

Содержание

Введение	2
Глава 1. Семантическая сеть	4
1.1. Введение в семантические сети	4

Введение

Достижения в области биологических наук наряду с нарастанием объемов доступной для использования информации повышают необходимость в интеграции разрозненных источников данных. К ним относятся и интернет, и ведомственные, и корпоративные системы:

1. Медстатистика
2. Результаты клинических испытаний
3. Электронная история болезни
4. Фармацевтика
5. Разработка лекарственных средств

В настоящее время в России разработано и используется большое количество разрозненных медицинских информационных систем, разнообразных баз данных с описанием лекарственных препаратов, результатами научных исследований; написано множество научных трудов и статей в области здравоохранения (медицина, фармацевтика, медицинское страхование и др.), хранящихся в специализированных электронных библиотеках в различных форматах. Однако эффективные механизмы извлечения из таких источников знаний, хранения и предоставления к ним широкого доступа отсутствуют.

Требуется фундаментальный сдвиг от единичных попыток интеграции к единой функциональной области. Для решения этой проблемы разработан стандарт публикации данных Linked Data. Одним из важнейших достоинств этой технологии является ее открытость - возможность объединения в общую семантическую сеть распределенных семантических хранилищ,

созданных различными организациями (органы управления здравоохранением, ВУЗы, НИИ, МО) и профессиональными сообществами (ассоциации кардиологов, анестезиологов, медицинских IT-специалистов и др.) на основе единых открытых стандартов. Как показывает международный опыт, это позволяет системе саморазвиваться, постоянно пополняя количество доступных знаний и повышая их качество.

Межресурсные ссылки дают исследователям возможность перемещаться между источниками данных и открывать связи, которые не были замечены ранее. Существуют универсальные инструменты, такие как семантические веб-браузеры и поисковые движки, которые могут использоваться для задач представления и поиска данных.

Основная цель моей работы - это создание семантического хранилища медицинских знаний.

Для достижения поставленной цели решались следующие задачи:

1. Скачивание и парсинг информации с ресурса Webapteka
2. Разработка онтологии лекарственных препаратов
3. Конвертация html данных в rdf представление
4. Разработка SPARQL-запросов для извлечения информации и выявления дополнительных связей в RDF-хранилище.
5. Разработка пользовательского интерфейса
6. Кеширование элементов приложения для повышения производительности

Глава 1

Семантическая сеть

1.1. Введение в семантические сети

Семантическая сеть (англ. Semantic Web) — это набор технологий, позволяющих представлять информацию в виде пригодном для машинной обработки: RDF, OWL, SPARQL. RDF используется для представления информации, SPARQL - для доступа к ней, OWL - добавляет метайнформацию, связи между концептами.

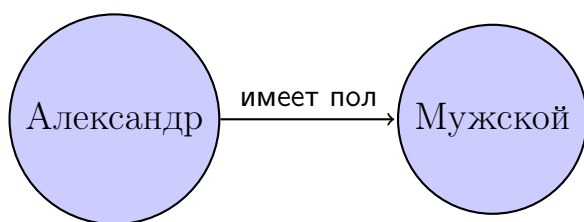
В RDF вся информация представляется в виде триплетов: субъект, предикат, объект. Триплеты по форме похожи на простое предложение. Например:

Субъект: Александр

Предикат: Имеет пол

Объект: Мужской

Триплет может быть выражен в виде графа



Субъекты и объекты могут быть представлены URI, либо литералом. URI - это уникальный идентификатор, который обозначает сущность: например URI для собаки может быть таким *'http://example.ru/animals/dog'*. Литерал - это просто строка, например *'Jack Nickolson'*, с возможными добавлениями, указывающими язык, тип данных (поддерживаемый XML, такие как integer и datetime). В идеале каждая между сущностями и URI

составлено взаимно однозначное соответствие: каждый URI принадлежит только одной сущности и каждая сущность имеет только один URI.