

ПРИМЕНЕНИЕ МЕТОДА АНАЛИЗА ФОРМАЛЬНЫХ КОНЦЕПТОВ В ОНТОЛОГИЧЕСКОМ ИНЖИНИРИНГЕ

Рябова Н.В.

Харьковский национальный университет радиоэлектроники, 61166 Харьков, пр.-т Науки, 14
nataliya.ryabova@nure.ua

Представление и работа со знаниями в веб-пространстве являются одним из ключевых направлений в разработке современных интеллектуальных систем поддержки принятия решений. Наиболее эффективным подходом для представления знаний в такого рода веб-системах стал онтологический подход, развиваемый в рамках онтологического инжиниринга (ontological engineering, OE). Особое значение приобретают методы, позволяющие осуществлять построение онтологий «с нуля» (from the scratch), а также методы для осуществления операций над онтологиями. Общепринятым определением онтология стало «эксплицитная спецификация концептуализации». Обычно это означает, что онтология описывает концепты (понятия) и отношения на выявленном множестве концептов, релевантных предметной области (ПрО). Однако, для обеспечения возможности обмена знаниями необходимо описать эти концепты и отношения более тщательным образом, чем простое упорядочивание их в таксономию. Тем не менее, в большинстве случаев разработка онтологий начинается и заканчивается построением таксономий. В связи с этим представляется целесообразным применение для построения, уточнения, пополнения и «заселения» онтологий метода анализа формальных концептов (Formal Concept Analysis - FCA), изначально предназначенного для анализа данных и позволяющего идентифицировать концептуальные структуры на множествах данных [1]. FCA также позволяет обнаруживать необходимость в новых онтологических концептах и отношениях, что, в свою очередь, приводит к более полной онтологии, содержащей описания этих сущностей таким образом, чтобы сделать возможным обмен знаниями и совместное использование (sharing) разработанной онтологии заинтересованными группами пользователей, работающими в одной или схожих ПрО. Например, концепт может быть описан не только его позицией в таксономической (is-a) иерархии, но также с помощью отношений, которые могут быть применены к этому концепту. Аналогичным образом, отношение может быть описано концептами, связанными этим отношением.

Целью данной работы является развитие подхода к построению онтологий, их обучению (ontology learning) и «заселению» (population), т.е. наполнению конкретными экземплярами концептов, на основе метода FCA. Такой подход позволит выявлять новые концепты и отношения, не заданные явным образом, и встраивать их в онтологическую структуру, представляя более полное и адекватное описание концептуальной модели ПрО для решения комплекса поставленных задач. Для достижения поставленной цели в данной работе решаются следующие задачи:

- анализ описаний онтологий в общем виде;
- описание метода FCA и его возможностей для онтологического обучения;
- исследование возможностей и использование метода построения онтологий, их обучения и «заселения» на основе FCA;
- практическая апробация FCA-метода онтологического обучения;
- очертить перспективы дальнейшего использования предлагаемого подхода в онтологическом инжиниринге при решении прикладных задач.

Построение таксономии, т.е. иерархии концептов, является одной из ключевых задач любой системы, основанной на знаниях, которая оперирует декларативными знаниями о предметной области (ПрО) и способна к логическому выводу на основе этих знаний. Иерархии концептов позволяют структурировать информацию в категории и тем самым облегчают ее поиск и повторное использование. Одним из возможных путей решения данной задачи является автоматическое обучение иерархии концептов с использованием в качестве источника знаний специально подобранных (релевантных ПрО) текстовых коллекций. В данной работе рассматривается проблема автоматизации получения таксономий из текстовых коллекций на основе метода анализа формальных концептов (Formal Concept Analysis - FCA). Модель контекста формируется в результате применения парсера (лингвистического анализатора) и построения векторного представления синтаксических зависимостей определенных терминов. Затем с помощью FCA строится решетка, которая конвертируется в иерархию концептов. Метод FCA, базирующийся на теории порядка, позволяет выявлять отношения наследования между объектами, описание которых задано в виде множества атрибутов (свойств или признаков объектов) с одной стороны, и самими атрибутами с другой стороны [1]. FCA является хорошей теоретической базой для автоматического порождения онтологий из коллекций объектов и их свойств. Далее рассмотрим процесс автоматического построения иерархии концептов с использованием текстовой коллекции в качестве источника информации [2]. Полученная в результате таксономия концептов рассматривается как

скелетная онтология, которая в дальнейшем может наращиваться и расширяться. Данный процесс включает несколько шагов лингвистического анализа и предобработки текстов.

Обобщенный алгоритм построения иерархии концептов на основе обработки текстов строится следующим образом [3,4]:

- первоначально осуществляется лексико-грамматический анализ текстовой коллекции с использованием деревьев разметки, а также синтаксический анализ для порождения дерева разбора для каждого предложения;
- затем из деревьев разбора извлекаются зависимости глагол /субъект, глагол/объект, глагол/предложная группа;
- таким образом, результатом процесса извлечения должны быть пары, состоящие из глагола и ядра (заголовка) субъекта, объекта или предложной группы, которые подлежат субкатегоризации;
- следующим шагом является нормализация (приведение к словарной форме) глаголов и существительных – ядер полученных синтаксических групп;
- во избежание разреженности данных применяется шаг сглаживания коллекции пар, т.е. пары, которые на появились в коллекции, оцениваются на базе частоты появления остальных пар;
- далее пары взвешиваются в соответствии с определенной статистической мерой;
- только пары, которые превысили определенный порог, трансформируются в формальный контекст, к которому применяется метод FCA;
- решетка, полученная таким образом (\mathcal{B}, \leq) , преобразуется в отношение частичного порядка (C', \leq') , которое ближе к иерархии концептов в традиционном смысле;
- поскольку применение метода FCA может привести к возрастанию количества концептов, отношение частичного порядка формируется на шаге подрезки (pruning) путем удаления абстрактных концептов;
- полученное отношение частичного порядка (C'', \leq'') составляет результирующую иерархию концептов.

Более формально процесс построения иерархии концептов с применением метода FCA может быть представлен с помощью следующего алгоритма:

```
/*construct a hierarchy for the terms in T on the basis of the documents in D*/
```

1. Parses = parse(POS-tag(D));
2. SynDeps = tgrep(Parses);
3. lemmatize(SynDeps);
4. smooth(SynDeps);
5. weight(SynDeps);
6. SynDeps' = applyThreshold(SynDeps);
7. K = getFormalContext(T, SynDeps');
8. $(\mathcal{B}, \leq) = \text{computeLattice}(K)$;
9. $(C', \leq') = \text{transform}(\mathcal{B}, \leq)$;
10. $(C'', \leq'') = \text{compact}(C', \leq')$;
11. return (C'', \leq'') .

REFERENCES

1. Ganter B., Stumme G., Wille R. Formal Concept Analysis. Theory and Applications // J.UCS Special Issue.- vol.10, No.8, July 2004.
2. Cimiano P. Ontology Learning and Population from Text. Algorithms, Evaluation and Applications.- Springer, 2006. – 347 p.
3. Рябова Н.В., Золотухин О.В. Побудова онтологічних баз знань із застосуванням методів інтелектуальної обробки текстів // Матеріали XIV Міжнар. наук. конф. «Інтелектуальні системи прийняття рішень та проблеми обчислювального інтелекту» ISDMCI'2018, 21-27 травня 2018 р., Залізний Порт, Україна. – Херсон, ХНТУ, 2018. – С. 200-201.
4. Золотухин О.В., Рябова Н.В. Обучение онтологий на основе методов интеллектуального анализа текстовых документов // Сучасні проблеми та перспективи соціально-економічного, інформаційного та науково-технічного розвитку підприємств України: Всеукр. науково-практ. конф. / Маріуполь (2 січня 2018 р.): тези доп. / редкол.: М.Г. Белопольський [та ін.]; ДВНЗ «ПДТУ». – Маріуполь: ПДТУ, 2018. – С.306-309.