

# ПРЕДМЕТНА ОНТОЛОГИЯ НА КОМПЮТЪРНИТЕ МРЕЖИ

Боряна Делийска

**Резюме:** С настъплението на Семантичната мрежа онтолозиите от теоретични изследвания все повече се превръщат в технология и ключов фактор за изграждането на интелигентни приложения. Те са многократно използвани бази знания в дадена област, основани на контролиран речник, таксономия и тезаурус за областта.

В статията се разглеждат видовете онтологии и връзките между тях, както и формалното описание на елементите им. Въз основа на известни методи са създадени таксономия и тезаурус на предметна онтология за компютърни мрежи. Онтологията е формализирана на език OWL в средата на редактор Protégé.

**Ключови думи:** предметна онтология, онтологично инженерство, таксономия, тезаурус, компютърна мрежа

## DOMAIN ONTOLOGY OF COMPUTER NETWORKS

Boriana Deliiska

**Abstract:** With the advent of the Semantic Web ontologies from theoretical studies are becoming a technology and key issue for intelligent applications building. They are reusable knowledge bases in given area, developed on controlled vocabulary, taxonomy and thesaurus in the area.

In the article common classification, relationships and elements of ontologies are given. On the base of certain methods a thesaurus and taxonomy of domain ontology of computer networks are developed. The ontology is presented by OWL language i Protégé editor environment.

**Keywords:** domain ontology, ontology engineering, taxonomy, thesaurus, computer network

### 1. Въведение

Онтолозиите дефинират семантиката на битието и са основен фактор за натрупване и придобиване на знания в съвременното информационно общество. В близкото минало те бяха чисто теоретични разработки, но в последните години придобиха широко практическо приложение – от комуникациите между хора и компютърни системи до автоматично извеждане и многократно използване на знания [10]. Нарастващото значение на онтолозиите в информационните системи доведе до възникването на нова дисциплина *технология на онтологии* или *онтологично инженерство* (ontology engineering), която се занимава с разработването и използването на онтологии.

В литературата по изкуствен интелект (ИИ) *онтология* е термин за означаване на формално представени знания въз основа на *концептуализация*. Концептуализацията е обобщение на характеристики и описание на

множество от обекти и понятия, знания и връзки между тях.

Структурно една онтология се състои от базови елементи и допълнителни (advanced) елементи. Базовите елементи са: *класове* (наричани още концепции, фрейми), *слотове* (или свойства, роли, атрибути), *фасети* (или ограничения на слотовете) и *екземпляри* (или инстанции, индивидуали), а допълнителните – *аксиоми* и *външни връзки* към други онтологии и тезауруси. Елементите се свързват с *релации*, които са два основни вида – йерерхични и нейерархични (асоциативни).

Базовите елементи са задължителни за изграждането на една онтология.

Изследвайки най-известните формализми за дефиниране на структурата на онтология, удачен за целите на настоящата теза е този на Bloehdorn S. et al. [1]:

$$(1) \quad O := (C, \leq C, R, \leq R, \sigma, I; iC, iR)$$

където:

$C$ ,  $R$  и  $I$  са непресичащи се (дизюнктивни) множества, чиито елементи се наричат класове (концепции), релации и екземпляри, съответно;

$\leq C$  е частично подреждане на  $C$ , наречено йерархия на класовете или таксономия;

функция  $\sigma : R \rightarrow C^+$ , е сигнатура (signature), описваща речника на онтологията;

$\leq R$  е частично подреждане на  $R$ , наречено йерархия на релациите (relation hierarchy);

$iC : C \rightarrow P(I)$  е функция, наречена дефиниране на екземпляри на класовете (concept instantiation);

$iR : R \rightarrow P(I^+)$  е функция, наречена дефиниране на екземпляри на релациите.

От друга страна, отчитайки не само синтаксиса, но и семантиката на една онтология, е удачен формализмът на Calvanese et al. [3]:

$$(2) \quad F = \{\Sigma, LC, LT, LA, Sem\},$$

където:

$\Sigma$ , LC, LT, LA са съответно азбука, концептуален език, TBox език и ABox език, дефиниращи синтаксиса на онтолозиите в  $F$ ;

Sem е семантична спецификация, описваща семантиката на онтолозиите, дефинирани в  $F$ .

ABox (Assertional Box) е множеството от твърдения за екземплярите в една онтология, а Tbox (Terminological Box) е множеството от декларации, дефиниращи класовете и ролите в логически език.

Според предмета на концептуализация [11] онтолозиите са: общи, представителни, предметни, приложни и онтолози на задачи. Връзките между видовете онтолози са изследвани от Guarino [7]. Схемата, предложена от него, тук е допълнена с включването на представителна онтология и релациите ѝ, която дефинира основните понятия за формализиране на всяка друга онтология (фиг. 1).



фиг. 1. Връзки между типове онтологии

Общите онтолози (common, general, core, generic or meta-ontologies) са минимални онтолози, съдържащи единствено общите за

много области класове, напр. състояние, събитие, процес, дейност, компонент и т.н.

Онтолозиите за представяне на знания или *представителните онтолози* (knowledge representation ontologies, representation ontologies) съдържат примитивите, необходими за формализиране на знание въобще. Те обясняват концептуализациите, залежали в основата на формализмите за знания, като осигуряват представителна рамка, без претенции за конкретно приложение. *Представителните онтолози* съдържат примитиви, използвани от предметните и общите онтолози.

*Приложните онтолози* (application ontologies) обхващат необходимото знание за моделиране на конкретно приложение. Обикновено приложните онтолози са смес от класове, взети от предметни и общи онтолози. Те не са многократно използвани. Примери за приложни онтолози са онтолози на учебни специалности, онтолози на информационни системи в дадена предметна област и т.н.

*Онтолозиите на задачи и методи* (task or method ontologies) са свързани с решаването на специфични задачи (зависими или независими от предметната област). Такива са онтолози за проектиране на сграда, парк, лаборатория или онтология за събиране на данни и т.н.

Няма общоприети методики относно създаването на онтолози, но всички известни методики [6] включват следните основни етапи:

- концептуализация, включително съставяне на таксономия и тезаурус;
- формализиране;
- верификация и публикуване;

*Предметните онтолози* (domain ontologies) представляват формални модели на определени области или пространства от знания (в някоя наука, общност, практика и т.н.) и съдържат понятията от тези области. Една област (domain)  $D$  може да се представи като двойка  $(\Delta D, \Phi)$  от непразно множество  $\Delta D$  от понятията (класове, индивидуали) в областта и множеството от имена на предикати  $\Phi$  [3]. Предикатите са релации или функции (изчисляващи стойност true или false), свързващи понятията.

Основно изискване към предметните онтолози е да бъдат многократно използвани в коанкретната област. В света на информационните системи предметната онтология е софтуер (или код на формален език), поръчан от определен клиент или приложна програма в характерен контекст [9].

Целта на настоящата работа е изграждане на предметна онтология на компютърните мрежи. Предназначението на тази онтология е

използването ѝ като база за развитие на приложни онтологии на задачи в областта на компютърните мрежи, както и за целите на обучението.

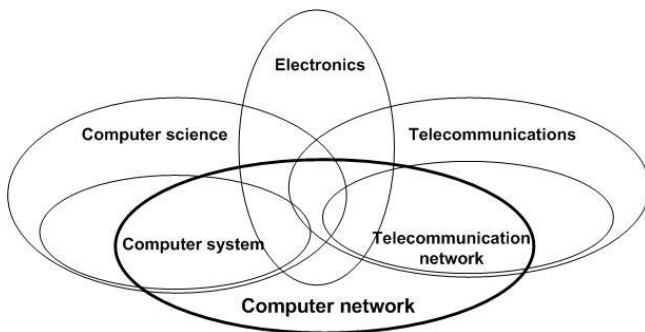
От прегледа на литературните източници бяха установени онтологии за мрежи от различен вид (транспортни, биологични, социални, семантични и др), но не бяха открити публикувани онтологии за компютърни мрежи.

## 2. Изграждане на таксономия и тезаурус на компютърните мрежи

*Таксономията* представлява йерархична класификация (йерархия от тип суперклас-субклас) на концепциите (класовете) в тази дадена област.

### 2.1. Изграждане на таксономия

Първа стъпка е установяване на мястото и връзките на предметната област с други сродни научни и/или приложни области. В случая на фиг.2 е показана обща схема на връзките на теорията и практиката на компютърните мрежи с компютърната наука, електрониката и телекомуникациите.



фиг.2. Връзка на предмета на компютърните мрежи с други предмети

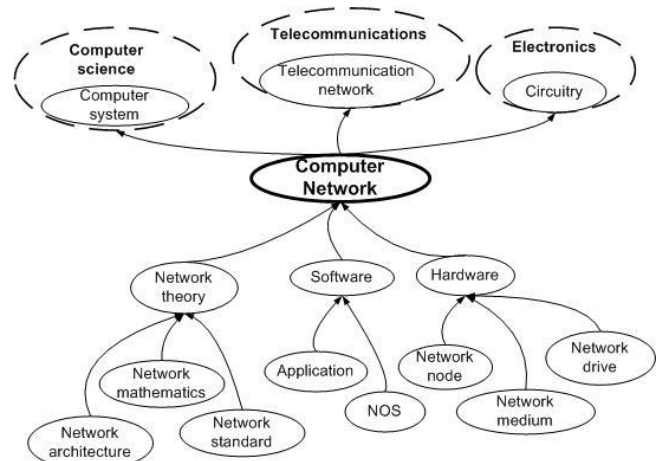
На втората стъпка се съставя на *контролиран речник* на термините в предметната област. Контролираният речник представлява подреден набор от препоръчани (стандартизирани) термини в предметна област с утвърдени дефиниции. Като такъв в случая се използва web-базираният речник по компютърни мрежи и web дизайн [4].

Третата стъпка е изграждане на схемата на таксономията на компютърните мрежи по метода "отгоре-надолу" за горните йерархични нива, а за долните – "отдолу-нагоре". Извадка от горните нива на тази таксономия са показани на фиг.3.

Най-старшият клас (top term, ТТ) *Computer network* (Компютърна мрежа), като научно-приложна дисциплина, включва теоретична и приложна част, в случая – подкласове *Network theory* (Теория на мрежите), *Software* (Софтуер) и *Hardware* (Хардуер).

По-надолу, на клас *Network theory* са подчинени класовете *Network mathematics* (Мрежова

математика), *Network architecture* (Мрежова архитектура) и *Network standard* (Мрежов стандарт). Клас *Software* има два подкласа – *NOS* (Network Operating System) и *Application* (Приложен софтуер). Клас *Hardware* включва класове *Network node* (Мрежов възел), *Network medium* (Мрежова среда) и *Network drive* (Мрежово устройство), и т.н.



фиг.3. Извадка от таксономия на компютърни мрежи

Релациите в една таксономия са йерархични и са два основни вида – таксономични (от тип *подклас-суперклас* или *is-a/has-a*) и мереологични (от тип *част-цяло* или *is-part-of/has-part*).

От друга страна, релациите с външни онтологии, са асоциативни (от вида *изисква/се-изисква-от* или *requires/is-required-by*). В структурата на настоящата онтология са включени такива релации с онтологии за телекомуникации, компютърна наука и електроника. В този смисъл може да се каже, че компютърните мрежи са вид телекомуникационни мрежи и компютърни системи, изградени на електронни схеми. На фиг.3 са показани само релациите на най-старшият клас *Computer Network* с класове *Computer system*, *Telecommunication network* и *Circuitry* от тези онтологии. В действителност, и останалите класове от онтологията може да имат асоциативни връзки с тези и други външни онтологии.

### 2.2. Проектиране и създаване на тезаурус

Съществуват различни дефиниции за тезаурус. Съгласно стандарта ANSI/NISO Z39.19-2005, *тезаурус* е контролиран речник, аранжиран в известен порядък и структуриран, така че различните релации между термините са ясно показани и обозначени със стандартизирани индикатори.

От друга страна, тезаурусът е база данни или списък от семантично ортогонални актуални

ключове за търсене (Wikipedia (<http://en.wikipedia.org/wiki/Thesaurus>)).

В областта на изкуствения интелект често тезаурусът се идентифицира с онтология.

Във всички случаи създаването на тезаурус за онтология е свързано с проектиране на база данни, включваща:

- класовете на таксономия на онтология - терминологията на тезауруса се съдържа в контролирания речник и е допълнена с термини от таксономията;
- слотовете на класовете – напр., слотове на най-старшия клас *Computer network* са: обхват на мрежата (локална, глобална, градска), скорост на обмен, нива на йерархия, конфигурация и т.н.;
- фасетите на слотовете, включително кардиналността (брой стойности на слот), типа на стойностите, областта и обхвата им;
- екземплярите на класовете – напр., екземпляри на клас NOS са LINUX, Windows NT, UNIX и др.;

В тезауруса се включват и нейерархичните релации, непрепоръчителните термини (синоними, близки синоними и лексикални варианти), както и етикети на възли.

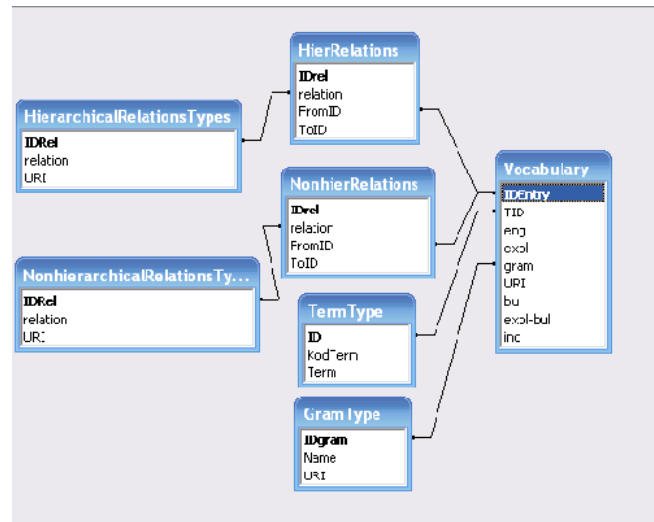
**Забележка:** Етикет на възел (node label) е термин, който не се индексира в тезауруса, но служи за структуриране на йерархия [van Assem et al 2004]. Например, терминът *Network classification* (Класификация на мрежи) е такъв етикет, който притежава под-етикети *Network by range* (Мрежи според обхвата) и *Network by topology* (Мрежи според топологията) *Network by access type* (Мрежи според типа на достъпа) и т.н..

Тезаурусът на компютърните мрежи е двуезичен – англо-български с водещ английски език, като може да се конвертира и в българо-английски.

**Забележка:** Броят и видът на езиците не влияе върху логическите конструкции в софтуера на тезауруса и лесно се променя с добавяне или отстраняване на съответната таблица в базата данни.

Схемата на базата данни на тезауруса (фиг.4) е вариант на публикуваната в [5]. Главната таблица Vocabulary съдържа елементите на онтологията на два езика – английски и български с дефиниции. Тя е свързана с таблиците-класификатори на йерархичните релации (HierRelations), нейерархичните релации (NonhierRelations), граматическите типове (GramType) и стандартните индикатори (TermType) на термините.

Понастоящем тезаурусът съдържа около 1300 лексикални единици и е достъпен от адрес [http://www.e-forestinformatics.org/projects/Ontologies/Ontology\\_networks/OntologyNetworks\\_bg.html](http://www.e-forestinformatics.org/projects/Ontologies/Ontology_networks/OntologyNetworks_bg.html).



фиг.4 Схемата на базата данни на тезауруса

Чрез динамични HTML страници, свързани с ASP скриптове [4] за достъп до съдържанието на базата данни, потребителят може да го разгледа и да потърси справки.

### 3. Формализация на предметна онтология за компютърни мрежи

Предметната онтология на компютърните мрежи е формализирана на езика OWL чрез конвертиране на тезауруса с VBA приложение на алгоритъма, изложен в [5].

Генерираната онтология се проверява синтактично чрез системи за верификация, напр., W3C Validator (<http://www.w3.org/RDF/Validator>) или Racer Pro [<http://www.racer-systems.com/>]. След отстраняване на евентуални грешки, се отваря в редактора Protégé [7].

В средата на редактора най-напред се добавя слой на външните връзки на онтологията с другите онтологии, посочени в таксономията ѝ (фиг.3). За целта предварително се търсят готови онтологии. От изследването на онтологичните сайтове бе установено, че част от необходимите онтологии за настоящата онтология се намират в библиотеката на семантичната търсачка Swoogle [<http://swoogle.umbc.edu/>].

В заглавният блок с метаданни на онтологията, освен префиксите на стандартните пространства на имената (namespaces) в XML(S) [2], RDF(S) и Dublin Core, се добавят и тези на намерените в семантичната търсачка Swoogle [<http://swoogle.umbc.edu/>] OWL онтологии, свързани с компютърните мрежи, а именно:

```
...xmlns comm="http://www.tssg.org/public/ontologies/omg/uml/2004/Communications.owl#"
```

Към тях се добавят пространствата на имената на онтологията за компютърни науки и онтологията за електроника, публикувани от

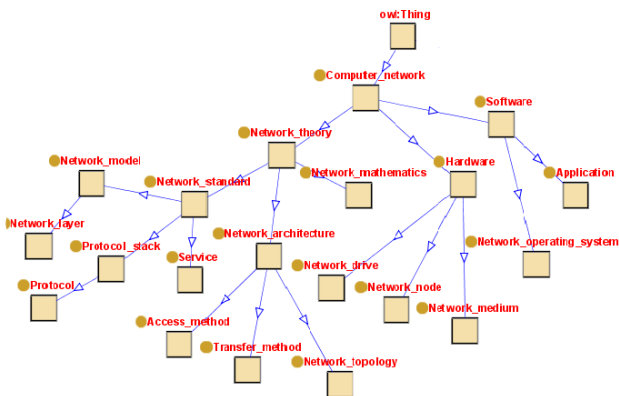
автора на адрес [http:// www.e-forestinformatics.org/projects/Ontologies/](http://www.e-forestinformatics.org/projects/Ontologies/).

```
xmlns:com="http:// www.e-forestinformatics.org/
projects/Ontologies/
OntologyComputerScience.owl#"
xmlns:com="http:// www.e-forestinformatics.org/
projects/Ontologies/ OntologyElectronics.owl#"

```

Текстът на онтологията по компютърни мрежи е публикуван на адрес: [http://www.e-forestinformatics.org/projects/Ontologies/Ontology\\_networks/OntologyNetworks.html](http://www.e-forestinformatics.org/projects/Ontologies/Ontology_networks/OntologyNetworks.html).

Графът на класификационното дърво на онтологията е създаден е с Jambalaya – една от графичните надстройки (plug-in) на редактора Protégé [8]. Извадка от него е е показана на фиг. 5.



фиг.5. Граф на горните нива на предметна онтология за компютърни мрежи

#### 4. Заключение и насоки за бъдеща работа

Настоящата предметна онтология за компютърни мрежи е не само опит за обобщаване на терминологията и релациите в областта, но и за използването ѝ в обучението по едноименната дисциплина. Извличането на приложна онтология на учебна програма от предметна онтология чрез онтологични операции за изображение и трансформация е разгледано в [4].

Обект на бъдещо изследване е сравнението на съществуващи учебни програми с генерирана от съответната приложна онтология структура на курс.

#### REFERENCES

- [1] Bloehdorn S., Haase P., Sure Y., Voelker J. *D.6.6.1 Report on the integration of ML, HLT and OM, EU-IST Integrated Project (IP)*, IST-2003-506826 SEKT, www document 2005, [www.sekt-project.org/rd/deliverables/wp06/sekt-d-6-6-1-Int.%20ML,%20HLT,%20OM.pdf](http://www.sekt-project.org/rd/deliverables/wp06/sekt-d-6-6-1-Int.%20ML,%20HLT,%20OM.pdf)
- [2] Bray T., Hollander D., Layman A., Tobin R. (eds) *Namespaces in XML*, www document 2006, <http://www.w3.org/TR/REC-xml-names/>
- [3] Calvanese et al. *Common Framework for Representing Ontologies*, 2006 Deliverable TONES-D08, <http://www.tonesproject.org/D08.CommonFramework.pdf>
- [4] Deliiska B. *Web-based system of dictionaries in area of computer science*, Proceedings of Conference of Computer Science'2004, Sofia, 06-08.12.2004
- [5] Deliiska B. *Thesaurus and domain ontology of geoinformatics*, J. Transaction in GIS, ISSN: 1361-1682, vol 12 issue 4 (11), 2007
- [6] Fernández-López M. (ed) *Ontoweb: A survey on methodologies for developing, maintaining, integrating, evaluating and reengineering ontologies*, IST Project IST-2000-29243, www document 2002, <http://ontoweb.org/About/Deliverables/D1.4-v1.0.pdf>
- [7] Guarino N. *Formal Ontology and Information Systems*, in Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998, Amsterdam, IOS Press, pp. 3-15
- [8] Horridge M., Knublauch H., Rector A., Stevens R., Wroe C. *A Practical Guide To Building OWL Ontologies Using The Protégé-OWL, Plugin and CO-ODE Tools Edition 1.0*, www document 2004, <http://www.co-ode.org/resources/tutorials/ProtegeOWLTutorial.pdf>
- [9] Smith B. *Ontology and Information Systems*, Technical Report, www document 2001, [http://ontology.buffalo.edu/ontology\(PIC\).pdf](http://ontology.buffalo.edu/ontology(PIC).pdf)
- [10] Uschold M., Gruninger M. *Ontologies: principles, methods and applications*, The Knowledge Engineering Review 1996, 11(2), pp. 93-136.
- [11] van Heijst G., Schreiber G., Wielinga B. *Using explicit ontologies for KBS development*, International Journal of Human-Computer Studies, 42(2/3):183-292, 1997