

1. Отчет за 2016 год

к.т.н. Черкашин Е.А.

Заявлено в планы на 2016 год. *Разработка методов извлечения онтологии, как грубой модели информационной системы, из неформализованных описаний системы.*

Разработана методика автоматизированного построения концептуальной модели предметной области, состоящей из трех этапов. В основу методики входит полисистемное онтологическое представление предметной области, где составляющие онтологии представляют собой систему слоев (категорий), отображаемых друг при помощи функторов. Полисистемное представление позволяет создавать концептуальный базис, общий для разных аспектов моделирования предметной области.

На *первом шаге* методики в текстах входных документов автоматически выделяется набор ключевых слов и производится построение иерархической классификаций входных документов согласно пространству, заданному выделенными ключевыми словами. На *втором шаге* осуществляется поименование узлов иерархической классификации наиболее часто встречающимися ключевыми словами в соответствующем подмножестве, характеризующем поддерево классификации. Алгоритмическое обеспечение для первого и второго шага в той или иной форме уже существует и адаптируется для обработки текстов предметной области. Результатом второго этапа является онтология, тезаурус, где задан словарь и отношение между концептами вида «is-a» («является»).

На *третьем шаге* производится сопоставление сгенерированного тезауруса терминологическому базису используемой полисистемы онтологий. В результате такого сопоставления тезаурус обогащается как новыми «горизонтальными» отношениями, так и производится его оценка на полноту относительно отношений, заданных в полисистеме онтологий. После анализа полученной онтологии производится ее интегрирование в полисистему в виде нового слоя.

Методика предусматривает повторение методики над частями текста более мелкого масштаба (переход от документам к их разделам, от разделов к отдельным таблицам и строкам) с целью дальнейшей декомпозиции структур предметной области.

Основная цель исследований – разработка методики рекурсивной декомпозиции предметной области и ее привязка к существующим слоям полисистемы онтологий. Такая привязка позволяет переносить существующие программные реализации теорий слоев на новые слои по образцу и подобию.

В 2016 году в рамках исследования разработана программная инфра-

структура для хранения слоев полисистемы с использованием существующих стандартных онтологий. Слой представляет собой часть онтологии (ядро), которая подвергается отображению в другой слой через функтор или является результатом такого отображения. Слои хранятся в онтологической базе в виде графов. Хранилище реализовано при помощи системы ClioPatria. Система ClioPatria реализована на языке Prolog, что позволяет создавать логические теории слоев непосредственно на сервере. Кроме того, декларативная сущность Prolog-а значительно упрощает задачу переноса реализаций теорий между слоями.

Разработанная распределенная система применена в решении задачи автоматизации анализа структуры текстов учебных программ технического вуза. Для этого произведена разметка абзацев учебных программ обозначениями концептов, задающих семантическое значение материалу, представленному в абзаце или таблице. Затем выполнено машинное обучение, направленное на распознавание концепта в зависимости от свойств, выделенных в абзаце (наличие ключевых слов, определенных словосочетаний и стилей представления текста). Результатом применения синтезированной системы знаний является семантическая разметка документа, которая отображается на слои полисистемы, что в дальнейшем позволяет применять к элементам документов те или иные алгоритмы обработки данных.