

МОДЕЛИРАНЕ НА ПРОЦЕСА НА СЪЗДАВАНЕ НА ОНТОЛОГИИ

Даниела Орозова, Евдокия Сотирова
orozova@bfu.bg, ensotirova@gmail.com

Abstract: *A theoretical generalized net (GN) model is proposed that describes process of ontology's development. The model is universal and independent of problem's area. It represents the process of ontology's development, based on life-cycle methodology. An example taken into consideration is the METHONTOLOGY approach, using Gruber principles.*

Key words: *Generalized nets, Ontology, Concept Dictionary, Concept Classification Trees, Binary Relations Diagrams, Intelligent database system.*

1. Методология за създаване на онтологии

Онтологите днес са изключително важни при изграждането на семантичната глобална мрежа (семантичен уеб), която има за цел да направи възможно откриването на ресурси по смисловото им съдържание, а не просто чрез търсене по ключови думи. Това се реализира чрез добавяне на семантична информация към мрежовите ресурси, наричана семантична анотация (semantic markup). За да бъде разбираема за хората и машините тази семантична информация трябва да бъде описана чрез общоприети концептуализации. За тази цел се използват онтологите [8]. Факт е, че онтологите могат да реферират други онтологии. Това в крайна сметка ще доведе до създаването на малко на брой независими от конкретната област (общи) онтологии и много на брой конкретни онтологии, които описват конкретните понятия и релации на дадена предметна област¹.

В момента на практика липсват утвърдени общи и верифицирани методологии, които да определят какви „процедури“ трябва да се изпълняват 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 [3]:

- **Яснота (Clarity)** – онтологията трябва ефективно да предава смисъла на въведените термини. Определенията трябва да бъдат обективни, като се използва ясен и фиксиран формализъм, Добре е 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.

¹ Изследването е свързано с работата по проект Д12/21.02.2013, финансиран от фонд „Научни изследвания“ при Бургаски свободен университет;

- **Разширяемост (Extendibility)** – онтологията трябва да бъде проектирана така, че да осигурява използването на разделяеми речници на термините, допускащи възможност за монотонно разширяване или специализация.

- **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 (Minimalontological commitment)** – онтологията трябва да съдържа само най-съществените предположения за моделирания свят, за да оставя свобода за разширяване и специализация. Целта на създаване и използване на онтологии се състои основно в това да „говорят“ за предметната област (за разлика от базите от знания, които съдържат знания, необходими за решаване на задачи 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) [4]. 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*.

1.1. Процедурите по управлението на проекта включват: *планиране, контрол и гаранция на качеството*.

- *Планирането* определя какви задачи трябва да бъдат изпълнени, как ще се организират, 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.

1.2. 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“.

- При процеса на *реализацията* изчислителният модел се програмира (кодира) на съответния език за представяне на знания.

1.3. 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, документирание и управление на конфигурациите*.

- *Придобиването на знания* представлява акумулиране на знания от дадена предметна област.

- *Оценката* е конструиране на технически решения с цел оценяване на онтологията, съответното програмно осигуряване и документацията в процеса на изпълнението на всяка от 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.

2. Модел на процеса на създаване на онтологии, следвайки методологията на „жизнения цикъл”

Методологията специфицира основните съставни части на онтоологиите: понятия (концепти), релации, функции, аксиоми, екземпляри, нефункционални свойства, импортирани онтологии и др. Основните елементи на всяка онтология са: понятия, екземпляри, релации и аксиоми:

- *понятията* се дефинират чрез: име, връзка с други понятия в йерархията от понятия и дефиниции на атрибути на понятието, като се разрешава наследяване на дефинициите на наследените атрибути;
- дефиницията на *екземпляр* се задава чрез име на екземпляра, името на понятието, към което принадлежи този екземпляр и описание на стойностите, които са свързани с атрибутите на екземпляра;
- релацията описва връзката между множеството от зададени параметри. Позволява се дефиниране на релации с произволен брой параметри, като за всеки параметър се указва типа му. Параметрите на релацията са строго подредени.
- аксиомите дефинират логически изрази, свързани с понятията.

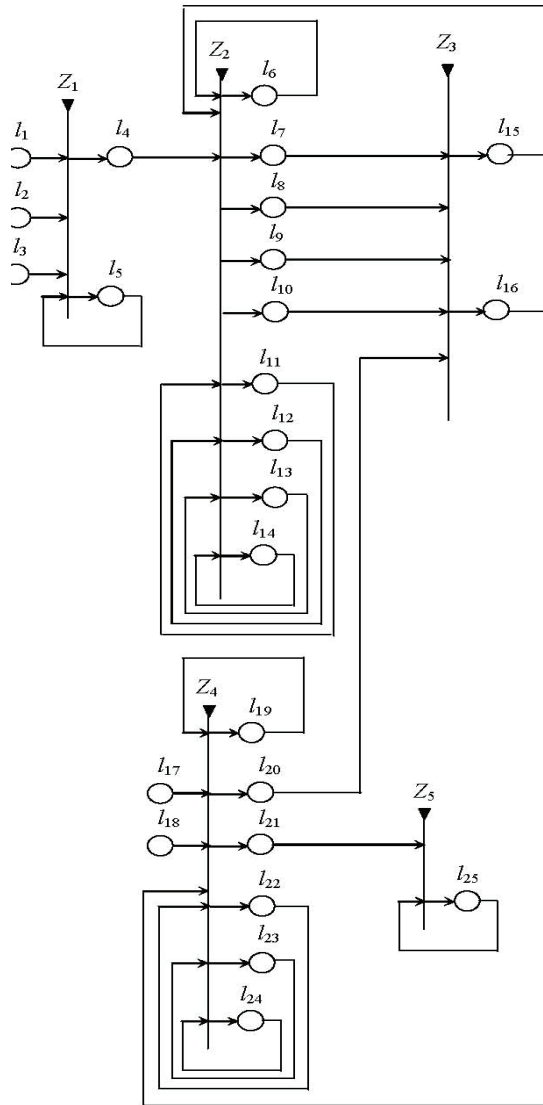
Създавайки една онтология трябва да бъде въведена информацията за всяко едно понятие и неговите атрибути, за всяка една релация и нейните параметри, за всеки един екземпляр и неговите стойности на атрибутите. Могат да бъдат дефинирани и логически изрази в аксиоми.

Обобщено-мрежовият модел [1,2] на процесите на взаимодействие на отделните компоненти при изграждане на предметно-ориентирана онтология е представен на Фигура 1.

Моделът, представящ методологията на изграждане на предметно-ориентирана онтология съдържа 5 прехода и 25 позиции.

Преходът Z_1 представя процеса на управление на проекта. През входните му позиции l_1 , l_2 и l_3 в мрежата постъпват ядра с начални характеристики съответно:

- „начални условия за планиране” в позиция l_1 ,
- „индикатори за контрол на планираните задачи” в позиция l_2 ,
- и „критерии за качество” в позиция l_3 .



Фигура 1. Обобщено-мрежови модел на процесите на изграждане на предметно-ориентирана онтологи

Преходът Z_1 има вида:

$$Z_1 = \langle \{ l_1, l_2, l_3, l_5 \}, \{ l_4, l_5 \}, r_1, \vee(\wedge(l_1, l_2, l_3), l_5) \rangle,$$

	l_4	l_5
l_1	false	true
$r_1 = l_2$	false	true
l_3	false	true
l_5	$w_{5,4}$	$w_{5,5}$

$w_{5,4}$ = „Изготвен е план за създаване на онтология”,

$w_{5,5} = \neg w_{5,4}$.

Ядрата, постъпващи в позиция l_5 не получават нови характеристики, докато ядрата, постъпващи в позиция l_4 получават характеристика:

„План за създаване на онтология”.

Преходът Z_2 представя процеса на същинска разработка на проекта. В позиция l_6 през цялото време на работата на обобщеномрежовия модел стои α -ядро с характеристика:

„Текущо състояние на същинската разработка на проекта”.

Цялата информация постъпваща в позиция l_6 ще се слива към текущата характеристика α -ядрото и ще се установява като текуща негова характеристика. В определени моменти от време това ядро ще генерира нови ядра, които ще постъпят в позиции l_7, l_8, l_9 или l_{10} .

Преходът Z_2 има вида:

$$Z_2 = \langle \{ l_4, l_6, l_{11}, l_{12}, l_{13}, l_{14}, l_{15} \}, \{ l_6, l_7, l_8, l_9, l_{10}, l_{11}, l_{12}, l_{13}, l_{14} \}, r_2, \vee(l_4, l_6, l_{11}, l_{12}, l_{13}, l_{14}, l_{15}) \rangle,$$

	l_6	l_7	l_8	l_9	l_{10}	l_{11}	l_{12}	l_{13}	l_{14}
l_4	true	false	false	false	false	false	false	false	false
l_6	true	$w_{6,7}$	$w_{6,8}$	$w_{6,9}$	$w_{6,10}$	$w_{6,11}$	$w_{6,12}$	$w_{6,13}$	$w_{6,14}$
$r_2 = l_{11}$	false	false	false	false	$w_{11,10}$	$w_{11,11}$	false	false	false
l_{12}	false	false	false	$w_{12,9}$	false	false	$w_{12,12}$	false	false
l_{13}	false	false	$w_{13,8}$	false	false	false	false	$w_{13,13}$	false
l_{14}	false	$w_{14,7}$	false	false	false	false	false	false	$w_{14,14}$
l_{15}	true	false	false	false	false	false	false	false	false

$w_{6,7}$ = „Изготвена е спецификация на модела”,

$w_{6,8}$ = „Изготвен е експлоатационен модел”,

$w_{6,9}$ = „Изготвен е изчислителен модел”,

$w_{6,10} =$ „Изготвен е кодиран модел”,
 $w_{11,10} =$ „Направена е реализация на модела”,
 $w_{11,11} = \neg w_{11,10}$,
 $w_{12,9} =$ „Направена е формализация на модела”,
 $w_{12,12} = \neg w_{12,9}$,
 $w_{13,8} =$ „Направена е концептуализация на модела”,
 $w_{13,13} = \neg w_{13,8}$,
 $w_{14,7} =$ „Изготвена е спецификация с целите на онтологията, потенциални приложения и потребители”,
 $w_{14,14} = \neg w_{14,7}$.

Ядрата, постъпващи в позиции l_7 , l_8 , l_9 и l_{10} получават характеристики съответно:

„Реализацията на онтологията” в позиция l_7 ,
 „Формален модел на онтологията” в позиция l_8 ,
 „Експлоатационен модел на онтологията” в позиция l_9 ,
 и „Спецификация на онтологията” в позиция l_{10} .

Ядрата, постъпващи и циклещи в позиции l_{11} , l_{12} , l_{13} и l_{14} имат характеристики:

„Текущо състояние на реализацията на онтологията” в позиция l_{11} ,
 „Текущо състояние на реализация на формален модел на онтологията”
 в позиция l_{12} ,
 „Текущо състояние на реализацията на експлоатационен модел на онтологията”
 в позиция l_{13} ,
 и „Текущо състояние на спецификацията на онтологията” в позиция l_{14} .

Преходът Z_3 има вида:

$$Z_3 = \langle \{ l_7, l_8, l_9, l_{10}, l_{20} \}, \{ l_{15}, l_{16} \}, r_3, \vee (l_7, l_8, l_9, l_{10}, l_{20}) \rangle,$$

	l_{15}	l_{16}
l_7	$w_{7,15}$	$w_{7,16}$
l_8	$w_{8,15}$	$w_{8,16}$
l_9	$w_{9,15}$	$w_{9,16}$
l_{10}	$w_{10,15}$	$w_{10,16}$
l_{20}	$w_{20,15}$	$w_{20,16}$

$w_{7,15} = w_{8,15} = w_{9,15} = w_{10,15} = w_{20,15} =$ „Има информация относно същинската разработка”.

$w_{7,16} = w_{8,16} = w_{9,16} = w_{10,16} = w_{20,16} =$ „Има информация относно процедурите по поддръжка”.

Ядрата, постъпващи в позиции l_{15} и l_{16} получават характеристики съответно:

„Информация относно същинската разработка” в позиция l_{15} ,
 и „Информация относно процедурите по поддръжка” в позиция l_{16} .

Преходът Z_4 представя процедурите по поддръжка на проекта. През входните му позиции l_{17} и l_{18} в мрежата постъпват ядра с начални характеристики съответно:

„нови знания за предметната област” в позиция l_{17} ,
и „критерии за оценка” в позиция l_{18} .

В позиция l_{19} през цялото време на работата на обобщеномрежовия модел стои β -ядро с характеристика:

„Текущо състояние на поддръжката на проекта”.

Цялата информация постъпваща в позиция l_{19} ще се слива към текущата характеристика β -ядрото и ще се установява като текуща негова характеристика. В определени моменти от време това ядро ще генерира нови ядра, които ще постъпят в позиции l_{20} или l_{21} .

Преходът Z_4 има вида:

$$Z_4 = \langle \{ l_{16}, l_{17}, l_{18}, l_{19}, l_{22}, l_{23}, l_{24} \}, \{ l_{19}, l_{20}, l_{21}, l_{22}, l_{23}, l_{24} \}, r_4, \vee(\wedge(l_{17}, l_{18}), l_{16}, l_{19}, l_{22}, l_{23}, l_{24}) \rangle,$$

където

	l_{19}	l_{20}	l_{21}	l_{22}	l_{23}	l_{24}
l_{16}	true	false	false	false	false	false
l_{17}	true	false	false	false	false	false
l_{18}	true	false	false	false	false	false
l_{19}	true	$w_{19,20}$	$w_{19,21}$	$w_{19,22}$	$w_{19,23}$	$w_{19,24}$
l_{22}	$w_{22,19}$	false	false	$w_{22,22}$	false	false
l_{23}	$w_{23,19}$	false	false	false	$w_{23,23}$	false
l_{24}	$w_{24,19}$	false	false	false	false	$w_{24,24}$

$w_{19,20}$ = „Има информация от процедурите по поддръжка”,

$w_{19,21}$ = „Необходимо е управление на конфигурациите”,

$w_{19,22}$ = „Необходима е интеграция на новата онтология”,

$w_{19,23}$ = „Необходимо е документиране на текущото състояние на създаването на онтологията”,

$w_{19,24}$ = „Необходимо е оценяване на текущото състояние на създаването на онтологията”,

$w_{22,19}$ = „Извършена е интеграция на онтологията”,

$w_{22,22}$ = $\neg w_{22,19}$,

$w_{23,19}$ = „Извършено е документиране на текущото състояние на създаването на онтологията”,

$w_{23,23}$ = $\neg w_{23,19}$,

$w_{24,19}$ = „Направена е оценка на текущото състояние на създаването на онтологията”,

$w_{24,24}$ = $\neg w_{24,19}$.

Ядрата, постъпващи в позиции l_{20} и l_{21} получават характеристики съответно:

„Информация, получена от процедурите по поддръжка” в позиция l_{20} ,
и „Има информация за архива на документиране на онтологията” в позиция l_{21} .

Ядрата, постъпващи и циклещи в позиции l_{22} , l_{23} и l_{24} имат характеристики:

„Текущо състояние на интеграцията на онтологията” в позиция l_{22} ,
„Текущо състояние на документацията на онтологията” в позиция l_{23} ,
и „Текущо състояние на конструирани технически решения за онагледяване на онтологията, програмно осигуряване и документация” в позиция l_{24} .

Преходът Z_5 има вида:

$$Z_5 = \langle \{ l_{21}, l_{25} \}, \{ l_{25} \}, r_5, \vee(l_{21}, l_{25}) \rangle,$$

$$r_5 = \begin{array}{c|c} & l_{25} \\ \hline l_{21} & true \\ l_{25} & true \end{array},$$

В позиция l_{25} през цялото време на работата на обобщеномрежовия модел стои γ -ядро с характеристика:

„Архив на документацията на онтологията”.

3. Приложения

Създаденият обобщено-мрежови модел показва връзките, начина на взаимодействие и процесите при изграждане на една онтология. Моделът може да бъде детайлизиран чрез прилагане на йерархичен оператор H_3 , който заменя преход или позиция от мрежата с нова мрежа, описваща по-детайлно съответен процес. Такъв модел намира приложение при създаване, изследване и оптимизиране на специализиран софтуер за работа с онтологии.

Литература:

1. Atanassov, K., Generalized Nets, World Scientific, Singapore, 1991.
2. Atanassov, K. On Generalized Nets Theory. Prof. M. Drinov Academic Publ. House, Sofia, 2007.
3. Gruber, T. R., A translation approach to portable ontology specification. Knowledge Acquisition, 5(2), pp. 199-220, (1993).
4. Grüninger M. and M.S. Fox, „Methodology for the Design and Evaluation of Ontologies,” Proc. Int'l Joint Conf. AI Workshop on Basic Ontological Issues in Knowledge Sharing, 1995.
5. Jecheva V., D. Orozova, Jecheva V., D. Orozova, Ontology-Based Electronic Test Result Evaluation, Third International Conference of Software, Services and Semantic Technologies S3T, Springer, 213-214, (2011).
6. McGuinness D. L., F. van Harmelen. „OWL Web Ontology Language Overview.” W3C Recommendation 10: 2004-2003.
7. Snae C., M. Brueckner, Ontology-Driven E-Learning System Based on Roles and Activities for Thai Learning Environment, Interdisciplinary Journal of Knowledge and Learning Objects, Vol.3, pp. 1-18, (2007).
8. Todorova M. Grid framework for e-learning services, Fourth International Conference Information Systems & Grid Technologies, May 28 – 29, 2010, Sofia, Bulgaria (pp 153-163).