

Субъективные метрики оценки онтологий¹

Гаврилова Т.А.¹, Горовой В.А.¹, Болотникова Е.С.², Горелов В. В.³

¹Высшая школа менеджмента, Санкт-Петербургский государственный университет,
Волховский пер., 1-3, г. Санкт-Петербург, 199004, Россия

²Санкт-Петербургский государственный политехнический университет,
ул. Политехническая д.29, 195251, г. Санкт-Петербург, Россия

³Белорусский государственный университет информатики и радиоэлектроники,
ул. П.Бровки д.6, 220013, г. Минск, Белоруссия

gavrilova@gsom.pu.ru, gorovoy@gsom.pu.ru,
bolotnikovakate@gmail.com, golen@bsuir.by

Аннотация. В статье рассматривается вопрос оценки качества онтологий. Приведена классификация существующих методов оценки онтологий и предложены субъективные метрики воспринимаемости онтологий человеком. Кроме того, предложен вариант автоматизации оценки когнитивной эргономичности на основе анализа графа.

Ключевые слова: онтология, онтологический инжиниринг, оценка онтологий, инженерия знаний

1 Введение

Вопросом оценки качества создаваемых онтологий является одной из актуальных проблем современного онтологического инжиниринга. Эта часть процесса разработки онтологий важна в практическом плане, и это является причиной того, что разными группами ученых разработано множество различных подходов в области оценки онтологий. В настоящее время известно более десятка методов, и задача выбора подходящей методики для решения конкретной задачи становится все более сложной.

Хорошие обзоры этих существующих методов и подходов в оценке онтологий изложены в работах [11, 6]. Качественные критерии оценки с позиций гештальт-психологии предложены в работах Гавриловой [2, 3].

Существующие методы оценки онтологий ставят перед собой одну из следующих целей:

- Полнота и точность словаря предметной области. Такую задачу решают подходы, описанные в работах [14] и [7].
- Адекватность структуры с точки зрения таксономии, отношений и т.п. Наиболее известна в этом отношении формальная онтология метасвойств OntoClean [10]. Другие методы изложены в статьях [7, 17, 14]
- Воспринимаемость (с когнитивной точки зрения). Гештальт-подход описан в работе [9]. Внимание этому аспекту уделено также в работе [8].
- Производительность при использовании в приложениях (см. [16]).
- Выбор лучшей онтологии из нескольких имеющихся. Как правило, такие работы используют различные метрики, например Ontometric [13]

¹ Работа частично поддержана грантами РФФИ и грантом СПбГУ

Оценка онтологий может применяться на разных стадиях разработки и использования онтологий. Возможны указанные ниже варианты:

- Разработка и прототипирование [10]
- Тестирование перед выпуском [13]
- Использование [16]

По степени автоматизации все методы оценки онтологий можно разделить на три группы:

1. Автоматические. Например, EvaLexon [17].
2. Полуавтоматические [8]
3. Ручные [13]

Объектами для анализа у существующих методов могут быть один или несколько из перечисленных ниже объектов:

4. Структура [8]
5. Словарь [7]
6. Эффективность практического использования ([16])

По средствам, используемым для анализа качества и зрелости онтологий, можно все методы разделить на несколько классов:

7. Основанные на данных (Data-driven) [7]
8. Экспертные оценки [13]
9. Исследования профилей использования [16]
10. Сравнение с "золотым стандартом" [14, 17]
11. Исследования топологии графа онтологии [5, 8].

Вопросы, относящиеся субъективному восприятию онтологий, т.е. к когнитивной эргономике (разделу эргономики, в котором изучаются интерфейсы взаимодействия человека с программным обеспечением, на основе изучения его мышления, памяти, восприятия и т.д.) при оценке онтологий, рассматриваются в работах [9] и [8]. Первая из этих работ послужила отправной точкой для формирования принципов, лежащих в основе воспринимаемости онтологий, изложенных ниже. Во второй работе предлагается несколько метрик, которые могут использоваться для оценки когнитивной эргономичности.

С точки зрения данной на рис. 1 классификации предлагаемая в работе модель оценки можно описать следующим образом:

- Цель: оценка воспринимаемости (с когнитивной точки зрения), выбор лучшей онтологии из нескольких имеющихся
- Объект анализа: структура онтологии
- Средство анализа: анализ топологии графа онтологии
- Степень автоматизации: автоматический, полуавтоматический (окончательное решение принимает эксперт на основе посчитанной автоматически модели)
- Стадия применения: разработка и прототипирование (подсчет может производиться на каждой следующей итерации разработки), тестирование перед выпуском

2 Исследования качества онтологии на основе анализа графа

Подход к оценке качества на основе топологии графа онтологии использован в работах [5, 8].

Во второй работе приводится достаточно много метрик, используемых для анализа качества онтологии, часть из которых рассчитывается на основе топологии графа онтологии. Приведем

здесь те из них, которые относятся к метрикам когнитивной эргономичности (в контексте этих метрик будем далее рассматривать только is-A дуги в качестве ребер графа):

- **Глубина онтологии.** Gangemi выделяет три метрики для подсчета глубины:
 - Абсолютная глубина. Вычисляется как сумма длин всех путей графа (где путем называется любая последовательность соединенных между собой вершин, начинающаяся от корневой вершины и заканчивающаяся листом графа).
 - Средняя глубина. Равняется абсолютной глубине деленной на количество путей в графе.
 - Максимальная глубина. Равняется максимальной длине пути.

Чем больше глубина, тем труднее граф поддается восприятию.

- **Ширина онтологии.**
 - Абсолютная ширина. Равняется сумме количества вершин для каждого уровня иерархии по всем уровням.
 - Средняя ширина. Вычисляется как абсолютная ширина, деленная на количество уровней иерархии.
 - Максимальная ширина. Равняется количеству вершин на самом большом по количеству вершин уровне.

Чем она меньше, тем лучше с точки зрения когнитивной эргономики.

- **Запутанность онтологии (tangledness).** Определяется как количество вершин графа онтологии, деленное на количество вершин, у которых есть несколько непосредственных суперклассов. Таким образом, в онтологиях, где нет множественного наследования (связи is-A), эта метрика будет равняться нулю. Чем меньше итоговое значение, тем лучше онтология с точки зрения когнитивной эргономики.
- **Отношение количества классов к количеству свойств.** Чем больше, тем легче воспринимать онтологию.
- **Количество анонимных классов.** Для улучшения воспринимаемости, лучше минимизировать их число.

Несмотря на полезность этих метрик, они покрывают лишь очень небольшую часть факторов, влияющих на наше восприятие и способность к запоминанию. В следующем разделе изложены принципы, влияющие на когнитивные способности человека, которые легли в основу предлагаемой модели метрик.

3 Принципы, лежащие в основе оценки когнитивной эргономичности

Основные принципы оценки визуальной воспринимаемости и понятности онтологий изложены в работах [2, 3, 9]. В их основе лежат взгляды Макса Вертгеймера в области гештальт-психологии [1]. Он рассматривал все задачи с точки зрения незавершенности или несовершенства структуры. Так основным принцип хорошего гештальта (хорошей формы) или закон прегнантности был сформулирован так:

«Организация любой структуры в природе или в сознании должна быть настолько хороша (регулярна, полна, сбалансирована или симметрична), насколько позволяют существующие условия».

Также полезными могут быть и другие когнитивно-перцептивные принципы:

- **Закон близости** – визуальные стимулы (объекты), находящиеся *близко* друг от друга воспринимаются как единое *целое*.
- **Закон сходства** – вещи, обладающие *одинаковыми свойствами*. Обычно воспринимаются как нечто *единое* (цельное).

- Закон включения В.Келера – есть тенденция воспринимать только большую фигуру, а не ту меньшую которую она включает.
- Закон парсимонии – самый простой пример является самым лучшим, известен как принцип «бритвы Оккама»: «не нужно умножать сущности без необходимости».

Для целей онтологического инжиниринга эти законы можно переформулировать и сделать применимыми для практического инженера по знаниям. Основная гипотеза может быть сформулирована как: «*Гармония = концептуальный баланс + ясность*».

При этом концептуальный баланс подразумевает, что

- Понятия одного уровня иерархии связываются с родительским концептом одним и тем же типом отношения (например, «класс-подкласс» или «часть-целое»).
- Глубина ветвей онтологического дерева должна быть примерно одинаковая (± 2).
- Общая картинка должна быть довольно симметричной.
- Перекрестные ссылки должны быть по возможности исключены.

Ясность включает

- Минимизацию. Так максимальное число концептов одного уровня или глубина ветви не должна превышать знаменитое число Ингве-Миллера (7 ± 2) [15].
- Прозрачность для чтения. Тип отношений должен быть по возможности очевиден, так чтобы не перегружать схему онтологии лишней информацией и опускать названия отношений.

Полученная в данной работе модель оценки качества онтологии основана на принципах, изложенных выше, и во многом является их формализацией. Целью применения модели является оценка сбалансированности и воспринимаемости онтологий пользователями.

4 Субъективные метрики оценки качества онтологии

К субъективным метрикам мы относим все метрики, оценивающие воспринимаемость онтологии человеком. На качество онтологии с когнитивной точки зрения влияют множество аспектов, многие из которых описаны в предыдущем разделе. Предлагаемая модель оценки состоит из нескольких метрик, на основе измерения которых, можно сделать некоторые выводы о качестве онтологий. Они рассмотрены ниже.

Примем следующие обозначения:

- g – граф, представляющий онтология. Концепты (классы и экземпляры) онтологии являются вершинами графа, отношения между концептами представлены в виде ребер графа
- G – множество всех вершин графа g
- E – множество всех ребер графа g

4.1 Метрики циклов

Наличие циклов не помогает восприятию. Желательно, чтобы их не было вообще. Ниже приводятся две метрики для измерения степени цикличности графа:

- Количество различных циклов в графе. В хорошей онтологии оно должно быть равно нулю
- Количество вершин, входящих в какой-нибудь цикл деленное на количество вершин в графе. Чем значение этой метрики меньше, тем лучше для онтологии, оптимальное значение равно нулю.

$$m = \frac{N_{v \in C}}{n_G},$$

где n_G — количество вершин графа, C — множество вершин графа, входящих в хотя бы один цикл, $N_{v \in C}$ — количество вершин графа, входящих в какой-нибудь цикл.

4.2 Метрики Ингве-Миллера

Основная идея этих метрик в том, что количество связей у одного концепта не должно превышать 7 ± 2 [15]. Таким образом, будем называть вершинами с нормальной степенью (степень вершины — количество дуг, для которых она является концевой) вершины, у которых суммарное число входящих и исходящих ребер не превышает 9. Предлагаются следующие метрики Ингве-Миллера:

- Отношение количества вершин с нормальной степенью по отношению ко всем вершинам. Чем ближе оно к 1, тем лучше онтология с точки зрения когнитивной эргономики.

$$m = \frac{N_{v \in GD}}{n_G},$$

где n_G — количество вершин графа, $GD = \{v \in G | \text{deg}(v) \leq 9\}$ — множество вершин с нормальной степенью, $N_{v \in GD}$ — количество вершин графа с нормальной степенью

- Средняя степень вершины графа.

$$m = \frac{\sum_{v \in G} \text{deg}(v)}{n_G} = \frac{2 * n_E}{n_G},$$

где $\sum_{v \in G} \text{deg}(v)$ — сумма всех степеней вершин графа, n_E — количество ребер графа.

- Медиана степени вершины графа.

$$m = \text{deg}(\widetilde{v}),$$

где $\text{deg}(\widetilde{v})$ — медиана степени вершины графа (возможное значение степени вершины графа, которое делит ранжированную совокупность вершин графа на две равные части: 50 % «нижних» единиц ряда данных будут иметь значение степени вершины не больше, чем медиана, а «верхние» 50 % — значения степени вершины не меньше, чем медиана).

- Среднее квадратичное отклонение степени вершины графа.

$$m = \frac{\sum_{v \in G} (\text{deg}(v) - \frac{\sum_{v \in G} \text{deg}(v)}{n_G})^2}{n_G - 1} = \frac{\sum_{v \in G} (\text{deg}(v) - \frac{2 * n_E}{n_G})^2}{n_G - 1}$$

4.3 Метрики разнообразия количества связей

Чем больше различных типов связей используется в онтологии, тем сложнее она для восприятия. Многие виды онтологии (генеалогии, например), строятся на одном типе связи. Но иногда встречается несколько типов связей и в этом случае влияние на восприятие можно оценить следующим образом:

- Количество различных типов связей в графе.

$$m = N_{t \in TE},$$

где $TE = \{type(e) | e \in E\}$ — множество всех типов связей графа, $N_{t \in TE}$ — количество различных типов связей.

- Нормированное количество различных типов связей, т.е. количество различных типов связей, деленное на количество концептов (вершин графа).

$$m = \frac{N_{t \in TE}}{n_G}$$

4.4 Метрики глубины

Если длины различных путей в графе сильно отличаются, то это вредит сбалансированности онтологии и качеству ее восприятия. Для измерения показателей, связанных с глубиной графа предлагаются следующие метрики:

- Метрики глубины, предложенные в работе [8]:
 - Абсолютная глубина.
 - Средняя глубина.
 - Максимальная глубина.
- Минимальная глубина.

$$m = N_{j \in P} \\ \forall i (N_{j \in P} \leq N_{i \in P}),$$

где $N_{j \in P}$ и $N_{i \in P}$ — длины пути j и i из множества путей P графа g .

- Медиана глубины.

$$m = \widetilde{N_{j \in P}},$$

где $\widetilde{N_{j \in P}}$ — медиана глубины графа (возможное значение глубины графа, которое делит ранжированную совокупность путей графа на две равные части: 50 % «нижних» единиц ряда данных будут иметь значение длины пути не больше, чем медиана, а «верхние» 50 % — значения длины пути не меньше, чем медиана).

- Среднее квадратичное отклонение глубины.

$$m = \frac{\sum_j^P (N_{j \in P} - \frac{\sum_j^P N_{j \in P}}{n_{P \subseteq g}})^2}{n_{P \subseteq g} - 1}$$

- Среднее квадратичное отклонение глубины по отношению к средней глубине

$$m = \frac{\frac{\sum_j^P (N_{j \in P} - \frac{\sum_j^P N_{j \in P}}{n_{P \subseteq g}})^2}{n_{P \subseteq g} - 1}}{\frac{\sum_j^P N_{j \in P}}{n_{P \subseteq g}}}$$

4.5 Метрики запутанности (tangledness) графа

Чем чаще множественное наследование используется в онтологии, тем хуже она с точки зрения когнитивной эргономики. Для того чтобы оценить степень его использования, предлагаются следующие метрики (в данном разделе мы опять рассматриваем только часть графа с *is-a* отношениями):

- Количество вершин с множественным наследованием по отношению ко множеству всех вершин графа (в работе [8] дается обратное соотношение, что менее удобно, т.к. значение целевой функции не лежит в интервале $[0, 1]$).

$$m = \frac{N_{v \in MI}}{n_G},$$

где $MI = \{v \in G | \exists a_1, a_2 (isa(v, a_1) \wedge isa(v, a_2))\}$ — множество всех вершин графа, с более, чем одной входящей дугой отношения *is-a*, $N_{v \in MI}$ — количество всех элементов этого множества.

- Среднее количество родительских вершин у вершины графа

$$m = \frac{1}{n_G} \sum_v^G N_{S_v \in G},$$

где $S_v = \{a \in G | isa(v, a)\}$ — множество всех родителей вершины v , $N_{S_v \in G}$ — количество всех родительских вершин вершины v .

4.6 Метрики измерения ветвистости (fan-outness) графа

Позволяют оценивать «распределение» вершин графа, в котором в качестве дуг рассматриваются только дуги отношения is-a (или другое отношение, в случае, когда оно является основным в онтологии). Можно разделять внутреннюю ветвистость, и ветвистость у листьев графа. Достаточно много метрик ветвистости предложено в работе [Gangemi, 2005]. Ниже представлены метрики ветвистости, предлагаемые в качестве дополнения авторами:

- Количество вершин, у которых есть и листья и нелистовые ноды в качестве детей, по отношению ко всем кол-ву вершин у которых есть листья среди детей.

$$m = \frac{N_{v \in S_{LEA \& SIB}}}{N_{v \in S_{LEA}}},$$

где $S_{LEA \& SIB}$ — множество всех вершин, которые являются одновременно и родителями листьев графа и родителями внутренних вершин, $N_{v \in S_{LEA \& SIB}}$ — количество таких вершин; S_{LEA} — множество всех вершин, у которых среди детей есть листья графа, $N_{v \in S_{LEA}}$ — количество таких вершин.

- Минимальное количество детей-листьев у предпоследних вершин в графе.

$$m = N_{j \in SIB}^{j \subseteq LEA}$$

$$\forall i (N_{j \in SIB}^{j \subseteq LEA} \leq N_{i \in SIB}^{i \subseteq LEA}),$$

где $N_{j \in SIB}^{j \subseteq LEA}$ — количество листьев набора j , имеющих общего родителя.

- Среднее квадратичное отклонение детей-листьев у предпоследних вершин в графе.

$$m = \frac{\sum_{j \in SIB_{LEA}} (N_{j \in SIB}^{j \subseteq LEA} - \frac{\sum_{j \in SIB_{LEA}} N_{j \in SIB}^{j \subseteq LEA}}{n_{SIB_{LEA}}})^2}{n_{SIB_{LEA}} - 1},$$

где $n_{SIB_{LEA}}$ — количество всех наборов листьев, имеющих общего родителя, $N_{j \in SIB}^{j \subseteq LEA}$ — количество листьев в группе детей, имеющих общую родительскую вершину.

Основным преимуществом всех представленных выше метрик, является возможность их автоматического вычисления на основе анализа графа онтологии. Основная модель их использования следующая:

1. Эксперт выбирает метрики, которые считает существенными для принятия решения о качестве онтологии
2. Метрики вычисляются
3. Для некоторых метрик срабатывают триггеры, указывающие на проблемы онтологии с точки зрения когнитивной эргономики
4. Для остальных метрик эксперт анализирует результаты всех метрик в совокупности и принимает решение об окончательной оценке качества.

5 Инструментарий автоматического расчета метрик модели

Существенная особенность всех рассматриваемых выше метрик — возможность их автоматического вычисления. Это может серьезным образом упростить работу эксперта при оценке онтологий с большим количеством концептов (если в онтологии меньше 50 концептов, то опытный эксперт может ее оценить практически одним взглядом с точки зрения когнитивной эргономичности).

Авторами был разработан инструмент COAT (Cognitive Ontology AssessmentT) автоматической оценки когнитивной эргономичности онтологий. Кроме непосредственного вычисления метрик у пользователя COAT есть возможность получить информацию об этих метриках, их назначении и трактовке значений из встроеного в инструменте словаря.

Инструмент COAT на данном этапе реализован в виде консольного приложения на языке Java. Поддерживаются онтологии в формате OWL, работа с которыми в программе осуществляется с помощью Jena Semantic Web Framework [12] – библиотеки Java классов для работы с RDF и OWL онтологиями.

6 Заключение

Предложенные метрики могут выступать как дополнительные и помогут экспертам в оценке качества онтологий в совокупности с другими метриками. В некоторых случаях предложенная в статье модель не применима, т.к. нет необходимости оценивать когнитивную эргономичность онтологии. Например, если важна только вычислительная эффективность, и с создаваемой онтологией будут иметь дело только программы, а не люди.

Кроме того, при использовании предложенных методов важно, чтобы были образцы для сравнения (хотя есть и метрики, для которых есть рекомендуемые абсолютные границы значений, которые желательно соблюдать при создании онтологий). Обычно такие примеры имеются, т.к. процесс разработки онтологий, как правило, итеративный. Предложенная модель оценки может помочь понять, что следует изменять в описании предметной области, чтобы оно стало лучше с точки зрения когнитивной эргономики. Таким образом, каждая следующая версия онтология будет лучше и быстрее восприниматься ее пользователями.

Методы оценки когнитивной эргономичности также можно применять для оценки онтологий одной и той же предметной области, сделанных разными людьми/командами. Посчитанные метрики помогут понять, какая из них лучше с точки зрения когнитивной эргономики и сделать выбор в пользу одной из них в случае, если оценки других важных критериев отличаются принципиально.

Оценка когнитивной эргономичности онтологий важна в случаях, когда онтология предназначена для обучения и для передачи знаний. При этом существующие на сегодняшний день модели оценки онтологий не позволяли в полной мере произвести их анализ с точки зрения качества и скорости ее восприятия людьми. Предлагаемая модель позволяет восполнить этот пробел.

Большая часть метрик, используемых в ее рамках, может считаться автоматически, что существенно уменьшает нагрузку на эксперта, принимающего решение об итоговой оценке качества онтологии.

Предложенная модель также позволяет разработчикам онтологий понимать, улучшается ли их онтология с точки зрения когнитивной эргономики, сравнивая последующие итерации и результатами предыдущих. Описанные результаты могут эффективно применяться в системах управления знаниями на предприятиях, в образовательном процессе и помогать специалистам и студентам создавать более качественные онтологии.

Список литературы

- [1] Вертгеймер М. *Продуктивное мышление*. М.: Прогресс, 1987.
- [2] Гаврилова Т. Об одном подходе к онтологическому инжинирингу // Ж. «Новости искусственного интеллекта», N3, 2005. – с.25-31.
- [3] Гаврилова Т. Гештальт-принципы построения онтологий // Тезисы докладов Второй международной конференции по когнитивной науке, Том 1, СПбГУ, 2006. – с.240-242.
- [4] Горовой В.А. Модель классификации методов оценки онтологий // Материалы 2-й международной молодежной конференции «Искусственный интеллект: философия, методология, инновации». Санкт-Петербург, 15-17 ноября 2007 – с. 307-310
- [5] Клещев А.С., Шалфеева Е.А. Каталог свойств онтологий. принципы организации каталога: Препринт, 2007. Владивосток: ИАПУ ДВО РАН, 2007. - 20 с.
- [6] Brank J., Grobelnik M., Mladenic D. A survey of ontology evaluation techniques. In In Proceedings of the Conference on Data Mining and Data Warehouses (SiKDD 2005), Ljubljana, Slovenia, 2005. <http://kt.ijs.si/dunja/sikdd2005/Papers/BrankEvaluationSiKDD2005.pdf>

- [7] *Brewster C.* et al. Data driven ontology evaluation. Proceedings of Int. Conf. on Language Resources and Evaluation, Lisbon, 2004.
- [8] *Gangemi A., Catenacci C., Ciaramita M., Lehmann J.* Ontology evaluation and validation. An integrated formal model for the quality diagnostic task. http://www.loa-cnr.it/Files/OntoEval4OntoDev_Final.pdf
- [9] *Gavrilova T., Gorovoy V.* Technology for ontological engineering lifecycle support. //International Journal "Information Theories & Applications" Vol.14 / 2007 – p. 19-25
- [10] *Guarino N., Welty C.*, Evaluating ontological decisions with OntoClean. Comm. of the ACM, 45(2):61–65, February 2002.
- [11] *Hartmann J.* Et al. Methods for ontology evaluation. Knowledge Web Deliverable D1.2.3, 2005. <http://www.starlab.vub.ac.be/research/projects/knowledgeweb/KWeb-Del-1.2.3-Revised-v1.3.1.pdf>
- [12] Jena Semantic Web Framework homepage. <http://jena.sourceforge.net/>
- [13] *Lozano-Tello A., Gomez-Perez A.*, Ontometric: A method to choose the appropriate ontology. J. Datab. Mgmt., 15(2):1–18) 2004.
- [14] *Maedche A., Staab S.*, Measuring similarity between ontologies. Proc. CIKM 2002. LNAI vol. 2473.
- [15] *Miller G.* The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. The Psychological Review, 1956. 63: 81–97.
- [16] *Porzel R. and Malaka R.* A Task-based Approach for Ontology Evaluation. Proc. of ECAI 2004.
- [17] *Spyns P.*, EvaLexon: Assessing triples mined from texts. Technical Report 09, STAR Lab, Brussels, Belgium, 2005.