

Электронные гипертекстовые тезаурусы в дистанционном обучении

И.А.Иванова, асп. каф. Информатики и информационного менеджмента

Тел.: (8452) 56-25-42. E-mail: aivanov@volgasoft.net

Поволжская академия государственной службы

This article presents thesaurus definition, stages and rules of thesaurus building. Hypertext thesaurus for computer science was created by Web-technology. Thesaurus structure and methods of its optimization are describing. Some variants and results of computer hypertext thesaurus using for distant education are getting.



Система дистанционного обучения включает разработку электронных учебников, специальных программ для организации практических работ, приложений для контроля знаний и др. При решении всех этих задач могут быть использованы электронные гипертекстовые тезаурусы.

Тезаурус – это смысловой словарь, представляющий совокупность знаний внутри какой-либо предметной области. Тезаурус содержит перечень понятий, терминов с заданными между ними смысловыми отношениями. Элементами, между которыми устанавливаются семантические соответствия, являются дескрипторы («описатели»), имеющие форму слов или словосочетаний. Перечень дескрипторов в тезаурусах должен быть упорядочен по алфавитному и систематическому принципу. Каждая словарная статья тезауруса, помимо основного элемента, содержит ссылки на другие дескрипторы или понятия, находящиеся с данным дескриптором в отношениях рода, вида, ассоциации, синонимии и других.

Разработка тезауруса может осуществляться группой специалистов по данной предметной области, и тогда метод построения тезауруса называется методом экспертной оценки. Часто этот метод сочетают со статистическим методом, использующим компьютерный анализ текстов по данной области знания.

Использование для реализации тезауруса средств Web-технологии делает возможными ссылки на информацию, имеющуюся как в самом тезаурусе, так и на любом узле Интернет, предусматривает вызов конкретной гипертекстовой страницы, содержащей какой-либо термин. Становится возможным использовать в словарной статье графическую, видео и другие виды информации. Значительно упрощается организация ссылок и интерфейса пользователя.

Учитывая все перечисленные преимущества, создан небольшой электронный гипертекстовый тезаурус по информатике, содержащий несколько десятков дескрипторов и около 200 терминов.

Для реализации тезауруса использован язык гипертекстовой разметки HTML. Структура HTML-документов проста, что позволяет с помощью разработанной модели создавать подобные гипертекстовые тезаурусы для других предметных областей.

Подбор лексики осуществлялся на основе курса лекций по информатике для студентов специальности «Информационные системы в управлении». После составления алфавитного списка для каждого из элементов (дескрипторов) будущего словаря были выделены синонимы, родовые, видовые, ассоциативные дескрипторы, также являющиеся элементами тезауруса.

Форма дескрипторной статьи в различных тезаурусах может быть неодинакова. Однако внутри одного тезауруса форма дескрип-

торной статьи должна быть единой. В данном тезауусе дескрипторная статья содержит следующие пункты: определение дескриптора, синоним, родовой дескриптор, видовой дескриптор, ассоциативные отношения, поле использования. Каждый пункт статьи может содержать текст, графику, а также ссылки на элементы данного тезаууса, тезаууса по смежной области знания, любые другие файлы на данном компьютере или компьютерах локальной сети, любые ресурсы Интернета. Файловая структура модели представлена на рис.1.

Каждой букве алфавита соответствует файл, содержащий перечень дескрипторов, начинающихся с данной буквы, после которого в этом же файле следуют соответствующие словарные статьи. В названии файлов, находящихся в папке «Систематический», отражена условная нумерация вершин классификационной схемы тезаууса.

Работа с тезауусом возможна с помощью любого Web-браузера. На рис.2 представлен вид экрана при обращении пользователя к алфавитному способу работы с тезауусом.

Окно просмотра разбито на два фрейма. В левом представлен алфавит. Щелчок на нужной букве позволяет получить в правом фрейме список имеющихся дескрипторов, также упорядоченных по алфавиту. Щелчок на нужном термине выводит в этом же правом фрейме соответствующую словарную статью. Общий вид словарной статьи также представлен на рис. 2.

Для упорядочения дескрипторов тезаууса по систематическому принципу необходимо построить классификационную схему и составить указатель иерархических отношений. Электронный гипертексто-

вый тезауус позволяет обращаться непосредственно к классификационной схеме, не задействуя указатель иерархических отношений в том виде, в каком он используется в бумажных тезауусах.

На практике применяются древовидные классификационные схемы (например, десятичные классификации, типа УДК), фасетно-блочные классификации и булевы классификации. При составлении тезауусов удобно использовать классификационные схемы первых двух типов. В данном тезауусе применена древовидная классификационная схема. Фрагмент ее представлен на рис. 3.

Рядом с названием каждой вершины дерева указано имя файла, в котором содержится перечень подчиненных ей вершин и список дескрипторов, по смыслу относящихся к данной вершине. Вид экрана при работе по систематическому принципу представлен на рис. 4.

Щелчок мыши позволяет либо просмотреть нужную словарную статью, либо перейти к разделу более низкого уровня.

Следует указать на имеющуюся взаимосвязь между алфавитным и систематическим способами работы с тезауусом. В дескрипторной статье, в поле для каждого дескриптора

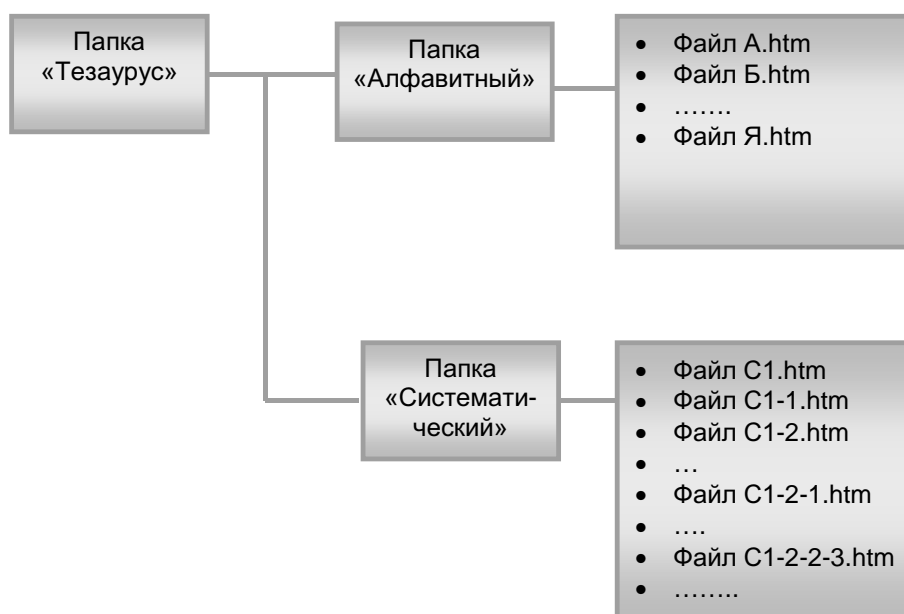


Рис. 1. Файловая структура электронного гипертекстового тезаууса

указывается ссылка на соответствующую вершину классификационного дерева.

Таким образом, пользователь свободно комбинирует два способа работы, а также может просто «скользить» по смысловым ссылкам непосредственно из словарных статей.

В процессе создания тезауруса могут возникнуть различные ошибки, ведущие к нарушению оптимальной структуры словаря, в частности:

- отсутствие у дескрипторов связей вообще;
- избыточное или слишком малое количество связей;
- наличие дескрипторов, не включенных ни в один раздел классификационной схемы;
- неполное, неправильное или одностороннее отражение родо-видовых, ассоциативных и других связей;
- наличие в дескрипторной статье ссылок на саму себя.

Для более эффективно и правильного составления тезауруса можно рассматривать его как графовую структуру, где каждому дескриптору соответствует вершина графа, а дугами, инцидентными ей, являются связи с другими дескрипторами (вершинами).

Для описания структуры тезауруса предлагается ввести соответствующие графу матрицы:

- ⇒ Матрицу весов A ;
- ⇒ Матрицу связности $A1$.

Для отражения порядка и типа связей дескрипторов друг с другом предлагается

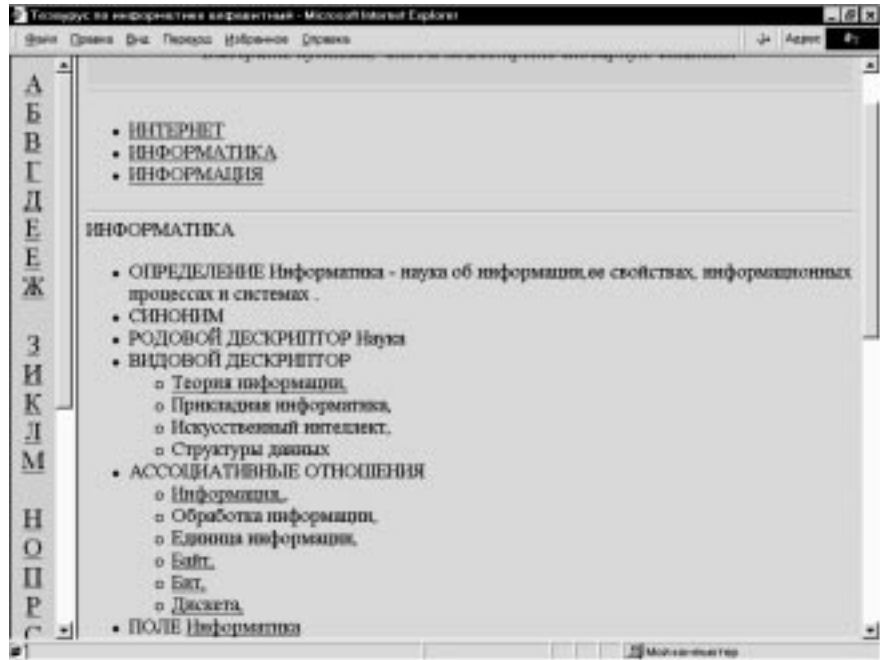


Рис. 2. Вид дескрипторной статьи и экрана при алфавитном способе работы с тезаурусом

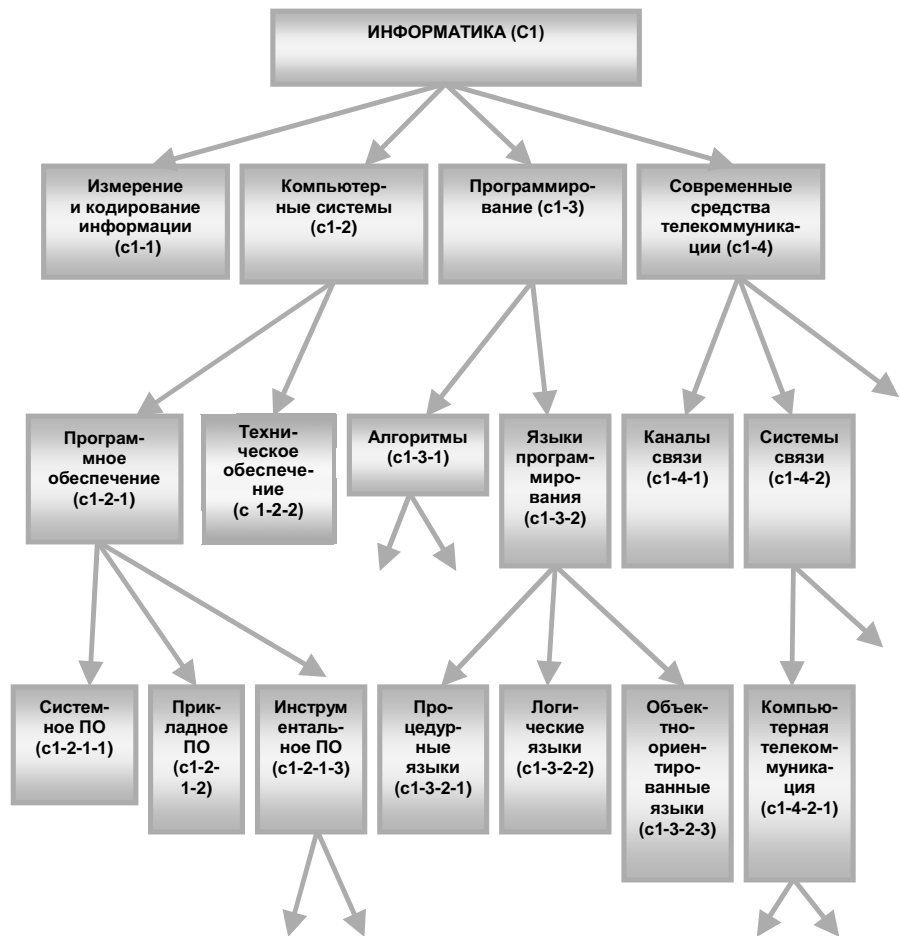


Рис. 3. Фрагмент классификационной схемы гипертекстового тезауруса по информатике

строить матрицу весов A . Это обязательно квадратная матрица размерности n , где n – количество дескрипторов в тезаурусе. В графе, ассоциированном с тезаурусом, каждую дугу можно охарактеризовать условным «весом», зависящим от типа соответствующей связи. Элемент матрицы весов $a(i,j)=2$, если дескриптор i связан с дескриптором j ассоциативными отношениями; $a(i,j)=3$, если дескриптор j является видовым для дескриптора i ; $a(i,j)=4$, если дескриптор j является родовым для дескриптора i .

Для отражения связей разделов классификационной схемы с дескрипторами строится матрица связности $A1$. Это матрица размерности $(m*n)$, где m – число вершин классификационного дерева (число подклассов), n – количество дескрипторов в тезаурусе. Элементы этой матрицы $a1(i,j)=1$, если дескриптор i связан с вершиной классификационного дерева j . В каждой строке матрицы $A1$ содержится информация о том, к каким вершинам относится дескриптор j . Каждый столбец матрицы $A1$ содержит столько единиц, сколько дескрипторов включены в подкласс j .

Автоматический анализ описанных матриц, составленных на этапе построения тезауруса, позволяет эффективно устранить ошибки (нерегулярности) в структуре тезауруса. Для этого следует придерживаться некоторых эмпирических правил и требований к матрицам A и $A1$.

Правило 1. Полнота смысловых связей между дескрипторами. Матрица A не должна содержать пустых строк и столбцов.

Правило 2. Симметричность родо-видовых и ассоциативных связей между дескрипторами. В матрице A должны выполняться условия:

- если $a(i,j)=2$, то $a(j,i)=2$;
- если $a(i,j)=3$, то $a(j,i)=4$;
- если $a(i,j)=4$, то $a(j,i)=3$;

Правило 3. Дескрипторы не должны содержать ссылок на самих себя (тезаурусная сеть не должна иметь петель). В квадратной матрице A диагональные элементы должны быть пусты.

Правило 4. Любой дескриптор должен быть включен в классификационную схему тезауруса (принадлежать хоть одной вершине классификационного дерева). Матрица $A1$ не должна содержать пустых строк.

Набор правил постоянно расширяется. Небольшой комплекс программ на VBA в среде EXCEL позволяет автоматически проводить анализ матрицы весов и матрицы связности. Использование матричной модели позволяет улучшить качество тезауруса, ускорить процесс его построения.

Электронный гипертекстовый тезаурус по информатике, прежде всего, служит для обучения основному смысловому ядру предметной области.

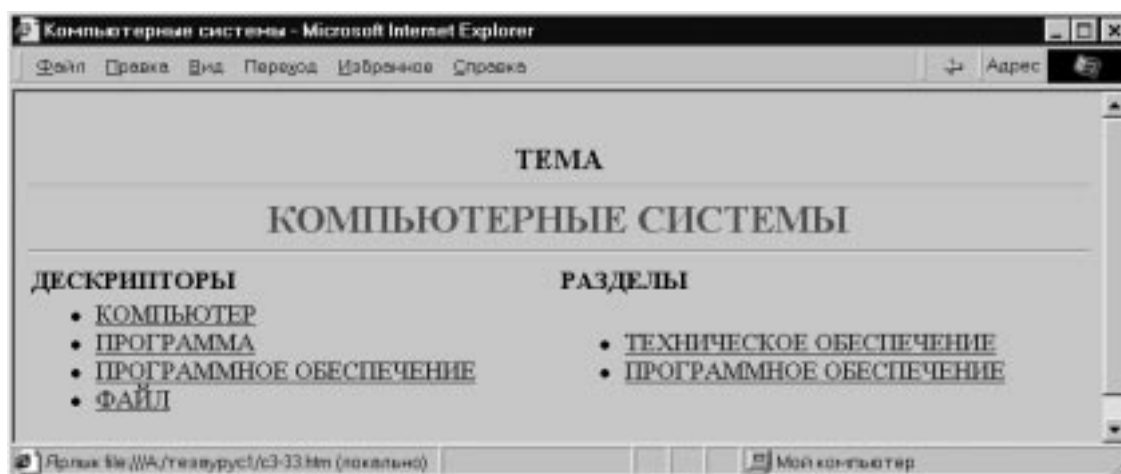


Рис. 4. Вид экрана при выборе систематического способа работы с тезаурусом.

Подобный тезаурус может стать полезным приложением (или заключительной главой) к любому электронному учебнику. Это позволит обучаемому систематизировать знания, полученные при ознакомлении с материалом учебника, а также быстро отыскать не только интересующие его новые понятия, но и изучить все основные связи этих понятий с другими. В данном случае в дескрипторную статью предлагается ввести ссылку на параграф в электронном учебнике, в котором излагается материал, связанный с данным дескриптором. И, наоборот, в текст учебника – указатель на соответствующий дескриптор. При составлении такого приложения может быть использован метод автоматического составления тезауруса.

С другой стороны, предлагается использовать электронный гипертекстовый тезаурус как основу для составления электронного учебника. В этом случае классификационная схема тезауруса может полностью соответствовать оглавлению будущего учебника. Понятия и терминологию предметной области, по которой составляется учебник, удобно подчинить «тезаурусным» правилам: выделить все возможные связи между понятиями и определить способ изложения в зависимости от типа этих связей и выявленной иерархии.

Электронный гипертекстовый тезаурус может применяться для контроля знаний. В этом случае, например, обучаемый получает описание структуры тезауруса, модель, состоящую из готовых информационных структур и интерфейса (оболочку). Указывается приблизительное число дескрипторов и терминов, которые должны войти в тезаурус. Контрольным заданием является самостоятельное составление тезауруса по изученной предметной области. Преподаватель, подводя итог самостоятельной работы, использует не только свои знания как эксперта, но и применяет описанную выше матричную модель для анализа

полноты и правильности связей, отраженных в «контрольных» тезаурусах.

Практическая апробация предлагаемых способов самостоятельного составления электронных гипертекстовых тезаурусов проводилась со студентами, изучающими курс «Лингвистические основы информатики», специальности «Информационные системы в управлении». Студентами были успешно выполнены курсовые работы по созданию гипертекстовых тезаурусов по экономике, биологии и другим областям знания. Особенно эффективным было использование этого способа в рамках заочного образования, широко применяемого в Поволжской академии государственной службы.

Литература

1. Тезаурус информационно-поисковый многоязычный. Правила разработки: ГОСТ 7.24-80.
2. Бельков С.А., Гольдштейн С.Л., Ткаченко Т.Я. Гипертекстовый тезаурус системных знаний // Научно-техническая информация. Серия 2. Информационные процессы и системы». 1998. № 10.
3. Майника Э. Алгоритмы оптимизации на сетях и графах М.: «Мир», 1984.
4. Соколов А.В. Информационно-поисковые системы. М., 1981.
5. Иванова И.А. Методика разработки и оценка качества электронных гипертекстовых тезаурусов / Тезисы доклада на 10-й международной научно-методической конференции «Наукоемкие технологии образования». п. Дивноморское (19-22 июня 2001).
6. Иванова И.А. Методология построения электронных гипертекстовых тезаурусов оптимальной структуры / Материалы третьей всероссийской конференции молодых ученых и аспирантов «Новые информационные технологии. Разработка и аспекты применения», Таганрог (15-16 ноября 2000).