|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РФ | | | | |
|  |  | | |  |
| Пермский государственный национальный  исследовательский университет | | | | |
|  |  | | |  |
|  | **разработка ОНТОЛОГИЧЕСКОЙ БАЗЫ ЗНАНИЙ**  *Практическая работа по дисциплине «Инструментальные средства создания оболочек экспертных систем»* | | |  |
|  |  | | |  |
|  |  |  | Работу выполнил  студент гр. ПМИ-3 4 курса механико-математического факультета  Скоробогатова М.М. \_\_\_\_(подпись)  «\_\_» \_\_\_\_\_\_ 2021 |  |
|  |  |  | Работу проверил  старший преподаватель кафедры МОВС  Леонтьева Т.А. \_\_\_\_\_ (подпись)  «\_\_» \_\_\_\_\_\_ 2021 |  |
|  |  |  |  |  |
| Пермь 2021 | | | | |

СОДЕРЖАНИЕ

[Введение 3](#_Toc91537032)

[Глава 1. Этап идентификации 4](#_Toc91537033)

[Глава 2. Этап концептуализации 5](#_Toc91537034)

[Глава 3. Этап формализации 9](#_Toc91537035)

[Глава 4. Построение онтологии 11](#_Toc91537036)

[Заключение 13](#_Toc91537037)

Введение

Цель: разработка онтологической базы знаний для проблемной области задачи, решаемой в рамках выпускной работы.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Выполнить этап идентификации.
2. Выполнить концептуализацию онтологической БЗ.
3. Выполнить формализацию онтологической БЗ.
4. Реализовать онтологическую БД в среде визуального инструментального средства построения онтологических БЗ ОНТОЛИС.

Отметим, что по условиям задания разрабатываемая онтологическая БЗ должна дополнять одну из двух предоставленных онтологических БЗ.

# Этап идентификации

В разрабатываемой онтологической БЗ должны быть представлены знания о проблемной области задачи, решаемой в рамках выпускной работы.

В рамках выпускной работы решается задача разработки настольное приложение BibReader, модули которого автоматизируют обработку метаданных публикаций. Следовательно, онтологическую БЗ необходимо разработать для проблемной области «Автоматизация обработки метаданных публикаций».

Выделим основные понятия, связанные с рассматриваемой проблемной областью.

# Этап концептуализации

Выделим основные понятия в проблемной области, опираясь на описание решаемой задачи, методов решения задачи и связанных с ними понятий.

Начнём с того, что разрабатывается настольное приложение BibReader, модули которого автоматизируют обработку метаданных публикаций. Следовательно, выделим три центральных понятия — «BibReader», «Модуль обработки метаданных» и «Публикация».

BibReader представляет из себя настольное приложение, написанное на языке программирования C#, поэтому вводим понятия «Настольное приложение», «Язык программирования», «C#».

В BibReader реализованы модули, предназначенные для следующих задач: чтение файла, скрининг, составление библиографической ссылки, фильтрация, сбор статистики, классификация. Добавим в список понятия, соответствующие данным задачам.

Модуль «Чтение файла» принимает на вход файлы формата .csv и .bib, полученные в результате поиска документов в электронных библиотеках. Таким образом, введем понятия «Файл», «.bib», «.csv», «Поиск документов», «Электронная библиотека».

В настоящее время BibReader выполняет работу с результатами поиска в следующих электронных библиотеках: ACM DL, ASTS, IEEE, Inspec, Scopus, SpringerLink, Web of Science. Введем для каждой из них соответствующее понятие.

Модуль «Чтение файла» на выходе выдает первичный корпус документов, который затем направляется в модуль «Скрининг», который состоит из двух этапов: поиск уникальных документов и поиск релевантных документов. Первый этап отдает на выходе корпус уникальных документов, который затем идет на вход второму этапу, после которого на выходе на основе требований релевантности получается корпус релевантных документов. Отсюда введем понятия «Корпус документов», «Корпус уникальных документов», «Корпус релевантных документов», «Поиск уникальных документов», «Поиск релевантных документов», «Требования релевантности».

Модуль «Составление библиографической ссылки» принимает на вход любой из корпусов документов и на выходе отдает библиографические ссылки различных стилей: ГОСТ, IEEE, APA и Harvard. Введем понятия «Библиографическая ссылка», «Стиль», «ГОСТ», «IEEE», «APA», «Harvard».

Модуль «Классификация» принимает на вход любой из корпусов документов и на основании полей метаданных «Аннотация», «Ключевые слова» и «Название» составляет облаков тегов. Введем понятия «Метаданные», «Аннотация», «Ключевые слова», «Название», «Облако тегов».

Модули «Сбор статистики» и «Фильтрация» также принимают на вход любой из корпусов документов и по полям метаданных «Год», «Источник», «Тип документа», «Журнал», «Конференция», «География» и «Количество авторов», которые используются в качестве статистики и фильтра, проводят сбор статистики и фильтрацию документов соответственно. В результате сбора статистики получается файл .xlsx со статистикой, разложенной по таблицам, в результате фильтрации возвращается отфильтрованный корпус документов. Введем понятия для приведенных выше полей метаданных, а также понятия «Статистика», «Фильтр», «.xlsx» и «Отфильтрованный корпус документов».

Дадим определения всем введённым понятиям.

Все введённые понятия и их определения могут быть структурированы в виде списка.

1. «.bib» – формат файла, содержащий метаданные публикации.
2. «.csv» – формат файла, содержащий метаданные публикации, разделенные запятыми.
3. «.xlsx» – формат файла, содержащий табличные данные.
4. «ACM DL» – электронная библиотека.
5. «APA» – стиль библиографической ссылки.
6. «ASTS» – электронная библиотека.
7. «BibReader» – настольное приложение, автоматизирующее обработку метаданных публикаций.
8. «C#» – объектно-ориентированный язык программирования.
9. «Harvard» – стиль библиографической ссылки.
10. «IEEE» – электронная библиотека.
11. «IEEE» – стиль библиографической ссылки.
12. «Inspec» – электронная библиотека.
13. «Scopus» – электронная библиотека.
14. «SpringerLink» – электронная библиотека.
15. «Web of Science» – электронная библиотека.
16. «Аннотация» – часть метаданных публикаций, соответствующее отрывку из текста публикации.
17. «Библиографическая ссылка» – совокупность библиографических сведений о документе.
18. «География» – поле метаданных публикации, соответствующее географической принадлежности публикации.
19. «ГОСТ» – стиль библиографической ссылки.
20. «Год» – поле метаданных публикации, соответствующее году выхода публикации.
21. «Журнал» – поле метаданных публикации, соответствующее названию журналу, в котором была опубликована публикация.
22. «Источник» – поле метаданных публикации, соответствующее электронной библиотеке, в которой была найдена публикация.
23. «Классификация» – модуль, формирующий облако тегов для корпуса документов по какому-либо полю метаданных.
24. «Ключевые слова» – поле метаданных публикаций, соответствующее ключевым словам в тексте публикации.
25. «Количество авторов» – поле метаданных публикации, соответствующее количеству авторов публикации.
26. «Конференция» – поле метаданных публикации, соответствующее названию конференции, на которой была представлена публикация.
27. «Корпус документов» – список публикаций внутри BibReader.
28. «Корпус релевантных документов» – корпус документов, из которого были удалены документы, не соответствующие требованиями релевантности.
29. «Корпус уникальных документов» – корпус документов, из которого были удалены дубликаты.
30. «Метаданные» – данные, характеризующие публикацию.
31. «Модуль» – часть приложения BibReader, выполняющая ту или иную задачу по обработке метаданных публикаций.
32. «Название» – поле метаданных публикаций, соответствующе названию публикации.
33. «Настольное приложение» – это программа, предназначенная для выполнения определенных пользовательских задач и рассчитанная на непосредственное взаимодействие с пользователем.
34. «Облако тегов» – визуальное представление списка категорий.
35. «Обработка метаданных публикаций» – набор задач, выполняемых модулями BibReader.
36. «Отфильтрованный корпус документов» – корпус документов, отфильтрованный по каким-либо из полей метаданных.
37. «Публикация» – документ, облечённый в письменную форму носитель информации, удостоверяющий наличие фактов определённого значения.
38. «Поиск документов» – получение списка публикаций, содержащихся в электронной библиотеке, по заданным пользователем критериям.
39. «Поиск релевантных документов» – этап скрининга, отбирающий релевантные публикации.
40. «Поиск уникальных документов» – этап скрининга, удаляющий дубликаты среди публикаций.
41. «Сбор статистики» – модуль, собирающий статистику по корпусу документов по каким-либо из полей метаданных.
42. «Скрининг» – модуль, выполняющий удаление дубликатов среди публикаций и отбирающий релевантные среди оставшихся.
43. «Составление библиографической ссылки» – модуль, составляющий библиографические ссылки для каждой из публикаций в корпусе.
44. «Статистика» – поле метаданных, по которому происходит сбор статистики.
45. «Стиль» – стиль библиографической ссылки.
46. «Тип документа» – поле метаданных публикации, соответствующее типу публикации.
47. «Требования релевантности» – требование того, что публикация должна быть полноценным документом объемом не менее трех страниц.
48. «Фильтр» – поле метаданных, по которому происходит фильтрация.
49. «Фильтрация» – модуль, производящий фильтрацию корпуса документов по каким-либо из полей метаданных.
50. «Файл» – именованная область данных на носителе информации, используемая как базовый объект взаимодействия с данными в операционных системах.
51. «Чтение файла» – модуль, производящий чтение входных файлов.
52. «Электронная библиотека» – упорядоченная коллекция разнородных электронных документов, снабжённых средствами навигации и поиска.
53. «Язык программирования» – формальный язык, предназначенный для записи компьютерных программ.

Определим связи между всеми введёнными понятиями.

# Этап формализации

Связи между всеми введёнными понятиями могут быть структурированы в виде таблицы (см. табл. 3.1). Так как разрабатываемая онтология будет интегрирована с предоставленной онтологией, то отобразим и внешние по отношению к разрабатываемой онтологии понятия (будут обозначены жирным выделением).

Таблица 3.1 — Связи между основными понятиями

|  |  |  |
| --- | --- | --- |
| Понятие 1 | Связь | Понятие 2 |
| Поиск документов | same\_as | **Document Searching** |
| Поиск документов | output | .csv |
| Поиск документов | output | .bib |
| ACM DL | use\_for | Поиск документов |
| ASTS | use\_for | Поиск документов |
| IEEE | use\_for | Поиск документов |
| Scopus | use\_for | Поиск документов |
| Web of Science | use\_for | Поиск документов |
| SpringerLink | use\_for | Поиск документов |
| Inspec | use\_for | Поиск документов |
| ACM DL | is\_a | Электронная библиотека |
| ASTS | is\_a | Электронная библиотека |
| IEEE | is\_a | Электронная библиотека |
| Scopus | is\_a | Электронная библиотека |
| Web of Science | is\_a | Электронная библиотека |
| SpringerLink | is\_a | Электронная библиотека |
| Inspec | is\_a | Электронная библиотека |
| .csv | is\_a | Файл |
| .bib | is\_a | Файл |
| BibReader | is\_a | Настольное приложение |
| C# | a\_part\_of | Настольное приложение |
| C# | is\_a | Язык программирования |
| Чтение файла | is\_a | Модуль обработки метаданных |
| Скрининг | is\_a | Модуль обработки метаданных |
| Составление библиографической ссылки | is\_a | Модуль обработки метаданных |
| Классификация | is\_a | Модуль обработки метаданных |
| Фильтрация | is\_a | Модуль обработки метаданных |
| Сбор статистики | is\_a | Модуль обработки метаданных |
| Чтение файла | a\_part\_of | BibReader |
| Скрининг | a\_part\_of | BibReader |
| Составление библиографической ссылки | a\_part\_of | BibReader |
| Классификация | a\_part\_of | BibReader |
| Фильтрация | a\_part\_of | BibReader |
| Сбор статистики | a\_part\_of | BibReader |
| .csv | input | Чтение файла |
| .bib | input | Чтение файла |
| Чтение файла | output | Корпус первичных документов |
| Корпус первичных документов | is\_a | Корпус документов |
| Поиск уникальных документов | a\_part\_of | Скрининг |
| Поиск уникальных документов | next | Поиск релевантных документов |
| Поиск релевантных документов | a\_part\_of | Скрининг |
| Требования релевантности | use\_for | Поиск релевантных документов |
| Корпус первичных документов | input | Поиск уникальных документов |
| Поиск уникальных документов | output | Корпус уникальных документов |
| Корпус уникальных документов | is\_a | Корпус документов |
| Корпус уникальных документов | input | Поиск релевантных документов |
| Поиск релевантных документов | output | Корпус релевантных документов |
| Корпус релевантных документов | is\_a | Корпус документов |
| Корпус документов | input | Составление библиографической ссылки |
| Корпус документов | input | Классификация |
| Корпус документов | input | Фильтрация |
| Корпус документов | input | Сбор статистики |
| Сбор статистики | output | .xlsx |
| .xlsx | is\_a | Файл |
| Классификация | output | Облако тегов |
| Фильтрация | output | Отфильтрованный корпус документов |
| Составление библиографической ссылки | output | Библиографическая ссылка |
| Библиографическая ссылка | has | ГОСТ |
| Библиографическая ссылка | has | IEEE |
| Библиографическая ссылка | has | APA |
| Библиографическая ссылка | has | Harvard |
| ГОСТ | is\_a | Стиль |
| IEEE | is\_a | Стиль |
| APA | is\_a | Стиль |
| Harvard | is\_a | Стиль |
| Год | is\_a | Метаданные |
| Источник | is\_a | Метаданные |
| Тип документа | is\_a | Метаданные |
| Журнал | is\_a | Метаданные |
| Конференция | is\_a | Метаданные |
| География | is\_a | Метаданные |
| Количество авторов | is\_a | Метаданные |
| Аннотация | is\_a | Метаданные |
| Ключевые слова | is\_a | Метаданные |
| Название | is\_a | Метаданные |
| Год | is\_a | Статистика |
| Источник | is\_a | Статистика |
| Тип документа | is\_a | Статистика |
| Журнал | is\_a | Статистика |
| Конференция | is\_a | Статистика |
| География | is\_a | Статистика |
| Количество авторов | is\_a | Статистика |
| Год | is\_a | Фильтр |
| Источник | is\_a | Фильтр |
| Тип документа | is\_a | Фильтр |
| Журнал | is\_a | Фильтр |
| Конференция | is\_a | Фильтр |
| География | is\_a | Фильтр |
| Количество авторов | is\_a | Фильтр |
| Фильтр | input | Фильтрация |
| Статистика | input | Сбор статистики |
| Аннотация | a\_part\_of | Классификация |
| Ключевые слова | a\_part\_of | Классификация |
| Название | a\_part\_of | Классификация |
| Публикация | has | Метаданные |
| Публикация | a\_part\_of | .csv |
| Публикация | a\_part\_of | .bib |
| Публикация | a\_part\_of | Корпус документов |

Представим выделенные понятия и связи между ними в виде связного графа. Этот граф будет представлять онтологию.

Для построения онтологии используем средство построения онтологических БЗ ОНТОЛИС.

Глава 4. Построение онтологии

На основе всех вышеприведённых этапов разработки онтологии в средстве построения онтологических БЗ ОНТОЛИС была разработана онтология для проблемной области задачи, решаемой в рамках выпускной работы.

Результат приведён на рис. 4.1-4.4.

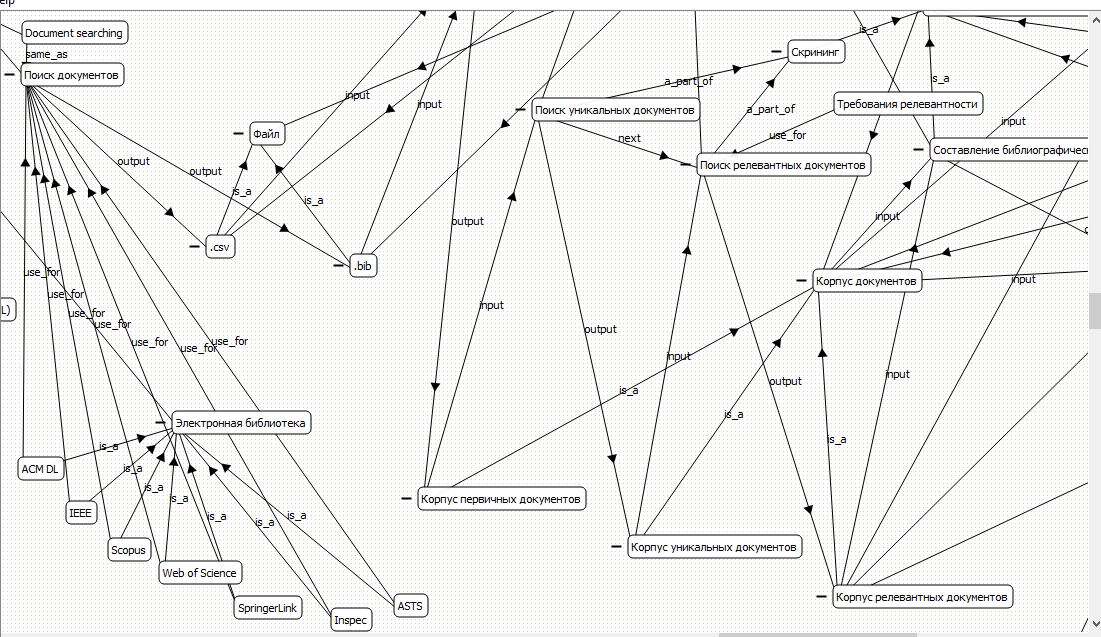


Рис. 4.1 — Часть онтологии №1

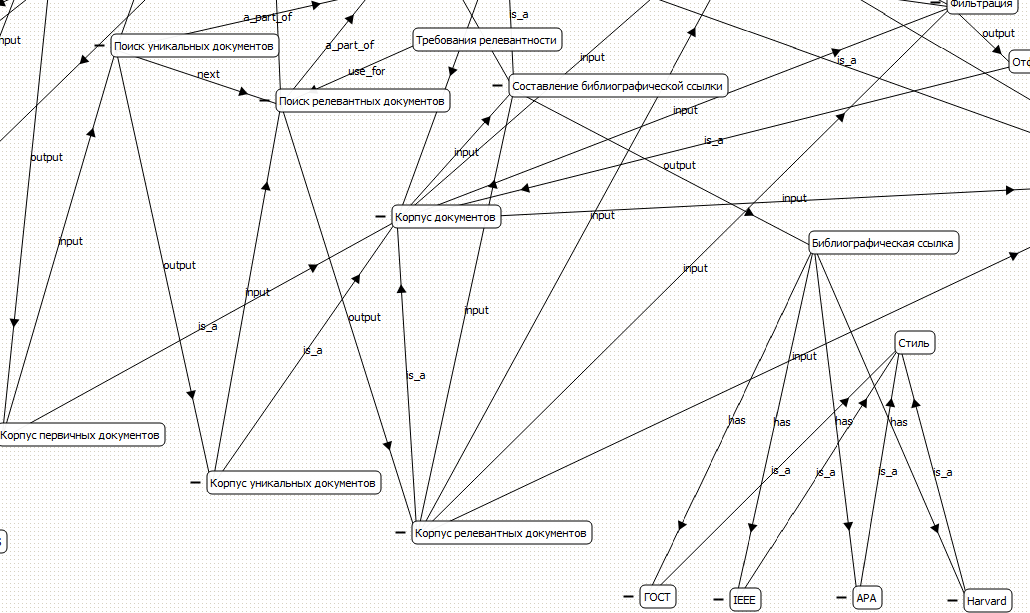


Рис. 4.2 — Часть онтологии №2

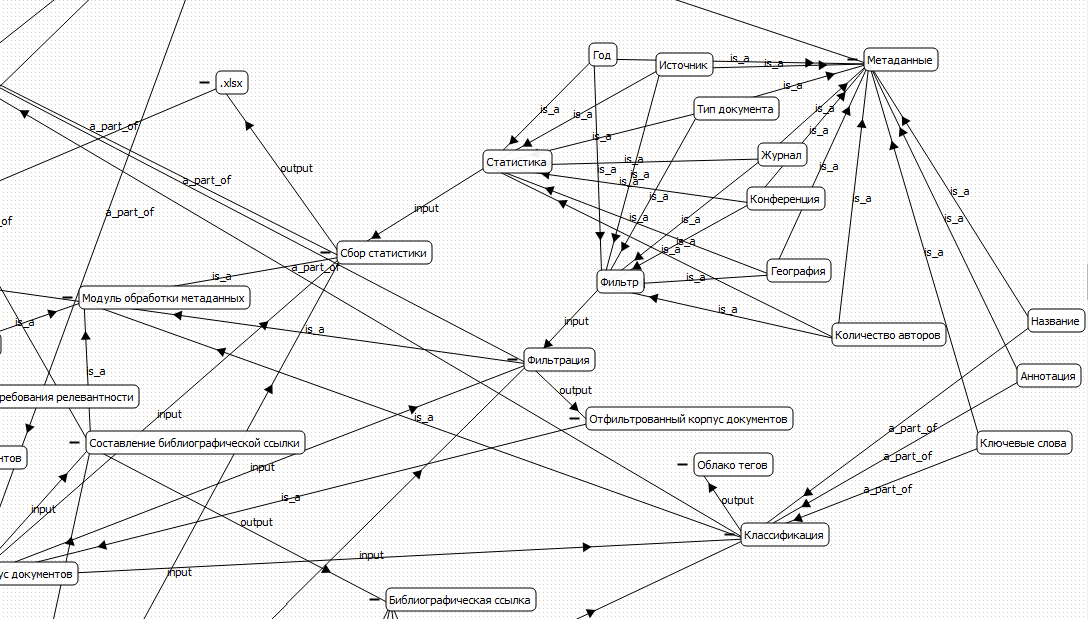


Рис. 4.3 — Часть онтологии №3

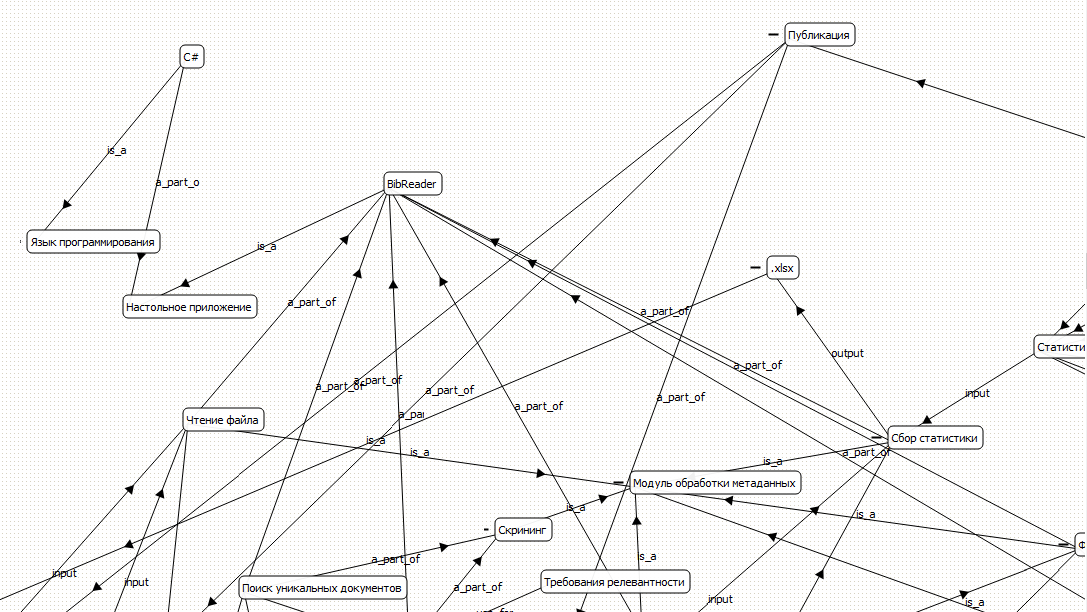


Рис. 4.4 — Часть онтологии №4

Заключение

В ходе данной работы была разработана онтология для проблемной области задачи, решаемой в рамках выпускной работы

Разработанная онтология полностью удовлетворяет предъявленным к ней требованиям.