

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: *онтоологиите 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.

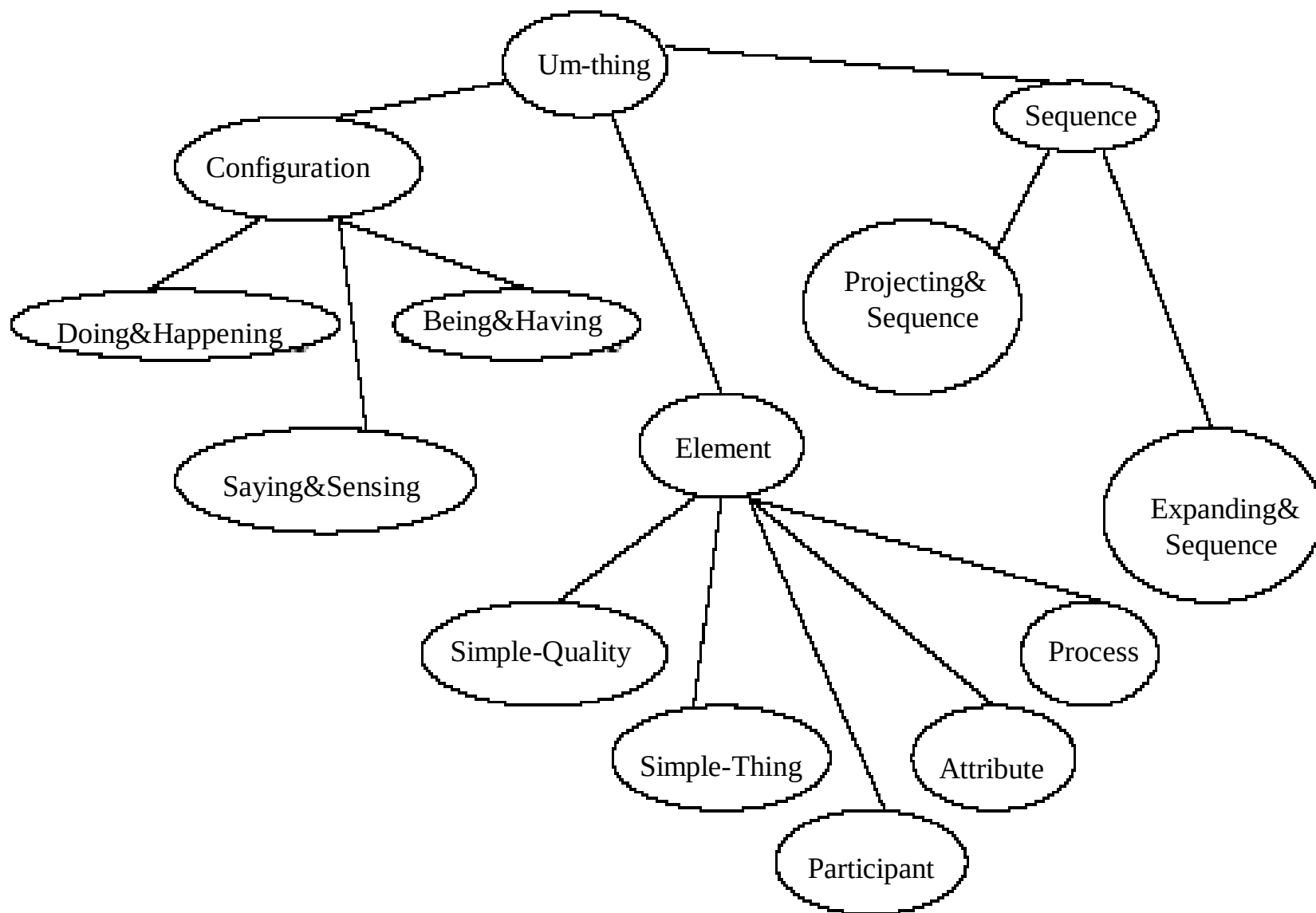
С други думи, предлага се класификация, според която онтологите се разделят на: *общи онтологии; онтологии, ориентирани към конкретна област; онтологии, ориентирани към конкретни задачи (приложни онтологии).*

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 Сусогр 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 СуsL.

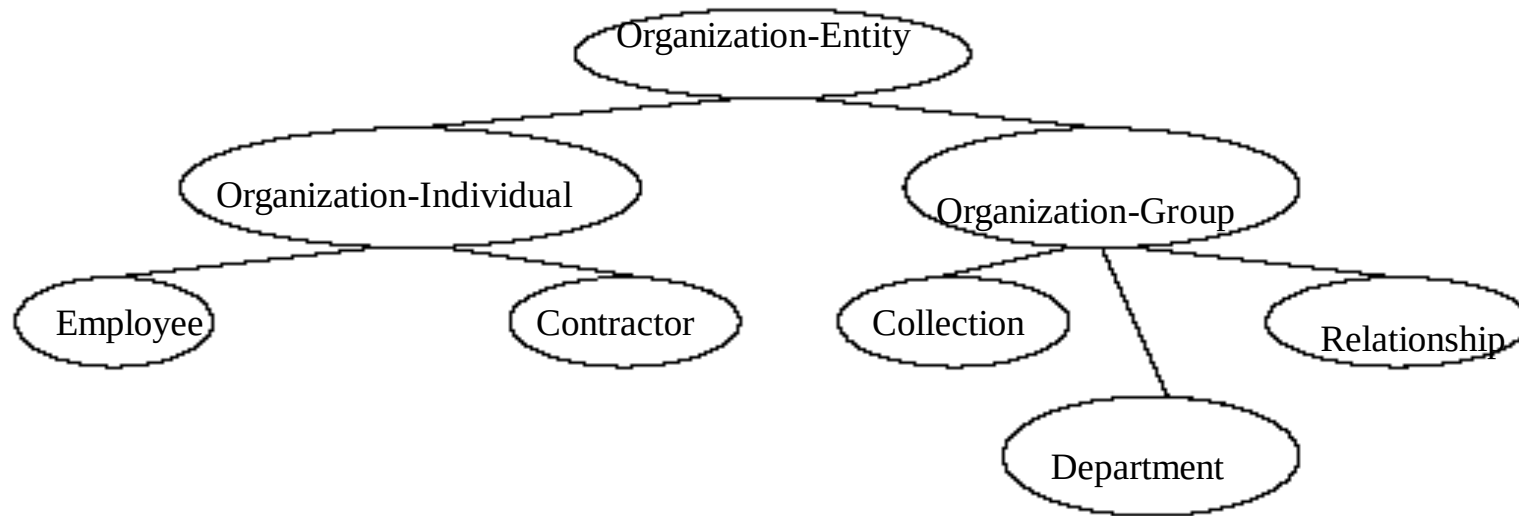
Друг пример за обща онтология е онтологията на системата Generalized Upper Model, ориентирана към поддръжка на процесите на обработка на естествен език: английски, немски и италиански. Нивото на абстракция на тази онтология се намира между лексическите и концептуалните знания, което се определя от изискванията 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 достатъчно добро 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 *степенна на детайлност, използвана за характеризиране 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 онтоологиите 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).

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 онтоологиите 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.