

УДК 004.5

**АВТОМАТИЗИРОВАННАЯ ОБРАБОТКА ТОВАРНЫХ ДАННЫХ ИЗ XML С
КАТЕГОРИЗАЦИЕЙ И СТАНДАРТИЗАЦИЕЙ ХАРАКТЕРИСТИК**

Трикула А.К. (ИТМО), Лаврова А.К. (ИТМО)

**Научный руководитель – доктор экономических наук, профессор Максимова Т.Г.
(ИТМО)**

Введение. Интернет-магазины регулярно получают обновленные каталоги товаров от поставщиков в формате XML. Эти файлы содержат наименования, описания, характеристики, изображения и иные атрибуты продукции. Однако различия в формате и структуре данных между поставщиками требуют обязательного приведения информации к единому стандарту для корректной загрузки на сайт. Традиционный процесс ручной обработки данных требует значительных временных и финансовых затрат, что обуславливает актуальность разработки эффективных методов автоматизации данного процесса [1].

Основная часть. В рамках исследования поставлена задача автоматизированной обработки входящих XML-каталогов, содержащих сведения о товарах для интернет-магазина. Исходные данные характеризуются разнообразием структур и перечней характеристик, поскольку каждый поставщик использует собственную систему наименований. Это приводит к несоответствиям при объединении информации и требует значительной корректировки при публикации на сайте. Для повышения удобства пользователей и улучшения навигации по ресурсу разработана унифицированная система категорий, подкатегорий и атрибутов товаров. Помимо базовых характеристик, в систему интегрированы дополнительные параметры. Реализация алгоритма обработки данных включает несколько последовательных этапов. На первоначальном этапе происходит анализ входных XML-файлов, выделение товарных позиций и извлечение соответствующих характеристик [2]. Далее применяется правило категоризации, основанное на использовании ключевых слов и специальных маркеров, что позволяет автоматически сопоставить товары с заранее определенной иерархией категорий. Особое внимание уделено проблеме стандартизации характеристик.

Выводы. Была решена проблема ручной обработки товарных данных, поступающих в формате XML от различных поставщиков.

Список использованных источников:

1. Миронов В.В., Шакирова Г.Р., Яфаев В.Э. Иерархическая модель персонализированных документов и ее XML-реализация // Вестник. – б. д. – т. 11, № 1(28). – С. 164–174.
2. Миронов В.В., Гусаренко А.С. Ситуационно-ориентированные базы данных: концепция управления XML-данными на основе динамических DOM-объектов // Вестник. – 2012. – т. 16, № 3(48). – С. 159–172.