

Švietimo ir mokslo terminų žodynas ir ontologija

IRENA MARKIEVICZ, ERIKA RIMKUTĖ

Vytauto Didžiojo universitetas

ESMINIAI ŽODŽIAI: tekstynas, švietimo ir mokslo terminai, ontologija, žodynas

ĮVADAS

2010–2012 m. Vytauto Didžiojo universiteto Kompiuterinės lingvistikos centre vykdytas Lietuvos mokslo tarybos finansuotas projektas *Švietimo ir mokslo terminų automatinis identifikavimas (ŠIMTAI 2)*. Šio projekto pagrindinis tikslas – išbandyti tekstynų lingvistikos metodus, kad būtų galima automatiškai ar pusiau automatiškai atpažinti terminus, automatizuoti terminų apibrėžtims suformuluoti reikalingos informacijos atranką tekстыne ir taip palengvinti terminologų darbą. Svarbiausi atlikto tyrimo rezultatai, be metodikų, yra specialusis švietimo ir mokslo (toliau ŠM) tekstynas (jo dydis – 4 mln. žodžių), šio specialiojo tekstyno pagrindu parengtas *Švietimo ir mokslo terminų žodynas* ir *Švietimo ir mokslo terminų ontologija*.

Straipsnyje pristatomas ŠM terminų žodynas ir ŠM terminų ontologija. Rašoma apie žodyno struktūrą, jo skirtumus nuo kitų terminų žodynų. Daugiau dėmesio skiriama ŠM terminų ontologijai. Paaiškinama, kaip skirtingose mokslo šakose apibrėžiamos ontologijos, kam jos reikalingos, kokiais metodais kuriamos. Išsamiai aptariama ŠM terminų ontologija, pristatoma jos sandara, pateikiama pavyzdžių, paaiškinama, kaip šis šaltinis gali būti pritaikomas.

Abu darbai – tiek ŠM terminų žodynas, tiek ŠM terminų ontologija – parengti remiantis ta pačia medžiaga, t. y. specializuotu ŠM tekstynu, naudojant tekstynų ir kompiuterinės lingvistikos metodologiją.

Šiuo darbu norima pristatyti tekstynų ir kompiuterinės lingvistikos galimybes terminologijos srityje, siekiama atkreipti terminologų, terminų žodynų ar duomenų bazių sudarytojų dėmesį į tai, kad galima bent iš dalies automatizuoti terminų tvarkybą, kad įmanoma suderinti preskriptyvų ir deskriptyvų terminų analizės būdą.

ŠVIETIMO IR MOKSLO TERMINŲ ŽODYNO SUDARYMO ETAPAI

ŠM terminų žodyno terminai nustatyti specializuotame tekстыne taikant lingvistinius metodus (plačiau žr. Grigonytė et al. 2011; Rimkutė 2012). Žodyno sudarymas truko dvejus metus, t. y. kol buvo vykdomas įvade minėtas projektas, tokiais etapais:

- 1) sudarytas 4 mln. žodžių specializuotas ŠM tekstynas;
- 2) tekstynas automatiškai morfologiškai anotuotas;
- 3) sukurtas automatiškai terminus atrenkantis įrankis, pagrįstas lingvistinėmis taisyklėmis;
- 4) apibrėžti galimų terminų atrankos kriterijai;
- 5) nustatyta apie 11 tūkst. galimų terminų, sudarytų nuo vieno iki penkiolikos žodžių;
- 6) pritaikyti papildomi galimų terminų atrankos kriterijai¹;
- 7) nustatyta 780 ŠM terminų (1–5 žodžių);
- 8) parengta terminų apibrėžimo, remiantis tekstynais, metodika;
- 9) parengti lingvistiniai ištekliai, reikalingi terminų apibrėžtims nustatyti;
- 10) apibrėžti 300 pasirinktų grupių terminų;
- 11) pateikti apibrėžtų terminų atitikmenys anglų, vokiečių, prancūzų kalbomis.

Neabejotina, kad kiekvieno terminų žodyno sudarymas reikalauja daug specialistų darbo laiko sąnaudų, finansinių išteklių. Šiame straipsnyje pristatomas ŠM terminų žodynas, nors ir nedidelis – jame pateikti 224 terminai (plačiau žr. kitą skyrių), bet metodologiškai labai sudėtingas. Kaip matyti iš aprašytų žodyno sudarymo etapų, net ir turint specializuotą reikiamos srities tekstyną², buvo reikalingi specialūs įrankiai³ terminams automatiškai atpažinti ir nustatyti.

¹ Analizuoti ne ilgesni nei iš septynių žodžių sudaryti terminai; atsižvelgta ir į dažnumą: kuo trumpesnis terminas, tuo dažniau jis turi būti vartojamas, pvz., analizuoti tik tie vienažodžiai terminai, kurie pavartoti bent 20 kartų, o septyniažodžiai terminai turėjo būti pavartoti ne mažiau nei du kartus.

² Tekstyno sudarymas – vienas iš svarbiausių ir ilgiausių viso tyrimo etapų, nes norint sudaryti reprezentatyvų tekstyną, kuriame būtų galima nustatyti terminus, reikėjo atidžiai atrinkti tekstus, juos žanriškai subalansuoti, techniškai sutvarkyti (iš pdf rinkmenų perkelti į tekstines, pakeisti kodavimą ir kt.).

³ Šiame projekte buvo sukurtas lietuvių kalbai pritaikytas terminams automatiškai nustatyti skirtas įrankis, nes, kaip matyti iš ankstesnių eksperimentų (žr. Grigonytė et al. 2011), dėl kalbų struktūros skirtumų lietuvių kalbai beveik neįmanoma pritaikyti kitoms kalboms skirtų terminų atpažinimo įrankių. Šiame įrankyje naudotos lingvistinės taisyklės: ieškota tam tikros gramatinės struktūros terminų, pvz., būdvardis + daiktavardis, daiktavardžio kilmininkas + daiktavardis, dalyvis + daiktavardis ir kt.

Panašiais metodais terminus nustato latvių kalbos tyrėjai (žr. Gornostay et al. 2012; Krugļevskis et al. 2005). Pasak jų, terminų nustatymas naudojant lingvistinius metodus (morfosintaksinę analizę) yra tinkamesnis nei analitinėms kalboms tinkantys statistiniai metodai.

Sudėtinga buvo ir iš visų automatiškai nustatytų prasmingų ir taisyklingos gramatinės struktūros junginių atsirinkti ŠM terminus. Sunkumų kilo dėl to, kad vadinamųjų galimų terminų iš 4 mln. žodžių tekstyno nustatyta net 11 tūkst. Kitame tyrimo ir žodyno sudarymo etape buvo svarbu iš visų junginių nustatyti ŠM terminus (apie šią problemą rašoma Rimkutė 2012).

Atkreiptinas dėmesys į tai, kad pristatomas ŠM terminų žodynas sudarytas tekstyno pagrindu. Tai reiškia, kad buvo naudotas ne tradicinis, iki šiol Lietuvoje vis dar paplitęs terminologų kompetencija paremtas, kartais gana subjektyvus, preskriptyvusis terminų tvarkybos būdas, o šiuolaikinėje terminologijoje pranašesniu laikomas deskriptyvusis terminų analizės ir apibrėžimo būdas, pagrįstas tekstynų duomenimis. Žodynams sudaryti tekstynai labai reikalingi, nes juose galima nustatyti tipiškiausius, dažniausius terminų vartosenos modelius, kurie svarbūs reikšmei apibrėžti, aktualiems pavyzdžiams parinkti.

Dėl sudėtingos terminų apibrėžimo metodologijos nuspręsta išsamiai apibrėžti 300 terminų. Atsirinkti tam tikrų grupių terminai: terminai su žodžiu *akademini* (pvz.: *akademine laisvė*, *akademine etika*, *akademine varda*), su žodžiu *darbas* (pvz.: *diplominis darbas*, *namų darbas*, *savarankiškas darbas*), su žodžiu *išsilavinimas* (pvz.: *universitetinis išsilavinimas*, *bendrasis išsilavinimas*, *aukštasis išsilavinimas*), su žodžiu *švietimas* (pvz.: *švietimo teikėjas*, *suaugusiųjų švietimas*, *neformalusis švietimas*) ir t. t. Apibrėžti ir keli dažnai vartojami vienažodžiai terminai: *studentas*, *mokslininkas*, *universitetas* ir kt.

Kaip ir kituose terminų žodynuose, ŠM terminų žodyne prie daugumos terminų nurodyti atitikmenys anglų, vokiečių ir prancūzų kalbomis. Jie rinkti iš patvirtintų apibrėžtos srities šaltinių, dokumentų, specializuotų terminų duomenų bazių: *Anglų kalbos švietimo terminų žodyno (Glossary of Educational Terms)*, *Aiškinamojo su studijomis susijusių terminų žodyno*, *Der Europäische Qualifikationsrahmen für lebenslanges Lernen*, *EuroVoc* (daugialkalbio Europos Sąjungos tezauro), *Lietuvos Respublikos terminų banko* ir kt.

Daugiausia remtasi Didžiosios Britanijos, Prancūzijos ir Vokietijos švietimo terminija, tačiau daugelio vertimo atitikmenų reikšmė minėtų šalių sistemos kontekste nėra visiškai tokia pati kaip originalių terminų, vartojamų Lietuvos švietimo ir mokslo srityje. Žodyne pateikti terminų atitikmenys užsienio kalbomis yra taikytini Lietuvos Respublikos mokslo ir švietimo sistemai ir ne visada (arba iš dalies) gali būti taikomi kitos šalies mokslo ir švietimo sistemai apibūdinti.

ŠVIETIMO IR MOKSLO TERMINŲ ŽODYNO SANDARA

ŠM terminų žodynas yra laisvai prieinamas internete adresu <http://daukantas.vdu.lt/moksliniai-terminai/>. Čia pateikti 224 terminai, nors anksčiau skyriuje rašyta, kad apibrėžta 300 terminų. Apibrėžtų ir žodyne pateiktų terminų skaičiai skiriasi todėl, kad pastebėta, jog kai kurie terminai (pvz.: *kvalifikacijos lygis / lygmuo; mokslo / mokslinis laipsnis; mokslo / mokslinė produkcija; magistrantūros / magistro studijos; studento mokymosi / darbo krūvis; valstybės / valstybinis mokslo institutas*) vartojami kaip sinonimai, jie yra panašios struktūros, todėl tokie terminai sujungti.

ŠM terminų žodyne yra visiems terminų žodynams būtinos dalys, t. y. pats terminas, jo apibrėžtis. Kaip minėta, nurodyti atitikmenys anglų, vokiečių ir prancūzų kalbomis. Taip pat pateiktos dalykinės sritys (pvz., aukštojo mokslo valdymas, aukštojo mokslo politika, aukštojo mokslo administravimas), kai kurių terminų santrumpos.

Žodynas nuo kitų terminų žodynų ar duomenų bazių skiriasi tuo, kad dar nurodytos visos terminų kaitybines formos su dažnumais ir bendrasis termino dažnumas ŠM tekстыne. Prie visų terminų galima rasti bent du vartosenos pavyzdžius iš ŠM tekстыno (pavyzdžiai neredaguoti, nekeisti, tik kartais sutrumpinti), kairiuosius ir dešiniuosius kolokatus⁴, konkordansą⁵.

Ne visiems terminams apibrėžti pakako informacijos iš tekстыno. Prie tų terminų, kurių apibrėžtims suformuluoti informacijos buvo mažoka (pvz.: *mokymo teikėjas, studento krepšelis, bazinis išsilavinimas*), pažymėtos žvaigždutės.

Kadangi terminai išrinkti iš tekстыno, buvo žinoma, kaip dažnai jie pavartoti, todėl prie visų terminų nurodytas sutartinai žymimas dažnumas kitų terminų atžvilgiu:

- ★★★ rodoma prie dažniausių terminų, kurie švietimo ir mokslo tekстыne pavartoti iki 500 kartų;
- ★★☆ rodoma prie retesnių terminų, kurie tekстыne pavartoti 100–499 kartus;

⁴ Tai statistiškai (pagal dažnumą) nustatomi analizuojamo termino kontekstiniai partneriai, dažnai atsidurian-tys šalia termino iš kairės ir dešinės pusės.

⁵ Tai visų pasirinkto termino gramatinių formų vartojimo atvejai iš tekстыno, pateikti vienos eilutės kontekste (konkordanso eilutė dažniausiai apima daugiau negu vieną sakinį). Konkordansas paprastai analizuojamas skaitant jį pirmiausia vertikaliai, nes tik taip išryškėja tipiška termino vartoseną. Analizuojamas terminas pateiktas konkordanso eilutės viduryje, jis yra paryškintas. Analizuojamam terminui iš kairės ir iš dešinės nurodyta maždaug po 50 simbolių konkordanso ištraukos. Norint pamatyti ilgesnę konkordanso ištrauką (maždaug po 150 simbolių iš kairės ir dešinės), reikia paspausti konkordanso eilutės gale esantį pliuso ženklą.

★☆☆ rodoma prie terminų, kurie tekstyne pavartoti 10–99 kartus;

☆☆☆ rodoma prie retų terminų, kurie tekstyne pavartoti mažiau nei 10 kartų.

Toliau kaip ŠM terminų žodyno iliustracija pateiktas termino *bakalauro studijos* straipsnis.

BAKALAURO STUDIJS ★☆☆

Apibrėžtis

Aukštojo mokslo pirmosios pakopos studijos, teikiančios bendrą universitetinį išsilavinimą ir pasirinktos studijų krypties žinių pagrindus bei kvalifikacinį bakalauro laipsnį.

Pavyzdžiai

JAV universitetuose įprasta, kad gabesniems studentams sudaromos sąlygos sujungti magistrantūrą su doktorantūra, tiksliau – iškart po bakalauro studijų stoti į doktorantūrą (tai tam tikru laipsniu paaiškina, kodėl magistrantūros ir doktorantūros studijos yra sujungiamos vienu terminu – graduate studies, kurios vykdomos taip vadinamoje graduate school) ir įgyti PhD laipsnį po 8 metų studijų universitete.
--

Rekomenduojamas periodas kartu su bakalauro studijomis – 5 metai
--

Bakalauro studijos (I pakopa) paprastai trunka 4 metus (160 kreditų, dieninės studijos).
--

Bakalauro studijos, magistrantūra ir doktorantūra baigiamos atitinkamai bakalauro darbo, magistro darbo ir disertacijos gynimu.

Bakalauro studijos – tai pirmosios pakopos studijos, teikiančios bendrą universitetinį išsilavinimą ir pasirinktos studijų krypties žinių pagrindus.
--

Dažnumas

bakalauro studijos	21
bakalauro studijomis	6
bakalauro studijose	3
bakalauro studijas	7
bakalauro studijų	14
Bendrasis dažnumas:	51

Atitikmenys

Anglų k.: *bachelor's studies*

Vokiečių k.: *Bachelorstudium*

Prancūzų k.: *études de licence*

Dalykinė sritis

aukštojo mokslo administravimas

Kolokatai

Kairieji kolokatai		Dešinieji kolokatai
PO	bakalauro studijų	
SU	bakalauro studijomis	
TRUMPINTI	bakalauro studijas	
PROGRAMINĖ ĮRANGA	bakalauro studijoms	
	bakalauro studijos	I. PAKOPA
	bakalauro studijos	MAGISTRANTŪRA IR DOKTORANTŪRA

Konkordansas⁶

[1]	darbinę veiklą arba toliau tęsti ir baigti	bakalauro studijas.	Aukštosios mokyklos atstovai išplatino pranešimą,	+
[2]	šalyse, siūlantys tokias bakalauro studijas.	Bakalauro studijų	IV kurso rudens semestre, o magistrantūros studijų	+
[3]	teologijos studijų programa – 5 metai) ; į	bakalauro studijas	buvo priimti 1 870 studentai, o 2009 m. Universitetai	+
[4]	universitetai ryžosi iki 3 ir 3, 5 metų trumpinti	bakalauro studijas,	informuoja, kad Mokslo ir studijų įstatymas	+
[5]	universitetai ryžosi iki 3 ir 3, 5 metų trumpinti	bakalauro studijas.	JAV bakalaurai, siekdami pradėti magistrantūros	+
[6]	komisijų reikalavimus : reikia būti baigus	bakalauro studijas,	jei stojama į tyrimus orientuotą programą,	+
[7]	m. Universitetai ryžosi iki 3 metų trumpinti	bakalauro studijas.	Keli universitetai ryžosi iki 3 ir 3, 5 metų	+
[8]	kreditų (4, 5 metų) bakalauro studijų.	bakalauro studijas,	trunkančias 4 metus (katalikų teologijos studijų	+
[9]	neteikiančios studijos. Rekomenduojamas periodas kartu su	bakalauro studijomis	- 5 metai (3 + 2 ; 3, 5 + 1, 5 ; 4 + 1). Bakalauro	+
[10]	Paprašytas palyginti šias studijas su savo	bakalauro studijomis	JAV, užsienietis pastebėjo, kad čia daug lengviau.	+

⁶ Čia pateikta tik dešimt pirmųjų konkordanso eilučių. Interneto svetainėje paspaudus pliuso ženklą parodomas ilgesnis konkordansas.

Prie kai kurių terminų pateikti keli tūkstančiai konkordanso eilučių. Peržiūrėti tokį informacijos kiekį reikia daug laiko, ne visiems šie duomenys svarbūs. Norintiems daugiau sužinoti apie termino vartoseną, kontekstą, kuriame vartojamas pasirinktas terminas, galbūt sudaryti kitą apibrėžtį, išsirinkti reikalingų pavyzdžių, susidaryti objektyvią nuomonę konkordansas suteikia daug naudos.

ONTOLOGIJŲ SAMPRATA IR KŪRIMO PRIEŽASTYS

Šiame ir tolesniuose skyriuose rašoma apie ŠM terminų ontologiją, jos sudarymo principus, šios ir kitų ontologijų panaudojimo galimybes. Kad būtų aiškiau, pateikta keletas vizualizuotų ŠM terminų ontologijos klasifikacijų ir rezultatų.

Ontologijos – formalios žinių konceptualizavimo priemonės, apibendrinančios natūraliąja kalba aprašytas žinias į struktūrizuotą, hierarchinį susijusių sąvokų modelį. Dėl informacijos apdorojimo kokybės ir terminų daugiareikšmiškumo problemų kyla poreikis kurti struktūrizuotų terminų bazes, su kuriomis kompiuterinės programos galėtų apdoroti natūraliąją žmonių kalbą (angl. *natural language processing*) ir interpretuoti jos semantiką. Dėl šios priežasties ir kuriamos ontologijos.

Tam tikrą kalbą mokantis žmogus lengvai suvokia jam pateikiamo natūraliosios kalbos teksto prasmę, o kompiuterinė programa interpretuoja tą pačią informaciją kaip tekstinių simbolių ir su jais susijusių metažymų rinkinį. Ontologijos yra ta jungiamoji grandis, paverčianti natūraliąja kalba parašytą informaciją į koncepcinį žinių modelį, kurį lengvai interpretuoja informacijos surinkimo ir apdorojimo robotai.

Ontologijų kūrimas yra tarpdisciplininis procesas, apimantis filosofijos, lingvistikos ir informatikos mokslų sritis (plačiau žr. Grigonytė 2010). Filosofijoje ontologija suvokiama kaip disciplina, nagrinėjanti būti ir egzistenciją. Tokią ontologijos apibrėžtį papildė lingvistai; ontologija lingvistikoje – atskirų sąvokų ir jas apibūdinančių terminų tinklas. Informatikoje ontologija yra loginis ir formalus tikrovės vaizdas.

Didžiausias ir lengvai prieinamas natūraliosios kalbos tekstų šaltinis yra internetas, dėl to ontologijos dažnai suvokiamos kaip viena iš semantinio saityno ir semantinės paieškos priemonių. W3C organizacija, kurianti internetinių svetainių standartus, apibūdina ontologiją kaip terminų rinkinį, aprašantį tam tikros srities žinias: „ontologijas naudoja žmonės, duomenų bazės ir taikomosios programos, kurios siekia pasidalyti tam tikros srities

informacija <...>. Ontologijos susieja kompiuterinėms programoms pritaikytas tam tikros srities sąvokas ir ryšius tarp jų“ (W3C).

Švietimo ir mokslo terminų ontologija (toliau ŠM terminų ontologija) sudaryta siekiant formalizuoti ŠM terminų rinkinį, sukurtą vykdant įvadinę projektą, ir jį papildyti semantiniais ryšiais, kuriuos gali interpretuoti taikomosios programos. Pagrindiniai ontologijos kūrimo darbai: kokybiškai aprašyti pasirinktos srities terminus bei sąvokas ir formaliai specifiuoti jų reikšmes. Šie darbai svarbūs sprendžiant terminų daugiareikšmiškumo problemą ir kuriant ŠM srities struktūrinį terminų modelį. ŠM terminų ontologija sukurta ne tik kaip ŠM terminų taksonomija, bet ir kaip vienas iš galimų semantinės paieškos resursų.

ŠM terminų ontologija prieinama internete adresu <http://meta.vdu.lt/webprotege/>, vizualizuojama naudojant *WebProtege* įrankį, su kuriuo galima peržiūrėti hierarchinį klasių medį, jų atributus ir sąvokų esinių pavyzdžius. Vartotojas, atidarydamas ar paslėpdamas naujas korteles ir blokus, gali susidėlioti patogią grafinę sąsają. Rekomenduojama naudoti *Chrome* ar *Mozilla Firefox* naršyklės. Tam, kad ryšiai tarp terminų būtų geriau pavaizduoti, rekomenduojama naudoti *Protege 4.3* su *OntoGraf* papildiniu.

ŠVIETIMO IR MOKSLO TERMINŲ ONTOLOGIJOS KŪRIMO METODOLOGIJOS PASIRINKIMAS

Apibendrinta ontologijų kūrimo koncepcija apibūdinama šiuo procesų metamodeliu (plačiau žr. Cimiano 2006): $M=\{D, LA, T, S, C, TR\}$, kuriame *D* reiškia dokumentų surinkimą, *LA* – lingvistines žymas, *T* – terminų atranką, *S* – sinonimų nustatymą, *C* – terminų ir jų sinonimų susiejimą į sąvokas ir *TR* – sąvokų semantinių ryšių apibrėžimą.

Naudojant skirtingas ontologijų kūrimo metodologijas, taikomos įvairios ontologijų kūrimo strategijos: remiantis vienomis metodologijomis, daugiau dėmesio skiriama ontologijos gyvavimo ciklui ir prototipų kūrimui (*Cyc*, *TOVE*, *Methontology*, *On-To-Knowledge*), kitose labiau akcentuojamas ontologijų panaudojimas taikomosiose programose (KAKTUS). Vienais metodais ontologijos kuriamos iš sukauptų tekstynų (*TOVE*, *On-To-Knowledge*), kitais – iš didelių bendrinių žinių ontologijų (*SENSUS*, *Cyc*) (plačiau žr. Corcho et al. 2003).

Paprastčiausiai ontologijos kuriamos šiais etapais: parenkama žinių sritis, apdorojama informacija (apibrėžiamos sąvokos, terminai, apribojimai),

nurodomi ontologijos plėtimo metodai ir įvertinimo kriterijai, parengiama ontologijos dokumentacija (plačiau žr. Uschold et al. 1996). Sudarant ontologijas tokiu metodu, nenumatoma, kad bus kuriami ontologijos prototipai ar kad ontologijos, kaip papildomi informacijos šaltiniai, bus naudojamos. Taikant šią metodologiją, sukurtos kelios ūkinės veiklos ontologijos, bet nenumatyta, kaip jas panaudoti taikomuosiose programose.

On-To-Knowledge projekte pasiūlyta metodologija, daugiausia dėmesio skirianti žinių valdymui ir būsimosios ontologijos naudojimo informacinėse sistemose galimybėms (plačiau žr. Staab et al. 2001). Atlikus galimybių studiją, apibrėžti informacijos šaltiniai, potencialūs ontologijos vartotojai, panaudojimo būdai.

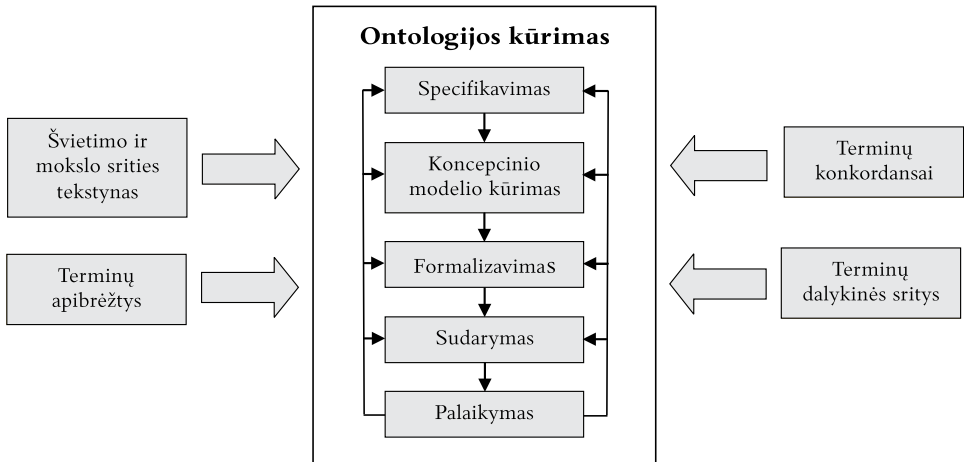
Tobulinant *On-To-Knowledge* ontologiją dalyvavo pasirinktos srities ekspertai, buvo apdorojama iš tekstų ir kitų žinių bazių sukaupta informacija. Ontologijos kokybė įvertinta tikrinant, ar ontologija atitinka reikalavimus ir pritaikymo galimybes. Visi šie ontologijos kūrimo ir tobulinimo etapai gali būti laikomi rekomendacijomis kitoms ontologijoms sudaryti.

Pristatytose metodologijose daugiausia dėmesio skiriama ontologijų kūrimo iš tekstinių informacijos šaltinių metodams. SENSUS ontologijos metodologija siūlo specialiosios srities ontologijos kūrimo procese naudoti bendrųjų žinių ontologijas (plačiau žr. Gomez-Perez et al. 2004). Metodologija numato terminų iš *OntoSaurus* žinių bazės grupavimą į sąvokų medį. Bandant susieti dvi sąvokų klases, ieškoma bendrinės sąvokos klasės (angl. *root class*).

Pasirenkant ŠM terminų ontologijos kūrimo metodologiją, atsižvelgta į turimus informacijos išteklius ir kalbos technologijų bei ontologijų kūrimo įrankių pritaikymo galimybes. Nuspręsta naudoti *Methontology* projekte sukauptą patirtį (plačiau žr. Mariano et al. 2004; Gomez-Perez et al. 2004; Corcho et al. 2003).

Methontology pritaikymas ŠM terminų ontologijai kurti ir ontologijos kūrimo ištekliams pavaizduoti 1 paveiksle. Specifikuojant nurodomas ontologijos kūrimo tikslas ir apimtys. Konceptinis ontologijos modelis kuriamas atsižvelgiant į specifikacijos metu išgrynintą sritį, vėliau jis formalizuojamas – paverčiamas formaliu ontologijos modeliu. Kūrimo proceso metu formalus modelis įgyvendinamas pagal pasirinktos ontologijų kūrimo kalbos reikalavimus. Net ir sukurta ontologija neretai būna keičiama, plečiamą ar atnaujinama.

1 pav. ŠM terminų ontologijos kūrimo procesai ir išteklių (remiantis *Methontology*)



Pritaikius *Methontology* modelį, sukurtas terminų glosarijus, sąvokų taksonomija, nustatyti ryšiai tarp atskirų ontologijos objektų, parengta pavyzdžių atributų, klasių, abstrakčių klasių bei jų atributų ir ryšių bazė, kurioje įvardijama, kokiais atvejais egzistuoja pasirinkta sąvoka, su kokiomis kitomis sąvokomis ji susijusi, taip pat aprašyti sąvokų esiniai.

Ontologijomis aprašomi skirtingi tos pačios specifinės srities dalykai: asmenys (pvz., *rektorius*, *studentas*), vietos (pvz., *universitetas*), procesai (pvz., *mokymas*, *dėstymas*), procesų rezultatai (pvz., *meno kūrinys*, *bakalauro darbas*) ir kt. Visi jie ontologijoje įvardijami kaip sąvokos (angl. *concept*), kurios gali egzistuoti kaip savarankiškas ontologijos informacijos vienetas arba gali būti susietos į sąvokų klases (angl. *class*).

Sąvokų taksonomija (angl. *concepts taxonomy*) – sisteminė sąvokų klasifikacija pagal pasirinktą koncepcinį ontologijos modelį – įprastai kuriamas iš terminų glosarijaus (angl. *glossary of terms*), kuris šiuo atveju suprantamas kaip formalus terminų ir jų apibrėžčių sąrašas (plačiau žr. Uschold et al. 2004).

Klasikinėje taksonomijoje klasė apibrėžiama kaip „panašių esinių grupė, besiskirianti nuo kitų individų grupių“ (plačiau žr. Mayr 1987). Kiekvieną ontologijos klasę apibūdina tik jai būdingų savybių (angl. *property*), ir jų apribojimų (angl. *restriction*) rinkinys, kurį perima visi tam tikros klasės esiniai (angl. *instance*, *individual*). Esinys yra konkretus pasirinktos sąvokos egzempliorius (pvz., *Zigmas Lydeka* yra sąvokos *rektorius* esinys).

ŠM terminų ontologijos terminai nustatyti specializuotame tekстыne naudojant lingvistinius metodus (plačiau žr. Grigonytė et al. 2011; Rimkutė 2012). Ontologijos sąvokos ir sąvokų klasės kurtos pagal mišrią iš apačios į viršų (angl. *down-top strategy*) ir iš viršaus į apačią (angl. *top-down strategy*) strategiją (plačiau žr. Uschold et al. 1996). Taikant iš apačios į viršų strategiją, pirmiausia identifikuojamos specifiškiausios sąvokos, vėliau jos keičiamos abstraktesnėmis. Taikant iš viršaus į apačią strategiją, pirmiausia apibrėžiamos abstrakčios sąvokos, vėliau jos tikslinamos. Taikant mišrią ontologijų sudarymo strategiją, pirmiausia nagrinėjamos pradinės sąvokos ir, atsižvelgiant į poreikius, jos arba tikslinamos, arba apibendrinamos.

Šiame ontologijos terminų taksonomijos kūrimo etape aktyviai dalyvavo ŠM srities ekspertas. Jis sugrupavo terminus pagal dalykines sritis ir sukūrė pirminį ontologijos koncepcinį modelį.

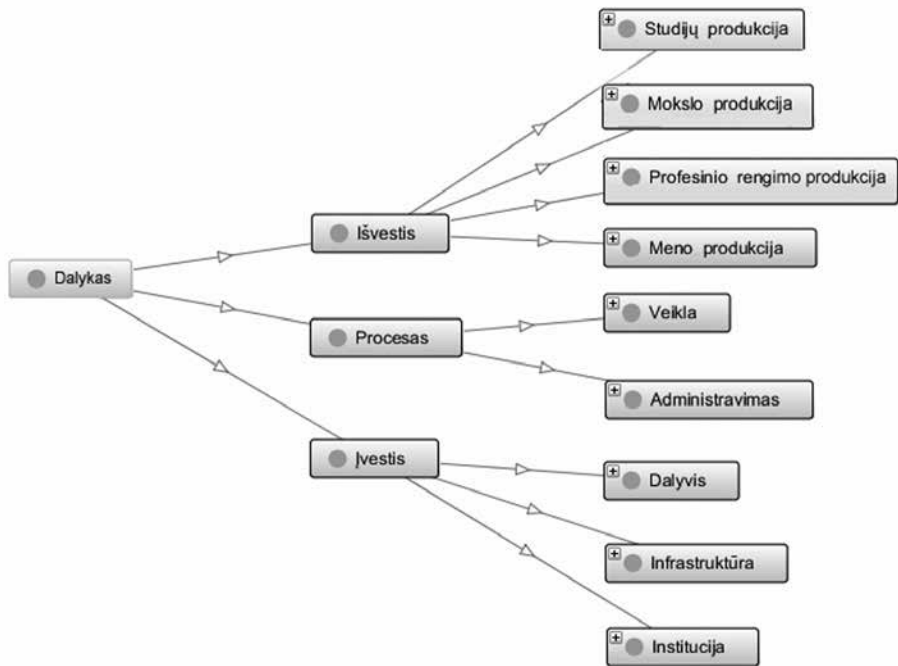
Kuriant sąvokų taksonomiją, rekomenduojama itin daug dėmesio skirti šiems etapams: sudaryti terminų sąrašą ir iš jo atrinkti reikalingus terminus, sukurti sąvokų modelį, parengti apibrėžtis ir įvertinti terminus (ISO 10241: 1992).

ŠVIETIMO IR MOKSLO TERMINŲ ONTOLOGIJOS STRUKTŪRA

Vienas iš ontologijų kūrimo proceso iššūkių – sukurti kokybišką duomenų struktūrą, kuri būtų universali ir pritaikoma skirtingoms informacijos apdorojimo sistemoms, suderinama su kitais žinių šaltiniais. Pagrindinis ontologijos elementas yra dalykas (angl. *thing*), abstrahuojantis visas ontologijos sąvokas. Šio elemento būtinumą paaikškina filosofinė ontologijos – kaip mokslo, nagrinėjančio būtį, – apibrėžtis. Ontologija formalizuoja visas tam tikros srities žinias, tad būtinas jos elementas yra abstrakti klasė, apibendrinanti visas specializuotas srities sąvokas.

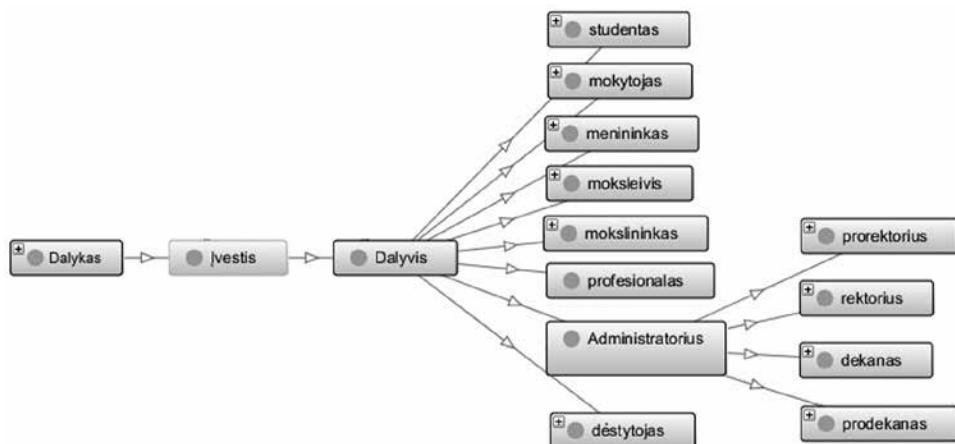
Sąvokos, terminai ir aksiomos yra pagrindiniai bet kurios ontologijos koncepciniai elementai. Svarbu skirti termino ir sąvokos reikšmes. LST ISO 1087-1 standarte (2005) *sąvoka* apibūdinama kaip „[p]ažinimo vienetą, kurį sudaro savitas požymių derinys“. Sąvoka nepriklauso nuo kalbos, jai įtakos gali turėti tik kontekstas. *Terminas* yra sąvokos identifikatorius, t. y. „[ž]odinis bendrosios sąvokos žymiklis specialioje dalykinėje srityje“ (LST ISO 1087-1: 2005). Aksiomomis apibūdinami specifinės srities apribojimai, kuriama ontologijos sąvokų hierarchinė struktūra. Formaliems struktūriniais ontologijos objektams priskiriamos sąvokos, sąvokų klasės, klasių ir sąvokų esiniai, klasių, sąvokų ir esinių savybės, visus ontologijos objektus siejantys ryšiai bei apribojimai.

2 pav. ŠM terminų ontologijos struktūra pagal verslo veiklų modelio sritis



Ekspertui rekomendavus, visi terminai sugrupuoti į tris abstrakčias *įvesties*, *proceso* ir *išvesties* klases (žr. 2 pav.), būdingas verslo veiklų modeliams (plačiau žr. ISO 9001: 2000). *Įvesties* klasėje nurodomi sistemos ar konkrečios veiklos reikalavimai, ištekliai ir duomenys. Dėl to šiai kategorijai priskiriamos *infrastruktūros*, *institucijos* ir *dalyvio* terminų grupės. Veiklos procesai skirstomi į esminius (klasė *veikla*), sukuriančius pridėtinę vertę, ir palaikančius procesus (klasė *administravimas*), kurie sustiprina pagrindinių procesų veiklą, užtikrina infrastruktūrą sukuriamiems procesams. *Išvestis* apibendrina proceso efektyvumą ir našumą – nurodo, kokius rezultatus galima buvo pasiekti, kokie buvo pasiekti ir kokie ištekliai buvo jiems panaudoti (*mokslo produkcijos*, *profesinio rengimo produkcijos*, *studijų produkcijos* ir *meno produkcijos* klasės). Kiekviena abstrakti dalykinės srities klasė gali turėti papildomų dalykinės srities poklasių (kaip matyti 3 paveiksle, klasės *dalyvis* poklasis yra *administratorius*, šis poklasis skirstomas dar smulkiau). Iš viso ŠM terminų ontologijoje yra 29 abstrakčiosios klasės. Pagrindinė jų paskirtis – kurti bendrąją ontologijos koncepcinį modelį pagal jau minėtą *iš viršaus į apačią* strategiją.

3 pav. *Dalyvio* klasės bendrosios sąvokos ir abstraktūs poklasiai ŠM terminų ontologijoje



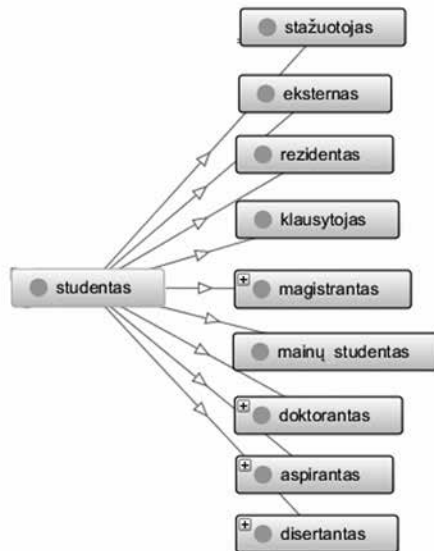
ŠM terminų ontologiją sudaro 465 bendrosios sąvokos⁷, kurių hierarchija sukurta pasitelkus apibrėžiamųjų kontekstų atpažinimo šablonus (plačiau žr. Bielinskienė et al. 2012). Apibrėžiamieji kontekstai iš dalies arba visiškai atitinka terminų apibrėžtis ir sudaromi iš hiperoniminių sąvokų. Pavyzdžiui, sąvoka *studentas* apibrėžiama kaip *asmuo, studijuojantis aukštojoje mokykloje pagal nuosekliųjų studijų programą*. O užsienio *studentas* yra *studentas, atvykęs į tam tikros šalies aukštąją mokyklą iš kitos šalies studijuoti visos nuosekliųjų studijų programos arba dalinių studijų*.

Mokslinių terminų ontologijoje įvardijami šie taksonomijos ryšiai:

- klasės poklasis (angl. *subclass-of*) – pvz., *magistrantas* yra klasės *studentas* poklasis;
- išskaidymas (angl. *disjoint-decomposition*) – pvz., *studijų produkcija* negali tuo pačiu metu būti ir *bakalauro darbas*, ir *disertacija*;
- visiškas išskaidymas (angl. *exhaustive-decomposition*) – pvz., klasei *dėstytojas* priklauso poklasiai *docentas*, *lektorius*, *profesorius* ir kt. Jei tam tikras objektas yra *dėstytojas*, jis turi priklausyti bent vienam iš išvardytų poklasių;
- padalijimas (angl. *partition*) – pvz., sąvoka *studijų semestras* apima sąvokas *pavasario semestras* ir *rudens semestras*. Jei kalbama apie *studijų semestrą*, jis turi priklausyti kuriam nors iš išvardytų poklasių.

⁷ Ontologijoje pateikti ir tie terminai, kurių nėra ŠM terminų žodyne, t. y. kurie nebuvo išsamiai apibrėžti.

4 pav. Sąvokos *studentas* taksonominė hierarchija ŠM terminų ontologijoje



Kadangi pagrindinis ontologijos kokybės kriterijus yra terminų viena-reikšmiškumas, kuriant ŠM terminų ontologiją daug dėmesio skirta, kad sąvokų savybių rinkinys būtų sudarytas kokybiškai. Klasių, sąvokų ir esinių savybių rinkiniui nustatyti tinka aprašomųjų kontekstų atpažinimo šablonai (plačiau žr. Bielinskienė et al. 2012). Aprašomieji kontekstai nusako analizuojamos sąvokos pobūdį ir jos ryšį su kitomis sistemos dalimis.

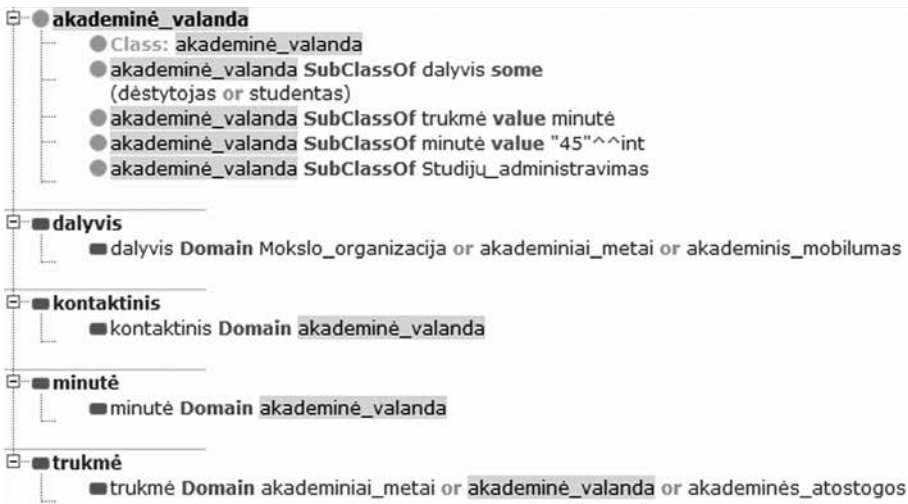
5 pav. Atributo *aktyvus* taikymo galimybės ŠM terminų ontologijoje⁸



Naudojant tą pačią savybę galima apibūdinti kelias skirtingas sąvokas. Pavyzdžiui, savybė *aktyvus* būdinga keturioms sąvokoms: *praktinis darbas*, *praktinis rengimas*, *praktinis užsiėmimas*, *pratybos* (žr. 5 pav.). Todėl svarbu akcentuoti, kad tik tam tikru unikaliu savybių deriniu galima kokybiškai apibūdinti sąvoką.

⁸ Naudojant *Protege 4.3* ontologijos kūrimo ir vizualizavimo įrankį, nebuvo galimybės vartoti lietuvių kalbą, todėl struktūriniai ontologijos elementai (pvz., *ObjectProperty*, *Domain*) pateikti anglų kalba.

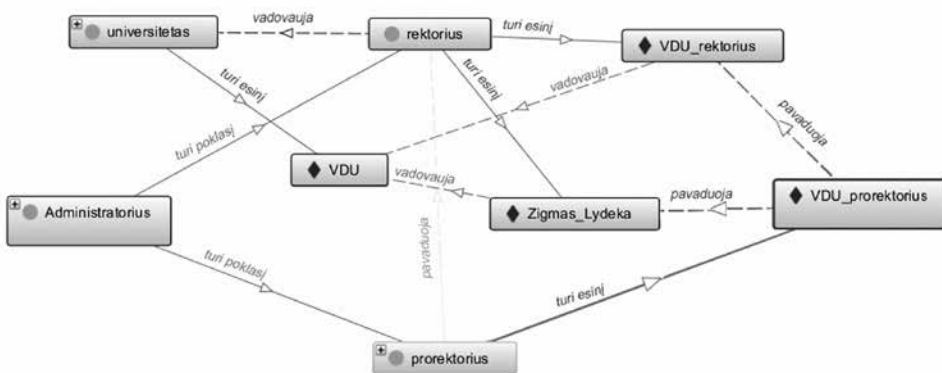
6 pav. Sąvokos *akademine valanda* savybių ir jų apribojimų rinkinys ŠM terminų ontologijoje



Kiekviena savybė gali turėti nurodytus jo apribojimus ar ryšius su kitomis sąvokomis ar sąvokų klasėmis. Pavyzdžiui, *akademine valanda* apibūdinama kaip *dėstytojų ir studentų kontaktinio darbo trukmės vienetą, lygus 45 minutėms*. Formali, ontologinė šios sąvokos apibrėžtis įvardija tokį savybių rinkinį: *akademine valanda* kaip *proceso* poklasis; *dalyviai* yra *studentai* ir *dėstytojas*; tai *kontaktnis* procesas, turintis *trukmę*, kuri apibrėžiama 45 minučių konstanta (žr. 6 pav.).

Klasių ir sąvokų savybes, jų apribojimus ir ryšius su kitomis savybėmis perima esiniai – konkretūs klasių ir sąvokų egzemplioriai, pasižymintys

7 pav. ŠM terminų ontologijos sąvokų ir esinių ryšių medis



tomis pačiomis savybėmis. Pavyzdžiui, jeigu ontologijoje nurodoma, kad *rektorius* vadovauja *universitetui*, tai šios sąvokos esinys – *Zigmas Lydeka* – vadovauja VDU, t. y. sąvokos *universitetas* esiniui (žr. 7 pav.).

Švietimo ir mokslo terminų ontologija sukurta OWL kalba su *Protege* 4.3 ontologijų kūrimo įrankiu, vizualizuota naudojant *WebProtege* įrankį ar *OntoGraf* papildinį⁹.

ŠVIETIMO IR MOKSLO TERMINŲ ONTOLOGIJOS ĮVERTINIMAS BEI PRITAIKYMAS

Šiuo metu ŠM terminų ontologijoje yra 29 abstrakčiosios sąvokų klasės ir 465 bendrosios sąvokos, 87 sąvokų, sąvokų klasių ir jų esinių savybės, 141 esinys, 296 ontologijos sąvokų klasių, sąvokų bei esinių ryšiai.

Formaliai ontologijos sudėtingumą galima įvertinti pasitelkiant aprašomosios logikos (angl. *description logic*) metrikas (plačiau žr. Corcho et al. 2003; Baader 2003). ŠM terminų ontologija įvertinta ALCHON(D) sudėtingumo lygmeniu, čia:

- *AL* – atributų kalba; ja galima susieti sąvokas, rašyti joms universalius apribojimus;
- *C* – nurodo, kad galima neigti sąvokų ryšius;
- *H* – nurodo, kad galima kurti vaidmenų hierarchiją;
- *O* – kiekį įvertinantys apribojimai (ryšiai: *tiksliai vienas, vienas iš daugelio*);
- *N* – kiekio apribojimai (ryšiai: *mažiau, daugiau*);
- (*D*) – duomenų tipų savybės.

Be ontologijos žodyno, taksonomijos ir semantinių ryšių vertinamos jos panaudojimo ir pritaikymo galimybės. Pasirinkta OWL ontologijų kūrimo kalba yra tinkama padaryti semantinėms išvadoms, ontologijai pildyti naujomis sąvokomis ar kelioms skirtingų sričių ontologijoms susieti. ŠM srities struktūrinis terminų modelis parengtas taip, kad kiekviena ontologijos sąvoka būtų aprašyta unikaliu savybių ir semantinių ryšių rinkiniu – toks sprendimas padeda spręsti terminų daugiareikšmiškumo problemą. Šis išteklius sukurtas ne tik kaip ŠM srities terminų taksonomija, bet ir kaip vienas iš galimų semantinės paieškos ir natūraliosios kalbos apdorojimo priemonių.

⁹ Daugiau informacijos: <http://protege.stanford.edu>.

Iki šiol daug dėmesio skirta tam, kaip ontologijos gali būti pritaikytos kompiuterizuojant natūraliąją kalbą, ją sisteminant, padarant suprantamą kompiuterinėms programoms. Galbūt iš pirmo žvilgsnio ne visai aišku, kuo ontologijos gali būti pravarčios lingvistams ar kitų sričių specialistams. Reikia pabrėžti, kad ontologijos ypač naudingos terminologams, nes, pasinaudojus ontologijos vizualizavimo priemonėmis, aiškiai matyti terminų tarpusavio ryšiai. Terminams iliustruoti galima parinkti ontologijoje pateiktus su tais terminais susijusius konkrečius žodžius (ontologijoje vadinamus *esiniais*).

Tiek ŠM terminų žodynas, tiek ontologija gali padėti ŠM srities dalyviams: dėstytojams, mokytojams, administracijos darbuotojams, studentams, moksleiviams, nes šiuose šaltiniuose matyti, kaip, pvz., klasifikuojami studentai, kokių rūšių būna studijų, iš ko sudaryta aukštųjų mokyklų administracija ir pan.

APIBENDRINAMOSIOS PASTABOS

Straipsnyje pristatytas *Švietimo ir mokslo terminų žodynas* ir *Švietimo ir mokslo terminų ontologija*. Tiek žodynas, tiek ontologija yra tobulinti. Didžiausias abiejų šaltinių privalumas ir išskirtinumas yra tai, kad jie parengti naudojant lietuvių terminologų dar menkai taikomus tekstynų ir kompiuterinės lingvistikos principus.

ŠM terminų žodynas yra prieinamas internete, jį sudaro 224 nuo vieno iki penkių žodžių sudaryti terminai. Nuo kitų terminų žodynų šis skiriasi tuo, kad be įprastinių terminų žodyno dalių (pačių terminų, apibrėžčių, atitikmenų užsienio kalbomis) jame pateiktos visos to termino kaitybinės formos su dažnumais, pavyzdžiai iš ŠM tekstyno, kolokatai ir konkordansas. Šie duomenys svarbūs tuo, kad žodyno naudotojai gali susidaryti išsamų vaizdą apie tam tikro termino vartoseną, kontekstą; gali rasti apibrėžtyje galbūt nepaminėtus svarbius termino požymius.

Internetu prieinama ŠM terminų ontologija atskleidžia, kokiais semantiniiais ryšiais susiję ŠM terminai, padaro šią informaciją suprantamą kompiuterinėms programoms. Ontologijos gali ypač praversti terminologams, klasifikuojantiems, aprašantiems tam tikros srities terminus, pateikiantiems konkrečių pavyzdžių.

Tikimasi, kad tiek ŠM terminų žodynas, tiek ontologija bus metodologiškai svarbūs resursai ir paskatins Lietuvos terminologus labiau pasinaudoti tekstynų teikiamomis galimybėmis tirti, aprašyti, klasifikuoti terminus.

LITERATŪRA

- Baader F. 2003: Appendix: description logic terminology. – *The description logic handbook: theory, implementation, and applications*, Cambridge: Cambridge university press.
- Bielinskienė A., Boizou L., Kovalevskaitė J., Utka A. 2012: Towards the Automatic Extraction of Term-defining Contexts in Lithuanian. – *Human Language Technologies. The Baltic Perspective: Proceedings of the Fifth International Conference Baltic HLT 2012*, Amsterdam, Berlin, Tokyo, Washington, DC: IOS Press, 18–26.
- Cimiano P. 2006: *Ontology Learning and Population from Text: Algorithms, Evaluation and Applications*, Karlsruhe: Springer.
- Corcho O., Fernández-López M., Gómez-Pérez A. 2003: Methodologies, tools and languages for building ontologies. Where is their meeting point? – *Data & knowledge engineering* 46(1), 41–64.
- Gomez-Perez A., Corcho O., Fernandez-Lopez M. 2004: *Ontological Engineering: with examples from the areas of Knowledge Management, e-Commerce and the Semantic Web*, London: Springer-Verlag.
- Gornostay T., Ramm A., Heid U., Morin E., Harastani R., Planas E. 2012: Terminology Extracion from Comparable Corpora for Latvian. – *Human Language Technologies. The Baltic Perspective: Proceedings of the Fifth International Conference Baltic HLT 2012*, Amsterdam, Berlin, Tokyo, Washington, DC: IOS Press, 66–73.
- Grigonytė G. 2010: *Building and Evaluating Domain Ontologies: NLP Contributions*, Logos Verlag Berlin GmbH.
- Grigonytė G., Rimkutė E., Utka A., Boizou L. 2011: Experiments on Lithuanian Term Extraction. – *Proceedings of the 18th Nordic Conference of Computational Linguistics NODALIDA 2011. NEALT Proceedings Series* 11, 82–89.
- ISO 10241: 1992 *International terminology standards – Preparation and layout*, Geneva: ISO.
- ISO 9001: 2000 *Quality management systems – Requirements*, Geneva: ISO.
- Krugļevskis V., Vancāne I. 2005: Term Extraction from Legal Texts in Latvian. – *Proceedings of the Second Baltic Conference Human Language Technologies*, 155–159.
- LST ISO 1087-1: 2005 *Terminoloģijas darbas. Aiškinamasis žodynas. 1 dalis. Teorija ir taikymas* (tpd ISO 1087-1: 2000).
- Mariano F., Gomez-Perez A., Juristo N. 2004: Methontology: From Ontological Art Towards Ontological Engineering. – *AAAI-97 Spring Symposium. Series on Ontological Engineering*, Stanford: AAAI Press, 33–40.
- Mayr E. 1987: The ontological status of species: scientific progress and philosophical terminology. – *Biology and Philosophy* 2(2), 145–166.
- Rimkutė E. 2010: Lietuvių kalbos terminų automatinio nustatymo galimybės. – *Kalbų studijos* 16, 71–78.
- Rimkutė E. 2012: Automatinis švietimo ir mokslo terminų nustatymas lingvistiniais metodais. – *Terminologija* 19, 54–69.
- Staab S., Studer R., Schnurr H. P. 2001: Knowledge Processes and Ontologies. – *Intelligent Systems* 16(1), 26–34.
- Uschold M., Gruninger M. 1996: Ontologies: Principles, methods and applications. – *Knowledge engineering review* 11(2), 93–136.
- Uschold M., Gruninger M. 2004: Ontologies and semantics for seamless connectivity. – *ACM SIGMod Record* 33(4), 58–64.
- W3C – OWL Web Ontology Language Use Cases and Requirements. Prieiga internete <http://www.w3.org/TR/2002/WD-webont-req-20020307/> (žiūrėta 2013-10-14).

ONTOLOGY AND DICTIONARY OF SCIENCE AND EDUCATION TERMS

The paper presents the dictionary and the ontology of science and education terms. The proposed approach of ontology and dictionary building uses the following Natural Language Processing techniques: linguistic pattern recognition, collocation extraction, morphological analysis, POS tagging, and others.

The dictionary of science and education is available online and contains 224 one word and multiword terms (their length is up to 5 words). The dictionary contains the

necessary dictionary fields (terms, definitions, translations to other languages), as well as some additional fields such as declined forms of terms with their frequencies, examples of term usage, collocations and concordances from the special science and education corpus.

The main focus of the paper is the ontology of terms in the domain of science and education. Authors present, how different research areas define ontologies, what are the possibilities of their use, what are ontology building methods. The structure of the ontology is illustrated by multiple examples. The hierarchical structure of concepts, their properties and relationships make this information readable and understandable for knowledge-based information systems. Ontologies can be particularly useful for terminology experts, who try to classify and describe terms, concepts and their instances in a domain specific field.

Both resources can be valuable for lecturers, teachers, administration management bodies, students, etc. While using the ontology, they can define and conceptualize the processes of studying, teaching or learning.

The paper introduces the possibilities of semi-supervised term extraction with mixed prescriptive and descriptive term analysis methods, which can be useful for terminology experts, lexicographers, and creators of lexical databases.

Gauta 2013-10-29

Irena Markievicz
Vytauto Didžiojo universitetas
K. Donelaičio g. 52, LT-44248 Kaunas
E. paštas i.markievicz@if.vdu.lt

Erika Rimkutė
Vytauto Didžiojo universitetas
K. Donelaičio g. 52, LT-44248 Kaunas
E. paštas e.rimkute@hmf.vdu.lt