

ОНТОЛОГИИ И ОНТОЛОГИЧНИ СИСТЕМИ

ОСНОВНИ ОПРЕДЕЛЕНИЯ

Онтология (от древногръцки: онтос – битие, съществуване; логос – учение, наука) е термин, определящ науката за битието, за съществуващото, в отличие от **гносеологията** – науката за познанието. Терминът “онтология” в съвременната философска литература се използва за означаване на определена система от категории, които са следствие от определена система от възгледи (определена гледна точка) за света.

В литературата по изкуствен интелект “онтология” е термин, който се използва за означаване на формално представени знания на основата на някаква **концептуализация**. Концептуализацията предполага описание на множество от обекти и понятия, знания за тях и връзки между тях.

Според Грубер **онтология се нарича експлицитната спецификация на концептуализацията**. Формално онтологията се състои от термини, организирани в таксономия, техните определения и атрибути, а също и свързаните с тях аксиоми и правила за извод.

Често множеството от предположения, съставлящи онтологията, има формата на логическа теория от първи ред, в която термините от речника са имена на унарни и бинарни предикати, наричани съответно *концепти (категории)* и *отношения*. В най-простия случай онтологията описва само йерархията на концептите, свързани с отношенията на категоризация. В по-сложните случаи към нея се добавят подходящи аксиоми за изразяване на други отношения между концептите и за ограничаване на предполагаемите им интерпретации.

В този смисъл онтологията е *база от знания, описваща факти, за които се предполага, че са винаги верни в рамките на определена взаимна общност* на основата на общоприетия смисъл на използвания речник.

Още по-конкретно е дефинирано понятието онтология в известния проект Ontolingua на Стенфордския университет. Там се предполага, че *онтологията е експлицитна спецификация на определена тема*. Този подход предполага формално и декларативно представяне на някаква тема, което включва речник (или списък на константи) за насочване към термините от предметната област, ограничения върху общността на термините, логически твърдения, които ограничават интерпретацията на термините и това, как те се отнасят един към друг.

В специализираната литература напоследък се налага следното определение: *онтологиите са БЗ от специален тип, които могат да се “четат” и разбират, да се отделят от разработчика и/или физически да се поделят между техните потребители.*

Всяка онтология се основава на някаква концептуализация, но една концептуализация може да служи за основа на различни онтологии и различни БЗ могат да отразяват една и съща онтология. В този смисъл понятието онтология се дефинира от някои автори в контекста на споделянето на знания. Така например според Грубер *онтологиите са споразумения (agreements) относно споделените концептуализации.*

Споделените концептуализации включват структури от понятия за моделиране на знанията за дадена област, протоколи със специфично съдържание за комуникация между взаимодействащите си агенти и споразумения за представянето на конкретни теории за дадена област. В контекста на споделянето на знания онтологиите са специфицирани под формата на дефиниции в представящи ги речници. В най-простия случай те се състоят от йерархии на типове, специфициращи класове, и релации между тях.

КЛАСИФИКАЦИЯ

Онтоологиите могат да бъдат класифицирани в зависимост от различни класификационни признаци. Някои автори разглеждат онтоологиите като ориентирани към понятията (концепциите), а други – към обектите в предметната област.

В литературата се говори и за зависимост и независимост на онтологията от конкретна предметна област или задача.

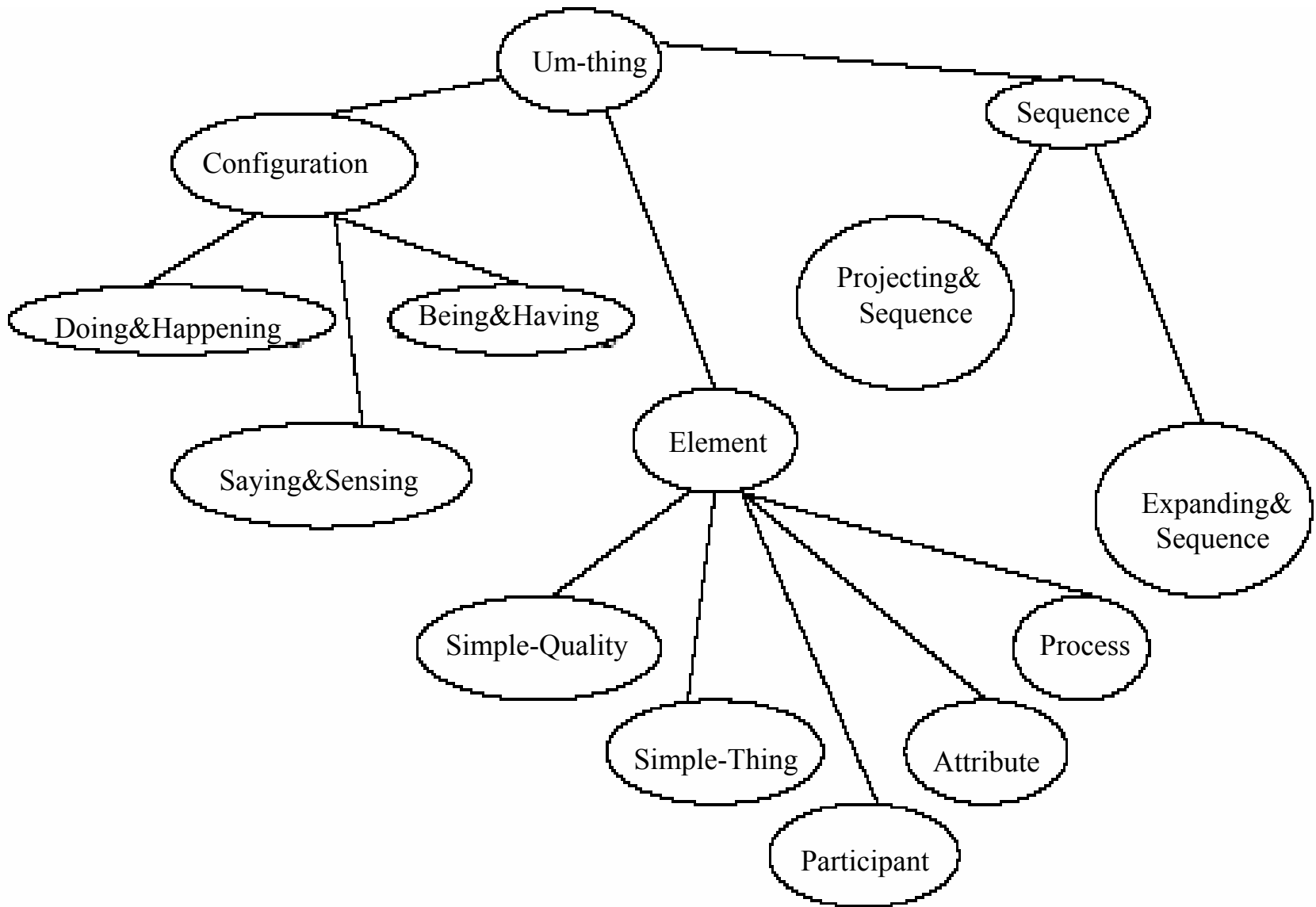
С други думи, предлага се класификация, според която онтологиите се разделят на: *общи онтологии; онтологии, ориентирани към конкретна област; онтологии, ориентирани към конкретна задача и приложни онтологии.*

Общите онтологии (онтологиите от най-високо ниво) са предметно независими онтологии, съдържащи top-level категории (концепти) като пространство, време, обект, събитие, действие, количество, мярка и т.н., които са независими от конкретната задача или област. Затова е целесъобразно те да бъдат унифицирани за големи общности от потребители.

Пример за такава обща онтология е Сус (Д. Ленат и др.). Едноименният проект Сус на Suscorp е ориентиран към създаването на мултиконтекстна база от знания и специална машина за извод. Основната цел на този огромен проект е построяването на база от знания за всички общи понятия (започвайки с понятия от типа на време, същност и т.н.), която включва семантичната структура на термините, връзките между тях и множество аксиоми.

Предполага се, че такава база от знания може да бъде достъпна за разнообразни програмни средства, работещи със знания, и ще играе ролята на база от “начални знания”. Според някои литературни източници в тази онтология вече са представени около 10^5 категории и 10^6 аксиоми. За предствяне на знанията в рамките на този проект е разработен специален език, наречен СуcL.

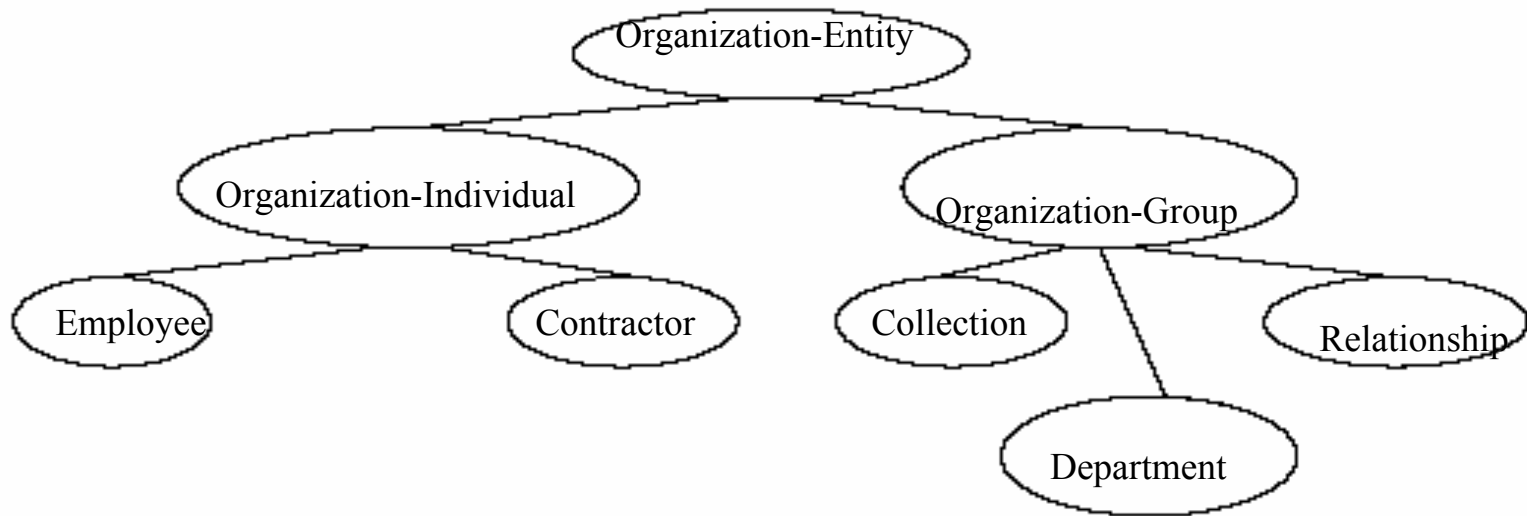
Друг пример за обща онтология е онтологията на системата Generalized Upper Model, ориентирана към поддръжка на процесите на обработка на естествен език: английски, немски и италиански. Нивото на абстракция на тази онтология се намира между лексическите и концептуалните знания, което се определя от изискванията за опростяване на интерфейсите с лингвистичните ресурси. Моделът на Generalized Upper Model включва таксономия, организирана под формата на йерархия между категориите (около 250 понятия) и отделна йерархия на връзките. Фрагмент от системата от понятия на тази онтология е показан на следващата фигура.



Като цяло може да се констатира, че независимо от постигането на отделни успехи, създаването на достатъчно общи онтологии от най-високо ниво е много сериозна задача, която все още няма задоволително решение.

Предметните онтологии и онтологиите на задачи описват съответно речник, свързан с предметната област (медицина, търговия и т.н.) или с конкретната задача или дейност (диагностика, продажби и т.н.) за сметка на специализация на термините, въведени в общата онтология. Примери за онтологии, ориентирани към определена предметна област и задача, са съответно TOVE и Plinius.

Онтологията в системата TOVE (Toronto Virtual Enterprise Project) е предметно ориентирана към представянето на модел на корпорация. Основната цел на нейната разработка е отговарянето на въпроси на потребители по реинженеринг на бизнес процеси, извличайки експлицитно (явно) представени в онтологията знания. При това системата може да извършва дедуктивен извод на отговорите. В тази онтология няма средства за интегриране с други онтологии. Формално онтологията се описва с помощта на фреймове. Таксономията на понятията на TOVE е показана на следващата фигура.



Приложните онтологии описват концепти, които зависят както от конкретната предметна област, така и от задачите, които ще се решават. Концептите в такива онтологии често съответстват на ролите, които играят обектите в предметната област в процеса на изпълнение на определена дейност.

Пример за такава онтология е онтологията на системата Plinius, предназначена за полуавтоматично извличане на знания от текстове в областта на химията. За разлика от другите споменати по-горе онтологии тук няма явна таксономия на понятията. Вместо това са определени няколко множества от атомарни концепти (като например химичен елемент, цяло число и т.н.) и правила за конструиране на останалите концепти. В онтологията са описани около 150 концепта и 6 правила. Формално онтологията на Plinius също се описва с помощта на фреймове.

За авторите, които поддържат концептуалната идея, от особена важност е *степента на детайлност, използвана за характеризирание на концепциите* в онтологиите. Тези с по-детайлно описание са по-добри, но за тях е необходимо използването на по-богат речник. По-опростените могат да бъдат разработвани с цел да бъдат споделяни между различните потребители, възприели и споразумели се предварително за намиращата се в основата на онтологията концептуализация.

Друга класификация, която се поддържа от някои автори, се основава на типа на включения в онтологията речник на понятията. В зависимост от типа на речника си онтологиите се разделят на четири типа: *силно неформални* (изразени в свободна форма на естествен език), *полунеформални* (изразени в ограничена и специално структурирана форма на естествен език), *полуформални* (изразени в термините на изкуствено създаден език), *формални* (с точно дефинирани термини и формално описана семантика, с теореми за пълнота и съответни доказателства).

СИСТЕМИ И СРЕДСТВА ЗА ПРЕДСТАВЯНЕ НА ОНТОЛОГИЧНИ ЗНАНИЯ

Понастоящем в целия свят се извършват сериозни изследвания, свързани с пространството на знанията в средата на Интернет. Един от ключовите аспекти на тези изследвания са алгоритмичните и програмните средства за представяне на онтологични знания и работа с онтологии. Обща цел на практически всички популярни проекти в тази област е разработването на нови подходи към построяването на пространството на знанията и средства за работа с тях, които осигуряват:

- използване на семантика за управление на процеса на изпълняване (отговаряне) на заявки;
- програмни възможности за построяване на отговори с добре определена семантика и прост синтаксис, които биха могли да бъдат “разбрани” и обработени от програмни агенти или други средства;
- възможности за хомогенен достъп до информация, която физически е разпределена и е хетерогенно представена в Интернет;
- получаване на информация, която не присъства явно сред фактите, намерени в мрежата, но може да бъде изведена от други факти и базови знания.

Инициативата (КА)² и инструментариумът на проекта Ontobroker

(КА)² означава Knowledge Annotation Initiative of the Knowledge Acquisition Community. Цел на работата по този международен проект е в крайна сметка създаването на средства за интелигентно търсене в средата на Интернет и автоматично натрупване на нови знания. В рамките на инициативата (КА)² могат да бъдат обособени три основни направления на изследванията:

- онтологичен инженеринг (ontological engineering);
- анотация на Web страници;
- заявки за търсене на информация от Web страници и извод на отговори на базата на онтологични знания.

Онтологичният инженеринг е едно от основните направления, в рамките на което се предполага, че съвместната общност на (КА)² трябва да създаде своя собствена и достатъчно обща онтология на основата на използване на средства на Ontolingua.

Към настоящия момент са разработени осем онтологии, които могат да се разглеждат като раздели на обща онтология: онтология на организация (organization ontology), на проект (project ontology), на личност (person ontology), направление на изследвания (research-topic ontology), публикации (publication ontology), събития (event ontology), изследователски продукти (research-product ontology) и изследователски групи (research-group ontology).

При това разработването на онтологии се осъществява и управлява от участници в проекта, които се наричат *provider agents*, а така създадените онтологии се помещават на Web страници. Такива страници се аотират с използване на нов тип тагове (команди) на HTML (ONTO), информацията в рамките на които се обработва от специален компонент, който работи на основата на онтологии – системата *Ontocrawler*. В рамките на този компонент, в зависимост от “богатството” на използваната онтология, може да се извежда нова информация, която е релевантна на заявките, но не присъства явно на наличните Web страници.

Самата система Ontocrawler се разработва в рамките на отделен проект от инициативата (КА)² – проекта Ontobroker, който е интересен от гледна точка на средствата за представяне и обработка на онтологични знания, които поддържа.

Ontobroker има три основни подсистеми: интерфейс за формулиране на заявки (query engine), машина за извод на отговори (inference engine) и машина за достъп до ресурси на Интернет – “червей” (Webcrawler), която се използва за натрупване на необходимите знания от тази среда.

За специфицирането на онтологии е разработен специален език за представяне на знания. Подмножество на този език е предназначено за формулиране на заявки, а езикът за аотиране е предназначен за обогатяване на Web документи с онтологична информация.

Формализмът на заявките е ориентиран към фреймово представяне на онтологиите, в рамките на което са определени понятията клас, екземпляр, атрибут и стойност.

В системата Ontobroker се поддържат два типа интерфейси при формирането на заявките: текстов (за експертите) и графичен (за обикновените потребители). Първият от тях предполага, че заявките се формулират непосредствено на входния език за описание на онтологиите. Потребителят формулира заявката си посредством диалог, управляван от системата (system-driven dialogue). Той може да избира от подходящи менюта компонентите на заявката си и връзките между тях.

Така се осигурява синтактична коректност и еднозначност на интерпретацията на потребителската заявка. Все пак за правилното формулиране на заявката е необходимо да се знае поне какви концепти са включени в онтологията и какви атрибути имат те. Затова всички системи за представяне и използване на онтологични знания предоставят на своите потребители средства за визуализиране на наличните онтологии и навигация по дадена онтология.

Проект SHOE – спецификация на онтоологиите и инструментариум

Проектът SHOE (Simple HTML Ontology Extensions) е ориентиран към решаване на проблема за добавянето на семантична информация към Web страници и съотнасянето ѝ с онтологии на съответни предметни области. Предполага се, че като използват тази информация, системите за търсене ще могат да осигуряват по-релевантни отговори на заявките на потребителите, отколкото е възможно това сега на базата на използване на машините за търсене, функциониращи в Интернет.

За поддръжка на процесите на аотиране в рамките на проекта SHOE се разработва специален набор от инструментални средства (suite of tools) на основата на специфичен език за Интернет-съвместимо представяне на знанията, чието име носи целият проект.

Понастоящем в проекта SHOE са обособени следните основни направления на изследванията:

- разработка на множество повторно използваеми онтологии (reusable ontologies) за концептите, които са най-често срещани във Web ресурсите;
- създаване на средства за проектиране на онтологии – анотатори на знания (knowledge annotators), които биха опростили този процес.

Концепцията на езика за представяне на знанията в SHOE е подобна на тази в (КА)² и отново е свързана с разширяването на HTML със специални тагове (команди). Тук обаче се предлага “пълномащабно” разширяване на HTML. За целта SHOE въвежда в стандарта на HTML множество нови тагове за спецификация на онтологии: ONTOLOGY, USE-ONTOLOGY, CATEGORY, DEF-CATEGORY, RELATION, DEF-RELATION, DEF-INFERENCE, DEF-CONSTANT, DEF-TYPE, ARG, DEF-ARG и др. За аотиране на HTML документи се използват част от изброените тагове и някои нови, например INSTANCE.

За определеност в рамките на спецификациите на езика SHOE се предполага, че онтологията се представя под формата на **is-a** йерархия на класове/категории, множество от атомарни отношения между категориите и множество от правила за извод под формата на прости клаузи на Хорн.

Декларациите на онтологии се задават в тялото на HTML документа и не могат да се припокриват с други тагове на HTML. В един документ могат да бъдат определени няколко онтологии, но такива определения не могат да се припокриват или да се влагат. Общата схема на дефиниране на онтология е следната:

```
<ONTOLOGY ID="идентификатор-на-онтологията"  
  VERSION="версия"  
  [BACKWARD-COMPATIBLE-WITH="списък-на-версии"]  
  [DESCRIPTION="текст"]  
  [DECLARATORS="списък-от-декларации-на-  
    екземпляри"]>
```

декларация-на-онтологията

```
</ONTOLOGY>
```

За посочване на това, че дадена онтология разширява друга, вече съществуваща, се използва специален таг:

```
<USE-ONTOLOGY ID="идентификатор-на онтология"  
  VERSION="версия" PREFIX="префикс" [URL="URL"]>
```

Вътре в определението на онтологията могат да бъдат специфицирани нови категории. За целта се използва специален таг от вида

```
<DEF-CATEGORY NAME="име-на-категорията"  
[ISA="списък-от-родителски-категории"]  
[DESCRIPTION="текст"] [SHORT="текст"]>
```

Аналогичен подход се използва и при дефинирането на отношения:

```
<DEF-RELATION NAME="име-на-отношението"  
  [DESCRIPTION="текст"] [SHORT="текст"]>
```

списък-от-аргументи

```
</DEF-RELATION>
```

Един от най-важните компоненти на дефиницията на една онтология са правилата за извод. В SHOЕ тези правила по същество са подобни на клаузите на Хорн, но се различават от тях по форма:

```
<DEF-INFERENCE [DESCRIPTION="текст"]>  
  <INF-IF> тяло </INF-IF>  
  <INF-THEN> глава </INF-THEN>  
</DEF-INFERENCE>
```

МЕТОДОЛОГИЯ ЗА СЪЗДАВАНЕ НА ОНТОЛОГИИ

Работата на разработчиците на онтологии е затруднена изключително много от обстоятелството, че към момента на практика липсват утвърдени общи и верифицирани методологии, които да определят какви “процедури” трябва да се изпълнят в процеса на създаването на една онтология. Понастоящем съществуват само няколко предметно-независими методологии, ориентирани към построяването на онтологии.

Тези методологии се основават на следните принципи на проектирането и реализацията на онтологии, предложени от Грубер:

- ***Ясноста (Clarity)*** – онтологията трябва ефективно да предава смисъла на въведените термини. Определенията трябва да бъдат обективни, въпреки че мотивацията на въвеждането на термините може да се определя от ситуацията или от изисквания за изчислителна ефективност. За обективизация на определенията трябва да се използва ясен и фиксиран формализъм, при това е целесъобразно определенията да се задават под формата на логически аксиоми.

- ***Съгласуваност (Coherence)*** – въведените определения трябва да са логически непротиворечиви, а всички твърдения, които се извеждат в онтологията, не трябва да противоречат на аксиомите.
- ***Разширяемост (Extendibility)*** – онтологията трябва да бъде проектирана така, че да осигурява използването на разделяеми речници на термините, допускащи възможност за монотонно разширяване и/или специализация без необходимост от ревизия на вече съществуващи понятия.

- *Минимално влияние на кодирането (Minimal encoding bias)* – концептуализацията, стояща в основата на създаваната онтология, трябва да бъде специфицирана на нивото на представянето, а не на нивото на символното кодиране. Този принцип е свързан с обстоятелството, че агентите, които поделят онтологията, могат да бъдат реализирани в различни системи за представяне на знания.

- ***Минимална онтологична обвързаност (Minimal ontological commitment)*** – онтологията трябва да съдържа само най-съществените предположения за моделирания свят, за да оставя свобода за разширяване и специализация. От тук следва, че онтологиите се основават на “слаби” теории, тъй като целта на тяхното създаване и използване се състои преди всичко в това, те да “говорят” за предметната област (за разлика от базите от знания, които могат и дори трябва да съдържат знания, необходими за решаване на задачи и/или отговори на въпроси).

Тук ще обсъдим методологията и “жизнения цикъл” на създаването на онтологии, използвайки като пример подхода METHONTOLOGY, при който са реализирани принципите на Грубер и също така е разработено подходящо програмно обкръжение за спецификация на онтологии, наречено ODE (Ontology Design Environment).

В рамките на този подход са обособени следните процедури (етапи) от “жизнения цикъл” на създаването на една онтология: *управление на проекта, същинска разработка и поддръжка на разработката.*

Процедурите по управлението на проекта включват *планиране, контрол и гаранция на качеството*. *Планирането* определя какви задачи трябва да бъдат изпълнени, как ще се организират те, какви ресурси ще са необходими за тяхното изпълнение. *Контролът* гарантира, че планираните задачи са изпълнени именно така, както се е предполагало. *Гаранцията на качеството* е необходима, за да се получи увереност, че компонентите и продуктът като цяло са на необходимото равнище.

Същинската разработка включва *спецификация, концептуализация, формализация и реализация.* *Спецификацията* определя целите на създаването на онтологията, нейните потенциални приложения и потребители. *Концептуализацията* осигурява структурирането на предметните знания под формата на значим експлицитен модел. *Формализацията* трансформира концептуалния модел във формален или “изчислителен”. Накрая в процеса на *реализацията* изчислителният модел се програмира (кодира) на съответния език за представяне на знания.

Процедурите по поддръжката включват действия, които се изпълняват едновременно с разработката и без които онтологията не може да бъде построена. Тук се включват процедури по *придобиване на знания, оценка, интеграция, документиране и управление на конфигурациите*. *Придобиването на знания* представлява акумулиране на знания в (от) дадена предметна област. *Оценката* е конструиране на технически решения с цел оценяване на онтологията, съответното програмно осигуряване и документацията както в процеса на изпълнението на всяка от фазите, така и между фазите.

Интеграцията е необходима, когато се изгражда нова онтология с използване на вече съществуващи такива. *Документирането* дава детайлна, понятна и изчерпателна информация за всяка фаза и за продукта като цяло. *Управлението на конфигурациите* е необходимо за архивирането на всички версии на документацията, програмното осигуряване и кода на онтологията, а също и за целите на контрола на измененията.

Процесът на създаване на една онтология се декомпозира на серия от подпроцеси, свързани със създаването на междинни представяния. При това изпълнението на отделните подпроцеси не е последователно, а се определя от пълнотата и точността на вече натрупаните знания. Все пак, както показва опитът, най-напред се построява речникът на термините (Glossary of Terms), след това – класификационните дървета на концептите (Concept Classification Trees) и диаграмите на бинарните отношения (Binary Relations Diagrams). Едва след това се построяват останалите междинни представяния.

Както вече беше посочено, най-напред се построява речникът на термините, включващ всички термини (концептите, техните екземпляри, атрибутите, действията и т.н.), които са важни за предметната област, и техните естествено-езикови описания.

Когато речникът на термините достигне “съществен” обем, се построяват класификационните дървета на концептите. Като правило при това се използват отношения от типа Subclass-Of и други подобни таксономически отношения. По такъв начин се идентифицират основните таксономии в предметната област, а всяка таксономия съгласно разглежданата методология в крайна сметка определя съответна онтология.

Следващата стъпка е построяването “Ad hoc” на диаграмите на бинарните отношения. Целта на тяхното създаване е фиксирането на отношенията между концептите на една или няколко онтологии. Ще отбележим, че по-нататък тези диаграми могат да послужат за основа за интегрирането на различни онтологии.

След построяването на представянията, фиксирани по-горе, за всяко класификационно дърво на концепти се построяват:

- Речник на концептите (Concept Dictionary), който съдържа всички концепти от предметната област, екземплярите на такива концепти, атрибутите на екземплярите, отношенията, чиито източници са съответните концепти, а също синонимите и акронимите на концептите.
- Таблица на бинарните отношения (Table of Binary Relations) за всяко “Ad hoc” отношение, изходният концепт на което се съдържа в класификационното дърво. За всяко отношение се фиксират неговото име, имената на концепта – източник и целевия концепт, “обратното” отношение и т.н.

- Таблица на атрибутите на екземпляра (Instance Attribute Table) за всеки екземпляр от речника на концептите. Основни характеристики тук са: име на атрибута, тип на стойността, мерна единица, точност, диапазон, стойност по подразбиране, атрибути, които могат да бъдат изведени с използване на текущо описвания, формула или правило за извод на атрибута и др.
- Таблица на атрибутите на класа (Class Attribute Table) за всеки клас от речника на концептите с аналогични характеристики.

- Таблица на логическите аксиоми (Logical Axioms Table), в която се включват определения на концептите чрез винаги верни за тях логически изрази. Определението на всяка аксиома включва нейното име, описание на естествен език, концепта, за който се отнася аксиомата, атрибутите, които се използват в аксиомата, логически израз, описващ формално аксиомата, и др.
- Таблица на константите (Constants Table), в която за всяка константа се включват името ѝ, описание на естествен език, типът на стойността ѝ, самата ѝ стойност, мерната единица и др.

- Таблица на формулата (Formula Table) за всяка формула, включена в таблицата на атрибутите на екземпляра. Всяка таблица от този вид трябва да специфицира формулата като такава: нейното име, естествено-езиково описание, математическия израз, който я определя, атрибута, изводим с нейна помощ, ограниченията, при които може да се използва тази формула, и др.

- Класификационни дървета на атрибутите (Attribute Classification Trees), които показват графически съответните атрибути и константи, приложими при извода на стойността на атрибута, съответен на корена, и формулите, приложими за целта. По същество тези дървета се използват за проверка на това, че всички атрибути, участващи във формулата, имат описания и не е пропуснат нито един атрибут.
- Таблица на екземплярите (Instance Table) за всеки “входен” елемент в речника на концептите. Тук се специфицират името на екземпляра, неговите атрибути и техните стойности.