

Изложены разработанные модели и методы, формирующие целостный подход к тестированию и позволяющие решить следующие задачи:

- автоматическое разбиение программной системы на модули с целью их дальнейшего параллельного тестирования;
- оценка полноты тестирования с помощью динамического анализа программной системы;
- генерация тестовых воздействий при модульном тестировании;
- генерация тестовых сценариев для интеграционного тестирования объектно-ориентированных программ.

Применение предложенных методов наглядно продемонстрировано на вычислительных экспериментах на реальном программном обеспечении.

Монография будет полезна для специалистов, ответственных за проектирование и производство сложных программных продуктов высокого качества, также может использоваться в качестве учебного пособия по тестированию программного обеспечения.

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ АНАЛИЗА И МОДЕЛИРОВАНИЯ ТЕКСТОВЫХ ДАННЫХ (монография)

Ломакина Л.С., Суркова А.С.

*Нижегородский государственный технический
университет им. П.Е. Алексеева, Нижний Новгород,
e-mail: llomakina@list.ru*

Важность, значимость и необходимость анализа и обработки текстовых и других слабоструктурированных информационных данных постоянно возрастают. В связи с широким распространением систем электронного документооборота, социальных сетей, блогов, сетевых информационных порталов, персональных сайтов это становится особенно важным и как техническая задача, и как значимая часть взаимодействия людей в современном информационном мире.

Одной из основных форм представления информации является текстовая форма, наряду с графической, звуковой, а также видео информацией. Если первоначально первостепенными проблемами считались задачи, связанные с обеспечением сбора, хранения, поиска и представления данных, то в последнее время, при упрощении доступа к разнообразным коллекциям текстовых документов, появляются новые задачи анализа и обработки текстовых данных. К традиционным проблемам добавляются новые, связанные, например, с большими объемами текстовых данных в различных социальных сетях и других информационных, поисковых и аналитических приложениях Интернета.

Область, занимающаяся проблемами обработки все увеличивающегося объема текстовой

информации, получила название *Text Mining*. На русский этот термин можно перевести как Интеллектуальный анализ текстов аналогично уже устоявшемуся понятию *Data Mining* – интеллектуальный анализ данных. Предметную область *Text Mining* как согласованную совокупность задач можно описать следующим образом. К числу традиционных проблем можно отнести задачи определения авторства, автоматического извлечения ключевых слов, аннотирования и реферирования, кластеризации и классификации по тематическим категориям и т.д. К недавно возникшим задачам относятся проблема анализа текстов в глобальной сети для обеспечения безопасности и выявления потенциально опасных или нежелательных сообщений, а также задачи, связанные с многоязыковыми текстами и проблемой «переводного» плагиата и заимствования.

При всем достаточно большом количестве книг и статей, посвященных задачам *Text Mining* необходимо отметить отсутствие литературы на русском языке, узкую направленность имеющихся материалов, при этом в большинстве случаев приходится каждый раз заново решать возникающие частные задачи.

Целью настоящей монографии является рассмотрение с единых позиций общих вопросов, связанных с подходами к моделированию и обработке текстовых данных при решении разнообразных прикладных задач. Также важной особенностью книги является обсуждение возможностей и перспектив современных информационных технологий при анализе текстовых данных.

Монография написана с позиций современных информационных технологий и содержит изложение различных аспектов методологии, технологии и реализации решения основных задач анализа и обработки текстовых данных. Таким образом, книга разделена на три части: методологические, теоретические и практические аспекты анализа и моделирования текстовых данных, в каждой части материал структурирован по главам.

Первая часть посвящена проблемам построения общей методологических аспектов анализа и моделирования текстовых данных. В главе рассмотрены особенности основных задач – кластеризации, классификации и идентификации текстов. Также определены и описаны основные принципы, использование которых позволяет построить системы анализа и обработки текстовых данных. К таким принципам относятся принцип системного представления текстов, принцип нечеткой логики и принцип обучающихся систем. Рассмотрение отдельных задач с учетом основных принципов позволило представить единый методологический подход к рассматриваемым проблемам анализа и моделирования текстовых данных.

Вторая часть дает общее представление о возможностях решения основных задач анализа текстовых данных на основе сформулированных принципов. Рассмотрены основные подходы к системному представлению текстов: статистические, информационные, структурно-иерархические и другие и показаны возможности использования потокового представления текстов и использования алгоритмов сжатия при решении задач анализа текстов. Также во второй части изложены принципы и методы нечеткой логики и аспекты обучающихся систем, необходимые для корректного использования в задачах анализа и моделирования текстов.

В третьей части в разделе «Алгоритмическое и программное обеспечение анализа и обработки текстов» приведены конкретные алгоритмы анализа текстовых данных, основанные на системном представлении текстов. Рассмотрены базовые нечеткие методы кластеризации, такие как *fuzzy c-means (FCM)*, *Kernel Fuzzy Clustering* и др. и предложенные модификации алгоритмов классификации и кластеризации на основе нечетких отношений. Также в данном разделе описаны алгоритмы решения основ-

ных задач анализа текстов на основе нейронных сетей, деревьев принятия решений и подхода Random forest.

Раздел «Практическая реализация алгоритмов решения основных задач» демонстрирует примеры практической реализации рассмотренных в предыдущем разделе алгоритмах, приведены результаты кластеризации, классификации и идентификации текстовых данных. При анализе полученных результатов, сделаны выводы о перспективности и направлениях дальнейших исследований.

Следует отметить также необходимость использования английских терминов при обсуждении практически всех аспектов рассматриваемых проблем. Поскольку многие задачи и подходы к их решению обсуждаются и развиваются в нескольких направлениях и нескольких научных областях, то возникают терминологические противоречия. В данной сфере еще не завершено окончательное формирование терминологический и понятийной базы, что осложняется и различными трактовками определений при переводе на русский язык. В связи с этим для большинства вводимых понятий и терминов приводятся английский аналог.

Филологические науки

READING FOR PLEASURE 1

(пособие по чтению на английском языке)

Журавлева Л.В., Кузовлева Н.Е.,
Григоревская Е.В., Идрисова Ф.К.,
Абжалимова Р.С., Ляпустина Л.И., Ищук Е.В.

ЧУД ПО «Центр гуманитарного образования
“Лингва”, Сургут, e-mail: lingua_omp@mail.ru

Данное пособие предназначено для совершенствования техники чтения младших школьников, а также развития таких механизмов чтения, как: структурная и содержательная антиципация, догадка, логическое понимание и др. Упражнения предполагают не только многократное, коммуникативно-оправданное обращение к тексту, но и усвоение содержания прочитанного, понимание его на уровне смысла.

Интересный фактический материал, изложенный доступно и занимательно, способствует расширению кругозора детей, развитию активности в познании окружающего мира, что позволяет организовать учебный процесс на основе принципов диалогичности и культуросообразности. Предлагаемые формы организации учебной деятельности придают процессу чтения гуманистический характер и помогают генерированию смысла. В ходе выполнения заданий учащиеся узнают о традиционных домах, в которых живут люди в разных частях света, национальной кухне народов мира, совершат виртуальное путешествие в Лондонский и Московский зоопарки, обсудят вопросы, связанные с вымиранием животных на планете, а также

познакомятся с семейным укладом жизни британцев и их ценностями, научатся сопоставлять факты родной и иностранной культуры.

READING FOR PLEASURE 2

(пособие по чтению на английском языке)

Журавлева Л.В., Кузовлева Н.Е.,
Григоревская Е.В., Яровикова Е.Н., Ищук Е.В.,
Тягунова Н.И., Абжалимова Р.С.

ЧУД ПО «Центр гуманитарного образования
“Лингва”, Сургут, e-mail: lingua_omp@mail.ru

Данное пособие предназначено для совершенствования техники чтения младших школьников, а также развития таких механизмов чтения, как: структурная и содержательная антиципация, догадка, логическое понимание и др. Упражнения предполагают не только многократное, коммуникативно-оправданное обращение к тексту, но и усвоение содержания прочитанного, понимание его на уровне смысла.

Интересный фактический материал, изложенный доступно и занимательно, способствует расширению кругозора детей, развитию активности в познании окружающего мира, что позволяет организовать учебный процесс на основе принципов диалогичности и культуросообразности. Предлагаемые формы организации учебной деятельности придают процессу чтения гуманистический характер и помогают генерированию смысла. В ходе выполнения заданий учащиеся узнают о скаутском движении,