

**Импорт данных из тезауруса в семантическую сеть**  
**Волненко Д.А. (ИТМО)**  
**Научный руководитель – старший преподаватель Цопа Е.А.**  
**(ИТМО)**

**Введение.** Семантические сети представляют собой ориентированные графы, в которых узлы обозначают понятия, а связи между ними отражают различные типы отношений, такие как гиперонимия (переход от частного к общему), синонимия (лексическая близость) и антонимия (противоположность значений). Они широко применяются в интеллектуальных системах обработки информации, включая машинный перевод, поиск и рекомендации, обеспечивая структурированное представление знаний [1].

Одним из ключевых вызовов при использовании семантических сетей является заполнение их данными. Готовые тезаурусы содержат ценные лексические связи, однако их интеграция требует значительных усилий, так как существующие ресурсы различаются по структуре, полноте и качеству данных [2]. Кроме того, многие тезаурусы требуют ручной обработки перед импортом, что делает процесс трудоемким и малоэффективным. Автоматизация этого процесса позволила бы значительно ускорить и упростить пополнение семантических сетей, повышая их связность и точность представления знаний.

**Основная часть.** Разработан алгоритм автоматизированного импорта данных в семантическую сеть. Процесс начинается с проверки доступности ресурсов для обращения к API тезауруса. При наличии ресурсов формируется и отправляется запрос, а полученные данные проходят валидацию на соответствие структуре и корректность формата.

После валидации осуществляется фильтрация данных: исключаются нерелевантные связи и дублирующие элементы. Оставшиеся данные преобразуются в узлы, которым присваиваются уникальные идентификаторы. Между узлами формируются связи в соответствии с их семантическими типами (гиперонимия, синонимия, антонимия и др.).

Перед интеграцией в семантическую сеть проводится дополнительная проверка корректности связей. Обнаруженные ошибки, включая несоответствие типов узлов или нарушение структуры сети, фиксируются и отправляются на повторную обработку. Корректные данные добавляются в хранилище семантической сети.

**Выводы.** Разработан и успешно протестирован алгоритм автоматизированного импорта данных в семантическую сеть. Реализованные механизмы валидации, фильтрации и интеграции обеспечивают корректное формирование узлов и связей, что повышает структурированность и полноту данных.

**Список использованных источников:**

1. R. Navigli, S.P. Ponzetto. BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. Artificial Intelligence Journal // 2012 / 12. С. 217-250.
2. Цопа Е.А., Клименков С.В., Харитонов А.Е., Письмак А.Е. Оценка семантической близости предложений на естественном языке методами математической статистики // Научно-технический вестник информационных технологий, механики и оптики // 2016 / 02.