

Русский язык как второй – где найти информацию преподавателю?

Голубева А.В., Максимова А.В., Левина Г.М.

Подготовительный этап работы по проекту включал экспертный анализ интерфейса следующих информационных систем для русистов: Национальный корпус русского языка, Справочно-информационная система "Русский язык в России и за рубежом", разработанная Петрозаводским ГУ в рамках ФЦП «Русский язык», "Архив Петербургской русистики". На основе данных анализа была произведена разработка опросников для фокус-группы, касающиеся удобства интерфейса и возможных ситуаций использования системы, и проведены 16 опросов, в которых приняли участие студенты (бакалавры и магистры), обучающиеся по направлениям «Лингвистика» и «Филология», аспиранты (специальности 13.00.02 - теория и методика обучения и воспитания (русский язык как иностранный) и 10.02.01 – русский язык), преподаватели трех возрастных групп. Всего в опросах приняли участие 112 человек. В результате опросов была выделена группа экспертов из 11 человек, представляющая различные типы пользователей. На основании опросов и экспертных оценок было составлено структурированное описание предметной области, выделены объекты, ее составляющие, и их существенные характеристики, смоделированы пользователи системы, выявлены их потребности при использовании информационной системы, установлены атрибуты, на основании которых пользователи осуществляют запросы. В качестве рабочего был выбран локальный подход к созданию информационной системы, т.к. он обеспечивает быструю отдачу, возможность работы удаленными малыми группами, возможность постепенного наращивания возможностей системы с учетом результатов ее внедрения и отзывов пользователей, а также развития программного обеспечения. В качестве последнего было принято решение опираться на типовые инструментальные средства как более экономичные и не требующие привлечения специалистов-программистов высокой квалификации.

Далее были сформулированы требования к системе в целом, ее функциям, уровню подготовки разработчиков, экспертов и пользователей системы.

В качестве требований к экспертам учитывались опыт научной и практической работы в области преподавания русского языка как

второго, знание современных методов и технологий преподавания иностранных языков, знание методов исследования и системного анализа, способность общения с специалистами разных уровней и профилей. Эксперты были распределены по типам пользователей в зависимости от количественной оценки соответствия требованиям: новичок, специалист, мастер.

Для сбора текстовых материалов экспертами предложены следующие методы: горизонтальный, предполагающий детальное изучение информации по одной теме в каждом подразделении описываемой предметной области, и вертикальный, предполагающий исследование подразделений объекта по которым происходит движение информационных потоков.

Методы анализа материалов исследования не предполагают использования экономико-математических методов, при этом они позволяют отображать организационные, функциональные, содержательные характеристики объекта, их взаимосвязи, а также хронологическую последовательность развития понятий и методов предметной области.

В рамках подготовительного этапа исследовательской группой было произведено: 1) разделение будущей информационной системы на компоненты на основе функционального и тематического подходов, 2) определен необходимый перечень задач, 3) определена необходимость решения конкретных прикладных задач с помощью вычислительной техники, 4) согласованы задачи между собой с выделением некоторых задач из разработок, объединением, упрощением задач.

Поставленные задачи отнесены разработчиками информационной системы к классу задач, в которых нет формализованного алгоритма.

Сформулированы требования к информационному обеспечению: 1) оно должно быть достаточным для выполнения всех функций, которые автоматизируются, 2) для кодирования информации должны использоваться классификаторы которые понятны конечному пользователю системы, 3) система должна быть совмещена с информационным обеспечением, которое взаимодействует с ним, 4)

форматы документов-исходных текстов должны отвечать стандартным требованиям разработчиков и быть доступны для пользователей, 5) форматы графических изображений также должны отвечать стандартным требованиям разработчиков и быть доступны для пользователей, 6) формы представления выходной информации должны быть оценены фокус-группой на стадии внедрения системы, 7) сокращения должны быть общеприняты в предметной области и оценены фокус-группой, 8) в информационной системе должны быть предусмотрены необходимые средства по контролю и обновлению данных в информационных массивах, контролю идентичности информации в БД.

Анализ требований фокус-групп к интерфейсу информационной системы позволил выстроить этапы процесса размещения данных на экране: решить какая информация должна появляться, определить главный формат информации, определить где она должна появиться, определить средства для выделения полей, разработать проект размещения данных на экране, оценить эффективность размещения.

При этом разработчики ориентировались на следующие принципы:

- 1) информация должна размещаться так, чтобы пользователь мог просматривать информацию в логической последовательности;
- 2) идентифицировать связанные группы информации;
- 3) отличать исключительные ситуации;
- 4) определять необходимые действия.

Основные этапы, которые предстоит решить в процессе апробации прототипа информационной системы: 1) на основе исследования определить какая информация нужна пользователю и какая должна сейчас находиться на экране; 2) разработчик должен определить размер областей вывода и атрибуты каждого поля; 3) определяется место рационального распределения информации на экране (информация размещается в иерархической последовательности: экран, видео кадр, окно, панель, поле; сформированное изображение для одновременного показа информации на экране; типы окон: первичное, вторичное, всплывающее; размещение информации на видео кадре: заголовок, инструкция по способу работы с информацией, предметная область, инструкция по тому что делать дальше, область сообщений, область

ввода команд, область функциональных клавиш; общие принципы размещения информации на экране: оставить свободной половину экрана, оставить пустым строку после каждой энной строки, оставить 4-5 пропусков между столбцами, текст размещается так чтобы просмотр перемещался в нужном направлении, содержание полей должно размещаться и выравниваться около горизонтальной и вертикальных осей, размещение данных слева направо и сверху вниз); 4) определение атрибутов которые привлекают внимание к некоторой части экрана или действию (атрибуты: цвет символов, мерцание, звук, уровень яркости, цвет фона; правила пользования цветом: использовать минимальное количество цветов, для больших панелей использовать цвет фона, для данных использовать более яркий цвет, для выделения двух областей использовать черный и белый, цвет нужно использовать в соответствии с представлениями пользователя, экспериментировать с разными оттенками); 5) разработка проекта экранной формы; 6) методы оценки размещения информации (прямоугольников- информация разбивается на логические группы и выделяется прямоугольниками, экран разделяется вертикальными и горизонтальными осями и число и размер прямоугольников оцениваются относительно выделенных осей, выделенных точек- позволяет выделить область экрана к которой обращается внимание пользователя).

Прототип системы в настоящее время оптимизирован для просмотра с помощью обозревателя Internet Explorer (версии 5.0 и выше). Кириллические символы не входящие в стандартную таблицу (знаки фонетической транскрипции и т.п.) отображаются с помощью шрифта Arial Unicode MS, входящего в стандартную поставку MS Office 2000 или доступного бесплатно. Для корректного чтения некоторых статей, в которых встречаются символы, отсутствующие в стандартных шрифтах, рекомендуется установить на компьютер шрифты [SILDoulosIPA](#), и [SIL Galatia](#) доступные бесплатно.

Второй этап работы над информационной системой заключался в формировании базы данных из открытых первоисточников по предметной области, предоставляемых органами государственной власти, научными институтами, аналитическими центрами,

общественными организациями и частными лицами объемом до 500 авт.л. , а также включение в базу тематических публикаций научных журналов, центральных и региональных СМИ. Для стандартизации описания объектов системы были выработаны внутрисистемные требования к библиографированию источников. Процесс библиографирования зависит, с одной стороны, от объема, формы и содержания его непосредственных документальных объектов, с другой - от конкретных целевых и пользовательских установок создаваемой библиографической информации. Подготовка библиографического описания любого назначения складывается из четырех этапов: подготовительного, включающего выбор и изучение выделенных подразделов в предметной области, выявления литературы; аналитического, связанного с общим библиографическим анализом, составлением библиографических описаний, аннотированием или реферированием, индексированием и предметизацией выявленных документов; синтетического, в рамках которого происходят окончательный отбор документов для базы данных и их группировка; заключительного, на котором осуществляются перевод документов в электронную форму, форматирование, редактирование и оформление базы данных, разметка для последующих поисковых инструментов. Нами использовались два вида библиографического отбора: ограничительный и качественный.

Как известно, ограничительный отбор выступает на практике в двух разновидностях: формальный и содержательный. Формальные ограничения могут быть пространственными - фонд библиотеки или фонды нескольких библиотек, издательский репертуар, книготорговый ассортимент; формально-издательскими - вид издания, тираж, объем, способ размножения и т.д.; хронологическими - учетный период, время написания или издания; территориальными, языковыми, авторскими и др. При таком отборе не затрагивается содержание, тематика документов. Поэтому формальный отбор был дополнен содержательным. Содержательные ограничения в отборе, по сравнению с формальными, более активны по отношению к пользователю, но четкой

дифференциации их пока не разработано. Поэтому в каждом конкретном случае составления существуют свои содержательные ограничения.

Важно, чтобы из всей массы имеющихся документов были отобраны те, которые связаны с определенной отраслью знания, проблемой, темой, жизнедеятельностью конкретного лица или коллектива и т.п. В нашем случае включались публикации по следующим темам:

общие вопросы методики,
языковая политика,
усвоение родного языка,
методика образования взрослых,
методика преподавания иностранного языка,
преподавание русского языка как иностранного,
преподавание русского языка билингуам,
русский язык в специальных целях,
тестирование и сертификация,
технические средства обучения,
межкультурная коммуникация,
страноведение,
организации,
учебные пособия.

Каждая из этих рубрик станет отдельным разделом базы данных. Качественный отбор связан с оценкой содержания документов, что особенно характерно в случае рекомендательной и имеет место в научно-вспомогательной библиографии, разновидностью которой можно считать создаваемую информационную систему. Составителю важно не только основательно знать предметную область и имеющуюся по разделам литературу, но и уметь осуществлять сравнительный анализ научной или практической значимости, ценности близких по тематике работ, учитывать уровень подготовки, специфику восприятия и особенности психологии различных групп потребителей информации. В целом важно иметь в виду, что все виды ограничений в процессе библиографического отбора практически могут сочетаться друг с другом, что и имеет место в нашем случае.

Последующий этап систематизации отобранной текстотеки включил три основных процесса: индексирование и предметизация, библиографическая группировка, оцифровка.

Согласно ГОСТ 7.0-84 библиография представляет собой упорядоченное множество библиографических записей, объединенных по какому-либо признаку. Правда, в свете такого определения (по одному признаку) мы получим описание с самой простой структурой, т.е. библиографический список. Помимо формальных способов систематизации в библиографии используются еще два вида систематизации содержательного характера: библиографическая классификация - по какой-либо схеме классификации документов; библиографическая предметизация - непосредственно по содержанию документа путем определения ключевых слов или дескрипторов. Как известно, в библиографической литературе и в ГОСТ 7.0-84 принят еще один вид библиографической систематизации - "библиографическая группировка", определяемый в очень ограниченном значении: "объединение библиографических записей в группы по заданным признакам" [ГОСТ.7.0-84. С. 13]. Чтобы облегчить библиографическую группировку в создаваемом библиографическом пособии, составитель должен предварительно осуществить индексирование и предметизацию каждой библиографической записи. При индексировании в аппарате библиографической записи указывается классификационный индекс - "условное обозначение деления какой-либо системы классификации, к которому относится библиографируемый документ" [ГОСТ 7.0-84. С. 5]. Учитывая использование в нашей стране нескольких систем классификации (УДК, ББК и др.), было принято решение опираться на универсальную десятичную классификацию (УДК) как принятую в качестве единой информационной платформы среди специалистов книжного дела. В ходе предметизации, т.е. выявления основного предмета (содержания), о котором идет речь в библиографируемом документе, для каждой библиографической записи определяется предметная рубрика: "краткая формулировка предмета (факта, события, аспекта и т.п.), рассматриваемого в документе" [Там же].

Предметные рубрики (заголовки) могут создаваться, по крайней мере, тремя способами. Во-первых, путем перевода классификационного индекса в соответствующую предметную рубрику используемой схемы классификации (в нашем случае УДК – там где оно имеется в оригинальном издании). Во-вторых, в качестве предметных рубрик могут быть использованы заглавия документов, в необходимых случаях уточненные исходя из контекста (содержания) библиографируемого документа (этот случай использовался для статей, тезисов и подобных материалов). В-третьих, основное смысловое содержание (предмет) документа может быть с достаточной степенью точности и полноты выражено соответствующим списком или сочетанием так называемых ключевых слов, т.е. наиболее существенных для этой цели слов и словосочетаний, обладающих назывной (номинативной) функцией в тексте библиографируемого документа. Последний способ является наиболее трудозатратным, т.к. на практике авторы представляющей интерес для настоящей работы предметной области редко описывают свои произведения через ключевые слова.

Ключевые слова могут быть выбраны непосредственно из содержания документа или носить формализованный, нормативный характер. В этом случае они превращаются в дескрипторы, т.е. предназначенные для предметного индексирования документов нормативные ключевые слова, которые по определенным правилам отобраны из основного словарного состава того или иного естественного языка и у которых искусственно (при помощи соответствующих отсылок и помет) устранены синонимия, полисемия и омонимия. Для этих целей создаются специальные дескрипторные словари, или тезаурусы. Они могут быть универсальными и отраслевыми. На их основе развивается такой вид предметизации, как координатное индексирование, при котором основное смысловое содержание текста документа представляется в виде сочетания ключевых слов или дескрипторов. Координатное индексирование считается одним из перспективных направлений в библиографической систематизации. В целом взаимосвязь между предметом (содержанием) документа, предметной рубрикой и предметным индексированием или,

иначе, между содержанием, выражающими его ключевыми словами и дескрипторами хорошо выразил известный американский специалист по информатике Ч.Л.Берньер: "Предмет - это то, чему посвящен разговор, дискуссия или статья. Предметный заголовок - это термин, который выбран для того, чтобы направить или привести потребителя к предмету, а предметное индексирование - это метод указания предмета потребителю. Предметизатор должен выявить один или более предметов, о которых идет речь, подобрать один или более терминов для обозначения этих предметов, после чего заменить данные термины стандартными, применяемыми при индексировании".

Независимо от способа предметного индексирования, полученные рубрики заданным образом должны быть сгруппированы (перечислительно, иерархически, интегрально) и как бы составить содержательную структуру будущей базы данных. Библиографическая группировка, проведенная по проекту, поможет пользователям лучше ориентироваться в пособии, быстро находить нужные материалы, видеть взаимосвязь между ними, а в ряде случаев и подсказывает порядок, последовательность знакомства с содержанием. Группировка включала три основные операции: разделение всего массива записей на определенные группы по избранному признаку; определение последовательности разделов, подразделов, рубрик и т.д.; выбор способа расположения записей внутри рубрик (подрубрик), т.е. внутри так называемых последних делений.

Таким образом, по проекту завершен предварительный этап, включавший формулировку требований к информационной системе на основании опросов фокус-групп, представлявших различные группы будущих пользователей информационной системы. Большое внимание на этом этапе уделялось требованиям к интерфейсу системы. С помощью экспертов из числа участников опросов произведены отбор объектов для базы данных, их оцифровка и выделены формальные и содержательные характеристика ее последующей систематизации. Осуществлено разделение всего массива записей на

определенные группы по заданным признакам; определена последовательность разделов, подразделов, рубрик и т.д.; выбор способа расположения записей внутри рубрик (подрубрик). Ведется постоянное пополнение базы и подготовка к размещению ее в сети. Проект выполняется при финансовой поддержке РГНФ (09-04-12118в).