

Отримані результати моделювання дозволяють стверджувати, що застосування нечіткої логіки при побудові моделі процесу нормалізації глікемії дало змогу внести плавність в коливання рівня глюкози, і тим самим наблизити їх до природніх процесів в організмі здорової людини. Ці характеристики є дуже важливими задля ефективного контролю інсулінозалежного цукрового діабету.

Розробки штучної підшлункової залози на сьогодні поки ще не закінчені, і хворі на цукровий діабет не мають змоги користуватись подібними приладами. Існуючі засоби ін'єкцій інсуліну дуже дорогі та все ще потребують від користувача ручного вводу даних. Впровадження моделей нормалізації глікемії на базі нечіткої логіки в поєднанні із засобами автоматичного вимірювання рівня глюкози дозволять покращити якість життя людям, котрі хворіють на цукровий діабет.

**Дмуховский Р.И., Орлова Е.В.**

**Науч. руководитель к.т.н., доц. Егошина А.А.**

*Донецкий национальный технический университет*

**Автоматическое построение онтологии на основе  
результатов предварительной кластеризации  
коллекции текстовых документов**

В настоящее время, можно наблюдать бурный рост и развитие технологий Semantic Web. Одним из перспективных направлений в исследованиях является использование онтологий для решения задач интеграции данных. Методы интеграции данных на основе онтологий показали на практике свою эффективность, однако построение онтологии требует экспертных знаний в исследуемой предметной области и занимает существенный объем времени, поэтому актуальной задачей является

автоматизация процесса построения онтологии.

Согласно определению Т. Грубера, онтология – это спецификация концептуализации предметной области [1]. Среди уже разработанных онтологий наиболее известными и объемными являются CYC (<http://www.cyc.com>) и SUMO (<http://www.ontologyportal.org/>).

Существуют множество подходов к автоматизации процесса построения онтологий [2-5]. Рассмотрим некоторые из них.

### **1. Построение семантической карты ресурса**

В данном методе для автоматизации процесса построения онтологии предлагается использовать текстовое содержание массива Веб ресурсов описательного характера определенной тематики [3].

Базовой является задача разработки алгоритма автоматического построения семантической карты веб ресурса с помощью анализа его текста. Семантическая карта ресурса – это отображение контента Веб ресурса в концептуализацию его содержания, представленное в виде OWL онтологии.

Семантическая карта ресурса строится на основе особенностей языка, которые позволяют вытягивать семантические конструкции из текста.

Семантическая карта строится в два этапа, на первом строится формальная семантическая OWL конструкция, на втором происходит привязка полученной конструкции к конкретной предметной области. Формулируются правила, использующие синтаксис языка. Правила синтаксического уровня, выявляют семантику на основе принципов построения словосочетаний и предложений.

Для того чтобы привязать полученную семантическую модель к интересующей предметной области, используется словарь соответствующей тематики.

В итоговой онтологии фиксируются только те семантические конструкции, в которых участвуют термины из словаря предметной области.

## **2. Автоматическое построение онтологии по коллекции текстовых документов**

В работе [5] предлагается подход к решению проблемы автоматического построения онтологий, преимущественно основанный на статистических методах анализа текстов на естественном языке.

Построение онтологий разделено на 3 этапа: предварительная подготовка коллекции, определение классов онтологии, определение отношений «is-a» и «synonym-of», построение иерархии классов.

На первом этапе построения онтологии требуется выделить входящие в ее состав классы. Данная задача сводится к определению терминов рассматриваемой предметной области.

Алгоритмы извлечения терминов из текстов на естественном языке можно разделить на две группы: статистические и лингвистические. Предполагается, что существующие статистические методы могут показать лучшие результаты, если дополнить их определенными эвристиками.

Этап выделения отношений между классами создаст наибольшие трудности. В связи с чем, первоначально имеет смысл говорить об автоматическом тезаурусе (таксономии с терминами). В качестве базовых отношений, действующих между терминами, определим отношения «is-a» и «synonym-of».

Для выделения отношения «is-a» можно воспользоваться количественным подходом к информации. Для этого было использовано предположение, что количество информации термина из нескольких слов больше, чем количество информации отдельных слов,

входящих в его состав.

Предложенный подход позволяет выделить только базовые отношения, однако предполагается, что возможно его расширение для выделения других отношений.

### **Предлагаемое решение**

На основе онтологии можно проводить интеграцию данных. В ходе исследований было выявлено, что интеграция на основе автоматически построенной онтологии проходит значительно проще, быстрее и качественнее. В частности, для интеграции текстовых данных, наиболее подходящим является метод построения онтологии по коллекции текстовых документов.

Также, в результате исследований было установлено, что на качество построения онтологии влияет предварительная подготовка текста, в частности, особенности коллекции документов. Кластеризация документов по общей тематике может сократить время, затрачиваемое на создание онтологии.

В качестве алгоритма кластеризации предлагается алгоритм LSA/LSI. Алгоритм LSA/LSI – это реализация основных принципов факторного анализа применительно ко множеству документов. Данный метод кластеризации позволяет успешно преодолевать проблемы синонимии и омонимии, присущие текстовому корпусу основываясь только на статистической информации о множестве документов/терминов.

### **Литература.**

1. Соловьев В.Д., Добров Б.В., Иванов В.В., Лукашевич Н.В. Онтологии и тезаурусы. Учебное пособие. Казань, Москва. 2006;
2. Крывый С.Л., Ходзинский А.Н. Автоматное представление онтологий и операции на онтологиях / Интернет-ресурс. –

Режим доступа: <http://shcherbak.net/avtomatnoe-predstavlenie-ontologij-i-operacii-na-ontologiyax>.

3. Рабчевский Е.А. Автоматическое построение онтологий / Интернет-ресурс. – Режим доступа: <http://shcherbak.net/avtomaticheskoe-postroenie-ontologij>.

4. Рабчевский Е.А. Автоматическое построение онтологий на основе лексико-синтаксических шаблонов для информационного поиска // Труды 11 Всероссийской науч. конференции «Электронные библиотеки: перспективные методы и технологии, электронные коллекции» – Петрозаводск, 2009. – С. 69-77.

5. Мозжерина Е. С. Автоматическое построение онтологии по коллекции текстовых документов // Электронные библиотеки: Перспективные Методы и Технологии, Электронные коллекции – Воронеж, 2011 – С. 293 – 298.