

М.М. Митев, Е.В. Рачева, Д.З. Димитров

Применение текстометрических исследований при разработке специализированных интерпретаторов

Представлены результаты разработки технологической последовательности целевого проведения текстометрических исследований проектирования специализированных интерпретаторов пользовательских интерфейсов, предназначенных для генерирования и настройки программного кода для создания и управления базами данных, связанными с информационным обеспечением дистанционного обучения.

The article presents the results of a technological sequence of the textometric research for the purpose of the design of specialized interpreters for user interfaces, designed to generate and configure the programming code for creating and managing the databases connected with the information support for E-learning.

Наведено результати розробки технологічної послідовності цільового проведення текстометричних досліджень проектування спеціалізованих інтерпретаторів інтерфейсів користувача, призначених для генерування та налаштування програмного коду для створення та керування базами даних, пов'язаними з інформаційним забезпеченням дистанційного навчання.

Введение. Разработка специализированных интерпретаторов для генерирования пользовательских интерфейсов путем ассемблирования и настройки программных классов связана с выполнением следующих технологических этапов:

- *Определение предметной области*, для которой предназначена разрабатываемая программная система. Это требует определения основных, дополнительных и вспомогательных функций, программно реализуемых и поддерживаемых системой, а также определения данных и их характеристик, обеспечивающих выполнение специфицированных функций и других особенностей приложения.

- *Определение профиля основных пользователей*. На данном этапе необходимо провести исследование функциональных характеристик потребителей, используемых языковых средств, структуры управления, методов идентификации, доступа и видов ответственности.

- *Определение диапазона характеристик и требований*, предъявляемых к программной среде разработанным специализированным интерпретатором.

- *Проведение текстометрических исследований* языковых характеристик потенциальных пользователей в целях определения множества понятий и частоты их использования, структуры и иерархии терминов, связей между понятиями и выполняемыми функциями, а также изучения коммуникации между данными и

пользователями, требований к данным, обеспечивающим определенные понятия и пр. [1].

- *Разработка семантических цепочек* понятий для выполнения определенных функций и данных, необходимых для их обеспечения.

- *Разработка множества классов* для генерации программного кода, связанного с созданием и поддержкой специализированной базы данных и исполнением определенных функций.

- *Разработка специализированного интерпретатора*, обеспечивающего пользовательский интерфейс, базирующийся на специфицированных понятиях, ассемблирование и настройку программных классов, управление и контроль над исполнением и др.

- *Разработка программного навигатора* и системы предоставления вспомогательной информации.

Во время проведения исследований и проектирования необходимо учитывать, что каждый из перечисленных этапов имеет свои специфические особенности, которые, в общем случае, непосредственно связаны с предметной областью. Эти этапы можно расширять и модифицировать в зависимости от конкретных требований пользователей и/или ограничений, накладываемых программной средой.

Цель исследований

Цель настоящей разработки – проведение текстометрических исследований языковых средств, используемых отдельными пользователями в рассматриваемой предметной облас-

ти. Информационные источники, представляющие собой объекты исследования, могут группироваться следующим образом:

- *Текстовые документы* – законы, нормативные и поднормативные источники, квалификационные характеристики потенциальных пользователей, описания основных видов деятельности, специфицированные функции для разработки технического задания, письменные мнения и точки зрения.

- *Онтология* предметной области (при наличии таковой).

- *Результаты разговоров* (исследований) с представителями управляющих, контролирующих и исполнительных органов в сфере применения.

- *Экспертные мнения.*

- *Текстовые документы, полученные при проектировании* подобных систем в той же или близкой предметной области.

- *Использованная терминология* – при проектировании, управлении и поддержке базы необходимых данных.

Текстометрические исследования

Для достижения цели были проведены два вида текстометрических исследований:

- *Стандартные исследования*, включающие в себя определение набора и частотных характеристик используемых понятий, их производных (с приставками и суффиксами), синонимов и омонимов, языковых конструкций, связей между понятиями – базовыми и полученными как следствие, структурой и иерархией понятий. Результаты установленных связей между понятиями имеют онтологический смысл.

- *Текстометрические исследования функций* предусматривают определение базовых понятий и их следствий, расположение понятий в логической последовательности по оси времени в соответствии с разработанной технологической схемой обработки, структурирование и оптимизацию понятий по местоположению и пр.

Формализация проблемы может быть сведена к следующему:

D – словарь терминов рассматриваемой предметной области, который может быть представлен как $D = \{d_i\}, i \in \{1, N\}$. Это множество неповторяющихся понятий, представляющих собой отдельные слова или комбинацию таких слов с самостоятельным значением. Часть из них – это базовые понятия, с помощью которых можно объяснять другие понятия, являющиеся их следствием. В контексте изложенного, понятия, представляющие собой следствие, могут рассматриваться как базовые для других понятий. Следовательно, множество D можно представить как составленное из двух взаимно пересекающихся подмножеств, т.е. $D = B \cup S$, где B – подмножество базовых понятий, а S – их следствий. В частности подмножества B и S могут полностью совпадать с D .

Пусть для заданной предметной области специфицирован набор функций $F = \{f_j\}, j \in \{1, M\}$. Выполнение каждой из них связано с определенной последовательностью понятий, необходимых для реализации основных технологических этапов обработки, например, таких как идентификация, селекция, сортировка или структурирование.

Следовательно, существует множество Q , элементы которого представляют собой цепочку понятий, связанных с выполнением функций из F . Между Q и F существует однозначное соответствие:

$$(\forall f_j \in F)(\exists q_j \in Q) \& (\forall q_j \in Q)(\exists f_j \in F).$$

В общем случае, цепочки могут пересекаться, объединяться и взаимно дополняться на основе включенных в них понятий.

При подобной постановке [2] для каждой цепочки понятий формально можно применить следующие алгоритмы структурирования:

- алгоритмы для упорядочивания понятий по оси времени при уменьшении «расстояния» между ними [3, 4]. Таким образом осуществляется оптимизация управления исполнением отдельных операций, связанных со специфицированной функцией;

- алгоритмов из [5–7] в целях определения начальных понятий при запуске каждой из функ-

ций, что, в общем случае, связано с предыдущими видами обработки.

Аналогично обеспечивается выполнение и следующих функций, а также их разновидностей [8, 9], реализованных программно в [10].

Проектирование пользовательского интерфейса

При дистанционном обучении коммуникация между обучающим и обучаемыми осуществляется посредством компьютерной системы и средств телекоммуникации. Интерфейс между обучающим и программной системой требуется в следующих случаях:

- При подготовке учебного материала и при управлении процессом обучения. Подготовка связана с двумя технологическими циклами:

- ✓ *Подбор учебного материала.* Логическая последовательность изложения может основываться на так называемой понятийной модели, в которой каждое базовое понятие должно предшествовать своим следствиям и «расстояния» между ними, измеренные по оси времени, должны быть минимальными. Следовательно, проблема сводится к структурированию понятий, т.е. к формированию цепочек понятий, соответствующих определенным методическим требованиям [11].

- ✓ *Процесс введения и актуализации учебного материала в базе данных может быть описан с помощью функционально связанных понятий, представляющих собой цепочки из команд и параметров.*

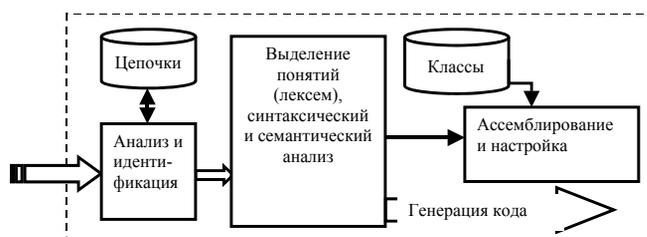
- ✓ *Дозирование учебных единиц, их изучение, оценка полученных знаний и переход к следующей учебной единице.* И здесь формирование учебных единиц базируется на подходящей селекции понятий и связанных с ними следствий. Конкретное содержание каждой учебной единицы определяется на основе предыдущего материала, уровня усвоения и целей следующего этапа обучения. Следовательно, здесь можно выделить два вида цепочек понятий:

- Цепочки понятий, непосредственно включенных в учебную единицу.

- Цепочки понятий (команд и параметров) для формирования учебных единиц и связан-

ных с ними действий по представлению, оценке и принятию последующих решений.

Возможным подходом при создании и поддержании пользовательских интерфейсов является разработка специализированных интерпретаторов. В отличие от классических вариантов, подобные интерпретаторы располагают набором цепочек понятий у себя на входе (рисунки).



Принципиальная схема специализированного интерпретатора

Такой вид специализированных интерпретаторов имеет набор предварительно разработанных цепочек понятий, связанных с выполнением конкретной специфицированной функции. Процесс интерпретации начинается с идентификации начала введенной цепочки и продолжается анализом ее развития. В общем случае цепочки могут прерываться, объединяться, разъединяться, пересекаться и т.д. Описание каждой цепочки связано с конкретным подлежащим анализу синтаксисом.

Для реализации интерпретации необходимо разработать библиотеку классов с унифицированным интерфейсом, связанных с выполнением определенных действий, порождаемых понятием или цепочкой понятий. Часть понятий может быть использована в качестве параметров для настройки классов и, соответственно, включена в генерируемый код.

Сфера приложения

Специализированный интерпретатор обрабатывает предварительно созданные цепочки понятий и с помощью библиотеки инструментальных средств ассемблирует, настраивает и генерирует код. Сфера приложения диктует вид и содержание цепочек, а также библиотеки классов, созданных на выбранном языке программирования.

Для целей дистанционного обучения специализированный интерпретатор может использоваться в случаях, когда:

- определена предметная область;
- разработан терминологический словарь используемых понятий;
- определены базовые понятия и их следствия;
- установлена иерархия понятий и их распределение по оси времени;
- разработаны функциональные спецификации системы обучения и самообучения;
- разработаны цепочки понятий в целях информационного и организационного обеспечения каждой функции;
- определены взаимоотношения между отдельными цепочками;
- разработана база данных с описанием цепочек и взаимодействия между ними;
- установлено соответствие между понятиями и библиотекой классов в целях их функциональной поддержки;
- выполнена настройка специальной части интерпретатора (анализатора и идентификатора цепочек понятий).

Заключение. В статье обсуждается возможность разработки цепочек понятий, связанных с процессом обучения и его управлением, при помощи целенаправленных текстометрических исследований учебного материала. Для этого предлагается примерная схема специализированного интерпретатора, который преобразует входной поток цепочек понятий в выполнимый код, обеспечивающий основные этапы дистанционного обучения.

Представленные результаты получены в процессе работы над научными проектами [12, 13], финансируемыми Фондом научных исследований Республики Болгария.

1. Власенко Н.А., Кузьминская Н.Л., Максименко А.А. Текстометрические исследования многоязычных научных текстов // УСиМ. – 2009. – № 2. – С. 43–47.
2. Митев М., Рачева Е., Маринова Г. Графоаналитический подход для разработки курсов дистанционного обучения // Четвертая Междунар. конф. «Новые информационные технологии в образовании для всех: инновационные методы и модели», г. Киев, Украина,

на, 24–26 нояб. 2009 г. сборник «*New information technologies in education for all*». – С. 127–136.

3. Рачева Е., Митев М., Калчева Е. Генетичен алгоритъм за структуриране на учебното съдържание за целите на дистанционно обучение // Списание «Компютърни науки и технологии», бр. 1/2008 г., ТУ-Варна, 15–20.
4. Митев М., Рачева Е. Разработка на алгоритъм и програма за структуриране на учебен материал // Там же, бр. 2/2008 г., ТУ – Варна, 15–20.
5. Митев М., Рачева Е., Маринова Г. Algorithm for Defining of Initial Package of Notions for Purpose of E-Learning. // 10th Intern. Conf. on Comp. Syst. and Technologies (CompSysTech'09), ACM ISBN 978-1-60558-986-2, ACM ICPS. – 433, P. IV. 9-1- IV. 9-5, ACM Press, N.Y. USA, 2009.
6. Митев М. Алгоритъм за определяне на финалния пакет понятия при дистанционно обучение // Годишник на ТУ – Варна, 2009. – С. 124–128.
7. Митев М., Рачева Е. Алгоритъм първоначално определя структура понятия при дистанционном обучении // Сб. докл. I Междунар. науч.-практ. конф. «Системный анализ. Информатика. Управление», г. Запорожье, Украина, 4–5 мар. 2010 г. – С. 127–128.
8. Митев М., Георгиев Д. A Priori Algorithm for Ordering Concepts in Distant Learning Courses // Intern. Conf. «Informatics in the Scientific Knowledge» – 2010, June 24–26, VFU «Chernorizets Hrabar», Varna, Bulgaria, 2010.
9. Митев М., Антонов А., Певев И. An Algorithm and Program for Management and Evaluation of Individual Distance Learning // 6th Intern. Scien. Conf. «e-Learning and Software for Education». – Bucharest. – 2010. – P. 261–266.
10. Митев М., Георгиев Д. Program Library for Development and Research of Distance Learning Courses // ICEST, Ohrid, Macedonia, 2010.
11. Михнина Н.В. Способы структурирования учебного материала как условие развития внимания // Современные наукоемкие технологии. – 2007. – № 7.
12. Проект № 483 Разработване на учебен курс по дисциплината «Синтез и анализ на алгоритми» за целите на дистанционно обучение с изследване на начините за представяне на учебния материал и създаване на специализирана база от данни с текстова и графична информация // Технически Университет – Варна, 2008.
13. Проект НП № 25 Теоретична разработка и реализация на програмна платформа за компютърно проектиране и изследване на курсове за дистанционно обучение // Технически Университет – Варна, 2009–2010.

© М.М. Митев, Е.В. Рачева, Д.З. Димитров, 2011