

# **Разработка рекомендательной модели для системы ведения проектов и портфолио**

*Ключевые слова:* рекомендательные системы, онтологии, открытое ПО, семантическая паутина, семантический граф

## **Аннотация**

С развитием идеи открытости знаний, и совместной работы над проектами, начинают появляться площадки и сообщества энтузиастов. Зачастую коллективы многочисленны, а в их рамках разрабатывается множество проектов, что приводит к появлению большого количества открытой неструктурированной информации и затрудняет обмен опытом. В данной работе рассмотрена задача построения системы для рекомендации участникам проектов и наоборот, основываясь на наличествующих и требуемых навыках, согласно семантике предметных областей. Приведены существующие подходы к решению задачи рекомендации на основе семантических данных. А так же приведены основные технологии для работы с онтологиями

## **Введение**

На базе университетов, открытых площадок и коворкингов часто собираются сообщества энтузиастов, занятых воплощением идей и созданием проектов. Вместе с тем идея открытого ПО трансформируется и начинает применяться для проектов в инженерной сфере, что приводит к появлению большого количества проектов с открытыми материалами (исходными кодами, чертежами и т.п.) С ростом количества проектов, растёт количество информации и знаний, которые могут пригодиться участникам. Это приводит к проблеме учёта проектов, участников, внутри сообществ, и обмен опытом в его рамках. Для решения этой проблемы внутри одного из таких сообществ была начата разработка система ведения проектов и портфолио энтузиастов. В данной работе рассматривается разработка одного из её модулей - рекомендательной системы.

## **Актуальность**

В системе ведения проектов существует проблема обмена опытом. В условиях большого числа проектов перед участником встаёт задача найти среди них тот, в котором ему будет интереснее или полезнее всего поучаствовать. Аналогичная задача стоит перед командами проекта в случае поиска новых участников. Т.о. рекомендательная система должна решить эту проблему, облегчить процесс обмена опытом, делая выводы о том на сколько участникам будет интересно либо полезно участвовать в том или ином проекте.

## **Проблематика**

В системе проекты и участники характеризуются навыками и знаниями. В случае участников можно говорить и имеющихся у них навыков и о их интересах, т.е. навыки, которые они хотели бы получить. В случае проектов, в них есть решённые и не решённые задачи, т.о. проекты можно трактовать и как источник навыков (объект рекомендации) и как субъект рекомендации.

Навыки и области знаний формулируются в терминах онтологии. Таким образом рекомендуемая система должна учитывать семантику знаний и предметных областей для решения задачи рекомендации.

## **Цель**

Разработка рекомендательной системы под нужды участников открытых сообществ-агрегаторов проектов.

## **Задачи**

- 1) Изучить существующие системы рекомендации
- 2) Изучить алгоритмы, пригодные для решения проблемы
- 3) Сформировать список инструментов для разработки прототипа

## **Обзор предметной области**

В данном разделе рассматриваются подходы к построению рекомендуемых систем в других работах из близких областей.

В таб. 1 приведено сравнение аналогов по критериям, которые будут сформулированы ниже.

## **Принцип отбора аналогов**

Рассматривались работы, затрагивающие создание рекомендательных систем на основе онтологий, других систем, затрагивающих вопросы построения и обработки онтологий и выражения сущностей в семантических терминах, а также рекомендательных систем, использующих информацию о специфике пользователя(субъект рекомендации) либо объекта рекомендации.

Формирование профиля пользователя. Конкретные приложения Использует статистические методы: - Учитывает схожесть объектов рекомендации (Методы, основанные на фильтрации контента)

- Учитывает схожесть субъектов рекомендации (Методы коллаборативной фильтрации)

- Использует контекст

должны учитывать экономические, психологические и другие факторы (рекомендующие системы 2-поколения)

- Использует семантику (рекомендующие системы 3-поколения)

Использование семантических категорий естественного языка для концептуализации, представления и использования знаний, вовлекаемых в процессы выработки решений  
- Вовлекает знания о предметной области (терминах семантических категорий)

### ***1) Recommender Systems: International Joint Conference on Artificial Intelligence [1]***

В данной работе рассматриваются классические системы рекомендаций. В ней рассматриваются методы, основанные на фильтрации контента (множества свойств), описывающего продукты/услуги, которые потенциально могут быть рекомендованы пользователю, а также методы коллаборативной фильтрации, в которых полагается, что текущие предпочтения пользователя будут сходны с предпочтениями других пользователей с похожими интересами по отношению к предлагаемому товару или услуге.

### ***2) Incorporating contextual information in recommender systems using a multidimensional approach[2]***

Данная работа ориентирована на детальный анализ сущности контекста, моделей его формального представления и учета в процессах выработки рекомендаций в рамках системы. Работа представляет многомерный подход к рекомендующим системам, ориентированный на детальный анализ сущности контекста, моделей его формального представления и учета в процессах выработки рекомендаций в рамках системы, помимо неопосредованной информации о предпочтении пользователя к товарам, либо о информации отношения групп пользователей к определённому товару.

### ***3) Personalized Web Search by Mapping User Queries to Categories[3]***

Работа формирует разделы интересов пользователя по множеству веб-документов его истории поиска и использует базу данных понятий Open Directory Project (ODP). Далее профиль пользователя представляется в виде множества найденных категорий с весами, пропорциональными частоте их появления в документах.

### ***4) Deriving Concept-Based User Profiles from Search Engine Logs[4]***

Работа концентрируется на формировании профиля пользователя для целей персонализации. В ней рассматриваются положительные и отрицательные предпочтения пользователя для более взвешенных решений о схожести/несхожести пользователей для целей кластеризации. По каждому запросу авторы сначала извлекают множество понятий на основе служебной информации, затем выполняется кластеризация найденных понятий по некоторой мере сходства и построение на этой основе персонифицированного профиля пользователя.

## **5) A New Ontology-Supported and Hybrid Recommending Information System for Scholars[5]**

Работа предлагает гибридную рекомендательную систему на основе онтологий для учёных. Её особенностью является возможность дополнять предметные области и извлекать значимые метрики для фильтрации и ранжирования, что позволяет легко интегрировать её в существующие системы

### **Критерии сравнения аналогов**

#### ***Учитывает схожесть объектов/субъектов рекомендации***

Методы рекомендации, учитывающие сходство объектов(напр.товаров) и субъектов(пользователей) входят в одно семейство, иногда называемое системами 1-го поколения. Данные методы являются статистическими, и слабо учитывают персональные предпочтения пользователя.

#### ***Использует контекст***

Семейство рекомендательных систем, учитывающих контекст объединены названием систем 2-го поколения. В качестве главных измерений (атрибутов) контекста рассматриваются обычно время, место и социальный контекст. Они способны вырабатывать рекомендации в контексте социальных сетей, использовать тэги в качестве источника знаний и профиле пользователя и т.п.

#### ***Использует семантику***

Системы, попадающие под это критерий, ориентируются на семантические аспекты моделей интересов и профиля пользователя, могут учитывать экономические, психологические и другие факторы, определяющие выбор пользователя, объединять в себе знания о предметной области рекомендаций и о персональном многомерном профиле пользователя.

#### ***Использует стороннюю информацию***

Данный критерий подразумевает привлечение сторонних источников информации для дачи рекомендаций, помимо профиля пользователя, либо вопросы построения профиля на основе сторонней информации из различных источников

### **Таблица 1 - сравнение по критериям**

++ - решающий критерий, акцент работы

Критерий\Аналог	[1]	[2]	[3]	[4]	[5]
Учитывает схожесть объектов/субъектов	+	+	-/(только субъект)	++	-
Использует контекст	-	+	-	-	-
Использует семантику	-	-+ идейный предшественник концепции использования семантики	+	-	++
Использует стороннюю информацию	-	-	+	+	-

## Выводы по итогам сравнения

Сравнение показало, что работы, представляющие интерес для решения задачи как правило являются гибридными, т.е. используют более одной модели. В частности прослеживается тенденция к привлечению семантики области для более аргументированной рекомендации при меньшем количестве данных, что потенциально имеет большое значение для предметной области данной работы. Так же было обнаружено, что в большинстве предметных областей возможно использование CF, CBF в качестве гибрида

Исследованные работы показали, что решения, использующие семантику предметной области наиболее подходят для решения поставленной задачи. Это связано с тем, что в разрабатываемой системе проекты(объекты рекомендации) и участники(их профили, субъекты) наиболее точно логически представимы в виде набора формальных понятий, семантически связанных друг с другом.

## Выбор метода решения

Имея представление о принятых подходах к задаче построения рекомендательных систем на основе онтологий, а так же набор требований, требуется отобрать подходящие решения. Исследование состоит в создании прототипов предложенных системы, а затем тестировании её на данных системы ведения проектов и портфолио.

Среди рассмотренных систем отмечено два основных подхода - knowledge-based[5], и подходы с использованием контекста коллаборативной фильтрации. Для последних необходимо выделение меры (рейтинга) объекта фильтрации у субъекта. В исходных данных отсутствует данная информация в чистом виде, что значит, что необходимо выделить метрики из семантики исходных данных[9,10].

## **Описание метода решения**

Базовая идея системы состоит в том, что понятие интерес пользователя рассматривается как экземпляр некоторой категории в иерархии категорий (понятий) онтологии, а профиль пользователя – как некоторая структура, заданная на множестве интересов-категорий онтологии, возможно, структурированных иерархически. Для построения онтологии предметной области рекомендаций и профиля пользователя используются одни и те же источники данных.

Такой набор исходных данных позволяет работать с ним как с учетом семантики интересов, предметных областей и пр. так и классическими рекомендательными системами, путём выделения метрик из онтологии, и использования их в качестве исходных данных

## **Используемые инструменты**

В этом разделе представлены инструменты, используемые для работы с онтологиями.

### **Средства разработки**

#### ***Python***

Для исследования работы моделей на онтологиях выбран ЯП Python из-за скорости прототипирования и разработки. Он используется для построения прототипов рассмотренных систем.

#### ***Java***

Менее подходит для прототипирования и исследований, однако некоторые из используемых инструментов предназначены для работы с java

### **Онтологии**

#### ***OWL***

OWL это формат для представления онтологий и работы с ними. Он поддерживается многими инструментами (некоторые перечислены далее), и основан на xml, что позволяет легко ссылаться на объекты с внешних ресурсов через uri.

#### ***Protégé[6]***

Открытый редактор онтологий и фреймворк для построения баз знаний. Предоставляет обширные возможности по построению, визуализации, обработке онтологий. Поддерживает работу с форматом OWL и ссылки на внешние онтологии

## **OwlReady[7]**

Библиотека для Python, предоставляет API для работы с онтологиями. Работает с форматом OWL и внутренним сериализуемым представлением.

## **Apache Jena[8]**

Веб-фреймворк для построения семантической паутины, предоставляющий API на Java. Имеет более широкий функционал по сравнению с библиотекой OwlReady: поддерживает основные форматы(языки) представления онтологий (OWL, RDF, OIL), поддерживает Семантический механизм рассуждений (Reasoner) - инструмент для построения логических цепочек, вывода новой информации по правилам формальной логики

## **Заключение**

В данной работе были сделаны первые шаги для построения рекомендующей системы для системы ведения портфолио и проектов. Были рассмотрены виды рекомендующих систем и подходы к их построению. Рассмотренные системы отнесены к одной и нескольким групп, по информации, участвующей в принятии решения: использование колаборативной фильтрации, контекста, семантики. Данный анализ дал понимание того, какие требования предъявляются к будущей системе рекомендации. Определено, что для решения задачи наилучшим образом подойдут системы, использующие семантику предметной области. Также выяснены преимущества гибридных систем, использующих более чем один подход. Для реализации гибридного подхода возможно выделение колаборативных метрик из семантического графа.

Был проведен обзор инструментов, для построения прототипов на основе рассмотренных решений и исследования результатов, чтобы сделать вывод о применимости подхода для решения задачи.

В качестве ЯП выбраны Python из за простоты прототипирования и Java из за специфичных инструментов. В качестве формата представления онтологий был взят OWL, как язык с поддержкой всеми рассмотренными инструментами для работы с семантической информацией.

Для исследований и проверки гипотез взято приложение Protégé, предоставляющее удобный инструментарий для манипулирования семантическими графами.

На данном этапе, имея постановку задачи в терминах онтологии, формирование требований к рекомендующей системе и формирование списка технологий позволяет приступить непосредственно к прототипированию.

## **Список литературы**

- 1) Jannach, D. Tutorial: Recommender Systems / D. Jannach, G. Friedrich // International Joint Conference on Artificial Intelligence.(Beijing, August 4, 2013).
- 2) Adomavicius, G. Incorporating contextual information in recommender systems using a multidimensional approach / G. Adomavicius, R. Sankaranarayanan, S. Sen, A. Tuzhilin // ACM Transactions on Information Systems

- 3) Liu, F. Personalized Web Search by Mapping User Queries to Categories / F. Liu, C. Yu, W. Meng // In Proc. of Intern. Conf. on Information and Knowledge Management (CIKM)
- 4) Leung, K.W.T. Deriving Concept-Based User Profiles from Search Engine Logs / K.W.T. Leung, D.L. Lee // IEEE Transaction on Data and Knowledge Engineering.
- 5) Yang, S.-Y., & Hsu, C.-L. (2010). A New Ontology-Supported and Hybrid Recommending Information System for Scholars. 2010 13th International Conference on Network-Based Information Systems. doi:10.1109/nbis.2010.27
- 6) Protégé 5. Документация [Электронный ресурс]. URL:<https://protegeproject.github.io/protege/>
- 7) Owlready2. Документация [Электронный ресурс]. URL: <https://owlready2.readthedocs.io/en/v0.35/>
- 8) Apache Jena. Документация [Электронный ресурс]. URL: <https://whimsy.apache.org/board/minutes/Jena.html>
- 9) Gan M., Dou X., Jiang R. From Ontology to Semantic Similarity: Calculation of Ontology-Based Semantic Similarity // The Scientific World Journal. 2013. Vol. 2013. P. 1–11. URL: <https://www.hindawi.com/journals/tswj/2013/793091/>
- 10) Guisheng Y., Qiuyan S. Research on Ontology-Based Measuring Semantic Similarity // 2008 International Conference on Internet Computing in Science and Engineering. IEEE, 2008. P. 250–253