

УДК 004.82

Е.А. Оробинская¹, Н.В. Шаронова²^{1, 2} НТУ «ХПИ», г. Харьков, Украина

МЕТОД FCA ДЛЯ ПОСТРОЕНИЯ ОНТОЛОГИИ НА ОСНОВЕ ТЕКСТОВОГО КОРПУСА

В статье рассматривается полуавтоматический метод построения онтологии на основе текстового корпуса для специальной предметной области. Предлагаемый метод основан на возможности классификации объектов, исходя из их общих свойств, с применением метода анализа формальных понятий АФП (FCA) для построения решетки понятий. Построенная решетка служит основанием для преобразования найденных терминов в концепты онтологии. Предлагается новая методика, которая позволяет дополнить построение онтологии определением внешних (трансверсальных) отношений между концептами, используя реляционный анализ понятий (RCA). Каждый концепт создаваемой онтологии описывается с помощью дескрипторной логики (ДЛ). В качестве предметной области применения методики была использована область понятий в радиологии.

ОНТОЛОГИЯ, КОНЦЕПТ, РЕШЕТКА ПОНЯТИЙ, АНАЛИЗ ФОРМАЛЬНЫХ ПОНЯТИЙ (FCA), РЕЛЯЦИОННЫЙ АНАЛИЗ КОНЦЕПТОВ (RCA)

Введение

Важным этапом развития любой предметной области является структурированное представление знаний, аккумулированных в данной области, в виде базы знаний (БЗ). Это позволяет специалистам области лучше манипулировать знаниями, обнаруживать новые закономерности, обмениваться опытом. Например, в области радиационной защиты, при медицинском облучении важным является понимание свойств радиоактивных изотопов и специфики их использования для диагностики и лечения различных заболеваний, от чего зависит как терапевтический эффект для пациентов, так и безопасность персонала.

Одним из эффективных способов представления знаний являются онтологии, к разработке которых в последнее время проявляется большой интерес [1]. Понятие онтологии в интерпретации для задач обработки информации наиболее кратко и выразительно определил Т. Грубер [2]. Он определил онтологию как «...явную спецификацию некоторой концептуализации». Под концептуализацией у Грубера понимается совокупность общих понятий, объектов и связей, характерных для некоторой предметной области (ПО). Построение онтологии предполагает определение классов объектов и описание их отношений с помощью одного из формальных языков, например, дескриптивной логики, что позволяет отвечать на запросы, или так называемые компетентные вопросы (*competency questions*), которые изначально составляются на естественном языке и затем «переводятся» на используемый формальный язык.

Исходя из этого, процесс построения онтологии («ручной», полуавтоматический или автоматический) предполагает исходное создание базовой онтологической структуры, представляющей основные понятия ПО и связи между ними. Традиционно для этих целей привлекаются эксперты

соответствующих предметных областей. Однако в последние годы специалисты в области искусственного интеллекта (ИИ) ставят перед собой задачу найти эффективные методы извлечения знаний из специализированных текстовых корпусов без дорогостоящего привлечения специалистов. Основанием для этого может служить то, что именно в текстах хранится значительная часть всех знаний в стабильной, относительно упорядоченной форме. При этом доля электронных текстов, доступных для непосредственной обработки, все время возрастает.

1. Постановка задачи

Целью данной работы является обоснование выбора языка описания концептов и разработка метода построения онтологии для некоторой предметной области, представленной текстовым корпусом. В данном случае был использован текстовый корпус, состоящий из статей, стандартов и других источников из области радиологии и радиационной защиты, доступных в электронном виде, допускающих использование инструментов терминологического анализа для автоматического обнаружения и определения свойств терминов [3]. Конкретные задачи, стоящие перед авторами, заключаются в следующем. Во-первых, было интересно предложить описание концептов с помощью дескриптивной логики (ДЛ) [4] с последующим сравнительным анализом описания. Во-вторых, предложить метод построения правил вывода для концептов базовой онтологии, позволяющих получить ответы на компетентные вопросы следующего типа:

- *Какие объекты относятся к тому же классу, что и «опухоль печени»?*
- *Принадлежат ли объекты «бронхогенная карцинома» и «опухоль желудка» одной группе?*
- *Является ли объект «легкие» органом?*

И, наконец, предложить новый подход — анализ реляционных понятий, являющийся расширением анализа формальных понятий (FCA), который позволяет описать так называемые внешние, или трансверсальные, отношения между концептами. Предложенный подход позволяет классифицировать объекты на основании отношений, которые они разделяют с другими объектами, чтобы получить ответы на вопросы типа:

- *Что разрушает радиация?*
- *Что диагностируется при помощи радиоактивных изотопов?*

Данная статья организована следующим образом. В разделе 2 описывается общая методика построения онтологии. Далее, в разделе 3 рассмотрены некоторые процедуры обработки текста, позволяющие перейти от текстового корпуса к входным данным, пригодным для построения онтологии. В разделе 4 описывается метод FCA для построения иерархии концептов. В разделе 5 представлен метод формирования внешних отношений между концептами онтологии. И, наконец, раздел 6 завершается выводами о возможных перспективах использования предлагаемого метода.

2. Методика построения онтологии

В работе предлагается использовать технологию «Methontology» [5, 6], разработанную в лаборатории искусственного интеллекта Политехнического университета Мадрида для построения прикладных онтологий.

Процесс построения онтологии на основе этой технологии, проиллюстрированный на рис.1, состоит из следующих этапов:

- извлечение терминов: этот этап состоит в том, чтобы обнаружить и извлечь из входного корпуса термины и их свойства. Для этого используются специализированные информационные ресурсы — глоссарий медицинских терминов и терминов по радиационной безопасности, синтаксические шаблоны. С помощью синтаксического анализатора извлекаются пары (*объект, свойство*) и триплеты (*объект, свойство, объект*), относящиеся к общим смысловым блокам (см. раздел 3);

- построение основы онтологии: на этом этапе найденные пары (*объект, свойство*) используются для построения иерархии концептов на основе метода FCA (см. раздел 4);

- извлечение внешних отношений: на этом этапе применяется реляционный анализ понятий для извлечения внешних отношений;

- в завершение, результаты выполнения двух предыдущих этапов объединяются для получения более полной онтологии.

3. Методы формирования концептов онтологии

Несмотря на разнообразие существующих подходов, предлагающих автоматическую группировку понятий, эта задача продолжает оставаться в значительной степени открытой. Одним из рациональных подходов, на взгляд авторов, является подход, основанный на гипотезе Харриса (Harris) [6]. Согласно ему, изучение синтаксических закономерностей в научном текстовом корпусе, т.е. состоящем из текстов, написанных на «подъязыке», характерном для данной ПО, позволяет обнаружить специфичные синтаксические структуры, формируемые терминами, которые отражают знания исследуемой ПО. Несколько похожих методов [7, 8], основанных на этой гипотезе, предлагают группировать термины в классы на основе их совместного появления в синтаксических структурах с одинаковыми группами глаголов. Использование одного из этих методов позволяет объединять названия объектов в классы в соответствии с группами глаголов, при которых они выступают в качестве подлежащего или дополнения. Например, объекты {*щитовидная железа, желудок*} объединены в один класс, так как они появляются в качестве подлежащего (или субъекта) при глаголе {*поглощать*} и в качестве дополнения (или объекта) при глаголе {*повреждать*}. С другой стороны, данный метод позволяет также извлекать отношения между разными объектами, появляющимися в качестве подлежащего или дополнения одного и того же глагола. Например, объект {*радиоактивный йод*} является дополнением, связанным с объектом {*щитовидная железа*}, который, в свою очередь, является подлежащим для глагола {*поглощать*}.

В результате обработки текстового корпуса требуется получить пары (*субъект, предикат*), (*объект, предикат*) и триплеты (*субъект, предикат, объект*), содержащие ключевые слова. Отбор терминов был выполнен с помощью программы Monosocp. Синтаксический анализ выполнен с помощью анализатора АОТ. Извлеченные пары и триплеты были

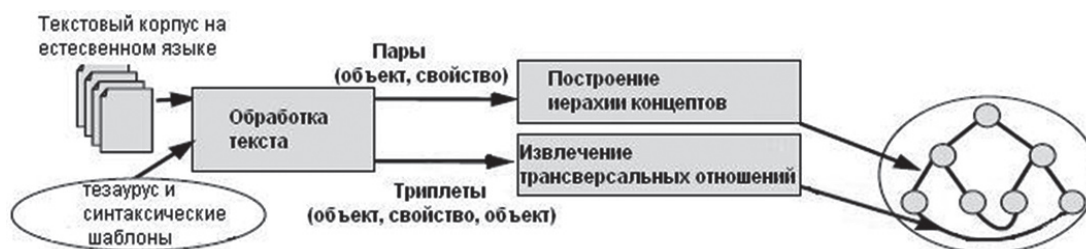


Рис. 1. Методика построения онтологии

предложены эксперту для отбора среди них наиболее релевантных. Слова в парах и триплетях приведены к нормальной форме.

Представим несколько примеров фрагментов текста из области медицинской радиологии:

1. «Радионуклиды разрушают клетки, повреждают органы и ткани и являются причиной скорой гибели организма, однако они же разрушают и злокачественные опухоли».

Из текста были извлечены пары (*ткани, повреждать*), (*органы, повреждать*), (*опухоли, разрушать*), (*радионуклиды, разрушать*).

2. «Щитовидная железа детей в три раза активнее поглощает попавший в организм радиоактивный йод».

Извлечены пары (*щитовидная железа, поглощать*) и (*радиоактивный йод, поглощать*).

3. «Радиоактивный йод использовался, чтобы диагностировать рак щитовидной железы и другие, связанные с ней заболевания».

Извлечены пары (*радиоактивный йод, диагностировать*), (*рак щитовидной железы, диагностировать*).

4. Построение основы онтологии с помощью метода FCA

В работе была использована идея Ф. Симиано [1], предлагающего строить онтологию на основе метода FCA [3]. Далее сформулирован переход от решетки свойств к онтологии и дано определение каждого концепта с помощью дескриптивной логики, что позволяет получить ответы на вопросы, сформулированные в разделе 1.

По собственному замечанию Ф. Симиано [7], наиболее узким местом (*knowledge acquisition bottleneck*) при построении онтологии является именно моделирование предметной области, т.е. определение отношений между ее концептами.

Основой онтологии является множество основных понятий или концептов предметной области, связанных между собой бинарными отношениями. Метод FCA позволяет визуализировать зависимости между объектами с помощью решетки формальных понятий [8]. Его можно использовать при анализе данных для обнаружения отношений между элементами (в данном случае концептами) некоторой системы (в данном случае текстового корпуса). Отношения обнаруживаются через атрибуты, описывающие свойства. В свою очередь, атрибуты должны быть близки к обычным категориям человеческого мышления и должны допускать наглядную и понятную интерпретацию. Таким образом, FCA можно рассматривать как технологию кластеризации концептов, которая позволяет определить интенционалы для отдельных блоков данных.

Основными в FCA являются понятия *формального контекста* и *формального концепта*.

Определение 1 (формальный контекст)

Множество (G, M, I) называется формальным контекстом, если G и M являются множествами, элементы которых связаны бинарным отношением $I: I \subseteq G \times M$. Элементы множества G называются объектами, элементы множества M называются атрибутами, а элементы множества I определяют инцидентность объектов и атрибутов или, другими словами, принадлежность атрибута объекту, определяя, таким образом, формальный контекст.

Определение 2

Пусть имеется формальный контекст (G, M, I) . Определим A' для $A \subseteq G$ следующим образом $A' := \{m \in M \mid \forall g \in A: (g, m) \in I\}$; аналогично определим B' для $B \subseteq M$ как $B' := \{g \in G \mid \forall m \in B: (g, m) \in I\}$.

Проще говоря, A' — это множество атрибутов, общих для всех объектов множества A , и B' — это множество объектов, обладающих всеми атрибутами из B .

Оператор «'» называется оператором дерирации и применяется для обозначения подмножеств множеств G и M .

На основании вышесказанного можно определить понятие *формального концепта* следующим образом:

Определение 3 (формальный концепт)

Пара (A, B) является формальным концептом (G, M, I) тогда и только тогда, когда $A \subseteq G$, $B \subseteq M$, $A' = B$ и $A = B'$.

Другими словами, (A, B) является формальным концептом, если множество всех атрибутов, описывающих объекты A , совпадает с B и, с другой стороны, все объекты A описываются всеми атрибутами B . В этом случае A называют *экстендом*, а B называют *интендом* формального концепта (A, B) .

Множество формальных концептов в данном контексте реализует отношение подчиненности, определенное следующим образом:

$$(A_1, B_1) \leq (A_2, B_2) \Leftrightarrow A_1 \subseteq A_2 (\Leftrightarrow B_2 \subseteq B_1).$$

Отношение подчиненности позволяет организовать формальные концепты в полную матрицу, называемую решеткой концептов, которая обозначается следующим образом: $B(G; M; I)$.

В качестве примера в табл. 1 приведен фрагмент формального контекста (G, M, I) для области медицинской радиологии, где G — это множество органов человека и их онкологических патологий, M — множество их свойств. I — это множество бинарных отношений между M и G , такое что $I(g, m)$ означает, что g является субъектом или объектом в текстовом корпусе.

Результующая решетка представлена на рис. 2. Такое размещение узлов решетки отражает наследование объектами свойств. Свойство обозначено над узлом решетки, а элемент, обладающий данным свойством — под узлом.

Все элементы, размещенные под некоторым элементом g , помеченным свойством m , наследуют это свойство. Аналогичным образом, все последователи, размещенные под узлом решетки, помеченным объектом g , являются более специфичными понятиями по отношению к g .

Таблица 1

Фрагмент формального контекста

	Разрушаемый	Диагностируемый	Повреждаемый	(Хорошо) поддающийся лечению
Опухоль желудка	×	×		
Бронхогенная карцинома	×	×		×
Легкие	×		×	
Опухоль щитовидной железы	×	×		×
Желудок	×		×	
Печень	×		×	
Щитовидная железа	×		×	
Опухоль печени	×	×		

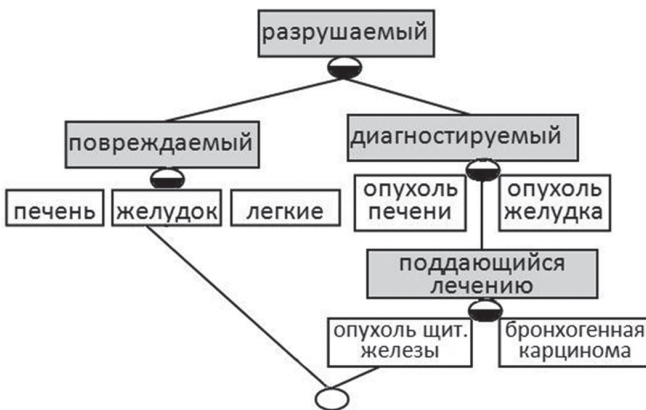


Рис. 2. Решетка, соответствующая заданному формальному контексту

Таким образом достигается расширение экстенции A концепта (A, B) по отношению ко всем объектам, появляющимся ниже узла n в решетке, и интенции B по отношению ко всем атрибутам наследников узла n . Такое представление позволяет идентифицировать свойства и экземпляры каждого концепта.

4.1. Переход от решетки к онтологии, разметка с помощью экспертов

Определим функцию преобразования: $\alpha: B(G, M, I) \rightarrow TBox \cup ABox$, где $B(G, M, I)$ – это решетка концептов, полученная на основе FCA. $TBox$ представляет терминологический словарь ПО и $ABox$ – это множество утверждений об экземпля-

рах, т.е. описание их свойств. Совокупность $TBox$ и $ABox$, терминов и утверждений составляет базу знаний. [4]. Формально $TBox$ и $ABox$ определены следующим образом определения 4,5).

Определение 4 (основа онтологии)

Основа онтологии представляет собой триплет $O := (C, \subseteq c, A)$ где C – это ансамбль концептов онтологии, $\subseteq c$ – отношение подчиненности (*is-a*) между концептами, являющееся транзитивным и несимметричным. A – это множество свойств или атрибутов концептов.

Определение 5 (база знаний)

База знаний для онтологии $O := (C, \subseteq c, A)$ представляет собой структуру:

$KB := (I, i_C, i_A)$, где I – это множество экземпляров, $i_C: C \rightarrow 2^I$ – это функция создания экземпляров, i_A – это функция создания атрибутов.

Функция преобразования α , формализующая этот переход, представлена с помощью табл. 2.

Таблица 2

Формализация перехода Решетка – Онтология

Решетка концептов	Онтология (описание в ДЛ)
Контекст K	Атомарный концепт $c \equiv \alpha(K)$
Свойство $m \in M$	Атомарная роль $\alpha(m) \equiv \exists m.T$ в $TBox$
Объект $g \in G$	Экземпляр $\alpha(g)$ в $ABox \perp (g)$
Элемент $(g; m) \in I$	Утверждение $\alpha(m)(\alpha(g))$ в $ABox$
Концепт $c = (X; Y) \in C$	Концепт определенный в $TBox$, т.е. $\alpha(c) \equiv \bigcap_{m \in Y} \alpha(m)$
$\forall (c, \bar{c}) \in C \times C$, так что, $c \prec \bar{c}$	Включение аксиом $\alpha(m) \subseteq \alpha(\bar{m})$ в $TBox$
Множество атрибутов концептов $(\bigwedge_{i=1}^n c_i)$	Конъюнкция концептов $\alpha(c_1 \prec \dots \prec \alpha(c_n))$

Полученная онтология предлагается эксперту, который должен определить соответствие общего концепта каждому множеству элементов, исходя из свойств, которыми обладают все элементы множества. Например, группа объектов, обладающих свойствами $\{разрушаемый, повреждаемый\}$, может быть промаркирована понятием *орган*; а группа, обладающая свойствами $\{разрушаемый, диагностируемый\}$ – понятием *опухоль*. Такая маркировка осуществляется для того, чтобы облегчить восприятие и чтение онтологии другими специалистами.

4.2. Представление концептов с помощью дескриптивной логики

Определение концепта в ДЛ получается путем конъюнкции его атрибутов квантором существования \exists . Результат преобразования решетки понятий в онтологию представлен на рис. 3. В табл. 3 даны определения каждого концепта для полученного фрагмента онтологии.

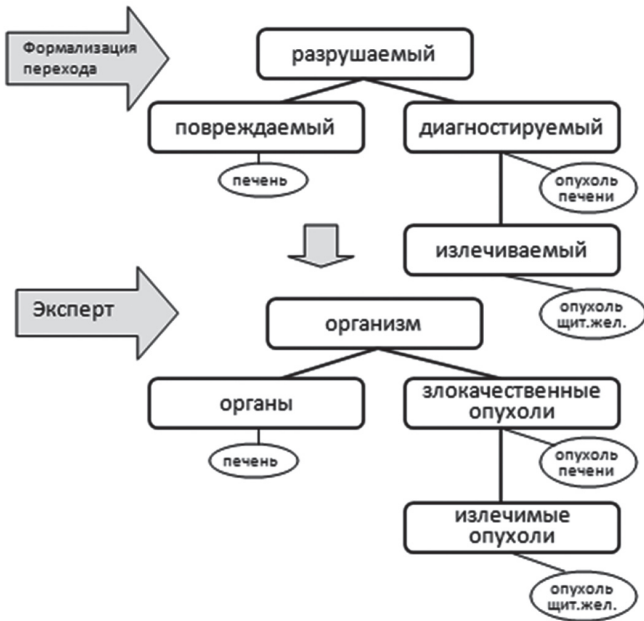


Рис. 3. Преобразование решетки в онтологию и определение концептов

Таблица 3

Определение концептов онтологии в ДЛ

Определение концепта	Определение концепта
Объект: $= \exists$ исследуемый	Злокачественная опухоль: $= \exists$ исследуемый $\cap \exists$ диагностируемый
Орган: $= \exists$ исследуемый $\cap \exists$ повреждаемый	Излечиваемая злокачественная опухоль: $= \exists$ исследуемый $\cap \exists$ диагностируемый $\cap \exists$ излечиваемый

Такое представление позволяет использовать возможности ДЛ для получения ответов на запросы нижеследующих типов.

Население онтологии (ontology population). Пусть o_1 – это объект, обладающий свойствами $\{a, b\}$. Экземпляром какого класса (каких классов) является объект o_1 ? Ответ может быть следующим: это самый верхний класс, который обладает свойствами $\{a, b\}$, т. е. класс $C_1 \subseteq \exists a. T \cap \exists b. T$ [4]. Например, на вопрос «Является ли опухоль щитовидной железы злокачественной опухолью, если этот объект обладает следующими свойствами {исследуемый, диагностируемый, излечиваемый}?» Ответ: соответствующий класс C_1 , обладающий теми же свойствами – это класс «излечиваемая злокачественная опухоль», который в свою очередь является подклассом класса «злокачественная опухоль». Поэтому опухоль щитовидной железы является злокачественной опухолью.

Сравнение экземпляров онтологии: Пусть имеются два объекта o_1 и o_2 . Принадлежит ли o_1 тому же классу, что и o_2 ? Ответом на этот вопрос будет определение того, каков класс C_1 , которому принадлежит o_1 , каков класс C_2 , которому принадлежит o_2 и последующая проверка: справедливо ли, что $C_1 \equiv C_2$. Например, на вопрос, какие объекты

принадлежат тому же классу, что и «печень»? Ответ: это множество объектов того же класса, что и объект «печень». Объект «печень» принадлежит классу «орган». Ко множеству других объектов этого же класса относятся «желудок», «щитовидная железа», «легкие».

Вопрос «Принадлежат ли объекты «печень» и «бронхогенная карцинома» одному и тому же классу?». Ответ: объект «печень» принадлежит классу $орган := \exists$ исследуемый $\cap \exists$ повреждаемый; объект «бронхогенная карцинома» принадлежит классу $излечиваемая злокачественная опухоль := \exists$ исследуемый $\cap \exists$ диагностируемый $\cap \exists$ излечиваемый.

В свою очередь, $орган \cap$ $излечиваемая злокачественная опухоль = \perp$. Следовательно, объекты «печень» и «бронхогенная карцинома» не принадлежат одному и тому же классу.

Однако представление объектов не ограничивается одним лишь описанием их свойств. Они могут быть определены через отношения, которые их связывают с другими объектами. Поэтому возникает необходимость расширить основу онтологии внешними отношениями между концептами (классами онтологии). В этом случае предлагается использовать дополнительный метод: реляционный анализ концептов (RCA) с тем, чтобы учесть внешние (трансверсальные) отношения в онтологии.

5. Извлечение внешних отношений

Извлечение внешних (трансверсальных) отношений позволяет давать концептам более точные определения. Благодаря этому концепт определяется не только по общим свойствам отдельных экземпляров, но и по отношениям, которые его связывают с другими концептами. Осенак-Жилье (Aussenac-Gilles) предлагает использовать метод обучения онтологии для синтаксических образцов [3]. На основе триплетов ($термин_1$, $отношение_1$, $термин_2$), извлеченных вручную из текстового корпуса, эксперты пытаются прежде всего обобщить $отношение_1$, извлекая триплеты типа ($термин_1$, $отношение_k$, $термин_2$). Определяется отношение R , наиболее общее для терминов ($термин_1$, $термин_2$). Затем извлекают все термины, связанные отношением R в формате ($термин_i$, R , $термин_j$). Таким образом, можно заменить одно частное отношение, связывающее экземпляры, другим, более общим. Однако в [9] не предлагается дополнительно использовать некоторую базовую онтологическую структуру, чтобы обобщить термины. Другой подход, предложенный в [10], основан на извлечении ассоциативных правил. Из текстового корпуса извлекаются пары ($термин_1$, $термин_2$), для них определяется ассоциативное правило ($термин_1 \Rightarrow термин_2$) с тем, чтобы сохранить лишь наиболее часто встречающиеся и наиболее достоверные пары. Метод, предложенный в [10], определяет наличие

общих отношений между концептами онтологии, однако сами отношения никак не определены. Известно только, что концепт C_1 связан с концептом C_2 , но не известно, каким отношением.

5.1. Анализ формальных концептов

В работе предлагается формальный метод извлечения внешних отношений, т.е. отношений между концептами. Этот метод позволяет приписать некоторый маркер каждому извлеченному отношению. Он требует также наличия базовой онтологии для обобщения терминов. Данный метод является расширением метода FCA, который, как было отмечено выше, позволяет группировать объекты не только по их общим свойствам, но и по отношениям, которые их связывают. Идея метода, описанного в [12], заключается в том, чтобы сконструировать формальный контекст для каждого внешнего (трансверсального) отношения, извлеченного из текстового корпуса. Элементами каждого контекста являются (см. определение 1) множество экземпляров, множество атрибутов, и связывающее их бинарное отношение. Такое расширение называется реляционным анализом концептов (ARC). Центральным понятием ARC является реляционное множество контекстов.

Определение 6 (ARC)

Реляционное множество контекстов представляет собой пару (K, R) , где

- K – это множество контекстов $K_i = (G_i, M_i, I_i)$, причем каждое множество экземпляров имеет один единственный контекст;
- R – это множество отношений $r_k \subseteq G_i \times G_j$, где G_i и G_j – это два множества экземпляров K .

Приведем пример для рассматриваемой предметной области – медицинской радиологии. Пусть имеются два контекста, сконструированных на основе метода FCA: $K_1 = (G_1, M_1, I_1)$, представленный в табл. 1 – множество органов человека и их опухолевых патологий, и $K_2 = (G_2, M_2, I_2)$, представленный в табл. 4 – множество радионуклидов, применяемых в медицине. Имеется также два отношения r_1 и r_2 , которые описаны в табл. 5.

Интеграция отношений r_1 и r_2 и решетки свойств для органов выполнена на основе процесса взвешивания,

подробно описанного в [12]. Результирующая решетка демонстрирует такие внешние отношения: *диагностируемый_с_помощью* между экземплярами «опухоль щитовидной железы» и «йод 131» и *излечиваемый_с_помощью* между экземплярами «бронхиальная карцинома» и «кобальт 60».

Таблица 4

Фрагмент формального контекста для радионуклидов

	Период полураспада < 2 часов	Период полураспада > 24 часов и < 10 суток	Период полураспада от 100 суток до 10 лет
Йод 131		×	
Кобальт 60			×
Углерод 11	×		

Таблица 5

Решетка для внешних отношений между концептами

Диагностируемый_с_помощью			
	Йод 131	Кобальт 60	Углерод 11
Опухоль щитовидной железы	×		
Опухоль печени			×
Излечиваемый_с_помощью			
	Йод 131	Кобальт 60	Углерод 11
Бронхиальная карцинома		×	
Опухоль щитовидной железы		×	

На основе внешних отношений структура онтологии может быть расширена и преобразована в более сложную онтологию (рис. 4), описанную с помощью ДЛ.

Определение 7 (полная онтология)

Полная онтология представлена следующим набором из пяти элементов $O := (C, \subseteq_C, A, R, \sigma)$, где (C, \subseteq_C, A) представляет основу онтологии, R – это множество отношений и σ – это сигнатуры отношений.

Определение 8 (база знаний)

База знаний, соответствующая полной онтологии типа $O := (C, \subseteq_C, A, R, \sigma)$ – это структура $CB := (I; i_C; i_A; i_R)$, где I – множество экземпляров, $i_C : C \rightarrow 2^I$ – так называемая функция установки концептов, i_A – функция установки свойств и $i_R : C \rightarrow 2^{I+}$ – это функция определения отношений.

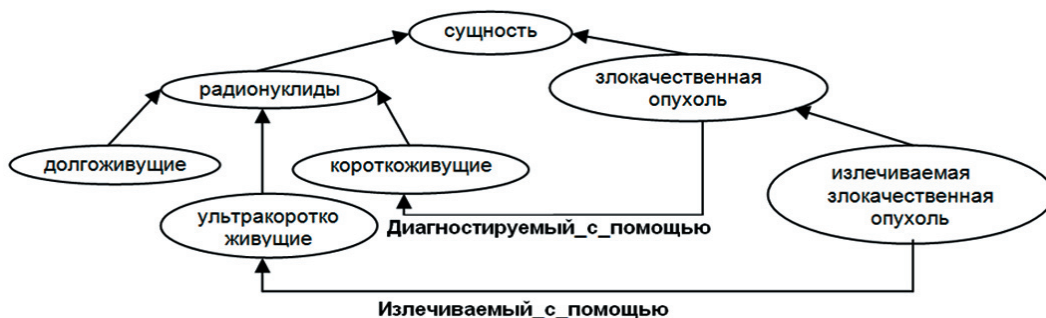


Рис. 4. Онтология, дополненная внешними отношениями

5.2. Представление концептов в дескриптивной логике

Извлечение внешних (трансверсальных) отношений позволяет улучшить и обогатить определение концептов на основе ДЛ. Два концепта «злокачественная опухоль» и «излечиваемая злокачественная опухоль» были переопределены с помощью двух отношений *диагностируемый_с_помощью* и *излечиваемый_с_помощью*:

– *злокачественная опухоль*: = \exists *исследуемый* \cap \exists *диагностируемый* \cap \exists *диагностируемый_с_помощью*. Радионуклиды короткоживущие;

– *излечиваемая злокачественная опухоль*: = \exists *исследуемый* \cap \exists *диагностируемый* \cap \exists *излечиваемый* \cap \exists *излечиваемый_с_помощью*. Радионуклиды долгоживущие.

Приведем еще один пример запроса: «Что диагностируется с помощью короткоживущих радионуклидов?» Ответ может быть получен из онтологии, обогащенной внешними отношениями между концептами, фрагмент которой представлен на рис. 4. Экземпляры концепта «злокачественная опухоль» являются подмножеством отношения *диагностируемый_с_помощью* и служат ответом на данный запрос.

Заключение

В работе описаны методы, позволяющие построить онтологию на основе текстового корпуса, представляющего некоторую предметную область. Показано, что на основе метода FCA можно корректно группировать объекты, рассматривая их общие свойства, и строить таксономию концептов, связанных транзитивным отношением подчиненности (*is-a*). Второй предлагаемый метод позволяет извлекать внешние отношения между концептами с помощью реляционного анализа концептов и, таким образом, расширить возможности использования онтологии, например получать ответы на более сложные запросы. Дескрипторная логика была выбрана в качестве языка описания онтологии благодаря относительной простоте построения правил вывода. Достоинством метода является его универсальность и независимость от области применения.

Дальнейшими задачами авторов являются следующие: исследовать возможности использования в данном методе других языков описания онтологий, в т.ч. алгебрологические средства; уточнить требования к текстовому корпусу, являющемуся информационным ресурсом для построения онтологии; усовершенствовать данный метод за счет добавления синтаксических шаблонов для терминов предметной области.

Список литературы: 1. Buitelaar P. Ontology Learning from Text: An Overview [Текст]: / P. Buitelaar, P. Cimiano, B. Magnini reviewed by C. Brewster; US, IOS Press, 2005. – 180 p. 2. Gruber T. Toward Principles for the Design of Ontologies Used for Knowledge Sharing [Текст] / T. Gruber // In International Journal Human-Computer Studies. – 1995. – Vol. 43. – P. 907-928. 3. Aussenac-Gilles N. Methodes ascendantes

pour l'ingenierie des connaissances. [Текст] : Habilitation a diriger des recherches / N. Aussenac-Gilles. – France: Universitit Toulouse Paul Sabatier. – 2005. – 226 p. 4. Description Logic Handbook: Theory, Implementation and Applications [Текст] / F. Baader, D. Calvanese, D. McGuinness and all. – UK.: Cambridge University Press, 2003. – 572 p. 5. Gomez-Perez A. Ontological Engineering [Текст] / A. Gomez-Perez, M. Fernandez-Lopez, O. Corcho. – London Ltd.: Springer, 2004. – 432 p. 6. Fernandez-Lopez M. Bilding a Chemical Ontology Using Methontology and the Ontology Gesign Environment [Текст] / M. Fernandez-Lopez, A. Gomez-Perez, J. P. Sierra // In IEEE intelligent systems & their applications. – 1999. – Vol. 1. – P. 37-45. 7. Cimiano Ph. Learning concept hierarchies from text corpora using formal concept analysis [Текст] / Ph. Cimiano, A. Hotho, S. Staab // In Journal of Artificial Intelligence Research. – 2005. – V. 24. – P. 305–339. 8. Ganter B. Formal Concept Analysis: Mathematical Foundations [Текст] / B. Ganter, R. Wille. – London Ltd.: Springer-Verlag, 1999. – 346 p. 9. Noun classification from predicate-argument structures [Текст]: proceedings of the 28th Annual Meeting of the ACL. – 1999. – p.268-275. 10. Discovering conceptual relation from text [Текст]: Proceeding of the 14th European Conference on artificial intelligence, 20-25 August 2000. Germany, Berlin. – p.321–325. 11. Extracting formal concepts out of relational data [Текст]: proceedings of the 4th Int. Conf. Knowledge Discovery and Discrete Mathematics, 3-6 September 2003. France, Metz. – p.37-49. 12. Proposal for combining formal concept analysis and description logics for mining relational data [Текст]: proceedings of the 5th Int. Conf., 12-16 February 2007. France, Clermont-Ferrand. – 327 p.

Поступила в редколлегию 15.06.2011.

УДК 004.82

Метод FCA для побудови онтології на основі текстового корпусу / О.О. Оробінська, Н.В. Шаронова // Біоніка інтелекту: наук.-техн. журнал. – 2011. – № 2 (76). – С. 129-135.

Розглянуто напівавтоматичний метод побудови онтології на основі текстового корпусу для спеціальної предметної області. Запропонований метод оснований на можливості класифікації об'єктів на базі їхніх загальних властивостей із застосуванням методу аналізу формальних понять – АФП (FCA) для побудови решітки понять. Побудована решітка використовується для перетворення знайдених термінів у концепти онтології. Розроблено методику, що дозволяє доповнити онтологію шляхом визначення зовнішніх (трансверсальних) відношень між концептами з використанням реляційного аналізу понять (RCA). Кожний концепт онтології описується за допомогою дескрипторної логіки (ДЛ). Метод був застосований на текстах з радіології.

Табл. 5. Іл. 4. Бібліогр.: 12 найм.

UDC 004.82

Ontology construction from text's corpus with FCA / O.O. Orobinska, N.V. Sharonova // Bionics of Intelligense: Sci. Mag. – 2011. – № 2 (76). – P. 129-135.

In this paper the semi-automatic method of ontology creation from the text corpora for special data domain is considered. With proposed method the objects classification is performed from their commons properties. The method of the formal concepts analysis (FCA) is on the base. The constructed grid is transformed in ontology concepts. New technique allows to add the exteriors (transversals) relations between the concepts is offered. This, technique is based on relational concepts analysis (RCA). The ontology concepts are described with a logic descriptive. The applied domain is the radiology.

Tab. 5. Fig. 4. Ref. 12 items.