПЗ, демонстрация такого свойства ФМПЗ как внутренняя наглядность ФМПЗ.

3.2.3. Обучение бизнесу

Тема: «Создание производственной базы за рубежом на базе международной компании».

В этом примере мы рассмотрим,

- как с помощью ФМПЗ моделируется способ организации деловой деятельности, какие факторы определяют принятие решения,
- какой продукт получается как результат учебной деятельности,
- что с этой моделью будут делать учащиеся.

Потребность в такой модели ФМПЗ диктуется тем, что для решения задачи вложения инвестиций за рубежом бизнесмен должен знать многое о стране, куда он планирует делать инвестиции. ФМПЗ, представляющий структуру необходимого для принятия решения знания, и дается в модели № 1.

Модель № 1. Создание производственной базы за рубежом на базе международной компании.

1. Производимый продукт
 - 1.1. технические параметры,
 - 1.2. функции,
 - 1.3. национальные стандарты.
2. Общая оценка экономического потенциала страны – потенциального объекта инвестирования
 - 2.1. население
 - 2.2. объем национального валового продукта
 - 2.3. национальная стратегия в отношении зарубежного капитала
3. Законодательная база
 - 3.1. возможности получения лицензий на совместное производство
 - 3.2. условия организации совместной производственной деятельности
 - 3.3. налоги на импортируемую продукцию
 - 3.4. возможности вывоза заработанного капитала
 - 3.5. требования регистрации в государственных органах
 - 3.6. наличие очереди на регистрацию в государственных органах
4. Политическая обстановка
 - 4.1. демократичность или избранность существующего режима законным путем
 - 4.2. стабильность власти
 - 4.3. планы и программы развития экономики страны
5. Уровень развития данной отрасли
 - 5.1. конкуренция
 - 5.2. степень присутствия данного продукта
 - 5.3. уровень потребления в динамике
 - 5.4. объем импорта
6. наличие местного делового контакта
 - 6.1. старый партнер
 - 6.1.1. уровень продаж
 - 6.1.2. уровень потребностей дополнительного инвестирования производства
 - 6.2. новый партнер
 - 6.2.1. успешный опыт работы на данном рынке
 - 6.2.2. наличие капитала для инвестиций в новое совместное дело

Содержание работы учащихся следующее. Коллектив учащихся (старшеклассники) делится на несколько групп экспертов по экономике, которые изучив модель, начинают наполнять слоты информацией на материале различных стран, потенциально привлекательных для инвестиций. Наполнением слотов станет ключевая, определяемая в рамках фрейма, информация о ситуации по разным странам, собранная учащимися из разных источников. Далее, эта информация будет доступна для участников ролевой игры – будущих бизнесменов, которые и будут выбирать страну для своих инвестиций на основании полученной информации.

Работа над этой задачей может соединить несколько групп учащихся, обучающихся и живущих в разных странах. Там они и будут собирать материал для наполнения этой модели содержанием, доступным им в стране проживания. Возможно при

этом, что языком общения будет английский. Далее этот материал должен собираться на сайте для просмотра и публикации в единой структуре. Это вариант организации дистанционного обучения.

Как творческий результат обучения на материале данного примера можно сказать, что может быть получена информация в заданной структуре но на материале различных стран, которая может быть ценной для бизнесменов, решающих аналогичные задачи.

Как делалась данная модель знаний? Были использованы материалы американского учебника [15]. В этом учебнике авторы ставили целью объяснить учащемуся, как делается успешный бизнес, разобрать интересные случаи такого бизнеса, описать их наиболее полно и системно. Фрейм можно было извлекать из этих текстов просто потому, что они уже сами по себе отвечали на основные вопросы, содержали в развернутой форме то, что требовалось свернуть в виде модели ФМПЗ. Это и было сделано.

Управляющая роль ФМПЗ в данном случае следующая. На конкретном примере в самом учебном материале авторы учебника показывают, что даже наличие комплекса определенных негативных факторов в экономической системе страны, может быть перевешено одним субъективным- наличием партнера – жителя этой страны, способного преодолеть влияние негативных факторов, зная систему и действуя изнутри по правилам данной системы и культуры.

Таким образом, можно сказать, что изучение такого материала и представления в целом полезно как повышающее степень осознания веса концептов как факторов риска, но это не алгоритм принятия решения.

3.2.4. Обучение ботанике

Тема «Жизнь растений». Цель работы – использование ФМПЗ, которое является стандартным форматом описания сущностных признаков растения для создания базы данных с детальным описанием определенных классов растений, которые представляют интерес в данном случае.

Организация работы. Учащиеся средней/старшей школы работают с материалами книги «Иллюстрированная энциклопедия растений» чешского автора Ф.А. Новака, издательство Артия, Прага, 1982 г., крупнейшего специалиста в области систематики растений, которое содержит детально иллюстрированное энциклопедическое описание всего растительного мира от низшего уровня растений- бактерий, до высшего уровня и самых совершенных форм. (Разумеется, возможно использовать любые материалы из любых источников). Основные признаки ФМПЗ преподаватель может определить сам, мы же в качестве примера предлагаем наиболее общие концепты этого объекта и модели:

- Название (тип, семейство по классификации)
- Внешняя форма и структура
- Условия жизни (температура, влажность, освещенность, почва, вредители)
- Размножение
- Уход (для культурных), обеспечивающий эффективный рост и цветение/плодоношение.

В качестве задач, решаемых учащимися может быть предложено следующее. Учащиеся, наполняют его слоты. Это они делают, извлекая информацию о конкретных видах растений из указанной энциклопедии. При этом каждый ученик или каж-

дый микроколлектив может иметь свою задачу. Полученный в результате извлеченный и систематизированный материал базы данных может использоваться для анализа и сопоставления, получения новых знаний о специфике жизни растений и составления аналитических докладов. Управляющая роль ФМПЗ в этой работе очевидна, т.к. учащиеся создают материал базы данных по стандартной структуре, заданной самим фреймом.

Чем будут наполнены слоты такого фрейма и сама база? Это будут высказывания, факты по каждому слоту (о температуре, составе почвы, вредителях и мероприятиях и др.), которая в сумме должна оптимально объяснить как реализуется и обеспечивается жизнедеятельность растения.

Более сложный вид задания, который может быть предложен старшим школьникам – это самостоятельное создание фрейма. Это возможно при работе с энциклопедическими текстами, т.к. их структура определяется типическим видением исследуемого предмета, выработанным в рамках данной науки. И концепты фрейма могут быть извлечены из анализа тематической структуры самого текста.

Еще более сложный вид работы для учащихся – ботаников – углубление содержания слотов или даже изменение самого фрейма в результате происходящего в науке сдвига парадигмы. Такую ситуацию можно создать следующим образом. В исследование дополнительной литературы включаются те источники, в которых содержатся научные споры и новые для данной науки понятия. Так, например, читая статью «На языке «Интеркорнета» и других наречиях» в журнале «Знание – сила», № 4, апрель, 2008 год [16], учащиеся смогут увидеть новую трактовку фрейма растения, т.к. в статье появляются категории, которые сближают представления о растениях с животным миром. На основании новых данных биологии и ботаники ученые делают выводы о том, что растения наделены почти теми же чувствами, что и животные, корни растений образуют громадную динамическую сеть связи, похожую на нервные волокна мозга в километры длиной, ряд ученых утверждает, что растения активно общаются друг с другом при помощи медиаторов и т.д. [16].

Анализ данного текста может показать, что появление новых нехарактерных для фрейма растения концептов (интеллект, поведение, коммуникация и др.) и нового наполнения их слотов говорит о том, что в этой области происходит сдвиг парадигмы. Учащимся предлагается проанализировать направление этого сдвига, отредактировать фрейм в соответствии с новыми данными науки, проанализировать «за» и «против» в научном споре. Такие задания приобщают учащихся к открытиям на уровне коллективного сознания, придают их учебной деятельности особую значимость, учат мыслить категориями современной науки.

Если мы рассматриваем фрейм как рациональный способ фиксации знания, безусловно, истинный на каком – то этапе развития науки, но требующий изменения на другом, попытка создать новый, измененный фрейм – это своего рода попытка поучаствовать в научной жизни ученого мира и создать своими силами экспликацию знаний проблемной области, в которой мы наблюдаем этот сдвиг. Это, на наш взгляд, возможное приобщение учащихся к творчеству и его проявление.

Аналогичные примеры и способы организации работы с ФМПЗ можно организовать, исследуя жизнь животных, насекомых. В научной и научно – популярной литературе сегодня много фактов, свидетельствующих о новых глубинах проникновения в жизнь живого и новых феноменах этой жизни, кардинально меняющих научный взгляд на вещи.

3.2.5. Обучение литературе

Тема «Литературные жанры – басня».

Задача преподавателя – использовать ФМПЗ в данном случае как способ моделирования поведения людей на материале поведения животных. Здесь ниже представлен пример фрейма, построенного на материале басни Эзопа «Лев и мышь». Модель развития событий представлена на рис. 8. Тело басни типа Проступок.

Модель эксплицирует последовательность событий, которая, будучи абстрагирована от исходной ситуации, дает возможность переосмысления и приложения к другой ситуации и другим героям. Здесь мы приводим пример басни, на основе модели которой и компьютер и учащийся могли сочинять новые басни, заменяя содержание слов, но не меняя модели отношений между действующими лицами.

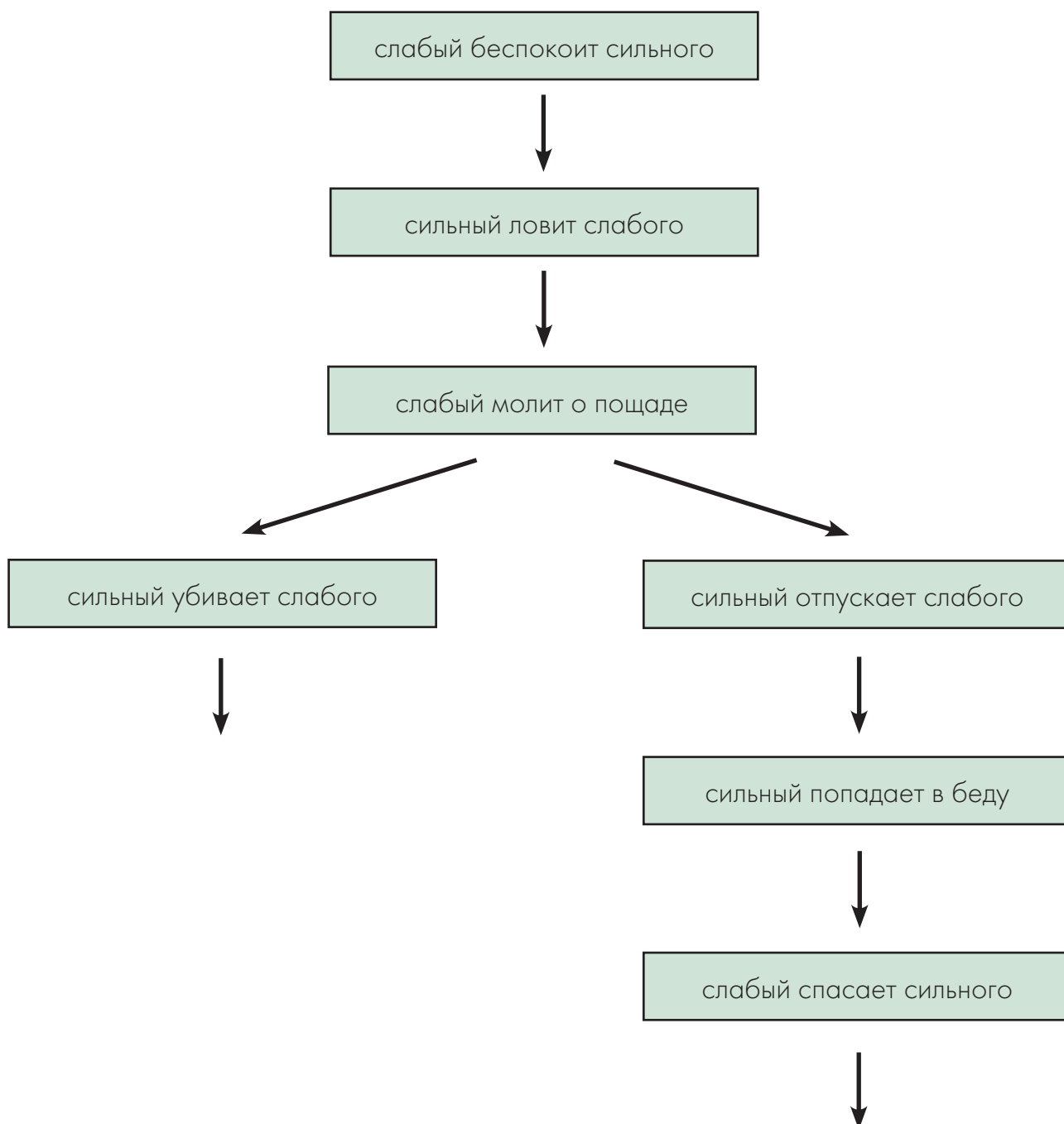


Рис. 8. «Тело басни типа – проступок»

Общепризнано, что басня является тем жанром, на материале которого можно изучать модели человеческих отношений. ФМПЗ наглядно это демонстрирует, такого представления достаточно, как показала наша практика, чтобы учащиеся стали сочинять текст. Творческим результатом изучения басен на основе ФМПЗ может быть сочинение аналогичных текстов, возможно даже и не по шаблону фрейма, а и с неожиданными поворотами событий.

Наличие ФМПЗ облегчает процесс сочинения, что и было подтверждено экспериментами (компьютерным и аудиторным). Аудиторный эксперимент также показал, что облегчается процесс чтения на иностранном языке. Поскольку модель представляет собой идеальный репрезентант смысла, который мгновенно воспринимается и обеспечивает успешную прогнозную работу мышления на аналогичном тексте.

Как делается подобный фрейм и кто может быть его автором? Практика показала, что этой работе можно научить учащихся старших классов или студентов. Это творческая работа, которая дает результат в развитии навыков владения как родным, так и иностранными языками.

3.2.6. Обучение грамматике английского языка

Автором данного исследования в настоящее время проводится эксперимент по обучению английской грамматике студентов технического ВУЗа (начальный этап обучения). В этом эксперименте ФМПЗ используется проектирования содержания – знаний, учебных материалов и тестов. Первоначально было разработано представление знаний в виде графов, которое и планировалась использовать. Практическая работа со студентами показала, что имеющееся представление нужно дополнить, включив следующие компоненты:

- Представление знаний языковой картины мира – входное, т.е. первично предъявляемое, в виде понятийной гештальт- картинкой системы концептов с количеством элементов в пределах 7+ – 2.
- Картинки – гештальт-концептосферы – внутренней формы концептов- метапонятий, также минимально- достаточной.
- Раскрытое содержания слотов в виде лексики, формул для описания грамматических структур, примеров, вербально изложенных правил, развернутого текстового описания грамматической системы – предмета изучения, развернутого текстового описания системы работы учащихся. (См. примеры на рис. №№ 9–11).



Рис. 9. Языковая картина мира (Английская грамматика - Морфология)



Рис. 10. Концептосфера «Качество»

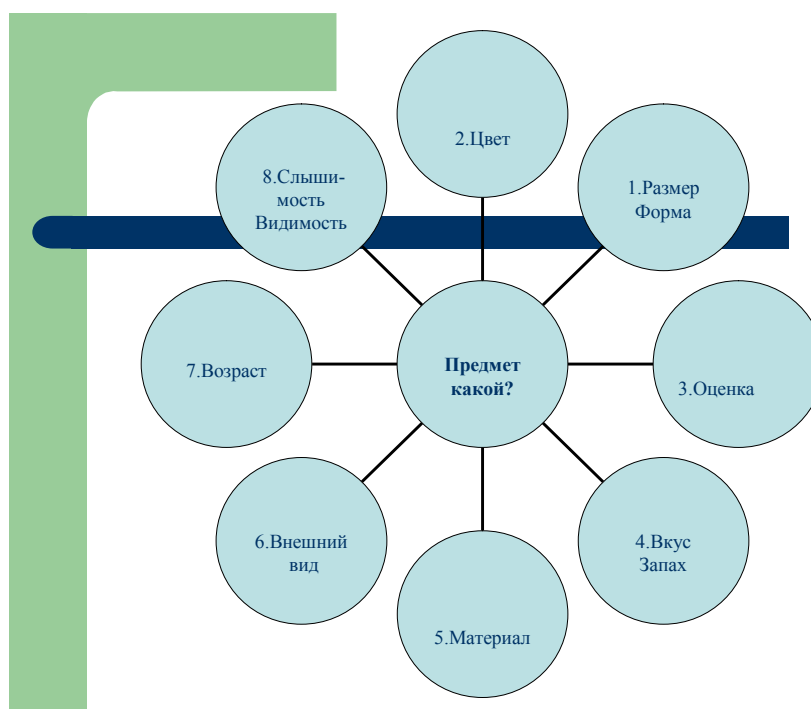


Рис. 11. Лексическая концептосфера «Качество»

Получилось развернутое представление, которое стало работать как ЭКС по принципам, изложенным в книге Гавриловой и Червинской, на которую мы ссылались выше, где изложены основы организации знаний для обеспечения самостоятельной работы пользователя в принятии решений.

3.2.7. Развитие коммуникативной компетенции учащихся на базе создания делового документа

Какую роль играют знания в анализе и синтезе текста учащимся? Мы рассмотрим этот вопрос на примере педагогического эксперимента работы со школьниками над созданием делового письма с помощью ЭКС «Деловая переписка». Использование этого материала в качестве экспериментального продиктовано тем, что лингвистическое обеспечение этой системы является работой автора исследования, как и его экспериментальная апробация.

Разработанное лингвистическое обеспечение представляет собой подход к трактовке лингвокомпозиционных знаний. Применение лингвистических знаний в данном подходе демонстрирует тип управления, который в методике обучения языкам называется «мягким». Учащийся получает возможность видеть модели знаний, эталоны текста. Уточним понимание этого термина применительно к условиям компьютерного применения: принятие решений учащегося в этом случае знаниями не детерминируется, поскольку способ решения задачи не алгоритмизируется.

Напомним, что система предназначена для консультирования пользователя по решению его проблемы. Она осуществляет поиск нужной пользователю информации – находит тип письма. Демонстрирует набор текстов-заготовок, фразеологический словарь, ранжированный по коммуникативным функциям, входящим в данный тип письма. Учащийся производит поиск нужного ему документа через иерархическую систему меню. Система предоставляет систематизированный материал для создания письма на ЭВМ методами компьютерной сборки, текстового редактирования. При использовании дополнительных приложений в системе возможны способы автоматической коррекции.

Система открыта для перенаполнения или наполнения любым содержанием. Допускает подключение других ПС и превращается в лабораторный учебный комплекс. Позволяет учащемуся решать лингвистические задачи, превосходящие по уровню его компетенцию. Ученик может решить проблему, т.е. получить консультацию и перейти в режим тренировки, если это требуется, не выходя из данной среды.

Информационное обеспечение включает:

- образцы, макеты и планы деловых и личных писем, фразеологические клише;
- правила правописания и оформления писем;
- средства поиска, сборки и редактирования писем.

Тематика писем разделена на одиннадцать классов:

- Реклама,
- Платежи,
- Организация поездок,
- Деловые контакты,
- Заказ,
- Поставки,
- Трудоустройство,
- Кредиты,
- Претензии и рекламации,
- Личные контакты,
- Цены.

Каждый класс включает подмножество различных видов писем, используемых в самых типичных ситуациях.

Например класс «Деловые контакты» включает следующие виды писем:

- Предложение сотрудничества,
- Письмо секретаря,
- Подтверждение договоренности,
- Отписка.

Каждый вид письма представлен своим подмножеством текстов. В комплект входят:

- справка об условиях использования;
- план содержания;
- типичная для вида фразеология;
- текст образца;
- макет письма.

Что может делать с этой базой знаний ученик – пользователь? Вот данные по результатам проведенного эксперимента, в состав испытуемых которого вошло около 50 человек. Ученик может найти необходимое письмо, вернее тип письма, соответствующий заданной ситуации (направление его поиска может быть задано учителем в процессе ролевой игры), изучить его структуру, языковые клише, функциональный план и деловые отношения. Используя имеющиеся заготовки, ученик может создать свое письмо методами компьютерного конструирования. Какие навыки ученик приобретает, чему учится?

Основной целью обучения деловому английскому языку может быть освоение практических навыков письма и чтения на английском языке по деловой проблематике:

1. Прочитать и понять текст делового характера, используя информационную базу курса;

2. Составить и правильно оформить текст письма делового содержания;
3. Проанализировать текст делового характера, не содержащийся в системе:
 - определить тип текста;
 - вычленить смысловые блоки и составить план;
 - выделить языковые клише и фразеологизмы;
 - выявить типологическое, сделать макет текста;
 - создать другой текст, используя составленные заготовки.

Работа учащегося с этой системой предусматривает не только создание текста с помощью информационной базы системы, но и просмотровое чтение (анализ и понимание) большого количества текстовой информации в процессе подготовки и поиска необходимой информации.

Задачей учащегося в работе с системой является структурирование письма методом компьютерной компиляции, о котором речь шла выше, на основе задачи, описанной как жизненная реальная ситуация. В задаче дается практический план содержания – Уведомите X. о поставке товара на сумму или Запросите Y об упаковках и маркировке товара и т.д. Учащийся сам должен ориентироваться в материалах системы, сравнивать имеющиеся эталоны со своей задачей.

Эксперимент показал, что деловое письмо с помощью информационной базы, описанной выше, могло быть составлено учащимся качественно, даже при полугодовом опыте в изучении английского языка.

За счет чего получался положительный результат в ситуации, когда учащиеся практически не владели английским языком как средством общения (напомним, что это были учащиеся с полугодом обучения английскому языку). Наблюдения показали, что учащиеся хорошо ориентировались в материале системы, что и являлось основным условием успеха в получении качественного результата. Предпосылками хорошей ориентации была проработка материала в системе и разработка специального русскоязычного интерфейса, когда системы основных понятий, подлежащих изучению и раскрытию, была дана на русском языке, т.е. чтобы дойти в процессе поиска до текста, учащемуся надо было оперировать русскими словами.

Другой предпосылкой для успешной работы мы считаем то, что учащиеся хорошо владели не английским, а французским языком, который был у них первым иностранным. А лингвистическая словарная общность этих двух языков вряд ли требует комментария. Таким образом, второй важнейшей предпосылкой мы считаем лингвистическую компетентность учащихся.

Третьей предпосылкой является то, что деловое письмо является достаточно жестко стандартизованным типом текста, где варианты могут значительно не расходиться заданными эталонами. Их степень расхождения – искусство постановки задачи преподавателем, который также и полностью владеет знаниями о содержании системы. Преподаватель может задавать степень расхождения с эталоном, выбирая и ставя задачи стандартные или требующие дополнительной вставки содержательных фрагментов в текст. Во многих случаях решение задачи складывается из собирания кусков из различных типов текста – уведомить о прибытии господина X, договориться о гостинице, дать дополнительную информацию о товаре и т.д.

Четвертой предпосылкой успеха был интерес учащихся к проблеме, их личная мотивация.

Хорошая ориентация проявлялась в понимании текстов при чтении, быстром схватывании их содержания, а также и в структурировании и грамотном подборе необходимой информации. Использование информационной базы при этом варьировалось: при чтении новых и сложных типов текста необходимо было для обеспечения их быстрого и качественного понимания демонстрировать экстралингвистическую, логическую информацию, все возможное о структуре самой деятельности вначале, и через это доходить до конкретного языкового материала. В условиях самостоятельной работы учащиеся сами регулировали получение того, что им было необходимо в нужной последовательности.

Эксперимент показал, что качество деятельности учащегося – скорость, безошибочность лингвистической работы зависит как для продукции, так и для рецепции материала от одинакового набора информации и от качества ее подготовки – качества моделирования любой информации, что позволяет концентрированно давать большие ее объемы, сводя их к таблицам, картам, деревьям и т.д.

Глава 4

Интегрированная лингводидактическая технология. Концепция

Понимание процесса построения нового информационного общества с управлением знаниями не может не включать в себя идеи реформирования образования, идеи формирования и воспитания человека, начиная уже со школы, где закладываются основы организации интеллектуального труда: его приемы и методы, модель реализации, даются инструменты и технологии применения ИТ и КТ. Эти основы должны помочь в дальнейшем сформировать и воспитать человека – специалиста, члена гражданского общества будущего, обладающего мировоззрением, знаниями, компетенциями, навыками, которые и обеспечивали ему в дальнейшем возможности самоусовершенствования и приобщения к творческому труду.

Определение целей и путей реформирования сегодня связано с трактовкой и пониманием сути происходящего кризиса в аспекте процессов технологизации и информатизации.

Кризис в системе образования сегодня происходит в связи с тем, что быстрый процесс технологизации образования не сопровождается пока адекватным педагогическим процессом эффективного встраивания ИТ и КТ в образовательный процесс.

Проблема технологизации и информатизации образования сложна тем, что бурно развивается информационно-ресурсная база: 1. такие феномены как ИНТЕРНЕТ – громадная постоянно стихийно и планомерно пополняемая область электронных текстовых и мультимедиа ресурсов, 2. в практику работы с текстовой информацией, и, соответственно, в образование, постоянно привносятся новые информационные технологии и программные продукты, требующие осмысления и интеграции, 3) информация поступает на различных языках, что ставит проблему преодоления языковых барьеров и подготовки учащихся, а также и самого учителя к полиязыковой деятельности 4) быстро меняется и сама идея учебника, который на сегодня стал комбинацией книги, интернет ресурсов, программного обеспечения, автоматизирующего работу по дисциплине.

Все происходящее – процесс бурного выброса информации на образовательный рынок, что дестабилизирует образовательную среду, требует поиска решений, обеспечивающих с одной стороны, интеграцию поступающей информации, а с другой понимания тех системных фильтров, которые направляли бы эту информацию в технологически – приемлемые концепции учебной деятельности, стабилизирующие образовательный процесс, снимающий с учителя ответственность за принятие решений в организации труда, которые не должны входить в его компетенцию, и обеспечивающие

гарантированно положительный результат в обучении. Это педагогические технологии, которые определяются нами как – рациональные способы организации учебной деятельности в компьютерной среде, пригодные для массового, промышленного применения и гарантированно дающие результат успешной реализации. Идеи таких технологий должны отвечать требованиям универсальности и междисциплинарности, решать проблемы простыми, доступными для массового потребителя способами, отвечать требованиям современного научного понимания когнитивного процесса, реализуемого как процесс человеко-машинного интеллектуального труда с возможностями его автоматизации и управления знаниями.

Мы выдвигаем идею такой технологии – это ИЛДТ.

ИЛДТ представляет собой рациональный педагогический междисциплинарный способ организации творческого обучения в компьютерной среде при управлении открытыми концентрированными знаниями в виде ФМПЗ.

Цель ИЛДТ сделать предметом обучения и объектом изучения естественный язык как знаковую систему, в которой в виде текстов фиксируется человеческий социальный опыт реализации деятельности, опыт социальной, культурно и языково-обусловленной коммуникации, сопутствующей этой деятельности, генерируется ноосфера – текстовая среда, продукция как процесс и результат коллективного осмысления и формирования коллективного научно- обоснованного сознания в предметном и проблемном поле данной деятельности.

4.1. Интеллектуальный уровень

ИЛДТ строится на основе фреймового представления знаний (ФМПЗ), которое является вероятностной моделью, обеспечивающей решение задач определенного класса, для чего модель и строится. ФМПЗ определяется как естественно- языковая иерархическая понятийная модель, верхний уровень которой – метазнания или прагматический уровень, является парадигматической языковой картиной мира конкретной предметной области. Каждое понятие – член парадигмы, далее подвергается структурированию или делению для выявления внутренней структуры понятия. Полученная концептосфера на конце деления имеет понятия- слоты, которые и содержат искомый для решения задачи материал- лексику, структуры, примеры высказываний и др.

Роль естественного языка рассматривается не только в связи с традиционными подходами и определениями его как средства мышления и коммуникации. Определение естественного языка в когнитивной лингвистике дает такое его определение как естественного средства накопления, хранения знаний в языке, язык сам может использоваться как средство построения формальных моделей для анализа самого себя. Для этих целей язык организован как иерархическая (гиперо-гипонимическая) сеть, имеющая свойства деривации, она ветвится по аналогии с нейронной сетью мозга. Использование этих свойств естественного языка для искусственного создания языковых моделей знаний предметных областей по аналогии с уже имеющимися компьютерными системами с нейросетевой навигацией и является необходимой идеей для обучения, заложенной в ФМПЗ.

Использование естественно-языкового моделирования рассматривается рациональным для решения двух взаимосвязанных задач:

1. для представления их в компактном, логически связанном виде, облегчающем восприятие и усвоение учащегося.

2. с другой стороны, доказана возможность применения естественно – языковых моделей знаний для их формального представления и использования в машинном интеллекте.

Сочетание этих подходов делает возможным организацию и реализацию обучения как человеко-машинного интеллектуального труда на единой основе представления знаний. Это представление должно быть открытым, чтобы обеспечивать не только автоматизированное скрытое для человека их использование, но и осознанное использование для управления процессами формирования памяти, решения учебно-практических задач, обеспечения творчества.

4.2. Теоретико-методический уровень

Данный аспект ИЛДТ учитывает то определение когнитивной деятельности человека, которое содержится в одной из современных гипотез о нем: человек рассматривается как ограниченный канал связи с ограниченными возможностями восприятия и памяти [Величковский 2007]. Исходя из этого основного положения и с учетом возрастающих объемов информации, которую необходимо человеку обрабатывать, для ИЛДТ требуется уточненное определение ФМПЗ. Определение ФМПЗ, на котором строится ИЛДТ следующее:

– фрейм – это структура – идеализированная картинка – гештальт древовидной структуры обработки знаний, которая уже в минимально свернутом запакованном виде (5 + – 2) обеспечивает большие и разнообразные возможности ее распаковывания, и при различном ее предъявлении имеет разные способы прочтения;

– эта структура, будучи логически связанной базовым отношением «род-вид», может распаковываться и запаковываться, создавая условия для различных форм ее методического использования;

– ФМПЗ – это способ концептуального моделирования и, если ФМПЗ строится преподавателем на основе онтологических семантических понятий, которые должны использоваться на уровне метазнаний и языковой картины мира, то учебное распаковывание ФМПЗ может войти в резонанс естественным процессом мышления с использованием тех же онтологических и семантических понятий и того же лингвистического ресурса- семантической памяти. В таком случае должно наступать такое состояние как «понимание», и семантическая память становится доступной для работы уже над учебным материалом как своим.

– понимание определяется как «вызванное внешними или внутренними воздействиями специфическое состояние сознания, фиксируемое субъектом как уверенность в адекватности воссозданных представлений и содержания воздействий. Без понимания как особого состояния субъекта невозможно продолжение общения, координации действий и какие – либо другие осмысленные действия и воздействия» [17],

– семантическая память определяется как «субъективный тезаурус, то организованное знание, которым обладает субъект о словах и других ...символах, об их значениях, об отношениях между ними и о правилах, формулах и алгоритмах, используемых для манипулирования этими символами, понятиями и их отношениями» [18].

ФМПЗ – это один из элементов в комплексе знаний, который может обеспечивать самостоятельную работу учащегося для творческой деятельности или такого уровня самостоятельной работы, который превышал бы уровень собственной компетенции самостоятельного пользователя. Для реализации концентрированного обеспечения само-

стоятельной работы учащегося необходимо также включить в комплекс концентрированных знаний такие элементы как: терминологическую лексику, правила ее употребления, сочетаемости, тематические и семантические выборки лексики, правила создания текста и лингвистическое обеспечение этого процесса, (см. также раздел 2.6.) и др.

Работа с текстовой информацией включает, таким образом, определенный и специфический комплекс операций, который может быть автоматизирован. См. далее Технологический аспект ИЛДТ и педагогический сценарий работы с ИПЭБ.

4.3. Технологический уровень

Автоматизация способов работы над текстом делится на два подраздела.

1) автоматизация тренировки и тестирования – текстовые манипуляции, которая предусматривает:

– автоматизацию разучивания лексики и разработку программного обеспечения типа «Мастер сортировки» для сортировки лексики по категориям.

2) автоматизацию операции построения линейных цепочек словосочетаний. Она предусматривает использование программного обеспечения типа «Мастер комбинаторики», которое генерирует слова для построения цепочек или же текстовые блоки (абзацы) для их упорядочивания.

3) автоматизацию работы над смыслом текста:

– автоматическое создание смыслового или тематического портрета текста или корпуса текстов из его ключевых слов,

– использование структурированных таблиц с концентрированными знаниями, которые предназначаются для предварительного предъявления в качестве управления знаниями в процессе чтения, аудирования, или других РД.

Важные технологические решения: реализация предъявления теоретического материала имеет концентрированное предъявление модели из ключевых слов, содержательное раскрытие которых в виде текстовых описаний, иллюстраций и др., реализуется в виде гиперссылок на документы – комментарии.

Важно также и то, что тренировочные упражнения репродуктивного характера реализуются в игровом режиме с помощью специальных программ – текстовых манипуляторов. Их применение позволяет сделать интерактивной работу по разучиванию учебного материала с реализацией пооперационного контроля. При этом значительно увеличивается общий объем работы учащегося, который проходит в интерактивном режиме.

Организация творческих видов работы связана с организацией видов деятельности, обеспечиваемых поддержкой специальных средств автоматического анализа текста или баз лингвистических данных и знаний, которые позволяют создавать аналитическую базу для самых различных видов работы с текстом по его распознаванию, переводу, конструированию. Это работа по подготовке самых различных текстовых документов, неотъемлемая часть работы специалиста.

Примером таких средств является **Text – Analyst**, возможности которого излагаются ниже.

Формирование ассоциативной сети. В основе обработки и представления информации в системе **TextAnalyst** лежит функция автоматического извлечения внутренней структуры анализируемого текста. Из текста автоматически извлекаются основные понятия и связи между ними. И те, и другие – с их числовыми характеристиками. На

основе этой информации строится ассоциативная сеть наподобие тех, что рисуют психолингвисты иллюстрируя структуру семантических представлений у человека. Визуализированная ассоциативная сеть позволяет познакомиться с основным содержанием предметной области, охватываемой текстом. Сочетания понятий, соединенных связью, вызывают у пользователя совершенно конкретные ассоциации, так как именно сочетания слов содержат основную информацию о модели мира. Таким образом, ассоциативная сеть – структурный портрет текста – вводит пользователя в предметную область, описываемую текстом.

Тематическое дерево. Если из ассоциативной сети извлечь минимальный древовидный подграф и упорядочить его вершины по весам, получится так называемое тематическое дерево. Это другой способ познакомиться с содержанием текста. Более привычный читателю. Это напоминает оглавление, в котором верхние вершины обозначают главы, более низкие – параграфы, и т.д. Тематическое дерево тоже позволяет пользователю представить себе предметную область, которая описывается текстом.

Гипертекстовая структура и средство навигации по ней. Поскольку мы имеем исходный текст и некоторое средство, сжато описывающее этот текст, мы можем, установив между одинаковыми понятиями гиперссылки, автоматически сформировать гипертекстовую структуру. На этот инструмент мы можем навешивать дополнительные бантики, например, можем ставить в соответствие каждому понятию сети множество предложений, их содержащих. Или каждой паре понятий. Это уже существенно более мощный инструмент. Используя гипертекстовую структуру, пользователь может знакомиться с текстом, обращаясь к контексту всякий раз, когда требуется более детальное представление о мысли автора. При этом ассоциативная сеть или тематическое дерево служат средством навигации по тексту. Первоначальное знакомство с понятиями текста осуществляется по ассоциативной сети или тематическому дереву, дальнейшее знакомство с текстом – отсылкой к предложениям текста, в том числе в самом тексте, содержащем эти понятия.

Автоматическое реферирование текста. Эта функция могла бы числиться первой, поскольку позволяет познакомиться с текстом быстрее всего. Но она может быть реализована только после создания ассоциативной сети. Ассоциативная сеть – понятия и их связи с их числовыми характеристиками – фактически формируют метрическое пространство, в котором и весь текст, и каждое его отдельное предложение являются вектором. Проранжировав предложения текста по весу (по длине вектора), и убрав предложения с весом меньшим порогового значения, мы можем получить что-то вроде реферата, если расставим оставшиеся предложения в порядке их появления в тексте.

Построение тематического реферата. Наряду с обычным рефератом можно построить так называемый тематический реферат. Для этого из ассоциативной сети выделяется подсеть, описывающая заданную запросом тему. Все предложения текста, содержащие понятия этой сети и имеющие вес выше порогового, составляют тематический реферат.

Автоматическая классификация текстов. Данная функция реализована технологией, но не включена в функциональность TextAnalyst'a. Можно сравнивать близость двух текстов, вычисляя степень корреляции между их ассоциативными сетями. Для этого вычисляется вектор пересечения этих текстов в декартовом произведении пространств каждого из текстов. Если имеется несколько тематических классов, входной текст может быть автоматически отнесен к одному из этих классов вычислением степени корреляции сети входного текста и сетей каждого тематического класса.

Автоматическая кластеризация текстов. Если разбить сеть на подсети удалением слабых связей, можно предположить, что подсети описывают некоторые предметные области – подобласти предметной области текста. Понятие, имеющее наибольший вес в данной подсети, обозначает данную предметную подобласть. Если мы имеем несколько документов, разбивая их общую сеть на части, мы можем отнести эти документы к соответствующим подсетям, сравнивая сети документов с сетями предметных областей.

Формирование персональной базы знаний. Используя вышеперечисленные функции можно обрабатывать сколько угодно текстов и врозь и вместе. Результаты такого анализа можно сохранить, формируя собственную базу знаний. Можно иметь столько папок, сколько предметных областей, их разделов, их объединений предпочитает иметь пользователь.

Дополнительно имеется еще несколько технических возможностей, делающих TextAnalyst удобной оболочкой для создания е-книги.

Редактирование ассоциативной сети и тематического дерева. Ассоциативная сеть, также как и тематическое дерево, может быть отредактирована. В ней можно заменить название понятия, исключить его и исключить связь.

Формирование словаря слов-предпочтений. Чтобы анализ текста был корректным, с точки зрения автора, он должен иметь возможность усилить анализ в интересующих его направлениях. Для этого он может сформировать свой словарь слов-предпочтений. Эти слова в процессе анализа будут учитываться в первую очередь.

Формирование словаря синонимов. Для корректного построения ассоциативной сети необходимо учесть синонимию текста. Необходимо в явном виде перечислить синонимы. Если этого не сделать, и если в тексте используются синонимы, вершины сети могут размножаться. Применение стандартных словарей синонимов в этом случае затруднено, поскольку синонимия конкретной предметной области – вещь очень тонкая.

Используя перечисленную функциональность, автор или редактор е-книги легко, в течение двух недель, может подготовить свою базу, а затем, зашить ее в оболочку. Для хорошего качества анализа необходимо соблюсти некоторые условия. Необходимо подбирать материал таким образом, чтобы он был тематически однородным. В случае книги – материал надо разбивать на главы. Можно собирать вместе несколько тематически однородных статей.

При желании можно навесить на соответствующие объекты текста (слова, предложения) информацию других мультимедийных модальностей: аудио, видео. Точно также можно иметь параллельно гипертекстовому представлению плоский текст с прямой отсылкой из гипертекста в текст и обратно.

Можно также гипертекстовую структуру создавать в гипертекстовых форматах, например, в.html. В этом случае можно использовать имеющиеся инструменты для оформления дизайна е-книги.

Практически система Text Analyst может выполнять функции обеспечения творческой работы учащегося по анализу, систематизации необработанного сырого текста, подготовке на этой основе собственных текстовых документов. Интеграция его в среду имеет огромное значение.

4.4. Педагогический сценарий работы с ИПЭБ

ИЛДТ – это технология организации речевой деятельности учащихся в компьютерной среде. Структура деятельности ИЛДТ включает слушание лекций, чтение учебной и другой литературы, выполнение письменных заданий, подготовку устных выступлений. В ИЛДТ предусматривается автоматизация всех видов речевой деятельности соответственно – чтения, письма, аудирования, говорения и перевода.

Большое значение имеет идея обеспечения речевой деятельности различными информационными (а также и) методическими ресурсами, далее ИР. Использование этих ресурсов обеспечивает, наряду с автоматизацией рационализацию и оптимизацию в выполнении видов речевой деятельности.

Интеллектуальная персональная библиотека преподавателя и учащегося – это электронный центр открытого самостоятельного доступа учебного заведения (класс, лингвистический центр), который по своей функции является не только электронной библиотекой, хранилищем разработок всего коллектива, отражающим специфику, уклон данного учебного заведения, но и электронной лабораторией по организации автоматизированной работы с текстовым материалом, лабораторией, в которой могут развиваться идеи новых технологий автоматизации интеллектуального труда, а также новые технологии обучения творчеству. Данная концепция ставит задачу показать структуру, функциональность и инновационные возможности ИПЭБ.

Концепция ИПЭБ – это концепция организации учебной работы с текстовым материалом в компьютерной среде, видов деятельности и способов их реализации. В общем виде данная концепция представлена в ниже. Она содержит отдельно представленные компоненты или следующие разделы:

- режимы работы (раздел А),
- тренировка (раздел Б)
- виды автоматизированного контроля (раздел В),
- программные средства автоматизации работы (раздел Г),
- информационные ресурсы (раздел Д),
- методическое обеспечение работы (раздел Е).
- педагогические технологии (раздел Ж)

ИПЭБ обеспечивает доступ:

- к справочному материалу по организации работы,
- к методическим и интеллектуальным компонентам обеспечения этой работы,
- к программным средствам автоматизации этой работы.

Раздел А – это раздел, содержащий сценарии организации работы по подготовке или реализации указанных видов учебной деятельности в условиях ИПЭБ. Рассмотрим их подробнее и в отдельности.

Лекционный режим работы. В этом режиме может быть реализовано три варианта работы. Вариант 1. Преподаватель готовит (обновляет) к лекции материал. В этом случае он нуждается в доступе к уже имеющимся материалам по дисциплине, сделанным ранее (разделы Д.1.). Он также нуждается во внешних источниках информации с выходом в другие (национальные, региональные или др.) библиотеки, где представлены данные по дисциплине или тематике (раздел Д2). Его работа по подготовке лекции, т.е. чтение, может быть автоматизирована, что может ускорить этот процесс за счет

автоматической обработки текстов. (См. раздел Б2 – чтение). Нужно также учесть, что работа преподавателя по составлению материалов учебного курса должна опираться на разработанные и/или утвержденные в коллективе (возможно и в сообществе специалистов, УМО или др.) концептуальные модели знаний. Они публикуются в библиотеке отдельно. (Раздел Е1). Если этого нет, то преподаватель готовит модели знаний самостоятельно. В последнем случае он нуждается в доступе к разделу Г – автоматизированные средства обработки текста и в доступе к инструкциям по организации работы преподавателя – раздел Е.5, где способы извлечения и представления знаний описаны.

Вариант 2. Преподаватель заранее готовит учащихся к своей лекции. Он может давать задания по подготовке к будущей лекции и использовать при этом материалы дополнительного чтения (раздел Д.1), материалы, помогающие в изучении словаря (гlossарии, словари, ведь лекция может быть и на иностранном языке).

В процессе чтения лекции преподаватель может использовать специальные методические форматы обеспечения речевой деятельности. Это электронные формы, электронный опорный конспект- матрица, которые содержат методические заготовки, обеспечивающие ускоренное прогнозное восприятие, корректное конспектирование, опору на модель знаний, вопросы по содержанию и др. Студент или учащийся конспектирует, используя компьютер на своем рабочем месте, создавая электронный документ, который затем становится частью его электронного портфеля. Даже если преподаватель предоставляет учащемуся свой электронный конспект или слайды, это конспект опорный, т.к. он краткий. Учащийся может делать свой конспект с индивидуальными акцентами на содержании.

Дискуссионно-презентационный режим работы. Это, по сути, миконференция, организуемая учащимися с презентацией самостоятельно – подготовленных докладов, и дискуссией по актуальной теме изучаемого курса или практики. Что будет востребовано в ИПЭБ в этом случае? Раздел А.4.3 по подготовке письменного текста (доклада) с использованием ресурсов ИПЭБ. В этом разделе должна содержаться информация по требованиям к написанию научного доклада, по структуре данного текста, его основным разделам, речевым клише, оформляющим эти разделы. В этой связи потребуются выход на разделы А 4. 5, Е.3.3 – разделы, в которых должны быть рекомендации по разработке БД или сами базы данных, содержащие описание и перечень клише, которые нужны учащемуся для обеспечения связности и логичности текста, требования или стандарты написания данного типа текста. Эти базы также могут содержать материал на различных языках.

Разработка или корректировка ФМПЗ. Данный вид учебной исследовательской деятельности в большей степени является прерогативой учителя. Однако в этот процесс могут вовлекаться и учащиеся. В ИПЭБ этот вид работы обеспечивается с помощью инструкций по организации работы учащегося и преподавателя (см. пункт А.4.1.), а также с помощью программного обеспечения обработки текстового материала для извлечения знаний. При этом используются два способа извлечения знаний и построения моделей – ручной и автоматический.

Анализ текста. Этот вид работы опирается на использование раздела «Инструкции по организации самостоятельной работы учащегося» Е.4., и его подраздел – Методы анализа текстовой информации. Среди этих методов указаны следующие:

- «Метод автоматического извлечения тематической или семантической информации»,
- «Метод анализа текста по типам информации: фактическая, субъективная, концептуальная, имплицитная»,
- «Метод содержательного (тема-рематического) анализа текста» и др.

Этот вспомогательный, учебный вид работы, как правило, сочетается с развитием аналитического чтения и оканчивается проверкой результата понимания и усвоения прочитанного в виде автоматизированного теста вопросно – ответной работы В2.

Перевод текста. Этот вид работы может быть связан с тем, что преподаватель дает учащимся задание прочитать статью по изучаемой тематике на иностранном языке, взятую из Интернета. В этом случае помочь в этой работе могут следующие компоненты ИПЭБ: глоссарии, тезаурусы, специализированные базы данных – все те разработки, которые могут содержать информацию о словаре предметной области. Эта информация, в случае острой актуальности и новизны проблемы, может и не быть еще отражена в существующих печатных словарях. Для этой цели и должны делаться специальные профессиональные, предметно- тематические накопления в виде словарей. Инструкции по разработке специализированных БД (включая словарные) должны находиться в разделе А5.

Раздел Чтение имеет в ИПЭБ 2 трактовки: аналитическое и просмотрное поисковое чтение. Учащийся или преподаватель может использовать автоматические возможности создания текстовой выжимки (реферата), словарной статистической выборки из изучаемого массива текстов, для формирования смыслового портрета текста и возможности его быстрого просмотра. Автоматические возможности изучения наиболее частотных слов через изучение их использования в контексте, а также автоматические возможности построения семантических и тематических деревьев также могут использоваться для реализации аналитического чтения. Технологии организации аналитического и просмотрного чтения должны содержаться в инструкциях по организации работы учащегося и преподавателя.

В экспериментальном порядке учащийся или преподаватель может использовать любые другие компоненты ИПЭБ.

Разучивание. Для реализации разучивания может быть использовано специальное программное обеспечение, работающее по общеизвестным принципам текстовых манипуляторов: вставки, восстановление текста, ответы на вопросы методом выбора из множества, конструирование словосочетаний, сортировки слов по категориям и др. Эти программные средства обеспечивают разучивание словаря (терминов), а также текстов определений, различных данных и др. Эти же средства используются и для организации автоматизированного тестирования.

Конспектирование. Функция Конспектирования обеспечивается в ИПЭБ в нескольких вариантах: конспектирование звучащего текста и конспектирование прочитанного текста. Для обеспечения конспектирования могут использоваться файлы мультимедийной текстотеки, модели знаний, а также методические форматы обеспечения тренировки и тестирования, содержащие компоненты вспомогательной информации. Кроме того, в доступе могут быть любые другие источники информации ИПЭБ.

Виды автоматизированного контроля. Для организации контроля понимания прочитанного может успешно применяться программа вопросно- ответного типа с выбором ответа из предлагаемого множества. Г2.

Для организации контроля усвоения учебных материалов можно применить практически все имеющиеся в ИПЭБ программы тренировки и тестирования: сортировку, вставку, вопросно-ответную программу и др. Программы фиксируют время выполнения, количество ошибок.

Для организации контроля качества письменных работ в ИПЭБ программа по автоматической обработке письменного текста извлекает из работ учащихся использованный словарь, ранжирует его по частотности, что позволяет выявить и оценить объем словаря, его наполнение и степень освоенности.

Информационные ресурсы ИР содержат (раздел 1) мультимедийные материалы для дополнительного чтения и просмотра, учебные курсы, разрабатываемые преподавателями или готовые, персональные ИР для преподавателя и учащегося.

ИПЭБ также создает контакты с удаленными ИР других ЭБ. Процесс получения текстов из удаленных ЭБ сопровождается автоматической классификацией текстов по рубрикам. Это облегчает проблему их хранения и нахождения.

Последний раздел – педагогические технологии – способы организации работы с использованием ИР и практическим результатом представлены в модели 2. См. ниже Модель 2. Концепция ИПЭБ.

Модель 2. Концепция ИПЭБ

А. Режимы работы

- А.1. Лекционный
- А.2. Дискуссионно-презентационный
- А.3. Лабораторно-практический
- 4.1. Разработка и корректировка ФМПЗ
- 4.2. Анализ текста
- 4.3. Подготовка статьи, реферата, аннотации, доклада
- 4.4. Перевод текста
- 4.5. Разработка специализированных БД

Б. Тренировка

- Б.1. Чтение
- Б.2. Разучивание
- Б.3. Конспектирование

В. Виды автоматизированного контроля

- В.1. Контроль понимания прочитанного
- В.2. Контроль усвоения учебного материала (терминов, дефиниций, ФМПЗ)
- В.3. Контроль качества письменных работ

Г. Программные средства автоматизации работы

- Г.1. Автоматизированные средства обработки текста
- Г.2. Автоматизированные средства организации тренировки и тестирования

Д. Информационные ресурсы (ИР)

- 2.1. Внутренние ИР
 - 2.1.1. мультимедийные текстотеки
 - 2.1.2. учебные курсы
 - 2.1.3. персональные ИР учащегося
 - 2.1.4. персональные ИР преподавателя
- 2.2. Удаленные ИР
 - 2.2.1. Национальные ЭБ
 - 2.2.2. Региональные ЭБ
 - 2.2.3. Научные ЭБ
 - 2.2.4. Издательские ЭБ
 - 2.2.5. Профессиональные ЭБ

Е. Методическое обеспечение работы

- 3.1. Модели знаний
- 3.2. Методические форматы обеспечения тренировки и тестирования
- 3.3 Базы данных
- 3.4. Инструкции по организации самостоятельной работы учащегося
- 3.5. Инструкции по организации работы преподавателя
- 3.6. Словари
- 3.7. Глоссарии
- 3.8. Экспертно–консультирующие системы
- 3.9. Ссылки на дополнительные учебники, сайты, ресурсы

Заключение

В данной работе была выдвинута идея библиотечного и методического комплекса – ИПЭБ и организации работы в нем на базе технологии (ИЛДТ).

Что было предложено?

ИЛДТ – способ организации учебной речевой деятельности, углубляющий знания предметной области за счет организации основных этапов учебной деятельности, объектом и предметом которой становится структурированное содержание основных понятий предметной области. Представление ядра учебной информации в виде понятийной, компактной, взаимосвязанной логически и выводной модели знаний – фрейма обеспечивает использование этого ядра – модели для разных целей и многократных повторений в процессе тренировки-разучивания, анализа и синтеза текста.

Что может быть достигнуто?

Использование ФМПЗ для моделирования знаний предметной области в виде естественных метапонятий – концептов и их структурированной формы- концептосфер и слотов, дает возможность представить предметную область в виде лингвистической модели естественного языка как перечень словаря, высказываний различной модальности, текстов, содержащих аргументацию и трактовки изучаемого феномена.

Что для этого надо сделать?

Технология ИЛДТ реализуется на базе интеллектуальной персональной электронной библиотеки (ИПЭБ), содержащей специальные программные средства для работы с текстовой информацией, а также модели знаний и методическое обеспечение для обеспечения и управления самостоятельной творческой работой учащегося.

Предлагаемая концепция ИЛДТ в силу этого неразрывно связана с лабораторным комплексом ИПЭБ, который и содержит необходимую когнитивную базу для обучения и творческой речевой деятельности учащихся с программным обеспечением открытого типа, позволяющим ее использовать для творческих целей. И именно эта библиотека и должна быть разработана как лабораторный комплекс программных средств и методического обеспечения с экспертно консультирующими возможностями.

Сегодня, когда интерес зарубежных исследователей, занимающихся проблемами организации обучения (самостоятельной работы учащегося) в компьютерной среде, также сфокусирован на структурном моделировании учебного материала, представлении знаний в виде концептуальных древовидных моделей (ontologies)¹, управления знаниями, следует особо отметить те дидактические свойства, которыми эти модели обладают. ФМПЗ является универсальным способом организации знаний для обучения в компьютерной среде и управления знаниями работой учащегося. Но, при этом, следует признать, что трактовка учебного содержательного материала через призму фрейма- лингвистической и когнитивной модели, требует изменений в подходе к тео-

¹ Среди многочисленных зарубежных авторов, которые занимаются проблемой структурной обработки учебного контента можно назвать следующих: Mizoguchi R., Koper R., Olivier B., Meisel H.. отметим здесь, что они понятием «фрейм» не пользуются, используется термин «онтология». То, что было описано нами как «фрейм» в главных чертах с их представлениями об онтологии соотносится. Вопрос установления терминологической синонимии – это проблема будущей разработки глоссария терминов.

ретической базе обучения. Требуется расширение этой базы и фактическое признание лингвистического аспекта предметной деятельности, которое становится необходимым, если мы хотим понять содержание всей той новой учебной деятельности, в которую мы вовлечены, работая в компьютерной среде. Из этого и вытекает потребность в обосновании новой лингвистической технологии.

В лингводидактическом понимании – фрейм – это структура – идеализированная картинка – гештальт древовидной структуры обработки знаний, которая уже в минимально свернутом запакованном виде (7 + – 2) обеспечивает большие и разнообразные возможности ее распаковывания, и при различном ее предъявлении имеет разные способы прочтения. Эта структура, будучи логически связанной базовым отношением «род- вид», может распаковываться и запаковываться, создавая условия для различных форм ее методического использования.

ФМПЗ – это способ концептуального моделирования и, если ФМПЗ строится преподавателем на основе онтологических семантических понятий, которые должны использоваться на уровне метазнаний и языковой картины мира, то учебное распаковывание ФМПЗ может войти в резонанс с естественным процессом мышления с использованием тех же онтологических и семантических понятий и того же лингвистического ресурса – семантической памяти. В таком случае должно наступать такое состояние как «понимание», и семантическая память становится доступной для работы уже над учебным материалом как своим.

Использование ФМПЗ дает компактное представление знаний, которое может использоваться в обучении в запакованном микро- и распакованном макро- виде. Это дает возможность использовать микро презентации фрейма на начальных этапах обучения, увеличивая объем знаний и расширяя представление фрейма в дальнейшем, на повторном цикле при концентрической подаче материала. Для качества усвоения важно то, что проходя путем от простого к сложному, учащийся остается в том же смысловом поле того же понятия, наращивая только связи и объем, но не создавая нового концепта для запоминания.

Использование модели предметной области в вербальном представлении открывает ряд важных перспектив для ее автоматической измеряемости. Становится возможным измерять объем и качество словаря учащегося относительно модели, производить анализ текста на соответствие его тематике и полноту раскрытия темы, проверять точность определений, качество прочитанного и усвоенного. Это осуществляется с помощью предложенного в концепции комплекса программных средств.

ФМПЗ обеспечивает методическую опору для опережающего анализа и синтеза речи, будучи использованным в процессе аудирования, чтения и письма.

ФМПЗ содержит перечень ключевых слов, которые могут в дальнейшем стать стандартом описания данной темы, проблемы и др., выполнять библиотечные функции, что очень важно для использования этой модели для поиска текста, знаний о предмете и идентификации материала по модели знаний.

Описание предметной области как языковой модели создает условия для автоматизированной тренировки с помощью программ типа «текстовые манипуляторы».

Внедрение данной технологии потребует разработки программного обеспечения, экспертно- консультирующих систем и лингвистических баз данных, учебных курсов для того, лабораторного экспертно- консультирующего комплекса, чтобы реализовать идею формирования лингвистической, социально – коммуникативной и интеллектуальной компетенций учащегося в условиях массового обучения.

Литература

1. Лобанов Ю.И. Представление знаний в открытом образовании: проблемы, концепции. – М., 2001. – 56 с. – (Новые информационные технологии в образовании: Аналитические обзоры по основным направлениям развития высшего образования / НИИВО; Вып. 11, ч. 1.)
2. Лобанов Ю.И., Ильченко О.А. Управление знанием в открытом образовании. Часть 2. – М., 2004. – 60 с. (Новые информационные технологии в образовании: Аналитические обзоры по основным направлениям развития высшего образования / НИИВО; Вып. 2.)
3. Гаврилова Т.А., Червинская К.Р. Извлечение и структурирование знаний для экспертных систем. М., 1992. С. 9.
4. Смутьсон М.Л. Средства. Модель задачи // КТО. С. 464.
5. Золотопуп О.Н., Колос В.В. Фрейм // КТО. С. 591.
6. Поспелов Г.С. Искусственный интеллект – основа новых информационных технологий. – М.: Наука, 1988. С. 33.
7. Крюкова О.П. Самостоятельное изучение иностранного языка в компьютерной среде. – М.: Логос, 1998. – 126 с.
8. Гурина Р.В., Соколова Е.Е. Фреймовое представление знаний. – М.: Народное образование: НИИ школьных технологий, 2005. – 173 с.
9. Калмыков А.А. Сравнительный системный анализ образовательных технологий // Педагогическая информатика. 2001. № 4. С. 37–46.
10. Падучева Е.В. О денотативном статусе именных групп в предложении // Логический анализ языка. Ментальные действия. – М.: Наука, 1993. Вып. 6. С. 48.
11. Величковский Б.М. Когнитивная наука. Основы психологии познания. Т.2. М.: ACADEMIA: Смысл, 2006. С. 48.
12. Величковский Б.М. Когнитивная наука. Основы психологии познания. Т.2. М.: ACADEMIA: Смысл, 2006. С. 50.
13. Rosch, E.H. On the internal structure of perceptual and semantic categories // In T.E. Moore (Ed.). Cognitive development and the acquisition of language. NY: Academic Press, 1973. P. 113–114.
14. Abelson R. Psychological status of the script concept // American Psychologist. Vol. 36. 1981. № 6. P. 37–49.
15. Boyd F. *Making Business Decisions. Real Cases form Real Companies.* Addison-Wesley. 1994.
16. Волков А. На языке «Интеркорнета» и других наречиях // Знание – сила. 2008. № 4. С. 4–12.
17. Психология. Словарь. – М.: Политиздат, 1990. С. 284.
18. Tulving E. Episodic and semantic memory // In: E. Tulving & W. Donaldson (Eds.). *Organization of memory.* NY: Academic press, 1972. P. 386.